

**MAPPING NEAR-SURFACE SEISMIC VELOCITIES AND Q VALUES WITH FIRST  
ARRIVAL TOMOGRAPHY**

by  
Xinwei Huang

A dissertation submitted to the department of earth and atmospheric sciences

in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

in Geophysics

Chair of Committee: Hua-Wei Zhou

Committee Member: Yingcai Zheng

Committee Member: Castagna John

Committee Member: Jonathan Liu

University of Houston

May 2021

Copyright 2021, Xinwei Huang

## ACKNOWLEDGMENTS

Firstly, I would like to express my special appreciation and thanks to my advisor Professor Dr. Hua-Wei Zhou, you have been a tremendous mentor for me ever since my graduate career began in Lubbock. I sincerely thank you for your care, guidance and supporting during both of my Master and Ph.D. study. Your profound knowledge and rigorous scientific spirit led and encouraged me to keep exploring the fascinating world of geophysics. Your advice on both research as well as on life have been benefit me a lot. Also, my special thanks to your wife, Mrs. Wendy, for every group party she organized, and it made our great research group like a big family.

I also owe my sincere appreciation to my committee members, professor Dr. Yingcai Zheng, professor Dr. Castagna John and Dr. Jonathan Liu for serving as my committee members through such hardship of worldwide epidemic of COVID-19. Those days spending on online meeting and exchange email to absorb your insightful comments and suggestions are priceless for me. It reminds of the beautiful old days when I was taught and guided by you face-to-face in the UH campus. I hope the haze of the epidemic can dissipate as soon as possible, and my best wishes to you and your family in the future.

I would like to tank BGP inc., for providing me such a great platform for utilizing high-performance computing to support my research works in China. In particular, I would like to thank Dr. Zhenbo Guo, Dr. Hequn Li and Dr. Yubo Yue for your generous helps and brilliant advices on my research topics. Your experience on programming and handling field data benefits me a lot.

I would also like to express gratitude to my friends and colleges, Zhonghan Liu, Boming Wu, Yao Yao, Fang yuan, Zhihui Zou, Luchen Li, Shuhang Tang, Zhehao Li, Boming Wu and many other people who had a great friendship with me and helped me a lot in both study and life in Lubbock

and Houston, making all these grateful memories vivid and always shining in my life.

My deepest gratitude goes to my parents and families. It was you who gave me life, educated and cultivated me to be an honest and optimistic person to never give up when facing challenges in study and life. Without your encouragement, patience, love and continuous support, I would not have finished this dissertation work and such long journey in academic.

The last but not the least, I would like to give my special thanks to Xiaoying Liu who I met one year ago. Thank you for allowing me to enter into your world and share your life with me. You are always the sunlight in those dim days.

## ABSTRACT

Near-surface model building is critical in exploration geophysics studies. Among various methods, first-arrival traveltimes (FAT) tomography is among the most popular solutions, especially in land surveys over complex geologic structures in highly noisy environments. Though this is a relatively mature method, there are still many practical challenges that have motivated my study.

The first part of my study focuses at finding a way to significantly improve the computational efficiency of FAT tomography, which has limited its usage for the increasingly massive data volumes in high density seismic surveys nowadays. Often, hundreds of millions of seismic picks are beyond the limit of standard FAT tomography methods in terms of computation memory and turnaround time, both are critical in practice. I have adopted an adjoint-state solution, which reduces the memory cost regardless of the quantity of input data. I have also devised a highly efficient FAT tomographic inversion scheme by combining the dimensionality reduction and the sparsity-promoting techniques based on a compressive sensing approach. The computation time cost is significantly reduced by taking randomly subsampled data for computation. The model update is regularized and many imaging artefacts induced by random subsampling are mitigated through exploiting its sparsity within learned dictionaries. My new inversion scheme enables to use just a small portion of a dataset to achieve results practically identical to those from standard FAT tomography methods using the full dataset.

The second part of my study is an attempt to reconstruct an  $Q$  distribution model, in order to compensate for the near-surface loss in the amplitude of seismic data. This is a significant factor for seismic image quality. My approach is based on the impact of path attenuation factor ( $t^*$ ) to the amount of amplitude attenuation on seismic first arrival waveforms. To accurately estimate  $t^*$ ,

I use a high-quality tomographic velocity model to guide the identification of the relevant first arrival waveforms. Then each  $t^*$  is estimated through an adaptive correction method based on linear regression of the logarithmic spectral ratio. Based on  $t^*$ , my adjoint-state Q tomographic inversion is able to reconstruct Q distribution model for the near-surface.

# TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
ABSTRACT	v
TABLE OF CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER 1 INTRODUCTION	1
1.1 Development of industrial seismic tomography	2
1.2 Challenges from near surface modeling using first arrival tomographic inversion	5
1.1 Computation efficiency for large dataset	6
1.2 Near-surface Q model reconstruction	8
1.3 Dissertation themes	9
1.3.1 Mapping velocity model: efficient first-arrival traveltimes tomography using large datasets	10
1.3.2 Mapping Q distribution model: first arrival attenuation tomography based on An Adjoint-state Method	11
1.4 Outline of dissertation	12
CHAPTER 2 FIRSTARRIVAL TRAVELTIME TOMOGRAPHY BASED ON ADJOINT-STATE METHOD	15
2.1 Introduction	15
2.2 Theory	19
2.2.1 Misfit function and its gradient	19
2.2.2 FAT tomographic inversion based on raytracing method	21
2.2.3 FAT tomographic inversion based on eikonal equation solver	23
2.3 Comparative study	26
2.3.1 2D synthetic model test	27
2.3.2 3D synthetic model test	34

2.4 Summary and conclusions	38
<b>CHAPTER 3 EFFICIENT FIRST ARRIVAL TRAVELTIME TOMOGRAPHIC INVERSION IN LARGE DATASET</b>	
3.1 Introduction	40
3.2 FAT tomographic inversion	41
3.2.1 Observation on misfit function	41
3.2.2 Linearized misfit function and Gauss Newton (GN) approach	43
3.2.3 Adjoint PDE and gradient of misfit function	47
3.4 Stochastic process	52
3.4.1 Sample average approximation (SAA)	53
3.4.2 Stochastic approximation (SA)	54
3.4.3 Reduced GN problem based on stochastic approximation	57
3.5 Sparsity-promoting by sparse orthonormal transformation (SOT)	57
3.5.1 Online orthonormal dictionary learning	58
3.5.2 Dictionary-based blockwise transformation	61
3.5.3 Practical implementation of online blockwise ODL	62
3.6 SA+ Method: a modified GN method with gradient optimization	68
3.6.1 Compressive sensing and signal reconstruction problem	70
3.6.2 LASSO-based approach and modified GN method	72
3.6.3 Practical implementation	75
3.7 Numerical test and real data application	78
3.7.1 Synthetic data test	78
3.8.2 Real data test	113
3.8 Summary and discussions	117
<b>CHAPTER 4 FIRST ARRIVAL ATTENUATION TOMOGRAPHY BASED ON ADJOINT- STATE METHOD</b>	
4.1 Introduction	119
4.2 Theory Background	122
4.2.1 Forward modeling: governing equation for $t^*$	123

4.2.2 Attenuation Tomography based on adjoint-state method	126
4.2.3 Estimation of attenuated time and adaptive correction method	129
4.3 Numerical test	134
4.3.1 2-D synthetic data test: two-layer model	134
4.3.2 2-D synthetic data test: simple model with topographic	155
4.3.3 2-D synthetic data test: Marmousi model	159
4.3.4 Field data test	163
4.4 Summary and discussions	171
CHAPTER 5 CONCLUSION AND DISCUSSION	176
5.1 Mapping near-surface velocities using efficient FAT tomography for large dataset	176
5.2 Mapping near-surface Q values using First arrival attenuation tomography based on adjoint state method	178
5.3 Outlook on future work	180
BIBLIOGRAPHY	183

## LIST OF TABLES

<b>Table 3-1:</b> The rms error and time cost for SAA, SA, SAA+ and SA+ method (clean data). .....	83
<b>Table 3-2:</b> The RMSE and time cost for SAA, SA, SAA+ and SA+ method (8% noise)....	83
<b>Table 3-3:</b> The RMS error and time cost for SA with proposed method (clean data).....	97
<b>Table 3-4:</b> The rms error and time cost for SA with proposed method (8% noise).....	97
<b>Table 3-5:</b> Comparison of computational efficiency in memory occupation and time cost. .....	110

## LIST OF FIGURES

<b>Figure 1-1:</b> Growth of pre-stack seismic data during last 20 years. ....	7
<b>Figure 2-1:</b> True velocity model.....	29
<b>Figure 2-2:</b> Linear smooth initial velocity model.. ....	29
<b>Figure 2-3:</b> Inversion results from raytracing-based FAT tomographic inversion. ....	31
<b>Figure 2-4:</b> Inversion results from FAT tomographic inversion based on raytracing. ....	31
<b>Figure 2-5:</b> Ray density distribution at last iteration of tomographic inversion based on ray tracing method. ....	31
<b>Figure 2-6:</b> Computational efficiency comparison .....	33
<b>Figure 2-7:</b> Overthrust 3D velocity model and initial velocity model.....	35
<b>Figure 2-8:</b> Estimated 3D velocity model at depth of 40m.....	36
<b>Figure 2-9:</b> Estimated 3D velocity model at depth of 100m.....	37
<b>Figure 2- 10:</b> Estimated 3D velocity model at depth of 150m.....	38
<b>Figure 3- 1:</b> Schematic sparsity promoting FAT tomography workflow.....	41
<b>Figure 3-2:</b> Learned dictionaries by ODL for signals with different gradient angle features. .....	63
<b>Figure 3-3:</b> Example of velocity model with high and low velocity anomalies . ....	65
<b>Figure 3-4:</b> Learned dictionaries by ODL with different values. ....	66
<b>Figure 3-5:</b> NLA test results from two model updates.....	68

<b>Figure 3-6:</b> Synthetic 2D model.....	80
<b>Figure 3-7:</b> First arrival picks in clean and noisy data.....	84
<b>Figure 3-8:</b> Error bar curves of clean data and 8% noisy data.....	85
<b>Figure 3-9:</b> Convergence history of SAA, SA and SA+ with clean and 8% noisy data ....	87
<b>Figure 3-10:</b> Inverted velocity models.....	90
<b>Figure 3-11:</b> Comparison of gradient before and after optimization by our proposed method .....	92
<b>Figure 3-12:</b> The velocity functions at the center of high and low velocity anomaly inverted by SA and SA+ .....	94
<b>Figure 3-13:</b> Inverted velocity models with 7% and 10% data decimation at 20%-time cost of standard method for clean and 10% noisy data test respectively. ....	99
<b>Figure 3-14:</b> The relationship between the noise level and the data-decimation percentage for a fixed RMSE at 100 m/s.....	100
<b>Figure 3-15:</b> Ray density distribution in velocity model of final iteration and the velocity variation of each cell in SA+ with 7% of clean data and 10% of noisy data.....	102
<b>Figure 3-16:</b> True velocity model and linear increasing initial velocity model.....	103
<b>Figure 3-17:</b> Velocity model reconstructed by standard tomographic inversion (100% data), SA (10%), SAA+ (10% data), SA+ (10% data) dataset and SA+ (50% data).....	107
<b>Figure 3-18:</b> Velocity profiles at depth of 50m and 100m. ....	108
<b>Figure 3-19:</b> The RMS travelttime residual for standard method (100% data), SA+ (10% data)	

and SAA+ (10% data).....	109
<b>Figure 3-20:</b> One of the gradients used for standard method (100% data), Standard method (10% data) and after gradient optimization (10% data).....	112
<b>Figure 3-21:</b> The original synthetic FAT picks blended with random noise ranged around [-0.3s, 0.3s]. and its reconstructed velocity model by SA+ with 10% dataset .....	113
<b>Figure 3-22:</b> Inverted results by standard method (100% data) and SA+ (20% data) .....	115
<b>Figure 3-16:</b> Stack image after static correction corresponding to the velocity model obtained from (a) standard method (100% data) and SA+ (20% data). .....	116
<b>Figure 4- 1:</b> Comparison of $t^*$ calculated by two forward modeling method. ....	126
<b>Figure 4-2:</b> Waveform and Spectrum of different type of waves. ....	132
<b>Figure 4-3:</b> Two-layer velocity model.....	135
<b>Figure 4-4:</b> Original shot gather. ....	135
<b>Figure 4-5:</b> Seismic records at 600 m and 1800 m.....	136
<b>Figure 4-6:</b> Direct and refraction waveform extracted from time window of 0.0 to 0.6s, and aligned by emergent onset point. ....	138
<b>Figure 4-7:</b> The comparison of seismic waveforms of original direct wave, time domain integral of original direct wave and refracted wave.....	140
<b>Figure 4-8:</b> The comparison of amplitude spectrums of original direct wave, time domain integral of original direct wave and refracted wave.....	141
<b>Figure 4-9:</b> True Q model.....	143

<b>Figure 4-10:</b> Comparison of shot gathers with Q attenuation and without Q attenuation.	143
<b>Figure 4-11:</b> Comparison of seismic traces at 200 m, 600 m and 1665 m (without Q attenuation and with Q attenuation).....	144
<b>Figure 4-12:</b> First arrivals of seismic trace at 200 m, 600 m and 1665 m (without Q attenuation and with Q attenuation).....	145
<b>Figure 4-13:</b> Amplitude spectrum of direct wave with and without Q attenuation and spectral ratio method result. ....	146
<b>Figure 4-14:</b> Amplitude spectrum of direct wave with and without Q attenuation and spectral ratio method result without <i>Ppf</i> and with <i>Ppf</i> . . . . .	148
<b>Figure 4-15:</b> Reference signals and similarity of their synthetic seismic trace. ....	149
<b>Figure 4-16:</b> Original shot gather (interference of direct and refracted wave). ....	151
<b>Figure 4-17:</b> Normalized first arrival fatten out.....	151
<b>Figure 4-18:</b> Estimated $t^*$ based on direct wave and refraction.....	152
<b>Figure 4-19:</b> Similarities calculated based on direct wave and refraction. ....	153
<b>Figure 4-20:</b> Final estimated $t^*$ . ....	153
<b>Figure 4-21:</b> Inverted $1/Q$ distribution model.....	155
<b>Figure 4-22:</b> A 2-D synthetic data test: the true velocity model. ....	156
<b>Figure 4-23:</b> The true attenuation model ( $1/Q$ )......	156
<b>Figure 4-24:</b> Comparison of estimated $t^*$ and theoretical $t^*$ .....	157

<b>Figure 4-25:</b> Inverted Q model using the true velocity model. ....	158
<b>Figure 4-26:</b> Inverted velocity model using first arrival tomography. ....	158
<b>Figure 4-27:</b> Inverted Q model derived by Q tomographic inversion using inverted velocity model. ....	159
<b>Figure 4-28:</b> 2-D synthetic data test on Marmousi model: (a) True velocity model; (b) true attenuation model; (c) inverted velocity model by first arrival travelttime tomography; (d) inverted Q model derived by true velocity model of (a); (e) inverted Q model derived by inverted velocity model of (c). ....	163
<b>Figure 4-29:</b> Survey geometry map: source and receiver position. ....	163
<b>Figure 4-30:</b> Elevation map of survey area. ....	164
<b>Figure 4-31:</b> Original shot gather. ....	165
<b>Figure 4-32:</b> Frequency division scanning of original shot gather: (a) 10-25Hz frequency range shot gather; (b) 25-60Hz frequency range shot gather. ....	166
<b>Figure 4-33:</b> Original CMP stacking image results: (a) Original CMP stacking; (b) amplitude spectrum of left panel in (a); (c) amplitude spectrum of right panel in (a). ....	167
<b>Figure 4-34:</b> Inverted velocity model derived by FAT tomography. ....	168
<b>Figure 4-35:</b> Inverted attenuation model (1/Q). ....	170
<b>Figure 4-36:</b> CMP stacking after attenuation compensation. (a) Attenuation compensated CMP stacking; (b) amplitude spectrum of left panel in (a); (c) amplitude spectrum of right panel in (a). ....	171



---

## CHAPTER 1 INTRODUCTION

The earliest seismic tomography was suggested by seismologist serving for the global scale studies of solid earth several decades ago (Aki and Lee, 1976; Aki et al., 1977; Dziewonski et al., 1977). Similar to medical computerized tomography (CT) technique, seismic tomography will generate a cross-sectional picture using observed data outside the geological target of interest. This data inference technique generally aims for generating a heterogenous seismic model that is consistent with observations by solving for a substantial inverse problem (Rawlinson et al., 2010). Following the similar manners of physics and mathematics, it has been adopted and extensively used for investigating on regional scale problems of hydrocarbon exploration and production from oil industry. From the perspective of the research area, most of industrial seismic tomography application could be classified as “local” tomography, which is using temporary deployments of receiver and artificial source (e.g., explosions, vibroseis and airguns) to investigate the targets within a certain geographical region ranged from crust to upper mantle. One possible earliest industrial application of seismic tomography is by Boris (1977), from which a travelttime tomography method developed for imaging 2D velocity structure in between a cross-hole. In the following three decades, a series of developments for industrial seismic tomography was soon caught up. Developments of seismic tomography were pondered in all details of forward modeling method, inversion algorithm and assessing solution non-uniqueness.

---

## 1.1 Development of industrial seismic tomography

At an early stage, various of the cross-hole tomography techniques using ray tracing and inversion scheme was proposed (McMechan, 1983, 1987; Bregman et al., 1989). In these applications, the back-projection inversion techniques were more prevailed than the gradient-based one due to its specific acquisition geometry. Reflection and wide-angle (or refraction) tomography (Pratt and Worthington, 1988; Song et al., 1995; Pratt and Shipp, 1999) are also widely used in the industrial seismic exploration. Bishop et al. (1985) parameterized a 2D subsurface model into constant velocity blocks cubic separated by spline interfaces, which allow to constrain both of velocity and interface variation during tomographic inversion. Later on, series of similar studies (e.g., Farra and Madariaga, 1987; Williamson, 1990) were conducted under similar framework. Similar to reflection tomography, wide-angle (or refraction) tomography adopted. Wide-angle (or refraction) tomography could adopt with much longer offset survey to preserve refraction signal at significant depth. The development of tomographic inversion is inseparable from the continuous updating of forward modeling method. The early work of forward modeling method is focusing on the two-point ray tracing algorithm to generate synthetic refraction traveltimes (White, 1989) and solve for both velocities and refractor depths. It earns success for many cases with a relatively smooth velocity structure (Červený, 1987), but often fails in refraction travel time calculation for some complex geological area as much wave effect involved in. An alternative forward modelling method is wavefront ray tracing method, which is characterizing local ray and minimum-traveltime

---

wavefront were developed to improve the accuracy and efficiency of travel time calculation in some complex velocity medium (Vidale, 1988; Fischer and Lees, 1993; Weber, 1995), especially in the rapid velocity variation or shadow zone. Many wavefront expansion methods were developed in this time of period, such as finite-difference (FD) method to solve the eikonal equation (Vidale, 1990; Qin et al., 1992; Hole and Zelt, 1995), wavefront expansion based on analytical solution (Vinje et al., 1993) or minimum-traveltime paths from graph theory (Saito, 1990; Moser, 1991; Zhang, 1998).

Into the 21<sup>st</sup> century, the grid-based solvers of eikonal equation method have been dramatically improved by many authors regarding to accurate and efficient travelttime calculation. There are currently two competing algorithms used to solve the eikonal equation: Fast Marching Method (FMM) (Rawlinson and Sambridge, 2004) and Fast Sweeping Method (FSM) (Tsai et al., 2003; Zhao, 2004; Taillander et al., 2009). One common improvement of two methods is that they could expand the wavefront in a monotonic manner to ensure the causality, other than evolving a non-physical square wavefront expansion as original FD eikonal equation solver (Aldridge and Oldenburg, 1992; Schneider et al., 1992).

Either raytracing or wavefront expansion forward calculation method is founded on the geometric ray theory. Under high frequency assumption, a robust travelttime approximation requires that the seismic wavelength is much smaller than the scale length of heterogenous target body, which is not always the case. It is an unphysical assumption since that the seismic signal possesses a certain frequency bandwidth. Eventually, the finite frequency effect will degrade the final image resolution. An intermediate method to solve this issue is to calculate

---

the travel time using wave-equation without high frequency assumption, while keeping the robust convergence of quasi-linear traveltimes inversion. Based on this, wave-equation tomography (WET) (Luo and Schuster, 1991; Wang et al., 2012) was developed to bridge the gap between the extremes of traveltimes inversion and full-wave inversion. WET utilizing both reflection and refraction data could offer a correct kinematics initial model to improve the convergence of FWI (Wang et al., 2013). For near surface application, early arrival WET (Zhou and Greenhalgh, 2003; Sheng et al., 2006) was proposed to include more general wave effects for better image while preserving robust convergence succeeded by early arrival misfit function.

Finite frequency tomography employs Born scattering theory to the frequency dependence of traveltimes (Dahlen et al., 2000), and improves the image quality in general heterogeneous media. Unlike ray-based method, its “banana doughnut” kernel (de Hoop and van der Hilst, 2005a, b; Marquering et al., 1999; Montelli et al., 2006) could depict the responses throughout the model space but not only along ray trajectories. It could make better use of both phase (Nolet, 2008) amplitude information (Sigloch et al., 2008) from seismic data. Literature and different specific subject areas in seismic tomography are monumental and such short review cannot cover every aspect of such large and diverse field. Especially, development in seismological field is not covered in this short section, from which many related techniques were proposed. A few excellent review articles and books include Stewart (1991), Lehmann (2007), Nolet (2008), Vensnaver (2010, 2013) and Rawlinson et al. (2008, 2010).

---

## 1.2 Challenges from near surface modeling using first arrival tomographic inversion

As summarized in the previous section, generally, there are two kinds of tomography depending on projection data: traveltimes and waveform tomography. Both ray and waveform modeling could describe seismic wave phenomenon. In areas with strong lateral variations in topography and near-surface velocities, FAT tomography is one of the most popular methods adopted. Generally speaking, FAT information is extracted from refraction data or diving waves. One significant advantage of utilizing FAT, especially for land survey, is that it might be the most feasible information to access, since that other secondary events (e.g., reflection) are often contaminated by noise to detect. In some poor-quality data, FAT is even the only available data to be identified clearly to reconstruct near-surface models.

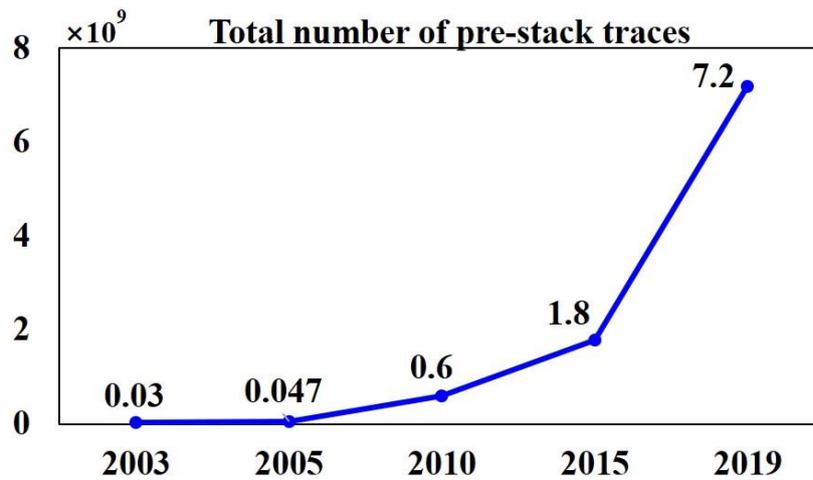
It is a well-established inverse modeling technique for determining shallow velocity structures, and a common approach to investigate deep structures. Good near-surface velocity models would also provide static corrections of reflection seismic data in complex media, and as initial velocity models for pre-stack depth migration and full waveform inversion. Though a solid theory background has already been built up for FAT, there are still some on-going challenges needed to be resolved. In my dissertation, we mainly focus on solving two common issues existing in near-surface modeling using FAT tomographic inversion: 1.) an efficient FAT tomographic inversion scheme for large datasets of current industrial problem size and 2.) attenuation distribution model building using first arrival tomographic inversion.

---

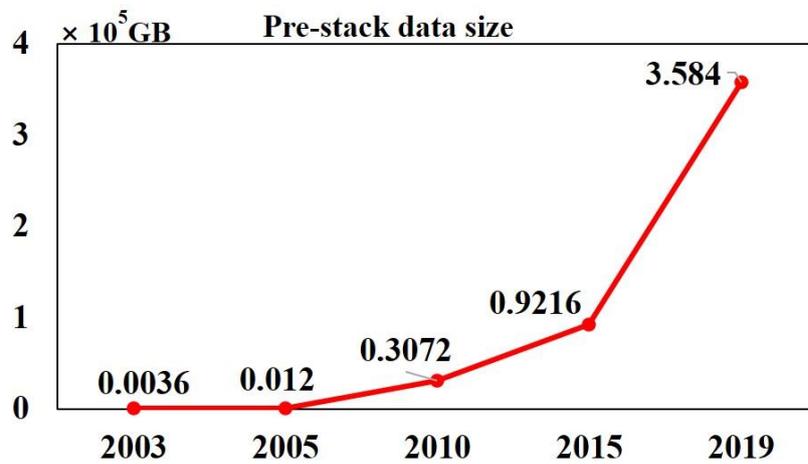
## 1.1 Computation efficiency for large dataset

From 1960s, digital computing and microprocessor technology started a rapid advance following the Moore's law (Moore, 1965), in which the number of transistors in microprocessor will approximately double every two years. This stunning development is still continuing as cluster and multiple core processors becoming more popular recently. Correspondingly, high-density seismic acquisition is becoming mainstream in seismic exploration technology (Matheny et al., 2009). The high density is realized by reducing the bin size and increasing spatial sample rate. It brings significant help on identifying finer geological structures (e.g., a small fault block, thin sand body and reservoir) and improving image resolution.

Generally, a high-density acquisition involves with vast quantities of seismic data recorded and archived. Figure 1-1 shows the noticeable growth of both seismic data size and the total number of pre-stack traces during past 20 years in land survey. As over 10 thousand channels seismograph popularization, a typical number of acquired trace could be simply over thousands of millions. It consequently results in millions of travel time picks and any type of parameterizations from a velocity model (Vesnaver, 2008; Noble et al., 2010; Sun and Zhang, 2017).



(a)



(b)

**Figure 1.1:** (a) Increase in the number of pre-stack traces ( $10^9$ ) and (b) data size (GB) of a typical land seismic survey during last 20 years.

Additionally, in some complex geological structure area, the near surface will exhibit strong variation in both lateral and vertical directions, which are usually giving rise to a short wavelength seismic wave propagation requesting a finer discretization of both datasets and model parameters. As a consequence, even though such large volume data will bring many benefits in image solution, such as wide aperture and more dense spatial sampling to improve imaging quality etc., the computation cost of tomographic method which is proportional to the

---

number of source and receiver deployments, will dramatically increase as burdens coming from both large memory occupations and long computation time respectively. Intuitively, one could subsample the seismic data and model parameters to relief such computational burdens. However, direct dimensionality reduction implementation to this problem would lead to either a loss of salient information or a poor resolution.

## **1.2 Near-surface Q model reconstruction**

Another issue that commonly involved with near-surface imaging is the appearance of attenuation effect quantified as the quality factor (Q) during propagation of seismic waves inside the subsurface medium. It is often caused by overlying unconsolidated soil or the overburden gas clouds, by which the seismic energy will be severely dissipated leading to the image solutions with degraded illumination and resolution. Since then, this Q factor should be compensated in order to improve image resolution and obtain correct amplitude and phase information to make their identification and interpretation more feasible.

A reliable estimation of the Q model is one of essential steps for a successful application of Q compensation imaging. The early attempt of using inverse Q filter using deconvolution method (Hargreaves & Calvert 1991; Varela et al. 1993; Wang 2002, 2006; Zhang & Ulrych 2007) could be incorporated into process flow to restore the resolution to some extent; however, it failed to handle with high geological complexity of large area in practical. A reflection-based attenuation tomography (Xin et al., 2008; Shen et al., 2016; He et al., 2016) is another common approach for estimating near-surface Q model, but it is often happened to be an issue for fully

---

utilizing reflection seismic data due to available offset information of shallow depth is limited. Since then, a first arrival Q tomography would be feasible, in considering of availability and stability of the first arrival wave in seismic record, especially for far offset data.

To apply the first arrival Q tomography, one major issue needed to be solved is how to accurately extract path attenuation factor  $t^*$  from first arrival field data, which is related to the amount of attenuation along the ray path. Q update amount could be obtained by minimizing the discrepancy between observed  $t^*$  of field data and synthetic  $t^*$  of model. A common approach of  $t^*$  extraction is the spectral ratio method (Brzostowki and McMechan, 1992; Calvaca and Fletcher, 2013). However, in a practical application, different waveforms (e.g., direct wave, turning wave or refraction) normally blend in the first arrival wave train, hence the accuracy of  $t^*$  estimation will be impacted leading to bias in retrieved Q distribution. A countermeasure for eliminating such influence is well-worth consideration in order to guarantee the accuracy and stability of the first arrival attenuation tomography. Additionally, its computation efficiency should be taken into account for a large seismic dataset as aforementioned.

### **1.3 Dissertation themes**

Based on two issues described above, there are two main research themes in this dissertation as briefly described in the following sections.

---

### **1.3.1 Mapping velocity model: efficient first-arrival traveltime tomography using large datasets**

The main objective of the first topic is to develop an efficient FAT tomographic inversion scheme solving for near-surface velocity model with large dataset as input, while preserving robustness and accuracy of retrieved velocity model.

To accomplish it, a grid-based eikonal solver combined with the adjoint-state formulation was adopted at first to improve the computing efficiency. It allows to implicitly evaluation of the gradient of misfit function, by which the computation cost is only proportional to the model size. Additionally, this eikonal solver utilized for conducting forward modeling by Fast Sweep Method (FSM) could also potentially circumvent the non-linearity of conventional ray-tracing based method in complex media.

Then the whole framework was developed underlying compressive sensing (CS) theory with a series of sparse-promoting based on online dictionary learning. With this sparse promoting scheme, the original L-2 norm misfit function of tomographic inversion problem will be reformed into a classical LASSO problem with a compressive sparsity constraint. A modified Gauss-Newton method is implemented to solve this hybrid norm problem. The sparsity level is in-situ adapting with residual data space and sparse domain for each iteration. It could ensure the sparsest representation of model update while dynamically adapting to variations of misfit between data and model space, hence guarantee the inversion convergency. One significant benefit of such approach is that the problem dimensionality could be substantially reduced by random subsampling a small portion of input data as guided by

---

Stochastic process. In this way, the memory occupation could be further reduced and eventually the overall computational efficiency has been remarkably promoted. This efficient FAT tomography is also feasible to implement with parallel computing.

### **1.3.2 Mapping Q distribution model: first arrival attenuation tomography based on An Adjoint-state Method**

The main objective of the second topic is to develop a first arrival Q tomographic inversion to map the near-surface Q distribution model for compensating attenuation effects and improving the quality of the seismic image. To accomplish this objective, one crucial issue needed to be handled with caution is to guarantee the accuracy of  $t^*$  estimation. Under the assumption of invariant ray path in weakly dissipative subsurface medium, the proposed method was organized as two cascading applications including velocity and Q model tomographic inversion. FAT tomography will be performed at first to provide velocity model and synthetic FAT as guidance for estimating  $t^*$ . In the second stage, the  $t^*$  will be estimated by a logarithmic spectral ratio linear regression method.

In the second stage, a time window with the proper length centered around synthetic FAT will be used for including complete first arrival waveforms related to  $t^*$ . In practice, such implementation is beneficial for obtaining stable synthetic FAT without introducing manual picking errors. This is also one aspect of proposed first arrival attenuation tomography superior to those of using seismic reflection data or other types of data. Given two categories of frequency-dependent propagation response from different wave types (direct, turning and

---

refraction wave), the  $t^*$  will be estimated twice through a logarithmic spectral ratio linear regression, respectively. Correspondingly, two synthetic attenuated seismic traces will be generated, and will be evaluated with the reference attenuated signal from field data by similarity coefficients. Through such process, an optimal  $t^*$  could be selected as input for following attenuation tomographic inversion.

As for attenuation tomographic inversion, A reformulated governing equation related to  $t^*$  is derived for implementation. The adjoint-state technique and grid-based eikonal equation solver using FSW is also adopted to address the computation efficiency issue of large-scale problem.

#### **1.4 Outline of dissertation**

My dissertation consists of five chapters:

- Chapter 1: Introduction

Introduction of research background on FAT tomography, current challenges and dissertation objectives.

- Chapter 2:

A comparative study of grid-based eikonal equation FAT tomography using adjoint state technique and conventional ray-tracing based FAT tomography is conducted in this chapter. Their imaging results and computation efficiency are compared and discussed in both theoretical formulations and numerical test. The proposed CS framework in Chapter 3 and Q tomographic inversion scheme in Chapter 4 will take this implementation as one of the essential

---

components.

- Chapter 3:

This chapter presents a highly efficient FAT tomographic inversion framework based on a sparse-promoting CS approach. At first, the original FAT tomographic inversion problem will be revisited, and the original misfit function is reformulated to evaluate the discrepancy between residual data space and model perturbation by a linearized eikonal equation. Then the sparse representation of a dictionary learning method will be described and modified into an online adaptive learning approach for FAT tomography. Following that, a sparse-promoting framework will be formulated with a combination of two previous topics. Three typical steps involve with CS approach, including data randomization, subsampling and sparsity promotion will be investigated respectively in the context of FAT tomography application. Eventually, a complete framework based on CS is formulated. A realistic 2D synthetic model and real field data are tested in this chapter to reveal the potential of the proposed method.

- Chapter 4:

A first arrival attenuation tomography based on adjoint-state method is proposed in this chapters. The methodology contains two cascading applications including velocity and Q model tomographic inversion. An adaptive correction method based on frequency-dependent propagation responses of different wave types is proposed to ensure the accuracy of  $t^*$  estimation. A new governing equation of the adjoint-state method allows to calculate the gradient of the misfit function for attenuation tomographic inversion will be introduced. The related numerical examples will demonstrate the robustness of this algorithm by comparison

---

with ray-tracing based algorithm. The theoretical basis in detail will be discussed and its performances on both synthetic and field data reveal the feasibility and potential of this method.

- Chapter 5

This is a conclusive chapter to summarize up previous chapters and provide potential directions of future work.

---

## CHAPTER 2 FIRSTARRIVAL TRAVELTIME TOMOGRAPHY BASED ON ADJOINT-STATE METHOD

### 2.1 Introduction

The first arrival traveltime (FAT) tomography is a well-established methodology can be used for investigating interior of subsurface ranged from near-surface scale to global scale (Aki and Lee 1976; Zelt and Smith 1992), using either active or passive sources. In the field of oil and gas seismic exploration, it is necessary to establish a more accurate near surface model to eliminate the influence on reflection signal, resulted from surface topography reliefs or near surface velocity anomalies. Especially for complex exploration areas such as mountains or deserts with varied near surface weathering zones, the quality of near surface model directly determines the final imaging quality. Due to the robustness of FAT and the high efficiency of traveltime tomography, near surface modeling by FAT tomographic inversion is one of mostly used method. The resulted near surface model provides macro-features of near surface seismic velocity, and it is usually used in tomography static correction (Zhu et al., 1992; Zhou et al., 2009), depth domain migration (Brooke, 2000), full-waveform inversion and other processing. The classical traveltime tomography inversion is based on raytracing-based algorithm, and most of them belong to kinematic approaches only requiring for analysis of the process of the motion in terms of time and space only (Červený, 2001). Normally, it is necessary to explicitly calculate the raypath and related traveltimes from source to receiver positions. Currently, two

---

categories of kinematic raytracing approaches are proposed: two-point method and wavefront tracking method. The two-point raytracing is a traditional way of tracing the corresponding ray path method, and its mainly includes shooting and bending methods. Raytracing by shooting method is an initial value problem of pursuing for an approximation solution of ray differential equation through an iterative process. Normally, a shooting method strategy will fix an initial incident point (source point) at first, and then an optimal take-off angle will be obtained through scanning until the desired end point is emerged through raypath. However, the strategy of choosing new take-off angle will sometimes become difficult due to the divergence of ill-positioned receiver locations (ending point). Such problem is happened more frequently in complex 3D case (Bishop et al., 1985). As for bending methods, it is a boundary problem of perturbing an initial guess path until it is satisfying the ray equations or Fermat's principle (Bulant, 1996). Though a raypath connecting between source and receiver location will always be obtained by bending method, the multipath propagation might be overlooked in some complicated structure due to the improper initial raypath guess. Both classical raytracing method are designed either for calculating the minimum traveltime or shortest raypath between source and receiver points, and some complex subsurface medium might lead to multivalued solution. Such convergence problem could result in low computation efficiency and local minimum trapping solution (Zhang et al., 2011).

Subsequently, the wavefront tracking method were proposed to overcome these issues happened in traditional raytracing method. It could be mainly categorized into two approaches: the shortest path method (SPM) (Klimes & Kvasnicka, 1994) and eikonal equation solution

---

method. The SPM algorithm was originated in network theory (Dijkstra 1959) in which the “starting” grids are selected with a shortest distance from all neighbors continuously until connecting source and receiver locations. Unlike the traditional raytracing method, one superior advantage of SPM method is that it can always obtaining a raypath meeting the requirement of minimum traveltime principle (Moser, 1991). However, in spite of some efforts made to improve the computation efficiency of SPM, such as schemes proposed by Klimes and Kvasnicka (1994) and Zhao et al. (2004), the potential higher computational cost is still an issue, comparing to the eikonal equation solution method.

The eikonal equation solution method is solved by finite-difference scheme while wave front expansion is evolving (Vidale, 199). The wavefront could be advanced by Fast marching method (FMM) using narrow band technique and sorting the solution path at each iteration (Sethian et al. 1999, 2001). Kim and Seongjai (2002) proposed an optimal variant of FMM by introducing group marching method (GMM) to update a group of wavefront points at once, instead of single point updates from original FMM. Hence, the computation efficiency is dramatically improved while maintaining the condition of wavefront expanding causality. Besides, a fast-sweeping method (FSM) originally proposed by Zhao (2005 and 2007) used WENO scheme for calculating wavefront based on eikonal equation. It evolves wavefront iteratively through sweeping and updates traveltime at each grid point simultaneously to guarantee the causality. Unlike FMM, the FSM does not require the update of wavefront conducted in a sequential order in a monotonic manner to maintain the causality. Since then, the implementation of FSM is more simplified and higher computational efficiency superior to

---

FMM even with appearance of rapid velocity changing in subsurface medium (Bak et al., 2010). Additionally, the causality of the partial differential equation in FSM is highly feasible to implement with efficient parallel computation in large number of FAT picks scenario. Since then, we will choose FSM as algorithm applied in our forward modeling and inversion strategy for our eikonal equation solver-based FAT tomography.

As for inversion algorithms, back projection (Humphreys and Clayton, 1988; Hole, 1992), LSQR (Zelt and Barton, 1998), SIRT and other methods are commonly used. Raytracing-based tomography has already been widely used in practical application. At present, the main interest of research on traveltimes tomographic inversion is to combine it with other types of higher accuracy inversion method with additional information. Raytracing method provides a natural approach for traveltimes calculation but can suffer from nonuniform ray sampling in presence of complex subsurface geology structure or long-offset acquisitions. Besides, most implementation of conventional traveltimes tomography often require to explicitly estimation of Fréchet matrix, which is high computation cost for large-scale dataset.

In addition to the above classical raytracing-based methods, there is another tomographic inversion method utilizing grid-based eikonal equation solver to circumvent the nonlinearity of raytracing approaches in complex media. Instead of calculating traveltimes along raypath, this type of method only needs to compute the first arrival time field. Moreover, with help of adjoint-state technique, it could obtain the model update for each iteration during inversion, without explicitly calculating Fréchet matrix. This method was developed based on control-

---

theory frame work (Lions, 1971) and was first adopted by Sei et al. (1994) with eikonal equation solver; and then Leung and Qian (2006) applied it in traveltime tomography problem. In recent years, the adjoint state technique is well known in field of geophysics study, due to large amount researches on waveform inversion (Brenders and Pratt, 2007; Brossier et al., 2009), and then extended into series of traveltime inversion related works (Leung and Qian, 2006). Taillandier et al. (2009) and Xie et al. (2014) carried out related research works based on fast sweeping method (FSM); Huang et al. (2012) proposed a joint transmission and reflection traveltime tomography using FSM and adjoint-state technique to improve both computation efficiency and image quality; Waheed et al. (2016) extended the grid-based isotropic eikonal solver into anisotropic case to conduct FAT tomographic inversion in anisotropic media.

In this chapter, we will review the theory background of both raytracing-based FAT tomography and a nonlinear FAT tomography based on adjoint-state method. A comparative study between these two types of FAT tomography will be conducted. In this study, the inversion accuracy, computation efficiency and memory occupations of these two methods will be quantitatively compared and analyzed through a 2D synthetic data test.

## **2.2 Theory**

### **2.2.1 Misfit function and its gradient**

If the regularization term is not considered, tomographic inversion is pursuing the best

---

fitting model between synthetic traveltine  $\mathbf{t}_s^{cal}$  and observed FAT  $\mathbf{t}_s^{obs}$  picked from observed data, and its L2 (least squares) norm of misfit function

$$O(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{ns} (\mathbf{t}_s^{cal}(\mathbf{m}) - \mathbf{t}_s^{obs})^T (\mathbf{t}_s^{cal}(\mathbf{m}) - \mathbf{t}_s^{obs}), \quad (2.1)$$

where  $\mathbf{m}$  is model parameter vectors of size  $nm \times 1$ , corresponding to different model parameterizations based on its physical meaning. As for FAT tomographic inversion, it will be parameterized as slowness, where  $nm$  is the number of medium parameters of model after grid discretization. It involves the squared error summed over  $ns$  total source number. The synthetic traveltine  $\mathbf{t}_s^{cal}(\mathbf{m})$  is a vector of size  $nr_s \times 1$  corresponding to all receivers, where  $nr_s$  is the total receiver number corresponding to the source  $s$ . Expanding the misfit function (2.1) about  $\mathbf{m}_0$  in Taylor expansion, and keep the term only up to second order

$$O(\mathbf{m}) \approx O(\mathbf{m}_0) + \left( \frac{\partial O(\mathbf{m})}{\partial \mathbf{m}} \Big|_{\mathbf{m}=\mathbf{m}_0} \right)^T \delta \mathbf{m} + \frac{1}{2} \delta \mathbf{m}^T \frac{\partial^2 O(\mathbf{m})}{\partial \mathbf{m} \partial \mathbf{m}^T} \Big|_{\mathbf{m}=\mathbf{m}_0} \delta \mathbf{m}. \quad (2.2)$$

The gradient of misfit function  $\mathbf{g}$  is the first derivative of the misfit relative to the medium parameter, written as

$$\mathbf{g} = \frac{\partial O(\mathbf{m})}{\partial \mathbf{m}}, \quad (2.3)$$

and its second derivative is  $\mathbf{H}$ , called Hessian matrix, and defined as

$$\mathbf{H} = \frac{\partial^2 O(\mathbf{m})}{\partial \mathbf{m} \partial \mathbf{m}^T}. \quad (2.4)$$

To obtain the minimum data residual, it is necessary to find out the stationary points of misfit function, where  $\mathbf{g} = 0$ . By deriving equation (2.2) and setting it to 0, we have

---


$$\mathbf{H}|_{m_0} \delta \mathbf{m} = -\mathbf{g}|_{m_0}, \quad (2.5)$$

$$\delta \mathbf{m} = -(\mathbf{H}|_{m_0})^{-1} \mathbf{g}|_{m_0}. \quad (2.6)$$

Equation (2.6) is an explicit expression of the model update at current inversion iteration. The nonlinear inversion can be conducted by iteratively updating model parameters in equation (2.6). Eventually, the final inverted model will be obtained when the data residual less than a certain preset small threshold. Since it is difficult to compute the inversed Hessian matrix directly, the linear equation (2.5) is usually solved by iterative inversion, and such type of method is called Gauss Newton method. For large-scale dataset problem, obtaining inversed Hessian matrix is very time consuming and infeasible to manage memory storage (e.g., the dimension size of  $\mathbf{H}^{-1}$  is  $nm \times nm$ ), hence an approximation approach is usually considered. For a large data acquisition aperture,  $\mathbf{H}$  sometimes could be approximated to be close to a unitary matrix, then  $\mathbf{H}^{-1}$  can be reduced to a weighing vectors scaling the steepest-descent directions (Pratt et al., 1998; Jang et al., 2009), and solved by steepest-descent method. If  $\mathbf{H}$  is approximated to a diagonal or band-limited matrix (Plessix, and Mulder, 2004; Pan et al., 2015), then it will be altered to a preconditioned steepest gradient method.

### 2.2.2 FAT tomographic inversion based on raytracing method

For FAT tomographic inversion based on raytracing, the raypath connecting each pair of source and receiver should be calculated. Then the first break time can be expressed as

$$t_{s,r} = \sum_{i=1}^{nm} m_i l_{s,r}^i, \quad (2.7)$$

---

where  $i$  is number of the grid;  $m_i$  is the slowness in the  $i$ th grid;  $l_{s,r}^i$  is the arc length of raypath in  $i$ th grid. It can be written in matrix form as

$$\mathbf{t}_s^{cal}(\mathbf{m}) = \mathbf{L}_s(\mathbf{m})\mathbf{m}. \quad (2.8)$$

Here  $\mathbf{L}_s$  is a forward operator, and it is containing the length of raypath corresponding to different receivers in the grid. Together,  $\mathbf{L}_s(\mathbf{m})$  is a function related with both medium parameters and raypath, that is to say, the synthetic traveltime  $\mathbf{t}_s^{cal}$  has a nonlinear relationship with slowness in our case.

Assuming that the model update in each iteration is not large enough to result in a dramatic change of raypath, meaning that  $\frac{\partial \mathbf{L}_s(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{0}$ . Based on such assumption, Equation (2.8) can be substituted into Equation (2.1), then combine with Equation (2.3), the misfit function gradient can be expressed as

$$\mathbf{g} = \sum_{s=1}^{ns} \mathbf{L}_s^T (\mathbf{L}_s \mathbf{m} - \mathbf{t}_s^{obs}), \quad (2.9)$$

From Equation (2.4), we will have Hessian matrix  $\mathbf{H}$  as

$$\mathbf{H} = \sum_{s=1}^{ns} \mathbf{L}_s^T \mathbf{L}_s. \quad (2.10)$$

If Hessian matrix is approximated by diagonal elements and substituted into Equation (2.5) and (2.6), then the basic formulation of back projection algorithm can be obtained

$$\delta m_i = \frac{\sum_s \sum_r \delta t_{s,r} l_{s,r}^i}{\sum_s \sum_r (l_{s,r}^i)^2} \quad (2.11)$$

---

Substituting equation (2.9) and (2.10) into equation (2.5), we have

$$\sum_s \mathbf{L}_s \delta \mathbf{m} = \sum_s \delta \mathbf{t}_s. \quad (2.12)$$

In above formulation,  $\delta \mathbf{t}_s$  is the traveltime residual corresponding to the  $s$ -th source, denoted by  $\delta \mathbf{t}_s = \mathbf{L}_s \mathbf{m}_0 - \mathbf{t}_s^{obs}$ . Equation (2.12) is linear, which is also equivalent to equation (2.5) and (2.6). A raytracing-based FAT tomography normally solves Equation (2.12) within each nonlinear iteration during inversion, to realize a similar effect as using Gauss-Newton (GN) method. According to different nonlinear approaches of solving for Equation (2.12), a series of algorithms have been developed, such as least square QR-factorization (LSQR) method (Zelt and Barton, 1998) and simultaneous iterative reconstruction technique (SIRT) method.

### 2.2.3 FAT tomographic inversion based on eikonal equation solver

Different from raytracing-based FAT tomography, the FAT tomography based on eikonal equation solver directly solve for the eikonal equation instead of ray tracing as forward modeling method during inversion. The propagation of the first arrival (or wavefront) and its traveltime in isotropic media satisfies the eikonal equation, which can be discretized by finite difference method using upwind scheme

$$(\mathbf{D}_x \mathbf{t}_s) \cdot (\mathbf{D}_x \mathbf{t}_s) + (\mathbf{D}_y \mathbf{t}_s) \cdot (\mathbf{D}_y \mathbf{t}_s) + (\mathbf{D}_z \mathbf{t}_s) \cdot (\mathbf{D}_z \mathbf{t}_s) = \text{diag}(\mathbf{m}) \mathbf{m}, \quad (2.13)$$

and its initial condition is

---


$$\mathbf{t}_s(i_s) = 0. \quad (2.14)$$

Here,  $\mathbf{t}_s$  represents FAT at all grids corresponding to each source  $s$ , which is a vector in size of  $nm \times 1$ . Same as discussed in previous section 2.2.1,  $nm$  is the total number of model parameters. The source location at certain grid point is  $i_s$ . At source position  $i_s$ , traveltime  $\mathbf{t}_s$  is equal to zero.  $\mathbf{D}_x$ ,  $\mathbf{D}_y$  and  $\mathbf{D}_z$  is upwind finite difference matrix along  $x$ ,  $y$  and  $z$  directions, respectively.  $\mathbf{m}$  is slowness vector in size of  $nm \times 1$ , and  $\text{diag}(\mathbf{m})$  is diagonal matrix that all slowness elements located along diagonal direction. Our forward modeling is conducted by fast sweeping method (FSM), and the related algorithm and practical implementation details referred to literatures from Zhao (2007).

Assuming operator can extract a synthetic first arrival traveltime  $\mathbf{t}_s^{cal}$  from  $\mathbf{t}_s$ , such as

$$\mathbf{t}_s^{cal} = \mathbf{P}\mathbf{t}_s. \quad (2.15)$$

In our case, term  $\mathbf{P}\mathbf{t}_s$  represents the FAT for each receiver at the acquisition surface. Substituting equation (1.15) into (1.1), and utilize equation (1.3) to obtain misfit function gradient

$$\mathbf{g} = \frac{\partial O(\mathbf{m})}{\partial \mathbf{m}} = \sum_{s=1}^{ns} \left( \frac{\partial \mathbf{t}_s^{cal}}{\partial \mathbf{m}} \right)^T (\mathbf{t}_s^{cal}(m) - \mathbf{t}_s^{obs}) \quad (2.16)$$

The Fréchet matrix  $\mathbf{F} = \frac{\partial \mathbf{t}_s^{cal}}{\partial \mathbf{m}}$  can be obtained by differentiating  $\mathbf{m}$  on both side of equation (2.13), expressed as

---


$$\frac{\partial \mathbf{t}_s^{cal}}{\partial \mathbf{m}} = \mathbf{P} \begin{pmatrix} \text{diag}(\mathbf{D}_x \mathbf{t}_s) \mathbf{D}_x + \\ \text{diag}(\mathbf{D}_y \mathbf{t}_s) \mathbf{D}_y + \\ \text{diag}(\mathbf{D}_z \mathbf{t}_s) \mathbf{D}_z \end{pmatrix}^{-1} \text{diag}(\mathbf{m}) . \quad (2.17)$$

The size of  $\mathbf{F}$  is the number of observed traveltimes multiplied by the number of model parameters. Substituting Equation (2.17) into (2.16), we have

$$\begin{aligned} \mathbf{g} &= \mathbf{F}^T (\mathbf{t}_s^{cal}(\mathbf{m}) - \mathbf{t}_s^{obs}) \\ &= \sum_{s=1}^{ns} \text{diag}(\mathbf{m}) \begin{pmatrix} \text{diag}(\mathbf{D}_x \mathbf{t}_s) \mathbf{D}_x + \\ \text{diag}(\mathbf{D}_y \mathbf{t}_s) \mathbf{D}_y + \\ \text{diag}(\mathbf{D}_z \mathbf{t}_s) \mathbf{D}_z \end{pmatrix}^{-1} \mathbf{P}^T (\mathbf{t}_s^{cal}(\mathbf{m}) - \mathbf{t}_s^{obs}) . \end{aligned} \quad (2.18)$$

A direct solving for the invert term in Equation (2.18) is almost computational prohibitive.

Then an iterative gradient for single source  $\mathbf{g}_s$  could be obtained by

$$\begin{pmatrix} \text{diag}(\mathbf{D}_x \mathbf{t}_s) \mathbf{D}_x^T + \\ \text{diag}(\mathbf{D}_y \mathbf{t}_s) \mathbf{D}_y^T + \\ \text{diag}(\mathbf{D}_z \mathbf{t}_s) \mathbf{D}_z^T \end{pmatrix} (\text{diag}(\mathbf{m}))^{-1} \mathbf{g}_s = \mathbf{P}^T (\mathbf{t}_s^{cal}(\mathbf{m}) - \mathbf{t}_s^{obs}) . \quad (2.19)$$

Equation (2.19) is equivalent to the Equation (2.12) from Taillandier et al. (2009), and this adjoint state system could be solved by the same strategy as used in forward modeling of eikonal equation solver. The adjoint state technique allows to directly compute misfit function gradient  $\mathbf{g}$  without introducing explicit computation of  $\mathbf{F}$ . Detailed mathematical developments and practical implementations can be found in Sei and Symes (1994), and Leung and Qian (2006). From Equation (2.4), the associated Hessian matrix can be expressed as

$$\mathbf{H} = \mathbf{F}^T \mathbf{F} = \left( \frac{\partial \mathbf{t}_s^{cal}}{\partial \mathbf{m}} \right)^T \left( \frac{\partial \mathbf{t}_s^{cal}}{\partial \mathbf{m}} \right) \quad (2.20)$$

Due to the existence of a large inversed matrix in equation (2.18), a direct method solving

---

for it is infeasible. Instead, we could obtain an equivalent form by substituting Equation (2.20) into (2.5)

$$\sum_{s=1}^{ns} \left. \frac{\partial \mathbf{t}_s^{cal}}{\partial \mathbf{m}} \right|_{\mathbf{m}_0} \delta \mathbf{m} - (\mathbf{t}_s^{cal}(\mathbf{m}) - \mathbf{t}_s^{obs}) = 0 \quad (2.21)$$

Similarly, a nonlinear iterative solving for above equation leads to type of GN method. If keep the diagonal elements dominated in Hessian matrix, it turns to be a Steepest descent method with pre-conditioning.

### 2.3 Comparative study

Based on the same misfit function, the basic formulas of tomographic iterative inversion of FAT based on ray tracing equation and eikonal equation are derived under the unified inversion framework. Comparing the theoretical derivations of the two methods in detail, they have the following similarities in theory:

- 1) It has the same theoretical basis. The ray theory based on high frequency approximation is used to calculate the theoretical first break travel time; the unified misfit function is used for the FAT travelttime inversion, and the basic formula of iterative inversion can be derived under the unified inversion framework.
- 2) The specific inversion algorithms are equivalent. The back-projection algorithm in raytracing equation method is equivalent to the preconditioned steepest descent method based on eikonal equation method; LSQR, SIRT and so on in ray tracing equation method are equivalent to Gauss Newton algorithm in path function equation method.

---

Due to the similarity in theory, both of them have similar resolution and ability to deal with complex problems, such as velocity anomalies and other complex exploration problems that cannot be solved by them.

In theory, the two methods have the following main differences:

- 1) They are based on different forward operators. The former is based on the ray tracing equation, and the travelttime is obtained by the integral of slowness along the ray path, while the latter is based on the eikonal equation solving for grid-based finite difference.
- 2) The two methods are different in calculating the gradient of misfit function and Hessian matrix. The raytracing-based one will explicitly solve them (equation 2.9 and 2.10), which is intuitive and has clear physical meaning; As for eikonal equation based one, the gradient corresponding to a single source can be obtained by solving for the PDE (equation 2.19) implicitly, and the related physical meaning is not clear.

### **2.3.1 2D synthetic model test**

In order to compare the difference between the two methods in inversion accuracy, calculation efficiency and memory occupation, the Amoco static correction benchmark test model 1994 is selected for numerical test, as shown in Figure 2-1. The model is originated from Amoco Tulsa Research Lab in 1994, which is firstly used by Mike O'Brien as one part of a project to investigate the static correction in land data. The model includes most of the common near surface geological structures, such as high-speed layer exposure, local high-speed, low-speed abnormal body, shallow low-speed layer, near-surface complex structure and extremely

---

shallow low-speed body, which can be used for comprehensive comparative analysis. Since the inversion depth of conventional near surface modeling is about 1km, the model shown in Figure 2-1 is the result of compression of the original model in the depth direction, which keeps the complexity of the original model and makes its depth consistent with the conventional near surface problem.

### **1. Comparison on inversion accuracy**

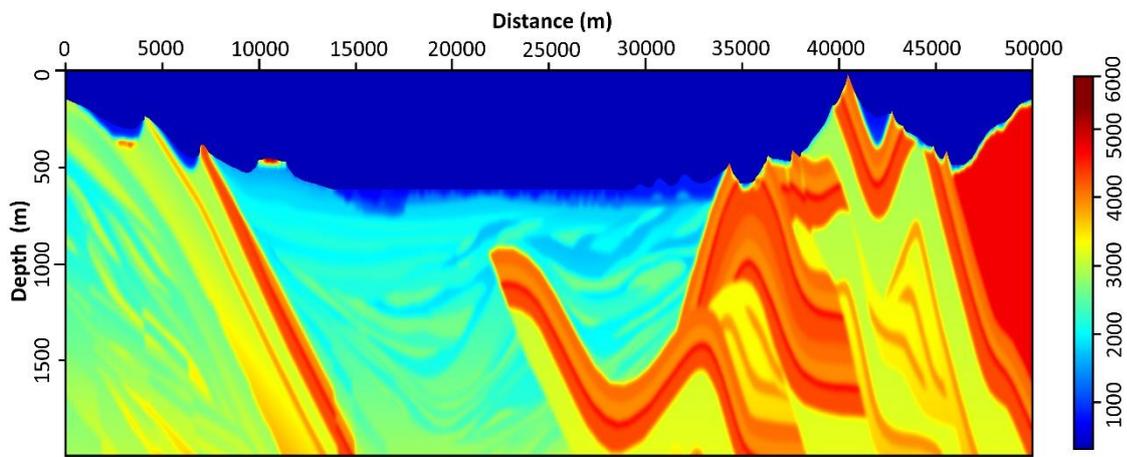
The accuracy of two types of tomographic inversions was compared by Amoco 94' synthetic model, in terms of image quality and model fitting. The forward operators from both of two inversion methods are realized by solving the eikonal equation numerically. For eikonal equation grid-based tomographic inversion method, the upwind finite difference algorithm based on FSW is adopted. For raytracing-based tomographic inversion, the raypath will be traced along the negative gradient direction of the first arrival timefield from the receiver position, then forward operator in Equation (2.8) will be constructed.

For the inversion scheme, the tomographic inversion based on eikonal equation adopts the preconditioned steepest gradient method. The tomography inversion method based on the raytracing equation adopts the back-projection algorithm corresponding to the steepest gradient method. Other than these differences, the two inversion methods adopted exactly the same processing and parameters settings are used in the numerical test.

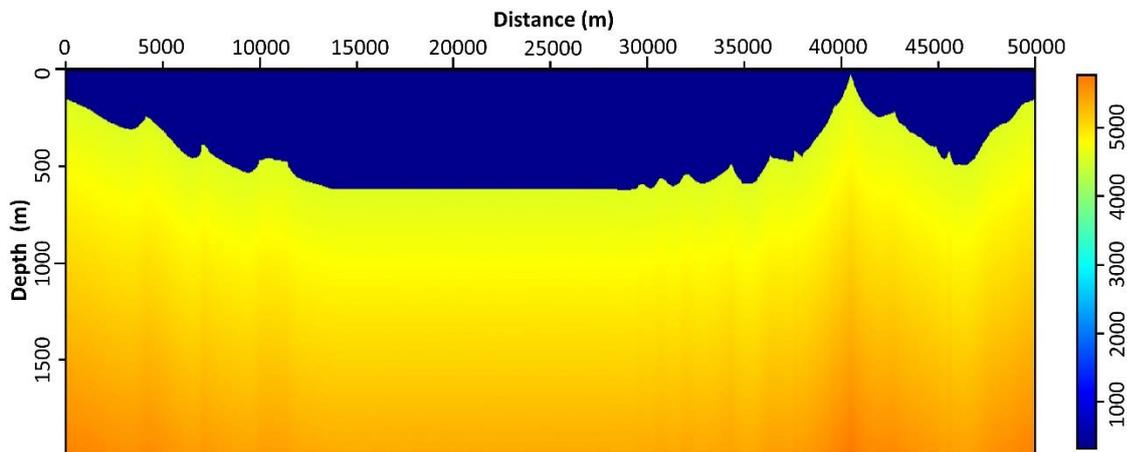
The synthetic observed FAT is simulated from this model with eikonal equation solver on 1998 sources. Each source is recorded up to 2500 receivers on either side to simulate a high-

---

density survey. The nearest offset is 10m and maximum offset is 7500m, where receiver interval is 20m. Numerical forward modeling is conducted to generate synthetic data for tomographic inversion. A linear increase model was shown in Figure 2-2, and it will be used for both two aforementioned methods as initial model. The velocity at the surface of the linear increasing model is 450 m/s, the gradient of velocity with depth is 0.5, and the velocity at the bottom of the model is about 5500 m/s.



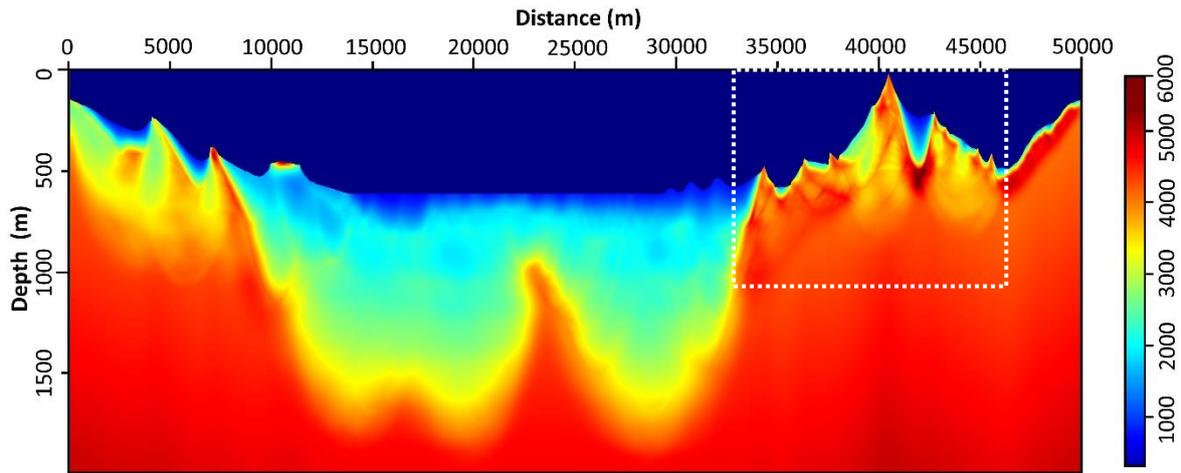
**Figure 2-1:** True velocity model.



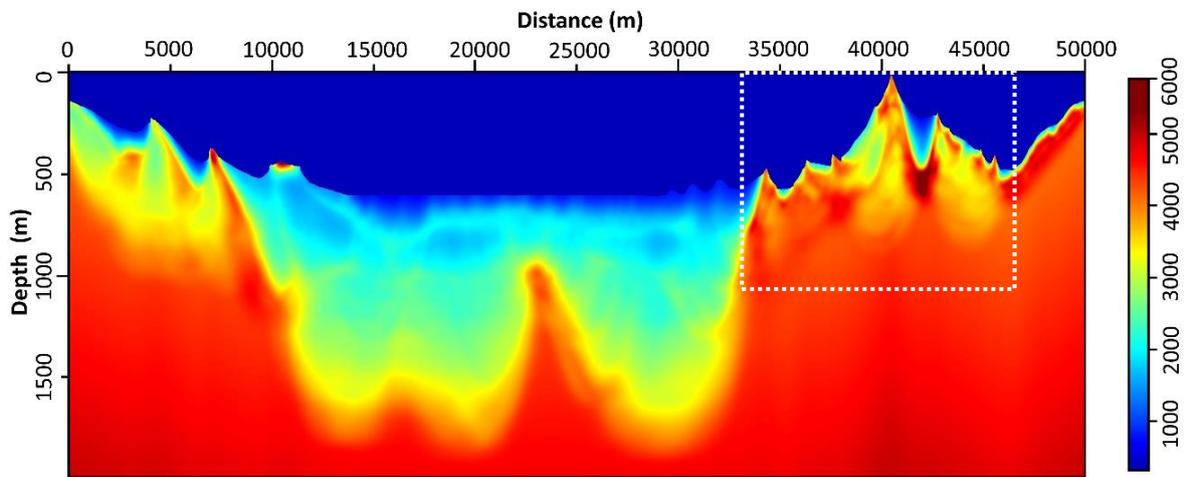
**Figure 2-2:** Linear smooth initial velocity model. The velocity is increasing from 450m/s at surface to 5500 m/s at bottom with gradient of 0.5.

---

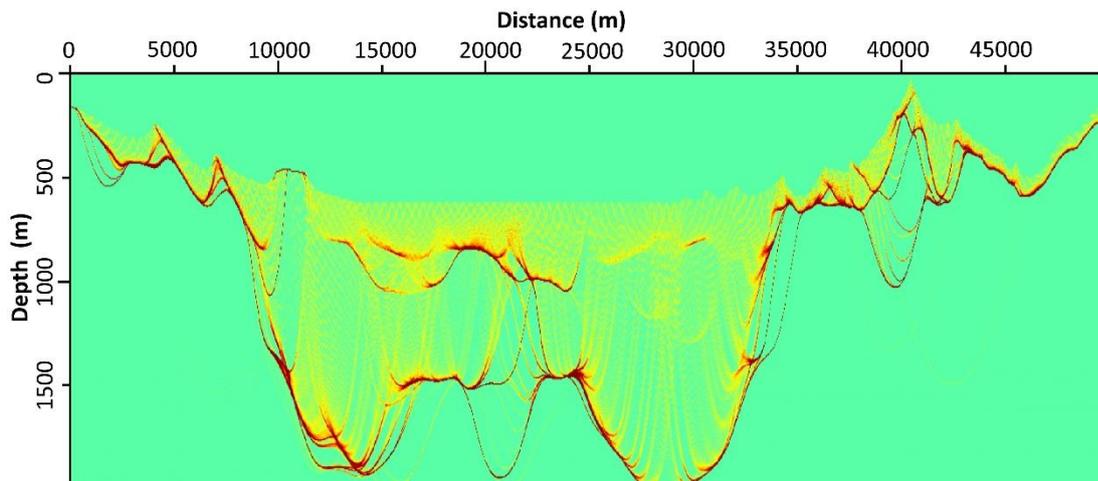
Figure 2-3 shows the tomographic inversion result generated by conventional raytracing-based tomography, and Figure 2-4 shows the results from method based on eikonal equation. For both methods, their reconstructed velocity models have good agreement with true model in near surface region (~500 to 700m), depending on location and corresponding ray coverage. The near surface geological structures such as high velocity layer outcrop, local high velocity and low velocity abnormal bodies, shallow low velocity layer and extremely shallow low velocity body have relatively good image quality. Due to the limitation of resolution of ray-based tomography, the complex structural area on the right side of the model has not achieved ideal results from both of two methods. In terms of results details, since the kernel function of tomographic inversion based on the eikonal equation is band limited similar to finite frequency characteristics, the inversion result is more stable, especially in the complex structural area, as labelled by the white dash line box in Figure 2-3 and 2-4. Unlike some significant smearing artifacts appeared in raytracing-based method result of Figure 2-3, the eikonal equation solver-based method tends to present smoother in Figure 2-4. It indicates that a high velocity anomaly existed in ray-based method result, while it is not happening in eikonal equation-based one leads to a more stable result. Since then, in terms of inversion accuracy, the two methods are very close, and the eikonal equation based one is slightly superior in some local details.



**Figure 2-3:** Inversion results from raytracing-based FAT tomographic inversion.



**Figure 2-4:** Inversion results from FAT tomographic inversion based on raytracing.



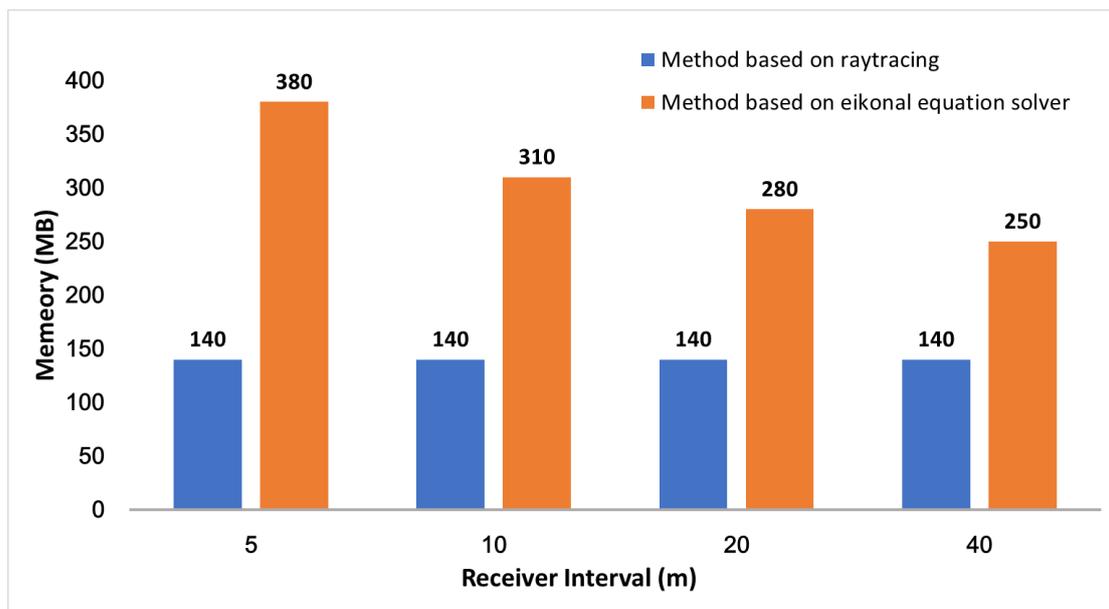
**Figure 2-5:** Ray density distribution at last iteration of tomographic inversion based on ray tracing method. It could bring helps on understanding the reliability of velocity inversion results in different

---

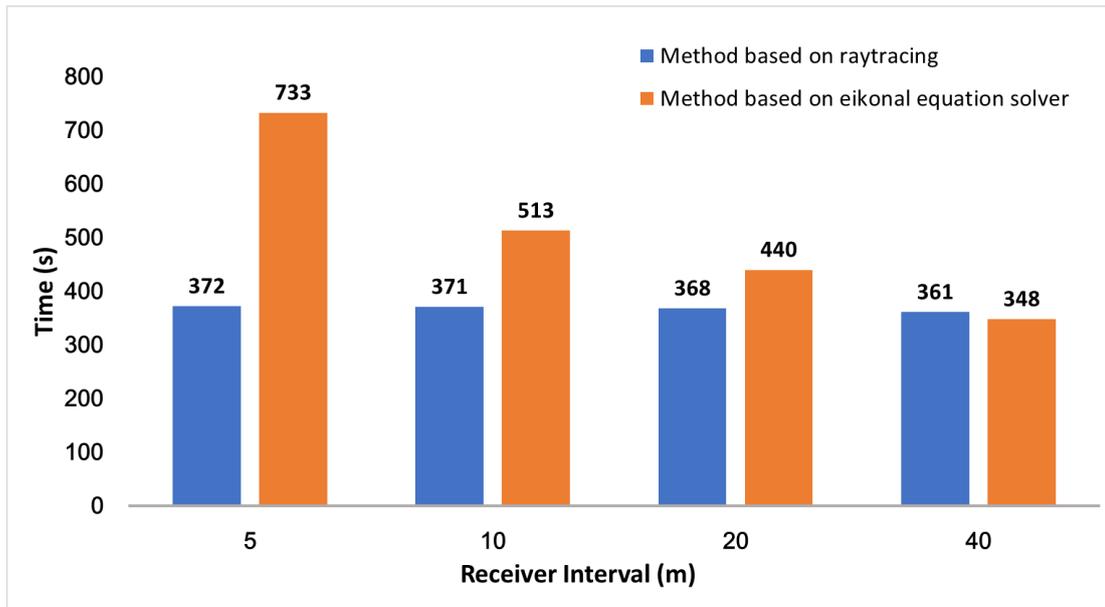
regions. However, in eikonal equation-based method, there is no such concept related to ray density, hence it is lack of effective tools for image quality control.

## 2. Comparison on computation efficiency and memory occupation

In order to make a more comprehensive comparison between the two methods in terms of memory occupation and computation efficiency, four different survey systems are designed for this comparison. As discussed in section 2.1, we know that the computation cost from eikonal equation-based method only relates to the size of model, while its proportional to the number of receivers by raytracing-based method. Based on this main difference, our test survey systems are set by different receiver interval with fixed total source number and maximum offset. The receiver intervals are set at 5m, 10m, 20m and 40m respectively.



(a)



(b)

**Figure 2-6:** Comparison of (a) memory occupation and (b) computation time of two methods: receiver intervals of 5m, 10m, 20m and 40m.

The test environment is Linux cluster with the CPU is Intel (R) Xeon (R) CPU e5-2670 (2.60GHz), MPI is used for parallel computation. The test conducted on 7 nodes in total, and each node uses 30 threads for 30 times inversion iteration of both two methods. The comparison of memory occupations for a single process and running time under different survey systems is shown in Figure 2-6a and Figure 2-6b, respectively. Since the memory occupation of raytracing-based method depends on the specific raypath of each source, the histogram only shows an average running time of several intermediate iterations. Noted that the memory occupation is an average value of intermediate iterations.

By comparing the items in Figure 2-6, it indicates that: 1) the memory occupation and computation time of eikonal equation-based method are independent of the number of receivers, but only related to the size of the current model; 2) the memory occupation and time cost of

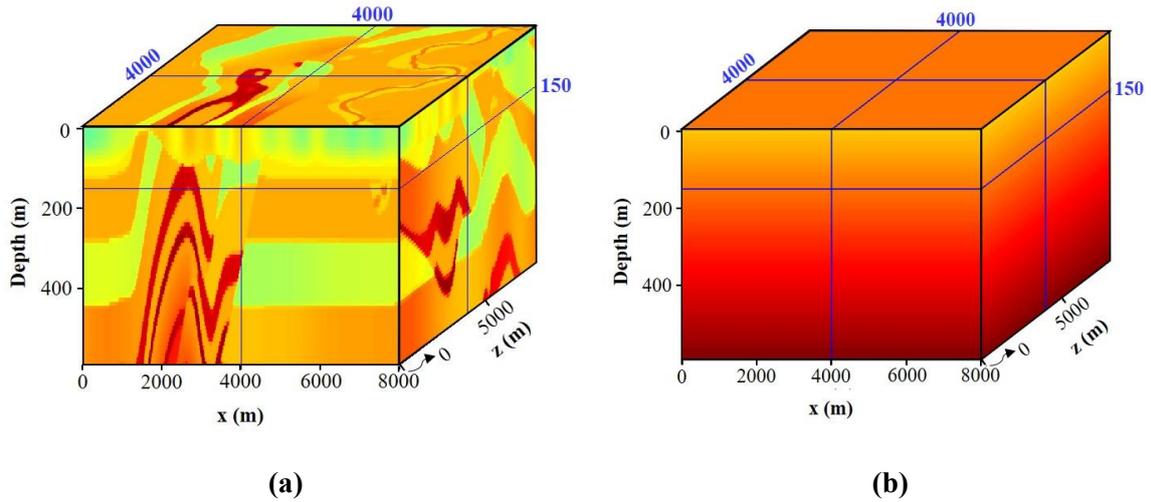
---

raytracing-based method are proportional to the number of receivers. With a greater number of receivers involved in, both of memory occupation and computation time will increase;

3) when the receiver array is dense as for some wide azimuth, high dense seismic data, the tomographic inversion method based on the functional equation has great advantages in memory consumption and computational efficiency. When the arrangement of receiver array is sparse (or large receiver interval), the tomography inversion method based on ray tracing equation has greater advantages.

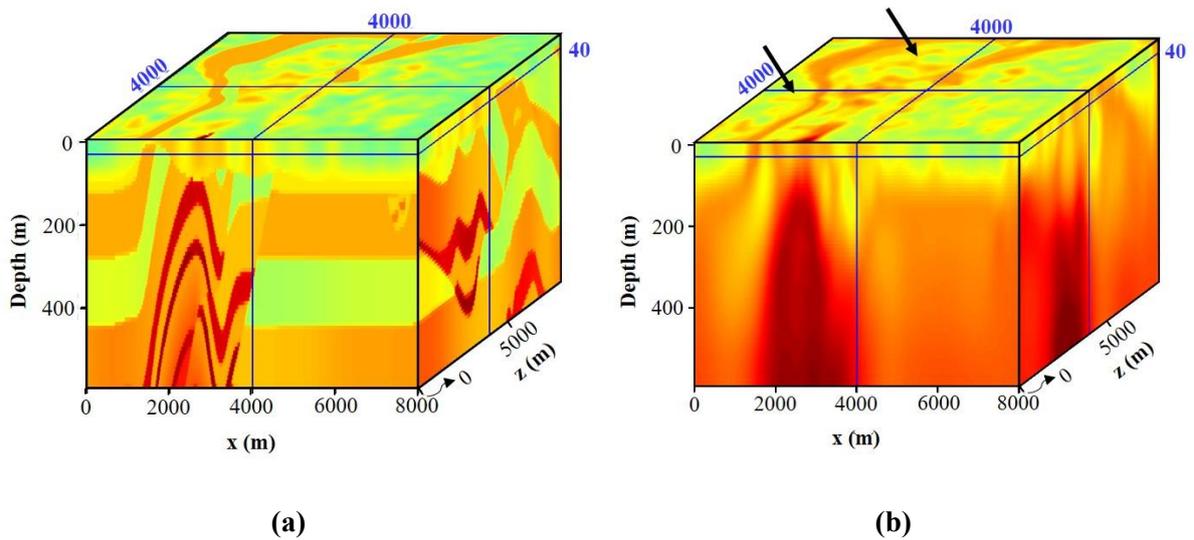
### **2.3.2 3D synthetic model test**

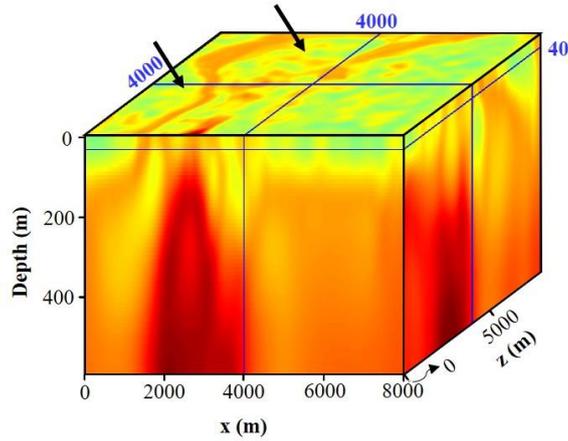
A 3D synthetic model test is conducted in this section, to further investigate the computation efficiency and image capability of FAT tomographic inversion based on adjoint state technique in 3D case. The model is a realistic overthrust 3D model as shown in Figure 2-7a, in which the near-surface region appears complex geological structures containing overthrust and channels. The velocity model has  $801 \times 801$  cells at horizontal surface, and vertical direction has 60 cells, where the cell interval is 10m. The corresponding acquisition system has 6400 synthetic shots in total, and the shot and receiver interval in both x and y direction is 90 and 10m, respectively. The maximum offset for a single shot is 2000m.



**Figure 2-7:** (a) Overthrust 3D velocity model and (b) Initial velocity model used for inversion.

The inverted models are first compared at depth of 40m as shown in Figure 2-8b and c. At this shallow depth, two method results are mostly identical, and their results all have good agreements with true velocity model from Figure 2-8a. In some specific areas, indicated by black arrows, the result of eikonal equation solver-based method have better image resolution than raytracing-based one.

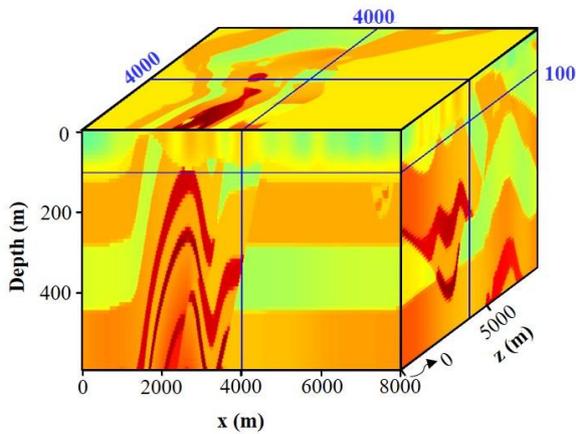




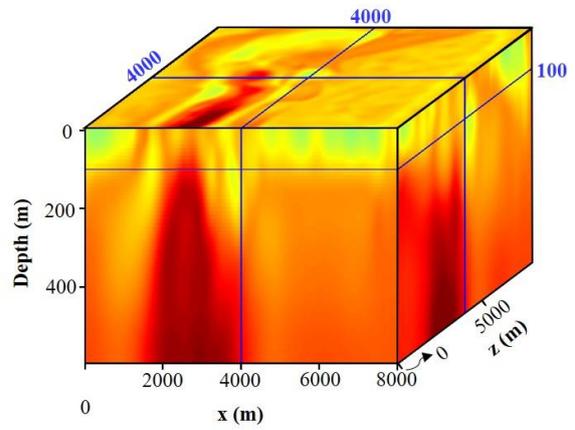
(c)

**Figure 2-8:** Estimated 3D velocity model at depth of 40m. (a) True velocity model, (b) Raytracing based method, (c) Eikonal equation solver-based method using adjoint-state technique.

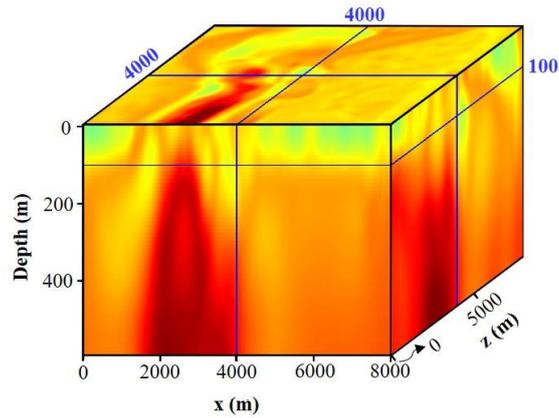
As depth increase to 100m as shown in Figure 2-9, both of two methods results in Figure 2-9b and c are identical and have good agreement with true model of Figure 2-9a. It indicates that these two methods all have capability of accommodating strong lateral variation.



(a)



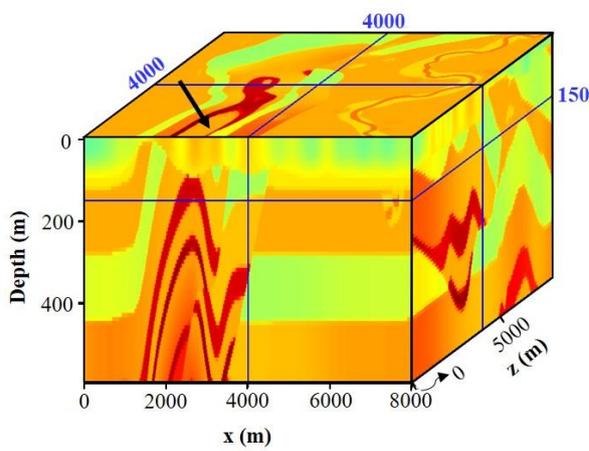
(b)



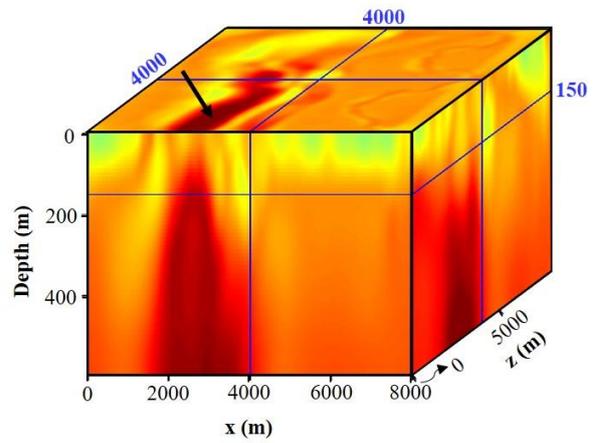
(c)

**Figure 2-9:** Estimated 3D velocity model at depth of 100m. (a) True velocity model, (b) Raytracing based method and (c) Eikonal equation solver-based method using adjoint-state technique.

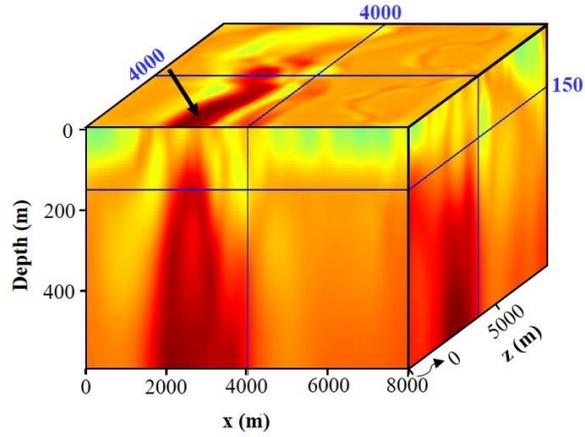
Eventually, at depth of 150m, where the reversed-velocity interfaces appear as indicated by black arrows in Figure 2-10b of true velocity, both of two methods in Figure 2-10b and c cannot reconstruct this structure, due to the low ray coverage.



(a)



(b)



(c)

**Figure 2-10:** Estimated 3D velocity model at depth of 150m. Reversed-velocity interfaces is indicated by black arrows. (a) True velocity model, (b) Raytracing based method and (c) Eikonal equation solver-based method using adjoint-state technique.

Overall, in near-surface of this 3D model, both of two methods could obtain satisfied result in both resolution and fidelity. However, for one iteration, the computation time cost by eikonal equation solver-based method is around 2 mins in average, which is only 1/2 of conventional raytracing-based method. The memory cost for a single thread is also around 1/2 of raytracing-based one at 1.2GB. Note that we only use 2000m as maximum offset for this test. If the offset used in inversion was further increased, the advantage of computation efficiency from the eikonal solver-based method would be more apparent.

## 2.4 Summary and conclusions

In this chapter, the performance of FAT tomography based on adjoint-state technique and conventional ray-tracing is compared in terms of their inversion accuracy and computational efficiency. The numerical test result shows that they could obtain almost identical results,

---

somehow the adjoint state method has better resolution in some geological complex area of near surface. The main reason for this is that gradient for one single source-receiver pair is slightly fatter than a ray, which is more sensitive for complicated velocity variation. The adjoint-state substantially has the same behavior as any classical raytracing-based tomography algorithm; hence their result will have no significant difference.

The major advantages of adjoint-state method tomography are high computational efficiency in terms of memory occupation and time cost. Its computation cost only related with the corresponding model size for each single source. Hence, FAT tomography based on adjoint-state method is very suitable for large-scale dataset scenario. Besides, it is feasible for parallel computing implementation, by which the computational efficiency could be further improved.

To further improve the computation efficiency, a stochastic process allowing to reduce problem dimensionality might be a good choice. Unfortunately, it somehow will induce in “noisy” artifacts caused by severe subsampling to degrade the image quality. Since then, we are looking forward to a gradient optimization method could suppress these artefacts. And this will be our main interest to investigate in next chapter.

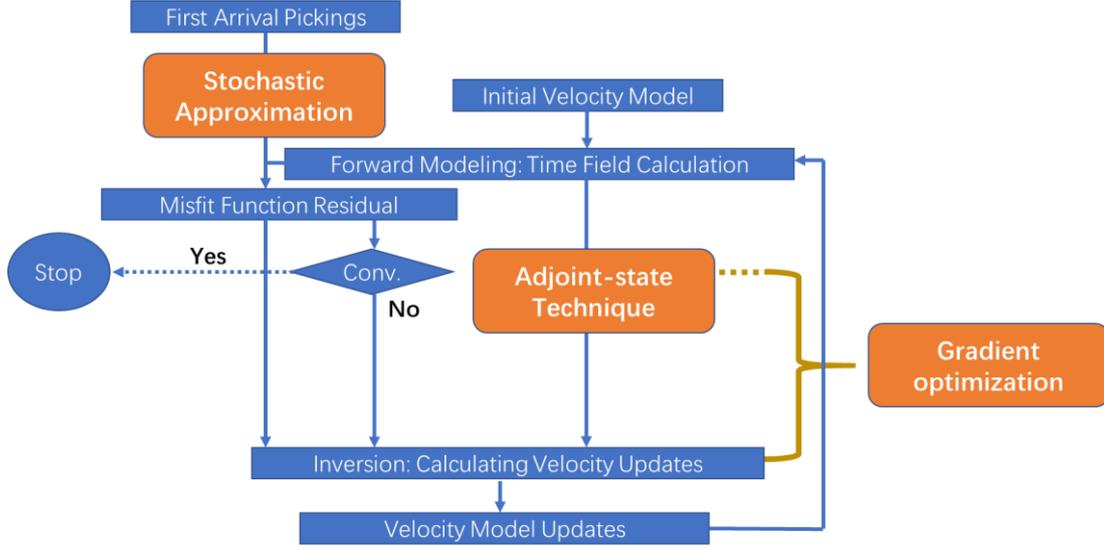
---

## CHAPTER 3 EFFICIENT FIRST ARRIVAL TRAVELTIME TOMOGRAPHIC INVERSION IN LARGE DATASET

### 3.1 Introduction

In this chapter, we establish an efficient FAT tomographic inversion scheme named SA+, while maintaining reasonable accuracy comparing to those from standard FAT tomographic inversion.

The basic workflow of SA+ method is illustrated in figure 3.1, offering a full picture as guidelines for the following discussions. It is well known that FAT tomographic inversion problem can be solved by a standard Gauss-Newton method of exploiting the convex-composite structure of Equation (2.1) using a linearized eikonal solver. Instead, we used a modified GN under compressive sensing framework combining stochastic process and gradient optimization to improve the computation efficiency while preserving the image quality. The stochastic process helps to reduce the problem dimensionality by using only a small subset of full dataset to conduct tomographic inversion. This gradient optimization allows to suppress the “noise” induced by random subsampling from stochastic process. It is realized by using a sparsity-promoting technique based on online orthonormal dictionary learning, which allows to exploit descent direction of misfit function in a sparse domain. Eventually, a modified GN method was proposed for a reduced misfit function regularized by the sparsity of model update.



**Figure 3- 1:** Schematic sparsity promoting FAT tomography workflow.

## 3.2 FAT tomographic inversion

### 3.2.1 Observation on misfit function

As discussed in chapter 2, a typical first-arrival traveltimes tomographic inversion is a data fitting procedure depending on the collection of series of first-arrival traveltimes picks and numerous computations to estimate the near-surface velocity model. The related nonlinear least-squares optimization problem is in the form

$$\text{minimize } J(\mathbf{m}) := \frac{1}{2} \sum_{s=1}^{N_s} |\mathbf{t}_s^{\text{obs}} - \mathcal{F}_{ek}^s(\mathbf{m})|^2. \quad (2.1)$$

Again, the  $N_s$  is the number of sources (batch size), and  $\mathbf{t}_s^{\text{obs}}$  composing as vectors represents the observed data of FAT picks corresponding to the  $s^{\text{th}}$  source; and  $\mathcal{F}_{ek}^s(\mathbf{m})$  is the forward operator for the  $s^{\text{th}}$  source. The unknown medium parameter (slowness for this case) is also organized as vectors and denoted by  $\mathbf{m}$ . The forward operator  $\mathcal{F}_{ek}(\bullet)$  acts linearly for

---

each source through the whole model space. And the receiver array geometry assumes for each source gather.

The observed data can be further organized as vectors:  $T^{obs} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{N_s}]$ , and the Equation (2.1) could be written as

$$J(\mathbf{m}) = \frac{1}{2} \| T^{obs} - \mathcal{F}_{ek}(\mathbf{m}) \|_F^2, \quad (2.2)$$

Where  $\| \bullet \|_F$  is the Frobenius norm. Here are several important observations on Equation (3.1), and it be utilized for our algorithm design guidance:

- First, the forward operator  $\mathcal{F}_{ek}(\bullet)$  is used for solving for a PDE with multiple right-hand-sides, and its work load is directly proportional to the data batch size  $N_s$ . Practically, the size of the data space and of the model space can be quite large (up to  $10^6$  FAT picks and  $10^5$  unknowns). Additionally, the batch size normally involves with millions of deployments of source and receiver pairs. Since then, it is necessary to consider dimensionality reduction technique implements at first place when we are working on this optimization problem.

- Second, it poses a convex-composite structure, though it is nonlinear and non-convex.

Therefore, we can write the objective function as

$$J(\mathbf{m}) = \rho(\mathcal{L}(\mathbf{m})), \quad (3.3)$$

where  $\rho(\bullet) = \frac{1}{2} \| \bullet \|_F^2$  is convex, and  $\mathcal{L}(\mathbf{m})$  is differentiable (smooth). Such convex-

composite structure allows for a natural design and analysis of iterative inversion

---

algorithms. To efficiently minimize a function with such structure, it is highly desirable to have its gradient available. However, a pure Newton method based on pursuing for second derivatives to solve for Equation (3.1) is often computationally infeasible. To address this issue, either a quasi-Newton (QN) method or a Gauss-Newton method taking advantage of the first-order derivatives can be used.

### 3.2.2 Linearized misfit function and Gauss Newton (GN) approach

The whole optimization process can be conducted by approaching the final optimized model iteratively (a 2-D case)

$$s_{k+1}(x, z) = s_k(x, z) + \delta s(x, z), \quad (3.4)$$

where  $\delta s(x, z)$  is the perturbation of the slowness at iteration  $k$ , and it will give rise to perturbation of the traveltimes. Such perturbation technique is also applied in acoustic anisotropic media to derive the linearized eikonal equation revealing the relationships of perturbation of anisotropic model parameters and arrival times. The total time field  $T_{total}(x)$  can be decomposed into two constituent parts: the reference time field  $T_{ref}(x)$  from background velocity model of current iteration; and the perturbed time field  $\delta T(x)$  from  $\delta\omega(x, z)$

$$T_{total}(x, z) = T_{ref}(x, z) + \delta T(x, z). \quad (3.5)$$

To obtain the relationship of the perturbed slowness  $\delta\omega(x, z)$  and corresponding variations in traveltimes  $\delta T_p$  of 2D case, in which direction of  $P$  is  $x$  and  $z$ , we have

---

the eikonal equation involving with perturbed time field expressed as

$$\begin{aligned} |\nabla T_p + \nabla \delta T_p|^2 &= |\nabla T_p|^2 + 2\nabla T_p \cdot \nabla \delta T_p + |\nabla \delta T_p|^2 \\ &= [s(x, z)]^2 + 2s(x, z) \cdot \delta s(x, z) + [\delta s(x, z)]^2, \end{aligned} \quad (3.6)$$

subtracting above equation by the unperturbed eikonal equation, we have

$$2\nabla T_p \cdot \nabla \delta T_p + |\nabla \delta T_p|^2 = 2s(x, z) \cdot \delta s(x, z) + [\delta s(x, z)]^2, \quad (3.7)$$

then dropping the high order term

$$\nabla T_p \cdot \nabla \delta T_p = |\nabla T_p| d\hat{l} \cdot \nabla \delta T_p = s(x, z) \cdot \delta s(x, z). \quad (3.8)$$

Here  $d\hat{l}$  as a dimensionless unit vector tangent to the raypath through subsurface model at certain position, thus we have  $\nabla T_p = |\nabla T_p| d\hat{l}$ . Now we define the directional derivative along  $d\hat{l}$  is  $d/d\hat{l} = d\hat{l} \cdot \nabla$ , which is only determined by the distribution of background velocity field, then Equation (3.8) could be simplified to

$$d\delta T_p/dl = \delta s(x, z). \quad (3.9)$$

After rearrangement, we have Fréchet derivative related to slowness variation

$$d\delta T_p/\delta s(x, z) = dl. \quad (3.10)$$

The Fréchet derivative is also called sensitivity kernel used for evaluating the sensitivity of the data to the corresponding slowness changes. For Equation (3.11), multiplying with ray-path integral from source point to receiver point located at  $(x_r, z_r)$ , then the formula could be expressed as

---


$$\delta T_p(x_r, z_r) = \int_{raypath} \delta s(x', z') dl. \quad (3.11)$$

The above equation indicates that the variation of traveltime induced by slowness change could be derived from integral of the slowness perturbation along the original raypath without calculating the new one. The raypath is determined by the implicit function of source point connecting to receiver point location.

Equation (3.11) is the basic formula of tomographic imaging using residual traveltime based on ray-tracing algorithm. It is an efficient computation since that we only need the original raypath generated at initial stage during the whole inversion process.

On the premise of guaranteeing the image quality, our main interest of developing such methodology is to improve the computation efficiency, since then we prefer to use finite-difference-based method to solve eikonal equation (forward modeling), and manipulate with adjoint-state technique as our inversion algorithm basics for calculating gradient of misfit function. The related computation efficiency analysis was already investigated in the first chapter.

For conciseness, assuming  $\frac{1}{v(\mathbf{x})^2} = w(\mathbf{x})$ , in which  $w$  is defined as slowness-squared, Equation (3.8) could be expressed as

$$2(\nabla T(\mathbf{x}) \cdot \nabla) \delta T(\mathbf{x}) = \delta \omega(\mathbf{x}) \quad \mathbf{x} \in \Omega \quad (3.12)$$

and boundary condition satisfies

---


$$\delta T(\mathbf{x}) = 0. \quad \mathbf{x} \in \partial\Omega \quad (3.13)$$

From Equation (3.12), we could build up a linearized relationship between the velocity perturbation  $\delta c$  and change of arrival time  $\delta T$  as follow

$$\delta T(x) = \frac{1}{2} (\nabla T(x) \cdot \nabla)^{-1} \delta \omega = \mathcal{F}_{\text{ek}}(\delta \omega) \quad (3.14)$$

We define linear operator  $\mathcal{F}_{\text{ek}}(\bullet) = \frac{1}{2} (\nabla T(x) \cdot \nabla)^{-1}(\bullet)$  as perturbed eikonal equation operator, and this linear operator  $(\nabla T(x) \cdot \nabla)$  acts as a ‘‘Jacobian tensor’’ to produce a matrix output, known as Jacobian matrix. It exhibits the sensitivity of perturbation of slowness giving rise to the change of arrival. Inserting Equation (3.14) to Equation (3.22), the misfit function reformulated to a Gauss-Newton (GN) subproblem (Li and Herman et al., 2011) involving with iterative linearization of  $\mathcal{L}(\mathbf{m})$  and solution of obtaining the optimum model update for each iteration of least-squares problems of the form

$$J(\delta \omega) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{\text{residual}}^{(k)}(\mathbf{x}) - \mathcal{F}_{\text{ek}}^{(k)}(\delta \omega)|^2, \quad (3.15)$$

where  $T_{\text{residual}}(\mathbf{x}) \triangleq T_{\text{obs}}(\mathbf{x}_s) - T_{\text{ref}}(\mathbf{x}_s)$  is the time residual at each iteration  $k$ . Specially, for tomographic inversion problem, the time residual is normally small and would not vary rapidly through the whole survey system, meaning that it will not perform strongly nonlinear at the solution. Hence, the GN methods can exhibits fast convergence (Courtier & Talagrand, 1987).

---

### 3.2.3 Adjoint PDE and gradient of misfit function

Minimization of  $J(\delta\boldsymbol{\omega})$  can obtain the optimal slowness perturbation  $\delta\boldsymbol{\omega}$ . In our approach, we use the adjoint-state technique to formulate a new adjoint PDE to obtain the gradient of the misfit function  $\frac{\partial J(\delta\boldsymbol{\omega})}{\partial \delta\boldsymbol{\omega}}$  without explicitly computing Fréchet derivatives

$\frac{\partial \delta T}{\partial \delta\boldsymbol{\omega}}$ . It will turn the inverse modeling problem into another forward modelling problem. In this section, we will derive the adjoint PDE of perturbed Eikonal equation through a Lagrange augmented functional technique and the model update can be efficiently calculated.

Now we introduce a Dirac masses  $\delta(\mathbf{x} - \mathbf{x}_r)$  by multiplying with misfit function  $J(\delta\boldsymbol{\omega})$  to derive its gradient more feasibly as follow

$$J(\delta\boldsymbol{\omega}) = \frac{1}{2} \int_{\Omega} |T_{residual}^{(k)}(\mathbf{x}) - \mathcal{F}_{ek}^{(k)}(\delta\boldsymbol{\omega})|^2 \cdot \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x}, \quad (3.16)$$

where  $\mathbf{x}_r$  denotes the position where the numerical sensors deploy at surface  $\partial\Omega$ . In practical, the above equation should be regarded as a summation, because that the residual time (or observed time) is only defined at the position where the receiver deploys.

Now we introduce a small amount change of slowness perturbation  $\widetilde{\delta\boldsymbol{\omega}}$  and its corresponding perturbation of perturbed arrival as  $\widetilde{\delta T}$ , by which the change of the misfit function is then given by

$$\delta J = \frac{1}{2} \int_{\Omega} \widetilde{\delta T} \cdot (T_{residual} - \delta T) \cdot \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x}. \quad (3.17)$$

From Equation (3.14), the perturbation in slowness is related to the perturbation of arrival time, and can be extended as following

$$\frac{\partial T}{\partial x} \frac{\partial \widetilde{\delta T}}{\partial x} + \frac{\partial T}{\partial y} \frac{\partial \widetilde{\delta T}}{\partial y} + \frac{\partial T}{\partial z} \frac{\partial \widetilde{\delta T}}{\partial z} = \frac{\widetilde{\delta \omega}}{2}, \quad (3.18)$$

by multiplying Equation (3.18) by Lagrange multiplier  $\lambda$  and adding it to Equation (3.17), we obtain a new expression for the change of energy (gradient of misfit function)

$$\begin{aligned} \delta J = & \frac{1}{2} \int_{\Omega} \widetilde{\delta T} \cdot (T_{residual} - \delta T) \cdot \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x} + \\ & \int_{\Omega} \widetilde{\delta T} \cdot \lambda \left[ (T_x \widetilde{\delta T}_x)_x + (T_x \widetilde{\delta T}_y)_y + (T_z \widetilde{\delta T}_z)_z + \frac{\widetilde{\delta \omega}}{2} \right] d\mathbf{x}. \end{aligned} \quad (3.19)$$

This arbitrary function  $\lambda$  will help to eliminate  $\partial \widetilde{\delta T}$ , whose evaluation is usually computationally infeasible for large scale problems. Then we apply integration by parts of second term over domain  $\Omega$  on the right-hand sides, and it yields

$$\begin{aligned} \delta J = & \frac{1}{2} \int_{\Omega} \widetilde{\delta T} \cdot (T_{residual} - \delta T) \cdot \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x} + \\ & \int_y \int_z \lambda T_x \widetilde{\delta T} \Big|_{x_{min}}^{x_{max}} + \int_x \int_z \lambda T_y \widetilde{\delta T} \Big|_{y_{min}}^{y_{max}} + \int_x \int_y \lambda T_z \widetilde{\delta T} \Big|_{z_{min}}^{z_{max}} + \\ & \int_{\Omega} \widetilde{\delta T} \left[ (\lambda \widetilde{\delta T}_x)_x + (\lambda \widetilde{\delta T}_y)_y + (\lambda \widetilde{\delta T}_z)_z \right] + \int_{\Omega} \frac{\lambda \cdot \widetilde{\delta \omega}}{2}. \end{aligned} \quad (3.20)$$

From above equation, we could obtain following two PDE equations as  $\lambda$  satisfying boundary condition as adjoint equation below

$$\lambda(\mathbf{x}) \vec{\mathbf{n}} \cdot \nabla T(\mathbf{x}) = T_{residual} - \delta T \quad \mathbf{x} \in \partial \Omega \quad (3.21)$$

where  $\vec{\mathbf{n}}$  is the unit vector perpendicular to the surface of numerical sensors  $\partial \Omega$ . And within

---

the subsurface  $\Omega$ ,  $\lambda$  is the solution of

$$\nabla \left( \lambda(\mathbf{x}) \cdot \nabla \widetilde{\delta T}(\mathbf{x}) \right) = 0. \quad \mathbf{x} \in \Omega \quad (3.22)$$

By combining two adjoint PDE equations as described above, we can calculate adjoint variable  $\lambda$  throughout the whole subsurface space  $\Omega$  by following procedure: Firstly, the arrival residual at the  $k$ -th iteration can be obtained by solving forward modeling through  $T_{obs}^{(k-1)}(\mathbf{x}_s) - T_{ref}^{(k-1)}(\mathbf{x}_s) - \mathcal{F}_{eikonal}^{(k-1)}(\delta\omega)$  of  $(k-1)$ -th iteration; Secondly, solving  $\lambda$  as its propagated back the residual along the ray-tube in to the source position of current model (Taillandier et al., 2009). The local scheme in algorithm of second step is adapted from the one proposed by Leung and Qian (2006), and its global scheme is identical to the FSW method originated from Zhao (2006). Eventually, the gradient of the misfit function could be obtained according to the last term on the right-hand side of Equation (3.20)

$$\nabla_{\delta\omega} J = \frac{\partial J(\delta\omega)}{\partial \delta\omega} = \int_{\Omega} \frac{\lambda(\mathbf{x})}{2}. \quad (3.23)$$

For a gradient-based iterative optimization such as GN approach, once we obtained the gradient of misfit function  $\nabla_{\delta\omega} J$ , the model update at each iteration could be performed following

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + \alpha_k \nabla_{\delta\omega} J, \quad (3.24)$$

where  $\alpha_k$  is step length computed by line-search method to guarantee a sufficient update towards decent direction of misfit function at each iteration.

In this chapter, our main purpose of adopting such adjoint-state technique is to obtain an

---

analytical solution of linearized adjoint operator  $\mathcal{F}_{ek}^\dagger: T_{residual} \rightarrow \delta J(\delta\omega)$  at  $k$ -th, and it can be expressed as

$$\nabla_{\delta\omega} J^{(k)} \triangleq \frac{\partial J(\delta\omega)}{\partial \delta\omega(\mathbf{x})} = - \sum_{s=1}^{Ns} \mathcal{F}_{ek}^{\dagger (k)} (T_{residual}^{(k)} - \delta T^{(k)}). \quad (3.25)$$

In fact, the term of  $T_{residual}^{(k)} - \delta T^{(k)}$  also represents the time residual for next iteration ( $k+1$ )-th, by which the misfit function gradient  $\nabla_{\delta\omega} J^{(k)}$  is derived from adjoint-state technique. Therefore, the Equation (3.25) can also be expressed as

$$\nabla_{\delta\omega} J^{(k-1)} = - \sum_{s=1}^{Ns} \mathcal{F}_{ek}^{\dagger (k)} (T_{residual}^{(k)}). \quad (3.26)$$

Now we have an analytical solution of misfit gradient  $\nabla_{\delta\omega} J$  equivalent to  $\mathcal{F}_{ek}^\dagger(T_{residual})$ , which is obtained in an efficient way by adjoint-state technique. Correspondingly,  $\mathbf{H}$  is Hessian matrix of  $\nabla_{\delta\omega} J$  with respect to  $\delta\omega(x)$  and  $\delta\omega(z)$

$$\mathbf{H} \triangleq \frac{\partial^2 J(\delta\omega)}{\partial \delta\omega(x) \partial \delta\omega(z)} = \sum_{s=1}^{Ns} \left( \mathcal{F}_{ek} \mathcal{F}_{ek}^\dagger - \frac{\partial \mathcal{F}_{ek}^\dagger}{\partial \delta\omega(z)} (T_{residual}^{(k)} - \delta T^{(k)}) \right). \quad (3.27)$$

As utilized in standard GN or quasi-Newton method, the second term in summation (3.27) is normally negligible and dropped, and the resulting approximated Hessian is

$$\widetilde{\mathbf{H}} \approx \sum_{s=1}^{Ns} \mathcal{F}_{ek}^\dagger \mathcal{F}_{ek}, \quad (3.28)$$

And the model update is approximated by

$$\delta\omega = - \left( \widetilde{\mathbf{H}} \right)^{-1} \cdot \nabla_{\delta\omega} J \approx - \left( \mathcal{F}_{ek}^\dagger \mathcal{F}_{ek} \right)^{-1} \cdot \nabla_{\delta\omega} J. \quad (3.29)$$

The class of GN method approximating the inversed Hessian matrix  $\widetilde{\mathbf{H}}^{-1}$  usually

---

involves the pseudo-inverse of the reduced Hessian as shown in Equation (3.28). This pseudo inverse of the reduced Hessian is given by the Jacobian operator  $\mathcal{F}_{ek}$  and its adjoint.

One benefit brought by GN method is that an explicit computation of the Hessian could be avoided, hence the computation efficiency could be guaranteed. However, the amount of solution of PDE's required in such GN subproblem at each iteration are still massive. Even though the manipulation of adjoint-state technique in tomographic inversion significantly improves the computation efficiency in large dataset application scenario, dimensionality reduction of problem scale still has its own practical significance. The main computational cost will now be brought up by solving eikonal equation for each source.

On the other hand, GN method assumes that the pseudo inverse Hessian matrix  $\widetilde{\mathbf{H}}^{-1}$  is diagonally dominant and semidefinite with ignoring the contributions from other off-diagonal elements, which only can be satisfied in some ideal case such that the acquisition aperture is wide enough. In practical, the acquisition aperture is often limited and ray coverage could vary dramatically leading to non-negligible contributions from off-diagonal elements. Since then, a standard GN method sometimes will not guarantee an optimal reconstruction of  $\delta\omega$ .

These two deficiencies will be two main concerns while we design a FAT tomographic inversion workflow using a GN method. To improve the computation efficiency on the premise of guaranteeing image quality, here are two major aspects of modifications we work on:

- 1.) Dimensionality reduction of least-square problem scale to relief computation burdens of forward calculation, meaning that a reduced GN subproblems;

- 
- 2.) Adopting sparsity regularization and compressive sensing technique to guarantee the sparsity-promoting and recoverability of reduced GN subproblems.

In the following sections, we will present attempts to solve above issues respectively, and eventually combine them together to present an online dictionary learning based sparsity promoting FAT tomography to achieve an optimize update of  $\delta\omega$  at each iteration.

### 3.4 Stochastic process

To deduce the number of required PDE to be solved in each inversion iteration, we utilize the idea from stochastic optimization to reduce input source. Such subsampling would introduce noise-like artifacts and cause the energy leakage due to the source interference. Since then, the weighed factor normally applied on the random source selection normally will be redraw to suppressing artifacts. The related two types of stochastic optimization adjusted for FAT tomography will be introduced in this section.

The idea of stochastic process is to define a batch size  $\tilde{K} \ll K$  of sub-dataset from full dataset, and solve the following inverse problem

$$\underset{\mathbf{m}}{\text{minimize}} \left\{ J(\mathbf{m}) = \mathbf{E} \left\{ \tilde{J}(\mathbf{m}; \mathbf{W}) \right\} \right\}, \quad (3.30)$$

where  $\mathbf{W}$  is a matrix with i.i.d. random entries with  $\tilde{K}$  columns. I will refer it to random batch encoding matrix, since it will randomly choose subset  $\tilde{K}$  out of full data collections. If  $\mathbf{W}$  is chose with unit covariance (i.e.,  $E\{WW^T\} = I$ ), we have

$$\mathbf{E} \left\{ \tilde{J}(\mathbf{m}; \mathbf{W}) \right\} = J(\mathbf{m}). \quad (3.31)$$

---

Instead of pursuing full misfit function and gradient,  $J$  and  $\nabla J$ , this approach intends to access to a “noisy” realization of reduced misfit function and its sub-gradient,  $\tilde{J}$  and  $\nabla\tilde{J}$ . The Equation (3.30) can be interpreted as a sample average process, by which an approximation of expectation with an error is depending on the batch size of  $\tilde{K}$ . In FWI application, such error often leads to source crosstalk, and it will decrease with larger  $\tilde{K}$ . In tomographic inversion, the artefact of ray tracing residual caused by sparse source sampling often contaminates the total gradient, and eventually will degrade the image quality. The energy from each single source cannot be balanced due to the uneven source sampling, especially in the vicinities of source position. Sample Average Approximation (SAA) and Stochastic Average (SA) are two main approaches to reduce problem dimensionality while suppressing the “noise” induced by subsampling.

### 3.4.1 Sample average approximation (SAA)

An intuitive approach is to pick a suitable batch size  $\tilde{K}$  that is large enough for equivalently replacing the expectation in Equation (3.31) by sample average, which is referred to Sample average approximation (SAA) (Nemirovski et al.,2009). In tomographic inversion application, the random batch encoding matrix  $W$  will actually turn to be a source number index vector, representing the selected sources that will be used in such stochastic optimization. The SAA approach will fix the batch selection  $W$  after it was drawn for all iteration. When some additional assumptions are satisfied, the optimal value of reduced misfit function will converge to the full one in most ideal case (Shapiro, 2003; Shapiro and Nemirovsky, 2005).

---

This approach is quite feasible and can be implemented with GN method to minimize the reduced misfit. However, the SAA approach is known to decay slowly while increasing  $\tilde{K}$ . And a proper batch size  $\tilde{K}$  that “large enough” to get rid of the noise is somehow hard to judge accurately.

### 3.4.2 Stochastic approximation (SA)

Another specialized stochastic optimization method to solve problem (3.31) directly, which is referring to the Stochastic Approximation (SA). It is an approach reminiscent of SA with independent weights  $W^k$  will be redrawn for each misfit function gradient update at every iteration  $k$  (Krebs et al. 2009; van Leeuwen et al., 2011a). The iterative update yielded by this approach is

$$\mathbf{m}^{k+1} = \mathbf{m}^k - \alpha_k \nabla \mathbf{s}^k, \quad (3.32)$$

where the search direction is a realization of the gradient:  $\mathbf{s}^k = \nabla \tilde{J}(\mathbf{m}^k; W^k)$ . The batch size is typically very small ( $K = \mathcal{O}(1)$ ) and  $\{\alpha_k\}$  are series of step length varied for each iteration. The step length here is a deterministic value that will be picked ahead of time. There are several conditions required to be hold for maintaining the convergency of SA algorithms to realize a “noisy” descent direction of  $\mathbf{s}^k + \sigma^k$ , as  $\sigma^k$  is a random noise term caused by stochastic process. These conditions were provided and proved by Betrsekas and Tsitsiklis (2000) in proposition 3, and was further reorganized by (Herrmann and Li et.al, 2012) as follows:

1.  $J$  is differentiable with  $\nabla J$  Lipchitz continuous. It limits the misfit function gradient  $\nabla J$

---

to the upper bound, by which the  $J$  owns global extreme value.

2. The expectation of random noise  $\mathbf{E}[\sigma] = 0$ , meaning that random noise is incoherent.
3. The expectations of the search directions are descent for  $J$ , i.e.,  $\nabla J(\mathbf{m}^k)^\top \mathbf{E}[\nabla \tilde{J}(\mathbf{m}^k; W^k)] < 0$ . It will guarantee the reduced misfit function  $\tilde{J}$  will decrease during iterative update.
4. There exist three positive deterministic constants  $c_1$ ,  $c_2$  and  $c_3$ , such that
  - i.  $c_1 \|\nabla J\|^2 \leq -\nabla J(\mathbf{m}^k)^\top \nabla(s^k + \sigma^k)$ ,
  - ii.  $\|s^k + \sigma^k\| \leq c_2(1 + \|\nabla J\|)$ , and
  - iii.  $\mathbf{E}[\|\sigma\|^2] \leq c_3(1 + \|\nabla J\|^2)$ .
5.  $\sum_{k=0}^{\infty} \gamma_k = \infty$ ,  $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$ . It will guarantee the step length will decrease during iterative optimization. A commonly used deterministic step size is  $\gamma_k \propto \frac{1}{k}$ .

Similar to the FWI application research indicates by (Herrmann and Li, et al., 2011), the modified GN algorithm is lacking rigorous convergence theory (above condition 3 cannot be guaranteed for each model update  $\delta\omega_k$  for our efficient FAT tomographic inversion. However, the SA method actually has better performance than SAA in terms of computational efficiency for many classical mathematical problem (Kim et al., 2015). We will combine SAA and our proposed gradient optimization, named as SAA+ method, while combing SA and gradient optimization and named as SA+ method. These two methods will be compared to decide which one is more appropriate for FAT tomography problem in the synthetic data test.

---

Sun and Zhang (2017) proposed a so called Fast Stochastic Approximation (FSAT) method to improve the computation efficiency by introducing SA theory. In their practical implementation, they managed to employ only a small percentage of data in inversion, by which a sub dataset of  $\widetilde{K}$  was established through randomly extracting  $\mathcal{L}$  observations (FAT picks) from full dataset. Hence the stochastic optimization will be reformulated to

$$\underset{\mathbf{m}}{\text{minimize}} \left\{ J(\mathbf{m}) = \mathbf{E} \left\{ \widetilde{J}(\mathbf{m}; \boldsymbol{\xi}^k) \right\} \right\}, k = 1, k_P. \quad (3.33)$$

Here  $\boldsymbol{\xi}^k$  is the  $L \times 1$   $k^{\text{th}}$  vector of FAT picks randomly extracted from full dataset. They set the random extraction obeying the uniform random number integer distribution with no repetition of picks in  $k^{\text{th}}$  iteration. Furthermore, they follow the same fashion of stochastic approximation to redraw a new  $\boldsymbol{\xi}^k$  vector for each iteration. The averaging over previous model iterates is needed in practical implementation of FAST, by which the optimal estimator of tomographic model can be calculated as

$$\mathbf{m}_{SA} = \frac{1}{k_P} \sum_{k=1}^{k_P} \mathbf{m}_{SA}^k. \quad (3.34)$$

$\mathbf{m}_{SA}^k$  is the model update estimated from each iteration  $k$ . They find that the  $k_P$  could be equal to 1, if an appropriate batch size of subset  $\boldsymbol{\xi}^k$  could be obtained. In this way, the reduced “noisy” realization of model update  $\mathbf{m}_{SA}$  could replace the realization of  $\mathbf{m}$  from full dataset approximation, i.e., such that  $\mathbf{m}_{FSAT} = \mathbf{m}_{SA}^1$ . However, they did not illustrate in detail that how to pre-define an “appropriate size” ahead of time of inversion. Despite the lack of rigorous mathematical proof, such argument is still reasonable as following the conclusion of literature

---

from Nemirovski et al. (2009). Besides, in a sense of practical application, they already showed the feasibility of redrawing random vectors  $\xi^k$  to improve the FAT tomographic inversion efficiency. However, the sole stochastic process will induce the artefacts that degrade the final image quality. Therefore, to improve image quality, we will further extend our work to incorporate sparsity promoting property as a remedy for modifying such dimensionality-reduced GN subproblems. In our proposed method, we will follow the same fashion of SA applied to our modified GN method imposing with sparsity promotion technique under the CS framework, to reduce the artefacts induced by random source encoding. However, we will use different random encoding vector  $\xi^k$  to meet the requirement of CS framework.

### 3.4.3 Reduced GN problem based on stochastic approximation

At this point, we could directly bring ideas from stochastic optimization with our original GN problem to formulate a reduced GN problem into a modified misfit function, given by

$$\tilde{J}(\delta\omega) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{residual}^W{}^{(k)}(\mathbf{x}) - \mathcal{F}_{ek}^W{}^{(k)}(\delta\omega)|^2 . \quad (3.35)$$

In the following section, an efficient sparse representation technique based on online orthonormal dictionary learning (ODL) will be introduced into model update  $\delta\omega$ , to further imposing sparsity promoting for this reduced GN problem.

### 3.5 Sparsity-promoting by sparse orthonormal transformation (SOT)

For seismic tomography, the traditional way of representation of model normally expanded

---

on an either directional or geometric wavelets, such as curvelet, shearlets and coutourlets, to provide a multigrid description of the model in different scales based on the properties of orthogonal or bi-orthogonal of their fixed dictionary basis. However, most of these methods are samples in regular spaced with predefined, nonadaptive and highly redundant function frames as imposing piecewise smoothness representation of smooth and discontinuous features of subsurface velocity structure. In contrast, the dictionary learning (DL) has advantages of offering to learn a set of atoms with sparser representation of model as a training set while adapting the transform to nonintuitive signal regularities.

In our proposed method, we will adopt a s SOT method based on DL following series works from Sezer et al. (2011 and 2015) and Zhu and Liu et al. (2015), to exploit the sparsity of  $\delta\omega$  with highly computation efficiency and feasibility beyond some other extensively used K-SVD DL algorithm (Chen et al., 1998; Mallat and Zhang, 1993; Tropp and Gilbert, 2007). The orthonormality imposing on dictionary could provide a natural approach for solving the L0 or L1 norm regularized optimization problem while reducing computation complexity. The whole sparse transform method composed of two parts: online orthonormal dictionary learning (ODL) and dictionary-based blockwise transformation.

### 3.5.1 Online orthonormal dictionary learning

The main idea of realizing such online ODL algorithm is to introduce an orthonormal dictionary  $\mathbf{D}$  such that  $\mathbf{D}^T \mathbf{D} = \mathbf{I}$ , and solving two sub-problems (L0 norm and L2 norm optimization) alternatively with an initial generic dictionary. In our implementation, we use an

---

initial arbitrary dictionary obtained by discretized cosine transformation (DCT). Imposing orthogonality on  $D$  could void repeated columns operations appeared in K-SVD algorithms (Aharon et al., 2006; Rubinstein et al., 2013), resulting in high computational demanding. It only involves with three matrix multiplications and one SVD operation to generate both the sparse coefficients and the learned dictionary at a single learning iteration. To adapt ODL with stochastic process, we use online learning technique proposed by Bousquet and Bottou (2007) to minimize an expected misfit function instead of empirical one. The workflow of online ODL is shown in box 4.1. The minimization of expected misfit function does not depend on the number of patches, but instead on the stochastic characteristics of the training patches (Zhu and Liu, et al., 2015). In each learning iteration, a batch of training patches will be extracted from either a full dataset (initial iteration) or a snapshot of a dynamic dataset. Meanwhile, the updated dictionary from previous learning iteration will anticipate in updating both sparse coefficients and dictionary learning at current iteration, which is referred to “online” manipulation. In this way, it will always maintain the sparse representation of dynamic data and add up the historical information to adaptively conduct our DL iterative process. Intuitively, online ODL is time efficiency since that it not requires a completely new dictionary to be generated at each learning iteration. Besides, the learning speed will dramatically increase along with progression of inversion.

---

**INITIALIZATION:**

1. Initializing overcomplete dictionary  $\mathbf{D}_0$  generated by DCT.
2. Initializing sparse coefficients matrix  $\mathbf{A}_0$ .

**OUTER LOOP:** <t times iteration (t=1, 2, ..., T)>

Stochastic approximation by drawing mini-batch from training dataset pool and obtain DBT of  $\delta\omega$  as training sample matrix  $\mathbf{Y}^{(t)}$

**INNER LOOP:** <Repeat until convergence meeting stopping rule>

**Sparse coding stage:**

- a. Optimizing  $\hat{\mathbf{A}}$  with hard-thresholding L1 norm optimization (sparse coding stage):

$$\hat{\mathbf{A}} = \underset{\mathbf{A}}{\operatorname{argmin}} (\|\mathbf{Y}^{(t)} - \mathbf{D}\mathbf{A}^{(t)}\|_F^2 + \lambda\|\mathbf{A}^{(t)}\|_0) \text{ s.t. } \|\mathbf{A}\|_0 \leq \lambda.$$

- Sparse coefficients from  $\mathbf{A}$  less than a certain thresholding value  $\lambda$  will be eliminated.
- $\mathbf{P}^{(t)} = \beta^{(t)}\mathbf{P}^{(t-1)} + \mathbf{X}^{(t)}[\mathbf{Y}^{(t)}]^T$  is gradually accumulating newer information from previous iteration by a certain learning rate  $\beta^{(t)}$ ;

**Dictionary update stage:**

Optimizing  $\hat{\mathbf{D}}$  by solving an orthogonal Procrustes problem using SVD:

$$\hat{\mathbf{D}} = \underset{\mathbf{D}}{\operatorname{argmin}} \|\mathbf{Y}^{(t)} - \mathbf{D}\hat{\mathbf{A}}^{(t)}\|_F^2 \text{ s.t. } \mathbf{D}^T \mathbf{D} = \mathbf{I}.$$

### 3.5.2 Dictionary-based blockwise transformation

Assuming that the length of dictionary atoms (learned from ODL) is  $N$ , which is same as localized patches (pixelized groups) length  $N_{a1} \times N_{a2}$  divided from original image (model update  $\delta\omega$ ). These patches will be directly used for dictionary training and later processing. For simplicity, we also assume that the image could be exactly divided into  $N_{patches} = N_{p1} \times N_{p2}$  patches. For the case that image size cannot be divided patch size, a zero or wrapping extended boundary will be applied to original image. The combination of extracting operator  $\mathbf{R}$  and its adjoint  $\mathbf{R}^\dagger$ , can perform an invertible process, either to recover the model update  $\delta\omega$  from local sparse coefficients  $\alpha$  by

$$\delta\omega = \sum_i^{N_{patches}} \mathbf{R}_i^\dagger [\mathbf{D}(\mathbf{R}_i(\alpha))] = \mathcal{D}(\alpha), \quad (3.36)$$

or to mapping  $\delta\omega$  into  $\alpha$  by

$$\alpha = \sum_i^{N_{patches}} \mathbf{R}_i^\dagger [\mathbf{D}^\dagger(\mathbf{R}_i(\delta\omega))] = \mathcal{D}^\dagger(\delta\omega). \quad (3.37)$$

where SOT synthesis operator  $\mathcal{D} \triangleq \sum_i^{N_{patches}} \mathbf{R}_i^\dagger \mathbf{D} \mathbf{R}_i$ , and SOT analysis operator

$\mathcal{D}^\dagger \triangleq \sum_i^{N_{patches}} \mathbf{R}_i \mathbf{D}^T \mathbf{R}_i^\dagger$ . This is so called dictionary-based block-wise transform that have been

widely used in different DL algorithms at present. Essentially, it is a patch processing technique for dictionary learning, and the extracting operator  $\mathbf{R}$  anticipating into the algorithm will have no influence on the algorithm convergency.

---

### 3.5.3 Practical implementation of online blockwise ODL

From box 4.1, there are several parameters needed to be selected cautiously to achieve good performance of online ODL coming into practical implementation.

#### 1. Learning rate $\beta$ in online ODL

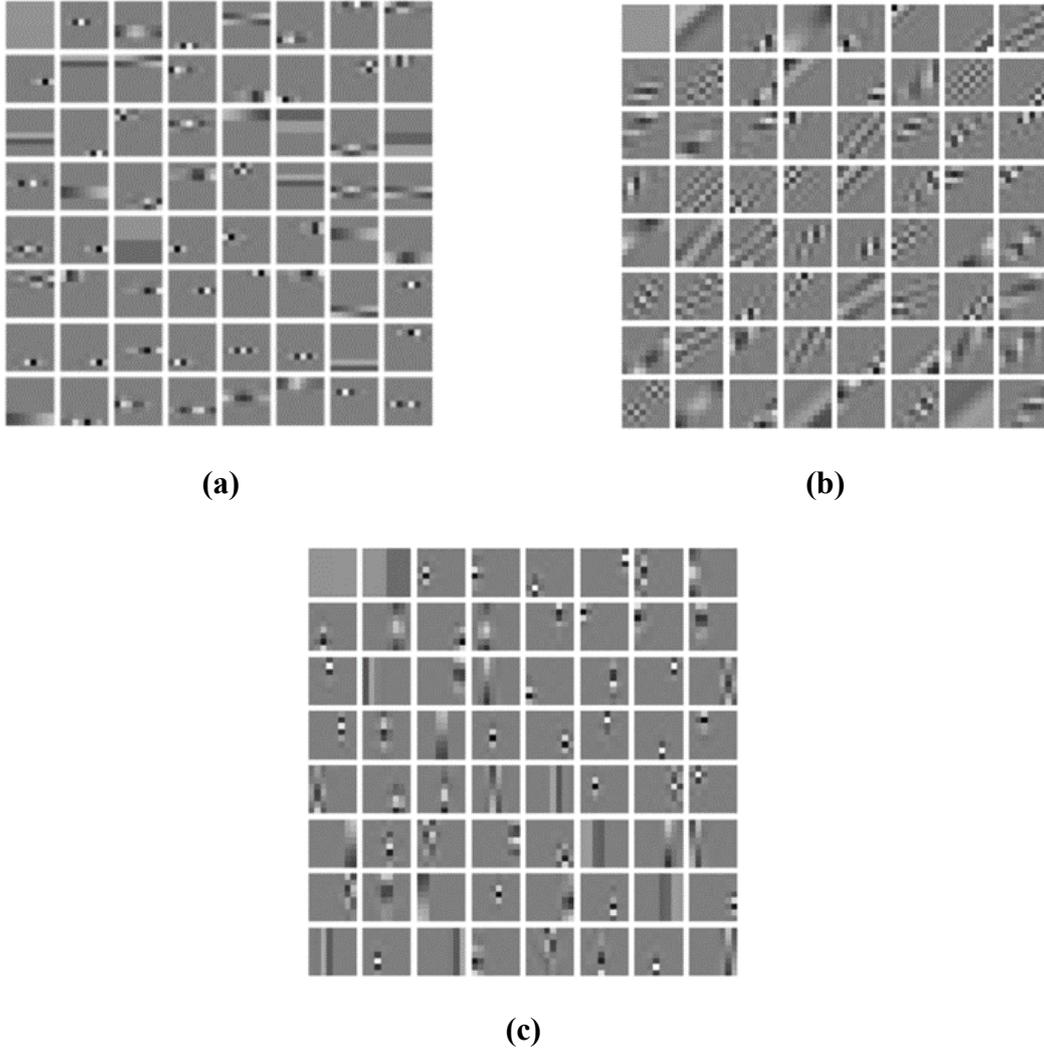
The algorithm will gradually rescale the older information from previous iterations during iterative learning. This scale factor is so called a “learning rate” parameter, playing a role as weight to amplify the contribution of older information as iteration carrying on. In practical application, its often relates with batch size that we used in stochastic optimization, and we normally set

$$\beta^{(t)} = \frac{t \times \text{batchsize} + 1 - \text{batchsize}}{t \times \text{batchsize} + 1},$$

where  $t$  is iteration number. In this way, the learning rate will eventually increase to 1, so that the large portion of accumulated old information will be dominant.

#### 2. Trained dictionary and sparse coefficients

Like other DL method, the ODL method is capable of adapting to nonintuitive signal regularities beyond piecewise smoothness (Zhu and Liu, et al., 2015). Here is a simple test shows the capability of ODL method capturing the specific signal patterns in sparse domain for given signals with gradient angle at (a) 0-degrees, (b) 45 degrees and (c) 90 degrees in Figure 3-2. The image of original signals has been transformed into blocks follow the dictionary-based block transform as described in section 3.5.2. Each block in Figure 3-2 represents an atom learned by ODL, and we could see that all of them maintain directional features of original signals with different gradient angles.



**Figure 3-2:** Learned dictionaries by ODL for signals with different gradient angle features at (a) 0-degree, (b) 45-degree and (c) 90-degree.

Similarly, all different signal patterns will be recognized and classified based on the oriented gradient in our implementation. The model update  $\delta\omega$  for each tomographic inversion iteration will be evaluated after divided into localized patches. Normalization is necessary step to balance the energy of model updates from each iteration. After that, two directional derivatives will be applied on all patches, to obtain both vertical and horizontal direction gradients for every single patch pixel  $(i,j)$  respectively, noted by  $\mathbf{g} = \{g_h, g_v\}$ .

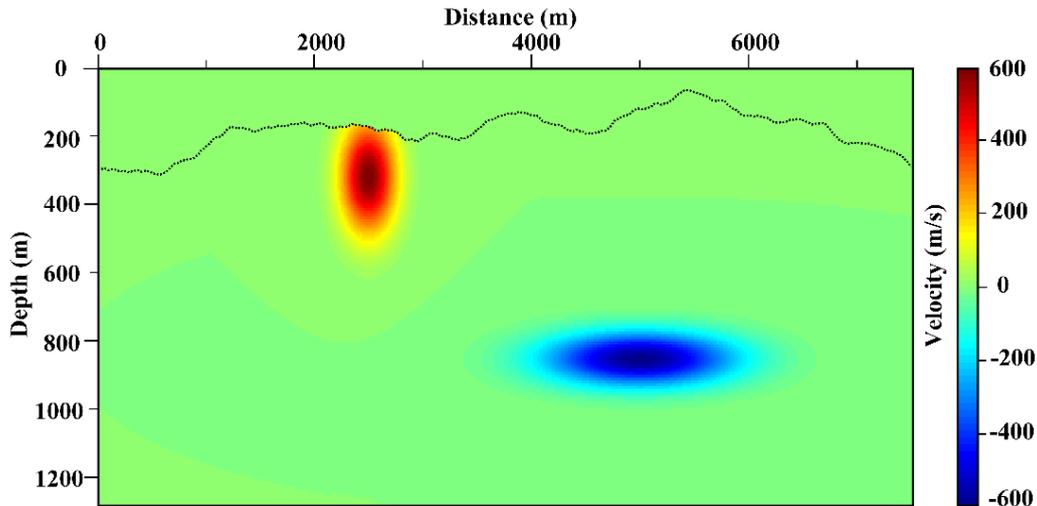
---

Then the gradient energy of these pixels will be computed by  $|\mathbf{g}(i,j)| = \sqrt{g_h^2(i,j) + g_v^2(i,j)}$ .

Correspondingly, the gradient angle can be derived from  $g_{angle} = \{(\arctan[g_h^2(i,j)/(g_v^2(i,j) + \epsilon)] + \pi) \times 180\}/\pi$ , where  $\epsilon$  is a small number to avoid zero in denominator. The gradient angle ranges from  $-90^\circ$  to  $90^\circ$ , where angle of  $0^\circ$  represents downward vertical direction point to subsurface model. In each localized patch, the gradient energy corresponding to all possible directional structures will be scanned and extracted. This structural or geometrical information are sparse in the model space, and it could be properly represented by a combination of dictionary and sparse coefficients. For those subsampling artifacts with no structural features, most of them represented by sparse coefficients in small values and eventually will be suppressed through ODL by certain sparse level. More details about the benefits brought by sparsity promotion will be discussed in later section.

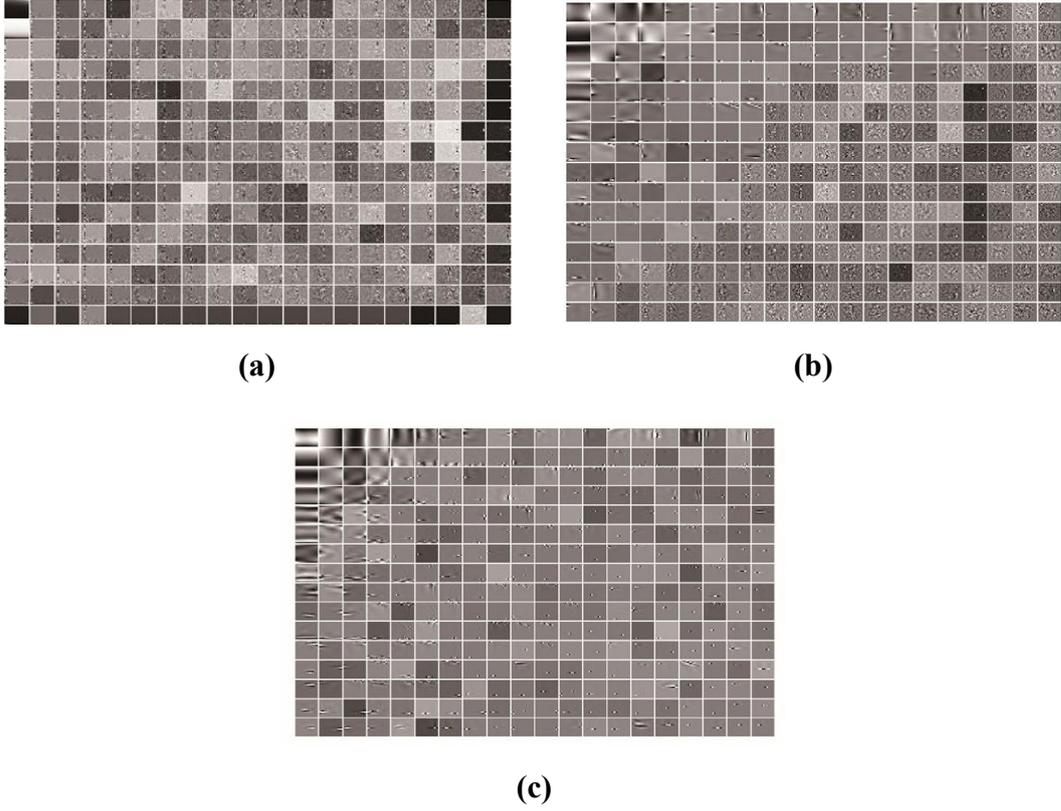
### 3. Choice of $\lambda$ : sparsity level of learned dictionary

During iterative update for both dictionary  $\mathbf{D}$  and sparse coefficients  $\boldsymbol{\alpha}$  by ODL, the Lagrange multiplier  $\lambda$  in Box 3.1 controls the tradeoff between speed of convergence and capability of sparse representations.  $\boldsymbol{\alpha}$  with absolute values smaller than  $\lambda$  will be hard threshold to zero.



**Figure 3-3:** Simple two-layer velocity model consist of smoothly varied velocity anomalies without sharp boundaries.

To illustrate the content of learned dictionaries, the trained dictionaries by ODL are shown for our simple toy model. As shown in Figure 3-3, this model is quite simple with only two smoothly varied velocity anomalies without sharp boundaries. Since then, one could only expect energy concentrated in a few atoms, and it is confirmed by Figure 3-4. The atoms with energy concentrations are mostly dominated by smooth-continuous features. Since learned dictionaries are adapted to specific data, they could more properly model data with sparsest coefficients than pre-determined dictionaries, such as wavelet or curvelet with prescribed frame. As value of  $\lambda$  increased, the number of active sparse coefficients is decreasing. Noting that for an extreme case of  $\lambda = 1$ , the learned dictionary will degrade to a unit matrix.



**Figure 3-4:** Learned dictionaries by ODL with different values of (a)  $\lambda= 0.8$ ,  $\lambda= 0.06$  and  $\lambda= 0.02$ .

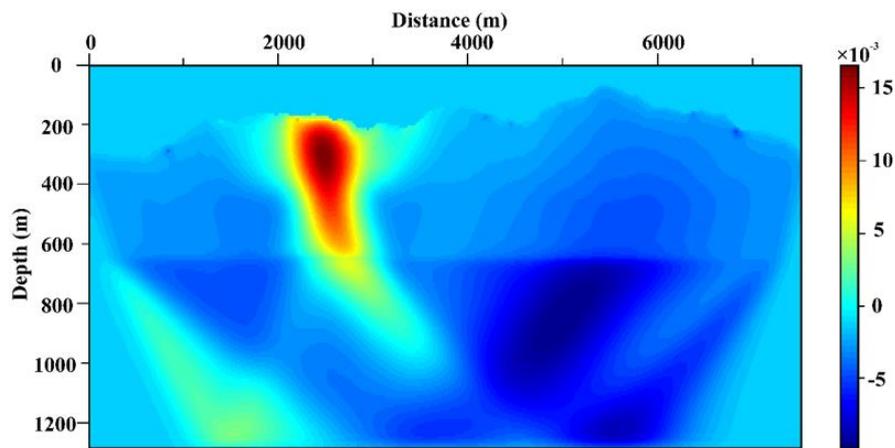
A nonlinear approximation (NLA) evaluation method could be performed to testify the recoverability of the learned orthonormal dictionary  $\mathbf{D}$  for a model update  $\delta\omega$ . This evaluation can be performed using only small portion of coefficients  $\tilde{\alpha}$ , and the mean square error (MSE) will be evaluated between the original and reconstructed signal as:

$$MSE(\delta\omega; \mathcal{D}, n) = \left\| \delta\omega - \mathcal{D}\tilde{\alpha} \right\|_2^2.$$

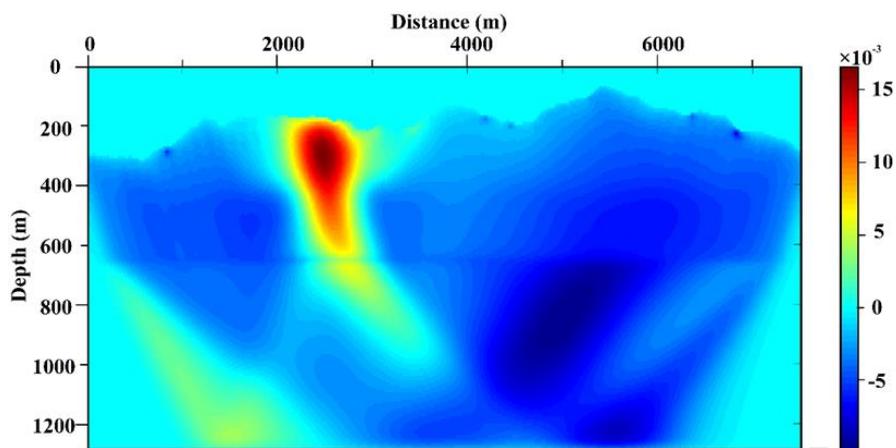
We conduct the similar NLA test to evaluate the performance of a learned orthonormal dictionary  $\mathbf{D}$  from a typical misfit function gradient for travelttime tomography. We choose one gradient (perturbation) update from tomographic inversion used as training model shown in Figure 3-6a. The initial dictionary is trained based on this training model. Then another gradient from Figure 3-6b is the test model for reconstruction using different  $\lambda$  values. The

---

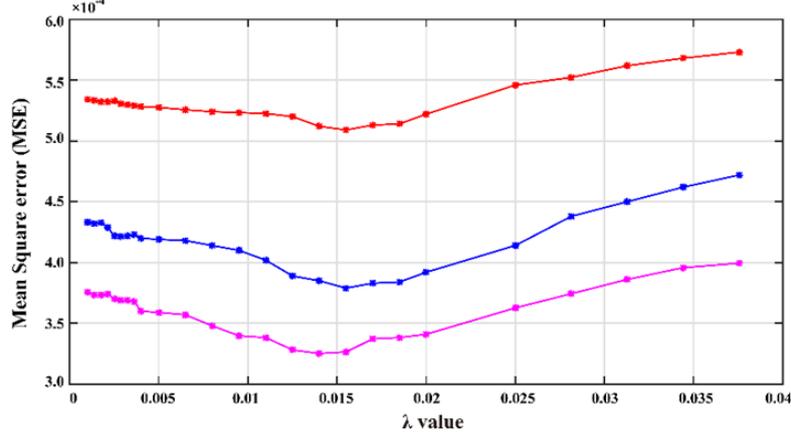
NLA curves in Figure 3-6c indicates the relationship between MSE and  $\lambda$  value for different number of sparse coefficients (5%, 10% and 15%) involved in DL reconstruction. All three curves show similar trends and optimal reconstructions delivered within range of  $\lambda \in (0.01, 0.02)$ . In practical, the sparsity level varied with different dataset, and it is always meaningful to conduct such NLA test for selecting ideal  $\lambda$  to balance the trade-off between computation efficiency and capability of sparse representation by learned dictionary.



(a)



(b)



(c)

**Figure 3-5:** NLA test results from two model updates. (a) Training Model. (b) Testing Model. (c) NLA curves for different number of sparse coefficients (5% in red, 10% in blue and 15% in magnet) used in ODL.

### 3.6 SA+ Method: a modified GN method with gradient optimization

At this point, we can reformulate the reduced GN subproblem into L1-norm sparsity constrained GN subproblem after imposing the orthonormal dictionary representation

$$J(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{residual}^{(k)}(\mathbf{x}) - \mathcal{F}_{ek}^{(k)}(\mathcal{D}^{(k)}(\boldsymbol{\alpha}))|^2 \text{ s.t. } \|\boldsymbol{\alpha}\|_1 \leq \tau, \quad (3.38)$$

and the optimized model update  $\widehat{\delta\boldsymbol{\omega}}^{(k)}$  could be obtained by

$$\widehat{\delta\boldsymbol{\omega}} = \mathcal{D}^{(k)}(\boldsymbol{\alpha}). \quad (3.39)$$

The sparsity level  $\tau$  provides L1-norm constraints on sparse coefficients  $\boldsymbol{\alpha}$  to enforce them to be smaller or equal to constant  $\tau$ . Then we impose SA method with  $W^{(k)}$ , the dimensionality reduced linearized time residual  $T_{residual}^{W^{(k)}}$  and operator  $\mathcal{F}_{ek}^{W^{(k)}}$  allows forward modeling on batches of reduced sources with much less PDE's solving to relief total computation cost. Now the sparsity-promotion misfit function reformulates into dimensionality

---

reduction one as following

$$J^W(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{residual}^{W^{(k)}}(\mathbf{x}) - \mathcal{F}_{ek}^{W^{(k)}}(\mathcal{D}^{(k)}(\boldsymbol{\alpha}))|^2 \text{ s.t. } \|\boldsymbol{\alpha}\|_1 \leq \tau. \quad (3.40)$$

The renewals of subsampling matrix  $W^{(k)}$  allows to realize the dimensionality reduction applied on both land and marine data, since that the receiver positions corresponding to a certain source is not required to be fixed during inversion. The stochastic process presented here does not strictly meet all conditions requested by convergency theory as presented in Proposition 3 from Betrsekas and Tsitsiklis (2000). However, the practical implementation results of FAT problem imply a significant convergency rate improvement as we resample the matrix  $W^{(k)}$  for each iteration. Additionally, the recovery is improved since that the bias introduced by fixed random sampling has been removed by different combinations of random subsampling.

One possible effort to suppress noise raised from SA or SAA relies on averaging model updates over previous iterations. Especially, in our case, the eikonal equation solver as forward modeling method sometimes will give rise the spikes in boundary values caused by an irregular topography, and the source or ray tracing signatures due to low dense sampling in filed survey. These unsmooth gradient spots may cause inversion unstable or render the model updates trapped in local minima (Huang et al., 2012). To guarantee the well-posedness of the inversion problem, one usually would apply some gradient preconditioner, e.g., a Gaussian filter applied on the gradient of each iteration to smooth out these artifacts (Leung and Qian, 2006). However, simple averaging process will also impair the salient information contained in gradient, leading

---

to unstable convergency and poor resolution in final inverted results. Hence, we are looking forward for preserving structural information from gradient in sparse domain. With aid of sparsity domain promotion on gradient, its main features from wavefront arrivals will be preserved and interferences caused by subsampling will be suppressed.

In the next section, we will make effort on giving some insight of connection between our sparsity regularization (or sparsity promotion) and compressive sensing theory.

### 3.6.1 Compressive sensing and signal reconstruction problem

The compressive sensing (CS) provides theory basis that one can recover a signal with compressibility in some sparse domain, from severe subsampling by solving a sparsity-promoting program (Candes,2006; Donoho,2006). The core problem is to solve an underdetermined linear system

$$\mathbf{y} = \Phi \mathbf{f}, \quad (3.41)$$

where  $\Phi$  is measurement matrix and  $\mathbf{f}$  is signal that could be sparsely represented in some transform domain, such as  $\mathbf{f} = \Psi \boldsymbol{\alpha}$ . In our case, signal (or model updates) could be expressed by a global SOT synthesis operator  $\mathcal{D}$  acting on sparse coefficients  $\boldsymbol{\alpha}$ , such that  $\mathcal{D}(\boldsymbol{\alpha})$ .

Hence, we have

$$\mathbf{y} = \Phi \mathcal{D}(\boldsymbol{\alpha}). \quad (3.42)$$

Denoting  $A = \Phi \mathcal{D}$ , the problem alternated to the sparsest solution of  $\mathbf{y} = A\boldsymbol{\alpha}$ , or

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad A\boldsymbol{\alpha} = \mathbf{y}, \quad (3.43)$$

---

where  $\|\cdot\|_0$  is the L0 norm given by the number of nonzero elements from all possible subset columns of  $A$ . The solution of equation (3.43) tends to have the fewest nonzero elements, or the sparsest solution while maintaining the recoverability. It is typically an “NP-hard” situation with almost prohibitive computation for large scale problem. The development of CS theory sheds light on solving this problem by proving that:  $A$  and  $\alpha$  should satisfies some condition, then the solution can be recovered by convex optimization

$$\min_{\alpha} \|\alpha\|_1 \text{ s.t. } A\alpha = \mathbf{y}, \quad (3.44)$$

And it is known as Basis Pursuit (BP). The success of solving this problem is depending on the sparsity level of  $\alpha$ , proper subsampling and low mutual coherence of  $A$ . The mutual coherence is the maximum off-diagonal entry of  $A^H A$ . Low mutual coherence could guarantee the capability of sparse signal recovery, since that it allows to separate two different signals with same sparsity level after random sampling. Low mutual coherence is equivalent to the well-known Restricted Isometry Property (RIP).

The original BP problem will alter to Basis Pursuit Denoise (BPDN) problem with noise add in

$$\min_{\alpha} \|\alpha\|_1 \text{ s.t. } \|A\alpha - \mathbf{y}\|_2 \leq \sigma \quad (3.45)$$

where  $\sigma$  is expected noise level in data. To solve it efficiently, it can be reformulated into two equivalent formulates: QP problem

$$\min_{\alpha} \|A\alpha - \mathbf{y}\|_2^2 + \epsilon \|\alpha\|_1, \quad (3.46)$$

---

and least absolute shrinkage and selection operator (LASSO) problem

$$\min_{\boldsymbol{\alpha}} \|A\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_1 \leq \tau \quad (3.47)$$

The solution of above three types of problems will be the same if every  $\sigma$  owned its unique  $\epsilon$  and  $\tau$ . Somehow it is unrealistic to determine these values ahead of time. Since then, many algorithms for solving problem of BPDN type is to gradually adjust  $\epsilon$  or  $\tau$  in each iteration.

### 3.6.2 LASSO-based approach and modified GN method

Among all these reconstruction methods, the LASSO-based approach has some advantages superior than other ones (Herrmann and Li et al., 2011). The formulation (3.47) could be efficiently solved by a Spectral Projected Gradient (SPG) method for a large-scale CS problem in seismic exploration (Hennenfent et al., 2008). The corresponding sparsity level  $\tau$  could be adaptively obtained follow its value function continuation analysis results from ‘‘Pareto trade-off curve’’ (van den Berg and Friedlander, 2008). It gives a systematic way to determine sparsity level  $\tau$  for each iteration. The practical implementation details on choice of  $\tau$  in our approach will be discussed in later sections.

Now we will make comparison in several aspects of modified GN subproblem with the LASSO problem to see the similarities among them, and eventually established the connection between sparsity promotion and CS technique. Recall the reduced GN subproblem based on stochastic optimization (3.35)

$$\tilde{J}(\delta\boldsymbol{\omega}) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{residual}^{W(k)}(\mathbf{x}) - \mathcal{F}_{ek}^{W(k)}(\delta\boldsymbol{\omega})|^2 \quad (3.48)$$

---

It preserves the same convex composite structure as the problem of full misfit. Hence, we could solve it following an original GN method by updating model iteratively

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + \alpha_k \delta \boldsymbol{\omega}_k, \quad (3.49)$$

where  $\alpha_k$  is step length solved by line search method. Here are two important aspects we need to look into:

1. **Mutual coherence of  $[\mathcal{F}_{ek}^{W(k)}]^T \mathcal{F}_{ek}^{W(k)}$ :** To follow the guidelines from CS theory, this dimensionality reduced eikonal equation Jacobian should be mutual coherence. The important difference between CS and our problem settings is that we need to solve for a tall system of equations to invert for GN updates. Since then, one could expect a low mutual coherence of  $[\mathcal{F}_{ek}^{W(k)}]^T \mathcal{F}_{ek}^{W(k)}$ , if it was near unitary with incoherent off-diagonals. For  $[\mathcal{F}_{ek}^W]^T \mathcal{F}_{ek}^W$ , there were massive successful research works related with steepest gradient or steepest descent method by assuming  $[\mathcal{F}_{ek}^W]^T \mathcal{F}_{ek}^W$  is either nearly unitary or diagonal matrix applied in tomographic inversion. Since then, though lacking of rigorous proof indicating the low mutual coherence of this term, we could still expect this CS-type of argument to be hold. As for random sampling matrix  $W$  and SOT operator  $\mathcal{D}$ , they all possessed presumptions of  $WW^T = \mathbf{I}$  and  $\mathcal{D}^T \mathcal{D} = \mathbf{I}$ , respectively. Hence, the low mutual coherence property would also hold for  $[\mathcal{F}_{ek}^{W(k)}]^T \mathcal{F}_{ek}^{W(k)}$  and  $[\mathcal{F}_{ek}^{W(k)} \mathcal{D}^{(k)}]^T [\mathcal{F}_{ek}^{W(k)} \mathcal{D}^{(k)}]$ .
2. **Sparsity of modified GN search direction in orthonormal dictionary:** As already noted, the gradient of misfit function  $\nabla_{\delta \boldsymbol{\omega}} J$ , could be optimally represented in a sparse space supported by the orthonormal dictionary. The solution or model update of a standard GN

---

method is given by

$$\delta\omega = - (\mathcal{F}_{ek}^\dagger \mathcal{F}_{ek})^{-1} \cdot \nabla_{\delta\omega} J. \quad (3.50)$$

The approximated Hessian represented by  $(\mathcal{F}_{ek}^\dagger \mathcal{F}_{ek})^{-1}$  could be treated as a scale factor, which also can be sparsely represented by orthonormal dictionary. Since then,  $\delta\omega$  also could be sparsely represented by orthonormal dictionary, no matter how far the current model from true solution.

Based on the observations above, we could argue that the LASSO problem is well-tailored to ideas related to sparsity promotion and CS as our modified GN subproblem based on stochastic optimization, expressed as:

$$J^W(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{residual}^W{}^{(k)}(\boldsymbol{x}) - \mathcal{F}_{ek}^W{}^{(k)}(\mathcal{D}^{(k)}(\boldsymbol{\alpha}))|^2 \text{ s.t. } \|\boldsymbol{\alpha}\|_1 \leq \tau. \quad (3.51)$$

Our approach now could be thought as pursuing for a sparse direction from a reduced dataset of severe subsampling. Consequently, the linearization of reduced misfit function, subsampling measurements and sparse signal recovery will rise a certain noise level for each subproblems. This unique noise level in each subproblem is corresponding to a sparsity level  $\tau$ . Therefore, the remain question is how to choose the right sparsity level  $\tau$  for approaching to an optimal solution of (3.40). The global convergency theory for this modified GN algorithm refers to literature from Herrmann and Li (2015). In their proof, each LASSO subproblem at  $k$ -th yields a descent direction for full nonlinear SAA problem for any  $\tau_k > 0$ . Nonetheless, a systematic way to select proper  $\tau_k$  in practical implementation. This issue will be addressed

in section 3.7.2.

### 3.6.3 Practical implementation

#### 1. Two-layer inversion strategy

Obviously, a basic GN method could not fulfill the requirement of problem (3.51) even the dimensionality was dramatically reduced by SA or SAA. In order to minimize the modified misfit function with sparsity promotion based on online ODL, one need to trace the optimal trade-off between the L2-norm residual and L1-norm solution  $\boldsymbol{\alpha}$  simultaneously. We will follow a limited-memory projected quasi-Newton (l-PQN) of two-layer inversion strategy proposed by Schmidt et al. (2009) and Zhu and Liu (2015).

#### a. Outer layer problem

The outer layer is aiming to solve a global minimization problem using the gradients and approximated Hessian, following Equation (3.25)

$$\nabla_{\delta\boldsymbol{\omega}} J^{(k)} \triangleq \frac{\partial J(\delta\boldsymbol{\omega})}{\partial \delta\boldsymbol{\omega}(\mathbf{x})} = - \sum_{s=1}^{N_s} \mathcal{F}_{ek}^{\dagger(k)} (T_{residual}^{(k)} - \delta T^{(k)}). \quad (3.52)$$

Imposing sparse promoting and stochastic process, the above equation reformulates to

$$\nabla_{\boldsymbol{\alpha}} J^{(k)} \triangleq \frac{\partial J(\delta\boldsymbol{\omega})}{\partial \delta\boldsymbol{\omega}(\boldsymbol{\alpha})} = - [\mathcal{D}^{(k)}]^\dagger \{ [\mathcal{F}_{ek}^{W(k)}]^\dagger (T_{residual}^{W(k)} - \mathcal{F}_{ek}^{W(k)}(\mathcal{D}^{(k)}(\boldsymbol{\alpha}))) \}. \quad (3.53)$$

The gradient and approximated Hessian matrix could be used for a limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) algorithm (Nocedal, 1980) to construct a

sequence of subproblems. The global problem in our layer is basically a GN problem that a 1-BFGS algorithm allows to derive a constrained quadratic approximation around the evaluation points.

### b. Inner layer problem

The inner layer is aiming to constrained sparse coefficients  $\alpha$  within the range of sparsity level  $\tau_k$  for each subproblem, to guarantee the optimal descent direction for solving the outer layer problem. We need to make sure that the solution  $\alpha$  is optimally constrained within an L1-norm ball with radius  $\tau_k$ , solving for:

$$\mathcal{P}_\tau(\alpha) = \underset{\beta}{\operatorname{argmin}} \|\alpha - \beta\|_2^2 \quad \text{s.t.} \quad \|\beta\|_1 = \tau, \quad (3.54)$$

where  $\mathcal{P}_\tau$  is the projection operator on to the L1-norm ball with radius and  $\beta$  is qualified sparse parameter after projection. The following box shows a simple procedure of solving for Equation (3.54).

**Input:**  $\alpha, \tau$

Sorting  $\alpha$  in descending order:  $\alpha_1 \geq \alpha_2 \geq \alpha_3 \cdots \geq \alpha_N$ ;

Find  $\gamma = \max \left\{ j: \alpha_j - \frac{1}{j} \left( \sum_{n=1}^j \alpha_n - \tau \right) > 0 \right\}$ ;

Define  $\theta = \frac{1}{\gamma} \left( \sum_{i=1}^{\gamma} \alpha_i - \tau \right)$ ;

**Box 3.2:** Projection of  $\alpha$  onto L1 norm ball

---

## 2. Choice of sparsity level $\tau$

The solution  $\boldsymbol{\alpha}$  of the LASSO problem (3.51) relies on our choice of the sparsity level  $\tau_k$  for each modified GN subproblem. Theoretically, any bounded positive sequence  $\{\tau_k\}$  could guarantee a descent direction for minimizing a reduced misfit function  $\tilde{J}^{(k)}$  at  $\delta\boldsymbol{\omega}^{(k)}$  (Hermann and Li et al., 2015). In practice, a systematic choice of  $\tau_k$  is necessary for optimally solving each subproblem. We could use a value function for the k-th subproblem, denoting as

$$V^{(k)}(\tau^{(k)}) = \frac{1}{2} \sum_{s=1}^{N_s} |T_{residual}^{W(k)}(\mathbf{x}) - \mathcal{F}_{ek}^{W(k)}(\mathcal{D}^{(k)}(\boldsymbol{\alpha}))|^2 \text{ s.t. } \|\boldsymbol{\alpha}\|_1 \leq \tau^{(k)}, \quad (3.55)$$

to evaluate  $\tau_k$  selection. Such type of value function mixing both L1 and L2-norm have been thoroughly studied in van den Berg and Friedlander (2008) by a ‘‘Pareto trade-off curve’’. The  $V^{(k)}(\tau^{(k)})$  could set to be approximately equal to  $V^{(k)}(0)$  as  $\tau^{(k)}$  is a very small positive value close to zero. By linear approximation of  $V^{(k)}(\tau^{(k)})$  at  $\tau^{(k)} = 0$ , we have

$$\tau^{(k)} \approx \frac{V^{(k)}(0) - V^{(k)}(\tau^{(k)})}{V^{(k)'}(0)} \approx \frac{V^{(k)}(0)}{V^{(k)'}(0)}. \quad (3.56)$$

Note that taking  $\tau^{(k)} = 0$  will force  $\delta\boldsymbol{\omega} = \mathcal{D}^{(k)}(\boldsymbol{\alpha}) = 0$ , which means that  $V^{(k)}(0) = \|T_{residual}^{W(k)}(\mathbf{x})\|_2^2$ . As suggested by van den Berg and Friedlander (2008) in Theorem 2.1, we could use closed expression form of  $V^{(k)'}(0)$  to write (3.56) as

$$\tau^{(k)} \approx \frac{\|T_{residual}^{W(k)}\|_2^2}{\|\mathcal{D}^{(k)\dagger}[\mathcal{F}_{ek}^{W(k)}]^\dagger T_{residual}^{W(k)}\|_\infty}, \quad (3.57)$$

taking adjoint operator from (3.26) and applied it on time residual from (3.57), we could obtain

---


$$\tau^{(k)} \approx \frac{\|T_{residual}^{(k)}\|_2^2}{\|D^{(k)\dagger} \nabla_{\delta\omega} J^{(k-1)}\|_{\infty}}. \quad (3.58)$$

### 3.7 Numerical test and real data application

#### 3.7.1 Synthetic data test

In the following sections, for simplicity, two stochastic process (SAA and SA) implemented with FAT tomography based on adjoint-stated method will be referred to SAA method and SA method, respectively. Furthermore, these two methods adopted with our proposed gradient optimization method will named as SAA+ and SA+, respectively.

##### 1. Two-layer model test

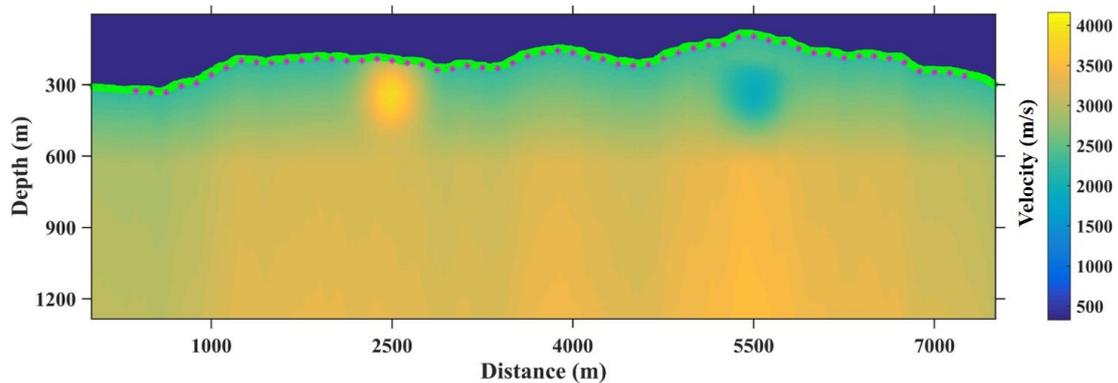
Apparently, either SAA method or SA method cannot significantly reduce computation cost, especially when batch size  $K$  is large. However, for a too small  $K$ , the stability and accuracy of inversion cannot be guaranteed. Therefore, in this case, our gradient optimization scheme is implemented with them to reduce time cost (small  $K$ ) while maintaining the inversion stability and accuracy of retrieved result.

This numerical test is divided into two stages. At first stage, we mainly focus on evaluating the applicability of SAA+ and SA+ for tomographic inversion, from which the optimal combinations of stochastic process and gradient optimization will be selected. At second stage, the performance of this optimal combination will be further evaluated regarding to both of time cost and robustness. It could provide important guiding significance for practice.

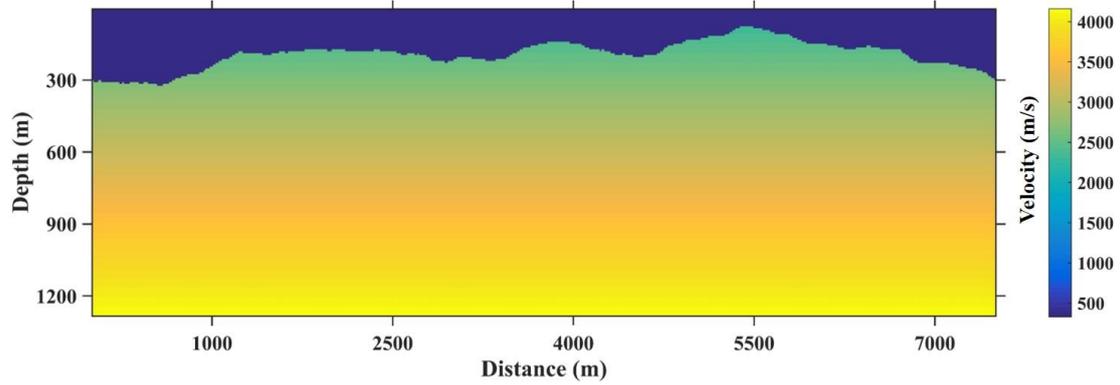
We designed a velocity model with topographic surface as shown in Figure 3-6 a. The

---

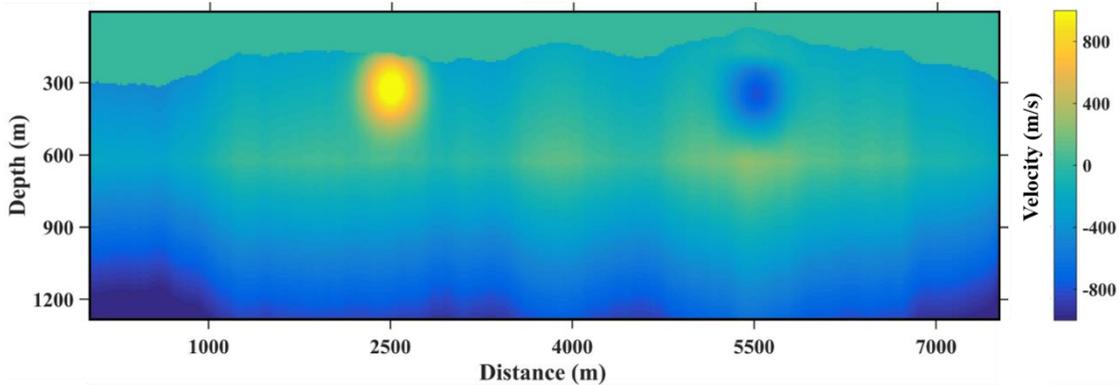
velocity above surface is equal to air velocity (330 m/s). This model is 7500 m in horizontal distance and 1280 m in deep, in which the grid cell size is  $12.5 \times 8$  m. The seismic survey in this model is following a fixed-spread acquisition system. The total receiver number is 600 in 12.5 m interval and 8 m depth spreading along the whole topographic surface. The total shot number is 100 in 62.5 m spacing and deployed in 8 m depth, as indicated by magenta stars in Figure 3-6a. The synthetic traveltime data is generated using FSW (Quan, 2004) method. The exact same forward modeling method was also used in the four tomographic inversion methods. For the tomography, we use a linearly increased velocity with depth as initial model as shown in Figure 3-6b. As seen in Figure 3-6a, the true velocity model contains two elliptical shaped velocity anomalous bodies (high and low for left and right, respectively) near surface, and the background velocity trend is quite smooth. The perturbation from difference between true and initial velocity model is shown in Figure 3-6c, which is the target to be recovered by tomographic inversion.



(a)



(b)



(c)

**Figure 3-6:** Numerical test model: (a) true velocity model, (b) initial velocity and (c) velocity perturbation.

### Applicability test of SA+ method

The performance of SAA, SA, SAA+ and SA+ is evaluated by this synthetic model test. In this test, the number of iterations for all methods are 30 to ensure that their inversions can achieve convergence, so as to obtain the optimal inverted results. Additionally, the K value for all methods is set to be 50, meaning that each inversion iteration will have 50 intermediate solutions. At first, we will use 10 different decimation percentages to conduct the test. Since all four inversion methods are based on adjoint-state technique, the data decimation percentage is only determined by the number of used shots out of the full dataset. For example, when the

---

decimation percentage is 10%, the number of used shots is 100 X 10%, 10 shots, while the corresponding number of receivers is 600 for all different decimation percentages. Especially, our interest is the inversion performance on small data decimation percentage (< 15%), so we conducted four tests with decimation percentage of 2%, 5%, 8% and 15%. Similarly, the other six tests are also performed on decimation percentage of 25%, 35%, 55%, 65%, 75% and 90%. For inversion results, we calculate their root mean square error (RMSE) for each test by equation below and show them in Table 3-1

$$RMSE(\mathbf{m})_P = \sqrt{\frac{1}{M} \|\mathbf{m}_P - \mathbf{m}^{std}\|_2^2}.$$

Where P represents four different methods: SAA, SA, SAA+ and SA+. Correspondingly, the RMSE of the inversion results derived by these four methods are  $RMSE(\mathbf{m})_{SA}$ ,  $RMSE(\mathbf{m})_{SAA}$ ,  $RMSE(\mathbf{m})_{SA+}$  and  $RMSE(\mathbf{m})_{SAA+}$ .  $M$  is the total number of model cells. Vector  $\mathbf{m}^{std}$  is the inversion results using full dataset, which is the standard FAT tomographic inversion based on adjoint state method. Since then, the RMSE value represents the error of results retrieved by four methods relative to those of standard method.

From Table 3-1, we observe that in case of small amount of data-decimation (e.g., 2%, 5% and 8%), the SA method gives a better accuracy than SAA as indicated by the RMSE value, while in the case of a relatively large amount of data-decimation ( $\geq 15\%$ ), the accuracies of their results are similar. After implemented with gradient optimization method, the accuracies of both SAA+ and SA+ are improved. Similarly, the accuracy of SA+ is still better than that of SAA+. Such improvement is more significant in the case of a small amount of data-decimation

---

percentage. This result proves that our gradient optimization method is helpful for improving the accuracy of inversion results retrieved by either SA or SAA. In addition, the combined method SA+ approach has more advantages than SAA+.

For the stability of the inversion process, we obtained the RMSE and its standard deviation to form the error bar curve of the 80 (as  $K = 80$ ) intermediate solutions during the inversion for all four methods, as shown in Figure 3-8a. Similar to the results presented in Table 3-1, the intermediate solutions of SA is better than that of SAA, no matter how the accuracy or the stability of the 80 intermediate solutions is, especially when data-decimation percentage is small. The implementation of our proposed method improves intermediate solutions reconstructed by SA or SAA in terms of both stability and accuracy. Among them, the improvement of SA+ is the most significant, which further proves that the combination of SA and proposed gradient optimization is the optimal integrated method.

Percentage of data	Rms error (m/s)			
	SAA	SA	SAA+	SA+
2%	143.55	115.61	128.76	97.86
5%	121.98	105.19	115.46	82.71
8%	98.38	77.24	85.77	58.34
15%	73.85	67.96	66.23	49.62
25%	54.34	56.34	53.21	32.74
35%	42.19	41.56	38.89	26.14
55%	33.80	34.23	32.89	21.68
65%	28.17	29.32	27.70	19.50
75%	21.74	22.70	21.26	18.47
90%	19.26	19.64	18.29	18.05

**Table 3-1:** The rms error for SAA, SA, SAA+ and SA+ method (clean data).

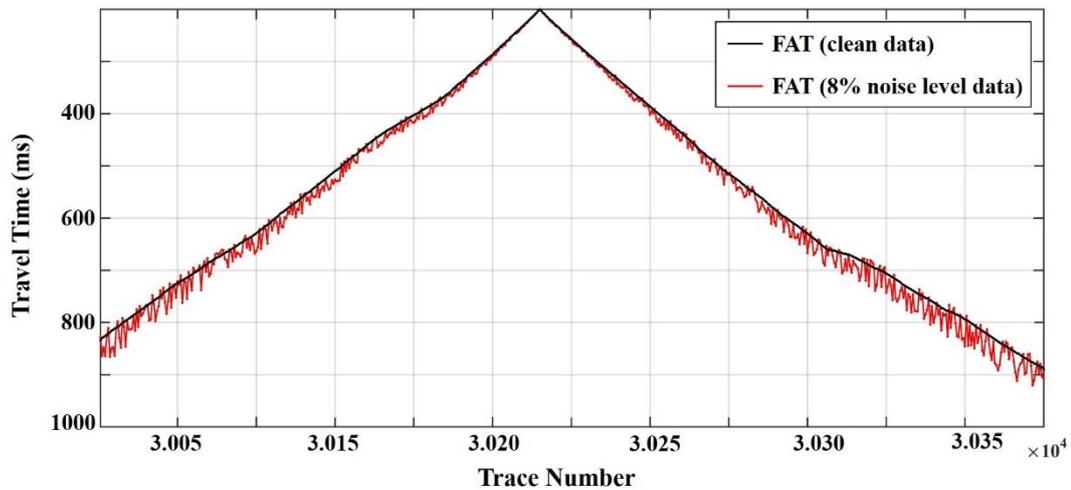
Percentage of data	Rms error (m/s)			
	SAA	SA	SAA+	SA+
2%	139.42	119.70	127.13	102.56
5%	124.08	107.31	111.29	91.24
8%	101.05	78.05	87.64	63.39
15%	71.53	71.38	64.93	52.38
25%	55.69	57.14	55.29	35.42
35%	43.07	43.61	40.17	28.74
55%	35.44	33.85	34.30	20.52
65%	28.14	30.24	28.55	21.17
75%	22.63	21.65	22.05	19.51
90%	19.18	20.02	17.31	18.65

**Table 3-2:** The RMSE for SAA, SA, SAA+ and SA+ method (8% noise)

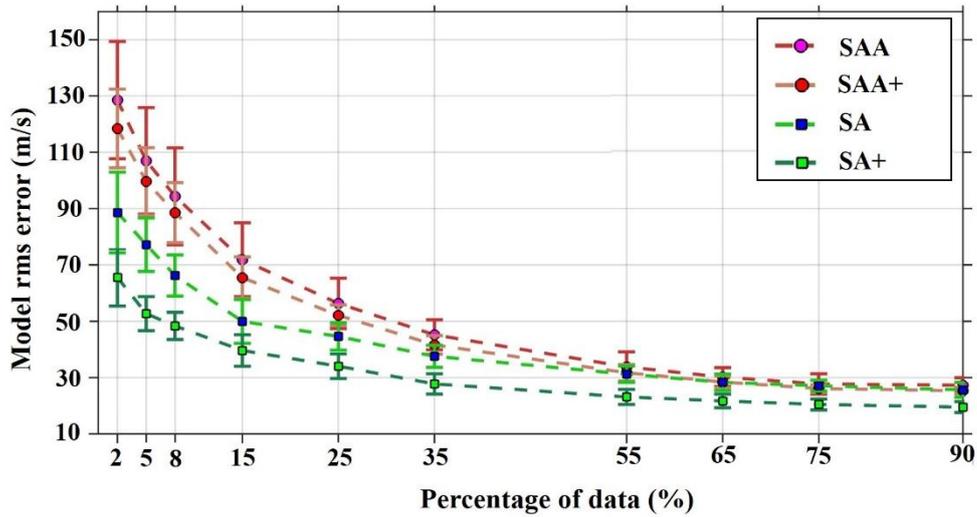
To further analyze the influence of noises in the application of the four methods, we add in 8% random noises in average to the true synthetic first arrival picks, as shown in red dot line in figure 3-7. In order to better simulate the actual situation of reality, the magnitude of the

---

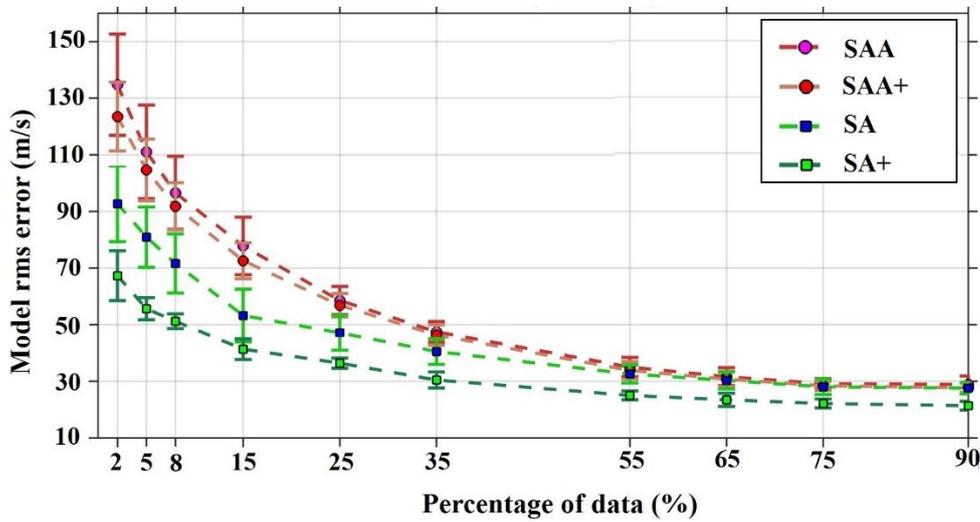
noise is set to be increasing with offset to mimic the offset-dependent signal-to-noise ratio or manual picking errors, which is often observed at large offset in practice. The exact same test on clean data was also performed on 8% noisy data, and its result is presented in Table 3-2. The results in clean (Table 3-1) and noisy (Table 3-2) data are basically consistent, indicating that such a noise level has no significant influence on the accuracy of all four methods. Besides, comparing to Figure 3-8a, the error bar curve from noisy data test in Figure 3-8b also does not change much. It implies that SA+ has best stability among all four methods. The above series of tests proves that SA+ is the best approach in terms of accuracy and stability. To avoid lengthy discussion, we will mainly focus our evaluation on the performance of SA and SA+ in the following tests.



**Figure 3-7:** First arrival picks: 8% noise level data (red line) and clean data (black line).



(a)

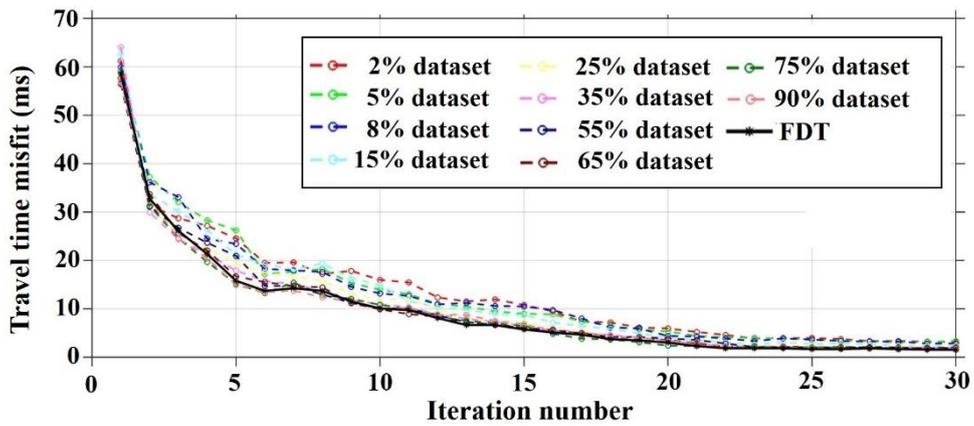


(b)

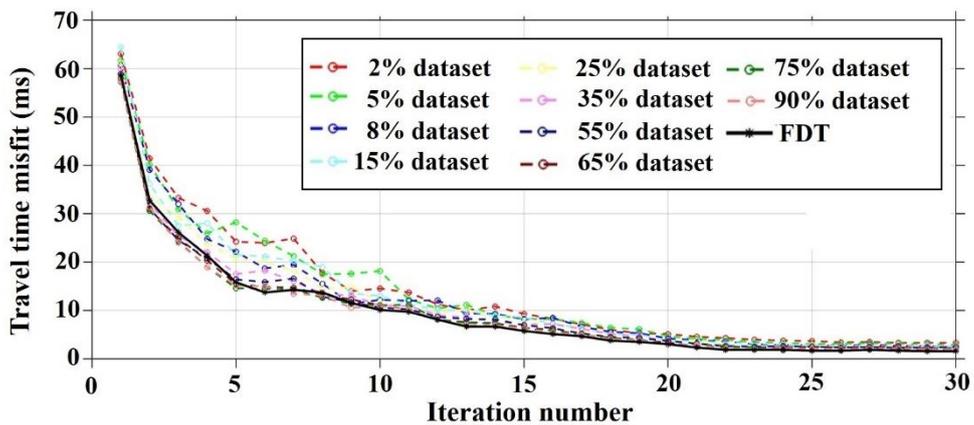
**Figure 3-8:** Error bar curves of (a) clean data and (b) noisy data (8% noise level).

We also show the inversion convergence history curve of SAA, SA, SAA+ and SA+ with either clean and noisy data, adopting with different data-decimation percentages as input. From Figure 3-9, we could see that their traveltimes misfits stabilize at about 25 iterations, meaning that such comparative test was carried out on the premise of obtaining the best solution for each method. Comparing to standard method, their convergence rate is reduced to some extent due to randomness of data decimation. Among them, the SA+ appear to have better convergence

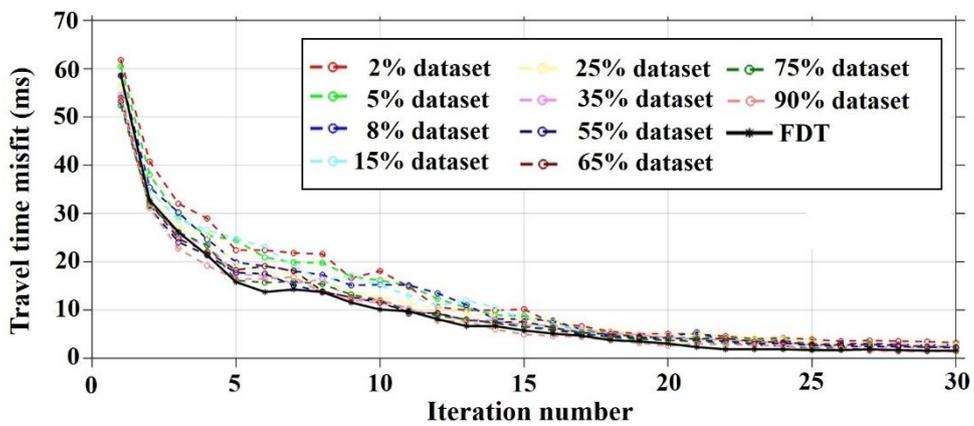
than SA in the early-to-middle stage of inversion. It is also one advantage of our proposed method in searching for the gradient for descent direction in sparse space, by which the convergence is stabilized.

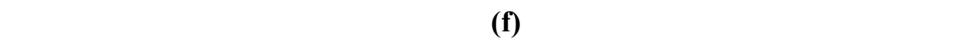
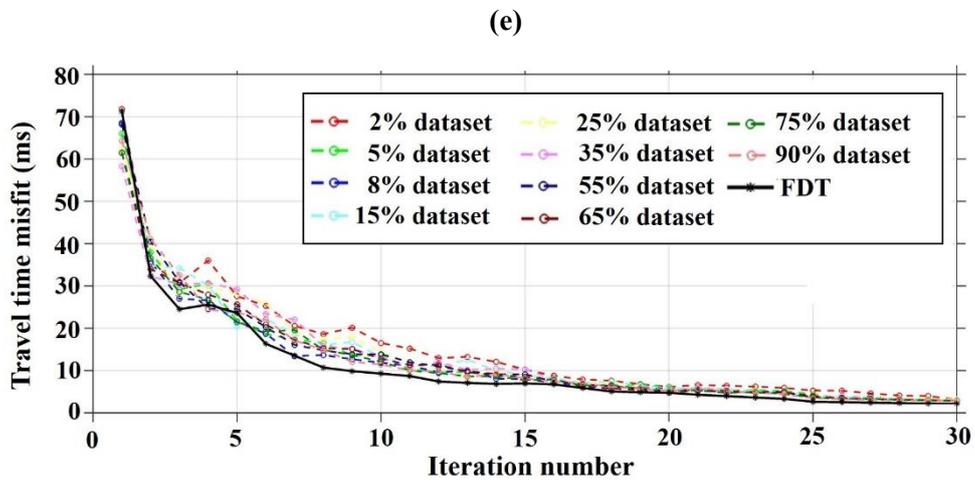
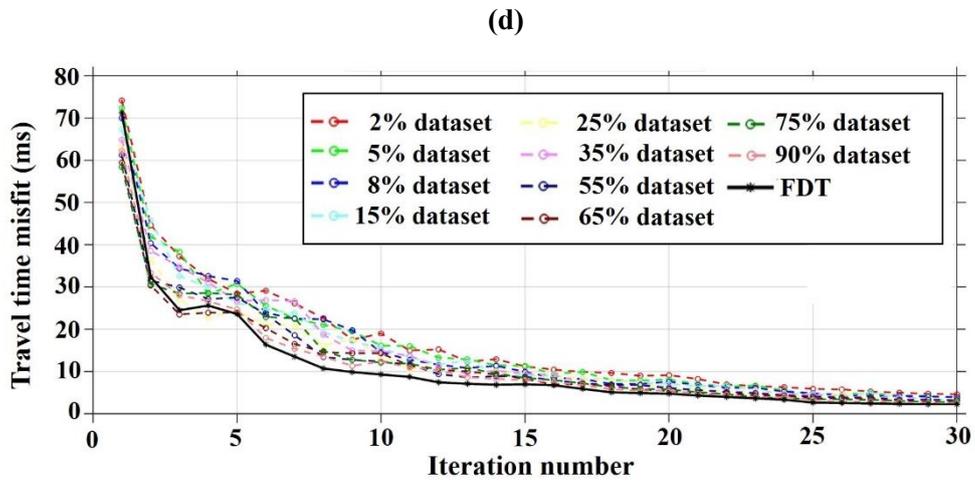
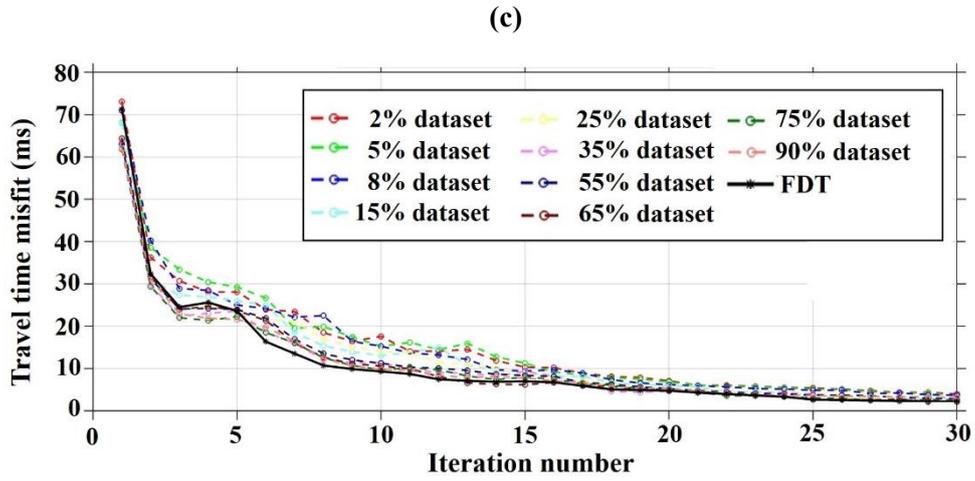


(a)



(b)



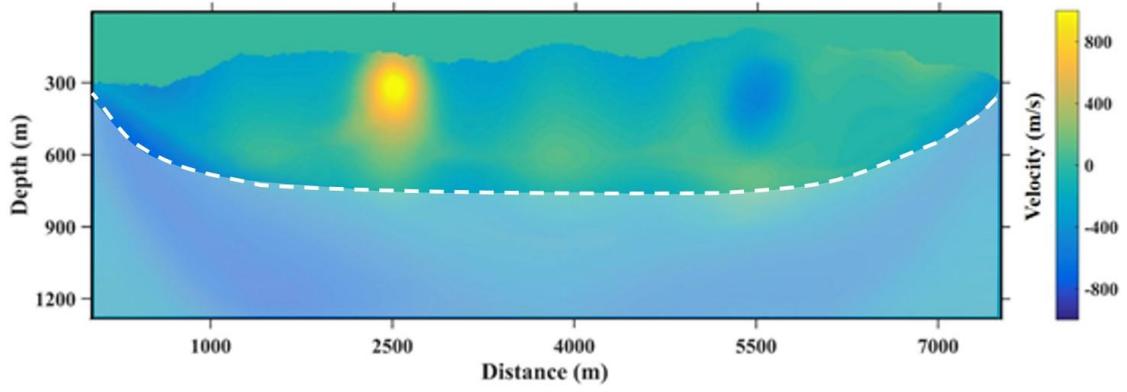


**Figure 3-9:** Convergence history of (a) SAA, (b) SA and (c) SA+ with clean data, Same of (d) SAA, (e) SA and (f) SA+ with noisy data (8% noise level).

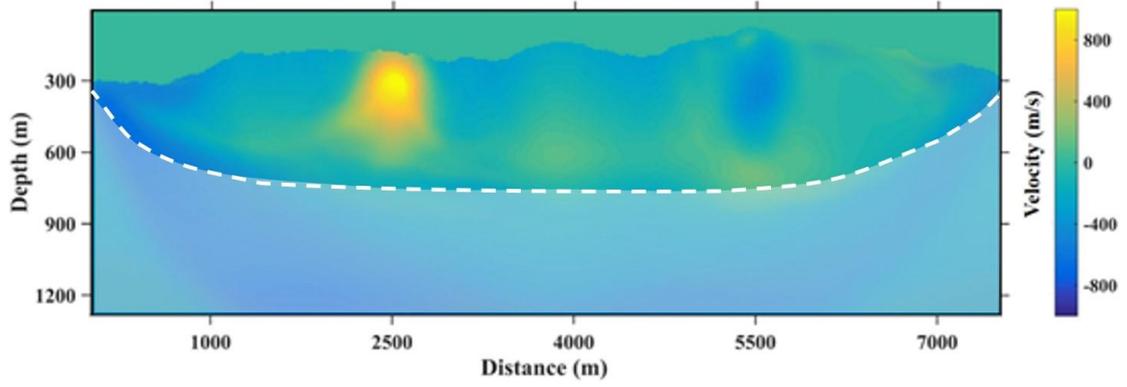
We further investigate the velocity models inverted by SA and SA+ method in both clean

---

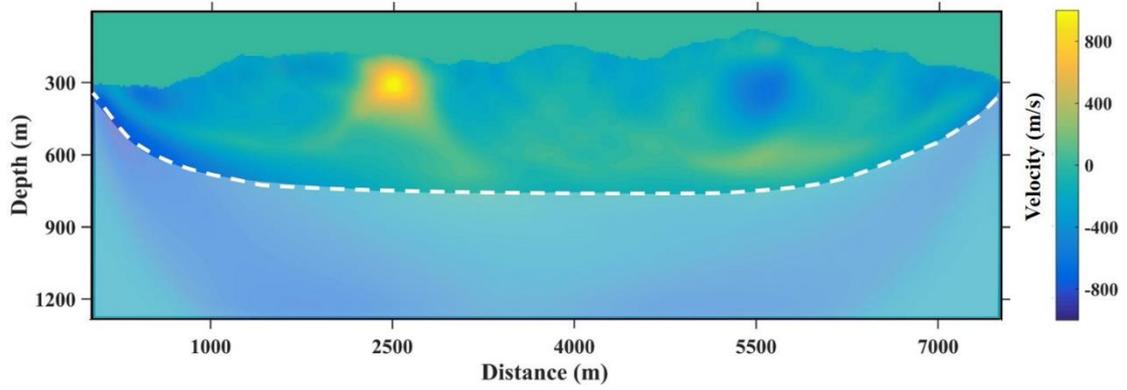
and noisy data cases, as displayed in Figure 3-9. Due to the limitation of ray coverage, we will mainly focus on observing trusted recovered region above white dash line, as indicated in Figure 3-9. The result of full dataset tomography (FDT) using standard method is shown in Figure 3-10a and 5b for the cases without and with noise in first arrival traveltime picks, respectively. Figure 3-9c and 3-9d displays SA results using 5% dataset for clean data test and 10% dataset for noisy data test. Figure 3-9e and 3-9f show SA+ results using exact same data-decimation percentage as SA for clean and noisy data test. By comparison, the two velocity anomalies are well depicted in both results of SA and SA+, and the main part of recovered velocity model looks identical to FDT. However, compared to SA+ in details, the SA results contain artificial residues similar to ray trajectories smearing in varying degrees from both clean and noisy data test. Such “smearing effect” is mainly caused by inversion gradient construction using a small data-decimation percentage as input. Apparently, to improve the performance of SA, one can further increase the K value to earn more average effect for final sampling average calculation of inverted model; or increase the smoothness of the gradient during inversion. However, these two approaches will bring negative effect on either computational efficiency or preserving accuracy of inverted velocity model. On contrary to SA, the SA+ can obtain smoother inversion results without gradient smoothing, and its result has more accurate depiction on both of two velocity anomalies. Such performance improvement comes from our gradient optimization method.



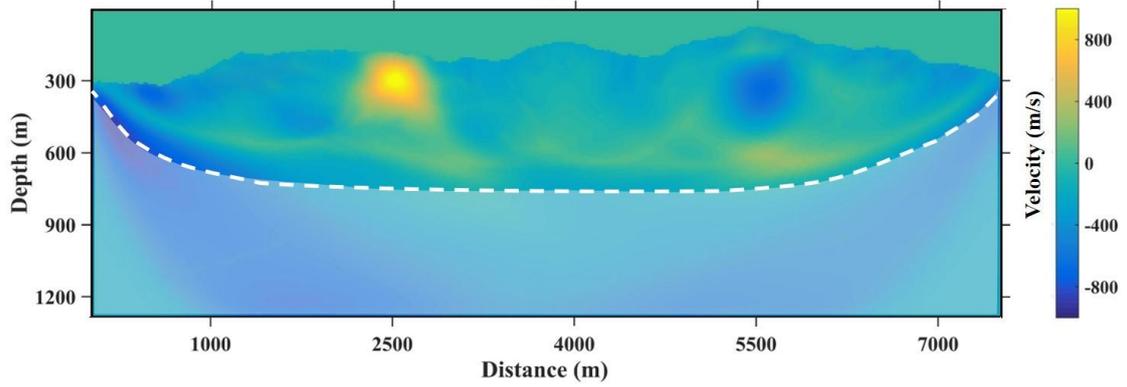
(a)



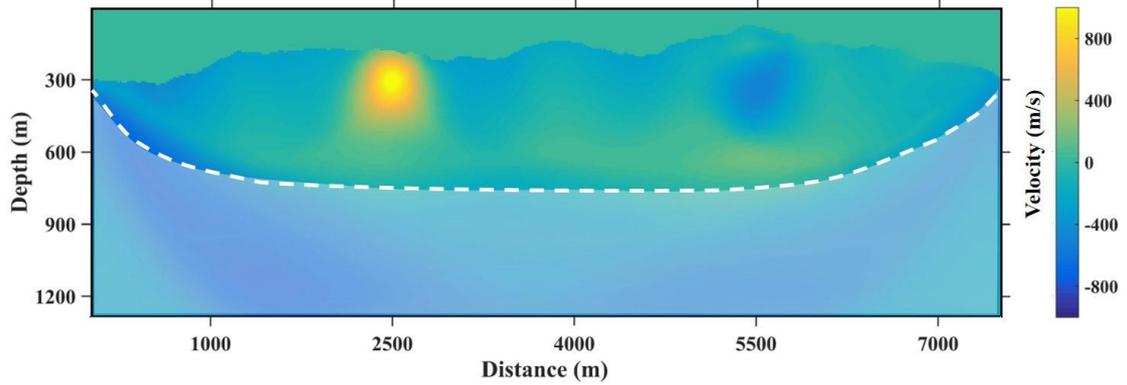
(b)



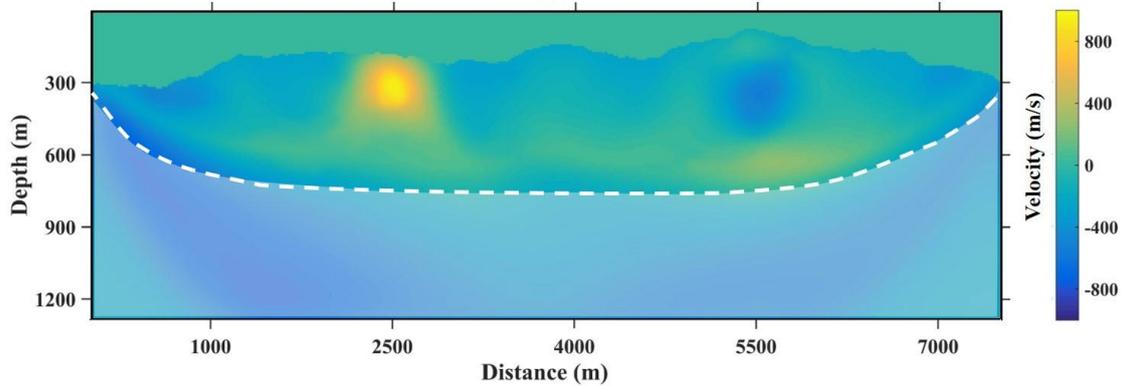
(c)



(d)



(e)



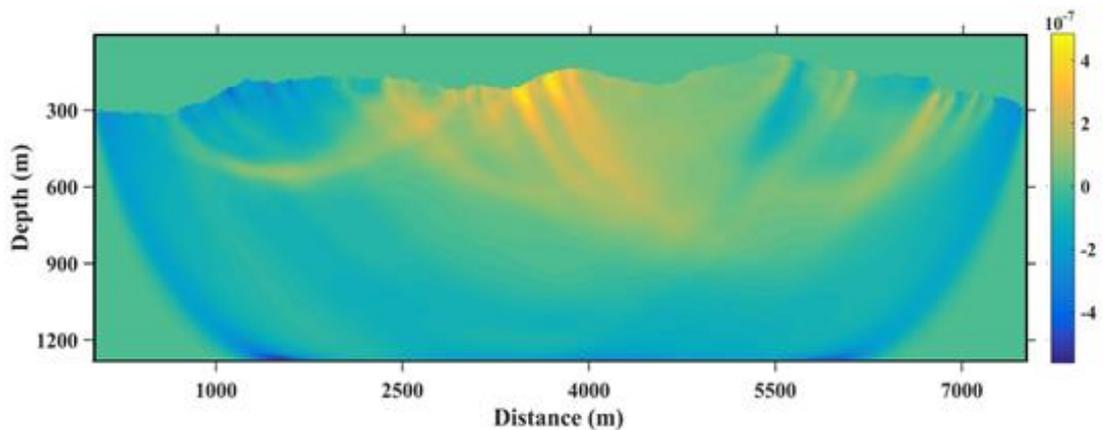
(f)

**Figure 3-10::** Inverted velocity models. The standard method results of (a) clean and (b) 8% noisy data. The SA method results of (c) clean and (d) 8% noisy data. The SA+ method results of (e) clean and (f) 8% noisy data. The region above white dash line is trusted recovered region.

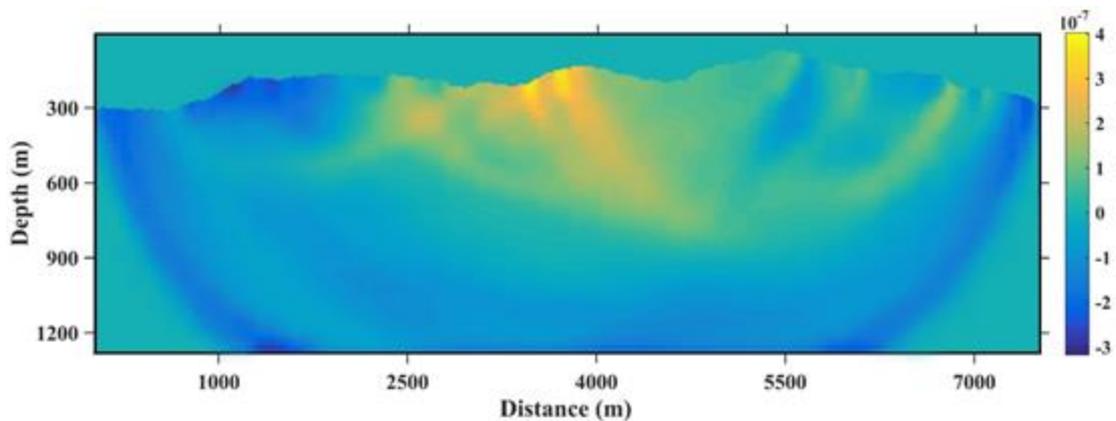
From Figure 3-10, it can be observed how our proposed gradient optimization improves the gradient from any one iteration of the SA inversion. Figure 3-10a displays the original gradient during SA, in which the interferences from ray trajectories residues are obvious,

---

especially in near surface areas. Figure 3-10b is the gradient after optimization, and Figure 3-10c is the difference. One can see that the interferences existed in the original gradient is significantly suppressed through the optimization, hence the smoothness of gradient is improved. Besides, there is a large amount of artificial interference signals left in the difference, as shown in Figure 3-10c), meaning that such interference in the original gradient has been suppressed by optimization. Though gradient optimization is quite similar to the general smoothing process in ray-based tomographic inversion, it can better preserve salient information contained in gradient while removing noise raised from random subsampling of SA.

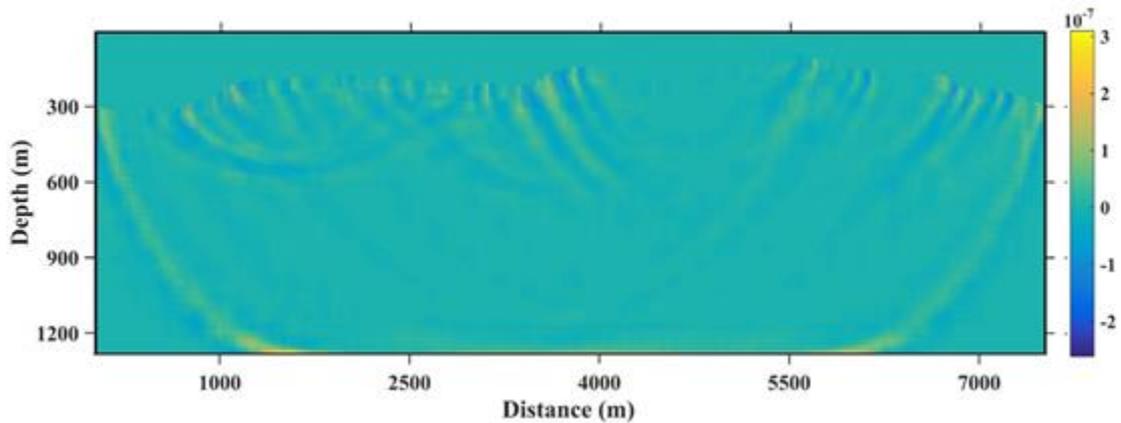


(a)



---

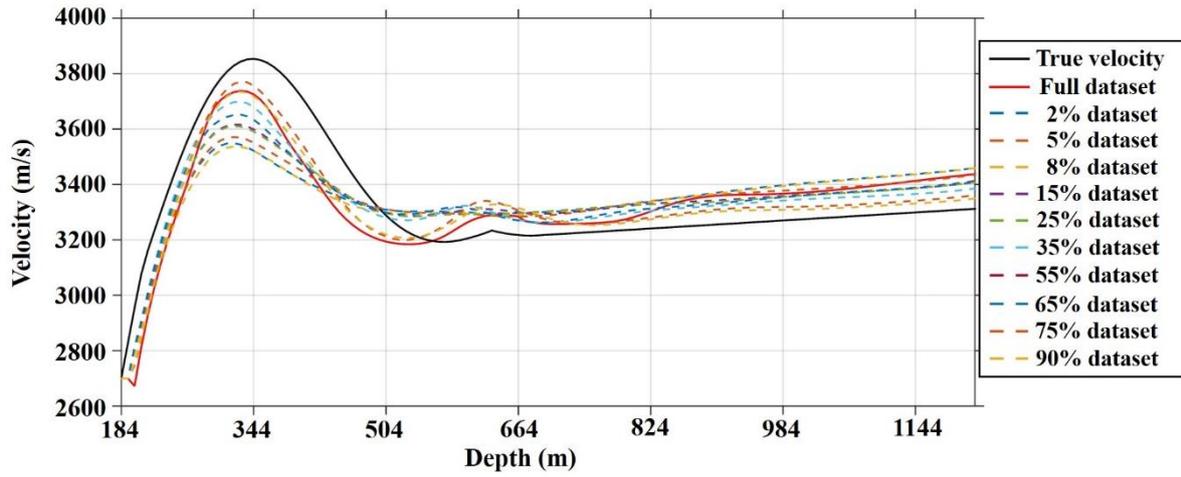
(b)



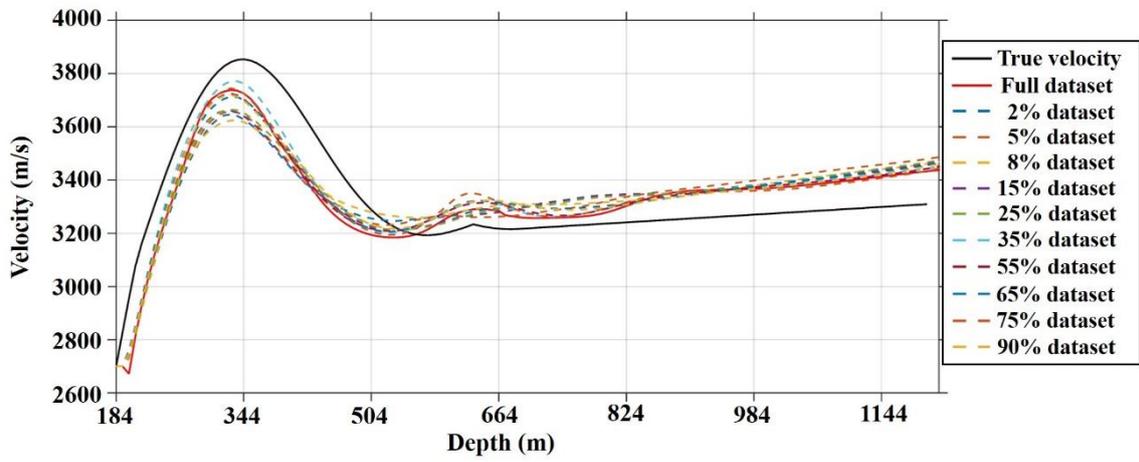
(c)

**Figure 3-11:** Comparison of gradient (a) before and (b) after optimization by our proposed method, and their (c) difference.

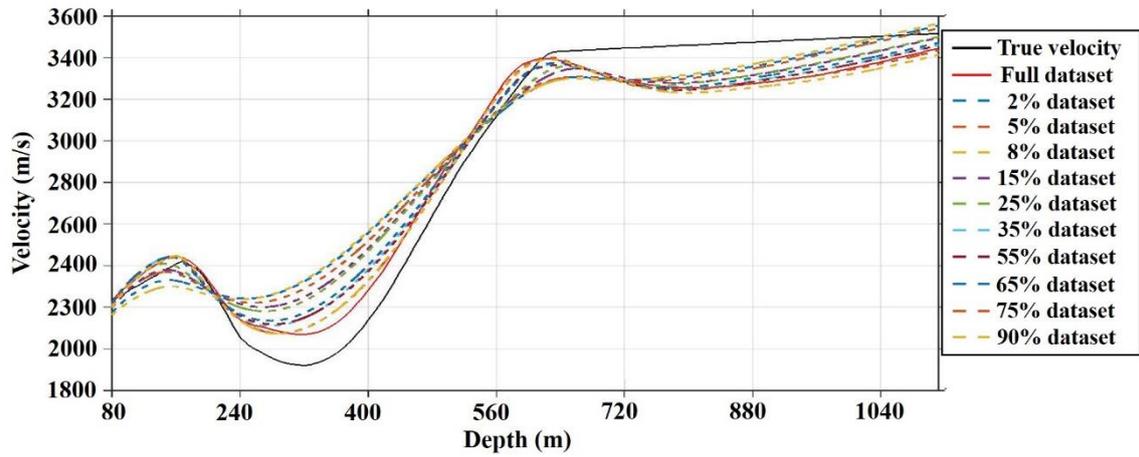
We further evaluate the performance of SA and SA+ by observing the vertical velocity functions at the center of the two velocity anomalous bodies. The exact horizontal positions are at 2500 m for high-velocity anomalous body, and 5500 m for low-velocity anomalous body. Here we only use noisy data with 8% noise level for different data-decimation percentage in this test. Figure 3-12a and c show the velocity functions at the high velocity anomaly for SA and SA+, respectively. Similarly, Figure 3-12d and e represents for low-velocity anomalies. We can see that the velocity function of SA+ is better focused around result of standard method full dataset than that of SA, especially for small data-decimation percentage. Again, this shows that the SA+ is superior to SA in terms of inversion accuracy.



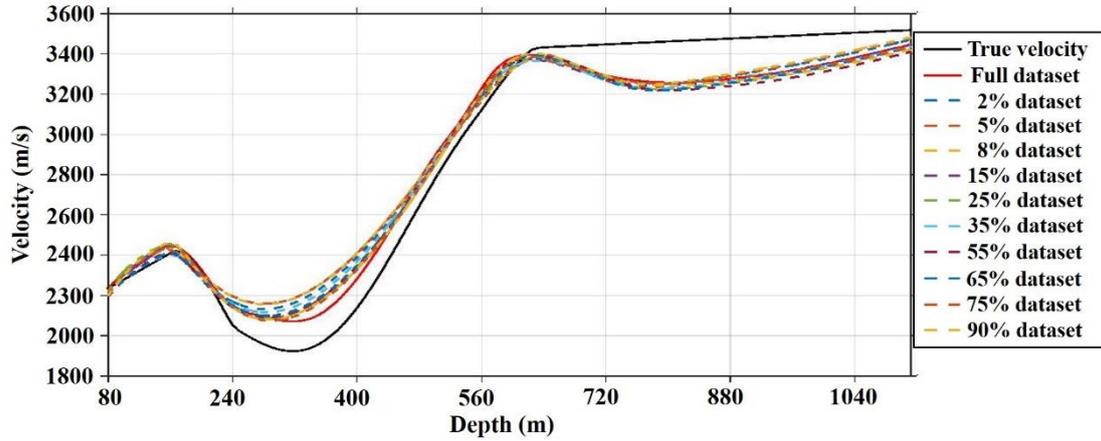
(a)



(b)



(c)



(d)

**Figure 3-12:** The velocity functions at the center of high velocity anomaly inverted by (a) SA and (b) SA+. The velocity functions at the center of low velocity anomaly inverted by (c) SA and (d) SA+.

From the results of series of tests above, the combination of SA and our proposed gradient optimization method shows great advantages in preserving inversion accuracy and stability, hence the performance of SA+ is quite promising. In the next section, we will focus on investigating improvement of computational efficiency of the SA+ method.

### Computation efficiency test of SA+ method

In this part, we will analyze the relationship between the accuracy and stability of SA+ method and other factors on the premise of improving the computational efficiency. These main factors mainly include two aspects: the data-decimation percentage and noise level of dataset. The same synthetic velocity model as first section will be used in this test.

The improvement of computational efficiency mainly refers to saving the time cost of inversion. In previous test, K value is set to 50 for all methods. Obviously, SAA, SA, SAA+ and SA using this K value cannot achieve the objectives of reducing computational time cost. For example, in ideal cases, the time cost of either SAA or SA method using only 2% data is

---

close to that of standard method using full dataset, which is 2.67 min. If the convergence rate reduction (as shown in Figure 3-9) is further considered, the time cost will certainly exceed the standard one. Therefore, for SA+ method, an appropriate K value and data-decimation percentage setting is important.

It should be noticed that the online dictionary learning involved in gradient optimization for SA+ will slightly increase the time cost for SA+ comparing to SA. However, only matrix multiplications and SVD operations are used and its time cost proportion will decrease rapidly as the inversion proceeds due to accumulation of past information in the updated dictionary. According to our practical experience, the ceiling of all the computational cost regarding the dictionary learning part is normally less than 15% of total time cost depending on data-decimation percentage used in SA+.

In this test, we obtained the RMSE of SA+ and SA using different data-decimation percentage as input under the time cost about 70%, 50%, 40%, 20% and 10% compared with the full dataset tomography (FDT), and list them in Table 3-3 and 3-4 for clean and noisy data test respectively. The K value meets the condition of  $K > 1$  for all cases, that is, the data is randomly selected at least twice to keep the basic idea of Stochastic process.

Besides, we assume that result is acceptable when the RMSE of result derived by either SA+ or SA is less than 100m/s, by which the subsequent data processing will not be significantly influenced. According to this standard, the green shaded blocks in Tables 3-3 and 3-4 are qualified results, while the red ones are unqualified. We can see that solutions retrieved from both of SA+ and SA cannot meet the qualification requirement ( $< 100$  m/s), when the

---

data-decimation percentage is as small as 2% in clean and noisy data test. For input of 5% dataset, SA failed at all time cost levels, but SA+ could achieve the optimal performance of using only about 40% of FDT computation time, which means that its time cost is 1.06 min. When the data-decimation percentage reaches 10%, both SA and SA + only take 20% of the FDT time cost within the RMSE of 100 m/s, which is about 0.53 min. However, the accuracy of SA+ result is much higher than that of SA in terms of RMSE.

In the case of higher data-decimation percentage ( $\geq 10\%$ ), results of both SA and SA+ are qualified, but the corresponding maximum time cost reducing is also more restricted (can only reduce as much as 20% of time cost). We performed the same test using 10% noisy data, and list its result in table 4. One can observe that the noise doesn't have much impact on the results derived by relatively high data-decimation percentage for both SA and SA+. However, in case of low data-decimation percentage such as 10%, the result of SA is closer to standard RMSE of 100 m/s at low time cost ratio, while the performance of SA+ is more reliable.

Time cost (compare to FDT)		~70%	~50%	~40%	~20%	~10%
Percentage of data		RMSE (m/s)				
SA+	2%	101.58	109.08	115.56	123.17	135.31
	5%	95.43	98.45	101.12	108.69	116.03
	10%	62.19	68.16	67.32	72.24	—
	15%	55.21	58.33	61.22	—	—
	20%	33.26	36.91	—	—	—
SA	2%	121.72	123.54	128.60	136.31	146.7
	5%	114.38	117.79	120.22	131.45	133.55
	10%	82.31	87.23	92.49	95.88	—
	15%	67.05	73.35	75.30	—	—
	20%	58.16	63.83	—	—	—

**Table 3-3:** The RMS error and time cost for SA with proposed method (clean data). The green shaded blocks represent that it is a qualified result, and RMSE is less than 100m/s. The red shaded blocks are unqualified result with RMSE more than 100m/s.

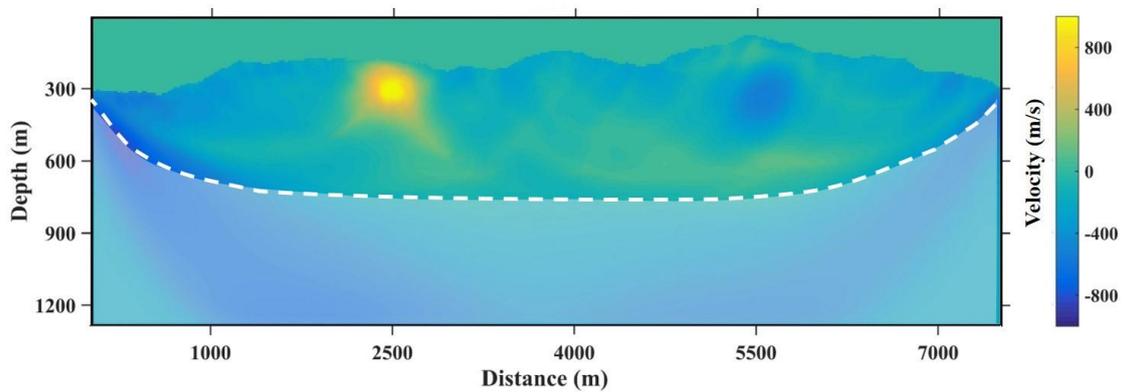
Time cost (compare to FDT)		~70%	~50%	~40%	~20%	~10%
Percentage of data		RMSE (m/s)				
SA+	2%	108.33	112.41	121.29	134.25	142.15
	5%	97.65	101.52	107.30	112.06	121.29
	10%	64.52	70.27	73.82	75.45	—
	15%	57.64	60.09	62.19	—	—
	20%	34.16	35.82	—	—	—
SA	2%	126.82	131.08	129.78	142.39	154.57
	5%	116.14	119.37	127.04	138.10	141.73
	10%	79.24	85.65	102.53	118.71	—
	15%	66.25	74.83	82.66	—	—
	20%	60.25	68.90	—	—	—

**Table 3-4:** The rms error and time cost for SA with proposed method (8% noise). The green and red shade block is same as table 3.

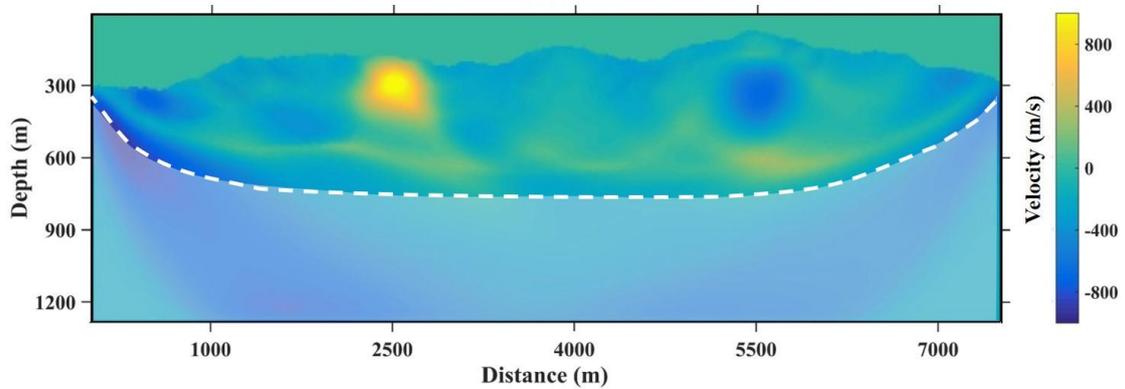
Now we look into some details of the exact velocity model retrieved by SA and SA+ in this test. Here the results are obtained from 7% and 10% data decimation as input at 20% of

---

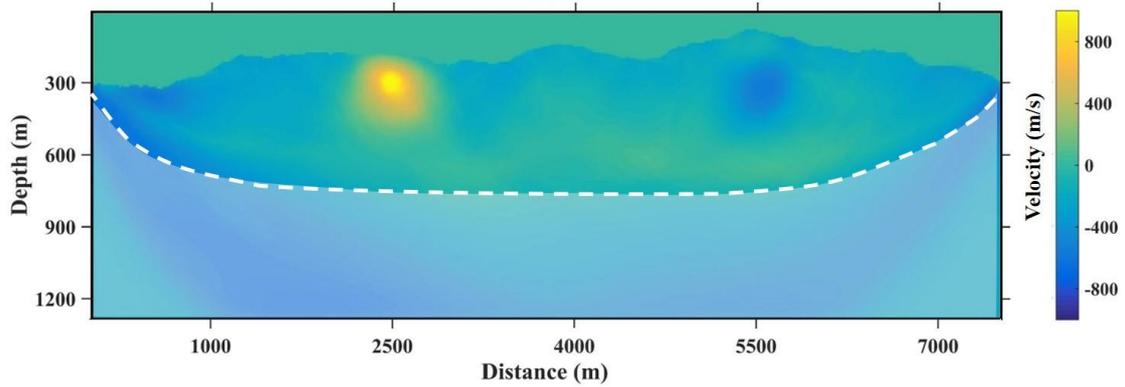
time cost of FDT for clean and 8% noisy data, respectively. Figure 3-13a and 7b display the clean data test results of SA and SA+ respectively, and Figure 3-13c and d are results from noisy data test. Similar to the observation from previous tests, the random sampling residues in SA result is suppressed in SA+ with more clean and smooth presentation. Comparing to results of FDT in Figure 3-13a, the SA+ obtains almost identical results, while the computational time cost is greatly reduced from 2.67 min to 0.53 min.



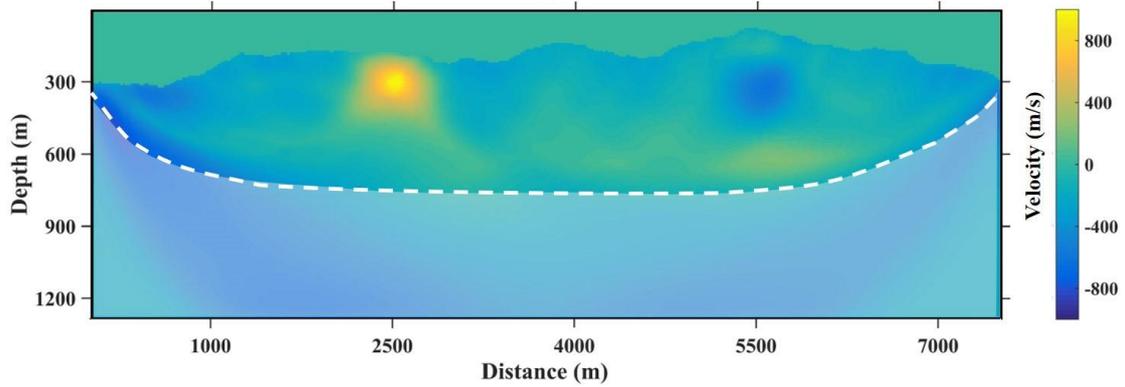
(a)



(b)



(c)



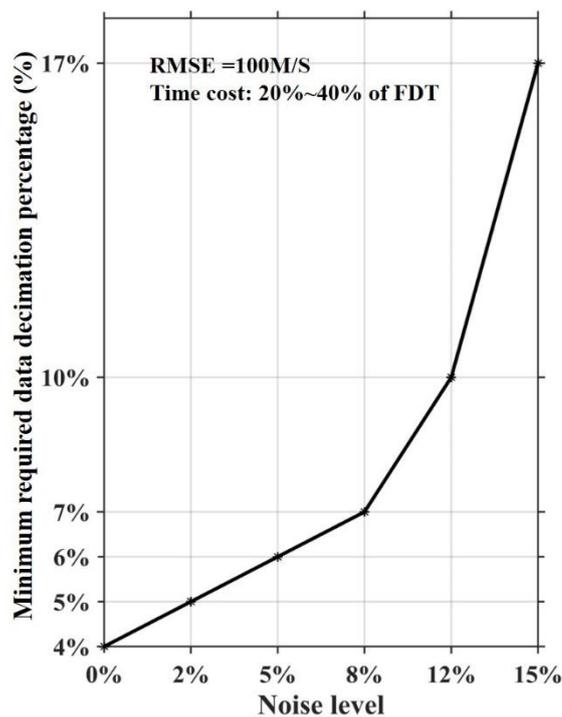
(d)

**Figure 3-13.** Inverted velocity models with 7% and 10% data decimation at 20%-time cost of standard method for clean and 10% noisy data test respectively. Clean data result: (a) SA and (c) SA+ method. 10% noisy data: (b) SA and (d) SA+ method. The region above white dash line is trusted recovered region.

To further understand the relationship between noise level and data-decimation percentage, we add 5%, 10%, 15%, 20% and 30% random noise into clean data to analyze the influence of different noise levels on data-decimation percentage for a fixed RMSE (100 m/s) in SA+ method. Each test is carried out on the premise of only using around 20% to 40% of FDT computation time cost, and corresponding data-decimation percentage is the minimum requirement under a certain noise level.

As shown in Figure 3-14, for the case of low noise level ( $< 8\%$ ), the minimum required data-decimation percentage for obtaining a qualified result is relatively stable, which can be

less than 10%. However, when the noise level is higher than 15%, in order to reach the qualified RMSE value, it is necessary to increase the data-decimation percentage. Especially, when the noise level is larger than 8%, the amount of decimation percentage increases dramatically (from 7% to 17%). This curve can be used as a guide for the application of SA + method in practice: 1.) when the data is of good quality and high SNR, we can only use around 8% data, meaning that we could use a more flexible K value to balance the trade-off between the goal of reducing time cost and preserving the image quality; 2.) when the data quality is poor, our primary goal is to ensure the accuracy of inversion results by introducing larger data-decimation percentage or increasing K value, and then consider the appropriately reduce the time cost.

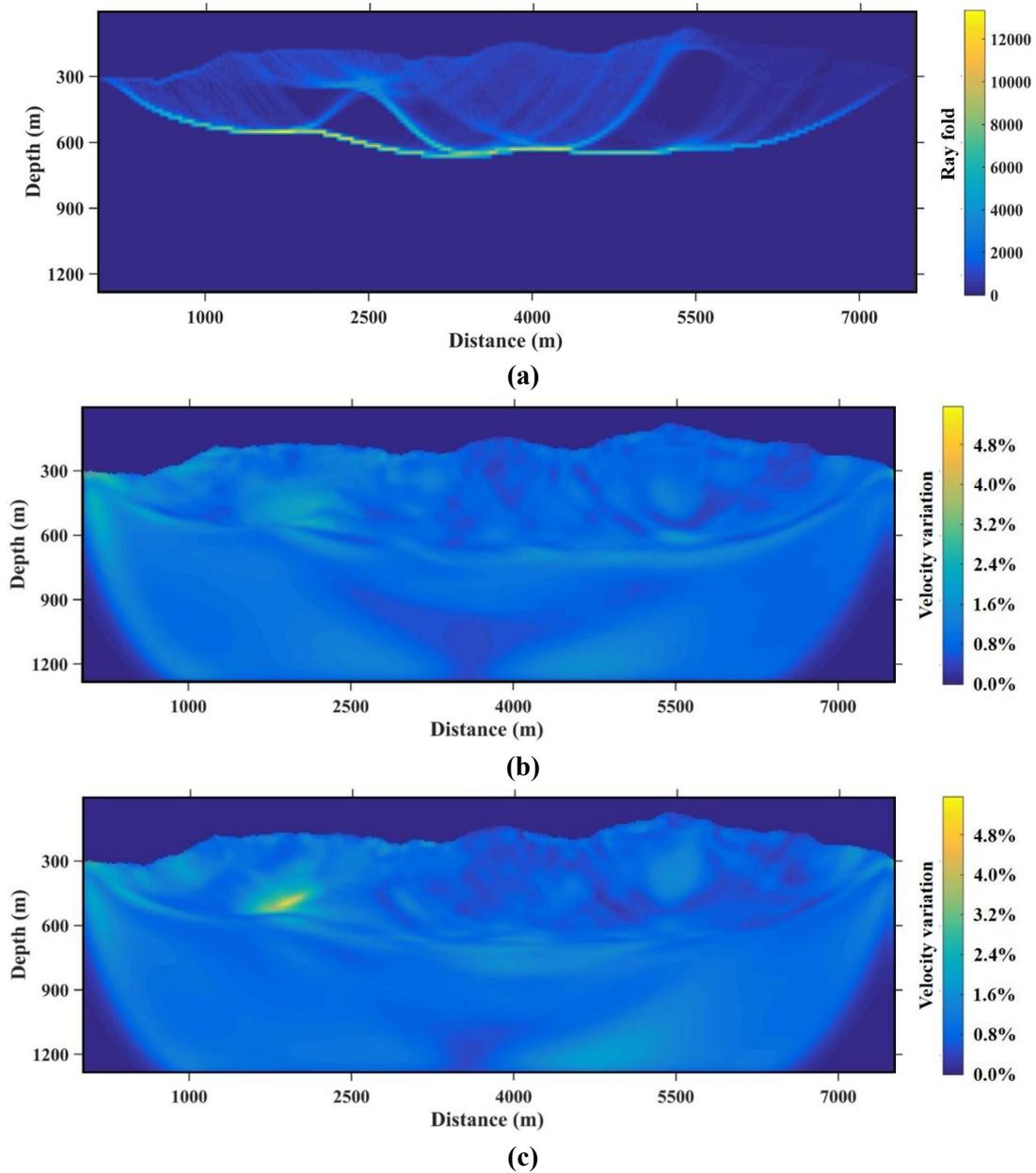


**Figure 3-14:** The relationship between the noise level and the data-decimation percentage for a fixed RMSE at 100 m/s. All time cost in this test is around 20% to 40% of FDT.

We also investigated the stability of SA+ using the recovered velocity at each cell in both clean

---

and 10% noisy data cases. This test is using a variation of the bootstrap method to estimate the standard deviation for each cell. The bootstrap was performed by repeating 80 times realizations of SA+ with different randomly selected data for each realization. Then we calculate the variance ratio of these 80 realization results in each cell  $i$ , which is expressed as  $\delta_i^m$ . We used 7% and 10% data-decimation percentages for clean data and 10% noisy data tests, and displayed the results in Figures 3-15b and c, respectively. The ray density corresponding to the velocity model of final iteration is shown in Figure 3-15a. We can observe that the large deviations are mainly distributed in areas with poor ray coverage, including the low-velocity and lower part of high-velocity anomalous body. Although the performance of inversion is slightly improved by adopting FSW method, the accuracy of this ray-based tomographic inversion is essentially restricted by ray density in the subsurface media. Moreover, for some regions with relatively sparse or mainly parallel ray paths, the velocity cells appeared to be some extent of instability, such as area below high-velocity anomalous body in Figure 3-15a. In practice, we need to have a basic understanding of subsurface geology, and then make corresponding adjustment for specific seismic data to ensure the accuracy and stability of the SA+ inversion. Overall, the noise does not have significantly influence on stability of recovered velocity model, which indicates that robustness of SA+.

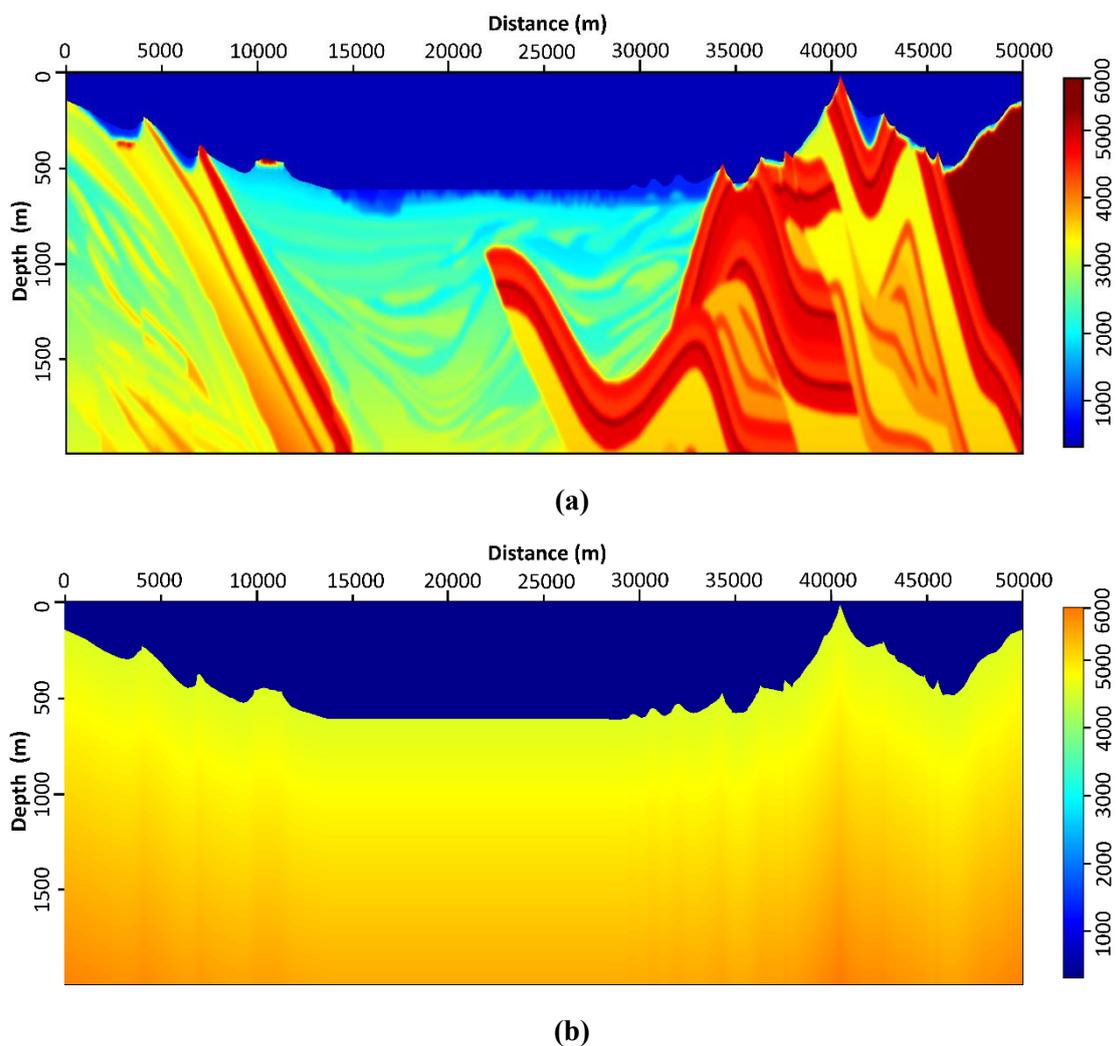


**Figure 3-15:** (a) Ray density distribution in velocity model of final iteration, the velocity variation of each cell in SA+ with (b) 7% of clean data and (c) 10% of noisy data.

---

## 2. Amoco 94' model test

To further test the performance of SA+ method in complex near-surface geological structure, we applied it on Amoco 94' synthetic model. The exact same model and survey settings are referred to section 2.3.2 in chapter 2. Figure 3-16a is the true velocity model, and Figure 3-16b is simple gradient initial model. This is a high density 2D synthetic data with 1998 sources and 2500 receivers, which is a large-scale dataset scenario favored by SA+ method.



**Figure 3-16:** (a) True velocity model and (b) linear increasing initial velocity model. The velocity

---

is increasing from 450m/s at surface to 5500 m/s at bottom with gradient of 0.5.

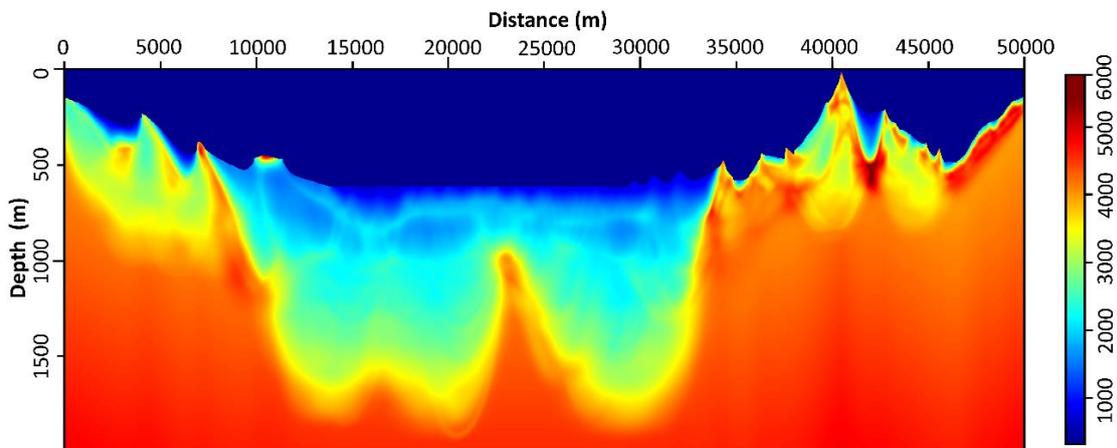
The standard tomography, referring to a FAT tomographic inversion based on adjoint state method adopted from Chapter 2, was conducted with full dataset and obtained the estimated velocity model as shown in Figure 3-17a. It will be used as a base criterion to evaluate image results of other methods presented in series of figures of Figure 3-17. From Figure 3-17b to c, we implemented only small fraction of full dataset (1998 sources) at 10% (200 sources) to conduct a series of tomographic inversion in different approaches.

As shown in Figure 3-17b, the SA method owns the poorest image quality among all other results. In the shallow velocity zone ranged from ~10000m to ~34000m, there are some significant smearing artefacts remains in the image caused by the sparse shot coverage. Particularly, in the right region of complex geological area, the estimated velocity model shows a strong non-convergence indication. The inverted velocity model by SAA+ method is shown in Figure 3-17c. Comparing to Figure 3-17b, the smearing artefact is suppressed substantially by imposing the sparsity promoting. However, the result is still unsatisfactory comparing with that of Figure 3-17a. When a fixed subset of shots is used during inversion, it will serve as a particular weighting matrix to induce noise and artefacts. Since then, such noise needed to be removed by some randomization process provided by SA technique. Figure 3-17d shows the result retrieved by SA+ method. The image quality is further improved comparing to that of Figure 3-17c. It approves that the random redrawing process (SA technique) clearly benefits the inversion to achieve better result in terms of final image resolution and quality. Additionally, the result from SA+ method obtained an almost identical estimated velocity model, comparing

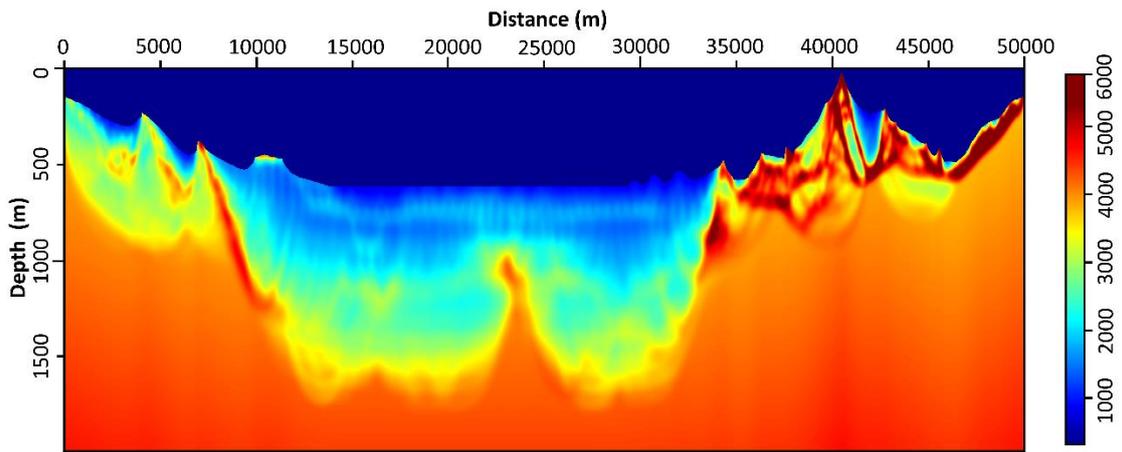
---

to the result of standard method using full dataset presented in Figure 3-17a, but with much higher computation efficiency.

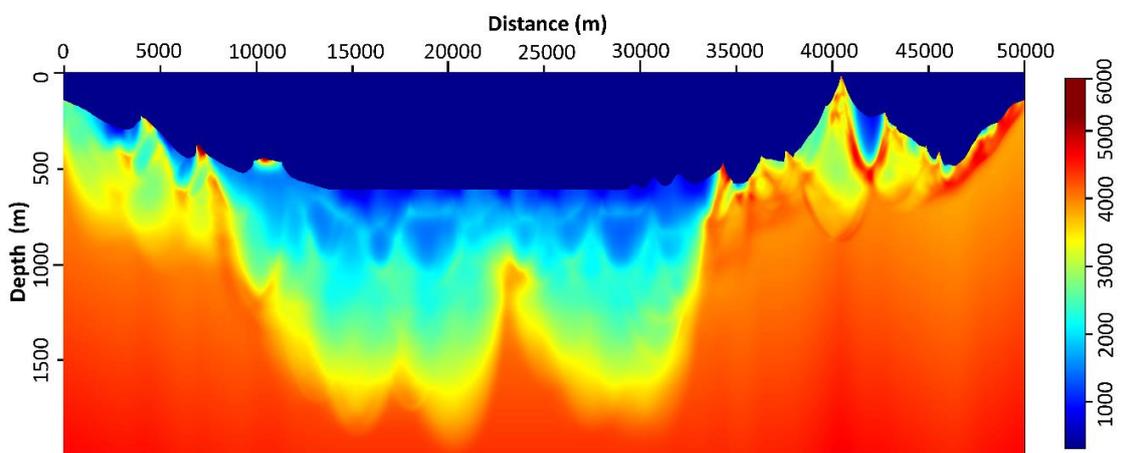
We further increase our subsampling batch size from 10% to 50% (~1000 sources) and using SA+ method to obtain result, as presented in Figure 3-17e. Overall, the results shown in Figure 3-17a, d and e are almost identical, confirming the robustness of SA+ method in different sufficient batch size of subsampling. By comparing Figure 3-17d and 3-17e, one could see the flexibility of SA+ on batch size selection, which is relatively insensitive. This property is favored in real field data application due to its feasibility. Besides, it is also another main reason that SA is superior to SAA. As discussed in section 3.4.1 and 3.4.2, to select a “large enough” batch size for optimal performance of SAA is very often difficult to decide.



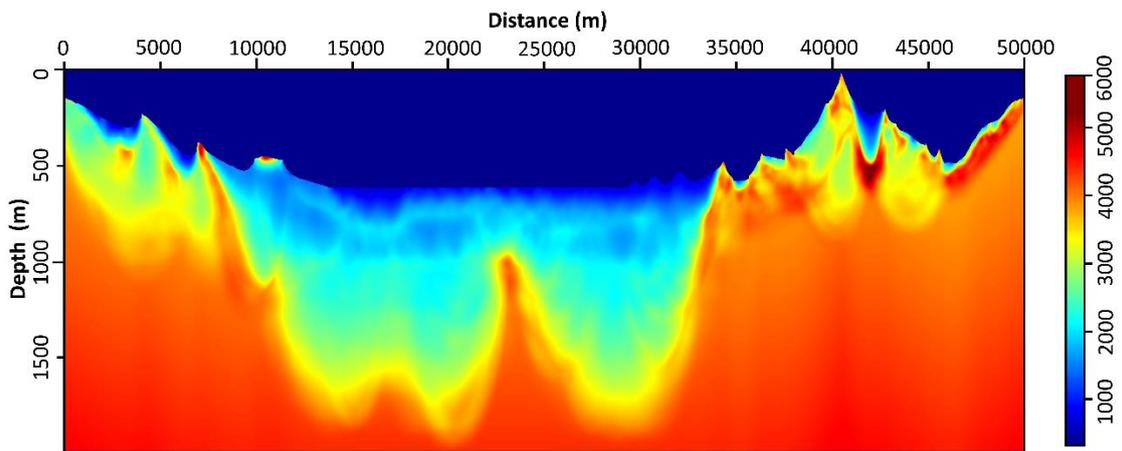
(a)



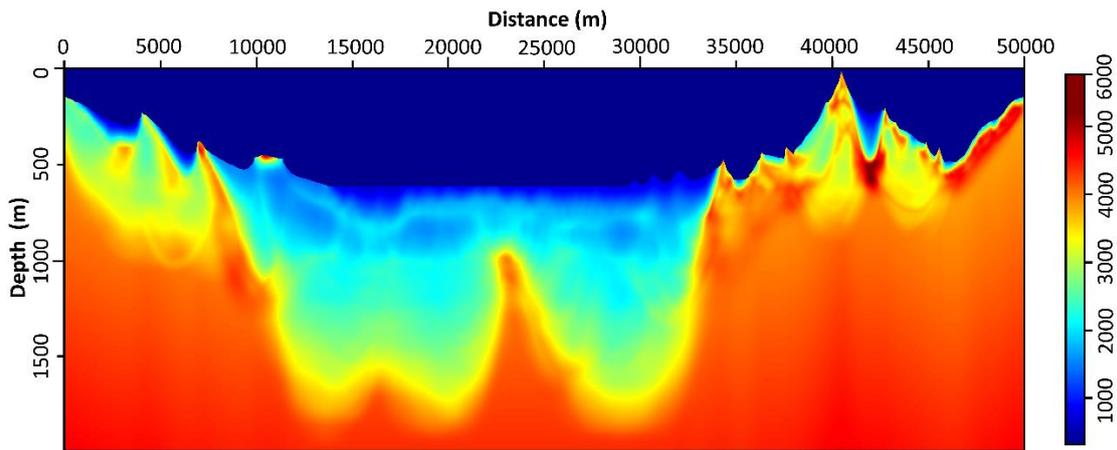
(b)



(c)



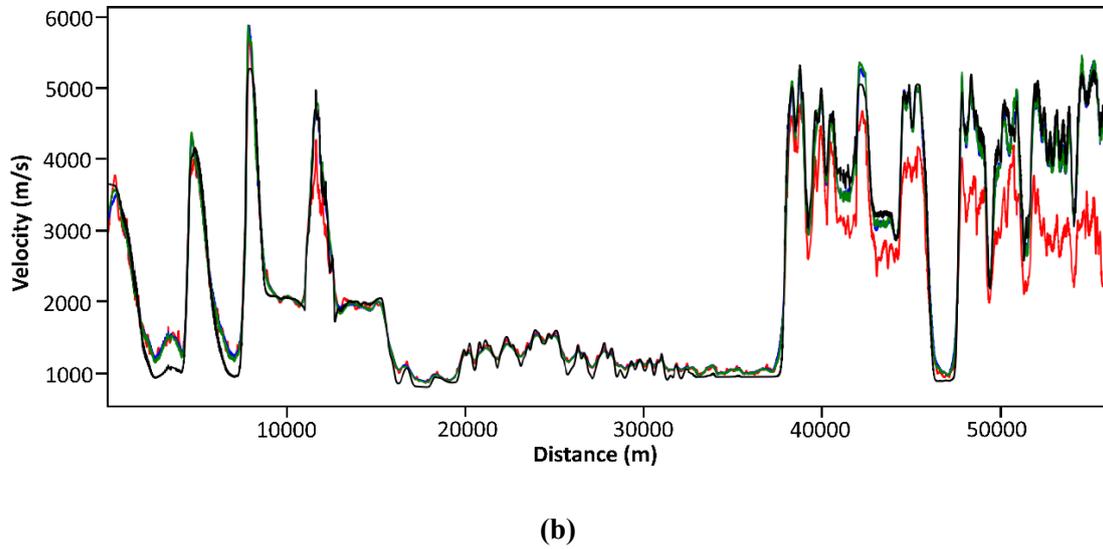
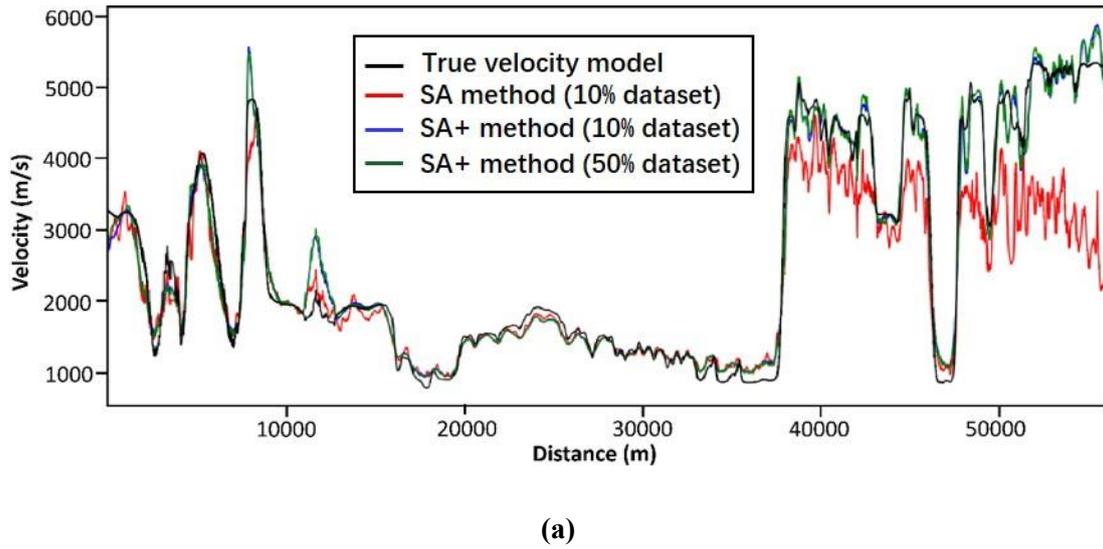
(d)



(e)

**Figure 3-17:** Velocity model reconstructed by (a) standard tomographic inversion using 100% dataset, (b) SA with 10% dataset, (c) SAA+ with 10% dataset, (d) SA+ with 10% dataset, (e) SA+ with 50% dataset.

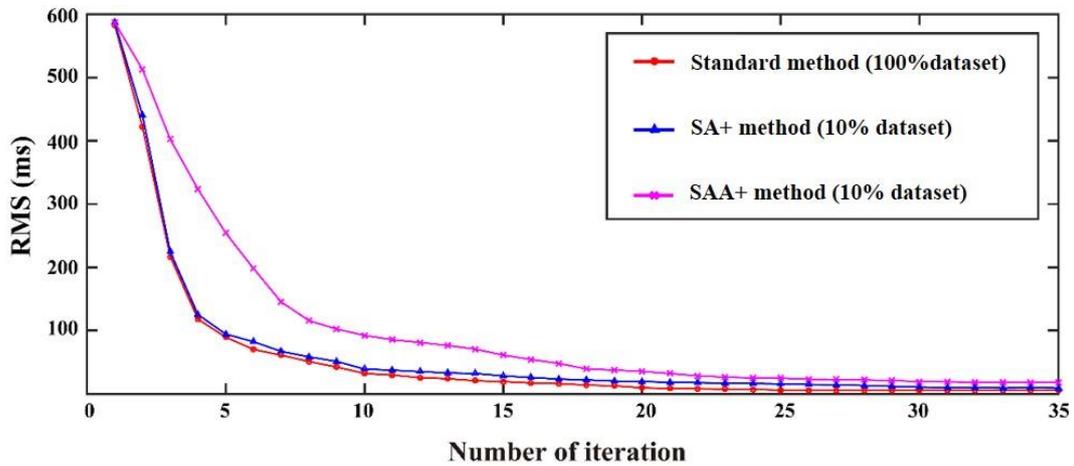
The velocity profile curves at depth of 50m and 100m below topographical surface are presented in Figure 3-18a and b, respectively. The true velocity is indicated by black line. The blue and green lines represent velocity trends from 10% and 50% dataset by SA+, respectively. The red line indicates the velocity trend of 50% dataset by SAA. Noted that the velocity trend of red line indicates that the SAA still failed in recovering complex velocity structure in right region of subsurface velocity model, even with 50% dataset used. These two figures confirm the inverted tomographic model results from Figure 3-17 in details. In the shallow depth, our reduced tomographic inversion method could obtain resulted image with quality as good as standard tomography method, but much higher computation efficiency. It demonstrated the capability of SA+ to image near-surface velocities with strong vertical and lateral variation at very low computation cost.



**Figure 3-18:** Velocity profiles at depth of (a) 50m and (b) 100m. Line color indicates different method employed with 10% and 50% dataset as labeled in the legend of (a).

The convergence history curve in terms of RMS variations was plotted in Figure 3-19 as function of number of iterations. The convergence reaches at RMS of 5.634ms, 7.862ms and 15.241ms for standard method (100% dataset), the SA+ (10% dataset) and SAA+ (10% dataset) respectively. We could see our proposed method are clearly benefitted from the renewals to maintain a continuous convergence and eventually obtained an acceptable RMS, meaning that the proposed method has a good model fit. However, for the same method without renewals,

its convergence rate and final model fit is unsatisfactory. Again, it proves the validity of redrawing subsample at each iteration in our sparse promoting approach.



**Figure 3-19:** The RMS travelttime residual for standard method (100% dataset), SA+ (10% dataset) and SAA+ (10% dataset) as labeled in figure legend.

Following the image quality comparison as presented in Figure 3.10, the corresponding comparison of computation efficiency regarding to memory and time cost is shown in table 3-5. Similar to the numerical test conducted in Chapter 2, all tomographic inversions are parallel computed in Linux Cluster with CPU of Intel Xeon at 2.60 GHz. For each inversion, there are 7 nodes used and each node has 30 threads. It should be noted that the memory cost records in Table 3-1 is single-threaded in average. From Table 3-5, we could see a significant improvement of computation efficiency in time cost wises due to substantial reduction of input shots brought by adopting SA+. Reduction of time cost are approximately in the same proportion of dataset reduction. The reason for slightly increase of memory cost that the dictionary and atoms matrix is needed to be restored iteratively during inversion. Still, the memory cost will not vary as inversion progress due to adjoint state technique imposed. Taking

the case of 10% dataset as example, the computation efficiency will be significantly improved due to a single evaluation of the reduced misfit function is 5 times cheaper than a full misfit evaluation. Though operations from online DL and inner layer subproblem solved by SPG-L1 as described in section 3.6.1 will slightly increase the computation cost, the improvement of computation efficiency is still impressive with at least 7~8 times faster than standard method.

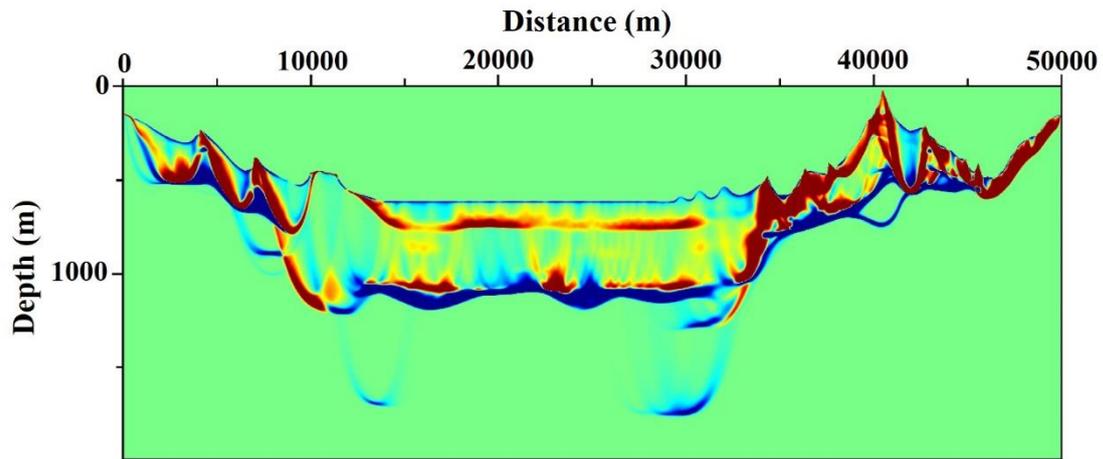
Method and used dataset	Memory Cost [Mb]	Time Cost [s]
Standard method (100% dataset)	140	440
SA (10% dataset)	146	89
SAA (10% dataset)	146	84
SA+ (10% dataset)	152	86
SA+ (50% dataset)	158	242

**Table 3-5:** Comparison of computation efficiency in memory and time cost corresponding

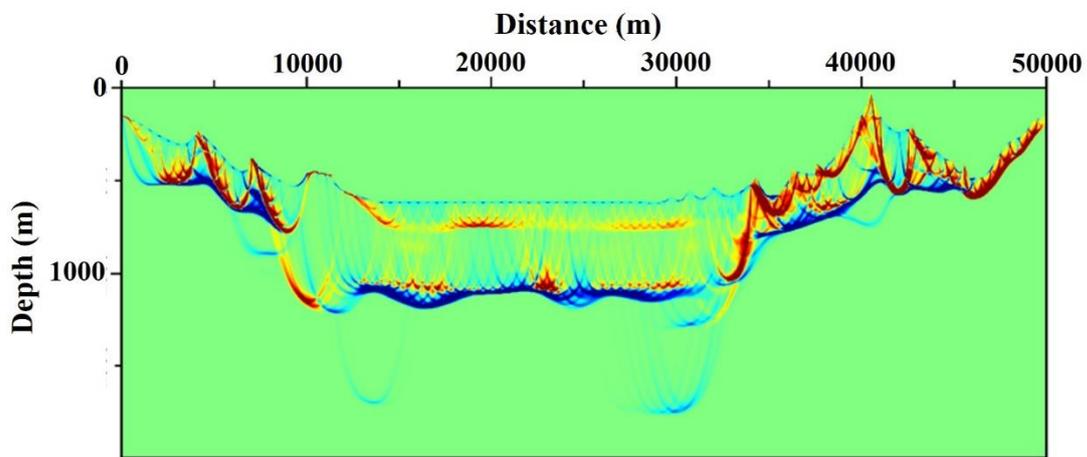
One of the gradients used during sparsity-promoting FAT tomographic inversion in Figure 3-18a perfectly explaining how the gradient was optimized through our sparsity promoting scheme by online ODL. The gradients in Figure 3-20a used 10% of dataset and had been optimized through our sparsity promoting technique. Same amounts of dataset were used in Figure 3-20b without sparsity promoting. The gradient in Figure 3-20c used full dataset to calculate from a standard tomographic inversion. The Figure 3-20a and c is similar and they all had relatively “smooth” perturbation in background. The sparsity promoting technique by online ODL from SA+ serves as “gradient preconditioning” via smoothing. However, it is different from those of deterministic approaches, since that we leverage ODL sparsity promotion, which preserves salient information from gradient update while removing other artifacts caused by subsampling. It can be clearly observed that the “smearing artefacts” in

---

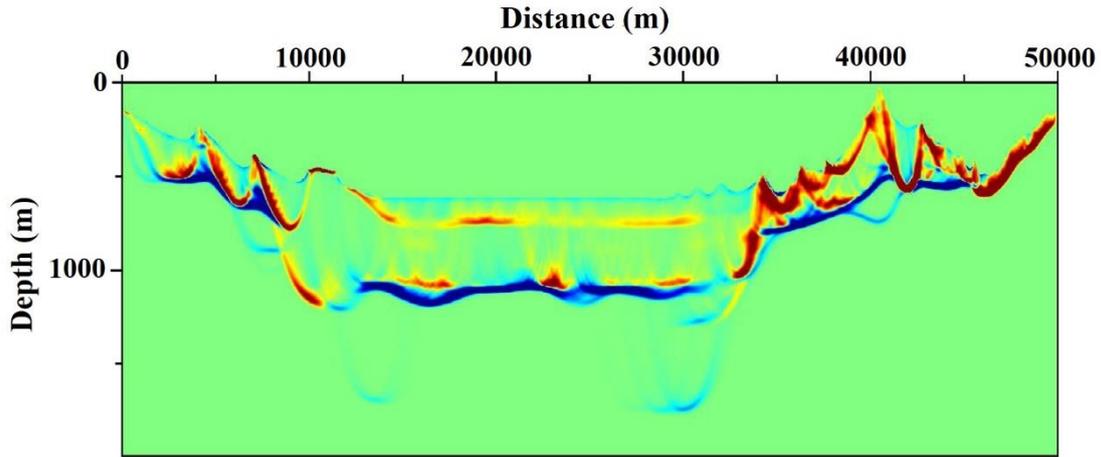
Figure 3.10a caused by severe subsampling were significantly suppressed in our optimized gradient as shown in Figure 3.10b. With aid of sparsity promoting, it is still safe to guarantee an optimal descent search, or to say, have a “correct” model update to guide the inversion process reach the convergency.



(a)



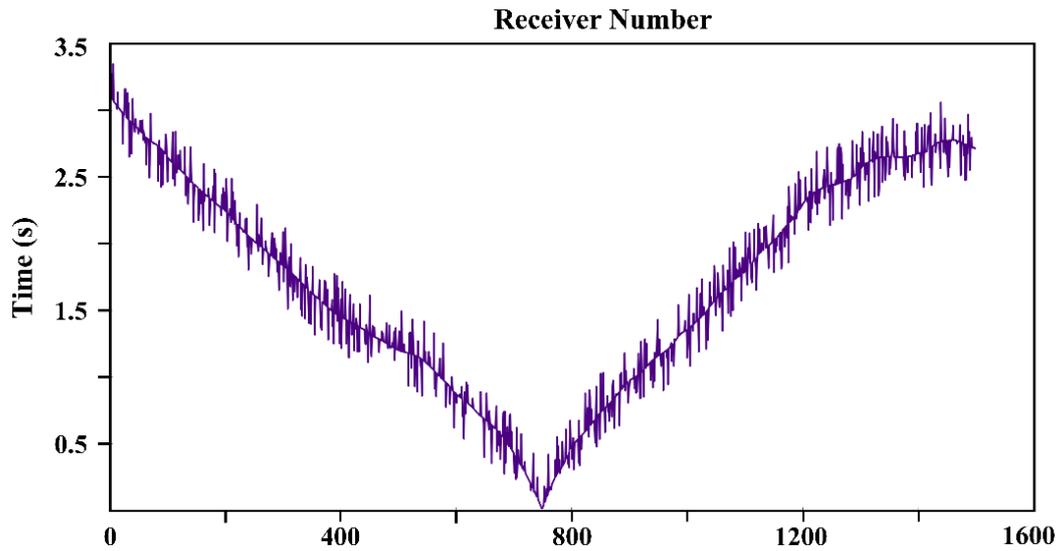
(b)



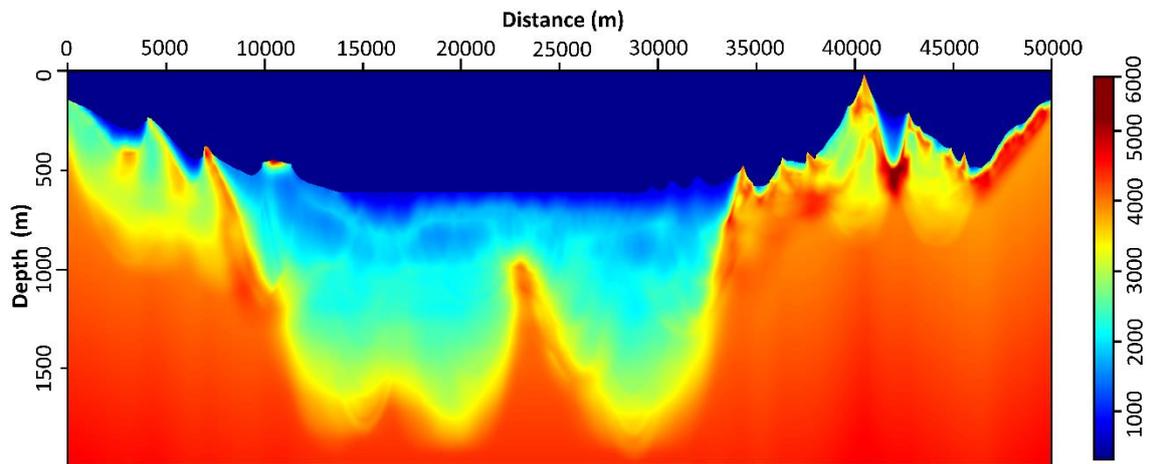
(c)

**Figure 3-20:** One of the gradients used for (a) standard method (100% dataset), (b) Standard method (10% dataset) and (c) after sparsity promoting (10% dataset).

We also testified the robustness of our proposed method on noisy dataset. For each single synthetic source gather, we randomly selected 40% numerical receivers to add in random noises within the range of  $[-0.3s, 0.3s]$  to generate noisy FAT picks as one of example shown in Figure 3-21a. In this case, we also only used 10% to implement with proposed method. The result is shown in Figure 3-21b, which is identical to the result from Figure 3.10c. And it proves the robustness of our proposed method in noisy data.



(a)



(b)

**Figure 3-21:** (a) The original synthetic FAT picks is blended with random noise ranged around  $[-0.3s, 0.3s]$ . (b) The reconstructed velocity model generated by the noisy data using SA+ with 10% dataset.

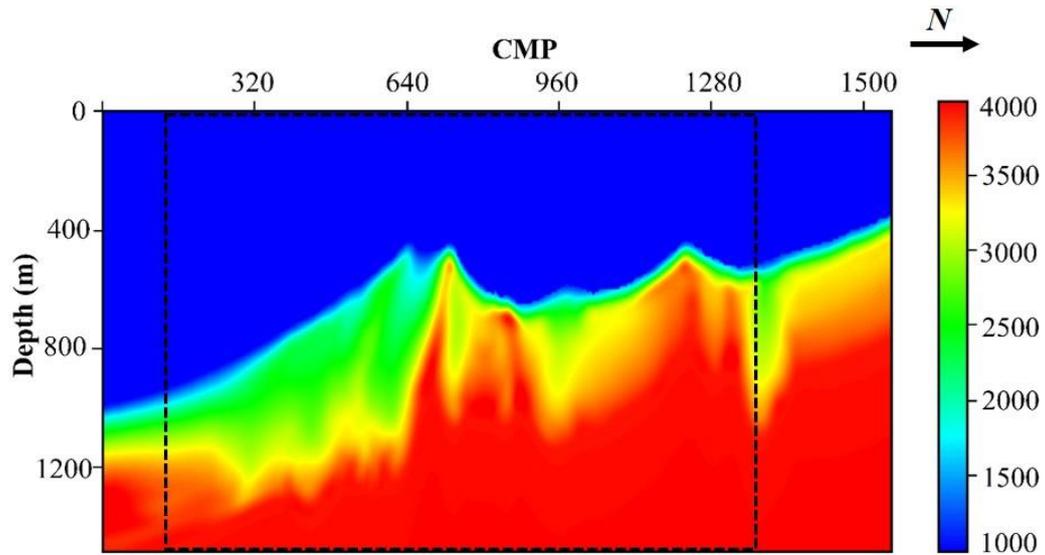
### 3.8.2 Real data test

The SA+ method is further tested on a real 3D land survey dataset from northwest in China. The field data is collected in a mountain area, where the weathered layer is extensively developed. The exposed area of sand and mudstone is relatively thin, and its thickness is

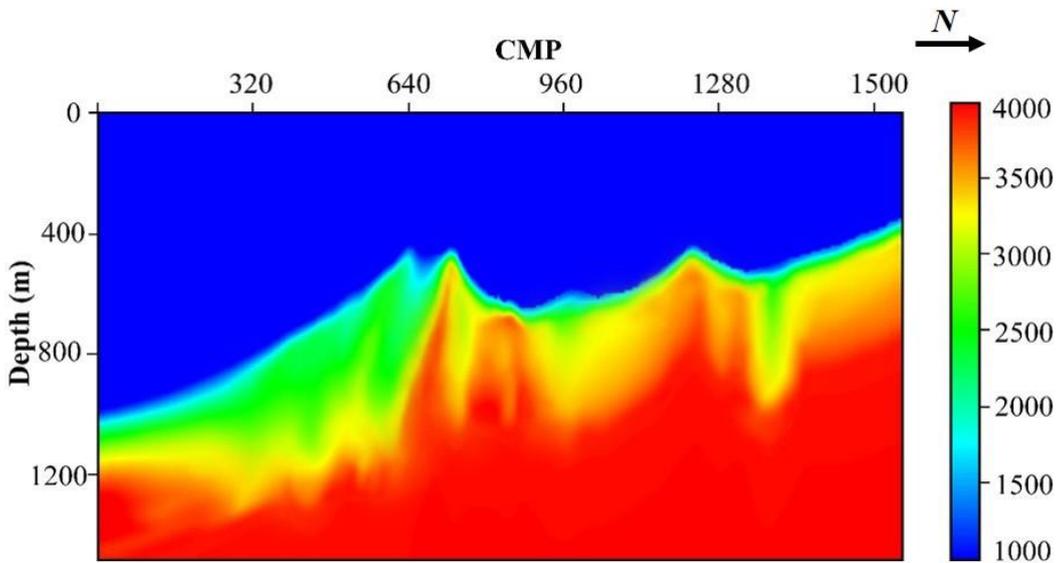
---

generally ranged from 10 to 20m in the north-central region. The gravel mountain is mainly located at south-central region of whole survey area, and it is appearing to be more severely weathered, in which its thickness is normally ranged from 20 to 40m, or even more than 100m in some areas. A vast alluvial fan develops at south side, the thickness of it is ranged from 50 to 170m at east-south region. The velocity change varies greatly, ranging from 1500m/s to 4000m/s. It is a typical area in urgently need for accurate near-surface velocity model estimation, by which a convincing static correction could be applied for mitigating the near-surface distortion effects and improving seismic image quality. A 2D test line is selected out for conducting this filed data experiment. In this 2D survey line, there are 523 shots and 1200 unique receivers. The max receiver number collected for a shot is 452 and the max offset is 7180m in this case. The seismic trace is recorded at 2ms sample rate.

The standard tomographic inversion method same as previous section, is employed with 100% dataset, and the reconstructed velocity model is shown in Figure 3-22a. The effective velocity range is in the dash line box. From the result, we could see the estimated velocity distribution, especially in the near surface, is basically correlated with observed geological characteristics. Then our proposed method applied on 20% dataset and obtain the result as shown in Figure 3-22b. Except for the deeper part at south side of insufficient ray coverage, two results are almost identical.



(a)

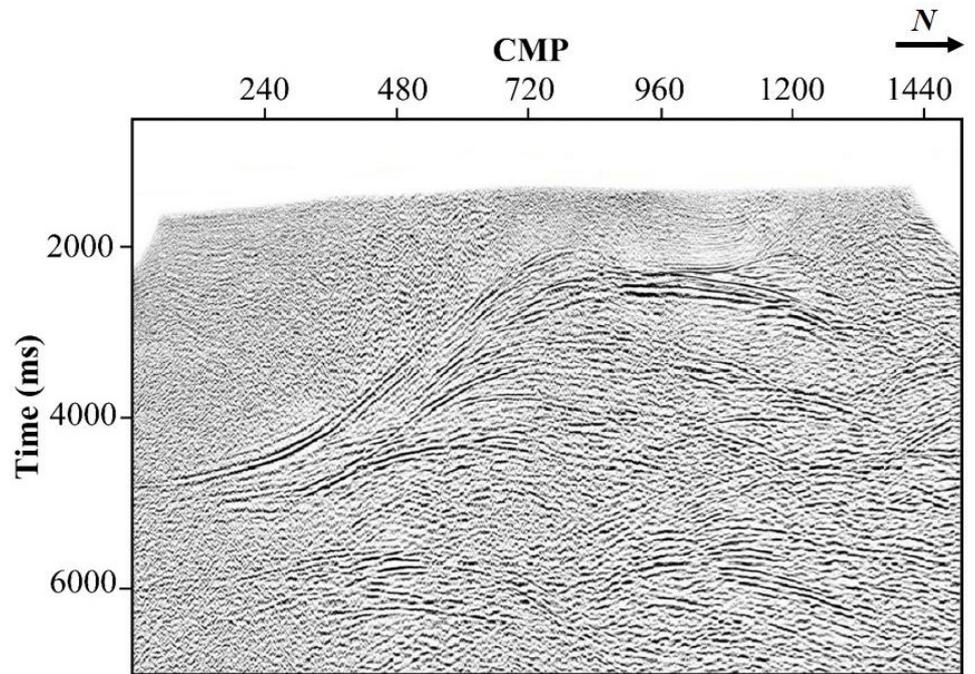


(b)

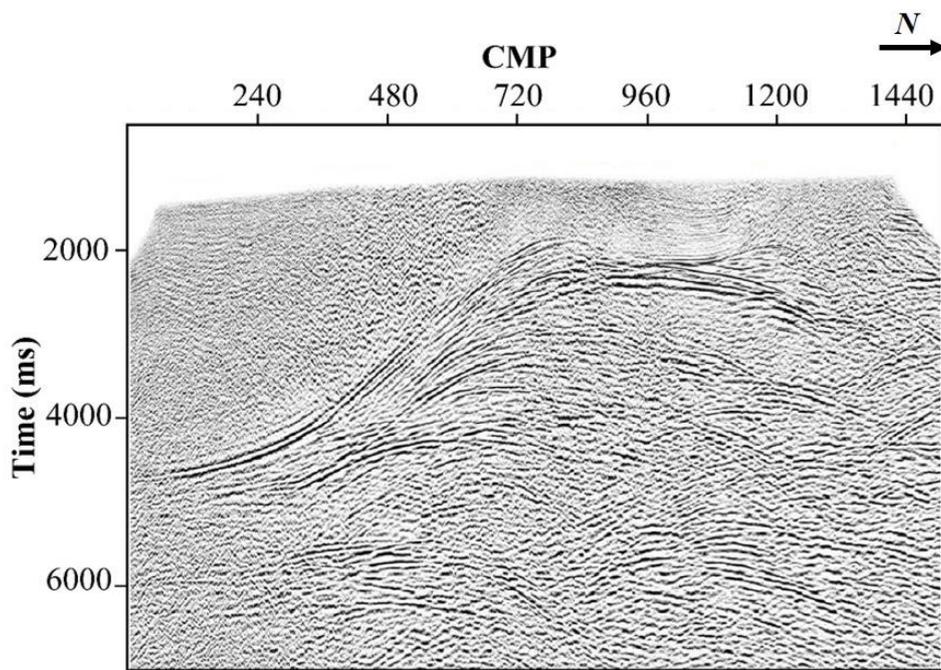
**Figure 3-22:** Velocity model reconstructed by (a) standard method with 100% dataset and (b) SA+ with 20% dataset.

To further evaluate the performance of our proposed method in real dataset, we used the estimated velocity model from Figure 3-22b for long wavelength static correction, with the standard method and proposed method result, as shown in Figure 3-23a and b. The average difference is small and within 2 sampling points and the subsequent data processing is not in

much difference. It is meaning that SA+ could generate an almost identical result as standard method with only small portion of full dataset at 20%.



(a)



(b)

**Figure 3-23:** Stack image after long-wavelength static correction corresponding to the inverted

---

velocity model obtained from (a) Standard method with 100% dataset and (b) SA+ with 20% dataset.

The field data test result proves that our proposed method could provide a satisfying and stable result, which is identical to standard method, but with much more higher computation efficiency. In this case, the computation cost of the standard method and our proposed method is 58s and 18s, respectively.

### **3.8 Summary and discussions**

In this chapter, an efficient FAT tomography method of SA+ is proposed. Using periodically random source redrawing in each inversion iteration, the computational cost is dramatically reduced. The method is built up on a modified Gauss-Newton algorithm. To mitigate the noise introduced by random subsampling of shots and maintain the convergency of inversion, we imposing online ODL to exploit convex-composite structure of the problem in sparse domain. The sparse representation provided by such lean-based dictionaries generated by training through inversion have benefit on tuning into sparser representation of patterns beyond piecewise smoothness. With this sparsity-promoting scheme, a LASSO subproblem with a compressive sparsity constraint is formed to be solved within our modified Gauss-Newton method using two-layer inversion strategy. SA+ method can dramatically reduce the computational complexity of the original tomographic inversion problem without degrading the image quality by regularizing the sparsity of model update over a predefined transform domain.

---

From synthetic data test, SA+ method obtained the optimal performance in terms of accuracy and robustness for different noise level. It also provides a practical guidance of selecting proper data decimation percentage with different noise level in first break picks. For instance, only ~10% dataset can provide a qualified result (RMSE within 100 m/s) at 10% noise level, and the computational cost is significantly reduced about 70%. Meanwhile, one should also be aware of geological complexity of target area. In practical, for an area with complicated near-surface or rapidly changed topography, larger data decimation percentage is necessary to ensure the accuracy and stability of SA+. For instance, when noise level rise above 10%, a data decimation percentage at least large than 15% should be a safe choice. As shown in real data example, though we directly use 20% dataset for SA+ to obtain a reasonable result, the computation cost is reduced around 70%. Overall, both of synthetic and field data test demonstrates that our proposed method could obtain an almost identical estimated velocity model as standard method but with much more higher computation efficiency during the inversion.

---

# CHAPTER 4 FIRST ARRIVAL ATTENUATION TOMOGRAPHY

## BASED ON ADJOINT-STATE METHOD<sup>1</sup>

### 4.1 Introduction

The amplitude and the phase of a propagating seismic waves will be distorted due to the inelasticity and heterogeneities of the subsurface medium (Bremaecker and Jean-Claude 1977; Müller et al. 2010). The quality factor ( $Q$ ) is designated to quantify the attenuation amount, and knowledge of the near-surface  $Q$  distribution will be help in improving image quality and correcting amplitude and phase information for accurately predicting reservoir properties (Best et al. 1994; Carcione et al. 2003). To derive  $Q$  distribution from surface seismic data, the frequency domain methods are mostly recommended, since that the time domain method often faces a challenge in separating intrinsic attenuation from spherical spread loss and transmission/reflection loss, etc. (Červený, 2001). Since then, amounts of tomographic generalization of the frequency-domain method (Rickett, 2006) are proposed for years, and they are often estimating  $Q$  distribution through inversion method from surface seismic data. One ultimate goal for such attenuation tomography method is to remove the elastic features, especially for the ones have contribution to amplitude attenuation (Calvaca and Fletcher, 2009). Besides, this is required to be done in certain type of attenuation model with the underlying

---

<sup>1</sup> A version of this chapter has been published. Xinwei Huang, Zhenbo Guo, Hua-Wei Zhou, Yubo Yue. First arrival  $Q$  tomography based on adjoint-state method. *Journal of geophysics and engineering*, 2020.

---

assumptions.

For a typical seismic exploration, many researchers have proved the effectiveness of frequency independent Q model over a certain seismic bandwidth either by real data application (e.g., Hargreaves, 1991; Brzostowski and McMechan, 1992; Quan and Harris, 1997) or experiment study (e.g., Futterman, 1962; Kjartansson, 1979). Considering such a model, the intrinsic attenuation of an anelastic medium can be characterized in terms of path attenuation factor ( $t^*$ ), which is related to the amount of attenuation along the ray path. It is proportional to the travel time and integrating the effects of both elastic and anelastic features of velocity ( $v$ ) and attenuation ( $Q$ ), respectively. Based on this relationship, to calculate the Q distribution, one can first obtain intermediate solution  $t^*$ , and then determine the Q distribution by inverting the linear system formed by  $t^*$ ,  $t$  and  $Q$  based on the assumption that ray paths are independent of  $Q$  for a given velocity model (Keers et al., 2012).

In this chapter, we proposed a two-step spectral ratio attenuation tomography to derive for Q distribution model, in which we first estimates  $t^*$  based on an adaptive correction method and then invert these path attenuation factor for the derivation of the attenuation ( $Q-1$ ) distribution. Though many attempts made to extract  $t^*$  from either reflection or transmission data (Nowack et al., 1997; Shen et al., 2016) it is still challenging to estimate Q model at near-surface by reflection-based Q tomographic inversion due to available offset information of shallow depth is limited (He et al. 2016). Since then, our proposed method is built on first arrival to build near-surface attenuation for better-utilizing information from far offset dataset.

---

In practice, comparing to transmission or reflection seismic data, one benefit of adopting first arrival of pre-stack surface data is relatively stable regarding to less prone to contaminated by other arrivals such as reflections, converted waves or interval multiples. The estimation of  $t^*$  in our proposed is done by exploiting amplitude variation at different arrivals using spectral ratio methods, which is quite similar to other previously proposed methods (Ganley et al. 1980; Brzostowski and McMechan, 1992; Cavalca and Fletcher, 2013 and 2015). However, unlike other  $t^*$  estimation approaches establishing on the same frequency response regarding the seismic attenuation effects, the influences of different frequency response of waveform, such as direct wave, turning wave and refraction were considered in our proposed method. Such influences were blended in the first arrivals, especially in the area of direct wave and refraction interfered each other, will eventually degrade the accuracy of  $t^*$  estimation. An adaptive correction method will be further adopted to evaluate the similarities of original first arrivals and synthetic waveforms generated by different estimated  $t^*$  based on different wave types, from which the optimal estimated  $t^*$  will be selected as observed  $t^*$  response of field data. Then, following a generalized tomographic inversion scheme by minimizing the discrepancy between observed  $t^*$  and synthetic  $t^*$ , the Q distribution can be achieved. To maintain the efficiency and feasibility of practical application on large seismic dataset, we adopt the same methodology of adjoint-state technique (Leung and Qian, 2006; Taillandier et al., 2009) into our tomographic inversion scheme as discussed in Chapter 2. An attenuate path factor  $t^*$  related governing equation is formulated to be implemented with adjoint-state technique, by which the explicit calculation of the gradient of misfit function could be avoided during tomographic inversion.

---

Our first arrival attenuation tomography algorithm is tested on synthetic and field data to demonstrate its feasibility and effectiveness of this method. A practical implementation about how to estimate  $t^*$  through spectral ratio and adaptive correction method in detail is also discussed in this chapter, especially for the case of different wave types (e.g., direct wave and refraction) interfered each other.

## 4.2 Theory Background

As discussed in previous section, we assume the wave propagation in a dissipative homogeneous medium can be described by a linear system for linear frequency attenuation model (Aki & Richards 2002). If the quality factor  $Q$  is frequency-independent — a constant  $Q$  model over the frequency band observed in seismic signal (Kjartansson 1979; Ganley and Kanasewich 1980). Ignoring velocity dispersion (i.e.,  $c(f) = c$ ), for a particular frequency  $f$ , the corresponding wavefield solution after propagating for a certain time can be determined by

$$\hat{U}(f, t) = \hat{U}_{ref}(f)GP(f, t)e^{i\phi(f,t)}e^{-f \cdot t^*}, \quad (4.1)$$

where  $\hat{U}_{ref}$  is the wavefield frequency response at reference location, and the  $G$  factor is assumed to be frequency-independent responsible for the effects of geometrical spreading, reflection/transmission coefficients, etc. Here  $P(f)$  is frequency-dependent amplitude term of wavefield response varied with different wave types propagating in the medium. The fourth and fifth terms in the fourth exponent are responsible for phase shift of wavefield response, and the fifth exponent is responsible for attenuation effects on amplitude decay (Ward and Toksöz, 1971). It is linearly proportional to frequency indicating that higher frequencies

---

contents suffer more attenuation than lower ones, and it can be characterized in terms of path attenuation factor  $t^*$  ((Romero et al. 1997; Cavalca and Fletcher, 2015), defined by

$$t^* = \int_{ray} \frac{\pi}{Q(l)c(l)} dl . \quad (4.2)$$

Where the integral is evaluated along the ray path  $l$  and two points source and receiver situated on it.  $Q(l)$  and  $c(l)$  are the quality factor and seismic velocity respectively, defined along each point of the ray path.

The first-arrival attenuation tomography requires estimation of attenuated traveltime based on Equation (4.1). Then the attenuated traveltime will be used to invert for Q factor model according to Equation (4.2) through tomographic inversion. Conventionally, the computation cost of ray-based tomography is almost linear with number of source-receiver pairs, whose ray path forward modeling is both time and memory consuming when coming for large seismic dataset. To improve the computation efficiency, we adopt our first arrival Q tomographic inversion with adjoint-state technique as introduced in Chapter 2. However, we first need to formulate a new governing (or Eikonal) equation to establish the relationship of Q distribution, velocity field and the corresponding time field.

#### 4.2.1 Forward modeling: governing equation for $t^*$

Converting Equation (4.2) from integral form to differential form, yields

$$\frac{dt^*}{dl} = \frac{\pi}{Qc} , \quad (4.3)$$

---

where  $d/dl$  is directional derivative along the direction of ray path. Keers *et al.* (2001) has proved that the attenuation only affects the waveform through the complex and frequency-dependent traveltime when attenuation is small ( $Q^{-1} \ll 1$ ), other than changing of ray paths. In other words, the ray path will remain the same regardless of whether the media is viscosity or not. Therefore, it is feasible for us to separate the velocity and Q tomographic inversion into two cascading applications in our proposed method.

In isotropic media, the propagation direction of ray is consistent with the gradient of the first arrival traveltime field. Once we obtained the inverted velocity model by travel-time tomography, we can calculate the first arrival by solving the Eikonal equation, which also provides the magnitude of the gradient of the first arrival traveltime field,

$$|\nabla t(\mathbf{x})| = \frac{1}{c(\mathbf{x})} . \quad (4.4)$$

For the convenience of derivation, denoting  $Q(\mathbf{x})^{-1}$  as  $m(\mathbf{x})$  to represent the parameterized attenuation model. The left side differential term  $dt^*/dl$  in Equation (4.3) yields  $\nabla t^*(\mathbf{x})$  as a Laplace operator applied on, and multiplied it with equation (4),

$$\nabla t^*(\mathbf{x}) \cdot \nabla t(\mathbf{x}) = m(\mathbf{x}) \frac{\pi}{c^2(\mathbf{x})} . \quad (4.5)$$

As we assumed that velocity field  $c(x)$  and its corresponding travel time field  $t(x)$  is already known, the Equation (4.5) is the governing equation of  $t^*$  and establish the relationship between  $t^*$  and Q distribution of  $m(x)$  for each spatial point  $x$  throughout the whole model

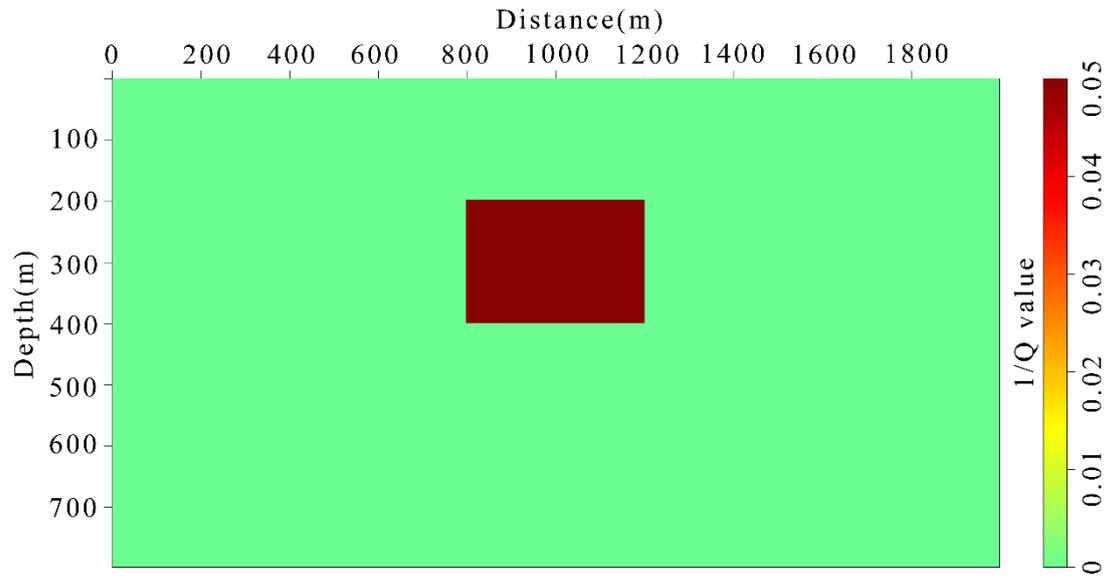
---

space. Similar to Chapter 2, this governing equation could be solved by FSM algorithm, which is finite difference method performed in the model parametric space, but with different parameterization. For a more detailed numerical implementations of FSW, we refer to Leung & Qian (2006). In this case, the computational cost of Equation (4.5) is linear with the size of model, rather than the number of shot-receiver pairs. Besides, the ray-path storage required by ray-tracing based algorithm is not required, hence the memory occupation will be reduced thoroughly. From Equation (4.5), we could see that  $t^*$  is linearized by the model parameter  $m(\mathbf{x})$ , which indicates that the corresponding attenuation tomographic inversion will also be transferred into a linear process.

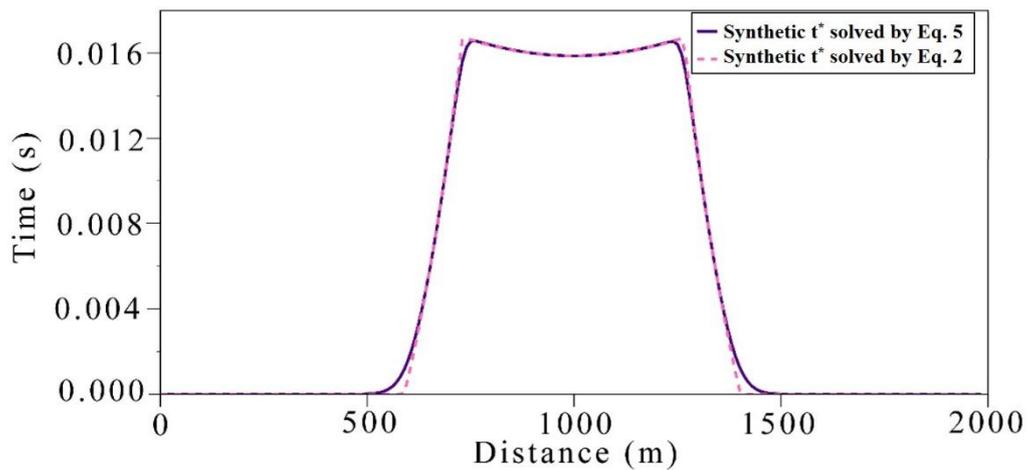
### **The accuracy of governing equation**

In order to verify the accuracy of numerical solutions obtained from Equation (4.5), we conducted a synthetic test to compare the forward modeling results generated by both of conventional ray-tracing algorithm based on Equation (4.2) and FSW algorithm based on Equation (4.5). The synthetic test is performed in a 2D constant background velocity model as shown in Figure 4.1a. The model width is 2000m and depth is 800m. The value of rectangular attenuation anomaly  $Q$  is unitless of magnitude 20. The source located right beneath  $Q$  anomaly at depth of 700 m and lateral position of 1000m. The receivers are all located along the flat top surface of the model. From Figure 4.1b, we can see that two calculated  $t^*$  are almost identical to each other, except the small misfit above the boundary corners of the  $Q$  anomalies. These errors are caused by the limited precision of finite difference method solving for Equation (4.5),

and they are acceptable without any significant impact on the final tomographic inversion result.



(a)



(b)

**Figure 4- 1:** Comparison of  $t^*$  calculated by two forward modeling method based on Eq.2 (ray-tracing) and Eq.5 (Eikonal equation solved by FSM): a) True attenuation model ( $Q^{-1}$ ); b)  $t^*$  calculated by equation 2 in solid blue line and  $t^*$  calculated by equation 5 in dashed magenta line.

#### 4.2.2 Attenuation Tomography based on adjoint-state method

Q tomography aims to obtain the estimated Q distribution model that minimizes the

---

difference between the observed and synthetic  $t^*$ . Given the model  $m$  ( $Q^{-1}$  in this case), a least square misfit function  $O(m)$  is chosen to quantify such difference and can be defined as

$$\begin{aligned}
O(m) &= \frac{1}{2} \sum_{s=1}^{ns} \sum_{r=1}^{nr} \left| w(s,r) \left[ t_{cal}^*(s,r) - t_{obs}^*(s,r) \right] \right|^2, \\
&= \frac{1}{2} \sum_{s=1}^{ns} \int_{\Omega} \left\{ w(\mathbf{x}) \left[ t_{cal}^*(s,\mathbf{x}) - t_{obs}^*(s,\mathbf{x}) \right] \right\}^2 \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x}
\end{aligned} \tag{4.6}$$

where  $s$  is shot number,  $r$  is receiver number,  $ns$  is the number of shots,  $nr$  is the number of receivers in each shot,  $\Omega$  is the whole model space,  $w(s,r)$  and  $w(\mathbf{x})$  is weighting coefficient for each receiver,  $t_{cal}^*$  is the calculated attenuated time and  $t_{obs}^*$  is the observed attenuated time.  $\mathbf{x}_r$  is the set of all the receiver locations in one shot.

Here we simply use the conjugate gradient method to invert for the attenuation model iteratively. The Q factor at  $k$ th iteration will be update by

$$m_{k+1} = m_k + \alpha m_k^{update}, \tag{4.7}$$

where

$$m_k^{update} = \begin{cases} -g & \text{if } k = 0 \\ -g_k + \beta m_{k-1}^{update} & \text{if } k > 0 \end{cases}. \tag{4.8}$$

To avoid intensive computational burden of Fréchet derivative matrix approximation, we adopted the adjoint-state technique to efficiently obtain the gradient  $g$  of misfit function (4.6) by solving for a new formulated adjoint Partial Differential Equation (PDE), and it turns original inverse modelling problem into a forward modelling problem. Here we derive the mathematical representation of  $g$  based on adjoint state technique (Plessix 2006; Huang &

---

Bellefleur 2012).

Suppose a perturbation  $\delta m$  of the attenuation model  $m$  induces a perturbation  $\delta t^*$  of attenuated time  $t^*$ , and also induces a perturbation  $\delta O(m)$  of objective function  $O(m)$ . By ignoring the higher order term of  $\delta t^*$ ,  $\delta O(m)$  can be expressed by

$$\delta O(\mathbf{m}) = \sum_{s=1}^{ns} \int_{\Omega} w^2(\mathbf{x}) \delta t^*(t_{cal}^* - t_{obs}^*) \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x}. \quad (4.9)$$

Since Equation (4.5) is a linear equation,  $\delta m$  and  $\delta t^*$  is also controlled by Equation (4.5). For expression convenience, we will only derive objective function perturbation for a single source as  $\delta O_s(m)$ , and the total perturbation  $\delta O(m)$  would be a simple summation from all sources. Introducing the relationship between  $\delta m$  and  $\delta t^*$  into Equation (4.9) by augmented function yields

$$\begin{aligned} \delta O_s(\mathbf{m}) = & \int_{\Omega} w^2(\mathbf{x}) \delta t^*(t_{cal}^* - t_{obs}^*) \delta(\mathbf{x} - \mathbf{x}_r) d\mathbf{x} \\ & + \int_{\Omega} \lambda(\mathbf{x}) \left[ \nabla \delta t^*(\mathbf{x}) \cdot \nabla t(\mathbf{x}) - \delta m(\mathbf{x}) \frac{\pi}{c^2(\mathbf{x})} \right] d\mathbf{x}, \end{aligned} \quad (4.10)$$

where  $\lambda$  is adjoint state variable. Simplifying Equation (4.10) through integration by parts for the second term yields

$$\begin{aligned} \delta O_s(\mathbf{m}) = & \int_{\Omega} \delta t^* \left[ (t_{cal}^* - t_{obs}^*) w^2(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_r) - \nabla \cdot (\lambda(\mathbf{x}) \nabla t) \right] d\mathbf{x} \\ & + \int_{\partial\Omega} \delta t^* \lambda(\mathbf{x}) \nabla t \cdot ds - \int_{\Omega} \lambda(\mathbf{x}) \delta m(\mathbf{x}) \frac{\pi}{c^2(\mathbf{x})} d\mathbf{x}. \end{aligned} \quad (4.11)$$

The terms in Equation (4.11) associated with  $\delta m$  should be kept and the other terms can be eliminated by choosing adjoint state variable  $\lambda$  that satisfies

---


$$\nabla \cdot (\lambda(\mathbf{x}) \nabla t) = (t_{cal}^* - t_{obs}^*) w^2(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_r), \quad (4.12)$$

and boundary condition

$$\int_{\partial\Omega} \delta t^* \lambda(\mathbf{x}) \nabla t \cdot \mathbf{ds} = 0, \quad (4.13)$$

then Equation (4.11) can be further simplified to

$$\delta O_s(\mathbf{m}) = - \int_{\Omega} \lambda(\mathbf{x}) \delta m(\mathbf{x}) \frac{\pi}{c^2(\mathbf{x})} d\mathbf{x}. \quad (1)$$

Now we build the relationship between  $\delta m$  and  $\delta O(m)$ , and the gradient can be expressed as

$$g(\mathbf{x}) = - \frac{\pi \lambda(\mathbf{x})}{c^2(\mathbf{x})}. \quad (4.15)$$

Adjoint state variable  $\lambda$  over the parametric space can be solved by Equation (4.12) and (4.13), and its physical meaning can be explained as the back propagation of the attenuated time residual from attenuation model to the source. Once we obtained the misfit function gradient  $g(\mathbf{x})$ , the Q distribution in terms of  $m(\mathbf{x})$  can be iteratively updated according to Equation (4.7), until the misfit  $O(m)$  reduces to a given tolerance level.

### 4.2.3 Estimation of attenuated time and adaptive correction method

To conduct first arrival attenuation tomography, two input of velocity model  $c(\mathbf{x})$  and path attenuation  $t^*$  are required. At the first stage of our first arrival attenuation tomography, an inverted velocity model  $c(\mathbf{x})$  can be obtained by any general traveltime tomography. It will

---

be used as one of inputs for the following Q tomographic inversion to calculate the misfit function gradient, as indicated by Equation (4.15). The synthetic first arrival traveltime for each receiver will be calculated according to  $c(\mathbf{x})$ . They will be utilized as guidance for constraining the estimation of attenuated traveltime within certain time windows from attenuated seismic traces. In practical, such implementation has benefit on obtaining stable synthetic first arrival traveltimes for all receivers without involving manual picking errors. This is also one aspect of our proposed first arrival attenuation tomography superior to those of using seismic reflection, refraction or other types of data.

Now we could properly locate a time window at synthetic first arrival traveltime in a single seismic trace, in which the path attenuation factor is included. The following issue is the main challenge of estimating  $t^*$  from windowed seismic trace, where the  $t^*$  is implicitly blended in the first arrival wave trains. Obviously, it is impossible to manually pick  $t^*$  from seismogram as conventional traveltime tomography process. Since then, we developed a feasible approach to extract  $t^*$  from the first arrival by utilizing its amplitude spectrum variation.

According to the propagation properties described by Equation (4.1), to avoid the complex and wrapping issues caused by phase shifting term, we simply use its amplitude information to estimate attenuated time. From Equation (4.1), we have the amplitude spectrum of attenuated wavefield response propagating for a certain time expressed as

$$A(f) = P_g A_{ref}(f) P_p(f) e^{-ft^*}. \quad (4.16)$$

---

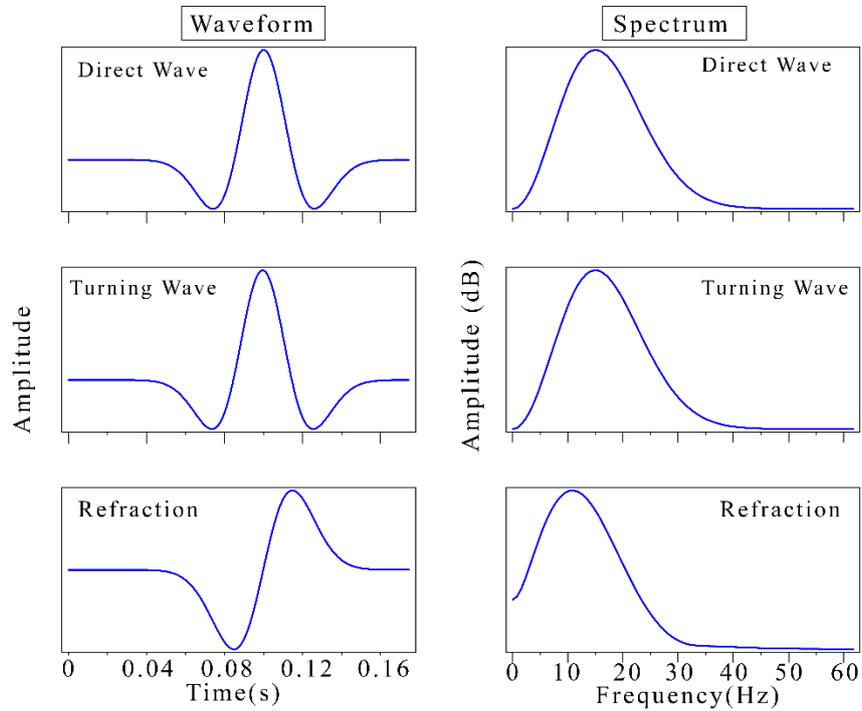
$A_{ref}(f)$  is the amplitude spectrum of wavefield at reference location,  $P_g$  is frequency-independent geometric factor including the effects of geometrical spreading,  $P_p(f)$  is frequency-dependent propagation response for a certain wave type, and  $e^{-ft^*}$  is the attenuation response.

To estimate  $t^*$ , we need to remove the effects from other frequency-dependent terms of both  $A_{ref}(f)$  and  $P_p(f)$ . Then the left attenuation response term  $e^{-ft^*}$  can be used for estimating  $t^*$ . Now we will discuss the exact forms of  $P_p(f)$  corresponding to different wave types. There are mainly three wave types existed in first arrival waveform: direct wave, turning wave and refraction. It is not easy to separate them directly in apparent way. Fortunately, the direct and turning wave have the same amplitude spectrum and waveform as source wavelet, while the waveform of refraction is the integral of source wavelet (Aki & Richards 2002). These frequency-dependent propagation responses can be expressed as

$$\begin{aligned}
 P_p(f) &= 1 && \text{for direct wave and turning wave} \\
 P_p(f) &= \frac{1}{2\pi f} && \text{for refraction}
 \end{aligned} \tag{4.17}$$

Figure 4-2 illustrates that how frequency-dependent propagation responses varied with these three wave types. All these waveforms are modeled by a 3D acoustic equation finite difference solver and their waveforms of first arrival were windowed out manually. The source wavelet we used in this experiment is Ricker wavelet at main frequency of 15 Hz. In the left column of Figure 4-2, we could see that the direct and turning wave have the identical waveform as original Ricker wavelet, while the refraction is different. One should note that the

refraction preserves a low peak frequency even the attenuation effect is not considered in this acoustic modelling case. Such low peak frequency phenomenon is caused by the frequency-dependent propagation responses as indicated by Equation (4.17).



**Figure 4-2:** Waveform and Spectra of different type of waves.

Inspired by the observation above, to identify and separate refraction from other wave types among first arrivals, we will calculate attenuated time twice by Equation (4.16) according to assumptions of direct/turning wave and refraction respectively. Here we use logarithmic spectral ratio method to estimate attenuated time. The natural logarithm applied to the both side of Equation (4.16) gives a linear function

$$\ln \frac{A(f)}{A_{ref}(f)P_p(f)} = -ft^* + \ln P_g. \quad (4.18)$$

---

Here  $\ln P_g$  can be treated as a constant. Thus, within selected frequency band, the attenuated time can be estimated by applying the least-square linear regression which calculates the average slope of spectrum ratio  $\ln \frac{A(f)}{A_{ref}(f)P_p(f)}$  with frequency  $f$ .

These two estimated attenuated times will be applied with reference signal  $A_{ref}$  for all selected frequency bands based on Equation (4.16), to generate synthetic attenuated seismic data  $S_{syn}$  in time domain. Then we can evaluate the similarity between synthetic and real attenuated seismic data  $S_{real}$  for each sample point ( $i=0,1,2,\dots,n.$ ) in time domain. The similarity coefficient is defined as

$$Similarity = \frac{\sum_{i=0}^n S_{real}(i) S_{real}(i)}{\sqrt{\sum_{i=0}^n S_{real}^2(i)} \sqrt{\sum_{i=0}^n S_{syn}^2(i)}}. \quad (4.19)$$

Among two estimated attenuation time, the one with larger similarity coefficient will be selected as optimal attenuated time for further first arrival Q tomography. Furthermore, similarity coefficient can be served for quantifying the reliability of the attenuated time. In practice, many of traces in seismic data contaminated with noises severely, from which is unreliable to be used as attenuated time extraction. Based on our experiment, we prefer to remove those noisy traces whose similarity coefficients are less than 0.5 and using the similarity coefficients as weighting factors in tomographic inversion for the remaining traces.

---

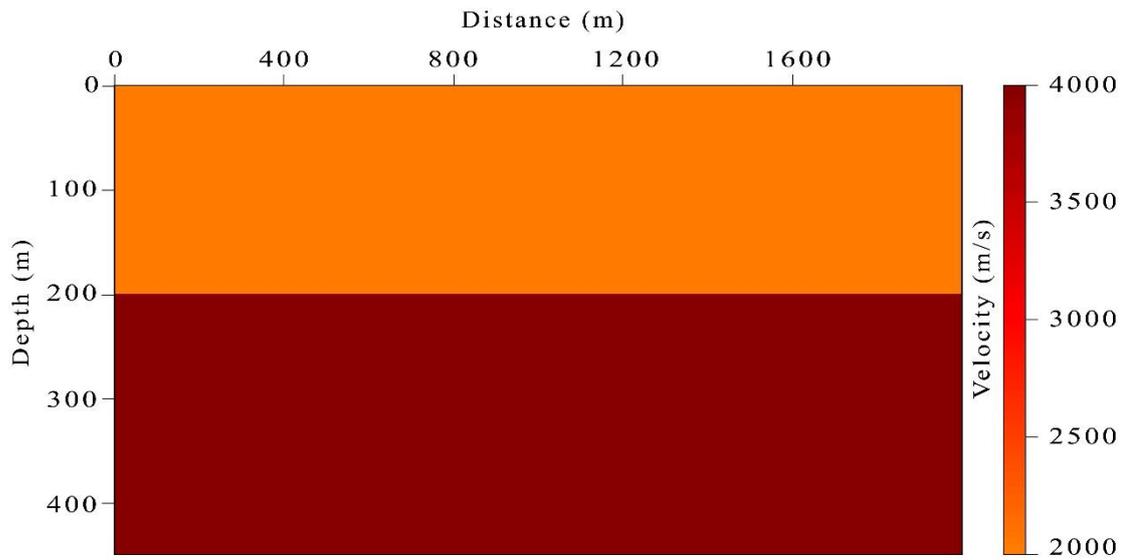
## 4.3 Numerical test

### 4.3.1 2-D synthetic data test: two-layer model

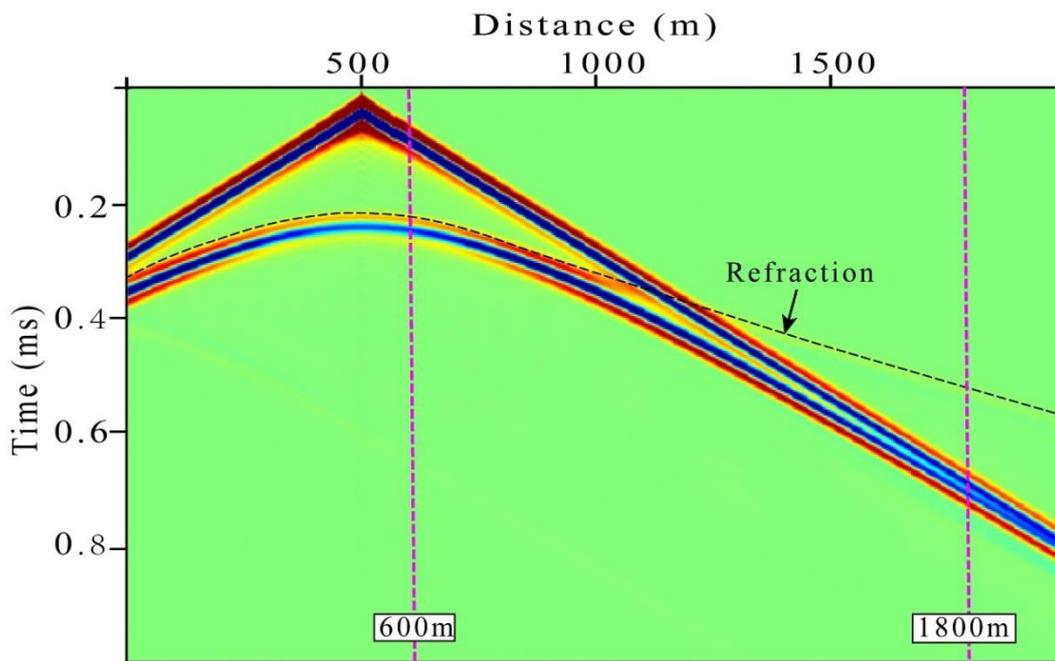
#### 1. Observation on direct wave and refraction

A two-layer synthetic model test is presented here to illustrate the influence of different wave types on estimating  $t^*$ . As shown in Figure 4.3, this half space velocity model consists of two constant velocity layers, the top layer is 2000m/s in velocity and 200 m in thickness, and the bottom layer is 4000 m/s in velocity.

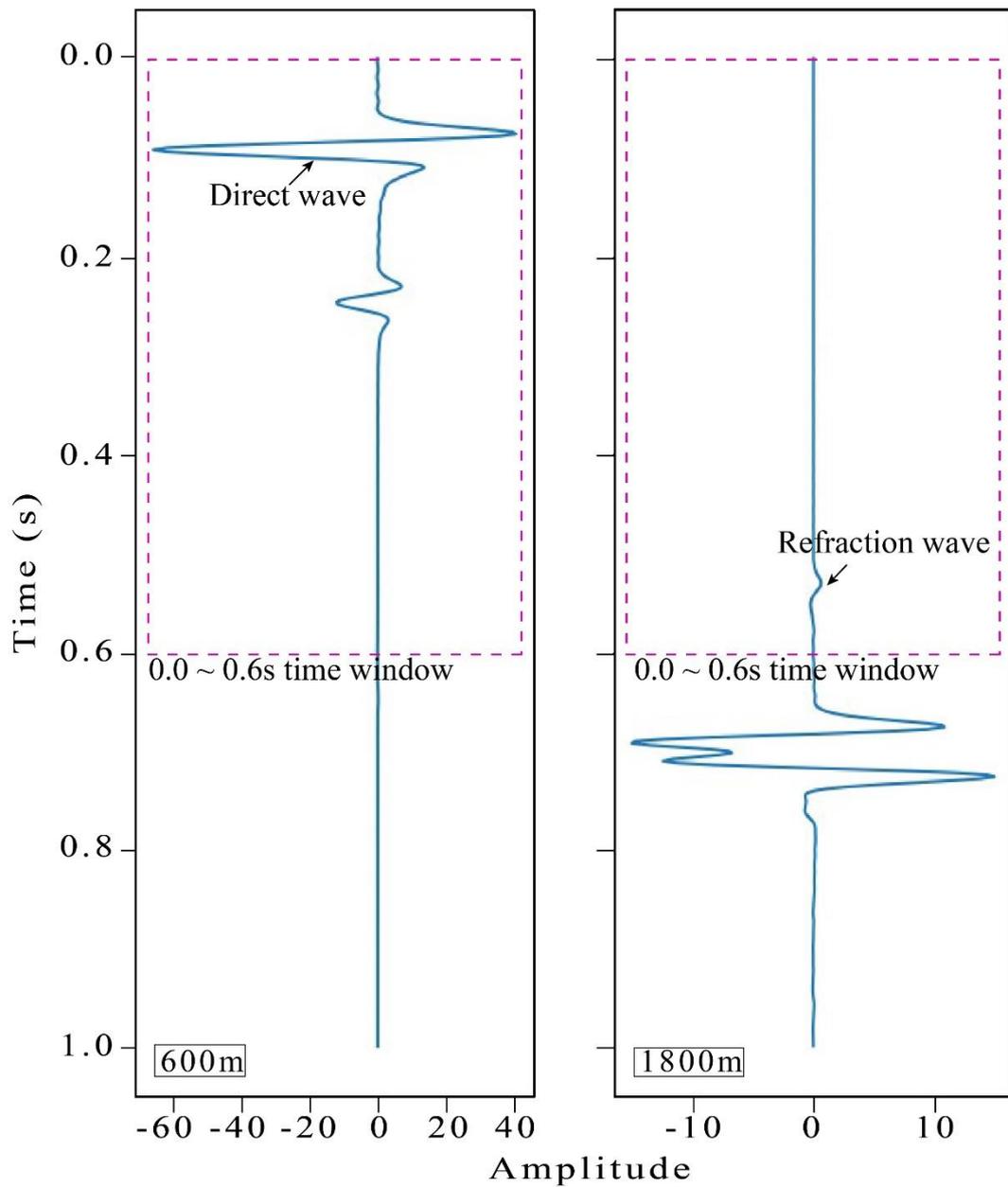
In this section, we will first examine the waveform behaviors and their amplitude spectra in the case of fully elastic assumption (without Q attenuation), and the difference of direct and refracted waves will be presented. The source wavelet is a Ricker wavelet with dominant frequency of 25 Hz. The original shot gather is shown in Figure 4.4. The first arrival waveform is dominated by direct wave from source location (at 500 m) to around 1300 m with an offset of 800 m. Then the refracted wave appeared to be recognizable first arrivals from distance of 1300m. Two seismic traces were extracted at distance 600 m and 1800 m as shown in Figure 4.5. In order to obtain accurate frequency spectrum of direct wave and refraction respectively, a time window from 0.0 s to 0.6 s is applied as indicated by rectangular box in magenta dash line of Figure 4.5, and the direct and refraction waveforms were extracted as shown in figure 6a and aligned based on the emergent onset point in figure 6b.



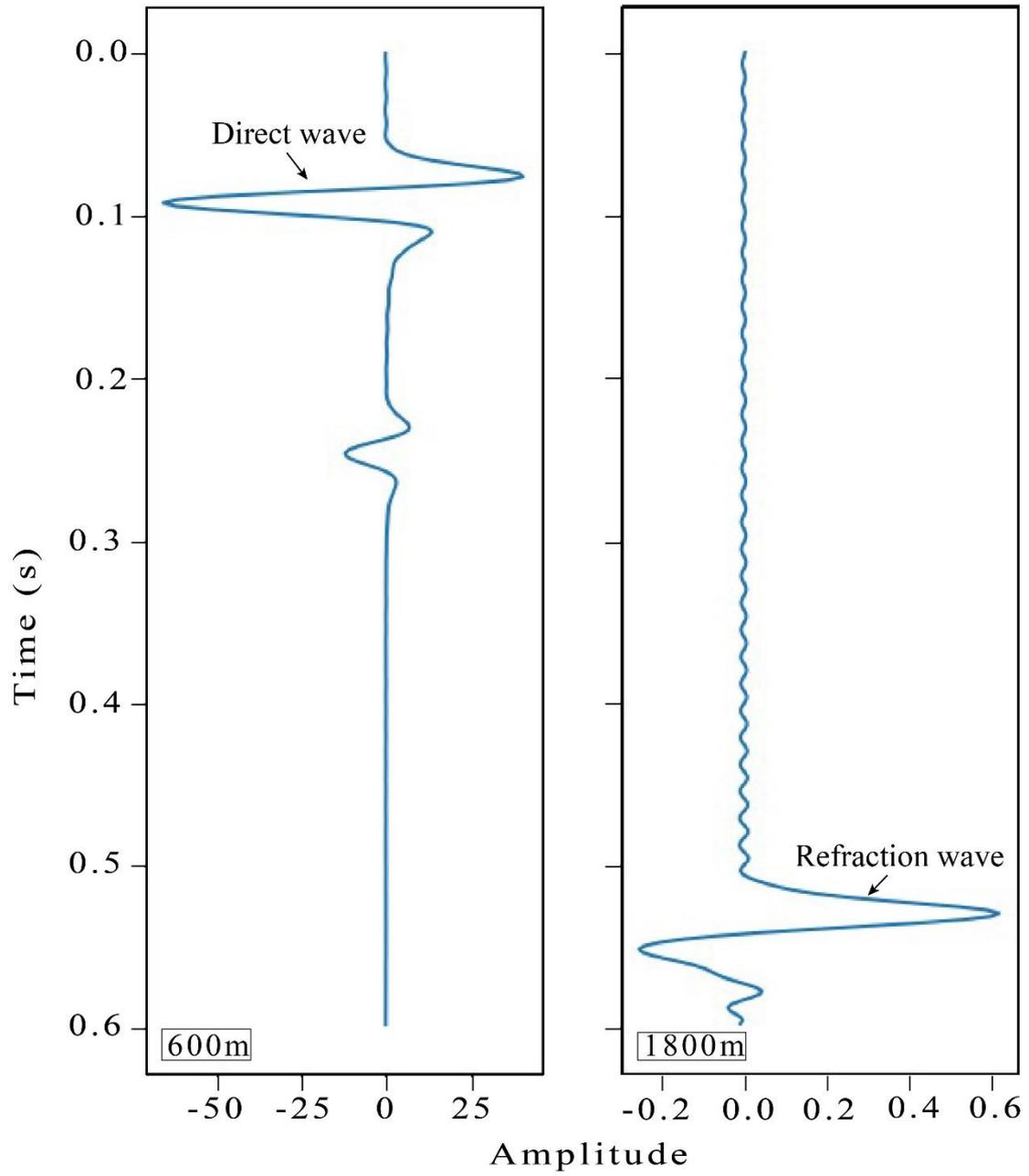
**Figure 4-3:** Two-layer velocity model.



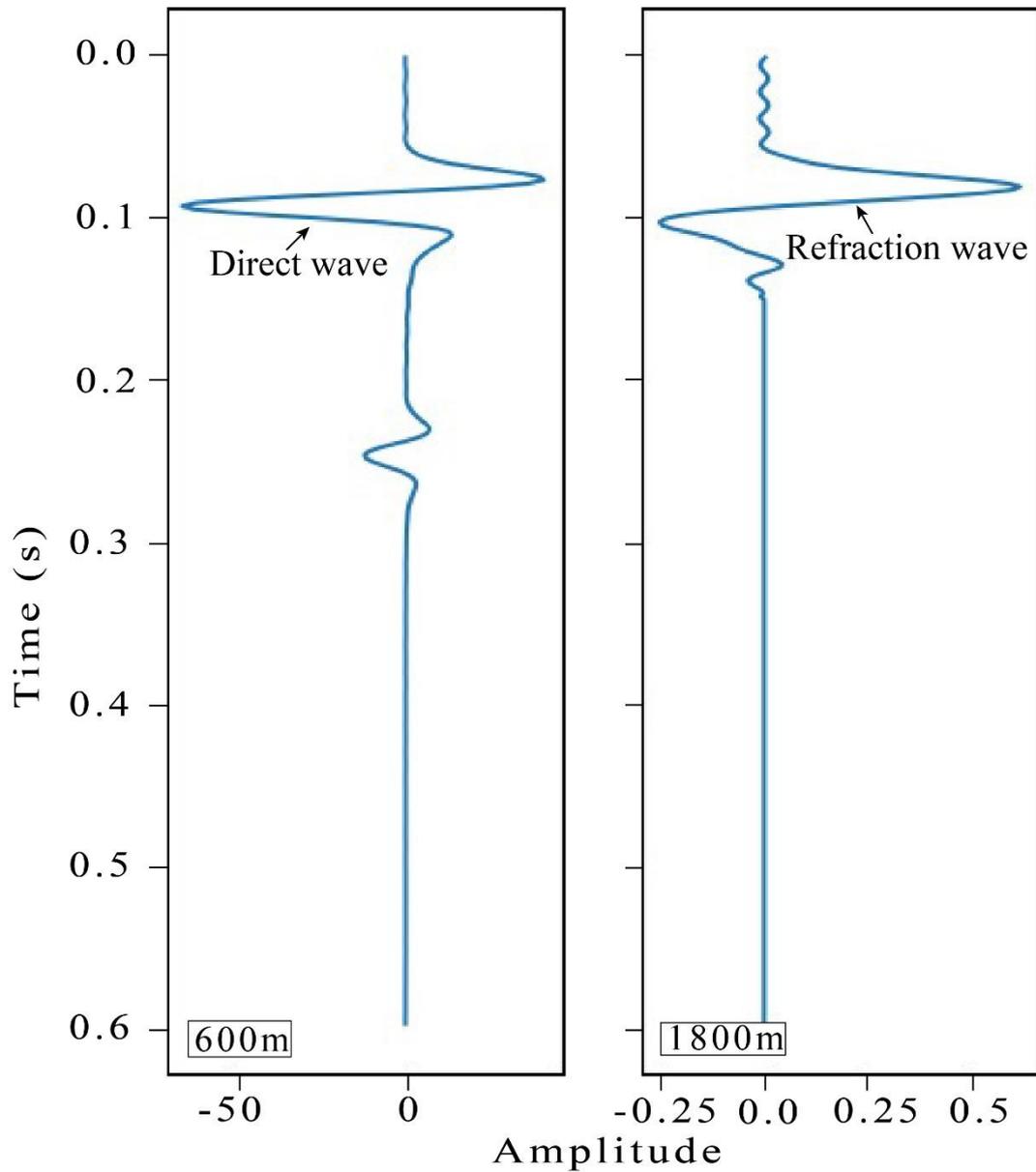
**Figure 4-4:** Original shot gather.



**Figure 4.5:** Seismic records at 600 m (left) and 1800 m (right).



(a)



(b)

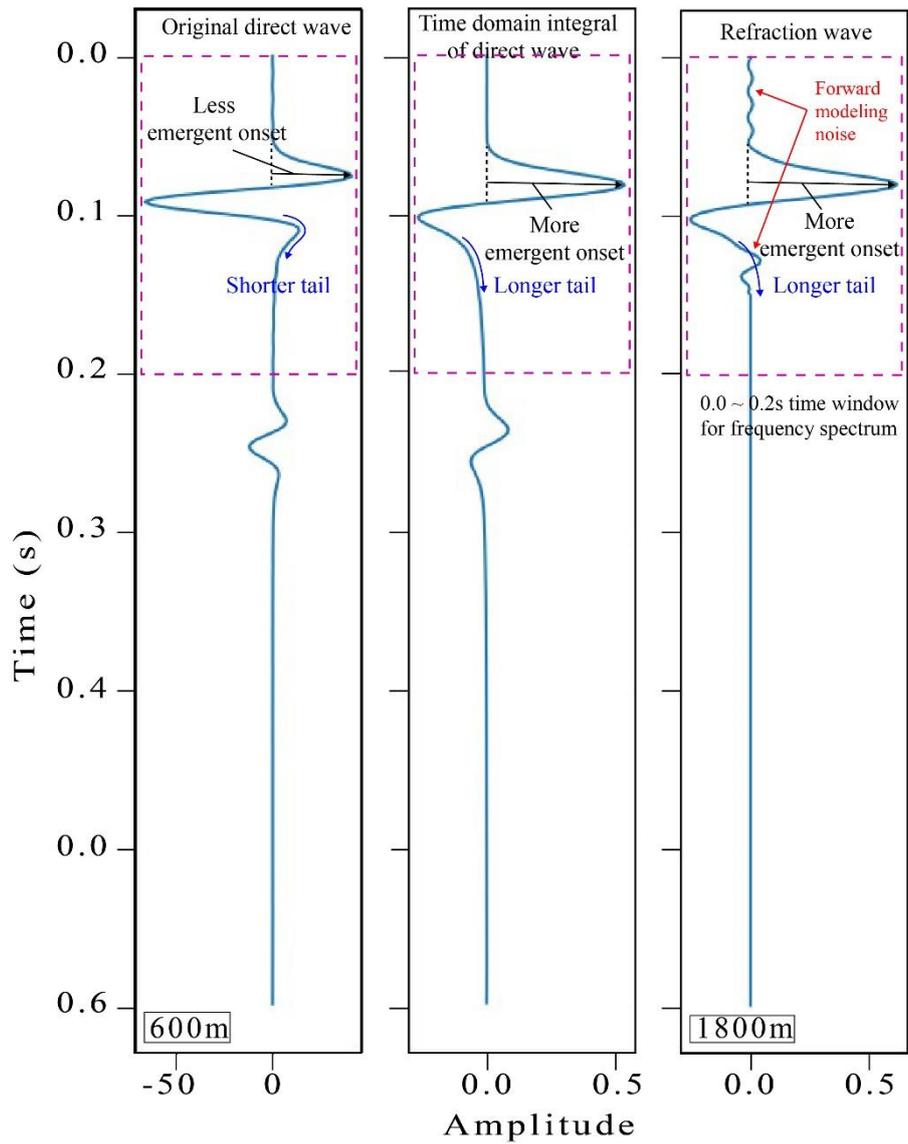
**Figure 4-6:** Direct (left column) and refraction waveform (right column) were (a) extracted from time window of 0.0 to 0.6s, and (b) aligned by emergent onset point.

As mentioned in Aki and Richards (2002) of box 6.4, the head wave possessed the same amplitude spectrum and waveform as original source wavelet, while the refraction waveform is time integral of source wavelet. In our case, the head wave consists of both of direct and

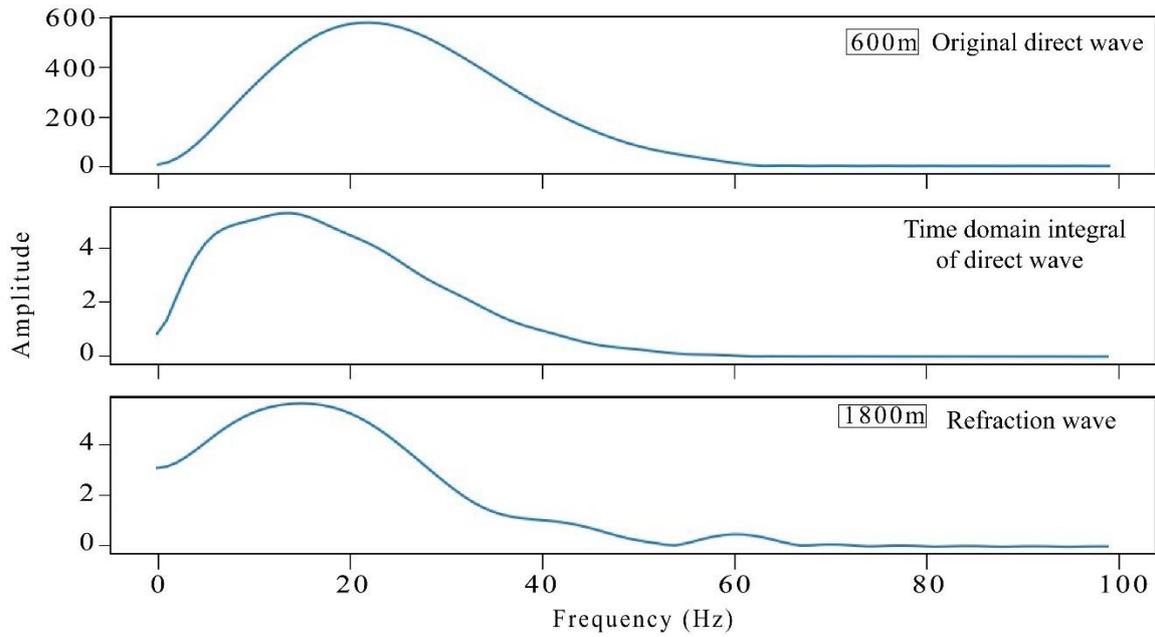
---

turning waves. In such fully acoustic synthetic model test, the waveform of direct wave and its amplitude spectrum should be identical to the original source wavelet, which is a 25 Hz Ricker wavelet. On the other hand, the waveform and amplitude spectrum of time domain integral of direct wave should be identical to those of refracted wave. Such conclusion is supported by the evidences observed from Figure 4.7 and 4.8.

In figure 8, the original direct wave appeared to be less emergent onset and shorter tail than both of its time domain integral and refraction waveforms, as labeled by black and blue arrow respectively. The waveform of time integral of direct wave in the middle column is similar to the one from right column of refracted wave, excluding the artificial noise raised by forward modeling as indicated by red arrows.



**Figure 4-7:** The comparison of seismic waveforms of original direct wave (left), time domain integral of original direct wave (middle) and refracted wave (right).



**Figure 4-8:** The comparison of amplitude spectrums of original direct wave (top), time domain integral of original direct wave (middle) and refracted wave (bottom).

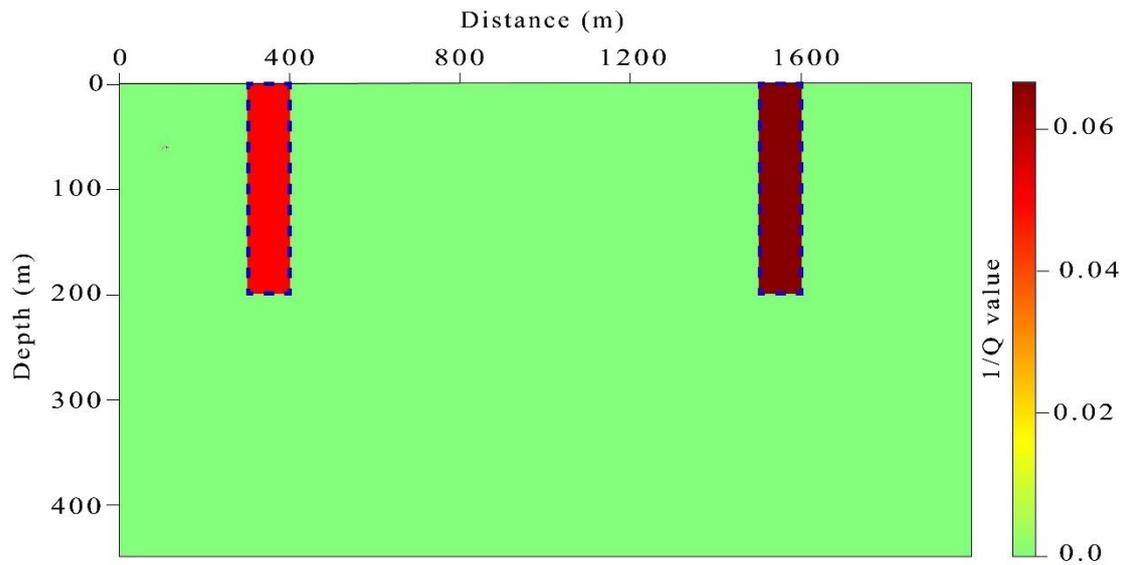
Then their amplitude spectrums were compared in Figure 4.8 within a time window from 0.0 to 0.2 s as indicated by magnet dash line box from Figure 4.7. From Figure 4.8, we could see the original direct wave preserves the amplitude spectrum of source wavelet with central frequency at 25 Hz. Meanwhile, both amplitude spectrum of its time domain integral and refraction appeared to have low peak central frequency shift toward around 18 Hz, and they were almost identical. The minor amplitude distortion of time domain integral around peak frequency was caused by the 0.2s time window truncation in time domain.

One should note that such observation is from a fully elastic model without introducing  $Q$  attenuation, and difference between the amplitude spectrum of direct wave and refraction had already shown up. Such difference is becoming more significant in consideration of  $Q$  attenuation, and eventually could be utilized to help us on estimation for the attenuation time

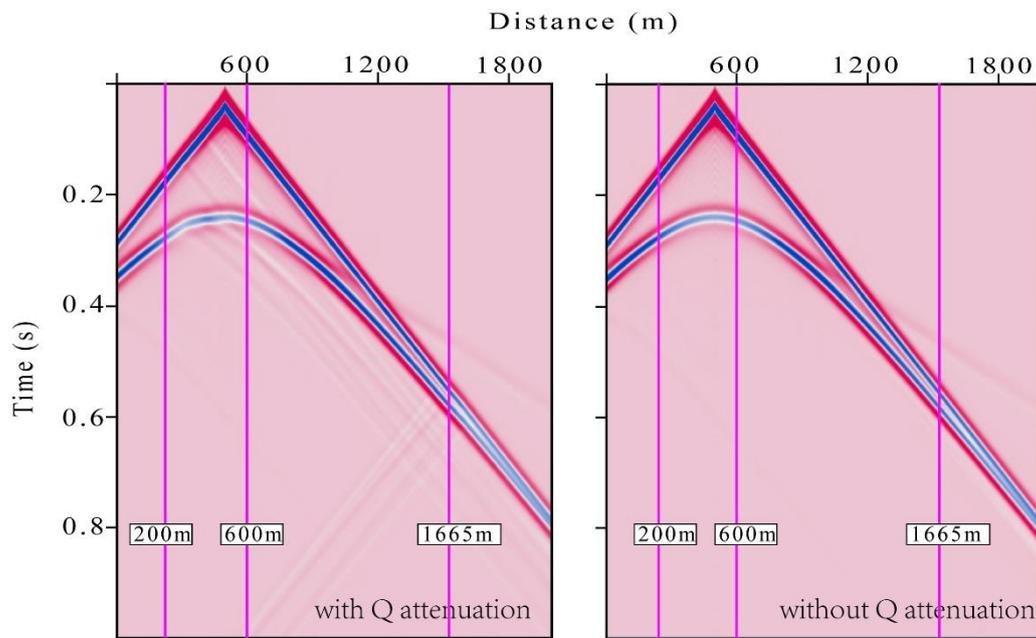
---

$t^*$ , as presented in the following test.

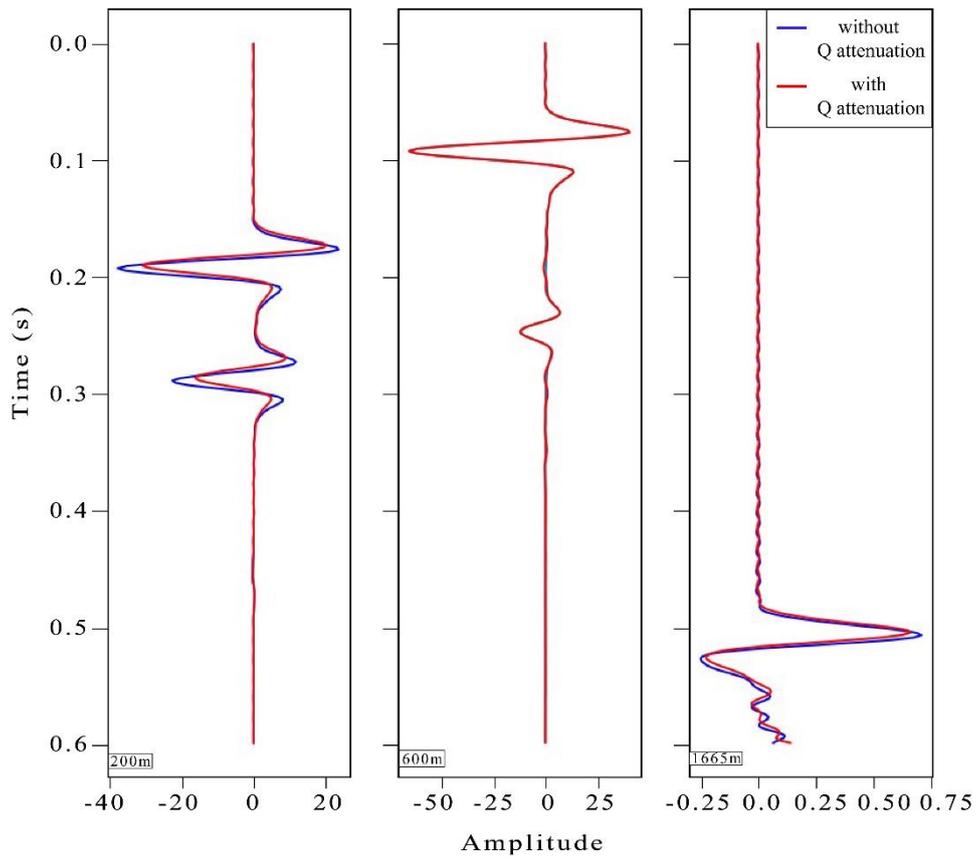
Here we use the same velocity model as shown in Figure 4.3, while introducing two attenuation anomalies as shown in Figure 4.9. The Q value of one attenuation anomaly is 20 and located within 300 to 400 m distance along the surface, with thickness of 200m. Another one is Q value of 15 within 1500 to 1600 m distance along the surface with same thickness. The blue dash line box depicted the boundaries of these two anomalies. The forward modeling method we used in this test is based on frequency domain finite difference visco-acoustic equation. The source wavelet is exact same ricker wavelet as previous test at 25 Hz. Correspondingly, a synthetic shot gather (source at 500 m) obtained from this Q model is shown in the left column in Figure 4.10, and the right column is the forward modeling result without Q attenuation. Three seismic traces at 200, 600 and 1665 m are extracted from these two shot gathers and labeled as blue and red line to indicate with and without Q attenuation in Figure 4.9. Some minor phase shift (delay) and amplitude distortion could be observed in blue seismic traces with Q attenuation.



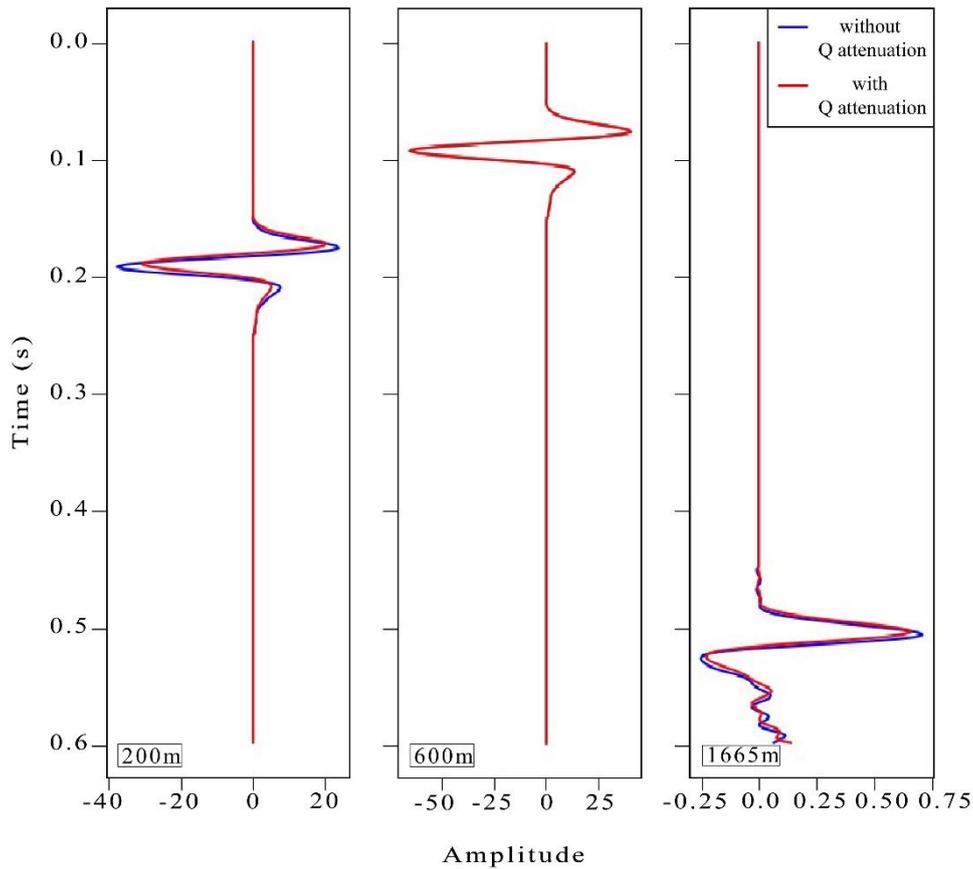
**Figure 4-9:** True Q model.



**Figure 4-10:** Comparison of shot gathers (source at 500 m) with Q attenuation (left) and without Q attenuation (right).



**Figure 4-11:** Comparison of seismic traces at 200 m (left), 600 m (middle) and 1665 m (right). The blue and red line indicate seismic trace without Q attenuation and with Q attenuation respectively.



**Figure 4-12:** First arrivals of seismic trace at 200 m (left), 600 m (middle) and 1665 m (right). The blue and red line indicate seismic trace without Q attenuation and with Q attenuation respectively.

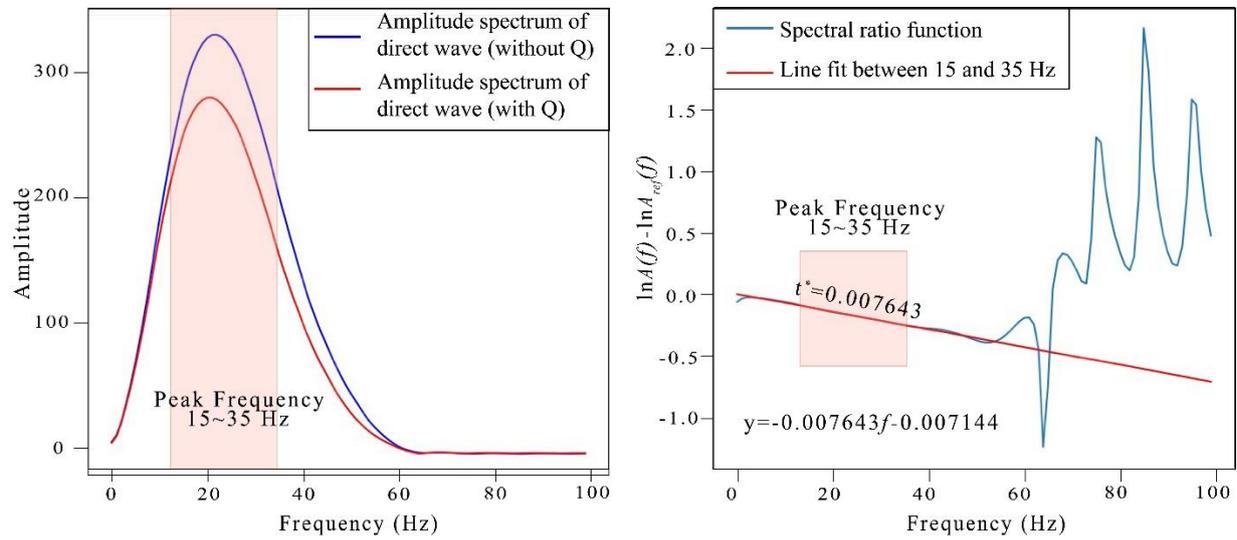
The corresponding first arrival is cut out from each seismic trace as shown in Figure 4.12.

Then the spectral ratio method was applied to estimate  $t^*$  in the following section.

## **2. Modified logarithm spectral ratio method**

Firstly, the spectral ratio method was applied on direct wave with Q attenuation from red seismic trace in left column of Figure 4.12. Here we choose the dominant frequency contents between 15 Hz and 35 Hz for linear fit. The spectral ratio result was shown as blue line in right column of Figure 4.13, and the red line indicates the linear fit result, indicating the estimated  $t^*$  from spectral ratio method is 0.007643 s. The true  $t^*$  could be derived from equation (4.2),

and its value is 0.007853 s ( $Q = 20$ , velocity = 2000m/s and travel distance = 100 m). We could see the estimated  $t^*$  obtained from spectral ratio method is very close to the true one, and the minor errors are caused by forward modeling, which is also acceptable. Instead of apparent  $Q$  value, the spectral ratio method was only implemented for obtaining the estimated  $t^*$ , which is the cumulative effect of  $Q$  anomalous body on the whole propagation process of seismic wave through subsurface media. The following tomographic inversion was conducted for deriving out the spatial  $Q$  value along the raypath based on  $t^*$ .



**Figure 4-13:** Amplitude spectrum of direct wave with and without  $Q$  attenuation (left) and spectral ratio method result (right).

The previous section is the case of only direct wave involved in the first arrival waveform trains, and the logarithm function for spectral ratio method without considering frequency-dependent propagation response  $P_p(f)$  can be expressed as:

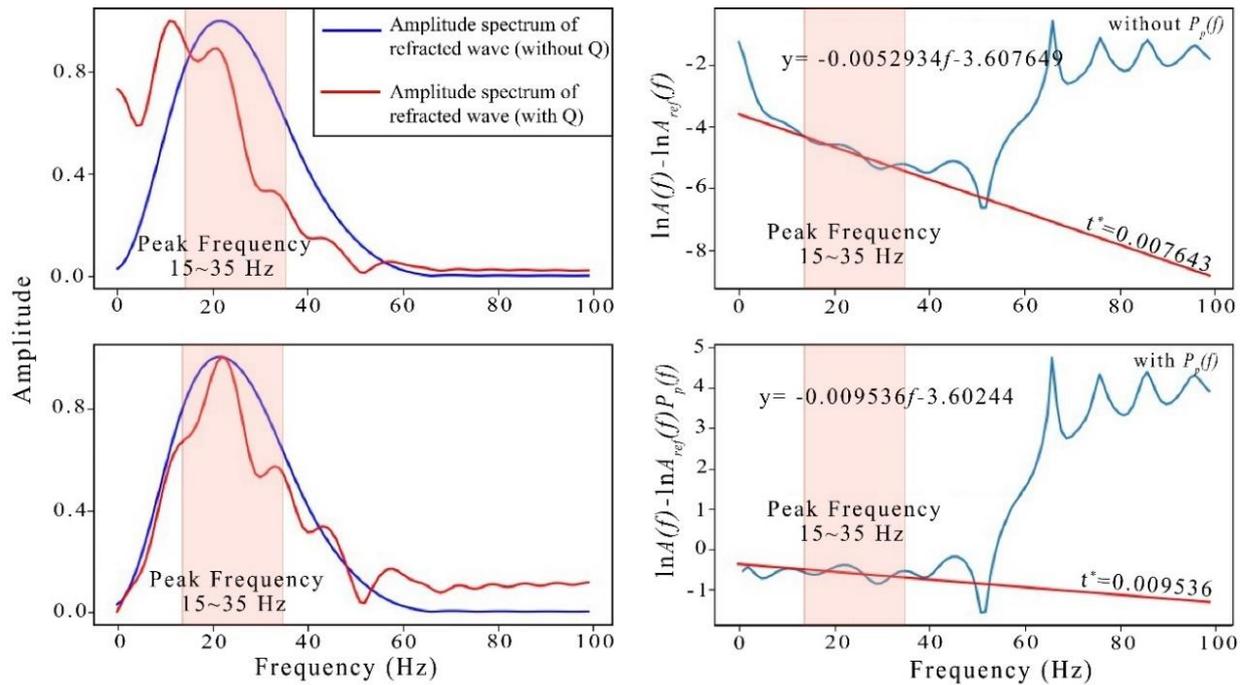
$$\ln \frac{A(f)}{A_{ref}(f)} = -ft^* + \ln P_s . \quad (4.20)$$

The equation (4.20) works well when its straightly applied in direct wave case, since that

---

the frequency response of original source wavelet will be preserved for direct wave as presented in section 4.1.3 (Figure 4-2). However, it will fail in the case of refraction without considering its wave type frequency response  $P_p(f)$ . In the following part of test, refraction will be introduced and the performance of with or without taking  $P_p(f)$  into account for estimating  $t^*$  will be compared.

Similarly, the true  $t^*$  could be calculated as shown in previous section with utilizing propagation mechanism of refraction, and is 0.01047 s. The spectral ratio method result is shown in the right column of Figure 4.14. The top right one is the result without considering  $P_p(f)$ , and estimated  $t^*$  is 0.007643 s, which is far from the true  $t^*$ . The bottom right figure shows the result considering  $P_p(f)$  using logarithm spectral ratio from Equation (4.18), and its estimated  $t^*$  is equal to 0.009536 s, which is very close to the true  $t^*$ . It is indicating that the accuracy of estimated  $t^*$  by spectral ratio method will be improved after introducing correct  $P_p(f)$  for certain wave type.



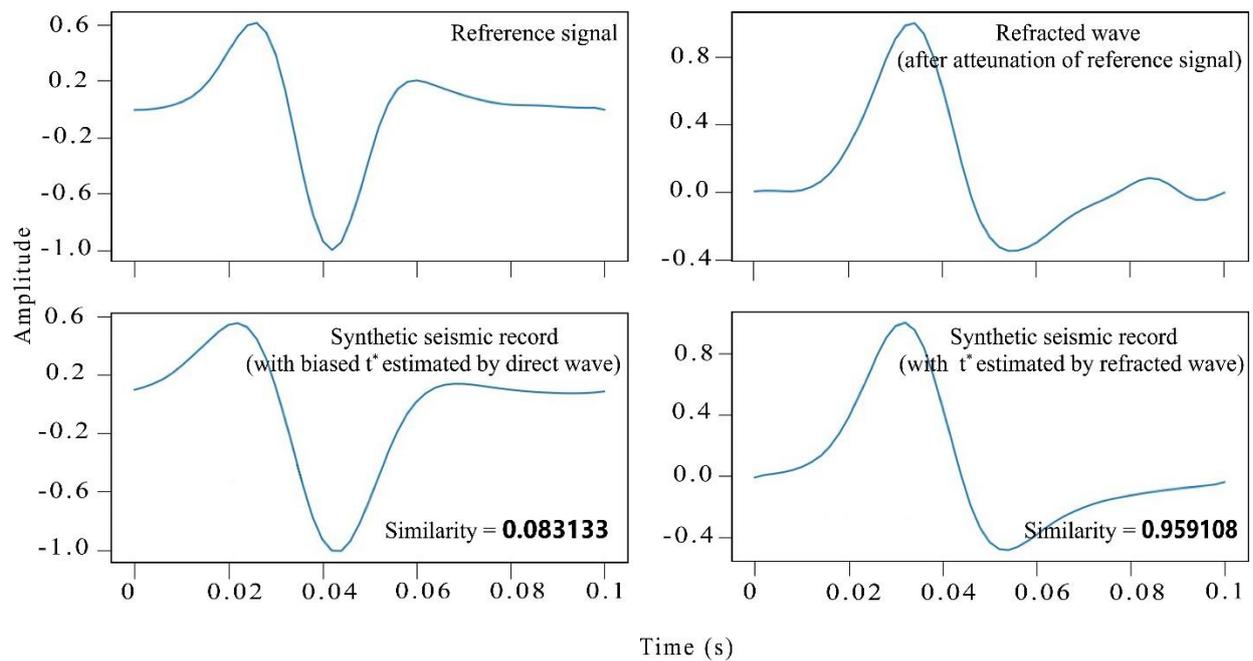
**Figure 4-14:** Amplitude spectrum of direct wave with and without Q attenuation (left column) and spectral ratio method result without  $P_p(f)$  (top right) and with  $P_p(f)$  (bottom right).

### **3. Adaptive correction method – determination of estimated $t^*$**

Now the question is how to determine estimated  $t^*$  derived from either direct or refracted wave for specific cases, since these two types of waveforms in the first arrival waveform trains cannot be distinguished automatically. To address this issue, we applied both estimated  $t^*$  to the reference signal (source wavelet) to generate two synthetic seismic traces and calculated their similarities with original seismic traces. Eventually, the one with higher similarity value will be selected as final estimated  $t^*$ .

Figure 4.15 showed one example of how the adaptive correction method worked. The top left figure is the original reference signal and the top right one is refracted wave after attenuation of reference signal. The bottom left is synthetic seismic record generated based on

estimated  $t^*$  assuming only direct wave containing in first arrival waveform trains. The bottom right one is result generated based on estimated  $t^*$  accounting for  $P_p(f)$  as refracted wave involved in first arrival waveform trains. Form their final similarity results, we could see that the refracted wave assumption will be better fit for this case (similarity = 0.959108), and its wave propagation factor  $P_p(f)$  could increase the accuracy of  $t^*$  estimation.



**Figure 4-15:** Reference signals (top row) and similarity of their synthetic seismic trace (bottom row).

#### **4. Interference of different wave types**

Practically, the first arrival is mostly stable maintain relatively higher SNR than other secondary wave types of raw seismic data. Still, it is required a cautious pre-process to suppress the contaminated noise while properly preserving the salient amplitude information of signal. Apart from noise contamination, interference of different wave types would also have significant impact on  $t^*$  estimation by spectral ratio method. To suppress the influence raised

---

from interference of different wave types or events, we will follow resolutions below in real data application:

a. Noise removed by waveform similarity judgement:

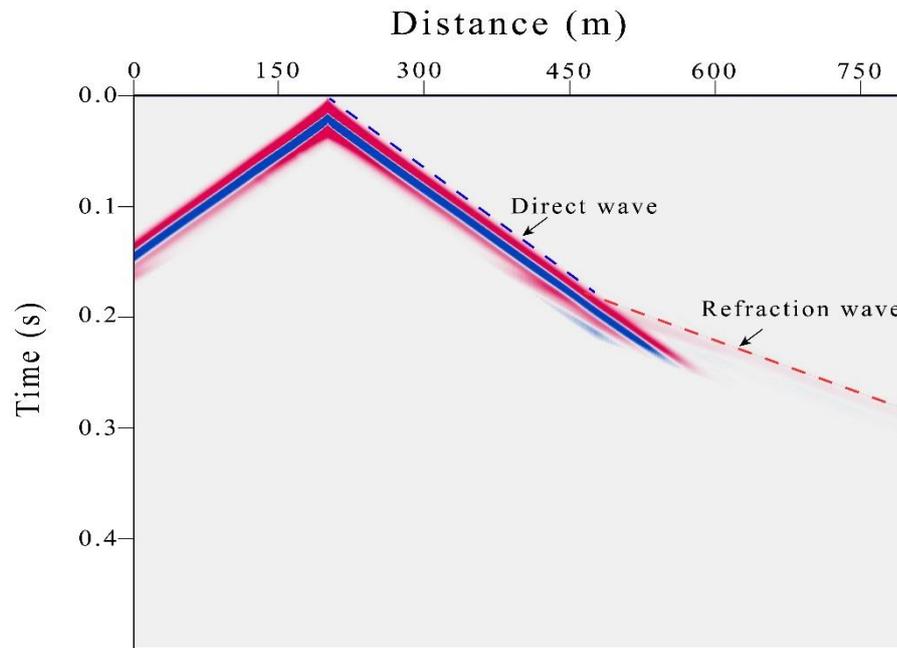
In the case of the energy level of the interference signal is quite close to that of the first arrivals (very low SNR), even the seismic waveform could be distorted severely. Such distortion could lead the original waveform to be much different from either direct wave or refracted wave. Then we could calculate its waveform similarity as described in section 4 to distinguish useful signal from noise.

Here we assume that the similarity between recorded seismic trace and reference signal (statistical source wavelet) is  $a$ . And the similarity between recorded seismic trace and the time integral of reference signal (similar to direct wave) is  $b$ . The manual threshold  $c$  here is set to be small enough, and the threshold of similarity difference is  $d$ . Then the recorded seismic trace should satisfy the following condition to be further utilized for estimating  $t^*$ :  $a > c \wedge b > c \wedge \text{abs}(a - b) > d$ .

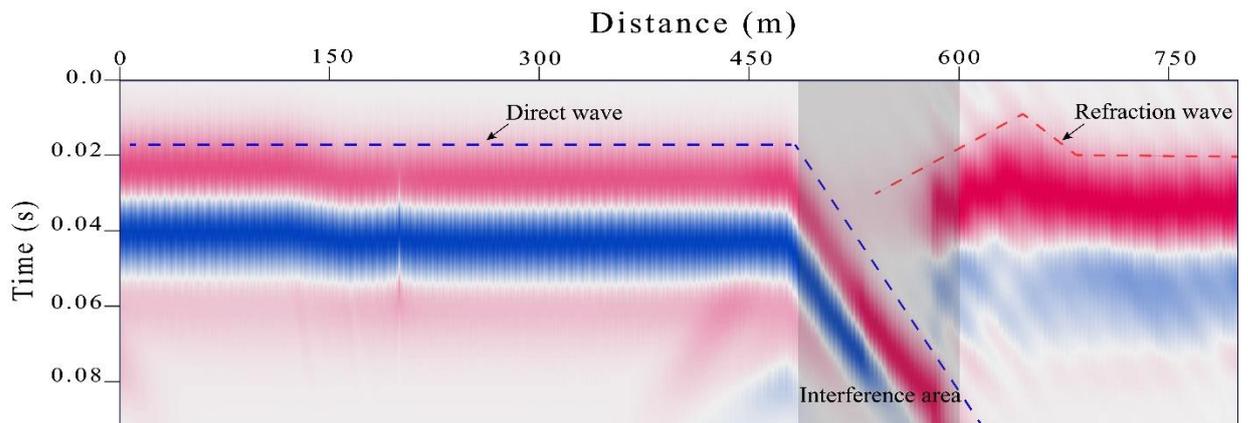
These series of conditions could guarantee that the noise contaminated signal is still available for estimating relatively reliable  $t^*$ . Those who do not meet the conditions will be removed during the process due to the redundancy of seismic data, especially in first arrival tomographic inversion.

Here is one example of how we handle with different wave types interference. The

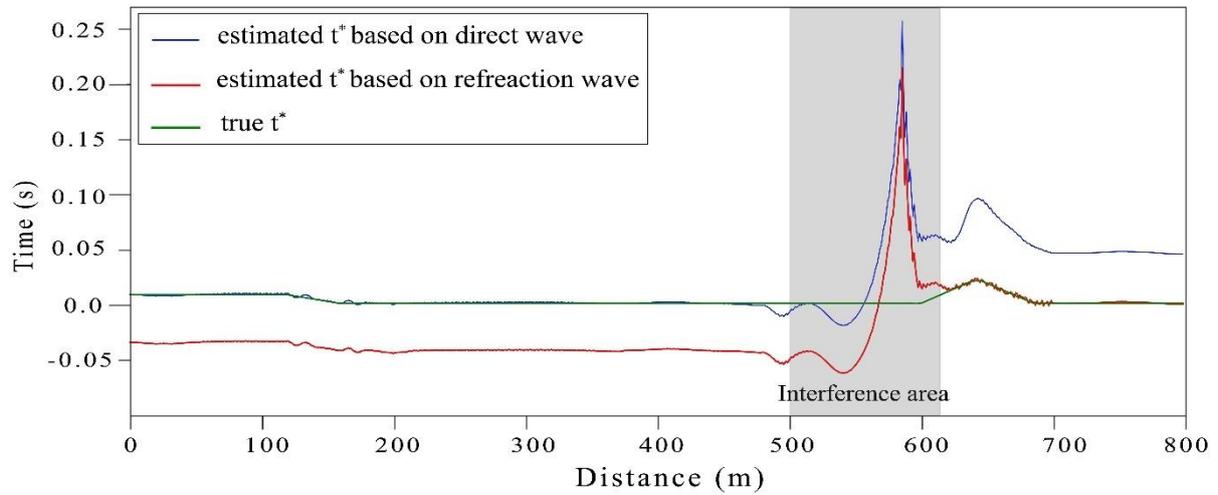
velocity and Q model is exactly same as used in section 4.1.1 and 4.4.2. The original shot gather is presented in Figure 4.16, and the direct wave and refraction wave was interfered around the horizontal distance of 450m. The first arrival is cut out and flattened out with amplitude normalization of each channel, as shown in Figure 4.17.



**Figure 4-16:** Original shot gather (interference of direct and refracted wave).

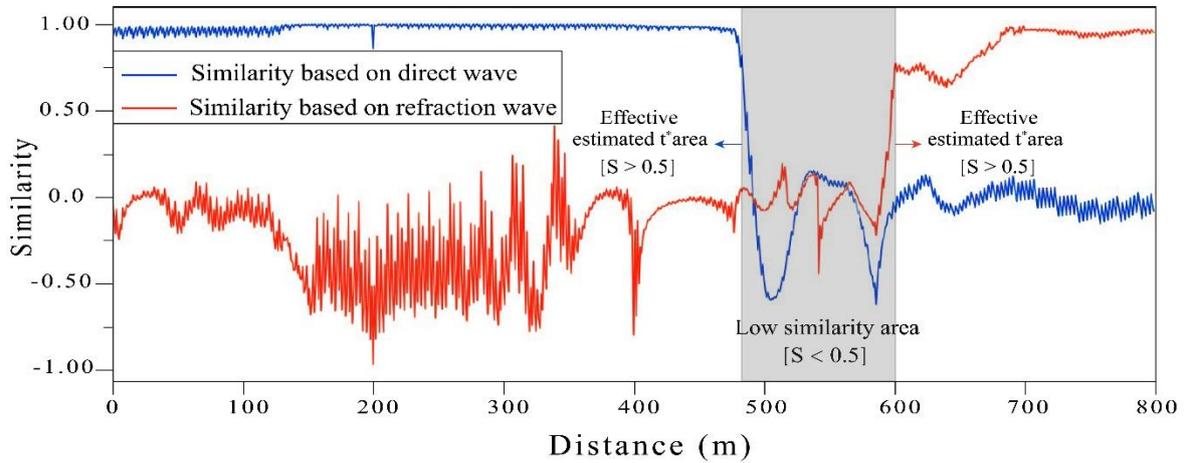


**Figure 4-17:** Normalized first arrival fatten out.



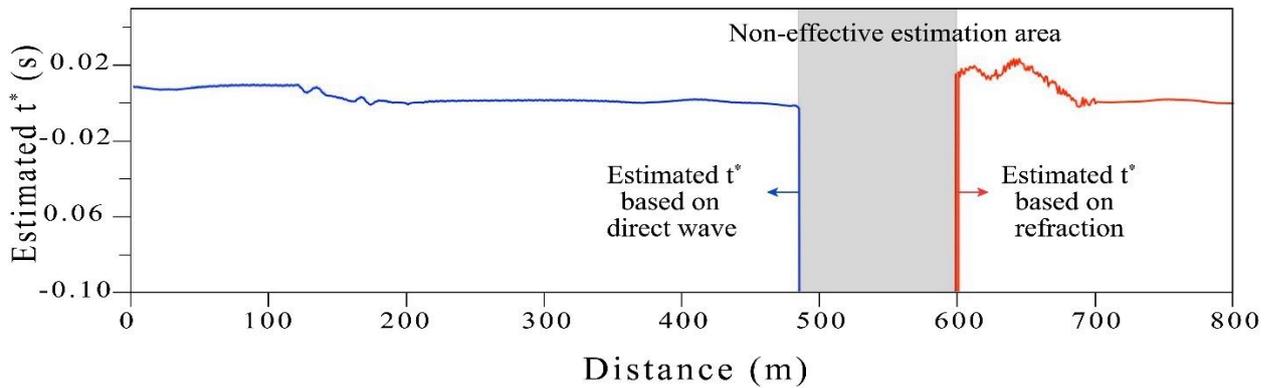
**Figure 4-18:** Estimated  $t^*$  based on direct wave and refraction.

From Figure 4.16, one can observed that the first arrival is only dominated by direct wave before interference area (around 550 m) and refracted wave after interference area. Since then, the wave propagation effect should come from direct wave and refracted wave before and after interference area, respectively. The results of estimated  $t^*$  confirmed such observation as shown in Figure 4.18. We could see that the estimated  $t^*$  based on direct wave (blue line) coincided quite well with true  $t^*$  (blue line) before interference area, and then the one based refracted wave (red line) is becoming more fit after interference area.



**Figure 4-19:** Similarities calculated based on direct wave and refraction.

This phenomenon could also be observed in similarity results as shown in Figure 4.19. Again, the low similarity area (similarity  $< 0.5$ ) is within the zone of interference. The direct and refracted wave preserve good similarities (similarity  $> 0.5$ ) before and after interference area, meaning that the corresponding estimated  $t^*$  are reliable.



**Figure 4-20:** Final estimated  $t^*$ .

Eventually, the final estimated  $t^*$  is shown in Figure 4.20. The non-effective estimation area is corresponding to interference area, and the estimated  $t^*$  in this area is negative, which

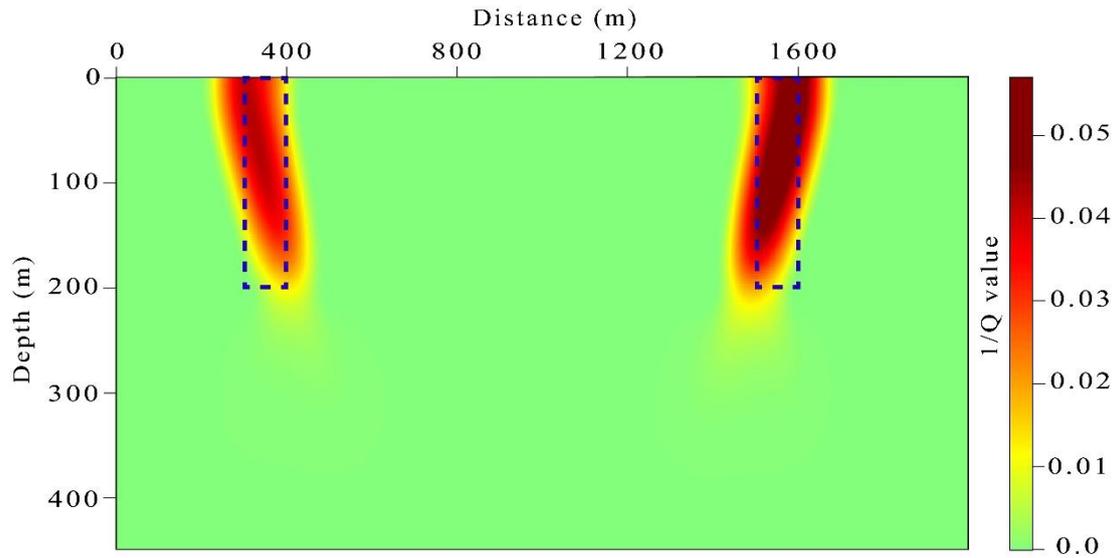
---

will be abandoned from this area to ensure the reliability of the results. Through this process, most of influence caused by waveform interference could be avoided.

b. Re-check the misfit of true  $t^*$  and forward modeling  $t^*$  in late stage of tomographic inversion iteration:

For other small amounts of waveform interference or noise cannot be removed by similarity judgement, we will re-check the  $t^*_{cal}$  obtained from forward modeling according to Q model of current iteration in the late stage of inversion, and compared it with estimated  $t^*_{est}$ . The estimated  $t^*_{est}$  leading to misfit larger than certain threshold (determined by current data residuals) will be treated as abnormal  $t^*_{est}$  and excluded in the following inversion process. It is quite intuitive to recognize that such abnormal value cannot be corrected through inversion process, since that it is not conformed with the principle of seismic wave propagation.

Based on these two treatments to interference signal, we obtained the final Q model as shown in figure 19. Both of inverted attenuation anomalous bodies were slightly deflected along the main raypath direction due to the simplex raypath contents in this simple model, which is a normal phenomenon in such ray-based tomographic inversion.

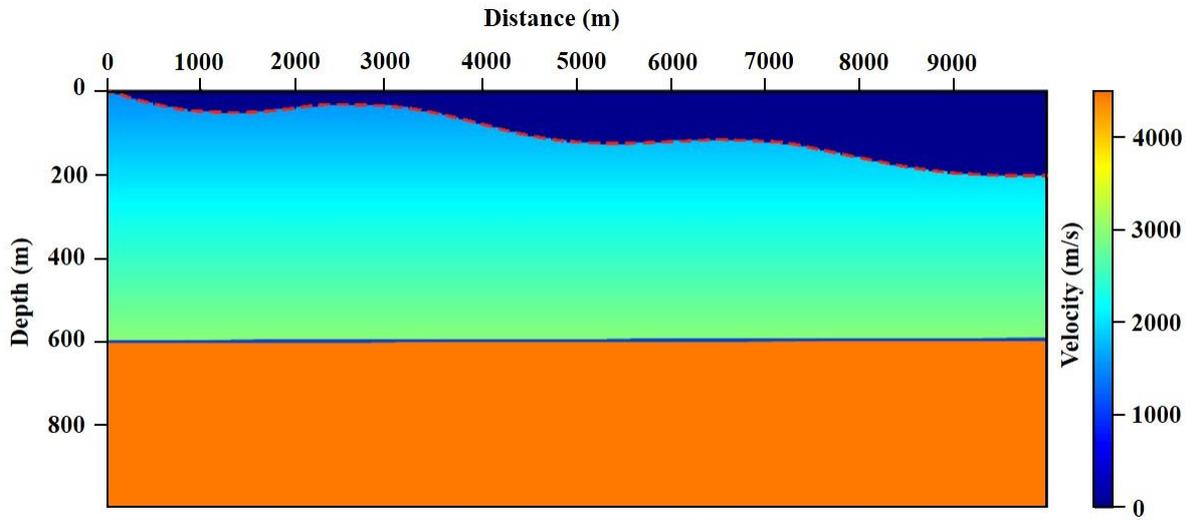


**Figure 4-21:** Inverted 1/Q distribution model

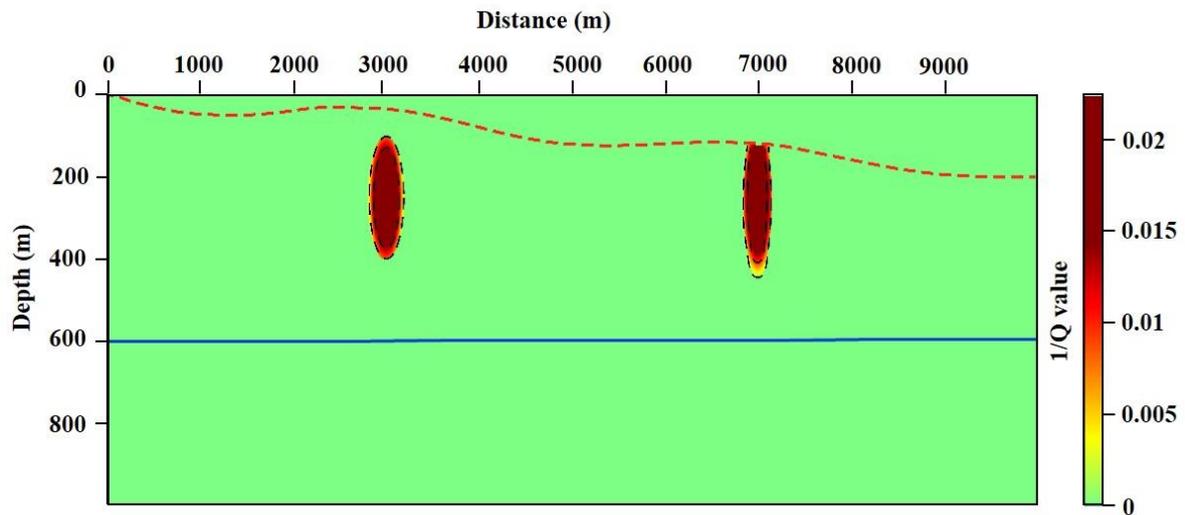
#### 4.3.2 2-D synthetic data test: simple model with topographic

A 2-D numerical seismic survey is simulated over a 10000m long by 1000m deep model comprising two layers (Figure 4-22). The velocity of top layer is linearly increasing with depth from 1500m/s to 3000m/s. The bottom layer becomes homogeneous with constant velocity of 4500m/s. The corresponding attenuation model is shown in Figure 4-23. The background attenuation values are zero, and two attenuation anomalies located beneath the surface. The shapes of anomalies are both ellipse with semi-major length of 200m and semi-minor axis length of 150m. The attenuation value is 0.02 ( $Q = 50$ ) at the center of the anomaly and reduces linearly to zero from center. total of 490 sources and 2000 receivers are deployed along the topographic surfaces with source and receiver intervals of 20 and 5m, respectively. The synthetic data was generated by a visco-acoustic wave equation finite difference solver, which is based on standard linear solid method (Carcione *et al.* 1998). The source wavelet is a Ricker

wavelet with 15Hz dominant frequency.



**Figure 4-22:** A 2-D synthetic data test: the true velocity model.

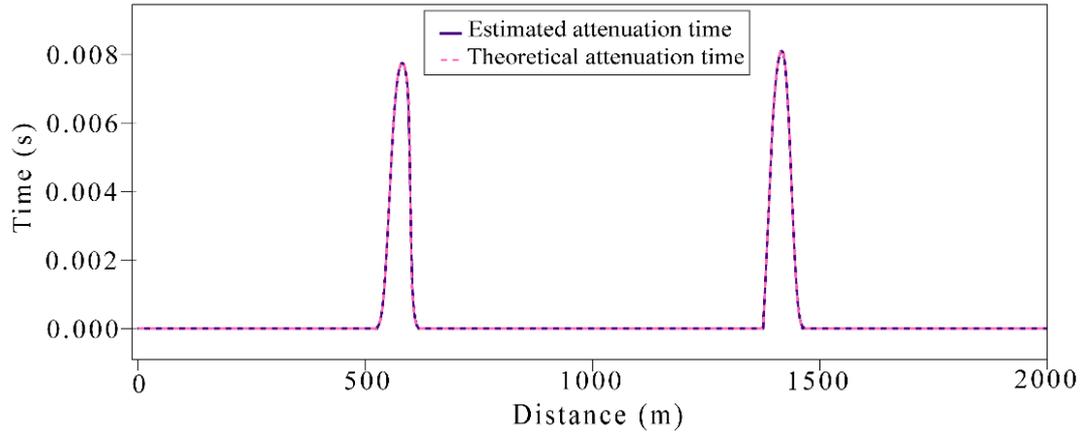


**Figure 4-23:** The true attenuation model ( $1/Q$ ).

We compared the attenuated traveltimes estimated by our proposed method with the theoretical attenuated traveltimes generated by solving equation (4-5) as shown in Figure 4-24. The almost identical result indicates the capability and validity of estimating attenuated traveltimes by our proposed method in such noise-free data. In practice, we will not use all

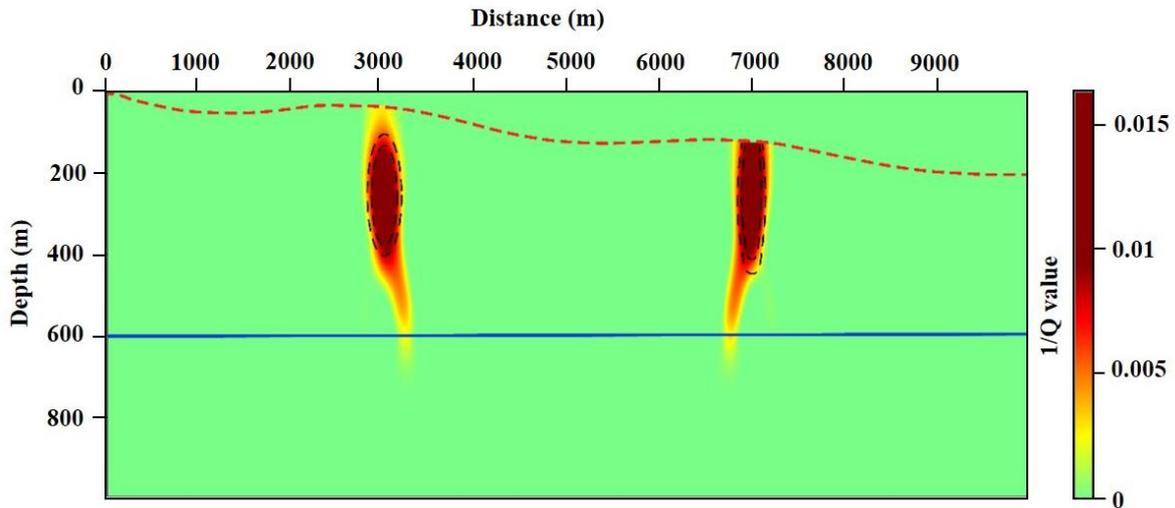
---

estimated attenuated traveltimes for following attenuation tomography. The amplitude spectrums of seismic traces will be affected by the superpositions of both turning waves and refraction. In such case, the corresponding similarity coefficients are quite small to lead zero weights in Q tomographic inversion.

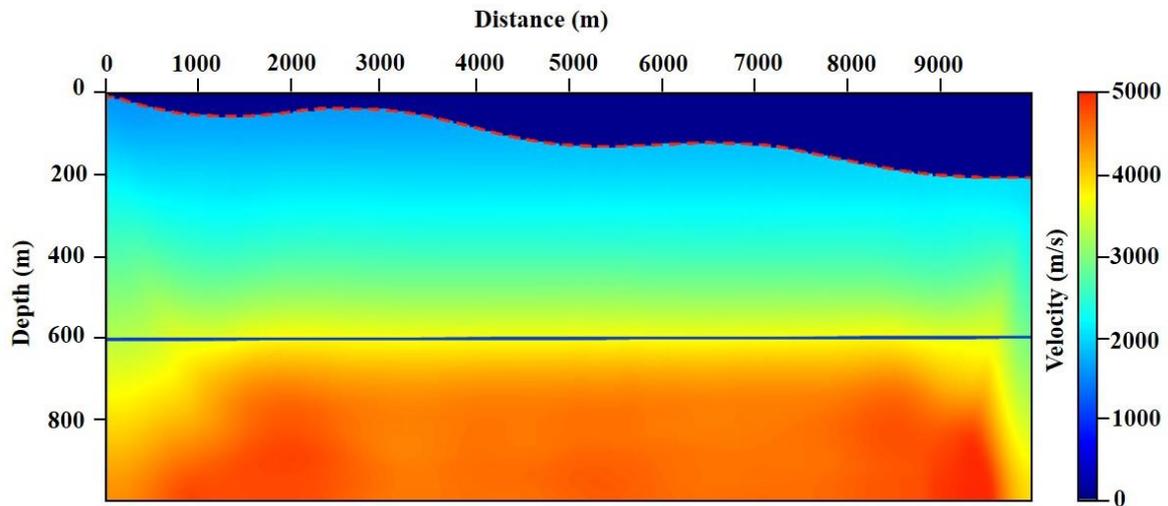


**Figure 4-24:** Comparison of estimated  $t^*$  and theoretical  $t^*$ .

We first use the true velocity to perform Q tomography. The inverted Q model using true velocity is shown in Figure 4-25. Compared with the true attenuation model (Figure 4-23), we can find the inverted Q model is almost same with true one except smearing “tails” below the anomalous bodies. Such artifacts are mainly caused by the limited acquisition and non-uniform directions of ray path near the high velocity layer.



**Figure 4-25:** Inverted Q model using the true velocity model.



**Figure 4-26:** Inverted velocity model using first arrival tomography.

We then use the inverted velocity model generated by travelt ime tomography (Figure 4-26) to perform Q tomography. The inverted velocity model is structurally consistent with true model but has a lower resolution. Velocities at the left and right boundaries is less reliable because of limited ray coverage. The attenuation model inverted by Q tomography using the inverted velocities is shown in Figure 4-27. The shape of the two attenuation anomalies is almost same as the one shown in Figure 4-25 except the appearance of insignificant “tails”

artifacts caused by the inaccurate velocity model. It is evident that the near-surface Q anomalies could be successfully reconstructed by our proposed two-steps first arrival Q tomography.

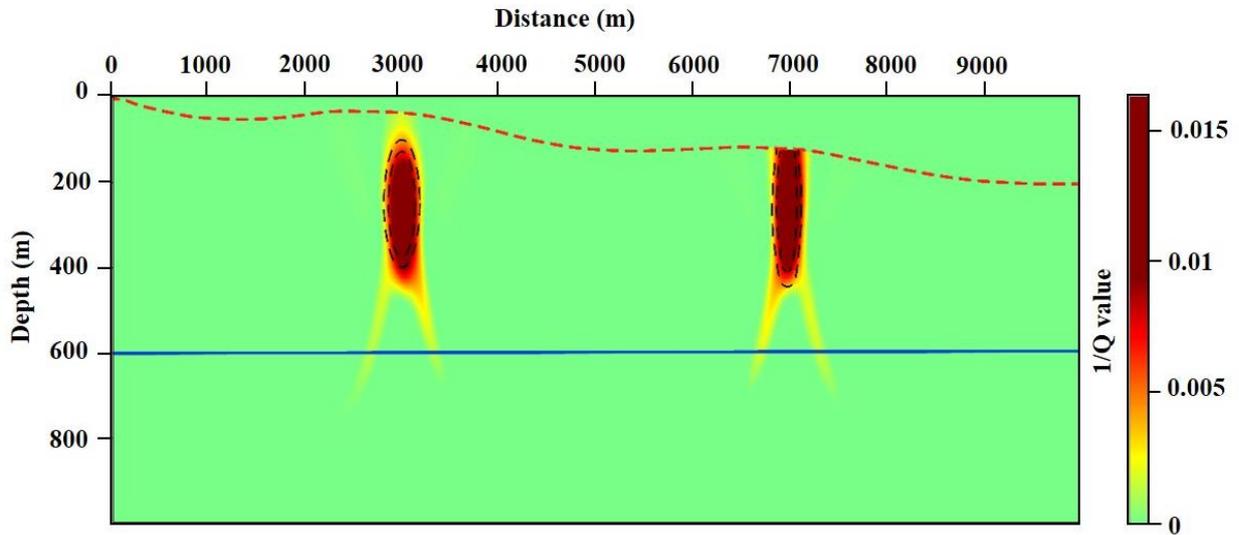


Figure 4-27: Inverted Q model derived by Q tomographic inversion using inverted velocity model.

#### 4.3.3 2-D synthetic data test: Marmousi model

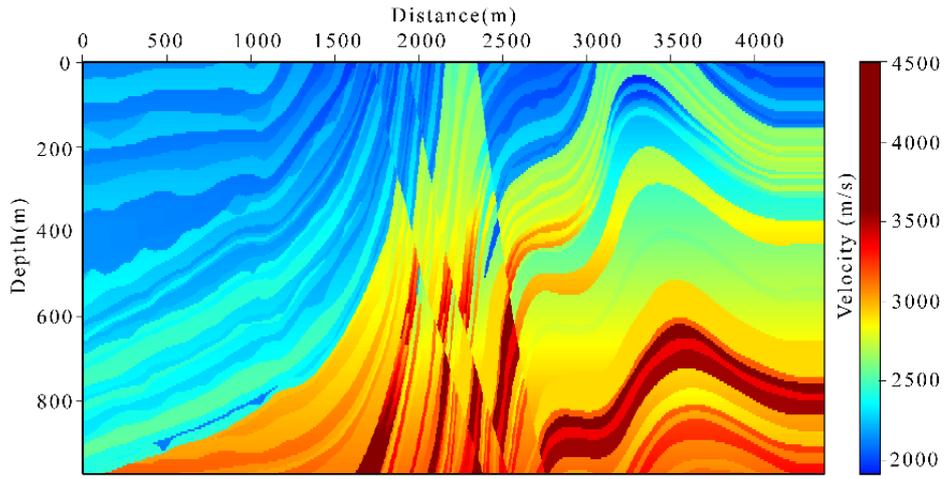
The first arrival Q tomographic inversion based on adjoint-state method is tested on Marmousi model with strong heterogeneous presence (Figure 4-28a). The model is 4500m wide and 1000m deep, and we use a fix-spread acquisition geometry on the surface with 221 sources and 884 receivers of intervals 20 and 5m, respectively. The source wavelet is same as in the previous simple 2-D model experiment, whose dominant frequency is 20Hz. The corresponding attenuation model is presented in Figure 4-28b, and it shows a relatively high attenuated trends in left-side region comparing to right-side.

As shown in Figure 4-28d, the inverted attenuation model estimated by the true velocity model produces a good Q tomographic inversion result in near surface, ranged from top surface

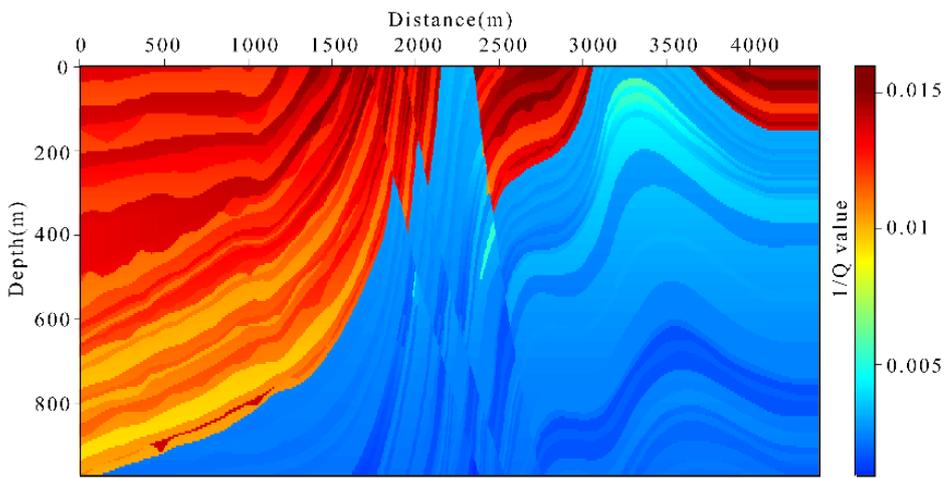
---

down to around depth of 300m. It preserves a similar trend comparing to the true answer from Figure 4-28b, and even some detailed features of attenuated anomalies are slightly reconstructed, e.g., in between distance of 2400 and 3000m. Relevantly, the result using inverted velocity model (Figure 4-28c) shows a similar near surface attenuation distribution as presented in Figure 4-28e. However, the result resolution is degraded as losing some edge and detailed feature depictions comparing to Figure 4-28d. It indicates that the accuracy of inverted model we used for Q tomography has impacts on the quality of inverted attenuation model by our proposed method. A precise velocity model could bring benefit to both of attenuated traveltimes estimation and Q tomographic inversion.

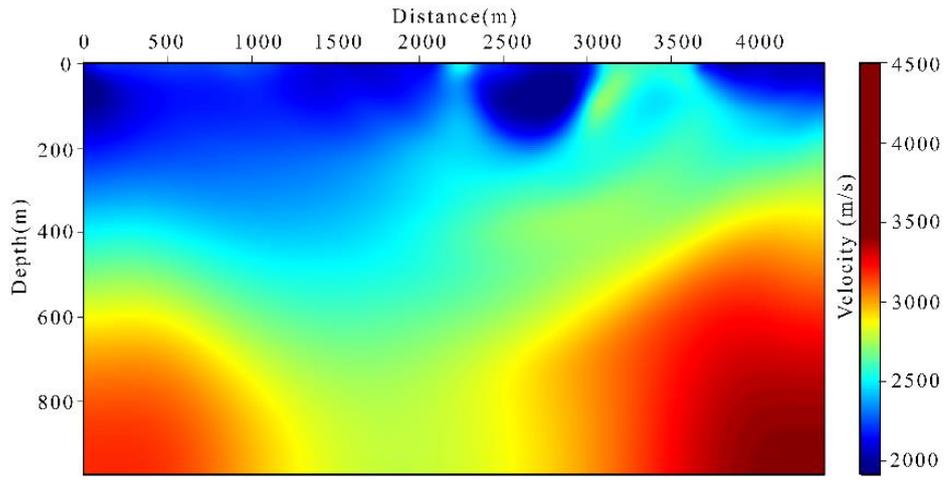
Have limited offset, we could only retrieve a reasonable near-surface attenuation structure based on first arrival Q tomography. The ray coverage on the edges of model is poor and becomes worse with increasing depth, leading to unreliable inverted results. This could be one limitation of our first arrival Q tomography. Other than that, there is good agreement true and inverted attenuation model through the near surface profile. This figure indicates that our tomographic methods is capable of accommodating such strong heterogenous.



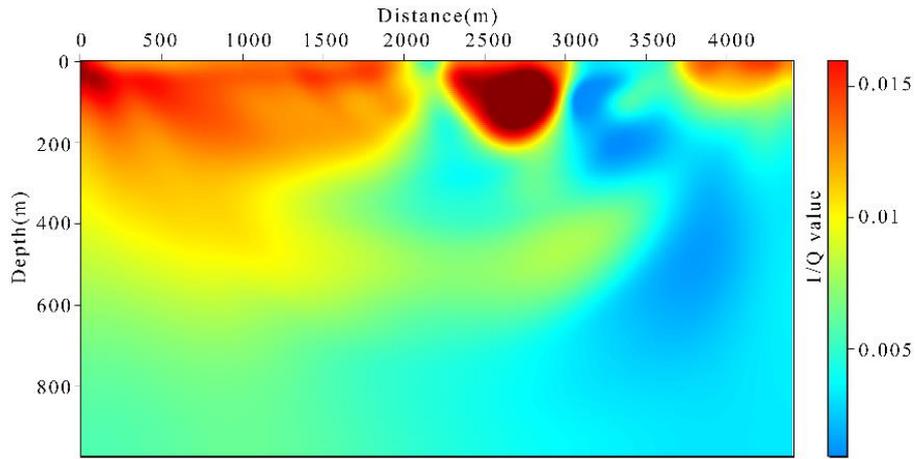
(a)



(b)



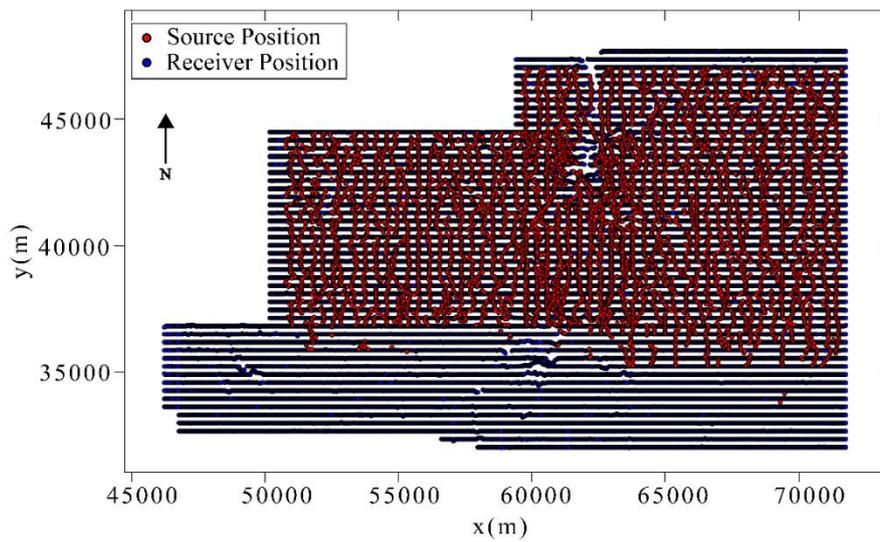
(c)



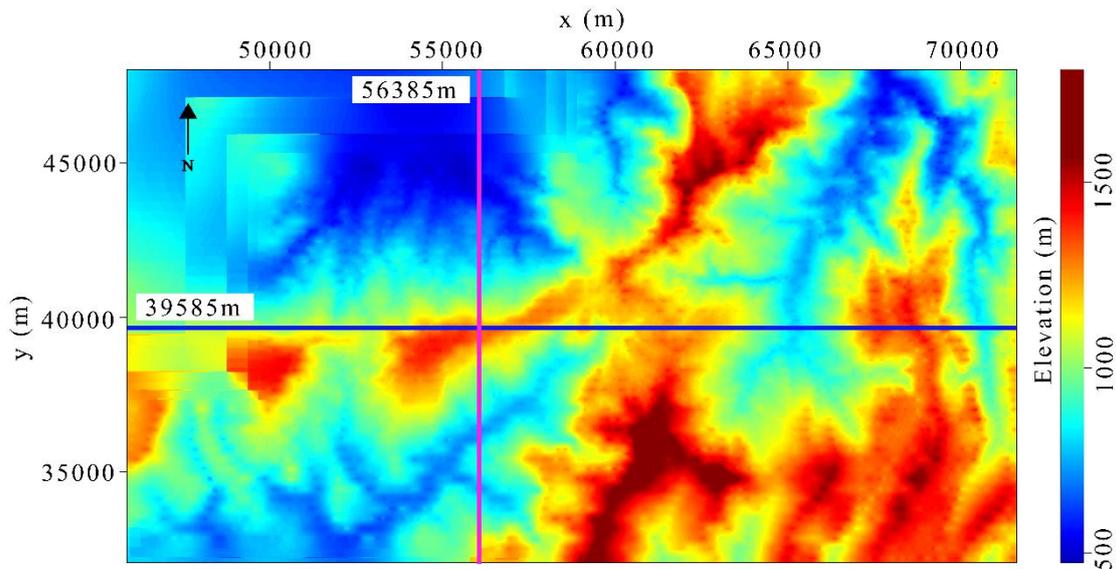
(e)

**Figure 4-28:** 2-D synthetic data test on Marmousi model. (a) True velocity model; (b) true attenuation model; (c) inverted velocity model by first arrival traveltim tomography; (d) inverted Q model derived by true velocity model of (a); (e) inverted Q model derived by inverted velocity model of (c).

#### 4.3.4 Field data test

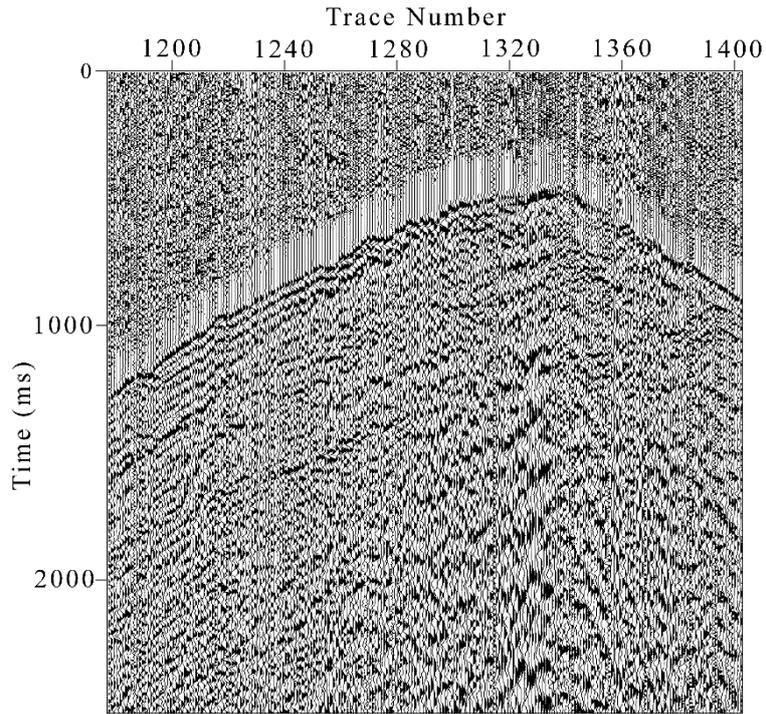


**Figure 4-29:** Survey geometry map: source (red) and receiver (blue) position.

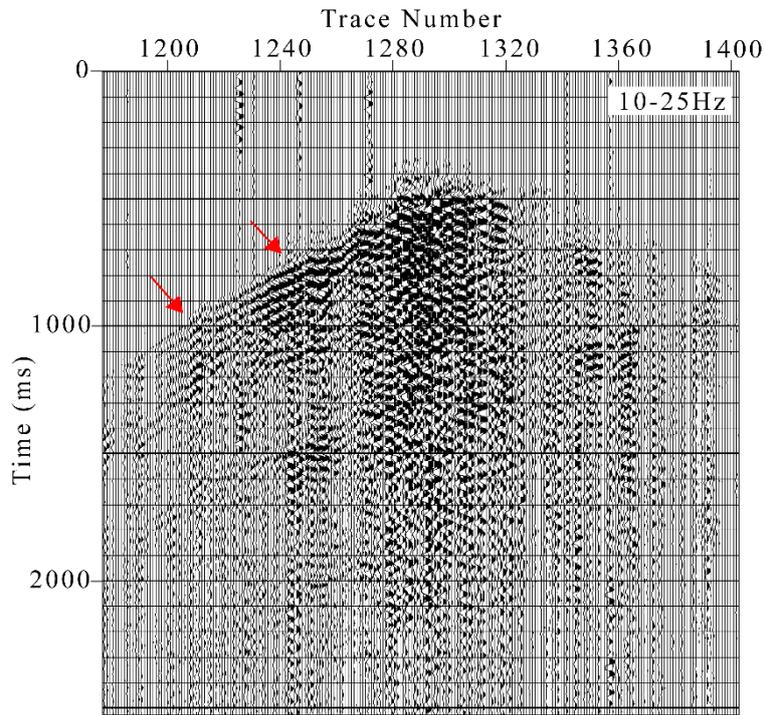


**Figure 4-30:** Elevation map of survey area: the magnet line indicates the position of both velocity and Q model profile presented in figure 15 and 16 along y direction at 56385m. The blue line is same but along x direction at 39585m.

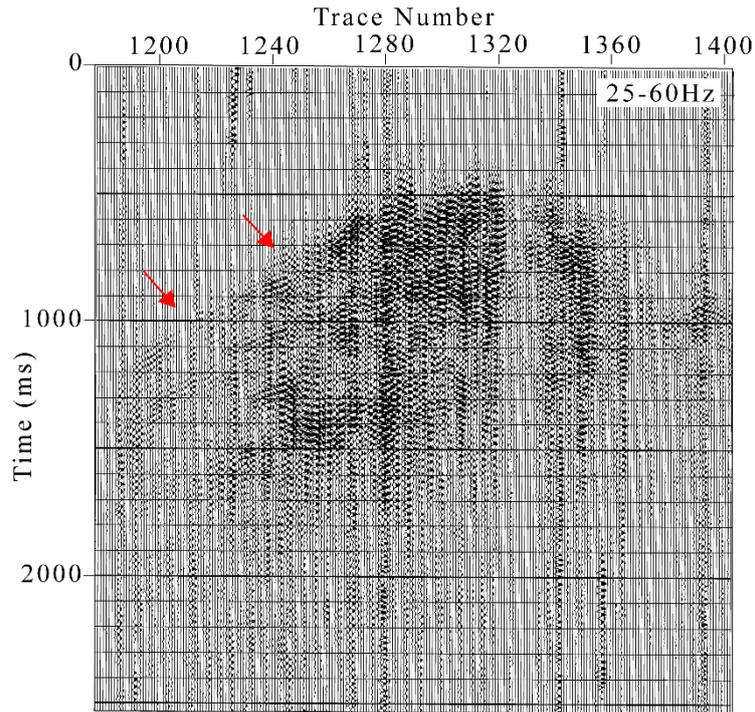
This example is an application on 3D land field dataset acquired at a mountainous area in southern China, and the whole seismic survey area is around 350km<sup>2</sup>. We selected one part of the data to test the feasibility and effectiveness of our proposed method. The acquisition geometry is shown in Figure 4.29. The total source and trace number are 7,407 and 32,558,862, respectively. The maximum offset is approximately 8500 m. The seismic recording length is 6s with sample rate of 2ms. The elevation map of our chosen survey area is shown in Figure 4-30, in which the maximum elevation difference is 1040m, including different complex landforms such as river channel deposit, piedmont alluvial fan and others. Since then, the velocity lateral variation and attenuation effect in in near surface is expecting to be intensive. Additionally, the poor condition of seismic data acquisition leads to a poor quality of seismic records with low SNR, which is very challenging to obtain a satisfied tomographic solution.



**Figure 4-31:** Original shot gather.



**(a)**

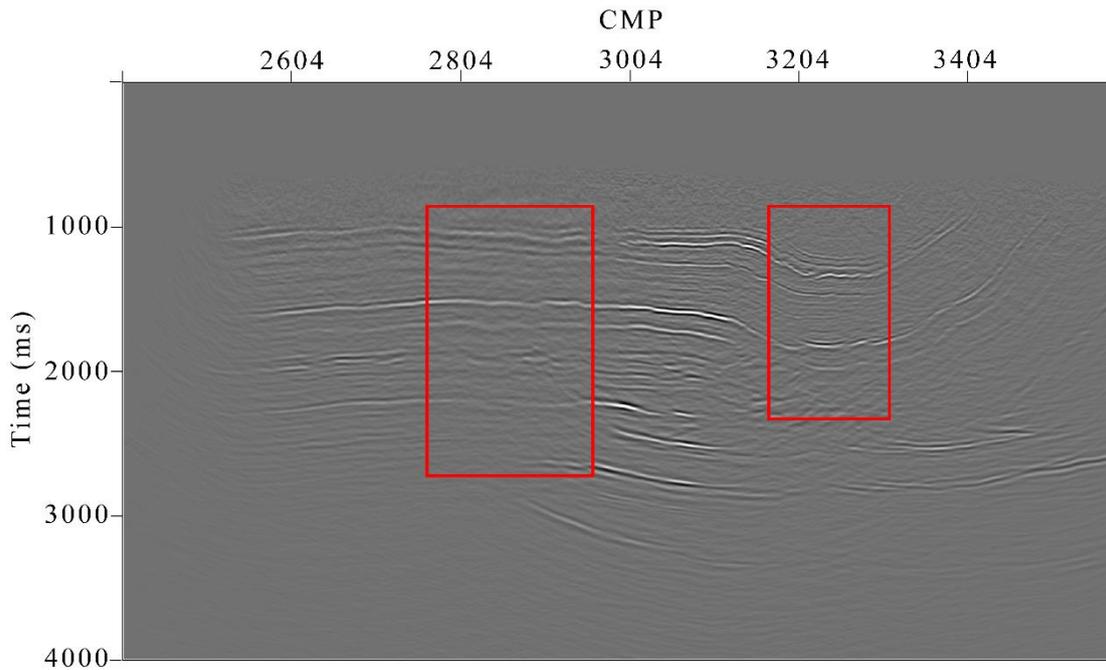


(b)

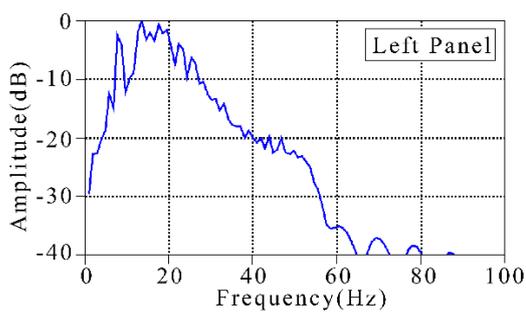
**Figure 4-32:** Frequency division scanning of original shot gather: red arrows indicate locations of shallow reflections. (a) 10-25Hz frequency range shot gather; (b) 25-60Hz frequency range shot gather.

One shot gather is selected as shown in Figure 4.31 to further illustrate the details of this dataset. The seismic traces are contaminated by noise and shallow reflections can be barely identified from original shot gather. A division frequency scanning analysis is performed on the record to compare the low frequency (10-25Hz) and high frequency (25-60Hz) features in Figure 4.32a and b, respectively. The shallow reflections (marked by red arrows) appeared at the low frequency band range of Figure 4.32a, and vanished at higher frequency band range of Figure 4.32b. It indicates the presence of some local attenuating absorption anomalies in the subsurface. The corresponding interferences caused by attenuation effects could also be observed in CMP stacking of Figure 4.33a, in which differences of both resolution and energy

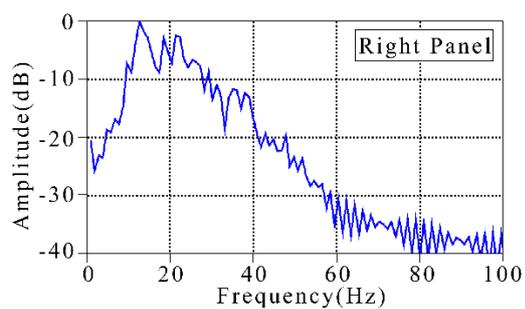
levels appeared in left and right panels. The lacking of high frequency components in left panel is more apparent than right one as shown in Figure 4.33b and 4.33c. All these observations prove the existence of some strong attenuation anomalies in this area. More specifically, the attenuation effects in the left region are much stronger than right region as depth increasing.



(a)



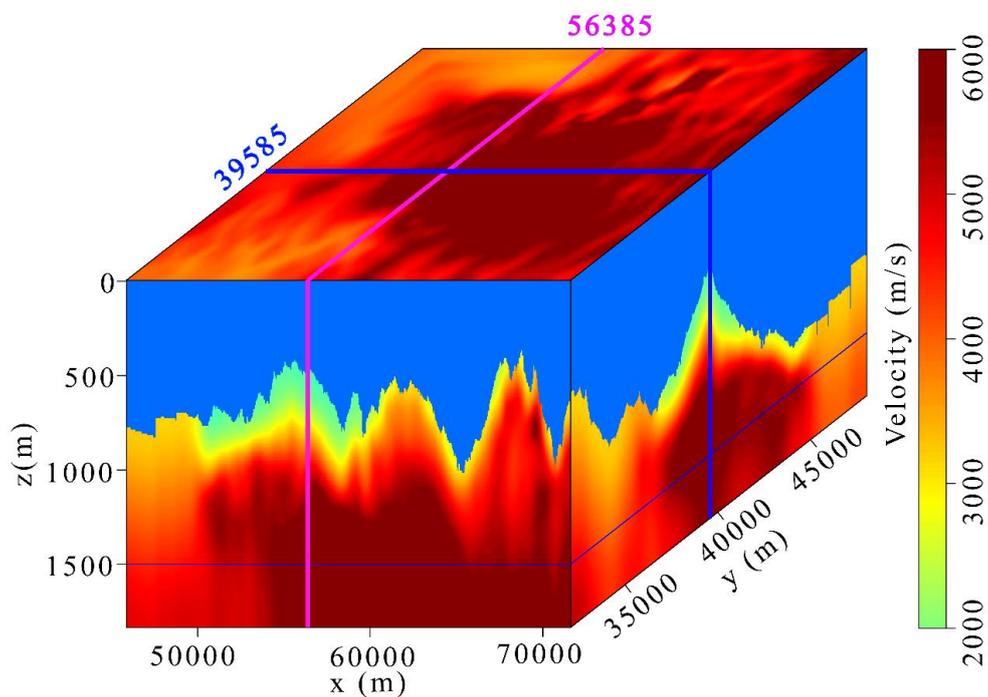
(b)



(c)

**Figure 4-33:** Original CMP stacking image results. (a) Original CMP stacking; (b) amplitude spectrum of left panel in (a); (c) amplitude spectrum of right panel in (a).

A 3-D near-surface velocity model is estimated by first arrival traveltimes tomography as presented in Figure 4.34. The most of recovered near-surface velocity structure of left region shows a low velocity anomaly, while high anomaly presenting on right. According to the relationship of strata velocity and medium Q factor, a low velocity rock has less compaction, and consequently, the attenuation effect will be stronger. This observation is consistent with the CMP stacking (figure 4.32a).



**Figure 4-34:** Inverted velocity model derived by FAT tomography.

The inverted velocity model is utilized for generating first arrivals, by which the attenuated first arrival waveforms will be estimated as described in section 4.2.3. The waveform length is determined by the dominant frequency from current seismic traces statistics. In this case of poor-quality (low SNR) data condition, to avoid unstable attenuated time estimation by single channel (trace), we employed a multi-channel weighted average function

---

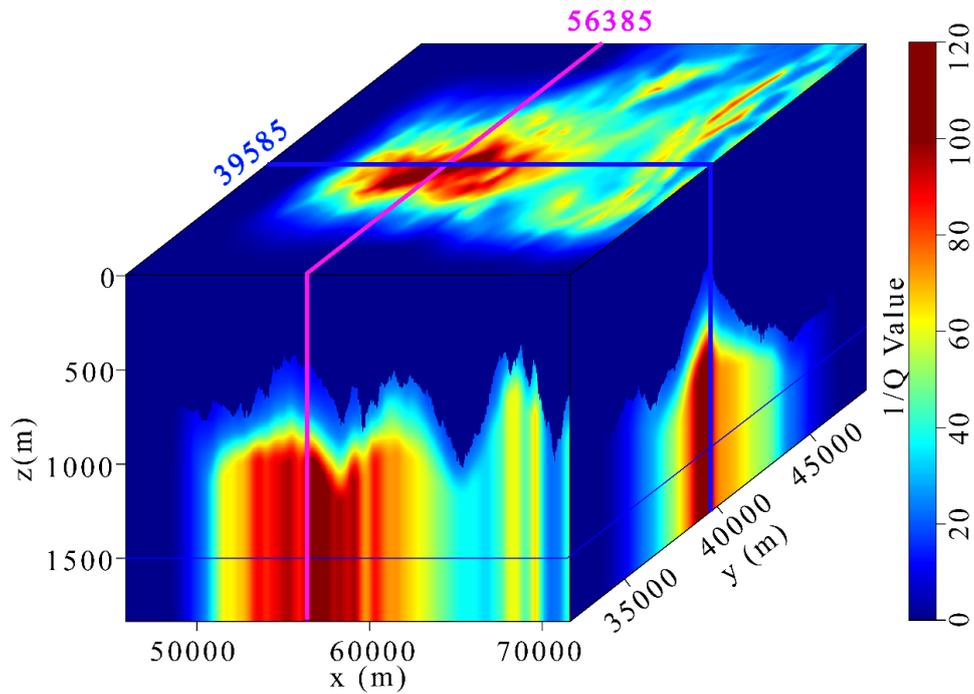
to apply on several high-quality reference traces. We choose 50 shot gathers with largest dominant frequencies, and then apply a weighted averaging on those near offset channels within 100m to maintain high quality of data. The averaged reference trace will be eventually used in  $t^*$  estimation through spectral ratio algorithm. However, the weighted averaging over multiple traces will diminish the local absorption anomaly features in a finer scale, which is leading to a global “smoothing effect” and degrade the resolution of inverted Q model.

The inverted near-surface Q distribution model is shown in Figure 4-35. The abrupt subsurface velocity variation in near-surface (as shown in figure 4-34) and limited acquisition aperture, resulting in a shallow effective imaging depth of inverted Q model. Below the effective imaging region, we simply fill in the same maximum Q value of each trace to enhance the visibility of resulted image. In inverted Q model, the low Q anomalies are mainly distributed in the left region with large thickness, and the high Q anomalies are in right region with small thickness. It is consistent with the velocity anomalies distribution (Figure 4-34) as we discussed previously.

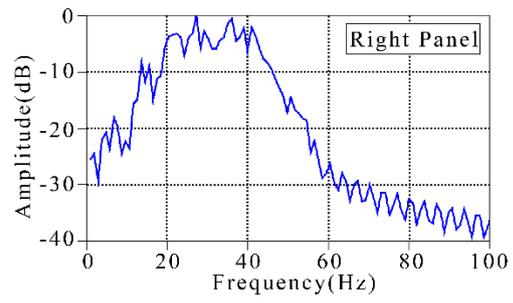
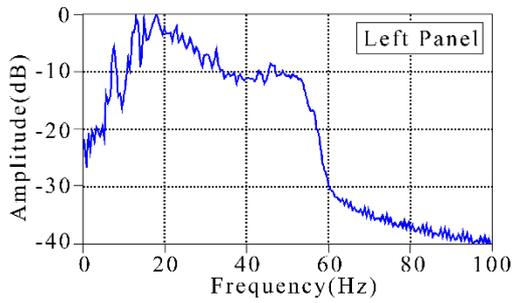
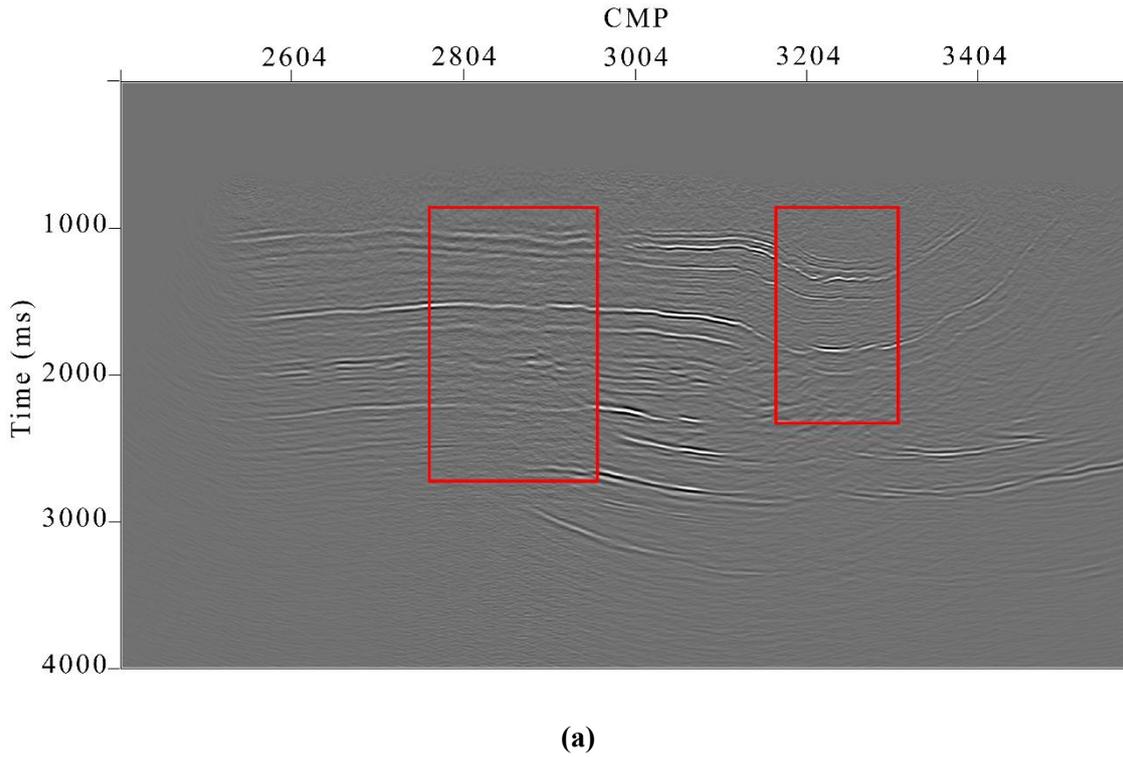
To further assess the validity of inverted Q model, we perform the following procedures. Firstly, the elevation of high velocity top interface is determined by the inverted velocity model. Above this interface, we assume that the outgoing seismic ray is traveling vertically, and hence the exact time -spatial attenuation effect could be evaluated by equation (4.1) and (4.2) in a CMP stacking. Eventually these attenuation effects are removed from original stacking, and the corrected CMP stacking is presented in Figure 4-36a. The lateral continuity of seismic

---

events is enhanced respecting to both amplitude and geological structural in lateral direction. The resolution of whole stacking image was improved, and the frequency bandwidths of two panel regions have been broaden by recovering of both low and high frequency contents, as shown in Figure 4-36b and c.



**Figure 4-35:** Inverted attenuation model ( $1/Q$ ).



**Figure 4-36:** CMP stacking after attenuation compensation. (a) Attenuation compensated CMP stacking; (b) amplitude spectrum of left panel in (a); (c) amplitude spectrum of right panel in (a).

#### 4.4 Summary and discussions

In this chapter, we developed a first arrival attenuation tomography based on adjoint-state method, which is highly computational efficiency and its computation cost is proportional to the size of model parameterization. A specific governing equation of the adjoint state method

---

for  $t^*$  is formulated to enable us to calculate the gradient of misfit function. It is feasible to handle with large dataset, especially for a common dense survey nowadays. The Q distribution model could be obtained from a tomographically inversion by minimizing the difference between observed  $t^*$  and synthetic  $t^*$  under a least-square framework.

To achieve a relatively accurate observed  $t^*$  from the first arrival of seismic data, we design a specific work flow with  $t^*$  estimation and optimal selection in terms of the spectral ratio and adaptive correction method, respectively. The modified logarithm spectral ratio method can provide all possible  $t^*$  estimation corresponding to the frequency response of different wave types in first arrival, such as direct wave, turning wave or refraction. An adaptive correction method can further select the optimal estimated  $t^*$  by evaluating their similarities between synthetic and real attenuated seismic data for each possible  $t^*$ . Hence, the one with correct frequency response of wave type will be screened out through these procedures and used as input for following tomographic inversion.

The effectiveness of our  $t^*$  estimation workflow is demonstrated in a synthetic test of a two-layer model from many perspectives in details. Besides, we design a standard process in practical implementation to handle with noise contamination and interferences of different wave types commonly existed in first arrival before and while conducting  $t^*$  estimation. It is advisable to estimate  $t^*$  from prestack seismic data that have not been processed by frequency filter methods (e.g., deconvolution), or more generally, any method that is possible to diminish the salient amplitude information related to source wavelet. Thanks to the redundancy of

---

seismic data, some non-effective  $t^*$  estimation derived from interference region could be directly abandoned in the tomographic inversion to ensure the reliability of the results. Under the guidance of this workflow, two other synthetic data tests are deployed and achieve promising results. In synthetic data test of a simple model with topographic surface, both of shallow and deep attenuation anomalies are reconstructed in a reasonable way. Its success is partly attributed to an accurate  $t^*$  estimation based on a good inverted velocity model. Results from Marmousi model demonstrate some benefits and robustness in recovering near surface attenuation anomalies in the presence of heterogeneous.

Accuracy of proposed method mainly depends on the accuracy of input velocity model and  $t^*$  estimation for attenuation tomographic inversion. The accuracy of inverted attenuation model is impacted by the precision of the velocity model we used in Q tomography. In practical, apart from following the workflow as presented in previous, we also use a statistic approach to obtain a reference signal to approximate source wavelet with different weight for multi-traces. The real data test results prove that this method works effectively in low SNR circumstance and brings benefit to stabilize tomographic solutions. Even though the quality of raw seismic record is poor and only partially offset data with limited acquisition were used, the inverted Q distribution in near-surface shows a reasonable consistency corresponding to the velocity model.

In essence, our tomographic inversion is ray theory based (high frequency approximation assumption) and only use first arrival, hence the resolution of the Q distribution model is

---

limited by the ray coverage and high fidelity only preserved in near surface region. To improve deeper subsurface imaging, a joint inversion of both transmission and reflection could be an efficient way to enhance the final inversion result. A tomographic inversion based on a wave equation would be another way to further improve the resolution if one had enough computing resource to afford heavy computational cost in practical applications. Besides, a proper time window to enclose an entire seismic wavelet at each time location predicted from inverted velocity model is also urgent for effectively applying the spectral ratio method. Time-frequency analysis, such as windowed time-variant spectral analysis or the continuous wavelet transform might be a helpful tool to address this issue, and it is one direction of our future work.

Other limitations of our method come from linear frequency attenuation model itself as the only intrinsic attenuation and non-dispersion velocity is assumed. The absorption property of the medium leading to attenuation effect on the amplitude of source wavelet is exactly what we concentrated on. However, any  $Q$  model can be viewed as effective mathematical tools rather than an ultimate self-explanation (Bourbié et al., 1987). The underlying assumption of our model will definitely introduce error to a certain extent varying from case to case. After all, the quality factor  $Q$  is a phenomenological quantity interpreted from real data or physical reality. For instance, there are many cells along raypath, and the absorption type could be different, such as pure intrinsic, media inhomogeneity or possibly some other unknown factors. Correspondingly, inverted types of  $Q$  would be intrinsic  $Q$ , effective  $Q$ , etc. Since then, a “ $Q$  distribution model” referring to apparent  $Q$  seems to be a proper nomination for our retrieved

---

model result. In practical, data should be the ultimate arbitrator (Morozov, 2009) and their analysis is useful for understanding the application scope of a certain method. Nevertheless, our research study on using first arrival to conduct attenuation tomographic inversion still has its practical significance, in considering of feasibility of the spectral ratio method and accessibility of first arrival from raw seismic data.

---

## CHAPTER 5 CONCLUSION AND DISCUSSION

Geometric ray theory related technique is a longstanding research area and has been playing an important role in exploration seismology for decades. Nowadays, travelttime-based tomography is still a prevailed choice considering its efficiency, robustness and feasibility in many application scenarios. Especially, when it comes to the near-surface modeling of land data. FAT tomography is still most commonly used inversion method from scratch. This dissertation aims to handle two challenges existed in the application of FAT tomography: improving its computational efficiency with large seismic dataset from densely sample survey and retrieving near-surface Q model from first arrival seismic data.

### **5.1 Mapping near-surface velocities using efficient FAT tomography for large dataset**

To handle large dataset and improve the computational efficiency of FAT tomography for retrieving near-surface velocity model, a modified GN method named SA+ is proposed under a compressive sensing framework. I first adopted the adjoint-state technique with FAT using gradient-based approach, and compared its performance with conventional ray-tracing based FAT tomography in both 2D and 3D synthetic model. The numerical test result indicated that adjoint-state method could achieve the identical result while improving the computational efficiency in terms of memory occupation and time cost. It allows to make the memory occupation of inversion algorithm only few times proportional to the whole model size. Hence, the adjoint-state method has great advantages for the seismic data from dense receiver array deployment.

---

To further improve computational efficiency in time cost wise, a SA+ method composed by stochastic process and gradient optimization is proposed. Two typical realization of Stochastic process, SAA and SA, was implemented with the adjoint state method. The stochastic processes allow to significantly reduce the problem dimensionality by using a small part of dataset during inversion, meaning that fewer PDE was to be solved hence less time cost for a single iteration. However, both of SAA and SA required extra number of iterations to obtain identical results to standard FAT tomography using full dataset, which is high time cost. On the other hand, limiting the number of iterations will induce artifacts from random subsampling of sources.

Since then, a gradient optimization method to mitigate artifacts by sparsity regularization is adopted. It is achieved by using an online ODL method to exploit the convex-composite structure of the model update in a sparse domain spanned by dictionary atoms. These atoms are formed in non-overlapping blocks through a sparse orthonormal transform, by which each block represents a localized featuring patches of model update learned over a dictionary from the previous iteration. There are two main advantages of adopting online ODL method: 1.) its low computational demanding due to orthogonal operation and 2.) its capability of adapting to non-intuitive signal regularities beyond piecewise smoothness, allowing to capture salient information from a “noisy” gradient of stochastic process. From synthetic test, we could see that such dynamic regularization on model updates eventually functioned as a “smoothing filtering” to suppress the subsampling artefacts and mildly smooth out gradient. However, different from commonly used filtering-based in FAT tomography, it allows to capture salient

---

information of a “noisy” gradient from a stochastic process based on data driven. Eventually, the original least-square problem reformulated into a LASSO problem and solved by a modified GN method that can trace the optimal trade-off between the L2-norm of residual and L1-norm of sparse coefficients simultaneously.

Through synthetic data test, we found that SA and our proposed gradient optimization were the optimal method combination (SA+) with the best performance in terms of accuracy and stability. In general, the time cost could be reduced about 70% for a relatively clean data from our experience. For noisy data, a practical guidance of data decimation percentage is also provided through this test: if the noise level for first arrival picks less than 10%, a data decimation percentage round 10% would be a safe choice; when noise level rise above 10%, a data decimation percentage at least large than 15%. At the same time, one should also be aware of geological complexity of the target area to choose a proper data decimation percentage to ensure the accuracy and stability of SA+. Following this guidance, we applied SA+ in a field data with 20% dataset as input and obtained an almost identical retrieved velocity model comparing to standard FAT tomography using full dataset. The difference of their long-wavelength statics is minor and acceptable for following process, by which further reveals the potential for SA+ method in practical application.

## **5.2 Mapping near-surface Q values using First arrival attenuation tomography based on adjoint state method**

To retrieve near-surface Q distribution model using first arrival, a first arrival Q

---

tomography was proposed. Under the assumption invariant ray path in weakly dissipative (high quality factor  $Q$ ) subsurface medium, this tomographic inversion approach can be integrated in two cascading processes consisting of FAT tomography and attenuation tomography. To ensure the computational efficiency, both of tomographic inversion were adopted with adjoint-state technique. To implement attenuation tomography with adjoint-state technique, a governing equation involved with path attenuation factor  $t^*$  is formulated and adopted with FSW algorithm. With same velocity model and  $t^*$  as input, the inverted  $Q$  distribution model from adjoint-state method is almost identical to that of raytracing-based tomography, confirming the accuracy of this algorithm.

The FAT tomography will be conducted at first to provide velocity model as one of input for the following attenuation tomography. Besides, the synthetic FAT derived from the inverted velocity model can provide a stable guidance in selecting time window of first arrival from seismic trace, wherein the potential attenuation information related with  $t^*$  is contained. For the following attenuation tomography, its success is heavily dependent on the accuracy of observed  $t^*$  estimated from attenuated first arrival.

The path attenuation factor  $t^*$  is related to the amount of amplitude attenuation along the ray path. Therefore, a modified logarithmic spectral rate method was proposed to estimate  $t^*$ . It improves the accuracy of  $t^*$  estimation by removing the effects of frequency-dependent propagation responses of different wave types existed in the first arrival, e.g., direct wave, turning wave and refraction. From numerical experiment, we observed that the amplitude spectra of direct and turning wave had identical waveforms as source wavelet, while the

---

refraction appearing to be a low peak frequency shift. This property can be utilized for identifying and separating refraction from other wave types blended in first arrival, and an adaptive correction method is proposed. Two estimated  $t^*$  will be obtained based on these two different frequency-dependent responses, and their synthetic attenuated seismic traces will be calculated. Then their similarities to observed attenuated seismic traces will be evaluated for selecting the optimal estimated  $t^*$ , and used as observed  $t^*$  for attenuation tomography. The Q distribution model could be retrieved from a by minimizing the discrepancy between observed  $t^*$  and synthetic  $t^*$  under a least-square inversion framework.

In practical, though first arrival generally has a higher quality than other secondary waves such as reflection or refraction seismic data in term of SNR. Noise contamination and interference of different wave types will inevitably have impact on  $t^*$  estimation. Since then, we also proposed a detailed practical implementation to handle with this issue. Redundancy of field data allows to directly reject some of unreliable estimated  $t^*$  obtained from interference region without degrading too much of quality of inverted Q distribution model. It is verified by the real data test with relatively low SNR from mountain field. The inverted Q model confirmed the consistency for the near surface velocity structure and geological outcrop along the topographic surface. The improvements of resolution and frequency bandwidth recovery of corrected CMP stacking image further prove the validity of an inverted Q distribution model.

### **5.3 Outlook on future work**

The work on highly efficient FAT tomographic inversion in large dataset is all tested in 2D

---

case. Nowadays, in reality, a 3D acquisition is the field of those classical algorithms might suffer from computational limitations, in terms of memory occupation, computation time, or implementation. In Chapter 2, the benefits brought by adjoint state technique adoption have previously been presented in the 3D synthetic overthrust model. It has great potential to extend our highly efficient FAT tomographic inversion into 3D case to further improve inversion computation efficiency. A 3D geometry could provide two more dimensions to subsample the data and the related dimensionality reduction problem should be more efficient than the 2D case. One important issue I should be focused on in this future work extension is that how to adapt 2D block-wise online ODL into 3D case. The patch selection should be rigorously manipulated along another two dimensions. Additionally, some other wave-equation based inversions also could be implemented with CS framework to further improve the computation efficiency. It should be a promising extension, since that dimensionality reduction could be more efficient in such sophisticated method. Besides, the update gradients used in these methods are mostly less “smooth” than ray-based one, from which the dictionary learning could capture more directional features in sparse space to have better constraint (or regularization) during inversion.

Another promising direction is counting on improving dictionary learning algorithm to better exploit convex-composite structure of the problem. In real world, the appearance of geological structure in subsurface is actually appearing to have many nonlocal similarities, meaning that similar structural features are repeated. Capturing these features could effectively reduce the freedom degrees in model space, and eventually bring benefit to improving the

---

inversion convergency. It is quite analogous to the idea of training dataset and test dataset in machine learning. Some geological patterns of many classical models (e.g., Marmousi, Amoco 94 or BP subsalt) at different places could be set up as external models to anticipate the dictionary training, serving as prior information to constrain the inversion.

The resolution of retrieved Q distribution model is limited by the input of inverted velocity model derived from previous ray-based FAT tomography. Since then, an inverted velocity model with higher image quality can lift a favor in improving the resolution and fidelity of the reconstructed Q distribution model. A joint tomographic inversion of both transmission and refraction seismic data could be an efficient way to provide a better velocity model with more constraints. Besides, wave equation-based tomography would be another way to further improve the resolution by providing high quality velocity model, if one had enough computing resource to afford heavy computational cost in practical applications.

To successfully apply the spectral ratio method for  $t^*$  estimation, one important underlying assumption is that the wavelet before and after absorption is known, meaning that an entire seismic wavelet should be enclosed in the time window when selecting first arrival. Since then, using a time-frequency analysis technique, such as windowed time-variant spectral analysis or the continuous wavelet transforms might be a helpful tool to solve this issue.

---

## BIBLIOGRAPHY

- Abercrombie, R. E., 1997. Near-surface attenuation and site effects from comparison of surface and deep borehole recordings. *Bulletin of the seismological society of America*, 87, 731-744.
- Aharon, M., M. Elad, and A. Bruckstein, 2006, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54, 4311 - 4322.
- Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, (2nd Ed). University Science Books, Mill Valley, California.
- Aki, K., & Lee, W. H. K., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first p arrival times from local earthquakes: 1. a homogeneous initial model. *Journal of Geophysical Research*, 81(23), 4381-4399.
- Aki, K., Christoffersson, A., Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, 82, 277–296.
- Aldridge, B. and D. Oldenburg, 1992. Refractor imaging using an automated wavefront reconstruction method. *Geophysics*, 57, 223-235.

- 
- Asakawa, E. & Kawanaka, T., 1993. Seismic ray tracing using linear travelttime interpolation. *Geophysical Prospecting*, 41, 99–112.
- Bak, S., McLaughlin, J. & Renzi, D., 2010. Some improvements for the fast-sweeping method. *Siam Journal on Scientific Computing*, 32(5), 2853 – 2874.
- Best, A.I., Mccann, C. & Sothcott, J., 1994. The relationships between the velocities, attenuations and petrophysical properties of reservoir sedimentary rocks. *Geophysical Prospecting*, 42, 151-178.
- Betrsekas, D. P., and J. N. Tsitsiklis, 2000, Gradient convergence in gradient methods with errors. *Siam Journal of Optimization*, 10, 627–642.
- Bishop, T. N., Bube, k. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, J. T., Spindler, D. A. and Wyld, H. W., 1985. Tomographic determination of velocity and depth in laterally-varying media, *Geophysics*, 50, 903-923.
- Bois, P., La Porte, M., Lavergne, M., Thomas, G., 1972. Well-to-well seismic measurements. *Geophysics* 37, 471–480.
- Bourbié, T., 1987. Acoustics of porous media. *Journal of Acoustical Society of America*, 91(5), 3080-3080.
- Bousquet, O., and L. Bottou, 2007. The tradeoffs of large-scale learning. *Advances in Neural Information Processing Systems*. MIT Press, 161-168.
- Bregman, N.D., Bailey, R.C., Chapmans, C.H., 1989. Crosshole seismic tomography. *Geophysics* 54, 200-215.
- Bremaecker, D. & Jean-Claude., 1977. Is the oceanic lithosphere elastic or viscous? *Journal of Geophysical Research*, 82, 2001-2004.

- 
- Brooke J. Carney. Building velocity models for steep-dip prestack depth migration through first arrival travelttime tomography[D]. Virginia Polytechnic Institute and State University, 2000.
- Brzostowski, M. A. & McMechan, G.A., 1992. 3-D tomographic imaging of near-surface seismic velocity and attenuation. *Geophysics*, 57, 396-403.
- Bube, K. B., R. T. Langan, and J. R. Resnick, Theoretical and numerical issues in the determination of reflector depths in seismic reflection tomography. *Journal of Geophysical Research*, 100, 12,449-12,458, 1995.
- Bulant, P., 1996. Two-point ray tracing in 3-d. *Pure and Applied Geophysics*, 148(3-4), 421-447.
- Carcione, J.M., Helle, H.B. & Pham, N.H., 2003. White's model for wave propagation in partially saturated rocks: Comparison with poroelastic numerical experiments. *Geophysics*, 68, 1389–1398.
- Carcione, J.M., Kosloff, D. & Kosloff, R., 1988. Viscoacoustic wave propagation simulation in the earth. *Geophysics*, 53, 769-777.
- Cavalca, M., Fletcher, R. & Riedel, M., 2013. Q compensation in complex media-ray-based and wavefield extrapolation approaches. 83th Annual International Meeting, SEG, Expanded Abstracts, 3831-3835.
- Červený, V., 1987. *Ray-Tracing Algorithms in Three-Dimensional Laterally Varying Layered Structures*. Springer Netherlands.
- Červený, V., 2001. *Seismic ray theory*. Cambridge Univ. Press, Cambridge.

- 
- Cavalca, M., Fletcher, R. P., & Du, X., 2015. Q-compensation Through Depth Domain Inversion. Eage Conference & Exhibition.
- Chavent, G., 1974. Identification of function parameters in partial differential equations. Joint Automatic Control Conference, 155–156.
- Chen, S. S., D. L. Donoho & M. A. Saunders, 1998, Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20, 33 – 61.
- Courtier, P., Talagrand, O., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation - part 2: numerical results. *Quarterly Journal of the Royal Meteorological Society*. 113(478), 1329-1347.
- D., A., & Waltham. (1988). Two-point ray tracing using fermat's principle. *Geophysical Journal*.
- Dahlen, F.A., Hang, S.H., Nolet, G., 2000. Fréchet kernels for finite frequency travel times. I. Theory. *Geophysical Journal of International* 141, 157–174.
- de Hoop, M.V., van der Hilst, R.D., 2005a. On sensitivity kernels for “wave equation” tomography. *Geophysical Journal of International* 160, 621–633.
- de Hoop, M.V., van der Hilst, R.D., 2005b. Reply to comment by F.A. Dahlen and G. Nolet on: “On sensitivity kernels for wave-equation transmission tomography”. *Geophysical Journal of International* 163, 952–955.
- de Hoop, M.V., van der Hilst, R.D., Shen, P., 2006. Wave-equation reflection tomography: annihilators and sensitivity kernels. *Geophysical Journal of International* 167, 1332–1352.
- Dijkstra, E.W., 1959. A note on two problems in connection with graphs. *Numerische Mathematics*, 1, 269–271.

- 
- Dziewonski, A.M., Hager, B.H., O'Connell, R.J., 1977. Large-scale heterogeneities in the lower mantle. *J. Geophysical Research*. 82, 239–255.
- Fabrice Jurado, Patrick Lailly & Andreas Ehinger, 1998. *Proceedings of SPIE, Mathematical Methods in Geophysical Imaging V*, 70-81.
- Farra, V., Madariaga, R., 1987. Seismic waveform modelling in heterogeneous media by ray perturbation theory. *J. Geophysical Research*. 92, 2697–2712.
- Fischer, R., Lees, J.M., 1993. Shortest path ray tracing with sparse graphs. *Geophysics* 58, 987–996.
- Futterman, W. I., 1962. Dispersive body waves. *Journal of Geophysical Research Atmospheres*, 67(13), 5279-5291.
- Ganley, D. C. & Kanasewich, E. R., 1980. Measurement of absorption and dispersion from check shot surveys, *Journal of Geophysical Research Solid Earth*, 85, 5219-5226.
- Hargreaves, D. & Calvert, A.J., 1991. Inverse Q Filtering by Fourier Transform. *Geophysics*, 56, 519.
- He, Y., Xie, Y. & Xin, K., 2016, Diving Wave Q Tomography for Compensating Absorption and Dispersion of Shallow Gas Cloud, Presented at the 78th Annual International Conference and Exhibition, EAGE.
- Hennenfent, G., E. van den Berg, M. P. Friedlander, and F. J. Herrmann, 2008, New insights into one-norm solvers from the Pareto curve: *Geophysics*, 73, no. 4, A23–A26.
- Herrmann, F. J., and X. Li, 2011a, Efficient least-squares imaging with sparsity promotion and compressive sensing.: Tech. rep., University of British Columbia, Vancouver.

- 
- Hicks, G., Pratt, R.G., 2001. Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit. *Geophysics* 66, 598–612.
- Hobro, J. W. D., Singh, S. C., & Minshull, T. A., 2003. Three-dimensional tomographic inversion of combined reflection and refraction seismic traveltimes. *Geophysical Journal International* (1), 79-93.
- Hole, J. A., 1992. Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research: Solid Earth*, 97.
- Hole, J.A. & Zelt, B.C., 1995. 3-D finite-difference reflection traveltimes, *Geophysical Journal of International*, 121, 427 - 434.
- Hu, W., Liu, J., Bear, L., & Marcinkovich, C., 2011. A robust and accurate seismic attenuation tomography algorithm, 81th Annual International Meeting, SEG, Expanded Abstracts, 2727-2731.
- Huang, J.W. & Bellefleur, G., 2012. Joint transmission and reflection traveltimes tomography using the fast sweeping method and the adjoint-state technique, *Geophysical Journal International*, 188, 570-582.
- Humphreys, E., & Clayton, R. W., 1988. Adaptation of back projection tomography to seismic travel time problems. *Journal of Geophysical Research Solid Earth*, 93(B2).
- Jianzhong, Z., Yueqin, H., Lin-Ping, S., & Qing-Huo, L., 2011. Fast and accurate 3-d ray tracing using bilinear traveltimes interpolation and the wave front group marching. *Geophysical Journal International* (3), 1327-1340.

- 
- Jun-Wei, Huang, Gilles, & Bellefleur., 2012. Joint transmission and reflection traveltime tomography using the fast sweeping method and the adjoint-state technique. *Geophysical Journal International*.
- Keers, H., Vasco, D.W. & Johnson, L.R., 2012. Viscoacoustic crosswell imaging using asymptotic waveforms, *Geophysics*, 6, 1569-1582.
- Kim, & Seongjai, 2002. 3-d eikonal solvers: first-arrival traveltimes. *Geophysics*, 67(4), 1225-1231.
- Kjartansson, E., 1979. Constant Q wave propagation and attenuation, *Journal of Geophysical Research Solid Earth*, 84, 4737–4748.
- Klimes, L. & Kvasnicka, M., 1994. 3-D network ray tracing. *Geophysical Journal of International*, 116, 726–738.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, 74, WCC177–WCC188.
- Lehmann B. 2007. *Seismic Traveltime Tomography for Engineering and Exploration Applications*. EAGE.
- Leung, S. & Qian, J., 2006. An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals: *Communications in Mathematical Sciences*, 4, 249-266.
- Lions, J., 1971. Optimal control of systems governed by partial differential equations: *Bulletin of the London Mathematical Society*, 1049.

- 
- Luo, Y. and Schuster, G.T., 1991, Wave-equation travelttime inversion: *Geophysics*, 56(5), 645–653.
- Lutter, W. J. and R. L. Nowack, 1990. Inversion for crustal structure using reflections from the PASSCAL Ouachita experiment, *Journal of Geophysical Research Solid Earth*. 95, 4633-46.
- Mallat, S., and Z. Zhang, 1993, Matching pursuits with time-frequency dictionaries: *Signal Processing, IEEE Transactions on*, 41, 3397 – 3415.
- Marquering, H., Dahlen, F.A., Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency travel times: the banana–doughnut paradox. *Geophysical Journal of International* 137, 805–815.
- Matheny, P., Sambell, R., Mahrooqi, S., Yarubi, S., & Abri, S., 2009. Evolution of the land seismic super crew. *SEG Technical Program Expanded*, 28: 81~85.
- Matthew, A., Brzostowski, George, A., & McMechan. (1992). 3-d tomographic imaging of near-surface seismic velocity and attenuation. *Geophysics*.
- McCaughey, M. & Singh, S.C., 1997. Simultaneous velocity and interface tomography of normal-incidence and wide-aperture seismic travelttime data, *Geophysical Journal International*, 131, 87–99.
- McMechan, G.A., 1983. Seismic tomography in boreholes. *Geophysical Journal of the Royal Astronomical Society*. 74, 601–612.
- McMechan, G.A., Harris, J.M. and Anderson, L.M. (1987). Crosshole tomography for strongly variable media with applications to scale model data. *Bulletin Seismological Society of America*. 77, 1945-1960.

- 
- Moghaddam, P. P., and F. J. Herrmann, 2010, Randomized full-waveform inversion: a dimensionality-reduction approach. SEG Technical Program Expanded Abstract, 977–982.
- Montelli, R., Nolet, G., Dahlen, F.A., 2006. Comment on ‘Banana–doughnut kernels and mantle tomography’ by van der Hilst and de Hoop. *Geophysical Journal International*, 167, 1204–1210.
- Moore, G.E., 1965. Cramming more components onto integrated circuits. *Electronics* 38, 114–117.
- Moser, T.J., 1991. Shortest path calculation of seismic rays. *Geophysics*, 56, 59–67.
- Müller, T. M., Gurevich, B. & Lebedev, M., 2010. Seismic wave attenuation and dispersion resulting from wave-induced flow in porous rocks—a review, *Geophysics*, **75**, 75A147–75A164.
- Nemirovski, A., A. Juditsky, G. Lan, and A. Shapiro, 2009, Robust stochastic approximation approach to stochastic programming: *SIAM Journal on Optimization*, 19, 1574 – 1609.
- Nercessian, A., A. Hirn, and A. Tarantola, Three-dimensional seismic prospecting of the Mont Dore volcano, France, *Geophysical Journal of the Royal Astronomical Society*, 76, 307–15, 1984.
- Noble M, Gesret A, Belayouni N, 2014. Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media. *Geophysical Journal International*, 199(3):1572-1585.

- 
- Noble, M., Thierry, P., Taillandier, C., & Calandra, H., 2010. High-performance 3D first-arrival traveltimes tomography. *Leading Edge*, 29(1), 86.
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: *Mathematics of computation*, 35, 773 – 782.
- Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and the Sun*. Cambridge University Press, Cambridge.
- Nowack R. L., and Michael P. M., 1997, Inversion of seismic attributes for velocity and attenuation structure: *Geophysics of Journal International*. 128,689-700.
- Oliveira, F.D.S., De Figueiredo, J.J.S., Oliveira, A.G., Schleicher, J. & Araújo, lury C.S., 2017. Estimation of quality factor based on peak frequency-shift method and redatuming operator: application in real data set, *Geophysics*, 82, N1-N12.
- Pan, W., K. A. Innanen, G. F. Margrave, and D. Cao, 2015, Efficient pseudo-Gauss-Newton full-waveform inversion in the  $\tau$ -p domain: *GEOPHYSICS*, 80, R225 – R14.
- Plessix, R.-E., and W. A. Mulder, 2004, Frequency domain finite-difference amplitude-preserving migration: *Geophysical Journal International*, 157, 975 – 987.
- Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal of the Royal Astronomical Society*, 167, 495-503.
- Pratt, R.G. & R.M. Shipp, 1999. Seismic waveform inversion in the frequency domain. Part 1. Theory and verification in a physical scale model. *Geophysics* 64, 888–901.
- Pratt, R.G., Worthington, M.H., 1988. The application of diffraction tomography to cross-hole seismic data. *Geophysics* 53, 1284–1294.

- 
- Qin, F., Y. Luo, K. B. Olsen, W. Cai, and G. T. Schuster, 1992. Finite-difference solution of the eikonal equation along expanding wavefronts, *Geophysics*, 57, 478-487.
- Quan, Y. & Harris, J.M., 1997. Seismic attenuation tomography using frequency shift method, *Geophysics*, 62, 895-905.
- Rawlinson, N. & Sambridge, M., 2004. Multiple reflection and transmission phases in complex layered media using a multistage fast marching method, *Geophysics*, 69, 1338 – 1350.
- Rawlinson, N., Hauser, J., & Sambridge, M., 2008. Seismic ray tracing and wavefront tracking in laterally heterogeneous media. *Advances in Geophysics*, 49, 203-273.
- Rawlinson, N., Pozgay, S., & Fishwick, S., 2010. Seismic tomography: a window into deep earth. *Physics of the Earth & Planetary Interiors*, 178(3-4), 101-135.
- Rice, J.A., Krohn, C.E. & Houston, L.M., 1991. Shallow near - surface effects on seismic waves, 61th Annual International Meeting, SEG, Expanded Abstracts, 10, 1646.
- Rickett J., 2006. Integrated estimation of interval-attenuation distributions: *Geophysics*, 71, A19-A23.
- Romain Brossier, Stéphane Operto, & Jean Virieux., 2009. Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion. *Geophysics*.
- Romero, A.E., McEvelly, T.V. & Majer, E.L., 1997. 3-D microearthquake attenuation tomography at the Northwest Geysers geothermal region, California, *Geophysics*, 62, 149-167.
- Rubinstein, R., T. Peleg, and M. Elad, 2013, Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model: *Signal Processing, IEEE Transactions on*, 61, 661 – 677.

- 
- Saito, H., 1990. 3-D ray tracing method based on Huygens' principle, in Soc. Exploration Geophys. 1990 Meeting, pp. 1024-1027, SEG, Tulsa, 01K, 1990, Expanded abstracts.
- Schmidt, M. W., E. Berg, M. P. Friedlander, and K. P. Murphy, 2009, Optimizing costly functions with simple constraints: A limited-memory projected quasi-Newton algorithm: International Conference on Artificial Intelligence and Statistics, 456 - 463.
- Schneider, W. A., K. A. Ranzinger, A. H. Balch, and C. Kruse, 1992. A dynamic programming approach to first arrival traveltimes computation in media with arbitrarily distributed velocities, *Geophysics*, 57, 39-50.
- Sei, A. & Symes, W.W., 1994. Gradient calculation of the traveltimes cost function without ray tracing, 64th Annual International Meeting, SEG, Expanded Abstracts, 1679.
- Sethian, J.A. & Popovici, A.M., 1999. 3-D traveltimes computation using the fast-marching method. *Geophysics*., 64, 516–523.
- Sethian, J.A., 2001. Evolution, implementation, and application of level set and fast marching methods for advancing fronts. *Journal of Computational Physics*, 169(2), 503-555.
- Sezer, O. G., 2011, Data-driven transform optimization for next generation multimedia applications: PhD thesis, Atlanta, GA, USA. (AAI3500608).
- Sezer, O., O. Guleryuz, and Y. Altunbasak, 2015, Approximation and compression with sparse orthonormal transforms: *Image Processing, IEEE Transactions on*, 24, 2328 - 2343.
- Shapiro, A., 2003, Monte carlo sampling methods, in *Stochastic Programming, Volume 10 of Handbooks in Operation Research and Management Science: North-Holland*.
- Shapiro, A., and A. Nemirovsky, 2005, On complexity of stochastic programming problems, in *Continuous Optimization: Current Trends and Applications: Springer, New York*.

- 
- Shen, Y., Biondi, B. & Clapp R., 2017. Q-model building using one-way wave-equation migration q analysis — part 1: theory and synthetic test, *Geophysics*, 83, 1-64.
- Sheng, J., Leeds, A., Buddensiek, M., & Schuster, G. T., 2006. Early arrival waveform tomography on near-surface refraction data. *Geophysics*, 71(4), U47-U57.
- Sigloch, K., McQuarrie, N., Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography. *Nature Geoscience*. 1, 458–462.
- Song, Z.M., Williamson, P.R., Pratt, R.G., 1995. Frequency-domain acoustic-wave modelling and inversion of cross-hole data. Part II. Inversion method, synthetic experiments and real-data results. *Geophysics* 60, 796–809.
- Stewart R.R. 1991. *Exploration Seismic Tomography: Fundamentals*. SEG. ISBN 1560800526.
- Stork, C. and R. W. Clayton, 1986. Analysis of the resolution between ambiguous velocity and reflector position for traveltimes tomography, *Extended Abstracts, 56th Annual International. SEG Meeting.*, 545-50.
- Sun, M., Zhang, J. & Sacchi M., 2017. Highly efficient 3D first-arrival traveltimes tomography by stochastic approximation, 83th Annual International Meeting, SEG, Expanded Abstract: 2676-2680.
- Taillandier, C., Noble, M., Chauris, H. & Calandra, H., 2009. First-arrival travel time tomography based on the adjoint state method, *Geophysics*, 74(6), WCB1 – WCB10.
- Traynin, P., Liu, J. & Reilly, J.M., 2008. Amplitude and bandwidth recovery beneath gas zones using kirchhoff prestack depth Q-migration, 74th Annual International Meeting, SEG, Expanded Abstracts, 27, 2412-2416.

- 
- Tropp, J., and A. Gilbert, 2007, Signal recovery from random measurements via orthogonal matching pursuit: *Information Theory, IEEE Transactions on*, 53, 4655 – 4666.
- Tsai, Y-H. R., Cheng, L-T., Osher, S. & Zhao, H-K., 2003. Fast sweeping algorithms for a class of Hamilton-Jacob Equations, *SIAM Journal on Numerical Analysis*, 53, 673 – 694.
- van den Berg, E., and M. P. Friedlander, 2008. Probing the pareto frontier for basis pursuit solutions: *SIAM Journal on Scientific Computing*, 31, 890–912.
- Van Leeuwen, T., Aravkin, A. Y., & Herrmann, F. J., 2011. Seismic waveform inversion by stochastic optimization. *International Journal of Geophysics*, 2011, (2011-6-26), 2011, 879-888.
- Vanelle, C. & Gajewski, D., 2002. Second-order interpolation of traveltimes. *Geophysical Prospecting*, 50, 73–83.
- Varela, C.L., Rossa A.L. & Ulrych T.J., 1993. Modeling of attenuation and dispersion, *Geophysics*, 58, 1167–11733.
- Vesnaver, A., 2010. Yardsticks for industrial tomography. *Geophysical Prospecting*, 56(4), 457-465.
- Vesnaver, A., 2013. Seismic tomography from the old to the new millennium. *Arabian Journal for Science & Engineering*, 38(1), 1-9.
- Vidale, J. E., 1990. Finite-difference calculation of traveltimes in three dimensions. *Geophysics*, 55(5), 521-526.
- Vidale, J.E., 1988. Finite-difference calculations of traveltimes. *Bulletin of Seismological Society of America*, 78, 2062–2076.

- 
- Vinje, V., E. Iversen, and H. Gjoystdal. 1993. Traveltime and amplitude estimation using wavefront construction, *Geophysics*, 58, 1157-1166.
- Waheed, U.B., Flagg, G. & Yarman, C.E., 2016. First-arrival traveltime tomography for anisotropic media using the adjoint-state method, *Geophysics*, 81, R147-R155.
- Wang, B. & Braile, L.W., 1996. Simultaneous inversion of reflection and refraction seismic data and application to field data from the northern Rio Grande rift, *Geophysical Journal International*, 125, 443–458.
- Wang, H., Singh, S.C., Jian, H. and Calandra, H., 2012. Integrated inversion of subsurface velocity structures using wave equation tomography and full waveform inversion: 82nd Annual International Meeting, SEG, Expanded Abstract.
- Wang, Y., 2002. A stable and efficient approach of inverse Q filtering, *Geophysics*, 67,657-663.
- Wang, Y., 2006. Inverse-filter for seismic resolution enhancement, *Geophysics*, 71, V51-V60.
- Ward, R.W. & Toksöz, M.N., 1971. Causes of regional variation of magnitude, *Bulletin of Seismological Society of America*, 61, 649–670.
- Weber, Z., 1995. Some improvement of the shortest path ray tracing algorithm. Springer Netherlands
- White, D.J., 1989. Two-dimensional seismic refraction tomography. *Geophysics Journal*. 97, 223–245.
- Williamson, P.R., 1990. Tomographic inversion in reflection seismology structure. *Geophysical Journal International*. 100, 255–274.

- 
- Xin, K., Hung, B., Birdus, S. & Sun, J., 2008. 3-D tomographic amplitude inversion for compensating amplitude attenuation in the overburden, 74th Annual International Meeting, SEG, Expanded Abstracts, 27, 3713.
- Xin, K., Xie, Y., He, Y., 2014. Adaptive centroid frequency shift Q tomography, Presented at 76th Annual International Conference and Exhibition, EAGE.
- Zelt, C. A., & Barton, P. J., 1998. Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe basin. *Journal of Geophysical Research: Solid Earth*, 103(B4), 7187.
- Zelt, C. A., & Smith, R. B., 1992. Seismic traveltime inversion for 2-d crustal velocity structure. *Geophysical Journal International*, 108(1).
- Zhang J, & Toksoz M. N., 1998. Nonlinear refraction traveltime tomography. *Geophysics*, 63(5):1726-1737.
- Zhang, C. & Ulrych, T.J., 2002. Estimation of quality factors from CMP records, *Geophysics*, 67, 15420.
- Zhang, J. Z., Huang, Y. Q., Song L. Q., Liu Q. H., 2015. Fast and accurate 3-D ray tracing using bilinear traveltime interpolation and the wave front group marching. *Geophysical Journal International* (3), 1327-1340.
- Zhang, J., Ten Brink, U. S., & Toksöz, M. N., 1998. Nonlinear refraction and reflection travel time tomography. *Journal of Geophysical Research*, 103(B12), 29743.
- Zhang, Y., Zhang, P. & Zhang, H., 2010. Compensating for viscoacoustic effects in reverse - time migration, 76th Annual International Meeting, SEG, Expanded Abstracts, 3160-3164.

- 
- Zhao, A., Zhang, Z. & Peng, S., 2004. Minimum travel-time tree algorithm for seismic ray tracing: improvement in efficiency. *Journal of Geophysical Engineering*, 245–251.
- Zhao, H., 2005. Fast sweeping method for eikonal equations. *Mathematics of Computation*, 74(250): 603-627.
- Zhao, H., 2007. Parallel implementations of the fast-sweeping method. *Journal of Computational Mathematics*, 25(004), 421-429.
- Zhou, B., & Greenhalgh, S. A., 2003. Crosshole seismic inversion with normalized full-waveform amplitude data. *Geophysics*, 68(4), 1320-1330.
- Zhou, H. W., Li, P., Yan, Z., & Liu, H., 2009. Constrained deformable layer tomostatics. *Geophysics*, 74(6), WCB35-WCB46.
- Zhu, L., Liu, E., & McClellan, J. H., 2015. Seismic data denoising through multiscale and sparsity-promoting dictionary learning. *Geophysics*, 80(6), WD45-WD57.
- Zhu, T., Harris, J.M. & Biondi, B., 2014. Q-compensated reverse-time migration, *Geophysics*, 79, S77-S87.
- Zhu, X., Sixta, D. P., & Angstman, B. G., 1992. Tomostatics: turning-ray tomography+ static corrections. *Leading Edge*, 11(12), 15-23.