Centre Eau Terre Environnement

DÉTECTION DE LA MICROSISMICITÉ INDUITE PAR L'EXPLOITATION MINIÈRE PAR APPRENTISSAGE PROFOND

Par

Sepideh Vafaei Shoushtari

Thèse présentée pour l'obtention du grade de
Philosophiæ doctor (Ph.D.)
en sciences de la Terre

Jury d'évaluation

Président du jury et examinateur interne	Mathieu J. Duchesne Commission Géologique du Canada
Examinateur externe	Dany Lauzon Polytechnique de Montréal
Examinateur externe	Adrien Dimech ÉTS Montréal
Directeur de recherche	Bernard Giroux INRS

REMERCIEMENTS

Je tiens d’abord à exprimer ma gratitude à mon superviseur, le professeur Bernard Giroux, pour ses conseils et son soutien tout au long de mes études à l’Institut national de la recherche scientifique. Sa volonté d’investir du temps dans la révision de ma thèse et de fournir une rétroaction inestimable a joué un rôle déterminant dans le développement de mes recherches, ce dont je suis profondément reconnaissant.

Je tiens également à remercier le professeur Erwan Gloaguen pour ses conseils d’expert concernant les discussions sur les réseaux neuronaux dans le cadre de ce projet. Son expertise a amélioré ma compréhension du sujet.

Je suis particulièrement reconnaissante du soutien financier pour cette recherche fourni par le Fonds de recherche du Québec – Nature et technologies (FRQNT) dans le cadre du programme de recherche en partenariat sur le développement durable du secteur minier-II (subvention 2020-MN-286604). De plus, je tiens à souligner le soutien essentiel de Calcul Québec et de l’Alliance canadienne de recherche numérique, qui a grandement facilité mes efforts de recherche. Nous remercions tout particulièrement ESG Solutions, en particulier D. Angus et D. Collins, pour leur générosité dans le partage des données réelles.

Je suis également reconnaissant envers les membres du jury, dont le professeur Dany Lauzon de Polytechnique Montréal, le Dr Mathieu J. Duchesne de la Commission géologique du Canada et le professeur Adrien Dimech de l’École de technologie supérieure. Leurs commentaires constructifs et leurs discussions engageantes ont considérablement enrichi ma thèse de doctorat.

Je remercie mon collègue, le Dr Maher Nasr, dont le soutien dans la simulation de données et les suggestions perspicaces ont été essentiels de ce projet.

Enfin, je tiens à exprimer ma profonde gratitude à ma famille et à mes amis. Vos encouragements et votre soutien indéfectible ont été une source de force tout au long de ce parcours. Je ne peux vraiment pas imaginer passer près de quatre ans à l’étranger et surmonter les différents défis sans votre amour et votre soutien. Pour ma mère et mon père merveilleux, vous avez été mes plus grands piliers de force et de soutien. À ma charmante sœur, je me sens incroyablement chanceuse de t’avoir à mes côtés en tout temps.

Québec, Canada, Été 2025 Sepideh Vafaei Shoushtari

RÉSUMÉ

L’exploitation minière souterraine et l’extraction des ressources minérales sont souvent étroitement liées à des événements microsismiques induits, qui peuvent servir d’indicateurs critiques de la stabilité des masses rocheuses. La surveillance continue de la microsismicité induite par l’exploitation minière à l’aide de réseaux sismiques est une norme de l’industrie pour atténuer les risques et améliorer la sécurité opérationnelle. Pour qu’une telle surveillance soit efficace, il est nécessaire de disposer d’un vaste catalogue d’événements microsismiques, allant d’événements de faible à forte magnitude, afin d’évaluer la réponse de la masse rocheuse à l’exploitation minière. Bien que les événements de forte magnitude soient relativement plus faciles à détecter, l’identification des événements de faible magnitude reste difficile, en particulier dans les environnements miniers où les niveaux de bruit de fond peuvent être importants. Les techniques conventionnelles de détection d’événements, y compris le prélèvement manuel et les approches automatisées telles que la moyenne à court terme/moyenne à long terme (STA/LTA) et le filtrage adapté, sont souvent inappropriés pour ces conditions bruyantes, ce qui conduit à des catalogues d’événements incomplets et à une sous-estimation potentielle des risques sismiques.

Les progrès récents de l’apprentissage profond et des méthodologies basées sur les données ont considérablement amélioré la détection des événements sismiques et la précision et l’efficacité de la sélection de phase. Les réseaux neuronaux convolutifs (CNN) entraînés sur de vastes ensembles de données sismiques ont démontré des performances supérieures dans la détection et la sélection automatiques des arrivées d’ondes P et S, en particulier dans les réseaux sismiques régionaux. Dans cette étude, nous évaluons la performance de PhaseNet, un modèle de sélection du temps d’arrivée basé sur l’apprentissage profond, pour la détection des événements microsismiques induits par l’exploitation minière dans des conditions de bruit variables. Contrairement aux méthodes conventionnelles, PhaseNet est conçu pour être généralisé à différents environnements sismiques et niveau de bruit, ce qui en fait un outil prometteur pour la surveillance microsismique en temps réel.

L’un des principaux défis de l’application de l’apprentissage profond à la sismicité induite par l’exploitation minière est le manque de jeux de données microsismiques de haute qualité, étiquetés, adaptés à l’entraînement des réseaux et à leur évaluation. Pour résoudre ce problème, nous avons généré des sismogrammes à trois composants synthétiques réalistes à l’aide d’une simulation de champ d’ondes élastiques avec la méthode des différences finies dans le domaine du temps (FDTD), en intégrant des conditions géologiques, des modèles de vitesse et une géométrie de réseau représentative d’une mine en Ontario, au Canada. La mine Nickel Rim South (NRS), située dans la ville du Grand Sudbury, qui comprend des exploitations de cuivre et de nickel, connaît fréquemment des événements sismiques induits par l’exploitation minière, dont certains posent des risques importants pour les mineurs et l’infrastructure Le jeu de données simulées comprend une variété de mécanismes de sources sismiques pour refléter les conditions de contrainte complexes dans la mine. Avec plus de 272 000 sismogrammes étiquetés contenant divers niveaux de bruit et magnitudes d’événements, le jeu de données fournit un terrain d’essai solide pour évaluer la capacité de PhaseNet à détecter des signaux faibles que les approches traditionnelles pourraient négliger.

Nos résultats démontrent que PhaseNet peut détecter avec précision les arrivées d’ondes P et S, même en présence d’un bruit important, surpassant ainsi considérablement la méthode conventionnelle (STA/LTA) dans la détection d’événements de faible magnitude. Ces résultats suggèrent que les méthodes basées sur l’apprentissage profond peuvent améliorer la surveillance sismique dans les mines, permettant une meilleure évaluation des risques et des systèmes d’alerte précoce. En plus de tester PhaseNet sur des ensembles de données synthétiques, nous avons également évalué sa performance à l’aide de données sismiques réelles dans divers scénarios. Cette approche globale met en évidence l’applicabilité et l’efficacité du modèle dans diverses conditions, validant davantage son potentiel de surveillance en temps réel dans les environnements miniers. De plus, les connaissances acquises dans le cadre de cette étude contribuent au domaine plus large de la surveillance sismique en démontrant le potentiel des ensembles de données synthétiques pour l’entraînement et la validation de modèles d’apprentissage automatique dans des environnements où les données réelles sont rares. Les travaux futurs se concentreront sur l’optimisation des modèles de prélèvement basés sur l’apprentissage profond pour le déploiement dans des applications minières en temps réel et l’intégration de contraintes géophysiques supplémentaires pour améliorer la robustesse et la généralisation à différents contextes miniers.

Mots-clés: Microsismicité induite par l’exploitation minière, Exploitation minière souterraine, Sismogrammes synthétiques, Moyenne à court terme/Moyenne à long terme (STA/LTA), Domaine temporel à différences finies (FDTD), Simulation de champ d’ondes élastiques, Réseau de neurones convolutifs (CNN), PhaseNet, Arrivées des ondes P et S, Événements de faible magnitude.

ABSTRACT

Underground mining and extraction of mineral resources are often closely linked to induced microseismic events, which can serve as critical indicators of rockmass instability. Continuous monitoring of mining-induced microseismicity using seismic networks is an industry standard to mitigate risks and enhance operational safety. For such monitoring to be effective, an extensive catalogue of microseismic events, spanning low to high-magnitude events, is needed to assess the response of the rock mass to mining. While high-magnitude events are relatively more straightforward to detect, the identification of low-magnitude events remains challenging, particularly in mining environments where background noise levels can be significant. Conventional event detection techniques, including manual picking and automated approaches such as Short-Term Average/Long-Term Average (STA/LTA) and matched filtering, often struggle in these noisy conditions, leading to incomplete event catalogues and potentially underestimating seismic hazards.

Recent advancements in deep learning and data-driven methodologies have significantly enhanced the accuracy and efficiency of seismic event detection and phase picking. Convolutional Neural Networks (CNNs) trained on extensive seismic datasets have demonstrated superior performance in automatically detecting and picking P- and S-wave arrivals, particularly in regional seismic networks. In this study, we evaluate the performance of PhaseNet, a state-of-the-art deep learning-based arrival-time picking model, in detecting mining-induced microseismic events under varying noise conditions. Unlike conventional methods, PhaseNet is designed to generalize across different seismic environments and noise levels, making it a promising tool for real-time microseismic monitoring.

A key challenge in applying deep learning to mining-induced seismicity is the lack of high-quality, labeled microseismic datasets suitable for training and evaluation. To address this, we generate realistic three-component synthetic seismograms using finite-difference time-domain (FDTD) elastic wavefield simulation, incorporating geological conditions, velocity models, and network geometry representative of the NRS mine located in Sudbury, Ontario, Canada. This mine, which includes copper and nickel operations, experiences frequent mining-induced seismic events, some of which pose significant risks to miners and infrastructure. The simulated dataset encompasses a range of seismic source mechanisms that reflect the complex stress conditions within the mine. With over 270,000 labeled seismograms containing diverse noise levels and event magnitudes, the dataset provides a robust testing ground for evaluating the ability of PhaseNet to detect weak signals that traditional approaches might overlook.

Our results demonstrate that PhaseNet can accurately detect P- and S-wave arrivals, even in the presence of substantial noise, and significantly outperforms the conventional method (STA/LTA) in identifying low-magnitude events. These findings suggest that deep learning-based methods can enhance seismic monitoring in mines, thereby enabling improved hazard assessment and the development of early warning systems. In addition to testing PhaseNet on synthetic datasets, we also evaluated its performance using real seismic data across various scenarios. This comprehensive approach highlights the model's applicability and effectiveness in diverse conditions, further validating its potential for real-time monitoring in mining environments. Furthermore, the insights gained from this study contribute to the broader field of seismic monitoring by demonstrating the potential of synthetic datasets for training and validating machine learning models in environments where real data is scarce. Future work will focus on further optimizing deep learning-based picking models for deployment in real-time mining applications and integrating additional geophysical constraints to enhance robustness and generalize them to various mining settings.

Keywords: Mining-induced microseismicity, Underground mining, Synthetic seismograms, Short-Term Average/Long-Term Average (STA/LTA), Finite-difference time-domain (FDTD), Elastic wavefield simulation, Convolutional Neural Network (CNN), PhaseNet, P- and S-wave arrivals, Low-magnitude events.

SOMMAIRE RÉCAPITULATIF

La sismicité induite par l'exploitation minière présente des risques importants pour la sécurité et l'efficacité des opérations minières profondes et à forte contrainte. Cette sismicité est souvent déclenchée par l'exploitation minière souterraine profonde et l'exploitation extensive de carrières en raison des contraintes locales et des conditions géologiques. Ces phénomènes peuvent entraîner des événements dangereux tels que des coups de terrain, qui sont responsables de nombreux décès dans l'exploitation minière en raison de la libération soudaine d'énergie dans les roches. Il est difficile de prédire les coups de terrain, car divers facteurs influencent ces événements, notamment les niveaux de contrainte et le développement de fractures. Malgré la mise en œuvre de systèmes de surveillance sismique pour analyser le comportement des masses rocheuses, il reste difficile d'obtenir une compréhension complète et un cadre théorique universel des mécanismes d'éclatement des roches. Il est particulièrement difficile de prédire avec précision le moment de l'explosion des roches, car les opérations minières s'enfoncent plus profondément dans des environnements sous contrainte, ce qui augmente la nécessité de technologies avancées. La recherche suggère que l'analyse des modèles fractals temporaux des événements microsismiques et de leur distribution spatiale peut aider à identifier les emplacements et les intensités potentiels des coups de terrain, ce qui pourrait permettre des systèmes d'alerte précoce.

La détection et la localisation des événements microsismiques reposent en grande partie sur des mesures précises du temps d'arrivée, un processus qui dépend généralement de l'expertise des géophysiciens. Cependant, le déploiement rapide des sismomètres a rendu de plus en plus difficile pour les analystes de gérer le flux de données, notamment lorsqu'il s'agit d'identifier avec précision les arrivées d'ondes S. Contrairement aux ondes P, qui sont reçues en premier, les ondes S émergent des ondes dispersées de la coda P. La détermination précise des temps d'arrivée des ondes S est essentielle pour traiter les compromis entre la distance et le temps d'origine et améliorer les prévisions de forts mouvements du sol. Bien que diverses techniques de sélection automatique de phase aient été explorées, y compris des méthodes statistiques et des réseaux neuronaux précoces, la méthode STA/LTA, qui est toujours largement utilisée, est sujette à des interférences de bruit et à des imprécisions de synchronisation, en particulier pour les ondes S.

L'intérêt croissant pour les techniques d'apprentissage automatique (ML) et d'apprentissage profond (DL) en sismologie découle de leur aptitude à gérer les subtilités des données sismiques.

Ces méthodes modernes surpassent les approches traditionnelles de sélection automatique de phase en déduisant des relations non linéaires directement à partir des données, éliminant ainsi le besoin de modèles explicites de processus complexes. Les techniques d'apprentissage automatique (ML) et d'apprentissage profond (DL) apparaissent comme des outils puissants pour détecter et classer les événements sismiques, y compris les cas de faible magnitude que les méthodes traditionnelles négligent fréquemment. Les modèles sophistiqués tels que les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux récurrents (RNN) sont particulièrement efficaces pour analyser les données sismiques brutes, offrant une précision exceptionnelle dans la surveillance des événements en temps réel et l'identification des événements en raison de leur conception complexe et stratifiée.

Dans cette thèse, le réseau d'apprentissage profond PhaseNet est utilisé pour prédire les temps d'arrivée des ondes P et S à l'aide de sismogrammes synthétiques qui simulent les conditions réelles d'une mine sous différents rapports signal/bruit (SNR). PhaseNet fait preuve d'une précision et d'un taux de rappel supérieurs à ceux des méthodes de prélèvement STA/LTA traditionnelles. Le modèle traite les formes d'onde non filtrées et prédit les distributions de probabilité pour les ondes P, les ondes S et le bruit à l'aide d'un ensemble d'apprentissage dérivé de la modélisation élastique dans le domaine temporel des différences finies (FDTD). La recherche porte sur une mine souterraine (NRS) de roche dure près de Sudbury, en Ontario, qui est exploitée entre 1100 et 1720 mètres de profondeur. Ce site présente divers types de roches et une composition géologique complexe, ce qui rend la surveillance sismique difficile. Équipée d'un système de surveillance microsismique comportant des géophones et des accéléromètres, la mine effectue une analyse détaillée des événements sismiques.

Un sous-domaine rectangulaire de la mine spécifique est considéré pour préparer les sismogrammes synthétiques pour l'entraînement du modèle d'apprentissage profond. Dans ce sous-domaine, la vitesse de l'onde P varie de 1700 m/s à plus de 6500 m/s. Le modèle de vitesse de l'onde S montre des valeurs allant de 700 m/s à 3600 m/s. La simulation utilise des informations réelles recueillies dans la mine étudiée, en intégrant des modèles de vitesse et la configuration spécifique des récepteurs dans le système de surveillance. Les paramètres des sources sismiques, y compris les emplacements, les signatures temporelles et les mécanismes focaux, sont principalement conjecturaux, mais soigneusement sélectionnés à des fins de modélisation pour reproduire les événements microsismiques réels. Les temps d'arrivée des ondes P et S sont calculées à l'aide d'un logiciel de tracé de rais. Les différentes combinaisons de paramètres de la source ont donné un total de 1 700 événements, obtenus par simulation de forme d’onde via le logiciel SeisCL. Les mécanismes au foyer contiennent 600 sources à force unique, 500 avec un mécanisme à double couple (DC), 300 avec un mécanisme isotrope, et 300 avec un mécanisme de dipôle vectoriel linéaire compensé (CLVD).

En plus des 65 capteurs réels du système de surveillance microsismique dans la mine, 95 capteurs supplémentaires sont envisagés pour augmenter la densité des récepteurs pour la simulation du champ d’ondes. Tous les capteurs sont supposés avoir trois composants. De cette façon, le champ d’ondes modélisé est mieux échantillonné. Les résultats démontrent un alignement acceptable entre les sismogrammes simulés et les temps de parcours calculés par tracé de rais. De plus, les sismogrammes synthétiques affichent une gamme de diffractions, de réfractions et d’interférences d’ondes, ce qui ajoute encore plus de complexité à l’ensemble de données et le rend plus réaliste. L’ensemble de données synthétiques contient un nombre total de 272 000 sismogrammes à trois composants. Les enregistrements synthétiques sont stockés dans un format SEGY. Chaque fichier contient l’une des trois composantes du champ d’ondes modélisé (E, N ou Z). Comme mentionné précédemment, une approche de tracé de rais est utilisée pour étiqueter les formes d’onde modélisées et calculer les temps de parcours des ondes P et S pour toutes les paires source-récepteur dans le sous-domaine proposé. L’implémentation de la méthode du chemin le plus court (SPM) dans le package ttcrpy est utilisée pour le calcul du temps de trajet, car cette méthode peut gérer les fortes variations de vitesse contenues dans les modèles P et S de la mine.

Comme mentionné précédemment, le réseau qui va être utilisé pour l’entraînement du modèle DL à des fins de prédiction d’arrivée est PhaseNet. Il est un réseau de neurones profonds développé à l’origine pour la surveillance des tremblements de terre, maintenant évalué pour des applications minières. PhaseNet repose sur une architecture U-Net dotée d’un cadre d’encodeur-décodeur avec des sections de contraction et d’expansion. L’encodeur capture les caractéristiques à l’aide de modules convolutifs et de sous-échantillonnage, tandis que le décodeur suréchantillonne les données à l’aide de modules déconvolutifs et de concaténation. Adapté aux formes d’ondes sismiques, PhaseNet traite les entrées unidimensionnelles des données sismiques à trois composantes (Z, N, E), produisant des canaux de sortie qui indiquent les probabilités d’arrivée P, d’arrivée S ou de bruit. Notamment, aucun filtrage de fréquence n’est appliqué aux formes d’onde d’entrée, et les performances du réseau sont testées avec des sismogrammes synthétiques contenant différents niveaux de bruit, où les pics de distributions de probabilité désignent les temps d’arrivée des ondes P et S. Lorsque des données microsismiques sont traitées dans PhaseNet, elles passent par une série d’étapes qui utilisent des opérations de convolution 1D et des fonctions d'activation ReLU. Les connexions par saut permettent la concaténation directe des sorties, ce qui améliore la convergence de l'entraînement. La convolution utilise une taille de sept points de données, avec une foulée de sous-échantillonnage de quatre, réduisant la longueur du canal à un quart. Une déconvolution par suréchantillonnage rétablit la longueur d'un facteur quatre, avec un remplissage pour maintenir la longueur de la séquence. La dernière couche applique la fonction softmax pour transformer la sortie du décodeur en probabilités de classification multiclasse.

L'évaluation de PhaseNet indique des performances élevées et constantes à différents niveaux de bruit, avec une précision approchant les 0,9, en particulier pour la détection des ondes S. La comparaison avec la méthode STA/LTA révèle les avantages de PhaseNet dans les environnements bruyants et sa capacité à caractériser efficacement les événements sismiques de faible magnitude. Des études de généralisation ultérieures démontrent que l'entraînement sur des données synthétiques et réelles renforce considérablement la précision et la robustesse des modèles dans des environnements miniers complexes. Dans l'ensemble, les résultats soulignent l'importance d'intégrer des données synthétiques et réelles pour améliorer la détection et l'analyse des événements sismiques.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABRÉVIATIONS

1C	One-Component
1D	One-Dimensional
3C	Three-Component
3D	Three-Dimensional
3DVM	Three-Dimensional Velocity Model
ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Auto-Regression
AR-AIC	Autoregressive Akaike Information Criterion
CAE	Convolutional Auto Encoder
CapsNet	Capsule Network
CLVD	Compensated Linear Vector Dipole
CNN	Convolutional Neural Network
ConvNet	Convolutional Network
CPML	Convolutional Perfectly Matched Layer
CPU	Central Processing Unit
CT	Contact Zone
CT II	Secondary Contact Zone
DC	Double Couple
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DSPM	Dynamic Shortest-Path Method
ER	Energy Ratio
FCN	Fully Convolutional Network

FD	Finite Difference
FDTD	Finite Difference Time Domain
FGN	Felsic Gneiss Unit
FM	Fast Marching
FN	False Negative
FP	False Positive
FSM	Fast-Sweeping Method
FW	Footwall Zone
FWB	Feature Weighted Beamforming
FWI	Full Waveform Inversion
FWM	Full Waveform Modeling
GAN	Generative Adversarial Network
GD	Gradient Descent
GPR	Ground Penetrating Radar
GPU	Graphics Processing Unit
GSI	Geological Strength Index
GR	Gutenberg Richter
HSF	Hyperion Seismic Format
LSTM	Long Short-Term Memory
MCM	Modified Coppen’s Method
MER	Modified Energy Ratio
ML	Machine Learning
MLM	Maximum Likelihood Method
MS	Monitoring System
MSE	Mean Squared Error
MTI	Moment Tensor Inversion
Mw	Moment Magnitude
NN	Neural Network
PDF	Probability Density Function
PML	Perfectly Matched Layer
ReLU	Rectified Linear Unit

RF	Random Forest
RMSProp	Root Mean Squared Propagation
RNN	Recurrent Neural Network
RT	Ray Tracing
RTM	Reverse Time Migration
SDBX	Sudbury Breccia Zone
SF	Single Force
SGD	Stochastic Gradient Descent
SGM	Strong Ground Motion
SNR	Signal-to-Noise Ratio
SPM	Shortest-Path Method
SPMD	Single-Program Multiple-Data
SSG	Standard Staggered Grid
STA/LTA	Short-Term Average to Long-Term Average Ratio
STF	Source Time Function
SVM	Support Vector Machine
TL	Transfer Learning
TN	True Negative
TP	True Positive
WGN	White Gaussian Noise

1 INTRODUCTION

1.1 Background

Seismic activity can greatly affect both the safety and efficiency of underground mining operations. In particular, the occurrence of rockbursts is the most serious and least understood problem in deep mining operations worldwide (Gibowicz et al., 1994). As mining operations become deeper and more extensive, the associated problems tend to escalate. It is, therefore, not surprising that the monitoring of seismic events triggered by mining has been ongoing for a long time. The approaches and techniques used to analyze mining-induced seismicity have largely been adapted from those used in earthquake seismology. While no consistent distinctions have been identified between major mine tremors and natural earthquakes so far, there are notable differences in how these two types of seismicity are observed. In mines, seismic networks are typically three-dimensional, with measurements taken directly from the source area (Gibowicz et al., 1994).

1.1.1 Induced Seismicity

This section begins with the definition of several key terms. The term "induced seismicity" refers to seismic events that occur as a result of changes in pore pressure and stress caused by human activities. "Potentially induced seismicity" refers to seismic events that occur concurrently with human activities. "Microseismicity" refers to small-scale seismic events directly associated with controlled deformation processes, typically having magnitudes between $-5 \leq M_w \leq 0$. A "tectonic earthquake" is characterized by the intense shaking of the Earth's surface due to the sudden release of energy and movement within the Earth's crust, often associated with natural occurrences. Additionally, in seismology, the term "trigger" refers to the initiation of a process that may ultimately lead to an earthquake. Triggered earthquakes can occur after a delay that varies from seconds to years following the initial event. The moment magnitude symbol, M_W, also refers to the energy released during a seismic event or earthquake, and M_L indicates local magnitude according to the Richter scale.

Documented instances of felt seismic events caused by human activities date back to 1894 in Johannesburg (McDonald, 1982) and were subsequently linked to gold production in the Witwatersrand region, which began in 1886 (Cook et al., 1966). Since then, numerous cases of induced seismicity have been recognized and are believed to be related to underground operations. Some of these cases are illustrated in Figure 1-1.

- Mining-induced seismicity (Gibowicz et al., 1994; Hudyma, 2004)
- Reservoir-induced seismicity (Gupta, 2002; Simpson et al., 1988)
- Hydraulic fracturing (Ellsworth, 2013; Zoback et al., 2012)
- Wastewater injection (Keranen et al., 2014; Rubinstein et al., 2015)
- Geothermal energy extraction (Häring et al., 2008; Majer et al., 2007)
- Carbon capture and storage (CCS)(Nicol et al., 2011; Zoback et al., 2012)
- Oil and gas extraction (Segall, 1989; van Thienen-Visser et al., 2015)
- Controlled explosions, as in nuclear waste disposal sites (Mendecki, 1996; Richards et al., 2007)
- Underground storage and withdrawal (Evans et al., 2009)
- Groundwater extraction (Holzer et al., 2005; Pacheco-Martínez et al., 2013)

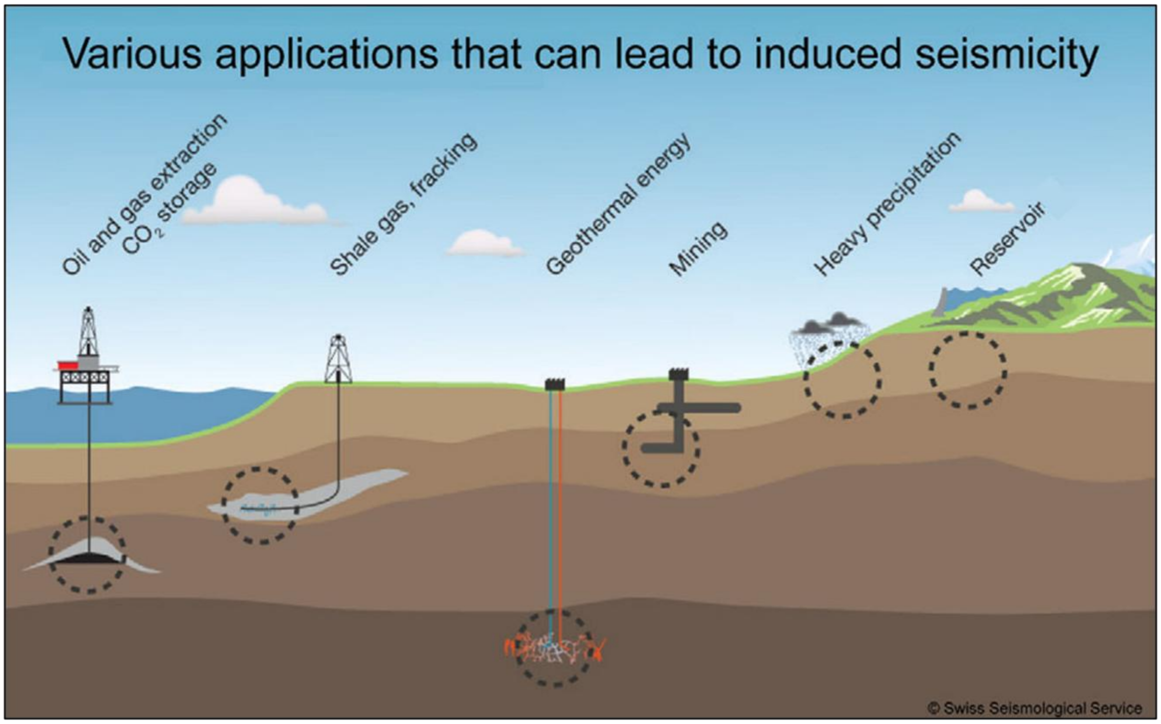


Figure 1-1: Examples of anthropogenic activities that trigger seismic events, as well as various sources contributing to these occurrences. Image modified from the Swiss Seismological Service website (SED, n.d.).

Seismic monitoring installations for mines first emerged in Europe in the late 1920s, particularly in Poland's Upper Silesia coal basin (Gibowicz et al., 1994). Since then, seismic monitoring systems have been regularly implemented for various purposes at local and regional levels. Regional seismic networks are predominantly utilized to locate earthquakes, primarily focusing on detecting tectonic events within the upper 100 kilometers of the crust. In contrast, local networks serve to identify hazards and support the development of reservoirs. However, integrating these two systems can enhance the accuracy of seismic event detection and localization and help clarify the causes of induced, potentially induced, and triggered seismicity related to faults.

Seismicity associated with underground mining is likely the most adverse phenomenon among the various types of triggered earthquakes, affecting both mining safety and productivity. Rockbursts are very often the primary cause of fatalities in mines. The problem becomes progressively more severe as the average depth and the extent of mining operations increase.

1.1.2 Seismicity in Underground Mining

Mining-induced seismicity poses a significant risk to deep and high-stress mining operations. Seismic events arise from a blend of local stress factors and geological conditions, influenced by various failure mechanisms. Large seismic occurrences (MR ≥ 1) can trigger rockbursts, threatening the safety of underground excavations, equipment, and personnel. By utilizing seismic monitoring systems, mining operations can gain valuable insights into the conditions and failure processes affecting the local rock mass. Seismic source parameters obtained from triaxial sensors allow for a quantitative description of the seismic population. These parameters encompass time, location, energy, moment, magnitude, and apparent stress. Different analytical techniques can uncover significant trends in these seismic source parameters. Trends associated with the source mechanism, local stress levels, and event frequency can help assess seismic hazards and aid in the creation of seismic hazard maps (Brown, 2015).

Figure 1-2 shows a typical configuration of a microseismic monitoring system (Collins et al., 2014). These monitoring systems typically comprise a network of a few dozen acquisition units, where the signals received by the seismic sensors are digitized and processed. Acquisition units typically include 6 channels, allowing them to process either 2 triaxial sensors and 6 uniaxial sensors or 1 triaxial sensor and 3 uniaxial sensors. A system can thus include several dozen sensors. Depending on the chosen configuration, the acquisition units continuously transmit data in real-time to a central unit or in segments if an event is detected at the acquisition unit. The central unit is typically a powerful PC equipped with a graphics processing unit (GPU) and a substantial storage capacity.

During the phenomenon of mining-induced seismicity, elastic energy is released along failure planes in the rocks or through sudden slips on pre-existing faults and fractures. This release is often related to the weight and stress of the overburden, propagating in the form of compressional

(P-wave) and shear (S-wave) waves. The arrival times of these P- and S-waves are detected by the sensors placed in the acquisition units. Figure 1-3 illustrates the P- and S-wave arrivals of a single microseismic real event recorded by a three-component geophone located in the monitoring system of the mine under study in this thesis.

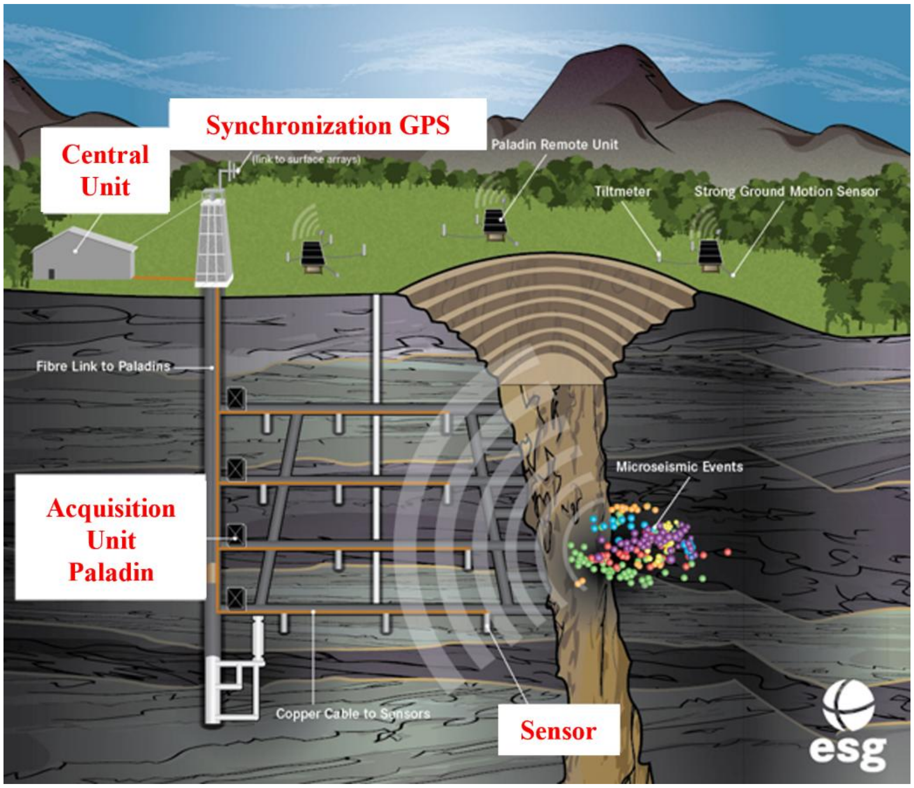


Figure 1-2: Schematic diagram of a typical microseismic monitoring configuration in an underground mine, courtesy of ESG Solutions (Collins et al., 2014).

The effective detection and localization of microseismic events largely depend on the quality and accuracy of arrival-time measurements. This phase-picking process, which involves measuring seismic arrival times, is primarily carried out by network analysts. These professionals leverage their expertise and extensive experience to conduct phase-picking procedures. However, the rapid proliferation of seismometers presents challenges in managing the overwhelming volume of data generated. Moreover, phase pickers encounter difficulties when identifying S-wave arrivals, as these waves often emerge from the dispersed waves of the P coda. Accurately determining the arrival times of S-waves is vital, as S-waves help mitigate the trade-off in depth origin that can occur when relying solely on P-wave arrivals for locating seismic events. Furthermore, a thorough understanding of S-wave structure is critical for precise predictions about strong ground motions.

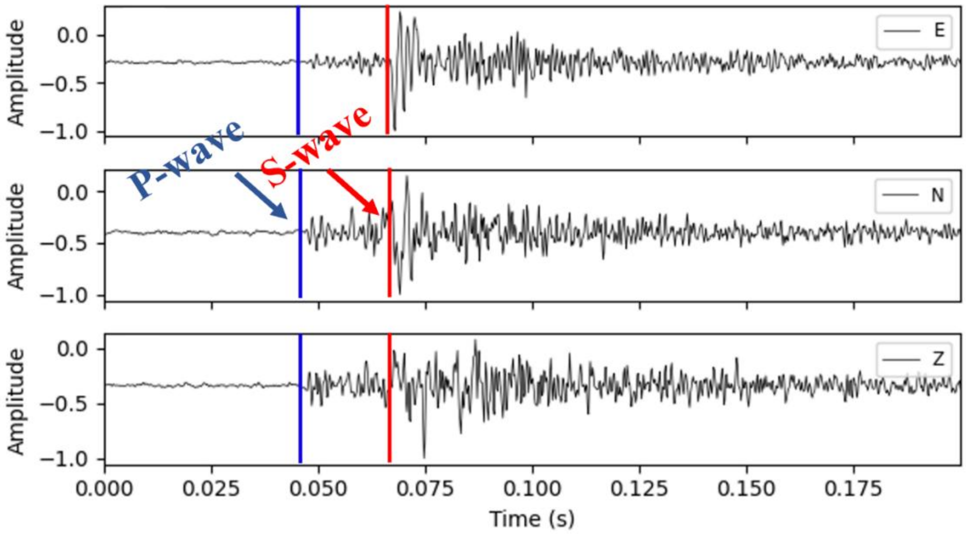


Figure 1-3: The waveforms of a real single microseismic event recorded by a three-component (3C) geophone at the proposed mine site of this study.

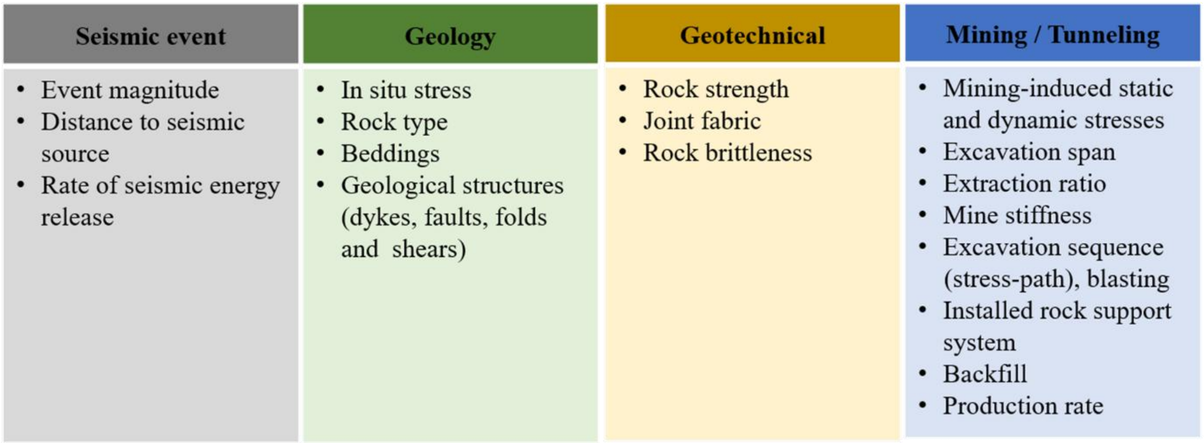
1.1.3 Rockburst

The difference between a seismic event in a mine, commonly known as a "mine tremor", and a rockburst is significant. Rockbursts are sudden and severe failures of rock that cause damage to underground structures (Dmowska et al., 1999). Therefore, only the seismic events that lead to damage in areas of the mine that are not easily accessible are classified as "rockbursts". While this distinction was created by mining engineers and may seem somewhat arbitrary, it holds substantial practical relevance. Out of the thousands of seismic events documented each year by monitoring networks in various mines, only a small number escalate into rockbursts. Rockbursts are merely a minor portion of the broader category of seismic events triggered by mining operations, and there are no definitive indicators to determine which specific seismic event will culminate in a rockburst (Gibowicz et al., 1994). Figure 1-4 shows some rockburst occurrences during the underground mining processes.

Figure 1-4: Examples of rockburst damages and risks in underground mining. (1) Hecla's Greens Creek Mine, (2) rockburst resulting from the failure of a pillar in a hard rock mine in Canada, (3) rockburst in a deep mine in the Western United States, and (4) collapse of the fill stope in a Canadian Mine.

While the primary factors influencing the occurrence and intensity of rockbursts have been generally recognized, it remains challenging to clarify and standardize the mechanisms behind rockbursts across various geological settings and excavation impacts. Researchers are working to better understand the conditions needed for rockbursts to form and their development processes by analyzing a range of engineering cases and performing physical model experiments (Wu et al., 2022). Kaiser et al. (2012) summarized the main factors influencing rockburst damage and its severity, as shown in Figure 1-5.

In situ microseismic monitoring reveals that most monitored rockbursts exhibit microseismic precursors (Xiao et al., 2016). When a rockburst occurs, the accumulated energy and the number of microseismic events reach a maximum (Ma et al., 2015). Using the energy rate, moment tensor analysis, and the P-wave development method based on actual microseismic information, supplemented by in situ macro failure characteristics of rockbursts and the identification of rock fracture modes (tension, mixing, and shear), is a promising way to interpret the rockburst mechanism (Feng et al., 2017). Zhang, Zhang, et al. (2020) revealed the timing patterns of rockbursts by analyzing microseismic data collected during a delayed rockburst and a uniaxial dead load test. The time behaviour of rockburst is the time behaviour of crack propagation under high stress. When the internal damage of rock mass accumulates to a critical stage, even if there is no obvious excavation disturbance, rock mass may still be unstable (Zhang, Ma, et al., 2020). Based on the thin plate theory, Yang et al. (2019) investigated the occurrence mechanism of rockbursts during mining of an irregular working face through microseismic monitoring, theoretical analysis, numerical simulation, and field measurement of working resistance and support. The results showed that the combined dynamic and static loads were the main factors inducing rockbursts. Although considerable research has been conducted on rockburst mechanisms, a unified theory has yet to be established. The complex and diverse influencing factors, along with the unclear nonlinear relationship, make it challenging to explain the formation mechanisms of rockbursts in different environments, and it is arduous to establish a universal mechanical theory for rockbursts (Waqar et al., 2023; Zhou et al., 2024).



Seismic event	Geology	Geotechnical	Mining / Tunneling
• Event magnitude • Distance to seismic source • Rate of seismic energy release	• In situ stress • Rock type • Beddings • Geological structures (dykes, faults, folds and shears)	• Rock strength • Joint fabric • Rock brittleness	• Mining-induced static and dynamic stresses • Excavation span • Extraction ratio • Mine stiffness • Excavation sequence (stress-path), blasting • Installed rock support system • Backfill • Production rate

Figure 1-5: The main factors influencing rockburst damage and severity (Kaiser et al., 2012).

Accurately predicting the timing of rockbursts is a critical yet unsolved challenge. As mining operations extend into deeper and more stressed conditions, the demand for advanced rockburst prevention technologies increases. Multiple studies indicate that the temporal fractal behaviour of microseismic event energies and the spatial distribution of these events could serve as indicators for potential rockburst locations and severity, offering a path toward early warning systems (Feng et al., 2015; Gibowicz et al., 1994; Kisslinger, 1976; Lama et al., 1998).

As mining progresses, the natural stress equilibrium of the rock mass is disrupted, leading to the accumulation of potential energy that can be released in the form of seismic events. Rockbursts can be classified based on their seismic triggering mechanism into two main types: those caused by mining activities and those triggered by distant seismic events. The likelihood of rockbursts occurring is influenced by several key factors. These include unfavorable stress conditions, such as the magnitude and alignment of principal stresses in relation to existing geological features or excavation shapes. Additionally, the characteristics of the rock mass are significant, particularly when there are varying geomechanical properties or geological irregularities, such as fractures. The intensity of the stresses that build up around underground voids also plays a crucial role as excavation continues (Keneti et al., 2018). Various approaches can be adopted to mitigate the risks related to mine seismic activity and rockbursts. Long-term monitoring focuses on comprehending alterations within the mine, which helps in adjusting extraction techniques and enhancing mine design, ultimately minimizing the risks linked to rockbursts. Some short-term measures include installing ground support systems to reinforce the structures, absorbing the kinetic energy released (Keneti et al., 2018) or using remotely operated equipment. For all approaches, understanding the mine seismicity is crucial; therefore, microseismic monitoring is the primary method for mitigating hazards associated with underground mining.

1.2 Literature Review

Determining arrival times is a crucial initial phase in processing microseismic data, essential for phase identification and subsequent analyses, such as accurately locating a microseismic hypocenter. Mistakes or misclassifications in these arrival times can significantly impact the calculated hypocenters and the overall interpretation of the findings, especially in cases where there is a consistent bias (Maxwell et al., 2010; Tan et al., 2016). As the volumes of microseismic data continue to grow, manually determining arrival times promptly becomes impractical and time-consuming, necessitating the use of automated methods. A range of automated algorithms has been created for detecting arrival times, applicable for time or frequency domain and suitable for both single-component and multi-component microseismic data (Akram et al., 2016). These methods have been developed over time and can be categorized into three main approaches:

1. **Arrival Picking** – Identifying arrival times of seismic P- and S-waves.
2. **Waveform Classification** – Using pattern recognition to distinguish seismic events from noise.
3. **Migration-Based Methods** – Back-projecting seismic waveforms to locate sources.

Each category offers unique advantages and limitations depending on data quality, network geometry, and computational efficiency. The following sections review the key techniques within each category.

Arrival Picking

This procedure determines the precise times at which different types of seismic waves arrive at a seismic station. Typically, three main types of waves are recorded during an earthquake: Primary (P) waves, Secondary (S) waves, and surface waves. Each wave travels at a different velocity and has distinct characteristics. In underground mining applications, surface waves are not encountered. By accurately identifying the arrival times of these waves, seismologists can calculate the distance to the source, estimate its magnitude, and provide timely information about the event. Various techniques, including visual inspection of seismograms and automated algorithms, are employed to enhance the accuracy of arrival picking, enabling a rapid response to seismic events.

Automated seismic arrival pickers are algorithms designed to detect the onset timing, identify phase types, and characterize the arrivals of seismic energy on seismogram waveforms. There is a substantial history of algorithms dedicated to automated seismic arrival detection, which effectively utilize variations in amplitude, frequency, standard deviation, polarization, and statistical approaches of seismogram waveforms (Akram et al., 2016; Allen, 1978; Baer et al., 1987; Bai et al., 2000; McEvilly et al., 1982; Withers et al., 1998). These automated pickers, often referred to as "classical" pickers, are primarily focused on rule-based algorithms developed by experts who utilize seismological knowledge to process seismogram waveforms and conduct subsequent analyses for detection, onset timing, and phase identification. The parameters of these pickers are often set and fine-tuned through trial and error or regular optimization methods to yield results that closely align with manually determined or other reference picks. Additionally, the outputs from classical pickers can be further utilized as inputs into deep neural networks to enhance the accuracy of the arrival timing and characterize the picks (Lomax et al., 2024; Yeck et al., 2021).

The majority of classical automated pickers start by transforming raw seismograms into characteristic function (CF) time series, which significantly enhance the key features of seismic phase arrivals, such as sudden shifts in amplitude, frequency, or polarization of the waveforms. Next, these pickers examine the CFs to identify and determine onset times, phase types, and other potential seismic energy arrival attributes while disregarding background signal fluctuations. The conversion of raw seismograms into picker CFs typically involves basic processes, such as mean removal and high-pass filtering, followed by calculations using a sliding window. Some of these commonly used classical approaches are described in the following paragraphs.

The short-term average to long-term average ratio (STA/LTA) algorithm, which computes the energy ratio between short-term and long-term windows, has been used widely and has gained widespread acceptance for its efficiency and effectiveness in identifying P- or S-wave arrivals (Allen, 1978; Baer et al., 1987; Stevenson, 1976). The process begins by determining the ratio of absolute amplitudes, energies, or the envelope function of the seismic trace across two distinct time windows, resulting in a CF curve. The arrival time is identified when the CF value surpasses a specified threshold set by the user. However, the STA/LTA method typically struggles with signals that exhibit a low Signal-to-Noise Ratio (SNR). This method can effectively detect microseismic events with a high SNR and accurately determine their arrival times. In contrast, it is less reliable in accurately picking the arrival times of low-SNR events. Additionally, the STA/LTA algorithm may produce incorrect arrival times when the data contains noise events with sudden increases in amplitude. This issue is pervasive in surface and mining microseismic monitoring, where various noise sources showing abrupt amplitude spikes can resemble real microseismic events. Examples include human footsteps, moving vehicles, and explosions. Many noise sources cannot be filtered out using standard bandpass filtering, a common pre-processing technique for raw microseismic data. Due to the waveform similarities between actual microseismic events and unwanted noise, these noise events often trigger the detection threshold, leading to their misidentification as microseismic events when the STA/LTA method is applied (Chen, 2020; Vaezi et al., 2015).

To overcome these flaws, the commonly used STA/LTA method was improved by incorporating the signal's envelope as a principal feature, which enhanced the identification of seismic phase arrivals (Baer et al., 1987). Instead of directly applying STA/LTA to the raw waveform, the algorithm first computes the signal envelope using the Hilbert transform. The envelope represents the amplitude variations of the waveform over time, smoothing out rapid oscillations and making arrivals more distinguishable from background noise. This adjustment minimized the likelihood of false detections due to noise variations and increased sensitivity to faint signals, making it particularly useful in noisy environments, such as those associated with mining-induced seismicity. The approach provided a more reliable determination of the onset of seismic waves, thereby enhancing the precision of automatic phase detection compared to the traditional STA/LTA method.

The AR-AIC algorithm, which combines autoregressive (AR) noise and signal modeling with the Akaike Information Criterion (AIC), was introduced to precisely detect the onsets of P- and S-waves in seismic signals. The algorithm generates the CF curve by calculating the AIC, which assesses the fit of joint autoregressive modeling for two stationary signal sequences across different time intervals. A microseismic event is identified when the AIC value reaches its lowest point (Leonard, 2000; Sleeman et al., 1999). The AR-AIC approach is based on various factors that differentiate seismic signals from ambient noise through specific characteristic functions. However, it struggles to provide reliable arrival picks when the underlying contrast is minimal, especially in conditions of low signal-to-noise ratios (Leonard et al., 1999).

Matched filtering, also known as template matching, is a highly sensitive technique to detect weak or hidden seismic events by cross-correlating continuous waveforms with known event templates. This method is particularly effective in identifying repeating seismic events, such as aftershocks, mining-induced microseismicity, or tremor-like signals, even in noisy environments. Using array-based waveform cross-correlation, a reference event is compared with continuous recordings to identify events that share significant similarities. The cross-correlation values generated from this process create a CF curve, highlighting potential events through marked increases in correlation. (Gibbons et al., 2006). This approach, when combined with autocorrelation, addresses the shortcomings of the STA/LTA method by utilizing information from the complete microseismic waveform throughout the geophone array for detection. This technique begins by selecting a master event, an exemplary waveform that serves as a template for comparison against continuous recordings from a seismic array. The master event, ideally representative of the expected signal characteristics, undergoes preprocessing that includes filtering and normalization. The core of the method involves cross-correlating the master waveform with real-time data from individual geophones, allowing for the measurement of similarity at each time step. Autocorrelation can enhance this process by identifying repeating waveforms within a single channel, particularly when a suitable master event is not available. The results from individual sensors are then stacked to form a coherent signal, boosting detection confidence. However, the effectiveness of matched filtering is contingent upon the quality of the master event; significant differences in source mechanism, location, or propagation path can hinder detection capabilities, indicating the method's limitations in varying seismic contexts.

Furthermore, progress has also been made using higher-order statistical measures, such as kurtosis and skewness, to identify seismic event onsets by detecting transitions from Gaussian to non-Gaussian distributions, thus improving detection amidst noise (Saragiotis et al., 2002). While most algorithms have concentrated on P-waves, Ross et al. (2016) have made strides in differentiating P- and S-waves through polarization analysis, enhancing the accuracy of S-wave detection.

Despite all the advancements in classical approaches, the accuracy of automated phase picking still does not match expert human analysis. This discrepancy largely stems from the inherent complexity of waveforms, which are influenced by several factors, including focal mechanisms, stress drop, scattering, phase transformations, and the prevalence of noise. Classical methods are limited in scope, relying on manually selected features and necessitating careful data preprocessing, such as bandpass filtering and threshold setting. In contrast, neural networks have the potential to autonomously identify relevant features from labeled data, capturing both noise and signal and thus offering a substantial improvement in analyzing complex seismic waveforms.

Recent progress in machine learning (ML) and deep learning (DL) techniques (Hinton et al., 2006) have improved the efficiency of automated seismology processes, leading to the creation of supervised and unsupervised ML algorithms, as well as Deep Neural Networks (DNNs), which can be designed to automate the detection of microseismic events. This improvement is due to the ability of DNNs to adopt more advanced inference techniques that mimic human cognitive processes (LeCun et al., 2015).

The increasing preference for ML and DL techniques in seismology, particularly for tasks such as phase-picking and detecting P- and S-wave arrivals, stems from their proven ability to handle the intricacies of seismic data effectively (Meier et al., 2019). Automated seismic arrival detection algorithms based on ML techniques have been developed over several years (Beyreuther et al., 2012; Gentili et al., 2006; Kong et al., 2019; Mousset et al., 1996). The early application of machine learning methods for automating the identification of microseismic events involved a non-linear and multilayer network, which has been utilized in various studies for event detection and arrival time picking (Dai et al., 1995, 1997; Gentili et al., 2006; Maity et al., 2014). In a recent study, a neural network-based method was introduced to utilize the probability values generated by the neural network as the cumulative frequency (CF) curve to identify events and determine arrival times (Akram et al., 2017). The distinction between early ML approaches and recent DL methods primarily lies in the complexity and capability of the models used. While early ML techniques typically employed simpler, non-linear, and multilayer networks, recent DL approaches leverage more complex architectures. This shift allows for automatic learning of hierarchical features from raw data, eliminating the extensive manual feature extraction often required in traditional ML. Additionally, DL methods excel at handling large datasets and can improve accuracy with vast amounts of labeled data, while also demonstrating enhanced robustness in low SNR and complex waveforms. Moreover, DL models tend to generalize better to unseen data, increasing their prediction capabilities across various seismic detection contexts. Overall, this transition signifies a move towards more sophisticated, automated, and effective methodologies in seismic arrival detection.

Integrating DNN algorithms for arrival picking has gained significant attention in research within the last few years, as these methods push the boundaries of traditional approaches by leveraging larger datasets, improved computational power, and advanced learning techniques. By incorporating more comprehensive and diverse datasets, DNNs can better capture complex patterns and nuances in arrival times, which traditional methods may overlook. Additionally, innovations in architectures, such as convolutional neural networks (CNNs) for spatial data or recurrent neural networks (RNNs) for sequential data, enable DNNs to improve accuracy and responsiveness. Researchers are also focusing on hyperparameter tuning, regularization techniques, and ensemble methods to further refine these models, ultimately leading to enhanced performance in arrival time predictions. Three primary neural network architectures have emerged as foundational for deep learning methodologies in determining arrival times, with numerous studies aimed at enhancing DNN performance through model refinement and the combination of established frameworks.

One prominent approach was proposed by Zheng et al. (2018), who employed a RNN utilizing long short-term memory (LSTM) units. Their architecture included seven hidden layers, specifically designed to effectively capture the arrival times of microseismic and acoustic emission events. Meanwhile, Ross et al. (2018) introduced a CNN composed of two key components with specific goals in mind. The first component, focused on feature extraction, aims to automatically identify and extract relevant patterns or features from the input data using CNN techniques. This is crucial for capturing the essential characteristics of the data that will feed into the model. The second component, which serves as a regressor with two fully connected layers, is intended to take the extracted features and utilize them to predict continuous output values. Essentially, the two components work together to enhance the model's ability to learn from the data effectively, improving its predictive performance.

Advancing from this work, Zhu et al. (2019) developed a novel variant of CNN known as PhaseNet, which was inspired by the fully convolutional network (FCN) framework introduced by Long et al. (2015). PhaseNet aimed to enhance the accuracy of automatic P- and S-wave arrival detection by replacing the fully connected layers in the regressor of Ross et al. (2018) with fully convolutional layers. PhaseNet gives probabilities of P- and S-wave components as output, and its architecture was also adapted from the classic U-net structure presented by Ronneberger et al. (2015), utilizing an encoder-decoder methodology to merge high-resolution features from the contracting path with the upsampled output. PhaseNet offers significant advantages by providing probabilistic outputs for seismic phase detection, which enhances the overall reliability of the results. The inclusion of probabilities allows for a quantification of uncertainty, giving users valuable insights into the confidence levels associated with each phase detection. This is particularly beneficial in seismology, where data quality can fluctuate, and low-magnitude events may easily be overlooked by traditional methods. Moreover, the probabilistic nature of the outputs supports adaptive thresholding, enabling better phase-picking accuracy in noisy environments. Additionally, these probabilities facilitate the integration of PhaseNet with other algorithms through ensemble methods, further improving detection reliability. Overall, the ability to provide probabilities not only enhances the precision of seismic event identification but also supports more informed decision-making and continuous model improvement over time.

In a further advancement, a comprehensive deep convolutional neural network (ConvNet) model is introduced, specifically designed to distinguish between earthquake signals and background noise (Dokht et al., 2019). This multilayer ConvNet was trained to recognize the general characteristics of noise and seismic signals in the time-frequency domain. Complementing this, an auxiliary network is employed in ConvNet, utilizing wavelet transformations, to differentiate between P- and S-waves while estimating their arrival times. The ConvNet method offers several advantages for distinguishing between earthquake signals and background noise. Its multilayer architecture allows for high accuracy in recognizing complex patterns in seismic data, particularly through time-frequency analysis that enhances the identification of relevant characteristics. Additionally, ConvNet enables automation in the detection and classification process, reducing reliance on human intervention and increasing processing speed. The integration of wavelet transformations in an auxiliary network further improves the model’s capability to differentiate between P- and S-waves, enhancing the precision of arrival time estimations. However, there are notable drawbacks as well. Training ConvNet requires substantial amounts of labeled data, which can be challenging to gather, and there is a risk of overfitting without proper regularization.

Furthermore, the computational intensity of these models may limit their applicability in real-time scenarios unless optimized.

EQTransformer (Mousavi et al., 2020) is another deep learning model specifically designed for the automated detection, phase-picking, and classification of seismic events. The model utilizes raw three-component seismic waveforms as input and employs preprocessing techniques such as normalization and windowing to optimize data quality. It first applies 1D CNNs to extract spatial-temporal features from the seismic signals, allowing for the effective detection of characteristic waveform patterns. To further enhance performance, EQTransformer incorporates a self-attention mechanism, a hallmark of transformer models, enabling the model to focus on relevant signal segments while disregarding irrelevant noise. This capability significantly improves the detection of weak seismic signals, even in noisy environments. The architecture facilitates multi-task learning, enabling the model to simultaneously detect seismic events, identify P- and S-wave arrival times, and classify waveforms as either noise or genuine earthquakes. The output includes precise P- and S-phase arrival times, uncertainty estimates, and a classification of whether the detected signals constitute noise or true seismic activity. Notably, EQTransformer stands out due to its end-to-end processing, which eliminates the need for separate models, its exceptional noise handling, and its scalability for efficiently processing large seismic datasets, positioning it as a promising advance in seismology.

While models like EQTransformer offer robust multi-task learning capabilities and handle noise exceptionally well, they also introduce complexity that may not be necessary for all tasks. PhaseNet, with its convolutional architecture, simplifies the detection process while still providing highly accurate P- and S-wave arrival predictions. Its ability to capture spatial information effectively through fully convolutional layers ensures that the model remains responsive to the intricate patterns present in seismic data. Additionally, the architectural similarities to established frameworks, such as the U-net, provide a level of familiarity and reliability, making it easier to implement and adapt for any specific needs. Thus, PhaseNet offers a pragmatic solution that is expected to meet the requirements for precision without the overhead of more complex models.

When applied to additional field data, DNN-based methods proved to be superior to traditional techniques, enabling the detection of a greater number of seismic events with significantly lower computational costs (Dokht et al., 2019). However, as supervised methods, they require a significant amount of labor to label a suitable training dataset. This labor-intensive process is crucial for training models that can accurately distinguish between different seismic signals and noise, thereby posing a significant hurdle in implementing these advanced techniques effectively.

Waveform Classification

Waveform classification involves applying pattern recognition techniques to differentiate between actual seismic events and background noise. Seismic data can often be inundated with various types of noise, including environmental vibrations from human activity, weather, or other natural phenomena. Advanced algorithms and machine learning models are utilized to analyze waveform data, enabling the accurate detection and classification of significant seismic events, including earthquakes, aftershocks, and volcanic activity. These systems learn to recognize the unique characteristics of different seismic events compared to noise patterns, thus improving event detection rates while reducing false alarms. From the standpoint of a single trace, microseismic data is fundamentally a form of time series data. Microseismic events or noise can be regarded as time series, allowing any classification method for time series to be adapted for microseismic event detection. Thus, the challenge of identifying microseismic events can effectively be framed as a classification problem within time series analysis. The classification of time series involves predicting the category label of a sequence over time by building a classifier based on a pre-labeled training dataset. This subject has gained considerable interest with the creation of repositories containing labeled datasets (Gupta et al., 2020). With the recent focus on ML and DL approaches to address the limitations inherent in traditional physics-oriented methods, it has become possible to detect microseismic events by classifying waveforms on an individual basis.

In machine learning, the technique of support vector machines (SVMs) has been widely used for the effective classification of microseismic events (Qu et al., 2018; Zhao et al., 2017). SVM is a supervised machine learning algorithm utilized for classification and regression tasks. It has proven particularly effective for seismic classification due to its ability to handle high-dimensional data and complex decision boundaries. In the realm of seismic classification, SVM operates by mapping input features derived from seismic signals, such as amplitude, frequency, and waveform characteristics, into a high-dimensional feature space. The algorithm identifies the optimal hyperplane that effectively separates different classes of seismic events, such as distinguishing earthquakes from background noise or categorizing various seismic wave types (e.g., P-waves and S-waves) (Vapnik, 1995). The fundamental principle of SVM revolves around maximizing the margin between the closest data points from different classes, known as support vectors. This focus on creating a wide margin enhances the classifier ability to generalize effectively to unseen data, making SVM robust against overfitting, especially in cases where the number of features exceeds the number of samples, an occurrence that is often characteristic of seismic datasets (Bishop et al., 2006; Vapnik, 1998). Furthermore, SVM can utilize various kernel functions to manage non-linear relationships within the data, enabling more accurate classification of complex seismic events. This flexibility means that SVM can outperform traditional linear classification methods, particularly in environments where the data exhibits complex structures (Schölkopf et al., 1999). When comparing SVM and CNN for arrival time picking in seismic data analysis, several factors come into play. SVMs are known for their effectiveness in handling high-dimensional data and are particularly good at finding hyperplanes that separate classes based on labeled datasets. They often perform well with smaller, well-defined datasets and are easier to interpret. However, SVMs may struggle with the complexities of seismic signals, which can be highly varied and noisy. On the other hand, CNNs excel in capturing intricate patterns and features within large volumes of data, making them better suited for the time-frequency analysis of seismic waveforms. Their multilayer architecture allows CNNs to automatically learn relevant features without extensive manual engineering, which is advantageous for distinguishing between P- and S-wave arrivals amidst background noise. While CNNs typically require larger datasets for training, they also offer scalability and higher accuracy for complex tasks like arrival time picking, especially in challenging seismic environments. Ultimately, the choice between SVM and CNN depends on the specific requirements of the task, the available dataset size, and the desired level of accuracy in arrival time estimations.

In other studies, this method is combined with multiscale singular spectrum entropy for high identification accuracy (Zhang, Zhao, et al., 2020). Additionally, Lin et al. (2019) proposed a hybrid method that combines DCNNs with SVMs, automating feature extraction and achieving an impressive accuracy rate of 98%. These approaches collectively highlight the effectiveness of integrating deep and machine learning techniques in seismic event classification.

In DL, earthquake detection has been formulated as a supervised classification task with the introduction of ConvNetQuake, the first CNN designed for detecting and locating earthquakes by distinguishing seismic signals from noise using a large labeled dataset (Perol et al., 2018). ConvNetQuake captures intricate patterns and spatial-temporal features within the signals by processing raw seismic waveform data. With its architecture that emphasizes feature extraction through multiple convolutional layers, ConvNetQuake can produce high accuracy in detecting seismic activities while significantly reducing the computational cost and time associated with traditional methods. The output indicates the likelihood of the input belonging to various seismic event categories, such as earthquakes or aftershocks, thereby improving detection accuracy in noisy environments. Various DL classifiers, including FCN and RNN, were explored, and a Generative Adversarial Network (GAN) was combined with Random Forest (RF) for earthquake early warning, achieving high precision and recall metrics with a substantial dataset (Meier et al., 2019). Capsule Networks (CapsNet) were also introduced to classify mining-induced microseismic signals, even when limited waveform samples are available (Peng et al., 2020; Sabour et al., 2017). They extracted 21 commonly utilized features from each 1D waveform sample, analyzing both time and frequency domains to train the CapsNet. When compared to CNN and traditional machine learning techniques, CapsNet demonstrated superior performance in waveform classification, achieving an impressive accuracy rate of 99.2%.

For the category of waveform classification, it can be discussed that event classification methods do not qualify as true event detection approaches since an "event" refers solely to a waveform segment from an individual trace, limiting its application for determining hypocenter locations. However, microseismic event waveform segments can be identified within each trace during a designated time window, resulting in a compilation of these segments. Techniques such as travel-time moveout and algorithms related to beamforming or migration can then be applied to these waveform segment collections to determine the locations of the microseismic events. Existing literature supports this approach, with Reynen et al. (2018) introducing a novel technique called Feature Weighted Beamforming (FWB) designed for real-time seismic event localization and cataloging. This method transforms component trace data into various features based on frequency content, STA/LTA, and kurtosis. A SVM is trained to categorize the phases as P-wave, S-wave, or noise by utilizing a training dataset that includes feature information about phase arrivals. The predicted travel time probabilities are aggregated for a range of potential origin times and locations. Peaks in the resulting stacking grid that exceed a specified threshold are identified as events. Implementing FWB in real-world settings has demonstrated its effectiveness over traditional STA/LTA methods, achieving a 75% reduction in false positives while maintaining sensitivity and decreasing the average discrepancy in event locations between automated and manually determined solutions (Reynen et al., 2018). In another study, the automatic detection of microseismic events was achieved by utilizing DL techniques to analyze microseismic signals (Saad et al., 2020). This was accomplished by employing a Convolutional AutoEncoder (CAE) to extract features from segments of waveform data within a training dataset. Following this, they implemented a K-means clustering algorithm to classify the raw microseismic recordings into waveform and non-waveform components. The identified waveforms were then processed using a waveform-based Reverse Time Migration (RTM) algorithm to locate the microseismic events accurately. This innovative approach demonstrated that the images of hypocenters generated from machine learning processed waveforms offered greater accuracy compared to those derived directly from noisy data.

Migration-Based Methods

Migration methods in seismology are employed to pinpoint the origin of seismic events by utilizing recorded seismic waveforms. This process involves backprojecting the seismic data through a geological Earth model to reconstruct the location of the source. Migration helps convert the observed wavefields into a visual representation of where seismic energy originated, often represented as images or "velocity models" of subsections within the Earth structure. This technique is particularly useful in complex geological settings where reflections and diffractions can complicate the interpretation of the data. By refining the model of the subsurface, migration-based methods enhance the accuracy of location estimates and can provide insights into the geological features surrounding seismic sources. Essentially, this method combines detection and localization, making it a favored choice in microseismic monitoring projects (Duncan, 2005; Gajewski et al., 2005). In this framework, microseismic events are distinctly recognized, and their locations can be accurately deduced from the images produced by the localization algorithm.

Most studies related to migration-based detection primarily delve into the selection of stacking functions and remain rooted in physics-based modeling. Many existing techniques have been designed to address variations in source radiation patterns, which result in different signal polarities across the recording array (Gharti et al., 2010; Liang et al., 2016), and the comparative analysis of migration-based location and detection methods for microseismic events (Trojanowski et al., 2017).

In recent advancements, DL techniques, specifically Fully Convolutional Networks (FCNs), were applied to automate the location of earthquakes triggered by oil and gas operations (Zhang, Zhang, et al., 2020). The developed system could differentiate between certain interfering events and those outside the monitoring area based on probability estimates. Notably, the approach could locate an event without requiring a velocity model or human input. However, its effectiveness is constrained by the accuracy of the ground-truth hypocenter locations found in catalog data, as well as the breadth of the training dataset. If these limitations are not addressed, the method may fail to produce accurate results. A SVM was also utilized to detect microseismic events from a time-reversed source image created through the microseismic source localization method, which involves auto-correlating adjoint wavefields. They extracted three key features, including coherency, local dip angle, and source energy from the raw source images to classify between source-imaging artifacts and actual microseismic events. Their results indicated that an SVM could successfully distinguish microseismic events from artifacts in source images (Song et al., 2020).

Although it is possible to identify the locations of microseismic events using source images and microseismic events, this approach does not accurately detect microseismic event segments within the traces or produce arrival time picks, as it classifies only the source image and not the raw seismograms.

To summarize the discussions on comparing different approaches, the arrival picking approach is crucial for seismic event detection, but it faces challenges with weak signals and overlapping arrivals. Waveform classification methods are effective for large datasets and the automation of source localization, but they require significant amounts of training data. Migration-based methods are strong in complex environments but depend on dense station coverage for accuracy.

Each approach has its strengths and weaknesses, with the choice depending on network geometry, noise levels, and computational resources. Although developing a new, fully automated microseismic event detection and location workflow is a compelling pursuit, this thesis primarily focuses on enhancing automated microseismic event detection through an arrival-picking approach, as it focuses specifically on the detection of seismic waves' first arrivals, which are often the most reliable indicators of microseismic events. This direct method allows us to timestamp the exact moment a seismic wave reaches the sensors, providing clear and precise data for event detection.

1.3 Objectives

The general objective of this thesis is to enhance the performance of microseismic event identification by implementing a dedicated DL neural network specifically tailored for the unique conditions of underground mines equipped with seismic monitoring systems. The following specific objectives are defined to reach the general objective:

- Create a synthetic dataset representative of the conditions found in a typical mine.
- Evaluate the performance of the network on the synthetic dataset and compare it to the commonly used detection algorithm.
- Assess the performance of the network on real data and determine the best training strategy.

By accurately detecting microseismic events and minimizing false positives, this research ultimately aims to facilitate the development of effective early warning systems for rockburst prevention, thereby improving safety protocols and operational efficiency in the mining sector.

The following chart in Figure 1-6 summarizes the workflow and steps for the current study.

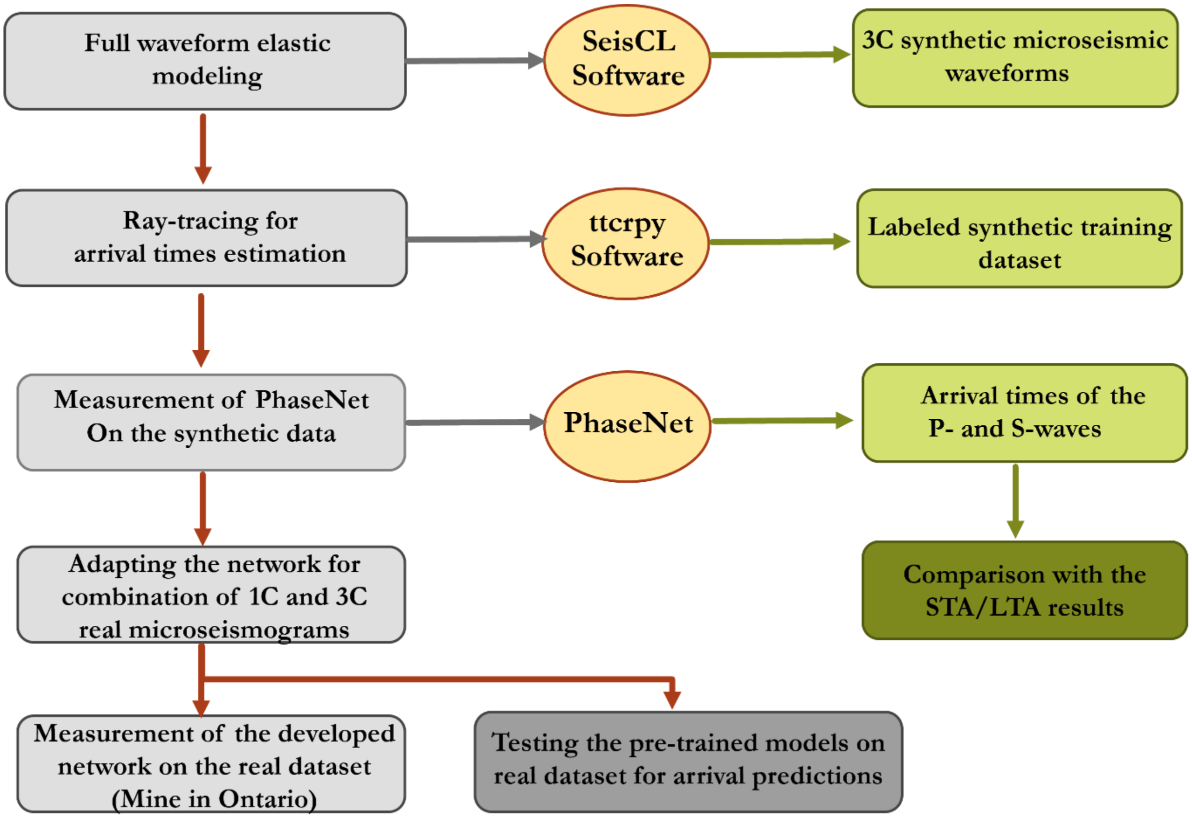


Figure 1-6: The workflow and procedures of the current study.

1.4 Proposed Methodology

Data-driven methods, such as machine learning and deep learning, enhance traditional physics-based modeling approaches. These methods are not only less computationally intensive but also quicker to develop, enabling automated detection of microseismic events in real time (Jinqiang et al., 2021). Both machine learning and deep learning can be implemented as either supervised or unsupervised approaches. However, supervised methods tend to be more appealing, as unsupervised techniques struggle to accurately identify seismic arrivals, particularly in scenarios with significant noise. It's important to note that supervised deep learning necessitates a considerably larger set of labeled training data compared to supervised machine learning (Goodfellow et al., 2016).

The task of microseismicity detection can be compared to the process of image classification. A three-component seismogram can be viewed as a 2D image, with each of the three components representing an RGB color channel (Ross et al., 2018). CNNs can automatically extract features from images using various kernels (LeCun et al., 2015) without the necessity for manual feature design. In contrast to the approach of Ross et al. (2018), which focuses solely on event detection, Zhu et al. (2019) developed an algorithm that determines arrival times in addition to detection. Their algorithm generates three probability distributions (one each for P-waves, S-waves, and noise) based on known probabilities used for training. Events are labeled with a probability of 1 at the corresponding arrival time, surrounded by a Gaussian distribution to represent uncertainty concerning the arrival time. Noise is assigned a probability reflecting the remaining likelihoods of the waves. The authors demonstrated that the time corresponding to the maximum probability given by the algorithm aligns with the wave arrival time. This methodology will be mirrored in our algorithm. The code is tested across various combinations of training data, including synthetic datasets or more conceptual models, with or without actual data provided by our partners. Performance metrics will encompass those traditionally used in deep learning (precision, recall, F1 score).

The proposed project in our study is structured into two main parts: modeling synthetic seismograms and determining arrival times using deep learning. A code utilizing CNN known as PhaseNet is leveraged, and the synthetic dataset is generated to evaluate and test the performance of the network for different training strategies and network parameters (e.g., convolution filter size, stride, and activation function). Synthetic data is created from a conceptual mine model representative of a typical mine context, such as the Nickel and Copper deposits. The mine under consideration is equipped with a microseismic monitoring system comprising 65 receivers, including a combination of 1C and 3C receivers that capture the wavefield. The parameters of the real receiver configuration are employed to generate the synthetic data, with all receivers assumed to have three components. Additionally, we incorporate 95 extra receivers to enhance the density of recordings. The placement of the assumed receivers is carefully determined to maintain a realistic configuration and mitigate unnecessary clustering.

This study adopts the finite-difference time-domain (FDTD) method for full waveform modeling due to its simplicity and capability for parallel execution, particularly utilizing SeisCL software on GPU systems, which efficiently handles computations at individual grid nodes following the SPMD model (Fabien-Ouellet et al., 2017). The FDTD method for solving the wave equation involves discretizing the subsurface into a grid, where seismic properties are assigned to each node. An iterative leapfrog scheme propagates the wavefield from the source to all nodes, injecting seismic energy at each step and updating particle velocities and stress fields using finite difference equations. To prevent artificial reflections and simulate finite propagation, an absorbing layer called a convolutional perfectly matched layer (CPML) is implemented at the domain boundaries (Roden et al., 2000).

The modeling results are compiled into an open-source PassiveSeis database containing over 270,000 labeled samples, each representing a synthetic three-component seismogram. Careful assumptions and precautions are applied to ensure the dataset is realistic and diverse, rendering it reliable for training neural networks. The seismic source parameters are thoughtfully selected to simulate real microseismic events, covering source locations, source time functions, and focal mechanisms. Arrival times for P- and S-waves are determined via the ray tracing method, serving as labels that demonstrate good consistency with the simulated seismograms. The Shortest-Path Method (SPM) from the ttcrpy Python package is used to calculate the P- and S-wave travel times for all source-receiver pairs (Giroux, 2021). This results in a synthetic dataset of labeled three-component waveforms, accurately reflecting the complex seismic behaviours like those found in real mine data.

Furthermore, the synthetic traces exhibit various diffractions, refractions, and wave interference patterns, enhancing the dataset diversity, complexity, and realism. PassiveSeis database could support a range of passive seismic projects harnessing deep learning methods for various objectives: (1) training, validating, and testing new models and neural network architectures for microseismic data processing; (2) augmenting datasets when real data are sparse; and (3) providing an external test dataset to evaluate the transferability and generalizability of deep learning models.

At the core, this study utilizes PhaseNet, a deep convolutional neural network, first applied to synthetic seismograms generated. The unique architecture of PhaseNet is leveraged to identify microseismic occurrences effectively. By utilizing synthetic data, we aim to surpass traditional STA/LTA automatic phase-picking methods, particularly excelling in the detection of S-wave arrivals. We expect the trained model to demonstrate exceptional capabilities in identifying low-magnitude events, yielding valuable microseismic data with a precision that consistently exceeds that of conventional mining techniques. Importantly, the proposed approach does not rely on denoising or data filtering, facilitating rapid microseismic data processing. Furthermore, we adapt the model trained on synthetic seismograms to develop a robust and generalized framework for predicting arrival times in real mine datasets. Experiments were conducted on real data through three distinct scenarios. In the first scenario, the model is trained solely using real data. The second scenario integrates both synthetic and real data during the retraining of the pre-trained model. Finally, the third scenario centers on retraining the pre-trained model exclusively with synthetic data. The performance of all these trained models will then be assessed using the real test dataset. Through this comprehensive methodology, we anticipate delivering highly accurate predictions for both P- and S-wave arrival times, ultimately enhancing the catalogue of microseismic events within the mine for thorough investigation.

1.5 Thesis Outline

The different chapters of the thesis are organized as follows:

- Chapter 1 provides the research problem and background information. It introduces the topic, outlines its history and explains its relevance in the broader field of study. By including key terms and a literature review relevant to the research topic, the introduction chapter provides a comprehensive overview that prepares the reader for the detailed research analysis presented in the subsequent chapters.

- Chapter 2 presents the microseismic synthetic modeling approach and the workflow for preparing the synthetic dataset for the proposed deep learning algorithm. It includes steps for generating the synthetic dataset, labeling the dataset for supervised learning, and splitting it into training, validation, and test datasets for the purpose of training the network.

- Chapter 3 describes the methodology of using a DL algorithm, specifically PhaseNet, as a convolutional neural network architecture for automatic microseismic event detection and phase picking of synthetic data generated based on the information of the specific mine under study.

- Chapter 4 presents the details of microseismic event detection using the conventional STA/LTA method and compares the statistical results between predictions obtained through PhaseNet, and those from the standard STA/LTA approach.

- Chapter 5 demonstrates the capability and accuracy of applying the microseismic event detector trained on a synthetic dataset to a real dataset. Three different scenarios are considered in this chapter, and the results demonstrate that the model re-trained with both synthetic and real seismograms yields the best performance.

- Chapter 6 concludes the thesis with a summary of the findings and a discussion of possible future studies and experiments.

Some of the important terms utilized in the current study can be defined as follows:

1. **Microseismicity:** Small-scale seismic events typically associated with mining activities that can provide critical information on subsurface conditions.
2. **Synthetic Seismograms:** Synthetic seismograms are generated from mathematical models of the Earth's subsurface structure, which incorporate various geological properties such as rock types, layering, and fluid content. These models are often derived from actual geophysical data, including seismic surveys, and use physical principles of wave propagation to simulate how seismic waves would travel through different materials. The generated waveforms can then mimic realistic seismic signals produced by natural events like earthquakes, allowing researchers to better understand seismic behavior and improve seismic hazard assessments.
3. **PassiveSeis Database:** This open-source repository contains labeled synthetic seismograms designed to support deep learning applications in mining seismology.
4. **Event Detection:** The process of identifying seismic events within a continuous stream of seismic data.
5. **Hypocenter Location:** Determining where a seismic event originates within the earth.
6. **Early Warning Systems:** Technologies and protocols developed to detect and alert personnel to potential hazardous events such as rockbursts in mining operations.
7. **Elastic waveform modeling:** This is a computational technique used to simulate the generation and propagation of seismic waves through an elastic medium, such as rocks and sediments in the Earth's subsurface.
8. **Ray tracing:** It is a powerful technique used in seismic analysis to compute the travel times of seismic waves as they propagate through complex geological structures. The method relies on the principles of geometrical optics, treating seismic waves as rays that

travel along straight paths but can be bent due to changes in the medium's properties, such as velocity variations.

9. **Velocity Models:** A velocity model is a comprehensive representation of how seismic wave velocities change with depth and across different horizontal positions within the subsurface. It plays a crucial role in accurately predicting the propagation of seismic waves during ray tracing, as it determines the travel speeds of these waves in various geological layers, thereby influencing their paths and the resulting subsurface imaging. A well-constructed velocity model is essential for interpreting seismic data and understanding geological structures.
10. **Ray Paths:** Seismic rays are traced from the source location to the receiver. Depending on the velocity model, rays may bend or reflect at interfaces where the seismic velocity changes. Depending on their interaction with geological layers, the paths can be categorized into direct, reflected, and refracted rays.
11. **Travel Time Calculation:** The travel time for each ray is computed based on the distance traveled through each layer and the velocity within those layers. The total travel time is the sum of the time to traverse each path segment.
12. **Deep Learning (DL):** A subset of machine learning employing neural networks with multiple layers to analyze complex data patterns and make predictions.
13. **PhaseNet:** A specific deep learning architecture for identifying seismic phases, namely P-waves and S-waves, from seismic data.
14. **Training Set:** A training set comprises examples used during the learning phase, enabling a model to adjust its parameters.
15. **Validation Set:** A validation set is a collection of examples used to fine-tune a network's hyperparameters, which refer to its architecture. It’s also known as the development set or "dev set." This dataset is utilized for regular evaluations of the model, allowing the model to see this data occasionally, but it does not learn from it.
16. **Training Dataset:** The training dataset includes both the training set, and the validation set combined.
17. **Test Dataset:** A test dataset, or test set, is employed to assess the classifier developed during the training phase. It features the same probability distribution as the training dataset but remains independent from it.
18. **Pre-trained model:** A pre-trained model refers to a machine learning model that has been previously trained on a large dataset and can be fine-tuned or used for different but related tasks. In the context of deep learning, pre-trained models are often employed to leverage

the learned features from massive datasets, allowing for improved performance on specific tasks with less data and training time.

1.6 Contributions of the Thesis

This thesis presents original contributions that significantly advance the detection and management of microseismic events relevant to early warning systems for gallery collapses and the safety of personnel in mining operations. Central to this work is the preparation of an extensive synthetic seismic database designed specifically for the application of deep learning algorithms in mining-induced microseismic event detection. This unique dataset implicitely integrates diverse geological scenarios and mining conditions, reflecting the complexities of real environments.

A key innovation is the meticulous design of synthetic seismic events that replicate the characteristics of microseismic emissions generated during mining activities. To enhance the accuracy of our dataset, we utilized wavefield simulation tools like SeisCL and ttcrpy for the seismic waveform's generation and labeling of P- and S-wave arrival times, providing precise annotations crucial for effective analysis. Additionally, we employed the PhaseNet as a deep learning network which has been trained for active seismic data in a different context to detect low-magnitude events within noisy mining environments.

The comprehensive dataset generated addresses industry challenges by allowing the development of more accurate detection algorithms without the constraints imposed by real-world data scarcity. We also leveraged a pretrained model in different scenarios to transfer the knowledge from the model to real data, optimizing a model capable of providing accurate predictions for low-magnitude events within the noisy environment of the mine with high precision.

On the other hand, to determine the optimal training sequence to be used with actual data. The performance of the deep learning algorithm has been compared to the classical algorithm currently used in the industry, and the results show a clear improvement in the detectability of low-magnitude events.

Ultimately, this contribution not only improves our understanding of microseismic activity within the mines but also establishes a precedent for utilizing synthetic datasets in other microseismic processing steps, with the potential to enhance safety measures and risk mitigation strategies in mining operations.

Finally, during my thesis, I submitted an article titled "Detection of Mining-Induced Microseismicity through a Deep Convolutional Neural Network" in the Journal of Applied Geophysics (ANNEXE II). Additionally, two extended abstracts were written and submitted for the 77th Canadian Geotechnical Conference, which presented the results of synthetic data simulations and the performance of the deep neural network on synthetic datasets. I also contributed to the conference La Journée des Sciences de la Terre et de l'Environnement (JSTE), highlighting the comparison between the detection results of a deep learning algorithm (PhaseNet) and an industry conventional approach (STA/LTA). Furthermore, I prepared an abstract for Geoconvention 2025, showcasing predictions on the real mine dataset via transfer learning (TL). The list of contributions and publications is as follows:

- Detection of mining-induced microseismicity through a deep convolutional neural network, Sepideh Vafaei Shoushtari, Bernard Giroux, Erwan Gloaguen, Maher Nasr, Submitted to Journal of Applied Geophysics, July 2025.

- Automatic event detection of mining-induced microseismic data based on PhaseNet: a case study of a Nickel-Copper Mine, Ontario, Sepideh Vafaei Shoushtari, Bernard Giroux, GeoConvention 2025, May 2025.

- Detection of mining-induced microseismicity through deep learning, Sepideh Vafaei, Bernard Giroux, Erwan Gloaguen, Maher Nasr, 77th Canadian Geotechnical Conference, GeoMontreal 2024, Extended Abstract, September 2024.

- PassiveSeis: A synthetic open-source database of mining-induced microseismic data, Maher Nasr, Sepideh Vafaei, Bernard Giroux, Erwan Gloaguen, 77th Canadian Geotechnical Conference, GeoMontreal 2024, Extended Abstract, September 2024.

- Mining-induced microseismicity detection using a deep convolutional neural network, la Journée des Sciences de la Terre et de l'Environnement (JSTE), Sepideh Vafaei, Bernard Giroux, Erwan Gloaguen, March 2024.

2 SYNTHETIC MICROSEISMIC DATA

2.1 Introduction

This study incorporates both microseismic synthetic and field datasets, which play vital roles in the training and evaluation processes of our Neural Network. Synthesizing seismic data enables the creation of a virtually limitless number of recordings, providing an ample dataset for training. It can be outlined in several key steps: first, creating noise-free synthetic seismograms; next, adding white Gaussian noise (WGN) or colored noise to these recordings to achieve various signal-to-noise ratios (SNRs); and finally, labeling the dataset for feature extraction.

This chapter describes the methodologies employed for generating synthetic microseismic data, which are integral to the overall work presented in this thesis.

2.2 Synthetic Microseismic Dataset Generation

The integration of deep learning methods into the mining sector to detect mining-induced microseismicity remains a work in progress, encountering several challenges. A significant obstacle is the insufficient availability of labeled data needed to train an optimized model effectively, coupled with various noise sources. To address this challenge and improve the accurate detection of microseismic events, a modeling approach for creating synthetic microseismic datasets for network training is a viable solution (Lehmann et al., 2024; Mendecki, 1996).

Our methodology utilizes a staggered-grid finite-difference approach to solve the wave equation, which is formulated in terms of velocity and stress (Fabien-Ouellet et al., 2017). The results will be synthetic microseismic datasets with varying levels of random noise, designed to assess the robustness and reliability of the proposed deep learning architecture in noisy environments. The wavefield simulation utilizes data collected from an underground mine located in Sudbury, Ontario, Canada. This information comprises the 3D velocity models of both P- and S-waves, as well as the receiver configuration relevant to the seismic monitoring volume within the mine. The parameters for the seismic sources have been primarily assumed and carefully selected to closely resemble real microseismic events. The subsequent subsections provide a detailed explanation of the various procedures employed in wavefield modeling.

2.2.1 Geology and Specifications of the Mine

The synthetic data simulation relies on information from a specific underground hard-rock mine, known as Nickel Rim South (NRS) located in Ontario, Canada. The mine has been in operation since mid-2009, extracting nickel and copper at depths ranging from 1,100 to 1,720 meters below the surface. It also produces minor amounts of cobalt, gold, silver, platinum, and palladium using transverse blast hole mining. The ore deposit includes a nickel envelope in the hanging wall and a copper envelope in the footwall. The main portion of the nickel is located at the top of the ore deposit, with a transition zone from nickel to copper in the middle, and the main part of the copper is hosted at the bottom of the ore deposit. The local geology of the monitored zone within the deposit is characterized by a predominance of igneous (Norite) and metamorphic (Gneiss) rocks. The stopes create sharp velocity variations and, therefore, add complexity to the monitoring volume in terms of the seismic properties of the rock mass. Figure 2-1 and Table 2-1 present geological domains and their relevant seismic velocities. Therefore, seismic monitoring of the volume or structure at the proposed mine is complex due to a significant variation in seismic properties of the rock matrix throughout the mine. The mine is equipped with a microseismic (MS) monitoring system, as in Figure 2-1, and a strong ground motion (SGM) system.

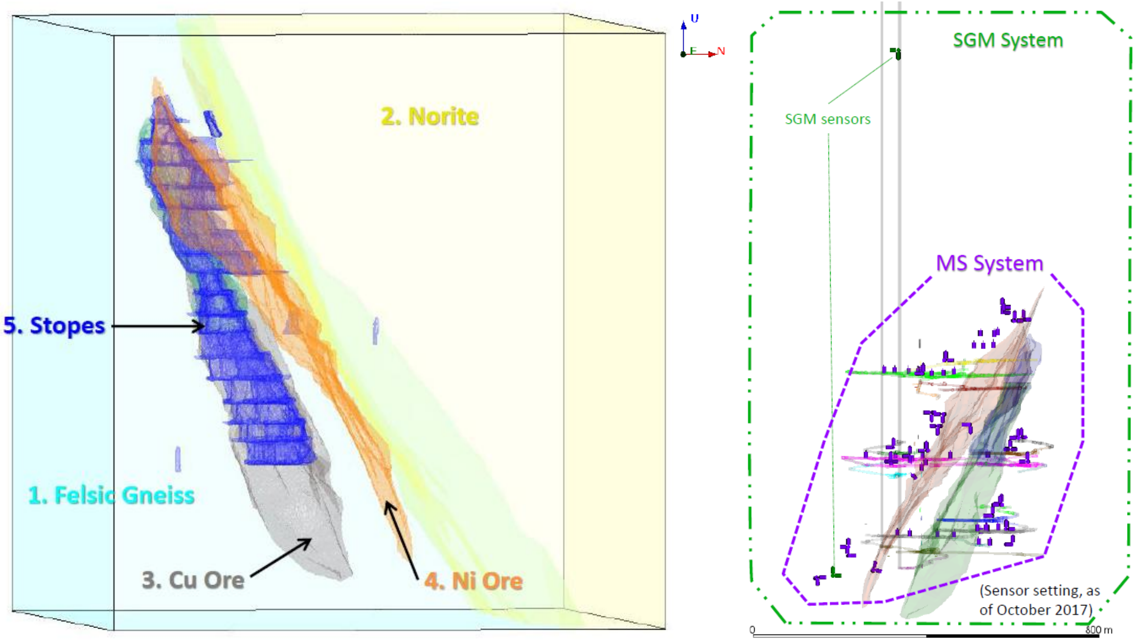


Figure 2-1: Three-dimensional (3D) geological model of the proposed mine (left) and the microseismic monitoring zone around the mining layout (right) (Adapted from ESG Solutions).

The MS system is used to determine event locations, while source parameters of large local seismic events are complemented by more accurate estimates determined by the SGM system. Recently, a potential expansion project has been considered to access deeper copper orebodies located about 2 km west of the current mine, at depths ranging from 2,000 to 2,700 m below surface. The primary mining method is sublevel open stoping with cemented hydraulic backfill. Extraction generally follows a primary–secondary transverse stoping sequence arranged into three overhand pyramids, although modifications were applied in some areas to accommodate local geology and orebody geometry. A time-stamped mine geometry model, showing both the current extraction state and the historical stoping sequence, is presented in Figure 2-2 (Yadav, 2024).

Jalbout et al. (2014) describe the geology, in situ stress conditions, rock mass properties, and major structural features of the site. Details of the mine’s microseismic monitoring system and its applications, such as assessing seismic hazards, identifying fault trends, evaluating rock mass response, and monitoring secondary pillar behavior, are reported in Simser et al. (2022).

The use of a dense seismic array capable of detecting smaller-magnitude events has enabled ongoing assessment of rock mass conditions and the seismic activity of key geological structures. Observing how seismic behaviour evolves with changing mine geometry has become a crucial tool for ground control engineers, supporting both strategic and tactical decision-making to enhance safety and effectiveness (Yadav, 2024).

Stopes are accessed from three main levels: 1280 m level for nickel, 1480 m level for nickel and copper, and 1660 m level for nickel zone in the specific monitoring system (MS), as shown in Figure 2-2.

Table 2-1: Geological domains with their seismic P- and S-waves velocities.

Geological Domain	Vp (m/s)	Vs (m/s)
Felsic Gneiss	5970	3513
Norite	6131	3564
Copper Ore	5350	3100
Nickel Ore	6520	3590
Stopes	1745	700

The deposit geometry, with access from both the hanging wall and footwall, is ideal for providing comprehensive seismic monitoring. The resulting high-density array comprises 65 sensors spaced approximately 120 meters apart, consisting of 39 uniaxial accelerometers and 26 triaxial geophones. Given the relatively compact mining volume (under 1 km^3), this setup provides dense coverage capable of recording moment magnitudes below -2. While small-magnitude seismic events rarely pose a direct safety risk, their cumulative occurrence serves as an excellent record of rock mass damage (Jalbout et al., 2014).

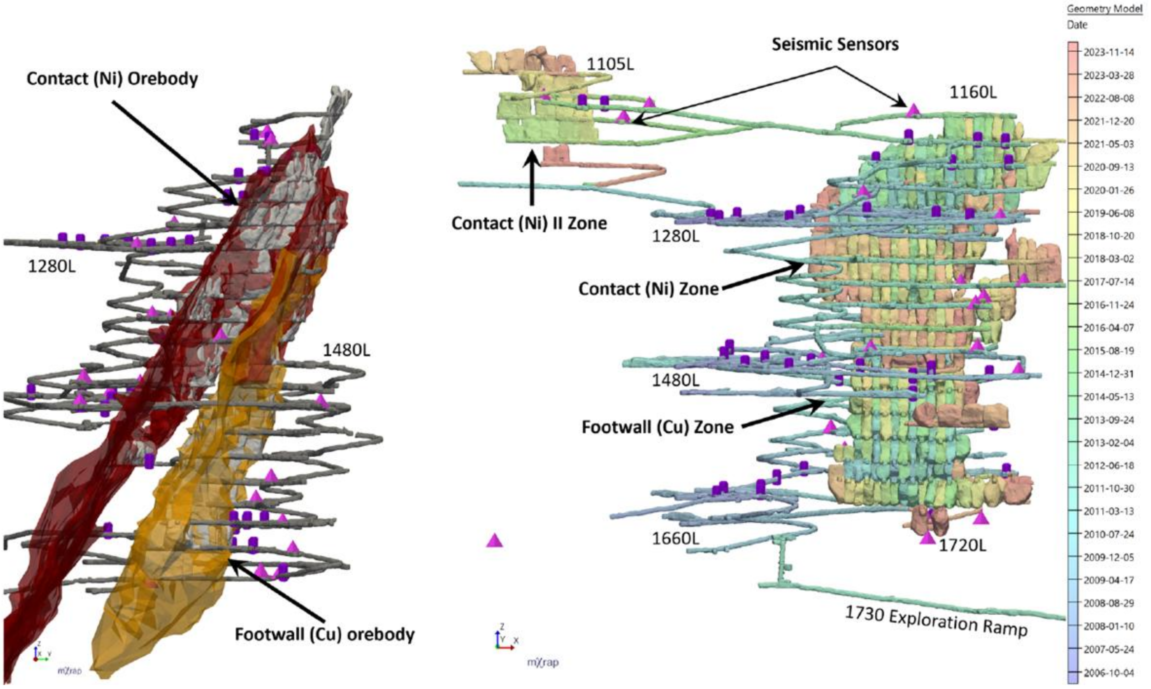


Figure 2-2: West-looking cross-section (left), and north-looking long section of the Nickel Rim South Mine illustrate the mined stopes, development openings, and the microseismic array as of May 2024 (right). Level designations represent depth in metres below surface, and excavations are colour-coded by extraction date. The cross-section also highlights the mineralized envelopes: the hanging-wall zone hosts a typical Sudbury-style massive sulphide nickel (Ni) deposit, while the footwall copper (Cu) zone comprises irregular copper veins and stringers enriched with significant precious metal content that were bulk mined.

A rectangular subdomain of dimension $600 \times 800 \times 730$ m^3 is considered to generate the synthetic seismograms (Figure 2-3). P-wave velocity in the specific subdomain ranges from 1700 m/s to more than 6500 m/s. The S-wave velocity model exhibits values ranging from 700 to 3600 m/s. Both V_P and V_S models were created based on a velocity analysis that utilized 41 calibration shots. 3D velocity models utilized for the synthetic modeling in the proposed subdomain are depicted in Figure 2-4. The resolution of the initial model is defined on a coarse regular grid of 10 m cubic cells (received from ESG Solutions) and is linearly interpolated to fit the resolution required for the 3D simulation steps ($\Delta x = 0.75$ m).

The accuracy of synthetic seismogram modeling improves when the density information of the considered domain is available. However, density information is incomplete, and the density model is approximated from P-wave velocity using Gardner's equation (Gardner et al., 1974):

$$\rho = 1.74 {V_p}^{0.25} \tag{2.1}$$

where ρ is the bulk density in g/cm3, and V_p is the P-wave velocity in km/s.

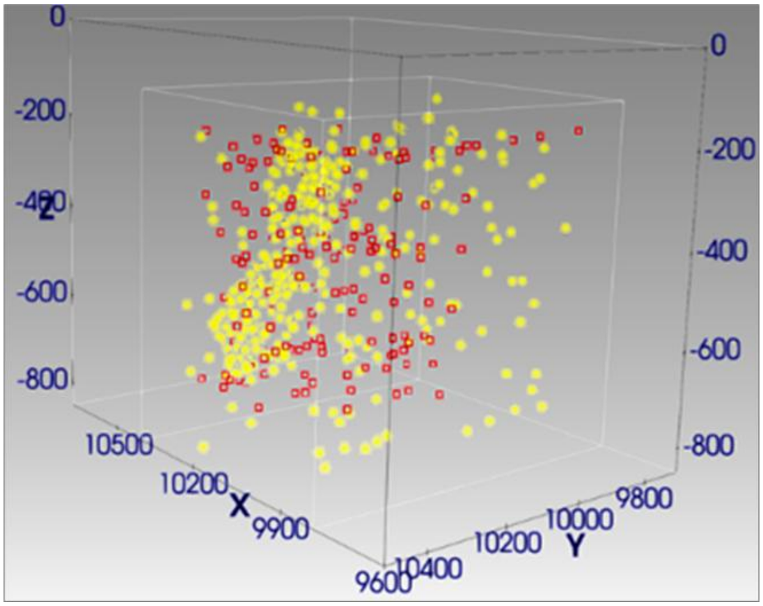


Figure 2-3: The rectangular subdomain of the monitoring volume (600 × 800 × 730 m^3). Red squares and yellow dots show the sensor configuration and microseismic event locations, respectively.

2.2.2 Seismic Sources

The characteristics of the simulated microseismic sources are established through a randomization process. Specifically, the coordinates of the hypocenters are derived from two different probability density functions (PDFs). The first set comprises half of the coordinates sampled from a uniform distribution within the defined subdomain, allowing for an even distribution of events across the volume. The second set of coordinates is sourced from a PDF that aligns with the spatial velocity gradient illustrated in Figure 2-5, a method that emphasizes areas of potential induced seismicity, i.e. the second sampling aligns with the boundaries of stopes and orebodies, which are identified as preferential regions for seismic activity due to their geological characteristics (see Figure 2-1). To highlight regions of elevated stress within the velocity gradient model, an iso-surface filter was applied in ParaView. This visualization delineates continuous 3D surfaces corresponding to threshold stress values, allowing clear identification of source positions in structurally significant high-stress zones. The distance of the sources to the nearest sensor is set to 30 meters to avoid near-field effects as well as P- and S-wave interferences.

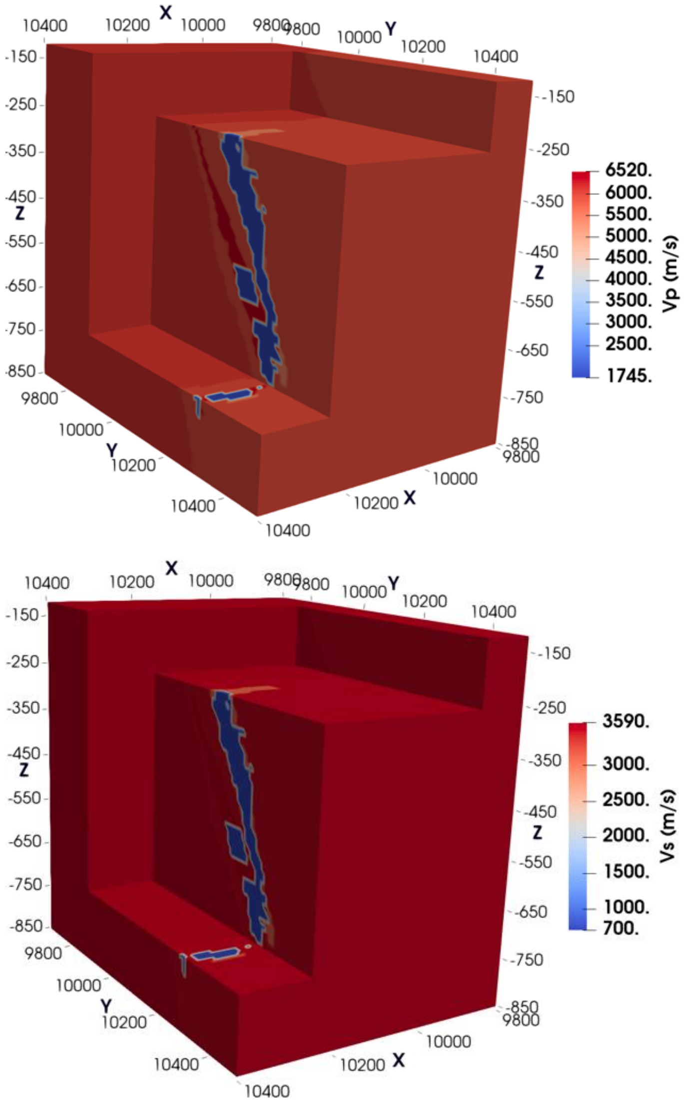


Figure 2-4: 3D Velocity models utilized for the waveform simulation in the proposed subdomain of the mine under study: P-wave velocity model (left) and S-wave velocity model (right).

Additional source parameters are selected to ensure that the seismic data accurately reflects the diverse range of microseismic waveforms recorded in actual mining operations. The frequency values for these synthetic sources are sampled uniformly within a range of 20 to 80 Hz, a range commonly observed in the literature on mining-induced seismicity. Moreover, variations in the focal mechanism and source time function (STF) are introduced to create diverse wavefields. In total, four distinct focal mechanisms are considered: (1) a single force with arbitrary orientation (SF), (2) a double-couple mechanism (DC) featuring randomized fault parameters such as rake, strike, dip, and magnitude, (3) an isotropic mechanism, and (4) a compensated linear vector dipole (CLVD) mechanism (Figure 2-6). These focal mechanisms are foundational in seismology as they represent different physical processes that can occur during seismic events, playing crucial roles in understanding geological processes (Richards et al., 2002).

- **Single force with arbitrary orientation (SF):** This mechanism is primarily associated with localized variations in stress within the Earth's crust. Typically observed in scenarios such as mining operations or nuclear detonations, the SF mechanism depicts how a singular event can generate seismic waves without conforming to standard faulting patterns. The seismic response in these cases arises from an explosion or collapse rather than traditional fault slip, highlighting the complexity of stress distributions caused by anthropogenic activities or geological anomalies (Richards et al., 2002).

- **Double-Couple (DC):** The DC mechanism stands as the predominant model for analyzing tectonic earthquakes, especially those occurring along fault lines. This mechanism characterizes the shear displacements along a fault plane through a configuration of two opposing forces. It effectively models various types of faulting, including strike-slip, reverse, and normal faulting, making it invaluable for evaluating the orientation and slip of geological faults during seismic events (Sibson, 1989). The double-couple model is pivotal for researchers in determining the seismic source characteristics and understanding the mechanical behavior of fault interactions within tectonic settings.

- **Isotropic:** In contrast to the more directional mechanisms, the isotropic mechanism represents seismic sources demonstrating uniform properties in all directions. This is typically linked to processes such as volcanic eruptions or explosions, where the energy released is symmetrical and radiates evenly across space. The uniform stress distribution implies that seismic energy propagates equally, allowing for simplified modeling of waveforms despite complex geological interactions. Isotropic mechanisms are particularly relevant in the study of large volcanic eruptions or geothermal activity, as they streamline the understanding of such events and their resulting seismic effects (Chouet, 1996).

- **Compensated linear vector dipole (CLVD):** The CLVD mechanism is utilized to describe seismic sources exhibiting both shear and volumetric changes, for instance modeling complex seismic activities linked to volcanic processes or fluid movement (e.g., geothermal systems). This mechanism captures the intricacies associated with faults that do not conform strictly to either shear or tensile categories, thus providing a nuanced understanding of seismic sources in contexts where traditional models fall short (Knopoff et al., 1970).

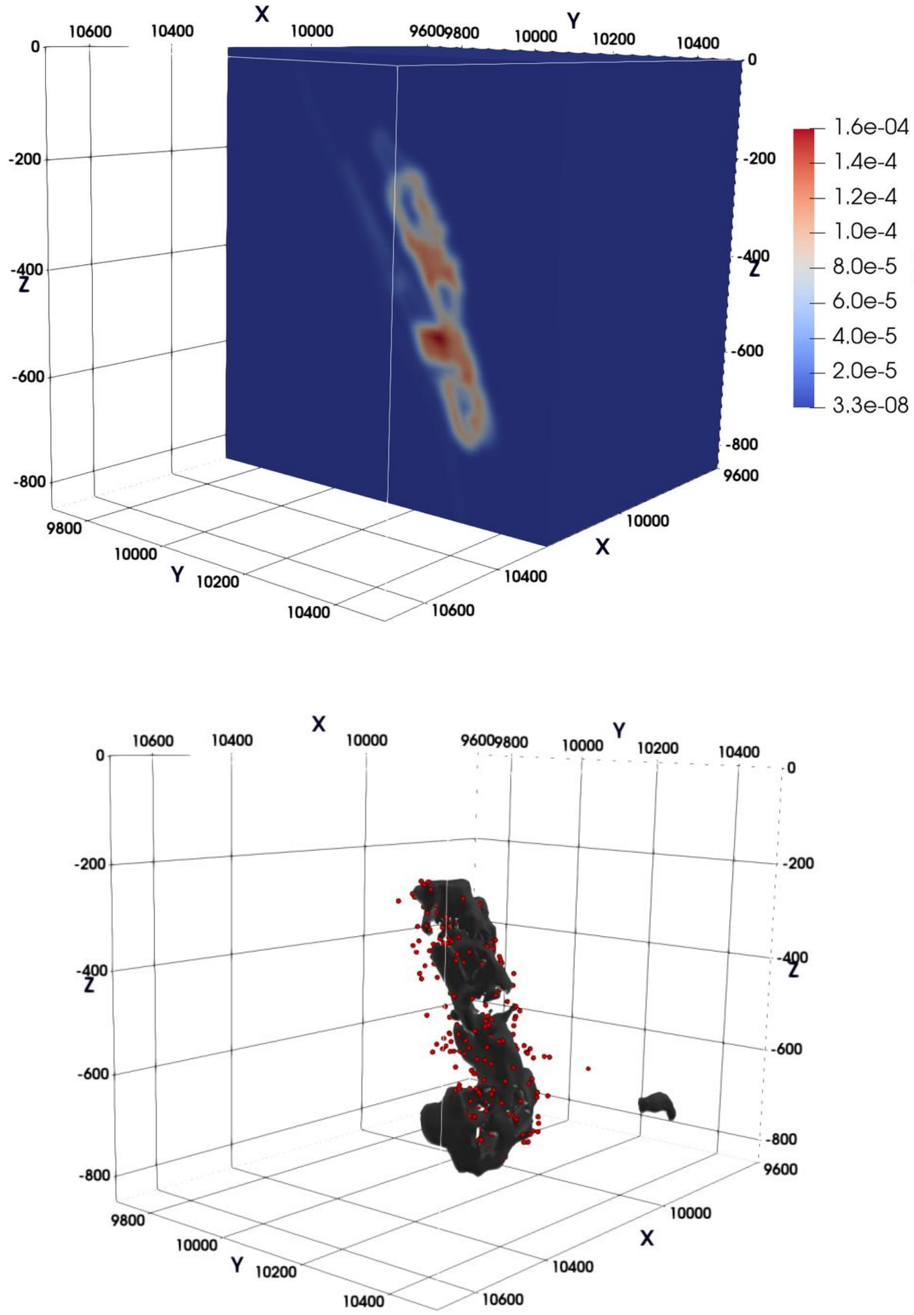


Figure 2-5: 3D PDF used to sample a part of seismic events computed from the velocity derivative (top) and iso-surface representation of 3D PDF with the microseismic events sampled from this PDF (bottom).

The source time functions employed in this modeling include the sine-cubed function, the Fuchs-Müller wavelet, and a kinematic STF introduced by Dreger et al. (2007). The various time functions are documented in Table 2-2, providing a reference for their specific temporal characteristics. The shapes of these source time functions, integral to the modeling process, are illustrated in Figure 2-7, demonstrating their characteristics in both time and frequency domains.

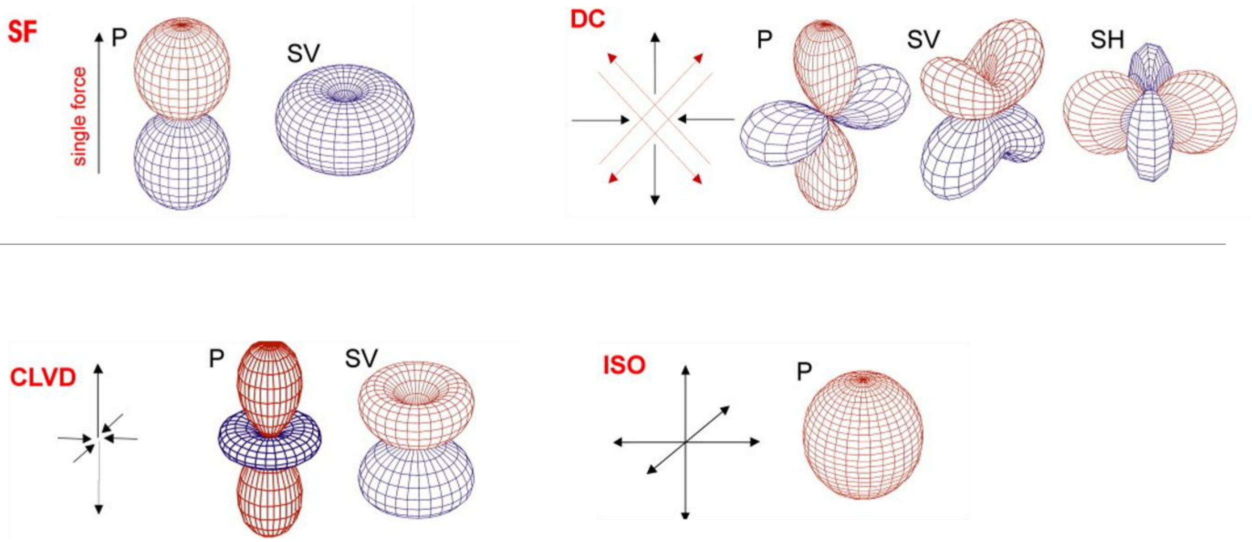


Figure 2-6: Different source mechanisms used for full-waveform modeling.

Table 2-2: Source time functions (STFs), f_c is the central frequency.

Source Time Function (STF)	**Functions**
Sine-cubed function	$S3(t) = \frac{3}{4}\,\pi\, f_c \sin^3(\pi\, t\, f_c))$
Fuchs-Müller wavelet	$f_m(t) = \sin(2\pi\, t\, f_c) - \frac{1}{2}\sin(4\pi\, t\, f_c)$
Kinematic STF	$s(t,\zeta,\tau) = t^{\zeta} e^{\frac{-t}{\tau}}$

Combining these diverse parameters for seismic source modeling has allowed generating 1,700 events. These comprise 600 events resulting from a single force source, 500 from the DC mechanism, 300 from isotropic sources, and 300 from CLVD mechanisms. These event distributions are designed to reflect the complexity and variability of seismic activity observed in real mining environments, contributing to a comprehensive understanding of induced seismicity in such settings.

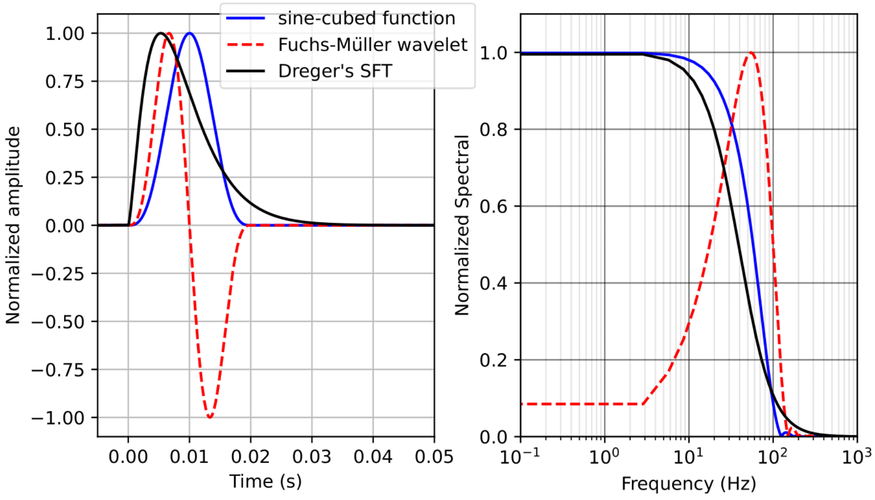


Figure 2-7: The shape of source time functions (STFs) utilized for full-waveform simulation.

2.2.3 Seismic Receivers

As mentioned previously, the specific mine under study features a microseismic monitoring system that includes 65 receivers, with a combination of 1C and 3C receivers employed to capture the wavefield. The synthetic data is prepared using the actual receiver configuration from the mine, with the assumption that all receivers possess three components. Additionally, we have included 95 extra receivers to enhance the receiver density, allowing for a more thorough recording of the modeled wavefield and increasing the number of data samples. The localizations of the additional receivers have been thoughtfully selected to maintain a realistic layout and to prevent unnecessary clustering (Figure 2-3).

2.2.4 Seismic Waveform Modeling

Seismic modeling has become an essential asset across various aspects of seismic investigations, playing a crucial role in establishing optimal recording parameters during data acquisition, testing processing techniques, and enhancing the understanding of seismic measurements through comparisons between synthetic and field seismograms. Additionally, modeling is a critical element in seismic inversion processes. Seismic waveform modeling can be categorized based on the underlying physical assumptions regarding the Earth's medium. Acoustic modeling treats the medium as an ideal fluid, facilitating the propagation of only compressional (P) waves and relying on the acoustic wave equation (Richards et al., 2002). In contrast, elastic modeling incorporates both P-waves and shear (S) waves, thereby making it more appropriate for simulating wave propagation in solid materials through the use of the elastic wave equation (Chapman, 2004). Viscoelastic modeling extends this concept by accounting for time-dependent strain, effectively capturing both elastic and viscous behaviors, which are critical for accurately representing attenuation and dispersion in sediments (Carcione, 1993). Viscodynamic modeling takes this a step further by incorporating the dynamic responses of materials under varying stress conditions (Mavko et al., 2020). Full waveform modeling (FWM) aims to simulate the entire seismic wavefield, encompassing all phases and interactions, which offers comprehensive insights into subsurface structures (Virieux et al., 2009). In contrast, ray tracing modeling simplifies the wavefield to calculate travel times and paths through various media, making it particularly useful for high-frequency approximations and velocity modeling (Červený et al., 2007). Numerical techniques are fundamental to many of these modeling approaches. Finite difference and finite element methods discretize wave equations over spatial grids or meshes, thus accommodating complex geometries (Graves, 1996). Spectral element modeling merges the geometric flexibility of finite elements with the accuracy of spectral methods, facilitating precise simulations of wave propagation in heterogeneous media (Komatitsch et al., 1999). Each modeling approach offers distinct advantages depending on the geological context, types of waves, and the computational resources available.

Two primary methods for computing the full waveform response in elastic media are analytical solutions and numerical simulations. Analytical solutions, like Green's function for an infinite half-space (Richards et al., 2002), are typically used for simpler models, such as homogeneous or layered media. In contrast, numerical solutions comprising methods such as finite-difference (Graves, 1996; Kelly et al., 1976), finite-element (Zienkiewicz et al., 1997), and spectral element methods (Tromp et al., 2008), are better suited for simulating wavefields in complex media but generally require more computational resources. All the different approaches can effectively simulate the wavefield with sources and receivers situated within a defined velocity model (Wong et al., 2011). Among the various FWM techniques, the finite-difference time-domain (FDTD) method is particularly interesting due to its adaptability in modeling wave propagation in complex media while maintaining computational efficiency. As modeling scales and complexities have increased, numerous strategies have been developed to enhance the computational efficiency and accuracy of FDTD (Bohlen, 2002; Fabien-Ouellet et al., 2017; Shi et al., 2018).

In the proposed study, full waveform simulation is carried out by solving the wave equation (2.2) in a three-dimensional elastic medium, utilizing the open-source software SeisCL. This program implements the first-order formulation of the wave equation through a FDTD method (Fabien-Ouellet et al., 2017) represented as follows:

$$\partial_t v_i - \frac{1}{\rho}\partial_j \sigma_{ij} = f_{v_i} \tag{2.2}$$

$$\epsilon_{kl} = \frac{1}{2}[\partial_k v_l + \partial_l v_k], \quad k,l = 1,2,3 \tag{2.3}$$

$$\sigma_{ij} = c_{ijkl}\epsilon_{kl} \tag{2.4}$$

where, f_{v_i} is the source term, ρ is the material density, v_i is the i^{th} component of wavefield particle velocity, σ_{ij} is the Cauchy stress tensor, and ∂_t is the first-order temporal derivative. ϵ_{kl} is an element of the linear strain tensor, ∂_k is the spatial derivative in the k^{th} direction and c_{ijkl} is the fourth-order stiffness tensor. Equation (2.4) denotes a constitutive relationship that describes the elastic material properties through c_{ijkl} (Weiss et al., 2013). More details and equations regarding the solution of the wave equation in elastic media are given in ANNEXE I. Solving the wave equation using the FDTD technique involves the following steps: (1) The subsurface domain is discretized using a regular grid, where the seismic properties (velocity, density) are defined at each node. Figure 2-8 illustrates a basic cell from the standard staggered grid, detailing the placement of each seismic variable (Fabien-Ouellet et al., 2017). (2) An iterative leapfrog algorithm is applied to propagate the wavefield from the source position to all nodes within the domain. In each iteration, seismic energy is introduced into the domain at the source location, according to the source time function. At the same time, the finite difference formulations of equations (2.2) to (2.4) are utilized to update the particle velocities and stress fields. (3) To simulate finite propagation and eliminate artificial reflections, seismic energy is absorbed at the domain boundaries by enclosing the simulation area with an absorbing frame. Common methods for this include using either a perfectly matched layer (PML) or a convolutional perfectly matched layer (CPML) (Roden et al., 2000). PML condition has the remarkable property of having a zero reflection coefficient for all angles of incidence and all frequencies before discretization, and has become widely used (Collino et al., 2001). However, this reflection coefficient is no longer zero after the discretization and becomes very large at grazing incidence. Therefore, an improved version of the PML condition has been developed: the CPML boundary condition, which is applied in this study for the purpose of synthetic modeling (Komatitsch et al., 2007).

In this study, the FDTD numerical approach is chosen for full waveform modeling due to its simplicity and compatibility with parallel computing architectures. SeisCL effectively leverages this feature, enabling GPU operation with clustering. The finite-difference method aligns well with

GPUs, as it relies on independent computations across grid nodes, which mirrors the SPMD (single-program, multiple-data) model. We opted for elastic modeling due to insufficient information regarding the seismic attenuation model of the monitored mine. Additionally, tests performed with viscoelastic modeling indicate that elastic simulations yield synthetic seismograms better aligned with the arrival times determined through the ray-tracing method. Elastic modeling is also considered for its reasonable computational efficiency. For the current experiment, the simulation was conducted over a duration of 0.35 seconds, using parameters set to a sampling rate of $\Delta t = 0.012$ ms and a spatial step of $\Delta x = 0.75$ m. An 8th-order finite difference operator based on Holmberg's coefficients was employed to mitigate numerical dispersion. Since all microseismic events and receivers are located deeply within the mine (over 120 meters), no free surface was assumed for the domain. Instead, a finite 3D domain was defined, enclosing the entire area within a CPML frame that is 17 meters thick.

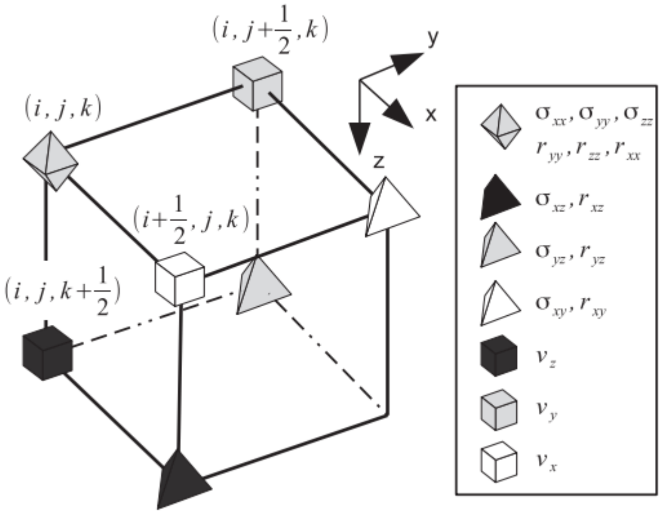


Figure 2-8: An elementary cell illustrating the positions of nodes for each seismic variable (Fabien-Ouellet et al., 2017).

2.2.5 Ray-Tracing

The ray-tracing approach to calculating travel times is a crucial technique in seismic imaging and modeling. It allows for estimating the time taken for seismic waves to propagate from a source to a receiver. This method is especially useful in complex geological settings, where variations in material properties can significantly impact wave propagation (Vidale, 1988).

To label the simulated dataset, the Shortest-Path Method from the open-source software ttcrpy is employed to compute the travel times of P- and S-waves for all source-receiver pairs within the designated subdomain (Giroux, 2021). The velocity models utilized for waveform modeling are used during this step.

ttcrpy is a software package designed for calculating travel times and ray tracing of seismic and electromagnetic waves in geophysical contexts. It can be used for applications like ray-based seismic and ground-penetrating radar (GPR) tomography, microseismic event location through joint hypocenter-velocity inversion, and migration analysis. The tool supports computations on both 2D and 3D rectilinear grids, as well as on 2D triangular and 3D tetrahedral meshes. To enhance its flexibility, it features three distinct algorithms: the Fast-Sweeping Method (FSM), the Shortest-Path Method (SPM), and the Dynamic Shortest-Path Method (DSPM). Additionally, it allows for parallel processing on multi-core systems (Giroux, 2021).

Within ttcrpy, there are two variants of the SPM. The first is the traditional grid-based SPM, which employs Dijkstra’s algorithm (Dijkstra, 1959), enabling the calculation of travel times at all grid nodes. This algorithm can be adapted to calculate seismic travel times by representing the subsurface as a network of nodes and connexions, where the connexion weights represent travel times based on seismic velocity and edge length, allowing for the finding of the shortest path (i.e., the minimum travel time) between a source and a receiver. The second variant is an improved version known as the Dynamic Shortest Path Method (DSPM), which employs a backward tracing technique to track the ray paths for each source-receiver pair and update the travel times along these paths. In DSPM, secondary nodes are temporarily added in the vicinity of the source, enhancing the efficiency of the ray tracing process, although backward ray tracing can also be conducted in the traditional SPM. In this work, we select the classical grid-based SPM due to its efficiency in computing travel times while accommodating model complexities. A notable feature of the ttcrpy SPM implementation is the distinction between primary and secondary nodes. Primary nodes are positioned at the corners of the cells, while secondary nodes are found along the edges and faces surrounding the cells. The inclusion of secondary nodes increases the accuracy and angular distribution of the discrete ray paths. However, it is important to note that the ray path remains an approximation and may differ from the actual one, as depicted in Figure 2-9 for a homogeneous model.

As the domain in this study is 3D, we refined the initial 3D rectangular grid used for domain discretization (Figure 2-3) by incorporating secondary nodes along each cell edge and face. This increase in node density enhances angular coverage and the resolution of ray paths, resulting in more accurate travel time calculations for each source-receiver pair (Giroux et al., 2013).

In this instance, 12 and 144 secondary nodes are added along each cell edge and face, respectively. For clarity, Figure 2-9 provides a different example, showing 5 and 20 secondary nodes along the edges and faces.

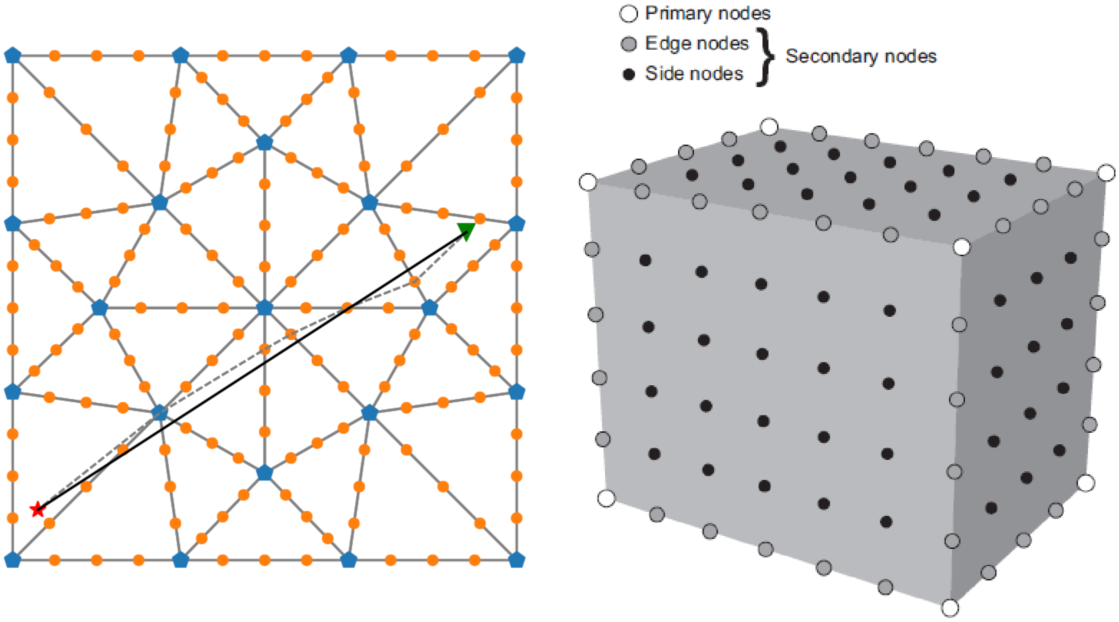
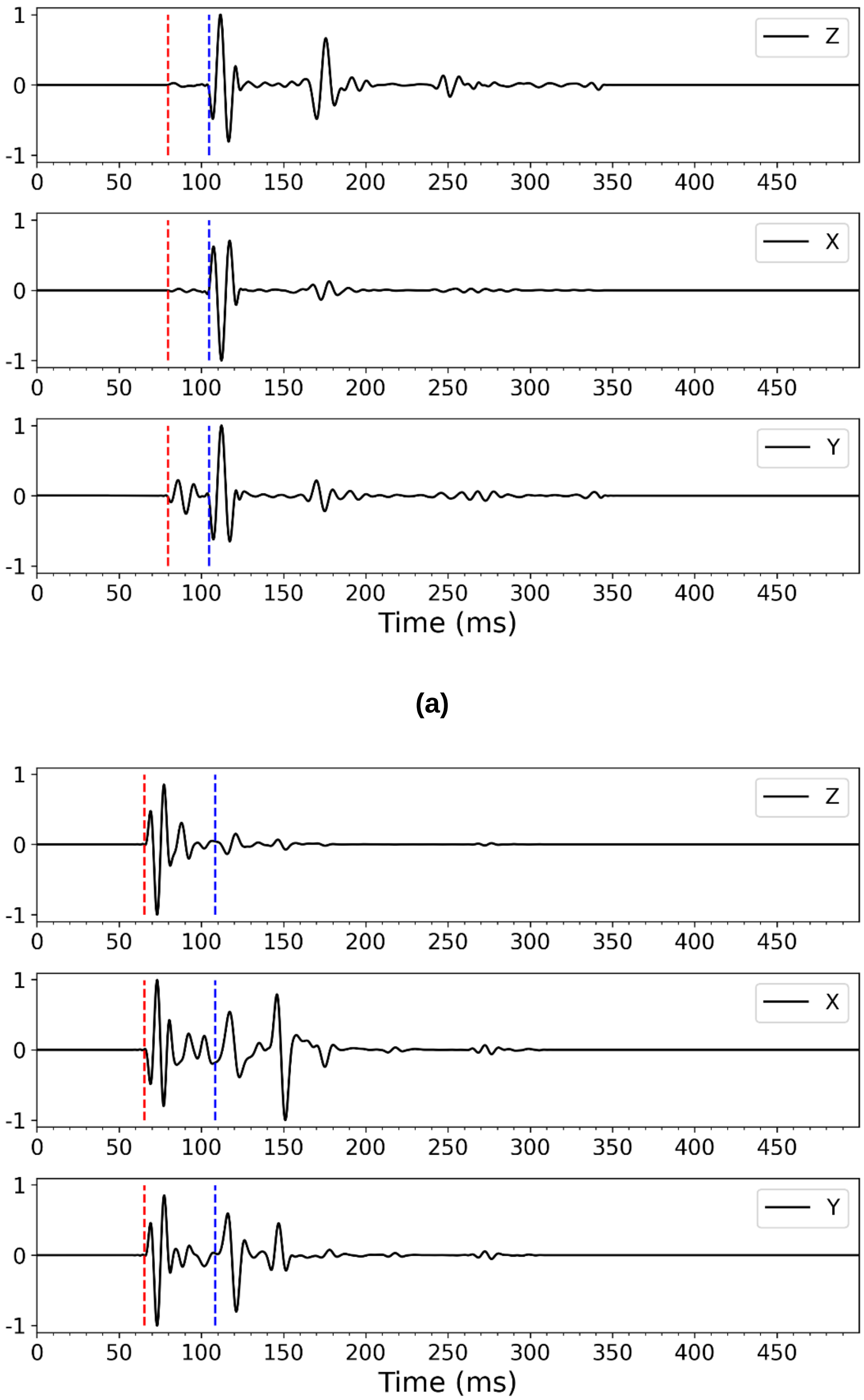


Figure 2-9: A cell with the location of primary and secondary nodes for SPM calculation: 2D (left) and 3D (right) (Giroux et al., 2013).

2.2.6 Noise-Free Synthetic Seismograms

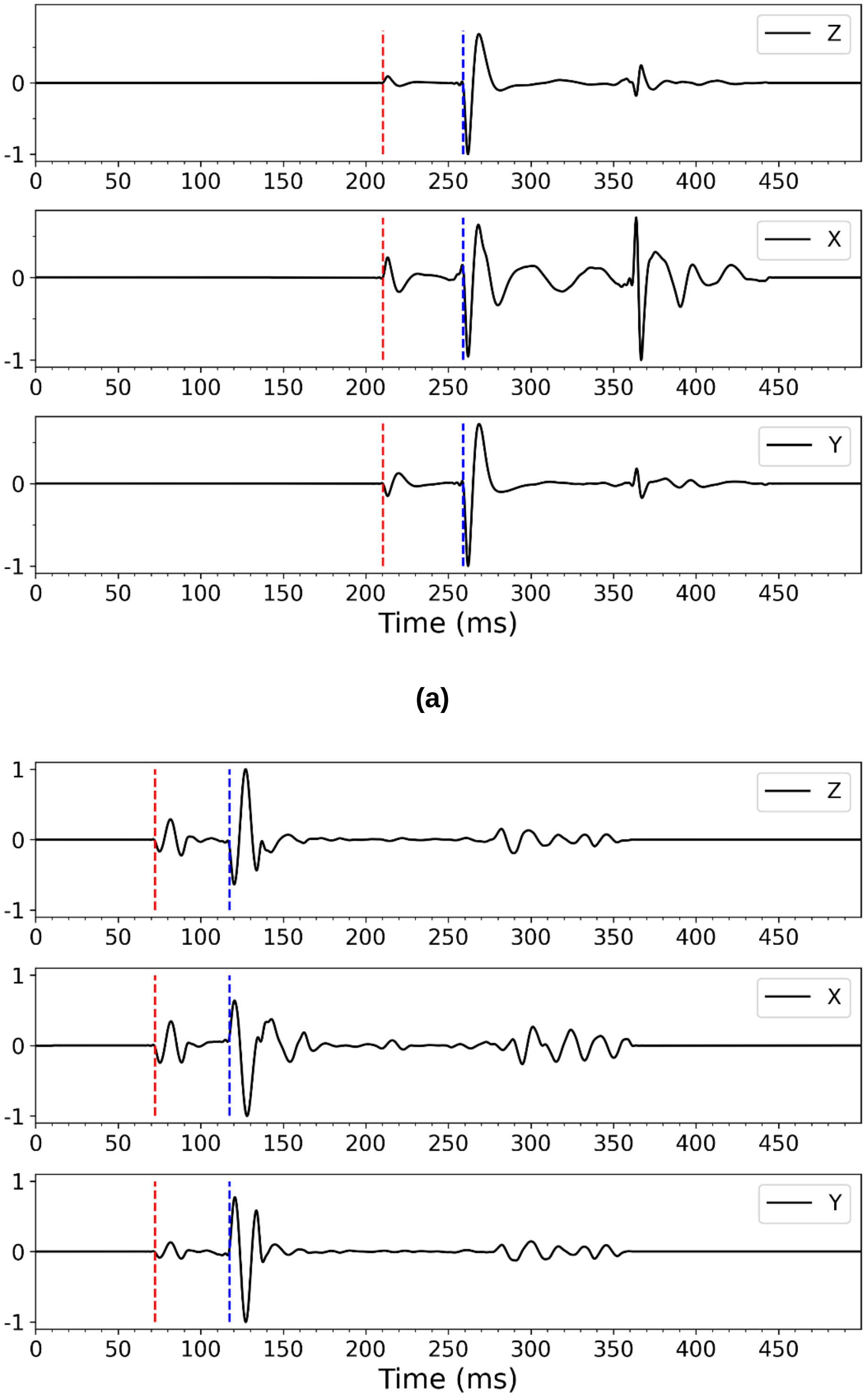
Visual inspection of the modeling results indicates that the quality of the data and information presented in the synthetic database can be considered satisfactory. The samples exhibit a strong correlation and consistency between the microseismic waveforms and the corresponding arrival times of P- and S-waves, as shown in Figure 2-10 and Figure 2-11 for all four different types of source mechanisms. A quantitative assessment of the reliability of the simulated data proves to be challenging due to the lack of analytical solutions for the specific velocity models employed. Nonetheless, the relative error is anticipated to remain below 1%, owing to the precision provided by the softwares used for the synthetic data simulation and travel-time calculation (SeisCL and ttcrpy). Furthermore, the synthetic seismograms exhibit a complexity comparable to the actual data recorded at the mine. This includes the presence of interferences, distortions, secondary arrivals, and wave conversions, as depicted in Figure 2-12. These phenomena can be attributed to variations in velocity models resulting from stopes within the subdomain of the considered mine.



(a)

(b)

Figure 2-10: Accuracy and consistency of the modeling and ray-tracing for two different source mechanisms used in synthetic data modeling: a) Single force (SF) and b) Isotropic. The red and blue vertical dashed lines are the P- and S-wave arrival times, respectively.



(a)

(b)

Figure 2-11: Accuracy and consistency of the modeling and ray-tracing for four different source mechanisms used in synthetic data modeling: a) Double-couple and (DC) and b) Compensated linear vector dipole (CLVD). The red and blue vertical dashed lines are the P- and S-wave arrival times, respectively.

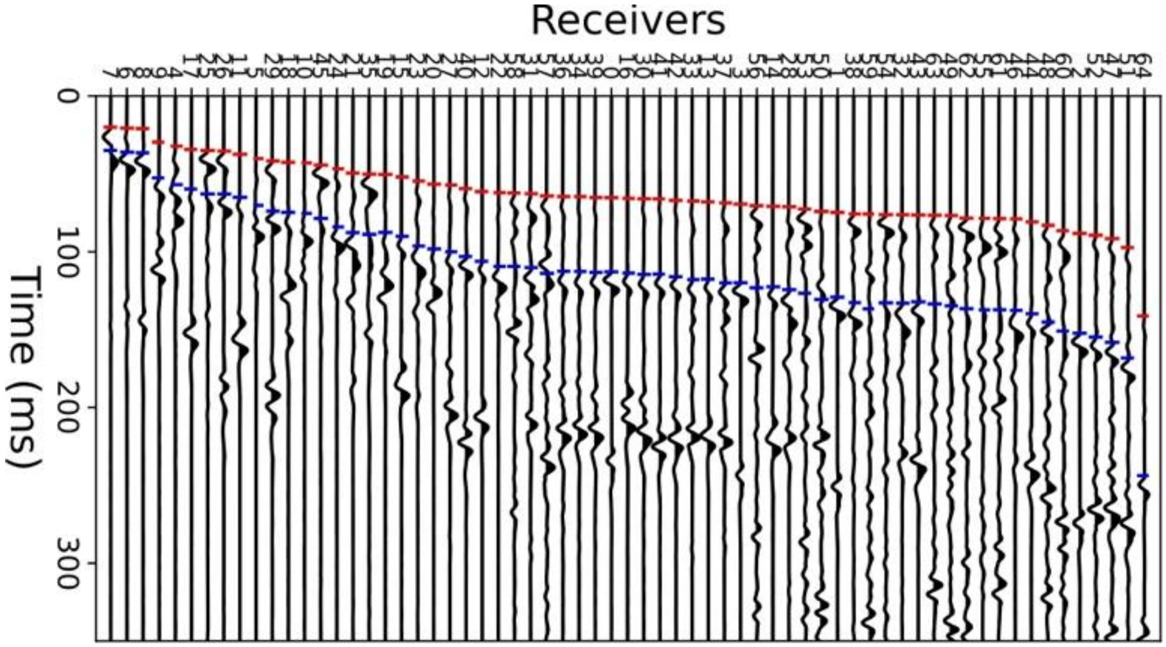


Figure 2-12: Synthetic seismograms for an array of 65 three-component sensors produced by FDTD elastic wavefield modeling for a single event: The noise-free Z-component. The red and blue vertical lines are P- and S-wave arrival times, respectively.

The entire workflow is summarized in Figure 2-13, which illustrates the specific task of generating and preparing the synthetic microseismic dataset, encompassing wavefield simulation and labeling. The resulting simulated synthetic dataset comprises 272,000 three-component labeled waveforms saved in SEGY format, where each file corresponds to one of the three components of the modeled wavefield. Consequently, the dataset comprises 1,700 SEGY files, featuring 160 sensors, each labeled with the corresponding event index and source mechanism.

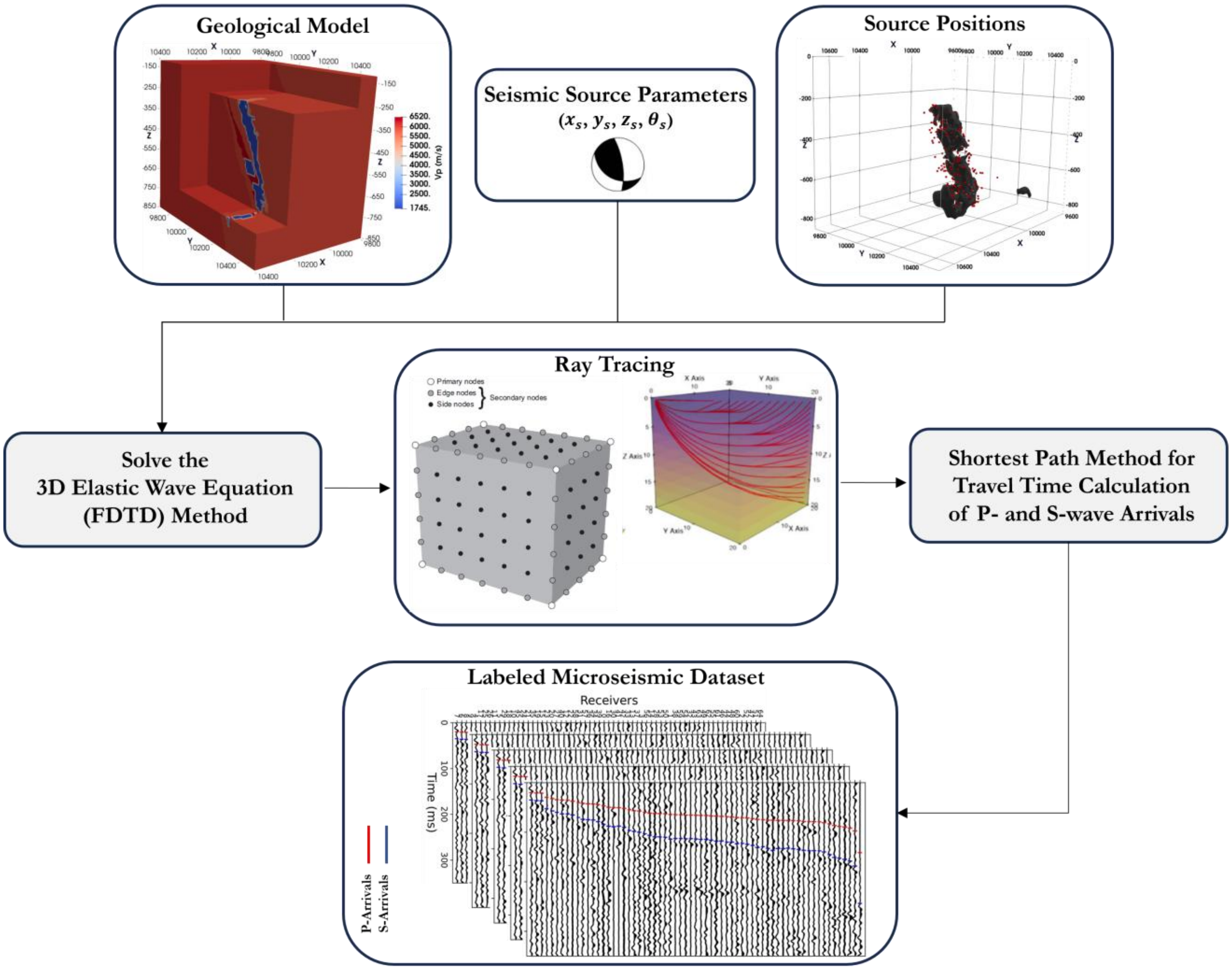


Figure 2-13: Workflow for seismogram modeling and arrival time labeling.

2.2.7 Adding Noise to Synthetic Seismograms

The simulated raw synthetic seismogram data is free of noise; therefore, it is necessary to introduce some noise to better emulate field conditions. In practice, noise can be represented as WGN or field noise extracted from actual recordings. In signal processing, white noise is characterized by its random signal properties, which maintain equal intensity across various frequencies, resulting in a constant power spectral density (Proakis, 2001). In discrete terms, white noise consists of a collection of samples that are independent of each other and come from the same probability distribution. WGN is a specific type of white noise generated using a random number generator, ensuring that all samples follow a Gaussian distribution. The noise level is usually assessed by the SNR, which is used in signal processing to quantify the level of the desired signal relative to the background noise. It is commonly expressed in decibels (dB) and is defined as (Allen, 1978):

$$SNR_{dB} = 10log_{10}\left(\frac{P_{Signal}}{P_{Noise}}\right) \tag{2.7}$$

where P_{Signal} represents the average power of the signal and P_{Noise} denotes the average power of the noise component. Alternatively, in waveform analysis, SNR can be computed as the ratio of the signal root-mean-square (RMS) amplitude to the noise.

$$SNR_{dB} = 20log_{10}\left(\frac{RMS_{Signal}}{RMS_{Noise}}\right) \tag{2.8}$$

Figure 2-14 shows arrays of one-component seismograms for a source location and 65 three-component sensors. The SNR values are set to $log_{10}(SNR) = 3.995$ and $log_{10}(SNR) = 2.6$. Notice that higher noise levels conceal the P-wave arrivals due to their lower amplitudes in microseismic data.

To enhance the realism of seismic data in our study, we also incorporated colored noise into the synthetic datasets that resembles some types of stochastic noise found in real seismic data. This process involves several key steps to ensure that the added noise closely mimics the characteristics of actual seismic noise encountered in field conditions. Detailed steps for adding colored noise to the dataset are as follows:

1. Generation of White Gaussian Noise:

The first step involves generating white Gaussian noise, characterized by a uniform power spectral density across all frequencies, resulting in a flat frequency response. In this context, we generate the white Gaussian noise using a pseudorandom number generator to ensure a statistically normal distribution with a mean of zero.

2. Application of a Low-Pass Filter:

To transform the white noise into colored noise, we apply a low-pass filter. This filter is designed to attenuate high-frequency components while preserving lower frequencies, which are typically more representative of actual seismic signals. The filtering process effectively alters the frequency spectrum of the white noise, reducing its high-frequency content and resulting in a smoother signal that resembles some types of stochastic noise found in real seismic data. This is accomplished by low-pass filtering the white noise, using a 4th order Butterworth filter with 0.5 normalized coefficients. The normalized coefficient of `0.5` indicates the cutoff frequency of the low-pass filter relative to the Nyquist frequency (half the sampling rate), allowing the filter to attenuate high-frequency components while preserving lower frequencies that are typical of seismic signals. This is essential for shaping the frequency response of the noise to better reflect the characteristics of actual seismic noise found in field data.

3. Scaling and Integration with the Original Signal:

Once we have the colored noise, the next step is to scale it appropriately to ensure its amplitude is consistent with that of the synthetic seismic signals. This scaling is crucial for maintaining the integrity of the underlying seismic events while effectively adding the desired noise components.

Finally, the scaled colored noise is added to the original synthetic waveforms, resulting in new datasets that incorporate the complexity of noise expected in operational seismic environments. The scaling factors of ‘0.05’, ‘0.06’ and ‘0.1’ are used to control the amplitude of the added colored noise, ensuring that it is proportionate to the peak amplitude of the original seismic data. This factor effectively influences the SNR of the resulting waveforms, helping to balance the expected noise level with the underlying seismic events. This integration enables a more robust assessment of the network performance under conditions that reflect real-world challenges in seismic phase detection.

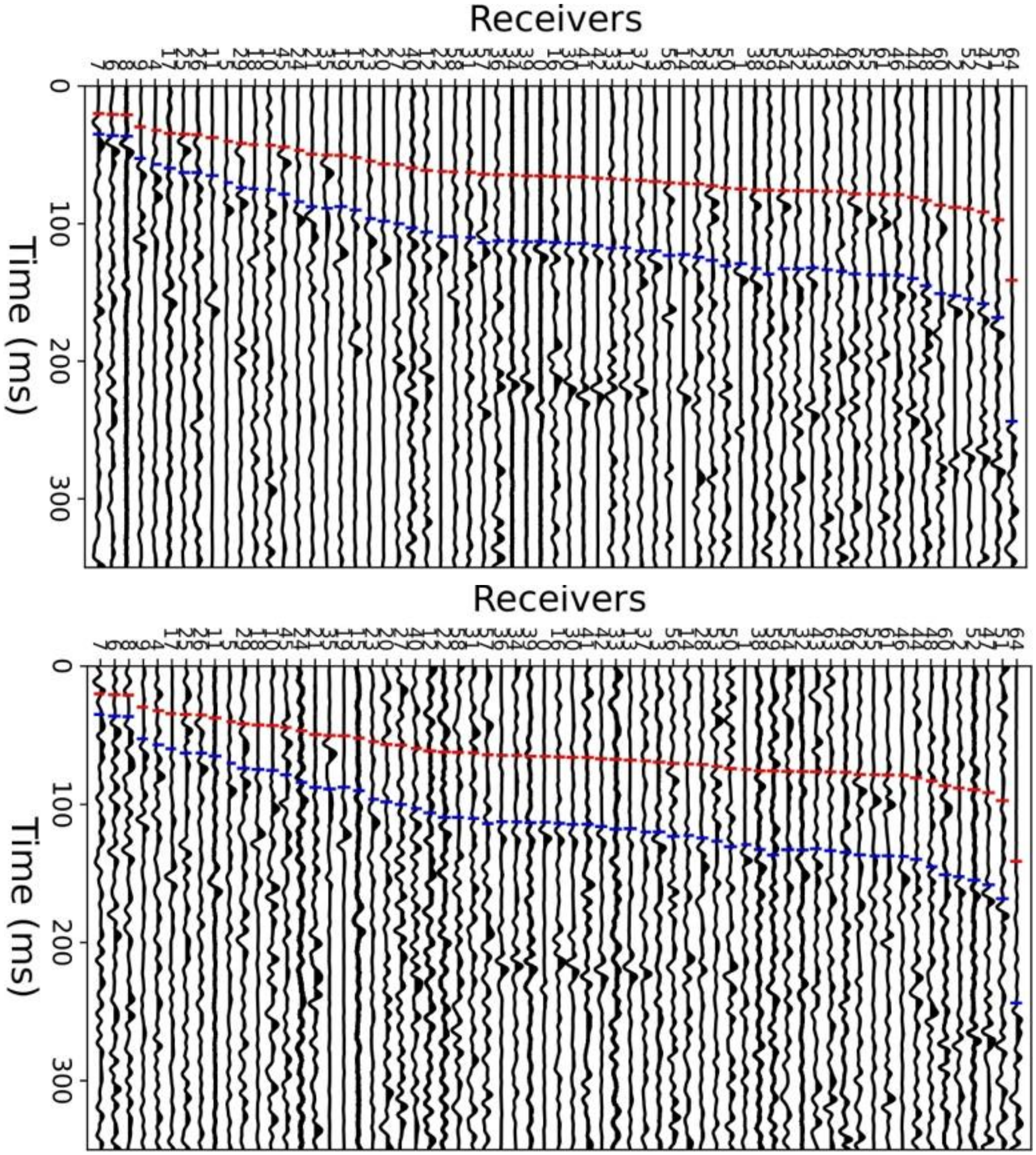


Figure 2-14: Synthetic seismograms for arrays of 65 three-component sensors produced by FDTD elastic wavefield modeling for a single event: The noise-added Z-component $log_{10}(SNR) = 3.995$ (on the top) and $log_{10}(SNR) = 2.6$ (on the bottom). The red and blue vertical lines are P- and S-wave arrival times, respectively.

2.3 Conclusion

For the forward modeling experiments conducted with SeisCL, a total of 1,700 simulation runs were performed. Each run required approximately 30–40 minutes, resulting in a cumulative computation time of about 900 hours, utilizing four GPUs in parallel. The modeling domain was discretized with a temporal step of 0.012 ms and a spatial step of 0.75 m, yielding nearly 500 million grid cells in total. From these simulations, approximately 272,000 three-component, noise-free seismograms were generated for subsequent analysis (Table 2-3).

The PassiveSeis dataset is compiled by collecting and storing the generated synthetic seismic data, along with the relevant metadata, in separate files. This database is freely accessible through https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP3/MVVP3M.

Table 2-3: Different information regarding the dataset simulation

Parameter	**Value**
Number of forward simulations	1,700
Computation time per run	30–40 minutes
Total computation time	~900 hours (4 GPUs in parallel)
Time step	0.012 ms
Spatial step	0.75 m
Modeling domain size	~500,000,000 grid cells
Number of seismograms generated	272,000 (3-component, noise-free)

The organization of the dataset is as follows:

1) Seismic Data: A total of 272,000 three-component (3C) seismograms have been generated. These synthetic records are saved in SEGY format, with three files dedicated to each simulated event. Each file represents one of the three components (X, Y, or Z) of the modeled seismic wavefield. Consequently, the dataset consists of 5,100 SEGY files, each labeled with event indices, components, and source mechanisms.
2) Arrival Times: The travel times for P- and S-waves are documented in text files, arranged in order of the receiver. A slight variation in travel times is observed compared to other source mechanisms. Thus, the arrival time files are categorized by event indices, seismic phases, and source mechanisms.

3) Metadata Files: These files provide details about the simulated seismic events, including indices, frequency, source time functions, and receiver configurations.

To summarize the outcomes of this chapter, synthetic and field microseismic data at various noise levels will be available for training and testing the deep learning networks. The results of training and testing the models are discussed in the following chapters.

3 DETECTION OF INDUCED MICROSEISMICITY WITH DEEP LEARNING

3.1 Introduction

This chapter introduces the theory of deep learning and the specific algorithms implemented in this study, with a focus on the convolutional neural network. The initial section outlines deep learning principles, addressing its advantages over machine learning in terms of feature extraction, redundancy and the capability to classify seismic events, including those of low magnitude that other approaches may fail to identify. Following that, the basics of convolutional neural networks are discussed, focusing on 1D CNNs used for time-series forecasting and the structure of PhaseNet, which is employed as a deep learning method to predict the P- and S-wave arrivals. Lastly, the predictions of the model trained on synthetic seismograms will be presented along with the evaluation metrics.

3.2 Deep Learning

Deep learning is a branch of machine learning, which itself is a part of artificial intelligence (Figure 3-1) (Aminzadeh et al., 2022). Artificial intelligence broadly encompasses methods that allow computers to replicate human-like behaviours. Machine learning consists of a collection of algorithms trained on data to facilitate these processes. Deep learning, in particular, is a form of machine learning that draws inspiration from the architecture of the human brain (LeCun et al., 2015).

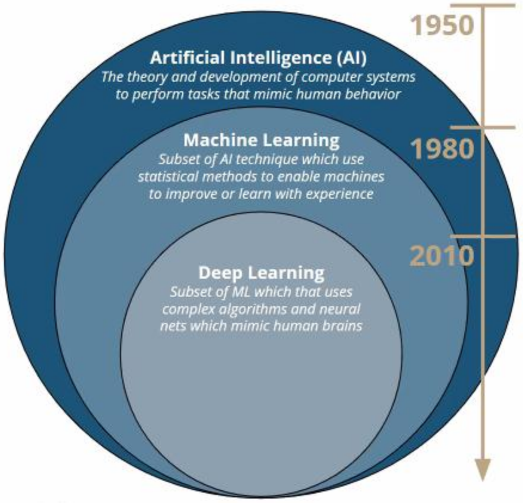


Figure 3-1: Artificial intelligence, machine learning, and deep learning (adapted from “Machine learning and deep learning” in Electronic Markets, inspired by (Goodfellow et al., 2016)).

Deep learning algorithms aim to conduct reasoning similar to that of humans by continually examining data organized within a specific logical framework. To accomplish this, deep learning employs a layered arrangement of computational units known as neural networks (NN), as in Figure 3-2. The first layer takes in the input $\vec{x}$, which consists of the data that the neural network uses for learning. The input layer contains a number of neurons that match the number of entries in the vector $\vec{x}$. This means that each input neuron corresponds to a specific element within the vector. The final layer is known as the output layer, which produces a vector $\vec{y}$, that denotes the neural network result. The elements in this vector correspond to the values of the neurons within the output layer. To generate a prediction vector $\vec{y}$, the network needs to execute specific mathematical calculations within the layers located between the input and output layers, which are referred to as the hidden layers.

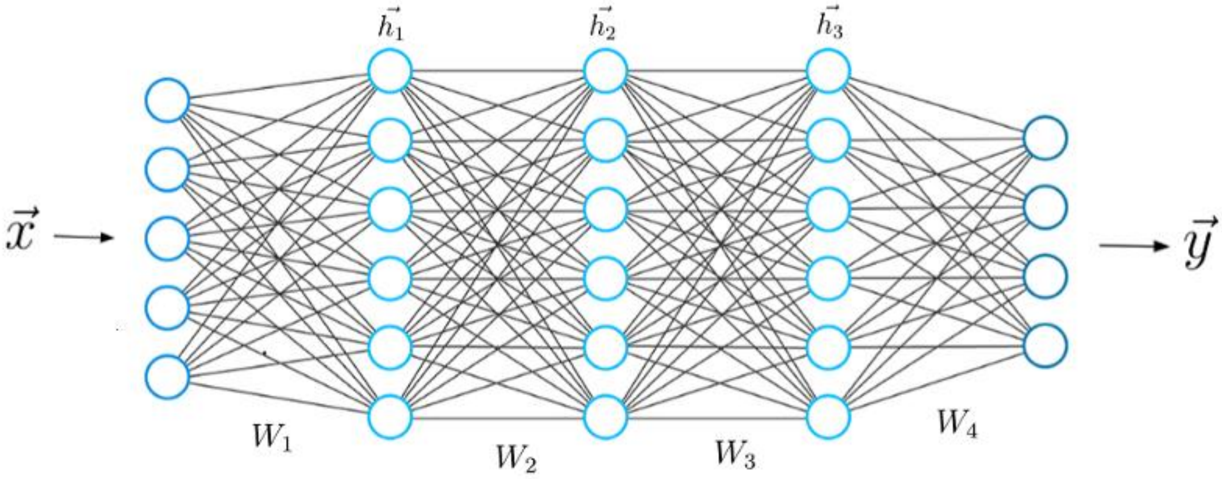


Figure 3-2: Architecture of a feedforward neural network. This diagram illustrates an input vector $\vec{x}$ with associated weights W_1, W_2 and W_3. The network processes the input through hidden layers represented by the neurons $\overrightarrow{h_1}$, $\overrightarrow{h_2}$ and $\overrightarrow{h_3}$, which apply activation functions to derive intermediate representations. The network ultimately produces an output vector $\vec{y}$ by propagating the input data through these interconnected layers, highlighting the relationships between inputs and outputs as influenced by the weights. Each parameter plays a crucial role in determining the network's ability to learn from data and make predictions.

The architecture of neural networks draws inspiration from the human brain, particularly in the way we recognize patterns and categorize information. Just as our brains process new information by comparing it with existing knowledge, neural networks are designed to perform similar functions in data analysis. Each layer of a neural network acts as a filter that processes information, moving from broad concepts to increasingly detailed distinctions. This hierarchical structure not only enhances the model ability to detect and generate results but also establishes a clear logical progression from one layer to the next, facilitating more refined output.

This hierarchical approach contributes significantly to the versatility of neural networks, enabling them to effectively tackle a range of tasks such as clustering, classification, and regression. For instance, when dealing with unlabeled data, these networks can autonomously organize information based on similarities between samples. In classification scenarios, labeled datasets can be leveraged to train the network, allowing for the categorization of samples into distinct classes. Importantly, while neural networks excel at replicating tasks performed by traditional machine learning algorithms, classical algorithms cannot manage the complex tasks that neural networks tackle. This distinction not only highlights the unique capabilities of artificial neural networks but also emphasizes their role in advancing deep learning methodologies that address challenges beyond the scope of classical machine learning.

A key advantage of deep learning over traditional machine learning is its ability to eliminate the need for manual feature extraction. Historically, before the rise of deep learning, practitioners depended on conventional techniques such as decision trees, support vector machines (SVMs), Bayes classifiers, and logistic regression. These methods, often referred to as "flat algorithms," require extensive preprocessing to transform raw data, whether it be text files or images, into formats suitable for analysis. This intricate preprocessing step, known as feature extraction, demands a deep understanding of both the specific problem area and the data itself, often requiring numerous iterations for refinement.

In contrast, deep learning models capitalize on artificial neural networks that bypass the need for manual feature extraction entirely. These networks are capable of autonomously learning direct and implicit representations of raw data. As data progresses through the various layers of an artificial neural network, increasingly abstract and compressed representations are generated. This compression is pivotal for achieving outcomes, such as effectively categorizing input data into distinct classes. For example, Figure 3-3 illustrates this process, showcasing how the transformation from raw input to compressed forms leads to insightful categorizations. By establishing this logical progression, the connections between these concepts become clearer, elucidating the comprehensive advantages of deep learning over traditional methods and reinforcing the significance of the neural network’s design in this evolutionary landscape.

Essentially, we can assert that feature extraction is inherently integrated into the workings of an artificial neural network. As the training progresses, the neural network fine-tunes this aspect to achieve the most effective abstract representation of the input data. This suggests that deep learning models require minimal to no manual intervention to perform and optimize the feature extraction process.

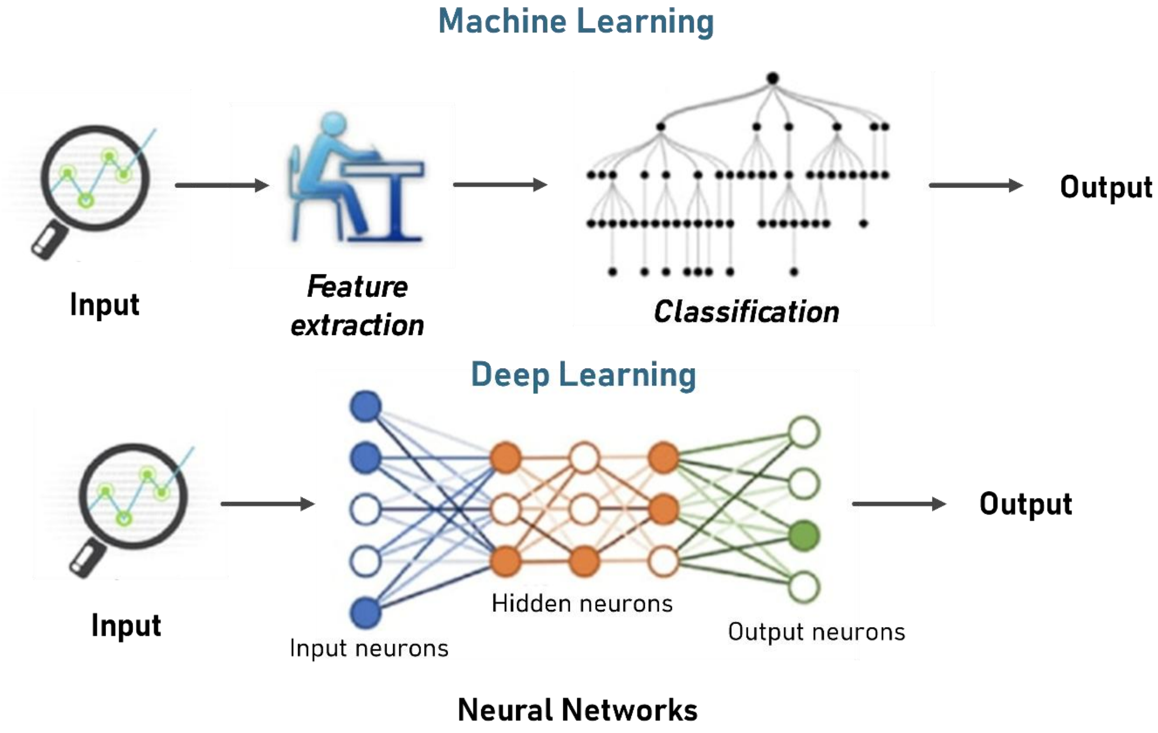


Figure 3-3: Comparison of ML and DL regarding the feature extraction procedure.

3.2.1 Convolutional Neural Networks (CNNs)

CNNs are a distinct type of artificial neural network specifically designed to handle data that is organized in a grid-like structure, such as one-dimensional time-series data or two-dimensional image pixels (Goodfellow et al., 2016). The origins of CNNs date back to the 1990s, when they were initially employed for text recognition tasks (LeCun et al., 1998). Different CNN architectures have been developed so far, with the most common ones being ResNet (He et al., 2016), UNet (Ronneberger et al., 2015), DensNet (Huang et al., 2017), RegNet (Radosavovic et al., 2020), and VGGNet (Simonyan et al., 2014). CNNs are particularly utilized in time-series forecasting due to their efficient training speeds, straightforward model designs, and independence from prior sequential dependencies (Wang et al., 2019). In a recent study, four key benefits of CNNs for one-dimensional signals have been highlighted as follows (Kiranyaz et al., 2021):

- The forward and backward propagation processes involve basic array operations instead of complex matrix calculations.
- Superficial networks, comprising a limited number of hidden layers and neurons, can effectively tackle intricate tasks.

- With only a couple of hidden layers (about 2) and fewer than 50 neurons, these networks can be trained efficiently on standard CPU systems.
- Their minimal computational demands make them ideal for real-time applications that require low-cost implementation.

Regardless of the specific architecture of a CNN, all CNN models share common, fundamental components. These components include convolutional layers, activation functions, pooling layers, and fully connected layers, which will be explored in detail next.

Convolutional layer

A convolutional layer consists of kernels or filters that interact with segments of the input signal to produce feature maps. The discrete convolution operation between a time-series segment $x(t)$ and a kernel $w(a)$ is defined as (Goodfellow et al., 2016):

$$s(t) = (x * w)(t) = \sum_{-\infty}^{\infty} w(t-a) \tag{3.1}$$

where $*$ represents the convolution operator. For a multi-channel input time series at channel $c \left(\mathbf{x}_{l-1}^{(c)}\right)$, the m^{th} output feature map $\left(\mathbf{x}_{l}^{(m)}\right)$ at layer l can be expressed as follows:

$$\boldsymbol{x}_l^{(m)} = \sum_{c=1}^{C} \boldsymbol{W}_l^{(c,m)} * \boldsymbol{x}_{l-1}^{(c)} + \boldsymbol{b}_l^{(m)} \tag{3.2}$$

where $\mathbf{W}_l^{(c,m)}$ represents the m^{th} filter at the channel c, and $\mathbf{b}_l^{(m)}$ denotes the m^{th} bias vector, and C indicates the total number of input channels or features. The result from Equation 3.2 is then fed into an activation function to incorporate non-linearity. Two frequently used non-linear activation functions are the sigmoid function and the rectified linear unit (ReLU) (Goodfellow et al., 2016):

$$ReLu(x) = \begin{cases} x, & if\ x > 0 \\ 0, & otherwise \end{cases} \tag{3.3}$$

$$Sigmiod(x) = \frac{1}{(1 + \exp(-x))} \tag{3.4}$$

The ReLU activation function outputs zero for negative inputs and remains linear for positive inputs. It is computationally efficient and helps mitigate the vanishing gradient problem, making it a preferred choice in DNNs and CNNs (Goodfellow et al., 2016). However, ReLU can suffer from the dying ReLU problem, where neurons become inactive if they receive only negative inputs (He et al., 2015). In contrast, the Sigmoid function maps inputs to a smooth range between 0 and 1, making it useful for probabilistic outputs. However, it suffers from the vanishing gradient problem, leading to slow learning in deep networks. While ReLU is commonly used in hidden layers due to its efficiency, Sigmoid is typically found in output layers for binary classification tasks. The two activation functions can be illustrated in Figure 3-4.

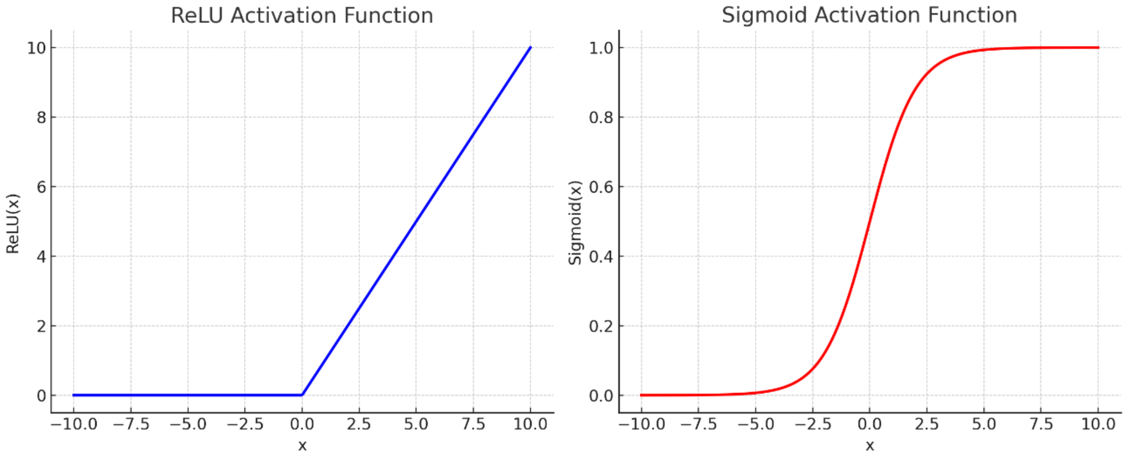


Figure 3-4: The ReLU (left) and sigmoid (right) activation functions.

Pooling layer

A pooling layer plays a crucial role in CNNs by reducing the spatial dimensions of feature maps (width and height) while preserving the most significant information. This process enhances the computational efficiency of the network and lowers the chances of overfitting (LeCun et al., 2015). Three main types of pooling layers are commonly used in CNNs:

- Max Pooling: This method selects the maximum value from each region of the feature map, effectively reducing its size while retaining the most prominent features. For example, in a 2x2 filter, the maximum value within each 2x2 section is taken.
- Average Pooling: Unlike max pooling, average pooling calculates the average value of each region of the feature map. This method smooths out the feature map, providing a broader representation of the data by averaging the values in the specified region.
- Global Average Pooling: This pooling technique takes the average of all values in each feature map, resulting in a single value per feature map. Global average pooling is often employed before the final classification layer in CNNs, which helps reduce the number of parameters and prevent overfitting.

Each of these pooling methods offers unique advantages depending on the specific architecture and objectives of the neural network.

Fully connected layer

Once a sequence of convolutional layers, non-linear activation functions, and pooling layers have been arranged, a fully connected layer, also known as a dense layer, commonly serves as the output layer in a CNN. In a fully connected layer, every neuron is connected to every neuron in the previous layer. This structure allows the network to learn complex relationships in the data, as each neuron can receive inputs from all activations of the previous layer and apply a learned weight to each input. The output vector of a fully connected layer $\mathbf{x}_l$ is:

$$\mathbf{x}_l = \sigma(\mathbf{W}_{\boldsymbol{l}}\mathbf{x}_{l-1} + \boldsymbol{b}_l) \tag{3.5}$$

In this context, the activation function σ which introduces non-linearity into the model, allowing the neural network to learn complex patterns in the data, is applied to the input vector $\mathbf{x}_l$, which is multiplied by the weight matrix $\mathbf{W}_{\boldsymbol{l}}$ and added to the bias vector $\boldsymbol{b}_l$.

The fully connected layer is crucial because it combines the features learned in the previous layers to make final predictions or classifications. This layer typically comes at the end of a network and summarizes the learned representations into the desired output format, such as classification scores for various classes in a classification task. While powerful, fully connected layers have a significant downside: they can introduce a large number of parameters, especially in scenarios with high-dimensional input data. This can lead to overfitting and increased computational expense. Therefore, it is common practice to use them judiciously, often following layers that reduce the dimensionality, such as convolutional and pooling layers (Goodfellow et al., 2016).

In general, neural networks establish a nonlinear mapping function comprising numerous parameters (potentially reaching several million in deep networks), which transform a set of input values, such as the amplitudes of a seismogram or engineered features, into a specified output (Figure 3-5) (Ross et al., 2018). This output can represent a prediction of a continuous variable, like a phase onset time, or a classification outcome, such as determining whether the initial movement is upward or downward. The mapping function is structured in sequential layers of neurons, with each neuron performing a simple operation on the incoming data and relaying the results to the subsequent layer. The coefficients associated with this mapping function are optimized through empirical methods using extensive datasets, ensuring that a specific set of input values yields an output that closely aligns with the desired outcome, thereby achieving accurate predictions or correct classifications across a broad validation dataset.

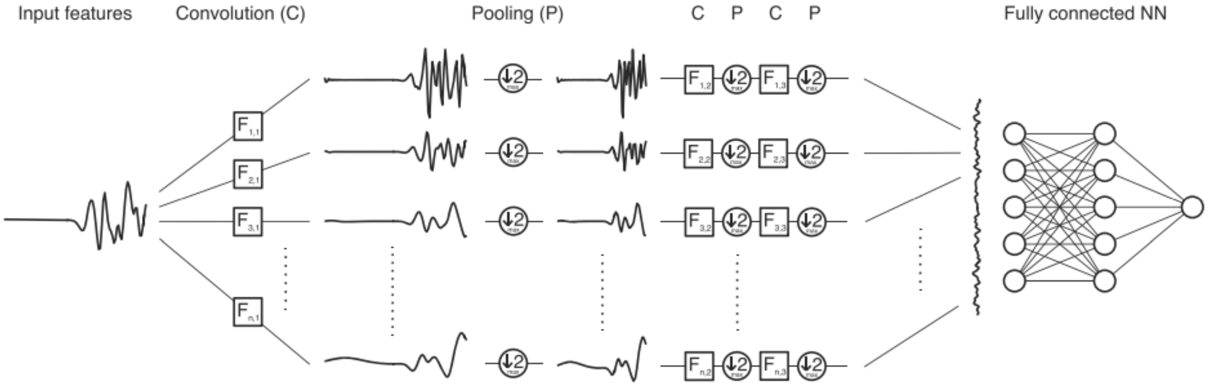


Figure 3-5: A CNN workflow for arrival time picking containing convolutional, pooling and fully-connected layers (Ross et al., 2018). The CNN acts as a trainable feature extraction system that collaborates with a fully connected network to handle classification and regression tasks.

3.2.2 Convolutional Neural Networks Training

Training a CNN involves several key steps to optimize the network parameters to minimize a given loss function. The process is typically structured as follows (Figure 3-6):

Data preparation

The first step is to prepare the dataset, which includes the process of identifying and extracting relevant characteristics or attributes from the input data, such as images or time-series data, that can help the model learn effectively and labeling. Input values, such as images or time-series data, are paired with corresponding target outputs (e.g., class labels or regression values). Data augmentation can be applied to increase the size and diversity of the training dataset, thereby enhancing the model generalization capabilities.

Forward propagation

During forward propagation, input data are passed through the network layer by layer. Each layer, consisting of convolutional, activation, pooling, and fully connected layers, transforms the input data. The output from the final layer is typically a probability distribution over classes (in classification problems) or a continuous value (in regression tasks).

Loss calculation

After obtaining the model predictions, the next step is to calculate the loss using a loss function, which measures the discrepancy between the predicted outcomes and the actual target values. Common loss functions include cross-entropy for classification tasks and mean squared error (MSE) for regression tasks. With training pairs (x_i, y_i) consisting of input values x_i and corresponding targets y_i, the optimal parameters, which include weights and biases, can be determined by minimizing the loss function L, which, for different approaches, might differ. Mathematically, we have

$$\underset{\theta}{argmin} \sum_i L\left(f_\theta(x_i), y_i\right) \tag{3.6}$$

where f_θ is a parametric family of mappings (e.g., CNNs) under the loss function L and θ is the optimal parameter, which includes the weights and biases (Lehtinen et al., 2018).

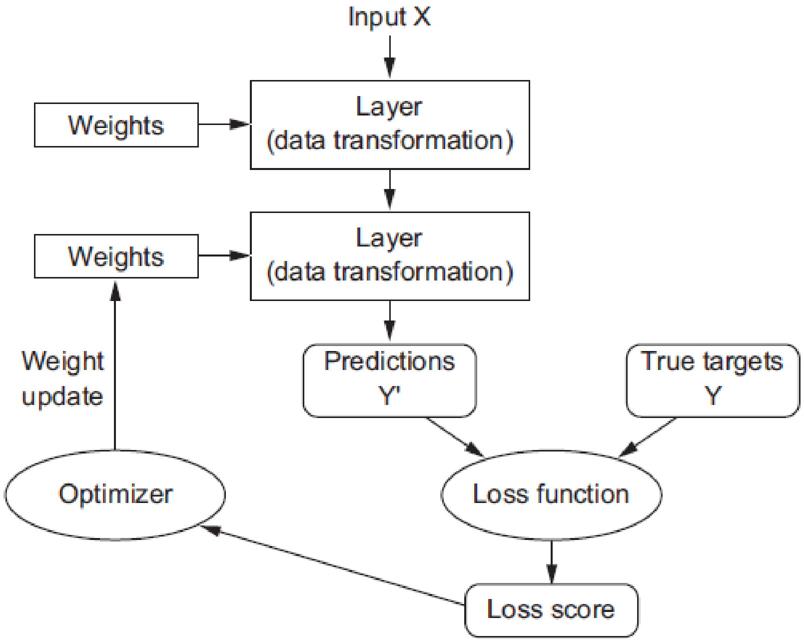


Figure 3-6: The workflow of training the neural network (Chollet, 2021).

Backpropagation

The network parameters (weights and biases) need to be updated to improve the model performance. Backpropagation calculates the gradient of the loss function with respect to each parameter by applying the chain rule. This step involves layer-by-layer computing derivatives, effectively propagating the error backward through the network (Rumelhart et al., 1986).

Parameter update

Optimization algorithms such as Stochastic Gradient Descent (SGD), Adam (utilized for the deep learning optimization in this study), or RMSprop (Root Mean Square Propagation) update the parameters based on the gradients computed during backpropagation. The parameters are adjusted to minimize the loss, gradually improving the model performance. The fundamental goal of Gradient Descent (GD) is to find the parameter values, e.g. weights in neural networks, $W^{(t)}$ at iteration (t) that minimizes the proposed loss function as:

$$W^{(t+1)} = W^{(t)} - \eta . \nabla L\left(W^{(t)}\right) \tag{3.7}$$

Where η is the learning rate that controls the step size and $\nabla L(W)$ is the gradient of the loss function L, i.e. the vector of partial derivatives of the loss function with respect to each parameter. This gradient indicates the direction and rate of change of the loss with respect to the parameters.

The learning rate plays a crucial role during the optimization. Although increasing the learning rate can accelerate convergence, it may also lead to oscillations in the cost function. Furthermore, if used without caution, there is a risk that the cost function diverges.

In Figure 3-7, the parameter b oscillates before ultimately settling at its minimum value. This behaviour is attributed to the nature of the gradient vector, which points perpendicularly to the contour lines. In an ideal scenario, as illustrated in Figure 3-7, we aim for a slow convergence in the vertical direction while achieving a rapid convergence horizontally.

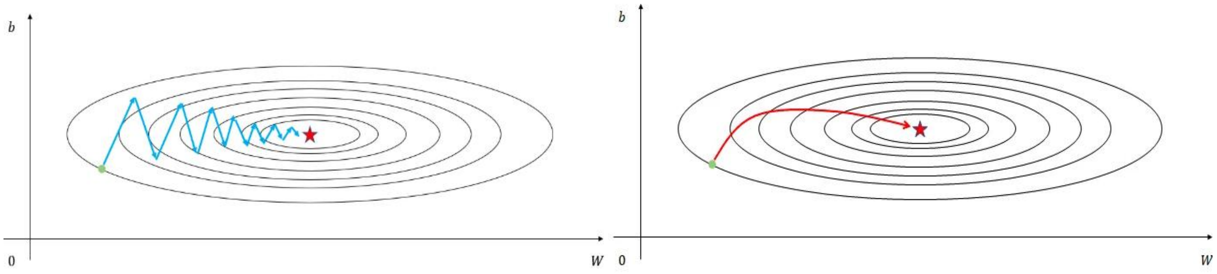


Figure 3-7: Gradient Descent convergence (left), and ideal convergence (right).

Optimizers that apply Gradient Descent with enhancements have been developed. Adam (Adaptive Moment Estimation) is an advanced optimization technique combining ideas from Gradient Descent with Momentum and RMSProp. In fact, Gradient Descent with Momentum serves as an enhancement. It adjusts the updates by slowing them down if the direction shifts frequently with each iteration and accelerates them when the direction remains consistent. RMSProp closely resembles Gradient Descent with Momentum, but instead, the learning rate is set for each parameter based on the magnitude of the gradient (Kingma et al., 2014).

With the Adam optimizer, the learning rate is adjusted for each parameter based on estimates of the first and second moments of the gradients. This means that parameters with high gradients will have their learning rates reduced, while parameters with lower gradients will have their learning rates increased, allowing for a more efficient convergence. The Adam optimizer relies on the concept of momentum by keeping an exponentially decaying average of past gradients (first moment) and the squared gradients (second moment). This helps the optimizer navigate the loss surface more effectively by smoothing out updates and reducing oscillations. Since the initial moments (first and second) are initialized to zero, they bias the estimates towards zero, especially in the early stages of training. Bias-correction terms are incorporated to adjust for this, ensuring that the estimates are accurate and stable early on. The first moment, which is similar to the velocity update in SGD with momentum, is defined as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla L(W) \tag{3.8}$$

The second moment tracks squared gradients to adjust learning rates dynamically, as in:

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\big(\nabla L(W)\big)^2 \tag{3.9}$$

The bias correction will be as follows:

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{3.10}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{3.11}$$

and the final Adam update term will be as follows:

$$W^{(t+1)} = W^{(t)} - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\widehat{m}_t \tag{3.12}$$

The developers of the ADAM algorithm recommend using the parameters of $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$ to achieve optimal performance (Kingma et al., 2014).

Iterative training

The forward propagation, loss calculation, backpropagation, and parameter update steps are repeated multiple times over epochs, with the training dataset often split into batches to facilitate efficient computation. The CNN performance is typically validated using a separate validation set to monitor for overfitting and ensure that the model generalizes well.

Hyperparameter tuning

Throughout the training process, various hyperparameters, such as learning rate, batch size, and network architecture, may need to be tuned. Techniques such as cross-validation can be used to identify the optimal set of hyperparameters that yield the highest performance on validation data.

All these structured approaches to CNN training are crucial for ensuring that the model learns meaningful patterns from the training data while remaining robust to unseen data during inference.

3.2.3 PhaseNet

PhaseNet is a deep neural network designed initially for earthquake monitoring, and a key element of the present work is to assess its performance for mining applications. PhaseNet is leveraging the principles of CNNs and the U-Net architecture, which was initially developed for image segmentation in the biomedical domain (Ronneberger et al., 2015). It is structured with contracting and expanding sections that mirror an encoder-decoder framework. The encoder comprises several convolutional modules that capture features from the input at different levels, followed by a max-pooling step for down-sampling. In contrast, the decoder features multiple deconvolutional modules for up-sampling and concatenation operations. This setup helps in projecting the high-resolution features obtained from the encoder into an up-sampled feature space, facilitating detailed classification (Zhu et al., 2019).

PhaseNet has been adjusted to work with seismic waveforms using one-dimensional (1D) inputs, enabling it to generate probability segmentations for P- and S-phases (Zhu et al., 2019). Figure 3-8 illustrates the detailed structure of the architecture, demonstrating the arrangement and interconnection of the various layers within the network. The input to PhaseNet consists of a three-component seismic waveform treated as a one-dimensional vector with three channels. This input produces three output channels, each representing the probability that a data point falls into one of three categories: P-arrival, S-arrival, or noise at each time step.

A key characteristic of the proposed approach is that no frequency filtering is applied to the input waveforms before they are fed into the network. The algorithm performance is tested using synthetic three-component seismograms simulated in the section 2.2 with various levels of added noise. The network is engineered so that peaks in the probability distributions correspond to the arrival times of both P- and S-waves.

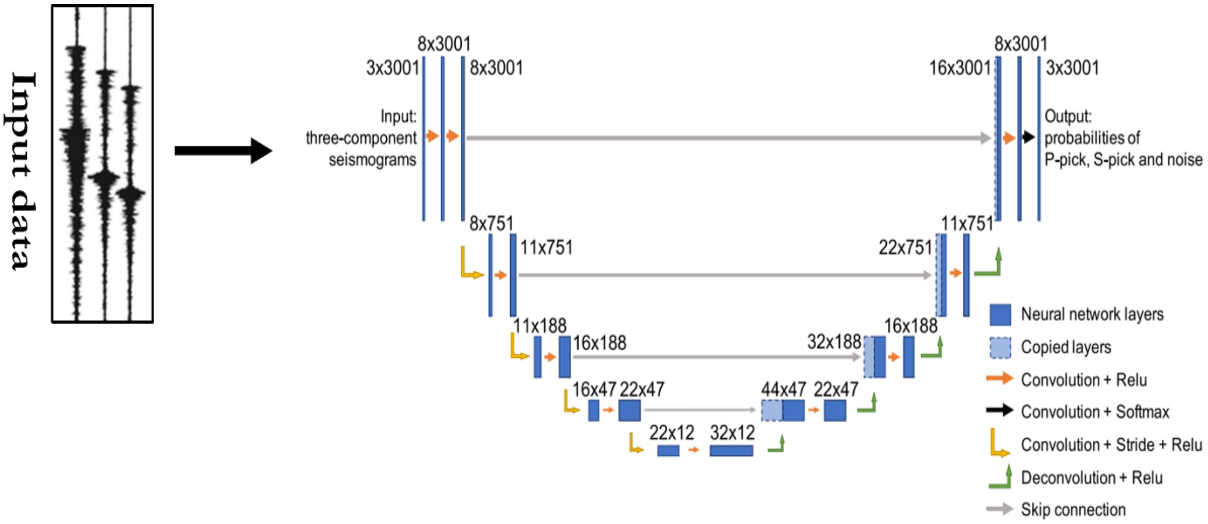


Figure 3-8: The PhaseNet architecture (Zhu et al., 2019).

When the microseismic data are input into PhaseNet, they undergo a sequence of processing phases. To be more specific, each phase utilizes 1D convolution operations and the ReLU activation function. The integration of skip connections enables the direct concatenation of outputs without passing through subsequent layers, thereby improving convergence during the training process. The convolution size is set to seven data points, while the down-sampling stride is configured to four data points, effectively reducing the channel length to one-quarter of its original size after each stride (Li et al., 2018). At the same time, the up-sampling deconvolution process (Noh et al., 2015) increases the condensed layers by a factor of four, restoring their original length, with padding added at both the front and back of each layer during convolutions to ensure that the input and output sequences remain of equal length. In the network, the final layer uses the softmax normalized exponential function to transform the output representation from the decoder into probabilities that fall within the range of [0, 1] (Goodfellow et al., 2016). In this context, the detection task is approached as a multi-class classification problem. Accordingly, the softmax function is defined as

$$q_i(x) = e^{z_i(x)} \Big/ \sum_{k=1}^{3} e^{z_k(x)} \tag{3.13}$$

where $i = 1, 2, 3$ indicates noise, P-arrival, and S-arrival, respectively. $z_i(x)$ is the unnormalized output from the final layer prior to applying the softmax function for the ith category. Subsequently, the cross-entropy function is utilized to compute the loss values L by comparing the true probability distribution with the predicted distribution (Murphy, 2012), i.e.

$$L(p,q) = -\frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n} p_i(x).\log\left(q_i(x)\right) \tag{3.14}$$

This loss function assesses the discrepancy between the two probability distributions. The terms $p_i(x)$ and $q_i(x)$ represent the actual values and the predicted probabilities for the ith category and xth data sample, respectively. The indices $i = \{1,2,3\}$ and $x = \{1,2,\dots,n\}$ are considered as category labels and sampling numbers.

Training a convolutional network requires significant computational resources; for this research, we employed an NVIDIA GPU. GPUs are particularly effective for the extensive parallel floating-point calculations required by deep learning. However, depending on the data volume to be handled, a multicore central processing unit may still be adequate for the forward prediction and classification of individual seismograms. Modern GPUs have large memory capacities, allowing for the simultaneous processing of numerous seismograms with minimal data transfer between the CPU and GPU. PhaseNet is executed within a TensorFlow environment and is trained and tested on a standard PC equipped with an Nvidia GeForce RTX 3060 GPU. The Adam optimizer is employed to enhance the model performance, and a dropout rate of 0.3 is applied in the dropout layers during the network training process.

3.2.4 Application of PhaseNet on Synthetic Dataset

The simulated dataset, which contains the single force mechanism, is used to train and test the DL model. The model is trained using 75% of the chosen dataset, which includes 72,570 three-component waveforms. 15% of the dataset is used as a validation dataset. Following training, the network is validated using 9,677 unseen 3C waveforms, which comprise 10% of the dataset. In these experiments, the synthetic dataset is sampled at a frequency of 4000 Hz, and arrival-time residuals under $\Delta t<0.002$ seconds are classified as true positives. In contrast, more significant residuals are categorized as false positives. Figure 3-9 shows a single 1D vector of the training waveform. The top-to-bottom plots represent the three-component seismic recordings.

White Gaussian noise and colored noise are added to the synthetic microseismic waveforms as mentioned in the previous chapter. The training, validation, and test datasets were converted into noise-contaminated data with the following signal-to-noise ratios: $\log_{10}(\mathrm{SNR})$ equal to 3.995, 2.6, 2.44, 2, 0.602, and 0.301.

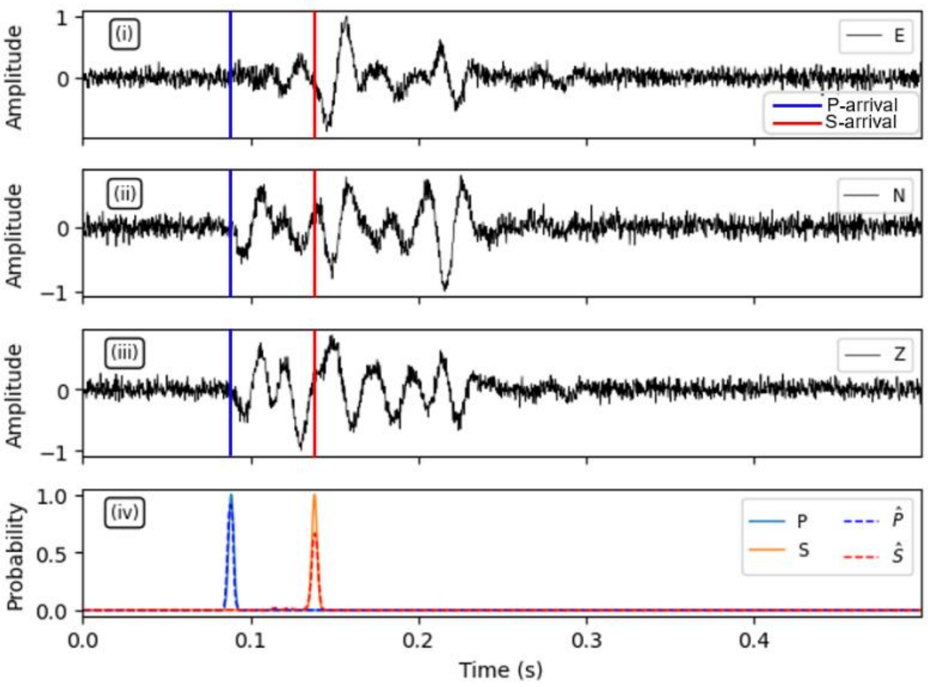


Figure 3-9: Example of synthetic microseismic data ($log_{10}(SNR)$=0.602), with the plots organized vertically to represent the three-component seismic recordings in east-west, north-south, and vertical orientations, along with the Gaussian predictions provided by PhaseNet. Blue and red vertical lines represent the initial arrival times of the P- and S-wave phases estimated using ray tracing. The final row displays the misfit of converted probability masks for P- and S-picks in the shape of a truncated Gaussian distribution.

Figure 3-10 illustrates the loss function and accuracy curves after optimizing the hyperparameters. The loss function for the training and validation datasets decreases consistently, resulting in a suitable misfit between the model predictions and the ground truth labels. This demonstrates the trained model effective learning and generalization capabilities.

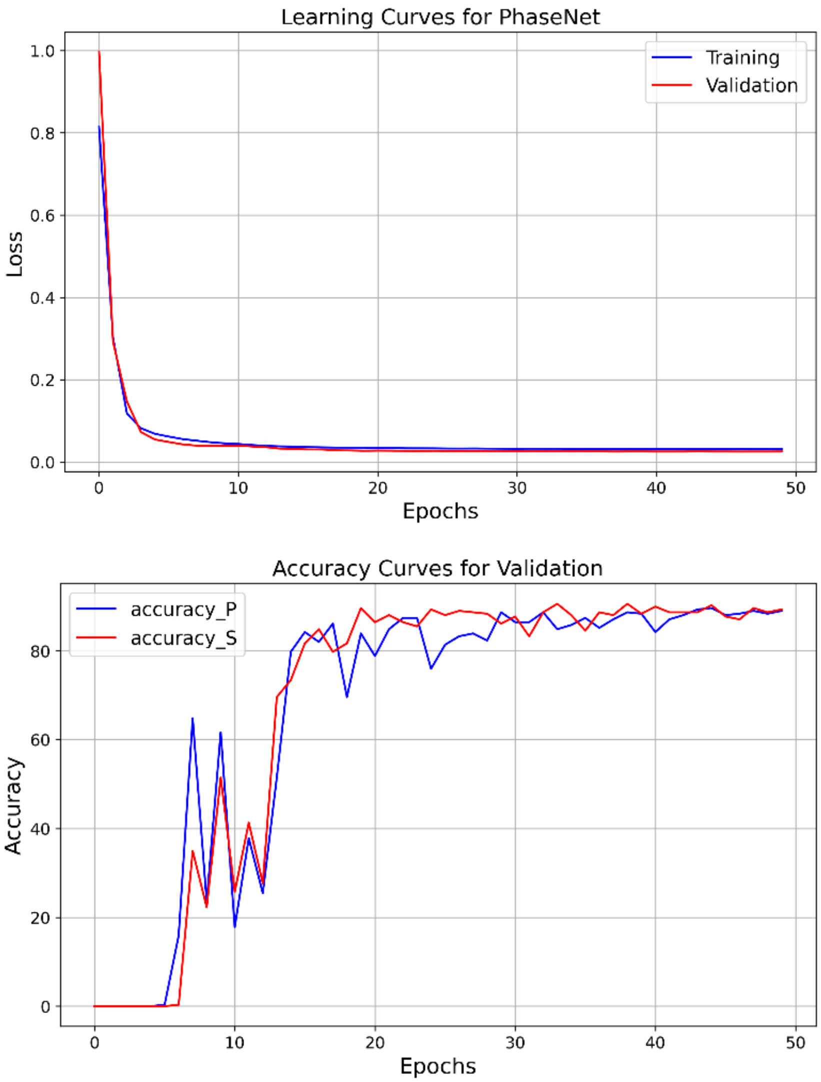


Figure 3-10: The loss functions observed during the training and validation processes (top) and the accuracy curves for P- and S-phases for the microseismic synthetic dataset (bottom).

Figure 3-11 displays the distribution of time residuals for the true versus predicted P- and S-wave arrival times, showing that both distributions exhibit no bias. Meanwhile, Figure 3-12 highlights the efficacy of microseismic event detection using confusion matrices in PhaseNet, indicating a low rate of misclassifying true events as the SNR in the synthetic dataset increases. In PhaseNet, a mask is applied around the manual selections to identify the P- and S-phases. This mask is a truncated Gaussian, characterized by a mean (μ) that indicates the arrival time and a standard deviation (σ) of 0.002 seconds. This method enables the synchronization of true arrival times with actual picks while incorporating a level of uncertainty.

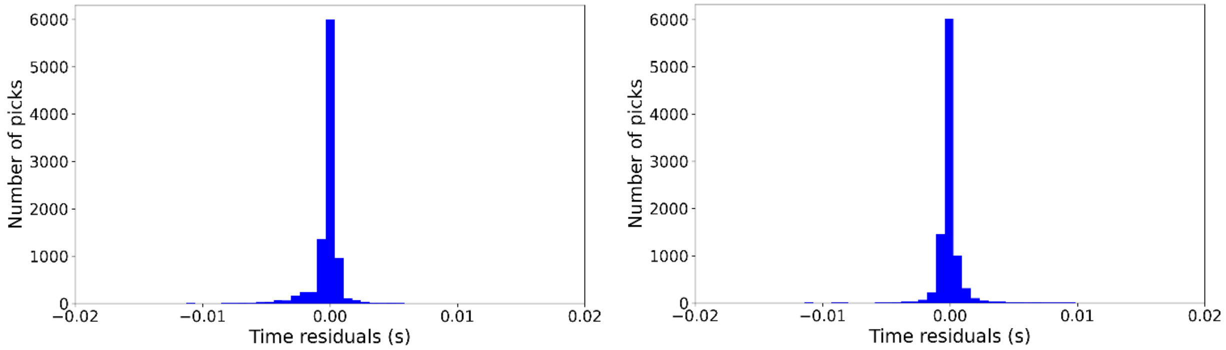
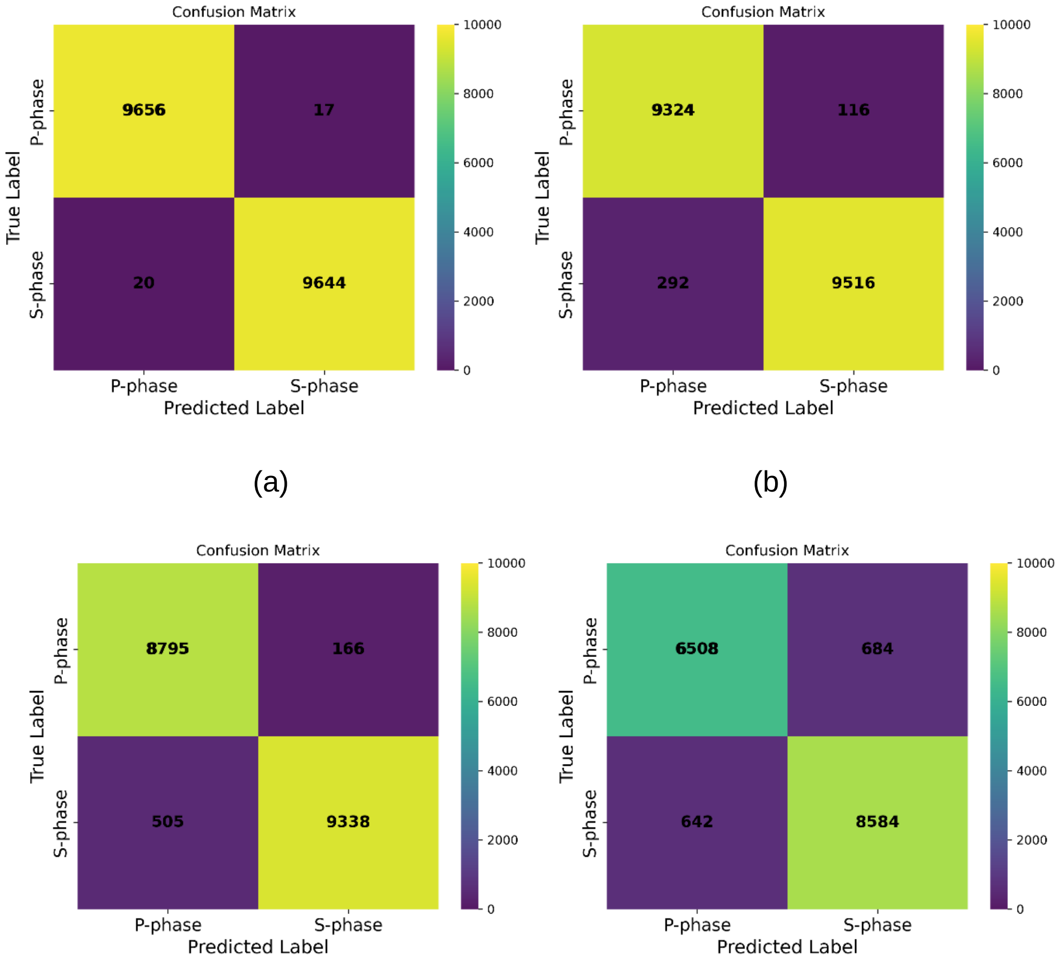


Figure 3-11: The distribution of picks residuals (Δt) of PhaseNet on the test dataset for P- (left) and S-wave arrival times (right).



(a) (b)

(c) (d)

Figure 3-12: Confusion matrices illustrating seismic event detection for the test dataset with varying input data for signal-to-noise ratios expressed as $log_{10}(SNR)$: (a) 3.995, (b) 2.6, (c) 0.602, and (d) 0.301.

Figure 3-13 and Figure 3-14 illustrate how the PhaseNet model performs for varying levels of noise added to randomly selected synthetic waveforms. In Figure 3-13, two different types of waveforms with the same noise level as $\log_{10}(\mathrm{SNR}) = 2.6$ are shown for the emphasis of the model predictions on more complex waveforms. In Figure 3-14, several waveforms containing the same level of noise as $\log_{10}(\mathrm{SNR}) = 0.301$ are illustrated to visually demonstrate how the trained model predicts the arrivals of both P- and S-phases in a more noisy environment. We compared the results of this model to those obtained using the STA/LTA method. The results from both techniques across various SNRs demonstrate that PhaseNet consistently delivers strong performance even as the noise level increases. As SNR rises, the performance metrics improve, with PhaseNet achieving a precision of almost 0.9 across all SNR levels. In contrast, the STA/LTA method struggles in high-noise conditions, particularly with P-waves, as low-magnitude events that are difficult to distinguish. The results of the comparison between the two different approaches are explained in detail in the next chapter.

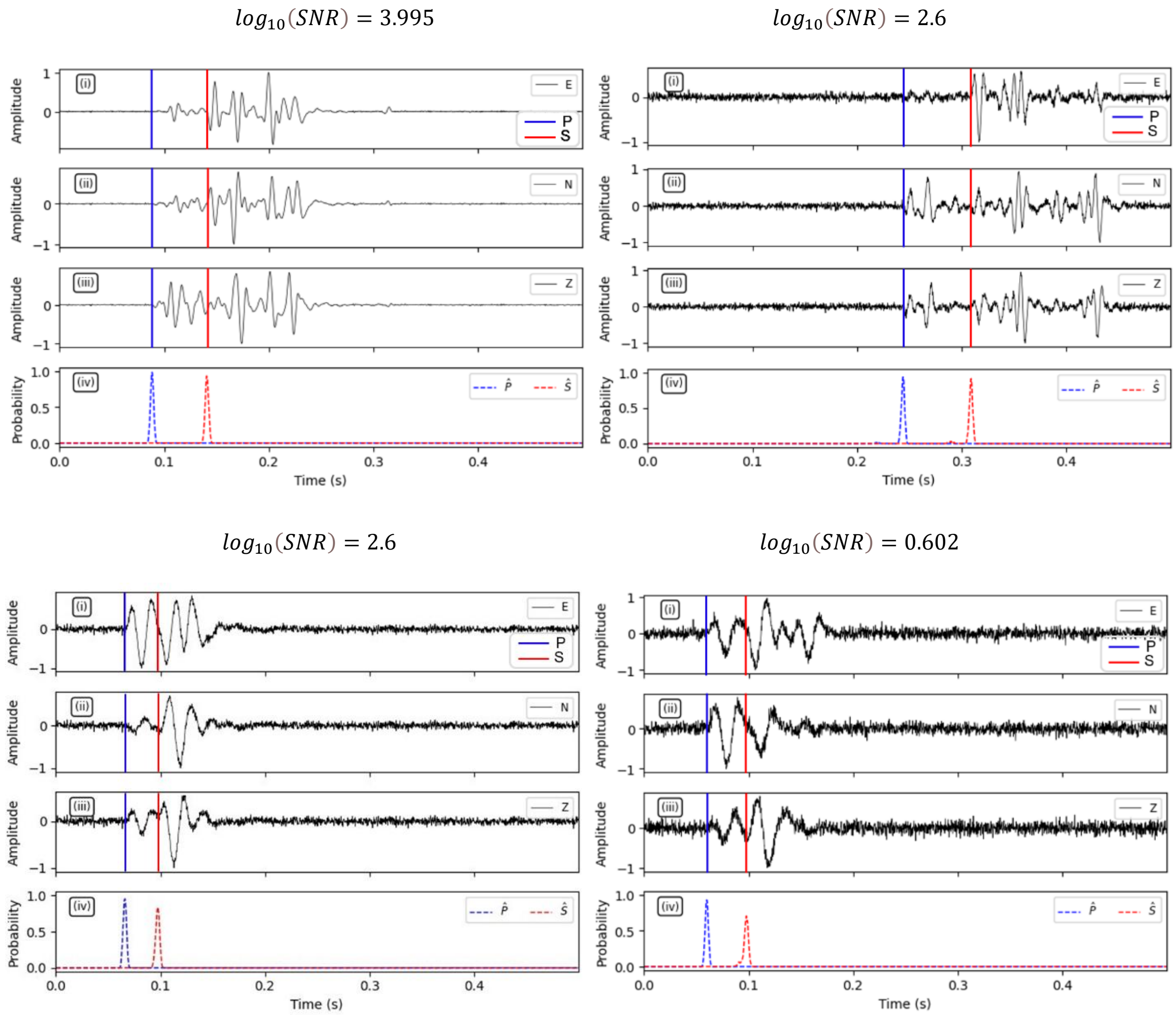
To evaluate the model performance, the confusion matrix is utilized, providing valuable insights into the effectiveness of the deep learning network in classifying the target classes within the dataset (Sammut et al., 2011). Subsequently, metrics such as precision, recall, and F1-score are computed (Powers, 2020):

$$Precision = \frac{TP}{TP + FP} \tag{3.15}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.16}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.17}$$

TP, TN, FP, and FN represent the counts of true positive samples, true negative samples, false positive samples, and false negative samples, respectively, indicating the instances where the model failed to accurately predict events according to their actual occurrence times within a given timeframe. In addition, the picking error, which refers to the time differences Δt between the model predictions and the actual labels, is used to evaluate the accuracy of the arrival time predictions.



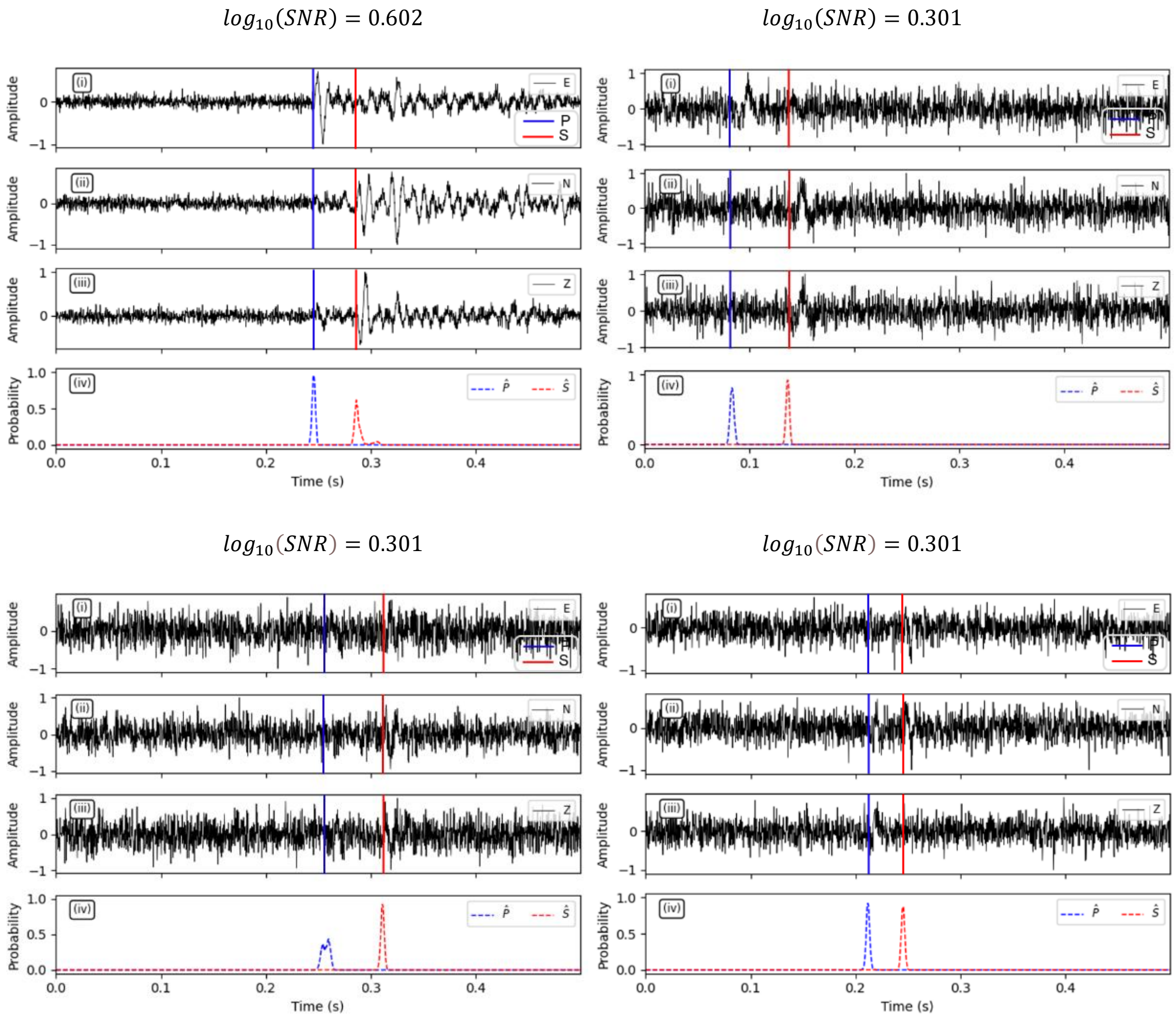
$log_{10}(SNR) = 3.995$

$log_{10}(SNR) = 2.6$

$log_{10}(SNR) = 2.6$

$log_{10}(SNR) = 0.602$

Figure 3-13: PhaseNet probabilistic pick predictions for P- (blue) and S- (red) arrival times on some synthetic three-component waveforms with higher signal-to-noise ratios and added white Gaussian noise.

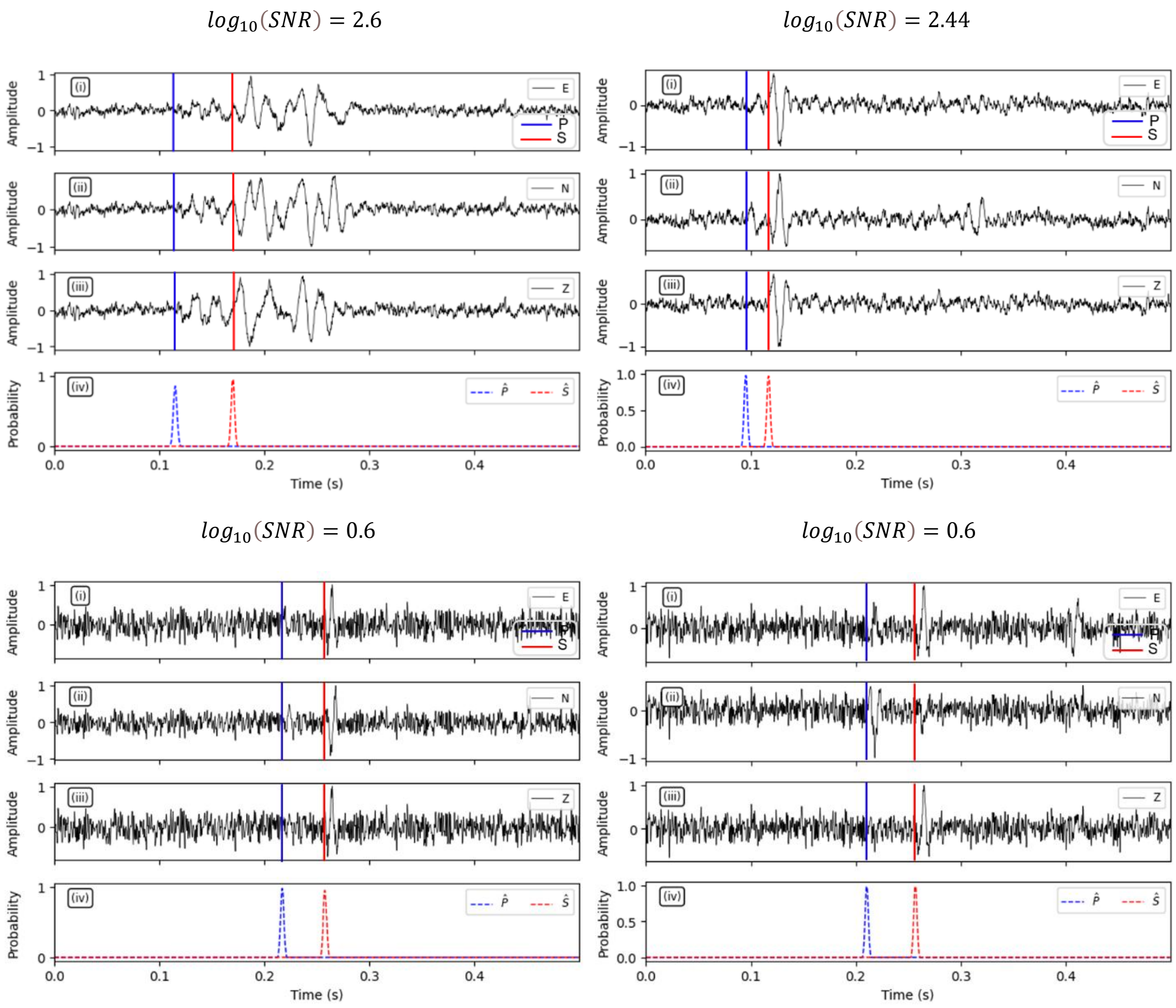

$log_{10}(SNR) = 0.602$ $log_{10}(SNR) = 0.301$

$log_{10}(SNR) = 0.301$ $log_{10}(SNR) = 0.301$

Figure 3-14: PhaseNet probabilistic pick predictions for P- (blue) and S- (red) arrival times on some synthetic three-component waveforms with lower signal-to-noise ratios and added white Gaussian noise.

As described in the section 2.2.7, colored noise was also added to the synthetic data to evaluate the robustness of the model. The results of the tests with colored noise are illustrated in Figure 3-15 for randomly selected waveforms. Again, PhaseNet performs well under the influence of varying levels of colored noise, offering insights into the model accuracy and reliability under more complex conditions.

Table 3-1 showcases the experimental results with respect to different levels of added white Gaussian noise. In addition to the metrics described above, the model performance is measured by the mean (μ) and standard deviation (σ) of the residuals between the detected and actual ground truth picks, reported in milliseconds. Overall, accuracy remains high, particularly at elevated SNR levels, where it approaches a value of 0.9958 for P-wave and 0.9948 for S-wave, indicating that the model makes correct predictions in most instances. Precision and recall values also reflect strong performance, often exceeding 0.9, indicating that when the model predicts a phase, it is generally accurate and effectively captures most actual instances. However, as the SNR decreases, a noticeable decline in these metrics occurs, demonstrating that the model has greater difficulty distinguishing between the phases amid increased noise. The F1-scores remain high, indicating that the model maintains a good balance between precision and recall. However, the timing differences reveal variability, especially at lower SNRs, where the predictions tend to be less reliable, as evidenced by higher standard deviation values. Thus, while the model performs robustly at elevated SNRs, its efficacy diminishes under poorer signal conditions. The trends in the diagrams of Figure 3-16 also show that as SNR increases, the performance metrics for the PhaseNet model improve significantly.



$log_{10}(SNR) = 2.6$

$log_{10}(SNR) = 2.44$

$log_{10}(SNR) = 0.6$

$log_{10}(SNR) = 0.6$

Figure 3-15: PhaseNet probabilistic pick predictions for P- (blue) and S- (red) arrival times on synthetic three-component waveforms with added colored noise.

The evaluation metrics indicate that PhaseNet effectively identifies both P and S seismic waves, with slightly superior performance observed for detecting S waves. The accuracy scores for both types of waves are notably high, with S-waves demonstrating a slight edge over P-waves.

Table 3-1: The performance of the first P- and S-phase arrival picking on the test dataset with different levels of SNRs for white Gaussian noise.

Evaluation Metric	P-phase	S-phase	P-phase	S-phase	P-phase	S-phase	P-phase	S-phase
$\log_{10}(\mathrm{SNR})$	**3.995**		**2.6**		**0.602**		**0.301**	
Accuracy	0.9958	0.9948	0.9353	0.9717	0.8638	0.9487	0.6307	0.8285
Precision	0.9979	0.9982	0.9696	0.9880	0.9457	0.9825	0.9102	0.9262
Recall	0.9978	0.9966	0.9635	0.9834	0.9089	0.9650	0.6725	0.8871
F1-Score	0.9979	0.9974	0.9666	0.9857	0.9269	0.9737	0.7735	0.9062
μ(∆t)(ms)	-0.1939	0.0635	-0.2598	0.0425	-0.2682	0.2314	-2.1175	-0.1440
σ(∆t)(ms)	0.4970	1.4599	2.5864	2.7525	3.7269	4.7992	11.5737	11.5478

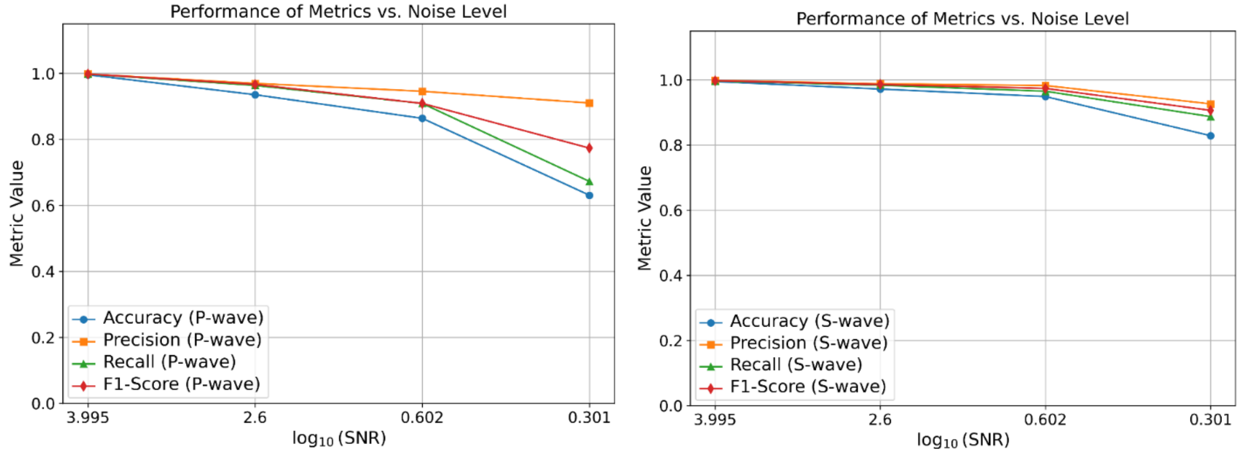


Figure 3-16: Performance of evaluation metrics for PhaseNet in P-wave (left) and S-wave (right) analysis: Accuracy, Precision, Recall, and F1-score across varying SNRs for white Gaussian noise.

Table 3-2 showcases impressive performance across several key criteria, including SNR (Signal-to-Noise Ratio), Accuracy, Precision, Recall, F1-score, and the mean (μ) and standard deviation (σ) of detection times (∆t) in milliseconds of the evaluation metrics for P- and S-phase detections. The SNR values, particularly 2.6 for the P-phase and 2.44 for the S-phase, indicate a strong signal-to-noise ratio, which is beneficial for accurate detection. However, the subsequent entry showing a lower SNR of 0.6 for the P-phase suggests a more challenging detection environment with increased noise. Accuracy remains notably high for the P-phase, with a value of approximately 0.9954. At the same time, the S-phase is slightly lower but still impressive, indicating that the model excels in identifying P-phase events. Both Precision and Recall metrics reinforce this trend, as they exceed 0.99 for both phases, indicating a high likelihood of correctness in detected events and capturing a substantial portion of actual occurrences, respectively. The F1-Score further underscores strong performance by balancing Precision and Recall, remaining consistently high across instances.

When examining mean detection times, the values present a mixture of positive and negative outcomes, with positive means such as 0.0807 ms for the first P-phase indicating quicker detections relative to a baseline. In contrast, negative values may indicate delays or timing misalignments. Standard deviation values, particularly higher ones such as 2.1304 ms for the S-phase, highlight variability in detection times, potentially suggesting inconsistent performance under varying conditions. Considering the impact of colored noise on these metrics is crucial, as performance may vary significantly. A drop in metrics such as accuracy, precision, and recall in the presence of colored noise compared to white noise would imply that the model may require adjustments to handle such noise types effectively. Overall, the metrics indicate that the detection system performs remarkably well, particularly in P-phase detection. However, further analysis under various noise conditions is necessary to assess the model's robustness and adaptability to real-world scenarios. Visual diagrams in Figure 3-17 could enhance the understanding of performance trends and identify areas for potential improvement.

Table 3-2: The performance of the first P- and S-phase arrival picking on the test dataset with different levels of SNRs for colored noise.

Evaluation Metric	P-phase	S-phase	P-phase	S-phase	P-phase	S-phase
$\log_{10}(\mathrm{SNR})$	**2.6**		**2.44**		**0.6**	
Accuracy	0.9954	0.9864	0.9871	0.9771	0.9777	0.9810
Precision	0.9977	0.9943	0.9937	0.9902	0.9890	0.9915
Recall	0.9976	0.9920	0.9933	0.9866	0.9884	0.9894
F1-Score	0.9977	0.9932	0.9935	0.9884	0.9887	0.9904
μ(∆t)(ms)	0.0807	-0.0422	-0.0807	0.1210	-0.1248	0.0001
σ(∆t)(ms)	0.8055	2.1304	0.8447	2.6329	1.1317	2.6087

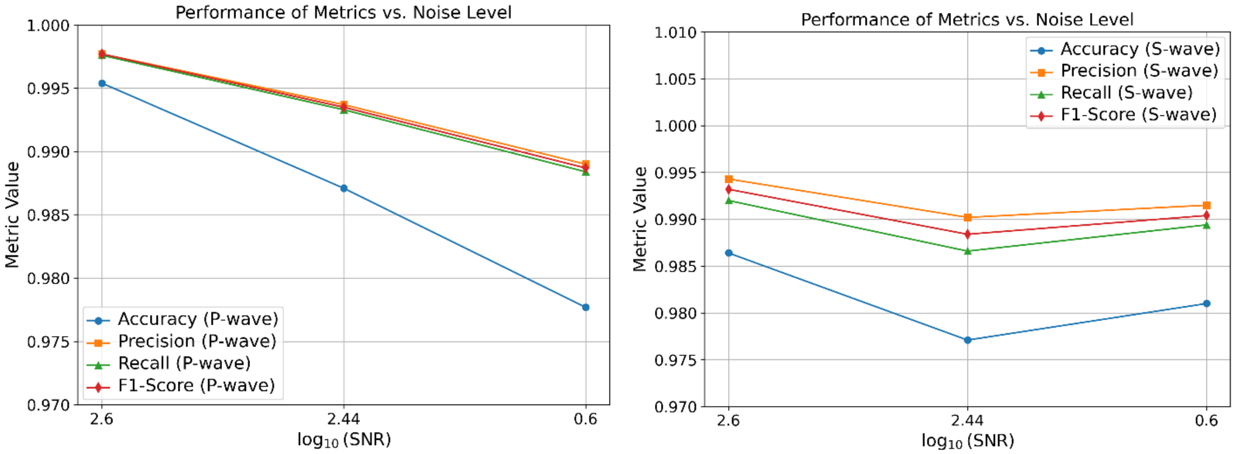
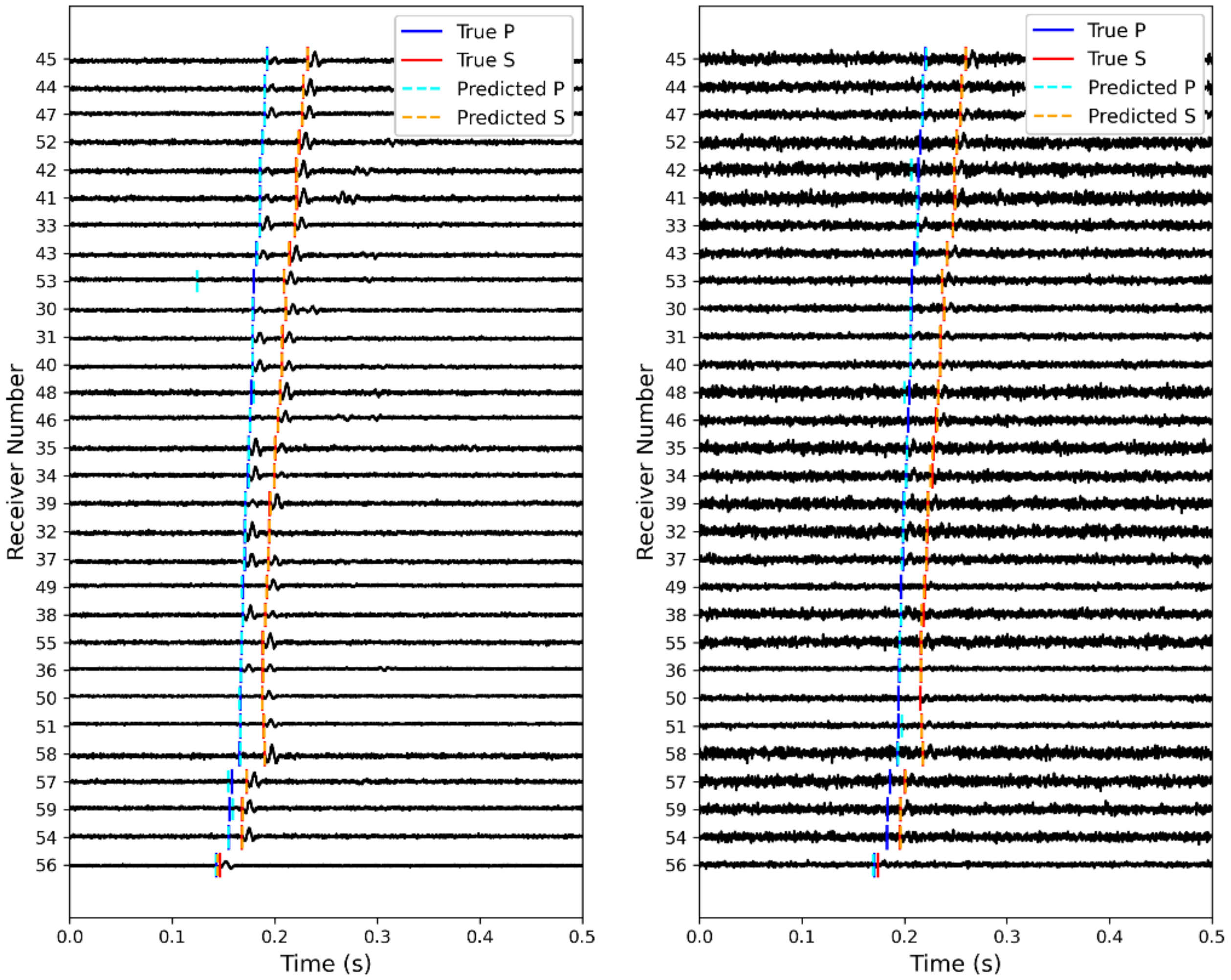


Figure 3-17: Performance of evaluation metrics for PhaseNet in P-wave (left) and S-wave (right) analysis: Accuracy, Precision, Recall, and F1-score across varying SNRs for colored noise.

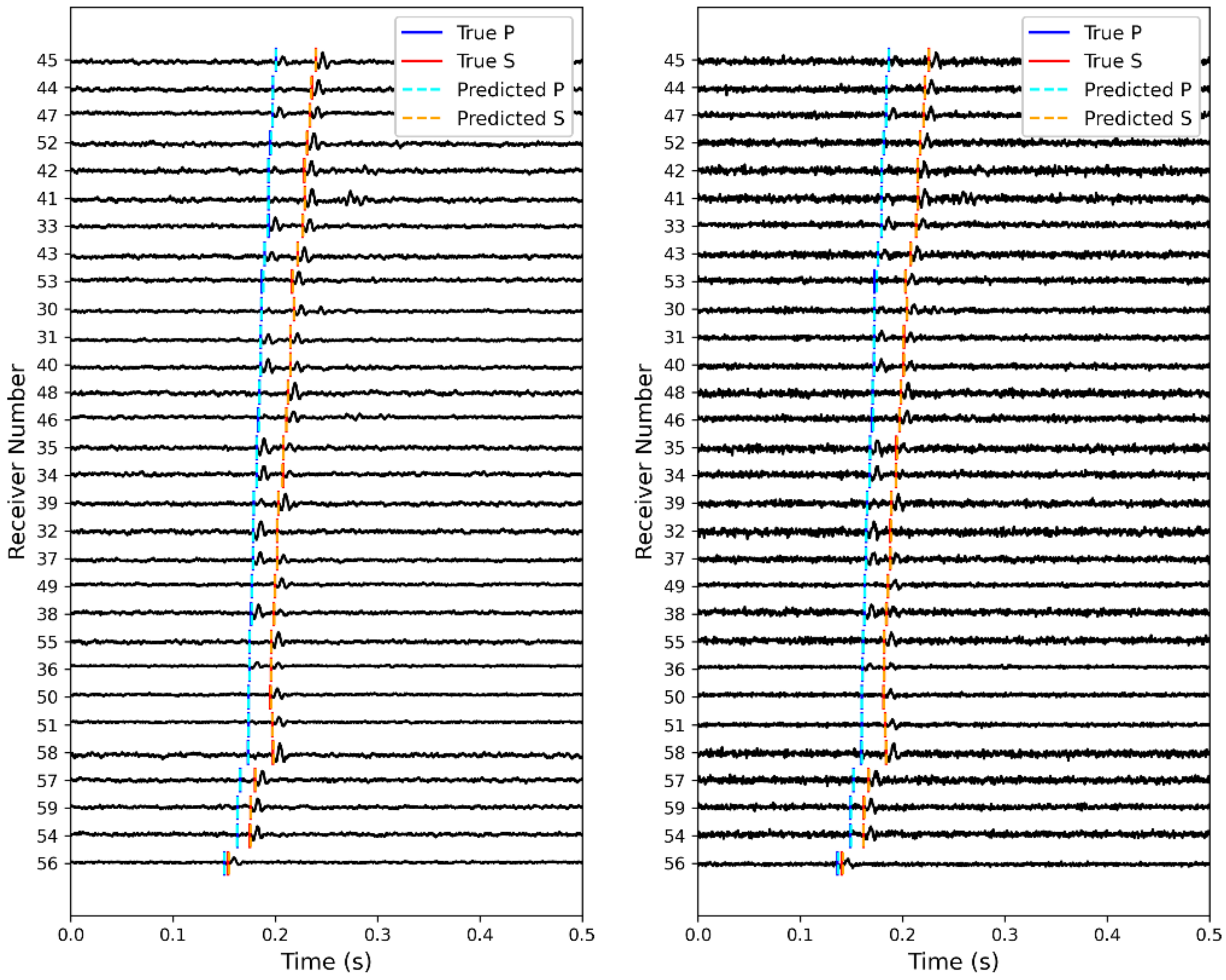
As illustrated in Figure 3-18 and Figure 3-19, the plots present the performance of the trained model on synthetic microseismograms from four distinct test datasets, all derived from the same seismic event but characterized by varying noise levels. In the two panels of Figure 3-18 we observe the predictions for P- and S-waves on synthetic data featuring $\log_{10}(\mathrm{SNR})$ values of 0.6 and 0.301, which correspond to white Gaussian noise conditions. These panels demonstrate the model's ability to capture the waveforms despite the challenges posed by the higher noise levels. Conversely, the two panels of Figure 3-19 showcase the model's performance on datasets exhibiting improved signal-to-noise ratios, with $\log_{10}(\mathrm{SNR})$ coefficient values of 2.44 and 2.6, associated with colored noise.



(a)

(b)

Figure 3-18: Performance of the trained model on synthetic microseismograms (z-component) across two different test datasets derived from the same event, each exhibiting varying noise levels. The two panels illustrate the P- and S-wave predictions on synthetic data with $log_{10}(SNR)$ values of (a) 0.6 and (b) 0.301, representing white Gaussian noise. The blue and red vertical lines represent the true P- and S-wave arrivals, and the light blue and orange dashed vertical lines represent the predicted arrivals of P- and S-wave by PhaseNet, respectively.



(a) **(b)**

Figure 3-19: Performance of the trained model on synthetic microseismograms (z-component) across two different test datasets derived from the same event, each exhibiting varying noise levels. The two panels illustrate P- and S-wave predictions on datasets with $\log_{10}(\mathrm{SNR})$ values of (a) 2.44 and (b) 2.6, corresponding to colored noise. The blue and red vertical lines represent the true P- and S-wave arrivals, and the light blue and orange dashed vertical lines represent the predicted arrivals of P- and S-wave by PhaseNet, respectively.

3.3 Conclusion

The evaluation of the model performance in the presence of white Gaussian noise reveals a noticeable decline as the signal-to-noise ratio decreases, specifically within the $\log_{10}(\mathrm{SNR})$ range of 3.995 to 0.301. At elevated SNR levels (approximately $\log_{10}(\mathrm{SNR}) \approx 4$), the model exhibits nearly perfect accuracy and F1-scores (exceeding 0.99) for both P- and S-phases, demonstrating strong confidence and reliability under clear signal conditions. However, as noise levels rise, particularly below $\log_{10}(\mathrm{SNR}) = 1$, detection accuracy suffers, with a more marked impact on P-phase identification. When observing the lowest tested SNR ($\log_{10}(\mathrm{SNR}) = 0.301$), the F1-score for P-phase detection drops to 0.77, while the S-phase score remains higher at 0.91, indicating that S-phase detection is more resilient to a higher amount of noise. This is attributed to the higher amplitude of S-waves compared to P-waves, which results in a localized higher SNR for S-waves. Additionally, uncertainty in arrival time increases with the rise in noise, as $\sigma(\Delta t)$ approaches ~11.5 ms for both phases at the lowest SNR. These findings highlight the anticipated sensitivity of arrival picking models to white noise, particularly affecting short-duration, high-frequency P-wave arrivals.

Furthermore, the assessment of the phase-picking model in the presence of increasingly intense colored noise ($\log_{10}(\mathrm{SNR})$ values between 2.6 and 2.0) reveals its ability to adapt to more realistic situations. F1-scores for both P- and S-phase detections remain above 0.99 at higher SNR levels, and they sustain values over 0.98 even at $\log_{10}(\mathrm{SNR}) = 2$. In comparison to its performance with white noise, the model shows greater resilience when faced with colored noise, reflected in lower standard deviations of arrival time errors ($\sigma(\Delta t)$ around 0.8–1.1 ms for P-phases and 2.1–2.6 ms for S-phases). This indicates the model efficacy in distinguishing seismic arrivals from low-frequency background noise. The slight decline in precision and recall across the SNR levels further underscores the model relevance in real-world conditions where colored noise prevails over theoretical expectations associated with white noise.

4 COMPARISON WITH CONVENTIONAL METHOD

4.1 Introduction

This chapter delves into the theoretical foundation of useful single-level conventional algorithms, detailing their parameters and outlining the processes for selecting P- and S-wave arrival times. Then, the Short-Term Average/Long-Term Average method, one of the most common techniques, is applied to the synthetic data for prediction comparison with PhaseNet.

4.2 Conventional Arrival-Time Picking Methods

Arrival-time picking algorithms are categorized into two main types: single-level and multilevel algorithms, also known as array processing algorithms (Akram et al., 2016). Single-level algorithms analyze data from a single receiver, either using single-component or multicomponent data without incorporating information from other receivers in the array. In contrast, multilevel algorithms leverage data across multiple receiver levels simultaneously, with several methods relying on cross-correlation for enhanced accuracy.

The single-level algorithms are also divided into window-based and non-window-based methods. Window-based algorithms utilize predefined window sizes and locations to compute microseismic data attributes, which are then used to set criteria for picking arrival times. Additionally, hybrid algorithms combine features from various methods to improve the accuracy and precision of arrival-time picks, recognizing that no single algorithm is universally optimal (Akram et al., 2016; Sharma et al., 2010). In the following, some of the most commonly used single-level algorithms are briefly described.

4.2.1 Short-Term Average / Long-Term Average Method

The STA/LTA ratio is analogous to the signal-to-noise ratio, with the STA being responsive to quick, meaningful changes in the amplitude of the time series, while the LTA offers insights into the background noise (Trnkoczy, 2009). It is crucial to configure the method parameters in a way that ensures the STA and LTA windows do not overlap, thereby maintaining statistical independence between the two values (Taylor et al., 2010). According to the principle of causality, the STA window must always precede the LTA window. Both STA and LTA windows can be set as pre-sample windows, meaning they occur before the time sample for which the STA/LTA is calculated, allowing arrival times to be determined from the maximum value of the derivative of the STA/LTA curve. Alternatively, using a post-sample STA window (which follows the time sample) alongside a pre-sample LTA window enables the identification of arrival times based on the peak of the STA/LTA curve. The generalized expressions for STA and LTA at the ith time sample are as follows (Akram et al., 2016; Taylor et al., 2010):

$$STA_{\{i\}} = \frac{1}{ns} \sum_{j=i-ns}^{i} CF_{\{j\}} \tag{4.1}$$

$$LTA_{\{i\}} = \frac{1}{nl} \sum_{j=i-nl}^{i} CF_{\{j\}} \tag{4.2}$$

$$r_{\{i\}} = \frac{STA_{\{i\}}}{LTA_{\{i\}}} \tag{4.3}$$

where ns and nl denote the number of samples within the short- and long-term windows, respectively. CF refers to a characteristic function that can signify the waveform energy, absolute amplitude, or any other mathematical combination of microseismic data and its derivatives (Allen, 1978; Baer et al., 1987; Trnkoczy, 2009; Vaezi et al., 2015). An STA/LTA ratio $r_{\{i\}}$ is calculated for each index i, and if the ratio surpasses a specified threshold, it signals the declaration of an arrival time (see Figure 4-1) (Earle et al., 1994). In our experiments, we adopt the energy function, which specifically emphasizes changes in amplitude.

When applying the STA/LTA method, it is crucial to select the window lengths and threshold value carefully to achieve optimal outcomes. It is advised that the STA window length be at least two to three times the dominant period of the signal, while the LTA window should be five to ten times the length of the STA. Additionally, the detection threshold should be sufficiently high to prevent false positives while still allowing for the detection of smaller magnitude events (Akram et al., 2016).

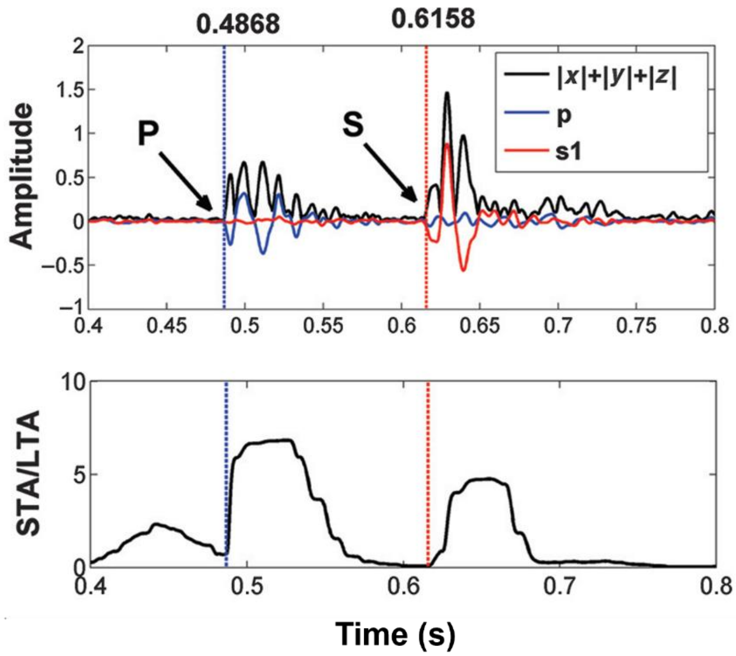


Figure 4-1: Input microseismic (stack of 3C) data (top) and characteristic function curve for STA/LTA (bottom). Manually picked P- and S-wave arrival times are shown with blue and red vertical lines, respectively.

The STA/LTA method is widely used in real-time seismic monitoring for its ease of implementation. However, it does have limitations, such as sensitivity to noise and the challenge of setting appropriate thresholds, which can impact its effectiveness in dense noise environments.

4.2.2 Modified Energy Ratio (MER) Method

Han et al. (2009) introduced the MER algorithm, which builds on the STA/LTA approach by utilizing pre-sample and post-sample windows of equal length (Gaci, 2013). The energy ratio (ER) at the ith time sample is calculated using the formula (Akram et al., 2016):

$$ER_{\{i\}} = \frac{\sum_{j=1}^{i+w} x_j^2}{\sum_{j=i-w}^{i} x_j^2} \tag{4.4}$$

where w represents the window length and x refers to the input data. The MER is then determined by:

$$MER_{\{i\}} = (ER_i|x_i|)^3 \tag{4.5}$$

In this method, since the ER function is derived from both post- and pre-sample windows, the time index at which the maximum value of the MER occurs indicates the estimated arrival time. Figure 4-2 illustrates the MER curve for the sample data, showing that the P- and S-wave arrivals align closely with the local maxima within their respective time intervals.

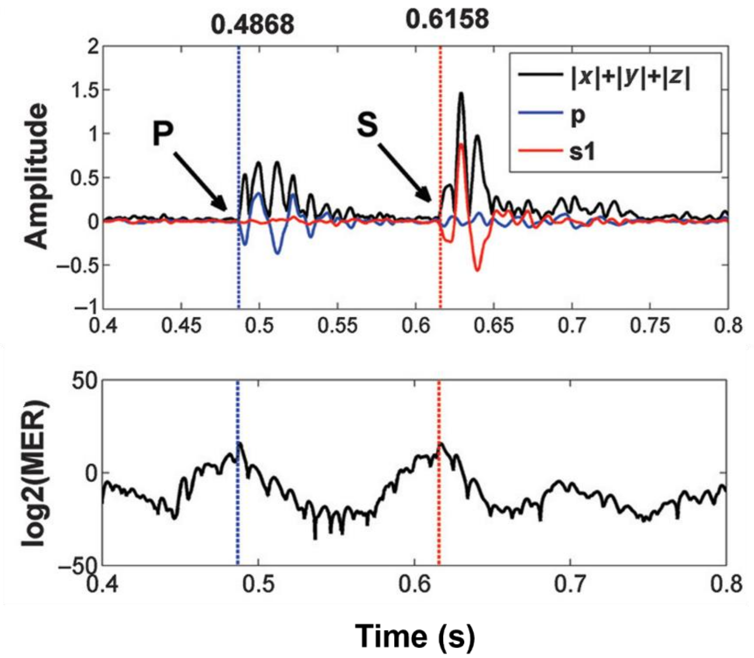


Figure 4-2: Input microseismic (stack of 3C) data (top) and characteristic function curve for MER (bottom). Manually picked P- and S-wave arrival times are shown with blue and red vertical lines, respectively (Akram et al., 2016).

Selecting the appropriate window size is crucial for obtaining meaningful results with the MER algorithm. Similar to the STA/LTA method, the window length should extend beyond several periods of the microseismic signal to mitigate false detections caused by noise fluctuations and to accurately identify changes in the signal.

4.2.3 Modified Coppens' Method (MCM)

A modified version of Coppens' method is referred to as MCM. Like other energy ratio algorithms, this method calculates energy across two windows, but the approach to selecting window sizes differs from that of the STA/LTA and MER methods. At the ith time sample, the energy windows are defined as follows (Akram et al., 2016; Coppens, 1985; Sabbione et al., 2010):

$$EW1_{\{i\}} = \sum_{j=i-nl+1}^{i} x_j^2 \tag{4.6}$$

$$EW2_{\{i\}} = \sum_{j=1}^{i} x_j^2 \tag{4.7}$$

Here, nl represents the length of the leading window. The second window is an increasing window that aids in providing a strong response when the first arrival occurs. The energy ratio is calculated as:

$$MCM_{\{i\}} = \frac{EW1_{\{i\}}}{\left(EW2_{\{i\}} + \beta\right)} \tag{4.8}$$

In this equation, β is a stabilization constant designed to reduce the rate of false detections by smoothing the fluctuations of the MCM curve. It is typically set to 0.2 for normalized input data ranging from -1 to 1. Figure 4-3 illustrates the MCM curve for the sample data. Like the STA/LTA and MER algorithms, the purpose of the leading window is to detect changes in the signal.

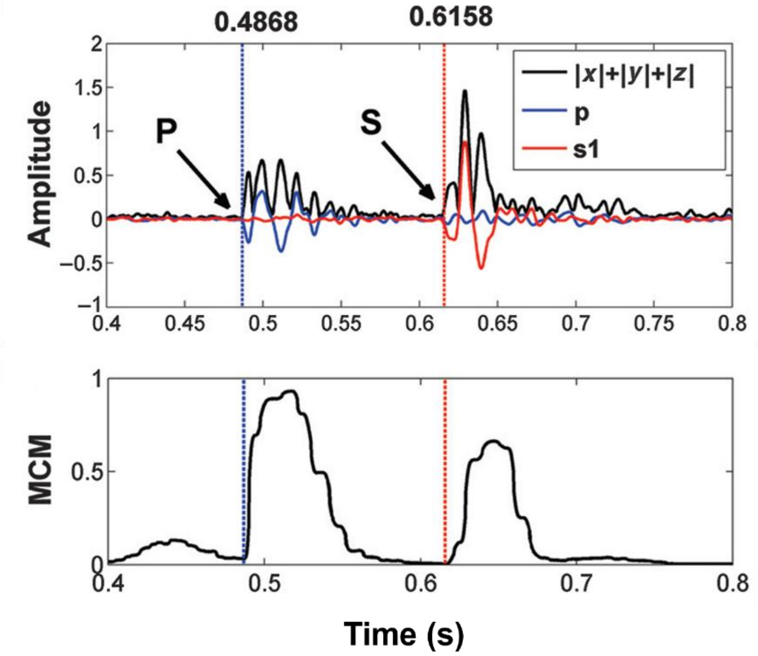


Figure 4-3: Input microseismic (stack of 3C) data (top) and characteristic function curve for MCM (bottom). Manually picked P- and S-wave arrival times are shown with blue and red vertical lines, respectively (Akram et al., 2016).

4.2.4 Akaike Information Criterion (AIC) Method

The Akaike Information Criterion (AIC) algorithm operates on the premise that microseismic signals are inherently nonstationary. To analyze these signals, the approach involves segmenting the observed waveform into smaller, locally stationary parts, each of which is regarded as an autoregressive process. For the kth sample in a microseismic waveform comprised of N samples, the AIC can be expressed as (Akram et al., 2016; Sleeman et al., 1999; Zhang et al., 2003):

$$AIC(k) = (k - M)\log\left(\sigma_{1,max}^2\right) + (N - M - K)\log\left(\sigma_{2,max}^2\right) + C_2 \tag{4.9}$$

where M denotes the order of the autoregressive model, $\sigma^2_{,max}$ represents the variances of the microseismic signals in two separate intervals not captured by the autoregressive process, and C_2 is a constant term. The selection of the autoregressive model order typically involves a trial-and-error approach applied to a data window that contains noise. The AIC calculated with this estimated model order quantifies how well the model fits the data, with the optimal separation of the two stationary time series (noise and signal) reflected by the time index where the AIC reaches its minimum value. The AIC can also be computed directly from the time series data without incorporating autoregressive model coefficients. In this formulation, AIC is expressed as (Maeda, 1985):

$$AIC(k) = k\,log(var\{x(1,k)\}) + (N-k-1)\log(var\{x(k+1,N)\}) \tag{4.10}$$

where k spans all samples of the input microseismic data and $var\{x\}$ refers to the variance function. Figure 4-4 illustrates the AIC response for the provided sample data, clearly marking the P-wave arrival, although the S-wave arrival is less discernible through AIC alone. This discrepancy arises because AIC identifies the onset of events as a global minimum, necessitating the use of arrival time window estimates to enhance accuracy, particularly when multiple arrivals are present (Zhang et al., 2003).

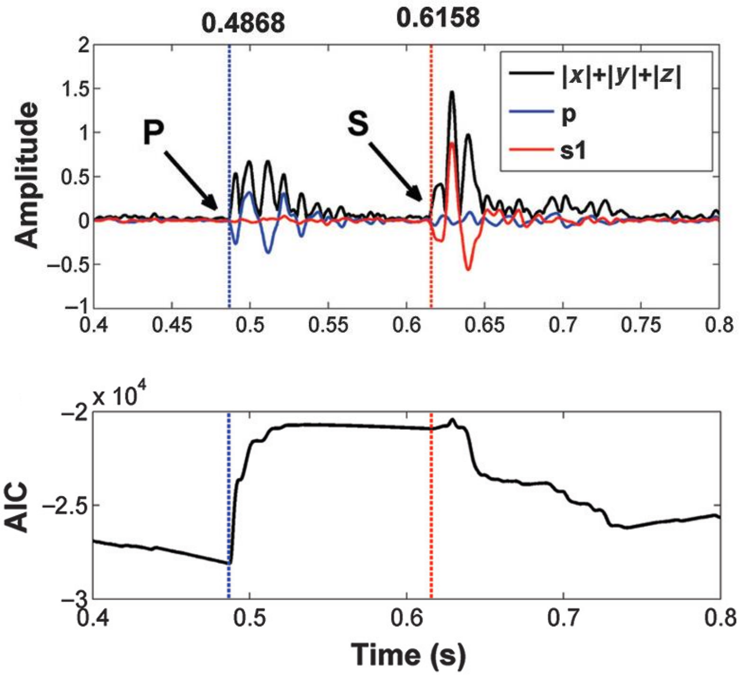


Figure 4-4: Input microseismic (stack of 3C) data (top) and characteristic function curve for AIC (bottom). Manually picked P- and S-wave arrival times are shown with blue and red vertical lines, respectively (Akram et al., 2016).

Table 4-1 reviews some of the key parameters of the four different methods for arrival-time picking described in this section.

Table 4-1: Conventional single-level arrival-picking methods, along with their key parameters described in this section (Akram et al., 2016).

Arrival-Picking Method	Type	Key Parameters	Reference
Short-Term Average/Long-Term Average (STA/LTA)	Window-based, Single-level	STA = 2–3 times the dominant period of the signal LTA = 5–10 times STA	(Allen, 1978)
Modified Energy Ratio (MER)	Window-based, Single-level	Window size = 2–3 times the dominant period of the signal	(Han et al., 2009)
Modified Coppens' Method (MCM)	Window-based, Single-level	Window size = 1–3 times the dominant period of the signal	(Sabbione et al., 2010)
Akaike Information Criterion (AIC)	Non window-based, Single-level	A priori information of time windows containing arrivals from other algorithms	(Sleeman et al., 1999)

4.3 Application of STA/LTA Method on Synthetic Seismograms

Instead of calculating STA/LTA separately for each component of the seismograms, we compute the vector magnitude of the three components (x, y, and z). By combining the three components, we enhance the sensitivity by capturing the overall energy of the signal, which is particularly beneficial for detecting weak or subtle events that may be less pronounced in individual channels. It has been shown to work better for P-waves (which are usually strongest on z component) and S-waves (which are often strong on x and y components) (Richards et al., 2002; Shearer, 2019). This vector magnitude approach also plays a critical role in noise reduction, as it suppresses random noise that could otherwise interfere with single-component measurements. Moreover, using a single measure simplifies processing, reduces complexity, and streamlines computations, which is particularly advantageous when dealing with large datasets or conducting real-time analyses. However, it should be mentioned that it inherently sacrifices directional information.

Figure 4-5 to Figure 4-7 illustrates the STA/LTA responses for the three components of several seismograms with three different levels of noise and source-dominant frequencies. The true P- and S-arrival times are indicated by pink and green vertical dashed lines, respectively. As the level of noise increases in synthetic seismograms, the efficacy of the STA/LTA method in detecting low-magnitude events, particularly P-waves, tends to diminish significantly. This is largely attributed to the inherent characteristics of the STA/LTA technique and the delicate balance required in parameter selection.

In these figures, a detection is declared when the STA/LTA ratio exceeds a predefined upper threshold (trigger on), which marks the onset of a potential seismic event. The detection continues until the ratio falls below a lower threshold (trigger off), indicating the return to background noise conditions.

This two-threshold approach is important because it stabilizes the detection process. Without distinct on and off thresholds, the trigger state could oscillate rapidly due to fluctuations in the ratio, especially in noisy data. In practice, the trigger on point generally corresponds to the arrival of the P- or S-phase, while the trigger off point indicates when the transient seismic energy has decayed. Together, these markers define the temporal window of the detected event and form the basis for subsequent picking or phase association. Trigger on and trigger off are illustrated by vertical red and blue lines, respectively, for both P- and S-arrival times detected by the STA/LTA approach.

The analysis window size is crucial for STA/LTA algorithms. A short STA window leads to noise variations, while a long LTA window can mask closely occurring events. The window sizes should be based on the frequency properties of the microseismic waveform. The STA window should exceed a few cycles of the signal's frequency, while the LTA window should cover several cycles of random noise. A longer LTA window also enhances sensitivity to weaker P-waves (Akram et al., 2016; Earle et al., 1994).

As an example, the dominant period of the specific signal on the top of the Figure 4-5 ($\tau_{dominant} =$ the sampling frequency of 1/65 Hz, corresponding to 0.015 s, or 2000 time samples with a sampling interval of 0.0075 ms), is used to set the STA and LTA window sizes. The STA/LTA responses are presented for an STA window size of 0.5*$\tau_{dominant}$ and an LTA window size of 3*$\tau_{dominant}$ through a trial-and-error approach for the suitable size of the windows. The STA/LTA analysis reveals unwanted fluctuations for short STA and LTA windows, which occurs because the LTA window is too brief to effectively capture the average local noise, making it overly reactive to noise variations. Notably, a slight increase in the LTA size to 5*$\tau_{dominant}$ leads to significant improvements, although it may still result in false picks when the data has low SNRs. Based on the observations and several experiments, it is advisable to set the STA window size between 0.5 and 1.5 times the $\tau_{dominant}$ and the LTA window size to be 3-5 times that of the STA size.

In environments with high noise levels, the background noise can obscure the smaller, low-magnitude seismic waves. As noise levels rise, the noise amplitude can influence both the STA and LTA calculations, potentially causing the STA to be elevated due to noise fluctuations. This results in a lower STA/LTA ratio, which may not surpass the predetermined threshold for event detection, leading to missed detections of P-wave arrivals in these examples, specifically for the seismograms with $log_{10}(SNR) = 0.301$ as illustrated in Figure 4-7.

The threshold value for the STA/LTA ratio is a critical parameter. When the noise level is high, a higher threshold may be necessary to avoid false positives. However, this higher threshold can inadvertently cause genuine low-magnitude P-wave arrivals to be overlooked. Finding the optimal threshold is a balancing act that is particularly challenging in noisy environments.

Furthermore, the configuration of the STA and LTA windows is another pivotal factor that impacts P- and S-wave detection. The recommended practice suggests that the STA window should be two to three times the dominant period of the signal, while the LTA should be five to ten times the length of the STA. In high-noise environments, if the window lengths are not properly calibrated, the STA may fail to capture the important fluctuations caused by actual seismic events, thereby impeding effective detection.

Shorter STA windows may enhance responsiveness to rapid changes in the signal but could also make the algorithm overly reactive to noise. On the other hand, longer STA windows might smooth out transient signals, reducing the likelihood of detecting low-magnitude events, as they may not present a sufficiently significant amplitude for detection. Due to the principle of causality inherent in the STA/LTA method, the STA window must always precede the LTA window. This temporal requirement can influence the detection of onset times. In a noisy environment, if noise spikes rapidly follow P-wave arrivals, the LTA window may average out these fluctuations, leading to an underestimation of the actual arrival time of the event. As a result, the onset detection accuracy suffers, affecting the overall reliability of the phase picks.

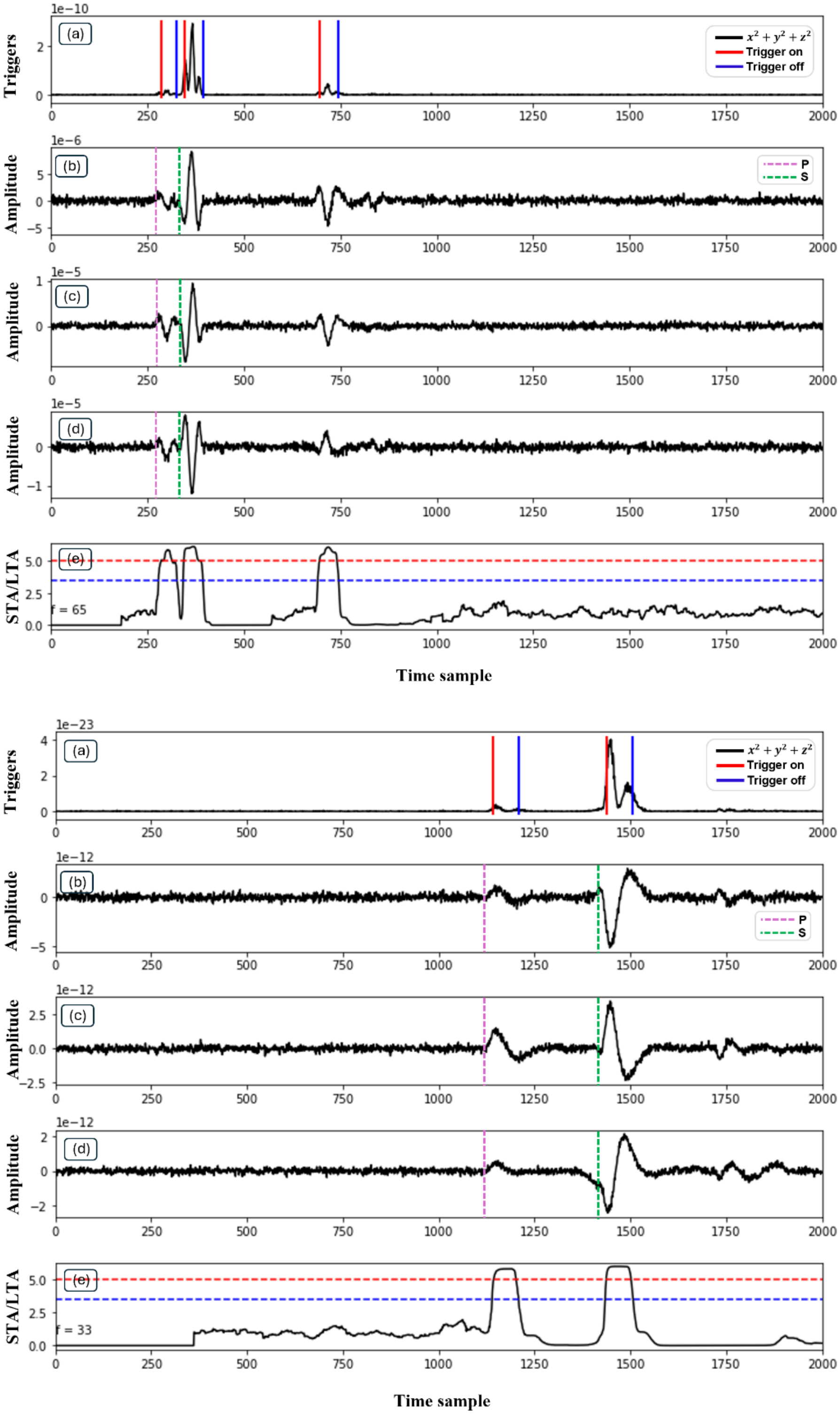


Figure 4-5: Application of STA/LTA on two synthetic seismograms with high (f=65 Hz) (top), and low (f=33 Hz) (bottom) source frequencies and $log_{10}(SNR) = 2.6$. Each panel includes: (a) the triggers from the STA/LTA applied to the calculated energy of the 3C noisy waveform (Trigger on and off are shown as vertical red and blue lines, respectively, for both P- and S-wave arrivals), (b horizontal east-west component, (c) horizontal north-south component, (c) the vertical component, and (e) calculated CF using a sliding window for STA/LTA. The true P- and S-wave arrivals are pink and green dashed vertical lines, respectively.

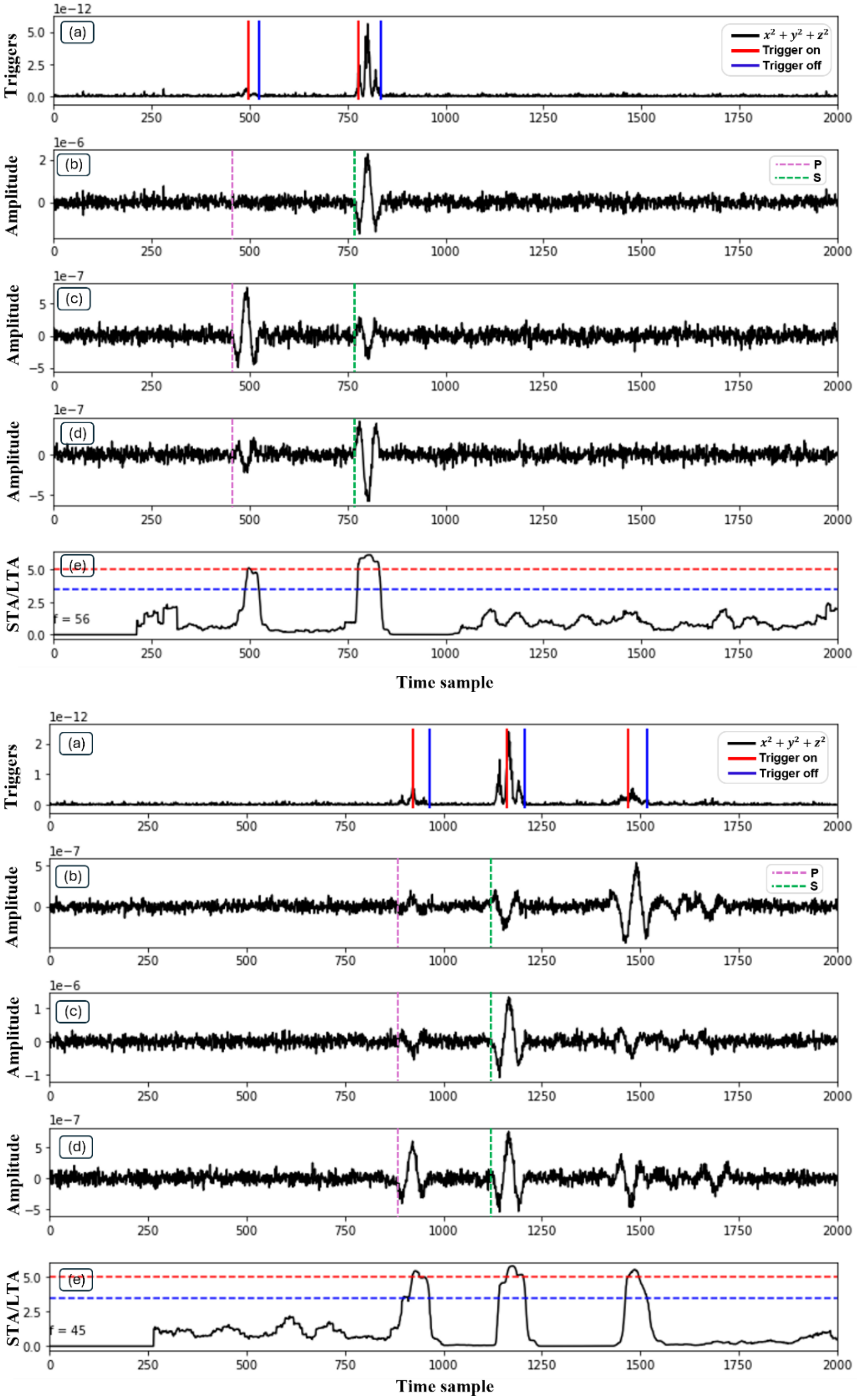


Figure 4-6: Application of STA/LTA on two synthetic seismograms with f=56 Hz (top), and f=45 Hz (bottom) source frequencies and $log_{10}(SNR) = 0.6$. Each panel includes: (a) the triggers from the STA/LTA applied to the calculated energy of the 3C noisy waveform (Trigger on and off are shown as vertical red and blue lines, respectively, for both P- and S-wave arrivals), (b horizontal east-west component, (c) horizontal north-south component, (c) the vertical component, and (e) calculated CF using a sliding window for STA/LTA. The true P- and S-wave arrivals are pink and green dashed vertical lines, respectively.

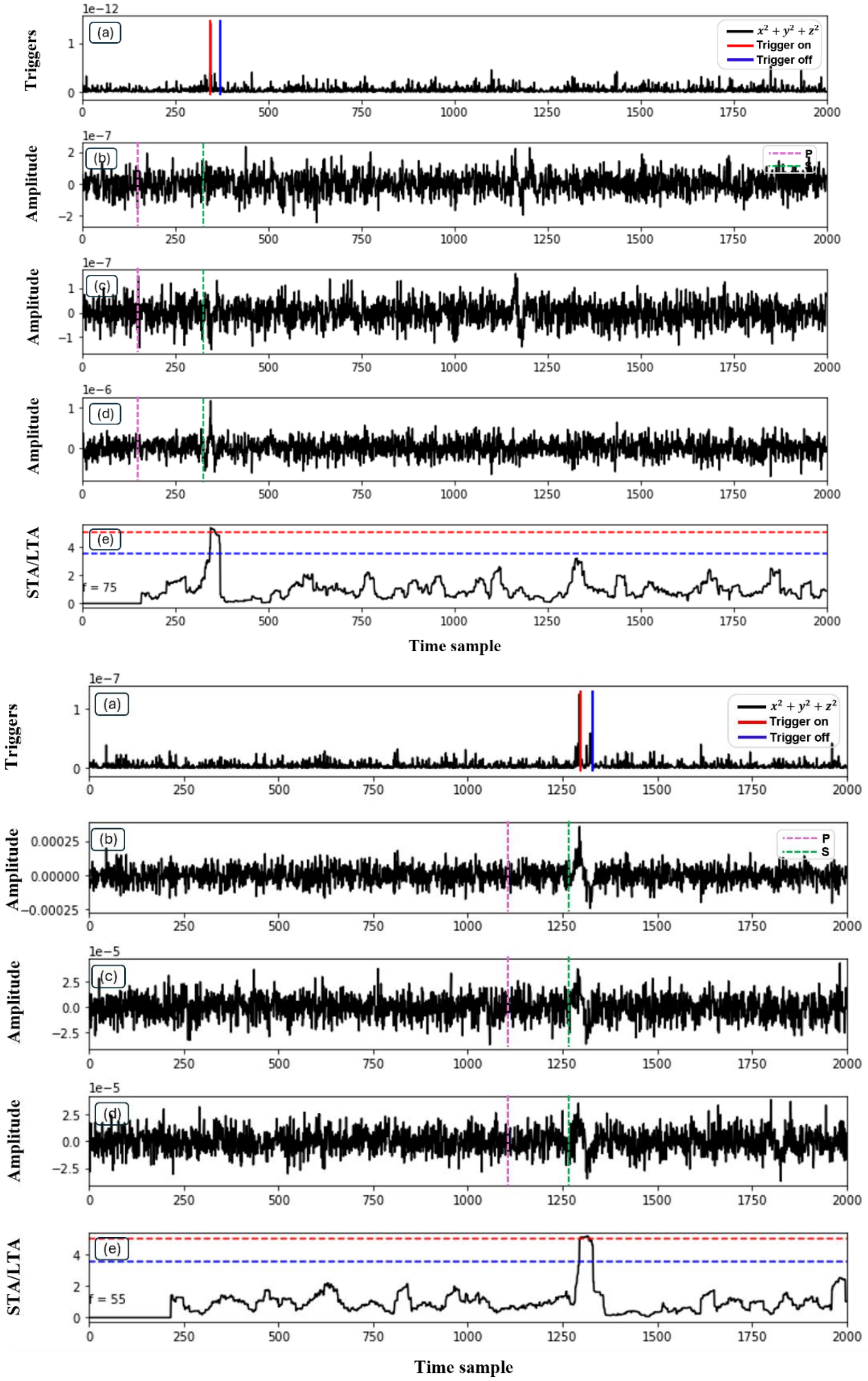


Figure 4-7: Application of STA/LTA on two synthetic seismograms with f=75 Hz (top), and f=55 Hz (bottom) source frequencies and $log_{10}(SNR) = 0.301$. Each panel includes: (a) the triggers from the STA/LTA applied to the calculated energy of the 3C noisy waveform (Trigger on and off are shown as vertical red and blue lines, respectively, for S-wave arrivals), (b horizontal east-west component, (c) horizontal north-south component, (c) the vertical component, and (e) calculated CF using a sliding window for STA/LTA. The true P- and S-wave arrivals are pink and green dashed vertical lines, respectively.

4.4 Comparison of PhaseNet with STA/LTA Results

While the STA/LTA method relies on optimal short and long sliding windows for each analysis, PhaseNet operates on fixed-length time series, eliminating the need for window parameterization. The output sequences presented in Figure 4-8 compare the two approaches on two synthetic seismic waveforms with different SNR levels, demonstrating the advantage of PhaseNet in detecting low-magnitude events in noisy environments.

The comparison between the STA/LTA method and PhaseNet for detecting seismic wave arrivals reveals notable differences in their performance, particularly under varying SNRs. As in the previous chapter, a temporal window centered around the actual phase-picking times is employed to assess the predicted arrivals of P- and S-phases, with this interval set to varying thresholds. Table 4-2 illustrates the effectiveness of both methods across various threshold (time interval) levels, indicating the acceptable differences between the predicted and actual seismic wave arrival times. PhaseNet consistently demonstrates high performance in P-wave picking, regardless of changes in the threshold for synthetic seismograms. In contrast, STA/LTA exhibits improved efficiency in terms of the number of true events detected with broader threshold ranges, implying that a more significant interval of error is necessary for event detectability. Figure 4-9 presents a comparative analysis of the performance of the two phase-picking algorithms across different threshold values. The thresholds, denoted on the x-axis, represent the allowable interval of discrepancies between predicted and true arrival times of seismic waves. In the left subplot, we observe that the P-wave picking accuracy of PhaseNet remains consistently high and largely unaffected by changes in threshold values, maintaining a level above 90%. In contrast, STA/LTA exhibits a positive correlation between threshold broadening and accuracy. For the S wave picking accuracy shown in the right subplot, a similar trend is observed; PhaseNet outperforms STA/LTA at tighter thresholds, with both methods converging at higher threshold values, although PhaseNet maintains a slight edge.

Figure 4-10 shows the results from both techniques across different SNRs, highlighting that PhaseNet consistently delivers strong performance even as the noise level increases. The dataset is divided into five categories based on $log_{10}(SNR)$ values, and the accuracy of true P- and S-picks is measured for each category. As the SNR increases, the metrics improve, with PhaseNet achieving an accuracy of nearly 0.9 across most SNR levels. In contrast, the STA/LTA method struggles in high-noise conditions, particularly with low-magnitude events that become hard to distinguish.

Table 4-2: A comparative evaluation of the effectiveness of PhaseNet and STA/LTA techniques for various detection threshold levels for P-wave and S-wave for waveforms with high levels of SNR ($log_{10}(SNR) = 2.6$).

Threshold			±0.002 (s)	±0.005 (s)	±0.02 (s)	±0.05 (s)
P-wave (9677)	**PhaseNet**	Number of picked events	8891	9324	9636	9674
		Percentage	**91.87%**	**93.53%**	**99.57%**	**99.96%**
		Number of false picks	787	292	42	4
	STA/LTA	Number of picked events	225	2253	5259	7890
		Percentage	2.32%	23.28%	54.34%	81.53%
S-wave (9677)	**PhaseNet**	Number of picked events	9129	9516	9650	9653
		Percentage	**94.33%**	**97.17%**	**99.72%**	**99.75%**
		Number of false picks	568	116	47	7
	STA/LTA	Number of picked events	289	2249	4044	4356
		Percentage	2.98%	23.24%	41.78%	45.01%

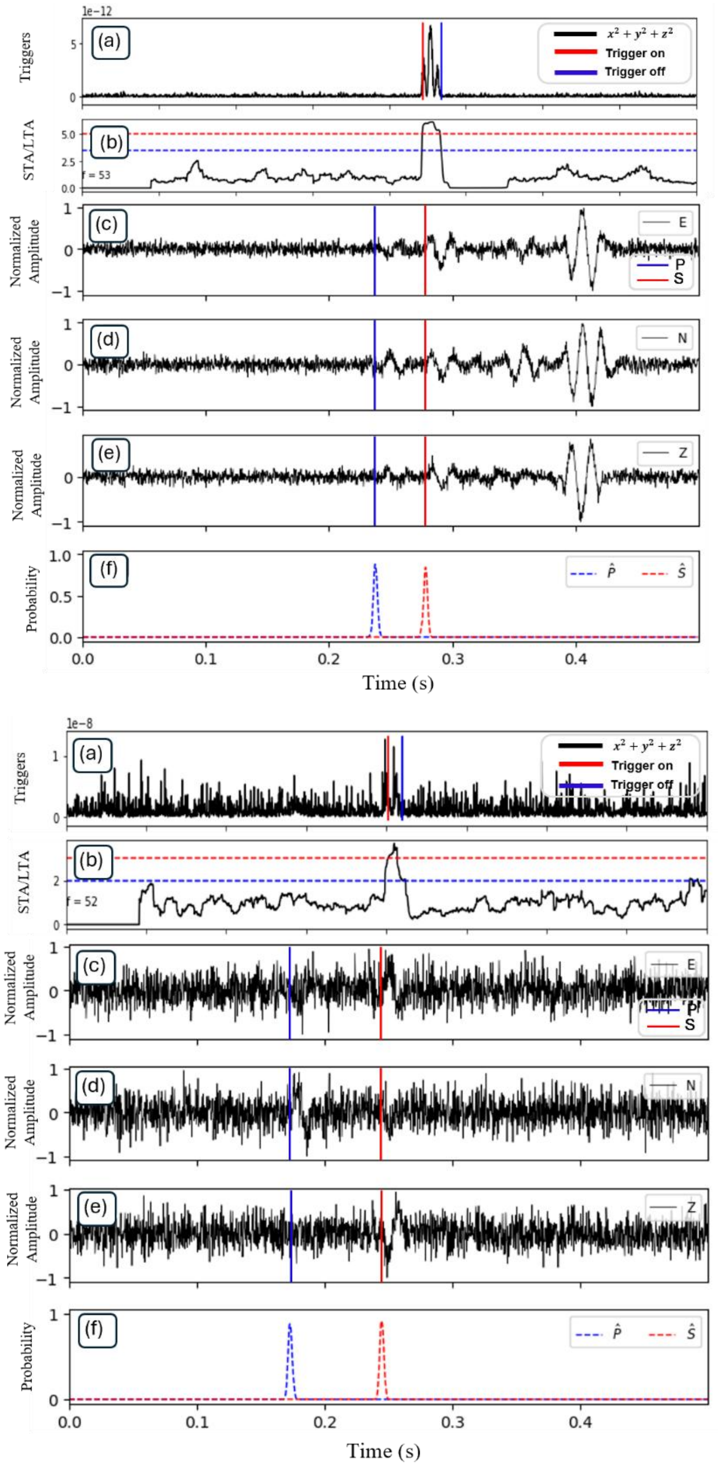


Figure 4-8: Comparison of the PhaseNet and STA/LTA for two waveforms with different levels of noise: (a) the output from the STA/LTA applied to the calculated energy of a specific noisy waveform (Trigger on and off are shown as vertical red and blue lines, respectively, for S-wave arrivals), (b) calculated CF using a sliding window for STA/LTA, (c) the waveform representing the horizontal east-west component, (d) the waveform of the horizontal north-south component, (e) the waveform of the vertical component, and (f) the output generated by PhaseNet. While the STA/LTA method processes the energy of all three components, PhaseNet also operates on these three components simultaneously. The simulated waveforms in this experiment are characterized by source frequencies of 52 Hz (bottom) and 53 Hz (top).

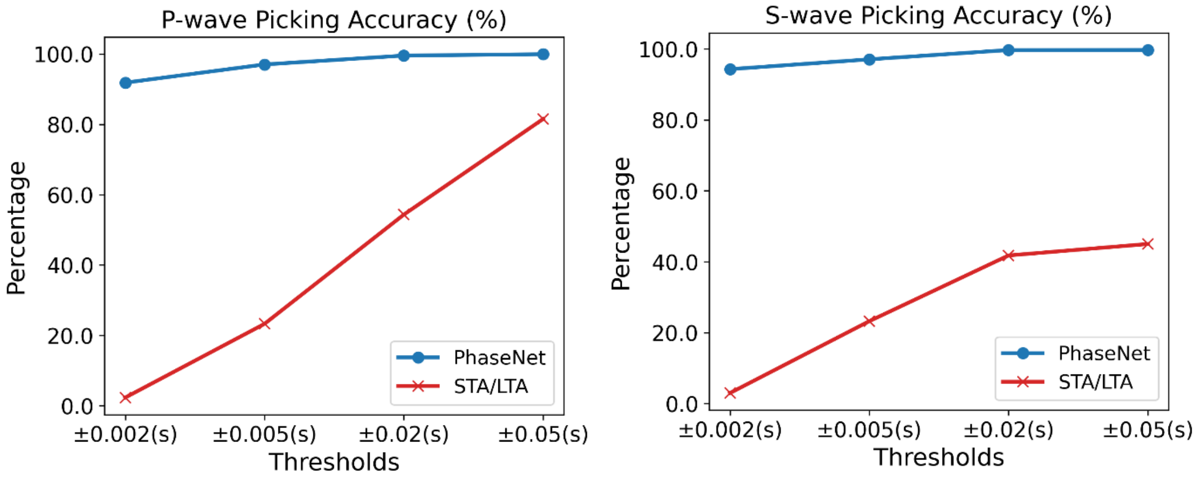


Figure 4-9: Comparative analysis of the performance of PhaseNet and STA/LTA methods across different threshold values for P-wave (left) and S-wave (right).

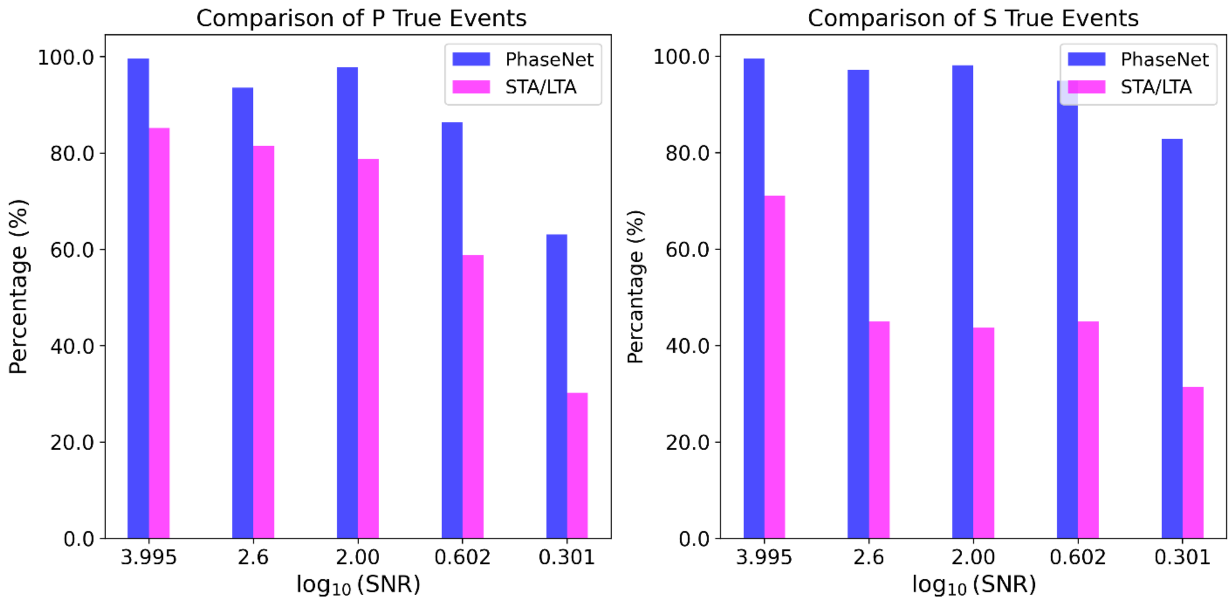


Figure 4-10: Comparison of PhaseNet and STA/LTA algorithm performances at different signal-to-noise ratios in terms of accuracy of predicted picks.

4.5 Conclusion

In summary, as noise levels increase in synthetic seismograms, the STA/LTA method faces significant challenges in detecting P-wave events. The interaction between noise and the parameters used in the STA/LTA calculation, including threshold values, window lengths, and inherent causality constraints, plays a pivotal role in the effectiveness of this specific conventional algorithm. The intricacies of different wave types must also be considered when discussing the detection capabilities of the STA/LTA method. While P-waves travel faster with higher frequencies, S-waves generally have lower frequencies and take longer to emerge. In a high-noise scenario, determining the precise moment when the P-wave transitions to the S-wave becomes increasingly difficult. The detection algorithms may struggle to distinguish the subtle changes in signal characteristics, making it challenging to accurately determine the P- and S-wave phases.

Furthermore, the comparison between the STA/LTA method and PhaseNet highlights that while both are valuable for seismic event detection, PhaseNet demonstrates superior performance, especially in noisy environments. Its capacity to maintain high accuracy without requiring intricate window parameterization makes it a more reliable option for detecting low-magnitude events. As seismic monitoring evolves, incorporating methods like PhaseNet could significantly enhance our ability to detect and analyze seismic activity accurately.

5 PERFORMANCE ON FIELD DATASET

5.1 Introduction

In the realm of deep learning, model generalization is a critical attribute that defines how effectively a trained model can apply its learned knowledge to new, unseen datasets. This capability is particularly relevant in practical applications, where the model robustness and reliability are usually tested under varying conditions. To rigorously evaluate the generalization performance of our model, which has been trained on synthetic microseismic data, we conduct a comprehensive assessment using a carefully selected subset of real seismograms. This real dataset consists of authentic waveform data collected from the mine that provided the foundation for our synthetic dataset, ensuring a relevant comparison. Importantly, some of the microseismic recordings are accompanied by quality-controlled arrival times, allowing us to benchmark the performance of our deep learning model against a reliable standard. By analyzing how the model performs in this real-world context, we can gain a deeper understanding of its strengths and weaknesses, ultimately providing valuable insights into its applicability for detecting microseismic events within different complex and noisy environments. This evaluation not only highlights the model ability to generalize across different datasets but also reinforces the potential of leveraging synthetic data for training robust deep learning models in seismic monitoring applications.

In this chapter, we aim to delve into the performance of our pre-trained deep learning model by rigorously examining its ability to generalize when applied to real microseismic data. Our primary objective is to assess how effectively the model, initially trained on synthetic datasets, translates its learned patterns and insights to actual field recordings collected from the proposed mine that served as the basis for generating our synthetic data. To achieve this, we will systematically evaluate the model performance on a subset of real seismograms, focusing on its accuracy in picking P- and S-wave arrival times in these recordings. Experiments were carried out using real data across three different scenarios. The first scenario involves training the model exclusively on real data. In the second scenario, both synthetic and real data are used in the retraining process of the pre-trained model. The third scenario focuses on retraining the pre-trained model using only synthetic data containing other types of source mechanisms. The effectiveness of each of these trained models will be evaluated using a real test dataset.

5.2 The Mine Deposit

The geology and characteristics of the ore deposit, as outlined in section 2.2.1, provide important context for understanding the specific mine being studied. Understanding the lithology and structure of the mine enables improved interpretation of the data, allowing operators to make informed decisions regarding mining practices and safety protocols. It creates a feedback loop where geological knowledge informs monitoring strategies, and monitoring results further refine our understanding of the geological conditions.

The copper-nickel mining operation is situated between Sudbury Regional Airport and Wanapitei Lake. This mine, part of the Sudbury Integrated Nickel Operations, has been active since 2009. A significant geological feature of the Sudbury area is the Sudbury Basin (Figure 5-1), which measures 60 by 25 kilometers and is the result of a meteoric impact that occurred during the Middle Proterozoic era. The mineralization in the region formed as molten rock cooled. The Sudbury structure intersects three distinct geological provinces: the southern edge of the Upper Archean Province, the northern margin of the Proterozoic Southern Province, and the northern boundary of the Proterozoic Grenville Province. The rock types within the Sudbury Structure can be categorized into three main groups: the hanging wall, the Sudbury Igneous Complex (SIC), and the breccias, along with sedimentary rocks that have filled the basin post-impact. The SIC has been central to exploration, mine development, and the production of Ni-Cu-PGE ores for over a century (McLean et al., 2005).

The mineral deposit was discovered in 2001 and became operational in 2009. While nickel and copper are the primary outputs, other byproducts include cobalt and various precious metals such as gold, silver, platinum, and palladium. The favorable geological conditions, known mineralization patterns, and geophysical data have facilitated effective exploration, leading to the discovery of the deposit located at depths ranging from 1,100 to 1,700 meters below the surface. This contemporary facility employs mining techniques in conjunction with backfilling to facilitate extraction. The deposit exhibits structural complexities, lacking ductile shear but featuring three significant brittle fault systems: steeply dipping north-south faults, low-angle faults, and steeply dipping faults oriented northeast and southwest. The interplay of high in situ stress and mining-related stresses leads to rock fracturing and seismic activity, which are common characteristics of the rock mass. Although seismic incidents are infrequent, they can occasionally result in stability issues, such as ground blows or falls due to micro-earthquakes, posing potential safety risks.

The mine resource comprises three primary mineralized zones that span approximately 720 meters along a northwest-southeast corridor. These zones include the main contact zone (CT), a secondary contact zone (CTII), and a principal footwall zone (FW) (Figure 5-2). The contact zones are primarily rich in nickel, while the footwall zones contain copper, nickel, and precious metals, including platinum group elements, gold, and silver. The designated level names indicate the depth in meters below the surface. Access to the mine is provided through a twin shaft system, with a hoisting shaft extending down to 1,730 meters and a ventilation shaft descending to 1,700 meters below the surface (Jalbout et al., 2014).

The primary contact zone has been thoroughly examined and characterized across an area of approximately 590 meters in plunge length and 360 meters in strike length. The thickness of the contact zone exhibits a range of 5 meters to 105 meters. This zone typically follows the SIC footwall contact along its base, with the most substantial intersections found in lower topographic areas or depressions in the footwall contact (Figure 5-3). In contrast to the more distinct lower contact, the upper boundary of the mineralization is generally less defined and more gradual (McLean et al., 2005).

5.2.1 Contact Zone

At Nickel Rim South, the contact mineralization is characterized by massive and semi-massive nickel-bearing sulfides. This steeply dipping zone, with an inclination of approximately 70 degrees, is accessed from the hanging wall. Development for access is executed within norite units (DNOR and FNOR), utilizing crosscuts through breccia units to reach the sulfides. Typically, the ore contact in the hanging wall is dictated by grade, resulting in a low-grade disseminated sulfide halo, whereas the footwall exhibits a more distinct and sharp contact. All rock units possess good mechanical quality, with a Geological Strength Index (GSI) exceeding 65. Although they are generally more fractured than the footwall units, which boast a GSI of around 80.

5.2.2 Footwall Zone

The mineralization style in the footwall area of the mine closely resembles that found in other Sudbury footwall deposits, commonly referred to as copper zones. A key distinction here is the proximity to the contact zone. Typically, these two mineralization types are separated by several hundred meters; however, at the considered mine, they converge at around a depth of 1,320 meters and are subsequently divided by only a few tens of meters below this horizon. The mineralization is notably associated with a zone of Sudbury Breccia (SDBX), while the footwall development occurs within a Felsic Gneiss unit (FGN). Both types of rock are fine-grained and hard granitic varieties, which, despite being highly competent, may be vulnerable to rockbursts at significant depths. The copper sulfides manifest as an irregular assemblage of veins, varying in thickness from a few centimeters to several meters. Overall, they reside within a steeply dipping (70-degree) mineralized envelope, which is also accessed using the blasthole stoping method. A significant disparity in strength exists between the intricate copper veins and the surrounding host rock, resulting in preferential fracturing and deformation of the veins. It is important to note that footwall zones tend to generate considerably more seismic activity when compared to the nickel ore (Jalbout et al., 2014).

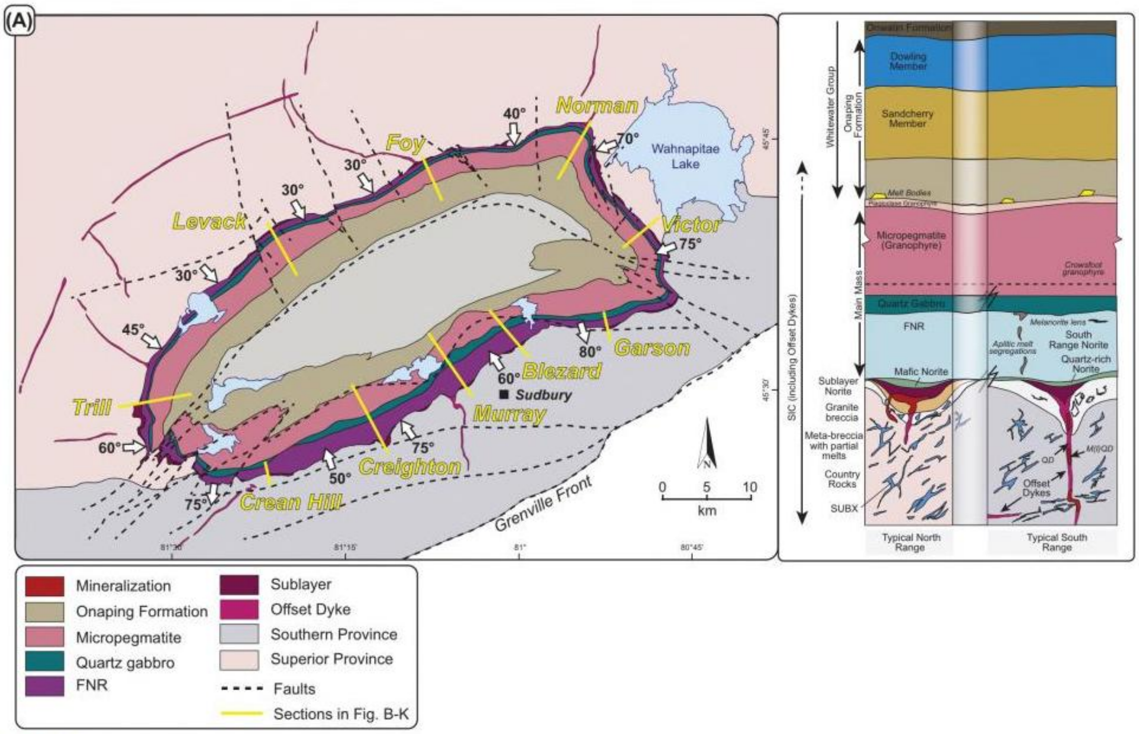


Figure 5-1: Map of Sudbury Igneous Complex (containing the specific mine under study) main mass lithologies (Lightfoot, 2016).

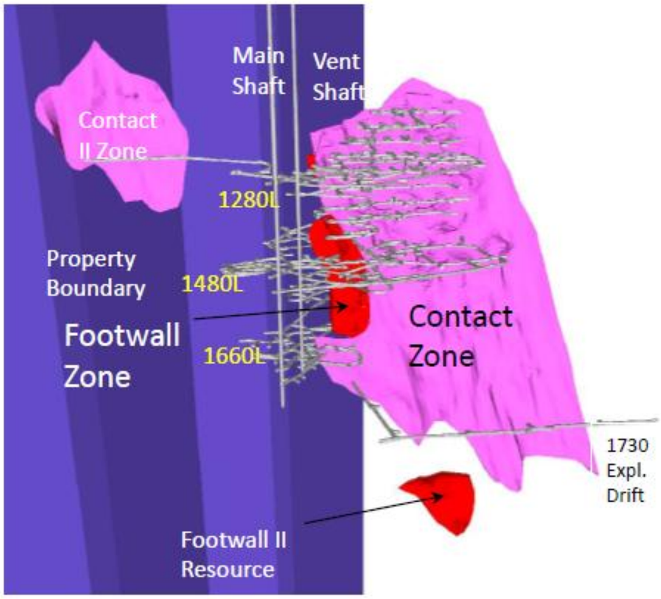


Figure 5-2: A longitudinal perspective of the geological framework for the mine under study: It reveals that the uppermost part of the deposit is situated roughly 1,150 meters beneath the surface, while existing mining operations are planned to reach depths of 1,710 meters. Ongoing exploration efforts are focused on identifying deeper targets, such as the footwall II resource, which lies below 2,000 meters.

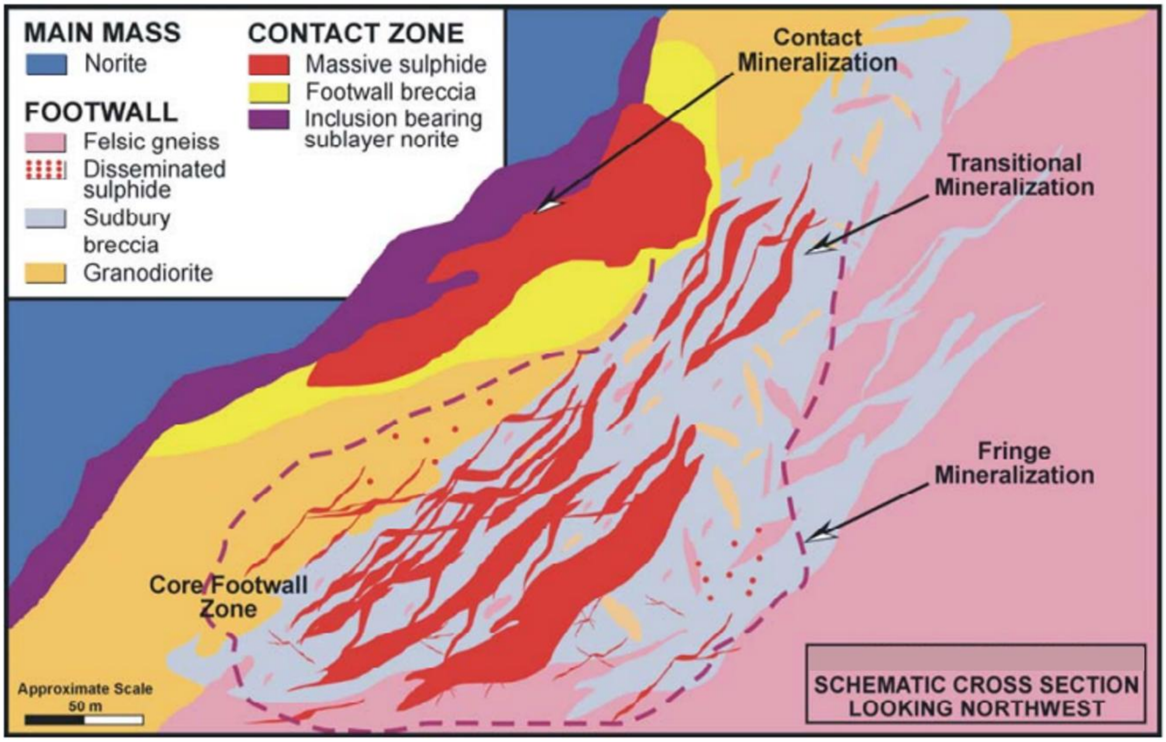


Figure 5-3: Schematic cross section of the mine deposit (McLean et al., 2005)

5.3 Microseismic Data

The microseismic database comprises data files in proprietary Hyperion Seismic Format (HSF), spanning a three-month period that includes events of varying magnitudes. During this period, the sensor array included 6 uniaxial, 26 triaxial geophones, and 39 accelerometers operational in the mine monitoring zone. The P- and S-wave onset times have been picked either by an advanced automatic processing algorithm or by a human interpreter, and all onsets are quality controlled through hypocenter estimation. Some traces of the real data are shown in Figure 5-4 and Figure 5-5.

We converted the microseismic data initially stored in HSF format to SEG-Y format for broader compatibility (Barry et al., 1975). Afterward, we further transformed the SEG-Y files into NPZ format for usage in the Python environment, in which PhaseNet is implemented. The seismograms from the active mine notably differ from the data used to train PhaseNet in this study in terms of the number of samples, input data length, and sampling rate. Table 5-1 illustrates the distinct features of both datasets, particularly highlighting that the real samples are shorter in time (measured in seconds) and include only the z component in some of the channels.

Table 5-1: Major characteristics and differences of the synthetic and real data used for training PhaseNet.

Dataset	Synthetic data	Real data
Type	Passive seismic	Passive seismic
Number of samples	96 761	5 134
Input data channels	3C (x, y, z)	3C (x, y, z) and 1C (z)
Input data length	0.5 (seconds)	0.2 (seconds)
Sampling rate	4000 (Hz)	10000 (Hz)
P- and S-wave labels	Ray-tracing estimations	Quality-controlled labels

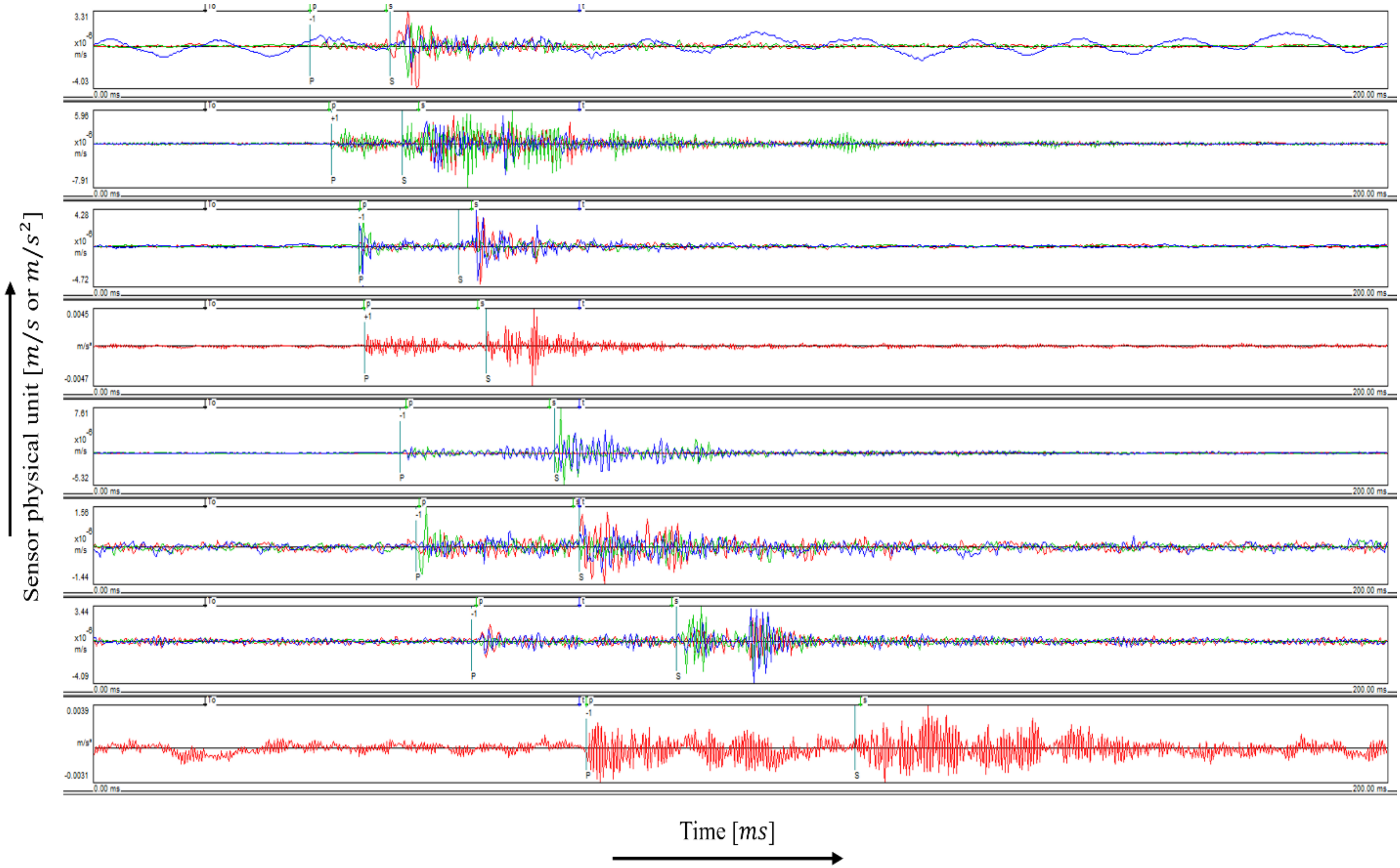


Figure 5-4: Illustration of one event recordings: The panel displays seismograms for an event with a magnitude of -1.8 (channels 4 and 8 are from 1C accelerometers, while the other channels are from 3C geophones).

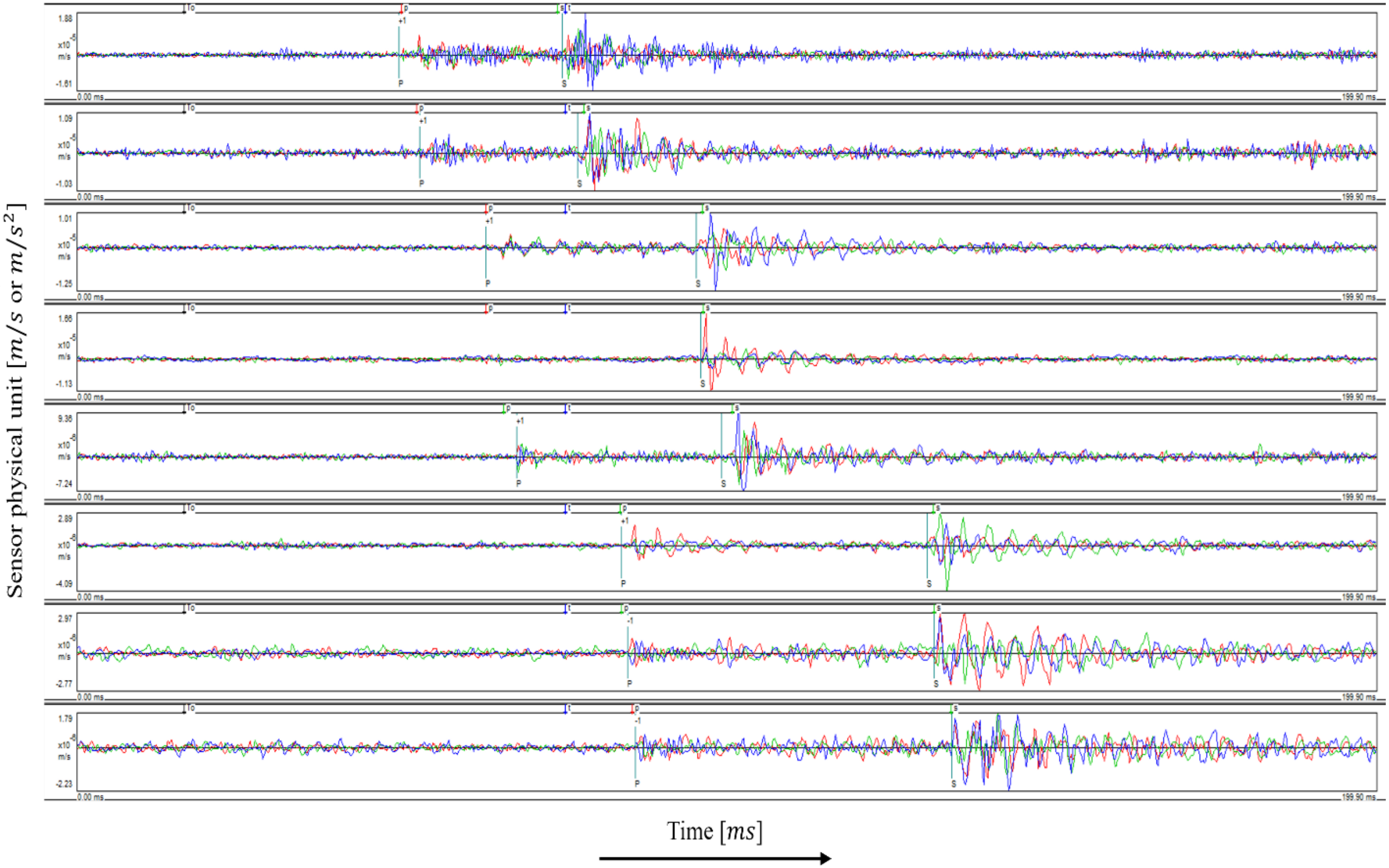


Figure 5-5: Illustration of one event recordings: The panel displays seismograms for an event with a magnitude of -1.3 (all 8 channels are from 3C geophones)

In seismology, the magnitude of seismic events can often be described using the Gutenberg-Richter (G-R) law, which states that there is a relationship between the number of events (N) and their magnitudes (M) (Gutenberg et al., 1944). This law can be mathematically expressed as:

$$log_{10}N(M) = a - bM \tag{5.1}$$

where $N(M)$ is the number of seismic events exceeding a particular magnitude M, a is a constant dependent on the dataset and is higher for more active regions, and b is called the b-value, which represents the relative proportion of small vs. large events. The b-value is a crucial parameter in this context. It represents the slope of the line in a log-log plot of the frequency-magnitude distribution. A higher b-value (typically around 1) indicates a relative abundance of smaller events compared to larger ones, whereas a lower b-value suggests a greater proportion of larger events. This relationship provides insights into the stress state of the geological environment. For instance, a decreased b-value can indicate increased tectonic activity or a more stressed system, as larger events are more frequent in such conditions. Conversely, a higher b-value may signify a more stable environment where smaller events occur more commonly.

Analyzing the magnitude distribution and calculating the b-value of microseismic events are valuable for understanding the seismic behavior of a geological setting, including potential hazards and the dynamics of rock mass failure.

The b-value is often estimated using the Maximum Likelihood Method (MLM) as:

$$b = \frac{log_{10}e}{M_{mean} - M_{min}} \tag{5.2}$$

where M_{mean} denotes the mean magnitude of events, M_{min} is the magnitude of completeness, denoting the smallest magnitude that can be reliably detected.

Figure 5-6 illustrates the magnitude distribution of total real events that occurred in a day during the mine operation. The distribution shows a linear trend on the log–linear plot, which is exactly what we expect. The b-value ≈ 1 suggests a “normal” scaling that each unit decrease in magnitude increases event frequency by ~10×. Since there are negative magnitudes, this indicates microseismic events (small local earthquakes). The completeness magnitude (Mc), the lowest magnitude above which the catalog is reliable, looks close to -1.6, since this is where the straight-line fit starts.

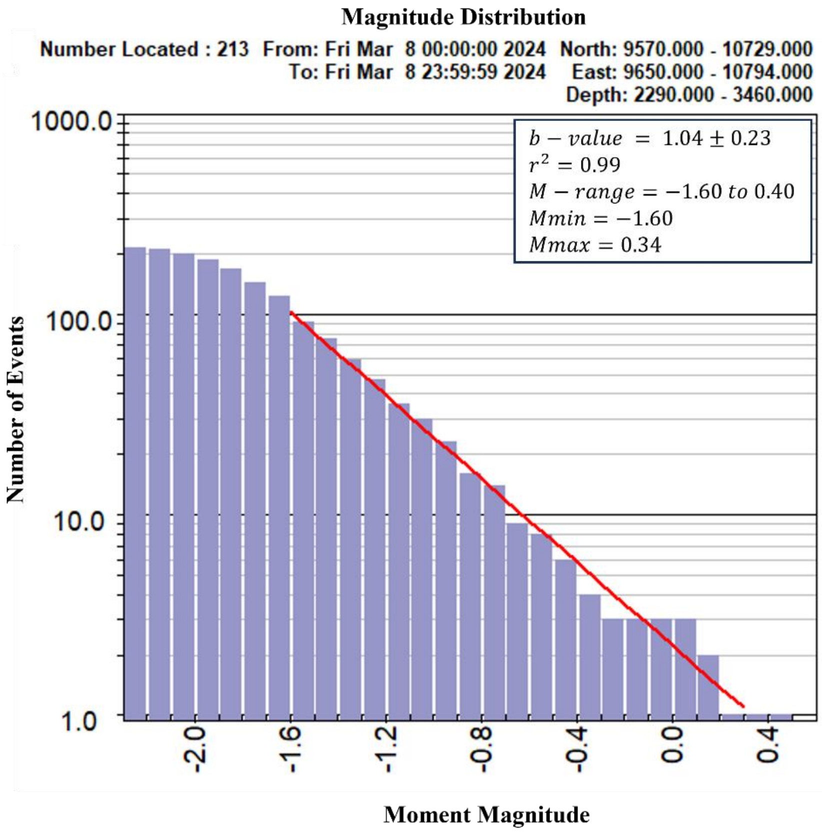


Figure 5-6: Magnitude distribution of the real seismograms.

5.4 Performance of the Pre-trained Model on the Real Dataset

To effectively assess the performance of the network, it is essential to construct a training dataset derived from mine data. The labels for this dataset are generated by utilizing QC-controlled arrival timestamps. In line with the methodology proposed by Zhu et al. (2019), this process involves creating Gaussian distributions that are centered around the arrival times, with a fixed standard deviation set at 10% of the total signal length. Additionally, the output of the network necessitates post-processing to extract the arrival times from the segmentation maps. This extraction is accomplished using a peak detection algorithm, which employs various parameters to determine the minimum peak height or the minimum interval between consecutive peaks. This process involves detecting local maxima within the output and applying specific thresholds and constraints to validate the picks.

Key parameters for the peak extraction include ‘prob threshold’, which establishes the minimum probability for a peak to be recognized as valid (generally ranging from 0.1 to 0.5, with a common default of 0.3); ‘distance’, which specifies the minimum number of samples between detected peaks to minimize duplicates; and ‘peak prominence’, an optional parameter ensuring that a peak is noticeably distinct from surrounding values, thereby filtering out low-amplitude noise. Furthermore, options such as ‘suppress multiple picks’ prioritize the highest peak when multiple peaks are in proximity, while ‘phase margin’ dictates the minimum time gap between P- and S-picks for coherent sequencing. The ‘window’ parameter permits a search duration following the first high-probability value, typically extending over a few seconds, and the ‘sampling rate’ influences how these thresholds and distances are interpreted in time, usually set at 100 Hz. In summary, carefully calibrating these parameters is crucial, especially in noisy environments like mining settings, to achieve a balance between sensitivity and precision in the detection of seismic events (Table 5-2).

Table 5-2: Key parameters for peak extraction

Parameter	**Description**	**Typical value**
Prob threshold	Minimum probability value for a peak to be considered a valid pick	Often 0.1–0.5 (default: **0.3**)
Distance	Minimum number of samples between detected peaks (prevents close/fake duplicates)	Depends on sampling rate, e.g., 20–40 samples
Peak prominence (optional)	How prominent a peak must be compared to the surrounding values	Useful to filter noisy low-amplitude peaks
Suppress multiple picks	Only return the highest peak if multiple peaks are close	Often enabled to avoid false picks
Phase margin	Minimum time between P and S picks (e.g., S can't come before P)	Optional logic for consistent sequencing
Window	Search window after the first high-probability value	Often ± a few seconds, depending on sampling rate
Sampling rate	Affects how thresholds and distances are interpreted in time	Typically, 100 Hz in PhaseNet papers and code

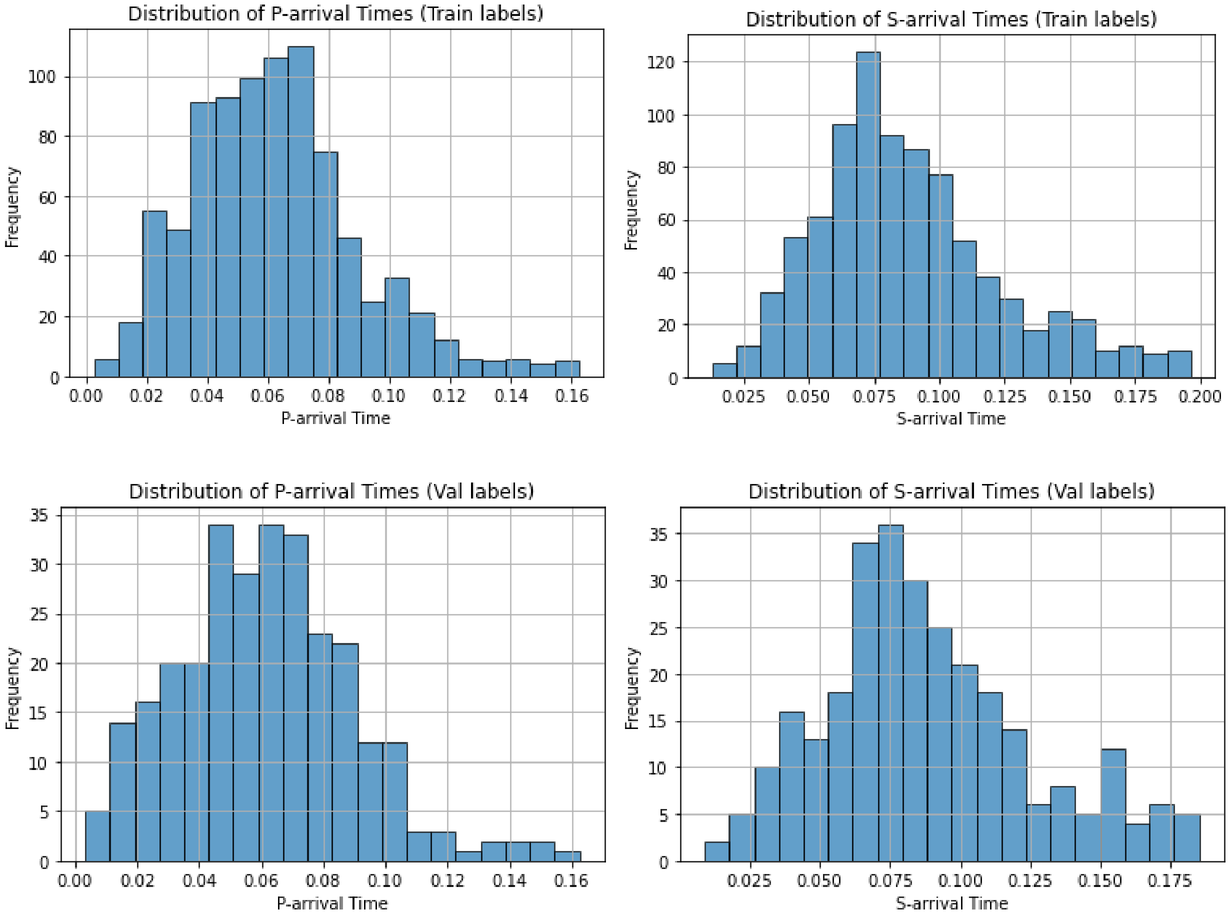


Figure 5-7: P- and S-wave arrival distributions for the training and validation sets of real mine data, used for retraining and fine-tuning the pre-trained model.

To create a balanced and representative dataset for training and validating the network while fine-tuning the pre-trained model with synthetic data, careful attention was given to the labeling of P- and S-wave arrivals. The distributions of these arrivals in both the training and validation datasets were designed to reflect the fundamental characteristics of real seismic data. In analyzing the overall dataset, samples were selected to maintain similar proportions of P- and S-wave arrival events across both the training and validation sets, which prevents bias towards one type of phase. This strategy ensures that the network can learn effectively from a variety of scenarios and generalize well to new, unseen datasets. The four distributions of P- and S-wave arrivals for training and validation exemplify this careful balance and underscore the commitment to preserving the integrity and representativeness of the data throughout the process, as illustrated in Figure 5-7.

To maximize the performance of the network, we aim to use the features already learnt by PhaseNet when trained on the synthetic dataset. However, this dataset is different from real dataset, in terms of scale, frequencies, experimental conditions, sensors, etc. We thus propose to use part of the original model features while retraining the network on our simulated and real data, a process referred to as transfer learning (TL) (Figure 5-8) (Chai et al., 2020). TL is a type of machine learning that uses a pre-trained neural network to speed up the learning process of a new task. To solve a problem, we need to have a pre-trained model of a similar problem. Instead of building a model from scratch to solve a similar problem, we use the model trained on another problem, usually a larger set of dataset, as a starting point (Pan et al., 2009).

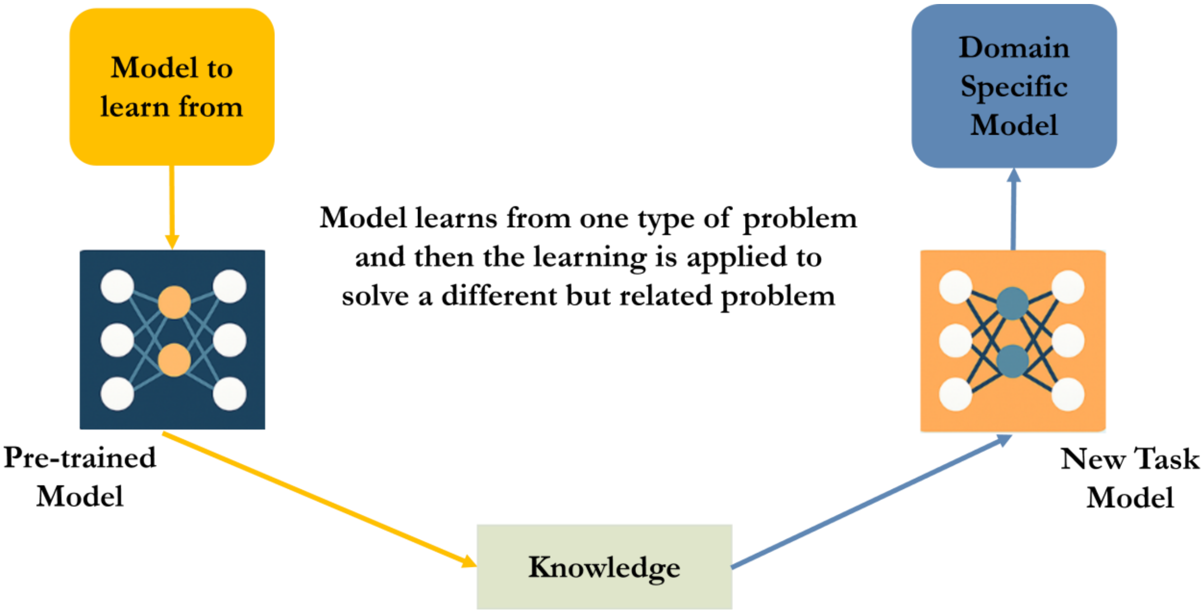


Figure 5-8: Transfer Learning (TL).

We use transfer learning to save time or to get better performance, as it is an optimization problem.

As in Figure 5-9, the "without transfer" curve serves as a baseline, demonstrating learning and performance in the absence of relevant prior experience. There are three possible benefits to look for when using transfer learning:

Higher start: The initial skill on the source model is higher than it otherwise would be.

Higher slope: The rate of improvement of skill during training of the source model. That is steeper than it otherwise would be.

Higher asymptote: The converged skill of the trained model is better than it otherwise would be.

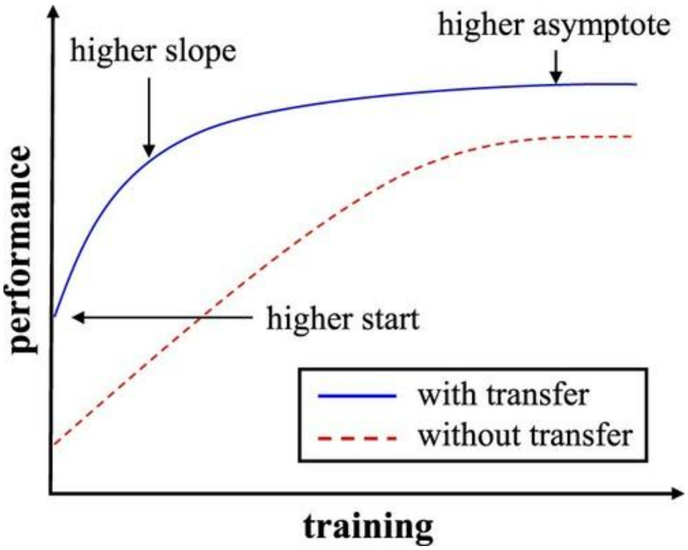


Figure 5-9: The curve illustrates the concept of transfer learning, showing how prior knowledge or training can impact subsequent learning and performance (Ko et al., 2021).

Three distinct strategies are employed to evaluate the impact of model pre-training on arrival-time estimations for the real dataset. The first approach (Scenario 1) involves training PhaseNet on the real dataset alone, where quality-controlled labeled arrivals are used to identify arrival patterns specific to the real mine. Although this approach adjusts model 1 to the real-world data, its effectiveness may be constrained by the size, quality, and diversity of the labeled dataset.

Scenarios 2 and 3 both involve using a pre-trained model to improve performance on a real mining dataset. In Scenario 2, the model was first trained on a synthetic seismic dataset explained completely in Chapter 3. It was then retrained using both synthetic and real seismic data before being fine-tuned on the real mining dataset. This final step allowed model 2 to adapt its learned features to the specific characteristics of actual mining-induced seismic events. In contrast, Scenario 3 relied solely on a pre-trained model that had been re-trained only on a smaller set of synthetic data from other source mechanisms, without any further re-training on real seismic events, which results in model 3 (Figure 5-10).

Figure 5-11 shows six different seismograms from a real dataset, along with their true and predicted probability distributions as the input and output segmentation maps according to Scenario 2.

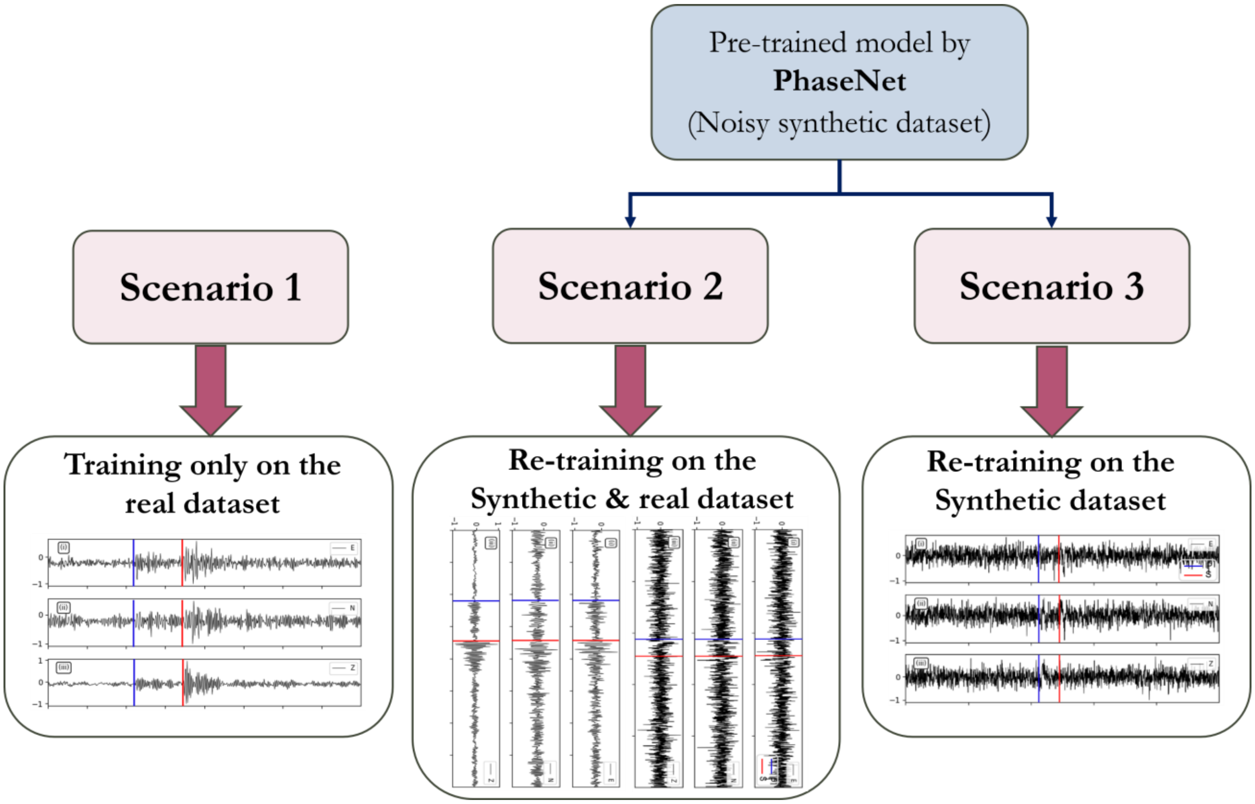


Figure 5-10: Three distinct strategies that are employed to evaluate the impact of model training and pre-training on arrival-time estimations for the real dataset.

The confusion matrices for the real test dataset, shown in Figure 5-12, reveal a clear advantage of using a pre-trained PhaseNet model compared to training directly on the real mining data. The number of timestamps of the real dataset is 5,134, of which 5% is considered as the test dataset, which contains 256 P-wave and 256 S-wave arrivals. The pre-trained model exhibits a higher number of True Positives (TP) and a reduced number of False Negatives (FN), indicating its ability to detect more seismic arrivals while minimizing missed events. The lower False Positives (FP) count reflects the model 2 improved precision, distinguishing between genuine arrivals and noise. These results highlight the advantage of synthetic data simulation and the pre-trained model ability to balance sensitivity and specificity, especially in noisy or ambiguous scenarios.

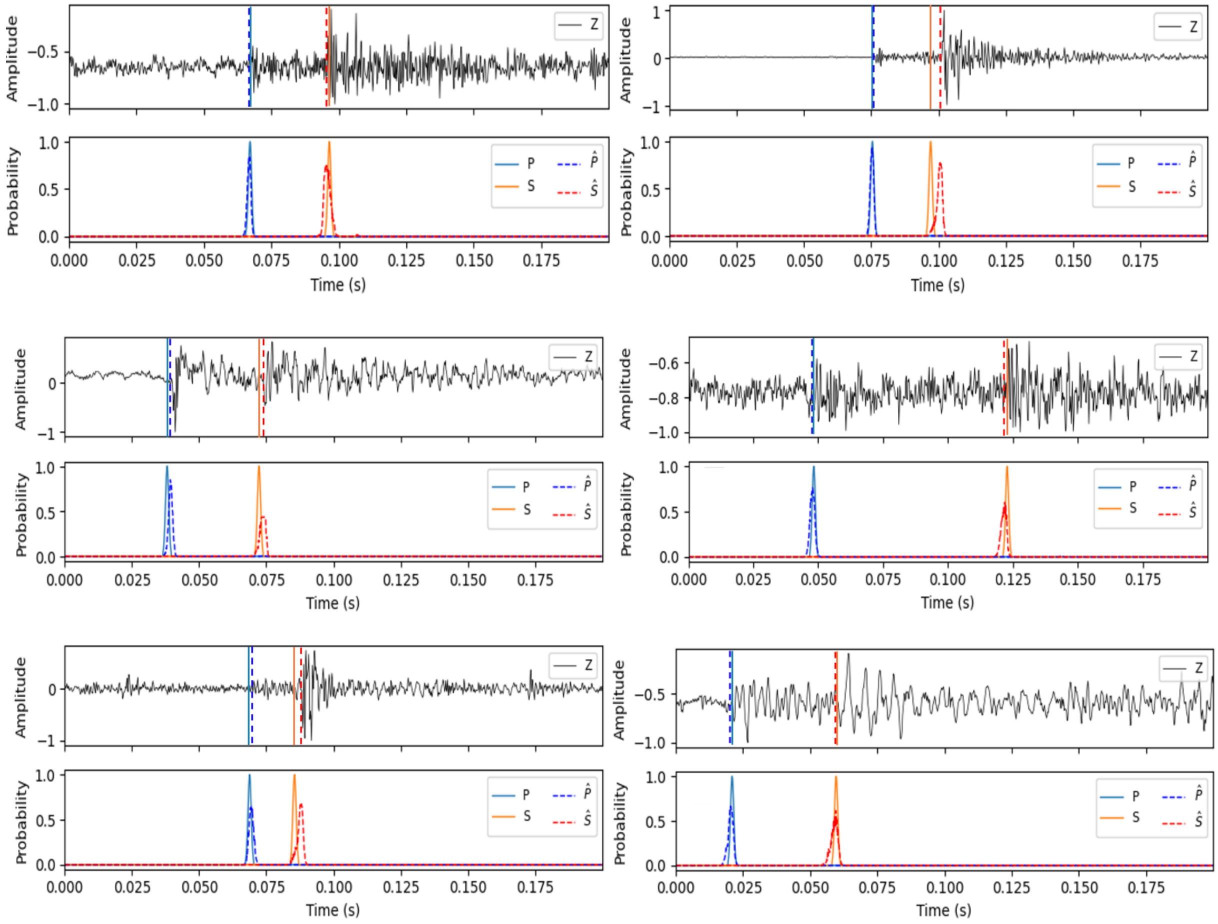


Figure 5-11: Input and output segmentation maps of test data for PhaseNet: Each panel displays a seismogram (only the Z-component) derived from the real mine dataset (top) and the input segmentation map for P- and S-waves (bottom), represented by continuous blue and orange lines, respectively. The output segmentation map predicted by the pre-trained model in PhaseNet is represented by dashed dark blue and red lines. The vertical lines indicate the manually picked arrivals, and the vertical dashed lines denote detected peaks from the segmentation maps.

(a) **(b)** **(c)**

Figure 5-12: Comparison of confusion matrices for prediction of P- and S-wave arrivals on the real dataset for (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

Residual analysis further underscores the benefits of pretraining, as illustrated in Figure 5-13. In Scenario 2, where model 2 was first trained on high-noise synthetic data and then fine-tuned with real mining data, the residual distribution is noticeably tighter. Most predicted arrival times cluster closely around the actual values, indicating enhanced accuracy and a reduction in outliers compared to Scenario 1, which lacked pretraining and displayed a wider spread of values. While P-wave residuals tend to be smaller and more consistent, the pre-trained model also shows significant improvements in the more challenging S-wave arrival predictions. Conversely, Scenario 3, which involves training solely on synthetic data without any fine-tuning with real data, yields a less refined residual distribution, underscoring the crucial role of real data in the training process.

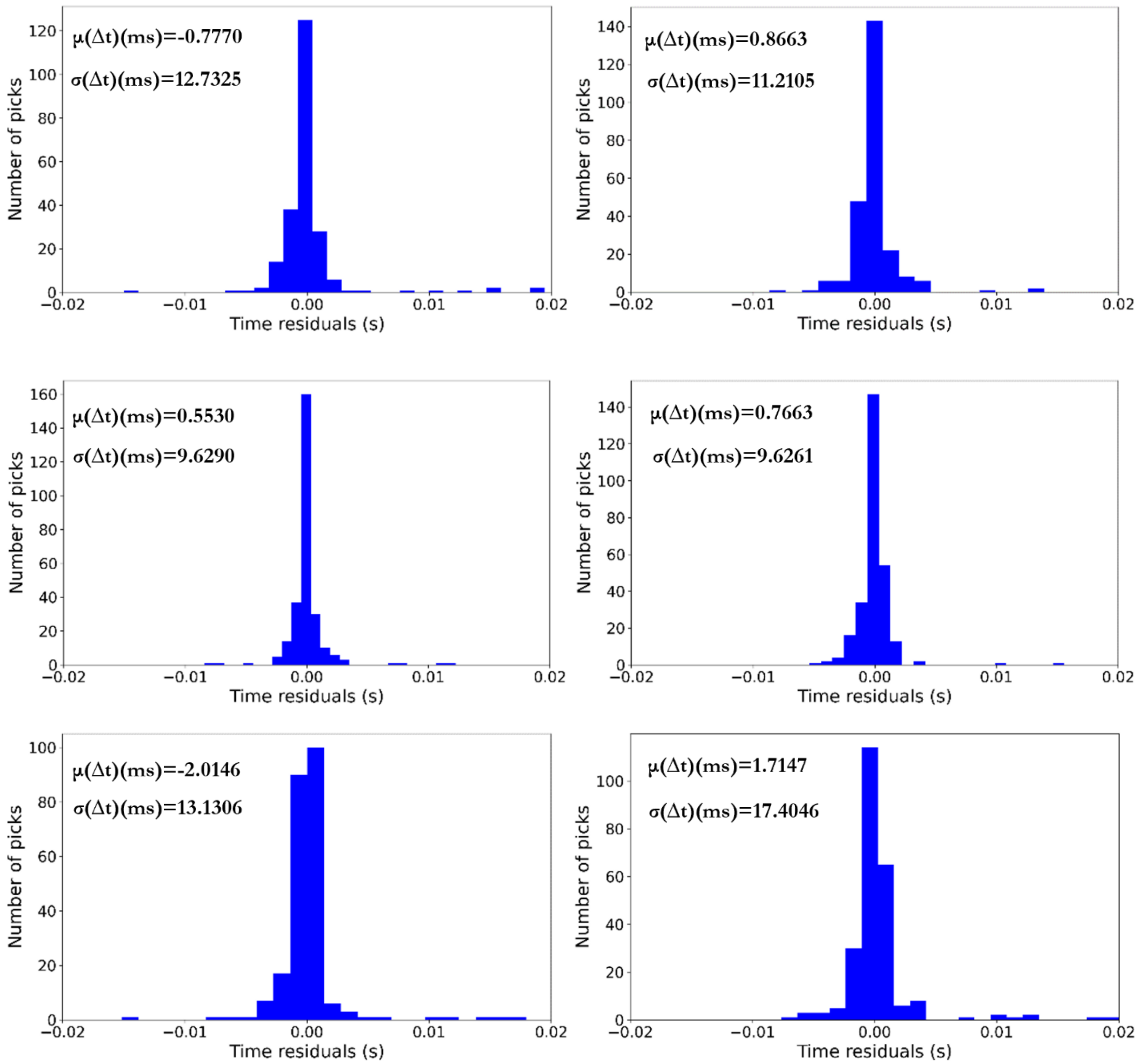


Figure 5-13: Comparison of the residual analysis of P- (left column) and S-wave (right column) arrival times for Scenario 1 (first row), Scenario 2 (second row), and Scenario 3 (third row).

The evaluation metrics presented in Table 5-3 provide a detailed comparison of the three training approaches, highlighting the impact of pretraining and synthetic data simulation on model performance. The pre-trained model in Scenario 2, which was first trained on high-noise synthetic data and later fine-tuned with real seismic data, achieves the highest accuracy, precision, recall, and F1-score for both P-wave and S-wave arrivals. This demonstrates its superior generalization ability and robustness in detecting seismic events. The use of synthetic data in this scenario plays a crucial role in improving the model's ability to handle diverse noise conditions and event characteristics, enabling it to learn essential seismic patterns even before being exposed to real mining data. The higher precision in Scenario 2 indicates a lower rate of false detections, while the higher recall reflects its capability to capture a greater proportion of true arrivals. The strong F1-score further validates its balance between precision and recall, making it the most reliable approach for detecting seismic arrivals in mining environments. In contrast, the model of Scenario 1 performs moderately well but lacks the benefits of pre-training with synthetic data, resulting in lower detection accuracy. The pre-trained model in Scenario 3, which was trained only on synthetic data without fine-tuning on real data, shows the weakest performance, struggling to adapt to real-world seismic patterns. These results underscore the importance of synthetic data simulation for initial training, as well as the necessity of incorporating real data for fine-tuning. The combination of both enables model 2 to bridge the gap between simulated and real seismic signals, significantly enhancing detection accuracy and reliability in complex mining environments. The PhaseNet prediction results for some waveforms of the real data are shown in Figure 5-14.

Table 5-3: Metrics for P- and S-wave arrival predictions using pre-trained and non-pre-trained models for a real dataset.

	P-wave Arrival			**S-wave Arrival**		
Evaluation Metrics	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**
Accuracy	0.621	0.813	0.561	0.713	0.832	0.584
Precision	0.845	0.915	0.784	0.851	0.927	0.780
Recall	0.712	0.893	0.664	0.782	0.901	0.699
F1-score	0.772	0.904	0.719	0.815	0.914	0.737
$\mu(\Delta t)$(ms)	-0.7770	0.5530	-2.0146	0.8663	0.7663	1.7147
$\sigma(\Delta t)$(ms)	12.7325	9.6290	13.1306	11.2105	9.6261	17.4046

Figure 5-15 illustrates the performance of the trained models across three mentioned distinct scenarios, utilizing unseen real microseismograms. The figure comprises three panels, each showcasing multiple waveforms captured by authentic sensors, aligned with the arrival times of

P-waves. Panel (a) corresponds to Scenario 1, while panel (b) depicts Scenario 2, and panel (c) represents Scenario 3. This visual representation highlights the variations in waveform characteristics among the different scenarios, providing insights into the model effectiveness in analyzing microseismic data under diverse conditions.

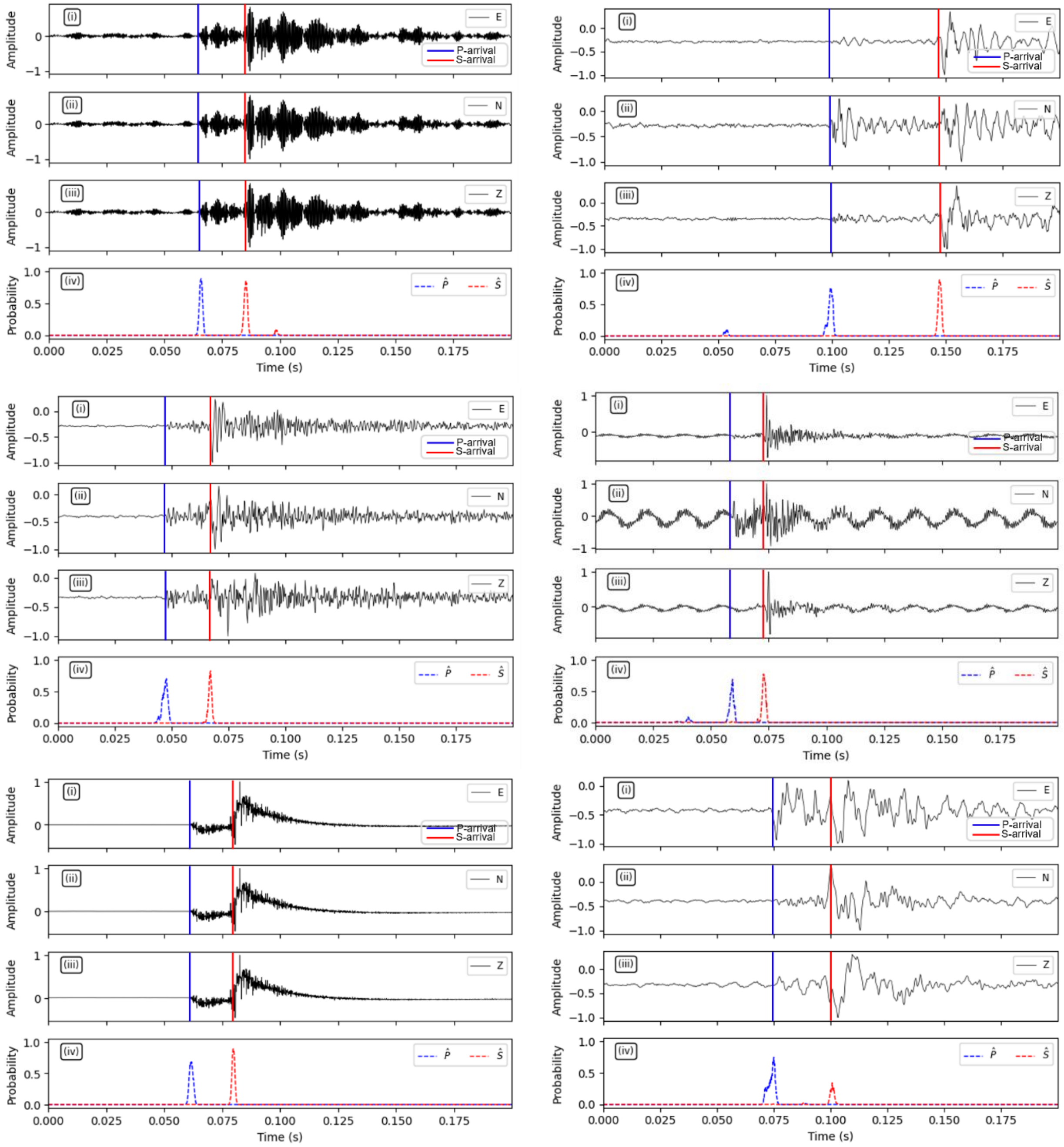
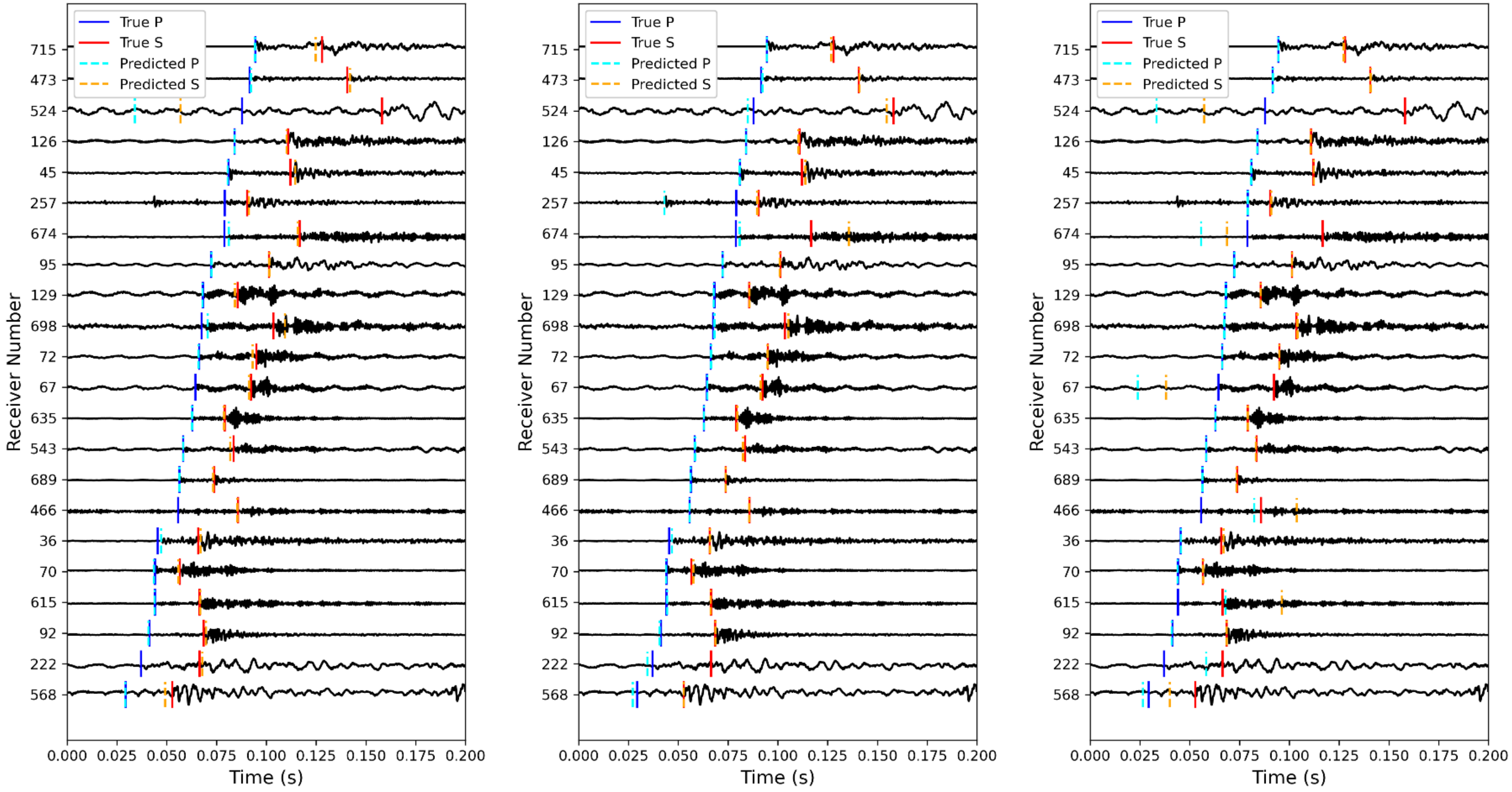


Figure 5-14: Visualization of some waveforms from real test datasets that are picked for P- and S-wave arrivals based on the re-trained model across scenario 2.



(a) **(b)** **(c)**

Figure 5-15: Performance of the trained models (1, 2, 3) across the three Scenarios on unseen real seismograms. The three panels display several waveforms received by true sensors with respect to the arrival times of P-waves: (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

5.5 Conclusion

The results presented in this chapter emphasize the effectiveness of the trained model, which was developed using synthetic seismograms generated within the mine area, particularly its ability to predict the arrival times of microseismic waves in real datasets after retraining. Initially, the model demonstrates strong performance in identifying low-magnitude events, which are often challenging to detect using traditional methods employed in mining. However, it is essential to understand that the synthetic data used for training may not perfectly represent the complexities of real microseismic data.

Despite variations in synthetic and actual seismograms, the retraining process with real data fine-tunes the model to accommodate these differences. This involves updating the model weights, specifically the PhaseNet model, to enhance its predictive capabilities when applied to real mine data. By retraining the model using real data, the predictions for both P- and S-wave arrival times are significantly improved. Retraining with just a small portion of real data containing 5134 labeled seismograms allows the model to learn the unique characteristics or noise patterns present in the actual dataset. In essence, while the initial training provides a strong foundation, the retraining process bridges the gap between synthetic and real data, enabling the model to achieve higher precision and reliability in practical applications.

In conclusion, while PhaseNet was initially designed for earthquake seismology data, we adapted its parameters to effectively process and analyze passive seismic data collected from our specific mining site. Our approach involved customizing the model to recognize and interpret the unique characteristics of microseismic waveforms generated by mining activities, enabling event detection.

6 CONCLUSION

6.1 Summary

The challenge of accurately picking arrival times in seismic data remains a critical issue in geophysical studies. Deep learning can play a pivotal role in this context by identifying and learning complex patterns within the data, ultimately enhancing the precision of phase picking. By utilizing deep learning techniques, it is possible to optimize conventional methods, enabling them to effectively process large and intricate datasets. However, integrating deep learning solutions fully into this workflow is challenging due to the complexity of the problem and the limited availability of labeled datasets. Automatic arrival time picking presents a promising and cost-effective approach to addressing this issue, yet the process of labeling thousands of seismograms for training deep neural networks is still a demanding and labor-intensive task.

For the purpose of this work, PassiveSeis was created. It is a novel seismic dataset that offers representative samples of mining-induced seismicity records. The open-source program SeisCL was used to solve the wave equation in a three-dimensional (3D) elastic medium using a finite difference time domain approach. The database is unique in its size, diversity, and meticulous realism. It contains a labeled dataset holding the true P- and S-picks, event positions, and seismic source parameters (source time function, focal mechanism type, and moment tensor). Hence, the dataset can be used to train diverse deep learning models to predict the mentioned features. P- and S-wave arrival times in PassiveSeis are calculated for all source-receiver pairs using the ttcrpy software, which implements various raytracing methods. The shortest path method (SPM) is selected for its accuracy, despite being slightly slower. ttcrpy features two SPM versions: the classical grid-based SPM utilizing Dijkstra's algorithm to compute travel times at all grid nodes and an enhanced Dynamic Shortest Path Method (DSPM) that updates travel times along raypaths for each source-receiver pair. The classical SPM is preferred for its ability to manage sharp velocity variations. The calculations are performed on a refined 3D rectangular grid used to discretize the mining domain, with additional secondary nodes added to improve accuracy and coverage. PassiveSeis contains a total of 272,000 three-component (3C) seismograms with their corresponding P- and S-wave arrival times. These synthetic records are saved in SEGY format, with each simulated event comprising three separate files, one for each of the modeled seismic wavefield components. This results in a dataset that includes 5,100 SEGY files, each labeled by event indices, components, and source mechanisms.

We propose a novel approach for detecting induced microseismic events in mines, leveraging deep learning by creating a dataset that closely resembles our real data in a high-stress underground mine. We demonstrate the potential of using deep learning methods, particularly the PhaseNet architecture, which is trained on synthetic seismograms that mimic data collected in a mine in Ontario, to accurately detect and label (as P or S) seismic arrivals. It is found that PhaseNet outperforms the standard STA/LTA conventional automatic picking in noisy environments for the synthetic data in PassiveSeis. In the proposed approach, the detection of low-magnitude events, a sought-after characteristic of microseismic data, is achieved with high precision, which has been a challenge using conventional approaches. The results indicate that the arrival time predictions of the trained model on the synthetic dataset are sufficiently accurate for data with low SNRs. Moreover, we experimented with various strategies on the simulated dataset. We identified the most effective scenario to assist in training a network that performs better for real-time arrival time prediction. The optimal strategy involved utilizing a pre-trained model on synthetic data and re-training the network with real data. The strategy of retraining the model on real data, leveraging the pre-trained model on synthetic data, achieved high precision in detecting the first P- and S-wave arrival times, ultimately facilitating the labeling of thousands of real seismograms through deep learning.

6.2 Discussion

Accurate P- and S-wave arrival times are essential for developing seismic velocity models, determining hypocenter locations, and performing moment tensor inversions. The precise identification of seismic event hypocenters, where ruptures initiate, enhances the monitoring and assessment of seismic activity, which is particularly important at mining sites. Specifically, accurate S-wave velocity models are integral for effective seismic source localization and moment tensor inversion (MTI) in microseismic studies. Determining V_S accurately enhances the ability to locate seismic events and characterize their source mechanisms, which is crucial for understanding subsurface activities during mineral extraction operations.

One major way to leverage V_S modeling is achieved by analyzing the arrival time differences between P- and S-waves. These disparities offer insights into subsurface velocities, enabling the estimation of V_S values at specific locations. By systematically analyzing the time windows from a dense network of seismometers, we can establish relationships between both wave types. This approach is particularly effective in heterogeneous settings, where geological features such as faults or fractures impact wave propagation, thereby refining spatial and depth-specific estimates of S-wave velocity.

Leveraging accurate arrival times for hypocenter localization involves using these measurements within seismic network data. Utilizing localization algorithms, such as the method of least squares or sophisticated approaches like Bayesian inversion, enables the derivation of the spatial coordinates of the seismic source. The accuracy of this process hinges on the number and quality of arrival time data, which is significantly enhanced by employing deep learning models, such as PhaseNet, in the current study.

Once hypocenter locations are established, the next step is to perform moment tensor inversion. This involves analyzing the distribution of seismic waveforms recorded at various stations to extract meaningful information about the source of the seismic event, to construct a tensor that describes the seismic source mechanism, thereby providing insights into the geological processes at play. More specifically, this step includes examining parameters such as the amplitude and polarity of seismic waves, as well as their arrival times at different locations. By doing so, researchers can identify the spatial orientation of the fault, the type of faulting (e.g., strike-slip, normal, or reverse), and the overall complexity of the rupture process. Through this detailed spatial analysis, scientists gain valuable insights into tectonic activities and the stress regime of the affected region. Accurate hypocenter locations enhance the quality of moment tensor inversions by reducing uncertainties in the source model, thereby providing higher constraints on the inversion process and resulting in more reliable moment tensor solutions. This is particularly important in mining operations, where understanding the mechanics of induced seismicity informs safety protocols and operational decisions (Linzer et al., 2015; Ma et al., 2018; Minson et al., 2008).

In summary, leveraging accurate P- and S-wave arrival times for hypocenter localization followed by moment tensor inversion represents a promising area of research. This approach enhances our understanding of induced seismicity and predictive capabilities, ultimately contributing to safer mining practices and more effective seismic monitoring systems.

Microseismic monitoring holds immense potential for enhancing safety in underground mining operations. By utilizing P- and S-wave arrivals, sophisticated decision-support tools can be developed to alert personnel to potential hazards such as rockfalls or seismic events. This research project aims to further this potential by automating the detection of risks related to ground instability that could lead to gallery collapses. The integration of an early warning system is crucial to this initiative, as it would provide timely alerts to miners, allowing them to take necessary precautions and seek shelter when risks are detected. Future developments could involve real-time monitoring systems that analyze seismic data on-site, delivering instant feedback to workers and enabling adjustments to activities in response to seismic events. This proactive approach not only aims to safeguard workers' lives but also seeks to minimize disruptions to mining operations caused by unforeseen collapses. The practical implications of this research are significant for the mining sector, which continues to grapple with safety and operational efficiency challenges.

Our research highlights promising avenues for further developments in microseismic data analysis through deep learning. The demonstrated superiority of PhaseNet over conventional automatic detection methods, such as STA/LTA, marks a significant advancement, particularly in detecting low-magnitude events that pose challenges due to low SNRs. This success paves the way for exploring how deep learning methodologies can further enhance seismic data processing and analysis in complex mines. As machine learning techniques evolve, exploring state-of-the-art architectures beyond PhaseNet may involve integrating ensemble methods or hybrid models that combine the strengths of multiple approaches. This could improve the precision and efficiency of microseismic event detection and phase picking. EQ-transformer, EPick and recurrent neural networks that have been used for earthquake data can be some choices to be tested (Li et al., 2022; Mousavi et al., 2020; Ross et al., 2018).

Additionally, the successful application of a pre-trained model on synthetic seismograms indicates a promising route for operationalizing deep learning in real-world scenarios. More work could expand the dataset to include diverse mining conditions or integrate data from multiple sources to refine the model further. Investigating PhaseNet adaptability across various geological settings could also enhance its reliability in different mining environments. Testing the model on data from different mines would provide valuable insights into its performance across various contexts. Each mining environment presents unique challenges, such as variations in rock properties and stress conditions. By evaluating the model performance on distinct datasets, we can identify potential weaknesses and areas for improvement, leading to more versatile and effective automatic arrival detection tools.

Enhancing the synthetic dataset by incorporating real noise characteristics is vital for improving model training. As mining seismic data is often contaminated with various noise types, incorporating realistic noise patterns in the synthetic data will train the model to be more resilient against disturbances, thereby further enhancing predictive accuracy in real-world settings. This approach not only strengthens the model against varying noise conditions but also pushes the boundaries of microseismic data simulation. Combining synthetic datasets enriched with real-world noise and extensive labeling will likely yield a more comprehensive training resource, enabling the exploration of advanced machine learning techniques, such as transfer learning, and allowing for adaptation to new mining environments with fewer examples.

Overall, integrating additional real data, applying realistic noise simulations, and extensive testing across various mining sites represents critical steps in improving the efficacy and reliability of automatic arrival detection through deep learning methodologies. This continuous refinement can lead to better safety protocols, more efficient operations, and informed decision-making in microseismic monitoring.

7 BIBLIOGRAPHIE

Akram, J., & Eaton, D. W. (2016). A review and appraisal of arrival-time picking methods for downhole microseismic data. *Geophysics, 81*(2), KS71-KS91.

Akram, J., Ovcharenko, O., & Peter, D. (2017). *A robust neural network-based approach for microseismic event detection.* Paper presented at the SEG International Exposition and Annual Meeting.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America, 68*(5), 1521-1532.

Aminzadeh, F., Temizel, C., & Hajizadeh, Y. (2022). *Artificial Intelligence and Data Analytics for Energy Exploration and Production*: John Wiley & Sons.

Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the seismological society of America, 77*(4), 1437-1445.

Bai, C.-y., & Kennett, B. (2000). Automatic phase-detection and identification by full use of a single three-component broadband seismogram. *Bulletin of the seismological society of America, 90*(1), 187-198.

Barry, K., Cavers, D., & Kneale, C. (1975). Recommended standards for digital tape formats. *Geophysics, 40*(2), 344-352.

Beyreuther, M., Hammer, C., Wassermann, J., Ohrnberger, M., & Megies, T. (2012). Constructing a Hidden Markov Model based earthquake detector: application to induced seismicity. *Geophysical Journal International, 189*(1), 602-610.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4): Springer.

Bohlen, T. (2002). Parallel 3-D viscoelastic finite difference seismic modelling. *Computers & Geosciences, 28*(8), 887-899.

Bohlen, T., De Nil, D., Köhn, D., & Jetschny, S. (2012). SOFI3D-seismic modeling with finite differences 3D-acoustic and viscoelastic version: Users guide. *KIT, Karlsrue*.

Bohlen, T., & Saenger, E. H. (2006). Accuracy of heterogeneous staggered-grid finite-difference modeling of Rayleigh waves. *Geophysics, 71*(4), T109-T115.

Brown, L. G. (2015). *Seismic hazard evaluation using apparent stress ratio for mining-induced seismic events.*

Carcione, J. M. (1993). Seismic modeling in viscoelastic media. *Geophysics, 58*(1), 110-120.

Červený, V., Klimeš, L., & Pšenčík, I. (2007). Seismic ray method: Recent developments. *Advances in geophysics, 48*, 1-126.

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., . . . Team, E. C. (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical research letters, 47*(16), e2020GL088651.

Chapman, C. (2004). *Fundamentals of seismic wave propagation*: Cambridge university press.

Chen, Y. (2020). Automatic microseismic event picking via unsupervised machine learning. *Geophysical Journal International, 222*(3), 1750-1764.

Chollet, F. (2021). *Deep learning with Python*: Simon and Schuster.

Chouet, B. A. (1996). Long-period volcano seismicity: its source and use in eruption forecasting. *nature, 380*(6572), 309-316.

Collino, F., & Tsogka, C. (2001). Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. *Geophysics, 66*(1), 294-307.

Collins, D., Toya, Y., Pinnock, I., Shumila, V., & Hosseini, Z. (2014). *3D velocity model with complex geology and voids for microseismic location and mechanism.* Paper presented at the Deep Mining 2014: Proceedings of the Seventh International Conference on Deep and High Stress Mining.

Cook, N. G. W., Hoek, E., Pretorius, J., Ortlepp, W., & Salamon, M. (1966). Rock mechanics applied to study of rockbursts. *Journal of the Southern African Institute of Mining and Metallurgy, 66*(10), 436-+.

Coppens, F. (1985). First arrival picking on common-offset trace collections for automatic estimation of static corrections. *Geophysical Prospecting, 33*(8), 1212-1231.

Dai, H., & MacBeth, C. (1995). Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical Journal International, 120*(3), 758-774.

Dai, H., & MacBeth, C. (1997). The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings. *Journal of Geophysical Research: Solid Earth, 102*(B7), 15105-15113.

Dijkstra. (1959). E W. A note on two problems in connexion with graphs. *Numerische Mathematik, 1*(1), 269-271.

Dmowska, R., & Saltzman, B. (1999). *Advances in Geophysics: Long-range Persistence in Geophysical Time Series* (Vol. 40): Academic Press.

Dokht, R. M., Kao, H., Visser, R., & Smith, B. (2019). Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismological Research Letters, 90*(2A), 481-490.

Dreger, D., Tinti, E., & Cirella, A. (2007). Slip velocity function parameterization for broadband ground motion simulation. *Seismol. Res. Lett., 78*(2), 308.

Duncan, P. M. (2005). Is there a future for passive seismic? *First Break, 23*(6).

Earle, P. S., & Shearer, P. M. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the seismological society of America, 84*(2), 366-376.

Ellsworth, W. L. (2013). Injection-induced earthquakes. *science, 341*(6142), 1225942.

Evans, D. J., & Chadwick, R. (2009). *Underground gas storage: Worldwide experiences and future development in the UK and Europe*.

Fabien-Ouellet, G., Gloaguen, E., & Giroux, B. (2017). Time-domain seismic modeling in viscoelastic media for full waveform inversion on heterogeneous computing platforms with OpenCL. *Computers & Geosciences, 100*, 142-155.

Feng, G.-L., Feng, X.-T., Chen, B.-r., Xiao, Y.-X., & Yu, Y. (2015). A microseismic method for dynamic warning of rockburst development processes in tunnels. *Rock Mechanics and Rock Engineering, 48*, 2061-2076.

Feng, X.-T., Liu, J., Chen, B., Xiao, Y., Feng, G., & Zhang, F. (2017). Monitoring, warning, and control of rockburst in deep metal mines. *Engineering, 3*(4), 538-545.

Gaci, S. (2013). The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Transactions on Geoscience and Remote Sensing, 52*(8), 4558-4563.

Gajewski, D., & Tessmer, E. (2005). Reverse modelling for seismic event characterization. *Geophysical Journal International, 163*(1), 276-284.

Gardner, G., Gardner, L., & Gregory, A. (1974). Formation velocity and density—The diagnostic basics for stratigraphic traps. *Geophysics, 39*(6), 770-780.

Gentili, S., & Michelini, A. (2006). Automatic picking of P and S phases using a neural tree. *Journal of Seismology, 10*, 39-63.

Gharti, H. N., Oye, V., Roth, M., & Kühn, D. (2010). Automated microearthquake location using envelope stacking and robust global optimization. *Geophysics, 75*(4), MA27-MA46.

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International, 165*(1), 149-166.

Gibowicz, S. J., & Kijko, A. (1994). *An introduction to mining seismology*: Elsevier.

Giroux, B. (2021). ttcrpy: A Python package for traveltime computation and raytracing. *SoftwareX, 16*, 100834.

Giroux, B., & Larouche, B. (2013). Task-parallel implementation of 3D shortest path raytracing for geophysical applications. *Computers & Geosciences, 54*, 130-141.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1): MIT press Cambridge.

Graves, R. W. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the seismological society of America, 86*(4), 1091-1106.

Gupta, A., Gupta, H. P., Biswas, B., & Dutta, T. (2020). Approaches and applications of early classification of time series: A review. *IEEE Transactions on Artificial Intelligence, 1*(1), 47-61.

Gupta, H. K. (2002). A review of recent studies of triggered earthquakes by artificial water reservoirs with special emphasis on earthquakes in Koyna, India. *Earth-Science Reviews, 58*(3-4), 279-310.

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the seismological society of America, 34*(4), 185-188.

Han, L., Wong, J., & Bancroft, J. (2009). Time picking and random noise reduction on microseismic data. *CREWES Research Report, 21*, 1-13.

Häring, M. O., Schanz, U., Ladner, F., & Dyer, B. C. (2008). Characterisation of the Basel 1 enhanced geothermal system. *Geothermics, 37*(5), 469-495.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.* Paper presented at the Proceedings of the IEEE international conference on computer vision.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science, 313*(5786), 504-507.

Holzer, T. L., & Galloway, D. L. (2005). Impacts of land subsidence caused by withdrawal of underground fluids in the United States.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Hudyma, M. (2004). Mining-induced seismicity in underground, mechanised, Hardrock mines-results of a World Wide Survey. In: Australian Centre for Geomechanics, The University of Western Australia.

Jalbout, A., & Simser, B. (2014). *Rock mechanics tools for mining in high stress ground conditions at Nickel Rim South Mine.* Paper presented at the Deep Mining 2014: Proceedings of the Seventh International Conference on Deep and High Stress Mining.

Jinqiang, W., Basnet, P., & Mahtab, S. (2021). Review of machine learning and deep learning application in mine microseismic event classification. *Mining of mineral deposits*.

Kaiser, P. K., & Cai, M. (2012). Design of rock support system under rockburst condition. *Journal of Rock Mechanics and Geotechnical Engineering, 4*(3), 215-227.

Kelly, K. R., Ward, R. W., Treitel, S., & Alford, R. M. (1976). Synthetic seismograms: A finite-difference approach. *Geophysics, 41*(1), 2-27.

Keneti, A., & Sainsbury, B.-A. (2018). Characterization of strain-burst rock fragments under a scanning electron microscope–an illustrative study. *Engineering Geology, 246*, 12-18.

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & Ge, S. (2014). Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. *science, 345*(6195), 448-451.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical systems and signal processing, 151*, 107398.

Kisslinger, C. (1976). A review of theories of mechanisms of induced seismicity. *Engineering Geology, 10*(2-4), 85-98.

Knopoff, L., & Randall, M. J. (1970). The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research, 75*(26), 4957-4963.

Ko, Y.-D., & Park, C.-S. (2021). Parameter estimation of unknown properties using transfer learning from virtual to existing buildings. *Journal of Building Performance Simulation, 14*(5), 503-514.

Komatitsch, D., & Martin, R. (2007). An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics, 72*(5), SM155-SM167.

Komatitsch, D., & Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International, 139*(3), 806-822.

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters, 90*(1), 3-14.

Lama, R., & Bodziony, J. (1998). Management of outburst in underground coal mines. *International Journal of Coal Geology, 35*(1-4), 83-115.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324.

Lehmann, F., Gatti, F., Bertin, M., & Clouteau, D. (2024). Synthetic ground motions in heterogeneous geologies from various sources: the HEMEW S-3D database. *Earth System Science Data, 16*(9), 3949-3972.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., & Aila, T. (2018). Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*.

Leonard, M. (2000). Comparison of manual and automatic onset time picking. *Bulletin of the seismological society of America, 90*(6), 1384-1390.

Leonard, M., & Kennett, B. (1999). Multi-component autoregressive techniques for the analysis of seismograms. *Physics of the earth and planetary interiors, 113*(1-4), 247-263.

Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. *Advances in neural information processing systems, 31*.

Li, W., Chakraborty, M., Fenner, D., Faber, J., Zhou, K., Rümpker, G., . . . Srivastava, N. (2022). EPick: Attention-based multi-scale UNet for earthquake detection and seismic phase picking. *Frontiers in Earth Science, 10*, 953007.

Liang, C., Yu, Y., Yang, Y., Kang, L., Yin, C., & Wu, F. (2016). Joint inversion of source location and focal mechanism of microseismicity. *Geophysics, 81*(2), KS41-KS49.

Lightfoot, P. C. (2016). *Nickel sulfide ores and impact melts: Origin of the Sudbury Igneous Complex*: Elsevier.

Lin, B., Wei, X., & Junjie, Z. (2019). Automatic recognition and classification of multi-channel microseismic waveform based on DCNN and SVM. *Computers & Geosciences, 123*, 111-120.

Linzer, L., Mhamdi, L., & Schumacher, T. (2015). Application of a moment tensor inversion code developed for mining-induced seismicity to fracture monitoring of civil engineering materials. *Journal of Applied Geophysics, 112*, 256-267.

Lomax, A., Bagagli, M., Gaviano, S., Cianetti, S., Jozinović, D., Michelini, A., . . . Giunchi, C. (2024). Effects on a Deep-Learning, Seismic Arrival-Time Picker of Domain-Knowledge Based Preprocessing of Input Seismograms. *Seismica, 3*(1).

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Ma, J., Dineva, S., Cesca, S., & Heimann, S. (2018). Moment tensor inversion with three-dimensional sensor configuration of mining induced seismicity (Kiruna mine, Sweden). *Geophysical Journal International, 213*(3), 2147-2160.

Ma, T., Tang, C., Tang, L., Zhang, W., & Wang, L. (2015). Rockburst characteristics and microseismic monitoring of deep-buried tunnels for Jinping II Hydropower Station. *Tunnelling and Underground Space Technology, 49*, 345-368.

Maeda, N. (1985). A method for reading and checking phase times in autoprocessing system of seismic wave data. *Zisin, 38*, 365-379.

Maity, D., Aminzadeh, F., & Karrenbach, M. (2014). Novel hybrid artificial neural network based autopicking workflow for passive seismic data. *Geophysical Prospecting, 62*(4-Vertical Seismic Profiling and Microseismicity Frontiers), 834-847.

Majer, E. L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B., & Asanuma, H. (2007). Induced seismicity associated with enhanced geothermal systems. *Geothermics, 36*(3), 185-222.

Mavko, G., Mukerji, T., & Dvorkin, J. (2020). *The rock physics handbook*: Cambridge university press.

Maxwell, S. C., Rutledge, J., Jones, R., & Fehler, M. (2010). Petroleum reservoir characterization using downhole microseismic monitoring. *Geophysics, 75*(5), 75A129-175A137.

McDonald, A. J. (1982). *Seismicity of the Witwatersrand basin.* University of the Witwatersrand,

McEvilly, T., & Majer, E. (1982). ASP: An automated seismic processor for microearthquake networks. *Bulletin of the seismological society of America, 72*(1), 303-325.

McLean, S., Straub, K., & Stevens, K. (2005). The discovery and characterization of the Nickel Rim South deposit, Sudbury, Ontario.

Meier, M. A., Ross, Z. E., Ramachandran, A., Balakrishna, A., Nair, S., Kundzicz, P., . . . Yue, Y. (2019). Reliable real-time seismic signal/noise discrimination with machine learning. *Journal of Geophysical Research: Solid Earth, 124*(1), 788-800.

Mendecki, A. J. (1996). *Seismic monitoring in mines*: Springer Science & Business Media.

Minson, S. E., & Dreger, D. S. (2008). Stable inversions for complete moment tensors. *Geophysical Journal International, 174*(2), 585-592.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications, 11*(1), 3952.

Mousset, E., Cansi, Y., Crusem, R., & Souchet, Y. (1996). A connectionist approach for automatic labeling of regional seismic phases using a single vertical component seismogram. *Geophysical research letters, 23*(6), 681-684.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*: MIT press.

Nicol, A., Carne, R., Gerstenberger, M., & Christophersen, A. (2011). Induced seismicity and its implications for CO2 storage risk. *Energy Procedia, 4*, 3699-3706.

Noh, H., Hong, S., & Han, B. (2015). *Learning deconvolution network for semantic segmentation.* Paper presented at the Proceedings of the IEEE international conference on computer vision.

Pacheco-Martínez, J., Hernandez-Marín, M., Burbey, T. J., González-Cervantes, N., Ortíz-Lozano, J. Á., Zermeño-De-Leon, M. E., & Solís-Pinto, A. (2013). Land subsidence and ground failure associated to groundwater exploitation in the Aguascalientes Valley, México. *Engineering Geology, 164*, 172-186.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering, 22*(10), 1345-1359.

Peng, P., He, Z., Wang, L., & Jiang, Y. (2020). Microseismic records classification using capsule network with limited training samples in underground mining. *Scientific reports, 10*(1), 13925.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances, 4*(2), e1700578.

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Proakis, J. G. (2001). *Digital signal processing: principles algorithms and applications*: Pearson Education India.

Qu, S., Verschuur, E., & Chen, Y. (2018). *Automatic microseismic-event detection via supervised machine learning.* Paper presented at the SEG International Exposition and Annual Meeting.

Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. (2020). *Designing network design spaces.* Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Reynen, A., Chambers, K., & Baturan, D. (2018). *Automatic event detection and location using feature weighted beamforming.* Paper presented at the SEG International Exposition and Annual Meeting.

Richards, P., & Aki, K. (2002). Quantitative seismology. *University Science Book*.

Richards, P. G., & Kim, W.-Y. (2007). Seismic signature. *Nature Physics, 3*(1), 4-6.

Roden, J. A., & Gedney, S. D. (2000). Convolution PML (CPML): An efficient FDTD implementation of the CFS–PML for arbitrary media. *Microwave and optical technology letters, 27*(5), 334-339.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation.* Paper presented at the Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.

Ross, Z. E., Ben-Zion, Y., White, M. C., & Vernon, F. L. (2016). Analysis of earthquake body wave spectra for potency and magnitude values: Implications for magnitude scaling relations. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, 207*(2), 1158-1164.

Ross, Z. E., Meier, M. A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth, 123*(6), 5120-5129.

Rubinstein, J. L., & Mahani, A. B. (2015). Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity. *Seismological Research Letters, 86*(4), 1060-1067.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature, 323*(6088), 533-536.

Saad, O. M., & Chen, Y. (2020). Automatic waveform-based source-location imaging using deep learning extracted microseismic signals. *Geophysics, 85*(6), KS171-KS183.

Sabbione, J. I., & Velis, D. (2010). Automatic first-breaks picking: New strategies and algorithms. *Geophysics, 75*(4), V67-V76.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in neural information processing systems, 30*.

Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning*: Springer Science & Business Media.

Saragiotis, C. D., Hadjileontiadis, L. J., & Panas, S. M. (2002). PAI-S/K: A robust automatic seismic P phase arrival identification scheme. *IEEE Transactions on Geoscience and Remote Sensing, 40*(6), 1395-1404.

Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., & Platt, J. (1999). Support vector method for novelty detection. *Advances in neural information processing systems, 12*.

Segall, P. (1989). Earthquakes triggered by fluid extraction. *Geology, 17*(10), 942-946.

Sharma, B., Kumar, A., & Murthy, V. (2010). Evaluation of seismic events detection algorithms. *Journal of the Geological Society of India, 75*, 533-538.

Shearer, P. M. (2019). *Introduction to seismology*: Cambridge university press.

Shi, P., Angus, D., Nowacki, A., Yuan, S., & Wang, Y. (2018). Microseismic full waveform modeling in anisotropic media with moment tensor implementation. *Surveys in Geophysics, 39*, 567-611.

Sibson, R. H. (1989). Earthquake faulting as a structural process. *Journal of structural geology, 11*(1-2), 1-14.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Simpson, D. W., Leith, W., & Scholz, C. (1988). Two types of reservoir-induced seismicity. *Bulletin of the seismological society of America, 78*(6), 2025-2040.

Simser, B., & Butler, T. (2022). The Value of Recording Small Mining Induced Microseismic Events with Examples from Glencore's Nickel Rim South Mine. *RaSim10 Rockbursts and Seismicity in Mines, Tucson, Arizona*.

Sleeman, R., & Van Eck, T. (1999). Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the earth and planetary interiors, 113*(1-4), 265-275.

Song, C., & Alkhalifah, T. (2020). *Identifying microseismic events in time-reversed source images using support vector machine.* Paper presented at the EAGE 2020 Annual Conference & Exhibition Online.

Stevenson, P. R. (1976). Microearthquakes at Flathead Lake, Montana: A study using automatic earthquake processing. *Bulletin of the seismological society of America, 66*(1), 61-80.

Tan, Y., & He, C. (2016). Improved methods for detection and arrival picking of microseismic events with low signal-to-noise ratios. *Geophysics, 81*(2), KS93-KS111.

Taylor, S. R., Arrowsmith, S. J., & Anderson, D. N. (2010). Detection of short time transients from spectrograms using scan statistics. *Bulletin of the seismological society of America, 100*(5A), 1940-1951.

Trnkoczy, A. (2009). Understanding and parameter setting of STA/LTA trigger algorithm. In *New manual of seismological observatory practice (NMSOP)* (pp. 1-20): Deutsches GeoForschungsZentrum GFZ.

Trojanowski, J., & Eisner, L. (2017). Comparison of migration-based location and detection methods for microseismic events. *Geophysical Prospecting, 65*(1), 47-63.

Tromp, J., Komatitsch, D., & Liu, Q. (2008). Spectral-element and adjoint methods in seismology. *Communications in Computational Physics, 3*(1), 1-32.

Vaezi, Y., & Van der Baan, M. (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, 203*(3), 1896-1908.

van Thienen-Visser, K., & Breunese, J. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge, 34*(6), 664-671.

Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*, 273-297.

Vapnik, V. (1998). Statistical learning theory. *John Wiley & Sons google schola, 2*, 831-842.

Vidale, J. (1988). Finite-difference calculation of travel times. *Bulletin of the seismological society of America, 78*(6), 2062-2076.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics, 51*(4), 889-901.

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics, 74*(6), WCC1-WCC26.

Wang, K., Li, K., Zhou, L., Hu, Y., Cheng, Z., Liu, J., & Chen, C. (2019). Multiple convolutional neural networks for multivariate time series prediction. *Neurocomputing, 360*, 107-119.

Waqar, M. F., Guo, S., & Qi, S. (2023). A comprehensive review of mechanisms, predictive techniques, and control strategies of rockburst. *Applied Sciences, 13*(6), 3950.

Weiss, R. M., & Shragge, J. (2013). Solving 3D anisotropic elastic wave equations on parallel GPU devices. *Geophysics, 78*(2), F7-F15.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., & Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the seismological society of America, 88*(1), 95-106.

Wong, J., Manning, P. M., Han, L., & Bancroft, C. (2011). Synthetic microseismic datasets. *CSEG RECORDER, 36*, 31-38.

Wu, M., Ye, Y., Wang, Q., & Hu, N. (2022). Development of rockburst research: a comprehensive review. *Applied Sciences, 12*(3), 974.

Xiao, Y.-X., Feng, X.-T., Hudson, J. A., Chen, B.-R., Feng, G.-L., & Liu, J.-P. (2016). ISRM suggested method for in situ microseismic monitoring of the fracturing process in rock masses. *Rock Mechanics and Rock Engineering, 49*, 343-369.

Yadav, P. (2024). *Geomechanical evolution of the Nickel Rim South Mine.* Paper presented at the Deep Mining 2024: Proceedings of the 10th International Conference on Deep and High Stress Mining, Australian Centre for Geomechanics, Perth.

Yang, Z., Liu, C., Zhu, H., Xie, F., Dou, L., & Chen, J. (2019). Mechanism of rock burst caused by fracture of key strata during irregular working face mining and its prevention methods. *International Journal of Mining Science and Technology, 29*(6), 889-897.

Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Ambruz, N. B., . . . Earle, P. S. (2021). Leveraging deep learning in global 24/7 real-time earthquake monitoring at the National Earthquake Information Center. *Seismological Society of America, 92*(1), 469-480.

Zhang, H., Thurber, C., & Rowe, C. (2003). Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the seismological society of America, 93*(5), 1904-1912.

Zhang, S., Ma, T., Tang, C. a., Jia, P., & Wang, Y. (2020). Microseismic monitoring and experimental study on mechanism of delayed rockburst in deep-buried tunnels. *Rock Mechanics and Rock Engineering, 53*, 2771-2788.

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific reports, 10*(1), 1941.

Zhang, X., Zhao, Z., Jia, R., & Cao, L. (2020). Identification of Microseismic Signals Based on Multiscale Singular Spectrum Entropy. *Shock and Vibration, 2020*(1), 6717128.

Zhao, Z., & Gross, L. (2017). *Using supervised machine learning to distinguish microseismic from noise events.* Paper presented at the SEG international exposition and annual meeting.

Zheng, J., Lu, J., Peng, S., & Jiang, T. (2018). An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks. *Geophysical Journal International, 212*(2), 1389-1397.

Zhou, J., Zhang, Y., Li, C., He, H., & Li, X. (2024). Rockburst prediction and prevention in underground space excavation. *Underground space, 14*, 70-98.

Zhu, W., & Beroza, G. C. (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International, 216*(1), 261-273.

Zienkiewicz, O. C., & Taylor, R. L. (1997). *The finite element method* (Vol. 3): McGraw-Hill London.

Zoback, M. D., & Gorelick, S. M. (2012). Earthquake triggering and large-scale geologic storage of carbon dioxide. *Proceedings of the National Academy of Sciences, 109*(26), 10164-10168.

8 ANNEXE I

Chapter 2 of this study outlines the approach to full-waveform modeling for generating synthetic microseismic data. In this appendix, we present more details regarding the elastic modeling technique that we employed, which is based on the framework articulated in the work of Fabien-Ouellet et al. (2017) that was originally developed for viscoelastic media. This methodology provided a robust foundation for our analysis and simulation of the phenomena relevant to our research.

Full Waveform Inversion (FWI) is a method employed to determine the elastic properties of the Earth by correlating ground motion data with solutions derived from the wave equation. Nonetheless, simulating realistic scenarios can be highly computationally intensive, which limits the applicability of FWI. Recent developments in hardware have enhanced the capabilities of parallel computing, necessitating new programming approaches to utilize diverse architectures effectively. The mentioned study presents SeisCL, a portable program created with OpenCL designed for 2D and 3D viscoelastic FWI in the time domain, which can compute both forward and adjoint wavefields using finite-difference techniques. Performance evaluations on Intel CPUs, NVIDIA GPUs, and Intel Xeon PHI demonstrate that GPUs running OpenCL can enhance computation speeds by almost two orders of magnitude compared to single-threaded CPU implementations. While OpenCL offers code portability, optimizations specific to each device are crucial for achieving optimal performance. Below is an elaboration of their approach for full waveform modeling, including relevant equations:

Wave Equations

During the passage of a seismic wave through rock, an infinitesimally small cube within the rock contains atoms, ions, or molecules that move in an unspecified manner. By analyzing these microscopic particle motions, we can predict the displacements that lead to seismic wave propagation. The principle of conservation of momentum governs the total momentum $\boldsymbol{P}$ of the rock and can be altered in three distinct ways:

1. Not a bit

$$\frac{\partial \boldsymbol{P}}{\partial t} = \frac{\partial}{\partial t}\left\{\int_V \rho(x,t)v(x,t)dV\right\} = 0 \qquad (A.1)$$

where t represents time, x signifies the position vector, ρ indicates the density, and v refers to the particle velocity distributions within the rock.

2. By either adding or removing particles from the cube, this results in a gain or loss of momentum in the cube (Bohlen et al., 2012):

$$\Phi = \int_{\mathrm{A}} \rho \mathrm{v}\,(\mathrm{v} \cdot \mathrm{dA}) \qquad (A.2)$$

3. By the forces applied to the cube's surface, such as hydrostatic pressure P, or, more broadly, the stress σ_{ij} (surface forces) along with the body forces f_i acting on all particles within the volume.

   After performing a brief calculation, we arrive at the following system of first-order linear differential equations:

$$\rho \frac{dv_i}{dt} + v_i \rho \boldsymbol{\nabla} v_i = -\frac{\partial \sigma_{ij}}{\partial x_j} + f_i \qquad (A.3)$$

This equation holds true for any volume, regardless of whether it contains a liquid, gas, or rock. So far, there has been no suggestion regarding how the medium deforms when subjected to external stress. In the case of an elastic medium, there is a linear relationship between stress and strain:

$$\sigma_{ij} = \lambda \theta \delta_{ij} + 2\mu \epsilon_{ij} \qquad (A.4)$$

Where λ and μ are Lame parameters and ϵ_{ij} is the deformation tensor, θ is the cubic delatation and δ_{ij} is the Kronecker symbol, as:

$$\epsilon_{ij} = \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), i, j = 1,2,3$$

$$\theta = \epsilon_{11} + \epsilon_{22} + \epsilon_{33}$$

$$\delta_{ij} = \begin{Bmatrix} 1 & for \quad i = j \\ 0 & else \end{Bmatrix}$$

where σ is the stress tensor, ρ is the density of the medium, and u is the displacement vector.

For a viscoelastic medium, the stress-strain relationship is more complicated as (Bohlen et al., 2012):

$$\dot{\sigma}_{ij} = \frac{\partial v_k}{\partial x_k} \pi (1 + \tau^s) - 2\mu(1 + \tau^s) + 2 \frac{\partial v_i}{\partial x_j}(1 + \tau^s) + \sum_{l=1}^{L} r_{ijl} \ \boldsymbol{if} \ i = j \qquad (A.5)$$

$$\dot{\sigma}_{ij} = \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right)\mu(1+\tau^s) + \sum_{l=1}^{L} r_{ijl} \;\; \boldsymbol{if} \; i \neq j$$

the memory variables are as:

$$\dot{r}_{ijl} = -\frac{1}{\tau_{\sigma l}}\left\{(\pi\tau^p - 2\mu\tau^s)\frac{\partial v_k}{\partial x_k} + 2\frac{\partial v_i}{\partial x_j}\mu\tau^s + r_{ijl}\right\} \; if \; i = j$$

$$\dot{r}_{ijl} = -\frac{1}{\tau_{\sigma l}}\left\{\mu\tau^s\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) + r_{ijl}\right\} \; if \; i \neq j$$

τ^s and τ^p are related to the seismic quality factors Q_s and Q_p for the S-wave and P-wave velocities, respectively, through

$$\tau^s = 2/Q_s$$

$$\tau^p = 2/Q_p$$

Solving the Elastic Wave Equation with Finite Differences

To solve equations (A.3) and (A.4) for an elastic medium, or equations (A.3) and (A.5) for a viscoelastic medium, the particle velocities $\vec{v}$, the stresses σ_{ij}, and the Lame parameters λ and μ are determined at discrete Cartesian coordinates defined as $x = i \times dx$, $y = j \times dy$, $z = k \times dz$ and at specific time intervals $t = n \times dt$ on a grid. Here, dx, dy, and dz represent the spacing between grid points in the x, y, and z directions respectively, while dt indicates the difference between successive time steps. The indices i, j, k, and n run through the sets $i \in N|[1, NX]$, $j \in N|[1, NY]$, $k \in N|[1, NZ]$, and $n \in N|[1, NT]$, where NX, NY, NZ, and NT correspond to the total number of spatial grid points and time steps. Ultimately, the partial derivatives in the equations are substituted with finite-difference operators. The derivative of a function y with respect to a variable x can be approximated using either a forward operator D^+ or a backward operator D^-:

$$D_x^+ y = \frac{y[i+1] - y[i]}{dh} \qquad (A.6)$$

$$D_x^- y = \frac{y[i] - y[i-1]}{dh}$$

To compute with a greater distance between grid points, the variables are arranged on a staggered grid as in Figure 8-1 (Bohlen et al., 2012; Virieux, 1986). It is important to note that Y represents the vertical axis, and the indices for the stress components are labeled accordingly.

To ensure the stability of the Standard Staggered Grid (SSG) code, the density (ρ) and the Lame parameter (λ) are averaged arithmetically and harmonically (Bohlen, 2002; Bohlen et al., 2006). Additional information on the boundary conditions and the implementation of the SeisCL software can be accessed in the study by Fabien-Ouellet et al. (2017).

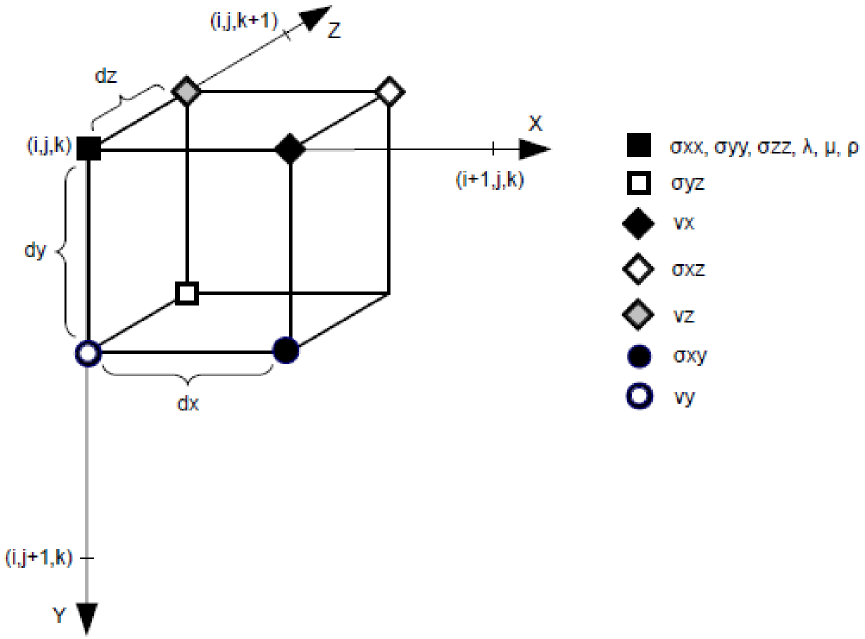
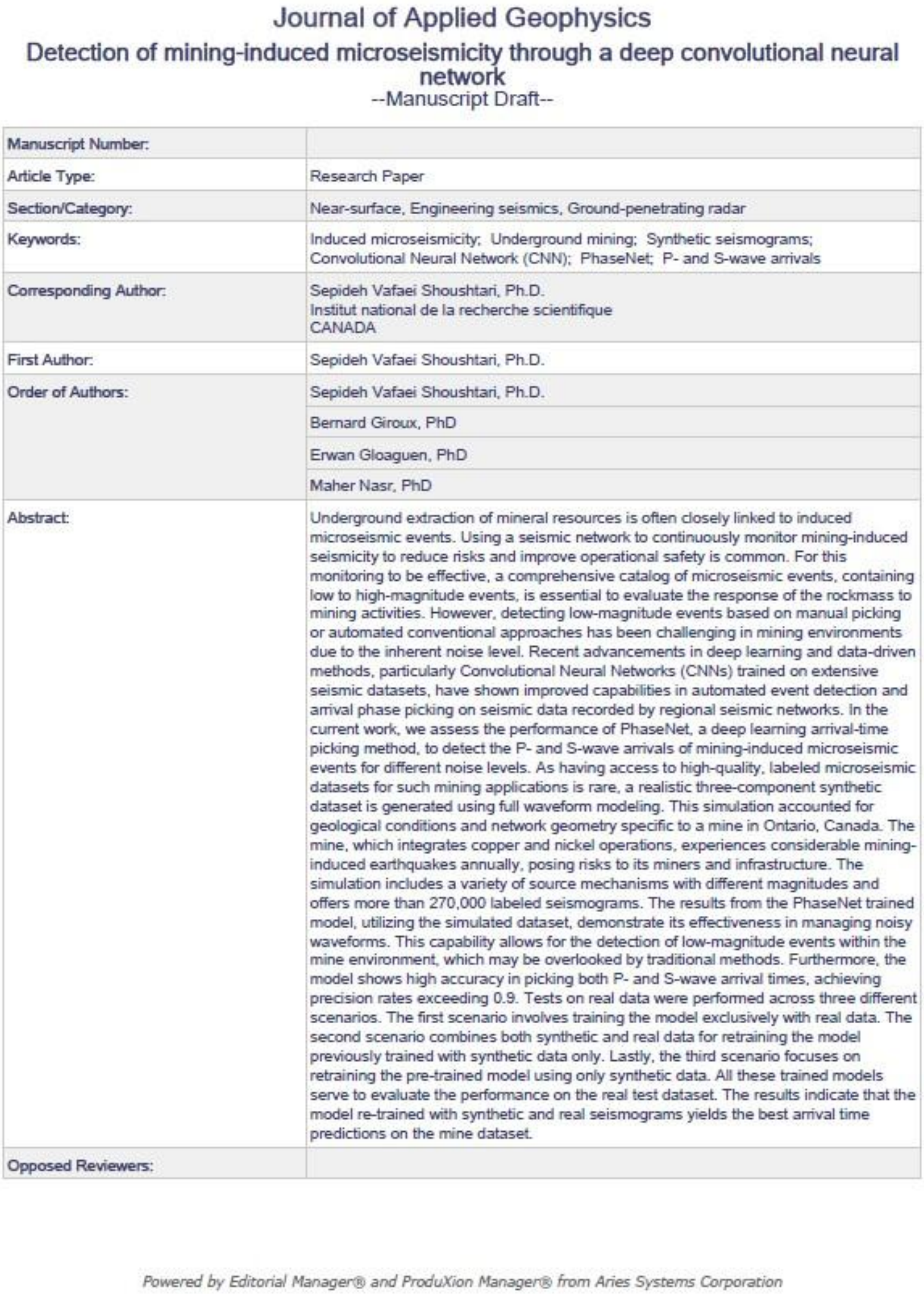


Figure 8-1: Geometry for a Standard Staggered Grid (Bohlen et al., 2012).

Conclusion

SeisCL demonstrates a comprehensive approach to modeling seismic wave propagation in viscoelastic media using time-domain numerical methods. The equations and computational strategies discussed make significant contributions to the field of geophysics, particularly in enhancing the resolution and accuracy of subsurface imaging through FWI techniques. Their innovative implementation using OpenCL further facilitates effective hardware utilization, thereby addressing the computational challenges associated with such complex modeling.

9 ANNEXE II



Journal of Applied Geophysics

Detection of mining-induced microseismicity through a deep convolutional neural network

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Section/Category:	Near-surface, Engineering seismics, Ground-penetrating radar
Keywords:	Induced microseismicity; Underground mining; Synthetic seismograms; Convolutional Neural Network (CNN); PhaseNet; P- and S-wave arrivals
Corresponding Author:	Sepideh Vafaei Shoushtari, Ph.D. Institut national de la recherche scientifique CANADA
First Author:	Sepideh Vafaei Shoushtari, Ph.D.
Order of Authors:	Sepideh Vafaei Shoushtari, Ph.D.
	Bernard Giroux, PhD
	Erwan Gloaguen, PhD
	Maher Nasr, PhD
Abstract:	Underground extraction of mineral resources is often closely linked to induced microseismic events. Using a seismic network to continuously monitor mining-induced seismicity to reduce risks and improve operational safety is common. For this monitoring to be effective, a comprehensive catalog of microseismic events, containing low to high-magnitude events, is essential to evaluate the response of the rockmass to mining activities. However, detecting low-magnitude events based on manual picking or automated conventional approaches has been challenging in mining environments due to the inherent noise level. Recent advancements in deep learning and data-driven methods, particularly Convolutional Neural Networks (CNNs) trained on extensive seismic datasets, have shown improved capabilities in automated event detection and arrival phase picking on seismic data recorded by regional seismic networks. In the current work, we assess the performance of PhaseNet, a deep learning arrival-time picking method, to detect the P- and S-wave arrivals of mining-induced microseismic events for different noise levels. As having access to high-quality, labeled microseismic datasets for such mining applications is rare, a realistic three-component synthetic dataset is generated using full waveform modeling. This simulation accounted for geological conditions and network geometry specific to a mine in Ontario, Canada. The mine, which integrates copper and nickel operations, experiences considerable mining-induced earthquakes annually, posing risks to its miners and infrastructure. The simulation includes a variety of source mechanisms with different magnitudes and offers more than 270,000 labeled seismograms. The results from the PhaseNet trained model, utilizing the simulated dataset, demonstrate its effectiveness in managing noisy waveforms. This capability allows for the detection of low-magnitude events within the mine environment, which may be overlooked by traditional methods. Furthermore, the model shows high accuracy in picking both P- and S-wave arrival times, achieving precision rates exceeding 0.9. Tests on real data were performed across three different scenarios. The first scenario involves training the model exclusively with real data. The second scenario combines both synthetic and real data for retraining the model previously trained with synthetic data only. Lastly, the third scenario focuses on retraining the pre-trained model using only synthetic data. All these trained models serve to evaluate the performance on the real test dataset. The results indicate that the model re-trained with synthetic and real seismograms yields the best arrival time predictions on the mine dataset.
Opposed Reviewers:	