POLYPUBLIE
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Trois étapes sensibles du travail de géologue de ressources
Auteur: Author:	Raphaël Dutaut
Date:	2021
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Dutaut, R. (2021). Trois étapes sensibles du travail de géologue de ressources [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5612/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/5612/
Directeurs de recherche: Advisors:	Denis Marcotte
Programme: Program:	Génie minéral

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

TROIS ÉTAPES SENSIBLES DU TRAVAIL DE GÉOLOGUE DE RESSOURCES

RAPHAËL DUTAUT
Département de Département des génies civil, géologique et des mines

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie minéral

MARS 2021

© Raphaël Dutaut, 2021.

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

TROIS ÉTAPES SENSIBLES DU TRAVAIL DE GÉOLOGUE DE RESSOURCES

présentée par **Raphaël DUTAUT**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Richard SIMON, président
Denis MARCOTTE, membre et directeur de recherche
Michel GAMACHE, membre
Serge SÉGURET, membre

DÉDICACE

À ma femme et mes enfants.

REMERCIEMENTS

Lorsque j'ai commencé comme étudiant en Ph.D, je travaillais déjà dans l'industrie comme personne qualifié en ressources depuis plusieurs années. C'est en travaillant sur un dossier professionnel avec M. Dominique Francois-Bongarcon que j'ai pour la première fois été initié aux géostatistiques avancées. A son contact, je me suis rendu compte de mon ignorance dans ce domaine. Quelques mois plus tard, j'ai eu la chance de pouvoir travailler avec M. Harry Parker, qui, amusé par ma curiosité, m'a suggéré mes premières lectures d'articles géostatistiques. Ce sont ces deux grands noms de la géostatistique minière, qui m'ont recommandé d'approfondir mes connaissances en géostatistiques et m'ont aiguillé vers un retour sur les bancs d'école. Je leur en suis profondément reconnaissant. C'est d'ailleurs M. Harry Parker qui m'a suggéré et orienté sur le sujet de mon premier article (présenté au chapitre 5).

Une fois la décision de faire un Ph.D prise, je me suis mis à la recherche d'un directeur de recherche pour m'accompagner sur ce chemin. J'ai alors rencontré celui qui devait devenir mon mentor, le professeur Denis Marcotte. Ce dernier, a compris ma démarche et a décidé de soutenir mon projet de retour aux études. *Denis* m'a évidemment beaucoup appris, il a toujours patiemment pris le temps de corriger mes nombreuses erreurs de *scripting* dans Matlab, et d'enrichir mes travaux de recherche. Les plus simples des mots portent souvent le plus de poids : merci *Denis*.

Durant les 6 années qu'a duré mon doctorat, j'ai eu un 2e enfant, déménagé à plusieurs reprises et progressé de quelques échelons dans ma carrière professionnelle. Tout ceci n'aurait jamais été possible sans ma moitié. Cette femme extraordinaire que j'ai la chance d'avoir épousé, m'a toujours encouragé dans cette aventure doctorale. Bien souvent elle a dû s'occuper seule de nos enfants, de la maison et de tout le reste, pendant que je travaillais sur mes recherches. Mais malgré cela, elle a toujours été là pour me soutenir dans les moments les plus difficiles, et il y en a eu beaucoup ! Sans elle, je n'en serais pas là dans ma vie : merci *Isabelle*.

RÉSUMÉ

En géologie minière, l'estimation des ressources par modèle de bloc est la norme. Afin de garantir des estimés de ressources de qualité, la personne qualifiée selon la norme canadienne NI43-101 s'appuie sur un ensemble de règles formelles et informelles : les *meilleures pratiques*.

Ces règles édictent notamment les approches à suivre concernant les recommandations de forages, l'utilisation d'une teneur d'écrêtage et la validation des ressources récupérables globales. Ces phases, consécutives, sont parmi les plus sensibles dans le processus d'estimé de ressources et peuvent induire des dépenses ou pertes pour des montants de plusieurs dizaines à centaines de millions de dollars et sont donc centraux pour l'avenir des projets ou mines en opérations. Cependant, les géologues de ressources n'utilisent pas, dans la grande majorité, des méthodes objectives et répétables pour s'assurer de l'efficacité ou de la qualité de leurs utilisations des *meilleures pratiques*.

Dans le cadre de cette thèse, nous proposons de nouvelles approches permettant de systématiser et de rationaliser les recommandations de forages et la définition des seuils d'écrêtage. Nous présentons aussi une étude afin de renforcer et documenter les meilleurs pratiques de validations de ressources récupérables globales par changement de support en fonction du type de variogramme.

Tout estimé de ressources commence par une banque de données de forages et se termine par une recommandation de forages additionnels. Une des problématiques concernant ces recommandations est qu'elles s'effectuent généralement de manière manuelle et donc subjectivement. Les campagnes de forages représentent des investissements pouvant aller jusqu'à plusieurs dizaines de millions de dollars et sont un enjeu significatif pour toutes les compagnies minières ou d'explorations. Pour ce problème de positionnement de forage, nous proposons de rationaliser les recommandations de forages grâce un nouvel outil de la recherche opérationnelle qui pour la première fois permet de prendre en compte de manière dynamique la couverture continue des blocs par les forages existants ou recommandés. L'algorithme est flexible tout en obtenant une efficacité proche de l'optimum obtenu par programmation linéaire. Il fonctionne pour des données de mines souterraines ou à ciel ouvert. Utilisé sur des données issues d'une mine en opération, l'algorithme a démontré qu'il était suffisamment rapide pour permettre l'optimisation de plusieurs projets simultanément. Cet avantage est particulièrement utile pour les exercices budgétaires, car il permet une rationalisation du gain des forages recommandés. L'impact de la fonction objectif utilisée sur plusieurs projets simultanément reste cependant à établir. Flexible et rapide, ce nouvel outil permet d'obtenir

des recommandations de forages objectives garantissant une couverture maximale en fonction du budget alloué.

Une fois les forages réalisés, l'analyse des carottes permet de commencer l'estimé des ressources. Un problème récurrent est alors la grande variabilité des analyses et la présence de valeurs extrêmes qui peuvent influer grandement sur l'estimation des ressources. D'importantes quantités de métal et donc de revenus potentiels étant associés à ces analyses, la gestion des valeurs extrêmes et la définition des seuils d'écrêtage associés sont des problèmes majeurs pour tous géologues de ressources. La pratique la plus courante consiste à interpréter visuellement l'histogramme des teneurs afin de définir le seuil d'écrêtage. Nous proposons une nouvelle approche qui utilise les données du contrôle de la qualité et assurance de la qualité (QaQc) afin d'établir un lien entre qualité d'échantillonnage et valeurs extrêmes. Le seuil d'écrêtage est sélectionné tel que la population écrêtée retourne le coefficient de variation estimé de la population sans erreur d'échantillonnage (établie grâce au coefficient de corrélation des duplicats). En utilisant des données issues de mines aurifère et nickélifère, la méthode a démontré des résultats cohérents avec les approches généralement utilisées dans l'industrie. L'utilisation d'un estimateur robuste du coefficient de corrélation des duplicats représente une amélioration possible. Grâce à cette approche originale qui valorise les données de QaQc, un nouvel outil de définition des valeurs d'écrêtage objectif et répétable est fourni aux géologues pour les assister dans ces décisions critiques.

Une fois les données écrêtées, la personne qualifiée peut passer à la phase d'interpolation du modèle de bloc. Le plus souvent, dans l'industrie minière canadienne, un estimateur linéaire est utilisé. Afin de choisir les paramètres d'interpolation, la validation des estimés de ressources récupérables par changement de support global est la méthode objective la plus utilisée. Pour obtenir les courbes tonnages-teneurs à la taille de support désirée, il est nécessaire d'obtenir des données sans valeur extrême afin d'en déduire : l'histogramme et le variogramme. Il existe quatre types de variogrammes fréquemment utilisés dans l'industrie minière (traditionnel, corrélogramme, relatif par paires et sur données gaussiennes). Le géologue-géostatisticien doit, par la suite, choisir le modèle de changement de support proprement dit. Afin de documenter les conséquences de ces choix, nous avons étudié l'impact des différents types de variogrammes sur les résultats de plusieurs modèles de changement de support. Grâce à des données simulées, nous avons corroboré les meilleures pratiques utilisées dans l'industrie. Il en ressort que le modèle discret gaussien est le plus robuste et ne présente pas de biais comparativement aux méthodes indirectes log normal qui exposent certains projets à plus de risques. Aussi, le variogramme relatif par paires ne semble pas fournir une bonne évaluation de la dispersion de blocs et doit donc être évité pour l'utilisation de changement de support. Les techniques de changement de support globales actuellement

disponibles sont basées sur l'hypothèse de stationnarité ; or, pour de nombreux projets miniers, cette hypothèse est faible. Le développement d'une méthode de changement de support globale non-stationnaire représente une piste intéressante pour des recherches futures. Cette étude permet d'identifier les *meilleures pratiques* en matière de validation des estimés de ressources récupérables.

Dans cette thèse, nous avons proposé plusieurs solutions originales afin d'apporter des améliorations aux *meilleures pratiques* utilisées dans l'industrie minière canadienne pour l'estimation de ressources. Grâce à des données simulées et réelles issues de projets ou mines en opérations, nous avons testé les algorithmes et méthodes proposées. Les résultats de nos études corroborent les pratiques actuelles tout en y apportant des améliorations robustes et objectives. L'utilisation et le respect des méthodes développées dans cette thèse devraient permettre d'améliorer les *meilleures pratiques*, de maximiser la valeur des projets ou mines en opérations et donc de diminuer le risque associé aux estimés de ressources.

ABSTRACT

In mining geology, block models are used for resource estimations. To guarantee quality of the resources estimates in accordance to the Canadian' standard NI43-101, the qualified person relies on a set of formal and informal rules : the *best practices*.

These rules lay down the approaches to be followed for significant subjects such as drilling recommendations, high grade capping and overall validation of recoverable resources. These successive steps are among the most important subjects during the resource estimate process. They can induce multi-millions additional expenses or losses up to hundreds of millions of dollars: these steps are therefore of central importance for projects or mines in operations. However, the vast majority of resource geologists do not use objective and repeatable methods to ensure the effectiveness or quality of their applications of the *best practices*.

In this thesis, we propose new approaches to systematize and rationalize the drilling recommendations and the definitions of capping thresholds. We also present a study to document and reinforce some of the textit best practices regarding the validation of the global recoverable resources using change of support methods according to the variogram type.

Any resource estimate begins with a drill hole database and ends with drill holes' recommendation. These recommendations are generally carried out manually and are therefore subjective. Drilling campaigns represent investments up to tens of millions of dollars and are a significant challenge for all mining or exploration companies. To this drilling positioning problem, we propose to rationalize the drilling recommendations with a new operational research tool. Flexible, it works for both underground and surface data sets. The algorithm dynamically takes into account the partial coverage of the blocks of existing or recommended drill hole. The algorithm is fast even for large data sets with an optimality gap minimum when compared to linear programming. Applied to an operating mine data, the algorithm allows the optimization of several projects simultaneously. This advantage is particularly useful companies' budgets definition, as it allows a rationalization of the different recommended drilling programs. In application on several projects simultaneously, the impact of the objective function parameters used remains to be established. Flexible and fast, this new tool makes possible to obtain objective drilling recommendations which guarantee maximum coverage according to the allocated budget.

Once the drilling campaign is completed, core assays allow the resource estimate to begin. A recurring challenge is then grades variability management and the presence of extreme values. Since a large quantity of metal and therefore potential income is associated with

extreme values, their management through capping threshold definitions is a major problem for all resource geologists. The most common practice consists of visually interpret assays histogram to define capping threshold. Thought the use of quality analysis quality control (QaQc) data, we propose a new approach linking sampling quality and extreme values. The capping threshold is selected such that the coefficient of variation of the capped assays is equal to the one estimated forward error-free population (established from the correlation coefficient of duplicates). Using two real data sets from gold and nickel mines, the method demonstrates results consistent with existing methods used in industry. The use of a robust estimator for the correlation coefficient of duplicates represents a possible improvement to the proposed methods. This original approach is objective and repeatable, it enhances the data from QaQc and link the capping threshold to the sampling quality. The proposed method provides a new tool to assist geologists in this critical decision.

Once data has been properly capped, the qualified person can proceed with block model interpolation. Most often in the Canadian mining industry a linear estimator is used. The interpolation parameters choice is often done thought the validation of the recoverable resource estimates by global change of support. The grade-tonnage curves at the desired block size are deduced from the data sets, without an extreme value, providing: the histogram and the variogram. There are four different variograms types commonly used in the mining industry (traditional, correlogram, pairwise and nscore). The geologist-geostatistician must then choose a change of support model. In order to document the consequences of these choices, we studied the impact of the different variograms types on the results of several change of support models. Using simulated data sets, we have corroborated some of the *best practices* used in the industry. We demonstrated the superiority of the Discrete Gaussian Model is the compared to the indirect log normal methods.

The Discrete Gaussian Model is more robust and does not present any bias for all tested grade distributions, as such, it is the recommended model to be used to reduce projects risks. The pairwise relative variogram do not provide a good estimate of block dispersion for change of support models and should therefore be avoided in this context. Existing global change of support models are based on the stationary assumption; however, in most mining projects, this assumption is weak. Therefore, the development of a non-stationary global change of support model represents an interesting direction for future research. This study helps the qualified persons to identify the approaches to be avoided and the ones recommended as *best practices* to be used for the validation of the resource estimates by global change of support methods.

In this thesis, we proposed several original solutions as improvements to the Canadian' re-

source estimation *best practices*. Using both simulated and real data sets from projects and operating mines, we assessed the proposed algorithms and methods. The results of our studies corroborate current practices while providing robust and objective improvements. The use and respect of the methods developed in this thesis improve the *best practices*, reducing the risk associated with resource estimates and maximizing the value of exploration projects and operating mines.

TABLE DES MATIÈRES

DÉDICACE iii

REMERCIEMENTS iv

RÉSUMÉ v

ABSTRACT viii

TABLE DES MATIÈRES xi

LISTE DES TABLEAUX xiv

LISTE DES FIGURES xv

LISTE DES SIGLES ET ABRÉVIATIONS xvii

CHAPITRE 1 INTRODUCTION 1
- 1.1 Éléments de la problématique 1
- 1.2 Définitions et concepts de base 4
  - 1.2.1 Optimisation des forages 4
  - 1.2.2 Seuil d'ecrétage 5
  - 1.2.3 Estimation globale et changement de support 6
- 1.3 Objectifs de recherche 7
  - 1.3.1 Objectifs généraux 7
  - 1.3.2 Objectifs spécifiques 7
- 1.4 Plan de la thèse 8

CHAPITRE 2 REVUE DE LITTÉRATURE ET MÉTHODES 10
- 2.1 Revue des *meilleures pratiques* trouvées dans des rapports NI43-101 10
- 2.2 Les recommandations de forages objectives 16
- 2.3 Les techniques d'écrêtage et les erreurs d'échantillonnage 19
  - 2.3.1 Les techniques d'écrêtage commune dans l'industrie minière 20
  - 2.3.2 L'erreur d'échantillonnage et lien avec les duplicatas 21
- 2.4 La géostatistique 23
  - 2.4.1 Le variogramme 23
  - 2.4.2 Géostatistique linéaire 28

2.4.3 Géostatistique non linéaire 28
2.5 Discussion 31

CHAPITRE 3 ARTICLE 1 : A NEW SEMI-GREEDY APPROACH TO ENHANCE DRILLHOLE PLANNING 33
3.1 Abstract 33
3.2 Keywords : 33
3.3 Introduction 33
3.4 Methodology 35
3.4.1 Objective Function 35
3.4.2 Semi-greedy Optimization 37
3.4.3 Semi-greedy algorithm 38
3.5 Case Study 40
3.5.1 Datasets Presentations 40
3.5.2 2D Application 40
3.5.3 Influence of dmin for Step or Linear Weight Functions 46
3.5.4 3D Application 47
3.5.5 Simultaneous optimization 51
3.6 Discussion and Conclusions 52
3.7 Aknowledgments 54

CHAPITRE 4 ARTICLE 2 : A NEW GRADE CAPPING APPROACH BASED ON COARSE DUPLICATES CORRELATION 55
4.1 Abstract 55
4.2 Keywords : 55
4.3 Introduction 55
4.4 Methodology 58
4.4.1 Theoretical background 58
4.4.2 Correlation between Duplicates 59
4.4.3 Capping based on correlation between duplicates 61
4.4.4 Effects of variable assay supports 62
4.4.5 Effects of departing from the lognormal assumption 63
4.4.6 Capping after or before compositing ? 66
4.4.7 Workflow 67
4.5 Case study 67
4.5.1 Gold case 67
4.5.2 Nickel case 69

4.6 Discussion 70
4.7 Conclusions 71

CHAPITRE 5 ARTICLE 3 : A COMPARISON OF INDIRECT LOGNORMAL AND DISCRETE GAUSSIAN CHANGE OF SUPPORTS METHODS FOR VARIOUS VARIOGRAM ESTIMATORS 72
5.1 Abstract 72
5.2 Keywords : 72
5.3 Introduction 73
5.4 Methods 75
5.4.1 Simulation and sampling method 75
5.4.2 The variogram estimators 76
5.4.3 Change of support models 77
5.5 Results 79
5.5.1 Change of support – Sensitivity to variogram type. 79
5.5.2 Exhaustive data sets 80
5.5.3 Grade tonnage curves for the different distributions 83
5.6 Change of support – Sensitivity to SMU size 85
5.7 Discussion 86
5.8 Conclusions 88

CHAPITRE 6 DISCUSSION GÉNÉRALE 89

CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS 92

RÉFÉRENCES 93

LISTE DES TABLEAUX

3.1 Block values basic statistics for the 2D dataset 41

3.2 Coverages for five drillholes as a function of *dmin* for the step and linear weight functions 47

3.3 Block values basic statistics for the 3D dataset 48

4.1 Sampled intervals and definitions for parameters of the lognormal distributions of Z_v and t 61

4.2 Cases considered to evaluate the robustness of the approach 64

4.3 Gold case cap results - The target CV (Eq. 4.8) is 1.76 68

4.4 Nickel case cap results - The target CV (Eq. 4.8) is 0.783 70

5.1 Variogram model (Gaussian) 79

5.2 Target population statistics 79

5.3 Mean dispersion variance for SMU size of 10x10 units 80

5.4 Mean Relative Unsigned Errors – 10x10 SMU size - 50 nodes data spacing ; Errors on : Volumes / Grades / Conventional Profit 85

LISTE DES FIGURES

2.1 Compilation des informations de bases de 40 rapports NI43-101 : informations générales 11

2.2 Compilation des informations de bases de 40 rapports NI43-101 : informations géostatistiques 15

3.1 Weighting scheme theoretical example. Upper row presents w_1 (left) and w_2 (right) for the point (50, 25) (black dot) accounting for existing or already selected drillholes (two red dots). Bottom row shows the updated weights for point (50, 75) once point (50, 25) has been selected. 37

3.2 Longitudinal 2D results view. Black dots correspond to existing drillholes with ore intersects. Red dots are the resulting drillholes from the optimization. BV considering only the existing drillholes is illustrated. 42

3.3 Coverage as a function of number of drillholes. Semi-greedy with *ntrial* = 100 and *nlist* = 50. 43

3.4 (a) Coverage for 20 drillholes as a function of *nlist*, using *ntrial* = 300. (b) Optimization coverage results for 1–200 drillholes for different *nlist* with *ntrial* = 300. 45

3.5 Coverage results for varying *ntrial* and number of drillholes, with *nlist* = 50. 46

3.6 Longitudinal 2D results for five drillholes and increasing dmin with step (a) and linear (b) weight functions. Existing drillholes (black dots) and drillholes selected (red dots) by semi-greedy algorithm with *nlist* = 50 and *ntrial* = 1000. Map of BVs considering only the existing drillholes is illustrated. 47

3.7 (a) A total of 178 existing drillholes composites (black dots) and blocks with $BV > 0.1$ (blue squares), within the selected area. (b) Trace of the 3,672 drillholes that fulfill constraints on drillpad locations (green dots), direction and dips. 49

3.8 Topography (green), existing drillholes (black), 10 selected new drillholes (red) and blocks with $OF > 0.1$ (blue). 50

3.9 3D coverage as a function of number of drillholes. 51

3.10 Coverage as a function of drilling budget (in meters) for the two datasets optimized simultaneously 52

4.1 (a) Theoretical (red line, Eq. 4.7) and simulated ($n = 1,000,000$ duplicates with 200 blue dots) ρ_{dup} against σ for β^2=1. (b) Theoretical ρ_{dup} against σ for various β values. 60
4.2 Theoretical CV and predicted CV using Eq. 4.8 for different lognormal parameters and an increasing number of duplicates (n=100, 200, 1,000, 10,000). 61
4.3 Percentile threshold against ρ_{dup} to have $CV^2_{cap} = CV^2_v$ for various β values 62
4.4 Duplicates' scattergrams (left column), error-free histograms (middle) and observed values histograms (right column) for cases 1 (top) to 5 (bottom) described in Table 4.2 65
4.5 Effect of capping before or after compositing. Bias on the mean against CV for 3 m composites using assays on 0.5 m, 1.5 m and 3 m ; $\alpha = 0$, $\beta = 1$, $\sigma = 0.8$; Target CV obtained from error-free composites. . . . 67
4.6 Gold duplicates' scattergram (left) and log probability plot for duplicates 1 and 2 (right) 68
4.7 Nickel duplicates' scattergram (left) and log probability plot for duplicates 1 and 2 (right) 69
5.1 COS results at a 10x10 SMU size for simulated lognormal, Bimodal and Negatively skewed distribution using exhaustive dataset : error boxplot for DGM (Blue), Traditional IndLog (Red), Emery's IndLog (Green) for each variogram estimator ; low cut-off (left column), high cut-off (right column) ; error on grade in grade units (first row), error on volume in % (second row), error on Conventional Profit in % (third row). 82
5.2 True Grade-Tonnage curves (black line) compare to COS results in the case of lognormal (a), Bimodal (b) and Negatively skewed (c) distribution sampled on regular grids of size 50 (first row) and 10 (second row) ; traditional variogram estimator used for fitting model. 84
5.3 Cumulative absolute error as a function of the ratio (block size – variogram range) for both COS models : DGM(bleu), traditional IndLog(red), Emery IndLog (green). Lognormal (a), Bimodal (b) in the middle and Negatively skewed (c) distribution on the last row. 86

LISTE DES SIGLES ET ABRÉVIATIONS

COS	Changement de support (*Change Of Support*)
CV	Coefficient de variation
DDH	Forage aux diamants (*Diamands Drill Hole*)
DG - MDG	Modèle discret gaussien
EDA	Analyses des données (*Exploratory Data Analysis*)
FFT-MA	Simulation par transformée de Fourier (*Fast Fouier Transformation - Moving Average*)
ICM	Institue canadien des mines
ID - IDw	Inverse de la distance
IL - IndLog	Méthode indirecte log normal
KNA	Analyse du voisinage du krigeage (*Kriging Neighbourhood Analysis*)
LUC	Conditionnement uniforme localisé (*Localised Uniform Conditioning*)
MIK	Krigeage des indicatrices multiples (*Multiple Indicator Kriging*)
NI43-101	Instrumentation nationale canadienne 43-101 (*Canadian National Instrument 43-101*)
PQ (QP)	Personne qualifiée (*qualified person*)
QaQc	Contrôle de la qualité et assurance de la qualité (*Quality Analysis Quality Control*)
KO - OK	Krigeage ordinaire
SMU	Taille des blocs (*Selective Mining Unite*)
SRF	Fonction aléatoire d'ordre deux (*Second order Random Function*)
TSX	Bourse de Toronto (*Toronto Stock Exchange*)
UC	Conditionnement uniforme (*Uniform Conditioning*)

κ	Teneur cible (dans la formule de Gy)
γ	Variogramme
ϕ	Fonction d'anamorphose
ρ	Coefficient de corrélation
σ_z^2	Variance des teneurs Z
a	Portée ou distance de corrélation du modèle de variogramme
b_κ^2	Variance relative de l'erreur fondamentale à la teneur κ
c	Seuil d'indicatrice
C	Fonction ou palier de covariance

$g(u)$	Fonction de densité gaussienne
h	Translation de distance h
$I(x_i; c)$	Indicatrice codée pour la valeur x_i par rapport au seuil c
m_z	Moyenne de Z
$N(h)$	Nombre de paires disponibles à une distance h
$N(\mu, \sigma)$	Distribution suivant une loi normale de moyenne μ et d'écart type σ
r	Coefficient de corrélation du changement de support
s_r^2	Erreur fondamentale (dans la formule de Gy)
t	Facteur d'erreur aléatoire multiplicatif
v	Taille du support
Z	Teneur d'une variable aléatoire

CHAPITRE 1 INTRODUCTION

La géologie minière a pour vocation de maximiser la valeur actuelle ou éventuelle des projets et des opérations minières. Pour ce faire, le géologue minier se doit de maîtriser une vaste gamme d'outils offerts par la géostatistique et la recherche opérationnelle.

1.1 Éléments de la problématique

Les projets miniers représentent des investissements pouvant aller jusqu'à plusieurs milliards de dollars canadien. Toutes les phases de planifications (minière ou financière) utilisent le modèle de bloc comme fondation. Une erreur aussi minime soit-elle dans l'estimé de ressources peut avoir des conséquences économiques catastrophiques. Parker [1] estime qu'une erreur de 10% sur la teneur, peut impacter la valeur nette d'un projet jusqu'à 60%. Afin de limiter les erreurs, standardiser les rapports et pratiques et donc contrôler le risque, les réglementations canadienne (NI43-101), australienne (JORC) ou sud-africaine (SAMREC) offrent un cadre qui protège le public et les compagnies rapportant des chiffres de ressources. Les géologues de ressources et les personnes qualifiées (PQ ou *qualified person* (QP) en anglais) appliquent alors un ensemble de règles formelles telles que les définitions de l'institut canadien des mines [2] pour la rédaction du rapport et d'autres règles informelles appelées les *meilleures pratiques* pour les choix des paramètres techniques. Ces « règles » traitent de tous les sujets abordés dans le processus de ressources : validations des données historiques, des méthodes de forages et transport des échantillons, des méthodes d'échantillonnages, pratiques sur le contrôle de la qualité et assurance de la qualité (QaQc), gestions des banques de données numériques, interprétation et modélisation géologique, création des domaines, analyses des données (EDA) et détermination du variogramme, taille des blocs (SMU : *selective mining unit*) et paramètres d'interpolation, type de validation des modèles, classification, réconciliations, recommandations de forages subséquents, etc... Ces « règles » couvrent aussi les explications à fournir dans les différents chapitres des rapports de type NI43-101 ou lors d'un processus de contrôle des teneurs en opérations. Bien qu'il existe plusieurs ouvrages de références abordant plus ou moins directement la géostatistique minière ([3], [4], [5], [6], [7] pour ne citer que quelques livres récents), il n'existe malheureusement aucune documentation minière *officielle* (CIM, JORC ou SAMREC) validant explicitement les *meilleures pratiques*, et le géologue se retrouve souvent sans critères objectifs pour le guider dans ses décisions techniques.

Tout exercice de ressources commence par une banque de données de forages, et se termine le

plus souvent par la recommandation d'une nouvelle campagne de forages. Le projet commence à un niveau de connaissances minimal et de risques maximaux correspondant à des ressources supposées ou à du potentiel géologique dans la norme NI43-101. Si le projet démontre un potentiel intéressant, l'analyste fait alors un ensemble de recommandations permettant de réduire le risque et autorisant le projet à passer aux étapes suivantes de ressources indiquées puis mesurées. Parmi ces recommandations, l'ajout de nouvelles informations sous la forme de forages est la plus commune. Plus tard, une fois le projet en opération, il est souvent nécessaire d'ajouter des forages dans les zones présentant les plus forts risques, ou dans le but de maximiser le processus de contrôle de teneur. Le plus souvent, le géologue identifie visuellement les zones qui, pour lui, représentent le plus haut risque. Il place alors manuellement un forage à l'endroit qui lui semble le plus adéquat. L'ensemble des forages ainsi proposés devient la recommandation de forage. Parfois un exercice de revue par les pairs permet de consolider et de diminuer la subjectivité de l'exercice, mais là encore, il s'agit d'une évaluation déterministe et non objective des besoins en forages. Les coûts de forages varient généralement entre quelques dizaines à plusieurs centaines de dollars par mètre. Ils dépendent du type de forages (aux diamants ou à circulation inverse), du type de sol (profile latéritique ou roche dure), de la localisation (en surface ou sous-terre), de l'accessibilité (proche de centre urbain ou à plusieurs centaines de kilomètres d'une route) et enfin de la longueur et du nombre de forages de la campagne. Dans tous les cas, les forages représentent un investissement significatif, souvent de plusieurs millions de dollars, permettant aux compagnies de réduire le risque sur leur ressources et/ou sur leur production. Les recommandations de campagnes de forages représentent donc l'une des tâches les plus significatives du géologue de ressources. Les recommandations de forages sont capitales dans la vie de tous les projets et représentent le premier objectif spécifique de cette thèse.

Une fois les campagnes de forages terminées, les données analytiques sont importées dans des banques de données et le processus de modélisation géologique peut commencer. Des domaines sont alors interprétés sur des critères géologiques ou économiques, de manière déterministe ou implicite ([8], [9]). Une fois les domaines définis, l'étude géostatistique peut avoir lieu. L'une des première étapes, l'analyse de données (EDA), consiste à considérer le traitement des valeurs extrêmes. La méthode la plus communément utilisée par les PQ pour gérer ce risque, consiste à sélectionner les valeurs extrêmes et à les réduire à une valeur plus faible [10]. La définition du seuil d'écrêtage est une étape cruciale dans la majorité des interpolations de modèle de blocs. En effet, dans les gisements à fort coefficient d'asymétrie, telle que la majorité des gisements de métaux précieux (or, argent), une forte proportion du métal est contenue dans la queue de la distribution. Ce métal impacte souvent significativement la valeur économique d'un projet et représente donc un enjeu significatif dans la

plupart des estimés de ressources. Dans la réglementation NI43-101, à travers les définitions posées par l'institut canadien des mines (ICM), le traitement des teneurs extrêmes doit être explicitement présenté et justifié [2] par la personne qualifiée. Les méthodes utilisées sont généralement empiriques, basées sur l'interprétation visuelle et souvent appuyée sur l'expérience de l'analyste, ce qui peut introduire un risque (positif ou négatif) sur la valeur économique d'un projet. De plus, les méthodes actuellement décrites dans la littérature ou utilisées dans les rapports techniques de type NI43-101, négligent les erreurs d'échantillonnages lors de la définition des seuils. Or intuitivement, pour la même distribution (et donc les mêmes valeurs extrêmes), la confiance associée aux très hautes valeurs d'un gisement présentant une faible erreur d'échantillonnage sera nettement supérieure à un gisement présentant une forte erreur d'échantillonnage. La création d'un outil d'évaluation du seuil d'écrêtage simple, répétable et objectif qui prend en compte l'erreur d'échantillonnage serait utile afin de réduire le risque associé aux valeurs extrêmes. Ceci constitue le deuxième objectif spécifique de ma thèse.

Une fois les données préparées et le traitement des données extrêmes effectué, s'en suit l'étude variographique. Pratiquement toutes les méthodes de la géostatistique nécessitent d'obtenir un modèle de la corrélation spatiale. Le variogramme est l'outil généralement privilégié par rapport à la fonction de covariance. En effet, il est plus général et il est défini pour toute fonction aléatoire intrinsèque, alors que la fonction de covariance nécessite une stationnarité d'ordre deux. De plus, le variogramme ne nécessite pas d'estimation de la moyenne, contrairement à la fonction de covariance. L'écrêtage des valeurs extrêmes permet généralement d'améliorer le comportement du variogramme expérimental et facilite son interprétation. Cependant, l'effet de pépite géologique [11], les erreurs d'échantillonnages et analytiques [12], les écarts de stationnarité ou hétérogénéités diverses [4] rendent parfois l'interprétation du variogramme expérimental difficile. Afin de pallier aux difficultés de calcul de la méthode classique d'estimé du variogramme expérimental, plusieurs estimateurs ont été proposés [4] et les plus courants sont présentés dans le chapitre suivant. Cependant, ces méthodes alternatives aboutissent à des modèles de variogrammes présentant des paramètres différents : effet de pépite, palier, portée. L'objectif de toutes ces méthodes est de permettre de modéliser la *vraie* corrélation spatiale, celle qui permettra aux méthodes géostatistiques de fournir des résultats optimums. Les modèles de variogrammes sont nécessaires lors de l'utilisation des méthodes de changement de support. L'effet de support décrit pour la première fois en mines par le Sud-Africain D. Krige [13] permet de faire le lien entre la variance des données et celle des blocs. Cette relation volume-variance permet d'inférer à l'avance la distribution de la variable étudiée pour toute sélectivité basée sur des petits blocs (SMU) : c'est le changement de support (COS). Là encore, plusieurs modèles mathématiques ont été proposés au fil des années et deux familles de méthodes sont principalement utilisées dans les mines : le

modèle discret gaussien [4] et le modèle *indirect log normal* [14]. Une pratique courante dans l'industrie consiste à utiliser les résultats du COS global pour *étalonner* les paramètres d'interpolation et mieux contrôler le risque sur ces paramètres centraux. Ceci nécessite donc un contrôle strict des paramètres et des méthodes utilisées pour le COS. Déterminer l'influence des différents types d'estimateurs de variogramme sur les deux méthodes principales de COS, le modèle discret gaussien (DG) et le modèle *indirect log normal* (IL), constitue le troisième objectif spécifique de cette thèse.

1.2 Définitions et concepts de base

Dans cette thèse, trois aspects récurrents et significatifs dans la vie d'un projet minier sont abordés : l'optimisation des forages, la définition des seuils d'écrêtage et l'estimation par changement de support global. Ces aspects ont été spécifiquement ciblés, car ils représentent de forts enjeux en termes économiques et que les méthodes ou approches actuellement utilisées laissent une forte place à la routine et à la subjectivité dans l'interprétation des résultats.

1.2.1 Optimisation des forages

L'une des principales tâches du géologue, consiste à assurer la gestion des campagnes de forages. Cette gestion peut se diviser en plusieurs sous-étapes : planification, déroulement de la campagne et conclusions. La conclusion d'une campagne de forage consiste généralement en une recommandation de forages additionnels avec un retour à l'étape de planification. Dans ce cas, les gestionnaires doivent alors décider de critères à atteindre pour la prochaine campagne. Il s'agit souvent d'ajouter des forages dans les zones considérées à risque ou afin de respecter une maille de forage définie. Ces ajouts sont faits afin de permettre la classification de ressources, suivant la norme NI43-101 par exemple, en Mesurées-Indiquées-Présumées (cette classification reconnue internationalement représente un degré de confiance décroissant dans les valeurs ou les paramètres d'un bloc [2]). Une distance de plus en plus petite entre les forages existants étant la méthode la plus communément utilisée. Cet exercice peut, dans le cas de certains projets, être chronophage, fastidieux et est surtout généralement subjectif. Afin de contrôler ces inconvénients, l'utilisation d'outils de la recherche opérationnelle [15] est une avenue développée depuis plusieurs années, mais encore rarement utilisée. Bilal fut un des premiers auteurs à approcher la problématique à travers la méthode du problème de recouvrement par ensemble [16]. Bilal a proposé une métaheuristiques pour résoudre ce problème d'optimisation classique.

Afin de pouvoir utiliser une méthode d'optimisation, il est nécessaire dans un premier temps

de définir une valeur à maximiser, c'est le rôle de la *fonction objectif*. Cette fonction peut être aussi simple qu'une distance entre les sondages existants et planifiés : une maille de forage systématique. Cependant, l'utilisation d'algorithmes puissants permet l'application de fonctions objectifs plus complexes prenant en compte la variance d'interpolation, des résultats de simulations, des paramètres géologiques ou encore économiques. Grâce à cette fonction, il est possible d'identifier les zones de risques maximaux et, pour un coût minimal par rapport à une grille systématique, de les forer. Il en résulte une optimisation de la localisation et des longueurs des forages recommandés. Deux hypothèses existent alors :

— le budget (nombre de forages-métrage) est fixe et connu d'avance, il s'agit alors de sélectionner la meilleure solution pour ce métrage ;
— le budget n'est pas encore fixé, il faut alors le fixer en se basant sur un pourcentage de couverture.

Enfin, dans les méthodes proposées par Bilal [16] ou Zagré [17] la couverture par la fonction objectif est plutôt rigide avec une fonction objectif basée sur la couverture discrete des blocs. En effet, la couverture d'une zone (blocs) et le calcul de la fonction objectif est binaire : soit un forage couvre la zone, soit ce n'est pas le cas. Or, intuitivement, pour un géologue, cette couverture est plus continue que discrète : proche d'un forage, le niveau de confiance (et donc de couverture) est maximal puis il décroît avec la distance.

La rapidité et la flexibilité de l'algorithme heuristique utilisé, la fonction objectif à maximiser ainsi que la couverture continue d'un bloc par un forage existant ou futur représentent certains des principaux défis actuels au problème d'optimisation de positionnement de forages et leur application sur les projets miniers. Le premier objectif de ma thèse vise à trouver une méthode heuristique rapide, flexible, permettant d'optimiser la position des recommandations de forages tout en respectant l'aspect continu de la couverture des sondages.

1.2.2 Seuil d'ecrétage

Lors des campagnes de forages, des échantillons sont envoyés au laboratoire pour analyses. Durant l'ensemble du processus, de nombreuses erreurs peuvent être générées. Leurs quantifications et leurs contrôles ont été largement étudiés par P.Gy dans la théorie d'échantillonnage [18]. Afin de contrôler et maîtriser ces erreurs, il est commun d'avoir un programme de contrôle de la qualité et de contrôle des analyses en place, c'est le *QaQc* qui lui aussi est traité par les *meilleures pratiques*. Ce programme peut être en place, soit au niveau du département géologie, soit au laboratoire lui même ; bien souvent aux deux. Les protocoles QaQc consistent généralement en un ensemble de reprises sur les échantillons (les duplicatas), de mesures de matériels certifiés (les standards) et de vérifications de la contamination grâce aux blancs

(matériel connue pour avoir une concentration nulle dans le ou les élément(s) recherché(s)). Malgré le respect de ces pratiques, il reste très fréquent d'observer des valeurs extrêmes dans les résultats. Elles peuvent être naturelles ou avoir pour origine une erreur analytique [11]. Ces valeurs peuvent représenter, dans un gisement aurifère par exemple, une forte proportion du contenu métal existant. Afin de contrôler le risque de sur-estimation, la pratique la plus largement utilisée dans l'industrie consiste alors à définir une teneur au-dessus de laquelle toutes les teneurs supérieures y sont ramenées : c'est le seuil d'écrêtage [10]. Il peut en résulter une diminution, parfois très importante, du contenu métal, et donc de la valeur économique d'un projet. Ce seuil d'écrêtage représente une des décisions les plus significatives et pourtant souvent subjective que le géologue analyste doit prendre. L'impact sur la valeur financière, couplé à son aspect subjectif en font un enjeu majeur de discussions et de risque.

Parmi les différentes méthodes proposées dans la littérature, aucune ne considère les résultats du QaQc afin d'aider à déterminer objectivement le seuil d'écrêtage. Il semblerait logique que pour une même distribution on décide d'un seuil plus élevé lorsque les analyses sont précises et d'un seuil plus bas lorsque les analyses le sont moins. Le second objectif spécifique de cette thèse s'attaque à cette question.

1.2.3 Estimation globale et changement de support

Une fois les données écrêtées, il est nécessaire de passer à l'exercice variographique car pratiquement toutes les méthodes de la géostatistique utilisent cet outil de mesure de la corrélation spatiale des données. Les différents types d'estimateurs de variogrammes expérimentaux généralement utilisés par les géologues dans l'industrie minière sont : le variogramme traditionnel (sur données brutes ou transformées) [19], le corrélogramme [14] et le variogramme relatif par paires [20]. Le but du géologue est d'obtenir un variogramme expérimental relativement facile à modéliser. Cet exercice est parfois difficile lorsque les résultats de QaQc sont médiocres ou que des valeurs extrêmes compliquent l'interprétation. L'écrêtage des valeurs extrêmes est le sujet visé par le second objectif spécifique. Une fois les variogrammes expérimentaux modélisés, le géologue analyste doit alors passer à l'interpolation du modèle de blocs. Dans la pratique, l'estimation des ressources demande de connaître la distribution des teneurs de petits blocs (SMU). Une pratique courante dans l'industrie consiste à ajuster les paramètres d'interpolation afin d'obtenir une distribution Tonnage-teneur des blocs interpolés proche de celle établie par COS [21]. Cette technique a pour but de mieux contrôler les paramètres d'interpolation et donc de réduire le risque lié au lissage d'interpolation qui peut impacter significativement la planification minière et la valeur d'un projet minier. Il est possible grâce à la relation de volume-variance, d'obtenir cette distribution : il s'agit des techniques de chan-

gement de support global. Les modèles de COS les plus populaires dans l'industrie minière sont les modèles discret gaussien (DG) ([4] et [22]) et les modèles *indirects log normal* (IL) ([14] et [23]). Ces modèles demandent de spécifier la variance des SMU et un histogramme des teneurs ponctuelles. L'histogramme des données ponctuelles est disponible pour un domaine donné, dès sa création. La variance des SMU, elle, doit être calculée à partir du modèle de variogramme ajusté sur le variogramme expérimental.

Peu d'études ont comparé les performances du DG et IL sur différentes distributions ponctuelles de teneurs ([24], [25], [26]). Aucune étude n'a comparé l'influence du type d'estimateur du variogramme sur les performances de ces méthodes. Le troisième objectif de ma thèse vise à déterminer l'impact des différents types de variogrammes expérimentaux sur les modèles de changements de support les plus couramment utilisés dans l'industrie minière pour plusieurs types de distribution et taille de SMU.

1.3 Objectifs de recherche

1.3.1 Objectifs généraux

Le but général de cette thèse est d'améliorer la robustesse et l'objectivité des estimations de ressources. Pour atteindre ce but, trois objectifs spécifiques seront poursuivis touchant chacun une phase distincte du processus suivi lors de l'estimation des ressources.

1.3.2 Objectifs spécifiques

A partir des problématiques décrites ci-dessus, trois objectifs spécifiques ont été définis :

— le premier objectif spécifique consiste à proposer un outil de recommandation de forage. Cet outil doit permettre l'optimisation de localisation de forage tout en permettant la comparaison de plusieurs gisements simultanément et en tenant compte de la couverture partielle continue des blocs en fonction des forages existants ou futurs. L'outil doit être objectif, flexible, rapide, et permettre de quantifier l'impact sur le taux de couverture de différentes solutions en considérant des budgets croissants. Le but est de maximiser la valeur de chaque forage recommandé de manière locale (dans le gisement) ou globale (sur plusieurs gisements) ;

— le second objectif spécifique vise à proposer une nouvelle approche de définition des seuils d'écrêtage qui soit simple, reproductible, objective et qui prenne en compte les erreurs d'échantillonnage disponibles grâce aux résultats de QaQc. Cette méthode doit permettre de réduire la subjectivité souvent existante dans ce domaine et être applicable sur un ensemble de substances minérales et pour différents types d'échan-

tillonnage (forages au diamants, forages de production, front de taille, échantillons de convoyeur, camions, tas, etc..) ;

— le troisième objectif spécifique a pour but de mesurer et d'interpréter l'impact du choix du type de variogramme expérimental sur les techniques de changements de support les plus utilisées dans l'industrie. Pour différents types de distribution, les erreurs introduites par les différentes méthodes seront quantifiées, permettant de définir les méthodes les plus robustes et les pratiques à éviter.

1.4 Plan de la thèse

Cette thèse est composée de six chapitres et présente trois articles. L'introduction (chapitre 1) présente les éléments des problématiques ainsi que les objectifs spécifiques de cette thèse.

Dans le chapitre 2, une revue de littérature des différentes méthodes et des outils traitant du sujet est réalisée. La première partie présente une revue de certaines des pratiques les plus communément trouvées dans des rapports techniques NI43-101 publics. Une revue des méthodes de recommandations objectives de forages est ensuite présentée. La problématique des erreurs d'échantillonnage, leur lien avec la corrélation des duplicatas, ainsi que les différentes techniques d'écrêtage sont présentés. Par la suite, sont discutés les méthodes de la géostatistique minière avec : les différents types de variogrammes expérimentaux et leur interprétation manuelle ou automatique d'une part, et les méthodes linéaires et non linéaires d'autre part. Enfin, une conclusion sur la revue de littérature est présentée. Les chapitres 3 à 5 présentent les articles publiés et soumis durant ce doctorat.

Dans le chapitre 3, un algorithme semi-glouton permettant l'optimisation de la localisation de forages est présenté. L'algorithme proposé est rapide, donne des résultats proches de l'optimum, fonctionne autant pour des projets à ciel ouvert ou souterrain, et permet durant l'optimisation et de manière dynamique de considérer la couverture continue des blocs par les forages. Ce chapitre publié dans [27] présente une application sur un jeu de données réelles issues d'une mine d'or. Cet article permet d'atteindre le premier objectif spécifique.

Dans le chapitre 4 une nouvelle approche de définition des seuils d'écrêtages prenant en compte la qualité des résultats de QaQc est proposée. Basée sur la corrélation des duplicatas de QaQc, le coefficient de variation des données sans erreur est estimé, et le seuil d'écrêtage est défini afin d'obtenir cette valeur. Cette nouvelle approche a été éprouvée sur deux jeux de données simulées et réelles issues de mines d'or et de nickel. Ce chapitre a été soumis au journal du *Southern African Institute of Mining and Metallurgy* en Septembre 2020 et permet d'atteindre le second objectif spécifique.

Le chapitre 5 correspond à un ensemble de tests effectués sur des données simulées permettant de quantifier les erreurs d'estimés des distributions tonnages-teneurs des modèles de changement de support global en fonction du type de variogramme expérimental choisi. Les différents modèles de COS et variogrammes ont été évalués pour plusieurs types de distributions et tailles de SMU. Ce chapitre a été publié dans [28] et concrétise l'atteinte du troisième objectif spécifique.

Dans le chapitre 6, une discussion sur les principaux résultats obtenus est réalisée. Elle est suivie par une analyse des limitations des outils et des méthodes proposées ainsi que des suggestions de recherches futures. Enfin, une conclusion générale vient clôturer le document.

CHAPITRE 2 REVUE DE LITTÉRATURE ET MÉTHODES

Les éléments chimiques possédant un intérêt pour l'extraction minière ont souvent des comportements compliqués à interpréter et à anticiper spatialement. Leur étude dans les années 50 en Afrique du Sud, ont permis à D. Krige de définir les bases de ce que G. Matheron allait formaliser pour devenir la géostatistique. Les méthodes de la géostatistique appliquées aux problématiques minières ont généralement pour but de réduire l'erreur d'estimation (estimation des valeurs du modèle de bloc) et de quantifier la confiance de l'estimation (définition des critères de classification, recommandations de forages). Dans ce chapitre, sont présentées une revue de la littérature critique ainsi qu'une présentation des principales méthodes et approches utilisées durant ce doctorat. La section revue de littérature commence par une revue de rapports techniques suivant l'instrumentation nationale canadienne (NI43-101), puis les concepts de bases de la géostatistique seront abordés : cadre mathématique, variographie, méthodes linéaires et non linéaires. Dans la section méthodologie, les concepts utilisés mais non présentés directement dans les articles sont décrits, notamment les outils de recherche opérationnelle heuristique, les problématiques d'échantillonnages, le QaQc, ainsi que les modèles de changement de support global et local.

Les compagnies canadiennes inscrites à la bourse du TSX sont soumises à la réglementation NI43-101. Cette réglementation se réfère aux normes et définitions de l'Institut Canadien des Mines (ICM) afin d'offrir un cadre légal notamment pour ce qui est de la publication d'estimation des ressources. Cette réglementation et les définitions de l'ICM offrent un cadre clair pour tout professionnel, cependant elle ne présente aucune recommandation technique spécifique. La personne qualifiée (PQ) doit se fier aux *meilleures pratiques* utilisées dans l'industrie. Ces règles, souvent non écrites, peuvent se retrouver dans les méthodes et les stratégies décrites dans les rapports techniques.

2.1 Revue des *meilleures pratiques* trouvées dans des rapports NI43-101

Tous les rapports NI43-101 sont publics et ils doivent être déposés sur le site internet : www.sedar.com Dans cette section est présentée une compilation de quarante rapports de ressources datant de 2016 à 2020 et traitant de projets ou de mines d'or, argent, cuivre, zinc, plomb, graphite, phosphate, tungstène et lithium.

La figure 2.1 présente certaines des informations de base de cette compilation. La majorité des rapports compilés ont été préparés par des firmes de consultation, seuls 28% des rap-

ports ont été préparés à l'interne. Ceci peut s'expliquer par la réglementation qui impose aux compagnies non productrices de s'orienter vers des auteurs indépendants pour leurs rapports NI43-101. Dans cette compilation, 35% des rapports concernaient des mines en opération, ce qui montre que certaines compagnies productrices préfèrent utiliser des consultants indépendants plutôt que de rendre public leurs propres rapports réalisés à l'interne. La majorité des rapports (57%) traite de projets aurifères. Par la suite, ce sont les projets touchant des dépôts d'argent qui sont les plus représentés. Noter que bien souvent, les gisements sont polymétalliques, dans ce cas, l'élément prépondérant a été utilisé pour classification. Bien que ces rapports traitent de gisements très différents tant d'un point de vue géologique que géostatistique, ils présentent de nombreuses similitudes dans les pratiques et les approches des analystes. Ceci est justement le but de la norme NI43-101, permettre des comparaisons simplifiées entre des gisements pourtant très différents.

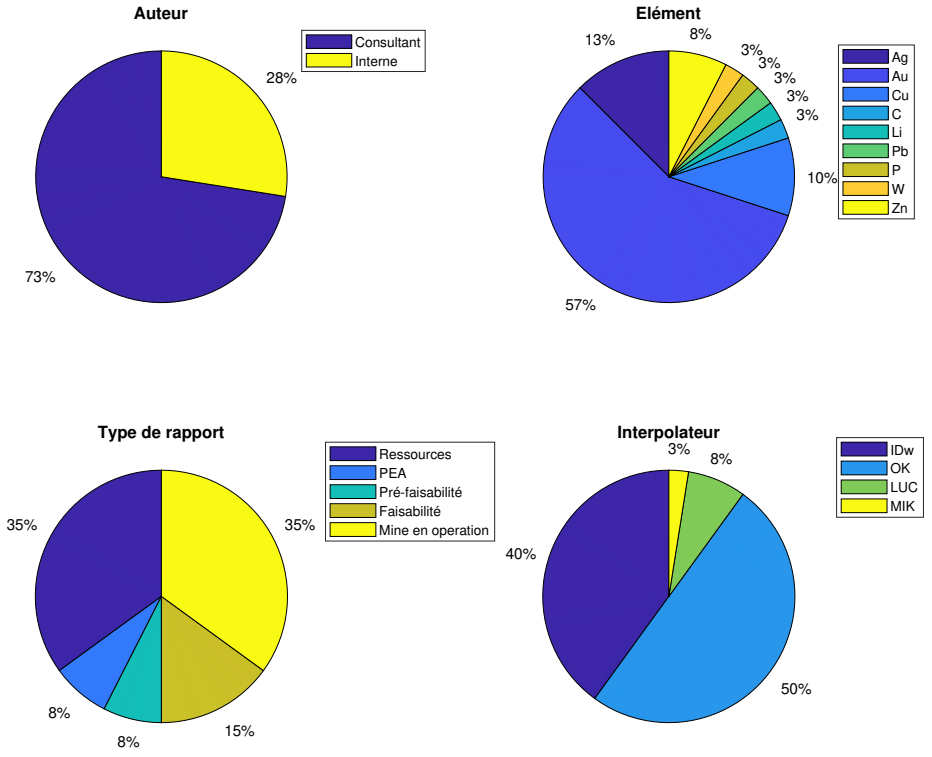


Figure 2.1 Compilation des informations de bases de 40 rapports NI43-101 : informations générales

Pour ce qui est des recommandations de forages, 90% des rapports en proposent et 3 des rapports n'en proposent pas. Il s'agit dans ces 3 cas de mines en opération qui possèdent pro-

bablement des programmes de forages de définition annuels non discutés dans les rapports. Malgré la présence de recommandations de forages dans pratiquement tous les rapports, seuls 5 présentent un phasage temporel des forages en fonction des résultats de la campagne précédente. Bien qu'il soit très probable que dans la majorité des cas, l'analyste ait un détail sur l'importance relative de chaque forage recommandé, cette notion pourtant très importante n'est jamais présentée dans les rapports techniques. Il existe probablement de nombreuses raisons à cela, et le manque de support visuel clair (graphique) en fait probablement partie. Dans aucun des rapports, les forages recommandés sont présentés ni justifiés de manière locale. Les justifications sont généralement de deux types : resserrement de maille de forages systématique ou tests des continuités latérales. Or, ces justifications générales ne donnent au lecteur aucune indication objective sur le motif des sondages, le taux de conversion des ressources (ex : de indiqué à mesuré) ou la diminution du risque après forages. Le développement d'un outil simple et accessible pour la majorité des analystes pourrait permettre de démocratiser des recommandations de forages montrant l'augmentation de la couverture proportionnellement au métrage recommandé. Une solution simple et rapide au problème des recommandations de forages est proposée dans le chapitre 3.

L'étape suivant les campagnes de forages est l'analyse des carottes. Durant ce processus, des erreurs d'échantillonnages [18] peuvent avoir lieu. Afin de quantifier et mesurer ces erreurs et d'autres, les compagnies minières mettent en place des protocoles de QaQc qui consistent notamment dans la mise en place de blancs, de standards et de duplicatas [29]. Les duplicatas ont pour but de mesurer la précision et la justesse des échantillons [30]. Les échantillons de QaQc peuvent représenter jusqu'à 20% des échantillons analysés [30], ce qui représente un coût très important pour une utilisation restreinte dans les NI43-101. Dans les rapports compilés, les résultats de QaQc ne sont utilisés qu'au fin de QaQc et il n'est pas fait mention de leur utilisation directe lors des étapes suivantes de l'estimé de ressources. La valorisation des informations, pourtant riches, contenues dans les résultats de QaQc représente un axe de recherche intéressant traité dans le chapitre 4 de cette thèse.

Parfois, pour des raisons d'économie ou de gain de productivité, les protocoles analytiques sont sous-optimaux [29] (par exemple une quantité trop faible de matériel est pulvérisée), résultant en une variance fondamentale de l'erreur d'échantillonnage élevée [18]). Ces variances ainsi que l'effet de pépite naturel [31] peuvent générer de fortes valeurs dans les résultats d'analyses. Quelles soient naturelles ou qu'elles résultent d'une erreur analytique, ces fortes valeurs représentent un risque pour l'estimé de ressources. Les risques peuvent être de diverses natures. Ils peuvent par exemple impacter le krigeage si un poids de krigeage négatif (même très faible) est associé à une très forte valeur ; alors, il est possible d'obtenir une valeur de blocs négative [32]. La présence de ces très hautes valeurs peut aussi masquer la continuité

des structures lors de l'interprétation des variogrammes [33]. Un autre risque significatif, notamment pour les populations à fort coefficient d'asymétrie, est la concentration d'une forte quantité de métal dans une très faible proportion des analyses. Afin de contourner ces problèmes, la pratique courante dans 98% des rapports compilés consiste à effectuer un écrêtage : une réduction de toutes les valeurs dépassant un certain seuil [10]. Seul le rapport traitant du gisement de phosphate n'applique pas cette technique. Dans ce rapport, le géologue a considéré que compte tenu de la distribution normale des résultats d'analyses, il n'était pas nécessaire de réduire la valeur des plus hautes teneurs. Dans tous les autres rapports, des seuils d'écrêtage ont été appliqués.

Dans près de 90% des cas, les personnes qualifiées font référence à une méthode graphique [34], comme principal outil de détermination des seuils. Bien que généralement présentés, ces graphiques offrent visuellement plusieurs options pour les seuils et ne sont pas objectifs ou répétables. Dans 4 rapports, la méthode Parish [35] a été utilisée, et seul un rapport fait état de l'utilisation d'une méthode de simulation. Ces méthodes sont objectives et répétables, même si l'utilisation de simulation est limitée par de nombreuses contraintes non abordées dans le rapport technique. Elles n'ont été utilisées cependant que sur des gisements aurifères. La méthode Parish a une tendance à recommander des résultats conservateurs (un maximum de 10% du métal contenu dans le dernier percentile des données). De plus, cette méthode étant basée sur les percentiles, un changement dans la taille de la population de référence (*i.e.* le nombre d'échantillons) entraîne des variations dans le seul d'écrêtage recommandé. Bien que plus de 75% des rapports discutent les résultats QaQc sur les duplicatas, seul 10% des rapports établissent un lien interprétatif entre le résultat de l'estimation et la qualité du QaQc. Dans ces rapports, la mauvaise qualité des résultats de QaQc, généralement sur des données historiques, a poussé l'analyste à classifier les ressources en *présumées* bien que la densité d'information aurait permise, toujours selon l'auteur de classer les ressources comme indiquées. Ces dé-classifications conservatrices et subjectives auraient peut-être pu être évitées si une gestion des seuils d'écrêtage en fonction des résultats de QaQc avait été utilisée. Cette compilation de rapports montre que l'absence d'outils objectifs de recommandation de seuil d'écrêtage en fonction des résultats du QaQc ne permet pas une gestion objective du risque associé. Dans le chapitre 4, nous proposons une méthode objective apportant une solution au problème de la définition de seuil d'écrêtage.

Durant l'exploration des données, seul 18% des rapports ne font pas mention d'études variographiques (figure 2.2). Dans les autres rapports, c'est le variogramme traditionnel qui reste le plus utilisé (50% des rapports y font référence). Ce chiffre est cependant à interpréter avec précaution, car à trois reprises le texte mentionnait « variogramme », alors que sur les graphiques on pouvait voir qu'il s'agissait en fait d'un autre type de variogramme (corrélo-

gramme ou relatif par paire). On peut supposer que les titres de certains graphiques n'ont pas toujours été correctement mis à jour. Le variogramme sur données transformées gaussiennes est présent dans 18% des rapports. Les variogrammes relatifs par paire et les corrélogrammes se retrouvent respectivement dans 8 et 5% des rapports, et enfin dans un rapport il est fait mention d'un variogramme sur données logarithmiques. Au total, on retrouve donc dans les rapports techniques compilés, un minimum de 34% de variogrammes autres, soit à peine moins que des variogrammes traditionnels.

Il est fréquent avec des données minières log normales d'avoir plusieurs valeurs extrêmes ainsi que de relativement fortes valeurs d'erreurs d'échantillonnages. Ces raisons entre autres poussent les analystes à utiliser des estimateurs autres que traditionnels afin de mieux modéliser la corrélation spatiale. Cependant, l'utilisation de ces différents variogrammes expérimentaux ne répond pas, d'un point de vue théorique, aux besoins des méthodes géostatistiques, qui requièrent l'utilisation du variogramme traditionnel (notamment le krigeage ordinaire (*KO*)). Il est à noter que dans aucun des rapports où les variogrammes sur données gaussiennes ont été utilisés il n'est fait mention d'une transformation du variogramme expérimental ou de son modèle, ce qui est pourtant requis.

Une fois les modèles de variogrammes définis, l'interpolation des blocs peut commencer. Près de la moitié des rapports utilisant une interpolation par inverse de la distance *ID* ont quand même effectué une étude variographique, soit pour valider la taille et l'orientation des ellipses de recherche, soit pour paramétrer une estimation par krigeage ou les outils géostatitistiques utilisés pour la validation, soit pour documenter l'utilisation d'une méthode non géostatistique (ID). Approximativement 50% des rapports utilisent le krigeage ordinaire comme méthode d'interpolation. Les méthodes utilisant l'inverse de la distance sont utilisées dans 40% des rapports. La justification de cette pratique étant que les variogrammes sont erratiques et la personne qualifiée a alors préféré utiliser une approche basée sur l'inverse de la distance. Les auteurs justifient ce choix par la volonté de limiter l'effet de lissage et de laisser les teneurs varier plus librement (une sorte de sacrifice du biais local au profit de la variance globale). Les méthodes non linéaires sont utilisées dans approximativement 10% des études seulement. Le conditionnement uniforme localisé est utilisé dans trois rapports, le krigeage par indicatrices multiple une fois. Bien que de très nombreuses techniques géostatistiques ont été développées dans la littérature, seul quelques unes sont rapportées par les géologues dans les rapports NI43-101.

Concernant les validations des estimés de blocs et de ressources, la réglementation NI43-101 ne recommande encore une fois aucune méthode spécifique, se reposant sur l'appréciation de l'analyste. Dans les projets étudiés, la vaste majorité (93%) présentaient une distribution log

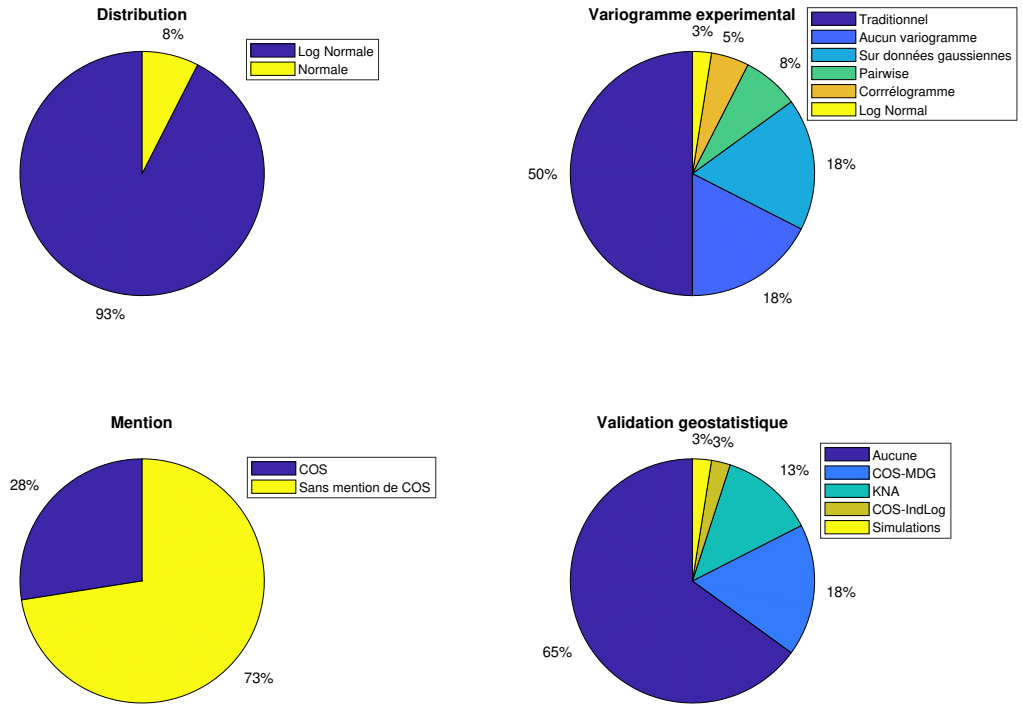


Figure 2.2 Compilation des informations de bases de 40 rapports NI43-101 : informations géostatistiques

normale pour les teneurs principales. Dans 28% des rapports, il est mentionné une validation par changement de support (COS), cependant très peu de rapports comparaient clairement les résultats du COS avec les résultats d'interpolation (par KO ou ID). La quantification d'un possible biais en fonction des teneurs de coupure oblige le lecteur à « faire confiance » à l'appréciation de l'analyste. Près de 65% des rapports compilés ne présentaient aucune validation géostatistique. Les seules mesures de validation présentées consistent généralement à une inspection visuelle des plans et des sections, à comparer différents types d'estimés (KO vs ID vs méthode polygonale (ou plus proche voisin)), à présenter les teneurs moyennes des analyses et estimés dans chaque domaine, à comparer visuellement les moyennes par lignes ou par plans pour mesurer les tendances (*SWATS*). Pour ce qui est des validations géostatistiques les plus communes, on retrouve les méthodes de changement de support dans 21% des rapports, avec respectivement 18% des rapports rapportant l'utilisation du modèle discret gaussien (*DG*, [4]) et dans un rapport (3%) la méthode *indirect log normal* (IL, [14]) (voir figure 2.2). Dans 13% des cas, les analystes ont décidé d'utiliser la technique d'analyse

du voisinage du krigeage (KNA, [36]) afin de choisir le voisinage de l'interpolation.

Souvent dans l'industrie minière, l'utilisation du KNA pour définir les paramètres d'interpolation est conceptuellement opposée à l'approche par COS ([37], [38]). Dans le premier cas, on recherche les paramètres offrant les meilleurs résultats pour le krigeage (résultant en un fort lissage), alors que dans le deuxième cas on cherche les paramètres offrant la distribution de blocs souhaités (généralement moins lissée). Enfin, on ne retrouve que dans un seul rapport la mention de simulation conditionnelle, pour la classification et une étude de risque sur les seuils d'écrêtage. Un examen détaillé des rapports montre que les méthodes de validation sont intimement liées aux logiciels utilisés par les analystes. Ainsi les variogrammes gaussiens et les validations par KNA sont retrouvés exclusivement dans les rapports mentionnant le logiciel *Supervisor* de Snowden, alors que les personnes utilisant une validation par COS-DG citent souvent le logiciel-module Herco (*HERmitian COrrection* développé dans Sgems ou le logiciel *Vulcan* de Maptek). L'utilisation de divers types de variogrammes expérimentaux et de plusieurs méthodes de changements de support sont des pratiques relativement courantes dans les rapports NI43-101 étudiés. Leur mauvais usage peut représenter un risque pour la qualité des estimés des ressources. Cette problématique est traitée dans le chapitre cinq de cette thèse.

Tous les rapports NI43-101 de ressources commencent avec une banque de données de forages. Les forages apportent toutes les informations, qu'elles soient géologiques, structurelles ou analytiques. C'est à partir de ces informations que le QaQc est formé, les domaines interprétés, l'exploration des données faite et les blocs interpolés. Les forages représentent aussi, comme le démontre la compilation de rapport NI43-101, une des recommandations les plus coûteuses et communes des géologues.

2.2 Les recommandations de forages objectives

La pratique la plus commune concernant les recommandations de forages consiste à effectuer manuellement et visuellement une évaluation des besoins d'information. Cette recommandation, basée sur l'expérience du géologue, est par définition subjective.

La *recherche opérationnelle* a pour but la recherche de méthodes rationnelles permettant de trouver le meilleur choix à un problème donné [39]. Elle a pour vocation de fournir une aide à la prise de décision pour des situations complexes qu'il faut dans un premier temps réduire à un problème de *minimisation* ou de *maximisation* d'une fonction objectif. Il existe de nombreuses méthodes mathématiques permettant d'optimiser la fonction objectif [40].

L'optimisation de position de forages est un problème *classique* de la recherche opérationnelle

appliquée à la géologie minière. Les premières références dans le domaine remontent à une quarantaine d'années ([41], [42] ou [43]). Ces études, réalisée en deux dimensions seulement, visaient à maximiser l'efficacité du krigeage. Une des limitations de l'utilisation de la variance de krigeage est qu'elle dépend seulement de l'espacement des données et du modèle de variogramme, mais n'est pas sensible aux variations locales des teneurs du gisement. Les méthodes gloutonnes, fonctionnent de manière itérative en sélectionnant le choix optimum local à chaque étape. Cependant, les méthodes gloutonnes sont sensibles aux minimums locaux et ne sont donc généralement pas utilisées pour l'optimisation de position de forages. Avec l'augmentation de la vitesse de calcul des ordinateurs, les méthodes heuristiques permettent de traiter de très larges volumes de données ainsi que les problèmes en trois dimensions. Soltani et *al.* [44] ont utilisé un algorithme génétique afin d'optimiser aussi la longueur des forages. Encore une fois, la variance de krigeage a été utilisée comme fonction objectif pour le positionnement des forages et bien qu'optimisant la longueur des forages, ces derniers étaient considérés comme verticaux. Cette limitation a été corrigée par les mêmes auteurs dans [45], en permettant d'orienter les forages dans une direction préférentielle grâce à la méthode du recuit simulé. Cependant, seule cette direction est investiguée, ce qui ne permet pas une optimisation de toutes les directions.

Dans deux articles ([16] et [46]), Bilal a proposé d'exprimer la problématique d'optimisation de position de forages sous la forme d'un problème de recouvrement partiel d'ensemble. Appliqué au problème d'optimisation des forages, il consiste, pour un budget donné, à déterminer le sous-ensemble de forages permettant de maximiser la valeur des blocs informés par ceux-ci. Bien que la programmation linéaire (algorithme du *simplexe*) permette de résoudre ce problème, elle n'est pas adaptée aux grand jeux de données. Pour contrer cette limitation, Bilal a proposé un algorithme heuristique hybride (détaillé plus loin) combinant l'algorithme génétique et la recherche itérative tabou afin d'effectuer respectivement une optimisation globale et locale. Cette approche permet de résoudre le problème en trois dimensions et ce, pour toutes les directions de forages possibles et approche le résultat optimum obtenu par programmation linéaire. Dans cette approche, les coûts de forages sont utilisés comme fonction objectif, et seule la distance entre les forages existants (ou proposés) est considérée, ce qui ne respecte pas les caractéristiques géologiques ou géostatistiques du gisement.

Un algorithme itératif a été utilisé dans Morshedy et *al.* [47] positionnant les forages dans les zones de métal maximum (teneur x épaisseur). L'algorithme quantifie l'amélioration de la variance des forages additionnels en ré-effectuant un krigeage pour chaque position possible. Cette méthode, de par son caractère itératif doit être particulièrement lente sur des jeux de données larges.

Soltani et *al.* [48] ont proposé d'utiliser les résultats de simulations conditionnelles sur les teneurs afin de définir la fonction objectif. Ils ont comparé les résultats de deux algorithmes d'optimisation (recuit simulé et essaim de particules) afin de démontrer que l'utilisation de plusieurs paramètres comme fonction objectif était supérieure à l'utilisation de la variance de krigeage seule. Dans leur approche, Pinheiro et *al.* [49] cherchent aussi à minimiser plusieurs critères obtenus par simulations conditionnelles. Ils utilisent la méthode de recuit simulé pour optimiser la position des forages à partir d'un ensemble de données géotechniques.

Une limitation importante des algorithmes d'optimisation de positionnement de forage est le temps de calcul. Afin d'améliorer les performances de la méthode d'optimisation par essaim de particules, Safa et *al.* [50] ont proposé de réduire le nombre de blocs grâce à une pré-sélection des blocs contenant l'information la plus significative. Leur approche a permis de réduire par trois le nombre de blocs utilisés pour l'optimisation sans altérer les résultats de positionnement des nouveaux forages. Cependant leur approche est sensible aux paramètres de pré-sélection et ajoute donc un risque significatif sur le positionnement des nouveaux forages.

Plus récemment, Zagré et *al.* [17] ont proposé d'améliorer la fonction objectif utilisée par Bilal, en y ajoutant notamment la notion d'incertitude sur la valeur du bloc couvert. Cette incertitude est ajoutée grâce à des simulations gaussiennes tronquées qui permettent de quantifier la confiance dans le modèle d'interprétation. Zagré et *al.* [17] utilisent aussi la recherche tabou [51]. Cette méthode métaheuristique permet une recherche des optimums locaux par exploration des solutions voisines, mais en s'interdisant de retourner sur une solution déjà testée pendant un certain nombre d'itérations. Cette interdiction assure ainsi un balayage large des solutions. Concrètement, l'algorithme procède de la manière suivante [17] :

— création d'une solution initiale grâce à la méthode gloutonne. Celle-ci sélectionne les forages présentant individuellement le plus fort recouvrement, sans tenir compte de leurs interactions ;
— recherche de l'optimum local par exploration du voisinage : les solutions proches sont vérifiées. Il s'agit soit d'ajouter un forage à la solution (par sélection gloutonne), d'en supprimer un, ou d'une combinaison de suppression et d'ajout d'un forage. Si une solution offrant une meilleure couverture est trouvée (*i.e.* la somme des valeurs de la fonction objectif calculée pour chaque bloc est supérieure à la solution précédente), cette solution est conservée ; sinon, l'algorithme se souvient des solutions testées et ne les re-testera pas pendant un certain nombre d'itération ;
— ré-itération de l'étape précédente en considérant la nouvelle solution optimale. Un ensemble de règles dites de *violations* permettent de pénaliser les solutions qui ne satisfont pas toutes les contraintes afin d'améliorer les résultats. S'il n'y a plus d'amé-

lioration, alors la solution est considérée quasi-optimale et conservée comme finale. L'approche de Zagré offre des résultats quasi-optimum (lorsque comparée à la programmation linéaire) et s'avère rapide sur des petits jeux de données, mais elle a pris plus d'une heure sur un jeu de données de taille intermédiaire, ce qui peut être problématique pour les projets de grandes tailles ou pour comparer plusieurs projets simultanément.

Les différentes approches heuristiques ([46], [17] ou [27]) ont toutes été comparées à la solution optimale obtenue par programmation linéaire dans le cas de la fonction de couverture discrète (seul cas pouvant être résolu par programmation linéaire). Lors de nos comparaisons dans [27], nous avons utilisé la fonction *intlinprog* du programme Matlab ([52]) pour résoudre le modèle linéaire de recouvrement, sous contrainte de budget, tel que formulé dans Zagré et *al.* [17].

Les méthodes de Bilal [46] et Zagré [17] ont utilisé une couverture de type binaire : soit un bloc est couvert entièrement, soit il ne l'est pas du tout. Or, cette notion est contre-intuitive pour le géologue. En effet, à l'instar de la variance de krigeage qui augmente avec la distance d'un point informé, la confiance dans la couverture d'un forage diminue lorsqu'on s'en éloigne. Il est donc nécessaire de proposer une approche où la valeur de recouvrement n'est pas binaire, mais partielle et que cette couverture soit ré-évaluée de manière dynamique durant le processus d'optimisation pour prendre en compte la nouvelle valeur de couverture partielle. Le premier objectif spécifique de cette thèse vise à couvrir cette lacune.

Une fois les campagnes de forages finalisées, l'analyse de données commence très généralement par le traitement des données extrêmes. En effet, ces données peuvent impacter négativement l'ensemble du projet pour deux raisons. Premièrement, elles peuvent représenter un risque économique significatif et deuxièmement, elles peuvent rendre difficile l'interprétation de certains outils géostatistiques, notamment le variogramme.

2.3 Les techniques d'écrêtage et les erreurs d'échantillonnage

Dans la plupart des campagnes de forages, les résultats d'analyses présentent quelques résultats avec des valeurs anormalement hautes. Ces valeurs extrêmes peuvent avoir pour origine des phénomènes naturels ou avoir pour origine des erreurs d'échantillonnage ou analytique. Afin de limiter le risque que représentent ces valeurs extrêmes, la pratique la plus courante trouvée dans les rapports NI43-101 consiste à y appliquer un seuil d'écrêtage.

2.3.1 Les techniques d'écrêtage commune dans l'industrie minière

Bien qu'il existe dans la littérature plusieurs techniques évitant d'appliquer un seuil d'écrêtage (par exemple, [53], [54], [55], [56], [57], [58], [59] ou [60]), leur utilisation n'est pas répandue dans les rapport NI43-101. Les *meilleures pratiques* recommandent de mettre en place un seuil d'écrêtage basé sur l'interprétation des données d'analyses. Certains auteurs parlent même de « ne pas croire dans un estimé sans écrêtage » [61], ce qui démontre à quel point la pratique d'écrêtage est un passage obligatoire pour la majorité des géologues.

Il existe plusieurs méthodes très largement utilisées dans l'industrie minière pour décider des teneurs d'écrêtages [10] :

— le dernier percentile [5] : cette méthode simpliste consiste à écrêter de manière autoritaire au 99e percentile le plus fréquemment. Cette méthode ne considère aucune des caractéristiques géologique, analytique ou géostatistique des données et représente une pratique historique peu justifiable. Une pratique similaire de certains géologues dans les gisements aurifères à haute teneur de l'Abitibi (nord-ouest du Québec) consistait à écrêter à une teneur d'une once par tonne (30 g/t) ;

— l'écrêtage de Parrish [35] : cette méthode rationnelle, basée sur le contenu en métal, cherche le seuil d'écrêtage qui permet d'obtenir respectivement moins de 40 % ou 10 % du métal au-dessus du 90e ou 99e percentile. Cette méthode objective dans son application est cependant subjective quant à ces définitions. Pourquoi 40 % du métal au-dessus du 90e percentile ? Pourquoi ne pas utiliser 60 % du métal ? De plus, bien que mise au point originellement pour des gisements aurifères, la méthode est particulièrement sévère. En effet, dans de nombreux gisements aurifères, notamment pour les mines souterraines, seuls quelques déciles de la population sont économiques et minés. Restreindre le seuil d'écrêtage au 90e percentile est très conservateur ;

— la courbe de coupure [62] : dans cette méthode, la moyenne de la population est calculée pour plusieurs teneurs d'écrêtage et représentée graphiquement. Le géologue doit alors chercher visuellement le point d'inflexion où la teneur d'écrêtage commence à impacter significativement la teneur moyenne. Cette méthode ne repose sur aucune logique statistique et est très subjective quant à la détermination visuelle du point d'inflexion ;

— les méthodes basées sur des simulations géostatistiques (Parker rapporté dans [5], [63] ou [64]) : ces méthodes plus ou moins élaborées utilisent les concepts de simulation afin de rechercher à quel moment les valeurs extrêmes commencent à représenter un risque pour l'interpolation ou les résultats de production. Bien que séduisantes, elles restent peu utilisées par la majorité des analystes, car elles sont complexes à mettre en œuvre

et reposent sur de nombreuses hypothèses parfois difficile à vérifier (notamment les problèmes de stationnarité). De plus, afin de générer les simulations conditionnelles, il faut assumer l'histogramme connu, or, c'est précisément la question de l'existence ou non des valeurs extrêmes. Il faut aussi fournir un modèle de variogramme qui est particulièrement sensible aux valeurs extrêmes. La méthode de Parker décrite par [5] a cependant été retrouvée dans un des rapports NI43-101 compilés ;

— interprétation de l'histogramme [65], [5] : sous une forme d'histogramme fréquentiel, de percentile, normale ou logarithmique, le géologue doit interpréter visuellement un décrochement ou un saut dans la queue de la distribution afin de définir la valeur d'écrêtage. Cette méthode est la plus largement utilisée dans les rapports techniques et correspond dans une certaine mesure à une *meilleure pratique*. Bien que les géologues arrivent souvent à un certain consensus sur l'interprétation de l'histogramme, cette méthode reste très largement subjective. De plus, il est fréquent qu'avec l'augmentation du nombre de données, le décrochement se décale vers une valeur plus haute, ce qui a une tendance à fournir des seuils d'écrêtage plus sévères pour les projets en début de vie et moins sévères par la suite.

Il est fréquent que les géologues utilisent une combinaison de plusieurs méthodes pour définir les valeurs d'écrêtage, car aucune de ces méthodes n'offrent une réponse objective et répétable. Une autre limitation de ces méthodes est le non respect de la qualité d'échantillonnage. En effet, intuitivement et statistiquement, si les analyses montrent une forte erreur d'échantillonnage alors il y aura statistiquement plus de valeurs extrêmes erronées. Inversement, si l'erreur d'échantillonnage est minimale, alors la confiance dans les teneurs extrêmes sera plus grande. Le second objectif spécifique de cette thèse vise à couvrir cette lacune.

Afin de pouvoir quantifier et mesurer l'impact des erreurs d'échantillonnages sur les valeurs extrêmes, il est nécessaire, dans un premier temps, de fournir un modèle reliant l'erreur statistique à une mesure d'erreur d'échantillonnage, ce problème est traité dans [29] et est résumé ci-dessous.

2.3.2 L'erreur d'échantillonnage et lien avec les duplicatas

Dans la réglementation canadienne NI43-101, il est obligatoire de mettre en place un processus de validation des protocoles analytiques, le plus souvent on le retrouve sous forme de contrôle qualité et assurance qualité : le QaQc. Des revues complètes des procédures et des méthodes de suivi du QaQc pour l'industrie minière sont présentées dans [30], [66], [67] ou encore [68].

Un des éléments du QaQc consiste à valider les variances analytiques telles que présentées dans la théorie d'échantillonnage de Gy [69]. Cette théorie originale a été peaufinée et ajustée

au fil du temps par Pierre Gy [18] et quelques autres auteurs notamment Pitard [70] et Francois-Bongarcon [71] pour ne citer que les principaux acteurs du domaine minier.

La formule de Gy [18] permet de calculer l'erreur fondamentale (s_r^2) de manière théorique en utilisant : une teneur cible, la distribution des tailles des particules et les caractéristiques principales des minéraux portant la minéralisation. Cette valeur théorique peut alors être corrélée avec celles des résultats de QaQc notamment la variabilité et la corrélation des duplicatas [29]. A partir de variance fondamentale d'erreur de Gy, et grâce à un modèle d'erreur simple de type $Z_o = Z_t.t$; où Z_o est la teneur observée avec erreur, Z_t est la teneur originale « sans erreur » (*ex* : le résultat d'une simulation) et t est le facteur d'erreur aléatoire multiplicatif, les auteurs montrent que :

$$\rho_{dup} = \frac{\sigma_z^2}{\sigma_z^2 + \kappa m_z b_\kappa^2} \tag{2.1}$$

où ρ_{dup} est la corrélation entre les duplicatas, σ_z^2 la variance des teneurs sans erreur, κ est la teneur cible (dans la formule de Gy), m_z est la teneur moyenne des analyses et b_κ^2 est la variance relative de l'erreur fondamentale calculée avec la formule de Gy à la teneur κ.

Développé dans le chapitre cinq, il est démontré que dans le cas log normal :

$$\rho_{dup} = \frac{e^{\beta^2} - 1}{e^{\beta^2 + \sigma^2} - 1} \tag{2.2}$$

avec β^2 étant la variance des teneurs sans erreurs et σ^2 la variance du modèle d'erreur t, t étant un bruit blanc de distribution $N(\mu, \sigma)$. Supposant qu'il n'y a pas de biais sur les teneurs on peut poser $\mu = -\sigma^2/2$.

Le coefficient de variation des valeurs *vraies* (CV_v) est directement lié à celui des valeurs observées (CV_o) par le coefficient de corrélation des duplicatas, avec :

$$CV_v^2 = \rho_{dup} CV_o^2 \tag{2.3}$$

Grâce à ces différentes équations, il est possible de :

- simuler des jeux de données respectant les erreurs d'échantillonnages souhaitées ;
- calculer, à partir des résultats de duplicatas observés, l'erreur d'échantillonnage théorique et aussi le coefficient de variation sans erreur de la distribution.

Ces relations offrent beaucoup de flexibilité afin de tester de manière théorique ou empirique l'impact des protocoles ou des résultats de QaQc. L'équation 2.3 permet de lier le résultat du QaQc (ρ_{dup}) à l'écrêtage des valeurs extrêmes (lien entre CV_o^2 et CV_v^2) permettant ainsi

d'atteindre le troisième objectif spécifique de cette thèse.

2.4 La géostatistique

La géostatistique est une branche des statistiques spatiales qui vise l'analyse des données numériques spatialisées issues des sciences de la Terre, c'est l'application des méthodes statistiques et probabilistes aux *variables régionalisées* [72]. Dans un contexte minier, l'approche théorique consiste à traiter les résultats d'analyses comme *une réalisation* ($z(x)$) d'une *fonction aléatoire* ($Z(x)$) inconnue au point x. Cette réalisation est considérée comme *ergodique* ce qui, pour simplifier, signifie que l'on peut inférer les premiers moments de la fonction aléatoire $Z(x)$ par des moyennes spatiales de la réalisation $z(x)$. Si ces moments existent, et qu'en tout point leur *espérance* est identique, on dit qu'ils sont invariables à une translation près (déplacement de h), la variable est alors dite *stationnaire*. Le cas où les deux premiers moments sont invariables à une translation h est dit *stationnaire d'ordre deux* (SRF ou *second order random function* en anglais), on a alors [4] :

$$E[Z(x)] = m, \tag{2.4}$$

$$E[Z(x) - m][Z(x+h) - m] = C(h) \tag{2.5}$$

où $C(h)$ est la fonction de covariance. Dans la pratique, les projets miniers sont très rarement stationnaires d'ordre deux et *l'hypothèse intrinsèque*, moins stricte, est généralement assumée. Cette hypothèse consiste à supposer que pour chaque translation h, seul l'incrément est stationnaire d'ordre deux [4], on a alors :

$$E[Z(x+h) - Z(x)] = 0, \tag{2.6}$$

$$Var[Z(x+h) - Z(x)]^2 = 2\gamma(h) \tag{2.7}$$

L'hypothèse intrinsèque couvre généralement les caractéristiques des données minières, au moins dans des sous-domaines homogènes, ce qui explique l'utilisation large de la géostatistique dans cette industrie.

2.4.1 Le variogramme

Quand on étudie une SRF, on peut utiliser la covariance pour décrire sa corrélation spatiale. Cependant, dans les applications minières, c'est le *variogramme* qui est le plus utilisé car il ne nécessite pas d'estimer la moyenne inconnue ni l'existence de la variance (palier

du variogramme) puisqu'il ne demande que la stationnarité des accroissements (hypothèse intrinsèque). Dans la pratique, l'analyse variographique est divisée en deux parties :

— calcul du variogramme expérimentale ;

— modélisation du variogramme expérimental.

L'objectif de l'étude variographique est d'obtenir une modélisation des corrélations spatiales pour tout vecteur distance h (module et orientation) et ce, à partir des données d'analyses disponibles. Ce modèle est par la suite utilisé dans pratiquement toutes les méthodes géostatistiques.

Le variogramme expérimental traditionnel [73], est défini comme :

$$\gamma(h) = \frac{1}{2N} \sum_{i=1:N(h)} (Z(x_i) - Z(x_i + h))^2 \tag{2.8}$$

avec $Z(x_i)$ la teneur au point x_i, h le vecteur distance séparant les deux points et $N(h)$ le nombre de paires disponibles à cette distance. Dans la pratique, une tolérance sur le vecteur h est appliquée, en module et en orientation, pour accroître le nombre de paires $N(h)$. Le variogramme traditionnel est sensible à la présence de valeurs extrêmes et aux différentes hétérogénéités présentes dans le domaine étudié.

Le variogramme sur données gaussiennes ([4], [74]) consiste tout simplement à utiliser les transformées gaussiennes des valeurs plutôt que les teneurs originales. Dans le champ gaussien, le *bruit* parfois créé par les fortes valeurs est atténué et le comportement du variogramme est généralement plus facile à interpréter. Il présente une variance normalisée à 1, ce qui peut, dans certaines applications, nécessiter un traitement particulier, par exemple pour les techniques utilisant un COS local comme le conditionnement uniforme.

Le variogramme relatif par paire [20] propose de normaliser la corrélation de chaques paires par la valeur moyenne de chacune de ces paires. Cette méthode permet de contenir les valeurs extrêmes et de contrôler partiellement la dérive.

$$\gamma(h) = \frac{2}{N(h)} \sum_{i=1:N(h)} \left(\frac{Z(x_i) - Z(x_i + h)}{(Z(x_i) + Z(x_i + h))} \right)^2 \tag{2.9}$$

Un des outils les plus utilisés dans les mines est le corrélogramme inversé [14]. Grâce à diverses normalisations, ce dernier serait moins sensible aux déviations et aux valeurs extrêmes, notamment pour les distributions log normales. Il offre un estimé de la variance normalisée à 1, ce qui peut dans certaines applications nécessiter un traitement particulier.

$$\rho(h) = \frac{1}{N(h)} \sum_{i=1:N(h)} Z(x_i)Z(x_i+h) - m_{-h}m_{+h} \tag{2.10}$$

avec,

$$m_{-h} = \frac{1}{N(h)} \sum_{i=1:N(h)} Z(x_i) \tag{2.11}$$

et,

$$m_{+h} = \frac{1}{N(h)} \sum_{i=1:N(h)} Z(x_i+h) \tag{2.12}$$

on peut alors calculer :

$$\gamma(h) = \frac{\rho(h)}{\sigma_{-h} \cdot \sigma_{+h}} \tag{2.13}$$

avec,

$$\sigma_h = \frac{1}{N(h)} \sum_{i=1:N(h)} Z(x_i)^2 - m_h^2 \tag{2.14}$$

et,

$$\sigma_{+h} = \frac{1}{N(h)} \sum_{i=1:N(h)} Z(x_i+h)^2 - m_{+h}^2 \tag{2.15}$$

avec m_{-h} ; m_{+h} et σ_{-h}, σ_{+h} les moyennes et écarts-types obtenues respectivement du côté tête et du côté queue des paires. Le fait que les moyennes et les écarts-types varient selon le côté de la paire et selon h permet, selon leurs auteurs ([75], [76]), de mieux tenir compte d'hétérogénéités ou de tendances dans le domaine étudié.

Le variogramme d'indicatrice est un autre outil souvent utilisé dans l'industrie minière. Il est insensible aux valeurs extrêmes et permet de décrire la corrélation spatiale pour plusieurs teneurs différentes [4]. Plusieurs méthodes géostatistiques nécessitent l'obtention de variogrammes d'indicatrices.

$$\gamma_(h) = \frac{1}{2N(h)} \sum_{i=1:N(h)} (I(x_i;c) - I(x_i+h;c))^2 \tag{2.16}$$

où $I(x_i;c)$ est l'indicatrice codée par rapport au seuil c suivant :

$$I(x_i;c) = \begin{cases} 1, & Z(x_i) <= c \\ 0, & Z(x_i) > c \end{cases} \tag{2.17}$$

Lorsque le variogramme expérimental est présenté pour différents vecteurs h, sans regard à

l'orientation il est dit *omnidirectionnel*. Si on l'affiche pour une direction particulière (généralement dans 3 directions orthogonales en 3 dimensions), on parle de variogramme *directionnel*. A partir de ces valeurs de variogrammes expérimentaux, il est nécessaire d'ajuster un modèle mathématique afin d'avoir une valeur de corrélation spatiale pour tout vecteur h. Si les variogrammes directionnels montrent des comportements différents selon la direction, on cherchera à les modéliser avec des modèles présentant des anisotropies géométriques ou zonales. Dans le cas contraire, on adoptera un modèle isotrope [77].

Lors de l'utilisation de technique d'interpolation telle que le krigeage, il est nécessaire de disposer d'un modèle de corrélation spatiale définie pour toute distance. Il faut donc modéliser ou *ajuster* une fonction mathématique au variogramme expérimental. Il existe de nombreuses fonctions mathématiques autorisées permettant de modéliser les variogrammes, la seule condition étant que cette fonction soit *conditionnellement positif semi-définie* (la condition est que la somme des poids dans la combinaison linéaire donne 0). Cette condition assure qu'une solution existe dans le système de krigeage et que la variance de krigeage soit positive ou nulle [20]. La fonction utilisée n'a généralement que peu d'importance, le plus important étant que la modélisation corresponde au mieux aux données expérimentales, en particulier aux petites distances et à la compréhension du phénomène étudié. Les modèles les plus couramment utilisées dans les mines sont [4] :

— le modèle sphérique qui présente un comportement quasi-linéaire à l'origine et ce jusqu'au palier, il est décrit comme :

$$\gamma(h) = \begin{cases} C(\frac{3h}{2a} + \frac{h^3}{2a^3}), & \text{si } h <= a \\ C, & \text{si } h > a \end{cases} \tag{2.18}$$

où C est le palier et a la portée ou distance de corrélation ;

— le modèle exponentiel, présente un comportement plus «arrondi» qui atteint le pallier de manière asymptotique. Dans la pratique, on considère que la portée à une valeur de 3a (ou 95% du palier), il est décrit par :

$$\gamma(h) = C\left(1 - exp(\frac{-3h}{a})\right) \tag{2.19}$$

— Le variogramme expérimental présente parfois une *discontinuité à l'origine*, cette dernière est référée comme l'effet de pépite. Afin d'ajuster correctement un modèle sur le variogramme expérimental, il est parfois nécessaire d'ajouter une première structure C_0 qui prend pour la valeur du palier de la discontinuité à une distance $h = 0$.

Il existe de nombreuses autres fonctions [4] : plus générale comme les fonctions de *Matern*,

d'autres montrant un comportement à l'origine plus continu comme les fonctions gaussienne, des fonction d'effets de trous qui permettent de modéliser une certaine périodicité dans les fluctuations, et bien d'autres. Mais ces fonctions sont moins utilisées dans l'industrie minière, notamment dû aux limitations dans les logiciels commerciaux.

Parfois, le variogramme expérimental montre une forme difficile à modéliser avec une seule fonction. Il est alors nécessaire d'utiliser un modèle *gigogne*. Il s'agit d'une combinaison linéaire de plusieurs fonctions autorisées (à coefficients positifs), où la valeur de la variance à une distance h est égale à la somme de plusieurs modèles admissibles, avec une forme :

$$\gamma(h) = \sum_{i=1:k} b_i \gamma_i(h) \tag{2.20}$$

avec $b_i > 0$, où b_i sont les coefficients des fonctions autorisées.

Il est très fréquent dans les projets miniers d'ajuster des modèles gigognes comprenant un effet de pépite suivis de deux structures (sphériques et/ou exponentielles). A l'exception de l'effet de pépite qui est isotrope, il est fréquent d'avoir une ou deux des structures qui présentent des caractéristiques anisotropes. Pour un modèle avec effet de pépite, et deux sphériques anisotropes par exemple, cela représente 15 paramètres à ajuster $(C_0, C_1, C_2, ax1, ay1, az1, ax2, ay2, az2, rx1, ry1, rz1, rx2, ry2, rz2)$, où les C sont les paliers, les a les portées et les r les rotations, ce qui est une tâche complexe.

L'ajustement du modèle de variogramme est une étape critique lors de l'exercice variographique [5]. Ce faisant, afin d'améliorer l'objectivité de l'exercice, le rendre plus rapide ou répétable, ou pour permettre la validation en série (pour des résultats de simulation par exemple) plusieurs auteurs ont proposé d'automatiser l'ajustement du modèle. Cressie [78] fut l'un des premiers auteurs à proposer une méthode d'ajustement automatique, c'est cette approche qui a été utilisée dans nos travaux [28]. A partir des points du variogramme expérimental, un modèle composé de 1 à k structures est ajusté par la méthode des moindres carrés pondérés par l'inverse de la distance et par le nombre de paires [79].

D'autres méthodes d'ajustement existent comme la validation croisée [80]. La maximisation de la vraisemblance [81] est particulièrement appréciée pour leur robustesse en présence de dérives. Plus récemment, une version itérative des moindres carrés couvrant notamment les cas multivariés et la géostatistique locale a été développée [82]. Contrairement à la plupart des autres approches, cet algorithme sélectionne lui-même le nombre et le type de structures à ajuster. Elle est notamment utilisée dans le logiciel *Isatis* [83] de la société Géovariances. Plus récemment, une approche utilisant les machines à vecteurs de support et un algorithme d'optimisation génétique a été proposée [84]. L'ajustement a pour but de trouver le modèle

de variogramme qui s'ajuste le mieux aux données expérimentales et qui fournit les meilleurs résultats pour le krigeage (validation croisée). Les auteurs y démontrent que la méthode par algorithme génétique surpasse celle des moindres carrés pondérés de Cressie.

2.4.2 Géostatistique linéaire

Les méthodes d'interpolation de la géostatistique linéaire ont pour but d'estimer en un point ou un bloc (volume) généralement non échantillonné la valeur d'un champ aléatoire. Le krigeage en est la méthode de base, il s'agit d'un estimateur utilisant le modèle de corrélation spatiale afin de minimiser la variance d'estimation de la variable aléatoire en effectuant une moyenne pondérée des données existantes [4]. Dans le cas du krigeage simple, la teneur moyenne (connue) de la variable est nécessaire, alors que dans le cas du krigeage ordinaire (KO), la somme des poids est forcée à 1, ce qui permet de ne pas devoir fournir la moyenne. Le KO est la forme la plus largement utilisée dans les mines où la teneur moyenne de la variable est rarement connue. Le krigeage ordinaire a de nombreuses propriétés intéressantes dans un contexte minier.

Il existe plusieurs autres méthodes de krigeage, notamment le krigeage multigaussien [85], krigeage avec dérive [4], le krigeage contraint (initialement proposé par [86] et plus récemment traité dans [87]) et les co-krigeages hétérotopiques ou multicolocalisés ([4], [88]) (pour les cas multivariés). Ces méthodes, bien qu'efficaces, sont moins utilisées dans le domaines minier (hormis les techniques de co-krigeage), notamment à cause de la limitation à la technique de KO de la plupart des logiciels miniers.

2.4.3 Géostatistique non linéaire

La géostatistique non linéaire est la branche de la géostatistique qui étudie les combinaisons non linéaires de la fonction aléatoire étudiée [89]. Elle peut se diviser en 3 familles [34] :

- les méthodes basées sur l'espérance conditionnelle, avec les fonctions indicatrices, comme la méthode de krigeage des indicatrices multiples (MIK) ;
- les méthodes d'estimations globales ou locales basées sur la relation volume variance : les techniques de changement de supports (e.g. le modèle discret gaussien (DG) et *indirect log normal*) et les méthodes de conditionnement uniforme (UC) ;
- les méthodes de simulations conditionnelles, dont les bandes tournantes, la méthode séquentielle gaussienne, transformée de Fourier rapide (FFT-MA) ou encore par décomposition Cholesky.

Les méthodes d'estimation utilisant l'espérance conditionnelle, telles que le MIK, sont dites non paramétriques car elles ne nécessitent aucune hypothèse sur la distribution des données

[4]. Elles utilisent une transformation par indicatrices (eq. 2.17) : pour plusieurs seuils de valeurs, la variographie est effectuée sur ces indicatrices, et un krigeage est fait pour chaque indicatrice. Pour chaque point interpolé, on obtient donc plusieurs valeurs interprétées comme la probabilité en ce point d'être au-dessus de chaque seuil, il s'agit donc d'une courbe de distribution locale, identique à celle obtenue par UC, à ceci près, que cette distribution est calculée au support ponctuel et qu'il faut donc y appliquer une technique de changement de support pour l'obtenir à la taille du SMU souhaitée. C'est généralement la méthode *affine* qui est retenue [14]. Bien qu'attrayante à premiere vue, la méthode MIK avec correction affine possède de nombreux inconvénients incluant un problème d'ordre (croissance non monotone des probabilités avec les seuils), une déstructuration des variogrammes à hautes teneurs et un manque de confiance dans la queue de distribution des hautes teneurs. De plus, la correction affine produit des résultats non réalistes dès que la réduction de variance est importante. Ces problèmes et d'autres font que certains auteurs ont émis des réserves quant à l'utilisation du MIK ([90] ou [91]). Elle reste cependant utilisée par certains géologues. Noter que la méthode de localisation [92] peut être appliquée au cas du MIK, elle est alors généralement référée sous le nom LIK (*localised indicator kriging*).

La relation de Cartier [4] stipule qu'un point de position aléatoire dans un bloc a une espérance égale à la valeur moyenne du volume (bloc). Ce postulat permet d'obtenir une relation entre l'histogramme des valeurs quasi-ponctuelles (analyses) et l'histogramme des valeurs sur une taille de support différentes (un bloc). La seule donnée manquante est la dispersion dans le volume du SMU, qui peut être calculée grâce au variogramme.

Le modèle discret gaussien global [93], et son application locale par conditionnement uniforme ([94], [4]), est la technique non linéaire la plus largement rapportée dans les rapports NI43-101 compilés. Son utilisation a regagné en popularité depuis le développement de la méthode de localisation [92]. De nombreux auteurs ont démontré ses avantages dans des contexte miniers ([95], [96], [97], [98], [99], [100], [101], [102] en sont quelques exemples récents).

Le modèle DG consiste, dans un premier temps, à créer une fonction d'anamorphose. Cette dernière peut être soit sous la forme des polynômes d'Hermites, soit sous une simple forme graphique (quantile-quantile). Sous l'hypothèse que les transformées gaussiennes au support ponctuel et à la taille du SMU ont une distribution binormale avec un coefficient de corrélation r. On peut ensuite retrouver la distribution pour tout support à partir de l'anamorphose en utilisant [94] :

$$Z_v = \phi_v(Y_v) = \int \phi(rY_v + (1-r^2)^{\frac{1}{2}}u)g(u)du \tag{2.21}$$

ϕ étant la fonction d'anamorphose, v le support ou SMU, $g(u)$ la fonction de densité gaussienne. La variance de bloc est donnée par :

$$Var(Z_v) = Var(Z_.) - \overline{\gamma}(v, v) \tag{2.22}$$

où $\overline{\gamma}(v, v)$ est la moyenne de $\gamma(h_0)$ pour toutes les paires de points dans v. Le coefficient r de l'équation 2.21 est sélectionné de sorte que $Var(Z_v)$, calculé à partir de l'équation 2.21 soit égal à $Var(Z_v)$ calculé à l'aide du variogramme par l'équation 2.22. Dans nos travaux [28], nous n'avons pas utilisé les polynômes d'Hermite en raison des instabilités numériques rencontrées et de la non monotonicité de la transformation obtenue, problèmes inexistants avec l'approche graphique.

Une des forces du DG est qu'il est facile de l'adapter pour pouvoir l'utiliser localement plutôt que globalement. On utilise alors les résultats du krigeage sur de larges panneaux (comparativement au SMU) pour poser $E[Z_v|Z_V] = Z_V^*$, où Z_V^* est la teneur krigée du panneau : c'est le conditionnement uniforme [4]. On obtient ainsi, pour chaque panneau, la distribution des teneurs de blocs. Cependant, afin de localiser ces teneurs à l'intérieur du panneau, une étape supplémentaire est nécessaire. Abzalov [92] a proposé une solution basée sur la discrétisation de la distribution basée sur l'ordre décroissant des valeurs krigées, c'est le *localised uniform conditionning*. Cette technique, bien que maintenant largement utilisée dans l'industrie, possède plusieurs défauts : elle peut notamment créer des artefacts de positionnement suivant le gradient du classement par krigeage ou encore *créer* des blocs de basses teneurs (théoriquement existant dans la distribution) dans des zones de hautes teneurs avérées.

Une autre technique de changement de support global utilisée dans l'industrie minière est la méthode *indirect log normal* ([14], [23]). Mathématiquement plus simple, cette méthode est adaptée aux distributions à fort coefficient d'asymétrie et utilise l'hypothèse de permanence de la distribution log normale pour *ajuster* la variance de la distribution à un support v tout en conservant la teneur moyenne inchangée.

Bien qu'existantes depuis les années 1970 et le début de la géostatistique ([103], [104]), les méthodes de simulations se sont surtout répandues à partir des années 2000 avec l'augmentation de puissance de calcul des ordinateurs. Alors que les estimations linéaires visent à fournir le meilleur estimé de la variable régionalisée en minimisant la variance d'estimation, les simulations, elles, visent à reproduire l'histogramme et la corrélation spatiale de la variable régionalisée. Grâce à un ensemble de réalisations, il devient possible de calculer plusieurs statistiques et tests permettant d'établir des mesures de risques [105], de définir les cas op-

timistes ou conservateurs ([106], [107]), quantifier l'effet d'information ou tester des mailles de forages ([108], [109]), ou encore quantifier les ressources récupérables pour diverses tailles de SMU [110]. Les simulations peuvent être faites sans données ou avec des données localisées, on parle alors de simulations conditionnelles. La plupart des techniques de simulations telles que les bandes tournantes [4] ou les transformées de Fourier (FFT-MA [111]) se décomposent de la façon suivante : transformation gaussienne (par anamorphose), estimation du variogramme gaussien, simulations non conditionnelles de la variable (avec reproduction du variogramme), conditionnement par krigeage simple ou ordinaire [112] et transformation des résultats par anamorphose inverse. Certaines techniques de simulations ne nécessitent cependant pas de conditionnement, c'est le cas des simulations gaussiennes séquentielles [34], des simulations par recuits [113] ou encore par décomposition de la matrice de covariance [114]. Chaque méthode possède des avantages et des inconvénients, mais toutes ces méthodes, bien que puissantes, sont difficiles à ajuster et nécessitent notamment un bon contrôle de l'histogramme et du variogramme de la variable simulée. L'hypothèse de stationnarité est aussi très forte, ce qui n'est que rarement acceptable dans les données minières. Bien que les simulations non stationnaires existent [4], elles sont encore pour le moment réservées au monde de la recherche. C'est pourquoi malgré leur forte valeur ajoutée, les simulations en général restent très peu utilisées dans la pratique courante des géologues de ressources ou minier.

2.5 Discussion

De toutes les recommandations que le géologue peut faire, les forages représentent l'une des plus importantes. Ces recommandations aident les compagnies à financer leurs futures campagnes, parfois pour des montants de plusieurs dizaines de millions de dollars. C'est aussi grâce à ces forages que les projets peuvent augmenter leurs ressources en indiquées ou mesurées. Les compagnies de grandes tailles ont souvent plusieurs projets qui sont en compétition pour une partie de l'enveloppe budgétaire de forages. Les budgets de forages des compagnies étant généralement coûteux et limités, l'utilisation de méthodes rationnelles et objectives est de la plus haute importance afin de maximiser le retour sur investissement que représentent les forages. Les outils existants de la recherche opérationnelle permettent de traiter le problème, mais les temps de calculs sont parfois très longs et ne permettent pas de traiter plusieurs projets simultanément. De plus, afin de correspondre aux besoins des géologues, il est nécessaire de développer un nouvel outil permettant une couverture des blocs continue plutôt que discrète. Le développement d'un nouvel outil répondant à ces besoins est le thème du premier objectif spécifique de cette thèse.

La valeur d'un projet minier, ou les profits d'une mine en opération, sont directement liés à la

vente d'un métal. L'évaluation du contenu métal est donc de la plus grande importance lors de l'estimation des ressources. Or, pour la plupart des projets revus dans les NI43-101, le fort coefficient d'asymétrie des données garantit qu'une très forte proportion du métal réside dans les valeurs les plus hautes. La distinction entre une valeur extrême vraie et une valeur extrême due à la variance d'échantillonnage est donc de la plus haute importance. L'utilisation de seuil d'écrêtage sur les projets miniers peut réduire la valeur du projet de plusieurs millions de dollars. Les méthodes actuellement utilisées dans les NI43-101 sont subjectives et ne prennent pas en compte l'erreur d'échantillonnage. Apporter une solution à cette lacune est le second objectif de cette thèse.

D'autres sujets importants auraient pu être choisis, mais ils sont généralement moins significatifs. Par exemple, le choix de l'interpolateur : krigeage ordinaire vs inverse de la distance. Les problèmes de longueur des composites (régularisation) sont eux aussi un choix important. Cependant, l'importance est surtout sur le contrôle du lissage des données. Peu importe la méthode d'interpolation choisie, la taille des composites ou des SMU, le plus important est de bien choisir les paramètres d'interpolation et d'ajuster judicieusement le degré de lissage. Ce thème correspond au troisième objectif spécifique.

A partir de la revue de rapports techniques publiques NI43-101, trois sujets particulièrement importants ont été identifiées : les recommandations de forages, la définition objective de seuils d'écrêtages et la meilleure pratique concernant l'utilisation du COS en fonction du type de variogramme. La revue des méthodes existantes a démontré que plusieurs lacunes existent dans les approches actuelles, elles sont souvent subjectives, non reproductibles, incomplètes, ou non documentées dans la littérature. Les chapitres suivants apportent des réponses concrètes et explicites à ces manques pour chacun des trois sujets.

CHAPITRE 3 ARTICLE 1 : A NEW SEMI-GREEDY APPROACH TO ENHANCE DRILLHOLE PLANNING

Historique de l'article : soumis le 20 janvier 2020, accepté le 28 mars 2020, publié dans *Natural Resources Research*.

Auteurs : Raphael Dutaut, Denis Marcotte

DOI : 10.1007/s11053-020-09674-8

3.1 Abstract

In advanced exploration projects or operating mines, the process of allocating capital for infill drilling programs is a significant and recurrent challenge. Within a large company, the different mine sites and projects compete for the available funds for drilling. To maximize a project value to its company, a drillhole location optimizer can be used as an objective tool to compare drilling campaigns. The fast semi-greedy optimizer presented here can allow for the obtention of close to optimal solutions to the coverage problem with up to three orders of magnitude less computing time needed than with integer programming. The heuristic approach is flexible as it allows dynamic updating of block values once new drillholes are selected in the solution, as opposed to existing methods based on static block values. The block values used for optimization incorporate kriging estimate and variance, estimate of indicator at cutoff grade and distances to existing or newly selected drillholes. The heuristic approach tends to locate new drillholes within the maximum risk areas, *i.e.*, within less informed zones predicted as being ore zones. Applied to different deposits, it enables, after suitable normalization, comparison of different drilling campaigns and allocation of budgets accordingly.

3.2 Keywords :

Geostatistics, Infill drilling, Risk, Semi-greedy optimization, Heuristic.

3.3 Introduction

During exploration and mining phases, the costs of infill drilling represent a significant portion of total project costs. Deposit definition is usually the result of a staged process with confidence in mineral resource estimates being gradually upgraded from inferred to indicated

and measured categories (Canadian National Instrument 43-101). Uncertainty on a deposit can represent a high risk for the company in terms of investment (e.g., construction of a new mine or extensions) or actual production. Parker [1] indicates that changes in the net present value (NPV) can be six times the variations on grade, so that a 10% error in grade estimation can result in a 60% variation in NPV. In order to decrease the risks associated with resource estimation, obtaining more information through further drilling is the standard practice. Face with limited budget, optimization of drillhole localization becomes of strategic importance for mining companies.

Infill drillhole planning is generally performed manually by an experienced geologist. The drilling plan is then submitted for approval. To reduce subjectivity inherent to the exercise, drillhole location optimization has been studied since the late 1970s ([41, 115]). These attempts focused mainly on selecting drillholes so as to minimize kriging variance (see [116–119]. The main limitation of kriging variance is that it depends only on data location and the variogram model, but not on actual grade estimates. Mohammadi et al. [118] proposed to use the product of kriging variance and kriged grade estimate to favor localization of new drillhole in high-grade areas with large uncertainty.

Conditional simulations have also been considered for drillhole location optimization. They allow generation of equiprobable realizations of the random function [4]. Both grade and domains (or lithologies) can be simulated with different algorithms such as sequential simulation [120], turning bands [121], moving average [4] or simulated annealing [122] among many other methods. For each node or block, multiple realizations can be generated, providing a measure of local uncertainty. Pilger et al. [123], Pinheiro et al. [49] or Zagré et al. [17] used various forms of conditional simulations to define the uncertainty parameter to be minimized.

The drill hole optimization problem was modeled by Bilal et al. [16] and Zagré et al. [17] as a set covering problem. In this approach, if a block distance to the closest drillhole is less than a selected distance threshold, it is considered fully covered ; otherwise, it is considered fully uncovered. It is then possible to formulate the best coverage for a set of blocks by a set of drillholes as an integer program as mentioned in Bilal et al. [46]. The coverage approach has two main drawbacks. First, it is very slow for large problems. Second, it uses a crude definition of coverage. More realistically, a block should be considered as only partially covered when some drillholes are found close to the block.

We propose two main improvements to the previous approaches. First, we define an objective function focusing on the riskier parts of the deposits and allowing easy updating of block values (BVs) without having to redo the kriging every time a drillhole is selected. Second, we present a new heuristic approach using a semi-greedy method that executes quickly and

allows, for the first time, a partial coverage of blocks. The proposed approach is applied to a real deposit where it is also compared to integer programming solution for execution time and optimality gap in the special case of a 0–1 weighting function.

3.4 Methodology

3.4.1 Objective Function

The proposed block value (BV) function is :

$$BV(x|h) = Z_k^*(x) \cdot \sigma_k^{2^*}(x) \cdot I_a^*(x) \cdot \Pi_{i=1}^n w_1(x|h_i) \tag{3.1}$$

where h is the set of existing drillholes and already selected drillholes, n is the cardinal of the set h, $BV(x|h)$ indicates that the BV depends of the set h, $Z_k^*(x)$ is the kriged grade estimated at block x, $\sigma_k^{2^*}(x)$ is the block kriging variance and $I_a^*(x)$ is the indicator value [124] estimated at block center using cutoff grade a. Only existing drillholes (not the selected ones) are used to compute $Z_k^*(x)$, $\sigma_k^{2^*}(x)$ and $I_a^*(x)$. Hence, these quantities have to be computed only once. Only the weighting function $w_1(x|h_i)$ varies, between 0 and 1, with the selected drillhole h_i. It is 1 for blocks outside the h_i, and it is 0 for a block x that is crossed at its center by drillhole h_i. Equation 3.1 gives high values to blocks with high-grade estimate $Z_k^*(x)$, high uncertainty $\sigma_k^{2*}(x)$ and high probability $I_a^*(x)$ to be above the cutoff grade. For a given grade estimate and uncertainty, it seems logical to prioritize covering of blocks that can be exploited at a lower cutoff grade (hence with higher $I_a^*(x)$). Once a given drillhole h_i is selected, the likelihood of selecting another drillhole in the same area diminishes drastically because of the function $w_1(h_i)$. This avoids clustering of drillholes in a highgrade area.

Similarly, we define the coverage of a candidate drillhole h_i as :

$$K(h_i) = \sum_{j=1}^{nb} BV(x_j|h) \cdot w_2(x|h_i) \tag{3.2}$$

where nb is the number of blocks and $w_2(x|h_i)$ is a decreasing function, which is equal to 1 when x is crossed at its center by h_i and is 0 for blocks outside the distance of influence of drillhole h_i. Note that w_1 diminishes the value of a block that is already informed by at least one nearby drillhole and w_2 decreases the coverage of a block that is located farther from the candidate drillhole. The classical approach used in coverage optimization (e.g. Zagré et al. [17]) implicitly uses w_1 as a 0–1 step function at the drillhole distance of influence and w_2 as a 1–0 step function at this same distance. Our formulation is, therefore, more general as

it allows continuous variation of the BV and drillhole influence.

The objective function to minimize is simply the sum of BVs in Eq. 3.1 given the existing and added drillholes, thus :

$$OF = \sum_{j=1}^{nb} BV(x_j | h \cup h_{add}) \tag{3.3}$$

In the semi-greedy approach, we do not seek to minimize exactly Eq. 3.3 but rather to find good solutions by sequentially choosing drillholes with high $K(h_i)$ values in Eq. 3.2 as described in the next section below.

The quality of a solution is best expressed by its coverage value :

$$C_{Sol} = \frac{OF_0 - OF_{Sol}}{OF_0} \tag{3.4}$$

where OF_0 is the value of the objective function using only the existing drillholes and OF_{Sol} is the value of the objective function obtained with the solution proposed. Figure 3.1 presents a theoretical 2D example of the dynamic updating of weighting schemes. The example area has an extent of 100 by 100 units, and two existing virtual data (red squares) were positioned at [0, 50] and [100, 50]. The two upper images present w_1 (left) and w_2 (right) for a first tested point positioned at [50, 25]. The red circle lines are drawn at 25 units from the existing and tested positions (distance of influence). The two bottom images show the updated weights at a second point (at [50, 75]) once the first point is accepted.

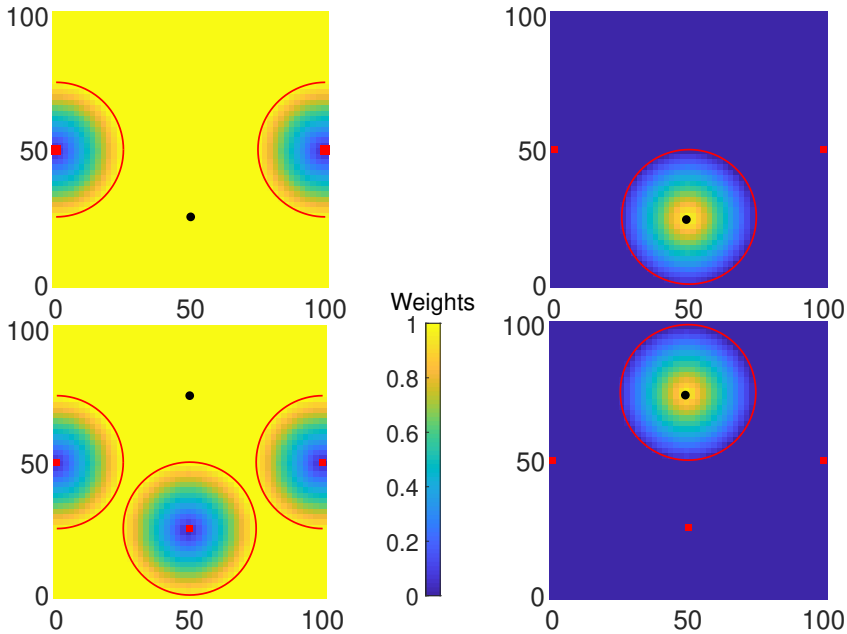


Figure 3.1 Weighting scheme theoretical example. Upper row presents w_1 (left) and w_2 (right) for the point (50, 25) (black dot) accounting for existing or already selected drillholes (two red dots). Bottom row shows the updated weights for point (50, 75) once point (50, 25) has been selected.

3.4.2 Semi-greedy Optimization

The drillhole positioning optimization problem belongs to the family of the "set covering problem" (Bilal [125]), one of the 21 NP-complete problems defined by Karp [15]. Some heuristic algorithm used for drillhole positioning optimization includes simulating annealing (Pinheiro et al. [49]), genetic algorithm (Soltani et al. [126]), bees colonies (Jafrasteh and Fathianpour [119]), particle swarm (Fatehi et al. [127]) and tabu search (Zagré et al. [17]). These algorithms are CPU time intensive for large 3D models. To circumvent this limitation, a fast and flexible semi-greedy heuristic algorithm (Hart and Shogan [128]) is proposed.

The algorithm was developed in MATLAB software and works for either 2D datasets (best suited for a single lens or a tabular deposit) or 3D datasets. In 3D, a series of possible collar positions, orientations, dips and lengths of drillholes are evaluated to provide close to an optimal solution.

3.4.3 Semi-greedy algorithm

The semi-greedy algorithm works sequentially, one drillhole at a time. A new drillhole is randomly selected among a list of *nlist* best drillhole candidates. A new selected drillhole is added to the set of already selected drillholes. The blocks values are updated, and the coverage function of the remaining candidate drillholes is recomputed and sorted to account for the newly selected drillhole. The sequential selection continues until the drilling budget is exhausted. Because there is a random component involved when selecting one drillhole from the sorted candidate list, the process can be repeated *ntrial* times, each time providing a different subset of selected drillholes. In the end, the best solution among the *ntrial* is kept. In the semi-greedy algorithm, there are three parameters to specify : the size *nlist* of the list of best drillhole candidates, the number *ntrial* of trials to run and the distance of influence of a drillhole (*dmin*).

- Initialization. The input consists of :
  - blocks to consider with their block-values $BV(x|h)$ (see Eq. 3.1)
  - existing drillholes
  - candidate drillholes
  - length of best candidate list (*nlist*)
  - number of trials (*ntrial*)
  - distance of influence of a drillhole (*dmin*)
  - available drilling budget
- For each trial, until drilling budget is exhausted :
  - compute $BV(x|h)$ per block x considering the existing and already selected drillholes h (see Eq. 3.1)
  - compute coverage value $K(h_i)$ (see Eq. 3.2) for the remaining candidate drillholes and sort it in decreasing order of coverage
  - randomly select one among the *nlist* best drillhole candidates
- Compute OF value (with Eq. 3.3) for current solution. The solution with the best coverage among the *ntrial* solutions (see Eq. 3.4) is kept.

When updating the blocks values, only function $w_1(x|h_i)$ modifies the BVs. Kriging estimates (grade and indicator) are left unchanged as the selected drillhole is not yet materialized. Although kriging variance can be recomputed to account for the reduction in uncertainty brought by the newly selected drillhole, this was not implemented in our approach. The weight function $w_1(x|h_i)$ essentially fulfills this purpose by reducing the value of a block already informed by a nearby drillhole (real or selected). This avoids repetition of kriging for each potential candidate drillhole, which drastically reduces the CPU time required.

The Role of *nlist* Parameter

The parameter *nlist* represents the length of the list of the current best candidates. When $nlist = 1$, we have the greedy solution, i.e., at each step we take the best drillhole available. When $nlist$ = number of candidate drillholes, the selection is purely random.

A simple example illustrates that the semigreedy approach provides better solutions than a greedy approach. Consider a series of potential drillholes with unit separation along a line and values : 1.1–2.1–3.1–4–3–2–1. To simplify, we use the step weighting function w_2 (with $dmin = 1.5$) and we suppose two drillholes are needed and can only be located at the above points. The greedy approach first selects point 4 (best) and then point 2.1 (as points 3.1 and 3 are already fully covered by point 4, their value becomes 0). The total coverage value is $1.1 + 2.1 + 3.1 + 4 + 3 = 13.3$. If we use a semi-greedy approach say with $nlist = 2$, points 3.1–4 are candidates for the first draw. Assume 3.1 is selected, the next draw will be between 3 and 2 (as 2.1 and 4 are already fully covered, their value is updated to 0). Suppose point 2 is selected. The total coverage is then $2.1 + 3.1 + 4 + 3 + 2 + 1 = 15.2$, which is a significant improvement over the greedy choice. Note that, in this simple example, the optimal choice is to select points 2.1 and 3, which give a total coverage of $1.1 + 2.1 + 3.1 + 4 + 3 + 2$=15.3. This optimal choice would be available in the semi-greedy approach with $nlist \geq 3$ and will eventually be selected provided we draw often enough (i.e., large *ntrial*).

Choice of Distance of Influence *dmin*

The choice of the distance of influence can be based on the geologists knowledge and experience with the consideration of the variogram effective correlation range, the geometric features of the deposit, similar outcrops or of already exploited areas. A noticeable advantage of our approach is that it enables to have a decreasing weight function from 1 for distance zero to 0 at the distance of influence, instead of only a step function as in Zagré et al. [17].

For linearly varying weight functions $w_1(x)$ and $w_2(x)$, increasing *dmin* diminishes $BV(x|h)$ for each block but increases the weights given to each block when computing drillhole coverage with Eq. 3.2. The solutions obtained are relatively robust to the choice of dmin for linearly varying weight functions (see next section).

3.5 Case Study

3.5.1 Datasets Presentations

The proposed methodology and algorithm were applied to two different datasets from the Rosebel Gold Mine (RGM) in Suriname (South America). The RGM is situated within the greenstone belt of the Guianese shield, and it is composed of multiple orogenic gold deposits (Daoust et al. [129]). RGM produced approximately 5 Moz of gold since the beginning of its commercial production in 2004.

To assess both 2D and 3D capabilities of the software, two partial datasets were extracted from two different deposits from the RGM complex. Multiplicative factors were applied to the gold grade values to protect confidentiality without altering any of the results or conclusions in the present research.

3.5.2 2D Application

The 2D dataset corresponds to a subzone of the RGM's deposit. The ore zone selected corresponds to a vertical shear zone extending approximately over 400 m along strike and 500 m vertically. In total, 96 drillholes crossing the ore zone were selected to provide initial information.

The gold grades of the intersecting portions of the drillhole, within the ore zone, were calculated as a single composite, and the horizontal width of the ore zone was saved. A 10 x 10 m regular grid was created to define blocks within the deposit. The same grid was used as possible definition drillhole locations. Variogram models were used to interpolate the variables grade, thickness and grade indicator above a cutoff grade of 1 g/t on the grid. The BV function Eq. 3.1 was then calculated for each block including ore thickness. The basic statistics of the initial BVs are presented in Table 3.1. Locations farther than 100 m from existing drillholes were considered too remote to be considered for definition drillholes. A crisp 1–0 coverage distance was applied to each drillhole to allow direct comparison with integer programming results. A *dmin* of 26 m, specific to the Rosebel case, was selected as a compromise : it is small enough to be compatible with full coverage but large enough to allow some overlapping between drillhole locations. The *dmin* to be used is expected to vary depending on geological continuity, drill spacing and smu size specific to the case under study. The impact of the *dmin* distance on results is assessed in the next section below.

Tableau 3.1 Block values basic statistics for the 2D dataset

Number of Blocks	Minimum	Maximum	Mean	Standard Deviation
873	0	100.2	27.0	22.9

Figure 3.2 presents a 2D longitudinal view of the existing drillholes (black points), and the BVs obtained considering the existing drillholes. The selected drillholes by semi-greedy algorithm (red points) for sets of 1, 5, 10 and 20 drillholes are indicated. The selected drillholes effectively cover the high values areas. It can be noted that the exact position varies with the number of drillholes recommended. Hence, for example, the best location when $N = 1$ is not necessarily retained for solutions corresponding to a larger N, which matches the expected behavior (Fig. 3.2).

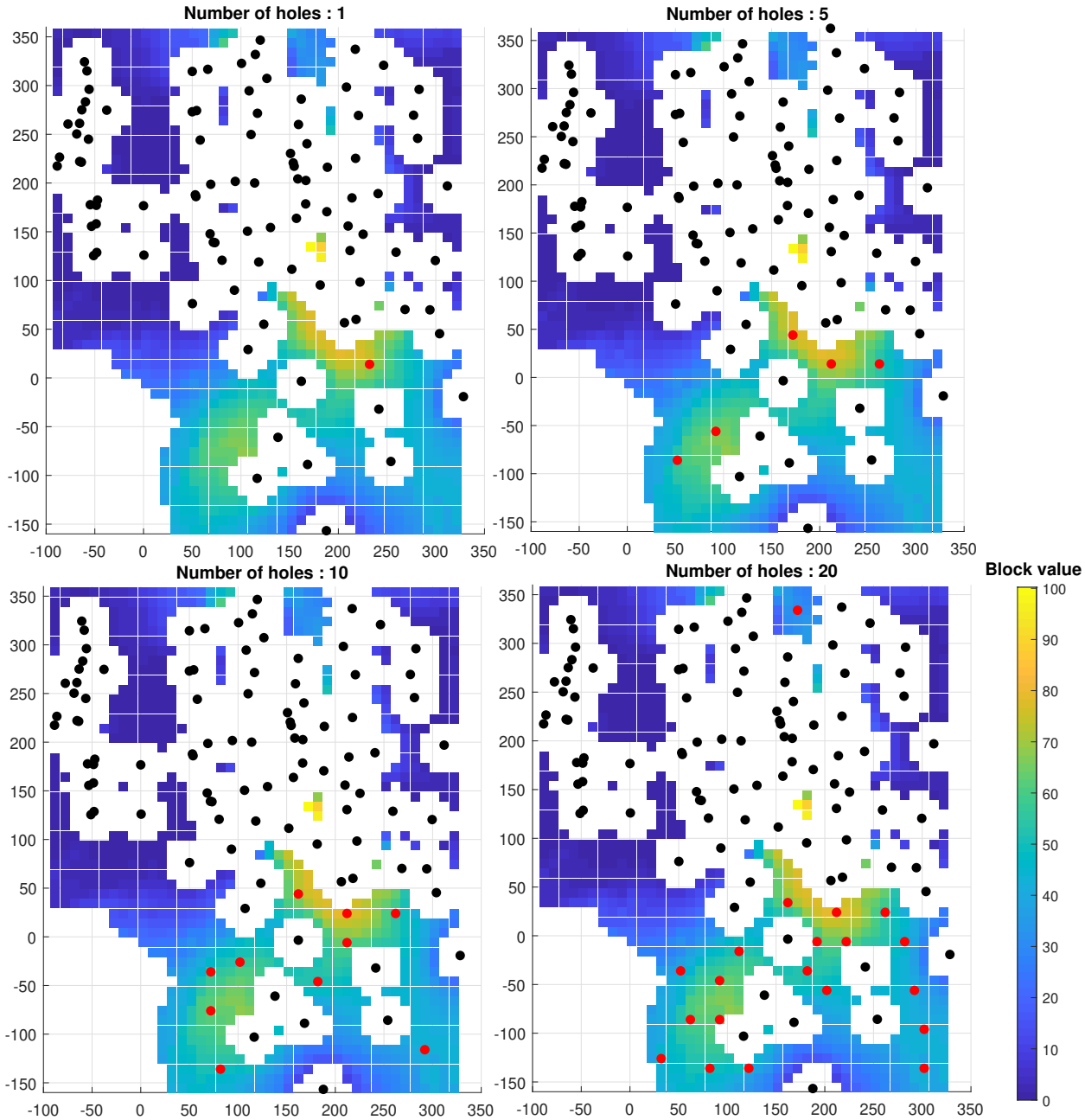


Figure 3.2 Longitudinal 2D results view. Black dots correspond to existing drillholes with ore intersects. Red dots are the resulting drillholes from the optimization. BV considering only the existing drillholes is illustrated.

With a 0–1 weight function, we can compare the best values of the semi-greedy algorithm to the exact optimal solution obtained by integer programming. For this, we use the linear program described in Zagré et al. [17]. We stress that the integer programming approach is limited to the step function weighting scheme and does not allow for partial coverage of

blocks, contrarily to our semi-greedy approach.

Figure 3.3 presents the coverage expressed in percentage of the total BVs, as a function of the number of drillholes optimized (a proxy for budget). With $dmin = 26$ m and a maximum distance of 100 m from existing drillholes, approximately 80 drillholes suffice to cover entirely the area. Both curves initially show a fast increase in coverage and then a slower rate of improvement, due to then filling the areas with lower BVs. The blue line represents the optimality gap for each solution expressed as the ratio $\frac{C_{sg}-C_{opt}}{C_{opt}}$ (C_{sg} is coverage with semi-greedy ; C_{opt} is coverage with integer programming). The maximal optimality gap reaches around 13% for N between 15 and 25. Afterwards, the optimality gap decreases progressively.

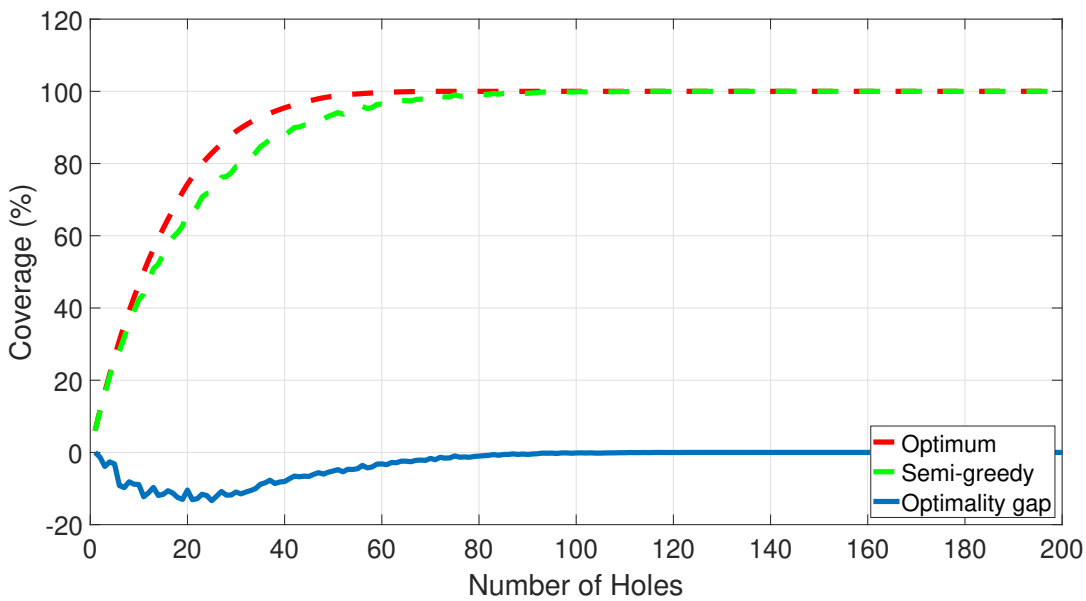
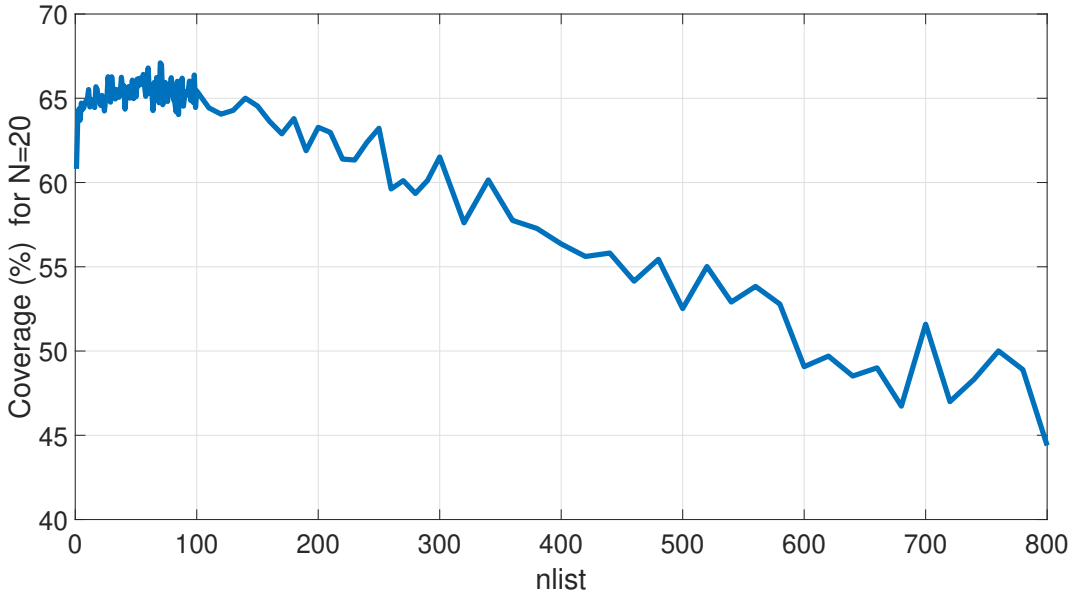


Figure 3.3 Coverage as a function of number of drillholes. Semi-greedy with $ntrial = 100$ and $nlist = 50$.

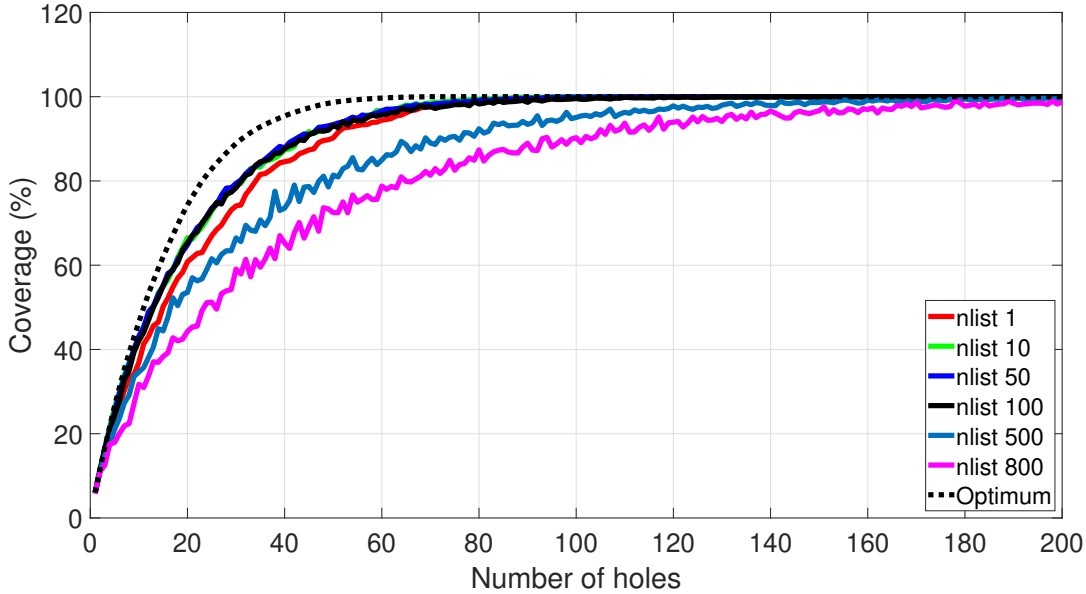
Two parameters mainly influence the quality of the results and the calculation time of the semi-greedy algorithm : *nlist* and *ntrial*. Figure 3.4(a) shows the semi-greedy coverage results for 20 drillholes as a function of nlist (using $ntrial = 300$). Increasing nlist rapidly improves the quality of the solution at the beginning, but when *nlist* becomes too large, the total coverage decreases. Indeed, setting *nlist* equal to the total number of drillholes available would be equivalent to randomly selecting N drillholes and keeping the best draw among *ntrial* trials, a poor method of optimization. Conversely, setting *nlist* to 1 corresponds to a purely greedy solution, again, not the best optimization approach. An easy method to select *nlist* is to simply run the algorithm with a series of increasing values for *nlist* until

the maximum has been identified. Because the semi-greedy algorithm is fast, this approach is tractable.

Figure 3.4(b) presents the results of applying the semi-greedy algorithm with different *nlist* for 1–200 drillholes (with *ntrial* = 300). The best results were obtained for *nlist* between 10 and 100 for all number of drillholes to be selected. This suggests that a good choice for *nlist* is relatively independent of the drilling budget represented here by the number of drillholes.



(a)



(b)

Figure 3.4 (a) Coverage for 20 drillholes as a function of *nlist*, using *ntrial* = 300. (b) Optimization coverage results for 1–200 drillholes for different *nlist* with *ntrial* = 300.

The other parameter that can be adjusted is the number of trials *ntrial*. Using a fixed *nlist* = 50, Fig. 3.5 demonstrates convergence to a decent solution after a relatively low number of trials (i.e., 25–50). Increasing *ntrial* to above 100 improves the results only marginally but

increases computation time linearly.

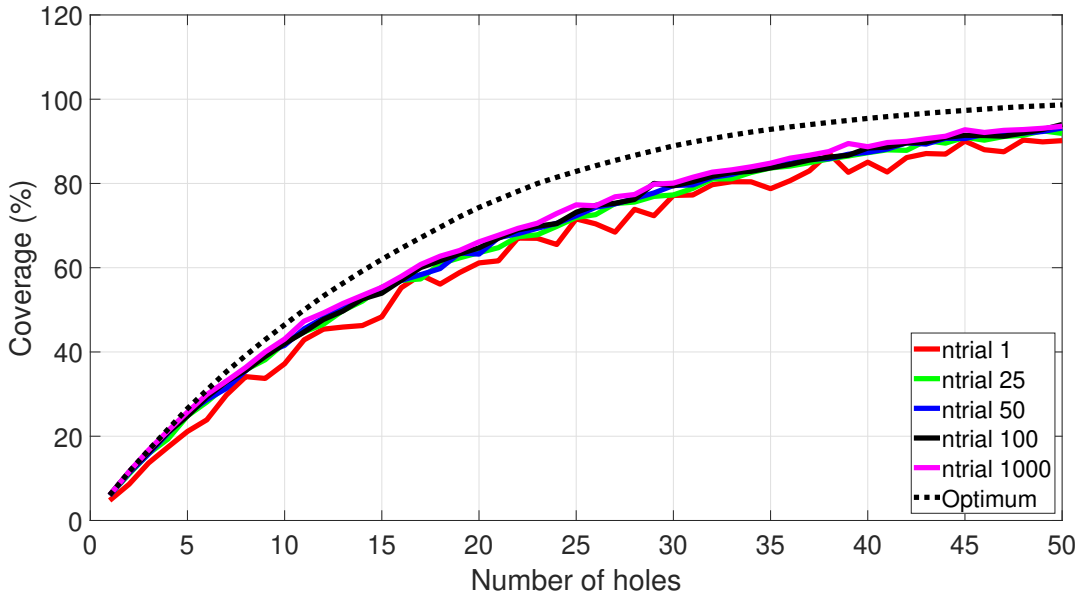


Figure 3.5 Coverage results for varying *ntrial* and number of drillholes, with $nlist = 50$.

3.5.3 Influence of dmin for Step or Linear Weight Functions

The use of step functions in the previous sections was done to allow direct comparison of the semi-greedy solutions to the optimal solutions obtained by integer programming. It is obviously more realistic to assume that the coverage of a block due to the presence of a drillhole decreases regularly as a function of distance. The influence of choice of dmin for both types of weight functions is examined.

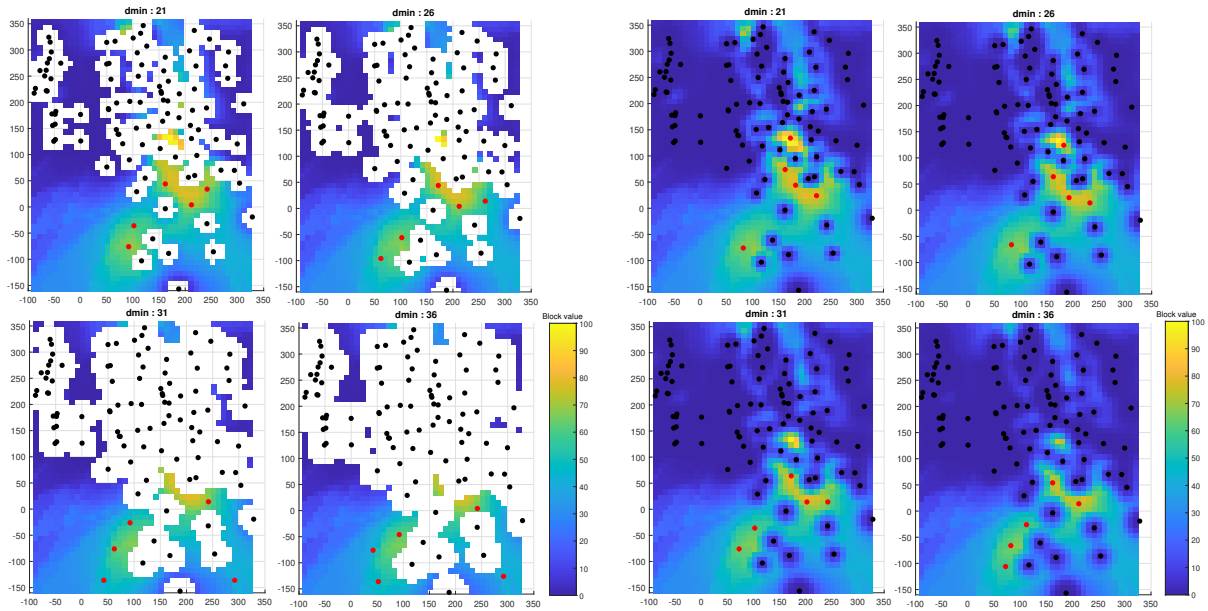
Results obtained for five new drillholes with *dmin* values of 21, 26, 31 and 36 used with step and linear functions are shown in Figure 3.6. The corresponding coverage value obtained (computed with Eq. 3.4) is given in Table 3.2. As expected, the choice of *dmin* is more influential for the step functions than for the linearly varying functions. For the step functions, increasing *dmin* to 26 m removes the drillhole located in the central part of the deposit as this area becomes covered by the existing drillholes. Increasing further *dmin* increases separation between drillholes and locate them more on the periphery. The coverage C_{sol} increases with *dmin* as total BV to be covered diminishes, and the area covered by each drillhole increases. We retained a value of $dmin = 26$ m because it is close to typical separation for definition drilling.

In the linear case, a drillhole remains selected in the central area with $dmin = 26$ m (contrary

to the step function). Further increasing *dmin* also tends to increase drillhole separation and place them more on the periphery but less so than as the step function. These results indicate lesser dependency of C_{sol} and drillhole location with the linear functions than with the step functions.

Tableau 3.2 Coverages for five drillholes as a function of *dmin* for the step and linear weight functions

dmin (m)	21	26	31	36
Step functions	12%	21%	29%	37%
Linear functions	4%	6%	8%	11%



(a) Step functions

(b) linear weight functions

Figure 3.6 Longitudinal 2D results for five drillholes and increasing dmin with step (a) and linear (b) weight functions. Existing drillholes (black dots) and drillholes selected (red dots) by semi-greedy algorithm with $nlist = 50$ and $ntrial = 1000$. Map of BVs considering only the existing drillholes is illustrated.

3.5.4 3D Application

The 3D dataset corresponds to one of the RGM active pits. The mineralization is composed of a succession of sheared zone with quartz veins of different orientations. The extension of the mineralization is about 1 km in easting and 200 m on northing with vertical extension of

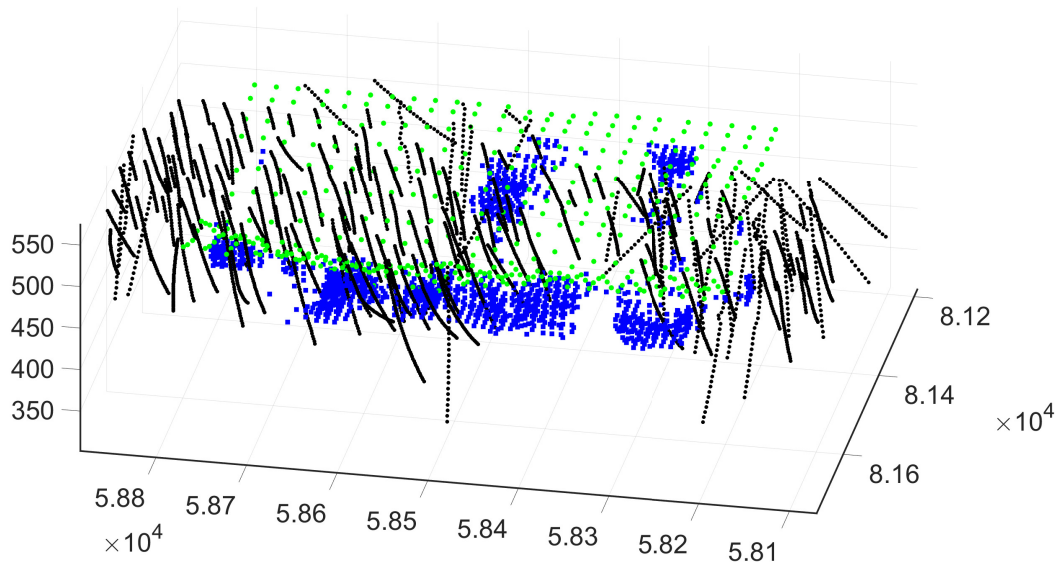
about 300 m. The mineralization has been partly mined out, but some pushbacks are planned in the coming years. The actual topography was used to generate possible drillpad locations on a regularly spaced 25-m pattern.

A restricted area of 400 m of extension (Easting) was selected within the entire 3D area. In total, 178 diamond drillholes totaling 23,635 m and 38,556 m composites were available in this zone. The deposit was discretized in blocks of size 10 x 10 m horizontally x 8 m vertically. BVs were calculated using Eq. 3.1 for all blocks with a cutoff of 0.3 g/t to define the ore indicator variable. The BVs statistics of the 3D case are presented in Table 3.3. They differ significantly from the statistics presented in the 2D case, because of the varying grade, spacing and economical cutoff (indicator). Moreover, thickness does not intervene in the 3D case.

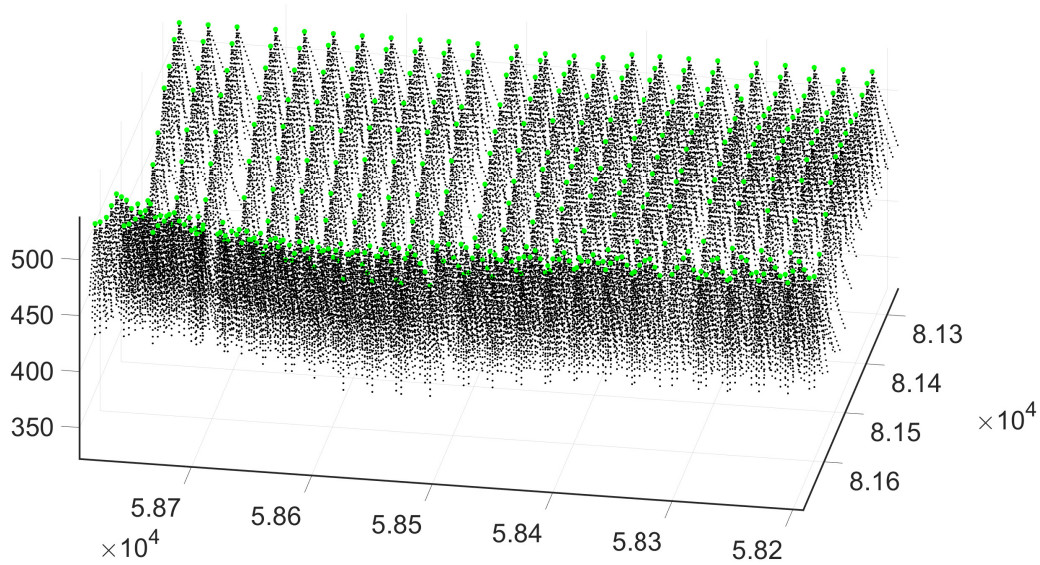
Tableau 3.3 Block values basic statistics for the 3D dataset

Number of blocks	Minimum	Maximum	Mean	Standard Deviation
19,447	0	1.67	0.031	0.078

The zone of influence of each drillhole was also set at 26 m for w_1 and w_2 functions. These were defined as step functions to allow comparison with integer programming results. Figure 3.7(a) shows the existing drillholes composites (in black) with $BVs > 0.1$ as blue squares and the possible collar positions extracted from topography in green. Figure 3.7(b) presents the evaluated drillpads as well as all the generated composites from drillholes. In order to account for mineralization orientation as well as existing drillholes, drillhole azimuth orientations of 165, 180 and 195 were forced, associated with plunges of 60, 75 and 90, respectively. A fixed drillhole length of 102 m was used. With these parameters, a total of 3,672 candidate drillholes were defined.



(a)



(b)

Figure 3.7 (a) A total of 178 existing drillholes composites (black dots) and blocks with $BV > 0.1$ (blue squares), within the selected area. (b) Trace of the 3,672 drillholes that fulfill constraints on drillpad locations (green dots), direction and dips.

Optimization result for $N = 10$ drillholes (totaling 1,020 m) is presented in Figure 3.8, using $nlist = 20$, $ntrial = 30$.The proposed drillholes (in red) are localized within area of high BV

and spaced by at least 26 m from each other and from existing drillholes.

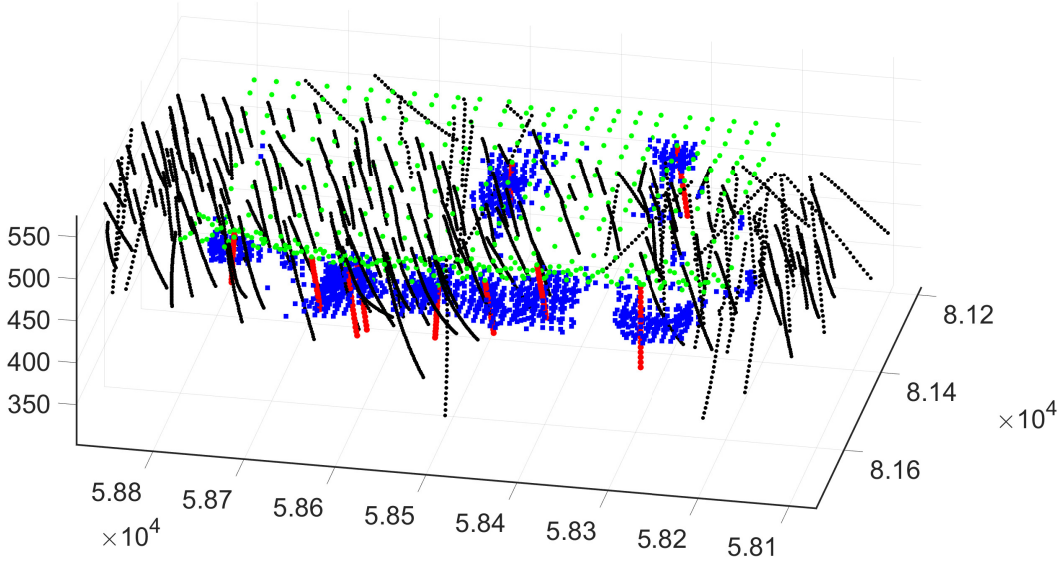


Figure 3.8 Topography (green), existing drillholes (black), 10 selected new drillholes (red) and blocks with $OF > 0.1$ (blue).

Figure 3.9 shows the coverage as a function of the number of drillholes for optimal integer programming solution compared to the semi-greedy solution for the 3D case dataset. The gap of optimality is generally less than 10%. The proposed algorithm ran all the cases shown in the figure in about 1,240 s, compared to the 455,403 s (about 127 h) it took for the integer programming.

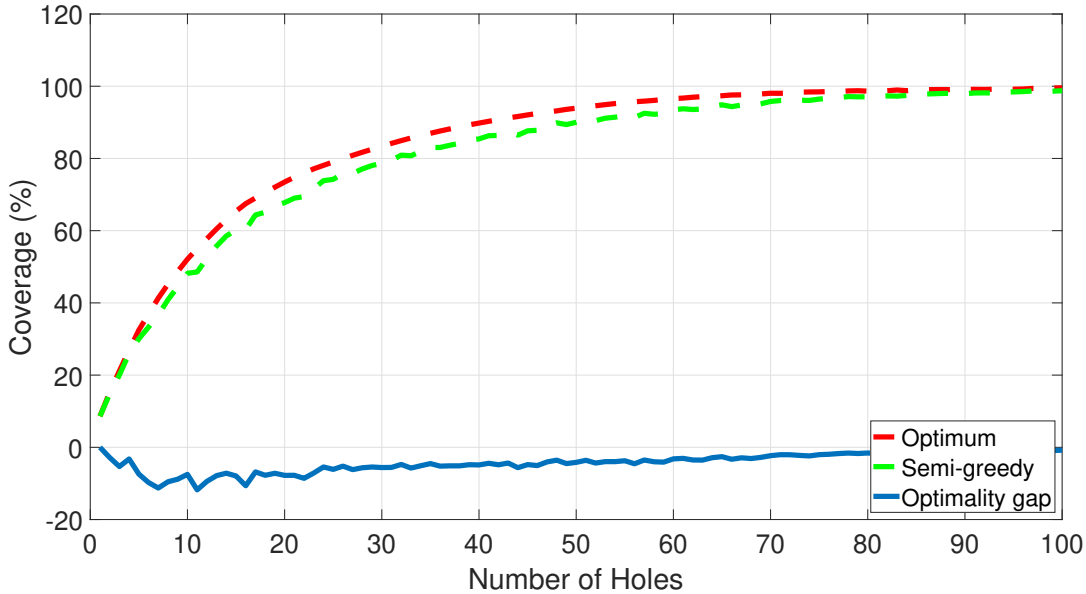


Figure 3.9 3D coverage as a function of number of drillholes.

3.5.5 Simultaneous optimization

An interesting application of the proposed method is to optimize two or more deposits simultaneously and use the results to support budget allocation. This requires a full or partial normalization of the components defining Eq. 3.1 for the BVs : grades, variograms sills and/or indicator cutoffs. It is then possible to optimize the coverage of the riskiest areas for both deposits simultaneously. The choice of the normalization depends on the main purpose of the definition campaign. As an example, Figure 3.10 presents the results of a simultaneous optimization for the previous 2D and 3D examples. Variograms sills were normalized to 1, and cutoffs were set to 1 and 0.3 g/t for the 2D and 3D datasets, respectively. In this example, the composite grades were normalized by dividing them by the mean grade of all composites in the respective datasets. A flat topography was added to the 2D dataset so that it could be treated as a 3D case. Both datasets were then combined, and the semi-greedy algorithm was applied to the combined dataset. In this application, $w1$ and $w2$ were set as a linear function [from 0 to 1 using $dmin = 26$ m].

Fig. 3.10 shows the total coverage obtained for both projects combined (red line) as a function of the allocated budget (in m) from 500 to 10,000 m. For a relatively small budget of up to 3,000 m, most of the added drillholes were positioned within the 3D zone. When the drilling budget exceeded 3000 m, then most of the additional drillholes were located in the 2D project as most blocks with high OF values in the 3D case were already partially covered by the

drillholes added in the initial steps.

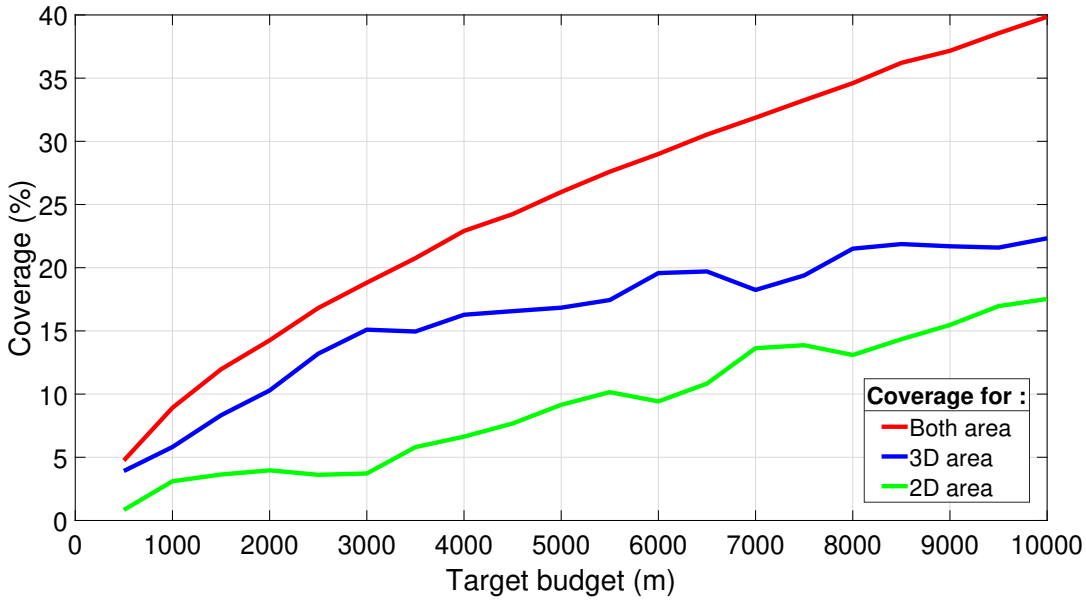


Figure 3.10 Coverage as a function of drilling budget (in meters) for the two datasets optimized simultaneously

3.6 Discussion and Conclusions

The most common practice for drillhole optimization consists of minimizing the kriging variance or statistics extracted from conditional simulations results. Often, professionals simply implement systematic and regular spacing regardless of any geostatistical parameters. These approaches do not capture entirely the economic risks of a project or operation, which generally lie within the high-grade ore zones. An excess or lack of drillholes within the areas of risk is possible. This would correspond to a loss for the company (either as a risk or opportunity). The proposed OF function incorporates economical risk by maximizing the coverage of additional definition drillholes. Other OF could have been used with our heuristic method. As an example, conditional variances obtained by simulation could replace kriging variance. Here, we limited our study to a reasonable OF reflecting the main goals to achieve for the Rosebel case study : avoiding drillhole clustering, targeting primarily the areas with expected high grades, and reflecting the economic factors by the probability to be ore material. More research is needed to compare the utility of different OF variants for decision purposes in a variety of geological and economical contexts.

In our test, the semi-greedy algorithm was close to three orders of magnitude faster than

integer programming. This computational advantage is expected to increase with larger instances. However, the main advantage relative to integer programming is to allow dynamic updating of blocks values as new drillholes are selected. We stress that both grade kriging and indicator kriging remain unchanged as updating is taken into account entirely by the weight functions w_1 and w_2. The use of w_2 function can be seen as a fast alternative to updating of kriging variances each time a new drillhole is tentatively selected. Updating of kriging variance would mean performing $(nfor - 1) \times ntrial$ local kriging on the blocks found in the neighborhood of each added drillhole. This would increase by many orders of magnitude the time required to run the algorithm. The semi-greedy approach allows partial covering of a block by a drillhole, something unavailable with integer programming. Partial covering can also be implemented in other heuristic approaches, but then computing time is expected to become an issue due to the iterative nature rather than sequential (as with the semi-greedy approach) of these algorithms.

Our results indicate that relatively small values of *nlist* and *ntrial* are required to get close to the best results reachable by using the semi-greedy algorithm. This finding is interesting as computing time increases with both *nlist* and *ntrial*.

For the particular case of a 0–1 coverage function, the optimality gap of the semi-greedy algorithm varies in the 0–13% range in our examples. This gap is deemed acceptable considering the lack of realism of the 0–1 coverage function, the long computing time of the integer programming approach, and the relative imprecision of the different components (kriging estimate, kriging variance and kriged indicator) appearing in the definition of the BV function.

Sensitivity to the choice of *dmin* was shown to be higher for the step functions than for the more realistic linear functions. This finding supports the use of partial covering introduced in this paper. For both step and linear weight functions, increasing *dmin* tends to spread the drillholes more on the periphery of the field. However, this effect is much stronger with the step functions than with the linear functions. The value of *dmin* was selected based on typical separation for definition drilling. It can be adjusted to account for the kind of deposit studied, the perceived geological continuity and the correlation range.

Our tests clearly indicate that our algorithm does not cluster drillholes in high-grade area despite the presence of grade estimate in the BV definition (Eq. 3.1). The weight function w_1 diminishes strongly the value of blocks surrounding a selected drillhole and thus prevents any risk of clustering.

A simple test showed that it is possible to consider two deposits at once to assist decision making for drilling budget allocation. This requires suitable normalization of the components defining the BVs. More research is needed to determine whether this can be extended to more

than two deposits.

The proposed approach can be customized to include additional knowledge about the deposit. For example, if a certain lithology is known to control mineralization, the BVs can be calculated only inside the 3D solid of the mineralization, the rest of the BVs being set to 0. Metallurgical values or geotechnical parameters can also be incorporated in the BVs' definition (Eq. 3.1).

The new simple semi-greedy algorithm proposed is proven to be an efficient and flexible tool for selection, under budget constraints, of drillholes providing good coverage. Being sequential rather than iterative, the algorithm is fast. It has allowed up to three orders of magnitude CPU time reduction compared to integer programming. It is also flexible as it allows partial coverage of blocks, which is more realistic than coverage functions used in previous studies. The proposed method provides a useful tool to support managers' decision about drillhole allocations in competing projects.

3.7 Aknowledgments

This research was made possible by National Research Council of Canada thru NSERC Grant (RGPIN-2015-06653). We thank IAMGOLD Corp. for providing data used in this study. The authors would like to thank Mehanaz Yakub for her help and advises.

CHAPITRE 4 ARTICLE 2 : A NEW GRADE CAPPING APPROACH BASED ON COARSE DUPLICATES CORRELATION

Historique de l'article : soumis le 10 septembre 2020 au *The Journal of the Southern African Institute of Mining and Metallurgy.*

Auteurs : Raphael Dutaut, Denis Marcotte

DOI : 10.17159/2411-9717/2019/v119n1a1

4.1 Abstract

In most exploration or mining grade data sets, the presence of outliers or extreme values represents a significant challenge to mineral resource estimators. The most common practice is to cap the extreme values at a predefined level. A new capping approach is presented that uses QA/QC coarse duplicate data correlation to predict the real data coefficient of variation (*i.e.* error-free CV). The cap grade is determined such that the capped data have a CV equal to the predicted CV. The robustness of the approach with regards to original core assay length decisions, departure from log-normality and capping before or after compositing is assessed using simulated data sets. Real case studies of gold and nickel deposits are used to compare the proposed approach to the methods most widely used in the industry. The new approach is simple and objective. It provides a cap grade that is determined automatically, based on predicted CV, and takes into account the quality of the assay procedure as determined by coarse duplicates correlation.

4.2 Keywords :

Geostatistics, outliers, capping, duplicates, QA/QC, lognormal distribution

4.3 Introduction

Outliers or extreme values are present in most mining grade data sets. They can reflect true small-scale spatial variability and/or sampling and analytical errors introduced by procedures. Extreme values are deemed undesirable in kriging as they could propagate assay errors over significant ore tonnages. It is common practice in the mining industry to reduce these grade values to (or cap them at) a lower top-cut value [10]. Cap value determination is

not straightforward and often remains subjective, especially for highly skewed distributions found in precious metal deposits.

Numerous methods have been proposed to address cap value determination, including :

— Top percentile [34] : Most likely the simplest method, where the cap value is a high percentile of the distribution, generally between the $99th$ to $99.9th$ percentile. Sometimes the cap value is based on *historical practices*, such as capping at 1 ounce of gold per ton (this practice was common in northern Quebec, Canada). Although simple, the choice of cap percentile is arbitrary and subjective, and it does not take into account the quality of the assays.

— Parrish capping [35] : The cap grade is selected such that the post-cap metal content of the assays above the 90th percentile represents less than 40% of the total metal and/or that of the assays above the 99th percentile represents less than 10% of the total. Although this method is repeatable, it can define overly conservative (i.e., very low) cap grades, especially for highly skewed deposits. In addition, the percentiles and the proportions used are arbitrary, and the method does not take into account the quality of the assay procedure.

— Log probability, [34] : A cumulative log probability plot, or more simply a data histogram, is used to identify a change in slope or a gap in the tail of the distribution. This is one of the most commonly used methods in the industry. It is relatively easy to interpret, but there are often multiple breaks/gaps in the distribution that make the choice of cap value subjective. In addition, the method is sensitive to the amount of data. For small data sets, the cap value tends to be fairly low. When more data is available, the gap generally appears at a much higher value. The method also does not take into account the quality of the assay procedure.

— Cutting curves [62] : The average grade for different cap values is plotted and an inflection point is visually selected by the practitioner. This method is fairly arbitrary, and the choice of inflection point is subjective.

— Central variation error from the cross validation of simulated local average [130] : A local cap value is selected from a volume of influence (determined based on data spacing and kriging parameters) such that the interpolated block average within the volume is equal to the simulated median. A given high grade can be capped at different values depending on the neighbouring data. The method is complex as it requires kriging and conditional simulation, both of which need a variogram, on raw data for kriging and on Gaussian grades for simulation. The raw data variogram is itself sensitive to extreme values. The Gaussian transform required for simulation is also sensitive to extreme values. Furthermore, the results are likely to be sensitive to the

choice of neighbourhood for both kriging and conditional simulation. In addition, this method requires coordinates, so it can't be used for conveyor, truck or grab samples, for example.

— Metal risk analysis using simulations (Parker, reported in [34]) : Monte Carlo simulations are used to simulate the high-grade distribution. This method makes it possible to generate a confidence level for a predefined (e.g., yearly) metal production volume. This simulation method suffers from basically the same shortcomings as the method of [130].

It is worth noting that some authors have proposed alternative methods to avoid capping ([131], [55], [53] and [59]). Although these methods are interesting from a theoretical point of view, they are rarely used in mining applications, where the capping of extreme values is still perceived as a best practice.

In this research, we propose the use of the correlation of the coarse duplicates to determine the cap value. The use of duplicates is now mandatory according to QA/QC guidelines and routine in the mining industry. Although Abzalov [132] and Rossi and Deutsch [34] describe various uses for duplicate samples, helping the determination of the cap level is not one of them. A rarely documented and supplementary question in resource assessments is whether capping should be performed before or after compositing.

Using a multiplicative lognormal error models, it is possible to determine the coefficient of variation of the true (unobserved) grades from the correlation between original and duplicate values (see section 2.2). We propose to determine the cap grade such that the newly capped population has the same CV as the CV determined from correlation between original and duplicate pairs.

After describing the multiplicative lognormal error model, the link between the lognormal parameters of the true grades, those of the errors, and the theoretical correlation between duplicates is derived. From this correlation, the predicted theoretical coefficient of variation of the grades is obtained and established as the target to be reached when determining the cap grade. The robustness to the sample length and the lognormal assumption is then assessed and the effects of capping before or after compositing is also analyzed. Finally, case studies of a gold and a nickel deposit are presented, and cap grades obtained using the proposed approach are compared to those obtained using some traditional methods.

4.4 Methodology

This section presents the multiplicative lognormal error model and the main results that are derived from it. Synthetic case studies are used to assess the method's robustness with regard to some assumptions and to measure the effects of capping before or after compositing.

4.4.1 Theoretical background

Assuming the following model is valid :

$$Z_o(x) = Z_v(x) \cdot t \tag{4.1}$$

where Z_o is the observed value at the x position, Z_v is the true value and t is the multiplicative error of the analysis (in this study both Z_v and t are assumed to be lognormal, a reasonable assumption for most precious and base metal deposits), Eq. 4.1 can be written :

$$Z_o(x) = exp(log(Z_v(x)) + log(t)) \tag{4.2}$$

with $log(Z_v) \sim N(\alpha, \beta)$ and $log(t) \sim N(\mu, \sigma)$, which gives

$$E[Z_o(x)] = e^{(\alpha+\mu+\frac{\beta^2+\sigma^2}{2})} \tag{4.3}$$

$$Var[Z_o(x)] = (e^{\beta^2+\sigma^2} - 1) \cdot e^{2(\alpha+\mu)+\beta^2+\sigma^2} \tag{4.4}$$

where α and β are the logarithmic mean and standard deviation of Z_v respectively, and μ and σ are the logarithmic mean and standard deviation of t respectively. Note that in the multiplicative lognormal error model, σ^2 is related to the variance of the duplicates' log-ratio :

$$\sigma^2 = \frac{1}{2} Var\left(log\left(\frac{Z_o}{Z'_o}\right)\right) \tag{4.5}$$

where Z_o and Z'_o are the observed original and duplicate assays (in short duplicates values). Hence, σ^2 can be estimated directly from the duplicates independently of the other parameters (α, β, μ). This result is not used hereinafter but is presented here for the sake of completeness.

If quality assurance and quality control (QA/QC) is performed and sampling theory guidelines are followed, there should be no or limited bias on mean grade. From Eq. 4.4 this implies :

$$\mu = -\frac{\sigma^2}{2} \tag{4.6}$$

4.4.2 Correlation between Duplicates

The theoretical correlation between duplicates in the multiplicative model is given by :

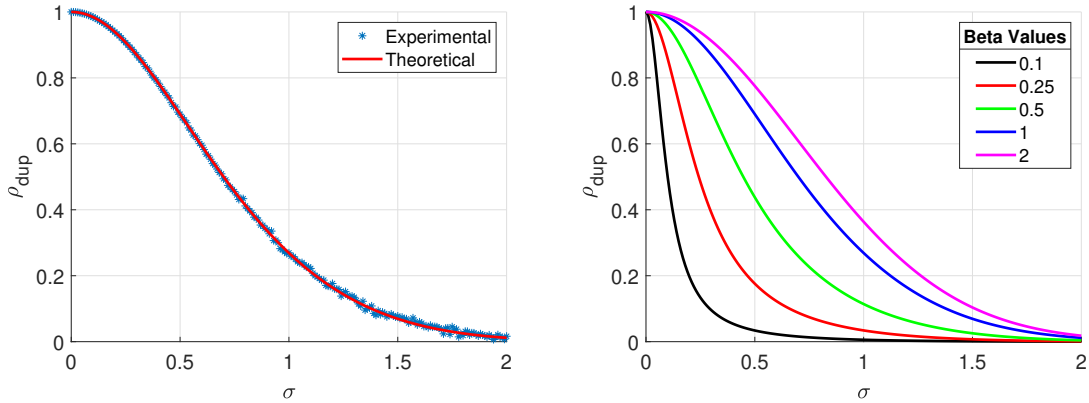
$$\rho_{dup} = \frac{(e^{\beta^2} - 1)}{(e^{\beta^2+\sigma^2} - 1)} = \frac{CV_v^2}{CV_o^2} \tag{4.7}$$

The first equality comes from $\frac{Var(Z_o)}{m^2} = \frac{Var(Z_o^{'})}{m^2} = (e^{\beta^2+\sigma^2} - 1)$ for the denominator and $\frac{Cov(Z_o,Z_o^{'})}{m^2} = (e^{\beta^2} - 1)$ in the numerator. The last equality in Equation 4.7 comes directly from the definition of coefficient of variation and Equations 4.3 and 4.4 for the denominator and the same equations with $\sigma^2 = 0$, for the numerator. With σ^2 estimated from Equation 4.5 and correlation between duplicates, it is possible to estimate β using Equation 4.7 and then α using Equation 4.3. Better yet, from Equation 4.7, the squared coefficient of variation of true and observed values are directly related to the duplicates' coefficient of correlation by :

$$CV_v^2 = \rho_{dup} CV_o^2 \tag{4.8}$$

where CV_v is the (unobserved) coefficient of variation of Z_v and CV_o is the (observed) coefficient of variation of Z_o. Hence, it is possible to estimate the coefficient of variation of the true grades using only correlation between duplicates and CV_o computed with all available samples.

Figure 4.1(a) shows the curve defined by Eq. 4.7, with $\beta = 1$, and simulated results. Figure 4.1(b) illustrates Eq. 4.7 for different β values.



(a) (b)

Figure 4.1 (a) Theoretical (red line, Eq. 4.7) and simulated ($n = 1,000,000$ duplicates with 200 blue dots) ρ_{dup} against σ for β^2=1. (b) Theoretical ρ_{dup} against σ for various β values.

Figure 4.2 compares the theoretical error-free coefficient of variation $CV_{theor} = \sqrt{exp(\beta^2) - 1}$ defined by β to CV_{pred} predicted using duplicates correlation ρ_{dup} and CV_{dup} in Eq. 4.8. Each point represents a different lognormal distribution with parameters uniformly drawn in intervals described in Table 4.1. As the number of duplicates increases, the spread of the points and the skewness of the conditional distribution of $CV_{pred}|CV_{theor}$ diminish. The number of duplicates required to estimate CV_{theor} to a given precision increases with CV value. Note that in Table 4.1, the minimum and maximum theoretical CV are $\sqrt{exp(0.5^2) - 1} = 0.53$ and $\sqrt{exp(1.5^2) - 1} = 2.91$, which cover most practical cases. Also, $\sigma \leq \beta$ and using Eq. 4.7, the minimum and maximum theoretical ρ_{dup} are 0.42 and 0.957 respectively. Using Eqs. 4.3 and 4.4, the minimum and maximum theoretical CV_o are 0.54 and 4.51 respectively, which cover most practical cases.

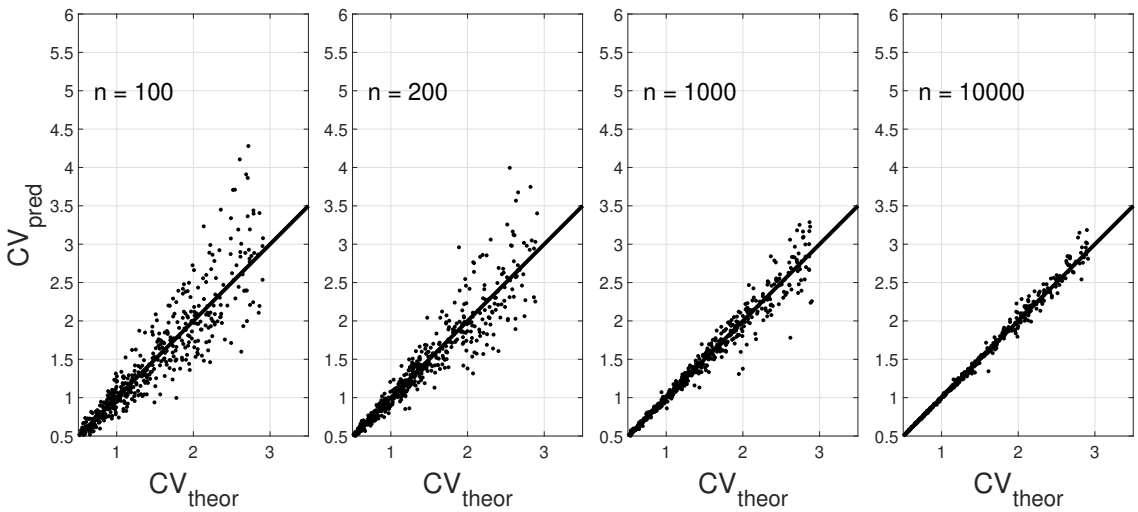


Figure 4.2 Theoretical CV and predicted CV using Eq. 4.8 for different lognormal parameters and an increasing number of duplicates (n=100, 200, 1,000, 10,000).

Tableau 4.1 Sampled intervals and definitions for parameters of the lognormal distributions of Z_v and t

Parameter	Interval or definition
α	[-0.5, 0.5]
β	[0.5, 1.5]
f	[0.2, 0.6]
σ	βf
μ	$-\sigma^2/2$

4.4.3 Capping based on correlation between duplicates

Equation 4.8 indicates the coefficient of variation of the observed duplicate grades is inflated by sampling errors. This suggests the following capping criterion : choose the threshold providing $CV_{cap} = CV_{pred}$ determined experimentally from the correlation between duplicates and the coefficient of variation computed using all available samples (original and duplicates) using Eq. 4.8. This criterion, although arbitrary, has the advantage of being objective, repeatable and simple to compute. It takes into account the quality of the assay procedure in determining the cap value, contrary to existing methods. Moreover, it does not require localization of the samples, contrary to the methods of Babakhni [130] and Parker [55].

Figure 4.3 shows the cap percentile applied to the distribution of Z_o that provides $CV_{cap} = CV_{pred}$ as a function of ρ_{dup} for different values of β. All curves are computed by numerical

integration of the truncated lognormal distribution. As expected, the cap percentile for each curve increases with the duplicates correlation, indicating that large error variances (lower duplicates correlation) require a lower cap value to obtain the desired coefficient of variation. Also, for a given correlation, the cap threshold increases with β, indicating a lower cap value is required for less skewed distributions of Z_v. Note that parameter α has no influence on determining the cap percentile.

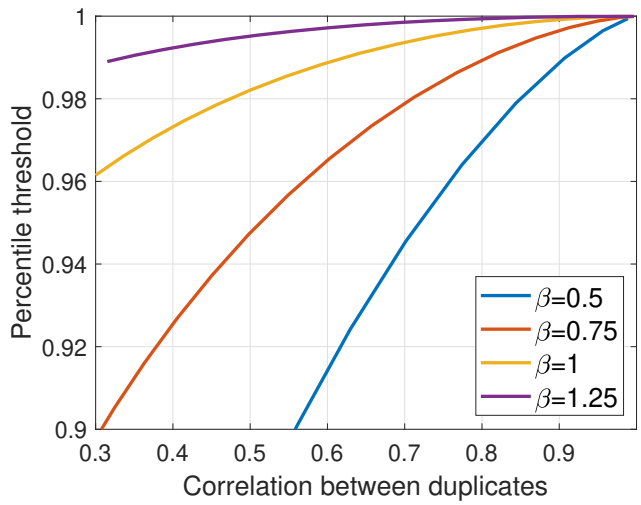


Figure 4.3 Percentile threshold against ρ_{dup} to have $CV_{cap}^2 = CV_v^2$ for various β values

4.4.4 Effects of variable assay supports

The results obtained so far assume the duplicate assays were on samples with identical supports. However, this might be not the case in practice for a variety of reasons (geological contacts, different sampling campaigns, technical drilling difficulties, etc...). It is therefore important to assess the effects of support variations on the statistics ρ_{dup} and β used to determine the cap percentile. Note that σ is the standard deviation of the error due to sample preparation and analysis ; hence, it is not related to the support. Only β varies with the support.

If the point variogram is available, it is possible to compute the variance of the regularized support for any support. Assuming that distribution remains lognormal with the same mean, it is possible to compute β_s, the logarithmic standard deviation associated with support s. The beta corresponding to a given support can be computed using classical geostatistical relations as :

$$\beta_s = \left\{ log\left(\frac{\tau_s^2}{m^2} + 1\right)\right\}^{1/2} \tag{4.9}$$

where $\tau_s^2 = \tau_0^2 - \overline{\gamma}(s,s)$, $m = exp(\alpha + \beta_0^2/2)$, $\tau_0^2 = m^2(exp(\beta_0^2) - 1)$, β_0^2 is the logarithmic variance at point support and $\overline{\gamma}(s,s)$ is the average variogram value within support s.

Example

Consider a typical gold deposit case with $\alpha = 0$, $\beta_0 = 1$, the mean $m = exp(0.5) = 1.65$ ppm, the point true grade variance $\tau_0 = m^2(exp(1) - 1) = 4.67$ ppm^2 and a spherical variogram with correlation range 20 m. The assay supports are half cores of lengths varying between 1 m and 5 m. The corresponding variances of core grades are 4.5 ppm^2 and 4.1 ppm^2 for 1 m and 5 m lengths respectively, corresponding to $\beta_{1m} = 0.99$ and $\beta_{5m} = 0.96$. Further computing duplicates correlation using Eq. 4.6 and $\sigma = 0.5$ for the two supports gives values of 0.69 and 0.68 respectively. Other values of σ provide similar differences in correlation and α has no impact on determining β_s and ρ_{dup}. As seen in this example, the effects of the support on β and ρ_{dup} are negligible when compared to the precision of the estimates of duplicates correlation and coefficient of variation.

4.4.5 Effects of departing from the lognormal assumption

Figure 4.3 indicates that realistic cap percentiles should be obtained for the lognormal case when the distributions are skewed (large β and CV) and the correlation between duplicates is high. When the skewness is low, the cap percentile tends to decrease, meaning that a significant percentage of the samples will be capped unless this is compensated for by a higher duplicates correlation corresponding to a better sampling preparation procedure in real deposits.

The lognormal assumption is realistic for many low-grade skewed distributions,which are typically encountered in precious and base metal deposits. However, nothing guarantees that the model applies to any particular data set. To get an idea of the sensitivity of the proposed approach to the lognormal assumptions, five different cases were simulated with a large quantity of data (1 Million). Table 4.2 describes the five simulated cases. Figure 4.4 shows the duplicates' scattergrams and the histograms of the error-free simulated Z_v and the observed Z_o. Despite the strong departure from lognormality and small CV_v of each case, the five cap values determined by our approach appear reasonable in terms of percentile of the observed grades. Moreover, the CV of the capped values match very well in each case with the CV of the error-free Z_v, indicating that Eq. 4.8 also approximates the true CV in the non-lognormal case well.

Tableau 4.2 Cases considered to evaluate the robustness of the approach

	Case 1	Case 2	Case 3	Case 4	Case 5
Description	Z_v lognormal, t normal	Z_v normal, t normal	Z_v normal, t lognormal	Z_v two lognormals, t lognormal	Z_v two lognormals, t lognormal, contamination
ρ_{dup}	0.87	0.78	0.67	0.75	0.77
Cap grade	49.4	5.0	8.7	5.3	4.9
Perc. cap	0.9996	0.946	0.935	0.940	0.940
Error-free CV	1.31	0.69	0.73	0.63	0.89
Capped CV	1.31	0.69	0.74	0.63	0.87

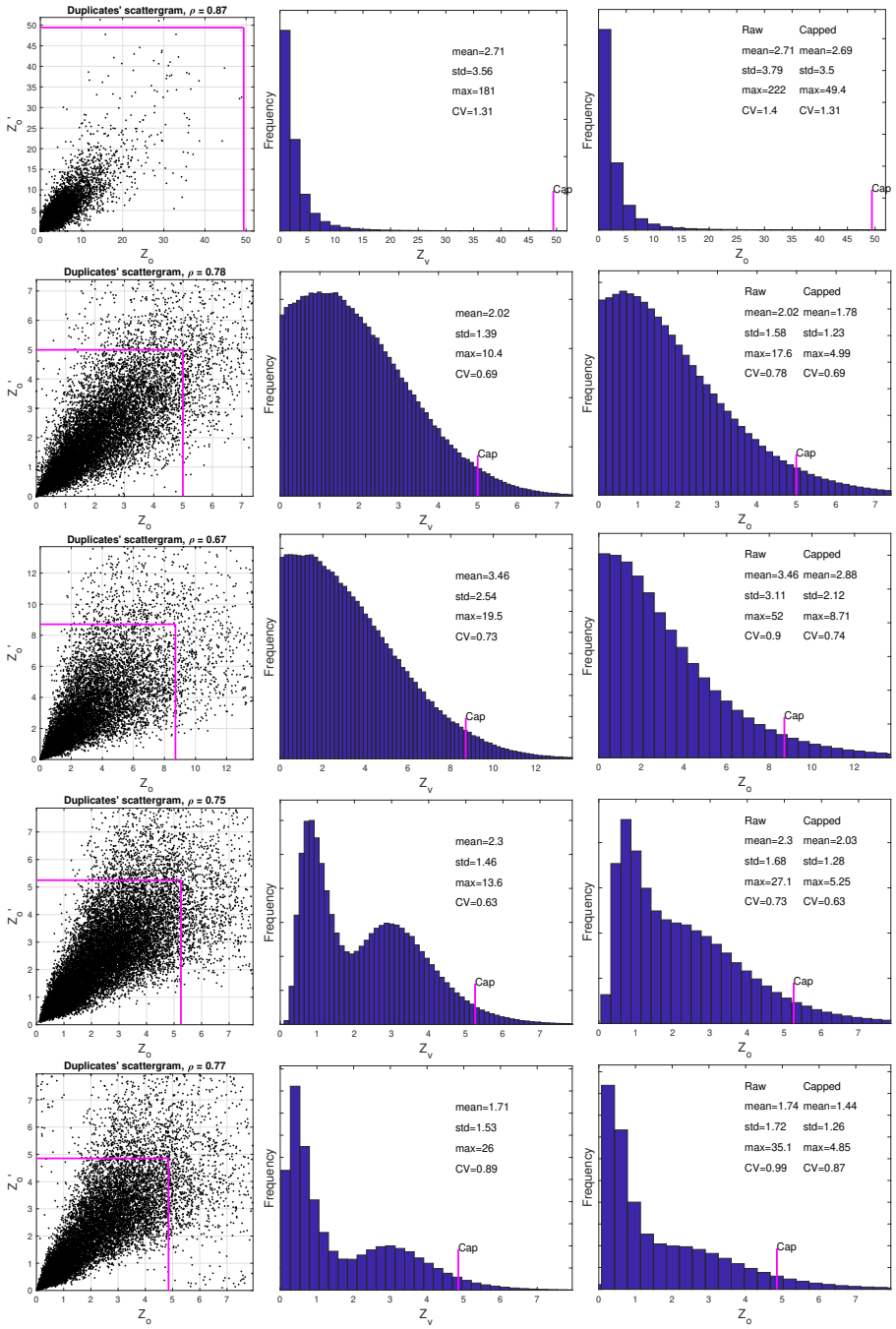


Figure 4.4 Duplicates' scattergrams (left column), error-free histograms (middle) and observed values histograms (right column) for cases 1 (top) to 5 (bottom) described in Table 4.2

. .

4.4.6 Capping after or before compositing ?

Some practitioners favor capping before compositing, others advocate doing so after. The argument of the 'before' camp is that possible outliers are averaged in the compositing process and can thus pass undetected, especially when the assay length is related to the presence or absence of visible mineralization. On the other hand, the 'after' camp maintains that assays over longer supports are already diluted, so it seems reasonable to dilute the outliers obtained on shorter supports to treat them like all other segments of the boreholes. But in fact, does it really matter ?

A simple experiment

A series of synthetic boreholes totalling 300 000 m are simulated at every 10 cm. The variogram model is spherical with range 20 m. The distribution is lognormal with $\alpha = 0$ and $\beta = 1$. Synthetic grades on variable lengths are computed according to the grade of the first 10 cm. The length is set to 0.5 m when $Z(x) > q_{90}$ (where q_{90} is the quantile 0.9 of the simulated distribution), 3 m when $Z(x) < q_{75}$ and 1.5 m otherwise. This scenario mimics preferential sampling of visible mineralization. The 'true' assays Z_v are obtained on the specified lengths by averaging the simulated values. Then a multiplicative error t (drawn from a lognormal distribution with $\sigma = 0.8$) is applied to each assay to form the set of observed assays Z_o.

In the 'before' case, a series of potential cap values are applied to observed assay grades. Then regular 3 m composites are formed using the capped assays. In the 'after' case, the 3 m composites are formed using the raw assays and then the cap values are applied to the composites. The target CV_vs for both cases are computed using the simulated Z_v at each 10 cm regularized over the variable length (in the 'before' case) or the composite length (in the 'after' case). Compositing is done on capped values for the 'before' case. The bias on mean in % (relative to simulated uncapped data) after capping the composites is computed. Results are shown in Fig. 4.5. The 'after' case presents slightly less bias on the mean for any given CV_{cap} than the 'before' case. Similar results (not shown) were obtained for different cases by varying the composite lengths, the variogram range, and parameters α, β and σ.

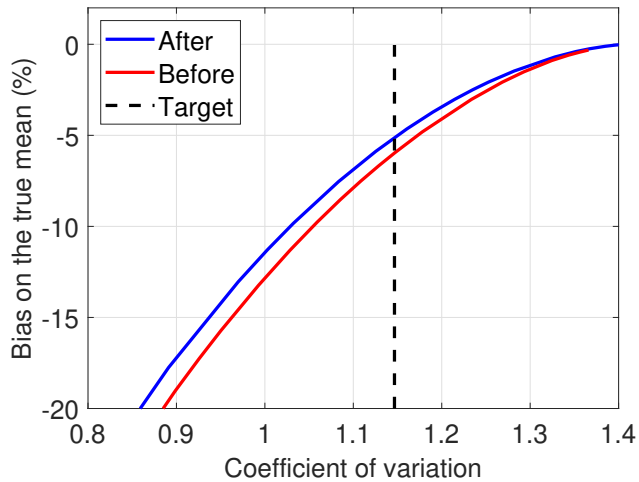


Figure 4.5 Effect of capping before or after compositing. Bias on the mean against CV for 3 m composites using assays on 0.5 m, 1.5 m and 3 m ; $\alpha = 0$, $\beta = 1$, $\sigma = 0.8$; Target CV obtained from error-free composites.

4.4.7 Workflow

To summarize, we propose to apply the following workflow to objectively determine the cap value based on duplicates :

- Compute correlation between coarse duplicates
- Compute the observed coefficient of variation (CV_o) using all data available
- Estimate CV_v, the target coefficient of variation of the error-free composites' grades, using Eq. 4.8
- Experimentally determine the cap value to apply to the composites to obtain $CV_{cap} = CV_v$

4.5 Case study

In the following section, two real duplicate data sets were used to compare the proposed capping strategy to some of the industry's most widely used methods.

4.5.1 Gold case

The first case study is on gold duplicates. Duplicate samples were obtained from 5 m blast holes (diameter 96 mm) using sampling of the cuttings in the cone around the drill string. Sample preparation consisted of pre-crushing 5 kg of cuttings to 2 mm and then pulverizing 250 g to 75 microns. Fire assaying was done on 30 g with an atomic absorption (FA-AA)

finish. The data set contains 1 786 samples for a total of 3 572 duplicate assays. The gold duplicates' scattergram (Figure 4.6) shows ρ=0.27, indicates an important overall sampling error. The artifact below $10^{-}2$ for both duplicate assays reflects the limits of detection used by the assay laboratory. It is also seen in the spike at low levels in the probability plot. This low correlation is undesirable but sadly common in practice, especially in blast hole samples. The log probability plot shows a gap in the upper tail around the 0.995 percentile value, which was retained as the log-probability plot capping level.

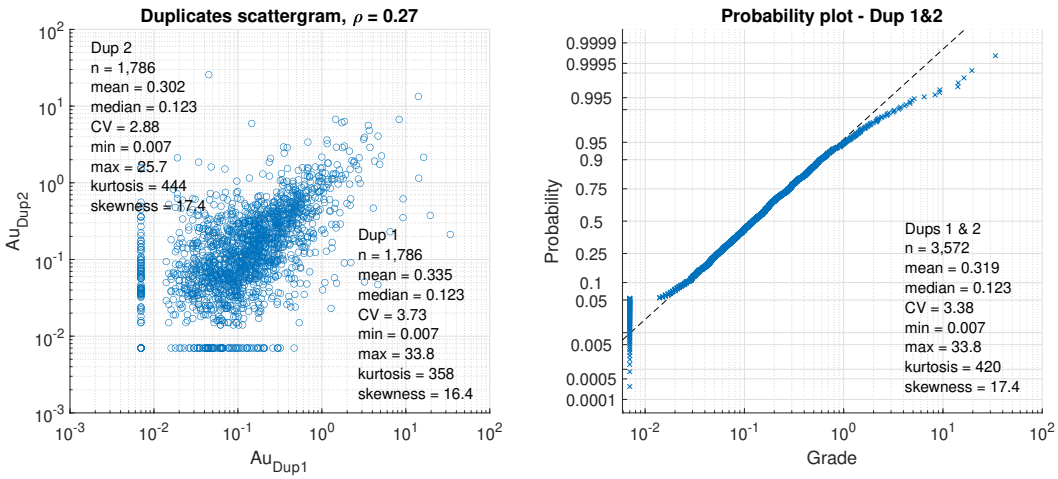


Figure 4.6 Gold duplicates' scattergram (left) and log probability plot for duplicates 1 and 2 (right)

Table 4.3 presents the cap results obtained using the different methods tested. The cap grades vary from 1.91 g/t to 4.49 g/t. The proposed method shows the second highest cap value at 3.69 g/t. The theoretical CV based on Eq. 4.8 is 1.76. By design, the proposed method returns the same CV. The 99th percentile method returns the next closest one. The log probability plot and Parrish methods suggest cap values that result in a strong overestimation and underestimation, respectively, of the target CV value.

Tableau 4.3 Gold case cap results - The target CV (Eq. 4.8) is 1.76

Capping method	Percentile	level	Capped CV level	Capped mean	Capped STD
Uncap data	n/a	n/a	3.38	0.32	1.08
99th percentile	0.990	3.11	1.67	0.27	0.45
Parish	0.981	1.91	1.43	0.25	0.36
Log probability plot	0.995	4.49	1.92	0.28	0.53
Duplicates	0.992	3.69	1.76	0.28	0.49

4.5.2 Nickel case

Drilling was done using the air core (AC) technique and 6 kg duplicates were created from the chips using a riffle splitter. Then 500 g of material were crushed to 2 mm and 250 g were pulverized to 75 microns. Fire assaying was done on 20 g with an XRF finish. The data set contains 698 samples. Each sample provided two coarse duplicates for a total of 1,396 duplicate assays.

The nickel duplicates present similar basic statistics (Fig. 4.7), with a much lower kurtosis value than for the gold data set. The high duplicates correlation coefficient (ρ=0.94) indicates good reproducibility of the assays as expected for a base metal deposit.

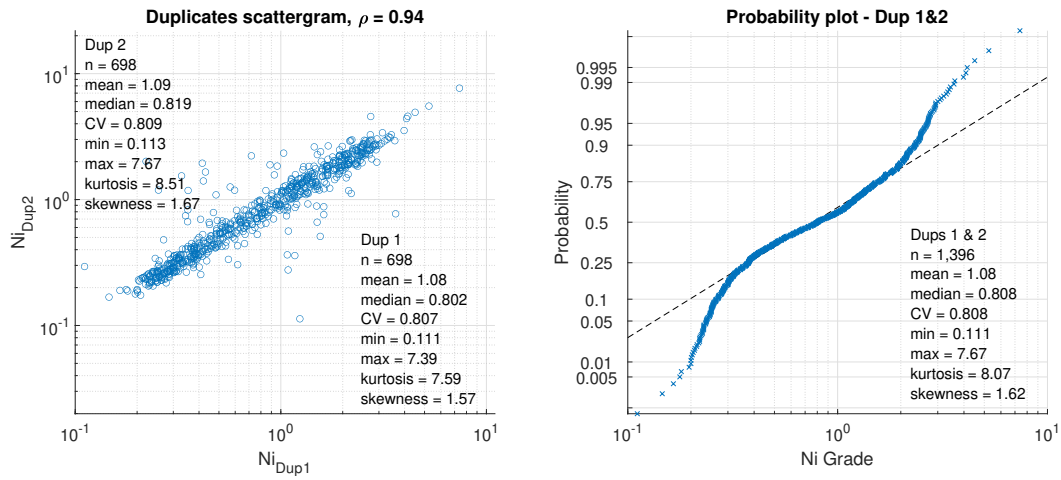


Figure 4.7 Nickel duplicates' scattergram (left) and log probability plot for duplicates 1 and 2 (right)

Table 4.4 shows the cap results obtained using the different methods. The proposed method presents the highest recommended cap value as a result of strong duplicates correlation. All other methods underestimate the target CV (0.783), a consequence of selecting an overly low cap value.

Tableau 4.4 Nickel case cap results - The target CV (Eq. 4.8) is 0.783

Capping method	Percentile	Capped level	Capped CV level	Capped mean	Capped STD
Uncap data	n/a	n/a	0.81	1.08	0.87
99th percentile	0.990	3.56	0.764	1.07	0.82
Parish	0.90	2.32	0.702	1.02	0.72
Log probability plot	0.991	3.60	0.764	1.07	0.82
Duplicates	0.996	4.59	0.783	1.08	0.84

4.6 Discussion

In mining applications, extreme values can often be encountered and must be controlled to avoid spreading high grade to large areas. The capping strategy can significantly impact the results of geostatistical and economic studies. In the exploration phase, an unduly low cap value can lead to an overly conservative resource estimate, possibly resulting in project rejection. On the other hand, too high of a cap value may lead to an overoptimistic economic valuation and important losses for mining companies. Various methods have been proposed in the past and are routinely used in the industry to determine cap values. Most are based on rather arbitrary empirical rules or on a subjective graphical interpretation of the cumulative distribution function. None use the important and mandatory coarse duplicate data to help determine a reasonable cap value that takes into account the quality of the assay procedure.

The proposed approach fills this gap. An unbiased multiplicative lognormal error model was used to derive the relationship between the observed and true grade CVs. The two CVs are simply linked through the duplicates correlation coefficient (see Eq. 4.8). One can predict the unobserved true CV from the observed CV with an estimate of the duplicates correlation. The proposed capping strategy is simple : select the cap value such that the CV of the capped samples is equal to the predicted true CV value.

Simulated data sets showed that Eq. 4.8 is unbiased for the true CV, even with a relatively small amount of duplicate data (Fig. 4.2). They also showed that the analyzed length has a limited impact on the proposed method, as both ρ and CV_o are relatively robust in response to this factor. Finally, Fig. 4.4 illustrates the good stability of the method relative to different departures from the lognormal assumption.

The question of capping before or after compositing was also examined. Simulated data sets with a preferential re-sampling approach (i.e., the intersections of high grade are analyzed on shorter lengths than low grade) were used. The differences between the 'before' and 'after'

cases are rather small but systematic. The 'before' case leads to more bias on the mean than the 'after' case for all CVs. We therefore recommend capping after compositing to minimize the bias on the mean.

The proposed approach was compared to some of the methods most widely used in the industry using two real data sets, one for gold and the other for nickel, where coarse duplicates were available. With the proposed approach, the CV computed using capped data was equal to the predicted true CV by design. Interestingly, the proposed approach had the highest cap value of all the methods tested when reproducible duplicate data were available (nickel case, $\rho = 0.94$) but not for duplicate data with low correlation (gold case, $\rho = 0.27$). This illustrates that the capping strategy of the proposed method, contrary to other methods, takes into account the quality and reliability of assay data measured by duplicates correlation. Other advantages of the proposed approach are its simplicity and objectivity, as it can easily be computed automatically from assays and duplicate data sets without requiring variogram modelling, kriging, simulation, or cdf plotting and gap/break interpretation like some of the other methods.

When following QA/QC recommendations, thousands of duplicates are typically obtained even before the pre-feasibility study. Hence, enough data are available to reliably estimate the correlation coefficient between duplicates. In earlier stages, when only a smaller quantity of duplicates is available, it might be interesting to consider a robust estimator of duplicates correlation. Many such robust estimators are described in the statistical literature. Further study is required to assess the influence of using different correlation robust estimators on the proposed method. Robust estimators of CV_o could also be considered but this appears less necessary as much more data are available for CV_o estimation.

4.7 Conclusions

A new capping strategy based on correlation coefficient between coarse duplicates is proposed. It incorporates the idea that the higher the correlation coefficient between duplicates, the greater the confidence level of the assay value, and accordingly the higher the cap value should be. The method proved to be robust to variations in sample support (assay length), departure from lognormality and capping before or after compositing. When applied to gold and nickel deposit data, the proposed approach provided cap grades that reflected the reliability of the assays and were generally higher than those obtained with the other methods tested. The proposed approach is simple, objective and repeatable. It makes it possible to automatically determine the cap grade from duplicate data.

CHAPITRE 5 ARTICLE 3 : A COMPARISON OF INDIRECT LOGNORMAL AND DISCRETE GAUSSIAN CHANGE OF SUPPORTS METHODS FOR VARIOUS VARIOGRAM ESTIMATORS

Historique de l'article : soumis le 11 décembre 2017, accepté le 7 décembre 2018, publié dans *The Journal of the Southern African Institute of Mining and Metallurgy*

Auteurs : Raphael Dutaut, Denis Marcotte

DOI : 10.17159/2411-9717/2019/v119n1a1

5.1 Abstract

In mining resources evaluations, geostatistical methods known as global change of support, allow prediction of the theoretical histogram and grade-tonnage curves prior to interpolations or simulations of grades. Two methods commonly used by professionals to guide the choice of interpolation parameters and assess results are the Discrete Gaussian Model (DGM) and the Indirect Lognormal correction (IndLog). These models rely upon an estimate of the dispersion variance of the blocks, which is derived by numerical integration of the variogram model over a discretized block. Due to difficulties in obtaining well-formed traditional experimental variograms (especially in the presence of outliers and limited clustered data), many professionals prefer to use "normalized" variograms such as correlogram (non-ergodic variogram), pairwise-relative or variogram of the normal score transform. A series of simulations with different grade distributions and variogram models are used to assess the performances and robustness of the various variogram estimators with respect to the DGM and IndLog global change of support. Our results show that the traditional variogram, the correlogram and normal score variogram have better performances, compared to pairwise, for both DGM and IndLog. Moreover, DGM provided better results than IndLog for the grade distributions that are not strictly lognormal. These findings provide valuable guides for the practitioners.

5.2 Keywords :

Geostatistics, Variogram type, Global change of support, Volume-Variance, Mining.

5.3 Introduction

When reporting resource estimates, practitioners are required to follow general standards described in the NI43-101 (CIM-guidelines), JORC or SAMREC codes. One can generally recognize three main parts directly related to the block model :

— Pre-processing : Exploratory Data Analysis (EDA), domaining, capping, compositing, declustering, variography.

— Processing : Block size, neighbourhood, interpolation types (linear or non-linear).

— Post processing : Classification, reporting.

Pre-processing aims to simplify and strengthen the processing step. Ideally, domaining, capping and compositing seeks to define a single homogenous population. During the processing phase, a common practice is to use a block size corresponding to the planned Selective Mining Unit (SMU). In most precious deposits (gold notably), the SMU size is frequently smaller than the recommended half data spacing [133], resulting in high estimation variance and either a high-degree of smoothing if the regression slope is managed or a high-degree of conditional bias if it is not. Most actual resource estimates are made using block sizes between a quarter to a sixth of the average data spacing, sometimes even smaller. In addition to the block size, the interpolation choices for the neighborhood selection and interpolator (e.g. inverse distance (ID) or ordinary kriging) are often based on the "Qualified Person's" (the QP) experience and some basic validation plots that often do not consider the conditional bias (ex : trends or SWATHs plots, which are designed to compare two sets of population using a one dimension graph).

Based on these findings, Rossi and Parker [134] proposed to "tune" the kriging plan so that the distribution of the interpolated estimate has a coefficient of variation close to the theoretical one obtained by a global change of support method (hereafter simply referred as the Change of Support or COS). Although criticized for the possible introduction of conditional bias ([135], [136]), this practice is becoming more common in technical reports. It is mostly applied as a visual check on the grade-tonnage curve or as a Q-Q plot of the grades. This simple technique ensures that the estimated blocks respect the SMU's theoretical global grade tonnage curves. In its simplest form, practitioners select the interpolant type (OK or ID2-ID3 most of the time), the minimum and maximum number of points (composites) used for the interpolation, and in some cases, the number of points per octant or maximum number of points per drill-hole. Isaaks [137], Rossi and Deutsch [5] and Nowak [138] stated that during the prefeasibility and feasibility studies, resource estimates should seek to respect the global recoverable resources (inferred from COS) rather than providing precise local estimates. Doing so, the estimates are necessarily conditionally biased [20]. Reliable

local estimates (free of conditional bias) can be later obtained during production from more abundant grade control data. Journel and Kyriakidis support the same idea but through the use of simulations [136].

In order to use the Rossi and Parker [134] "tuning method", the practitioner must first select an experimental variogram model and then choose a volume-variance correction algorithm (i.e. a COS model). Even though variogram models and main COS methods are well-known and widely used in the industry for decades, there are only a few publications that have experimentally compared the most widely used methods. Only some authors have compared the available methods from an experimental or practical point of view. Demange et al. [24] assessed four COS models on skewed datasets but didn't use the IndLog model. Srivastava and Parker [76] compared the robustness of different variogram type's to a lognormal heteroscedastic distribution and concluded that pairwise relative variograms are easier to interpret than traditional variograms, however they did not measure the impact of the resulting variogram model on estimation or COS. Rossi and Parker [134] tested global and local COS models, and concluded to the superiority of DGM, under the hypothesis that real ("perfect") variance correction factor was known. Curriero et al. [139] compared the impact of two variogram types on kriging results ; they found that the traditional variogram outperformed the correlogram. Emery [23] demonstrated some of the limitations of the IndLog, but without direct comparison to the DGM or the impact of variogram estimator. Emery and Torres [140] demonstrated that DGM outperform other COS model for lognormal distribution, but the IndLog model was not assessed. More recently, Chiles [26] compared two different DGMs and their validity range but the variogram types were not considered.

We aim to assess the impact of the grade distributions and the choices of the variogram estimator type on the performances of both DGM and IndLog COS models (the latter in two different versions). Three different grade distributions (lognormal, unimodal negatively skewed, and bimodal) and four different variogram estimators (traditional, pairwise relative, correlogram and normal scores) are compared, using different sampling densities. The goals pursued are to determine for the different distributions considered :

— which variogram estimator leads to estimated models providing more accurate block distribution predictions,
— which COS provides the best predictions when based on the estimated variogram models from the various variogram estimators,
— how does sampling density impact the performances of variogram estimators and COS models.

We also examine the effect of SMU size on the performances of COS models. We stress that, contrary to most previous papers on the subject, we do not assume the true variogram is

known but rather we fit the models to the different experimental variogram estimators. The same fitting criterion and algorithm is used in all cases studied. Hence, the impact of the choice of the fitting method is not considered and certainly constitutes a limitation to the study. Also, for simplicity, only 2D cases are considered.

The databases used for comparison are obtained by simulation using the FFT-MA method [141]. The methodology section presents briefly the FFT-MA method, the variogram estimators and the change of support models used. Then for various sampling spacing of simulated datasets, results of the change of support for different variograms types are examined. Tests on the COS model's sensitivity to block size and variogram range ratios are also studied. Discussions and conclusions follow.

5.4 Methods

We use a series of simulated data sets to assess the performance of different variogram estimators for DGM and IndLog. These two COS models are the most widely used in the mining industry. Simulated datasets enable us to compute the experimental block distributions of the different realizations and allows the comparison between the "true" results to the estimated results predicted by COS with limited sample data. Four main estimators of variograms are generally found within technical reports : traditional, correlogram (non-Ergodic), pairwise relative and normal score. Each of these variograms were computed on both exhaustive and partial datasets, with the latter used to mimic different densities of exploration drill holes.

5.4.1 Simulation and sampling method

Simulations to generate the exhaustive datasets were performed by Fast Fourier Transformation with Moving Average (FFT-MA) [141]. The mathematical concept behind FFT-MA simulations show that a random function can be written as a weighted average of white noises where the weighting function used guarantees the reproduction of the covariance model. In FFT-MA, most of the calculations are done in spectral domain [142], which allows fast generation of a large number of Gaussian simulations. To simulate different distributions, post-simulation transformations were applied : an exponential transformation for lognormal distribution and Gaussian anamorphosis, based on ranking, for bi-modal and negatively skewed distributions. Lognormal, bi-modal and negatively skewed distributions cover typical distribution types found in mineral deposits (especially in precious commodities). FFT-MA simulations were performed in 2D using a unit node spacing and a large square field of one thousand units aside. The 2D results could be interpreted as a large bench in an open pit

operation, or as a tabular vein-type gold deposit.

To simulate production or definition drill holes patterns, realizations were virtually sampled over regular grids at 10 and 50 units spacing. The regular grid removes the need to decluster data, although we stress that declustering is an important step that should be performed prior to any mining estimation. Most widely used declustering methods are cell-declustering [143], polygonal or voronoi [4] and kriging weights [144]. Moreover, no measurement error was added to the selected data. For interested readers, Journel and Kyriakidis [136] describes the impacts of sampling error on mining geostatistical methods.

5.4.2 The variogram estimators

There are 4 variogram estimators usually found in technical reports (the term variogram is used here for experimental semi-variogram, and more generally for any tool describing the correlation of the data as a function of the distance) :

— The traditional variogram [73] : defined as half the average squared difference between points separated by a distance h.

— The correlogram or non-ergodic variogram [14] is a standardized variogram type. The standardization is performed using the averages of both the head and the tail and the standard deviations of the pairs. It is considered to be more robust to outliers and heteroscedastic datasets [76]. It is one of the most commonly used variogram estimators in mining application.

— David [20] described two forms of relative variograms : the general relative and the pairwise relative. Both are standardized versions of the traditional variogram. In this paper, the pairwise version was used. In the pairwise relative variogram, the standardization is done with the average of the two values for each pair.

— The normal score variogram ([4], [74]) is identical to traditional variograms except that the data is first transformed to a Gaussian distribution. The Gaussian transformation reduces the impact of outliers on the variogram and eases the modelling. The variogram model must then be back transformed to the original space.

In addition to the nugget effect, it is a common practice in the mining industry to fit one or two variogram structures. Some practitioners may use more, but most resource geologists use two structures, one for short range and the other one for long range, but both oriented in the same directions. The spherical model is a common model used in public reports. Hence, we choose to fit a sum of two spherical variogram models to the experimental variogram of the raw Z-variable, *i.e.* obtained after transformation of simulated Gaussian values to grade. The fitting was done automatically using a weighted least squares method, similar

to Cressie [78]. Each spherical component was considered anisotropic with known directions of anisotropy. The automatic fitting procedure was applied with no manual control to limit the bias and subjectivity. Although some automatically fitted models ended up far from the true Z-exhaustive variograms, they were kept in all computations to avoid favoring any of the estimators.

5.4.3 Change of support models

Methods of change of support (COS), also known as volume-variance correction or global estimates, seek to derive the global histogram and grade-tonnage curves at SMU size from point distribution (composites). Regrettably, COS are not systematically included in resources estimates in technical reports, although they are required to obtain SMU's grade-tonnage curves necessary to guide block simulation/interpolation. There are numerous COS approaches including : affine correction [133], Gaussian approach [145], bi-Gaussian approach [146], mosaic correction [24], indirect Lognormal ([14], [23]), Lognormal with three parameters [147], Discrete Gaussian ([148], [94], [22]) and LU simulation [114]. However, the Indirect Lognormal correction (IndLog) and the Discrete Gaussian Model (DGM) methods are commonly used by practitioners of Canadian resource QPs.

The Indirect Lognormal method is suited for positively skewed distributions (not necessarily lognormal) observed for most precious metal deposits. It assumes the permanence of the lognormal distribution between data at the point support (Z) and SMU's histograms (Z_{SMU}). The distribution at the SMU's size is inferred from the point data by correcting the variance of the population. We have in this model :

$$Z_{SMU} = aZ^b \tag{5.1}$$

where the equality stands in distribution. Two different approaches can be used to determine the coefficients a and b. The first approach is described in Isaaks and Srivastava [14], hereafter referred to as "traditional IndLog". One determines first b :

$$b = \sqrt{\frac{\ln\left(fCV^2 + 1\right)}{\ln\left(CV^2 + 1\right)}} \tag{5.2}$$

with f the variance correction factor given by :

$$f = 1 - \frac{D^2(.|SMU)}{D^2(.|DOMAIN)} = \frac{D^2(SMU|DOMAIN)}{D^2(.|DOMAIN)} \tag{5.3}$$

and where D^2 is the dispersion variance and CV is the coefficient of variation of the point data (Z) ; f can be inferred from the variogram model as the ratio of the SMU block variance to the quasi-point variance within the domain. As it is a ratio, it is insensitive to the true sill value (*i.e.* normalised variograms can be used without rescaling to population variance).

Parameter a is then chosen to ensure equality of the mean for SMU and point distributions. However, as remarked by Isaaks and Srivastava [14], multiplication by $a \neq 1$ does modify the variance of the SMU, thus the correction is not completely consistent.

A better and entirely consistent approach was recently proposed by Emery [23], referred in this article as "Emery IndLog". First determine b :

$$\frac{E[Z_i^{2b}]}{E[Z_i^b]^2} - \left(1 + \frac{\sigma_v^2}{m^2}\right) = 0 \tag{5.4}$$

where $m = E[Z]$. Then, solve the following equation for a :

$$a = \frac{m}{E[Z^b]} \tag{5.5}$$

It is easy to verify that these choices ensure the point's mean is preserved and the right SMU variance is recovered. Despite the theoretical superiority of the Emery [23] approach, most softwares use the Isaaks and Srivastava [14] approach, in our research we assess both methods.

The DGM is slightly more evolved mathematically and interested readers will find more details in Armstrong and Matheron [149], Rivoirard [94], Emery [22], Chiles and Delfiner [4] or Rossi and Deutsch [134]. The method makes no assumption on the point marginal distribution (Gaussian, lognormal, multimodal or any other distribution type). Although it is usually presented in textbooks in link with Hermite polynomial expansions to express the transformation from Gaussian to grade variable, this is not a requirement for the method and it can be equally applied using only a graphical transformation. The DGM starts from the Cartier relation :

$$E[Z(x)|Z_v] = Z_v \tag{5.6}$$

where $Z(x)$ is the grade of a point randomly selected within block v. Then, assuming that the Gaussian variables Y and Y_v associated to the grades Z and Z_v follow jointly a zero mean unit variance bi-Gaussian distribution with correlation r, one can compute (Rivoirard [94]) :

$$Z_v = \phi_v(Y_v) = E[Z(x)|Z_v] = \int \phi(rY_v + (1-r^2)^{\frac{1}{2}}u)g(u)du \tag{5.7}$$

where u is an integration variable, $g(u)$ is the Gaussian density function evaluated at u and $\phi(.)$ is the transformation function from Gaussian Y to grade Z at the point scale. Eq. 5.7 enables to associate a block grade value Z_v to every value of a $N(0,1)$ distribution (Y_v) using simple numerical integration. Note that $\phi(.)$ was defined by a simple graphical relation in the present document. In all cases, the only unknown in Eq. 5.7 is the correlation coefficient r. The coefficient r is selected such as to ensure that the transformed Z_v variables have the desired SMU variance.

5.5 Results

5.5.1 Change of support – Sensitivity to variogram type.

A total of 100 Gaussian simulation realizations are obtained on a grid of 1,000 by 1,000 points. The large size of the field (>15 times the correlation range) allows to obtain more stable variograms ; the purpose of this study is not to assess variograms stability as in Srivastava and Parker [76], but rather the impact of variogram estimators on COS performances. The same variogram is used for all realizations of the Gaussian variable. The variogram is anisotropic with parameters shown in Table 5.1.

Tableau 5.1 Variogram model (Gaussian)

Structure	Variance	Range (x-y) (in grid units)	Orientation
Nugget effect	0.1	-	-
1	0.9	60-30	45

Using Gaussian anamorphosis, three different target distributions for point grade are used : lognormal, bimodal (considered as a single homogeneous bimodal distribution) and negatively skewed. Target distribution statistics are presented in Table 5.2.

Tableau 5.2 Target population statistics

Distribution	Mean	Variance	Skewness
Lognormal	1.65	4.64	5.84
Bimodal	1.32	1.44	0.50
Negatively skewed	3.00	0.94	-4.17

Using all simulated points, the four variogram estimators (Traditional, Correlogram, Pairwise, Nscore) were automatically fitted by assuming a spatial structure composed of a nugget effect

and two anisotropic spherical models, both oriented in the same direction. We stress that the fitted model has two structures instead of a single one in the theoretical model as indicated in Table 5.1. This adds flexibility in modelling to account for the effect of the Gaussian anamorphosis on the variogram model. Although a spherical model is used to simulate the Gaussian variable, the variogram of the transformed variable is no more spherical. For each of the four variogram types, DGM and both versions of IndLog change of support were tested using a SMU of the size of 10x10 units. In addition to exhaustive datasets, each realization was subsampled to imitate regular drill hole patterns at respectively 1, 10 and 50 units. Then, for each pattern, experimental variograms and COS models were recalculated. This resulted in 3 patterns x 4 estimators x 3 distributions x 3 COS = 108 possibilities for each cut-off.

5.5.2 Exhaustive data sets

COS models rely upon an estimate of the dispersion variance to derive distribution at different support size. Using the exhaustive database, Table 5.3 compares the mean dispersion variances calculated experimentally from the simulated nodes (real), to the dispersion variances estimated from the different variogram types 10x10 units SMU size. Generally, all variogram types provide close estimates of the true block dispersion variance ; although the Pairwise variogram underestimates systematically the dispersion variance.

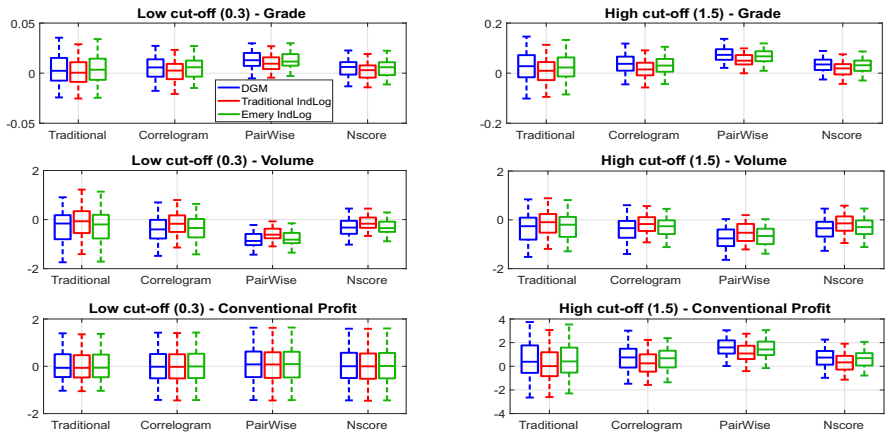
Tableau 5.3 Mean dispersion variance for SMU size of 10x10 units

Distribution	Simulated/real	Traditional	Correlogram	Pairwise	Nscore
Lognormal	2.79	2.85	2.83	2.69	2.84
Bimodal	0.93	0.91	0.91	0.77	0.91
Negatively	0.66	0.62	0.62	0.54	0.64

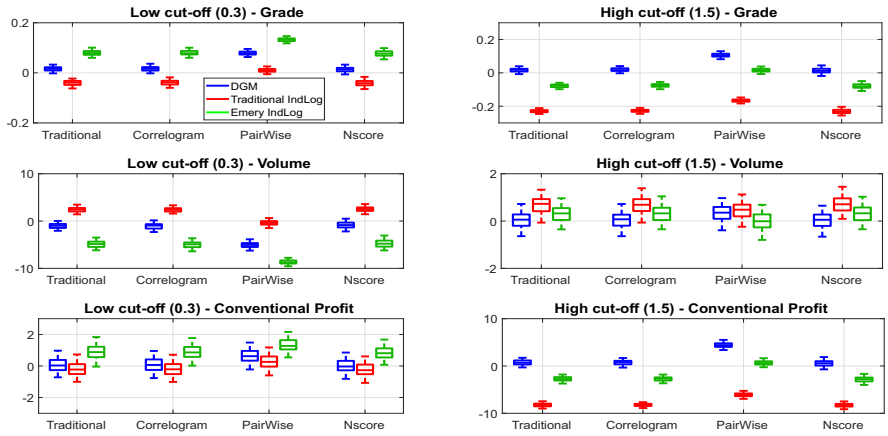
The difference between real (simulated) grades and COS results, were then calculated for cut-off grades varying from 0 to 3.3 with a 0.3 increment. Realistic mining cut-offs of 0.3 and 1.5 were selected for plotting (e.g. 0.3 and 1.5 g/t cut-off would be reasonable for a gold open pit and a large underground mine respectively). Box plots in Figure 5.1 present the Grade, Volume and conventional profit errors (the conventional profit is defined as $P(c) = V(c)(m(c) - c)$ where c is the considered cut-off and $P(c)$, $V(c)$ and $m(c)$ are respectively the conventional profit, the volume and the ore grade at that cut-off ; it corresponds to the amount of metal left after paying for the costs represented by the cut-off for the 3 distributions with exhaustive sampling). DGM (Blue), traditional IndLog (Red) and Emery IndLog (Green) are shown. The errors are calculated by subtracting the estimated value to the true value for

each realization so that a positive error corresponds to an underestimation of the parameter. On each boxplot, the central mark indicates the median, the box limits correspond to the 25th and 75th percentile, and the whisker extends to two standard deviations ($\pm 2\sigma$).

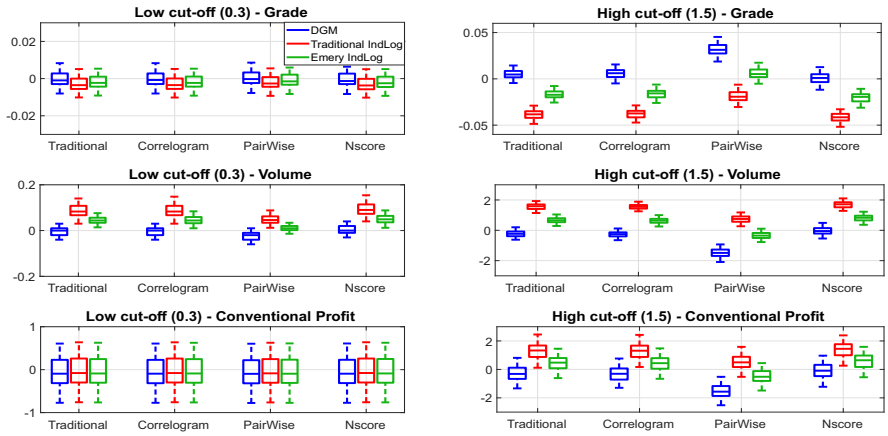
For the Lognormal distribution, all COS models perform relatively well, while the traditional IndLog shows errors more centred around 0 and with a slightly smaller spread. Minimum errors (mean and spread) are obtained with Nscore variograms whereas worst results are associated to the pairwise variograms. For distributions that are not strictly lognormal, the DGM appear unbiased, when conversely, both IndLog COS appear biased for most of the statistics considered. Emery IndLog achieves better results than the traditional lognormal correction. Variogram types performed equivalently, except for the Pairwise which return slightly biased results.



(a) Lognormal distribution



(b) Bimodal distribution

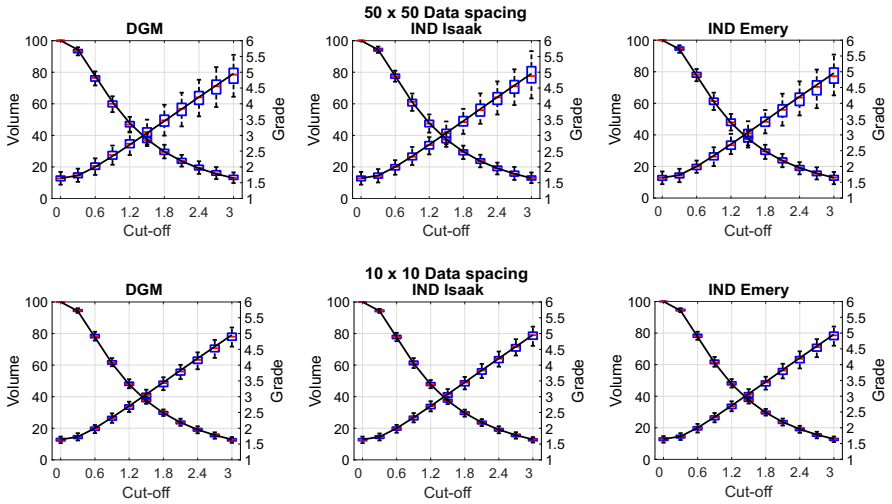


(c) Negatively Skewed distribution

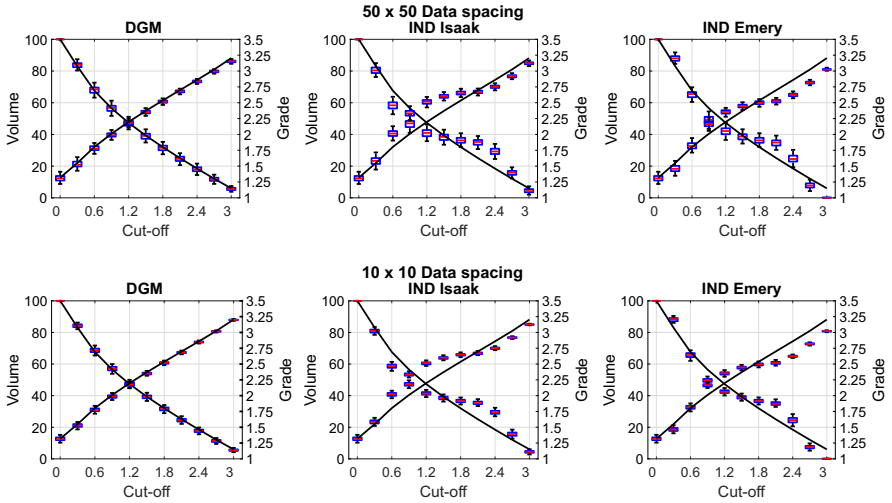
Figure 5.1 COS results at a 10x10 SMU size for simulated lognormal, Bimodal and Negatively skewed distribution using exhaustive dataset : error boxplot for DGM (Blue), Traditional IndLog (Red), Emery's IndLog (Green) for each variogram estimator ; low cut-off (left column), high cut-off (right column) ; error on grade in grade units (first row), error on volume in % (second row), error on Conventional Profit in % (third row).

5.5.3 Grade tonnage curves for the different distributions

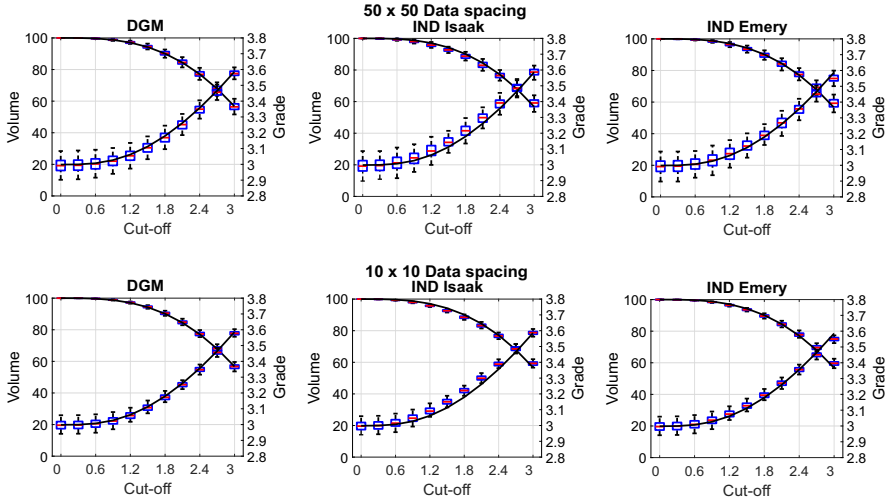
Figure 5.2 presents, as a boxplot, the results of simulated COS model for various cut-offs for each distribution at 10 and 50 node spacing using traditional variogram. These spacings were selected because they could correspond to the drilling spacing for grade control and long-term models in gold projects. As expected, there is a decrease in confidence (larger box plot spread) with the increase of data spacing, for both grade and volume. With increasing cut-offs, confidence decreases for grade, while confidence in tonnage is more constant. The error distribution is relatively symmetrical for all COS models. As previously noted, the DGM present results more robust to the different distribution. For Bimodal and negatively skewed distributions, Emery IngLog errors are more centred on the "real" grate tonnage curves (no systematic bias) than the traditional IndLog model. This systematic bias of the traditional IndLog is particularly critical in mining application.



(a) Lognormal distribution



(b) Bimodal distribution



(c) Negatively Skewed distribution

Figure 5.2 True Grade-Tonnage curves (black line) compare to COS results in the case of lognormal (a), Bimodal (b) and Negatively skewed (c) distribution sampled on regular grids of size 50 (first row) and 10 (second row) ; traditional variogram estimator used for fitting model.

To define an overall measure of performance, the Mean Relative Unsigned Errors were calculated over all realizations and all cut-off grades as follows :

$$MRUE = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{nc}\sum_{i=1}^{nc}\left(\frac{100 * abs(Estimated_{i,j} - Real_{i,j})}{Real_{i,j}}\right) \tag{5.8}$$

where n is the number of simulations and nc the number of different cut-offs considered.

Table 5.4 presents the Mean Relative Unsigned Errors (MRUE) calculated using Eq. 5.8 on the 50 node spacing datasets and using traditional variogram estimator. DGM's outperforms the other models for Bimodal distribution and shows a slight advantage for Negatively skewed distribution. For the Lognormal distribution, all models performed similarly as expected.

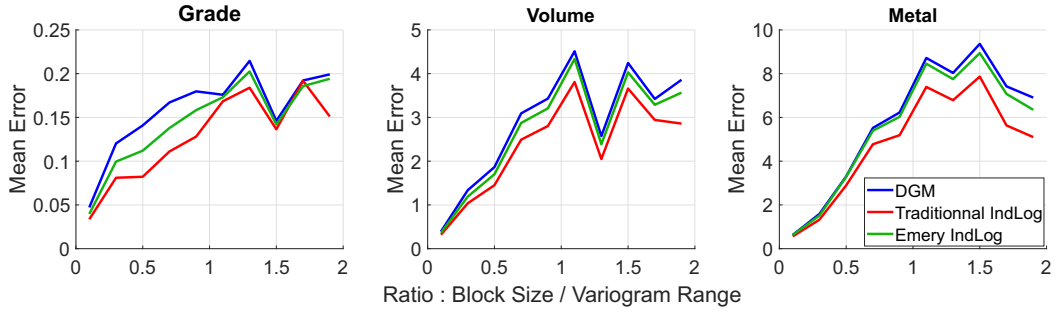
Tableau 5.4 Mean Relative Unsigned Errors – 10x10 SMU size - 50 nodes data spacing ; Errors on : Volumes / Grades / Conventional Profit

Distribution	DGM	IND Isaaks	IND Emery
Lognormal	5.0 / 5.2 / 11.6	5.2 / 4.8 / 10.9	5.3 / 4.9 / 11.1
Bimodal	5.1 / 1.5 / 9.1	20.4 / 6.8 / 19.8	23.3 / 10.9 / 23.7
Negatively	1.0 / 0.8 / 2.4	1.6 / 1.1 / 2.6	1.4 / 0.9 / 2.4

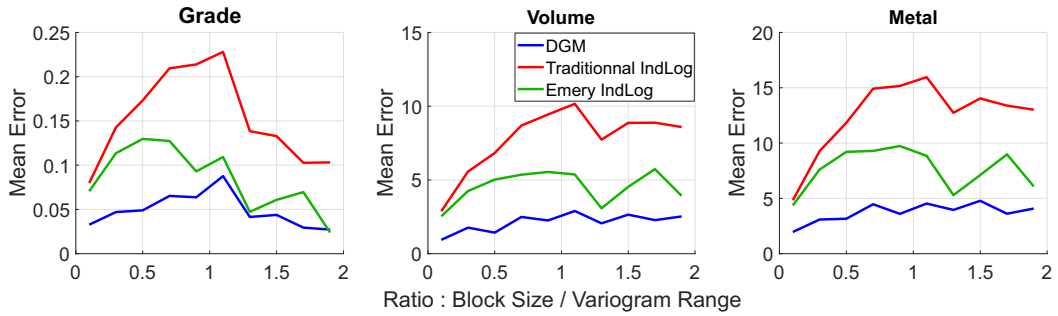
5.6 Change of support – Sensitivity to SMU size

For resource estimates, practitioners must consider the SMU size. This parameter is usually defined as a function of the geological setting, selected mining equipment (Shovel-Truck size) and other practical resource constraints such as reasonable minimum mining width or software limitations (sub-blocks or partial/percent blocks). The sensitivity of COS models to block size and variogram range was evaluated for several (block size/variogram range) ratios. Datasets were simulated using FFT-MA with a fixed variogram model (isotropic spherical model with a range of 100 units). The exhaustive traditional variograms were computed and automatically fitted. COS models were then compared to simulated SMUs. The exercise was repeated for block sizes ranging from 10 to 200 units, corresponding to ratios 0.1 to 2. Figure 5.3 presents the grade and volume mean absolute error (at all cut-offs) for each COS model as a function of the ratio. For Lognormal distribution, COS models present relatively similar results, with an error that comprehensively increases with the block size. For Bimodal and Negatively skewed distribution, the DGM perform the best, followed by Emery IndLog models and then the traditional IndLog. The increase of error with size of support is expected as the true block distribution departs more and more from the point distribution. Also, differences

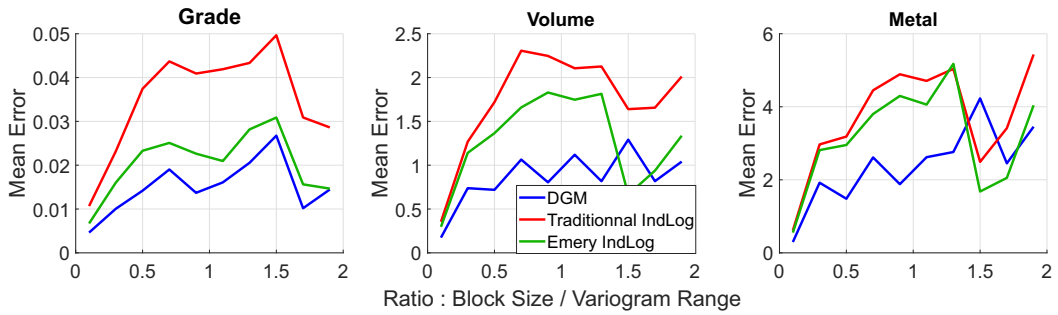
between COS models are stronger for the bimodal and negatively skewed distribution where DGM outperforms the other COS models for most ratios.



(a) Lognormal distribution



(b) Bimodal distribution



(c) Negatively Skewed distribution

Figure 5.3 Cumulative absolute error as a function of the ratio (block size – variogram range) for both COS models : DGM(bleu), traditional IndLog(red), Emery IndLog (green). Lognormal (a), Bimodal (b) in the middle and Negatively skewed (c) distribution on the last row.

5.7 Discussion

We compared experimentally the performances of two widely used COS models with respect to choice of variogram estimator, type of grade distribution, sampling density and SMU size.

Three different representative distributions were considered, and four different variogram estimators were automatically fit for a series of sampling patterns. For computational reasons, the tests were run in 2D rather than 3D and a single variogram model was used for the simulated Gaussian variable. Only regular sampling patterns were considered in order to avoid introducing the declustering method as an additional factor to consider.

Keeping in mind the limitations of this experimental study, our results displayed in Figure 5.1 indicate that the traditional variogram estimator, the correlogram proposed by Isaaks and Srivastava [14] and Normal score variogram provide the best, or close to best, COS results for all the tested distributions. Our results indicate no clear advantage for the correlogram compared to the traditional estimator, which are the two most widely used variograms estimators in technical reports. Pairwise variograms provide biased results when used for COS purposes, a consequence of the underestimation of block dispersion variance evidenced in Table 5.3 for this variogram estimator. For Lognormal distribution, the normal score estimator returns COS results with the smallest spreading. This is probably due to its robustness to outliers.

Figure 5.2 clearly indicate that the DGM returns more precise and less biased results than IndLog, particularly in its traditional version and when the grade distribution are not lognormal (Figures 5.2(b) and 5.2(c)). These findings concur with Emery and Torres [140] conclusions. As demonstrated in Figures 5.2(a) and 5.3(a), in the case of a strictly lognormal distribution, all COS models perform equivalently. DGM and Emery IndLog proved more robust than traditional IndLog to the (SMU size/ variogram range) ratio for other distributions (Figures 5.2 and 5.3(a) and 5.3(b)).

COS models seek to estimate the *in-situ* grade distribution at the SMU scale but the precision of the results depends of the data availability. When less data are available, both the histogram and the variogram estimations are less precise which in turn adversely affects the COS performances. However, DGM proved to be unbiased for all sampling patterns (Figure 5.2) and all SMU sizes (Figure 5.3). In our tests DGM provides grades-tonnages Mean Relative Unsigned Errors (MRUE) varying from 1 to 5% for all distribution types, while errors with Indirect lognormal methods can go up to 20% for bimodal distribution (Table 5.4). In the lognormal case for all COS models, MRUE can stretch up to more than 10% at a higher cut-off. These errors represent a best case scenario, without any domaining, sampling, analytical or declustering errors. In real case, practitioners will probably have higher errors during COS associated with an interpolation process. Therefore, it is recommended to use COS curves as guidelines rather than strict constraints for recoverable resource estimates.

During mining operations, the decision to classify a block as “ore” or “waste” is based on the future estimated grade at the time of mining, not the true unknown block grade. The more

information available, the better the classification. This (usually small) information effect can be incorporated within the COS (Roth and Deraisme [150]) by simply increasing the variance correction factor (or decreasing the degree of selectivity). Although DGM proved to be more efficient in our tests, it should not be applied blindly. The quality of results obtained also depends on important and difficult decisions about domaining (Emery and Ortiz [140], Romary et al. [151]) and declustering (Deutsch [143], Rossi and Deutsch [5]). Moreover, the quality of raw data sampling obviously has an impact (Francois-Bongarcon and Gy [71]). Any bias in the point histogram or variogram will adversely affect the performances of the DGM or other COS method (Journel and Kyriakidis [136], Pyrcz et al. [152], Carrasco [153]). Despite these reservations, DGM should be recommended preferably to IndLog because of its lack of bias and greater robustness to the sampling density, the type of distribution and the SMU size.

We did not consider the use of conditional simulations in this study. Despite its possible merits, this method is rarely used in technical reports. One reason is that it is much more computationally intensive than DGM or IndLog. Moreover, the variogram model in simulations is for the Gaussian transformed variable rather than the grade variable as in DGM and IndLog. This choice is therefore expected to have a bigger influence with simulations than with DGM and IndLog. It is also complicated by other sensitive decisions regarding the neighborhood to use for the conditioning.

5.8 Conclusions

Our study indicates a clear superiority of the Discrete Gaussian Model compared to the traditional Indirect Lognormal approach. The improvement proposed by Emery [23] increases the quality of the indirect lognormal model, but remains less robust than the DGM with respect to distribution types. DGM is unbiased, robust to sampling density, SMU size and distribution types. Our study also reveals that both traditional, Normal score variogram and correlogram estimators provide comparable estimates of the dispersion variance used for Change of Support correction. On the contrary, pairwise estimator should be used with caution as it is more sensitive to the grade distribution.

CHAPITRE 6 DISCUSSION GÉNÉRALE

Un nouvel algorithme rapide et quasi-optimum a été développé pour répondre aux besoins de quantification des budgets de sondages. L'algorithme est adapté aux projets de surface ou souterrain et permet d'optimiser la position des forages potentiels en considérant, de manière dynamique, la couverture des forages précédemment sélectionnés. De plus, la fonction de pondération (w) permet d'obtenir une variation linéaire de la couverture de chaque forage (existant ou planifié), ce qui n'est pas possible avec les modèles de programmation linéaire présentés dans la littérature. Nous avons proposé une fonction objectif à maximiser qui respecte les contraintes géométriques et économiques, ce qui permet de localiser les zones de risque maximum à forer. Normalisée, cette fonction objectif permet de comparer simultanément plusieurs gisements, ce qui peut être utilisé pour diminuer le risque de l'ensemble des ressources de la compagnie. Cette approche permet de répondre à la première problématique de *recommandation objective de forages* décrite dans la section 1.

Concernant les problèmes de seuil d'écrêtage, une méthode originale a été proposée en abordant cette problématique sous un jour nouveau. Dans la plupart des projets, un protocole QaQc d'analyses des duplicatas existe. Il sert à s'assurer de la précision et de la répétabilité des analyses ; établissant ainsi la confiance pour les étapes subséquentes du processus de ressources. La plupart des méthodes utilisées pour la sélection de seuil d'écrêtage dans les rapports NI43-101 sont basées sur une forme d'interprétation subjective de l'histogramme des données. En plus d'être subjectives, ces méthodes ne parviennent pas à capturer le risque associé à l'erreur d'échantillonnage. Afin de répondre à ces lacunes, notre méthode propose d'utiliser le coefficient de corrélation des duplicatas pour estimer le coefficient de variation des données sans erreurs d'échantillonnages. Cette information nous permet par la suite de définir un seuil d'écrêtage objectif. Cet article a aussi permis de montrer que le choix d'appliquer l'écrêtage avant ou après avoir calculé les composites, avait un certain impact et qu'il était préférable d'écrêter après avoir formé les composites puisque le calage du CV laissait plus de métal disponible dans ce cas. La gestion objective du risque associé aux hautes teneurs en tenant compte des résultats de QaQc, correspond au deuxième objectif spécifique de cette thèse et notre méthode y apporte une solution originale.

Lors d'un estimé de ressources, la méthode de validation géostatistique la plus commune consiste à comparer les courbes tonnages teneurs des blocs à celles obtenues grâce à une méthode de changement de support global produit à partir de l'histogramme et du variogramme des analyses. L'existence de plusieurs méthodes de changement de support, les plus

fréquemment utilisées étant les modèles discrets gaussien et *indirect log normal*, confronte les géologues au choix de la meilleure méthode pour son type de distribution de teneurs. De plus, il existe plusieurs outils d'analyses des corrélations spatiales (*variogrammes expérimentaux*) et leurs impacts sur les modèles de changements de support n'étaient pas décrits. Dans notre publication, une comparaison des différents impacts, en fonction de l'approche sélectionnée, permet d'améliorer la robustesse des courbes de changements de support, et donc de réduire le risque sur les interpolations de modèles de blocs. Notre étude démontre la supériorité du modèle gaussien discret et la faiblesse du variogramme relatif par paires utilisé pour l'estimation globale par COS. L'identification des *meilleures pratiques* concernant l'utilisation des modèles de COS permet de réduire le risque associé à l'utilisation de cet outil de validation et permet de répondre à la troisième problématique spécifique de cette thèse.

L'algorithme d'optimisation de campagne de forage proposé permet de maximiser pour un nombre de forages donné la fonction objectif. Une limitation de cette approche est l'aspect purement financier, en effet, notre algorithme considère implicitement que chaque mètre de forage a le même coût. Or, il est fréquent que le coût d'un forage varie en fonction de sa position : coût de mobilisation, distance aux points d'axes existants, défrichement nécessaire, gestion des eaux, etc... Une piste d'amélioration serait d'ajouter pour chaque localisation possible, un coût qui serait décroissant en fonction du nombre de sondage effectué depuis une même plate-forme. Il faudrait alors pondérer la fonction objectif en fonction du coût de chaque sondage ou plate-forme. La fonction de pondération (w) permet d'obtenir une pondération régulière de *l'effet de couverture* de chaque sondage. Dans notre article, une fonction linéaire a été utilisée, cependant, d'autres types de fonction de pondération pourraient être utilisées. Bien que nous pensons que cette fonction de pondération est une amélioration par rapport aux fonctions booléennes (0-1) utilisées par la programmation linéaire, sa relation avec la distance *dmin* définie par l'utilisateur reste à approfondir. La sélection des paramètres *nlist* et *ntrial*, qui permettent de maximiser la qualité des solutions, mais augmentent les temps de calculs, reste un enjeu significatif. Pour le moment, l'utilisateur doit rechercher manuellement les meilleurs paramètres pour chaque jeu de données, une automatisation de la sélection de ces paramètres rendrait l'algorithme plus facile à utiliser. L'algorithme permet d'optimiser plusieurs projets simultanément. Durant notre étude de cas, nous avons fait l'hypothèse qu'une normalisation du variogramme suffisait pour homogénéiser les fonctions objectifs et donc pouvoir optimiser simultanément les gisements. Une évaluation de cette hypothèse nécessite davantage d'investigation dans le futur. Une approche multivariée permettant de comparer les cas polymétalliques représente aussi un axe de recherche intéressant.

Dans notre étude, nous avons utilisé les duplicatas de rejets afin de calculer un seuil d'écrêtage intégrant l'erreur d'échantillonnage. Cependant, il existe plusieurs types de duplicatas

possible (terrain, sur rejet, sur pulpes). Une évaluation de l'impact des différents types de duplicata sur les recommandations de seuil d'écrêtage permettrait d'améliorer la confiance dans cette nouvelle méthode. Le coefficient de corrélation (ρ) des duplicatas est un élément central de la méthode proposée. Bien que cette grandeur soit parfois reconnue comme un estimateur « robuste », elle n'en reste pas moins sensible à la présence de valeurs extrêmes et peut présenter des biais quand peu de données sont disponibles. Il existe plusieurs autres estimateurs de ρ, l'utilisation de la corrélation de Spearman sur les rangs, au lieu de Pearson, pourrait présenter une amélioration ; l'utilisation de simulation Monte Carlo [154] quant à elle pourrait permettre d'associer une marge d'erreur à l'estimation du seuil d'écrêtage. Aussi, plutôt que de définir un seuil d'écrêtage unique, il est possible d'appliquer une fonction correctrice décroissante sur l'ensemble des teneurs, réduisant ainsi proportionnellement (ou suivant une fonction à définir) l'ensemble des valeurs afin d'obtenir le CV_v. Moins pénalisante sur les très hautes teneurs, cette approche permettrait de conserver les zones les plus riches des gisements, tout en réduisant significativement le risque sur le métal total.

Nous avons discuté de l'impact du choix du type de variogramme sur des modèles de changement de support globaux. Afin de calculer le variogramme expérimental ainsi que la dispersion de bloc, une hypothèse de stationnarité, ou intrinsèque est faite. Or, les données minières présentent souvent des comportements non stationnaires à l'échelle globale. Depuis plusieurs années, le développement des méthodes non stationnaires prend de plus en plus de place dans le domaine de la recherche, et commence à apparaître dans certains rapports techniques de ressources. Afin de pouvoir continuer à utiliser les méthodes de changement de support globaux comme outils de validation, il semble nécessaire de développer des méthodes de COS répondant aux cas non stationnaires.

CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS

Lors de l'estimation des ressources, la personne qualifié doit s'assurer de contrôler et minimiser le risque associé au projet minier grâce à l'utilisation des *meilleures pratiques*, un ensemble de règles non écrites et souvent subjectives couvrant l'ensemble de la pratique.

Parmi les nombreuses étapes techniques que le géologue doit effectuer pour produire un rapport NI43-101, trois ont été identifiées comme particulièrement importantes considérant leurs impacts économiques et sur l'avenir des projets. Pour chacun de ces trois axes, des lacunes dans les méthodes et connaissances ont été identifiées et des solutions originales permettant de répondre aux objectifs spécifiques ont été apportées.

Dans cette thèse nous avons développé de nouveaux outils qui permettent de réduire la subjectivité dans les décisions des QP. Notre algorithme d'optimisation de positionnement de forage est rapide, objectif et permet de positionner dans les zones les plus risquées les nouveaux forages. Notre outil d'évaluation des seuils d'écrêtages permet d'apporter une réponse objective à ce problème difficile tout en prenant compte de la qualité analytique permettant de faire un lien entre la théorie d'échantillonnage et les paramètres d'estimations. Enfin, nous avons évalué la sensibilité des outils de changement de support pour différents variogrammes afin de conforter les QP dans leurs utilisations des *meilleures pratiques*.

L'utilisation et l'amélioration des meilleures méthodes disponibles pour les estimés de ressources sont des obligations morales pour toutes les personnes qualifiées. A travers des outils et méthodes simples et rapides, nous avons fournis aux professionnels des solutions objectives permettant de clarifier certaines des *meilleures pratiques* associées au processus technique d'évaluation des ressources.

RÉFÉRENCES

[1] H. M. Parker, "Reconciliation principles for the mining industry," *Mining Technology*, vol. 121, n°. 3, p. 160–176, sep 2012.

[2] CIM Mineral Resources & Reserves Committee, "CIM estimation of mineral resources and mineral reserves best practice guidelines, 2nd draft," Institue Canadien des Mines, Rapport technique, 2019.

[3] J. Coombes, *The art and science of resource estimation : A practical guide for geologists and engineers*. Coombes Capability, 2008.

[4] J.-P. Chiles et P. Delfiner, *Geostatistics*. John Wiley and Sons, Inc., mar 2012.

[5] M. E. Rossi et C. V. Deutsch, *Mineral resource estimation*. Springer Science & Business Media, 2013.

[6] M. Abzalov, *Applied mining geology*. Springer, 2016, vol. 12.

[7] X. Emery et S. A. Séguret, *Geostatistics for the Mining Industry : Applications to Porphyry Copper Deposits*. Taylor & Francis, 2020.

[8] T. Romary *et al.*, *Domaining by Clustering Multivariate Geostatistical Data*. Dordrecht : Springer Netherlands, 2012, p. 455–466. [En ligne]. Disponible : https://doi.org/10.1007/978-94-007-4153-9_37

[9] X. Emery et J. M. Ortiz, "Defining geological units by grade domaining," Technical report, Universidad de Chile, Rapport technique, 2004.

[10] O. Leuangthong et M. Nowak, "Dealing with high-grade data in resource estimation," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 115, n°. 1, p. 27–36, 2015.

[11] I. Clark, "Statistics or geostatistics ? sampling error or nugget effect ?" *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 110, n°. 6, p. 307–312, 2010.

[12] D. François-Bongarçon, "Theory of sampling and geostatistics : an intimate link," *Chemometrics and intelligent laboratory systems*, vol. 74, n°. 1, p. 143–148, 2004.

[13] H. Parker, "Geostatistics. 5. volume-variance relationship-useful tool for mine planning," *E&MJ-Engineering and Mining Journal*, vol. 180, n°. 10, p. 106, 1979.

[14] E. H. Isaaks et R. M. Srivastava, "Applied geostatistics : Oxford university press," *New York*, vol. 561, 1989.

[15] R. M. Karp, "Reducibility among combinatorial problems," dans *Complexity of Computer Computations*. Springer US, 1972, p. 85–103.

[16] N. Bilal, P. Galinier et F. Guibault, “A new formulation of the set covering problem for metaheuristic approaches,” *ISRN Operations Research*, vol. 2013, 2013.

[17] G. E. Zagre *et al.*, “New tabu algorithm for positioning mining drillholes with blocks uncertainty,” *Natural Resources Research*, vol. 28, n°. 3, p. 609–629, sep 2018.

[18] P. Gy, *Sampling of particulate materials theory and practice*. Elsevier, 2012, vol. 6.

[19] J.-P. Chiles et P. Delfiner, *Geostatistics : modeling spatial uncertainty*. John Wiley & Sons, 2009, vol. 497.

[20] M. David, “Geostatistical ore reserve estimation : Elsevier sci,” *Publ Co, New York*, 1977.

[21] M. Rossi et H. Parker, “Estimating recoverable reserves : Is it hopeless ?” dans *Geostatistics for the next century*. Springer, 1994, p. 259–276.

[22] X. Emery, “On some consistency conditions for geostatistical change-of-support models,” *Mathematical geology*, vol. 39, n°. 2, p. 205–223, 2007.

[23] X. Emery, “On the consistency of the indirect lognormal correction,” *Stochastic Environmental Research and Risk Assessment*, vol. 18, n°. 4, p. 258–264, 2004.

[24] C. Demange *et al.*, “Global recoverable reserves : testing various changes of support models on uranium data,” dans *Geostatistical case studies*. Springer, 1987, p. 187–208.

[25] X. Emery et J. Ortiz, “Estimation of mineral resources using grade domains : critical analysis and a suggested methodology,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 105, n°. 4, p. 247–255, 2005.

[26] J.-P. Chilès, “Validity range of the discrete Gaussian change-of-support model and its variant,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, n°. 3, p. 231–235, 2014.

[27] R. Dutaut et D. Marcotte, “A new semi-greedy approach to enhance drillhole planning,” *Natural Resources Research*, 2020.

[28] R. Dutaut et D. Marcotte, “A comparison of indirect lognormal and discrete gaussian change of support methods for various variogram estimators,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 119, n°. 1, p. 1–9, 2019.

[29] D. Marcotte et R. Dutaut, “Linking Gy’s formula to QA/QC duplicates statistics,” *Mathematical Geosciences*, p. 1–13, 2020.

[30] M. Abzalov, “Quality control of assay data : a review of procedures for measuring and monitoring precision and accuracy,” *Exploration and Mining Geology*, vol. 17, n°. 3-4, p. 131–144, 2008.

[31] I. Clark, “Statistics or geostatistics ? sampling error or nugget effect ?” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 110, n°. 6, p. 307–312, 2010.

[32] A. J. Sinclair et G. H. Blackwell, *Applied mineral inventory estimation.* Cambridge University Press, 2006.

[33] D. G. Krige et E. J. Magri, “Studies of the effects of outliers and data transformation on variogram estimates for a base metal and a gold ore body,” *Journal of the International Association for Mathematical Geology*, vol. 14, n°. 6, p. 557–564, 1982.

[34] M. E. Rossi et C. V. Deutsch, *Mineral resource estimation.* Springer Science & Business Media, 2013.

[35] O. Parrish, “Geologist’s gordian knot : to cut or not to cut,” *Mining Engineering*, vol. 49, n°. 1, p. 45–49, 1997.

[36] J. Vann, S. Jackson et O. Bertoli, “Quantitative kriging neighbourhood analysis for the mining geologist-a description of the method with worked case examples,” dans *5th international mining geology conference*, vol. 1. Bendigo, Australia. Melbourne : Australian Inst Mining & Metallurgy, 2003, p. 215–223.

[37] D. Krige, “A basic perspective on the roles of classical statistics, data search routines, conditional biases and information and smoothing effects in ore block valuations,” dans *Proceedings of the Regional APCOM, Slovenia*, 1996.

[38] J. Deutsch, J. Szymanski et C. Deutsch, “Checks and measures of performance for kriging estimates,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, n°. 3, p. 223–223, 2014.

[39] R. Faure, B. Lemaire et C. Picouleau, “Précis de recherche opérationnelle,” *Dunod (Paris 1979)*, 1996.

[40] H. A. Taha, *Operations research : an introduction.* Pearson/Prentice Hall Upper Saddle River, NJ, USA, 2011, vol. 790.

[41] Y. C. Kim, D. E. Myers et H. Knudsen, “Advanced geostatistics in ore reserve estimation and mine planning (practitioner’s guide),” Arizona University, Rapport technique, 1977.

[42] D. E. Scheck et D.-R. Chou, “Optimum locations for exploratory drill holes,” *International Journal of Mining Engineering*, vol. 1, n°. 4, p. 343–355, 1983.

[43] M. Gershon, L. E. Allen et G. Manley, “Application of a new approach for drillholes location optimization,” *International Journal of Surface Mining, Reclamation and Environment*, vol. 2, n°. 1, p. 27–31, 1988.

[44] S. Soltani *et al.*, “Use of genetic algorithm in optimally locating additional drill holes,” *Journal of Mining Science*, vol. 47, n°. 1, p. 62–72, 2011.

[45] S. Soltani et A. Hezarkhani, “Proposed algorithm for optimization of directional additional exploratory drill holes and computer coding,” *Arabian Journal of Geosciences*, vol. 6, n°. 2, p. 455–462, 2013.

[46] N. Bilal, P. Galinier et F. Guibault, “An iterated tabu search heuristic for a variant of the partial set covering problem,” *Journal of Heuristics*, vol. 20, n°. 2, p. 143–164, 2014.

[47] A. H. Morshedy et H. Memarian, “A novel algorithm for designing the layout of additional boreholes,” *Ore Geology Reviews*, vol. 67, p. 34–42, 2015.

[48] S. Soltani-Mohammadi, M. Safa et H. Mokhtari, “Comparison of particle swarm optimization and simulated annealing for locating additional boreholes considering combined variance minimization,” *Computers & Geosciences*, vol. 95, p. 146–155, 2016.

[49] M. Pinheiro *et al.*, “Boreholes plans optimization methodology combining geostatistical simulation and simulated annealing,” *Tunnelling and Underground Space Technology*, vol. 70, p. 65–75, nov 2017.

[50] M. Safa et S. Soltani-Mohammadi, “Distance function modeling in optimally locating additional boreholes,” *Spatial Statistics*, vol. 23, p. 17–35, 2018.

[51] F. Glover, “Tabu search—part i,” *ORSA Journal on computing*, vol. 1, n°. 3, p. 190–206, 1989.

[52] MATLAB, *version 9.8.0 (R2020a).* Natick, Massachusetts : The MathWorks Inc., 2010.

[53] J. Rivoirard *et al.*, “A top-cut model for deposits with heavy-tailed grade distribution,” *Mathematical geosciences*, vol. 45, n°. 8, p. 967–982, 2013.

[54] A. Journel et A. Arik, “Dealing with outlier high grade data in precious metals deposits,” dans *Proceedings of the First Canadian Conference on Computer Applications in the Mineral Industry.* Balkema, Rotterdam, 1988, p. 161–171.

[55] H. Parker, “Statistical treatment of outlier data in epithermal gold deposit reserve estimation,” *Mathematical Geology*, vol. 23, n°. 2, p. 175–199, 1991.

[56] A. Arik et Y. Kim, “Outlier restricted kriging : a new kriging algorithm for handling of outlier high grade data in ore reserve estimation,” dans *Proceedings of the 23rd International Symposium on the Application of Computers and Operations Research in the Minerals Industry.* Society of Mining Engineers, Littleton, CO, 1992, p. 181–187.

[57] J. F. Costa, “Reducing the impact of outliers in ore reserves estimation,” *Mathematical geology*, vol. 35, n°. 3, p. 323–345, 2003.

[58] R. Machado *et al.*, “Field parametric geostatistics-geostatistical rigorous solutions dealing with extreme grades during resource estimation,” dans *Proceedings of the 35th International Symposium on the Application of Computers and Operations Research in the Minerals Industry*. Australasian Institute of Mining and Metallurgy, Melbourne, 2011, p. 191–201.

[59] M. Maleki, N. Madani et X. Emery, “Capping and kriging grades with long-tailed distributions,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, n°. 3, p. 255–263, 2014.

[60] A. Fourie, C. Morgan et R. Minnitt, “Limiting the influence of extreme grades in ordinary kriged estimates,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 119, n°. 4, p. 391–401, 2019.

[61] R. Kehmeier *et al.* Management of extreme high grade assays. [En ligne]. Disponible : https://www.rpmglobal.com/wp-content/uploads/mp/files/resources/files/rpm-perspectives-2016-131.pdf

[62] W. Roscoe, “Cutting curves for grade estimation and grade control in gold mines,” *98th Annual General Meeting, Edmonton, Alberta*, vol. 1, n°. 1, p. 8, 1996.

[63] M. Babakhani, “Geostatistical modeling in presence of extreme values,” Mémoire de maîtrise, University of Alberta, 2014.

[64] A. Chiquini et C. Deutsch, “A simulation approach to determine the cutting level,” 2017.

[65] J. Deutsch et C. Deutsch, “Fitting probability plot to identify multiple population and outliers,” *CCG Annual Report*, vol. 12, n°. 310, p. 1–6, 2010.

[66] M. Abzalov, “Quality control and assurance (QaQc),” dans *Applied Mining Geology*. Springer, 2016, p. 135–159.

[67] A. S. Méndez, “A discussion on current quality-control practices in mineral exploration,” dans *Applications and Experiences of Quality Control*. Citeseer, 2011.

[68] S. Dominy, “Importance of good sampling practice throughout the gold mine value chain,” *Mining Technology*, vol. 125, n°. 3, p. 129–141, 2016.

[69] P. Gy, *L’Échantillonnage des minerais en vrac : Théorie générale. Mémoires (France. Bureau de recherches géologiques et minières)*. Société de l’industrie minérale :[puis] Bureau de recherches géologiques et . . . , 1967.

[70] F. F. Pitard, *Pierre Gy's sampling theory and sampling practice : heterogeneity, sampling correctness, and statistical process control.* CRC press, 1993.

[71] D. Francois-Bongarcon et P. Gy, "The most common error in applying Gy's formula'in the theory of mineral sampling, and the history of the liberation factor," *Journal of the South African Institute of Mining and Metallurgy*, vol. 102, n°. 8, p. 475–479, 2002.

[72] G. Matheron, *Les variables régionalisées et leur estimation : une application de la théorie des fonctions aléatoires aux sciences de la nature.* Masson et CIE, 1965.

[73] G. Matheron, "Les principes de la géostatistique," *CG, École des Mines de Paris. Rapport*, vol. 88, 1963.

[74] B. Wilde et C. V. Deutsch, "A new approach to calculate a robust variogram for volume variance calculations and kriging," 2005.

[75] R. M. Srivastava, "A non-ergodic framework for variograms and covariance functions," Mémoire de maîtrise, Stanford University Stanford, California, 1987.

[76] R. M. Srivastava et H. M. Parker, "Robust measures of spatial continuity," dans *Geostatistics.* Springer, 1989, p. 295–308.

[77] E. Gringarten et C. V. Deutsch, "Teacher's aide variogram interpretation and modeling," *Mathematical Geology*, vol. 33, n°. 4, p. 507–534, 2001.

[78] N. Cressie, "Fitting variogram models by weighted least squares," *Journal of the international Association for mathematical Geology*, vol. 17, n°. 5, p. 563–586, 1985.

[79] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, n°. 2, p. 431–441, 1963.

[80] D. Marcotte, "Generalized cross-validation for covariance model selection," *Mathematical geology*, vol. 27, n°. 5, p. 659–672, 1995.

[81] N. Cressie et S. N. Lahiri, "Asymptotics for reml estimation of spatial covariance parameters," *Journal of Statistical Planning and Inference*, vol. 50, n°. 3, p. 327–341, 1996.

[82] N. Desassis et D. Renard, "Automatic variogram modeling by iterative least squares : univariate and multivariate cases," *Mathematical geosciences*, vol. 45, n°. 4, p. 453–470, 2013.

[83] Géovariances s.a.s., "Isatis," 2018. [En ligne]. Disponible : https://www.geovariances.com

[84] C. Han *et al.*, "New variogram modeling method using mggp and svr," *Earth Science Informatics*, vol. 9, n°. 2, p. 197–213, 2016.

[85] G. Verly, “The multigaussian approach and its applications to the estimation of local reserves,” *Journal of the International Association for Mathematical Geology*, vol. 15, n°. 2, p. 259–286, 1983.

[86] N. Cressie, “Aggregation in geostatistical problems,” dans *Geostatistics Troia’92*. Springer, 1993, p. 25–36.

[87] N. Mery, D. Marcotte et R. Dutaut, “Constrained kriging : An alternative to predict global recoverable resources,” *Natural Resources Research*, p. 1–15, 2019.

[88] S. A. Séguret et S. De La Fuente, “Diamond Drill Holes, Blast Holes and Cokriging,” dans *4th International Seminar on Geology for the Mining Industry - 8-10 July 2015 - Antofagasta Chili*. Antofagasta, Chile : Gecamin, juill. 2015. [En ligne]. Disponible : https://hal-mines-paristech.archives-ouvertes.fr/hal-01307095

[89] Wikipedia contributors, “"géostatistique non linéaire",” 2013. [En ligne]. Disponible : https://fr.wikipedia.org/wiki/Géostatistique_non_linéaire

[90] X. Emery et J. M. Ortiz, “Shortcomings of multiple indicator kriging for assessing local distributions,” *Applied Earth Science*, vol. 113, n°. 4, p. 249–259, 2004.

[91] J. Vann, D. Guibal et M. Harley, “Multiple indicator kriging–is it suited to my deposit,” dans *Proceedings of the 4th International Mining Geology Conference, Coolum, Qld, Australia*, 2000.

[92] M. Z. Abzalov, “Localised uniform conditioning (luc) : a new approach for direct modelling of small blocks,” *Mathematical Geology*, vol. 38, n°. 4, p. 393–411, 2006.

[93] G. Matheron, *Forecasting block grade distributions : the transfer functions*. Springer, 1976.

[94] J. Rivoirard, *Introduction to disjunctive kriging and non-linear geostatistics*. Clarendon Press,, 1994, n°. 551.021 R626i.

[95] C. De-Vitry, J. Vann et H. Arvidson, “A guide to selecting the optimal method of resource estimation for multivariate iron ore deposits,” dans *Proceedings of the Iron Ore Conference*. Citeseer, 2007, p. 67–77.

[96] J. Deraisme, J. Rivoirard et P. Carrasco, “Multivariate uniform conditioning and block simulations with discrete Gaussian model : application to chuquicamata deposit,” dans *VIII international geostatistics congress, GEOSTATS*, 2008, p. 69–78.

[97] J. Deraisme *et al.*, “Localised uniform conditioning in the multivariate case-an application to a porphyry copper gold deposit,” dans *Proceedings of the 35th APCOM Conference*. Australasian Institute of Mining and Metallurgy, Melbourne, 2011.

[98] W. Assibey-Bonsu *et al.*, “Production reconciliation of a multivariate uniform conditioning technique for mineral resource modelling of a porphyry copper gold deposit,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, n°. 3, p. iv–v, 2014.

[99] C. Bessin, J. Deraisme et D. Renard, “Choice of a methodological approach for the estimation of recoverable resources with uranium deposits in central jordan,” *Applied Earth Science*, vol. 124, n°. 2, p. 95–103, 2015.

[100] V. Govindsammy, “The practical implementation of uniform conditioning at AngloGold Ashanti African operations, and a case study as applied for potential underground mining at Nyankanga pit, Geita gold mine, Tanzania,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 116, n°. 7, p. 621–628, 2016.

[101] E. Maritz, “The appropriateness of the localised uniform conditioning technique for high-nugget Birimian-style gold deposits,” Thèse de doctorat, University of the Witwatersrand, School of Mining Engineering, South Africa, 2017.

[102] K. Hansmann, “When should uniform conditioning be applied ?” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 116, n°. 7, p. 645–654, 2016.

[103] G. Matheron, “The intrinsic random functions and their applications,” *Advances in applied probability*, vol. 5, n°. 3, p. 439–468, 1973.

[104] A. G. Journel, “Geostatistics for conditional simulation of ore bodies,” *Economic Geology*, vol. 69, n°. 5, p. 673–687, 1974.

[105] J. Vann, O. Bertoli et S. Jackson, “An overview of geostatistical simulation for quantifying risk,” dans *Proceedings of Geostatistical Association of Australasia Symposium" Quantifying Risk and Error*, vol. 1, 2002, p. 1.

[106] M. Armstrong *et al.*, “Scenario reduction applied to geostatistical simulations,” *Mathematical Geosciences*, vol. 45, n°. 2, p. 165–182, 2013.

[107] J. Deraisme, O. Bertoli et P. Epinoux, “Multivariate block simulations of a lateritic nickel deposit and post-processing of a representative subset,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, n°. 8, p. 673–680, 2014.

[108] E. J. Englund et N. Heravi, “Conditional simulation : practical application for sampling design optimization,” dans *Geostatistics Tróia’92*. Springer, 1993, p. 613–624.

[109] M. Abzalov *et al.*, “Optimisation of the grade control procedures at the yandi iron-ore mine, western australia : geostatistical approach,” *Applied Earth Science*, vol. 119, n°. 3, p. 132–142, 2010.

[110] D. Krige et W. Assibey-Bonsu, “Valuation of recoverable resources by kriging, direct conditioning or simulation,” dans *Proceedings of 29th APCOM Symposium : Beijing*, 2001, p. 3–8.

[111] M. Le Ravalec, B. Noetinger et L. Y. Hu, “The FFT moving average (FFT-MA) generator : An efficient numerical method for generating and conditioning Gaussian simulations,” *Mathematical Geology*, vol. 32, n°. 6, p. 701–723, 2000.

[112] X. Emery, “Conditioning simulations of Gaussian random fields by ordinary kriging,” *Mathematical Geology*, vol. 39, n°. 6, p. 607–623, 2007.

[113] C. V. Deutsch et P. W. Cockerham, “Practical considerations in the application of simulated annealing to stochastic simulation,” *Mathematical Geology*, vol. 26, n°. 1, p. 67–82, 1994.

[114] M. W. Davis, “Production of conditional simulations via the lu triangular decomposition of the covariance matrix,” *Mathematical geology*, vol. 19, n°. 2, p. 91–98, 1987.

[115] D. E. Scheck et D.-R. Chou, “Optimum locations for exploratory drill holes,” *International Journal of Mining Engineering*, vol. 1, n°. 4, p. 343–355, Dec 1983.

[116] A. McBratney, R. Webster et T. Burgess, “The design of optimal sampling schemes for local estimation and mapping of of regionalized variables .1. theory and method,” *Computers and Geosciences*, vol. 7, n°. 4, p. 331–334, jan 1981.

[117] M. Gershon, L. E. Allen et G. Manley, “Application of a new approach for drillholes location optimization,” *International Journal of Surface Mining, Reclamation and Environment*, vol. 2, n°. 1, p. 27–31, 1988.

[118] S. S. Mohammadi, A. Hezarkhani et A. E. Tercan, “Optimally locating additional drill holes in three dimensions using grade and simulated annealing,” *Journal of the Geological Society of India*, vol. 80, n°. 5, p. 700–706, nov 2012.

[119] B. Jafrasteh et N. Fathianpour, “Optimal location of additional exploratory drillholes using a fuzzy-artificial bee colony algorithm,” *Arabian Journal of Geosciences*, vol. 10, n°. 9, p. 213, May 2017.

[120] A. G. Journel, “Geostatistics for conditional simulation of ore bodies,” *Economic Geology*, vol. 69, n°. 5, p. 673–687, aug 1974.

[121] G. Matheron, “The intrinsic random functions and their applications,” *Advances in Applied Probability*, vol. 5, n°. 3, p. 439–468, dec 1973.

[122] S. Kirkpatrick, C. D. Gelatt et M. P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, n°. 4598, p. 671–680, may 1983.

[123] G. G. Pilger, J. F. C. L. Costa et J. C. Koppe, “Additional samples : Where they should be located,” *Natural Resources Research*, vol. 10, n°. 3, p. 197–207, Sep 2001.

[124] A. Journel, “The indicator approach to estimation of spatial distributions,” dans *Proceedings of the 17th APCOM International Symposium, New York*, 1982, p. 793–806.

[125] N. Bilal, “Métaheuristiques hybrides pour les problèmes de recouvrement et recouvrement partiel d’ensembles appliqués au problème de positionnement des trous de forage dans les mines,” Thèse de doctorat, Polytechnique Montreal, 2014.

[126] S. Soltani *et al.*, “Use of genetic algorithm in optimally locating additional drill holes,” *Journal of Mining Science*, vol. 47, n°. 1, p. 62–72, Jan 2011.

[127] M. Fatehi, H. A. Haroni et A. H. Morshedy, “Designing infill directional drilling in mineral exploration by using particle swarm optimization algorithm,” *Arabian Journal of Geosciences*, vol. 10, n°. 22, nov 2017.

[128] J. P. Hart et A. W. Shogan, “Semi-greedy heuristics : An empirical study,” *Operations Research Letters*, vol. 6, n°. 3, p. 107–114, 1987.

[129] C. Daoust *et al.*, “Geological setting of the paleoproterozoic Rosebel gold district, Guiana shield, Suriname,” *Journal of South American Earth Sciences*, vol. 32, n°. 3, p. 222–245, oct 2011.

[130] M. Babakhani, “Geostatistical modeling in presence of extreme values,” Mémoire de maîtrise, University of Alberta, Canada, 2014.

[131] M. David, *Geostatistical ore reserve estimation.* Elsevier, 1977.

[132] M. Abzalov, “Sampling errors and control of assay data quality in exploration and mining geology,” dans *Applications and experiences of quality control.* IntechOpen, 2011.

[133] A. G. Journel et C. J. Huijbregts, *Mining geostatistics.* Academic press London, 1978, vol. 600.

[134] M. Rossi et H. Parker, “Estimating recoverable reserves : Is it hopeless ?” dans *Geostatistics for the next century.* Springer, 1994, p. 259–276.

[135] D. Krige, “Block kriging and the fallacy of endeavouring to reduce or eliminate smoothing,” dans *Proceedings of the Regional APCOM, Moscow*, 1997.

[136] A. G. Journel et P. C. Kyriakidis, *Evaluation of mineral reserves : a simulation approach.* Oxford University Press, 2004.

[137] E. Isaaks, “The kriging oxymoron : a conditionally unbiased and accurate predictor,” dans *Geostatistics Banff 2004.* Springer, 2005, p. 363–374.

[138] M. Nowak et O. Leuangthong, “Conditional bias in kriging : Let’s keep it,” dans *Geostatistics Valencia 2016*. Springer, 2017, p. 303–318.

[139] F. C. Curriero *et al.*, “A statistical evaluation of non-ergodic variogram estimators,” *Environmental and Ecological Statistics*, vol. 9, n°. 1, p. 89–110, 2002.

[140] X. Emery et J. F. S. Torres, “Models for support and information effects : a comparative study,” *Mathematical geology*, vol. 37, n°. 1, p. 49–68, 2005.

[141] M. Le Ravalec, B. Noetinger et L. Y. Hu, “The FFT moving average (FFT-MA) generator : An efficient numerical method for generating and conditioning Gaussian simulations,” *Mathematical Geology*, vol. 32, n°. 6, p. 701–723, 2000.

[142] M. Liang, D. Marcotte et P. Shamsipour, “Simulation of non-linear coregionalization models by FFTMA,” *Computers & Geosciences*, vol. 89, p. 220–231, 2016.

[143] C. Deutsch, “Declus : a FORTRAN 77 program for determining optimum spatial declustering weights,” *Computers & Geosciences*, vol. 15, n°. 3, p. 325–332, 1989.

[144] R. A. Olea, “Declustering of clustered preferential sampling for histogram and semivariogram inference,” *Mathematical Geology*, vol. 39, n°. 5, p. 453–467, 2007.

[145] G. Matheron, “L’estimation globale des réserves récupérables, notes de cours c-75,” *Centre de Ge “ostatistique, Ecole des Mines de Paris, Fontainebleau*, p. 28, 1978.

[146] D. Marcotte et M. David, “The bi-Gaussian approach : a simple method for recovery estimation,” *Journal of the International Association for Mathematical Geology*, vol. 17, n°. 6, p. 625–644, 1985.

[147] D. G. Krige et D. Krige, *Lognormal-de Wijsian geostatistics for ore evaluation*. South African Institute of mining and metallurgy Johannesburg, 1981.

[148] G. Matheron, “A simple substitute for conditional expectation : the disjunctive kriging,” dans *Advanced geostatistics in the mining industry*. Springer, 1976, p. 221–236.

[149] M. Armstrong et G. Matheron, “Disjunctive kriging revisited : Part i,” *Mathematical Geology*, vol. 18, n°. 8, p. 711–728, 1986.

[150] C. Roth et J. Deraisme, “The information effect and estimating recoverable reserves,” dans *Proceedings of the sixth international geostatistics congress. Cape Town, South Africa*, 2000, p. 776–787.

[151] T. Romary *et al.*, “Domaining by clustering multivariate geostatistical data,” dans *Geostatistics Oslo 2012*. Springer, 2012, p. 455–466.

[152] M. J. Pyrcz *et al.*, “Representative input parameters for geostatistical simulation,” *AAPG Special Volumes*, 2006.

[153] P. Carrasco, “Nugget effect, artificial or natural ?” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 110, n°. 6, p. 299–305, 2010.

[154] G. L. Shevlyakov et H. Oja, *Robust correlation : Theory and applications.* John Wiley & Sons, 2016, vol. 3.