UNIVERSITY OF CALGARY

Traveltime tomography for land seismic data

by

Bernard Ki-Yun Law

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

JULY, 2022

Abstract

Full wave form inversion (FWI) uses the full recorded wavefield to determine the physical properties of the subsurface. Due to the lack of wide aperture and low-frequency data, a reliable low-frequency starting velocity model is often required. A velocity model derived from pre-stack depth migration (PSDM) is typically used as the starting model for FWI. Iterative PSDM is an expensive process, and it also requires a velocity macro model. Traveltime tomography methods are efficient algorithms to construct a smooth velocity model using refraction and reflection traveltimes. However, with limitations caused by acquisition constraints, data quality and assumptions used in traveltime tomography methods, the velocity model determined by these methods can be sub-optimal and results in degradation in the depth image. In this thesis, I address several strategies to improve tomography and to incorporate error measurements from refraction and reflection waveforms into the tomographic inversion kernels. My goal is to help to alleviate the inherent limitations in traveltime tomography methods.

Acknowledgements

Bernie Law completed his thesis while fighting a terminal illness but passed away just before his thesis defense and was not able to fill in this acknowledgement section. I thank CREWES students on behalf of Bernie for sharing with him his enthusiasm for science. Also, as Bernie's advisor, I thank Bernie for all his hard work and help during the years he was a student in CREWES. All of Bernie's colleagues are grateful to him for being always ready to help and do any extra work required. We will miss the excellent discussions about science and his enthusiasm for being once again (after a long side trip in industry) a part of the student community. Bernie will be always remembered as an example of dedication and strength. He used his last days of life to finish this thesis and achieved his goal by being granted his Ph.D. on his last day.

Dedication

To my mother, Son-You and my five children Rhonda, Linda, Tina, Anita and Brian.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1

Introduction

1.1 Seismic imaging and velocity inversion

Seismic experiment records the ground motions caused by the propagation of controlled seismic sources. We can use the time delays, amplitude, and phase characteristics to estimate the physical properties of the subsurface. Traditional velocity inversion uses the time delays and the kinematics of the wave propagation between the seismic sources and receivers to determine a velocity model that can best match the measured time delays of the seismic data. Seismic imaging (migration) transforms seismic data to an image of lithological boundaries (impedance) by reversing the wave propagation using the supplied model of the physical parameters. Analyzing the seismic waveform of the migrated data can quantify the accuracy and errors of the model parameters. This allows the model parameters to be updated and results in a higher resolution model. Analyzing the seismic waveform of recorded seismic data can also reveal the absorption and dispersion properties of the lithological units.

The accuracy of our estimation of the reflectivity and rock properties depends greatly on the recorded seismic data and how we model the wave field propagation. Traditional velocity inversion methods use time-delay characteristics of the seismic data to capture the large-scale features of the velocity model. Seismic migration methods accurately use the traveltime, amplitude and phase of the wave propagation to produce a highly resolved seismic image. Although seismic migration velocity analysis can update the velocity model using curvatures in migrated image gathers, an accurate starting velocity model still plays an important role in producing a successfully migrated image. Full waveform inversion updates the model parameters, including velocity, by minimizing the difference between the recorded and modelled waveform. FWI can produce a higher resolution velocity model than traveltime tomography (Tarantola, 1984). Since FWI uses the full waveform, including the high-frequency components, it can have many local minima. Without an initial model that is reasonably close to the long-wavelength component of the true model, FWI can result in cycle skipping. Bunks et al. (1995) address this problem by iterating FWI from low-frequency band to high-frequency band. However, this requires very low frequency and/or wide aperture data (Virieux and Operto, 2009), which are often missing in seismic data. Recently source extension methods (Warner and Guasch 2014, Huang and Symes 2015) have been proposed to address the cycle skipping issue in FWI.

Claerbout (1985) summarized the accuracy and resolution for velocity analysis and seismic imaging and pointed out that the traditional velocity analysis had resolution of up to 2.5 Hz in apparent frequency and seismic imaging had a resolution of 10 to 100 Hz. Lambaré and Allemand (2015) added that ray-based traveltime tomography had increased the resolution of velocity inversion to about 6 Hz; while broadband data had expanded the resolution bandwidth of seismic imaging (Figure 1.1). Both FWI and seismic migration algorithms continue to improve the resolution for velocity inversion and seismic imaging. Traveltime tomography also continues to play an important role in providing an accurate initial velocity model for both FWI and seismic migration. However, with limitations caused by acquisition constraints, data quality and assumptions used in traveltime tomography methods, the velocity model determined by traveltime tomography methods can be sub-optimal and results in degradation in the reflection image. Furthermore, in order to build the most accurate velocity model from the surface to the target depth, it is important to understand the advantages and limitations of each velocity inversion method including refraction tomography, reflection tomography as well as FWI.

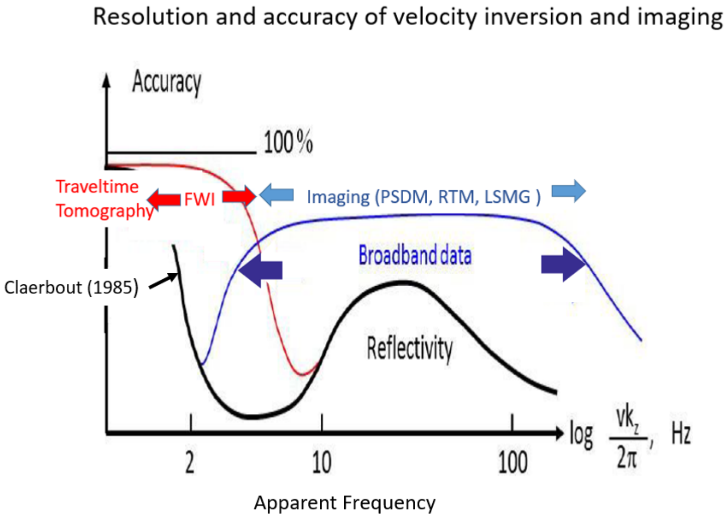


Figure 1. 1. Summary of resolution and accuracy of velocity inversion and imaging. (Lambare and Allemand, 2015)

1.2 Thesis proposal

The primary goal of this research is to first investigate and understand the effects of errors from traveltime tomography on the processed seismic data and then to develop strategies to quantify these errors and to incorporate them in the tomography inversion kernels to alleviate the inherent limitations in traveltime tomography methods. Second, I propose to expand the scope of traveltime tomography to include measurable information from the seismic waveform to improve the accuracy and efficiency of traveltime tomography.

1.3 Dissertation structure and overview

The main content of the dissertation is organized into 7 chapters. Chapter 2 reviews the tomography method. I compare medical tomography and seismic tomography. Medical tomography and seismic tomography share some similar algorithms and technologies. However, medical tomography has the benefit of small targets of investigation and not having to deal with the vast problem of acquisition limitations in seismic tomography, and results in much higher resolution images than that from seismic tomography.

Chapter 3 reviews the forward problem of tomography. In chapter 3, I review the wave equation for modelling seismic wave propagation and the seismic ray theory for traveltime computation. I also review the basic principles of different traveltime computation methods including the finite difference methods, ray shooting method, wavefront construction and paraxial method. I analyze their differences and similarities to investigate the effectiveness of these methods in traveltime tomography and seismic imaging.

Chapter 4 includes our paper “Robust refraction statics solution and near-surface velocity model building using feedback from reflection data” published in Geophysics Volume 83, no. 6, U63-U77. It reviews the inversion problem of refraction tomography and how to use feedback from the residual statics measurement of the deeper reflection data to improve the near-surface velocity model and the refraction statics solution. The long-wavelength components of the reflection residual statics measured from deeper reflection data do not suffer the same acquisition and data limitation of the refraction data. They are used to compute the model and data weights for the new refraction tomography kernel.

Chapter 5 includes our paper “Near-surface velocity model building and statics correction for blended land data” published in the Canadian Journal of Exploration Geophysics Volume 45, no. 1, spring 2021. It evaluates the problems of blended seismic data acquisition imposed on refraction and reflection arrival times. It proposes a robust refraction arrival separation method that uses amplitude burst suppression and the sparse Radon transform to enhance the first break quality of the blended data. It also demonstrates that after refraction statics correction, we can perform normal moveout velocity analysis and surface consistent residual statics prior to deblending because of the passive separation property of the blended data.

Chapter 6 reviews slope tomography, a reflection tomography method. In addition to the reflection traveltime, slope tomography also uses the slope of locally coherent events from the shot and geophone gathers to improve the results of reflection tomography. Each slope tomography pick will reconstruct a ray pair that connects a scatter point to a shot and receiver; therefore, slope tomography is also called stereo-tomography. I apply the slope tomography method to the Hussar 2D survey and confirm its accuracy with well-logs and depth migration.

Chapter 7 uses machine learning to addresses the major problem with refraction tomography: first breaking picking. First break picking is done using automated first break picking algorithms followed by laborious editing by trained technicians who are familiar with the near-surface geology and the first arrival energy waveform. I review two automatic trace-by-trace first break picking algorithms. I also apply an unsupervised clustering algorithm to reject mis-picks in first break time picks. Finally, I use a supervised UNET to train the network with first arrival energy images and the first break masks to automate the first break picking.

Chapter 8 contains the conclusions and discussions of future works.

Chapter 2

Tomography

2.1 Introduction

Tomography comes from the Greek word "tomos" meaning "section" or "slice" and "graphia" meaning "describing". Thus, tomography is a process that describes the material properties within the body of investigation. Seismic tomography methods share some similar physical and mathematical principles with medical tomography. Both methods seek to determine the interior distribution of values of physical properties (the integrant) from the projections (the integral or the sums of some interior value) measured outside of an object (Stewart 1991). In seismic tomography seismic energy propagated through the medium and are received at the receivers on the surface or in the borehole. An example of seismic tomography is traveltime tomography (Fig. 2.1b):

$$t_L = \int_L s(x)dl\,, \tag{2.1}$$

where the measured travel time t_L for the raypath, L is the integration of $s(x)dl$, and the objective of traveltime tomography is to determine the integrant $s(x)$, the inverse of the medium velocity (called slowness in geophysics).

In medical tomography, a CT scanner transmits a fan of X-ray beams through the target of the investigation and the detector elements record the attenuated radiation intensity (Fig. 2.1a). As X rays pass through the material of greater density, more of the rays are absorbed. Each radiation intensity reading represents the accumulated attenuation along the X-ray beam (Fig. 2.1a). Using Beer's law of absorption, the intensity of an attenuated X-ray beam that has travelled a distance d can be expressed as:

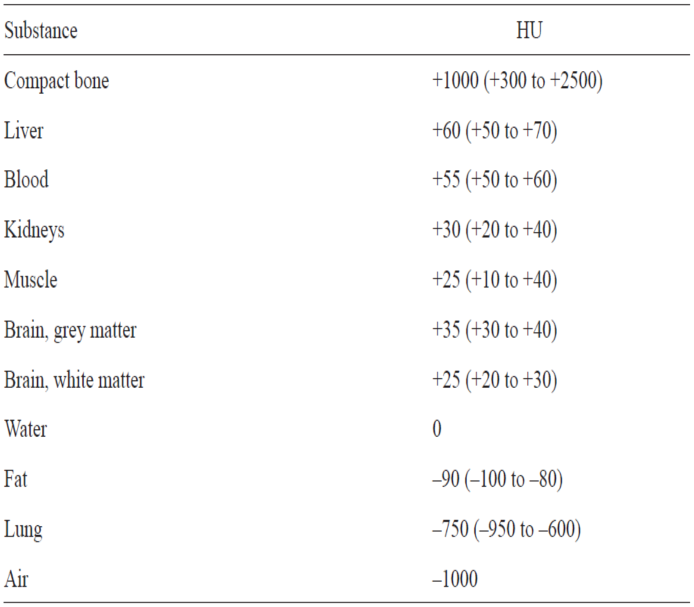
$$I(d) = I_0 e^{-\int_0^d \mu(x)dx} \quad \text{, or}$$

$$ln \frac{I_o}{I(d)} = \int_0^d \mu(x)dx \qquad (2.2)$$

where $\boldsymbol{I_0}$ is the initial beam intensity, $\boldsymbol{I(d)}$ is the attenuated intensity, $\boldsymbol{d}$ is the target thickness and $\boldsymbol{\mu(x)}$ is the attenuation coefficient at grid location $\boldsymbol{x}$. Equations (2.1) and (2.2) are similar. What the two equations measure is the result of integration. What they seek to determine is the integrant. For traveltime tomography, the integrant is the slowness of the medium. For CT scan, the integrant is the attenuation coefficients, which are later transformed to Hounsfield unit for subsequent interpretation:

$$HU_{material} = \frac{\mu_{material} - \mu_{water}}{\mu_{water}} * 1000. \qquad (2.3)$$

Table. 2.1 shows the typical HU values for different tissues and materials.



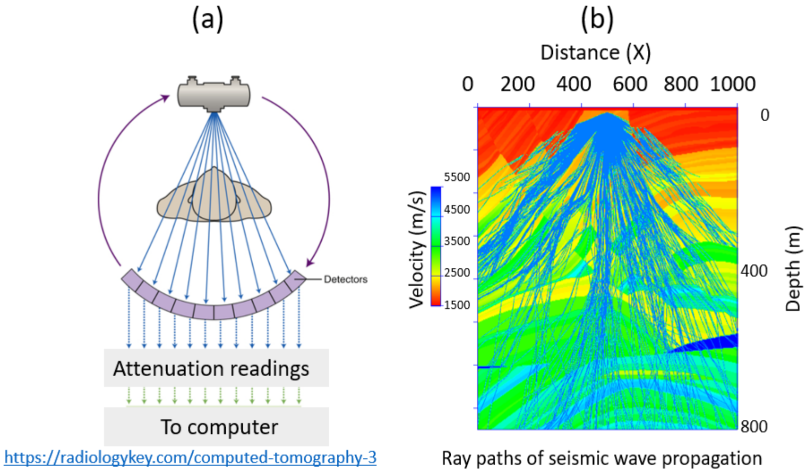
Substance	HU
Compact bone	+1000 (+300 to +2500)
Liver	+60 (+50 to +70)
Blood	+55 (+50 to +60)
Kidneys	+30 (+20 to +40)
Muscle	+25 (+10 to +40)
Brain, grey matter	+35 (+30 to +40)
Brain, white matter	+25 (+20 to +30)
Water	0
Fat	–90 (–100 to –80)
Lung	–750 (–950 to –600)
Air	–1000

Table 2.1. Hounsfield unit for some material and tissues.

.

2.2 Medical tomography

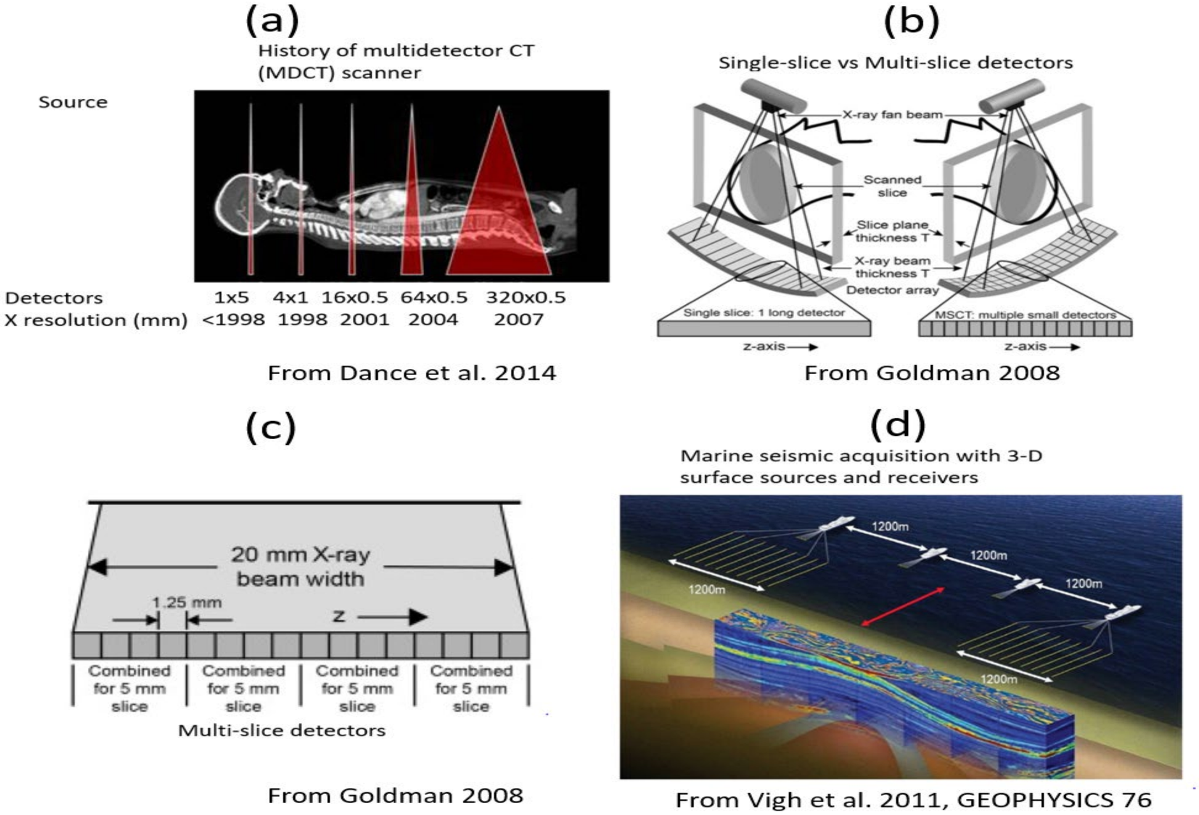
Computed Tomography (CT) is also known as Computed Axial Tomography (CAT). British engineer Godfrey Hounsfield invented the first dedicated head CT scanner based on x-ray computed tomography in 1967. The first CT scanner took up to several hours to acquire the data for a single slice and several days to reconstruct the image. Newer multidetector CT (MDCT) scanners have seen increases in both speed and resolution over the years (Fig. 2.2.a), and full-body scans can now be processed in real-time. Similarly, seismic acquisition systems also have evolved to include multiple sources and receivers (Fig. 2.2.b). CT differs from seismic tomography in scale, complexity and acquisition geometry. CT scan has a resolution of less than 1 mm and 360° angular coverage. With straight ray paths and continuous angular coverage, CT can reconstruct the image in the object space effectively by transforming the recorded data between the Radon, Fourier and object domains.



https://radiologykey.com/computed-tomography-3

Ray paths of seismic wave propagation

Figure 2. 1 (a) CT scanner sends a fan of X-ray beams from the source to the detectors placed on the other side of the target. (b) Ray paths of seismic wave propagation are more complex than the X-ray beams in CT scans because of large velocity variations.



From Dance et al. 2014

From Goldman 2008

From Goldman 2008

From Vigh et al. 2011, GEOPHYSICS 76

Figure 2. 2. (a) Development of the number of detectors and resolution of MDCT. (Adapted from Dance et al, 2014.) (b) Multi-slice detectors allow multiple CT slices to be recorded simultaneously. (c) A 16-slice scanner can be used for 16 1.25 mm slices or 4 5 mm slices. (d) The 3D seismic method employs multiple sources and receivers; however, seismic sources and receivers are placed on the surface and lack the 360° coverage of the CT scan method

2.2.1 Medical tomography reconstruction methods

Series expansion methods and transform methods are two groups of methods that can be used in medical tomography. Series expansion method developed by Kaczmarz in 1937 iteratively determines the model function. Series expansion methods include the algebraic reconstruction technique (ART) and the simultaneous iterative reconstruction technique (SIRT). Transform methods include the simple back-projection method and filtered back-projection method. With straight ray paths and 360° coverage, the CT method can utilize the transform methods effectively.

Most medical tomography reconstruction methods assume parallel-beam geometry; while all modern CT scanners use fan-beam projection. Therefore, it is necessary to rebin CT scan data from fan-beam geometry to parallel-beam geometry (Borsdorf et al. 2008).

Rebinning from fan-beam projection to parallel-beam projection

For a fan-beam described by a central angle of $\boldsymbol{\alpha}$ with the *x*-axis, and a fan-beam angle of $\boldsymbol{\beta_1}$ $(\textbf{Fig. 2.3})$, the equivalent parallel projection angle is:

$$\boldsymbol{\theta_1 = \alpha + \beta_1}, \tag{2.4}$$

and the orthogonal distance to the isocenter for a CT scanner with radius $\boldsymbol{R_S}$ is:

$$u_1 = R_s sin\beta_1. \tag{2.5}$$

We can use equation (2.4) and (2.5) to rebin the equiangular $P(\alpha, \beta)$ data to $P(u, \theta)$, and interpolate $P(u, \theta)$ to an equidistance and equiangular grid. Alternatively, we can use the following relationship to construct an equidistance $P(u_1, \theta_1)$ from the fan-beam data $P(\theta, \beta)$:

$$\beta_1 = \sin^{-1}\frac{u_1}{R_s} \tag{2.6}$$

Therefore, each fan-beam projection data can be rebinned to the parallel-beam Cartesian grid.

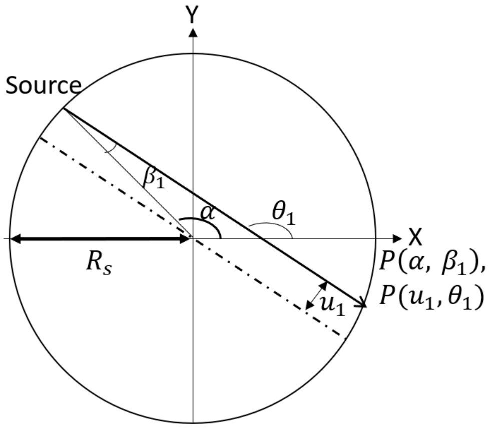
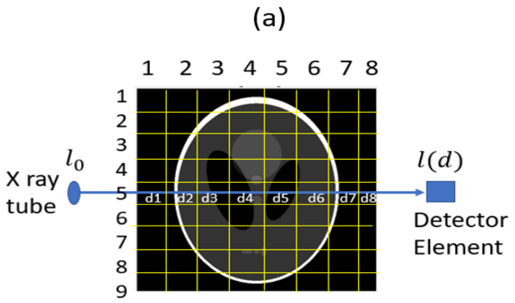


Figure 2. 3. Rebinning from fan-beam geometry data $\boldsymbol{P(\alpha, \beta)}$ to parallel-beam geometry data $\boldsymbol{P(u, \theta)}$.

Algebraic Reconstruction Technique (ART)

The first reconstruction method to review is ART. Figure 2.4b illustrates the concept of ART. ART uses the X-ray path and the attenuated intensity for each angle to build a set of simultaneous linear equations. The equations are then solved for the attenuation coefficients. Since ART uses the ray path characteristics to build the equations, it takes into account ray bending because of velocity gradient. However, there may be inconsistence in the equations, and the system of equations can be under-determined. The solution may also be sensitive to measurement errors and noise. Furthermore, with the increasing requirement of better resolution, the number of equations grows. Therefore, it is not feasible to use ART in clinical practice (Dance et al., 2014). Since CT data have continuous angular coverage, the back-projection method and transform techniques (Stewart 1991, Lo and Inderwiesen 1994) are better suited than ART.



(b)

$$\mu_{51}d1 + \mu_{52}d2 + \mu_{53}d3 + \mu_{54}d4 + \mu_{55}d5 + \cdots = \ln\left(\frac{l(d)}{l_0}\right)$$

Figure 2. 4 (a)The principle of attenuation of an X-ray beam. The path of the X-ray beam is discretized by the image grid. (b) Each ray path constitutes one equation. Equations from all the projection angles can be solved for the attenuation coefficients using Algebraic reconstruction.

Back-projection (without filter)

In the simple back-projection method, for each incident bean angle θ, the CT scanner projects the attenuated intensity of the X-ray beams to the detector elements (Fig. 2.5.a.b). The recorded projection, $P(u, \theta_i)$ is then back-projected to the object space using the incident angle θ_i (Fig. 2.5.c). Each back-projection is summed to reconstruct the image $u(x, z)$ of the object space (Fig. 2.5.d). Therefore, back-projection is really an average or sum of the recorded $P(u\,, \theta_i)$ mapped back to the object space at the incident angle of the X-ray beams. However, the reconstructed image is strongly blurred. The blurring of the image in this simple back-projection method is the result of the operation of the point spread function, 1/r (Stewart 1991).

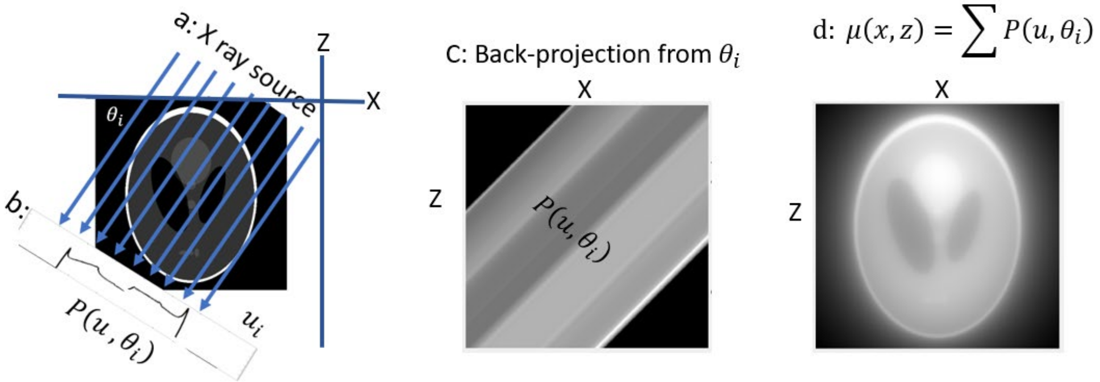


Figure 2. 5 (a) Incident X-ray beams propagate through the chest of a patient. (b) Attenuated radiation intensities recorded by the detector elements. (c) $\boldsymbol{P(u_i, \theta_i)}$ is back-projected to the object space using the incident angle. (d) Summing all the back-projections of different angles to reconstruct the object space.

Fourier-transform method

The Fourier transform method (Stewart 1991, Lo and Inderwiesen 1994) uses the projection slice theorem to reconstruct the image directly. Projection slice theorem states that 1-D Fourier transform of a projection of an object at an angle θ constitutes a slice of the 2-D Fourier transform of the object, where the slice makes the same angle with the k_x axis. Therefore, by performing 1-D Fourier transform on many projections of different angles, we can construct the 2-D Fourier transform of the object. We can then interpolate for empty grid point in the 2-D Fourier domain and reconstruct the object by 2-D inverse Fourier transform. Figure 2.6 illustrates the concept of the projection slice theorem for one angular projection. Figure 2.6a shows the detector elements recording the attenuated X rays, $P(u, \theta_i)$ in the Radon space.

The recorded projection is related to the line integral:

$$P(u, \theta_i) = \int_{ray} \mu(x, y) dv, \text{ or}$$

$$P(u, \theta_i) = \int_{-\infty}^{+\infty} \mu(x, y) dv. \tag{2.7}$$

where $\mu(x, y)$ is the unknown attenuation function, x and y are the axes of the object space, u is the projection axis and v is the axis parallel to the incident ray. The u-v coordinate system for the detector elements is related to the x-z coordinate system by:

$$\begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{2.8}$$

$P(u, \theta_i)$ can be mapped directly to the Radon space as a line at θ_i location (Figure 2.6b).

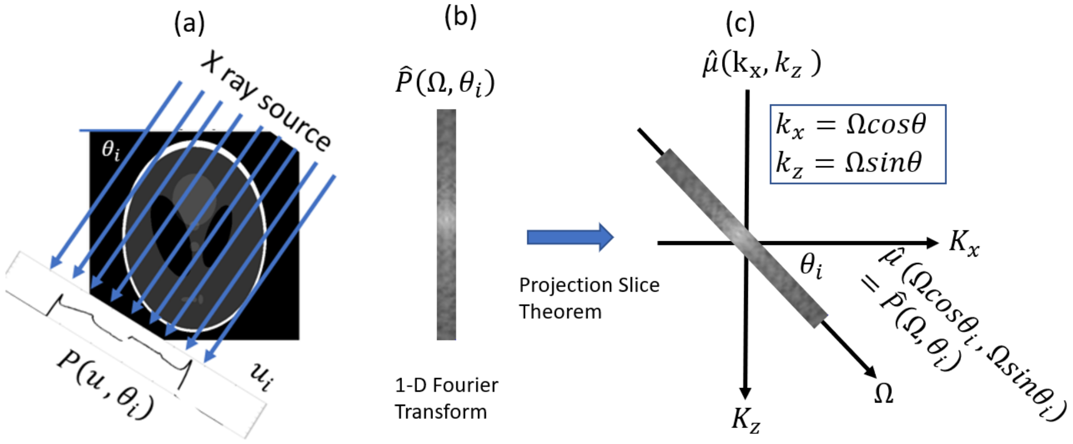


Figure 2. 6. An illustration of projection slice theorem. (a) Projection of an object to the Radon space. (b) 1-D Fourier transform of the projection slice (c) is equivalent to a slice in the 2-D Fourier transform of the object space at the same angle with the $\boldsymbol{k_x}$ **axis.**

Now, applying 1-D Fourier transform to $P(u, \theta)$ yields:

$$\hat{P}(\Omega, \theta) = \int_{-\infty}^{+\infty} P(u, \theta) e^{-i(\Omega u)} \, du. \tag{2.9}$$

Using the project slice theorem, we have:

$$\hat{\mu}(k_x, k_z) = \hat{P}(\Omega, \theta), \tag{2.10}$$

where $\hat{P}(\Omega, \theta)$ makes an angle θ to the k_x axis in the 2-D Fourier space $\hat{\mu}(k_x, k_z)$ (Fig. 2.6c). The two coordinate systems are related by:

$$k_x = \Omega cos\theta, \text{and } k_y = \Omega sin\theta. \tag{2.11}$$

Therefore, we can map each $\hat{P}(\Omega, \theta)$ to the (k_x, k_z) grid:

$$\hat{\mu}(\Omega cos\theta, \Omega sin\theta) = \hat{P}(\Omega, \theta). \tag{2.12}$$

By repeating this process for all the projection angles, we can populate the $\hat{\mu}(k_x, k_z)$ grid with $\hat{P}(\Omega, \theta)$. We can then compute the attenuation coefficients $\mu(x, z)$ by 2-D inverse Fourier transform of $\hat{\mu}(k_x, k_z)$. However, this process creates $\hat{\mu}(k_x, k_z)$ in the (k_x, k_z) that is denser near the zero wave numbers and sparser further away (Fig. 2.7). This can lead to artifacts in the inverse 2-D Fourier transform. This brings us to the filtered back-projection method.

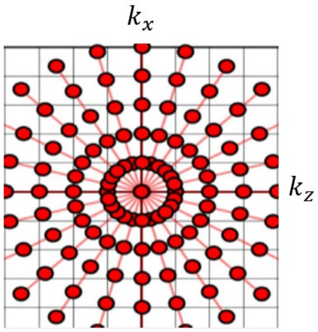


Figure 2. 7 Transforming $\boldsymbol{\hat{P}(\Omega, \theta)}$ to $\boldsymbol{\hat{\mu}(k_x, k_z)}$ results in denser population of data points near the zero wavenumber and sparser further away (from Dance et al, 2014)

Filtered back-projection method

The inverse 2-D Fourier transform of $\hat{\mu}(k_x, k_z)$ in the cartesian coordinate is:

$$\mu\,(k_x, k_z) = \frac{1}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \hat{\mu}(k_x, k_z)\; e^{i(k_x x + k_z z)} dk_x dk_z\ , \tag{2.13}$$

where $k_x = \Omega cos\theta, and\ k_z = \Omega sin\theta.$

Changing variables from $(k_z, k_z)\ to\ (\omega, \theta)$ give the integral in polar coordinates:

$$\mu\,(k_x, k_z) = \frac{1}{4\pi^2}\int_{0}^{\pi}\int_{-\infty}^{+\infty} \hat{\mu}(\Omega cos\theta, \Omega sin\theta)\; e^{i(k_x x + k_z z)} \Omega d\Omega d\theta\ . \tag{2.14}$$

Substitute 2.13 into 2.14 gives:

$$\mu\,(k_x, k_z) = \frac{1}{4\pi^2}\int_{0}^{\pi}\int_{-\infty}^{+\infty} \hat{P}(\Omega, \theta)\,|\Omega|\; e^{i(k_x x + k_z z)} d\Omega d\theta\ , \tag{2.15}$$

where $P'(u,\theta) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{P}(\Omega,\theta)\,|\Omega|\; e^{i(k_x x + k_z z)} d\Omega$ is the inverse transform of the 1-D FT of projection $P(u,\theta)$ filtered by $|\Omega|$, and $\frac{1}{2\pi}\int_{0}^{\pi} P'(\mu,\theta) d\theta$ is the inverse Radon transform. $|\Omega|$ is sometimes called the ramp filter or rho filter.

Fig. 2.8 illustrates these steps:

a,b,c: Acquire projection of attenuated X-ray intensity for projection angle θ_i.

d: Use equation (2.9) to transform each projection slice $P(u,\theta_i)$ to $\hat{P}(\Omega,\theta_i)$

e: Apply $|\Omega|$ to high pass filter the 1-D FT projection slice, and repeat for projection slices of all incident angles.

f: Inverse 1-D Fourier transform of all high pass filtered 1-D FT projection slices to form the high pass filtered Radon space.

g: Inverse Radon transform to reconstruct the image space.

2.3 Seismic tomography

Lo and Inderwiesen (1994) categorizes seismic tomography into seismic ray tomography and seismic diffraction tomography (Devaney 1982, Wu and Toksoz 1987) accordingly to the forward modelling method used. When the scale of the medium inhomogeneities is much larger than the seismic wavelength, seismic ray tomography uses ray theory to model the wave propagation as rays. When the scale of the medium inhomogeneities is comparable with the seismic wavelength, diffraction tomography uses wave scattering theory to model the wave propagation. Another situation that seismic diffraction tomography is used is when the scale of the medium inhomogeneities is larger than the seismic wavelength, but the velocity contrast is small. Seismic ray tomography includes traveltime tomography. Seismic diffraction tomography includes the method known as full waveform inversion (FWI). In this thesis, I will focus on traveltime tomography. Seismic ray tomography will be discussed in detail in the subsequent chapters.

Summary

With the advantage of 360° transmission coverage, medical tomography can use transform methods effectively to invert for the attenuation coefficients. To alleviate the blurring problem caused by the effect of the power spreading function, clinic medical tomography uses the filtered back-projection (FBP) method to apply a high pass filter to the recorded CT scan data. Figure 2.9 compares reconstructed images from 3, 6, 18, 45, 60, 90, 180 and 1800 angles.

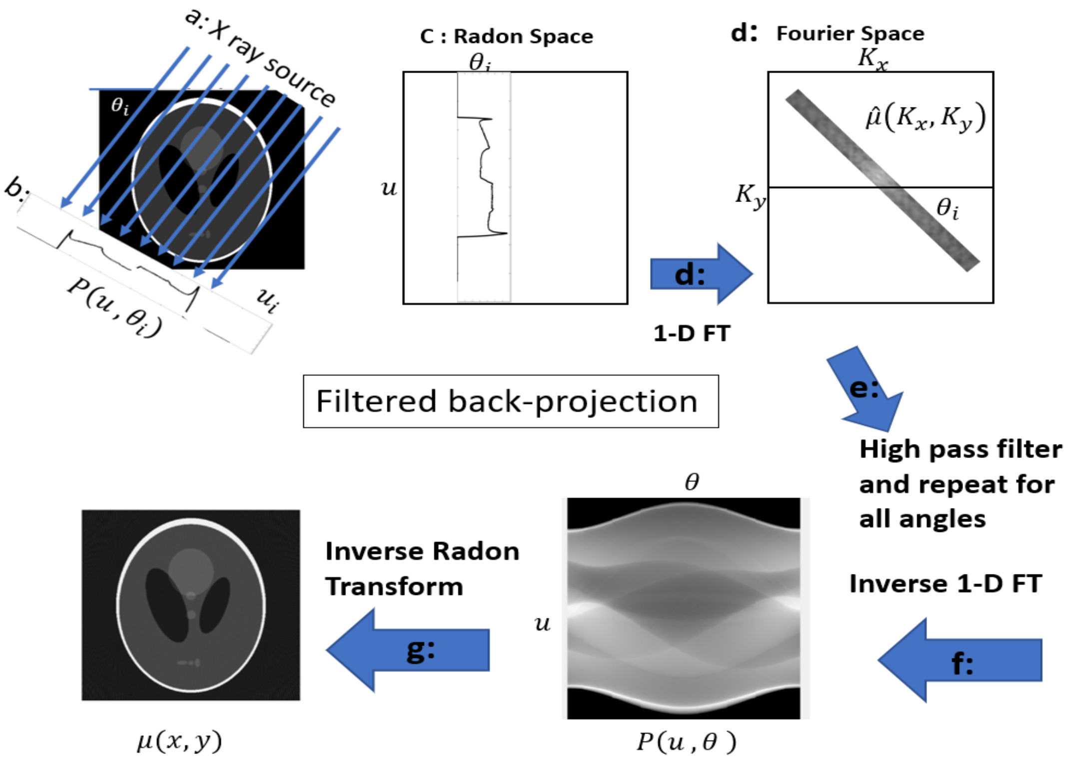


Figure 2. 8. (a) X-ray is transmitted through the object space at an incident angle $\boldsymbol{\theta}$, and (b) recorded by the detectors. (c) The projection corresponds with one line on the Radon space. (d) 1-D Fourier transform results in an angular line in Fourier space. (e) High pass filter and repeat (a) to (e) for angles. (f) 2-D inverse Fourier transform to Radon space. (g) Inverse Radon transforms back to object space.

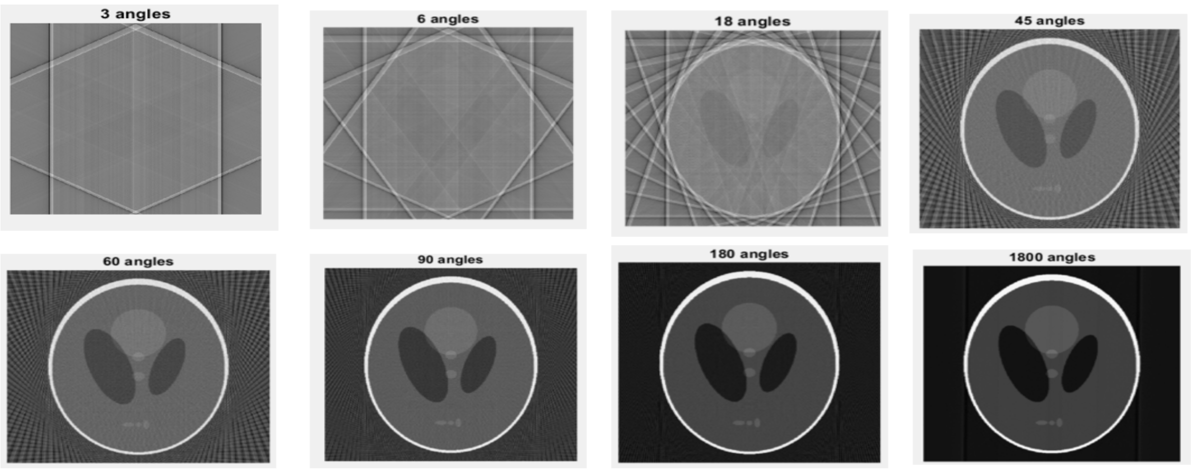


Figure 2. 9 Results of filtered back-projections with 3, 6, 18, 45, 60, 90, 180 and 1800 angles.

Chapter 3

Forward problem

Inversion problem begins with the generation of modelled data $\mathbf{d}$, involving a forward modelling operator **L** acting on the parameter **m** (Claerbout 1992):

$$d = L\,m. \qquad (2.1)$$

Velocity inversion methods estimate the optimal velocity model by minimizing the difference between the modelled and observed wavefield attributes, including time delays, amplitude, and phase. Full waveform inversion methods minimize the differences in amplitude and phase, while traveltime tomography methods minimize the differences in time delays. In this chapter, I review the wave equation for modelling seismic wave propagation and the seismic ray theory for traveltime computation. I also review the basic principles of different traveltime computation methods including the finite difference methods, ray shooting method, wavefront construction and paraxial method. I analyze their differences and similarities to investigate the effectiveness of these methods in traveltime tomography and seismic imaging. I compare the travel times from these methods to a finite-difference synthetic shot record of the Marmousi model and find travel times from all three methods are accurate except at the area where rays diverge. I also used the travel time from the fast-marching method in the refraction tomography processing of the Hussar 2D dataset (Margrave et al., 2012). The CDP stack from the refraction tomography processing is more coherent and better resolved than the CDP stack with datum static correction only.

3.1 Introduction

Full waveform inversion (FWI) methods model the seismic wave propagation by solving the wave equation. FWI then updates the velocity model by minimization of the differences between the modelled and observed seismic wavefield. Traveltime tomography estimates the optimal velocity model by minimizing the differences between the modelled and observed traveltimes. Core principles of most traveltime and ray tracing algorithms are derived from the seismic ray theory. High-frequency approximation of the solution of the elastodynamic equation leads to solutions in different forms. For kinematic ray tracing, the solution leads to the eikonal equation and the ray equations. The high-frequency approximation requires the velocity of the media to varying smoothly. Vidale (1988,1990) presented a grid-based traveltime computation scheme that solves the eikonal equation by finite difference approximation. Vidale's work leads to subsequent studies and developments by Qin (1992), Sethian and Popovici (1999) and other authors, and resulted in more robust algorithms that can better handle rapid velocity variations. The results of these algorithms are traveltime from source to regularly spaced grid points. Vidale (1988) proposed the construction of the ray paths by tracing the steepest traveltime gradient from the receiver back to the source. Matsuoka and Ezaka (1992) presented a ray path reciprocity method that traces the minimum time of summed shot and receiver traveltime tables. Alternate to grid-based traveltime computation schemes are kinematic and dynamic ray tracing (Červený and Hron,1980; Beydoun and Keho,1987), and wavefront construction method (Vinje et al.,1993). These methods involve tracing the ray path by computing the solutions to the ray equations at each ray step. Some geophysical applications such as Kirchhoff migration only require traveltime from a source or receiver to a subsurface point, while other applications such as refraction tomography require both first arrival time and ray path between a source and receiver. The purpose of this chapter is to review the basic principles of the fast-marching method, paraxial method and wavefront construction method, and to evaluate their accuracy and effectiveness when applied in refraction tomography and depth imaging.

3.1 Wave equation

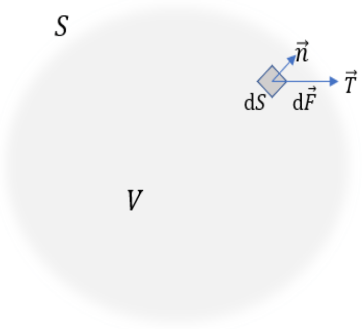


Figure 3. 10. Traction $\vec{\boldsymbol{T}}$ due to the element of force $\boldsymbol{d\vec{F}}$ working on element surface dS. $\vec{\boldsymbol{n}}$ is the unit vector normal to the $\boldsymbol{dS}$.

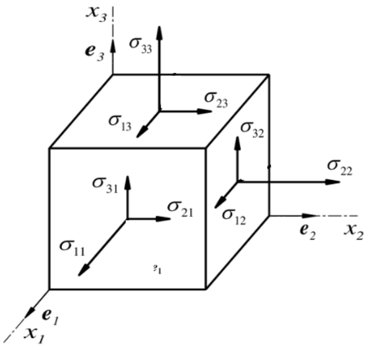


Figure 3. 2. Stress $\boldsymbol{\sigma_{ij}}$ and normal vector $\boldsymbol{e_j}$

Traction $\vec{\boldsymbol{T}}$ (Figure 3.1) is a contact force exerted by the material on the positive side of surface element $\boldsymbol{dS}$ due to the element of force $\boldsymbol{dF}$:

$$\vec{\boldsymbol{T}} = \boldsymbol{dF/ds}\text{ , or} \qquad (3.1)$$

In terms of stress $\boldsymbol{\sigma_{ij}}$:

$$\boldsymbol{T_i} = \sum_{j=1}^{3} \boldsymbol{\sigma_{ij} n_j} = \boldsymbol{\sigma_{ij} n_j} \qquad (3.2)$$

i refers to the direction of traction components, j refers to the direction of the normal of the face on which the traction is acting. On the right, I use Einstein's notation (repeated indexes imply summation). Applying Newton's second law of motion, we can write:

$$\int_s T dS + \int_V f dV = \int_V \rho \ddot{u} dV \ , \qquad (3.3)$$

where the first term represents surface forces acting on S, and the second term represents the body forces adding on the volume, u is the displacement. Applying equation (3.2) to (3.3) gives the equation for direction i:

$$\int_s \sum_{j=1}^{3} \boldsymbol{\sigma_{ij}}\, \boldsymbol{n_j} + \int_V f_i dV = \int_V \rho \ddot{u}_i dV \qquad (3.4)$$

Using divergence theorem, $\int_V \sum_{j=1}^{3} \frac{\partial A_j}{\partial x_j} dV \; = \int_S \sum_{j=1}^{3} A_j \, n_j \; dS$, gives:

$$\int_V \sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j} dV + \int_V f_i dV = \int_V \rho \ddot{u}_\iota dV, \text{ or} \tag{3.5}$$

$$\sigma_{ij},_j + \; f_i = \rho \ddot{u}_\iota \tag{3.6}$$

Applying Hooke's law,

$$\sigma_{ij} = c_{ijkl} \epsilon_{kl}, \tag{3.7}$$

where c_{ijkl} is the stress tensor and ϵ_{kl} is the strain tensor,

$$\epsilon_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right), \tag{3.8}$$

wee obtain the stress-strain relationship:

$$\left(c_{ijkl} u_{k,l}\right),_j + f_i = \rho \ddot{u}_i. \tag{3.9}$$

For a wavefield propagation in general (for example, far away from any source):

$$\left(c_{ijkl} u_{k,l}\right),_j \; = \rho \ddot{u}_i. \tag{3.10}$$

Equation (3.10) is often referred to as the elastodynamic equations. For an inhomogeneous, isotropic medium described by Lam**é** parameters $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and density $\boldsymbol{\rho}$, the stiffness tensor, $\boldsymbol{c_{ijkl}}$, can be written as:

$$\boldsymbol{c_{ijkl}} \; = \boldsymbol{\lambda \delta_{ij} \delta_{kl}} + \boldsymbol{\mu}\left(\boldsymbol{\delta_{ik} \delta_{jl}} + \boldsymbol{\delta_{il} \delta_{jk}}\right). \tag{3.11}$$

Substitute in $\boldsymbol{c_{ijkl}}$, ignore $\boldsymbol{f_i}$ for force from afar, and take the derivative of $\boldsymbol{u}$ gives:

$$(\lambda + \mu)\nabla \cdot (\nabla \cdot u) + \mu \nabla^2 u \; = \rho \ddot{u} \tag{3.12}$$

With $\nabla^2 u = \nabla \cdot (\nabla \cdot u) - \nabla \times \nabla \times u$:

$$(\lambda + 2\mu)\nabla \cdot (\nabla \cdot u) - \mu \nabla \times (\nabla \times u) = \rho \ddot{u}. \tag{3.13}$$

Apply divergence $(\nabla \cdot)$, and substitute in $V_p = \frac{\lambda + 2\mu}{\rho}$ give the scalar wave equation for P-wave:

$$\nabla^2 u = \frac{1}{V_p^2} \ddot{u} \tag{3.14}$$

This is a particular solution for the case of the scalar wave equation. In general, the displacements will be vectors and there will be solutions for acoustic and shear waves.

3.2 Seismic ray method and eikonal equation

The seismic ray method is based on an asymptotic high-frequency solution of the elastodynamic equation. A ray represents the movement of energy in the wavefield and is a curve perpendicular to the wavefronts as defined in the next section. In the previous section, we describe the acoustic wave equation can be written as:

$$\left(\boldsymbol{\nabla}^{2}-\frac{1}{c^{2}}\frac{d^{2}}{dt^{2}}\right)\boldsymbol{u}=-\boldsymbol{f}, \tag{3.15}$$

where $\boldsymbol{c}=\sqrt{\frac{k}{\rho}}$, $\boldsymbol{k}=\boldsymbol{bulk\ modulus};\ \boldsymbol{\rho}=\boldsymbol{density}.$

In the frequency domain:

$$\left(\boldsymbol{\nabla}^{2}-\frac{1}{c^{2}}\boldsymbol{\omega}^{2}\right)\boldsymbol{u}=-\boldsymbol{f}, \tag{3.16}$$

Going back to the general description in equations (3.5) and (3.6, we can substitute $\boldsymbol{u_i(\vec{x})}\ \mathbf{exp}\{-\boldsymbol{i\omega}[\boldsymbol{t}-\boldsymbol{T}(\vec{\boldsymbol{x}})]\}$ in those equations, giving the Kelvin-Christoffel equation:

$$(\boldsymbol{c_{ijkl,j}\,A_{k,j}+c_{ijkl}A_{k,lj}})\boldsymbol{\omega}^{-2}-\boldsymbol{c_{ijkl}T_{,j}\,T_{,l}\,A_k}=-\boldsymbol{\rho\delta_{ik}A_k} \tag{3.17}$$

For high-frequency approximation, we drop the $\boldsymbol{\omega}^{-2}$ term:

Define: $\boldsymbol{B_{ik}}=\boldsymbol{T_{,j}\,T_{,l}}$ gives:

$$(\ \boldsymbol{B_{ik}}-\boldsymbol{\rho\delta_{ik}})\boldsymbol{A_k}=\mathbf{0},\text{ or in matrix form }[\boldsymbol{B}-\boldsymbol{\rho I}]\boldsymbol{A}=\mathbf{0}. \tag{3.18}$$

The solution to the eigenvalue problem defined by equation (3.18) can be obtained by:

$$\text{Det }[\boldsymbol{B}-\boldsymbol{\rho I}]=\mathbf{0}. \tag{3.19}$$

Expanding equation (3.19) yields: $\left(\boldsymbol{T_{,k}T_{,k}}-\frac{1}{\alpha^{2}}\right)\left(\boldsymbol{T_{,k}T_{,k}}-\frac{1}{\beta^{2}}\right)=\mathbf{0}.$ (3.20)

Equation (3.20) is often expressed as $\left(\vec{\boldsymbol{\nabla}}\,\boldsymbol{T}\right)^{2}=\frac{1}{c^{2}}.$ (3.21)

Equation (3.21) is known as the eikonal equation and its solution provides the traveltime $\boldsymbol{T(x, y, z)}$ for the subsurface. There is a fully elastic form for anisotropic media which I don't present here (see Cervený, 2001).

3.3 Ray equations

To trace the position of a ray, we have to define some properties of rays and wavefronts (Figure 3.3) and express their relationships as a set of ray equations. Wavefronts are defined by the surfaces T(x,y,z)=constant. Slowness vector $\vec{\boldsymbol{q}}$ equals $\vec{\boldsymbol{\nabla}}\boldsymbol{T}$ and is tangential to the ray and normal to the wavefronts.

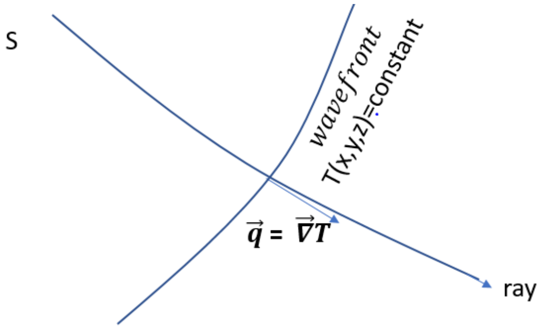


Figure 3. 3. Relationship between ray and wavefront

From equation (3.21), we have $\left|c\,\vec{\nabla}T\right| = 1$ being a unit vector normal to the wavefront. Therefore, ray can be defined by $\frac{d\vec{x}}{ds} = c\,\vec{\nabla}T = c\vec{q}$, or $\frac{dx_i}{ds} = cq_i$, i=1,2,3 (3.22)

where $\frac{d\vec{\boldsymbol{x}}}{d\boldsymbol{s}}$ is a unit vector tangential to the ray.

From $\vec{q} = \vec{\nabla}T$, we can obtain $\frac{d\vec{q}}{ds} = \frac{d}{ds}\vec{\nabla}T$ (3.23)

Substitute (3.22) into (3.23), we obtain the ray equations for the slowness vectors:

$$\frac{d\vec{q}}{ds} = \frac{d}{ds}\left[\frac{1}{c}\frac{d\vec{x}}{ds}\right] = \vec{\nabla}\left[\frac{1}{c}\right] \tag{3.24}$$

Ray equations can also be expressed in terms of T instead of arc length, s:

$$\frac{d\vec{x}}{dT} = c^2\vec{q}, \tag{3.25}$$

and

$$\frac{d\vec{q}}{dT} = c\vec{\nabla}\left[\frac{1}{c}\right] \tag{3.26}$$

These equations form the kinematic ray tracing system. Solution of equation (3.33) or (3.25) represents the trajectory $\vec{x}$, while the solution of equation (3.24) or (3.26) represents the slowness vector $\vec{q}$ along the ray as function of arc length or time.

3.4 Finite difference solution to the eikonal equation and grid-based method

Grid-based travel time computation algorithms use the eikonal equation (3.21) to solve for T(x,y,z). Vidale (1988) presented a method that uses a first-order finite difference approximation scheme to propagate geometric rays from three corners to the fourth corner of a square grid as shown in figure 3.4. Equation (3.12+16) and (3.12+17) are the average finite difference approximation of $\frac{\partial T}{\partial x}$ and $\frac{\partial T}{\partial z}$ respectively.

$$\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2 = s(x,z)^2 \tag{3.27}$$

$$\frac{\partial T}{\partial x} = \frac{1}{2h}\left(t_0 + t_2 - t_1 - t_3\right) \tag{3.28}$$

$$\frac{\partial T}{\partial z} = \frac{1}{2h}\left(t_0 + t_1 - t_2 - t_3\right) \tag{3.29}$$

Substitute equations (3.28) and (3.29) into equation (3.27):

$$t_3 = t_0 + \sqrt{2(h\bar{s})^2 - (t_2 - t_1)^2} \tag{3.30}$$

Where: s(x,z) is the slowness,

t_0, t_1 **and** t_2 **are computed travel times,**

t_3 **is travel time to be computed.**

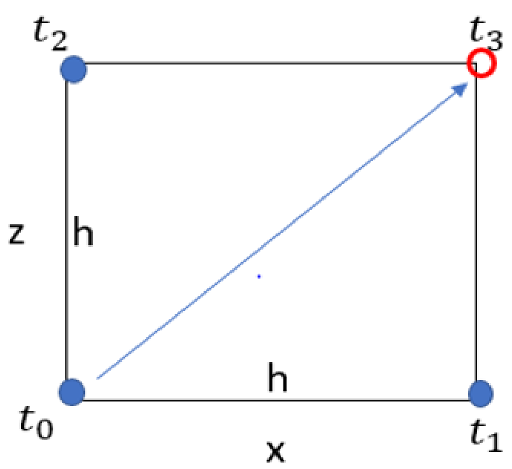


Figure 3. 4. Using t_0, t_1 **and** t_2 to compute t_3.

The procedure starts at the source and expands outward as square rings (Figure 3.5). Points on the square ring are sorted from minimum traveltime to maximum traveltime, and the new traveltime is computed starting from the point with minimum travel time.

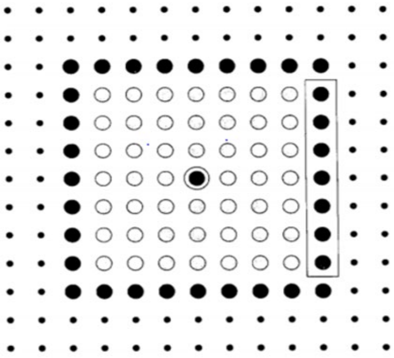


Figure 3. 5. The double circle shows the source point. Empty circles are timed locations. Filled circles are locations to be timed. Large filled circles are the square wavefront to be timed. Points on each edge are timed from location of minimum time to maximum time. (Adapted from Vidale 1988)

2.4.1 Expanding wavefront method

Qin (1992) showed that the expanding square strategy is not appropriate for a model with moderate to large velocity contrast and can lead to a negative value in the square root term in equation (18). Qin proposed an expanding wavefront method that can preserve the causality by expanding the wavefront only at points adjacent to the point with global minimum traveltime (shown as a double circle in Figure 3.6). This method ensures a ray associated with a point to be considered is completely timed up to that point. However, it is computationally expensive at $\mathbf{O(N^3)}$ algebraic operations, because sorting is required to establish the new global minimum after each wavefront point is added.

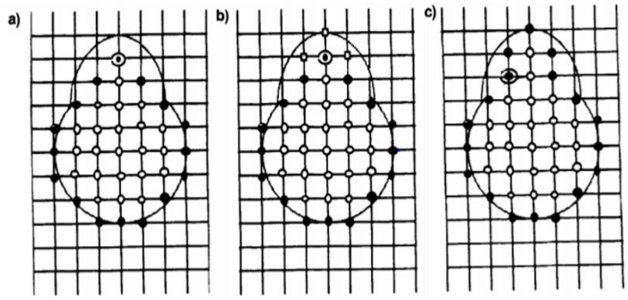


Figure 3. 6. (a) Filled circles mark the outer circumference of timed locations. The double circle shows the location of minimum time on the current time wavefront. (b) new locations to be timed (empty circles next to double circle). (c) New locations in (b) are timed and the new minimum time of the current wavefront is marked as a double circle. (Adapted from Qin 1992)

3.4.2 Fast marching method

Sethian and Popovici (1999) showed that propagating a triangular wavefront with unit speed using central difference approximation to the travel time gradient results in instabilities at the bend of the triangular wavefront. Rapid changes in velocity can result in similar instabilities. These instabilities are resolved by applying entropy-satisfying upwind differences schemes introduced by Osher and Sethian (1988):

$$\boldsymbol{\Psi_x^2 \approx [\max(D_i^{-x}\,\Psi, 0)^2 + \min(D_i^{+x}\,\Psi, 0)^2\,]} \tag{3.31}$$

A more convenient upwind scheme from Rouy and Tourin (1992) is used in Sethian and Popovici's fast marching method:

$$\boldsymbol{\Psi_x^2 \approx \max(D_i^{-x}\,\Psi, -D_i^{+x}\,\Psi, 0)^2} \tag{3.32}$$

where $\mathbf{D^- \text{ and } D^+ \text{are backward and forward difference operators}:}$

$$D_i^{-x}\Psi = \left(\frac{\Psi_i - \Psi_{i-1}}{h}\right)$$

$$D_i^{+x}\Psi = \left(\frac{\Psi_{i+1} - \Psi_i}{h}\right)$$

Ψ_i is the value of Ψ at grid point i and h is the grid spacing

The upwind scheme chooses grid points in terms of the direction of the flow of information. Sethian and Popovici (1999) express the eikonal equation as the following and apply the upwind finite difference scheme:

$$|\nabla t(x, y, z)| = s(x, y, z) \tag{3.33}$$

$$|\nabla t| \approx \left[\max\left(D_{ijk}^{-x} t, -D_{ijk}^{+x} t, 0\right)^2 + \max\left(D_{ijk}^{-y} t, -D_{ijk}^{+y} t, 0\right)^2 + \max\left(D_{ijk}^{-z} t, -D_{ijk}^{+z} t, 0\right)^2\right]^{\frac{1}{2}} = S_{ijk} \tag{3.34}$$

where S_{ijk} **is the slowness at grid point(i, j, k).**

To solve for t_{ijk}, we expand equation (3.34) to a quadratic equation in the form of

$$at^2 + bt + c = 0$$

t_{ijk} can now be solved explicitly as the root to a quadratic equation using $t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

The fast march algorithm also stores the traveltime values on a heap with the minimum time on top of the heap to reduce the sorting effort. This reduces the computationally cost to $\mathbf{O(NlogN\ )}$ algebraic operations (Sethian and Popovici, 1999). The fast marching algorithm is outlined in the following steps:

First, compute traveltimes at locations around the source point and tag these locations as ACCEPTED. Then tag as CLOSE all points one grid point away. Finally, tag as FAR all other grid points.

1) Begin Loop: Let TRIAL be the point in CLOSE with the smallest traveltime
2) Add the point TRIAL to ACCEPTED; remove it from CLOSE.
3) Tag as CLOSE all neighbours of TRIAL that are not ACCEPTED. If the neighbour is in FAR remove it from that list and add it to the set CLOSE.
4) Recompute traveltimes at all neighbours according to equation (3.34).
5) Return to 1.

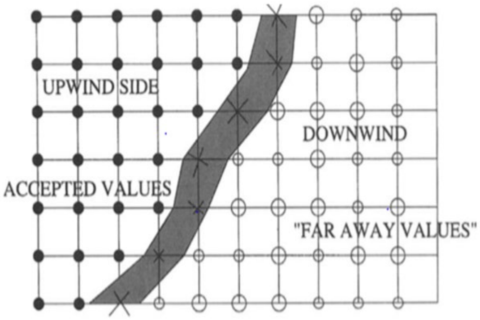


Figure 3. 7. Fast marching scheme. Filled circles are timed locations. X's are CLOSE locations to be tested for the minimum time. Empty circles are FAR locations have not been times.

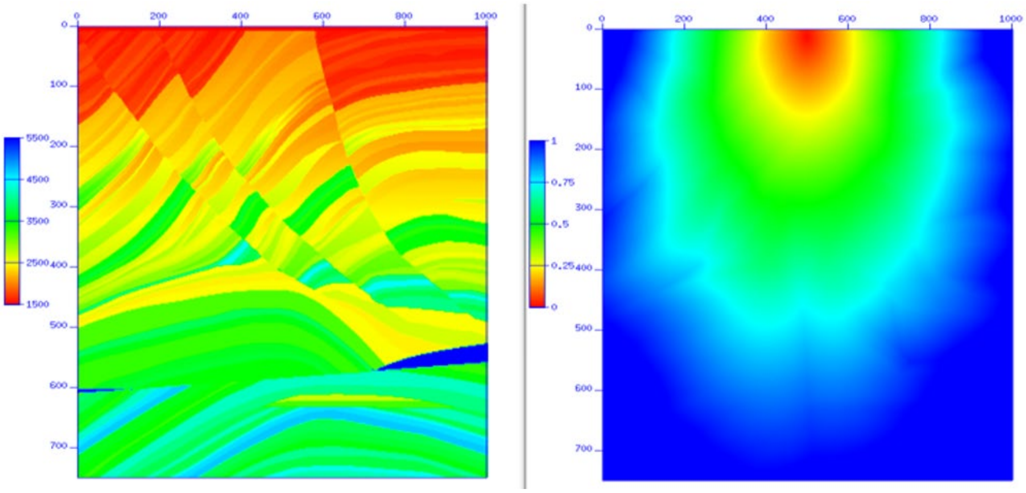


Figure 3. 8. Input velocity model and minimum traveltime from fast marching method (Sethian and Popovici, 1999).

3.5 Ray Shooting method

The ray shooting method (Figure 3.9) shoots a series of rays through the medium with starting vertical angle $\boldsymbol{\theta_i}$ and horizontal angle $\boldsymbol{\phi_i}$, and uses the ray equations 3.35 and 3.36 to compute the trajectory of the ray paths. These ray equations contain the definition and conservation of slowness (this conservation is equivalent to Snells' law). Travel times along the ray paths are then computed by integrating through the velocity model. Finally, the computed travel times are mapped to a regular grid map of the subsurface by interpolation across rays.

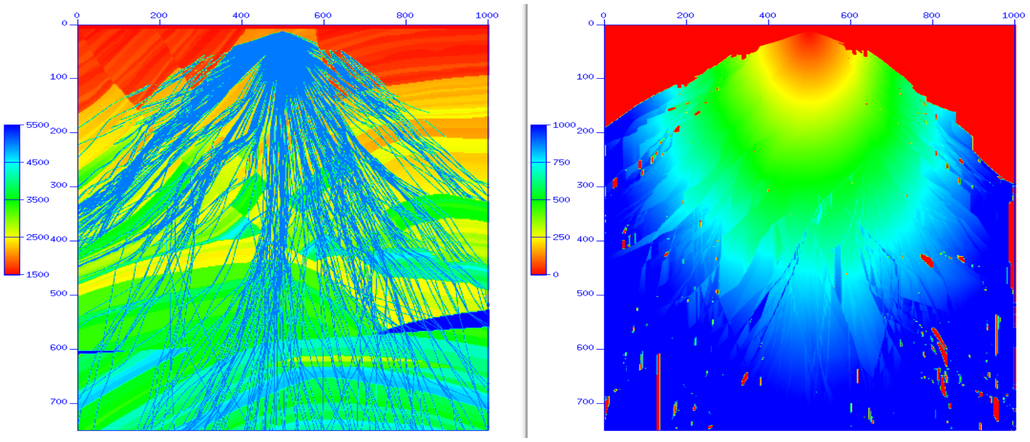


Figure 3.9. a) Input velocity model and rays, b) interpolated traveltime.

The initial value of equations (3.22) for an isotropic medium is:

$$\frac{d\vec{x}}{ds} = (sin\theta_i\ cos_i\phi, sin\theta_i\ sin\phi_i, cos\theta_i) \tag{3.35}$$

The initial value for the ray parameter for an isotropic medium is:

$$\vec{q} = \frac{1}{c(X_s, Y_s, Z_s)} \frac{d\vec{x}}{ds} \tag{3.36}$$

Trajectory $\vec{x}$ of the ray is computed with the following steps:

1: Solve ODE (3.22) for $\vec{x}$

2: Solve ODE (3.24) for $\vec{q}$.

Repeat steps 1 and 2 for all depth steps.

Repeat for all starting angles $\theta_i\ and\ \phi_i$

3.6 Wavefront Construction

Wavefront construction (WFC) (Vinje et al., 1993) is a natural extension of the ray shooting method. WFC uses localized ray tracing to construct a wavefront of constant traveltimes. The amplitude of rays can be computed from the ratio of the cross-sectional area of rays of adjacent wavefronts. The initial wavefront is constructed by shooting a series of short ray segments of equal time steps from the source. The end points of the ray segments on the wavefront are then propagated for another time step to construct a new wavefront. Coordinates of position and components of slowness vector of the ray segments are computed using the same procedure as the ray shooting method. When the wavefront crosses an interface with rapid velocity changes, the ray segments diverge and create a gap or shadow zone, that is subsurface points without illumination. The ray segments can also cross over and create caustics or multi-values (Figure 3.10a). Caustics in seismic refer to concentration of energy along a path that produces infinite values of amplitude. Multivalues are subsurface points where rays cross each other producing more than one possible value of the wavefield. To address the problem of shadow zones and to ensure sufficient ray density, additional ray segments can be interpolated (Figure 3.10b). For minimum traveltime ray tracing, caustics can be removed (Figure 3.10c). Figure 3.10d shows gridded minimum traveltime after caustics are removed.

In this example, upgoing rays are disabled for depth imaging. However, if caustics are to be removed, upgoing rays can be enabled for refraction ray tracing.

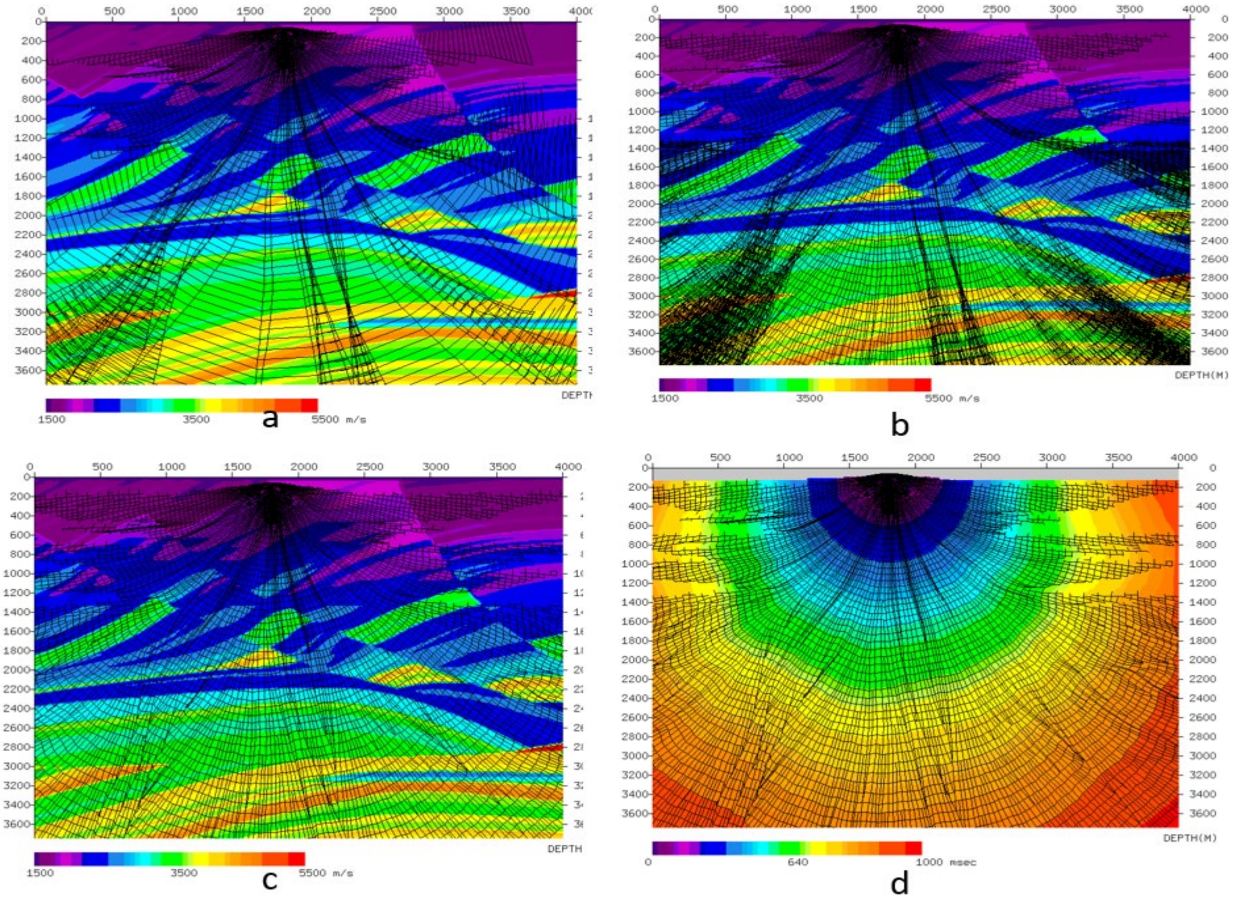


Figure 3. 10. a) Rays and wavefronts without interpolation, b) Rays and wavefronts with third-order interpolation along wavefronts, c) Rays and wavefronts with caustics removed, d) travel time gridded from wavefronts (Vinje et al., 1993).

3.7 Paraxial method

The paraxial method (Beydoun and Keho1987) is a dynamic ray-tracing method in ray coordinate system (γ_1, γ_2, u) or ray-centred coordinate system (q_1, q_2, q_3). The following discussion refers to the ray coordinate system shown in figure 3.11a. Paraxial rays are rays in the vicinity of a central ray (Figure 3.11b). γ_1 *and* γ_2 are ray parameters. They can be take-off angles i_0 *and* ϕ_0, or they can be components of slowness vector. They specify the initial direction of the ray in isotropic media. For anisotropic media, they specify the initial direction of the slowness vector. The third ray parameter u is a monotonically changing parameter along the ray. It can be arc length s or travel time T.

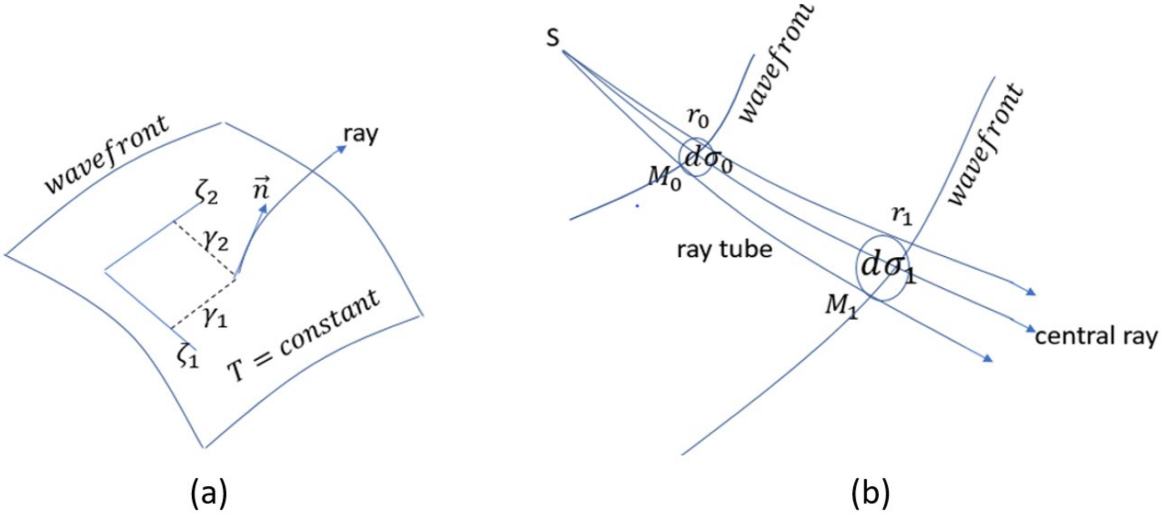


Figure 3.11. a) ray coordinates $\boldsymbol{\zeta_1}, \boldsymbol{\zeta_2}$; ray parameters $\boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}$ and wavefront $\boldsymbol{T}$. b) paraxial rays are rays in the vicinity of the central ray. $\boldsymbol{d\sigma_0}$ is the cross sectional area of the paraxial ray, $\boldsymbol{r_0}$ is radius of curvature of the wavefront at $\boldsymbol{M_0}$. Similarly, for $\boldsymbol{d\sigma_1}$ ***and*** $\boldsymbol{r_1}$.

Paraxial rays can have different properties than the central ray. These properties can be travel times or amplitude. The paraxial method determines these properties by differentiating the ray equations with respect to γ_1 *and* γ_2. If we choose u=s, we can start with ray equations in the form of:

$$\frac{dx_i}{ds} = c\,p_i \tag{3.38}$$

$$\frac{dp_i}{ds} = \frac{d}{dx_i}\left[\frac{1}{c}\right] \tag{3.39}$$

3.7.1 Dynamic ray tracing equations

We define:

$$Q_i \equiv \frac{\partial x_i}{\partial \gamma}, P_i \equiv \frac{\partial p_i}{\partial \gamma} \tag{3.40}$$

To derive the dynamic ray tracing equations, we take the derivatives of (3.38) and (3.39):

$$\frac{\partial}{\partial y}\frac{dx_i}{ds} = \frac{d}{ds}\frac{\partial x_i}{\partial y} = \frac{dQ_i}{ds} = \frac{\partial c}{\partial y}p_i + cP_i$$

$$\frac{dQ_i}{ds} = \frac{\partial c}{\partial x_k}\frac{\partial x_k}{\partial y}p_i + cP_i = c_{,k}\,Q_k p_i + cP_i \tag{3.41}$$

$$\frac{\partial}{\partial y}\frac{dp_i}{ds} = \frac{d}{ds}\frac{\partial p}{\partial y} = \frac{dP_i}{ds} = \frac{\partial}{\partial y}\left(\frac{d}{dx_i}\left[\frac{1}{c}\right]\right)$$

$$\frac{dP_i}{ds} = \left(\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_i}\left(\frac{1}{c}\right)\right)\frac{\partial x_k}{\partial \gamma} = \frac{\partial^2}{\partial x_i \partial x_k}\left(\frac{1}{c}\right)Q_k \tag{3.42}$$

Equation (3.41) and (3.42) are dynamic ray tracing equations and are used to compute Q_i **and** P_i for the central ray (Cervený and Hron,1980, Cervený., 2001).

3.72 Paraxial ray tracing equations

We define δx_i ***and*** δp_i as parameters that connect a paraxial ray to the central ray using the following approximation:

$$\boldsymbol{\delta x_i \approx \frac{\partial x_i}{\partial \gamma} d\gamma = Q_i\, d\gamma} \qquad (3.43)$$

$$\boldsymbol{\delta p_i \approx \frac{\partial p_i}{\partial \gamma} d\gamma = P_i\, d\gamma} \qquad (3.44)$$

Multiplying equation (29) and (30) with $\boldsymbol{\delta\gamma}$ and apply equation (3.43) and (3.44) yields:

$$\frac{dQ_i}{ds}\delta y = c_{,k}\, Q)k\; p_i\delta y + cP_i\delta y$$

$$\frac{d}{ds}\delta x_i = c_{,k}\,\delta x_k p_i + c\delta p_i \qquad (3.45)$$

$$\frac{dP_i}{ds}\delta y = \frac{\partial^2}{\partial x_i \partial x_k}\left(\frac{1}{c}\right) Q_k\delta y$$

$$\frac{d}{ds}\delta p_i = \frac{\partial^2}{\partial x_i \partial x_k}\left(\frac{1}{c}\right)\delta x_k = \left(\frac{2c_{,i}c_{,k}}{c^3} - \frac{c_{,ik}}{c^2}\right)\delta x_k \qquad (3.46)$$

Equation (3.45) and (3.46) are paraxial ray tracing equations and are used to compute $\boldsymbol{\delta x_i}$ **and** $\boldsymbol{\delta p_i}$ for paraxial ray from $\boldsymbol{c, \vec{\nabla} c}$ **and** $\boldsymbol{\vec{p}}$ (Beydoun and Keho, 1987, Cervený., 2001).

3.7.3 Geometrical spreading factor

Geometrical spreading can be computed from the ratio of cross-sectional areas. Equation (3.43) shows that the cross-sectional area of paraxial ray can be computed directly from $\boldsymbol{Q_1}$ **and** $\boldsymbol{Q_2}$ and the ray parameters $\boldsymbol{\gamma_1}$ ***and*** $\boldsymbol{\gamma_2}$ (Cervený., 2001):

$$\boldsymbol{d\sigma = \left|\frac{\partial \vec{x}}{\partial \gamma_1} \mathrm{x} \frac{\partial \vec{x}}{\partial \gamma_2}\right| d\gamma_1 d\gamma_2 = Q_1 Q_2 d\gamma_1 d\gamma_2} \qquad (3.47)$$

3.7.4 Paraxial ray traveltimes

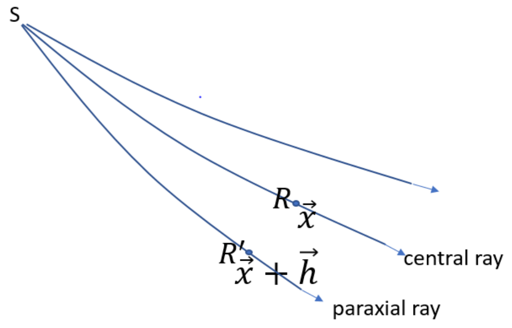


Figure 3. 12. Paraxial ray and traveltime (Beydoun and Keho, 1987)

As shown in figure 3.12, a point R is at $\vec{\boldsymbol{x}}$ on the central ray and a point R' is at $\vec{\boldsymbol{x}} + \vec{\boldsymbol{h}}$ on a nearby ray. Using the 3D Taylor series to relate R and R', we obtain:

$$T(\vec{x} + \vec{h}) = T(\vec{x}) + T_{,j}(\vec{x})h_j + \frac{1}{2}T_{,jk}(\vec{x})h_j h_k, \qquad (3.48)$$

where $\boldsymbol{T}(\vec{\boldsymbol{x}})$ is travel time at R, $\boldsymbol{T}_{,\boldsymbol{j}}$ **is the first derivative of traveltime and equals** $\boldsymbol{p}_{\boldsymbol{j}}$. $\boldsymbol{T}_{,\boldsymbol{jk}}$ is the second derivative of traveltime and can be computed from:

$$T_{,jk} = \frac{\partial p_j}{\partial x_k} = \frac{\partial p_j}{\partial \gamma_n}\frac{\partial \gamma_n}{\partial x_k} = P_{jn}Q_{nk}^{-1} \qquad (3.49)$$

$$\text{where } Q_{nk}^{-1} = \frac{\partial \gamma_n}{\partial x_k}$$

Or in matrix form: $\boldsymbol{T}_{,\boldsymbol{jk}} = \mathbf{T} = \mathbf{PQ^{-1}}$ (3.50)

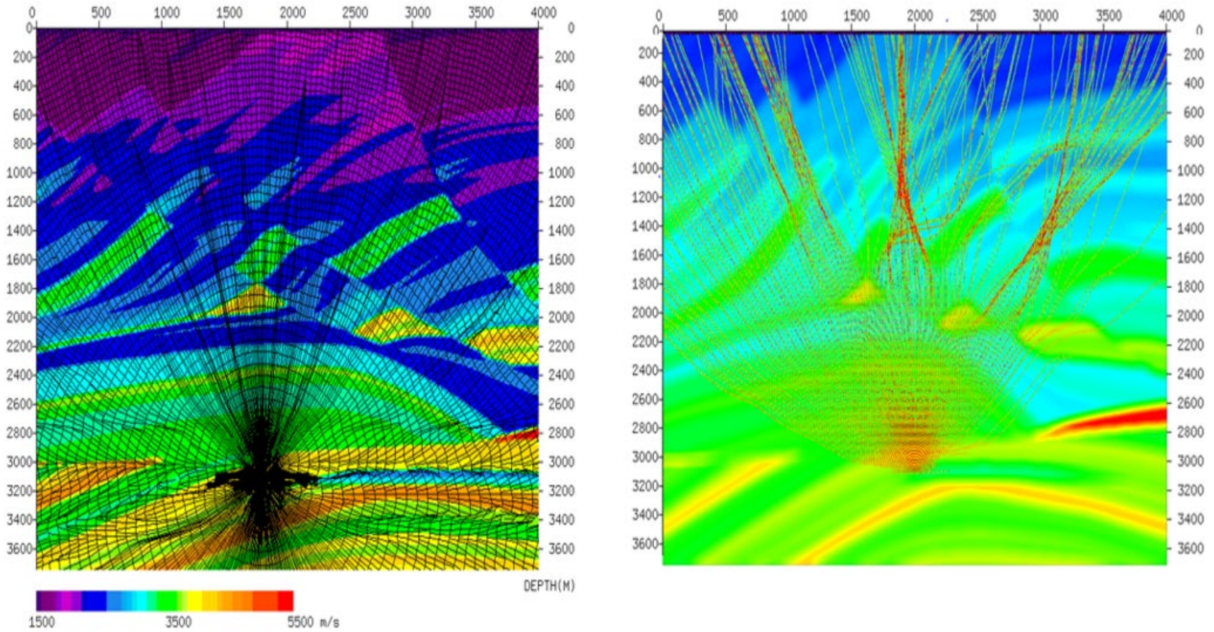
The paraxial ray tracing algorithm is outlined in the following steps:

1) Shoot a ray through the medium with starting vertical angle $\boldsymbol{\theta_i}$ and horizontal angle $\boldsymbol{\phi_i}$
2) Solve ODE 3.38 for displacement $\boldsymbol{x_i}$ for the central ray
3) Solve ODE 3.39 for slowness vector $\boldsymbol{p_i}$ for the central ray
4) Solve ODE 3.41 and 3.42 for $\boldsymbol{P_i}$ **and** $\boldsymbol{Q_i}$
5) Use equation 3.49 to compute $\boldsymbol{T}_{,\boldsymbol{jk}}$
6) Use equation 3.48 to compute paraxial travel time for paraxial rays near the central ray

Repeat steps 1 to 6 for all starting angle $\boldsymbol{\theta_i}$ and horizontal angle $\boldsymbol{\phi_i}$

3.8 Comparisons of travel times from ray-tracing methods

To verify and compare the accuracy of the travel times computed from WFC (Figure 3.13a), fast marching and paraxial method (Figure 3.13b), we use the Marmousi model (Brougois et al, 1990) with a buried source placed at the depth of 2500m and compute the travel times from these three methods. A second-order finite difference shot record was created using the Seismic Unix module sufdmod2. Travel times at the surface are plotted on the shot record with travel times from WFC plotted in red, fast marching plotted in blue and paraxial method plotted in yellow (Figure 3.14). As shown in figure 3.14, the travel times at the surface from WFC and fast marching are almost identical. Travel times at the surface from the paraxial method that uses the shortest ray path agree with the other two methods at most locations except at locations where rays diverge. This test demonstrated all three methods result in very similar travel times that agree with the finite difference shot record. Both WFC and fast marching methods produce smooth and stable minimum travel times. Rays in the paraxial method may diverge and create large gaps that can result in inaccurate travel times.



(a)

(b)

Figure 3. 13. a) Ray paths and wavefronts from the WFC method, b) Ray paths from the Paraxial method.

WFC and paraxial methods also show that rays can cross over in an area with a complex velocity structure. These cross-over ray paths result in multi-arrivals at the same grid point. Furthermore, WFC computes geometric spreading amplitude using cross-sectional area ratio at the starting and end points of ray segment and the paraxial method computes amplitude from dynamic ray tracing equations. Therefore, WFC and paraxial method can be used when multi-arrivals or different branches of traveltime including most energetic arrivals is desirable.

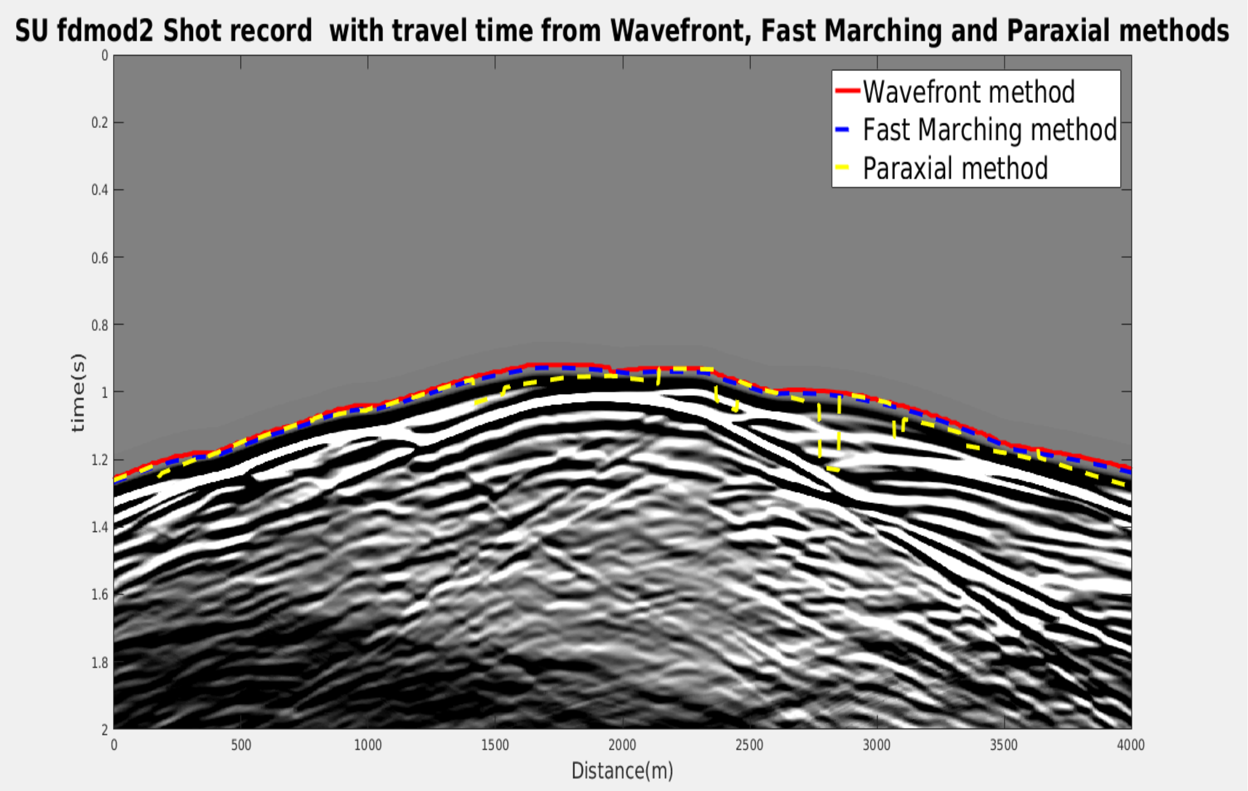


Figure 3. 14. Finite difference synthetic shot record with first arrival times from WFC, fast marching and paraxial method

3.9 Summary of ray tracing methods

As discussed before, there are many different ray tracing algorithms, all with applications in different areas of research. In the list below, I present the main advantages and disadvantages of each method.

Fast marching method:

- Advantages
    - Unconditionally stable
    - Can handle turning rays. Does not have a shadow zone problem.
    - Computes the first arrival time for every grid point without interpolation.
    - Excellent algorithm for refraction tomography

- Disadvantages
    - Does not compute ray paths directly but the wavefronts.
    - Does not compute multi-arrivals.
    - Does not compute amplitude.
    - Can be slow for a large output grid.

Wavefront Construction method:

- Advantages
    - Stable if appropriate velocity smoothing parameter is used; however, accuracy can decrease with increasing smoothing
    - Can handle turning rays. Does not have a shadow zone problem.
    - Can compute multi-arrivals and amplitude
    - Can be faster than the fast marching method, if a larger step size is used.
    - A good algorithm for refraction tomography as well as depth imaging

- Disadvantages
  - Ray paths from interpolated ray segments may not be accurate enough for tomographic inversion.

Paraxial method:

- Advantages
  - A good tradeoff between speed and accuracy.
  - More accurate travel time interpolation in the vicinity of the central ray than the classical ray shooting method.
  - Can compute multi-arrivals and amplitude
  - A good algorithm for depth imaging

- Disadvantages
  - Cannot handle turning ray. Not suitable for refraction tomography
  - Can have problems with ray path divergence and shadow zone in areas with complex structures.

Ray shooting method:

- Advantages
  - Fast and accurate.
  - Can compute multi-arrivals
  - A good algorithm for depth imaging

- Disadvantages
  - Cannot handle turning ray. Not suitable for refraction tomography
  - Travel time interpolation is not as accurate as the paraxial method in the vicinity of the central ray
  - Does not compute amplitudes
  - Can have problems with ray path divergence and shadow zone in areas with complex structure.

Application of ray tracing methods in depth imaging

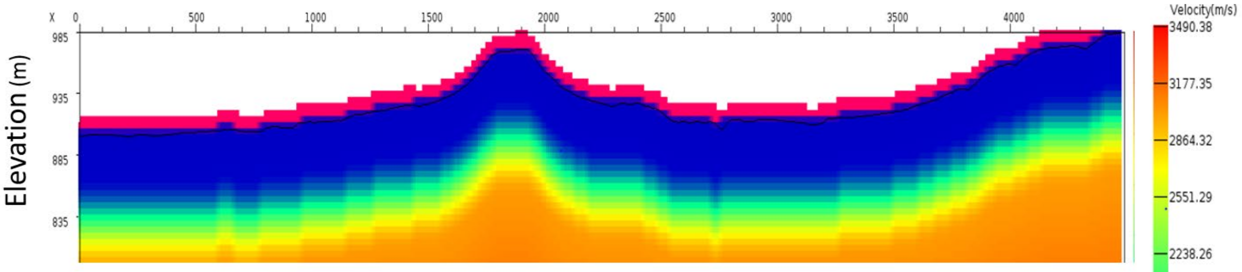
All methods tested show similar accuracy; while WFC and the paraxial method are capable of computing multi-value traveltimes. This poses a challenge in determining which arrival times to use as well as storage and computational resources in retrieving these values. However, when minimum time or shortest path is not the optimal approach, the multi-values capability of WFC and paraxial methods can improve the imaging result.

I did not perform a comprehensive analysis of the effects of these ray tracing methods in depth imaging. However, based on the observed geometry of the ray paths in our tests using the Marmousi model; I believe the proper application of the multi-arrivals, amplitude and ray path distance information from WFC and paraxial method can have a significant impact on the quality of the final depth image. This statement is based on the property of WFC and paraxial methods to create smooth traveltime tables, which help imaging algorithms like Kirchhoff to produce clean images.

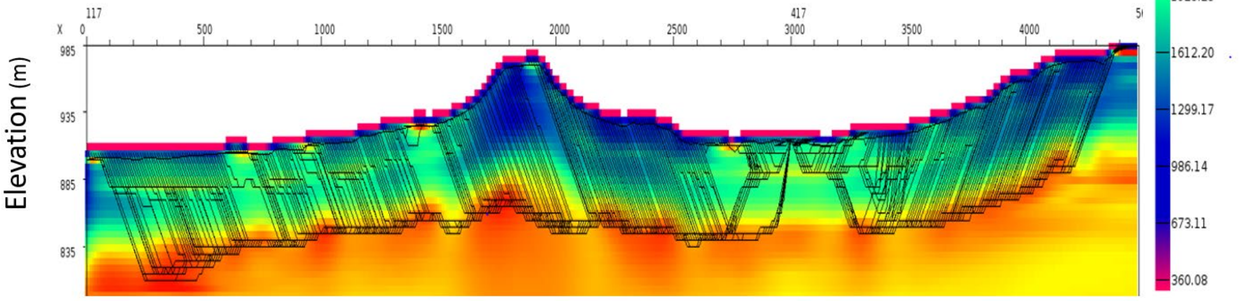
Application of ray tracing methods in refraction traveltime tomography

Refraction traveltime tomography involves forward modelling of first arrival times and using the differences between the modelled times and the actual first arrival time picks to update the velocity model along the ray path. I use different ray tracing methods in refraction tomography to invert for the near-surface velocity model. Ray shooting method and paraxial method are not suitable for ray tracing refraction ray paths because of their inability to handle up-turning rays. Both WFC and fast marching methods can handle up-turning rays; therefore, they are more suitable for refraction tomography. We used the fast marching method for forward modelling in refraction tomography and apply the refraction tomography process to the Hussar 2D line acquired in 2011 by CREWES of the University of Calgary (Margrave et al., 2012). We compared the CDP stack with the tomographic statics correction to the CDP stack with GLI (Hampson and Russel 1984) weathering statics correction. GLI is one of the delay time methods and has found great success when the near-surface can be approximated by layers with a distinct difference in velocity but has a problem with gradational velocity changes and rough topography.

A velocity model with a constant velocity gradient between layers and with the depth of the layer boundaries following the recording surface as shown in figure 3.15a is used as the starting model. Figure 3.15b shows the updated velocity model and refraction ray paths for shot location 417 after 10 iterations. Shown in figure 3.16 is the comparison of modelled refraction arrival times and the actual first arrival picks before and after tomographic inversion for shot location 417 as well as the RMS error for all time picks after each iteration. Fig 3.16b shows that modelled refraction arrival times from the final velocity model match the actual first arrival picks.



a) Starting Model



b) After 10 iterations

Figure 3. 15. a) Starting model for tomographic inversion. b) Velocity model after 10 iterations using traveltimes from fast marching method. Ray paths from shot location 417 are shown.

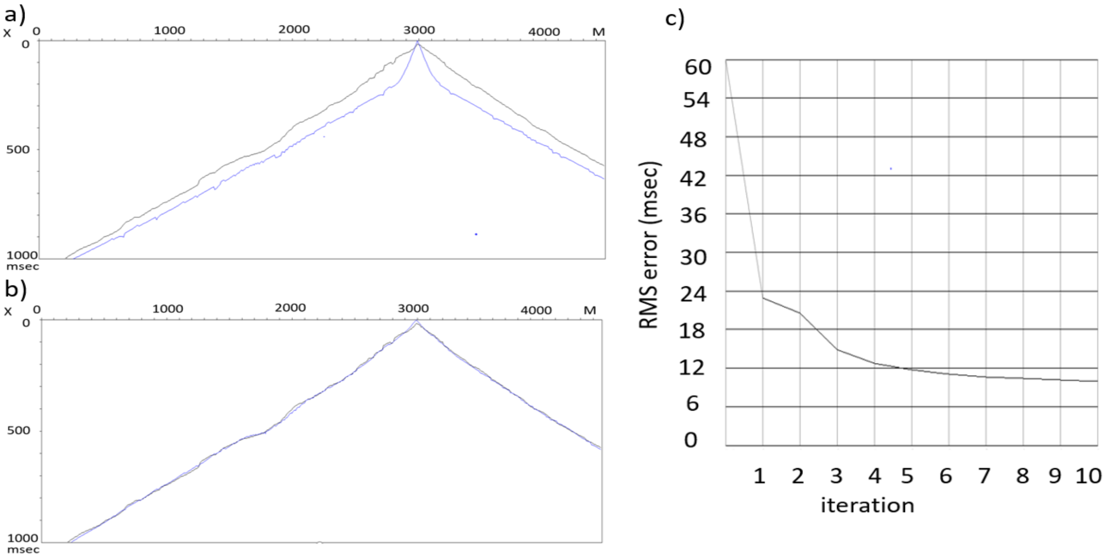


Figure 3. 16. a) Actual first arrival times from shot location plotted in black, minimum travel times from starting model plotted in blue, b) Minimum travel times from velocity model after 10 iterations, c) RMS error at each iteration.

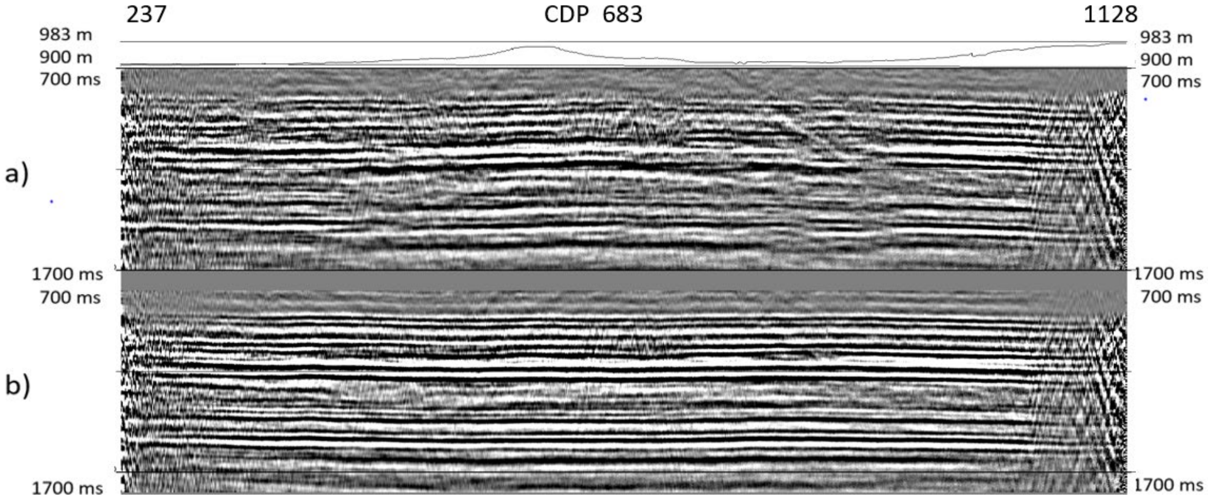


Figure 3.17. (a) CDP stack with datum statics correction, (b) CDP stack with tomographic weathering statics correction.

Figure 3.17 compares the CDP stacks with datum statics correction only and with tomographic weathering statics correction. The images from the tomographic weathering statics corrected CDP stack are more coherent and better resolved than the datum statics corrected CDP stack. These results demonstrate that the fast marching method is accurate and is effective when used in refraction tomography.

3.10 Conclusion

Fast marching, WFC, ray shooting method and paraxial method are all based on the principles of high-frequency ray theory. They all produce accurate travel times when the velocity model varies smoothly. Similar to the classical ray shooting method, the paraxial method has the problem of diverging ray paths and shadow zones in areas of complex structure. WFC alleviates this problem by interpolating additional ray segments along wavefronts to ensure sufficient ray density. Both paraxial and WFC methods can produce multi-arrivals as well as amplitude tables, which are important for true amplitude migration. However, the fast marching method can only produce minimum traveltime values. Therefore, WFC and paraxial methods are better suited for depth imaging of complex structures, where multi-arrivals are necessary to recreate the complexity of the wavefields by using several wavefronts. The fast marching method expands the wavefront and computes traveltime from the source to each grid cell without additional interpolation by solving the eikonal equation. Both fast marching and WFC methods can handle up-turning rays; therefore, they can be used in refraction tomography. We used the fast marching method in the refraction tomography processing of the Hussar 2D lines. The CDP stack image from the refraction tomography processing is more coherent and better resolved than the CDP stack with datum statics correction. Therefore, refraction travel times computed from the fast marching method are accurate and the velocity model from the refraction tomography is reliable and can potentially be used as starting model for full-waveform inversion and depth imaging. We did not perform a comprehensive analysis of the effects of these ray tracing methods in depth imaging. However, based on the observed geometry of the ray paths in our tests using the Marmousi model, we believe the proper application of the multi-value traveltime, amplitude and ray path distance information from WFC and paraxial method can have a significant impact on the quality of the final depth image.

Chapter 4

Trave-time tomography inversion problem using feedbacks from reflection data

This chapter combines the review of the inverse problem with a proposal of using feedbacks from reflection data to improve the near-surface velocity model. I include the following published paper in the chapter (reproduced with Permision from Society of Exploration Geophysics).

Robust refraction statics solution and near-surface velocity model building using feedback from reflection data
Bernard Law and Daniel Trad, 2018
Published in Geophysics, **83**, P. U63-77

Abstract

An accurate near-surface velocity model is critical for weathering statics correction and initial model building for depth migration and full waveform inversion. However, near-surface models from refraction inversion often suffer from errors in refraction data, insufficient sampling and over-simplified assumptions used in refraction algorithms. Errors in refraction data can be caused by picking errors resulting from surface noise, attenuation and dispersion of first arrival energy with offset. These errors are partially compensated later in the data flow by reflection residual statics. Therefore, surface consistent residual statics contain information that can be used to improve the near-surface velocity model. We present a new dataflow to automatically include median and long wavelength components of surface consistent reflection residual statics. This technique can work with any model-based refraction solution, including grid-based tomography

methods and layer-based methods. We modify the cost function of the refraction inversion by adding model and data weights computed from the smoothed surface consistent residual statics. By using an iterative inversion, these weights allow us to update the near-surface velocity model and to reject first arrival picks that do not fit the updated model. In this nonlinear optimization workflow, the refraction model is derived from maximizing the coherence of the reflection energy and minimizing the misfit between model arrival times and the recorded first arrival times. This approach can alleviate inherent limitations in shallow refraction data by using coherent reflection data.

4.1 Introduction

The earth near-surface is known to have localized variations in material properties, which can introduce time varying delays to reflection traveltimes. Furthermore, all ray paths travelling through un-accounted for near-surface velocity anomalies will be affected and this will result in imaging errors. For 2D split-spread or symmetrical 3D receiver recording patch, these imaging errors extend to an area that is within half cable length or receiver patch to either side of the velocity anomalies (Jones, 2012). Therefore, accurate determination of near-surface velocity variations is essential to the successful imaging of deeper reflection events. Seismic acquisition is designed to optimally illuminate deeper targets. It often does not provide sufficient shallow reflection data for near-surface velocity modeling. First arrivals of refracted waves from seismic reflection surveys have been used to create near-surface velocity models for initial static corrections. There are many different methods to obtain near-surface velocity models from refraction arrival times. Refraction methods that use a layered model include delay time method (Gardner, 1939, Barry, 1967), plus-minus method (Hagedoorn, 1959), generalized reciprocal method (Palmer, 1981), generalized linear inversion (GLI) method (Hampson and Russell, 1984) and weathering layer tomography

(Docherty, 1992). Some of these methods compute the velocity and thickness of a few near-surface layers directly using refraction travel time equations and shot and receiver geometry. GLI and weathering layer tomography use inversion schemes to compute the thickness and velocity parameters by minimizing the misfits between the modeled and actual first arrival times. Turning-ray refraction tomography methods (White 1989; Zhu, Sixta and Angsman 1992; Belfer and Landa, 1996) discretize the near-surface velocity model into rectangular cells. Rays are traced through these cells between source and receivers by solving the ray equations (Červený, 2001; Langan, Lerche and Cutler,1984; Vinje et al.,1993) or the eikonal equation (Vidale, 1990; Qin, 1992; Sethian and Papovici,1999). Turning-ray refraction tomography back propagates the misfits between the actual and ray-traced first arrival times along the ray paths to update the velocity grid. The GLI method can provide stable solutions; however, it is limited by the assumption of layered-based refraction model with velocity increasing in depth. In areas of complex geological structures and surface terrain, and when the velocity model is better represented by velocity gradients, the simplified assumption used in the GLI methods is often violated. Turning-ray tomography methods are not limited by this simple assumption, and therefore can model near-surface velocity changes with higher resolution than the GLI method. However, it suffers from instability when ray-density is low, especially at large shot point gaps and at edges of 2-D and 3-D surveys. The starting model, grid size and smoothing parameters can also influence the final model of turning-ray tomography. In addition, refraction solutions and corrections are also affected by the quality of refraction data and acquisition geometry. Therefore, refraction statics corrections often contain errors caused by the quality of the refraction data, numerical errors of the refraction solution and the inability of the refraction algorithm to model the actual physical properties of the near-surface. This can result in unsatisfactory statics corrections and reflection images. These problems are often addressed by revising the parameters and constraints of the refraction algorithm and by using deeper reflection data for the surface consistent residual statics. Using a surface-consistent hypothesis, Taner et al. (1974) showed that surface consistent residual statics correction can yield optimally stacked CDP section with lateral signal continuity representative of real geologies. Ronen and Claerbout (1985) demonstrated that surface-consistent residual statics can be estimated by stack-power maximization. They also pointed out that statics estimation is effectively a velocity analysis of the near-surface, but they do not use the surface-consistent residual statics derived from more coherent and better sampled reflection data in refraction inversion.

High resolution near-surface velocity models from turning-ray refraction tomography are often integrated with reflection velocity models and used in depth imaging and depth model building processes. Uncertainties in near-surface velocity models derived from refraction data alone tend to accumulate and adversely affect the velocity building process of the deeper layers. Kosloff et al. (1997) and Pecholcs et al. (1997) used depth errors from subsurface image gathers to update velocity and thickness in the near-surface layers. Birdus et al. (2013) used a velocity model from reflection tomography as the starting point for their iterative joint refraction/ reflection tomography workflow. The integrity of the reflections is the priority of this joint inversion.

The objective of this work is to use surface-consistent reflection residual statics detected by correlation of reflection data to alleviate limitations and errors in refraction data and refraction solution algorithms and to harmonize near-surface velocity models with reflection events. We introduce new cost functions for GLI and turning-ray refraction tomography that incorporate model space and data space regularization using medium and long wavelength residual statics derived from reflection correlation. We describe how these model and data weights for the new cost functions are computed and test the procedure with a synthetic dataset. We propose a nonlinear optimization workflow for refraction statics using the new cost functions. We use a 2D field dataset to show that the new method can improve the structural integrity of the reflection events and the reliability of the near-surface velocity model.

4.2 Theory

4.2.1 Linear Inversion

A linear inverse problem can be cast as the inversion of a parameter $\boldsymbol{m}$ using a linear operator $\boldsymbol{L}$ and data $\boldsymbol{d}$:

$$\boldsymbol{d} = \boldsymbol{L}\,\boldsymbol{m}\,, \qquad (4\text{-}1)$$

The parameter $\boldsymbol{m}$ can be computed by minimizing the objective function J:

$$J = \|\, \boldsymbol{d} - \boldsymbol{Lm} \,\|^2. \qquad (4\text{-}2)$$

The linear least squares solution of equation (4-2) is (Claerbout, 1992):

$$\boldsymbol{m} = (\,\boldsymbol{L}^T\boldsymbol{L} + \mu I)^{-1}\boldsymbol{L}^T\,\boldsymbol{d} \qquad , \qquad (4\text{-}3)$$

where μ is the stabilization parameter.

Alternatively, this linear inverse problem can be posed in terms of the Fréchet derivatives $\boldsymbol{G}$, changes in the model parameter $\boldsymbol{\delta m}$, and differences between the initial model response and the observed data $\boldsymbol{\delta d}$ (Lines and Treitel, 1984):

$$\boldsymbol{\delta d} = \; \boldsymbol{G}\,\boldsymbol{\delta m}. \qquad (4\text{-}4)$$

Fréchet derivatives are the partial derivatives of the modelled response with respect to the model parameters. The least squares solution of equation (4-4) is:

$$\boldsymbol{\delta m} = (\,\boldsymbol{G}^T\boldsymbol{G} + \mu I)^{-1}\boldsymbol{G}^T\boldsymbol{\delta d}. \qquad (4\text{-}5)$$

4.2.2 Nonlinear Inversion

For refraction inversion, the problem is nonlinear because the operator $\boldsymbol{L}$ is a function of parameters $\boldsymbol{m}$. The cost function becomes:

$$J = \| \boldsymbol{d} - \boldsymbol{L}(\boldsymbol{m}) \|^2, \tag{4-6}$$

and the model parameter $\boldsymbol{m}$ cannot be obtained directly using equation (4-3). This non-linear problem can be solved by a sequence of linear least squares estimates of $\boldsymbol{\delta m}$ using equation (4-5) and updating the model parameters iteratively (Lines and Treitel, 1984):

$$\boldsymbol{m^k} = \boldsymbol{m^{k-1}} + \boldsymbol{\delta m} \, , \tag{4-7}$$

where k is the iteration number. The iteration stops when the modelled response fits the observations within a selected convergence criterion.

In the case of refraction inversion, $\boldsymbol{m}$ is the near-surface model parameters, $\boldsymbol{d}$ is the first arrival time picks and $\boldsymbol{L}$ is the forward modelling operator that maps $\boldsymbol{m}$ into $\boldsymbol{d}$. $\boldsymbol{\delta m}$ contains the model perturbations between iterations, $\boldsymbol{\delta d}$ is the data difference between the modelled and the observed first arrival times and $\boldsymbol{G}$ is the matrix that contains the Fr$\acute{e}$chet derivatives of the modelled first arrival times with respect to the model parameter $\boldsymbol{m}$. $\boldsymbol{G}$ maps $\boldsymbol{\delta m}$ into $\boldsymbol{\delta d}$.

4.2.3 GLI method

A layer-based refraction model with two layers is depicted in Figure 4.1, where the total travel time, T, from the source (S) to the receiver (R) can be computed from the ray path SBCR in Figure 4.1:

$$T = 2 \, \frac{Z_1 \; cos\theta_c}{V_0} + \frac{X}{V_1}, \tag{4-8}$$

where Z_1 is the thickness of the layer, V_0 is the velocity of the layer, θ_c is the critical angle, X is the offset distance from S to R and V_1 is the velocity of the underlying refractor. Equation (4-8) can be written for the total travel time for an n layer media:

$$T_n = \int_{k=1}^{n} 2 \, \frac{Z_k \; cos\theta_{c_k}}{V_{k-1}} \; + \frac{X}{V_n} \; , \tag{4-9}$$

where Z_k and θ_{c_k} are the thickness and critical angle for the k^{th} layer; V_{k-1} and V_n are velocities for layers *k-1* and *n*.

Hampson and Russell (1984) presented a first-break interpretation method that uses the Generalized Linear Inversion technique to iteratively update the model parameters of a near-surface velocity model. The GLI method computes the model perturbation via first-order Taylor expansion and relates the errors in T_n to the model perturbations in $V_k\ and\ Z_k$ using the following sets of linear equations:

$$\Delta T = B\Delta m \tag{4-10}$$

$$B = \partial T/\partial m \tag{4-11}$$

where ΔT are the changes in ray-traced time between iterations, Δm are the model updates between iterations; and $\partial T/\partial m$ are the partial derivative of travel time with respect to the model parameters $V_k\ and\ Z_k$. The least squares solution for a model update Δm is:

$$\Delta m = (B^T B + \mu I)^{-1} B^T\ \Delta T \tag{4-12}$$

Equation (4-10) and (4-12) are equivalent to equation (4-4) and (4-5), and B in equation (4-10), (4-11) and (4-12) is equivalent to G in equation (4-4) and (4-5).

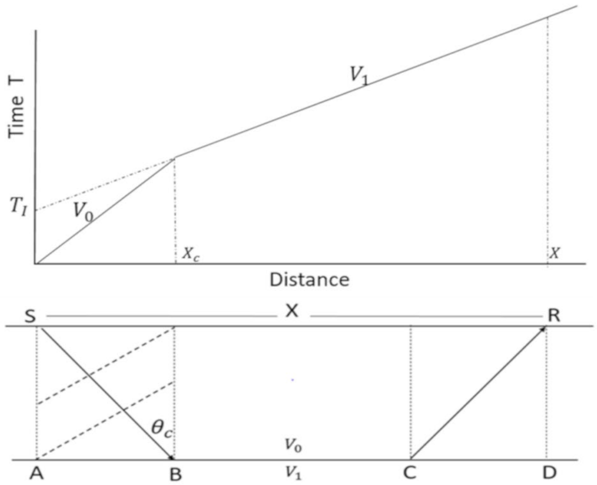


Figure 4. 1. Time-Distance plot and refraction raypath. $\boldsymbol{V_0}$ is the velocity of the first layer, $\boldsymbol{\theta_c}$ is the critical angle. X is the offset distance from S to R and $\boldsymbol{V_1}$ is the velocity of the underlying refractor. SBCR is the refraction raypath. $\boldsymbol{T_I}$ is the intercept time for $\boldsymbol{V_1}$ and $\boldsymbol{X_c}$ is the critical distance ABCDEFG.

4.2.4 Turning-ray refraction tomography

Turning-ray refraction tomography methods discretize the near-surface velocity model into a grid of rectangular cells. Figure 4.2 shows a near-surface velocity grid and the relationship between ray path geometry and travel times. Rays are traced through the velocity cells between source and

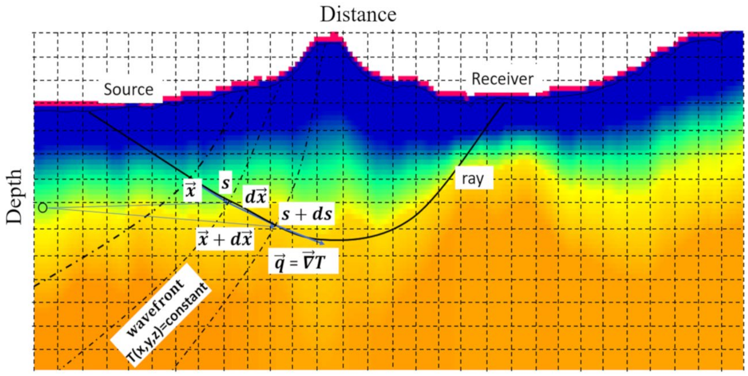


Figure 4. 2. Relationship between ray path geometry and travel time. $\boldsymbol{\vec{x}}$ and $\boldsymbol{\vec{x} + d\vec{x}}$ are the positions of 2 points along the ray path on two adjacent wavefronts of constant travel times separated by the distance vector $\boldsymbol{d\vec{x}}$. $\boldsymbol{s}$ and $\boldsymbol{s + ds}$ are the ray segment lengths from the source to

$\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{x}} + \boldsymbol{d}\vec{\boldsymbol{x}}$. $\vec{\boldsymbol{q}}$ is the slowness vector at $\vec{\boldsymbol{x}} + \boldsymbol{d}\vec{\boldsymbol{x}}$ and $\boldsymbol{\nabla T}$ is the travel time gradient at the same location.

receivers using diving rays by solving the ray equations (4-13) and (4-14) or the eikonal equation (4-15):

$$\frac{d\vec{x}}{ds} = c\,\vec{q} \quad , \tag{4-13}$$

$$\frac{d\vec{q}}{ds} = \vec{\nabla}\left[\ \frac{1}{c}\ \right] , \tag{4-14}$$

$$\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2 = \frac{1}{c^2} , \tag{4-15}$$

where c is the velocity, $\vec{q}$ is the slowness vector, $\frac{d\vec{x}}{ds}$ is a unit vector tangential to the ray, and T is the travel time. If equations (4-13) and (4-14) are used to trace the ray, travel times can be computed by integrating the slowness model along the ray path. If the eikonal equation is used to compute the travel time, the ray paths can be traced along the paths of the maximum travel time gradient (Vidale,1988) or along the paths of minimum time (Matsuoka,1992). Ray path distances l_{ij} within velocity cells computed from the ray tracing process form the Fréchet derivative G in equation (4-4) and (4-5):

$$\boldsymbol{G} = \begin{bmatrix} \frac{\partial T_1}{\partial m_1} & \frac{\partial T_1}{\partial m_2} & .. & \frac{\partial T_1}{\partial m_J} \\ \frac{\partial T_2}{\partial m_1} & \frac{\partial T_2}{\partial m_2} & .. & \frac{\partial T_2}{\partial m_J} \\ .. & .. & .. & .. \\ \frac{\partial T_I}{\partial m_1} & \frac{\partial T_I}{\partial m_2} & .. & \frac{\partial T_I}{\partial m_J} \end{bmatrix} , \tag{4-16}$$

where T_i is the modelled travel time for ray path i, $\frac{\partial T_i}{\partial m_j} = l_{ij}$ and is the ray segment length for ray path i and cell j. The $\boldsymbol{\delta m}$ vector contains the slowness update ΔM_j^k for the k^{th} iteration and can be solved iteratively using equation (4-5).

Lo and Inderwiesen (1994) proposed the Simultaneous Iterative Reconstruction Technique (SIRT) to solve for slowness update ΔM_j iteratively without matrix operations:

$$\Delta M_j^k = \frac{1}{W_j} \sum_{i=1}^{I} \, l_{ij} \delta d_i / \sum_{j'=1}^{J} L^2{}_{ij'} \, , \qquad (4\text{-}17)$$

where i is the observation number, I is the total number of observations, j is the model cell to update, j' to J is the range of model cells that the ray path has traversed, and l_{ij} is the ray segment length for observation number i and model cell j , W_j is the total number of rays intersecting the model cell j and k is the iteration number. SIRT is computationally more efficient than solving equation (4-5) directly and is used in the refraction tomography inversion in this paper.

4.3 Reflection residual statics and near-surface velocity model update

Errors in the refraction solution arise when the modelling operator $\boldsymbol{L}$ is unable to model the observed first arrival times because of poor refraction data quality, numerical errors of the refraction solution, and the inability of the refraction algorithm to model the actual physical property of the near-surface. We classify these errors *as* data error ϵ_d, model error ϵ_m and algorithm error ϵ_p. The data error is equivalent to poor pick quality. The model error is the difference between our current estimates and the actual subsurface parameters. The algorithm error is the error in the method because of several approximations like discrete derivative evaluations and coarse parameterization. These errors contaminate the refraction correction $\mathrm{C_{wx}}$ and are often revealed on CDP stack sections as deterioration in the reflection coherence and structural integrity. Some of these errors are often compensated by applying surface consistent residual statics corrections derived from correlation of reflection data. In conventional refraction and reflection residual statics workflows, the surface consistent residual statics are applied to the reflection data to compensate for deficiencies in the near-surface velocity model caused by these errors; however, the near-surface velocity model is not updated. If the surface consistent residual statics are caused by the deficiencies in the near-surface velocity model and contain these errors, we can back propagate these errors vertically as model weights to update the near-surface velocity model. For layer-based models, these errors can result in velocity or thickness error, or both. The following deviation assumes these errors contribute to only velocity or thickness error. Consider the layered model shown in Figure 4.3. The weathering statics correction can be computed by

$$T = \sum_{i=0}^{n}(\frac{1}{Vr} - P_i\,)\,Zs_i + (\frac{1}{Vr} - P_i\,)\,Zr_i \qquad , \tag{4-18}$$

where Zs_i and Zr_i are the thickness of layer i at source and receiver location, P_i is the slowness for layer i and V_r is the replacement velocity for weathering statics correction. Let us define

$$T_i = \frac{Z_i}{V_r} \; - Z_i P_i \,, \tag{4-19}$$

$$E \; = smoothed \;\; residual\, statics \;\; , \tag{4-20}$$

$$E_i = E\,(\;\frac{Z_i}{Total\, thickness}\,), \tag{4-21}$$

$$Z_i = Thickness\, for\; layer\; i \;, \tag{4-22}$$

$$P_i = \frac{1}{V_i}, \tag{4-23}$$

$$W_{mi}(slowness) = slowness\; model\; weight\; for\; layer\; i, \text{and} \tag{4-24}$$

$$W_{mi}(thickness) = thickness\; model\; weight\; for\; layer\; i\;, \tag{4-25}$$

where T_i is the weathering statics correction for layer i. Adding a smoothed surface consistent statics correction E_i to the weathering statics correction T_i, and updating only P_i with $W_{mi}(slowness)$ yields:

$$T_i + E_i = \frac{Z_i}{V_r} \; - Z_i P_i W_{mi}(slowness)\,. \tag{4-26}$$

Replacing T_i with $\frac{Z_i}{V_r} - Z_iP_i$ gives:

$$\frac{Z_i}{V_r} - Z_iP_i + E_i = \frac{Z_i}{V_r} - Z_iP_iW_{mi}(slowness) \text{ , and} \tag{4-27}$$

$$W_{mi}(slowness) = 1 - \frac{E_i}{Z_iP_i} \tag{4-28}$$

Similarly, updating only Z_i with $W_{mi}(thickness)$ yields:

$$T_i + E_i = (\frac{Z_i}{V_r} - Z_iP_i)\, W_{mi}(thickness) \tag{4-29}$$

Replacing $\frac{Z_i}{V_r} - Z_iP_i$ with T_i gives:

$$T_i + E_i = E_i\, W_{mi}(thickness),\ and \tag{4-30}$$

$$W_{mi}(thickness) = 1 + \frac{E_i}{T_i} \tag{4-31}$$

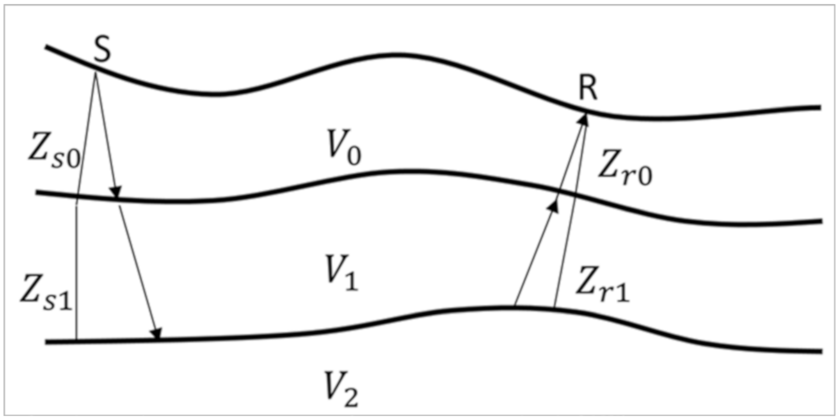


Figure 4. 3. Refraction model and weathering statics correction. $\boldsymbol{Z_{s0}}$, $\boldsymbol{Z_{r0}}$, $\boldsymbol{Z_{s1}}$ and $\boldsymbol{Z_{r1}}$ are thickness of layer $\boldsymbol{0}$ and 1 at source and receiver location S and R. $\boldsymbol{V_0}$, $\boldsymbol{V_1}$ and $\boldsymbol{V_2}$ are the velocity for layer 0, 1 and 2.

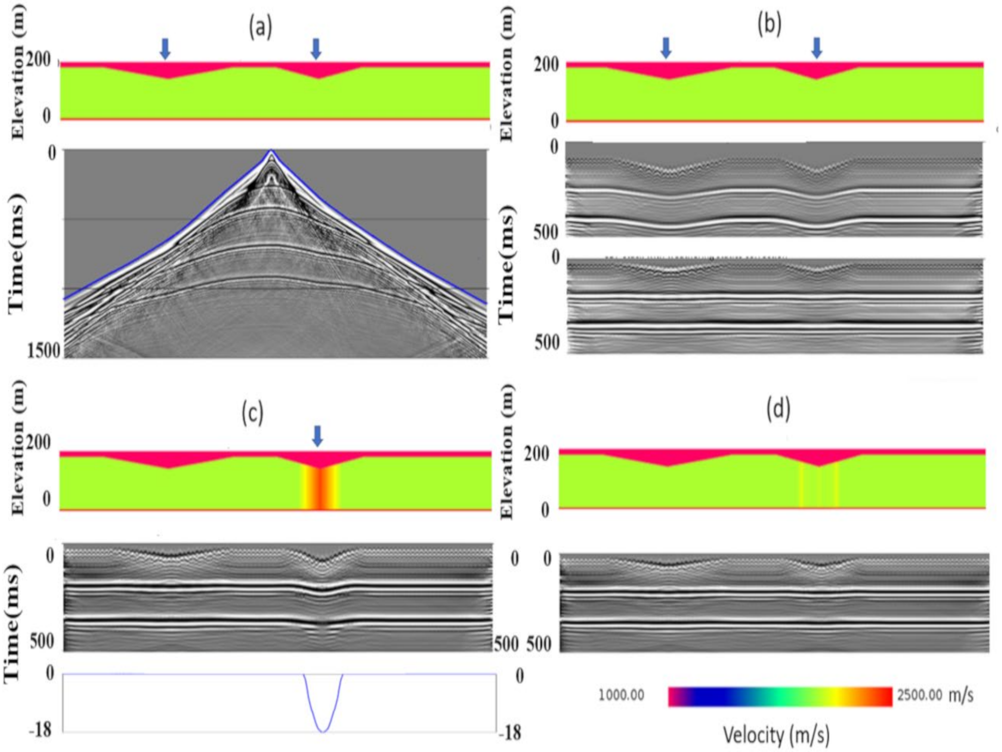


Figure 4. 4. (a) Finite difference synthetic data with velocity variations (marked by blue arrows) in the near-surface. (b) (top) near-surface velocity model, (middle) CDP stack without weathering statics corrections, (bottom) CDP stack with weathering statics correction. (c) (top) Error (marked by the blue arrow) introduced to near-surface velocity, (middle) CDP stack with weathering statics correction from the model with error, (bottom) surface consistent residual statics from reflection data. (d) (middle) Modified near-surface velocity model using model weights.

To illustrate the concept of model weights, we create a finite-difference synthetic dataset with a velocity model with 6 layers of velocities 1000, 2000, 2500, 3000, 3500 and 4000 m/sec. Both receiver spacing and depth step are 5 m. Two weathering pockets in the model are centred at stations 251 and 601. Distortion to reflection events caused by the weathering pockets can be seen in figure 4.4a. Figure 4.4b compares the CDP stacks before and after weathering statics correction. Figure 4.4c shows the near-surface velocity model with an arbitrary error introduced, the CDP stack with reduced coherence caused by weathering statics correction using the erroneous model and the surface consistent residual statics computed from reflection data. Figure 4.4d shows the updated velocity model $W_m m$, and the CDP stack with weathering statics correction from the updated velocity model. This synthetic data test demonstrates that incorrect near-surface velocity can reduce the coherence and structural integrity of the reflection stack. It also demonstrates that the surface consistent reflection residual statics process can detect near-surface statics errors and that the model weight W_m can be computed from the surface consistent reflection residual statics.

4.4 Application of data weight and model

Data weight W_d and model weight W_m in the cost function are a commonly used approach in geophysics (Claerbout 1992). Application of data weight and model weight is equivalent to data space and model space regularization. Regularization in the data space helps to reduce the effects of outliers in data picks on the solution. Regularization in the model space stabilizes the solution and provides a means of applying a priori information into the inversion (Zhou et al., 2003; Trad et al.,2003). We include the model weight W_m and data weight $W_d$in the cost function of the inversion problem,

$$J = \| \mathbf{W_d d} - \mathbf{W_d L W_m m'} \|^2, \text{ and } \mathbf{m} = \mathbf{W_m m'} \tag{4-32}$$

For layer-based method, E_i can be caused by errors in the slowness and/or thickness. Any combination of contributing factors can be used. When updating both slowness and thickness with equal distribution, equal factors of 0.5 can be applied to E_i and W_{mi} (slowness) and W_{mi} (thickness) are:

$$W_{mi}(slowness) = 1 - 0.5\frac{E_i}{Z_i P_i} \quad , \tag{4-33}$$

$$W_{mi}(thickness) = 1 + 0.5\frac{E_i}{T_i} \quad , \quad \text{and} \tag{4-34}$$

$$W_d = \{ \begin{matrix} 0 & \text{for } E \geq \varepsilon \text{ and } \delta t > K \times std(\delta \mathrm{t}) \\ 1 & otherwise \end{matrix} \tag{4-35}$$

where δt is the difference between observed and modelled first arrival time, $std(\delta t)$ is the standard deviation of δt, ε and K are thresholds used for W_d. We use equations (33) and (34) to compute W_{mi} (slowness) and W_{mi}(thickness) for the GLI algorithm. W_d corrects for data errors and is computed from the misfit between d and $L\ W_m m$ **.**

For turning-ray tomography, W_m is typically applied to slowness only. However, if applying W_m results in unreasonable velocity values, we may have to apply W_m to thickness as well. This is done by stretching the distance between the surface and the intermediate datum and remapping the velocity values. Using an equal distribution of model weights for slowness and thickness, equal factors of 0.5 can be applied to E_i and W_m (slowness) and W_m (thickness) can be computed as:

$$W_m \text{ (slowness)} = 1 - 0.5E/T\ , \tag{4-36}$$

$$W_m \text{ (thickness)} = 1 + 0.5E/T\ , \text{and} \tag{4-37}$$

$$T = \sum_{iz=1}^{idatum} \left(\frac{1}{Vr} - P_{iz} \right) dz\ , \tag{4-38}$$

where $idatum$ is the number of depth steps to the intermediate datum and P_{iz} is slowness at depth step iz. It is important to review the initial model after applying the model weight W_m (slowness) and W_m (thickness) to confirm the proper distribution of the model weights is used.

W_m and W_d can be incorporated into equation (4-4) for turning-ray refraction tomography as:

$$W_d\ G\ W_m\ \delta m = W_d\ \delta d. \tag{4-39}$$

$W_d\ G\ W_m$ represents the new G matrix that is updated on each external iteration. W_m is applied to the Fréchet derivatives computation in G matrix, and W_d is used to reject data points that do not agree with the new model. The Fréchet derivatives in this G matrix are used in the computation of $\boldsymbol{\delta m}$ during the internal iterations with SIRT.

4.5 Inversion procedure

The conventional refraction statics and reflection residual statics processing flow are outlined in Figure 4.5. In this conventional processing flow, the refraction model $\boldsymbol{m}$ is computed by minimizing the square of the misfit between the first arrival picks $\boldsymbol{d}$ and the modelled first arrival times $\boldsymbol{Lm}$. Refraction statics corrections are then calculated from the computed model $\boldsymbol{m}$.

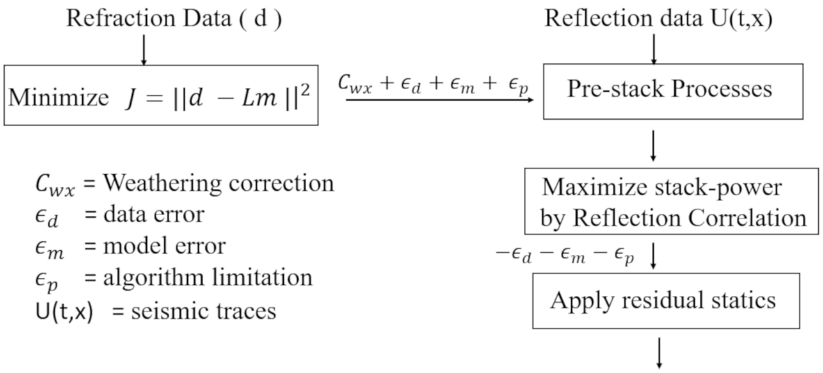


Figure 4. 5. Conventional refraction and reflection statics workflow. C_{wx} is the weathering correction computed by refraction inversion that solves for model parameter m by minimizing the cost function $J = \| d - Lm \|^2$. ϵ_d, ϵ_m and ϵ_p are data error, model error and algorithm limitation associated with the refraction inversion.

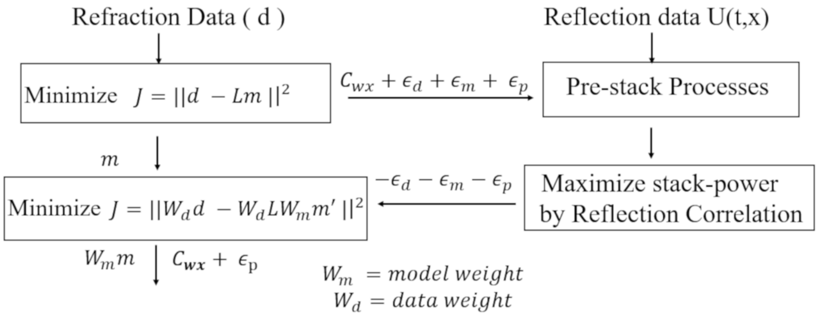


Figure 4. 6. Nonlinear optimization of the near-surface velocity model. C_{wx} is the weathering correction computed by refraction inversion that solves for model parameter m by minimizing the cost function $J = \| d - Lm \|^2$. ϵ_d, ϵ_m and ϵ_p are data error, model error and algorithm limitation

associated with the refraction inversion. W_m and W_d are the model and data weights for the new cost function.

However, if there are errors in the refraction's solution, these errors will be applied to the reflection data. These errors are partially compensated later in the data flow by surface consistent reflection residual statics. The proposed nonlinear optimization of near-surface velocity model processing flow using model and data weights described above is outlined in Figure 4.6. In this proposed processing flow, the refraction model $\boldsymbol{m}$ is computed by minimizing the original cost function. Weathering statics corrections are computed and applied to the reflection data. Surface consistent reflection residual statics are then computed using the cross-correlations of the reflection data. The smoothed surface consistent reflection residual statics $\boldsymbol{E}$ is then used to compute $\boldsymbol{W_m}$ and $\boldsymbol{W_d}$ for the new cost function. The weathering statics correction $\boldsymbol{C_{wx}}$ computed from the initial updated model $\boldsymbol{W_m m}$ is equivalent to applying smoothed surface consistent residual statics $\boldsymbol{E}$ to the seismic data. Subsequent iterations of minimizing $\| W_d d - W_d L\, W_m m' \|^2$ will produce a near-surface velocity model that is in harmony with the refraction and reflection data and can produce better imaging results. This processing flow is also outlined in the following steps:

1. Minimize $J = \| d - L\, m \|^2$ and apply weathering statics correction to seismic data.
2. Compute surface consistent reflection residual statics.
3. Compute smoothed surface consistent residual statics $\boldsymbol{E}$, $\boldsymbol{W_m}$ and $\boldsymbol{W_d}$.
4. If required, pick again first arrival times using modelled first arrival times $W_m L\, m$ as constraints.
5. Minimize $J = \| W_d d - W_d L\, W_m m' \|^2$.

6. Iterate 2 to 5 until there are no significant improvements in the near-surface velocity model and the reflection image.

Comparing to the conventional processing flow (Figure 4.5), the additional computation cost for this new processing flow is the cost for running another GLI or refraction tomography inversion as well as reflection residual statics. GLI inversion is computationally efficient even for large 3D surveys; therefore, computational cost should not be a concern. For refraction tomography, we can reduce the computational cost by using fewer iterations because we are starting with a refined model. For seismic surveys challenged by near-surface problems caused by near-surface conditions or acquisition limitations, we often try to enhance the reflection image by re-computing the near-surface refraction solution, involving manually revising the inversion parameters, editing the starting model and first arrival picks. Therefore, comparing to the manual approach, this new processing flow does not increase the computational cost. Moreover, it improves the manual approach by automatically updating the starting model and rejecting outlying picks that do not agree with the reflection coherence. This new processing flow assumes that surface consistent residual statics makes a significant improvement to the coherence and structural integrity of the reflection image. Therefore, choosing the optimal parameters for the reflection residual statics process to overcome acquisition limitations is very important. For example, it is important to use reflection residual parameters that allow large smoothing or macro-binning radius at edges or gaps of the seismic survey where the CDP fold is low. For this new method to make a noticeable difference to the near-surface velocity model and the reflection image, there must be a significant amount of medium or long-wavelength components in the surface consistent residual statics, because these components determine the magnitude of the model weight $\boldsymbol{W_m}$ and data weight $\boldsymbol{W_d}$ for the new cost function.

4.6 Field data example

CDP stack sections are created using near-surface velocity models computed from the conventional and the new non-linear optimization workflow. The data used in this example are the vertical component of the dynamite shot records from a 4.5 Km 2D 3C survey acquired at Hussar, Alberta in September 2011. The seismic survey was acquired for a broadband experiment (Margrave et al., 2012).

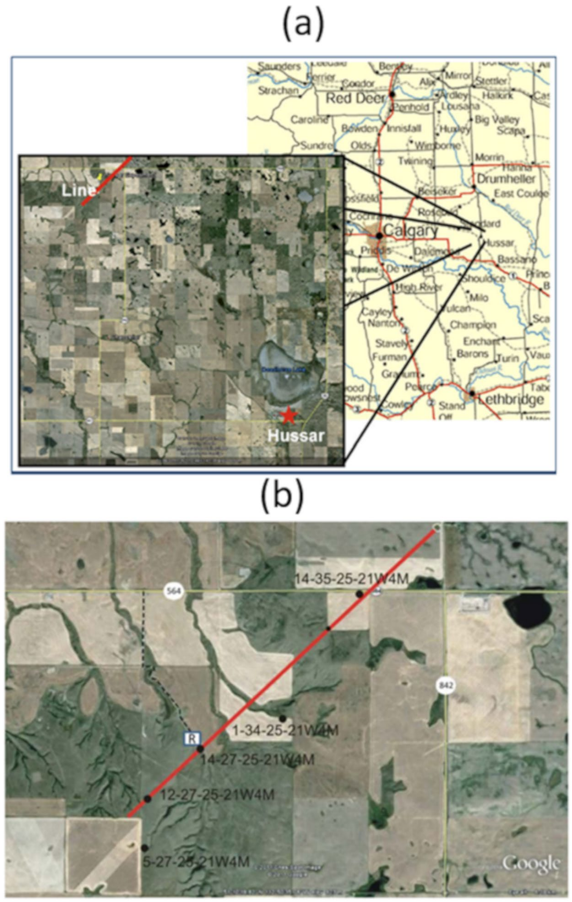


Figure 4. 7. Hussar 2D broadband experiment (a) location map, (b) seismic line layout.

Figures 4.7a and 4.7b show the location and the layout of the seismic line. The seismic line runs NE-SW with a topographic relief of about 80 m. The receiver interval is 10 m and the shot point interval is 20 m. The 448-channel split-spread geometry gives a nominal maximum offset of 2240 m for standard spread and a maximum offset of 4480 for off-end shots. The nominal CDP fold for offsets 0 to 1500 m is 80. First arrivals were picked for all traces and offsets; however, only first arrivals with an offset less than 3000 m were used in the refraction solution. Figure 4.8 shows the time-distance plot of the first arrival picks with a distinct difference between layer velocities V1 and V2. This is a good indication that layer-based refraction inversion methods such as GLI can produce a stable solution. We created the common-receiver stack (Figure 4.9) and the CDP stack (Figure 4.10) with datum statics corrections only. Effects of near-surface time delays are obvious on the common receiver stack. The incoherency of seismic events on the CDP stack is likely the result of the same near-surface time delays. To test the proposed nonlinear optimization workflow for the near-surface velocity model, we follow the steps outlined in figure 4.6. We create the first near-surface velocity model by minimizing $|| d - Lm ||^2$ for both GLI and turning-ray refraction tomography. We then use the weathering statics corrected gathers from both methods to compute surface consistent reflection residual statics. We smooth the surface consistent residual statics and use them to compute the model weight W_m. The smoothing length determines the smoothness of the model weight. A small smoothing length can introduce rapid changes to the near-surface velocity model and result in unreasonable near-surface velocity model and erroneous data weights. A large smoothing length can reduce the effectiveness of the model and data weights. We chose a smoothing length of 31 receiver stations because the smoothed residual statics represents a reasonable medium and long wavelength residual statics and should provide reasonable model weights to start the next refraction iteration. A threshold parameter of 2 times the standard deviation of traveltime residuals is used for W_d. The near-surface velocity model is then updated iteratively by minimizing the cost function $\| W_d d - W_d L W_m m' \|^2$.

The GLI solutions, residual error analysis and error distributions from the two processing flows are plotted in Figure 4.11, 4.12 and 4.13 respectively. CDP stacks with weathering statics correction from the two GLI solutions (Figure 4.11a and 4.11c) are plotted in Figure 4.14. Figure 4.12 shows the misfits between the first arrival picks and the modelled first arrival time for the two processing flows at iteration 0, 10 and 20. The vertical and horizontal alignment of misfits in Figure 4.12a represents receiver anomalies and shot anomalies in the near-surface velocity model prior to the GLI inversion. The misfits are reduced as the GLI solutions converge with iterations. The misfits after 10 iterations of minimizing the new cost function (Figure 4.12e) is smaller than the misfit after 20 iterations of minimizing the original cost function (Figure 4.12c). Figure 4.13a and 4.13b show the error distribution for the first and second layers after 10 iterations of the original GLI inversion; while figures 4.13c and 4.13d show the error distribution of the first and second layers after 10 iterations of the new GLI inversion. Errors are reduced after the new GLI inversion.

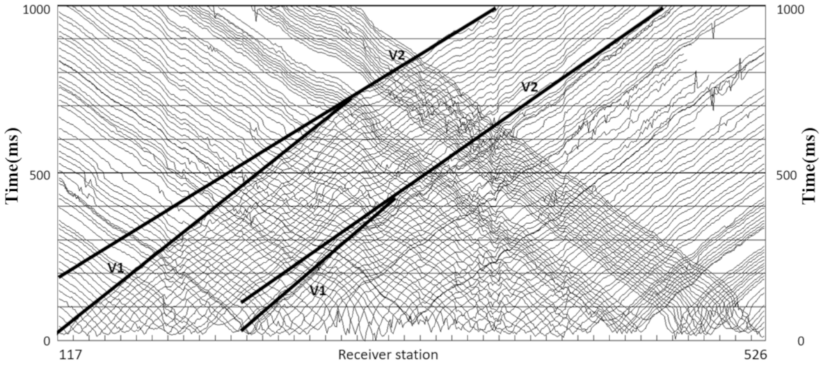


Figure 4. 8. First arrival picks for Hussar 2D survey

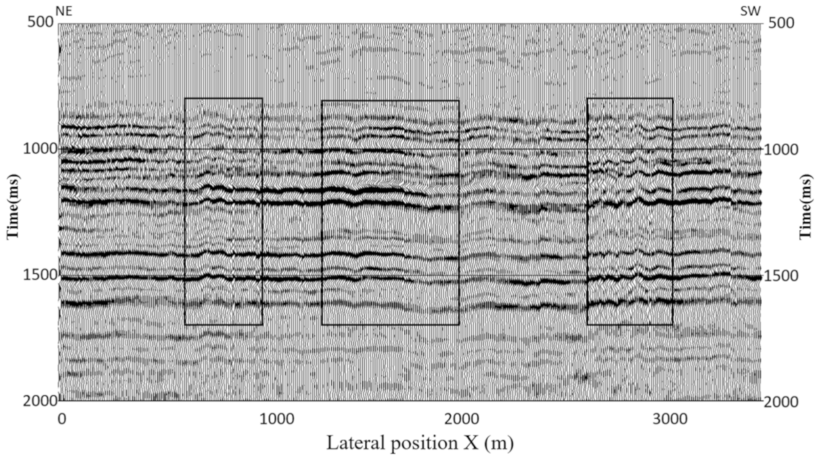


Figure 4. 9. Common receiver stack showing near-surface time delays

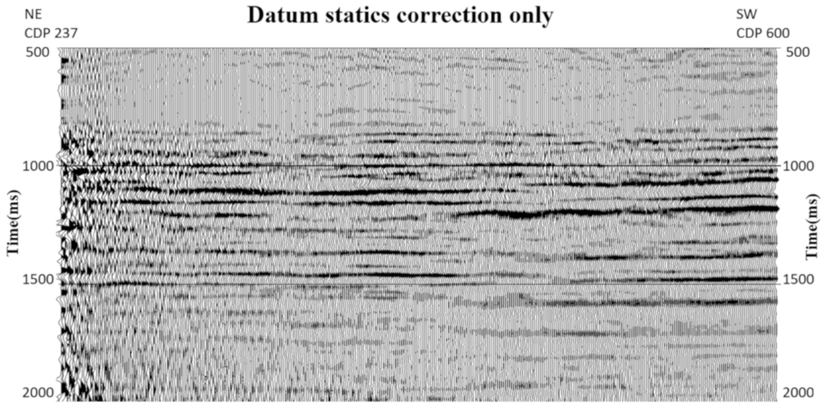


Figure 4. 10. CDP stack with datum statics correction only

The smoothed surface-consistent reflection residual statics (Figure 4.11b) from gathers corrected with the GLI solution from the original cost function are in the range of -2.7 to 3.3 msec. They are small; however, a small long-wavelength trend dipping from SW to NE can be observed at the NE end of the seismic line. The difference between the GLI solutions (Figure 4.11a and 4.11c) is small and occurs mostly at the two ends of the profile. CDP stacks with weathering statics correction from the two GLI methods (Figures 4.14a and 4.14b) show significant improvement in coherence and structural integrity when compared to the CDP stack with datum statics correction only. However, there is no significant difference in coherence between the two CDP stacks with different GLI corrections. This test confirms that small smoothed residual statics will not result in significant improvement in the near-surface velocity model or reflection image. For seismic surveys that are affected by acquisition limitations, such as 3D surveys with large shot line and receiver line spacing, and in areas where near-surface velocity is better represented by a velocity gradient, or in areas where first arrival picking is prone to cycle skipping errors we would expect more difference from the two processing workflows.

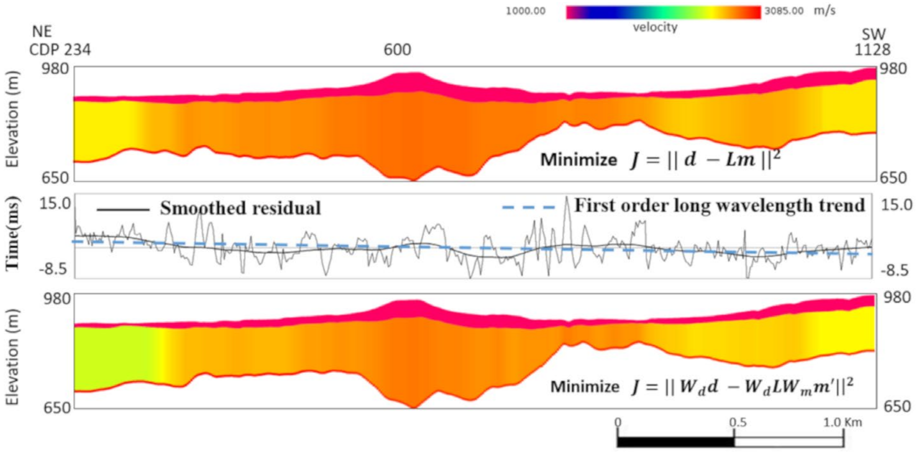


Figure 4. 11. (a) GLI near-surface velocity model computed from minimizing the original cost function. (b) Surface-consistent residual statics from reflection correlation, smoothed residual statics and first-order long-wavelength trend. (c) Near-surface velocity model computed from the new cost function with model and data weights derived from smoothed surface consistent reflection residual statics.

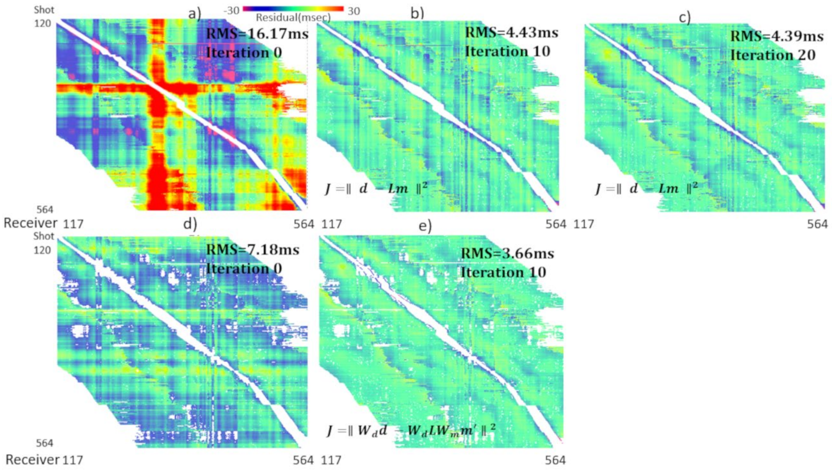


Figure 4.12. Top: misfits between first arrival picks and modelled first arrival times (a) prior to GLI inversion, (b) after 10 iterations, (c) after 20 iterations. Bottom: misfits between first arrival picks and modelled first arrival times (d) after applying $\boldsymbol{W_m}$ **and** $\boldsymbol{W_d}$, (e) after 10 iterations of the new GLI iterations.

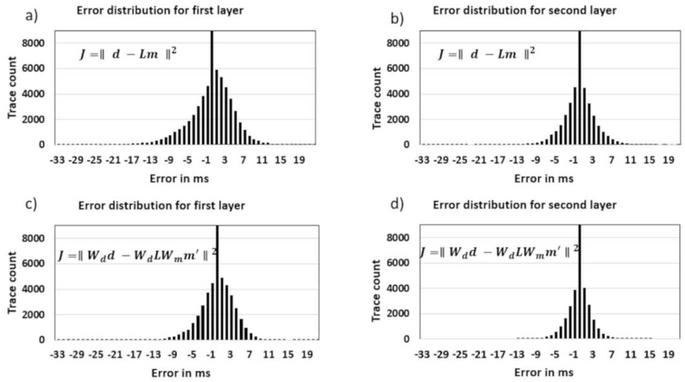


Figure 4.13. Error distribution after minimizing the original cost function for the first layer (a) and the second layer (b). Error distribution after minimizing the new cost function for the first layer (c) and the second layer (d).

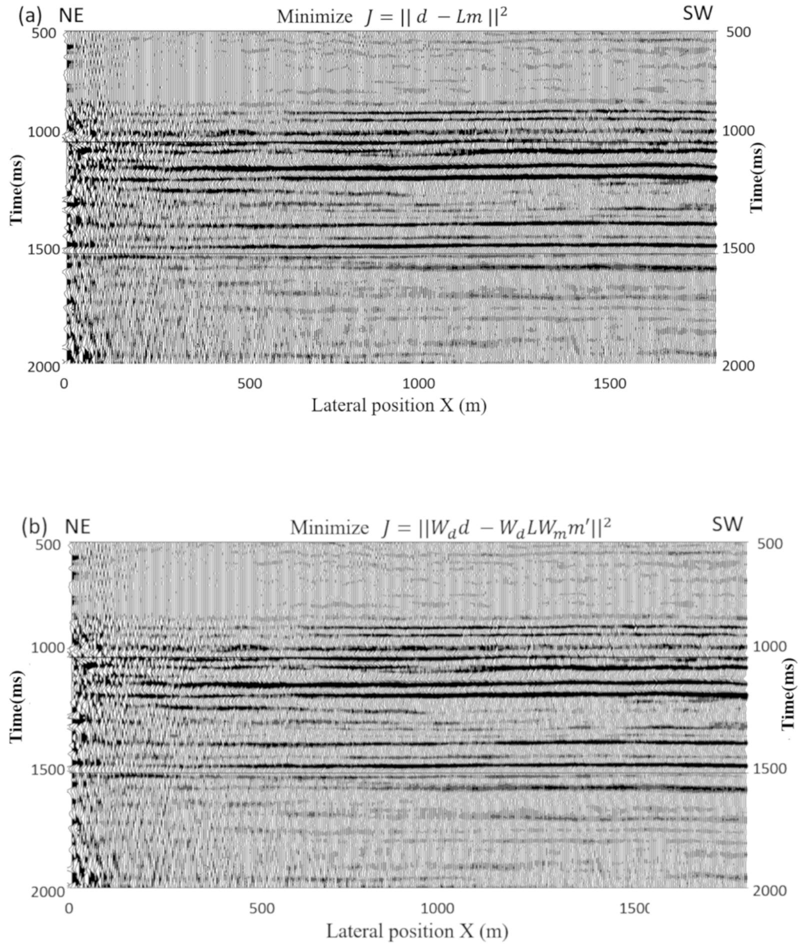


Figure 4.14. CDP stack section with weathering statics correction using GLI velocity model computed from minimizing (a) the original cost function, (b) the new cost function with model weight and data weight derived from surface constant reflection residual statics.

To confirm the effectiveness of the new approach for the GLI method in areas where first arrival picking is prone to cycle skipping errors, we imposed a 30 msec picking error to two-third of the first arrival picks greater than 300 msec and between receiver 250 and 300. The imposed picking errors and GLI solutions are plotted in Figure 4.15. The original first arrival picks and the first arrival picks with imposed errors over the receiver range of 117 to 327 are displayed in Figures 4.15a and 4.15c. The comparisons of the GLI solutions from input with and without imposed first arrival picking error are displayed in Figures 4.15b and 4.15d. The blue rectangle in Figure 4.15d marks the receiver ranges where imposed picking errors are added. Within this receiver range, we can see a thickening of the second layer in the GLI model (Figure 4.15d). We compute weathering statics corrections from the two solutions and their CDP stacks (Figure 4.16).

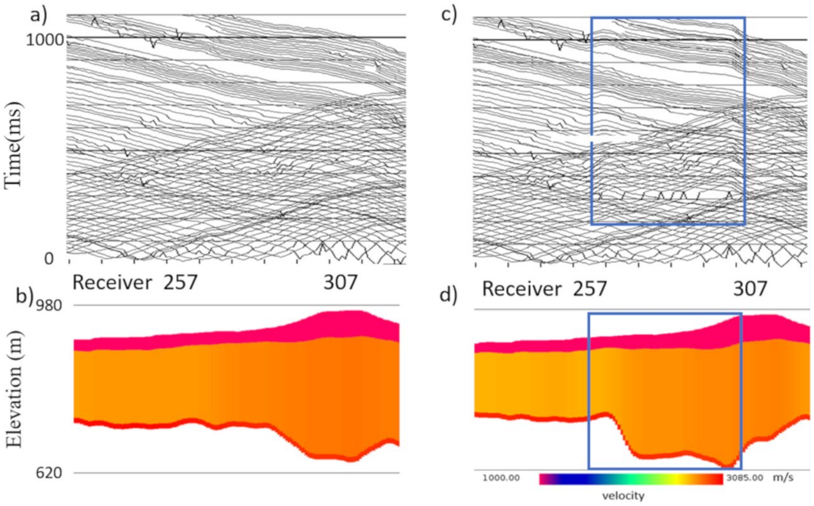


Figure 4.15. (a) Original first arrival pick over the receiver range of 117 to 327 (b) GLI solution from the original first arrival picks, (c) first arrival picks with 30ms errors added to 2/3 of the shots over the receiver range of 250 to 300, (d) GLI solution with imposed first arrival pick errors. The blue rectangle marks the receiver range where first arrival pick errors are added.

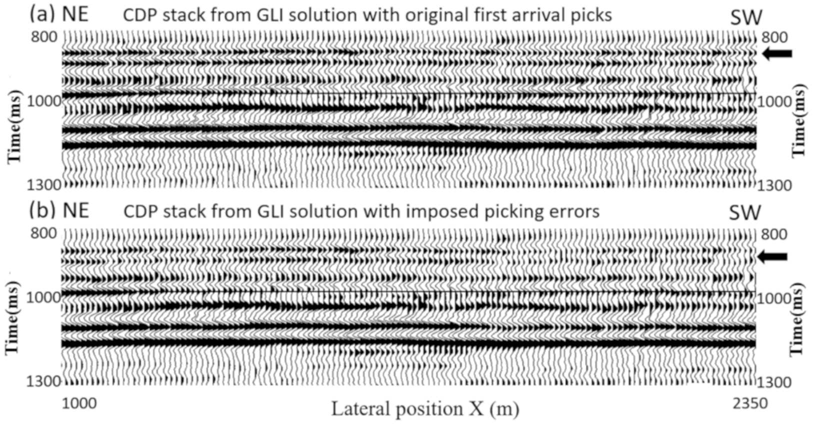


Figure 4.16. (a) CDP stack section with weathering statics correction using GLI solution without imposed first arrival pick errors (b) CDP stack section with weathering statics correction using GLI solution with imposed first arrival pick errors.

Deterioration of coherence can be seen in the CDP stack (Figure 4.16b) from the GLI solution with imposed pick errors. We use the weathering statics corrected gathers from the GLI solution with first arrival pick errors (Figure 4.15d) to compute surface consistent reflection residual statics. We use the same smoothing length of 31 receiver stations as in the previous test to compute the smoothed surface consistent reflection residual statics. The threshold parameter of 2 times the standard deviation of traveltime residuals is used for W_d. The near-surface velocity model is then updated iteratively by minimizing $\| W_d d - W_d L \, W_m m' \|^2$. The original GLI solution with first arrival pick errors, surface-consistent reflection residual statics, the new GLI solutions and the new surface-consistent residual statics after applying the new GLI solution are plotted in Figure 4.17. The smoothed reflection residual statics (Figure 4.17b) are in the range of -4 to 8 msec and correlate with the thickening of the original GLI solution caused by picking errors. The GLI solution from the new approach (Figure 17c) is closer to the solution without imposed first break picking errors shown in Figure 4.15b. Figure 4.17d shows the surface consistent residual statics from the correlation of reflection data after applying the new GLI solution. The new GLI solution does not completely remove the medium wavelength components of surface-consistent reflection residual statics. The smoothed reflection residual statics after applying the new GLI solution are in the range of -3 to 3 msec. This can be the limit of the sensitivity and resolution of the refraction inversion and agrees with the observations we found in the first GLI test.

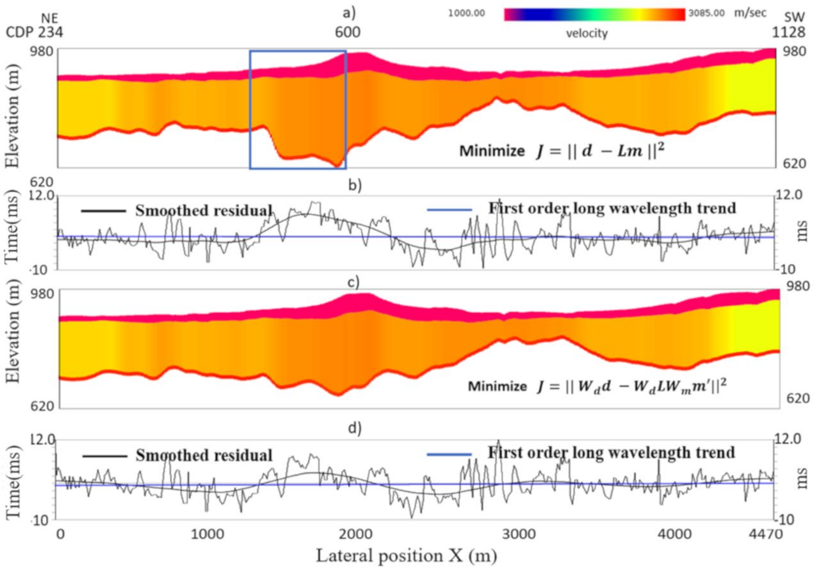


Figure 4. 17. (a) GLI near-surface velocity model computed from minimizing the original cost function. The blue rectangle marks the receiver range where first arrival pick errors are added. (b) Surface-consistent residual statics from reflection correlation, smoothed residual statics and first-order long-wavelength trend. Large positive smoothed residual correlates with the area with the first arrival pick errors. (c) Near-surface velocity model computed from the new cost function with model and data weights derived from smoothed surface consistent reflection residual statics. The excess thicknesses caused by first arrival pick errors are corrected. (d) Surface-consistent residual statics from reflection correlation after applying the new GLI solution.

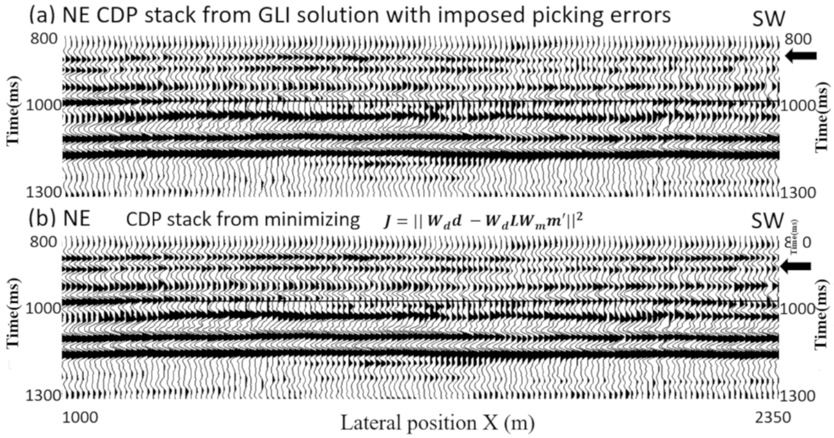


Figure 4. 18. CDP stack section with weathering statics correction using GLI velocity model computed from minimizing (a) the original cost function, (b) the new cost function with model weight and data weight derived from surface constant reflection residual statics.

CDP stacks are created using the two GLI solutions. The windows of data near the marked area, where picking errors were introduced, are plotted in Figure 4.18. Noticeable improvement in the reflection coherence is observed in the new CDP stack (Figure 4.18b). This test confirms that for a GLI solution that is negatively impacted by systematic picking errors due to poor first arrival pick quality, correlation of reflection data can be used to construct the model and data weights to guide and constrain the GLI solution in this new approach. However, large smoothed surface consistent reflection residual statics are required for improvements to be significant. In the above test, this was achieved with smoothed residual statics of 8 msec. The residual error analysis and error distribution from the two processing flows are plotted in Figures 4.19 and 4.20. The misfits between first arrival picks and modelled first arrivals time before and after GLI inversion for input with imposed first arrival picking error are plotted in Figures 4.19a and 4.19b respectively. The picking errors appear as receiver consistent anomalies marked by blue arrows in figure 4.19a.

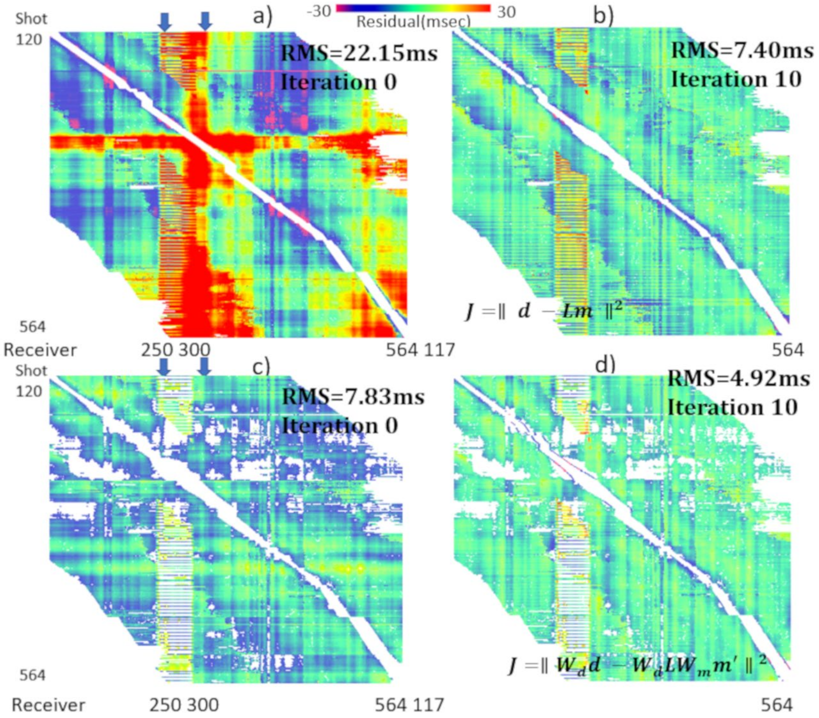


Figure 4. 19. (a) misfits between first arrival picks and modelled first arrival times prior to GLI inversion showing the effects of imposed errors, (b) misfits after GLI inversion showing most imposed errors are removed by the final GLI solution. Bottom: (c) misfits between first arrival picks and modelled first arrival times after applying W_m and W_d ,(d) after 10 iterations of minimizing the new cost function.

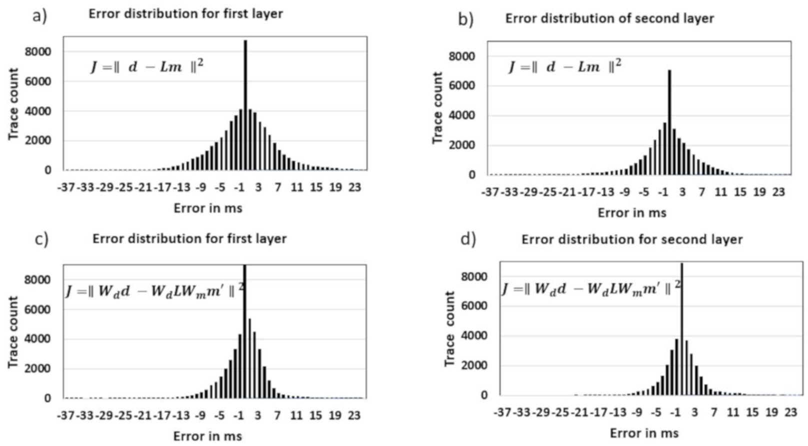


Figure 4. 20. Error distributions for GLI test with imposed pick errors. Top: After minimizing the original cost function for the first layer (a) and the second layer (b). Bottom: after minimizing the new cost function for the first layer (c) and the second layer (d).

Figure 4.19b shows the picking errors were corrected for by the final GLI solution. This results in the overestimation in layer thickness shown in figure 4.17a. Smoothed surface-consistent residual statics (Figure 4.17b) are used to compute the model and data weights for the new cost function || $W_d d - W_d L\, W_m m'$ ||^2. The misfits after applying the model weight and data weights are displayed in Figure 4.19c. The misfits after 10 iterations of the new GLI inversion are displayed in Figure 4.19d. First arrival picks rejected by the data weights are shown as missing picks and are shown in white. The RMS of the misfits after 10 iterations is 4.92 msec and is comparable to the RMS of 4.43 msec from the original GLI solution without the imposed first arrival pick errors. The error distributions from minimizing the cost functions || $d - Lm$ ||^2 and || $W_d d - W_d L\, W_m m'$ ||^2 are plotted in Figure 4.20. Similar to the first GLI test errors are reduced after minimizing the cost function || $W_d d - W_d L\, W_m m'$ ||^2. However, with larger smoothed surface-consistent residual statics of 8 msec, improvement to the near-surface velocity and the coherence of the reflection image is noticeable.

We also compare the turning-ray refraction solutions from the two approaches. The results of the turning-ray refraction tomography test are summarized in Figures 4.21a to 4.21e. As shown in the ray density plot (Figure 4.21b), not all cells are covered by ray paths. The minimum non-zero ray density of 600 and maximum ray density of 38000 seem high. However, for a 448-channel recording with 269 shots and 488 receivers, ray coverage immediate below a shot point can be 448 channels plus 448 receivers. If a velocity cell is covered by every ray path, the maximum possible ray density is 269 shots × 448 channels. Therefore, the ray density range of 600 to 38000 is reasonable. Velocity values at cells with no ray coverage cannot be updated and velocity values at cells with insufficient ray coverage can be unreliable. The smoothed surface-consistent reflection residual statics (Figure 4.21c) from gathers corrected with a tomographic solution from the original cost function are in the range of -4.9 to 2.5 msec. We do not see a similar long-wavelength trend at the NE end of the seismic line as observed in the GLI solutions; however, a long-wavelength trend can be observed at a location between 1.0 Km to 2.0 Km from the start of the seismic line. Applying the model weight to the velocity values results in very slow velocity in some velocity cells. To maintain the same time term corrections from the smoothed surface-consistent residual statics, we choose to update the model thickness by stretching the velocity model between the surface and the intermediate datum. The updated velocity model $W_m m$ and the final velocity model after 7 iterations of the new kernel are plotted in Figures 4.21d and 4.21e. The CDP stacks created with the two different solutions are plotted in Figure 4..22. The CDP stack with the new turning-ray refraction tomography solution (Figure 4.22b) shows significant improvement in coherence at a location between 1.0 Km to 2.0 Km from the start of the seismic line. Amplitudes of seismic events at around 1.0 second are slightly weaker on the CDP stack with the new solution. This suggests NMO velocity may have to be revised after the new statics solution. On the CDP stack with the original turning-ray refraction tomography solution (Figure 4.21a), there is a long-wavelength trend dipping from NE to SW. This trend is reduced on the CDP stack with the new solution. In the GLI tests, we found smoothed surface consistent residual statics of 3.3 msec was not sufficient to show noticeable improvement in the near-surface velocity model and reflection image; while smoothed surface consistent residual statics of 8 msec was sufficient to show noticeable improvements. Therefore, with the smoothed surface consistent residual statics of -4.9 to 2.5 msec, we do not believe additional iterations of the new processing flow will result in meaningful changes.

4.7 Discussion

Surface-consistent reflection residual statics derived from correlation of reflection data optimize the stacking response of the reflection data. Using these statics corrections as undetected errors in the near-surface refraction analysis and back projecting these errors to the near-surface velocity model can produce weathering statics corrections that give the same stacking response as applying the surface-consistent reflection residual statics. Using only the smoothed surface-consistent reflection residual statics helps to alter the medium to long-wavelength variations in the near-surface velocity model. Applying the model weight to the velocity can sometimes place the velocity into an unreasonable range. It may be necessary to review the initial updated velocity model to determine the proper combination of model weight for velocity and thickness. We only apply a constant correction to all depth steps below the same surface location; however, this correction serves only as a priori information to guide the refraction inversion toward time delays that agree with the reflection data. With proper selection of the data weight threshold, an optimal solution can be achieved by rejecting outlying picks. Data weight threshold of 1 standard deviation keeps 68% of the data; while data weight threshold of 2 standard deviations keeps 95% of the data. We suggest a data weight threshold between 1.5 and 2. We found the smoothing length of 31 to 51 receiver stations or 300 m to 500m results in a reasonable medium wavelength reflection residual statics for the proposed algorithm. However, these parameters should be tested. The number of non-linear iterations required depends on the quality of the first solution. The first GLI test results in a stable near-surface velocity model and only minor medium wavelength surface-consistent residual statics are derived from the reflection corrections. The GLI solution from the proposed non-linear optimization workflow only shows small differences in the near-surface velocity model and reflection coherence; therefore, more iterations will not result in meaningful improvement. However, for a more challenging seismic survey, improvement in the near-surface velocity model can lead to improved coherence and structural integrity of the reflection image. Therefore, subsequent iterations can potentially further enhance the near-surface velocity model. The second GLI test with imposed picking errors results in cycle skipping in first arrival picks. These errors generate medium wavelength residual statics in the range of -4 to 8 msec. The proposed nonlinear optimization workflow corrects for the errors caused by the imposed first arrival picking errors and results in a noticeable improvement to the near-surface velocity model and the coherence of the reflection image. However, the medium wavelength components derived from the new GLI solution are in the range of -3 to 3 msec. As observed in the first GLI test, a medium wavelength residual of 3 msec will not result in meaningful improvement from additional iteration.

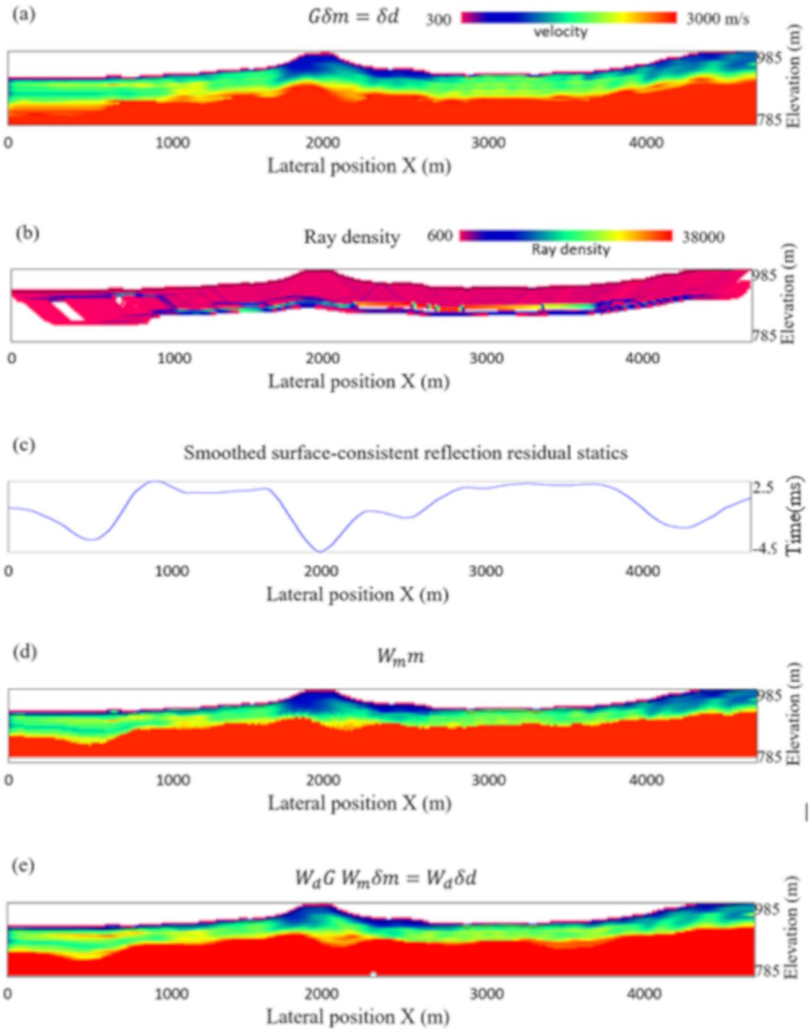


Figure 4. 21. Turning-ray refraction tomography: (a) final model by solving $G\ \delta m = \delta d$, (b) ray density plot shows low coverage at edges of the model, (c) smoothed surface-consistent reflection residual statics, (d) updated model $W_m\ m$, (e) final model by solving $W_d G W_m \delta m = W_d\ \delta d$.

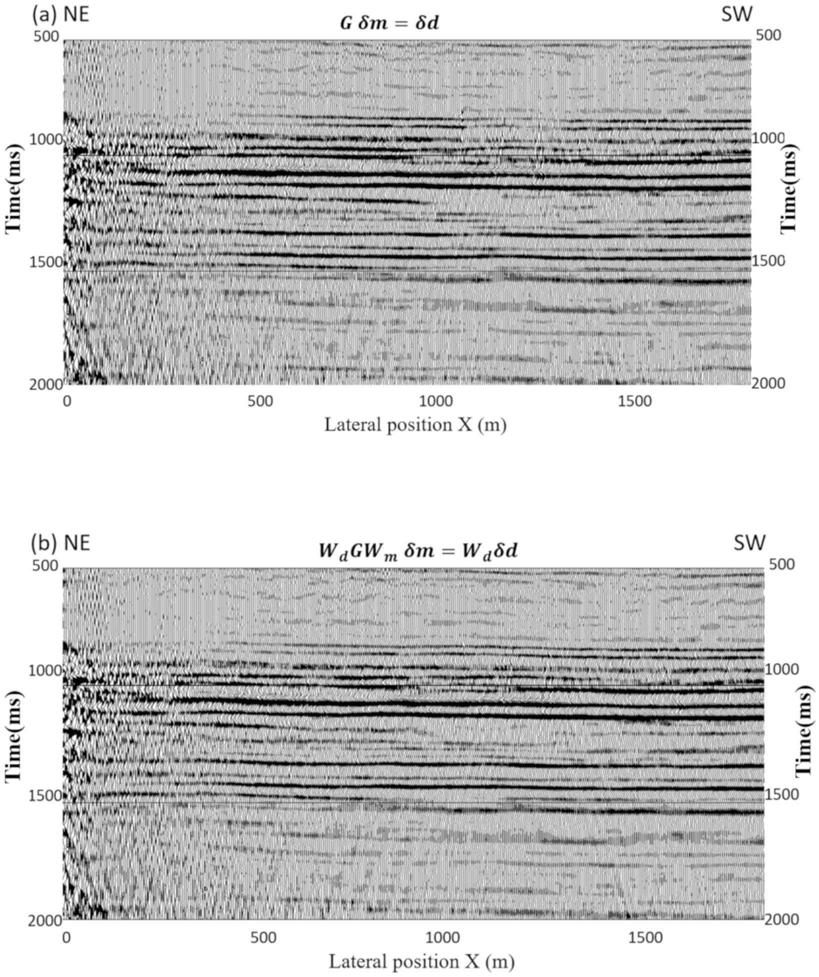


Figure 4. 22. CDP stack section with weathering statics correction using turning-ray refraction tomography velocity model computed from solving (a) the original cost function, (b) $W_d G\ W_m \delta m = W_d \delta d$ with model weight and data weight derived from surface constant reflection residual statics.

4.9 Conclusion

Conventional refraction inversion using first arrival times alone suffers from data errors, numerical errors, and algorithm errors inherent in refraction data and refraction methods. Surface-consistent residual statics using correlation of reflection data can compensate for some of these deficiencies in the near-surface velocity model by maximizing the stack response of the reflection data; however, the near-surface velocity model is left compromised by these errors. These deficiencies in the new surface velocity model tend to accumulate in the deeper reflectors during subsequent reflection velocity model building processes. In contrast to the conventional refraction inversion that uses first arrival times alone, the new nonlinear optimization scheme also uses the surface-consistent reflection residual statics that maximizes the stacking response as a priori information in the refraction inversion. This is implemented by modifying the cost function of refraction inversion to include model weight and data weight. We applied this scheme to GLI and turning-ray refraction tomography methods. Test results from the Hussar 2D dataset confirm that the proposed nonlinear optimization refraction solution workflow is robust and converges to a near-surface velocity model that is harmonized with the surface consistency of the reflection data. However, the significance of the improvement depends on the magnitude of the smoothed surface-consistent residual statics. We found that for the GLI method smoothed surface-consistent residual statics of 8 msec. can produce noticeable improvements in the near-surface velocity model and the coherence of the reflection image. We also found that for the refraction tomography test, smoothed surface consistent residual statics in the range -4.9 to 2.5 msec. can produce noticeable improvements.

Chapter 5

Near-surface velocity model building and statics correction for blended land data

This chapter addresses the problems of blended land data in new-surface and statics correction. It evaluates the problems of blended seismic data acquisition imposed on refraction and reflection arrival times. It proposes a robust refraction arrival separation method that uses amplitude burst suppression and the sparse Radon transform to enhance the first break quality of the blended data. It also demonstrates that after refraction statics correction, we can perform normal moveout velocity analysis and surface consistent residual statics prior to deblending because of the passive separation property of the blended data. I include the following published paper in the chapter (reproduced here with permission from Canadian Society of Exploration Geophysics).

Near-surface velocity model building and statics correction for blended land data
Bernard Law and Daniel Trad, 2021
Published in Canadian Journal of Exploration Geophysics Volume 45, no. 1, spring 2021

Bernard Law[1], *and Daniel Trad*[1]

[1]*University of Calgary, Department of Geoscience, Calgary, Canada*

Abstract: An accurate near-surface velocity model allows proper correction of near-surface variation on land seismic data. This process often uses first arrival time picks of refraction energy. Simultaneous source data acquisition, also called seismic blending, increases the spatial sampling and/or reduces the acquisition cost. However, interfering shots also contaminate the refraction arrivals and the quality of first arrival time picks for land seismic data. We use synthetic and

numerically blended land data to assess the interfering noises from seismic blending. We propose a robust refraction arrival separation method that uses amplitude burst suppression and the sparse Radon transform algorithm to enhance the first break quality of blended data. We also demonstrate that after refraction statics correction, we can perform normal moveout velocity analysis and surface consistent residual statics prior to deblending because of the passive separation property of blended data. Furthermore, we show that the suppression and removal of blending noise can be done using the computational efficient parabolic Radon transform after statics and moveout correction.

5.1 Introduction

Conventional seismic data acquisition deploys sources with a large time delay to minimize interference between sources. With advances in computing capacity and imaging algorithms, seismic data with longer offset, wider azimuth, denser source and receiver spacing have resulted in better seismic images. However, these improvements demand higher data density and increase the acquisition cost. Over the past two decades, acquisition and processing techniques have been developed to increase acquisition efficiency (Beasley et al., 1998; Berkhout, 2008). Beasley et al. (1998) propose a simultaneous source firing operation involving two or more sources firing at the same time. They demonstrate that, with sufficient blended source separation, processes that require correct geometry information, for example, NMO and stack, can suppress blending interference noises directly without direct separation of the blended shots. This happens because interference noises have incorrect geometry information. Another group of deblending methods is denoising-based deblending. Hampson et al. (2008) introduce small random firing-time delays between sources and show that after removing the firing-time delay (a process called “pseudo-deblending”), blending interference noise is coherent in the common shot domain and incoherent in other domains. Processes such as migration and CDP stack can effectively suppress blending interference. Mahdad et al. (2011) separate blended data by iteratively removing blending noise in pseudo-deblended receiver gathers via F-K filtering. Trad et al. (2012) use Apex Shifted Radon transform to separate blended shots in the shot gather domain. Another approach is to treat blending interference as a signal and simultaneously separate the blended shots by inversion (Moore et al. 2008, Wapenaar et al. 2012, Ibrahim and Sacchi 2013, 2015, Abma et al. 2015). Ibrahim and Trad (2020) show that an inversion-based approach results in better deblending than the denoising-based approach. However, inversion-based methods inverse all the shots at the same time, while the denoising-based approach works on one gather at a time. For large 3D, memory and computational requirements can be a challenge for the inversion-based approach because of the need for changing sort order on each iteration.

Blended land seismic data are generally noisier than marine data and surface-related statics has to be addressed at the beginning of the processing steps (Moore et al. 2008, Manning and Ahmad 2013). Manning and Ahmad (2013) show source separation by sparse inversion enhances first break quality. In this study, we use synthetic data and numerically blended land data to assess the effects of seismic blending on first arrival picking and statics computation. We also use sparse Radon transform to remove blending interference in pseudo-deblended receiver gathers.

5.1.1 Deblending vs passive separation

The typical processing flow of blended data is to separate the blended shots by using a denoising-based or inversion-based method and to follow that with the conventional processing flow (Figure 5.1a). An alternate processing flow is to process the blended data without separation by refining processing algorithms to make use of the characteristic of the blended data (Figure 5.1b). Algorithms such as velocity analysis, NMO and migration enhance the signal from primary shot points with proper survey geometry, while suppressing the signal from interfering shots through stacking or summing. For example, NMO aligns reflection energy according to the source to receiver offset and moveout velocity. Stacking of moveout corrected data will enhance the signal from the primary shots that have the correct source to receiver offset and will suppress signal from interfering shots. Migration sums the reflection energy according to the forward modelled ray paths. Similar to NMO stacking, migration will enhance the reflection signal from primary shots and suppress the reflection signal from the interfering shots.

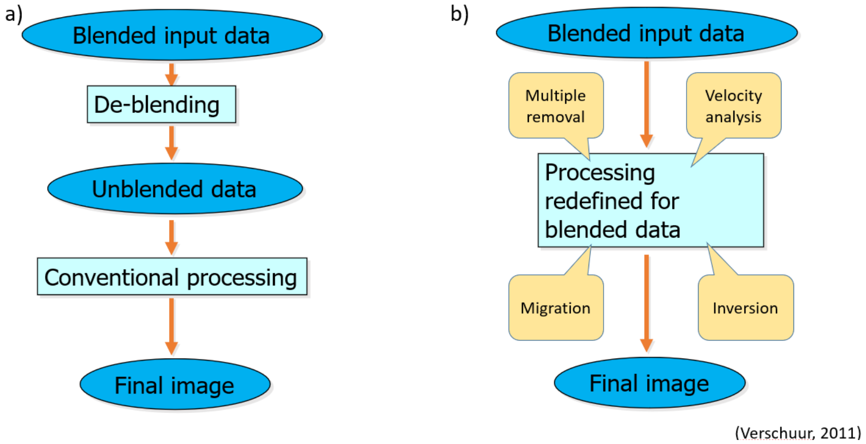


Figure 5. 1. Processing flow of blended data: (a) Deblending followed by conventional processing, (b) Modified processing algorithms using the characteristics of the blended data without direct deblending of blended data (image courtesy of Eric Verschuur).

5.1.2 Statics correction

Denoising-based deblending, inversion-based deblending and passive separation processing methods all require weathering statics and surface consistent residual statics corrections at the front end of the processing flow (Figure 5.2) for land seismic surveys. Weathering statics correction involves refraction signal and surface consistent residual statics involves reflection signal. Both refraction and reflection signals are contaminated by interfering shots in seismic blending. In this study, we use the denoising-based method with pseudo-deblended data to separate the refraction and reflection signal for weathering statics and surface consistent residual statics correction computation.

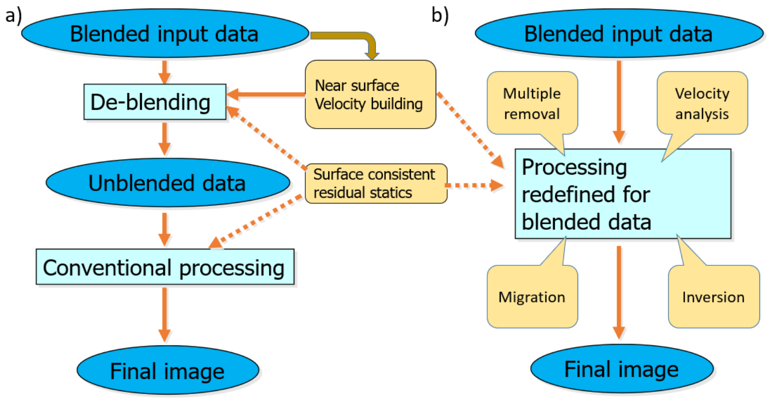


Figure 5. 23. Statics corrections for blended data processing flow.

5.2 Theory

5.21 Seismic blending and de-blending

Seismic source blending can be represented by the source matrix **S**, blending matrix $\mathbf{\Gamma}$, and the blended source matrix $\mathbf{S}_{bl}$:

$$\mathbf{S}_{bl} = \mathbf{S\Gamma}. \qquad (5\text{-}1)$$

For a 9 shot points seismic survey with a blending fold of 3 and regular shot increment, these matrices are:

$$\mathbf{S} = \begin{bmatrix} S_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & S_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & S_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & S_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & S_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_9 \end{bmatrix}, \qquad (5\text{-}2)$$

$$\mathbf{\Gamma} = \begin{bmatrix} e^{-j\omega dt_1} & 0 & 0 \\ 0 & e^{-j\omega dt_2} & 0 \\ 0 & 0 & e^{-j\omega dt_3} \\ e^{-j\omega dt_4} & 0 & 0 \\ 0 & e^{-j\omega dt_5} & 0 \\ 0 & 0 & e^{-j\omega dt_6} \\ e^{-j\omega dt_7} & 0 & 0 \\ 0 & e^{-j\omega dt_8} & 0 \\ 0 & 0 & e^{-j\omega dt_9} \end{bmatrix}, \tag{5-3}$$

$$\mathbf{S_{bl}} = \begin{bmatrix} S_1 e^{-j\omega dt_1} & 0 & 0 \\ 0 & S_2 e^{-j\omega dt_2} & 0 \\ 0 & 0 & S_3 e^{-j\omega dt_3} \\ S_4 e^{-j\omega dt_4} & 0 & 0 \\ 0 & S_5 e^{-j\omega dt_5} & 0 \\ 0 & 0 & S_6 e^{-j\omega dt_6} \\ S_7 e^{-j\omega dt_7} & 0 & 0 \\ 0 & S_7 e^{-j\omega dt_8} & 0 \\ 0 & 0 & S_9 e^{-j\omega dt_9} \end{bmatrix}, \tag{5-4}$$

where $\mathrm{dt_i}$ is the randomized firing-time delay for i^{th} shot point. Each column of $\mathbf{\Gamma}$ contains the blending parameters for a blended source, and each column of $\mathbf{S_{bl}}$ is a blended source. Randomized firing-time delays result in temporal inconsistency. By interchanging rows in the blending matrix $\mathbf{\Gamma}$, additional spatial inconsistency can be added to the seismic blending.

For blended seismic data $\mathbf{P}_{bl}$, it can be represented by the blending matrix $\mathbf{\Gamma}$ and the unblended data $\mathbf{P}$ as:

$$\mathbf{P}_{bl} = \mathbf{P\Gamma}. \tag{5-5}$$

Because $\mathbf{\Gamma}$ is not a square matrix, it is not invertible. Equation (5-5) cannot be directly inverted to get $\mathbf{P}$. $\mathbf{P}$ is instead computed by minimizing the following objective function :

$$J = \| \mathbf{P}_{bl} - \mathbf{P\Gamma} \|^2 + \lambda \| \mathbf{P} \|^2 , \tag{5-6}$$

where λ is a tradeoff constant, $\boldsymbol{\Gamma}^{\mathrm{H}}$ is the conjugate transpose of $\boldsymbol{\Gamma}$, and the least-squares solution to (6) is:

$$\mathbf{P} = \mathbf{P}_{bl}\,\boldsymbol{\Gamma}^{\mathrm{H}}(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{H}} + \lambda\mathrm{I}\,)^{-1}. \tag{5-7}$$

The term $\mathbf{P}_{bl}\,\boldsymbol{\Gamma}^{\mathrm{H}}$, also referred to as pseudo-deblending (Mahdad et al., 2011), expands the blended data $\mathbf{P}_{bl}$ into the number of sources that would be obtained without blending and corrects for the fire-time delay for each shot record within a blended shot. After pseudo-deblending, the signal appears coherent, while blending noises appear incoherent in receiver and CDP domains (Figure 5.4). This property leads to the removal of blending noises in the pseudo-deblended receiver and CDP gathers. Algorithms that use this property include FK transform (Madhad et al., 2011, Abma et al., 2015) and sparse Radon transform (Moore et al.,2008; Ibrahim and Sacchi., 2013).

5.2.2 Sparse Radon transform

The Radon transform (RT) maps data $u(t,h)$, in time t, distance h space to Radon space $U(\tau,p)$, in zero offset time τ and slowness p, according to the basis function $T(\tau,p,h)$:

$$U(\tau,p) = \iint d(t,h)\delta\big(t - T(\tau,p,h)\big)dtdh. \tag{5-8}$$

Linear, hyperbolic and parabolic basis functions are often used in seismic data processing for coherent and non-coherent noise attenuation :

Linear: $T(\tau,p,h) = \tau + ph.$ (5-9)

Hyperbolic: $T(\tau,p,h) = \tau + \sqrt{\tau^2 + p^2h^2}.$ (5-10)

Parabolic: $T(\tau,q,h) = \tau + q\,h^2,$ (5-11)

where q is not the slowness, but simply a coefficient in the transform (Hampson 1986). q is also referred to as curvature.

Apex shifted (AS) RT (Trad et al., 2012) incorporates apex h_0 of each source of blended shot record:

AS Hyperbolic: $T(\tau, p, h) = \tau + \sqrt{\tau^2 + p^2(h - h_0)^2}$. (5-12)

Equation (5-8) is used directly in velocity stacks for moveout velocity analysis. However, slope/curvature/velocity filtering requires an inversion process. Thorson and Claerbout (1985) use stochastic inversion in time domain hyperbolic RT. Hampson (1986) proposes fast frequency domain parabolic RT, that minimize the cost function:

$$\mathbf{J} = \| \mathbf{L}\,\mathbf{u} - \mathbf{d} \|^2 , \tag{5-13}$$

with the solution:

$$\mathbf{u} = (\mathbf{L}^{\mathbf{T}}\,\mathbf{L})^{-\mathbf{1}}\ \mathbf{L}^{\mathbf{T}}\,\mathbf{d} , \tag{5-14}$$

where L is the forward radon modelling operator for frequency $\boldsymbol{\omega}$, offset h_i and q_k:

$$L_{i,k} = e^{-\omega q_k h_i^2} \tag{5-15}$$

Ng and Perz (2004) outline a time-domain Gauss-Seidel iteration algorithm with sparseness constraints in p and tau direction using thresholding as well as prioritizing the computation sequence of the q traces according to their energy from the previous iteration.

5.3 Synthetic Example

The first arrival picking on blended data has not been widely addressed in the literature. To assess the impact of blending noise on the first arrival and reflection energy we generate a finite-difference synthetic dataset using a layered wedge model and acquisition geometry with a receiver and shot spacing of 24 m. We use a 5 shots pattern acquired simultaneously with randomized firing-time delays (Figure 5.3). The first shot in each pattern has zero firing delay, while the other 4 shots have randomized firing delays between 50 and 200 ms. Figure 5.4 shows the pseudo-deblended shot, receiver and CDP gather. Blending noise appears coherent in the shot gather; while it appears incoherent in receiver and CDP gather. The amplitudes of the first arrival energy from interfering shots are stronger than the reflection signal. We use sparse linear Radon transform (Ng and Perz, 2004) to model the first arrival energy on windowed pseudo-deblended receiver gathers (Figure 5.5a). The windowed pseudo-deblended received gather is transformed to p and tau domain. Sparseness is introduced by computing the p traces in the descending order of its energy level from the previous iteration and by rejecting p traces with low semblance value. Hence, the strongest event will be modelled first, and weak events will be omitted. The modelled first arrival energy is then constructed by the forward transform of the sparse p traces to the distance-time domain. The p range is 0.5 ms/m to 1 ms/m, and the semblance threshold is 0.0001 for this test. Figures 5.5b and 5.5c display the receiver and CDP gather with modelled first arrival energy removed. Strong residual first arrival energy can be seen near the sources; however, there are significant improvements in the reflection arrival between 1 and 1.5 seconds. Figures 5.6 and 5.7 compare the normal moveout (NMO) velocity analysis using CDP gather with and without the removal of modelled first arrival energy. Although the reflection event at 1100 ms is contaminated by strong first arrival from interfering shots (Figure 5.6a), the velocity semblance scan (Figure 5.6b) can focus at the right velocity because of the passive separation property of the NMO and stacking process that suppresses the blending interference with an incorrect source to receiver offset. After NMO correction, seismic reflections are better aligned, and some blending interference appears as large-amplitude bursts (Figure 5.7c). We use a 200 ms time-varying median filter to lower the amplitude level of the amplitude bursts in NMO corrected pseudo-deblended CDP gathers(Figure 5.8a) that is greater than 2 times the median amplitude level. However, the interferences are still significant and appear coherent on shot gathers (Figure 5.8c). Using the coherence of the moveout corrected seismic reflection, we perform sparse parabolic RT on the moveout corrected and amplitude burst suppressed CDP gather (Figure 5.9a). Figure 5.9b and 5.9c show the moveout restored CDP and shot gathers after sparse parabolic RT.

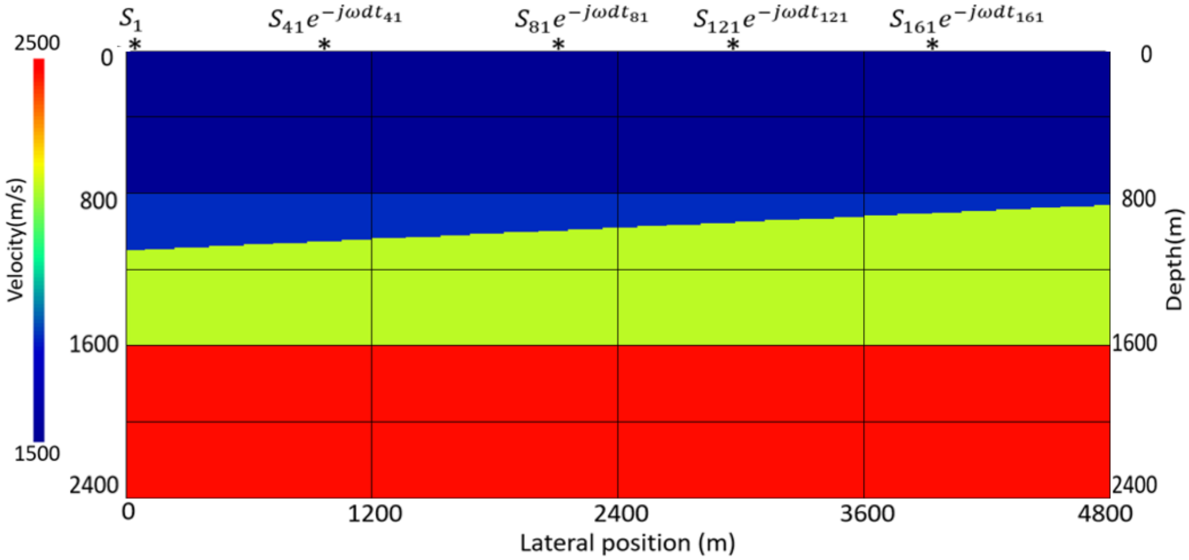


Figure 5. 3. Velocity model and 5 shots blending design pattern. The firing-time delay for this example is zero for the first source of each blended shot.

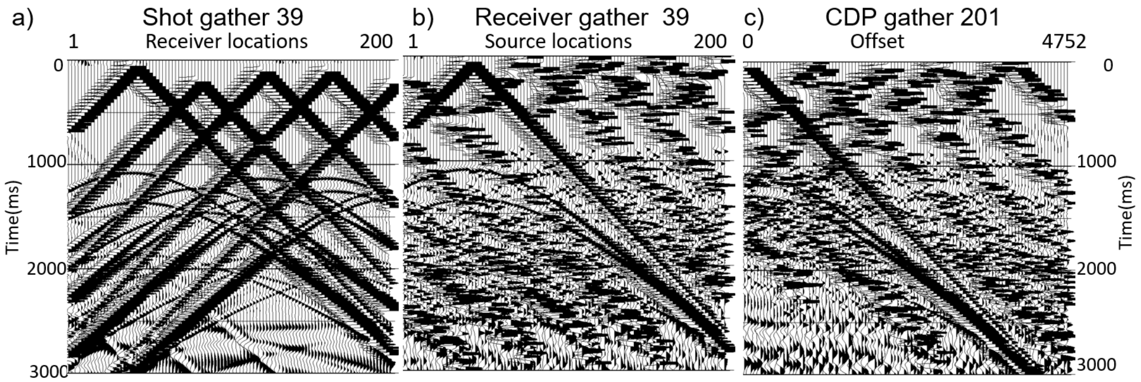


Figure 5. 4. (a) Pseudo-deblended shot gather showing coherent blending interference, (b) pseudo-deblended receiver gather and (c) pseudo-deblended CDP gather showing incoherent blending interference.

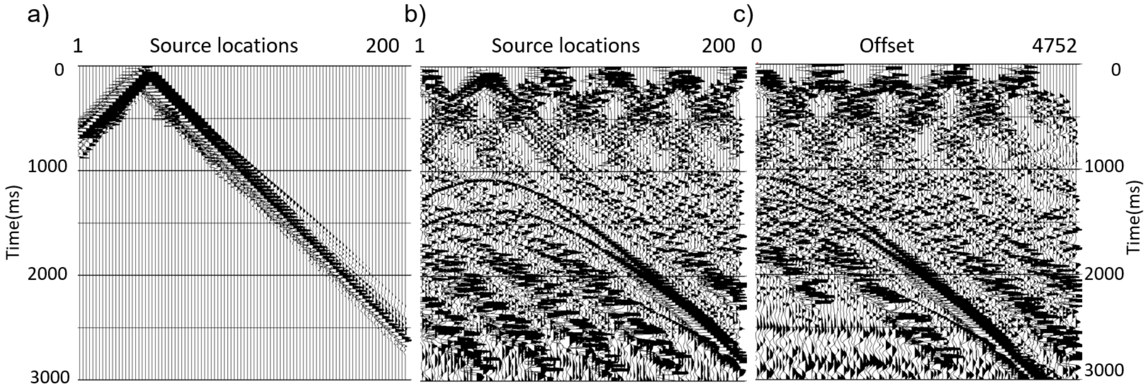


Figure 5. 5. (a) Refraction arrival modelled with sparse linear Radon transform of pseudo-deblended receiver gather 39, (b) pseudo-deblended receiver gather and (c) pseudo-deblended CDP gather after removal of all modelled refraction arrival.

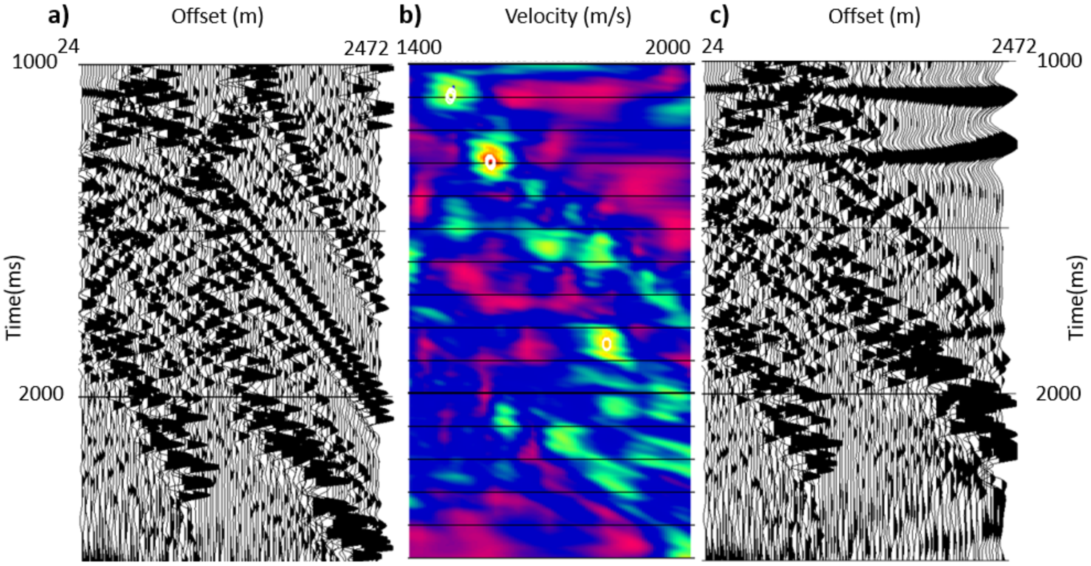


Figure 5. 6. Pseudo-deblended CDP gather 201, (b) NMO velocity analysis, (c) moveout corrected CDP gather.

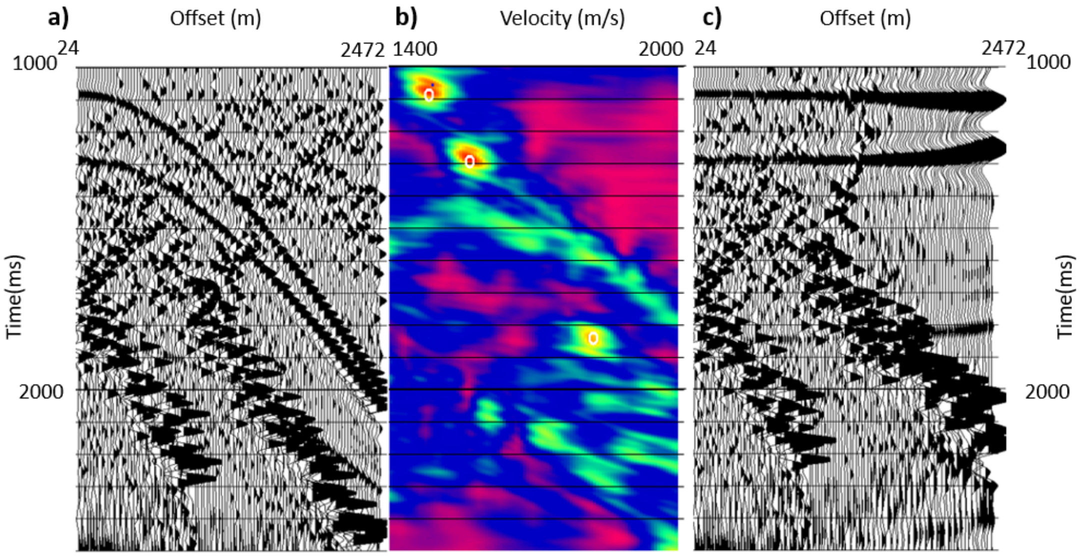


Figure 5. 7. (a) Pseudo-deblended CDP gather 201 after removal of modelled refraction arrival, (b) NMO velocity analysis, (c) moveout corrected CDP gather.

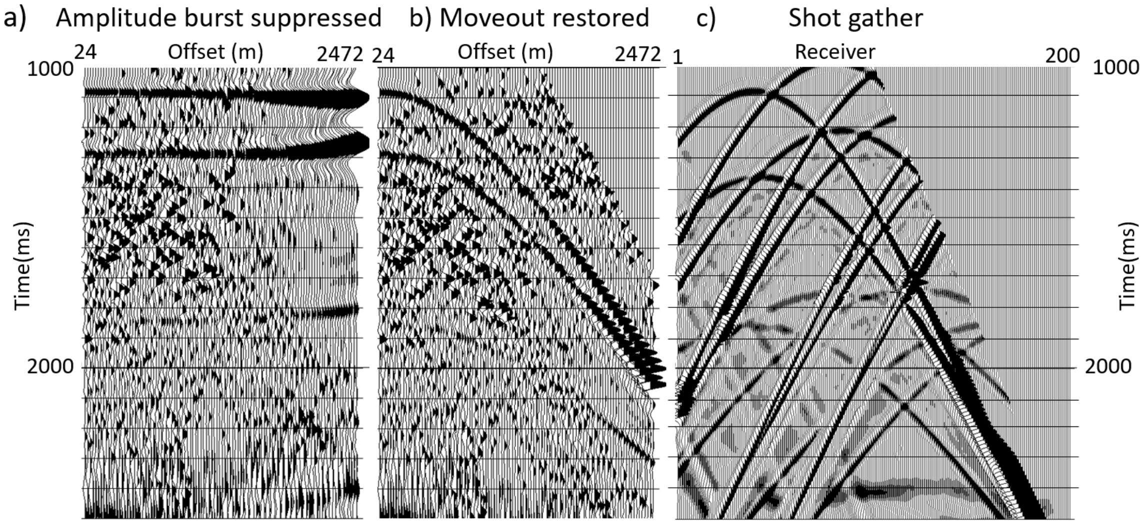


Figure 5. 8. (a) Pseudo-deblended CDP after moveout correction and amplitude burst suppression, (b) moveout restored, (c) shot gather after amplitude burst suppression in CDP gathers.

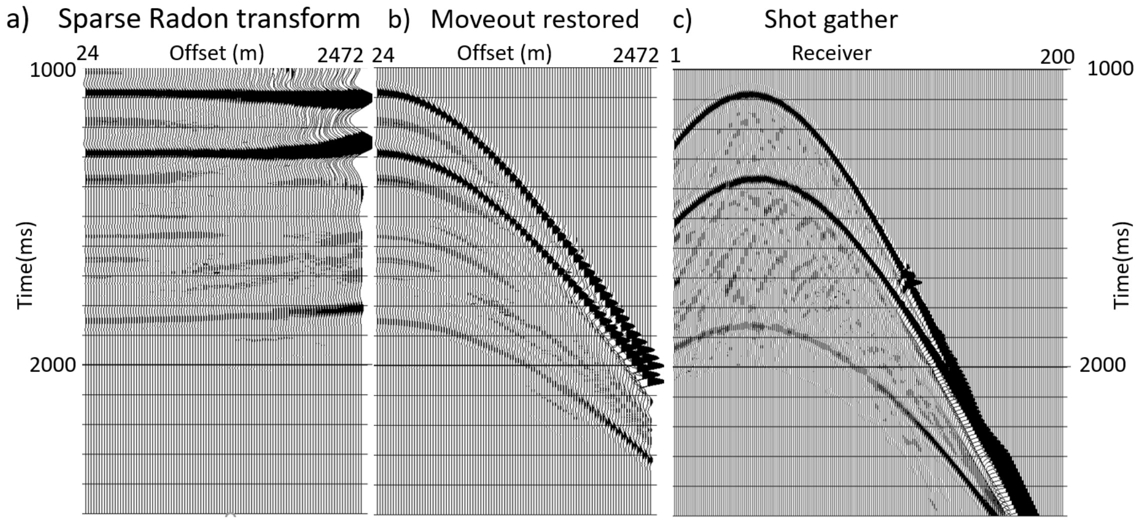


Figure 5. 9. (a) Pseudo-deblended CDP after moveout correction, amplitude burst suppression and sparse parabolic Radon transform (b) moveout restored, (c) shot gather after amplitude burst suppression and sparse parabolic Radon transform in CDP gathers.

These test results demonstrate that the denoising-based method using sparse linear Radon transform can effectively model and separate first arrival energy on blended data. This is computationally more efficient than the inversion-based approach. Separating first arrival energy on blended land data is important for near-surface velocity modelling using the first arrival time picks or waveform. These tests also show that moveout velocity analysis can be performed on pseudo-deblended CDP gather before separation because of the passive separation property of the NMO and stacking process. Furthermore, sparse parabolic Radon transform on moveout corrected data can be an effective denoising-based deblending method.

5.4 Numerically blended 2d land data example

We use numerically blended 2D land data to test the effectiveness of the first arrival separation and its significance on near-surface velocity model building and statics correction. The 3-shots blending pattern has zero firing delay time for the first shot and a randomized firing delay time that ranges from 1000 to 1500 ms. The data used in this example are the vertical component of the dynamite shot records from a 4.5 Km 2D 3C survey acquired at Hussar, Alberta in September 2011 (Margrave et al., 2012). The 2D line runs NE-SW with a topographic relief of about 80 m (Figure 5.10). Examination of blended shot records (Figure 5.11) shows surface elevation effects and blending interference from the ground roll and coherent energy of interfering shot points. We address the surface elevation effects by applying a static time shift that corrects the surface elevation to a smooth surface. We also low-cut filter the blended record to suppress the ground roll interference on the first arrival energy (Figure 5.12). In the pseudo-deblended receiver gather, refraction and reflection arrivals appear as a coherent signal; while blending interference appears as large amplitude burst and incoherent noise (Figure 5.13).

We apply low-cut and median filters to suppress the ground roll and large amplitude burst, and linear sparse Radon transform to suppress the incoherent blending noise in the pseudo-deblended receiver gathers over a time window that includes the first arrival energy (Figure 5.14). The processed receiver gathers are sorted to shot domain for first arrival picking (Figure 5.15). The process does not completely remove the blending noise; however, it improves the first arrival energy for the first break picking. Figures 5.15b and 5.15c display the processed first arrival energy and the original non-blended seismic shot record with the first arrival time picks from the processed pseudo-deblended data displayed in blue.

We use the first arrival time picks to compute a new surface velocity model using refraction tomography (Figure 5.16). The refraction tomography starts with a layered-based GLI refraction inversion with the top layer velocity fixed at 1000 m/s. Refraction tomography will then update the velocity model according to ray path geometry. There is also a lower velocity limit of 1000 m/s used in the refraction tomography solution. Figures 5.17 and 5.18 compare the CDP stack with datum elevation correction only with the CDP stack with weathering statics correction. The latter shows an overall significant improvement in both coherence and structural integrity. However, there are strong blending noises on the CDP stack. Moveout velocity functions for the CDP stacks are determined from weathering statics corrected CDP gathers (Figure 5.19). We also computed the surface consistent residual statics using the weathering statics and moveout corrected CDP gathers. CDP gathers after surface consistent residual statics (Figure 5.20b) shows significant improvement in the alignment of reflection energy. We use the smoothed surface consistent residual statics to update the refraction tomography model (Law and Trad, 2018). Figure 5.21 displays the updated refraction tomography model and the CDP stack with the updated weathering statics. Significant improvement in the lateral coherence can be seen when compared to the CDP stack with weathering statics computed from the original refraction tomography solution (Figure 5.18). Figure 5.22 summarizes the workflow for refraction tomography and surface-consistent residual statics for blended land data.

The CDP stack after weathering and surface-consistent residual statics still shows blending interference. We use a median filter and parabolic sparse Radon transform on statics and moveout corrected receiver gathers (Figure 5.23a) to remove the remaining coherent blending noises. Figure 5.23b displays the processed receiver gather. Figure 5.24a displays the CDP stack after the removal of blending noises in receiver gathers. It shows the strong blending noises have been removed. Figure 5.24b displays the CDP stack of the original nonblended data with statics correction and it shows a good correlation with the CDP stack from the denoising-based deblending workflow. However, some blending noises still noticeable, especially at the edges of the CDP stack. Figure 5.25a and 5.25b compare the statics corrected shot gather 373 before and after the removal of blending noises. Figure 5.25c shows the differences between the blended shot gather and the deblended shot gather. The refraction arrivals at the top of the gather are lost during the NMO and UNMO processes; therefore, they dominate the difference display. Figure 5.25d shows the original non-blended shot gather, it compares well with the deblended shot gather in figure 5.25b. However, it also shows leakage from the refraction arrivals from the blended shots. This study demonstrates parabolic sparse Radon transform with statics and moveout corrected pseudo-deblended receiver gather as an effective denoising-based deblending method.

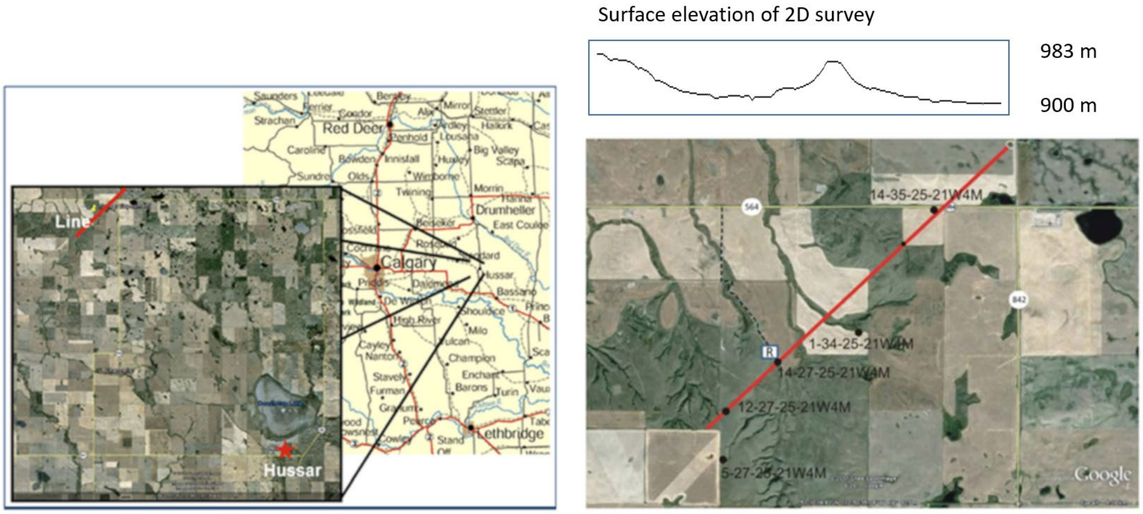


Figure 5. 10. Hussar 2D broadband experiment location map and seismic line layout (Margrave et al, 2012).

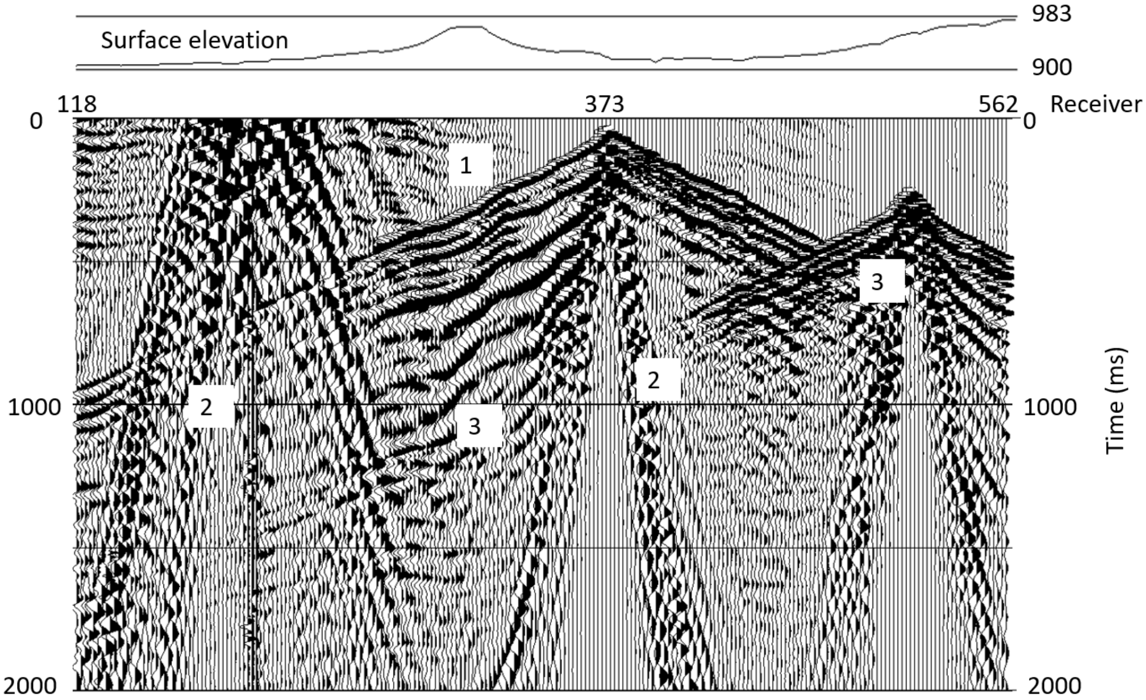


Figure 5. 11. Numerically blended shot record showing (1) surface elevation effect, (2) ground roll interference and (3) coherent first arrival interference.

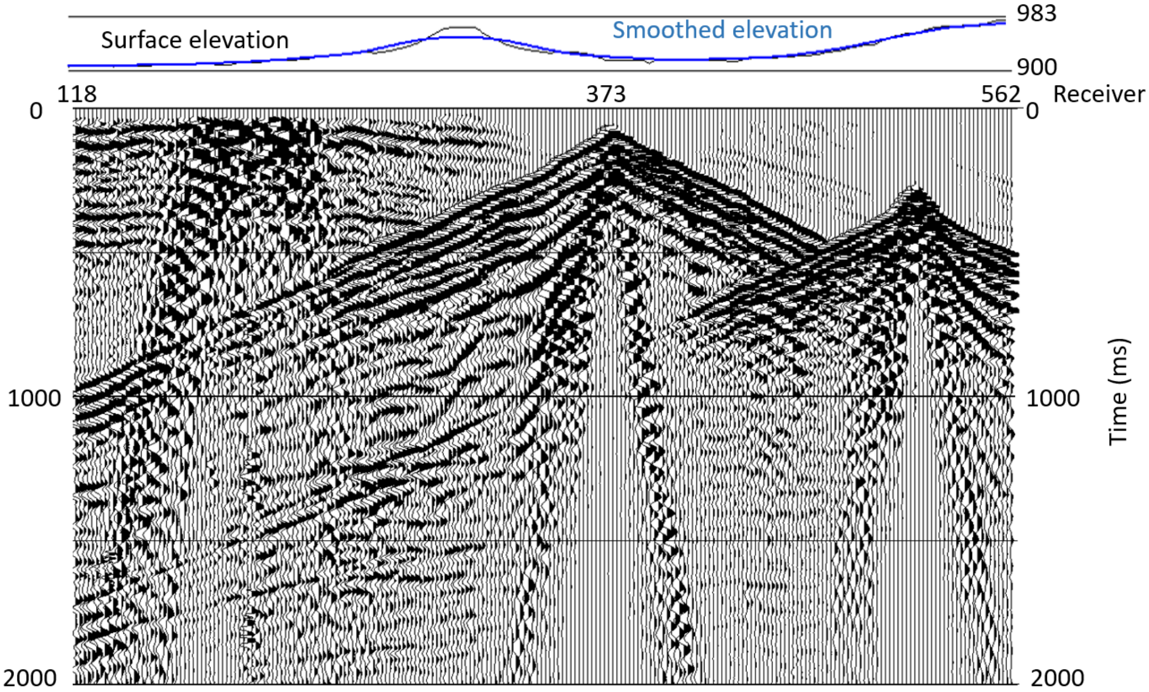


Figure 5. 12. Numerically blended shot record after smooth elevation time adjustments and low-cut filter.

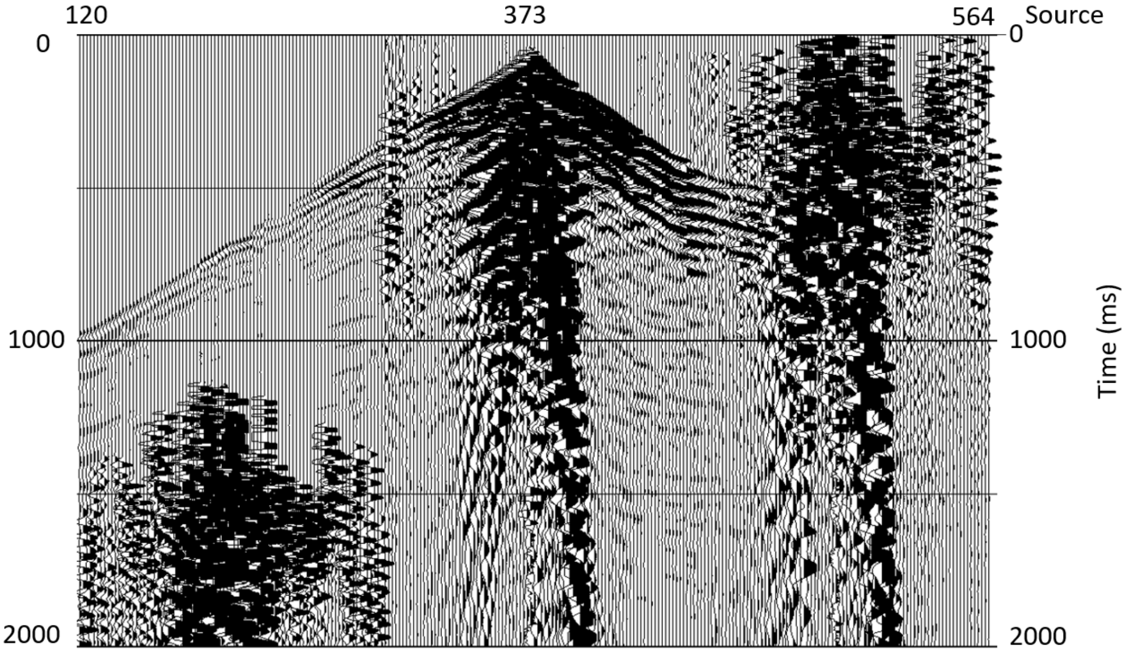


Figure 5. 13. Pseudo-deblended receiver gather.

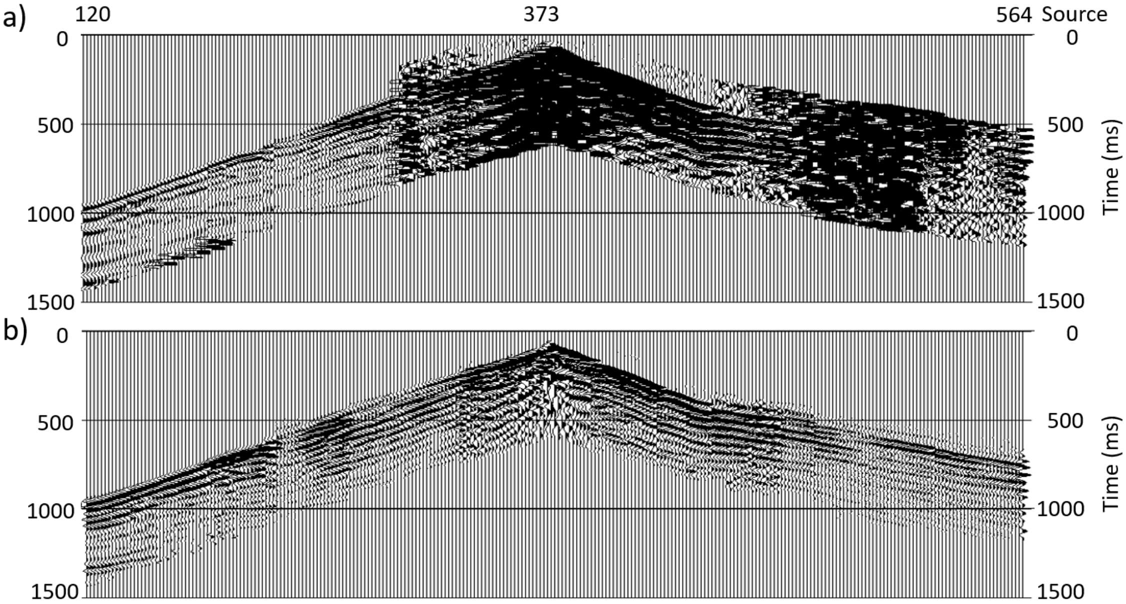


Figure 5. 14. Pseudo-deblended and surface elevation corrected receiver gather, b) after the low-cut filter, amplitude burst suppression and sparse linear Radon transform.

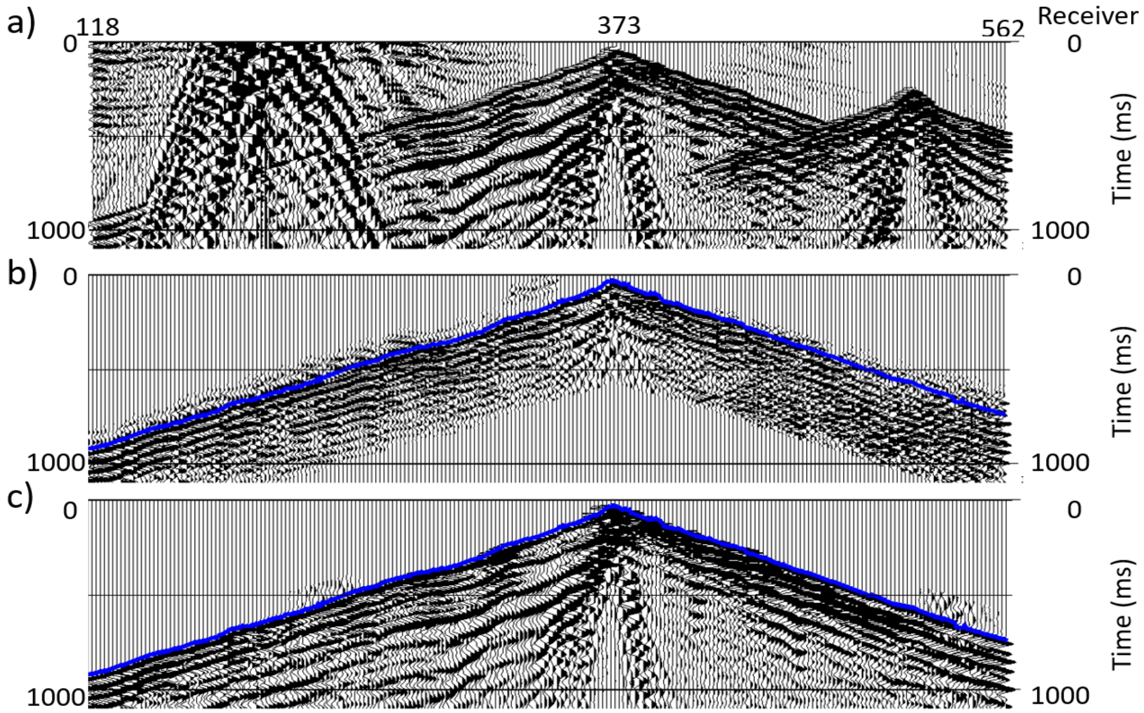


Figure 5. 15. Numerically blended shot gather, b) windowed, low-cut filter, amplitude burst suppression and sparse linear Radon transform in pseudo-deblended receiver domain, c) original non-blended shot gather. (First arrival time picks from processed pseudo-deblended data are displayed in blue)

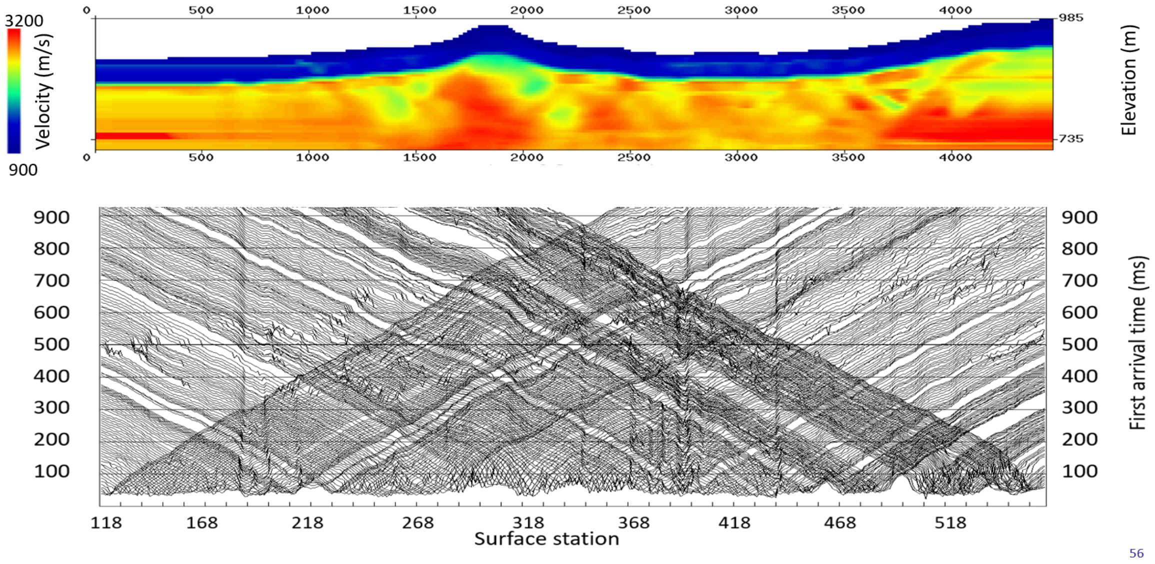


Figure 5. 16. First arrival time picks from deblended data and refraction tomography solution.

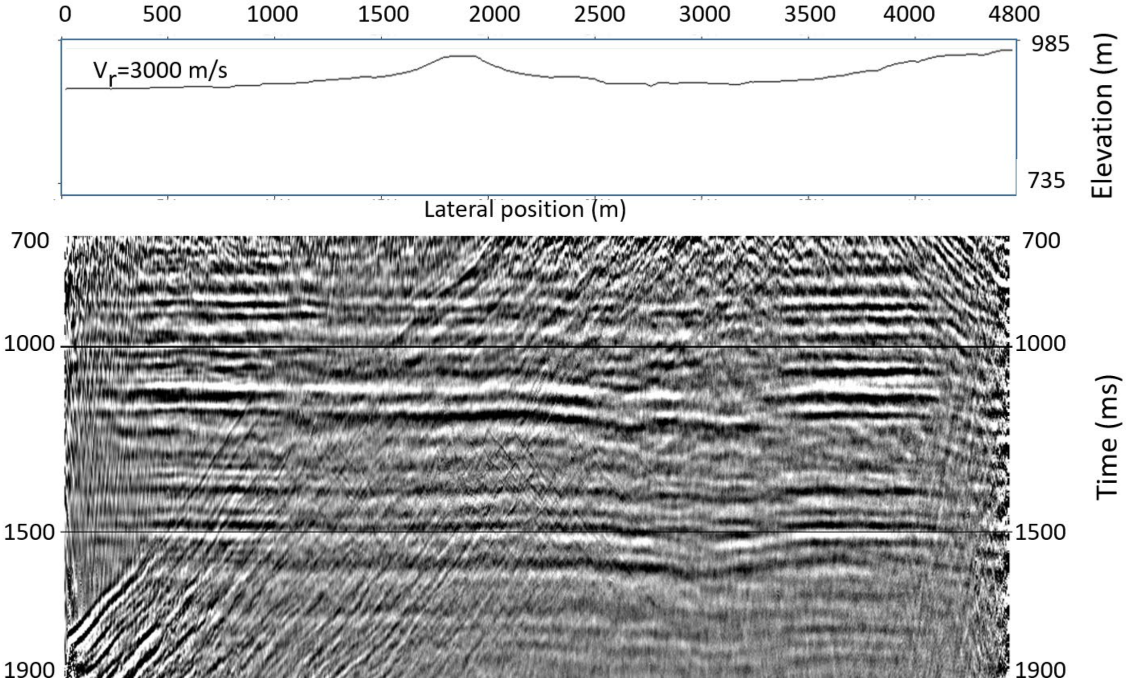


Figure 5. 17. CDP stack with datum elevation correction only.

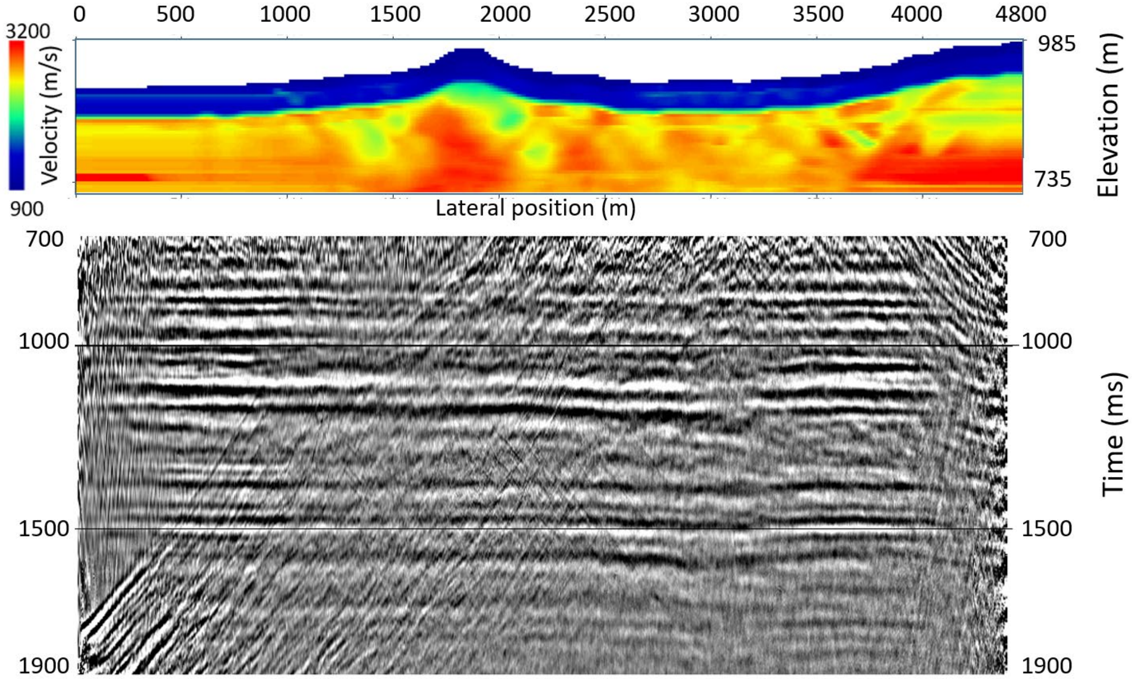


Figure 5. 18. CDP stack with datum elevation and weathering statics correction.

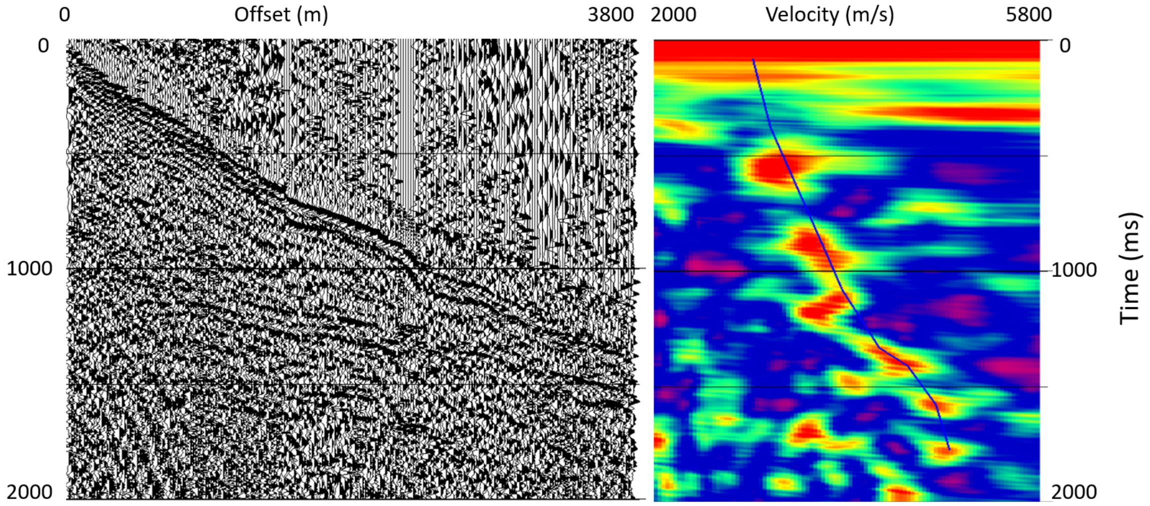


Figure 5. 19. Pseudo-deblended CDP gather with weathering statics correction and NMO velocity semblance.

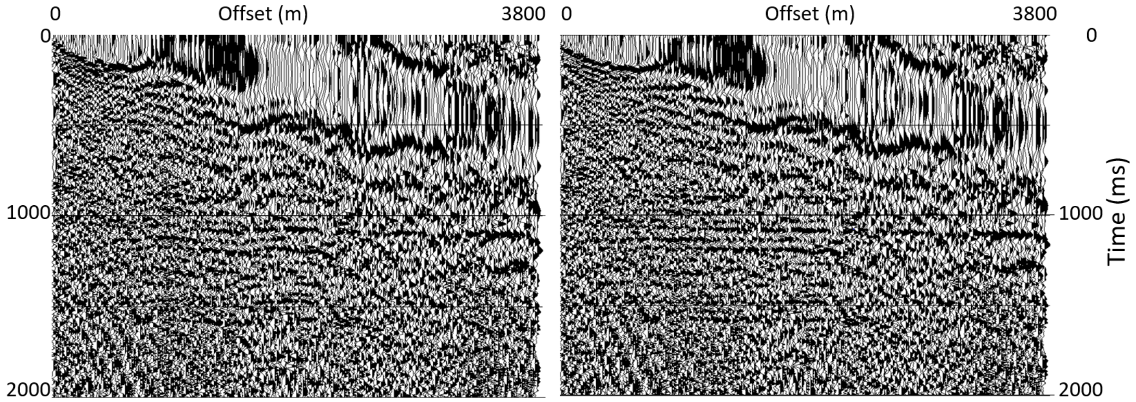


Figure 5. 20. a) Pseudo-deblended CDP gather with NMO correction, (b) pseudo-deblended CDP gather with NMO correction and surface consistent residual statics correction.

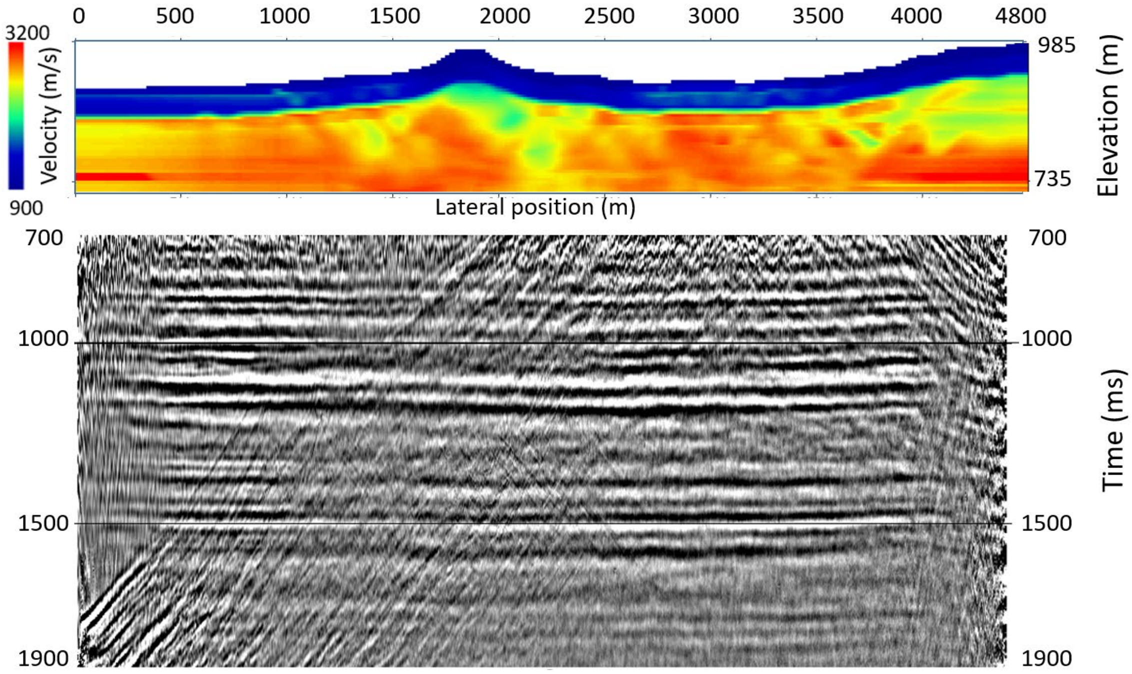


Figure 5. 21. CDP stack with weathering statics computed from model updated with reflection residual statics.

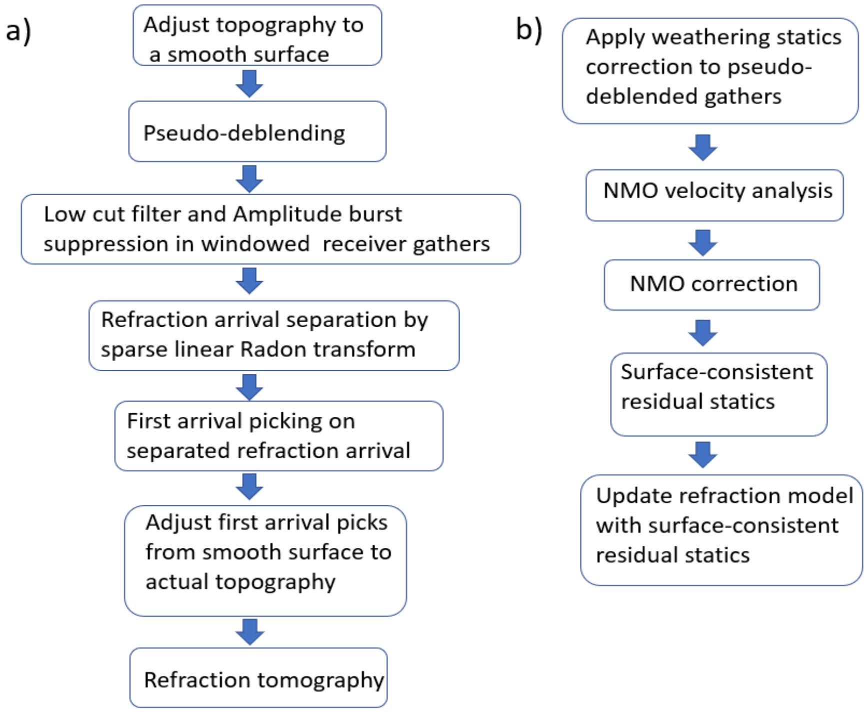


Figure 5. 22. a) Refraction tomography workflow for blended land data, b) Surface-consistent residual statics workflow for blended land data.

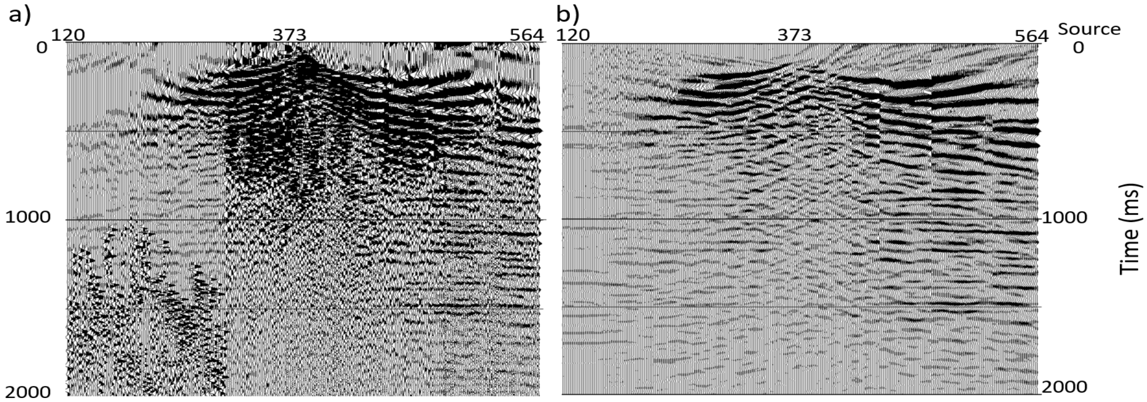


Figure 5. 23. a) Pseudo-deblended receiver gather with weathering statics computed from model updated with reflection residual statics, b) after the median filter and sparse parabolic radon transform.

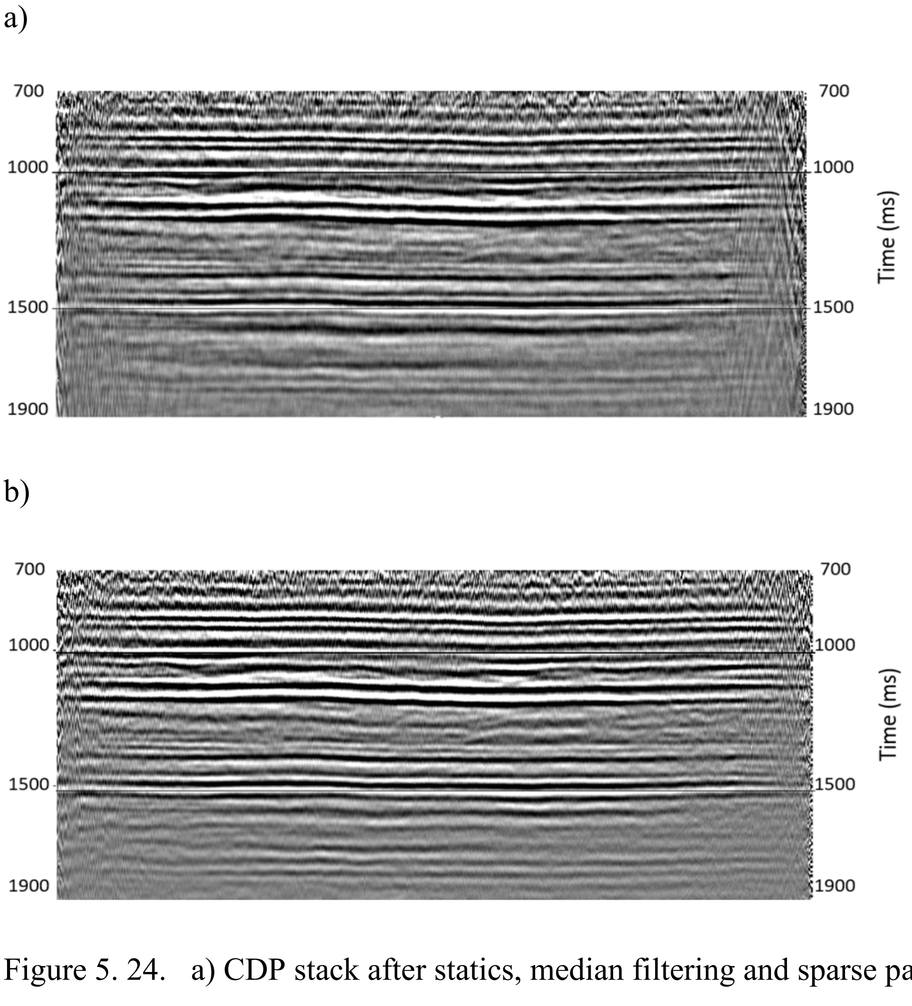


Figure 5. 24. a) CDP stack after statics, median filtering and sparse parabolic Radon transform in receiver gathers, b) CDP stack of original non-blended data with statics corrections.

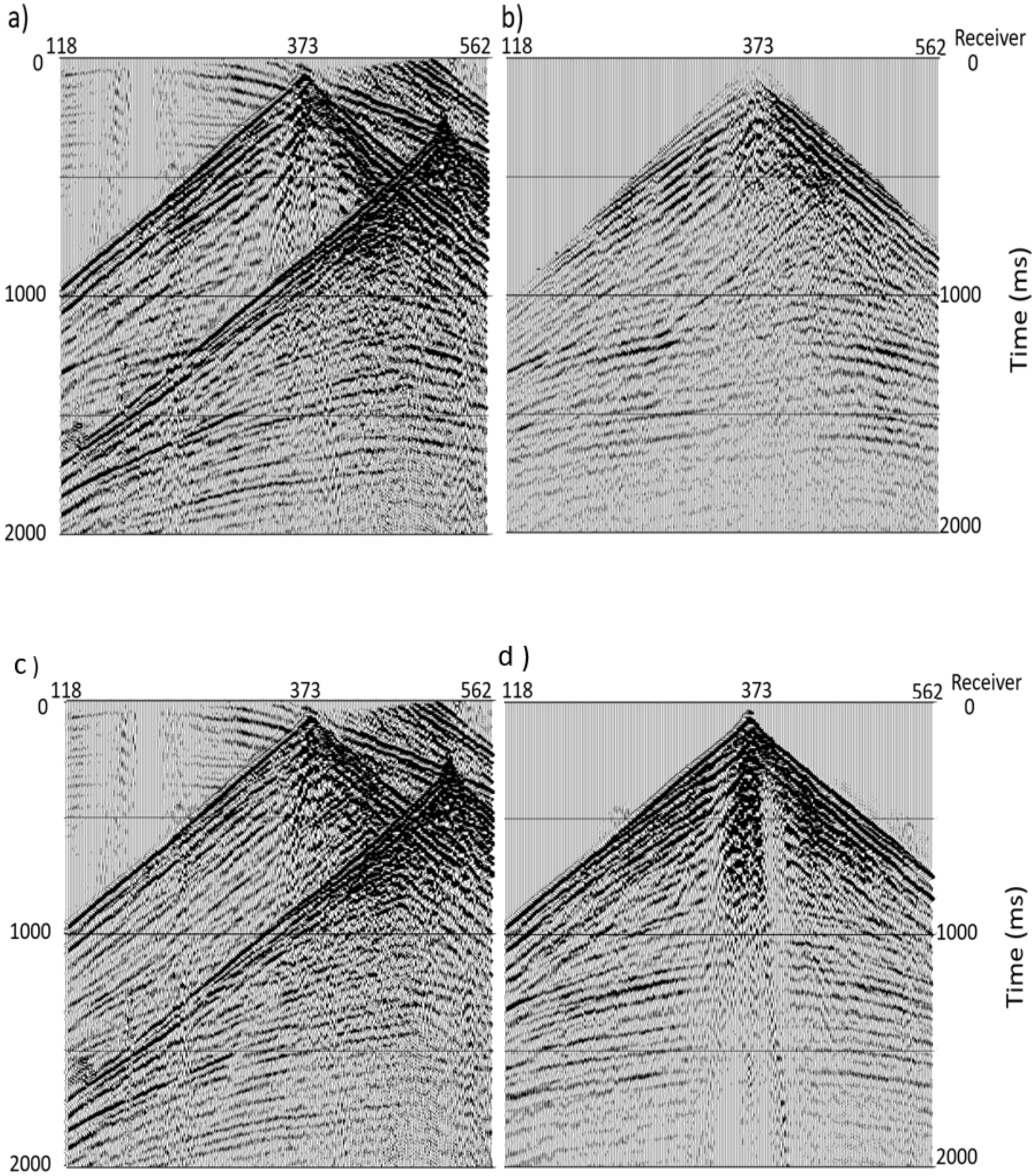


Figure 5. 25. a) Pseudo-deblended shot gather with statics correction, b) shot gather after sparse parabolic Radon transform on statics and moveout corrected pseudo-deblended receiver gathers. c) (a)-(b), d) original non-blended shot gather with statics correction.

5.5 Conclusion

Seismic blended interference degrades the quality of refraction arrival that is critical for near-surface velocity building and statics correction for blended land data. We have shown that linear sparse Radon transform can effectively model refraction arrival energy on blended data. For blended land data, it is also necessary to suppress ground roll and large amplitude bursts using low-cut and median filters. We also showed that normal moveout velocity analysis can be performed on weathering statics corrected blended data prior to separation. Subsequent surface consistent residual statics can then be computed on weathering statics and moveout corrected data. We also used the smoothed surface constant residual statics computed from the correlation of reflection energy of the blended data to update the near-surface velocity model. Furthermore, we have demonstrated that a parabolic sparse Radon transform on statics and moveout corrected pseudo-deblended receiver gather is an effective deblending method.

5.6 Acknowledgements

We are grateful to CSEG and the Canadian Journal of Seismic Exploration editor for granting us permission to reproduce this paper here. We thank Dr. Sam Gray for insightful discussions and suggestions. The authors also like to thank CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 46119-13 and CRDPJ 543578-19 funding this work.

Chapter 6

Application of stereo-tomography reflection tomographic methods to the Hussar 2D survey

Reflection tomography uses reflection arrival times to estimate the subsurface velocity. Using deeper reflection arrival times, reflection tomography can provide velocity information of deeper depth than the previously discussed refraction tomography. Stereotomography is one of the reflection tomography methods. Unlike classical reflection tomography, stereo-tomography does not require picking of continuous refection events and it uses additional information of the apparent slopes of local coherent events in the common shot and receiver gathers.

6.1 Introduction

Stereotomography belongs to the family of slope tomography methods (Sword, 1987). Slope tomography characterizes each reflection ray path with its two-way traveltimes and apparent slopes or ray parameters of the reflection event on the corresponding shot and receiver gathers (Figure 6.1). In a shot gather the apparent slope of a reflection event recorded at a geophone represents the ray parameter P_g of the wavefield at that location, and it determines the ray path between the reflection point and the geophone. Similarly, in a geophone gather the apparent slope of the same reflection event represents the ray parameters P_s of the shot, and it determines the ray path between the reflection point and the shot. Therefore, it is necessary to evaluate both the shot gather and receiver gather of a reflection event in slope tomography to establish the unique shot and geophone ray pairs. Hence, the name stereo-tomography was used by Billette and Lambaré

(1998) to emphasize the shot and receiver ray segment pair for each localized reflection event. Figure 6.1a shows traces around the shot location $\boldsymbol{S}$ and geophone location R at the two-way traveltime T_{sr}. If the velocity of media is known, the shot and receiver ray segments can be reconstructed with P_s, P_g and T_{sr} (Figure 6.1b). The ray parameters P_s and P_g can be picked by tracking the reflection events or by automatic picking using the semblance of the localized shot and receiver slant stacks.

The advantages of slope tomography over classical reflection tomography (Bishop et al., 1985; Chiu and Stewart, 1987) include the additional data measurements of shot and receiver ray parameters and the elimination of the requirement to pick continuous reflection events on pre-stack data; hence, making automated picking easier. Sword (1987) developed the first slope tomography method, also called CDR (Rieber 1936; Riabinkin 1957) tomography. This method reconstructs the shot and receiver ray segments by shooting rays from shot and receiver at the surface using the picked P_s , P_g and ending the ray tracing when the sum of the traveltime of the shot ray segment and receiver ray segment equals the two-way traveltime, T_{sr}. The velocity of the media V is estimated by minimizing the position errors X_{err} of the endpoints of the ray segments (Figure 6.2a). However, this method is sensitive to the picking errors and can be unstable because the accuracy of the forward modelling depends greatly on the picked P_s and P_g. Stereotomography (Billette et. al, 1998, 2003) remedied this instability using the generalized formulation of the slope tomography method. The forward modelling of stereo-tomography involves ray tracing from a scatter point X toward the S and R at the surface and is independent of P_s , P_g and T_{sr}. Therefore, it is independent of picked data and remedies the instability of the original slope tomography method. However, besides the media velocity, this approach also requires the estimations of the scatter position and the ray path geometry parameters for each local reflection event. This results in a more complex multi-parameter inversion problem (Figure 6-2b). The model space of stereo-tomography includes V, X, the shooting angle θ_s and traveltime T_s for the shot ray segment, and the shooting angle θ_g and traveltime T_r for the receiver ray segment. The data space includes S, R, P_s, P_g and T_{sr}. Picking of reflection arrival times and slopes are flexible in stereo-tomography and can be based on the semblance of the shot and geophone slant stacks. However, for noisy data and in areas of complex structure, picking can still be a challenge. We will validate the accuracy of the stereo-tomography method using synthetic data created with a wedge model and the Marmousi model (Brougouis at all, 1990). We will also apply the stereo-tomography method to a 2D land data set acquired in the Hussar area of Alberta to demonstrate the data preparation, the picking procedure and the quality of the Stereotomography solution.

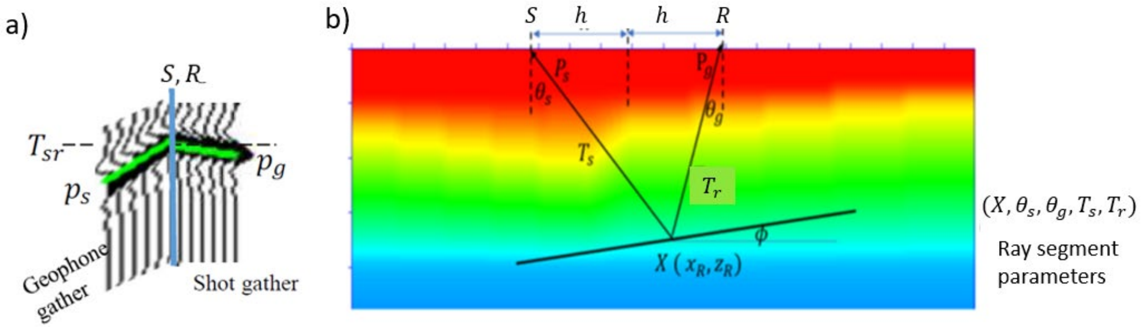


Figure 6. 1 (a) Relationship between $T_{sr}P_s, P_g$ of a localized coherent event. (b) The event is characterized by the traveltime T_{sr} and the ray parameters p_s and p_g and is associated with a ray segment pair in the velocity model. Reflector dip ϕ and ray segment parameters including the scatter point location $\mathbf{X}$, ray shooting angles θ_s and θ_g can be estimated from the half-offset h, the ray parameters, and two-way traveltime T_{sr}.

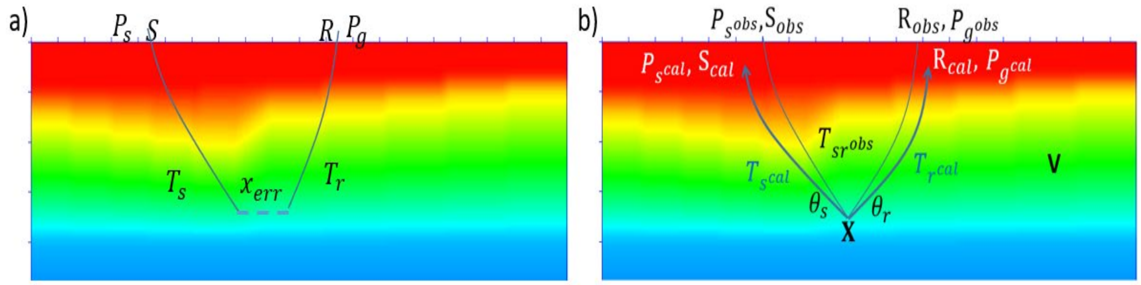


Figure 6. 2 (a) Controlled Direction Reception (CDR) tomography shoots a ray from the surface using picked ray parameters P_s, P_g and estimates the velocity by minimizing the position error Xerr of the ray segment endpoints. (b) Stereo-tomography shoots rays from an estimated scatter point X to the shot and receiver, and estimates the velocity **V** and ray segment parameter **X**, θ_s , θ_g , T_s *and* T_r by minimizing the misfits of the data space parameters.

6.2 Theory

6.21 Forward and inverse problem

In classical reflection tomography, the forward modelling of the traveltime tomography can be represented by:

$$t_{raypath} = \int_{raypath} s(x,z)dl, \tag{6-1}$$

where the measured traveltime t is the integral of slowness $s(x,z)$ along the ray path.

If the line integral equation (6-1) defines a linear system, it can be represented in matrix form as:

$$\mathbf{d} = \mathbf{L}\,\mathbf{m}\,, \tag{6-2}$$

where **d** is the travel time of a raypath, **m** is the slowness model, L is a matrix that contains the physical relationship between the measurements **t** and the model parameter, typically slowness

($\mathbf{s}$). **L** is the Kernel, often equivalent to the Jacobian, Fréchet derivative or sensitivity matrix. For the line integral equation 1, **L** is of the dimension of the number of data points by the number of velocity cells. Each row of **L** contains the ray path segment length for each cell that a ray path has traversed to create a traveltime measurement.

The cost function or misfit function for equation (6-2) is:

$$J(m) = \| \mathrm{d} - \mathbf{L}\,\mathrm{m} \|^2 \tag{6-3}$$

The linear least-squares solution of equation (6-3) is

$$\mathbf{m} = (\mathbf{L}^{\mathbf{T}}\mathbf{L}\,)^{-1}\mathbf{L}^{T}\mathbf{d}\,. \tag{6-4}$$

Since the ray path is a function of the slowness and line integral equation represent a non-linear system, equation 2 becomes:

$$\mathbf{d} = \mathbf{L}(\mathbf{m}), \tag{6-5}$$

and the cost function for equation (6-4) is:

$$J(m) = \| \mathbf{d} \; - \mathbf{L}\,(\mathbf{m}) \|^2. \tag{6-6}$$

Because of the non-linearity of Equations (6-5) and (6-6), the slowness model cannot be obtained directly using Equation 4. However, the traveltime **d** is picked from the data; it is invariant or model independent. We can exactly calculate the cost function in equation (6-6); Therefore, the non-linear problem of classical reflection tomography can be linearized by iteratively solving:

$$\mathbf{\Delta d} = \mathbf{L}\,\mathbf{\Delta m}, \tag{6-7}$$

where $\mathbf{\Delta m}$ is the model update vector between iterations, $\mathbf{L}$ is the Frechét derivative matrix $\partial L(m)/\partial m$, the partial derivatives of the modelled response with respect to the model parameters and $\mathbf{\Delta d}$ is the differences between the modelled and the observed traveltimes.

6.22 Stereo-tomography

Unlike classical traveltime tomography that has only traveltimes in the data space and slowness in the model space, stereo-tomography is a multiparameter problem (Figure 6-2b). To account for data uncertainties, we include the data covariance $\mathbf{C_d}$ in the cost function:

$$J(m) = \left(\mathbf{d} - \mathbf{L}(\mathbf{m})\right)^{\mathbf{T}} \mathbf{C}_{\mathbf{d}}^{-\mathbf{1}} \left(\mathbf{d} - \mathbf{L}(\mathbf{m})\right) \ . \tag{6-8}$$

When the data covariance is uncorrelated, $\mathbf{C_d}$ **i**s a diagonal matrix with the diagonal elements being the square of the standard derivation $\boldsymbol{\sigma}$ of the data, and $\mathbf{C}_{\mathbf{d}}^{-\mathbf{1}}$ is also a diagonal matrix with the diagonal element being $1/\boldsymbol{\sigma}^{\mathbf{2}}$. Therefore $\mathbf{C_d}$ can be chosen according to the standard deviation of the data measurements. It is important to choose the appropriate unit for the data covariance so that the data misfit of different data types is scaled accordingly. The data space of stereo-tomography includes S, R, T_{sr}, P_s and P_g (see Figure 6.2b). Data misfit Δd_i for each data point in equation 6.7 is:

$$\Delta d_i = ((S_{obs} - S_{cal})C_{d_S} \,, (R_{obs} - R_{cal})C_{d_R} \,, \left(T_{sr_{obs}} - T_{sr_{cal}}\right)C_{d_{Tsr}} \,,$$
$$\left(P_{s_{obs}} - P_{s_{pred}}\right) C_{d_{P_s}} , \left(P_{g_{obs}} - P_{g_{pred}}\right) C_{d_{P_g}}). \tag{6-9}$$

The model space includes V, X_c , θ_s , T_s , θ_g and T_g (Figure 6.2b). For each data point, the Fréchet derivative L_i in equation 6.7 is the combination of the derivatives of data space element with respect to the model space element:

$$L_i = \frac{\partial(\mathbf{S},\mathbf{R},\mathbf{T}_{\mathbf{sr}},\boldsymbol{P_s},\boldsymbol{P_g})}{\partial(\mathbf{X}_{\mathbf{c}},\boldsymbol{\theta}_{\mathbf{s}},\boldsymbol{\theta_g},\boldsymbol{T_s},\boldsymbol{T_g},\boldsymbol{V})}\,. \tag{6-10}$$

Each element of the Fréchet derivative L_i can be computed during paraxial raytracing (Cervený et al, 1977). With the data misfit Δd_i and the Fréchet derivative L_i for each data point established, model update $\mathbf{\Delta m}$ can be computed iteratively by solving equation (6-7) using a conjugate gradient method (Hestenes, and Stiefel, 1952, Nocedal and Wright, 2006).

6.3 Synthetic data example

To validate the accuracy of the stereo-tomography method, we create synthetic data sets using a wedge model and the Marmousi model. The wedge model (Figure 6.3a) consists of four constant velocity layers with the second layer thins out with increasing surface location. 200 shot gathers (Figure 6.3b) were created using acoustic finite-difference modelling with 4th order in space and second order in time, and a Ricker wavelet of 25Hz dominant frequency. We picked a rough moveout velocity function (Figure 6.3c) at the middle of the model and created a near-trace stack (Figure 6.3d) using traces with an offset less than 1000 m. Reflection boundaries were picked on the near offset stacks and were used to track the reflection arrival times automatically on the moveout corrected CDP gathers (Figure 6.3e). Moveout correction was then removed from the picked reflection arrival times. Figure 6.3f shows a shot gather with picked reflection arrivals displayed in blue. We use the reflection arrival times to compute the shot and geophone ray parameters for stereo-tomography inversion. Figure 6.4a shows a shot gather with the two-way times and geophone slope picks. Figure 6.4b shows the quality control panels for each analysis location. Figure 6.4c shows the initial estimates of the scatter positions for all the stereo-tomography picks using straight ray and homogeneous media assumptions. The green lines in Figure 6.4c mark the dip bars computed from the 3 stereo-tomography picks.

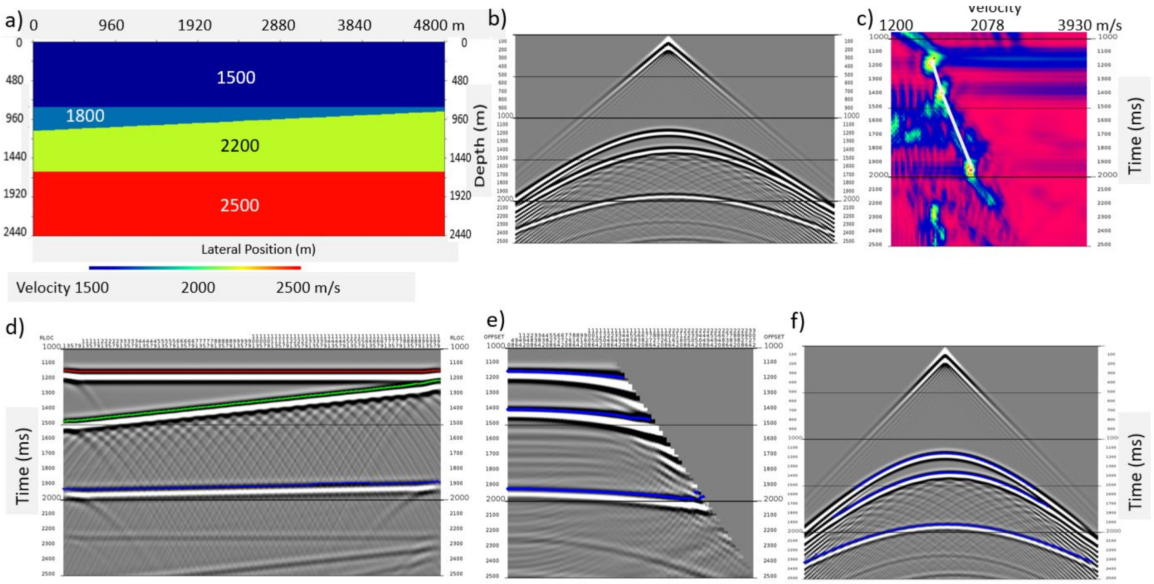


Figure 6. 3. (a) Wedge model. (b) Finite difference shot gather. (c) Moveout velocity scans. (d) Near trace stack. (e) A moveout corrected CDP gather with refraction arrival times picked by correlation with the near-trace stack. (f) Shot gather with reflection arrival times.

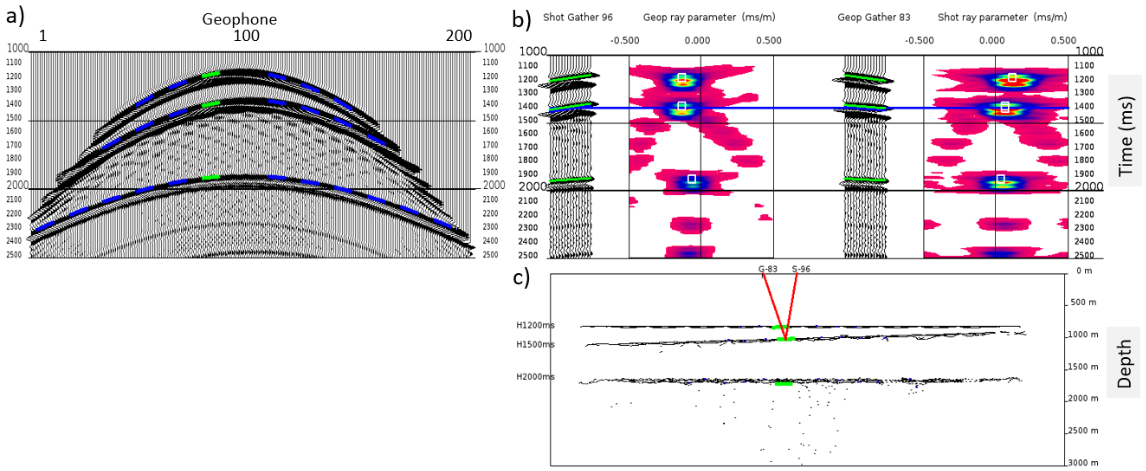


Figure 6. 4. (a) A shot gather with two-way traveltime and geophone ray parameter picks. Green lines mark the picks for the active analysis location. (b) Quality control panels for an analysis location, (c) Initial estimates of the scatter positions. Green lines represent dip bars computed from stereo-tomography picks at the active analysis location.

We use the stereo-tomography picks to estimate the velocity V and ray segment parameters X, θ_s, T_s, θ_g and T_g. Figure 6.5a shows the final velocity solution with the velocity values at 6 locations displayed in white. Velocity values of the true model are displayed in black. Figure 6.5b shows the estimated scatter positions. Velocity solution from stereo-tomography does not capture the blocky characteristics of the wedge model, but it resembles velocity gradients centred around the true velocity values. The scatter positions match the velocity boundaries of the wedge model; hence, it is possible to develop a hybrid method that uses the scatter positions to establish the layer boundaries and modify the stereo-tomography algorithm to estimate a layered based velocity model. We estimate the velocity boundaries by linear fitting the computed scatter positions. The average velocity is computed by averaging stereo-tomography solution between layer boundaries

(Figure 6.5c). Figure 6.5d displays the true velocity model with linear fitted scatters for comparison.

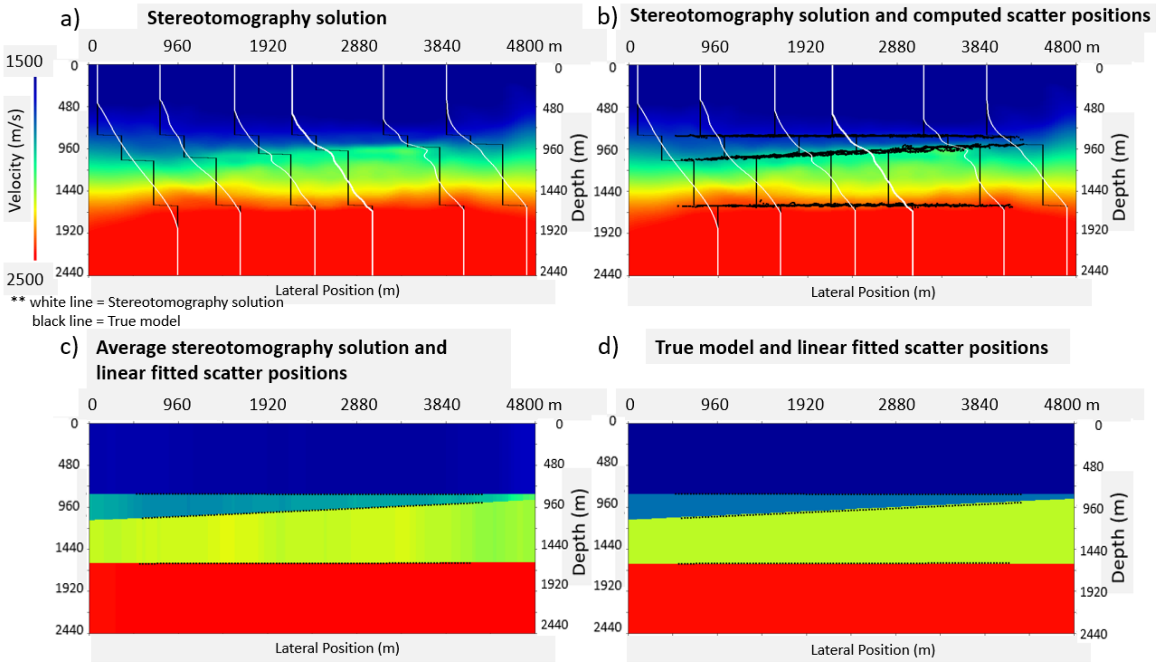


Figure 6. 5. (a) Stereotomography solution of the wedge model synthetic dataset. Velocity values at 6 locations are displayed in white for the stereo-tomography solution and in black for the true model. (b) Scatter positions are displayed as black dots. (c) Velocity layer boundaries are computed by linear fitting the computed scatter positions and displayed as black dots. Stereotomography solution is averaged between computed layer boundaries. (d) Linear fitted scatters positions displayed as black dots on the true velocity model.

We also use the Marmousi model (Figure 6.6a) to create 261 synthetic shot records with 96 traces per shot (Figure 6.6b). Both the shot spacing and geophone spacing are 25m. Because of the complex structures of the model and reflection signal, it is necessary to pick the reflection arrival times, shot and geophone ray parameters using semblance of the shot and geophone slant stacks. Figure 6.7a shows the picked reflection events on shot 248. Figure 6.7b shows the picked events and semblance at shot location 248 and geophone location 170. Figure 6.7c shows the initial estimates of the scatter positions for all the stereo-tomography picks using straight ray and homogeneous media assumptions. We use a constant velocity gradient as the starting model for stereo-tomography inversion. Figure 6.8a shows the final velocity solution, and figure 6.8b overlays the final scatter positions on the velocity solution. Comparing the stereo-tomography solution to the true model in figure 6.6a, the velocity solution captures the long-wavelength trend of the true model up to about 2Km, and the scatter positions match the velocity boundaries in the true model. Figure 6.8c displays the velocity values from the stereo-tomography solution in white, and the true model in black at 6 locations. This further confirms that the stereo-tomography solution does capture the long-wavelength trend of the true model.

To evaluate the effectiveness of the stereo-tomography solution as a starting model for high-resolution inversion methods such as FWI, we perform FWI on the Marmousi data set using starting model from a constant vertical velocity gradient (Figure 6.9a) and from the stereo-tomography solution (Figure 6.9c). Figures 6.9b and 6.9d compare the FWI solution from the two starting models. FWI solution using the constant vertical velocity gradient as the starting model only captures some of the high-frequency velocity changes in the true model. FWI solution using the stereo-tomography solution as the starting model recovers most of the velocity features up to about 2Km. Details below 2Km are missing in the FWI solution. Potential improvement can be using higher-order finite-difference propagation in FWI and using the higher resolution multi-grid FWI approach proposed by Trad 2020, in which the high-frequency field data will be shaped to the lower frequency predicted data between iterations.

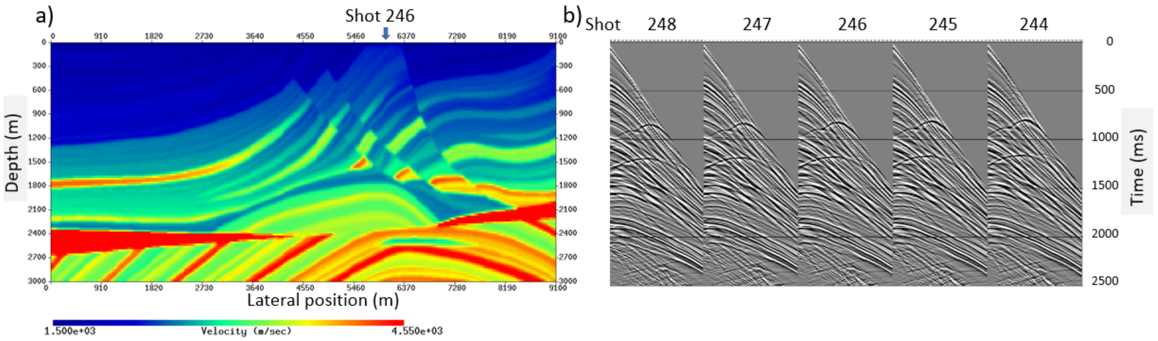


Figure 6. 6. (a) Marmousi model. (b) Synthetic shot records.

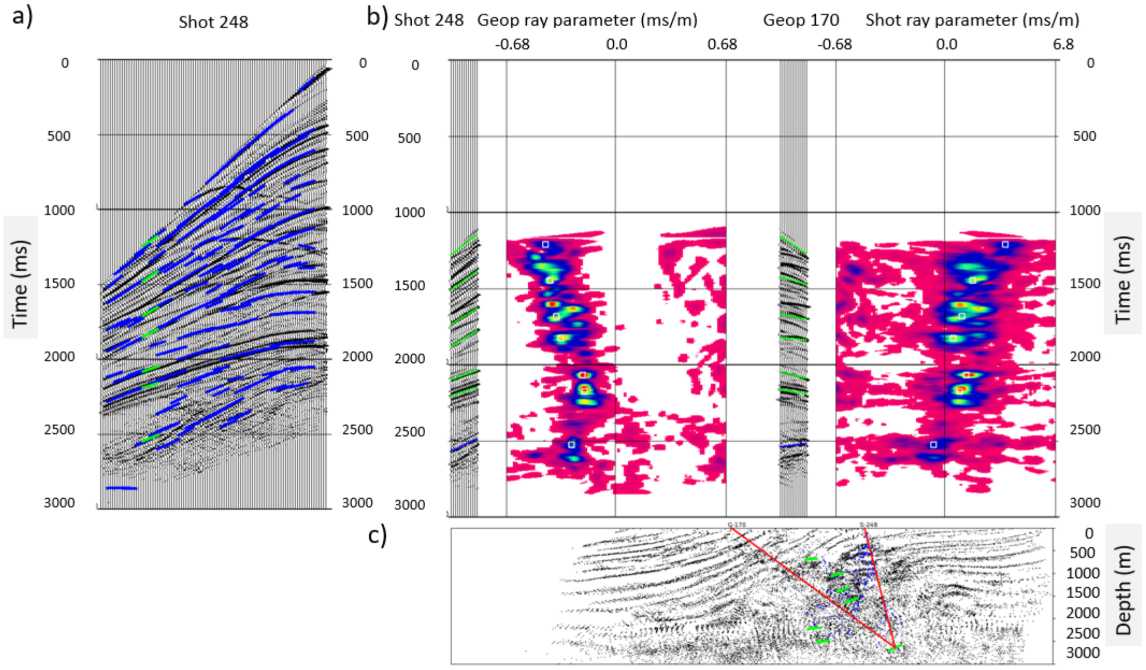


Figure 6. 7. (a) A shot gather with two-way traveltime and geophone ray parameter picks determined by the maximum semblance of the shot and geophone slant stacks. Green lines mark the picks for the active analysis location. (b) Quality control panels for an analysis location, (c) Initial estimates of the scatter positions. Green lines represent scatter positions and dips computed from stereotomography picks at the active analysis location.

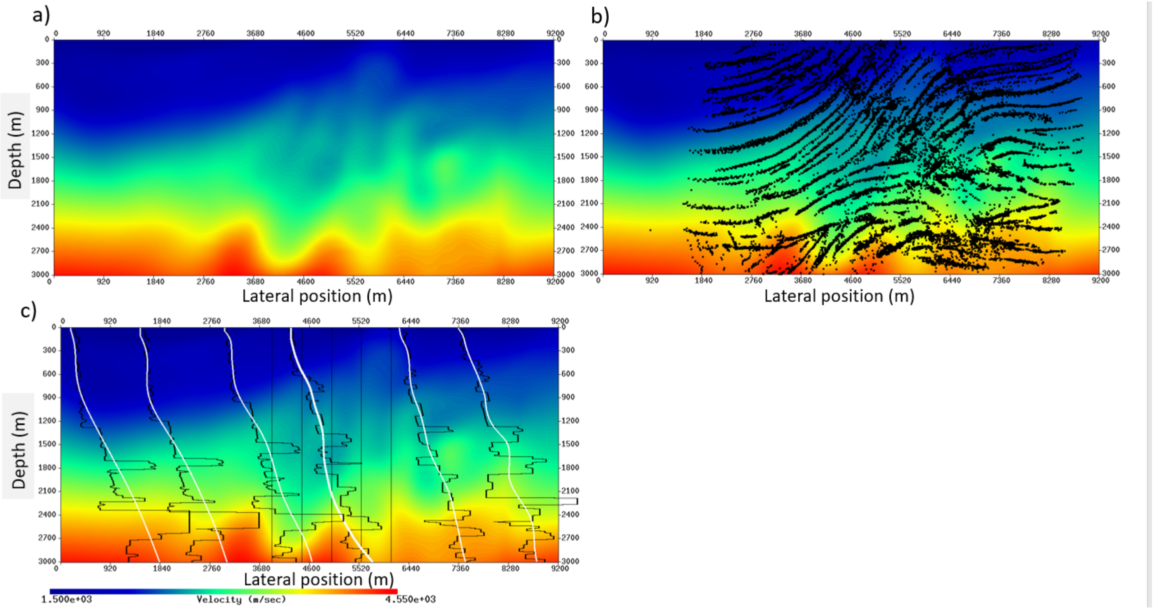


Figure 6. 8. (a) Stereotomography solution of the Marmousi synthetic data set. (b) Scatter positions are displayed as black dots. (c) Velocity values at 6 locations are displayed in white for the stereotomography solution and in black for the true model.

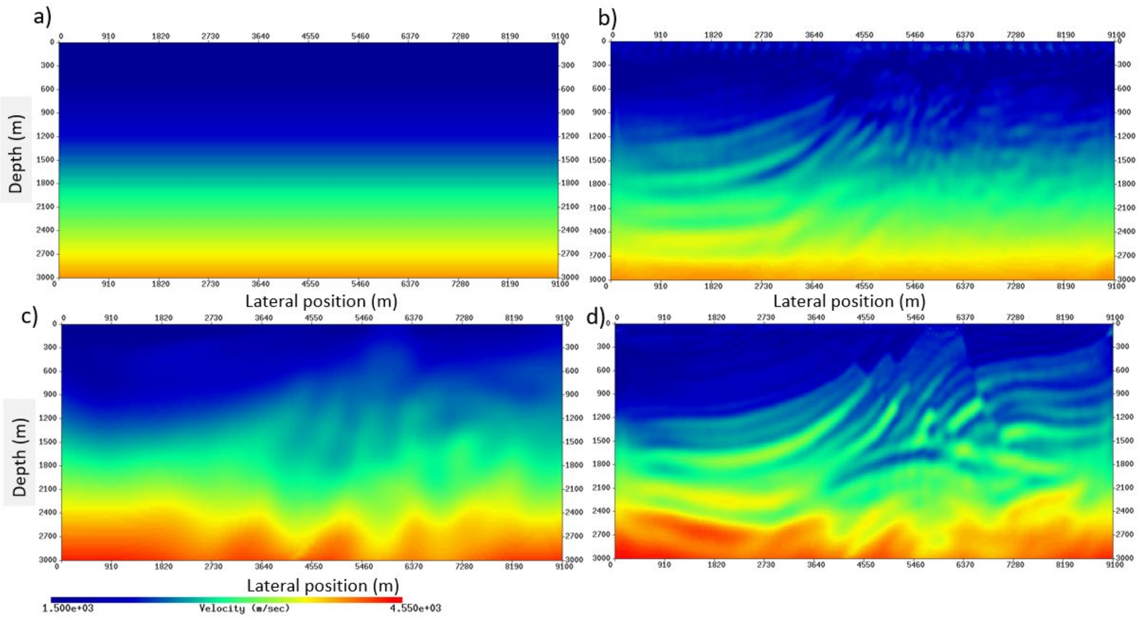


Figure 6. 9. (a) Constant vertical velocity gradient model. (b) FWI solution using (a) as the starting model. (c) Stereotomography solution. (d) FWI solution using (c) as the starting model.

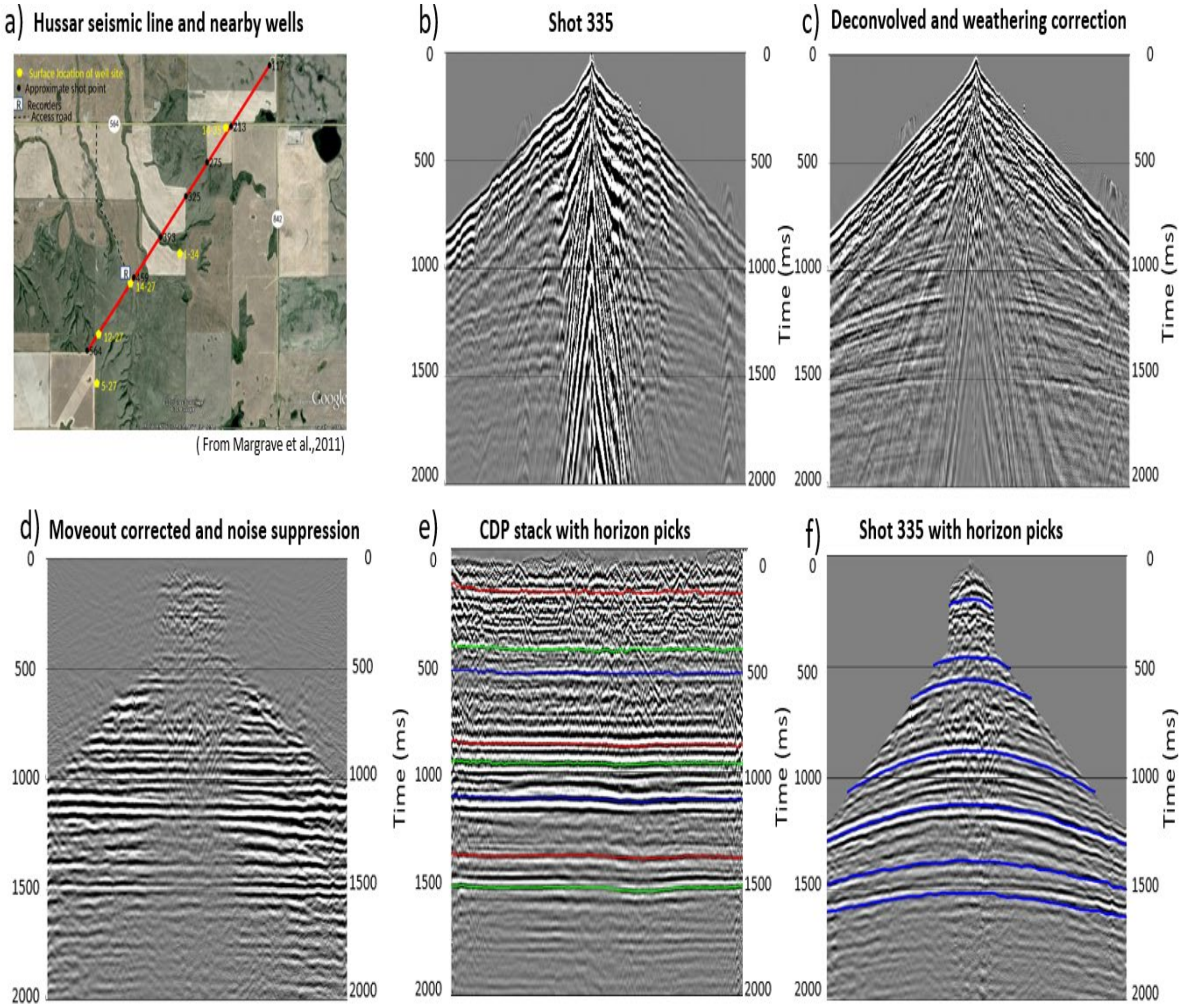
6.4 Field data example

We apply similar approaches of the wedge model and the Marmousi model in stereo-tomography to a 2D land dataset. The data used in the example are the vertical component of the dynamite shot records from a 4.5 Km long 2D 3C survey acquired at Hussar, Alberta in September 2011. The seismic survey was acquired for a broadband experiment (Margrave et al., 2012). Figure 6.10a shows the location and the layout of the seismic line and some nearby wells. Figure 6.10b shows a spherical divergence corrected shot gather, and figure 6.10c shows the deconvolution and weathering statics correction of the same shot records. To remove the ground roll interference and to improve the lateral coherence, we use Radon transform filter on moveout corrected CDP gathers. Figures 6-10d shows the same shot records after the Radon transform filter in the CDP domain. We use the same approach that we took in the wedge model to automatically pick the refraction arrival time by correlating a CDP stack (Figure 6.10e) with the moveout corrected and noise attenuated CDP gathers. The moveout correction times are then backed out from the picked reflection arrival times. Figure 6.10f shows the same shot records with picked reflection arrival times. In order to identify the tie between the well logs and the seismic events, we compare the CDP stack in time with the 8-12-45-55 Hz. synthetic seismograms (Figure 6.11). In specific, we tie the CDP stack with the Belly River, Basal Belly River, Base Fish scales and the Mannville formations. The Belly River reflection is quite noisy, we choose to add it for better control on the shallow. We also attempt to do some picking at 200 ms and below 1000 ms.

The reflection arrival times are then used to compute the geophone and shot ray parameters. To remove errors in the reflection arrival times, the geophone and shot ray parameters are picked again using the maximum semblance of the shot and geophone slant stacks (Figure 6.12a). The final reflection arrival times and ray parameters are then used for stereo-tomography inversion.

The final stereo-tomography solution is displayed with P-P velocity from well 01-34-025-21W4 and 14-34-025-21W4 (Figure 6.12b). The velocity values from the stereo-tomography solution at the well locations agree with the long-wavelength trend of the P-P velocity from well logs. Figure 6.6-12c shows the scatter position solution. The scatter positions align with Belly River, Basal Belly River, Base Fish Scales and Mannville. However, the picks do not cluster as tightly as the two synthetic models. This is an indication of some picking error.

To validate the accuracy of the stereo-tomography solution, we depth migrate the Hussar 2d data set using the stereo-tomography solution as the velocity model. Figure 6.13a shows the depth image gathers and figure 6.13b shows the depth migration result with P-P velocity logs in depth. This result confirms that the stereo-tomography velocity solution accurately migrates the Hussar 2D data in depth. A further enhancement is possible through residual curvature analysis using the depth image gathers.



(From Margrave et al.,2011)

Figure 6. 10. (a) Hussar seismic line and the location of nearby wells. (b) Shot 335 with spherical divergence correction. (c) Shot 335 deconvolved with weathering statics correction. (d) Noise suppression with Radon transform on moveout corrected data. (e) CDP stack with 8 picked horizons. (f) Reflection arrival times picked by correlating (d) and (e).

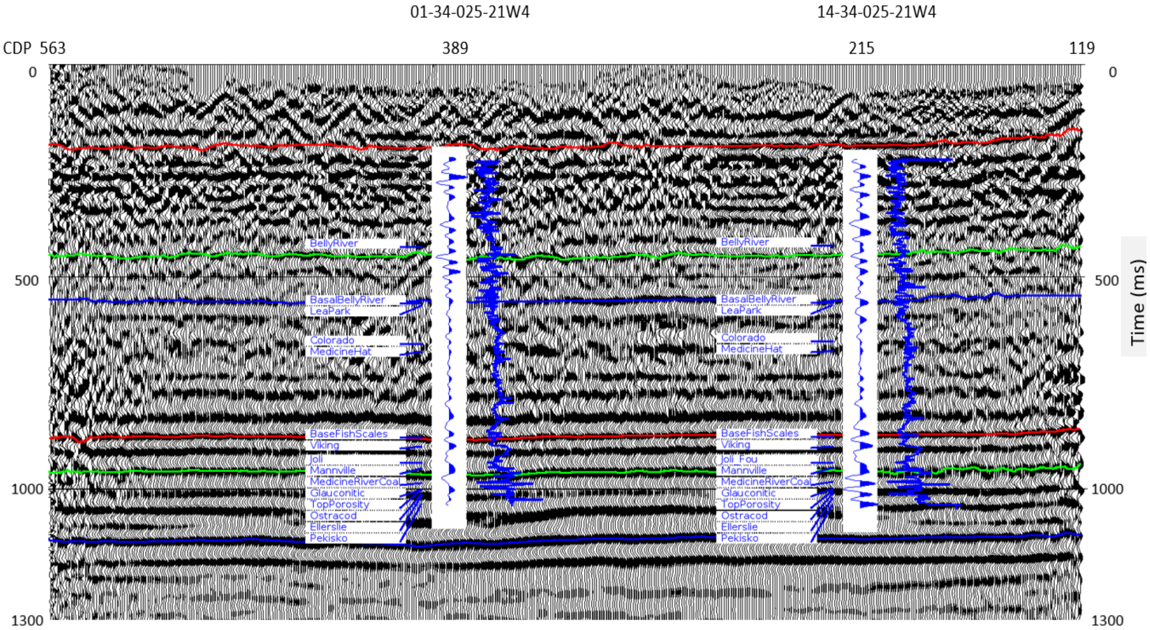
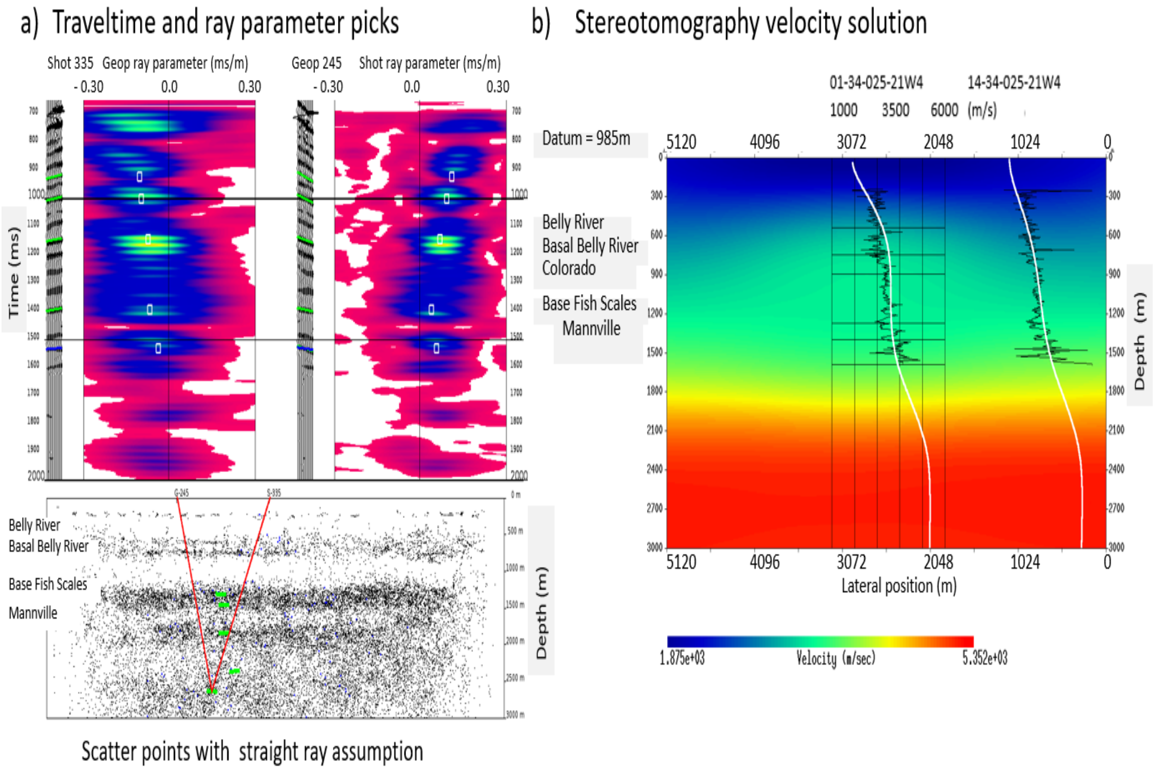


Figure 6. 11. CDP time stack with well ties and synthetic 8-12-45-55 Hz. Seismograms to identify reflection events with well tops.



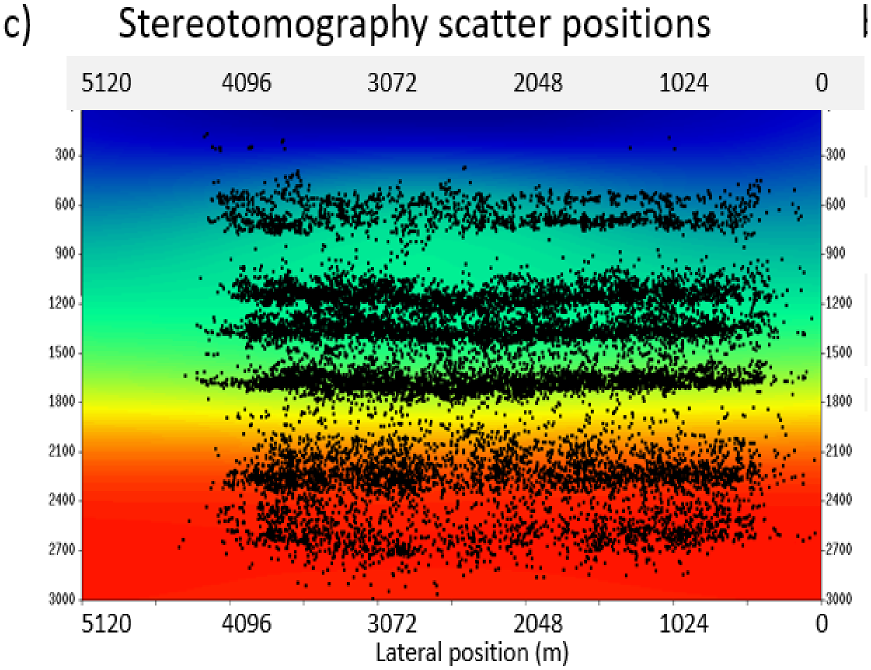
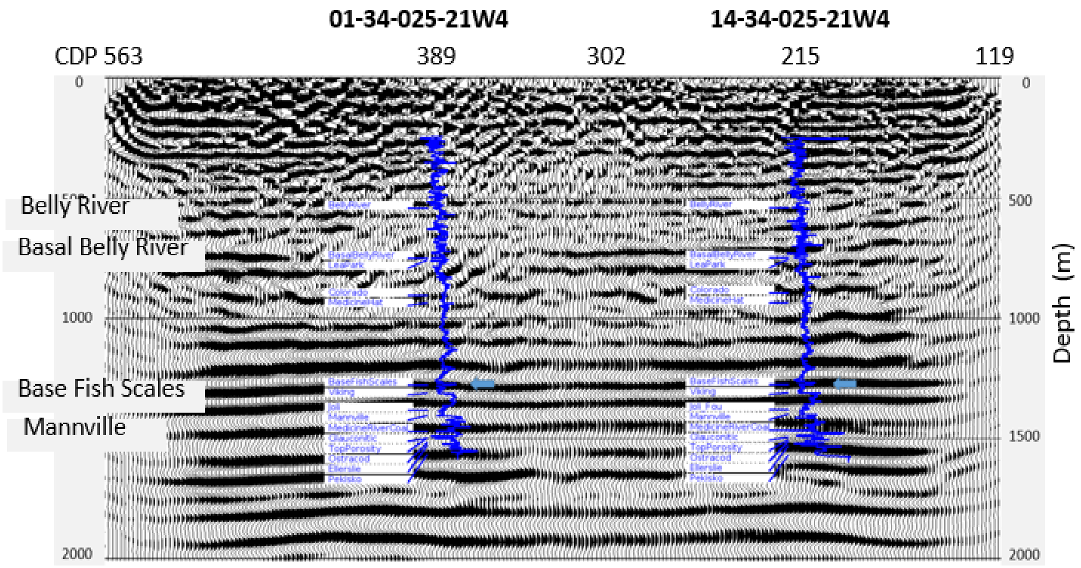


Figure 6. 12. (a) Geophone and shot ray parameters refined by the semblance of shot and geophone slant stacks, and scatter point positions computed from all the stereo-tomography picks using straight-ray and homogeneous velocity assumption. (b) Final stereo-tomography velocity solution

and well ties. Smooth white lines are stereo-tomography velocity values at well locations. (c) Scatter position solution.


b) Depth migration using stereotomography solution

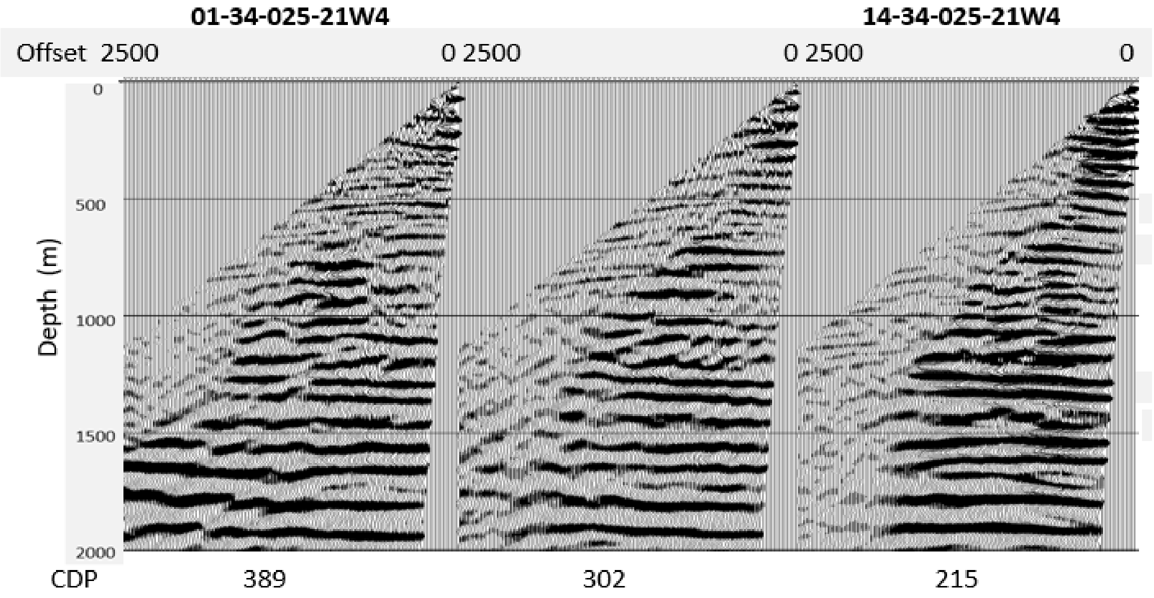

a) Common image gathers

Figure 6. 13. (a) Images gathers at the well locations and one image gather in between the well locations. (b) Depth migration using velocity from stereo-tomography displayed with P-P velocity logs at well locations 01-34-025-21W4 and 14-34-025-21W4.

6.5 Conclusion

We have reviewed the stereo-tomography method and verified its accuracy and characteristic with a wedge model and the more complex Marmousi model. We showed that the velocity solution from stereo-tomography does not capture the blocky characteristics of the wedge model, but it resembles velocity gradients centred around the true velocity values. The scatter positions match the velocity boundaries of the wedge model; hence, it is possible to develop a hybrid method that uses the scatter positions to establish the layer boundaries and modify the stereo-tomography algorithm to estimate a layered based velocity model. The Marmousi model test showed that the solution of stereo-tomography captured the long-wavelength velocity model that helped FWI to converge to a high-resolution model. We noticed that the FWI solution below 2Km was not able to capture the details of the Marmousi model. A higher-order finite-difference propagation in the FWI and adaptive multi-grid FWI approach can potentially improve the resolution of the FWI solution below 2Km in the Marmousi model. We apply the stereo-tomography approaches used in the blocky wedge model and the complex Marmousi model to the Hussar 2D dataset. We first track the reflection arrival times on moveout corrected and noise attenuated CDP gathers, and then remove the moveout corrections from the reflection arrival time picks. To correct for the picking errors, shot and geophone ray parameters are picked again automatically using the semblance of the shot and geophone slant stacks. The stereo-tomography solution was found to be accurately migrating the Hussar 2D data set to a depth section that agrees with the well logs. A further enhancement is possible through residual curvature analysis using the depth image gathers.

Chapter 7

Traveltime Tomography: First break picking and machine learning

One of the most laborious and problematic tasks in refraction tomography is the first arrival traveltime or first break (FB) picking. Many automated FB picking methods determine the arrival time by the difference in amplitude, phase, or frequency characteristics between the data before and after the FB and are often done on a trace-by-trace basis. Spatial correlation between adjacent traces is only used for subsequent editing of mis-picks. The final step in FB picking is to confirm or manually modify the FB picks by trained technicians. With experiences from a large number of datasets with different topography and near-surface geological setting, experienced technicians can recognize the relationship between the FB and the complex waveform of the first arrival energy and various interfering noises. With increasing data density, this has become a very time-consuming and expensive process.

Machine learning is a fast-developing science that teaches computers to learn from data and human experiences. There are two potential applications of machine learning in automatic FB picking. One application of machine learning is automated editing of outlying picks by clustering. Another more important application is deep learning by training the networks with manually edited FB and classifying the first arrival energy waveforms as pre-FB and post-FB. With a catalogue of images of trained models, the deep neural works will be able to classify the first arrival energy waveforms of a new datasets as pre-FB and post-FB as accurately as the trained technicians. In this chapter, I will review some of the automatic FB picking methods, clustering applications and one deep-learning application.

7.1 Automated first arrival picking

During a seismic experiment, ground motions are recorded after a seismic source is activated. A seismic record contains the refraction and reflection seismic signals caused by the seismic source, as well as surface-related seismic noises (Figure 7.1) caused by ambient noise, human and animal activities. First arrival energy is characterized by the relatively weak surface noises, followed by the stronger refraction seismic signal. I will review and compare two automatic first arrival picking methods that use these amplitude characteristics. The signal to noise ratio (SNR) method defines the sum of peaks, $P(1{:}\,k-1)$, of a potential first arrival time pick k, as noise, and the peak at k, as the signal, and $S2N$ at k as:

$$S2N(k) = \frac{P(k)}{Sum(P(1{:}k-1))} \tag{7-1}$$

Another method that uses the characteristics of the amplitude levels before and after the first arrival is the Akaike's information criterion (AIC) (Akaike, 1973). AIC is defined as:

$$AIC(k) = k * \log\,(var\big(y(1{:}\,k)\big) + (nsamp - k - 1) * (\log\,(var(y(k+1{:}\,nsamp)), \tag{7-2}$$

where $var(y(i{:}\,j))$ is the variance for the time series y from sample i to k.

$$var\big(y(i{:}\,j)\big) = \frac{sum\Big(\;v\big(y(i{:}j) - \overline{y(\imath{:}\jmath)}\;\;\big)^{2}\Big)}{j-i}_{ii=(i:j} \tag{7-3}$$

When k is less than the first arrival time, $var\big(y(1{:}\,k)\big)$ is small, and $(var\big(y(k+1{:}\,nsamp)\big)$ is large. Similarly, when k is greater than the first arrival time, $(var\big(y(1{:}\,k)\big)$ becomes larger, and $var(k+1{:}\,nsamp)$ becomes smaller. When k is at the first arrival time, $AIC(k)$ is at its minimum.

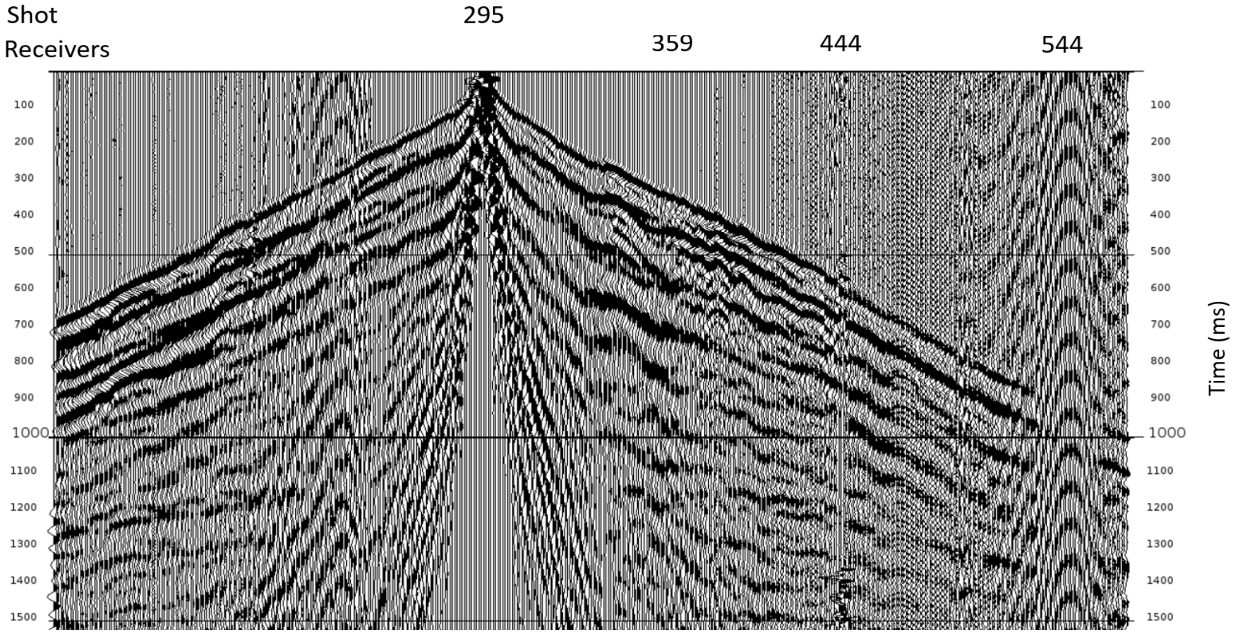


Figure 7. 1. Seismic record with varying surface noise conditions. Receivers 359, 444 and 544 are identified with an increased noise level.

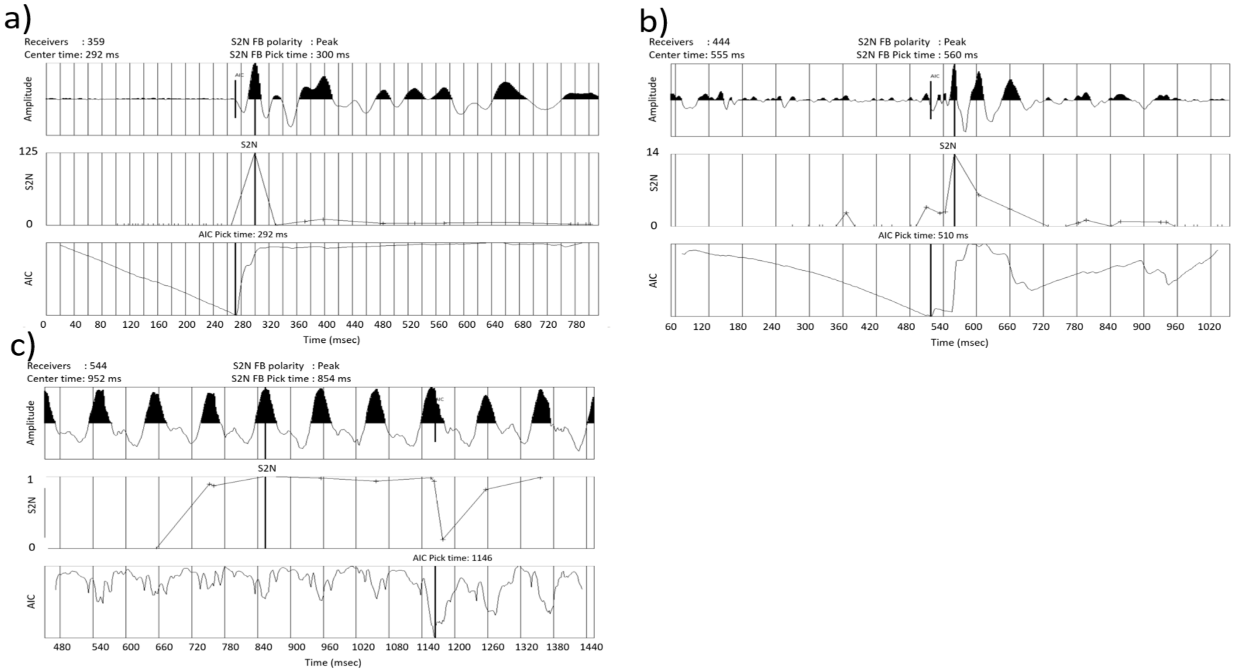


Figure 7. 2. Results of S2N and AIC methods with increasing noise level (a to c). Top panel: first arrival energy. Middle panel: S2N analysis. Bottom panel: AIC analysis.

Figure 7.2a shows the first arrival energy with weak surface noise, followed by S2N and AIC analysis. Using positive FB polarity. The first arrival pick is chosen at a peak with the maximum $\boldsymbol{S2N}(\boldsymbol{k})$. The polarity of the S2N analysis method is peak; therefore, it picks a peak that meets the S2N criteria and arrives later than the zero-crossing picked by the AIC method. With a moderate increase in noise level (Figure 7.2b) both S2N and AIC methods are able to detect the consistent FB picks. However, with significantly stronger surface noises that overwhelm the first arrival energy, both S2N and AIC methods fail.

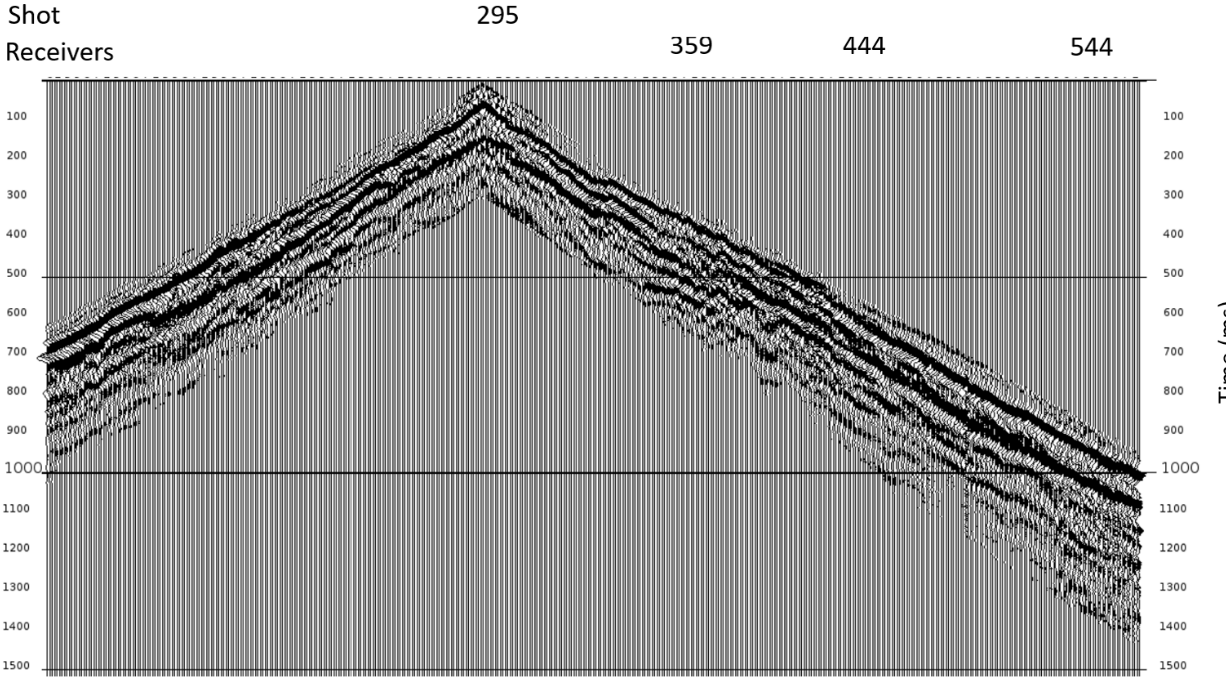


Figure 7. 3. Seismic shot 295 high-resolution linear Radon transform in windowed first arrival energy in receiver gather domain.

I sort the seismic records into common receiver gather domain and apply high-resolution Sparse Linear-Radon transform to remove the surface noise from windowed first arrival energy. Figure 7.3 shows the shot record 295 after being sorted back to the shot domain. This shows that high-resolution Sparse Linear-Radon transform in common receiver gather domain is an effective algorithm to remove surface noises that appear coherent within the shot gathers. Figure 7.4 shows both S2N and AIC methods can effectively determine the FB picks after the removal of the strong surface noise. However, this approach also creates some pre-first arrival artifacts that can negatively affect the FB picking accuracy. In general, AIC is superior to the S2N method; however, additional preconditioning processes may still be required to remove strong surface noises. Moreover, carefully editing of automatically picked FBs is often required by trained technicians with experience in identifying FB through complex first arrival waveforms and surface noises.

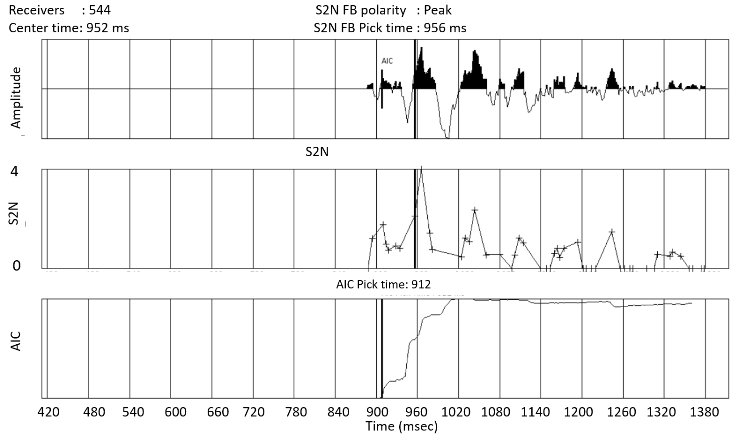


Figure 7. 4. Results of S2N and AIC methods at shot 295 and receiver 544 after high-resolution linear Radon transform in windowed first arrival energy in common receiver gather domain.

7.2 Application of machine learning in automatic first arrival picking

Unlike the two trace-by-trace automated first arrival picking algorithms, machine learning algorithms explore the spatial relationship between data points. When technicians perform quality control of the computer-picked FBs, they reject FB picks that vary rapidly with respect to the neighbouring picks. Clustering is a machine learning technique that groups the data points according to their attributes. It has the potential of automating the human efforts in rejecting and modifying the FB picks that vary too rapidly. Three commonly used clustering algorithms are K-Means, Gaussian Mixture Models (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (see Géron, 2019).

7.2.1 K-Means

K-Means (Lloyd, 1982) assumes the data points distribution to be Euclidean or circular. The Euclidean distance square is defined as:

$$\Delta^2 = (x - \mu)^T (x - \mu), \tag{7-4}$$

where x is the dimensional matrix, μ is the mean matrix.

The following steps outline the K-Means algorithm:

1. Starts with randomly placing the centroids of chosen N clusters (an initial guess).

2. For each data point, calculates the Euclidean distance between the data point and each of the centroids.

3. To find the clusters, assign the data point to the nearest centroid.

4. Recompute the coordinates of the centroids using the mean coordinates of the clusters.

5. Repeat steps 2, 3, and 4 until convergence.

7.2.2 Gaussian-Mixture-Models

GMM assumes the data points distribution to be Gaussian and is less restrictive than K-means. The Gaussian distribution function is defined as:

$$f(x) = \frac{1}{2\pi \left|\sum\right|^{\frac{1}{2}}} \exp[-\frac{1}{2}(x - \mu)^T \sum^{-1}(x - \mu)] \tag{7-5}$$

where x is the dimensional matrix, μ is the mean matrix and $\sum$ is the covariance matrix. For a two dimensions case and μ equals 0:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad 7\text{-}6$$

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad 7\text{-}7$$

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \qquad 7\text{-}8$$

$$\sum^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix} = \begin{bmatrix} a & -b \\ -b & c \end{bmatrix}. \qquad 7\text{-}9$$

The Mahalanobis or statistical distance is different from the Euclidean distance by the inclusion of the inverse covariance matrix:

$$\begin{aligned} \Delta &= (x-\mu)^T \textstyle\sum^{-1}(x-\mu)] \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & -b \\ -b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= ax_1^2 - 2bx_1x_2 + cx_2^2 \qquad 7\text{-}10 \end{aligned}$$

For the special case of $a = c$, and $b = 0$, this is the same as Euclidean distance. For the case of $a > c$, and $b = 0$, this is a horizontal ellipse. For the case of $a < c$, and $b = 0$, this is a vertical ellipse. For the case of $a \neq c$, and $b \neq 0$, this is a rotated ellipse. Hence, Mahalanobis distance is more flexible and can handle elongated clusters better than Euclidean distance.

The following steps outline the algorithm that uses the Mahalanobis distance:

1. Starts with randomly placing the centroids of N clusters.

2. Compute the mean matrix μ and the covariance matrix $\sum$ of the clusters.

3. Compute the Mahalanobis distance between the data point and each of the centroids.

3. To find the clusters, assign the data point to the nearest centroid.

4. Repeat steps 2, 3, and 4 until convergence.

GMM available in Scikit-learn achieves the same result but is implemented differently.

The following steps outline the GMM algorithm:

1. Starts with randomly initializing the Gaussian distribution parameters μ and $\sum$ for each cluster.
2. GMM iterates until convergence using the Expectation-Maximization (EM) algorithm.

7.2.3 Density-Based Spatial Clustering of Application with Noise

DBSCAN is a density-based clustering algorithm that forms clusters of dense regions of data points and ignores the low-density areas by considering them as noise. Hence, DBSCAN has an advantage in handling clusters with irregular shapes and data points with noisy outliers.

DBSCAN uses two parameters eps, and $\min\_samples$. eps defines the maximum distance between two points for them to belong to the same cluster. $\min\_samples$ defines the minimum number of data points a cluster must-have. The following outlines the DBSCAN algorithm:

1. For each data point calculate its distance from all other points. If the distance is within eps, it is a neighbour of the corresponding data point. If the data point has a number of neighbours greater than or equal to $\min\_samples$, it is considered a core point.

2. For each core point that has not been assigned to a cluster create a new cluster. For this core point, find all its neighbouring points and assign them to the same cluster.

3. Continue step 2 until all the non-core points are covered.

I will demonstrate the effectiveness of the three clustering algorithms in recognizing trends and rejecting outliers using FB from shot records from the Hussar 2D (Margrave et.al 2012). The shot record is linear moveout (LMO) corrected with time correction of 0 ms at 0 m offset and 1600 ms at 4500 m offset, and bulk shifted by 500 ms (Figure 7.5). Automatic FB picking is done using the AIC algorithm. Three groups of mis-picks are introduced at around sequential receiver locations 120, 220 and 320. The FB picks are input to the K-Means algorithms, GMM algorithms and DBSCAN algorithms. The parameter for the K-Means and GMM algorithms is 20 clusters. The parameters for the DBSCAN algorithm are $eps = 0.08$ and min_samples=3. Figures 7.6a to 7.6c compare the results from the 3 algorithms. All three algorithms capture the trend of the FB picks; however, only DBSCAN can reject the mis-picks at around sequential receiver 120, 200 and 320. Figure 7.6c displays the input FB as black dots and the interpolated DBSCAN picks as a blue line. Figure 7.7a displays the input FB picks and interpolated DBSCAN in the same scale as the LMO shot gather 203. Figure 7.7b overlays the interpolated DBSCAN picks on LMO shot gather 203. This shows DBSCAN is a good algorithm for rejecting outlying picks in FB picks.

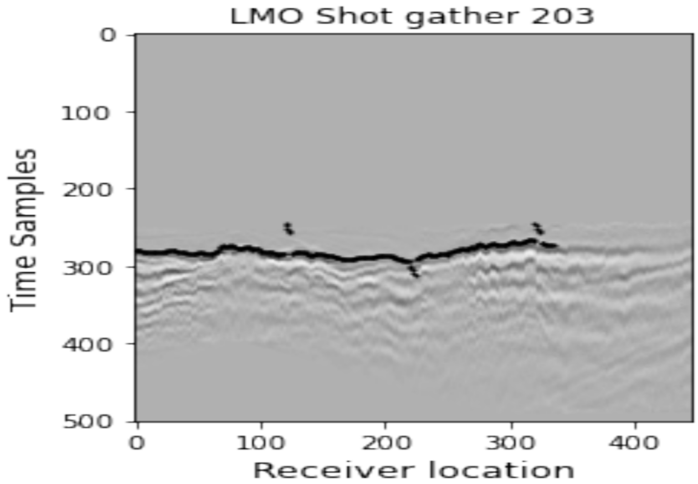


Figure 7. 5. Linear Moveout (LMO) and bulk shifted shot records with first break picks displayed as black dots.

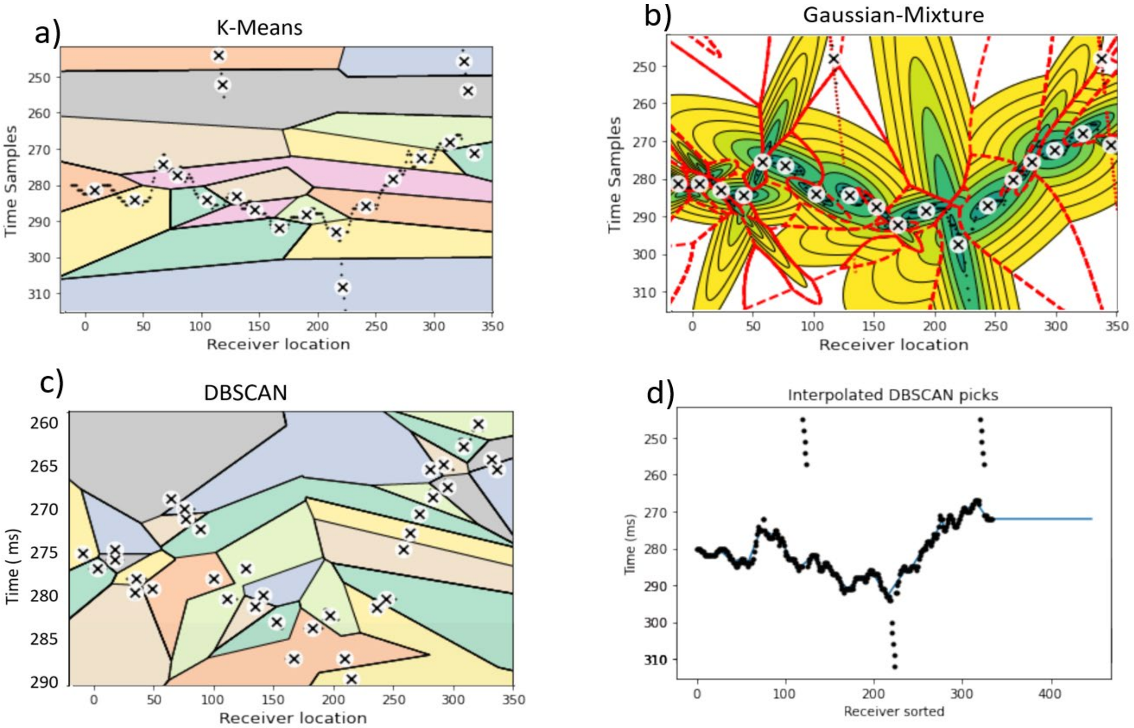


Figure 7. 6. (a) K-Means cluster boundaries and centroid, b) GMM cluster boundaries and centroids, c) DBSCAN cluster boundaries and centroids, d) Interpolated DBSCAN centroids displayed as a blue line and input FB picks displayed as black dots.

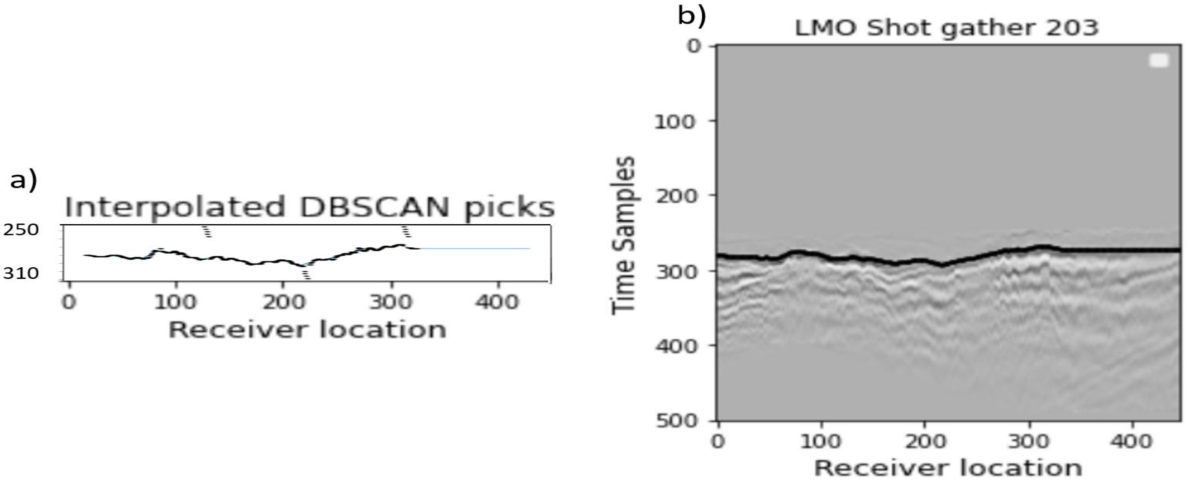


Figure 7. 7. a) Interpolated DBSCAN centroids displayed as a blue line and input FB picks displayed as black dots in the same scale as the LMO shot gather, b) LMO shot gather 203 overlays with interpolated DBSCAN picks.

7.2.4 Deep learning with UNET

Another more important machine learning application is supervised deep learning by training the networks with images of the first arrival energy waveform and pre-FB and post-FB masks created by automatically picked, and manually edited FB (Figure 7.8). With a catalogue of images of trained models from the regions of similar near-surface geology, the deep neural networks will be able to classify the first arrival energy waveforms of a new dataset as pre-FB and post-FB as accurately as the trained technicians. The ultimate goal of the network is to classify each pixel of an input image according to the class to which it belongs.

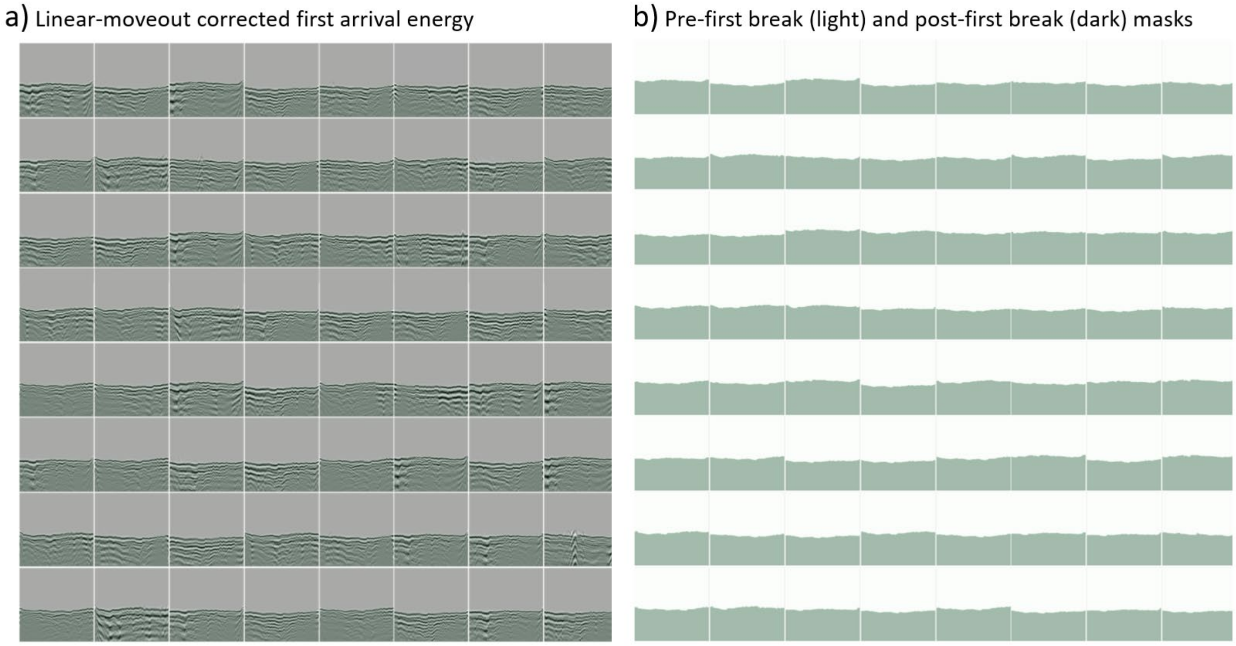
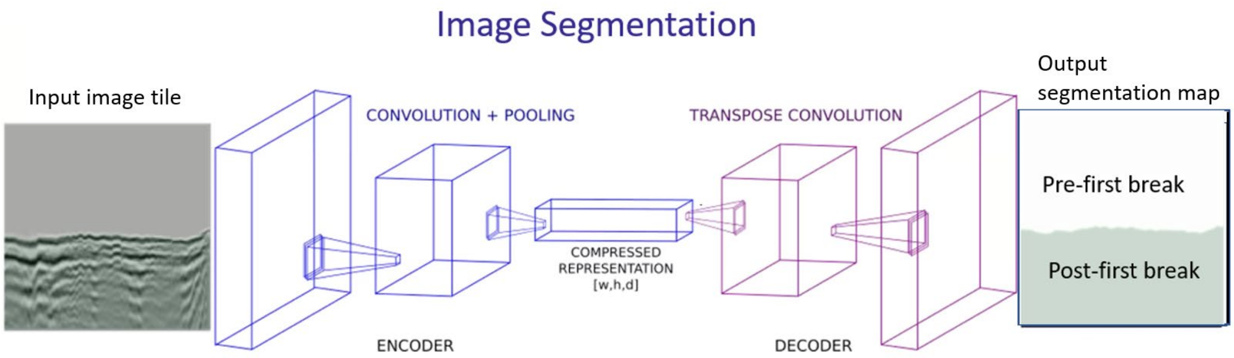


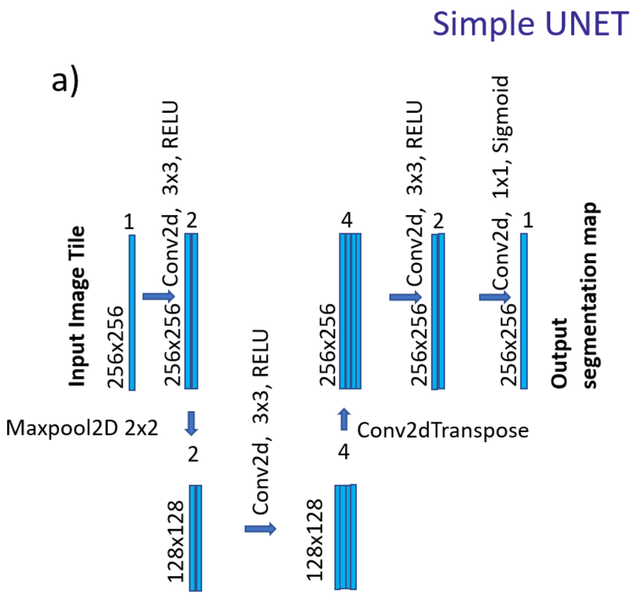
Figure 7. 8. Linear-moveout corrected shot records, b) Corresponding pre-first break and post-first break masks.

This can be solved as an image segmentation problem using UNET (Ronneberger et al., 2015). One-half of the UNET increases the depth of the feature maps and downsizes the resolution of the image using convolutional filters during the encoding process, while the other side of UNET increases the resolution of the output using transpose convolution filters during the decoding process (Figure 7.9). The input image size of an LMO corrected shot record is 258 traces and 501 samples. This is resized to 256 traces by 256 samples for the UNET. The first break mask is either pre-FB or post-FB; therefore, has a depth of 1. Figure 7.10 describes the simple UNET used in the problem. The Conv2D filter has a 3x3 filter width and height and 2 feature maps. A 2x2 Maxpooling reduces the image size to 128x128. Another Conv2D filter with a kernel size of 3x3 increases the number of feature maps to 4. The other half of the UNET uses a transpose Conv2D filter to increase the image size back to 256 by 256. The subsequent Conv2D filter reduces the depth of the output segmentation map back to 1.



* Modified from CREWES data science lab : Convolutional Neural Networks

Figure 7. 9. First, break picking as an image segmentation problem.



b)

Layer	Output Shape	Param #
Input	(256,256,1)	0
Conv2D(2,3)	(256,256,2)	20
Maxpool2D(2,2)	(128,128,2)	0
Conv2D(4,3)	(128,128,4)	76
Conv2D_transpose(2,2,2)	(256,256,4)	34
Conv2D(2,3)	(256,256,2)	74
Conv2D(1,1)	(256,256,1)	3

Figure 7. 10. Simple UNET for the first break image segmentation problem.

7.2.5 Field data example

The vertical component of the Hussar 2D multicomponent seismic survey is used for the UNET test. The first arrival energy is linear moveout corrected to reduce the data size, and the first arrival picking is done automatically using the AIC method, followed by manual editing. The edited FB times are used to create pre-FB and post-FB masks. Convolutional neural networks require input images to be of the same dimensions; therefore, we extract data from 0 to 2230 meters offset from each shot. The result is 258 positive and negative 2D spreads of 224 traces and 501 samples each. The input images are then resized to 256 traces by 256 samples. The 258 images are separated into 206 training images and 52 validation images. The number of training images is also expanded to 824 images by augmentation. A simple UNET is a setup as described in figure 7.10. A validation test using 60 epochs is run. Figures 7.11a and 7.11b show the convergence history for the validation test. The network converges after 15 epochs with 96.8% accuracy. Figure 7.12a shows the superimposed classification predictions on the first arrival energy. Figure 7.12b shows the classification predictions. The prediction results are good but not perfect because a small percentage of the traces have predicted the earlier FB than what is expected according to the first arrival images. This may be the result of the lack of training images for the UNET. If more training images are available, the prediction results are expected to improve.

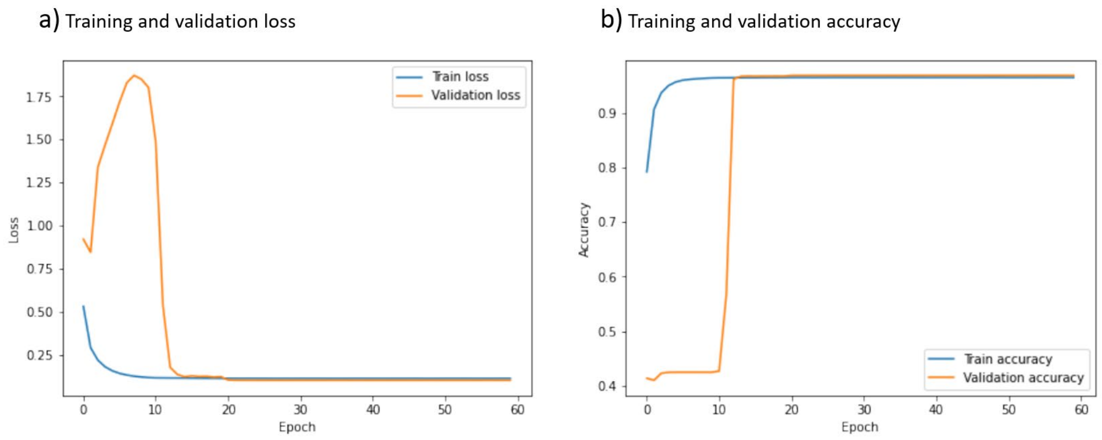


Figure 7. 11. (a) Training and validation loss displayed in blue and green respectively, b) training and validation accuracy displayed in blue and green respectively.

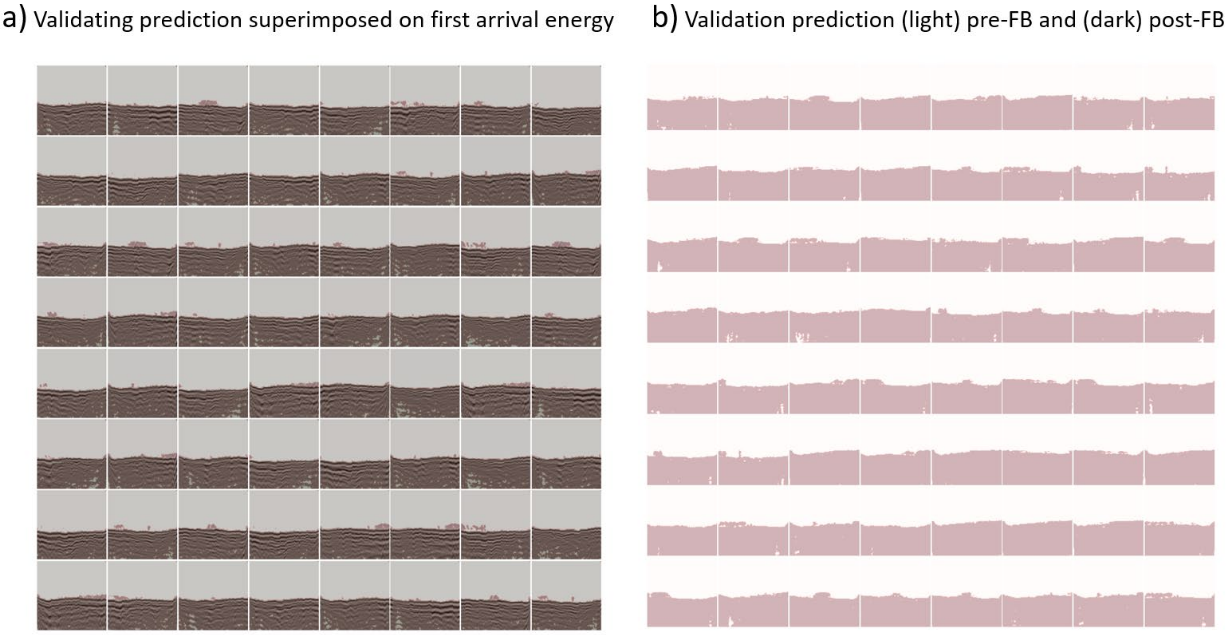


Figure 7. 12. a) Validation prediction results and first arrival energy displayed together, b) Validation prediction results.

7.3 Conclusion

When the first arrival energy is contaminated with noise, experienced technicians are required to confirm, or modified FB picks based on their experience in the regional near-surface geology and in recognizing FB trends through the complex first arrival waveform mixed with noises. With increasing data density, this has become a very time-consuming and expensive process. I used the vertical component of the multi-components Hussar 2D survey to demonstrate the application of machine learning in the most important part of near-surface velocity model building: FB picking. The first application is to reject outlying FB picks using three clustering algorithms, K-Means, GMM and DBSCAN. Only DBSCAN can reject outlying FB picks introduced in the test. The more important application of machine learning in FB picking is supervised deep learning. I define 2 classifications for the first arrival energy: pre-FB and post-FB. FB are picked automatically using the AIC method and edited interactively to create the training masks. With the Hussar dataset, I created 258 training images and masks of equal size. I use a simple UNET to model the first arrival events. The network converges after 15 epochs with 96.8% accuracy. We only used 258 training images in this test, while in a production processing environment one will have more than thousands of training images. Therefore, a better match between the FB mask and first arrival energy can be achieved in a production processing environment with more training images.

Chapter 8

Discussion and Conclusions

The algorithms and workflows for using additional information from reflection waveform for traveltime tomography in this thesis have been demonstrated to provide improved seismic images using synthetic and real field data.

Using traveltimes alone in refraction tomography limits the accuracy and resolution of the near-surface velocity model and statics corrections by the errors in the data, the model and the algorithms. It can be observed that the degradation to the near-surface velocity model and statics corrections deteriorate the coherence and the structural integrity of the deeper reflection data. The common procedure in conventional seismic data processing is to compute the residual statics corrections that can maximize the CDP stack power and apply the corrections to the reflection data. I presented a new workflow and a new refraction tomography kernel that use the long-wavelength component of the reflection residual statics to compute the data weight and model weight for the new refraction tomography algorithm. I verified the accuracy of this new method with both synthetic and real 2D field data.

With an increasing demand for higher data density, wider aperture and deeper depth, seismic blending has become a common practice in seismic acquisition. There have been many works done in direction deblending by inversion. However, the weathering corrections are required for land data for these inversion methods. I analyzed and effects of the seismic blending on refraction data, as well as, the effects of the refraction energy from blended shots on reflection data. I presented a robust workflow to model and remove refraction and reflection blending noises from blended shots. This workflow uses high-resolution sparse linear radon transform in the common receiver domain to model the refraction arrival and remove blending noises. It also uses high-resolution sparse parabolic Radon transform in the common CDP domain to remove blending noises on reflection data. I verified the accuracy of this workflow with synthetic and numerically blended 2D field data.

For reflection traveltime tomography, I investigated and verified the accuracy of the slope tomography. Unlike the classical reflection traveltime tomography that requires picking of continuous reflection events, slope tomography uses picks of reflection traveltimes and apparent slopes from the common shot and receiver gathers of any locally coherent events. This uses additional information from the shot and receiver ray parameters, and it also has the operational efficiency of not having to pick continuous reflection events. To verify the accuracy and efficiency of slope tomography, I used the Marmousi model to create synthetic data and automatically picked the traveltimes, shot and geophone ray parameters for all locally coherent events based on semblance. The picks were used to invert for a grid-based starting model for FWI. The result of FWI is satisfactory up to the depth of 1.6Km. The reason for the problem at greater depth is potentially caused by the quality of the picks at greater depth and the accuracy of the FWI algorithm. For non-structural data, I took a different approach. I automatically tracked the reflection arrival times for selected horizons using near-offset CDP stack and moveout corrected CDP gathers. The shot and geophone ray parameters were then computed using the arrival time picks. This approach was first tested on a synthetic dataset created using a simple model that consists of four constant velocity layers and with one of the layers has a non-zero dip. The velocity of the slope tomography inversion matched the true model, and the scatter points also matched the reflection boundaries. I used a similar approach with the vertical component of the Hussar 2D survey. Two nearby well logs were used in this test. Well tops were correlated with CDP stacks to identify the important reflectors. The velocity from the slope tomography inversion matched the long-wavelength trend of the well logs. Depth migrated section of the field data using the velocity model from slope tomography also tied the well logs.

Finally, I use machine learning to address the biggest challenge for refraction tomography: first break picking. Machine learning methods can be divided into unsupervised and supervised learning methods. I used unsupervised learning to learn the trend of the first break picks within a shot record. Three methods, K-Mean, GMM and DBSCAN were used to reject some outlying picks, only DBSCAN was able to reject the outliers. The most important application of machine learning for first break picking is to program the unsupervised deep learning network to learn from training models and masks and to automatically pick new datasets. I used the AIC method to automatically pick the first breaks for the Hussar 2D dataset, and then manually edited all the picks. The first break masks are created for each shot with two classifications, pre-first break and post-first break. This resulted in 258 training images and masks. 80% of this was used as training images and 20% of this was used as validation images. A simple UNET was set up to run the validation test with 60 epochs. The network converged after 15 epochs with 96.8% accuracy. Other metrics may be required to asses the quality of the results, but in general, higher accuracy is expected when more training images are available.

8.1 Future Work

Three potential future lines of work remain on traveltime tomography using feedbacks from reflection waveform and machine learning.

Most of my works have been evolved around 2D p-wave refraction tomography with the exception of slope tomography. Therefore, the first line of work is to extend the methods I presented to 3D converted wave. The second line of work is to investigate further slope tomography, including new and more efficient methods. The third line of work is to use more training images in the UNET validation test.

REFERENCES

Abma, R., Howe, D., Foster, M., Imtiaz, A., Tanis, M., Zhang, Q., Arogunmati, A., & Alexander, G. (2015). Independent simultaneous source acquisition and processing. *Geophysics, 80*(6), WD37-WD44. https://doi.org/10.1190/geo2015-0078.1.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255-265.

Barry, K. M. (1967). Delay time and its application to refraction profile interpretation. In A.W. Musgrave, (Eds.), *Seismic refraction prospecting* (pp. 348-361). Society of Exploration Geophysicist. https://doi.org/10.1190/1.9781560802679.

Beasley, C.J., Chambers, R.E., & Jiang, Z. (1998). A new look at simultaneous sources. *SEG Technical Program Expanded Abstracts,* 133-135. https://doi.org/10.1190/1.1820149.

Belfer, I., & Landa, E. (1996). Shallow velocity-depth model imaging by refraction tomography. *Geophysical Prospecting, 44,* 859-870. https://doi.org/10.1111/j.1365-2478.1996.tb00177.x.

Berkhout, A.J. (2008). Changing the mindset in seismic data acquisition*: The Leading Edge, 27*(7), 924-938. DOI: 10.1190/1.2954035.

Beydoun,W.B., & Keho,T.H. (1987). The paraxial ray method. *Geophysics, 52*(12),1601-1721. https://doi.org/10.1190/1.1442281.

Bilette, F., & Lambaré, G. (1998). Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International, 135,* 671-690. https://doi.org/10.1046/j.1365-246X.1998.00632.x.

Bilette, F., Le Bégat, S., Podvin, P., & Lambaré, G. (2003). Practical aspects and applications of 2D stereotomography. *Geophysics, 68*(3), 782-1103. https://doi.org/10.1190/1.1581072.

Birdus, S., Criddle, D., Artyomov, A., & Tang, Q. (2013). Successful application of joint reflection/refraction tomographic inversion in a shallow water marine environment. *ASEG Extended Abstracts, 2013*(1), 1-4. https://doi.org/10.1071/ASEG2013ab036.

Bishop, T.N., Bube, K.P., Cutler, R.T., Langan, R.T., Love, P.L., Resnick, J.R., Shuey, R.T., Spindler, D.A. & Wyld, H.W. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics, 50*(6), 892-1047. https://doi.org/10.1190/1.1441970

Borsdorf, A., Raupach, R., Flohr, T., & Hornegger, J. (2008). Wavelet based noise reduction in CT-images using correlation analysis. *IEEE transactions on medical imaging*, *27*(12), 1685-1703.

Brougois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P., & Versteeg, R. (1990). Marmousi, model and data. In *EAGE workshop-practical aspects of seismic data inversion* (pp. cp-108). European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.201411190.

Bunks, C., Saleck, F. M., Zaleski, S., & Chavent, G. (1995). Multiscale seismic waveform inversion: Geophysics.

Červený , V. (2001). *Seismic ray theor*y. Cambridge University Press. https://doi.org/10.1017/CBO9780511529399.

Červený , V., & Hron, F., (1980). The ray series method and dynamic ray tracing system for three-dimensional inhomogeneous media. *Bulletin of the Seismological Society of America, 70*(1), 47-77. https://doi.org/10.1785/BSSA0700010047.

Červený, V., Pretlová, V., & Pšenčik, I. (1977). Computation of ray amplitudes of seismic body waves in vertically inhomogeneous media. *Studia Geophysica et Geodaetica*, *21*(3), 248-255.

Claerbout, J. F. (1985). *Imaging the earth's interior* (Vol. 1). Oxford: Blackwell scientific publications.

Claerbout, J. (1992). Earth soundings analysis: *Processing versus inversion.* Blackwell Scientific Publications.

Chiu, S.K.L., & Stewart, R.R. (1987). Tomographic determination of three-dimensional seismic velocity structure using well logs, vertical seismic profiles, and surface seismic data. *Geophysics, 52*(8), 1033-1165. https://doi.org/10.1190/1.1442374.

Dance, D.R., Christofides, S., Maidment, A.D.A., McLean, I.D., & Ng, K.H. (2014). *Diagnostic radiology physics: A handbook for teachers and students.* International Atomic Energy Agency.

Devaney, A.J. (1982). A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic Imaging, 4*(4), 336-350. https://doi.org/10.1016/0161-7346(82)90017-7.

Docherty, P. (1992). Solving for the thickness and velocity of the weathering layer using 2-D refraction tomography. Geophysics, *57*(10), 1260-1380. https://doi.org/10.1190/1.1443198.

Gardner, L.W. (1939). An aerial plan of mapping subsurface structure by refraction shooting: Geophysics, *4*, 247-259. https://dx.doi.org/10.1190/1.1440501.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, Inc.

Hagedoorn, J.G. (1959). The plus-minus method of interpreting seismic refraction sections. *Geophysical Prospecting, 7*(2)*,* 158-182. https://doi.org/10.1111/j.1365-2478.1959.tb01460.x.

Hampson, D., & Russell, B. (1984). First-break interpretation using generalized linear inversion. *Journal of the Canadian Society of Exploration Geophysicists, 20*(1)*,* 40-54.

Hampson, D. (1986). Inverse velocity stacking for multiple elimination. *SEG Technical Program Expanded Abstracts,* 422-424. https://doi.org/10.1190/1.1893060.

Hampson, G., Stefani, J., & Herkenhoff, F. (2008). Acquisition using simultaneous sources. *SEG Technical Program Expanded Abstracts,* 2816-2821. https://doi.org/10.1190/1.3063930.

Hestenes, M.R., & Stiefel, E. (1952). Methods of conjugate gradients for solving. *Journal of Research of the National Bureau of Standards*, *49*(6), 409-436. https://dx.doi.org/10.6028/jres.049.044.

Huang, G., & Symes, W. W. (2015, October). Full waveform inversion via matched source extension. In *2015 SEG Annual Meeting*. OnePetro.

Ibrahim, A., & Saachi, M. D. (2013). Simultaneous source separation using a robust Radon transform. *Geophysics, 79*(1). https://doi.org/10.1190/geo2013-0168.1.

Ibrahim, A., & Saachi, M. D. (2015). Fast simultaneous seismic source separation using Stolt migration and demigration operators*. Geophysics, 80*(6), WD27-WD36, doi: 10.1190/geo2015-0044.1.

Ibrahim, A., & Trad, D. (2020). Inversion-based deblending using migration operators. *Geophysical Prospecting, 68*(8), 2459-2470. https://doi.org/10.1111/1365-2478.13015.

Jones, I.F. (2012). Tutorial: Incorporating near-surface velocity anomalies in pre-stack depth migration models. *First Break, 30*, 47-58. https://doi.org/10.3997/1365-2397.2011041.

Kosloff, D., Pecholcs, P., Zackhem, U., & Koren, Z. (1997). Estimation of long wavelength near surface velocity and low-relief structure anomalies Part 2: A new near surface reconstruction method. *SEG Technical Program Expanded Abstracts, 16,* 1063- 1065.

Lambare, G., & Allemand, T. (2015, June). Strength and Limitation of FWI-What Can Tomographic Approaches Bring?. In *77th EAGE Conference and Exhibition-Workshops* (Vol. 2015, No. 1, pp. 1-4). European Association of Geoscientists & Engineers.

Langan, R.T., Lerche, I., Cutler, R.T., Bishop, T.N., & Spera, N.J. (1984). Seismic tomography: The accurate and efficient tracing of rays through heterogenous media. *SEG Technical Program Expanded Abstracts,* 713-715. https://doi.org/10.1190/1.1894309.

Law B.K., & Trad, D. (2018). Robust refraction statics and near surface velocity model building using feedback from reflection data. *Geophysics, 83*(6), U63-U77. https://doi.org/10.1190/geo2018-0060.1.

Lines, L.R., & Treitel, S. (1984). A review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting, 32*, 159-186. https://doi.org/10.1111/j.1365-2478.1984.tb00726.x.

Lo, T.W., & Inderwiesen, P. (1994). *Fundamentals of seismic tomography.* Society of Exploration Geophysicists.

Lloyd, S. (1982). Least squares quantization in PCM. Institute of Electrical and Electronics Engineers, *28*(2), 129-137. https://doi.org/10.1109/TIT.1982.1056489.

Mahdad, A., Doulgeris, P., & Blacquiere, G. (2011). Separation of blended data by iterative estimation and subtraction of blending interference noise. *Geophysics, 76*(3), Q9-Q17. https://doi.org/10.1190/1.3556597.

Manning, T., & Ahmad, K. (2013). The impact of ISS® blending and source separation on refraction statics modeling. *SEG Technical Program Expanded Abstracts,* 114- 118. https://doi.org/10.1190/segam2013-1194.1.

Margrave, G., Bertram, M., Bertram, K.L., Hall, K., Innanen, K., Lawton, D., Mewhort, L., & Phillips, T.M. (2012). A low-frequency seismic field experiment. *SEG Technical Program Expanded Abstracts,* 454-457. https://doi.org/10.1190/SEGAM2012-0859.1.

Matsuoka, T., & Ezaka, T. (1992). Ray tracing using reciprocity. *Geophysics, 57*, 326-333. https://doi.org/10.1190/1.1443246.

Moore, I., Dragoset, B., Ommundsen, T., Wilson, D., Ward, C., & Eke, D. (2008). Simultaneous source separation using dithered sources. *SEG Technical Program Expanded Abstracts,* 2806-2810. https://doi.org/10.1190/1.3063928.

Ng, M., & Perz, M. (2004). High resolution Radon transform in t-x domain using “intelligent” prioritization of the Gauss-Seidel estimation sequence. *SEG Technical Program Expanded Abstracts,* 2160-2163. https://doi.org/10.1190/1.1845211.

Nocedal, J., & Wright, S.J. (2006). Conjugate gradient methods. (Eds.), *Numerical optimization* (pp.101-134). Springer New York, NY. https://doi.org/10.1007/978-0-387-40065-5.

Osher, S., & Sethian, J.A. (1988). Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics, 79*(1), 12-49. https://doi.org/10.1016/0021-9991(88)90002-2.

Palmer, D. (1981). An introduction to the generalized reciprocal method of seismic refraction interpretation. *Geophysics, 46*(11), 1508-1518. https://doi.org/10.1190/1.1441157.

Pecholcs, P.I., Nguyen, S., Kosloff , D., & Litvin, A. (1997). Estimation of long wavelength near surface velocity and low-relief structure anomalies Part 1: A case history in central Saudi Arabia. *SEG Technical Program Expanded Abstracts,*1059-1062. https://doi.org/10.1.1.1073.6724.

Qin, F., Luo. Y., Olsen, K.B., Cai, W., and Schuster, G.T. (1992). Finite-difference solution of the eikonal equation along expanding wavefronts. *Geophysics, 57*(3), 478-487. https://doi.org/10.1190/1.1443263.

Riabinkin, L.A. (1957). Fundamentals of resolving power of controlled direction reception (CDR) of seismic waves. In G.H.F. Gardner & L. Lu (Eds.), *Slant-Stack Processing* (pp. 3-36).

Rieber, F. (1936). A new reflection system with controlled direction sensitivity. *Geophysics, 1*(1), 97-106. https://doi.org/10.1190/1.1437082.

Ronneberger, O., Fischer, P., & Brox, T. (Eds.). (2015). U-net: Convolutional networks for biomedical image segmentation. In: N.Navab, J. Hornegger, W. Wells, & A. Frangi. (Eds). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234-241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28.

Ronen, J., & Claerbout, J.F. (1985). Surface-consistent residual statics estimation by stack-power maximization. *Geophysics, 50*(12), 2759-2767. https://doi.org/10.1190/1.1441896.

Rouy, E., & Tourin, A. (1992). A viscosity solution approach to shape-from-shading. *SIAM Journal on Numerical Analysis, 29*(3), 867-884. https://doi.org/10.1137/0729053.

Sethian, J.A., & Popovici, A.M. (1999). 3-D traveltime computation using the fast marching method. *Geophysics, 64*(2), 516-523. https://doi.org/10.1190/1.1444558.

Stewart, R.R. (1991). *Exploration Seismic Tomography: Fundamentals*. Society of Exploration Geophysics. https://doi.org/10.1190/1.9781560802372.

Sword C. H. (1987). *Tomographic determination of interval velocities from reflection seismic data: The method of controlled directional reception* (Publication No. 8801043) [Doctoral dissertation, Stanford University]. Stanford University ProQuest Dissertations Publishing.

Taner, M.T., Koehler, F., & Alhilali, K.A. (1974) Estimation and correction of near-surface time anomalies. *Geophysics, 39*(4), 441-463. https://doi.org/10.1190/1.1440441.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics, 49*(8), 1259-1266. https://doi.org/10.1190/1.1441754.

Thorson, J.R., & Claerbout, J.F. (1985). Velocity-stack and slant-stack stochastic inversion. *Geophysics, 50*(12), 2727-2741, https://doi.org/10.1190/1.144893.

Trad, D., Ulrych, T., & Saachi, M. (2003). Latest views of the sparse Radon transform. *Geophysics, 68*(1), 386-399. https://doi.org/10.1190/1.1543224.

Trad, D., Siliqi, R., Poole, G., & Boelle, J.L. (2012). Fast and robust deblending using apex shifted Radon transform. *SEG Technical Program Expanded Abstracts,* 1-5. https://doi.org/10.1190/segam2012-0703.1.

Trad D.O. (2020). A multigrid approach for time domain FWI. *CREWES Research Report*, *32*(54), 20.

Verschuur, E. (2011). DELPHI Sponsor Meeting, The Hague, Holland.

Vidale, J.E. (1988). Finite-difference calculation of travel times**.** *Bulletin of the Seismological Society of America, 78*(6), 2062-2076. https://doi.org/10.1785/BSSA0780062062.

Vidale, J. E. (1990). Finite-difference calculation of traveltimes in three dimensions. *Geophysics, 55*(5), 521–526. https://doi.org/10.1190/1.1442863.

Vinje, V., Iversen, E., & Gjøystdal, H. (1993). Traveltime and amplitude estimation using wavefront construction. *Geophysics, 58*(8), 1157-1166. https://doi.org/ 10.1190/1.1443499.

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

Wapenaar, K., van der Neut, J., & Thorbecke, J. (2012). Deblending by direct inversion. *Geophysics, 77*(3), A9-A12. https://doi.org/10.1190/geo2011-0497.1.

Warner, M., & Guasch, L. (2014, June). Adaptive waveform inversion-FWI without cycle skipping-theory. In *76th EAGE Conference and Exhibition 2014* (Vol. 2014, No. 1, pp. 1-5). European Association of Geoscientists & Engineers.

White, D. (1989). Two-dimensional seismic refraction tomography. *Geophysical Journal International, 97*(2), 223-245. https://doi.org/10.1111/j.1365-246X.1989.tb00498.x.

Wu, R. S., & Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, *52*(1), 11-25.

Zhou, H., Gray, S.H., Young, J., Pham, D., & Zhang, Y. (2003). Tomographic residual curvature analysis: The process and its components. *SEG Technical Program Expanded Abstracts,* 666-669. https://doi.org/10.1190/1.1818018.

Zhu, X., Sixta, D.P., Angstman, B.G. (1992)., Tomostatics: Turning-ray tomography + static corrections. *The Leading Edge, 11*(12), 15-23. https://doi.org/10.1190/1.1436864