POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Modèles d'apprentissage automatique et d'interpolation spatiale utilisant les mesures prises durant le forage dans les mines à ciel ouvert

YAKIN HAJLAOUI
Département de mathématiques et de génie industriel

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

Août 2025

© Yakin Hajlaoui, 2025.

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Modèles d'apprentissage automatique et d'interpolation spatiale utilisant les mesures prises durant le forage dans les mines à ciel ouvert

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

Agard BRUNO, président
Michel GAMACHE, membre et directeur de recherche
Richard LABIB, membre et codirecteur de recherche
Jean-François PLANTE, membre et codirecteur de recherche
Dany LAUZON, membre
Nadia MERY, membre externe

DÉDICACE

Louange à Dieu, source infinie de savoir et de sagesse,
pour Ses bienfaits et Son accompagnement tout au long de ce chemin.
À mes parents, Ali et Salha, dont l'amour et les sacrifices ont rendu ce travail possible.
À mes sœurs, Rihab et Rawnak, et à mon frère, Rayhan, pour leur soutien indéfectible et leur précieuse présence.
À mes nobles enseignants et encadrants, dont les conseils ont éclairé mon chemin vers le savoir et la sagesse.
À mes amis, pour leur encouragement inestimable tout au long de ce parcours . . .

REMERCIEMENTS

La réussite de cette thèse de doctorat est le fruit du soutien et de la guidance exceptionnels de nombreuses personnes et organisations. Je tiens à exprimer ma profonde reconnaissance à mon directeur de recherche, Michel Gamache, dont l'encadrement et la disponibilité ont été cruciales pour l'avancement de mes travaux. Je suis également très reconnaissant envers mes codirecteurs, Richard Labib et Jean-François Plante, pour leurs contributions significatives et leurs perspectives enrichissantes.

Mes sincères remerciements s'adressent aux membres du jury, Dany Lauzon, Nadia Mery et Bruno Agard, pour le temps précieux consacré à l'évaluation de ma recherche et pour leurs remarques constructives.

Je remercie chaleureusement notre partenaire industriel, Peck Tech Consulting, pour sa précieuse collaboration et son soutien constant.

Un grand merci à mes collègues doctorants, Mohamed Maftah, Younes Aalian, Gilles Éric Palobdé Zagre et Mostafa Dadkhah, ainsi qu'à mes amis, Ali Fradi, Amine Abouda, Akrem Dhahri, Haithem Mzoughi, Anis Bouabid, Mohamed Younsi, Oussema Farhat et Gaith Mtiri, dont le soutien moral et les échanges stimulants ont enrichi mon parcours doctoral.

Enfin, du fond du cœur, je remercie ma famille et mes amis. Votre soutien indéfectible, votre patience infinie et vos encouragements constants ont été le fondement sur lequel j'ai pu construire cette thèse.

À vous tous qui avez participé de près ou de loin à cette aventure, je vous exprime ma plus sincère gratitude.

RÉSUMÉ

L'industrie minière actuelle se confronte à des défis croissants pour optimiser l'exploration et l'exploitation des ressources géologiques, notamment en raison de la complexité des formations rocheuses et de la variabilité des données de forage, influencée par des facteurs tels que l'usure du trépan, les conditions opératoires et les vibrations. Cette thèse propose des solutions innovantes pour la délimitation des massifs rocheux et l'interpolation spatiale en s'appuyant sur les données de mesures prises durant le forage de production, appelé en anglais *'measurements while drilling'* (MWD). Structurée autour de trois articles de recherche, cette thèse explore la valorisation de ces données par l'application des techniques d'apprentissage automatique et d'interpolation avancées.

Le premier article vise à estimer un indicateur de l'énergie nécessaire pour la fragmentation de la roche, appelé en anglais le *blastability index* (BI) pour des régions inexplorées. Pour ce faire, une méthode d'interpolation spatiale basée sur la rétropropagation a été développée. Cette approche intègre des techniques de paramétrisation pour tenir compte des propriétés spatiales, notamment l'anisotropie géométrique. Elle aboutit au modèle *Gradient-Based Inverse Distance Weighting with Leave-Hole-Out Cross-Validation* (GBM-IDW-LHOCV), qui se révèle robuste face aux anomalies géologiques et dont les performances sont comparables à celles du modèle processus Gaussien imbriqué (GP-nested).

Le deuxième article se concentre sur la définition de zones spatiales continues en apprenant la structure sous-jacente des données MWD, spécifiquement liée à la dureté des roches, par l'utilisation de l'empilement des autoencodeurs avec le modèle GBM-IDW-LHOCV développé dans le premier article. L'objectif est d'extraire des caractéristiques latentes, appelées pseudo-BI, qui représentent efficacement la dureté des roches. Le modèle de *clustering* résultant a démontré sa capacité à améliorer la précision des domaines spatiaux en atténuant les effets des valeurs atypiques et en assurant une meilleure contiguïté spatiale des zones géologiques.

Le troisième article présente un modèle d'interpolation multicouche (ML-IDW), inspiré des réseaux de neurones, et qui permet une ajustabilité selon la complexité des données. L'objectif de ce modèle est d'optimiser l'efficacité d'apprentissage tout en intégrant les structures de dépendance spatiale inhérentes aux données.

Ensemble, ces travaux montrent comment l'application des techniques d'apprentissage automatique et d'interpolation avancées peut offrir des solutions prometteuses pour l'analyse des données recueillies par les capteurs installés sur les équipements de forage de production afin d'améliorer la connaissance du massif rocheux, tout en soulevant des défis liés à la qualité

des données et à la complexité computationnelle. Des perspectives pour des travaux futurs incluent l'exploration d'approches de transfert d'apprentissage et l'intégration d'architectures hybrides pour améliorer l'évolutivité et la robustesse des solutions proposées.

ABSTRACT

The current mining industry faces increasing challenges in optimizing the exploration and exploitation of geological resources, primarily due to the complexity of rock formations and data variability. This thesis proposes innovative solutions for rock mass domaining and spatial interpolation, leveraging Measurements While Drilling (MWD) data. Structured around three research articles, this study explores the application of machine learning techniques and advanced interpolation methods.

The first article aims to estimate the Blastability Index (BI) for unexplored regions by developing a spatial interpolation method based on backpropagation, incorporating parameterization techniques to account for spatial properties, particularly geometric anisotropy. This approach results in the Gradient-Based Inverse Distance Weighting with Leave-Hole-Out Cross-Validation model (GBM-IDW-LHOCV), which proves robust against geological anomalies and exhibits performance comparable to that of the Nested Gaussian Process model (GP-nested).

The second article focuses on defining continuous spatial zones by learning the underlying structure of MWD data, specifically related to rock hardness, through the use of stacked autoencoders with the GBM-IDW-LHOCV model developed in the first article. The objective is to extract latent features, termed pseudo-BI, that effectively represent rock hardness. The resulting clustering model has demonstrated its ability to enhance the accuracy of spatial domains by mitigating the effects of outlier values and ensuring better spatial contiguity of geological zones.

The third article presents the Multilayer Interpolation (ML-IDW) model, inspired by neural networks, allowing for customization according to data complexity. The objective of this model is to optimize learning efficiency while integrating the inherent spatial dependency structures of data.

Together, these studies demonstrate how the application of machine learning techniques and advanced interpolation can offer promising solutions for analyzing data collected by sensors installed on production drilling equipment to enhance the understanding of the rock mass. However, they also highlight challenges related to data quality and computational complexity. Future research directions include exploring transfer learning approaches and integrating hybrid architectures to improve the scalability and robustness of the proposed solutions.

TABLE DES MATIÈRES

DÉDICACE iii

REMERCIEMENTS iv

RÉSUMÉ v

ABSTRACT vii

TABLE DES MATIÈRES viii

LISTE DES TABLEAUX xi

LISTE DES FIGURES xii

LISTE DES ANNEXES xv

CHAPITRE 1 INTRODUCTION 1
1.1 Domaine de recherche 1
1.2 Contexte 1
1.3 Éléments de la problématique 2
1.4 Objectifs de recherche 3
1.5 Plan du thèse 4

CHAPITRE 2 REVUE DE LITTÉRATURE 6
2.1 MWD et leurs applications générales 6
2.1.1 La mécanique des roches 6
2.1.2 La reconnaissance des roches et identification lithologique 7
2.2 L'essor de l'IA et de l'apprentissage automatique dans l'industrie minière . . 8
2.2.1 Les méthodes d'apprentissage automatique utilisées pour résoudre les problèmes miniers en général 9
2.2.2 Classification des méthodes utilisées 11
2.2.3 Les méthodes d'apprentissage automatique appliquées au données MWD 13
2.3 Modèles d'interpolation spatiale vs modèles d'apprentissage automatique dans le domaine de la géoscience. 14

CHAPITRE 3 VUE D'ENSEMBLE DES PROJETS DE LA THÈSE 17

3.1 Projet 1 : Inférence basée sur la rétropropagation pour l'interpolation spatiale afin d'estimer le BI dans une mine à ciel ouvert 17
3.2 Projet 2 : Utilisation des autoencodeurs variationnels et de l'interpolation spatiale pour la délimitation des masses rocheuses 18
3.3 Modèle multicouche de pondération inverse de la distance pour l'interpolation spatiale 18
3.4 Conclusion 19

CHAPITRE 4 ARTICLE 1 : BACKPROPAGATION-BASED INFERENCE FOR SPATIAL INTERPOLATION TO ESTIMATE THE BLASTABILITY INDEX IN AN OPEN PIT MINE 20
4.1 Abstract 20
4.2 Introduction 20
4.3 Models 23
4.3.1 Models based on Gaussian Processes (GPs) 24
4.3.2 Inverse Distance Weighting based models 32
4.4 Experiments and results 36
4.4.1 Data 37
4.4.2 Models' comparison 37
4.4.3 The effect of regularization on IDW 39
4.4.4 Interpreting the coefficients in M-IDW-LHOCV and GBM-IDW-LHOCV 41
4.4.5 Analyzing two-dimensional cross sections of an interpolated volume . 42
4.5 Conclusion 44

CHAPITRE 5 ARTICLE 2: INTEGRATING VARIATIONAL AUTO-ENCODERS (VAEs) AND SPATIAL INTERPOLATION FOR IMPROVING ROCK MASS DOMAINING IN OPEN PIT MINES 47
5.1 Abstract 47
5.2 Introduction 48
5.3 Theory and methodology 50
5.3.1 Auto-Enocders (AEs) 50
5.3.2 Proposed methodology 52
5.4 Experiments and results 57
5.4.1 Data and the studied mining complex 57
5.4.2 Analyzing the Relation Between Pseudo-BI, Pseudo-parameters and Original Mechanical Parameters 58
5.4.3 Collar detection 63

5.4.4 Rock Mass Domaining of Patterns with no Collar Using Pseudo-BI Based Clustering 67
5.4.5 Correspondance of Some Examples with Lithology 70
5.5 Conclusion 75

CHAPITRE 6 ARTICLE 3: ENHANCING SPATIAL INTERPOLATION: A MULTI-LAYER INVERSE DISTANCE WEIGHTING MODEL FOR COMPLEX REGRESSION AND CLASSIFICATION TASKS IN SPATIAL DATA ANALYSIS 77
6.1 Abstract 77
6.2 Introduction 78
6.3 Theory and Methodology 79
6.3.1 Conventional IDW: a brief background 79
6.3.2 Gradient Based IDW 80
6.3.3 Single Layer IDW 80
6.3.4 Mutli-Layer IDW for regression 81
6.3.5 Mutli-Layer IDW for classification 82
6.4 Experiments and results 83
6.4.1 Data description 83
6.4.2 Training efficiency 83
6.4.3 Results on regression tasks 84
6.4.4 Results on classification 86
6.5 Conclusion 90

CHAPITRE 7 CONCLUSION 93
7.1 Synthèse des travaux 93
7.2 Limitations de la solution proposée 95
7.3 Améliorations futures 97

RÉFÉRENCES 99

ANNEXES 118
A.1 Parameters' constraints : rules of thumb 118
A.2 Algorithms 119
A.3 Derivation of the matrix notation for IDW models 121
A.3.1 The gradient based conventional IDW model 121
A.3.2 The gradient based modified IDW model 121
A.4 Sufficiency of distance equation for modeling anisotropy 122

LISTE DES TABLEAUX

Table 5.1	Models parameters	57
Table 6.1	Geostatistical characteristics of the simulated datasets	84
Table 6.2	Models efficiency	84
Table 6.3	The Mean Square Error test (MSE) for each model and each dataset.	87
Table 6.4	The F1 score in percentage for each model and each dataset.	92

LISTE DES FIGURES

Figure 4.1 Vertical semivariogram with angle tolerance 0.1° and lag tolerance 1 cm, computed in a train–test scenario using real BI data from an operating open-pit mine. Red points denote the empirical semi-variogram, and the blue curve is the fitted theoretical model. 28

Figure 4.2 Horizontal semivariogram computed in a train–test scenario using real BI data from an operating open-pit mine. Red points denote the empirical semi-variogram, and the blue curve is the fitted theoretical model. 29

Figure 4.3 Partial sill/range vs azimuth: Partial sill and range are monitered by computing several semi-variograms for different azimuth angles, using a train-test scenario of real BI data extracted from an operating open pit mine. 29

Figure 4.4 Data consisting of BI measurements of the drilled holes presented in a three-dimensional coordinates system. The intensity of the color-bar indicates the BI values. 37

Figure 4.5 A two-dimensional cross section of BI values at altitude 1240 cm. . . 38

Figure 4.6 Train-test scenarios, in each scenario, the circled set of holes presents the test set and the remaining holes are the training set. 39

Figure 4.7 The evolution of the RMSE_{test} of all scenario from the worst performed model (higher RMSE) to the best performed model (lower RMSE) (boxplots) along with the correspondant variance (orange curve). . . 40

Figure 4.8 Effect of regularization on the error, the curves monitor the change in the test error (blue) and the train error (orange curve) when increasing regularization in IDW by adding constraints in IDW-C and then by adopting LHOCV inference in IDW-LHOCV. 41

Figure 4.9 The infered parameters' values in IDW-C (blue curve), IDW-LHOCV (orange curve), and GBM-IDW-LHOCV (gray curve). The figure captures their changes when moving from one IDW variant to another. . 42

Figure 4.10 The curves present the coefficients c_i in the abscissa axis and the correspondent observed BI, t_i in the ordinate axis for M-IDW-LHOCV (blue curve) and GBM-IDW-LHOCV (gray curve). The interpolation points with coefficients that are above 1.5, correspond to BI above 80 (orange line). 43

Figure 4.11 The three-dimensional spatial distribution of the coefficients of the GBM-IDW-LHOCV model (left), and the corresponding BI values (right). 44
Figure 4.12 A three-dimensional cross-section at an altitude of 1240 cm with a thickness of 1 cm. 45
Figure 4.13 Models interpolation of a cross section at altitude 1240 cm. 45
Figure 5.1 VAE architecture 52
Figure 5.2 VAE with stacked GBM-IDW-LHOCV architecture 53
Figure 5.3 Adding spatial information by encoding spatial coordinates and adding a correlation constraint 56
Figure 5.4 The reconstructed Pseudo-parameters (orange scatter) and original parameters (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with FC layers on the pattern (P1). For the rainbow scatters more intense and hot colors indicate a higher concentration of data points. 59
Figure 5.5 The reconstructed Pseudo-parameters (orange scatter) and original parameters (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with Conv layers on the pattern (P1). . 60
Figure 5.6 The reconstructed Pseudo-parameters (orange scatter) and original parameters (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with RBF layers and Gaussian basis functions on the pattern (P1). 61
Figure 5.7 The Rec-BI (orange scatter) and BI (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with FC layers on the pattern (P1). 62
Figure 5.8 Two clusters provided by Birch and agglomerative clustering applied on Pseudo-BI learned using the model with FC layers (left) vs those applied on BI (right) for the pattern (P1). 64
Figure 5.9 Two clusters provided by Birch and agglomerative clustering applied on Pseudo-BI learned using the model with Conv layers (left) vs those applied on BI (right) for the pattern (P1). 65
Figure 5.10 Two clusters provided by Birch and agglomerative clustering applied on Pseudo-BI learned using the model with RBF layers (left) vs those applied on BI (right) for the pattern (P1). 66

Figure 5.11 Box plots showing the Domain Accuracy for domains provided by performing clustering on BI, Pseudo-BI learned using the model with FC layers, Conv layers, RBF-Gaussian, RBF-inv-quad, and RBF-spline. . 67

Figure 5.12 Boxplots of BI for Each Cluster in Pattern (P1) Using FC Layers. . . 69

Figure 5.13 Subfigures illustrating the Pooled Standard Deviation and Standard Deviation of PSD for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models. 70

Figure 5.14 Subfigures illustrating Domain Accuracy (DA) Boxplots and Standard Deviation of DA for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models. 70

Figure 5.15 Comparison of Lithology for Pattern (P1) Using Pseudo-BI Clustering with FC Layers. 72

Figure 5.16 Comparison of Lithology for Pattern (P2) Using Pseudo-BI Clustering with Conv Layers. 73

Figure 5.17 Comparison of Lithology for Pattern (P3) Using Pseudo-BI Clustering with RBF-Gaussian Layers 74

Figure 6.1 (a) Training time of conventional IDW vs. number of possibilities per parameter. (b) E^{LOOCV} error vs. number of possibilities per parameter for conventional IDW. 85

Figure 6.2 Top: Simulated dataset, training set, and test set. Bottom: Kriging, Grad-IDW, and Deep-IDW predictions. 86

Figure 6.4 Mean Square Error for OK, Grad-IDW, and DeepIDW vs. data size on a simple variogram model. 89

Figure 6.5 Top: Simulated dataset, training set, and test set for classification. Bottom: IK, Simplical IK with SVC, and Deep-IDW predictions. . . 90

Figure 6.6 F1 score for IK, Simplical IK with SVC, and DeepIDW vs. data complexity on a 10,000-point simple variogram dataset. 91

Figure A.1 Illustrative example 124

Figure A.2 Ellipse for ranges 125

LISTE DES ANNEXES

Annexe A Annexe du chapitre 4 118

CHAPITRE 1 INTRODUCTION

1.1 Domaine de recherche

Depuis plus d'une dizaine d'années, de nouvelles technologies permettent la captation, la gestion et la valorisation des données d'opérations minières en temps réel. Entre autres, les équipements miniers (foreuses, camions, pelles, etc.) sont maintenant munis de plusieurs capteurs qui permettent de prendre des mesures sur différentes caractéristiques opérationnelles. Ces avancées offrent des ressources qui peuvent être utilisées pour transformer les activités d'extraction traditionnelles en interventions plus autonomes.

La notion de l'automatisation englobe plusieurs fonctionnalités clés. La première consiste à calibrer en temps réel l'opération de forage. La seconde consiste à prédire le comportement des foreuses en fonction des variations des conditions du site. Finalement, la troisième fonctionnalité permettra de mettre à jour régulièrement le modèle géologique grâce à des analyses géostatistiques sophistiquées. En outre, une valorisation de ces données permettra d'optimiser les patrons de forage et de dynamitage pour obtenir une meilleure fragmentation et pour concevoir et coordonner les trajets des foreuses afin d'améliorer l'efficacité de l'opération de forage. Parmi ces aspects, l'intégration des données de forage joue un rôle important, permettant une compréhension des caractéristiques locales du gisement.

1.2 Contexte

Le développement de capteurs pouvant être attachés à la foreuse a rendu possible l'acquisition des mesures pendant le forage (*Measurements While Drilling*-MWD). Ces données capturent divers paramètres mécaniques du processus de forage, tels que le taux de pénétration (*Rate Of Penetration* - ROP), qui mesure la vitesse à laquelle le trépan avance dans la roche (généralement en mètres par minute), la force sur le trépan (*Weight On Bit* - WOB), représentant la force axiale appliquée au trépan (mesurée en kilonewtons), le nombre de rotations par minute (*Rotations Per Minute* - RPM) du trépan, et le couple (*Torque* - TRQ), qui quantifie la force de rotation exercée sur la tige de forage (mesurée en kilonewton-mètres). La disponibilité de ces données permet leur analyse en temps réel pendant le processus de forage. Ces mesures sont enregistrées à chaque niveau de profondeur au fur et à mesure du forage. Elles offrent ainsi une image détaillée de la performance des machines tout au long du processus. Exploiter efficacement ces informations contribue à améliorer l'efficacité des opérations minières et permet des prises de décision plus rapides et mieux informées.

Cette recherche vise à développer de nouvelles méthodes pour l'analyse et la prédiction des données issues de l'exploitation minière à ciel ouvert, en se concentrant particulièrement sur les données MWD.

Une utilisation courante de ces données consiste à calculer l'énergie spécifique de forage (*Specific Energy of Drilling* - SED), qui quantifie l'énergie nécessaire pour pénétrer la roche. Cette énergie est souvent normalisée pour obtenir un indice de fragmentation connu en anglais sous le terme *Blastability Index* (BI), un indicateur clé de la dureté des roches. L'idée est simple : plus l'énergie nécessaire pour forer est élevée, plus la roche est dure. Cependant, étant donné que la dureté des roches est une propriété géologique intrinsèque, les données MWD peuvent présenter des corrélations spatiales importantes qui doivent être examinées de près.

Toutefois, le BI, défini comme une fonction explicite des variables MWD, peut être fortement influencé par des facteurs externes non géologiques (réglages de l'opérateur sur le WOB/RPM, mode de forage, pression d'air, usure du trépan, vibrations/bruit, fréquence d'échantillonnage). Ces effets introduisent de la variabilité et des biais (valeurs non directement comparables entre engins, motifs ou campagnes), si bien que le BI ne reflète pas toujours fidèlement la dureté ou la lithologie ; de plus, il ne tient pas compte explicitement de la dépendance spatiale ni de l'incertitude. Pour surmonter ces défis, des approches plus avancées, telles que l'apprentissage profond, sont nécessaires pour révéler les structures cachées au sein de ces données. C'est dans ce cadre que des méthodes d'interpolation spatiale compatibles avec les réseaux neuronaux sont proposées et étudiées dans notre recherche. Ces techniques visent à préserver les dépendances spatiales tout en offrant une grande flexibilité en termes de complexité et de modélisation, répondant ainsi aux besoins spécifiques des opérations minières.

1.3 Éléments de la problématique

Le BI est crucial pour les opérations ultérieures, comme le dynamitage, car il permet de déterminer la quantité d'explosifs nécessaire pour obtenir la fragmentation souhaitée. Cependant, une fois qu'un trou est foré, il doit être dynamité, laissant peu de marge pour ajuster l'énergie utilisée. Cette contrainte souligne la nécessité d'estimer le BI avant le début du forage. Avec de telles estimations, les mineurs peuvent prendre des décisions éclairées sur les prochaines zones à cibler, permettant ainsi un meilleur contrôle de la consommation d'énergie.

Cette problématique est abordée en estimant, *avant forage*, le BI dans les zones non forées — objet du premier projet (Chapitre 4). La difficulté tient à l'absence de mesures MWD directes

dans ces zones, malgré leur proximité avec des forages déjà réalisés dans la même région, ainsi qu'à l'incertitude associée aux données disponibles. Dans ce contexte, nous proposons un cadre d'estimation préalable du BI qui mobilise l'information issue des forages existants et les relations spatiales observées, afin d'éclairer la planification des opérations tout en tenant compte des incertitudes.

Une approche de base pourrait consister à utiliser une moyenne pondérée des observations précédentes de BI. Cependant, étant donné que le BI est lié à la dureté des roches, une propriété géologique présentant une continuité spatiale, cette corrélation doit être prise en compte dans le processus d'estimation. Ce sujet est exploré en détail dans notre premier projet, où nous présentons des approches pour aborder ce problème de manière constructive.

Définir des zones pour le dynamitage après le forage, en se basant sur le BI comme indicateur de dureté, nécessite que le regroupement des données présente une certaine contiguïté spatiale, de sorte que chaque groupe puisse être considéré comme un domaine spatial. Toutefois, la variabilité du BI, influencée par des facteurs externes tels que les réglages de l'opérateur et l'état du trépan, combinée au fait que le BI ne tient pas compte de la corrélation spatiale inhérente à la géologie, rend cette tâche particulièrement complexe. Cette problématique a motivé notre deuxième projet, axé sur la définition de ces zones en apprenant la structure sous-jacente des données MWD, spécifiquement liée à la dureté des roches, tout en intégrant l'interpolation spatiale pour capturer la similarité géologique et l'intégrer aux caractéristiques sous-jacentes. La méthodologie proposée pour ce projet est détaillée au chapitre 5 de cette thèse.

L'approche adoptée dans cette thèse, qui combine l'interpolation spatiale avec une architecture de réseau de neurones en un seul modèle de bout en bout, nous a conduits à développer des interpolateurs spatiaux dont la formulation mathématique et l'approche d'inférence (rétropropagation), ainsi que la complexité ajustable, s'inspirent des réseaux de neurones multicouches. Cependant, contrairement aux réseaux de neurones traditionnels, nos modèles, dénommés modèles de pondération par distance inverse multicouche, modélisent la structure de dépendance spatiale et prennent en compte l'anisotropie spatiale géométrique. Cette classe de modèles est largement discutée et comparée aux modèles d'interpolation spatiale conventionnels dans le chapitre 6.

1.4 Objectifs de recherche

Compte tenu des défis précédemment évoqués, l'objectif global de cette thèse est de développer des modèles et des approches adaptés aux mesures en cours de forage (MWD) dans le

contexte de l'exploitation minière à ciel ouvert. Ces modèles doivent tenir compte des caractéristiques spécifiques des données MWD, telles que l'interpolation spatiale et l'intégration des contraintes spatiales et géologiques. Cet objectif global peut être divisé en trois sous-objectifs principaux :

- Sous-objectif 1 : Développer un modèle d'interpolation spatiale adapté pour la prédiction du BI. Cela impliquera l'introduction d'un nouveau cadre de modélisation pour l'interpolation spatiale, incorporant une inférence par rétropropagation, rendant le processus plus orienté vers la tâche.
- Sous-objectif 2 : Généraliser les modèles d'interpolation spatiale développés dans le sous-objectif 1, afin de fournir une classe de modèles ajustables en termes de complexité et capables de modéliser la non-linéarité, tout en résolvant des problèmes d'interpolation spatiale plus complexes. Une étude détaillée de leurs performances et caractéristiques sera incluse.
- Sous-objectif 3 : Développer un modèle de délimitation des masses rocheuses en empilant des architectures de réseaux neuronaux, notamment des autoencodeurs variationnels (VAEs), qui sont bien adaptés pour apprendre les structures sous-jacentes. Cela sera intégré avec les modèles d'interpolation spatiale développés dans les sous-objectifs 1 et 2 pour prendre en compte la corrélation spatiale et les contraintes géologiques.

1.5 Plan du thèse

Cette thèse est structurée en six chapitres, chacun abordant des aspects clés de la problématique et des objectifs de recherche développés dans cette étude.

Le chapitre 1, intitulé *Introduction et contexte*, présente le cadre général de la recherche, en mettant en lumière les problématiques liées aux mesures pendant le forage et leur importance dans l'exploitation minière à ciel ouvert. Il définit également les objectifs principaux et les sous-objectifs de la thèse.

Le chapitre 2, *Revue de littérature*, propose une analyse approfondie des travaux existants sur l'utilisation des données MWD et les méthodes d'apprentissage automatique appliquées à l'industrie minière. Il commence par une exploration des applications générales des MWD, notamment leur rôle dans l'évaluation des propriétés mécaniques des roches et l'identification des formations lithologiques. Ensuite, l'essor des techniques d'apprentissage automatique dans le domaine minier est examiné, avec une attention particulière portée aux approches dédiées aux données MWD. Enfin, une comparaison entre les méthodes traditionnelles d'interpolation spatiale et les modèles d'apprentissage automatique est effectuée, mettant en évidence leurs avantages respectifs dans la modélisation géologique.

Le chapitre 3, *Vue d'ensemble des projets de la thèse*, expose l'organisation générale de la recherche en détaillant la progression des projets développés dans le cadre de cette étude. Il met en évidence leur contribution à l'amélioration des techniques d'interpolation spatiale et de modélisation des données MWD.

Le chapitre 4, *Inférence basée sur la rétropropagation pour l'interpolation spatiale afin d'estimer l'indice de blastabilité dans une mine à ciel ouvert*, présente le premier projet de recherche. Ce travail consiste à développer un modèle d'interpolation spatiale permettant d'estimer le BI. Les méthodes conventionnelles sont comparées à un nouveau cadre de modélisation basé sur l'inférence par rétropropagation, visant à améliorer la précision des estimations.

Le chapitre 5, *Intégration des autoencodeurs variationnels et de l'interpolation spatiale pour améliorer la délimitation des massifs rocheux dans les mines à ciel ouvert*, se concentre sur le deuxième projet de recherche. Ce projet propose une approche combinant l'apprentissage automatique et l'interpolation spatiale pour définir des zones de dynamitage en fonction de la dureté des roches. L'objectif est de modéliser la structure sous-jacente des données MWD tout en intégrant les contraintes géologiques et spatiales.

Le chapitre 6, *Amélioration de l'interpolation spatiale : un modèle multicouche de pondération à distance inverse pour des tâches de régression et de classification complexes dans l'analyse de données spatiales*, propose une généralisation des modèles d'interpolation spatiale en les rendant compatibles avec les architectures de réseaux neuronaux. Ce modèle vise à améliorer la flexibilité et la précision des interpolations en intégrant des paramètres optimisés par apprentissage. Les performances de cette approche sont évaluées et comparées aux techniques d'interpolation classiques.

Enfin, le chapitre 7, *Conclusion*, synthétise les principaux résultats obtenus au cours de cette recherche. Il met en évidence les contributions scientifiques de l'étude et propose des perspectives pour des travaux futurs en matière de modélisation des données MWD et d'approches avancées d'interpolation spatiale.

Nota bene — Publications : Les chapitres 4, 5 et 6 correspondent à des articles scientifiques évalués par les pairs (publiés/acceptés/soumis), respectivement dans *Computers and Geosciences.*, *Computers and Geosciences.* et *World Academy of Science, Engineering and Technology International Journal of Mathematical and Computational Sciences.*

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 MWD et leurs applications générales

Les MWD jouent un rôle crucial dans l'acquisition de données en temps réel pendant les opérations de forage, apportant une meilleure compréhension des caractéristiques géologiques sous-jacentes. Cette section explore deux domaines d'application principaux des MWD : la mécanique des roches et la reconnaissance ainsi que l'identification lithologique. D'une part, les MWD permettent d'analyser les propriétés mécaniques des formations rocheuses, essentielles pour optimiser les performances de forage et la stabilité des excavations. D'autre part, ces données facilitent l'identification des lithologies traversées, offrant des informations précieuses pour la modélisation géologique et la prise de décisions en exploitation minière.

2.1.1 La mécanique des roches

La mécanique des roches consiste à étudier les masses rocheuses sous diverses conditions physiques et environnementales, entre autres l'évaluation de la dureté des roches, de sa fracturation, sa résistance, sa stabilité, et sa perméabilité. Des études ont montré que les paramètres mécaniques dérivés des données MWD sont efficaces pour fournir des informations sur la mécanique des roches [35, 124, 197]. L'association entre les données MWD et les propriétés des roches est souvent explorée à travers des techniques d'apprentissage automatique et d'analyse de données. Une grande partie des recherches s'est concentrée sur la classification et la caractérisation des masses rocheuses à l'aide des données MWD. Par exemple, [35] ont utilisé des séquences de mesures MWD provenant de trous de forage pour prédire les classes de roches environnantes. De même, [197] ont relié les mesures MWD aux conditions lithologiques, comme le marbre et le marbre fracturé. [124] ont combiné les données MWD avec la photogrammétrie pour classer les roches en tant que faibles, fracturées ou compétentes.

De plus, les données MWD se sont montrées efficaces pour évaluer la qualité des roches et déterminer les conditions *in situ* afin d'assurer un soutien adéquat des roches dans les projets de tunnels, comme le montre [193]. L'association entre les données MWD et la perméabilité des roches a également été étudiée en utilisant des estimations de pression de pore et de porosité dérivées des paramètres mécaniques de forage [105], et en employant le couple et la pression sur le trépan en temps réel pour l'évaluation de la perméabilité [158].

En outre, de nombreux travaux ont été réalisés pour tirer parti des données MWD dans l'estimation de la résistance des roches afin d'évaluer la stabilité des puits. Des études telles

que [198], [51], [33] et [139] se sont concentrées sur l'utilisation des paramètres MWD pour prédire la résistance des roches et améliorer les évaluations de stabilité.

2.1.2 La reconnaissance des roches et identification lithologique

La reconnaissance des roches et l'identification lithologique se concentrent sur l'identification et la classification des différents types de roches ainsi que leurs caractéristiques géologiques et géophysiques à l'aide des données MWD. La reconnaissance des roches consiste à évaluer la dureté des roches en utilisant les données MWD, ce qui implique généralement le développement de systèmes de classification et d'indices qui relient cette propriété aux variables mécaniques fournies par MWD. Ces indices dérivés des données MWD sont souvent appelés "indices basés sur le forage", par opposition aux indices basés sur des études géotechniques traditionnelles [10].

Une étude importante menée par [104] dans une mine de cuivre à ciel ouvert en Colombie-Britannique a exploré la corrélation entre la performance des foreuses rotatives et le facteur de poudre (la quantité d'explosifs nécessaire pour le dynamitage). Cette recherche a conduit au développement d'une équation reliant le facteur de poudre à l'indice de qualité de la roche (RQI) [10, 104] :

$$\ln(\text{Powder factor}) = \frac{\text{RQI} - 24.9}{7.1},$$

où RQI=WOB/ROP, où WOP signifie *Weight on bit* et ROP *Rate of Penetration*. Cependant, les auteurs ont constaté que le RQI n'était pas un indicateur fiable des masses rocheuses et que l'approche n'avait été testée que sur un seul type de foreuse [10, 104]. D'autres études utilisant des méthodologies similaires ont développé différentes équations reliant le facteur de poudre aux données MWD via une analyse de corrélation [80, 211]. Dans une étude réalisée à la mine BOLIDEN Aitik en Suède en utilisant les données MWD du système de forage Aquila (aujourd'hui appelé CAT Terrain™ Drills), [135] ont trouvé une corrélation entre la fragmentation et l'énergie spécifique de forage (SED) dérivée des données MWD. Cette recherche a motivé un projet collaboratif entre Teck Metals et Peck Tech Consulting Ltd. en 2008 à la mine de cuivre Highland Valley (HVC) près de Kamloops, en Colombie-Britannique [10]. Le projet a abouti au développement d'un algorithme pour l'évaluation du BI, prenant en compte la présence de fractures, tel que développé par Peck Tech. Le BI est essentiellement une version normalisée du SED [10, 148]. Dans l'identification lithologique, l'objectif est généralement d'associer les données MWD avec des lithologies connues étudiées sur le terrain. Cette tâche peut aider à comprendre la géologie souterraine, à guider les opérations de fo-

rage et à améliorer la précision des modèles géologiques. De nombreux chercheurs se sont appuyés sur des indices développés pour établir de telles associations. Par exemple, [107] ont mené une analyse de l'énergie de forage pour classer les roches en catégories telles que dures, molles et moyennes. De même, [212] ont pris en compte l'indice de perçabilité des roches et l'énergie spécifique de forage pour la reconnaissance de la résistance des roches. Dans [107], la résistance à la compression uniaxiale (UCS) et l'énergie spécifique de forage ont été utilisées pour prédire différentes catégories de roches.

Cependant, comme l'ont souligné [175], les paramètres MWD et les indices développés ne correspondent pas toujours directement aux propriétés physiques des roches, car plusieurs facteurs (tels que la résistance, la dureté et la porosité) se combinent avec les différents réglages de forage pour produire les résultats MWD. De plus, des facteurs externes peuvent influencer les mesures MWD, comme l'usure du foret, les changements des conditions *in situ*, et les réglages des opérateurs humains. Par exemple, un opérateur peut augmenter le couple tandis qu'un autre augmente la pression sur le trépan (WOB), ce qui conduit à des résultats MWD différents pour des lithologies identiques. Cette variabilité rend difficile la distinction entre la lithologie et les propriétés intrinsèques des roches par rapport à d'autres facteurs influents [175].

En conséquence, de nombreux chercheurs ont exploré l'utilisation de l'apprentissage automatique et de l'intelligence artificielle pour améliorer la reconnaissance lithologique à partir des données MWD [96, 134, 181, 198, 224–226].

2.2 L'essor de l'IA et de l'apprentissage automatique dans l'industrie minière

Les récents progrès en apprentissage automatique (ML) et en intelligence artificielle (IA) au cours de la dernière décennie ont introduit un potentiel considérable pour améliorer les processus de prise de décision dans l'industrie minière. La capacité à exploiter automatiquement les données massives historiques et à apprendre à partir d'informations passées — appelée apprentissage automatique — permet de détecter des motifs cachés dans les données, aidant ainsi à améliorer les expériences passées ou à en découvrir de nouvelles. Cela dépasse les capacités de la programmation explicite traditionnelle et de la modélisation basée sur des règles physiques. En conséquence, l'industrie minière se détourne progressivement des décisions basées sur des tests empiriques de caractérisation et sur l'intuition des opérateurs, où il est difficile de gérer efficacement de grandes quantités d'informations traitées. Désormais, les entreprises minières se tournent de plus en plus vers des modèles prédictifs basés sur le ML pour traiter de vastes ensembles de données et prendre des décisions d'ingénierie éclairées, suivies par une optimisation opérationnelle.

2.2.1 Les méthodes d'apprentissage automatique utilisées pour résoudre les problèmes miniers en général

Le ML est devenu un outil crucial dans diverses applications minières, réparties en trois grands domaines : l'exploration, l'exploitation et la réhabilitation (gestion environnementale). Dans chacun de ces domaines, les méthodes de ML ont démontré leur capacité à traiter de grands ensembles de données, à détecter des motifs et à fournir des prédictions qui améliorent l'efficacité opérationnelle, réduisent les coûts et augmentent la sécurité. Cette section présente les principales études dans ces domaines, en expliquant comment le ML a été utilisé pour relever les défis spécifiques de l'industrie minière.

Dans la phase d'exploration, les techniques d'apprentissage automatique sont largement utilisées pour l'exploration minérale, la détection d'anomalies géomécaniques, la cartographie géologique et l'analyse minérale. Les études dans ce domaine se concentrent principalement sur l'utilisation des données d'exploration pour détecter des anomalies géomécaniques et réaliser une cartographie géologique en se basant sur les caractéristiques des roches. Par exemple, [217] ont appliqué des approches d'apprentissage automatique pour identifier des structures géologiques subtiles, tandis que [20], [21], [166], et [167] ont utilisé des techniques de classification et de cartographie pour analyser les données géologiques et établir des liens entre les propriétés des roches et les caractéristiques géomécaniques. De plus, l'analyse minérale a bénéficié de ces techniques en utilisant les données de forage pour prédire la composition des minéraux. Des études telles que [1], [156], [74], [45], et [85] ont permis de classifier et d'analyser la composition des minéraux, améliorant ainsi la précision de l'estimation des ressources minérales. L'apprentissage automatique joue également un rôle clé dans la planification et l'évaluation des mines, avec des modèles développés pour estimer les coûts et optimiser la conception. Par exemple, [61], et [142] ont construit des modèles pour estimer les coûts opérationnels des mines. Des études comme [3], [146], et [90] ont exploré des modèles prédictifs pour assister la planification minière, l'optimisation de la conception des mines et l'estimation des réserves [8, 129].

La phase d'exploitation des mines, qui consiste à extraire les ressources, utilise largement l'apprentissage automatique pour optimiser les processus tels que le forage, le dynamitage, le transport et la gestion des sols. Un domaine d'intérêt majeur est le forage et le dynamitage, où ces techniques sont appliquées pour prédire des résultats comme la surpression d'air [68, 141], fracturation arrière (en anglais *backbreak*) [89, 165, 194], et les projections de roches [6, 61, 117, 121, 131]. La prédiction des vibrations du sol a également fait l'objet d'une grande attention, avec des modèles développés pour prédire les niveaux de vibration en fonction des conceptions de dynamitage et des paramètres géotechniques [49, 137, 157, 170, 203, 207, 219].

De plus, l'apprentissage automatique a été utilisé pour prédire la fragmentation des roches, un facteur essentiel pour une extraction efficace des ressources [67,123], ainsi que pour évaluer les dommages aux structures environnantes causés par le dynamitage [140, 193, 229].

En plus du dynamitage et du forage, la gestion des équipements dans le secteur minier a grandement bénéficié de l'apprentissage automatique. Des études ont appliqué ces techniques pour diagnostiquer les défaillances d'équipements [43, 220] et optimiser le transport en surveillant les modes de fonctionnement des équipements [14, 15, 122, 161, 179, 204]. Les stratégies de maintenance prédictive basées sur l'apprentissage automatique ont été essentielles pour identifier les défaillances potentielles des équipements avant qu'elles ne surviennent, évitant ainsi des temps d'arrêt coûteux et améliorant l'efficacité opérationnelle [18, 83, 160].

La mécanique des roches est un autre domaine clé où l'apprentissage automatique joue un rôle vital, en particulier pour assurer la stabilité du sol dans les mines souterraines et à ciel ouvert. Des études ont appliqué ces techniques pour prédire la stabilité des piliers dans les mines souterraines, aidant à prévenir les effondrements et à assurer des opérations sûres [46, 84, 205]. La prédiction de la séismicité des excavations profondes qui vise à prédire les risques de micro-séismes qui occasionnent des éclats de roches, a également fait l'objet de recherches, les chercheurs utilisant des données de séries temporelles pour prédire les risques sismiques [19, 70, 147, 227, 228]. La prédiction de la stabilité des pentes a été abordée par des modèles de l'apprentissage automatique pour évaluer et surveiller les déformations des pentes, assurant la sécurité des opérations dans les mines à ciel ouvert [29, 47, 78, 120, 153]. En outre, des méthodes ont été utilisées dans la conception de remblais, aidant les ingénieurs à optimiser la résistance et la composition des matériaux de remblayage [118, 151, 152, 215].

La sécurité des mines est d'une importance capitale dans les opérations minières, et l'apprentissage automatique a été essentiel pour prédire et atténuer les risques. Des études ont développé des modèles pour prédire les niveaux d'eau souterraine et contrôler le drainage, réduisant ainsi les risques associés aux inondations dans les zones minières [64, 65, 201, 222]. La prédiction de la qualité de l'air et la gestion du contrôle de la poussière dans les mines ont également été abordées pour assurer la santé et la sécurité des travailleurs [28, 81, 154]. D'autres recherches ont exploré la prédiction des tremblements de terre en utilisant des données sismiques pour prévoir la probabilité d'activités sismiques dans les régions minières [27, 54], tandis que des études sur le contrôle des gaz ont appliqué l'apprentissage automatique pour prédire les émissions de gaz dangereux comme le méthane dans les mines souterraines [127,174,216]. En outre, des études sur la sécurité au travail ont utilisé l'apprentissage automatique pour prédire les probabilités d'accidents, permettant aux opérateurs de mettre en place de meilleures mesures préventives [23, 55, 87].

Dans la réhabilitation des parcs à résidus miniers, les applications de l'apprentissage automatique sont orientées vers la restauration des terres et la gestion environnementale après les opérations minières. La surveillance de la couverture terrestre joue un rôle crucial dans la réhabilitation, et des modèles ont été appliqués pour détecter les changements de surface terrestre causés par les activités minières. Ces modèles permettent de cartographier la couverture terrestre et de surveiller les altérations de surface afin de guider efficacement les efforts de réhabilitation. Des études telles que [149], [196], [136], et [108] ont démontré l'efficacité de ces approches pour surveiller et analyser les changements de la couverture terrestre.

L'évaluation des risques miniers est également un domaine important dans la réhabilitation, visant à prédire et à gérer les risques environnementaux. Les modèles ont été utilisés pour évaluer les risques liés aux changements géographiques, y compris la prédiction des glissements de terrain et de l'affaissement des mines [79,103,106,119,178,183]. La pollution des sols et de l'eau causée par les activités minières est également une préoccupation majeure, et des études telles que [7] et [22] ont utilisé des techniques d'apprentissage automatique pour prédire les niveaux de contamination et aider à guider les efforts de remédiation des sols et de l'eau, garantissant ainsi l'atténuation des risques environnementaux.

2.2.2 Classification des méthodes utilisées

Les méthodes d'apprentissage automatique appliquées aux domaines de l'exploration, de l'exploitation et de la réhabilitation minière peuvent être classées en deux grandes catégories : les méthodes supervisées et les méthodes non supervisées. Ces techniques permettent d'analyser et de prédire des phénomènes complexes en exploitant les vastes ensembles de données géologiques et opérationnelles collectées tout au long du cycle de vie des projets miniers. Ci-dessous, nous décrivons ces deux types de méthodes en détaillant leurs principes et applications dans le domaine minier.

Les méthodes supervisées sont utilisées pour des tâches de classification et de régression où un modèle est entraîné sur des données étiquetées, c'est-à-dire des données pour lesquelles la sortie attendue est connue. Dans ce contexte, des approches telles que les forêts aléatoires (*Random Forest* - RF) sont couramment utilisées, comme l'indique [167], qui ont appliqué cette méthode pour estimer des variables géochimiques manquantes dans des projets d'exploration minière. De même, les machines à vecteurs de support (SVM) sont souvent employées dans des tâches de classification géologique, notamment pour différencier les types de roches, comme le montrent [20], qui ont utilisé ces techniques pour classifier des lithologies dans une mine à ciel ouvert sur la base d'images de véhicules aériens sans pilote (UAVs) et de données de forage. En outre, les réseaux de neurones ont été utilisés dans l'étude de [156]

pour l'association entre des caractéristiques d'imagerie et des signatures de spectrométrie de fluorescence des rayons X (XRF) des roches dans les mines souterraines, permettant ainsi une identification plus précise des types de roches.

D'autre part, les méthodes non supervisées n'exigent pas de données étiquetées et sont particulièrement utiles pour explorer des structures ou des motifs cachés dans des ensembles de données complexes. [217] ont utilisé une approche de clustering combinée à un autoencodeur pour détecter des anomalies géochimiques dans les données d'exploration minière. Cette méthode permet de découvrir des structures dans des ensembles de données géochimiques sans supervision préalable. De plus, [166] ont utilisé des processus gaussiens pour classifier des images hyperspectrales dans le but d'extraire des caractéristiques géologiques, contribuant ainsi à la classification des types de roches à partir de données géophysiques. Les méthodes de réduction de dimensionnalité, bien que non explicitement mentionnées, sont souvent intégrées dans ces études, comme dans le cas de [167], qui ont réduit la complexité des données géochimiques avant d'appliquer des techniques de classification comme les forêts aléatoires.

Les méthodes hybrides, combinant des approches supervisées et non supervisées, ont également montré leur efficacité dans l'optimisation des prédictions et l'analyse des données. [74] ont combiné à la fois des techniques supervisées et non supervisées pour relier les roches protolithes aux équivalents altérés, améliorant ainsi la prédiction des types de roches et des modifications géologiques dans les zones d'exploration. Une autre approche hybride est celle utilisée par [217], qui ont intégré un réseau autoencodeur pour la réduction de dimensionnalité et un *clustering* spatial basé sur la densité pour organiser les données restantes, afin de détecter des anomalies géochimiques dans les projets d'exploration minière. En outre, [61] ont combiné un réseau de neurones profonds avec un algorithme d'optimisation pour évaluer les risques liés aux opérations de dynamitage, comme la prédiction du *flyrock*, en exploitant à la fois l'apprentissage supervisé et des méthodes d'optimisation heuristique pour améliorer la précision des résultats.

Étant donné que nous nous intéressons spécifiquement aux données MWD, notre recherche se situe dans la catégorie de l'exploitation minière. Les données MWD sont utilisées pour extraire des informations sur les propriétés mécaniques des roches en temps réel, ce qui permet d'améliorer les processus de forage et de dynamitage ainsi que la gestion des équipements. Cette catégorie met en évidence l'utilisation des techniques de forage pour optimiser les opérations d'extraction tout en garantissant la sécurité et l'efficacité dans les environnements miniers. Les modèles d'apprentissage automatique appliqués aux données MWD jouent un rôle clé dans la prédiction des résultats et l'optimisation des opérations.

2.2.3 Les méthodes d'apprentissage automatique appliquées au données MWD

Les méthodes d'apprentissage automatique appliquées aux données MWD ont été explorées dans plusieurs études afin de mieux comprendre et caractériser les propriétés du massif rocheux pendant les opérations de forage. Parmi ces études, des approches supervisées et non supervisées ont été développées pour tirer parti des données MWD et améliorer la reconnaissance des types de roches.

Dans l'apprentissage supervisé, des modèles tels que les réseaux neuronaux profonds (DNN) et les forêts aléatoires (RF) sont utilisés pour des tâches de classification et de prédiction. Par exemple, [35] ont appliqué un réseau neuronal profond BILSTM pour prédire la classification des masses rocheuses à partir des données MWD, ce qui aide à identifier la résistance et la qualité des roches pendant le forage. De même, des modèles comme les SVM sont utilisés pour prédire la capacité de forage et la résistance des roches, fournissant des informations utiles sur l'efficacité du forage et la sécurité. [124] ont utilisé ces modèles supervisés pour caractériser les masses rocheuses en combinant les données MWD avec la photogrammétrie, améliorant ainsi l'exactitude de la classification des roches.

Les techniques non supervisées, telles que le clustering et l'analyse en composantes principales (PCA), sont également essentielles pour traiter les données MWD. [197] ont utilisé une analyse multivariée pour optimiser les stratégies d'échantillonnage dans la classification des roches, en appliquant des techniques non supervisées pour extraire des motifs des données de forage sans nécessiter d'échantillons étiquetés. Cette approche permet une meilleure utilisation des données, en particulier lorsque les données étiquetées sont rares.

Les modèles hybrides, qui combinent des méthodes supervisées et non supervisées, se révèlent particulièrement efficaces pour la caractérisation des masses rocheuses. Par exemple, [134] ont utilisé un modèle hybride pour apprendre les structures géologiques 3D à partir des capteurs de la foreuse, combinant le *clustering* pour l'extraction de caractéristiques et l'apprentissage supervisé pour la classification des masses rocheuses. Ces modèles se montrent efficaces pour créer des représentations plus complètes du sous-sol, améliorant ainsi la précision et la fiabilité des prédictions des masses rocheuses.

Une autre application notable est la surveillance en temps réel et la maintenance prédictive. Les études [5] et de [56] ont appliqué des méthodes de classification et de régression en utilisant des réseaux neuronaux et des forêts aléatoires, ainsi que des réseaux neuronaux profonds et des techniques de régression, pour détecter des anomalies dans les forages horizontaux et prédire les logs soniques en temps réel. Ces modèles prédictifs aident non seulement à anticiper les problèmes comme les défaillances des équipements, mais aussi à optimiser les paramètres de

forage, réduisant ainsi les coûts et améliorant la sécurité.

Enfin, des techniques de prédiction, telles que les processus gaussiens (GP) et les projections d'espace de caractéristiques en ondelettes, ont été utilisées pour la prédiction des performances du forage. [226] ont employé une classification GP combinée avec des projections d'ondelettes pour améliorer la précision des prédictions des propriétés des roches à partir des données MWD, optimisant ainsi la performance du forage.

En conclusion, bien que les méthodes d'apprentissage automatique appliquées aux données MWD aient été largement adoptées pour des tâches de classification et de prédiction des propriétés des roches, il est important de souligner que les données MWD présentent souvent une dépendance spatiale significative. Cependant, peu d'études ont comparé les modèles d'interpolation spatiale avec les modèles d'apprentissage automatique appliqués sur les données MWD, malgré l'importance de cette dépendance spatiale. Cela ouvre la voie à une nouvelle discussion dans la section suivante : modèles d'interpolation spatiale vs modèles d'apprentissage automatique dans le domaine de la géoscience.

2.3 Modèles d'interpolation spatiale vs modèles d'apprentissage automatique dans le domaine de la géoscience.

Dans le domaine de la géoscience, le débat entre l'utilisation des modèles d'interpolation spatiale traditionnels, notamment les méthodes géostatistiques comme le krigeage, et les modèles d'apprentissage automatique (ML), prend de plus en plus d'ampleur. Cette controverse repose principalement sur la capacité respective de ces deux approches à capturer les phénomènes spatiaux complexes et à prédire des variables environnementales. Les méthodes géostatistiques, longtemps considérées comme la référence en matière de modélisation spatiale, sont largement reconnues pour leur aptitude à intégrer des dépendances spatiales explicites à travers des techniques comme le krigeage ordinaire ou universel. D'un autre côté, les modèles d'apprentissage automatique se sont imposés par leur capacité à capturer des relations non linéaires complexes dans des jeux de données volumineux, notamment grâce à des algorithmes tels que les forêts aléatoires, les réseaux neuronaux et les méthodes de *boosting*.

D'un point de vue traditionnel, les méthodes géostatistiques sont préférées dans les situations où la continuité spatiale est cruciale et où l'incertitude des prédictions doit être estimée avec précision. Le krigeage, par exemple, est souvent utilisé pour estimer des surfaces ou des volumes tout en tenant compte de la structure spatiale sous-jacente des données, grâce à des variogrammes qui permettent de modéliser ces relations spatiales [44, 168]. Cependant, les limites de ces méthodes apparaissent lorsqu'il s'agit de traiter des ensembles de données à

grande échelle ou des phénomènes non linéaires, où elles peuvent devenir inefficaces ou moins précises que les approches d'apprentissage automatique [191].

Les modèles d'apprentissage automatique, quant à eux, ont démontré une capacité exceptionnelle à modéliser des phénomènes complexes, comme le montrent plusieurs études récentes. Par exemple, dans la cartographie géochimique, les modèles d'arbres de décision, tels que les forêts aléatoires, ont surpassé les méthodes géostatistiques en termes de précision de prédiction [110]. Cependant, l'une des critiques principales envers l'apprentissage automatique est son incapacité à modéliser directement les relations spatiales inhérentes aux données géoscientifiques. Les modèles d'apprentissage automatique classiques ignorent souvent la dépendance spatiale, ce qui peut conduire à des résultats incohérents dans des applications où cette dépendance est cruciale, comme dans la cartographie des sols ou la prédiction de la hauteur de la canopée forestière [185]. En revanche, des modèles hybrides, combinant des méthodes géostatistiques avec des algorithmes de machine learning, ont émergé pour combler cette lacune. Ces approches, telles que le *Random Forest Kriging* ou le *Neural Network Kriging*, offrent une combinaison des forces des deux méthodes : la capacité des modèles de ML à capturer des relations non linéaires complexes et la capacité de la géostatistique à modéliser la continuité spatiale [34, 73].

Ainsi, la question se pose de savoir quelle approche est la plus adaptée selon les cas d'application. Pour des données fortement spatiales avec des relations globalement linéaires, les méthodes géostatistiques sont souvent privilégiées, car elles modélisent finement la structure spatiale et fournissent des incertitudes mieux maîtrisées. À l'inverse, pour des ensembles volumineux et complexes, l'apprentissage automatique offre souvent la meilleure précision prédictive, mais au prix d'une interprétation spatiale moins directe ; de plus, l'incertitude qui en découle est plus difficile à calibrer et peut être moins fiable, en particulier en présence de dépendance spatiale [2].

C'est pourquoi, dans notre recherche, nous visons à développer de nouvelles classes de modèles qui combinent la capacité des réseaux neuronaux, typiques de l'apprentissage automatique, tout en maintenant la dépendance spatiale essentielle. Contrairement aux approches actuelles qui dissocient souvent ces deux aspects, nos modèles sont conçus pour être à la fois compatibles avec les réseaux neuronaux et capables de capturer explicitement la structure spatiale des données. Cette orientation vise non seulement à améliorer la précision des prédictions, mais aussi à préserver la cohérence spatiale, enjeu critique dans les données géoscientifiques comme les mesures en forage (MWD).

Cette nouvelle approche est motivée par les limites observées dans l'application des méthodes d'apprentissage automatique classiques, qui, bien qu'efficaces pour capturer des relations

complexes et non linéaires, ne tiennent pas compte de la continuité spatiale intrinsèque [110, 185]. De plus, les méthodes géostatistiques classiques, comme le krigeage, se concentrent principalement sur la modélisation de cette continuité spatiale, mais ne peuvent pas gérer efficacement des volumes de données complexes ou des relations non linéaires [168].

En intégrant les avantages des deux mondes - l'apprentissage profond pour la modélisation non linéaire et la géostatistique pour la dépendance spatiale, nous espérons combler ces lacunes. Nos efforts s'inscrivent dans la lignée des approches hybrides telles que celles décrites par [73] et [191], mais avec l'intention d'aller plus loin en rendant les nouvelles méthodes entièrement compatibles avec les techniques modernes d'apprentissage automatique.

CHAPITRE 3 VUE D'ENSEMBLE DES PROJETS DE LA THÈSE

Ce chapitre présente une vue d'ensemble des trois projets de recherche inclus dans cette thèse, chacun apportant une contribution unique à l'amélioration des modèles d'interpolation spatiale et à l'analyse des données MWD dans l'exploitation minière à ciel ouvert. Ces travaux, tout en étant fondés sur des concepts communs, abordent progressivement des problématiques de plus en plus complexes, tout en intégrant des techniques d'apprentissage automatique et d'interpolation spatiale avancées.

3.1 Projet 1 : Inférence basée sur la rétropropagation pour l'interpolation spatiale afin d'estimer le BI dans une mine à ciel ouvert

Le premier projet propose une approche novatrice pour l'estimation de BI à différentes profondeurs en utilisant uniquement les coordonnées spatiales et les mesures de BI observées lors du forage de puits précédents. Cette étude se concentre sur l'amélioration des techniques d'interpolation spatiale, en introduisant une nouvelle méthode pour le traitement des processus gaussiens (GP) et de la pondération inverse de la distance (IDW). L'intégration de la variographie assure un meilleur ajustement entre les données et la composante spatiale, en tenant compte de l'anisotropie observée. L'optimisation par descente de gradient avec rétropropagation permet de renforcer la performance des modèles GP et IDW pour prédire le BI.

Lien avec les autres projets : Cette première approche fournit la base méthodologique pour les projets suivants, qui introduisent des architectures plus sophistiquées, comme les autoencodeurs variationnels (VAE) et les modèles multicouches. L'inférence basée sur la rétropropagation pour les modèles GP et IDW sert de point de départ pour la compréhension de l'interpolation spatiale dans un contexte géologique complexe, ouvrant la voie à des techniques d'apprentissage plus profondes.

Statut de publication : Cet article a été accepté et publié dans la revue *Computers and Geosciences*.

3.2 Projet 2 : Utilisation des autoencodeurs variationnels et de l'interpolation spatiale pour la délimitation des masses rocheuses

Le deuxième projet explore le concept de la délimitation des masses rocheuses en utilisant des données MWD pour créer des domaines spatiaux contigus, caractérisés par une dureté rocheuse homogène. Ce projet propose l'utilisation d'un modèle d'autoencodeur variationnel (VAE) couplé avec un modèle d'interpolation spatiale compatible avec des réseaux neuronaux. L'objectif est d'apprendre la structure sous-jacente des données MWD, en ciblant spécifiquement la dureté des roches, tout en capturant la similarité spatiale pour assurer la cohérence des domaines. Le VAE, en tant qu'outil d'apprentissage, permet de réduire la dimensionnalité tout en conservant des informations importantes, et les résultats sont évalués à l'aide des métriques pour mesurer la précision des domaines (*Domain Accuracy* - DA) et la consistance de la dureté rocheuse (*Pooled Standard Deviation* - PSD). Ce modèle permet non seulement de créer des domaines spatiaux, mais aussi de détecter des lithologies réelles, identifiées par des experts, et d'étendre ces résultats à une modélisation en 3D.

Lien avec les autres projets : Ce projet s'appuie directement sur les résultats du premier en empilant des modèles d'interpolation capables de préserver les relations spatiales, tout en intégrant un VAE pour apprendre la représentation sous-jacente commune des MWD.

Statut de publication : Cet article a été soumis et est en attente de révision dans *Computers and Geosciences*.

3.3 Modèle multicouche de pondération inverse de la distance pour l'interpolation spatiale

Le troisième projet introduit un modèle multicouche de pondération inverse de la distance (ML-IDW), qui combine les avantages des réseaux neuronaux multicouches (ML-NN) et du modèle d'interpolation inverse de la distance (IDW). Ce modèle tire parti des capacités de traitement des ML-NN, caractérisées par des fonctions non linéaires apprenables appliquées aux caractéristiques d'entrée, et intègre la capacité du modèle IDW à apprendre les dépendances spatiales anisotropiques. Le modèle ML-IDW est optimisé en utilisant la descente de gradient et la rétropropagation, et ses performances sont comparées aux modèles classiques d'interpolation spatiale tels que le Krigeage et l'IDW standard, sur des tâches de régression et de classification utilisant des ensembles de données spatiales simulées de complexité variable. Les résultats montrent que le modèle ML-IDW surpasse ces méthodes classiques, particuliè-

rement dans la gestion de données spatiales complexes, avec une erreur quadratique moyenne plus faible en régression et un score F1 plus élevé en classification.

Lien avec les autres projets : Le modèle ML-IDW constitue une évolution et une généralisation des modèles développés dans le premier projet en intégrant les concepts de réseaux neuronaux multicouches pour mieux capturer les relations spatiales complexes et non linéaires. Il complète ainsi les travaux précédents en proposant une solution plus robuste et flexible pour l'interpolation dans des contextes géologiques plus complexes.

Statut de publication : Cet article a été accepté et publié dans la revue de *World Academy of Science, Engineering and Technology International Journal of Mathematical and Computational Sciences*.

3.4 Conclusion

Ces trois projets illustrent l'avancement progressif des techniques d'interpolation et de modélisation, allant des méthodes classiques à des approches plus complexes et adaptées aux défis géologiques spécifiques rencontrés dans l'exploitation minière.

CHAPITRE 4 ARTICLE 1 : BACKPROPAGATION-BASED INFERENCE FOR SPATIAL INTERPOLATION TO ESTIMATE THE BLASTABILITY INDEX IN AN OPEN PIT MINE

Cet article a été publié dans la revue *Computers & Geosciences* le 6 novembre 2024, coécrit avec Richard Labib, Jean-François Plante et Michel Gamache.

Aperçu du chapitre

Ce chapitre présente une approche pour estimer le BI en exploitant l'inférence basée sur la rétropropagation appliquée aux techniques d'interpolation spatiale. En combinant l'optimisation par descente de gradient avec des modèles de pondération inverse de la distance (IDW) et des processus gaussiens (GP), cette méthode vise à mieux capturer les relations spatiales complexes et à améliorer la précision des estimations. L'étude intègre également une analyse variographique pour ajuster les modèles aux caractéristiques spatiales des données MWD.

4.1 Abstract

The blastability index (BI) is a measure that indicates the resistance of rock to fragmentation when blasting. With novel technologies, miners are now able to collect and calculate BI at different depths while drilling. In this research, we propose an approach to estimate the BI at multiple depths for new areas using only spatial locations and observed BI measurements of previously drilled holes. Spatial interpolation techniques are investigated. This study introduces a novel treatment for Gaussian Processes (GPs) and Inverse Distance Weighting (IDW). Variography is leveraged to ensure an appropriate fit between the data and the spatial component. The parameters controlling anisotropy are constrained to intervals chosen to reflect the observed anisotropy. Gradient descent with back-propagation is used for optimization. The proposed approach improves the performance of GP and IDW at predicting BI. The similarities between the IDW variant proposed and a single-layer neural network are discussed.

4.2 Introduction

Rock hardness is a critical factor in open-pit mining operations, directly influencing the efficiency of the blasting phase and overall productivity. The Blastability Index (BI) quantifies

rock hardness, representing the energy required for effective fragmentation during blasting [11]. This index plays a crucial role in optimizing drilling operations and reducing associated costs.

In this study, the BI was derived using data from sensors embedded in drilling machines, which recorded key mechanical parameters such as Rate of Penetration (ROP), Weight on Bit (WOB), Rotation per Minute (RPM), and Torque (TRQ). The BI was then calculated by normalizing the Specific Energy of Drilling (SED), following the method described in [11]. Compared to Uniaxial Compressive Strength (UCS), the BI is a more cost-effective measure for assessing rock hardness in the context of blasting, taking into account factors like fragmentation and response to explosives. This makes it an economically advantageous choice for mining operations aiming to optimize resources [225].

Recent studies have highlighted the significance of Monitoring While Drilling (MWD) data, particularly SED, for lithology identification, drilling performance evaluation, and the derivation of rock mechanical properties [42, 97, 112, 144, 214].

The MWD data in this study is similar to the data in a study in which our industrial partner played a key role [148]. This project developed an algorithm for BI that incorporated the impact of fractures in rock formations. The BI algorithm was based on the normalization of SED, with BI essentially being a normalized version of SED ($\mathrm{BI} = \mathrm{SED}/\mathrm{BI}_{\mathrm{norm}}$) [11,148]. This work was inspired by earlier studies at the BOLIDEN Aitik mine in Sweden, where MWD data proved valuable in correlating blastability with SED [135], further demonstrating the potential of MWD data to optimize mining operations.

While the BI provides valuable insights for previously drilled areas, predicting the BI for undrilled or unexplored locations holds significant potential for enhancing mining operations. Accurate BI predictions for new locations can inform decision-making in several ways. Firstly, it allows mining engineers to strategically place future boreholes in areas where rock hardness is expected to be more suitable for efficient drilling and blasting operations. This optimized placement reduces the need for unnecessary drilling, thereby lowering operational costs and minimizing delays. Moreover, predictive BI modeling can help identify zones of varying rock hardness, which is crucial for tailoring the blasting process to the specific geological conditions. In softer rock areas, less explosive material may be required, reducing costs and improving safety. Conversely, in harder rock zones, enhanced drilling techniques can be deployed in advance, ensuring that the equipment and resources are appropriate for the conditions. By extending BI predictions to untested locations, mining operations can also better manage the lifecycle of open-pit projects. It allows for more efficient long-term planning, where drilling and blasting schedules can be adapted to anticipated geological chal-

lenges. This forward-looking approach not only reduces downtime but also enhances overall productivity by ensuring that equipment and manpower are utilized effectively. Additionally, predictive models using BI can contribute to environmental sustainability. By reducing overdrilling and optimizing resource allocation, mining companies can minimize their ecological footprint, lowering the amount of energy consumed during operations and limiting the environmental disruption associated with excessive drilling and blasting. Ultimately, predicting BI for new locations represents a significant step forward in the intelligent automation of mining operations. The integration of spatial interpolation techniques and machine learning into these predictions further enhances accuracy, ensuring that the BI model aligns with the geological and spatial relationships of the mining complex. This forward-looking capability leads to smarter, more efficient, and cost-effective mining practices [48, 66].

Existing literature on BI estimation primarily focuses on utilizing geophysical or geomechanical parameters [9, 209, 210]. However, these parameters are typically available only after drilling has occurred. Limited attention has been given to predicting the BI itself before drilling, and some studies explore predicting the Formation Drillability (FD) based on various parameters, including a three-dimensional coordinate system and acoustic properties of the formation [52]. Machine learning, deep learning, and traditional geostatistical methods have been employed to enhance the accuracy of such predictions [52, 53, 88, 200].

Spatial interpolation techniques, such as Gaussian Processes (GPs) and Inverse Distance Weighting (IDW), are commonly used for spatial prediction based on coordinates [169, 189]. Many studies have used and compared these models for spatial prediction problems applied in various fields. In mining, they have been used for lithology prediction and rock mass quality prediction [4, 182]. In environmental applications, these models have been used to predict groundwater quality and spatial rainfall information [57, 173]. These studies often conduct comparative analyses with respect to the data, use cases, and underlying assumptions [77]. Geostatistics, particularly Kriging, is a standard method for spatial interpolation, employing variography for covariance model selection. GPs, considered the counterpart to Kriging for surrogate modeling and statistical learning incorporating the maximization of the marginal log-likelihood to derive its hyperparameters [41]. However, existing studies often overlook the efficiency of variography in covariance model selection and anisotropy detection [150]. In this context, our study proposes an integrated approach that combines variography and GP modeling for BI prediction, aiming to improve accuracy by incorporating spatial structures and geological information into the predictive model.

Conventional IDW methods often rely on grid search for estimating parameters, such as the exponent, to minimize Leave One Out Cross Validation (LOOCV) error [12, 114, 116].

Some studies simplify the model by fixing the exponent to two [38, 111, 208]. Our treatment of IDW exploits the differentiability of weights as functions of parameters, employing continuous optimization with back-propagation and gradient descent instead of grid search. Geological information and anisotropy are incorporated by constraining the domain search using variography.

In our study we consider a dataset comprising measurements of the BI at different three-dimensional coordinates. For each measurement $i = 1, ..., N$, t_i represents the BI recorded at location $\mathbf{s}_i = [x_i, y_i, z_i]^T$, specifically in a hole located at (x_i, y_i) with altitude z_i. Let H_j denote the set of indices of measurements for hole $j = 1, ..., M$. To evaluate predictive performance, models need to be tested on data not used during training. In a general scenario, let $\mathcal{H} \subset \{1, \ldots, M\}$ represent the set of hole indices designated for testing. The set of indices $\mathcal{T} = \cup_{j \in \mathcal{H}} H_j$ identifies data points in the test set, while its complement $\overline{\mathcal{T}}$ constitutes the training set.

For a test datum with index $i \in \mathcal{T}$, the model predicts a BI of $\widehat{t}_i$ at location s_i. The Root-Mean-Square Error (RMSE) between the observed and predicted BI serves as the performance criterion:

$$RMSE_{test} = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} (\widehat{t}_i - t_i)^2}. \tag{4.1}$$

The literature review sets the stage for understanding the current state of BI prediction, highlighting the importance of spatial coordinates, geological information, and the potential improvements offered by advanced modeling techniques. In the following sections, we present the models and methodologies in detail, including the handling of data points (datums) and the validation process.

4.3 Models

In general, spatial interpolation methods derive the estimation $\widehat{t}_j$ at location $\mathbf{s}_j$ for $j \in \mathcal{T}$ using the weighted average of the observed values t_i, $i \in \overline{\mathcal{T}}$ of the training data [100]:

$$\widehat{t}_j = \sum_{i \in \overline{\mathcal{T}}} w_{ji} t_i, \tag{4.2}$$

where w_{ji} is the weight that is associated with t_i (the target of the i^{th} observation) to estimate the true value t_j in location $\mathbf{s}_j$.

When the weight w_{ji} takes anisotropy into account, the distance between locations $\mathbf{s}_i$ and $\mathbf{s}_j$

denoted $d_a(\mathbf{s}_i, \mathbf{s}_j)$ is written

$$d_a(\mathbf{s}_i, \mathbf{s}_j) = \sqrt{\frac{(x_i - x_j)^2}{\ell_x^2} + \frac{(y_i - y_j)^2}{\ell_y^2} + \frac{(z_i - z_j)^2}{\ell_z^2}}, \tag{4.3}$$

where ℓ_x is called the range, or length-scale, of the x principal direction, and similarly for ℓ_y and ℓ_z in the principal direction of y and z respectively.

The idea of spatial interpolation relies on the hypothesis that the closer the distance between BI measures, the more similar they are. When ℓ_x, ℓ_y, and ℓ_z are unequal, the similarity also depends on the direction, a situation described as directional anisotropy, or geometric anisotropy [150]. We demonstrate in the Appendix A.4 that this equation is sufficient for accounting for anisotropy when modeling both horizontal and vertical variograms in the context of drilling data.

The value of the length-scale of each input component is inversely proportional to the influence of the associated dimension on the output result [213]. Modeling anisotropy using Eq.(4.3) is known in the Gaussian Processes literature as Automatic Relevance Determination (ARD) [159, 213].

Since Gaussian Processes (GP) and Inverse Distance Weighting (IDW) are two of the most widely used spatial interpolation algorithms in geostatistics, they will be introduced and explored in this study. We aim to improve these conventional interpolation methods for enhanced accuracy and performance by incorporating learnable parameters through back-propagation and integrating advanced geostatistical techniques.

4.3.1 Models based on Gaussian Processes (GPs)

GPs examine interpolation from a probabilistic perspective. The deterministic target values t_i at the location $\mathbf{s}_i$ are considered as realizations of the random variables $T_i = T(\mathbf{s}_i)$. To predict t_j for a point in the test set, $j \in \mathcal{T}$, we will use the observations from the training set. For convenience, we denote by $\mathbf{T} = [T_i]$, the $n \times 1$ random vector containing the training targets, hence $n = |\overline{\mathcal{T}}|$. GPs are based on the assumption that the prior probability distribution of $\mathbf{T}$ is a multivariate normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{K} + \sigma_0^2 \mathbf{I}_n$ [159]. Symbolically,

$$\mathbf{T} \sim \mathcal{N}(\mathbf{m}, \mathbf{K} + \sigma_0^2 \mathbf{I}_n), \tag{4.4}$$

where $\mathbf{m} = \mathrm{E}[\mathbf{T}] = [m_i]$ is the $n \times 1$ mean vector, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and σ_0^2 is the noise variance. The values in $\mathbf{K} = [K_{ij}]$ are derived from the model and correspond to the covariance between T_i and T_j. In the context of GPs, that covariance only depends on

the distance between the two locations, here $\mathbf{s}_i$ and $\mathbf{s}_j$. For $i, j \in \overline{\mathcal{T}}, i \neq j$,

$$\text{cov}[T_i, T_j] = K_{ij} = k(d_a(\mathbf{s}_i, \mathbf{s}_j)), \tag{4.5}$$

where $k()$ is known as the *kernel function*. Although several functions $k()$ could be used to model the covariance, Radial Basis Function (RBF) is the most common. Based on a variography described in the next section, the exponential kernel appears more appropriate for our data. We formulate the kernel so that it will correspond to the exponential semi-variogram model described in Eq.4.14 [26]:

$$k_{\exp}(d_a(\mathbf{s}, \mathbf{s}')) = \sigma_f \exp(-3d_a(\mathbf{s}, \mathbf{s}')), \tag{4.6}$$

where $\mathbf{s}$ and $\mathbf{s}'$ are arbitrary locations and σ_f is the output-scale (the partial sill in Kriging). This function also depends on the length-scales included in the distance calculation $d_a(\mathbf{s}, \mathbf{s}')$ in Eq.(4.3).

We will use constant-mean GPs, where the mean is assumed constant, in such a case $\text{E}[T_i] = m$ for all i, and $\text{E}[\mathbf{T}] = m\mathbf{1_n}$, with $\mathbf{1}_n = [1, \ldots, 1]^\top \in \mathbb{R}^n$. We denote by $\theta = (\sigma_f, \ell_x, \ell_y, \ell_z, \sigma_0, m)$ the vector of hyper-parameters of the model. They are inferred from the data and then plugged-in to yield a Bayesian model for the distribution of new t's given the observations. The inference for θ relies on maximizing the log-likelihood function,

$$l(\theta) = -\frac{1}{2}(\mathbf{t} - m\mathbf{1}_n)^\top (\mathbf{K} + \sigma_0^2 \mathbf{I}_n)^{-1} (\mathbf{t} - m\mathbf{1}_n) - \frac{1}{2}\log|\mathbf{K} + \sigma_0^2 \mathbf{I}_n| - \frac{n}{2}\log(2\pi), \tag{4.7}$$

where $\mathbf{t} = [t_i]$, $i \in \overline{\mathcal{T}}$ is the vector containing the observed target values. Generally, the optimization of this log-likelihood is a non-convex problem. A gradient-based optimization algorithm is used to infer the hyper-parameters, and back-propagation is employed for complex setups.

Using the maximum likelihood estimate (MLE) for θ as a plugin estimate, the target at a new location $\mathbf{s}_j = [x_j, y_j, z_j]^\top$ may now be obtained from the posterior probability distribution for T_j given $(t_i, \mathbf{s}_i)$ for $i \in \overline{\mathcal{T}}$, which is $N(\widehat{t}_j, \widehat{\sigma}_j^2)$ where:

$$\widehat{t}_j = \widehat{m} + \mathbf{k_j}^T (\mathbf{K} + \widehat{\sigma}_0^2 \mathbf{I}_n)^{-1} (\mathbf{t} - \widehat{m}\mathbf{1_n}); \tag{4.8}$$

$$\widehat{\sigma}_j^2 = k(0) - \mathbf{k_j}^T (\mathbf{K} + \widehat{\sigma}_0^2 \mathbf{I}_n)^{-1} \mathbf{k_j}, \tag{4.9}$$

where $\mathbf{k}_j$ is the vector of covariances between the training data and T_j. Although distances and covariances depend on the estimate $\widehat{\theta}$, we omit to indicate it explicitly for simplicity of notation. In this case $\mathbf{w_j} = (\mathbf{K} + \widehat{\sigma}_0^2 \mathbf{I}_n)^{-1} \mathbf{k_j} = [w_{ji}]_i$, is a $n \times 1$ vector containing the

interpolation weights. Using these weights, Eq.(4.8) may be rewritten as

$$\widehat{t}_j = \widehat{m} + \sum_{i \in \overline{\mathcal{T}}} w_{ji}(t_i - \widehat{m}), \tag{4.10}$$

a form more similar to the formulas for Ordinary Kriging [126].

Variography and variogram modeling are commonly used in Kriging to infer the hyper-parameters. Inspired by this treatment, we combine variography with GPs to select the appropriate kernel function, and to detect and model anisotropies.

Variogram modeling and variography

In this section, variography and geostatistical exploratory analysis were conducted to gain deeper insight into the spatial relationships between data points and to verify hypotheses regarding anisotropy. Later, we demonstrate how our modeling framework can incorporate this information to constrain the models, ensuring they align realistically with the observed spatial dependencies.

The theoretical variogram is a measure of dissimilarity (variability) of the random field $T()$ expressed as a function of $\mathbf{h}$, the oriented vector separating the positions $\mathbf{s}$ and $\mathbf{s}' = \mathbf{s}+\mathbf{h}$ [40]. The semi-variogram is defined as [40]:

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}[(T(\mathbf{s}+\mathbf{h}) - T(\mathbf{s}))^2]. \tag{4.11}$$

By expanding Eq.(4.11), the following relation between the kernel function $k()$ and the theoretical semi-variogram is obtained:

$$\gamma(\mathbf{h}) = \mathrm{Var}[T] - \mathrm{cov}[T(\mathbf{s}+\mathbf{h}), T(\mathbf{s})] = k(0) - k(d_a(\mathbf{s}, \mathbf{s}+\mathbf{h})). \tag{4.12}$$

Experimentally, the variogram is calculated using the realizations t_i, $i \in \overline{\mathcal{T}}$ and corresponds to the average dissimilarity of all pairs (t_i, t_j), $(i,j) \in \overline{\mathcal{T}} \times \overline{\mathcal{T}}$ within a neighborhood of $\mathbf{h}$. Consider the following set of indices corresponding to lag tolerance 2ϵ and angle tolerance α

$$D_{\mathbf{h}} = \left\{ (i,j) \;\middle|\; |\, \|\mathbf{s}_i - \mathbf{s}_j\| - \|\mathbf{h}\| \,| \leq \epsilon \quad \text{and} \quad \left| \frac{(\mathbf{s}_j - \mathbf{s}_i)^\top \mathbf{h}}{\|\mathbf{s}_i - \mathbf{s}_j\| \, \|\mathbf{h}\|} \right| \geq \cos(\alpha) \right\}.$$

Then the experimental semi-variogram [126] is defined as :

$$\widehat{\gamma}(\mathbf{h}) = \frac{1}{2|D_{\mathbf{h}}|} \sum_{(i,j)\in D_h} (t_i - t_j)^2. \tag{4.13}$$

For a fixed direction represented by the unit vector $\mathbf{u}$, the variography is the plot of the experimental semi-variogram $\widehat{\gamma}(h) = \widehat{\gamma}(h\mathbf{u})$ as a function of the lag distance $h = \left\|\mathbf{h}\right\|$. Notably, $\mathbf{u} = [0, 0, 1]^\top$ yields the vertical variogram, $\gamma^{(v)}$ which is shown in Fig.4.1. A slightly different definition of $D_{\mathbf{h}}$ yields the horizontal variogram, $\gamma^{(h)}$ in Fig.4.2. The latter leverages a working assumption of horizontal isotropy and defines lag tolerance on the difference in altitude. The semi-variogram models (blue lines) that are used to fit the experimental variograms are:

$$\gamma^{(h)}(h) = \sigma_0 + \sigma_f^{(h)}(1 - \exp(-3h/a_h)), \tag{4.14}$$

$$\gamma^{(v)}(h) = \sigma_0 + \sigma_f^{(v)}(1 - \exp(-3h/a_v)), \tag{4.15}$$

where $a_h = \ell_x = \ell_y$ is the horizontal range, $\sigma_f^{(h)}$ is the horizontal partial sill, and similarly for $a_v = \ell_z$ and $\sigma_f^{(v)}$ in the vertical direction.

We use the traditional Weighted Least Square (WLS) approach developed by [39] to determine parameter values by minimizing

$$\sum_{\mathbf{h}\in\mathcal{V}} |D_{\mathbf{h}}| \left\{ \frac{\widehat{\gamma}(\mathbf{h})}{\gamma(\mathbf{h})} - 1 \right\}^2, \tag{4.16}$$

where $\mathcal{V}$ is a set of vectors that are carefully chosen to make the sets $D_{\mathbf{h}}$ a partition of the data. For variograms with a fixed direction, notably $\gamma^{(v)}$, or for $\gamma^{(h)}$ that assumes isotropy, the sets may simply be represented by scalar values $h = ||\mathbf{h}||$.

The variograms and their estimated parameters are computed on the training data. In the following sections, six different scenarios for train-test are considered. Note that Fig.4.1 and Fig.4.2 were computed based on one of these scenarios which also yields the following estimates

$$\begin{aligned}
\widehat{\sigma}_0 &= 12, && \text{the nugget (i.e noise variance),}\\
\widehat{\sigma}_f^{(v)} &= 45.79, && \text{the partial sill of the vertical semi-variogram,}\\
\widehat{\sigma}_f^{(h)} &= 36.78, && \text{the partial sill of the horizontal semi-variogram,}\\
\widehat{a}_v &= 9.62, && \text{the range of the vertical semi-variogram,}\\
\widehat{a}_h &= 18.23, && \text{the range of the horizontal semi-variogram.}
\end{aligned}$$

Note that the changes in the variograms and their parameter values are not noticeably different from one train-test scenario to another. The experimental semi-variograms show different ranges for the horizontal direction (direction perpendicular to the holes), and the vertical direction (parallel to the holes), which leads us to conclude to the existence of directional anisotropy. Similarly, we notice different sills on these semi-variograms, which indicates the existence of zonal anisotropy. In Fig.4.3 we monitored the range and the sill values for different azimuth angles varying in the interval $[0^{\circ}, 180^{\circ}]$ in the horizontal plane. We observe that the partial sill and the range are quasi-constant with respect to the azimuth. The assumption on the semi-variograms isotropy on the horizontal plane (x, y) could reasonably be considered valid.

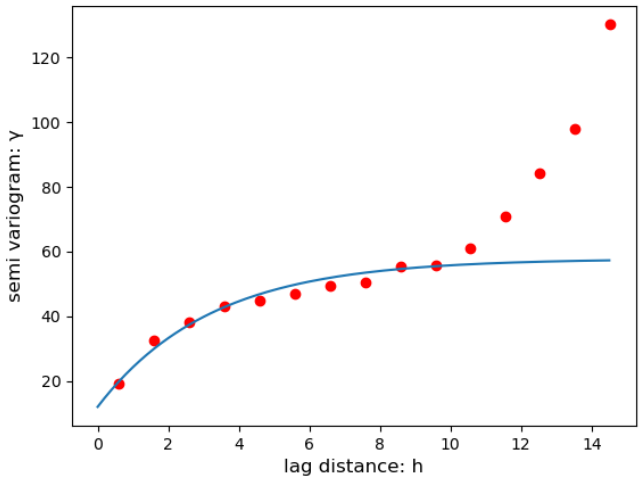


Figure 4.1 Vertical semivariogram with angle tolerance 0.1° and lag tolerance 1 cm, computed in a train–test scenario using real BI data from an operating open-pit mine. Red points denote the empirical semi-variogram, and the blue curve is the fitted theoretical model.

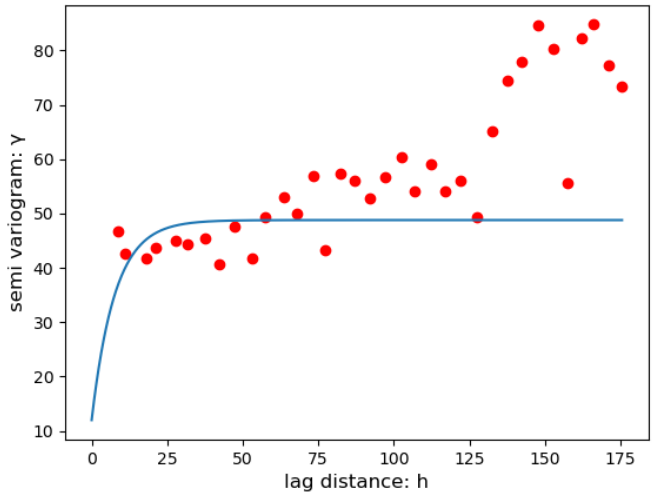


Figure 4.2 Horizontal semivariogram computed in a train–test scenario using real BI data from an operating open-pit mine. Red points denote the empirical semi-variogram, and the blue curve is the fitted theoretical model.

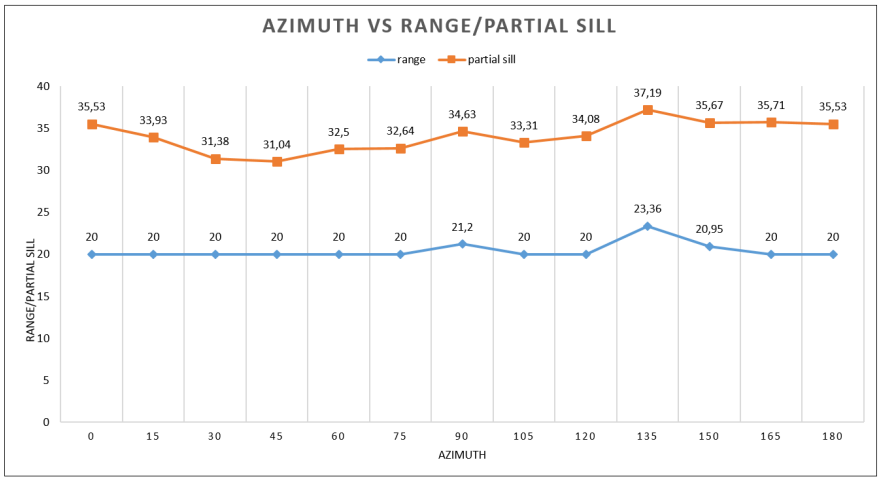


Figure 4.3 Partial sill/range vs azimuth: Partial sill and range are monitered by computing several semi-variograms for different azimuth angles, using a train-test scenario of real BI data extracted from an operating open pit mine.

In this section, we used variography to identify some spatial properties such as directional and zonal anisotropies. We will leverage these characteristics as well as the observed parameter values to determine a domain therefore in our proposed approach.

Combining Gaussian Processes and variography information

The originality of this study is to leverage the information extracted from variography to make better choices in GP modeling. In addition to selecting kernels that match the structure unveiled by the variograms, we add constraints to the optimization of the GP to ensure that the parameters with physical interpretations are constrained to values that are empirically compatible with those variograms.

In Sect.4.3.1, we discuss the ability of GP-based (Gaussian Process) machine learning to infer scales automatically, thereby capturing geometric anisotropy through Automatic Relevance Determination (ARD). This feature allows the model to adaptively weight different input dimensions, which is particularly advantageous in spatial interpolation where anisotropic patterns are present. By doing so, GPs can effectively model complex spatial dependencies and directional trends in the data. However, while this flexibility is highly desirable, it introduces numerical challenges that can lead to solutions misrepresenting the underlying geometry of the data. Specifically, the model may become more sensitive to outliers and atypical data points. This sensitivity can result in overfitting, where the GP model fits the noise in the data rather than the true underlying spatial structure, as noted by [30]. Such numerical instability underscores the importance of careful model calibration and validation to ensure that the inferred geometric anisotropy accurately reflects the true spatial patterns in the data.

Additionally, these numerical challenges can complicate the optimization process, potentially leading to convergence issues or suboptimal parameter estimates. Therefore, while GP-based models with ARD offer powerful tools for capturing anisotropy in spatial data, they require robust handling of numerical issues to fully realize their potential in accurately modeling spatial phenomena. In such case, variography in Kriging could be used to make GP more robust [186]. It is recognized as an effective tool in covariance-model selection and in detecting both geometric and zonal anisotropies [92](see Sect.4.3.1). In addition to using variography to select an appropriate covariance model, we use the same representations to manually extract intervals and bound the hyper-parameters of the GP model such as the output-scale, and the length-scales. A parametrization approach is applied in the gradient descent algorithm to enforce those domain constraints yet retain some automatic inference. For a generic parameter θ that we would like to constrain to the interval $[\theta_{\text{lower}}, \theta_{\text{upper}}]$, the

reparametrization approach of [94] consists in expressing θ as a differentiable function of the unrestricted raw parameter $\widetilde{\theta} \in \mathbb{R}$:

$$\theta = g(\widetilde{\theta}) = (\theta_{upper} - \theta_{lower})\frac{1}{1 + e^{-\widetilde{\theta}}} + \theta_{lower}, \tag{4.17}$$

ensuring that for $\widetilde{\theta} \in \mathbb{R}$, we have $\theta = g(\widetilde{\theta}) \in [\theta_{lower}, \theta_{upper}]$.

The likelihood function in Eq.(4.7) comprises six parameters, namely $\theta = (\sigma_f, \sigma_0, \ell_x, \ell_y, \ell_z, m)$. Among these six parameters, m remains unconstrained, and σ_0 will be fixed. To add some flexibility, instead of fixing σ_f, we preferred to optimize it over a small interval bounding all partial sills displayed by the variograms. Appendix A.1 shows how this interval is determined. The remaining optimizable parameters are reparametrized, leading to $\widetilde{\theta}$ and the likelihood function may be written as

$$l(\theta) = l(g(\widetilde{\theta})). \tag{4.18}$$

In general, variography requires some subjective choices in the sense that different experts will normally come to similar but not identical conclusions. Similarly, we followed intuition to select appropriate intervals from the variograms. Guidelines for those choices are discussed in Appendix A.1, but expert knowledge could also be leveraged.

Back-propagation algorithm is applied on $\widetilde{\theta}$ in this case instead of θ. Algorithm 3 in Appendix A.2 presents the main steps of the inference approach.

In the previous section, exponential variogram models were appropriate choices to fit the empirical variograms, which yields the study of a simple GP model with a simple exponential kernel function. However, in Sect.4.3.1, we noticed that the spatial structure exhibits a zonal anisotropy in which the partial sill changes with directions. Since simple kernel functions such as the exponential kernel are unable to model such cases, a more complex composed model is required. This model namely the nested model will be studied as well to account for zonal anisotropy.

The simple model We model the covariance matrix using the simple exponential kernel function:

$$K_{ij} = k(d_a(\mathbf{s}_i, \mathbf{s}_j)) = \sigma_f \exp(-3d_a(\mathbf{s}_i, \mathbf{s}_j)). \tag{4.19}$$

In the simple exponential, only directional anisotropy is modeled through $d_a()$ using ARD. Zonal anisotropy however is not modeled because the output-scale (partial sill) is invariant when changing directions. The hyper-parameters $\theta = (\sigma_f, \sigma_0, \ell_x, \ell_y, \ell_z, m)$ of the likelihood function are inferred using Algorithm 3. The domain for the vertical partial sill (i.e the

maximum partial sill) $\sigma_f^{(v)}$ will be used to constrain σ_f.

The nested model We propose to model zonal anisotropy by adding two simple exponential kernel functions leading to the following nested model:

$$K_{ij} = \left(\sigma_f^{(v)} - \sigma_f^{(h)}\right) \exp\left(-3\sqrt{\frac{(z_i - z_j)^2}{\ell_z^2}}\right) + \sigma_f^{(h)} \exp\left(-3\sqrt{\frac{(x_i - x_j)^2}{\ell_x^2} + \frac{(y_i - y_j)^2}{\ell_y^2} + \frac{(z_i - z_j)^2}{\ell_z^2}}\right), \tag{4.20}$$

where $\sigma_f^{(v)}$ is the partial sill along the vertical direction, and $\sigma_f^{(h)}$ is the partial sill along the horizontal direction. In the horizontal direction, $z_i = z_j$, and Eq.(4.20) is reduced to a simple exponential kernel with partial sill $\sigma_f^{(h)}$ and nugget $\sigma_f^{(v)} - \sigma_f^{(h)}$:

$$\begin{aligned} K_{ij} &= \sigma_f^{(v)} - \sigma_f^{(h)} + \sigma_f^{(h)} \exp\left(-3\sqrt{\frac{(x_i - x_j)^2}{\ell_x^2} + \frac{(y_i - y_j)^2}{\ell_y^2}}\right) \\ &= \sigma_f^{(v)} - \sigma_f^{(h)} + \sigma_f^{(h)} \exp\Big(-3\, d_a(\mathbf{s}_i, \mathbf{s}_j)\Big). \end{aligned} \tag{4.21}$$

In the vertical direction, we have that $x_i = x_j$ and $y_i = y_j$, hence Eq.(4.20) is reduced to a simple exponential kernel with partial sill $\sigma_f^{(v)}$:

$$K_{ij} = \sigma_f^{(v)} \exp\left(-3\sqrt{\frac{(z_i - z_j)^2}{\ell_z^2}}\right) = \sigma_f^{(v)} \exp(-3 d_a(s_i, s_j)). \tag{4.22}$$

This GP-nested approach models the zonal anisotropy and is compatible with the variography in Sect.4.3.1. The hyper-parameters $\theta = (\sigma_f^{(v)}, \sigma_f^{(h)}, \sigma_0, \ell_x, \ell_y, \ell_z, m)$ of the likelihood function are inferred using Algorithm.3.

4.3.2 Inverse Distance Weighting based models

Background

Unlike GPs which are based on probabilities, IDW is a deterministic (non-probabilistic) interpolation approach. The weights then associated with our targets, w_{ji}, $(i, j) \in \overline{\mathcal{T}} \times \mathcal{T}$ are parametric functions of the distance [100, 171]:

$$w_{ji} = \frac{d_a(\mathbf{s}_i, \mathbf{s}_j)^{-p}}{\sum\limits_i d_a(\mathbf{s}_i, \mathbf{s}_j)^{-p}}, \tag{4.23}$$

with weight functions depending on the exponent parameter p and the length-scales ℓ_x, ℓ_y, and ℓ_z used implicitly to compute the distances $d_a(\mathbf{s}_i, \mathbf{s}_j)$ (see Eq.(4.3)). The studies that account for geometric anisotropy rely on variography to infer ℓ_x, ℓ_y, and ℓ_z [169,189,208]. While some studies fix $p = 2$ claiming that it is suitable for geological interpolation [38, 111, 208], others infer p and Leave One Out Cross Validation (LOOCV) is the most common approach therefor [12, 114, 116]. The method consists of searching over a discretized domain $G = \{1, 1.5, 2, ...\}$ to choose p that minimizes the LOOCV Error given by $E^{LOOCV} = (1/|\overline{\mathcal{T}}|) \sum_{j \in \overline{\mathcal{T}}} (\widehat{t}_j - t_j)^2$ where $\widehat{t}_j = \sum_{(i,j) \in \overline{\mathcal{T}}^2, i \neq j} w_{ji} t_i$ is the IDW estimation of the real observation t_j using all the other observations and leaving t_j out. Algorithm 1 in Appendix A.2 shows the detailed steps of the LOOCV method used on the training data $D_{train} = \{(s_i, t_i),\ i \in \overline{\mathcal{T}}\}$ to infer the parameter p. To improve IDW, [38] proposed a modification of the conventional approach. The errors $e_i = (\ \widehat{t}_i - t_i)^2$, $i \in \overline{\mathcal{T}}$ between the estimations $\widehat{t}_i$ and the real values t_i, are used to compute coefficients c_i, expressed as:

$$c_i = 1 + \frac{e_i}{\max\limits_{v \in \overline{\mathcal{T}}} e_v},\ i \in \overline{\mathcal{T}}, \tag{4.24}$$

and used to adjust the interpolation weights as follows:

$$\widetilde{w_{ji}} = \frac{c_i . w_{ji}}{\sum\limits_{u \in \overline{\mathcal{T}}} c_u . w_{ju}}. \tag{4.25}$$

The interpolation is performed using the modified weights $\widetilde{w_{ji}}$ instead of w_{ji}:

$$\widehat{t}_j = \sum_{i \in \overline{\mathcal{T}}} \widetilde{w_{ji}} t_i, \text{ for all } j \in \mathcal{T}. \tag{4.26}$$

[38] noticed that if $e_i = \max_{v \in \overline{\mathcal{T}}} e_v$, then $c_i = 2$, and the influence of t_i when interpolating is doubled. However, if $e_i = 0$, then $c_i = 1$, and the influence of t_i is unchanged. They concluded that the coefficients c_i emphasize low-accuracy (high-error) interpolation points.

Algorithm 2 in Appendix A.2 presents the inference of p of the modified approach using LOOCV.

Gradient based inference for IDW

In this section, we propose a new IDW treatment that is inspired by neural networks and leverages a back-propagation algorithm. The strategy deployed for GPs that consists in constraining the length-scale parameters using variography information will be used here as

well.

The Gradient-based conventional IDW: We present a gradient-based inference framework for the IDW model. Conventional anisotropic IDW and the modified IDW depend on the parameters $\theta = (p, \ell_x, \ell_y, \ell_z)$. In our approach, these parameters are inferred using gradient-based optimization. The main idea is to model IDW as a single-layer neural network and use back-propagation to infer the IDW parameters.

In the LOOCV approach, the error $E^{LOOCV}(\theta) = (1/|\overline{\mathcal{T}}|)\sum_{j\in\overline{\mathcal{T}}}(\widehat{t}_j - t_j)^2$ must be minimized, where $\widehat{t}_j = \sum_{i\in\overline{\mathcal{T}}, i\neq j}\widetilde{w_{ji}}t_i$ is the IDW estimation of the real observation t_j of the training data, using all the other observations and leaving t_j out. In our approach, we use the matrix notation to exploit parallelism in the calculation of $\widehat{t}_j$ for all j. We show in Appendix A.3.1 that the LOOCV applied on IDW could be equivalently performed by minimizing the following error:

$$E^{LOOCV}(\theta) = \frac{1}{|\overline{\mathcal{T}}|}\left\|\widehat{\mathbf{t}} - \mathbf{t}\right\|^2 = \frac{1}{|\overline{\mathcal{T}}|}\left\|\mathbf{W}(\theta)\mathbf{t} - \mathbf{t}\right\|^2, \tag{4.27}$$

where $\mathbf{t} = [t_i]$, and $\widehat{\mathbf{t}} = [\widehat{t}_i]$ are the vectors containing respectively the true values and the IDW predicted values for $i \in \overline{\mathcal{T}}$, $\mathbf{W}$ is the matrix with a zero diagonal and the IDW weights w_{ji} elsewhere. As usual, $\left\|.\right\|$ is the Euclidean norm.

Since $\mathbf{W} = \mathbf{W}(\theta)$ is differentiable with respect to the parameter vector θ, $\widehat{\mathbf{t}} = \mathbf{W}(\theta)\mathbf{t}$ could be seen as a forward step of a single-layer neural network that has input vector $\mathbf{t}$ and output vector $\widehat{\mathbf{t}}$. With this setup, we infer θ by minimizing Eq.(4.27) using back-propagation. As with neural networks, the inference algorithm consists of two main steps (see e.g. [58]):

1. forward step : compute $E^{LOOCV}(\theta) = \frac{1}{|\overline{\mathcal{T}}|}\left\|\mathbf{W}(\theta)\mathbf{t} - \mathbf{t}\right\|^2$;
2. backward step: $\theta \leftarrow \theta - \alpha\frac{\partial E^{LOOCV}(\theta)}{\partial\theta}$, where α is the learning rate.

In anisotropic IDW, variography could be used to restrict the domain search of the length-scales ℓ_x, ℓ_y, and ℓ_z. The exponent parameter p must be positive and its domain search must be restricted to $[0, +\infty[$. The reparametrization technique is used to model these restrictions. As indicated in Eq.(4.17), a function $g()$ and an unrestricted vector of raw parameters $\widetilde{\theta} = (\widetilde{\ell_x}, \widetilde{\ell_y}, \widetilde{\ell_z}, \widetilde{p})$ are introduced so that $\theta = g(\widetilde{\theta})$ lies in the desired domain. The LOOCV error in this case is a differentiable function of $\widetilde{\theta}$:

$$E^{LOOCV}(\theta) = E^{LOOCV}(g(\widetilde{\theta})) = \frac{1}{|\overline{\mathcal{T}}|}\left\|\mathbf{W}(\mathbf{g}(\widetilde{\theta}))\mathbf{t} - \mathbf{t}\right\|^2. \tag{4.28}$$

The inference is then applied with respect to $\widetilde{\theta}$ instead of θ:

1. forward step : compute $E^{LOOCV}(g(\widetilde{\theta})) = \frac{1}{|\overline{\mathcal{T}}|}\left\|\mathbf{W}(\mathbf{g}(\widetilde{\theta}))\mathbf{t} - \mathbf{t}\right\|^2$;

2. backward step: $\widetilde{\theta} \leftarrow \widetilde{\theta} - \alpha \frac{\partial E^{LOOCV}(g(\widetilde{\theta}))}{\partial \widetilde{\theta}}$, where α is the learning rate.

Algorithm 4 in Appendix A.2 shows the detailed steps of IDW's gradient-based inference approach.

The gradient-based modified model: The modified IDW approach introduced in Sect.4.3.2 consists of adjusting the IDW weights w_{ji} using the coefficients c_i defined in Eq.(4.24). As illustrated previously, the LOOCV error in the case of the modified approach could be computed in one single forward step:

$$E^{LOOCV}(g(\widetilde{\theta})) = \frac{1}{|\overline{\mathcal{T}}|} \left\| \widehat{\mathbf{t}} - \mathbf{t} \right\|^2 = \frac{1}{|\overline{\mathcal{T}}|} \left\| \frac{\mathbf{W}(\mathbf{g}(\widetilde{\theta}))[\mathbf{c} \odot \mathbf{t}]}{\mathbf{W}(\mathbf{g}(\widetilde{\theta}))\mathbf{c}} - \mathbf{t} \right\|^2, \tag{4.29}$$

where $\mathbf{c} = [\mathbf{c_i}]$ is the vector that contains the coefficients for all $i \in \overline{\mathcal{T}}$; and the operator $\odot$ is the element-wise product ($\mathbf{c} \odot \mathbf{t} = [c_i t_i]$). The division on vectors is also performed element-wise ($\mathbf{c}/\mathbf{t} = [c_i/t_i]$) through this paper. The detailed derivation of Eq.(4.29) is presented in Appendix A.3.2. Notice that $E^{LOOCV}(g(\widetilde{\theta}))$ is differentiable with respect to $\widetilde{\theta}$. Algorithm 4 in Appendix A.2 could thus be used to infer $\widetilde{\theta}$.

The coefficients vector could be considered as a parameter instead of being calculated using Eq.(4.24). In this case, $E^{LOOCV}(g(\widetilde{\theta}))$ is differentiable with respect to $\mathbf{c}$. Next, we propose a novel Gradient Based Modified IDW (GBM-IDW) in which the coefficients $\mathbf{c} = [c_i]$ are derived using back-propagation along with θ. By considering that the coefficients are restricted to the interval $[1, 2]$, the reparametrization approach will be employed. A parameter vector for unrestricted raw coefficients $\widetilde{\mathbf{c}} \in \mathbb{R}^n$ and a differentiable function h are therefore defined such that $\mathbf{c} = h(\widetilde{\mathbf{c}}) \in [1, 2]^n$. The inference of $\mathbf{c}$ and θ comprises two main steps:

1. forward step : Compute $E^{LOOCV}(\theta, \mathbf{c})$:
   $E^{LOOCV}(\theta, \mathbf{c}) = E^{LOOCV}(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}})) = \frac{1}{|\overline{\mathcal{T}}|} \left\| \frac{\mathbf{W}(\mathbf{g}(\widetilde{\theta}))[\mathbf{h}(\widetilde{\mathbf{c}}) \odot \mathbf{t}]}{\mathbf{W}(\mathbf{g}(\widetilde{\theta}))[\mathbf{h}(\widetilde{\mathbf{c}}) \odot \mathbf{1}]} - \mathbf{t} \right\|^2$;
2. backward step : $\widetilde{\theta} \leftarrow \widetilde{\theta} - \alpha \frac{\partial E(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}}))}{\partial \widetilde{\theta}}$, $\widetilde{\mathbf{c}} \leftarrow \widetilde{\mathbf{c}} - \alpha \frac{\partial E(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}}))}{\partial \widetilde{\mathbf{c}}}$.

Algorithm 5 in Appendix A.2 shows the detailed process of inference in the new GBM-IDW. Because the coefficients are considered parameters, GBM-IDW might be more flexible than the conventional modified approach proposed by [38].

Leave Hole Out Cross Validation Interpolation: In the previous sections, we demonstrated that using our approach, LOOCV could be performed in one single forward step by computing $E^{LOOCV}(\theta) = 1/|\overline{\mathcal{T}}| \left\| \mathbf{W}(\theta)\mathbf{t} - \mathbf{t} \right\|_2^2$. In many cases, using LOOCV could underestimate the generalized error. K-Fold Cross Validation (K-FCV) is then preferred over LOOCV for spatial non-structured data [188]. However, geological and ecological data often show

internal spatial or temporal dependence structures [162]. In our case, the data is spatially structured as blocks of holes. It is not suitable for our case to separate the data randomly using K-FCV. This problem is addressed by Blocking [162], where blocks (in our case holes) are considered as Folds. In Leave Hole Out Cross Validation (LHOCV), each point in a hole is interpolated using only the observed points belonging to other observed holes during inference.

LHOCV could also be computed in a single forward step by eliminating the contributions of the weights w_{ji} if $\mathbf{s}_i$ and $\mathbf{s}_j$ belong to the same hole. This means working with the weights $\widetilde{w_{ji}} = w_{ji} * a_{ji}$ where,

$$a_{ji} = I\{(x_i - x_j)^2 + (y_i - y_j)^2 \neq 0\} = \begin{cases} 1 & \text{If } (x_i - x_j)^2 + (y_i - y_j)^2 \neq 0 \\ 0 & \text{Otherwise} \end{cases},$$

is the indicator function. The weights $\widetilde{w_{ji}}$ are still differentiable with respect to θ. In the case of LHOCV, $E^{LHOCV}(\theta)$ is minimized instead of $E^{LOOCV}(\theta)$:

$$E^{LHOCV}(\theta) = \frac{1}{|\overline{\mathcal{T}}|} \left\| [\mathbf{W}(\theta) \odot \mathbf{A}]\mathbf{t} - \mathbf{t} \right\|^2 = \frac{1}{|\overline{\mathcal{T}}|} \left\| \widetilde{\mathbf{W}}(\theta)\mathbf{t} - \mathbf{t} \right\|^2, \tag{4.30}$$

where: $\mathbf{A} = [a_{ji}]$ is the matrix that contains a_{ji} as elements. With this treatment, we obtain a sparser parameter matrix $\widetilde{\mathbf{W}}$. Note that this is a form of regularization since it is very analogous to dropout-regularization in neural networks [177]. By using LHOCV, we foresee a better regularization, giving in turn more accurate predictions.

LHOCV could be applied to all previously developed IDW approaches. Algorithm 6 in Appendix A.2 shows GBM-IDW's inference using LHOCV, denoted GBM-IDW-LHOCV.

4.4 Experiments and results

In this section, we will compare the models introduced previously on different train-test split scenarios of data consisting of BI measurements and their corresponding three-dimensional coordinates extracted from an operating open pit mine. We use the minimum RMSE (see Eq.(4.1)) as a performance criterion.

All the models are implemented using Pytorch (a machine-learning framework) in python. The tests were performed on a computer with an Intel(R) Core(TM) i7-8550 CPU @ 1.80GHz processor and 8 GB of RAM. The presented models are tested on a real-world data set provided by our industrial partner from an operating open pit mine.

4.4.1 Data

We have data consisting of BI measurements and their corresponding (x, y, z) coordinates from 72 boreholes, leading to 4719 data points. Fig.4.4 is a three-dimensional plot representing the data points in the three-dimensional coordinate systems named CoordX, CoordY, and CoordZ. The color bar represents the BI values of each data point. Fig.4.5 shows a two-dimensional cross section of BI values at altitude 1240 cm.

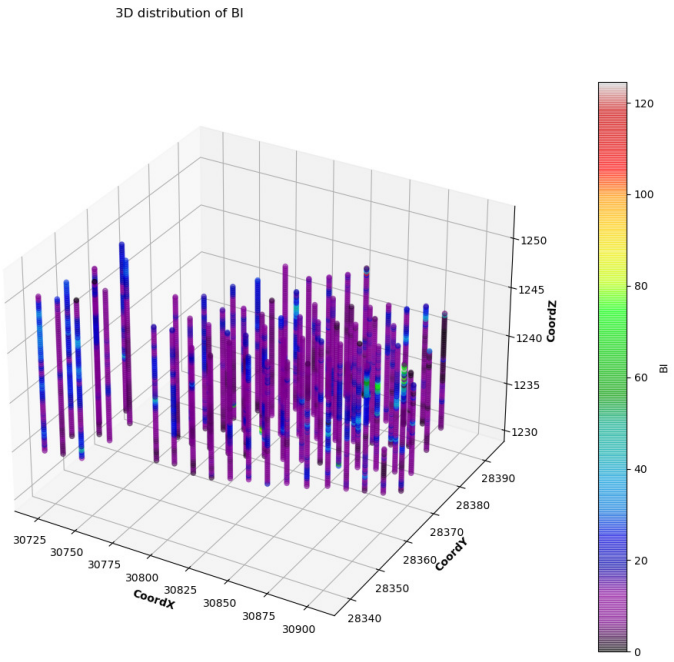


Figure 4.4 Data consisting of BI measurements of the drilled holes presented in a three-dimensional coordinates system. The intensity of the color-bar indicates the BI values.

4.4.2 Models' comparison

In this section, we present the results and the interpretations of the proposed approaches. The approaches are tested and compared using six extrapolation scenarios. A scenario consists of choosing five boreholes as testing data, and the rest are considered as training data. The circled boreholes in Fig.4.6 present the test boreholes of each scenario. The choice of the test data in each scenario to be in sequence and located in the border areas was intended to emulate the possible drilling patterns. According to experts and mining operators, the machine drills holes in series and it is convenient to consider the interior areas as drilled and the border areas as not drilled.

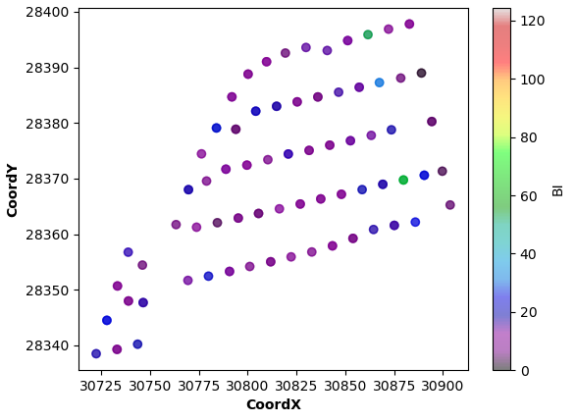


Figure 4.5 A two-dimensional cross section of BI values at altitude 1240 cm.

The minimization of the Root-Mean-Square Error (RMSE) defined in Eq.(4.1) is used as a performance criterion. We chose RMSE because it allows us to assess both bias and variance in the model's predictions simultaneously.

To compare models, an average RMSE of all scenarios is calculated for each model. Fig.6.6 shows the average RMSE_{test} of each model. The approaches that are compared are presented in the abscissa axis with the following notations:

- Mean: Using the mean of the training data as estimator.
- GP: The conventional Gaussian Process with a simple exponential kernel.
- GP-C: The conventional Gaussian Process with a simple exponential kernel and constrained hyper-parameters.
- GP-nested: The nested model with constrained hyper-parameters.
- IDW: The gradient-based conventional IDW model defined in Sect.4.3.2.
- IDW-C: The gradient-based conventional IDW model defined in Sect.4.3.2 with constrained parameters.
- IDW-LHOCV: The gradient-based conventional IDW model defined in Sect.4.3.2 with LHOCV inference.
- M-IDW-LHOCV: The gradient-based IDW-modified approach with LHOCV inference.
- GBM-IDW-LHOCV: The gradient-based IDW modified approach with learnable coefficients and LHOCV inference.

According to Fig.4.7, the Unconstrained GP (Gaussian Process) performed worse than the baseline, but it exhibited less variance in $RMSE_{test}$. This can be attributed to the Unconstrained GP being more sensitive to outliers and prone to overfitting. By introducing constraints and incorporating realistic physical ranges from variography, both the average

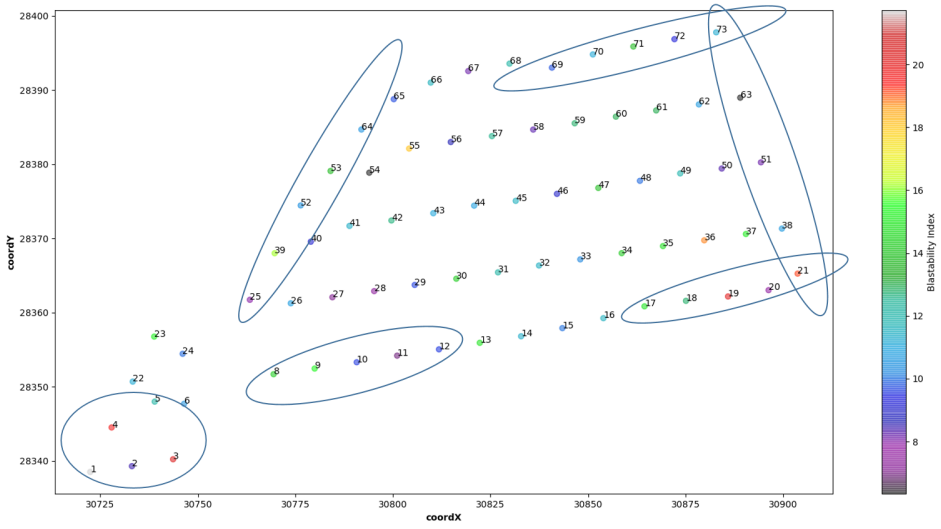


Figure 4.6 Train-test scenarios, in each scenario, the circled set of holes presents the test set and the remaining holes are the training set.

$RMSE_{test}$ and the variance in $RMSE_{test}$ decreased. A similar pattern was observed with IDW and IDW-C. The application of LHOCV inference regularization resulted in lower test errors.

Comparing GP-nested and GBM-IDW-LHOCV, both methods were quite comparable. While GP-nested had a better average $RMSE_{test}$, GBM-IDW-LHOCV showed less variance in $RMSE_{test}$. This suggests that GBM-IDW-LHOCV is more robust to outliers and exhibits less variability.

4.4.3 The effect of regularization on IDW

In Sect.4.3.2, we discussed that by adding constraints and using LHOCV, the error E^{LHOCV} estimates the test error better compared to the error E^{LOOCV} used in LOOCV. This could clearly be seen in Fig.4.8 where the more the IDW model is regularized, the closer the gap between the training error and the average test error.

LHOCV regularization has also an effect on the parameters p, ℓ_x, ℓ_y, and ℓ_z of the IDW model. Fig.4.9 shows the changes on the parameters when moving from IDW-C to IDW-LHOCV by considering LHOCV inference. And then, from IDW-LHOCV to GBM-IDW-LHOCV by introducing the modified model with learnable coefficients. Moving from IDW-C to IDW-LHOCV, ℓ_x and ℓ_y increased to improve the contribution of points belonging to nearby holes.

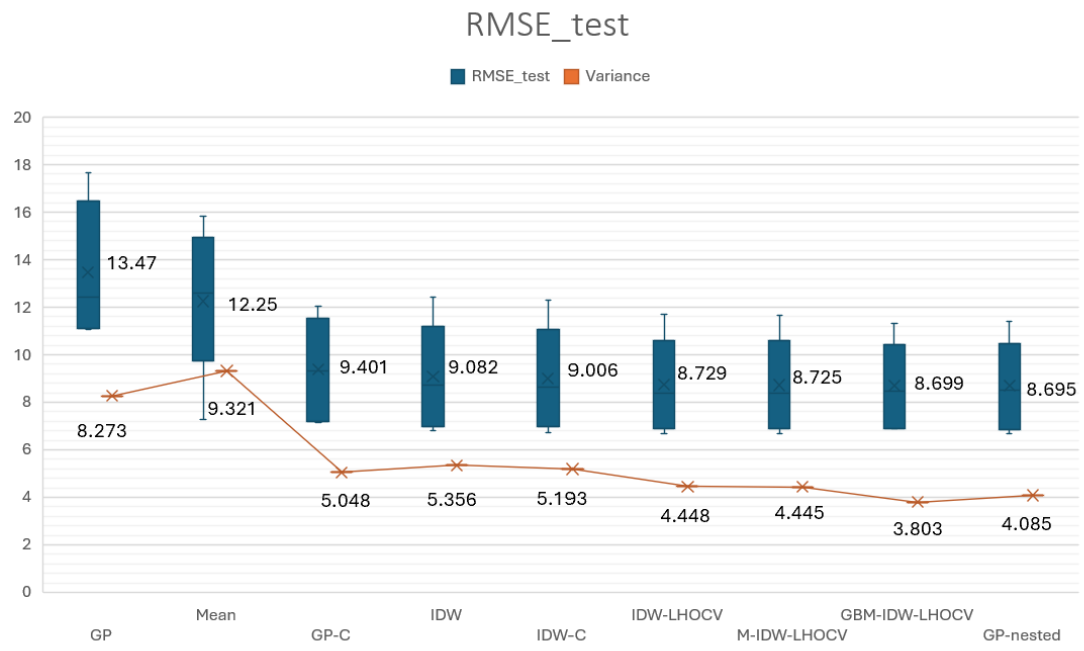


Figure 4.7 The evolution of the RMSE_{test} of all scenario from the worst performed model (higher RMSE) to the best performed model (lower RMSE) (boxplots) along with the correspondant variance (orange curve).

The vertical range ℓ_z decreased to reduce the contribution points belonging to the same holes because in real situations, they are not available for a non-drilled location to be predicted, and hence could not be used in interpolation. The parameter p is increased to limit the radius of influence to a closer boreholes, rending to a more localized and less smooth interpolation. By introducing the coefficients when moving from IDW-LHOCV to GBM-IDW-LHOCV, the range parameters ℓ_x, ℓ_y, and ℓ_z are approximately unchanged. However, as a reaction of emphasizing the contribution of some points, the power parameter p increased to make the interpolation more localized.

From this analysis, we conclude that there is a significant interaction between the anisotropic parameters and the power parameter. This contrasts with the traditional approach used in most geographic and geoscience studies [169, 189, 208], where these parameters are optimized in separate steps: anisotropy is first determined through variography, followed by the optimization of the exponent parameter using LOOCV. Our optimization scheme, however, integrates variography directly into the process, allowing for the concurrent optimization of both parameters. This differs from other methods that employ ellipsoidal search techniques, where interpolation is performed only for points within the ellipsoid, adhering to the interpo-

lation equation [128,145]. While these methods respect the geometric anisotropy of the data, they often demand more computational resources due to the complexity of the ellipsoidal search process.

In addition, our study demonstrates that integrating Leave Hole-Out Cross Validation (LHOCV) or other regularization techniques is straightforward, offering the potential to improve accuracy depending on the specific case. In summary, this framework provides a valuable, realistic, and automated solution to geospatial problems in geography and geoscience, ensuring a more holistic and efficient optimization of interpolation parameters.

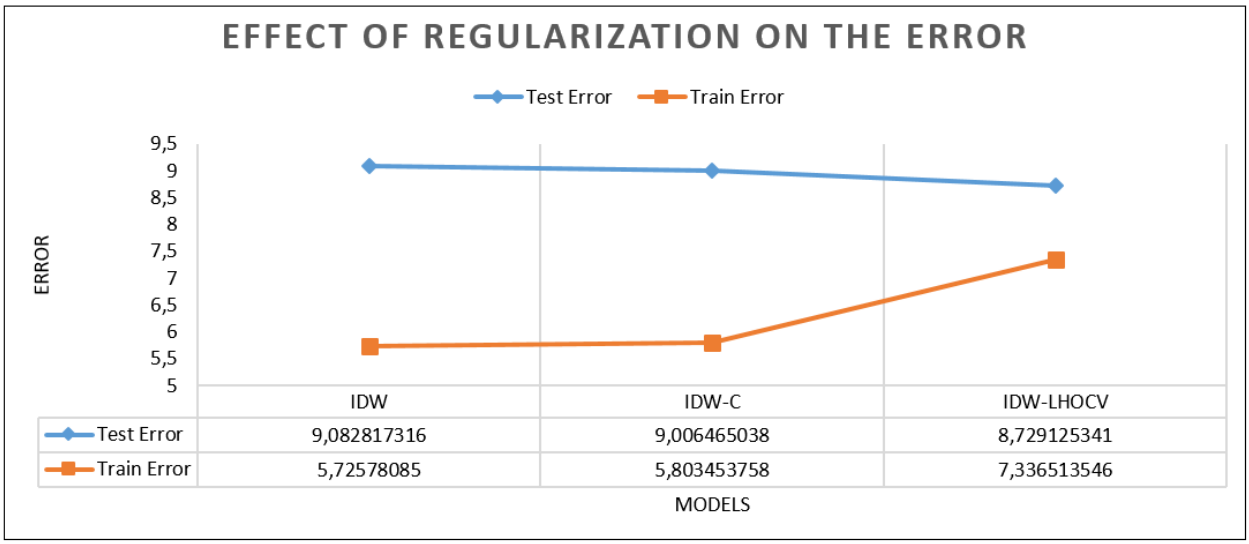
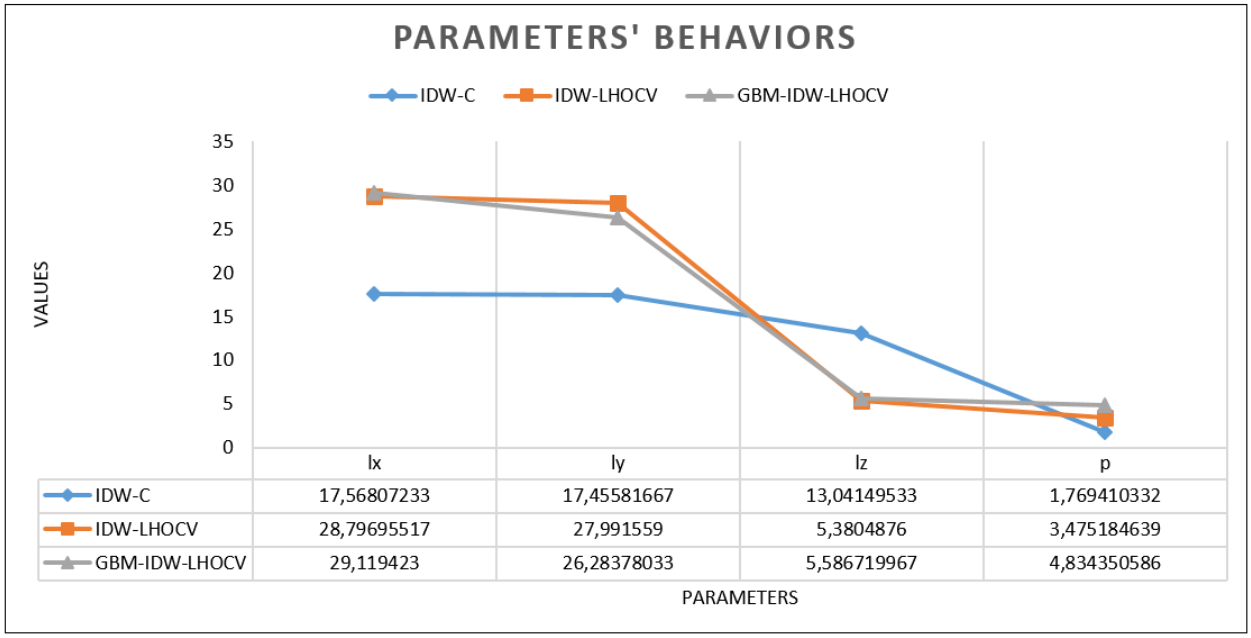


Figure 4.8 Effect of regularization on the error, the curves monitor the change in the test error (blue) and the train error (orange curve) when increasing regularization in IDW by adding constraints in IDW-C and then by adopting LHOCV inference in IDW-LHOCV.

4.4.4 Interpreting the coefficients in M-IDW-LHOCV and GBM-IDW-LHOCV

In M-IDW-LHOCV, the coefficients are calculated using Eq.(4.24). However, in GBM-IDW-LHOCV the coefficients are considered learnable parameters. In this section, we interpret these coefficients. We present in Fig.4.10 the coefficients c_i in the abscissa axis and the correspondent observed BI, t_i in the ordinate axis. According to Eq.(4.24) the coefficients of M-IDW-LHOCV (blue curve) are increasingly linear with the error (the method emphasizes high error points). The interpolation points with coefficients that are above 1.5, correspond to BI above 80 (orange line). They are associated with high error, so their contribution is increased. However, these points are considered minorities and may be outliers. GBM-IDW-LHOCV (gray curve) does not give large coefficients to these rare points. The coefficients



	lx	ly	lz	p
IDW-C	17,56807233	17,45581667	13,04149533	1,769410332
IDW-LHOCV	28,79695517	27,991559	5,3804876	3,475184639
GBM-IDW-LHOCV	29,119423	26,28378033	5,586719967	4,834350586

Figure 4.9 The infered parameters' values in IDW-C (blue curve), IDW-LHOCV (orange curve), and GBM-IDW-LHOCV (gray curve). The figure captures their changes when moving from one IDW variant to another.

values' scheme is more complex and aims to minimize global LHOCV error through gradient descent. GBM-IDW-LHOCV might be less sensitive to outliers which explains why it performed slightly better than M-IDW-LHOCV.

Fig.4.11 shows the spatial distribution of the coefficients for the GBM-IDW-LHOCV model and the corresponding BI distribution. We can visually note that the coefficients depict more variability. The role of the learnable coefficients is to make the IDW model more flexible in order to capture abrupt changes. This is achieved through the automatic optimization of the Leave-Hole-Out Cross-Validation (LHOCV) error, introducing variability as needed to cope with abrupt changes that the conventional IDW model would not handle effectively. This aspect will be seen more clearly when analyzing the three-dimensional cross-sections of the interpolated values of the proposed models.

4.4.5 Analyzing two-dimensional cross sections of an interpolated volume

To conduct a visual analysis of the spatial interpolation volume of the models, we focus on the three-dimensional cross-section at an altitude of 1240 cm with a thickness of 1 cm. Fig.4.12 displays the borehole data for this cross-section, which we consider as test data, while the remaining data serves as training data. We employ the models GP-C, IDW-C, GP-

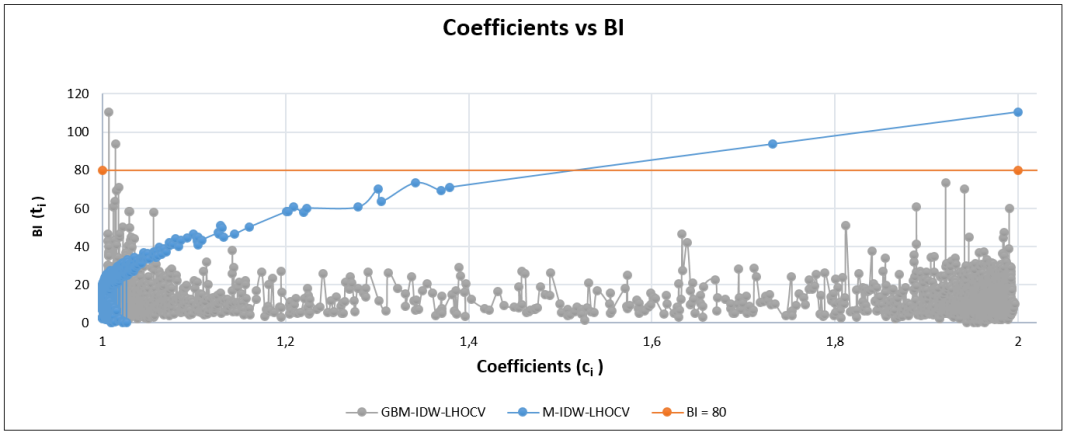


Figure 4.10 The curves present the coefficients c_i in the abscissa axis and the correspondent observed BI, t_i in the ordinate axis for M-IDW-LHOCV (blue curve) and GBM-IDW-LHOCV (gray curve). The interpolation points with coefficients that are above 1.5, correspond to BI above 80 (orange line).

nested, and GBM-IDW-LHOCV to interpolate across the volume that includes and extends beyond the existing boreholes. Fig.4.13 illustrates the two dimensional cross section of the interpolation volume for the aforementioned models.

The GP-C interpolation model shows continuous high-intensity zones characterized by BI values above 60, which do not match well with the ground truth data presented in the cross-section. GP-C is more sensitive to high BI values, which are less prevalent in the data. In some boundary regions, especially those lacking observational data, purple regions appear, indicating that the model tends towards the mean BI value of around 12.

The GP-nested model minimizes the regions of high BI and aligns more closely with the ground truth. It predicts some regions with high BI but not as extreme as GP-C. IDW-C shows more continuous purple regions (closer to the mean of the data). Although it performs better than the mean, it fails to account for some abrupt changes where higher BI values are present (blue regions). GBM-IDW-LHOCV resolves this issue. The learnable coefficients increase the flexibility of IDW-C, allowing it to handle abrupt changes from darker to lighter regions more effectively, making it closer to the ground truth. Because all the models are tested on only one single dataset we have to be cautious and not jump directly into generalizing the models' behaviors. The current results, however, indicate that the developed models are promising for spatial interpolation.

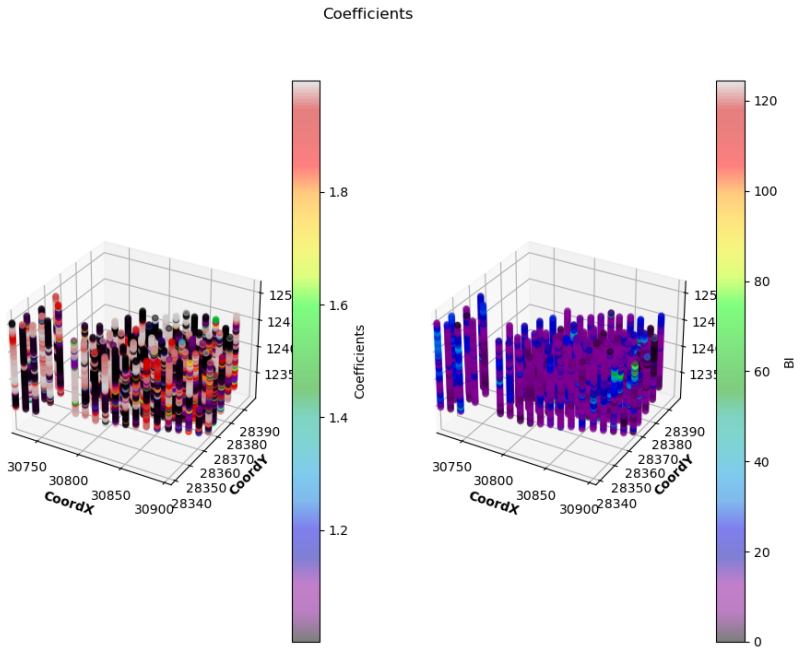


Figure 4.11 The three-dimensional spatial distribution of the coefficients of the GBM-IDW-LHOCV model (left), and the corresponding BI values (right).

4.5 Conclusion

In this paper, we used gradient-based machine learning (back-propagation) on conventional interpolation approaches. By employing parametrization techniques, we introduced constraints to account for spatial properties, such as geometric anisotropy, and improve predictions in the case of the Gaussian Process (GP) model. The GP-nested model, a more sophisticated variant, was specifically designed to account for zonal anisotropy. Additionally, we proposed a novel inference treatment for Inverse Distance Weighting (IDW), leading to the development of GBM-IDW-LHOCV, whose performance was found to be comparable to that of GP-nested. The results indicated that the Leave-One-Hole Cross Validation (LHOCV) error was closer to the average test error compared to the Leave-One-Out Cross Validation (LOOCV) error. This transition from LOOCV to LHOCV was achieved by zeroing out the weights associated with the hole to be predicted. This can be seen as a form of regularization, as it mirrors dropout regularization in neural networks [177]. By increasing horizontal ranges and decreasing vertical ranges, LHOCV regularization emphasizes interpolation in nearby boreholes while reducing the influence of points within the same hole that would not be available in an actual predictive scenario.

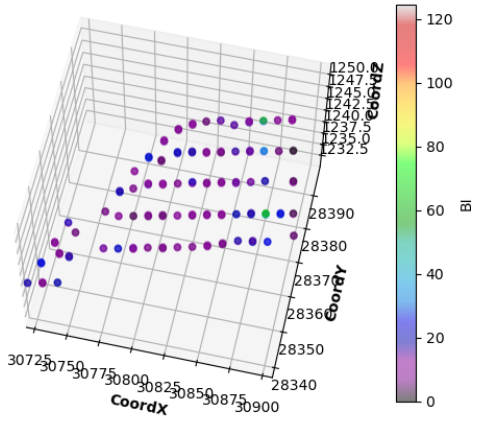


Figure 4.12 A three-dimensional cross-section at an altitude of 1240 cm with a thickness of 1 cm.

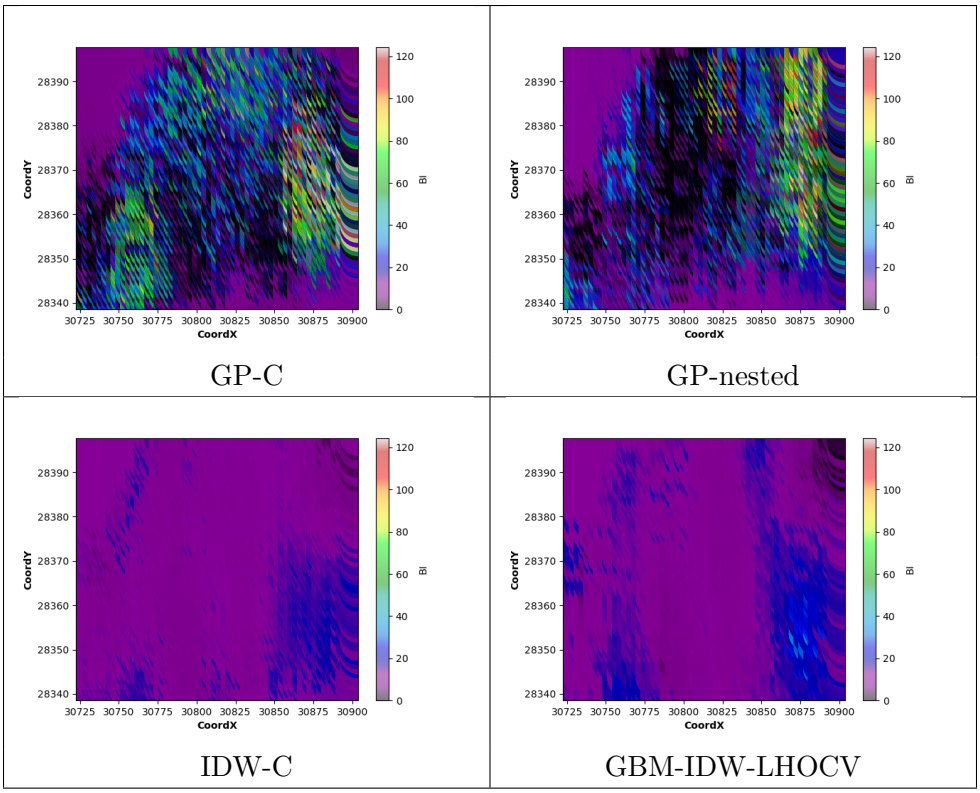


Figure 4.13 Models interpolation of a cross section at altitude 1240 cm.

Moreover, we demonstrated that GBM-IDW-LHOCV is potentially more robust to outliers compared to M-IDW-LHOCV, as it does not overly emphasize the contribution of rare or isolated interpolation points. This added robustness is particularly beneficial in geospatial data, where outliers can significantly distort predictions.

As future work, we propose extending this novel inference framework for IDW by exploring deeper architectures, analogous to deep neural networks. Since the current model, based on back-propagation, functions similarly to a single-layer neural network, it holds promise for solving more complex interpolation problems. By developing a deeper structure for our IDW variant, we could capture more intricate spatial relationships and dependencies.

Additionally, we envision stacking the IDW model with deep learning architectures such as Variational Autoencoders (VAE). This integration would allow the model to learn more complex latent structures while preserving spatial dependency, an essential factor in geoscience applications. In particular, such a framework could be applied to rock mass domaining and lithology recognition, improving both geological modeling and rock hardness predictions while maintaining the spatial continuity of the data. This would not only enhance the robustness of spatial interpolation but also provide deeper insights into the underlying geological structures of mining complexes, leading to better decision-making in mining operations.

Apport du chapitre à la thèse

Les résultats obtenus dans ce travail montrent que l'inférence basée sur la rétropropagation améliore considérablement la performance des modèles IDW et GP pour l'estimation du BI. En optimisant les paramètres du modèle via l'apprentissage automatique, cette approche surpasse les méthodes classiques en prenant en compte l'anisotropie et les structures spatiales sous-jacentes. Les modèles développés dans ce chapitre sont compatibles avec les réseaux de neurones, ouvrant ainsi la voie à l'intégration de techniques de réduction de dimension basées sur ces réseaux. Cela permet de concevoir des modèles mieux adaptés aux tâches spécifiques, telles que la délimitation des masses rocheuses, comme exploré dans les chapitres suivants.

CHAPITRE 5 ARTICLE 2: INTEGRATING VARIATIONAL AUTO-ENCODERS (VAEs) AND SPATIAL INTERPOLATION FOR IMPROVING ROCK MASS DOMAINING IN OPEN PIT MINES

Cet article a été soumis le 06 avril 2025 à la revue *Computers & Geosciences* et est en attente de révisions, coécrit avec Richard Labib, Jean-François Plante et Michel Gamache.

Aperçu du chapitre

L'identification des massifs rocheux dans les mines à ciel ouvert repose sur la capacité à regrouper des zones homogènes en termes de dureté et de comportement mécanique. Ce chapitre propose une méthode combinant les autoencodeurs variationnels (VAE) et l'interpolation spatiale pour améliorer la délimitation des domaines géologiques à partir des données MWD. L'objectif est d'extraire une représentation latente pertinente des MWD qui capture à la fois la dureté des roches et la continuité spatiale, permettant ainsi une classification plus robuste des unités géologiques.

5.1 Abstract

This study focuses on rock mass domaining, using measurement while drilling (MWD) data to create contiguous spatial domains characterized by consistent rock hardness. The proposed method employs a stacking variational encoder (VAE) combined with a spatial interpolation model compatible with neural networks. The VAE is designed to learn the underlying shared structure of the MWD data, specifically targeting rock hardness, while the spatial interpolator captures spatial similarity to ensure consistency. The resulting feature, termed Pseudo-BI, serves as the primary input for clustering via BIRCH algorithm (Balanced Iterative Reducing and Clustering Using Hierarchies). Three different VAE layer types——fully connected (FC), convolutional (Conv), and radial basis function (RBF)—are explored. The models' performances are evaluated by assessing contiguity using Domain Accuracy (DA) and rock hardness consistency through Pooled Standard Deviation (PSD). Beyond achieving spatial domaining, the models successfully detect true lithologies identified by experts in 2D and extend these findings into 3D, demonstrating their potential as powerful geomodeling tools for mining applications.

5.2 Introduction

Bench rock classification poses a significant challenge, requiring the identification of spatial zones with homogeneous rock properties for effective mining operations. To achieve a targeted fragmentation size distribution, it is essential to apply similar explosive energy to zones with comparable rock hardness. Rock mass domaining, which involves delineating contiguous spatial zones with similar rock hardness, is crucial for optimizing explosive recipes and subsequent operations such as grinding.

Rock hardness, quantified by the Blastability Index (BI), is derived by normalizing the Specific Energy of Drilling (SED). This process relies on mechanical parameters recorded by Monitoring While Drilling (MWD) sensors, which are attached to the drilling machines. These parameters include the Rate of Penetration (ROP), which measures the speed at which the drill bit advances through the rock (typically in meters per minute), the Weight on Bit (WOB), representing the axial force applied to the drill bit (measured in kilonewtons), the Rotations Per Minute (RPM) of the drill bit, and the Torque (TRQ), which quantifies the rotational force exerted on the drill string (measured in kilonewton-meters). The SED is computed using the following equation:

$$\mathrm{SED} = \frac{\mathrm{WOB}}{A} + \frac{2\pi \times \mathrm{RPM} \times \mathrm{TRQ}}{A \times \mathrm{ROP}}, \tag{5.1}$$

where A represents the cross-sectional area of the wellbore [11, 225]. The BI is then obtained by normalizing the SED to account for variations in drilling conditions and machine settings, ensuring a more consistent representation of rock hardness.

Traditional geological surveys, while informative on surface rock composition, are costly and time consuming [224]. These surveys lack the ability to provide comprehensive 3D rock recognition along the wells, limiting insight into rock characteristics. To overcome these limitations, recent studies propose employing Machine Learning (ML) techniques to relate rock properties to drilling measurements [175]. The existing literature categorizes these studies into two main approaches: supervised and unsupervised methods [175, 224].

Supervised methods rely on labeled data, where each MWD data point is assigned a corresponding rock class label. Several ML approaches have been explored, including Neural Networks [50, 62, 82, 98, 99, 125, 130, 192, 214], Boosting [62, 82, 214], Gaussian Processes [176, 223, 224, 226], Fuzzy Logic [82], Decision Trees [62, 96, 214], and Conditional Random Fields (CRF) [132, 133]. Challenges arise due to labels availability and the complexity of predicting rock types from MWD, resulting from the lack of a one-to-one correspondence between lithology and MWD measurements. Addressing these challenges, researchers propose

innovative solutions such as mapping MWD to the Adjusted Penetration Rate (APR) as a rock hardness measurement [224]. Entropy analysis, as proposed by [225], claims to eliminate inherent variability in ROP measurements.

Unsupervised learning approaches are introduced to analyze unlabeled data, primarily through clustering algorithms. However, their application to drilling measurements remains limited, with most studies employing unsupervised methods for feature learning as a precursor to classification [93, 225]. Spatial dependency is often overlooked, with notable exceptions being the work of [132–134], where boosted decision trees and Conditional Random Fields (CRF) are used to capture spatial relationships among neighboring measurements. Nevertheless, these approaches do not account for anisotropy, particularly geometrical anisotropy, which refers to the directional variation in spatial continuity, where the range of spatial correlation differs depending on direction. Ignoring anisotropy in spatial models can lead to inaccurate geological representations, particularly in structurally complex formations where rock properties exhibit directional dependencies. Despite its critical role in geostatistical modeling, anisotropy remains largely unexplored in machine learning applications to drilling data, highlighting a key gap in the current research landscape.

Nowadays, the baseline clustering relies on BI (similar to SED) as a measure that characterizes rock hardness of the zones. However, BI inherently captures the variability in mechanical parameters arising from measurement errors, vibration, bit wear, and operator and/or auto-drill setpoints. Its clusters, as we will show, lack provision for spatial domains. Addressing these challenges, our research focuses on developing a machine learning model leveraging Variational Auto-Encoders (VAEs) and spatial interpolation techniques. VAEs are employed to learn an underlying structure namely pseudo-BI from MWD parameters; it is considered as an indication of the rock hardness as a characteristic shared by all four mechanical parameters. Spatial interpolation techniques are then applied to learn spatial similarity and anisotropy, ensuring contiguity in space and mitigating observation variabilities. This integrated approach yields a 3D spatial mapping of pseudo-BI for every point in the minefield.

The spatial contiguity and similarity concerning geology are guaranteed by the spatial interpolator. Moreover, this methodology addresses data scarcity by using the same interpolator to estimate pseudo-BI in spatial locations lacking observational data. The final step involves clustering based on the pseudo-BI for the augmented data, defining contiguous spatial domains with consistent rock hardness. This integrated approach promises to significantly improve the efficiency of mining operations, marking a substantial step forward in the field of bench rock classification.

The following sections provide a detailed overview of the models and methodologies, including experimental procedures and visual illustrations. Section 2 covers the theoretical framework and methodology, Section 3 presents the experiments and results, and Section 4 discusses the conclusions and findings.

5.3 Theory and methodology

5.3.1 Auto-Enocders (AEs)

AEs represent a specific type of neural network (NN) designed to encode input into a meaningful latent representation with lower dimension and then decode it back to ensure that the reconstructed input closely resembles the original input [16,17]. Numerous variations of AEs, such as VAE [95], Correlational Neural Networks (CorrNets) [32], and Variational Canonical Correlation Analysis (VCCA) [199], have been developed to address challenges posed by conventional dimensionality reduction techniques, including Principal Component Analysis (PCA), Probabilistic Principal Component Analysis (PPCA) [184], Canonical Correlation Analysis (CCA) and Factor Analysis (FA). These variations cater to specific tasks, whether it is performing dimensionality reduction, learning a probabilistic latent representation, or finding a representation that maximizes the correlation of two random vectors.

In the domain of mining, geology and geoscience, recent studies have explored VAEs for geochemical pattern recognition [202], digital rock reconstruction [36], rock burst assessment [113]. A geologically constrained VAE showed promise in mineral prospectivity mapping [230] and recognizing geochemical anomalies [221]. VAE exhibited fair modeling flexibility, accommodating physical and geological constraints. Most studies introduced a geology-related loss, which is added to the VAE loss [221, 230]. In our study, we consider extending this approach by stacking a gradient-based spatial interpolator onto the VAE to account for spatial dependency, anisotropy, and spatial smoothness, thereby enhancing spatial contiguity.

Since VAEs serve as the foundational components of our proposed models, this section provides a comprehensive explanation of VAEs. Let $\mathbf{t}_i$, for $i = 1, \ldots, N$, denote the input data, where each vector $\mathbf{t}_i \in \mathbb{R}^d$ represents a single observation of the MWD data. Similar to AEs, VAEs encode each input vector, denoted by $\mathbf{t}$, into a latent representation $z_{\mathbf{t}}$ through an encoder network. This latent variable is then decoded back to $\hat{\mathbf{t}}$, a reconstruction of the original input, through a decoder network. What differentiates VAEs from AEs is the probabilistic framework they employ, particularly variational Bayesian inference. Rather than directly learning a deterministic mapping from input to latent variables, VAEs assume that the latent space follows a probabilistic distribution.

In VAEs, the model learns a variational approximation to the intractable Bayesian posterior over the latent variables. Specifically, it models a variational normal distribution $q(z_{\mathbf{t}}, \theta_{\text{enc}}) = \mathcal{N}(z_{\mathbf{t}}|\mu_{\mathbf{t}}, \sigma_{\mathbf{t}}^2)$ as an approximation to the true posterior distribution $p(z_{\mathbf{t}}|\mathbf{t})$, which is generally intractable. Here, θ_{enc} represents the parameters of the encoder network, which learns a probabilistic mapping from the input $\mathbf{t}$ to the mean $\mu_{\mathbf{t}} = f^{\mu}_{\theta_{\text{enc}}}(\mathbf{t})$ and the standard deviation $\sigma_{\mathbf{t}} = f^{\sigma}_{\theta_{\text{enc}}}(\mathbf{t})$ of the approximate posterior. The decoder learns the conditional normal likelihood $p(\mathbf{t}|z_{\mathbf{t}}, \theta) = \mathcal{N}(\mathbf{t}|\hat{\mathbf{t}}, \sigma_{\mathbf{t}}^2\mathbf{I})$, where $\hat{\mathbf{t}} = f_{\theta_{dec}}(z_{\mathbf{t}})$ is the reconstructed input. This conditional likelihood assumes that the reconstruction of the input $\mathbf{t}$ is normally distributed around $\hat{\mathbf{t}}$ with variance $\sigma_{\mathbf{t}}^2$.

The parameters of both the encoder and decoder networks are learned by minimizing the VAE loss function, which is derived from the Evidence Lower Bound (ELBO). To enable efficient training, VAEs use the reparameterization trick to sample the latent variable $z_{\mathbf{t}}$ while maintaining differentiability. Instead of sampling directly from $\mathcal{N}(\mu_{\mathbf{t}}, \sigma_{\mathbf{t}}^2)$, we express it as $z_{\mathbf{t}} = \mu_{\mathbf{t}} + \sigma_{\mathbf{t}} * \epsilon, \quad \text{where } \epsilon \sim \mathcal{N}(0, 1)$.

The ELBO provides a tractable lower bound on the log-likelihood of the observed data $p(\mathbf{t}_1, \ldots, \mathbf{t}_N)$, since computing the exact log-likelihood is computationally intractable. The ELBO consists of two terms: the first term is the reconstruction loss, which encourages the decoder to produce accurate reconstructions of the input; the second term is the KL divergence, which regularizes the encoder by minimizing the difference between the approximate posterior $q(z_{\mathbf{t}} \mid \mathbf{t})$ and the prior $p(z_{\mathbf{t}})$, usually a standard normal distribution. Here, the Kullback–Leibler (KL) divergence is $D_{\text{KL}}(q\|p) = \mathbb{E}_q\Big[\log \frac{q(z_t)}{p(z_t)}\Big]$, a non-negative measure of dissimilarity between distributions that equals zero if and only if $q = p$. This regularization helps ensure that the learned latent space is well-structured. Mathematically, the VAE loss function can be written as:

$$L_{\text{VAE}} = -\sum_{i=1}^{N} \left(\frac{1}{2} \left(\ln \left((\sigma_{\mathbf{t}_i})^2 \right) - (\mu_{\mathbf{t}_i})^2 - (\sigma_{\mathbf{t}_i})^2 + 1 \right) + \mathbb{E}_{q(z_{\mathbf{t}_i}, \theta_{\text{enc}})} \left[\ln p(\mathbf{t}_i|z_{\mathbf{t}_i}, \theta) \right] \right) \tag{5.2}$$

In this expression, the first part of the loss corresponds to the KL divergence between the approximate posterior $q(z_{\mathbf{t}}|\mathbf{t})$ and the prior $p(z_{\mathbf{t}})$, and the second part corresponds to the expected log-likelihood of the data given the latent variables. By minimizing this loss, the VAE learns both an efficient latent representation and a generative model that can produce realistic data reconstructions, while also ensuring that the latent space adheres to the chosen prior distribution.

Figure 5.1 shows the basic architecture of a VAE, in which the four mechanical parameters ROP, RPM, TRQ, and WOB are regarded as inputs that are encoded into one-dimensional

latent representation and then decoded back into pseudo-ROP, pseudo-WOB, pseudo-RPM, and pseudo-TRQ.

.

5.3.2 Proposed methodology

Stacking gradient-based spatial interpolator and VAE

In this study, we address a dataset containing MWD data at various three-dimensional coordinates. Each measurement, denoted as $i = 1, ..., N$, is represented by a four-dimensional vector $\mathbf{t}_i$, encapsulating ROP, WOB, RPM, and TRQ, recorded at the specific location $\mathbf{s}_i = [x_i, y_i, z_i]^T$ within a borehole situated at coordinates (x_i, y_i) and depth z_i. The set of of measurements for a hole j $(j = 1, ..., M)$ is denoted as H_j. We employ a VAE model to encode each input feature vector $\mathbf{t}_i$ into a one-dimensional latent representation $z_{\mathbf{t}}^{(i)}$. To mitigate variability, account for spatial dependency, and enhance spatial contiguity, we integrate a spatial interpolator and compute the interpolation $\hat{z}_{\mathbf{t}}^{(i)}$ using observations of $z_{\mathbf{t}}$ other than the datum i. In previous work, we developed a gradient-based spatial interpolator based on the Modified Inverse Distance Weighting (IDW) model, referred to as the Gradient-Based Modified Inverse Distance Weighting (GBM-IDW) [63]. This method integrates backpropagation-based optimization into the conventional IDW framework, making it compatible with neural networks. For validation, we employed Leave Hole Out Cross Validation (LHOCV), which differs from the standard Leave One Out Cross Validation (LOOCV). In LHOCV, the interpolation at a given location $\mathbf{s}_i$ within the kth borehole $(i \in H_k)$ is performed using observations from all boreholes except the kth one. This approach accounts for spatial dependency between boreholes and has been shown to improve interpolation accuracy compared to LOOCV, which excludes only the individual point being interpolated from the training data. In our context, employing such interpolation to compute $\hat{z}_{\mathbf{t}}^{(i)}$ would increase smoothness, spatial contiguity, and address data scarcity, as it enables the estimation of $\hat{z}_{\mathbf{t}}$ be-

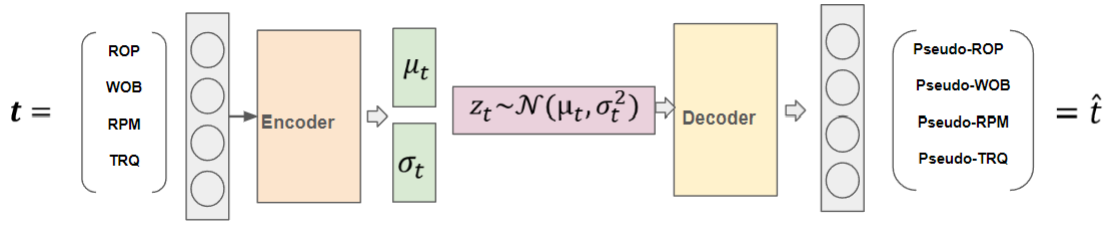


Figure 5.1 VAE architecture

tween boreholes where MWD data are unavailable. Subsequently, the GBM-IDW model with LHOCV interpolation is utilized as the spatial interpolator to compute $\hat{\mathbf{z}}_\mathbf{t} = [\hat{z}_\mathbf{t}^{(1)}, ..., \hat{z}_\mathbf{t}^{(N)}]^T$ using the inputs $\mathbf{z}_\mathbf{t} = [z_\mathbf{t}^{(1)}, ..., z_\mathbf{t}^{(N)}]^T$:

$$\hat{\mathbf{z}}_\mathbf{t} = \frac{[\mathbf{W}(\rho) \odot \mathbf{A}][\mathbf{c} \odot \mathbf{z}_\mathbf{t}]}{\mathbf{W}(\rho)[\mathbf{c} \odot \mathbf{1}]} \tag{5.3}$$

Here, $\mathbf{W}$ is an $N \times N$ matrix that has zeros on its diagonal, and IDW's weights $w_{ji} = d_a(\mathbf{s}_i, \mathbf{s}_j)^{-p} / \sum_i d_a(\mathbf{s}_i, \mathbf{s}_j)^{-p}$ elsewhere.

The Euclidean distance $d_a(\mathbf{s}_i, \mathbf{s}_j) = \sqrt{(x_i - x_j)^2/\ell_x^2 + (y_i - y_j)^2/\ell_y^2 + (z_i - z_j)^2/\ell_z^2}$ accounts for anisotropy by incorporating ℓ_x as the range, or length-scale, of the x principal direction, and similarly for ℓ_y and ℓ_z in the principal direction of y and z respectively. The set $\rho = \{p, \ell_x, \ell_y, \ell_z\}$ comprises the learnable parameters of the IDW weights, and the vector $\mathbf{c} = [c_1, ..., c_N]$ contains learnable coefficients constrained to the domain $[1, 2]^N$ to maintain model stability and prevent overfitting, ensuring reasonable adjustment of the spatial data without excessive sensitivity to outliers [63]. To perform LHOCV, the matrix $\mathbf{A} = [a_{ji}]$ is used to eliminate the contribution of points belonging to the same borehole in the interpolation (a_{ji} is set to zero if there is a hole k such that $i, j \in H_k$, it is set to one otherwise). The element-wise product is denoted by $\odot$ (e.g., $\mathbf{c} \odot \mathbf{t} = [c_i t_i]$), and vector division is performed element-wise (e.g., $\mathbf{c}/\mathbf{t} = [c_i/t_i]$) throughout this study. Using the interpolation equation, each $\hat{z}_\mathbf{t}^{(i)}$ is computed as follows:

$$\hat{z}_\mathbf{t}^{(i)} = \sum_{j \neq i} a_{ji} w_{ji} c_j z_\mathbf{t}^{(j)} \tag{5.4}$$

The measurements $\hat{\mathbf{z}}_\mathbf{t} = [\hat{z}_\mathbf{t}^{(1)}, ..., \hat{z}_\mathbf{t}^{(N)}]^T$ are referred to as pseudo-BI measurements, representing the structure aimed to be learned. They are decoded back using a VAE decoder to reconstruct the mechanical parameters, denoted as pseudo-ROP, pseudo-WOB, pseudo-RPM, and pseudo-TRQ, represented by the vector $\hat{\mathbf{t}}$. Figure 5.2 shows the VAE with the stacked GBM-IDW-LHOCV spatial interpolator.

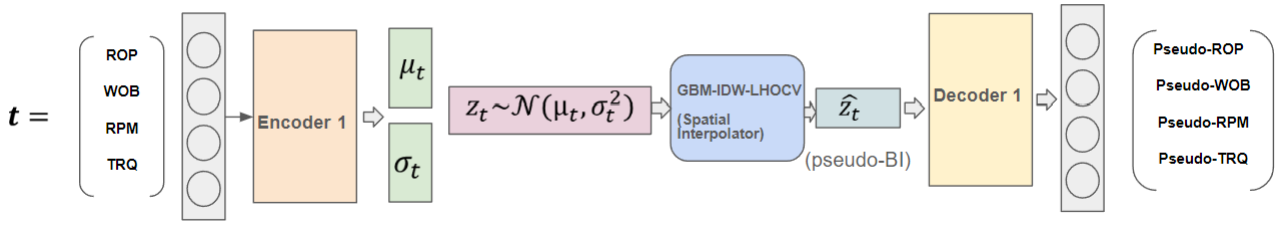


Figure 5.2 VAE with stacked GBM-IDW-LHOCV architecture

Add spatial information using VAE and correlation loss

To incorporate spatial feature information into pseudo-BI measurements, we introduce a second VAE that encodes the input spatial features, represented by $\mathbf{s}$, into a one-dimensional latent representation $z_{\mathbf{s}}$. To ensure consistency between spatial and mechanical representations, we constrain both VAEs to maximize the sample correlation coefficient between the interpolated measurements $\hat{\mathbf{z}}_{\mathbf{t}} = [\hat{z}_{\mathbf{t}}^{(1)}, \ldots, \hat{z}_{\mathbf{t}}^{(N)}]^T$ and the spatial latent variables $\mathbf{z}_{\mathbf{s}} = [z_{\mathbf{s}}^{(1)}, \ldots, z_{\mathbf{s}}^{(N)}]^T$. This is achieved by introducing a correlation loss term, L_{corr}, to the VAE objective function, defined as:

$$L_{\text{corr}} = 1 - |\text{Corr}(\hat{\mathbf{z}}_{\mathbf{t}}, \mathbf{z}_{\mathbf{s}})|. \tag{5.5}$$

The correlation loss term would enhance the spatial-awareness in the learned representations, aligning with techniques used in [199, 221, 230].

The general loss function L is the sum of the first VAE loss, $L_{\text{VAE}}^{(1)}$, the second VAE loss, $L_{\text{VAE}}^{(2)}$, and the correlation loss, L_{corr}. For VAE losses, we introduce a tunable parameter β to control the trade-off between reconstruction accuracy and latent space regularization, improving the efficiency of the VAE. This approach is known as the Beta-VAE variant [72]:

$$L_{\text{VAE}}^{(1)} = -\sum_{i=1}^{N} \left(\frac{\beta}{2} \left(1 + \ln((\sigma_{\mathbf{t}_i})^2) - (\mu_{\mathbf{t}_i})^2 - (\sigma_{\mathbf{t}_i})^2 \right) + \mathbb{E}_{q(z_{\mathbf{t}}^{(i)}, \theta_{\text{enc}}^{(1)})} \left(\ln(p(\mathbf{t}_i | z_{\mathbf{t}}^{(i)}, \theta^{(1)})) \right) \right), \tag{5.6}$$

$$L_{\text{VAE}}^{(2)} = -\sum_{i=1}^{N} \left(\frac{\beta}{2} \left(1 + \ln((\sigma_{\mathbf{s}_i})^2) - (\mu_{\mathbf{s}_i})^2 - (\sigma_{\mathbf{s}_i})^2 \right) + \mathbb{E}_{q(z_{\mathbf{s}}^{(i)}, \theta_{\text{enc}}^{(2)})} \left(\ln(p(\mathbf{s}_i | z_{\mathbf{s}}^{(i)}, \theta^{(2)})) \right) \right), \tag{5.7}$$

$$L = \alpha_1 L_{\text{VAE}}^{(1)} + \alpha_2 L_{\text{VAE}}^{(2)} + \alpha_3 L_{\text{corr}}, \tag{5.8}$$

Here, α_1, α_2, and α_3 are hyperparameters that can be tuned to optimize training performance. In particular, with this approach, the general loss is differentiable with respect to the parameters of the first VAE, $\theta^{(1)}$, the parameters of the second VAE, $\theta^{(2)}$, as well as the parameters of the GBM-IDW-LHOCV model, ρ. Consequently, we achieved an end-to-end model, as depicted in Figure 5.3, in which all parameters are trained simultaneously using backpropagation. In this figure, Rec-CoordX, Rec-CoordY, and Rec-CoordZ represent the reconstructed spatial coordinates.

This approach provides a comprehensive framework for integrating spatial information into the pseudo-BI measurements, enabling efficient training and optimization of the VAEs and the GBM-IDW-LHOOV model.

Investigating three types of NN layers for VAEs

In this study, three types of neural network architectures were investigated for modeling VAEs: Fully Connected (FC) networks, Convolutional (Conv) networks, and Radial Basis Function (RBF) networks.

The FC networks establish connections between every neuron in a layer, resulting in a dense connection matrix. This architecture allows for non-local interactions, enabling each neuron to influence every other neuron in the subsequent layer. While this provides a high representational capacity that can be advantageous in capturing complex representations, it can also lead to overfitting in some cases [59].

In contrast, Conv networks are characterized by parameter sharing and local connectivity, where the same weights are applied across different regions of the input, and each neuron is connected only to a local subset of the input space. This structure has been particularly effective in image processing due to the strong correlation among neighboring pixels [102]. Compared to FC networks, which learn independent weights for each connection, Conv networks impose structured constraints that can improve generalization when meaningful local dependencies exist. In the context of MWD data, the correlation structure arises primarily from the four mechanical parameters (WOB, RPM, TRQ, and ROP). While spatial adjacency is explicitly modeled using a spatial interpolator, this study examines whether Conv networks can inherently detect and leverage the correlation structure among the mechanical parameters.

Lastly, RBF networks employ a set of radial basis functions, each associated with a center and width parameter. This architecture can be viewed as a form of kernel-based learning that inherently considers the similarity and correlation of data points. Inputs are mapped to a high-dimensional feature space to enhance separability [24], making RBF networks particularly well-suited for clustering tasks. In this study, to determine the most suitable basis function, we investigate three different types: Gaussian RBF, defined as $\phi(r) = e^{-r^2}$; inverse quadratic RBF, given by $\phi(r) = 1/(1 + r^2)$; and thin plate spline RBF, expressed as $\phi(r) = r^2 \ln(r + 1)$. Here, r represents the Euclidean distance from the center of the basis function.

These distinct characteristics highlight the potential applications of each network type in the

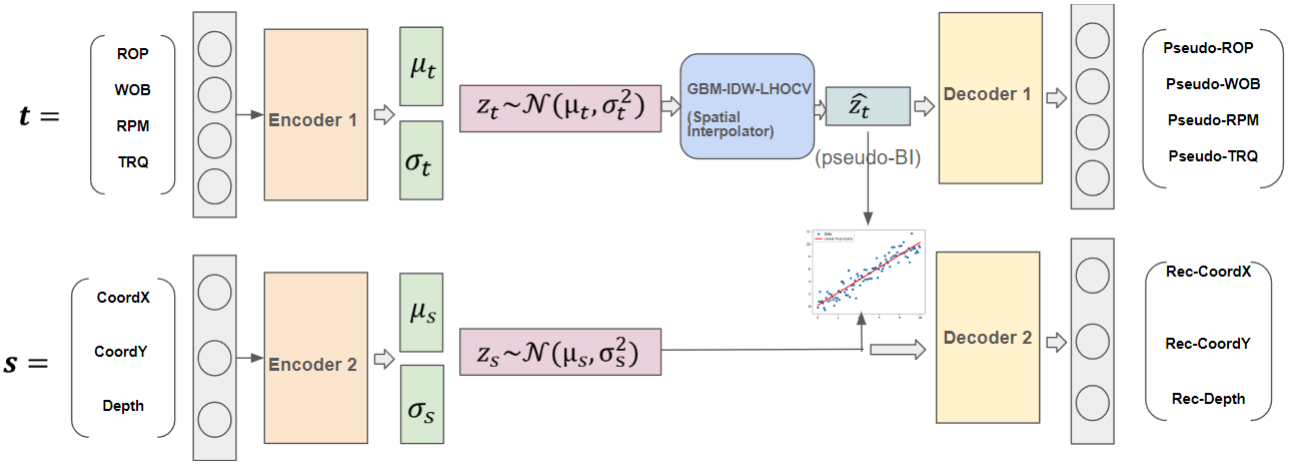


Figure 5.3 Adding spatial information by encoding spatial coordinates and adding a correlation constraint

context of VAE modeling, providing insights into their ability to capture and represent rock hardness from MWD data.

Clustering using pseudo-BI

After training the models, pseudo-BI measurements $\hat{\mathbf{z}}_\mathbf{t} = [\hat{z}_\mathbf{t}^{(1)}, \ldots, \hat{z}_\mathbf{t}^{(N)}]^T$ are used for clustering. In this study, we leverage the BIRCH algorithm (Balanced Iterative Reducing and Clustering Using Hierarchies) as an initial clustering step to generate compact representations of the data. Specifically, BIRCH incrementally constructs a hierarchical clustering feature (CF) tree, which summarizes the dataset into a set of potential cluster centers based on density and distribution properties. These centers serve as refined initialization points, improving the stability and efficiency of the subsequent clustering step. Following this, agglomerative clustering is applied to further refine the clusters. By starting with well-separated cluster centers provided by BIRCH, the agglomerative approach ensures better-defined spatial domains while preserving geological consistency. This two-step clustering strategy enhances both computational efficiency and cluster stability, making it particularly suitable for structuring the pseudo-BI feature space [218].

When clustering is applied solely to observed boreholes, there is a geometric restriction on the cluster centers, confining them to these specific holes. To mitigate this limitation, data augmentation is employed using the trained spatial interpolator to estimate pseudo-BI between the existing ones. This is achieved by sampling the locations $\mathbf{s}^*$ from a normal distribution $\mathcal{N}(\mathbf{s}^*|\mathbf{S}, \sigma^2\mathbf{I})$, where $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_N]^T$ represents the locations of the wells drilled and $\sigma^2\mathbf{I}$ is the covariance matrix. The estimated pseudo-BI values, $\hat{z}_\mathbf{t}^{(*)}$, are determined using

Equation 3. Using augmented pseudo-BI measurements would enhance spatial coverage and ensure that cluster centers are not restricted solely to observed boreholes.

5.4 Experiments and results

5.4.1 Data and the studied mining complex

A drilling pattern refers to a specific arrangement or sequence of holes within a designated area of the mining site. These holes are drilled in a predetermined sequence and layout, with the intention of loading them with explosives for a synchronized blast. The goal is to break the rock efficiently and control fragmentation, ensuring that the material is suitable for subsequent processing. In this study, we investigated the MWD data for 40 such drilling patterns, each consisting of a series of holes planned to be blasted together as part of the mining operation. The mining complex where these patterns are implemented is managed by an iron mine site in northern Quebec, and the data is provided by our industrial partner, Peck Tech Consulting.

Throughout the experimentation section, we will visually illustrate the results for a single pattern, denoted (P1), to demonstrate the underlying methodology, and we will present the appropriate key performance indicators for all 40 patterns. The architecture and parameter count of the FC, Conv, and RBF models used in these experiments are presented in Table 5.1.

Table 5.1 Models parameters

Block / model	FC (# parameters)	Conv (# parameters)	RBF (# parameters)
Encoder 1	input layer (500)	input layer (72)	input layer (20)
	hidden layers (10100)	hidden layers (272)	hidden layers (40)
	mean layer (101)	mean layer (49)	mean layer (5)
	variance layer (101)	variance layer (49)	variance layer (5)
Decoder 1	hidden layers (200)	hidden layers (168)	hidden layers(28)
	output layer (404)	output layer (56)	output layer (56)
Encoder 2	input layer (400)	input layer (42)	input layer (12)
	hidden layers (10100)	hidden layers (156)	hidden layers (24)
	mean layer (303)	mean layer (37)	mean layer (4)
	variance layer (303)	variance layer (37)	variance layer (4)
Decoder 2	hidden layers (400)	hidden layers (102)	hidden layers (18)
	output layer (303)	output layer (33)	output layer (24)
Spatial interpolator	60	60	60
Total parameters	23275	1133	284

5.4.2 Analyzing the Relation Between Pseudo-BI, Pseudo-parameters and Original Mechanical Parameters

In this section, we aim to demonstrate that the mechanical parameters reconstructed from the pseudo-BI—denoted as pseudo-ROP, pseudo-WOB, pseudo-RPM, and pseudo-TRQ—capture the general trends of the mechanical parameters while avoiding variability and outliers. The spatial interpolator played an important role in capturing spatial similarities and replacing extreme points and outliers with reasonable values that follow the general trend of the data. This can be seen in Figure 5.4, which shows the relationships between pseudo-BI with pseudo-ROP, pseudo-WOB, and pseudo-RPM (orange scatter) and the original mechanical parameters ROP, WOB, RPM, and TRQ (represented by rainbow scatters) using the model with FC layers in the pattern (P1). In this figure, the more intense and hot colors indicate a higher concentration of data points.

The models with Conv and RBF layers also succeeded in reconstructing pseudo-parameters from the learned latent representations that are smoother than the original mechanical parameters and follow their general trend. Figures 5.5 and 5.6 show the relationships using Conv layers and RBF layers with Gaussian basis functions, respectively, for the same pattern (P1).

Since the models are able to reconstruct the general behavior of the four mechanical parameters using only information about pseudo-BI, we can conclude that this latent structure represents a shared characteristic of the four mechanical parameters. This shared characteristic is presumably correlated with the hardness of the rock as a common physical property.

To obtain real physical and interpretable values that can be used in the industry, we computed the Reconstructed Blastability Index (Rec-BI) as a normalization of the Reconstructed Specific Energy of Drilling (Rec-SED) using the SED formula and replacing the mechanical parameters with the pseudo-parameters:

$$\text{Rec-SED} = \frac{\text{pseudo-WOB}}{A} + \frac{2\pi \times \text{pseudo-RPM} \times \text{pseudo-TRQ}}{A \times \text{pseudo-ROP}}, \tag{5.9}$$

Figure 5.7 shows Rec-BI and BI as functions of pseudo-BI for the model with FC layers, applied to the pattern (P1). We could see that Rec-BI presents the same physical units as BI, and its scatter (represented in orange) follows the general trend of the BI scatter plot in rainbow colors. However, Rec-BI is smoother and discards the less frequent data points (outliers) presented in cold colors, thus reducing the variability introduced by extrinsic factors and focusing on the intrinsic properties of the rock. As we will see later, this measurement provides better separation and less overlap when interpreting the spatial domains.

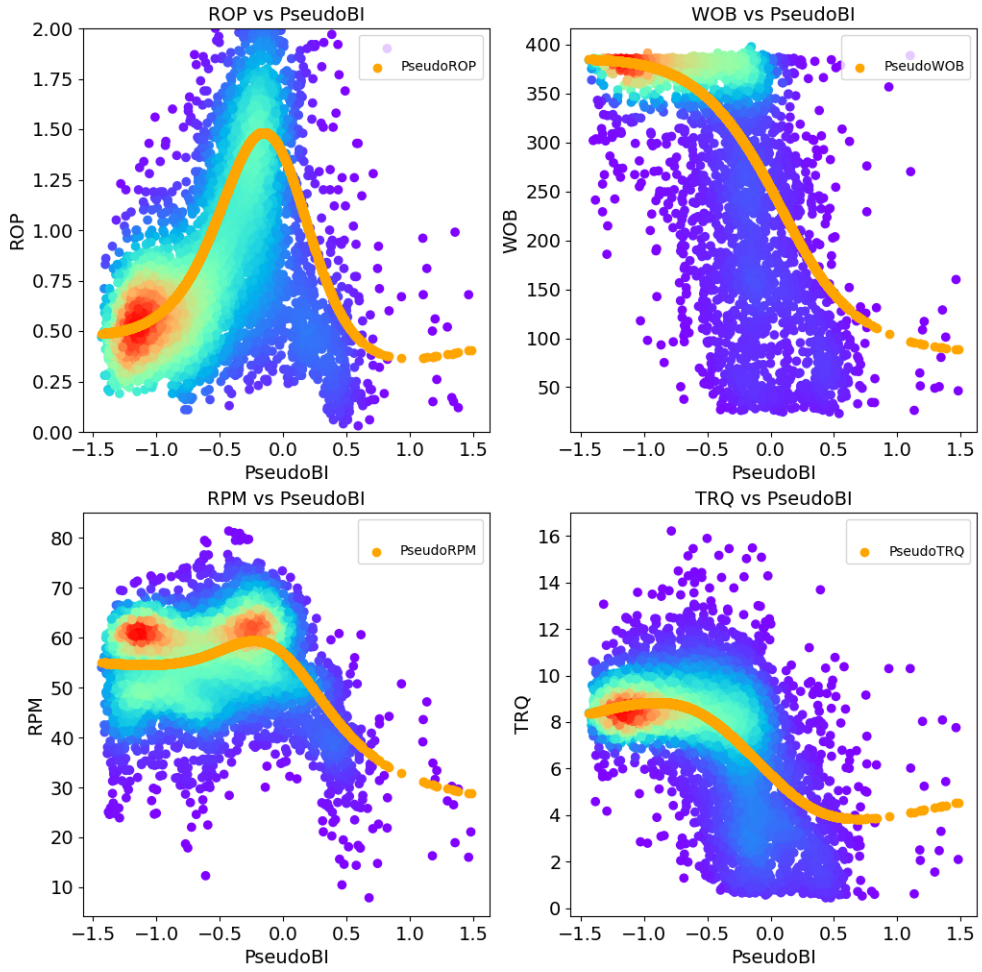


Figure 5.4 The reconstructed Pseudo-parameters (orange scatter) and original parameters (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with FC layers on the pattern (P1). For the rainbow scatters more intense and hot colors indicate a higher concentration of data points.

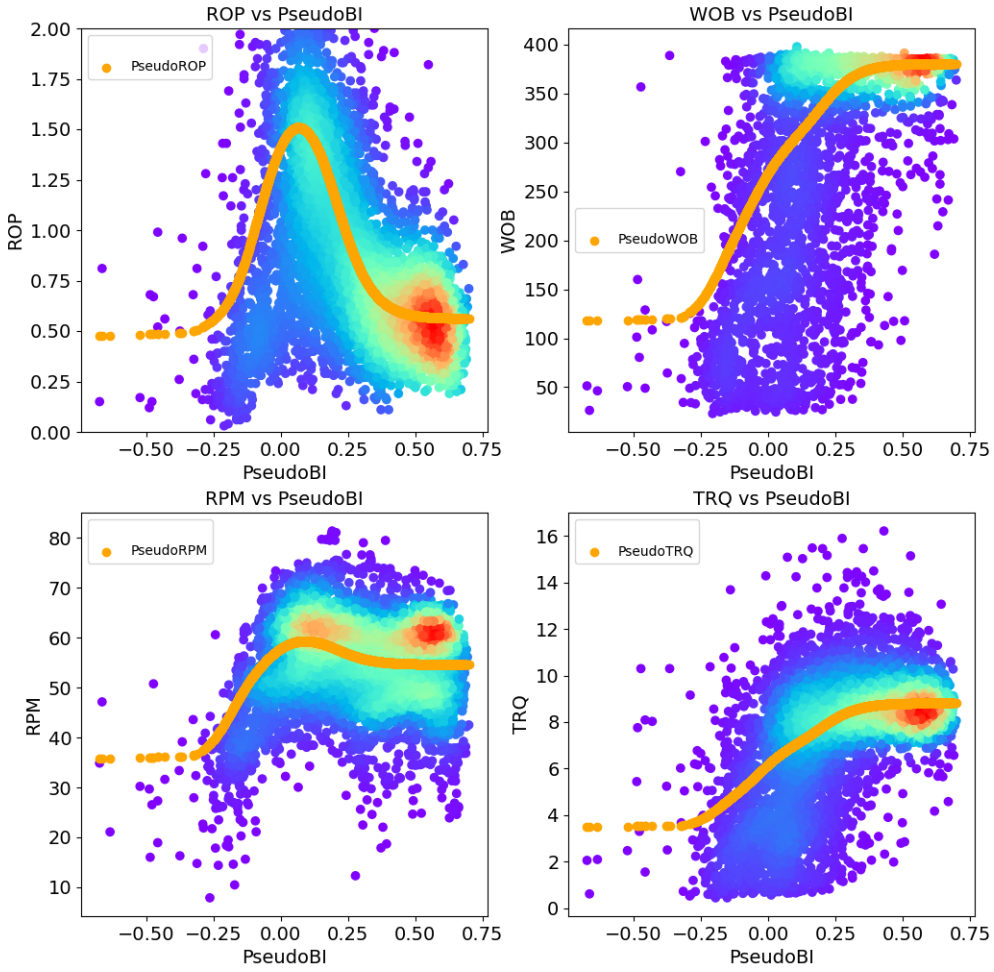


Figure 5.5 The reconstructed Pseudo-parameters (orange scatter) and original parameters (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with Conv layers on the pattern (P1).

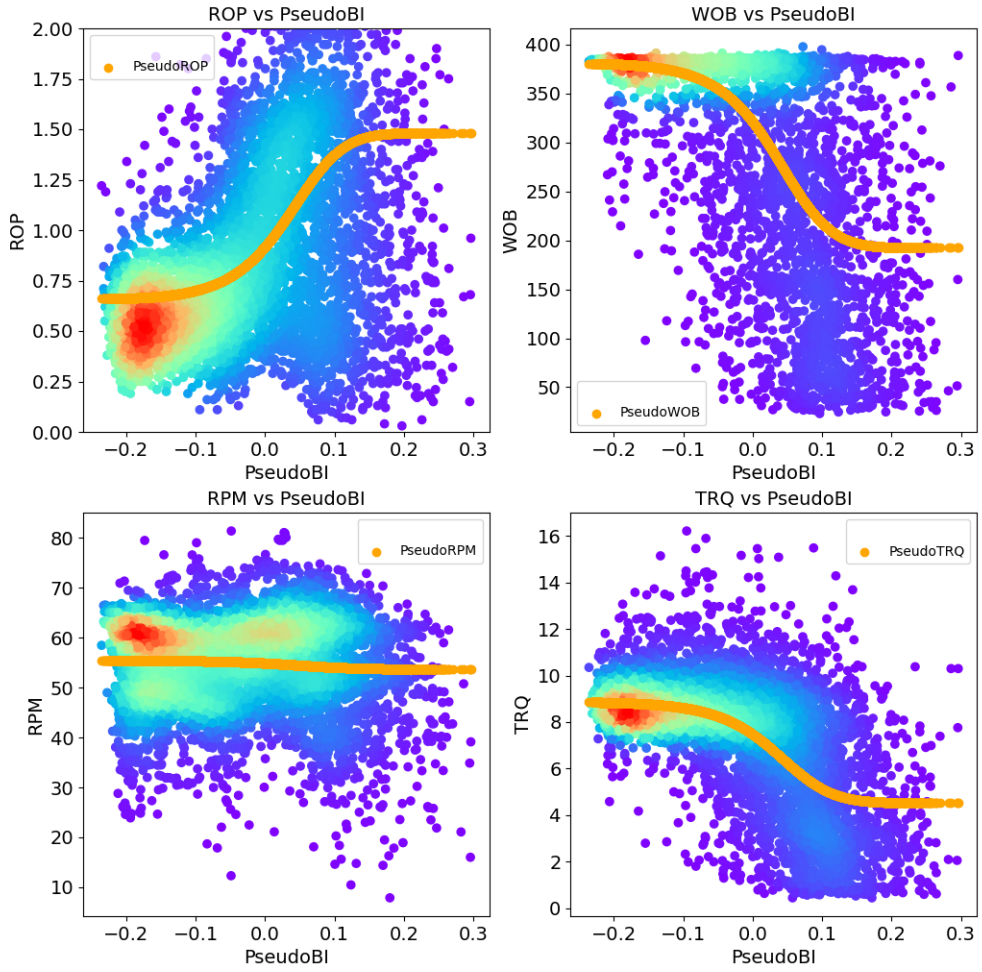


Figure 5.6 The reconstructed Pseudo-parameters (orange scatter) and original parameters (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with RBF layers and Gaussian basis functions on the pattern (P1).

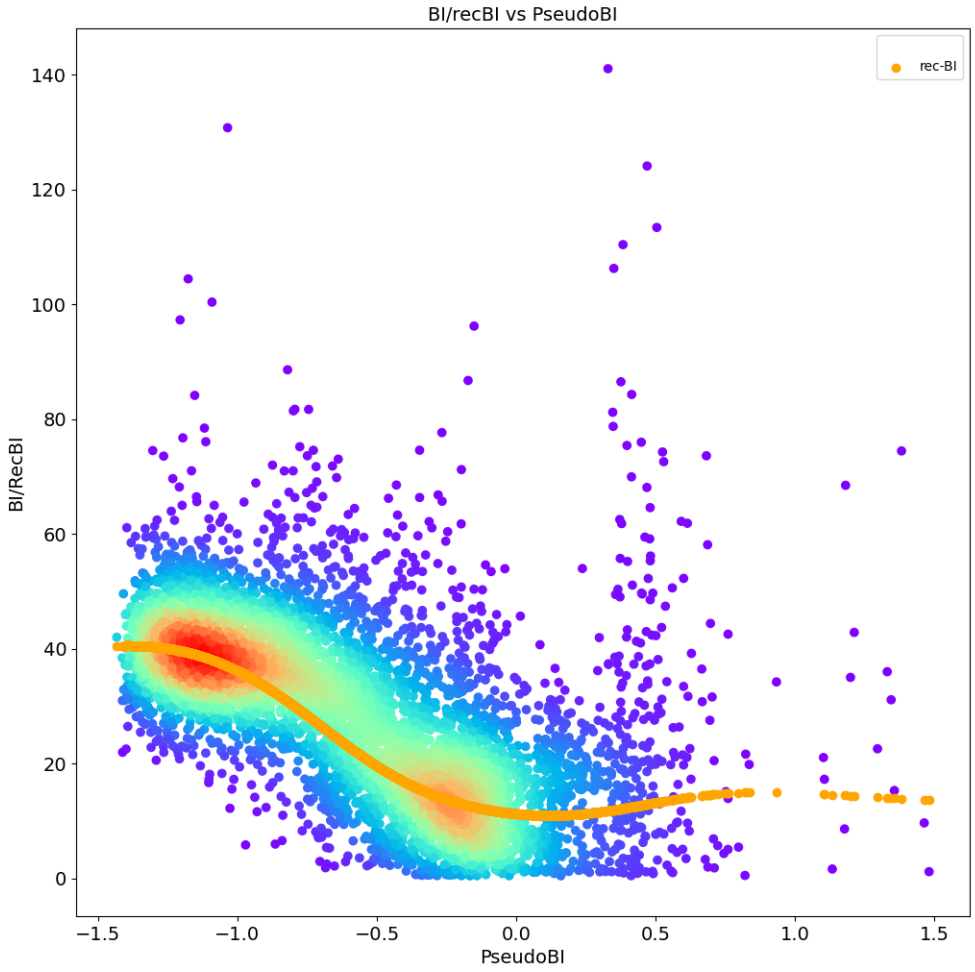


Figure 5.7 The Rec-BI (orange scatter) and BI (rainbow scatter) as a function of the learned latent feature (Pseudo-BI) using the model with FC layers on the pattern (P1).

5.4.3 Collar detection

Before starting drilling operations, operators often begin blasting the surface of the previous bench to facilitate access to the ore body, resulting in a zone comprising broken rocks known as the collar. Due to the need for caution while traversing this zone during drilling, MWD within this area do not accurately reflect the true hardness of the rock or the energy required for drilling. The collar is estimated to extend to a depth of approximately 2 meters. Identifying and subsequently excluding this collar zone from the dataset is vital to obtain precise insight into the genuine rock hardness during drilling operations.

Our objective is to compare the clusters obtained using the blastability index (BI) with those generated using the pseudoblastability index (pseudo-BI) as features. We aim to discern the effectiveness of each in detecting the broken material zone, particularly the collar. This comparison will aid in determining which feature representation yields better clustering results to accurately identify the collar zone and distinguish it from the surrounding rock formations.

Figure 5.8 shows the two clusters after using Birch and agglomerative clustering on the pseudo-BI learned using the model with FC layers versus those provided using clustering applied to the BI itself. The blue group has a lower median BI (requiring less energy to be blasted), indicating softer material, whereas the red group indicates harder material. Since the collar is a broken material, it will eventually be considered soft. Visually, we can see that collar detection is more accurate when using pseudo-BI. Due to the variability of the BI inherited from the original mechanical parameters, its clusters are not spatially contiguous and separable, showing more overlap and making it difficult to clearly identify the collar zone. However, pseudo-BI clusters are more continuous and spatially separable, accurately representing the collar zone as experts presumed to be around the first two meters.

Figures 5.9 and 5.10 show the cases for the pseudo-BI learned using, respectively, the models with Conv and RBF layers. We notice that visually, all models achieved good collar detection.

To numerically measure the performance of a given feature in providing spatially contiguous and separable clusters, we used an SVM linear classifier to perform a classification task on the provided clusters, using only the spatial coordinates as features. The classification accuracy indicates how well these clusters are spatially linearly separable and non-overlapping. We refer to this metric as DA.

Figure 5.11 presents boxplots showing DA for all patterns, comparing the performance of clustering on BI with clustering on Pseudo-BI learned from the proposed models (FC, Conv, RBF-Gaussian, RBF-inv-quad, RBF-spline). As shown, BI resulted in the lowest DA values compared to all proposed models. Among the proposed models, the RBF-inv-quad exhibited

the highest median DA value, indicating that it generally produces domains with the best spatial characteristics. However, all proposed models showed comparable performance in terms of DA, with some models exhibiting more variability than others. This aspect will be discussed in greater detail in the next section when comparing the three proposed models.

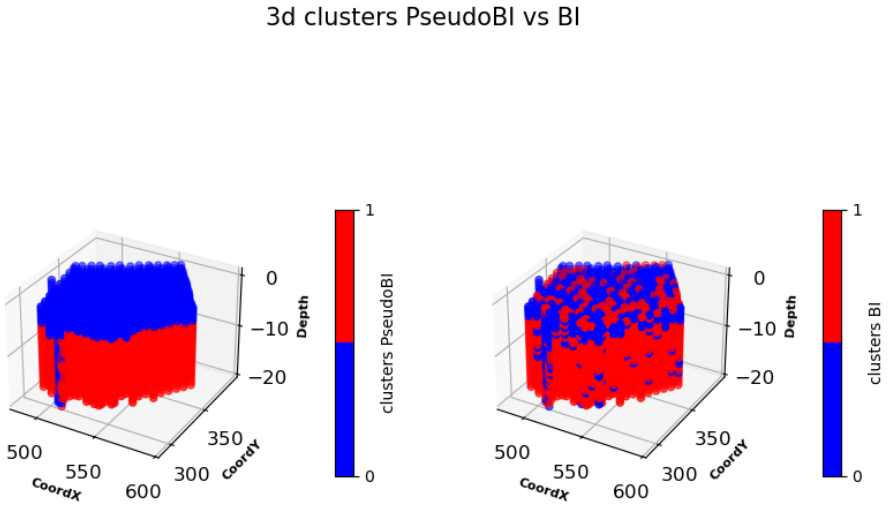


Figure 5.8 Two clusters provided by Birch and agglomerative clustering applied on Pseudo-BI learned using the model with FC layers (left) vs those applied on BI (right) for the pattern (P1).

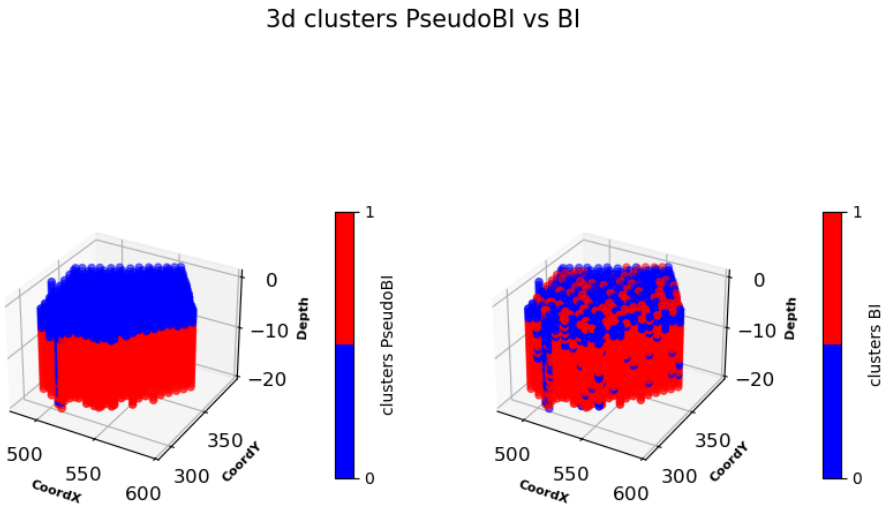


Figure 5.9 Two clusters provided by Birch and agglomerative clustering applied on Pseudo-BI learned using the model with Conv layers (left) vs those applied on BI (right) for the pattern (P1).

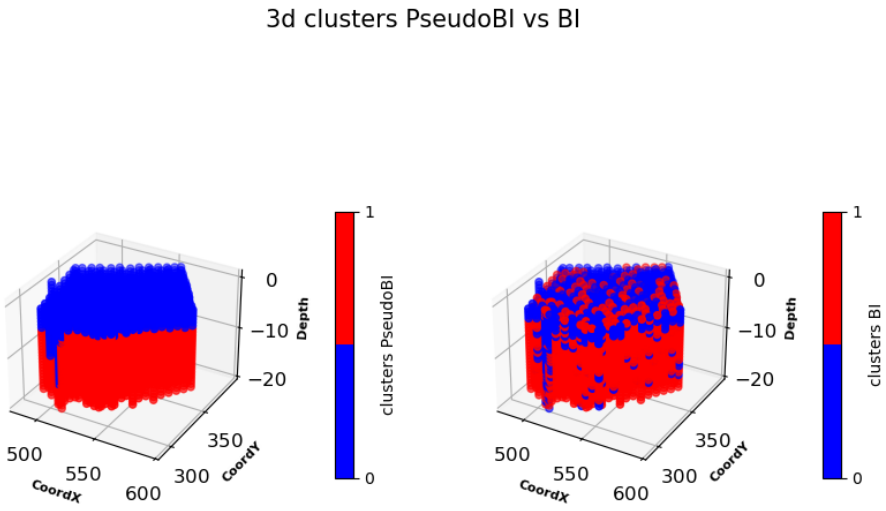


Figure 5.10 Two clusters provided by Birch and agglomerative clustering applied on Pseudo-BI learned using the model with RBF layers (left) vs those applied on BI (right) for the pattern (P1).

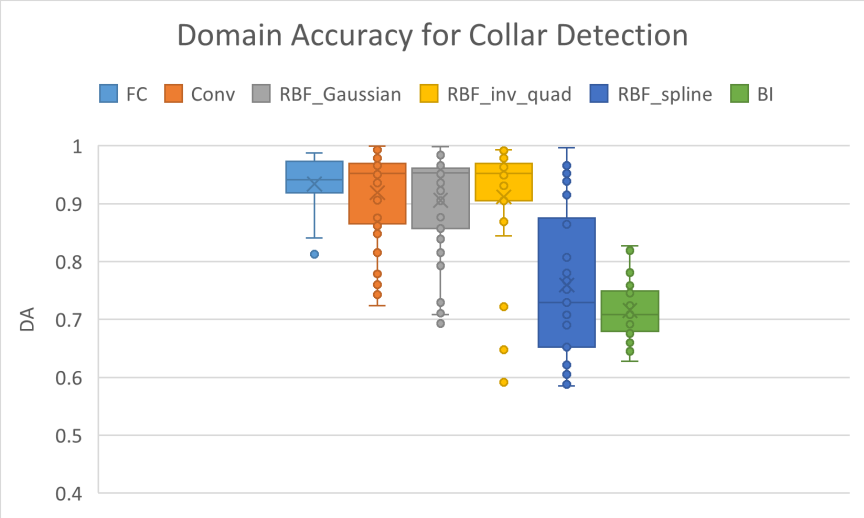


Figure 5.11 Box plots showing the Domain Accuracy for domains provided by performing clustering on BI, Pseudo-BI learned using the model with FC layers, Conv layers, RBF-Gaussian, RBF-inv-quad, and RBF-spline.

5.4.4 Rock Mass Domaining of Patterns with no Collar Using Pseudo-BI Based Clustering

After excluding the collar zone from each pattern dataset, we apply clustering on the pseudo-BI learned by each of the designed models (FC, RBF with variants Gaussian, Inverse Quadratic and Spline, and Conv). These clusters are considered domains with consistent rock hardness. To assess rock hardness in general, BI is typically used as an indicator. However, we demonstrate that Rec-BI is a better indicator in this case. Since Rec-BI is computed using the pseudo-parameters, it effectively handles variability and outliers. Moreover, it aligns with spatial and geological constraints, as these parameters are reconstructed from the pseudo-BI.

To assess the efficiency of domaining and ensure that rock hardness is consistent within each domain, we use two metrics: Domain Accuracy, as mentioned in the previous section, to assess spatial separability and contiguity; the Pooled Standard Deviation (PSD) with respect to Rec-BI (or BI if BI is used to assess rock hardness). A low PSD indicates that the Rec-BI measurements within each cluster are relatively close to their cluster centers, ensuring consistent rock hardness. In contrast, a high PSD implies greater variability within clusters, which can lead to less distinct grouping with respect to rock hardness.

Figure 5.12(a) shows an example of how the domains are spatially conceived when performing three clusters using pseudo-BI as the feature learned by the model with FC layers. We used the same pattern, (P1), that was previously considered. The DA is 0.84 in this example, which

is considered good because, as we see, the three clusters are spatially linearly separable and show some contiguity.

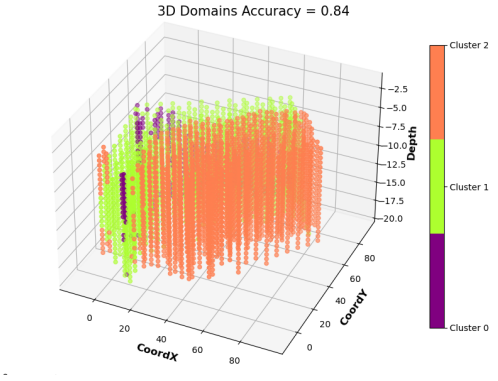
Figures 5.12(b) and 5.12(c) provide boxplots showing the Rec-BI and the BI for each cluster, respectively, along with their PSD metric values. As we can see, Rec-BI shows greater cluster separation and lower PSD, as it ignores variabilities, extreme values, and outliers. This is expected since it is reconstructed from pseudo-BI, which takes spatial geological similarities into account. Hence, we suggest using Rec-BI as a physical measure to assess and interpret the rock hardness of the provided domains.

Figures 5.13(a) and Figure 5.13(b) show the boxplots for PSD values for all patterns and the associated PSD standard deviation, respectively, using 3 clusters for domaining for each of the designed models.

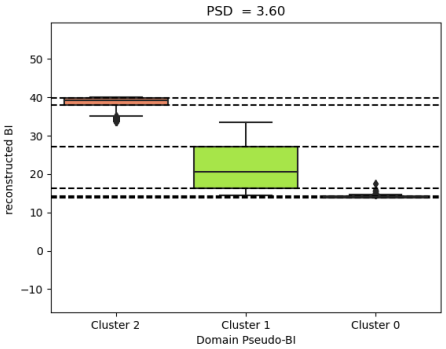
Figures 5.14(a) and Figure 5.14(b) show the boxplots for the DA values for all the patterns and the associated DA standard deviation, respectively, using 3 clusters for domaining for each of the designed models.

On average, the model with the RBF Spline variant layers achieved better results in terms of PSD, as evidenced by Figure 5.13(a). The results also show lower PSD variability compared to models with Conv and FC layers, as demonstrated in Figure 5.13(b). This suggests that the RBF Spline variant domains exhibit a more distinct grouping with respect to Rec-BI. Additionally, from the visual analysis of both figures, it is evident that the RBF-Spline variant consistently demonstrates tighter clustering with minimal variability across patterns. Rec-BI values remain close to cluster centers, reinforcing the consistency and reliability of the domains identified by this model. These findings are further supported by the overall lower PSD standard deviation shown in Figure 5.13(b), emphasizing the robustness of the RBF Spline architecture in achieving spatially consistent and geologically meaningful domains. Regarding spatial separability, the RBF models on average achieved a higher DA compared to FC and Conv. However, the DA standard deviation is highest for RBF-Spline, indicating that, for some patterns, its DA is compromised over Rec-BI similarity compared to other models. RBF-inv-quad shows a good balance between DA and DA consistency. In contrast, the FC model shows comparable DA with very low variability, but PSD is compromised, since it is highest for the FC model. This suggests a potential trade-off between spatial separability (DA) and rock hardness consistency within domains (PSD). Although RBF-Spline excels in achieving high DA, it may sometimes sacrifice the internal homogeneity of the resulting domains, as measured by Rec-BI. RBF-Gaussian and RBF-inv-quad appear to strike a better balance between these two objectives. However, FC models prioritize consistent Rec-BI within clusters but might result in less spatially distinct and potentially

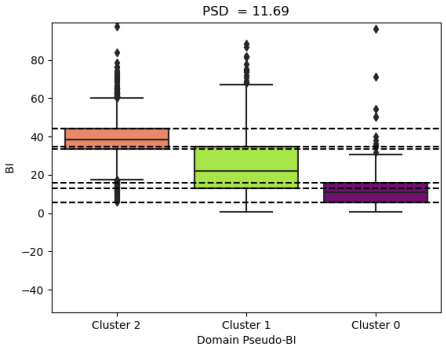
more fragmented domains. In conclusion, the RBF-spline model shows a strong performance in identifying domains with consistent rock hardness (low PSD and low PSD variability). However, it exhibits higher variability in its DA performance compared to other models, suggesting a potential trade-off between spatial separability and rock hardness consistency. The RBF-inv-quad appears to offer a more balanced approach, achieving reasonable DA with lower variability and maintaining good control over PSD. The choice of model will depend on the specific priorities of the application, whether it is more critical to prioritize spatially distinct domains or to ensure high internal homogeneity within the domains.



(a) Example of 3D Spatial Domains Obtained by Pseudo-BI Clustering for Pattern (P1) Using FC Layers.

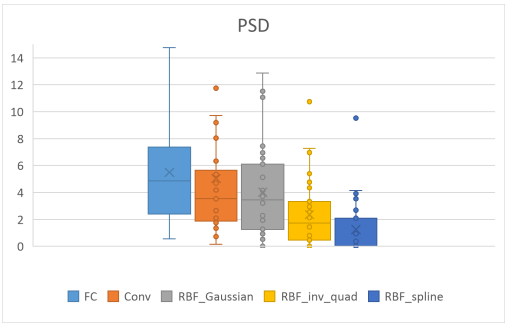


(b) Boxplots of Rec-BI for Each Cluster in Pattern (P1) Using FC Layers.

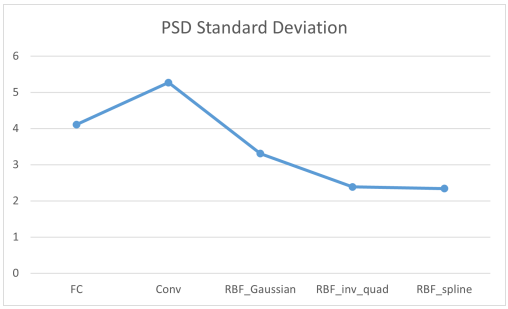


(c) Boxplots of BI for Each Cluster in Pattern BW-648-035 Using FC Layers.

Figure 5.12 Boxplots of BI for Each Cluster in Pattern (P1) Using FC Layers.

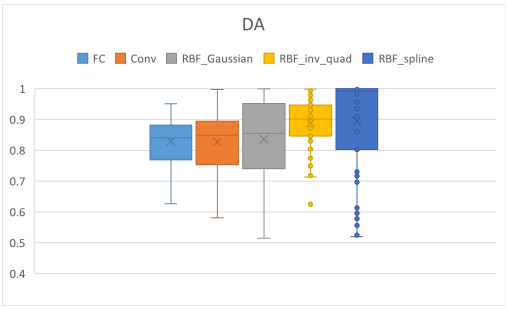


(a) Pooled Standard Deviation (PSD) Boxplots for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models.

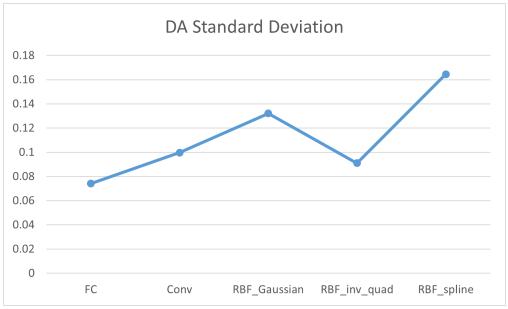


(b) Standard Deviation of PSD for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models.

Figure 5.13 Subfigures illustrating the Pooled Standard Deviation and Standard Deviation of PSD for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models.



(a) Domain Accuracy (DA) Boxplots for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models.



(b) Standard Deviation of DA for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models.

Figure 5.14 Subfigures illustrating Domain Accuracy (DA) Boxplots and Standard Deviation of DA for Pseudo-BI Domaining Across All Patterns Using FC, Conv, and RBF Models.

5.4.5 Correspondance of Some Examples with Lithology

In this section, we present examples where the domains identified by the developed models correspond to lithologies observed by geologists during surface tests of each borehole. As mentioned above, there is no one-to-one correspondence between rock hardness and lithologies. However, in many cases where the lithologies exhibit significantly different hardnesses, the similarity in patterns becomes visibly clear. The fact that the provided domains capture

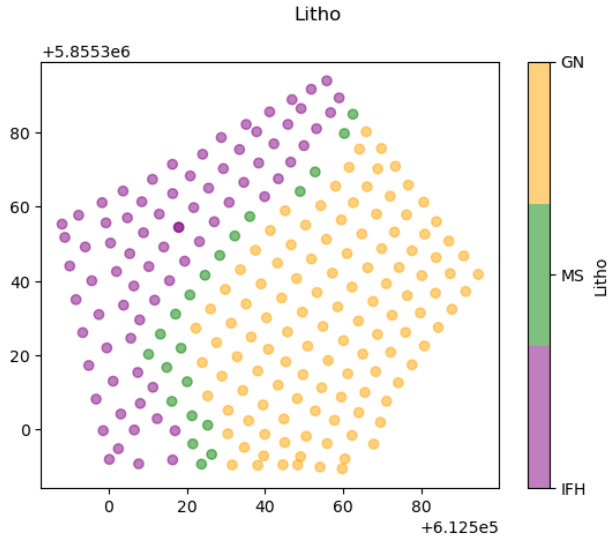
such similarities is encouraging. It reassures that the domains can distinguish certain formations using only MWD data, without relying on expert geologists. This also helps extend the understanding of formations beyond the surface, by incorporating the subsurface component, as MWD data is available along boreholes. Such models could serve as powerful tools in geomodeling software for mining applications, not only to divide the mining field into domains characterized by consistent rock hardness but also to gain insight into the three-dimensional lithology and geological formations beneath the surface.

Figure 5.15 shows the lithology of the pattern (P1) (left graph) and the three-dimensional domains obtained using clustering in pseudo-BI derived from the model with FC layers (right graph). According to the boxplot shown in the previous section of Rec-BI values for each cluster, cluster 0 (purple), cluster 1 (green) and cluster 2 (orange) are in increasing order of rock hardness. For this particular pattern, geologists detected three types of lithology on the surface: IFH (purple), MS (green), and GN (orange). The domaining results align with these formations. Specifically, IFH and MS are predominantly represented by Cluster 0 (purple) and Cluster 1 (green), indicating that they are softer compared to GN, which is represented by Cluster 2 (orange).

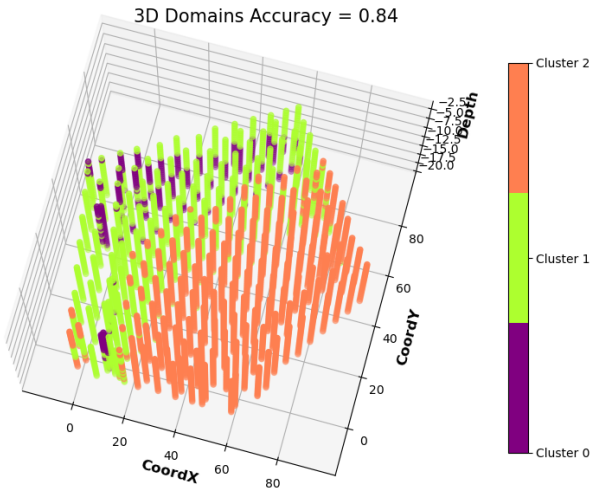
In Figure 5.16, we present the case of a pattern (P2), characterized by the same three lithologies as in the previous example. The plot in the left panel shows the three-dimensional domains obtained through clustering on pseudo-BI derived from the model with Conv layers. Similarly to the previous example, IFH and MS are predominantly represented by cluster 0 (purple) and cluster 1 (green), indicating that they are softer compared to GN, which is represented by cluster 2 (orange). From these two examples, we observe that domaining effectively differentiates GN from IFH and MS, suggesting that GN is considerably harder.

Figure 5.17 presents the case for a pattern (P3), where the RBF-Gaussian model is used for domaining. In this example, the three formations are represented by the three clusters, and the patterns are very close to the lithologies shown in the right-hand plot. Thus, we can associate IFH with cluster 0, MS with cluster 1, and GN with cluster 2. Based on our knowledge of the hardness of each cluster, we can conclude that IFH, MS, and GN are in increasing order of hardness.

Using our approach, domaining provided valuable insight into the three-dimensional formation and helped to assess the hardness of the lithologies determined by expert geologists.

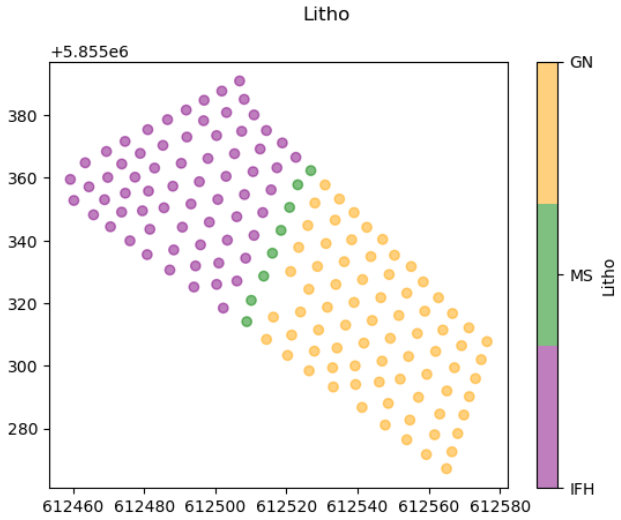


(a) Lithology for Pattern (P1).

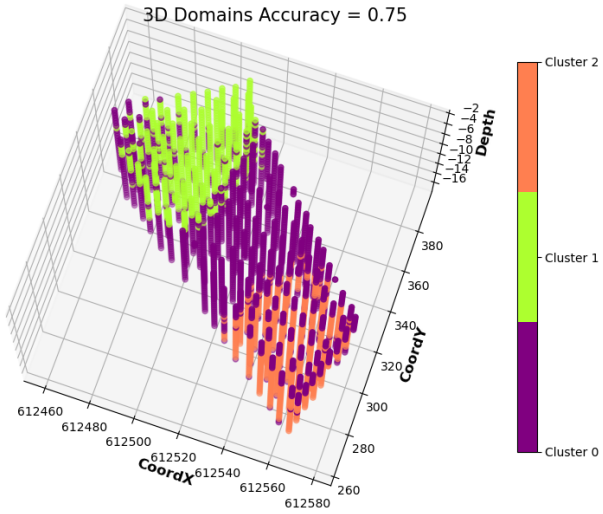


(b) 3D Spatial Domains for Pattern (P1) Using Pseudo-BI Clustering with FC Layers.

Figure 5.15 Comparison of Lithology for Pattern (P1) Using Pseudo-BI Clustering with FC Layers.

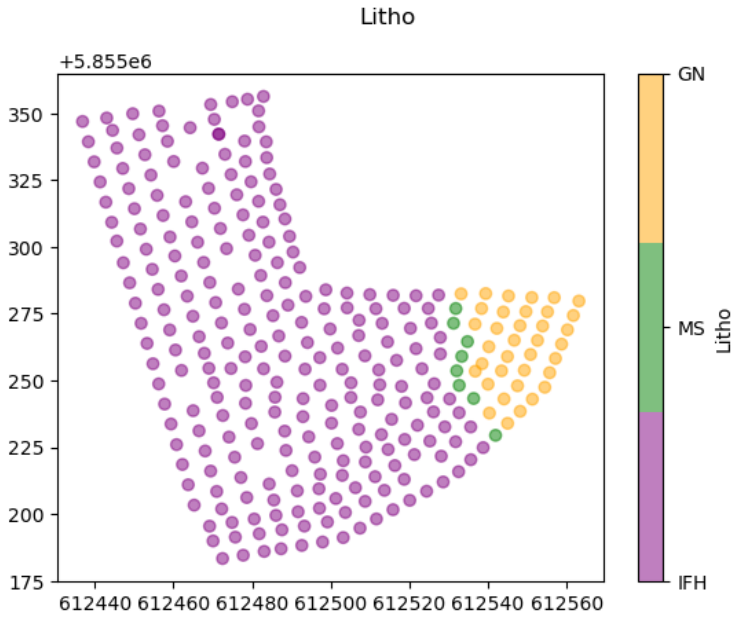


(a) Lithology for Pattern (P2).

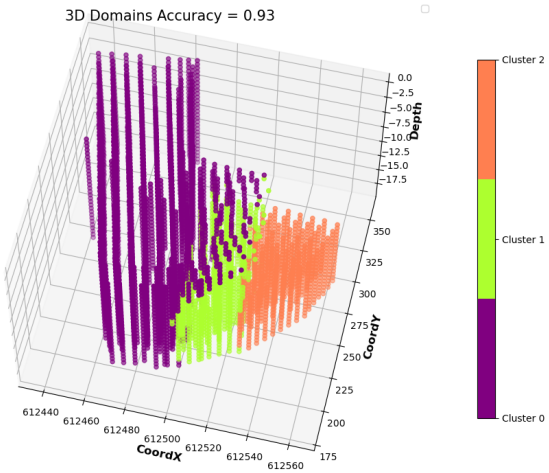


(b) 3D Spatial Domains for Pattern (P2) Using Pseudo-BI Clustering with Conv Layers.

Figure 5.16 Comparison of Lithology for Pattern (P2) Using Pseudo-BI Clustering with Conv Layers.



(a) Lithology for Pattern (P3).



(b) 3D Spatial Domains for Pattern (P3) Using Pseudo-BI Clustering with RBF-Gaussian Layers.

Figure 5.17 Comparison of Lithology for Pattern (P3) Using Pseudo-BI Clustering with RBF-Gaussian Layers

5.5 Conclusion

This study introduced a novel approach to rock mass domaining, using MWD data to identify contiguous spatial domains with consistent rock hardness. Using a stacked VAE and a spatial interpolation model, we were able to extract a latent feature, pseudo-BI, that effectively represents rock hardness. The pseudo-BI, when used for clustering, demonstrated remarkable capability in reconstructing mechanical parameters (pseudo-ROP, pseudo-WOB, pseudo-RPM, and pseudo-TRQ) that capture general trends while mitigating variability and outliers.

Our analysis showed that the spatial interpolator played a crucial role in ensuring spatial contiguity, leading to more reliable clustering results. The models with different VAE network types (FC, Conv, and RBF) all succeeded in reconstructing smoother and more consistent pseudo-parameters compared to the original mechanical parameters. The proposed Rec-BI metric, derived from pseudo-BI, provided better separation and less overlapping of clusters, offering a more accurate representation of rock hardness.

The study also highlighted the effectiveness of pseudo-BI in detecting the collar zone, where pseudo-BI-based clustering outperformed traditional BI-based clustering in identifying and excluding this zone. Collar detection was more accurate and spatially separable, particularly with the model using RBF network with inverse quadratic basis function.

In terms of rock mass domaining, pseudo-BI-based clustering proved to be a robust tool, consistently providing spatially contiguous and geologically meaningful domains across various patterns. The models were able to identify lithologies that correspond to those observed by geologists, even extending this correspondence into 3D, thus offering valuable insight into the subsurface formations.

Our analysis further revealed that the RBF-inv-quad model demonstrated strong performance in terms of DA, indicating its ability to generate spatially distinct and separable domains. Although RBF-Spline showed competitive DA performance, it exhibited higher variability in DA across different patterns. In terms of rock hardness consistency within domains, the RBF-Spline model demonstrated the lowest PSD values and the lowest PSD variability, suggesting that it effectively identifies domains with consistent rock hardness. However, this performance comes at the potential cost of a lower and more variable DA. RBF-Gaussian and RBF-inv-quad appear to offer a more balanced approach, achieving a reasonable DA with lower variability and maintaining good control over PSD.

Overall, this approach not only enhances the precision of rock mass domaining but also provides a powerful tool for geomodeling in mining applications. By effectively handling

variability and outliers and aligning with geological constraints, the pseudo-BI-based models offer a promising avenue for better understanding and managing mining fields. Future work could explore the integration of these models into broader geomodeling frameworks, potentially improving the accuracy and reliability of mining operations.

Apport du chapitre à la thèse

Dans ce chapitre, nous avons étudié l'application des VAEs couplés à l'interpolation spatiale pour la segmentation des masses rocheuses à partir des données MWD. Les résultats montrent que cette approche améliore la définition des domaines géologiques en garantissant à la fois une continuité spatiale et une homogénéité de la dureté des roches. Elle constitue une avancée significative dans la modélisation des données MWD et ouvre la voie à des méthodes d'interpolation plus flexibles et adaptées aux défis complexes de la géologie minière.

Dans le chapitre suivant, nous approfondissons cette démarche en introduisant un modèle multicouche de pondération inverse de la distance (ML-IDW), qui s'inspire des capacités des réseaux neuronaux multicouches et des modèles de pondération inverse de la distance pour mieux capturer les relations spatiales complexes et non linéaires.

CHAPITRE 6 ARTICLE 3: ENHANCING SPATIAL INTERPOLATION: A MULTI-LAYER INVERSE DISTANCE WEIGHTING MODEL FOR COMPLEX REGRESSION AND CLASSIFICATION TASKS IN SPATIAL DATA ANALYSIS

Cet article a été publié dans la revue *World Academy of Science, Engineering and Technology — International Journal of Mathematical and Computational Sciences* le 23 octobre 2024, coécrit avec Jean-François Plante, Richard Labib et Michel Gamache.

Aperçu du chapitre

Les modèles d'interpolation spatiale traditionnels sont souvent limités par leur rigidité et leur incapacité à capturer des relations complexes dans des environnements géologiques hétérogènes. Ce chapitre présente une nouvelle architecture de pondération inverse de la distance multicouche (ML-IDW), inspirée des réseaux neuronaux. En combinant la flexibilité des modèles neuronaux avec la robustesse des techniques d'interpolation classiques, cette approche améliore la précision aussi bien en régression qu'en classification spatiale.

6.1 Abstract

This study introduces the Multi-Layer Inverse Distance Weighting Model (ML-IDW), inspired by the mathematical formulation of both multi-layer neural networks (ML-NNs) and Inverse Distance Weighting model (IDW). ML-IDW leverages ML-NNs' processing capabilities, characterized by compositions of learnable non-linear functions applied to input features, and incorporates IDW's ability to learn anisotropic spatial dependencies, presenting a promising solution for nonlinear spatial interpolation and learning from complex spatial data. We employ gradient descent and backpropagation to train ML-IDW. The performance of the proposed model is compared against conventional spatial interpolation models such as Kriging and standard IDW on regression and classification tasks using simulated spatial datasets of varying complexity. Our results highlight the efficacy of ML-IDW, particularly in handling complex spatial dataset, exhibiting lower mean square error in regression and higher F1 score in classification.

6.2 Introduction

Deep learning has emerged as a transformative tool for discerning intricate structures within complex datasets, proving its efficacy across a spectrum of tasks from regression to classification in diverse domains such as natural language processing and computer vision [71]. By leveraging the inherent compositional properties of learnable non-linear functions applied to inputs, multi-layer neural networks (ML-NNs) have demonstrated remarkable prowess in pattern recognition [25]. This capability is underpinned by parameter optimization through gradient descent, facilitated by the efficient computation of gradients using the back-propagation algorithm [164].

Nevertheless, feedforward ML-NNs, like many machine learning models, often assume that data observations are independently sampled from a given distribution, thereby overlooking the spatial dependency structures inherent in the data. Conversely, spatial interpolation models like Inverse Distance Weighting (IDW) excel at capturing such dependencies, but they frequently lack the flexibility and the learning capacity of ML-NNs.

Numerous studies have compared machine learning against spatial interpolation (SI) models [31,91,155,164], showcasing the advantages of ML models in accurately depicting nonlinear relationships [91]. However, discussions on the limitations of ML models in modeling spatial relationships have led to a preference for spatial interpolations in scenarios where data exhibit significant spatial correlations [143]. Additionally, by employing linear interpolation, SI models furnish smooth predictions that risk overlooking abrupt changes and non-linearities in the data [206]. In response to these constraints, many researchers have proposed the fusion of both approaches [37, 76, 86, 109, 180, 190].

For instance, some studies have combined SI with artificial neural networks (ANNs), utilizing SI to augment data by interpolating in locations lacking observational features, subsequently leveraging ANN that uses the provided feature maps as dependent variables for prediction [190]. Other studies combined Regression Kriging with machine learning models like Support Vector Machine (SVM) where the ML model predicts the drift, while Kriging performs spatial interpolation on the residuals, treating them as corrections to the predicted values [109, 138]. This approach has also been extended to classification tasks, where Simplical Indicator Kriging is combined with SVM for classification purposes [138, 187].

Despite the effectiveness of many ML models, they still lack the representational capacity, computational efficiency, and flexibility of deep neural networks, in which model complexity is customizable relative to the datasets. Proposed architectures such as Residual Neural Networks have been introduced to safely use deeper models without risking the vanishing

gradient problem [69]. To our knowledge, no study has yet been conducted to introduce a spatial interpolation model that has similar benefits. We aim to address this gap through our study by introducing such a model: the ML-IDW model. This model applies compositions of learnable non-linear functions to input data and uses IDW interpolation with learnable parameters. Trained using gradient descent and back-propagation, it is compatible to be stacked with other ANNs models or architectures. This would allow us to leverage the power of deep learning while simultaneously accounting for spatial dependency structures.

6.3 Theory and Methodology

6.3.1 Conventional IDW: a brief background

Let the data consist of targets t_i, $i = 1, ..., N$, which are spatially correlated and located in 2D spatial coordinates $\mathbf{s}_i = [x_i, y_i]^T$. We denote by $\mathcal{T}$ the set of indices in the test data and by $\overline{\mathcal{T}}$ the set of indices in the training data. Conventional Inverse Distance Weighting (IDW) performs a prediction at an unsampled location s_j, $j \in \mathcal{T}$, by computing an average weighted sum of targets t_i, $i \in \overline{\mathcal{T}}$:

$$\widehat{t}_j = \sum_{i \in \overline{\mathcal{T}}} w_{ji} t_i, \tag{6.1}$$

where $w_{ji} = d_a(\mathbf{s}_i, \mathbf{s}_j)^{-p} / \sum_k d_a(\mathbf{s}_k, \mathbf{s}_j)^{-p}$ are the IDW weights that depend on the power parameter p. The term $d_a(\mathbf{s}_i, \mathbf{s}_j) = \sqrt{(x_i - x_j)^2/\ell_x^2 + (y_i - y_j)^2/\ell_y^2}$ represents the Euclidean distance accounting for anisotropy in the spatial correlation, where ℓ_x is called the range, or length-scale, of the x principal direction, and similarly for ℓ_y the principal direction of y [101, 172]. In the conventional approach, the set of parameters to be learned, denoted $\theta = \{p, \ell_x, \ell_y\}$, are derived using the Leave One Out Cross Validation (LOOCV) algorithm [13, 115]. This algorithm involves conducting a grid search G of possible parameter candidates in order to choose the set of parameters that minimizes the LOOCV error: $E^{LOOCV} = (1/|\overline{\mathcal{T}}|) \sum_{i \in \overline{\mathcal{T}}} (\widehat{t}_i - t_i)^2$. This error consists of the average of the squared errors between every target t_i in the training data and its IDW estimate $\widehat{t}_i$, computed using IDW performed on all the training data, leaving the i^{th} observation out.

This training approach has a time complexity scaling as $O(|\overline{\mathcal{T}}|^2 * |G|)$ and a space complexity of $O(|\overline{\mathcal{T}}| + |G|)$, where $|\overline{\mathcal{T}}|$ represents the size of the training dataset without the omitted observation, and $|G| = m^3$ denotes the size of the grid, with m being the number of parameter possibilities. This method faces challenges in scalability due to the potentially large size of the grid $|G|$. As m increases to explore more parameter possibilities, both time and space complexity grow significantly. While a higher m might lead to improved accuracy by

considering a finer-grained search over parameter space, it also compromises scalability.

6.3.2 Gradient Based IDW

To address the challenges posed by conventional Inverse Distance Weighting (IDW) interpolation, we introduced a Gradient-Based IDW (GB-IDW) approach in a previous study. In GB-IDW, we leveraged the matrix formulation of an IDW model to compute the leave-one-out cross-validation (LOOCV) error in a single forward pass: $E^{LOOCV}(\theta) = 1/|\overline{\mathcal{T}}| \left\|\widehat{\mathbf{t}} - \mathbf{t}\right\|^2 = 1/|\overline{\mathcal{T}}| \left\|\mathbf{W}(\theta)\mathbf{t} - \mathbf{t}\right\|^2$.

Here, $\mathbf{t}$ and $\widehat{\mathbf{t}}$ are vectors containing the true values and the IDW predicted values, respectively, for $i \in \overline{\mathcal{T}}$. $\mathbf{W}$ is a matrix with zeros on the diagonal and GB-IDW weights w_{ji} elsewhere. $\|.\|$ denotes the Euclidean norm.

We exploited the differentiability of this error with respect to the parameter set θ to apply gradient descent and backpropagation for minimizing $E^{LOOCV}(\theta)$. The time complexity of this training approach scales as $O(|\overline{\mathcal{T}}|^2 \times |\mathcal{I}|)$, where $|\mathcal{I}|$ denotes the number of training iterations. This is because the backpropagation algorithm scales linearly with the number of parameters ($O(3 \times |\mathcal{I}|)$) [163], making the time complexity dominated by $O(|\overline{\mathcal{T}}|^2 \times |\mathcal{I}|)$. This represents a significant reduction in time complexity compared to conventional methods, especially when $|\mathcal{I}| \ll |G|$.

Additionally, the space complexity is dominated by $O(|\overline{\mathcal{T}}|^2)$. While time and space complexity could be a drawback for large data sizes, leveraging the power of Graphics Processing Units (GPUs) in parallelism and memory efficiency can mitigate this challenge significantly.

By exploiting the optimized matrix operations inherent in GPUs, we can achieve substantial improvements in computational efficiency [60, 75, 195]. GPUs are designed to handle parallel operations efficiently, particularly in deep learning tasks, making them well-suited for accelerating the training process of GB-IDW. This can effectively address the scalability challenges encountered in large-scale interpolation tasks.

6.3.3 Single Layer IDW

GB-IDW, trained with gradient descent and backpropagation, is compatible with neural networks and can be stacked with an Artificial Neural Network (ANN) for further processing of the targets $\mathbf{t}$. We introduce a Single Layer IDW (SL-IDW) by adding a single processing layer to the vector $\mathbf{t}$ before using GB-IDW for interpolation.

Let X denote dependent covariates that may contain other dependent features along with the

spatial locations $\mathbf{S} = [\mathbf{s}_i]$. We employ a simple feedforward Neural Network, denoted as NN, which takes X as input and produces two vectors $\mathbf{c} = [c_i]$ and $\mathbf{b} = [b_i]$ representing the slope and intercept for $\mathbf{t}$, respectively. A simple linear regression is performed on $\mathbf{t}$ to compute the hidden vector $\mathbf{h} = \mathbf{c} \odot \mathbf{t} + \mathbf{b}$ using these two parameters, where $\odot$ denotes elementwise product and $+$ denotes elementwise sum ($\mathbf{c} \odot \mathbf{t} + \mathbf{b} = [c_i t_i + b_i]$). We then apply an activation function $g^{(1)}$ to introduce nonlinearity and compute the hidden layer $\mathbf{a}^{(1)} = g^{(1)}(\mathbf{h})$.

Spatial interpolation is then applied to $\mathbf{a}^{(1)}$ by computing $\widehat{\mathbf{t}} = \mathbf{W}(\theta)\mathbf{a}^{(1)}$. The LOOCV error $E^{LOOCV}(\theta, \varphi) = (1/|\overline{\mathcal{T}}|) \left\| \mathbf{W}(\theta)\mathbf{a}^{(\mathbf{1})} - \mathbf{t} \right\|^2$ remains differentiable with respect to all parameters of the model, comprising $\theta = \{p, \ell_x, \ell_y\}$ as GB-IDW parameters, and φ, as the parameters of the feedforward neural network NN.

In summary, the interpolation equation of the Single Layer IDW is:

$$\widehat{\mathbf{t}} = \mathbf{W}(\theta)\mathbf{a}^{(1)},$$

$$\mathbf{a}^{(1)} = g^{(1)}(\mathbf{h}),$$

$$\mathbf{h} = \mathbf{c} \odot \mathbf{t} + \mathbf{b}, \quad \mathbf{c}, \mathbf{b} = NN(X).$$

6.3.4 Mutli-Layer IDW for regression

Multi-Layer IDW (ML-IDW) for regression extends Single Layer IDW (SL-IDW) by incorporating multiple processing layers for $\mathbf{t}$. Each layer k takes $\mathbf{a}^{(k-1)}$ from the previous layer as input and computes the hidden vector $\mathbf{h}^{(k)} = \mathbf{c}^{(k)} \odot \mathbf{a}^{(k-1)} + \mathbf{b}^{(k)}$, where $\mathbf{c}^{(k)}, \mathbf{b}^{(k)} = NN^{(k)}(X)$ represent the intercept and slope provided by the feedforward neural network for layer k, $NN^{(k)}$. An activation function $g^{(k)}$ is then applied to compute $\mathbf{a}^{(k)} = g^{(k)}(\mathbf{h}^{(k)})$.

Inspired by Residual Neural Networks (ResNet), which learn a nonlinear residual to facilitate identity map learning and address vanishing gradients [69], we introduce a skipping connection layer such that $\mathbf{a}^{(k)} = g^{(k)}(\mathbf{h}^{(k)}) + \mathbf{h}^{(k-2)}$. Spatial interpolation is performed on the output of the last layer L: $\widehat{\mathbf{t}} = \mathbf{W}(\theta)\mathbf{a}^{(L)}$.

In summary, the interpolation equation of ML-IDW is:

$$\widehat{\mathbf{t}} = \mathbf{W}(\theta)\mathbf{a}^{(L)},$$

$$\mathbf{a}^{(k)} = g^{(k)}(\mathbf{h}^{(k)}) + \mathbf{h}^{(k-2)}, \quad k = 2, ..., L,$$

$$\mathbf{a}^{(1)} = g^{(1)}(\mathbf{h}^{(1)}),$$

$$\mathbf{a}^{(0)} = \mathbf{t},$$

$$\mathbf{h}^{(k)} = \mathbf{c}^{(k)} \odot \mathbf{a}^{(k-1)} + \mathbf{b}^{(k)},$$
$$\mathbf{c}^{(k)}, \mathbf{b}^{(k)} = NN^{(k)}(X), \quad k = 1, ..., L.$$

It's worth noting that with this setup, the model will also be able to learn through spatial interpolation on the values $\mathbf{t}$, any unknown mapping $\mathbf{u}$ of $\mathbf{t}$: $\mathbf{u} = f(\mathbf{t})$. A special case could be explored for classification in which we transform $\mathbf{t}$ into categorical variables $\mathbf{u}$ using the function $u_i = f(t_i) = 1$ if $t_i \geq \alpha$, $u_i = f(t_i) = 0$ otherwise, where α is a threshold. Such a setup will make our model competitive with nonlinear spatial interpolation methods such as Simplicial Indicator Kriging.

6.3.5 Mutli-Layer IDW for classification

In ML-IDW for classification, we aim to predict categorical spatial targets $u_i \in \{1, ..., C\}$, where C is the number of possible categories obtained through categorizing the targets t_i. We extend ML-IDW for regression to perform classification tasks by introducing a feedforward neural network FN and applying a softmax function to each value $\widehat{t}_i$ to obtain the vector of probabilities $\mathbf{p}_i = [p_i^{(1)}, ..., p_i^{(C)}]$ containing the probability of each category. The classification rule in this case is assigning the class characterized by the highest probability: $\hat{u}_i = \text{argmax}(\mathbf{p}_i)$. The cross-entropy loss function is utilized as the minimization criterion: $L(\mathbf{U}, \mathbf{P}) = 1/|\overline{\mathcal{T}}| \sum_{i=1}^{|\overline{\mathcal{T}}|} \sum_{l=1}^{C} u_i^{(l)} \ln(p_i^{(l)})$. Here, $\mathbf{U} = [u_i^{(l)}]$ contains the ground truth labels (1 if sample i belongs to class l, 0 otherwise), and $\mathbf{P} = [p_i^{(l)}]$ contains the predicted probabilities that sample i belongs to class l. In summary the interpolation equation of ML-IDW in case of classification is:

$$\mathbf{P} = \text{softmax}(FN(\mathbf{W}(\theta)\mathbf{a}^{(L)})),$$

$$\mathbf{a}^{(k)} = g^{(k)}(\mathbf{h}^{(k)}) + \mathbf{h}^{(k-2)}, \quad k = 2, ..., L,$$

$$\mathbf{a}^{(1)} = g^{(1)}(\mathbf{h}^{(1)}),$$

$$\mathbf{a}^{(0)} = \mathbf{t},$$

$$\mathbf{h}^{(k)} = \mathbf{c}^{(k)} \odot \mathbf{a}^{(k-1)} + \mathbf{b}^{(k)}, \quad \mathbf{c}^{(k)},$$
$$\mathbf{b}^{(k)} = NN^{(k)}(X), \quad k = 1, ..., L.$$

6.4 Experiments and results

6.4.1 Data description

To assess our models, we employed the Turning Band method [150] for geostatistical simulation to generate 2D spatial data across varying sizes, ranging from 100 to 10,000 data points. We utilized two types of variogram models for simulation to provide spatially correlated datapoints: a simple model composed of a Gaussian variogram and a nested model combining Gaussian and exponential variograms. Across all datasets, the partial sill was set at 20 with a nugget of 4. Anisotropy was introduced by establishing the range along the y-axis as one third of the range along the x-axis ($\ell_y = \ell_x/3$). We introduced complexity variations by considering different range values ℓ_x: 20, 50, and 80. A smaller range implies a more localized correlation, introducing local abrupt changes, thus increasing the complexity of the data. Table 6.1 summarizes the simulated datasets.

These datasets serve both regression and classification purposes. For classification tasks, categorical targets ($\mathbf{u}$) were derived from continuous targets ($\mathbf{t}$) by setting $u_i = 1$ if $t_i \geq \alpha$, $u_i = 0$ otherwise. Here, the threshold α is set as the median of the observations $\mathbf{t}$ to ensure data balance.

6.4.2 Training efficiency

In this section, we explore the training efficiency of GB-IDW and ML-IDW using gradient descent and backpropagation, comparing them to the conventional approach of training an IDW model via a search in a discretized grid. As previously noted, time complexity can be a significant challenge for conventional IDW, it varies polynomially with the number of possibilities per parameter m, and exponentially with the number of parameters. Fig. 6.1 illustrates different values of m versus the training time and the training error E^{LOOCV}, utilizing dataset 6 comprising 1000 data points. As anticipated, employing a finer grid leads to lower training error; however, the training time increases significantly.

Table 6.1 Geostatistical characteristics of the simulated datasets

Datasets	Variogram model	Size	Nugget	Partial Sill	Range	Anisotropy
dataset 1	Gaussian	100	4	20	20	$\ell_y = \ell_x/3$
dataset 2	Gaussian	100	4	20	50	$\ell_y = \ell_x/3$
dataset 3	Gaussian	100	4	20	80	$\ell_y = \ell_x/3$
dataset 4	Gaussian	1000	4	20	20	$\ell_y = \ell_x/3$
dataset 5	Gaussian	1000	4	20	50	$\ell_y = \ell_x/3$
dataset 6	Gaussian	1000	4	20	80	$\ell_y = \ell_x/3$
dataset 7	Gaussian	10000	4	20	20	$\ell_y = \ell_x/3$
dataset 8	Gaussian	10000	4	20	50	$\ell_y = \ell_x/3$
dataset 9	Gaussian	10000	4	20	80	$\ell_y = \ell_x/3$
dataset 10	Gaussian + Exponential	100	4	20	20	$\ell_y = \ell_x/3$
dataset 11	Gaussian + Exponential	100	4	20	50	$\ell_y = \ell_x/3$
dataset 12	Gaussian + Exponential	100	4	20	80	$\ell_y = \ell_x/3$
dataset 13	Gaussian + Exponential	1000	4	20	20	$\ell_y = \ell_x/3$
dataset 14	Gaussian + Exponential	1000	4	20	50	$\ell_y = \ell_x/3$
dataset 15	Gaussian + Exponential	1000	4	20	80	$\ell_y = \ell_x/3$
dataset 16	Gaussian + Exponential	10000	4	20	20	$\ell_y = \ell_x/3$
dataset 17	Gaussian + Exponential	10000	4	20	50	$\ell_y = \ell_x/3$
dataset 18	Gaussian + Exponential	10000	4	20	80	$\ell_y = \ell_x/3$

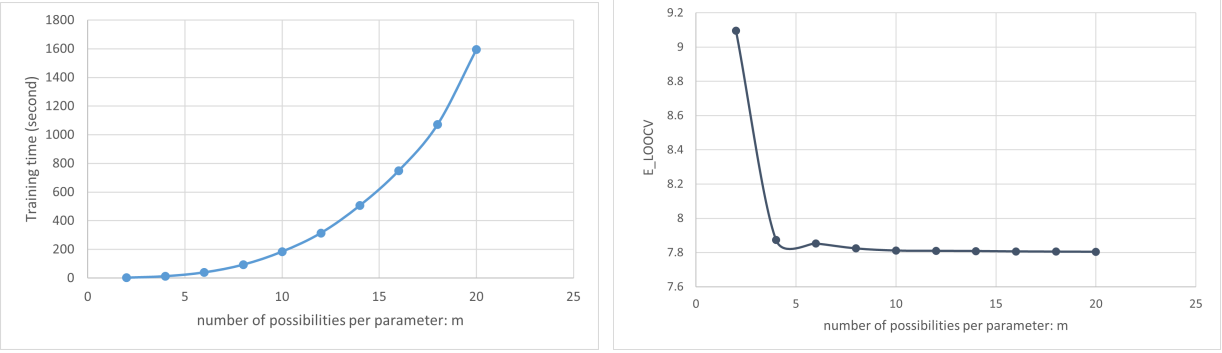
Table 6.2 presents the E^{LOOCV} values for IDW, GB-IDW, and ML-IDW, along with the number of gradient descent iterations, the grid size for conventional IDW, and the total training time. GB-IDW achieves a comparable training error in a shorter duration, while ML-IDW with 5 layers achieves a lower training error in a very short time. For larger datasets, such as those with 10,000 data points, the time complexity will significantly increase, making it exceedingly challenging to train conventional IDW within a reasonable timeframe while considering all three parameters $\{p, \ell_x, \ell_y\}$.

Table 6.2 Models efficiency

Model	Number of iterations $\lvert\mathcal{I}\rvert$ or grid size $\lvert G\rvert$	E^{LOOCV}	Training time (seconds)
Conventional IDW	$\lvert G\rvert = 20^3 = 8000$	7.804	1594.709
GB-IDW	$\lvert\mathcal{I}\rvert$ =1000	7.804	**12.49**
ML-IDW (5 layers)	$\lvert\mathcal{I}\rvert$ =1000	**6.680**	136.234

6.4.3 Results on regression tasks

To investigate model performance on regression tasks, we partitioned each dataset so that seventy percent of the data served as training data, while the remaining thirty percent served



(a) Training time vs grid size

(b) E^{LOOCV} error vs. grid size

Figure 6.1 (a) Training time of conventional IDW vs. number of possibilities per parameter. (b) E^{LOOCV} error vs. number of possibilities per parameter for conventional IDW.

as test data. We began with datasets containing 10,000 data points and monitored the prediction performance of Ordinary Kriging, Grad-IDW, and ML-IDW with 5 layers with Relu activation function (referred to as Deep-IDW) while varying complexity (specifically, varying the range).

Fig. 6.2 depicts, from the upper left figure to bottom right figure, the simulated data (dataset 7 characterized by high complexity with a range of 20), the training data, the test data, predictions from Ordinary Kriging, predictions from Grad-IDW, and predictions from Deep-IDW.

On datasets characterized by high complexity, Deep-IDW exhibits higher resolution, with predicted values closer to the ground truth values of the test data. Grad-IDW demonstrates lesser resolution. However, Ordinary Kriging appears very smooth and is unable to capture abrupt changes in the data.

Fig. 6.3a and Fig. 6.3b illustrate the Mean Square Error (MSE) for the three models concerning complexity respectively for datasets with simple and nested variogram models. It is evident that Deep-IDW excels in higher complexity scenarios (range=20), exhibiting lower MSE_{test}. However, the results are remarkably comparable in medium complexity situations (range=50), with Grad-IDW demonstrating the lowest test error in Fig 3a. On low complexity datasets (range=80), Ordinary Kriging performs slightly better.

The advantage of ML-IDW in this context lies in its customised nature: we can adjust the model complexity by adding or removing layers according to data complexity. It is worth noting that Grad-IDW is a variant of ML-IDW with no layers, making it a simpler alternative in certain scenarios.

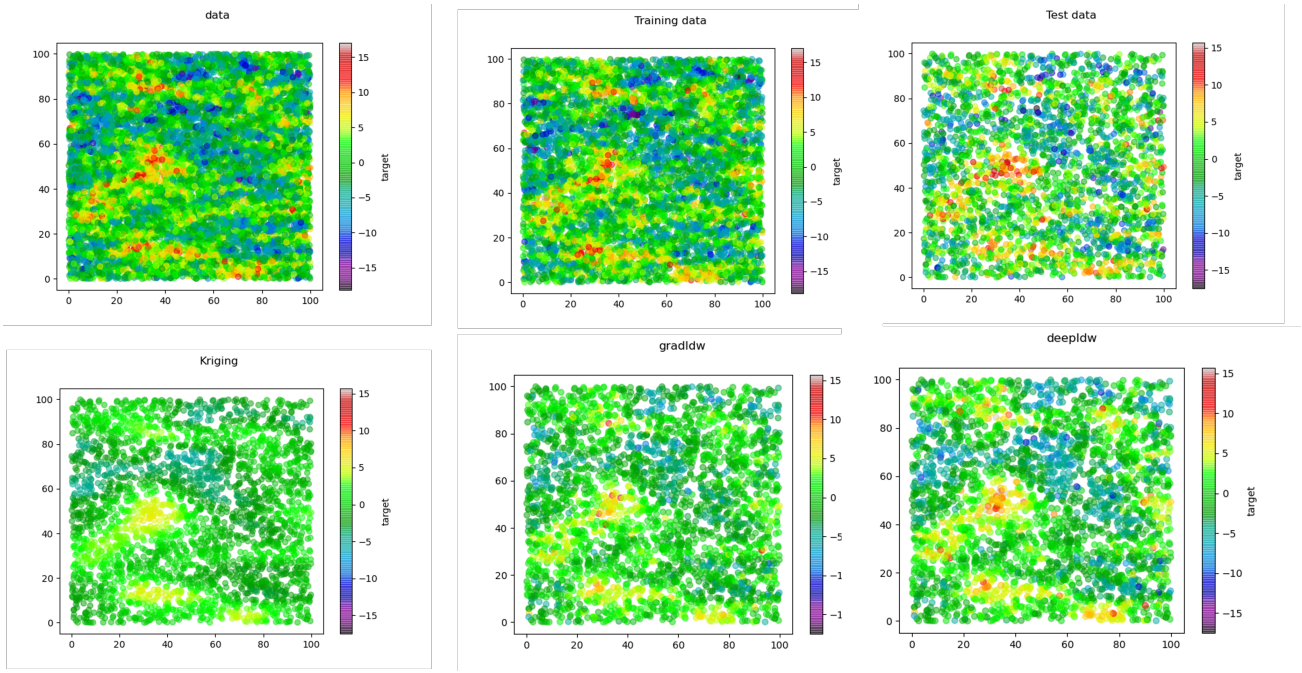


Figure 6.2 Top: Simulated dataset, training set, and test set. Bottom: Kriging, Grad-IDW, and Deep-IDW predictions.

Fig. 6.4a and Fig. 6.4b display the Mean Square Error (MSE) test results for the three models across different data sizes, specifically focusing on high complexity (range=20) and low complexity (range=80) scenarios, respectively.

Table 6.3 shows the MSE for all dataset. Observing the results while fixing complexity and varying data size, it is evident that the performance of all models declines as the data size decreases. Notably, Deep-IDW exhibits better performance than OK, and comaparable performance with Grad-IDW in high complexity scenarios, while Ordinary Kriging exhibits slightly superior performance in low complexity situations.

6.4.4 Results on classification

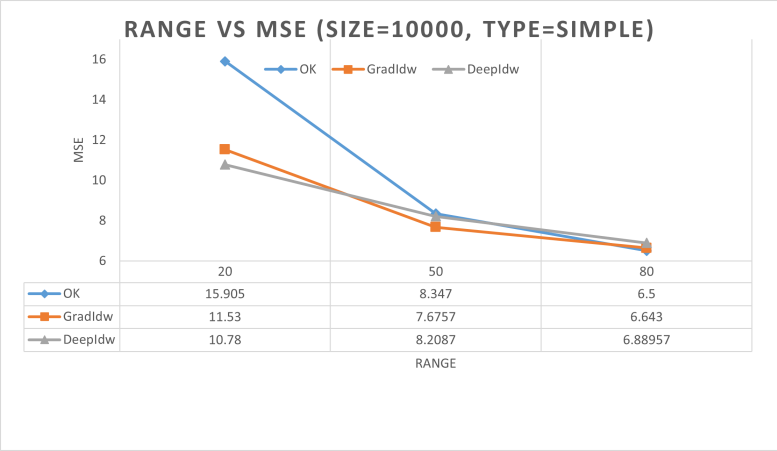
In the classification task, we compare a ML-IDW classifier with 5 layers, referred to as Deep-IDW-Classifier, against Indicator Kriging (IK) and Simplical Indicator Kriging (SIK) combined with Support Vector Machine for Classification (SVC). We employ the F1 score as the performance criterion (higher values indicating better performance) to evaluate the prediction capabilities on the test data. Similar to the regression task, we divide the data into training and test sets.

Fig. 6.5 illustrates the training and test data, along with the predictions of the three models

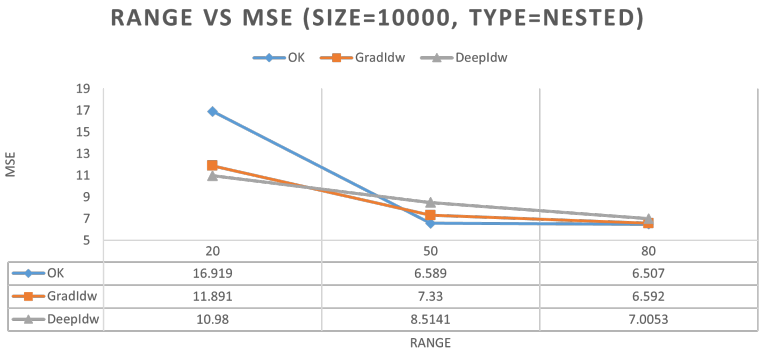
Table 6.3 The Mean Square Error test (MSE) for each model and each dataset.

Datasets	Complexity	OK	Grad-IDW	Deep-IDW (5 layers ML-IDW)
dataset 1	High	29.218	24.6919	**24.455**
dataset 2	Medium	**23.64497**	26.115776	26.46
dataset 3	Low	**16.72**	18.425	18.41
dataset 4	High	20.748	**16.26**	16.5218
dataset 5	Medium	**10.339**	12.695	12.443
dataset 6	Low	**8.6799**	9.9069	9.991
dataset 7	High	15.905	11.53	**10.78**
dataset 8	Medium	8.347	**7.6757**	8.2087
dataset 9	Low	**6.5**	6.643	6.88957
dataset 10	High	39.19	**36.139**	36.54
dataset 11	Medium	**22.829**	25.23613	25.397
dataset 12	Low	**15.705**	16.7	16.63
dataset 13	High	23.047	**17.45**	18.89
dataset 14	Medium	**9.6436**	13.549	14.19
dataset 15	Low	**8.133**	12.393	12.7872
dataset 16	High	16.919	11.891	**10.98**
dataset 17	Medium	**6.589**	7.33	8.5141
dataset 18	Low	**6.507**	6.592	7.0053

Note: Lower MSE indicates closer predictions to the ground truth. The **bold** values are the lowest for each dataset.



(a) Simple variogram



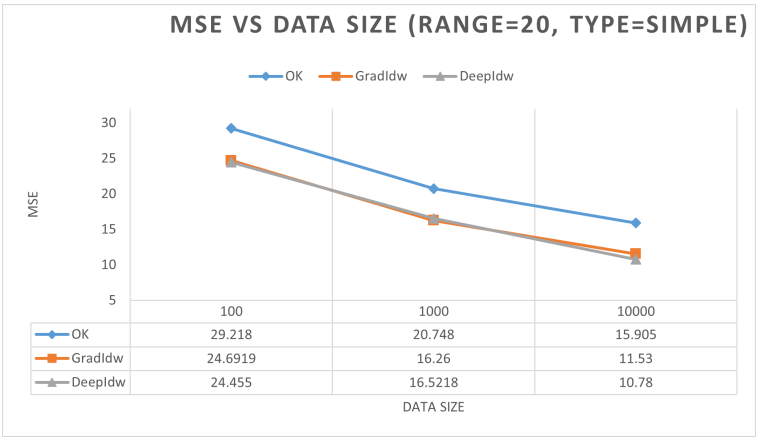
(b) Nested variogram

Figure 6.3 Mean Square Error for OK, Grad-IDW, and DeepIDW on 10,000-point dataset: (a) Simple variogram; (b) Nested variogram.

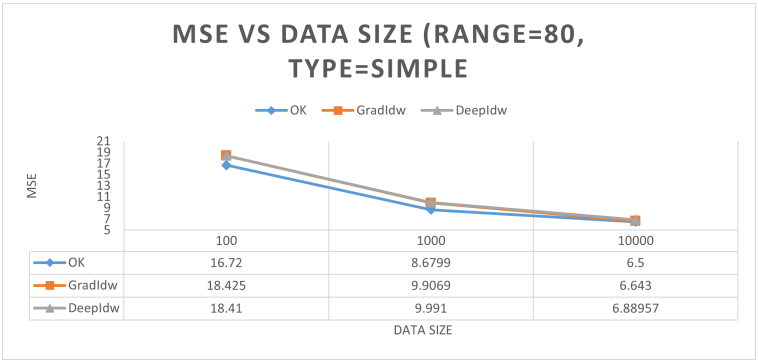
for dataset 7. It is evident that Deep-IDW can capture nonlinearities and abrupt changes in the data. Simplical Indicator Kriging combined with SVC also performs well, while Indicator Kriging struggles to mimic the behavior of the test set.

Fig. 6.6 presents the F1 score for each model with respect to complexity. In this case, Deep-IDW-Classifier outperforms Simplical Indicator Kriging with SVC and Indicator Kriging in high and medium complexity scenarios, whereas Indicator Kriging performs better in low complexity scenarios.

Table 6.4 summarizes the results for all datasets. We observe that Deep-IDW-Classifier excels in most cases and its results where at least comparable with the best classifier.



(a) High complexity



(b) Low complexity

Figure 6.4 Mean Square Error for OK, Grad-IDW, and DeepIDW vs. data size on a simple variogram model.

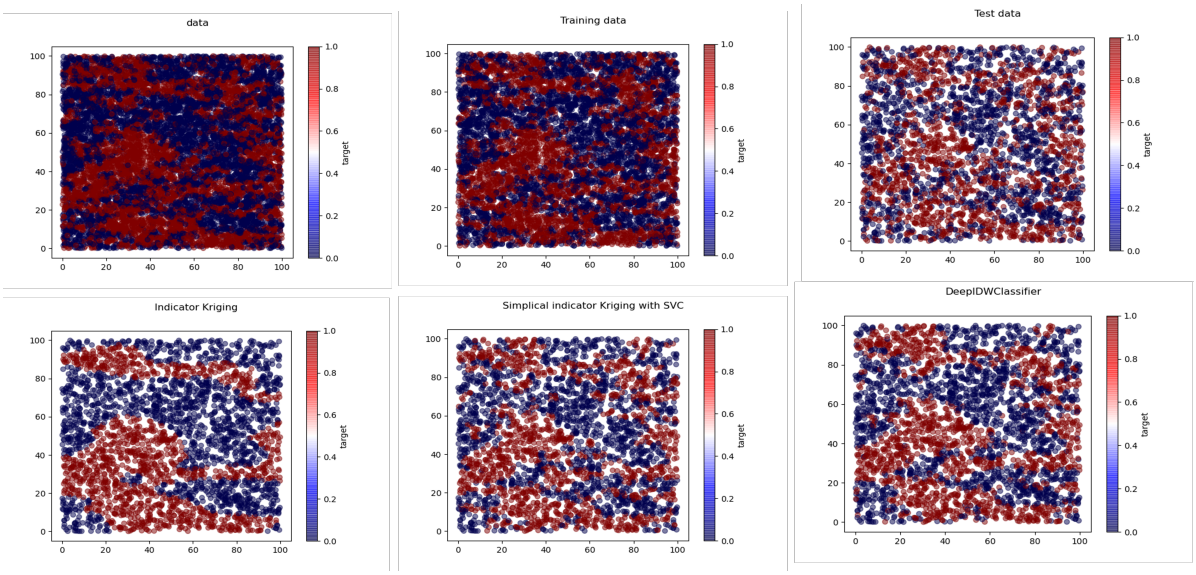


Figure 6.5 Top: Simulated dataset, training set, and test set for classification. Bottom: IK, Simplical IK with SVC, and Deep-IDW predictions.

6.5 Conclusion

In this study, we proposed the Multi-Layer Inverse Distance Weighting Model (ML-IDW), a spatial interpolation model inspired by both the formulation of multi-layer neural networks—as compositions of learnable functions—and the inverse distance weighting (IDW) approach, which utilizes an average weighted sum of inverse distance functions. The goal was to enhance spatial interpolation capabilities, enabling the model to learn from complex data without compromising the inherent spatial dependency structure.

Our results demonstrated that ML-IDW outperforms conventional geostatistical approaches, such as Ordinary Kriging, Indicator Kriging, and Simplical Indicator Kriging, in both regression and classification tasks when data complexity is high. In scenarios with lower complexity, the results were comparable. Notably, ML-IDW is a customizable model, allowing for the addition or removal of layers according to the data's complexity.

Furthermore, ML-IDW exhibited higher training efficiency compared to traditional IDW training methods. By utilizing gradient descent and backpropagation for training, the model is compatible with other neural network architectures.

For future work, we plan to stack ML-IDW neural network projection models to learn spatial dependencies on latent representations.

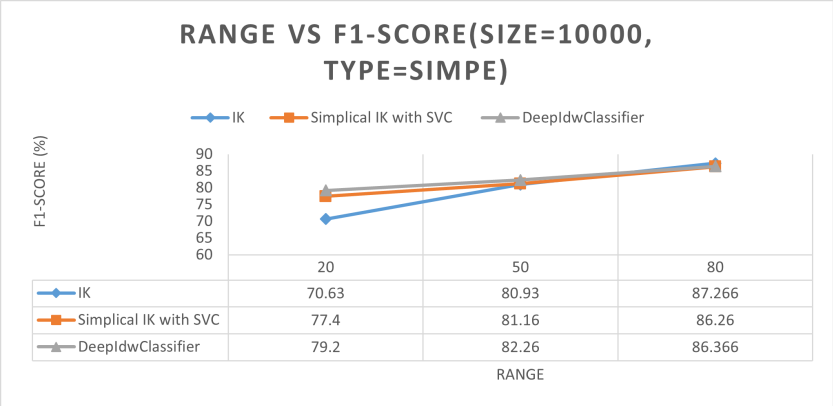


Figure 6.6 F1 score for IK, Simplical IK with SVC, and DeepIDW vs. data complexity on a 10,000-point simple variogram dataset.

Apport du chapitre à la thèse

Ce chapitre a exploré le développement des ML-IDW, une approche qui s'inspire des réseaux neuronaux et des modèles d'interpolation spatiale. L'étude démontre que ce modèle améliore significativement la précision des interpolations en exploitant des fonctions non linéaires apprenables et en optimisant les poids des relations spatiales. Les résultats obtenus confirment que cette approche surpasse les techniques classiques, notamment pour des jeux de données présentant des structures complexes. Cette avancée ouvre la voie à une utilisation plus large des réseaux neuronaux pour l'analyse des données spatiales en géosciences, tout en offrant des perspectives d'amélioration et d'adaptation à d'autres contextes miniers et géologiques.

Table 6.4 The F1 score in percentage for each model and each dataset.

Datasets	Complexity	IK	SIK with SVC	Deep-IDW-Classifier
dataset 1	High	50	56.66	**60**
dataset 2	Medium	**76.66**	53.33	**76.66**
dataset 3	Low	**63.33**	50	**63.33**
dataset 4	High	69.66	67	**70.33**
dataset 5	Medium	**78.66**	73.66	75.33
dataset 6	Low	**81**	78.33	79
dataset 7	High	70.63	77.4	**79.2**
dataset 8	Medium	80.93	81.16	**82.26**
dataset 9	Low	**87.266**	86.26	86.366
dataset 10	High	**80**	56.66	70
dataset 11	Medium	**66.66**	53.33	43.33
dataset 12	Low	**70**	60	40
dataset 13	High	50.33	66	**69.66**
dataset 14	Medium	**76.66**	76	76.33
dataset 15	Low	**83**	81.66	82.33
dataset 16	High	74.5	79.6	**80.2**
dataset 17	Medium	85.56	85.2	**85.66**
dataset 18	Low	87.766	89.23	**89.9**

Note: Higher F1 score indicates better classification. The **bold** values are the highest for each dataset.

CHAPITRE 7 CONCLUSION

7.1 Synthèse des travaux

Cette thèse vise à explorer et à développer des méthodes d'interpolation spatiale et d'apprentissage automatique pour analyser les données acquises lors du forage, soit les données MWD, afin d'améliorer la modélisation géologique et la délimitation des domaines rocheux. Chaque chapitre apporte des contributions spécifiques, répondant aux objectifs fixés pour optimiser les méthodes d'interpolation géospatiale et de reconnaissance des structures sous-jacentes en géoscience.

Au chapitre 4, une approche innovante a été mise en œuvre, combinant la rétropropagation à des techniques d'interpolation spatiale classiques pour répondre aux défis des données complexes. En intégrant des contraintes paramétriques, comme l'anisotropie géométrique, aux GPs et aux IDWs, les modèles développés permettent de mieux rendre compte des structures spatiales dans des environnements géoscientifiques. Le modèle GP-nested, une version sophistiquée du GP, a été conçu spécifiquement pour capter l'anisotropie zonale, un facteur géologique clé dans de nombreuses formations rocheuses. En affinant l'interpolation des données, ce modèle permet d'améliorer les prédictions des formations géologiques sous-jacentes.

Le modèle GBM-IDW-LHOCV, une variante du modèle IDW introduite dans ce chapitre, utilise la validation croisée *Leave-Hole-Out* (LHOCV), qui s'est avérée mieux adaptée que la méthode traditionnelle *Leave-One-Out* (LOOCV) pour les données de forage. En ignorant les contributions des points de prédiction dans le trou à prédire, LHOCV imite une régularisation de type *dropout*, bien connue dans les réseaux neuronaux, et assure ainsi une prédiction plus robuste et réaliste. Cette méthode favorise une continuité spatiale en réduisant l'influence des points dans un même trou de forage, offrant ainsi une prédiction plus conforme à un contexte réel où les trous de forage peuvent être indépendants. Par ailleurs, en réduisant les influences des observations atypiques, GBM-IDW-LHOCV améliore les prédictions dans des contextes où des points de mesure rares ou isolés peuvent fausser les résultats, renforçant ainsi sa pertinence dans des applications géospatiales complexes.

Dans la continuité, le chapitre 5 propose une approche de délimitation des massifs rocheux basée sur les données MWD pour délimiter des domaines spatiaux homogènes en dureté de roche. Cette technique utilise un empilement de VAE couplé à un modèle d'interpolation pour extraire une caractéristique latente, le pseudo-BI, qui représente la dureté de la roche.

L'intégration du pseudo-BI dans le processus de clustering a permis de recréer les tendances générales des paramètres mécaniques, comme le pseudo-ROP, pseudo-WOB, pseudo-RPM, et pseudo-TRQ, tout en atténuant la variabilité excessive et les observations atypiques. Le pseudo-BI répond à un besoin d'amélioration de la précision géologique des modèles en ajustant les *clusters* de manière à éviter les ambiguïtés et les chevauchements entre eux, créant ainsi des délimitations de domaines plus fiables.

Une découverte majeure a été la capacité du *clustering* basé sur le pseudo-BI à détecter efficacement les zones de collerette, où la séparation géologique est souvent complexe et floue. En effet, cette approche a surpassé les méthodes traditionnelles basées sur l'indice BI pour identifier et exclure ces zones, offrant une amélioration marquée de la détection et de la séparation spatiale. En ajustant les modèles de clustering pour qu'ils prennent en compte la contiguïté spatiale et en excluant les zones de collerette, les modèles basés sur le pseudo-BI permettent de mieux comprendre la structure sous-jacente des formations rocheuses. Ces résultats démontrent que cette méthodologie peut s'étendre avec succès en 3D et fournir des *insights* géologiques qui correspondent aux observations faites par les géologues.

En utilisant des autoencodeurs variés (FC, Conv, RBF) pour obtenir des paramètres pseudo-optimaux, nous avons démontré la flexibilité de cette approche en fonction des propriétés spécifiques des données. Cette capacité à adapter les architectures aux données spécifiques répond aux objectifs de la thèse, en offrant une méthode adaptable pour la modélisation des massifs rocheux et l'identification de lithologies correspondantes dans des environnements de forage.

Enfin au chapitre 6, nous avons développé un modèle d'interpolation par pondération multi-couches, appelé le *Multi-Layer IDW* (ML-IDW), inspiré par les réseaux de neurones multicouches et les techniques géostatistiques traditionnelles comme l'*Indicator Kriging* et le *Simplical Indicator Kriging*. Le ML-IDW permet une personnalisation selon la complexité des données, ce qui le rend pertinent pour des environnements de données variés allant de la simple à la haute complexité. Cette approche a permis de surmonter plusieurs limitations des modèles IDW traditionnels, notamment en ajoutant une structure multicouche qui améliore la capacité d'interpolation et permet au modèle de mieux s'adapter aux nuances géologiques spécifiques.

En utilisant la rétropropagation et la descente de gradient pour ajuster les paramètres de ML-IDW, le modèle présente une compatibilité unique avec d'autres architectures de réseaux neuronaux, renforçant ainsi la flexibilité et l'adaptabilité de l'interpolation spatiale dans les contextes géologiques. ML-IDW, tout en maintenant la dépendance spatiale, offre des résultats comparables à ceux de méthodes plus traditionnelles comme le Krigeage ordinaire,

en particulier lorsque la complexité des données est plus élevée — variogramme imbriqué à faibles portées, indiquant une hétérogénéité marquée à courte distance et des structures multiéchelles. Les performances accrues de ce modèle en termes d'efficacité d'apprentissage en font un outil prometteur pour la modélisation géologique dans les environnements d'exploitation minière.

En somme, les méthodes développées au sein de cette thèse répondent pleinement aux objectifs en améliorant la précision des approches d'interpolation géologique, en intégrant des techniques d'apprentissage profond et en adaptant les modèles d'interpolation spatiale classiques pour répondre aux exigences des données MWD. Les modèles GBM-IDW-LHOCV et ML-IDW, combinés aux techniques d'autoencodeurs, offrent une représentation géologique plus fidèle et une capacité accrue de prédiction de la dureté des roches. Ces innovations permettent d'améliorer les connaissances des formations géologiques, même dans des environnements de haute variabilité, en garantissant des domaines spatiaux continus et une robustesse face aux observations atypiques. En facilitant l'identification de zones géologiquement homogènes et en offrant une représentation spatiale fiable, cette recherche fournit des outils précieux pour les applications minières, répondant aux objectifs de la thèse d'une manière intégrée et innovante qui combine les techniques géostatistiques et l'intelligence artificielle appliquée au contexte géoscientifique.

7.2 Limitations de la solution proposée

Les solutions proposées dans cette thèse, bien que novatrices et performantes pour la délimitation des domaines rocheux et l'interpolation spatiale en géosciences, présentent certaines limitations qui méritent d'être discutées afin de contextualiser les résultats obtenus et de guider les recherches futures.

Tout d'abord, bien que le BI soit largement utilisé pour quantifier la dureté de la roche, cette mesure peut être influencée par des variabilités indépendantes des caractéristiques intrinsèques de la roche, ce qui en limite l'exactitude. En effet, des facteurs externes, comme le paramétrage du patron de forage ou encore les ajustements de l'opérateur, introduisent des biais dans le BI, ce qui affecte et perturbe la corrélation spatiale. Par exemple, des fluctuations dans la vitesse de rotation, la pression exercée, ou d'autres ajustements non standard modifient les valeurs de BI enregistrées, bien que ces ajustements ne soient pas directement liés à la lithologie. Cette sensibilité rend difficile la création de corrélations spatiales précises, car les indices enregistrés peuvent ne pas refléter la vraie dureté de la roche. Ainsi, des interprétations fondées sur le BI risquent d'être biaisées par des paramètres extérieurs au matériau rocheux, ce qui compromet la validité des modèles géologiques basés sur ce paramètre.

De plus, l'application de modèles de type autoencodeurs (AE) dans le cadre du domaining rocheux peut être limitée par la complexité des structures géologiques réelles, notamment dans les environnements hétérogènes ou à forte anisotropie spatiale. Bien que les modèles développés soient capables d'apprendre des caractéristiques latentes pertinentes, telles que le pseudo-BI pour la dureté des roches, cette représentation latente dépend fortement des données d'entraînement. En présence de données rares ou incomplètes, typiques des environnements de forage, la qualité de l'apprentissage peut en être affectée, limitant ainsi la précision et la fiabilité des domaines spatiaux générés. Par ailleurs, l'apprentissage supervisé, bien que performant, peut impliquer un biais en fonction des paramètres sélectionnés et de la qualité de l'échantillonnage, ce qui pourrait influencer les *clusters* identifiés et donc fausser l'interprétation géologique.

Les modèles d'interpolation multicouches, tels que ML-IDW, bien qu'efficaces pour l'adaptation à la complexité croissante des données, augmentent la demande computationnelle en raison de la nécessité d'une optimisation par rétropropagation sur plusieurs couches. Cela peut poser un défi pour les entreprises minières et géologiques ne disposant pas de ressources informatiques suffisantes ou de compétences en modélisation avancée. L'intégration de cette approche dans un flux de travail standard d'exploration ou d'exploitation minière peut ainsi exiger un investissement substantiel en infrastructure et en formation, ce qui pourrait freiner son adoption.

Enfin, bien que les techniques basées sur l'apprentissage profond et les méthodes d'interpolation avancées permettent de mieux capter la variabilité géologique et de traiter des données bruitées, elles peuvent introduire des biais dans les prédictions si elles sont mal calibrées. En effet, l'intégration de ces méthodes dépend de la qualité des données d'entrée MWD et des choix de paramétrage. Si les données d'entrée présentent des biais intrinsèques, par exemple dus à des erreurs de capteur ou à des variations de calibration de la machine de forage, cela pourrait altérer la qualité des prédictions et réduire la fiabilité des domaines générés.

En conclusion, bien que les solutions proposées offrent des avancées importantes en termes de précision et de robustesse dans la modélisation géologique, elles restent sensibles aux limitations inhérentes à l'apprentissage automatique, à la qualité des données, et à la complexité computationnelle. Il est donc essentiel de considérer ces limitations dans une optique d'amélioration continue, en développant des méthodologies adaptées aux contextes géologiques spécifiques et en explorant des solutions hybrides ou simplifiées pour permettre une adoption plus large dans l'industrie minière.

7.3 Améliorations futures

Pour répondre aux limitations identifiées et renforcer l'efficacité des méthodes proposées, plusieurs pistes de recherche future peuvent être envisagées. Premièrement, l'intégration de l'apprentissage par transfert pourrait améliorer la robustesse des modèles en permettant l'utilisation de connaissances acquises dans d'autres contextes géologiques pour enrichir les prédictions. Par exemple, des modèles entraînés sur des données provenant de sites géologiquement similaires pourraient être adaptés et utilisés pour des forages en cours ou de nouveaux sites présentant des caractéristiques géologiques analogues. On pourrait également utiliser l'information géologique et des données de forage provenant des niveaux supérieurs où les blocs de minerai ont été extraits. L'apprentissage par transfert permettrait ainsi de réduire le besoin de données d'entraînement étendues et spécifiques, rendant les modèles plus efficaces et plus rapides à déployer dans divers contextes miniers.

En outre, l'amélioration de la scalabilité des modèles par l'intégration de techniques d'optimisation et de simplification, comme le sous-échantillonnage intelligent et la réduction dimensionnelle avancée, pourrait significativement alléger le coût computationnel des calculs d'interpolation spatiale pour des bases de données volumineuses. L'exploration de techniques hybrides combinant des modèles classiques (ÌDW, Krigeage, GP) et des réseaux neuronaux profonds pourrait également offrir un compromis entre précision et efficacité. Par exemple, utiliser un modèle classique pour capter la tendance globale et réserver des architectures neuronales aux zones les plus variables : (i) estimer une surface de tendance à large échelle (p. ex., krigeage universel) ; (ii) mesurer la variabilité locale (écart-type/variance de prédiction, résidus, portées courtes du variogramme) et définir un masque de "zones à forte variabilité géologique" ; (iii) n'appliquer des modèles plus complexes (réseaux neuronaux compatibles avec la structure spatiale) que dans ces zones ciblées ; (iv) fusionner les sorties en conservant la tendance globale dans les zones stables et en corrigeant localement dans les zones complexes. Ce découplage permet d'optimiser les ressources tout en préservant une précision élevée là où elle est la plus nécessaire.

Enfin, des architectures avancées et complémentaires peuvent enrichir l'interpolation spatiale. D'une part, les réseaux de neurones sur graphes (p. ex., *Graph Convolutional Networks*) représentent explicitement la connectivité spatiale en construisant un graphe $G = (V, E)$ où les nœuds sont des échantillons MWD (ou des blocs) et où les arêtes relient des voisins proches avec des pondérations anisotropes w_{ij} basées sur une distance géologique (portées différentes selon les directions). La propagation de messages fait circuler l'information au sein du massif et capture les corrélations locales tout en respectant les discontinuités. D'autre part, des modèles de type *Transformer* permettent de traiter des dépendances non locales via un

mécanisme d'attention muni d'encodages positionnels relatifs (différences de coordonnées, distances, azimut/dip). En pratique, ces briques peuvent se combiner aux autoencodeurs (qui apprennent un code latent robuste des signaux MWD) et aux schémas d'interpolation existants : le code latent est diffusé spatialement par le graphe ou l'attention, puis décodé en cartes de propriétés (p. ex., indices de dureté) avec un terme de lissage contrôlé par la géométrie (anisotropie, voisinage). Ce couplage vise à mieux capter simultanément (i) la structure à courte portée (continuité locale) et (ii) les relations à longue portée (structures multiéchelles), afin d'accroître la robustesse et la précision dans des contextes géologiques complexes.

RÉFÉRENCES

[1] I. C. C. Acosta *et al.*, "A machine learning framework for drill-core mineral mapping using hyperspectral and high-resolution mineralogical data fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4829–4842, 2019.

[2] M. Ahmadi *et al.*, "Comparison of kriging, machine learning algorithms and classical thermodynamics for correlating the formation conditions for co2 gas hydrates and semi-clathrates," *Journal of Natural Gas Science and Engineering*, vol. 84, p. 103659, 2020.

[3] A. D. Ajak, E. Lilford, and E. Topal, "Application of predictive data mining to create mine plan flexibility in the face of geological uncertainty," *Resources Policy*, vol. 55, pp. 62–79, 2018.

[4] O. Akyildiz, H. Basarir, and S. L. Ellefmo, "The development of a lithology prediction model using measurement while drilling data in a quartzite quarry," *International Journal of Mining, Reclamation and Environment*, pp. 1–17, 2024.

[5] A. Alsaihati *et al.*, "Use of machine learning and data analytics to detect downhole abnormalities while drilling horizontal wells, with real case study," *Journal of Energy Resources Technology*, vol. 143, no. 4, p. 043201, 2021.

[6] H. Amini *et al.*, "Evaluation of flyrock phenomenon due to blasting operation by support vector machine," *Neural Computing and Applications*, vol. 21, pp. 2077–2085, 2012.

[7] A. Aryafar *et al.*, "Heavy metal pollution assessment using support vector machine in the shur river, sarcheshmeh copper mine, iran," *Environmental earth sciences*, vol. 67, pp. 1191–1199, 2012.

[8] E. Asa, "Machine learning characterization of a two-seam coal deposit," *Journal of Mining Science*, vol. 47, pp. 761–770, 2011.

[9] Y. Azimi *et al.*, "Prediction of the blastability designation of rock masses using fuzzy sets," *International Journal of Rock Mechanics and Mining Sciences*, vol. 47, no. 7, pp. 1126–1140, oct 2010.

[10] M. Babaei Khorzoughi, "Use of measurement while drilling techniques for improved rock mass characterization in open-pit mines," Ph.D. dissertation, University of British Columbia, 2013.

[11] M. Babaei Khorzoughi, "Use of measurement while drilling techniques for improved rock mass characterization in open-pit mines," Ph.D. dissertation, University of

British Columbia, 2013. [Online]. Available : https://open.library.ubc.ca/collections/ubctheses/24/items/1.0073823

[12] O. Babak and C. V. Deutsch, "Statistical approach to inverse distance interpolation," *Stochastic Environmental Research and Risk Assessment*, vol. 23, no. 5, pp. 543–553, mar 2008.

[13] O. Babak and C. V. Deutsch, "Statistical approach to inverse distance interpolation," *Stochastic Environmental Research and Risk Assessment*, vol. 23, pp. 543–553, 2009.

[14] J. Baek and Y. Choi, "Deep neural network for ore production and crusher utilization prediction of truck haulage system in underground mine," *Applied Sciences*, vol. 9, no. 19, p. 4180, 2019.

[15] J. Baek and Y. Choi, "Deep neural network for predicting ore production by truck-haulage systems in open-pit mines," *Applied Sciences*, vol. 10, no. 5, p. 1657, 2020.

[16] D. H. Ballard, "Modular learning in neural networks," in *Proceedings of the sixth National Conference on artificial intelligence-volume 1*, 1987, pp. 279–284.

[17] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *Machine learning for data science handbook : data mining and knowledge discovery handbook*, pp. 353–374, 2023.

[18] D. Bansal, D. J. Evans, and B. Jones, "A real-time predictive maintenance system for machine systems-an alternative to expensive motion sensing technology," in *2005 Sensors for Industry Conference.* IEEE, 2005, pp. 39–44.

[19] P. M. S. Basnet, S. Mahtab, and A. Jin, "A comprehensive review of intelligent machine learning based predicting methods in long-term and short-term rock burst prediction," *Tunnelling and Underground Space Technology*, vol. 142, p. 105434, 2023.

[20] F. Beretta *et al.*, "Automated lithological classification using uav and machine learning on an open cast mine," *Applied Earth Science*, vol. 128, no. 3, pp. 79–88, 2019.

[21] C. L. Bérubé *et al.*, "Predicting rock type and detecting hydrothermal alteration using machine learning and petrophysical properties of the canadian malartic ore and host rocks, pontiac subprovince, québec, canada," *Ore Geology Reviews*, vol. 96, pp. 130–145, 2018.

[22] G. D. Betrie *et al.*, "Predicting copper concentrations in acid mine drainage : a comparative analysis of five machine learning techniques," *Environmental monitoring and assessment*, vol. 185, pp. 4171–4182, 2013.

[23] A. Bewley and B. Upcroft, "Background appearance modeling with applications to visual object detection in an open-pit mine," *Journal of Field Robotics*, vol. 34, no. 1, pp. 53–73, 2017.

[24] C. M. Bishop, *Neural networks for pattern recognition.* Oxford university press, 1995.

[25] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[26] G. Bohling, "Introduction to geostatistics and variogram analysis," *Kansas geological survey*, vol. 1, no. 10, pp. 1–20, 2005.

[27] M. Boullé, "Predicting dangerous seismic events in coal mines under distribution drift," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2016, pp. 221–224.

[28] X.-N. Bui *et al.*, "Estimating pm10 concentration from drilling operations in open-pit mines using an assembly of svr and pso," *Applied Sciences*, vol. 9, no. 14, p. 2806, 2019.

[29] X.-N. Bui *et al.*, "Prediction of slope failure in open-pit mines using a novel hybrid artificial intelligence model based on decision tree and evolution algorithm," *Scientific reports*, vol. 10, no. 1, p. 9939, 2020.

[30] S. Castric *et al.*, "Modeling pollutant emissions of diesel engine based on kriging models : a comparison between geostatistic and gaussian process approach," *IFAC Proceedings Volumes*, vol. 45, no. 6, pp. 1708–1715, 2012.

[31] I. Chahrour and J. Wells, "Comparing machine learning and interpolation methods for loop-level calculations," *SciPost Physics*, vol. 12, no. 6, p. 187, 2022.

[32] S. Chandar *et al.*, "Correlational neural networks," *Neural computation*, vol. 28, no. 2, pp. 257–285, 2016.

[33] W. Chao *et al.*, "A prediction of borehole stability while drilling preliminary prospecting wells based on seismic impedance," *Petroleum Exploration and Development*, vol. 42, no. 3, pp. 427–433, 2015.

[34] L. Chen *et al.*, "A comparative assessment of geostatistical, machine learning, and hybrid approaches for mapping topsoil organic carbon content," *ISPRS International Journal of Geo-Information*, vol. 8, no. 4, p. 174, 2019.

[35] X. Cheng *et al.*, "Bilstm-based deep neural network for rock-mass classification prediction using depth-sequence mwd data : A case study of a tunnel in yunnan, china," *Applied Sciences*, vol. 13, no. 10, p. 6050, 2023.

[36] P. Chi *et al.*, "Reconstruction of 3d digital rocks with controllable porosity using cvae-gan," *Geoenergy Science and Engineering*, vol. 230, p. 212264, 2023.

[37] D. Cho, J. Im, and S. Jung, "A new statistical downscaling approach for short-term forecasting of summer air temperatures through a fusion of deep learning and spatial interpolation," *Quarterly Journal of the Royal Meteorological Society*, vol. 150, no. 760, pp. 1222–1242, 2024.

[38] K. Choi and K. Chong, "Modified Inverse Distance Weighting Interpolation for Particulate Matter Estimation and Mapping," *Atmosphere*, vol. 13, no. 5, p. 846, may 2022.

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

[40] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 2015.

[41] T. Cui, D. Pagendam, and M. Gilfedder, "Gaussian process machine learning and kriging for groundwater salinity interpolation," *Environmental Modelling & Software*, vol. 144, p. 105170, 2021.

[42] D. Curry *et al.*, "Technical limit specific energy—an index to facilitate drilling performance evaluation," in *SPE/IADC Drilling Conference and Exhibition*. SPE, 2005, pp. SPE–92 318.

[43] T. D'Angelo *et al.*, "Deep learning-based object detection for digital inspection in the mining industry," in *2019 18th ieee international conference on machine learning and applications (ICMLA)*. IEEE, 2019, pp. 633–640.

[44] S. De Iaco, D. T. Hristopulos, and G. Lin, "Geostatistics and machine learning," *Mathematical Geosciences*, vol. 54, no. 3, pp. 459–465, 2022.

[45] J. F. Díez-Pastor *et al.*, "Machine learning algorithms applied to r aman spectra for the identification of variscite originating from the mining complex of g avà," *Journal of Raman Spectroscopy*, vol. 51, no. 9, pp. 1563–1574, 2020.

[46] H. Ding *et al.*, "Prediction of pillar stability for underground mines using the stochastic gradient boosting technique," *IEEE Access*, vol. 6, pp. 69 253–69 264, 2018.

[47] S. Du *et al.*, "A new machine-learning prediction model for slope deformation of an open-pit mine : an evaluation of field data," *Energies*, vol. 12, no. 7, p. 1288, 2019.

[48] M. Esmaeili *et al.*, "Application of PCA, SVR, and ANFIS for modeling of rock fragmentation," *Arabian Journal of Geosciences*, vol. 8, no. 9, pp. 6881–6893, oct 2014.

[49] Q. Fang *et al.*, "Prediction of blast-induced ground vibration in open-pit mines using a new technique based on imperialist competitive algorithm and m5rules," *Natural Resources Research*, vol. 29, no. 2, pp. 791–806, 2020.

[50] G. L. Finfinger, *A methodology for determining the character of mine roof rocks*. West Virginia University, 2003.

[51] H. Gamal *et al.*, "Rock strength prediction in real-time while drilling employing random forest and functional network techniques," *Journal of Energy Resources Technology*, vol. 143, no. 9, p. 093004, 2021.

[52] C. Gan *et al.*, "Spatial estimation for 3d formation drillability field : A new modeling framework," *Journal of Natural Gas Science and Engineering*, vol. 84, p. 103628, Dec. 2020.

[53] C. Gan *et al.*, "A new spatial modeling method for 3D formation drillability field using fuzzy c-means clustering and random forest," *Journal of Petroleum Science and Engineering*, vol. 200, p. 108371, may 2021.

[54] Y. Geng *et al.*, "Seismic events prediction using deep temporal convolution networks," *Journal of Electrical and Computer Engineering*, vol. 2019, no. 1, p. 7343784, 2019.

[55] J. M. Gernand, "Machine learning classification models for more effective mine safety inspections," in *ASME International Mechanical Engineering Congress and Exposition*, vol. 46637. American Society of Mechanical Engineers, 2014, p. V014T08A020.

[56] S. Glubokovskikh *et al.*, "Machine learning algorithms for real-time prediction of the sonic logs based on drilling parameters and downhole accelerometers," in *SEG International Exposition and Annual Meeting*. SEG, 2020, p. D031S049R008.

[57] G. Gong, S. Mattevada, and S. E. O'Bryant, "Comparison of the accuracy of kriging and idw interpolations in estimating groundwater arsenic concentrations in texas," *Environmental research*, vol. 130, pp. 59–69, 2014.

[58] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[59] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[60] Q. Guan, P. C. Kyriakidis, and M. F. Goodchild, "A parallel computing approach to fast geostatistical areal interpolation," *International Journal of Geographical Information Science*, vol. 25, no. 8, pp. 1241–1267, 2011.

[61] H. Guo *et al.*, "Deep neural network and whale optimization algorithm to assess flyrock induced by blasting," *Engineering with Computers*, vol. 37, pp. 173–186, 2021.

[62] I. Gupta *et al.*, "Looking ahead of the bit using surface drilling and petrophysical data : Machine-learning-based real-time geosteering in volve field," *SPE Journal*, vol. 25, no. 02, pp. 990–1006, 2020.

[63] Y. Hajlaoui *et al.*, "Backpropagation-based inference for spatial interpolation to estimate the blastability index in an open pit mine," *Computers & Geosciences*, p. 105756, 2024.

[64] A. N. Hasan and B. Twala, "Improving single classifiers prediction accuracy for underground water pump station in a gold mine using ensemble techniques," in *IEEE EUROCON 2015-International Conference on Computer as a Tool (EUROCON)*. IEEE, 2015, pp. 1–7.

[65] A. N. Hasan, B. Twala, and T. Marwala, "Moving towards accurate monitoring and prediction of gold mine underground dam levels," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 2844–2849.

[66] M. Hasanipanah *et al.*, "Feasibility of PSO–ANFIS model to estimate rock fragmentation produced by mine blasting," *Neural Computing and Applications*, vol. 30, no. 4, pp. 1015–1024, Dec. 2016.

[67] M. Hasanipanah *et al.*, "Feasibility of pso–anfis model to estimate rock fragmentation produced by mine blasting," *Neural Computing and Applications*, vol. 30, pp. 1015–1024, 2018.

[68] M. Hasanipanah *et al.*, "Prediction of air-overpressure caused by mine blasting using a new hybrid pso–svr model," *Engineering with Computers*, vol. 33, pp. 23–31, 2017.

[69] K. He *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[70] M. He *et al.*, "Rockburst laboratory tests database—application of data mining techniques," *Engineering Geology*, vol. 185, pp. 116–130, 2015.

[71] J. Heaton, "Ian goodfellow, yoshua bengio, and aaron courville : Deep learning : The mit press, 2016, 800 pp, isbn : 0262035618," *Genetic programming and evolvable machines*, vol. 19, no. 1, pp. 305–307, 2018.

[72] I. Higgins *et al.*, "beta-vae : Learning basic visual concepts with a constrained variational framework." *ICLR (Poster)*, vol. 3, 2017.

[73] J. Hoffimann *et al.*, "Geostatistical learning : Challenges and opportunities," *Frontiers in Applied Mathematics and Statistics*, vol. 7, p. 689393, 2021.

[74] S. B. Hood, M. J. Cracknell, and M. F. Gazley, "Linking protolith rocks to altered equivalents by combining unsupervised and supervised machine learning," *Journal of Geochemical Exploration*, vol. 186, pp. 270–280, 2018.

[75] F. Huang *et al.*, "Opencl implementation of a parallel universal kriging algorithm for massive spatial data interpolation on heterogeneous systems," *ISPRS international journal of geo-information*, vol. 5, no. 6, p. 96, 2016.

[76] Z. Huang, H. Wang, and R. Zhang, "An improved kriging interpolation technique based on SVM and its recovery experiment in oceanic missing data," *American Journal of Computational Mathematics*, vol. 2, no. 1, pp. 56–60, Mar. 2012.

[77] M. N. Ikechukwu *et al.*, "Accuracy assessment and comparative analysis of idw, spline and kriging in spatial interpolation of landform (topography) : an experimental study," *Journal of Geographic Information System*, vol. 9, no. 3, pp. 354–371, 2017.

[78] S. Jiang *et al.*, "Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope," *Complexity*, vol. 2018, no. 1, p. 1048756, 2018.

[79] X. Jiang *et al.*, "Quantitative evaluation of mining geo-environmental quality in northeast china : comprehensive index method and support vector machine models," *Environmental earth sciences*, vol. 73, pp. 7945–7955, 2015.

[80] C. L. Jimeno *et al.*, "Drilling änd blasting of rocks," *USA CRS Press*, vol. 41, p. 35947, 1995.

[81] B. Jo and R. M. A. Khan, "An internet of things system for underground mine air quality pollutant prediction based on azure machine learning," *Sensors*, vol. 18, no. 4, p. 930, 2018.

[82] A. Kadkhodaie-Ilkhchi *et al.*, "Rock recognition from mwd data : a comparative study of boosting, neural networks, and fuzzy logic," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 680–684, 2010.

[83] A. Kale *et al.*, "Methodology for optimizing operational performance and life management of drilling systems using real time-data and predictive analytics," in *SPE Digital Energy Conference and Exhibition*. SPE, 2015, p. D021S009R003.

[84] M. Kamran, N. M. Shahani, and D. J. Armaghani, "Decision support system for underground coal pillar stability using unsupervised and supervised machine learning approaches," *Geomechanics and Engineering*, vol. 30, no. 2, pp. 107–121, 2022.

[85] U. E. Kaplan and E. Topal, "A new ore grade estimation using combine machine learning algorithms," *Minerals*, vol. 10, no. 10, p. 847, 2020.

[86] R. Karami and P. Afzal, "Estimation of elemental distributions by combining artificial neural network and inverse distance weighted (idw) based on lithogeochemical data in kahang porphry deposit, central iran," *Universal Journal of Geoscience*, vol. 3, no. 2, pp. 59–65, 2015.

[87] A. V. Kashnikov and L. Levin, "Applying machine learning techniques to mine ventilation control systems," in *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*. IEEE, 2017, pp. 391–393.

[88] M. Khandelwal and D. J. Armaghani, "Prediction of Drillability of Rocks with Strength Properties Using a Hybrid GA-ANN Technique," *Geotechnical and Geological Engineering*, vol. 34, no. 2, pp. 605–620, Dec. 2015.

[89] M. Khandelwal and M. Monjezi, "Prediction of backbreak in open-pit blasting operations using the machine learning method," *Rock mechanics and rock engineering*, vol. 46, pp. 389–396, 2013.

[90] A. Khorolskyi *et al.*, "Models and methods to make decisions while mining production scheduling," *Mining of Mineral Deposits*, vol. 13, no. 4, pp. 53–62, 2019.

[91] J. Kim *et al.*, "A comparative study of machine learning and spatial interpolation methods for predicting house prices," *Sustainability*, vol. 14, no. 15, p. 9056, 2022.

[92] J. Kim and C. Deutsch, "Calculation and modeling of variogram anisotropy," *Geostatistics Lessons. Retrieved from https ://geostatisticslessons. com/lessons/variogramanisotropy*, 2022.

[93] R. L. King, M. A. Hicks, and S. P. Signer, "Using unsupervised learning for feature detection in a coal mine roof," *Engineering Applications of Artificial Intelligence*, vol. 6, no. 6, pp. 565–573, 1993.

[94] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv :1312.6114*, 2013.

[95] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations (ICLR)*, 2014.

[96] N. Klyuchnikov *et al.*, "Data-driven model for the identification of the rock type at a drilling bit," *Journal of Petroleum science and Engineering*, vol. 178, pp. 506–516, 2019.

[97] P. Kolapo, "Investigating the effects of mechanical properties of rocks on specific energy and penetration rate of borehole drilling," *Geotechnical and Geological Engineering*, vol. 39, no. 2, pp. 1715–1726, 2021.

[98] D. LaBelle, *Lithological classification by drilling.* Carnegie Mellon University, The Robotics Institute, 2001.

[99] D. LaBelle, J. Bares, and I. Nourbakhsh, "Material classification by drilling," in *17th International Symposium on Automation and Robotics in Construction.* Citeseer, 2000.

[100] N. S.-N. Lam, "Spatial interpolation methods : A review," *The American Cartographer*, vol. 10, no. 2, pp. 129–150, jan 1983.

[101] N. S.-N. Lam, "Spatial interpolation methods : a review," *The American Cartographer*, vol. 10, no. 2, pp. 129–150, 1983.

[102] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[103] S. Lee and I. Park, "Application of decision tree model for the ground subsidence hazard mapping near abandoned underground coal mines," *Journal of environmental management*, vol. 127, pp. 166–176, 2013.

[104] J. C. Leighton, "Development of a correlation between rotary drill performance and controlled blasting powder factors," Ph.D. dissertation, University of British Columbia, 1982.

[105] W. Lesso Jr and T. Burgess, "Pore pressure and porosity from mwd measurements," in *SPE/IADC Drilling Conference and Exhibition*. SPE, 1986, pp. SPE–14 801.

[106] P. Li *et al.*, "Time series prediction of mining subsidence based on a svm," *Mining Science and Technology (China)*, vol. 21, no. 4, pp. 557–562, 2011.

[107] S.-c. Li *et al.*, "Identifying the geological interface of the stratum of tunnel granite and classifying rock mass according to drilling energy theory," *Arabian Journal of Geosciences*, vol. 9, pp. 1–11, 2016.

[108] X. Li *et al.*, "A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ziyuan-3 stereo satellite imagery," *Remote sensing*, vol. 8, no. 6, p. 514, 2016.

[109] X. Li *et al.*, "Combining regression kriging with machine learning mapping for spatial variable estimation," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 1, pp. 27–31, 2019.

[110] Y. Li, S. Li, and J. Zhang, "Application of Pattern Recognition Techniques geochemical mapping of arsenic using geostatistics and machine learning methods, year = 2023, howpublished = SSRN preprint, doi = 10.2139/ssrn.4618916,."

[111] Z. Li *et al.*, "Improvement on inverse distance weighted interpolation for ore reserve estimation," in *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 4. IEEE, 2010, pp. 1703–1706.

[112] H. Liang *et al.*, "Research on lithology identification method based on mechanical specific energy principle and machine learning theory," *Expert Systems with Applications*, vol. 189, p. 116142, 2022.

[113] S. Lin *et al.*, "Imbalanced rock burst assessment using variational autoencoder-enhanced gradient boosting algorithms and explainability," *Underground Space*, 2024.

[114] Z.-N. Liu *et al.*, "The influence of distance weight on the inverse distance weighted method for ore-grade estimation," *Scientific Reports*, vol. 11, no. 1, jan 2021.

[115] Z.-N. Liu *et al.*, "The influence of distance weight on the inverse distance weighted method for ore-grade estimation," *Scientific Reports*, vol. 11, no. 1, p. 2689, 2021.

[116] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," *Computers & Geosciences*, vol. 34, no. 9, pp. 1044–1055, sep 2008.

[117] X. Lu *et al.*, "Orelm : a novel machine learning approach for prediction of flyrock in mine blasting," *Natural Resources Research*, vol. 29, pp. 641–654, 2020.

[118] X. Lu *et al.*, "Ensemble learning regression for estimating unconfined compressive strength of cemented paste backfill," *IEEE access*, vol. 7, pp. 72 125–72 133, 2019.

[119] X. Luo *et al.*, "Mine landslide susceptibility assessment using ivm, ann and svm models considering the contribution of affecting factors," *PloS one*, vol. 14, no. 4, p. e0215134, 2019.

[120] P. Lv, S. Wu, and W. Jin, "Real-time monitoring and prediction method to slope displacement in transition from open pit to underground," *Electronic Journal of Geotechnical Engineering*, vol. 20, no. 16, pp. 9649–9656, 2015.

[121] K. Manoj and M. Monjezi, "Prediction of flyrock in open pit blasting operation using machine learning method," *International Journal of Mining Science and Technology*, vol. 23, no. 3, pp. 313–316, 2013.

[122] S. S. Mansouri *et al.*, "Towards mav navigation in underground mine using deep learning," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2018, pp. 880–885.

[123] S. Manzoor *et al.*, "Predicting rock fragmentation based on drill monitoring : A case study from malmberget mine, sweden," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 122, no. 3, pp. 155–165, 2022.

[124] S. Manzoor *et al.*, "Rock mass characterization using mwd data and photogrammetry," in *Mining Goes Digital*. CRC Press, 2019, pp. 217–225.

[125] J. E. J. Martin Gonzalez, "Application of pattern recognition techniques to Monitoring-While-Drilling on a rotary electric blasthole drill at an open-pit coal mine," Master's thesis, Queen's University, Kingston, Ontario, Canada, 2007.

[126] G. Matheron, "Principles of geostatistics," *Economic geology*, vol. 58, no. 8, pp. 1246–1266, 1963.

[127] Q. Meng, X. Ma, and Y. Zhou, "Application of the pso-svm model for coal mine safety assessment," in *2012 8th International Conference on Natural Computation*. IEEE, 2012, pp. 393–397.

[128] V. M. Merwade, D. R. Maidment, and J. A. Goff, "Anisotropic considerations while interpolating river channel bathymetry," *Journal of Hydrology*, vol. 331, no. 3-4, pp. 731–741, 2006.

[129] J. Ming *et al.*, "Analysis models of technical and economic data of mining enterprises based on big data analysis," in *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE, 2018, pp. 224–227.

[130] A. Moazzeni and M. A. Haffar, "Artificial intelligence for lithology identification through real-time drilling data," *Journal of Earth Science & Climatic Change*, vol. 6,

no. 3, pp. 1–4, 2015.

[131] M. Monjezi, H. Amini Khoshalan, and A. Yazdian Varjani, "Prediction of flyrock and backbreak in open pit blasting operation : a neuro-genetic approach." *Arabian Journal of Geosciences*, vol. 5, no. 3, 2012.

[132] S. Monteiro, F. Ramos, and P. Hatherly, "Learning CRF models from drill rig sensors for autonomous mining," in *NeurIPS 2009 Workshop : Learning from Multiple Sources with Applications to Robotics*, Vancouver, Canada, 2009, workshop paper.

[133] S. T. Monteiro, F. Ramos, and P. Hatherly, "Conditional random fields for rock characterization using drill measurements," in *2009 International Conference on Machine Learning and Applications*. IEEE, 2009, pp. 366–371.

[134] S. T. Monteiro *et al.*, "Learning 3d geological structure from drill-rig sensors for automated mining," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, no. 3, 2011, p. 2500.

[135] S. Mozaffari, "Measurement While Drilling system in Aitik mine," Luleå, Sweden, 2007.

[136] J. Mukherjee *et al.*, "A novel index to detect opencast coal mine areas from landsat 8 oli/tirs," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 3, pp. 891–897, 2019.

[137] A. F. Mulaba-Bafubiandi and L. Tartibu, "A predictive approach for vibration analysis in underground mining operation," in *2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. IEEE, 2019, pp. 101–105.

[138] B. S. Murphy, "Pykrige : development of a kriging toolkit for python," in *AGU fall meeting abstracts*, vol. 2014, 2014, pp. H51K–0753.

[139] M. Nabaei and K. Shahbazi, "A new approach for predrilling the unconfined rock compressive strength prediction," *Petroleum science and technology*, vol. 30, no. 4, pp. 350–359, 2012.

[140] J. Navarro *et al.*, "Detection of potential overbreak zones in tunnel blasting from mwd data," *Tunnelling and Underground Space Technology*, vol. 82, pp. 504–516, 2018.

[141] H. Nguyen and X.-N. Bui, "Predicting blast-induced air overpressure : a robust artificial intelligence system based on artificial neural networks and random forest," *Natural Resources Research*, vol. 28, no. 3, pp. 893–907, 2019.

[142] H. Nourali and M. Osanloo, "A regression-tree-based model for mining capital cost estimation," *International Journal of Mining, Reclamation and Environment*, vol. 34, no. 2, pp. 88–100, 2020.

[143] G. T. Nwaila *et al.*, "Spatial interpolation using machine learning : from patterns and regularities to block models," *Natural Resources Research*, vol. 33, no. 1, pp. 129–161, 2024.

[144] O. Oloruntobi and S. Butt, "Application of specific energy for lithology identification," *Journal of Petroleum Science and Engineering*, vol. 184, p. 106402, 2020.

[145] I. Ongarbayev and N. Madani, "Anisotropic inverse distance weighting method : an innovative technique for resource modeling of vein-type deposits," *Journal of Mining and Environment*, vol. 13, no. 4, pp. 957–972, 2022.

[146] C. Paduraru and R. Dimitrakopoulos, "Responding to new information in a mining complex : Fast mechanisms using machine learning," *Mining Technology*, 2019.

[147] D. Papadopoulos and A. Benardos, "Combining machine learning and numerical modelling for rockburst prediction," *Geomechanics and Geoengineering*, vol. 19, no. 2, pp. 183–198, 2024.

[148] T. Peck, "Functional requirements definition for the compensated blastability index (cbi) algorithm," *Internal Technical Document. Peck Tech Consulting*, 2009.

[149] G. P. Petropoulos, P. Partsinevelos, and Z. Mitraka, "Change detection of surface mining activity and reclamation based on a machine learning approach of multi-temporal landsat tm imagery," *Geocarto International*, vol. 28, no. 4, pp. 323–342, 2013.

[150] J.-P. C. Pierre Delfiner, *Geostatistics : Modeling Spatial Uncertainty*. WILEY, Mar. 2012. [Online]. Available : https://www.ebook.de/de/product/15356462/pierre_delfiner_jean_paul_chiles_geostatistics_modeling_spatial_uncertainty.html

[151] C. Qi *et al.*, "Pressure drops of fresh cemented paste backfills through coupled test loop experiments and machine learning techniques," *Powder Technology*, vol. 361, pp. 748–758, 2020.

[152] C. Qi *et al.*, "An intelligent modelling framework for mechanical properties of cemented paste backfill," *Minerals Engineering*, vol. 123, pp. 16–27, 2018.

[153] C. Qi and X. Tang, "A hybrid ensemble method for improved prediction of slope stability," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 42, no. 15, pp. 1823–1839, 2018.

[154] C. Qi *et al.*, "Particulate matter concentration from open-cut coal mines : A hybrid machine learning estimation," *Environmental pollution*, vol. 263, p. 114517, 2020.

[155] D. Radočaj *et al.*, "Spatial prediction of heavy metal soil contents in continental croatia comparing machine learning and spatial interpolation methods," *Geodetski list*, vol. 74, no. 4, pp. 357–372, 2020.

[156] A. Rahman *et al.*, "A machine learning approach to find association between imaging features and xrf signatures of rocks in underground mines," in *2015 IEEE SENSORS*. IEEE, 2015, pp. 1–4.

[157] K. Ram Chandar, V. Sastry, and C. Hegde, "A critical comparison of regression models and artificial neural networks to predict ground vibrations," *Geotechnical and geological engineering*, vol. 35, pp. 573–583, 2017.

[158] J. Rasmus and D. G. Stephens, "Real-time pore-pressure evaluation from mwd/lwd measurements and drilling-derived formation strength," *SPE drilling engineering*, vol. 6, no. 04, pp. 264–272, 1991.

[159] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[160] H. Reckmann *et al.*, "Mwd failure rates due to drilling dynamics," in *SPE/IADC Drilling Conference and Exhibition*. SPE, 2010, pp. SPE–127 413.

[161] K. Ristovski *et al.*, "Dispatch with confidence : Integration of machine learning, optimization and simulation for open pit mines," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 1981–1989.

[162] D. R. Roberts *et al.*, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017.

[163] D. E. Rumelhart *et al.*, "Backpropagation : The basic theory," in *Backpropagation*. Psychology Press, 2013, pp. 1–34.

[164] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[165] M. Sari, E. Ghasemi, and M. Ataei, "Stochastic modeling approach for the evaluation of backbreak due to blasting operations in open pit mines," *Rock mechanics and rock engineering*, vol. 47, pp. 771–783, 2014.

[166] S. Schneider *et al.*, "Classification of hyperspectral imagery using gps and the oad covariance function with automated endmember extraction," in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*. IEEE, 2011, pp. 579–584.

[167] N. Schnitzler, P.-S. Ross, and E. Gloaguen, "Using machine learning to estimate a key missing geochemical variable in mining exploration : Application of the random forest algorithm to multi-sensor core logging data," *Journal of Geochemical Exploration*, vol. 205, p. 106344, 2019.

[168] A. Sekulić *et al.*, "Random forest spatial interpolation," *Remote Sensing*, vol. 12, no. 10, p. 1687, 2020.

[169] S. Shahbeik *et al.*, "Comparison between ordinary kriging (OK) and inverse distance weighted (IDW) based on estimation error. Case study : Dardevey iron ore deposit, NE Iran," *Arabian Journal of Geosciences*, vol. 7, no. 9, pp. 3693–3704, jun 2013.

[170] Y. Shang *et al.*, "A novel artificial intelligence approach to predict blast-induced ground vibration in open-pit mines based on the firefly algorithm and artificial neural network," *Natural Resources Research*, vol. 29, no. 2, pp. 723–737, 2020.

[171] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, 1968, pp. 517–524.

[172] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, 1968, pp. 517–524.

[173] Y. Shi, L. Li, and L. Zhang, "Application and comparing of idw and kriging interpolation in spatial rainfall information," in *Geoinformatics 2007 : Geospatial Information Science*, vol. 6753. SPIE, 2007, pp. 539–550.

[174] M. Sikora *et al.*, "Application of a hybrid method of machine learning for description and on-line estimation of methane hazard in mine workings," *Journal of Mining Science*, vol. 47, pp. 493–505, 2011.

[175] K. L. Silversides and A. Melkumyan, "Machine learning for classification of stratified geology from mwd data," *Ore Geology Reviews*, vol. 142, p. 104737, 2022.

[176] K. Silversides and A. Melkumyan, "Multivariate gaussian process for distinguishing geological units using measure while drilling data," in *Mining goes Digital : Proceedings of the 39th International Symposium'Application of Computers and Operations Research in the Mineral Industry'(APCOM 2019), Wroclaw, Poland, 4–6 June 2019*. CRC Press Boca Raton, FL, USA, 2019, p. 94.

[177] N. Srivastava *et al.*, "Dropout : a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[178] Q. Su *et al.*, "Comparative assessment of three nonlinear approaches for landslide susceptibility mapping in a coal mine area," *ISPRS International Journal of Geo-Information*, vol. 6, no. 7, p. 228, 2017.

[179] X. Sun *et al.*, "The use of a machine learning method to predict the real-time link travel time of open-pit trucks," *Mathematical Problems in Engineering*, vol. 2018, no. 1, p. 4368045, 2018.

[180] J. Tan *et al.*, "Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with multiple-covariates," *Journal of Hydrology*, vol. 598, p. 126270, 2021.

[181] T. Tao *et al.*, "Blastability classification of fractured rock mass based on weighted clustering analysis," *Chemical Engineering Transactions*, vol. 62, pp. 469–474, 2017.

[182] Z. Tao *et al.*, "Rock mass quality prediction of open-pit gold mine slope based on the kriging interpolation method," *Geotechnical and Geological Engineering*, vol. 38, pp. 5851–5865, 2020.

[183] D. Tien Bui *et al.*, "Land subsidence susceptibility mapping in south korea using machine learning algorithms," *Sensors*, vol. 18, no. 8, p. 2464, 2018.

[184] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society Series B : Statistical Methodology*, vol. 61, no. 3, pp. 611–622, 1999.

[185] K. Tiwari and L. L. Narine, "A comparison of machine learning and geostatistical approaches for mapping forest canopy height over the southeastern us using icesat-2," *Remote Sensing*, vol. 14, no. 22, 2022. [Online]. Available : https://www.mdpi.com/2072-4292/14/22/5651

[186] F. Tobar, E. Cazelles, and T. de Wolff, "Computationally-efficient initialisation of gps : The generalised variogram method," *arXiv preprint arXiv :2210.05394*, 2022.

[187] R. Tolosana-Delgado, V. Pawlowsky-Glahn, and J. Egozcue, "Simplicial indicator kriging," *Journal of China University of Geosciences*, vol. 19, no. 1, pp. 65–71, 2008.

[188] H. Trevor, T. Robert, and F. Jerome, *The elements of statistical learning : data mining, inference, and prediction.* Spinger, 2009.

[189] V. D. Trong, T. D. Bao, and S. Fomin, "Ordinary kriging comparison and inverse distance weighting for quality assessment of vietnam cement limestone deposits," *Proceedings of the International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, Albena, Bulgaria*, vol. 29, 2017.

[190] T. Tunçay *et al.*, "Application of regression kriging and machine learning methods to estimate soil moisture constants in a semi-arid terrestrial area," *Computers and Electronics in Agriculture*, vol. 212, p. 108118, 2023.

[191] P. Tziachris *et al.*, "Comparing machine learning models and hybrid geostatistical methods using environmental and soil covariates for soil ph prediction," *ISPRS International Journal of Geo-Information*, vol. 9, no. 4, p. 276, 2020.

[192] W. K. Utt, "Neural network technology for strata strength characterization," in *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)*, vol. 6. IEEE, 1999, pp. 3806–3809.

[193] J. van Eldert *et al.*, "Measurement while drilling (mwd) technology for blasting damage calculation," in *12th International Symposium on Rock Fragmentation by Blas-*

ting (Fragblast 12), 11-13 June, 2018, Luleå, Sweden. Luleå University of Technology, 2018, pp. 139–148.

[194] J. van Eldert *et al.*, "Application of measurement while drilling technology to predict rock mass quality and rock support for tunnelling," *Rock Mechanics and Rock Engineering*, vol. 53, no. 3, pp. 1349–1358, 2020.

[195] P. J. van Oosterom *et al.*, "Proceedings-the 26th agile international conference on geographic information science spatial data for design : Preface," in *26th AGILE Conference on Geographic Information Science, AGILE 2023 : Spatial data for design*. Copernicus, 2023.

[196] Y. Vasuki *et al.*, "The spatial-temporal patterns of land cover changes due to mining activities in the darling range, western australia : A visual analytics approach," *Ore Geology Reviews*, vol. 108, pp. 23–32, 2019.

[197] V. S. Vezhapparambu, J. Eidsvik, and S. L. Ellefmo, "Rock classification using multivariate analysis of measurement while drilling data : Towards a better sampling strategy," *Minerals*, vol. 8, no. 9, p. 384, 2018.

[198] S.-f. Wang *et al.*, "Strength prediction and drillability identification for rock based on measurement while drilling parameters," *Journal of Central South University*, vol. 30, no. 12, pp. 4036–4051, 2023.

[199] W. Wang *et al.*, "Deep variational canonical correlation analysis," 2016. [Online]. Available : https://arxiv.org/abs/1610.03454

[200] Q. Wangde, W. Guojun, and L. Haojie, "Formation drillability prediction by using cascade model based on well logging data in the deep drilling process," *Petroleum*, May 2021.

[201] W. Wei *et al.*, "Prediction of mine water inflow based on support vector machine," in *2011 Workshop on Digital Media and Digital Content Management*. IEEE, 2011, pp. 326–329.

[202] Y. Xiong *et al.*, "A physically constrained variational autoencoder for geochemical pattern recognition," *Mathematical Geosciences*, pp. 1–24, 2022.

[203] S. Xu *et al.*, "Optimization of blasting parameters for an underground mine through prediction of blasting vibration," *Journal of Vibration and Control*, vol. 25, no. 9, pp. 1585–1595, 2019.

[204] X. Xue, W. SUN, and R. LIANG, "A new method of real time dynamic forecast of truck link travel time in open mines," *Journal of China Coal Society*, vol. 37, no. 8, pp. 1418–1422, 2012.

[205] A. Yadav, G. Singh, and B. Behera, "A machine learning model for evaluation of chain pillar stability in deep longwall workings in india," *Mining, Metallurgy & Exploration*, vol. 40, no. 6, pp. 2119–2137, 2023.

[206] J. K. Yamamoto, "Correcting the smoothing effect of ordinary kriging estimates," *Mathematical geology*, vol. 37, pp. 69–94, 2005.

[207] H. Yang *et al.*, "Intelligent prediction of blasting-induced ground vibration using anfis optimized by ga and pso," *Natural Resources Research*, vol. 29, pp. 739–750, 2020.

[208] A. B. Yasrebi *et al.*, "Application of an inverse distance weighted anisotropic method (IDWAM) to estimate elemental distribution in eastern kahang cu-mo porphyry deposit, central Iran," *International Journal of Mining and Mineral Engineering*, vol. 7, no. 4, p. 340, 2016.

[209] H. Yenice, "Determination of Drilling Rate Index Based on Rock Strength Using Regression Analysis," *Anais da Academia Brasileira de Ciências*, vol. 91, no. 3, 2019.

[210] H. Yenice, M. V. Özdoğan, and M. K. Özfırat, "A sampling study on rock properties affecting drilling rate index (DRI)," *Journal of African Earth Sciences*, vol. 141, pp. 1–8, May 2018.

[211] K. Yin and H. Liu, "Using information extracted from drill data to improve blasting design and fragmentation," *Fragblast*, vol. 5, no. 3, pp. 157–179, 2001.

[212] B. Yu *et al.*, "Real-time rock strength determination based on rock drillability index and drilling specific energy : an experimental study," *Bulletin of Engineering Geology and the Environment*, vol. 80, pp. 3589–3603, 2021.

[213] Y. Yuan *et al.*, "Quantum chemical calculations with machine learning for multipolar electrostatics prediction in rna : An application to pentose," *Journal of Chemical Information and Modeling*, vol. 62, no. 17, pp. 4122–4133, 2022.

[214] G. Zagré *et al.*, "Machine learning algorithms for real-time coal recognition using monitor-while-drilling data," *International Journal of Mining, Reclamation and Environment*, vol. 38, no. 1, pp. 27–52, 2024.

[215] B. Zhang *et al.*, "Strength prediction and application of cemented paste backfill based on machine learning and strength correction," *Heliyon*, vol. 8, no. 8, 2022.

[216] C. Zhang *et al.*, "Methane gas density monitoring and predicting based on rfid sensor tag and cnn algorithm," *Electronics*, vol. 7, no. 5, p. 69, 2018.

[217] S. Zhang *et al.*, "Integration of auto-encoder network with density-based spatial clustering for geochemical anomaly detection for mineral exploration," *Computers & Geosciences*, vol. 130, pp. 43–56, 2019.

[218] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch : an efficient data clustering method for very large databases," *ACM sigmod record*, vol. 25, no. 2, pp. 103–114, 1996.

[219] X. Zhang *et al.*, "Novel soft computing model for predicting blast-induced ground vibration in open-pit mines based on particle swarm optimization and xgboost," *Natural Resources Research*, vol. 29, no. 2, pp. 711–721, 2020.

[220] Y. Zhang *et al.*, "Support vector machine of the coal mine machinery equipment fault diagnosis," in *2013 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 2013, pp. 1141–1146.

[221] B. Zhao *et al.*, "Recognition of multivariate geochemical anomalies using a geologically-constrained variational autoencoder network with spectrum separable module–a case study in shangluo district, china," *Applied Geochemistry*, vol. 156, p. 105765, 2023.

[222] D. Zhao *et al.*, "Using random forest for the risk assessment of coal-floor water inrush in panjiayao coal mine, northern china," *Hydrogeology Journal*, vol. 26, no. 7, pp. 2327–2340, 2018.

[223] H. Zhou *et al.*, "Spectral feature selection for automated rock recognition using gaussian process classification," in *Proceedings of Australian Conference on Robotics and Automation*, 2009.

[224] H. Zhou *et al.*, "Automatic rock recognition from drilling performance data," pp. 3407–3412, 2012.

[225] H. Zhou *et al.*, "An adaptive data driven model for characterizing rock properties from drilling data," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1909–1915.

[226] H. Zhou *et al.*, "Automated rock recognition with wavelet feature space projection and gaussian process classification," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 4444–4450.

[227] J. Zhou, X. Li, and H. S. Mitri, "Classification of rockburst in underground projects : comparison of ten supervised learning methods," *Journal of Computing in Civil Engineering*, vol. 30, no. 5, p. 04016003, 2016.

[228] J. Zhou, X. Li, and X. Shi, "Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines," *Safety science*, vol. 50, no. 4, pp. 629–644, 2012.

[229] J. Zhou, X. Shi, and X. Li, "Utilizing gradient boosted machine for the prediction of damage to residential structures owing to blasting vibrations of open pit mining," *Journal of Vibration and Control*, vol. 22, no. 19, pp. 3986–3997, 2016.

[230] R. Zuo *et al.*, "A geologically constrained variational autoencoder for mineral prospectivity mapping," *Natural Resources Research*, vol. 31, no. 3, pp. 1121–1133, 2022.

ANNEXE A ANNEXE DU CHAPITRE 4

.

A.1 Parameters' constraints : rules of thumb

Our approach involved using variography to constrain physical parameters of the model. In this Appendix, we share intuitions and rules of thumb that we developed with experience.

Direction plays a role in the evaluation of the partial sill parameter (i.e output-scale), σ_f. Computing variography in several directions as in Fig.6.4.3 allows to observe a range of reasonable values, from which we extract the maximum and the minimum, denoted $\sigma_f^{\max}$ and $\sigma_f^{\min}$ respectively. We bound σ_f with these two observed values (i.e $\sigma_f \in [\sigma_f^{\min}, \sigma_f^{\max}]$). Although this intuitive choice leverages sill values from multiple semi-variograms, it does not present all possible directions, hence an additional tolerance may be used, a safety margin that we implemented by widening the interval to generously rounded values. In cases where expert knowledge is available, leveraging it would seem appropriate as well.

For the length-scale parameters denoted generally by ℓ, setting the constraints $\ell \in [\widehat{\ell} - \widehat{\ell}/4, \widehat{\ell} + \widehat{\ell}/4]$ yielded a good predictive performance. Here, $\widehat{\ell}$ is the estimate extracted from the semi-variograms. In our case and according to the values of $\widehat{\ell}$, intervals with length $\frac{\widehat{\ell}}{2}$ were appropriate. In a more general setup, interval length could take the form $\widehat{\ell}/r$, where r is a hyper-parameter to be tuned by experimentation. While a large r may lead to ignoring some potential solutions, a very small r yields a very large interval that may not fix the convergence challenges that the method attempts to solve. A good r is a trade-off between exploration (small r) and exploitation (large r), a compromise often met in operations research and in machine-learning problems. With our semi-variograms, we concluded that horizontal isotropy was a valid assumption, that is $\ell_x = \ell_y$. To regain some flexibility, we however let both parameters take different values in the optimization but we constrained them with the same intervals ensuring some accordance with the observed anisotropy.

A.2 Algorithms

Algorithm 1 Infer p

1- Construct a grid G contains a set of values of p to be chosen from : $G = \{1, 1.5, 2, ...\}$.

2- Loop for p in G :

2.1- Loop for (s_j, t_j) in D_{train} :

2.1.1- Apply IDW interpolation on s_j using all other observed targets : $\widehat{t}_{j,p} = \sum_{i=1, i\neq j}^{|\overline{\mathcal{T}}|} w_{ji}(p)t_i$, $w_{ji}(p) = \frac{d(s_i,s_j)^{-p}}{\sum_i d(s_i,s_j)^{-p}}$.

2.1.2- Compute the error between the predicted and the true target : $e_{j,p} = (\widehat{t}_{j,p} - t_j)^2$.

2.2 - Compute the LOOCV error for the parameter p : $E_p^{LOOCV} = \frac{1}{|\overline{\mathcal{T}}|} \sum_{j=1}^{|\overline{\mathcal{T}}|} e_{j,p}$.

3- Choose the power parameter p from G that have the minimum LOOCV error E_p^{LOOCV}.

Algorithm 2 Infer p for the modified IDW

1- Construct a grid G contains a set of values of p to be chosen from : $G = \{1, 1.5, 2, ...\}$.

2- Loop for p in G :

2.1- Loop for (s_j, t_j) in D_{train} :

2.1.1- Apply IDW interpolation on s_j using all other observed targets : $\widehat{t}_{j,p} = \sum_{i=1, i\neq j}^{|\overline{\mathcal{T}}|} w_{ji}(p)t_i$, $w_{ij}(p) = \frac{d(s_i,s_j)^{-p}}{\sum_i d(s_i,s_j)^{-p}}$.

2.1.2- Compute the error between the predicted and the true target : $e_{j,p} = (\widehat{t}_{j,p} - t_j)^2$.

2.2- Compute the coefficients : $c_j = 1 + \frac{e_j}{\max_v e_v}$, $j = 1, ..., n$.

2.3- Loop for (s_i, t_i) in D_{train} :

2.3.1- Update the weights : $\widetilde{w_{ji}}(p) = \frac{c_i . w_{ji}(p)}{\sum_{u,u\neq j} c_u . w_{ju}(p)}$.

2.3.2- Interpolate using $\widetilde{w_{ji}}$: $\widehat{t}_{j,p} = \sum_{i=1, i\neq j}^{|\overline{\mathcal{T}}|} \widetilde{w_{ji}}(p)t_i$

2.3.4- Compute the error between the predicted and the true observation : $e_{j,p} = (\widehat{t}_{j,p} - t_j)^2$.

2.4 - Compute the LOOCV error for the parameter p : $E_p^{LOOCV} = \frac{1}{|\overline{\mathcal{T}}|} \sum_{j=1}^{|\overline{\mathcal{T}}|} e_{j,p}$.

3- Choose the power parameter p from G that have the minimum LOOCV error E_p^{LOOCV}.

Algorithm 3 Infer GP raw parameters $\widetilde{\theta}$

1- Start with initial value $\widetilde{\theta} = \widetilde{\theta}_0$.

2- Fix a learning rate α

3- Loop for $i = 1$....number of iterations :

3.1 - Forward step : compute the log-likelihood $l(\theta) = l(g(\widetilde{\theta}))$

3.3 - Backward step : $\widetilde{\theta} \leftarrow \widetilde{\theta} + \alpha \frac{\partial l(g(\widetilde{\theta}))}{\partial \widetilde{\theta}}$

Algorithm 4 Gradient based inference for IDW

1- Start with initial value $\widetilde{\theta} = \widetilde{\theta}_0$.

2- Fix a learning rate α.

3- Loop for $i = 1$....number of iterations :

3.1 - Compute $\mathbf{W}(g(\widetilde{\theta}))$.

3.2 - Forward step : Compute $E^{LOOCV}(g(\widetilde{\theta}))$.

3.3 - Backward step : $\widetilde{\theta} \leftarrow \widetilde{\theta} - \alpha \frac{\partial E^{LOOCV}(g(\widetilde{\theta}))}{\partial \widetilde{\theta}}$.

Algorithm 5 Gradient based inference for GBM-IDW

1- Start with initial value $\widetilde{\theta} = \widetilde{\theta}_0$, $\widetilde{\mathbf{c}} = \widetilde{\mathbf{c}}_0$.

2- Fix a learning rate α.

3- Loop for $i = 1$....number of iterations :

3.1 - Compute $\mathbf{W}(g(\widetilde{\theta}))$.

3.2 - Forward step : Compute $E^{LOOCV}(\theta, \mathbf{c}) = E^{LOOCV}(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}})) = \frac{1}{|\overline{\mathcal{T}}|} \left\| \frac{\mathbf{W}(\mathbf{g}(\widetilde{\theta}))[\mathbf{h}(\widetilde{\mathbf{c}}) \odot \mathbf{t}]}{\mathbf{W}(\mathbf{g}(\widetilde{\theta}))[\mathbf{h}(\widetilde{\mathbf{c}}) \odot \mathbf{1}]} - \mathbf{t} \right\|^2$.

3.3 -Backward step : $\widetilde{\theta} \leftarrow \widetilde{\theta} - \alpha \frac{\partial E(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}}))}{\partial \widetilde{\theta}}$, $\widetilde{\mathbf{c}} \leftarrow \widetilde{\mathbf{c}} - \alpha \frac{\partial E(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}}))}{\partial \widetilde{\mathbf{c}}}$.

Algorithm 6 Gradient based inference for GBM-IDW using LHOCV

1- Start with initial value $\widetilde{\theta} = \widetilde{\theta}_0$, $\widetilde{\mathbf{c}} = \widetilde{\mathbf{c}}_0$.

2- Fix a learning rate α.

3- Loop for $i = 1$....number of iterations :

3.1 - Compute $\mathbf{W}(g(\widetilde{\theta}))$, $\mathbf{A}$.

3.2 - Forward step : Compute $E^{LHOCV}(\theta, \mathbf{c}) = E^{LHOCV}(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}})) = \frac{1}{|\overline{\mathcal{T}}|} \left\| \frac{(\mathbf{W}(\mathbf{g}(\widetilde{\theta})) \odot \mathbf{A})[\mathbf{h}(\widetilde{\mathbf{c}}) \odot \mathbf{t}]}{(\mathbf{W}(\mathbf{g}(\widetilde{\theta})) \odot \mathbf{A})[\mathbf{h}(\widetilde{\mathbf{c}}) \odot \mathbf{1}]} - \mathbf{t} \right\|^2$.

3.3 -Backward step : $\widetilde{\theta} \leftarrow \widetilde{\theta} - \alpha \frac{\partial E(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}}))}{\partial \widetilde{\theta}}$, $\widetilde{\mathbf{c}} \leftarrow \widetilde{\mathbf{c}} - \alpha \frac{\partial E(g(\widetilde{\theta}), \mathbf{h}(\widetilde{\mathbf{c}}))}{\partial \widetilde{\mathbf{c}}}$.

A.3 Derivation of the matrix notation for IDW models

A.3.1 The gradient based conventional IDW model

Consider $\overline{\mathcal{T}} = \{1, ..., n\}$.

In Sect.4.3.2, we claimed that the LOOCV error method for the IDW base model could be forwardly computed using a matrix notation as follows :

$$E^{LOOCV}(\theta) = \frac{1}{n} \left\| \widehat{\mathbf{t}} - \mathbf{t} \right\|^2 = \frac{1}{n} \left\| \mathbf{W}\mathbf{t} - \mathbf{t} \right\|^2$$

where :

- $\mathbf{t}$ is the vector that contains all the observed values of the target variables t_i, $i = 1, ..., n$;
- $\mathbf{W}$ is the matrix containing 0 at the diagonal, and the IDW weights $w_{ji} = \frac{d_a(s_i, s_j)^{-p}}{\sum\limits_i d_a(s_i, s_j)^{-p}}$ if $i \neq j$;
- $\widehat{\mathbf{t}}$ is the vector that contains all the IDW predicted values $\widehat{t_i}$, $t_i = 1, ..., n$;
- $\left\| . \right\|$ is the Euclidean norm.

We could prove that $\mathbf{W}\mathbf{t} = \widehat{\mathbf{t}}$:

$$\mathbf{W}\mathbf{t} = \begin{bmatrix} 0 & w_{21} & . & . & w_{n1} \\ w_{12} & . & . & . & . \\ . & . & . & w_{ji} & . \\ . & . & . & . & . \\ w_{1n} & . & . & . & 0 \end{bmatrix} \begin{bmatrix} t_1 \\ . \\ t_j \\ . \\ t_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1, i\neq 1}^{n} w_{ji} t_i \\ . \\ \sum\limits_{i=1, i\neq j}^{n} w_{ji} t_i \\ . \\ \sum\limits_{i=1, i\neq n}^{n} w_{ji} t_i \end{bmatrix} = \begin{bmatrix} \widehat{t_1} \\ . \\ \widehat{t_j} \\ . \\ \widehat{t_n} \end{bmatrix} = \widehat{\mathbf{t}}.$$

The $E^{LOOCV}(\theta)$ as defined in Sect.4.3.2, is the average of the errors between target points and their correspondent prediction values :

$$E^{LOOCV}(\theta) = \frac{1}{n} \sum_{j=1}^{n} (t_j - \widehat{t_j})^2 = \frac{1}{n} \left\| \widehat{\mathbf{t}} - \mathbf{t} \right\|^2$$

A.3.2 The gradient based modified IDW model

In Eq.(4.29), we defined the matrix notation of the IDW-modified approach interpolation as :

$$\widehat{\mathbf{t}} = \frac{\mathbf{W}[\mathbf{c} \odot \mathbf{t}]}{\mathbf{W}\mathbf{c}}$$

- $\odot$ is the element-wise product : $\mathbf{c} \odot \mathbf{t} = [c_i t_i]$;

— $\div$ is the element-wise subtraction : $\frac{\mathbf{c}}{\mathbf{t}} = [\frac{c_i}{t_i}]$.

Starting with the nominator : $\mathbf{W}[\mathbf{c} \odot \mathbf{t}] = \begin{bmatrix} 0 & w_{21} & . & . & w_{n1} \\ w_{12} & . & . & . & . \\ . & . & . & w_{ji} & . \\ . & . & . & . & . \\ w_{1n} & . & . & . & 0 \end{bmatrix} \left(\begin{bmatrix} c_1 \\ . \\ c_j \\ . \\ c_n \end{bmatrix} \odot \begin{bmatrix} t_1 \\ . \\ t_j \\ . \\ t_n \end{bmatrix} \right) =$

$$\begin{bmatrix} 0 & w_{21} & . & . & w_{n1} \\ w_{12} & . & . & . & . \\ . & . & . & w_{ji} & . \\ . & . & . & . & . \\ w_{1n} & . & . & . & 0 \end{bmatrix} \begin{bmatrix} c_1 t_1 \\ . \\ c_j t_j \\ . \\ c_n t_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1, i\neq 1}^{n} c_i w_{ji} t_i \\ . \\ \sum_{i=1, i\neq j}^{n} c_i w_{ji} t_i \\ . \\ \sum_{i=1, i\neq n}^{n} c_i w_{ji} t_i \end{bmatrix} \quad (1).$$

In the denominator, we have :

$$\mathbf{Wc} = \begin{bmatrix} 0 & w_{21} & . & . & w_{n1} \\ w_{12} & . & . & . & . \\ . & . & . & w_{ji} & . \\ . & . & . & . & . \\ w_{1n} & . & . & . & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ . \\ c_j \\ . \\ c_n \end{bmatrix} = \begin{bmatrix} \sum_{u=1, u\neq 1}^{n} c_u w_{ju} \\ . \\ \sum_{u=1, u\neq j}^{n} c_u w_{ju} \\ . \\ \sum_{u=1, i\neq n}^{n} c_u w_{ju} \end{bmatrix} \quad (2).$$

The element-wise subtraction of (1) and (2) gives the following :

$$\frac{\mathbf{W}\left[\mathbf{c} \odot \mathbf{t}\right]}{\mathbf{Wc}} = \begin{bmatrix} \sum_{i=1, i\neq 1}^{n} \frac{c_i w_{ji}}{\sum_{u=1, u\neq 1}^{n} c_u w_{ju}} t_i \\ . \\ \sum_{i=1, i\neq j}^{n} \frac{c_i w_{ji}}{\sum_{u=1, u\neq j}^{n} c_u w_{ju}} t_i \\ . \\ \sum_{i=1, i\neq n}^{n} \frac{c_i w_{ji}}{\sum_{u=1, u\neq N}^{n} c_u w_{ju}} t_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1, i\neq 1}^{n} \widetilde{w_{ji}} t_i \\ . \\ \sum_{i=1, i\neq j}^{n} \widetilde{w_{ji}} t_i \\ . \\ \sum_{i=1, i\neq n}^{n} \widetilde{w_{ji}} t_i \end{bmatrix}$$

A.4 Sufficiency of distance equation for modeling anisotropy

In this section, we show that for data in the form of boreholes, the distance modeled in Eq.(4.3) is sufficient to account for anisotropy in variograms. We consider two points, $A = (x_A, y_A, z_A)$ and $B = (x_B, y_B, z_B)$, lying in two different boreholes, with B at a higher altitude. Let $P = (x_B, y_B, z_A)$ be the projection of B onto the horizontal plane containing A. Fig.A.1

illustrates this example.

We use the assumption of horizontal isotropy validated from variography in Sect.4.3.1 so that the geometric anisotropy will depend only on the vertical angle θ, the vertical range a_v, and the horizontal range a_h.

Using the Pythagorean theorem, the square of the distance $d(A,B)^2$ is the sum of the squared horizontal distance $d(A,P)^2 = (x_A - x_B)^2 + (y_A - y_B)^2$ and the squared vertical distance $d(P,B)^2 = (z_A - z_B)^2$:

$$d(A,B)^2 = d(A,P)^2 + d(P,B)^2.$$

To account for anisotropy, if two points lie in the horizontal plane, we have to divide the Euclidean distance by the horizontal range :

$$d_a(A,P)^2 = \frac{d(A,P)^2}{a_h^2}.$$

When two points lie on the vertical line, we have to divide by the vertical range to account for anisotropy :

$$d_a(P,B)^2 = \frac{d(P,B)^2}{a_v^2}.$$

In the context of drilling holes, a_v is the minimal range, and a_h is the maximum range. To account for anisotropy when a point is neither on the horizontal plane nor on the vertical line, as is the case for AB, the distance should be divided by a_θ :

$$d_a(A,B)^2 = \frac{d(A,B)^2}{a_\theta^2},$$

where a_θ should respect the ellipse equation with a_h as the major axis and a_v as the minor axis, as illustrated in Fig.A.2 :

$$\frac{a_\theta^2 \cos^2(\theta)}{a_h^2} + \frac{a_\theta^2 \sin^2(\theta)}{a_v^2} = 1.$$

We can express a_θ as a function of a_v, a_h, and θ :

$$a_\theta^2 = \frac{a_h^2 a_v^2}{a_v^2 \cos^2(\theta) + a_h^2 \sin^2(\theta)}.$$

Replacing a_θ with its value when accounting for anisotropy in AB, we get :

$$d_a(A,B)^2 = \frac{d(A,B)^2}{a_\theta^2} = \frac{d(A,B)^2(a_v^2\cos^2(\theta) + a_h^2\sin^2(\theta))}{a_h^2 a_v^2} = \frac{a_v^2(d(A,B)cos(\theta))^2 + a_h^2(d(A,B)sin(\theta))^2}{a_h^2 a_v^2}.$$

Using the fact that $d(A,B)\cos(\theta) = d(A,P)$ and $d(A,B)\sin(\theta) = d(B,P)$, we get :

$$\frac{d(A,B)^2}{a_\theta^2} = \frac{a_v^2 d(A,P)^2 + a_h^2 d(P,B)^2}{a_v^2 a_h^2} = \frac{d(A,P)^2}{a_h^2} + \frac{d(P,B)^2}{a_v^2}.$$

Since $d(A,P)^2 = (x_A - x_B)^2 + (y_A - y_B)^2$ and $d(B,P)^2 = (z_A - z_B)^2$, the equation could be reduced as :

$$\frac{d(A,B)^2}{a_\theta^2} = \frac{(x_A - x_B)^2}{a_h^2} + \frac{(y_A - y_B)^2}{a_h^2} + \frac{(z_A - z_B)^2}{a_v^2}.$$

By setting $\ell_x = \ell_y = a_h$ and $\ell_z = a_v$, this would be equivalent to the anisotropic distance defined in Eq.(4.3).

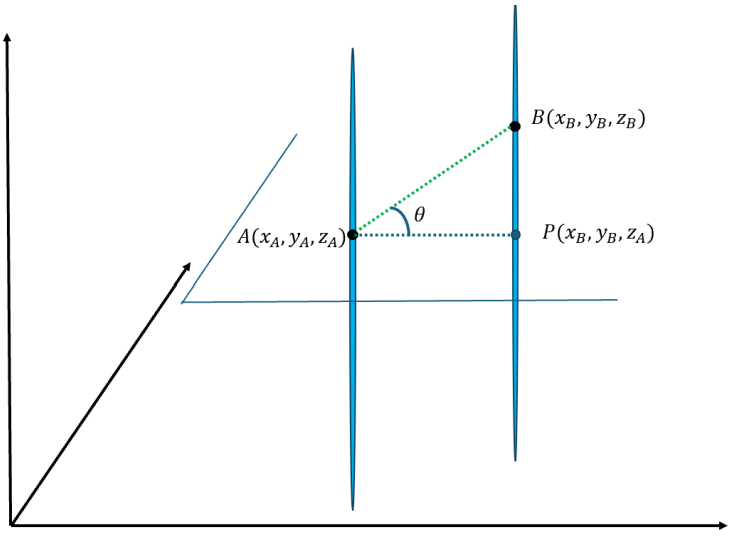


FIGURE A.1 Illustrative example

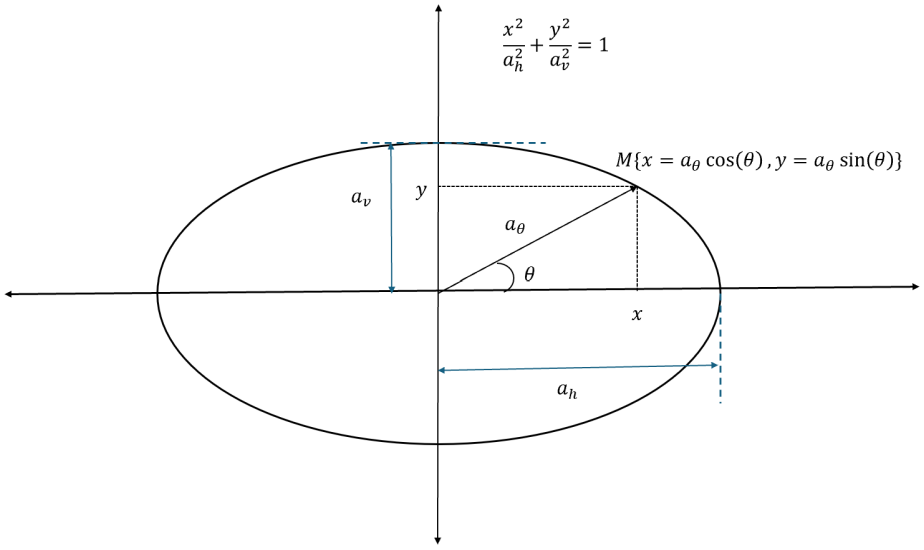


FIGURE A.2 Ellipse for ranges