POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Algorithmes d'apprentissage automatique appliqués au forage minier pour la reconnaissance de la lithologie et la détection de l'état du tricône

ZAGRÉ GILLES ERIC PALOBDE

Département de mathématiques et de génie industriel

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

Mathématiques

Avril 2025

© Gilles Eric Zagré, 2025.

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

Algorithmes d'apprentissage automatique appliqués au forage minier pour la reconnaissance de la lithologie et la détection de l'état du tricône

présentée par **Gilles Eric ZAGRÉ**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Bruno AGARD, président
Michel GAMACHE, membre et directeur de recherche
Richard LABIB, membre et codirecteur de recherche
Nelson MORALES, membre
Chahid AHABCHANE, membre externe

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de cette thèse.

Tout d'abord, je remercie sincèrement mon directeur de recherche, Michel Gamache, et mon co-directeur, Richard Labib, pour leur confiance, leur patience et la grande liberté académique qu'ils m'ont accordée tout au long de ce parcours. J'ai énormément apprécié nos échanges, en particulier nos discussions stimulantes d'après-midi, qui ont souvent débouché sur des pistes nouvelles et enrichissantes.

Mes remerciements vont également à Jonathan Peck, mon mentor et leader chez Peck Tech, pour ses encouragements constants et sa capacité à toujours me pousser à donner le meilleur de moi-même. Son soutien a été une source majeure de motivation et de dépassement personnel.

Je tiens à remercier toute l'équipe de Peck Tech, et tout spécialement Viktor, Aurélien et Richard, pour leur soutien technique, leur disponibilité et l'esprit d'équipe qui m'a accompagné durant toute cette collaboration.

Cette thèse, a été réalisée en étroite collaboration avec l'industrie, s'est vue grandement enrichie par ce contexte appliqué. J'ai eu la chance de visiter plusieurs sites miniers, ce qui m'a permis de mieux comprendre les enjeux du terrain et de donner un sens concret aux modèles développés.

Enfin, je remercie du fond du cœur ma famille pour leur soutien indéfectible, leur patience et leur amour qui m'ont permis d'aller au bout de ce projet.

À toutes et à tous, merci!

RÉSUMÉ

Cette thèse aborde les défis de l'optimisation du forage minier (mine à ciel ouvert) à l'aide de l'apprentissage automatique, en exploitant les données de mesure en continu de performance de la foreuse pendant le forage (MWD : *Measurement While Drilling*). Le but principal est d'améliorer la précision de l'identification lithologique et de mieux anticiper l'usure des trépans (outil de coupe utilisé dans les opérations de forage minier), deux éléments cruciaux pour optimiser l'efficacité opérationnelle et réduire les coûts associés aux opérations de forage.

L'étude explore des méthodes d'apprentissage supervisés et non supervisés pour analyser les signaux MWD. Ces données sont souvent bruitées, influencées par les conditions de forage et les propriétés géologiques, ce qui complique leur interprétation. La recherche propose des approches innovantes basées sur l'intégration de plusieurs modèles de classification supervisés pour détecter les changements lithologiques et un modèle de réseau de neurones récurrent pour l'usure des tricônes, en intégrant des séquences temporelles pour capturer les patrons évolutifs dans le forage.

Les travaux sont organisés autour de trois axes principaux, chacun abordé dans un chapitre distinct. Un ensemble de modèles de classification supervisé a été élaboré pour détecter les couches de charbon à partir des données MWD. Cette approche a d'ailleurs été validée dans une mine de charbon canadienne, montrant un potentiel de gains d'efficacité opérationnelle.

Pour enrichir l'analyse, une approche non supervisée tenant compte des données des trous de forage adjacents est proposée. L'approche recommandée est basée sur un algorithme de partitionnement spectral (*clustering spectral*) qui utilise une représentation des données sous forme de graphe. Ce modèle spatial renforce la stabilité des classifications lithologiques, en fournissant une vue cohérente des propriétés géologiques dans les zones de forage proches.

Une architecture LSTM (*Long Short-Term Memory*) a été adaptée pour prédire l'état d'usure des trépans de type tricônes en fonction des signaux MWD en temps réel. Ce modèle capture les dépendances temporelles dans les données, permettant d'estimer continuellement le degré d'usure et d'anticiper le moment optimal pour remplacer les trépans. Les résultats montrent que l'approche basée sur les LSTM est particulièrement efficace pour modéliser la progression de l'usure des tricônes et fournir des prévisions de qualité dans le cadre des opérations minières.

L’utilisation d’un indice continu d’usure (BWi-LSTM) apporte une flexibilité qui réduit le besoin d’interventions humaines fréquentes. Cela contribue à la réduction des arrêts non planifiés et à une meilleure gestion des équipements.

Cette thèse propose une avancée dans l’application de l’intelligence artificielle aux opérations de forage miniers, en développant des modèles qui apportent une valeur ajoutée notamment pour les étapes subséquentes du processus minier comme le dynamitage. Ces travaux démontrent comment l’intelligence artificielle peut non seulement améliorer la précision des décisions en temps réel, mais aussi ouvrir la voie vers des plateformes de forage autonomes et plus efficaces. L'algorithme de classification de charbon, en cours de déploiement dans une mine canadienne, illustre la portée industrielle de ces modèles et leurs impacts potentiels en termes d’efficacité et de réduction des coûts.

ABSTRACT

This thesis addresses the challenges of optimizing production drilling processes in open-pit mines by leveraging Measurement While Drilling (MWD) performance data and the application of machine learning techniques. The primary objective is to enhance the accuracy of lithological identification and better anticipate tricone bit wear (cutting tool used in drilling operations), two crucial elements for optimizing operational efficiency and reducing costs associated with production drilling operations.

The study explores both supervised and unsupervised learning methods to analyze MWD data, which are often noisy, influenced by drilling conditions and geological properties, that complicate interpretation. The research proposes innovative approaches by integrating multiple supervised classification models to detect lithological changes, as well as a recurrent neural network model for tricone wear detection.

The work is organized around three main themes, each covered in a distinct chapter.

A set of supervised classification models was developed to detect the presence of coal seams from MWD data. This approach was validated in a Canadian coal mine, located in British Columbia, where the results demonstrated potential for significant operational efficiency gains to be enabled.

To enrich the analysis, an unsupervised approach that incorporates data from adjacent drill holes was proposed. The recommended approach is based on a partitionnement spectral algorithm using a graph-based data representation. This spatial model enhances the stability of lithological classifications, providing a more coherent view of geological properties in nearby drilling areas.

An LSTM (Long Short-Term Memory) network architecture was implemented to predict the wear rate and state of tricone drill bits based on real-time MWD data. This model captures temporal dependencies within the data, allowing for continuous wear estimation and optimal timing of bit replacements. Results show that the LSTM-based approach is particularly effective in modeling tricone wear progression and delivering high-quality predictions for use in mining operations.

Using a continuous bit wear index (BWi-LSTM) provides flexibility, reducing the need for frequent human intervention. This contributes to fewer unplanned downtime periods and better overall equipment management.

This thesis offers a significant advancement in applying artificial intelligence (AI) to production drilling operations, in open-pit mines, by developing and applying models that add value, particularly in subsequent stages of the mining process, such as blasting. The research demonstrates how AI can not only improve real-time decision-making accuracy but also pave the way for more efficient and autonomous drilling platforms. The coal classification algorithm, currently being deployed in a Canadian mine, illustrates the industrial relevance of these models and their potential impact on efficiency and cost reduction.

TABLE DES MATIÈRES

CHAPITRE 1 INTRODUCTION 1

1.1 Contexte du projet 1

1.2 Éléments de la problématique 2

1.3 Objectifs de recherche 3

1.4 Contributions du projet de recherche 3

1.5 Plan de la thèse 4

CHAPITRE 2 NOTIONS DE BASE 6

2.1 Notions de base sur le forage de production minière 6

2.1.1 Forage et dynamitage dans le contexte minier 6

2.1.2 Reconnaissance du type de roche 9

2.1.3 Données recueillies : Measurement-While-Drilling (MWD) 10

2.1.4 Trépans de forage de type tricônes et usure des tricônes 13

2.2 Définitions et concepts mathématiques utiles 16

2.2.1 Apprentissage automatique 16

2.2.2 Modèles d'apprentissage supervisé pour la classification 17

2.2.3 Réseaux de neurones artificiels (régression et classification) 20

2.2.4 Modèles d'apprentissage non supervisé pour la classification 24

2.2.5 Analyse en composantes principales (ACP) 31

CHAPITRE 3 REVUE DE LITTÉRATURE 33

3.1 État de l'art sur les méthodes d'identification de la lithologie basée sur les MWD 33

3.1.1 Indicateurs de classification géologique basés sur les MWD 33

3.1.2 Méthodes d'identification de la géologie basées sur l'apprentissage automatique 35

3.2 État de l'art sur la détection de l'usure des tricônes basée sur les MWD.....37

3.2.1 Synthèse de la revue.....39

CHAPITRE 4 DÉMARCHE ET ORGANISATION DE LA THÈSE..... 40

CHAPITRE 5 APPRENTISSAGE SUPERVISÉ POUR LA RECONNAISSANCE DU CHARBON 42

5.1 Introduction 42

5.1.1 Contexte et enjeux.....42

5.1.2 Objectif spécifique 43

5.2 Méthodologie et hypothèses 44

5.2.1 Approche générale.....44

5.2.2 Étude de cas.....45

5.2.3 Hypothèses principales.....46

5.2.4 Identification des couches de charbon à partir des relevés géophysiques.....47

5.3 Analyse des données.....48

5.3.1 Preprocessing 48

5.3.2 Exploration des données.....50

5.3.3 Partition des données.....53

5.4 Modèle de classification supervisés pour la prédiction du charbon 53

5.4.1 Choix des modèles de classification.....53

5.4.2 Ingénierie et sélection des caractéristiques 55

5.4.3 Résultats des tests.....58

5.4.4 Analyse de sensibilité et importance de combiner plusieurs modèles 62

5.5 Modèle de régression pour la prédiction de la densité 64

5.5.1 Réseau de neurones artificiels pour prédiction de la densité 64

5.5.2 Correction de l'erreur d'alignement entre les MWD et les relevés géophysiques...67

5.6 Détection de la position de la couche de charbon 68

5.7 Discussion sur la généralisation des modèles 71

5.8 Conclusion 72

CHAPITRE 6 APPRENTISSAGE NON SUPERVISÉ POUR LA PRISE EN COMPTE DES TROUS DE FORAGE VOISINS 73

6.1 Introduction 73

6.1.1 Contexte et enjeux 73

6.1.2 Objectif spécifique 74

6.1.3 Difficulté du problème et défis spécifiques 75

6.2 Approche proposée 76

6.2.1 Idée générale 77

6.2.2 Sélection des trous de forage 77

6.2.3 Variables d'entrées (attributs) et de sortie 78

6.3 Application des modèles de partitionnement 79

6.3.1 Méthode des k-moyennes 79

6.3.2 Modèle de mélange de gaussiennes 80

6.3.3 Partitionnement spectral avec les données MWD et les prédictions 80

6.4 Exemple de mise en pratique 84

6.4.1 Description des données 85

6.4.2 Principales hypothèses 89

6.4.3 Analyse des résultats 89

6.5 Conclusion et recommandations 93

CHAPITRE 7 DÉTECTION DE L'USURE DU TRICÔNE EN UTILISANT LES MWD PRIMAIRES 95

7.1 Introduction 95

7.1.1 Contexte et enjeux 95

7.1.2 Objectif spécifique 96

7.1.3 Difficulté du problème et défis spécifiques 96

7.2 Description des différentes approches étudiées 98

7.3 Description et exploration sommaire des données 98

7.3.1 Hypothèses principales 98

7.3.2 Exploration des données 99

7.4 Approche basée sur l'énergie de forage cumulée 104

7.5 Approche basée sur les variations de distribution du SED 105

7.6 Modèle d'apprentissage supervisé (LSTM) 112

7.6.1 Variables d'entrée, variable cible et données d'entrainement 112

7.6.2 Architecture du modèle et hyperparamètres 114

7.6.1 Analyse des résultats de tests 115

7.7 Discussion sur l'utilisation de l'indice d'usure du trépan 119

7.8 Conclusion et recommandations 120

CHAPITRE 8 CONCLUSION ET RECOMMANDATIONS 122

8.1 Synthèse des travaux 122

8.2 Contributions majeures et originalité 123

8.3 Pistes d'améliorations futures 123

LISTE DES TABLEAUX

Tableau 1 – Avantages et inconvénients des modèles de classification supervisés proposés..... 54

Tableau 2 – Performance des modèles sur l'ensemble de test..... 58

Tableau 3 – Spécifications du meilleur modèle de réseau de neurones artificiel 64

Tableau 4 – Avantages et inconvénients des méthodes de partitionnement proposées 77

Tableau 5 – Performance des modèles de classification non supervisée proposés..... 93

Tableau 6 – Comparaison des méthodes d'évaluation de fin de vie des tricônes 119

LISTE DES FIGURES

Figure 2.1 Foreuse Rotative Minière 6

Figure 2.2 Paramètres importants pour le dynamitage d'un patron de forage 8

Figure 2.3 Paramètres importants pour le dynamitage d'un patron de forage 10

Figure 2.4 Exemple de MWD fournis par le DAU pour un trou de forage (200 Hz) 12

Figure 2.5 Trépan de type tricône 13

Figure 2.6 Principaux modes de défaillance et d'usure des tricônes 15

Figure 2.7 Représentation d'un réseau de neurones artificiel 22

Figure 2.8 Représentation d'une cellule LSTM (Zhang, Lipton, Li, & Smola, 2023)..... 23

Figure 5.1 – Chargement d'explosifs dans l'extraction du charbon..... 43

Figure 5.2 - Processus d'entrainement des modèles de prédiction 45

Figure 5.3 – Données MWD brutes pour un trou de forage et relevé de densité associée..... 48

Figure 5.4 – Inférence du charbon à partir de relevé géologique pour un trou de forage..... 49

Figure 5.5 – Exemple d'agrégation et de lissage des MWD pour un trou de forage 50

Figure 5.6 – Diagrammes Box-plot et histogrammes pour chaque signal MWD 51

Figure 5.7 – Coefficients de corrélation entre les MWD transformés et la densité (complet)..... 56

Figure 5.8 – Coefficients de corrélation entre les MWD transformés et la densité (partiels)..... 57

Figure 5.9 – Comparaison des prédictions pour le trou de forage A 60

Figure 5.10 – Comparaison des prédictions pour le trou de forage C..... 61

Figure 5.11 – Analyses de sensibilité pour les différents modèles de classification 63

Figure 5.12 – Évolution de l'entraînement/validation pour le modèle de densité 65

Figure 5.13 – Prédictions de densité pour les trous de forages tests..... 66

Figure 5.14 - Ajustement de l'alignement entre les MWD et les relevés géophysiques..... 67

Figure 5.15 – Identification de la profondeur de la couche de charbon pour les trous tests 69

Figure 6.1 Exemple d'une couche de charbon non prédite dans un trou de forage (faux négatif) 74

Figure 6.2 Exemple d'un trou de forage A et des trous voisins B et C 75

Figure 6.3 Représentation du graphe 81

Figure 6.4 Trou de forage A (interprétation difficile de la probabilité de charbon) 85

Figure 6.5 Trou de forage A avec les trous voisins B et C 87

Figure 6.6 MWD associés aux trous de forage B, A et C 88

Figure 6.7 Classification finale obtenue pour le modèle des k-moyennes 90

Figure 6.8 Classification finale obtenue pour le modèle de mélange de gaussiennes 92

Figure 6.9 Classification finale obtenue pour le modèle de Partitionnement spectral 93

Figure 7.1 Exemple de données MWD pour un tricône 101

Figure 7.2 Box-plot des données MWD pour différentes catégories d'état des tricônes 103

Figure 7.3 Évolution de l'énergie totale cumulée et énergie totale normalisée pour un tricône. 105

Figure 7.4a – Valeurs de $P50$, $P75$ et $P90$ pour chaque tricône (fenêtre mobile de 120 m) 107

Figure 7.5a – Valeurs de BWi calculées pour chaque tricône (fenêtre mobile de 120 m) 110

Figure 7.6 – Exemple de différentiel de BWi (variable cible δBWi) 113

Figure 7.7 – Exemple de différentiel de BWi (variable cible) 113

Figure 7.8 - Fonction de perte pour l'entrainement et la validation du modèle LSTM 114

Figure 7.9 – Exemple de différentiel de BWi prédit par le modèle test 115

Figure 7.10 – Valeurs de BWi prédites par le modèle LSTM pour les tricônes tests 117

LISTE DES SIGLES ET ABRÉVIATIONS

ACP Analyse en Composantes Principales

ANN Artificial Neural Network

BI Blastability Index

BI-LSTM-VAE Bidirectional Long Short-Term Memory Variational Autoencoder

CBI Compensated Blastability Index

FRF Fracture reduction Factor

GMM Gaussian Mixture Model

LSTM Long Short-Term memory

MAPE Mean Absolute Percentage Error

MSE Mean Squared Error

MWD Measurement-While-Drilling (Mesures Pendant le Forage)

PD PullDown Pressure (Pression de Poussée)

PRAUC Precision Recall Area Under the Curve

RNN Recurrent Neural Network

ROC Receiver Operating Characteristic

ROP Rate Of Penetration (Taux de pénétration)

RPM Rotation Per Minute (Tours par minutes)

SED Specific Energy of Drilling (Énergie spécifique de forage)

TRQ Torque (Couple)

WOB Weight-On-Bit (Poids sur le trépan)

CHAPITRE 1 INTRODUCTION

1.1 Contexte du projet

De nos jours, il est largement reconnu que l'automatisation dans l'industrie minière est essentielle pour améliorer la productivité, l'efficacité opérationnelle, la sécurité et les impacts environnementaux dans l'industrie minière. Les foreuses rotatives modernes pour le forage minier sont équipées de divers capteurs permettant la surveillance en temps réel des paramètres de l'opération de forage. Ces mesures sont identifiées comme des MWD (*Measurement-While-Drilling*) et incluent divers paramètres de performance tels que la vitesse de rotation dénotée RPM (*Rotation Per Minute*), la vitesse de pénétration dénotée ROP (*Rate of Penetration*), le couple dénoté TRQ (*Torque*), la force totale sur le tricône dénotée WOB (*Weight-On-Bit*) ou la pression de poussée dénotée PD (*PullDown Pressure*).

Ces données MWD peuvent à leur tour être utilisées pour déterminer des informations importantes pour les opérations subséquentes. Avec la résurgence des méthodes d'intelligence artificielle telles que les réseaux de neurones et les algorithmes modernes d'apprentissage automatique, ces grandes quantités de données historiques peuvent être exploitées et utilisées pour construire des modèles de prédiction riches et fiables qui permettront par la suite d'optimiser les opérations d'extraction minière et permettront également l'automatisation des processus et de la prise de décision. Le travail effectué dans cette recherche s'inscrit dans ce contexte général et se concentre particulièrement sur les opérations de forage minier. Des nouvelles solutions basées sur les modèles d'apprentissage automatique (*machine learning*) sont étudiées et proposées pour permettre la reconnaissance du type de roche rencontrée en temps réel pendant le forage, et aussi pour détecter le niveau d'usure des trépans de forage (tricônes) en temps réel à partir des MWD.

Initialement, ce projet de recherche a été développé dans le cadre d'un stage de recherche MITACS initié avec la compagnie Peck Tech Consulting Ltd., en partenariat avec Polytechnique Montréal. Peck Tech a continué à fournir les données à travers ses partenaires industriels miniers, influençant les grandes orientations de la recherche et participant aux réflexions sur la valeur ajoutée des thèmes abordés dans cette thèse.

1.2 Éléments de la problématique

La connaissance de la lithologie dans un patron de forage minier en cours d'exploitation est importante pour les opérations de production minières. Cela aide à optimiser le dynamitage, ainsi que les opérations subséquentes du processus d'exploitation minière (extraction, transport, stockage, concassage et broyage…). Selon Mason (2011), des explosions incorrectes dans une mine sur dix en Australie coûtent l'équivalent d'une mine inutile. Pour connaître le type de géologie, les ingénieurs en dynamitage se fient souvent aux résultats des diagraphies géophysiques. Cependant, ces mesures sont rares (seulement un faible pourcentage des trous forés sont analysés), coûteuses et sujettes à des incertitudes de profondeur. De plus, elles n'arrivent pas nécessairement à temps pour la prise de décision.

Une approche proactive de gestion de l'usure des tricônes est cruciale pour minimiser les coûts et maximiser la productivité dans les mines. La connaissance en temps réel de l'usure du tricône, permettrait non seulement d'améliorer l'efficacité des opérations de forage, de maximiser la productivité du forage, d'éviter les bris catastrophiques de tricônes et de normaliser les indices de dureté de la roche calculés à partir des MWD primaires. Pour évaluer le niveau d'usure des tricônes et déterminer le moment opportun pour les remplacer, les opérateurs se basent parfois sur leur ressenti pendant le forage, sur une inspection visuelle des tricônes ou sur une mesure de la longueur des dents à l'aide de jauges manuelles. Toutes ces méthodes sont sujettes à la subjectivité de l'opérateur et ne permettent pas de standardiser le processus de prise de décision.

Les données (MWD) fournissent des mesures de plusieurs paramètres de performance de forage pendant le forage. Ces informations sont disponibles en temps réel. La géologie locale et l'état du tricône affecte directement la réponse de la foreuse et donc les mesures MWD. Ces données MWD pourraient donc ensuite être utilisées pour inférer la géologie locale dans un trou de forage (identification du type de roche) et aussi l'état (niveau d'usure) des tricônes en temps réel.

Les techniques modernes d'intelligence artificielle, telles que l'apprentissage automatique supervisé et non supervisé, permettraient la prédiction en temps réel du type de roche et de l'état du tricône en utilisant les données MWD. Il existe plusieurs approches, chacune ayant ses avantages et ses inconvénients. Ces méthodes peuvent être judicieusement combinées pour améliorer les résultats globaux.

Cette thèse s'articule donc autour des questions suivantes :

- Quelles techniques d'apprentissage automatique peuvent être adoptées pour permettre la détection de la géologie (du type de roches) en cours de forage à l'aide des données MWD de base?
- Quelles techniques d'apprentissage automatique peuvent être utilisées pour inférer en temps réel le niveau d'usure des tricônes pendant le forage à l'aide des données MWD de base?

Étant donné que les variables géologie et état du tricône affectent simultanément la réponse de la foreuse, les solutions proposées pour chaque aspect doivent être suffisamment robustes pour tenir compte de ces interactions complexes et assurer une prédiction précise et fiable.

1.3 Objectifs de recherche

Les objectifs de cette recherche sont de proposer des modèles mathématiques basés sur l'adaptation des méthodes existantes d'apprentissage automatique afin de permettre la détection du type de roche et de l'usure du trépan en quasi-temps réel. Un défi supplémentaire consiste à développer des solutions prometteuses qui ne nécessitent pas l'ajout de capteurs additionnels (vibration, pression, etc.).

Par ailleurs, il est à noter que ce travail de recherche s'intéresse principalement aux algorithmes permettant la prise de décision (*output*) à partir des paramètres d'entrée (MWD). Une fois ces algorithmes identifiés et adaptés, le déploiement de ces modèles de prédiction sur les foreuses nécessitera une expertise en architecture des systèmes et une connaissance des infrastructures et solutions informatiques en place. Ce travail ne se concentre donc pas sur les méthodes de déploiement des algorithmes.

1.4 Contributions du projet de recherche

Au niveau opérationnel, ce projet de recherche apportera des solutions nouvelles à des problèmes concrets rencontrés dans l'industrie minière en général, et en particulier pour les opérations de forage et dynamitage (détection rapide du type de roche forée et de l'état du tricône). Les solutions proposées constituent une avancée vers le forage entièrement autonome, car elles permettent une prise de décision rapide, optimale et sans intervention humaine à partir d'informations déjà disponibles pour la plupart des foreuses actuelles.

D’un point de vue académique et mathématique, ce projet propose plusieurs solutions innovantes en combinant des algorithmes d’apprentissage supervisés et non supervisés pour obtenir des prédictions plus robustes. Alors que traditionnellement, plusieurs des modèles étudiés dans cette recherche sont utilisés de manière isolée, l’intuition derrière ce travail est qu’il peut être judicieux de combiner plusieurs approches. Cette thèse propose également une nouvelle façon de modéliser le problème et de trouver les paramètres pour certaines de ces approches, notamment pour l’apprentissage non supervisé.

1.5 Plan de la thèse

Cette thèse est organisée de la manière suivante :

Le chapitre 2 introduit les notions de nécessaires à la compréhension de cette thèse. Il s’agit de notion de base sur le forage de production minière et des notions mathématiques sur l’apprentissage automatique supervisé et non supervisé.

Le chapitre 3 passe en revue les travaux de recherche antérieurs et les développements dans le domaine. Ce chapitre propose notamment une revue des indicateurs géologiques basés sur les MWD. Ensuite, il explore les méthodes basées sur les algorithmes d'apprentissage automatique. Le chapitre se poursuit avec une revue de la recherche sur l'identification de domaines géologiques basée sur une approche d'apprentissage automatique non supervisé. Enfin, il aborde la recherche sur la détection de l'usure du tricône à l'aide des données MWD.

Le chapitre 4 décrit la méthodologie générale de la thèse, expliquant comment la recherche est organisée et structurée et le lien entre les différents sous-thèmes de la recherche. Ce chapitre propose un aperçu des démarches suivies pour atteindre les objectifs de la recherche et les étapes principales de l'étude.

Le chapitre 5 présente une approche basée sur l'apprentissage supervisé pour la reconnaissance des couches de charbon. Il commence par une introduction au contexte spécifique et la problématique, suivi des objectifs précis et un résumé de la méthodologie. Il se poursuit par une application, à l’aide données in-situ, des différents modèles d’apprentissage supervisé proposés.

Le chapitre 6 traite de l'utilisation de l'apprentissage non supervisé pour analyser les interactions spatiales entre les points de données provenant de trous de forages voisins. Le chapitre commence

par définir les objectifs et les hypothèses, ainsi que les difficultés associées à cette approche de classification plus globale. Ensuite, il définit et compare trois méthodes d’apprentissage non supervisés à l’aide d’une étude de cas. La conclusion offre des recommandations basées sur les résultats de chacun des modèles proposés.

Le chapitre 7 explore la détection de l'usure du tricône à l'aide des données MWD. Il introduit les objectifs et les hypothèses de cette partie de l'étude, ainsi que les difficultés potentielles. L'approche proposée est décrite de manière progressive et une analyse des résultats obtenus est proposée. Le chapitre se termine par des recommandations pour l'application pratique du modèle proposé.

Le dernier chapitre résume les principales conclusions de la thèse et offre des recommandations pour les recherches futures et les applications industrielles. Il met en évidence les contributions majeures du projet et discute des implications pratiques des résultats obtenus.

CHAPITRE 2 NOTIONS DE BASE

Cette section définit les différentes notions nécessaires à la compréhension de cette thèse. Ces notions sont regroupées autour de deux thèmes : les notions du domaine du forage de production minière dans les mines à ciel ouvert, et celles du domaine mathématique, plus précisément de l'apprentissage automatique.

2.1 Notions de base sur le forage de production minière

Figure 2.1 Foreuse Rotative Minière [1]

2.1.1 Forage et dynamitage dans le contexte minier

Dans le contexte d'une mine à ciel ouvert, les opérations de forage et de dynamitage sont cruciales pour la fragmentation de la roche afin de faciliter son extraction et son traitement ultérieur. C'est une suite d'opérations techniquement complexes et interdépendantes qui nécessitent une bonne planification et une exécution précise pour assurer la sécurité, l'efficacité et la rentabilité des opérations minières.

[1] Note : Image provenant du site web : https://www.cat.com/en_US/products (Caterpillar Inc., 2024)

Foreuse rotative minière

La foreuse rotative minière (voir la Figure 2.1) est un équipement essentiel dans les opérations minières, surtout dans les mines à ciel ouvert. Elle est utilisée pour percer des trous dans la roche à des fins d'exploration ou de dynamitage. Cette recherche s'intéresse particulièrement aux foreuses électriques qui sont équipées d'un moteur électrique qui entraîne un tricône rotatif utilisé pour percer des roches dures avec précision et efficacité en utilisant différents niveaux de pression et de rotation pour s'adapter aux caractéristiques de la roche. Les foreuses électriques sont généralement plus efficaces et moins polluantes que les modèles fonctionnant au diesel.

Banc de mine (Bench)

Le banc de mine désigne un niveau horizontal dans une mine à ciel ouvert. Les bancs sont créés par des explosions successives et sont utilisés pour organiser le processus d'extraction de manière à maximiser l'efficacité et à minimiser les risques de glissements de terrain ou d'autres problèmes structurels. Chaque banc est une section de la mine qui a été excavée à une certaine hauteur, permettant ensuite la planification de patrons pour les opérations de forage et de dynamitage nécessaires pour l'extraction des minéraux.

Patron de forage (Pattern)

Le patron de forage se réfère au plan ou à la configuration selon laquelle les trous de forage sont placés sur un banc de mine (voir Figure 2.2). Ce plan permet d'optimiser la couverture de l'explosion et assurer une fragmentation efficace de la roche. Le patron de forage est déterminé en fonction de plusieurs facteurs, y compris la géologie de la zone, la profondeur du banc, et le type de minerai à extraire. Les types de patrons les plus courants incluent le quadrillé, le carré ou le rectangulaire, chacun adapté aux des conditions spécifiques de la mine. La Figure 2.2 présente les paramètres importants pour le dynamitage d'un patron de forage. Le recouvrement correspond à la distance entre un trou de mine et la face libre, ou entre deux rangées de trous de forage. L'espacement désigne la distance entre deux trous de mine adjacents sur la même rangée. La crête représente le sommet du banc. Le surforage (*sub-drill*) est la profondeur supplémentaire forée sous le niveau du banc pour assurer un bon fractionnement de la roche. Le bourrage supérieur consiste en un matériau non explosif placé en haut du trou pour éviter les pertes d'énergie et confiner l'explosion. Le bourrage intermédiaire est une séparation de bourrage parfois utilisée entre plusieurs charges explosives dans un même trou, permettant du dynamitage en plusieurs phases.

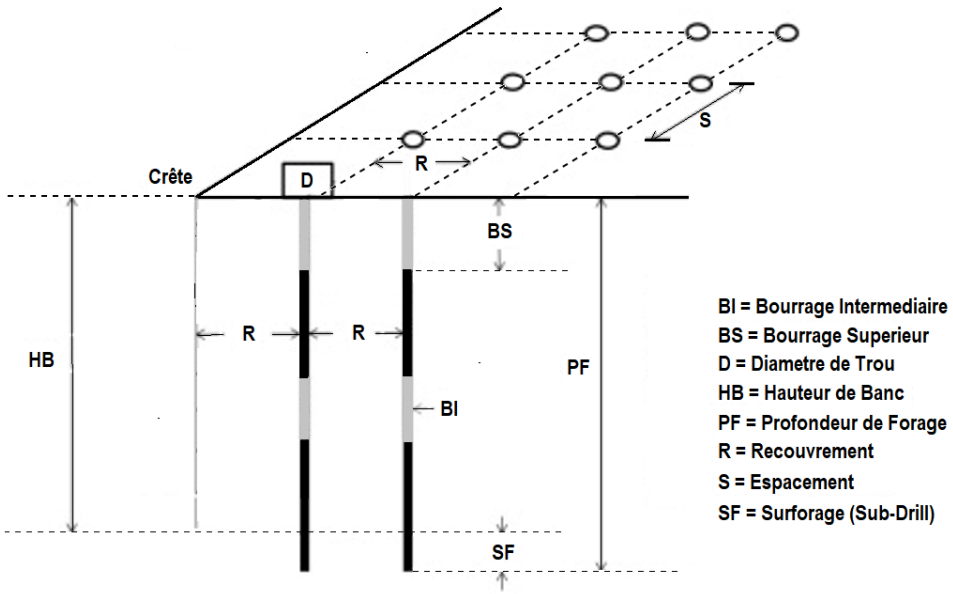


Figure 2.2 Paramètres importants pour le dynamitage d'un patron de forage[2]

Trou de forage (Blasthole)

Un trou de forage est le résultat direct de l'utilisation de la foreuse. Il s'agit d'une cavité cylindrique creusée dans le banc de mine, où les explosifs seront ensuite placés pour le dynamitage. La profondeur, le diamètre et l'espacement des trous de forage dépendent du patron de forage et sont conçus pour maximiser l'efficacité de l'explosion tout en minimisant la consommation des explosifs et les impacts négatifs sur l'environnement.

Opération de dynamitage

L'opération de dynamitage commence par le chargement des trous de forage avec des explosifs. Une fois les explosifs en place, ils sont détonés de manière séquentielle selon un plan précis qui contrôle la direction et l'ampleur de l'explosion. Cette séquence est cruciale pour contrôler la taille des fragments de roche, la stabilité du terrain et minimiser la projection de roches. Le dynamitage bien planifié permet d'extraire les matériaux de manière efficace et sécuritaire, facilitant ainsi le transport du minerai vers les usines de traitement et réduisant les coûts d'opération.

[2] Note : Image provenant de Singh et al. (2016)

2.1.2 Reconnaissance du type de roche

Domaine géologique

Un domaine géologique fait référence à une zone ou un volume spécifique dans lequel les caractéristiques de la minéralisation sont relativement uniformes. Le processus d'identification ou de modélisation des domaines géologiques est connu en anglais sous le nom de *domaining*. Le domaining joue un rôle crucial dans le forage-dynamitage minier en permettant une planification et une exécution plus précises des opérations. En créant des domaines géologiques distincts, les ingénieurs et les géologues peuvent identifier les zones où les caractéristiques de la roche et de la minéralisation sont uniformes, facilitant ainsi les quantités et types d'explosifs appropriés. Cela permet de maximiser l'efficacité de la fragmentation et prévenir les erreurs de dynamitage qui pourraient entraîner des pertes de minerai ou des contaminations indésirables du minerai.

Identification et classification de la lithologie

La lithologie est l'étude des caractéristiques physiques et de la composition des roches. La caractérisation de la roche consiste à identifier et à décrire les propriétés et la composition des roches afin de les classifier en différents types lithologiques.

La compréhension et l'identification de la lithologie ou du type de roches sont cruciales pour les opérations de forage et de dynamitage minier. La dureté et la densité des roches influencent la sélection des outils de forage (type et catégorie de tricônes) ainsi que les paramètres opérationnels du forage tels que la vitesse de rotation et la pression de poussée. Une identification de la lithologie permet d'anticiper les défis potentiels tels que l'usure des tricônes, ce qui permet d'optimiser les coûts et d'améliorer la productivité.

Diagraphies géophysiques

Les diagraphies géophysiques sont des techniques de mesure utilisées pour obtenir des informations sur les caractéristiques des formations rocheuses dans un trou de forage. Elles consistent à descendre des sondes équipées de capteurs dans des trous de forage pour enregistrer diverses propriétés physiques des roches, telles que la densité, la résistivité, la vitesse des ondes acoustiques, la radioactivité naturelle et le magnétisme en fonction de la profondeur (voir la Figure 2.3). Cela permet d'établir un profil géologique des formations traversées par le forage. Cette opération est coûteuse et chronophage, car les opérateurs doivent parfois, plus ou moins

manuellement, faire descendre les sondes dans les trous déjà forés. Elle est en général effectuée seulement pour un petit sous-ensemble de trous de forage.

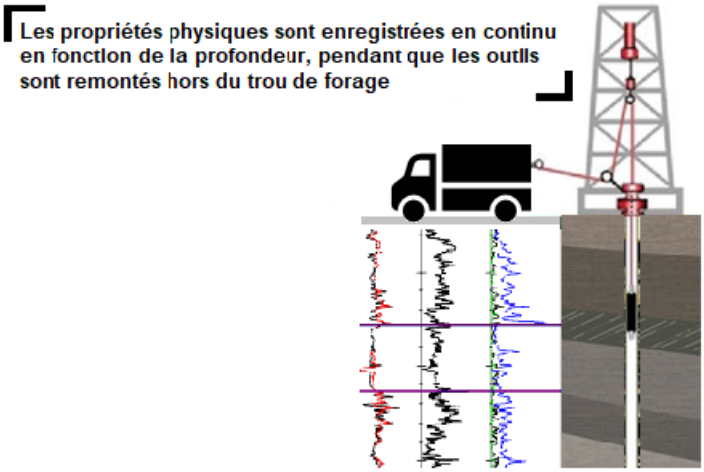


Figure 2.3 Paramètres importants pour le dynamitage d'un patron de forage[3]

2.1.3 Données recueillies : Measurement-While-Drilling (MWD)

Dans un sens large, le terme MWD désigne toutes les mesures relevées pendant les opérations de forage. Les principaux paramètres MWD d'intérêt sont ceux qui se rapportent aux paramètres d'opération de la foreuse et de performance globale du forage. En surveillant et en ajustant ces paramètres d'opération en temps réel, les opérateurs peuvent améliorer l'efficacité du forage. Il s'agit essentiellement de :

- La vitesse de rotation (RPM - *Rotations Per Minute*) : Il s'agit du nombre de rotations complètes que le tricône effectue en une minute. Le RPM influence directement l'efficacité de la coupe et la durée de vie du tricône. Une vitesse RPM appropriée permet de maximiser la pénétration dans la roche tout en minimisant l'usure du tricône et les risques de défaillance mécanique.
- La vitesse de pénétration (ROP - *Rate of Penetration*): La vitesse à laquelle le tricône avance dans le sol, généralement mesurée en mètres par heure ou en pieds par heure. C'est un indicateur clé de la performance de forage. Une ROP élevée peut indiquer une roche

[3] Note : Image provenant de Mohammed (2018)

tendre ou une efficacité de forage élevée, tandis qu'une ROP faible peut signaler des formations rocheuses dures ou des problèmes opérationnels.

- Le couple (TRQ - *Torque*) : Il s'agit de la force de rotation appliquée au tricône, mesurée en Newton-mètres (Nm). Le couple est crucial pour comprendre la résistance de la roche et l'efficacité du forage. Un couple élevé peut indiquer une roche dure ou des problèmes de tricône, tandis qu'un couple bas peut indiquer une roche plus tendre ou une performance de tricône optimale.
- La force totale sur le tricône (WOB - *Weight-On-Bit*) : Il s'agit de la force appliquée au tricône pour le pousser dans la roche, mesurée en newtons (N). Le WOB influence directement la ROP et l'efficacité du forage. Une force adéquate permet de maximiser la pénétration tout en minimisant l'usure du tricône. Une force trop faible ou trop élevée peut entraîner une inefficacité de forage ou des dommages au matériel.
- La pression de poussée (PD - *PullDown Pressure*) : Il s'agit de la pression (exprimé sous forme de force en *kN*) appliquée pour forcer le tricône à pénétrer dans la roche. Elle aide à ajuster la force exercée sur le tricône pour s'adapter aux variations de la dureté de la roche et garantir une performance de forage optimale. Le WOB inclut l'effet du PD et le poids du cylindre et toute la structure au-dessus du tricône (*Static Weight*).
- La pression (d'air) de purge (*Bailing Air Pressure*) : C'est la pression d'air envoyée à l'interface trépan/roche afin d'expulser la poussière et les débris de forage hors du trou de forage pour assurer un forage efficace et refroidir également le trépan.

En résumé, pour la plupart des foreuses électriques rotatives, RPM et WOB sont des commandes d'entrée définies par l'opérateur ou le système de contrôle de la foreuse et ajustées par des consignes (*setpoints*) pour optimiser le forage, tandis que ROP et TRQ sont des réponses qui fournissent des informations sur l'efficacité et les conditions de forage.

Plusieurs autres données MWD supplémentaires peuvent exister (vibration ou température), selon les capteurs installés sur une foreuse, tels que les capteurs de vibration ou de température. La fréquence à laquelle sont relevés les MWD varie en fonction des systèmes et des types de foreuses. Certaines foreuses sont équipées de systèmes qui fournissent des données à une fréquence de 10 Hz. Il existe certaines solutions secondaires, comme l'Unité d'Acquisition de Données (DAU) de Peck Tech Consulting, qui sont installées sur une foreuse et permettent la capture des données MWD à des fréquences 10 à 20 fois plus élevées (voir Figure 2.4).

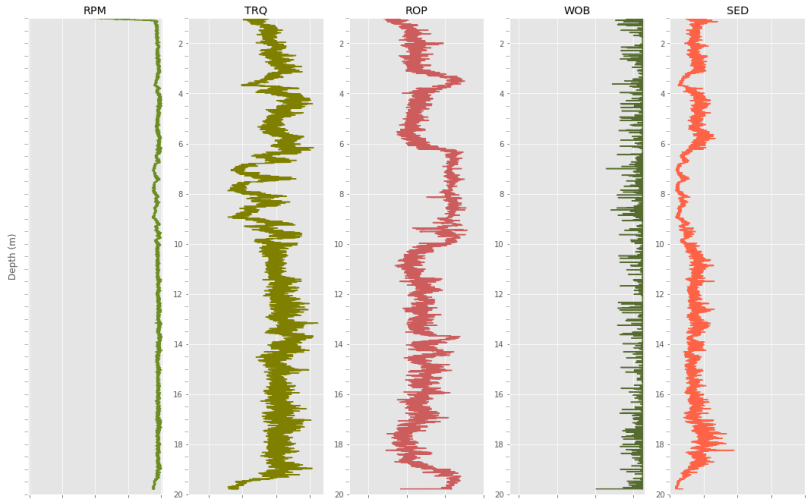


Figure 2.4 Exemple de MWD fournis par le DAU pour un trou de forage (200 Hz)

Le DAU fournit également d'importants MWD secondaires tels que :

- L'énergie spécifique de forage (SED - *Specific Energy of Drilling*) : Il s'agit d'un indicateur proposé par Teale (1965) qui mesure de l'énergie nécessaire pour fracturer une roche donnée et se définit comme suit :

  $$SED = \frac{TRQ \times RPM + WOB \times ROP}{A \times ROP} \quad (1)$$

  où A est le diamètre du trou (m²). SED est en fait exprimé sous forme de pression (en kPa ou MPa).

- L'indice de blastabilité (BI - *Blastability Index*) : Un indice normalisé dérivé du SED qui désigne la résistance de la roche à la fragmentation par explosion (Babaei Khorzoughi, 2013) :

  $$BI = \frac{SED}{BI_{NORM}} \times 100 \quad (2)$$

  où BI_{NORM} est une constante de normalisation permettant d'obtenir un indice compris entre 0 et 100.

- Le facteur de réduction des fractures (FRF - *Fracture Reduction Factor*): Il s'agit d'un coefficient utilisé pour quantifier la présence de fractures préexistantes dans la roche forée (Peck Tech, 2009). Plus une roche est fracturée, moins elle requiert d'énergie pour la fragmenter car les fractures préexistantes facilitent la propagation des ondes de choc.

- L'indice de blastabilité compensé (CBI – *Compensated Blastability Index*) : C'est une version ajustée du BI, intégrant l'effet des fractures préexistantes dans la roche à travers le facteur FRF (Peck Tech, 2009). Il est défini par.

$$CBI = BI \times 100 \quad (3)$$

2.1.4 Trépans de forage de type tricônes et usure des tricônes

Trépans de type tricônes

Les trépans, parfois appelés forets, sont les outils de perçage pour le forage minier. Cette recherche s'intéresse particulièrement aux trépans de type tricônes (ou trépans à trois cônes roulants), qui sont couramment utilisés dans les opérations de forage pour l'exploitation minière (voir la Figure 2.5). Ils sont composés de trois rouleaux coniques, chacun équipé de rangées de « dents » ou d'éléments coupants (souvent en carbure de tungstène), conçus pour écraser et briser les formations rocheuses au fur et à mesure que le tricône tourne. Les cônes sont généralement montés sur des roulements qui leur permettent de tourner indépendamment pendant que le tricône lui-même est en rotation.

Figure 2.5 Trépan de type tricône[4]

Classification IADC des tricônes

L'*International Association of Drilling Contractors* (IADC) propose un système d'identification et catégorisation des tricônes en fonction de leur conception et de leur utilisation prévue selon les

[4] Note : Image provenant de *Roller Cone Dull Grading Manual* de Smith Technologies (2008)

conditions de forage spécifiques. Ce système utilise un code en trois parties pour décrire les caractéristiques du tricône (voir Thomas (2008)) :

- Série du Tricône : Chiffre de 1 à 8 qui indique le type général de formation pour lequel le tricône est conçu.
- 1-3 : Formations tendres (ex : argiles, schistes)
- 4-6 : Formations moyennes
- 7-8 : Formations dures (ex : quartzite)
- Type d'élément coupants : Chiffre de 1 à 4 qui décrit le type de structure de coupe sur le tricône (sélectionné en fonction de la dureté et l'abrasivité des roches).
- *1* : Tricônes à dents en acier pour formations tendres
- *2* : Tricônes à dents en acier pour formations moyennes
- *3* : Tricônes à dents en acier pour formations dures
- *4* : Tricônes à inserts en carbure de tungstène
- Type de Roulements : Chiffre de 1 à 7 qui indique le type de roulement, de scellé et d'autres caractéristiques spéciales du tricône.
- *1-2 :* Roulements à rouleaux standard
- *3-4 :* Roulements à rouleaux standard avec protection de la jauge
- *5-6 :* Roulements à rouleaux scellés
- *7 :* Roulements à journal scellés

Des codes de lettres supplémentaires peuvent être ajoutés au code de base en trois parties pour indiquer des caractéristiques spéciales ou des modifications du tricône.

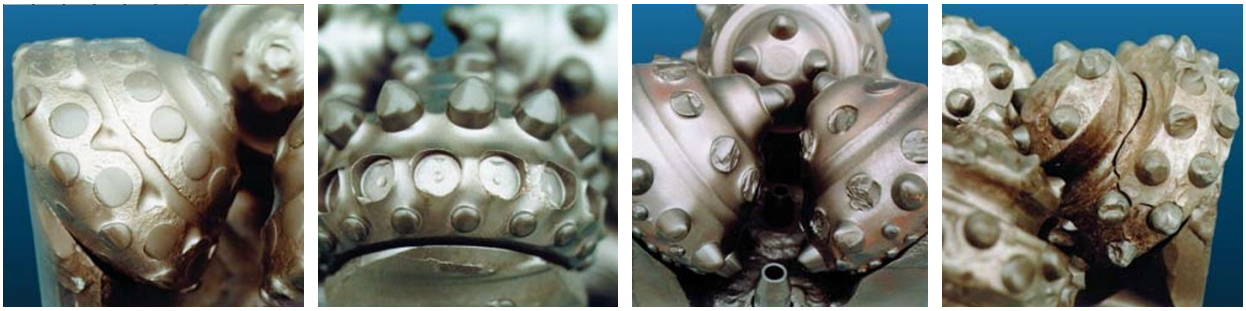
Usure et modes de défaillances des tricônes

Comprendre les modes de défaillance des tricônes est essentiel pour améliorer leur longévité et l'efficacité du forage. Les modes de défaillance les plus courants des tricônes incluent :

- Dents usées (*Worn teeth*) : Il s'agit de la réduction de la hauteur des éléments coupants due à l'action de forage. C'est le mode d'usure normal et attendu (voir la Figure 2.5-a).
- Dents perdues (*Lost teeth*) : Il s'agit de la perte d'un ou de plusieurs éléments coupants. La perte d'un élément coupant dans une rangée peut réduire les forces de rétention des éléments coupants restants (voir la Figure 2.5-b).

- Dents cassées (*Broken teeth*) : Lorsque plus du tiers d'un élément coupant est cassé. C'est une usure typique pour les trépans à dents en carbure de tungstène dans certaines conditions (voir la Figure 2.5-c).
- Cône fissuré (*Cracked cone*) : Condition où un ou plusieurs cônes sur un tricône développent des fissures (voir la Figure 2.5-d).

Plusieurs autres modes de défaillance existent. Ces différents types de défaillance et d'usure ne sont pas tous mutuellement exclusifs.



a- Dents usées b- Dents perdues c- Dents cassées d- Cône fissuré

Figure 2.6 Principaux modes de défaillance et d'usure des tricônes[5]

L'IADC propose également un système de classement de l'usure pour systématiser l'évaluation et l'enregistrement de l'usure et des dommages d'un tricône. Il s'agit d'une matrice comprenant plusieurs catégories pour décrire l'état du tricône et qui fournit un aperçu complet de l'état du tricône. Cela aide à analyser la performance de forage et à identifier les problèmes récurrents.

Connaître le niveau d'usure et l'état des tricônes est essentiel pour optimiser les opérations de forage ainsi que pour garantir la sécurité et l'efficacité des opérations minières. En effet, un tricône usé peut entraîner une diminution de la performance du forage, des coûts supplémentaires liés à des temps d'arrêt imprévus pour le retrait et le remplacement du tricône. En outre, les fragments métalliques provenant de tricônes brisés peuvent se retrouver dans les étapes ultérieures du processus minier. La présence de ces débris peut causer des dommages sévères aux équipements de broyage et de concassage, entraînant des arrêts de production coûteux et des réparations onéreuses. De plus, un tricône usé peut affecter la précision des données recueillies lors du forage notamment des indices sur la dureté de la roche, ce qui peut compromettre l'exactitude des analyses géologiques et des décisions opérationnelles subséquentes.

[5] Note : Image provenant de *Roller Cone Dull Grading Manual* de Smith Technologies (2008)

2.2 Définitions et concepts mathématiques utiles

2.2.1 Apprentissage automatique

L'apprentissage automatique (ou *machine learning* en anglais) est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur le développement d'algorithmes et de techniques permettant aux ordinateurs d'apprendre à partir de données et de faire des prédictions ou de l'inférence sans être explicitement programmés pour cela.

Problème de classification : Le problème de classification en apprentissage automatique consiste à assigner des étiquettes (catégories) de type $\mathrm{y} \in \{0,1,2,\dots,C-1\}$, où C est le nombre de classes distinctes, à des variables d'entrée de type $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

Dans le cas de la classification binaire ($C = 2$), l'objectif est de trouver une fonction $f : \mathbb{R}^n \rightarrow \{0,1\}$ qui prédit la probabilité que $\mathbf{x}$ appartienne à la classe $\mathrm{y} = 1$. Cette probabilité est souvent modélisée par une fonction sigmoïde :

$$\hat{\mathrm{y}} = P(\,y = 1 \mid \mathbf{x}\,) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b) \tag{4}$$

où $\sigma(z) = \frac{\mathbf{1}}{\mathbf{1}+e^{-z}}$ est la fonction sigmoïde qui transforme les entrées en valeurs entre 0 et 1,

$\mathbf{w}$ est le vecteur de poids et b est le biais.

Lorsque des données étiquetées sont disponibles (apprentissage supervisé), il est possible de définir une fonction de perte (ou fonction de coût) qui est une mesure de l'écart entre les prédictions du modèle et les valeurs ou étiquettes réelles des observations (données d'entraînement). La fonction de perte couramment utilisée pour les problèmes de classification est l'entropie croisée définie par:

$$L(\hat{\mathrm{y}}, \mathrm{y}) = -[\mathrm{y}\log(\hat{\mathrm{y}}) + (1-\mathrm{y})\log(1-\hat{\mathrm{y}})] \tag{5}$$

où $\hat{\mathrm{y}}$ représente les classes prédites et y les classes observées.

L'objectif est de trouver un modèle et les paramètres associés qui minimisent la fonction de perte.

Problème de régression : Les modèles de régression en apprentissage automatique sont utilisés pour prédire des valeurs continues plutôt que des catégories. Cela consiste à prédire une variable continue y à partir d'un ensemble de variables d'entrée $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Le but est de trouver une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ qui approxime la relation entre $\mathbf{x}$ et y de manière à minimiser l'erreur (fonction de perte) entre les valeurs prédites $\hat{y}$ et les valeurs réelles y (observations).

Une fonction de perte couramment utilisée pour la régression est l'erreur quadratique moyenne MSE (*Mean Squared Error*) définie par :

$$L(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (6)$$

où N est le nombre total d'échantillons dans l'ensemble de données (données d'entrainement),

y_{i} est la valeur réelle pour l'exemple i,

$\hat{y}_{\mathrm{i}}$ est la valeur prédite par le modèle.

2.2.2 Modèles d'apprentissage supervisé pour la classification

L'apprentissage supervisé est une méthode d'apprentissage automatique dans laquelle un modèle est formé en utilisant un ensemble de données étiquetées (ou observations). Le modèle apprend à associer les entrées aux sorties en minimisant l'erreur entre les prédictions du modèle et les étiquettes réelles à travers plusieurs itérations (fonction de perte). L'objectif final est de permettre au modèle de généraliser à de nouvelles données non vues et de faire des prédictions précises sur celles-ci.

Les sous-sections qui suivent présentent quelques modèles d'apprentissages supervisés qui seront utilisés dans cette recherche pour la classification (prédiction de catégories).

2.2.2.1 Régression logistique

La régression logistique (Hosmer Jr, Lemeshow, & Sturdivant, 2013) est l'un des plus anciens et des plus largement utilisés algorithmes de classification. Elle est basée sur la fonction logit :

$$Log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n. \quad (7)$$

où $P = P(y = 1)$ est la probabilité d'appartenir à la classe 1, β_0 est l'ordonnée à l'origine et β_i est le coefficient associé à la variable d'entrée x_i. Son avantage réside dans sa simplicité et le fait qu'elle ne nécessite pas beaucoup de données pour l'entraînement. Cependant, elle peut être limitée lorsqu'il s'agit de détecter des relations non linéaires complexes entre les variables (Tu, 1996).

2.2.2.2 Arbres de décision

Les arbres de décision (Meisel & Michalopoulos, 1973) sont des outils de classification avec une structure en forme de graphe, où chaque nœud applique une règle de séparation basée sur un test sur l'une des caractéristiques (variable d'entrée). Les nœuds terminaux (feuilles) sont assignés à une étiquette de classe. Le grand avantage des méthodes basées sur les arbres de décision est qu'elles peuvent produire des résultats même en présence de valeurs manquantes dans les paramètres d'entrée. Cependant, comme observé par Klyuchnikov et al. (2018), ils sont de mauvais interpolateurs de données et ne généralisent pas bien dans les régions de l'espace des caractéristiques où la densité des données d'entraînement est faible.

2.2.2.3 Gradient boosting

Le *Gradient Boosting* (Friedman, 2002) est une technique consistant à construire un classificateur fort en attribuant itérativement des poids plus élevés aux erreurs des points de données mal classés. Le *Gradient Boosting* est très utile lorsqu'il y a un fort déséquilibre des classes. *L'Extreme Gradient Boosting* est une implémentation particulière du *Gradient Boosting* (Chen, He, Benesty, Khotilovich, & Tang, 2015) qui intègre plusieurs optimisations et améliorations pour accroître la vitesse et les performances. Cependant, cette méthode nécessite un ensemble de données plus large que d'autres modèles plus simples.

2.2.2.4 Bagging

Le *bagging* ou *bootstrap aggregating* (Breiman, 1996) consiste à construire un classificateur fort en agrégeant ou en moyennant les prédictions de plusieurs classificateurs individuels (comme des arbres de décisions) qui ont été entraînés en utilisant différents sous-ensembles de l'ensemble de données d'entraînement. Cette méthode produit un classificateur par ensemble plus stable et fiable, mais nécessite également un grand ensemble de données pour fonctionner correctement.

2.2.2.5 k-Nearest Neighbors

Le *k-Nearest Neighbors* (k-NN) (Peterson, 2009) consiste à assigner à un point de donnée la classe correspondant à la majorité de ses *k* plus proches voisins en distance dans l'espace des caractéristiques dans l'ensemble d'entraînement. C'est un outil simple mais puissant qui ne nécessite pas beaucoup de données d'entraînement et peut gérer des relations non linéaires complexes entre

les caractéristiques. Cependant, il est naturellement sujet au surapprentissage et peut être trompeur dans les zones de l'espace des caractéristiques avec des densités de classe similaires.

2.2.2.6 Machines à Vecteurs de Support (SVM)

Les machines à vecteurs de support (SVM) (Cortes & Vapnik, 1995) sont des algorithmes de classification puissants qui cherchent à trouver l'hyperplan optimal séparant les différentes classes dans un espace de caractéristiques. Cet hyperplan maximise la marge entre les classes, ce qui minimise le risque de mauvaise classification. Les SVM peuvent être adaptés à des problèmes non linéaires en utilisant des noyaux (kernels) qui transforment les données d'entrée en un espace de dimension supérieure où un hyperplan linéaire peut être trouvé. Les SVM sont particulièrement efficaces dans des espaces de grande dimension et conviennent aux problèmes où le nombre de dimensions est supérieur au nombre d'échantillons. Cependant, ils peuvent être moins performants sur des ensembles de données très larges en raison de leur complexité de calcul.

2.2.2.7 Mesure de la performance des modèles de classification

Cette section décrit les métriques de performance proposées pour évaluer la performance des modèles de classification du charbon.

La matrice de confusion de la classification est constituée des éléments suivants:

- Les Vrais Positifs (VP) représentent les points de données correspondant au charbon qui ont été correctement classifiés comme Charbon.
- Les Vrais Négatifs (VN) représentent les points de données correspondant à la roche clastique (Pas Charbon) qui ont été correctement classifiés comme tels.
- Les Faux Positifs (FP) représentent les points de données correspondant au charbon qui ont été incorrectement classifiés comme Charbon.
- Les Faux Négatifs (FN) représentent les points de données correspondant à la roche clastique (Pas Charbon) qui ont été incorrectement classifiée comme Charbon.

Ces éléments permettent de calculer les indicateurs suivants :

- **L'exactitude** (*Accuracy*) : c'est une mesure de l'erreur systématique ou du niveau de proximité entre une prédiction et les valeurs réelles.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

La Précision et le Rappel (Saito, 2015) sont deux autres indicateurs qui permettent de juger certaines qualités spécifiques d'un modèle de classification.

- La **précision** (*Precision*) représente la capacité du modèle à être précis lorsqu'il classe une donnée de .
- $Precision = \frac{TP}{TP+FP}$ (9)
- Le **rappel** (*Recall*) mesure la capacité du modèle à ne manquer aucune instance de la classe d'intérêt.
- $Recall = \frac{TP}{TP+FN}$ (10)

La Précision et le Rappel peuvent-être parfois insuffisantes pour évaluer la qualité d'un modele sur tout en situation de classes débalancées. Un bon compromis est le F1-Score (Davis & Goadrich, 2006), qui est la moyenne harmonique de la précision et du rappel et représente la qualité globale d'un modèle:

$$F1 - Score = 2\frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

Pour identifier le meilleur seuil de probabilité à adopter pour la classification, le critère considéré est le PRAUC : *Precision Recall Area Under Curve* (Saito, 2015). Le PRAUC est une mesure d'évaluation utilisée pour quantifier la performance d'un modèle de classification binaire, en particulier lorsque les classes sont déséquilibrées. Elle correspond à l'aire sous la courbe tracée entre la précision et le rappel pour différents seuils de décision. Cette aire varie entre 0 (mauvais modèle) et 1 (modèle parfait). Elle se concentre sur la performance du modèle à détecter la classe positive, ce qui la rend plus informative dans les cas où la classe positive est peu fréquente.

2.2.3 Réseaux de neurones artificiels (régression et classification)

Les réseaux de neurones (Rumelhart, Hinton, & Williams, 1986) sont des modèles d'apprentissage automatique inspirés par la structure et le fonctionnement du cerveau humain. Ils peuvent être utilisés pour la régression et la classification.

2.2.3.1 Réseaux de neurones simples

Dans leur forme la plus simple (voir Figure 2.7), ils sont composés de couches de neurones artificiels (unités de calcul) interconnectés par des poids. Ces couches comprennent généralement une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Les données d'entrée

passent à travers le réseau couche par couche (propagation avant). Chaque neurone effectue une opération de pondération et applique une fonction d'activation pour produire une sortie qui devient l'entrée pour les neurones de la couche suivante.

L'entrainement de ces modèles peut se faire en deux phases : la propagation avant et la rétropropagation.

Dans la propagation avant, chaque neurone k dans une couche donnée j reçoit les entrées $a_1^{(j-1)}, a_2^{(j-1)}, \dots, a_n^{(j-1)}$ des n neurones de la couche précédente ($\mathbf{x} = x_1, x_2, \dots, x_n$ pour la première couche), et calcule la somme pondérée z_k^j en utilisant les poids $w_{k,1}^j, w_{k,2}^j, \dots, w_{k,n}^j$:

$$z_k^j = \sum_{i=1}^{n} w_{k,i}^j a_i^{j-1} + b_k^j \tag{12}$$

Cette somme pondérée z_k^j est ensuite transformée par une fonction d'activation non linéaire σ pour produire la sortie du neurone :

$$a_k^j = \sigma(z_k^j) = \frac{1}{1+e^{-z_k^j}} \tag{13}$$

Où σ est une fonction d'activation comme la sigmoïde qui transforme l'entrée en une sortie entre 0 et 1; La *ReLU* (*Rectified Linear Unit* : $\max(0, z)$), ou une tangente hyperbolique sont d'autres exemples de fonctions d'activation.

Le processus se propage ainsi jusqu'à la sortie finale $\hat{y}$.

Après la propagation avant, l'erreur totale entre l'ensemble des sorties prédites par le réseau $\hat{\mathbf{y}}$ et les valeurs ou étiquettes réelles $\mathbf{y}$ est calculée à l'aide d'une fonction de perte $L(\hat{\mathbf{y}}, \mathbf{y})$. Cette erreur est utilisée pour ajuster les poids des connexions dans le réseau (rétropropagation). L'erreur est ainsi propagée à rebours à travers le réseau, de la couche de sortie à la couche d'entrée, et les poids sont ajustés en utilisant des techniques d'optimisation comme la descente de gradient pour minimiser l'erreur.

$$w_{k,1}^j = w_{k,1}^j - \eta \frac{\partial L}{\partial w_{k,1}^j} \tag{14}$$

où η est le taux d'apprentissage, un hyperparamètre qui contrôle l'ampleur des ajustements des poids $w_{k,i}^j$ à chaque itération.

Les réseaux de neurones, notamment les réseaux profonds (plusieurs couches internes) ont la capacité de modéliser des relations non linéaires complexes. Ils offrent une flexibilité et une adaptabilité à différents types de données et de tâches. Ils requièrent toutefois beaucoup de données et sont sujets à des risques de surapprentissages. Le choix des hyperparamètres et de l'architecture des réseaux de neurones peut également être complexe. Les réseaux de neurones peuvent être utilisés pour la classification comme pour la régression.

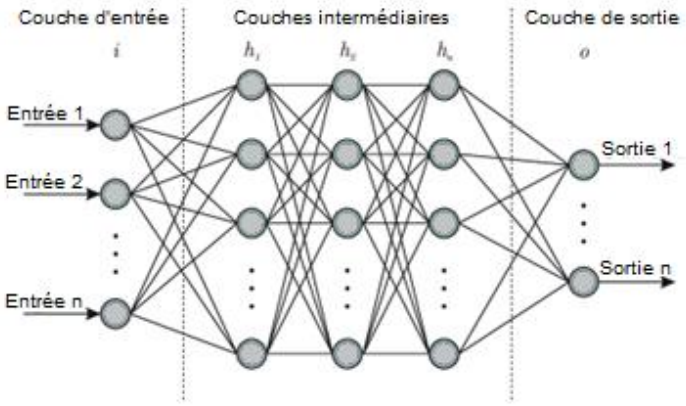


Figure 2.7 Représentation d'un réseau de neurones artificiel

2.2.3.2 Réseaux Long Short-Term Memory (LSTM)

Les réseaux de neurones récurrents (RNN) sont une classe de réseaux de neurones conçue pour traiter des données séquentielles, comme le texte et les séries temporelles. Contrairement aux réseaux de neurones classiques (ANN) qui traitent chaque donnée indépendamment, les RNN possèdent des boucles dans leur architecture qui permettent de retenir l'information d'un pas de temps à l'autre. Cela confère aux RNN la capacité de mémoriser les informations des étapes précédentes. Cependant, tel que discuté dans Bengio et al. (1994), les RNN traditionnels souffrent de problèmes de gradient lors de l'apprentissage. Lors de l'entraînement par rétropropagation à travers le temps, les gradients des erreurs de couches successives sont multipliés les uns par les autres. Si ces gradients sont des valeurs inférieures à *1*, ils diminuent au fur et à mesure qu'ils sont propagés dans le réseau, et après un certain nombre d'étapes, ils ne permettent plus d'ajuster efficacement les poids des couches éloignées. À l'inverse, si les gradients sont supérieurs à *1*, ils peuvent s'accumuler rapidement, causant des valeurs extrêmement grandes dans le réseau et le rendant instable et incapable de converger.

Les LSTM introduits par Hochreiter et Schmidhuber (1997) ont été conçus pour résoudre les limitations des RNN standards. La structure interne d'un LSTM repose sur une cellule mémoire qui transporte les informations de manière continue, régulée par des portes spécifiques, permettant ainsi un flux d'informations contrôlé. Chaque cellule LSTM contient quatre composants principaux (voir Figure 2.8) : la porte d'oubli (F_t), la porte d'entrée (I_t), la porte de sortie (O_t) et l'état de cellule (C_t).

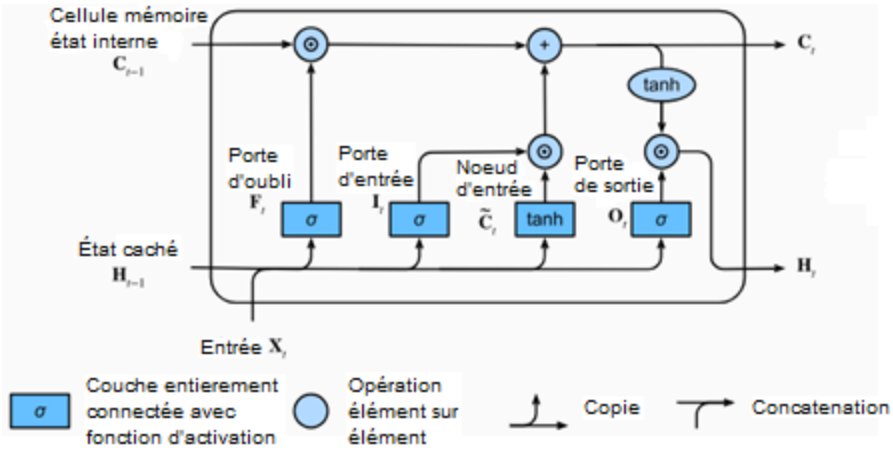


Figure 2.8 Représentation d'une cellule LSTM (Zhang, Lipton, Li, & Smola, 2023)

À partir d'une entrée donnée $\mathbf{x}_t$, un état caché précédent h_{t-1}, et un état de cellule précédent C_{t-1}, la cellule LSTM suit les étapes suivantes :

- La porte d'oubli (F_t): décide des informations de l'état de cellule précédent C_{t-1} à oublier ou conserver et est calculée comme suit:

  $$F_t = \sigma\left(W_f \cdot [H_{t-1}, \mathbf{x}_t] + b_f\right) \tag{15}$$

  où F_t est le vecteur de la porte d'oubli (valeurs entre 0 et 1), W_f et b_f sont les poids et biais associés au réseau, et σ est la fonction sigmoïde. Si F_t est proche de 0, l'information est oubliée.

- La porte d'entrée (I_t) : sélectionne les nouvelles informations qui seront ajoutées à l'état de cellule en deux étapes :

  Calcul de l'actualisation candidate : $\tilde{C}_t = tanh(W_C \cdot [H_{t-1}, \mathbf{x}_t] + b_C)$ (16)

  Décision de mise à jour : $I_t = \sigma(W_i \cdot [H_{t-1}, \mathbf{x}_t] + b_i)$ (17)

L'état de cellule est mis à jour en combinant les informations retenues par la porte d'oubli et les nouvelles informations ajoutées par la porte d'entrée :

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{18}$$

- La porte de sortie (O) : régule les informations qui vont être sorties par la cellule et est calculée comme suit :

$$O_t = \sigma(W_o \cdot [H_{t-1}, \mathbf{x}_t] + b_o) \tag{19}$$

- L'état caché actuel H_t est défini en appliquant une fonction $tanh$ sur l'état de cellule et en le multipliant par O_t:

$$H_t = O_t * tanh\,(C_t) \tag{20}$$

La sortie finale de la cellule LSTM pour cette étape est donc h_t, et elle est transmise en tant qu'état caché en entrée de l'étape suivante, de même que l'état de cellule C_t.

Les LSTM sont capables de maintenir des informations pertinentes sur des intervalles de temps prolongés, grâce à ces mécanismes de portes et elles peuvent sélectionner quelles informations garder, oublier, ou transmettre à la sortie. En raison de leur capacité à apprendre des séquences complexes, les LSTM sont largement utilisés dans les domaines de la génération de texte, et la prédiction de séries temporelles. Toutefois le nombre élevé de paramètres rend l'entraînement des LSTM coûteux surtout pour des séquences longues. De plus, en raison de leur structure complexe, il est souvent difficile de comprendre exactement comment les LSTM traitent l'information.

2.2.4 Modèles d'apprentissage non supervisé pour la classification

L'apprentissage non supervisé est une méthode de classification (voir section 2.2.1) où le modèle est formé sans utiliser de données étiquetées $\boldsymbol{y}$. Le modèle explore les structures et les motifs intrinsèques de l'ensemble $X = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$ des données d'entrée pour découvrir des regroupements ou des associations cachés. Les algorithmes courants incluent le clustering (comme la méthode des k-moyennes). L'apprentissage non supervisé est couramment utilisé pour la segmentation, la détection d'anomalies et l'analyse exploratoire de données permettant de révéler des structures sous-jacentes sans connaissance *a priori* sur les données. Il existe plusieurs méthodes de clustering. Dans cette recherche, trois méthodes seront sélectionnées et adaptées pour améliorer

l'interprétation des résultats du type de roche (charbon) prédit par les modèles supervisés. Ces trois méthodes sont décrites dans les sections ci-dessous.

2.2.4.1 Méthode des k-moyennes

La méthode des k-moyennes (*k-means*) (MacQueen, 1997) est un algorithme de clustering non supervisé utilisé pour partitionner un ensemble $X = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$ de M données, en k groupes ou clusters $(C_1, C_2, \dots, C_k)$, où chaque groupe C_i est défini par son point central $\boldsymbol{\mu}_i$ (ou centroïde). L'objectif est généralement de minimiser la variance intra-cluster, c'est-à-dire la somme des distances au carré entre chaque point de données et le centre du cluster auquel il appartient.

La fonction de coût que l'algorithme minimise peut se définir comme suit :

$$J = \sum_{i=1}^{k} \sum_{\mathbf{x}^t \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 \tag{21}$$

Où : $\boldsymbol{J}$ est la somme des erreurs quadratiques,

C_i représente le i-ème cluster,

$\mathbf{x}^{\boldsymbol{t}}$ est un point de données dans le cluster C_i, $(t \in \{1, \dots . M\})$

$\boldsymbol{\mu}_i$ est le centroïde du cluster C_i,

$\|\mathbf{x}^{\boldsymbol{t}} - \boldsymbol{\mu}_i\|$ est la distance (euclidienne) entre le point $\mathbf{x}^{\boldsymbol{t}}$ et le centroide $\boldsymbol{\mu}_i$.

L'objectif de l'algorithme est de trouver les centroïdes $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_k$ qui minimisent J.

La fonction de distance peut prendre plusieurs formes. Dans l'exemple ci-dessus, la distance euclidienne est utilisée.

L'algorithme commence par la sélection initiale d'un nombre prédéfini de k points initiaux parmi l'ensemble de données, qui serviront de centroïdes initiaux pour les k clusters.

À chaque itération, chaque point de données est assigné au cluster C_i dont le centre est le plus proche en termes de distance dans l'espace des caratéristiques.

$$C_i = \left\{\mathbf{x}^{\boldsymbol{t}} : \|\mathbf{x} - \boldsymbol{\mu}_i\| \leq \left\|\mathbf{x}^{\boldsymbol{t}} - \boldsymbol{\mu}_j\right\|, \forall j = 1, \dots, k\ ,, \forall t = 1, \dots, M\right\} \tag{22}$$

La position de chaque centroïde est mise à jour après chaque itération (moyenne des points de données assignés à ce cluster) :

$$\boldsymbol{\mu}_i = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{x}^{\boldsymbol{t}} \tag{23}$$

Les étapes d'assignation et de mise à jour sont répétées jusqu'à convergence (les centroïdes ne changent plus de manière significative).

Le nombre de clusters k est déterminé à l'avance. Il peut être évalué par essais et erreurs en évaluant le coefficient de silhouette moyen (voir Section 2.2.4.3) pour tous les points de données.

L'avantage de la méthode des *k-moyennes* est sa simplicité et sa facilité de mise en œuvre. Elle est également efficace pour les grands ensembles de données. Cependant, tel que mentionné dans Arthur & Vassilvitskii (2007), elle peut nécessiter de spécifier le nombre de clusters k à l'avance et peut être sensible aux valeurs initiales des centroïdes. Elle peut également être influencée par les points aberrants (*outliers*) et peut ne pas bien fonctionner pour des clusters de tailles très différentes.

2.2.4.2 Modèle de mélange gaussiens

La méthode du mélange de modèles gaussiens (GMM) est une technique de partitionnement qui modélise la distribution des données comme un mélange de plusieurs distributions gaussiennes. Contrairement à des méthodes de clustering plus rigides telles que la méthode des k-moyennes, GMM offre une plus grande flexibilité en supposant que les données suivent une distribution probabiliste et peuvent appartenir à un nombre prédéfini (k) de clusters avec différentes probabilités (Bishop, 2006).

Chaque cluster C_i est représenté par une distribution gaussienne paramétrée par une moyenne $\boldsymbol{\mu}_i$ et une covariance $\boldsymbol{\sigma}_i$.

Soit un ensemble de M observations $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$. La densité de probabilité de la distribution gaussienne d'un cluster C_i évaluée en $\mathbf{x}^t$ est définie par la fonction gaussienne suivante:

$$\mathcal{N}(\mathbf{x}^{\boldsymbol{t}}|\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\sigma}_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}^{\boldsymbol{t}} - \boldsymbol{\mu}_i)^T \boldsymbol{\sigma}_k{}^{-1}(\mathbf{x}^{\boldsymbol{t}} - \boldsymbol{\mu}_i)\right) \tag{24}$$

où $\boldsymbol{\mu}_i$ est la moyenne du cluster C_i,

$\boldsymbol{\sigma}_i$ est la matrice de covariance du cluster C_i,

d est la dimension des données.

La probabilité d'un point de données est une combinaison linéaire des distributions gaussiennes, pondérée par des coefficients qui représentent les poids des clusters (π_i):

$$P(\mathbf{x}^{\boldsymbol{t}}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}^{\boldsymbol{t}} | \boldsymbol{\mu_i}, \boldsymbol{\sigma_i}) \tag{25}$$

où k est le nombre de clusters,

π_i est la probabilité qu'une donnée appartienne au cluster C_i, avec la contrainte $\sum_{i=1}^{k} \pi_i = 1$.

La fonction de vraisemblance pour l'ensemble des M observations $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$ est donnée par :

$$L(\theta) = \prod_{t=1}^{M} \sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}^{\boldsymbol{t}} | \boldsymbol{\mu_i}, \boldsymbol{\sigma_i}) \tag{26}$$

où $\theta = \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\sigma_i}\}$ est l'ensemble des paramètres à estimer (les poids, les moyennes et les matrices de covariance pour les clusters C_i).

En pratique, il est plus facile de maximiser la *log*-vraisemblance que la fonction de vraisemblance directement (transformation du produit en somme):

$$\log L(\theta) = \sum_{t=1}^{M} log\left(\sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}^{\boldsymbol{t}} | \boldsymbol{\mu_i}, \boldsymbol{\sigma_i})\right) \tag{27}$$

L'algorithme d'*Expectation-Maximisation* (Dempster, Laird, & Rubin, 1977) est utilisé pour ajuster les paramètres des distributions gaussiennes (moyennes, covariances, et poids des clusters) en maximisant la probabilité des données observées. L'algorithme procède en deux étapes :

1- Le calcul de la probabilité d'appartenance de chaque point de données à chaque cluster en utilisant les paramètres actuels du modèle (*Expectation*):

$$\gamma(z_{\mathbf{x}k}) = \frac{\pi_k P(\mathbf{x} | \boldsymbol{\mu}_\mathbf{k}, \boldsymbol{\sigma}_\mathbf{k})}{\sum_{j=k}^{C} \pi_j P(\mathbf{x} | \boldsymbol{\mu}_\mathbf{j}, \boldsymbol{\sigma}_\mathbf{j})} \tag{28}$$

où $\gamma(z_{\mathbf{x}k})$ est la probabilité que le point $\mathbf{x}$ appartienne au cluster k.

2- La mise à jour les paramètres des distributions gaussiennes en maximisant la fonction de vraisemblance (Maximisation):

Mise à jour des moyennes des clusters k: $$\boldsymbol{\mu}_\mathbf{k} = \frac{\sum_{\mathbf{x} \in k} \mathbf{x} \gamma(z_{\mathbf{x}k})}{\sum_{\mathbf{x} \in k} \gamma(z_{\mathbf{x}k})} \tag{29}$$

Mise à jour des matrices de covariance: $$\boldsymbol{\sigma}_\mathbf{k} = \frac{\sum_{\mathbf{x} \in k} \gamma(z_{\mathbf{x}k})(\mathbf{x} - \boldsymbol{\mu}_\mathbf{k})(\mathbf{x} - \boldsymbol{\mu}_\mathbf{k})^T}{\sum_{\mathbf{x} \in k} \gamma(z_{\mathbf{x}k})} \tag{30}$$

Mise à jour des poids des clusters: $$\pi_k = \frac{1}{|k|} \sum_{\mathbf{x} \in k} \gamma(z_{\mathbf{x}k}) \tag{31}$$

Ces étapes sont répétées jusqu'à ce que la fonction de vraisemblance converge.

Contrairement à la méthode des *k-moyennes*, les GMM reposent sur un cadre probabiliste, ce qui permet à chaque point une probabilité d'appartenance à chaque cluster, plutôt qu'une assignation stricte. Cela permet de modéliser des clusters qui se chevauchent.

En revanche, GMM nécessite l'estimation de beaucoup plus de paramètres, notamment les matrices de covariance, ce qui peut rendre l'entraînement du modèle plus complexe et plus long, surtout avec des données de haute dimension.

Comme pour les *k-moyennes*, l'algorithme GMM peut converger vers des solutions locales optimales en fonction de l'initialisation des paramètres, et les points aberrants peuvent affecter significativement les paramètres des distributions gaussiennes, notamment les matrices de covariance, rendant la méthode moins robuste.

2.2.4.3 Mesure de la performance des modèles de partitionnement

Le coefficient de silhouette est une mesure de la qualité du partitionnement des données (Shahapure & Nicholas, 2020). Pour chaque point de données $\mathbf{x}^i$, on a :

$$s(\mathbf{x}^i) = \frac{b(\mathbf{x}^i) - a(\mathbf{x}^i)}{\max\left(a(\mathbf{x}^i), b(\mathbf{x}^i)\right)} \tag{32}$$

Où : $a(\mathbf{x}^i)$ est la distance moyenne entre $\mathbf{x}^i$ et tous les autres points dans le même cluster.

$b(\mathbf{x}^i)$ est la distance moyenne entre $\mathbf{x}^i$ et tous les points du cluster le plus proche.

Le coefficient de silhouette varie entre -1 et 1. Une valeur proche de *1* indique le point est dans la bonne partition, une valeur proche de 0 suggère qu'il est à la frontière entre deux partitions et une valeur proche de -1 suggère qu'il est dans la mauvaise partition.

Pour évaluer la qualité globale du clustering, on calcule le coefficient de silhouette moyen :

$$s = \frac{1}{|M|} \sum_{i=1}^{M} s(\mathbf{x}^i) \tag{33}$$

2.2.4.4 Partitionnement spectral (spectral clustering)

Le partitionnement spectral (Shi & Malik, 2000) est une méthode de clustering utilisée en apprentissage automatique et en théorie des graphes pour diviser les données en groupes (ou clusters) en fonction de leur structure sous-jacente, représentée par un graphe. Cette méthode est particulièrement utile pour des données complexes où les points de données ne sont pas

naturellement séparables en utilisant des techniques de clustering traditionnelles telles que les k-moyennes.

Représentation par un graphe

Soit $G = (V, E)$ un graphe non orienté constitué d'un ensemble de sommets V et d'un ensemble d'arêtes E. Chaque sommet représente un point de données ($V = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$), et les arêtes représentent les relations ou similitudes entre les sommets (basée par exemple sur la distance euclidienne). L'arête entre deux nœuds $\mathbf{x}^i$ et $\mathbf{x}^j$est donc pondérée par une fonction de similarité w_{ij} (par exemple la distance euclidienne). On peut ainsi définir la matrice de similarité S telle que :

$$S_{ij} = \begin{cases} w_{ij} & s'ily\ a\ une\ arête\ entre\ \mathbf{x}^i\ et\ \mathbf{x}^j \\ 0, & sinon \end{cases} \tag{34}$$

Si les poids non nuls sont tous *1.0*, cette matrice est la matrice d'adjacence du graphe G.

Objectif du partitionnement spectral

Le but du partitionnement spectral est de diviser les sommets V en deux sous-ensembles V_1 et V_2 tout en minimisant le poids total des arêtes qui relient les sommets de V_1 à ceux de V_2, et en maximisant la cohésion au sein de chaque partition (poids des arêtes intra-cluster).

Le coût de la coupe entre deux partitions V_1 et V_2 est défini comme :

$$Cut(V_1\ et\ V_2) = \sum_{i \in V_1, j \in V_2} S_{ij} \tag{35}$$

Matrice Laplacienne et décomposition spectrale

Une des clés du partitionnement spectral est l'utilisation de la matrice Laplacienne L du graphe, qui capture les relations entre les nœuds et la structure globale du graphe. La matrice Laplacienne est définie comme :

$L = D - S$ (Laplacien simple)

où D est la matrice diagonale des degrés des nœuds, définie par :

$D_{ii} = \sum_j a_{ij}$ (où $a_{ij} = 1$ s'il existe une arête entre $\mathbf{x}^i$ et $\mathbf{x}^j$, *0* sinon).

La matrice Laplacienne a des propriétés intéressantes qui facilitent la résolution du problème de partitionnement de graphe, notamment en exploitant ses valeurs propres et vecteurs propres qui sont utiles pour découper le graphe de manière optimale, car ceux-ci révèlent des propriétés du

graphe son apparent. Les vecteurs et les valeurs propres de la matrice Laplacienne sont les solutions de l'équation caractéristique :

$$\det(A - \lambda I) = 0 \text{ et } Lv = \lambda v \tag{36}$$

Chaque vecteur propre a une taille de M, correspondant aux nombres de points de données. Par exemple pour un bi-partitionnement du graphe ($k = 2$), le vecteur propre $\boldsymbol{v_1}$ correspondant à la plus petite valeur propre non nulle λ_1 ($\boldsymbol{v_1} =$ vecteur de Fiedler) peut être pour diviser aisément les sommets en deux groupes formant des cluster cohérents : les points de données associés aux valeurs positives de $\boldsymbol{v_1}$ forment une partition et les points de données associées aux valeurs négatives forment une deuxième partition.

Certaines implémentations considèrent le Laplacien normalisé :

$$L_N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \tag{37}$$

ou le Laplacien de Ng, Jordan & Weiss (Ng, Jordan, & Weiss, 2001) :

$$L_{NJW} = D^{-\frac{1}{2}} S D^{-\frac{1}{2}} \text{ où } S_{i,i} = 0 \tag{38}$$

Le choix du type de Laplacien va dépendre de plusieurs éléments comme la taille des données, le type de fonction de similarité utilisé.

Partitionnement du graphe en k clusters

Pour $k > 2$, les k premiers vecteurs propres $\boldsymbol{v_1}$ à $\boldsymbol{v_k}$ correspondant aux k premières valeurs propres par ordre croissant λ_1 à λ_k non nulles de L (en ignorant $\lambda_0 = 0$) peuvent être utilisés comme valeurs d'attributs, et un simple algorithme de clustering comme k-moyennes peut être appliqué en utilisant les valeurs des k vecteurs propres comme attributs des données. La méthode garantit qu'une méthode de partitionnement simple suffit pour un bon partitionnement des données.

Le partitionnement spectral offre une méthode puissante pour le clustering en exploitant les propriétés spectrales d'un graphe. Il est particulièrement utile pour les données complexes et n'a pas d'*a priori* sur la forme des clusters. Toutefois, il peut s'avérer complexe à implémenter, coûteux en temps de calculs pour de larges matrices avec une grande quantité de points de données et la qualité du clustering va beaucoup dépendre de la fonction de similarité utilisée.

2.2.5 Analyse en composantes principales (ACP)

L'analyse en composantes principales (Jolliffe, 2002) est une technique de réduction de dimensionnalité qui permet de transformer un ensemble de variables $X = \{X_1, X_2, \dots, X_p\}$ corrélées en un nouvel ensemble de p variables $Z = \{Z_1, Z_2, \dots, Z_p\}$, de variables non corrélées appelées composantes principales. La première composante principale Z_1 capture la plus grande variance possible des données, la seconde Z_2 capture la plus grande partie de la variance restante, et ainsi de suite.

L'ACP est largement utilisée en apprentissage automatique, en reconnaissance de formes et en traitement de signaux pour explorer les relations entre les variables et réduire la dimensionnalité des problèmes.

La méthode commence par le centrage des variables X en soustrayant la moyenne pour chaque variable: $X' = X - \mu$.

La matrice de covariance S est ensuite calculée pour comprendre comment les différentes variables sont corrélées entre elles : $S = \frac{1}{n-1} X'^T X'$ (où X'^T est la transposée de la matrice centrée X').

Les valeurs propres $(\lambda_1, \lambda_2, \dots, \lambda_p)$ et les vecteurs propres $(\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_p})$ de la matrice de covariance S sont ensuite calculés en trouvant les solutions de l'équation caractéristique :

$\det(S - \lambda I) = 0$ et $S\boldsymbol{v_i} = \lambda_i \boldsymbol{v_i}$

Les valeurs propres λ_i représentent la variance expliquée par chaque composante principale et les vecteurs propres $\boldsymbol{v_i}$ fournissent la direction de ces composantes principales (orthogonaux car la matrice S est symétrique).

Les nouvelles variables $Z_1, Z_2, \dots, Z_p$, ou composantes principales, sont des combinaisons linéaires des variables d'origine, telles que $Z_i = X'\boldsymbol{v_i}$

Si les k premières composantes capturent une grande partie de la variance (exprimée en pourcentage), les données peuvent être projetées dans un espace de dimension réduite k, offrant ainsi une simplification tout en préservant l'information essentielle.

La proportion de la variance totale expliquée par les k premières composantes principales est donnée par :

$$\%Variance\ expliquée = 100 \times \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \tag{39}$$

L'ACP est relativement simple à mettre en œuvre et bien adaptée à de nombreux types de données.

Elle souffre cependant d’une perte d’interprétabilité : une fois les données projetées dans l'espace des composantes principales, les nouvelles variables (composantes) ne correspondent plus directement aux variables d'origine, ce qui rend difficile l'interprétation des résultats dans des termes simples ou pratiques.

L'ACP repose sur l'hypothèse que les relations entre les variables sont linéaires. Cela peut limiter son efficacité dans les jeux de données avec des relations non linéaires. Par ailleurs, si les variables ne sont pas correctement normalisées ou standardisées, les résultats peuvent être faussés, car certaines variables avec des valeurs plus élevées domineront les composantes principales.

CHAPITRE 3 REVUE DE LITTÉRATURE

3.1 État de l'art sur les méthodes d'identification de la lithologie basée sur les MWD

Les différents types de roches souterraines ont des propriétés physiques et chimiques différentes, de sorte que les paramètres de fonctionnement de la foreuse changeront avec toute altération significative de la lithologie. Par exemple, le passage du *siltstone* au charbon peut voir une augmentation du taux de pénétration, car ce dernier est plus facile à forer. Des recherches précoces telles que celles de Scoble & Peck (1987) et Scoble et al. (1989) ont démontré qu'il existe une corrélation entre les MWD et le type de roche souterraine.

3.1.1 Indicateurs de classification géologique basés sur les MWD

Il existe une relation évidente ou intuitive entre le type de roche (géologie) rencontrées durant le forage et les signaux MWD. Segui et Higgin (2002) ont montré que le couple (Torque) est généralement proportionnel à la dureté de la roche, tandis que le taux de pénétration (ROP) est inversement proportionnel à la dureté de la roche. Hatherly et Leung (2015) ont effectué une évaluation détaillée du lien entre les données MWD des trous de forage et les relevés géophysiques dans les mines de charbon. Cette étude confirme qu'il existe une bonne corrélation entre les données MWD et le type de roche ainsi que d'autres propriétés de la roche telles que la vitesse sonique et la résistance de la roche. Le défi est de savoir comment utiliser ces corrélations apparentes pour une classification fiable en temps réel de la roche forée. Leung & Scheding (2015) ont également observé que lorsque les forets pénètrent dans une couche de charbon, le taux de pénétration augmente tandis que le couple diminue simultanément.

Depuis les premières recherches sur ce sujet MWD, plusieurs auteurs ont conçu et proposé des indicateurs visant à discriminer les différents types de roches. Ces indicateurs sont calculés à partir des MWD primaires : RPM la vitesse de rotation exprimée en rad/s; (WOB/PD) la force (N) exercée sur le tricône; (ROP) le taux de pénétration exprimé en m/s; et (TRQ) le couple (N-m). Dans le cadre de cette recherche, ces indicateurs peuvent être considérés comme des MWD composites provenant de capteurs « virtuels » (ou « *soft sensors* »). Parmi ces auteurs, Teale (1965)

a proposé l'énergie spécifique de forage (SED) comme indicateur. L'énergie spécifique est une mesure de l'énergie nécessaire pour fracturer une roche donnée et est définie comme suit :

$$SED = \frac{TRQ \times RPM + WOB \times ROP}{A \times ROP} \tag{40}$$

où A est le diamètre du trou (m²).

SED est en fait exprimé sous forme de pression (en kPa ou MPa). Bien qu'il existe une forte corrélation entre SED et le type de roche rencontrée (par exemple dans une mine de charbon), cet indicateur seul n'est pas toujours suffisant pour détecter de manière fiable le type de roche rencontrée, car il nécessite parfois le contexte et sa valeur peut être affectée le type de fractures présentes dans la roche. Dans cette recherche, le SED est utilisé conjointement avec d'autres données MWD primaires comme variables d'entrée pour plusieurs des modèles de prédictions proposés.

Zhou et al. (2012) ont proposé un nouvel indicateur appelé le taux de pénétration ajusté (APR) calculé à partir des données MWD :

$$APR = \frac{ROP}{PP\sqrt{RP}} \tag{41}$$

où PP (*Pulldown Pressure*) est la pression de poussée et RP (*Rotation Pressure*) est la pression de rotation. L'APR est conçu pour être une forme de normalisation du taux de pénétration et présente une forte corrélation avec certaines propriétés de la roche telles que la dureté relative de la roche. Cet indicateur ne permet pas à lui seul de faire une classification du type de roche, car des références sont nécessaires.

Leung & Scheding (2015) ont développé un nouvel indice appelé énergie spécifique modulée (SEM) qui combine le SED et le rapport entre la puissance de rotation (RPM) et la poussée (WOB). Le SEM s'est avéré être un meilleur indicateur de présence du charbon (classification binaire) que les indices précédents comme l'APR. L'article met de l'avant la précision (voir Section 2.2.2.7) de l'indicateur SEM qui est de 72,6 %. Cependant l'équilibre entre la précision et le rappel n'est pas clairement établi (en utilisant des métriques telles que le F-Score), ce qui est d'autant plus important car il s'agit d'un problème de classification débalancé (où une classe est largement majoritaire).

Zhang et Hou (2016) proposent une méthode d'évaluation d'un indice de *drillability* (« forabilité »), *Id*, est utilisé comme indicateur du type de roche et calculé à partir des paramètres de

fonctionnement d'une foreuse de type *drifter*. Cet indice est calculé en fonction de paramètres MWD tels que la vitesse de pénétration (*ROP*), le diamètre du trou de forage (*D*), la force de poussée (*PD*), le couple du tricône (*TRQ*) et la vitesse de rotation du tricône (*RPM*) :

$$I_d = \gamma \left(\frac{D \times PD}{TRQ}\right)^{\alpha} \left(\frac{ROP}{DRPM}\right)^{\beta} \tag{42}$$

Où γ, α et β sont les paramètres à déterminer. I_d est ensuite utilisé pour classifier le type de roche en utilisant plusieurs seuils par exemple $Id \geq 2{,}4$ pour le charbon.

Niu & Zhang (2019) utilisent le même indice de « forabilité » pour déterminer la classification du type de roche dans un cadre expérimental. Cependant, dans les deux études, les performances réelles de l'indice *Id* en tant que discriminant fiable du type de roche ne sont pas quantifiées et comparées avec d'autres indices.

Utiliser des indicateurs composites comme outils de classification de la lithologie requiert une étape supplémentaire qui est la construction de tables de références qui définissent les seuils correspondants à chaque catégorie de roche d'intérêt. Cela rajoute de l'incertitude et parfois des délais à la reconnaissance de la géologique. Pendant longtemps, l'indisponibilité d'outils adéquats pour traiter les données a limité l'utilisation des MWD pour la caractérisation souterraine et la reconnaissance en temps réel des types de roches. De nos jours, plusieurs études visent à démontrer que la classification des types de roches peut être réalisée en temps réel en utilisant des algorithmes d'apprentissage automatique avec les données MWD comme variables d'entrée.

3.1.2 Méthodes d'identification de la géologie basées sur l'apprentissage automatique

En lieu et place des indicateurs, qui nécessitent de définir des seuils et des tables de références pour classifier la roche forée, la plupart des travaux récents se concentrent sur les nouveaux outils d'apprentissage automatique qui peuvent exploiter la vaste quantité de données MWD disponibles.

LaBelle *et al.* (2000) et LaBelle (2001) ont utilisé un réseau de neurones pour classifier la lithologie en utilisant les données MWD comme variables d'entrée. Cinq catégories d'argile ont été classifiées avec succès. Les résultats ont suggéré que les MWD peuvent être utilisés pour classifier le type de roche, mais ils n'ont pas réussi à généraliser pour classifier avec précision le charbon et le schiste

lors des expériences préliminaires sur le terrain. Le manque de données appropriées était le principal obstacle dans cette recherche.

Dans une recherche similaire, Kryuchkova *et al.* (2018) proposent plusieurs modèles d'apprentissage supervisé pour prédire la lithologie au niveau du tricône en utilisant comme variables les données MWD et LWD (*Logging While Drilling*) et des caractéristiques d'ingénierie. L'objectif de cette étude était d'optimiser le forage directionnel dans les puits de pétrole. Les modèles utilisés pour une classification binaire (sable ou schiste et roche) sont la régression logistique, le *gradient boosting* et un réseau de neurones (*feedforward* et LSTM). Les métriques de qualité utilisées étaient l'exactitude, la courbe ROC (*Receiver Operating Characteristic*) et PRAUC (*Precision Recall Area Under the Curve*), cette dernière étant la plus pertinente en raison de la présence des classes débalancées. Le *gradient boosting* et le réseau de neurones *feedforward* se sont avérés être les meilleurs modèles dans ce contexte avec un PRAUC maximum de 0,625. Les auteurs ont souligné qu'un moyen d'améliorer les performances du modèle est d'utiliser des données provenant de capteurs supplémentaires.

Kumar et Vardan (2013) ont utilisé un modèle de régression multiple et des modèles de réseaux de neurones artificiels (perceptron multicouche avec fonctions de base radiale) pour prédire diverses propriétés des roches telles que la densité en utilisant comme variables d'entrée la vitesse et le diamètre du tricône, le taux de pénétration et le niveau sonore du forage. L'erreur moyenne absolue en pourcentage (MAPE) est utilisée comme indicateur de performance. Les erreurs de pourcentage absolues moyennes pour les prédictions de densité de la roche varient de 9,96 % pour le modèle de perceptron multicouche à 15,77 % pour le modèle de régression multiple.

Dans une étude récente, Zhong *et al.* (2020) ont également utilisé plusieurs modèles d'apprentissage supervisé (régression logistique, SVM, réseau de neurones artificiels, *random forest* et *gradient boosting*) pour inférer la localisation des couches de méthane houiller. Des techniques de suréchantillonnage aléatoire (suréchantillonnage aléatoire naïf et technique de suréchantillonnage synthétique des minorités) sont utilisées pour atténuer les ratios déséquilibrés des occurrences de couches de méthane houiller ciblées dans les données d'entraînement. Bien que les résultats semblent prometteurs, cette étude n'a pas tiré parti de l'opportunité de combiner les différentes méthodes d'apprentissage automatique pour améliorer les résultats.

Dans la plupart des recherches précédentes, un ou quelques modèles d'apprentissage sont sélectionnés, appliqués et comparés. Différents modèles ont différentes façons de séparer les données, et plus important encore, différentes façons de surapprendre pendant l'entraînement. Il pourrait y avoir des avantages à les combiner judicieusement. L'approche proposée au chapitre 5 compare plusieurs modèles et les combine ensuite pour obtenir une meilleure décision.

Zhou *et al.* (2012) dans leurs travaux définissant l'APR (voir Équation 2.2) utilisent l'apprentissage non supervisé (clustering gaussien) avec l'APR comme variable d'entrée pour effectuer la classification de la roche. Ce processus a été appliqué sur un cas test d'une mine de fer et a réussi à classifier la dureté de la roche, mais son efficacité pour une approche quasi-temps réel reste à prouver, car elle nécessite beaucoup de paramétrisation et des références géologiques.

3.2 État de l'art sur la détection de l'usure des tricônes basée sur les MWD

Abbas (2018) fournit une revue approfondie des mécanismes d'usure des trépans de forage pétrolier, notamment les forets à cône roulant (tricônes) et les forets en compact de diamant polycristallin. L'étude discute des modèles empiriques d'usure et des techniques de prédiction de l'usure pendant le forage, y compris l'analyse par éléments finis et la méthode des éléments discrets. Ces modèles et techniques permettent une meilleure compréhension du comportement des trépans et de leur conception, mais comportent des hypothèses qui limitent son applicabilité. L'étude conclut que bien que des progrès significatifs aient été réalisés dans la compréhension et la quantification de l'usure des trépans, il est essentiel d'intégrer les techniques avancées de modélisation avec les méthodes conventionnelles pour améliorer la fiabilité et la précision des prédictions de l'usure.

Les méthodes d'analyse des signaux de forage se concentrent sur l'analyse des données générées par l'équipement de forage, telles que le couple et les signaux de vibration du trépan, pour inférer l'état d'usure du trépan. Liu *et al.* (2022) ont extrait des caractéristiques temporelles et entropiques des signaux de vibration à proximité du trépan pour identifier le degré d'usure du trépan. Rafezi & Hassani (2021) ont étudié la relation entre les vibrations de forage, les signaux de courant du moteur et l'usure du trépan, en introduisant des fréquences de vibration des signaux sous différentes conditions d'usure. Rafezi et Hassani (2023) proposent une nouvelle approche utilisant la décomposition en paquets d'ondelettes des MWD (vibration) et des réseaux de neurones artificiels

pour classer les états d'usure des tricônes en cinq niveaux, avec une précision d'environ 71,6 %. Bien que prometteur ce système de classification ne s’appuie principalement que sur la vibration.

Le suivi de l'usure des trépans basé sur l'analyse des signaux propose des modèles de grande précision ; cependant, ces méthodes présentent également plusieurs inconvénients. Par exemple, comme les trépans fonctionnent dans des conditions de haute température et de haute pression souterraines, il y a une multitude de types et de quantités de bruit dans les données de vibration près du trépan. Les méthodes d'analyse des signaux nécessitent des algorithmes de filtrage complexes et des capteurs spécialisés. Par conséquent, interpréter avec précision les signaux de forage est difficile et coûteux, rendant cette méthode moins adaptée à une application généralisée.

Les modèles d’apprentissage supervisés nécessitent des étiquettes d'usure pour guider l’entraînement du modèle. Cependant, les ingénieurs n'annotent l'état d'usure du trépan qu'à la fin de son utilisation et l'évolution dynamique de l'usure pendant le processus de forage est inconnue. Par conséquent, il est impraticable de fournir des étiquettes d'usure correspondantes pour les données MWD enregistrées, ce qui limite grandement la quantité de données d'entraînement pour l’apprentissage supervisé. Luu *et al.* (2021) et Zhan *et al.* (2023) ont adopté des méthodes d'entraînement non supervisé pour surmonter ce défi. Ils utilisent le modèle BI-LSTM-VAE (*Bidirectional Long Short-Term Memory Variational Autoencoder*), un modèle hybride qui combine les réseaux de mémoire à long terme bidirectionnels et les auto-encodeurs variationnels, pour extraire des caractéristiques bidimensionnelles représentatives des grades d'usure des trépans, permettant ainsi une surveillance de l'usure pendant le processus de forage.

Agostini et Sampaio (2020) ont employé des réseaux de neurones convolutifs et un réseau de neurones probabiliste (PNN) pour extraire les caractéristiques des signaux de couple (Torque) et établir un modèle de surveillance de l'usure des trépans, pour un classement binaire de l’état des trépans (bon ou mauvais).

En résumé, pour la surveillance de l'usure des trépans, les modèles empiriques présentent une faible précision, tandis que l'analyse des signaux et les modèles intelligents nécessitent beaucoup de données de haute qualité et peinent à s'adapter aux changements des conditions de forage et des propriétés géologiques.

3.2.1 Synthèse de la revue

L'analyse des travaux de recherche existants met en évidence plusieurs limites dans les modèles actuels utilisés pour l'identification de la géologie et la détection de l'usure des trépans à l'aide des données MWD. Les approches traditionnelles, bien que prometteuses, souffrent de certaines faiblesses, notamment une grande dépendance aux indicateurs empiriques et un manque d'intégration efficace du contexte (par exemple, l'information issue des trous de forage voisins qui peut servir pour le trou courant). Par ailleurs, les modèles d'apprentissage automatique existants appliqués à ces problématiques ont souvent été développés de manière isolée, ne permettant pas une exploitation des différentes approches.

L'un des défis majeurs identifiés est la nécessité de combiner efficacement des approches supervisées et non supervisées pour améliorer la robustesse des prédictions. La prise en compte des relations spatiales entre les trous de forage ainsi que l'intégration d'informations temporelles restent des aspects peu explorés qui pourraient améliorer la fiabilité des modèles. Ce travail de recherche apporte une contribution importante en proposant de nouvelles méthodes basées sur l'apprentissage automatique pour surmonter ces limites, en intégrant des modèles hybrides combinant classification supervisée et apprentissage non supervisé, pour fusionner les données de plusieurs trous de forage. Un modèle récurrent est également exploré pour mieux prédire l'évolution de l'usure des trépans.

CHAPITRE 4 DÉMARCHE ET ORGANISATION DE LA THÈSE

Cette thèse est organisée comme suit. Le chapitre 1 a introduit le thème de cette recherche, définit les concepts clés, la problématique et les objectifs de recherche, alors que le chapitre 2 a traité des concepts mathématiques et des concepts de forage de productions utiles. Le chapitre 3 présente une revue de la littérature existante et des recherches précédentes dans le domaine. Les solutions proposées dans la suite de cette thèse sont présentées aux chapitres 5, 6 et 7 qui sont interreliés par une démarche de réflexion commune. Ces trois chapitres traitent de l'utilisation de différents algorithmes d'apprentissage automatique avec les données MWD afin d'identifier de manière adéquate et rapide la lithologie et l'état d'usure du tricône, dans le but d'optimiser l'efficacité et les coûts des opérations de forage et de dynamitage.

Chacun de ces chapitres aborde une problématique particulière et propose une solution. Chaque chapitre propose une introduction décrivant le contexte, définissant la problématique spécifique et les objectifs du chapitre. L'approche proposée comme solution est ensuite détaillée de manière théorique et appliquée en utilisant des données réelles. Les résultats sont ensuite discutés et des recommandations sont formulées en conclusion, en proposant des pistes de recherches ultérieures.

Le chapitre 5 explore l'utilisation de l'apprentissage automatique pour la reconnaissance (classification) en temps réel de la lithologie pendant le forage, en utilisant les données MWD. Le contenu de ce chapitre a fait l'objet d'un article (Zagré, Gamache, Labib, & Shlenchak, 2024) publié dans le journal *International Journal of Mining*. Ce chapitre propose donc une revue de l'approche proposée dans cet article qui est basée sur la combinaison de plusieurs algorithmes d'apprentissage supervisé (régression logistique, arbres de décision, *gradient boosting*, etc.). Le problème est abordé sous l'angle d'une classification binaire (minerai ou non-minerai) et est appliqué sur les données d'un cas réel provenant d'une mine de charbon (avec le charbon comme classe d'intérêt (minerai)). Plusieurs modèles basés sur l'apprentissage supervisé sont proposés. Une des conclusions de ce chapitre est que qu'il serait possible d'utiliser l'information (MWD) provenant des trous de forages voisins (prédiction globale) pour améliorer la fiabilité de la prédiction locale courante particulièrement dans certaines situations.

Le chapitre 6 traite des méthodes d'apprentissage non supervisé pour une classification plus globale de la lithologie (en opposition à la classification plus locale effectuée dans un seul trou de forage à

intervalles réguliers de profondeur). Ce chapitre présente la formulation du problème de partition des données (ou clustering) en domaines géologiques et compare plusieurs modèles non supervisés pour un partitionnement binaire du domaine formé par un trou de forage et les trous voisins. Les méthodes proposées sont comparées à l'aide d'un exemple provenant de la même mine de charbon.

Le chapitre 7 propose une méthode basée sur l'apprentissage non supervisé pour détecter le seuil d'usure d'un tricône à partir des données MWD. Cette méthode propose différentes façons d'isoler et d'évaluer un indicateur de l'efficacité du tricône. Ce chapitre se termine sur une discussion sur les méthodes de correction de la prédiction de la lithologie présentée au chapitre 4, prenant en compte la prédiction d'usure du tricône, ainsi qu'une méthode de normalisation de certaines mesures composites MWD, telles que les indicateurs de dureté de la roche (SED, BI, etc.).

CHAPITRE 5 APPRENTISSAGE SUPERVISÉ POUR LA RECONNAISSANCE DU CHARBON

5.1 Introduction

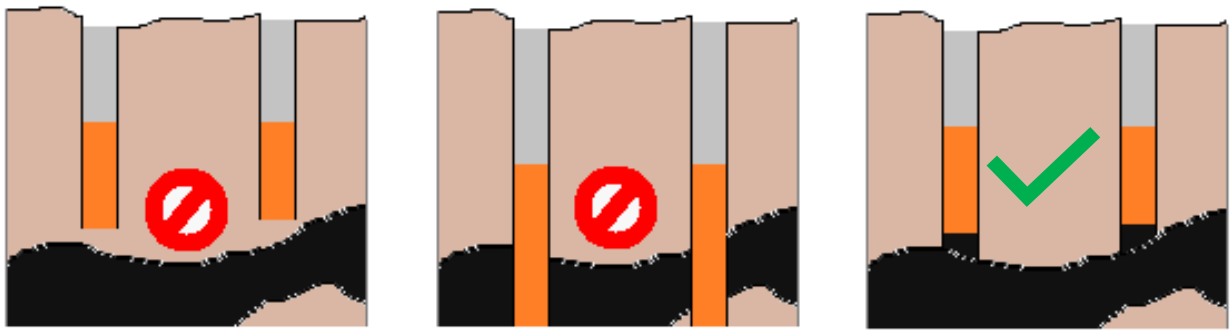
5.1.1 Contexte et enjeux

Il est aujourd'hui largement reconnu que l'automatisation peut améliorer l'efficacité opérationnelle dans l'industrie minière (Leung & Scheding, 2015). Les foreuses minières modernes sont équipées de divers capteurs permettant le suivi en temps réel des paramètres de forage. Ces mesures, identifiées par le terme MWD (voir section 2.1.3), incluent divers paramètres de performance tels que la vitesse de rotation (RPM), la vitesse de pénétration (ROP), le couple (TRQ), la pression exercée sur le trépan (WOB) ou la pression de poussée (PD). Avec la résurgence des algorithmes d'intelligence artificielle tels que les l'apprentissage automatique, ces grandes quantités de données historiques peuvent être exploitées pour construire des modèles riches et fiables qui seront utiles pour les processus d'exploitation minière, facilitant ainsi l'automatisation.

Un domaine d'intérêt urgent dans l'automatisation de l'extraction de charbon à ciel ouvert est la détection rapide, précise et fiable des couches de charbon. Le processus d'extraction du charbon nécessite l'élimination de la roche au-dessus de la couche de charbon par une succession de forages et de dynamitage. Cela consiste à forer des trous jusqu'à une certaine profondeur et les charger d'explosifs. Si les explosifs sont placés trop haut au-dessus de la couche de charbon, l'explosion est inefficace (Figure 5.1a). Si le trou est foré au-delà de la couche de charbon, la détonation pourrait endommager et contaminer la couche de charbon (Figure 5.1b). Selon Mason (2011), les pertes associés aux explosions incorrectes ou inefficaces dans dix mines de charbon en Australie équivaudraient au coût d'une mine (inutile).

La meilleure pratique actuelle consiste à forer au-delà de la couche de charbon, puis à remplir la portion du trou au niveau de la couche avec du charbon avant de charger les explosifs (Figure 5.1c). Cela entraîne des coûts et des délais supplémentaires. L'idéal serait d'arrêter le forage dès que la couche de charbon est détectée. Actuellement cette détection des couches de charbon est plus ou moins subjective donc sujette à des erreurs de jugement. Bajram et al. (2015) souligne que la

détection précise de la couche de charbon est cruciale, car cela pourrait permettre de réduire le taux de dilution du charbon de 66 % et abaisser le taux de perte de charbon de 13 % à 4 %.



a-Explosifs placés trop haut au-dessus de la couche de charbon

b-Explosifs placés trop haut au-dessus de la couche de charbon

c-Meilleure pratique courante

Figure 5.1 – Chargement d'explosifs dans l'extraction du charbon

Les relevés géophysiques post-forages utilisant des outils à source radioactive fournissent une estimation de l'emplacement de la couche de charbon mais seulement pour un petit sous-ensemble des trous de forage. Cette opération est coûteuse et chronophage, car les opérateurs doivent descendre manuellement des sondes dans des trous déjà forés. Par exemple, sur une mine de charbon Canadienne, il est estimé qu'environ seulement 1% des trous de forage sont l'objet de relevés géophysiques. Une détection automatique des couches de charbon pendant le forage présenterait donc une grande valeur.

5.1.2 Objectif spécifique

Les différents types de roches souterraines ayant des propriétés variées, les paramètres de fonctionnement de la foreuse changeront en cas de modification significative de la lithologie. Par exemple, passer du grès au charbon peut augmenter la vitesse de pénétration, le charbon étant plus facile à forer. Des recherches antérieures, telles que celles de Scoble & Peck (1987) et Scoble et al. (1989), ont démontré qu'il existe une corrélation entre les données MWD et le type de roche souterraine. Pendant longtemps, l'absence d'outils adéquats pour traiter ces données a limité l'utilisation du MWD pour la caractérisation et l'identification en temps réel des types de roches. Aujourd'hui, plusieurs études visent à démontrer que la classification des types de roches en temps réel peut être réalisée grâce aux algorithmes d'apprentissage automatique appliquées avec les MWD comme entrées.

L’objectif de ce chapitre est le développement et la comparaison de plusieurs algorithmes d’apprentissage automatique capables pour détecter automatiquement la présence ou l’approche de la couche de charbon pendant le forage. Les modèles sont construits à partir de données in situ MWD et les relevés géophysiques associées. Des algorithmes d’apprentissage supervisé tels que la régression logistique, les arbres de décision, l’extreme-gradient-boosting et les k-plus-proches voisins sont entraînés et utilisés pour classifier les points de données MWD de manière binaire : « Charbon » ou « Pas Charbon ». Les performances des différents modèles sont comparées en utilisant des métriques de qualité appropriées telles que la Précision, le Rappel et le F1-Score. Un modèle de réseau neuronal est également proposé pour la prédiction de la densité.

Ces outils permettront de gagner du temps et de réduire les coûts associés aux relevés géophysiques, d’obtenir une meilleure cartographie géologique des dépôts souterrains de charbon et de soutenir les efforts d’automatisation du forage minier.

5.2 Méthodologie et hypothèses

5.2.1 Approche générale

L'approche proposée dans cette recherche consiste à créer un ensemble d'outils pour la classification en temps réel de la roche souterraine, la prédiction de la densité de la roche et la détection de l'approche d'une veine de charbon.

Le premier outil est basé sur des algorithmes d'apprentissage supervisés pour une classification binaire. Ces modèles utilisent les données MWD comme caractéristiques. Les sorties de prédiction de différents modèles sont combinées par moyenne pondérée (combinaison linéaire avec des poids différents) et la performance de ce nouveau modèle intégré est évaluée.

Le deuxième outil est basé sur l'estimation de la densité par régression. Un réseau de neurones artificiels entièrement connecté et entraîné par rétropropagation pour prédire la densité. Le modèle utilise les données MWD comme caractéristiques d'entrée et plusieurs architectures sont testées (nombre de couches cachées, fonctions d'activation, etc.). La meilleure architecture de réseau de neurones trouvée pour la prédiction de la densité est ensuite utilisée pour corriger l’erreur d’alignement entre les MWD et les relevés géophysiques si nécessaire. Les modèles peuvent

ensuite être mis à jour en cas de correction des erreurs l'alignement. La Figure 5.2 schématise le processus de préparation des modèles.

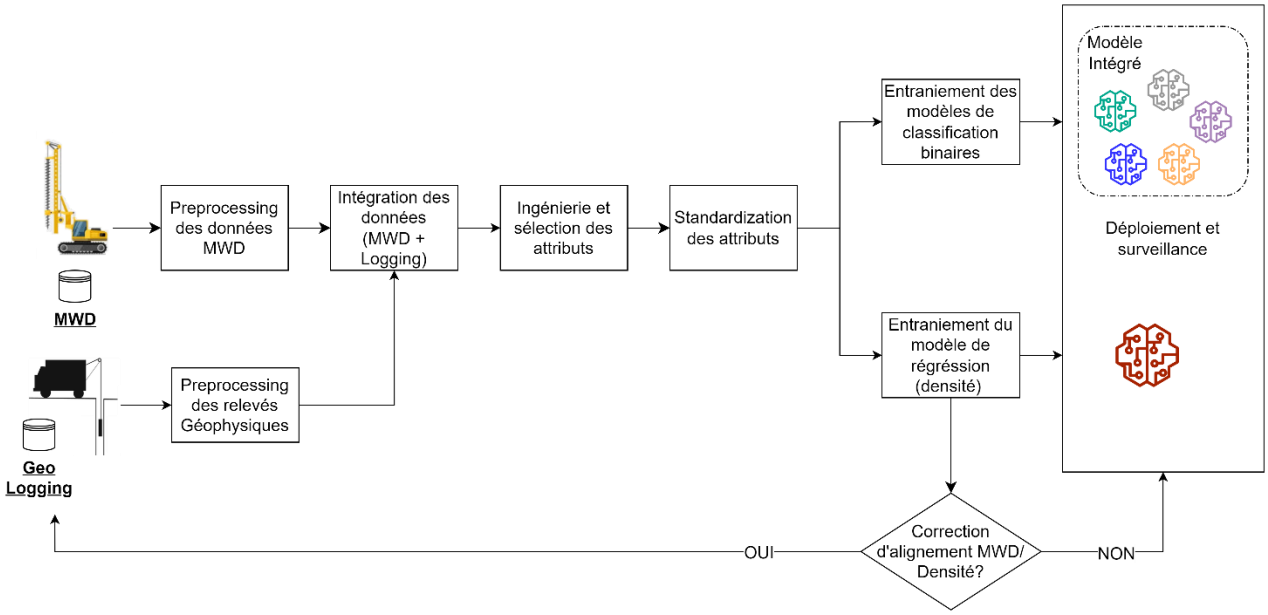


Figure 5.2 - Processus d'entrainement des modèles de prédiction

5.2.2 Étude de cas

L'étude de cas utilisée dans cette recherche repose sur des données in-situ provenant d'une mine de charbon à ciel ouvert Canadienne. Les données MWD ont été fournies par Peck Tech Consulting, qui a installé des unités d'acquisition de données (DAU) sur plusieurs foreuses. Les données analysées proviennent d'une unique foreuse et ont été enregistrées pour un ensemble de 41 trous de forage d'environ 20 m de profondeur, répartis sur deux bancs distincts. Les paramètres de fonctionnement enregistrés par la DAU de Peck Tech comprenaient les éléments suivants à une fréquence d'échantillonnage élevée de 200 Hz (toutes les 0,005 secondes):

- Profondeur
- Vitesse de rotation (RPM)
- Taux de pénétration (ROP)
- Couple (TRQ)
- Pression sur le trépan (WOB)
- Énergie spécifique de forage (SED)

Les données de diagraphie géophysique (sondes gamma/densité) associées incluaient entre autres les variables suivantes, mesurées tous les 0,02 m :

- Profondeur de référence
- Densité (g/cm³)

5.2.3 Hypothèses principales

Une classification fiable des types de roches ne peut être obtenue que si les données MWD utilisées représentent des mesures fiables, et la performance des algorithmes proposés dépend de la qualité des données MWD. Sachant qu'un certain niveau d'incertitude est toujours associé aux mesures de valeurs physiques, un prétraitement de base des données est appliqué pour lisser les données et éliminer les valeurs aberrantes. Il est reconnu qu'un décalage de profondeur entre les données MWD et les relevés géophysiques peut exister, et une méthode est proposée dans cette étude pour calculer ce décalage possible pour chaque trou de forage (voir section 5.5.2).

Les opérateurs de forage utilisent généralement des points de consigne pour certains paramètres de fonctionnement tels que le RPM et le WOB. Les modèles proposés fonctionnent bien uniquement dans des conditions où les principaux paramètres de réponse ne sont pas contraints manuellement par l'opérateur et où la foreuse est en mode de contrôle automatique (PDC activé). Dans ces conditions, les autres paramètres de forage sont des réponses directes aux conditions souterraines. Il est supposé que la plupart des données MWD disponibles ont été enregistrées lorsque la foreuse fonctionnait en mode automatique avec des points de consigne habituels. La Section 8 examine la capacité de généralisation de l'approche proposée dans différentes conditions et propose une méthode pour évaluer les cas où de nouvelles conditions nécessitent un nouveau modèle.

Le critère d'acceptation pour la détection des couches de charbon est défini comme l'identification de la profondeur de la couche de charbon avec une marge d'erreur de profondeur de ±0,20 m, conformément à Mason (2011). Cette hypothèse est très conservatrice car la marge d'erreur proposée est relativement faible. À titre de comparaison, une marge d'erreur de ±0,5 m est considérée acceptable pour un site minier de charbon Canadien.

L'état du trépan influence grandement l'efficacité du forage et plusieurs paramètres de forage. Cela pourrait affecter la généralisation des modèles proposés et leur performance dans des contextes différents (tricône neuf ou usé). L'état d'usure du tricône n'est pas connu dans les données utilisées

pour ce chapitre, et n’est donc pris en compte dans ce chapitre. Le Chapitre 6 propose une méthode de prédiction de l’état du tricône en fonction des MWD.

D'autres facteurs susceptibles de limiter la capacité des modèles proposés à généraliser incluent les actions et comportements des opérateurs, ainsi que la configuration de la foreuse. En l'absence de données suffisantes pour discriminer ces facteurs ou normaliser les données en conséquence, il est supposé que leur effet est négligeable pour cette recherche.

Il est également supposé que la dynamique du forage ne change pas en fonction de la profondeur. Comme expliqué par Leung & Scheding (2015), cette hypothèse est raisonnable pour une telle étude, car les foreuses modernes sont équipées d'un contrôleur PID qui régule leur dynamique et corrige toute variation linéaire potentielle des paramètres en fonction de la profondeur.

5.2.4 Identification des couches de charbon à partir des relevés géophysiques

La localisation de couches de charbon exploitables à partir des relevés géophysiques est un processus complexe. Les géologues utilisent les lectures de densité, de rayons gamma et de *caliper*, en plus de leurs connaissances préalables sur la géologie (données d’exploration et relevés de patrons voisins) pour inférer l'emplacement des couches de charbon. En combinant tous ces éléments avec leur expertise, les géologues peuvent déterminer si une couche de charbon exploitable se trouve dans les futurs trous de forage d’un patron et la location approximative de cette couche de charbon. Lorsqu'une faible densité est enregistrée, une valeur élevée du *caliper* pourrait indiquer qu'il ne s'agit pas réellement de roche tendre mais plutôt d'une augmentation du diamètre du forage. Le *caliper* est également utilisé pour ajuster ou corriger les mesures de rayons *gamma*.

La présence de géologies complexes complique davantage le problème (couches de charbon fin, zones de transitions remplies de *mudstone* de faible densité plutôt que de grès, et transitions graduelles entre le charbon et des sédiments rocheux différents).

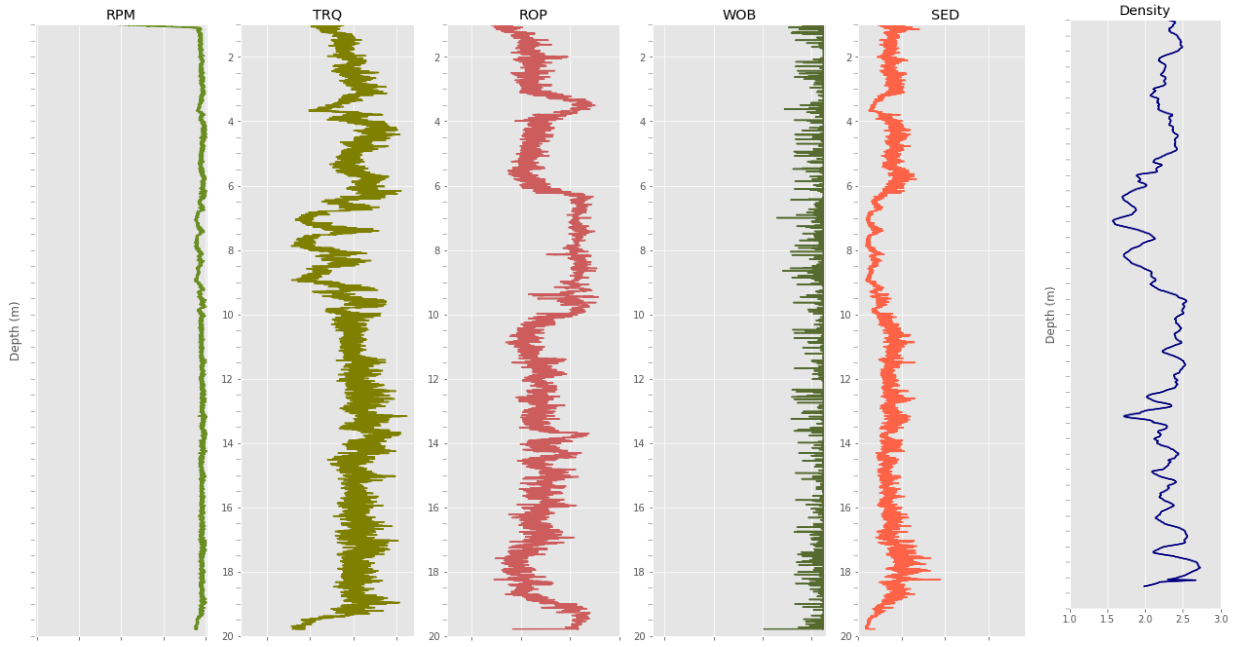
Dans le cadre de cette étude, un simple seuil de densité est utilisé pour déduire la présence de charbon à partir des relevés géologiques. Chaque point de données MWD qui est associé à une densité inférieure à 1,65 g/cm³ (valeur de référence utilisée sur un site spécifique) est considéré comme du charbon. Ce seuil est une recommandation des géologues dans le site minier à l’étude. Les autres points de données sont considérés comme des roches clastiques (grès, *mudstones* et

siltstones). Le seuil de 1,65 g/cm³ est conservateur lorsque comparé au seuil utilisé dans d'autres travaux similaires. Par exemple, dans l'étude de Hatherly et al. (2015), un seuil de densité plus élevé de 1,85 g/cm³ est utilisé pour distinguer le charbon des autres roches dans les relevés géophysiques. Les valeurs de densité comprises entre 1,85 et 2,2 g/cm³ étaient classées comme du matériel carboné, et les densités plus élevées comme des roches clastiques (grès et *siltstones*). Avec un seuil de 1,65 g/cm³, le nombre d'entrées de charbon est réduit, mais la confiance dans l'exactitude de cette classification est accrue. Cette étude se concentre sur les couches de charbon larges et exploitables (épaisseur de plus de 80 cm), les pics aléatoires de charbon dans les données de références sont lissés ou simplement ignorés.

5.3 Analyse des données

5.3.1 Preprocessing

Les données MWD ont été échantillonnées à haute fréquence (200 Hz), tandis que la densité a été enregistrée tous les 0,02 mètres. La Figure 5.3 présente un exemple de tracé des signaux MWD d'origine et des relevés de densité fournis pour un trou de forage.



a – Signaux MWD b - Densité

Figure 5.3 – Données MWD brutes pour un trou de forage et relevé de densité associée

Les premiers 2,0 m de chaque trou de forage sont considérés comme un sol potentiellement fragmenté et les données correspondantes sont supprimées. Les données correspondant à la montée de la mèche (ROP négatif) sont également retirées.

Les relevés de densité servent à créer la classification binaire de type de roche de référence pour chaque point de données en utilisant le seuil de 1,65 g/cm³:

- Type de roche = **1** pour la classe *Charbon* (densité ≤ 1,65 g/cm³)
- Type de roche = 0 pour la classe *Pas-Charbon* (roche dure avec densité > 1,65 g/cm³).

La Figure 5.4 montre les zones de couches de charbon identifiées pour un trou de forage.

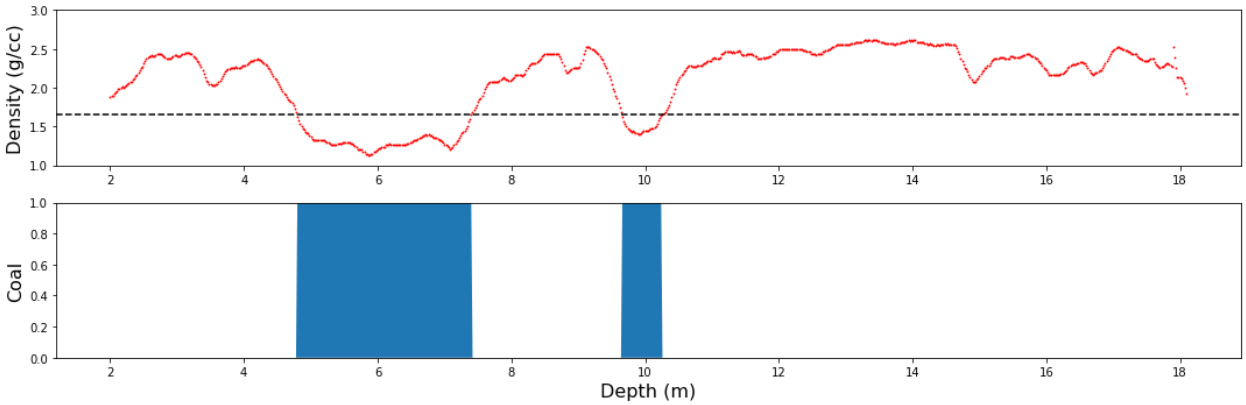


Figure 5.4 – Inférence du charbon à partir de relevé géologique pour un trou de forage

Afin de tirer parti de la granularité élevée des données MWD, chaque signal MWD est lissé à l'aide d'une moyenne mobile triangulaire sur une fenêtre mobile de 201 éléments (suffisant pour couvrir une profondeur de 0,02 m). Les poids sont normalisés et la fenêtre est centrée sur le point milieu de la fenêtre (poids le plus élevé). Soit X_i l'entrée i (classée par profondeur croissante) d'un signal MWD donné, et X_i^{av} le signal pondéré correspondant à cette entrée i, alors les poids associés sont:

$$X_i^{av} = \sum_{k=i-100}^{i+100} w_k X_k \tag{43}$$

Où $\sum_{k=i-100}^{i+100} w_k = 1$ et $w_k = \frac{4}{(n-1)^2}\left(\frac{n-1}{2} - \left|\frac{n+1}{2} - k\right|\right)$

Cette agrégation lisse le signal, réduisant ainsi la variabilité et atténuant les valeurs aberrantes potentielles, comme illustré dans la Figure 5.5. La Section 5.4.2 explique comment la variabilité des signaux est calculée et utilisée comme caractéristique d'entrée.

Après cette étape, les signaux MWD individuels sont filtrés pour correspondre aux relevés de densité en utilisant la profondeur (seuls les points de données correspondant aux entrées de profondeur des relevés de densité sont conservés pour chaque trou).

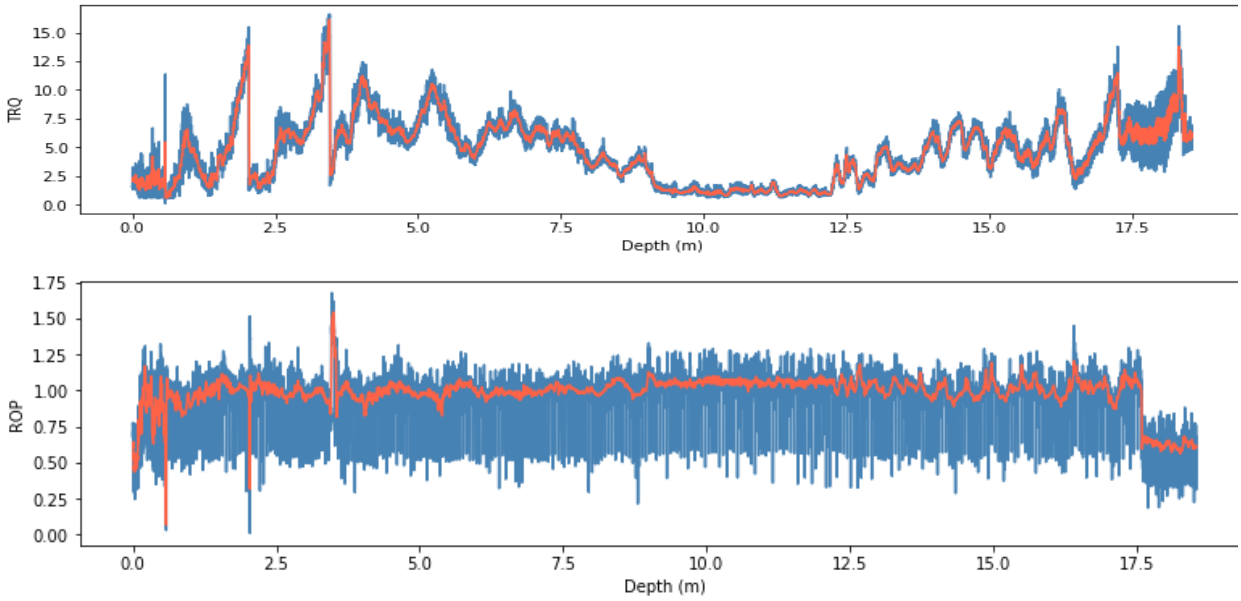
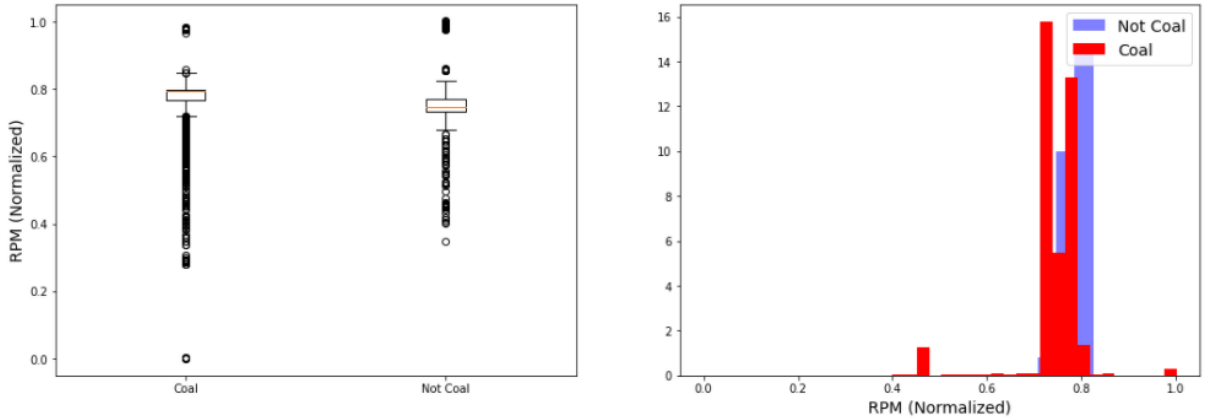


Figure 5.5 – Exemple d’agrégation et de lissage des MWD pour un trou de forage

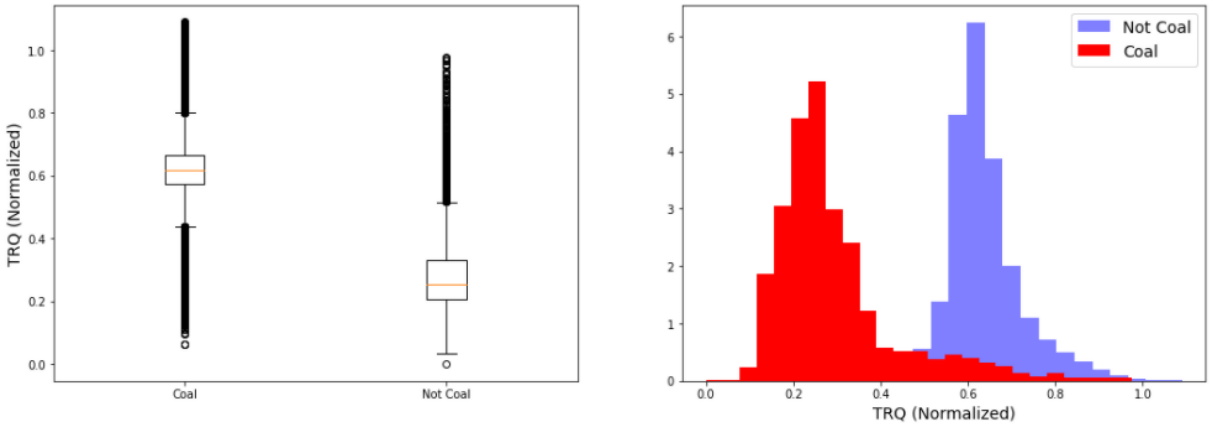
5.3.2 Exploration des données

L'intérêt se porte sur toute indication montrant qu'un certain signal MWD pourrait être un bon discriminateur de charbon. Une première évaluation est réalisée en comparant les diagrammes en boîte et les histogrammes de chaque signal MWD (voir Figure 5.6).

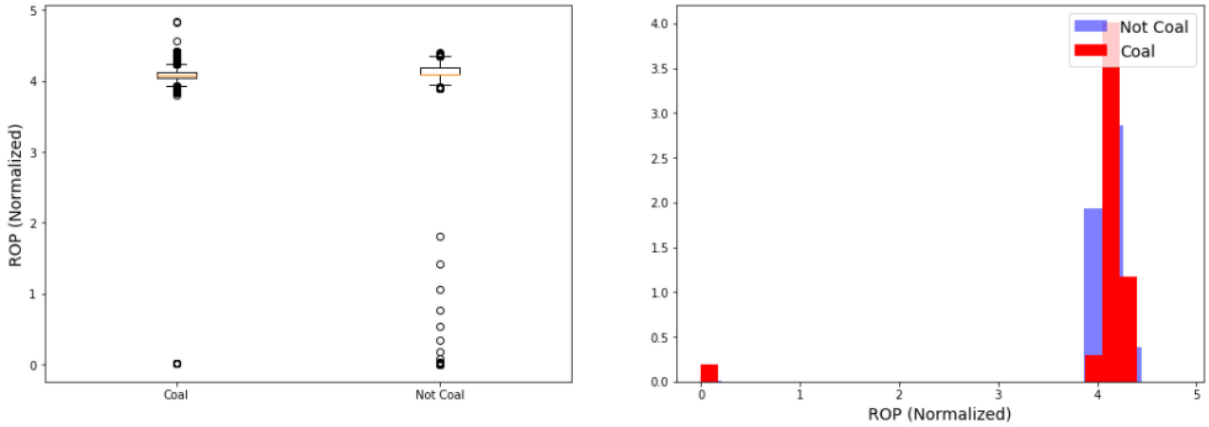
La Figure 5.6-a montre que le RPM n'est pas un discriminateur clair de charbon, car les valeurs médianes de RPM dans les deux classes sont similaires et les histogrammes montrent que les moyennes des deux distributions sont proches. En revanche, le couple (TRQ, Figure 5.6-b), le taux de pénétration (ROP, Figure 5.6-c), et l'énergie spécifique de forage (SED, Figure 5.6-e) sont de bons discriminateurs entre charbon et le non-charbon car pour chacune de ces variables les médianes des classes sont différentes, et les histogrammes suggèrent des distributions distinctes. La pression sur le trépan (WOB, Figure 5.6-d) semble être un faible discriminateur, car les valeurs médianes sont très proches pour les deux classes.



a- RPM standardisé (*Charbon* vs *Pas de charbon*)

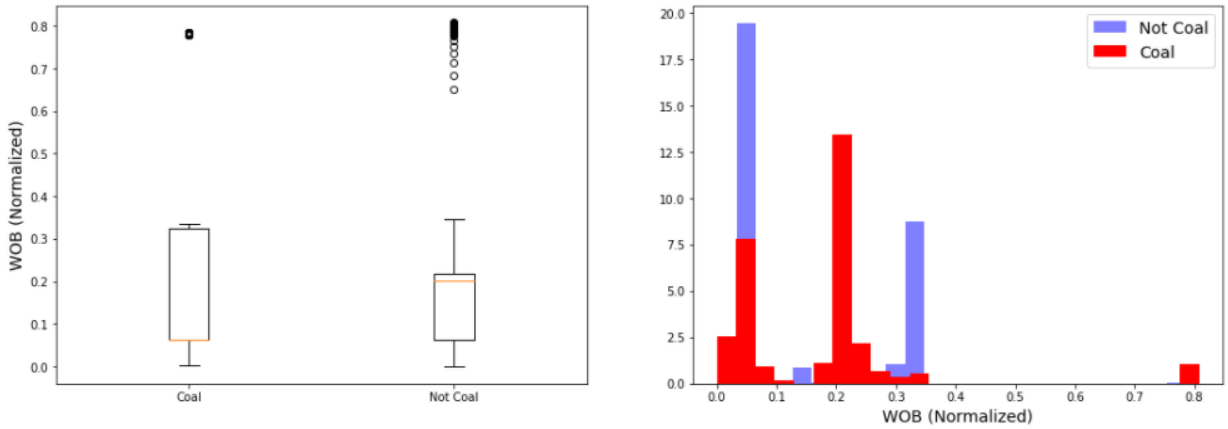


b- Torque standardisé (*Charbon* vs *Pas de charbon*)

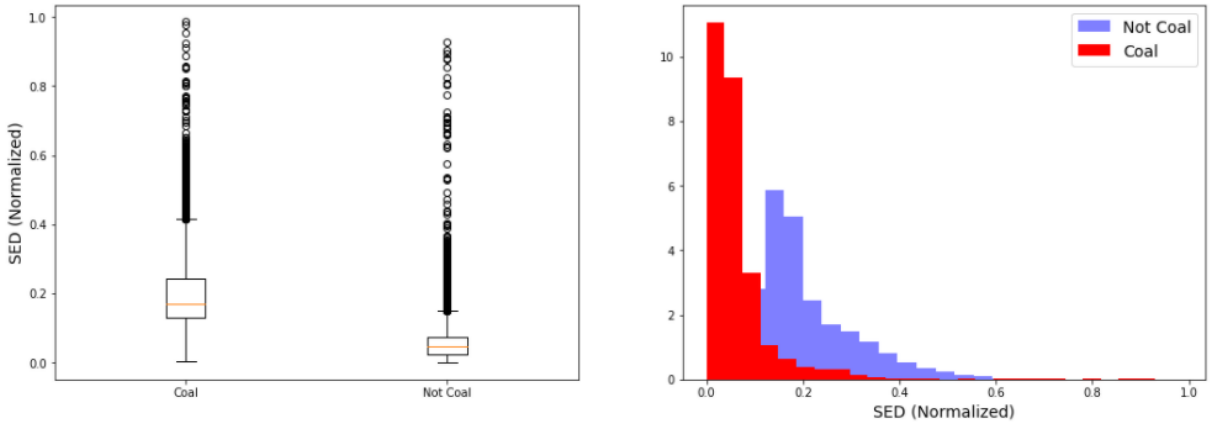


c- ROP standardisé (*Charbon* vs *Pas de charbon*)

Figure 5.6 – Diagrammes Box-plot et histogrammes pour chaque signal MWD



d- WOB standardisé (*Charbon* vs *Pas de charbon*)



e- SED standardisé (*Charbon* vs *Pas de charbon*)

Figure 5.6 – Diagrammes Box-plot et histogrammes pour chaque signal MWD - Suite

5.3.3 Partition des données

Les données correspondant à six trous de forage contenant des couches de charbon ont été mises de côté pour des tests sur l'emplacement de la couche de charbon supérieure.

Les données comportent moins de 10% (6,3 %) d'instances de la classe *Charbon*. Afin de réduire le déséquilibre de classes dans l'ensemble de données d'apprentissage, une approche de sous échantillonnage aléatoire dans la classe sur-représenté est utilisée : plusieurs instances de la classe *Pas-Charbon* sont aléatoirement éliminées, ce qui aboutit à un déséquilibre de classes de 25 % de *Charbon* – 75 % de *Pas-Charbon* (comparé à l'initial).

Un ensemble de 80% de ces points de données aléatoirement est dédié au processus d'entraînement-validation, le reste étant attribué à l'ensemble de test pour le calcul des indicateurs de performance.

5.4 Modèle de classification supervisés pour la prédiction du charbon

5.4.1 Choix des modèles de classification

La section 2.2.2 décrit les différents algorithmes de classifications supervisés sélectionnés. Le Tableau 1 résume les avantages et les inconvénients des différents modèles.

Étant donné le fort déséquilibre des classes (75 % des données représentent la classe *Non-Charbon* dans les données d'entraînement), l'indicateur de performance habituel, l'*exactitude* (voir section 2.2.2.7) serait non fiable comme mesure de la performance des modèles (un algorithme trivial attribuant toujours la classe *Non-Charbon* aurait un score d'exactitude de 75 % !). Le F1-Score (voir section 2.2.2.7), qui est la moyenne harmonique de la *précision* et du *rappel*, est utilisé pour évaluer la qualité globale des modèles de classifications.

Une fois les prédictions individuelles effectuées, ces prédictions seront ensuite intégrées en une prédiction unique par moyenne pondérée, afin d'obtenir un modèle plus robuste.

Tableau 1 – Avantages et inconvénients des modèles de classification supervisés proposés

Modèle	**Avantages**	**Inconvénients**	**Notes**
Régression Logistique (RL)	▪ Simple ▪ Ne nécessite pas beaucoup de données.	▪ Pas toujours efficace pour des relations non linéaires ou complexes entre les variables	▪ Choix pratique car le nombre de points de données n'est pas très élevé et que la dimensionnalité des données est relativement faible (6 variables).
Gradient Boosting (XGB)	▪ Performance accrue ▪ Utile en cas de fort déséquilibre des classes	▪ Nécessite un ensemble de données plus large que les modèles plus simples	▪ Choix intéressant car les données sont très débalancées.
Arbre de Decision (DT)	▪ Peut donner des résultats même en cas de valeurs manquantes dans les paramètres d'entrée	▪ Ne généralise pas bien dans les régions de l'espace des attributs ou il n'y a pas eu de données	▪ Choix intéressant car un des signaux peut parfois être erroné ou en dehors de l'intervalle rencontré dans les données d'entrainement.
Bagging avec Arbres de Décision (BDT)	▪ Stable et fiable	▪ Nécessite un ensemble de données plus large	▪ Améliore la performance des arbres de décision simple.
k-Plus Proches Voisins (kNN)	▪ Simple mais puissant ▪ Ne nécessite pas beaucoup de données d'entraînement ▪ Peut gérer des relations non linéaires complexes entre variables	▪ Naturellement sujet au surapprentissage ▪ Peut être trompeur dans des zones de l'espace des caractéristiques où les densités de classes sont faibles	▪ Choix intéressant car toutes les variables sont continues et la taille de l'ensemble de données est relativement petite.
Machines à Vecteurs de Support (SVM)	▪ Simple mais très polyvalent Particulièrement efficace dans les espaces de haute dimension	▪ Très sensible au surapprentissage	▪ Choix intéressant à explorer pour un cas de classification binaire

5.4.2 Ingénierie et sélection des caractéristiques

Les signaux MWD de base fournis peuvent être utilisés comme caractéristiques d'entrée dans les modèles supervisés. Plusieurs transformations des variables MWD d'origine sont également explorées. Parmi ces nouvelles variables figurent les coefficients de variation des signaux MWD, qui mesurent la variabilité des signaux. L'idée est que certaines variables peuvent fluctuer différemment lors du forage dans des roches tendres par rapport aux roches dures. Pour capturer cette caractéristique, une nouvelle variable est introduite : le coefficient de variation (CV) de la variable, qui est le rapport de l'écart-type sur la moyenne :

$$X_{CV_i} = \frac{\sqrt{\sum_{k=i-100}^{i+100}\left(X_k - X_i^{avg}\right)^2}}{X_i^{avg}} \tag{44}$$

Dans cette équation, l'écart-type est calculé en utilisant l'ensemble des données MWD brutes dans la fenêtre correspondante à la moyenne (moyenne mobile).

En plus du coefficient de variation, plusieurs transformations de chaque signal MWD (valeur carrée, inverse, racine carrée) sont ajoutées comme variables d'entrée potentielles. L'objectif est qu'une transformation spécifique d'un signal puisse être plus corrélée avec la présence de charbon ou de densité que le signal d'origine. Le coefficient de corrélation de Pearson est utilisé pour mesurer le degré de corrélation entre deux variables. La Figure 5.7 présente la matrice des coefficients de corrélation entre toutes les variables et les références géologiques (densité de référence et charbon de référence). L'intérêt se porte sur les variables ayant la plus forte corrélation (positive ou négative) avec la densité et le charbon de référence provenant des relevés géophysiques. L'objectif est de sélectionner un sous-ensemble de variables avec une intercorrélation limitée et une forte corrélation avec la densité et le charbon.

La Figure 5.8f montre le coefficient de corrélation entre les variable originales. Elle confirme que le couple (TRQ) et l'énergie spécifique de forage (SED) sont les plus corrélés avec la densité et le charbon. Les autres parties de la Figure 5.8 présentent les coefficients de corrélation entre plusieurs transformations d'une même variable. Par exemple, le coefficient de variation du ROP (ROP_{cv}) est significativement plus corrélé au Charbon de Référence que le ROP. De même, l'inverse de la SED (SED_{inv}) présente une meilleure corrélation avec la densité que la SED. Après analyse des matrices de coefficients de corrélation, le sous-ensemble de variables sélectionné comme caractéristiques d'entrée pour les modèles supervisés est : RPM, TRQ, ROP_{cv}, WOB et SED_{inv}.

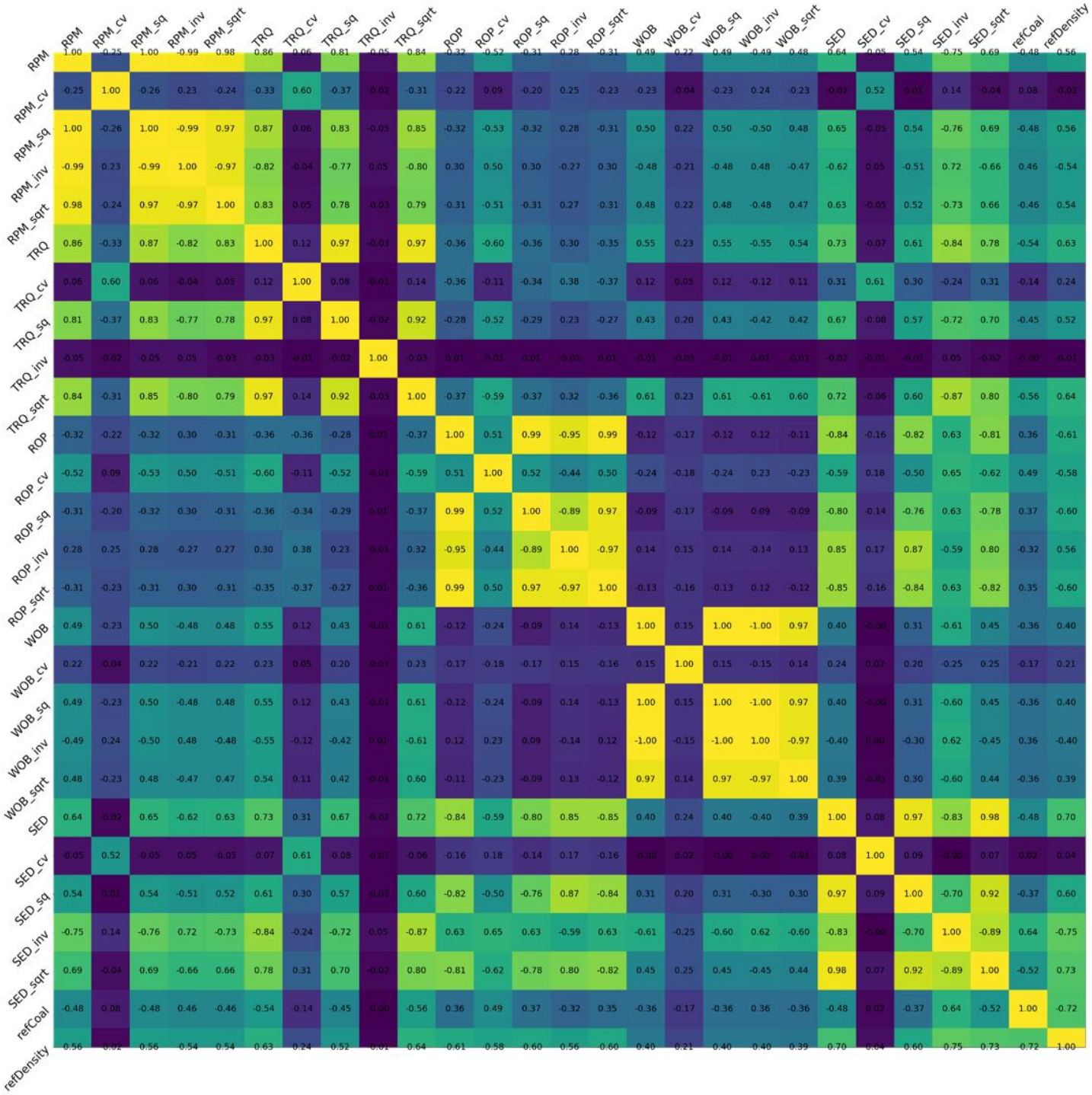
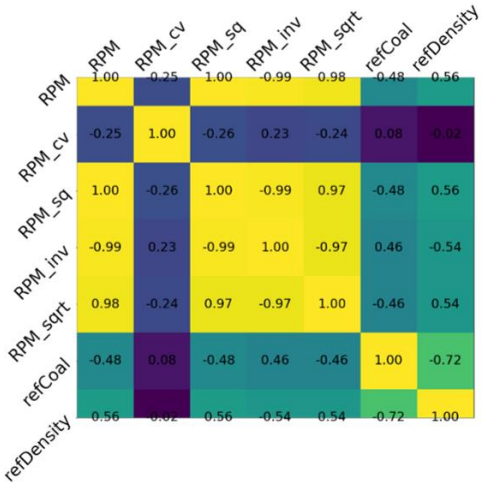
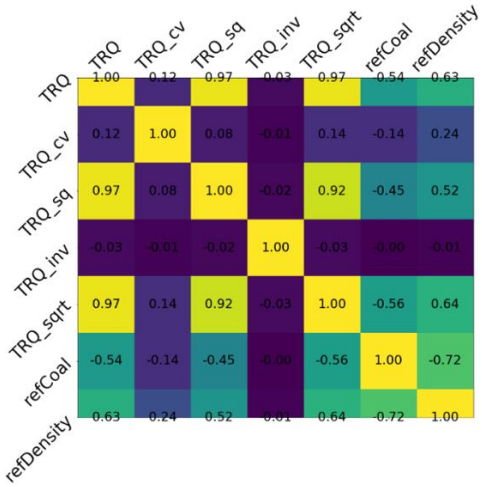


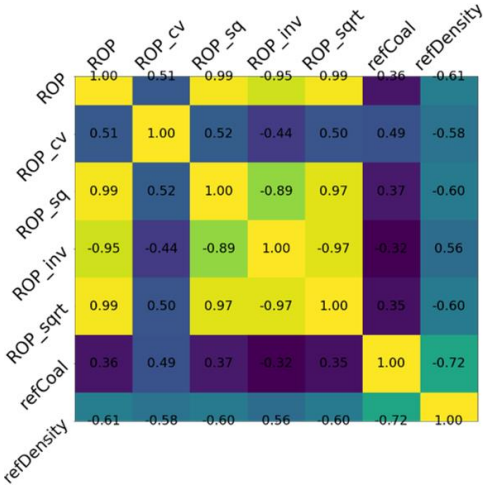
Figure 5.7 – Coefficients de corrélation entre les MWD transformés et la densité (complet)



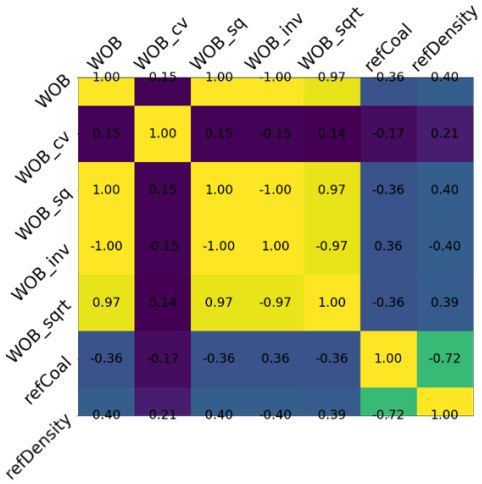
a- Vitesse de rotation (RPM)



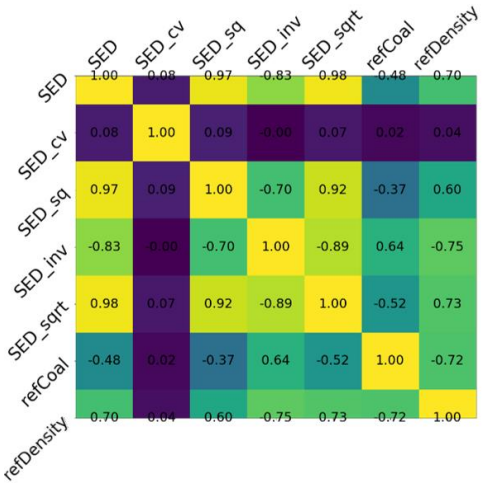
b- Couple (TRQ)



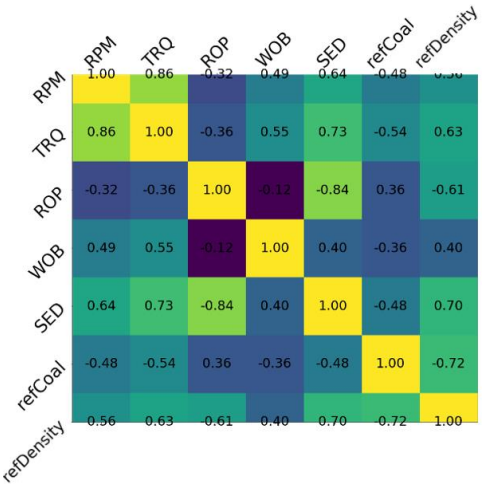
c- Vitesse de pénétration (ROP)



d- Force totale sur le tricône (WOB)



e- Énergie spécifique



f- Variables MWD originales

Figure 5.8 – Coefficients de corrélation entre les MWD transformés et la densité (partiels)

5.4.3 Résultats des tests

Les différents modèles sont entraînés en utilisant le même ensemble de données d'entraînement et de validation pour chaque modèle. Les meilleurs hyperparamètres pour chaque modèle sont sélectionnés par une stratégie de gridsearch. Le Tableau 2 résume la performance des modèles sur l'ensemble de test. Le F1-score donne un aperçu de la qualité globale des modèles. La Régression Logistique offre la meilleure qualité globale avec un F1-Score de 85,71 %. Lorsque le Bagging est appliqué, la performance du modèle d'Arbre de Décision simple est améliorée de 10 %. Bien que les modèles basés sur les Arbres de Décision aient une qualité globale inférieure, ils restent un choix intéressant lorsqu'un signal est manquant ou erroné.

Tableau 2 – Performance des modèles sur l'ensemble de test

Modèle	*Precision*	*Recall*	**F1-Score**
Regression Logistique (RL)	*88.97%*	*82.69%*	**85.71%**
Extreme Gradient Boosting (XGB)	*84.28%*	*85.90%*	**85.08%**
Arbre de décision (DT)	*77.78%*	*69.14%*	**73.20%**
Bagging sur Arbres de décision (BDT)	*78.36%*	*82.72%*	**80.48%**
k-plus proches voisins (KNN)	*86.21%*	*80.65%*	**83.33%**
Support Vector Machine (SVM)	*92.17%*	*78.83%*	**84.44%**

Pour tirer parti des capacités des différents modèles, un modèle intégré est construit en faisant une moyenne pondérée des prédictions de charbon des modèles RL, XGB, BDT, KNN et SVM. Le poids attribué à chaque modèle est proportionnel à sa qualité. Les différents modèles ont également été testés sur les trous mis de côté qui contiennent des larges couches de charbon (trou A et trou C). La Figure 5.9 et la Figure 5.10 présentent les prédictions de charbon et les probabilités de charbon pour ces trous.

Dans le trou A, le sommet réel de la couche de charbon est situé à une profondeur de 10,24 m (grande couche unique). Les modèles identifient le sommet de la couche de charbon à des profondeurs allant de 10,22 m à 10,24 m, ce qui correspond à une erreur de profondeur absolue maximale de 0,04 m, bien en deçà de la marge d'acceptation de ±0,20 m. Le modèle intégré identifie cette couche de charbon à 10,24 m. Tous les modèles prédisent une petite couche de charbon (non exploitable) à environ 9,26 m. Cette couche de charbon n'est pas présente dans la relevé

géophysique, cependant, en observant la courbe de densité, une diminution de la densité peut expliquer la présence de cette couche de charbon erronée dans les prédictions.

Le trou C comporte deux couches de charbon distinctes. La première, située à une profondeur de 7,82 m avec une largeur de 0,28 m, peut être classée comme une couche mineure ou erronée. La seconde, plus significative, se trouve à 9,16 m et a une largeur de 3,16 m. Les prédictions des différents modèles pour cette grande couche de charbon vont de 9,22 m (Régression Logistique) à 9,28 m (Arbre de Décision), soit une erreur absolue maximale de 0,12 m, ce qui reste dans la marge d'acceptation de ±0,20 m. Le modèle intégré identifie la couche de charbon à 9,22 m (erreur absolue de 0,06 m).

Plusieurs des modèles individuels testés dans cette section apparaissent comme de bons discriminateurs de charbon et peuvent être utilisés pour déduire la position de la couche de charbon avec une précision satisfaisante (c'est-à-dire dans la marge d'acceptation de ±0,20 m).

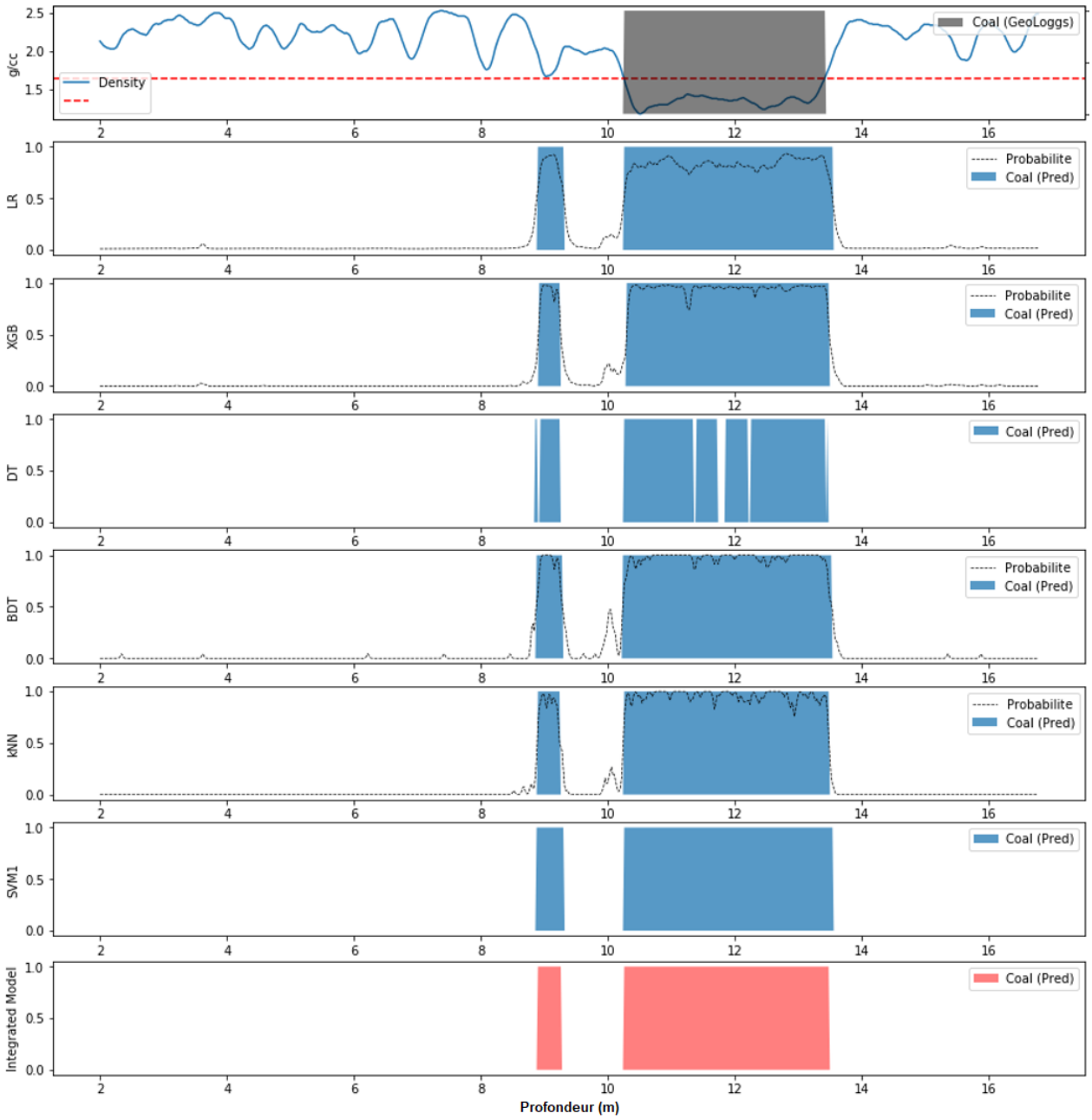


Figure 5.9 – Comparaison des prédictions pour le trou de forage A

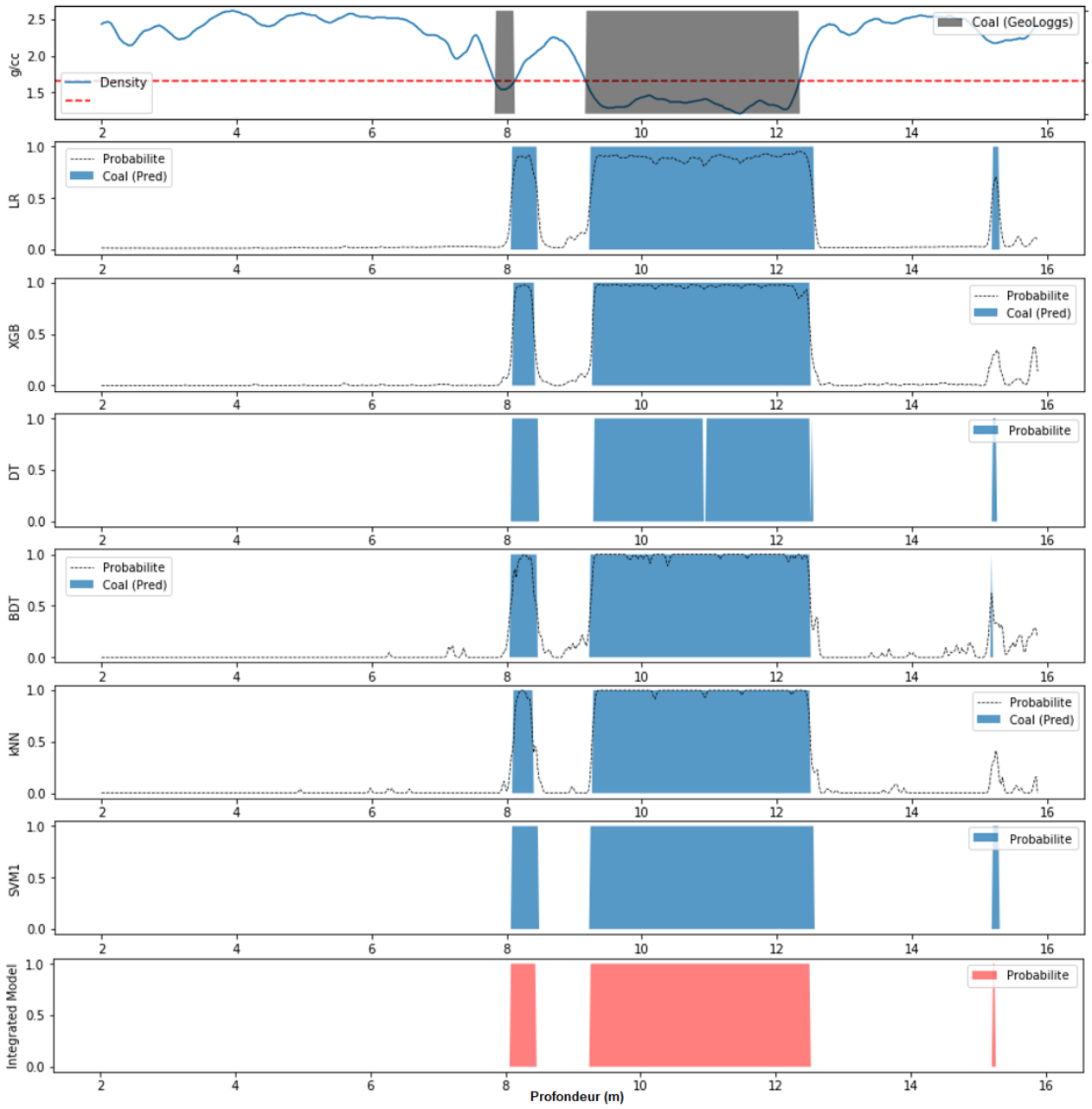


Figure 5.10 – Comparaison des prédictions pour le trou de forage C

5.4.4 Analyse de sensibilité et importance de combiner plusieurs modèles

Dans cette section, une analyse de sensibilité est utilisée pour démontrer l'importance de disposer de plusieurs classificateurs individuels et de les combiner. Comme mentionné à la section 5.4.3, certains modèles, comme l'arbre de décision (DT et BDT), même avec un F1-Score plus faible dans ce contexte, restent importants pour gérer les situations où une valeur MWD peut être erronée ou hors de portée.

Même lorsque des valeurs MWD sont erronées ou en dehors de leur intervalle habituel, les modèles d'apprentissage de classification peuvent toujours produire une prédiction de charbon. Ces prédictions doivent être signalées, et une analyse pour comprendre comment les valeurs hors normes de MWD affectent la probabilité de charbon devrait être réalisée.

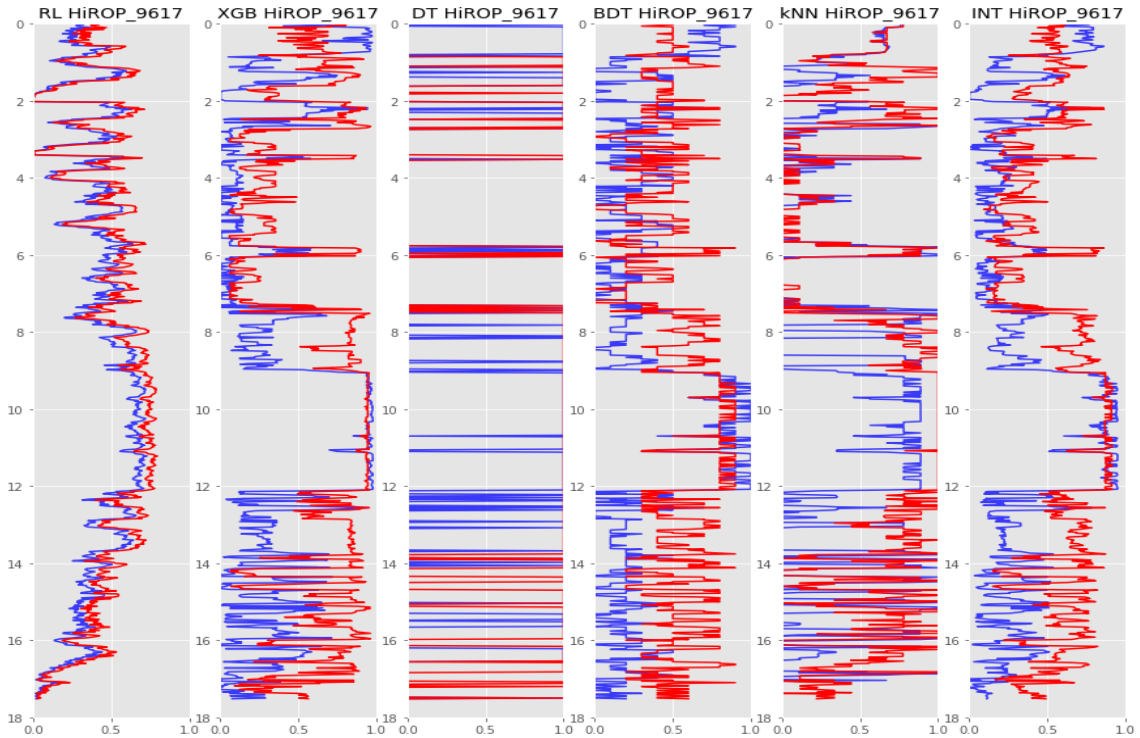
Dans ce test de sensibilité, un cas de référence est sélectionné (un seul trou de forage) avec des variations relativement élevées dans les signaux MWD. Plusieurs cas tests sont ensuite évalués, dans lesquels une mesure MWD est remplacée par une valeur artificielle (et erronée), élevée ou basse, hors de portée pour ce MWD. Pour illustrer l'impact, des cas test avec ROP anormalement élevé, WOB anormalement élevé et SED anormalement faible ont été choisis.

La Figure 5.11 présente les résultats des cas de sensibilité. Les prédictions des classificateurs individuels et du modèle intégré pour les cas de sensibilité sont comparées à la prédiction de référence. Il est montré que, bien que la prédiction intégrée pour le cas de sensibilité soit acceptable (ou proche de la prédiction de référence), certaines prédictions des classificateurs individuels diffèrent considérablement. Pour une valeur de ROP élevée, l'impact est plus important sur les modèles RL et KNN, tandis que les modèles XGB et BDT semblent plus robustes. Les modèles XGB et BDT paraissent également plus robustes face à un WOB élevé, tandis que les modèles KNN et RL sont fortement impactés. Les modèles RL et KNN sont toutefois plus robustes face à une valeur de SED basse et erronés.

Cette analyse montre que la performance des différents modèles est variablement affectée par des erreurs dans les différents MWD en entrée. Ainsi, on peut associer des poids plus élevés à certains modèles dans certaines conditions de valeurs MWD.

Bleu: Prédiction avec MWD réel - ***Rouge***: Prédiction avec erreur sur MWD (cas de sensibilité)

Case :High ROP



Bleu: Prédiction avec MWD réel - ***Rouge***: Prédiction avec erreur sur MWD (cas de sensibilité)

Case: Low SED

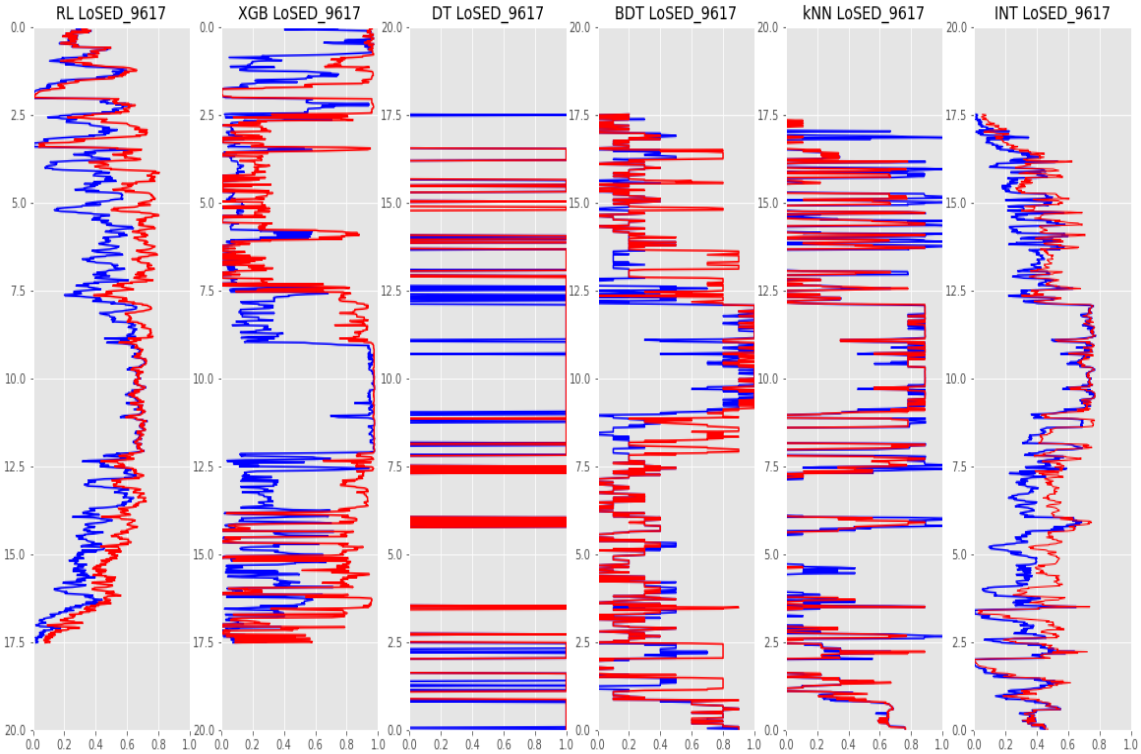


Figure 5.11 – Analyses de sensibilité pour les différents modèles de classification

5.5 Modèle de régression pour la prédiction de la densité

Dans cette section, un réseau de neurones artificiel est utilisé pour la régression de la densité. La prédiction de la courbe de densité fournit un deuxième outil pour la classification du charbon (en utilisant le seuil de densité) mais, plus important encore, permet d'identifier ou d'anticiper les transitions charbon/roche dure en suivant les variations de densité.

5.5.1 Réseau de neurones artificiels pour prédiction de la densité

La complexité du réseau de neurones est limitée par la taille relativement faible de l'ensemble des données disponibles pour cette recherche. Un réseau de neurones entièrement connecté simple est proposé pour prédire la densité à partir des données MWD.

Le même ensemble de variables que celui de la section 5 est utilisé comme caractéristiques d'entrée : RPM, TRQ, ROP_cv, WOB et SED_inv. Dans ce cas, l'ingénierie des caractéristiques proposée peut améliorer la performance du modèle, ce qui est particulièrement crucial compte tenu du faible nombre de points de données dans l'ensemble d'entraînement.

Pour suivre le processus d'apprentissage, la métrique de qualité utilisée est l'erreur quadratique moyenne (MSE). 80 % de l'ensemble d'apprentissage est utilisé pour l'entraînement, et le reste pour la validation. Les données de 6 trous de forages ont mis de côté pour les tests.

Différentes architectures et hyperparamètres de réseau de neurones artificiels ont été testés. Le Tableau 3 présente le meilleur ensemble d'hyperparamètres trouvé pour le modèle optimal en termes de MSE.

Tableau 3 – Spécifications du meilleur modèle de réseau de neurones artificiel

Nombre de couches cachées	**Nombre de neurones dans les couches cachées**	**Nombre d'itérations d'entraînement/validation (époques)**	**Optimiseur**	**Taux d'apprentissage**
4	8	1000	RMSprop	0.001

La Figure 5.12 montre l'évolution du processus d'apprentissage-validation. Après environ 200 itérations, il est observé que la MSE s'est stabilisée pour la courbe d'entraînement et la courbe de validation, ce qui suggère une convergence. Le fait que la courbe de validation ne réaugmente pas après la convergence de l'entraînement (MSE de validation plus élevée) suggère que le modèle ne fait pas de surapprentissage sur les données d'entraînement.

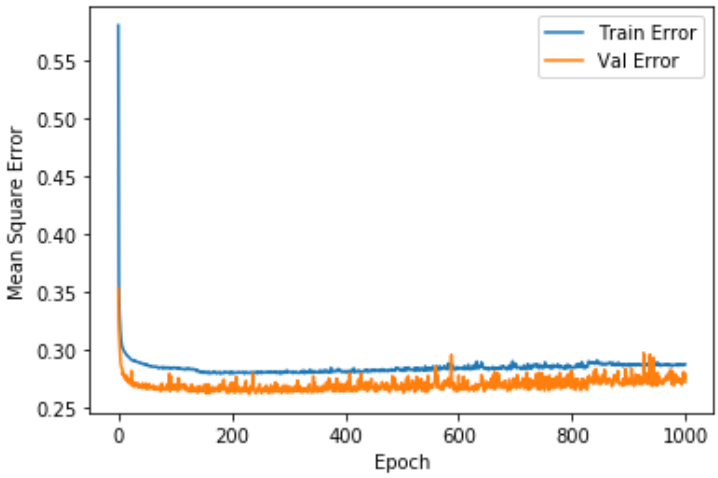


Figure 5.12 – Évolution de l'entraînement/validation pour le modèle de densité

Le modèle entraîné est ensuite utilisé pour les tests en utilisant les données des 6 trous mis de côté. La Figure 5.13 montre les prédictions de densité pour les 6 trous de test différents. Pour chaque trou, la ligne bleue représente la densité mesurée à partir des relevés géophysiques, et la ligne rouge est la prédiction du modèle. La MSE pour ces prédictions de densité varie entre 7,8 % pour le trou F et 1,7 % pour le trou D, la plupart des trous ayant une MSE autour de 3,5 %.

Pour tous ces trous, les prédictions du modèle montrent un bon ajustement bien que non parfait avec la densité observée dans les relevés géophysiques. Le modèle, sans être un prédicteur exact de la densité, est un excellent indicateur des tendances dans la densité (montée et descente de la densité de la roche forée) et peut être utilisé comme indicateur de densité en temps réel pour la détection du sommet de la couche de charbon, comme discuté dans la section 5.6.

À la différence des modèles de classification proposés à la section 5.4, ce modèle de prédiction de la densité sera moins performant s'il est utilisé uniquement pour déduire la présence de charbon en appliquant la règle de 1,65 g/cm³. Par exemple, dans la Figure 5.13e, bien que le modèle prédise les variations de densité en temps opportun, il est observé qu'il n'aurait pas pu classer clairement les points de données de la couche de charbon comme instances de charbon, car les faibles valeurs de densité prédites oscillent autour du seuil de 1,65 g/cm³. Le modèle de réseau de neurones artificiel est bon pour prédire la tendance générale de la densité mais moins efficace pour prédire précisément les faibles densités. Une explication est que le forage traverse des roches très tendres et les variations des signaux ne sont pas suffisamment nettes pour y évaluer précisément la densité.

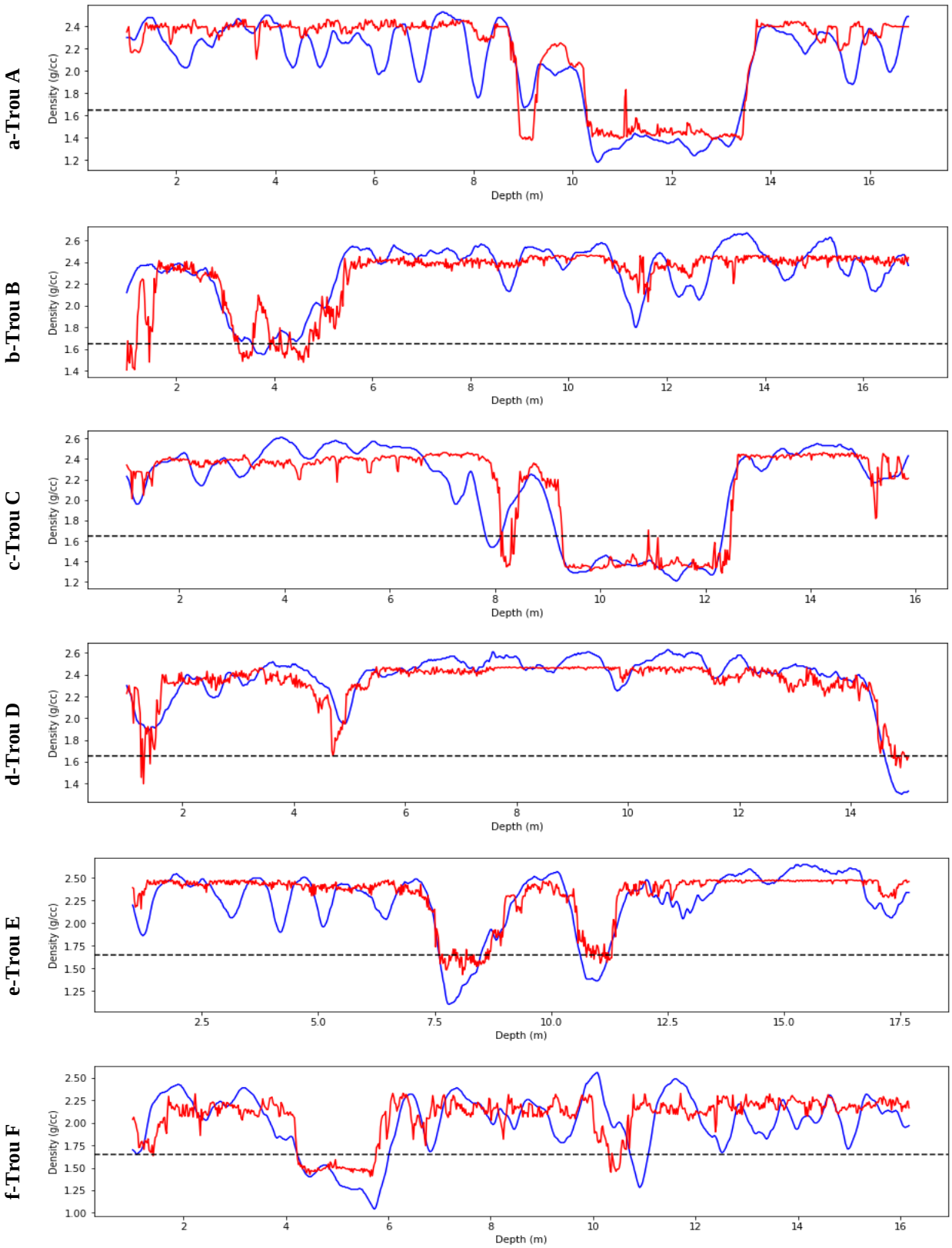


Figure 5.13 – Prédictions de densité pour les trous de forages tests

5.5.2 Correction de l'erreur d'alignement entre les MWD et les relevés géophysiques

Bien que la densité ne soit pas un bon prédicteur binaire du charbon, elle peut être utilisée pour calculer la différence potentielle d'alignement en profondeur entre les données MWD et les relevés géophysiques dont l'ajustement est parfois manuel. Cela se fait en trouvant l'écart de profondeur optimal qui donne la meilleure corrélation entre la densité prédite et la densité géophysique, comme illustré dans la Figure 5.14. Dans le premier exemple montré à la Figure 5.14, en reculant la profondeur de la densité de 20 cm, on obtient une meilleure corrélation (95%) entre la densité prédite à partir des données MWD et la densité mesurée.

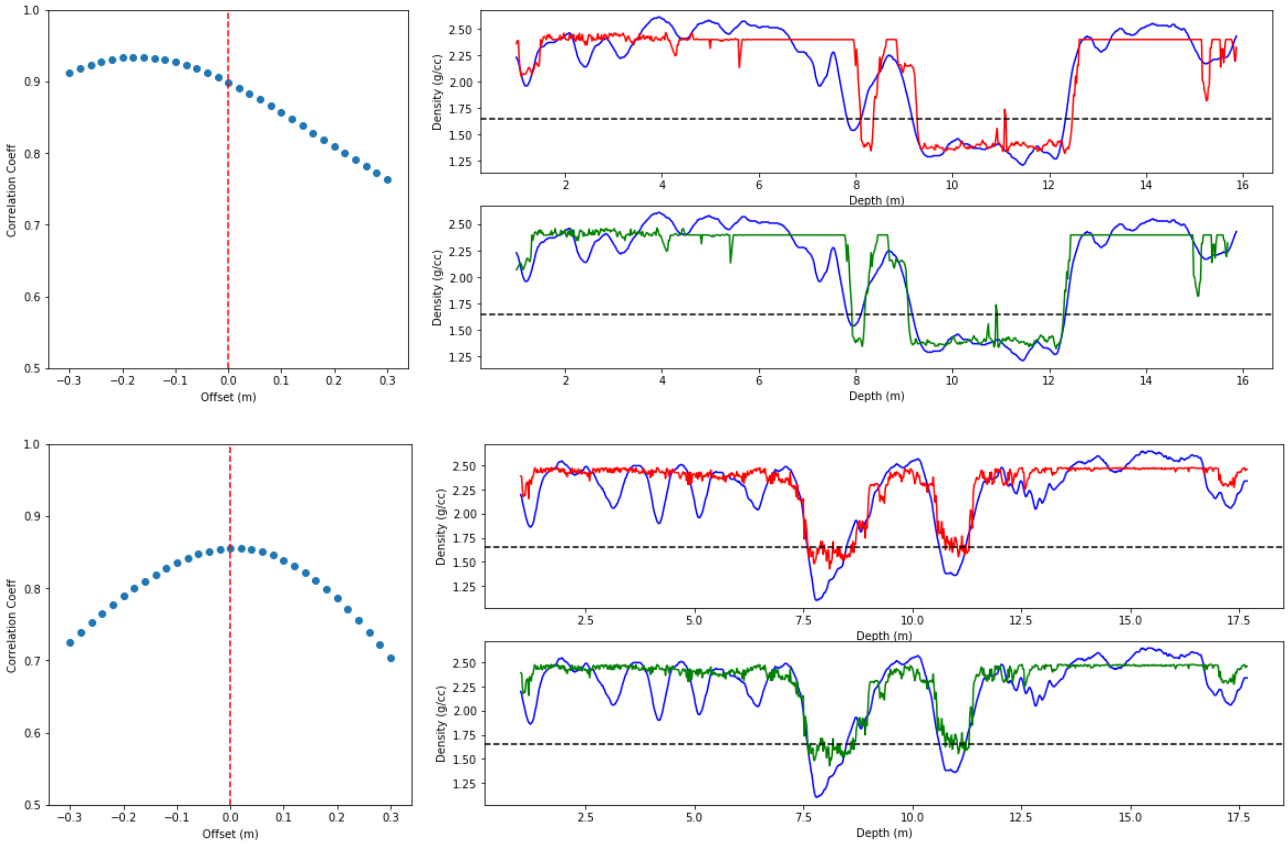


Figure 5.14 - Ajustement de l'alignement entre les MWD et les relevés géophysiques

5.6 Détection de la position de la couche de charbon

Les courbes de densité prédites montrent une corrélation plus élevée avec la densité mesurée que la probabilité de charbon issue des classificateurs binaires. Cela est attendu, car les classificateurs individuels n'ont jamais été conçus pour imiter les variations de densité, mais se limitent plutôt à déduire, à partir des valeurs MWD, s'il s'agit de charbon ou non (ou, alternativement, si la densité est faible ou élevée). Ils n'offrent donc pas une perspective plus riche sur l'évolution des conditions souterraines ou l'approche d'une couche de charbon.

Une règle de seuil simple peut être utilisée pour déduire l'emplacement du sommet de la couche de charbon en utilisant la prédiction de. Dans la Figure 5.15 la densité prédite varie autour d'une valeur élevée (2,4 g/cm³) lorsqu'il s'agit de roche dure, avant de chuter à une valeur plus basse en entrant dans le charbon. La règle de décision proposée pour identifier la transition est de suivre la densité prédite et de marquer la profondeur lorsqu'elle descend en dessous de 1,8 g/cm³ pour une période prolongée ou distance d'attente de 16 cm. Le critère de distance d'attente est ajouté pour disqualifier les filons mineurs de charbon (couches de charbon insignifiantes). Les deux paramètres (seuil de 1,8 g/cm³ et distance d'attente de 16 cm) peuvent être ajustés selon un compromis entre la sensibilité (détection en temps opportun) et la précision de l'identification du sommet de la couche de charbon.

La Figure 5.15 montre la prédiction du sommet de la couche de charbon pour les trous de test lorsque ces règles sont appliquées, ainsi que la comparaison avec les emplacements réels des sommets de charbon. Sur les 9 sommets de charbon réels, un seul n'a pas été identifié à temps (deuxième couche de charbon dans la Figure 5.15-E). Les autres sommets ont été détectés dans la marge acceptable de ±20 cm.

Deux faux sommets de charbon (faux positifs) ont été identifiés (Figure 5.15-A et B) ; Ces faux sommets peuvent être écartés en combinant les deux outils : le modèle de prédiction de densité et le modèle de classification du charbon, ou plus efficacement, en incluant des données contextuelles (connaissance préalable de l'emplacement attendu des couches à partir des trous de forage voisins).

Le modèle de prédiction de densité peut ainsi être utiliser pour détecter les transitions roche dure/charbon avec une précision acceptable.

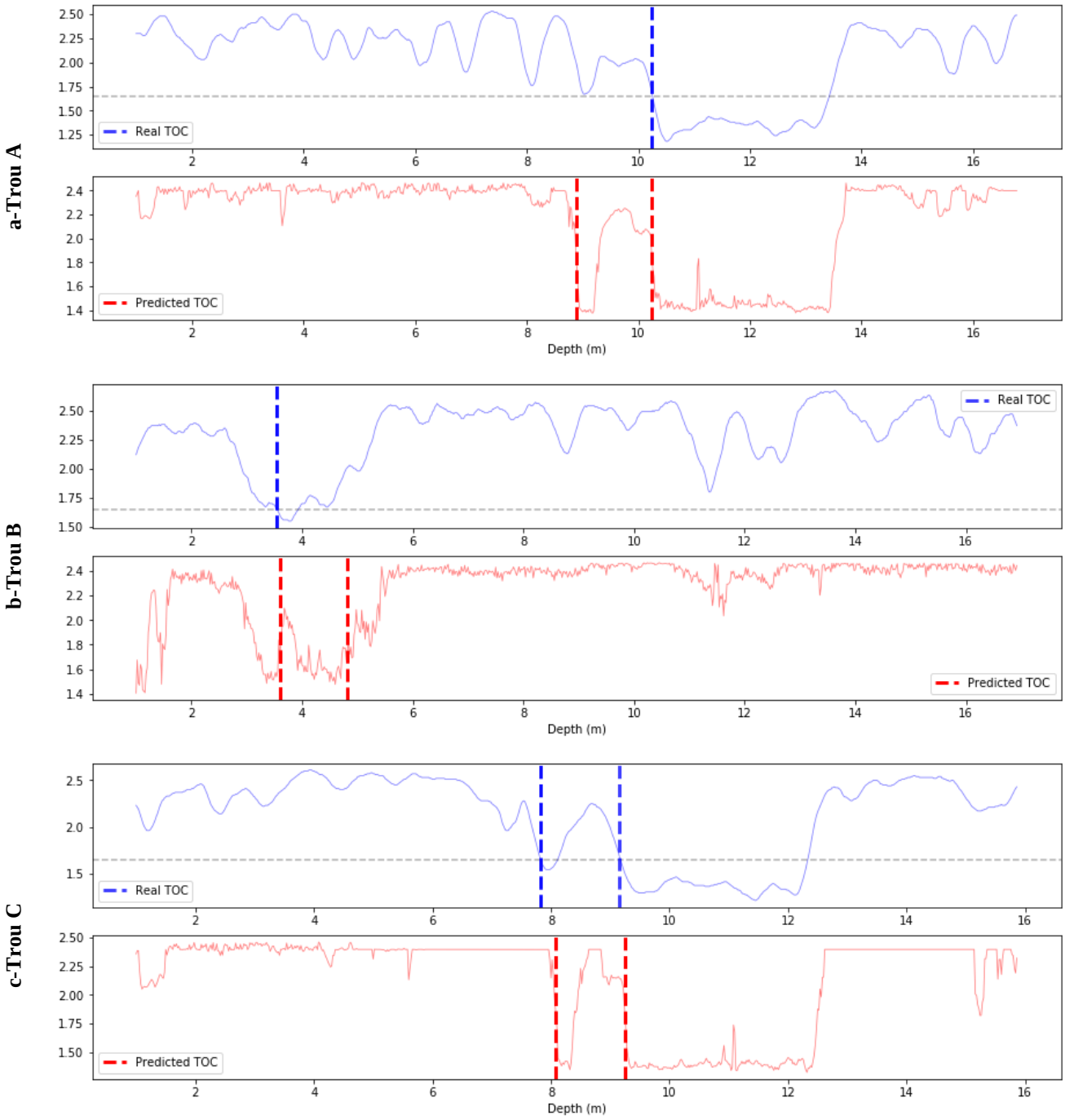


Figure 5.15 – Identification de la profondeur de la couche de charbon pour les trous tests

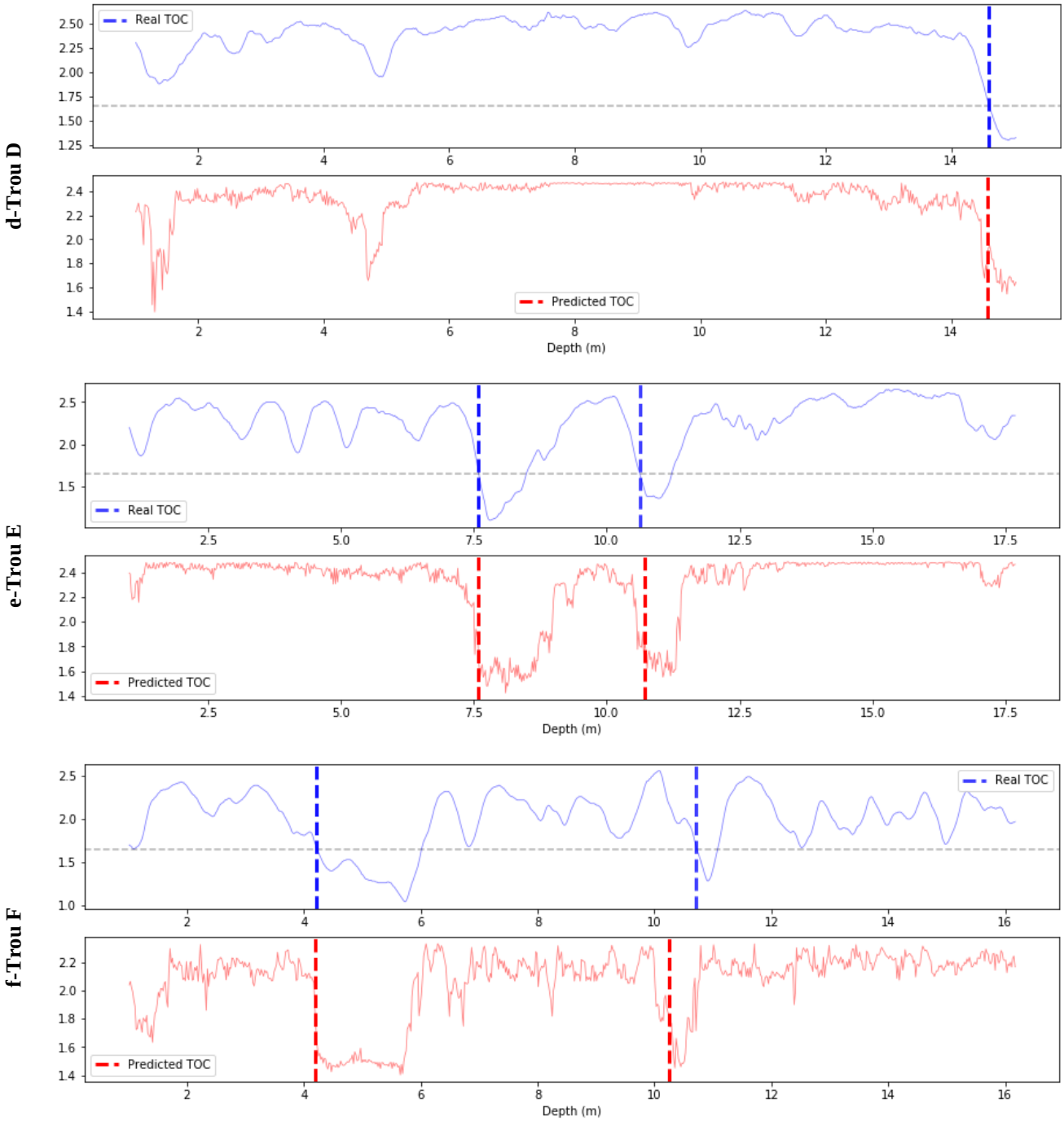


Figure 5.15 – Identification de la profondeur de la couche de charbon pour les trous tests - Suite

5.7 Discussion sur la généralisation des modèles

Les modèles proposés dans cette étude sont basés sur des données d'entraînement recueillies pour une seule foreuse dans des conditions contrôlées, et plusieurs hypothèses sont faites pour simplifier le problème. Cela soulève la question de la bonne généralisation de ces modèles : Que se passerait-pour une autre foreuse qui seraient opérée dans des conditions différentes, par exemple avec des points de consigne différents?

Les variables MWD d'entrée sont mises à l'échelle et standardisées en fonction des données disponibles. Cela permet de garantir une certaine généralisation des modèles, il suffirait de recueillir suffisamment de données et de standardiser les signaux MWD en fonction de leur contexte.

Des changements drastiques dans les conditions de forage pourraient survenir, nécessitant une mise à jour (réentraînement) ou l'utilisation d'un nouveau jeu de modèles. En général, les bonnes pratiques en matière de développement et de maintenance des modèles d'apprentissage automatique suggèrent de surveiller la performance du modèle et de procéder à des mises à jour lorsque l'on observe un changement significatif dans leur performance. Cependant, cela n'est possible que si des observations des classes réelles (relevés géophysiques) sont disponibles et accessibles en temps opportun. Dans le cas présent, les relevés géophysiques sont coûteux, et ne permettent pas une évaluation continue de la performance des modèles en temps réel. Dans ces conditions, il est proposé de surveiller la distribution des données MWD dans l'espace des caractéristiques.

Une approche basée sur la divergence de Kullback–Leibler (Wu, et al., 2019) est proposée pour indiquer quand un réentraînement du modèle est nécessaire :

$$D_{KL}(f\|g) = \sum_{x\in\Omega} f(x)\log\big(f(x)g(x)\big) \tag{45}$$

La divergence de Kullback–Leibler, ou entropie relative entre deux fonctions de densité de probabilité, mesure la différence entre une distribution de probabilité et une seconde, prise comme référence. Considérons deux distributions de probabilité f (distribution observée de MWD_i au temps t_1) et g (distribution observée de MWD_i au temps t_2) définies sur le même espace de probabilité Ω. Lorsque cette valeur change au fil du temps ou atteint un seuil prédéfini, cela indique qu'il est temps de réentraîner les modèles pour mieux refléter les conditions changeantes.

5.8 Conclusion

Ce chapitre traite des problèmes de classification du charbon et d'identification du sommet de la couche de charbon, en utilisant les données MWD. Pour la classification plusieurs classificateurs d'apprentissage automatique supervisé sont appliqués, leurs prédictions sont combinées pour améliorer la performance et la fiabilité de la classification. En utilisant le même ensemble de données et de variables, un réseau de neurones artificiel entièrement connecté est entrainé pour prédire la densité, ce qui s'avère être un bon outil pour prédire l'approche d'une couche de charbon. La performance de cet outil pour la détection d'une couche de charbon pourrait être améliorée en combinant ses sorties avec la classification du charbon et en rajoutant plus de contexte (prédictions dans les trous voisins).

Cette étude est limitée par le nombre relativement faible de données disponibles, notamment les relevés géophysiques, ce qui a empêché l'utilisation de modèles de réseaux de neurones artificiels plus sophistiqués. Les travaux futurs dans ce domaine bénéficieront de la disponibilité de plus de données et permettront de développer des modèles plus précis avec des architectures de réseaux de neurones plus riches. La performance des modèles futurs sera également améliorée par l'utilisation de nouvelles caractéristiques fournies par des capteurs additionnels pour différentes variables telles que la vibration de la tige de forage, le son, etc. L'utilisation de données contextuelles, telles que les connaissances préalables sur la géologie locale ou l'emplacement des couches de charbon dans les trous environnants, améliorera également considérablement la fiabilité des résultats. Plusieurs approches seront étudiées dans le Chapitre 5. Les modèles futurs devront prendre en compte différents réglages de la foreuse et l'état du trépan de forage (Chapitre 6), qui influent fortement sur la performance du forage.

CHAPITRE 6 APPRENTISSAGE NON SUPERVISÉ POUR LA PRISE EN COMPTE DES TROUS DE FORAGE VOISINS

6.1 Introduction

6.1.1 Contexte et enjeux

Dans le chapitre précédent, les données MWD sont utilisées pour inférer la présence de charbon en quasi-temps réel dans un seul trou de forage (identifié dans ce chapitre par prédiction locale), à l'aide de différents algorithmes d'apprentissage automatique supervisé (classificateurs binaires). Ces informations sont ensuite utiles pour le chargement approprié des explosifs, un processus crucial dans les opérations de dynamitage et de chargement. Cependant, l'identification du type de roche dans un seul trou de forage (identification locale) est sujette à un certain niveau d'incertitude en raison de plusieurs facteurs tels que les erreurs inhérentes dans les mesures MWD, les variations géologiques autour des transitions charbon/roche dure, l'usure du trépan tricône qui affecte la réponse du forage, la configuration de la foreuse, les actions de l'opérateur, etc.

Par ailleurs, il est parfois difficile de connaitre avec certitude l'emplacement du sommet du charbon dans un seul trou de forage, car la transition entre les types de roche n'est pas toujours nette. De plus, quand la probabilité de charbon est très proche du seuil de 50%, il est difficile de conclure si on est réellement en présence d'une couche de charbon ou non. La Figure 6.1 montre un exemple de faux négatif où le modèle supervisé intégré n'a pas identifié une large couche de charbon que l'on peut inférer par les relevés géologique (densité), car la probabilité de charbon était légèrement inférieure au seuil de classification de 50% (près de 28% des points de données dans cet exemple ont une probabilité de charbon comprise entre 40% et 60%). La prédiction aurait été excellente si le seuil de classification avait été abaissé à 40%.

L'état du trépan peut affecter considérablement la réponse du forage. Les trépans neufs coupent très efficacement même dans les roches plus dures, ce qui pourrait être interprété comme des roches moyennes ou tendres si l'état du trépan n'est pas pris en compte. Les trépans usés ont tendance à augmenter les vibrations ce qui occasionne une perte de rendement du forage (notamment en vitesse de pénétration), et cela peut notamment conduire les modèles supervisés à interpréter ces situations comme étant des zones de roche dure.

Par ailleurs, les changements de points de consignes des paramètres de la foreuse lors du forage d'un même patron affectent la précision des prédictions. Les classificateurs des modèles supervisés fonctionnent mieux lorsque la réponse de la foreuse est non contrainte (mode de forage automatique) et donc uniquement affectée par la géologie, contrairement aux situations où l'opérateur modifie les paramètres pour augmenter sa productivité.

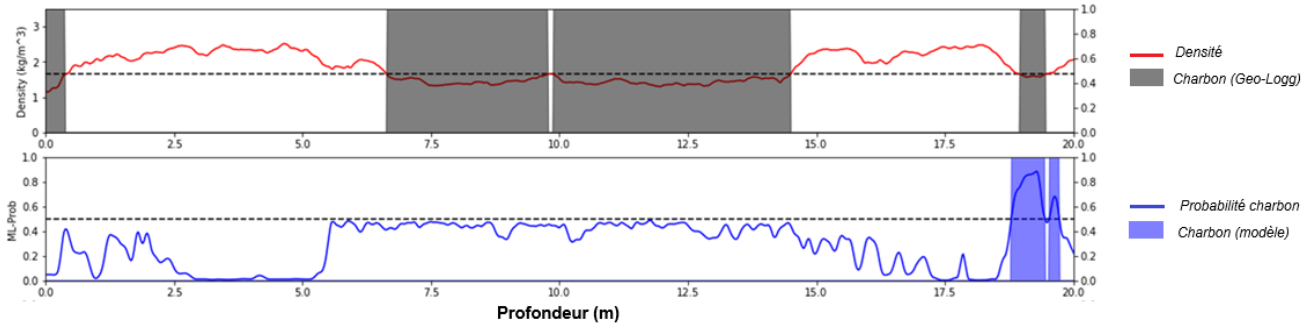


Figure 6.1 Exemple d'une couche de charbon non prédite dans un trou de forage (faux négatif)

Une des limitations majeures de la prédiction locale (dans un seul trou de forage à la fois) est qu'elle n'exploite pas les données MWD et les prédictions faites dans les trous voisins ni les connaissances contextuelles préalablement disponibles. Les prédictions se font par silo pour chaque trou de forage, ne tenant pas compte des résultats trouvés dans les trous précédents voisins. Ceux-ci comportent souvent de l'information utile quand la prédiction pour le trou de forage courant semble incertaine, ou trop proches du seuil de classification binaire (50%).

Les modèles basés sur l'apprentissage non supervisé, notamment le clustering (partitionnement de données) permettent une segmentation du champ en domaines géologiques (*domaining*), qui permettent une interprétation plus contextuelle du type de géologie. L'idée de ce chapitre est qu'en adaptant et en appliquant ces méthodes de partitionnement à un trou de forage et son entourage (si celui-ci propose de meilleures prédictions), il est possible d'améliorer l'interprétation de la prédiction des modèles supervisés.

6.1.2 Objectif spécifique

L'objectif de ce chapitre est de développer une approche basée sur des modèles d'apprentissage automatique non supervisé pour améliorer l'interprétation des prédictions du trou de forage courant en considérant les données et des résultats de prédiction obtenus dans les trous de forage voisins.

Les avantages des approches locales et globales (voisinage) seraient ainsi exploités pour améliorer l'interprétation de la géologie. Alors que les méthodes locales basées sur l'apprentissage supervisé fournissent une précision locale, les approches globales basées sur l'apprentissage non supervisé offrent une vue d'ensemble plus cohérente et peuvent servir à améliorer ou corriger la prédiction locale.

En résumé il s'agit d'améliorer les prédictions locales dans certaines conditions d'incertitudes en les combinant avec une prédiction globale obtenue à l'aide des techniques d'apprentissage non supervisé en utilisant les MWD dans le forage courant et les MWD issus des trous de forages voisins.

6.1.3 Difficulté du problème et défis spécifiques

La Figure 6.2 présente un trou de forage A et ses voisins (trou précédent B et trou suivant C). Dans cet exemple, il est facile d'inférer visuellement, à partir des prédictions des trous voisins B et C, qu'il y a du charbon entre 3 et 4 mètres dans le trou A même si les probabilités de charbon ne dépassent pas le seuil de 50%, ce qui est confirmé par les relevé géophysique du trou A. L'enjeu est de trouver une méthode qui permette d'inférer la présence de cahrbon de manière systématique et automatique quand nécessaire, sachant que les relevés géophysiques ne sont généralement pas disponibles. Il est à noter qu'il est difficile de conclure pour la zone comprise entre 6 à 8 m dans les trous B et A sans les relevés géophysiques.

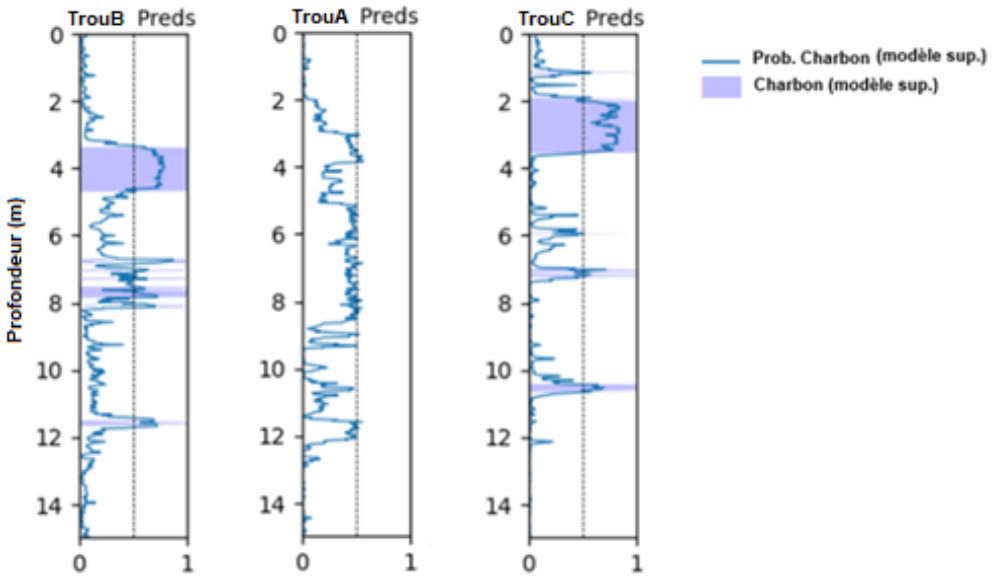


Figure 6.2 Exemple d'un trou de forage A et des trous voisins B et C

Faire simplement la moyenne des valeurs MWD ou des prédictions à travers les trous voisins pour obtenir des données MWD plus « robustes » n'est pas toujours une approche adéquate car les couches ne sont pas nécessairement horizontales (alignées en profondeur). De plus, les trous sont souvent forés selon un motif en serpentin; deux trous peuvent ainsi être spatialement proches mais chronologiquement éloignés et donc forés dans des conditions très différentes (différents opérateurs, tricônes dans un état différent et même foreuses différentes).

Différentes approches d'interpolation spatiale tel que le krigeage (technique d'interpolation spatiale) et la pondération par inverse des distances peuvent être utilisées pour corriger les prédictions locales. Les prédictions les plus fiables peuvent servir de base pour l'interpolation des types de roche aux emplacements spatiaux où les prédictions sont incertaines ou moins fiables. Cependant, cette méthode présente des limitations en raison de la rareté des prédictions faibles dans les trous de forages voisins (parfois, seulement certaines portions des trous peuvent avoir des prédictions jugées fiables). De plus les méthodes comme le krigeage ou les simulations géostatistiques peuvent-être couteuses en temps de calcul et nécessitent parfois la connaissance d'un variogramme approprié qui décrit la covariance spatiale.

Les approches basées sur le clustering (partitionnement) peuvent-être utilisées pour segmenter le domaine formé par les trous en plusieurs zones différentes. La difficulté repose sur le fait de combiner comme attributs des données géospatiales (coordonnées spatiales des points de données dans les forages) avec des propriétés mécaniques de la foreuse (MWD). En effet la plupart des approches de partitionnement reposent sur l'évaluation d'une mesure de similarité, qui peut souvent être par exemple la distance (euclidienne) entre les attributs observés entre deux points. Il est possible de regrouper les segments de forage ayant des propriétés géologiques similaires. La qualité des résultats va beaucoup dépendre du type de modèle utilisé et de la fonction de distance adoptée.

6.2 Approche proposée

L'approche générale proposée dans ce travail est d'appliquer une méthode d'apprentissage machine non supervisé en utilisant comme attributs les données du trou courant et des trous voisins pour améliorer l'interprétation de la présence de charbon dans le trou courant.

6.2.1 Idée générale

Quand un trou présente une classification incertaine provenant du modèle supervisé, il est possible de remplacer ou de corriger cette classification par une classification par partitionnement provenant d'un modèle non supervisé qui prend en compte les données du trou courant et celles des trous voisins. En appliquant une méthode de partitionnement, il est possible de regrouper les segments de forage ayant des propriétés géologiques similaires, facilitant ainsi l'interprétation des données et la classification. Les attributs considérés pour ce modèle non supervisé sont entre autres les données MWD, mais aussi les prédictions (c'est-à-dire la probabilité de présence de charbon dans le trou courant et dans les trous voisins. Différentes méthodes de partitionnement ont été présentées à la section 2.2.4 : méthode des k-moyennes, méthode de mélange de gaussiennes et partitionnement spectral. Le Tableau 4 résume les avantages et les inconvénients associés à chacune de ces méthodes. Ces méthodes seront appliquées au problème courant et leur performance comparées afin de suggérer la meilleure méthode.

Tableau 4 – Avantages et inconvénients des méthodes de partitionnement proposées

Méthode	Avantages	Inconvénients
k-moyennes	▪ Simple et rapide à exécuter ▪ Efficace pour des clusters de forme sphérique ▪ Facile à implémenter	▪ Sensible aux points aberrants ▪ Clusters de forme limitée
Mélange de gaussiennes	▪ Modèle probabiliste donnant une appartenance douce ▪ Flexible dans l'ajustement de densités	▪ Convergence lente pour de grands ensembles ▪ Sensible aux initialisations
Partitionnement spectral	▪ Efficace pour des clusters de formes complexes ▪ Peut modéliser diverses formes de relations globales entre les données	▪ Complexité élevée pour des grands ensembles ▪ La matrice des similarités peut être coûteuse en mémoire

6.2.2 Sélection des trous de forage

Le trou A est sélectionné en se basant sur son incertitude soit parce que les prédictions de charbon pour ce trou semblent très différentes de trous voisins, soit parce que les valeurs de probabilités de

présence de charbon dans ces trous sont trop proches du seuil de décision tel qu'illustré à la Figure 5.1. Le processus présenté va permettre de remplacer et améliorer ces prédictions.

Pour qu'un trou B puisse être considéré, il doit-être situé dans un certain « rayon d'influence » ou portée du trou A où il est raisonnable de supposer que la géologie dans les deux trous sont fortement corrélés. Ces paramètres sont prédéfinis par l'utilisateur selon le contexte. Dans cette recherche, un trou est supposé voisin du trou de forage courant (trou A) s'il précédé ou succède le trou A sur une ligne de patron de forage.

Par souci de simplification, la méthode proposée est discutée en utilisant le cas de deux trous de forage, le trou courant identifié par trou A et un trou voisin précédent identifié par trou B. Il est à noter que la méthode proposée ne limite pas le nombre de trous voisins pouvant-être utilisés. La limite peut se trouver au niveau de la taille du graphe initial. Au besoin, et selon la résolution initiale en profondeur des données (2 cm), une agrégation préalable des données pourrait être effectuée afin de limiter la taille du problème, en effectuant une simple moyenne des variables par segments de profondeur.

6.2.3 Variables d'entrées (attributs) et de sortie

Les variables suivantes (et attributs du modèle non supervisé) sont considérées pour chaque point de données:

- Les variables MWD standardisées $\widehat{RPM}, \widehat{WOB}, \widehat{ROP}, \widehat{TRQ}, \widehat{SED}$ où $\widehat{MWD} = \frac{MWD - \mu_{MWD}}{\sigma_{MWD}}$
- La probabilité de charbon P_c (calculé par l'approche proposée au chapitre 4)
- La profondeur Z (pour certaines des méthodes, elle sera ramenée à l'échelle unitaire [0,1], soit $\hat{Z}$)

Analyse en composantes principales

Les variables MWD ($\widehat{RPM}, \widehat{WOB}, \widehat{ROP}, \widehat{TRQ}$) et la probabilité de charbon sont fortement corrélées, (Zagré, Gamache, Labib, & Shlenchak, 2024). Les méthodes de partitionnement considérées reposent sur le calcul de la similarité entre deux points de données. Afin de mesurer la similarité entre deux points une mesure de distance entre les valeurs MWD de ces deux points sera évaluée. Cette mesure pourrait-être biaisée par la corrélation entre les variables. Afin d'obtenir une meilleure mesure de distance entre les points de données dans l'espace des attributs tout en gardant l'information pertinente véhiculée par les données, les quatre variables MWD seront remplacés par

les trois premières composantes principales fournies par l'ACP (voir la description de cette approche à la section 2.2.5), soient les nouvelles variables *COMP1*, *COMP2* et *COMP3*. Il sera confirmé que la proportion de la variance totale expliquée par ces 3 premières composantes est significative (typiquement autour de 96% de la variabilité pour les données courantes).

Soit M la matrice des variables MWD standardisées : $M = [\widehat{RPM}, \widehat{WOB}, \widehat{ROP}, \widehat{TRQ}, \widehat{SED}, P_c]$

La matrice de covariance symétrique $S_{\overline{MWD}}$ est calculée comme suit : $S_{\overline{MWD}} = \frac{1}{n-1} M^T M$

Les valeurs propres $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et les vecteurs propres $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}$ de la matrice $S_{\overline{MWD}}$ sont calculés. Les composantes principales $COMP_1$, $COMP_2$ et $COMP_3$, sont calculées comme suit: $COMP_i = M\boldsymbol{v_i}$ pour $i = 1, 2, 3$. Ces composantes sont ensuite ramenées à l'échelle unitaire [0,1].

À la fin de cette étape, chaque point de données a maintenant les attributs supplémentaires suivants : $\widehat{COMP}_1, \widehat{COMP}_2, \widehat{COMP}_3$ dont les valeurs varient entre 0 et 1.

Agglomération de clusters pour former deux catégories : charbon/pas charbon

À la fin du processus de partitionnement par l'une des méthodes sélectionnées, chaque point de donnée est assigné à une des k partition. Ces k partitions seront regroupées dans deux classes (charbon/pas charbon) en fonction de la classe dominante préalablement observée dans les points de données de ces partitions.

6.3 Application des modèles de partitionnement

Les sous-sections suivantes décrivent l'application des différentes méthodes de partitionnement proposées. La méthode de partitonnement spectral, étant la moins courante dans la littérature du domaine, elle est décrite de façon plus détaillée et une approche spécifique de modélisation est proposée.

6.3.1 Méthode des k-moyennes

Cette méthode de partitionnement est décrite à la section 2.2.4.1.

Les attributs considérés pour ce modèle sont les composantes principales provenant de l'ACP et la profondeur ramenés à l'échelle [0,1] : $\widehat{COMP}_1, \widehat{COMP}_2, \widehat{COMP}_3, \hat{Z}$

La fonction de distance proposée est donc une combinaison de la variable et de la profondeur. Afin de dissimuler l'influence de la variable $\hat{Z}$, elle sera multipliée par un coefficient α pour augmenter ou diminuer son importance relative par rapport aux variables de l'ACP. La meilleure valeur utilisée pour les tests de ce chapitre est de 5.0.

Ainsi la distance entre deux points de données ou un point de données $\mathbf{x}^t$ et un centroïde $\boldsymbol{\mu}_i$ est de la forme :

$$D(\mathbf{x}^{\boldsymbol{t}}, \boldsymbol{\mu}_i) = \|\mathbf{x}^{\boldsymbol{t}} - \boldsymbol{\mu}_i\| = \sqrt{\sum_{l=1}^{3}(\widehat{COMP}_l[\mathbf{x}^{\boldsymbol{t}}] - \widehat{COMP}_l[\boldsymbol{\mu}_i])^2 + \alpha(\hat{Z}[\mathbf{x}^{\boldsymbol{t}}] - \hat{Z}[\boldsymbol{\mu}_i])^2} \quad (46)$$

Le nombre idéal de clusters k à utiliser est déterminé par essais et erreurs en utilisant le coefficient de silhouette moyen s_{sil} tel que défini à la section 2.2.4.1. Les clusters obtenus par la méthode des *k-moyennes* sont par la suite étiquetés et regroupés selon deux catégories : charbon/pas charbon.

6.3.2 Modèle de mélange de gaussiennes

Cette méthode de partitionnement est formulée mathématiquement et décrite à la section 2.2.4.2.

Pour ce modèle la variable $\hat{Z}$ qui correspond à la profondeur n'est pas considérée dans cette méthode, car on ne peut pas supposer qu'elle suit une distribution gaussienne. Les attributs considérés sont : $\widehat{COMP}_1, \widehat{COMP}_2, \widehat{COMP}_3, \hat{Z}$

Le nombre de clusters k est déterminé en utilisant le coefficient de silhouette moyen s_{sil} pour plusieurs valeurs de k. Les clusters obtenus par cette méthode sont par la suite étiquetés selon deux catégories : charbon/pas charbon.

6.3.3 Partitionnement spectral avec les données MWD et les prédictions

Cette méthode de partitionnement est décrite et formulée mathématiquement à la section 2.2.4.4.

Dans le contexte du problème actuel, les nœuds du graphe représentent des forages ou des segments de forages, et les poids des arêtes peuvent être basés sur des mesures telles que la vitesse de pénétration (ROP), le couple (TRQ), ou la force sur le trépan (WOB). En appliquant le partitionnement spectral, il est possible de regrouper les segments de forage ayant des propriétés géologiques similaires, facilitant ainsi l'interprétation des données et l'optimisation des opérations de forage. La représentation sous forme de graphe permet de modéliser les relations spatiales

souhaitées entre les points de données : par exemple, une arrête peut exister entre deux points de différents forages seulement si ceux-ci ont une proximité spatiale (profondeur similaire).

L'adaptation d'algorithme de partitionnement spectral proposé peut être résumée par les étapes suivantes :

- *Étape 1 :* Construire le graphe à partir des trous de forage sélectionnés
- *Étape 2 :* Faire l'analyse en composantes principales des MWD
- *Étape 3 :* Créer la matrice de similarité
- *Étape 4 :* Effectuer une décomposition spectrale (calculer la matrice Laplacienne ainsi que les vecteurs propres de la matrice Laplacienne)
- *Étape 5 :* Partitionner les données (k premiers vecteurs propres) à l'aide de la méthode des k-moyennes

Les sous-sections suivantes décrivent en détail certaines des étapes de cette méthode.

<u>Étape 1 : Représentation des trous de forage par un graphe non orienté G=(V,E)</u>

Cette section décrit les composantes du graphe $G = (V, E)$ de départ, où V est l'ensemble des sommets et E est l'ensemble des arêtes. La Figure 6.3 illustre le graphe G tel que décrit dans les étapes ci-dessus.

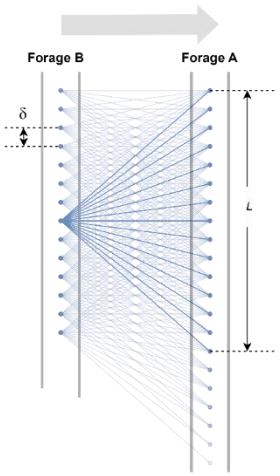


Figure 6.3 Représentation du graphe

Définition de l'ensemble des sommets V

Soit V_A l'ensemble des n_A points de données agrégées du trou de forage A : $V_A = \{\mathbf{x}_1^A, \mathbf{x}_2^A, \dots, \mathbf{x}_{n_A}^A\}$

De même, pour le trou B on a : $V_B = \{\mathbf{x_1^B}, \mathbf{x_2^B}, \dots, \mathbf{x_{n_B}^B}\}$

Dans la représentation sous forme de graphe, l'ensemble des sommets V est l'union de ces deux ensembles (les points de données désormais identifiés par i ou $j = 1, 2, \dots, n$) :

$$V = V_A \cup V_B = \{1, 2, \dots, n\} \text{ où } n = n_A + n_B \quad (47)$$

Définition de l'ensemble des arêtes E

Dans la formulation courante du problème, une arête représente le lien de proximité entre deux sommets. L'approche proposée pour la sélection des arêtes mets l'accent sur la proximité spatiale. Une arête existe entre deux sommets s'il est envisageable que ces deux sommets fassent partie de la même couche ou continuité géologique du fait de leur position spatiale. La corrélation géologique entre deux sommets sera modélisée par la fonction de similarité (poids des arêtes) discutée à la section 6.3.1.

Deux sommets peuvent faire partie du même cluster (partition) même s'il n'y a pas d'arête entre ces sommets. Il suffit qu'il y ait une chaîne (succession d'arêtes) entre ces deux sommets.

Arêtes verticales : pour les arêtes verticales, seules les arêtes entre deux sommets consécutifs en profondeur dans un même trou sont considérées.

Soit E_1 l'ensemble des arêtes verticales entre deux sommets consécutifs dans un même trou de forage. Sachant que les cordonnées X et Y restent constantes le long d'un trou de forage, E_1 peut-être formulé comme suit :

$$E_1 = \{(i,j) |\ i,j \in V, X(i) = X(j)\ \&\ Y(i) = Y(j), Z(j) = Z(i) + \delta\} \quad (48)$$

où δ est la résolution en profondeur (2 cm).

Arêtes entre deux sommets de trous différents : pour les arêtes entre deux sommets de trous différents, il est possible d'ajouter une contrainte de profondeur pour refléter l'idée qu'un point de données à une profondeur h_1 n'est plus corrélé au point de donnée à une profondeur $h_1 + L$, où L représente la portée verticale, soit le seuil de distance verticale à partir duquel deux sommets ne sont plus corrélés (une portée verticale de L=2 m sera considérée dans cette recherche).

L'ensemble E_2 des arêtes entre deux sommets de deux trous différents peut dont être formulé comme suit :

$$E_2 = \left\{(i,j) \middle| \; i,j \in V, X(i) \neq X(j) \,\&Y(i) \neq Y(j), Z(i) - \frac{L}{2} \leq Z(j) \leq Z(i) + \frac{L}{2}\right\} \tag{49}$$

L'ensemble des arêtes E est l'union de ces deux sous-ensembles d'arêtes précédemment :

$$E = E_1 \cup E_2$$

L'ensemble des arêtes peut être représenté par une matrice d'adjacence :

$$A_{ij} = \begin{cases} 1 \;\; s'il\ y\ a\ une\ arete\ entre\ i\ et\ j \\ 0, \;\; sinon \end{cases} \tag{50}$$

Calcul de la matrice de similarité

La matrice de similarité S contenant les poids associés aux arêtes (i,j) est ensuite calculée. Le poids w_{ij} de l'arête (i,j) représente la similitude, ou distance entre deux sommets. Une mesure de distance comprise entre 0 (données très similaires) et 1 (données très différentes) est adoptée. S'il n'existe pas d'arête entre deux sommets :

$$S_{ij} = \begin{cases} w_{ij} \;\; s'il\ y\ a\ une\ arête\ entre\ i\ et\ j\ \left(i.e.si\ A_{ij} = 1\right) \\ 0, \;\; sinon\ \left(si.e.i\ A_{ij} = 0\right) \end{cases} \quad \forall i,j \in V \tag{51}$$

<u>Calcul des poids w_{ij}:</u>

Cas 1 : Sommets avec une probabilité de charbon acceptable

Si les deux sommets i et j ont une probabilité de charbon qui n'est pas entre *40%* et *60%*, ces probabilités sont jugées suffisamment fiables et sont utilisées pour calculer la similarité entre les points. Dans ce cas, le poids sera simplement la différence entre les deux probabilités :

$$w_{ij} = 1 - |P_c(i) - P_c(j)| \tag{52}$$

Cas 2 : Un des sommets a une probabilité de charbon non fiable (entre *40%* et *60%*)

Dans ce cas on utilisera une fonction de distance basée sur les composantes principales (normalisées) provenant de l'ACP des variables MWD telle que décrite à la section 6.2.3 :

$$w_{ij} = \frac{1}{\sqrt{3}} \|i - j\|_{ACP} = \frac{1}{\sqrt{3}} \sqrt{\sum_{t=1}^{3} (\widehat{COMP}_t[i] - \widehat{COMP}_t[j])^2} \tag{53}$$

Où $\|i - j\|_{ACP}$ est la distance entre les points de données i et j dans l'espace des composantes principales.

Les deux fonctions ainsi définies ont le même intervalle d'arrivée (entre 0 à 1). La valeur de $w_{ij} = 1$ quand deux points de données sont proches et $w_{ij} = 0$ ou tendent vers 0 quand deux points sont différents.

Décomposition spectrale

D est la matrice diagonale représentant les degrés de chaque sommet est calculé à l'aide de la matrice d'adjacence A en posant : $D_{ii} = \sum_{j \in V} A_{ij}, \forall i \in V$

Et la matrice Laplacienne normalisée peut ensuite être calculée en utilisant : $L_N = D^{\frac{1}{2}} L D^{-\frac{1}{2}}$,

Les valeurs propres λ et les vecteurs propres associés $\boldsymbol{v}$ de la matrice Laplacienne sont calculées en trouvant les solutions de l'équation l'aide de l'équation : $L\boldsymbol{v} = \lambda \boldsymbol{v}$. Les k premiers vecteurs propres sont ensuite utilisés comme attributs pour partitionner les données en k parties.

Le nombre de partitions k est sélectionné en fonction du nombre de zones ou couches lithologiques distinctes attendues observés dans le trou courant et les trous voisins (soit k =3 dans l'exemple présenté à la section 6.4).

6.4 Exemple de mise en pratique

Afin d'illustrer et comparer leur performance, les différents modèles de partitionnement présentés dans ce chapitre sont appliqués à un cas de forage réel provenant d'une mine de charbon pour laquelle les modèles de classification supervisés du chapitre 4 ont été développés. Pour les besoins de cette recherche il s'agissait de sélectionner un forage pour lequel l'application du modèle supervisé intégré de prédiction de charbon a fourni des résultats incertains et pour lequel il existe un relevé géophysique. Une instance de trou de forage est sélectionnée en se basant sur les critères suivants :

- Le trou présente une prédiction (modèle supervisé intégré) incertaine, où l'interprétation charbon/pas charbon est difficile ou ambiguë.
- Il existe un relevé géologique utilisé comme *ground truth* pour ce trou et qui permettra d'évaluer si les prédictions des modèles non supervisés proposés sont justes.
- Il existe des forages voisins à ce trou avec des prédictions plus fiables, qui ont la capacité d'améliorer les prédictions du trou courant ou d'y apporter plus de contexte.

La présence d'un relevé géologique pour le trou à l'étude permet de calculer le F1-Score qui est une mesure de la qualité de la classification (voir Section 2.2.2.7, Équation $F1 - Score = 2\frac{Precision \times Recall}{Precision + Recall}$ (11). Le F1-Score sera évalué pour chacun des modèles proposés.

6.4.1 Description des données

La Figure 6.4 présente le trou de forage sélectionné, identifié comme Trou A. Près de 31% des prédictions du modèle supervisé pour le forage A présentent des probabilités de charbon comprises entre 40% et 60%. Notamment, entre 3 et 4 m de profondeur, le modèle supervisé ne prédit pas de charbon alors que les probabilités sont proches de 50%. Cependant, les relevé géophysique suggèrent bien la présence de charbon entre 2.8 et 4.1 m, car la densité est en dessous du seuil de 1.65 g/cm^3. De même, entre 5 m et 8.5 m les probabilités sont proches du seuil de 50% et le modèle supervisé ne prédit pas de charbon, ce qui est confirmé cette fois par le relevé géophysique, même si on observe une baisse de densité entre 6 et 8 m, celle-ci demeure au-dessus du seuil de 1.65 g/cm^3. En résumé l'interprétation de la présence de charbon dans ce trou de forage uniquement à partir des probabilités du modèle supervisé est difficile.

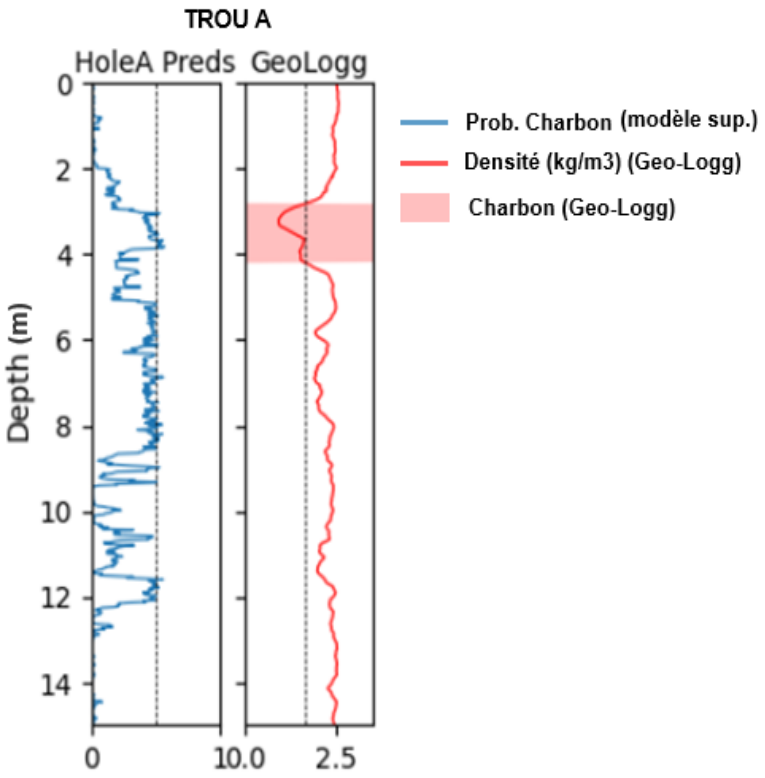


Figure 6.4 Trou de forage A (interprétation difficile de la probabilité de charbon)

La Figure 6.5 présente les prédictions associées aux trous de forage B, A et C. En observant les prédictions dans le trou de forage A, il est assez facile de conclure que les probabilités de charbon entre 3 et 4 m correspondent probablement à du charbon, car on remarque la présence d'une couche de charbon correspondante dans le trou de forage B, prédite aux alentours de 3.2 m à 4.8 m, et aux alentours de 2 à 3.8 m dans le trou de forage C.

Les données MWD associées à ces trous de forages (voir Figure 6.6) permettent de mieux comprendre les probabilités de charbon observées et leur évolution d'un trou à l'autre. Pour le trou voisin B, les RPM et WOB sont maintenus relativement constants (points de consigne), mais les changements dans les variables de réponse (TRQ et ROP) influencent la prédiction de charbon. Cependant, à partir du trou de forage A, le mode d'opération semble avoir changé : de grandes variations dans la variable WOB sont observées, alors que le ROP montre moins de variations pendant de longues portions du forage. Ceci peut être dû à un mode de fonctionnement automatique ou le système essaie de maintenir le ROP. Ces changements dans les conditions de forage peuvent affecter les modèles de prédiction supervisé et engendrer des prédictions moins fiables.

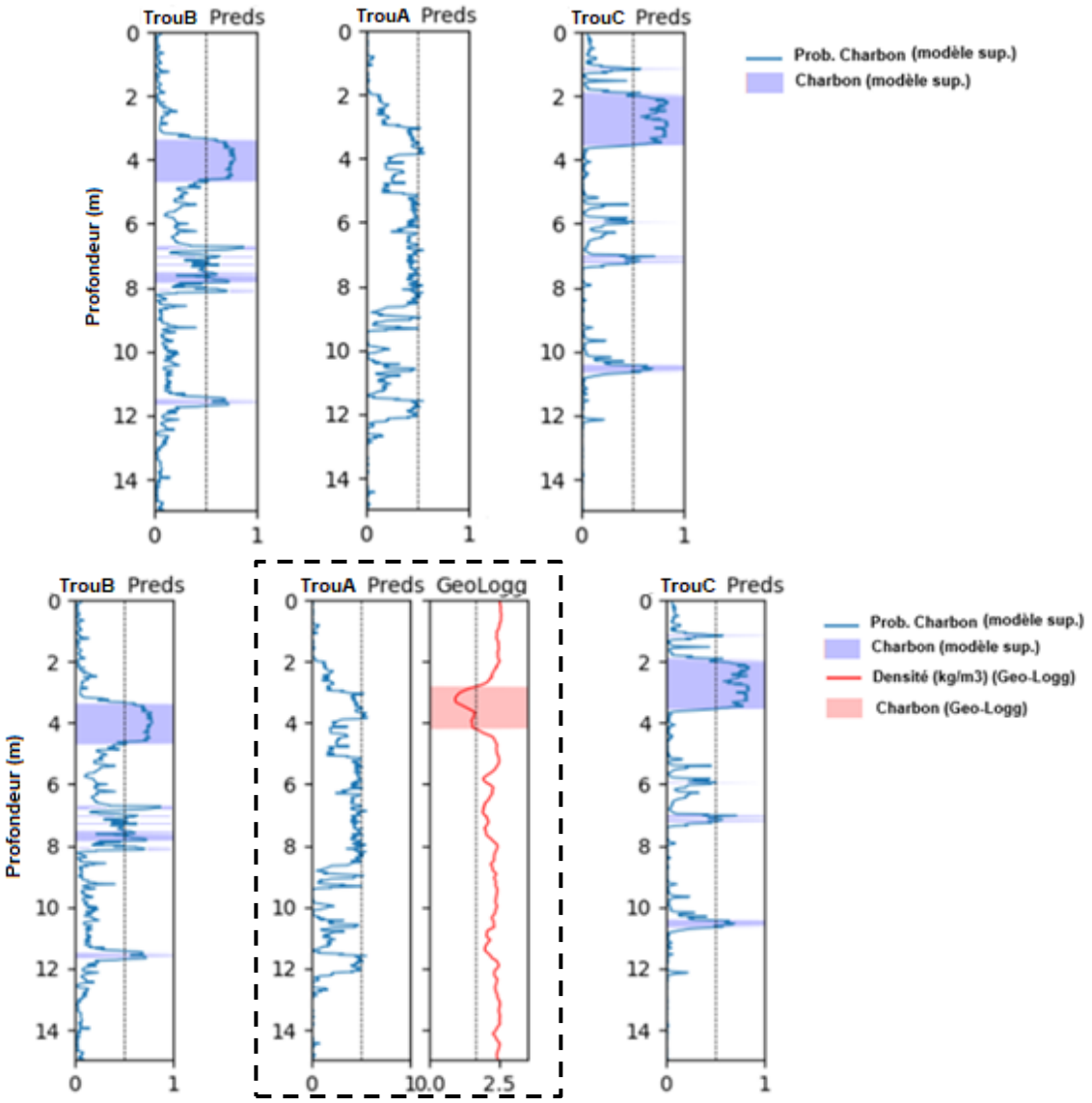


Figure 6.5 Trou de forage A avec les trous voisins B et C

En utilisant un modèle de classification non supervisé qui considère toutes ces données (prédictions de charbon et valeurs MWD dans le trou courant et dans les trous voisins), il est possible de « connecter » les zones similaires dans les différents trous et délimiter localement des domaines géologiques continus dans les trois trous de forage. Ceci permettra une meilleure interprétation de la présence de charbon dans le trou courant A.

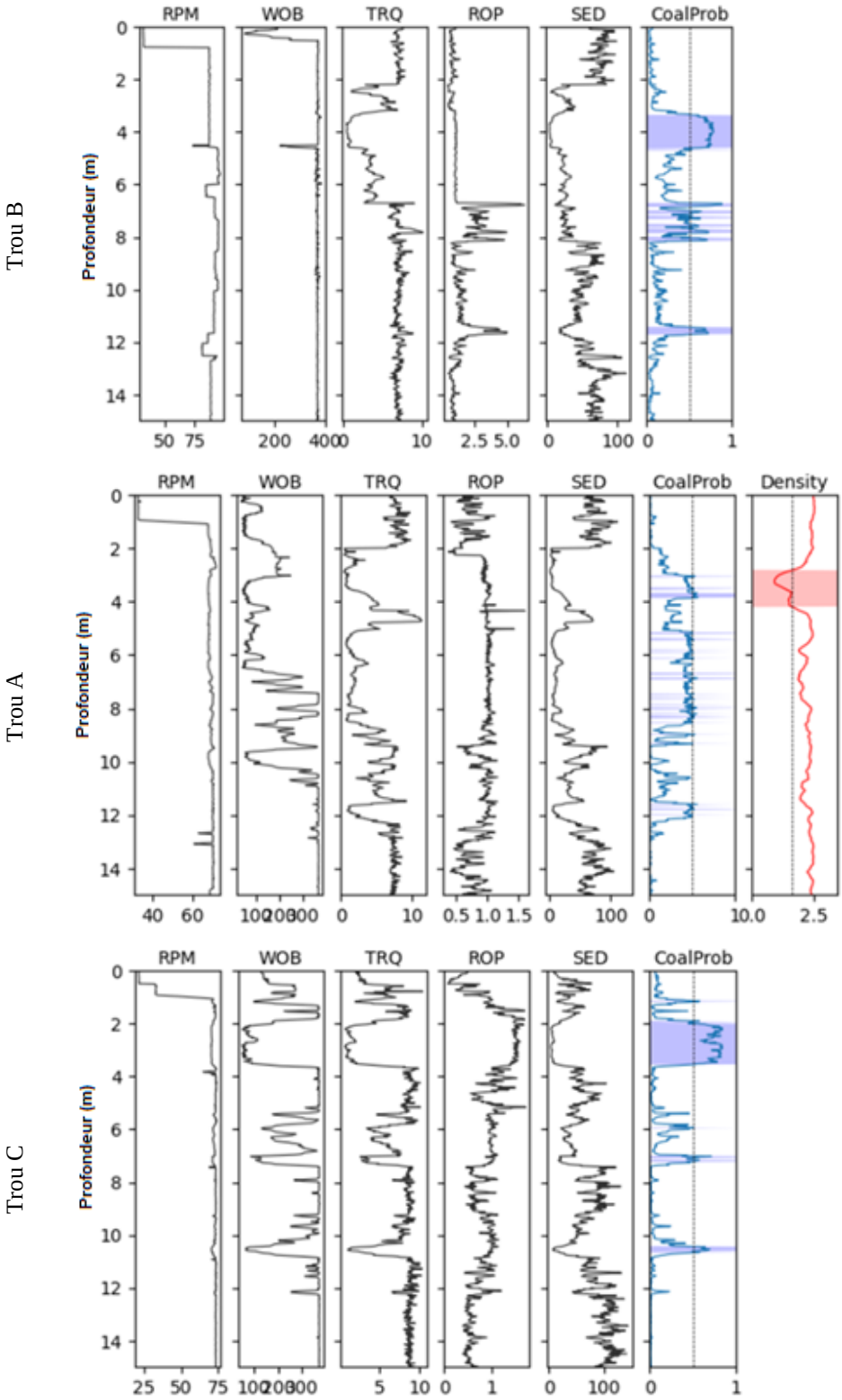


Figure 6.6 MWD associés aux trous de forage B, A et C

6.4.2 Principales hypothèses

La principale hypothèse qui est émise est que le relevé géologique est exact et représentatif de la réalité. De manière simplifiée, comme dans le chapitre 4, la densité est donc utilisée comme critère pour établir la présence de charbon (en dessous du seuil de 1.65 g/cm^3). Pour l'exemple à l'étude, cela implique qu'il existe bien une couche de charbon dont la présence peut être à une profondeur approximative de 3 à 4 m dans le trou A.

Il est également admis que les données MWD et le relevé géologique sont bien alignés en profondeur pour le trou A, soit que l'incertitude d'alignement (en profondeur) entre les deux sources de données est négligeable.

6.4.3 Analyse des résultats

6.4.3.1 Modèle des k-Moyennes

La Figure 6.7 présente la classification finale obtenue avec la méthode des k-moyennes telle que décrite à la section (meilleur modèle avec $k = 6$ et un poids $\alpha = 5.0$ pour la variable $\hat{Z}$). Le modèle prédit une couche de charbon entre 3 et 5 m dans le trou A et cette couche est également présente dans les trous B et C. Cependant cette couche de charbon est plus élargie dans le trou B (2.8 à 5 m) par rapport à la prédiction originale (3.2 à 4.8 m).

Ce modèle prédit également une couche de charbon aux alentours de 6 à 8 m dans les trois trous de forage, certainement influencé par les probabilités proches de 50% dans le trou A et les fluctuations de probabilité dans cette zone dans le trou B. L'existence de cette deuxième couche n'est toutefois pas corroborée par le relevé géophysique (seulement une légère baisse de la densité est observée entre 6 et 8 m).

Ce modèle a un F1-Score de 42.1% pour la prédiction dans le trou courant A (en considérant le relevé géophysique comme *ground truth*). Le modèle a tendance à surestimer la présence de charbon. La faiblesse de ce modèle réside dans la nécessité de combiner des variables de différentes natures (MWD transformés et profondeur) dans une fonction de distance et également le fait que le type de partitions cibles (couches de charbon et pas charbon) ne sont pas nécessairement de forme sphérique dans l'espace des attributs (particulièrement si la profondeur est un attribut) et ont des tailles très différentes du fait du débalancement des classes (généralement lorsque moins de 10% d'un trou correspond à du charbon).

Modèle Supervisé (original)

Modèle des k-Moyennes

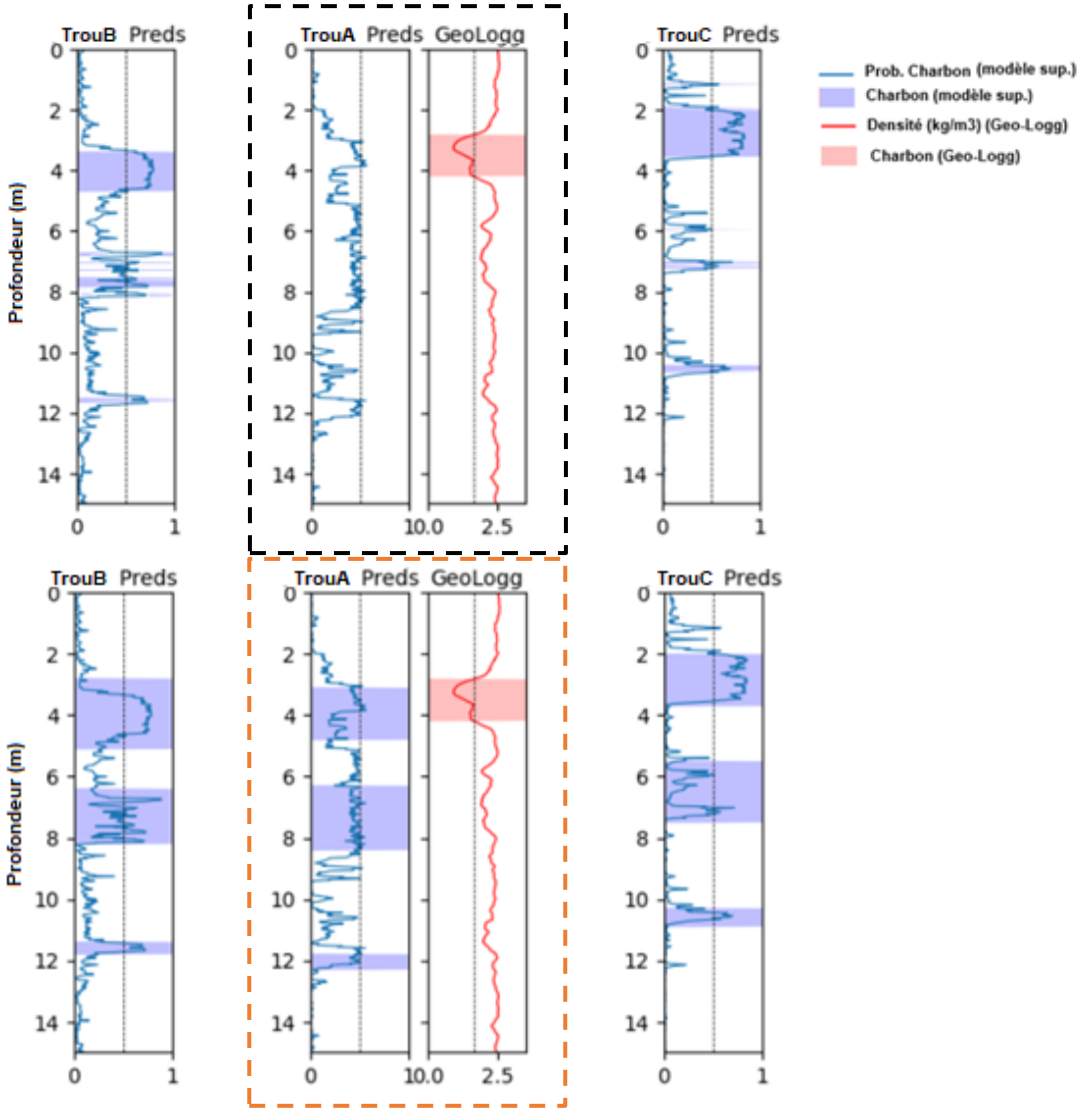


Figure 6.7 Classification finale obtenue pour le modèle des k-moyennes

6.4.3.2 Modèle de mélange de gaussiennes

La Figure 6.8 présente la classification finale obtenue avec le modèle de mélange de gaussiennes (meilleur modèle obtenu avec $k = 4$). Le modèle prédit une large couche de charbon dans le trou B (2 à 8 m) qui se prolonge en rétrécissant dans le trou A (3 à 7.9 m) et dans le trou C (2 à 3.8 m).

Ce modèle a un F1-Score de 35.5% pour la prédiction dans le trou courant A (en considérant le relevé géophysique comme *ground truth*). Ce modèle a tendance à surestimer la présence de charbon, mais aussi à agglomérer les zones potentielles de charbon (en comparaison avec le modèle des k-Moyennes).

6.4.3.3 Méthode de partitionnement spectral

La Figure 6.9 présente la classification finale obtenue avec la méthode de *partitionnement spectral*. Le modèle prédit une couche de charbon entre 3 et 4.5 m dans le trou A, qui correspond à la couche de charbon observée dans le relevé géophysique avec un léger décalage (2.8 à 4.1 m dans le relevé géophysique). La couche de charbon dans le trou B est légèrement élargie (de 2.9 à 5m) comme observé dans le modèle des k-Moyennes. La couche dans le trou C n'est pas significativement changée.

Ce modèle a un F1-Score de 81.1% pour la prédiction dans le trou courant A (en considérant le relevé géophysique comme *ground truth*), ce qui est une performance acceptable pour ce problème. Le léger décalage observé entre les prédictions de ce modèle pour le trou A et le relevé géophysique est probablement attribuable à un défaut d'alignement ou décalage d'alignement en profondeur entre les données MWD et le relevé géophysique. Cette incertitude d'alignement est assez courante et peut-être compensée, tel que discuté à la Section 5.5.2. Une correction de ce décalage améliorerait davantage le F1-Score de ce modèle.

Le Tableau 5 résume la performance (F1-Score) des différentes méthodes proposées. Le modèle de Partitionnement spectral est significativement meilleur pour cette application que les deux autres méthodes envisagées. Cela est dû à la flexibilité au niveau de la modélisation des relations (possibilité de tenir compte de la distance spatiale entre les données) sans nécessairement les inclure dans une même fonction de distance que les données MWD.

Modèle Supervisé (original)

Modèle de mélange de gaussiennes

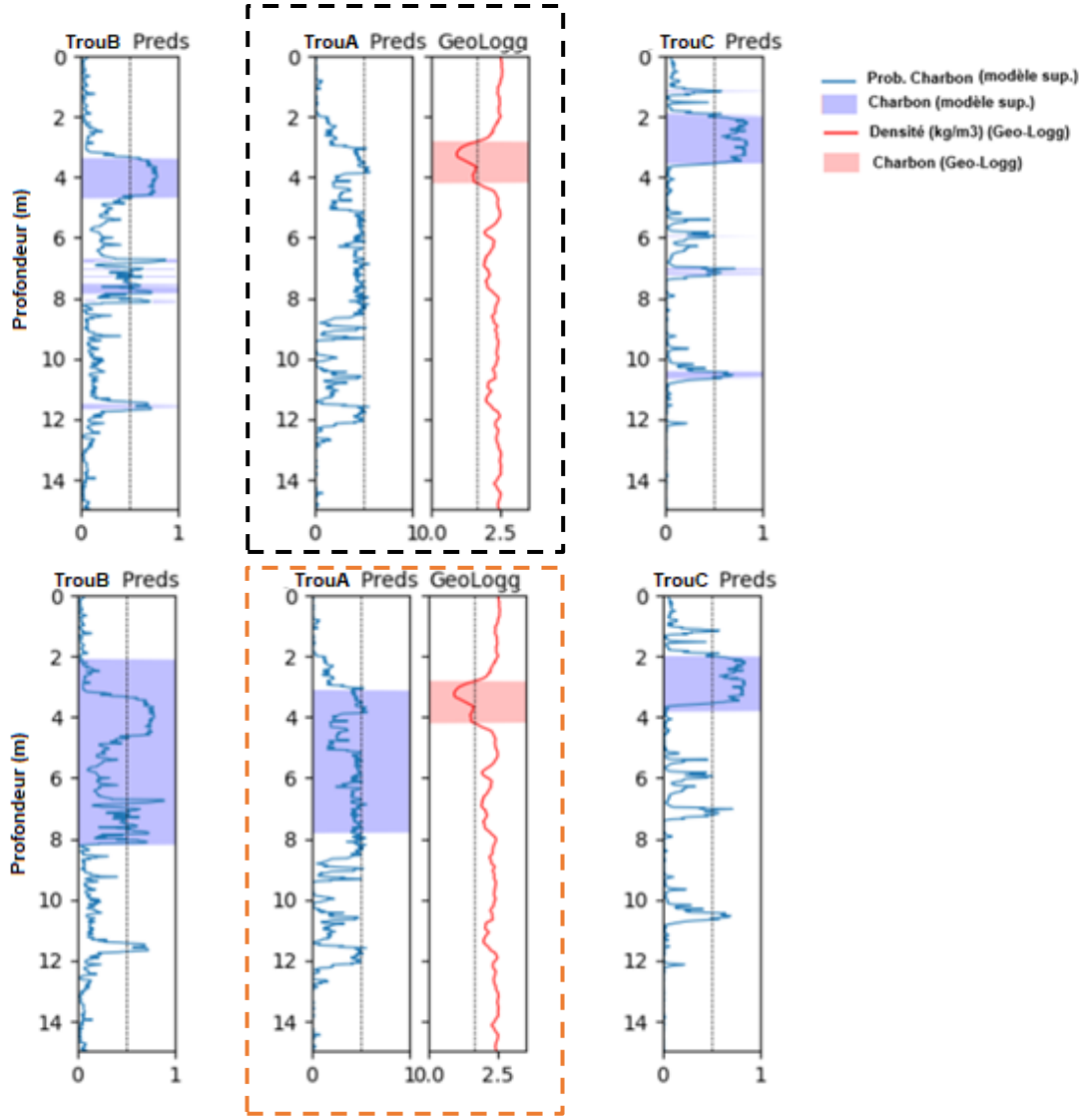


Figure 6.8 Classification finale obtenue pour le modèle de mélange de gaussiennes

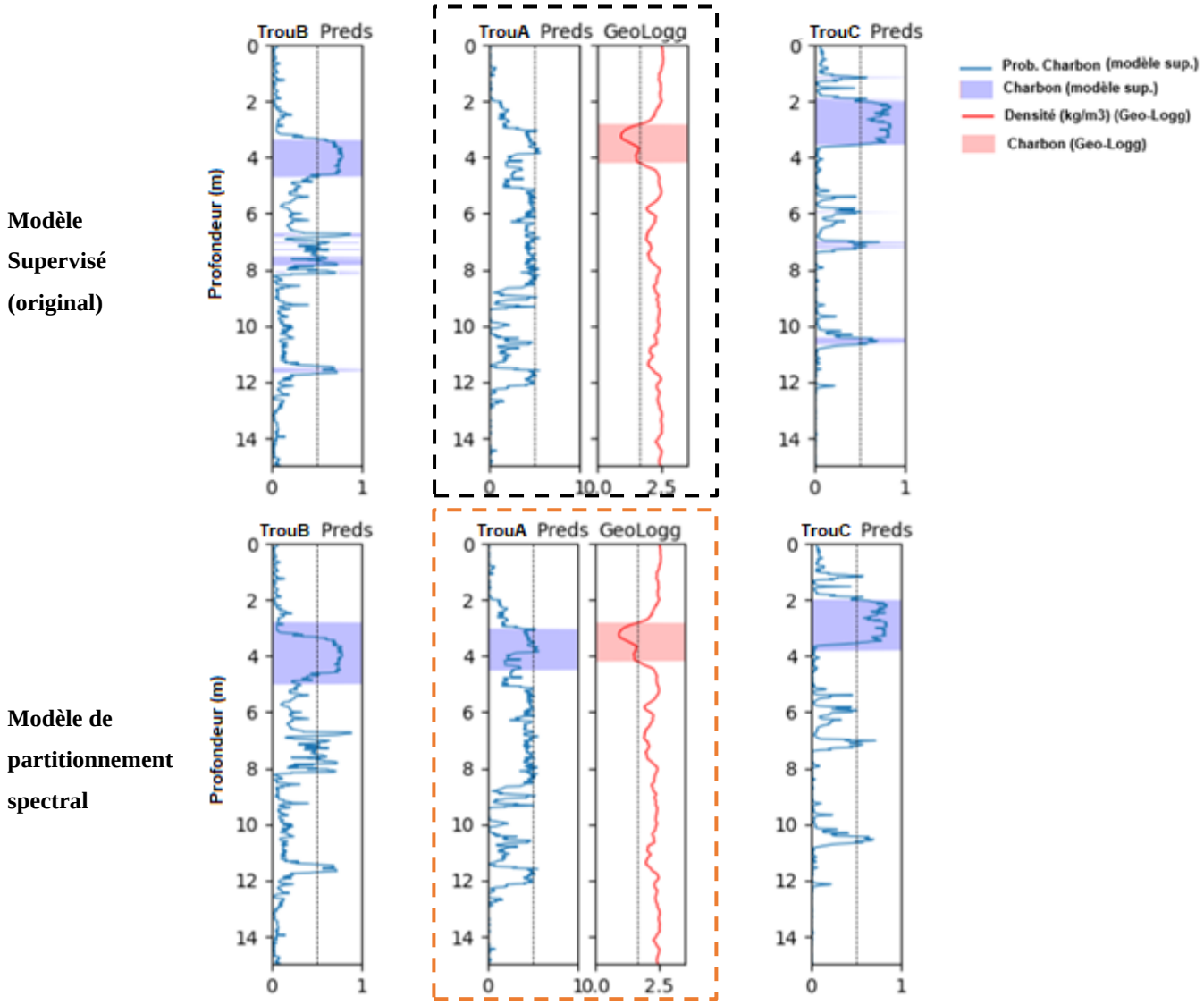


Figure 6.9 Classification finale obtenue pour le modèle de Partitionnement spectral

Tableau 5 – Performance des modèles de classification non supervisée proposés

Modèle	***k-Moyennes***	***Mélange de gaussiennes***	***Partitionnement spectral***
F1-Score	42.1%	35.5%	81.1%

6.5 Conclusion et recommandations

Dans ce chapitre, trois modèles non supervisés de partitionnement ont été adaptés et comparées pour une classification plus globale (forage courant et voisinage) du type de roche à partir des

MWD. Chacune de ces méthodes a des avantages et des inconvénients, toutefois après comparaison, la méthode du *partitionnement spectral* est l'approche recommandée pour intégrer l'information (MWD et probabilités de charbon) des trous voisins pour améliorer l'interprétation dans le trou courant.

Le partitionnement spectral offre une méthode puissante pour le clustering en exploitant les propriétés spectrales d'un graphe. Il est particulièrement utile pour les données complexes et non linéaires, comme celles issues du forage, où la simple application de techniques de clustering comme les k-moyennes serait insuffisante pour capturer la structure sous-jacente des données. Il faut noter toutefois que l'algorithme de *clustering spectral* nécessiterait une puissance de calcul importante pour une utilisation en temps réel.

Les conclusions de ce chapitre auraient bénéficier d'exemples supplémentaires pour confirmer et comparer ses modèles. Il a été difficile de trouver un trou de forage pour lequel les prédictions du modèle supervisé remplissent les critères discutés à la section 6.4 et pour lequel il existe un relevé géophysique.

CHAPITRE 7 DÉTECTION DE L'USURE DU TRICÔNE EN UTILISANT LES MWD PRIMAIRES

7.1 Introduction

7.1.1 Contexte et enjeux

Dans l'industrie minière, l'efficacité des opérations de forage est cruciale pour optimiser les coûts et maximiser la production. Les trépans de type tricônes, largement utilisés dans l'industrie minière subissent une usure progressive au cours des opérations de forage, ce qui affecte leur performance, la productivité et la qualité des forages.

Jusqu'ici l'effet de l'usure du tricône a été ignoré dans les différents modèles proposés pour identifier le type de roche à partir des MWD. Toutefois l'état du tricône a un impact sur la fiabilité de ces mesures et par conséquent sur la fiabilité des modèles de classification. Par exemple, un trépan usé nécessite plus de poussée (WOB) et de rotation (RPM) pour forer le même type de roche qu'un tricône neuf. De plus, un trépan usé peut entraîner des risques de défaillance qui peuvent non seulement endommager les équipements, mais aussi poser des risques de sécurité pour les opérateurs. Un tricône brisé lors du forage nécessite l'arrêt du forage afin de le rechercher et de retirer ses morceaux pour que ceux-ci ne se retrouvent dans les étapes subséquentes du processus comme le moulin et le concasseur ce qui pourrait endommager ces équipements et provoquer des arrêts coûteux pour la mine.

La connaissance précise de l'état d'usure des tricônes en temps réel permet d'optimiser les opérations de forage, de minimiser les temps d'arrêt pour les remplacements et d'éventuellement de maximiser la durée de vie des trépans. Cela permettrait également d'ajuster ou corriger certains des paramètres MWD, notamment les indicateurs d'énergie du forage. À terme, cela permettra une meilleure interprétation de la géologie et bénéficierait aux opérations de dynamitage.

La capacité de prédire l'usure en temps réel des tricônes permet de prendre des décisions proactives pour prévenir les incidents (changement du tricône).

7.1.2 Objectif spécifique

L'objectif de ce chapitre est de développer un modèle prédictif indiquant l'état d'usure des tricônes basés sur les données MWD et prédire automatiquement le moment où il faut procéder à leur retrait. Ces mesures étant collectées en temps réel, elles fournissent des informations précieuses sur les conditions de forage et offrent une opportunité d'évaluer l'état des trépans en temps réel. L'idée sous-jacente est qu'en intégrant les techniques d'apprentissage automatique appropriés, il est possible de développer des systèmes robustes et efficaces pour la prédiction de l'usure des trépans.

7.1.3 Difficulté du problème et défis spécifiques

La prédiction de l'usure des trépans à l'aide des données MWD et de l'apprentissage automatique présente plusieurs défis spécifiques :

Usure finale des tricônes et différents modes de défaillance

Les trépans ne sont généralement pas utilisés jusqu'à la fin de leur vie utile pour éviter les risques de bris et aussi pour maintenir un certain niveau de productivité. Il n'y a donc pas de certitude qu'un tricône qui a été changé l'a été en raison d'une usure excessive du tricône.

Par ailleurs, il existe différents modes de défaillance pour les tricônes tels que discuté à la section 2.1.4. L'usure des dents en carbure de tungstène est un mode de défaillance parmi d'autres (bris ou perte de dents, problèmes de roulements, etc.).

Ainsi dans le cadre de cette recherche, on définit l'usure comme étant une perte d'efficacité du tricône quel qu'en soit la modalité. Les approches étudiées se concentreront sur l'isolation et l'évaluation des changements dans la performance du tricône, sans se préoccuper du type de défaillance réel associé.

Variabilité des conditions de forage

Les conditions de forage peuvent varier considérablement d'un forage l'autre, même sur le même patron. Les variations géologiques, les différentes configurations des foreuses et les comportements des opérateurs introduisent une complexité supplémentaire dans l'analyse des MWD et il est difficile de dissocier leurs effets de celui de l'état du tricône.

Incertitude des données MWD

L’incertitude associée aux mesures MWD complique le développement de modèles fiables et l’interprétation des données. Au cours de la durée de vie d’un tricône, il arrive parfois que les mesures relevées sont erronées. Ainsi, il a été noté parfois des niveaux de vitesse (ROP) 10 à 20 fois plus élevé que la normale. Ces données doivent être identifiées comme des données aberrantes (*outliers*).

Usure non linéaire

L'usure des trépans ne suit pas nécessairement un modèle linéaire simple. Les modèles d'apprentissage automatique doivent être capables de capturer les relations complexes et non linéaires entre les données MWD et l'état d'usure des trépans.

Données d'entraînement limitées

La formation de modèles précis nécessite des données d'entraînement de haute qualité. Cependant, les étiquettes d'usure sont souvent limitées ou inexistantes. Les tricônes se sont généralement inspectés et annotés manuellement qu’après avoir été remplacés. Il est donc impossible de connaître l’évolution de leur usure ou de leur état avec une granularité intéressante. Cela pose un défi pour l'entraînement des modèles supervisés.

Adaptation en temps réel

Pour être utiles, les prédictions d'usure doivent être fournies en temps réel. Les algorithmes doivent donc être efficaces en termes de calcul pour fonctionner sur des données en continu et fournir des prédictions quasi-instantanées.

Plusieurs méthodes proposées dans la littérature nécessitent l’ajout de capteurs supplémentaires comme les capteurs de vibrations. En effet, tel que montré par Rafezi et Hassani (2023), le niveau de vibration à certaines fréquences peut-être un bon indicateur de l’état du tricône. Un défi supplémentaire posé dans cette recherche est de parvenir à un modèle acceptable en se basant uniquement sur les MWD traditionnels (ROP, WOB, TRQ, ROP et SED).

7.2 Description des différentes approches étudiées

Plusieurs méthodes pour détecter l'état d'usure ou d'efficacité des tricônes sont abordés dans ce chapitre :

- Méthode basée sur l'énergie de forage cumulée : Cette méthode consiste à évaluer l'énergie totale de forage cumulée pour un tricône. L'idée sous-jacente est que cette énergie peut-être un bon indicateur de l'état du tricône et du seuil à partir duquel il devrait être remplacé.
- Méthode basée sur l'évolution de la performance de forage dans les roches dures : Au fur et à mesure que le forage se poursuit avec un même tricône, l'efficacité s'amenuise. Cette efficacité est particulièrement plus facile à détecter dans la roche plus dure. Un indicateur basé sur l'évolution des percentiles du SED est proposé pour refléter l'état du tricône.
- Méthode basée sur un réseau de neurone récurrent (LSTM) : Un modèle de régression basé sur une architecture LSTM est proposé pour prédire non pas directement l'état usure, mais l'évolution du taux d'usure ou dégradation en fonction du temps.

Ces méthodes sont comparées en évaluant le niveau de variabilité lors du retrait des différent tricônes à l'étude.

7.3 Description et exploration sommaire des données

Pour ces travaux de recherche, les données provenant d'une minière canadienne ont été collectées. Il s'agit de données MWD (RPM, WOB, TRQ, ROP, SED) historiques collectées durant toute la vie utile (installation jusqu'au retrait) de 19 tricônes utilisés sur une même foreuse. Les données sont fournies ont une résolution en profondeur de 1 cm (0.01 m).

La Figure 7.1 présente un exemple de données MWD collectées pour un tricône utilisé pour forer une distance totale d'environ 500 m. Les chutes de RPM et de WOB représentent généralement les changements de trous de forage.

Les rapports d'inspection postérieur des tricônes n'étant disponibles que pour une minorité de ces tricônes (six tricônes), cette information n'a pas pu être utilisée dans le cadre de cette recherche.

7.3.1 Hypothèses principales

Les principales hypothèses formulées pour cette recherche sont :

- Les MWD relevés son exactes et on ignore l'incertitude associée à leur mesure.
- Les tricônes sont de même type : À défaut d'informations supplémentaires, il est considéré que la majorité des tricônes considérés pour cette étude sont du même type.
- Les tricônes ont été retirés au moment où ils ont atteint leur maximum d'utilité c'est-à-dire qu'il existe un critère commun plus ou moins représentatif du retrait de la majorité de ces tricônes. Ce retrait n'est donc pas fait de façon entièrement subjective.

Pour renforcer la validité de ces deux dernières hypothèses, quelques tricônes qui ont eu une durée de vie (distance totale de forage) trop élevée ou trop faible par rapport à la norme ne sont pas considérés. Selon les opérateurs, la distance de forage totale pour un tricône sur ce site minier varie généralement autour de 500 à 600 m. Les tricônes qui ont une durée de vie inférieure à 350 m ne sont donc pas considérés dans cette recherche, car la raison du retrait peut être due à une défaillance anormale ou soudaine du tricône et ces tricônes ne seraient donc pas représentatifs d'un cycle de vie normal de tricône.

De même, les tricônes ayant une durée de vie supérieure à 800 m sont également déconsidérées, car une durée de vie inhabituellement longue suggère un tricône d'un type différent et donc non représentatif. En appliquant ces deux critères, cinq tricônes ayant des durées de vies totales de 190 m, 264 m, 345 m, 1209 et 1494 m ont été abandonnés. Ceci ramène le nombre de tricônes pour cette étude à 14.

7.3.2 Exploration des données

7.3.2.1 Distance de forage totale et comparaison des critères de retrait.

Malgré l'application des critères sur la distance finale au moment du retrait des tricônes retenus varie grandement (entre 374 et 780 m). Le Tableau 6 présente la distance totale de forage correspondant à chacun des 14 tricônes considérés. La moyenne est de 680 m.

Un coefficient de variation normalisé (CV_X^{Norm} dans la formule qui suit) sera utilisé afin de comparer différents critères de retrait sur une base commune. Ce coefficient est calculé comme suit pour un critère ou indicateur X normalisé (ramené à une échelle de 0 à 1) :

$$CV_X^{Norm} = \sigma_{X_{NORM}} \times \frac{1}{\mu_{X_{NORM}}} \quad \text{où} \quad X_{NORM} = \frac{X - Min(X)}{Max(X) - Min(X)} \tag{54}$$

Plus ce coefficient de variation normalisé est élevé, moins ce critère est fiable pour représenter le moment de retrait observé dans la population des tricônes considérés. Idéalement on veut trouver un critère pour lequel la variabilité serait presque nulle et cohérente d'un tricône à l'autre.

À titre d'exemple, le coefficient de variation normalisé pour le critère de la distance totale forée est de 82,9%. En d'autres termes, si la décision de retrait des tricônes était basée sur la distance totale, alors le niveau de variabilité associé à ce critère serait de 82.9% en se basant sur la population de tricônes à l'étude.

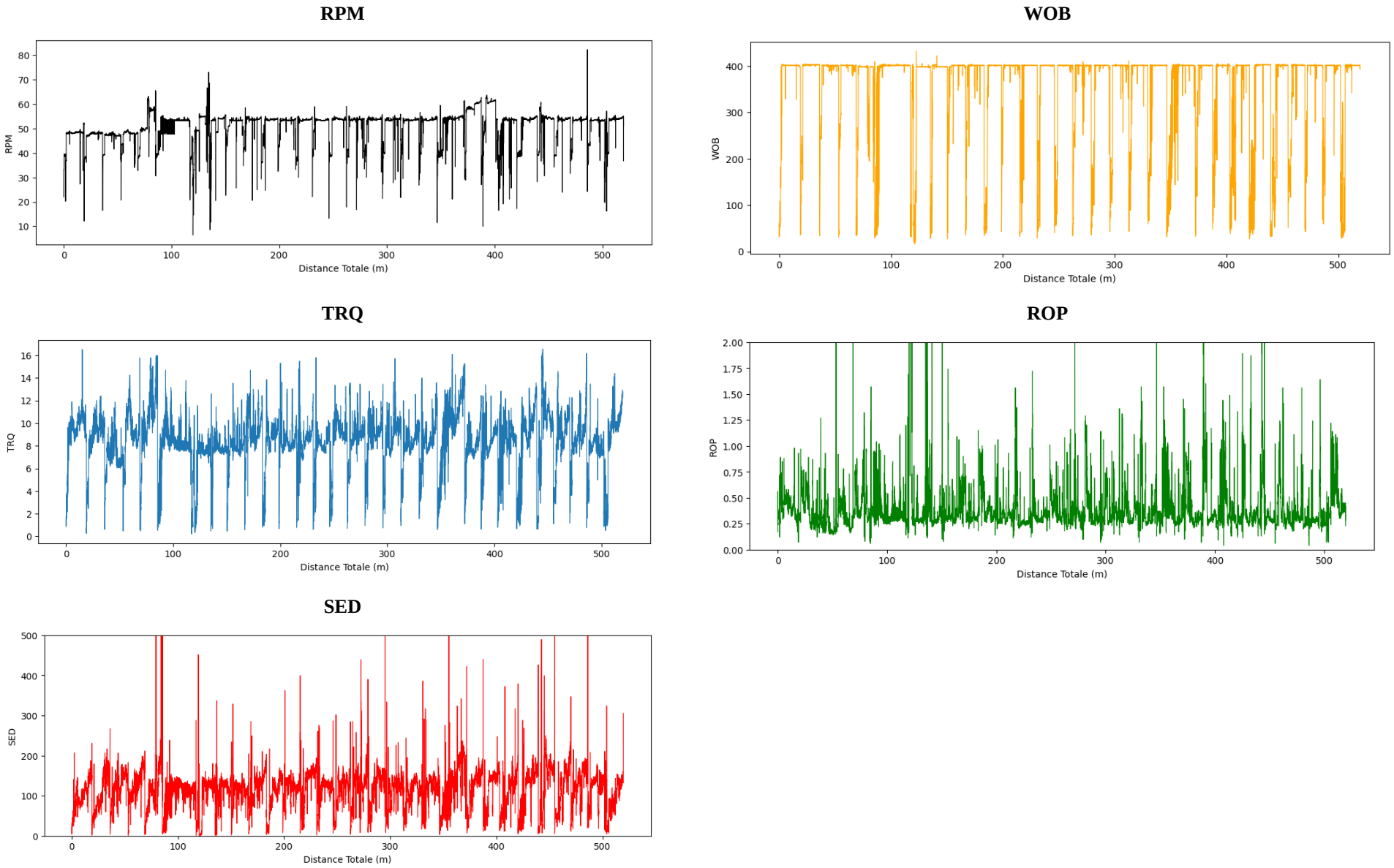


Figure 7.1 Exemple de données MWD pour un tricône

7.3.2.2 Évolution de la distribution des MWD

Afin de vérifier s'il existe des différences significatives dans les MWD entre un tricône à l'état neuf et un tricône à l'état usé proche de sa fin de vie utile, les données ont été regroupées et subdivisées comme suit :

- Tricône **neuf** : les données correspondant aux premiers 15% de la distance totale forée pour chaque tricône sont regroupées dans cette catégorie.
- Tricône en **milieu de vie** : les données correspondant à 40 à 60% de la distance totale forée pour chaque tricône sont regroupées dans cette catégorie.
- Tricône en **fin de vie** : les données correspondant aux derniers 15% de la distance totale forée pour chaque tricône sont regroupées dans cette catégorie.

En comparant les données de ces différentes catégories, il est possible d'identifier de façon macroscopique des tendances sur l'évolution des MWD entre les différents états des tricônes. Ces regroupements permettent aussi d'éliminer ou d'atténuer l'effet des autres variables incontrôlées telles que la géologie et les conditions d'opérations (manuelle, automatique, etc.).

La Figure 7.2 présente les diagrammes box-plot des variables MWD pour les trois catégories d'état de tricône identifiés ci-dessus. Cette figure permet de tirer les observations suivantes.

Les variables RPM et WOB ne changent pas significativement d'un état à l'autre. En effet les médianes de ces deux variables restent à peu près constantes, de même que les 3e quartile (75e percentile). Ceci s'explique par le fait que ces variables sont habituellement des paramètres d'opération de la foreuse entrés par consignes, contrairement au ROP et au TRQ qui sont des réponses qui varient en fonction de la géologie. L'écart interquartile est légèrement plus élevé pour le WOB quand le tricône est à l'état usé. Cela peut s'expliquer par le fait que l'opérateur essaie de varier la consigne du WOB pour maintenir une vitesse de forage (ROP) acceptable.

La distribution de la variable TRQ change avec le niveau d'usure : la médiane ainsi que le 1er et 3e quartiles baissent avec le niveau d'usure ou de dégradation du tricône. Cela peut s'expliquer par le fait que plus le tricône est usé, moins ses dents s'insèrent dans la roche, créant moins de friction et de résistance à la rotation appliquée.

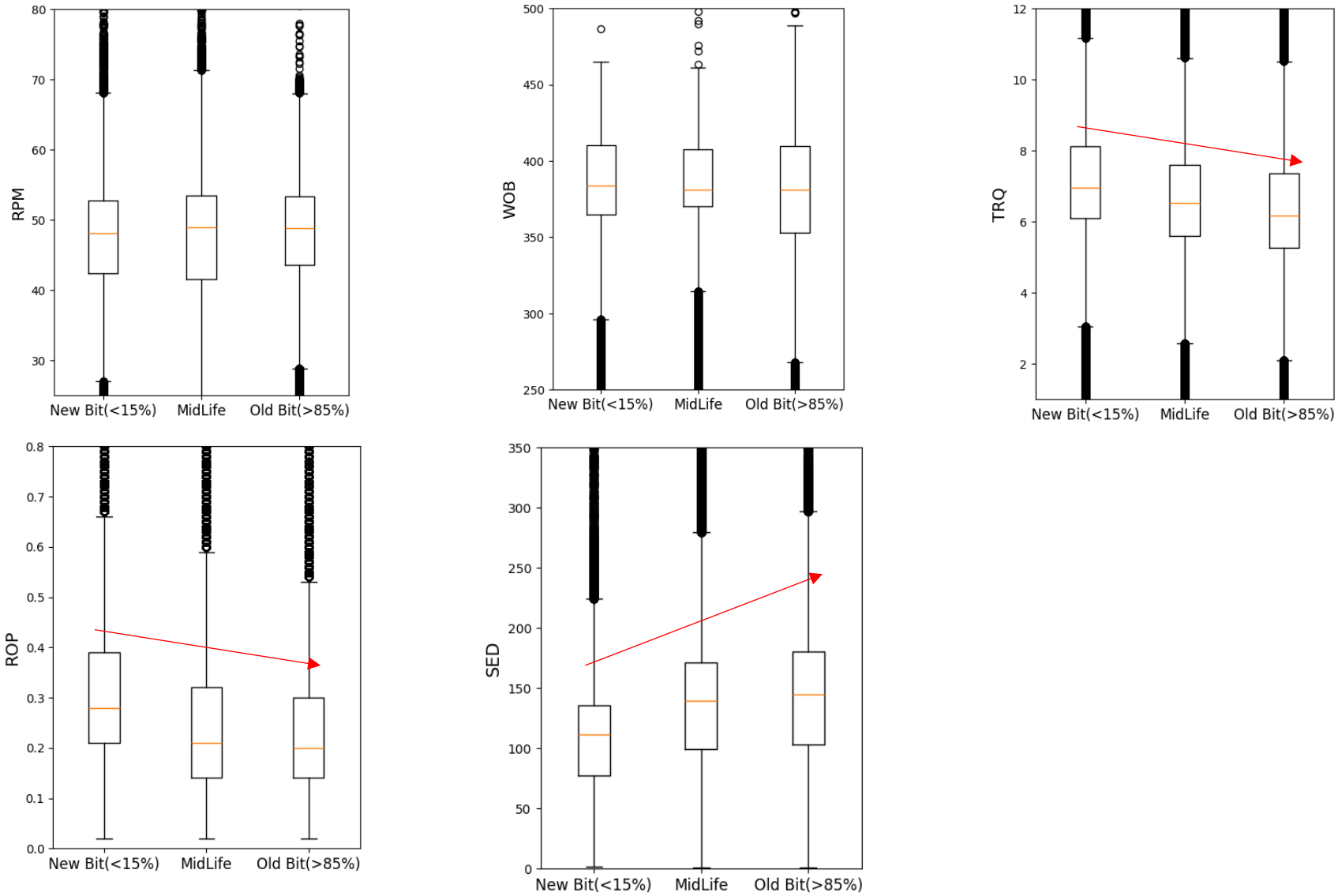


Figure 7.2 Box-plot des données MWD pour différentes catégories d'état des tricônes

À l'instar du TRQ, la distribution du ROP change avec l'état du tricône. La médiane ainsi que le 1er et 3e quartile ont une tendance baissière. Cet effet est surtout accentué au niveau du 3eme quartile. Ceci s'explique par le fait qu'au fur et à mesure que le trépan se dégrade la performance baisse, pour le même niveau d'énergie appliquée on obtient moins de vitesse pénétration.

Cette tendance baissière est également observable au niveau du SED, particulièrement dans les percentiles plus élevés. Le SED est habituellement interprété comme un indicateur de la dureté de la roche (énergie nécessaire pour briser la roche), cela confirme qu'il faut de plus en plus d'énergie pour forer au fur et à mesure que le tricône s'use.

7.4 Approche basée sur l'énergie de forage cumulée

La première méthode envisagée est celle de l'énergie de forage cumulée au cours de l'utilisation du tricône. Intuitivement, il semble raisonnable de supposer que plus un tricône va accumuler de l'énergie de forage, plus le tricône va accumuler de l'usure et sa performance se dégrader. Il est à noter que l'appellation énergie spécifique de forage est un abus de langage, car le paramètre SED est en réalité une mesure de pression (voir équation 33). Toutefois cette approximation est acceptable ici vu que la surface (aire du trou) est constante et l'écart de profondeur entre les mesures consécutives de SED est constant (0.01 m).

Soit x la distance cumulée à chaque mesure des MWD ($\mathrm{x} = 0.1, 0.2, \ldots, \mathrm{Distance_Totale}$). $SED(\mathrm{x})$ donne l'énergie spécifique de forage calculée en fonction de la distance totale cumulée.

$$SED_{Cumule}(\mathrm{x}) = \sum_{t=0}^{\mathrm{x}} SED(\mathrm{t}) \tag{55}$$

La Figure 7.3a montre un exemple de la valeur du SED cumulé pour un tricône. La valeur finale au moment du retrait du tricône est autour de $5\mathrm{x}10^6$.

Le Tableau 6 donne les valeurs finales de SED cumulé pour chaque tricône. En considérant le SED cumulé comme critère de retrait des tricônes, le coefficient de variation normalisé serait de 102.4%, ce qui est pire que celui de la distance totale (82.9%). La variabilité observée en fin de vie pour cet indicateur est trop grande. Une hypothèse qui expliquerait l'inefficacité de ce critère est qu'il ne serait pas toujours représentatif à cause de la déviation dans la distribution du SED tel qu'observé dans la Figure 7.2

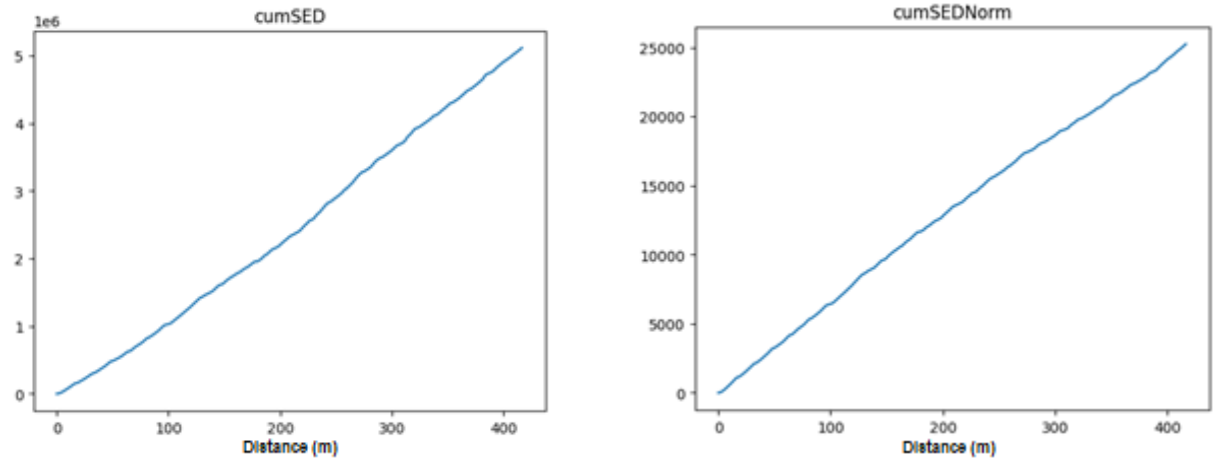
Une idée explorée est de normaliser le SED pour chaque trou, en divisant sa valeur par le maximum rencontré dans le trou. En lieu et place du maximum, le 95e percentile ($soit\ P_{95}()$) peut être utilisé pour éviter les valeurs trop élevées.

$$SED^{Norm}(\mathrm{x}) = \frac{SED(\mathrm{x})}{P_{95}(SED_{Trou})} \text{ et } SED^{Norm}_{Cumule}(\mathrm{x}) = \sum_{t=0}^{\mathrm{x}} SED^{Norm}(\mathrm{t}) \quad (56)$$

Où x représente la distance totale forée et SED_{Trou} est l'ensemble des valeurs SED dans le trou courant où se situe x.

La Figure 7.3b montre un exemple de SED normalisé cumulé pour un tricône. La valeur finale au moment du retrait du tricône est autour de 25x10^{3}.

Le Tableau 6 donne les valeurs finales pour le SED normalisé cumulé pour chaque tricône. En considérant cet indicateur comme critère de retrait des tricônes, le coefficient de variation normalisé observé serait de 83.6%, ce qui est une légère amélioration par rapport au 102.4% du SED cumulé, mais similaire à celui de la distance totale (82.9%). La variabilité observée en fin de vie avec cet indicateur est encore trop grande.



a- Énergie de forage cumulée

b- Énergie de forage normalisée cumulée

Figure 7.3 Évolution de l'énergie totale cumulée et énergie totale normalisée pour un tricône

7.5 Approche basée sur les variations de distribution du SED

Les approches basées sur l'énergie de forage cumulée ne sont pas concluantes. Cependant, tel qu'expliqué à la section 7.3.2.2, la Figure 7.2 montre que l'Énergie de forage SED croît bien avec

l'état d'usure du tricône et cet effet est plus accentué au niveau des percentiles élevés du SED. Une approche avec plus de granularité permettrait de suivre l'évolution des percentiles élevés du SED et serait une bonne indication du niveau d'usure des tricônes. Une fenêtre mobile de 12000 entrées (120 m) est utilisée pour évaluer la médiane ($P_{50}(SED(x))$), le 75e percentile ($P_{75}(SED(x))$) et le 90e percentile ($P_{90}(SED(x))$) du SED pour chaque incrément de distance totale (à partir de 120 m).

À chaque incrément de distance forée, les percentiles sont calculés sur les données au cours des 120 mètres précédents. La fenêtre de 120 m garantit qu'il y a assez de données pour que les percentiles soient significatifs et n'aient pas une évolution trop volatile.

Afin de faciliter la comparaison entre ces indicateurs basés sur différents percentiles, les valeurs de percentiles sont normalisées par rapport à la première valeur de percentile calculé :

$$\overline{P_{100\alpha}(SED(x))} = \frac{P_{100\alpha}(SED(x))}{P_{100\alpha}(SED(x=120m))} \tag{57}$$

où x représente la distance totale forée et α =0.5, 0.75 ou 0.90 et pour $x \leq 120$m. On considère que $\overline{P_{100\alpha}} = 1.0$

La Figure 7.4 montre les valeurs de $\overline{P_{50}}$, $\overline{P_{75}}$, et $\overline{P_{90}}$ pour chacun des tricônes. Une tendance haussière est visible pour la majorité des tricônes : plus la distance augmente, plus la médiane, le 75e percentile et le 90e percentile ont tendance à augmenter. Les baisses temporaires dans les percentiles normalisés peuvent s'expliquer par des changements géologiques : pendant 120 m le tricône se trouve dans une roche moins dure que d'habitude. Sachant que l'état des tricônes ne s'améliore pas, pour construire un indicateur à partir de ces signaux, il suffira de considérer le maximum cumulatif (ou minimum cumulatif pour un indicateur décroissant).

La croissance est plus stable pour le 75e percentile. Le 90e percentile (normalisé) semble exagéré les tendances comme pour le tricône 10 (entre 150 et 250 m) alors que le 50e percentile sous-évalue parfois la tendance comme pour le tricône 7 (entre 200 et 275 m). La valeur finale $\overline{P_{75}}$ oscille autour de 1.6.

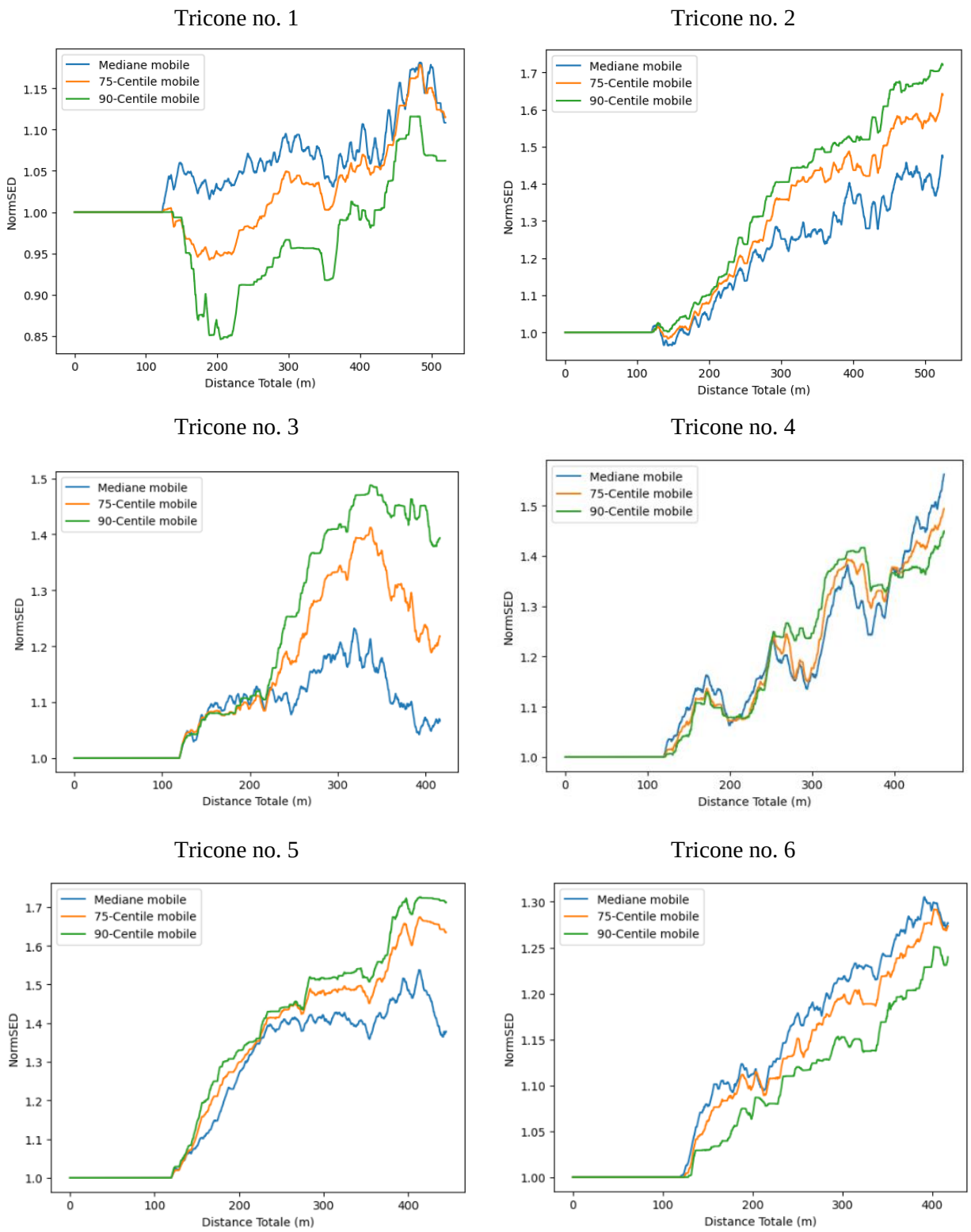


Figure 7.4a – Valeurs de $\overline{P_{50}}$, $\overline{P_{75}}$ et $\overline{P_{90}}$ pour chaque tricône (fenêtre mobile de 120 m)

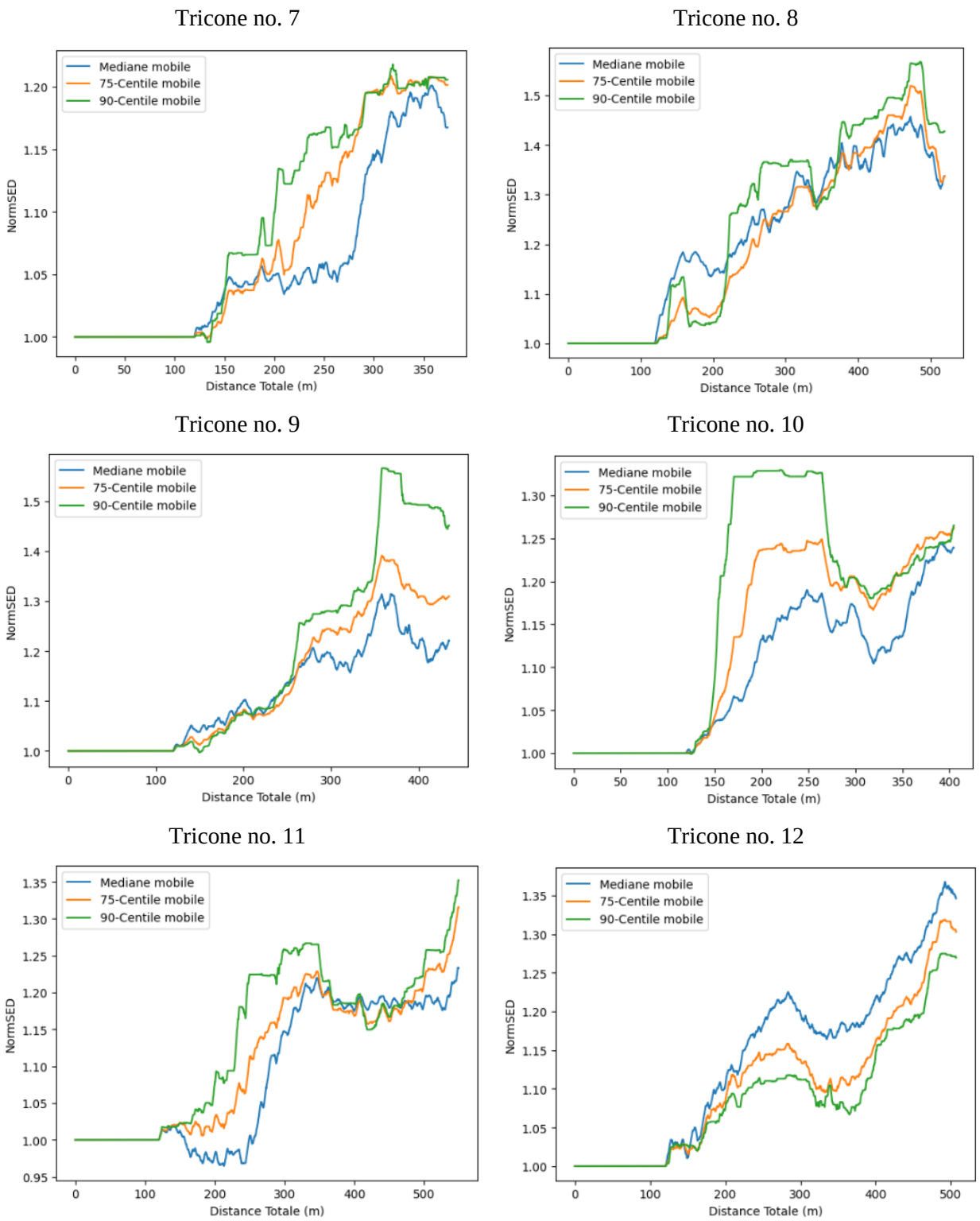


Figure 7.4b – Valeurs de $\overline{P_{50}}$, $\overline{P_{75}}$ et $\overline{P_{90}}$ pour chaque tricône (fenêtre mobile de 120 m) - Suite

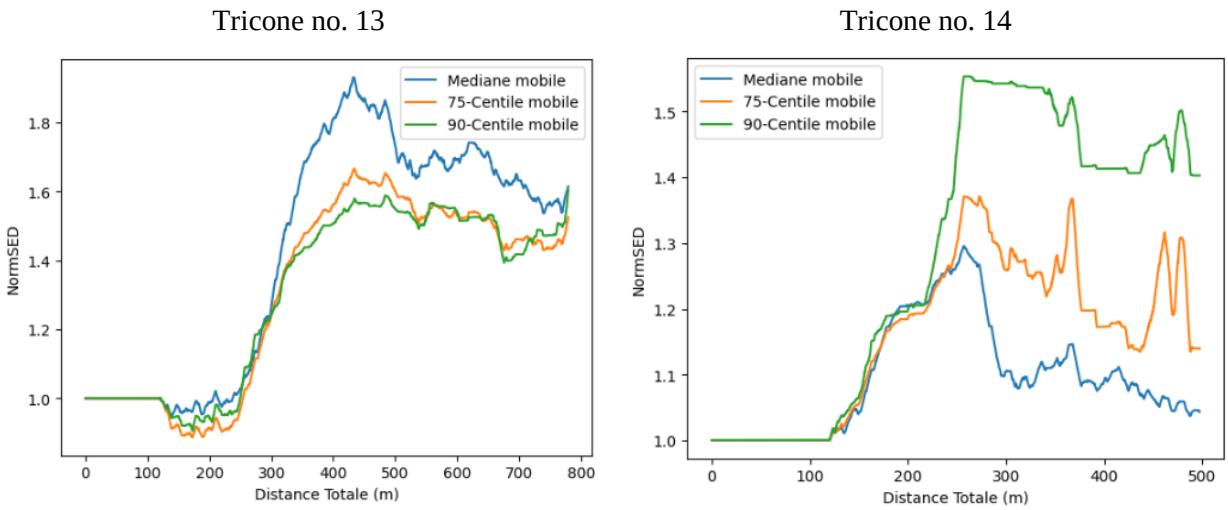


Figure 7.4c – Valeurs de $\overline{P_{50}}$, $\overline{P_{75}}$ et $\overline{P_{90}}$ pour chaque tricône (fenêtre mobile de 120 m) - Suite

Le 75e percentile mobile du SED (normalisé) est utilisé pour construire un indicateur d'efficacité du tricône appelé BWi (*Bit Wear index*), calculé comme suit :

$$\mathrm{RawBWi(x)} = \left(\overline{\mathrm{P_{75}(SED(x))}} - 0.625\right) \times 2.4 + 0.1 \tag{58}$$

où x représente la distance totale forée. RawBWi(x) est une simple transformation de $\overline{\mathrm{P_{75}(SED(x))}}$. L'indicateur ainsi défini commence à la valeur 1.0 quand $\overline{\mathrm{P_{75}}} = 1.0$ (soit 100% pour un tricône neuf) et le *BWi* prend la valeur de 0.1 quand $\overline{\mathrm{P_{75}}} = 1.6$ (soit 10% pour un tricône proche de la fin).

Afin de ne considérer que le minimum cumulatif et ignorer les variations, on pose :

$$\mathrm{BWi(x)} = Min(\,\{\mathrm{RawBWi(t)} | \mathrm{t} \leq \mathrm{x}\}\,) \tag{59}$$

La Figure 7.5 présente l'indicateur BWi pour chaque tricône. L'indicateur commence à 100% pour un tricône neuf et la tendance baissière débute à partir de 120 m et se poursuit jusqu'au retrait, avec parfois des périodes plateau qui pourrait suggérer que le tricône a été dans de la roche beaucoup moins dure que d'habitude et n'a pas accumulé beaucoup d'usure.

Le Tableau 6 résume les valeurs finales de BWi pour chaque tricône. Ces valeurs varient entre 60% et 3%. Le coefficient de variation normalisé associé à cet indicateur est de 32.2%. Ce qui est une nette amélioration par rapport aux indicateurs précédents.

Cet indicateur est un bon outil pour capturer les tendances de dégradation du tricône et offre une meilleure façon de définir un seuil de retrait des tricônes. Cependant, elle néglige l'usure potentiellement accumulée pendant les 120 premiers mètres de forage, et elle engendre dans certain cas de longues périodes de forage où on observe aucune dégradation (car le 75e percentile dans la fenêtre mobile ne fournit pas de nouvelle valeurs maximales), ce qui est rigoureusement peu réaliste.

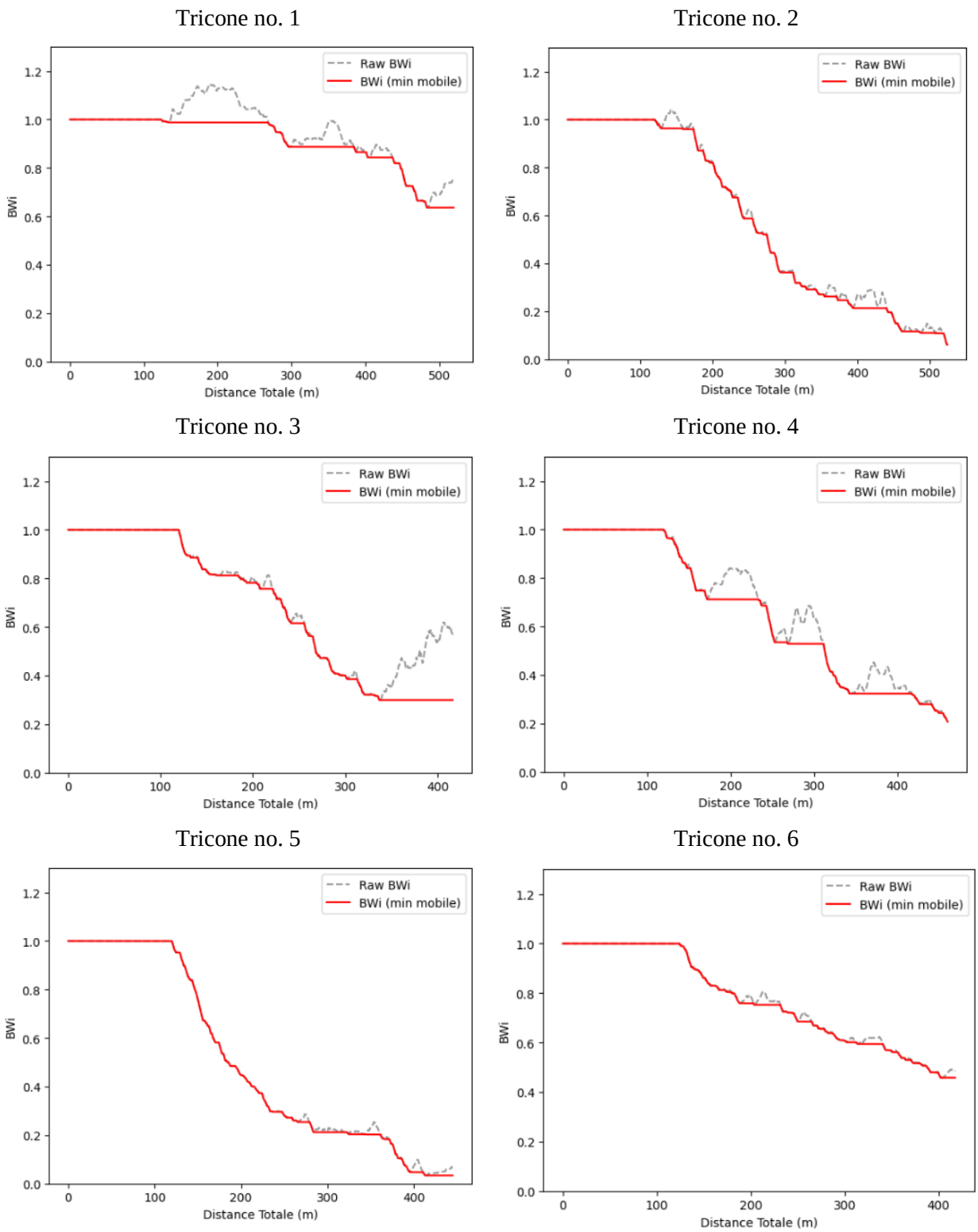
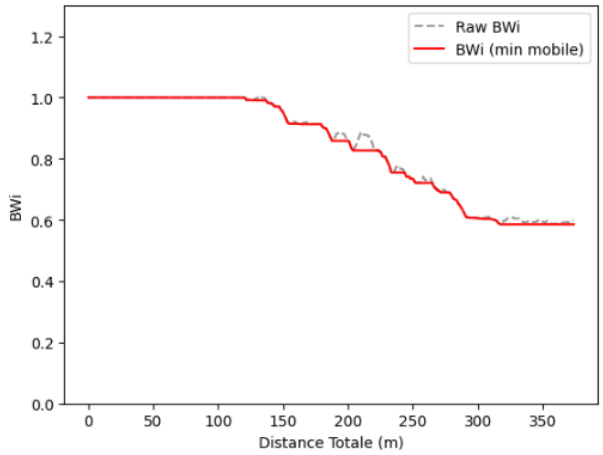
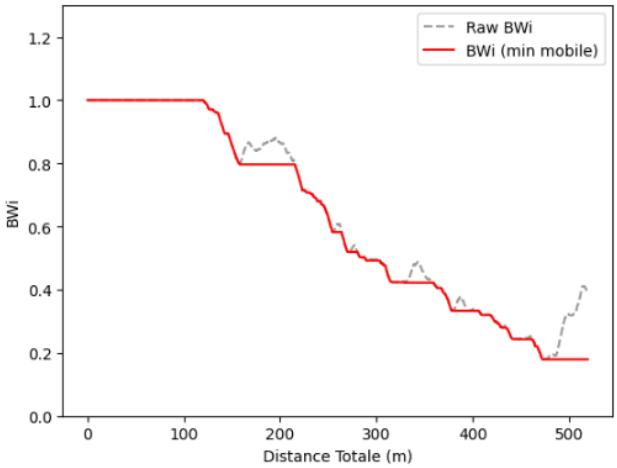


Figure 7.5a – Valeurs de *BWi* calculées pour chaque tricône (fenêtre mobile de 120 m)

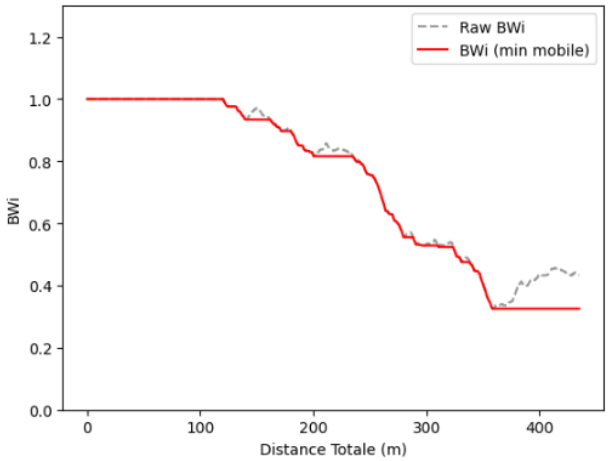
Tricone no. 7



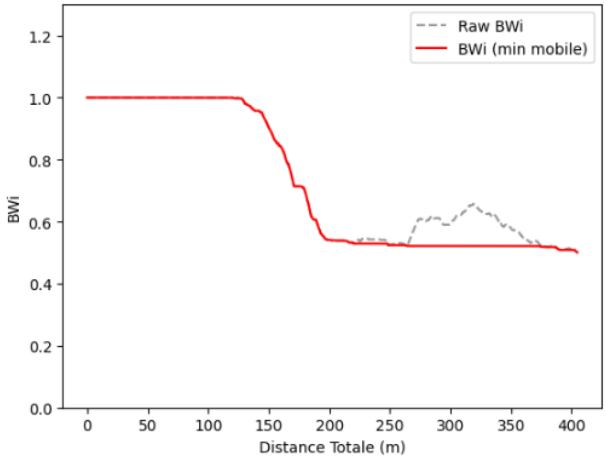
Tricone no. 8



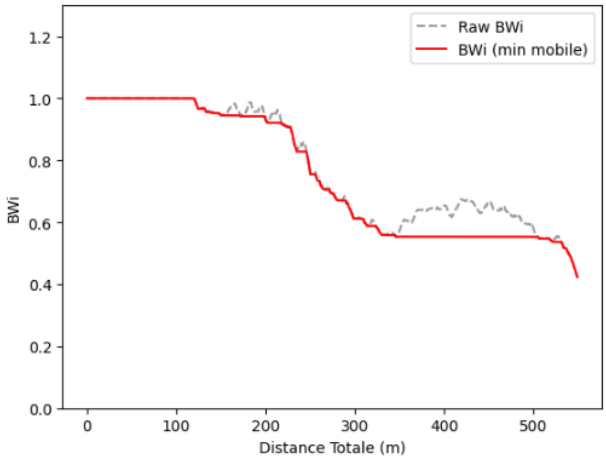
Tricone no. 9



Tricone no. 10



Tricone no. 11



Tricone no. 12

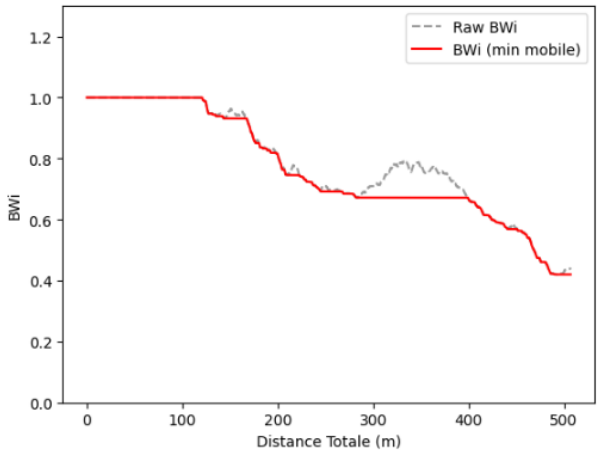


Figure 7.5b – Valeurs de *BWi* calculées pour chaque tricône (fenêtre mobile de 120 m) - Suite

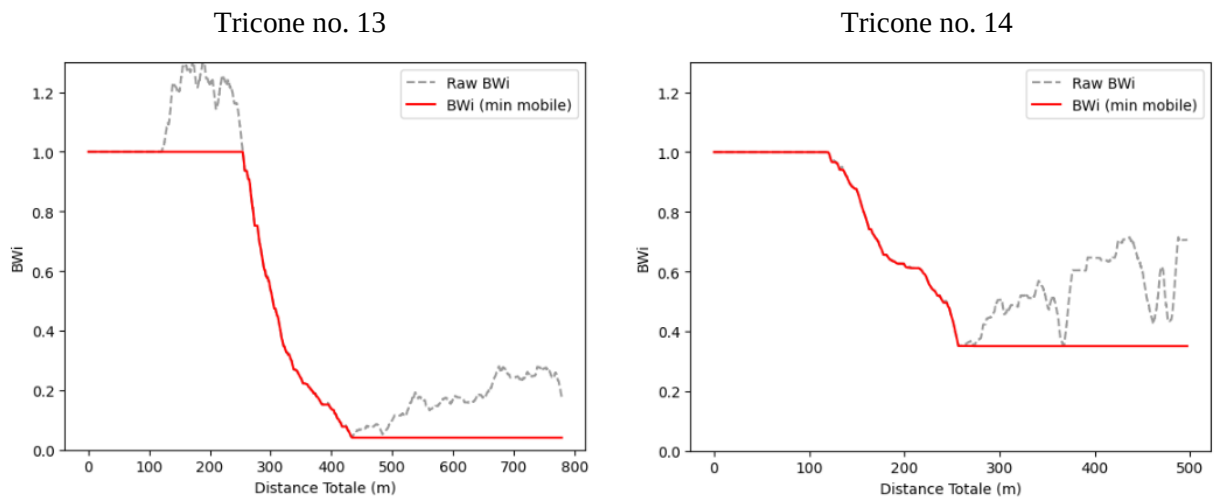


Figure 7.5c – Valeurs de *BWi* calculées pour chaque tricône (fenêtre mobile de 120 m) - Suite

7.6 Modèle d'apprentissage supervisé (LSTM)

Dans cette section, un modèle de régression est proposé pour calculer un indice similaire au *BWi, soit le BWi-LSTM* mais qui ne nécessiterait pas d'attendre 120 m pour une prédiction et qui n'engendrerait pas de longues périodes où aucune usure ou dégradation n'est prédite.

Le modèle proposé utilise les MWD pendant une fenêtre donnée (10 mètres) pour prédire non pas l'état du tricône, mais le taux d'usure. Le problème étant chronologique, le choix d'un réseau de neurone de type récurrent s'est imposé. Une architecture basée sur les LSTM est sélectionnée pour leur capacité à capturer et à exploiter les dépendances temporelles dans les données séquentielles avec leurs mémoires à court et long terme.

7.6.1 Variables d'entrée, variable cible et données d'entrainement

Les variables d'entrées du modèle sont une séquence de MWD standardisés ($\widehat{\mathrm{RPM}}$, $\widehat{\mathrm{WOB}}$, $\widehat{\mathrm{TRQ}}$, $\widehat{\mathrm{ROP}}$, $\widehat{\mathrm{SED}}$) sur une fenêtre mobile de 10 m (soit une séquence de 1000 mesures MWD consécutives). La variable cible est le différentiel de l'indice BWi calculé par la méthode des percentiles et ramené à l'échelle de 1 à 0 : $\delta_{BWi}(\mathrm{x}) = \overline{BWi}(\mathrm{x}) - \overline{BWi}(\mathrm{x} - 0.01)$ où x est la distance totale forée. Ce différentiel représente un indice de « taux d'usure » ou « taux de dégradation » progressive entre deux relevés de MWD (séparés de 0.01 m). Cette intuition du taux d'usure vient du fait que $\overline{BWi}(\mathrm{x})$ à chaque distance x est une mesure de l'état d'efficacité du trépan, elle peut

donc également être interprétée comme un cumul de la dégradation ou de la perte d'efficacité du trépan au fil de l'avancement forage.

Le BWi est ramené à l'échelle 1 à 0 ($\overline{BWi}$) pour chaque tricône, car la valeur finale d'un indice parfait aurait dû être nul en fin de vie pour chacun des tricônes. Ainsi le profil de taux d'usure est conservé pour chaque tricône, mais il est ajusté pour que l'indice d'usure accumulé au moment soit 0 (100% d'usure). La Figure 7.6 présente un exemple de différentiel $\overline{BWi}$ ou du taux d'usure. On remarque des périodes où ce différentiel est de 0 et des périodes de gros changements. Idéalement, s'il n'y a pas de surapprentissage, le modèle LSTM aura pour effet de lisser ou répartir cette dégradation sur plusieurs périodes.

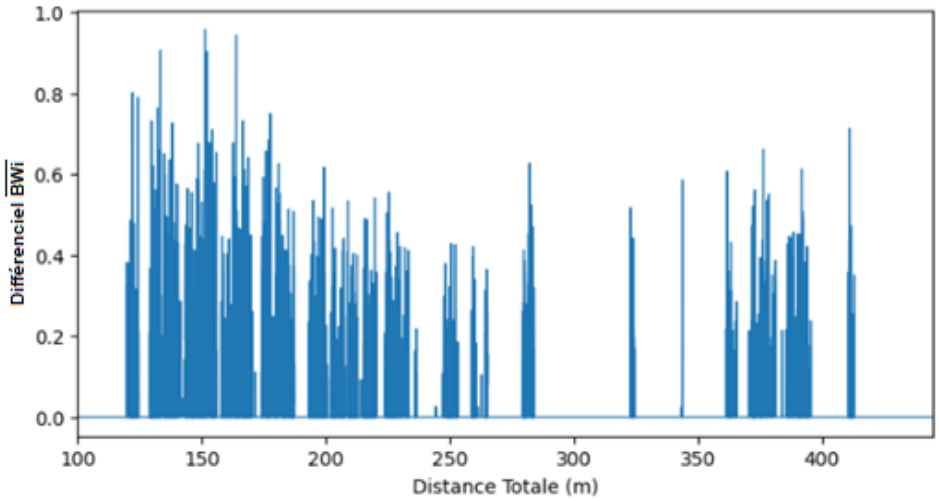


Figure 7.6 – Exemple de différentiel de BWi (variable cible δ_{BWi})

La Figure 7.9 présente un exemple de variable d'entrée (séquence de MWD) et de variable de cible correspondant (différencie de BWi).

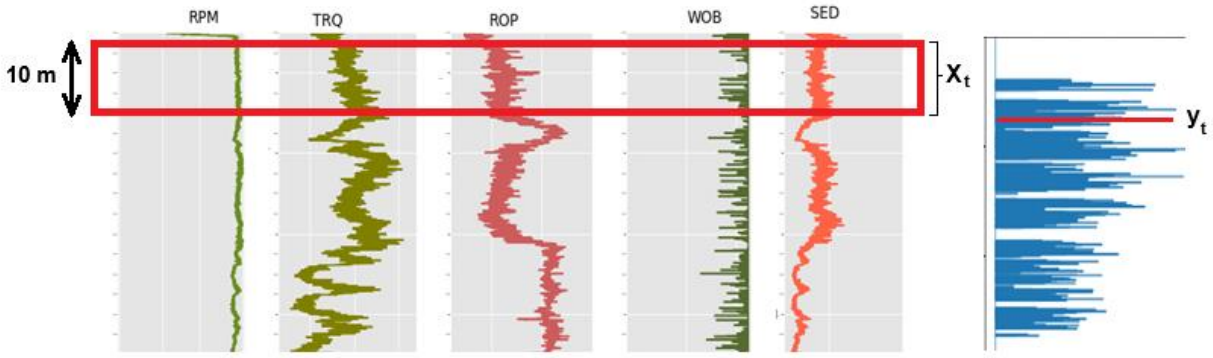


Figure 7.7 – Exemple de différentiel de *BWi* (variable cible)

Les 5 tricônes suivants ont été sélectionnés aléatoirement pour l'entraînement et la validation du modèle : $\{5, 7, 8, 11, 14\}$. De plus, 30% des séquences entrée/sortie sont gardées pour la validation du modèle.

Le reste des 9 tricônes est gardé comme ensemble test : $\{1, 2, 3, 4, 6, 9, 10, 12, 13\}$.

7.6.2 Architecture du modèle et hyperparamètres

Le modèle proposé est décrit par les éléments suivants :

- Une couche d'entrée et de masquage (*Masking*) pour ignorer les valeurs nulles.
- Une couche LSTM avec 128 neurones. La fonction d'activation dans la couche LSTM est la fonction *Sigmoïde* pour les portes d'entrée, d'oubli et de sortie, et la fonction *tanh* pour l'état de la cellule.
- Une couche de sortie *Dense* (*Fully Connected*) avec une seule unité et une fonction d'activation linéaire pour les prédictions finales.
- La fonction de perte utilisée est l'erreur quadratique moyenne MSE *(Mean Squared Error*)
- L'optimiseur Adam est sélectionné avec un taux d'apprentissage $\eta = 0.0001$.
- Le nombres d'époques pour l'entrainement est de 100 et un critère d'arrêt *EarlyStopping* est utilisé pour arrêter l'entraînement si la perte de validation ne diminue pas après 10 époques.

La Figure 7.8 montre les courbes de perte d'entraînement et de validation (MSE en fonction des *epochs*). Cette courbe suggère un entraînement réussi, avec un bon équilibre entre la capacité d'apprentissage du modèle et sa généralisation. Le modèle semble bien ajusté, sans indication de surapprentissage ou de sous-apprentissage. Le MSE pour les deux courbes est proche de 0.

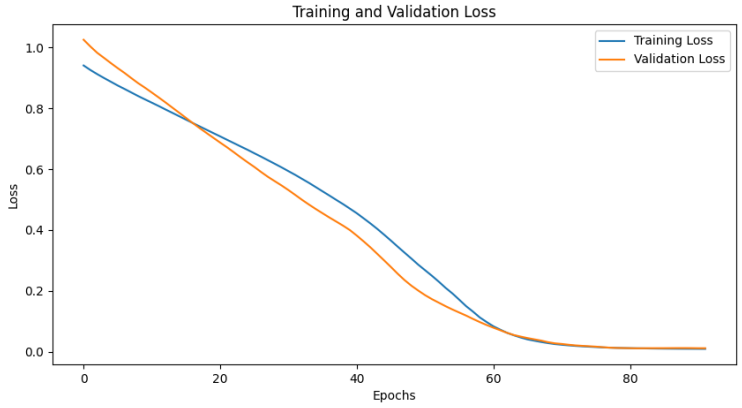


Figure 7.8 - Fonction de perte pour l'entrainement et la validation du modèle LSTM

7.6.1 Analyse des résultats de tests

La Figure 7.9 montre un exemple de prédiction du taux d'usure δ_{BWi} fournit par le modèle pour un tricône, et la compare au δ_{BWi} calculé à partir des BWi calculés à l'aide des percentiles. Le taux d'usure de la méthode des percentiles présente des piques et de longues périodes nulles comme pour les premiers 120 m, tandis que le taux d'usure prédit par le modèle LSTM est plus réparti sur la durée de vie du tricône. Les sommes cumulatives des deux taux d'usure sont comparables.

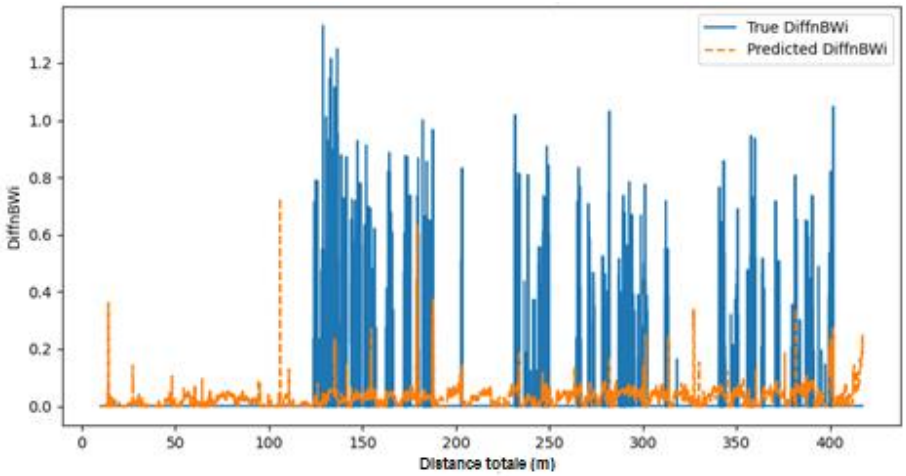


Figure 7.9 – Exemple de différentiel de BWi prédit par le modèle test

Une fois que le taux d'usure est prédit par le modèle LSTM, il est facile de reconstituer la prédiction d'état d'usure du tricône à chaque distance totale de forage x, soit BWi_{LSTM} :

$$\mathrm{BWi}_{LSTM}(\mathrm{x}) = 1 - \sum_{t=0}^{\mathrm{x}} \delta_{BWi_{LSTM}}(t) \tag{60}$$

La Figure 7.10 présente les prédictions de BWi_{LSTM} pour chacun des tricônes dans l'ensemble test. Le BWi prédit par le modèle commence à 1.0 (état de 100%) et décroit graduellement à partir de 10 m. Pour plusieurs tricônes, l'indice BWi à la fin est effectivement proche de 0.0 (tricône 1, 9, 10 et 12) qui est donc le seuil auquel il faudrait remplacer le tricône suivant cet indice.

Le Tableau 6 montre le BWi final prédit pour chacun des tricônes tests. Pour plusieurs tricônes, le BWi final est en deça de 0, ce qui signifie que le modèle estime que le tricône aurait dû être remplacé plus tôt. Malgré cela, ce modèle a un coefficient de variation normalisé de 26.7%, qui est le plus bas de toutes les approches proposées, ce qui suggère que cette approche est meilleure lorsque comparée aux données observées.

Le modèle LSTM a ainsi tendance à prédire un remplacement anticipé des tricônes par rapport aux observations de distance totale. Ce biais du modèle aurait pu être corrigé en lui fournissant plus de données d'entraînement. Toutefois, il est à noter que ce biais va dans le sens conservateur : il est préférable d'avoir un modèle qui prédit une usure anticipée plutôt que tardive, au risque de générer des bris de tricône en cours d'opération ou de la perte de productivité significative dû à un tricône inefficace.

La dernière colonne du Tableau 6 fournit la distance à laquelle le modèle prédit un BWi égal à 0.0. Pour 6 cas sur 9, cette distance totale prédite est en bas de 12% de ce qui a été observations. Pour les 3 autres cas, le modèle prédit une distance totale jusqu'à 33% moins longue que ce qui a été observé.

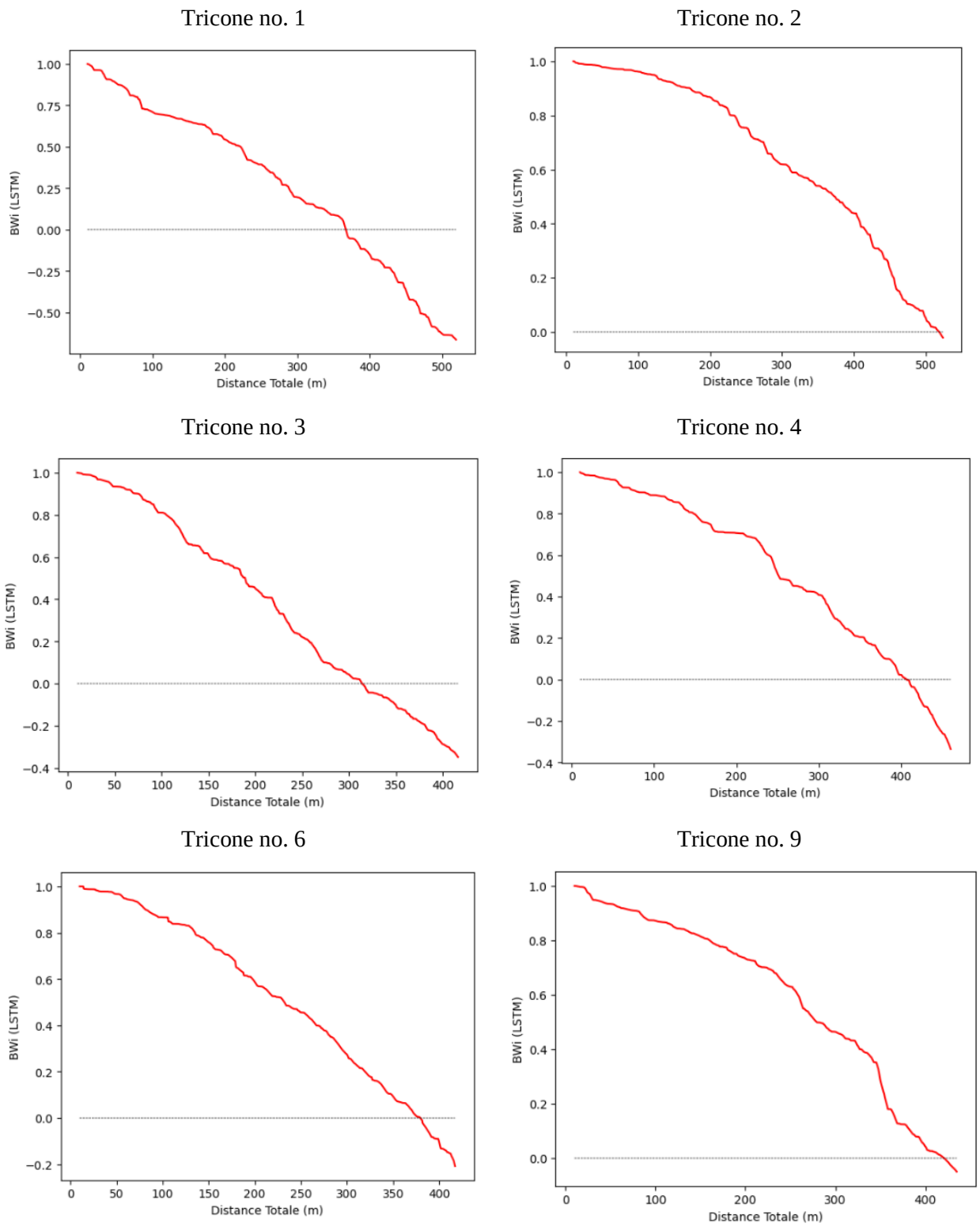


Figure 7.10 – Valeurs de BWi prédites par le modèle LSTM pour les tricônes tests

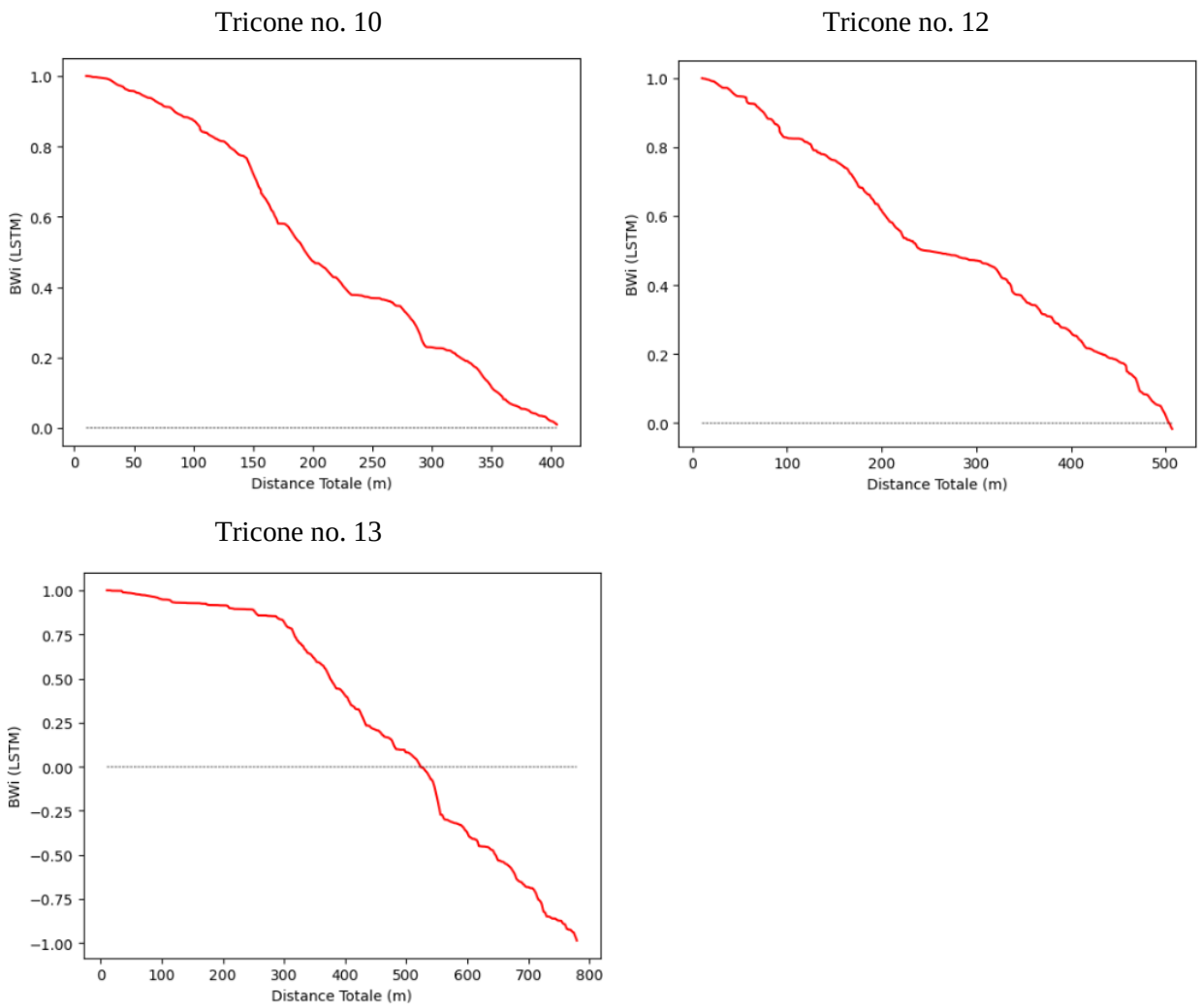


Figure 7.10 – Valeurs de BWi prédites par le modèle LSTM pour les tricônes tests - Suite

Tableau 6 – Comparaison des méthodes d'évaluation de fin de vie des tricônes

Bit no.	Distance de forage totale (m)	SED Cumulé final	NormSED Cumulé final	BWi final (par 75e percentile)	LSTM	
					BWi final	Distance totale prédite (BWi=0) (m)
1	519.6	5892622.3	35881.2	0.64	-0.66	367.2 (-29.3%)
2	523.8	5788289.7	31904.6	0.06	-0.02	518.6 (-1.0%)
3	416.4	5115087.2	25242.5	0.30	-0.35	314.0 (-24.6%)
4	460.3	5552371.9	28856.5	0.21	-0.33	406.7 (-11.6%)
5	444.8	5307297.9	28464.5	0.03	-	-
6	417.4	5465237.4	30062.1	0.46	-0.21	379.6 (-9.1%)
7	373.9	5741141.2	26110.1	0.59	-	-
8	519.5	6583250.1	34766.6	0.18	-	-
9	435.0	5341552.0	29154.7	0.33	-0.05	421.0 (-3.2%)
10	404.7	5365363.1	29582.7	0.50	0.01	404.7 (0.0%)
11	549.2	8247737.4	38690.5	0.42	-	-
12	506.9	6673796.6	36351.0	0.42	-0.02	503.6 (-0.7%)
13	779.1	9055894.8	51569.9	0.04	-0.99	523.6 (-32.8%)
14	497.1	7275147.5	35112.7	0.35	-	-
CV_{Norm}	***82.9%***	***102.4%***	***83.6%***	***32.2%***	***26.7%***	

7.7 Discussion sur l'utilisation de l'indice d'usure du trépan

L'indice d'état de performance du tricône (BWi) proposé dans ce chapitre est utilisé pour prédire le moment optimal de remplacement des tricônes. Cependant, l'utilisation d'un indice continu plutôt que d'un indice catégorisé, comme dans les travaux précédents, présente plusieurs avantages supplémentaires.

Correction des MWD : Les indicateurs de dureté de la roche, tels que le SED, peuvent être normalisés en fonction du niveau d'usure prédit du trépan. En effet, l'énergie totale nécessaire pour briser la roche dépend non seulement des propriétés de la roche elle-même, mais aussi de l'état d'usure du tricône. Avec une évaluation continue de l'état du tricône, il devient possible de normaliser le SED pour qu'il reflète plus fidèlement la dureté de la roche, ou de développer un autre indice capable d'isoler l'effet de l'état du tricône.

Bien que davantage de données soient nécessaires pour calibrer ce facteur de correction, celui-ci pourrait être formulé de la manière suivante :

$$SED_{norm} = f(BWi) \times SED \quad (61)$$

où $f(BWi)$ représente une fonction dépendant de l'indice d'usure du tricône (BWi) et serait à déterminer.

Amélioration des modèles de prédiction du charbon : L'indice BWi proposé peut également contribuer à améliorer la prédiction de la présence de charbon par les modèles de classification supervisés du chapitre 4. Actuellement, ces modèles utilisent un seuil fixe de 50 % pour déterminer, à partir de la probabilité de charbon calculée par le modèle, s'il s'agit de charbon ou non. Cependant, comme l'état d'usure du trépan peut influencer les MWD et modifier la perception de la dureté de la roche forée, ce seuil de probabilité pourrait être ajusté dynamiquement en fonction de l'état d'usure du tricône.

Par exemple, lorsque le trépan est usé, le seuil de probabilité pourrait être abaissé pour compenser la tendance des modèles à interpréter les données comme représentant une roche plus dure, et donc à réduire la probabilité de charbon qui sera prédite. En ajustant le seuil en fonction de l'indice BWi, on pourrait ainsi corriger ce biais lié à l'usure du trépan et améliorer la précision des prédictions de charbon.

7.8 Conclusion et recommandations

Ce chapitre a exploré et comparé plusieurs méthodes de prédiction de l'usure des tricônes en utilisant les données MWD collectées en continu pendant les opérations de forage. Parmi les approches étudiées, l'utilisation d'un modèle basé sur les réseaux de neurones récurrents LSTM s'est avérée la plus performante. Grâce à sa capacité à capturer les dépendances temporelles dans les données séquentielles, l'architecture LSTM a permis de prédire un taux indice de taux d'usure puis d'estimer un indice d'usure (BWi-LSTM) qui reflète de manière continue l'état de dégradation du tricône. Le LSTM a obtenu un coefficient de variation normalisé qui est le plus faible (26.7 %) parmi les approches testées, ce qui indique une meilleure stabilité et cohérence dans la prédiction du moment optimal de remplacement.

L'utilisation d'un indice continu pour représenter l'usure du trépan, plutôt qu'un indicateur catégorisé, ouvre des possibilités pour ajuster dynamiquement certains paramètres de forage (comme le SED) en fonction de l'état du trépan. En tenant compte de l'usure du trépan à travers

l'indice BWi, les modèles de prédiction de la présence de charbon peuvent être améliorés en ajustant leurs seuils de probabilité en fonction de l'état d'usure, ce qui pourrait réduire les biais dans l'identification des zones riches en charbon.

Bien que le modèle LSTM prédise une usure parfois anticipée par rapport aux observations, ce biais conservateur peut être bénéfique pour la sécurité et la protection de l'équipement. Il serait toutefois pertinent de déterminer des seuils d'usure optimaux basés sur les données de terrain. Pour améliorer davantage la précision et la robustesse du modèle LSTM, il serait souhaitable de collecter plus de données MWD, y compris des annotations fréquentes de l'état des tricônes.

Pour maximiser les bénéfices de l'approche LSTM, l'intégration de ce modèle dans un système de suivi en temps réel serait idéale. Cela permettrait aux opérateurs de surveiller l'état des tricônes et de prendre des décisions proactives concernant le remplacement, minimisant ainsi les arrêts imprévus et améliorant la productivité.

CHAPITRE 8 CONCLUSION ET RECOMMANDATIONS

8.1 Synthèse des travaux

Cette thèse s'inscrit dans le contexte de l'optimisation des opérations minières, avec un accent sur l'identification en temps réel de la lithologie et la détection de l'usure des trépans dans des environnements de forage exigeants. Pour répondre aux défis de précision, de rapidité et d'automatisation posés par l'industrie minière, cette recherche propose des modèles d'apprentissage automatique, notamment des architectures basées sur les réseaux LSTM, pour analyser les données MWD (*Measurement While Drilling*) en temps réel et ainsi améliorer les décisions opérationnelles.

Le chapitre 4 a proposé un modèle intégré pour la détection en temps réel des couches de charbon en utilisant les données MWD. Ce modèle est basé sur l'intégration d'un ensemble de classificateurs individuels supervisés. L'approche permet également d'améliorer la prédiction en corrigeant le décalage de profondeur entre les données des relevés géophysiques et les MWD. Les résultats montrent que l'identification en temps réel du charbon réduit le besoin de diagraphies géophysiques coûteuses, avec une identification du sommet des couches de charbon précise à ±20 cm, améliorant ainsi l'efficacité du dynamitage et réduisant les coûts liés aux opérations de dynamitage et de forage.

Le chapitre 5 a introduit une approche d'apprentissage non supervisé, intégrant des données de trous de forage adjacents pour améliorer la stabilité de la classification lithologique, en prenant en compte la proximité spatiale des données. Trois modèles d'apprentissage non supervisés sont comparés et la méthode recommandée est celle basée sur le *partitionnement spectral*. Cette méthode a permis une meilleure cohérence des prédictions dans des régions géologiquement similaires. Cependant, cette approche pourrait être limitée par des exigences de calcul élevées, rendant son utilisation en temps réel difficile.

Finalement au chapitre 6, un modèle LSTM a été appliqué pour la détection de l'usure des tricônes en utilisant les données MWD en continu. Grâce aux capacités des LSTM à capturer les dépendances temporelles dans les données séquentielles, le modèle a su prédire l'état d'usure du trépan de manière continue, permettant ainsi de déterminer des moments optimaux de

remplacement. Bien que prometteur, le modèle nécessite un volume de données important pour une précision optimale et peut être sensible au bruit dans les données MWD. Une amélioration de la filtration des données, ainsi que l'ajout de facteurs additionnels liés à l'usure, pourraient renforcer la robustesse du modèle.

8.2 Contributions majeures et originalité

Cette recherche représente une avancée importante vers l'automatisation des décisions de maintenance et la détection en temps réel des ressources minérales. Les modèles proposés contribuent à la réduction des coûts d'usure des équipements, à l'optimisation des opérations de forage et à une meilleure efficacité opérationnelle dans le secteur minier.

Cette thèse apporte des contributions notables, notamment par l'application innovante des réseaux LSTM pour la détection de l'usure des tricônes et par le développement d'un modèle de classification adaptatif pour la détection en temps réel de la lithologie. Ces approches démontrent la capacité des modèles d'apprentissage automatique à optimiser les processus de forage minier, ouvrant la voie à des solutions intelligentes et autonomes.

L'algorithme de classification du charbon, déployé dans une mine de charbon canadienne, confirme l'impact de ces travaux. Son implémentation permet d'anticiper des gains significatifs en efficacité opérationnelle, validant ainsi la pertinence et l'applicabilité industrielle des modèles développés.

8.3 Pistes d'améliorations futures

Pour renforcer la robustesse et la précision des modèles proposés, plusieurs pistes de recherche et recommandations pour des applications industrielles sont suggérées :

- Extension des sources de données : L'intégration de données supplémentaires, telles que les vibrations, pourrait permettre d'affiner la précision des prédictions et mieux comprendre l'interaction entre l'usure du trépan et les propriétés géologiques.
- Validation en milieu industriel diversifié : Tester et valider les modèles dans des contextes miniers variés (différents types de mines et de formations géologiques) permettraient d'assurer les modèles proposés ont une bonne capacité de généralisation et prouver leur

robustesse. L’application de ces modèles à d’autres ressources minérales que le charbon et à d’autres équipements de forage constituerait également une voie de recherche prometteuse.

- Utilisation de modèles plus avancés : avec la collecte de données supplémentaires, il deviendrait possible d’entraîner des modèles plus sophistiqués, tels que les transformers. Ces modèles, réputés pour leur capacité à capturer des dépendances complexes à longue portée dans les séquences de données, pourraient améliorer la précision des prédictions et offrir des interprétations plus fines des données de forage. Les transformers, en particulier, pourraient traiter des volumes de données MWD massifs et hétérogènes, permettant ainsi d'identifier la lithologie pour des géologies plus complexes et améliorer la détection de l’usure des tricônes, ouvrant ainsi des perspectives d’automatisation encore plus poussées dans les opérations minières.

RÉFÉRENCES

Abbas, R. K. (2018). A review on the wear of oil drill bits (conventional and the state of. *Engineering Failure Analysis, 90*, 554-584.

Agostini, C. E., & Sampaio, M. A. (2020). Probabilistic Neural Network with Bayesian-based, spectral torque imaging and Deep Convolutional Autoencoder for PDC bit wear monitoring. *Journal of Petroleum Science and Engineering, 193*. doi:10.1016/j.petrol.2020.107434

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035). Society for Industrial and Applied Mathematics.

Babaei Khorzoughi, M. (2013). *Use of measurement while drilling techniques for improved rock mass characterization in open-pit mines.* Doctoral dissertation, University of British Columbia.

Bajram, E., Hudson, D., & Adam, S. (2015). *RAB drill rig top of coal detection while drilling, ACARP C21005.* Tech. rep., CRCMining.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157-166.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

Breiman, L. (1996). Bagging predictors. *Machine learning, 24*, 123–140.

Caterpillar Inc. (2024, 07 04). *Drills MD6310*. Consulté le July 4, 2024, sur Caterpillar site web officiel: https://www.cat.com/en_US/products/new/equipment/drills/rotary-drills/1000032468.html

Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1–4.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297. doi:10.1007/BF00994018

Davis, J., & Goadrich, M. (2006). The Relationship between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233–240). New York, NY, USA: Association for Computing Machinery. doi:10.1145/1143844.1143874

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1-38.

Facundo, B., Gimenez, J., & Fachinotti, V. (2017). Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks. *Energy and Buildings, 158*.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*, 367-378. doi:https://doi.org/10.1016/S0167-9473(01)00065-2

Hatherly, P., Leung, R., Scheding, S., & Robinson, D. (2015). Drill monitoring results reveal geological conditions in blasthole drilling. *International Journal of Rock Mechanics and Mining Sciences, 100*, 144–154.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735 - 1780.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Jolliffe, I. (2002). *Principal Component Analysis.* Springer.

Klyuchnikov, N., Zaytsev, A., Gruzdev, A., Ovchinnikov, G., Antipova, K., Ismailova, L., . . . Koroteev, D. (2018). Data-driven model for the identification of the rock type at a drilling bit. *Data-driven model for the identification of the rock type at a drilling bit*.

Kumar, B. R., Vardhan, H., Govindaraj, M., & Vijay, G. S. (2013). Regression analysis and ANN models to predict rock properties from sound levels produced during drilling. *International Journal of Rock Mechanics and Mining Sciences, 58*, 61–72.

LaBelle, D. (2001, 1). *Lithological Classification By Drilling.* Tech. rep., Carnegie Mellon University, Pittsburgh.

LaBelle, D., Bares, J., & Nourbakhsh, I. (2000). Material classification by drilling. *Proceedings of the 17th International Symposium on Automation and Robotics in Construction.(International Association for Automation and Robotics in Construction), 10*.

Leung, R., & Scheding, S. (2015). Automated coal seam detection using a modulated specific energy measure in a monitor-while-drilling context. *International Journal of Rock Mechanics and Mining Sciences, 75*, 196–209.

Liu, Z., Hong, H., Gan, Z., Hou, Y., & Zeng, Z. (2022). Design and analysis of an optical zoom system-based on super-elastic film liquid lens. *2021 International Conference on Optical Instruments and Technology: Optical Systems, Optoelectronic Instruments, Novel Display, and Imaging Technology},, 12277*. International Society for Optics and Photonics. doi:10.1117/12.2616337

Luu, T. P., Bomidi, J. A., Magana-Mora, A., Alalsayednassir, A., & Zhan, G. D. (2021). Bi-Directional Long Short-Term Memory Variational Autoencoder for Real-Time Bit-Wear Estimation. *SPE Asia Pacific Oil and Gas Conference and Exhibition*. doi:10.2118/205627-MS

MacQueen, J. B. (1997). Some Methods for classification and Analysis of Multivariate Observations. . *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1*, pp. 281-297.

Mason, I. (2011). *MWD for super-resolution control of R.A.B. drilled blast-hole stand-off distances.* Tech. rep., University of Sydney, ACARP Project C19023.

Meisel, W. S., & Michalopoulos, D. A. (1973). A partitioning algorithm with application in pattern classification and the optimization of decision trees. *IEEE Transactions on Computers, 100*, 93–103.

Mohammed, R. S. (2018). Well Logging Analyzing in Misan Province South of Iraq with Estimation of Logging Expenses. *Journal of Advanced Research in Dynamical & Control Systems, 10*(11).

Ng, A., Jordan, M., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems, 14*.

Niu, G., Zhang, K., Yu, B., Chen, Y., Wu, Y., & Liu, J. (2019). Experimental Study on Comprehensive Real-Time Methods to Determine Geological Condition of Rock Mass along the Boreholes while Drilling in Underground Coal Mines. *Shock and Vibration*, 17. doi:10.1155/2019/1045929

Peck Tech. (2009). *Functional Requirements Definition for the Compensated Blastability Index (CBI) Algorithm.* Internal Technical Document. Peck Tech Consulting.

Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia, 4*, 1883.

Rafezi, H., & Hassani, F. (2021). Drilling signals analysis for tricone bit condition monitoring. *International Journal of Mining Science and Technology2021, 31*(2), 187-195. doi:https://doi.org/10.1016/j.ijmst.2020.12.025

Rafezi, H., & Hassani, F. (2023). Drill bit wear monitoring and failure prediction for mining automation. *International Journal of Mining Science and Technology, 33*(3), 289-296.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533-536.

Saito, T. (2015). The Relationship between Precision-Recall and ROC Curves.

Scoble, M. J., & Peck, J. (1987). A technique for ground characterization using automated production drill monitoring. *International Journal of Surface Mining, Reclamation and Environment, 1*, 41–54.

Scoble, M. J., Peck, J., & Hendricks, C. (1989). Correlation beween rotary drill performance parameters and borehole geophysical logging. *Mining Science and Technology, 8*, 301–312.

Segui, J. B., & Higgins, M. (2002). Blast Design Using Measurement While Drilling Parameters. *Fragblast, 6*, 287-299. doi:10.1076/frag.6.3.287.14052

Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, (pp. 747-748). Sydney.

Shi, J., & Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8).

Singh, P., Roy, M., Paswan, R., Sarim, M., Kumar, S., & Jha, R. R. (2016). Rock fragmentation control in opencast blasting. *Journal of Rock Mechanics and Geotechnical Engineering, 8*(2), 225-237. doi:https://doi.org/10.1016/j.jrmge.2015.10.005

Smith Technologies. (2008). *Roller Cone Dull Grading Manual.* Houston: Smith International, Inc.

Thomas, T. (2008). *IADC CODES.* IADC CODES.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology, 49*, 1225–1231.

Wu, G., Zhang, H., He, Y., Bao, X., L., L., & Hu, X. (2019). Learning Kullback-Leibler Divergence-Based Gaussian Model for Multivariate Time Series Classification. *IEEE Access, 7*, 139580-139591.

Zagré, G., Gamache, M., Labib, R., & Shlenchak, V. (2024). Machine learning algorithms for real-time coal recognition using monitor-while-drilling data. *International Journal of Mining, Reclamation and Environment, 38*(1), 27--52. doi:10.1080/17480930.2023.2243783

Zhan, G. D., Dossary, M. J., Luu, T. P., Xu, H., Furlong, T., & Bomidi, J. (2023). On Field Implementation of Real-Time Bit-Wear Estimation with Bit Agnostic Deep Learning Artificial Intelligence Model Along with Physics-Hybrid Features. *SPE/IADC Middle East Drilling Technology Conference and Exhibition.* doi:10.2118/214603-MS

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into Deep Learning, Interactive deep learning book with code, math, and discussions.* Cambridge University Press.

Zhang, K., Hou, R., Zhang, G., Zhang, G., & Zhang, H. (2016). Rock drillability assessment and lithology classification based on the operating parameters of a drifter: case study in a coal mine in China. *Rock Mechanics and Rock Engineering, 49*, 329–334.

Zhong, R., Johnson, J., & Chen, Z. (2020, 6). Using Machine Learning Methods To Identify Coal Pay Zones from Drilling and Logging-While-Drilling (LWD) Data. *SPE Journal, 25*, 1241-1258. doi:10.2118/198288-PA

Zhou, H., Hatherly, P., Monteiro, S. T., Ramos, F., Oppolzer, F., Nettleton, E., & Scheding, S. (2012). Automatic rock recognition from drilling performance data. *2012 IEEE International Conference on Robotics and Automation*, (pp. 3407–3412).