

**FIRST ARRIVAL PICKING USING TWO
DIFFERENT METHODS: MDPE AND DBSCAN**

BY

Abdullah Abobakr Al-Mashhor

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

GEOPHYSICS

November 2020

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by Abdullah Abobakr Al-Mashhor under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN GEOPHYSICS.**



Prof. Abdullatif Al-Shuhail
(Advisor)



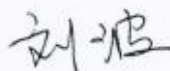
Dr. John Humphrey
Department Chairman



Dr. Sherif Mahmoud
(Co-Advisor)



Dr. Suliman S. Al-Homidan
Dean of Graduate Studies



Dr. Bo Liu
(Member)

Date

© Abdullah Abobakr Al-Mashhor

2020

| I dedicate my work to my parents, brother, and sisters |

ACKNOWLEDGMENTS

I begin with the name of Allah, the most beneficent, the most merciful. May Allah bestow peace on our beloved Prophet Mohammed (peace and blessings of Allah be upon him), his family, his noble companions, and all those who follow them with righteousness until the Day of Judgment. I would like to thank the Almighty who gave me strength, guidance, and patience to complete this thesis.

I am very grateful to my family; my parents, my brother, and my sisters for their support, patience, prayers, and encouragement during my academic journey.

I also would like to express my deepest appreciation to my advisor and academic mentor, Prof. Abdullatif Al-Shuhail, for his valuable and fatherly guidance, suggestions, supervision, knowledge, and support that he has provided me in my thesis work and other projects throughout my study at KFUPM. I thank Dr. Sherif Mahmoud for his expertise support, suggestions, and for his unforgettable kindness during the period of this thesis preparation. I thank Dr. Bo Liu for his contribution, encouragement, and useful suggestions.

Finally, I thank all faculty and staff of the Geosciences Department at King Fahad University of Petroleum & Minerals for their support and encouragement through my enjoyable learning experience. I acknowledge the generosity of the Saudi Arabian Government, through the Ministry of Higher Education, for providing a scholarship to complete my master's program.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES.....	IX
LIST OF FIGURES.....	X
ABSTRACT	XIV
ملخص الرسالة	XVI
CHAPTER 1 INTRODUCTION.....	1
1.1 Introduction	1
1.2 Hilbert Transform and The Complex Seismic Trace	7
1.3 Machine Learning	8
1.4 Clustering Algorithm.....	9
1.5 Statement of The Problem.....	10
1.6 Objective	11
1.7 Justification of The Study.....	12
1.8 Benefits of The Study.....	12
1.9 Structure of Thesis	13
CHAPTER 2 LITERATURE REVIEW	15
2.1 Previous Studies of First Arrival Traveltime Picking	15
2.2 History of The Complex Seismic Trace	18
CHAPTER 3 METHODOLOGY.....	21
3.1 First Arrival Traveltime Picking Based on MDPE Method	21

3.1.1	Hilbert Transform and Envelope	21
3.1.2	Sliding Window Median	25
3.1.3	Maximum Difference Between Adjacent Elements (MDPE)	25
3.2	First Arrival Traveltime Picking Based on DBSCAN Method	27
3.2.1	Selecting A Sliding Window	27
3.2.2	Linear Moveout Shift (LMO) And Stacking	28
3.2.3	DBSCAN Clustering Algorithm.....	33
CHAPTER 4 RESULT AND DISCUSSION		36
4.1	First Arrival Traveltime Picking Based on MDPE Method	36
4.1.1	Results.....	36
4.1.2	Discussion.....	57
4.2	First Arrival Traveltime Picking Based on DBSCAN Method	59
4.2.1	Results.....	59
4.2.2	Discussion.....	71
CHAPTER 5 CONCLUSION AND RECOMMENDATIONS.....		73
5.1	First Arrival Traveltime Picking Based on MDPE Method	73
5.1.1	Conclusion	73
5.1.2	Recommendations	74
5.2	First Arrival Traveltime Picking Based on DBSCAN Method	75
5.2.1	Conclusion	75
5.2.2	Recommendations	76
REFERENCES.....		77
VITAE		83

LIST OF TABLES

Table 4.1: Minimum, maximum, mean, median, and standard deviation errors in percentage for all synthetic and real data sets used in both MDPE and DBSCAN methods.	42
--	----

LIST OF FIGURES

Figure 1.1: Model of the near-surface before applying static correction shows a seismic raypath between the source (S) and receiver (R _n) with two underlying layers.	2
Figure 1.2: Model of the near-surface after applying static correction where the source (S') and receive (R' _n) are placed on the datum after removing the near-surface effects. ...	3
Figure 1.3: A seismic wave with an indication of the first-arrival time, t_0 , location.	6
Figure 2.1: The trace envelope (blue) of a real seismic trace (black) and imaginary seismic trace (red) is the maximum value that the trace can achieve.	20
Figure 3.1: A complex seismic trace $z(t)$ that contains a real part $x(t)$, which is the real seismic trace, and an imaginary part $y(t)$, which is calculated by the Hilbert transform of the real seismic trace (Taner et al., 1979).	24
Figure 3.2: Workflow of the MDPE method for picking the first arrival.	26
Figure 3.3: (a) A synthetic seismic gather with a linear event. LMO shift of the traces using: (b) optimum velocity, (c) high velocity, (d) low velocity. (e) a stacked gather for attempted velocity values where the maximum amplitude (back circle) output is achieved at the optimum velocity. Color indicates normalized amplitude values of stacked traces.	32
Figure 3.4: A data set that is divided into different clusters after applying DBSCAN clustering algorithm. Each color represents one cluster and outliers are shown in back color (Kassambara, 2017).	35
Figure 4.1: Velocity model used to generate the synthetic data set. Color bar shows velocities for the three different layers (400-2800 m/s).	38
Figure 4.2: Sample shot gathers from the synthetic seismic data: (a) Shot gather 1; (b) Shot gather 56.	39
Figure 4.3: The same synthetic shot records shown in Figure 4.2 showing first arrival picks from the MDPE method (red crosses) and manual picks (blue circles): (a) Shot gather 1; (b) Shot gather 56. (c) Comparison of the MDPE method (dash-dot line) and STA/LTA method (dashed line) versus manual picking (solid line) for trace 110 of shot gather 1.	40

Figure 4.4: Shot gather 1 from the same synthetic seismic data shown in Figure 4.2 (a) showing first arrival picks from the MDPE method (red crosses) and manual picks (blue circles) after adding random Gaussian noise with zero mean and 0.01 standard deviation to the raw amplitudes.	41
Figure 4.5: Sample shot gathers from the real data set from southeast Texas: (a) shot gather 7; (b) shot gather 17.	45
Figure 4.6: Sample traces from the real data set from southeast Texas: (a) trace 31 from shot 7; (b) trace 29 from shot 2.	46
Figure 4.7: The envelope of the traces in Figure 4.6. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.	47
Figure 4.8: The envelope in decibel units for the traces in Figure 4.6. Note the spike near the first arrival time position. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.	48
Figure 4.9: Median smoothing with a 50-sample sliding window of the envelope for the traces in Figure 4.6. Note the smoothness near the first arrival. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.	49
Figure 4.10: Difference between adjacent points of the envelope after applying moving median smoothing. The first arrival is always picked at the point with maximum difference. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.	50
Figure 4.11: Comparison of the MDPE method (dash-dot line) and P-phase picker method (dotted line) for the traces in Figure 4.6. The manual picking is shown by the dashed line. (a) Application on this trace gives almost the same accuracy for the two compared method; (b) Application on this trace shows more agreement between the MDPE method and manual picking than the P-phase picking method.	51
Figure 4.12: Results of applying the proposed MDPE method on 54 selected traces from the southeast Texas data set. Trace (T) and shot (S) numbers are indicated on the x-axis while the first arrival travelttime position of each trace is indicated on the y-axis. The blue filled circles indicate first arrival estimated by the MDPE method on each of the traces indicated on the x-axis while the error bar indicates its deviation from manual	

picking. The red concentric circle represents the median first arrival traveltime of all 54 traces together with the median error from manual picks.....	52
Figure 4.13: Results of applying the P-phase picking method on the same 54 traces shown in the previous figure. The blue filled circles indicate first arrival estimated by the P-phase picking method on each of the traces (T) indicated on the x-axis while the error bar indicates its deviation from manual picking. The red concentric circle in the middle of Figure represents the median error from manual picks. Note the high errors compared to the proposed method.	53
Figure 4.14: Raw shot records selected from Yilmaz (2001). (a) Shot record 6; (b) Shot record 25.	55
Figure 4.15: The same two Yilmaz's shot records shown in Figure 4.14 with a comparison of first arrival picks between the MDPE method (crosses) and the manual picks (circles). (a) Shot gather 6; (b) Shot gather 25.	56
Figure 4.16: Velocity model used to generate the DBSCAN synthetic data set.....	61
Figure 4.17: Sample shot gather generated from the synthetic seismic data.	62
Figure 4.18: Sample selected window with a size of 10 traces wide and 30 samples long.	63
Figure 4.19: Stacked gather for the selected window (Figure 4.18) after applying a linear-moveout (LMO) shift for each velocity value (200-800 m/s).	64
Figure 4.20: Stacked gather (Figure 4.19) after applying 50% threshold to remove low amplitude output energies.	65
Figure 4.21: The output of the DBSCAN clustering algorithm with one cluster. The centroid (dot symbol) represents the optimum velocity value and the first arrival traveltime pick.	66
Figure 4.22: The same shot gather shown in Figure 4.17 with a comparison of first arrival picks between the DBSCAN method (crosses) and the true picks (circles).....	67
Figure 4.23: A comparison between the manual (red) and the DBSCAN method (blue) picking.....	68
Figure 4.24: The differences in picking between the manual and DBSCAN method.	69

Figure 4.25: Results of applying the DBSCAN method on the synthetic shot gather (Figure 4.17) traces. Filled circles indicate first arrival estimated by the DBSCAN method while the error bar indicates its deviation from manual picking. 70

|

ABSTRACT

Full Name : [Abdullah Abobakr Al-Mashhor]

Thesis Title : [First Arrival Traveltime Picking Using Two Different Methods:
MDPE and DBSCAN]

Major Field : [Geophysics]

Date of Degree : [Nov. 2020]

First arrival traveltime is essential to be picked accurately to enhance the efficiency of the static correction methods for both reflection and refraction data processing. The purpose of this thesis is to apply two new techniques of first arrival traveltime picking on both real and synthetic data to evaluate the advantages and limitations of each technique. The first proposed method is based on the maximum difference between adjacent points of the envelope (MDPE) of a seismic trace. This method starts by calculating the Hilbert transform of the seismic trace to generate a complex seismic trace, which consists of the real and imaginary parts. After generating the complex seismic trace, the instantaneous amplitude or the envelope is extracted and then used for picking the first arrival traveltime. We first test the MDPE method on synthetic shot records and then compare the automatically picked first arrival traveltimes based on it with those picked manually by an experienced processor. Furthermore, we perform the same test on two real shot records (6 and 25) from Yilmaz's published assortment of worldwide real seismic data. Finally, we apply the MDPE method on 58 traces extracted from a real 2D land seismic data set from southeast Texas and compare the MDPE method and the P-phase picker method against manual picks.

The second method is based on the density-based spatial clustering of applications with noise (DBSCAN). This proposed method starts by selecting a sliding window with a certain number of traces (or wavelengths) wide and samples (or periods) long. Then, a linear moveout (LMO) shift is applied for a range of velocities to compute a single time shift for each trace within the selected window and remove the dipping of first arrival events. After that, all shifted traces for each velocity value are stacked to get the output energy where the maximum output energy value is the location of the optimum velocity for the selected window. Finally, DBSCAN clustering algorithm is applied to divide the stacked gather into different clusters and select the centroid point of the maximum cluster to be the optimum apparent velocity and the first arrival traveltime location. The DBSCAN method is tested on a synthetic shot record and then compared with those picked manually by an experienced processor. |

ملخص الرسالة

الاسم الكامل: عبدالله أبوبكر عبدالله علي المشهور

عنوان الرسالة: انتقاء وقت اشارة الوصول الأولى للبيانات الزلزالية باستخدام طريقتين: MDPE و DBSCAN

التخصص: الجيوفيزياء

تاريخ الدرجة العلمية: نوفمبر 2020

يعد وقت اشارة الوصول الأولى للبيانات الزلزالية ضروريًا ليتم انتقاؤه بدقة لتعزيز كفاءة طرق التصحيح الزمني للبيانات الزلزالية لمعالجة كلا من بيانات الانعكاس والانكسار. الغرض من هذه الأطروحة هو تطبيق تقنيتين جديدتين لانتقاء وقت الوصول الأولي على كل من البيانات الحقيقية والاصطناعية لتقييم مزايا وقيود كل تقنية. الطريقة الأولى المقترحة تعتمد على الحد الأقصى للاختلاف بين النقاط المجاورة لغلaf التسجيل الزلزالي (MDPE). هذه الطريقة تبدأ بحساب تحويل هيلبرت للتسجيل الزلزالي لتوليد تسجيل زلزالي معقد يتكون من الجزء الحقيقي والخيالي. بعد إنشاء التسجيل الزلزالي المعقد، يتم استخراج السعة اللحظية (أو غلاف التسجيل الزلزالي) ثم استخدامها لانتقاء وقت اشارة الوصول الأولى. يتم اختبار طريقة MDPE على سجلات البيانات الاصطناعية ومن ثم مقارنتها مع أوقات اشارة الوصول الأولى المنتقاة بشكل يدوي. بعد ذلك، تم اختبار الطريقة نفسها على بيانات حقيقية وتم عمل مقارنة بين كل من طريقة MDPE و P-phase picker مع أوقات اشارة الوصول الأولى المنتقاة بشكل يدوي. الطريقة الثانية المقترحة تعتمد على خوارزمية التجميع DBSCAN. تبدأ هذه الطريقة باختيار نافذة متحركة بها عدد معين من التسجيلات الزلزالية وعدد من العينات. بعد ذلك، يتم تطبيق طريقة فرق أزمنة الوصول الخطية (LMO) على نطاق محدد من السرعات لإزاحة التسجيلات الزلزالية وإزالة الانحدار الخطي لأحداث اشارة الوصول الأولى. بعد ذلك، يتم تجميع كل التسجيلات الزلزالية التي تمت ازاحتها باستخدام LMO لكل قيمة سرعة حيث إن القيمة القصوى للطاقة الناتجة عبارة عن موقع السرعة المثالية داخل النافذة المتحركة. أخيرًا، يتم استخدام خوارزمية DBSCAN لتقسيم البيانات داخل النافذة المتحركة وتحديد النقطة الوسطى داخل المجموعة الكبرى كموقع لإشارة وقت اشارة الوصول الأولى. تم تطبيق طريقة DBSCAN على سجلات البيانات الاصطناعية ومن ثم تم مقارنتها مع أوقات اشارة الوصول الأولى المنتقاة بشكل يدوي.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Static correction is considered as an important step during processing of land seismic data. To understand the objective of static correction, Figure 1.1 depicts a raypath from source S to receiver R_n with two underlying layers: the weathering and subweathering layers. The datum is also shown in the Figure. The idea behind static correction is to move the source and receiver vertically down to the datum at S' and R'_n, respectively (Figure 1.2). The raypaths in Figure 1.2 under the weathering layer are now a bit different from the ones in Figure 1.1. Therefore, the aim of static correction is to account for near surface effects whether they are due to the difference in elevations between sources and receivers or variations in the weathering layer (Ali, 2019).

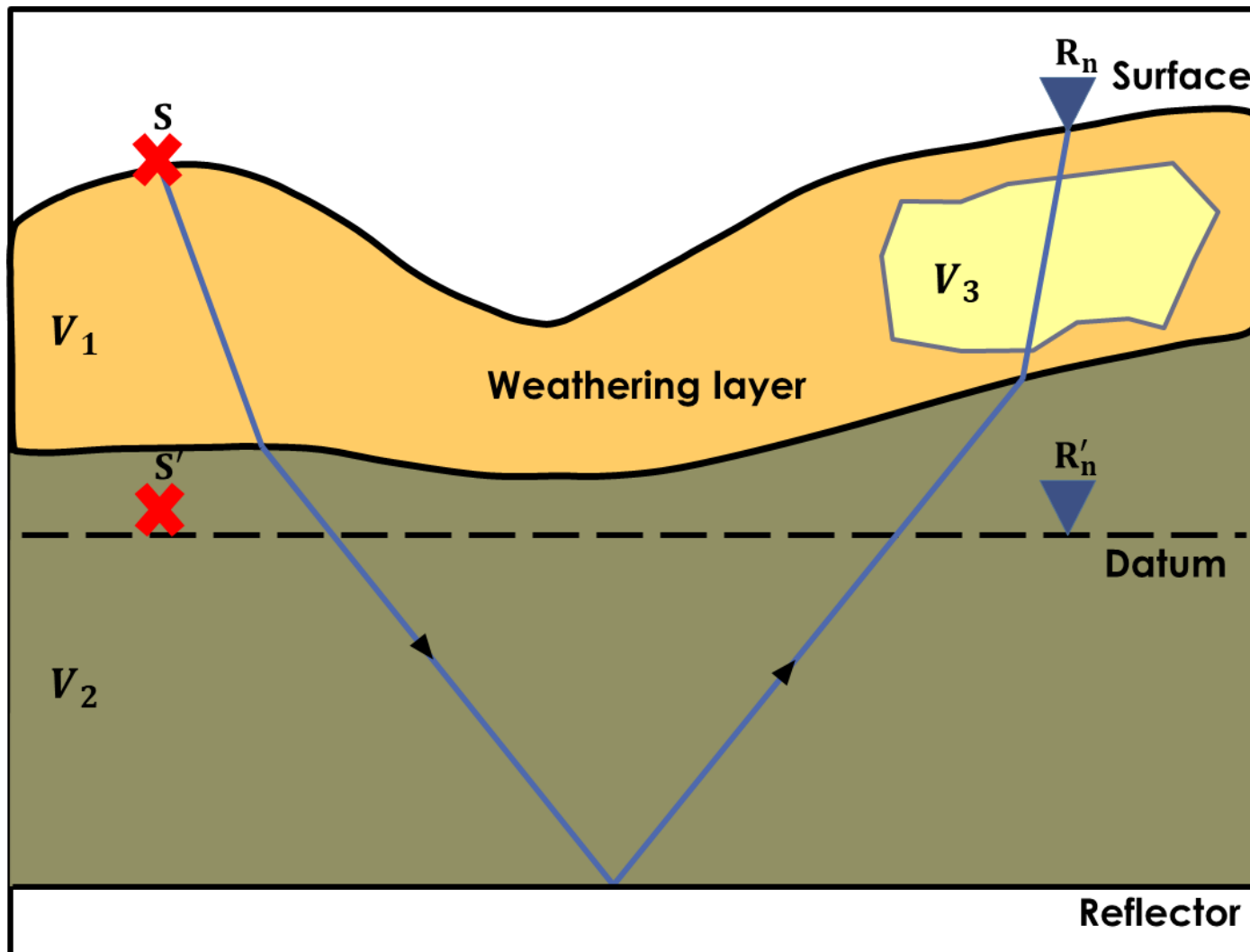


Figure 1.1: Model of the near-surface before applying static correction shows a seismic raypath between the source (S) and receiver (R_n) with two underlying layers.

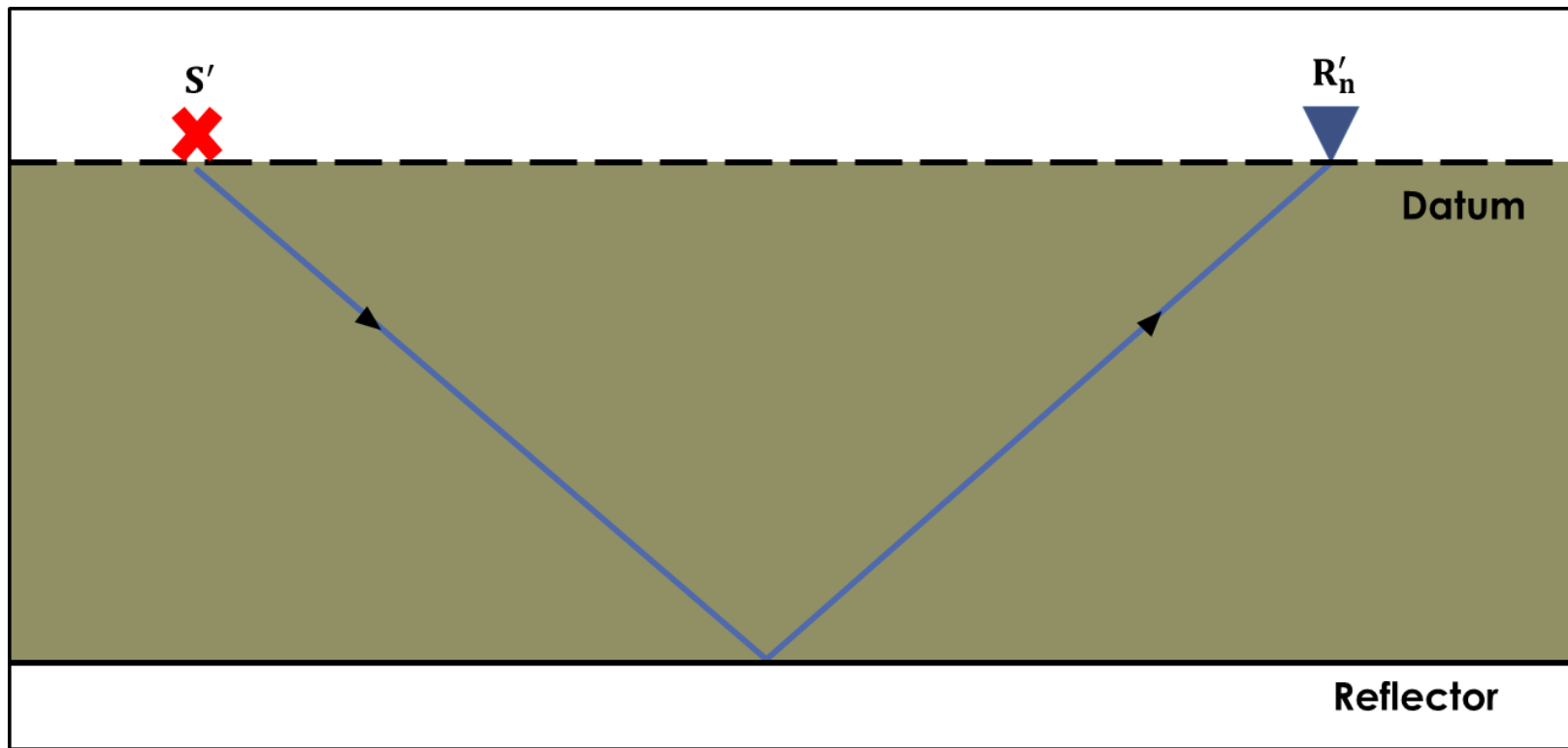


Figure 1.2: Model of the near-surface after applying static correction where the source (S') and receive (R'_n) are placed on the datum after removing the near-surface effects.

The elevation static correction is applied to remove the difference in elevations between sources and receivers (topography). This correction involves bringing sources and receivers to a common datum. It is preferable to bring this datum below the elevation of the lowest source or receiver (Figure 1.2). The residual static correction is usually applied after the elevation correction by correcting for the lateral variation of the velocity and/or thickness of the weathering layer. Static correction calculation may use seismic tomography techniques to get velocity models of the near-surface layer. The general idea behind tomography is using the misfit between observed and calculated first arrival traveltimes to correct an initial velocity model (Becerra et al., 2009).

There are some challenges that could face the static correction calculation such as in areas with complex near-surface geology, weak source wavelet penetration, or data with a low signal-to-noise ratio (SNR). In such cases, it is important to enhance the efficiency of the static correction methods for both reflection and refraction data processing.

The first arrival or first break, t_0 , can be defined as the onset arrival on a shot gather that is constituted from the refracted energy associated with the base of the weathering layer (Yilmaz, 2001; Figure 1.3). The shape of first arrivals is affected by different factors. It can be influenced by ground condition and coupling at both source and receiver locations (Cox, 1999). Moreover, different source types, such as vibroseis and dynamite, can affect the shape of first arrivals because vibroseis data include a component of autocorrelation function of the sweep signal and the recording filter after the sweep removal process (Cox, 1999). This means that a causal wavelet, such as one obtained from an explosive source, is not associated with correlated vibroseis data. First arrival picking of any refraction data depends on source type, near-surface structure, and SNR (Senkaya and Karsli, 2014). The first arrival traveltime picking becomes difficult especially when there is a complexity in the near-surface or data has a low SNR.

First arrival traveltime picking is divided into two main techniques: manual and automatic (Cox, 1999). Manual determination of the first arrival traveltime is done by visual inspection of the amplitudes by an experienced processor. In general, manual picking of first arrival traveltimes gives good results, but it is time-consuming because of the large volumes of seismic data and the need for real-time seismic data processing (Senkaya and Karsli, 2014). Furthermore, if the data is noisy, first arrival traveltimes may be difficult to pick. As a result, the picking procedure takes 20-30 % of the total processing time, especially when the data set is large (Sabbione and Velis, 2010). On the other hand, using advanced algorithms, machines can perform first arrival traveltime picking automatically. These automatic techniques are more efficient than manual picking but may fail or give inaccurate results under unfavorable conditions (Mousa et al., 2011).

Near-surface seismic data usually has low SNR especially at far offset, which makes the first arrival traveltime picking a challenge. The variation in geological features in the subsurface and velocity results in variations in travel times of seismic data. So, there is a need to enhance the accuracy of the first arrival traveltime picking, especially in challenging seismic data, by using new techniques.

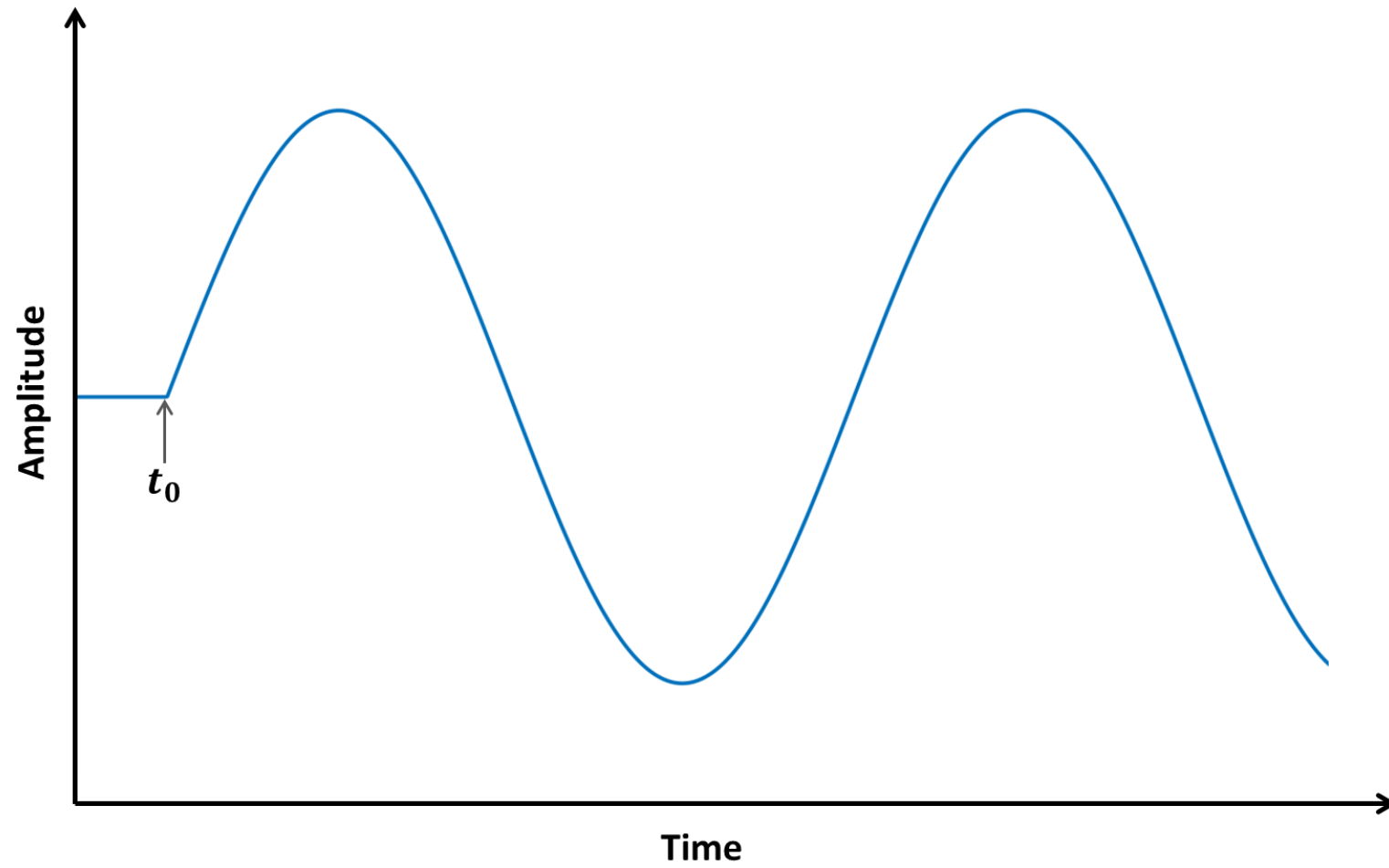


Figure 1.3: A seismic wave with an indication of the first-arrival time, t_0 , location.

1.2 Hilbert Transform and The Complex Seismic Trace

Signal processing is a subfield of electrical engineering which focuses on the analysis, modification, and synthesis of signals such as sound, images, and scientific measurements. Today, it is a rapidly growing region because it is needed in various applications in life. To make it possible to apply theoretical knowledge in an effective manner, special signal processors have been developed. Signal processors are also widely used in radio, transport, medical, and industrial equipment (Johansson, 1999).

Euler's formula, named in 1743 after Leonard Euler, was stated as follows:

$$e^{jz} = \cos(z) + j \sin(z). \quad (1.1)$$

Later, to introduce the complex notation of harmonic waveforms, physicist Arthur E. Kennelly and scientist Charles P. Steinmetz used Euler's formula using equation (1.2):

$$e^{j\omega t} = \cos(\omega t) + j \sin(\omega t). \quad (1.2)$$

Finally, German scientist David Hilbert showed that the $\sin(\omega t)$ function is the Hilbert transform of $\cos(\omega t)$. Therefore, the Hilbert transform is a specific linear operator that takes a function of a real signal and produces another function as an imaginary part of that signal (Barnes, 2007). The importance of calculating the imaginary part of a signal is because of the impossibility of extracting information such as instantaneous amplitude (or envelope), frequency, and phase from a real signal. Section 3.1.1 will discuss with more details the steps used to calculate the imaginary part (complex seismic trace) of a seismic trace.

There are several applications of the complex seismic trace attributes such as using polarization analysis to yield the wave type of arrivals (Vidale, 1986), using the complex seismic trace for time-lapse seismic surveys (Zhang and Bentley, 2000), instantaneous phase (or frequency) analysis of broadband seismograms (Bai and Kennett, 2001), use of geosteering phase attributes to detect discontinuities in seismic images (Yuan et al., 2018). This study will use the instantaneous amplitude (or envelope) to pick first arrival traveltimes.

1.3 Machine Learning

Machine Learning (ML) as a general definition is a part of computer science that teaches computer how to learn and act as humans. It is a subset of artificial intelligence (AI) that allows computer systems to learn and improve through experience (Mitchell, 1996; Goodfellow et al., 2016). Machine learning first applications and debates date back to the 1950's and its acceptance in the last 10 years has increased dramatically. It is used in a broad range of real-life applications such as image recognition, speech recognition, traffic prediction, and email spam and malware filtering.

There are mainly two types of machine learning algorithms: supervised learning and unsupervised learning. The kind of algorithm that a scientist prefers to use depends on what kind of data.

- **Supervised machine learning:** It is one of the most widely used and an efficient kind of machine learning. Supervised learning algorithms construct a mathematical model of data subset, known as training data, that contains both the inputs and the desired outputs. The goal of this algorithm is to learn the mapping function from inputs to the desired outputs. The purpose is to approximate the mapping function so well that when there is an input of new

data, the mapping function can predict the desired output for that data. In this type of machine learning, both the algorithm's input and desired output are specified (Russell and Norvig, 2002). There are two types of supervised machine learning. The first type is classification where the goal is to predict a class label. It is a binary classification when there is a distinguishing between two classes and a multiclass classification where the classification is between more than two classes. The second type is regression where the aim is to predict a continuous value. A simple way to differentiate between classification and regression is to ask if the input has some form of continuity. If there is a continuity between potential results, then it is a regression problem.

- **Unsupervised machine learning:** This type of machine learning includes algorithms that train on unlabeled data. In unsupervised learning, some of the most common algorithms used include:
 - Clustering: Separating the data set into similarity-based clusters. Examples of clustering algorithms are k-means clustering and density-based spatial clustering of applications with noise (DBSCAN) clustering.
 - Anomaly detection: Detecting of uncommon data points in a data set.
 - Association mining: Finding sets of objects that often occur together in a data set.
 - Dimensionality reduction: Decreasing the number of variables in a data set.

1.4 Clustering Algorithm

In general, clustering is an unsupervised machine learning algorithm that is used to identify similar objects in any dataset with a given similarity criterion. These methods can be divided into three

groups: hierarchical, partitioning, and density-based clustering. Hierarchical clustering algorithms seek to build a hierarchy of clusters. They start with some initial clusters and gradually converge to the solution. Partitioning-based clustering algorithms divide the dataset into initial ‘K’ clusters and iteratively improve the clustering quality based on an objective function (Sander et al., 1998).

The third group of clustering algorithm is density-based clustering algorithms, where dense regions in space are assumed to be clusters and separated by low-density regions (Lu and Feng, 2018). A dense cluster is a region, which is “density connected”. That is, the density of objects in that region is greater than a minimum. These algorithms can generate arbitrary shapes of clusters because it is based on dense connectivity. An example of density-based clustering algorithms is the DBSCAN method. Section 3.2.3 will discuss with more details the concept and steps used in DBSCAN clustering algorithm.

1.5 Statement of The Problem

First arrival traveltime picking is a necessary step in seismic data processing to create traveltime tomogram to be used for near-surface applications, static correction, etc. Due to the complexity of near-surface and the variations of the geologic features vertically and laterally in the subsurface, most near-surface seismic data has low signal-to-noise ratio (SNR). These could pose challenges in first arrival traveltime picking, so there is a need for novel methods to enhance traveltime picking.

There are several studies used the complex seismic trace attributes such as polarization analysis (Vidale, 1986), and phase identification and attribute analysis (Bai and Kennett, 2001). However,

complex seismic trace attributes are seldom used attributes to pick first arrival traveltime. Therefore, this study uses the envelope (or instantaneous amplitude) attribute to pick first arrival traveltime on both synthetic and real seismic data sets.

Moreover, many studies recently used machine learning algorithms for first arrival traveltime picking by using convolutional neural network (Xie et al., 2018), reinforcement learning (Ma and Luo, 2018), support vector machine (Yalcinoglu and Stotter, 2018), and deep semi-supervised neural network (Tsai et al., 2018). Most of these studies used supervised or reinforcement learning algorithms, but they rarely used unsupervised algorithms to pick first arrivals. Thus, the second proposed method in this study focuses on DBSCAN method to pick first arrival traveltimes and test the method on synthetic seismic data set.

1.6 Objective

The following are the objectives of the study:

1. Develop a new method based on the maximum difference between adjacent points of the envelope (MDPE) of a seismic trace to pick first arrival traveltimes.
2. Test MDPE method on both synthetic and real datasets.
3. Compare MDPE method with P-phase picker method and manual picking to test its picking accuracy.
4. Apply linear-moveout (LMO) shift and DBSCAN clustering algorithm on synthetic data to locate first arrival traveltime.

Compare DBSCAN method with manual picking to test its picking accuracy and limitations.

1.7 Justification of The Study

There are two justifications of the study as follows:

1. Previous studies, such as Vidale (1986) and Bai and Kennett (2001), used the complex seismic trace attributes to analyze polarization and identify and analyze instantaneous phase. However, complex seismic trace attributes were rarely used to pick first arrival traveltimes. Therefore, this study uses the instantaneous amplitude attribute for traveltimes picking.
2. DBSCAN clustering algorithm is used in this study for first arrival traveltimes picking, instead of supervised, reinforcement or support vector machine algorithms.

1.8 Benefits of The Study

As mentioned in the first section of this chapter, static correction is an important step for processing land seismic data. Due to the complexity of near-surface and low SNR, the purpose of static correction is to account for elevation difference between source and receiver and lateral variation of velocity and thickness of the weathering layer. To enhance the efficiency of static correction, the accuracy of picking traveltimes is important. During this study, both methods MDPE and DBSCAN will be tested on different datasets and compare them with manual picking to evaluate their accuracy and limitations. Moreover, both MDPE and DBSCAN methods do not require much

prior knowledge of data like many other methods. Therefore, this will add a value to the first arrival picking methods especially after specifying parameters of each method carefully.

1.9 Structure of Thesis

This thesis is divided into five chapters as follows:

Chapter 1 starts with an introduction of static correction and its importance in seismic data processing. It discusses the basic concepts of the Hilbert transform and the complex seismic trace. Then, it introduces concepts of machine learning and clustering algorithm briefly. It also covers objectives, motivation, and benefits of the study.

Chapter 2 mentions the history of complex seismic trace. Also, it covers previous studies and the literature review of picking first arrival traveltime from past to recent studies.

Chapter 3 contains two sections that cover the methodology steps of the two proposed methods used throughout this study. The first section starts with MDPE method and covers the main concepts of the Hilbert transform and steps used to calculate the instantaneous amplitude from the complex seismic trace. The second section shows the methodology steps of DBSCAN method. It starts with specifying parameters of the sliding window until identifying the first arrival traveltime picking location.

Chapter 4 has two sections. The first section shows and discusses application of MDPE method on both synthetic and real datasets, while the second one discusses the application of DBSCAN clustering algorithm on synthetic data. Both methods, MDPE and DBSCAN, will be compared with manual picks.

Chapter 5 concludes results and discussions of the two proposed methods and suggests some recommendations for future studies.

CHAPTER 2

LITERATURE REVIEW

2.1 Previous Studies of First Arrival Traveltime Picking

One of the early picking methods were done by correlation of adjacent traces to find delay times between adjacent arrival times (Peraldi and Clement, 1972). Hatherly (1982) proposed another first arrival traveltimes technique by determining both the first inflection point and first kick. This can be done for each shot by first determining the mean difference between the time to the corresponding first kick points and the first inflection points. Then, this difference is subtracted from the times to the first inflection points. Ervin (1983) described a computer algorithm by first summing several closely spaced shots to suppress noise and then convolve the stacked trace with a box operator to convert first arrival traveltime to a peak that can be detected by computer. A similar method has been developed by using several traces instead of one trace (Gelchinsky and Shtivelman, 1983).

Three types of traveltime picking algorithms were described by Akram and Eaton (2016): Single-level, multilevel, and hybrid traveltime picking algorithms. The single-level algorithms perform on single- and multi-component recordings. These algorithms are further categorized as window-based (e.g., energy ratio methods) and non-window-based (e.g., Akaike information criterion) methods. The window-based algorithms require specification of window size and position. The MDPE method is considered as a window-based single-level traveltime picking algorithm.

Window-based methods include: the short-term-average to long-term-average ratio (STA/LTA) method (Allen, 1978; Earle and Shearer, 1994), Coppens' energy ratio method (Coppens, 1985), and various modification of Coppens' method (e.g., Al-Ghamdi, 2007; Han et al., 2009;; Sabbione and Velis, 2010; Mousa and Al-Shuhail, 2012; Gaci, 2013). Earle and Shearer (1994) proposed the automatic phase-picking STA/LTA algorithm taken along an envelope function generated from the seismogram. Despite its additional computational cost, the envelope function has the advantage of remaining positive at various phase arrivals. Similarly, energy ratio over offset gathers can be used to pick first arrival traveltimes (Coppens, 1985). Al-Ghamdi (2007) modified Coppens' method by selecting a threshold value depending on the energy ratio curve slope. Sabbione and Velis (2010) developed an algorithm for picking first arrival traveltimes using three methods: a modified Coppens' method, an entropy-based method, and fractal dimensional analysis. Mousa and Al-Shuhail (2012) developed an enhanced first arrival traveltime picking for data with a low SNR by using the τ -p transform on energy ratio seismic records. In all ratio-based picking methods, the assumption is that the short-term window is sensitive to rapid signal variations while the long-term window represents background variations. This ratio minimizes effects of local signal variations and acts as a measure of local SNR (Trnkoczy, 2009). Some examples of algorithms used the Akaike information criterion (non-window-based methods) include Akaike (1973), Takanami and Kitagawa (1991), Sleeman and Van Eck (1999), Leonard (2000), Zhang et al. (2003), and Diehl et al. (2009).

The second type is the multilevel traveltime picking algorithms. These algorithms (e.g., cross-correlation approaches) make simultaneous use of data on multiple receiver levels inside the array (Akram and Eaton, 2016). Examples of these algorithms are Molyneux and Schmitt (1999), Raymer et al. (2008), De Meersman et al. (2009), and Senkaya and Karsli (2014). Hybrid

algorithms are a combination of more than one single-level algorithm. The benefit of this combination is that none of the individual algorithms are optimal in all situations to obtain a precise and accurate arrival-time picking (Sharma et al., 2010). Examples of these algorithms include Akazawa (2004) and Akram et al. (2013).

Additional automatic picking methods include those based on fractals (Boschetti et al., 1996; Jiao and Moon, 2000), digital image segmentation (Mousa et al., 2011), and high-order statistics (Yung and Ikelle, 1997; Saragiotis et al., 2002; Saragiotis et al., 2004; Küperkoch et al., 2010; Tselentis et al., 2012).

Recently, picking first arrival traveltimes focused on using machine learning, such as applying neural network algorithms (Murat and Rudman, 1992; Kusuma and Fish, 1993; McCormack et al., 1993; Li et al., 2013), using fully convolutional neural network (Xie et al., 2018; Yuan et al., 2018; Tsai et al., 2018; Hollander et al., 2018; Zhu and Beroza, 2019; Mezyk and Malinowski, 2018; Duan et al., 2018; Hu et al., 2019; Duan et al., 2020), reinforcement learning (Ma and Luo, 2018), combination of machine learning and traditional attribute-based computation (Hollander et al., 2018), support vector machine (Yalcinoglu and Stotter, 2018), and deep semi-supervised neural network (Tsai et al., 2018).

Some of these are based on single-level algorithms such as Murat and Rudman (1992), McCormack et al. (1993), Yalcinoglu et al. (2011), Li et al. (2013), Mezyk and Malinowski (2018), and Zhu and Beroza (2019). On the other hand, Yuan et al. (2018) used a multilevel algorithm based on convolutional neural network. Additional algorithms include those based on a combination of convolutional neural network with energy ratio methods (e.g. Hollander et al., 2018; Duan et al., 2018; Duan et al., 2020).

The proposed DBSCAN method was inspired by the work of Lu and Feng (2018) where they proposed a method that specifies a time window automatically by using machine learning methods for supervirtual interferometry (SVI). The time window containing first arrivals is specified using convolutional neural network (CNN) and DBSCAN). Then, SVI is used to enhance the SNR within the specified window. Synthetic data set is used with different levels of noise to test the performance of both CNN and DBSCAN. They suggested using CNN with extremely low SNR data sets, while the DBSCAN algorithm is preferable in most cases because it does not require a training data set.

2.2 History of The Complex Seismic Trace

Taner and Sheriff (1977) and Taner et al. (1979) showed that attributes such as trace envelope (a.k.a. instantaneous amplitude), phase rotation angle (a.k.a. instantaneous phase), and instantaneous frequency are provided by the complex seismic trace. To calculate the complex seismic trace, they used the Hilbert transform. The seismic trace amplitude represents the real part of the complex seismic trace, while the imaginary part of the signal is its Hilbert transform. Barnes (2007) explains that the trace envelope is the maximum value that the trace can achieve through a constant phase rotation (Figure 2.1). He defined the phase rotation angle as the angle needed to rotate the trace to its maximum amplitude. Finally, the instantaneous frequency is calculated as the time derivative of the phase.

Chopra and Marfurt (2005) mentioned the importance of each one of the three attributes as follows:

- **The trace envelope (or instantaneous amplitude)** is sensitive to acoustic impedance changes and, therefore, to porosity, thin-bed tuning, hydrocarbons, and lithology.
- **Instantaneous phase** is helpful for reflector continuity tracking and thus for detecting faults, unconformities, and changes in stratigraphy laterally.
- **Instantaneous frequency** is helpful in recognizing thin-bed tuning and abnormal attenuation.

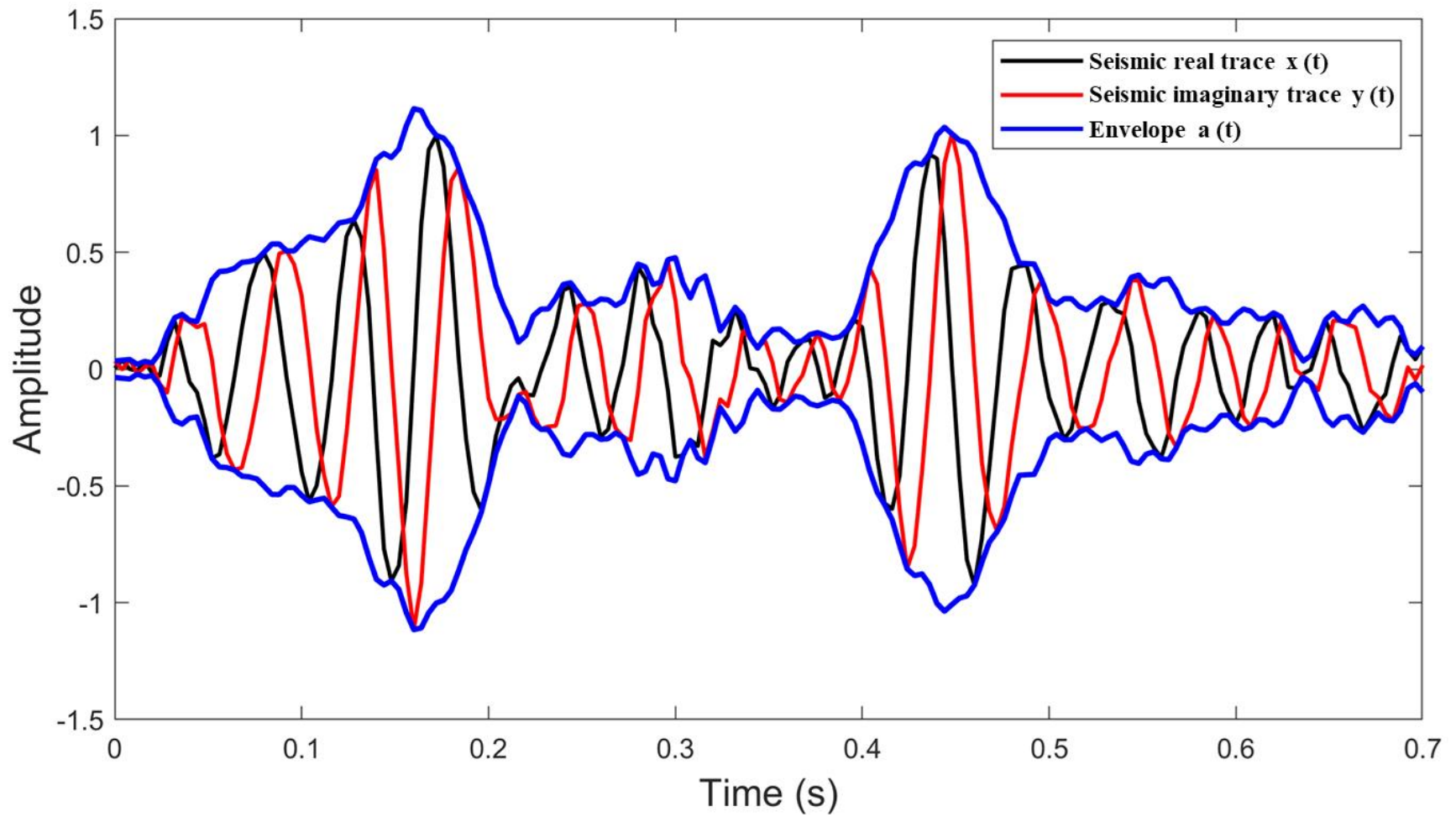


Figure 2.1: The trace envelope (blue) of a real seismic trace (black) and imaginary seismic trace (red) is the maximum value that the trace can achieve.

CHAPTER 3

METHODOLOGY

3.1 First Arrival Traveltime Picking Based on MDPE Method

The proposed MDPE method consists of the following steps:

1. Calculate the Hilbert transform to generate the envelope from the complex seismic trace.
2. Convert the envelope readings into decibels.
3. Calculate the moving median of the envelope within a sliding window.
4. Calculate the differences between adjacent points of the median along the trace envelope.

Pick the first arrival traveltime at the point with the maximum difference between adjacent points.

3.1.1 Hilbert Transform and Envelope

The Hilbert transform is an alternative method for extracting the instantaneous amplitude, frequency and phase information from signal data. Figure 3.1 illustrates the concept of the complex seismic trace $z(t)$, which consists of a real part, denoted by $x(t)$, and an imaginary part, denoted by $y(t)$, as defined by equation (3.1):

$$z(t) = x(t) + iy(t). \quad (3.1)$$

The real part $x(t)$ is the real (input) seismic trace, $y(t) = h(t) * x(t)$ is the imaginary seismic trace, and $h(t)$ is the Hilbert operator. The Hilbert operator $h(t)$ has the form:

$$h(t) = \frac{1}{\pi t} \quad (3.2)$$

and its Fourier transform $H(f)$, given by:

$$H(f) = -i \operatorname{sgn}(f) = \begin{cases} -i & f > 0 \\ +i & f < 0 \\ 0 & f = 0 \end{cases} \quad (3.3)$$

where sgn is the sign function.

The complex seismic trace is calculated by applying the following steps:

- Take the Fourier transform of the real signal to convert it from the time domain to frequency domain.
- Apply a -90° phase shift to all frequency components of the phase spectrum.
- Apply the inverse Fourier transform.

By applying the previous steps, the complex seismic trace in equation (3.1) will be generated with the real and imaginary parts of the data. Instantaneous amplitude, or envelope $a(t)$, can be calculated as:

$$a(t) = \sqrt{x^2(t) + y^2(t)}. \quad (3.4)$$

To emphasize amplitude jumps, the envelope of the complex seismic trace is converted to the decibel scale using the following equation:

$$\text{Envelope (dB)} = 20 \times \log_{10}[a(t)]. \quad (3.5)$$

By applying equation (3.5) to the envelope, a spike on the envelope of the complex seismic trace occurs when we approach the first arrival traveltime. However, to obtain an accurate value of the first arrival traveltime, the envelope needs to be smoothed by calculating the median in a moving sliding window of a user-defined length as explained in the next section.

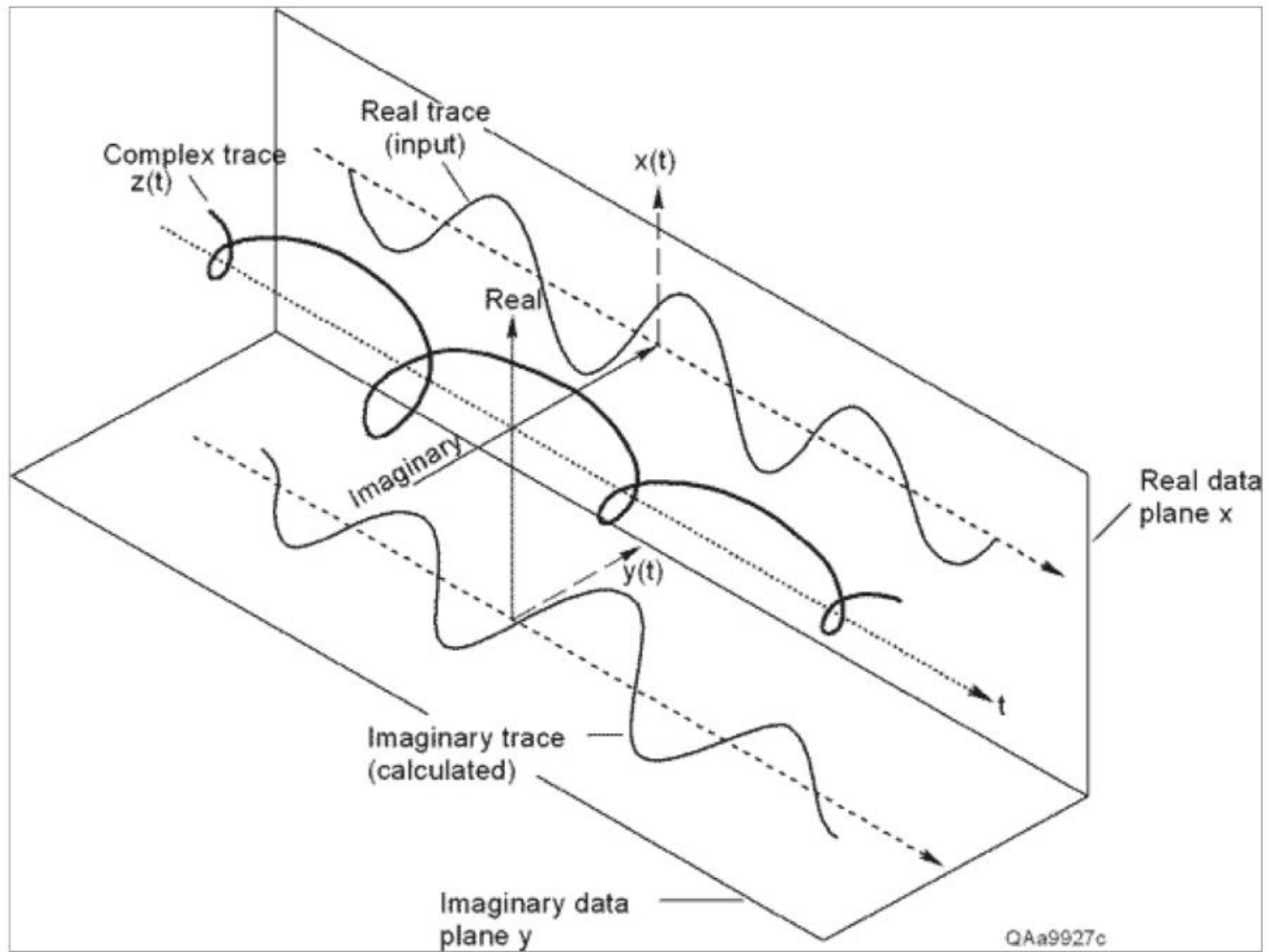


Figure 3.1: A complex seismic trace $z(t)$ that contains a real part $x(t)$, which is the real seismic trace, and an imaginary part $y(t)$, which is calculated by the Hilbert transform of the real seismic trace (Taner et al., 1979).

3.1.2 Sliding Window Median

A sliding-window median is applied to smooth the decibel-scaled trace envelope and sharpen the spike at the first arrival travelttime. In general, the median is the middle value of an ordered odd-numbered list in an array. If the count of the list is even, the median will be the mean of the two middle values in the list. It is also possible to take a median within a sliding window of a user-defined length (k). The window moves from the beginning to the end of the array one sample at a time. The user selects the sliding window length carefully depending on the signal's dominant frequency, time sampling interval and the number of time samples per trace. After calculating the sliding-window median of a certain trace, a spike should appear at the first arrival travelttime position.

3.1.3 Maximum Difference Between Adjacent Elements (MDPE)

The slope between any two adjacent points along the trace envelope can be found by calculating the ratio of the vertical change to the horizontal change between the two points:

$$Slope = (y_2 - y_1)/(x_2 - x_1). \quad (3.6)$$

The difference between adjacent elements uses almost the same concept as the slope with a small change. It is assumed that the distance between any two adjacent points (x_1, x_2) equals one, simplifying the slope to the difference in envelope values between each two adjacent points. After applying this to the whole trace envelope, the position of the maximum envelope difference is selected as position of the first arrival travelttime for that specific trace. The overall workflow of the proposed MDPE method is shown in Figure 3.2.

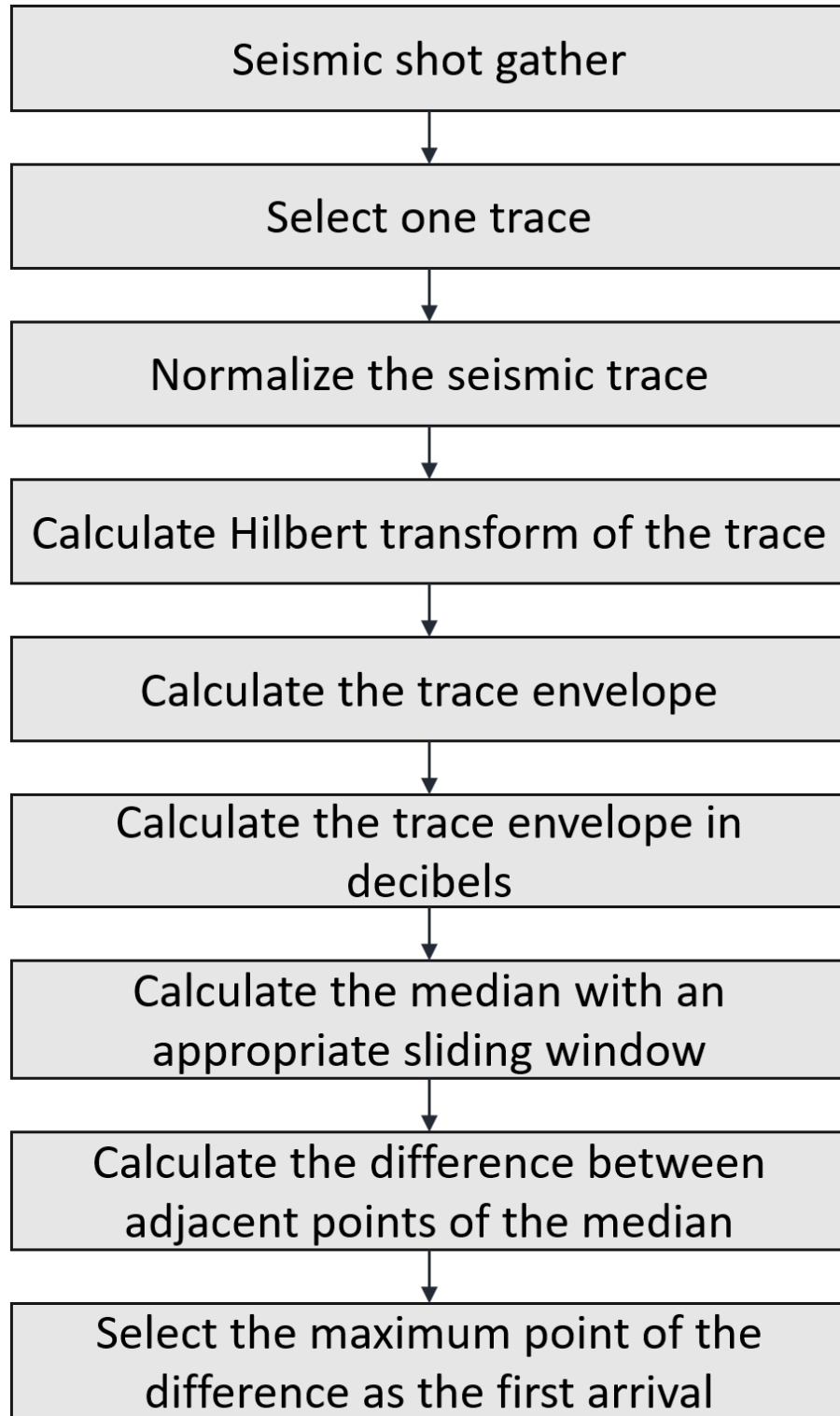


Figure 3.2: Workflow of the MDPE method for picking the first arrival.

3.2 First Arrival Traveltime Picking Based on DBSCAN Method

The proposed DBSCAN clustering method is described in the following steps:

1. A range of expected velocity values of the first arrival event is provided.
2. A sliding window that includes several traces with the same number of samples is selected.
3. A linear-moveout shift (equation 3.7) is applied to traces in the selected window (step 2) using a constant velocity value (step 1).
4. Shifted traces (step 3) are stacked for each velocity value (step 1) and the amplitude value of the stacked trace is calculated.
5. Apply DBSCAN clustering algorithm to divide the stacked gather (step 4) into different clusters.
6. The centroid of the first cluster identify the optimum apparent velocity and the first arrival traveltime pick.
7. Steps 2 to 6 are repeated for all possible windows in the recorded data set.

3.2.1 Selecting A Sliding Window

The first step in the second proposed method is to select an appropriate user-defined sliding window with specified numbers of traces and samples. This selected sliding window will be used throughout the whole recorded data set. The number of traces and samples within the sliding window differ from one data set to another. Number of samples should be selected carefully to avoid missing first arrivals, while number of traces should be chosen enough to get the maximum amplitude readings within first arrivals location after applying linear-moveout shift and stacking. Furthermore, the sliding window has a time shift parameter along the time axis to cover all needed samples (the ones that include first arrivals) within the specified traces, as well as a trace shift

parameter on offset axis to cover all needed traces in the recording data set. By default, the time and trace shifts should be set to one sample and one trace, respectively, to ensure that all first arrivals are covered in the recorded data. On the other hand, these two parameters could be increased to speed up the process and then interpolate between picked first arrivals but interpolated first arrivals could have larger picking errors than non-interpolated first arrivals.

3.2.2 Linear Moveout Shift (LMO) And Stacking

There are some seismic events, such as first arrivals for the direct and head waves, which have time-offset characteristics that can be represented by a straight line (Figure 3.3a). To project these events horizontally, a linear-moveout (LMO) shift algorithm is applied to eliminate the effect of offset on travel time and remove the dipping of first arrival events as shown in Figure (3.3b). The LMO uses a constant linear velocity to compute a single time (or sample) shift applied to each trace in the recorded data set. Therefore, the LMO shift is typically used to “level” an event that is linear with offset on a shot gather.

Mathematically, the LMO can be defined as:

$$t_{\text{LMO}} = t_0 + \frac{dx}{v} \quad (3.7a)$$

where t_{LMO} is the shift applied to the n th trace, t_0 is two-way travelttime corresponding the zero offset, dx is the distance between the n th trace and the 1st trace in of the selected window, and v is the event’s velocity. The LMO-corrected time of the linear seismic event is defined as:

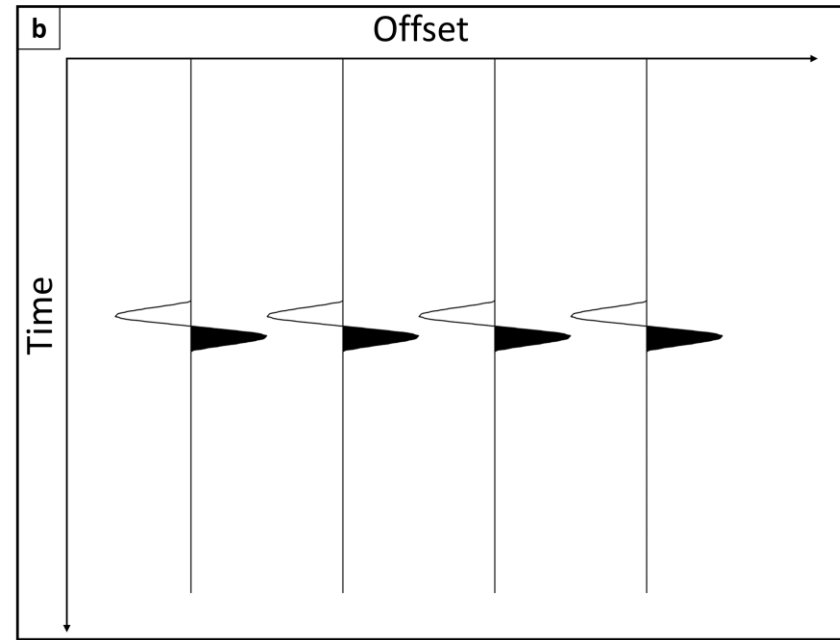
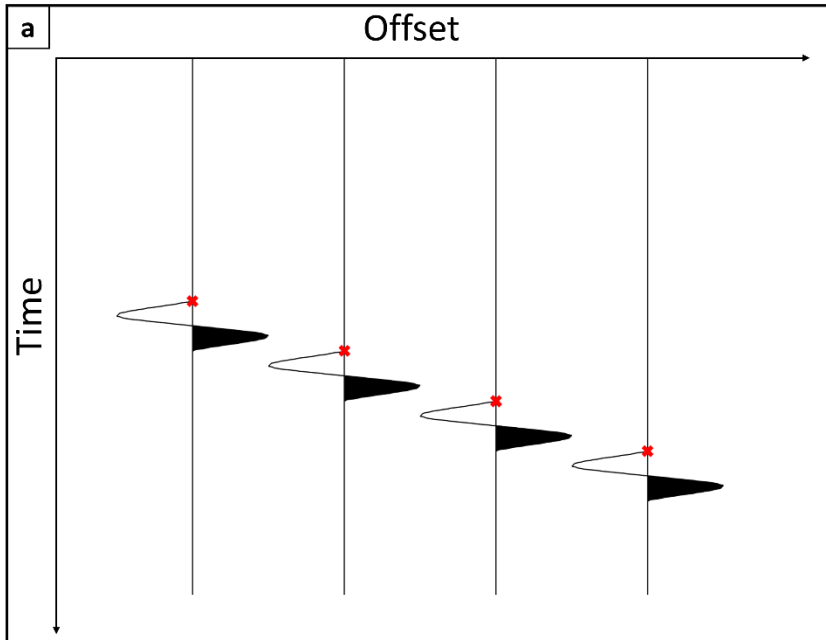
$$\Delta t_{\text{LMO}} = t_{\text{LMO}} - \frac{dx}{v} \quad (3.7b)$$

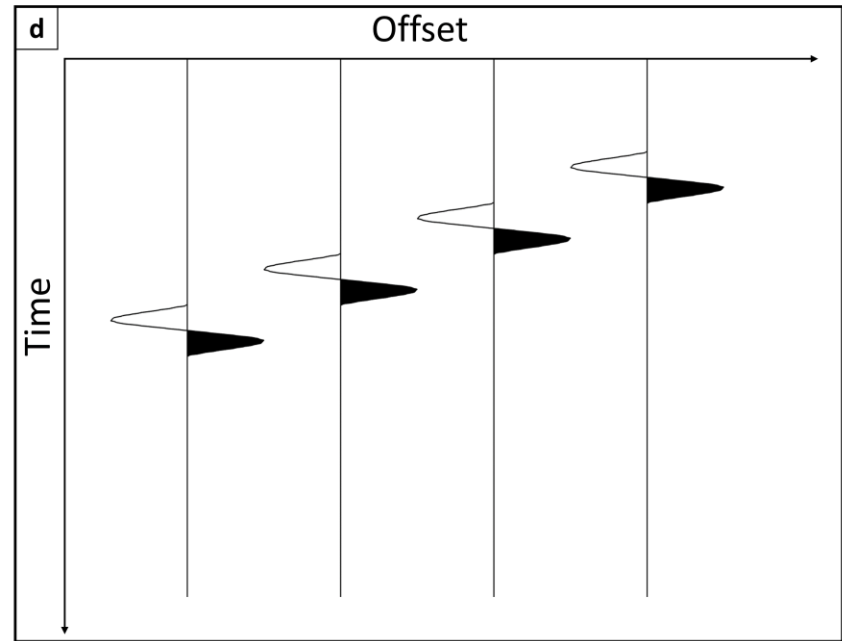
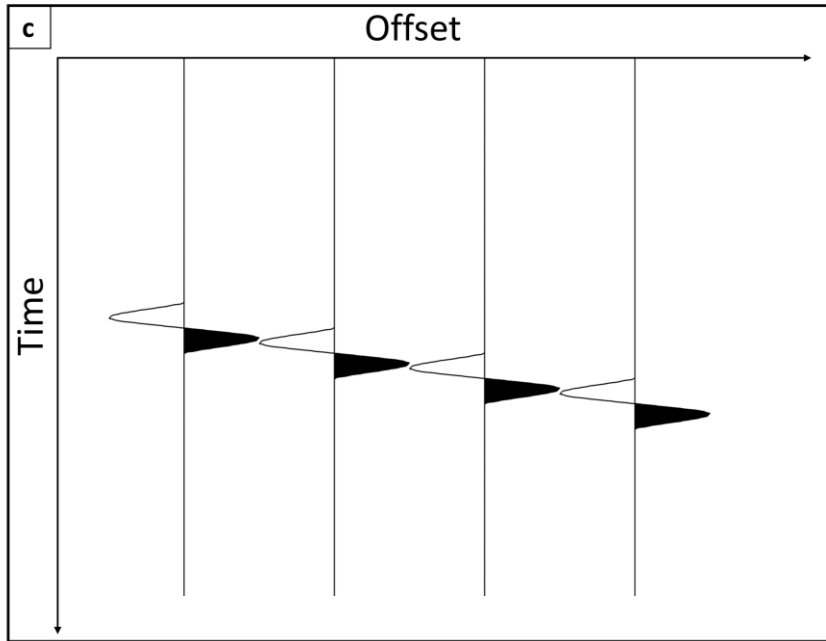
There are three different cases of applying velocity values in equation (3.7b):

1. Optimum (or best fit) velocity value (v) will level the linear event as shown in Figure 3.3b.

2. High velocity value (v) will under-correct the linear event with offset as shown in Figure 3.3c.
3. Low velocity value (v) will over-correct the linear event with offset as shown in Figure 3.3d.

The next step is to stack shifted traces within the sliding window for each selected velocity. The idea behind this step is to get the output amplitude for all used velocities in a stacked gather, where the maximum amplitude value location is related to the optimum velocity as shown in Figure 3.3e.





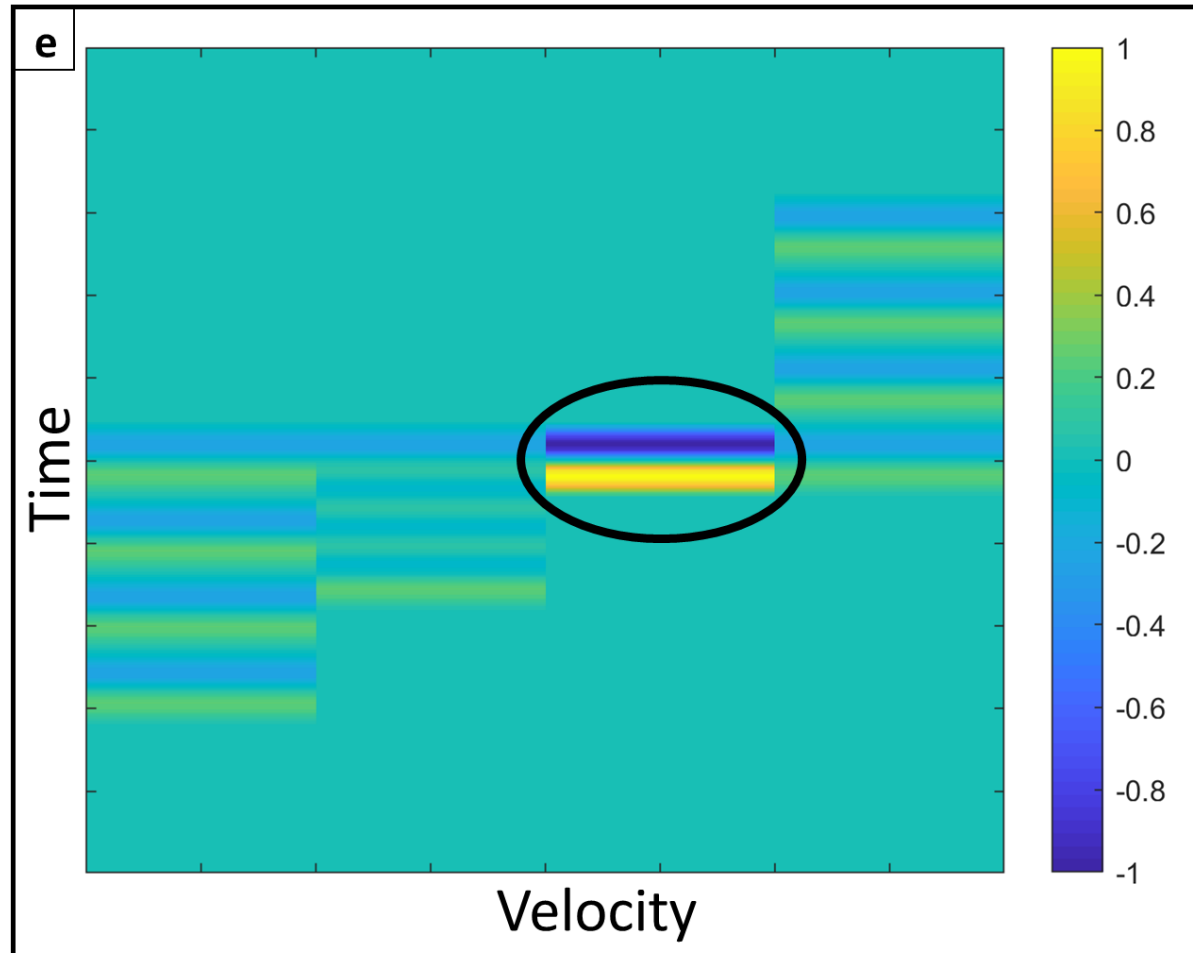


Figure 3.3: (a) A synthetic seismic gather with a linear event. LMO shift of the traces using: (b) optimum velocity, (c) high velocity, (d) low velocity. (e) a stacked gather for attempted velocity values where the maximum amplitude (back circle) output is achieved at the optimum velocity. Color indicates normalized amplitude values of stacked traces.

3.2.3 DBSCAN Clustering Algorithm

As mentioned in the introductory part, the DBSCAN is an unsupervised density-based clustering algorithm. It makes a cluster of points within high-density regions (points with close neighbors) and designates an outlier for points that have low-density regions (points with far neighbors) as shown in Figure 3.4. DBSCAN algorithm depends on two important parameters: epsilon-neighborhood (ϵ) and minimum objects (MinPts). Parameter ϵ defines the radius of neighborhood around an object x . MinPts objects are the minimum number of neighbors within epsilon (ϵ) radius.

This algorithm can identify three different object types:

1. **Core object:** An object x in data set with neighbor points count greater or equal to minimum objects (MinPts).
2. **Border object:** If the neighbor's number of objects y is less than minimum objects, but it belongs to the epsilon-neighborhood of core object x , it is called a border object.
3. **Outlier object:** An object that is neither a core nor a border object.

As mentioned by Ester et al., (1996), DBSCAN algorithm starts with calculating the distance between an initial object x with other objects in the data set by finding all neighbor objects within radius epsilon (ϵ) from the initial object x . All neighbor objects that count greater than or equal to minimum objects are marked as core objects. After that, create a new cluster for each core object that is not assigned to a cluster and find recursively all the density connected objects and assign them to the same cluster as core objects. Subsequently, iterate through the remaining objects in the data set. Finally, all objects that are neither within radius epsilon from core objects nor count greater than or equal to minimum objects are marked as outlier objects.

After creating the stacked gather with shifted traces for all provided velocities, a threshold parameter is applied to mute energies below this threshold. The importance of this parameter is to make DBSCAN algorithm work easily and more accurately, since the purpose of DBSCAN algorithm is to group data within the sliding window based on high amplitude output energies. Subsequently, DBSCAN algorithm will be applied to identify number of clusters within the same window. There are two cases after identifying the number of clusters as follows:

1. One or two clusters: the algorithm will automatically consider this window as an important one and the centroid of the maximum cluster will be selected as the optimum velocity value. In this case, the centroid point is the first arrival traveltime picking for the first trace in the selected window.
2. Two or more clusters: the algorithm will automatically consider this window as a noisy window. In this case, the algorithm will look for a different window with different samples and same traces until it finds an important window (case 1).

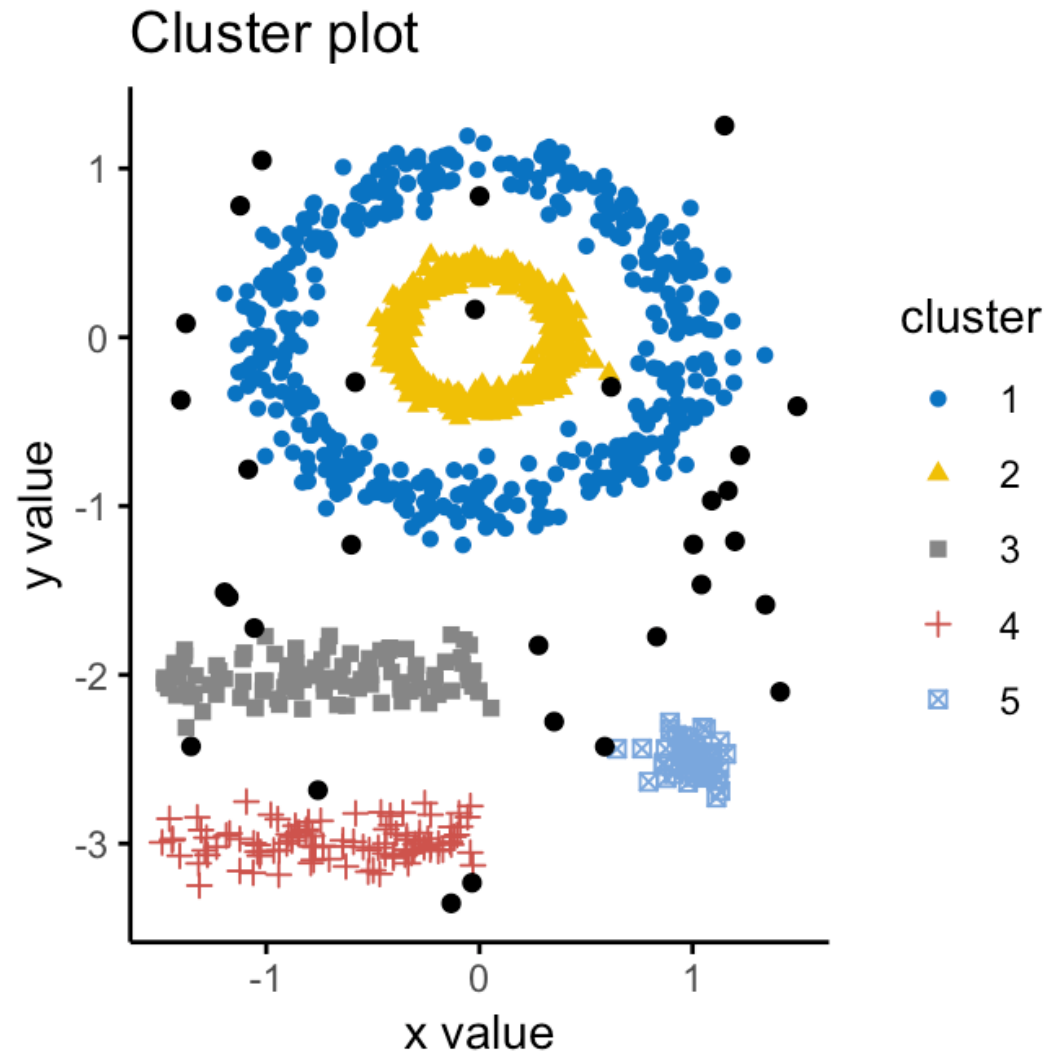


Figure 3.4: A data set that is divided into different clusters after applying DBSCAN clustering algorithm. Each color represents one cluster and outliers are shown in black color (Kassambara, 2017).

CHAPTER 4

RESULT AND DISCUSSION

4.1 First Arrival Traveltime Picking Based on MDPE Method

4.1.1 Results

This section is divided into three subsections. The first section shows the application of the MDPE method on a synthetic data set. The second section shows the results of applying the MDPE method on a 2-D real surface seismic data from southeast Texas and compares the performance of MDPE with the P-phase picker method (Kalkan, 2016). The last section shows the results of applying our method on two of Yilmaz's worldwide assortment of shot records (Yilmaz, 2001).

4.1.1.1 Synthetic Data

A 3-layer 2D velocity model (Figure 4.1) is used to generate a synthetic data set using acoustic finite difference forward modeling. The velocity model is 2400m wide and 200m deep. The first layer has a constant velocity of 400 m/s, while the 2nd and 3rd layers have velocities increasing with depth. The velocities range from 1200 to 1370 m/s in the second layer and from 2600 to 2800 m/s in the third layer. The data is generated using 110 receivers with 10 m receiver interval, 0.001 s time interval and 1 s total recording time.

Figure 4.2 shows synthetic shot gathers 1 and 56, which contain several seismic events (e.g., head waves and primary reflections). Figures 4.3(a and b) show results of applying the MDPE picking method (red) compared with manual picking (blue) for each shot record. Both shot records give results with picking accuracy (within 20 ms of manual picks) of 94% for shot gather 1 and 75% for shot gather 56. Median errors of the synthetic shot gathers 1 and 56 are 1% and 2%, respectively (Table 4.1).

To test the effects of noise, we added random Gaussian noise with zero mean and 0.01 standard deviation to the raw amplitudes of synthetic shot gather 1. Results show that picking accuracy (within 20 ms of manual picks) decreased from 94% to 69%. We also test the proposed method with different lengths of the sliding window on synthetic shot gather 1. Accuracy of picks (within 20 ms of manual picks) decreased from 94% (when an optimum window of 50 samples was used) to 59% (with a 25-sample window) and 73% (with a 75-sample window). Finally, we compare the STA/LTA and MDPE methods on synthetic shot gather 1. Results show that both methods give almost the same result when compared to manual picking. Figure 4.3(c) shows results of this analysis on trace 110 of this shot. STA/LTA estimates the first arrival at 0.533 s and MDPE estimates it at 0.530 s, which are both very close to the manual pick at 0.532 s.

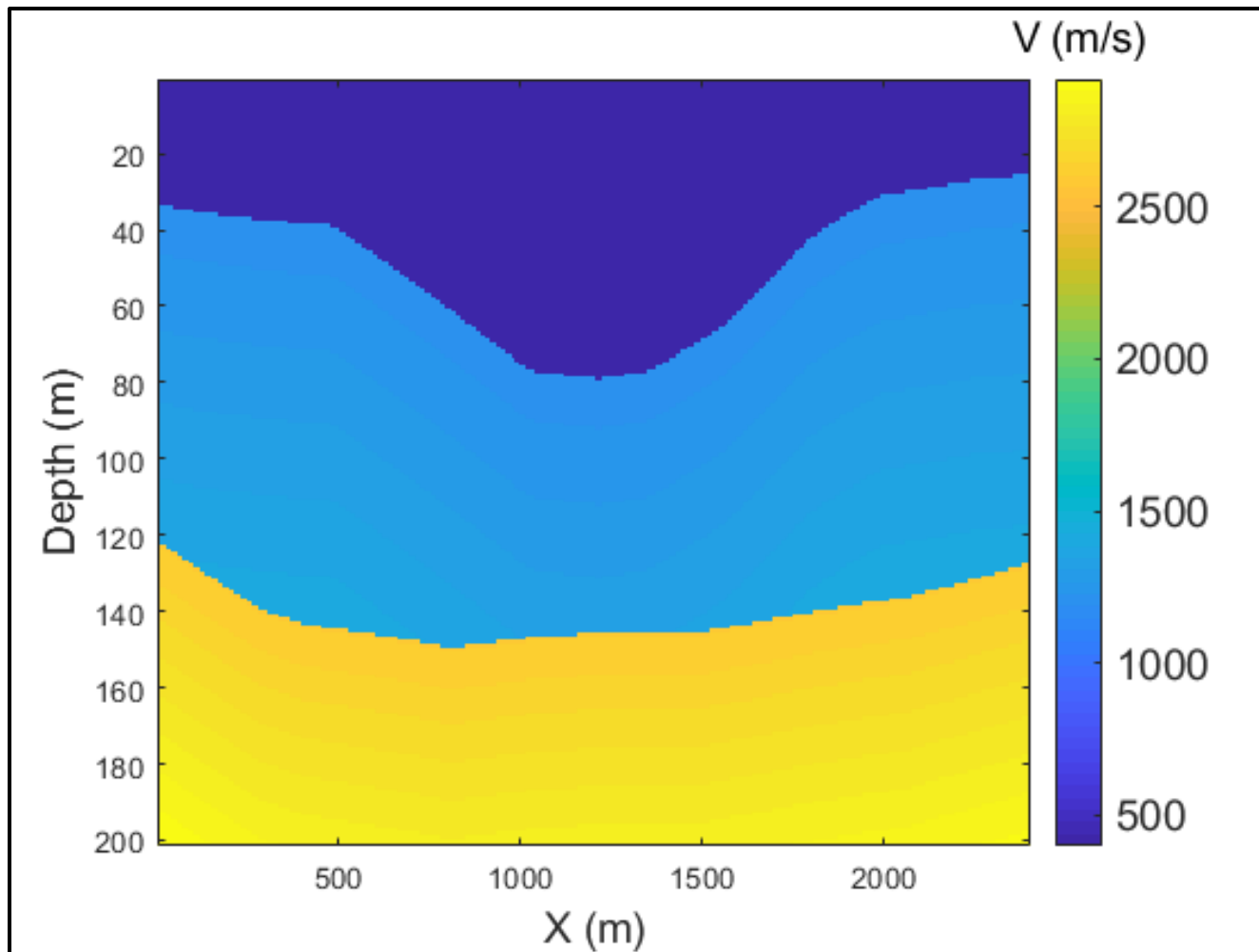


Figure 4.1: Velocity model used to generate the synthetic data set. Color bar shows velocities for the three different layers (400-2800 m/s).

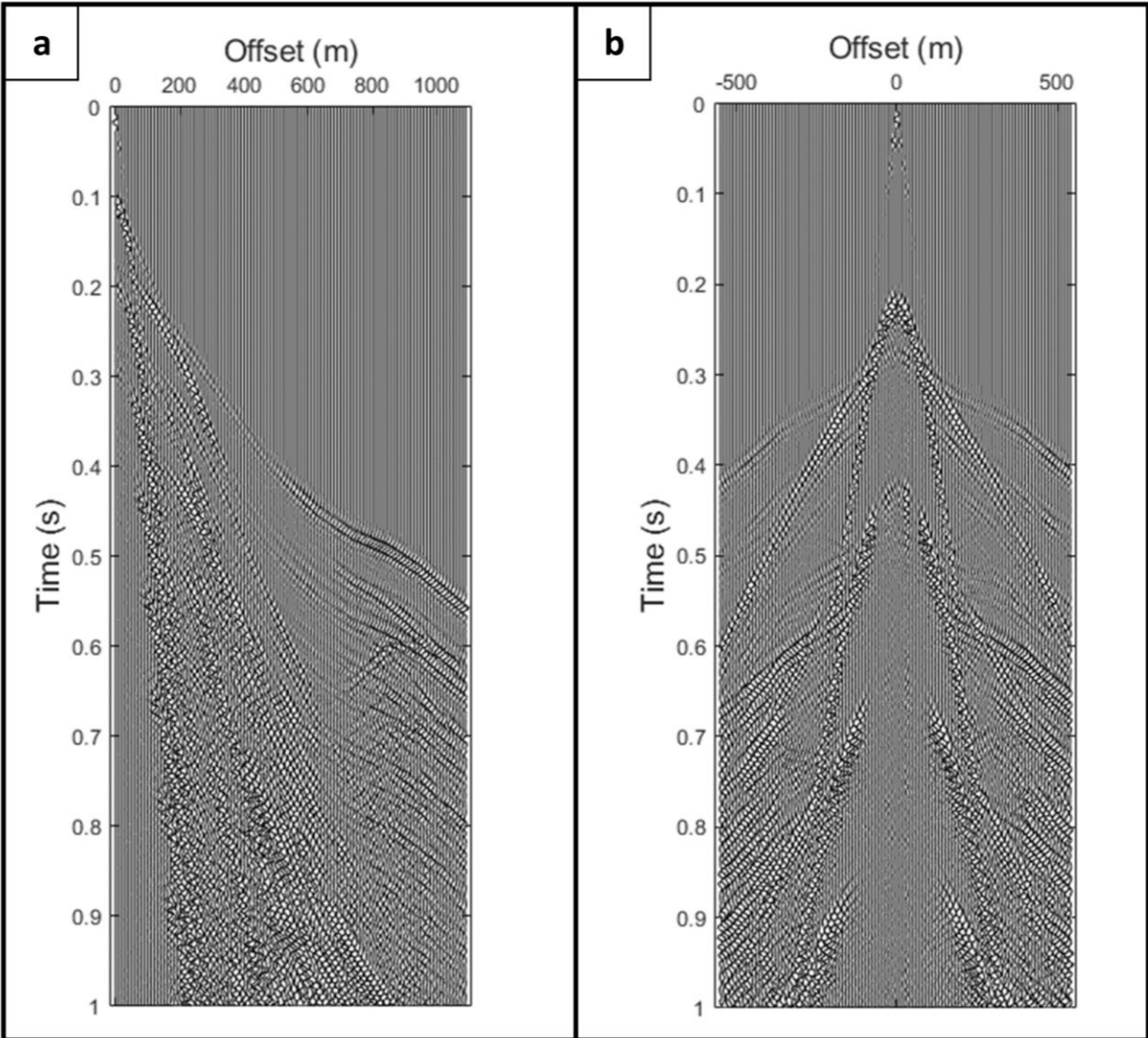


Figure 4.2: Sample shot gathers from the synthetic seismic data: (a) Shot gather 1; (b) Shot gather 56.

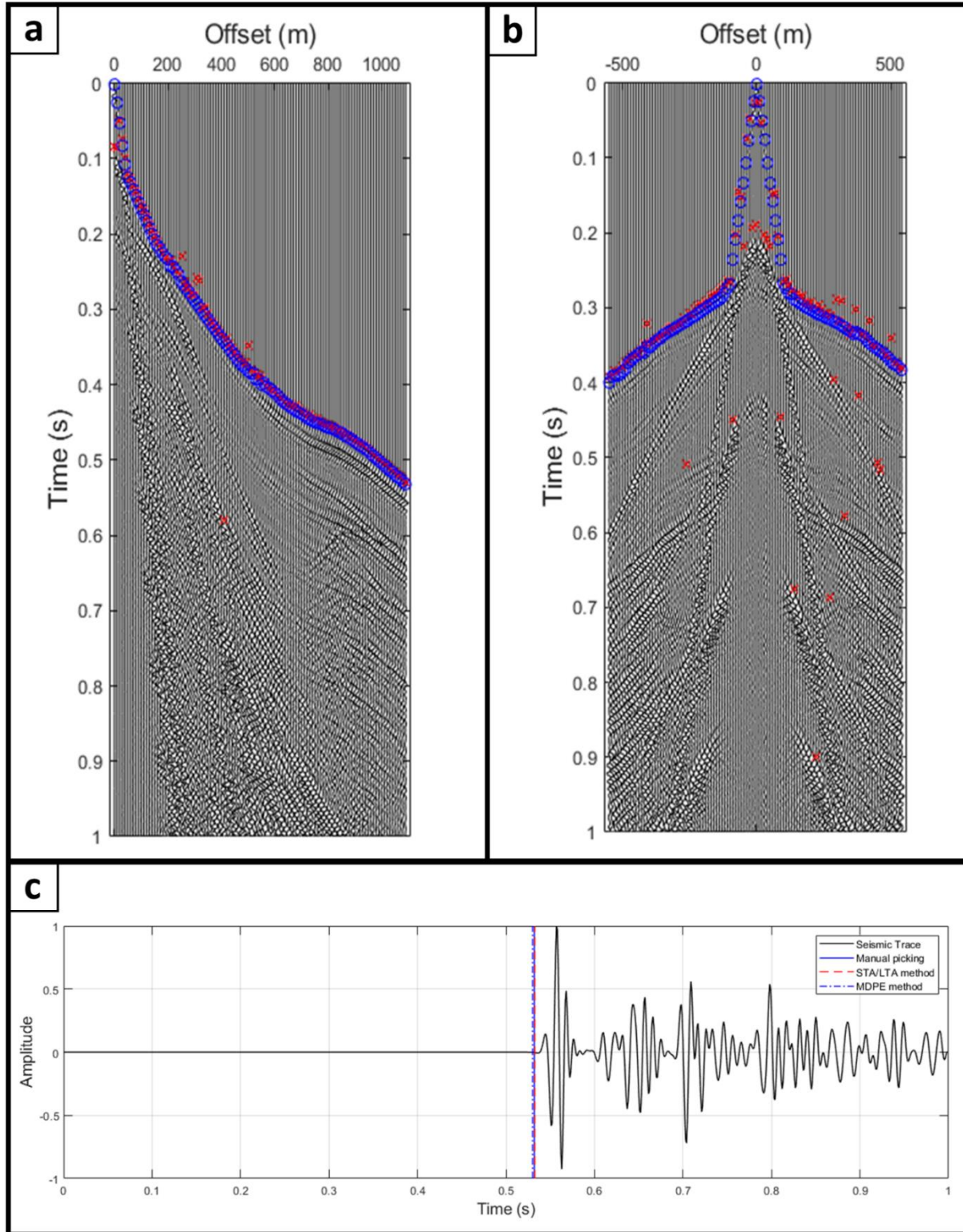


Figure 4.3: The same synthetic shot records shown in Figure 4.2 showing first arrival picks from the MDPE method (red crosses) and manual picks (blue circles): (a) Shot gather 1; (b) Shot gather 56. (c) Comparison of the MDPE method (dash-dot line) and STA/LTA method (dashed line) versus manual picking (solid line) for trace 110 of shot gather 1.

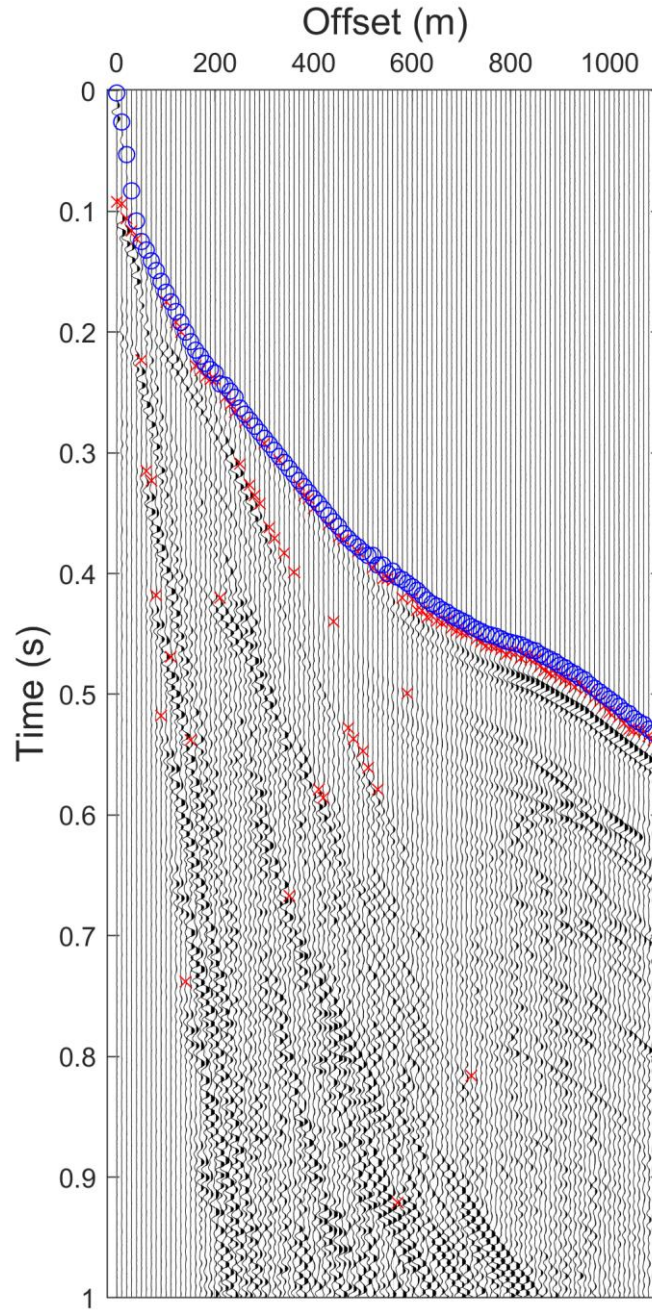


Figure 4.4: Shot gather 1 from the same synthetic seismic data shown in Figure 4.2 (a) showing first arrival picks from the MDPE method (red crosses) and manual picks (blue circles) after adding random Gaussian noise with zero mean and 0.01 standard deviation to the raw amplitudes.

Table 4.1: Minimum, maximum, mean, median, and standard deviation errors in percentage for all synthetic and real data sets used in both MDPE and DBSCAN methods.

MDPE method					
Error (%)	Minimum error (%)	Maximum error (%)	Mean error (%)	Median error (%)	Standard deviation error (%)
Synthetic data set					
Shot gather 1	0%	98%	4%	1%	12%
Shot gather 56	0%	9350%	102%	2%	891%
Real data set					
Southeast Texas	0%	2%	0%	0%	1%
Yilmaz's shot gather 6	0%	351%	13%	0%	65%
Yilmaz's shot gather 25	0%	302%	3%	0%	31%
DBSCAN method					
Synthetic data set	0%	145%	9%	6%	20%

4.1.1.2 Real Data

1. Real data set from southeast Texas

The MDPE method was applied to a data set published by (Mousa and Al-Shuhail, 2011). The data set is a 2-D set from southeast Texas and has the following parameters:

- 18 shot gathers
- 33 traces in each shot
- Dynamite source
- Receiver interval = 67 m
- Time sampling interval = 0.002 s
- Number of time samples per trace = 1501

Figure 4.5 shows sample shot gathers (7 and 17) from this data set. To illustrate how the MDPE method works, we apply the workflow on two selected traces of this data set: trace 31 from shot 7 (with high SNR) and trace 29 from shot 2 (with low SNR) and compare the results with those from the P-phase method and the manual picking. Figure 4.6 shows the details of the raw traces. The envelope of these two traces is calculated (Figure 4.7) and expressed in decibel units (Figure 4.8). A smoothed version of the traces in Figure 4.8 is calculated using a 50-sample moving median (Figure 4.9). Finally, the difference between successive points of the smoothed envelopes is calculated (Figure 4.10). The first arrival traveltimes is picked at the position of the maximum difference. The first arrival traveltimes estimated from the MDPE method and P-phase picker method are compared with manual picks. Figure 4.11 shows that application of these two methods on trace 31 from shot 7 estimates the manual pick accurately by both methods (Figure 4.11(a)).

On the other hand, application on trace 29 from shot 2 shows that the MDPE method estimates the manual pick more accurately than the P-phase method (Figure 4.11(b)).

Furthermore, Figure 4.12 shows the result of applying the proposed MDPE method on 54 traces selected from this data set. The proposed method gives almost the same first arrival traveltime picks as the manual picks. On the other hand, Figure 4.13 shows the result of applying the P-phase picker method on the same 54 traces which results in 65% picking accuracy within 20 ms of manual picks. Median error of the 54 traces selected from this data set is 0% as shown in Table 4.1.

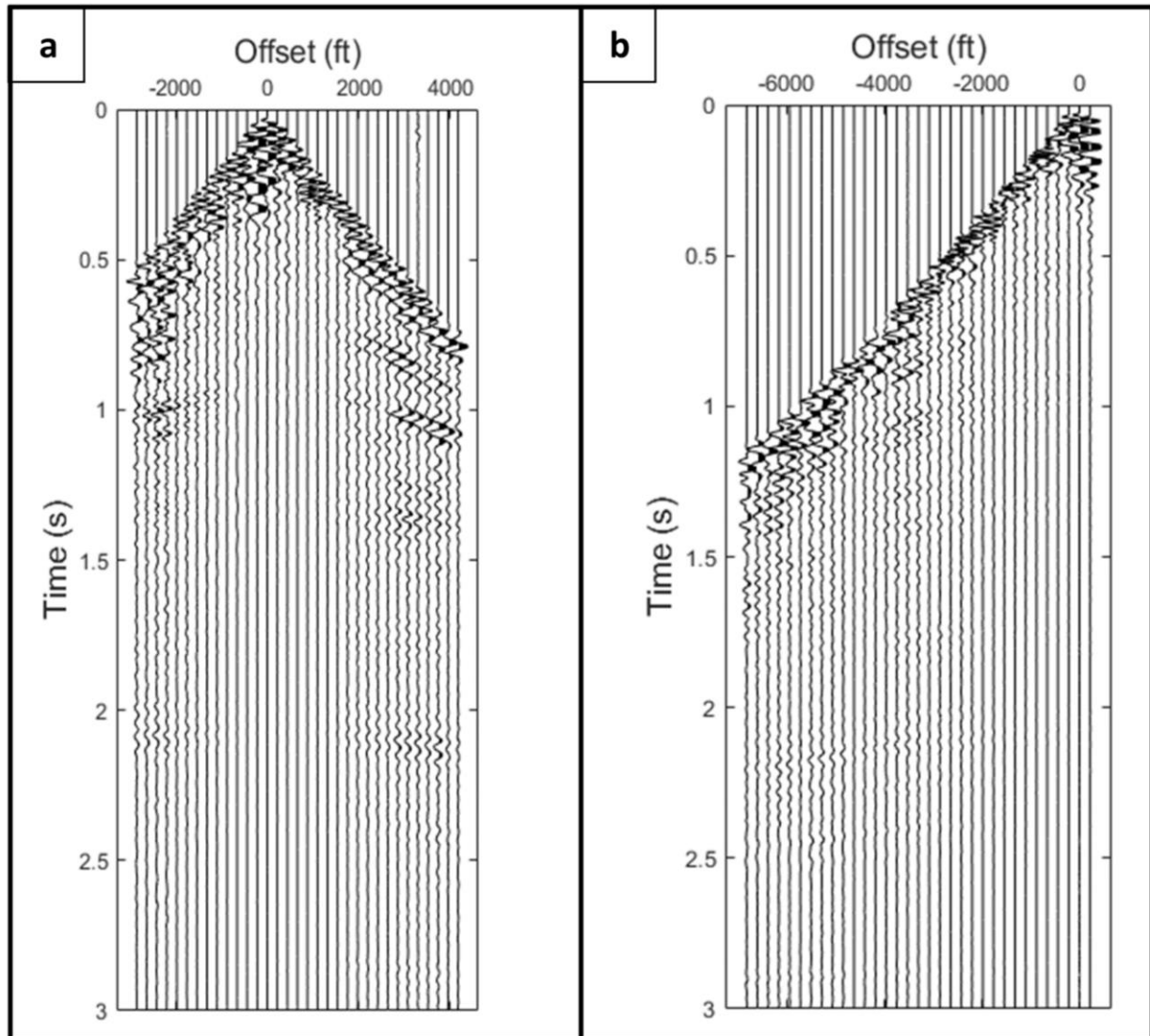


Figure 4.5: Sample shot gathers from the real data set from southeast Texas: (a) shot gather 7; (b) shot gather 17.

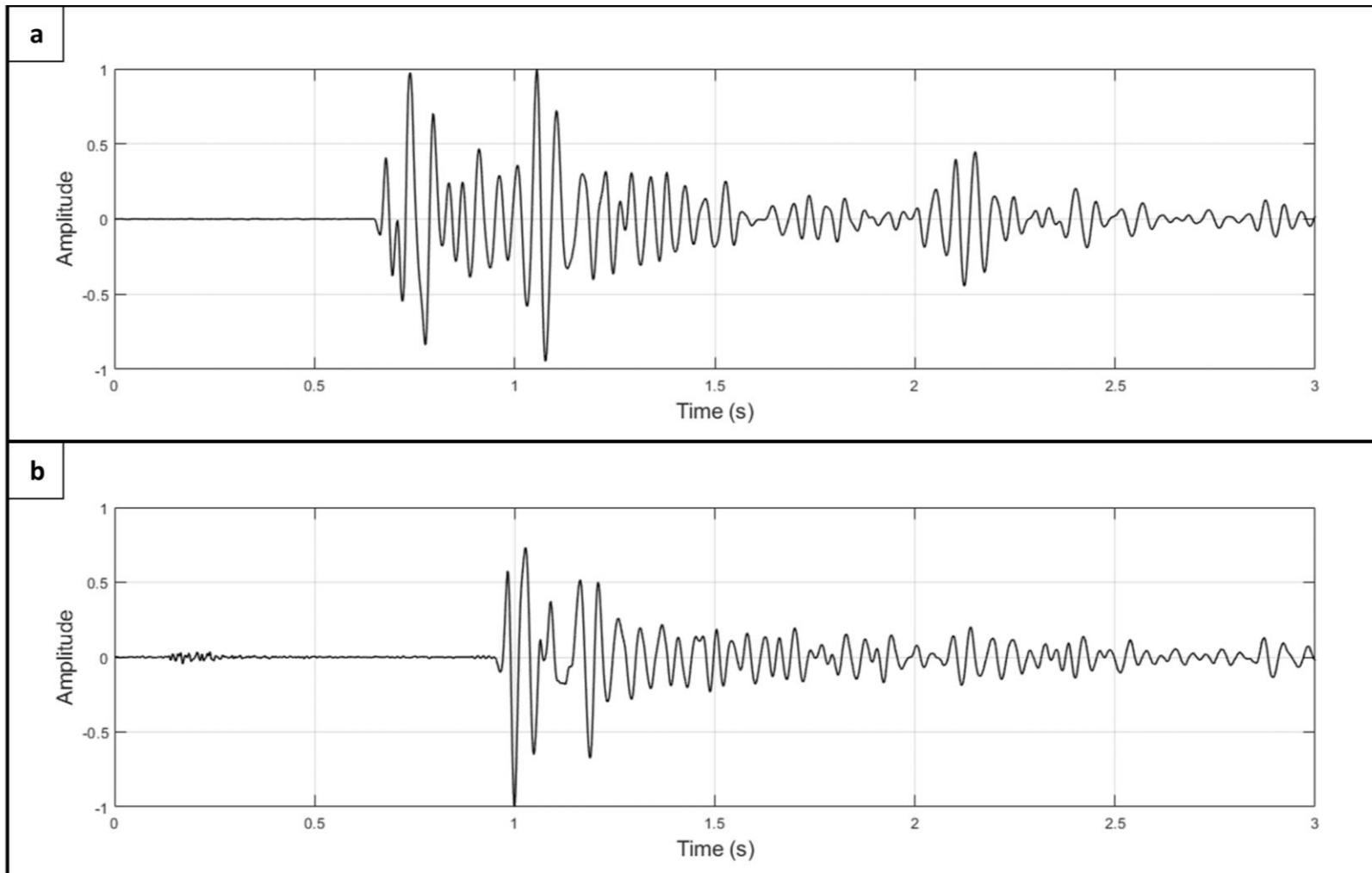


Figure 4.6: Sample traces from the real data set from southeast Texas: (a) trace 31 from shot 7; (b) trace 29 from shot 2.

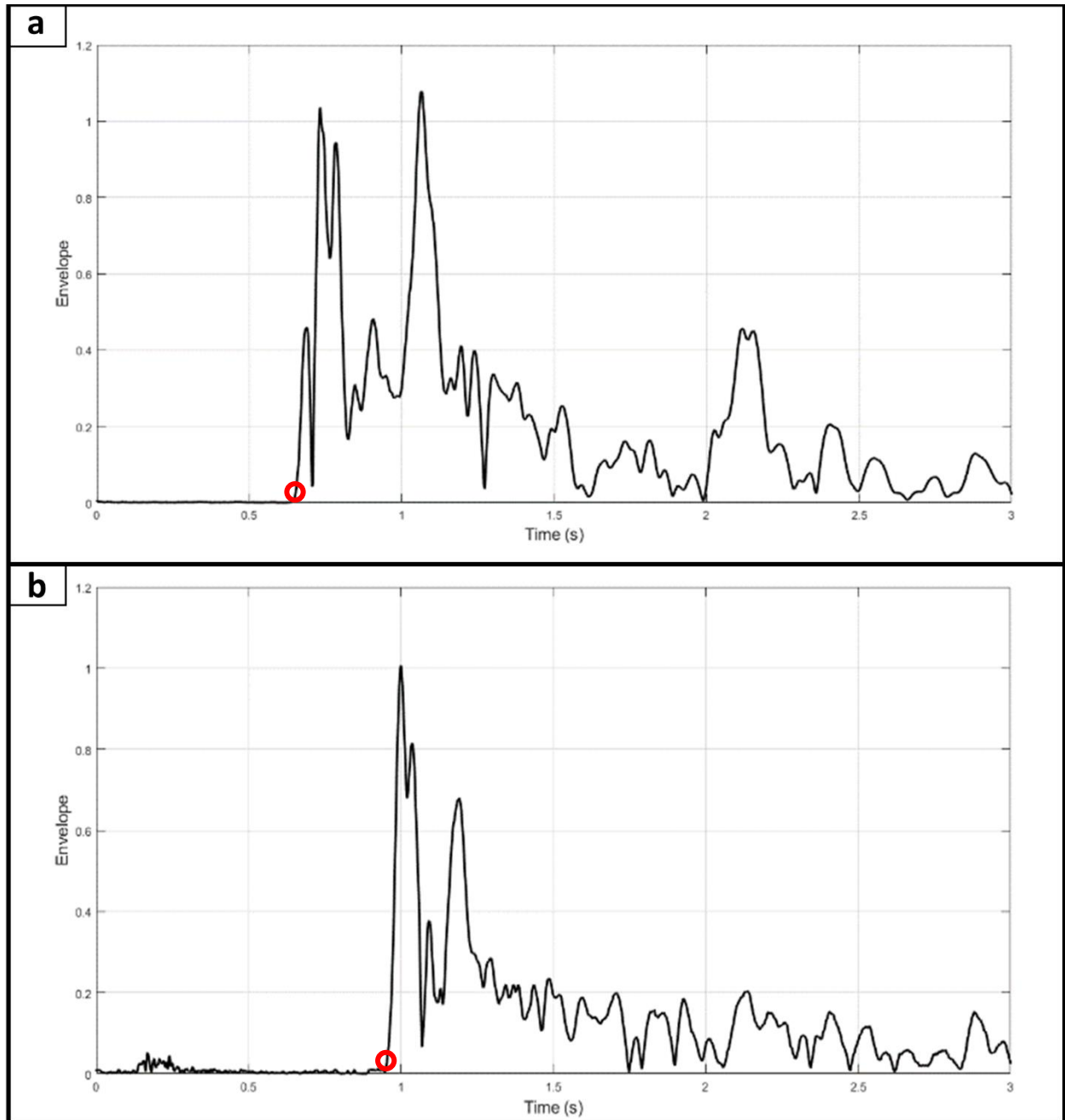


Figure 4.7: The envelope of the traces in Figure 4.6. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.

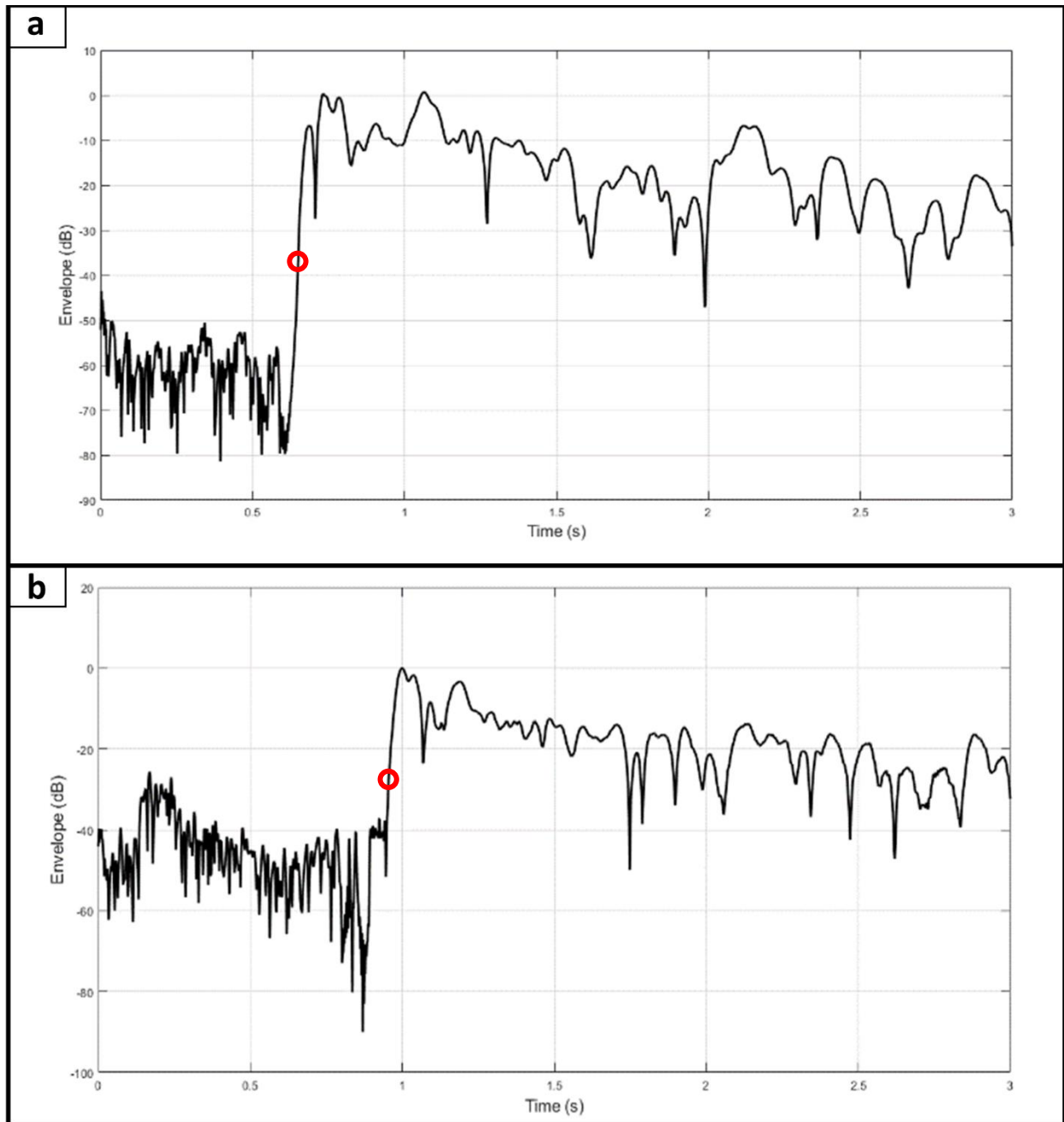


Figure 4.8: The envelope in decibel units for the traces in Figure 4.6. Note the spike near the first arrival time position. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.

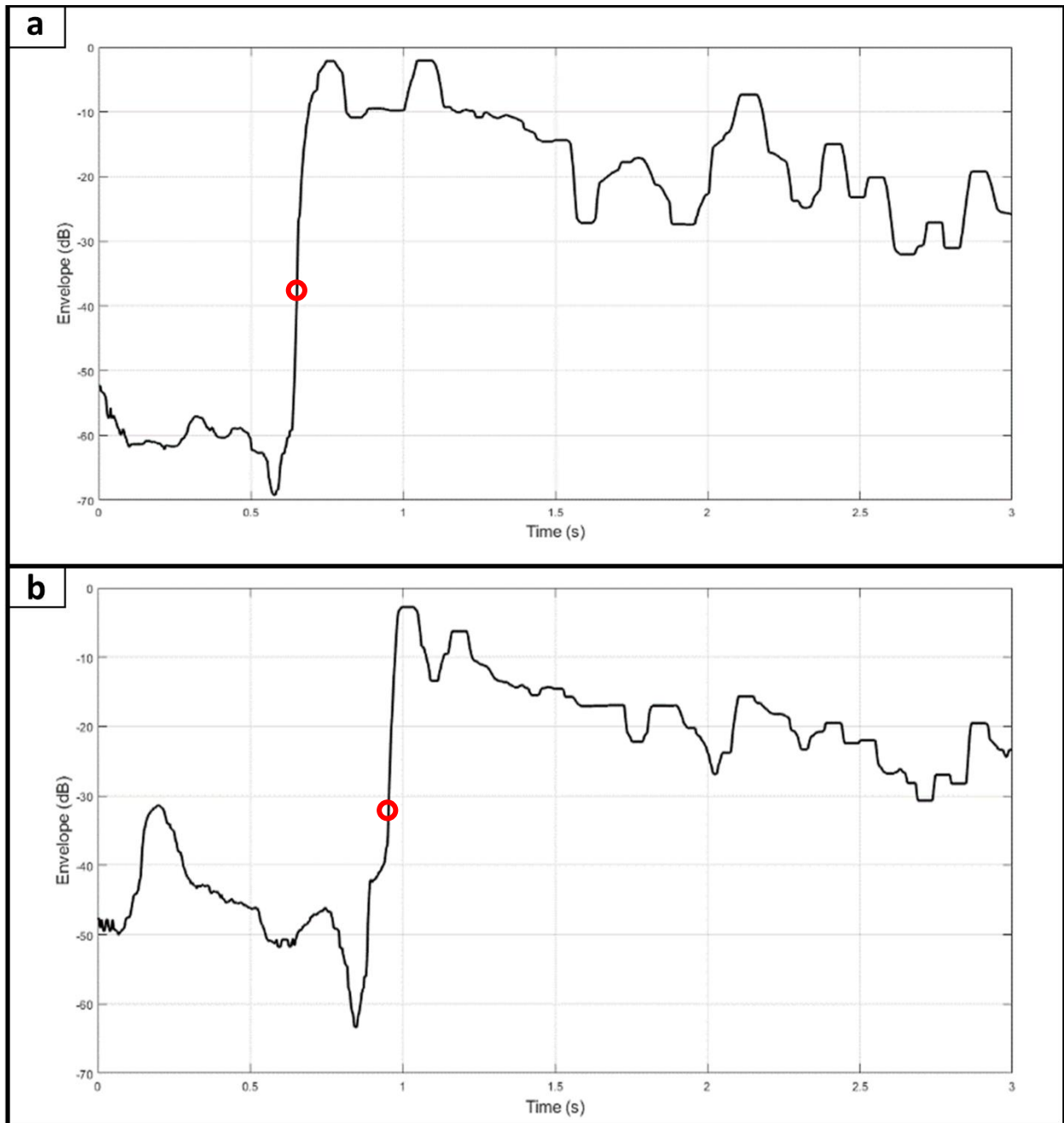


Figure 4.9: Median smoothing with a 50-sample sliding window of the envelope for the traces in Figure 4.6. Note the smoothness near the first arrival. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.

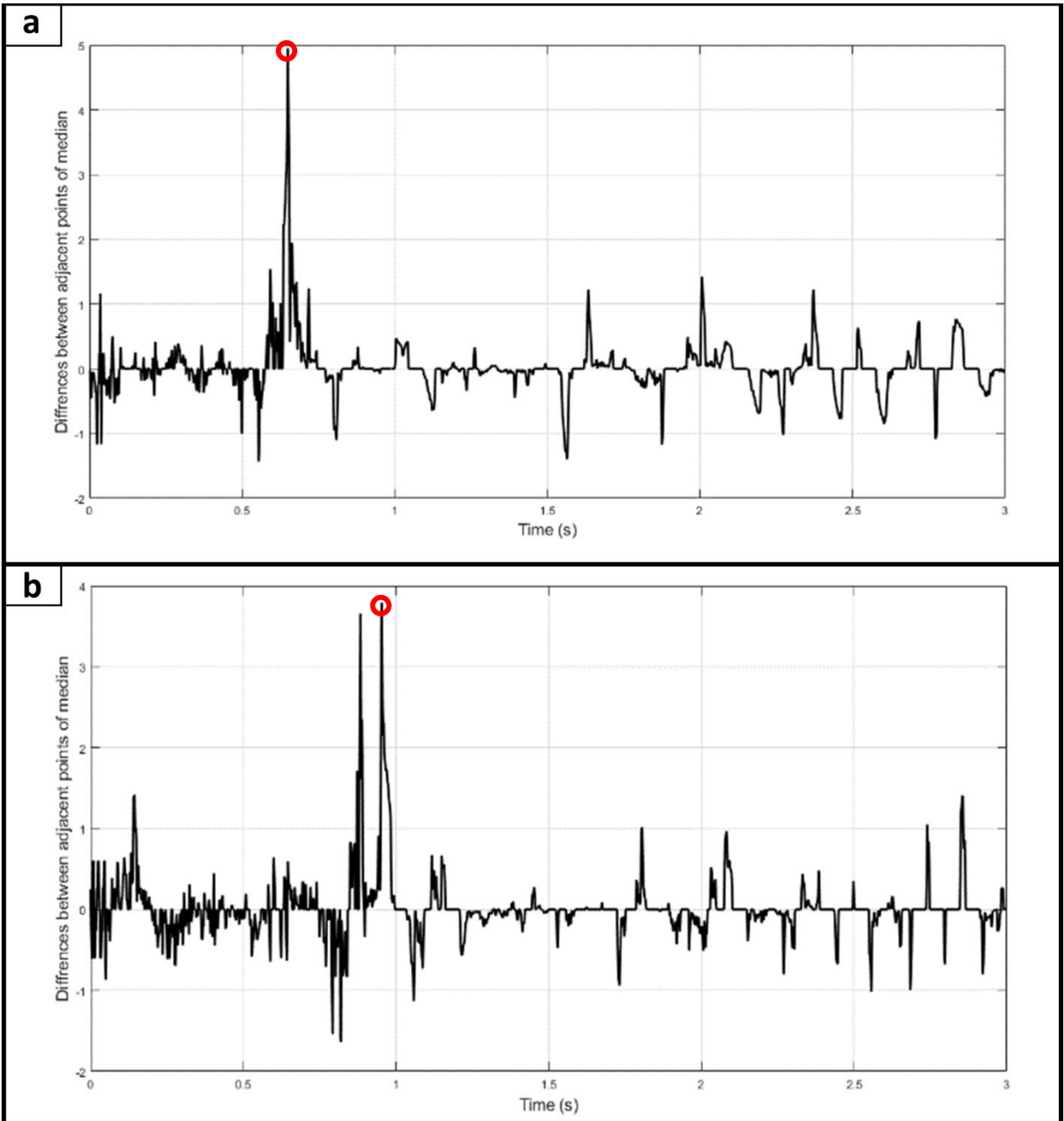


Figure 4.10: Difference between adjacent points of the envelope after applying moving median smoothing. The first arrival is always picked at the point with maximum difference. The red circle indicates the manual first arrival picking of: (a) trace 31 from shot 7; (b) trace 29 from shot 2.

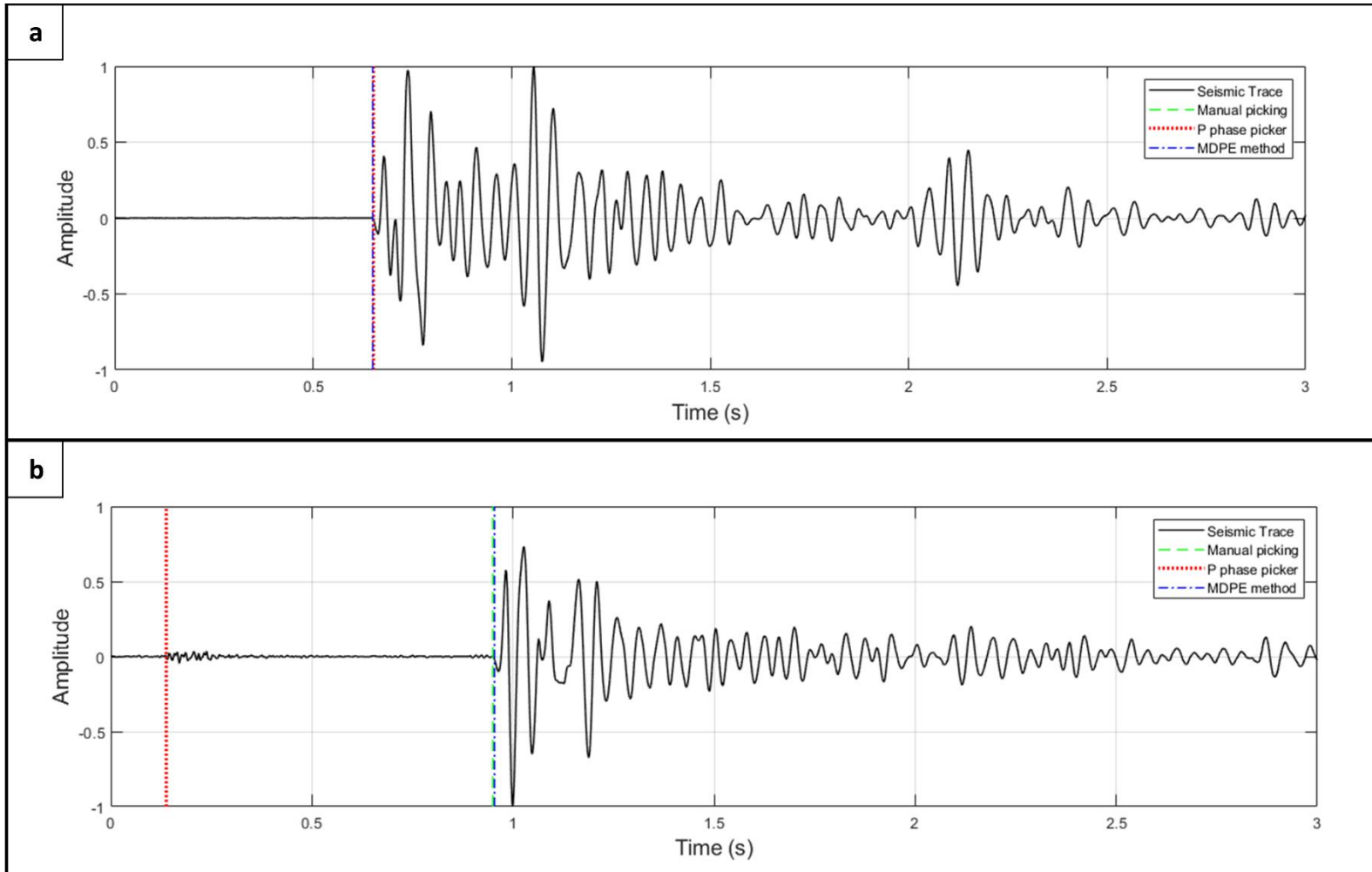


Figure 4.11: Comparison of the MDPE method (dash-dot line) and P-phase picker method (dotted line) for the traces in Figure 4.6. The manual picking is shown by the dashed line. (a) Application on this trace gives almost the same accuracy for the two compared method; (b) Application on this trace shows more agreement between the MDPE method and manual picking than the P-phase picking method.

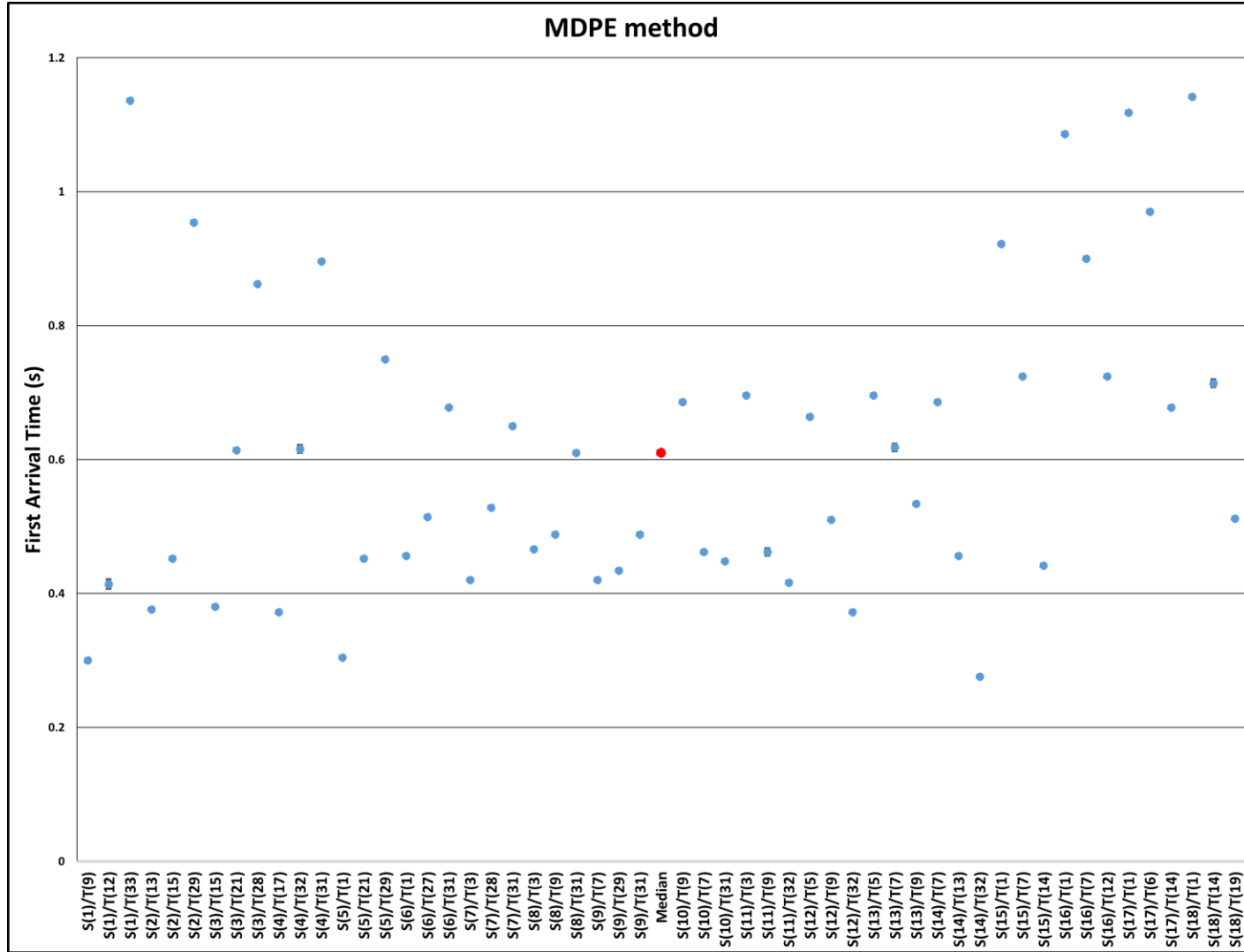


Figure 4.12: Results of applying the proposed MDPE method on 54 selected traces from the southeast Texas data set. Trace (T) and shot (S) numbers are indicated on the x-axis while the first arrival traveltime position of each trace is indicated on the y-axis. The blue filled circles indicate first arrival estimated by the MDPE method on each of the traces indicated on the x-axis while the error bar indicates its deviation from manual picking. The red concentric circle represents the median first arrival traveltime of all 54 traces together with the median error from manual picks.

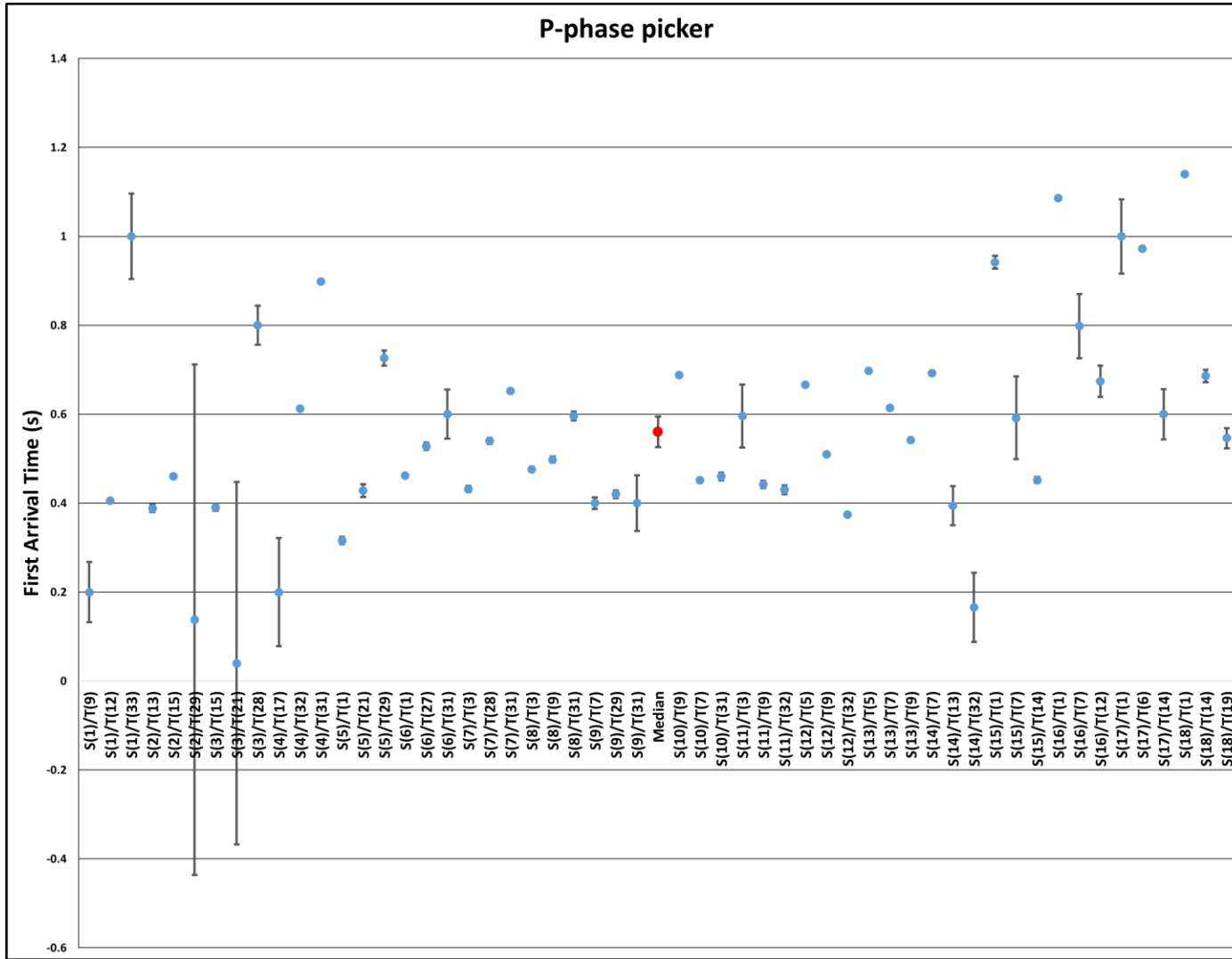


Figure 4.13: Results of applying the P-phase picking method on the same 54 traces shown in the previous figure. The blue filled circles indicate first arrival estimated by the P-phase picking method on each of the traces (T) indicated on the x-axis while the error bar indicates its deviation from manual picking. The red concentric circle in the middle of Figure represents the median error from manual picks. Note the high errors compared to the proposed method.

2. Yilmaz's real shot gathers

I have chosen two real shot gathers (6 and 25) from Yilmaz's published assortment of worldwide real seismic data (Yilmaz, 2001) to test the MDPE method and compare it with manual picking. Figure 4.14(a) shows shot gather 6 that contains 48 traces with a time sample interval of 0.004 s and a receiver interval of 100 m using a dynamite source. The first arrival traveltimes determined by MDPE method (crosses) is shown in Figure 4.15(a) compared with manual picks (circles). The picks from the MDPE method have 83% accuracy from manual picking (within 20 ms of manual picks).

Shot gather 25 is shown in Figure 4.14(b) where data contains 96 traces with 0.002-s time sampling interval and 50-m receiver interval using a dynamite source. The first arrival traveltimes determined by the MDPE method (crosses) are shown in Figure 4.15(b) compared with manual picks (circles). The picks from the MDPE method are very close to the manual picks with picking accuracy 91% (within 20 ms of manual picks) except for few traces at very near offsets requiring a sliding window length less than 50 samples and at very far offsets due to the very weak first arrivals. Median errors of both Yilmaz's shot gather 6 and 25 are 0% as shown in Table 4.1.

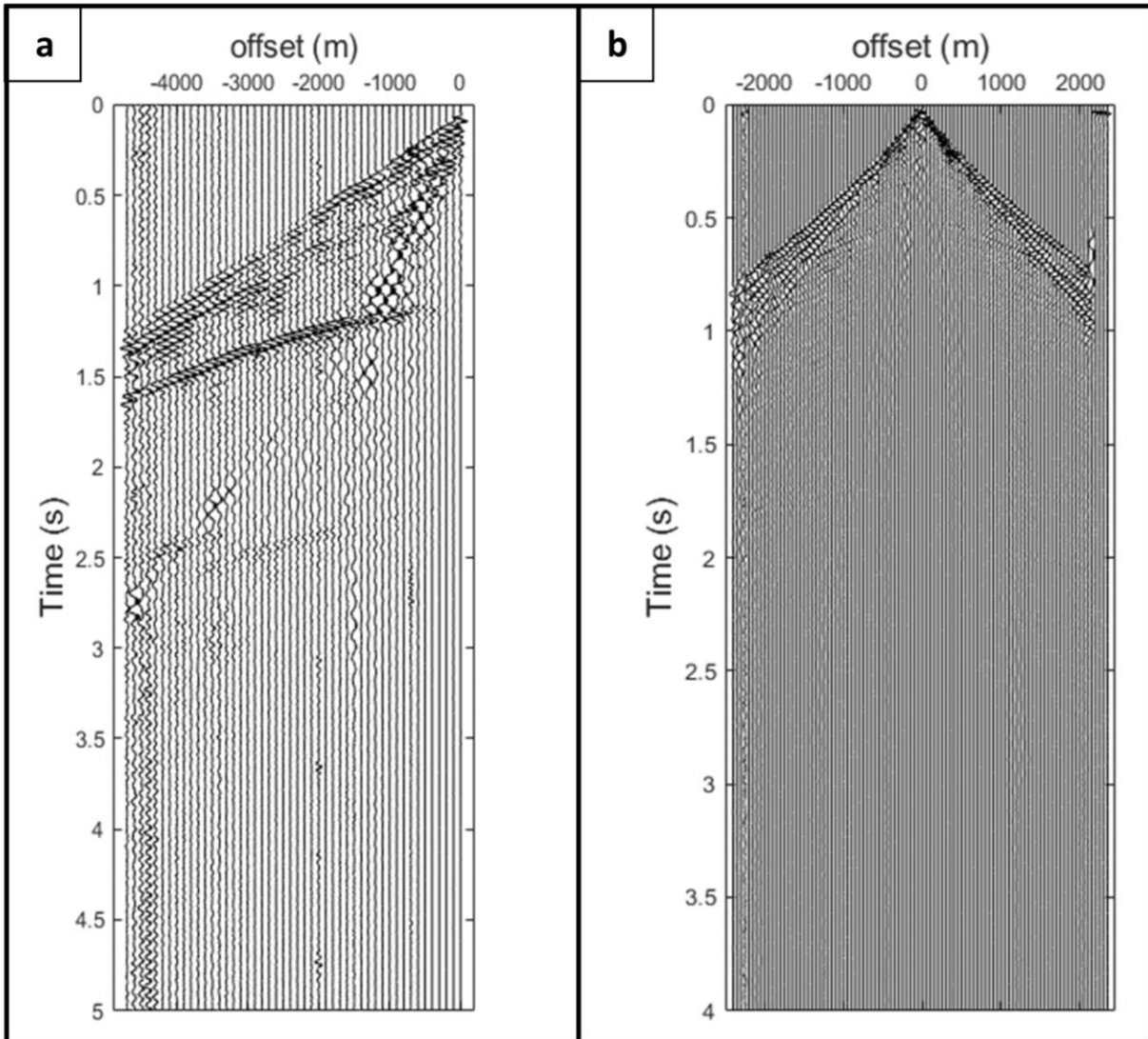


Figure 4.14: Raw shot records selected from Yilmaz (2001). (a) Shot record 6; (b) Shot record 25.

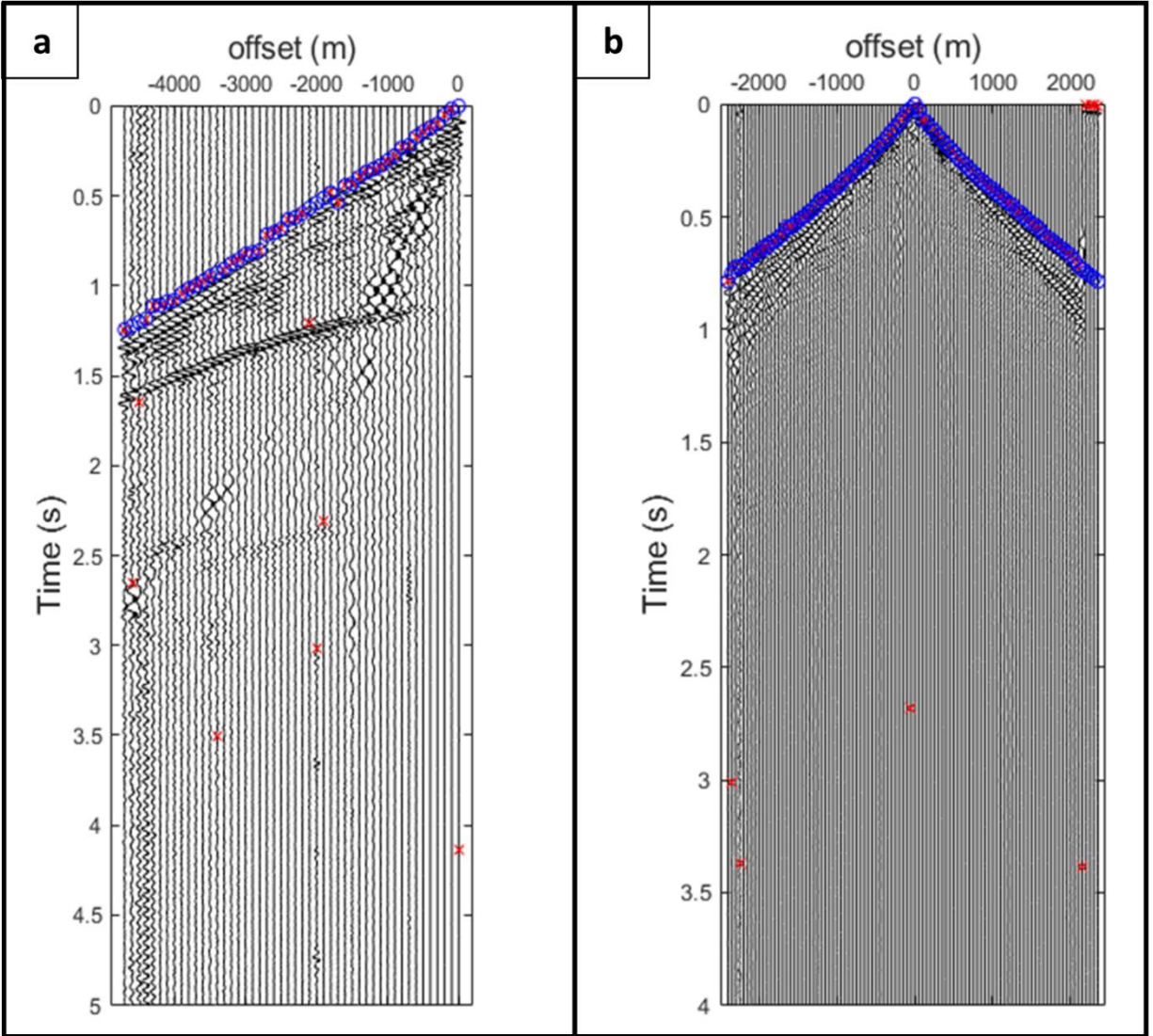


Figure 4.15: The same two Yilmaz's shot records shown in Figure 4.14 with a comparison of first arrival picks between the MDPE method (crosses) and the manual picks (circles). (a) Shot gather 6; (b) Shot gather 25.

4.1.2 Discussion

The MDPE method works with raw data requiring no type of gain to the data. It can be used for picking weak traces that pose difficulties for other methods. Another benefit of the method is that it needs only the sliding window length (k) from the user, while the rest is calculated automatically. To compute the differences between the adjacent points along the time samples, it is important to convert the envelope of the complex seismic trace into decibel units.

An initial choice for the length of the median's sliding window is the period of the dominant signal, although this value depends on the data itself. In general, an optimum window length should produce a spike in the data near the first arrival travelttime in the envelope (in dB scale) after calculating the median. This is important to obtain the maximum difference at the correct first arrival position.

Moreover, there is a possible problem with the MDPE method in the near offset traces (e.g., direct wave events). Near offset traces need smaller sliding windows to produce a median with the correct trace spike although more fluctuations in the data will be generated. Therefore, one must assure that the envelope after computing the median looks smooth and the spike is still at the same position as in the original envelope. In this case, it is preferable to choose a different sliding window for the direct wave events than head wave events to enhance the accuracy of picking first arrival events.

Figure 4.10(b) shows the importance of picking first arrival travelttime as the maximum point of the envelope in decibel rather than picking first arrival travelttime directly as the first jump of the envelope. In this example, trace 29 shot gather 2 from southeast Texas dataset, the P-phase picker method failed to pick the first arrival travelttime because of the noisy event before the main event.

On the other hand, the MDPE method was capable to differentiate between the first arrival traveltimes event and the noisy event before it.

There are some picked traces that are considered as outliers in Yilmaz's shot records. Some of these traces are bad traces (e.g., traces numbers 2, 3, 27, and 28 in Yilmaz's shot gather 1 and traces numbers 4, 92, 93, 94, 95, and 96 in Yilmaz's shot gather 25), while some of them are near offset traces (e.g., trace number 48 in Yilmaz's shot gather 1 and 25) where the first arrival picking sample is less than the sliding window (50 samples).

All real data used in this work have a dynamite source producing a minimum-phase wavelet. In case we need to apply the MDPE method on a data set with non-minimum phase wavelet (e.g., Vibroseis source), it is important to convert the wavelet to its minimum-phase equivalent using an appropriate spectral decomposition method (Yilmaz, 2001). Finally, the MDPE method needs more testing with different data sets to understand its limitations.

4.2 First Arrival Traveltime Picking Based on DBSCAN Method

4.2.1 Results

A velocity model (Figure 4.16) is used to generate a synthetic data set using a least-squares finite-difference solution to the elastic wave equation. A Ricker wavelet with a 100-Hz peak frequency is used in this synthetic data set as the source wavelet, and the generated shot gather has 96 receivers with 0.5 m receiver interval, 0.001 s time sampling interval, and 0.25 s total recording time (Figure 4.17). This data mimics a field data set recorded in northeast Saudi Arabia on a small geological outcrop as shown in Figure 4.16.

To understand how the DBSCAN method works, I show Figures of the workflow for one selected window from the data set. The method is applied on the generated shot gather where the sliding window has a size of 10 traces (5 m) wide and 30 samples (30 ms) long as shown in Figure 4.18. Trace shift and sample shift equal to one trace (0.5 m) and one sample (1 ms) long, respectively. A range of velocities, between 200 m/s and 800 m/s with a 10 m/s velocity interval, is applied on equation (3.7) to calculate the linear moveout (LMO) shift. Figure 4.19 shows the result of the stacked gather after applying LMO. To remove low amplitude output energies, a 50% threshold is applied to the stacked gather (Figure 4.20).

The DBSCAN's parameters are selected to be $\epsilon=5$ and $\text{MinPts}=10$. The result of applying the DBSCAN algorithm on the stacked gather is shown in Figure 4.21 where the algorithm identified only one cluster in this case and the centroid point is selected to be the first arrival traveltime for the first trace in the selected window. The first arrival traveltimes estimated from the DBSCAN method is compared with manual picks. Figure 4.22 shows the application of the DBSCAN method

and manual picks on the generated shot gather. Figure 4.23 shows the comparison between the DBSCAN method picking and manual picking, while Figure 4.24 and 4.25 show the difference between them, respectively. The results of the DBSCAN approach show an error between the DBSCAN method and manual picking ranging between -1.16 and 7 ms with an average value of 1.98 ms (Figure 4.24), which is smaller than the period of the first arrival traveltime wavelet (10 ms). Median error of the DBSCAN method picks using the synthetic data set is 6% as shown in Table 4.1.

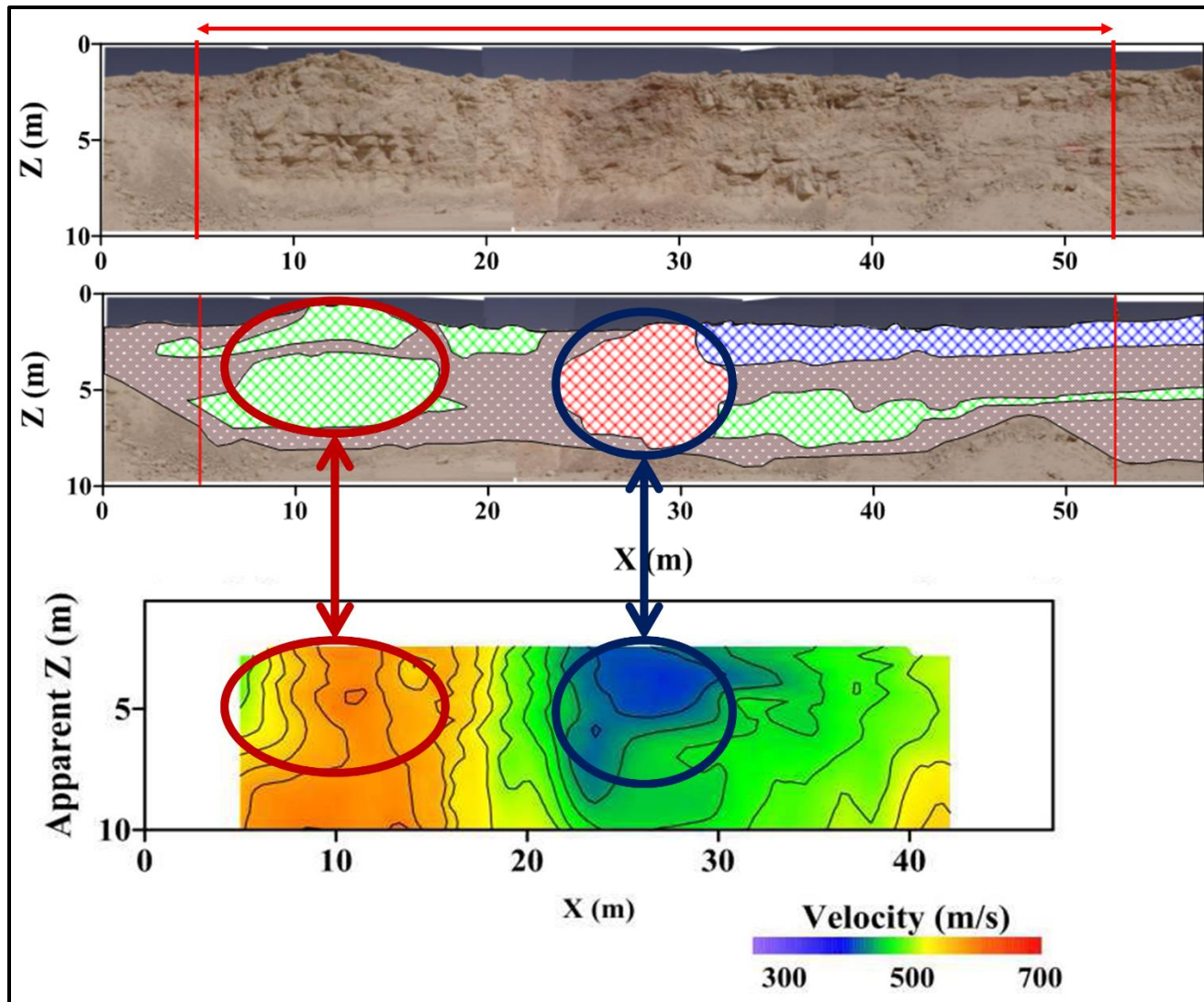


Figure 4.16: Velocity model used to generate the DBSCAN synthetic data set.

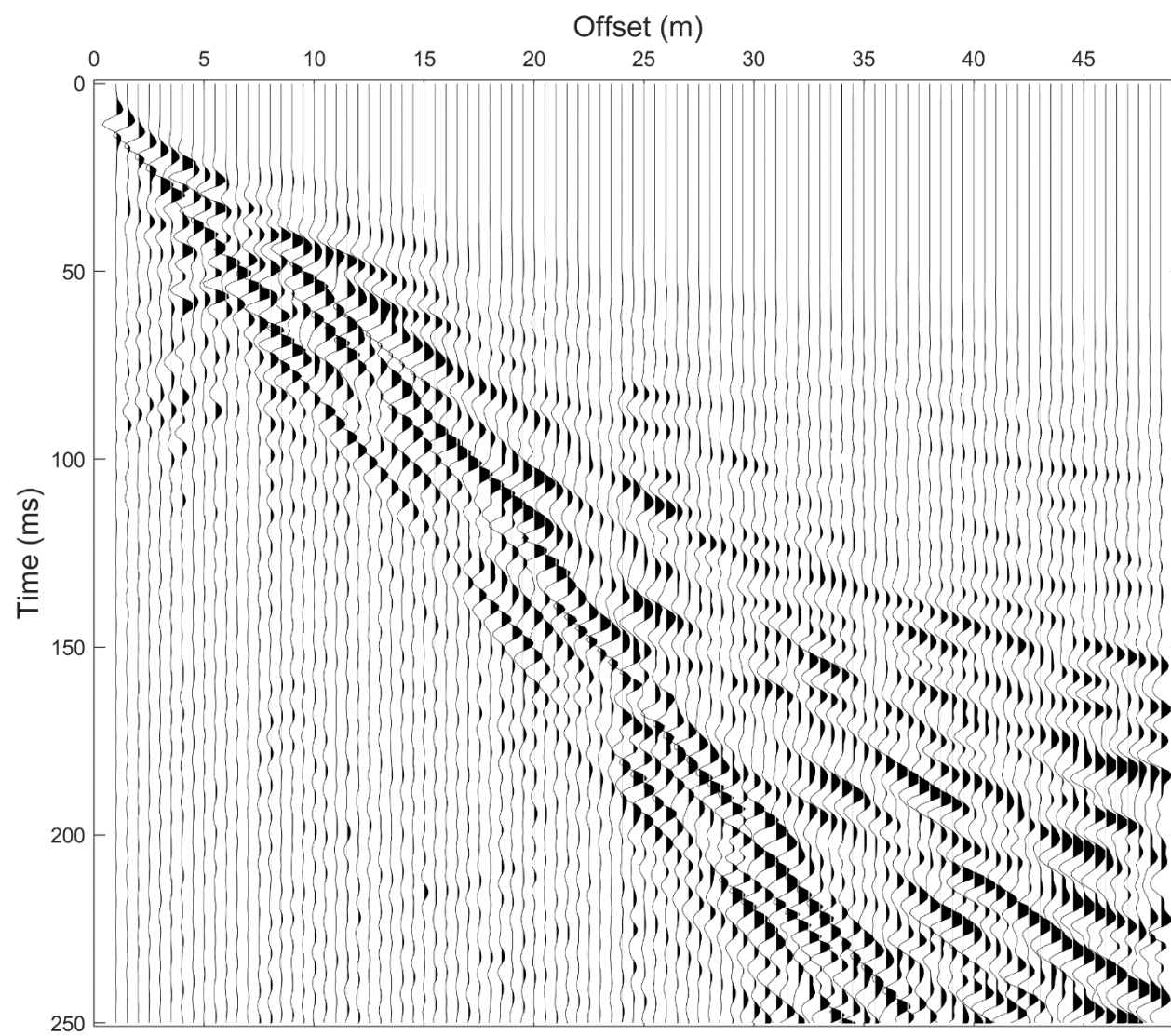


Figure 4.17: Sample shot gather generated from the synthetic seismic data.

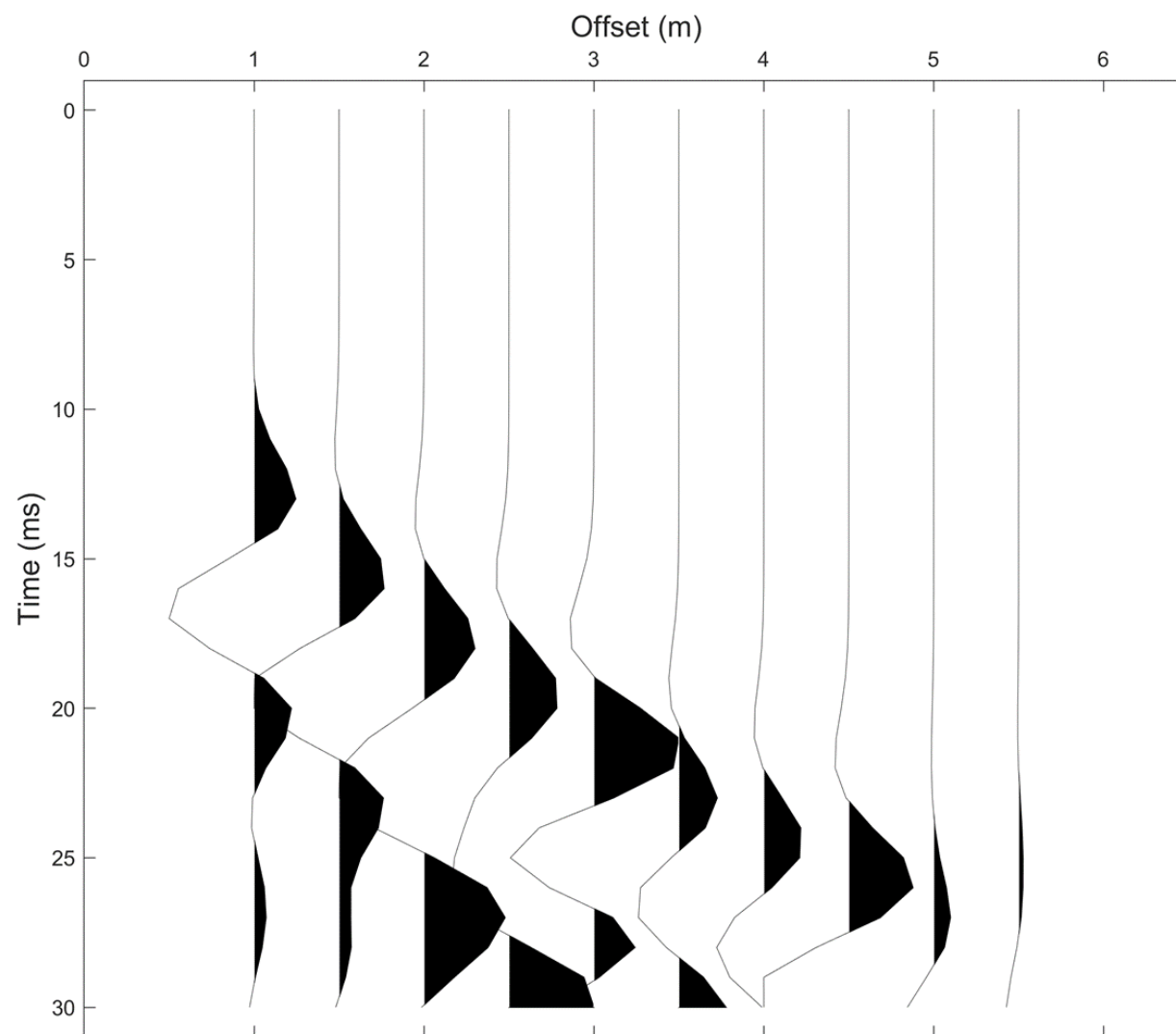


Figure 4.18: Sample selected window with a size of 10 traces wide and 30 samples long.

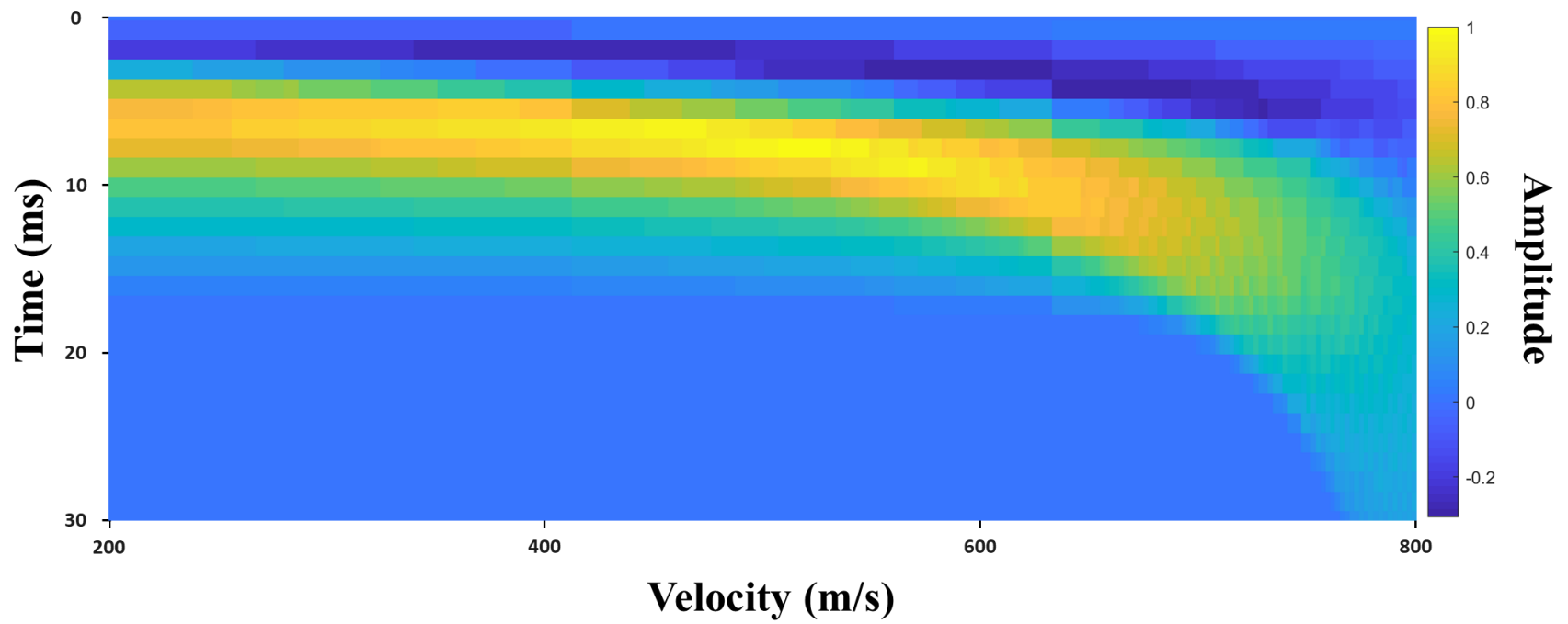


Figure 4.19: Stacked gather for the selected window (Figure 4.18) after applying a linear-moveout (LMO) shift for each velocity value (200-800 m/s).

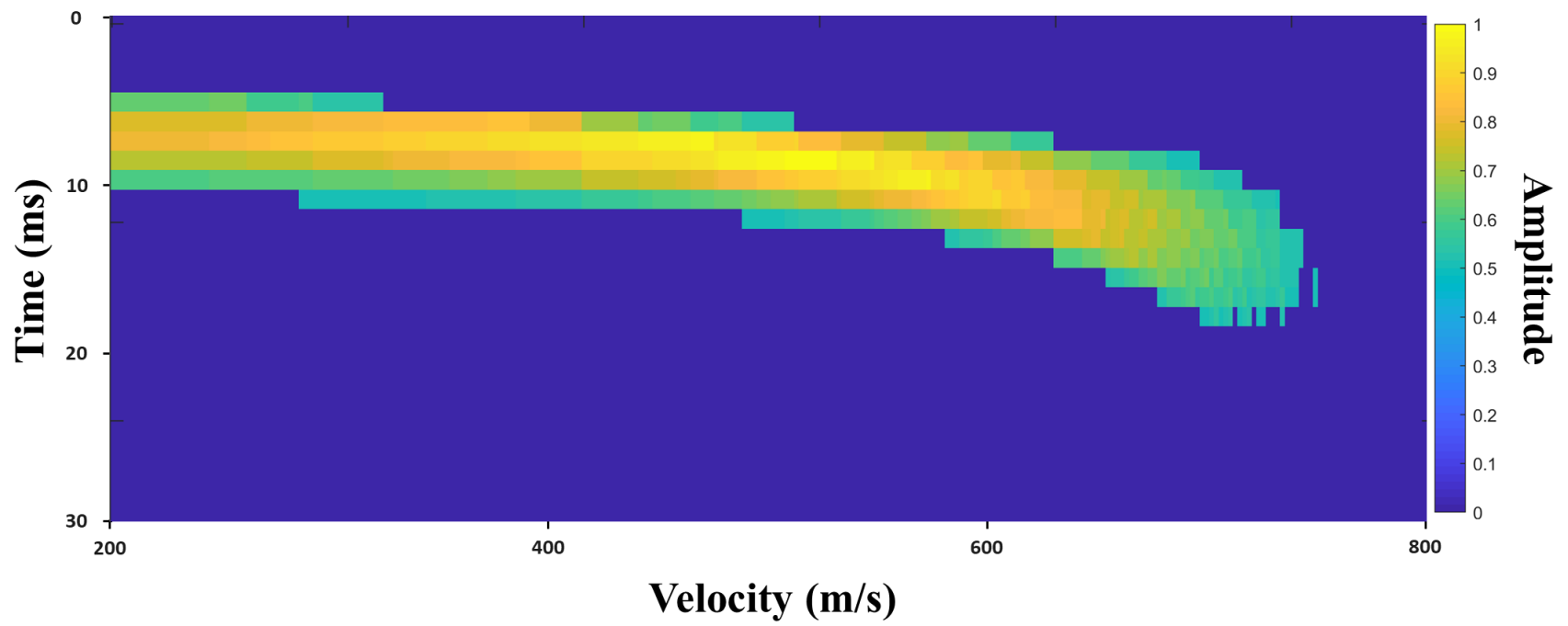


Figure 4.20: Stacked gather (Figure 4.19) after applying 50% threshold to remove low amplitude output energies.

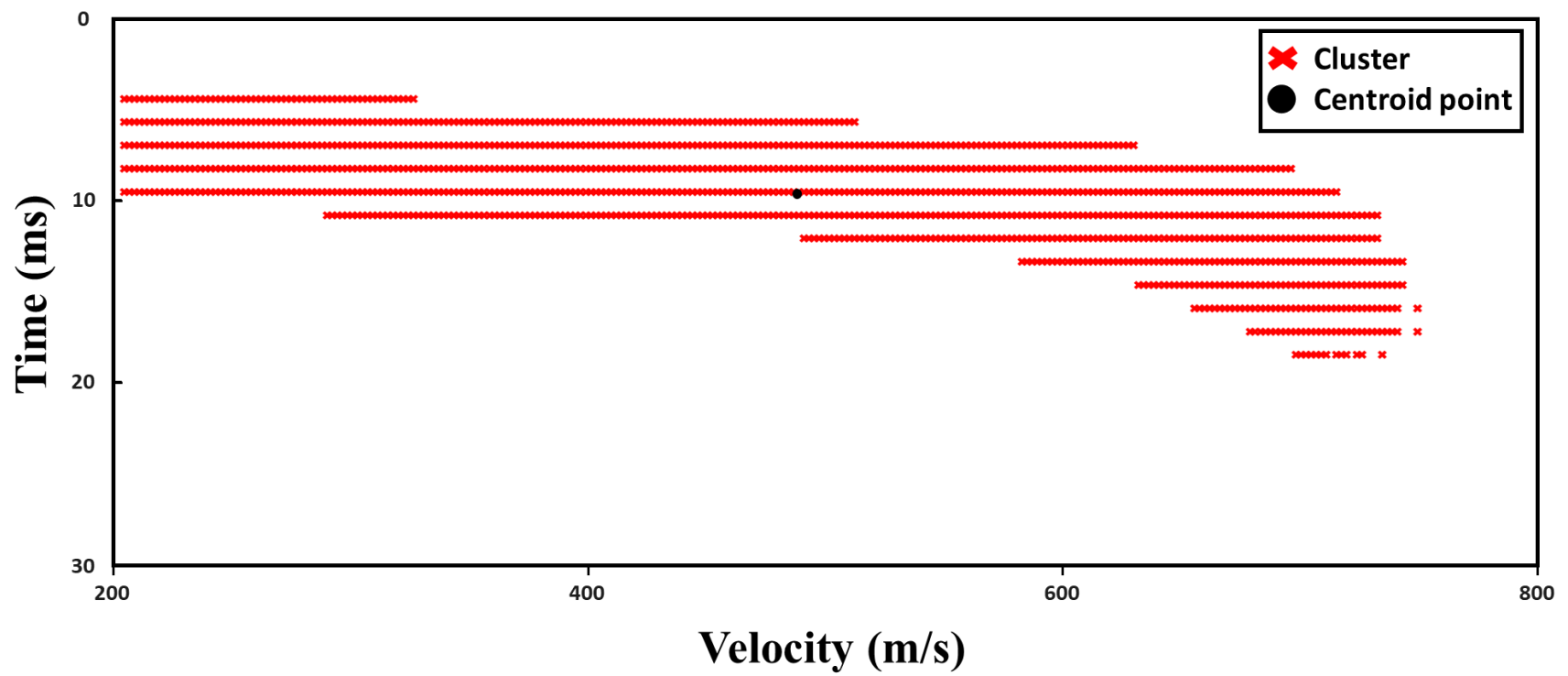


Figure 4.21: The output of the DBSCAN clustering algorithm with one cluster. The centroid (dot symbol) represents the optimum velocity value and the first arrival traveltime pick.

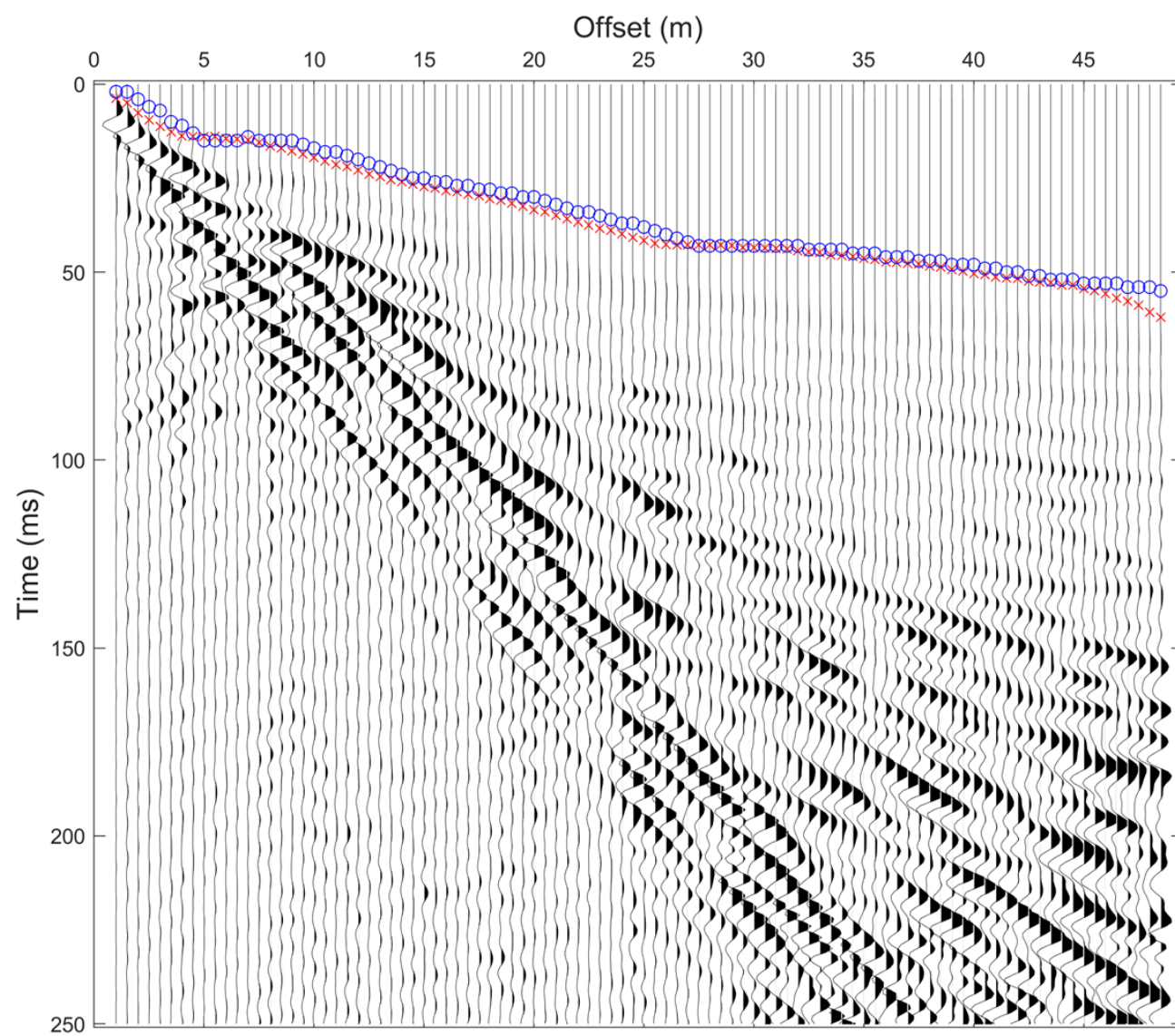


Figure 4.22: The same shot gather shown in Figure 4.17 with a comparison of first arrival picks between the DBSCAN method (crosses) and the true picks (circles).

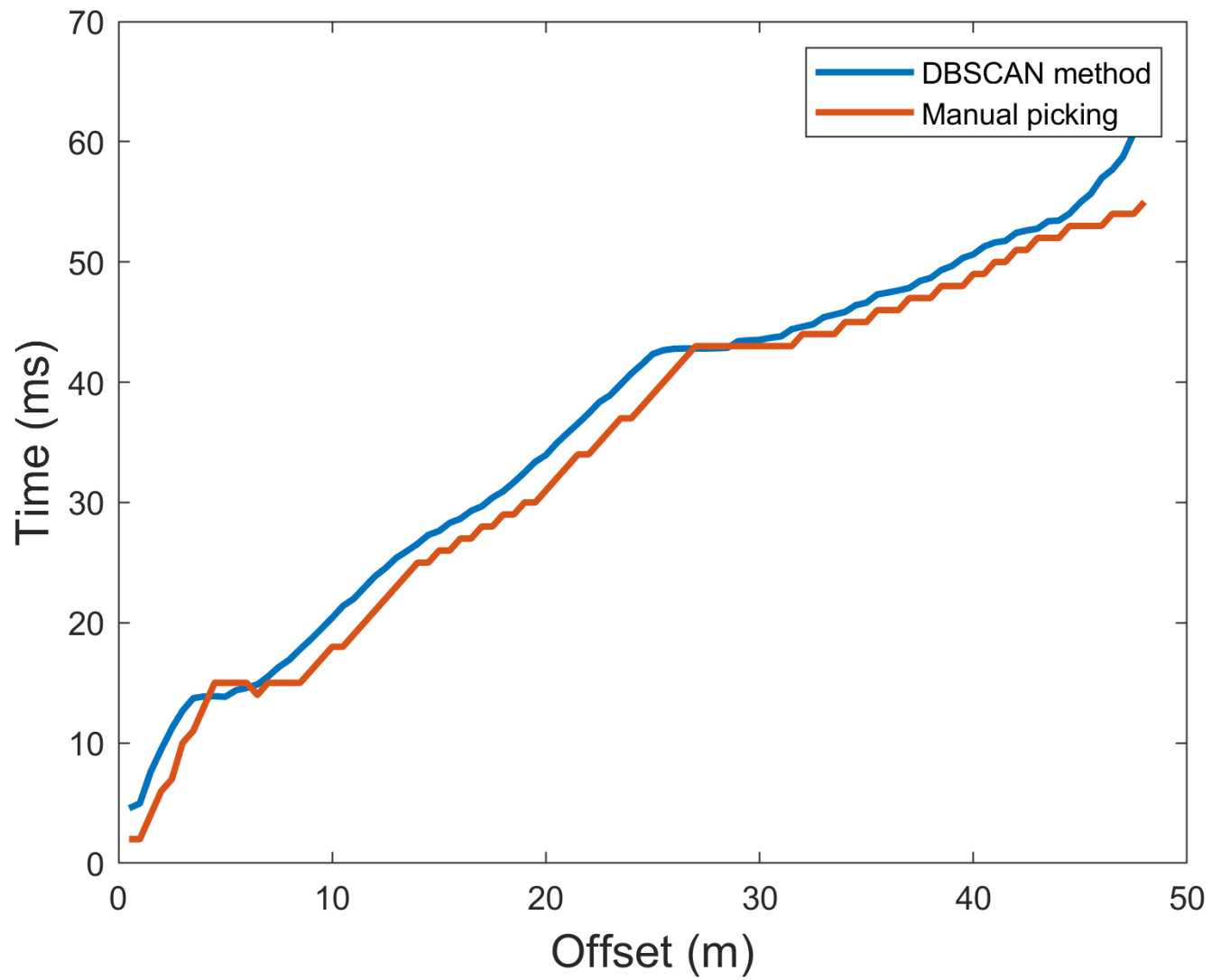


Figure 4.23: A comparison between the manual (red) and the DBSCAN method (blue) picking.

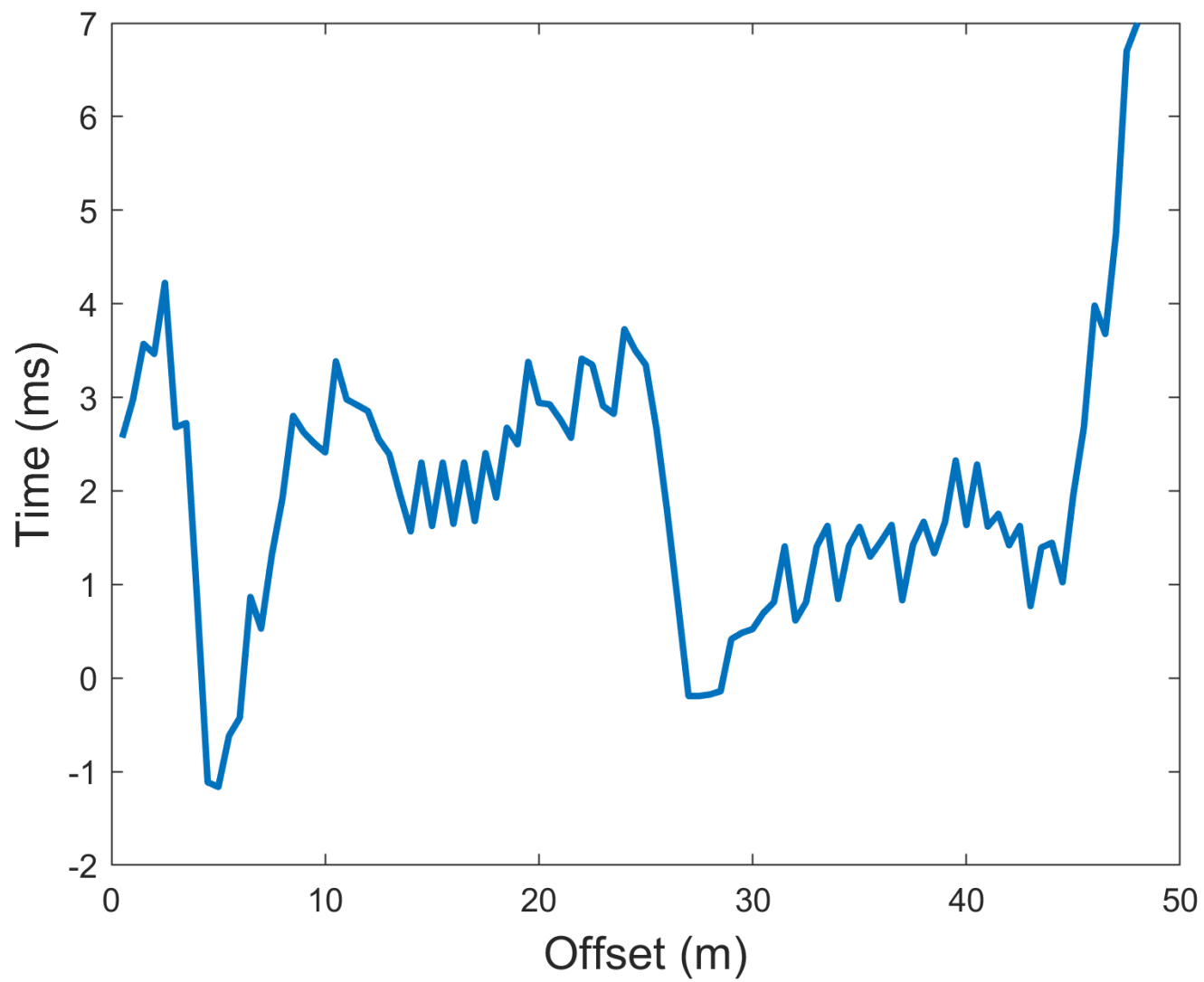


Figure 4.24: The differences in picking between the manual and DBSCAN method.

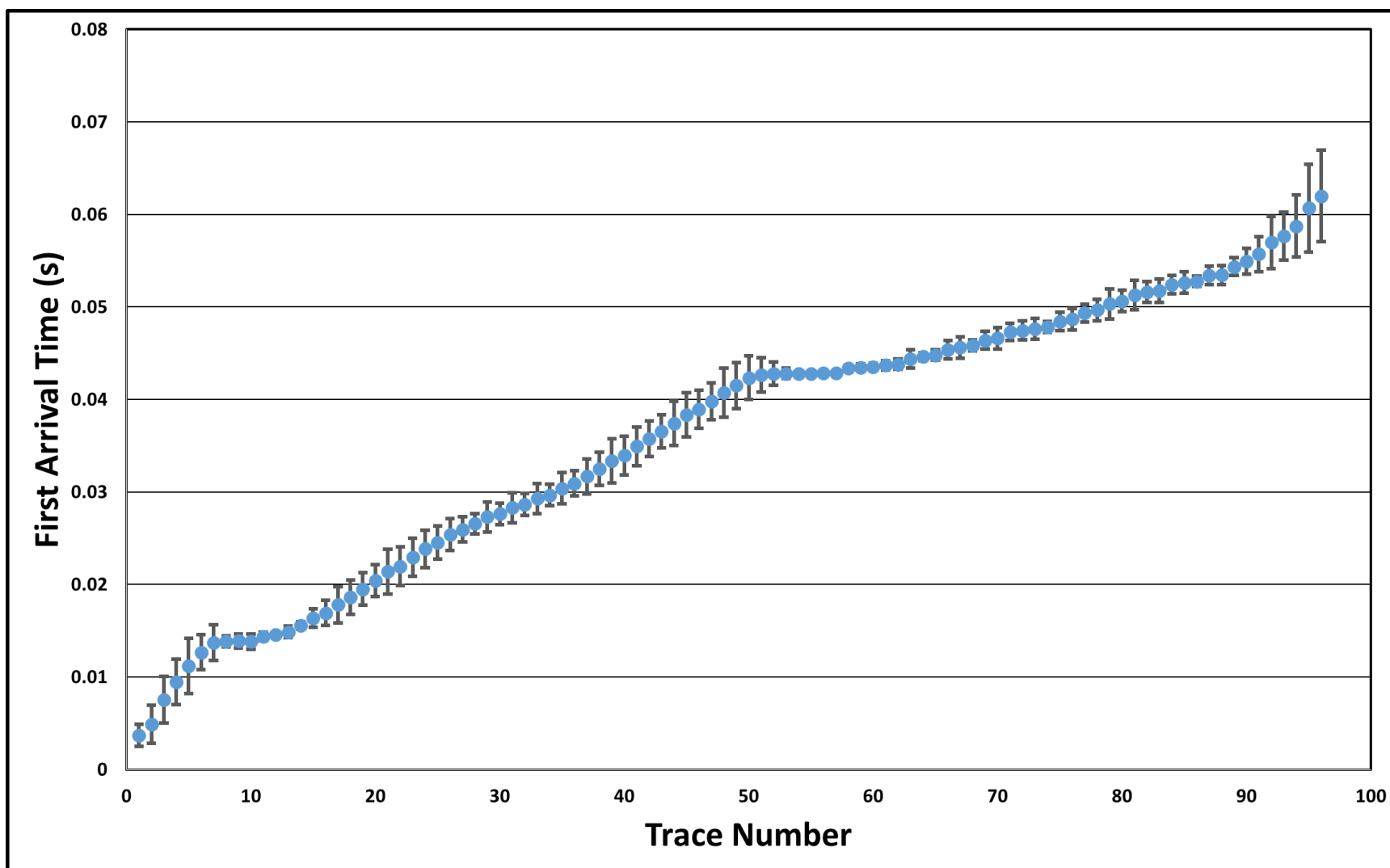


Figure 4.25: Results of applying the DBSCAN method on the synthetic shot gather (Figure 4.17) traces. Filled circles indicate first arrival estimated by the DBSCAN method while the error bar indicates its deviation from manual picking.

4.2.2 Discussion

The DBSCAN method requires a careful examination of the recorded data set to select a range of apparent velocities that include optimum velocity values for all traces in the recorded data. The importance of these optimum velocity values is to level first arrival events for both direct and head waves within the recorded data set. Therefore, it is necessary to choose a range of velocities, 200-800 m/s in this study, that include all needed values to level events for the whole recorded data set. In case the difference in velocity values between each layer and the next one is too large, it is preferable to specify a range of velocities for each layer separately to increase the effectiveness of the DBSCAN method and enhance the accuracy of picking first arrivals.

The DBSCAN method, unlike MDPE method, requires a type of gain such as AGC used in this study to the raw data before applying any step, especially for a data set that has weak events. This is an important step to make sure that the DBSCAN algorithm picks the centroid of the maximum cluster as the first arrival event.

Moreover, there is a decrease in picking accuracy in the last traces, 8 traces in this study, in the recorded data set (Figure 4.24). The reason for that is the decrease of traces number in the selected window and therefore decrease in the amplitude values within the first arrival event in the stacked gather which leads to reducing the picking accuracy. To overcome this issue, the selected window could be flipped for the last traces to choose enough traces from previously selected windows to increase amplitude values in the stacked gather and therefore enhance the first arrival picking accuracy.

Finally, to enhance the picking accuracy, window size parameters, and DBSCAN algorithm parameters should be carefully selected. Window size parameters, 10 traces (5 m) wide and 30

samples (30 ms) long in this study, should be selected carefully to avoid missing first arrival events location. On the other hand, for the DBSCAN parameters, it is recommended to choose a large value for MinPts and a small value of ϵ .

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 First Arrival Traveltime Picking Based on MDPE Method

5.1.1 Conclusion

A new automatic first arrival traveltime picking method is proposed. The proposed MDPE method picks the first arrival traveltime automatically based on the maximum difference between adjacent points of the envelope of a complex seismic trace. At the beginning, the study introduced the Hilbert transform and how we apply it on the seismic trace to produce the imaginary part of the complex seismic trace. After that, the complex seismic trace envelope has been calculated as the magnitude of the complex seismic trace. Moreover, the envelope has been converted to decibel scale to obtain a spike near the position of the first arrival traveltime. Next, a sliding median filter of the envelope was calculated to reduce its fluctuations and make it smoother. Then, the difference between the envelope values of adjacent points was calculated. Finally, the first arrival traveltime was picked at the maximum difference.

The proposed method was tested on two synthetic shot records and compared with the manual picking and found that the method estimated the manual picks with more than 75% picking accuracy. Moreover, I used the same test on two of Yilmaz's published real seismic shot gathers and found that the method estimated the manual picks with a picking accuracy of more than 83%. The proposed method was tested on real data from southeast Texas and compared with the P-phase

picker method. The results from that data show that the proposed method performs better than the P-phase picker method in most cases.

The MDPE method, in general, is working properly with datasets (e.g., Yilmaz shot gathers 6 and 25 and Southeast Texas datasets) where they have a dynamite source producing a minimum-phase wavelet. So, in the case of having a non-minimum phase wavelet (e.g., the synthetic dataset with zero phase wavelet), the method needs to be adjusted or convert the wavelet to its minimum-phase equivalent to enhance picking accuracy.

Another point to be considered is that the MDPE method is sensitive to noise since it is applying the methodology steps trace-by-trace instead of several adjacent traces. Therefore, to enhance picking accuracy in low SNR using the MDPE method, an extra step should be applied to data by selecting a window of adjacent traces and then stack the traces within this selected window.

5.1.2 Recommendations

From the results of this study, the following are some recommendations for future research related to the MDPE method study:

1. All real data used in this study have a dynamite source with a minimum-phase wavelet. It would be nice to see how the MDPE method performs in other types of wavelets (e.g., zero or mixed phase).
2. The proposed MDPE method is used only on land seismic data in this study. It would be also nice to try it on marine data.
3. The proposed MDPE method is used to pick the first arrival events in this study. The method could be adapted to pick later arrivals (e.g., picking reflections).

5.2 First Arrival Traveltime Picking Based on DBSCAN Method

5.2.1 Conclusion

An unsupervised machine learning technique is proposed in this study to pick the first arrival traveltime. The proposed DBSCAN method picks the first arrival traveltime automatically based on the density-based spatial clustering of applications with noise (DBSCAN) and the linear-moveout (LMO) shift. At the beginning, a sliding window was selected on both spatial and time directions to divide the recorded data into smaller parts. After that, the recorded traces in each window were shifted using a LMO approach for a constant velocity value. Then, the shifted traces were stacked to produce one trace corresponding to that velocity value. The process was repeated for different velocity values and a stacked gather is then generated. Moreover, DBSCAN was used to arrange the amplitude values of the stacked gather into different clusters. Finally, the centroid of the first cluster represented the optimum velocity and the first arrival traveltime. The proposed method was tested on a synthetic shot record and compared with manual picks. The result of the synthetic example found an error between the manual and DBSCAN picking that ranges between -1.16 and 7 ms with an average value of 1.98 ms.

The advantage of the DBSCAN method over supervised machine learning techniques is that the proposed method does not require prior labels or training. This means that the proposed DBSCAN method can be applied in both small and large datasets without much prior knowledge.

Another point to be considered is that the DBSCAN method can be applied on a dataset with random noise because the SNR of the random noisy data will be enhanced after applying LMO shift and stacking the shifted traces. On the other hand, the method might not work properly in a

dataset with a coherent noise since SNR will be decreased after applying LMO shift and stacking on shifted traces. Therefore, to compare both MDPE and DBSCAN methods, the DBSCAN is recommended to be applied on a dataset with a random noise.

5.2.2 Recommendations

From the results of this study, the following are some recommendations for future research related to the DBSCAN method study:

1. The proposed DBSCAN method is used only with synthetic data set. It would be nice to see how the method performs in real data set with different seismic sources (e.g., land and marine sources).
2. The proposed DBSCAN method is used in synthetic data set without adding noise. It would be nice to try it with different levels of noise to test its performance and its limitation.
3. The proposed DBSCAN method is used to pick the first arrival events in this study. The method could be adapted to pick later arrivals (e.g., picking reflections).

References

- [1] Akaike, H.; "Information theory and an extension of maximum likelihood principle," *In Proc. 2nd Int. Symp. on Information Theory*, pp. 267-281, 1973.
- [2] Akazawa, T; "A technique for automatic detection of onset time of P-and S-phases in strong motion records," *In Proc. of the 13th World Conf. on Earthquake Engineering*, August. 2004
- [3] Akram, J.; Eaton, D.W.; "A review and appraisal of arrival-time picking methods for downhole microseismic dataArrival-time picking methods," *Geophysics*, vol. 81, no. 2, pp.KS71-KS91, 2016.
- [4] Akram, J.; Eaton, D.; Onge, A.S.; "Automatic event-detection and time-picking algorithms for downhole microseismic data processing," *In 4th EAGE passive seismic workshop*, European Association of Geoscientists & Engineers, pp. cp-337, March. 2013
- [5] Al-Ghamdi, A. S.; "Automatic first arrival picking using energy ratios," *M.S. thesis*, King Fahd Univ. Petroleum Minerals, Dhahran, Saudi Arabia, 2007.
- [6] Ali, H. Z.; "Application of Static Correction Technique on 2D Land Seismic Data in the South of Iraq," *Iraqi Journal of Science*, vol. 60, no. 4, pp. 830-839, 2019.
- [7] Allen, R.V.; "Automatic earthquake recognition and timing from single traces," *Bulletin of the Seismological Society of America*, vol. 68, no. 5, pp.1521-1532, 1978.
- [8] Bai, C.Y.; Kennett, B.L.N.; "Phase identification and attribute analysis of broadband seismograms at far-regional distances," *Journal of seismology*, vol. 5, no. 2, pp.217-231, 2001.
- [9] Barnes, A.E.; "A tutorial on complex seismic trace analysis," *Geophysics*, vol. 72, no. 6, pp.W33-W43, 2007.
- [10] Becerra, C.; Agudelo, W.; Guevara, S.; "Uncertainty analysis in statics corrections obtained by tomographic inversion: Application in a mountainous zone in Catatumbo (Colombia)," *The Leading Edge*, vol. 28, no. 2, pp. 212-215, 2009.
- [11] Boschetti, F.; Dentith, M. D.; List, R. D.; "A fractal-based algorithm for detecting first arrivals on seismic traces," *Geophysics*, vol. 61, no. 4, pp. 1095-1102, 1996.
- [12] Chopra, S.; Marfurt, K.J.; "Seismic attributes—A historical perspective," *Geophysics*, vol. 70, no. 5, pp.3SO-28SO, 2005.

- [13] Coppens, F.; "First arrival picking on common-offset trace collections for automatic estimation of static corrections," *Geophysical Prospecting*, vol. 33, no. 8, pp. 1212-1231, 1985.
- [14] Cox, M.; "Static corrections for seismic reflection surveys," *Society of Exploration Geophysicists*, 1999.
- [15] De Meersman, K.; Kendall, J.M.; Van der Baan, M.; "The 1998 Valhall microseismic data set: An integrated study of relocated sources, seismic multiplets, and S-wave splitting," *Geophysics*, vol. 74, no. 5, pp.B183-B195, 2009.
- [16] Diehl, T.; Deichmann, N.; Kissling, E.; Husen, S.; "Automatic S-wave picker for local earthquake tomography," *Bulletin of the Seismological Society of America*, vol. 99, no. 3, pp.1906-1920, 2009.
- [17] Duan, X.; Zhang, J.; "Multi-trace first break picking using an integrated seismic and machine learning method," *Geophysics*, vol. 85, no. 4, pp.1-58, 2020.
- [18] Duan, X.; Zhang, J.; Liu, Z.; Liu, S.; Chen, Z.; Li, W.; "Integrating seismic first-break picking methods with a machine learning approach," *In SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, pp. 2186-2190, 2018.
- [19] Earle, P. S.; Shearer, P. M.; "Characterization of global seismograms using an automatic-picking algorithm," *Bulletin of the Seismological Society of America*, vol. 84, no. 2, 366-376, 1994.
- [20] Ervin, C. P.; McGinnis, L. D.; Otis, R. M.; Hall, M. L.; "Automated analysis of marine refraction data: A computer algorithm," *Geophysics*, vol. 48, no. 5, 582-589, 1983.
- [21] Ester, M.; Kriegel, H. P.; Sander, J.; Xu, X.; "A density-based algorithm for discovering clusters in large spatial databases with noise," *In Proceedings of the second international conference on knowledge discovery and data mining*, vol. 96, no. 34, pp. 226-231, 1996.
- [22] Gaci, S.; "The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp.4558-4563, 2013.
- [23] Gelchinsky, B.; Shtivelman, V.; "Automatic picking of first arrivals and parameterization of traveltimes curves," *Geophysical Prospecting*, vol. 31, no. 6, pp. 915-928, 1983.
- [24] Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y.; "Deep learning," vol. 1, 2016.
- [25] Han, L.; Wong, J.; Bancroft, J.C.; "Time picking and random noise reduction on microseismic data," *CREWES Research Report*, vol. 21, pp.1-13, 2009.

- [26] Hatherly, P. J.; "A computer method for determining seismic first arrival times," *Geophysics*, vol. 47, no. 10, pp. 1431-1436, 1982.
- [27] Hollander, Y.; Merouane, A.; Yilmaz, O.; "Using a deep convolutional neural network to enhance the accuracy of first-break picking," In *SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, pp. 4628-4632, 2018.
- [28] Hu, L.; Zheng, X.; Duan, Y.; Yan, X.; Hu, Y.; Zhang, X.; "First-arrival picking with a U-net convolutional network," *Geophysics*, vol. 84, no. 6, pp.U45-U57, 2019.
- [29] Jiao, L.; Moon, W.M.; "Detection of seismic refraction signals using a variance fractal dimension technique," *Geophysics*, vol. 65, no. 1, pp.286-292, 2000.
- [30] Johansson, M.; "The hilbert transform," *Mathematics Master's Thesis, Växjö University, Suecia. Disponible en internet: http://w3.msi.vxu.se/exarb/mj_ex.pdf, consultado el*, vol. 19, 1999.
- [31] Kalkan, E.; "An automatic P-phase arrival-time picker," *Bulletin of the Seismological Society of America*, vol. 106, no. 3, pp.971-986, 2016.
- [32] Kassambara, A.; "Practical guide to cluster analysis in R: Unsupervised machine learning," *Sthda*, vol. 1, 2017.
- [33] Küperkoch, L.; Meier, T.; Lee, J.; Friederich, W.; EGELADOS Working Group; "Automated determination of P-phase arrival times at regional and local distances using higher order statistics," *Geophysical Journal International*, vol. 181, no. 2, pp.1159-1170, 2010.
- [34] Kusuma, T.; Fish, B.C.; "Toward more robust neural-network first break and horizon pickers," In *SEG Technical Program Expanded Abstracts 1993*, Society of Exploration Geophysicists, pp. 238-241, 1993.
- [35] Leonard, M.; "Comparison of manual and automatic onset time picking," *Bulletin of the Seismological Society of America*, vol. 90, no. 6, pp.1384-1390, 2000.
- [36] Li, Z.; Sheng, R.; Xu, W.; Yang, J.; "The improvement of neural network cascade-correlation algorithm and its application in picking seismic first break," *Advances in Petroleum Exploration and Development*, vol. 5, no. 2, pp.41-51, 2013.
- [37] Lu, K.; Feng, S.; "Auto-windowed Super-virtual Interferometry via Machine Learning: A Strategy of First-arrival Traveltime Automatic Picking for Noisy Seismic Data," In *SEG 2018 Workshop: SEG Maximizing Asset Value Through Artificial Intelligence and Machine Learning, Beijing, China, 17-19 September 2018*, Society of Exploration Geophysicists and the Chinese Geophysical Society, pp. 10-14, 2018.

- [38] Ma, Y.; Luo, Y.; "Automatic first-arrival picking with Reinforcement Learning," *In International Geophysical Conference, Beijing, China, 24-27 April 2018*, Society of Exploration Geophysicists and Chinese Petroleum Society, pp. 493-497, 2018.
- [39] McCormack, M.D.; Zaucha, D.E.; Dushek, D.W.; "First-break refraction event picking and seismic data trace editing using neural networks," *Geophysics*, vol. 58, no. 1, pp.67-78, 1993.
- [40] Mezyk, M.; Malinowski, M.; "Deep Neural Network and Multi-pattern Based Algorithm for Picking First-arrival Traveltimes. " *In 80th EAGE Conference and Exhibition 2018*, European Association of Geoscientists & Engineers, vol. 2018, no. 1, pp. 1-5, June. 2018
- [41] Mitchell, T.; "Machine Learning," *McCraw Hill*, 1996.
- [42] Molyneux, J.B.; Schmitt, D.R.; "First-break timing: Arrival onset times by direct correlation. " *Geophysics*, vol. 64, no. 5, pp.1492-1501, 1999.
- [43] Mousa, W.A.; Al-Shuhail, A.A.; "Processing of seismic reflection data using MATLAB™." *Synthesis Lectures on Signal Processing*, vol. 5, no. 1, pp.1-97, 2011.
- [44] Mousa, W. A.; Al-Shuhail, A. A.; "Enhancement of first arrivals using the τ -p transform on energy-ratio seismic shot records," *Geophysics*, vol. 77, no. 3, pp. V101-V111, 2012.
- [45] Mousa, W. A.; Al-Shuhail, A. A.; Al-Lehyani, A.; "A new technique for first-arrival picking of refracted seismic data based on digital image segmentation" *Geophysics*, vol. 76, no. 5, pp. V79-V89, 2011.
- [46] Murat, M.E.; Rudman, A.J.; "Automated first arrival picking: A neural network approach 1," *Geophysical Prospecting*, vol. 40, no. 6, pp.587-604, 1992.
- [47] Peraldi, R.; Clement, A.; "Digital processing of refraction data study of first arrivals" *Geophysical Prospecting*, vol. 20, no. 3, pp. 529-548, 1972.
- [48] Raymer, D.; Rutledge, J.; Jaques, P.; "Semiautomated relative picking of microseismic events," *In SEG Technical Program Expanded Abstracts 2008*, Society of Exploration Geophysicists, pp. 1411-1414, 2008.
- [49] Russell, S.; Norvig, P.; "Artificial intelligence: a modern approach," 2002.
- [50] Sabbione, J. I.; Velis, D.; "Automatic first-breaks picking: New strategies and algorithms" *Geophysics*, vol. 75, no. 4, pp. V67-V76, 2010.
- [51] Sander, J.; Ester, M.; Kriegel, H. P.; Xu, X.; "Density-based clustering in spatial databases: The algorithm gdbscan and its applications," *Data mining and knowledge discovery*, vol. 2, no, 2, pp. 169-194, 1998.

- [52] Saragiotis, C.D.; Hadjileontiadis, L.J.; Panas, S.M.; "PAI-S/K: A robust automatic seismic P phase arrival identification scheme," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 6, pp.1395-1404, 2002.
- [53] Saragiotis, C.D.; Hadjileontiadis, L.J.; Rekanos, I.T.; Panas, S.M.; "Automatic P phase picking using maximum kurtosis and/spl kappa/-statistics criteria. " *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 3, pp.147-151, 2004.
- [54] Senkaya, M.; Karsli, H.; "A semi-automatic approach to identify first arrival time: the cross-correlation technique (CCT)," *Earth Sciences Research Journal*, vol. 18, no. 2, pp. 107-113, 2014.
- [55] Sharma, B.K.; Kumar, A.; Murthy, V.M.; "Evaluation of seismic events detection algorithms," *Journal of the Geological Society of India*, vol. 75, no. 3, pp.533-538, 2010.
- [56] Sleeman, R.; 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, no. 1-4, pp.265-275, 1999.
- [57] Takanami, T.; Kitagawa, G.; "Estimation of the arrival times of seismic waves by multivariate time series model," *Annals of the Institute of Statistical mathematics*, vol. 43, no. 3, pp.407-433, 1991.
- [58] Taner, M. T.; Koehler, F.; Sheriff, R. E., "Complex seismic trace analysis," *Geophysics*, vol. 44, no. 6, pp. 1041-1063, 1979.
- [59] Taner, M.T.; Sheriff, R.E.; "Application of amplitude, frequency, and other attributes to stratigraphic and hydrocarbon determination: Section 2. Application of seismic reflection configuration to stratigraphic interpretation," pp. 301-327, 1977.
- [60] Trnkoczy, A.; "Understanding and parameter setting of STA/LTA trigger algorithm," *In New Manual of Seismological Observatory Practice (NMSOP)*, Deutsches GeoForschungsZentrum GFZ, pp. 1-20, 2009.
- [61] Tsai, K. C.; Hu, W.; Wu, X.; Chen, J.; Han, Z.; "First-break automatic picking with deep semisupervised learning neural network" *In SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, pp. 2181-2185, 2018.
- [62] Tselentis, G.A.; Martakis, N.; Paraskevopoulos, P.; Lois, A.; Sokos, E.; "Strategy for automated analysis of passive microseismic data based on S-transform, Otsu's thresholding, and higher order statistics" *Geophysics*, vol. 77, no. 6, pp.KS43-KS54, 2012.
- [63] Vidale, J.E.; "Complex polarization analysis of particle motion," *Bulletin of the Seismological society of America*, vol. 76, no. 5, pp.1393-1405, 1986.

- [64] Xie, P.; Boelle, J.; Blais, C; "Automatic Seismic First Arrival Picking With Deep-Learning" *In First EAGE/PESGB Workshop Machine Learning*, European Association of Geoscientists & Engineers, vol. 2018, no. 1, pp. 1-3, November. 2018
- [65] Yalcinoglu, L.; Stotter, C.; "Can Machines Learn To Pick First Breaks As Humans Do?," *In First EAGE/PESGB Workshop Machine Learning*, European Association of Geoscientists & Engineers, vol. 2018, no. 1, pp. 1-3, November. 2018
- [66] Yalcinoglu, L.; Millahn, K.; Fruhwirth, R.; "Automatic first break detection using support vector machine classification," *In 73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011*, European Association of Geoscientists & Engineers, pp. cp-238, May. 2011
- [67] Yilmaz; "Seismic data analysis: Processing, inversion, and interpretation of seismic data," *Society of exploration geophysicists*, Ö., 2001.
- [68] Yuan, S.; Liu, J.; Wang, S.; Wang, T.; Shi, P.; "Seismic waveform classification and first-break picking using convolution neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp.272-276, 2018.
- [69] Yuan, S.; Su, Y.; Wang, T.; Wang, J.; Wang, S.; "Geosteering phase attributes: A new detector for the discontinuities of seismic images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp.145-149, 2018.
- [70] Yung, S.K.; Ikelle, L.T.; "An example of seismic time picking by third-order bicoherence," *Geophysics*, vol. 62, no. 6, pp.1947-1952, 1997.
- [71] Zhang, H.; Thurber, C.; 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, no. 5, pp.1904-1912, 2003.
- [72] Zhang, J.J.; and Bentley, L.R.; "Complex seismic trace analysis and its application to time-lapse seismic surveys," *CREWES, Geology and Geophysics, University of Calgary*, 2000.
- [73] Zhu, W.; Beroza, G.C.; "PhaseNet: a deep-neural-network-based seismic arrival-time picking method," *Geophysical Journal International*, vol. 216, no. 1, pp.261-273, 2019.

Vitae

Name: Abdullah Abobakr Al-Mashhor

Nationality: Yemen

Date of Birth: May 22, 1993

Email: abdullahalmashhor@gmail.com

Address: KFUPM Box 6552
Dhahran 31261
Saudi Arabia

Al-Rawdha street, Tarim, Hadhramout
Tarim 0000
Yemen

Academic Background: King Fahd University of Petroleum and Mineral (KFUPM),
B.S. in Geophysics,
January 2018.

King Fahd University of Petroleum and Mineral (KFUPM),
M.Sc. in Geophysics,
November 2020.

Publication: Sherif M. Hanafy, **Abdullah Al-Mashhor**, Umair bin Waheed, & Kai Lu. "First Arrival Picking of Seismic Data based on the DBSCAN Clustering Algorithm". *Artificially Intelligent Earth Exploration Workshop*, Virtual Workshop, 23–26 November 2020.

Al-Mashhor, A. A., Al-Shuhail, A. A., Hanafy, S. M., & Mousa, W. A.; "First Arrival Picking of Seismic Data Based on Trace Envelope" *IEEE Access*, vol. 7, pp. 128806-128815, 2019.