

# Probabilistic Graphical Modelling of Seismic Data Processing in Mining

by

Cornel du Toit



*Dissertation presented for the degree of Doctor of  
Philosophy (Physics) in the Faculty of Science at  
Stellenbosch University*

Supervisor: Prof. H. C. Eggers

March 2020

# Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: .....

Copyright © 2020 Stellenbosch University  
All rights reserved.



# Abstract

Mining has long been characterised by deep shafts and dangerous conditions. Accurate monitoring and prediction of seismic activity and rockfalls are matters of life and death. The Institute of Mine Seismology (IMS) is the world's largest independent organisation that provides worldwide mine seismic data processing using human data processors. Approximately 35000 seismic events are processed per day by a team of 65 data processors (24 hours a day, 365 days a year) in order to provide rapid data assessments to the mine, typically within minutes of the event being recorded by the seismic network. This aim is achievable only with the assistance of automatic, computer-based, data processing.

While automatic processing is common in natural earthquake seismology, in mining-induced seismology the problem is more complex, and an automatic processor is yet to be developed. In mine seismology, classification of the recorded data is essential as there are many sources of noise in mines. Furthermore, with dense seismic sensor arrays in seismically active mines, multiple signals associated with both seismic events and noise sources may be conflated into a single seismogram. The matching of a Pressure (P)- and Shear (S)-wave for a specific seismic event in the presence of multiple sensors is not a simple task, even when analysed by an experienced seismologist.

In this dissertation, an automatic method based on probabilistic graphical models for both the event classification (seismic event, blast or rejected event) and the determination of the phase arrival times (P- and S-wave) is investigated. This machine learning approach has led to higher reliability, faster availability of results, more satisfied clients, less organisational load as well as a financial advantage to IMS and its clients.

By using Hidden Markov Models (HMM) as classification tool, different characteristics of the wave can be analysed for classification. By identifying the most likely hidden states (P-wave and S-wave) using the Viterbi algorithm combined with standard short-term average (STA) and long-term average (LTA) analysis, the candidate phase arrivals for each sensor are determined. The probability of each candidate phase arrival being the true arrival is seen as a parameter, expressed as a mixing weight, through the introduction of latent variables. The latent variables, together with the seismic event location parameters (3D multi-sensor and origin

time), are written as a probabilistic graphical model (PGM) which turns out to be a hierarchical Bayesian network. In most cases, the maximum a posterior (MAP) estimates of the latent variables are the true phase arrivals. In cases where the optimisation technique failed to deliver the MAP estimates e.g. got stuck in local maxima, outlier detection techniques are used to identify spurious events.

Of a total of 80 mines, the automatic processor which forms the subject of this dissertation is currently being tested on the 25 most seismically active ones. Of an average 35000 daily events (based on all 80 mines), 60% can be successfully processed. The average quality control score of the automatic processor is slightly higher than the average human quality score at a fraction of cost.

# Opsomming

Mynbou word lank reeds gekenmerk deur diep skagte en gevaarlike toestande. Akkurate monitering en voorspelling van seismiese aktiwiteit en rotsbarste kan gesien word as sake van lewe en dood. Ter verduideliking: Hier word die term “seismic event” eenvoudig vertaal as “skudding”. Die Instituut van Myn Seismologie (IMS) is die wêreld se grootste onafhanklike organisasie wat wêreldwye myn-verwante seismiese dataverwerking aanbied as een van sy dienste deur gebruik te maak van dataprocesseerders. Ongeveer 35000 skuddings word daaglik verwerk deur ’n span van 65 dataprocesseerders (24 uur per dag, 365 dae per jaar) met die doel om die verwerkte data in die myn se databasis op te dateer, minute nadat die skudding plaasgevind het. Hierdie doel kan net bereik word met die hulp van ’n outomatiese, rekenaargebaseerde dataverwerker.

Terwyl outomatiese verwerkers in natuurlike aardbewing seismologie reeds bestaan, is die probleem in mynbou-geïnduseerde seismisiteit meer kompleks en daar is nog nie ’n outomatiese verwerker ontwikkel nie. Klassifikasie van data is noodsaaklik in myn seismologie aangesien daar baie bronne van geraas is. Verder, in die geval van aktiewe myne met digte seismiese netwerke, kan daar dikwels meervoudige seine, geassosieer met beide skuddings en geraasbronne, in ’n enkele seismogram aangeteken word. Dit is dan selfs vir ervare seismoloë moeilik om te bepaal watter druk (P)- en skuif (S)-golwe by mekaar pas wanneer dit gegengereer is deur ’n spesifieke skudding en aangeteken word by meervoudige seismiese sensore.

’n Outomatiese metode, gebaseer op probabilistiese grafiese modelle vir beide die klassifikasie (skudding, ontploffing of verwerpte rekord) en die bepaling van die fase-aankomstye (P- en S-golf) word ondersoek in hierdie proefskrif. Hierdie masjienleerbenadering het gelei tot hoër betroubaarheid, vinniger beskikbaarheid van resultate, meer tevrede kliënte, minder organisatoriese betrokkenheid, sowel as finansiële voordele vir IMS en sy kliënte.

“Hidden Markov Models” (HMM) word gebruik as klassifikasie instrument en verskillende kenmerke van die golf word ontleed. Die kandidaat fase-aankomstye van die P- en S- golwe vir sensore word bepaal deur: a) identifisering van die mees moontlike verborge toestande deur gebruik van die Viterbi-algoritme en b) kombinasie met die standaard korttermyn gemiddelde (STA) en langtermyn (LTA) analise. Die waarskynlikheid dat elke kandidaat fase aankoms die ware aankoms is,

word as parameters gesien, uitgedruk as 'n menggewig deur die gebruik van latente veranderlikes. Die latente veranderlikes tesame met die liggingparameters (3D-middelpunt en oorsprongstyd), word geskryf as 'n probabilistiese grafiese model (PGM) en blyk dan om 'n hiërargiese Bayesiese netwerk te wees. In die meeste gevalle is die maksimum a posteriori (MAP) skattings van die latente veranderlikes die ware fase aankomste. In gevalle waar die optimaliseringstegniek misluk het om die MAP-ramings te lewer, d.w.s. waar dit vasval in plaaslike maksima, word uitskieter-opsporingstegnieke gebruik om verdagte skuddings te identifiseer.

Die outomatiese verwerker waaroor hierdie proefskrif handel word tans getoets op 25 van die mees seismies-aktiewe myne. 60% van 'n gemiddelde 35000 daaglikse aangetekende rekords (gebaseer op al 80 myne), word suksesvol verwerk deur die outomatiese prosesseerder. Die gemiddelde kwaliteitskontrole telling van die outomatiese verwerker is effens bokant die gemiddelde kwaliteit-telling van die menslike prosesseerders teen fraksie van koste.

# Acknowledgements

I would like to thank the following people for their support:

- My brother-in-law, Dr Ludwig Schwardt, for hours of discussion on Bayesian statistics and probabilistic graphical models, and for helping me format my dissertation. Your insight and generous time spent assisting me are greatly appreciated!
- My supervisor at the University of Stellenbosch, Prof. Hans Eggers, for providing me with the opportunity to do my dissertation despite having a full programme! Without your help, my dissertation would not have materialised.
- My parents, who spent hours editing and proofreading my dissertation. I love you!
- My boss and chairman of IMS, Dr Alex Mendecki, for affording me the opportunity to do this dissertation.
- Lastly, my wife, Dr Lauren Mongie, who by humorously poking fun at me spurred me on to achieve this goal. Hopefully, now you won't be the only one in the family with a PhD!



# Contents

|   |          |
|---|----------|
| Declaration   | i        |
| Abstract  | iii      |
| Opsomming   | v        |
| Acknowledgements                                      | vii      |
| Contents  | viii     |
| List of Figures                                       | x        |
| List of Tables  | xii      |
| List of Acronyms                                      | xiii     |
| List of Symbols                                       | xiv      |
| <b>1 Introduction</b>                                 | <b>1</b> |
| 1.1 Overview . . . . .                                | 1        |
| 1.2 Contributions of This Work . . . . .              | 5        |
| 1.3 Organisation of Dissertation . . . . .            | 6        |
| <b>2 Background</b>                                   | <b>9</b> |
| 2.1 Introduction . . . . .                            | 9        |
| 2.2 Seismic Source Parameters . . . . .               | 11       |
| 2.3 Physics of Seismic Waves . . . . .                | 12       |
| 2.4 Radiation Pattern . . . . .                       | 14       |
| 2.5 Triggering and Association . . . . .              | 15       |
| 2.6 Recorded Data at Mines . . . . .                  | 19       |
| 2.7 Location of Seismic Events . . . . .              | 23       |
| 2.8 Data Processing Carried Out by IMS . . . . .      | 25       |
| 2.9 Importance of Fast Automatic Processors . . . . . | 27       |

|   |            |
|---|------------|
| 2.10 Summary . . . . .  | 30         |
| <b>3 Automatic Classification</b>                             | <b>31</b>  |
| 3.1 Introduction . . . . .                                    | 31         |
| 3.2 Blast Discriminator . . . . .                             | 32         |
| 3.3 Hidden Markov Models . . . . .                            | 35         |
| 3.4 Classification of Recorded Data . . . . .                 | 52         |
| 3.5 Summary . . . . .   | 60         |
| <b>4 PGM of Phase Arrivals</b>                                | <b>63</b>  |
| 4.1 Model construction . . . . .                              | 63         |
| 4.2 The Specific Bayesian Network . . . . .                   | 76         |
| 4.3 PGM Factorisation . . . . .                               | 78         |
| 4.4 Posterior Distributions via MCMC . . . . .                | 85         |
| 4.5 Maximum A Posteriori Estimation . . . . .                 | 88         |
| 4.6 Examples of the Likelihood of Sensors . . . . .           | 91         |
| 4.7 Examples and Necessity of Multi-Sensor Analysis . . . . . | 97         |
| 4.8 Examples of the Posterior Distributions . . . . .         | 107        |
| 4.9 Outlier Detection . . . . .                               | 113        |
| 4.10 Results: Human Versus Automatic Processing . . . . .     | 114        |
| 4.11 Chapter Summary . . . . .                                | 117        |
| <b>5 Conclusion</b>   | <b>119</b> |
| <b>List of References</b>                                     | <b>121</b> |

# List of Figures

|      |  |    |
|------|--|----|
| 1.1  | Example of a waveform with multiple signals. . . . .                         | 3  |
| 2.1  | Illustration of a seismic event. . . . .                                     | 10 |
| 2.2  | The log-probability map of a synthetic event. . . . .                        | 11 |
| 2.3  | Example of a seismogram containing a double event. . . . .                   | 14 |
| 2.4  | Radiation pattern of a double-couple event. . . . .                          | 15 |
| 2.5  | Typical seismic event recorded at a mine. . . . .                            | 20 |
| 2.6  | Typical production blast recorded at a mine . . . . .                        | 20 |
| 2.7  | Typical development blast recorded at a mine. . . . .                        | 21 |
| 2.8  | Typical noise event. . . . .   | 21 |
| 2.9  | Drilling . . . . .   | 22 |
| 2.10 | Mine noise induced by ore pass noise (rolling rock). . . . .                 | 22 |
| 2.11 | Seismogram of QC versus original picks. . . . .                              | 26 |
| 2.12 | Histogram of all events versus manual events. . . . .                        | 28 |
| 3.1  | Representation of an HMM as a PGM. . . . .                                   | 37 |
| 3.2  | Example of segments of a waveform. . . . .                                   | 40 |
| 3.3  | Example of a seismic event with $E_S/E_P$ ratio equal to 20. . . . .         | 43 |
| 3.4  | Histograms of the normalised energy given different states. . . . .          | 43 |
| 3.5  | STA/LTA variable as a function of time for a typical seismic event. . .      | 44 |
| 3.6  | Histograms of STA/LTA variable given different states. . . . .               | 45 |
| 3.7  | Example of an event with the maximum amplitude near the S-wave onset. .      | 46 |
| 3.8  | Histograms of distance to max. amplitude variable given different states. .  | 47 |
| 3.9  | Typical seismogram belonging to the reject model. . . . .                    | 48 |
| 3.10 | Number of segments with a large amplitude feature example. . . . .           | 48 |
| 3.11 | Histograms of the percentage of segments with large maximum amplitudes. .    | 49 |
| 3.12 | Example of picks of the standard and customised Viterbi algorithm. . .       | 50 |
| 3.13 | Further example of picks of the stand. and cust. Viterbi algorithm. . .      | 51 |
| 3.14 | Picks when the maximum S-P separation is ignored. . . . .                    | 52 |
| 3.15 | Example of recorded event that was correctly classified as an accept. . .    | 54 |
| 3.16 | Example I of a recorded event that was correctly classified as a reject. . . | 55 |
| 3.17 | Example II of an event that was correctly classified as a reject. . . . .    | 55 |

|      |  |     |
|------|--|-----|
| 3.18 | Example of an event that was barely correctly classified as an accept. . . | 56  |
| 3.19 | Example of an event that was barely correctly classified as a reject. . .  | 56  |
| 3.20 | Example I of an event that was incorrectly classified as an accept. . . .  | 57  |
| 3.21 | Example II of an event that was incorrectly classified as an accept. . .   | 57  |
| 3.22 | Example of an event that was incorrectly classified as a reject. . . . .   | 58  |
| 3.23 | Proposed HMM classification histogram. . . . .                             | 59  |
| 3.24 | Example of a seismic event with stronger P-wave than S-wave arrivals.      | 60  |
|      |  |     |
| 4.1  | Example I of the picks of a single seismogram picker. . . . .              | 64  |
| 4.2  | Example II of the picks of a single seismogram picker. . . . .             | 65  |
| 4.3  | Example III of the picks of a single seismogram picker. . . . .            | 65  |
| 4.4  | Schematic of multiple events in a buffer. . . . .                          | 66  |
| 4.5  | Connections. . . . .   | 70  |
| 4.6  | A probabilistic graphical model with latent variables. . . . .             | 77  |
| 4.7  | Example of an unclear P-wave arrival. . . . .                              | 80  |
| 4.8  | Example I of the P-wave likelihood function for a waveform. . . . .        | 93  |
| 4.9  | Example I of the likelihood function of the S-wave. . . . .                | 93  |
| 4.10 | Example II of the P-wave likelihood function for a waveform. . . . .       | 94  |
| 4.11 | Example II of the S-wave likelihood function for a waveform. . . . .       | 94  |
| 4.12 | Example III of the P-wave likelihood function for a waveform. . . . .      | 95  |
| 4.13 | Example III of the S-wave likelihood function for a waveform . . . . .     | 95  |
| 4.14 | Example IV of the P-wave likelihood function for a waveform. . . . .       | 96  |
| 4.15 | Example IV of the S-wave likelihood function for a waveform. . . . .       | 96  |
| 4.16 | Example I of the single- and multiple arrival time distributions. . . . .  | 98  |
| 4.17 | Example II of the single- and multiple arrival time distributions. . . . . | 99  |
| 4.18 | Example III of the single- and multiple arrival time distributions. . . .  | 100 |
| 4.19 | Example IV of the single- and multiple arrival time distributions. . . .   | 101 |
| 4.20 | Example V of the single- and multiple arrival time distributions. . . . .  | 102 |
| 4.21 | Example VI of the single- and multiple arrival time distributions. . . .   | 103 |
| 4.22 | Example VII of the single- and multiple arrival time distributions. . . .  | 104 |
| 4.23 | Example VIII of the single- and multiple arrival time distributions. . .   | 105 |
| 4.24 | Example IX of the single- and multiple arrival time distributions. . . .   | 106 |
| 4.25 | Example X of the single- and multiple arrival time distributions. . . . .  | 107 |
| 4.26 | Example I of posterior distribution of relevant parameters. . . . .        | 109 |
| 4.27 | Example II of posterior distribution of relevant parameters. . . . .       | 110 |
| 4.28 | Example III of posterior distribution of relevant parameters. . . . .      | 111 |
| 4.29 | Example IV of posterior distribution of relevant parameters. . . . .       | 112 |
| 4.30 | Example of the inaccurate picking mistakes and QC score. . . . .           | 115 |

# List of Tables

|     |  |     |
|-----|--|-----|
| 1.1 | Characteristics of a general data processing system. . . . .             | 2   |
| 2.1 | Table showing the P-wave velocity (in m/s) for different rock types. . . | 13  |
| 2.2 | Table showing the S-wave velocity (in m/s) for different rock types. . . | 13  |
| 2.3 | Number of recorded data per hour outliers. . . . .                       | 29  |
| 4.1 | Table showing the plate notation. . . . .                                | 71  |
| 4.2 | Comparison of human and automatic processing at Mine1. . . . .           | 116 |
| 4.3 | Comparison of human and automatic processing at Mine2. . . . .           | 116 |

# List of Acronyms

|        |  |
|--------|--|
| AIC    | Akaike Information Criterion           |
| ANN    | Artificial Neural Network              |
| AR     | Auto-Regressive                        |
| ARMA   | Auto-Regressive Moving Average         |
| CNN    | Convolutional Neural Network           |
| DAG    | Directed Acyclic Graph                 |
| DC     | Direct Current                         |
| EM     | Expectation Maximisation               |
| HMM    | Hidden Markov Model                    |
| IMS    | Institute of Mine Seismology           |
| LTA    | Long-Term Average                      |
| M      | Moduli                                 |
| NEIC   | National Earthquake Information Center |
| MAP    | Maximum A Posteriori                   |
| MCMC   | Markov Chain Monte Carlo               |
| PGM    | Probabilistic Graphical Model          |
| P-wave | Pressure Wave                          |
| QC     | Quality Control                        |
| QCQ    | Quality Control Query                  |
| S-wave | Shear Wave                             |
| SNR    | Signal To Noise                        |
| STA    | Short-Term Average                     |
| TTE    | Travel Time Equation                   |

# List of Symbols

|                              |  |
|------------------------------|--|
| $\alpha$                     | fixed hyper-parameters of Dirichlet distribution                                 |
| $d_i$                        | distance between hypocentre and sensor $i$                                       |
| $E$                          | seismic energy   |
| $E_P$                        | seismic energy of P-wave   |
| $E_S$                        | seismic energy of S-wave   |
| $\varepsilon^P$              | travel time modelling error of the P-wave  |
| $\varepsilon^S$              | travel time modelling error of the S-wave  |
| $f_0$                        | corner frequency   |
| $K$                          | bulk modulus or number of STA/LTA time windows                                   |
| $k_i$                        | total number of possible arrival time pairs of sensor $i$                        |
| $n$                          | number of seismic sensors  |
| $n_j^{LTA}$                  | number of samples of LTA of $j$ th time window                                   |
| $n_j^{STA}$                  | number of samples of STA of $j$ th time window                                   |
| $n_i^P$                      | number of possible P-wave arrivals of sensor $i$                                 |
| $n_i^S$                      | number of possible S-wave arrivals of sensor $i$                                 |
| $n_i^{SP}$                   | number of possible P- and S-wave arrival time pairs of sensor $i$                |
| $M$                          | seismic moment   |
| $m_i$                        | number of observations of sensor $i$   |
| $\mu$                        | rigidity or shear modulus  |
| $\omega_{ij}^{P\theta}$      | Laplace distribution of P-wave for $j$ th arrival of sensor $i$ at $\theta$      |
| $\omega_{ij}^{S\theta}$      | Laplace distribution of S-wave for $j$ th arrival of sensor $i$ at $\theta$      |
| $\omega_{ij}^{\theta\Delta}$ | Laplace likelihood with time offset for $j$ th arrival of sensor $i$ at $\theta$ |
| $P$                          | seismic potency  |
| $\psi_i^{\text{high}}$       | highest probability of segments of sensor $i$ used in classification             |
| $\psi_i^{\text{low}}$        | lowest probability of segments of sensor $i$ used in classification              |
| $R_j^P(t_i)$                 | STA/LTA ratio of $j$ th time window for P-wave at time $t_i$                     |
| $R_j^S(t_i)$                 | STA/LTA ratio of $j$ th time window for S-wave at time $t_i$                     |
| $\rho$                       | density  |
| $\rho_i$                     | time off set parameter of sensor $i$   |

|                           |   |
|---------------------------|---|
| $\boldsymbol{\rho}$       | vector of time off set parameters of all sensors                        |
| $\phi_{ij}$               | transition probability of hidden state $i$ to $j$                       |
| $\Phi$                    | transition matrix   |
| $\varphi_{ij}^P$          | normalised STA/LTA of P-wave for $j$ th candidate arrival of sensor $i$ |
| $\varphi_{ij}^S$          | normalised STA/LTA of S-wave for $j$ th candidate arrival of sensor $i$ |
| $\boldsymbol{\pi}$        | mixing weights of all sensors or initial state probabilities            |
| $\boldsymbol{\pi}_i$      | mixing weights of sensor $i$  |
| $\pi_{ij}$                | mixing weight of $j$ th candidate arrival of sensor $i$                 |
| $s$                       | number of hidden states   |
| $t_0$                     | origin time of seismic event  |
| $\mathbf{t}^{\text{raw}}$ | all sample times of all sensors   |
| $\mathbf{t}$              | possible arrival time pairs of all sensors                              |
| $\boldsymbol{\theta}$     | hypocentre and origin time or general parameter                         |
| $\mathbf{t}_i$            | vector of the possible arrival time pairs of sensor $i$                 |
| $\mathbf{t}_{ij}$         | the $j$ th candidate arrival time pair of sensor $i$                    |
| $t_i^P$                   | arrival time of the P-wave at sensor $i$                                |
| $t_{ij}^P$                | P-wave arrival of pair $j$ of sensor $i$                                |
| $t_i^S$                   | arrival time of the S-wave at sensor $i$                                |
| $t_{ij}^S$                | S-wave arrival of pair $j$ of sensor $i$                                |
| $t_i^{\text{off}}$        | the time offset of sensor $i$   |
| $u_i$                     | uniform distribution indicating false alarm of sensor $i$               |
| $v^P$                     | P-wave velocity   |
| $v^S$                     | S-wave velocity   |
| $x, y, z$                 | hypocentre of seismic event   |
| $\mathbf{x}$              | data vector or observations   |
| $\mathbf{y}$              | STA/LTA values of all sensors   |
| $\mathbf{y}^P$            | STA/LTA values of P-wave of all sensors                                 |
| $\mathbf{y}^S$            | STA/LTA values of S-wave of all sensors                                 |
| $\mathbf{y}_i^P$          | STA/LTA values of P-wave of sensor $i$                                  |
| $\mathbf{y}_i^S$          | STA/LTA values of S-wave of sensor $i$                                  |
| $\mathbf{y}_{ij}^P$       | STA/LTA value of P-wave of $j$ th possible arrival of sensor $i$        |
| $\mathbf{y}_{ij}^S$       | STA/LTA value of S-wave of $j$ th possible arrival of sensor $i$        |
| $z_i$                     | categorical variable/latent variable of sensor $i$                      |
| $z_t$                     | hidden state at time $t$  |
| $\mathbf{z}$              | hidden states vector of latent variable vector of all sensors           |





# Chapter 1

## Introduction

In this chapter, a brief introduction to the context and the topics is provided which will be covered in more detail in subsequent chapters. Technical terms, which will be defined in detail later, can for the moment be found in the List of Acronyms.

### 1.1 Overview

Seismic monitoring in mines has five components: seismic array design, data acquisition, signal processing and classification of the recorded data, analysis of seismic events and finally quantification and interpretation of the seismic response to the rock-mass in mining activity. The ultimate goal of such monitoring is ultimately to predict, and possibly prevent, large seismic events and to evacuate mining areas for safety reasons. Mendecki [1] provides a comprehensive guide to each step.

By establishing a reliable location of current seismic events in high stressed areas as they occur; better-informed decisions can be made timeously, such as which areas to avoid and where to mine. Up till now, human operators have processed data. Human processing is slow and often inadequate, especially for so-called data bursts where multiple events occur within a short time-span. These data bursts are strongly correlated with hazardous conditions. To illustrate: large mines with dense seismic arrays averaging 5000 recorded events per day can peak to 20000. Should there be simultaneous bursts at different mines, there might be days worths of data backlog, which can have life-threatening consequences.

Table 1.1 illustrates the characteristics of a general data processing system conforming to high standards. The importance of integrity and speed is indisputable. While staying within its financial constraints, IMS needs to employ as many human processors as possible to achieve this. Employing full-time staff is expensive because of overtime in a 24/7 service environment. For this reason the data processing service is contracted out globally. These processors are paid on a per event basis

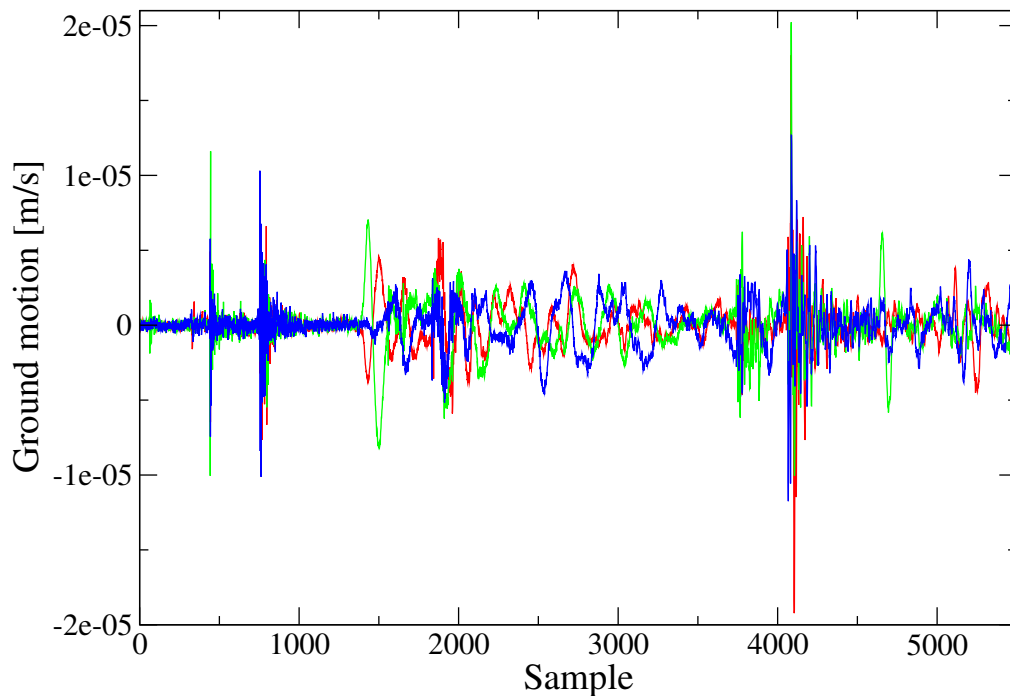
to speed up processing, but in processing as fast as possible, quality is compromised. For the reasons mentioned above, the use of a fast computer-based processor (automatic processor) is essential.

**Table 1.1:** Characteristics of a general data processing system.

|                             | Human                  | Auto                        |
|-----------------------------|------------------------|-----------------------------|
| Integrity of processing     | 60%                    | 95%                         |
| Burst handling/speed        | too slow               | fast enough                 |
| Location accuracy           | satisfactory           | satisfactory                |
| Location error accuracy     | adequate<br>(possible) | good<br>(readily available) |
| Reproducibility             | poor                   | excellent                   |
| Future improvement/learning | limited                | good                        |
| Management of backlogs      | difficult              | easy                        |
| General management          | demanding              | easy                        |
| Labour Unrest               | possibility            | never                       |
| Financial cost              | expensive              | moderate running cost       |

At present, there is no automatic processor for mine data available on the market. A successful automatic data processor will perform on a par with a human processor concerning the classification and phase arrival determination of seismic events, but will come with the added benefit of an order-of-magnitude faster speed.

While there are similarities, mining seismology is distinctly different from the seismology of natural earthquakes. Phase arrival determination in natural earthquakes is an established field. In natural earthquake seismology, classification is not essential, as only earthquakes can generate waves triggering sensors, spatially separated over thousands of kilometers. In natural earthquake seismology, recorded seismic waves consist of both body waves and surface waves, which are comprised of Rayleigh waves and Love waves. These make the determination of the phase arrivals of natural earthquakes more complicated, as more types of waves are involved. In mine seismology, by contrast, recorded seismic waves consist of only body waves. However, a recorded mining-induced, high resolution seismic array seismogram frequently consists of multiple signals separated close in time and signals of different seismic events combined with signals generated by non-seismic sources; see Figure 1.1. In these types of scenarios, determining the signals for each sensor caused by the same event is a problem unique to high resolution arrays. In summary, the techniques and algorithms developed in natural earthquake seismology to determine phase arrivals automatically are applicable in mine seismology, but are not sufficient.



**Figure 1.1:** Example of a seismogram recorded by a tri-axial geophone with a sampling rate of 6000 Hz. The red, blue and green signals represent the three spatial directions of movement. Three seismic events are visible plus a signal most probably generated by ore pass noise (signal starting around sample 1500). These four signals all occurred within 0.8 seconds from one another. Determining which signals from each seismogram are causally related is not a trivial exercise.

Seismic body waves are either pressure waves or shear waves and are correspondingly termed P- and S-waves. Automatic P-wave phase arrival pickers, using an automatic phase arrival determination algorithm, are simpler than S-wave pickers. Allen [2, 3] describes one of the first automatic phase arrival detection algorithms based on an STA to LTA ratio; see Subsection 2.5.1 for the definition of STA/LTA ratio. Baer and Kradolfer [4] used an adaptive threshold in order to make the automatic picker more sensitive to amplitude changes. A more sophisticated statistical technique assumes that the seismic wave is an autoregressive (AR) model. The assumption is made that the seismogram can be divided into locally stationary segments and the intervals before and after the onset of the P-wave phase arrival can be seen as two different stationary AR processes. The Akaike Information Criterion (AIC) as described in Akaike [5] was used to detect P-wave onsets in Sleeman and van Eck [6]; Leonard and Kennett [7] and Leonard [8]. The AIC picker was combined with wavelet analysis to detect P-wave onsets in Zhang *et al.* [9].

There are many automatic S-wave phase arrival pickers in literature. Differences in the polarisation properties of the P- and S-wave were, for example, utilised by

Cichowicz [10]. A discrete wavelet transform was used by Geldron *et al.* [11] and Ebel *et al.* [12] to detect and classify seismic event signals. Wang and Teng [13] used a neural network approach based on several signal attributes to detect phase arrival onsets, while Akazawa [14] combined the AR-AIC algorithm and STA/LTA ratio in detecting both P- and S-wave arrivals. All the mentioned pickers associated with natural earthquake seismology have some characteristic function obtained from the waveforms with the highest (or lowest) value being the most likely phase arrival. These P-wave and S-wave pickers act only on a single seismogram and are referred to as single seismogram pickers.

Classification of recorded data in mine seismology has not been documented extensively to date. Many studies have been carried out classifying regional (seismic networks of up to few 1000 kilometers) seismic events and explosions/blasts. Fah and Koch [15] used multivariate statistical analysis to derive a general discriminant function, while Tiira [16] discriminated between nuclear explosions and earthquakes using artificial neural networks (ANN). AllamehZahed [17] used ARMA coefficients determination by ANN to discriminate between earthquakes and man-made events, and Kortstrom *et al.* [18] made use of an automatic classifier for regional seismic events, spurious events and nearby mine blasts, based on spectrograms and support vector machine algorithm. Giacco *et al.* [19] used support vector machine algorithm for automatic classification of seismic signals at a volcano.

Beyreuther and Wassermann [20] constructed a hidden semi-Markov model (HMM) as an algorithm to reduce false alarms for small seismic arrays. The authors classified signals from continuous data which introduces complexities not applicable to our triggered waveform classification problem - see Section 2.5 for definition of triggered data. Benitez [21] used an HMM with log energies as feature vector to classify, amongst others, long period seismic events, tremors and noise generated by a volcano in Antarctica. The training data set used by the authors was small and non-standard training techniques were applied; our data sets are large and these optimisation techniques are not needed. An HMM has the added advantage of being a single seismogram picker. It utilises the Viterbi algorithm to obtain the most likely hidden states.

In this dissertation, the classification of triggered waveforms are carried out by using a customised hidden Markov model, and the feature extraction proposed consists of four parameters. A first-order HMM with generic Viterbi decoding cannot directly model the differences between polarisation properties of the P-wave and S-wave arrival. An HMM with a customised Viterbi decoding algorithm was developed which can indirectly model the orthogonal directivity of particle motions of the P- and S-wave. In addition, other shortcomings of the HMM using the standard Viterbi decoding algorithm are also addressed. A classification function is proposed.

After recorded data has been classified as a seismic event, the phase arrivals of the P- and S-wave must be determined for each waveform recorded by the seismic

sensors. As pointed out, in mine seismology, the waveforms of the classified seismic event may contain signals from multiple events- and/or noise signals; thus it is important which signals from each seismogram are linked to the same source, or at least, the phase arrivals generated by the most energetic seismic event. First, sets of P- and S-wave candidate arrivals must be identified for each seismogram using the single seismogram picker(s). The value of the characteristic function tends to scale with the probability of the phase arrivals: high values of the characteristic function are positively correlated with the phase arrival of the most energetic seismic event. Then, the problem of determining the phase arrivals generated by the most energetic seismic event can be solved only by viewing it as multivariate dependent. In this dissertation, this is done in a multivariate Bayesian framework.

Previous work on multivariate Bayesian analysis in automatic processing has been carried out. Tarantola and Valette [22] suggested, in a Bayesian framework, that each segment of seismogram is assigned a probability of a phase arrival. The probability of these candidate arrivals, or segment, can be seen as a mixture weight, all of which sums to 1. Martinsson [23] used multiple candidate arrivals to do Bayesian inference through MCMC methods to obtain posterior 3D hypocentre distributions of blasts at a mine in Europe. Martinsson showed that the automatic event locations produced in his paper were more accurate than the seismic event locations processed by humans. However, using the characteristic function of a single seismogram picker to obtain estimates for the mixture weights is sub-optimal when accurate phase arrival estimates are needed. Note that Martinsson [23] was interested in only the 3D hypocentre distribution, so treating the mixture weights as known values was not critical.

In this dissertation, it will be shown that the mixture weights estimates of both P-wave and S-wave candidates, found from single seismogram pickers, can be improved significantly. The mixture weights should be modelled as random variables and not as fixed numbers. A probabilistic graphical model (PGM) is proposed to express the conditional dependence structure of candidate arrivals and the origin time and 3D hypocentre. This is the first time that the determination of the phase arrivals of seismic waves is modelled in a multivariate way by using a probabilistic graphical model. The specific PGM is a hierarchical Bayesian network with latent variables.

## 1.2 Contributions of This Work

The data, algorithms and methods set out in this dissertation rely on significant contributions made by others at IMS. This dissertation documents the following original contributions made by me. While they are mentioned throughout the dissertation, they are listed here for easy referencing.

1. The P-coda hidden states and S-coda hidden states of the HMM.
2. The specific form of the transition matrix of the HMM.
3. The feature extraction of the HMM. The features are: the normalised energy, the STA/LTA feature, the distance to maximum amplitude and the number of spikes.
4. A customised Viterbi decoding algorithm that indirectly models the orthogonal directivity of the particle motion of the P- and S-wave. In addition, other shortcomings of the standard Viterbi decoding algorithm are addressed.
5. A function that classifies recorded data as accepted or rejected.
6. A modified STA/LTA measurement determining the candidate phase arrivals of a sensor.
7. Expressing the conditional dependence structure of candidate arrivals and the origin time and 3D hypocentre as a PGM.
8. A method to find the starting values of the latent variables as well as the starting value of the origin time and hypocentre. The proposed starting value of the origin time and hypocentre has a large probability of being in the basin of attraction of the global maxima. These starting values are sufficient to ensure that the MAP estimates of the latent variable are in most cases the correct arrival time estimates.
9. Incorporating the STA/LTA and HMM information into Bayesian network.
10. Factorising the conditional distributions to solve the statistical inference of the specific Bayesian network.

To the best of my knowledge, I am the originator of the contributions mentioned above.

### 1.3 Organisation of Dissertation

The main body of this dissertation is summarised as follows. In Chapter 2 an introduction to relevant mine seismology topics is given. These include seismic source parameters, the physics of seismic waves, triggering and association of data, examples of recorded waveforms of seismic events, blasts and noise, the mathematics involved in locating seismic events, data processing and the importance of automatic processing.

Chapter 3 discusses in detail automatic classification of data recorded at mines. Because it captures important predictable structures of the wave, I show why modelling the recorded waves as an HMM is preferred. The Viterbi decoding algorithm falls short with respect to a few important features; hence a customised Viterbi algorithm is proposed.

Chapter 4 forms the heart of the dissertation where automatic phase arrival determination for seismic events is discussed. A PGM is proposed to represent the interplay of seismic waves recorded at different sensors generated by an unknown source. The PGM is a Bayesian network and it is shown how to factorise the joint distribution and doing inference of the phase arrivals and the origin time and 3D hypocentre parameters. A comparison (classification and accuracy of phase arrivals) between human processing and the automatic processor is also shown.

Chapter 5 concludes.





# Chapter 2

## Background

### 2.1 Introduction

In earthquake seismology the tectonic motions are slow (few centimeters per year), but continuous, whereas mining usually proceeds with discrete blasts and irregular timing. The magnitude distribution of earthquakes or seismic events follows a power law. This means that there are significantly more smaller seismic events than larger seismic events. The source size of the largest events that can be induced by mining is of the order of hundreds of meters, while for natural large earthquakes hundreds of kilometers. The potentially damaging event that can be induced by mining ranges in Richter Magnitude from 1.0 to 4.0, while destructive natural earthquakes range in magnitude from 6.0 to 9.0 on the Richter Scale.

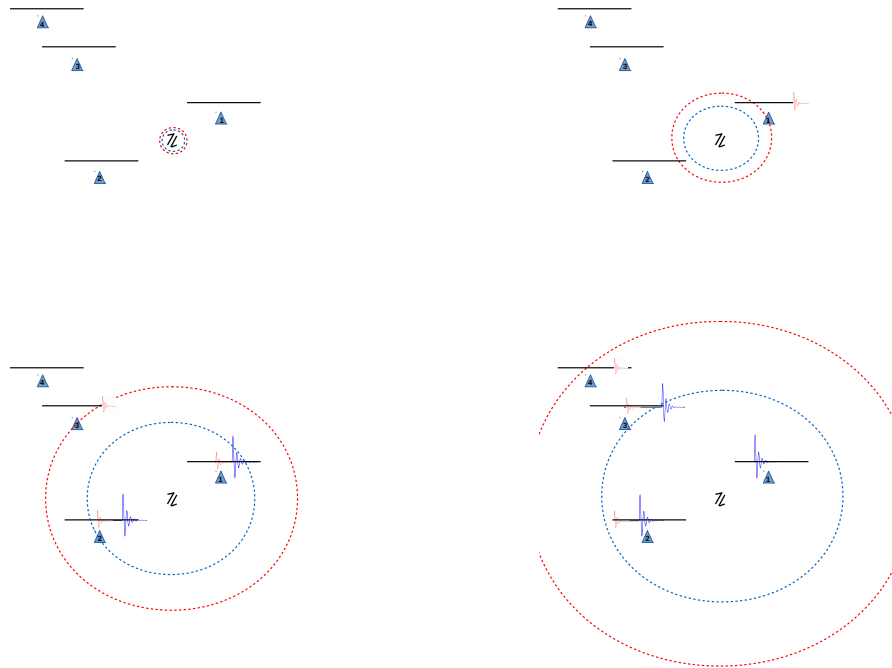
Seismic monitoring started during the middle of the 1970s. For mining, work-horse sensors are miniature geophones which are sensitive above their natural frequencies, 4.5 Hz where possible and 14 Hz or 28 Hz where the installation conditions are challenging. These allow reasonable coverage up to magnitude 3. Piezoelectric accelerometers can extend the magnitude range down to -3. All are best installed underground to avoid wave filtering and distortion in the surface layers.

For regional networks, which have diameters up to a few hundred kilometres, there is some overlap with 1 Hz geophones being the sensors of choice. Earthquake seismologists refer to these as short-period sensors, which are effective up to about magnitude 5.

There is a lot of noise with periods of around 6 seconds, usually attributed to ocean waves, so lower frequencies were historically monitored with long-period (> 30 s) instruments. In the last 50 years broadband sensors have gradually taken over, covering the range from 300 s to 0.05 s. Their response is flat to ground velocity at low frequencies, usually changing over to acceleration near the high end. These work for global seismology and detecting nuclear explosions.

Another common application of seismic sensors is strong ground motion meas-

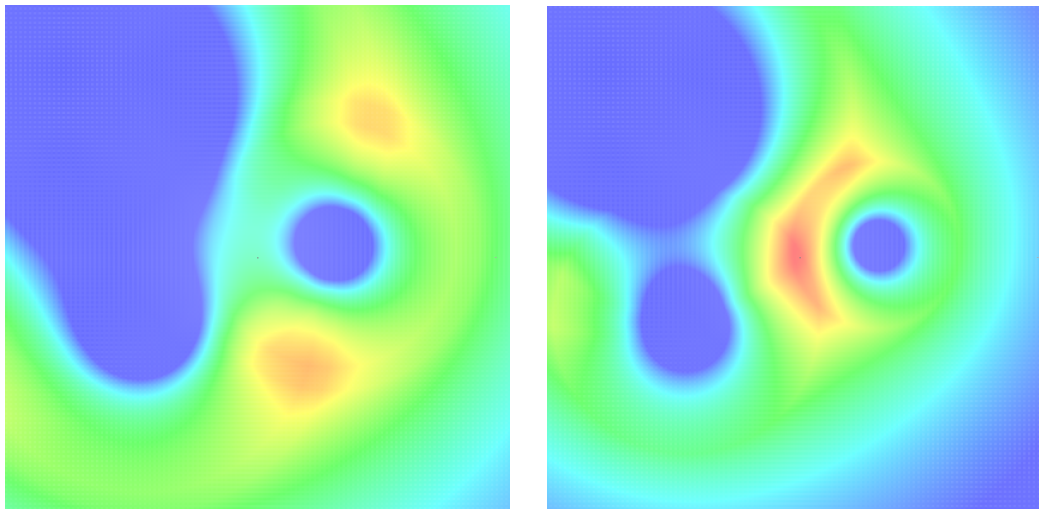
ured with force balance accelerometers. These are sensitive down to DC to record permanent deformation, and usually up to at least 100 Hz to cover the range which can cause damage to structures, as specified in building codes.



**Figure 2.1:** Illustration of a seismic event that generates seismic waves recorded by seismic sensors at different times: the seismic waves consist of the faster pressure wave (P-wave) indicated by the red dotted rings and the slower shear wave (S-wave) indicated by the blue dotted rings. The seismic sensors are indicated by blue triangles. The seismograms as time elapses (from left to right and from top to bottom) are shown above each sensor. The seismic event is indicated by the double harpoon symbol.

Seismic events generate seismic waves (see Section 2.3) that are recorded by seismic sensors as illustrated in Figure 2.1. The phase arrivals of the different types of waves need to be identified or “picked” by a data processor, which can be human or automatic. In a mining environment, the true seismic wave velocities are not known due to unknown geology, voids, excavations etc. A velocity model is assumed, which typically has a 3% – 5% error. For the assumed velocity model, the mismatch, in terms of its 3D location and time, between the picked phase arrivals and the expected phase arrivals at each single sensor can be calculated. The 3D location and time which minimise the average mismatch over all sensors is then an estimate of the hypocentre (the point within the earth from where the seismic waves originated) and origin time (the start time of the seismic rupture obtained

from the seismic sensors which are synchronised by a single GPS timing source) of the seismic event. As set out in Section 2.7) and illustrated in Figure 2.2, the mismatch can be converted into a probability.



**Figure 2.2:** The log-probability map of the synthetic event of Figure 2.1 at times: a) 50 milliseconds before or after the origin time (left) and b) at the origin time (right). This illustrates how the hypocentre and origin time of a seismic events are estimated. The assumed velocity model has a 5% error, the P-wave velocity is assumed to be 5500 m/s, while the assumed S-wave velocity is 3500 m/s. The phase arrival error is assumed to be 1 millisecond. The log-probability map has a surface area of 1 square kilometer. The highest probability is close to where and when the true hypocentre and origin time of the synthetic event are.

Sections 2.2-2.7 cover aspects of seismology needed to understand the concepts of this dissertation. Only the basics are covered. These sections may be skipped by a reader familiar with seismology. The job description of human processors and the quality control procedure at IMS are explained in Section 2.8. The same job description should apply to the automatic processor with one addition: to be able to decide when a recorded event cannot be processed successfully. The importance of an automatic processor is discussed in detail in Section 2.9.

## 2.2 Seismic Source Parameters

Seismic potency  $P$ , seismic moment  $M$ , radiated energy  $E$  and corner frequency  $f_0$  are the basic source parameters inferred from recorded seismic waveforms:  $P$  is proportional to the low-frequency plateau of displacement spectra,  $E$  is proportional to the integral of velocity-squared spectra and  $f_0$  is the frequency of maximum radiated

energy (Mendecki [1]). Seismic potency  $P$  is representative of inelastic deformation, expressed in cubic meters, at the source of a seismic event. Seismic moment  $M = \mu P$  is another measure of the size of the event, where rigidity  $\mu$  (Keylis-Borok [24]) is measured in pascals (Pa) while  $M$  has the dimensions of torque units of Newton meters. Radiated energy is representative of the corresponding work done (or energy) radiated as elastic waves; it has units of Joules. In practice, the inversion of seismic potency is more robust than radiated energy, and for this reason, the logarithm of seismic potency  $\log P$  is probably the simplest, most reliable measure of seismic event size for the mining environment. Other notable source parameters (derived from seismic potency, radiated energy and/or corner frequency) include source radius  $r = 2.34v^S/2\pi f_0$ , where  $v^S$  is S-wave velocity (Brune [25]), static stress drop  $\Delta\sigma = 7\mu P/16r^3$  and apparent volume  $V_A = \mu P^2/E$  (Mendecki [26]).

## 2.3 Physics of Seismic Waves

Seismic waves are waves of energy travelling through rock and are generated by the sudden inelastic deformation caused by earthquakes, volcanoes, landslides and so forth. Seismic waves are recorded by a seismic sensor and can be of the velocity (geophone) or acceleration (accelerometer) type. The velocity of seismic waves depends on rock density and elasticity. Velocity tends to increase with depth and ranges from approximately 2 to 8 km/s within the crust of the earth.

Earthquakes or seismic events create distinct types of waves each having a different velocity. The different types of seismic waves can be divided into body waves and surface waves. Body waves travel through the interior of the earth, while surface waves travel across the surface. Body waves, travelling in three dimensions, attenuate over shorter ranges than surface waves. In mine seismology, surface waves do not play a role, as seismic sensors are installed kilometers under the surface of the earth. Only the P- and S-body waves discussed in the next section are of concern.

### 2.3.1 Types of Body Waves

**P-waves** Primary waves (P-waves) are compressional waves longitudinal in nature. Of all seismic waves, P-waves travel the fastest through all mediums. The velocity of pressure waves is 330 m/s in air, 1450 m/s in water and between 4500 and 6000 m/s in typical hard rock when generated by seismic events. The velocity of P-waves in a homogeneous isotropic medium is given by

$$v^P = \sqrt{\frac{K + \frac{4}{3}\mu}{\rho}}, \quad (2.3.1)$$

where  $K$  is the bulk modulus (the modulus of incompressibility),  $\mu$  is the shear modulus and  $\rho$  is the density of the material through which the wave propagates. The elastic moduli P-wave modulus,  $M$  is defined so that  $M = K + 4\mu/3$  and thereby  $v^P = \sqrt{M/\rho}$ . Table 2.1 lists the P-wave velocities for different types of rock.

**Table 2.1:** Table showing the P-wave velocity (in m/s) for different rock types.

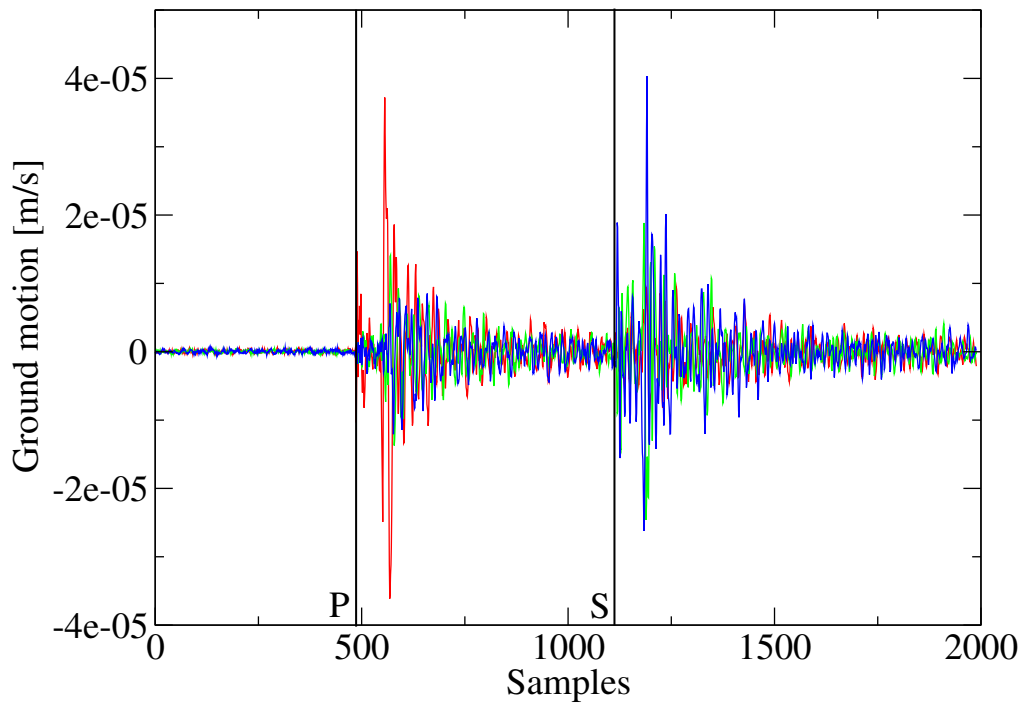
| Rock type | Velocity, [m/s] |
|-----------|-----------------|
| Sandstone | 4600 – 5800     |
| Shale     | 1800 – 4900     |
| Limestone | 5800 – 6400     |
| Dolomite  | 6400 – 7300     |
| Anhydrite | 6100 – 20000    |
| Granite   | 5800 – 6100     |
| Gabbro    | – 7200          |

**S-waves** Secondary waves (S-waves) are shear waves that are transverse in nature. S-waves travel at about 60% of the speed of P-waves, but their speed varies depending on rock type. The radiated S-wave energy is typically larger than the radiated P-wave energy and can easily be ten times higher (Koyama [27]). In Table 2.2 the S-wave velocities is given for different types of rock.

**Table 2.2:** Table showing the S-wave velocity (in m/s) for different rock types.

| Rock type | Velocity, [m/s] |
|-----------|-----------------|
| Limestone | 2000 – 3300     |
| Dolomite  | 1900 – 3600     |
| Granite   | 2500 – 3300     |

For the P-wave, the ground/particle motion is parallel to the direction of wave propagation, while for the S-wave, the ground/particle motion is perpendicular to the direction of wave propagation. This is useful in distinguishing between P- and S-waves when two or more events arrive at a sensor at nearly the same time as illustrated in Figure 2.3.



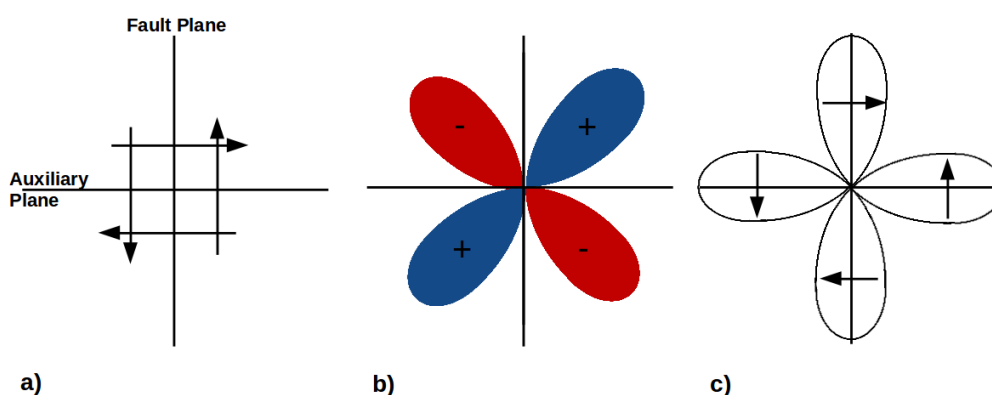
**Figure 2.3:** Example of a seismogram, with a sampling rate of 6000 Hz, containing a double event. The red, blue and green signals represent the three spatial directions of movement. The directivity of the particle motion of the P-wave and S-wave is orthogonal. When this directivity is not taken into account, the second P-wave arrival could then easily be seen as the S-wave of the first event.

## 2.4 Radiation Pattern

The *radiation pattern* describes the spatial distribution of energy radiated during an earthquake; it differs depending on the mechanics of the seismic source. Rock slip on a fault plane is described by a pair of force couples, which are pairs of forces in opposite directions, but separated by a small distance. Due to the conservation of angular momentum, force couples do not exist in isolation. If one couple is oriented in the slip direction, then another must exist in the perpendicular auxiliary plane. This pair of force couples is referred to as double-couples.

The direction of ground motion of a P-wave at a specified location, the so-called polarity of a P-wave, can be: a) compression (up or push) or b) dilation (down or pulled). The respective radiation patterns for the P- and S-wave of a double-couple source are shown in Figure 2.4. The blue quadrants indicate compression, while the red quadrants indicate dilation. The arrows indicate the particle motion direction of the S-wave. Note: the angle between the maximum radiated energies of the P-wave and S-wave is 45 degrees. Furthermore, both sets of double-couples would describe

the P-wave polarity observed in the recorded seismograms in the four quadrants around the seismic source with two possible nodal (slip) planes. The ambiguity may be resolved through a-priori geological information (e.g. pre-existing faults or mining geometry) or the distribution of aftershocks. Note: the idea of equivalent forces is only a simple model of a potentially complex faulting process.



**Figure 2.4:** Radiation patterns resulting from a double couple event. The radiation pattern describes the spatial distribution of energy radiated during an earthquake. The double couple shown in a) produces P-wave and S-wave radiation patterns as shown in b) and c), respectively. In b) the blue lobes indicate that the first particle motion for the P-wave is directed away from the seismic source, while the red lobes indicate motion towards the source. In c) the arrows indicate the direction of the first particle motion for the S-wave in the corresponding lobes.

## 2.5 Triggering and Association

Subsections 2.5.1 and 2.5.2 are extracts from Mendecki [1] and explain how recorded data is generated.

### 2.5.1 Triggering

Generally, automatic detection of the presence of a seismic signal is desirable. When this detection takes place in real time and is used to initiate further action from the system such as recording and association, it is referred to as triggering. For low frequencies and distant, long duration events, continuous recording may be justified, but in the mining environment where the relatively high frequencies of interest necessitate a correspondingly high sampling rate, and the nearby events



produce ground motion of short duration, triggering is necessary to reduce the amount of data recorded and to initiate the process which culminates in producing a report. In situations which might defeat a simple trigger algorithm, and especially where further processing must proceed automatically because of the volume of data, it is useful to have a further validation phase which verifies that the entire segment of recorded data conforms to a single processable seismic event.

In the mining environment, the prime concern has historically been the quantification of damaging events, which naturally have a very high signal to noise ratio. With the shift in emphasis to ever more sensitive networks in the search for precursive activity, no shortage of events with good signal to noise ratio has been experienced; hence the processing routines are still designed to function under these circumstances, and the triggering algorithms need only look for a sudden increase in amplitude.

The simplest trigger is simply a fixed threshold, and a trigger is declared if a single sample exceeds this value. This does not work well where the noise is environmental as the level tends to change with mining activity. The noise level may be represented by a long term average (LTA) of some estimate of the instantaneous signal amplitude. This is then compared with a short term average (STA) of the amplitude, the trigger criterion being  $STA/LTA > R_t$ , where  $R_t$  is the trigger ratio. There are several parameters which characterise this algorithm. The period over which the LTA is taken represents a cut-off between the shorter period signal and the longer period noise envelope variations; the period of the STA could be seen as the minimum duration of a valid event, and the trigger ratio is the minimum signal to noise ratio for a valid recording. Many sources of interference such as lightning, switching transients in power lines and single bit digital errors are impulsive in nature. Although of short duration, these can give rise to large amplitudes, which may, even when averaged over the STA period, be sufficient to push the STA over the threshold. One common variation on this algorithm is thus to use the median, rather than the mean, of samples within the STA window. The signal envelope  $E$  may be calculated by combining with a copy of the signal  $S$  which has been delayed by a  $90^\circ$  phase shift, performed by a Hilbert transformer  $E = \sqrt{S^2 + S_H^2}$  where  $S_H$  is the output of the Hilbert transformer (Hilbert [28]). A useful approximation is to add a fraction of the derivative to the original signal. This has the effect both of smoothing the envelope compared with simply taking the absolute value, and also of responding to frequency changes (Allen [2]).

## 2.5.2 Association

In the initial analysis chain, the so-called associator must decide which triggers from individual sensors correspond to the same event. For association between two triggers, the difference in times must be less than or equal to the seismic wave travel time  $\Delta t_{ij}$  between the two sensors:

$$|T_i - T_j| \leq \Delta t_{ij} = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}}{v^W}, \quad (2.5.1)$$

where  $T_i$  is the trigger time of the  $i$ th sensor at position  $(x_i, y_i, z_i)$  and  $v^W$  is the wave propagation velocity for the phase on which it is assumed the sensors triggered. Equation 2.5.1 describes a test only for pairs of triggers, and further rules are required to find all triggers from an event, especially when spurious triggers may occur which result in a situation where trigger A associates with trigger B and trigger B with trigger C, but where trigger A does not associate with trigger C. A popular strategy, especially in systems with direct transmission, is to open a time window upon receipt of the earliest trigger, with a duration equal to the longest travel time between any two sensors in the network, and to accept any triggers occurring within this window as part of the same event (Lee and Stewart [29]). If fewer than some minimum number of triggers are received within this window, the event and at least its initial trigger are discarded, and the process repeated for the succeeding trigger. This method is sometimes referred to as a “network trigger” rather than association, by analogy with individual sensor triggers. An event defined in this way may still contain spurious triggers, and it is then worthwhile to apply Equation 2.5.1 exhaustively to all pairs to define the largest subset of these triggers which is mutually consistent, before proceeding with location or further processing (Lawrence [30]).

In order to unambiguously separate the triggers caused by sequential events, there must be no triggers for a period equal to the travel time between the most widely separated sensors under consideration. The associator may use various heuristics to improve its performance when this criterion is not met, such as including each sensor only once in each event, but any sustained activity which violates this principle will, in the presence of spurious triggers, result in triggers from different events being associated or triggers from the same event being separated. There is thus a relationship between the diameter of a group of sensors and the maximum rate at which events on which they trigger can occur and be reliably associated. The word “group” is deliberately used in this

instance as it may comprise a subset of the sensors in a network. The defining property is that only members of the group be considered for mutual association. If the maximum travel time between any pair of sensors within the group is  $\Delta t_{\max}$ , then events must be separated by at least  $2\Delta t_{\max}$  so that the last trigger of the first event is separated by  $\Delta t_{\max}$  from the first trigger of the second event. If we assume a Poisson (random) distribution of events with time, then the probability of no events occurring during the period  $2\Delta t_{\max}$  is related to the average rate of occurrence of events  $r$  by (Mood and Graybill [31]):

$$p = e^{-2r\Delta t_{\max}}. \quad (2.5.2)$$

If we use a probability of failure  $f = 1 - p$  to express the probability that our criterion will not be met, take natural logs on both side of Equation 2.5.2, and use the approximation  $\ln(1 + x) \simeq x$  for small  $x$ , then

$$f = 2r\Delta t_{\max}. \quad (2.5.3)$$

so

$$r \simeq \frac{f}{2\Delta t_{\max}} \ll \frac{1}{2\Delta t_{\max}}, \quad (2.5.4)$$

i.e. for a low probability of interference between events, the average event rate must be much less than the reciprocal of the minimum separation time. This problem is compounded by the fact that mining often does not proceed continuously, but in discrete steps caused by periodic blasts. The event rate following such a blast greatly exceeds the daily average for a long time compared with the travel times.

For example, consider a mining network with a longest travel time of 0.5 s which gathers 1000 events per day. The average time between events over this period is then 86.4 s. If we make the common assumption that 90% of events occur in 10% of the time after a blast, the average interval during this period reduces to 10.4 s, and from Equation 2.5.3 the probability of two events occurring within the 1 s minimum spacing is 10%. Considering that each such overlap could result in two erroneously associated sets of triggers, this network is already a candidate for introducing groups for association purposes. The actual situation is even worse, as the event rate is not constant for two hours after the blast, but decreases, meaning that immediately after the blast the rate would be far higher than the average one event per 10.4 s rate used above. Also limits on system throughput might result in many events being lost during this time and not reflected in the 1000 recorded events per day which was the initial assumption.

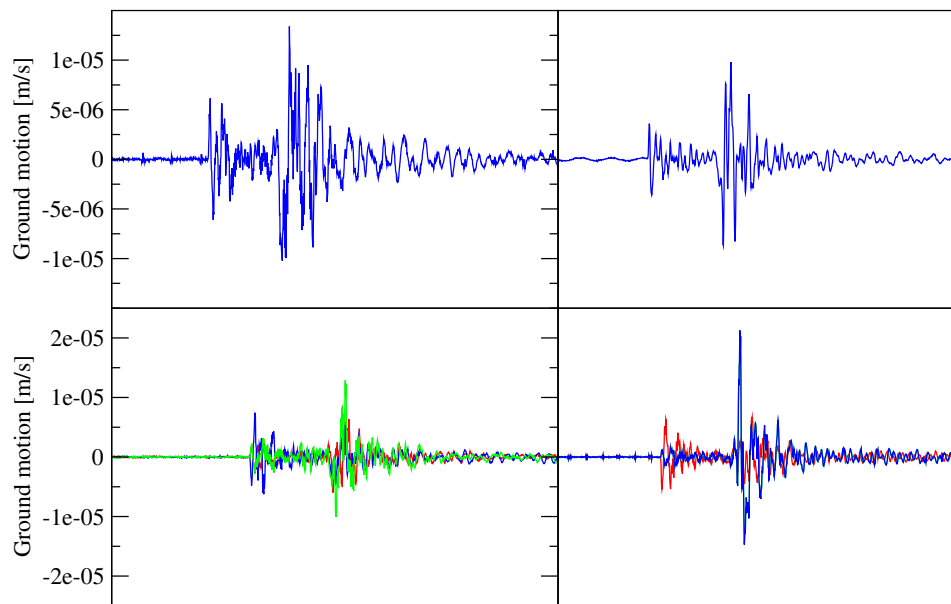
This analysis assumes that only information available to the associator is the arrival time of a single seismic phase. This is generally the case because the associator (or network trigger) decides whether a sensor trigger should be recorded for further processing or discarded. If resources are available for the temporary storage of, and further parameter extraction from a potentially spurious waveform, then maximum amplitude, dominant frequency, polarization parameters, the arrival of a second phase, or a full location, from a single sensor or two or more closely spaced sensors may be used to improve the associator performance when events occur almost simultaneously in different parts of the network.

## 2.6 Recorded Data at Mines

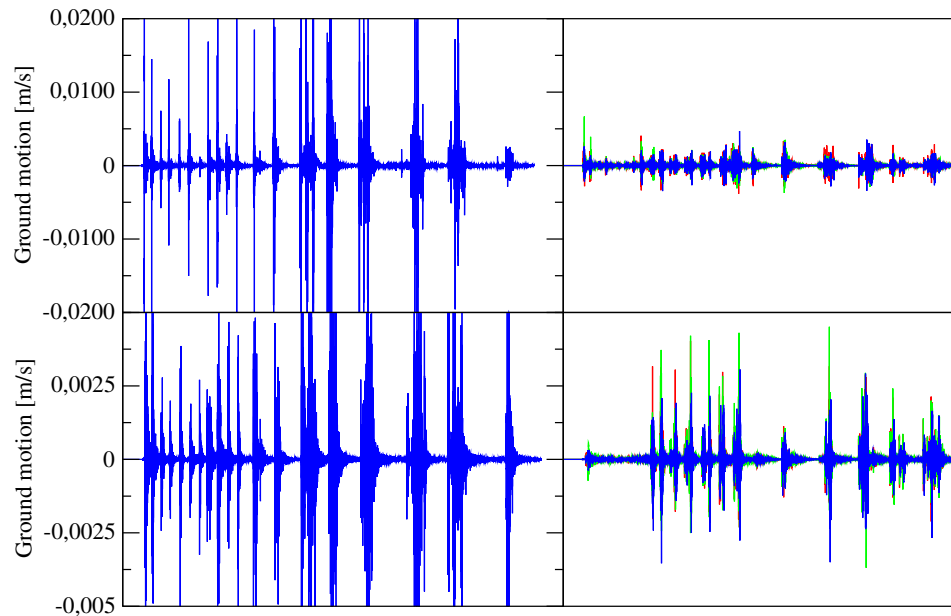
The three classification groups (seismic events, blasts and rejected (noise) events) have distinct features. As shown by example in Figure 2.5, seismic events always have visible (amplitudes much larger than the background noise) P- and S-wave arrivals. Most mining-induced seismic events have a smaller amplitude P-wave arrival and a larger S-wave arrival.

Production blasts, by contrast, typically occur at certain times of day and have multiple arrivals separated at fixed times (see Figure 2.6). As blasts are explosive in nature, in most cases only P-waves will be generated or at least weak S-wave amplitudes. Development blasts typically occur at certain times of day and have single arrivals with no clear S-wave arrival. The duration of the blast signal tends to be longer than that of a typical seismic event (see Figure 2.7).

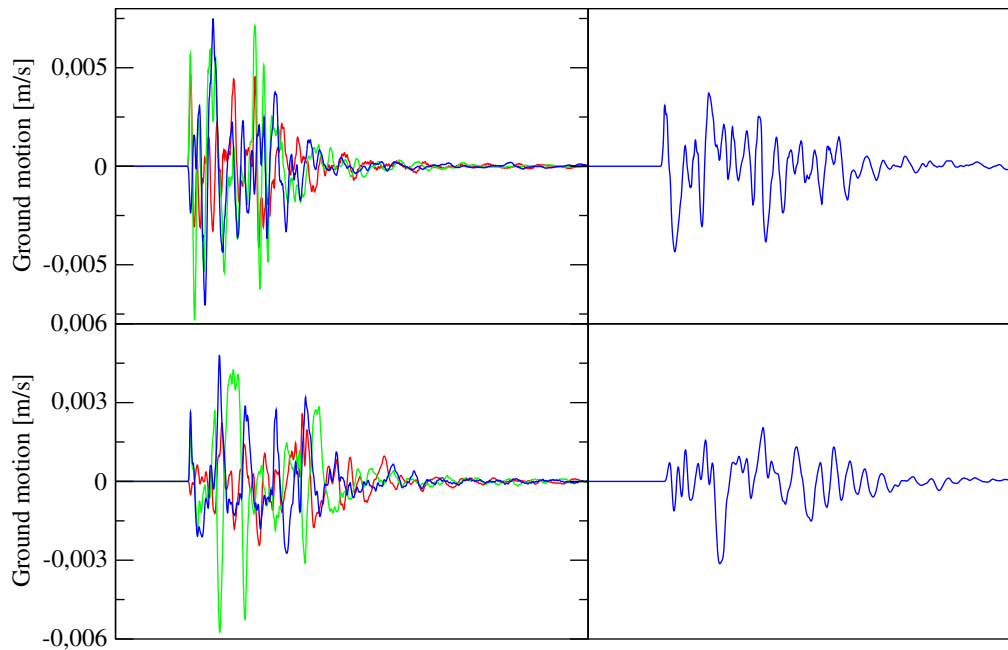
Noise events have many different characteristics and must be treated on a 'per mine basis'. Fortunately, the three most common types of noise, which are random noise, drilling and rock falling, are characterised by specific seismogram patterns. Multiple arrivals, distributed randomly in time and in amplitude (see Figure 2.8) characterise random noise. The drilling of holes in rock generates a distinctive pattern in that the inter-event time of the drill signals is constant and the amplitudes are similar (see Figure 2.9), while the falling of rock through an ore pass results in low frequency and small amplitudes (see Figure 2.10).



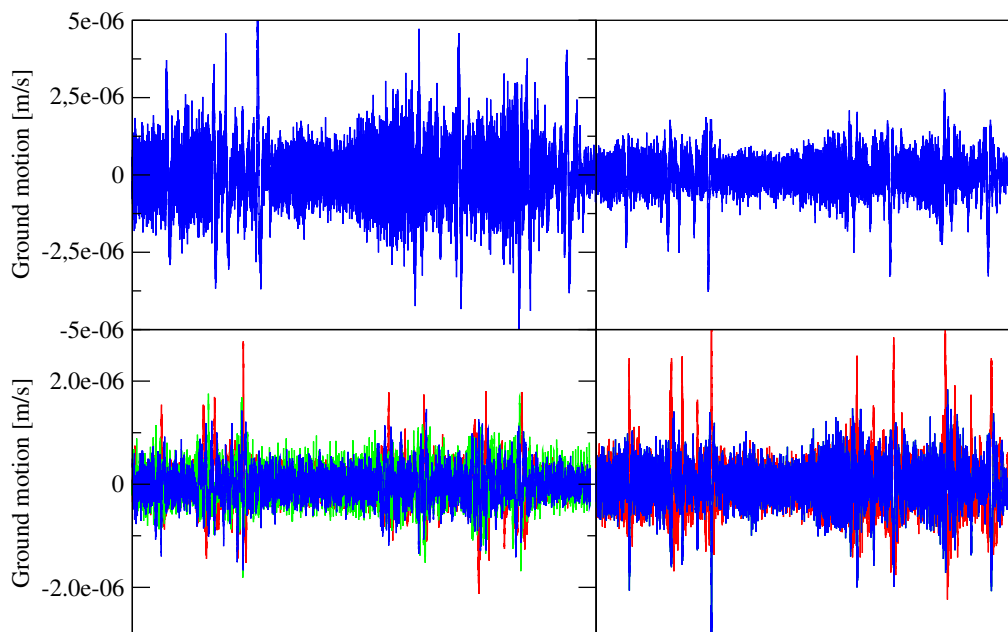
**Figure 2.5:** Typical seismic event recorded at a mine. In this example four seismograms are shown recorded by different sensors; see Subsection 2.5 for further discussion. The seismograms clearly exhibit the arrival of the faster P-wave followed by the slower but more energetic S-wave. Typically, the S-wave has larger amplitude than the P-wave.



**Figure 2.6:** Typical production blast recorded at a mine. The signals consist mostly of P-waves which are separated at fixed intervals.



**Figure 2.7:** Typical development blast recorded at a mine. The signal consists mostly of a P-wave arrival.



**Figure 2.8:** Typical noise event.

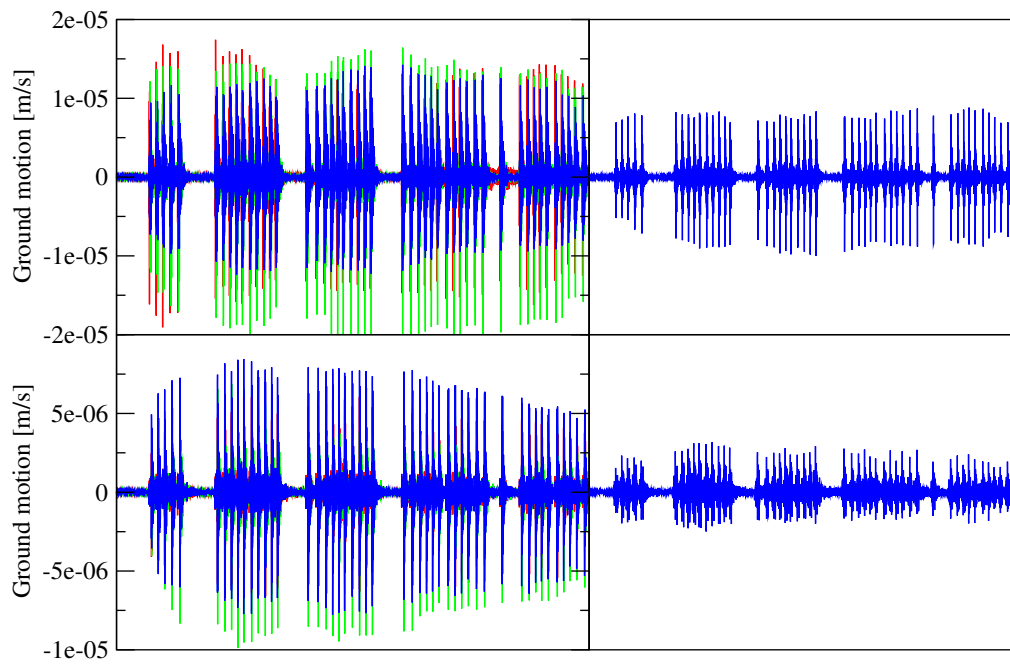


Figure 2.9: Drilling

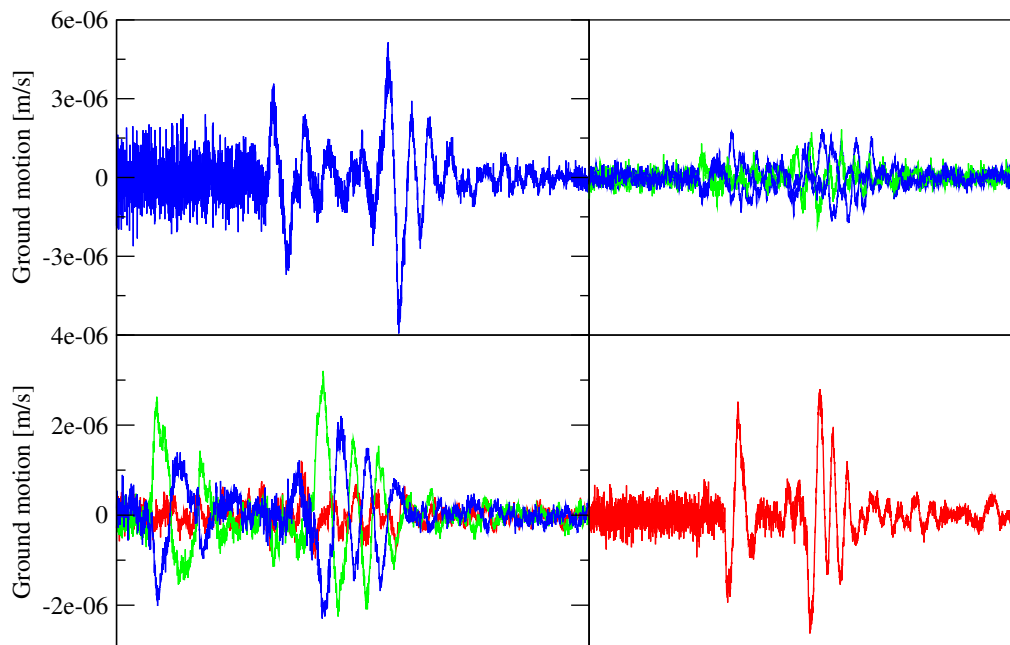


Figure 2.10: Mine noise induced by ore pass noise (rolling rock).

## 2.7 Location of Seismic Events

Consider a seismic event that occurs where the seismic waves generated are recorded by  $n$  sensors. The apparent velocity of the P-wave for sensor  $i$  is  $v_i^P$ , while the apparent velocity of the S-wave for sensor  $i$  is  $v_i^S$ . The apparent velocity is derived from the travel time measured between two points, assuming that ray-paths travel in a straight line; in reality seismic waves will travel in a straight line only when the rock mass is homogeneous and competent i.e. not fractured. The arrival time of the P-wave at sensor  $i$  is  $t_i^P$ , while the arrival time of the S-wave for sensor  $i$  is  $t_i^S$ . Furthermore, assume that the seismic event occurred at an unknown hypocentre of  $x, y, z$  and at time  $t_0$ . Assuming that the wave propagates in a straight line between the hypocentre of the seismic event and a sensor  $i$ , then the travel time of the P-wave is simply  $d_i/v_i^P$ , where  $d_i$  is the distance between the hypocentre of the seismic event and sensor  $i$ . Similarly, the travel time of the S-wave is  $d_i/v_i^S$ . The travel time equation (TTE) of the P-wave of the seismic event with unknown hypocentre and unknown origin time as recorded at sensor  $i$  is

$$t_i^P = t_0 + \frac{d_i}{v_i^P} + \varepsilon_i^P \quad (2.7.1)$$

and for the S-wave

$$t_i^S = t_0 + \frac{d_i}{v_i^S} + \varepsilon_i^S, \quad (2.7.2)$$

where  $\varepsilon_i^P$  is the travel time modelling error of the P-wave and  $\varepsilon_i^S$  is the travel time modelling error of the S-wave, both having an expected value of zero. The travel time equation is valid for each recorded sensor. To estimate the unknown hypocentre and origin time, a cost function can be formulated as the sum of the absolute difference between the observed arrival time  $t_i^P$  and expected arrival times  $t_0 + (d_i/v_i^P)$  for all  $i$ . The same holds for the S-wave. The cost function to be minimised in terms of the hypocentre and origin time is

$$\text{Cost}(x, y, z, t_0) = \sum_{i=1}^n \left\{ \left| t_i^P - t_0 - \frac{d_i}{v_i^P} \right| + \left| t_i^S - t_0 - \frac{d_i}{v_i^S} \right| \right\}. \quad (2.7.3)$$

In most scenarios not all P-wave and S-wave phase arrivals can be picked accurately; consequently noisy picks are disabled. Assume  $n^P$  P-picks are enabled and  $n^S$  S-picks are enabled, then the cost function to be minimised becomes

$$\text{Cost}(x, y, z, t_0) = \sum_{i=1}^n \left\{ I(P_{\text{en},i}) \left| t_i^P - t_0 - \frac{d_i}{v_i^P} \right| + I(S_{\text{en},i}) \left| t_i^S - t_0 - \frac{d_i}{v_i^S} \right| \right\}, \quad (2.7.4)$$

where  $I(P_{\text{en},i})$  is 1 if the P-wave pick is enabled for the sensor  $i$  and 0 otherwise; similarly  $I(S_{\text{en},i})$  is 1 if the S-wave pick is enabled for the sensor  $i$  and 0 otherwise.



Minimisation of the cost function given by Equation 2.7.4 is unstable and optimisation methods, for instance the Olsson and Nelson [32] or differential evolution (Storn and Price [33]) algorithm, may become trapped in local minima. To make the cost function more stable, the  $t_0$  parameter is eliminated in the following way (Mendecki and Sciocatti [34]): from Equation 2.7.1 it follows that the expected value of  $t_0$  denoted by  $\hat{t}_0$  is

$$\hat{t}_0 = t_i^P - \frac{d_i}{v_i^P} \quad (2.7.5)$$

which holds for all sensors  $i$  and is true for the S-wave as well. Hence

$$\hat{t}_0 = \frac{1}{n^P + n^S} \sum_{i=1}^n \left\{ I(P_{\text{en}}) \left( t_i^P - \frac{d_i}{v_i^P} \right) + I(S_{\text{en}}) \left( t_i^S - \frac{d_i}{v_i^S} \right) \right\} \quad (2.7.6)$$

which is simply  $\hat{t}_0 = \bar{t} - \bar{\tau}$ , where  $\bar{t}$  is the average arrivals times of enabled picks, that is  $\bar{t} = \frac{1}{n^P + n^S} \sum_{i=1}^n \{ I(P_{\text{en}}) t_i^P + I(S_{\text{en}}) t_i^S \}$  and  $\bar{\tau}$  is the average travel time of enabled picks, that is  $\bar{\tau} = \frac{1}{n^P + n^S} \sum_{i=1}^n \left\{ I(P_{\text{en}}) \frac{d_i}{v_i^P} + I(S_{\text{en}}) \frac{d_i}{v_i^S} \right\}$ . A more stable cost function that is minimised to obtain estimates of the hypocentre is

$$\begin{aligned} \text{Cost}(x, y, z, t_0) &= \sum_{i=1}^n I(P_{\text{en}}) \left| (t_i^P - \bar{t}) - \left( \frac{d_i}{v_i^P} - \bar{\tau} \right) \right| \\ &\quad + \sum_{i=1}^n I(S_{\text{en}}) \left| (t_i^S - \bar{t}) - \left( \frac{d_i}{v_i^S} - \bar{\tau} \right) \right|. \end{aligned} \quad (2.7.7)$$

Rather than minimising the cost function given by Equation 2.7.4, one could also maximise a probability distribution. As we are using the  $L_1$  norm, the appropriate distribution is the Laplace distribution and the probability distribution to maximise is

$$\frac{e^{-\text{Cost}}}{\prod_{j=1}^{n^P} \sigma_{t_j^P} \prod_{j=1}^{n^S} \sigma_{t_j^S}} \quad (2.7.8)$$

with

$$\text{Cost} = \left( \sum_{i=1}^n \left\{ I(P_{\text{en},i}) \frac{\left| t_i^P - t_0 - \frac{d_i}{v_i^P} \right|}{\sigma_{t_i^P}} + I(S_{\text{en},i}) \frac{\left| t_i^S - t_0 - \frac{d_i}{v_i^S} \right|}{\sigma_{t_i^S}} \right\} \right), \quad (2.7.9)$$

where  $\sigma_{t_i^P}$  and  $\sigma_{t_i^S}$  are given by Equation 4.3.11 in Section 4.3. Note that the logarithm of Equation 2.7.9 is the logarithmic probability map shown in Figure 2.2.

## 2.8 Data Processing Carried Out by IMS

IMS has 65 human data processors that are contracted out from various countries who work independently from home and who do processing services for about 80 mines located throughout the world. As a result, IMS has to deal with an average of 35000 recorded events per day. This daily total far exceeds that of the National Earthquake Information Center (NEIC) based in the USA, which locates only about 50 earthquakes per day or approximately 20000 a year.

A typically experienced data processor can process around 120 events per hour. Thus on average 10-15 processors should be on shift at any given time in order to process events within one hour of occurrence.

Seismic data from IMS customers is processed by decentralised human processors using a modified version of a data processing software called Trace. A work list of events is automatically generated and made available for downloading and processing via Trace. The processor is required to process each downloaded event accurately and timeously.

Upon receiving an event, the processor is tasked to classify the event into one of three types:

1. Noise
2. Blast
3. Seismic event

This classification is based on the character of the waveforms and time of day as discussed in Section 2.6. Noise events are to be rejected, while events classified as seismic must be picked. This entails selecting the P- and S-wave arrivals using the following guidelines issued to processors:

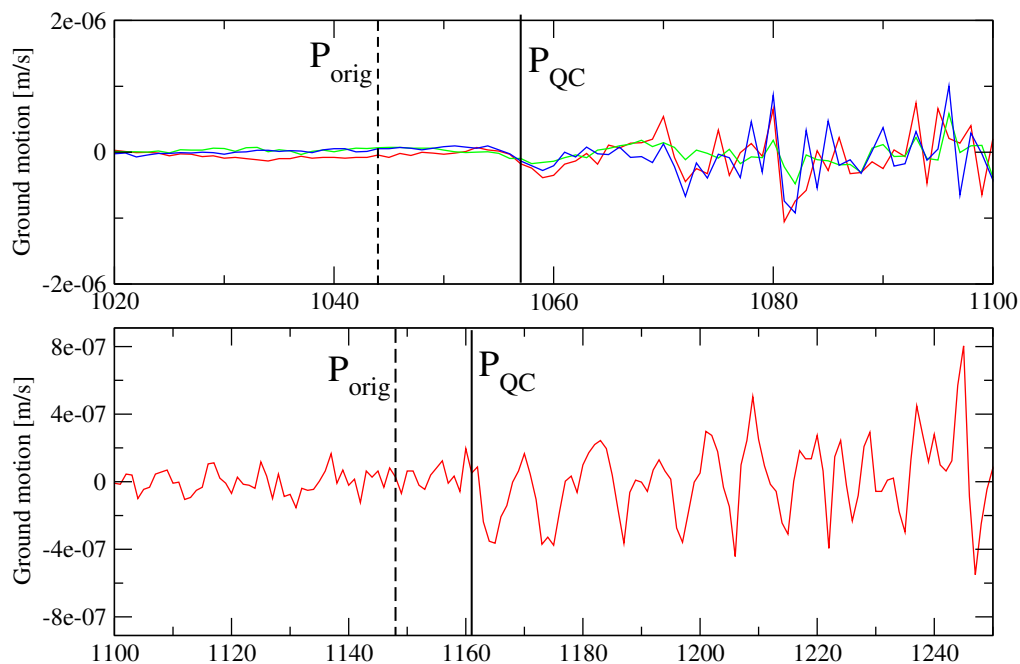
1. For an event to be classified seismic, the typical arrival waveform should be visible in at least six seismograms and must include at least one P-pick and at least one S-pick. The location residual (see Equation 4.9.1 for definition) must be below 3%.
2. Make use of the expand button to zoom into the arrivals.
3. Use the travel time window as a guide to where an arrival may be.
4. Preferably use tri-axial (seismic sensor with three components) waveforms above uni-axials (seismic sensor with only one vertical component) where possible.
5. After picking a P-arrival on a tri-axial sensor, use the rotate button to help identify the S-arrival.

6. Preferably use the first seismograms where possible.
7. Before disabling picks or rejecting a seismogram, try the S-P mode.
8. Use the velocity curve button for noisy accelerometer signals.

Mine-specific rules are made available to the processor via pop-ups. These must be followed as closely as possible. Once the event has been processed to within the mine specific rules, the event must be saved.

After processing, the processor's judgement and data is subjected to IMS quality control (QC) procedures.

Ten percent of all data is reprocessed by the more experienced data processors for QC purposes. Every QC event has an initial score of 100 and marks are deducted for any misclassification (difference between original classification and QC classification) or any phase arrival differences (picking differences) that are outside the error bounds of the arrivals points (see Figure 2.11).



**Figure 2.11:** Seismogram showing the original picks (black vertical dashed lines) and the QC picks (black vertical solid lines). In the top and bottom figures, it is clear that the original P-wave picks are too early. The original processor will therefore lose marks for this event. If the average score for the QC session is less than 85%, the processor does not get paid for that session.

## 2.9 Importance of Fast Automatic Processors

Contractually, the turn-around time for large events must be less than 10 minutes and for smaller events less than an hour. The processing service will be more useful if these turn-around times can be improved. The turn-around time consists of 6 stages:

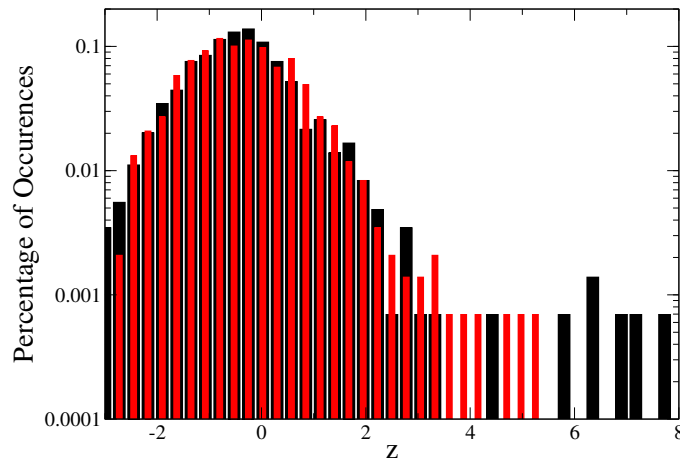
- The time it takes the recorded event to be sent, via the Internet, from a mine somewhere in the world to the IMS office located in Somerset West, South Africa.
- The time it takes the event to be sent, via the Internet, from the Somerset West office to the data processors in South Africa, China and India.
- The waiting time for processing other events already in the human processor's work list; this depends on how busy the processor is.
- The time it takes the human data processor to classify an event ( $\pm 5$  seconds) and, if the recorded data was classified as a seismic event, the time it takes to pick phase arrivals ( $\pm 1$  minute).
- The time it takes the event to arrive, via the Internet, at IMS offices after being processed.
- The time it takes the event to be sent, via the Internet, from IMS offices back to the mine.

Typically, about 70% of the total turn-around time is taken up by the human processor.

IMS needs to optimise the number of human data processors scheduled to work at any given time. Having too few processors often results in processed data not getting back to the mine in time, which may compromise safety. On the other hand, having too many processors on shift puts pressure on wages and will lead to individual processors not earning a competitive income, resulting in high turnover or even a high percentage of processors resigning simultaneously. There is an optimal number of processors to be appointed. The predictability of the number of events per hour per week is therefore critical.

Unfortunately, however, there are large fluctuations in the number of events per hour; events occur in bursts, and the number of recorded events per hour is therefore fat-tailed. The aftershock distribution follows the Omori-law (Omori [35]) which is a power-law distribution. Noise events such as underground drilling or hydro-hammering, typically also occur in bursts. The fat-tailed nature of recorded events at mines makes it difficult to schedule the correct number of data processors at any given hour. Without an automatic processor, the risk associated with over- or under-scheduling is high.

To illustrate the importance of reducing the highly variable event rate to be processed by humans, a histogram of the hourly number of events for data recorded from 01 June 2016 until 01 August 2016 is shown in Figure 2.12. The kurtosis of the hourly rate distribution is 34.8, indicating that the hourly number of events is not forecastable with the necessary accuracy and that predicting the required number of data processors on a shift is almost impossible.



**Figure 2.12:** Histogram of the hourly standardised number of events for all events (black bars) and manual events processed by humans (red bars). The standardised number of events is  $z_{ij} = (x_{ij} - \mu_i)/\sigma_i$ , where  $x_{ij}$  is the number of events in hour  $i$  for the  $j$ th day,  $\mu_i$  is the expected number of events for the  $i$ th hour and  $\sigma_i$  the standard deviation of the number of events for hour  $i$ . The manual events (all events that cannot be automatically processed successfully) have fewer outliers making scheduling of human processors (when automatic processing is carried out) much easier than when auto-processing is not done. The kurtosis is 34.8 for all events, but only 4.6 for manually processed events. It is impossible to have enough processors on shift each hour owing to this fat-tailed behaviour.

To quantify the utility of fast processing, one may define the available information  $I(t)$  at time  $t$  as the normalised total number of events that was processed between time 0 and time  $t$ . The normalisation is taken as the ratio of the processed events and the sum of processed and unprocessed events. Further, the information gain function,  $IG(t)$ , is the derivative of  $I(t)$ . We have that

$$I(t) = \int_0^t IG(\tau) d\tau \quad (2.9.1)$$

and specifically  $I(\infty) = \int_0^\infty IG(\tau) d\tau = 1$ . Assuming exponential decay, then the utility of a processor,  $U$ , may be defined as

$$U = \int_0^\infty IG(\tau) e^{-\tau} d\tau. \quad (2.9.2)$$

The maximum value of the utility function is 1, when all events are processed immediately.

Here is a typical example. On 9 July 2016 between 13h-14h GMT+2, 4808 events occurred, while the expected number of events for that hour was only 1700. This means that twenty-one processors would have been on shift if no auto-processor was available. These twenty-one processors, each working at a rate of 100 events/hour, would have had to work for 2.29 hours in order to process those events. The auto-processor, as described in this dissertation, could process all events in 20 minutes. Therefore, the auto-processor has  $I(t) = 3t$  and the human processors  $I(t) = 0.43t$ . Hence we have a utility of 0.85 for the automatic processor,

$$U = 3 \int_0^{1/3} e^{-t} dt = 3 - 3e^{-1/3} = 0.85, \quad (2.9.3)$$

compared to a lower utility of 0.39 of the human processor

$$U = 0.43 \int_0^{2.29} e^{-t} dt = 0.43 - 0.43 e^{-2.29} = 0.39.$$

Outliers contained in the recorded data for the period 01 June 2016 – 15 August 2016 are shown in Table 2.3. There were 6 hours for the time period 1 June – 10 August 2016 during which many more events occurred than expected. The automatic processor was able to process these bursts successfully enough so that processing did not fall behind. The average process rate of manual processors is assumed to be 100 per hour.  $\mu_{all}$  is the average number of events that will occur in each hour, while  $\mu_{man.}$  is the expected number of events that the automatic processor will not be able to handle, i.e. the number of events that needs to be processed by human data processors. The  $z_{obs}$  and  $U_{obs}$  are the actual standardised score and utility, respectively, of the hybrid set-up (the automatic processor and the manual processors). The  $z_{man. only}$  and  $U_{man. only}$  entries are the expected standardised score and utility assuming no automatic processor was available.

**Table 2.3:** Number of recorded data per hour outliers.

| Date    | Time | #events | $\mu_{all}$ | #man. | $\mu_{man.}$ | #auto. | $z_{obs}$   | $z_{man. only}$ | $U_{obs}$ | $U_{man. only}$ |
|---------|------|---------|-------------|-------|--------------|--------|-------------|-----------------|-----------|-----------------|
| 15 June | 20h  | 3134    | 1815        | 829   | 718          | 2305   | 0.82        | <b>6.01</b>     | 0.87      | 0.48            |
| 21 June | 12h  | 3893    | 1721        | 1008  | 777          | 2885   | 1.89        | <b>7.30</b>     | 0.85      | 0.41            |
| 21 June | 13h  | 4866    | 1665        | 1228  | 710          | 3638   | 3.55        | <b>6.91</b>     | 0.83      | 0.32            |
| 21 June | 14h  | 3966    | 1544        | 1240  | 720          | 2726   | 3.86        | <b>8.47</b>     | 0.82      | 0.37            |
| 08 July | 13h  | 4683    | 1665        | 1513  | 711          | 3170   | <b>5.50</b> | <b>6.51</b>     | 0.78      | 0.34            |
| 08 July | 18h  | 3621    | 1832        | 1354  | 840          | 2267   | 3.20        | <b>6.51</b>     | 0.80      | 0.45            |

## 2.10 Summary

The three different classes of recorded data at mines are seismic events, rejected events and blasts. The phase arrivals of seismic events need to be determined by processors, be it manual or automatic, with a normalised distance residual typically smaller than 3%. The assumption is made that the seismic waves travel in a straight line according to the travel time equation (TTE)

$$t_i^P = t_0 + \frac{d_i}{v_i^P}$$

for the P-wave, and for the S-wave

$$t_i^S = t_0 + \frac{d_i}{v_i^S}$$

In this dissertation these equations are used throughout. In order to deliver timeous data for mine managers to make informed decisions, an automatic processor is essential. A successful automatic processor first needs to classify and then determine phase arrivals for events that were classified as seismic. The classification by the automatic processor is discussed in the following chapter.

# Chapter 3

## Automatic Classification

### 3.1 Introduction

The first task of a data processor, be it a manual or an automatic processor, is to classify recorded data as accepted events, rejected events or blast events. In natural earthquake seismology all recorded data consists of seismic events, so that classification arises only in mine seismology and perhaps in regional seismology. Automatic classification in mine seismology is a relatively new field. Dong *et al.* [36] carried out research on automatic discrimination of seismic events occurring at mines and blasts. As mentioned in Chapter 1, all other research was carried out on large blasts and regional earthquakes. In this chapter, an automatic classification scheme is derived. The proposed method is based on a probabilistic graphical model, namely a Hidden Markov Model (HMM) using a customised Viterbi decoding algorithm. An HMM can be used to model processes which consist of stages (or states) that occur in a predictable statistical order. It relies on knowledge that time series exhibit sequences typical of some feature of the physical phenomenon under scrutiny, in the present case one or more of the features set out in Subsection 3.3.3. HMMs have been employed successfully in areas such as speech, handwriting and gesture recognition and in bioinformatics.

The pre-existing knowledge used for our HMM includes the following features. The P-wave of a seismic event is faster than the S-wave. Therefore a seismogram of an event will start with noise, followed by the P-wave arrival followed by the so-called P-wave coda, followed in turn by the S-wave arrival and lastly the S-wave coda. While this sequence may be repeated and/or overlaid in the case of multiple events recorded in a single seismogram, we shall for the moment assume that only a single event sequence is being analysed. This expected sequence can be modelled as an HMM. The short-fall of first-order HMM as a classification technique is that it cannot directly model the relationship between the P-wave and the S-wave arrivals. The directivity of the particle motion of the P- and S-



wave is orthogonal. To model this orthogonality, higher-order HMM needs be used, but because the separation between the P-wave and S-wave is unknown, the exact dimension cannot be determined in advance. Furthermore, higher-order HMM is too slow to classify the large amount of data that IMS receives on a daily basis. The customised Viterbi algorithm derived below significantly decreases the joint probability of the sequence of observations in cases when the most likely state of the HMM is an S-wave arrival, not orthogonal to the already chosen P-wave state. In addition to testing for orthogonality, other checks will be discussed in this chapter. The proposed HMM works well as a pattern recognition tool, but fails to capture certain visually obvious features. While many more such tests could be added to the Viterbi algorithm, they slow down the process too much. For this reason only the most important tests are built into the Viterbi decoding algorithm and the rest of the post-pick checks are carried out once the HMM has been completed. This is set out in Section 4.9.

Section 3.2 covers the blast discriminator previously developed by IMS. Section 3.3 discusses hidden Markov models in detail, while in Section 3.4 the exact classification scheme is derived and examples of the automatic classification are shown for different recorded data. The chapter is concluded with a summary.

## 3.2 Blast Discriminator

The blast discriminator forms one component in the analysis as set out in this dissertation. It has been part of the IMS software since 2010. This section is therefore mostly an extract from an internal IMS research report and also published in Malovichko [37].

Extensive investigations have been carried out in seismology to build algorithms to discriminate between nuclear explosions and natural (tectonic) earthquakes. The following features of seismic signals or seismic source parameters were found to be useful for this task:

- the ratio of P- to S-wave amplitudes in various frequency ranges
- the ratio between amplitudes of body and surface waves
- the depth of the source and
- the mechanism of the source

Usually the discrimination takes into account several of these features.

A popular discrimination technique is a multivariate Bayesian Gaussian classifier. A vector  $\underline{v}$  of measured features is considered in the technique, where the dimension of the vector corresponds to the number of features. It is assumed that the population of blasts ( $X$ ) has features  $\underline{v}$  belonging

to a multivariate Gaussian distribution with mean feature vector  $\boldsymbol{\mu}_B$  and covariance matrix  $\Sigma_B$ , while the features of population of normal events ( $Q$ ) belong to a Gaussian distribution with mean vector  $\boldsymbol{\mu}_E$  and covariance matrix  $\Sigma_E$ . The following discrimination function could then be used for separating the populations of blasts and normal events:

$$Q(\mathbf{v}) = \left[ -\frac{1}{2}(\mathbf{v} - \boldsymbol{\mu}_B)^T \Sigma_B^{-1}(\mathbf{v} - \boldsymbol{\mu}_B) \right] - \left[ -\frac{1}{2}(\mathbf{v} - \boldsymbol{\mu}_E)^T \Sigma_E^{-1}(\mathbf{v} - \boldsymbol{\mu}_E) \right] + \frac{1}{2} \ln \frac{|\Sigma_E|}{|\Sigma_B|} \quad (3.2.1)$$

The function  $Q$  can be understood as a log Bayes factor of the two likelihoods.

Given the feature  $\mathbf{v}$  of a particular event, a positive  $Q(\mathbf{v}) > 0$  would classify it as a blast and a normal event otherwise. Utilisation of the technique includes two steps:

- Calibration: the vectors  $\boldsymbol{\mu}_B$  and  $\boldsymbol{\mu}_E$ , and matrices  $\Sigma_B$  and  $\Sigma_E$  are determined from a representative set of known blasts and known normal events.
- Application: for the event of interest, the features  $\mathbf{v}$  are calculated and expression 3.2.1 is evaluated.

The features included in  $\mathbf{v}$  are (see Section 2.2 for definitions of source parameters mentioned below):

- *Time of Blast*: Many mines have prescribed blasting times. During these, the probability of an event being a blast is high, while in other time intervals it is low. The feature *Time of Blast* which measures the difference of the time of the event with the nearest blasting time was found to be efficient in several mines.
- *Clustering in Time*: Blasts are frequently characterized by a repetition of similar signals within a short time interval. Signals from individual firings are within one buffer (seismogram). It is this signature of waveforms which is generally used by processors for classification of the event as a *blast*. In other cases the seismograms may contain seismic signals only from a single firing, making visual classification of the event difficult. To quantify the repetition effect, the similarity of seismic waveforms of the analysed event with preceding and subsequent waveforms is calculated within 20 seconds intervals. During such a calculation, seismic signals both in the same buffer and in the neighbouring events are considered.

A median value of the 5 highest stacked cross-correlation amplitudes is calculated, and the logarithm of this value is considered as the feature *Clustering in Time*. In general the feature varies between -2 and 0. The values close to 0 are indicative of the presence of multiple similar signals in the proximity of the analysed event. The opposite case (values below -1.0) shows either an absence of neighbouring signals or poor similarity of their waveforms with the waveforms of the considered event.

- *High versus Low Frequency of Seismic Waves*: Blasts (explosions) radiate waves at higher frequencies in comparison with normal seismic events. As mentioned by Bormann [38], “as compared to tectonic earthquakes the duration of the source process of explosions and the so-called rise time to the maximum level of displacement is much shorter (milliseconds as compared to seconds up to a few minutes) and more impulsive.” Accordingly, explosions of comparable body wave magnitude excite more high frequency oscillations (cf. seismic source spectra). Such a difference may be captured using various combinations of seismic parameters. For instance, it may be a ratio of seismic moment and corner frequency. We decided to use an equivalent characteristic: apparent stress  $\sigma_A$  which is proportional to the ratio of seismic energy over seismic moment ( $\sigma_A \propto E_S/M$ ) as the evaluation of seismic energy is more robust than that of corner frequency. The logarithm of this ratio is considered as the feature *High versus Low Freq Waves*. Generally blasts have higher  $\log\sigma_A$  than normal events.
- *Radiation Pattern*: There is a fundamental difference in radiation (discussed in Section 2.4) of seismic waves by blasts and that of normal events. A blast radiates predominantly P-waves which have positive polarity for all sensors, i.e. the first motion of each sensor is directed radially outward from the event source. Dynamic rupture of shear crack, on the other hand, radiates S-waves which are stronger than P-waves. The polarity of P- and S-waves varies depending on the direction of the source relative to the spatial orientation of the sensor. The basic characteristics of radiation are captured by a seismic moment tensor. The latter represents a set of fictitious dipoles (components of moment tensor) acting on a point in the source area which is considered to be elastic and reproduces the low-frequency amplitudes and polarities of seismic waves. Inversion of the moment tensor from the observed waveforms and analysis of its components makes it possible to discriminate blasts from slip-type and even from crush-type events. The discrimin-

ating feature based on the source mechanism is sensitive to the evaluation of polarities of first motions in P- and S-waves. The latter depends on the orientation of the sensors. For this reason, we restricted the radiation pattern signature analysis to the ratio of amplitudes of P- and S-waves. Polarities of the waves and spatial distribution of the sites are not taken into account here. This enhances the stability of the feature, but reduces its resolving power. A generalized relation between amplitudes of P- and S-waves is provided by comparison of  $M_P$  (seismic moment estimated from P-waves) with  $M_S$  (seismic moment estimated from S-waves). The routine procedure of source parameters calculation in the IMS software assumes a shear crack model, imposing a corresponding radiation pattern correction to P- and S-waves. Thus, for the normal events we expect  $M_P \simeq M_S$ , while for other source processes the applied radiation pattern correction is not valid, and as a result  $M_P = M_S$ . The logarithm of the ratio of  $M_P$  to  $M_S$  was selected as a feature *Radiation Pattern*. This gives  $\log(M_P/M_S) \simeq 0$  for normal events and  $\log(M_P/M_S) > 0$  for the blasts.

### 3.3 Hidden Markov Models

Subsection 3.3.1 sets out the mathematics of the HMM along with some useful algorithms as used in probabilistic graphical models, with the view of solving the inference problem. Subsection 3.3.2 shows how an HMM framework can be applied to seismic waves. Subsection 3.3.3 details the proposed feature extraction (the derived observations from the measured data), as well as the parameter estimation used in the HMM. Subsection 3.3.4 points out the shortcomings of modelling seismic waves as an HMM with the normal Viterbi algorithm, which leads to the detailed discussion of the customised Viterbi algorithm.

#### 3.3.1 HMM in General

In order to aid the reader who may not be an expert in probability theory, I start with a general and straightforward example of an HMM, leading to the formal definitions.

**Example of a Simple HMM:** An example of an HMM is the following game related to drawing balls from urns of infinite size (sampling with replacement). Imagine a group of people on a stage, and you are in the front seat. You can see who and how many people are on stage. A number of urns are placed before each

person. Each person (called a model or class in HMM theory) may have a different number of urns (called hidden states in HMM theory).

Each urn contains balls of different colours in different proportions (called the observation probability given the state): for instance, one urn may have only white balls and another urn an even number of black and yellow balls. The people, the number of urns in front of each person and the distribution of coloured balls of each urn are memorised by you. A divider is then put between you and those on stage such that you cannot see them anymore. The divider blinds you to certain aspects, hence the word hidden. Only one person behind the divider will play the game, but you do not know which one. One person on stage will be chosen at random and they will draw balls from the urns in front of them. The other people's urns will be removed from the stage. The balls (called observations) are thrown, one by one, over the divider for you to observe. The first draw can come from any urn, but you know the prior probabilities (called the initial state probabilities) of each urn for each person. The person on stage will also stay or move from one urn to another with a given probability, depending on the current urn from which he/she is drawing. Note that, for every person, these probabilities are known to you (this is called the transition probabilities and is presented as a transition matrix).

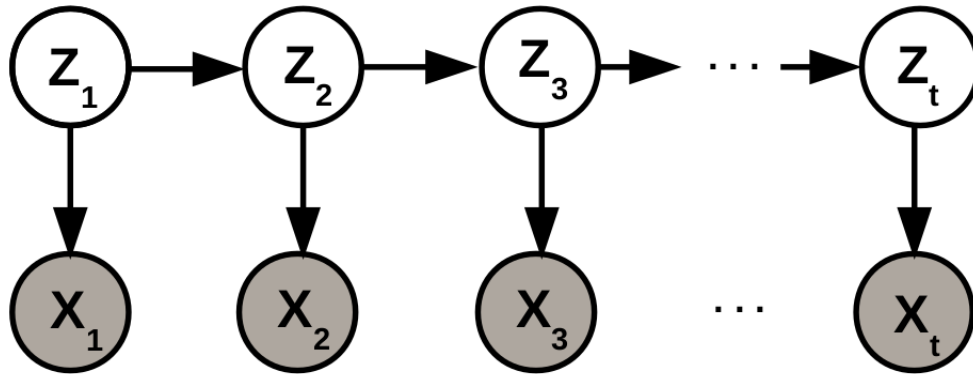
After receiving  $n$  balls and recording their colors, your task is to answer, as far as possible, three questions:

1. If I knew which person had drawn the balls, what is the joint probability of that sequence of balls observed by me? (Sampling probability)
2. If I knew the identity of the drawing person, what is the most likely urn (state) from which each draw was made?
3. If I did not know which person had drawn the balls, what is the probability, given the observed balls, that person  $i$  drew them? (Posterior)

An extension of the example to a multivariate one would be to expand the drawing of balls to drawing of balls, circles, triangles, etc.

**Theory:** A Graphical Model or Probabilistic Graphical Model (PGM) is a probabilistic model in which a graph expresses the conditional dependence structure among the relevant random variables. It is commonly used in probability theory, particularly Bayesian statistics, and machine learning. Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or factorised representation of a set of independences that hold in the specific distribution. The HMM is a particular example of a Bayesian network, which in turn is a particular case of a PGM. If the network structure of the model is a so-called directed acyclic graph (DAG), the model represents a factorization

of the joint probability of all random variables. The direction of the arrows determines the factorisation of the joint probability distribution. A Hidden Markov Model (HMM) is a statistical Markov Model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. We use the  $t$  subscript because HMM is typically used on a time series, where the observations are made at certain time intervals. An HMM can be presented as the simplest dynamic Bayesian network shown in Figure 3.1.



**Figure 3.1:** Representation of an HMM as a PGM. Each  $X_t$  is an observation, while the  $Z_t$  is a hidden state that produces the observation. Both  $X_t$  and  $Z_t$  are random variables.

The mathematics of HMMs was developed by Baum and Eagon [39]. The data is a sequence of observations. In many cases of machine learning applications, or problems, not all the measured data is relevant to the specific problem and is redundant. Certain relevant features (see Section 2.6) of the data can be extracted or derived from the measured data. This modified data is called the observations or the feature. The process is called feature extraction and is discussed in Subsection 3.3.3 for our application. The observations  $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_n)$  are not independent and each observation was produced by a hidden unknown state. The assumption is made that the latent variables  $z_t$  are discrete and that every  $z_t$  can be in  $s$  possible states. For an HMM, each state  $z_t$  which produced  $\mathbf{x}_t$  is dependent only on the previous state  $z_{t-1}$  and not any states before that. Thus, the hidden states have the Markov property

$$p(z_t | z_1, z_2, \dots, z_{t-1}) = p(z_t | z_{t-1}). \quad (3.3.1)$$

The HMM further assumes that  $\mathbf{x}_t$  is only dependent on  $z_t$  and on no other states before or after  $t$ , i.e.

$$p(\mathbf{x}_t | z_1, z_2, \dots, z_t) = p(\mathbf{x}_t | z_t). \quad (3.3.2)$$

The probability density function of  $z_t|z_{t-1}$  can be represented by a transition matrix of size  $s \times s$ . The transition matrix is

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1s} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{s1} & \phi_{s2} & \cdots & \phi_{ss} \end{bmatrix}, \quad (3.3.3)$$

where  $\phi_{ij} = p(z_t=j | z_{t-1}=i)$  for all  $t$ . The transition matrix is assumed to be independent of time. The probability distribution of the initial states  $z_0$  can be represented by a vector

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_s), \quad (3.3.4)$$

where  $\pi_i = p(z_0 = i)$ . The transition matrix and prior probability values, which we collectively call  $\boldsymbol{\theta} = (\Phi, \boldsymbol{\pi})$ , are unknown parameters and may be estimated on a training data set using Bayesian analysis or maximum likelihood estimates.

The incomplete or marginal likelihood function  $\mathcal{L}(\boldsymbol{\theta}, \mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta})$  can be written as

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta}), \quad (3.3.5)$$

where the summation is over all possible state sequences. Evaluating Equation 3.3.5 directly is not practical in most scenarios. In mine seismology, there are typically 2000 observations and 8 hidden states for each sensor. Given a transition matrix such as that shown in 3.3.13 below, at least 2 states are accessible at each time. Hence, there are at least  $2^{2000}$  possible state sequences: a number much too large to evaluate each possible sequence.

The Viterbi algorithm (Baum *et al.* [40]; Viterbi [41]) solved this problem as set out in this section. Firstly, by exploiting the Markov property, the factorisation of the complete likelihood function of the HMM model  $p(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta})$  can be written as

$$\begin{aligned} p(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta}) &= p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z}|\boldsymbol{\theta}) \\ &= p(z_1|\boldsymbol{\theta}) p(\mathbf{x}_1|z_1, \boldsymbol{\theta}) \prod_{t=2}^n p(z_t|z_{t-1}, \boldsymbol{\theta}) p(\mathbf{x}_t|z_t, \boldsymbol{\theta}). \end{aligned} \quad (3.3.6)$$

Baum *et al.* [40] and Viterbi [41] then solved the following three problems:

- Probability of the observed data given parameters  $\boldsymbol{\theta}$ , i.e.  $p(\mathbf{x}|\boldsymbol{\theta})$ .
- Given the observation sequence  $\mathbf{x}$ , how do we choose a corresponding state sequence  $\mathbf{z}$  which is optimal in some meaningful sense (i.e. best “explains” the observation).
- Finding the maximum likelihood estimates of  $\boldsymbol{\theta}$ .

The first two problems are solved by belief propagation (Pearl [42]). Specifically, the first problem is solved by the sum-product algorithm and the second problem

solved by the max-product algorithm (Weiss and Freeman [43]). The third problem is solved via the Expectation-Maximisation (EM) algorithm (Dempster *et al.* [44]). These algorithms are inference techniques for PGMs. The EM algorithm is not used as all our states in the training data sets are known/labeled; hence obtaining accurate initial estimates of  $\theta$  are obtained without the need for EM.

### The Forward Algorithm (Instance of Sum-Product Algorithm)

This algorithm obtains the joint distribution of  $\alpha_t(j) = p(z_t=j, \mathbf{x}_1, \mathbf{x}_1, \dots, \mathbf{x}_t | \theta)$  recursively from generic probability theory as

$$\alpha_t(j) = p(\mathbf{x}_t | z_t = j, \theta) \sum_{i=1}^s \alpha_{t-1}(i) \phi_{ij}, \quad (3.3.7)$$

where  $\alpha_1(j) = p(\mathbf{x}_1 | z_1 = j, \theta) \pi_j$  starts with the prior. It follows that

$$p(\mathbf{x} | \theta) = \sum_{i=1}^s \alpha_m(i), \quad (3.3.8)$$

where m is the number of observations.

### The Viterbi Algorithm (Special Case of the Max-Product Algorithm)

This algorithm specifies a way to obtain the most likely sequence of hidden states or latent variables. To find the most likely latent variable sequence  $\hat{\mathbf{z}} = \hat{z}_1, \hat{z}_2, \dots, \hat{z}_m$  given the  $\mathbf{x}$  and  $\theta$ , we may define

$$\delta_t(j) = \max_{z_1, z_2, \dots, z_{t-1}} p(z_1, z_2, \dots, z_t = j, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t | \theta) \quad (3.3.9)$$

and then by induction it follows that

$$\delta_{t+1}(j) = p(\mathbf{x}_{t+1} | z_{t+1} = j, \theta) \max_{1 \leq i \leq s} \delta_t(i) \phi_{ij}, \quad (3.3.10)$$

where  $\delta_1(j) = p(\mathbf{x}_1 | z_1 = j, \theta) \pi_j$ . Further, define

$$\psi_{t+1}(j) = \arg \max_{1 \leq i \leq s} \delta_t(i) \phi_{ij}, \quad (3.3.11)$$

then

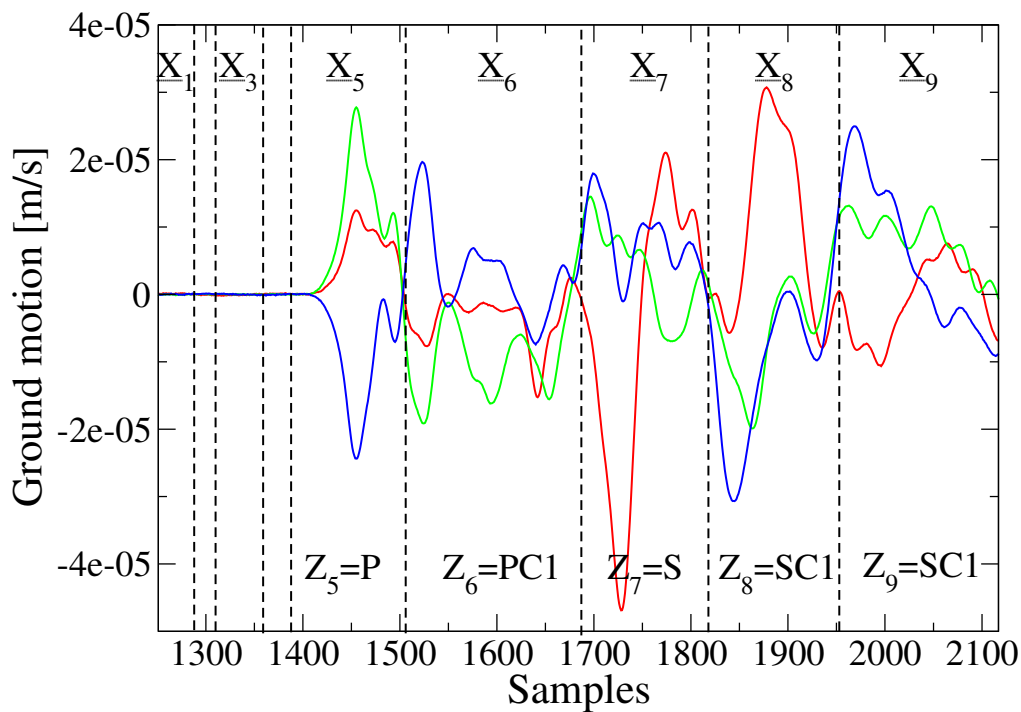
$$\hat{z}_t = \psi_{t+1}(\hat{z}_{t+1}). \quad (3.3.12)$$

The Viterbi algorithm is very similar to the forward algorithm with the  $\alpha_t$  replaced by  $\delta_t$ , i.e. the summation is replaced by maximisation. The Viterbi algorithm gives the mode of  $p(\mathbf{z} | \mathbf{x}, \theta)$ .



### 3.3.2 HMM in Mine Seismology

Although at first, it might not seem evident that HMM analysis can be applied to mine seismology, an HMM can describe our problem reasonably well. First of all, we observe data and want to know whether it originated from the rejected event-, blast- or seismic event model. Had the data been generated, for example, by the seismic event model, the onset of the P-wave and S-wave needs to be identified. The recorded wave can be partitioned into segments. The segments are not of fixed length, but are instead a function of the number of zero crossings in such a way that each segment must contain a few periods of the waveform as illustrated in Figure 3.2.



**Figure 3.2:** A waveform consisting of 866 samples, sampled at 6000 Hz. There are 9 segments with different numbers of samples. For example, the first segment consists of 34 samples and the sixth segment of 181 samples, with the length determined by the condition that the zero line must be crossed three times. Features  $\mathbf{x}_t$  are extracted for each segment  $t$  (see Subsection 3.3.3). The hidden states are also shown as described later in this subsection. For example, the 5th observation was generated by a P-wave hidden state ('P') and the last observation by the first S-wave coda hidden state ('SC1').

Note that the balls in the urn example are equivalent to the segments. For a seismic event model, the hidden states  $z_t$  are:

- a noise state (row 1 of the matrix shown in Equation 3.3.13),
- a P-wave onset state (row 2),
- P-wave coda state(s) (row 3),
- an S-wave onset state (row 4), and
- S-wave coda state(s) (row 5).

From previous experience, we assign a value 0.9 or larger to the probability that a noise state is followed by another noise state, with the remaining 0.1 probability being the probability that a noise state is followed by a P-wave onset state. From the physics, it is also known that the probability that a noise state is followed by one of the other states is 0, a P-wave onset state can be followed only by P-wave coda states, a P-wave coda state can be followed only by another P-wave coda state or by an S-wave onset state and further that an S-wave onset state can be followed only by an S-wave coda state and that an S-wave coda state can again be followed only by S-wave coda states or by a noise state. A simplified transition state matrix may hence take the form

$$\Phi = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.3 & 0 & 0 & 0 & 0.7 \end{bmatrix}, \quad (3.3.13)$$

where, from top to bottom, the states are noise, P-wave onset, P-wave coda, S-wave onset, S-wave coda respectively. In this dissertation, three different P-wave coda states and three S-wave coda states are used to model the seismic wave data. The initial state probabilities would typically look like  $\boldsymbol{\pi} = (0.95, 0.01, 0.01, 0.01, 0.01)$ , where the small probabilities assigned to non-noise states take account of the fact that de-triggering may fail at times, for instance during the signal of the seismic event.

A noise model is simpler than an event model as the noise state is the only hidden state. A blast model can also be built using an HMM, but a successful blast discriminator (as discussed in Section 3.2) was already developed by IMS in 2010 and hence will not be discussed here.

A question that remains is what observations (colour and shape of the balls in the urn example) should or could be used in mine seismology to classify recorded data and to determine accurate P-wave- and S-wave onsets. This is addressed in the next section.

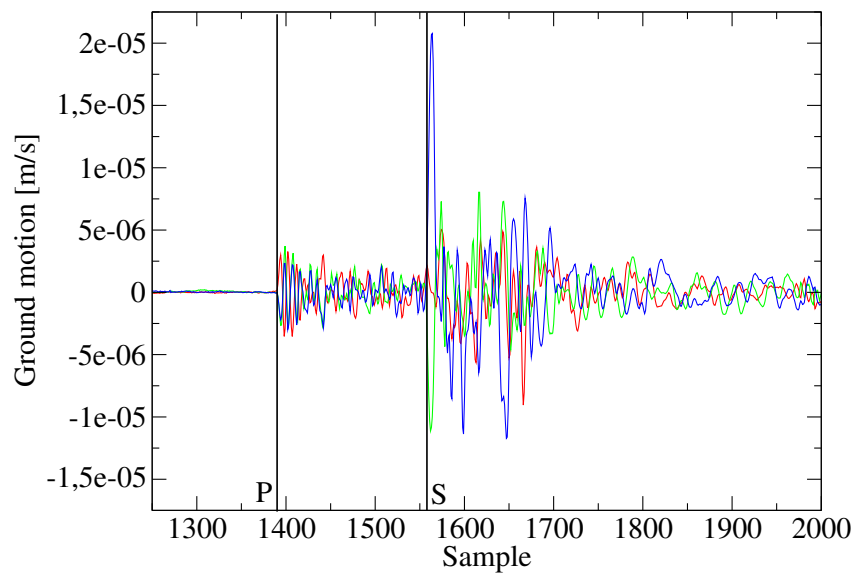
### 3.3.3 Feature Extraction, Observation Probabilities and Transition Matrix

In many cases, including the present mining situation, the practitioner is free to choose which of the many available observables to use. The choice of observation variables is critical because an HMM will be successful in classification and correctly determining the most likely hidden state only if these variables can successfully discriminate between models. To phrase it differently: given that the observations are purportedly realised from a specific type of event or model (reject, blast or seismic event), we want the probability of the observation sequence given for any alternative candidate model to be low. Fortunately, it is possible to make a good choice of observations which does discriminate between models and states by combining some basic physics principles with visual inspection.

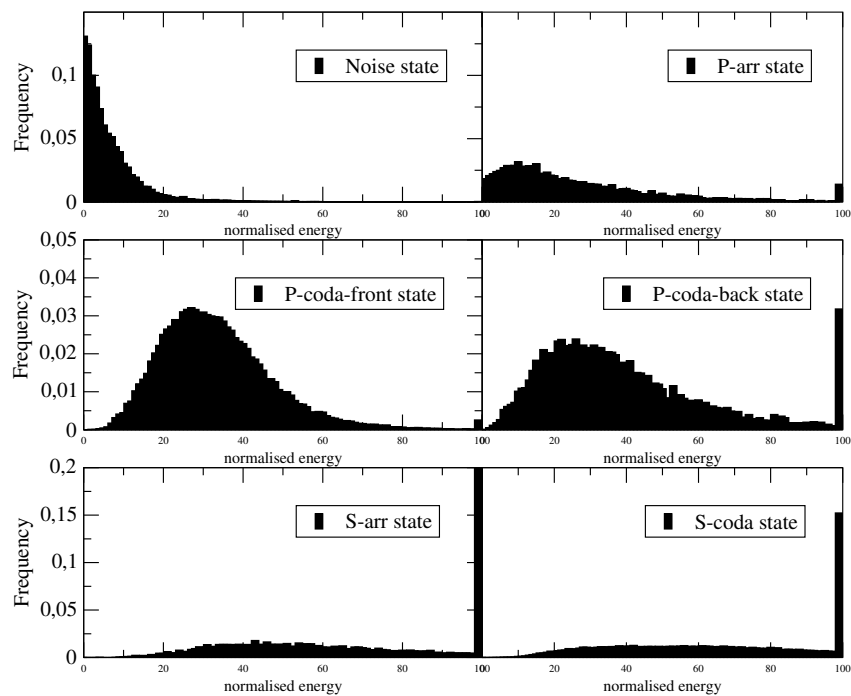
To obtain the observation probabilities for seismic events (or rejected events), a data training set processed by experienced QC processors is used. The P-wave and S-wave onsets are known. The noise states before the P-wave onset are also known, as are the P-wave coda states between the P-wave onset and S-wave onset. The S-wave coda is not exactly known, but can be estimated as the consecutive segments above the noise level following the S-wave onset. Therefore, from QC training data, all states are known or can be estimated. For instance, all known P-wave onset states for all seismic events in the training set are used in obtaining the histogram of the normalised energy. This histogram will be an estimate for the normalised energy given the P-wave onset state distribution. This can be applied for all models, all states and all observations to obtain the observation probabilities. However, it is only applied to the seismic event model as this model is the only one used in the classification (see Section 3.4). The initial state probabilities are easily obtained by calculating the percentage of time the first segment is a noise state or P-wave onset state. The non-zero elements in the transition probabilities can be calculated in the same way.

I now list and discuss the variables and the histograms used.

1. The **normalised energy** is an obvious choice for an observation. For the seismic event model, the energy observation will typically be the largest at or close to the S-wave onset state. The P-wave onset will tend to carry an energy  $E_P$  between 1/20th to 1/50th (Koyama [27]) of the S-wave energy  $E_S$  at onset; see Figure 3.3. For a clear event, the SNR will be high and for a rejected event (noise model) the SNR will tend to be low. Hence, the probability of the normalised energy (energy divided by maximum energy) being close to one for the noise state of the noise model is larger than that of the seismic event model. Figure 3.4 shows the histogram of the normalised energy feature given the different hidden states.

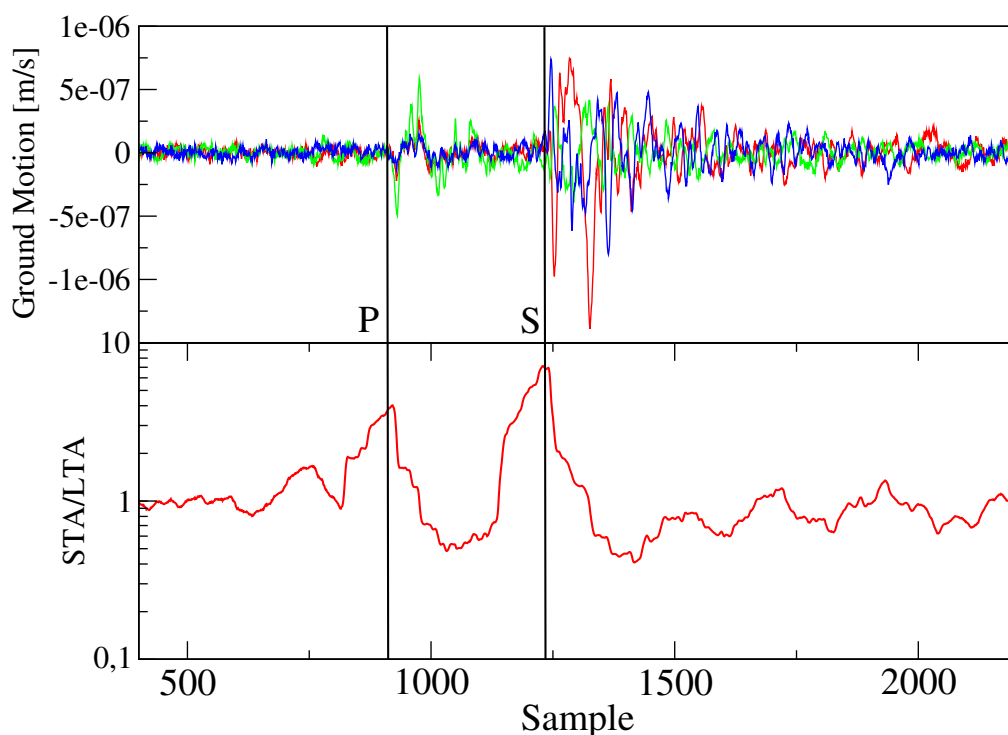


**Figure 3.3:** Typical events recorded at mines have larger S-wave energies than P-wave energies. In the above example the  $E_S/E_P$  ratio is 20, as implied by the much larger amplitude of ground motion in the S-wave. The normalised energy can thus be used as a feature: high values (0.8-1.0) tend to coincide with S-wave arrivals, while values between 0.05 and 0.5 tend to coincide with P-wave arrivals.

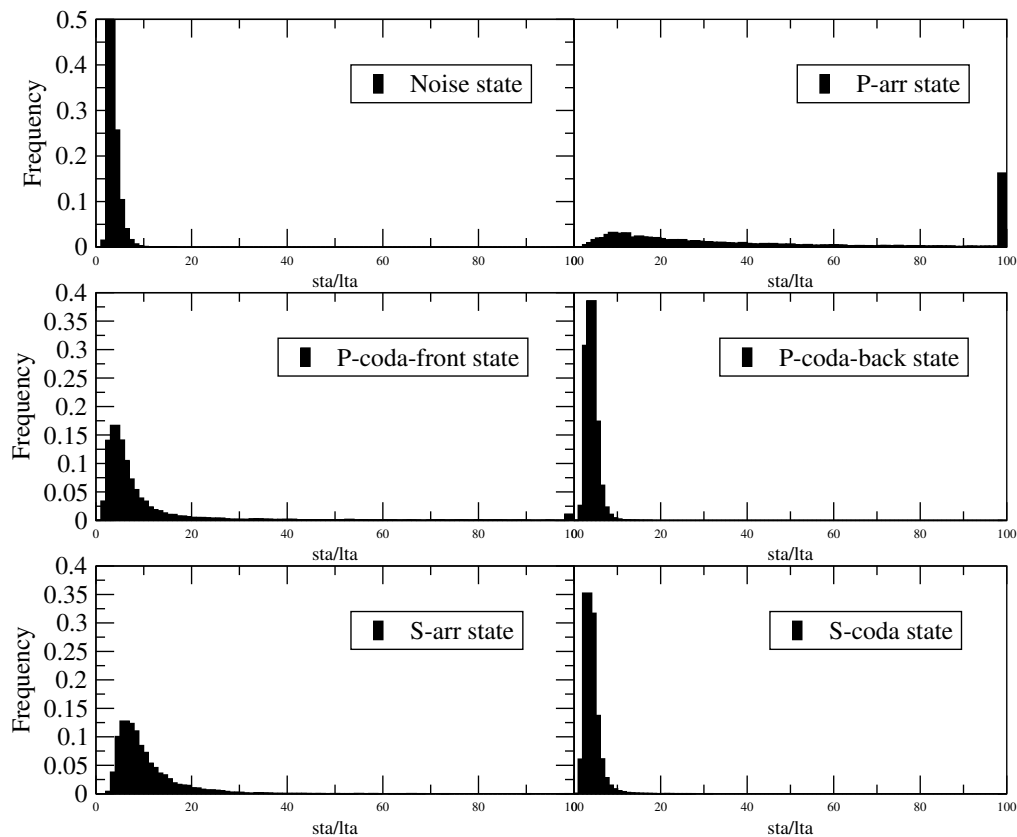


**Figure 3.4:** Histograms of the normalised energy given different hidden states. The distributions differ significantly in shape, which helps to minimise the misclassification error.

2. The next obvious observation is the ratio of short-term to long-term average of ground motion **STA/LTA**. The STA can be seen as the power of the wave taken over a short time-period, while the LTA is the power of the wave taken over a longer time-period as explained in Section 2.5. Note that the STA (as discussed in this Chapter) is taken over half a wave period and the LTA is taken over one wave period. For noise states the expected (or mode) value of the STA/LTA parameter will be close to 1, while for the S-wave onset the expected value will be larger than 1 and even larger for the P-wave onset state. For the first P-wave coda state the expected value should be smaller than 1, while for the second P-wave coda the expected value should be 1. For all the S-wave coda states the expected value of the STA/LTA value will again be smaller than 1. Figure 3.5 shows the STA/LTA values of a typical event. Figure 3.6 shows the histogram of the STA/LTA feature given the different hidden states.



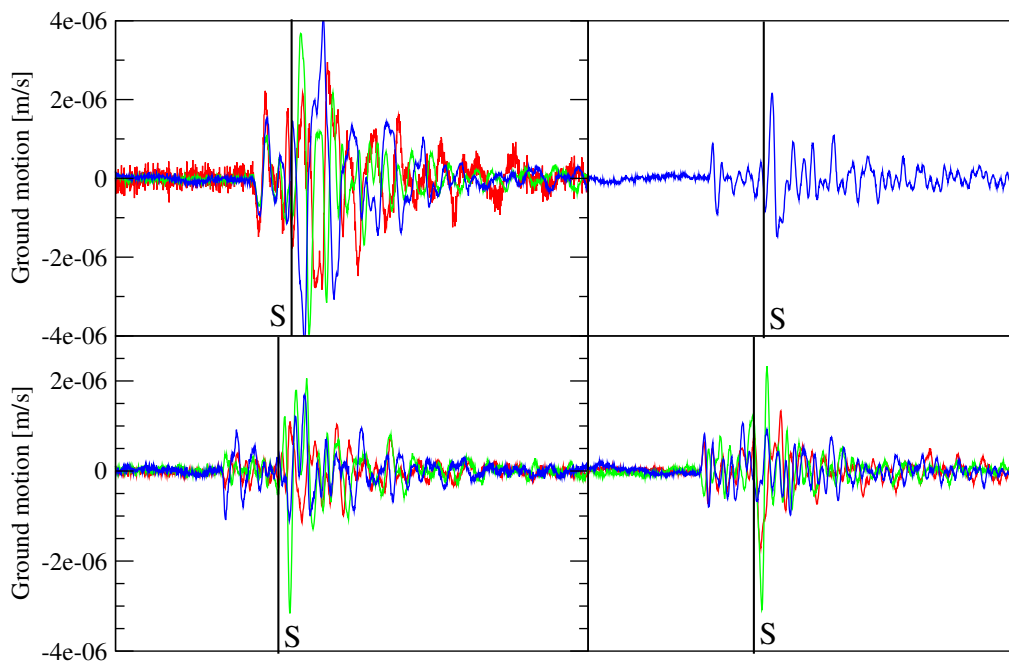
**Figure 3.5:** STA/LTA variable as a function of sample (time). Clearly the STA/LTA graph has maxima close to the P-wave and S-wave onsets; note the logarithmic scale. In the noise, the STA/LTA values are close to 1, while just after the P-wave (first P-wave coda state) the values drop below 1. Most of the S-wave coda states are also below 1 and then slowly revert to the noise state with STA/LTA 1.



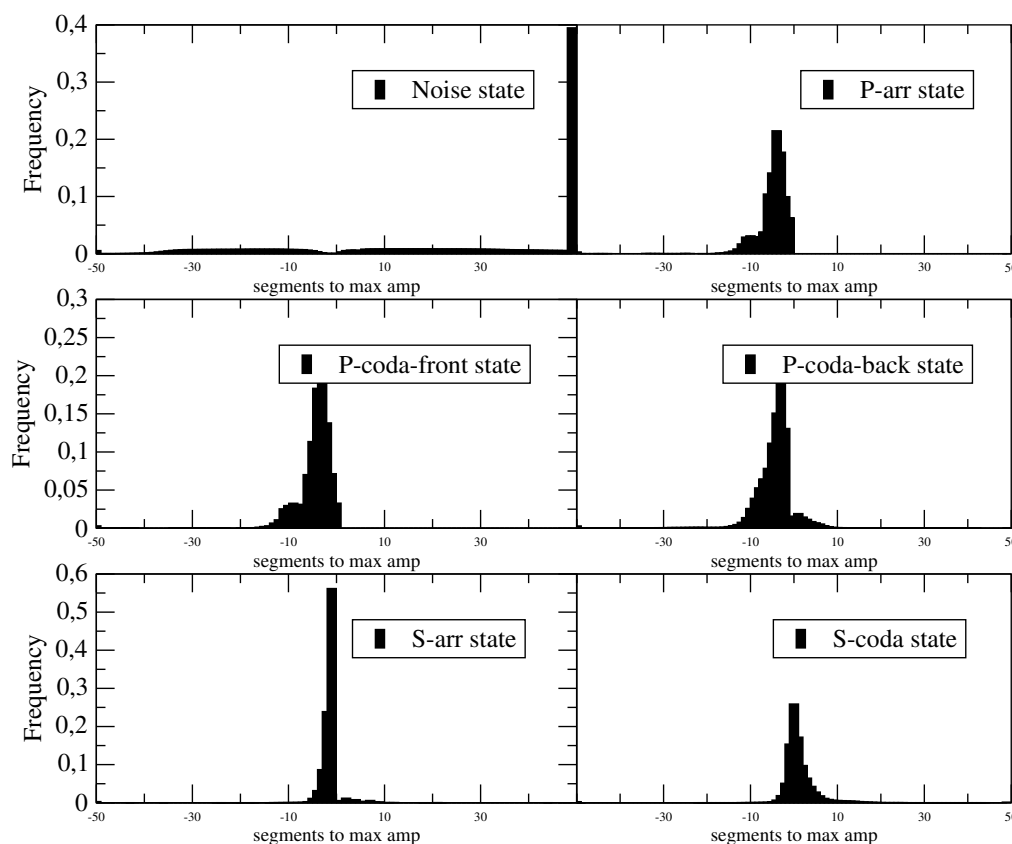
**Figure 3.6:** Histograms of the STA/LTA variable given different hidden states. This observable distinguishes among different hidden states.

- 3. Time difference between segments:** Ignoring regional seismicity and neighbouring mines, the maximum separation of the P-wave arrival and the S-wave arrival is determined by the size of the mine. It will be advantageous to incorporate this fact into the HMM. However, the P-wave onset and the S-wave onset are hidden states and it is therefore not possible to use the time separation of the P-wave and the S-wave directly as an observation. In order to capture some waveform before the P-wave and also to capture the S-wave coda, a typical seismogram must be longer than the maximum possible time separation of the P-wave and the S-wave (given that the seismic event occurred on the mine). Hence, there is a large probability that the HMM mis-identified the true P-wave and/or S-wave in cases where the HMM classified the most likely P-wave state early in the buffer and the S-wave state late in the buffer. The maximum amplitude is typically inside the segment of the S-wave arrival; therefore an observation that may be used is the difference (in samples or in time) between the current segment and the segment of the maximum amplitude; see Figure 3.7. Segments that are close to the

segment with the maximum amplitude are more likely to contain the P-wave than segments further away from it. This might not be an obvious choice for an observation, but the probability distributions of this observation given the different states, differ significantly for the different hidden states. Figure 3.8 shows the histogram of the time difference between segments feature given the different hidden states.



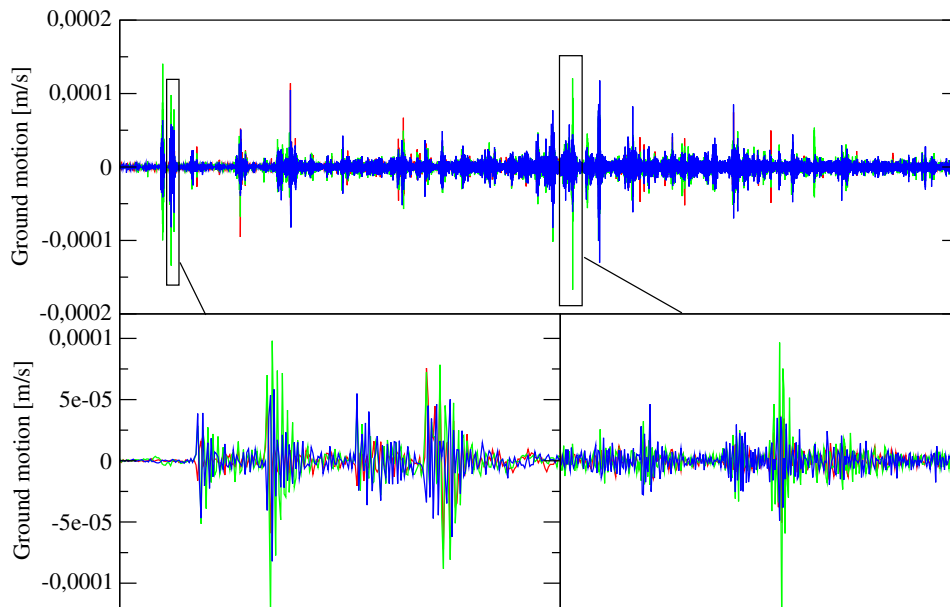
**Figure 3.7:** The maximum amplitude tends to be within the S-wave onset segment or at least close to it. The distance to the maximum amplitude is an important feature. Multiple events in a buffer or events where P-wave onsets have the maximum amplitude occur frequently enough, so that these events are not classified incorrectly as reject model data.



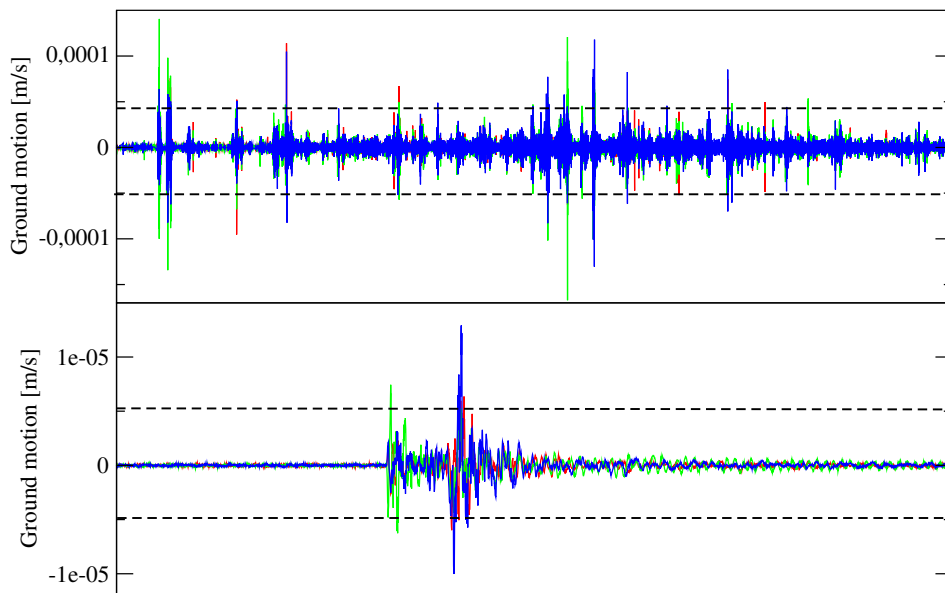
**Figure 3.8:** Histograms of distance to maximum amplitude variable for the different hidden states. It is clear that the distributions differ, which is preferable for minimising the misclassification error.

4. **The number of spikes.** The last observation is to use the number of spikes within a segment to determine whether a seismic event has been recorded or not. In Figure 3.9, the top panel shows two typical situations, parts of which have been enlarged in the lower two panels. In the left-hand panels, the seismic event is “hidden” in the noise (see Figure 3.9) and should not be classified as a seismic event model, while the right-hand panels show a clear signal-to-noise ratio (SNR) above 1 that should be classified as a seismic event model. The distinction is made on the basis of the number of spikes, i.e. the number of segments with a maximum amplitude larger than some fraction of the global maximum amplitude. For an event to be clearly seismic, the P-wave onset should have zero spikes, the P-wave coda states also zero spikes, the S-wave arrival one spike and lastly S-wave coda states a few spikes. For a noisy event the number of spikes will be much higher than the clear event; see Figure 3.10. Figure 3.11 shows the histogram of the number of spikes feature given the different hidden states.

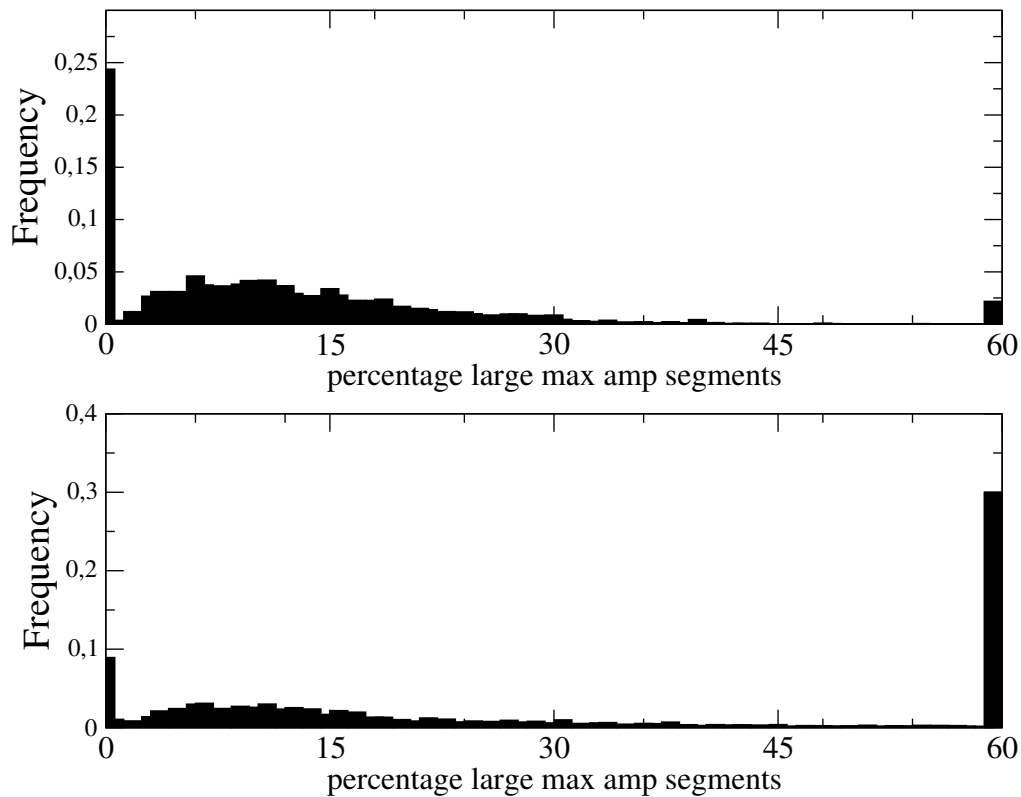




**Figure 3.9:** Multiple events in a buffer. Not a single event stands out. These types of seismograms belong to the reject model. The zoomed-in part indicates that there are seismic events in the recorded data. The number of segments with large amplitude features will distinguish between this type of data and data being part of the seismic event model.



**Figure 3.10:** The number of segments with a large amplitude feature will distinguish between reject model data, as shown in the top panel, seismic event model data shown in the bottom one.



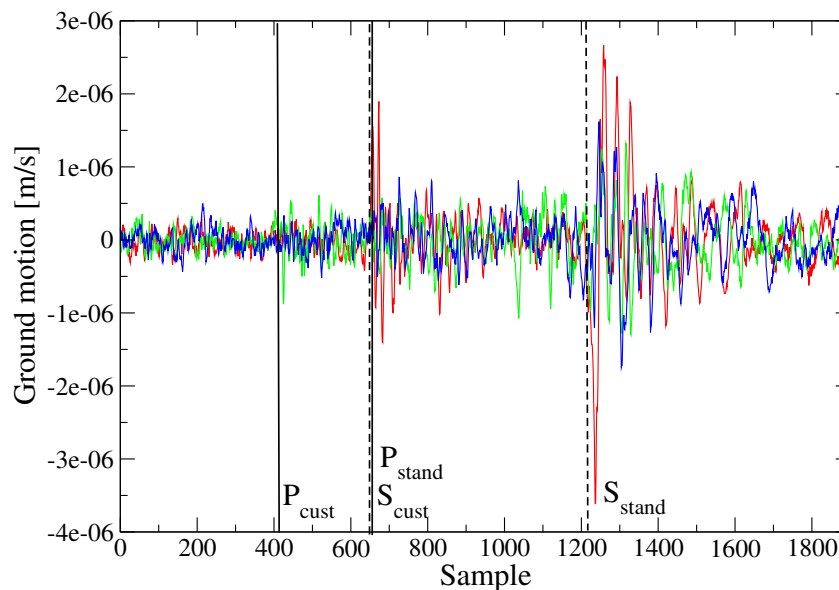
**Figure 3.11:** Histograms of the percentage of segments with large maximum amplitudes of seismic events (top) and rejected events (bottom). There is a significant difference between these distributions, indicating that this variable can be used to differentiate between a reject and a seismic event model.

### 3.3.4 HMM Using a Customised Viterbi Algorithm

An HMM cannot implicitly model the direction of local particle displacement of the P-wave and the S-wave. This is problematic, because according to the HMM, the most likely P-wave onset state and S-wave onset state may, in fact, both be P-wave onsets or both S-wave onsets. For waveforms recorded by uni-axial sensors, this is not relevant as it is impossible to distinguish between P-wave onsets and S-wave onsets using the direction of local particle displacement. On tri-axial waveforms, when the Viterbi algorithm decides on the most likely hidden state sequence, one can at least identify whether the candidate P-wave and S-wave patches/states are orthogonal to each other. This is not ideal, but given the time constraints of the classification algorithm, the HMM using the customised Viterbi algorithm, as described in this section, at least models local displacement of the P-wave and the S-wave in some sense.

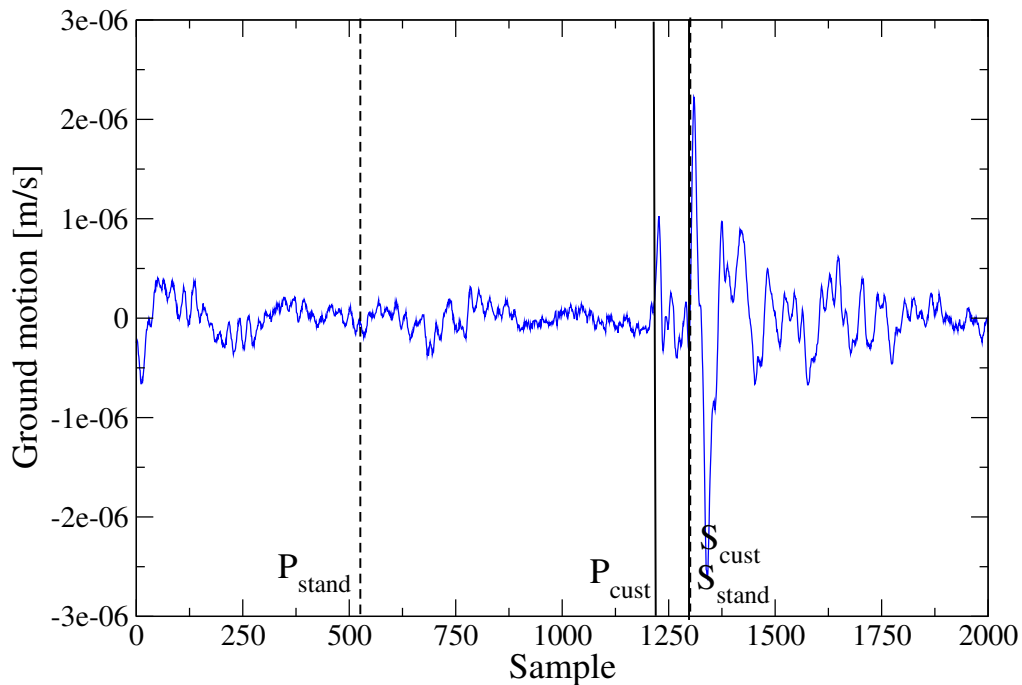
The “sanity checks” listed below are added to the standard Viterbi algorithm, forming, what I call, the customised Viterbi algorithm. Probabilistic graphical models can model these features explicitly, but unfortunately take too long to run. However, in the future, with faster computers being available, waveforms may be modelled using a more realistic PGM. In this dissertation, this aspect is not considered.

1. The customised Viterbi algorithm adjusts the joint observation probability by significantly lowering the observation probability given the S-wave onset state when the directivity of particle motion of the possible S-wave onset waveform is not orthogonal to the currently identified most likely P-wave onset. Hence, no P-wave onset state and S-wave onset state will have parallel directivity of particle motion. This scheme will work only if the most likely P-wave onset state was indeed the true P-wave onset. As a result, when the incorrect P-wave onset was identified, no S-wave onset state will be identified. Consequently, the data may be misinterpreted as consisting of noise states only. In such cases the HMM is run again with a significantly lower probability of a P-wave onset at the segment where it was incorrectly identified in the previous run (see Figure 3.12).



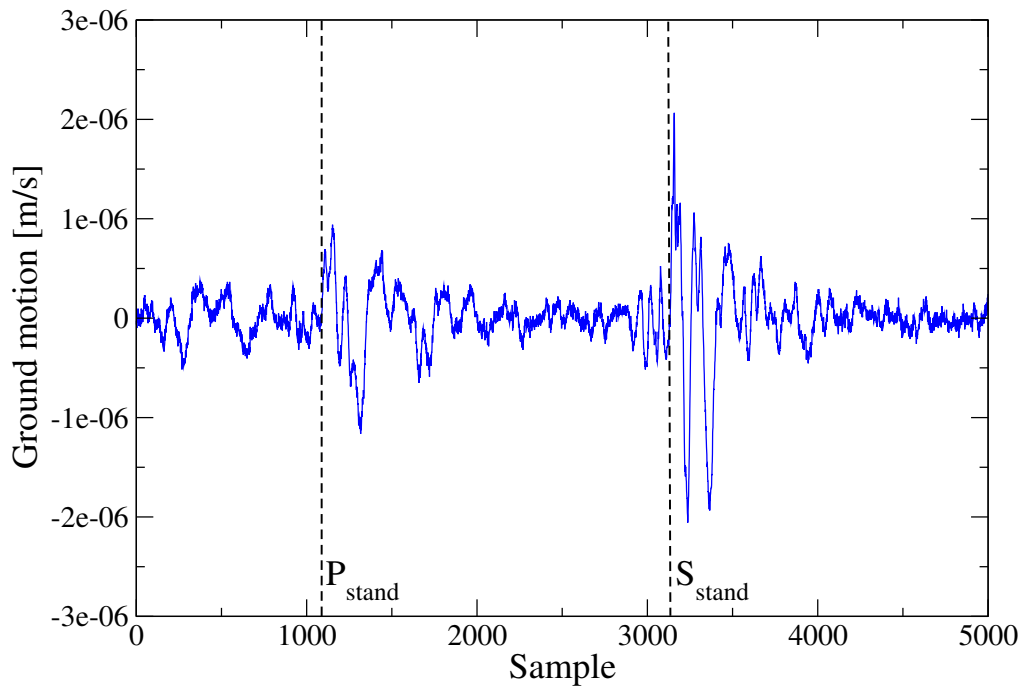
**Figure 3.12:** Picks ( $P_{stand}$  and  $S_{stand}$ ) of the HMM using the standard Viterbi algorithm and picks ( $P_{cust}$  and  $S_{cust}$ ) of the HMM using the customised Viterbi algorithm. When the orthogonal directivity of the P-wave onset and S-wave onset check is not built into the Viterbi algorithm, then the picks may be wrong, as signified in the above example. The correct P-wave arrival is at sample 405 and the correct S-wave arrival is at sample 653 as the customised Viterbi decoding algorithm model suggests.

2. Vibrations in segments of P-wave coda should be mutually parallel, as should vibrations in S-wave coda segments. A segment can hence be classified as a P-wave coda (or S-wave coda) only if its vibrations are parallel to the those of previous wave segment. This is again achieved by lowering the observation probability given the P-wave coda (S-wave coda) state when the directivity of particle motion of the possible P-wave coda (S-wave coda) waveform is orthogonal to the currently identified most likely P-wave (S-wave) onset.
3. Another useful “sanity check” is that the physics of seismic waves stipulates that there should be no significant change in the dominant period of oscillation. When consecutive segments have a significant difference, which are not both noise states, then the observation probabilities given all non-noise states are significantly lowered. An example is shown in Figure 3.13.



**Figure 3.13:** Picks ( $P_{\text{stand}}$  and  $S_{\text{stand}}$ ) of the HMM using the standard Viterbi algorithm and picks ( $P_{\text{cust}}$  and  $S_{\text{cust}}$ ) of the HMM using the customised Viterbi algorithm. This example shows that the picks may be wrong if the frequency of the P-wave onset and S-wave onset is not built into the Viterbi algorithm. The correct P-wave arrival is at sample 1220 and the correct S-wave arrival is at sample 1300 as the HMM using the customised Viterbi algorithm suggests. The standard Viterbi decoding algorithm gives a P-arrival onset estimate which is incorrect, as it is clearly visible that it has the same frequency as the background noise.

4. The last check is the enforcement of a maximum allowable distance between the most likely P- and S-wave onset states. Even though this is already a feature of the HMM ('Time difference between segments'), it fails at times and the identified P- and S-wave onset states are too far apart. Figure 3.14 shows an example where the maximum allowable distance between the most likely P- and S-wave onset states ensures identification of correct hidden states.



**Figure 3.14:** Picks ( $P_{\text{stand}}$  and  $S_{\text{stand}}$ ) of the HMM using the standard Viterbi algorithm. When the maximum separation (in this example 800 samples or 0.133 s) of the P-wave onset and S-wave onset is enforced, events with false arrivals are more likely to be rejected. The HMM using the Viterbi algorithm will classify this as a seismic event model, while the HMM using the customised Viterbi decoding algorithm correctly rejects this seismogram.

### 3.4 Classification of Recorded Data

A classification scheme can be derived from the parameter estimates  $\theta_a = (\Phi_a, \pi_a)$  and  $\theta_r = (\Phi_r, \pi_r)$ , where the subscripts  $a$  and  $r$  indicate respectively an accepted seismic event model and a rejected noise event model, where the acceptance and rejection criteria are obtained from the training data set. The forward algorithm is used to obtain  $p(\mathbf{x}_i|\theta)$  for each sensor  $i$  for data recorded at  $n$  sensors. Note that a subscript is used for each sensor, so  $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im_i})$  is the observed data segment sequence for sensor  $i$  and  $m_i$  is the number of observations of sensor  $i$ .

To derive a classification scheme for data recorded at multiple sensors, a technique first needs to be devised to classify a recorded seismogram as produced by a rejected noise model or produced by an accepted seismic event model. It was initially thought that it is adequate simply to use the standard HMM technique for classifying a seismogram as noise if and when the likelihood based on the rejection model exceeds that of the acceptance one, i.e.  $p(\mathbf{x}_i|\boldsymbol{\theta}_r) > p(\mathbf{x}_i|\boldsymbol{\theta}_a)$ . This simple approach does not work well, however. The size of the space of possible noise states is very large compared to the size of the space for accept states, even though the individual probability of a noise state may be much smaller than that of an accept state. For a typical seismogram, the seismic signal forms less than 10% of recorded waveform. For the accept model, we are of course more interested in the non-noise states of the waveform, so the noise states needs to be ignored. On the other hand, totally ignoring the rest of the seismogram is not optimal, as low probability noise states (the most likely states though) should preferably be part of the noise model.

The implemented classification scheme uses the highest average probability of the  $u - l$  consecutive segments of non-noise states and a few representative noise states with lowest probabilities both within the accept model, where  $l$  is the index of the segment of the P-wave onset state and  $u$  is the index of the segment of the last S-wave coda state. The probability of segments that are used for classification for seismogram  $i$  can be written as

$$\psi_i^{\text{high}} = \max_{u,l} \left\{ I(\hat{z}_{i,l-1} = l_\epsilon) I(\hat{z}_{i,u+1} = l_\epsilon) \frac{1}{u-l+1} \sum_{k=l}^u (1 - I(\hat{z}_{ik} = l_\epsilon)) \ln p(\mathbf{x}_{ik}|\hat{z}_{ik} = l_k, \boldsymbol{\theta}_a) \phi_{l_{k-1}l_k} \right\}, \quad (3.4.1)$$

where  $l_\epsilon$  indicate a noise state,  $l_k$  indicate the most likely state of the  $k$ th segment,  $\hat{z}_{ik}$  is the most likely state of the  $k$ th segment taken over the consecutive segments of highest average probability for non-noise states and  $\psi_i^{\text{low}}$  is a low (typically 2.5th) percentile of the noise state probabilities. At first glance, the equation for  $\psi_i^{\text{high}}$  might not seem to be the maximum average probability taken over consecutive non-noise states, but the reader can verify that this is, in fact, the case.

The smaller  $\psi_i^{\text{high}}$  and  $\psi_i^{\text{low}}$  are, the more likely it is that the seismogram was generated by a noise model. Note that only the accept event model is used in the classification scheme. A seismogram is classified as being generated from an accept event model if both the high-probability exceeds an upper threshold and the low-probability exceeds a separate lower threshold,

$$(\psi_i^{\text{high}} > \psi_{\text{threshold}}^{\text{high}}) \cap (\psi_i^{\text{low}} > \psi_{\text{threshold}}^{\text{low}}), \quad (3.4.2)$$

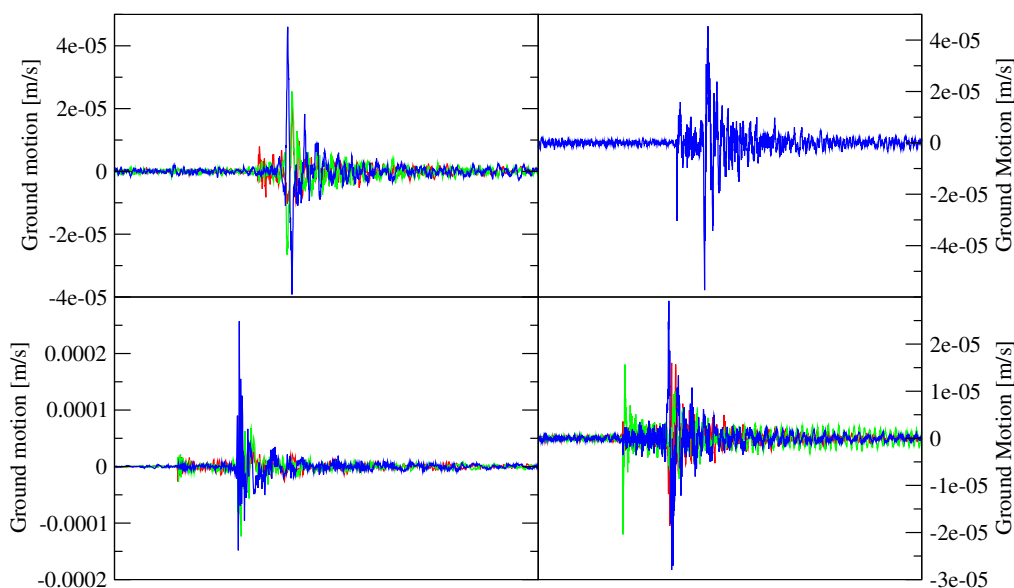
otherwise it is classified as being generated from a noise model. In the case of multiple sensors (seismograms), recorded data is classified as an accepted event model if the number of classified accepted seismograms is larger than some pre-

defined threshold  $\psi_{\text{threshold}}$ , that is

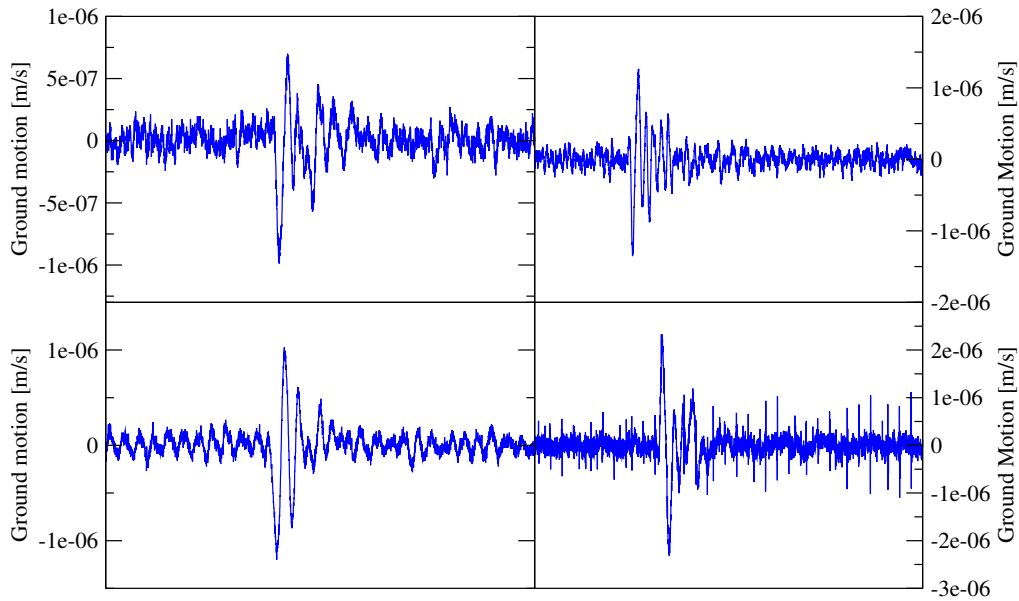
$$\sum_{i=1}^n I \left[ (\psi_i^{\text{high}} > \psi_{\text{threshold}}^{\text{high}}) \cap (\psi_i^{\text{low}} > \psi_{\text{threshold}}^{\text{low}}) \right] > \psi_{\text{threshold}}, \quad (3.4.3)$$

where  $\psi_{\text{threshold}}$  is larger than 3 and smaller than  $n$ . The exact value of  $\psi_{\text{threshold}}$  depends on the accuracy of the classification or on how sensitive we are to making rejection errors.

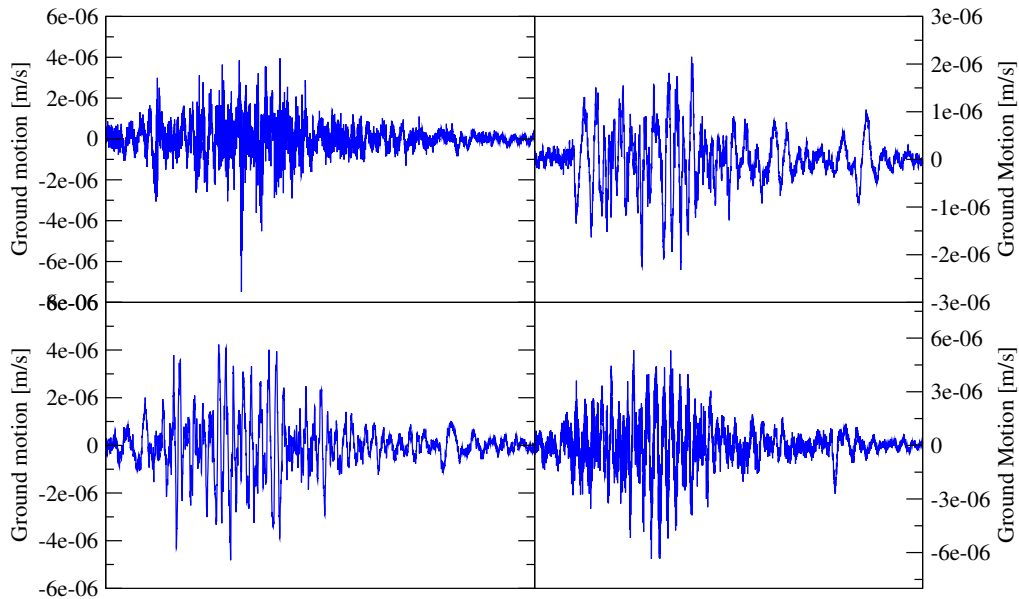
I now present some graphical illustrations of the above considerations and criteria. In Figure 3.15 an example of a recorded event is shown that is safely classified as an accepted event model, while Figures 3.16 and 3.17 show events that are comfortably classified as reject event models. In Figure 3.20 an example of a recorded event is shown that barely passed the classification rule of an accepted event model, while Figure 3.22 shows an event that just failed the rule of an accepted event model. Figures 3.18 and 3.21 show examples of recorded events that are incorrectly classified as an accept event model, while Figure 3.19 shows an event that is incorrectly classified as a reject event model. In Figure 3.23 a 3D histogram is shown of  $\psi_i^{\text{high}}$  and  $\psi_i^{\text{low}}$  for seismograms of a test set as classified by experienced processors for two mines.



**Figure 3.15:** Example of a recorded event showing four seismograms recorded by geophones that was correctly classified as an accept event model. The median seismogram had  $\psi_i^{\text{high}} = -8.47$  and  $\psi_i^{\text{low}} = -13.84$ . A seismogram is classified as being generated by the accept event model if  $\psi_i^{\text{high}} > -12 \cap \psi_i^{\text{low}} > -14$ .

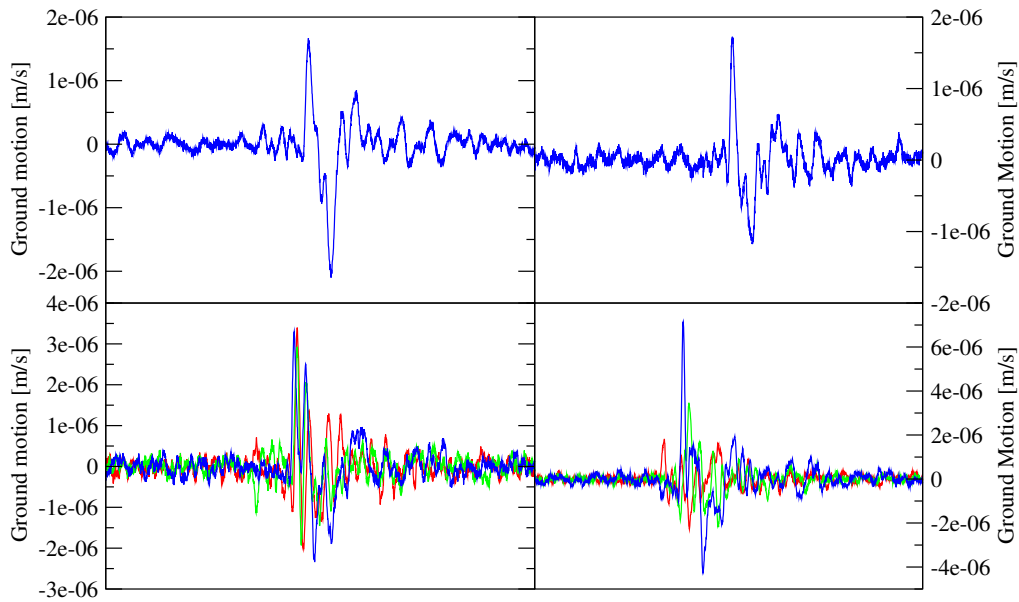


**Figure 3.16:** Example I of a recorded event showing four seismograms recorded by geophones that was correctly classified as a reject event model. There are no P-wave arrivals and these signals were generated by an unknown source. The median seismogram had  $\psi_i^{\text{high}} = -23.1$  and  $\psi_i^{\text{low}} = -13.84$ .

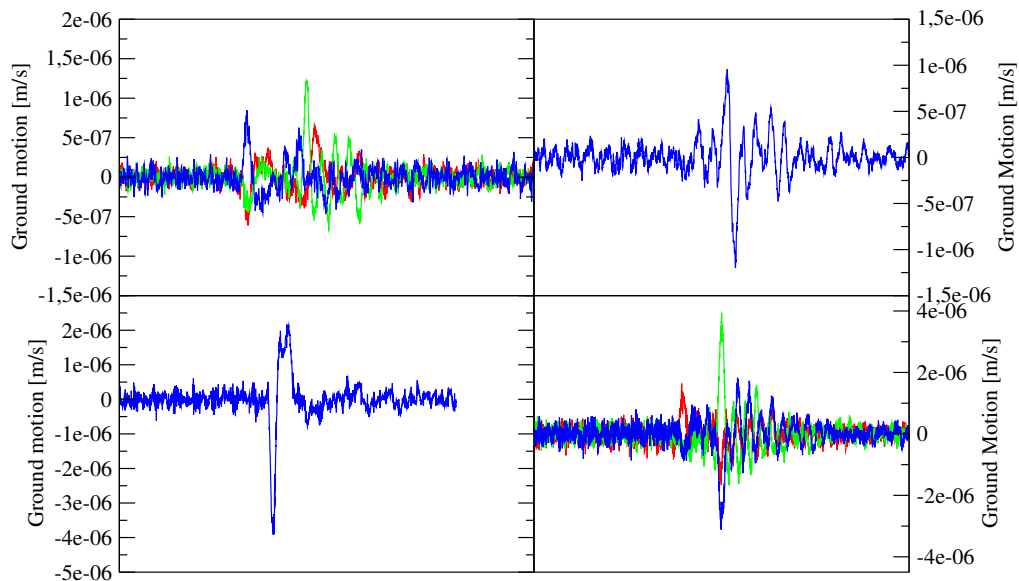


**Figure 3.17:** Example II of a recorded event showing four seismograms recorded by geophones that was correctly classified as a reject event model. The median seismogram had  $\psi_i^{\text{high}} = -18.73$  and  $\psi_i^{\text{low}} = -14.1$ . A seismogram is classified as being generated by the accept event model if  $\psi_i^{\text{high}} > -12 \cap \psi_i^{\text{low}} > -14$ .

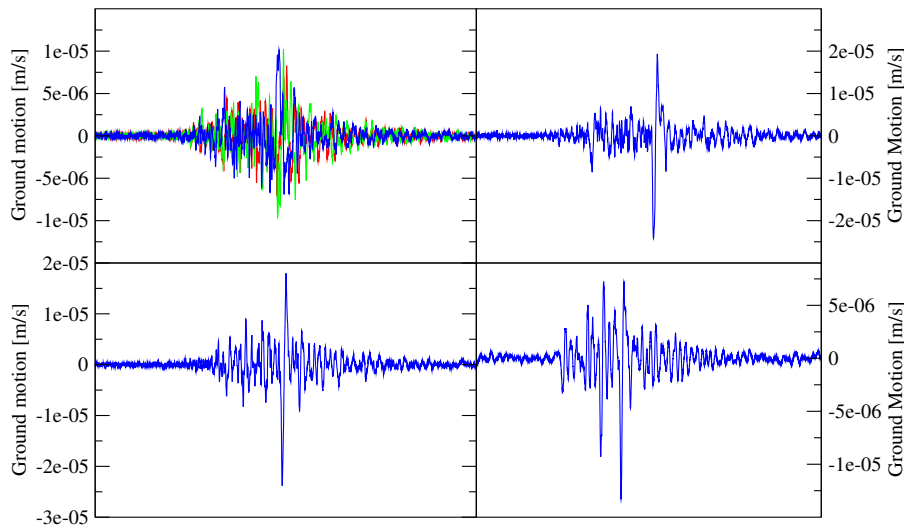




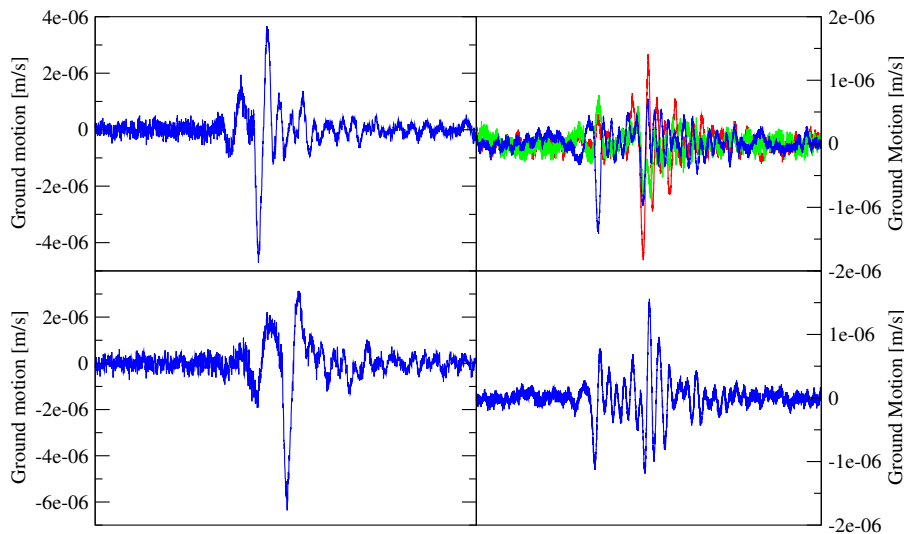
**Figure 3.18:** Example of a recorded event showing four seismograms (upper panel uni-axials and lower panel tri-axials) recorded by geophones that was correctly classified as an accept event model, but only just. The median seismogram had  $\psi_i^{\text{high}} = -11.32$  and  $\psi_i^{\text{low}} = -13.84$ . The two tri-axial seismograms have P-wave arrivals, but the uni-axials have no clear P-wave arrivals.



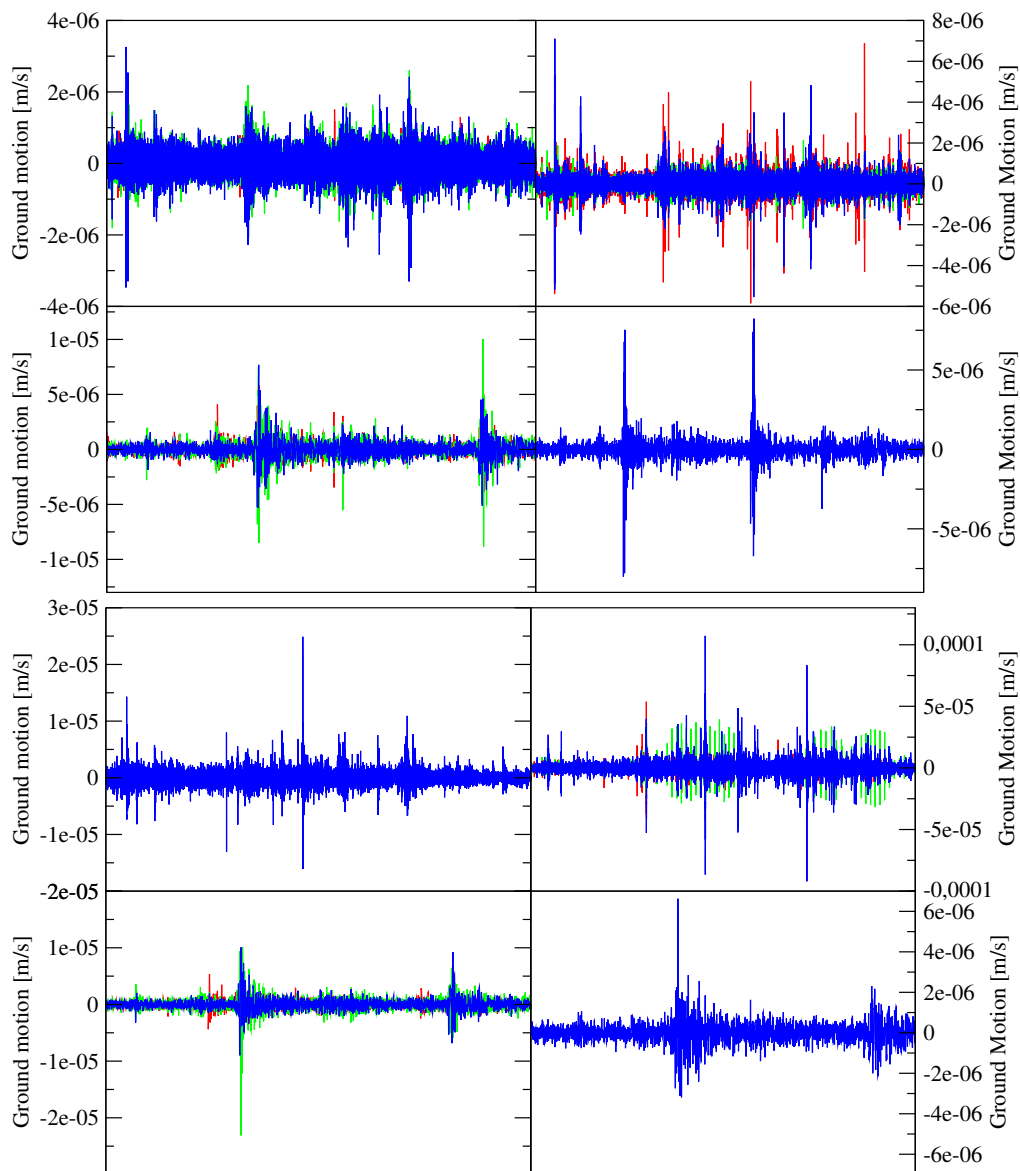
**Figure 3.19:** Example of a recorded event showing four seismograms (two uni-axials and two tri-axials) that was correctly classified as a reject event model, but only just. The median seismogram had  $\psi_i^{\text{high}} = -12.68$  and  $\psi_i^{\text{low}} = -13.84$ . The tri-axial signals seem to have P-wave and S-wave arrivals, but the signal-to-noise ratio is low. These signals are most probably ore pass noise.



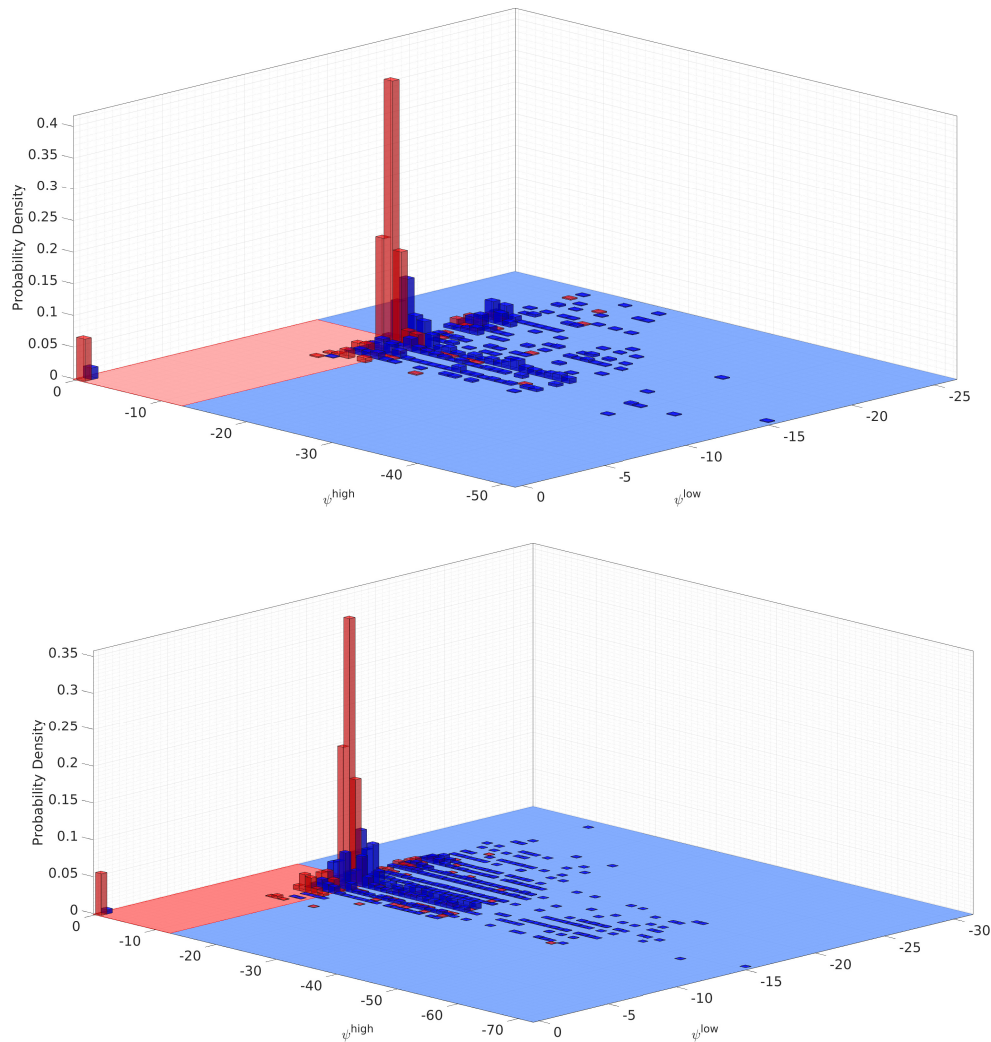
**Figure 3.20:** Example I of a recorded noise event which was incorrectly classified, based on the appearance of the three uni-axial seismograms shown in blue, as an accept event model. The multicoloured tri-axial seismogram shows that this is not a seismic event. Recorded events such as this one are uncommon and would give inexperienced human data processors difficulty also. The median seismogram had  $\psi_i^{\text{high}} = -11.74$  and  $\psi_i^{\text{low}} = -13.84$ .



**Figure 3.21:** Example II of a recorded event showing four seismograms (three uni-axials and one tri-axial) that was incorrectly classified as an accept event model. The median seismogram had  $\psi_i^{\text{high}} = -10.48$  and  $\psi_i^{\text{low}} = -13.84$ . The automatic classifier was relatively “sure” that it is a seismic event, but it was found to be noise. The dominant frequency is unnatural (too low) for the ground motion. These seismograms probably signify ore pass noise. Different event models based on dominant frequency might be built to solve these types of mis-classification. In this dissertation however, the post-pick classifier (see Section 4.9) will reject these types of events. The proposed classification scheme works relatively well, but makes errors as in this example.



**Figure 3.22:** Example of a recorded event that was incorrectly classified as a reject event model. There are a few tri-axial components (seismograms 3 and 7) that are clear accepts. However, there are a few very noisy seismograms (seismograms 1 at top left, 2 at top right, 5 and 6). Seismograms 4 and 8 (bottom right) are ambiguous and can be accepted or rejected. The human processor accepted (processed) seismograms 3,7,4 and 8. The median seismogram had  $\psi_i^{\text{high}} = -9.01$  and  $\psi_i^{\text{low}} = -19.94$ . A seismogram is classified as being generated by the accept event model if  $\psi_i^{\text{high}} > -12 \cap \psi_i^{\text{low}} > -14$ . The automatic classifier identified a clear seismic event ( $\psi_i^{\text{high}} = -9.01$ ), but it was rejected because the classifier thought it was too noisy. These types of mistakes are not serious though, as other manual processors might also have rejected this event.

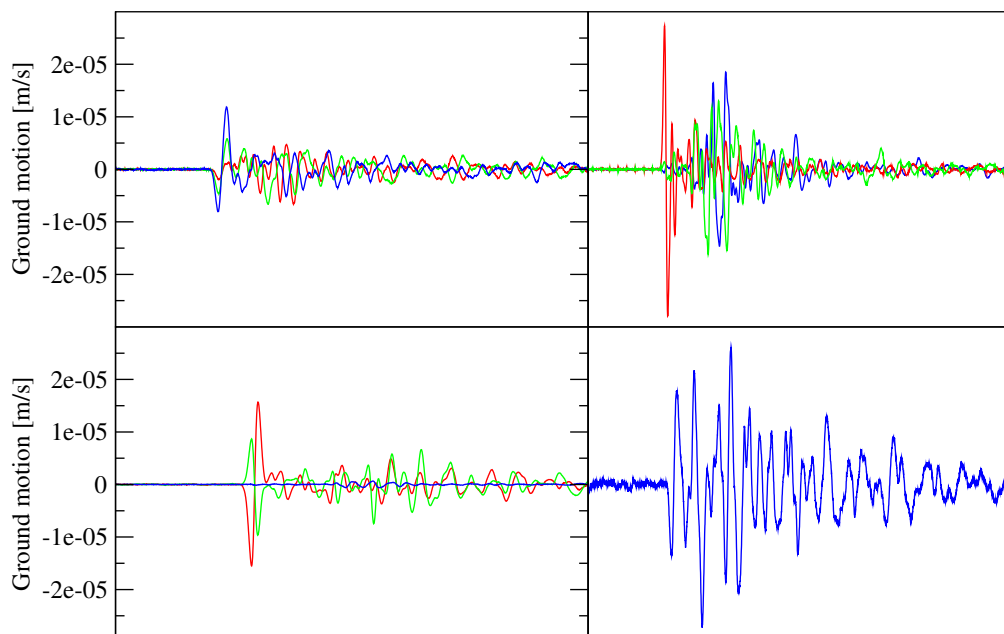


**Figure 3.23:** Histogram of  $\psi_i^{\text{high}}$  and  $\psi_i^{\text{low}}$  on a testing data set of accepted events (red bars) and rejected events (blue bars) for Mine1 (top) and Mine2 (bottom). The numerical value of the  $(\psi_i^{\text{high}}, \psi_i^{\text{low}})$  pair is obtained from the geophone of the median  $\psi_i^{\text{high}}$ . The areas shaded red, defined by  $(\psi_i^{\text{high}} > -12) \cap (\psi_i^{\text{low}} > -14)$ , contain events classified as accepted and the areas shaded blue contain events classified as rejected. This classification scheme has an error rate for accepted events of less than 12% for Mine1 and less than 10% for Mine2.

### 3.5 Summary

In this chapter an automatic classification scheme was discussed. The PGM used to explain the seismic waves captures some important features of seismic waves. The HMM using the standard Viterbi decoding algorithm fails in modelling some important features such as the orthogonality of the directivity of particle motion between P- and S-waves. An HMM using a customised Viterbi algorithm and classification scheme is proposed which addresses these shortcomings and indirectly models these critical features. This HMM performs relatively well, but fails in some visually obvious examples. The Viterbi algorithm allows for the calculation of the most likely states.

In general, an HMM is a natural way of modelling waveforms generated in the mining environment. However, there are some types of events, for instance pillar failures, where the P-wave onset generates most of the energy (see Figure 3.24) and where the waveform differs from a typical shear event. Because the seismic event model assigns a low probability to such observations, these type of events may be classified as a reject model. One can easily build a pillar failure type event model to overcome this problem. In essence, an HMM will work only if the variability of seismic events is not too large.



**Figure 3.24:** Example of a seismic event with stronger P-wave arrivals than S-wave arrivals. In some cases this type of event may be classified as a reject model. One could build a strong P seismic event model to overcome this misclassification. Note, that I did not build such a model in this dissertation.

---

Other features such as the shape of the sensor's amplitude response, signal kurtosis, rectilinearity and planarity should improve the performance of the customised HMM and this is future work.

The most likely P- and S-wave onset states of a waveform as generated by a seismic event can be used, along with other methods, as a rough estimate of the phase arrival distribution (Tarantola and Valette [22]). These initial rough estimates of the phase arrivals play a crucial role in determining a first estimate of the origin time and hypocentre for the seismic event. The hypocentre time and location estimates are also used for updating the distribution of the phase arrivals. Chapter 4 addresses this.



# Chapter 4

## PGM of Phase Arrivals

### 4.1 Model construction

Once recorded data has been classified as a seismic event, the second task of a data processor is to determine the phase arrivals of the P- and S-wave. When processing is done automatically, determining these phase arrivals is not simple. How to determine phase arrivals using machine-learning techniques is the topic of this chapter. Unlike earlier work, I do so by joint multivariate analysis of multiple sensors and their seismograms.

This section is structured as follows: Subsection 4.1.1 shows why multi-sensor analysis is important. Subsection 4.1.2 deals with problems related to automatic detection of phase arrivals and goes into more detail as how to solve these problems; Subsection 4.1.3 describes how phase arrivals are detected when information contained in a single seismogram is used; Subsection 4.1.4 discusses Bayesian networks; in Subsection 4.1.5 the parameters, data and latent variables of our problem are discussed and lastly, Subsection 4.1.6 carries out a literature review of previous studies.

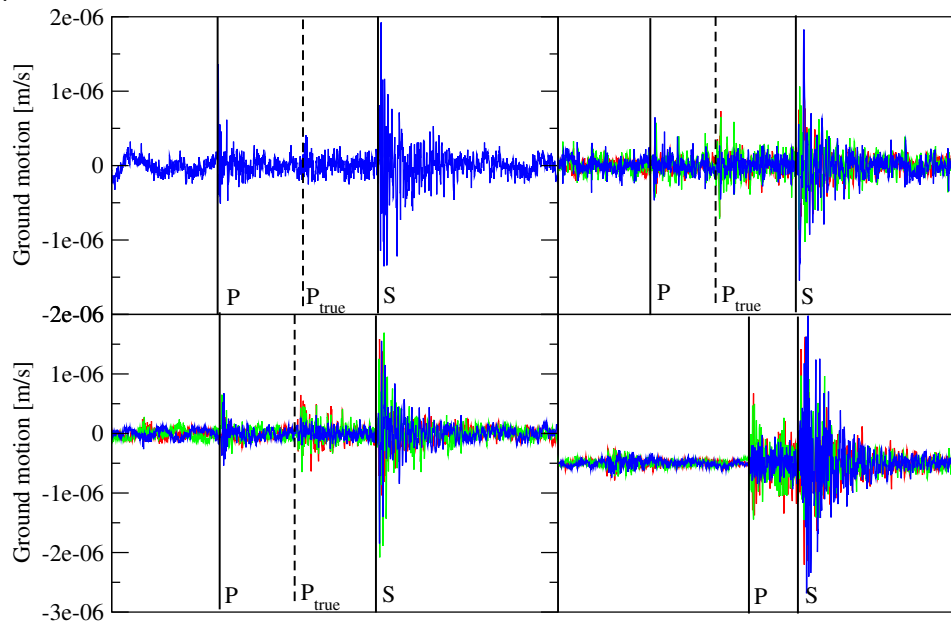
Section 4.2 gives the specific Bayesian network used to solve the problem of the phase arrival detection; the factorisation of the PGM and other relevant conditional distributions are given and derived in Section 4.3; Section 4.4 shows how to obtain posterior distributions of the parameters; Section 4.5 shows how to obtain maximum a posteriori (MAP) estimates of the parameters; Section 4.6 gives examples of the likelihood function of our model for different seismograms; Section 4.7 shows examples of the relevant conditional distributions and these examples show the necessity of using multi-sensor analysis; Section 4.8 covers examples of the posterior distributions of the parameters; in Section 4.9 I examine outlier detection techniques used to identify suspicious automatic processing and picking; in Section 4.10 a comparison between the quality of human processing and the automatic processor is given. Lastly, Section 4.11 summarises this chapter.



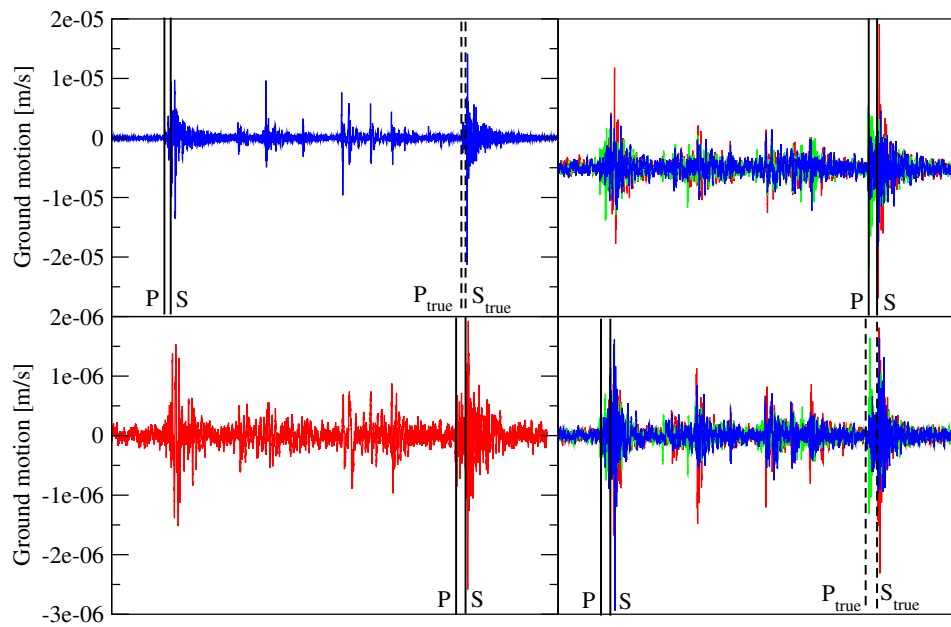
### 4.1.1 From single-sensor to multi-sensor analysis

In this subsection, I show by example why the multi-sensor analysis introduced in this dissertation is so important. I remind the reader that, while such single-seismogram pickers are adequate in high-SNR situations, single seismogram pickers fail when seismograms are noisy or when multiple seismic events are recorded in a buffer. Figures 4.1–4.3 show by example what can and does go wrong.

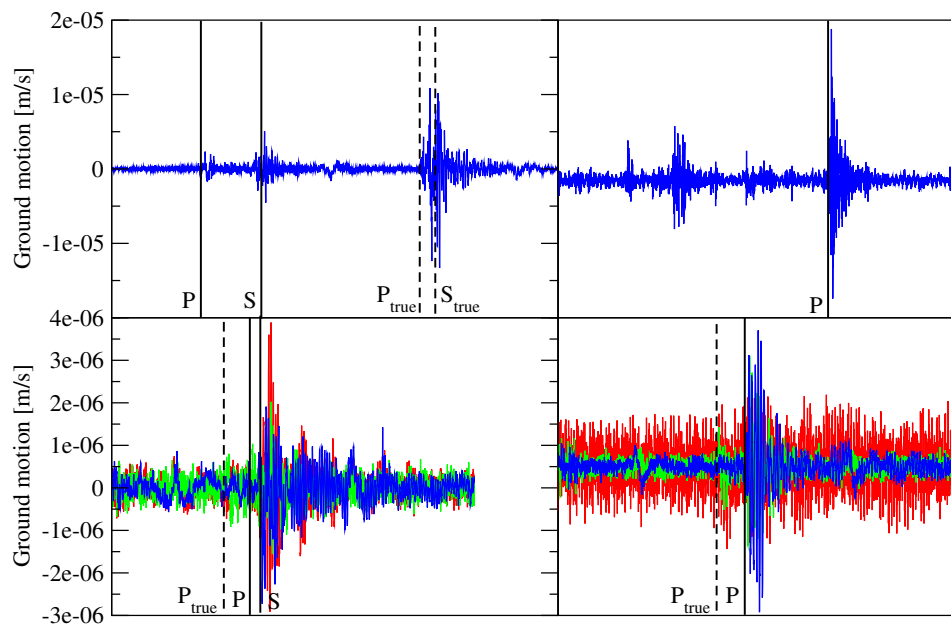
The underlying reason why multiple-sensor analysis is crucial is the seismic event itself. The *same* origin time and hypocentre information pertains to all pertinent seismograms, which thereby interlinks their analysis. By contrast, it is not optimal to determine the wave onsets on a per seismogram basis, as information about the origin time and hypocentre is ignored. Viewing it in a probabilistic way: better arrival time estimates will be achieved if, instead of simply considering the probability distribution of the arrival times given individual seismogram information, the probability distribution of the arrival times given the origin time and hypocentre of the seismic event in question is also utilised. The hypocentre and origin time link arrival times of the different seismograms. Figure 4.4 illustrates the importance of knowledge of the hypocentre and origin time in determining arrival times.



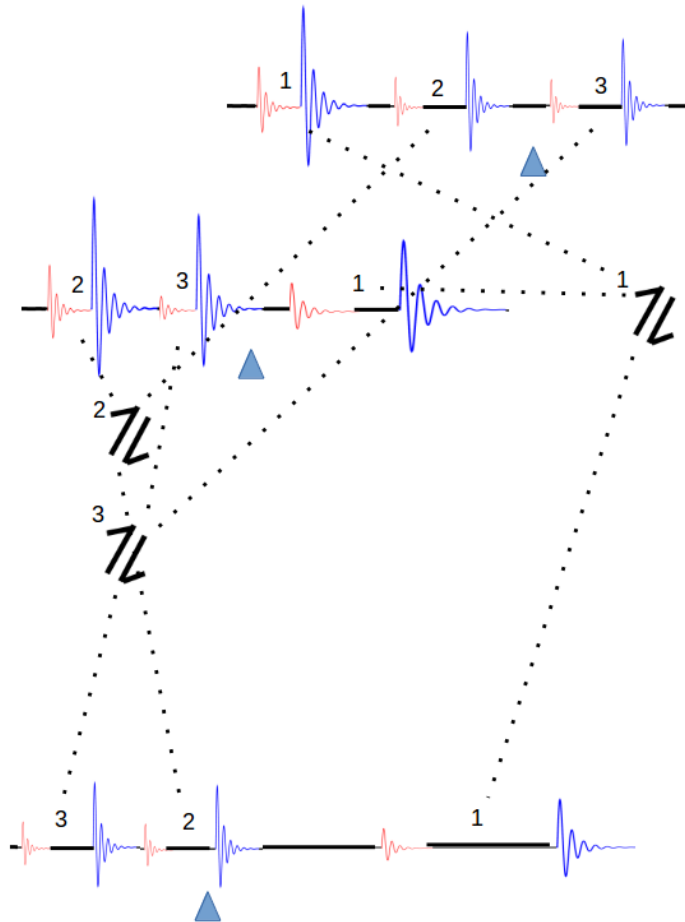
**Figure 4.1:** Failure of single-seismogram pickers I. The vertical solid lines indicate the phase arrivals as determined by the single seismogram picker. The dotted vertical lines indicate the correct (true) phase arrivals. Three of the four seismograms were picked incorrectly by the automatic single seismogram picker. Note that this event was correctly picked by the automatic processor as developed in this dissertation.



**Figure 4.2:** Failure of single-seismogram pickers II. Solid lines indicate the phase arrivals as determined by the single seismogram picker; dotted ones the true ones. The single-seismogram picker processed the first seismic event in some cases and in other cases picked the last seismic event. This event was correctly picked by the automatic processor as developed by me.



**Figure 4.3:** Failure of single-seismogram pickers III. The single picker typically fails when the seismogram has significant noise or when the SNR ratio of the P-wave is low; it picks the P-wave on the true S-wave arrival. This event was again correctly picked by the automatic processor described in this dissertation.



**Figure 4.4:** Schematic drawing of multiple events in a buffer. Three seismic events marked as harpoons are recorded in three locations marked as blue triangles. Depending on relative proximities, even the order of arrival of signals can change. Without information of the location, it is almost impossible to automatically pick the arrivals correctly. The only way to obtain accurate/correct arrivals is to approach this as a joint system of locations and pickings.

From these examples, it is clear that the determining of the phase arrivals must be modelled as a multivariate dependent problem. My method uses both single- and multiple-seismogram analysis. The single seismogram picker is used to identify candidates or possible arrivals, with the multiple-seismogram picker finalising the choice. I first consider the criteria and formulae used for the single-seismogram

picks in Section 4.1.3, followed in Subsection 4.1.4 by the multiple-seismogram theory.

### 4.1.2 The Problem (and Solution) of Automatic Picking

When processing recorded data, the seismologist identifies a single pair of arrivals consisting of a P-wave and an S-wave for each sensor. When processing is done automatically, rather than identifying a single P-wave and S-wave pair, multiple possible pairs of arrival times are considered. Of these multiple possible pairs of arrival times, only one pair is correct: all other pairs are false. Initially, for a recorded seismogram of  $n$  samples, and ignoring the physics of waves, each of the  $\binom{n}{2}$  possible arrival time pairs has the same probability of being the actual arrival time. These probabilities change as physical constraints are applied and information is received and processed. For instance, when the STA/LTA set out in Subsection 4.1.3 is carried out on a waveform, most of the original possible arrivals are dismissed right away, because no burst in energy (high STA/LTA) is associated with them, and only a few candidate arrival pairs are left with non-zero probabilities. The original seismogram is the raw data and the processed data are the candidate arrivals. The problem is simple:

How to determine the correct phase arrival times out of a set of possible (or candidate) arrivals for each sensor. Of the set of possible arrivals, only one pair is correct; all other arrivals are fake or incorrect.

Probabilistic graphical models (PGM), of which a Bayesian network is one, provide a framework to combine all relevant information (variables and data) and to determine the joint distribution and conditional distributions of these variables. The candidate arrivals are modelled as so-called latent variables. One of the more important conditional distributions is the probability distribution of the candidate pairs/latent variables conditioned on specific variables and is derived in Section 4.3.

### 4.1.3 Single Seismogram Pickers and STA/LTA Analysis

The automatic phase arrival determination of seismic waves is called first break picking or simply automatic picking or, as in this dissertation, single seismogram picking. There are various types of statistical and mathematical methods for determining the phase arrivals with the STA/LTA method of Baer and Kradolfer [4] being the most well-known. Other methods are based on fractal-dimension (Boschetti *et al.* [45]), auto-regressive (AR) type models (Zhang *et al.* [9]; Martinsson [46]) and neural networks or multi-window algorithms (Chen and Stewart [47]). I have also proposed an HMM method in Chapter 3. All these automatic pickers attempt to find phase arrivals based on the information contained in a single

seismogram. Most automatic pickers are single seismogram pickers. In the case of waveforms with a single recorded event and a high signal-to-noise ratio (SNR), single seismogram picking should be adequate. Even in cases of noisy single events, filtering can increase the SNR significantly, resulting in accurate picks. A popular single seismogram picker is described in Akazawa [14]. The method makes use of the AR property of seismic waves and that of Akaike information criteria (AIC), to determine first the P-wave arrival and thereafter the S-wave arrival. The technique described in that particular paper was implemented and tested by me and is the single seismogram picker shown in Figures 4.1–4.3. The method works well for single events recorded in a buffer and without too much noise.

For the single-sensor picks, our main criterion is that there should be a burst of energy at each P- and S-wave phase arrival. These bursts can be identified by simply running an STA/LTA characteristic function over the raw data. Also, the standard STA/LTA definition used in the literature has been modified; in this dissertation the modified STA/LTA as defined in Equations 4.1.3 and 4.1.4 is preferably used. The probability of a phase arrival is a function of the STA/LTA values (large STA/LTA values indicate high probability of phase arrivals). There are, however, a number of complications. High STA/LTA values may also be associated with noise or other less energetic seismic events. These non-zero probabilities can be seen as false alarms. For noisy events, the number of possible arrivals will be high. Also, there is a small probability that true arrivals have a low STA/LTA value (say  $< 1.5$ ). The model which I describe in Subsection 4.1.5 can handle such complications, but the majority of sensors should at least have consistent large STA/LTA values, based on the TTE equations (Equation 4.1.6 and Equation 4.1.7). Candidate arrival times are in our case defined as all sample times that have:

1. an associated STA/LTA larger than some threshold value, typically set to 1.5, and
2. the associated STA/LTA is a local maximum within half a period to the left and right.

For a given sensor  $i$ , multiple time windows,  $j = 1, 2, \dots, K$ , of the STA and LTA are used to determine candidate arrival times of a seismogram; this should remove obvious false phase arrivals. To obtain the final STA/LTA values for each sample, the values of the STA/LTA for all time windows are multiplied. True phase arrivals should be associated with high STA/LTA values for all time windows used. For the P-wave, the STA/LTA at time  $t_i$  with STA time window  $n_j^{\text{STA}}$  and LTA time window  $n_j^{\text{LTA}}$  is defined in terms of the STA/LTA ratio  $R$  and the signal amplitude  $A_t$  at time  $t$  as

$$R_j^P(t_i) = \sqrt{\frac{\frac{1}{n_j^{\text{STA}}} \sum_{t=t_i}^{t_i+n_j^{\text{STA}}-1} A_t^2}{\frac{1}{n_j^{\text{LTA}}} \sum_{t=t_i-n_j^{\text{LTA}}+1}^{t_i} A_t^2}}. \quad (4.1.1)$$

For the S-wave, the STA/LTA at time  $t_i$  is defined as

$$R_j^S(t_i) = \frac{\frac{1}{n_j^{\text{STA}}} \sum_{t=t_i}^{t_i+n_j^{\text{STA}}-1} A_t^2}{\sqrt{\frac{1}{n_j^{\text{LTA}}} \sum_{t=t_i-n_j^{\text{LTA}}+1}^{t_i} A_t^2}}. \quad (4.1.2)$$

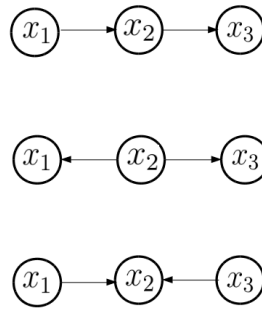
Note that the modified P-wave and S-wave STA/LTA formulas differ. This is done to make the STA/LTA of the S-wave more sensitive to the larger amplitudes near the S-wave onset. It was tested on a large set of events and improves the quality. The final STA/LTA value of the P-wave at time  $t_i$  is then the product of all segments,

$$R^P(t_i) = \prod_{j=1}^K R_j^P(t_i) \quad (4.1.3)$$

and for the S-wave

$$R^S(t_i) = \prod_{j=1}^K R_j^S(t_i). \quad (4.1.4)$$

Furthermore, when the  $R^P(t_i)$  value has a maximum value close (within half a period) to the maximum  $R^S(t_i)$  value and the last candidate arrival of the S-wave is close (within a quarter period) to this maximum value, then it is likely that the time of maximum value of  $R^P(t_i)$  corresponds to the true S-wave arrival. In such cases the maximum value of  $R^P(t_i)$  is lowered. On the other hand, when the  $R^P(t_i)$  value is maximised close (within half a period) to maximum  $R^S(t_i)$  value and the last candidate arrival of the S-wave is not close (more than a quarter period) to this maximum value, then it is likely that the time of maximum value of  $R^S(t_i)$  corresponds to the true P-wave arrival. In such cases the maximum value of  $R^S(t_i)$  is lowered. To increase the SNR of the phase arrivals of events with original low SNR, a  $0.25 \cdot f_{\text{dom}}$  high pass filter is applied to these low SNR seismograms, where  $f_{\text{dom}}$  is the dominant frequency at the maximum amplitude of the seismogram.



**Figure 4.5:** The three possible connections of a DAG: a chain connection (top), a fork connection (middle) a collider connection (bottom).

#### 4.1.4 Bayesian Networks

Signals recorded on multiple seismograms are interdependent when caused by a specific seismic event. In this situation, Bayesian network can adequately model the interplay between the recorded seismograms and further, find the optimal parameters indicating the correct phase arrivals. A Bayesian network (David [48]) is a probabilistic directed acyclic graphical (DAG) model whose nodes represent random variables in the Bayesian sense: they may be observable quantities, latent variables or unknown parameters. There are three types of connections and nodes that are useful in factorisation of a DAG. The three connections are a fork connection, chain connection and a collider connection. Every DAG is made up of a combination of fork-, chain- and collider connections. Firstly, the factorisation of a chain connection (see Figure 4.5 top) is given for the simplest example of three variables by the product rule

$$p(x_1, x_2, x_3) = p(x_1) p(x_2|x_1) p(x_3|x_2),$$

where node  $x_2$  is the chain node, secondly the factorisation of a fork connection (see Figure 4.5 middle) is given by

$$p(x_1, x_2, x_3) = p(x_1|x_2) p(x_3|x_2) p(x_2),$$

where node  $x_2$  is the fork node and lastly the factorisation of a collider connection (see Figure 4.5 bottom) is given by

$$p(x_1, x_2, x_3) = p(x_1) p(x_3) p(x_2|x_1, x_3),$$

where node  $x_2$  is the collider node. The following rules apply:

1. If the chain node  $x_2$  is known or given, then the other two nodes  $x_1$  and  $x_3$  are mutually independent, i.e.  $p(x_1|x_2, x_3) = p(x_1|x_2)$ .

2. If the fork node  $x_2$  is known or given, then the other two nodes  $x_1$  and  $x_3$  are mutually independent, i.e.  $p(x_1|x_2, x_3) = p(x_1|x_2)$ .
3. If the collider node  $x_2$  is unknown, then the other two nodes  $x_1$  and  $x_3$  are mutually independent, i.e.  $p(x_1|x_3) = p(x_1)$ .






Node  $x_1$  is called the parent node and node  $x_2$  is called the child node if  $x_1$  points to  $x_2$ . The factorisation of a DAG consisting of nodes  $x_1, x_2, \dots, x_n$  is then

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | Pa_{x_i}), \quad (4.1.5)$$

where  $Pa_{x_i}$  is the set of parent nodes of node  $x_i$ .

In a Bayesian network, or when Bayesian inference is carried out, plate notation (Buntine [49]) can be used as a convenient representation of the graphical model. This notation is useful when variables or parameters are repeated; a plate or rectangle is used to group these repeated variables into a subgraph, and a number is drawn on the plate to represent the number of repetitions of the subgraph in the plate. Also, plate notation is suited for distinguishing among unknown variables, observed variables (data), fixed parameters and latent variables. Table 4.1 summarises the different variables as used in plate notation.

**Table 4.1:** Table showing the plate notation. Note that if no bracket is indicated in the circle or square, it is then assumed that the variable is a scalar.

| Notation  | Description   |
|---|---|
|  | plate with $n$ repetitions                            |
|  | fixed parameter $\alpha$                              |
|  | latent variable $z_i$ with $k_i$ possible outcomes    |
|  | data vector $\mathbf{x}_i$ of length $d_i$            |
|  | unknown variable $\boldsymbol{\pi}_i$ of length $k_i$ |



### 4.1.5 Variables, Latent Variables, Parameters and Data

As usual, assume a seismic event is recorded at  $n$  sensors. The raw data is the time and amplitude of the waveform at each sample and at each sensor.  $\mathbf{t}^{\text{raw}} = (\mathbf{t}_1^{\text{raw}}, \mathbf{t}_2^{\text{raw}}, \dots, \mathbf{t}_i^{\text{raw}}, \dots, \mathbf{t}_n^{\text{raw}})$  denote the sample times of all sensors, where  $\mathbf{t}_i^{\text{raw}}$  is the vector of sample times of sensor  $i$ ,  $i \in [n] = \{1, 2, \dots, n\}$ . Running any kind of single seismogram picker on the raw seismogram will result in the processed seismogram times (referred to as data), consisting of all times that have a non-zero probability (or at least, in our case, a probability above some STA/LTA threshold) of being a P-wave or S-wave arrival - see Subsection 4.1.3.

Assume that we have  $n_i^{\text{P}}$  possible P-wave arrivals and  $n_i^{\text{S}}$  possible S-wave arrivals at sensor  $i$ . Candidate arrival time pairs consisting of a P-wave and an S-wave time can then be formed as follows: only pairs where the P-wave arrival is earlier than the S-wave arrival are considered; furthermore, only those pairs whose directivity of particle motion of the candidate P-wave and the candidate S-wave is orthogonal are considered (for tri-axial sensors). Assume there are  $n_i^{\text{SP}}$  such pairs at sensor  $i$ .

Also, as was discussed in Section 2.7, not all waveforms have clear P-waves and/or S-wave arrivals. The two main reasons are:

- 1) some waveforms are noisy and the P-wave arrival is hidden in the noise and
- 2) the sensor did not record the P-wave or S-wave, owing to the sensor location relative to the radiation pattern (see Figure 2.4) of the seismic event (Kwiatek and Ben-Zion [50]).

To accommodate these, we introduce in addition another  $n_i^{\text{P}}$  arrival pairs at sensor  $i$  consisting of P-wave arrivals, but without associated S-wave arrivals and another  $n_{S_i}$  arrival pairs at sensor  $i$  which consist of S-wave arrivals, but without associated P-wave arrivals. Thus, in total there are  $k_i = n_i^{\text{SP}} + n_i^{\text{P}} + n_i^{\text{S}}$  possible arrival time pairs. Due to the above mentioned orthogonality condition, there are fewer possible candidates for tri-axial than for uni-axial sensors: for a uni-axial sensor  $n_i^{\text{SP}} = \binom{n_i^{\text{P}} + n_i^{\text{S}}}{2}$ , while for a tri-axial  $n_i^{\text{SP}} \leq \binom{n_i^{\text{P}} + n_i^{\text{S}}}{2}$ . All candidate or possible arrival time pairs of all sensors will be denoted as  $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_i, \dots, \mathbf{t}_n)$ , where  $\mathbf{t}_i = (\mathbf{t}_{i1}, \mathbf{t}_{i1}, \dots, \mathbf{t}_{ik_i})$  is a vector containing the possible arrival time pairs of sensor  $i$ ,  $\mathbf{t}_{ij} = (t_{ij}^{\text{P}}, t_{ij}^{\text{S}})$  is the  $j$ th candidate arrival time pair of sensor  $i$  and  $j \in [k_i]$ ,  $t_{ij}^{\text{P}}$  is the associated P-wave arrival of pair  $j$  of sensor  $i$ , while  $t_{ij}^{\text{S}}$  is the associated S-wave arrival time of pair  $j$  of sensor  $i$ . Each  $t_{ij}^{\text{P}}$  and  $t_{ij}^{\text{S}}$  will have an associated  $R_k^{\text{P}}(t_{ij}^{\text{P}})$  and  $R_k^{\text{S}}(t_{ij}^{\text{S}})$  value respectively. We further denote the STA/LTA set of variables as follows: let  $\mathbf{y} = (\mathbf{y}^{\text{P}}, \mathbf{y}^{\text{S}})$ , where  $\mathbf{y}^{\text{P}} = (\mathbf{y}_1^{\text{P}}, \mathbf{y}_2^{\text{P}}, \dots, \mathbf{y}_n^{\text{P}})$ ,  $\mathbf{y}^{\text{S}} = (\mathbf{y}_1^{\text{S}}, \mathbf{y}_2^{\text{S}}, \dots, \mathbf{y}_n^{\text{S}})$ , consisting of sub-vectors  $\mathbf{y}_i^{\text{P}} = (y_{i1}^{\text{P}}, y_{i2}^{\text{P}}, \dots, y_{ik_i}^{\text{P}})$ ,  $\mathbf{y}_i^{\text{S}} = (y_{i1}^{\text{S}}, y_{i2}^{\text{S}}, \dots, y_{ik_i}^{\text{S}})$ , which in turn are made up of individual variables  $y_{ij}^{\text{P}} = R_k^{\text{P}}(t_{ij}^{\text{P}})$  and  $y_{ij}^{\text{S}} = R_k^{\text{S}}(t_{ij}^{\text{S}})$ . The processed  $\mathbf{t}$  and  $\mathbf{y}$  represent our data.

The arrival at each sensor is modelled using a latent variable. Latent variables are used when variables are hidden or in our case to link observable data in the real world to symbolic data in the modelled world. In our case, the latent variable for each sensor represents all possible phase arrivals at that sensor. Each sensor has a latent selection variable  $z_i$  (modelled as an unknown variable) that is a categorical variable in the sense that a value  $z_i = 0$  indicates that none of the  $k_i$  candidate arrivals is considered true, while any other  $z_i = j$  (with  $1 \leq j \leq k_i$ ) marks the arrival  $j$  as the perceived true one, associated with the relevant arrival time  $t_{ij}$  and corresponding STA/LTA values  $y_{ij}^P$  and  $y_{ij}^S$ . Only one candidate arrival can be true, so all other candidates are false; these false arrivals of both P-waves and S-waves are assumed to have been generated by a uniform distribution with lower bound  $t_{i1}^{\text{raw}}$  and upper bound  $t_{in_i}^{\text{raw}}$  i.e. domain equal to the length of the seismogram recorded at sensor  $i$ . This uniform distribution assumption will become clear in Section 4.3 when the mathematics of the PGM is explained. While there may be other pairs of arrivals that are seismic in nature, these are assumed false as we are only interested in the seismic event with the largest energy. The categorical variable  $z_i$  is equivalent to saying that each possible arrival pair is a Bernoulli trial equalling 1 if the candidate pair is the true arrival or 0 otherwise, under the constraint that the sum of all Bernoulli trials must equal 1. Correspondingly, each possible outcome  $j$  for a given  $z_i$  has an associated probability  $\pi_{ij}$  (called a mixing weight), with  $\sum_{j=0}^{k_i} \pi_{ij} = 1$  for all sensors  $i$ . The Bernoulli probability vector for all sensors is  $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_n)$ . All the component individual sensor mixing weights  $\boldsymbol{\pi}_i = (\pi_{i0}, \pi_{i1}, \dots, \pi_{ik_i})$  are modelled as unknown variables. The prior distribution for  $\boldsymbol{\pi}_i$  is modelled by a fixed hyper-parameter  $\boldsymbol{\alpha}$ .

Note that in most cases latent variables are eliminated through summation in Bayesian analysis; in our case the latent variable is the most important variable and will not be eliminated.

To summarise: For each sensor  $i$ , there are  $k_i$  possible arrival pairs, and the aim is to determine a single most likely arrival pair for the latent variable  $z_i$ , which can take on any integer value between 0 and  $k_i$ .

Each integer  $j$  represents a candidate arrival and each candidate arrival has

- 1) a corresponding arrival time pair (data) that consists of a P-wave arrival time  $t_{ij}^P$  and S-wave arrival time  $t_{ij}^S$  and
- 2) corresponding STA/LTA values (data)  $y_{ij}^P$  and  $y_{ij}^S$ .

In our case, the probability of each phase arrival is not modelled as fixed, unlike that described in the Bayesian analysis of Tarantola and Valette [22] and Martinsson [23]. The data of our model are the times of the possible arrivals for both P- and S-wave of each seismogram as determined by the methods of Section 4.1.3. All other recorded data (sample times with associated STA/LTA < 1.5) is assumed to be generated by non-seismic processes and is discarded. The data therefore consists of true phase arrivals or incorrectly identified arrivals generated by the random process.

The next set of variables and parameters that forms part of our PGM follows from considering the travel time equation (TTE). This equation will be used every time the joint distribution of all variables is evaluated. As stated in Section 2.7, the TTE is

$$t_i^P = t_0 + \frac{d_i(x, y, z)}{v_i^P} \quad (4.1.6)$$

for the P-wave and

$$t_i^S = t_0 + \frac{d_i(x, y, z)}{v_i^S} \quad (4.1.7)$$

for the S-wave, where  $t_0$  is the unknown origin time,  $v_i^P$  is the P-wave velocity between the seismic event and sensor  $i$  (modelled as a fixed parameter),  $v_i^S$  is the S-wave velocity between the seismic event and sensor  $i$  (modelled as a fixed parameter) and  $d_i(x, y, z)$  is the distance between the unknown location of the seismic event and the sensor location  $i$ . The variable  $\theta = (x, y, z, t_0)$  is modelled as an unknown parameter.

There is an additional complication. The sample times of waveforms from sensors are synchronised by a single GPS timing source. Because of technical constraints in the timing signal transfer mechanisms, some sensors may exhibit a time offset. Let  $t_i^{\text{off}}$  be the time offset of sensor  $i$ , then the travel time equations become  $t_i^P = t_0 + t_i^{\text{off}} + \frac{d_i(x, y, z)}{v_i^P}$  for the P-wave and  $t_i^S = t_0 + t_i^{\text{off}} + \frac{d_i(x, y, z)}{v_i^S}$  for the S-wave. A variable  $\rho_i$  is introduced, which is 0 if  $t_i^{\text{off}} = 0$  or otherwise 1. The  $\rho_i$  variable is modelled as an unknown variable.

### 4.1.6 Literature Review

Signals recorded on multiple seismograms are interdependent when caused by a specific seismic event. Modelling or determining phase arrivals in a multivariate

way should outperform phase arrival pickers that use information contained in only one seismogram. Bayesian analysis to obtain posterior distributions for the hypocentre of seismic events has been carried out in literature (e.g. Tarantola and Valette [22]; Martinsson [23]), but few, if any, PGM-based methods were used to determine phase arrival onsets. Tarantola and Valette [22] as well as Tarantola [51] considered the seismograms as the observable parameters (data), and modeled the information of the physical correlations between the unknown parameters (origin time and 3D hypocentre) and the observable parameters. This information can then be combined using Bayesian statistics. The author of that article continues by explaining that for each seismogram and for each sample, a probability of an arrival (for all types of waves) needs to be specified.

The probability density function of the arrival times might be multi-modal and/or asymmetric. Martinsson [23] modelled the origin time and location parameters in a fully Bayesian way with multiple possible candidate arrivals. The likelihood function is modelled as a mixture model. The mixture weights were however kept fixed, and the only parameters of the author's model were the origin time and hypocentre. Myers *et al.* [52] modelled the phase labels (similar to mixture weights) and the origin time and location parameters in a full Bayesian way for multiple types of waves. The phase labels were modelled as random variables, but the number of candidate arrivals was limited to the number of types of waves. Thus, the resulting likelihood function was not a mixture model, but a product of Gaussian distributions. Note that these authors had other objectives in mind for the statistical inference which are not exactly applicable in this study. Our objective is to obtain the correct phase arrivals from a set of candidate arrivals by utilising the information of the physical relationship among the phase arrivals and the origin time and hypocentre of the seismic event. For each seismogram, all but one P-wave and one S-wave arrival of the set of candidate arrivals are in our case eliminated.

As stated, I model the mixture weights as unknown variables; by contrast, Martinsson [23] assumes that the mixture weights are fixed parameters and thereby uses Bayesian analysis to obtain the posterior distribution of the seismic event location. The posterior predictive equation, or similar methods, may then be used for obtaining the most likely phase arrivals. Martinsson [23] mentions that the likelihood function is a standard mixture model. He uses a likelihood function that integrates (sums) all the possible arrivals in order to determine the posterior distribution of the location. The uncertainty/variance of the location could be reduced by using only the 'correct' arrival time in the likelihood function instead of assuming that all mixing weights are equal and fixed.

In the case of automatic processing, initially the correct phase arrivals are not known precisely, so this should not be interpreted as observed data. In actuality, the P-wave and S-wave phase arrivals are random variables with the candidate arrivals (data) as possible outcomes (having associated probabilities). There is

only one true phase arrival with the rest being false alarms. The information of the origin time and hypocentre can be used to identify these false alarms (outcomes with very low probability). The distribution of the arrivals given an origin time and seismic event location, may be used to identify these false alarms. However, the false alarms will be correctly identified as such only if an accurate origin time and hypocentre are used.

To solve this inference problem, I propose that a probabilistic graphical model be used. As will be shown, this specific PGM problem can be written as a latent variable model, which is a Bayesian network.

## 4.2 The Specific Bayesian Network

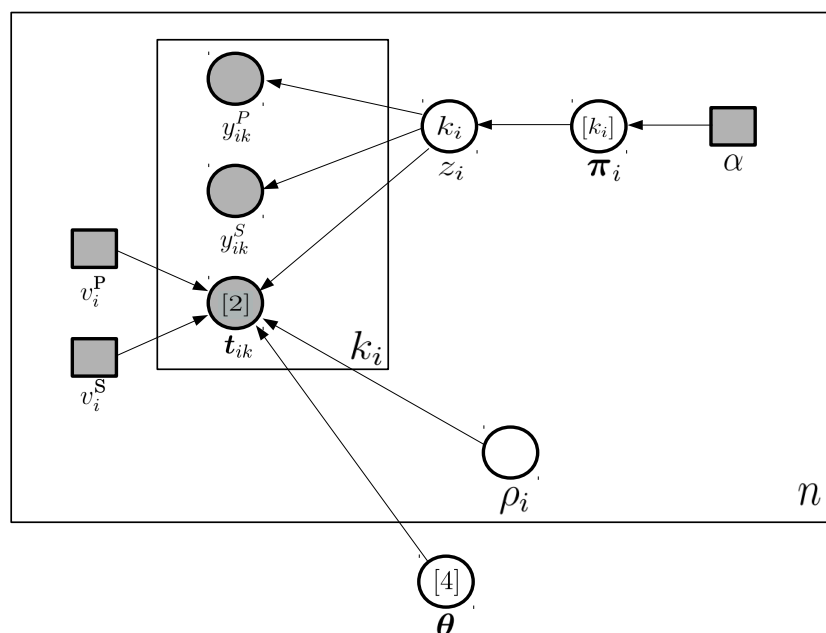
In PGM analysis, the parent node is typically considered the hypothetical cause and the child node is typically the effect. This is not a requirement, but models generally follow this casual intuition (Koller [53]). In Bayesian networks, the direction of arrows always points from the variables towards the observed data. To see this, note that the posterior distribution is proportional to the likelihood function (which is the probability distribution of the data conditioned on the parameters) times the prior distributions. Thus the data node is a collider node. Arrows that connect variables with other variables may point in any direction, but typically the direction is based on how easily the conditional probabilities can be written. Variable nodes can be chain nodes or fork nodes.

For our seismic data processing PGM, we have the following nodes and interactions, which are then summarised in graphical form in Figure 4.6.

- Firstly, let us consider the observed variable  $\mathbf{t}_i$  at sensor  $i$ : the origin time and the hypocentre of the seismic event  $\boldsymbol{\theta}$  is the cause (the parent node) that effects the arrival times  $\mathbf{t}_i$  (the child node). This holds for all sensors, making  $\boldsymbol{\theta}$  a fork node. This also suggests that the arrival times at different sensors are mutually conditionally independent given  $\boldsymbol{\theta}$ . This intuitively makes sense.
- $v_i^P$  (the known seismic P-wave velocity variable) and  $\boldsymbol{\theta}$  are independent, except if  $\mathbf{t}_i$  is known. This suggests that  $v_i^P$ ,  $\boldsymbol{\theta}$  and  $\mathbf{t}_i$  is a collider connection, with  $\mathbf{t}_i$  the collider node. The same holds for the known seismic S-wave velocity variable  $v_i^S$ .
- $\rho_i$  (the time offset variable) and  $\boldsymbol{\theta}$  are independent, except if  $\mathbf{t}_i$  is known. This suggests that  $\rho_i$ ,  $\boldsymbol{\theta}$  and  $\mathbf{t}_i$  is a collider connection, with  $\mathbf{t}_i$  the collider node.
- $\rho_i$  and  $v_i^P$  are independent, except if  $\mathbf{t}_i$  is known. This suggests that  $\rho_i$ ,  $v_i^P$  and  $\mathbf{t}_i$  is a collider connection, with  $\mathbf{t}_i$  the collider node.

- The latent variable  $z_i$  at sensor  $i$  is the parent node of the children data nodes,  $t_i$ ,  $\mathbf{y}_i^P$  and  $\mathbf{y}_i^S$ . This makes  $z_i$ ,  $t_i$ ,  $\mathbf{y}_i^P$  and  $\mathbf{y}_i^S$  a fork connection with  $z_i$  the fork node. This also makes  $t_i$ ,  $\mathbf{y}_i^P$  and  $\mathbf{y}_i^S$  all conditionally (on  $z_i$ ) independent of one another.
- The  $\pi_i$  variable is the parent node of the child node  $z_i$  of sensor  $i$ . The reason for the arrow being in this direction is that  $p(z_i|\pi_i)$  is known. This makes  $\pi_i$ ,  $z_i$  and  $t_i$  ( $\mathbf{y}_i^P$  and  $\mathbf{y}_i^S$ ) a chain connection with  $z_i$  the chain node. Also, this suggests that  $\pi_i$  is independent of  $t_i, \mathbf{y}_i^S$  and  $\mathbf{y}_i^P$  given  $z_i$ .
- The fixed parameter  $\alpha$  is the parent node of  $\pi_i$ .

These considerations suggest the probabilistic model represented in Figure 4.6; see Table 4.1 for the symbols used. There are  $n$  square plates, one for each sensor  $i$ . The node  $\theta$  for the hypocentre and origin time of the seismic event is connected to all of these plates. Figure 4.6 correctly models the interplay of the candidate arrivals recorded at different sensors.



**Figure 4.6:** Our probabilistic graphical model using plate notation. It models the interplay of the candidate arrivals recorded at different sensors of a seismic event with hypocentre and origin time,  $\theta$ .

### 4.3 PGM Factorisation

The factorisation rule of Equation 4.1.5, together with the connection rules, may be used to derive distributions of the random variables defined in Section 4.1.5. The posterior distribution, the prior distributions, the likelihood distribution(s) and some conditional distributions will now be considered.

**The posterior distribution of  $\theta, z, \pi, \rho$ .** As argued above, the joint distribution of  $\theta, t, y, z, \pi, \rho$  can be factorised as

$$\begin{aligned}
 p(\theta, t, y, z, \pi, \rho) &= p(\theta) \prod_{i=1}^n \left\{ p(t_i | \theta, z_i, \rho_i, v_i^P, v_i^S) p(y_i^P | z_i) p(y_i^S | z_i) \right. \\
 &\quad \left. p(z_i | \pi_i) p(\pi_i | \alpha) p(\rho_i) \right\} \\
 &= p(\theta) \prod_{i=1}^n \left\{ \left[ \prod_{j=0}^{k_i} p(t_{ij} | \theta, z_i, \rho_i, v_i^P, v_i^S) p(y_{ij}^P | z_i) p(y_{ij}^S | z_i) \right] \right. \\
 &\quad \left. p(z_i | \pi_i) p(\pi_i | \alpha) p(\rho_i) \right\}. \tag{4.3.1}
 \end{aligned}$$

This formula provides the most compact and accurate summary of my model.

Note that the fixed parameters are ignored, while they are technically part of the joint distribution. The posterior distribution of  $\theta, z, \pi, \rho$  is proportional to Equation 4.3.1 and, as can be seen, is a hierarchical Bayesian model. The maximum a posteriori probability (MAP) estimate of  $z$  is important as it answers the question posed in 4.1.2.

I now discuss the component probabilities one by one.

1. **The prior distribution of  $\rho_i$ .**  $p(\rho_i)$  is the prior distribution of  $\rho_i$  and can be estimated from the observed percentage of sensors that do have timing issues. Since for a typical mine, the percentage of sensors with timing issues is 0.01, we have

$$p(\rho_i) = \begin{cases} 0.99 & \text{if } \rho_i=0, \\ 0.01 & \text{if } \rho_i=1. \end{cases} \tag{4.3.2}$$

2. **The prior distribution of  $\pi_i$ .** The prior for  $\pi_i$  is assumed to be the Dirichlet distribution with hyper-parameter  $\alpha$ . The Dirichlet distribution is a conjugate prior for the categorical distribution and further, we choose  $\alpha = \mathbf{1}$  such that the prior is equivalent to a multivariate uniform distribution. Hence,

$$p(\pi_i | \alpha = \mathbf{1}) = \Gamma(k_i). \tag{4.3.3}$$

3. **The conditional distribution of  $z_i|\boldsymbol{\pi}_i$ .**  $p(z_i|\boldsymbol{\pi}_i)$  is a categorical distribution with

$$p(z_i|\boldsymbol{\pi}_i) = \begin{cases} \pi_{i0} & \text{if } z_i = 0 \\ \pi_{i1} & \text{if } z_i = 1 \\ \vdots & \vdots \\ \pi_{ik_i} & \text{if } z_i = k_i \end{cases} \quad \text{with} \quad \sum_{j=0}^{k_i} \pi_{ij} = 1. \quad (4.3.4)$$

4. **The likelihood function of  $y_{ij}^P|z_i$  and  $y_{ij}^S|z_i$ .** In general the larger the STA/LTA values, the higher the probability of a phase arrival. The probability of a phase arrival therefore scales with the normalised STA/LTA values. Define

$$\varphi_{ij}^P = \frac{R^P(t_{ij})}{\sum_{j=1}^{k_i} R^P(t_{ij})} \quad \text{and} \quad \varphi_{ij}^S = \frac{R^S(t_{ij})}{\sum_{j=1}^{k_i} R^S(t_{ij})}. \quad (4.3.5)$$

Then it follows that

$$p(y_{ij}^P|z_i) = \begin{cases} \varphi_{ij}^P & \text{if } z_i = j, j \in [1, n_i^{\text{SP}} + n_i^{\text{P}}] \\ \frac{1}{t_{in_i}^{\text{raw}} - t_{i1}^{\text{raw}}} & \text{otherwise} \end{cases} \quad (4.3.6)$$

and

$$p(y_{ij}^S|z_i) = \begin{cases} \varphi_{ij}^S & \text{if } z_i = j, j \in [1, n_i^{\text{SP}}] \\ \varphi_{ij}^S & \text{if } z_i = j, j \in (n_i^{\text{SP}} + n_i^{\text{P}}, k_i] \\ \frac{1}{t_{in_i}^{\text{raw}} - t_{i1}^{\text{raw}}} & \text{otherwise} \end{cases}. \quad (4.3.7)$$

Note that  $z_i \in [1, n_i^{\text{SP}} + n_i^{\text{P}}]$  corresponds to arrivals where there is a P-wave arrival, but without a corresponding S-wave arrival, hence:  $p(y_{ij}^S|z_i) = \varphi_{ij}^P$  and  $p(y_{ij}^P|z_i) = (t_{in_i}^{\text{raw}} - t_{i1}^{\text{raw}})^{-1}$ . Also,  $z_i \in (n_i^{\text{SP}} + n_i^{\text{P}}, k_i]$  corresponds to arrivals where there is an S-wave arrival, but without a corresponding P-wave arrival, hence:  $p(y_{ij}^P|z_i) = (t_{in_i}^{\text{raw}} - t_{i1}^{\text{raw}})^{-1}$  and  $p(y_{ij}^S|z_i) = \varphi_{ij}^S$ .

5. **The likelihood function of  $t_{ij}|\boldsymbol{\theta}, z_i, \rho_i, v_i^P, v_i^S$ .**  $p(t_{ij}|\boldsymbol{\theta}, z_i, \rho_i, v_i^P, v_i^S)$  is the likelihood of  $t_{ij}^W$ , where  $W$  is the P-wave or the S-wave, being the arrival time given that the arrival is expected to occur at

$$t_i^W(\boldsymbol{\theta}) = t_0 + \frac{d_i(x, y, z)}{v_i^W} + \varepsilon_i, \quad (4.3.8)$$

where  $v_i^W = v_i^P$  or  $v_i^S$ ,  $t_0$  the origin time,  $d_i(x, y, z) \equiv d_i$  the distance between the hypocentre of the event and the  $i$ th sensor. The candidate arrival times



of the P-wave (or S-wave) are the observed data, but have some known uncertainty which is modelled by  $\varepsilon_i \sim \mathcal{N}(0, \sigma_{\text{pick}})$ , where  $\sigma_{\text{pick}}$  is typically between 1 and 3 ms (see Figure 4.7). From calibration blasts and similar studies (Du Toit and Lynch [54]), the standard deviation of  $v_i^W$  is between  $0.03v_i^W$  and  $0.05v_i^W$ . The standard deviation of  $t_{ij}^W$  can now be computed as (substituting  $1/v_i^W = u_i^W$ )

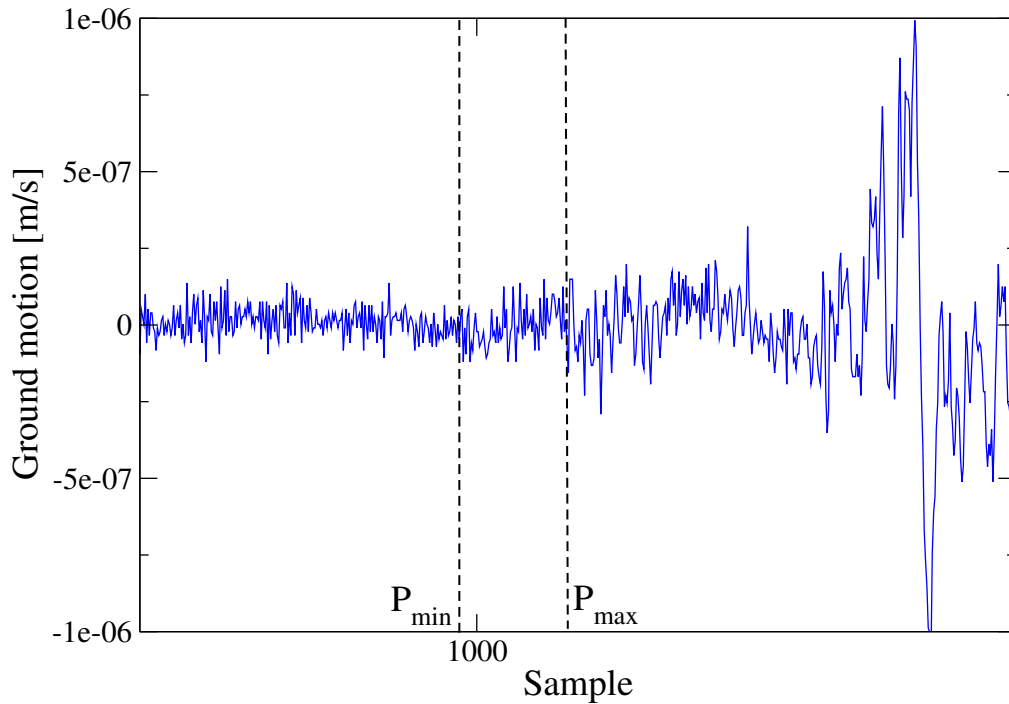
$$\sigma_{t_{\theta i}^W}^2 = \sigma_{\text{pick}}^2 + d_i^2 \sigma_{u_i^W}^2, \quad (4.3.9)$$

where  $u_i^W$  is called the slowness.  $\sigma_{u_i^W}^2$  is usually not known, so by Taylor expansion it follows that

$$\sigma_{u_i^W}^2 \simeq \frac{1}{(v_i^W)^4} \sigma_{v_i^W}^2. \quad (4.3.10)$$

We can therefore rewrite Equation 4.3.9 as

$$\sigma_{t_{\theta i}^W}^2 \simeq \sigma_{\text{pick}}^2 + \left( \frac{d_i}{v_i^W} \right)^2 \left( \frac{\sigma_{v_i^W}}{v_i^W} \right)^2. \quad (4.3.11)$$



**Figure 4.7:** The P-wave arrival of this seismogram is not clear. The  $P_{\min}$  and  $P_{\max}$  are a soft minimum and maximum of the P-wave arrival. Most arrivals are clearer than the arrivals of this seismogram.

As expected, from Equation 4.3.11 it follows that the variance of  $t_i^W$  increases as a function of the distance between the recorded sensor and the hypocentre of the event and also increases as a function of picking error. Note that if  $t_i^{\text{off}} \neq 0$ , then instead of considering  $t_i^W$ , we use  $\Delta t_i = t_i^S - t_i^P$  with

$$\Delta t_i(\boldsymbol{\theta}) = d_i(x, y, z) \left( \frac{1}{v_i^S} - \frac{1}{v_i^P} \right)$$

and

$$\sigma_{\boldsymbol{\theta}\Delta t_i}^2 = \sigma_{\boldsymbol{\theta}t_i^S}^2 + \sigma_{\boldsymbol{\theta}t_i^P}^2,$$

which effectively gets rid of the time offset, but introduces larger variances. A Laplace distribution may be used as the likelihood function. To improve legibility, define

$$\begin{aligned} u_i &= \frac{1}{t_{in_i}^{\text{raw}} - t_{i1}^{\text{raw}}}, \\ \omega_{\boldsymbol{\theta}ij}^P &= \frac{1}{\sigma_{t_i^P\boldsymbol{\theta}}} \exp\left(-\frac{|t_{ij}^P - t_i^P(\boldsymbol{\theta})|}{\sigma_{t_i^P\boldsymbol{\theta}}}\right), \\ \omega_{\boldsymbol{\theta}ij}^S &= \frac{1}{\sigma_{\boldsymbol{\theta}t_i^S}} \exp\left(-\frac{|t_{ij}^S - t_i^S(\boldsymbol{\theta})|}{\sigma_{\boldsymbol{\theta}t_i^S}}\right), \\ \omega_{ij}^{\boldsymbol{\theta}\Delta} &= \frac{1}{\sigma_{\boldsymbol{\theta}\Delta t_i}} \exp\left(-\frac{|\Delta t_{ij} - \Delta t_i(\boldsymbol{\theta})|}{\sigma_{\boldsymbol{\theta}\Delta t_i}}\right). \end{aligned}$$

We then define the arrival probability as the product of the respective P- and S-wave case probabilities ( $\omega_{ik}^{P\boldsymbol{\theta}}$  and  $\omega_{ik}^{S\boldsymbol{\theta}}$  or  $\omega_{ik}^{\boldsymbol{\theta}\Delta}$  in cases when  $\rho_i=1$ ). Specifically, we have that

$$p(\mathbf{t}_{ij} | \boldsymbol{\theta}, z_i, \rho_i, v_i^P, v_i^S) = \begin{cases} (\omega_{ij}^{P\boldsymbol{\theta}})(\omega_{ij}^{S\boldsymbol{\theta}}) & \text{if } \rho_i=0 \text{ and } z_i = j, j \in [1, n_{SPi}] \\ (\omega_{ij}^{P\boldsymbol{\theta}})(u_i) & \text{if } \rho_i=0 \text{ and } z_i = j, j \in (n_{SPi}, n_{SPi}+n_i^P] \\ (u_i)(\omega_{ij}^{S\boldsymbol{\theta}}) & \text{if } \rho_i=0 \text{ and } z_i = j, j \in (n_{SPi}+n_i^P, k_i] \\ \omega_{ij}^{\boldsymbol{\theta}\Delta} & \text{if } \rho_i=1 \text{ and } z_i = j, j \in [1, n_{SPi}] \\ u_i^2 & \text{otherwise} \end{cases} \quad (4.3.12)$$

- The prior distribution of  $\boldsymbol{\theta}$ .** The k-means clustering algorithm is used to determine the optimal hypocentres and standard deviation for a given number of clusters (mixing components). The normalised number of events (density) in each cluster can be used as an estimate of the mixture weights in a Gaussian mixture model. This is similar to the prior of  $\boldsymbol{\theta}$  as described in Martinsson [23].

7. **The likelihood function of sensor  $i$ .** The likelihood function is the distribution of the data given the parameters, i.e.  $p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S | \boldsymbol{\theta}, z_i=k, \rho_i, v_i^P, v_i^S)$ . Let  $c = u_i^{4(k_i-1)}$ , then applying the connection rules for  $\rho_i=0$  and  $k \in [1, n_i^{\text{SP}}]$ , we have

$$\begin{aligned} p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S | \boldsymbol{\theta}, z_i=k, \rho_i) &= p(\mathbf{t}_i | \boldsymbol{\theta}, z_i=k, \rho_i=0) p(\mathbf{y}_i^P | z_i=k, \rho_i) p(\mathbf{y}_i^S | z_i=k, \rho_i) \\ &= c p(\mathbf{t}_{ik} | \boldsymbol{\theta}, z_i=k, \rho_i=0) p(y_{ik}^P | z_i=k, \rho_i) p(y_{ik}^S | z_i=k, \rho_i) \\ &= c(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}}) \end{aligned} \quad (4.3.13)$$

in terms of the above definitions. For  $k \in (n_i^{\text{SP}}, n_i^{\text{SP}} + n_i^{\text{P}}]$  and  $\rho_i=0$

$$p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S | \boldsymbol{\theta}, z_i=k, \rho_i) = c(u_i^2)(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}}) \quad (4.3.14)$$

and for  $k \in (n_i^{\text{SP}} + n_i^{\text{P}}, k_i]$  and  $\rho_i=0$

$$p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S | \boldsymbol{\theta}, z_i=k, \rho_i) = c(u_i^2)(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{S}}). \quad (4.3.15)$$

For  $\rho_i=1$  and  $k \in [1, n_i^{\text{SP}}]$

$$p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S | \boldsymbol{\theta}, z_i=k, \rho_i) = c(\omega_{ik}^{\boldsymbol{\theta}\Delta})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}}). \quad (4.3.16)$$

We have that:  $k_i = n_i^{\text{SP}} + n_i^{\text{P}} + n_i^{\text{S}}$ ; to improve legibility define:  $n_i = n_i^{\text{SP}}$ ;  $m_i = n_i^{\text{SP}} + n_i^{\text{P}}$ . The likelihood function of sensor  $i$  marginalised over  $z_i$  is

$$\begin{aligned} u_i^4 \pi_{i0} + \sum_{j=1}^{n_i} (\omega_{ij}^{\text{P}\boldsymbol{\theta}})(\omega_{ij}^{\text{S}\boldsymbol{\theta}})(\varphi_{ij}^{\text{P}})(\varphi_{ij}^{\text{S}})(\pi_{ij}) \\ + (u_i^2) \sum_{j=n_i}^{m_i} (\omega_{ij}^{\text{P}\boldsymbol{\theta}})(\varphi_{ij}^{\text{P}})(\pi_{ij}) + (u_i^2) \sum_{j=m_i+1}^{k_i} (\omega_{ij}^{\text{S}\boldsymbol{\theta}})(\varphi_{ij}^{\text{S}})(\pi_{ij}). \end{aligned} \quad (4.3.17)$$

8. **The distribution of  $\boldsymbol{\theta}, \mathbf{t}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\rho}$  marginalised over  $\mathbf{z}$ .** Assuming  $\rho_i=0$  for all  $i$ , the marginal distribution of  $\boldsymbol{\theta}, \mathbf{t}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\rho}$  is proportional to

$$\begin{aligned} p(\boldsymbol{\theta}) \prod_{i=1}^n \left[ u_i^4 \pi_{i0} + \sum_{j=1}^{n_i} (\omega_{ij}^{\text{P}\boldsymbol{\theta}})(\omega_{ij}^{\text{S}\boldsymbol{\theta}})(\varphi_{ij}^{\text{P}})(\varphi_{ij}^{\text{S}})(\pi_{ij}) + u_i^2 \sum_{j=n_i+1}^{m_i} (\omega_{ij}^{\text{P}\boldsymbol{\theta}})(\varphi_{ij}^{\text{P}})(\pi_{ij}) \right. \\ \left. + u_i^2 \sum_{j=m_i+1}^{k_i} (\omega_{ij}^{\text{S}\boldsymbol{\theta}})(\varphi_{ij}^{\text{S}})(\pi_{ij}) \right]. \end{aligned} \quad (4.3.18)$$

9. **The conditional distribution of  $z_i | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i$  (single seismogram arrival time distribution).** The distribution of  $z_i | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i$  is the probability distribution of the arrival time when only information contained in the individual seismogram is used. This is the probability distribution of the single

seismogram picker given that  $\boldsymbol{\pi}_i$  is equal to the expected value of  $\boldsymbol{\pi}_i | \alpha = 1$ , i.e.  $\boldsymbol{\pi}_i = \left( \frac{1}{k_i}, \frac{1}{k_i}, \dots, \frac{1}{k_i} \right)$ . Using the connection rules it follows that for  $k \in [1, n_i^{\text{SP}}]$ ,

$$\begin{aligned}
p(z_i=k | \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i) &= \frac{p(\mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}} | z_i=k, \boldsymbol{\pi}_i) p(z_i=k | \boldsymbol{\pi}_i)}{\sum_{j=0}^{k_i} p(\mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}} | z_i=j, \boldsymbol{\pi}_i) p(z_i=j | \boldsymbol{\pi}_i)} \\
&= \frac{p(\mathbf{y}_i^{\text{P}} | z_i=k, \boldsymbol{\pi}_i) p(\mathbf{y}_i^{\text{S}} | z_i=k, \boldsymbol{\pi}_i) \pi_{ik}}{\sum_{j=0}^{k_i} p(\mathbf{y}_i^{\text{P}} | z_i=j, \boldsymbol{\pi}_i) p(\mathbf{y}_i^{\text{S}} | z_i=j, \boldsymbol{\pi}_i) \pi_{ij}} \\
&= \frac{\prod_{j=0}^{k_i} \{p(y_{ij}^{\text{P}} | z_i=k, \boldsymbol{\pi}_i) p(y_{ij}^{\text{S}} | z_i=k, \boldsymbol{\pi}_i)\} \pi_{ik}}{\sum_{j=0}^{k_i} \prod_{l=0}^{k_i} \{p(y_{il}^{\text{P}} | z_i=j, \boldsymbol{\pi}_i) p(y_{il}^{\text{S}} | z_i=j, \boldsymbol{\pi}_i)\} \pi_{ij}} \\
&= \frac{u_i^{2(k_i-1)} (\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})(\pi_{ik})}{\sum_{j=0}^{k_i} u_i^{2(k_i-1)} p(y_{ij}^{\text{P}} | z_i=j, \boldsymbol{\pi}_i) p(y_{ij}^{\text{S}} | z_i=j, \boldsymbol{\pi}_i) \pi_{ij}} \\
&= \frac{(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})(\pi_{ik})}{u_i^2(\pi_{i0}) + \sum_{j=1}^{n_i} (\varphi_{ij}^{\text{P}})(\varphi_{ij}^{\text{S}})(\pi_{ij}) + u_i \sum_{j=n_i+1}^{m_i} (\varphi_{ij}^{\text{P}})(\pi_{ij}) + u_i \sum_{j=m_i+1}^{k_i} (\varphi_{ij}^{\text{S}})(\pi_{ij})}
\end{aligned} \tag{4.3.19}$$

and assuming that  $\boldsymbol{\pi}_i = \left( \frac{1}{k_i}, \frac{1}{k_i}, \dots, \frac{1}{k_i} \right)$ , then

$$p(z_i=k | \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i) = \frac{(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})}{u_i^2 + \sum_{j=1}^{n_i} (\varphi_{ij}^{\text{P}})(\varphi_{ij}^{\text{S}}) + u_i \sum_{j=n_i+1}^{m_i} \varphi_{ij}^{\text{P}} + u_i \sum_{j=m_i+1}^{k_i} \varphi_{ij}^{\text{S}}} \tag{4.3.20}$$

For  $k \in (n_i^{\text{SP}}, n_i^{\text{SP}} + n_i^{\text{P}}]$ ,

$$p(z_i=k | \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i) = \frac{(u_i)(\varphi_{ik}^{\text{P}})(\pi_{ik})}{u_i^2(\pi_{i0}) + \sum_{j=1}^{n_i} (\varphi_{ij}^{\text{P}})(\varphi_{ij}^{\text{S}})(\pi_{ij}) + u_i \sum_{j=n_i+1}^{m_i} (\varphi_{ij}^{\text{P}})(\pi_{ij}) + u_i \sum_{j=m_i+1}^{k_i} (\varphi_{ij}^{\text{S}})(\pi_{ij})} \tag{4.3.21}$$

and for  $k \in (n_i^{\text{SP}} + n_i^{\text{P}}, k_i]$ ,

$$p(z_i=k | \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i) = \frac{(u_i)(\varphi_{ik}^{\text{S}})(\pi_{ik})}{u_i^2(\pi_{i0}) + \sum_{j=1}^{n_i} (\varphi_{ij}^{\text{P}})(\varphi_{ij}^{\text{S}})(\pi_{ij}) + u_i \sum_{j=n_i+1}^{m_i} (\varphi_{ij}^{\text{P}})(\pi_{ij}) + u_i \sum_{j=m_i+1}^{k_i} (\varphi_{ij}^{\text{S}})(\pi_{ij})} \tag{4.3.22}$$

10. **The conditional distribution of  $z_i | \boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i, \rho_i$  (multiple seismogram arrival time distribution).** This is the probability distribution of the arrival time when all relevant information is used. Using the connection

rules and the results in previous distribution, it follows that  $p(z_i=k|\boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=0)$ , for  $k \in [1, n_i^{\text{SP}}]$ , is equal to

$$\begin{aligned}
& \frac{p(\mathbf{t}_i | \boldsymbol{\theta}, z_i=k, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=0) p(z_i=k | \boldsymbol{\theta}, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i)}{\sum_{j=0}^{k_i} p(\mathbf{t}_i | \boldsymbol{\theta}, z_i=j, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=0) p(z_i=j | \boldsymbol{\theta}, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i)} \\
&= \frac{p(\mathbf{t}_i | \boldsymbol{\theta}, z_i=k, \rho_i=0) p(z_i=k | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i)}{\sum_{j=0}^{k_i} p(\mathbf{t}_i | \boldsymbol{\theta}, z_i=j, \rho_i=0) p(z_i=j | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i)} \\
&= \frac{\prod_{j=0}^{k_i} \{p(\mathbf{t}_{ij} | \boldsymbol{\theta}, z_i=k, \rho_i=0)\} p(z_i=k | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i)}{\sum_{j=0}^{k_i} \left\{ \prod_{l=0}^{k_i} p(\mathbf{t}_{il} | \boldsymbol{\theta}, z_i=j, \rho_i=0) \right\} p(z_i=j | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i)} \tag{4.3.23} \\
&= \frac{u_i^{2(k_i-1)} p(\mathbf{t}_{ik} | \boldsymbol{\theta}, z_i=k, \rho_i=0) p(z_i=k | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i)}{\sum_{j=0}^{k_i} u_i^{2(k_i-1)} p(\mathbf{t}_{ik} | \boldsymbol{\theta}, z_i=j, \rho_i=0) p(z_i=j | \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i)} \\
&= \frac{(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})(\pi_{ik})}{u_i^4 \pi_{i0} + \sum_{j=1}^{n_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \varphi_{ij}^{\text{S}} \pi_{ij} + u_i^2 \sum_{j=n_i+1}^{m_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \pi_{ij} + u_i^2 \sum_{j=m_i+1}^{k_i} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{S}} \pi_{ij}}
\end{aligned}$$

and assuming that  $\boldsymbol{\pi}_i = (\frac{1}{k_i}, \frac{1}{k_i}, \dots, \frac{1}{k_i})$ , then  $p(z_i=k|\boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i)$  is equal to

$$\frac{(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})}{u_i^4 + \sum_{j=1}^{n_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \varphi_{ij}^{\text{S}} + u_i^2 \sum_{j=n_i^{\text{SP}}+1}^{n_i^{\text{SP}}+n_i^{\text{P}}} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} + u_i^2 \sum_{j=m_i+1}^{k_i} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{S}}}. \tag{4.3.24}$$

For  $k \in (n_i^{\text{SP}}, n_i^{\text{SP}} + n_i^{\text{P}}]$  and  $\rho_i=0$  it follows that  $p(z_i=k|\boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=0)$  is equal to

$$\frac{(u_i^2)(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\pi_{ik})}{u_i^4 \pi_{i0} + \sum_{j=1}^{n_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \varphi_{ij}^{\text{S}} \pi_{ij} + u_i^2 \sum_{j=n_i+1}^{m_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \pi_{ij} + u_i^2 \sum_{j=m_i+1}^{k_i} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{S}} \pi_{ij}} \tag{4.3.25}$$

and for  $k \in (n_i^{\text{SP}} + n_i^{\text{P}}, k_i]$  and  $\rho_i=0$  it follows that  $p(z_i=k|\boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=0)$  is equal to

$$\frac{(u_i^2)(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{S}})(\pi_{ik})}{u_i^4 \pi_{i0} + \sum_{j=1}^{n_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \varphi_{ij}^{\text{S}} \pi_{ij} + u_i^2 \sum_{j=n_i+1}^{m_i} \omega_{ij}^{\text{P}\boldsymbol{\theta}} \varphi_{ij}^{\text{P}} \pi_{ij} + u_i^2 \sum_{j=m_i+1}^{k_i} \omega_{ij}^{\text{S}\boldsymbol{\theta}} \varphi_{ij}^{\text{S}} \pi_{ij}}. \tag{4.3.26}$$

For  $\rho_i=1$  and  $k \in [1, n_i^{\text{SP}}]$  it follows that  $p(z_i=k | \boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=1)$  is equal to

$$\frac{(\omega_{ik}^{\theta\Delta})(\varphi_{ik}^P)(\varphi_{ik}^P)(\pi_{ik})}{u_i^4(\pi_{i0}) + \sum_{j=1}^{n_i} (\omega_{ik}^{\theta\Delta})(\varphi_{ij}^P)(\varphi_{ij}^S)(\pi_{ij})}. \quad (4.3.27)$$

The mode of this distribution is a point-estimator of the phase arrivals. This distribution is important as it can be used to answer (as does the MAP of  $\mathbf{z}$ ) the question posed in Subsection 4.1.2 when  $\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{MAP}}$  and  $\boldsymbol{\pi}_i = \boldsymbol{\pi}_{i\text{MAP}}$ . This distribution will also be used in the Gibbs sampler discussed next in Section 4.4.

## 4.4 Posterior Distributions via MCMC

### 4.4.1 Motivation

Markov Chain Monte Carlo (MCMC) method is a well-known technique of obtaining (or approximating) a complex posterior distribution in Bayesian analysis; for a comprehensive review of these methods see Neal [55]. Ideally, one would want to obtain the full posterior distribution over the parameters and variables of interest, but this takes too long to calculate in a situation where fast processing is of the essence. Fortunately, the maximum a posteriori (MAP) estimate, which is the mode of the posterior distribution, is often sufficient for the purpose. There are techniques for MAP estimation that will be discussed in Subsection 4.5 which do not rely on approximating the full posterior distribution. In this section, posterior distributions of parameters will be approximated for a few seismic events in order to test the logical and mathematical structure of our model which the production algorithms approximate. Examples of the posterior distribution for a few seismic events are shown in Section 4.8.

Two sampling techniques, namely Metropolis-Hastings (Metropolis *et al.* [56]; Hastings [57]) and Gibbs sampling (Geman and Geman [58]) will be employed to sample the posterior distribution  $p(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\rho} | \mathbf{t}, \mathbf{y})$ . For simplicity, we will assume  $\rho_i=0$  for all  $i$ . There are two types of Metropolis-Hastings sampling algorithms:

- 1) the global algorithm, where all the parameters are updated simultaneously and
- 2) the local algorithm, where each parameter is updated systematically in turn.

The local Metropolis-Hastings algorithm (also called variable-at-a-time Metropolis-Hastings or Metropolis-within-Gibbs algorithm (Geyer [59])) will be used to sample  $\boldsymbol{\theta}$  and  $\boldsymbol{\pi}$ . The algorithm specifies a way to generate samples using a proposal

density that rejects or accepts proposed new samples of the posterior distribution. The proposal distribution of a new  $\boldsymbol{\theta}^{(t)}$  will be the symmetric random walk model centered at the previous  $\boldsymbol{\theta}^{(t-1)}$ , i.e.  $\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)} \sim \mathcal{N}(\boldsymbol{\theta}^{(t-1)}, h)$ , where  $h$  is the standard deviation of the jumps, while the proposal distribution of  $\boldsymbol{\pi}_i^{(t)}$  will be the Dirichlet prior of  $\boldsymbol{\pi}_i \sim \text{Dir}(\mathbf{1})$  which is independent of the current state  $\boldsymbol{\pi}_i^{(t-1)}$ ; this is discussed further in Subsection 4.4.2. Further, the  $z_i$  variable will be marginalised over when sampling  $\boldsymbol{\pi}_i$ , which is called a collapsed Gibbs sampler (Liu [60]). The Gibbs sampler will be used to generate samples for each  $\mathbf{z}$ . Gibbs sampling generates samples from the conditional posterior distributions of each variable, i.e. the distribution of a single parameter conditioned on all other parameters and data. The conditional distributions of  $z_i|\boldsymbol{\theta}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i$  are given by Equation 4.3.23 - Equation 4.3.26. As shown in Section 4.4.2, sampling from  $z_i|\boldsymbol{\theta}^{(t)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i^{(t-1)}$  can be done directly and easily. The following step-by-step procedure will produce a sequence of observations that is approximately distributed according to the posterior distribution of  $\boldsymbol{\theta}, \boldsymbol{\pi}, \mathbf{z}$ .

#### 4.4.2 Procedure to Sample From Posterior Distribution

The sampler is initialised:  $\boldsymbol{\theta}^{(0)}, \boldsymbol{\pi}_i^{(0)}, z_i^{(0)}$  for each  $i \in [n]$ . The following steps are taken for each  $i$  to sample  $\boldsymbol{\pi}_i^{(t)}$  for a given  $\boldsymbol{\theta}^{(t-1)}$  and  $\boldsymbol{\pi}_i^{(t-1)}$  at time step  $t$ :

- Draw a proposal state  $\boldsymbol{\pi}_i \sim q(\boldsymbol{\pi}_i|\boldsymbol{\pi}_i^{(t-1)}) = q(\boldsymbol{\pi}_i)$ , where  $q(\boldsymbol{\pi}_i) = \text{Dir}(\mathbf{1})$ .
- Calculate the probability for the proposal state  $\boldsymbol{\pi}_i$ ,

$$p'(\boldsymbol{\pi}_i|\boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S) = (u_i^4)(\pi_{i0}) + \sum_{j=1}^{n_i} (\omega_{ij}^{P\boldsymbol{\theta}})(\omega_{ij}^{S\boldsymbol{\theta}})(\varphi_{ij}^P)(\varphi_{ij}^S)(\pi_{ij}) \\ + u_i^2 \sum_{j=n_i+1}^{m_i} (\omega_{ij}^{P\boldsymbol{\theta}})(\varphi_{ij}^P)(\pi_{ij}) + u_i^2 \sum_{j=m_i+1}^{k_i} (\omega_{ij}^{S\boldsymbol{\theta}})(\varphi_{ij}^S)(\pi_{ij}).$$

Note that  $p(\boldsymbol{\pi}_i|\boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S) \propto p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\pi}_i) p(\boldsymbol{\pi}_i|\boldsymbol{\theta}^{(t-1)}) \propto p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\pi}_i) = \sum_k p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S|\boldsymbol{\theta}, z_i = k) p(z_i = k|\boldsymbol{\pi}_i)$

- Obtain  $r = \min \left\{ 1, \frac{p'(\boldsymbol{\pi}_i|\boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S)}{p'(\boldsymbol{\pi}_i^{(t-1)}|\boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S)} \right\}$ .
- If  $\alpha \sim \text{Uni}(0, 1) < r$  with  $\text{Uni}$  the uniform distribution, then accept the proposed state,  $\boldsymbol{\pi}_i^{(t)} = \boldsymbol{\pi}_i$ ; otherwise set  $\boldsymbol{\pi}_i^{(t)} = \boldsymbol{\pi}_i^{(t-1)}$ .

Once a  $\boldsymbol{\pi}_i^{(t)}$  is drawn for all  $i$ , then the following steps will sample  $z_i$  for a given  $\boldsymbol{\theta}^{(t-1)}$  and  $\boldsymbol{\pi}_i^{(t)}$ :

- Draw a  $z_i$  from  $p(z_i | \boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i^{(t)})$ . This can be done by generating  $\alpha \sim \text{Uni}(0, 1)$ , then  $z_i^{(t)} = k^*$  if

$$\sum_{j=0}^{[k^*-1]} p(z_i = j | \boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i^{(t)}) < \alpha < \sum_{j=0}^{[k^*]} p(z_i = j | \boldsymbol{\theta}^{(t-1)}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i^{(t)}).$$

To evaluate the logarithm of Equation 4.3.1 for a proposed  $\boldsymbol{\theta} = (x, y, z, t_0)$ , where  $\boldsymbol{\theta}$  was generated by perturbing  $\boldsymbol{\theta}^{(t-1)}$  with a standard deviation of 10 meters for  $x^{(t-1)}$ ,  $y^{(t-1)}$  and  $z^{(t-1)}$  and 1 millisecond for the origin time, and for a given  $\boldsymbol{\pi}_i^{(t)}$  and  $z_i^{(t)}$ , the following steps are taken:

- Calculate the distance  $d_i$  between  $(x, y, z)$  and the coordinate of the sensor.
- Determine the travel time between  $(x, y, z)$  and the coordinate of the sensor, i.e.  $(d_i/v_i^P)$  for P-wave and  $(d_i/v_i^S)$  for S-wave.
- Calculate the expected arrival time  $t_i^P(\boldsymbol{\theta}) = t_0 + \frac{d_i}{v_i^P}$  and  $t_i^S(\boldsymbol{\theta}) = t_0 + \frac{d_i}{v_i^S}$ .

- Calculate the variances

$$\sigma_{t_{\boldsymbol{\theta}_i}^P}^2 = \sigma_{\text{pick}}^2 + \left(\frac{d_i}{v_i^P}\right)^2 \left(\frac{\sigma_{v_i^P}}{v_i^P}\right)^2 \quad \text{and} \quad \sigma_{t_{\boldsymbol{\theta}_i}^S}^2 = \sigma_{\text{pick}}^2 + \left(\frac{d_i}{v_i^S}\right)^2 \left(\frac{\sigma_{v_i^S}}{v_i^S}\right)^2$$

- The sensor probability for possible arrival time pair  $\mathbf{t}_{iz_i}$  given  $\rho_i=0$  is proportional to:

$$p'(\boldsymbol{\theta}, z_i=k^*, \boldsymbol{\pi}_i) = \begin{cases} (u_i^4)(\pi_{k^*i0}) & \text{if } z_i=0 \\ (\omega_{ik^*}^{P\boldsymbol{\theta}})(\omega_{ik^*}^{S\boldsymbol{\theta}})(\varphi_{ik^*}^P)(\varphi_{ik^*}^S)(\pi_{ik^*}) & \text{if } z_i \in [1, n_i] \\ (u_i^2)(\omega_{ik^*}^{P\boldsymbol{\theta}})(\varphi_{ik^*}^P)(\pi_{ik^*}) & \text{if } z_i \in (n_i, m_i] \\ (u_i^2)(\omega_{ik^*}^{S\boldsymbol{\theta}})(\varphi_{ik^*}^S)(\pi_{ik^*}) & \text{if } z_i \in (m_i, k_i] \end{cases} \quad (4.4.1)$$

- Obtain  $\ln p(\boldsymbol{\theta}, \boldsymbol{\pi}, \mathbf{z}) = \sum_{i=1}^n \ln p'(\boldsymbol{\theta}, z_i=k_i^*, \boldsymbol{\pi}_i) + \ln p(\boldsymbol{\theta}_{loc})$ .
- Compute  $\tau = \min\left\{0, \ln p(\boldsymbol{\theta}, \boldsymbol{\pi}, \mathbf{z}) - \ln p(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\pi}^{(t-1)}, \mathbf{z}^{(t-1)})\right\}$ .
- With probability  $\exp(\tau)$ , set  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$ ; otherwise  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$ .

Note that care has been taken to approximate  $p(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\rho}|\mathbf{t}, \mathbf{y})$ . To draw samples of  $\mathbf{z}$  and  $\boldsymbol{\theta}$  from the posterior distribution is straightforward, as we can directly sample  $\mathbf{z}$ , and  $\boldsymbol{\theta}$  is only four-dimensional. Sampling  $\boldsymbol{\pi}_i$  from the posterior distribution, on the other hand, is more problematic:

- The Markov chain can take extremely long (hours to days) to sample the target distribution when  $k_i > 10$ . For this reason, the proposed  $\pi_{ij}$  is set to zero for any  $j$ , in cases when  $p(z_i=j|\boldsymbol{\theta}_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S, \boldsymbol{\pi}_i, \rho_i=1) < 0.05$ .



We therefore do not propose values other than zero if we know that the distribution of  $\pi_{ij}$  will peak at zero. In essence, only proposed samples near the mode are allowed; low probability regions are ignored. With this scheme the number of independent samples  $n_o = 500,000$  and to run the sampler takes about 5 minutes.

- The lag at which the autocorrelation of the MCMC samples of  $\boldsymbol{\pi}$  starts to be weakly dependent was calculated, i.e. where the Pearson correlation coefficient became smaller than some  $\varepsilon$  as  $\eta_\varepsilon$ . The autocorrelation at lag 5 is below 0.01. For our problem we thus have  $\eta_{0.01} = 5$ ; hence every 5th sample can be assumed to be practically independent.
- Burn-in periods are not an issue here as the starting values of the parameters are the MAP values.
- The number of optimal independent samples,  $n_o$ , was obtained by determining when the standard deviation of  $\langle \pi_{ij} \rangle = \frac{1}{n_o} \sum_{t=0}^{n_o} \pi_{ij}^{(t)}$ , i.e.  $\{[n_o(n_o-1)]^{-1} \sum_{t=0}^{n_o} (\pi_{ij}^{(t)} - \langle \pi_{ij} \rangle)^2\}^{1/2}$  is smaller than some critical value  $\lambda$  for all  $i$  and  $j$ . In our case,  $\lambda$  was set to be 0.0005.
- With this scheme, the acceptance ratio in sampling  $\boldsymbol{\pi}_i$  is around 0.66.

## 4.5 Maximum A Posteriori Estimation

As will be shown in this section, it turns out that the maximum a posteriori (MAP) estimates of  $z_i$ ,  $\boldsymbol{\pi}_i$  and  $\rho_i$  can be determined analytically given  $\boldsymbol{\theta}_{\text{MAP}}$ , while  $\boldsymbol{\theta}_{\text{MAP}}$  itself can only be derived numerically by using a nonlinear optimisation algorithm such as Nelder-Mead (Olsson and Nelson [32]). In this dissertation, we are more interested in the MAP estimates, but in Section 4.8 I do compare the posterior distributions of  $z_i$  and  $\boldsymbol{\pi}_i$  to  $p(z_i=k|\boldsymbol{\theta}_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i, \rho_i)$ . It turns out that these distributions are similar.

For a known  $\boldsymbol{\theta}$ , obtaining MAP estimates (technically these are MAP estimates only if  $\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{MAP}}$ ) for  $z_i$ ,  $\boldsymbol{\pi}_i$  and  $\rho_i$  for  $i \in [n]$  is straightforward. The MAP estimate of  $z_i \equiv \hat{k}_i$ , where  $\hat{k}_i \equiv k_i(\boldsymbol{\theta}, \rho_i) \in [k_i]$ , for a known  $\boldsymbol{\theta}$  and  $\rho_i=1$  is

$$\begin{aligned} & \arg \max_{k \in [k_i]} \prod_{j=0}^{k_i} p(\mathbf{t}_{ij} | \boldsymbol{\theta}, z_i=k, \rho_i, v_i^{\text{P}}, v_i^{\text{S}}) p(y_{ij}^{\text{P}} | z_i=k) p(y_{ij}^{\text{S}} | z_i=k) \\ & = \arg \max_{k \in [k_i]} p(\mathbf{t}_{ik} | \boldsymbol{\theta}, z_i=k, \rho_i, v_i^{\text{P}}, v_i^{\text{S}}) p(y_{ik}^{\text{P}} | z_i=k) p(y_{ik}^{\text{S}} | z_i=k) \end{aligned} \tag{4.5.1}$$

The MAP estimate of  $\boldsymbol{\pi}_i$  is:  $\pi_{i\hat{k}_i} = 1$  and  $\pi_{ij} = 0$  for all  $j \in \{n \in [k_i] | n \neq \hat{k}_i\}$ . The MAP estimate for  $\rho_i$  is 0 if the expression

$$\max_{k \in [k_i]} \left( u_i^4, \max_{k \in [n_i]} [(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})] , \max_{n_i < k \leq m_i} [(u_i^2)(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\pi_{ik})] , \right. \\ \left. \max_{m_i < k \leq k_i} [(u_i^2)(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{PS}})(\pi_{ik})] \right) p(\rho_i=0)$$

is larger than

$$\max_{k \in [k_i]} \left( u_i^4, \max_{k \in [n_i]} [\omega_{ik}^{\boldsymbol{\theta}\Delta} \varphi_{ik}^{\text{P}} \varphi_{ik}^{\text{S}}] \right) p(\rho_i=1)$$

and otherwise it is 1.

The MAP estimate of  $\boldsymbol{\theta}$  is the maximum of  $2n + 1$  factors

$$\boldsymbol{\theta}_{\text{MAP}} = \left[ \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) \right] \prod_{i=1}^n \left[ F_{0i} p(\rho_i=0) , F_{1i} p(\rho_i=1) \right] \quad (4.5.2)$$

with

$$F_{0i} = \max_{k \in [k_i]} \left[ u_i^4 , \max_{k \in [n_i]} [(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})] , \right. \\ \left. \max_{n_i < k \leq m_i} [(u_i^2)(\omega_{ik}^{\text{P}\boldsymbol{\theta}})(\varphi_{ik}^{\text{P}})(\pi_{ik})] , \max_{m_i < k \leq k_i} [(u_i^2)(\omega_{ik}^{\text{S}\boldsymbol{\theta}})(\varphi_{ik}^{\text{PS}})(\pi_{ik})] \right] \quad (4.5.3)$$

$$F_{1i} = \max_{k \in [k_i]} \left[ u_i^4, \max_{k \in [n_i]} [(\omega_{ik}^{\boldsymbol{\theta}\Delta})(\varphi_{ik}^{\text{P}})(\varphi_{ik}^{\text{S}})] \right]. \quad (4.5.4)$$

Firstly, starting locations of  $\boldsymbol{\theta}$  as close as possible to the true location (in the basin of attraction of Equation 4.5.2) are desirable. The location found by minimising the travel time equations TTE (Equation 4.1.6 and Equation 4.1.7), using the most likely P-wave onset and S-wave onset found from the HMM as arrival times, can be used as a starting point in most scenarios, with the origin time starting point the difference between the average arrival times and average travel times of all sensors. The coordinates of the first sensor that triggered can also be a starting point (for a relatively dense seismic array). A starting point of the origin time is the trigger time of the first sensor location.

Secondly, we want the search space of the maximisation algorithm to be as small as possible. The probability of  $\boldsymbol{\theta}_{\text{MAP}}$  falling into this volume should be high. The radius of this volume can be obtained by calculating the distance between the median first possible/candidate arrival time (assume this is P-wave onset) and the median last possible arrival time (assume this is S-wave onset). Using the TTE,

this distance can be calculated by subtracting the travel time of the S-wave from the travel time of the P-wave, i.e.

$$t_S - t_P = d \left( \frac{1}{v^S} - \frac{1}{v^P} \right) \quad (4.5.5)$$

or

$$d = \frac{v^S v^P (t^S - t^P)}{v^P - v^S}. \quad (4.5.6)$$

The larger distances associated with noisy signals are still acceptable, as in most cases the optimisation algorithm converges to the global maximum. Though this distance may be large in very noisy signals, it at least shrinks the search area for the MAP estimation.

The Nelder-Mead optimisation method is used because of its speed. Multiple starts are generated within a radius  $d$  from the possible starting points.

### 4.5.1 Procedure to Obtain MAP Estimation

Evaluating the logarithm of Equation 4.3.1 for a specific  $\boldsymbol{\theta}_{\text{spec}} = (x_{\text{spec}}, y_{\text{spec}}, z_{\text{spec}}, t_{0,\text{spec}})$  for MAP estimation is done as follows:

- Calculate the distance  $d_{\text{spec},i}$  between  $(x_{\text{spec}}, y_{\text{spec}}, z_{\text{spec}})$  and the coordinate of sensor  $i$ .
- Calculate the travel time between  $(x_{\text{spec}}, y_{\text{spec}}, z_{\text{spec}})$  and the coordinate of the sensor, i.e.  $(d_{\text{spec},i}/v_i^P)$  for P-wave and  $(d_{\text{spec},i}/v_i^S)$  for S-wave.
- Calculate the expected arrival time  $t_i^P(\boldsymbol{\theta}_{\text{spec}}) = t_{0,\text{spec}} + (d_{\text{spec},i}/v_i^P)$  and  $t_i^S(\boldsymbol{\theta}_{\text{spec}}) = t_{0,\text{spec}} + (d_{\text{spec},i}/v_i^S)$ .
- Calculate the variance  $\sigma_{t_i^P}^2 = \sigma_{\text{pick}}^2 + (d_{\text{spec},i}/v_i^P)^2 (\sigma_{v_i^P}/v_i^P)^2$  and  $\sigma_{t_i^S}^2 = \sigma_{\text{pick}}^2 + (d_{\text{spec},i}/v_i^S)^2 (\sigma_{v_i^S}/v_i^S)^2$ .
- The likelihood of sensor  $i$  for possible arrival time pair  $\mathbf{t}_{ik}$  given  $\rho_i=0$  is proportional to  $\mathcal{L}_{\max ik}(\boldsymbol{\theta}_{\text{spec}}, \rho_i=0)$  which equals  $\max[u_i^4, (\omega_{,ik}^{P\boldsymbol{\theta}})(\omega_{ik}^{S\boldsymbol{\theta}})(\varphi_{ik}^P)(\varphi_{ik}^S), u_i^2(\omega_{ik}^{P\boldsymbol{\theta}})(\varphi_{ik}^P), u_i^2(\omega_{ik}^{S\boldsymbol{\theta}})(\varphi_{ik}^S)]$ .
- The likelihood of sensor  $i$  for possible arrival time pair  $\mathbf{t}_{ik}$  given  $\rho_i=1$  is proportional to  $\mathcal{L}_{\max ik}(\boldsymbol{\theta}_{\text{spec}}, \rho_i=1) = \max[u_i^4, (\omega_{ik}^{\boldsymbol{\theta}\Delta})(\varphi_{ik}^P)(\varphi_{ik}^S)]$ .
- Obtain  $p_{ik} = \max[\mathcal{L}_{\max i}(\boldsymbol{\theta}_{\text{spec}}, \rho_i=0) p(\rho_i=0), \mathcal{L}_{\max i}(\boldsymbol{\theta}_{\text{spec}}, \rho_i=1) p(\rho_i=1)]$ .
- Obtain the maximum  $p_{ik}$  of all  $k_i$  possible arrivals time pairs for sensor  $i$ ,  $p_{\max i}$ .

- Obtain  $\sum_{i=1}^n \ln p_{maxi}$  and add this term to the logarithm of the prior seismic event location probability  $\ln p(\boldsymbol{\theta}_{loc})$  to obtain the logarithm of the posterior distribution.

This step-by-step algorithmic procedure can be used to find the MAP estimates of the parameters when using a Nelder-Mead algorithm.

## 4.6 Examples of the Likelihood of Sensors

For any arbitrary  $\boldsymbol{\theta}$  there are corresponding expected arrival times  $t_i^P(\boldsymbol{\theta})$  and  $t_i^S(\boldsymbol{\theta})$ . Of scientific interest are the graphs of the likelihoods of S- and P-arrivals, Equation 4.3.14 versus  $t_i^P(\boldsymbol{\theta})$  and Equation 4.3.15 versus  $t_i^S(\boldsymbol{\theta})$  for  $z_i = \hat{k}_i$  for different seismograms. Equations 4.3.14 and 4.3.15, which are dependent only on the candidate P-wave arrival times and S-wave arrival times respectively, are used for visualisation purposes: plotting these equations, together with the raw seismogram, can give insight into the sample times  $\hat{k}_i$ , at which the preferred arrival time changes from one  $z_i$  to another. For example: for expected arrivals far in time from any candidate arrivals,  $\hat{k}_i$  will equal zero. Expected values of arrival times close to candidate arrival times will result in  $\hat{k}_i \neq 0$ . As mentioned, only  $z_i \in [n_i, m_i]$  (P-no-S) is considered for the graph of Equation 4.3.14 versus  $t_i^P(\boldsymbol{\theta})$  for  $z_i = \hat{k}_i$  and only  $z_i \in [m_i, k_i]$  (S-no-P) is considered the graph of Equation 4.3.15 versus  $t_i^S(\boldsymbol{\theta})$  for  $z_i = \hat{k}_i$  as visualisation of other pairs, where there are P- and S-wave arrivals is nonsensical.

The variance of the distribution,  $\sigma_{t_{\boldsymbol{\theta}_i}^P}^2 = \sigma_{\text{pick}}^2 + \left(\frac{d_j}{v_i^P}\right)^2 \left(\frac{\sigma_{v_i^P}}{v_i^P}\right)^2$ , depends on the distance between the source (hypocentre) and recorded sensor. The larger the distance, the flatter the distribution (i.e. the higher the variance) will be; the shorter the distance, the more peaked the distribution. This makes sense, as the velocity error aggregates with distance. The graphs of Equation 4.3.14 versus  $t_i^P(\boldsymbol{\theta})$  for  $z_i = \hat{k}_i$  and Equation 4.3.15 versus  $t_i^S(\boldsymbol{\theta})$  for  $z_i = \hat{k}_i$  at different distances  $d_j$  are also of scientific interest. For flatter likelihood functions, i.e. for larger  $d_j$ , the preferred arrival time tends to be arrival times with a larger STA/LTA value: that is  $\varphi_{ij}^P$  and  $\varphi_{ij}^S$  dominate. On the other hand, for  $\boldsymbol{\theta}$  close to the sensor, the preferred P-wave arrival (S-wave arrival) will tend to be close to  $t_i^P(\boldsymbol{\theta})$  ( $t_i^S(\boldsymbol{\theta})$ ). This is evident in Figures 4.8 to 4.15, where the likelihood distribution is flatter, assuming a distance between the hypothetical event location and the sensor coordinate of 2000 m as opposed to 50 m. The more complex the waveforms, the more possible arrivals there are and hence, the more local maxima.

Figures 4.8 to 4.15 present a set of typical results and examples of real seismic events of the likelihood functions for P-wave and S-wave arrival times.

I now summarise the figures and their properties as follows.

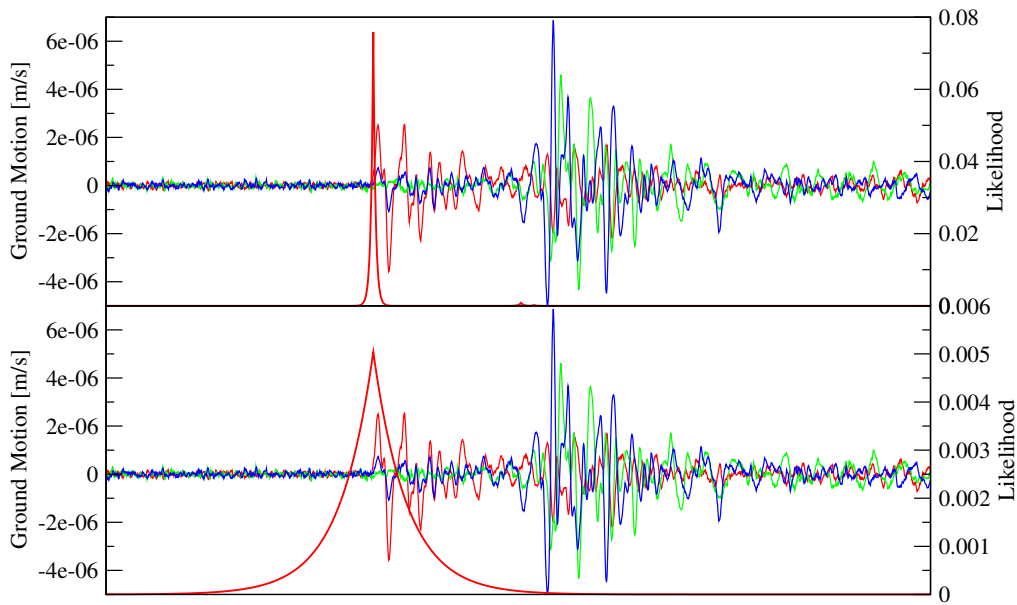
- In all cases, the horizontal axis represents time (sampled at 6000 Hz).

- The seismogram amplitudes are shown on the left for the  $x$ (red)-  $y$ (green)- and  $z$ (blue)-components, while the likelihood function values are shown on the right which quantify the smooth red likelihood curves.
- The shapes of the likelihood functions resemble Laplace mixture models. This should be of no surprise. Let  $\varphi_{ik}^P \pi_{ik} = p_{ik}$ ; then the marginal likelihood of Equation 4.3.14 after summation of  $z_i$  is given by

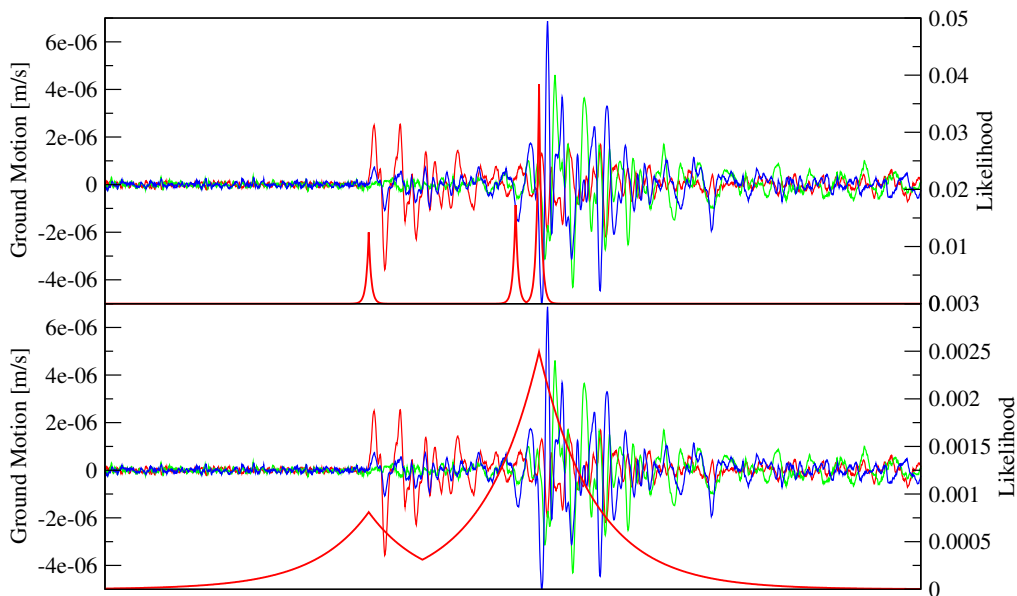
$$p(\mathbf{t}_i, \mathbf{y}_i^P, \mathbf{y}_i^S | \boldsymbol{\theta}, \rho_i) \propto \sum_{k=n_{i+1}}^{m_i} (\omega_{ik}^{P\boldsymbol{\theta}})(\varphi_{ik}^P)(\pi_{ik}) = \sum_{k=n_{i+1}}^{m_i} (p_{ik})(\omega_{ik}^{P\boldsymbol{\theta}}), \quad (4.6.1)$$

which is the same mathematical expression as a Laplace mixture model. Instead of summation, our likelihood function is  $\max_k (\omega_{ik}^{P\boldsymbol{\theta}})(\varphi_{ik}^P)$  evaluated at the most likely candidate arrival. Note that even though the shape of the likelihood is similar to a mixture model, the interpretation is fundamentally different. Where standard mixture models have modes at the mixture component means, the likelihood function has modes at the candidate arrival times. In standard mixture models, the dimension of the parameters (number of mixture components) determines the number of modes, but in our case, any given  $\boldsymbol{\theta}$  corresponds to or represents only one mode, namely the expected arrival at each sensor. Hence, technically, our model is not a mixture model, even though the shape of the distribution resembles one. Martinsson [23] claims it is a mixture model, but as I have explained here, that is not the case.

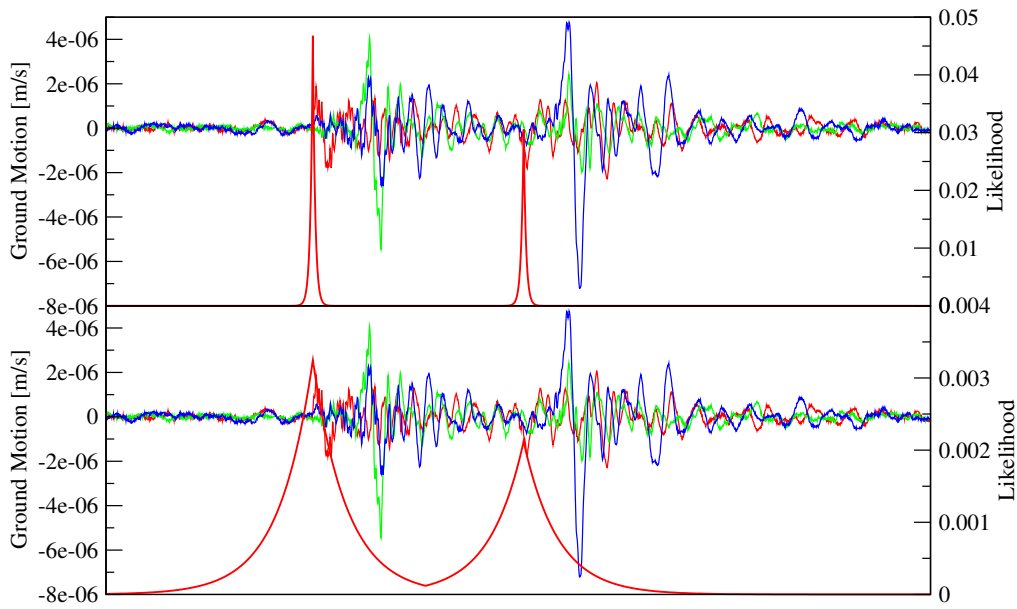
- As mentioned previously, the modes of the likelihood functions shown in the figures correspond to only P-wave (or S-wave) candidate arrival times and will therefore typically differ from the modes of the posterior distribution of  $z_i$ . On the other hand, the modes of the likelihood functions of Equation 4.3.13, which have corresponding P-wave and S-wave arrivals, will be similar to the modes of the posterior distribution of  $z_i$ , but cannot be superimposed on the seismograms.
- The larger the distance (bottom panels), the flatter the distribution, i.e. the higher the variance will be; the shorter the distance (top panels), the more peaked the distribution. This makes sense, as the velocity error aggregates with distance.
- The main reason for these figures is to illustrate (in cases of multiple possible arrivals) at what sample times  $\hat{k}_i$ , the preferred arrival time changes from one  $z_i$  to another. Many examples are shown so that the reader can appreciate the complexities in the likelihood function, the local maxima it produces and the natural variability of seismograms producing the corresponding candidate arrivals.



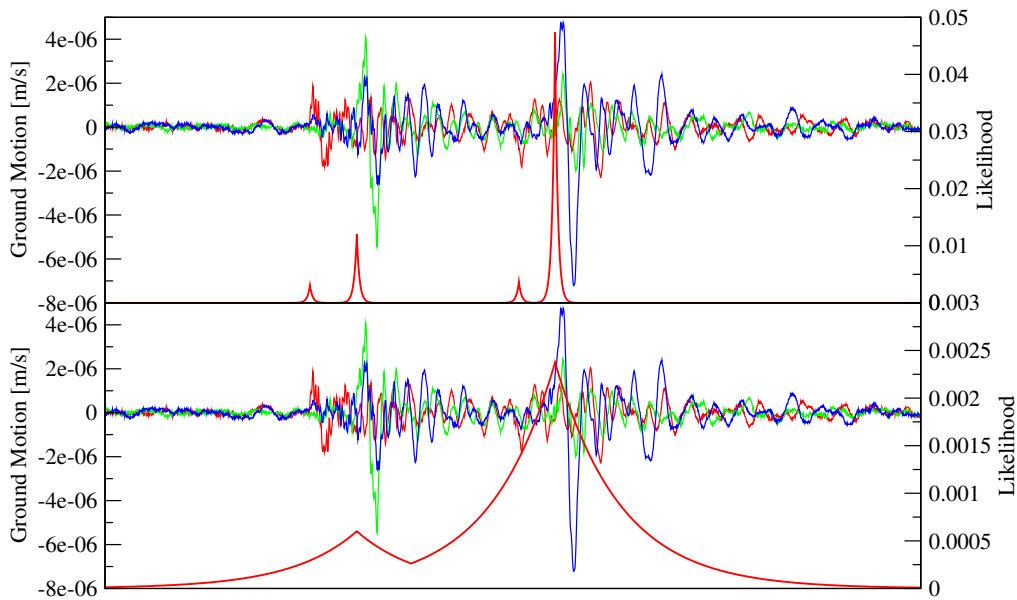
**Figure 4.8:** Seismograms and superimposed likelihood functions of the P-wave arrival for given distances, 50 m (upper panel) and 2000 m (lower panel), between the hypothetical event location and the sensor location. The larger the distance between source and sensor, the larger the variance, as the error in the velocity plays a more significant role. In this example there is only one likely P-wave arrival, and finding the global maximum as discussed in Subsection 4.5.1 should not be too sensitive to the initial starting points of  $\theta$ .



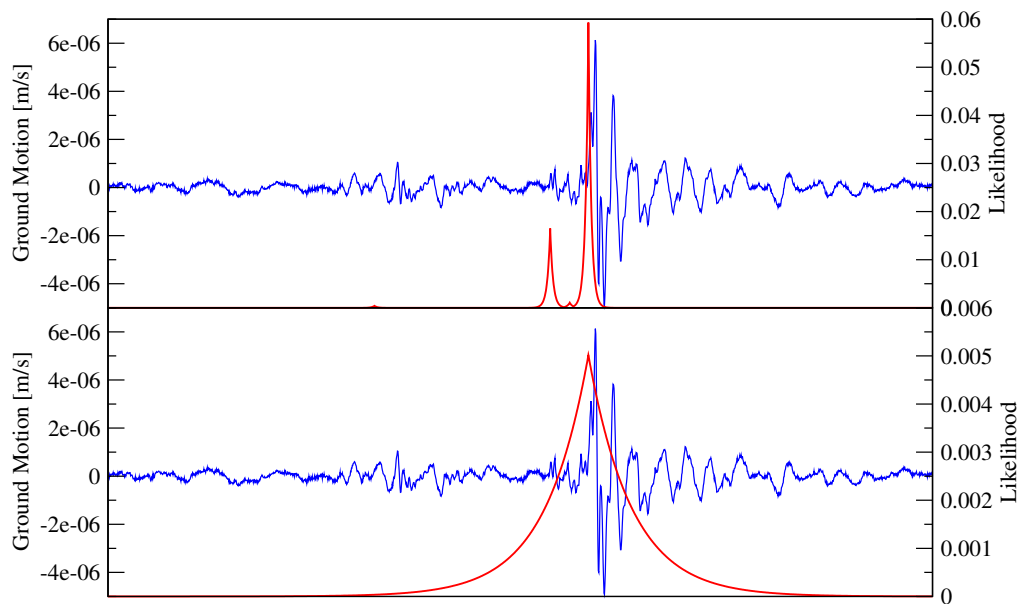
**Figure 4.9:** Likelihood functions of the S-wave arrivals for a seismogram given for the same distances used in Figure 4.8. For the upper 50 m distance seismogram, there are three possible S-wave arrivals. The first arrival peak is clearly wrong as it corresponds to the P-wave arrival, and the likelihood amplitude of this is correctly small. The largest peak and the most likely arrival time correctly corresponds to the true S-wave arrival.



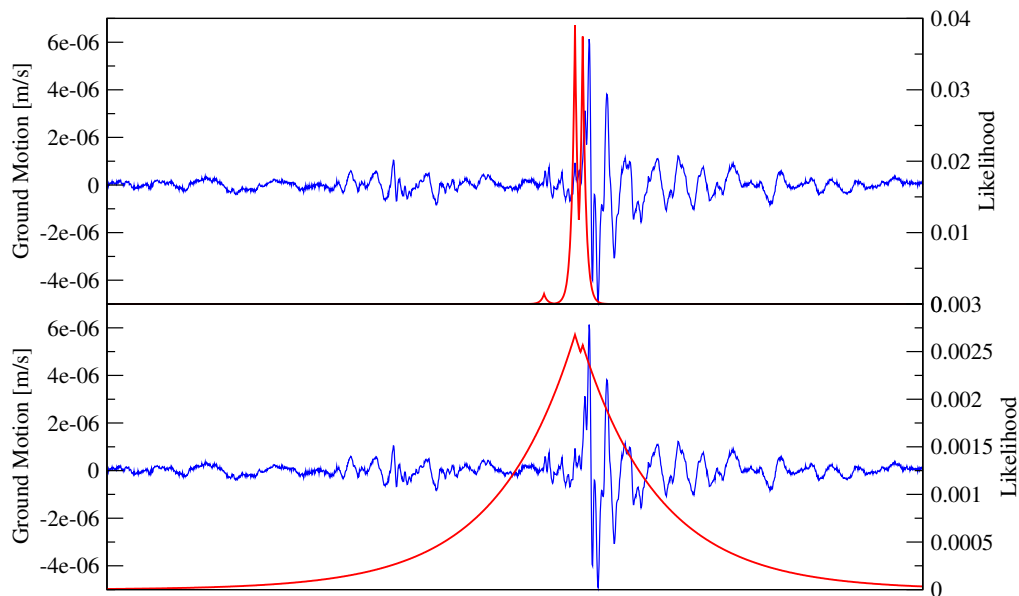
**Figure 4.10:** Likelihood functions of the P-wave arrivals for a seismogram at 50 m (upper panel) and 2000 m (lower panel) distances. While there are two possible P-wave arrivals, the second arrival is clearly wrong as it corresponds to the S-wave arrival. The most likely arrival again corresponds to the true P-wave arrival.



**Figure 4.11:** Likelihood functions of the S-wave arrivals for a seismogram with two seismic events at 50 m (upper panel) and 2000 m (lower panel) distances. In this example there are four possible S-wave arrivals: two incorrect arrivals that correspond to the P-wave arrivals of the two events and then also the two most likely arrivals corresponding to the S-wave arrivals. The second event has a slightly higher probability. The event with the larger maximum amplitude should be picked if the global maximum is found. In this example, with multiple possible arrivals, finding the global maximum as discussed in Subsection 4.5.1 should become more sensitive to the initial values of  $\theta$ .

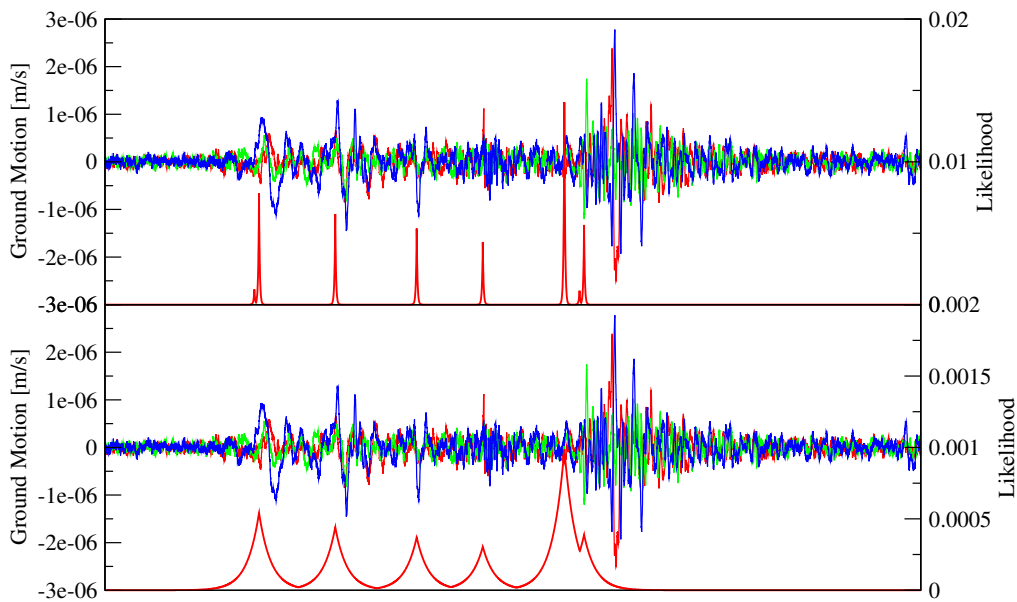


**Figure 4.12:** The likelihood function of the P-wave arrivals for a seismogram at 50 m (upper panel) and 2000 m (lower panel) distances. In this example there are two possible P-wave arrivals, but the most likely P-wave arrival is incorrectly identified at the onset of the S-wave arrival. This will not be of any concern as the optimisation occurs over multiple sensors.

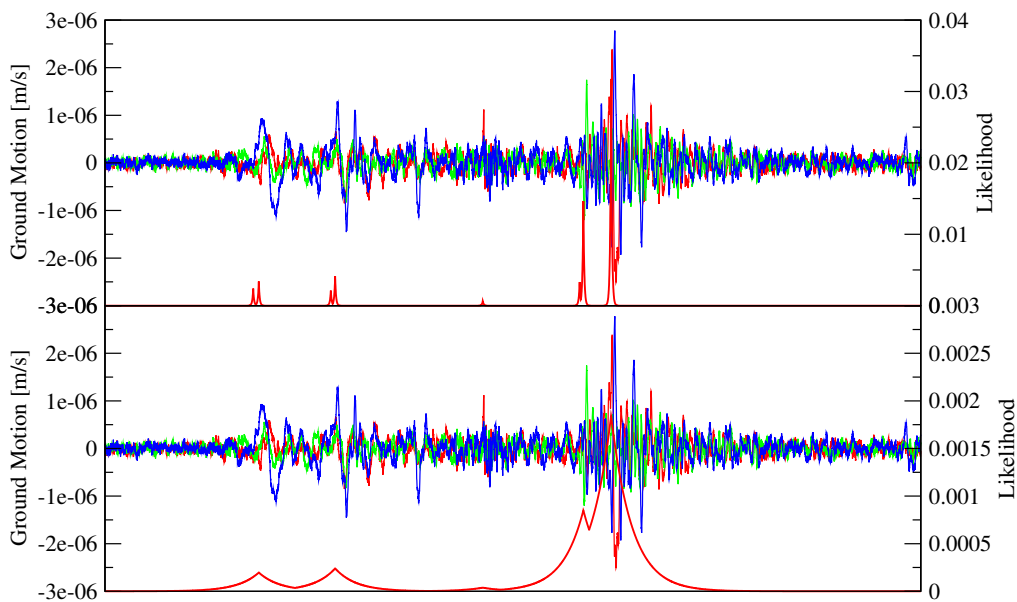


**Figure 4.13:** The likelihood function of the S-wave arrivals for the same seismogram of Figure 4.12. In this example there are only two likely S-wave arrivals close to each other. As the S-wave arrival has been correctly identified, it should solve the problem of the incorrect P-wave identification in the previous figure.





**Figure 4.14:** The likelihood function of the P-wave arrivals for a seismogram. There are many possible arrivals, but the most likely arrival has been correctly identified as the one with the highest likelihood peak. Because optimisation algorithm can easily become stuck in a local maximum, multiple starts are carried out as discussed in Section 4.5.1. Note that the most likely P-wave arrivals was identified correctly.



**Figure 4.15:** Last example of the likelihood function of the S-wave arrivals for a seismogram. This is the same seismogram as the previous example and again there are multiple arrivals for the S-wave. Note that the most likely S-wave arrival was also identified correctly.

## 4.7 Examples and Necessity of Multi-Sensor Analysis

As mentioned in Chapter 2, recorded data needs to be processed within 5 seconds to be of any value. Obtaining posterior distributions of the parameters is thus not practical, but at least MAP estimates of  $z_i$  in particular are quick. However, the posterior distribution of  $z_i$  or  $\pi_i$  can indicate how many likely candidate arrival pairs there are. This may be useful in identifying seismic events that may be potentially incorrectly processed by the automatic processor in order for these to be automatically forwarded to human processors for further assessment.

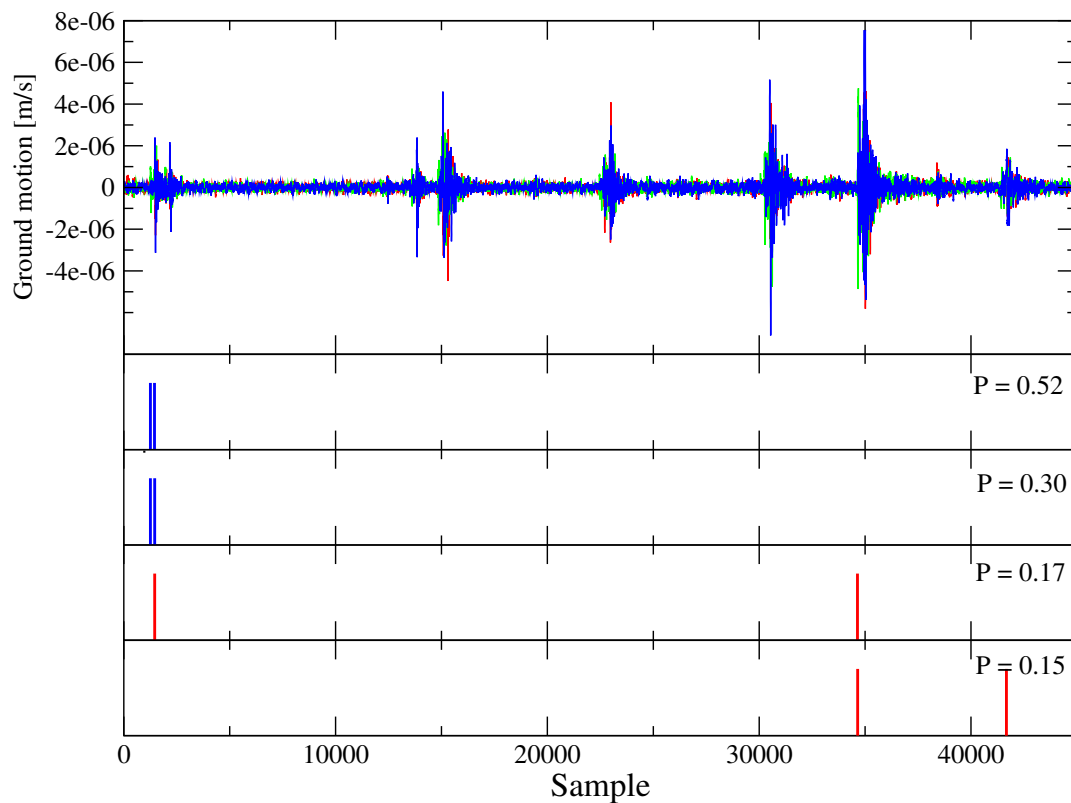
As stated in Section 4.5, the conditional distribution of  $z_i$  given the values of  $\theta_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i^{\text{prior}}, \rho_i$  (distribution 10 in Section 4.2), where  $\boldsymbol{\pi}_i^{\text{prior}} = \boldsymbol{\pi}_i = (\frac{1}{k_i}, \frac{1}{k_i}, \dots, \frac{1}{k_i})$ , is similar to the posterior distribution of  $z_i$  and is quick to evaluate; see Section 4.8 for the similarities. This conditional distribution can therefore be used as an approximation for the posterior distribution. To appreciate the effect of knowing what  $\theta_{\text{MAP}}$  has on the distribution of the arrival, a comparison between the modes of  $z_i | \theta_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i^{\text{prior}}, \rho_i$  (blue bars) and the single seismogram picker  $z_i | \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \boldsymbol{\pi}_i^{\text{prior}}$  (red bars) is shown in Figures 4.16 to 4.25.

I summarise the figures and their properties as follows.

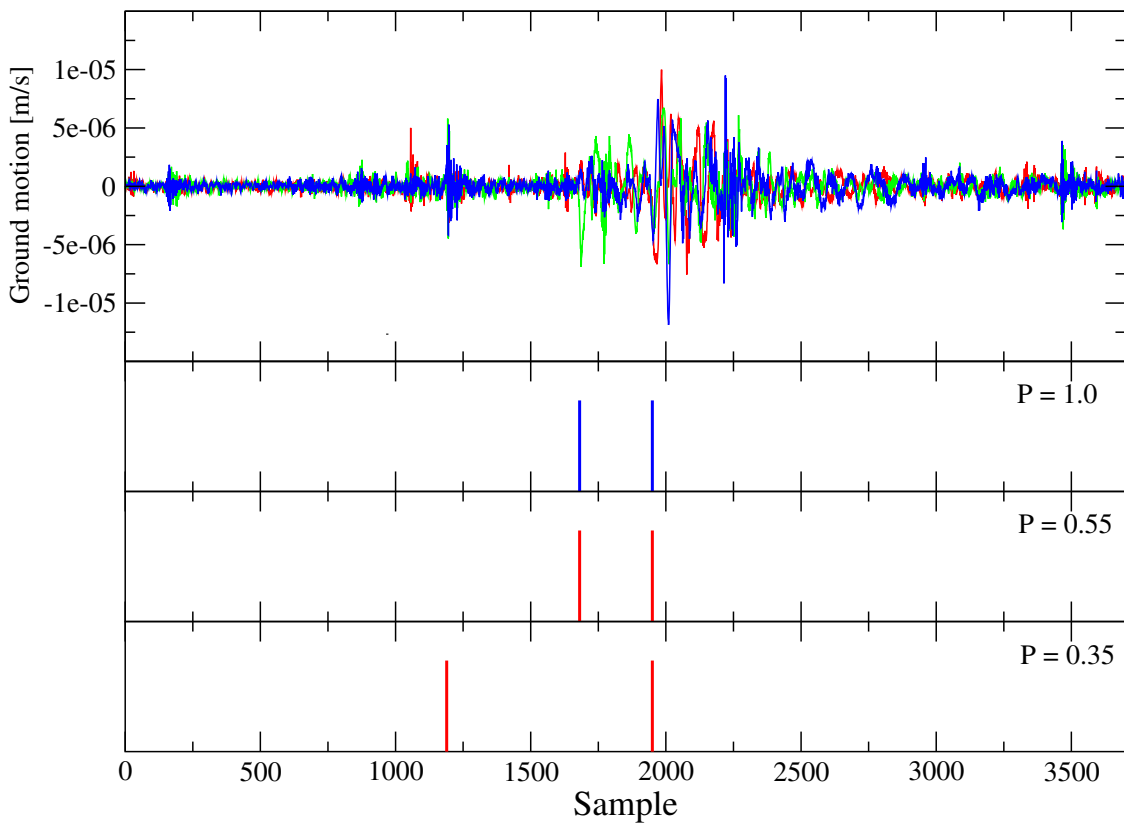
- In all cases, the horizontal axis represents the sample times (sampled at 6000 Hz). These seismic events are real events that occurred on a mine serviced by IMS.
- Only one figure per page is displayed due to its size and length of the caption.
- In all figures, the upper panel displays the seismogram amplitudes on the left for the  $x$ (red)-  $y$ (green)- and  $z$ (blue)-components, while the lower panels represent the modes of the multi-sensor arrival times (blue bars) and the modes of the single seismogram picker (red bars). P-values indicate the probability of the specific arrival time pair being shown. Note that there may be many red bars and blue bars (candidate arrivals), but only the most likely ones are shown. Also, note that if there are two or more modes for the multi-sensor arrival times, the possible arrival pairs will be close (in time) to one another — see for example Figure 4.16.
- Only one (or a few) seismogram in the examples is shown, but many seismograms were recorded and used to determine the multi-sensor arrival distribution.
- The main reason for these examples being shown is to illustrate the necessity of using multi-sensor analysis (fully Bayesian analysis) in phase arrival detection. In most examples the algorithm performs well. Note that the mode of

the multi-sensor arrival times shown in Figure 4.25 is wrong. This is not a rare occurrence and the automatic processor may get stuck in local maxima in up to about 30% of cases. Outlier detection techniques, as discussed in Section 4.9, are used to possibly identify these types of mistakes.

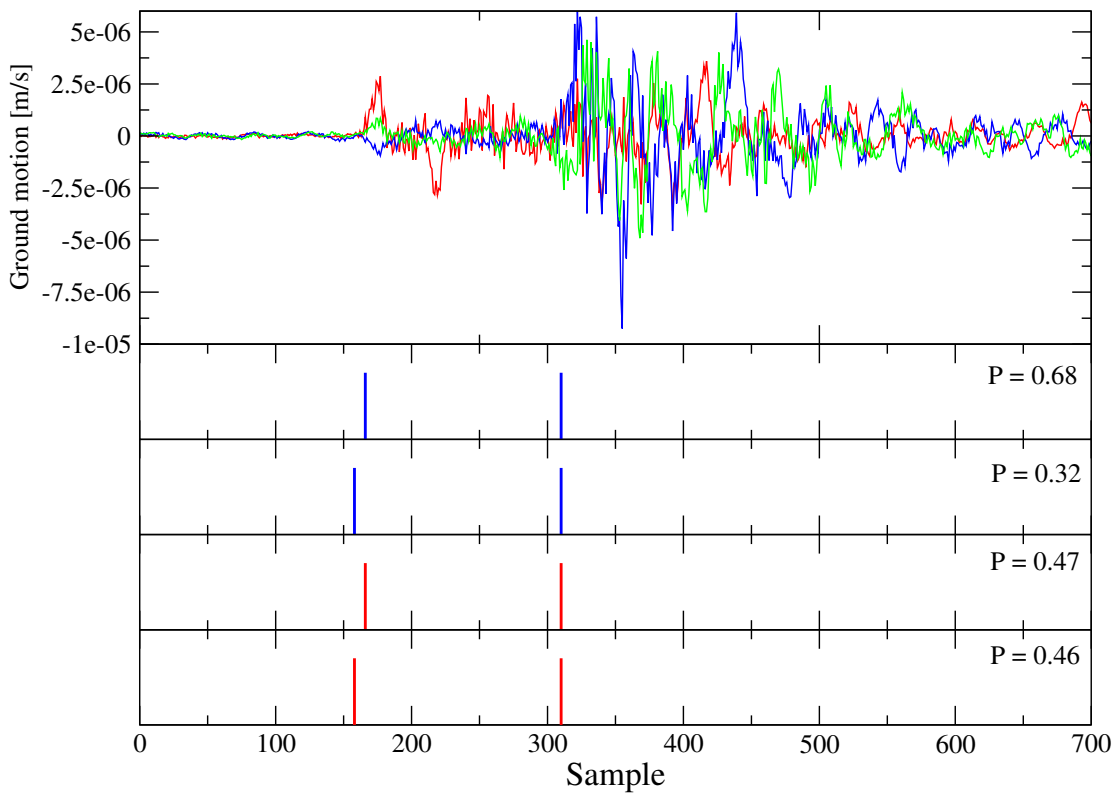
- The phase arrivals shown in the Figures were verified by human data quality control processors.



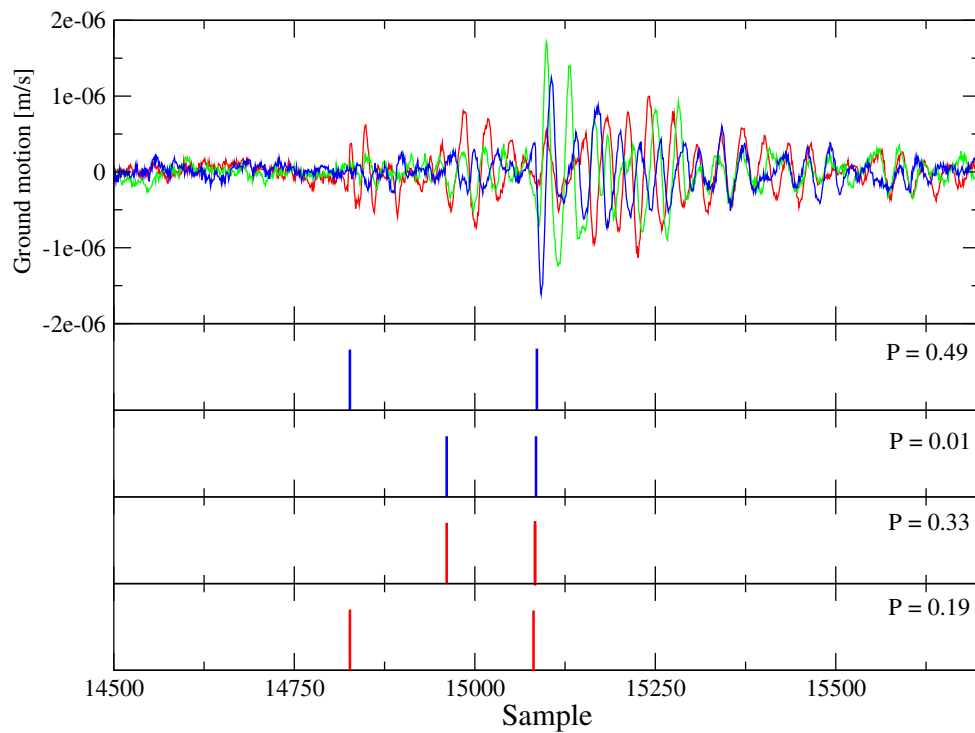
**Figure 4.16:** Example I of the single seismogram arrival time distribution (red bars) and the multiple seismogram arrival time distribution (blue bars). In this example there are multiple events from which one of the smaller seismic events was picked: of all the sensors that recorded the seismic event, this event was the dominant one. The mode of the single seismogram arrival time distribution does not contain the correct P-wave arrival.



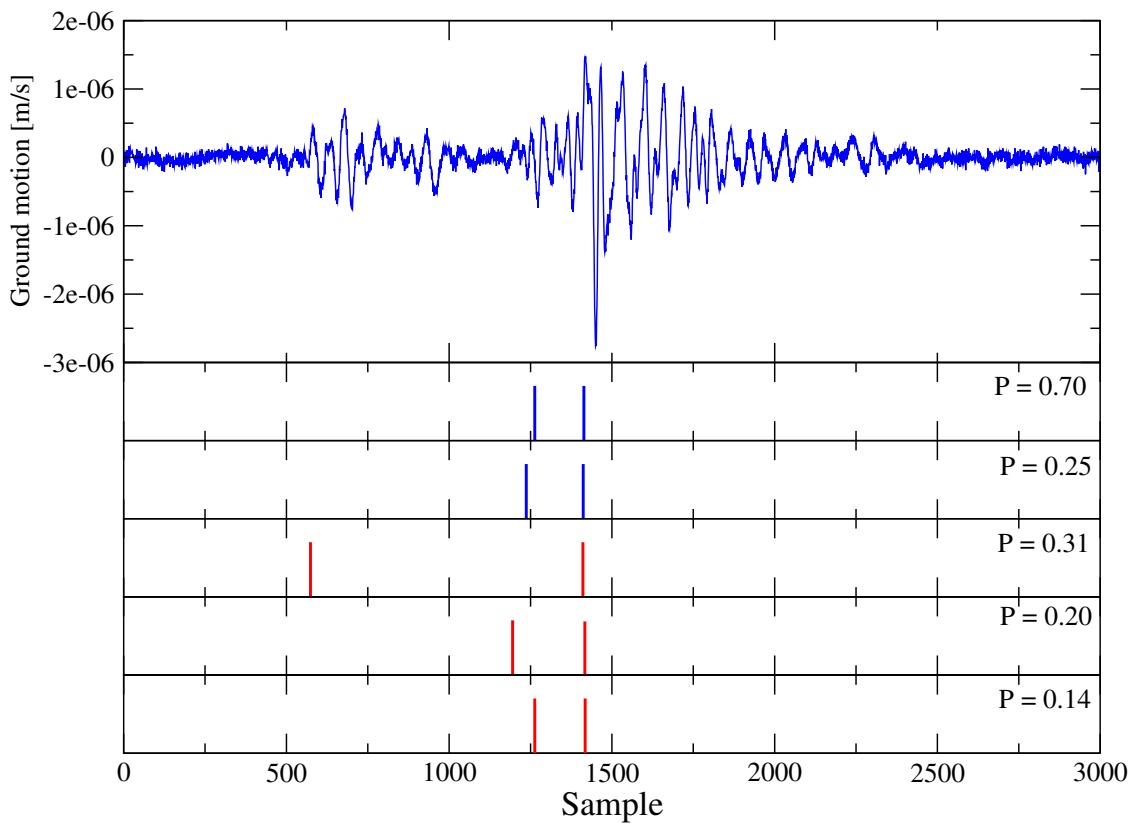
**Figure 4.17:** Example II of the single seismogram arrival time distribution (red bars) and the multiple seismogram arrival time distribution (blue bars). This is a relatively easy event in that the mode of the single seismogram arrival time distribution correctly contains both the true P-wave and true S-wave arrivals.



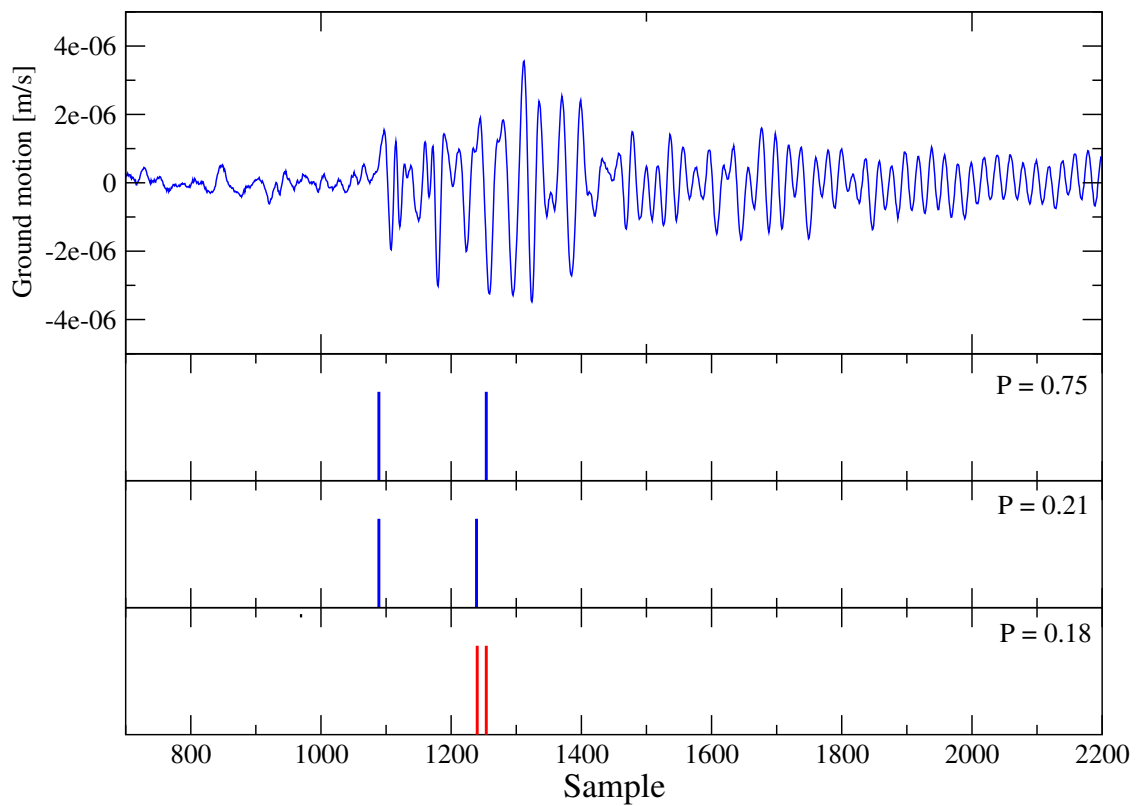
**Figure 4.18:** Example III of the single seismogram arrival time distribution (red bars) and the multiple seismogram arrival time distribution (blue bars). The mode of the single seismogram arrival time distribution contains the true S-wave arrival.



**Figure 4.19:** Example IV of the single seismogram arrival time distribution (red bars) and the multiple seismogram arrival time distribution (blue bars). The mode of the single seismogram arrival time distribution does not contain the P-wave arrival and corresponds to P-wave coda. The second most likely arrival pair of the single seismogram picker corresponds to the true P-wave and S-wave arrival.

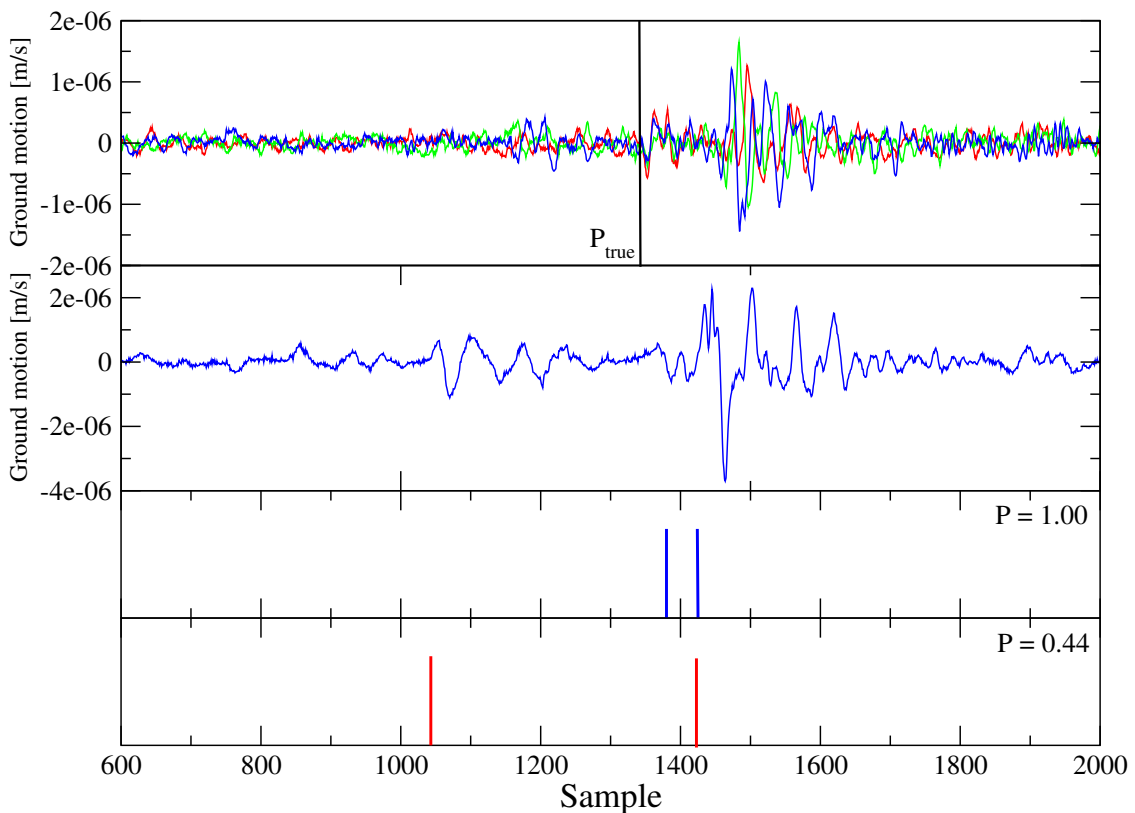


**Figure 4.20:** Example V of the single seismogram arrival time distribution (red bars) and the multiple seismogram arrival time distribution (blue bars). The mode at 600 of the single seismogram arrival time distribution is not the P-wave arrival and corresponds to noise.

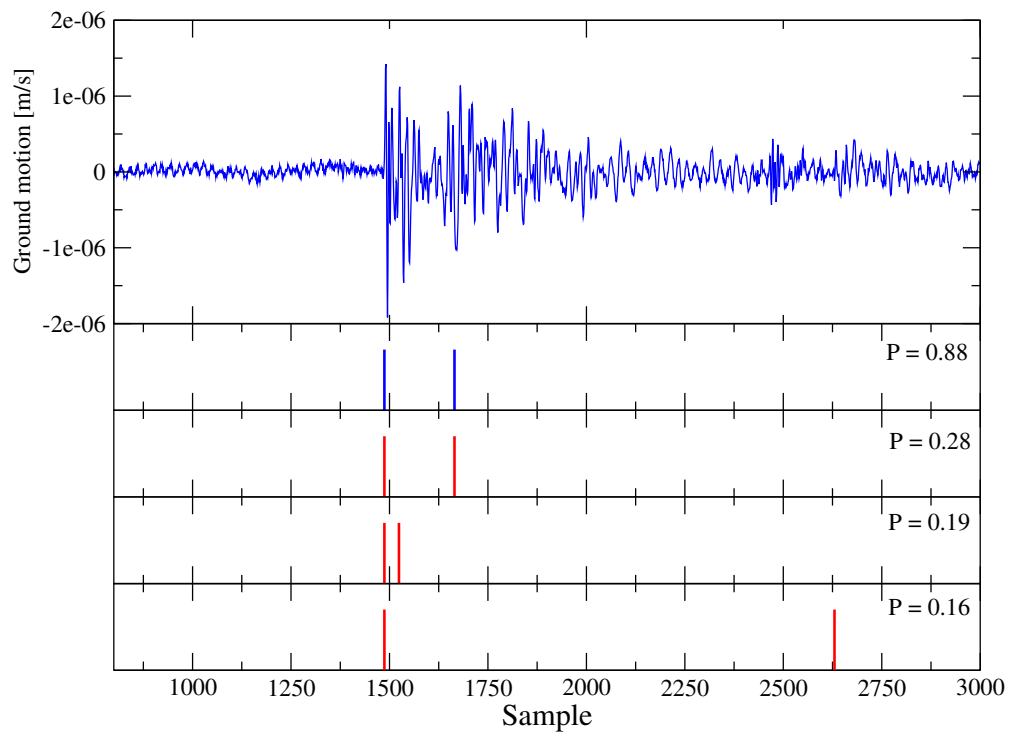


**Figure 4.21:** In this example the mode of the single seismogram arrival time distribution does not contain the P-wave arrival and corresponds to coda. Here the mode corresponds to an incorrectly identified P-wave arrival close to the true S-wave arrival.

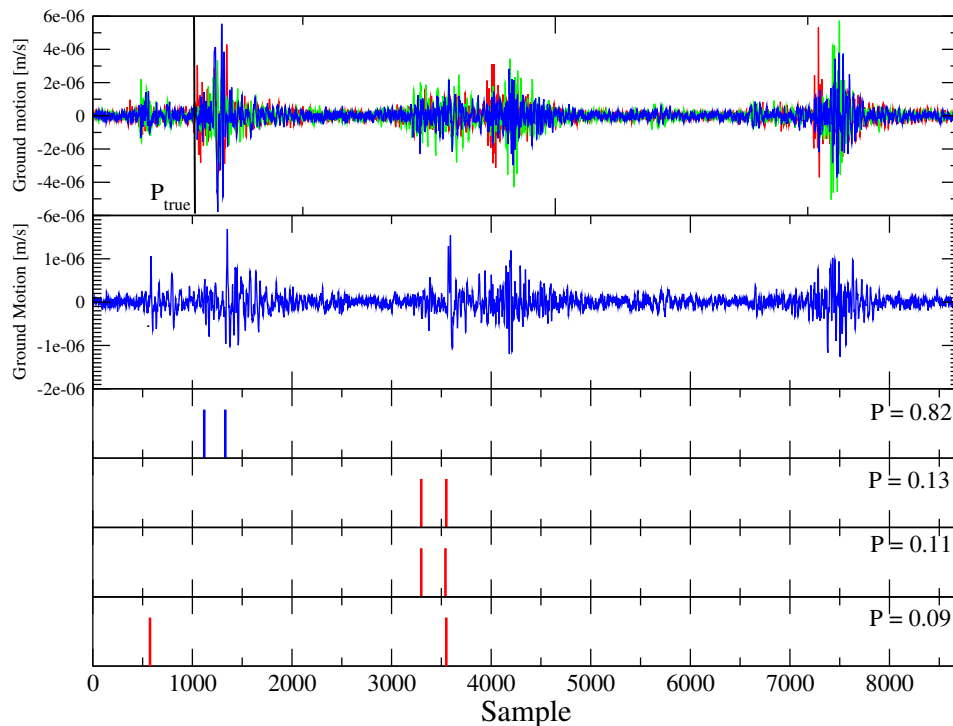




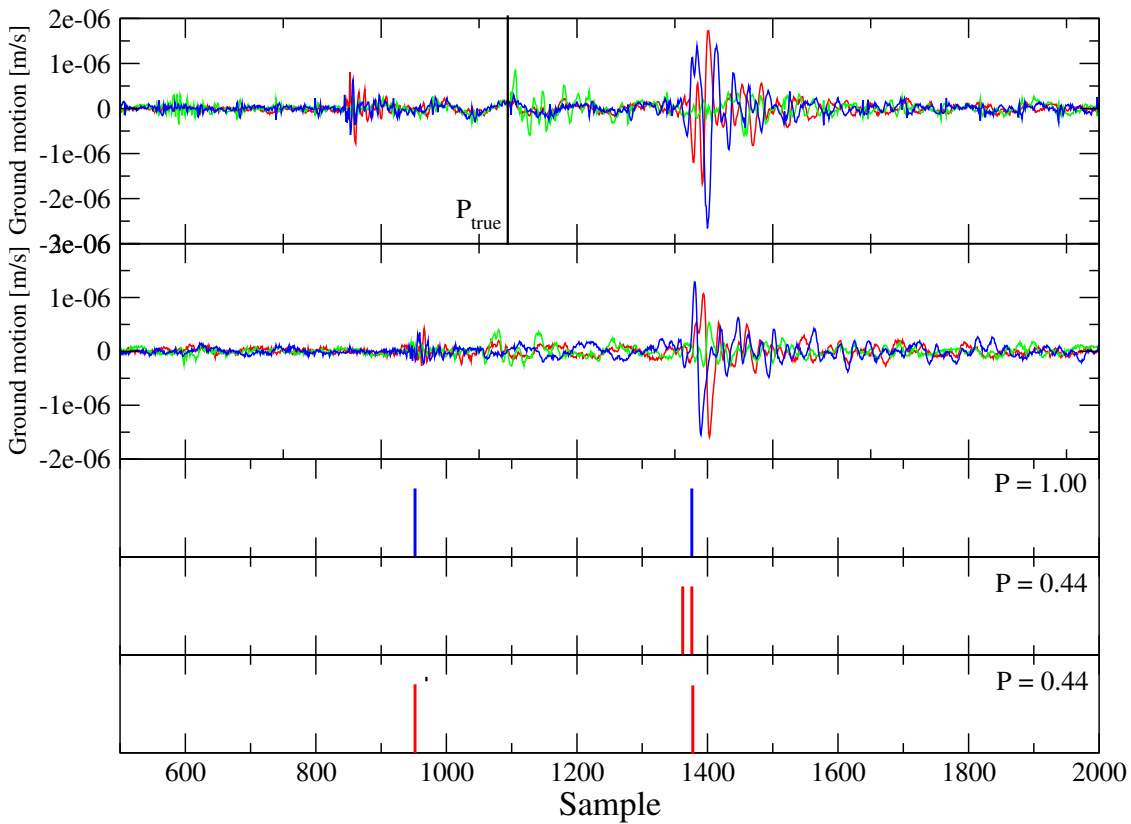
**Figure 4.22:** Example VII of a uni-axial waveform (second panel, blue) for which it is almost impossible to identify the correct P-wave arrival considering only this individual seismogram. The top panel shows the recorded seismogram of the same event by a different tri-axial sensor, where the true P-wave arrival is clear. By combining all the information from multiple seismograms, the automatic processor identified the correct P-wave and S-wave (blue bars).



**Figure 4.23:** In this example the mode of the single seismogram arrival time distribution contains the true P-wave arrival. The S-wave arrival is not clear, but the mode of the single and multiple seismogram arrival time distribution does however correctly identify the true S-wave arrival.



**Figure 4.24:** Example IX of the seismogram arrival time distributions. The top graph is the recorded seismogram from a different sensor, where it clearly shows the P-wave arrival. From the uni-axial it is not clear which event is the most energetic or which arrivals are correct. As mentioned in other examples, considering it as a multivariate problem, combining the information contained in all sensors is critical in solving this problem of identifying the correct arrivals on the uni-axial.



**Figure 4.25:** The top graph is the recorded seismogram from a different sensor, where it clearly shows the P-wave arrival. The correct P-wave arrival is almost totally hidden in the noise and not identified as a possible P-wave arrival. The nearest candidate P-wave arrival to the true arrival (leftmost blue bar) was chosen as the estimate of the P-wave arrival. Outlier detection techniques, as discussed in Section 4.9, are used to possibly identify the type of mistakes made by the algorithm in this example.

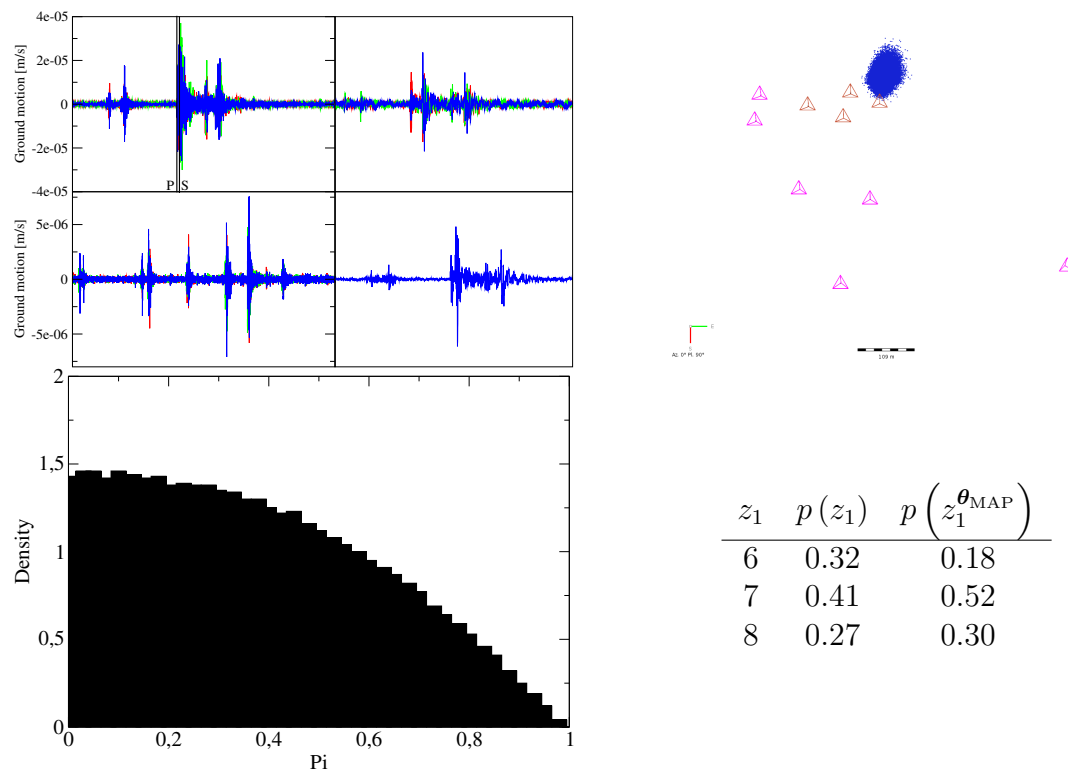
## 4.8 Examples of the Posterior Distributions

The posterior distribution of parameters can be obtained as described in Section 4.4. In Figures 4.26 to 4.29 the posterior distributions of  $\theta$ ,  $\pi_{1k}$  and  $z_1$  are shown for a few seismic events, where  $k$  corresponds to the mode of the arrival time of the nearest sensor. Note the difference in the shape of the marginal posterior distribution of  $\pi_{1k}$  depending on the number of modes of  $z_i$ . In general, if the posterior distribution of  $z_i$  has only one mode at the  $k$ th possible arrival pair, then the marginal posterior distribution of  $\pi_{ik}$  will be linearly increasing with a mode at 1 (with density 2) and a minimum value at 0 (with density 2); see Figures 4.27 and 4.28. Hence, the mode of the marginal posterior distribution of  $\pi_{ik}$  will be equal to the  $k$ th entry of the

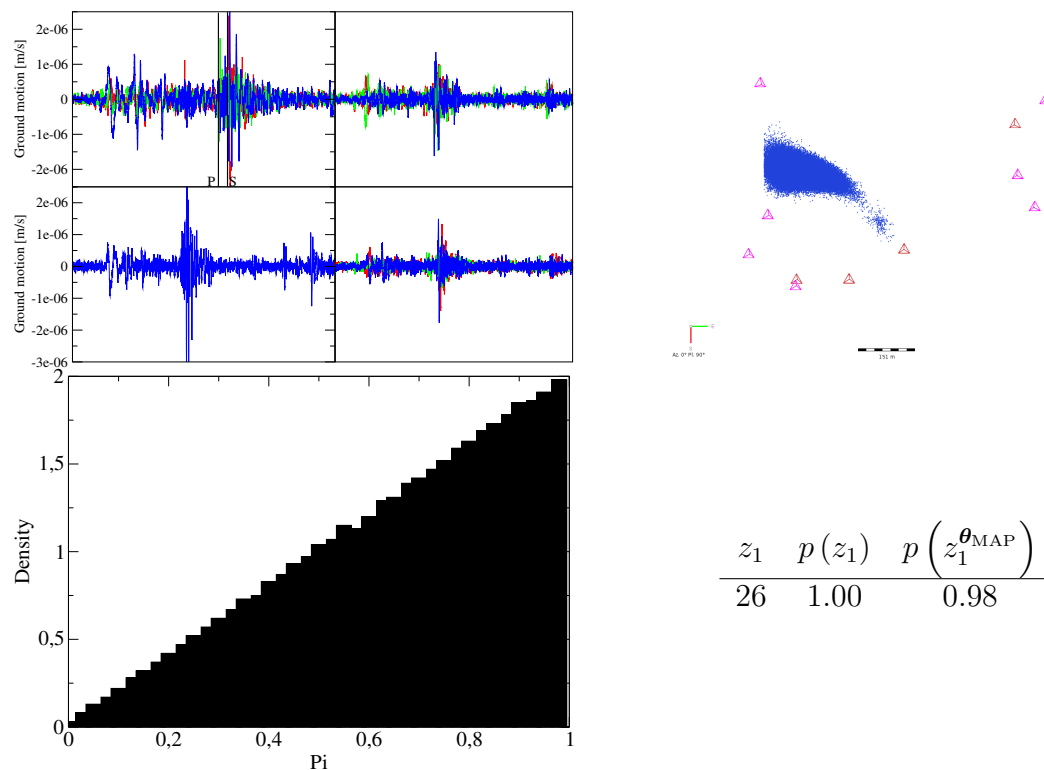
MAP estimate of  $\pi_i$ . When the posterior distribution of  $z_i$  has multiple modes, then the posterior distribution of  $\pi_{ik}$  can take on more general shapes, as shown by example in Figures 4.26 and 4.29.  $p(z_1|\theta_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \pi_i^{\text{prior}}, \rho_i)$  written shortly as  $p(z_1^{\theta_{\text{MAP}}})$  in the figures below are also given as reference and, as can be seen, are similar to the posterior probabilities of  $z_1$  in most cases.

I now summarise the figures and their properties as follows.

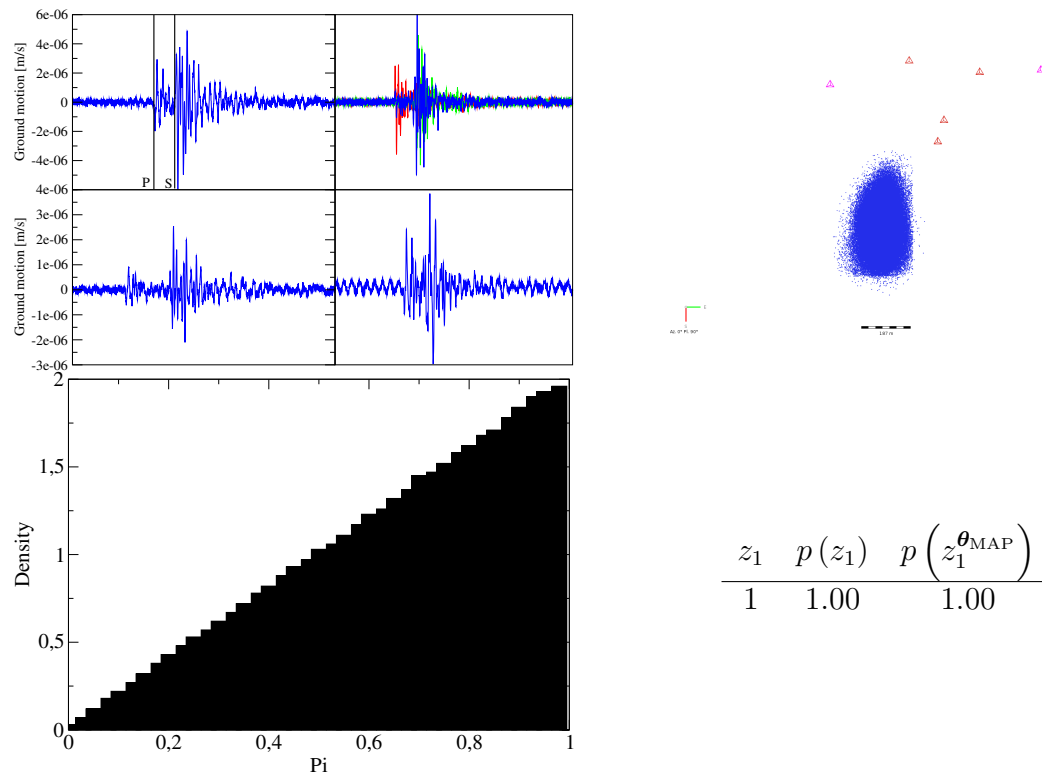
- These seismic events are real events that occurred on a mine serviced by IMS. The arrivals times were verified by experienced human processors. The top left panel shows the four seismograms that triggered (as explained in Section 2.5.1) first. I indicate the MAP arrivals of the first triggered seismogram (top left seismogram) with solid lines and P and S symbols at the bottom of the seismogram. The posterior distribution of the hypocentre (indicated by the blue dots) is shown in the top right panel. The pink and red triangles indicate the sensors that recorded the seismic events (the red triangles show the sensors of the seismograms in the top right panel). The posterior distribution of the mixing weight of the MAP arrival is shown in bottom left panel. Lastly, the posterior distribution of the arrival times (solid lines with P and S symbols of top left seismogram) of the nearest sensor is shown in the bottom right panel.
- The distance scale for Figure 4.26 equals 109 m, for Figure 4.27 the distance scale is 191 m, 197 m for Figure 4.28 and lastly, 67 m for Figure 4.29. Note the wide range of possible shapes of the posterior of the hypocentre and that the variance of the distribution tends to increase as the average hypocentre distance between the seismic event and the sensors increases. It also increases the less surrounded the seismic event is by seismic sensors.
- For very clear seismograms and only one signal in a buffer, the probability of the MAP of  $z_1$  equals one - for example Figure 4.28. In such cases the marginal posterior distribution of  $\pi_{ik}$  will be linearly increasing with a mode at 1 (with density 2) and a minimum value at 0 (with density 0). For more complex waveforms - for instance many signals in a buffer (Figure 4.26) or lower SNR (Figure 4.29) - there will be more than one mode in the posterior of  $z_1$ . In these cases, the marginal posterior distribution of  $\pi_{ik}$  deviates from linearly increasing with a mode at 1 (with density 2) and a minimum value at 0 (with density 0).
- Note how similar in most cases  $p(z_1|\theta_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \pi_i^{\text{prior}}, \rho_i)$  is to the posterior probabilities of  $z_1$ .  $p(z_1|\theta_{\text{MAP}}, \mathbf{t}_i, \mathbf{y}_i^{\text{P}}, \mathbf{y}_i^{\text{S}}, \pi_i^{\text{prior}}, \rho_i)$  is thus a good approximation of the posterior counterpart and has the added advantage of being quick to calculate.



**Figure 4.26:** Example I of the posterior distribution of the hypocentre (top right) indicated by blue dots, mixing weight of the MAP arrival (bottom left) and the arrival time at the nearest sensor (bottom right) of a seismic event with recorded seismograms shown in top left. The pink and red triangles in the top right figure indicate the sensors that recorded the seismic event. The red triangles show the sensors of the seismograms in the top right panel. Only the arrival time pairs that have non-zero posterior densities are shown in the bottom right table. There is more than one mode in the posterior of  $p(z_1^{\theta_{\text{MAP}}})$  and  $z_1$ ; hence the deviation in the marginal posterior distribution of  $\pi_{ik}$  from linearly increasing with a mode at 1 (with density 2) and a minimum value at 0 (with density 0). In this example the deviation is significant.

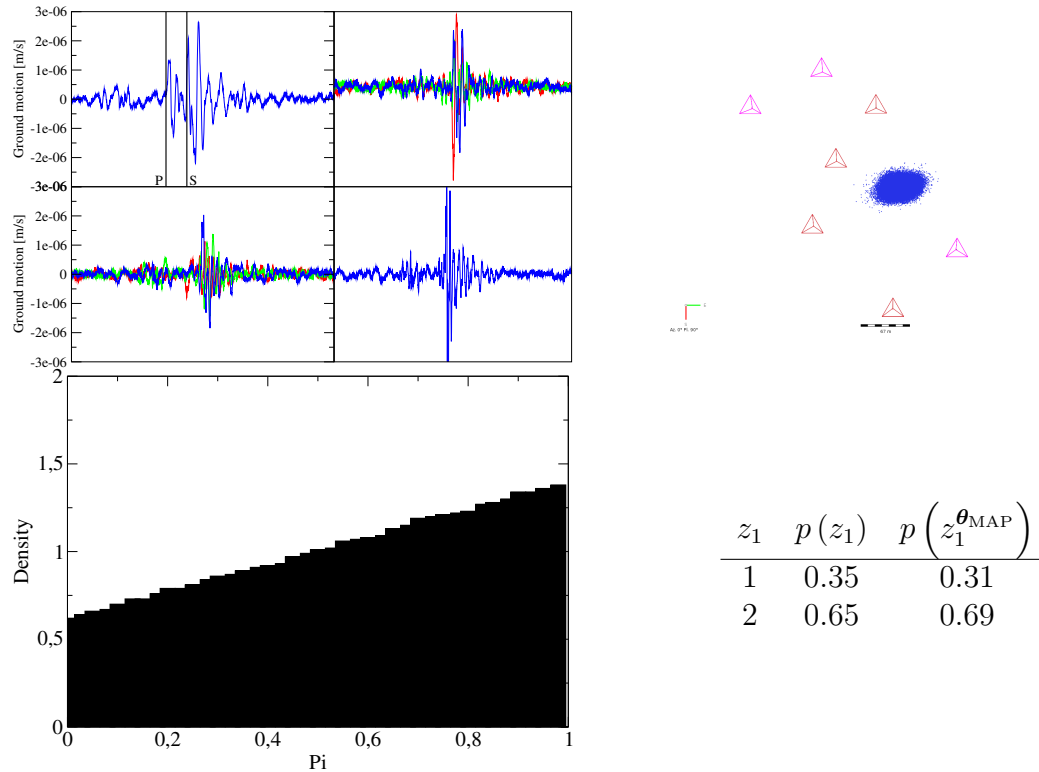


**Figure 4.27:** Example II of the posterior distribution of the hypocentre (top right) indicated by blue dots, mixing weight of the MAP arrival (bottom left) and the arrival time at the nearest sensor (bottom right) of a seismic event with recorded seismograms shown in top left. The pink and red triangles in the top right figure indicate the sensors that recorded the seismic event. The red triangles show the sensors of the seismograms in the top right panel. Only the arrival time pairs that have non-zero posterior densities are shown in the bottom right table. The probability of the MAP of  $z_1$  equals one; hence the marginal posterior distribution of  $\pi_{ik}$  is linearly increasing with a mode at 1 (with density 2) and a minimum value at 0 (with density 0).



**Figure 4.28:** Example III of the posterior distribution of the hypocentre (top right) indicated by blue dots, mixing weight of the MAP arrival (bottom left) and the arrival time at the nearest sensor (bottom right) of a seismic event with recorded seismograms shown in top left. Similar to Figure 4.27, the probability of the MAP of  $z_1$  equals one; hence the marginal posterior distribution of  $\pi_{ik}$  is linearly increasing with a mode at 1 (with density 2) and a minimum value at 0 (with density 0).





**Figure 4.29:** Last example of the posterior distribution of the hypocentre (top right) indicated by blue dots, mixing weight of the MAP arrival (bottom left) and the arrival time at the nearest sensor (bottom right) of a seismic event with recorded seismograms shown in top left. There is more than one mode in the the posterior of  $z_1$ ; interestingly, while the marginal posterior distribution of  $\pi_{i_k}$  is linearly increasing with a mode at 1 and a minimum value at 0, the densities are not equal to 2 and 0 respectively.

## 4.9 Outlier Detection

In the cases where the phase arrival MAP estimates may be incorrect the vast majority of those cases the algorithm would have got stuck in a local maximum. The MAP estimates that result from the local maximum will typically be due to a few outliers. To identify these outliers, the ACE algorithm is applied to the seismic event. The ACE algorithm as described below adheres to the same rules as followed by manual processors before a seismic event is allowed to be “saved” in the data base. The normalised distance residual must be below 3%, and at some mines all per site distance residuals must be below some threshold (typically 50 m). The average hypo-central distance is defined as  $AHD = (1/n) \sum d_i$  and the normalised average distance residual  $\text{res}_{\text{dist}}(\boldsymbol{\theta})$  is defined as

$$\text{res}_{\text{dist}}(\boldsymbol{\theta}) = \frac{1}{\text{AHD (in percent)}} \times \dots \quad (4.9.1)$$

$$100 \min \left( \sum_{i=1}^n \left\{ I(P_{\text{en},i}) v_i^P \left| t_i^P - t_0 - \frac{d_i(x,y,z)}{v_i^P} \right| + I(S_{\text{en},i}) v_i^S \left| t_i^S - t_0 - \frac{d_i(x,y,z)}{v_i^S} \right| \right\} \right).$$

A residual of 3% will typically result in a accurate hypocentre. A simple, yet effective, identification of accurate phase arrivals is achieved by using the following algorithm (ACE algorithm):

- Calculate all the site-specific P-wave and S-wave residuals.
- Obtain the maximum per-site residual.
- If the per site residual is larger than 50 m or the normalised residual larger than 0.03, disable the largest residual arrival and iterate; else exit.

The classification carried out in Chapter 3 did not use all arrival onset information in the decision-making process. With arrival time onset information available, post-pick classification can be carried out. The picking information is needed for verifying polarities of the P-wave and S-wave arrival onset, SNR and other types of sanity checks. This can also be used for post-pick blast discrimination as described in Section 3.2. P-picks and S-picks with low SNR are disabled. With the pick information, some single shot blasts can be identified by dominant frequency and maximum amplitudes, time duration and amplitude of the coda. Some waveforms are not seismic in nature, but missed by the HMM (as shown for example in Figures 3.18 and 3.21) and can similarly be identified and rejected.

Despite all these sanity checks, some visually obvious but incorrect automatically processed events may still survive. Even though an informative prior of hypocentres is used, in many cases the locations of these incorrectly processed seismic events are spatially far from any cluster. In our algorithm, such events which have an automatic location far from clusters should not be saved in the seismic

database using the automatic processing, but should rather be selected for human processing. This cluster check for outliers is most successful.

In addition, source parameter outliers (such as suspiciously high/low energy events given the moment, suspiciously high/low ground motion given the distance to sensor and suspiciously high/low moments given energy) should not be automatically processed and are left for human processing.

The Wadati analysis is one of the oldest in seismology (Wadati [61]) and useful in identifying unreliable phase arrivals. The difference between the modelled travel times of the P-wave of two sensors, assuming a homogeneous velocity medium, is  $\Delta t^P = \Delta d/v^P$  and similarly for the S-wave is  $\Delta t^S = \Delta d/v^S$ . From these equations it follows that

$$\Delta t^S = \frac{v^P v^S}{\Delta d} t^P, \quad (4.9.2)$$

which simply states a linear relationship between differences in S-wave travel times and P-wave travel times, with a slope equal to the ratio of the P-wave and S-wave velocity and intercept of zero. Automatically processed seismic events that have data points that differ statistically significantly from the expected line are not saved in the seismic database and are rather selected for human processing.

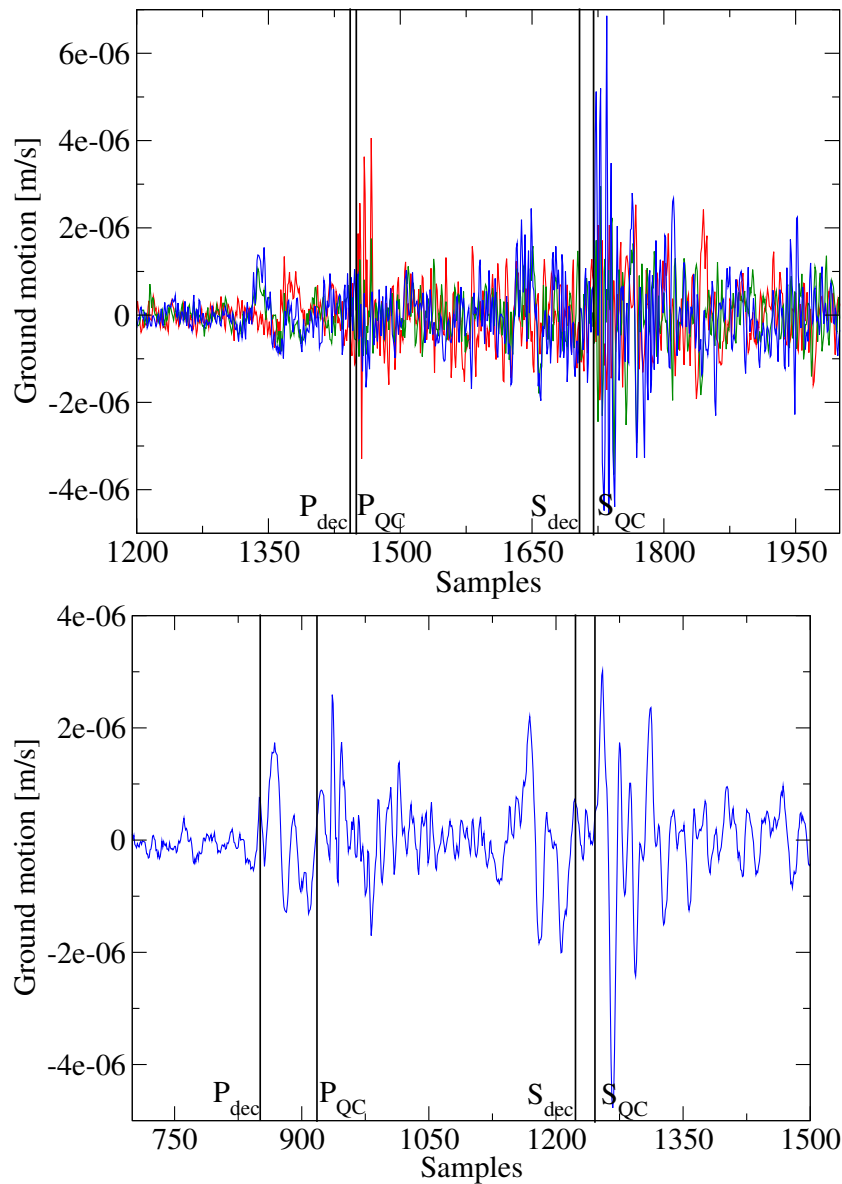
## 4.10 Results: Human Versus Automatic Processing

In this subsection, I consider how the algorithms developed in this dissertation compare to the current best available methods - those of normal and of elite quality control human pickers. I specifically compare the classification and picking accuracy of the automatic processor and decentralised human processors on the one hand to the corresponding processing of the quality control (QC) processors; see Subsection 2.8 for definitions.

Every event that is quality controlled has an initial score of 100 and marks are deducted for any misclassification (difference between original classification and QC classification) or any phase arrival differences (picking differences) that are outside the error bounds of the arrivals points. A score of zero is given when the original processing is rejected and the QC processing is accepted, while a score of 50 is given when the original processing is accepted and the QC processing is rejected. The deduction formula for incorrect P-wave picks for a seismogram  $i$  is  $(100/n)[1 - \exp(-|P_{dec,i} - P_{QC,i}|/\sigma_{P_{QC,i}})]$ , where  $n$  is the number of seismograms recorded by the sensors and  $\sigma_{P_{QC,i}}$  is a function of the SNR of the P-wave pick of the QC processor; the lower the SNR, the larger  $\sigma_{P_{QC,i}}$ . A similar formula holds for incorrect S-wave picks.

An example of two seismograms that were picked by a decentralised processor ( $P_{dec}$  and  $S_{dec}$ ) and a QC processor ( $P_{QC}$  and  $S_{QC}$ ) is shown in Figure 4.30. The

QC processor picked the seismograms correctly and the decentralised processor incorrectly. The S-wave pick by the decentralised processor in the seismogram in the top graph is too early, while the P-wave pick of the seismogram in the bottom graph is too early. Both these incorrect picks resulted in marks being deducted.



**Figure 4.30:** Comparison of arrival picks by normal decentralised human processors ( $P_{dec}$  and  $S_{dec}$ ) and quality control processors ( $P_{QC}$  and  $S_{QC}$ ). See text for discussion. For this specific event, the QC score was 61.92.

As the QC processing can sometimes be incorrect or where the classification of certain events are ambiguous (for example low SNR events), a quality control query (QCQ) system is in place, where all events that did not score 100 are double-checked by experienced in-house processors employed by IMS. Incorrectly classified events by the QC processor are corrected and ambiguous events are also given a score of 100.

For quantitative comparison of many events, I analysed a total of 5000 quality-controlled events sourced from the two highest activity mines which we shall refer to as Mine1 and Mine2. These events were recorded between 01 January 2018 and 01 January 2019. The QC scores before and after QCQ for the automatic processor and decentralised processing are shown in Table 4.2 for Mine1 and Table 4.3 for Mine2. Note that each decentralised processor processed a similar number of events, i.e. 71 events for each mine. At Mine1 the auto-processor can process 81.96% of data and at Mine2 80.90% of data.

**Table 4.2:** Comparison of human and automatic processing at Mine1. 'Auto' is the automatic algorithm developed in this dissertation and 'Human' is the decentralised human processors

|                      | Before QCQ |         | After QCQ |          |
|----------------------|------------|---------|-----------|----------|
|                      | Auto., %   | Human,% | Auto., %  | Human, % |
| QC score             | 87.25      | 85.87   | 96.43     | 95.50    |
| Classification score | 89.56      | 88.94   | 98.73     | 98.58    |
| Picking score        | 93.57      | 91.18   | 93.57     | 91.18    |

**Table 4.3:** Comparison of human and automatic processing at Mine2.

|                      | Before QCQ |         | After QCQ |          |
|----------------------|------------|---------|-----------|----------|
|                      | Auto., %   | Human,% | Auto., %  | Human, % |
| QC score             | 91.87      | 92.63   | 97.74     | 97.26    |
| Classification score | 92.73      | 94.08   | 98.60     | 98.69    |
| Picking score        | 94.33      | 91.26   | 94.33     | 91.42    |

As can be seen from the results in this section, the automatic processor has more or less the same QC score as the average decentralised processor and performs considerably better than the worst decentralised processor. The automatic processor has the added advantage of being much faster and cheaper.

## 4.11 Chapter Summary

Since this Chapter contains the essence of the original work done by me, and since it is so long, I end with a summary of its contents. A PGM-based method was developed that utilises the information contained in the single seismogram automatic picker(s) recorded at all sensors with the information contained in the hypocentre and origin time of the event. A single seismogram automatic picker that uses modified STA/LTA measurements to determine the candidate phase arrivals of a sensor was developed. The PGM is a Bayesian network and in particular a hierarchical Bayesian network. Using a probabilistic graphical model enables us to derive the joint probability density function of all the variables at play. Latent variables were introduced to model the phase arrivals. The mode of the latent variable distribution is a point-estimator of the phase arrivals.

Section 4.5 outlined how to obtain, with a high probability, the MAP estimates of the posterior distribution. Starting points for  $\theta$  that fall with a high probability into the basin of attraction of the joint distribution were discussed. I found analytical solutions for the arrival distribution where  $\theta_{\text{MAP}}$  is known. The solutions are quick to obtain and similar to the posterior distribution of the phase arrivals.

The effectiveness of the proposed PGM was demonstrated using examples. Techniques and conditional distributions to solve the statistical inference of this PGM were found.

Outliers, based on the processing of the automatic processor, as discussed in Section 4.9, are not saved in the seismic database, but are selected for human processing instead.

In Section 4.10, a comparison between the quality of the automatic processor and the decentralised (human) processing was given. The automatic processor has a classification accuracy and picking accuracy comparable to the average decentralised processor.



# Chapter 5

## Conclusion

Strong changes in the rate of seismic activity, so-called bursts, are due to a step-loading caused by production blasts or large events. When these bursts occur, it is critical to ascertain as quickly as possible where the mainshock(s) (main seismic event) and aftershocks have occurred with a view to order withdrawal of underground personnel. For re-entry purposes of workers, it is also important to know when the level of seismic activity has returned to an acceptable level. These bursts can occur at a rate of up to 100 times the normal background activity, in which case the number of human processors on shift during these times will typically not be sufficient. As human processing takes long and it may take hours to catch up, activity bursts will lead to longer delays before workers can re-enter the mine. Automatic processors are therefore a critical matter, in terms of safety and financial efficiency.

This dissertation documents my development of a successful automatic processor. The processor can classify (Chapter 3) and determine phase arrivals (Chapter 4) at the same level of quality as human processing, but is much faster. The phase arrival determination algorithm can be described by a Bayesian network (Figure 4.6) with a joint distribution given by Equation 4.3.1; this formula provides the most compact and accurate summary of my model. The automatic processor can also determine when it is unsure of a specific event and pass the event back to the decentralised human team (Section 4.9).

The method documented here can successfully process more than 50% of all recorded data at mines serviced by IMS, that is on average more than 20,000 recorded events per day. In addition to handling bursts, it has the added advantage of making processing costs cheaper for both IMS and its customers.

The automatic processor in its current state is constrained by business considerations as there was a need to produce an automatic processor that has a quality comparable to human processors as quickly as possible. There is future work to be carried out discussed below.



The HMM derived in this dissertation is linear and sequential. In deep learning, a recurrent neural network is also a sequential model, but has the advantage of being nonlinear and has long-term dependency. Also, convolutional neural networks have had much success in image classification. Such models have become very popular in recent years, but these deep learning models were not tested in this dissertation. Future work will include testing these new algorithms for classification of seismic data recorded at mines and compare the success rate (QC scores) to that of the HMM using the customised Viterbi algorithm as developed in this dissertation.

The MAP estimates of the Bayesian network should be the correct phase arrivals from a set of possible (or candidate) arrivals for each sensor. A simple STA/LTA algorithm was used to identify only candidate or possible arrivals for each sensor, the so-called single seismogram picker. To improve the current automatic processor, it would be helpful if the single seismogram picker could produce fewer false alarms. There is currently promising research being carried out in deep learning, for example such as convolutional neural networks described in Zhu and Beroza [62] and Wang *et al.* [63]. If these types of single seismogram pickers could reduce false alarms, far fewer local maxima and a smoother objective function would be achieved. The number of successfully processed seismic events will thus increase. This, too, is work for the future.

# List of References

- [1] Mendecki, A.J. (ed.): *Seismic Monitoring in Mines*. Chapman and Hall, 1997.
- [2] Allen, R.: Automatic earthquake recognition and timing from single traces. *Bulletin of Seismological Society of America*, vol. 68, pp. 1521–1532, 1978.
- [3] Allen, R.: Automatic phase pickers: their present use and future prospects. *Bulletin of Seismological Society of America*, vol. 72, pp. 225–242, 1982.
- [4] Baer, K. and Kradolfer, U.: An automatic phase picker for local and teleseismic events. *Bulletin of Seismological Society of America*, vol. 77, pp. 1437–1445, 1987.
- [5] Akaike, H.: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [6] Sleeman, R. and van Eck, T.: Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and Planetary Interiors*, vol. 113, pp. 265–275, 1999.
- [7] Leonard, M. and Kennett, B.L.N.: Multi-component autoregressive techniques for the analysis of seismograms. *Physics of the Earth and Planetary Interiors*, vol. 113, no. 247-264, 1999.
- [8] Leonard, M.: Comparison of manual and automatic onset time picking. *Bulletin of Seismological Society of America*, vol. 90, pp. 1384–1390, 2000.
- [9] Zhang, H., Thurber, C. and Rowe, C.: Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the Seismological Society of America*, vol. 93, pp. 1904–1912, 2003.
- [10] Cichowicz, A.: An automatic S-phase picker. *Bulletin of Seismological Society of America*, vol. 83, pp. 180–189, 1993.
- [11] Geldron, P., Ebel, J. and Manolakis, D.: Rapid joint detection and classification with wavelet bases via Bayes theorem. *Bulletin of Seismological Society of America*, vol. 90, pp. 764–774, 2000.

- [12] Ebel, J., Geldron, P. and Manolakis, D.: Development of a seismic event detection and identification algorithm based on wavelet transform. *Seismological Research Letter*, 1996.
- [13] Wang, J. and Teng, T.: Identification and picking of S-phase using an artificial neural network. *Bulletin of Seismological Society of America*, vol. 87, pp. 95–106, 1997.
- [14] Akazawa, T.: A technique for automatic detection of onset time of P- and S-phases in strong motion records. *13th World Conference on Earthquake Engineering*, vol. 13, no. 786, 2004.
- [15] Fah, D. and Koch, K.: Discrimination between earthquakes and chemical explosions by multivariate statistical analysis: A case study for Switzerland. *Bulletin of the Seismological Society of America*, vol. 92, pp. 1795–1805, 2002.
- [16] Tiira, T.: Discrimination of nuclear explosions and earthquakes from teleseismic distances with a local network of short period seismic stations using artificial neural network. *Physics of the Earth and Planetary Interiors*, vol. 97, pp. 247–268, 1996.
- [17] AllamehZahed, M.: Discrimination analysis of earthquakes and man-made events using arma ccoefficient determination by artificial neural networks. *Natural Resources Research*, vol. 4, pp. 367–375, 2011.
- [18] Kortstrom, J., Uski, M. and Tiira, T.: Automatic classification of seismic events within a regional seismograph network. *Computers and Geosciences*, vol. 87, 2015.
- [19] Giacco, F., Esposito, A., Scarpetta, S. and Marinaro, M.: Support vector machines and MLP for automatic classification of seismic signals at Stromboli volcano. *Frontiers in Artificial Intelligence and Applications*, vol. 204, pp. 116–123, 2009.
- [20] Beyreuther, M. and Wassermann, J.: Hidden semi-Markov model based earthquake classification system using weighted finite-state transducers. *Nonlinear Processes in Geophysics*, 2011.
- [21] Benitez, M.: Continuous HMM-based seismic-event classification at Deception island, Antarctica. *Geoscience and remote sensing*, vol. 45, no. 1, 2007.
- [22] Tarantola, A. and Valette, B.: Inverse problems = quest for information. *Journal of Geophysics*, vol. 50, pp. 159–170, 1982.
- [23] Martinsson, J.: Robust Bayesian hypocentre and uncertainty region estimation: the effect of heavy-tailed distributions and prior information in cases with poor, inconsistent and insufficient arrival times. *Geophysical Journal International*, vol. 192(3), pp. 1156–1178, 2012.
- [24] Keylis-Borok, V.I.: On estimation of the displacement in an earthquake source and of source dimensions. *Annali di Geofisica*, vol. 12, no. 2, pp. 205–214, 1959.

- 
- [25] Brune, J.N.: Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research*, vol. 75, no. 26, pp. 4997–5009, September 1970.
- [26] Mendecki, A.J.: Real time quantitative seismology in mines: Keynote Address. In: Young, R.P. (ed.), *Proceedings 3rd International Symposium on Rockbursts and Seismicity in Mines, Kingston, Ontario, Canada*, pp. 287–295. Balkema, Rotterdam, 1993.
- [27] Koyama, J.: *The Complex Faulting Process of Earthquakes*. Springer Science & Business Media, 2013.
- [28] Hilbert, D.: Grundzuge einer allgemeinen Theorie der linearen Integralgleichungen. *Chelsea Pub. Co.*, 1953.
- [29] Lee, W. and Stewart, S.: Principles and applications of microearthquake networks. *Academic Press, New York*, 1981.
- [30] Lawrence, D.: Seismicity in the Orange Free State gold-mining district. In: Gay, N. and Wainwright, E. (eds.), *Proceedings of the 1st International Congress on Rockbursts and Seismicity in Mines*, vol. 6, pp. 121–130. S. Afr. Inst. Min. Metal., Johannesburg, 1984.
- [31] Mood, A. and Graybill, F.: *Introduction to the Theory of Statistics*. McGraw-Hill International, 1974.
- [32] Olsson, D. and Nelson, L.: The Nelder-Mead simplex procedure for function minimization. *Technometrics*, vol. 17, 1975.
- [33] Storn, R. and Price, K.: Differential evolution, a simple and efficient adaptive scheme for global optimization over continuous spaces. *Technical Report TR-95-012* downloadable from <http://www.icsi.berkeley.edu/techreports/1995.abstracts/tr-95-012.html>, 1995.
- [34] Mendecki, A.J. and Sciocatti, M.: Location of seismic events. In: *Seismic Monitoring in Mines*. Chapman and Hall, London, 1997.
- [35] Omori, F.: On the aftershocks of earthquakes. *Journal of the College of Science*, vol. 7, pp. 111–200, 1894.
- [36] Dong, L., Wesseloo, J., Potvin, Y. and Li, X.: Discrimination of mine seismic events and blasts using the Fisher classifier, naive Bayesian classifier and logistic regression. *Rock Mechanics and Rock Engineering*, vol. 45, 2015.
- [37] Malovichko, D.: Discrimination of blasts in mine seismology. In: *Deep Mining*. Australian Centre for Geomechanics, 2012.
- [38] Bormann, P. (ed.): *NEW Manual of Seismological Observatory Practice*. Electronically published, 2002.

- [39] Baum, L. and Eagon, J.: An inequality with applications to statistical prediction for function of Markov processes and to model for ecology. *Bulletin of American Mathematical Society*, vol. 73, pp. 360–363, 1963.
- [40] Baum, L., Petrie, T., Soules, G. and Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, vol. 41, no. 164-171, 1970.
- [41] Viterbi, A.: Error bounds for convolution codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.
- [42] Pearl, J.: Reverend Bayes on inference engines: A distributed hierarchical approach. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence.*, pp. 133–136, 1983.
- [43] Weiss, Y. and Freeman, T.: On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 736–744, February 2001.
- [44] Dempster, A., Laird, N. and Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Series B.*, vol. 39, no. 1, pp. 1–38, 1977.
- [45] Boschetti, F., Dentith, M. and List, R.: A fractal-based algorithm for detecting first arrivals on seismic traces. *Geophysics*, vol. 61, no. 4, pp. 1095–1102, 1996.
- [46] Martinsson, J.: Estimating the underlying signal waveform, noise co- variance and synchronization jitter from unsynchronized measurements. *Measurement Science and Technology*, vol. 19, 2008.
- [47] Chen, Z. and Stewart, R.: Multi-window algorithm for detecting seismic first arrivals. *CSEG National Convention*, 2005.
- [48] David, B.J. (ed.): *Bayesian Networks: An Introduction*. Wiley, 2009.
- [49] Buntine, W.: Operations for learning with graphical models. *Journal of Artificial Intelligence Research 2*, pp. 159–225, 1994.
- [50] Kwiatek, G. and Ben-Zion, Y.: Assessment of P and S wave energy radiated from very small shear-tensile seismic events in a deep south african mine. *Journal of Geophysical Research*, vol. 118, no. 7, pp. 3630–3641, July 2013.
- [51] Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
- [52] Myers, S., Johannesson, G. and Hanley, W.: Incorporation of probabilistic seismic labels into a Bayesian multiple-event seismic locator. *Geophysics Journal International*, vol. 177, pp. 193–204, 2009.

- 
- [53] Koller, D.: *Probabilistic Graphical Models Principles and Techniques*, chap. The Bayesian Network Representation, pp. 45–102. MIT Press, 2009.
- [54] Du Toit, C. and Lynch, R.: Study of seismic array sensitivity analysis, Kiruna mine, LKAB, 2011.
- [55] Neal, R.: Probabilistic inference using Markov Chain Monte Carlo methods. Tech. Rep., Department of Computer Science, univeristy of Toronto, 1993. Available at: <http://www.cs.utoronto.ca/~radford/ftp/review.pdf>
- [56] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E.: Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.
- [57] Hastings, W.: Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, vol. 57, pp. 97–109, 1970.
- [58] Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 721–741, 1984.
- [59] Geyer, C.: Introduction to Markov Chain Monte Carlo. In: Brooks, S., Gelman, A., Jones, G. and Meng, X. (eds.), *Handbook of Markov Chain Monte Carlo*, chap. 1. Chapman & Hall, 2011.
- [60] Liu, J.: The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, vol. 89, pp. 958–966, 1994.
- [61] Wadati, K.: On the travel time of earthquake waves. *Geophysics Magazine*, vol. 7, pp. 101–111, 1933.
- [62] Zhu, W. and Beroza, G.: Phasenet: A deep-neural-network-based seismic arrival time picking method. *Geophysical Journal International*, 2018.
- [63] Wang, J., Zhuowei, X., Change, L. and Zhenxing, Y.: Deep-learning for picking seismic arrival times. *JGR Solid Earth*, 2019.