Improving the Quality of Seismic Images by Deterministic Inversion and Machine Learning Methods

Dissertation by

Yuqing Chen

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

June, 2019

EXAMINATION COMMITTEE PAGE

The dissertation of Yuqing Chen is approved by the examination committee

Committee Chairperson: Gerard Schuster
Committee Members: Daniel Peter, Xiangliang Zhang, Sherif M. Hanafy

ABSTRACT

Improving the Quality of Seismic Images by Deterministic Inversion and Machine Learning Methods

Yuqing Chen

This thesis develops five novel methods for seismic imaging and inversion to improve both their computational efficiency and accuracy. Three of them improve the accuracy of the final inverted images by novel preconditioning strategies, and the other two are machine learning (ML) methods applied to seismic data.

1. Conventional viscoacoustic least-squares reverse time migration, also denoted as Q-LSRTM, suffers from slow-convergence and low-resolution problems due to the attenuative property of the adjoint Q propagators. To mitigate these problems, I propose a viscoacoustic deblurring filter (DF) as a preconditioner for Q-LSRTM. Moreover, to avoid the usage of the attenuative adjoint Q propagator, I propose the application of a hybrid deblurring filter to acoustic reverse time migration (RTM) images to correct for attenuation distortions. Numerical tests demonstrate that both deblurring filter strategies can produce images with higher resolution than Q-LSRTM and much cheaper in computation.

2. The deblurring filter is less effective when the migration image contains strong migration artifacts. To mitigate this problem, I develop a novel support vector machine-based (SVM) filtering method which employs the features of coherency, amplitude, and dipping angle from selected dip-angle angle-domain common-image gathers (ADCIGs) to automatically distinguish signals from artifacts. Our numerical results show that SVM filtering can efficiently remove the aliasing artifacts and improves the image quality.

3. The accurate imaging of subsurface requires the correct estimates of the velocity model. Here, I present the strategy of multiscale reflection phase inversion (MRPI) with a deblurring filter for inverting the low-wavenumber components of the velocity model. This method employs amplitude replacement, trace integration and offset-selection of traces to mitigate the cycle-skipping problem. It also uses deblurring filters as an inexpensive alternative to LSRTM to compute the perturbation image. Numerical results show that MRPI + DF can efficiently recover the low-wavenumber velocity model and is less prone to getting stuck in local minima compared to conventional reflection inversion method.

4. Non-linear inversion gets stuck in a local minimum because the data are very complex (i.e, wiggly in time), which means that the objective function is characterized by many local minima. To avoid this problem, I present a wave-equation inversion method that inverts for the subsurface velocity model from data skeletonized by a machine learning method. The skeletonized representation of the seismic traces consists of the low-rank latent-space variables predicted by a well-trained autoencoder. The input data to the autoencoder are the seismic traces, while the implicit function theorem is used to determine the formula for the Fréchet derivative used in the gradient calculation. Empirical results suggest that the cycle-skipping problem is largely mitigated by replacing the waveform differences with the latent-space parameters. We denote this method as Newtonian machine learning because it unites, for the first time, the parameter inversion of the governing equations of Newtonian physics with the dimensional reduction properties of a neural network.

ACKNOWLEDGEMENTS

I would like to thank my mentor, Prof. Gerard T. Schuster, for his teaching, encouragement, and support through my Ph.D. study at the King Abdullah University of Science and Technology (KAUST). He taught me geophysics theories in an intuitive and solid way which is unforgettable. He provides me freedom in my research directions and is lenient with my failures. His passion and rigorous attitude to science will always motivate me. I am also grateful to the members of my dissertation committee: Prof. Daniel Peter, Prof. Xiangliang Zhang and Prof. Sherif Hanafy for taking their time, patience and for their insights and suggestions which tremendously benefited my thesis.

I am grateful to Los Alamos National Laboratory (LANL), the U.S. for offering me a great internship during my Ph.D. I extremely appreciate Dr. Lianjie Huang, who is a great advisor for me at LANL. I would also like to thank the guidance and helps that I received from Dr. Kai Gao, Dr. Benxi Chi, and Dr. Yunsong Huang during my internship in LANL.

I also thank all of my colleagues from the Center for Subsurface Imaging and Modeling (CSIM) for their discussion and assistance. I benefited from my discussions with Dr. Abdullah AlTheyab, Dr. Gaurav Dutta, Dr. Bowen Guo Dr. Mrinal Sinha, Dr. Zongcai Feng, Dr. Jing Li, Dr. Lei Fu, Dr. Han Yu, Kai Lu, Zhaolun Liu, and Shihang Feng. I feel very lucky to be a CSIMer, and I appreciate the friendships that I made in the CSIM family.

Last but not least, I thank my parents for their love and support during all these years. I want to especially thank my fiancée Xi Zhang who is a beautiful, smart and independent lady. I appreciate her love, understanding, support and sacrifice for me. I am sorry I can only remotely and digitally take care of her for all these years.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

The goal of exploration seismology is to estimate the Earth's subsurface rock parameters by recording, processing and inverting seismic data Wikipedia (2019). These rock parameters will vary depending on their fluid content and lithology. If oil and gas are present, then they will often leave a marked signature in the seismic images. Our task is to find these subtle signatures.

There are two types of seismic images: the migration image is a map of the Earth's reflectivity distribution while the tomogram is a rendering of the subsurface velocity model. The former reveals the structural geometry of the layer interfaces while the latter provides information about the fluid content and lithology. My dissertation presents five novel methods I have developed that will improve the quality of these seismic images. These methods are based on either deterministic inversion or machine learning techniques, or a combination of the two.

I will first introduce three seismic imaging methods, which focus on the topics of migration of attenuate data and removing migration artifacts, respectively. In the first topic, I propose two types of deblurring filters: the first is denoted the viscoacoustic deblurring filter and the second is denoted as the hybrid deblurring filter. These deblurring filters are used to improve the image resolution of conventional Q-least square reverse time migration (Q-LSRTM), and reduce the computational costs by approximating the inverse of the viscoacoustic and hybrid Hessian operator. The computational costs of these matching filters are at least an order-of-magnitude less expensive than standard Q-LSRTM. I also show that the deblurred images of these

two deblurring filters are better in image quality than the standard Q-LSRTM. The deblurring filters are less effective when the migration image contains strong migration artifacts. To mitigate this problem, I will introduce a support vector machine (SVM) filtering method to distinguish between the signal and noise points for different dip-angle angle-domain common image gathers (ADCIGs). The advantage of the SVM filtering methods over conventional semblance filtering method is that multiple criteria can be used to distinguish signal from noise, which achieves a much higher classification accuracy.

Accurate imaging also relies on the accurate estimates of the subsurface velocity model. In this thesis, I will introduce two velocity model building methods. The first method employs the strategy of multiscale reflection phase inversion to mitigate the cycle-skipping problem of conventional reflection waveform inversion (RFWI). This strategy largely prevents the inversion from getting stuck in a local minimum by trace integration and offset selection, and focus on the phase mismatch instead of amplitude mismatch. However, the phase difference misfit function still requires the wiggle by wiggle matches which still can lead the inverted model get stuck in a local minimum. To mitigate this problem, a skeletonized representation of the data such as first-arrival traveltimes (Luo and Schuster, 1991a,b) can be inverted to obtain the low-to-intermediate wavenumber details of the background velocity model (Lu et al., 2017). In the second velocity-model building, I propose to skeletonize the data by an unsupervised machine learning method. The skeletonized representation of the seismic traces consists of the low-rank latent-space variable computed by a well-trained autoencoder neural network. Then, the implicit function theorem is used to determine the perturbation of the skeletonized data with respect to the velocity perturbation is used to compute the misfit gradient. This method, denoted as the Newtonian machine learning method, unites for the first time adjoint-based inversion methods with data skeletonized by machine learning.

1.1 Seismic Attenuation and Attenuation Compensation Migration

Problems : Ideally, the subsurface is considered as a perfectly elastic medium, so the total energy of the seismic waves is conserved as mechanical motion (Aki and Richards, 1980). However, in the Earth, seismic waves traveling in the earth experience amplitude loss and velocity dispersion associated with strong subsurface attenuation. This attenuation transforms the mechanical energy of wave propagation into the thermal energy associated with the sloshing of a viscous fluid in porous rock. These attenuation effects can be quantified by a dimensionless quantity known as the quality factor Q

$$Q = \frac{2\pi E}{\Delta E}, \tag{1.1}$$

where E represents the total energy of the seismic wave and ΔE indicates the energy loss per cycle. Therefore, a small Q value indicates the strong attenuation and a large Q value means the opposite. As an example, Figure 1.1 compares the acoustic and viscoacoustic wavefields at different propagation times with the same velocity model ($v = 2800m/s$) but different Q models ($Q = 50$ & $Q = 18$). The source wavelet excited at the center of the model is a Ricker wavelet with a peak frequency of 20 Hz. The sub-figures plot the snapshots of both acoustic and viscoacoustic wavefields. Figures 1.1a, 1.1c and 1.1e clearly show that the amplitude loss and phase shift both increase with increasing propagation distance. Moreover, Figures 1.1b, 1.1d and 1.1f show that the viscoacoustic waves in the model with $Q = 18$ suffers more from attenuation compared with the viscoacoustic waves shown in Figures 1.1a, 1.1c and 1.1e. Therefore, strong subsurface attenuation can significantly distort the amplitudes and phases of the seismic waves (Aki and Richards, 1980).

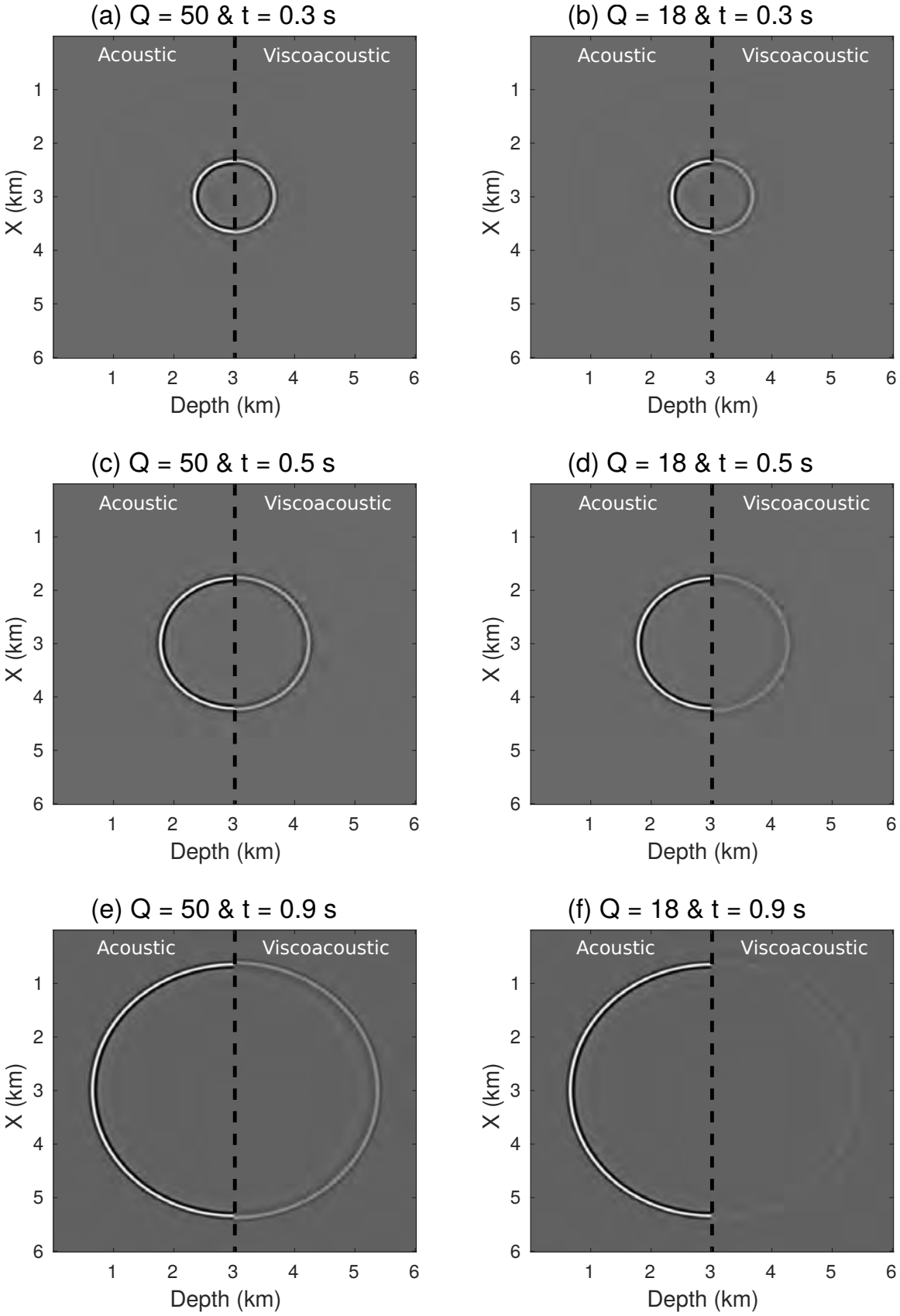


Figure 1.1: The comparison of wavefield time slices for acoustic and viscoacoustic wavefields with (left) $Q = 50$ and (right) $Q = 18$.

Conventional acoustic migration methods cannot correct for these attenuation distortions because they do not consider the anelastic nature of wave propagation. Figure 1.2a shows a poor quality migration image of an offshore Brunei dataset computed by Kirchhoff prestack depth migration (PSDM) (Gamar et al., 2015), where no attenuation effects are considered during migration. The white dotted circles indicate the areas where the gas clouds exist. It clearly shows that the seismic reflectors below these gas clouds are weak in amplitude and their resolution is low. Figure 1.2b shows a compensated image migrated by the Q-PSDM method using ray-tracing, which achieves more balanced amplitudes and better resolution by taking into account the attenuation effects (Gamar et al., 2015).

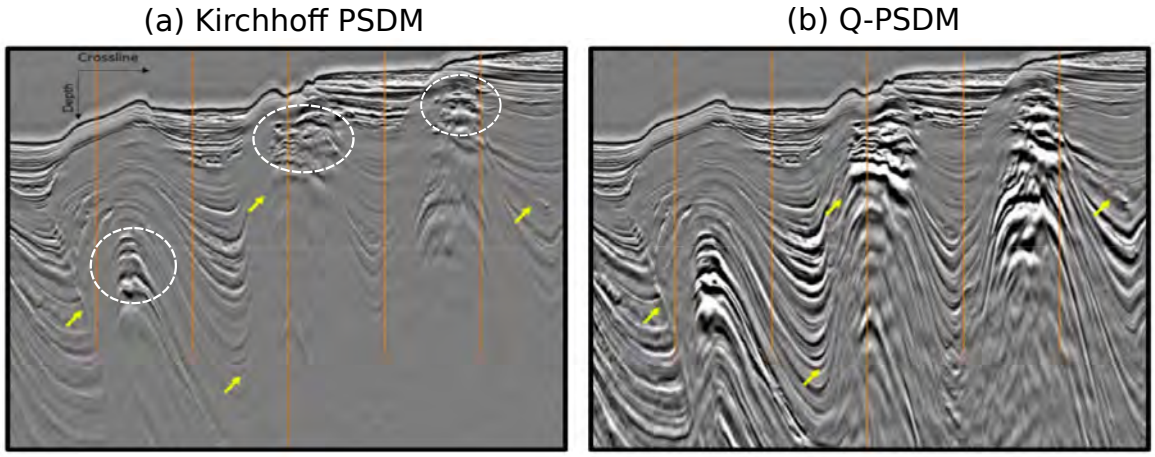


Figure 1.2: The migration image of offshore Brunei dataset computed by (a) Kirchhoff PSDM and (d) Q-PSDM (Gamar et al., 2015).

Currently, wave-equation based attenuation compensation methods are more commonly used than ray-based methods as they do not require a high-frequency approximation. For Q-reverse time migration (Q-RTM), Zhang et al. (2010), Suh et al. (2012) Zhu et al. (2014) and Zhu and Harris (2015) proposed a constant Q viscoacoustic wave equation which employs a fractional Laplacian operator to separately correct for phase and amplitude losses. Application of this equation to the lossy

data can correct the attenuation distortions by reversing the sign of the attenuation operator while preserving the sign of the dispersion operator. However, the disadvantages are that high-frequency noise will be amplified by backpropagation and the numerical solution to the viscoacoustic equation is expensive. Another method for compensating for the attenuation effects is Q least-squares reverse time migration (Q-LSRTM) (Dutta and Schuster, 2014a; Dai et al., 2015a). This method computes the source-side wavefield by the solution to the time-domain viscoacoustic wave equation derived from the standard linear solid (SLS) model. However, the adjoint Q propagator used for backpropagating the data residual is attenuative. Thus the inverted images from Q-LSRTM with a small number of iterations are often observed to have low resolution.

Solutions: To increase the resolution and accelerate the convergence of Q-LSRTM, I propose two solutions as follows:

(1) In Chapter 2, I propose a viscoacoustic deblurring filter (DF) as a preconditioner for Q-LSRTM. This viscoacoustic DF can approximate the inverse of the viscoacoustic Hessian matrix, which is estimated by matching a simulated migration image to its reference reflectivity model. Numerical tests on synthetic and field data demonstrate that Q-LSRTM combined with viscoacoustic DFs can produce images with higher resolution and more balanced amplitudes than images from Q-LSRTM when there is strong attenuation in the background medium. In addition, this filter significantly speeds up the convergence of Q-LSRTM at least by a factor of 2. The results of this Q-LSRTM + viscoacoustic DF approach are presented in Chapter 2, and it is published in Geophysics (Chen et al., 2017).

(2) In Chapter 3, I propose applying a hybrid deblurring filter to acoustic reverse time migration (RTM) instead of directly deblurring Q-LSRTM images. The

advantage of this approach is that it avoids the attenuative characteristics of attenuative backpropagation and produces higher-resolution migration image than Q-LSRTM. Moreover, it only requires one solution to the viscoacoustic wave equation compared to more than a dozen such solutions for iterative Q-LSRTM. This method is more than an order-of-magnitude reduction in computational costs.

I also compare the performance between the hybrid and viscoacoustic DF in terms of both image quality and computational cost. For weak attenuation, the performance of the viscoacoustic DF is about 5% to 10% better than the hybrid approach. However, as the attenuation increase, the performance of the two methods become similar. However, the hybrid DF approach is at least 30% faster and saves at least 1/3 of the storage space compared to the viscoacoustic DF approach. The results of this method are presented in Chapter 3, and it is published in Geophysics in (Chen et al., 2019).

1.2 Migration Artifacts

Problems : Both migration and deblurring filters tend to behave poorly when the migration image contains strong migration artifacts. Such migration artifacts are generated if the source and receiver distributions are sparse. These artifacts degrade amplitudes and reduce an interpreter's ability to identify important geological features of interest. As an example, Figure 1.3a shows a migration image of a marine dataset from offshore West Africa (Dutta et al., 2017), where the image quality is severely degraded by strong artifacts. The geology structures at the depth between 3.0 km to 5.0 km are especially hard to identify. Figure 1.3b shows the migration image after removing the artifacts where some important geology structures are revealed.

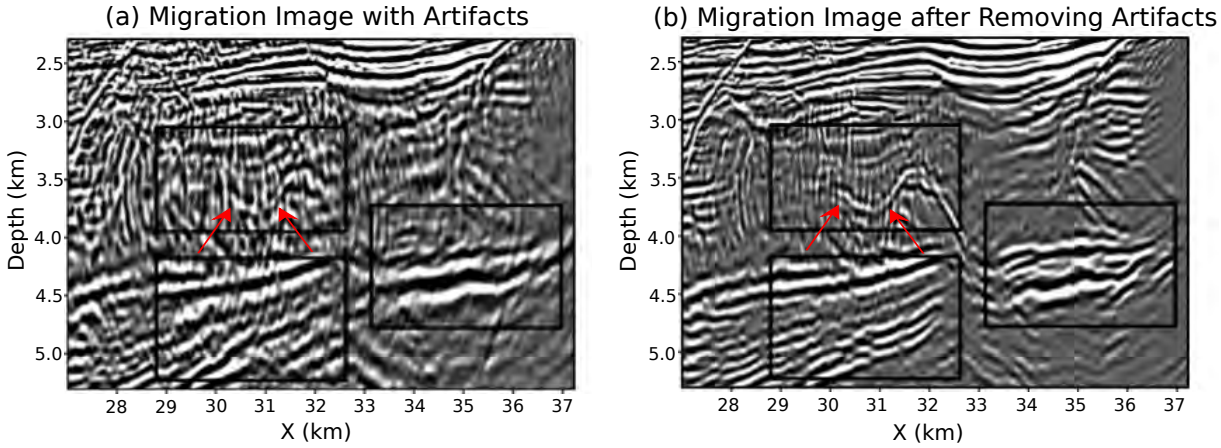


Figure 1.3: The migration image (a) with artifacts and (b) after removing artifacts (Dutta et al., 2017).

To alleviate this problem, least-square reverse time migration (LSRTM) with preconditioning or regularization is commonly used to suppress these artifacts (Cabrales-Vargas and Marfurt, 2013; Dutta and Schuster, 2015; Lin et al., 2016; Dutta, 2017). However, least-squares-based methods require a large number of expensive iterations to enhance image quality. Dafni and Symes (2016a,b) and Liu and Zhang (2018) designed a specularity filter based on the semblance equation to preserve the signal and suppress the migration artifacts in the dip-angle angle-domain common image gather (ADCIG). This assumes that the valid signal has a higher semblance score but the artifact has a small score. However, this assumption fails for an image with coherent artifacts.

Solution: In Chapter 4, I propose a supervised support vector machine (SVM) filtering method to remove the coherent noise in the dip-angle ADCIGs in order to reduce the migration artifacts. The input data vector for the SVM contains the skeletal features of selected dip-angle ADCIGs such as coherency, amplitude, and dip-angle to distinguish the migration signal from coherent noise. The results show that this SVM approach achieves a much higher classification accuracy than the

conventional semblance approach. This supervised machine learning method only uses 1% to 10% of the points in the selected dip-angle ADCIGs and requires just minutes of training time. The results of applying this method to marine data are shown in Figure 1.4. Here, the strong artifacts in Figure 1.4a are mostly removed by the SVM filtering method. This work is now submitted to Geophysics for review.

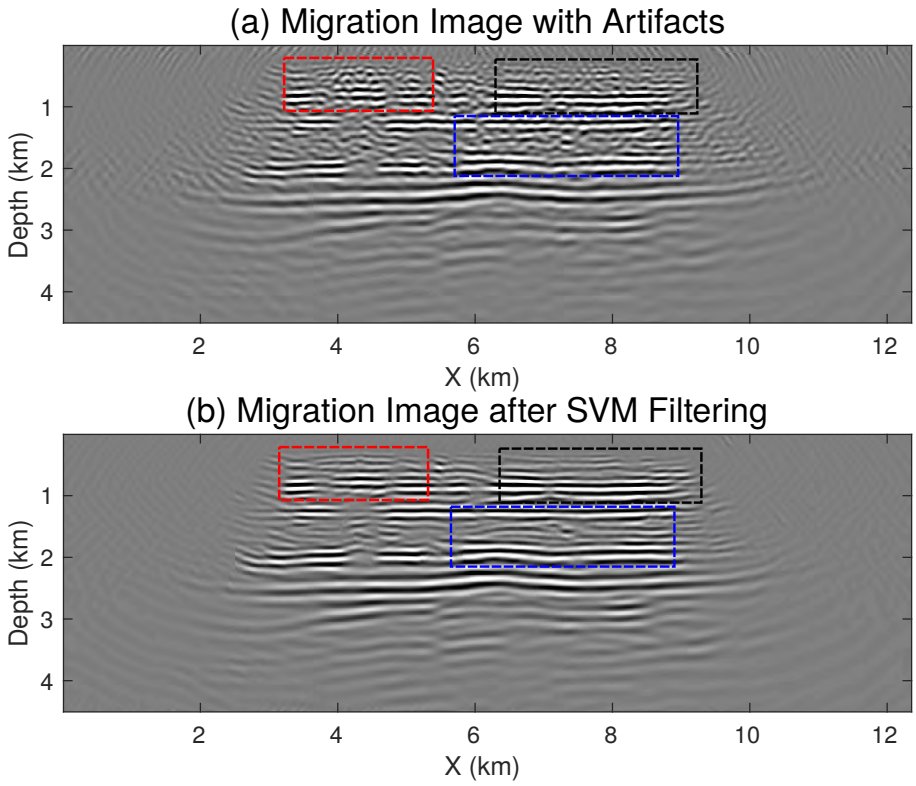


Figure 1.4: Migration image of North Sea data (a) with artifacts and (b) after SVM filtering.

1.3 Estimates of the Background Velocity Model

Problems: Accurate migration of seismic reflections relies on the accuracy of the estimated velocity model. To invert for a reliable velocity model below the penetration depth of diving waves, reflection waveform inversion is the method of choice. In this method, reflection full waveform inversion (RFWI) separates the subsurface velocity v into a background model (low-wavenumber component) v_0 and a perturbation model

(high-wavenumber component) δv (Xu et al., 2012). The background model and the perturbation model largely determine the kinematic and dynamic characteristics of the reflections, respectively. These two components are alternately updated at each iteration. For example, RFWI first computes the perturbation model, also denoted as the migration image, using the current background model. The calculated migration image is then used for Born modeling of the reflection data. The differences between the Born modeled data and the observed data are computed, which are then back-propagated to generate the source- and receiver reflection wavepaths for updating the background velocity model (Xu et al., 2012; Wang et al., 2013; Brossier et al., 2015).

Conventional RFWI requires a least-squares reverse time migration (LSRTM) image with balanced amplitudes and high-resolution interfaces. However, LSRTM is computationally expensive because it requires a large number of iterations to get the desired uplift in image quality. Moreover, conventional RFWI uses the waveform-difference misfit function which is more likely to fall into a local minimum. The reason is that the waveform-difference misfit function is highly nonlinear and is very sensitive to the amplitude mismatches between the synthetic and the observed data.

Solution: In Chapter 5, I develop a multiscale reflection phase inversion with deblurring filters (DFs) method to mitigate the problems of getting stuck in a local minimum and the high-computational cost of conventional RFWI. This method partly relieves the heavy computation costs of RFWI by using the DFs as an inexpensive alternative to LSRTM. To mitigate the cycle-skipping problem, MRPI uses a multiscale approach which combines the trace integration, amplitude replacement, and offset selection techniques to remove the cycle-skipped traces and focus the inversion on inverting phases instead of amplitudes. I denote this method as MRPI + DF. Figure 1.5 compares the migration images of marine data using the inverted tomograms from conventional RFWI and MRPI + DF, respectively. It is clear that the image computed using the MRPI + DF tomogram reveals many more details compared to

conventional RFWI migration. To confirm this observation, Figure 1.6 shows the common image gather in the suboffset domain. The most accurate velocity model should be the one which focuses the reflections on the zero-suboffset region. It is obvious that the MRPI + DF method achieves this goal much more than the conventional RFWI method. This work was submitted to Geophysics and the paper was accepted subject to minor revision in April of 2019.

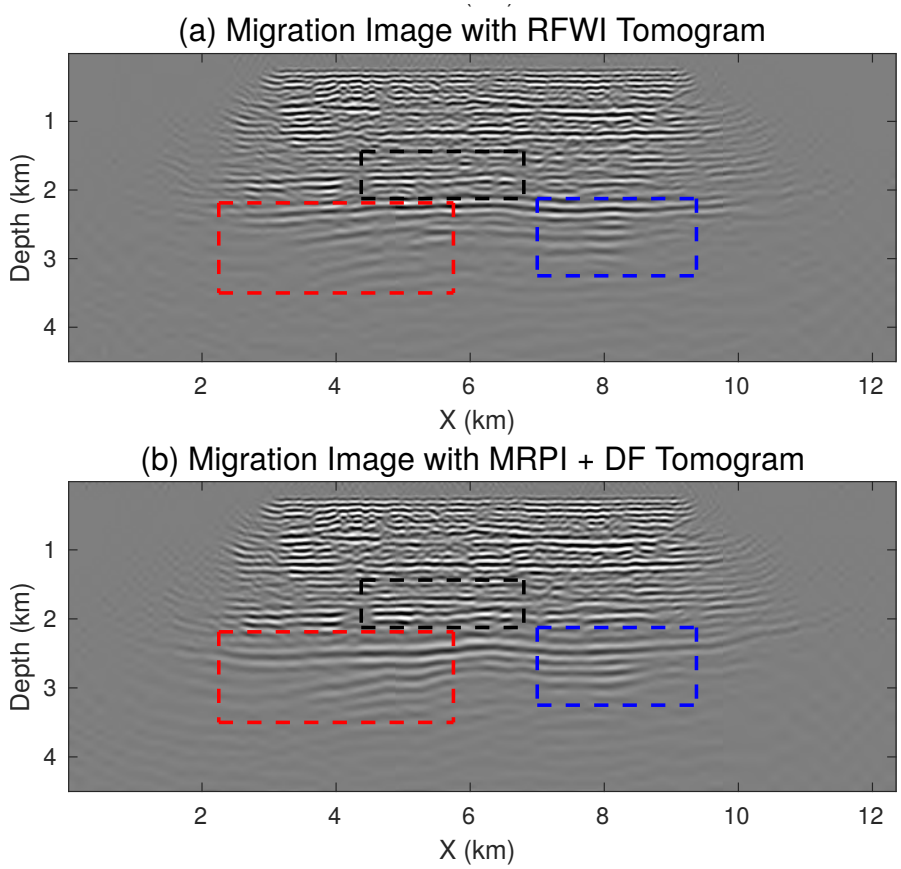


Figure 1.5: Comparisons of the migration images computed from the (a) RFWI tomogram and (b) MRPI + DF tomogram.

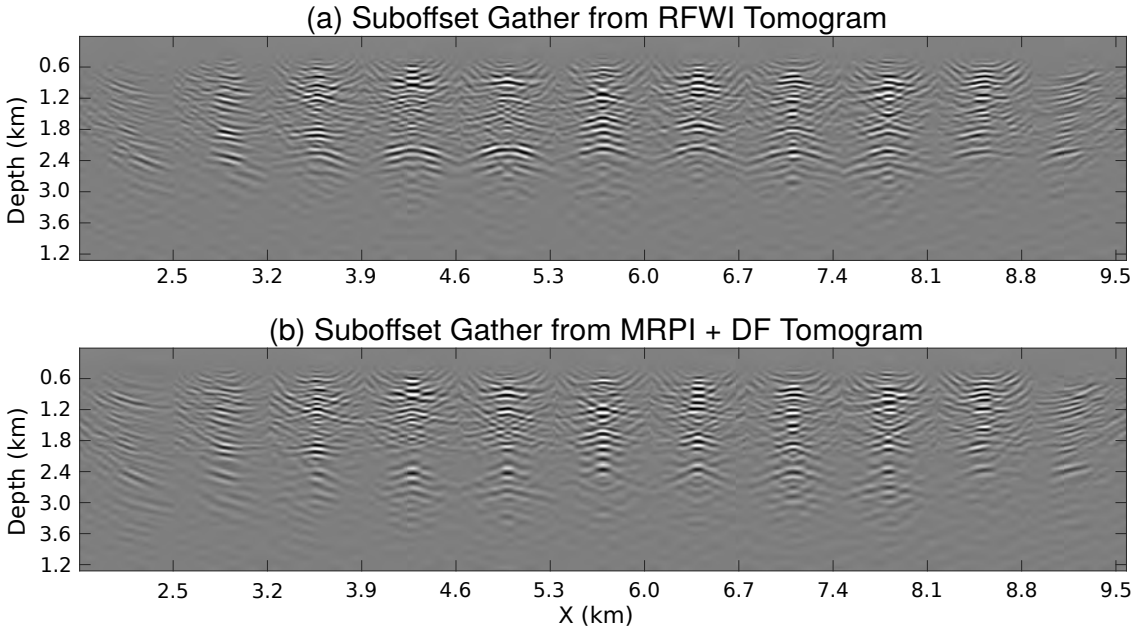


Figure 1.6: Comparisons of the suboffset common image gather images computed from the (a) RFWI tomogram and (b) MRPI + DF tomogram.

1.4 Why Skeletonized Inversion?

Problems: The waveform-difference misfit function is both a blessing and a curse for seismic inversion. On one hand, by minimizing the wiggle to wiggle difference between the observed and predicted data, full waveform inversion (FWI) method can, theoretically, recover a high-resolution estimate of the subsurface velocity model. On the other hand, the waveform-difference misfit function is highly nonlinear and is characterized by multiple minima. Therefore, the success of FWI heavily relies on a good initial model that is close to the true model, otherwise, the cycle-skipping problem will lead FWI into a local minimum (Bunks et al., 1995a). Moreover, coherent noise in the traces, and insufficient consideration of physics in the wave equation will lead to waveform mismatches between the observed and synthetic data.

To avoid these problems, Luo and Schuster (1991a,b) proposed a skeletonized inversion method which combines the skeletonized representation of seismic data with the implicit function theorem to accelerate convergence to the vicinity of the global

minimum Lu et al. (2017). Figures 1.7a shows the local minima associated with the waveform misfit function. In contrast, the much simpler traveltime misfit function has just one minimum that facilitates fast convergence by an iterative gradient descent method. Therefore, simplification of the data by skeletonization reduces the complexity of the misfit function and reduces the number of local minima (Luo and Schuster, 1991a; Feng and Schuster, 2016; Dutta and Schuster, 2016; Li et al., 2016, 2017; Liu et al., 2018). Inversion of the skeletonized data can be fast and robust, and lead to a velocity model with intermediate resolution. This velocity model can then be used as a starting model for RFWI or FWI. However, one of the key problems with skeletonized inversion is that the skeletonized data must be picked from the original data, which can be labor intensive for large data sets. For higher model resolution, several skeletonized parameters per trace must be tediously identified and inverted, as long as they are sensitive to changes in the velocity model.

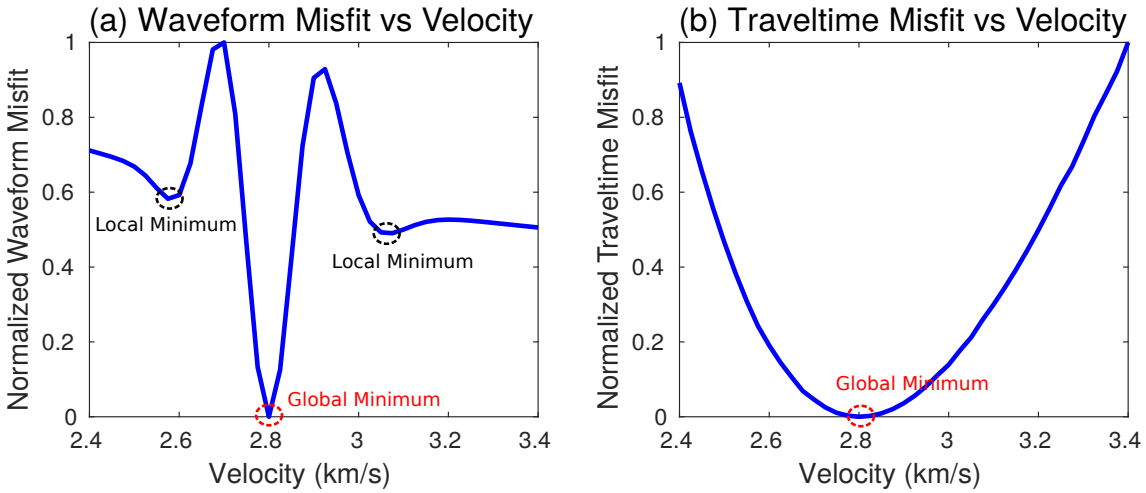


Figure 1.7: The (a) waveform and (b) traveltime misfit functions.

Solution: In Chapter 6, I propose obtaining the skeletonized data from an autoencoder to avoid the manual picking of traveltimes, and then use solutions to the wave equation to invert such data for the model of interest. The strategy I propose is illustrated in Figure 1.8 where the seismic data are inserted into an autoencoder. The

latent variable in the center layer for the observed and predicted seismic data are then used to form a misfit function. The implicit function theorem is then used to derive the formula for the Fréchet derivate, which can then be used to compute the misfit gradient for velocity inversion. The gradient employs solutions to the acoustic wave equation so there is no high-frequency approximation. I denote this method as Newtonian machine learning because it unites, for the first time, the deterministic PDEs of Newtonian physics with the dimensional reduction properties of a neural network. This methodology can be used to account for the physics of any classical physics phenomena such as gravity, electromagnetic wave propagation, magnetism, seismic wavefields, and thermal fields.

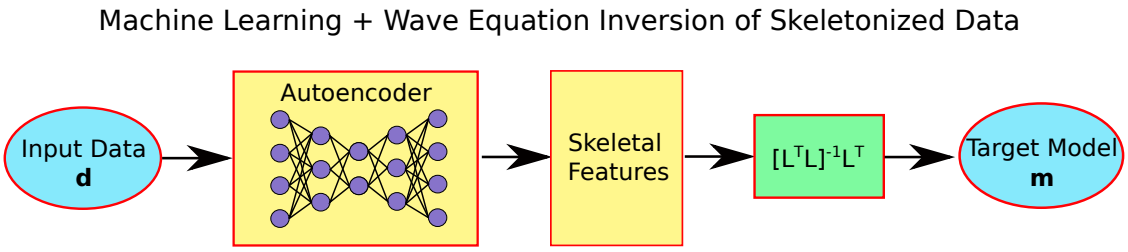


Figure 1.8: The strategy for inverting the skeletonized latent variables.

As an example, Figures 1.9a and 1.9b show a true velocity model which has three high-velocity horizontal layers and initial model. The inverted velocity tomogram is shown in Figure 1.9c, where the three layers are mostly well recovered. This work is now published in the online journal arXiv.

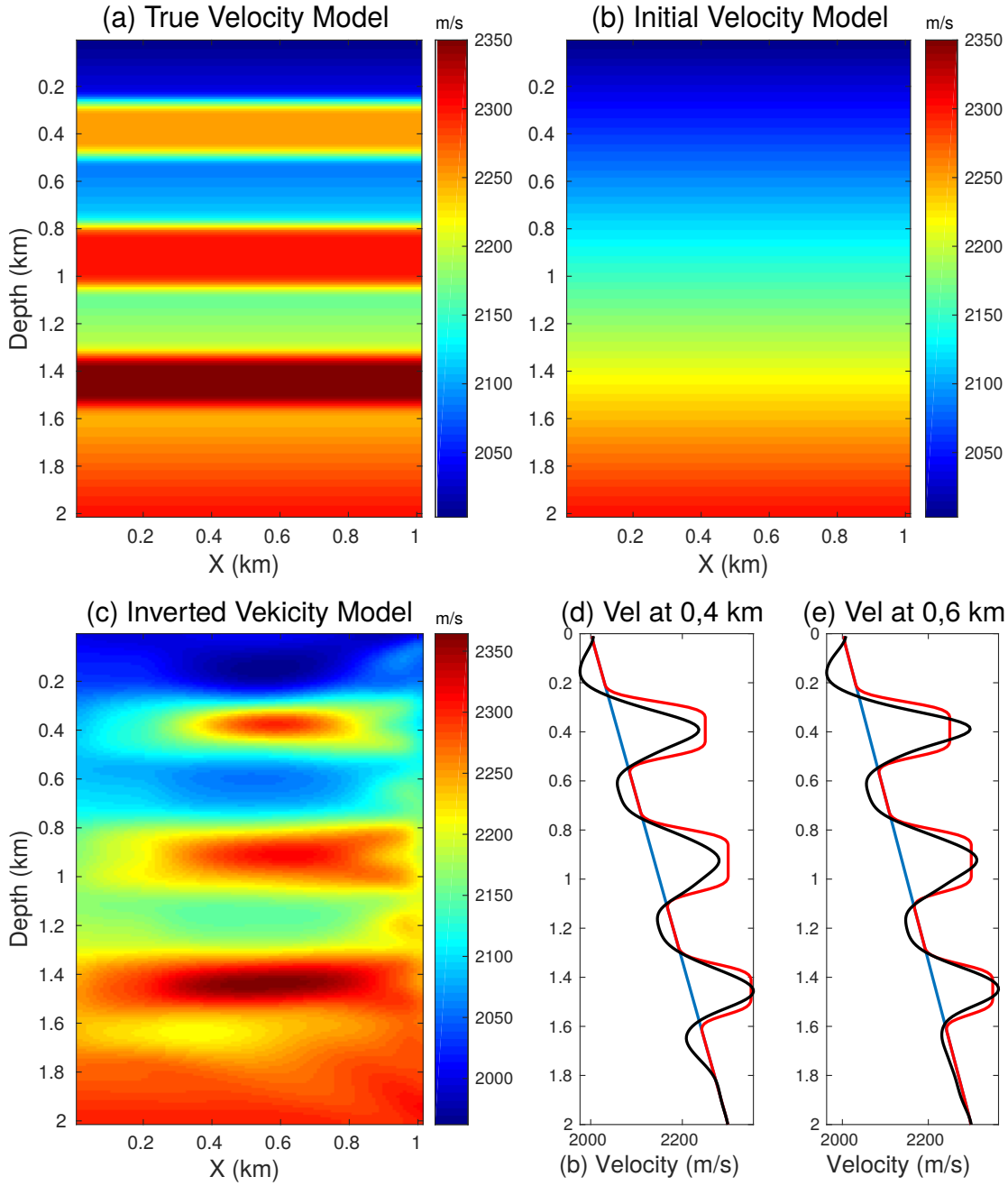


Figure 1.9: The (a) true velocity model, (b) linear increasing initial model, (c) inverted velocity and the comparison of their vertical profiles at (d) x=0.4 km and (e) x=0.6 km.

1.5 Technical Contributions

The main technical contributions of this dissertation are summarized below.

• Chapter 2: A Q-LSRTM + viscoacoustic deblurring filter (DF) method is developed to improve the image resolution and accelerate the convergence of Q-LSRTM by approximating the inverse of the viscoacoustic Hessian matrix. The viscoacoustic DF successfully mitigates the low-resolution and slow-convergence problem of Q-LSRTM and speeds up 2D Q-LSRTM imaging by a factor of 2.

• Chapter 3: An acoustic RTM + hybrid DF method is developed to avoid computing the attenuative adjoint Q propagator for the field data. The lossy data are migrated by acoustic RTM and the attenuation distortions are then corrected by the hybrid DFs. This hybrid approach produces a high-resolution image and significantly reduces computation cost and storage requirements compared to Q-LSRTM. This procedure requires less than 1/3 of the storage space and is $O(N-1)$ times faster compared to Q-LSRTM.

• Chapter 4: A supervised support vector machine (SVM) filtering method is developed to remove the coherent noise in the dip-angle ADCIGs. The SVM filtering method employs multi-skeletal features to distinguish the valid signal from coherent noise in the dip-angle ADCIG. Numerical tests show that SVM filtering has a much higher classification accuracy when compared with the conventional semblance approach.

• Chapter 5: A multiscale reflection phase inversion (MRPI) with DF strategy is developed to estimate the subsurface background velocity model. This method relieves the heavy computation costs of RFWI by using the DFs as an inexpensive alternative to LSRTM. To mitigate the cycle-skipping problem, MRPI uses a multiscale approach which combines the trace integration, amplitude replacement, and offset selection techniques together to remove the cycle-skipped traces.

• Chapter 6: A wave-equation inversion method is developed that inverts skeletonized data for subsurface velocity model. The skeletonized representation of the seismic data is the low-rank latent-space variables predicted by a well-trained au-

toencoder neural network. The inputs to the autoencoder are the recorded seismic traces, and the implicit function theorem is used to determine the perturbation of the skeletonized data with respect to the velocity perturbation. This provides a general framework for using solutions to any governing PDE to invert skeletal data generated by any type of neural network.

- Chapter 7: I summarize my results and discuss both the advantages and limitations of my proposed methods. I also suggest future research topics based on the results of this dissertation.

Chapter 2

Q-least-squares Reverse Time Migration with Viscoacoustic Deblurring Filters

Viscoacoustic least-squares reverse time migration, also denoted as Q-LSRTM, linearly inverts for the subsurface reflectivity model from lossy data. It can compensate for the amplitude loss in the migrated images due to strong subsurface attenuation and can produce reflectors that are accurately positioned in depth. However, the adjoint Q propagators used for backward propagating the data residual is attenuative which causes the low-resolution problem of Q-LSRTM. To increase the resolution and accelerate the convergence of Q-LSRTM, we use the viscoacoustic deblurring filters (DFs) as a preconditioner for Q-LSRTM. These filters can be estimated by matching a simulated migration image to its reference reflectivity model. Numerical tests on synthetic and field data demonstrate that Q-LSRTM combined with viscoacoustic DFs can produce images with higher resolution and more balanced amplitudes than images from acoustic RTM, acoustic LSRTM, and Q-LSRTM when there is strong attenuation in the background medium. The viscoacoustic deblurring filter also greatly improves the convergence rates of Q-LSRTM.

2.1 Introduction

Migration deconvolution (MD) is used to deblur migration images corrupted by acquisition footprint, strong velocity contrasts, and uneven subsurface illumination (Hu and Schuster, 1998; Hu et al., 2001). The migration Green's function (Schuster and Hu,

2000), sometimes denoted as a point spread function (PSF) (Jansson and Richardson, 1997), can be efficiently computed by invoking a layered-medium assumption (under this assumption, the migration response can be horizontally invariant) localized around the trial image point. Therefore, the migration image can be approximated as a lateral convolution between the reflectivity model and the migration Green's function. Thus, the MD operation can be used as an approximation to the inverse of the Hessian to deblur the migration image.

Hu and Schuster (1998) and Hu et al. (2001) designed a MD operator in the space-wavenumber domain to suppress the migration artifacts of poststack migration. Yu et al. (2006) extended the application of MD from poststack to prestack depth migration. Instead of using MD in the space-wavenumber domain, Guitton (2004) approximated the inverse of the Hessian in the space domain with a bank of matching filters. These filters are similar to deblurring filters (DFs) and have been used as a preconditioner for conventional least-squares migration (LSM) (Aoki and Schuster, 2009). For multisource LSM, Dai and Schuster (2009) and Dai et al. (2011) used DFs to reduce the crosstalk noise and accelerate the convergence of multisource LSM.

The previous work on MD assumed a lossless background medium. However, strong subsurface attenuation can significantly distort the amplitudes and phases of seismic waves (Aki and Richards, 1980). In this case, conventional acoustic reverse time migration (RTM) and least-squares reverse time migration (LSRTM) cannot completely correct for the attenuation loss.

To account for attenuation, Blanch and Symes (1994) and Blanch and Symes (1995) used the adjoint state method and proposed a linearized inversion method to recover the short wavelength model components in a viscoacoustic medium in the $\tau - p$ domain. Ribodetti et al. (1995) developed an analytical inversion in the frequency domain for viscoacoustic media using the Born approximation and asymptotic solutions for Green functions. Dai and West (1994), Yu et al. (2002), Wang (2007) and Valen-

ciano et al. (2011) used the one-way wave-equation migration in the frequency domain for attenuation compensation. For reverse time migration, Zhang et al. (2010), Suh et al. (2012), Fletcher et al. (2012), Zhu et al. (2014) and Zhu and Harris (2015) proposed different viscoacoustic wave equations with separate controls over phase and amplitude to compensate for the attenuation loss. Viscoacoustic least-squares reverse time migration, also denoted as Q-LSRTM, has been shown to compensate for the attenuation loss and produce images with more balanced amplitudes and accurately positioned reflectors than standard migration techniques (Dutta and Schuster, 2014b; Dai et al., 2015b; Sun et al., 2016). However, the inverted images from Q-LSRTM sometimes have lower resolution when compared to the benchmark acoustic LSRTM images. This is because the adjoint Q propagators used for backpropagating the data residual are also attenuative. Hence, a large number of least-squares iterations are required to get the desired uplift in the image quality, which makes the Q-LSRTM technique computationally expensive when compared to standard RTM.

To mitigate these problems, we propose using viscoacoustic DFs as a preconditioner for Q-LSRTM. A reference reflectivity model is first constructed using a uniform distribution of point scatterers while the background velocity and Q models are kept the same. The viscoacoustic data generated from these background models are then migrated by viscoacoustic reverse time migration (Q-RTM) to obtain a reference migration image. The viscoacoustic DFs are then estimated for different parts of the migration image using local matching filters to transform the simulated migration image into its reference reflectivity model. These filters are then used as a preconditioner during the Q-LSRTM iterations. The estimation of these local filters can be done in parallel using the Message Passing Interface (MPI) language, which makes the preconditioning process very efficient. Numerical results with these DFs show a much faster convergence rate and a much improved image for Q-LSRTM. These benefits only require the extra computational cost of constructing the viscoacoustic

DFs, which is no more than one migration of the data.

This chapter is divided into four sections. After the introduction, the second section presents the theory of Q-LSRTM with viscoacoustic DFs. Numerical tests on synthetic and field data are then used to demonstrate the advantages of the proposed preconditioning method. The conclusions are in the last section.

2.2 Theory

In an acoustic medium, the forward modeling of the data can be represented mathematically as

$$\mathbf{d} = \mathbf{L}\mathbf{m}_0, \tag{2.1}$$

where $\mathbf{L}$ represents a linear modeling operator, $\mathbf{d}$ is the data vector, and $\mathbf{m}_0$ is a vector that represents the subsurface reflectivity model. The migration image $\mathbf{m}_{mig}$ is computed by applying the migration operator $\mathbf{L}^T$ to the observed data to give

$$\mathbf{m}_{mig} = \mathbf{L}^T\mathbf{d} = \overbrace{\mathbf{L}^T\mathbf{L}}^{blurring\ operator}\ \mathbf{m}_0, \tag{2.2}$$

where $\mathbf{L}^T$ denotes the adjoint of the forward modeling operator $\mathbf{L}$. Since the adjoint operator $\mathbf{L}^T$ in equation 2.2 is not the inverse of the forward modeling operator, the computed migration image $\mathbf{m}_{mig}$ is a blurred version of the true reflectivity model $\mathbf{m_0}$. Here, the blurring operator is defined as the $\mathbf{L}^T\mathbf{L}$ operator in equation 2.2. The blurring operator blurs the true reflectivity model $\mathbf{m}_0$ to give the migration image $\mathbf{m}_{mig}$, which often suffers from artifacts because of uneven illumination. The blurring effects can be mostly corrected by applying the inverse of $\mathbf{L}^T\mathbf{L}$ to the migration image as

$$\mathbf{m}_0 = \overbrace{(\mathbf{L}^T\mathbf{L})^{-1}}^{deblurring\ operator}\ \mathbf{m}_{mig}. \tag{2.3}$$

However, computing the direct inverse of $\mathbf{L}^T\mathbf{L}$ is computationally prohibitive for practical seismic imaging problems. One possible solution is iterative least-squares migration (Lailly, 1983; Nemeth et al., 1999), which typically requires an order-of-magnitude more computations than standard migration. To accelerate the convergence, deblurring operators can be designed to approximate the inverse of the Hessian operator $(\mathbf{L}^T\mathbf{L})^{-1}$, which can then be used as a preconditioning operator (Hu et al., 2001; Guitton, 2004; Yu et al., 2006; Aoki and Schuster, 2009; Dai and Schuster, 2009).

Guitton (2004) and Aoki and Schuster (2009) proposed localized DFs which deblur the migration image in the space domain. The DFs are estimated using a reference model and its migration image. Following Aoki and Schuster (2009), the reference reflectivity model is constructed using a uniform distribution of point scatters as shown in Figure 2.1a. The reference data $\mathbf{d}_{ref}$ are generated from these reference reflectivity and background velocity models, which are then migrated to get a reference migration image $\mathbf{m}_{mig\_ref}$ as

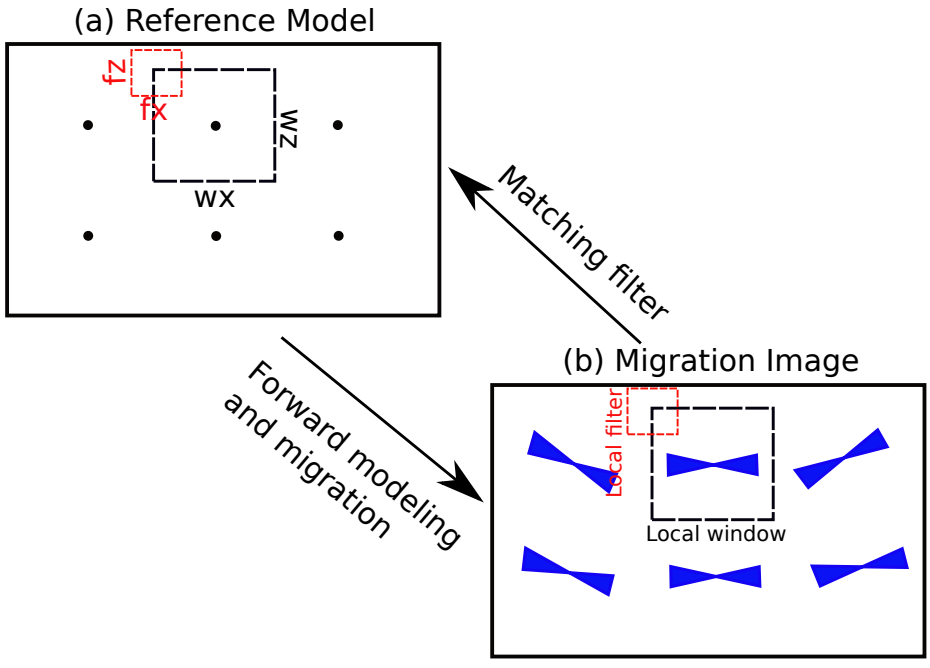


Figure 2.1: Illustration of local matching filters that transform the migration image to its reference model.

$$\mathbf{m}_{mig\_ref} = \mathbf{L}^T\mathbf{L}\mathbf{m}_{ref} = \mathbf{L}^T\mathbf{d}_{ref}. \tag{2.4}$$

As shown in Figure 2.1, $\mathbf{m}_{ref}$ and $\mathbf{m}_{mig\_ref}$ are then divided into subdomains. A subdomain or a local window is chosen such that a point scatterer is at the center of the window and the size of the window is large enough to cover the main part of the migration butterfly (Hu et al., 2001). Within the local window, the migration Green's function can be assumed to be approximately invariant with respect to the scatterer's location. The DF for each local window is then estimated by locally matching the reference migration image with the true reference reflectivity model as (Aoki and Schuster, 2009):

$$m_{ref}(x,y,z)_i = \int_{V_0} F(x-x_0, y-y_0, z-z_0)_i \, m_{ref\_mig}(x_0,y_0,z_0)_i \, dV_0, \tag{2.5}$$

which can also be written in the matrix-vector notation:

$$[\mathbf{m}_{ref}]_i = [\mathbf{F}]_i \otimes [\mathbf{m}_{ref\_mig}]_i, \tag{2.6}$$

where $\otimes$ denotes spatial convolution, and i indicates the ith local window. Here, $[\mathbf{F}]_i$, $[\mathbf{m}_{ref}]_i$ and $[\mathbf{m}_{ref\_mig}]_i$ denote the DF, the reference reflectivity model and the reference migration image within the ith local window, respectively. Similar to the localized migration Green's function, each DF is also shift invariant within its local window. Since the convolution operation is commutative, we can write

$$[\mathbf{F}]_i \otimes [\mathbf{m}_{ref\_mig}]_i = [\mathbf{m}_{ref\_mig}]_i \otimes [\mathbf{F}]_i. \tag{2.7}$$

To numerically estimate the DF, equation 2.6 needs to be transformed into matrix

multiplication form

$$[\mathbf{M}_{ref\_mig}]_i[\mathbf{f}]_i = [\mathbf{m}_{ref}]_i, \tag{2.8}$$

which requires transforming the ith reference migration image $[\mathbf{m}_{mig\_ref}]_i$ (shown as the black dashed square in Figure 2.1) into a convolution matrix $\mathbf{M}_{ref\_mig}$ with a size of $M \times N$. Also, the corresponding ith DF (shown as the red dash square in Figure 2.1) and the reflectivity model needs to be flattened as a vector $\mathbf{f}_i$ of size $N \times 1$ and a vector of $\mathbf{m}_i$ of size $M \times 1$, respectively. Here, $N = fx \times fz$ and $M = wx \times wz$, where fx and fz are the filter lengths in the horizontal and vertical directions, respectively, and wx and wz are the window lengths in the horizontal and vertical directions, respectively. This transformation is shown in 2.2, where the yellow migration "butterfly" indicates the migration response $[\tilde{\mathbf{m}}_{mig\_ref}]_i$ of a point scatterer in the ith local window. The red square with the dimensions $win_x \times win_z$ represents the local window which should effectively cover the migration butterfly. The blue square with the dimensions $f_x \times f_z$ represent the DF. In principle, the size of the filter should entirely cover the migration butterfly. However, in that case, a large DF is needed which dramatically increases the computational cost. To reduce this cost we use a DF that is no larger than several dominant wavelengths in the migrated data, so this DF mitigates high-and intermediate-wavenumber artifacts, but not the low-wavenumber one. The DF can be estimated by solving equation 3.8 by the least-squares method: (Aoki and Schuster, 2009; Dai and Schuster, 2009)

$$[\mathbf{M}_{ref\_mig}]_i^T[\mathbf{M}_{ref\_mig}]_i[\mathbf{f}]_i = [\mathbf{M}_{ref\_mig}]_i^T[\mathbf{m}]_i, \tag{2.9}$$

and the estimated filters can then be applied to the real migration image $\mathbf{m}_{mig}$ to remove the blurring effects:

$$[\mathbf{m}_{deblur}]_i \approx [\mathbf{F}]_i \otimes [\mathbf{m}_{mig}]_i. \tag{2.10}$$

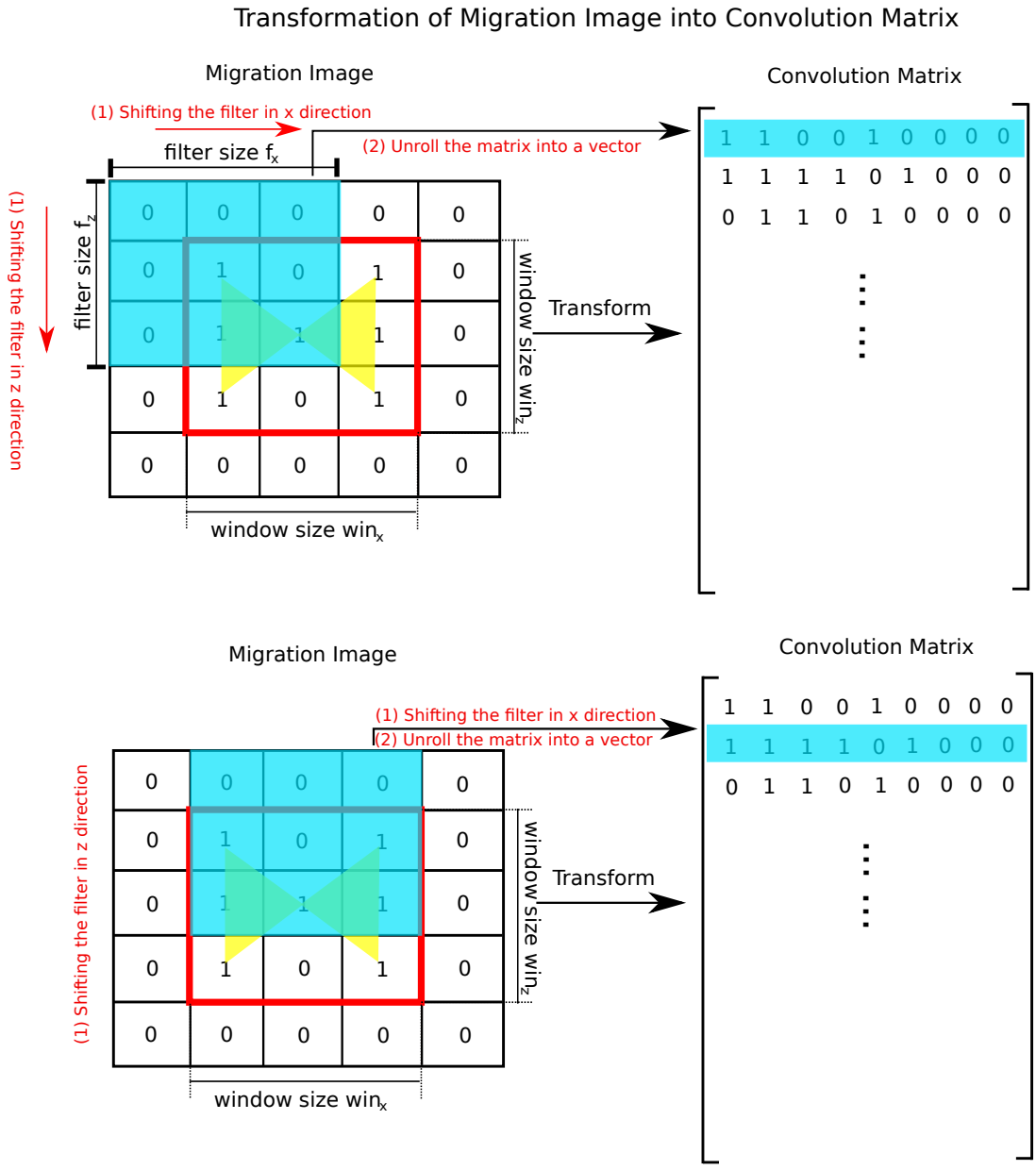


Figure 2.2: Migration response of a point scatterer in the ith local window and the convolution matrix associated with the image.

2.2.1 Deblurring filters in a viscoacoustic medium

Previous DF research concentrated on lossless media. However, strong subsurface attenuation can significantly distort the amplitudes and phases of seismic waves (Aki and Richards, 1980). To mitigate this problem, Q-LSRTM (Dutta and Schuster, 2014b) was developed to generate migration images with more balanced amplitudes and accurately positioned reflectors than standard migration techniques. However, the inverted images from Q-LSRTM sometimes tend to have lower resolution when compared to the benchmark acoustic LSRTM images using lossless data because the adjoint Q propagators used for backward propagating the data residual are also attenuative.

This loss of resolution can be explained by analyzing the migration Green's function for a viscoacoustic medium. For a homogeneous medium with velocity v_0 and a monochromatic point source at $\mathbf{x_s} = (x_s, z_s)$ with angular frequency ω, the acoustic Green's function $G(\mathbf{x}, \mathbf{x_s})$ is given by (Schuster, 2017)

$$G(\mathbf{x}, \mathbf{x}_s) = \frac{\exp\{i\omega\frac{|\mathbf{x}_s - \mathbf{x}|}{v_0}\}}{|\mathbf{x}_s - \mathbf{x}|}. \tag{2.11}$$

If the medium is lossy, the viscoacoustic Green's function can be derived by replacing the acoustic phase velocity v_0 with the complex phase velocity given by (Aki and Richards, 1980):

$$v(\omega) = v_0[1 + \frac{1}{\pi Q}\ln\frac{\omega}{\omega_0}](1 - \frac{i}{2Q}), \tag{2.12}$$

where Q is the quality factor used to quantify the attenuation in the subsurface and ω_0 is the reference frequency. After substituting equation 2.12 into equation 2.11, we get the viscoacoustic Green's function as

$$G(\mathbf{x},\mathbf{x}_s)=\frac{1}{|\mathbf{x}_s-\mathbf{x}|}\overbrace{\exp\{\frac{i\omega|\mathbf{x}_s-\mathbf{x}|}{v_0\xi}\}}^{phase\ distortion}\quad\overbrace{\exp\{-\frac{\omega|\mathbf{x}_s-\mathbf{x}|}{2Qv_0\xi}\}}^{amplitude/frequency\ attenuation}\quad, \tag{2.13}$$

where $\xi=[1+\frac{1}{\pi Q}(ln(\frac{\omega}{\omega_0}))][1+\frac{1}{4Q^2}]$. The first exponential term is the phase distortion term and the second exponential term represents the amplitude/high-frequency loss term. Therefore, the acoustic migration Green's function are given by

$$\Gamma_{acou}=\int_\omega d\omega\sum_s\sum_r\frac{\exp\{i\frac{\omega}{v_0}(\frac{rr}{\xi}-rr_0)\}\exp\{-\frac{\omega}{2Qv_0\xi}(rr)\}}{|\mathbf{x_s}-\mathbf{x}|\cdot|\mathbf{x_r}-\mathbf{x}|\cdot|\mathbf{x_s}-\mathbf{x_0}|\cdot|\mathbf{x_r}-\mathbf{x_0}|}, \tag{2.14}$$

and the viscoacoustic migration Green's function can be approximated by

$$\Gamma_Q=\int_\omega d\omega\sum_s\sum_r\frac{\exp\{i\frac{\omega}{v_0\xi}(rr-rr_0)\}\exp\{-\frac{\omega}{2Qv_0\xi}(rr+rr_0)\}}{|\mathbf{x_s}-\mathbf{x}|\cdot|\mathbf{x_r}-\mathbf{x}|\cdot|\mathbf{x_s}-\mathbf{x_0}|\cdot|\mathbf{x_r}-\mathbf{x_0}|}, \tag{2.15}$$

where $rr=|\mathbf{x_s}-\mathbf{x}|+|\mathbf{x_r}-\mathbf{x}|$ and $rr_0=|\mathbf{x_s}-\mathbf{x_0}|+|\mathbf{x_r}-\mathbf{x_0}|$. Equation 2.14 shows that if acoustic migration (acoustic adjoint operator) is used to migrate viscoacoustic data, the reflection energy will be focused at the wrong subsurface location $\mathbf{x}$ instead of $\mathbf{x_0}$. If viscoacoustic migration (viscoacoustic adjoint operator, i.e. Q-RTM) is used, the reflector will be imaged at the right subsurface location $\mathbf{x_0}$. However, the backward propagated receiver wavefield in Q-RTM is further attenuated because of the $\exp\{-\frac{\omega}{2Qv_0\xi}(rr_0)\}$ term on the RHS of equation 2.15. Thus, the inverted images from Q-LSRTM will have lower resolution when compared to the images computed by acoustic LSRTM on acoustic data.

To increase the resolution of the Q-LSRTM images and accelerate the convergence of the least-squares iterations, we propose the use of local viscoacoustic DFs to approximate the inverse of the Hessian in a viscoacoustic medium. Assuming the standard linear solid (SLS) model (Christensen, 1982; Carcione et al., 1988; Blanch et al., 1995) and the Born approximation, the observed data $d_Q(\mathbf{x}_r,t;\mathbf{x}_s)$, excited by

a point source at $\mathbf{x_s}$ and recorded by a receiver at $\mathbf{x_r}$ in a viscoacoustic medium can be represented as

$$d_Q(\mathbf{x}_r,t;\mathbf{x}_s) = \int_{V_0} -\delta K(\mathbf{x}_0)\Big\{(\tau(\mathbf{x}_0)+1)\big[g_P(\mathbf{x}_r,t;\mathbf{x}_0,0) * \nabla\cdot \mathbf{v}(\mathbf{x}_0,t;\mathbf{x}_s)\big] + \frac{\tau(\mathbf{x}_0)}{\tau_\sigma(\mathbf{x}_0)}\big[g_{r_p}(\mathbf{x}_r,t;\mathbf{x}_0,0) * \nabla\cdot \mathbf{v}(\mathbf{x}_0,t;\mathbf{x}_s)\big]\Big\}\mathrm{d}V_0, \tag{2.16}$$

where K represents the bulk modulus, $\mathbf{v}$ is the particle velocity vector, $g_P(\mathbf{x}_r,t;\mathbf{x}_0,0)$ and $g_{r_p}(\mathbf{x}_r,t;\mathbf{x}_0,0)$ are the pressure and memory variable Green's functions, respectively. Here, τ_σ and τ_ϵ represent the stress and strain relaxation times, respectively, and $\tau = \frac{\tau_\epsilon}{\tau_\sigma} - 1$, which is also related to the quality factor Q. Using a matrix-vector notation, equation 2.16 can also be written as

$$\mathbf{d}_Q = \mathbf{L}_Q\mathbf{m}_0, \tag{2.17}$$

where $\mathbf{L}_Q$ is a linear viscoacoustic modeling operator, $\mathbf{d_Q}$ represent the lossy data, and $\mathbf{m}_0$ is the reflectivity of the medium. The perturbation in the image, δm, is associate with the perturbation in the bulk modulus, δK, which in turn can be obtained by zero-lag cross-correlation of the adjoint fields with the background wavefields

$$\delta m \approx \delta K = \int_0^t (1+\tau)(\nabla\cdot\mathbf{v})q + \frac{\tau}{\tau_\sigma}(\nabla\cdot\mathbf{v})s\,dt, \tag{2.18}$$

where, q and s are the adjoint-state variables of pressure P and memory variable r_p, respectively (Blanch and Symes, 1994; Dutta and Schuster, 2014b, 2016). Equation 2.18 can also be represented by the matrix-vector notation

$$\mathbf{m}_{mig} = \mathbf{L}_Q^T\mathbf{d}_Q, \tag{2.19}$$

where the detailed derivation of equations 2.16 to 2.19 are shown in the Appendix B.

Similar to the acoustic case, the viscoacoustic DF is estimated by defining a uniform distribution of point scatterers as the reference reflectivity model. The viscoacoustic synthetic data are generated using this reference reflectivity model and the background velocity and Q models. The data are then migrated by Q-RTM to obtain a reference Q-RTM image. Viscoacoustic DFs for different subdomains of the Q-RTM image are then estimated using matching filters as described in the previous section (detailed derivations are shown in Appendix C). The viscoacoustic DF for the ith local window are estimated in the space domain by solving the system of equations given by

$$[\mathbf{F}_Q]_i \otimes [\mathbf{L}_Q^T \mathbf{L}_Q \mathbf{m}_{ref}]_i \approx [\mathbf{m}_{ref}]_i, \tag{2.20}$$

where $[\mathbf{F}_Q]_i$ and $[\mathbf{m}_{ref}]_i$ represent the viscoacoustic DF and reference reflectivity model in the ith local window, respectively. As the reference migration image is generated from the same source-receiver configuration as the original field experiment and by using the same velocity and Q models, the application of these DFs to the migration image is an acceptable approximation to the true inverse Hessian operator $[\mathbf{L}_Q^T \mathbf{L}_Q]^{-1}$.

2.2.2 Q-LSRTM using viscoacoustic deblurring filters

The misfit function for Q-LSRTM is given by (Dutta and Schuster, 2014b)

$$\epsilon = \frac{1}{2} \|\mathbf{L}_Q \mathbf{m}^{(k)} - \mathbf{d}_Q^{obs}\|^2, \tag{2.21}$$

where $\mathbf{d}_Q^{obs}$ denotes the observed data that have suffered from attenuation, $\mathbf{m}^k$ represents the migration image at the k^{th} iteration and $\mathbf{L}_Q$ is the linearized viscoacoustic forward modeling operator. The Gauss-Newton gradient for this misfit function is

given by

$$(\mathbf{L}_Q^T\mathbf{L}_Q)\boldsymbol{\Delta}\mathbf{m}^{(k)} = \mathbf{L}_Q^T[\mathbf{L}_Q\mathbf{m}^{(k)} - \mathbf{d}_Q^{obs}] = \mathbf{g}^{(k)}. \tag{2.22}$$

It can be seen from this equation that the update at each iteration $\mathbf{g}^{(k)}$ is a blurred version of the desired update $\Delta\mathbf{m}^{(k)}$ caused by the viscoacoustic migration Green's function $\mathbf{L}_Q^T\mathbf{L}_Q$. Thus, a preconditioner for the gradient in equation 2.22 can be written as the deblurring approximation $[\mathbf{L}_Q^T\mathbf{L}_Q]^{-1} \approx \mathbf{F}_Q$, so that the preconditioned gradient in equation 2.22 can be used in the iterative update equation

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} - \alpha\mathbf{F}_Q \otimes (\mathbf{L}_Q^T(\mathbf{L}_Q\mathbf{m}^{(k)} - \mathbf{d}_Q^{obs})), \tag{2.23}$$

where α is the step length.

2.3 Numerical results

The effectiveness of Q-LSRTM with viscoacoustic DFs is now demonstrated with synthetic data generated from the Marmousi II model and the 3D SEG/EAGE Overthrust model. Also, we apply this method to the benchmark viscoacoustic data released by Schlumberger for the BP2004Q model and field data recorded in a crosswell experiment in Friendswood, Texas.

2.3.1 Marmousi II model

The preconditioned Q-LSRTM method is first tested on the Marmousi II model. Figures 3.6a and 3.6b shows the true velocity and Q models, respectively, used for generating the observed data. The migration velocity and Q models are shown in Figures 3.6c and 3.6d, respectively, which is smoothed by a 5×5 window. The Q model is chosen such that the attenuation layers overlie the deeper anticlines. A time-domain viscoacoustic finite-difference modeling algorithm is used with one standard

linear solid model for both data simulation and migration. A Ricker wavelet with a peak frequency of 15 Hz is used as the source wavelet. A fixed-spread acquisition geometry is employed where there are 150 sources evenly distributed on the surface at an interval of 50 m. The data are recorded by 800 receivers for each shot uniformly distributed every 10 m on the surface.

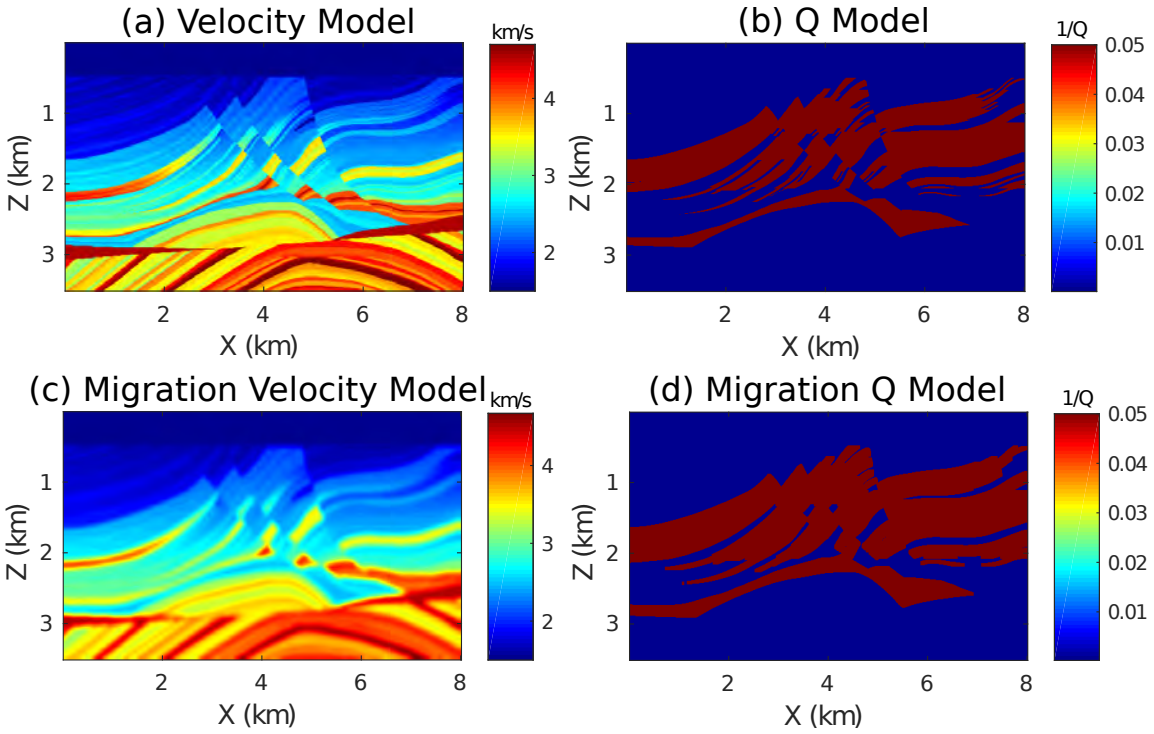


Figure 2.3: The Marmousi model: (a) true velocity model, (b) true Q model, (c) migration velocity model, and (d) migration Q model.

Conventional acoustic RTM and LSRTM images obtained from the viscoacoustic data are shown in Figures 2.4a and 2.4b, respectively. Both these images fail to recover the amplitudes of the reflectors at the deeper parts. The Q-LSRTM image, shown in Figure 2.4d, shows improvement in the deeper layers when compared to the acoustic migration results. However, the Q-LSRTM image has lower resolution for the reflectors below the Q anomaly when compared to the benchmark acoustic LSRTM image, shown in Figure 2.4f, that has been obtained from acoustic data generated using the same velocity model in Figure 3.6a. As discussed in the previous section,

this low resolution problem with Q-LSRTM is due to the attenuation properties of the adjoint operator $\mathbf{L}_Q^T$. However, the preconditioned Q-LSRTM image in Figure 2.4e computed with the viscoacoustic DFs has better resolution when compared to the Q-LSRTM image in Figure 2.4d. The amplitudes in the preconditioned Q-LSRTM image are also better balanced when compared to the acoustic RTM and LSRTM images in Figures 2.4a and 2.4b, respectively.

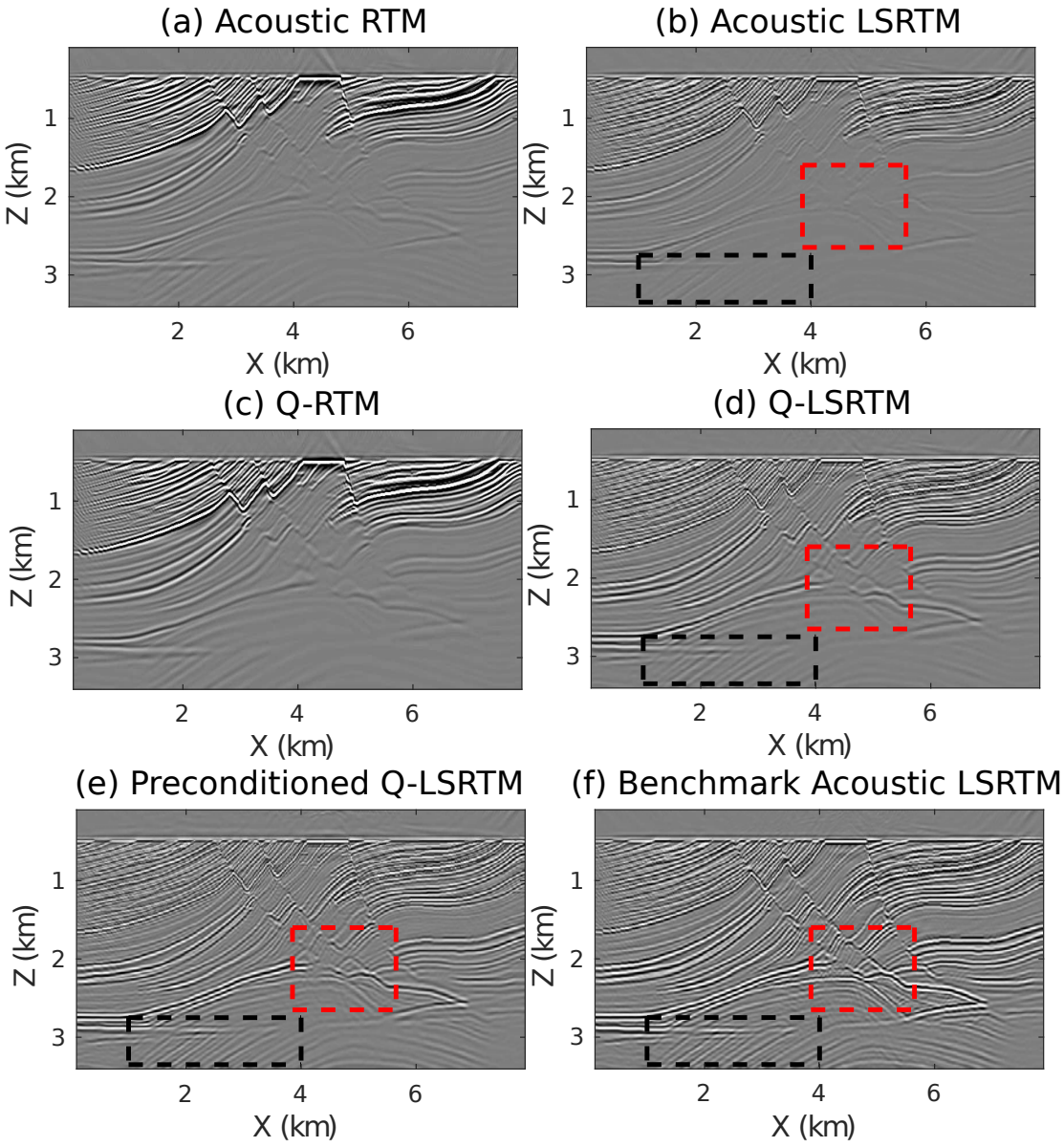


Figure 2.4: Comparison between images computed from viscoacoustic data by (a) acoustic RTM, (b) acoustic LSRTM, (c) Q-RTM, (d) Q-LSRTM and (e) Q-LSRTM using viscoacoustic DFs as a preconditioner, and (f) Acoustic LSRTM for lossless acoustic data, which is used as the benchmark image. 20 least-squares iterations are carried out in all the cases.

The magnified views of these images are compared in Figures 2.5 and 2.6, where the black arrows point to the areas in which noticeable improvements in resolution can be seen. Figure 2.5e shows the wavenumber spectrum of a vertical slice at $x = 3.11$ km. The wavenumber spectrum clearly shows the improvement in resolution with

preconditioned Q-LSRTM, which is also confirmed in Figure 2.6e which shows the wavenumber spectrum of a vertical slice at $x = 5.13$ km. It is evident from these plots that the high-wavenumber details in the image are successfully recovered in the preconditioned Q-LSRTM image and these images have a similar resolution as the benchmark image obtained from applying acoustic LSRTM to lossless acoustic data.

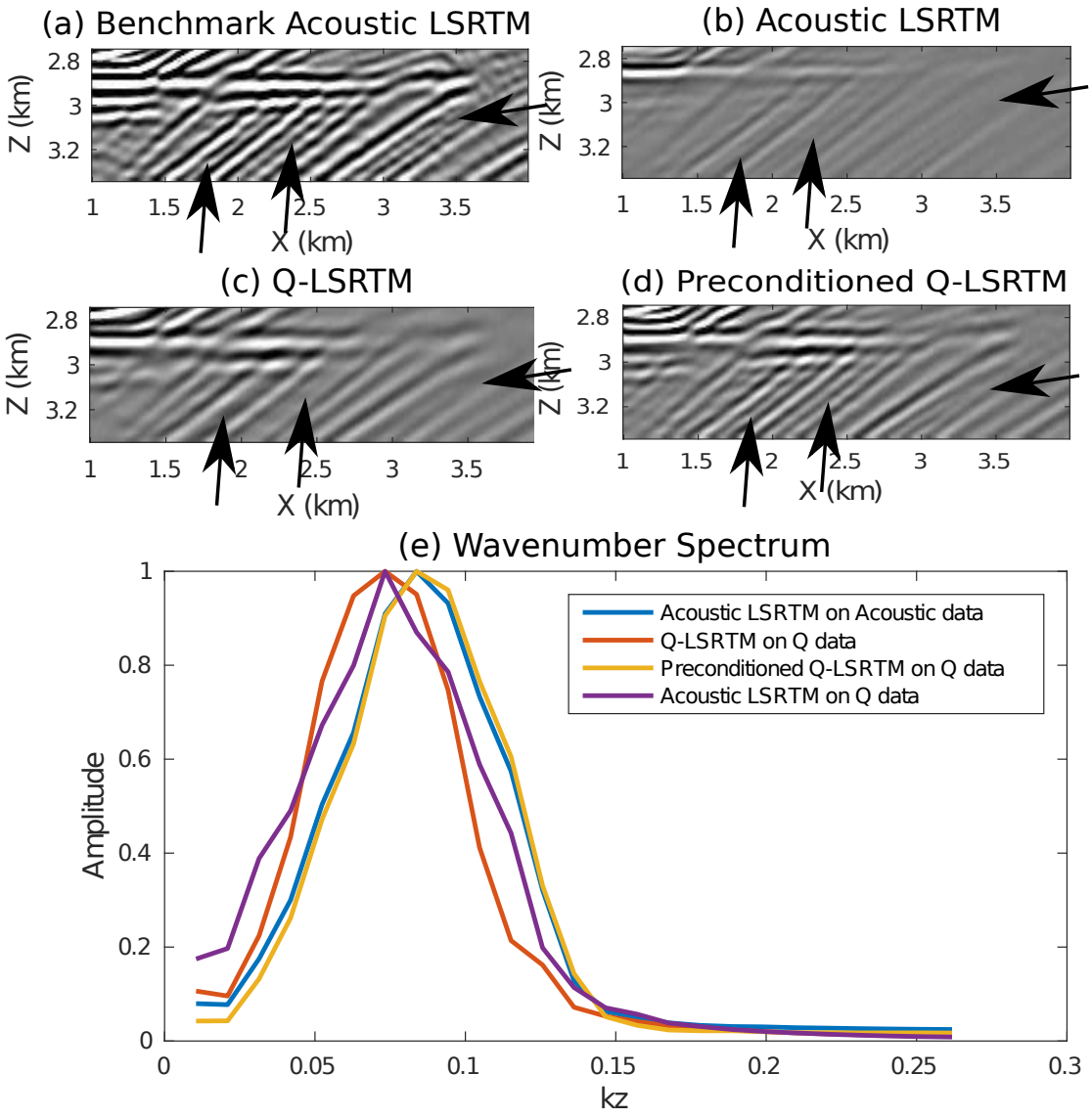


Figure 2.5: Magnified views of the black boxes in Figures 2.4. The black arrows point to the areas in which improvements can be seen, and Figure 2.5e shows the k_z wavenumber spectrum of a vertical slice at $x = 3.11$ km in the above four pictures.

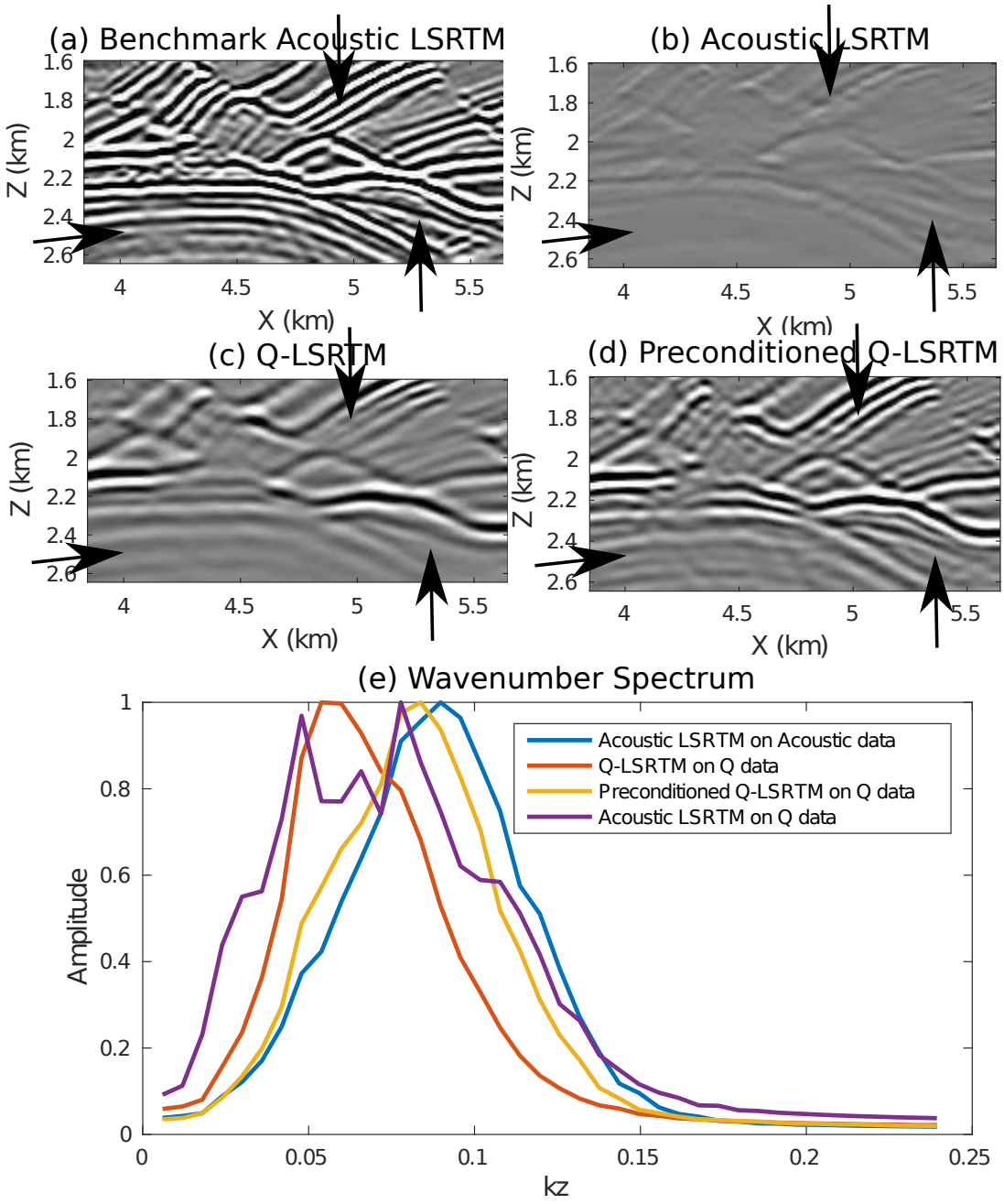


Figure 2.6: Magnified views of the red boxes in Figures 2.4. The black arrows point to the areas in which improvements can be seen. Figure 2.6e shows the k_z wavenumber spectrum of a vertical slice at $x = 5.13$ km in the above four pictures.

The data residual as a function of iteration number for LSRTM, Q-LSRTM and preconditioned Q-LSRTM is plotted in Figure 2.7. This plot shows that the convergence rate for preconditioned Q-LSRTM is much faster than Q-LSRTM especially at

the early iterations. The residuals in the 3rd and 9th iterations of preconditioned Q-LSRTM are almost equal to the residuals at the 6th and 20th iterations of Q-LSRTM, respectively. Thus, a speedup of around 50% can be seen with the proposed preconditioning method. The convergence rate for preconditioned Q-LSRTM is better than that of standard Q-LSRTM because the viscoacoustic DFs are an effective approximation of the inverse Hessian operator.

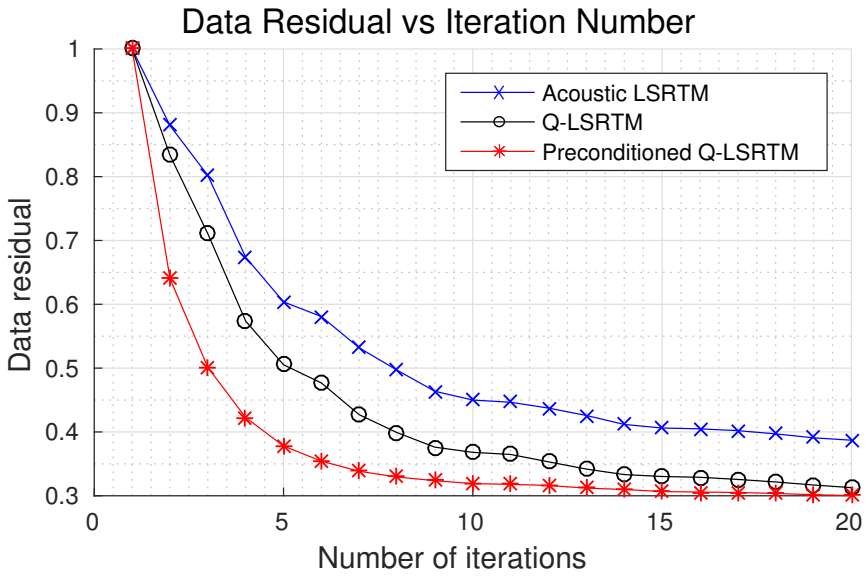


Figure 2.7: Data residual vs number of iterations for acoustic LSRTM, Q-LSRTM and preconditioned Q-LSRTM applied to data associated with the Marmousi II model.

2.3.2 Sensitivity of deblurred images with respect to errors in the Q model

To check the sensitivity of the preconditioned Q-LSRTM method to errors in the migration Q model, numerical tests are carried out for different background Q models. Errors are introduced in the background Q model by varying the Q value at the anomalies. In Figure 2.8, as the errors in the background Q model increase, the amplitudes of the reflectors below the Q anomalies become weaker with increasing errors in the Q model. On the contrary, the PSFs, which are the migration response of the point scatterers, in the same areas become stronger as the Q value increases

and thus, they are not able to take into account the effect of Q during deblurring. Therefore, a fairly accurate estimation of the background Q model is required to see noticeable improvements in the image quality with the proposed preconditioned Q-LSRTM method.

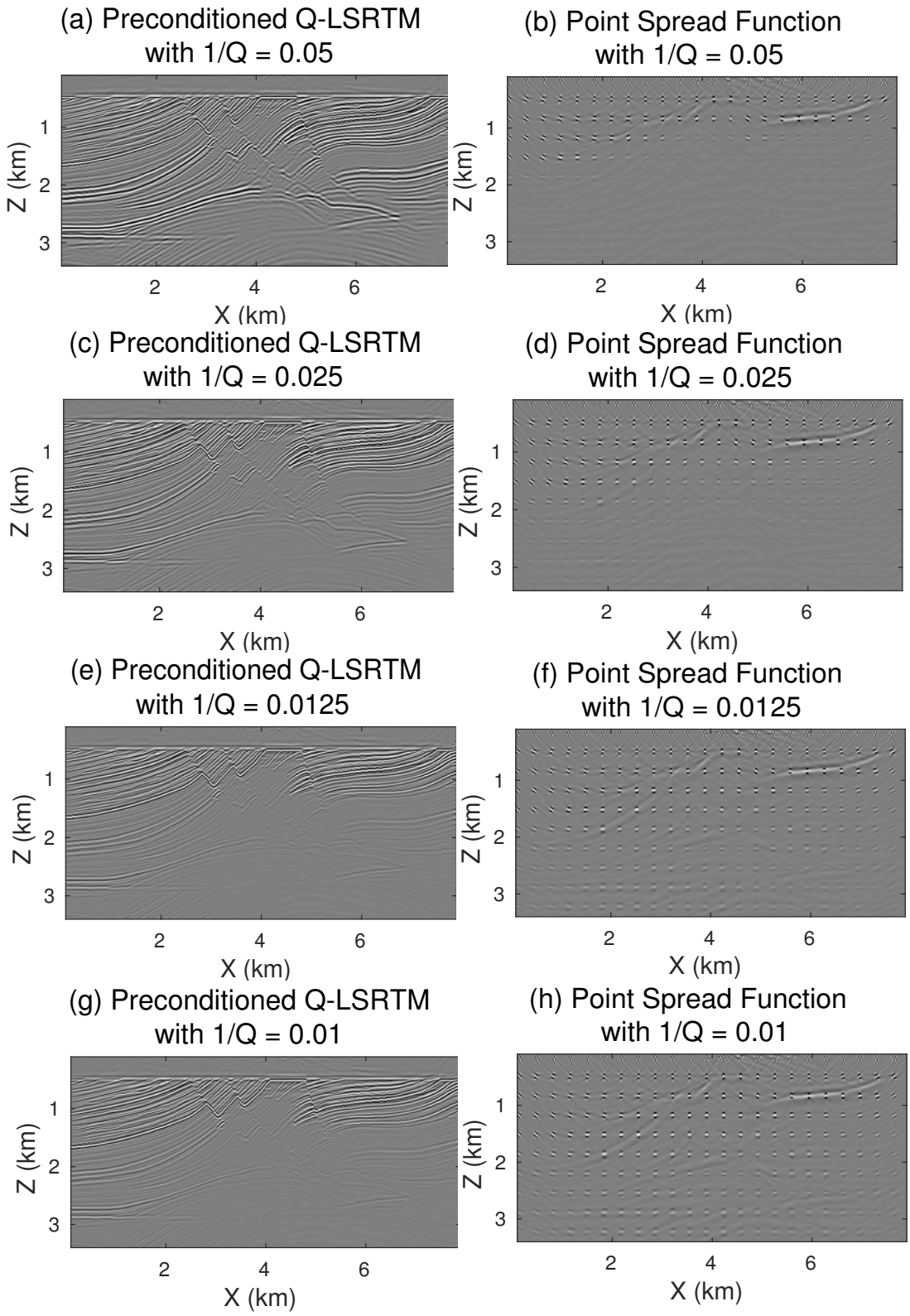


Figure 2.8: Sensitivity of preconditioned Q-LSRTM to errors in the migration Q model. The left panel shows the preconditioned Q-LSRTM images while the right panel shows the corresponding point spread functions.

2.3.3 BP viscoacoustic benchmark data

The preconditioned Q-LSRTM method is now tested on the B2004Q data generated by Schlumberger (Billette and Brandsberg-Dahl, 2005; Cavalca et al., 2013), where the velocity and Q models used for migration are shown in Figures 2.9a and 2.9b, respectively. The observed data are generated using a Ricker wavelet with a peak frequency of 19 Hz, where the original data consist of 1348 shot gathers and each shot is recorded by 2401 receivers. The shot spacing is 50 m while the receivers are uniformly distributed on both sides of a shot at a spacing of 12.5 m. The sources and receivers are placed at a depth of 12.5 m.

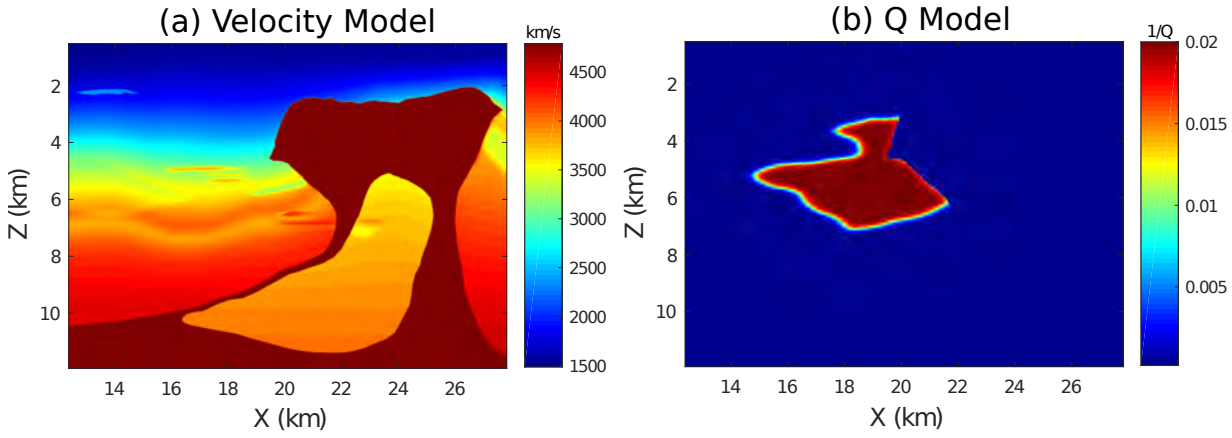


Figure 2.9: BP2004Q model: (a) true velocity and (b) true Q models.

For our numerical tests, we only use 236 shots. The true reflectivity model in Figure 2.10f is used as the groundtruth for the acoustic and Q-LSRTM images. In the acoustic LSRTM image in Figure 2.10b, it is difficult to delineate the reflectors near the salt flank. The Q-LSRTM image in Figure 2.10d has better balanced amplitudes than the acoustic LSRTM image. However, the preconditioned Q-LSRTM image in Figure 2.10e has reflectors with better balanced amplitudes and better resolution than the standard Q-LSRTM and acoustic LSRTM images. Magnified views of these images in Figure 2.11a-2.11d, also illustrate the same. Figure 2.11e shows the wavenumber spectrum of a vertical slice at $x = 17.05$ km, and the red curve

represents the spectrum of the preconditioned Q-LSRTM image that clearly shows an improvement in resolution when compared with Q-LSRTM indicated by the blue line.

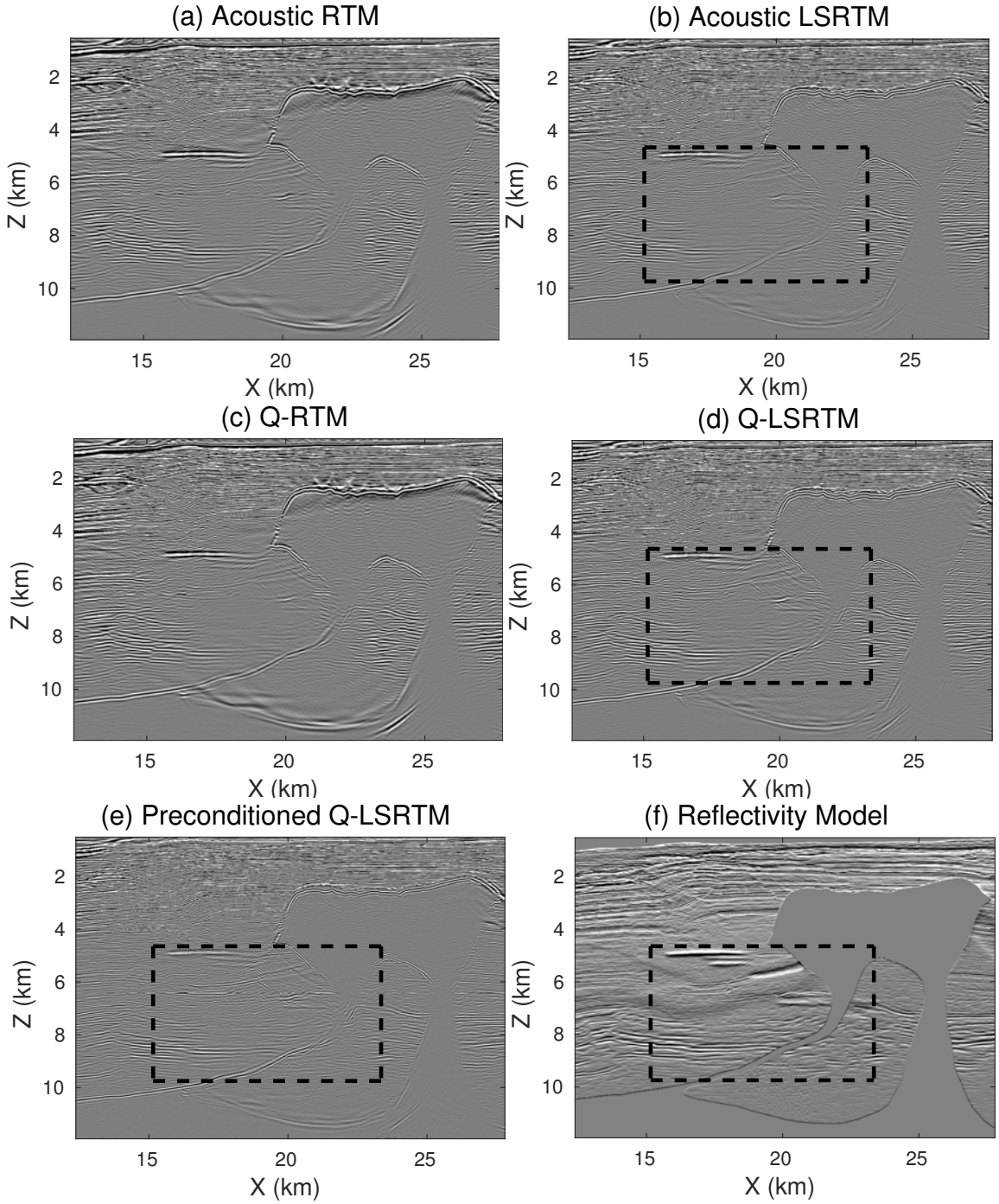


Figure 2.10: Comparison between images from (a) acoustic RTM, (b) acoustic LSRTM, (c) Q-RTM, (d) Q-LSRTM, (e) preconditioned Q-LSRTM and (f) the true reflectivity model. The black boxes indicate the areas for magnified views.

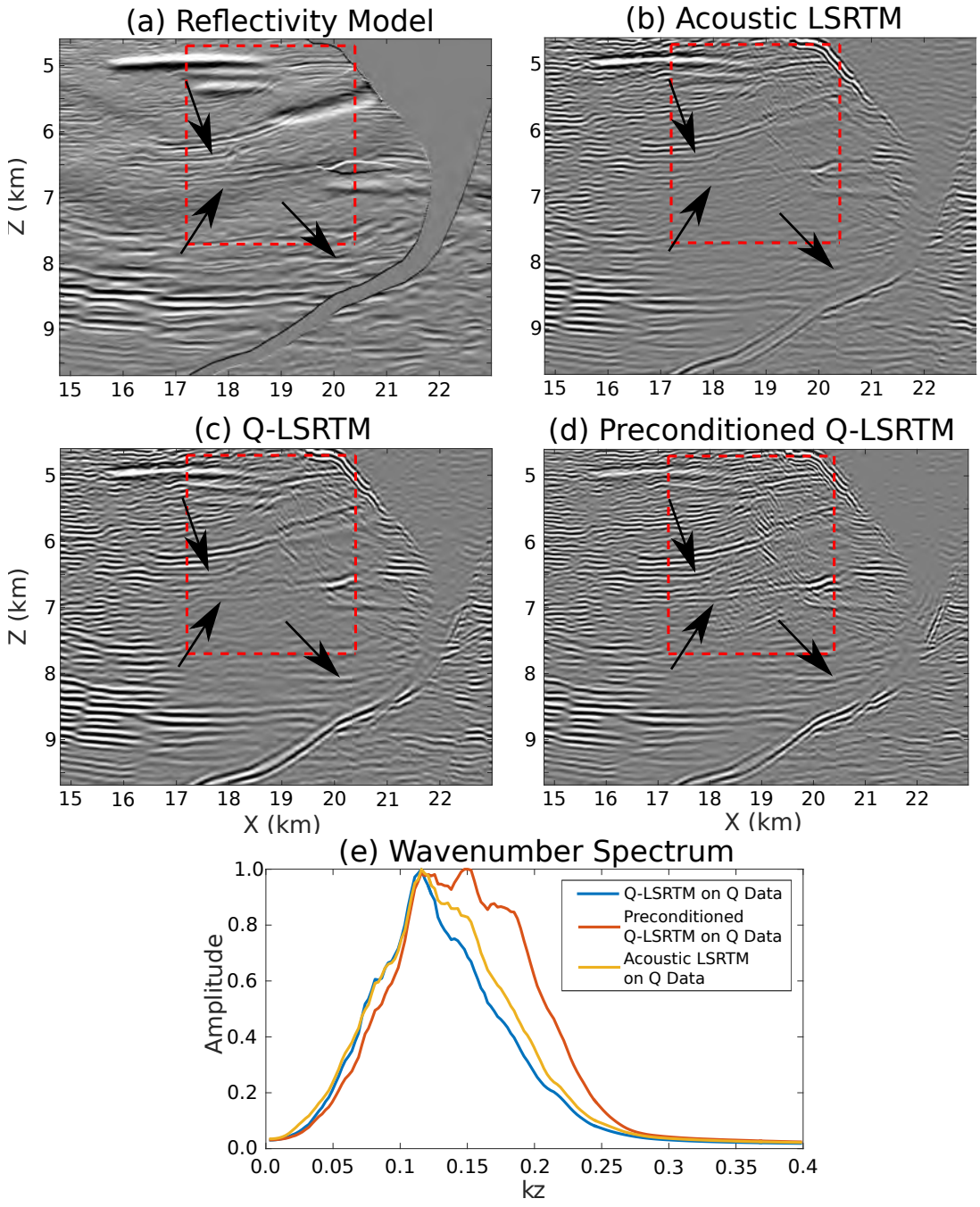


Figure 2.11: Magnified views of the black boxes in Figure 2.10. The black arrows point to the reflectors below the high attenuation area where improvements from the preconditioned Q-LSRTM method can be seen. Figures 2.11e shows the wavenumber spectra of a vertical slice at $x = 17.05$ km in the above four pictures.

The zoomed views especially in the red box also reveal some artifacts parallel to the salt boundary that become amplified in the preconditioned Q-LSRTM image. These

artifacts can also be seen in the acoustic RTM and LSRTM images. The residual as a function of iteration number for LSRTM, Q-LSRTM and preconditioned Q-LSRTM is plotted in Figure 2.12. The convergence rate for preconditioned Q-LSRTM is much faster than Q-LSRTM and acoustic LSRTM, especially for the first few iterations.

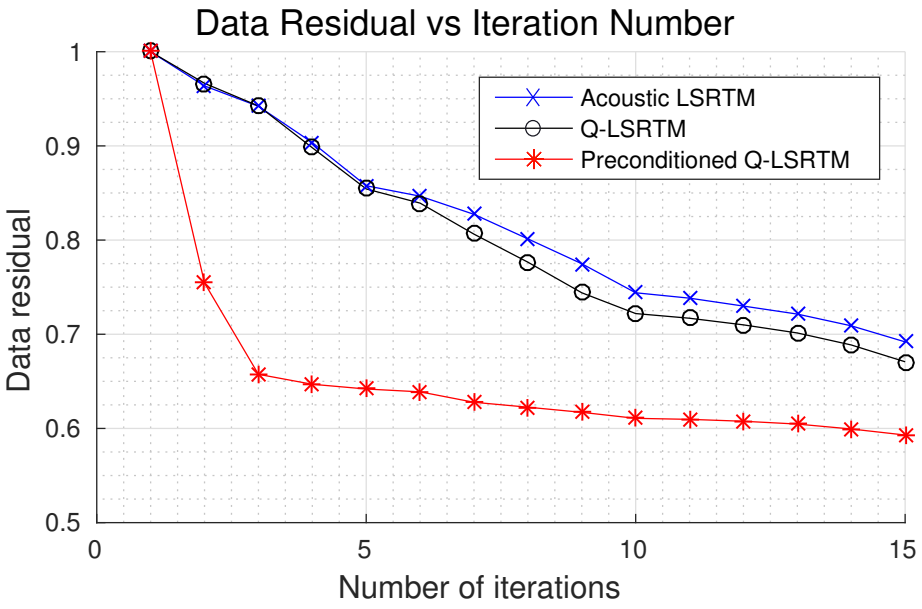


Figure 2.12: Data residuals vs number of iterations for LSRTM, Q-LSRTM and preconditioned Q-LSRTM applied to the BP2004Q benchmark data.

2.3.4 3D SEG/EAGE Overthrust Model

To further illustrate the effectiveness of viscoacoustic DFs, we apply the method to data generated from the 3D SEG/EAGE Overthrust model. Similar to the 2D examples, we use 3D viscoacoustic DFs as a preconditioner for the 3D Q-LSRTM. The number of grid points in the original model is $187 \times 801 \times 801$. For our numerical tests, we only use part of the model with a size of $110 \times 240 \times 240$ grid points. Figures 2.13a and 2.13b shows the true velocity and Q models, respectively, used for generating the observed data. A Ricker wavelet with a peak frequency of 15 Hz is used as the source wavelet. A fixed spread acquisition geometry is used where there are 4225 sources evenly distributed on the surface at an interval of 40 m in the

inline and crossline directions. The data are recorded by 130 receiver lines uniformly distributed at an interval of 20 m on the surface. Each receiver line has 130 receivers with a spacing of 20 m between each receiver.

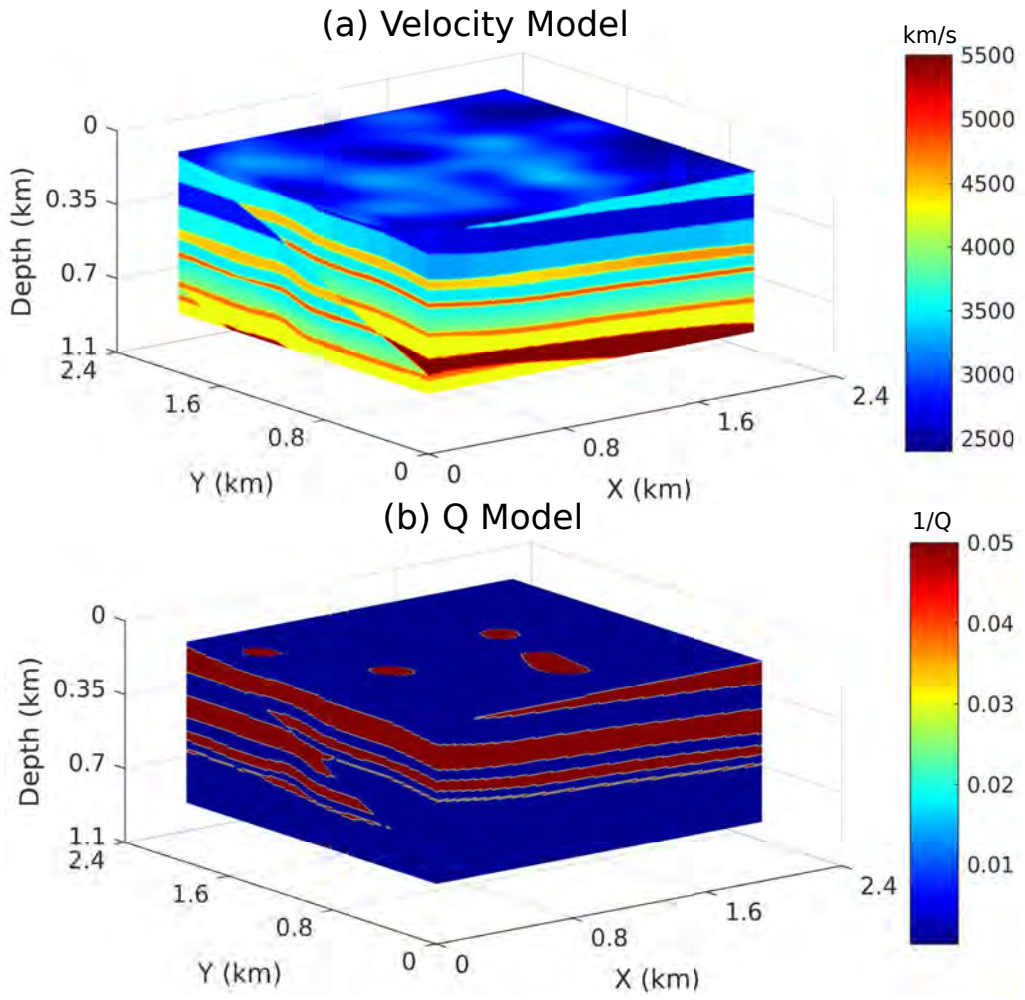


Figure 2.13: SEG/EAGE Overthrust model: (a) true velocity and (b) true Q models.

Figure 2.14 compares the images from 3D Q-LSRTM and 3D preconditioned Q-LSRTM. The 3D Q-LSRTM and preconditioned Q-LSRTM images are obtained after 6 and 2 iterations, respectively. The preconditioned Q-LSRTM image in Figure 2.14b has reflectors with more balanced amplitudes and better resolution than the standard Q-LSRTM in Figure 2.14a. The residual as a function of iteration number for the 3D Q-LSRTM and preconditioned Q-LSRTM images is plotted in Figure 2.15. The

convergence rate for preconditioned Q-LSRTM is much faster than Q-LSRTM. The residuals in the second iterations of preconditioned Q-LSRTM are even smaller than the residuals for the 6th iteration of Q-LSRTM. This fast convergence further proves that the estimated visco DF is a good approximation of the true Hessian inverse.

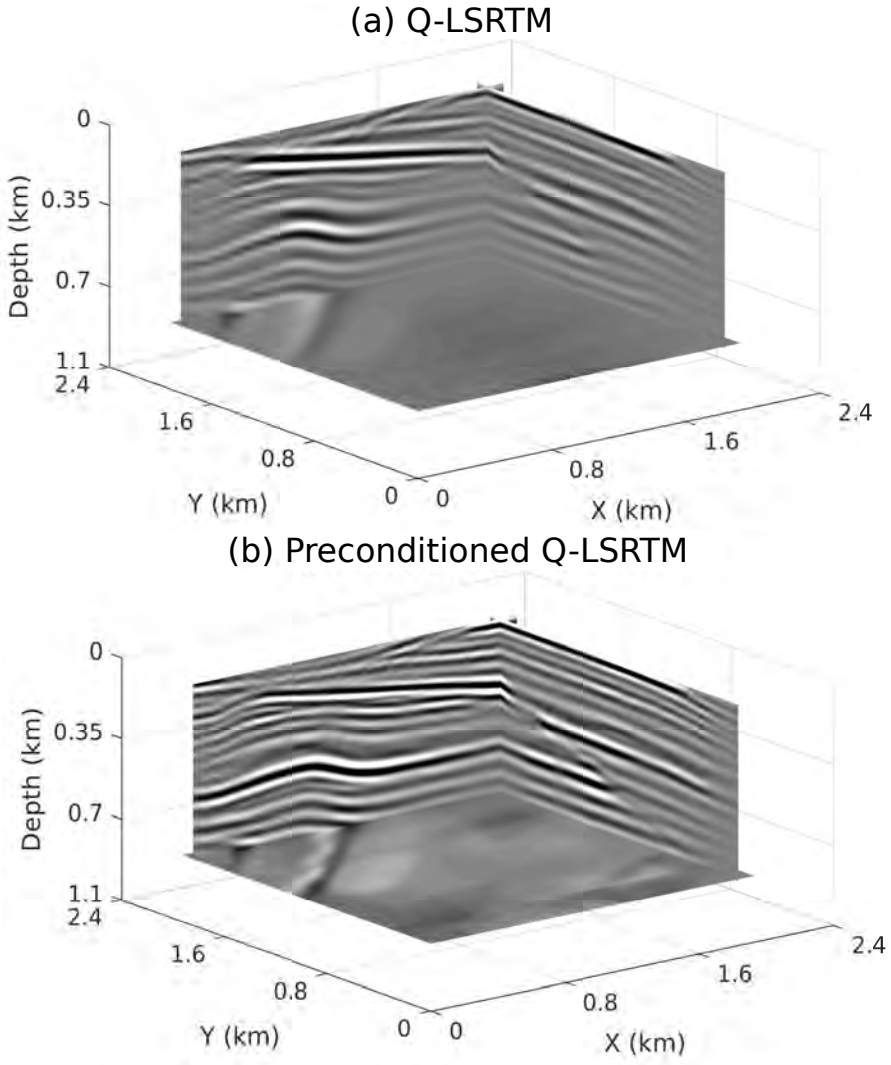


Figure 2.14: Comparison between images computed from viscoacoustic data by (a) 3D Q-LSRTM and (b) 3D preconditioned Q-LSRTM.

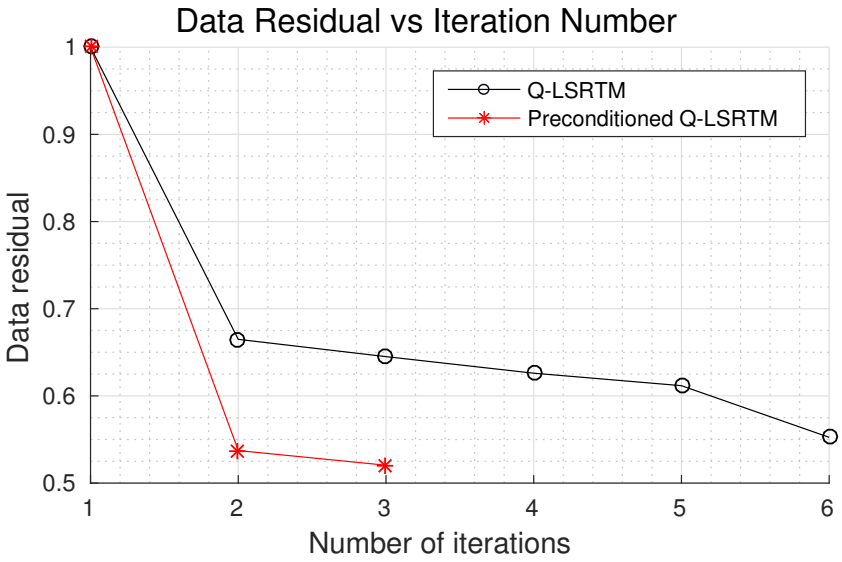


Figure 2.15: Data residual vs number of iterations for 3D Q-LSRTM and 3D preconditioned Q-LSRTM applied to data associated with the 3D SEG/EAGE Overthrust model.

2.3.5 Friendswood crosswell field data

As a final example, we test the preconditioned Q-LSRTM method on the Friendswood crosswell field data set. Two 305-m-deep cased wells separated by 183 m were used as the source and receiver wells. Downhole explosive charges were fired at intervals of 3 m from 305 m to 9 m in the source well, and the receiver well had 96 receivers placed at depths ranging from 293 m to 3 m. The data were recorded with a sampling interval of 0.25 ms for a total recording time of 0.375 s (Chen et al., 1990). During processing. the data are Wiener-filtered to transform the original wavelet to a Ricker wavelet with a 200-Hz peak frequency. For these data, the migration velocity and Q models are shown in Figures 3.10a and 3.10b, respectively. The migration velocity model is estimated by early-arrival waveform inversion and the migration Q model is estimated by wave-equation Q tomography (Dutta, 2016; Dutta and Schuster, 2016).

The comparison between the acoustic LSRTM and Q-LSRTM images after 20 iterations are shown in Figures 2.17a and 2.17b, respectively. Similar to the synthetic examples, the amplitudes are more balanced in the Q-LSRTM image in Figure 2.17b

than in the acoustic LSRTM images in Figures 2.17a, while the resolution becomes lower. The preconditioned Q-LSRTM image is shown in Figure 2.17c. When compared to the standard Q-LSRTM image, the preconditioned LSRTM image has better resolution at depths of 10-80 m. Magnified views of these areas are shown in Figure 2.18 that further validate the improvement in resolution with preconditioning. The black arrows in these figures depict the areas that have become more pronounced in the preconditioned Q-LSRTM image.

As a sanity check, reflectivity slices from the preconditioned Q-LSRTM image are compared to the well log data taken at a distance of 12 m from the source well. The comparison between the synthetic seismogram computed from the well-log profile and the standard and the preconditioned Q-LSRTM reflectivity profiles is shown in Figures 2.19a and 2.19b, respectively. It is evident that the well log agrees better with the preconditioned Q-LSRTM image than with the Q-LSRTM image. The black arrows in this figure highlight some of these agreements.

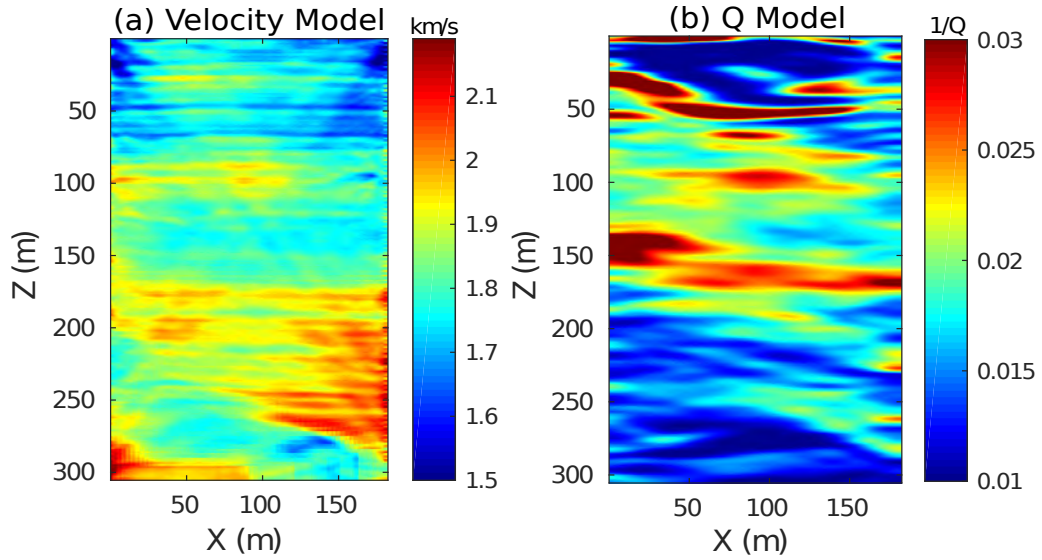


Figure 2.16: (a) The estimated migration velocity and (b) Q models for the Friendswood crosswell data.

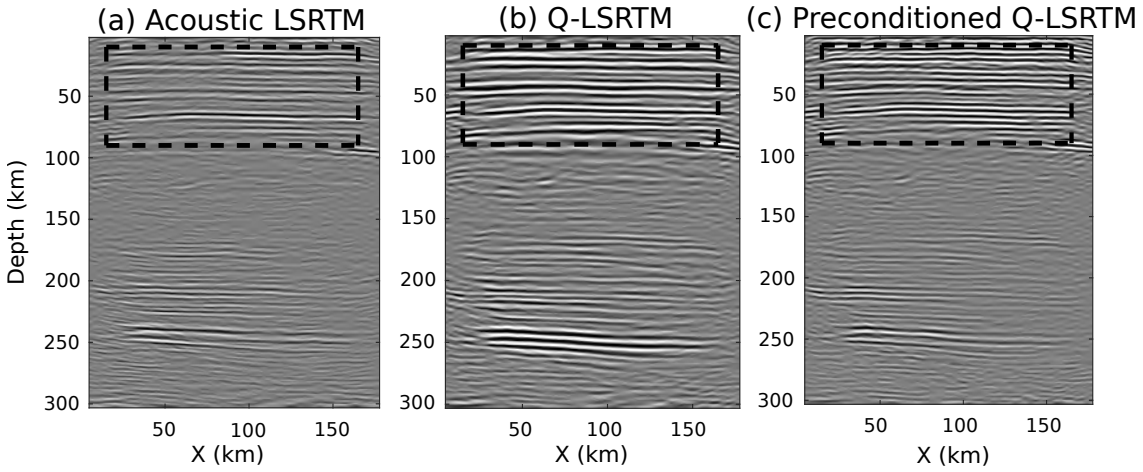


Figure 2.17: Comparison between images from (a) acoustic RTM, (b) acoustic LSRTM, (c) Q-RTM, (d) Q-LSRTM and (e) preconditioned Q-LSRTM. Twenty iterations are carried out in b), d) and e).

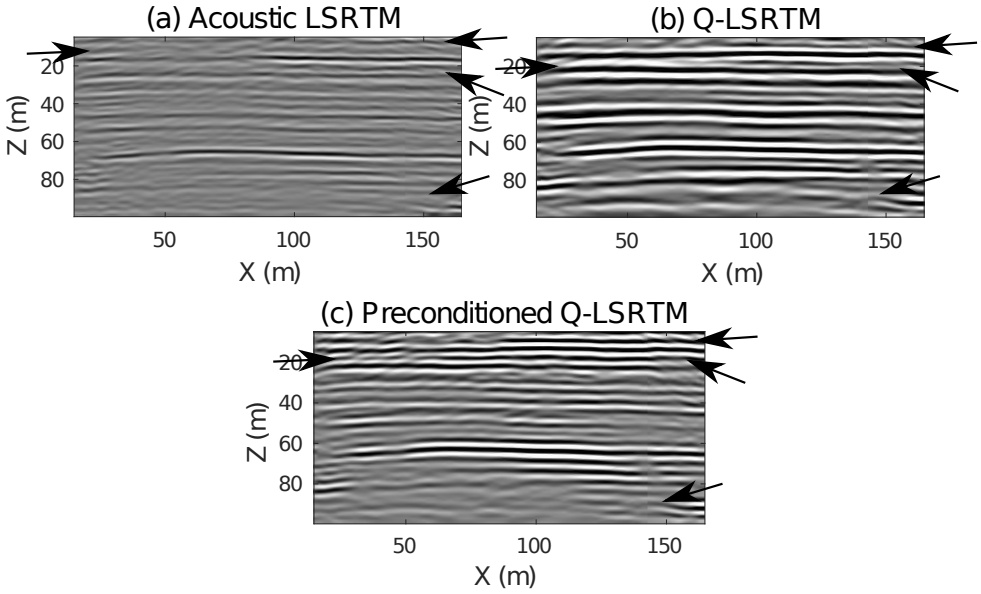


Figure 2.18: Magnified views of the black boxes in Figure 2.17. The black arrows point to the reflectors where the improvement in resolution can be seen from the preconditioned Q-LSRTM method.

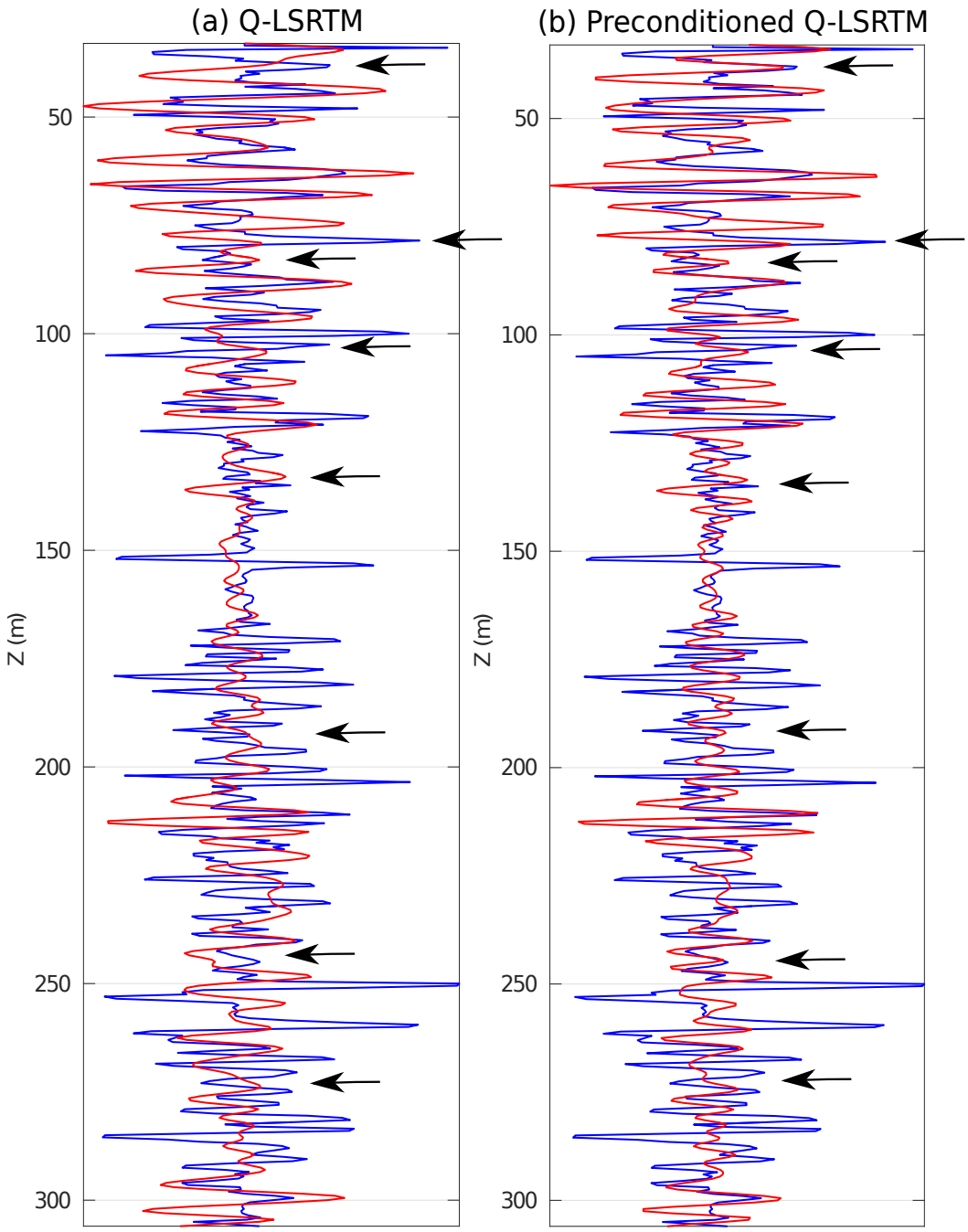


Figure 2.19: Comparison between the true reflectivity obtained from a well log (represented by the blue line) and the inverted reflectivity (represented by the red line) from (a) Q-LSRTM and (b) preconditioned Q-LSRTM. The well log is at a distance of 12 m from the source well

2.4 Conclusion

A preconditioned Q-LSRTM method is presented that uses viscoacoustic DFs to compensate for the amplitude and resolution losses due to strong subsurface attenuation. Numerical tests on synthetic and field data validate that the proposed preconditioning method mitigates the problem of low resolution associated with standard Q-LSRTM and can produce images with better balanced amplitudes and better resolution than acoustic RTM and LSRTM. The viscoacoustic DFs are estimated from a reference model with evenly distributed point-scatterers and its Q-RTM image is estimated by local matched filters. The proposed preconditioning method is also shown to improve the convergence rate of iterative LSM by more than 50 percent in some cases. Similar to standard Q-LSRTM, a fairly accurate estimation of the background Q model is required to see noticeable improvements in the image quality with the preconditioned Q-LSRTM method. The limitation of this procedure is that the localized DF will not reduce artifacts associated with strong migration artifacts far from the scattering point. In this case, the DFs must be computed by using much wider windows, which can significantly increase the computational costs.

Chapter 3

Migration of Viscoacoustic Data Using Acoustic Reverse Time Migration with Hybrid Deblurring Filters

Viscoacoustic least-squares reverse time migration (Q-LSRTM) can compensate for the amplitude loss and phase distortion in migration images computed from highly attenuated data. However, the adjoint Q propagator of Q-LSRTM is attenuative which leads the low-resolution problem of Q-LSRTM. To mitigate this problem, we use acoustic reverse time migration (RTM) instead of Q-LSRTM to migrate the viscoacoustic data, and then correct the amplitude and phase distortion by hybrid deblurring filters (DFs) in the image domain. Numerical tests on synthetic and field data demonstrate that acoustic RTM combined with hybrid DFs can compensate for the attenuation effects and produce images with high resolution and balanced amplitudes. This procedure requires less than 1/3 of the storage space and is $O(N-1)$ times faster compared to the viscoacoustic migration. Here, N represents the number of iterations used for the least-square migration method. This method can be extended to 3D migration at even a greater cost saving.

3.1 Introduction

Subsurface attenuation distorts both the amplitudes and phases of propagating waves. To correct for these distortions, several attenuation compensation methods were proposed to produce images with balanced amplitudes and accurately positioned reflectors (Dai and West, 1994; Blanch et al., 1995; Wang, 2007; Zhang et al., 2010;

Valenciano et al., 2011; Suh et al., 2012; Fletcher et al., 2012; Dutta and Schuster, 2014b; Dai et al., 2015b; Sun et al., 2016; Chen et al., 2017).

In chapter 2, we introduced the viscoacoustic deblurring filters (DFs) as a preconditioner for Q-LSRTM to improve the image resolution and accelerate the convergence rates of Q-LSRTM. In this chapter, we use acoustic RTM instead of Q-LSRTM to migrate the lossy data in order to avoid the usage of the attenuative adjoint Q propagator, but which results in a distorted image as acoustic RTM is unable to correct for the attenuation effects. The hybrid DFs are then used to correct these amplitude and phase distortions in the acoustic RTM image. These filters (Aoki and Schuster, 2009) can be calculated by the following procedure: (1) Construct a reference reflectivity model as a uniform distribution of point scatterers; (2) Generate the reference viscoacoustic data with the reference reflectivity model while the background velocity and Q models are the same as the ones used for migration; (3) Migrate the reference data by acoustic RTM to obtain a reference migration image; (4) Solve the hybrid DFs by matching the reference migration image with the reflectivity model. If N iterations are required by Q-LSRTM, the hybrid deblurring approach is $O(N-1)$ times faster because no iterative model updates are required and this algorithm only computes solutions to the acoustic wave equation after the hybrid filter is computed.

This chapter is divided into four sections. After the introduction, we present the theory for applying the hybrid DFs to images computed by acoustic migration. Numerical tests on synthetic and field data are then used to demonstrate the advantages of the proposed method. The conclusions are presented in the last section.

3.2 Theory

Acoustic RTM is a computational friendly method because it is easy to implement and numerically stable. Migrating viscoacoustic data using the acoustic migration

can be formulated as

$$\tilde{\mathbf{m}}_{mig} = \mathbf{L}^T \mathbf{d}_Q = \mathbf{L}^T \mathbf{L}_Q \mathbf{m}_0, \tag{3.1}$$

where $\mathbf{L}^T$ is the acoustic migration operator and $\tilde{\mathbf{m}}_{mig}$ is the migration image. The tilde symbol indicates that the migration image $\tilde{\mathbf{m}}_{mig}$ is incorrect as the acoustic migration operator is not able to compensate for the phase distortion and amplitude loss in the migration image. Equation 3.1 can be re-written as

$$\tilde{m}_{mig}(\mathbf{x}) = \sum_{\mathbf{x}_0} \Gamma(\mathbf{x}|\mathbf{x_0}) m(\mathbf{x}_0), \tag{3.2}$$

where, $\Gamma_{mig}(\mathbf{x}|\mathbf{x}_0)$ can be explicitly derived in a homogeneous medium with constant velocity $\mathbf{v}_0$ as

$$\begin{aligned} \Gamma_{mig}(\mathbf{x}|\mathbf{x}_0, \omega) &= \int_\omega d\omega \sum_s \sum_r G^*(\mathbf{x}_g|\mathbf{x}) G^*(\mathbf{x}|\mathbf{x}_s) G_Q(\mathbf{x}_g|\mathbf{x}_0) G_Q(\mathbf{x}_0|\mathbf{x}_s) \\ &= \int_\omega d\omega \sum_s \sum_r \frac{\exp\{i\frac{\omega}{v_0}(\frac{rr}{\xi} - rr_0)\} \exp\{ -\frac{\omega}{2Qv_0\xi}(rr)\}}{|\mathbf{x}_s - \mathbf{x}| \cdot |\mathbf{x}_r - \mathbf{x}| \cdot |\mathbf{x}_s - \mathbf{x}_0| \cdot |\mathbf{x}_r - \mathbf{x}_0|}, \end{aligned} \tag{3.3}$$

where $rr = |\mathbf{x}_s - \mathbf{x}| + |\mathbf{x}_r - \mathbf{x}|$, $rr_0 = |\mathbf{x}_s - \mathbf{x}_0| + |\mathbf{x}_r - \mathbf{x}_0|$ and the symbol $*$ indicates complex conjugation. Equation 10 suggests that if acoustic migration is used to migrate viscoacoustic data, the reflection energy will be focused at the wrong location where $\frac{rr}{\xi} - rr_0 = 0$. Furthermore, the image is blurred by $\mathbf{L}^T\mathbf{L}_Q$. To compensate for the attenuation effects and insure that the imaged reflectors are at the right location in the $\tilde{\mathbf{m}}_{mig}$ image, the inverse of the Hessian operator $(\mathbf{L}^T\mathbf{L}_Q)^{-1}$ needs be applied to the migration image, so that

$$\begin{aligned} \mathbf{m}_{deblur} &= (\mathbf{L}^T\mathbf{L}_Q)^{-1}\tilde{\mathbf{m}}_{mig} \\ &= (\mathbf{L}^T\mathbf{L}_Q)^{-1}(\mathbf{L}^T\mathbf{L}_Q)\mathbf{m}_0 \\ &= \mathbf{m}_0. \end{aligned} \tag{3.4}$$

However, it is computationally prohibitive to compute the direct inverse of $\mathbf{L}^T\mathbf{L}_Q$ for practical seismic imaging problems. In addition, $\mathbf{L}^T\mathbf{L}_Q$ is often ill-conditioned so that its inverse is prone to numerical errors.

We now propose a hybrid DF to efficiently approximate the Hessian inverse operator $(\mathbf{L}^T\mathbf{L}_Q)^{-1}$. To estimate the hybrid DF, we first define a uniform distribution of point scatterers as the reference reflectivity model. The viscoacoustic synthetic data $\mathbf{d}_Q$ are generated using this reference reflectivity model with the background velocity and Q models:

$$\mathbf{d}_{ref\_Q} = \mathbf{L}_Q\mathbf{m}_{ref}. \tag{3.5}$$

The data $\mathbf{d}_{ref\_Q}$ are then migrated by acoustic RTM to obtain a reference migration image

$$\tilde{\mathbf{m}}_{mig\_ref} = \mathbf{L}^T\mathbf{d}_{ref\_Q}. \tag{3.6}$$

The hybrid DFs $\mathbf{F}_i$ for different subdomains of the RTM image are then estimated by locally matching the reference migration image $\tilde{\mathbf{m}}_{mig_{ref}}$ with the true reference reflectivity model $\mathbf{m}_{ref}$ as

$$[\mathbf{m}_{ref}]_i = [\mathbf{F}]_i \otimes [\tilde{\mathbf{m}}_{mig\_ref}]_i, \tag{3.7}$$

where $\otimes$ denotes spatial convolution and i indicates the ith local window. Here, $[\mathbf{F}]_i$, $[\mathbf{m}_{ref}]_i$, and $[\tilde{\mathbf{m}}_{mig\_ref}]_i$ denote the DF, the reference reflectivity model, and the reference migration image within the ith local window, respectively. To numerically estimate the DF, equation 3.7 needs to be transformed into matrix multiplication form

$$[\mathbf{M}_{ref\_mig}]_i[\mathbf{f}]_i = [\mathbf{m}_{ref}]_i. \tag{3.8}$$

This requires flattening the two-dimensional DF $[\mathbf{F}]_i$ into the vector format $[\mathbf{f}]_i$ and transforming the migration image $[\mathbf{m}_{mig\_ref}]_i$ into a convolution matrix $[\mathbf{M}_{ref\_mig}]_i$ (this transformation has explained in chapter 2). The hybrid DF can be estimated by solving equation 3.8 by the least-squares method: (Aoki and Schuster, 2009; Dai and Schuster, 2009)

$$[\mathbf{f}]_i = ([\mathbf{M}_{ref\_mig}]_i^T [\mathbf{M}_{ref\_mig}]_i)^{-1} [\mathbf{M}_{ref\_mig}]_i^T [\mathbf{m}_{ref}]_i, \tag{3.9}$$

and the estimated filters can then be applied to the real migration image $\tilde{\mathbf{m}}_{mig}$ to correct for the attenuation and blurring effects:

$$[\mathbf{m}_{deblur}]_i \approx [\mathbf{F}]_i \otimes \tilde{\mathbf{m}}_{mig}. \tag{3.10}$$

3.3 Workflow

The workflow for migration of viscoacoustic data using acoustic RTM with a hybrid DFs consists of the following steps:

1. Compute $\tilde{\mathbf{m}}_{mig}$ by migrating the viscoacoustic observed data using acoustic RTM.

2. Build a reference reflectivity model $\mathbf{m}_{ref}$ with point scatterers evenly distributed in the model. The reference viscoacoustic data $\mathbf{d}_{ref\_Q}$ are generated by the viscoacoustic Born modeling based on this reference reflectivity model and the background velocity and Q models.

3. Compute the reference migration image $\tilde{\mathbf{m}}_{mig\_ref}$ by migrating the reference data using acoustic RTM.

4. The hybrid DFs $\mathbf{F}_i$ for different subsections of the RTM image are estimated by locally matching the reference migration image $[\tilde{\mathbf{m}}_{mig\_ref}]_i$ with the reference

reflectivity model $[\mathbf{m}_{ref}]_i$.

5. Apply the calculated filters $\mathbf{F}_i$ to the migration image $\tilde{\mathbf{m}}_{mig}$ to correct for the attenuation distortions.

3.4 Numerical results

The effectiveness of migrating viscoacoustic data by the combination of acoustic RTM and the hybrid DFs is now demonstrated with synthetic and field data examples.

3.4.1 Point-Scatterer Model

A point-scatterer model is used to test the hybrid filter's effectiveness in correcting for image distortion. The size of the model is 201 by 201 grid points with a single point scatterer located in the middle of a homogeneous model with $v = 2500$ m/s and $Q = 25$; and the seismic data are computed using viscoacoustic Born modeling with a 15-Hz Ricker wavelet. Figure 3.1a shows the migration image using acoustic RTM to migrate the lossy data. Compared with the actual scatterer location, which is indicated by the red star, the migration response of the point scatterer is mis-located. After applying the hybrid DFs, these errors have been successfully corrected in Figure 3.1b.

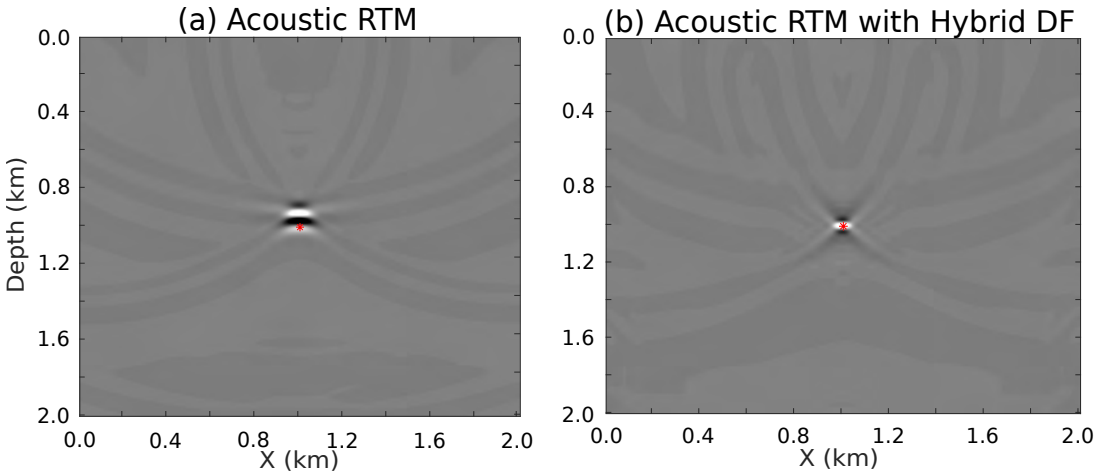


Figure 3.1: Comparison between images from (a) acoustic RTM and (b) acoustic RTM with a hybrid DF.

3.4.2 Sensitivity of Deblurred Images with Respect to Errors in the Attenuation Model

To check the sensitivity of the hybrid DFs to errors in the estimated Q model, numerical tests are carried out based on the previous scatterer model for different background Q values.

Figure 3.2a shows the migration image migrated by acoustic RTM and the red star indicates the correct point-scatterer location. The attenuation distortions in Figure 3.2a are corrected in Figure 3.2b by applying the hybrid DFs to the acoustic RTM image. In this case, these filters are estimated based on the correct background Q model. However, the hybrid DFs become less effective as the errors of the estimated Q model increase.

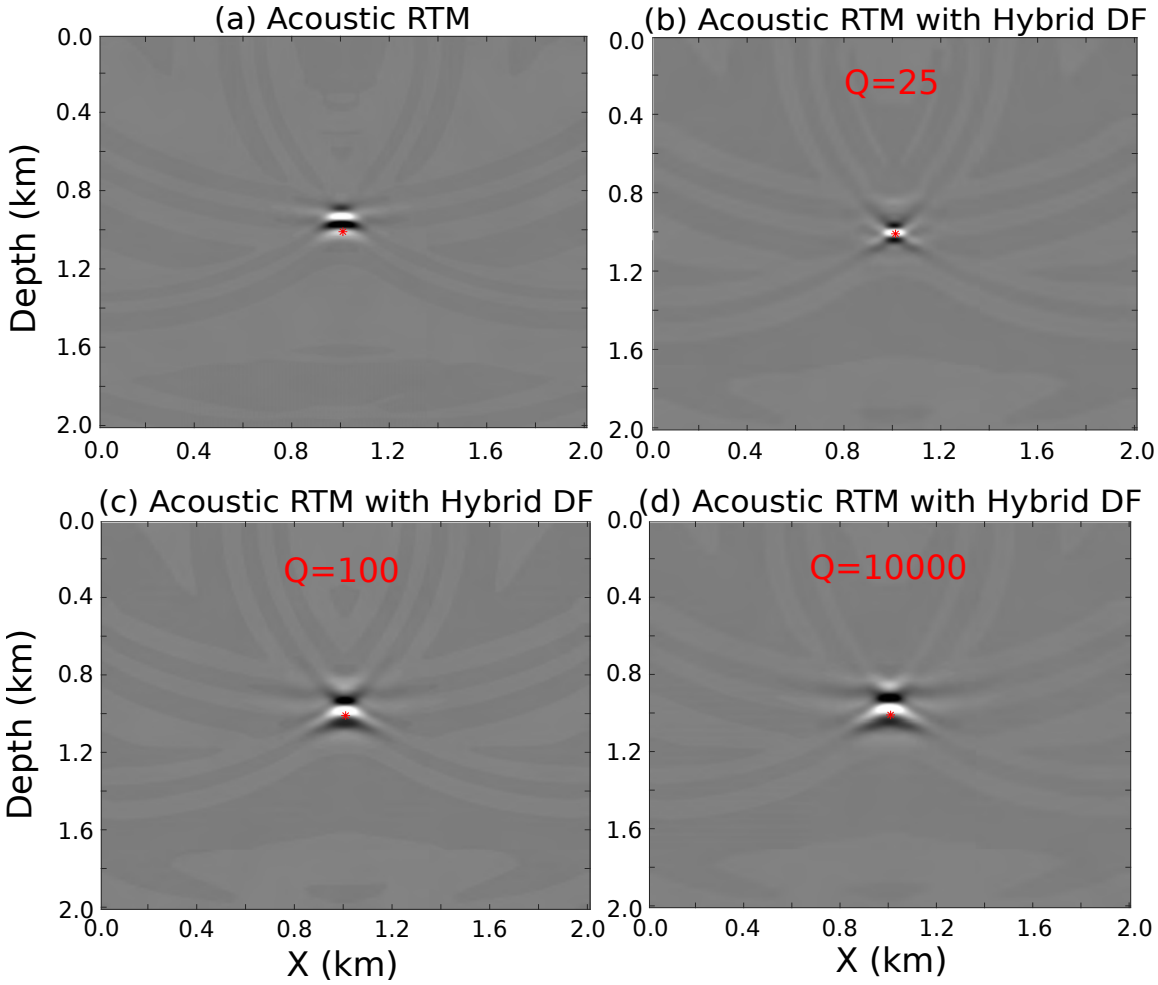


Figure 3.2: Comparison between (a) acoustic RTM image, deblurred images with (b) Q=25, (c) Q=100 and (d) Q=10000, respectively.

To illustrate this point, Figure 3.3 shows the performance of the DF with respect to different errors in the estimated Q model by comparing the deblurred image to the reference image. The reference image is obtained by applying the acoustic LSRTM to the acoustic data. As the errors in the Q model increase, the similarity between the deblurred image and the reference image decreases quadratically, where the similarity is quantified as the correlation between the deblurred image and the reference image. We can see that when the error is larger than 40 percent, the hybrid DF becomes less effective in correcting the attenuation distortions.

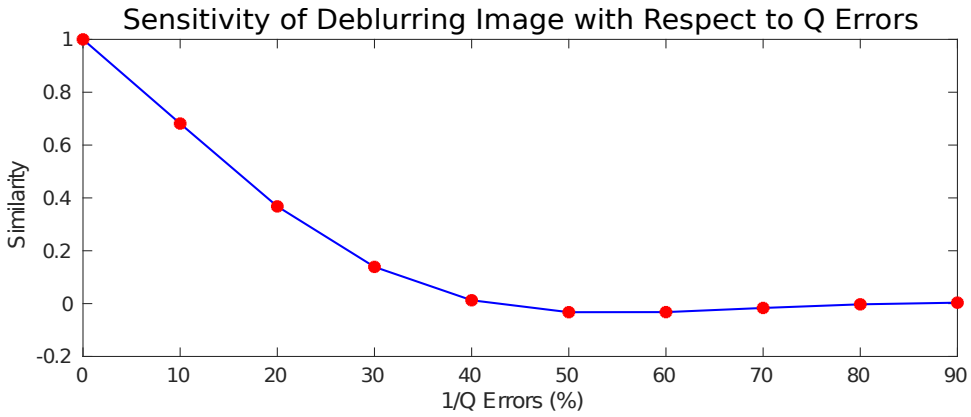


Figure 3.3: Sensitivity of deblurred image with respect to the errors in the attenuation model.

Therefore, a reliable Q model is required for the hybrid compensation method. For data-domain methods, Quan and Harris (1997) used the centroid-frequency shifts between the predicted and the observed traces and smeared the shifts along raypaths to update the Q model. For full waveform inversion (FWI)-like algorithm, Bai et al. (2014) and Cheng* et al. (2015) used the waveform difference to invert for the Q model. Dutta and Schuster (2016) developed a skeletonized wave-equation Q inversion method that finds the Q model that minimizes the differences of the peak frequencies between the observed and the predicted transmission arrivals. This skeletonized method is less prone to the cycle skipping compared to the waveform difference misfit function. For image-domain algorithms, Shen* et al. (2014) invert for the Q model by minimizing the spectral ratio difference between a Q-image and a target image, which is attenuation-free. The above Q inversion methods provide a varity of options for estimating the Q model that is needed for the hybrid compensation method.

3.4.3 Effectiveness and Limitations of the Hybrid Compensation

We now test the DF on migration images generated from different Q models. Figures 3.4a, 3.4c, 3.4e and 3.4g show the acoustic migration results with the observed datasets generated from the homogeneous attenuation models with the $1/Q = 0.01$, 0.04, 0.07 and 0.1, respectively. The amplitudes and phase distortions in the acoustic migration image become increasingly severe as $1/Q$ increases. We then apply the hybrid DFs to these images to correct the attenuation distortions and the results are shown in Figures 3.4b, 3.4d, 3.4f and 3.4h. In all four cases, the amplitude and phase distortions in these images are well recovered. The blue curve in Figure 3.5 represents the similarity between the hybrid compensated images and the reference image. Here, the reference image is obtained by applying acoustic LSRTM to acoustic data. We can see that as the attenuation increases from $1/Q = 0.05$ to $1/Q = 0.1$, the performance of the hybrid DF decreased by 22%, which is acceptable compared to a migration image without compensation. The reason for the performance degradation is that as the attenuation increase, the spatial resolution and the amplitude of the migration Green's function decrease, so that solving for the hybrid DFs becomes more ill-conditioned.

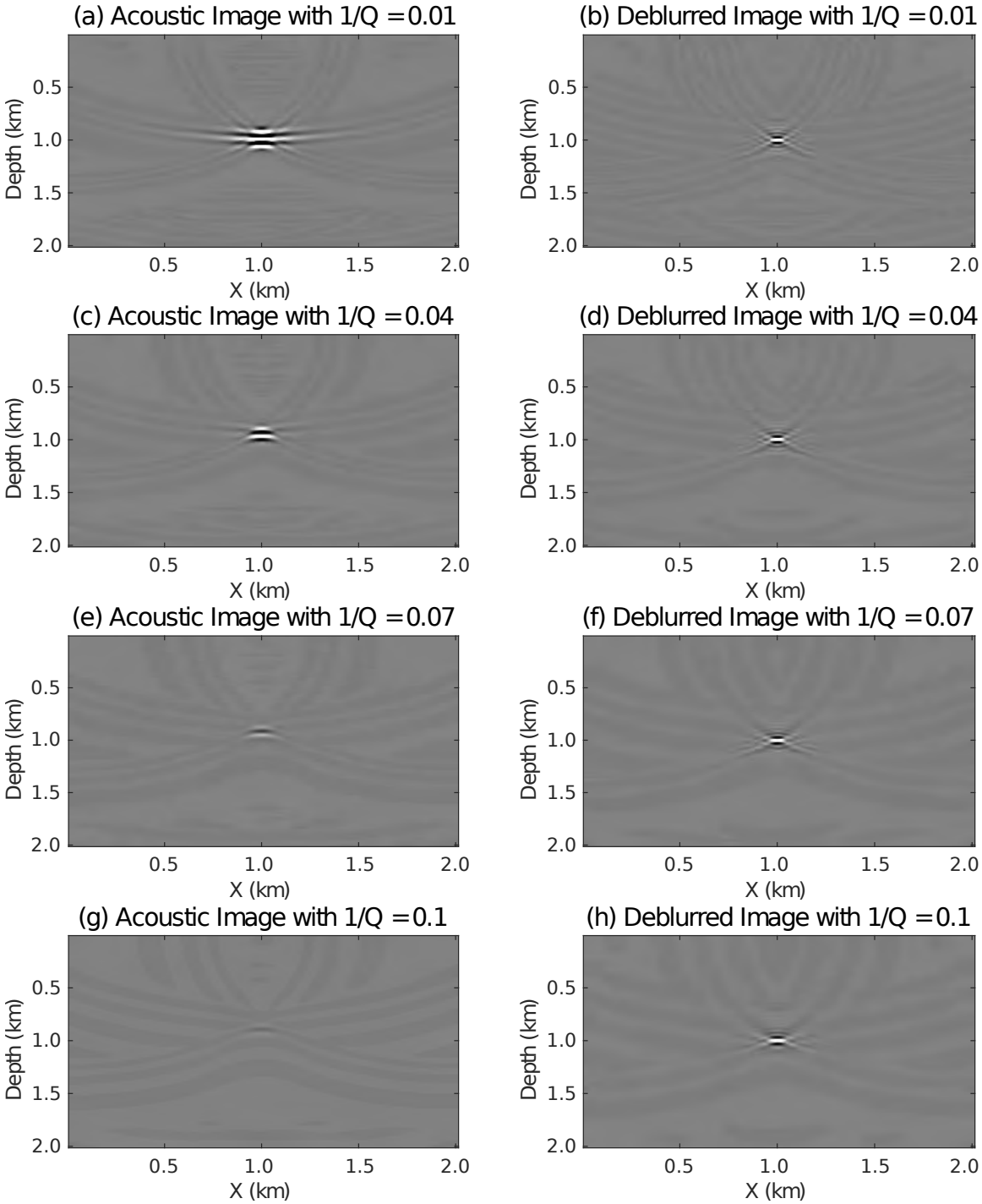


Figure 3.4: Acoustic migration results with the observed datasets generated by the homogeneous attenuation model with 1/Q = (a) 0.01, (c) 0.04, (e) 0.07 and (g) 0.1, respectively. The corresponding deblurred results are shown in (b), (d), (f) and (h).

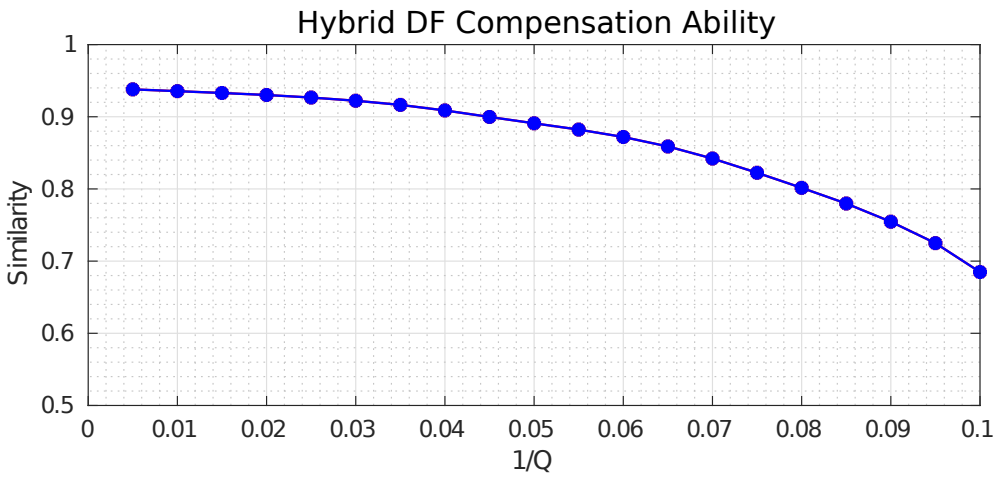


Figure 3.5: Effectiveness of the hybrid deblurring and the attenuation factor 1/Q.

A key assumption with the DF is that the migration Green's function is spatially invariant in the filter window. This assumption is violated in a geophysical medium with a strong spatial heterogeneity, so appropriate remedies must be used. One such remedy is to overlap the adjacent filter windows and to make the windows smaller. However, these windows cannot be too small otherwise the filters will not be able to compensate for intermediate-wavenumber distortions. In this paper, the reference model is a set of uniformly distributed point scatterers. There are other options. For example, a reference model which follows the geology of the model may have the potential benefit over the uniform one (Aoki and Schuster, 2009).

3.4.4 Marmousi II Model

Acoustic RTM with hybrid DFs is now tested on viscoacoustic data generated for the Marmousi II model. Figure 3.6 shows the true velocity and 1/Q models. A Ricker wavelet with a peak frequency of 15 Hz is used as the source wavelet and a fixed-spread acquisition geometry is employed where 150 sources are evenly distributed on the surface at a source interval of 50 m. The data are recorded by 800 receivers for each shot, where the receivers are uniformly distributed every 10 m on the surface.

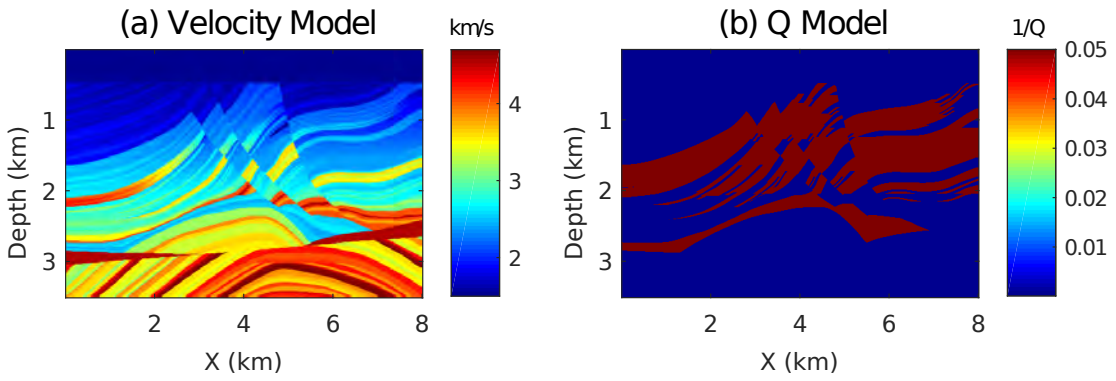


Figure 3.6: The Marmousi model: (a) true velocity model and (b) true Q model.

Figure 3.7a shows the conventional acoustic RTM image computed from viscoacoustic data. This image fails to recover the correct amplitudes of the deeper reflectors, and some reflectors are located at the wrong position. The Q-LSRTM image, shown in Figure 3.7b, shows an improvement in the deeper layers when compared to the acoustic RTM result. However, the Q-LSRTM image has a lower resolution for those reflectors below the Q anomaly compared to the benchmark migration image shown in Figure 3.7d. More iterations are needed for Q-LSRTM to get a more balanced image with higher resolution. Here, the benchmark image is the result of acoustic LSRTM applied to acoustic data generated from the same velocity model in Figure 3.6a. To mitigate the amplitude-loss and phase-shift problems in the acoustic RTM image, we apply the hybrid DFs to the image in Figure 3.7a. The result is shown in Figure 3.7c, which has better resolution with more balanced amplitudes, and the reflectors are at the correct locations.

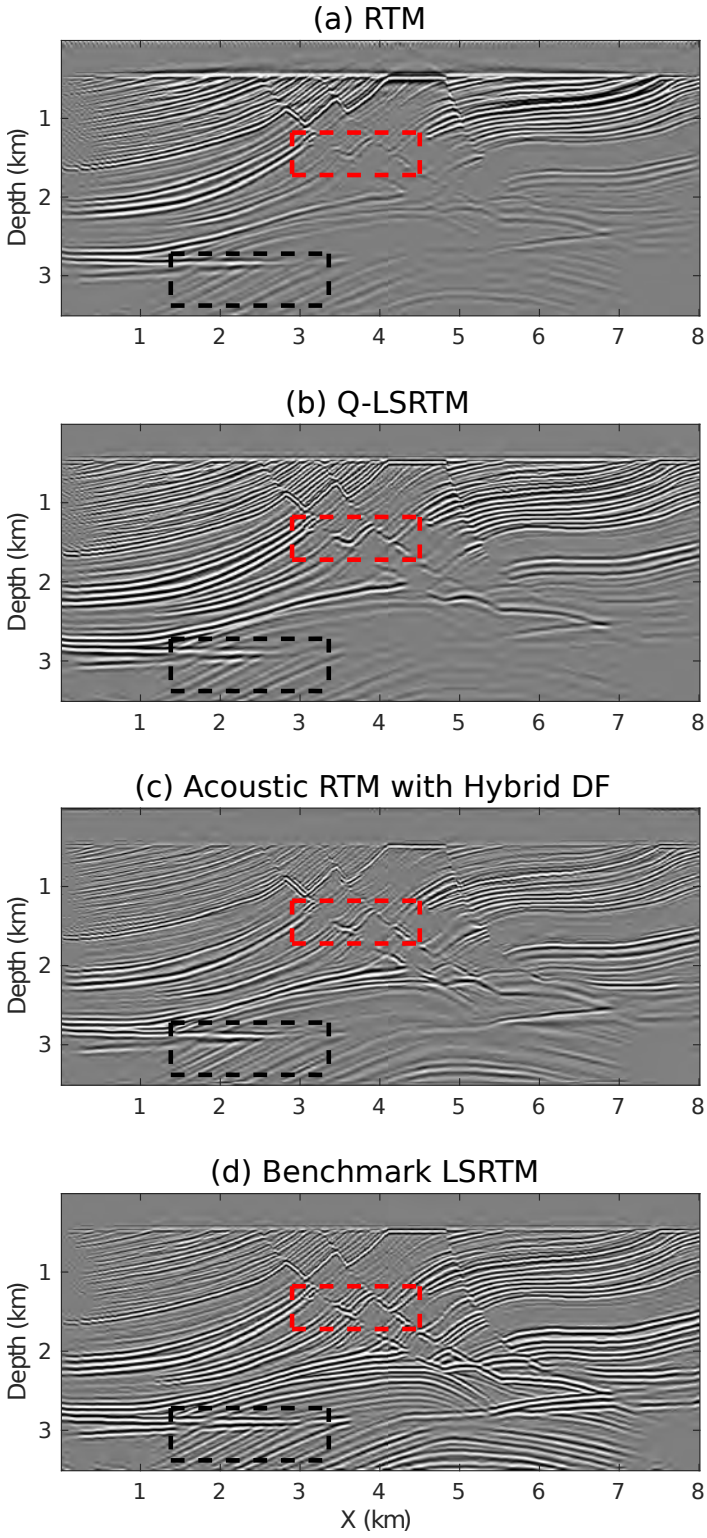


Figure 3.7: Comparison between images computed from the viscoacoustic data by (a) acoustic RTM, (b) Q-LSRTM, (c) acoustic RTM with hybrid DFs and (d) acoustic LSRTM for lossless acoustic data, which is used as the benchmark image. For these least-squares images, 20 iterations are carried out in all cases.

Magnified views of these images are compared in Figures 3.8 and 3.9, where the red arrows point to the areas with noticeable improvements. The correct location of the reflectors are indicated by the blue dashed lines in Figures 3.8 and 3.9, where the reflectors in Figures 3.8a and 3.9a are at the wrong places compared to the other images. This is due to the fact that the acoustic migration operator cannot correct the phase-shift problem in the lossy data. In contrast, the hybrid DF successfully corrects for phase and amplitude distortions as shown in Figures 3.8c and 3.9c.

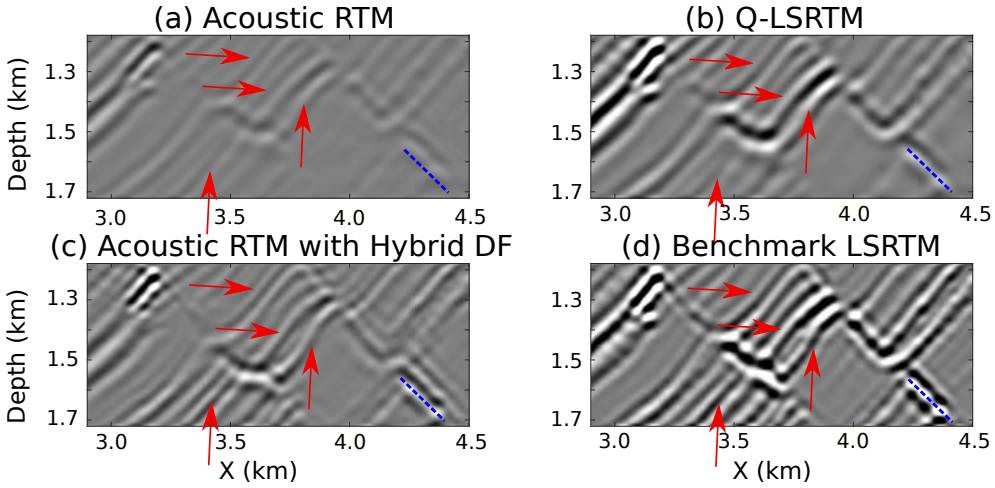


Figure 3.8: Magnified views of the red boxes in Figure 3.7. The red arrows point to the areas with noticeable improvements and the blue dashed line indicates the true locations of the reflectors.

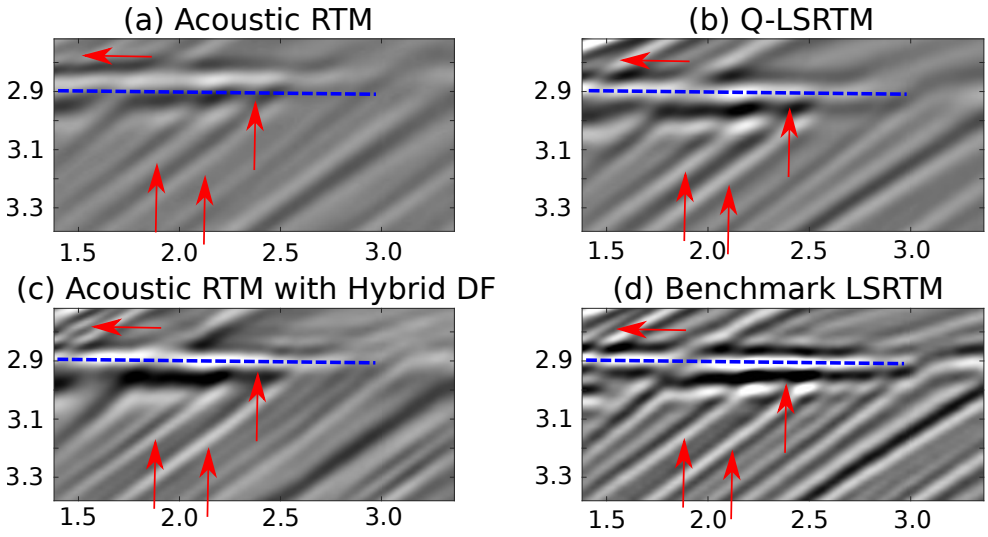


Figure 3.9: Magnified views of the black boxes in Figures 3.7. The red arrows point to the areas with noticeable improvements and the blue dashed line indicates the true locations of the reflectors.

3.4.5 Friendswood Crosswell Field Data

We now test the hybrid DF method on the Friendswood crosswell field data set. Two 305-m-deep cased wells separated by 183 m were used as the source and receiver wells. Downhole explosive charges were fired at intervals of 3 m from 9 m to 305 m in the source well, and the receiver well had 96 receivers placed at depths ranging from 3 m to 293 m. For these data, the migration velocity and Q models are shown in Figures 3.10a and 3.10b, respectively. The migration velocity model is estimated by early-arrival waveform inversion and the migration Q model is estimated by wave-equation Q tomography (Dutta and Schuster, 2016).

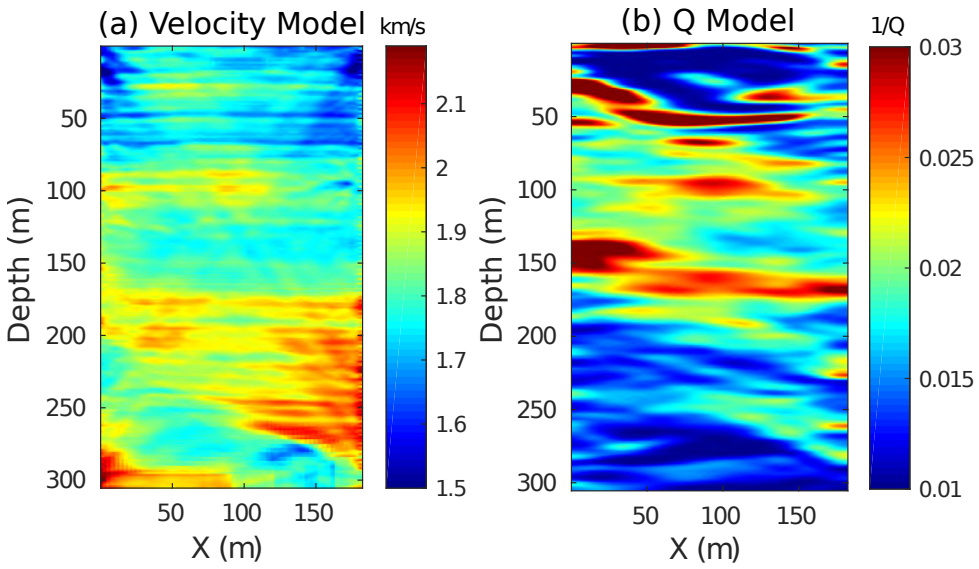


Figure 3.10: (a) The estimated migration velocity and (b) Q models for the Friendswood crosswell data.

The comparison between the acoustic RTM and Q-LSRTM images are shown in Figures 3.11a and 3.11b, respectively. Similar to the synthetic examples, the amplitudes are more balanced in the Q-LSRTM image in Figure 3.11b than those in the acoustic RTM image in Figure 3.11a. At the depths of 25 m and 48 m, two strong and continuous reflectors are shown in the Q-LSRTM image, which is barely visible in the RTM image. By applying the hybrid DF to Figure 3.11a, these reflectors are recovered in the deblurred image. Moreover, Figure 3.11c also suggests that the amplitudes in the deblurred image also become more balanced and the phase problem is largely mitigated.

On the other hand, there are also some artifacts in the deblurred image. At the depth of 250 m, there are three continuous reflectors in the Q-LSRTM image in Figure 3.11b, and yet these reflectors can hardly be seen at the right part of the deblurred image in Figure 3.11c. The possible reason is that the estimated Q model is not accurate enough in this area.

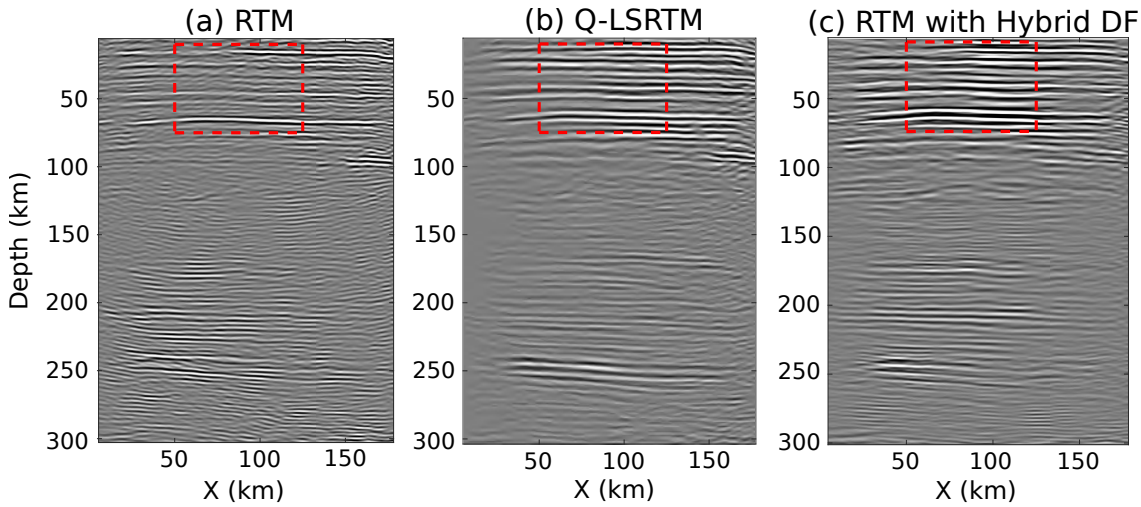


Figure 3.11: Comparison between the images from (a) acoustic RTM, (b) Q-LSRTM and (c) acoustic RTM with hybrid DFs.

Figure 3.12 depicts the magnified views of the black box in Figure 3.11, where the arrows point to the areas with noticeable improvements. Some missing reflectors in the RTM image are well recovered by the hybrid DF in the deblurred image shown in Figure 3.12c. Figure 3.12a shows a deviation of the migrated reflector from the correct location (marked by the red line). This shift is corrected in Figure 3.12c by the hybrid DF. The comparison of the vertical profiles in Figure 3.11 at x=68 m are shown in Figure 3.13, where the acoustic migration result (blue line) is weakly correlated with the Q-LSRTM image (red line) in Figure 3.13a. However, after applying the hybrid DF to the acoustic migration image, the deblurred image (black line) correlates well with the Q-LSRTM image (red line) in Figure 3.13b.

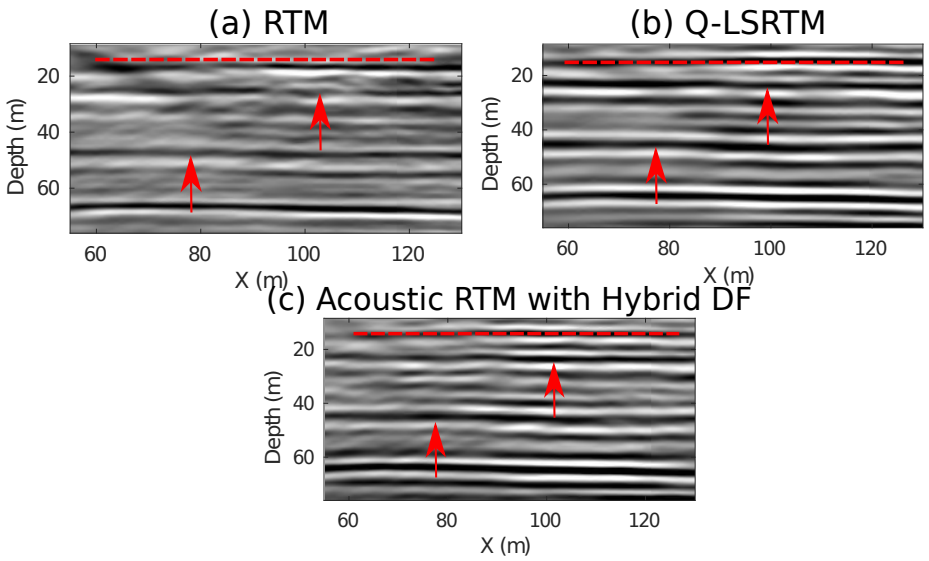


Figure 3.12: Magnified views of the black boxes in Figure 3.11.

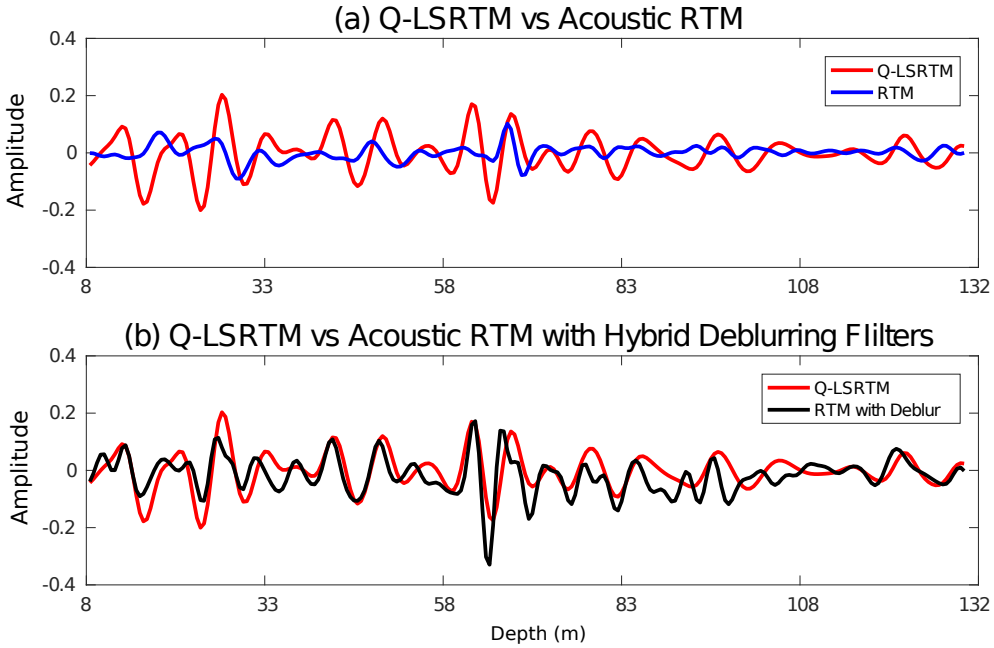


Figure 3.13: Comparison of the vertical profiles in Figure 3.11 at x=68 m.

3.5 Discussion

In this section, we compare the performance between the hybrid and viscoacoustic DF in terms of both image quality and computational cost. Numerical tests are carried

out based on the previous scatterer model for different background Q values. Figure 3.14 show the comparison result of the image quality between the hybrid deblurring approach and the visco deblurring method. The red curve in Figure 3.14 depicts the similarity between the visco deblurred image and the reference image, which decreases as the attenuation increases. The blue curve is the similarity between the hybrid deblurred image and the reference image. For weak attenuation, the performance of the viscoacoustic DF is about 5% to 10% better than the hybrid approach. However, as the attenuation increase, the performance of the two methods become similar.

For the computation cost comparison, a fixed-spread acquisition geometry is employed where 99 sources are evenly distributed on the surface. Each shot has 199 receivers on the surface. The source wavelet is a Ricker wavelet with the peak frequency of 15 Hz. The simulation time is 2s with a time interval of 0.001s. We use the viscoacoustic wave-equation which is derived based on the standard linear solid model (SLS) with one relaxation mechanism. The computation time of the hybrid and viscoacoustic DF approach is 139 s and 200 s, respectively. Therefore, the hybrid approach is about 30% faster than the visco approach. For a 3D case, or for a viscoacoustic wave-equation with more than one relaxation mechanisms, more computation time can be saved by using the hybrid deblurring approach.

3.6 Conclusion

An acoustic hybrid RTM method is presented that uses hybrid DFs to compensate for the amplitude losses and phase shifts due to strong attenuation in the subsurface. Numerical tests on synthetic and field data show that the proposed method can significantly mitigate the problem of amplitude loss and phase shift in acoustic RTM images when migrating lossy data. This technique also increases spatial resolution in the migration image and can reduce the computational costs and storage requirements by a factor of $O(N-1)$ times compared to iterative Q-LSRTM. Here, N is the number

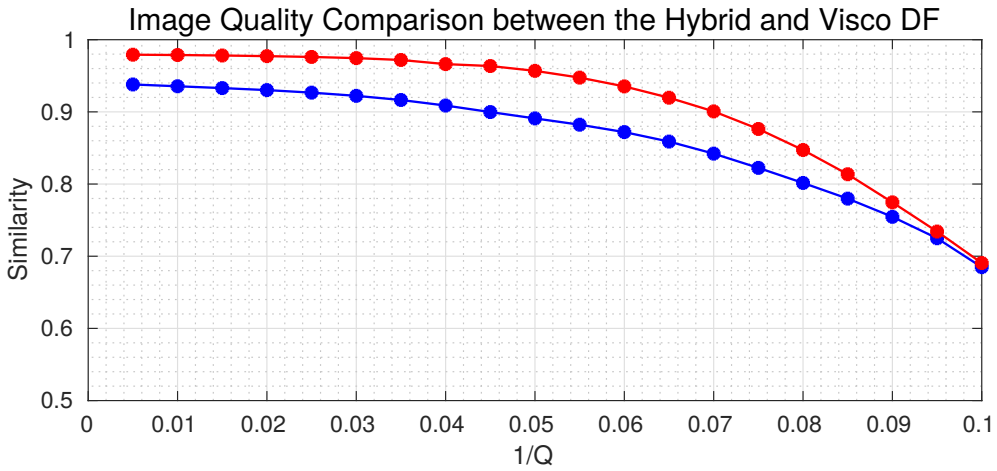


Figure 3.14: The relationship between the performence of the hybrid and viscoacoustic DF.

of iterations.

The limitations of this method are that the hybrid DFs are more suitable for correcting high- and intermediate-wavenumber distortion errors in a migration image caused by inadequate compensation for attenuation distortion. Compensating for the low-wavenumber errors with a larger DF will significantly increase the computational costs. For strong distortions, such as using acoustic migration on strongly anisotropic data, the hybrid DFs are not able to correct for these effects. The performance of the hybrid DF relies on the accuracy of the estimated Q model, otherwise, will degrade in performance with decreased accuracy in the Q model.

Chapter 4

Suppressing Migration Image Artifacts Using a Support Vector Machine Method

Reverse-time migration (RTM) produces images of the subsurface reflectivity, but migration images may deteriorate when the source and/or receiver spacings are irregular and their distributions are sparse. The resulting migration artifacts can obscure the identification of important geological features of interests. We develop a new filtering method using the support vector machine (SVM) to reduce these migration artifacts. This filtering method distinguishes noise from signal in migration images automatically. The SVM filtering predicts a muting window in the dip-angle angle-domain common image gathers (ADCIG) to suppress the artifacts. Once the SVM is sufficiently trained, it can be applied to out-of-the-training set data to remove migration artifacts in the dip-angle ADCIGs. The numerical results show that the SVM filtering produces migration images with better signal-to-noise ratios (S/N) and fewer aliasing artifacts compared with those obtained using the conventional RTM method.

4.1 Introduction

Reverse-time migration (RTM) produces images contaminated with migration artifacts if the source and receiver distributions are sparse. These artifacts degrade amplitude versus offset analysis and interfere with the ability of an interpreter to identify important geological features of interests. Such migration artifacts can also slow down the convergence rate of iterative least-squares migration (LSM).

To alleviate this problem, least-squares migration (LSM) with preconditioning or regularization is often used to suppress migration artifacts. Cabrales-Vargas and Marfurt (2013) developed a preconditioned LSM method with an additional roughness penalty term added to the L_2 misfit function. The method preserves the amplitudes in migration images and reduces aliasing artifacts compared with the conventional LSM. Dutta and Schuster (2015) used a sparse least-squares reverse-time migration (LSRTM) method with the seislet transform as a change of basis for the reflectivity. Along with a dip-constrained preconditioner, this method can produce an image with more meaningful structural updates. Lin et al. (2016) incorporated a L_p-norm-based compressive-sensing term in the misfit function of LSRTM to improve migration imaging of sparse seismic data. Dutta (2017) introduced a local Radon-based preconditioned LSRTM method to alleviate the problems of low S/N data and image aliasing. However, least-squares-based methods require a large number of expensive iterations to produce the desired enhancement in the image quality. Yu and Hornby (2008) implemented a stereographic local beam imaging scheme to attenuate migration noise caused by limited acquisition apertures in VSP data. Zhou et al. (2010) used a local-angle-domain correlation imaging condition to attenuate migration artifacts. Dafni and Symes (2016a,b) and Liu and Zhang (2018) designed a specularity filter based on the semblance equation to preserve the signal and suppress migration artifacts in the dip-angle angle-domain common image gather (ADCIG). The specular reflections are assumed to have higher semblance scores compared with those of migration artifacts as the formers are more horizontally coherent, defined as the "semblance criterion." The normalized semblance scores should be close to one for the signals so the specularity filter can better preserve their amplitudes. By contrast, the semblance scores for migration artifacts should be nearly zero in order to identify them as noise. However, this assumption fails for an image with severe coherent artifacts, because such noise can appear as a signal in the dip-angle ADCIG

incorrectly. In such a case, using the semblance criterion is not enough to accurately distinguish between the signals and migration artifacts. The semblance scores could also approach to 1 for the artifact points. Therefore, the specularity filter based on the semblance equation has the potential to consider strong coherent noise as signal.

In this chapter, we develop a support vector machine (SVM) filtering method to suppress the migration artifacts. The migration image is first transformed to the dip-angle ADCIGs where the signals and artifacts are separable from each other. Then the SVM algorithm is used to distinguish between the signal and artifacts. The input data vector for the SVM method contains the skeletal features of a dip-angle ADCIG such as coherency, amplitude, and dip-angle. More features are used indicate more strict rules are applied to separate the signals from artifacts in the dip-angle ADCIG, which reduce the possibilities of misclassification. The SVM filtering method predicts a weighting coefficient for each point in the dip-angle ADCIG and the weighting coefficient approaches the value of 1 for a migration signal point and approaches -1 for a migration artifact point. Therefore these weighting coefficients can be used to preserve the amplitude information of the real image components and effectively suppress migration artifacts. Once the filter is applied to the dip-angle ADCIG, the cleaned dip-angle ADCIG is inverse transformed back to the image domain to produce the final migration image which is free from coherent noise. Both synthetic and field seismic data are used to verify the effectiveness of the SVM filtering method. In the synthetic tests, we mainly focus on reducing the migration artifacts caused by a sparse source-receiver geometry.

This chapter is organized as follows. After the introduction, we introduce the theory and workflow of the SVM filter. We then present numerical results in both the synthetic- and field-data examples, discuss the limitations of the method and give conclusions.

4.2 Theory

In the dip-angle angle-domain common image gather (ADCIG), a seismic reflector image is focused around a specular dip-angle as a spot-like response. This focusing distinguishes signal from any artifacts events, and the spot-like response in the dip-angle domain has a horizontal phase orientation no matter how steep a reflector is dipping (Dafni and Symes, 2016a,b). To compute the dip-angle ADCIG, the subsurface-offset extended image is computed firstly using an extended image condition. Then the forward and inverse local Radon transforms are used to transform the subsurface-offset extended image to and back from the dip-angle domain, respectively (Dafni and Symes, 2016a,b).

Figure 4.1 shows an RTM image with 4 reflectors, each with different shapes. Figure 4.2a shows the suboffset gather at x = 3.5 km indicated by the red dashed line, where the reflector's energies focus around zero suboffset as we migrate the data using the correct velocity. The corresponding dip-angle ADCIG is shown in Figure 4.2b, where these reflectors are locally coherent with respect to the specular dip-angle. This feature enables us to remove the unwanted artifacts by keeping the spot-like energy around the specular dip-angle.

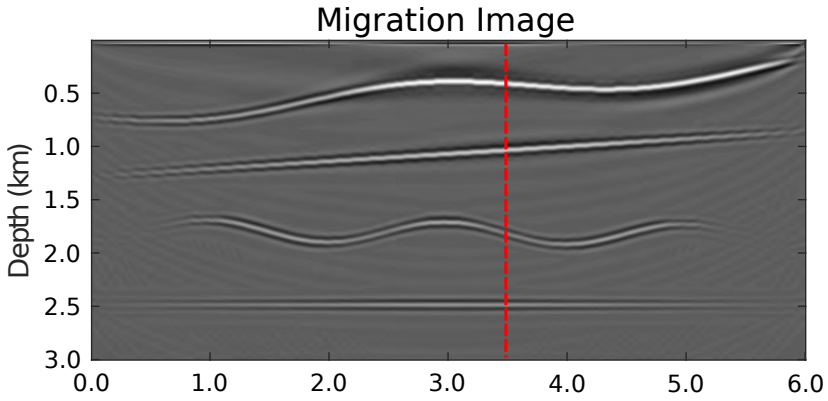


Figure 4.1: Migration image of a five-layer model.

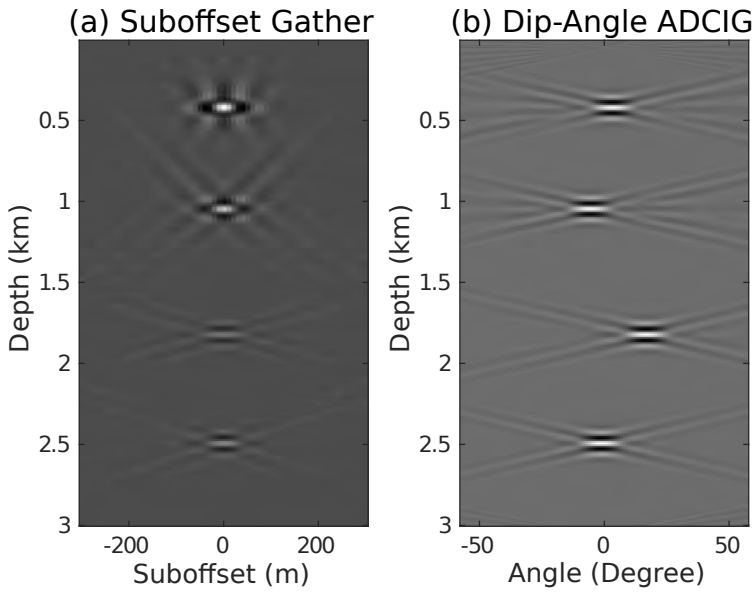


Figure 4.2: The suboffset gather (a) and corresponding dip-angle ADCIG (b) at x = 3.5 km in Figure 4.1.

4.2.1 Semblance Filtering

The semblance equation (Neidell and Taner, 1971) is commonly used to evaluate the local coherency for each image point in the dip-angle ADCIG (Dafni and Symes, 2016a; Liu and Zhang, 2018)

$$S(z,\theta) = \frac{1}{2N_\theta + 1} \frac{\sum_{iz=z-N_z}^{iz=z+N_z} \Big(\sum_{i\theta=\theta-N_\theta}^{i\theta=\theta+N_\theta} ADCIG(iz, i\theta) \Big)^2}{\sum_{iz=z-N_z}^{iz=z+N_z} \sum_{i\theta=\theta-N_\theta}^{i\theta=\theta+N_\theta} ADCIG^2(iz, i\theta)}, \tag{4.1}$$

where N_z and N_θ are the half-length of the windows along the depth and the dip angle, respectively. The choice of the window size depends on the typical size of a spot-like response in the dip-angle ADCIG. For each image point in the dip-angle ADCIG, a local area is first selected based on the window size. Equation (4.1) is then used to measure the coherence of each local window and the semblance scores are further used as the weighting coefficients to filter out artifacts and preserve the signals. This assumes that the signals in the dip-angle ADCIG have much higher

semblance scores compared with those of artifacts.

The semblance filtering method assumes that only reflectors are locally coherent in the dip-angle ADCIG. However, in the case of severe coherent artifacts, the artifacts might masquerade as false structures. Such structures are also locally coherent in the dip-angle ADCIG but often focus at the large dip-angle range (not the specular dip angle). In this case, these artifacts also receive high semblance scores and cannot be removed using the semblance filtering method. Moreover, the semblance scores are not always one for signal points and zero for artifact points, which might degrade SVM's ability to remove coherent noise.

4.2.2 Support Vector Machine Filtering

To alleviate the coherent noise problem, we employ support vector machine (SVM) filtering to separate signals from artifacts in the dip-angle ADCIG.

Support Vector Machine

The SVM is a supervised machine learning algorithm that sorts data into two categories (Cortes and Vapnik, 1995). Supervised learning by SVM uses a set of training data consisting of N training samples

$$(\mathbf{x}^{(i)}, y^{(i)}), \tag{4.2}$$

where $i = 1, 2, ..., N$, $\mathbf{x}^{(i)}$ is the i^{th} training data instance vector with dimension $D \times 1$ and $y^{(i)} = 1$ or -1 is the binary label for $\mathbf{x}^{i}$. D indicates the number of features for each data instance. For a set of two-dimensional $\mathbf{x}^{(i)}$, the goal of SVM is to determine the dashed line that has the greatest separation from the red and blue dots in Figure 4.3. The optimal dashed line has the fattest margin width and is equidistant from the solid lines that are parallel to one another. This dashed line, also known as the decision line, is mathematically defined by the normal vector $\mathbf{w}$ and the intercept

b, where any vector $\mathbf{x}$ on the decision line satisfies $w_1x_1 + w_2x_2 + b = 0$. The red and blue dots intersecting with the black solid line are known as the support vectors. These support vectors define the points with the closest perpendicular distance to the decision line. The decision lines in Figures 4.3a and 4.3b are also able to separate the two classes of data, but they are sub-optimal compared with that in Figure 4.3c. The margin thicknesses of the decision lines in Figure 4.3a and 4.3b are thinner than the one in Figure 4.3c. The thinner the margin thickness, the more likely it is to misclassify the feature vector $\mathbf{x}^{(i)}$.

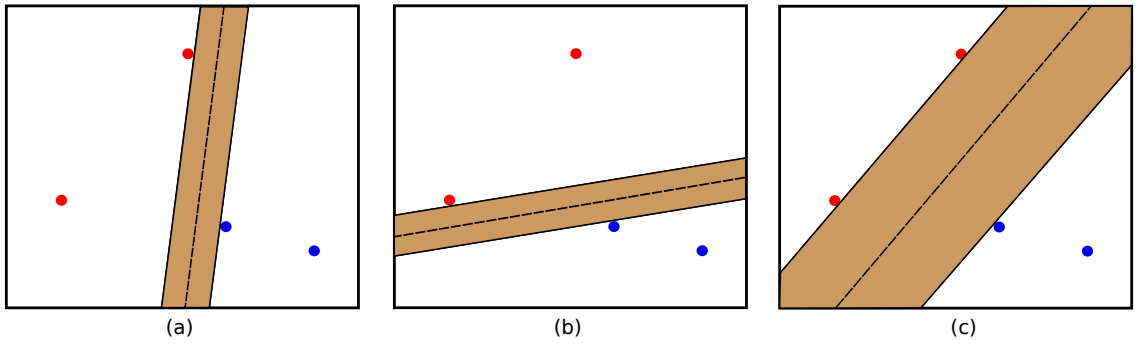


Figure 4.3: Two classes of data with different decision lines, where the red and blue dots represent different classes. Even though the decision boundaries in (a) and (b) cleanly separate the two datasets, they are not optimal separators compared with the decision boundary in (c) which has the fattest thickness margin.

Once the normal vector $\mathbf{w}$ and the intercept b are determined after sufficient training, the SVM classifies new points $\tilde{\mathbf{x}}^{(i)}$ that are out of the training set. Figure 4.3 depicts samples in a two-dimensional case. In a more general case where the samples are in a D-dimensional space, SVM learns a D-1 dimension hyperplane to separate different classes of data. The samples $\mathbf{x}$ on this hyperplane satisfy

$$\mathbf{w} \cdot \mathbf{x}^{(i)} = w_1x_1 + w_2x_2 + w_3x_3 + ... + w_Dx_D + b = 0, \tag{4.3}$$

where $\mathbf{w}$ is determined by solving a minimization problem

$$arg\min_{\mathbf{w},b} d = \frac{1}{2}\|\mathbf{w}\|^2,$$
$$\text{subject to } y^{(i)}(\mathbf{w}\cdot\mathbf{x}^{(i)} + b) \geq 1, \tag{4.4}$$

where $i = 1, 2, ..., N$.

Solving the minimization problem determines the optimal line with the fattest thickness of margin. Inequality (4.4) ensures that $\mathbf{w}$ cleanly separates the training set $\mathbf{x}^{(i)}$. A numerical solution to the above minimization problem obtained using, e.g. quadratic programming (Wright and Nocedal, 1999), results in the fattest margin as shown in Figure 4.3.

The workflow for Support Vector Machine Filtering

We use SVM to largely filter out the migration noise artifact points in the dip-angle ADCIG. For a signal point, SVM predicts a weighting coefficient close to 1 compared with the weighting coefficient of -1 for an artifact point. We then reset -1 to 0 to better remove the artifacts by muting it in the dip-angle ADCIG. The detailed workflow is described in the following:

(1) The element of the input 3x1 feature vector consists of the semblance S, the amplitude A of the image in the dip-angle ADCIG domain and the associated dip angle θ. The amplitude A is computed by calculating the envelope of each trace in the dip-angle ADCIG as shown in Figure 4.4b. This assumes that the signal points in the dip-angle ADCIG have a higher amplitude value compared with the coherent noise.

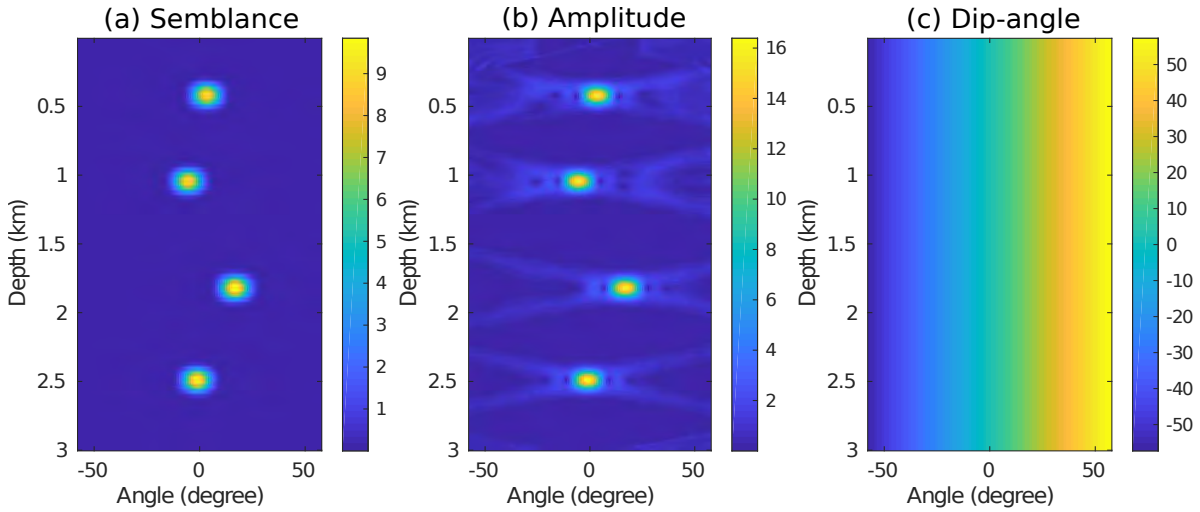


Figure 4.4: The feature map of Figure 4.2b: (a) semblance, (b) amplitude and (c) dip-angle features.

Figure 4.4c shows the image with the dip-angle features. Each column in this feature has the same value that represents the specular dip-angle information, which assumes that the signal responses in the dip-angle ADCIG only distribute over a certain specular angle range. So far, each image point in the dip-angle ADCIG has three features. All three features are used to determine whether an image point in the dip-angle ADCIG is a signal or an artifact point, which dramatically increases the accuracy and confidence in classification.

(2) Building a training set. The training set is a subset of the data for training the model (model weights and biases term). The model learns from this dataset by solving the minimization problem in equation (4.4). We define the training set as $(\mathbf{x}_{train}^{(i)}, y_{train}^{(i)}) = (A^{(i)}, \theta^{(i)}, S^{(i)}, y^{(i)})$ for about 1% of the picked points in the dip-angle ADCIG as shown in Figure 4.5a. The red and blue dots indicate the picked signal and artifact points, respectively. Each point in this training set is manually classified as either signal $y^{(i)} = 1$ or noise $y^{(i)} = -1$. I then reset

-1 to 0 in order to remove the noise more easily. Figure 4.6 shows the training set in the feature domain, where the effective signal and artifact points are cleanly separable from one another. We then use the SVM to determine the 3×1 vector $\mathbf{w} = (w_1, w_2, w_3)$ and the intercept b that cleanly separates the labeled signals from artifacts. This vector and intercept term map a surface in feature space as shown in Figure 4.6b, where the colored plane defines the decision plane in the 3D feature space.

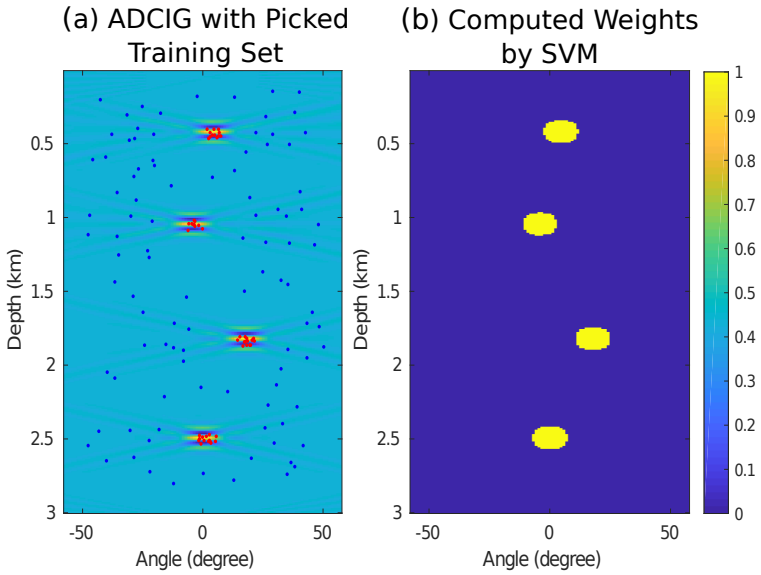


Figure 4.5: (a) The dip-angle ADCIG with the picked training set. The red and blue points indicate the picked signal and artifact points, respectively. (b) The decision value of SVM.

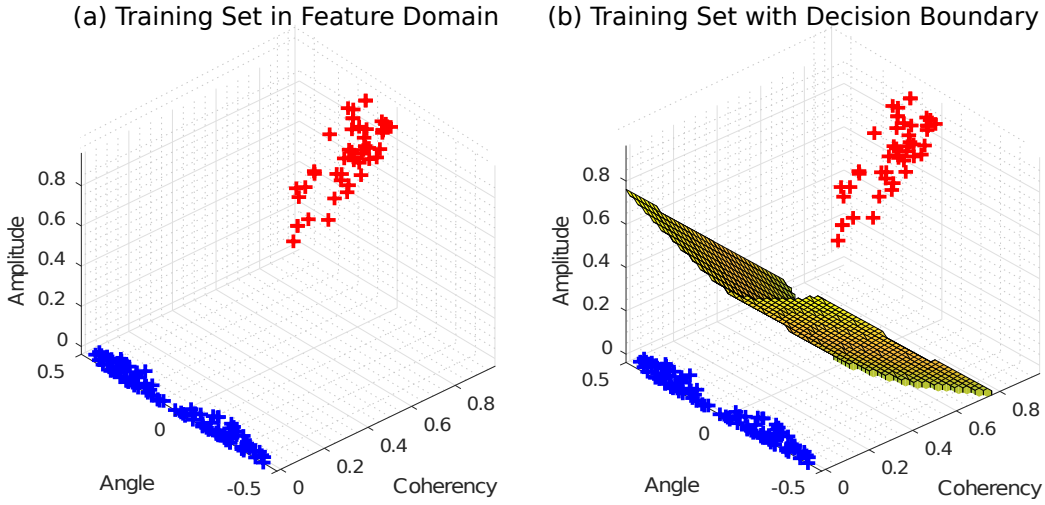


Figure 4.6: (a) 3D plots of the training set in the 3D feature domain, where the red and blue crosses indicate the signal and artifact points, respectively. The signal points and the artifact points are separable in this 3D feature domain. (b) The signal and artifact points in the training set are well separated by a colored hyperplane calculated using the SVM algorithm.

(3) Once the normal vector $\mathbf{w}$ and the intercept b are determined, we compute $\mathbf{w} \cdot \mathbf{x} + b$ for all points in all dip-angle ADCIGs, and output weighting coefficients as shown in Figure 4.5b for Figure 4.5a. It can be noticed that the weighting coefficients for the signals and artifacts are 1 and 0, respectively. In this case, these weighting coefficients can preserve the amplitude information of the signals in Figure 4.5a while cleanly suppressing the artifact points.

In practice, a migration image with a simple geological structure varies gently along the horizontal direction. In other words, the features for each dip-angle ADCIGs are similar. In this case, the trained SVM model from one dip-angle ADCIG is often accurate enough to predict all of the weighting coefficients for every image point in every dip-angle ADCIGs. However, a complicated structure is characterized by features at each dip-angle ADCIG that vary dramatically. In this case, we often use one dip-angle ADCIG for SVM training and then use the trained SVM model to predict the weighting coefficients only for the nearby dip-angle ADCIGs. This

procedure is repeated for the rest of the dip-angle ADCIGs.

The computational costs of the SVM filtering method are inexpensive compared with the least-squares-based method because the number of data points in the training set is small. Usually, we use less than 10% of the image points (often around $100 \sim 300$ data points) of the dip-angle ADCIG for training, and the training progress takes only a few seconds of CPU time. After training is finished, the computation of weighting coefficients for the rest of the image points are much cheaper compared with the training because we only need to calculate the vector multiplication $\mathbf{w}^T \cdot \mathbf{x} + b$.

Discussions of Support Vector Machine Filtering

The practical details for implementing SVM filtering are described in the following.

1. Feature scaling is a very important step before using the SVM filtering method. Figure 4.4 shows that the dip-angle feature θ is highly varying in magnitude and range compared with the semblance feature S and amplitude feature A. But since SVM filtering and most of machine learning algorithms use the Euclidian distance between two data points in their calculation, the features with high magnitudes dominate in the distance calculation compared with the features with low magnitude. Without the feature scaling, the SVM algorithm might be mainly affected by the dip-angle feature θ and neglect the other two features, which can lead to significant misclassification.

   Figures 4.7a and 4.7b compare the predicted weighting coefficients computed by SVM without and with feature scaling, respectively. It clearly shows that without feature scaling, the dip-angle feature dominates the classification result. In this paper, we use the mean normalization method for feature scaling:

$$\tilde{\mathbf{x}}^{(i)} = \frac{\mathbf{x}^{(i)} - mean(\mathbf{x}^{(i)})}{max(\mathbf{x}^{(i)}) - min(\mathbf{x}^{(i)})} \tag{4.5}$$

where $\tilde{\mathbf{x}}^{(i)}$ is the ith feature vector after feature scaling. Figure 4.8 shows the scaled result of Figure 4.4, in which the magnitude range of every feature varies from -1 to 1. In this paper, we apply the feature scaling for each dip-angle ADCIG separately.

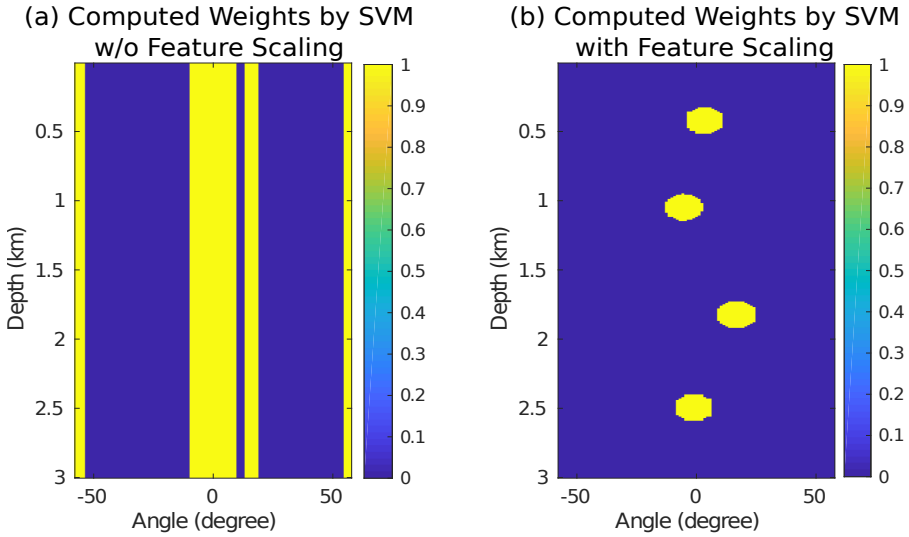


Figure 4.7: The computed weighting coefficients by SVM (a) without and (b) with feature scaling.

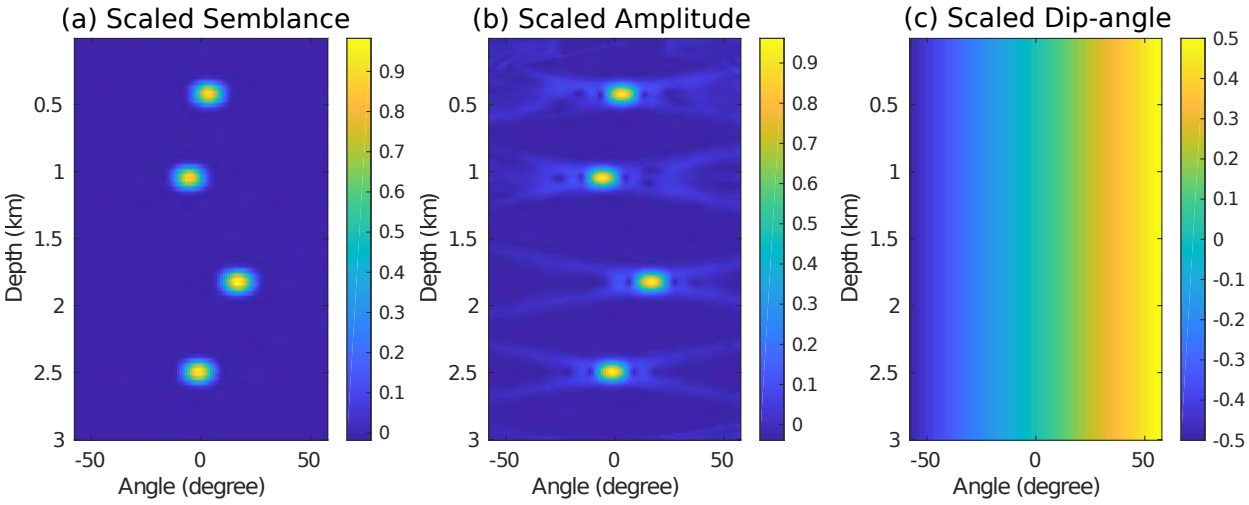


Figure 4.8: The scaled (a) coherency, (b) amplitude and (c) dip-angle features, respectively.

2. How to choose the features? The features for SVM or other machine learning algorithms greatly affect the training efficiency and classification accuracy. A

good feature should be informative, discriminating and independent. we choose the semblance S, amplitude A and dip-angle θ as the features because their values vary dramatically between the signals and artifacts, which can help the SVM algorithm "understand" the difference between the signal and artifacts very quickly.

To demonstrate this characteristics, we choose a new feature type for training and classification. For every image point in the dip-angle ADCIG, we open a small window around the point with a size of $win_z = 25$ by $win_x = 25$ as shown in Figure 4.9. Therefore every pixel in this small patch is a feature and each image point in the dip-angle ADCIG has $win_z \times win_x$ features. Figure 4.10 shows the weighting coefficients computed by SVM using the small image patches as the features, which is less promising compared with the result in Figure 4.5b.

3. How many features are needed to separate the signal and artifact points in the dip-angle ADCIG? Theoretically, the more the features, the higher dimension is the feature space, and in principle, we can more easily separate the signal points from artifacts. The reason is that by analyzing the features, each point in a dip-angle ADCIG can map into a high-dimensional space, where the signal and artifact points are cleanly separable from one another. However, this conclusion is based on the assumption that each feature is independent of, or not linearly related to, one another in the feature space. For example, if we introduce the semblance feature $2S$ as a new feature in the SVM filtering method, this new feature is not able to increase the dimension as it is linearly related to the semblance feature. One simple method to increase the number of features is to apply a nonlinear operation to the existing features, such as take the second power of the semblance feature S^2. To analyze the redundancy of the features, one can use the principal component analysis (Wold et al., 1987).

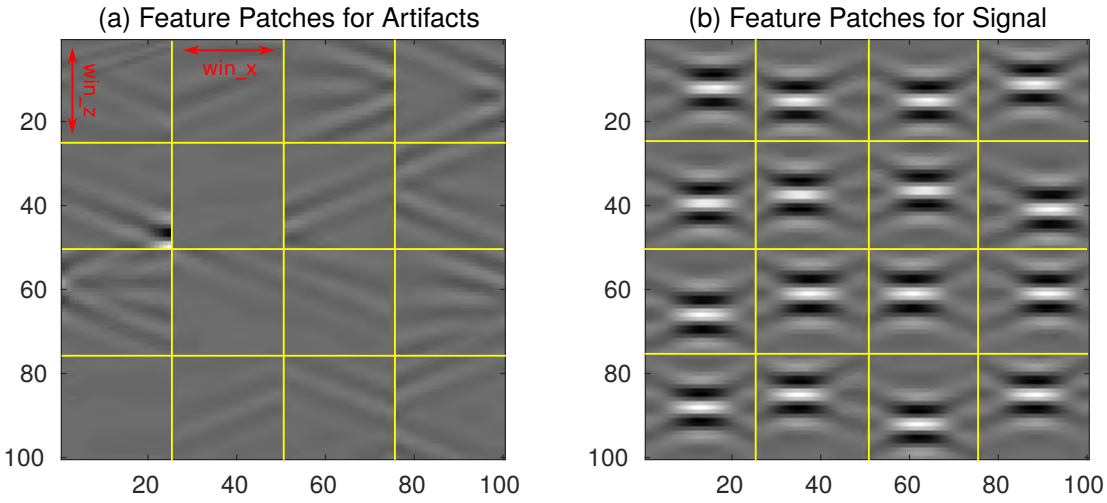


Figure 4.9: The feature patches for (a) artifacts and (b) signal.

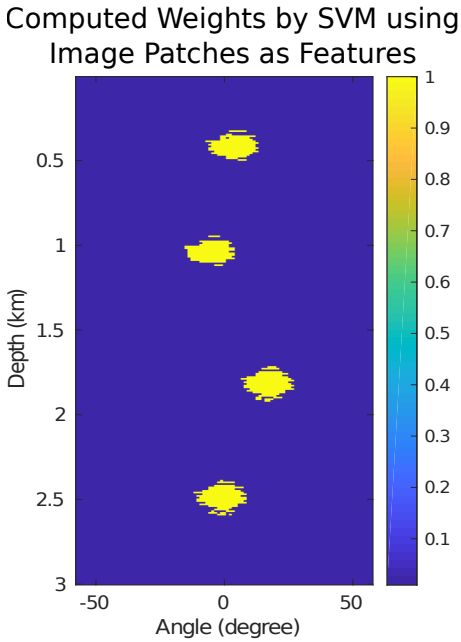


Figure 4.10: The computed weighting coefficients by SVM using the small image patched as features.

4. How to build a proper training set? A proper training set should include most types of signals and artifacts in the dip-angle ADCIG. In this case, the SVM method can classify the rest of the image points in the dip-angle ADCIG with

the reduced chances of misclassification. In the machine learning world, the percentage of the training set is often selected from more than 70% of the entire dataset. However, in our case, we choose less than 10% of the entire image points in the dip-angle ADCIG for training, which is already representative of all types of the signals and artifacts in the dip-angle ADCIG.

4.3 Numerical results

We study the effectiveness of the SVM filtering of migration artifacts for synthetic data and a recorded marine dataset. The marine data are OBC traces recorded in a North Sea seismic survey.

4.3.1 Marmousi Model

We test the SVM filtering using synthetic data for the Marmousi model. Figure 4.11 is the true velocity model used to generate synthetic seismic data. The true velocity model is smoothed with a 5×5 window for migration imaging. The source time function is a Ricker wavelet with a peak frequency of 20 Hz. The synthetic data are computed for 40 sources and 40 receivers on the surface with 200 m of source and receiver intervals, respectively, which is about $2 \sim 3$ times of that of the minimum wavelength.

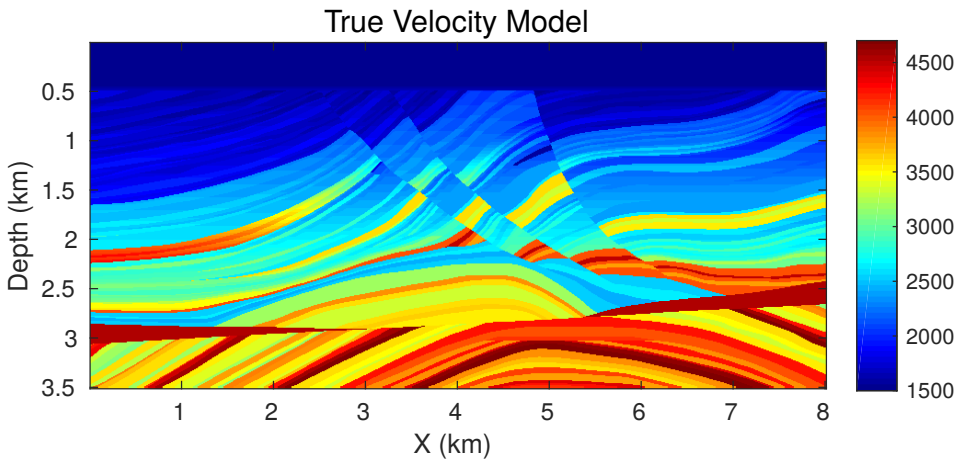


Figure 4.11: Marmousi velocity model.

The conventional RTM image in Figure 4.12a contains severe migration artifacts caused by the sparse data acquisition geometry. The artifacts have strong amplitudes that severely interfere with the true reflectors. In the black box, the strong artifacts break the continuity of the reflectors and cut them into small pieces. The artifacts at the left part of the red box overlap the true reflectors, which make them appear as false structures. These artifacts obscure the actual geology and interfere with the ability of an interpreter to identify important geological features of interests. In the horizontal range between $0 \sim 3$ km, the reflectors in this area vary gently along the horizontal direction. So we choose only the dip-angle ADCIG at $x = 1.5$ km for training. The trained model is then used to compute the weighting coefficients for all of the dip-angle ADCIGs between $0 \sim 3$ km. Similarly, we choose the dip-angle ADCIG at $x = 6.5$ km for training and use the trained model to calculate the weighting coefficients for the dip-angle ADCIGs between $5.5 \sim 8$ km. As the geology structure varies dramatically between $3 \sim 5.5$ km on both the horizontal and vertical directions, we choose the dip-angle ADCIG at $x = 2.97, 3.28, 3.86, 4.5$ for training and only use the trained model to compute the weighting coefficients for nearby dip-angle ADCIGs.

Figure 4.12b shows an improved image where the aforementioned artifacts are

largely suppressed using SVM filtering. Magnified views of these images are compared in Figures 4.13, 4.14 and 4.15, respectively, where there is a noticeable improvement in the image quality. Figure 4.13b shows that reflectors previously covered by strong artifacts now appears coherently again. The reflectors in Figures 4.13b, 4.14b and 4.15b become more continuous and the S/Ns of these images are largely improved. However, there are still some artifacts remaining above the sea layer as denoted by the red arrows. This might be caused by insufficient data in the training set. In this paper, we only used 5% of the image points in a dip-angle ADCIG for training. One possible solution for removing these artifacts is to increase the amount of data in the training set. Another possible solution to introduce a new feature, such as the depth feature, so the SVM algorithm can detect the possible depth range of extant reflectors. As the Marmousi model is geologically complicated, we set the weighting coefficient of the artifact points to 0.1 instead of 0 to avoid the discontinuity on the image caused by misclassification.

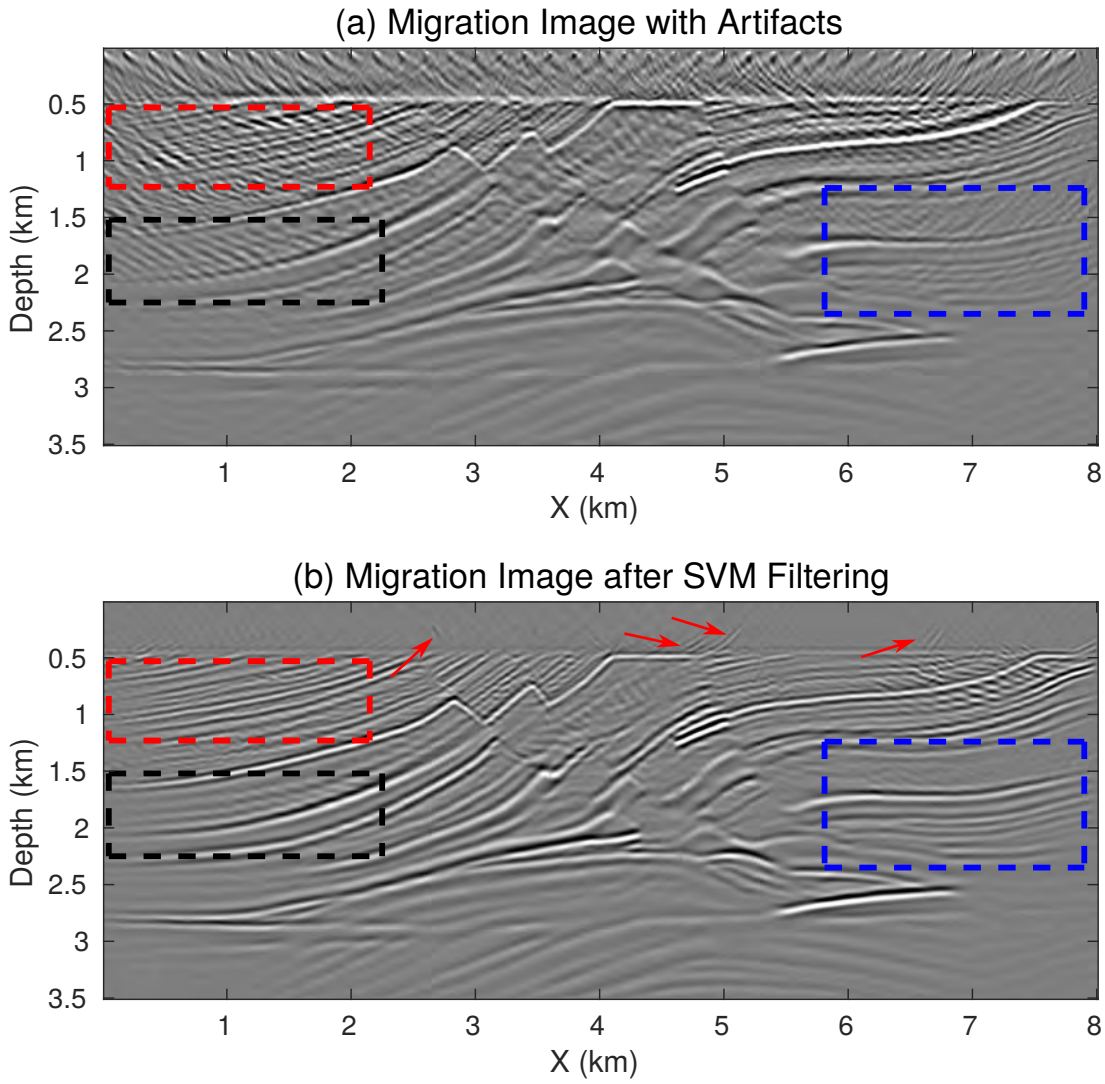


Figure 4.12: Migration images (a) with artifacts and (b) after SVM filtering.

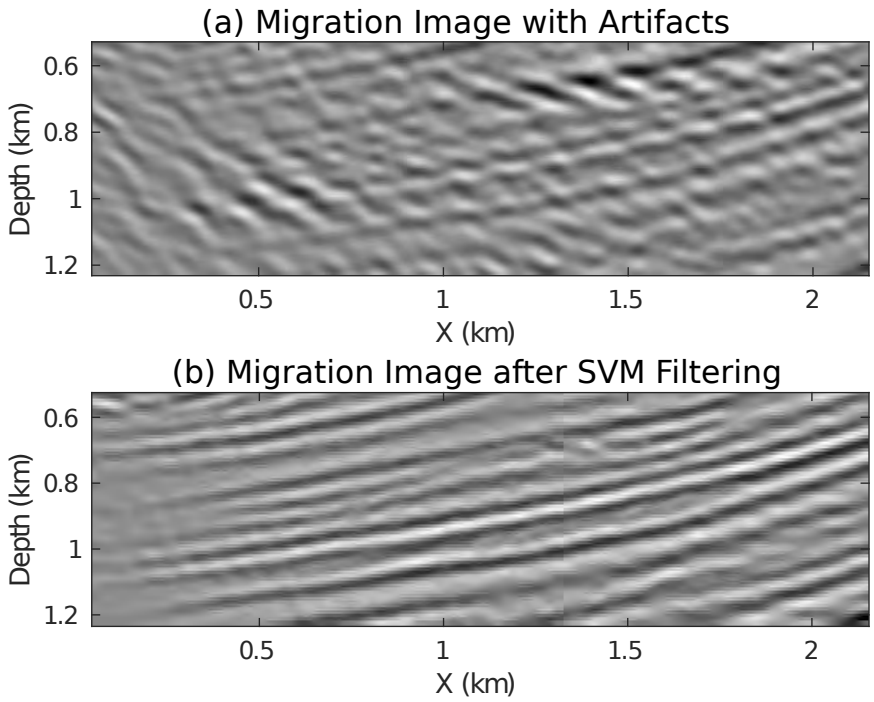


Figure 4.13: The zoom views of the red box in Figure 4.12 (a) with artifacts and (b) after SVM filtering.

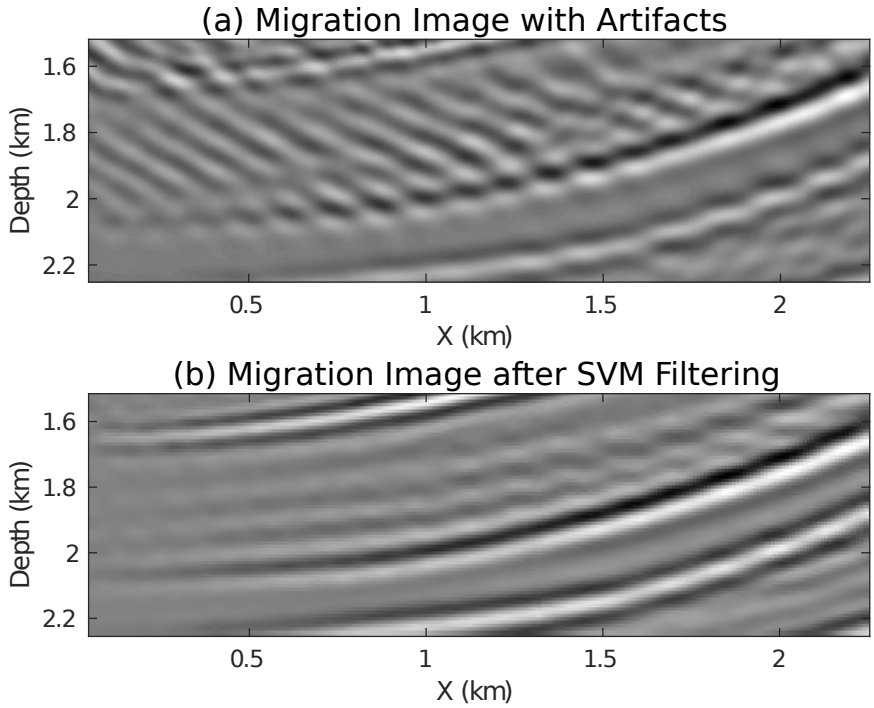


Figure 4.14: The zoom views of the black box in Figure 4.12 (a) with artifacts and (b) after SVM filtering.

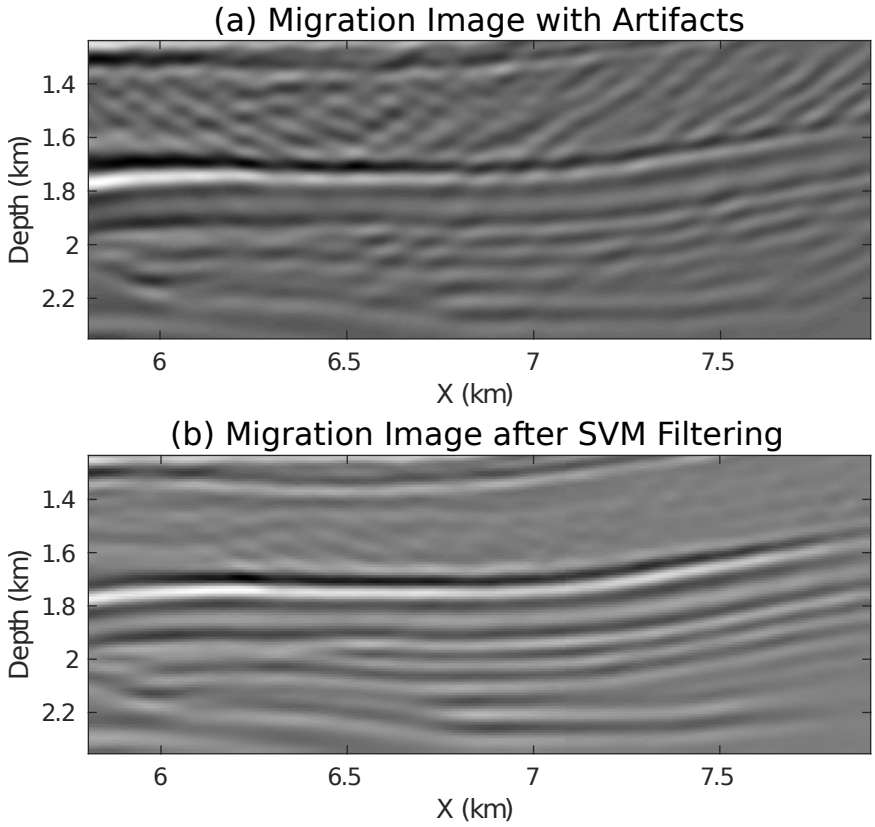


Figure 4.15: The zoom views of the blue box in Figure 4.12 (a) with artifacts and (b) after SVM filtering.

4.3.2 Volve OBC Data

We test the effectiveness of SVM filtering using a 2D slice of inline shot gathers extracted from the 3D Volve OBC data (Fu et al., 2017). The 3D dataset is acquired using 12 parallel cables with 400 m separation between the cables, and each cable contains 240 receivers. The 2D dataset is extracted from line 6 which is in the middle of the survey. The recorded data consist of 240 common-shot gathers with a shot interval of around 50 m, and the number of receivers for each shot ranges from 77 to 237 with a receiver interval of 50 m. The data sampling interval is 4 ms, and the total recording time of 7.504 s. During processing, the data are low-pass filtered to 14 Hz and resampled to 1 ms. The traces are muted after 4.5 s as there are no reflections after 4.5 s. Figure 4.16 shows a comparison between the raw and processed CSG.

The migration velocity in Figure 4.17 is estimated using a multiscale reflection phase inversion method.

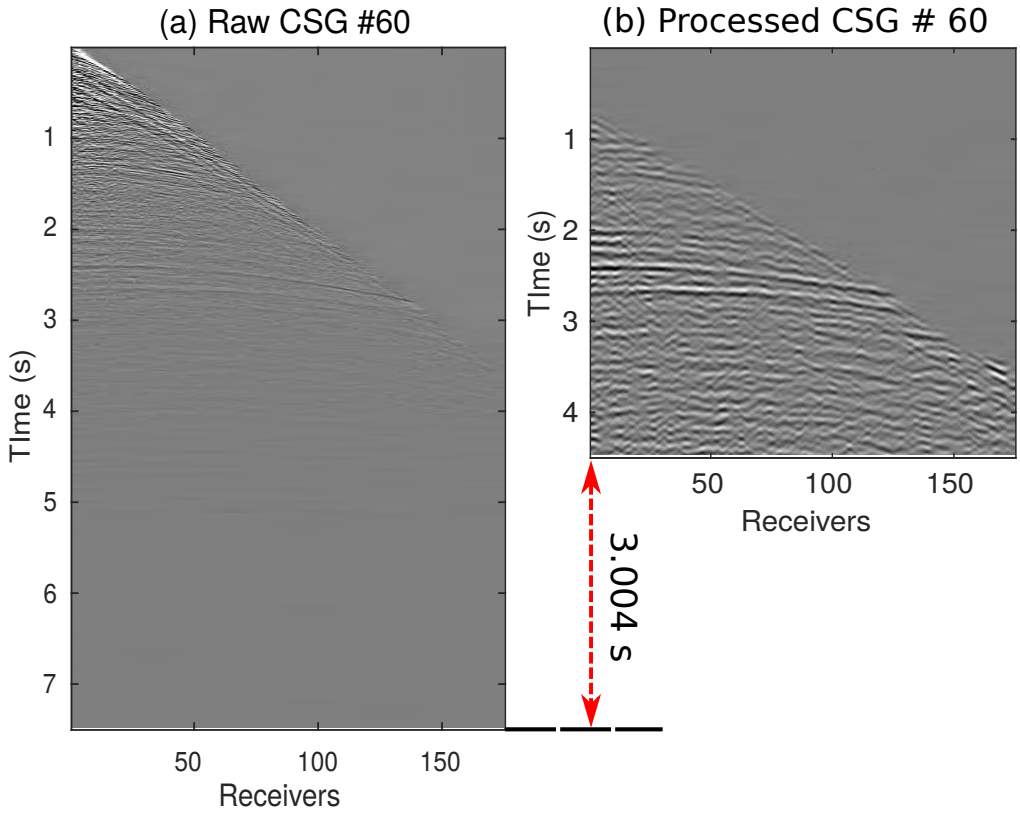


Figure 4.16: CSG gather for the (a) raw data and (b) processed data.

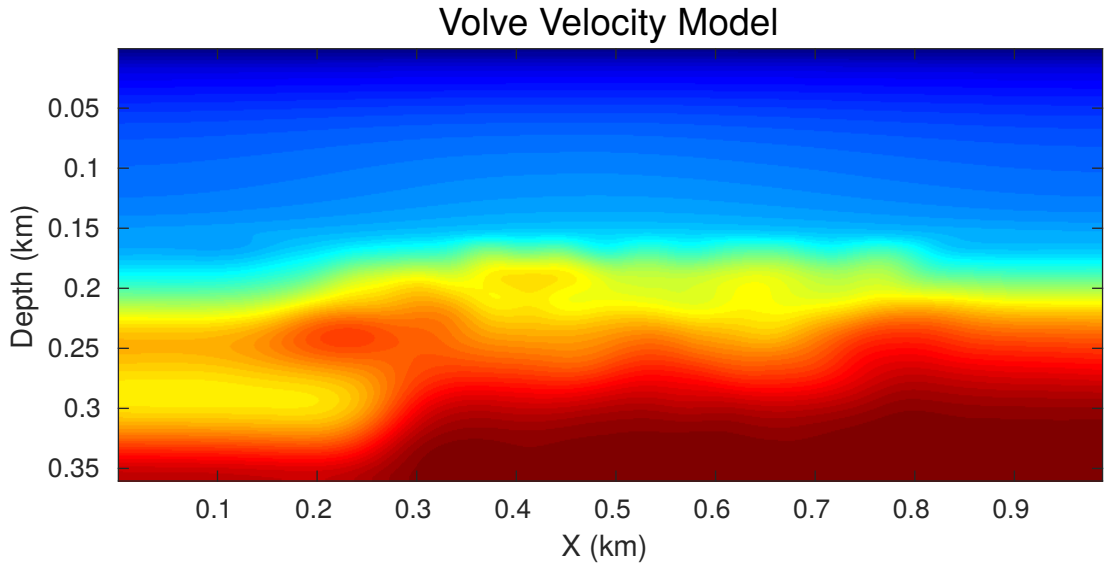


Figure 4.17: The Volve velocity model.

Figure 4.18 shows the conventional RTM image with severe artifacts, particularly at the shallow region above $z = 1.5$ km. These artifacts result from the sparse geometry and the poor quality of the seismic data. This model has a layered structure with a small dip-angle variation along the horizontal direction. Therefore we choose only the dip-angle ADCIGs at $x = 3.42, 4.59, 6.22$ km for the training set to compute the weighting coefficients for the dip-angle ADCIGs at the ranges of $0 \sim 3.79$ km, $3.79 \sim 4.95$ km and $4.95 \sim 9.88$ km, respectively. To avoid the discontinuity on the SVM filtered image that caused by misclassification, we set the weighting coefficient to be 0.1 for the artifact points instead of 0. The migration image after SVM filtering is shown in Figure 4.18b, where a noticeable improvement in the image quality, particularly in the shallow layers when compared with the conventional RTM image. The reflectors disrupted by the migration artifacts now become more continuous and the image S/N ratio is also improved after SVM filtering. Magnified views of these images are compared in Figures 4.19, 4.20 and 4.21, where the red arrows point to the areas with noticeable improvement. It is apparent that the SVM filtered image contains well defined geological structures.

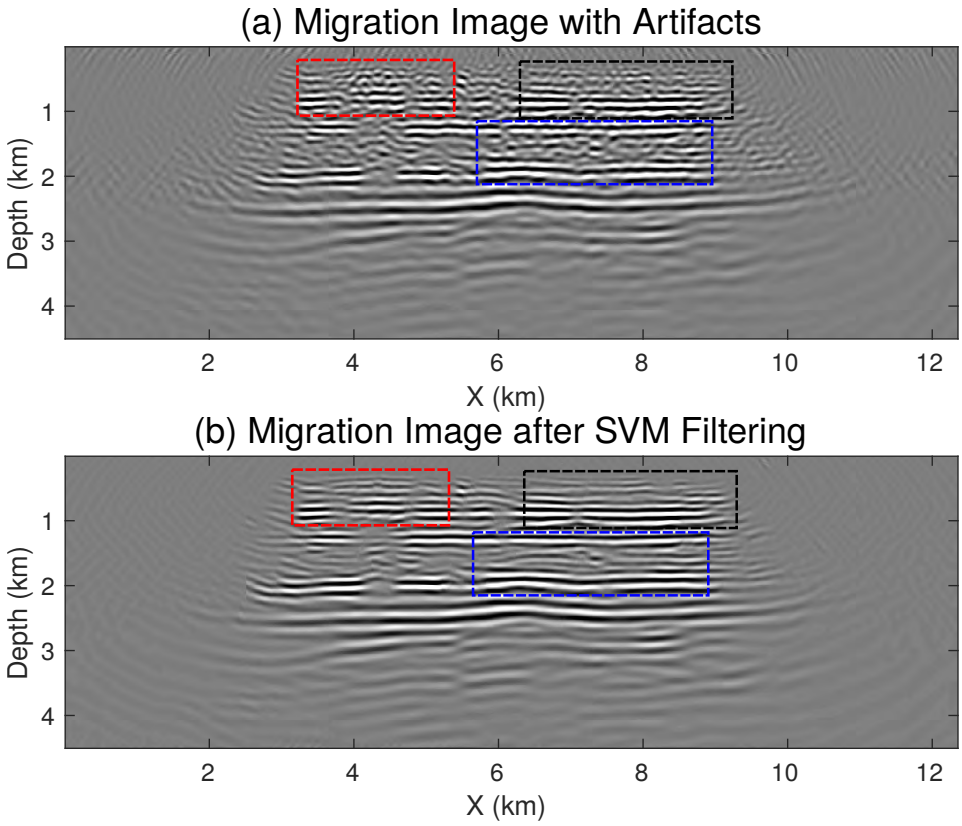


Figure 4.18: Migration image (a) with artifacts and (b) after SVM filtering.

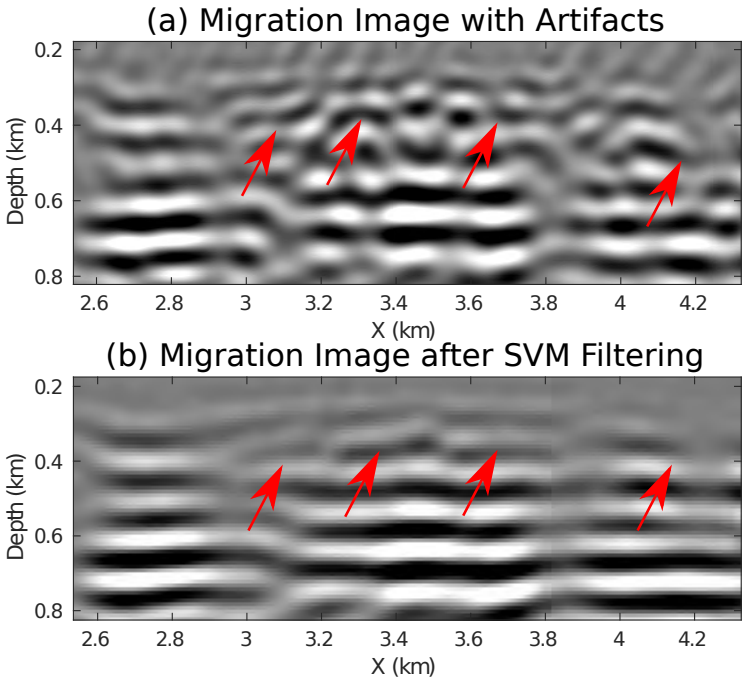


Figure 4.19: The zoom views of the red box in Figure 4.18 (a) with artifacts and (b) after SVM filtering.

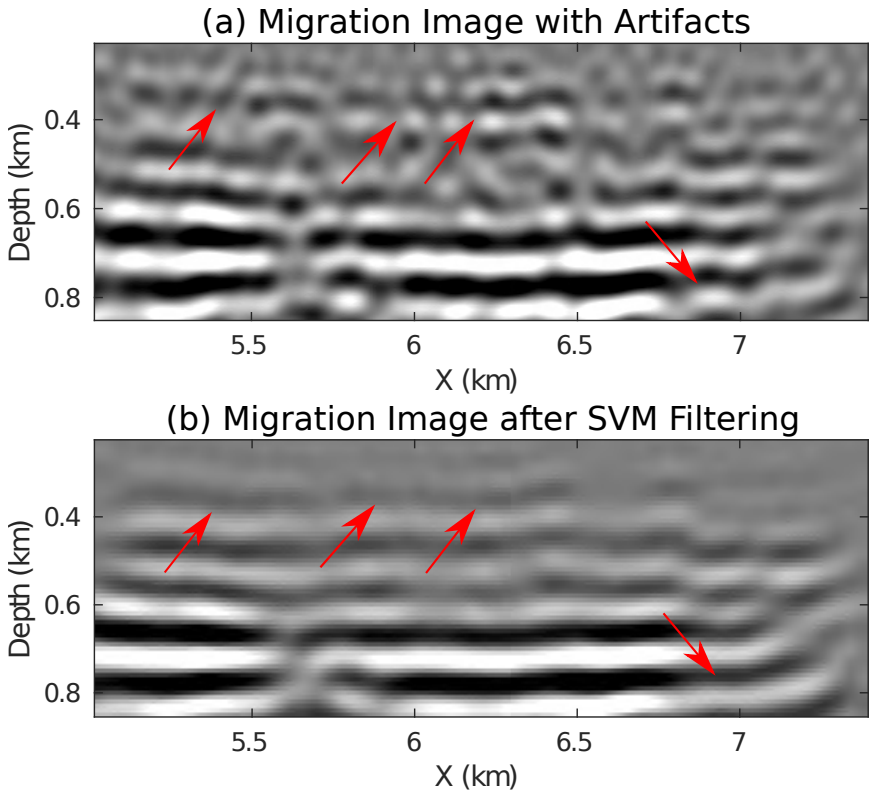


Figure 4.20: The zoom views of the black box in Figure 4.18 (a) with artifacts and (b) after SVM filtering.

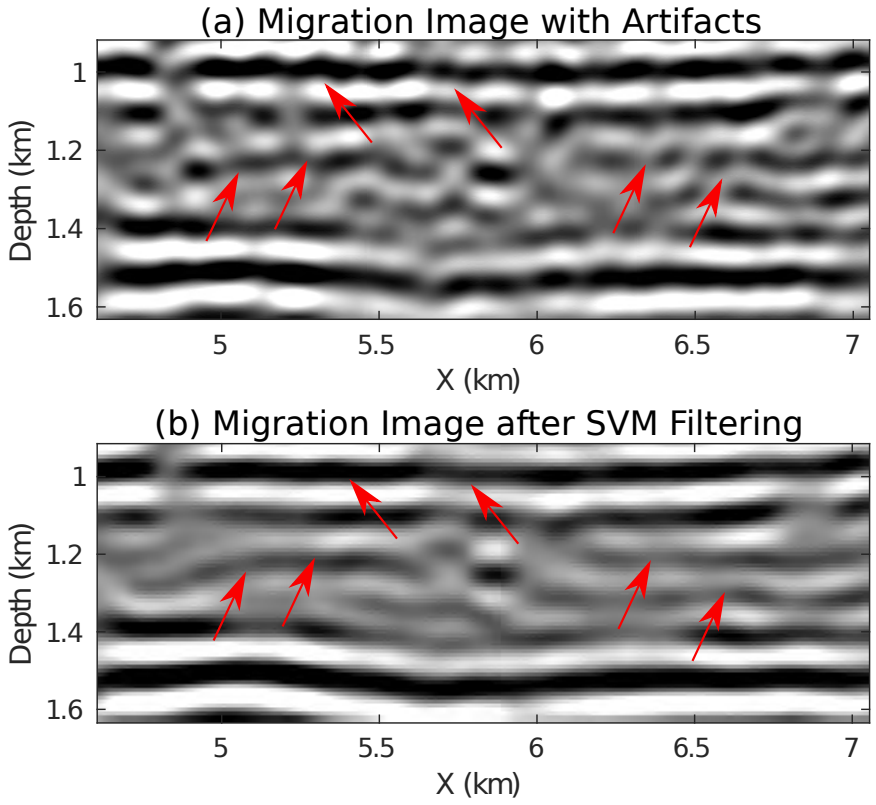


Figure 4.21: The zoom views of the blue box in Figure 4.18 (a) with artifacts and (b) after SVM filtering.

4.4 Conclusions

We have developed a novel support vector machine-based filtering method for suppressing migration artifacts resulting from aliased/sparse data. The SVM filtering method employs the features of coherency, amplitude, and dipping angle from selected dip-angle angle-domain common-image gathers (ADCIGs) to distinguish the signal points from artifact points. For each migration image, only a few dip-angle ADCIGs are needed for training and the trained model is used to compute the SVM weighting coefficients for all dip-angle ADCIGs. The method uses only 1% $\sim$ 10% of the points in the selected dip-angle ADCIGs for training. We have tested the effectiveness of the SVM filtering using both synthetic and field seismic data. In the numerical examples, fewer than 1% of the total dip-angle ADCIGs are used for

training. The additional computational cost of SVM filtering is trivial compared with some other preconditioned LSRTM methods.

Chapter 5

Multiscale Reflection Phase Inversion with Deblurring Filter

Reflection full waveform inversion (RFWI) can recover the low-wavenumber components of the velocity model along with the reflection wavepaths. However, this requires an expensive least-square reverse time migration (LSRTM) to construct the perturbation image and, even so, RFWI can still suffer from cycle-skipping problems. As an inexpensive alternative to LSRTM, we propose the use of deblurring filter (DF) with RFWI. To mitigate cycle-skipping problems, we introduce a multiscale reflection phase inversion (MRPI) strategy which boosts the low-frequency data and only needs to explain the phase information in the recorded data, not its amplitude spectrum. To mitigate cycle-skipping problems, we use the rolling-offset strategy which gradually extends the offset range of data with an increasing number of iterations. Numerical results show that the MRPI + DF method can efficiently recover the low-wavenumber components of the velocity model and is less prone to getting stuck in local minima compared to conventional RFWI.

5.1 Introduction

Full waveform inversion (FWI) has been shown to accurately invert seismic data for high-resolution estimates of the Earth's velocity distribution (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). However, the success of FWI heavily relies on the availability of either low-frequency or long-offset diving waves and refractions in the seismic data (Xu et al., 2012; Wang et al., 2013). With a narrow-aperture acquisition

geometry, FWI can only reconstruct the high-wavenumber components of the deep-velocity model illuminated by the deep reflections. Deep reflections recorded over a narrow aperture are unable to recover the long-wavelength components of the velocity model (Wang et al., 2013). To remedy this problem, Xu et al. (2012) developed a reflection full waveform inversion (RFWI) method that aims to retrieve the low-wavenumber components of the velocity model. RFWI splits the velocity model into a background (low-wavenumber component) and a perturbation model (high-wavenumber component). These two components are alternately updated at each iteration. For example, RFWI first computes the perturbed velocity model using the current background model. The calculated migration image is then used for Born modeling of the reflection data. The residuals between the Born-modeled traces and the observed traces are computed, which are then backpropagated to generate the source- and receiver-side reflection wavepaths for updating the background velocity model (Xu et al., 2012; Wang et al., 2013; Brossier et al., 2015).

Conventional RFWI uses the waveform-difference misfit function which is more likely to fall into local minima for poor starting models. This is because the misfit function is highly nonlinear and is very sensitive to the amplitude mismatch between the synthetic and the observed data. To alleviate these problems, Chi et al. (2015) developed a correlation-based misfit function which measures the kinematic differences between the observed and synthetic reflection events. Similarly, Feng and Schuster (2019) used the traveltime-difference misfit function to invert for both velocity and anisotropic parameters. Guo and Alkhalifah (2017) incorporated the elastic properties of RFWI to invert for both the P- and S-wave velocities. Wang et al. (2017) combined the elastic properties and the traveltime-based misfit function together to update both P- and S-wave velocities. Bozdağ et al. (2011); Wu et al. (2014); Chen et al. (2018) introduced a reflection multiscale envelope inversion method so that the signal's envelope carries ultra-low frequency information missing in the original sig-

nal. Furthermore, a set of methods have been developed to exclude the cycle-skipped events in different domains (Bunks et al., 1995b; Asnaashari et al., 2012; Bi and Lin, 2014; AlTheyab and Schuster, 2015).

In this chapter, we employ a multiscale reflection phase inversion (MRPI) strategy to mitigate cycle-skipping problems in conventional RFWI. We temporally integrate the seismic traces several times to both attenuate its high-frequency components and boost its low-frequency information. These low-boosted seismograms tend to mitigate cycle-skipping problems so that the solution is less prone to getting stuck in a local-minimum. We then replace the amplitude spectrum of the synthetic data with the amplitude spectrum of the observed data. This modification relieves the inversion procedure of having to explain the magnitude spectrum in the observed data. Therefore, the misfit function focuses more on the phase mismatch than on the amplitude mismatch (Sun and Schuster, 1993; Fu et al., 2017). This phase mismatch is related to the kinematic errors associated with the low-wavenumber components of the velocity model.

The problem with multiscale phase inversion described above is that the far-offset reflection traces can still suffer from cycle-skipping problems. To mitigate this problem, we also apply the rolling-offset strategy of AlTheyab and Schuster (2015). In the rolling-offset strategy, we only use the near-offset traces at the early iterations which are not cycle-skipped, and later iterations gradually include larger offset traces into the migration and inversion. We also gradually reduce the order of the temporal integration to introduce higher-frequency information into the inverted model. As will be seen with the numerical examples, these strategies increase the effectiveness and robustness of reflection inversion.

Moreover, conventional RFWI requires a least-squares migration image with balanced amplitudes and high-resolution interfaces. However, LSRTM is computationally expensive because it requires a large number of iterations to get the desired uplift

of image quality. To alleviate this expense, we use deblurring filter (DF) introduced in chapter 2 as an inexpensive alternative of LSRTM, which can produce a migration image with image quality comparable to LSRTM without iterations.

This chapter is organized into the following sections. After the introduction, we show how to incorporate DF with multiscale reflection phase inversion. We then present the numerical results for both synthetic and OBC data from the North Sea. The field data presents a challenge for estimating the low-wavenumber parts of the velocity model. The last section provides a summary and conclusions. The appendix provides a detailed derivation of the imaging condition of the multiscale reflection phase inversion using the adjoint-state method.

5.2 Theory

Reflection full waveform inversion (RFWI) splits the subsurface velocity model v into a background model (low-wavenumber component) v_0 and a perturbation model (high-wavenumber component) δv (Xu et al., 2012)

$$v = v_0 + \delta v. \tag{5.1}$$

At each iteration, a migration image for δv is first calculated under the current background model v_0. The estimated migration image δv is then used for Born modeling of reflection data. RFWI computes the data residual by subtraction of the Born data and recorded traces. The resulting residual traces are then backpropagated to generate the source- and receiver-side reflection wavepaths. These wavepaths are used to update the low-wavenumber components of the velocity model v_0, while v_0 and δv are alternatively updated as the iterations proceed.

5.2.1 Deblurring Filter

A migration image with balanced amplitudes and sharp interfaces is essential for generating strong reflection wavepaths. We use Born modeling to generate 100 shot gathers based on the true reflectivity and velocity model shown in Figure 5.1a and 5.1b, respectively. Figure 5.1c shows a reverse time migration (RTM) image with an illumination problem, in which the reflector amplitudes in the central region are much higher than in the side region. Figure 5.1e shows the least-squares reverse time migration (LSRTM) image at the 6th iteration, which has a more balanced amplitudes and a sharper interface compared to the RTM image. The RFWI gradients computed from the RTM and LSRTM images are shown in Figures 5.1d and 5.1f, respectively. It is obvious that the RFWI gradient generated from the LSRTM image is more suited to update the low-wavenumber parts of the velocity model, where the true velocity model is shown in Figure 5.1b with a small Gaussian perturbation in the middle. Despite its accuracy in reconstructing the correct velocity model, LSRTM is computationally expensive because it requires many iterations to get the desired uplift in image quality. However, the problem with LSRTM is that it is more than an order-of-magnitude more expensive than standard RTM. To reduce this cost, DF can provide image quality similar to LSRTM (see Figure 5.1g) but at the cost of about 1 RTM. As an example, Figure 5.1h shows the accurate recovery of the central velocity anomaly using DF with RFWI. Therefore, we use the DF as the deblurring operator to compute the migration image for RFWI.

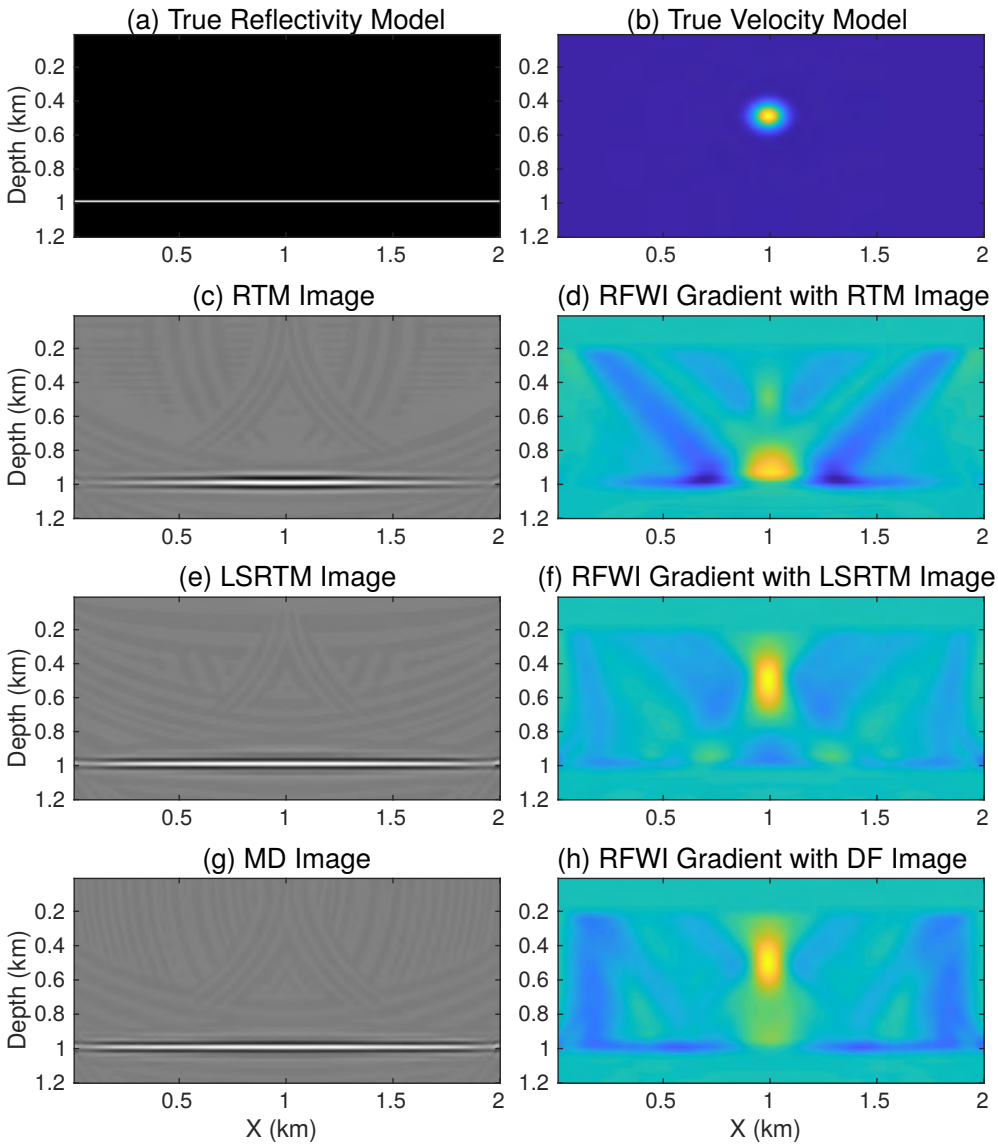


Figure 5.1: The (a) true reflectivity model, (b) true velocity model, (c) reverse time migration, (e) least-squares reverse time migration and (g) deblurred image. The RFWI gradients based on the (d) RTM, (f) LSRTM, and (h) DF images.

5.2.2 Amplitude Replacement

Currently, acoustic waveform inversion is the most widely used inversion method in the industry due to its low-computation costs. However, the acoustic assumption neglects elastic, attenuation, and anisotropic effects, which can severely compromise the inversion results. In some cases, a matching filter can be used to reduce the

non-acoustic effects of the seismic data (Agudo et al., 2018), but the matching filters need to be iteratively updated and can be prone to errors.

Sun and Schuster (1993) proposed amplitude replacement as a one-step method for compensating for some of the non-acoustic effects in the data. To eliminate the need to explain the magnitude spectrum, the amplitude spectra of the synthetic traces are replaced by those from the corresponding recorded traces. For any point source at $\mathbf{x}_s$ and a receiver at $\mathbf{x}_r$, both the observed traces $d_{obs}(\mathbf{x}_r, t|\mathbf{x}_s)$ and the synthetic traces $d(\mathbf{x}_r, t|\mathbf{x}_s)$ are Fourier transformed into the frequency domain to obtain

$$\dot{D}_{obs}(\mathbf{x}_r|\mathbf{x}_s) = \mathscr{F}(d_{obs}(\mathbf{x}_r, t|\mathbf{x}_s)) = A_{obs}(\mathbf{x}_r, \omega; \mathbf{x}_s) exp[i\theta_{obs}(\mathbf{x}_r, \omega; \mathbf{x}_s)], \tag{5.2}$$

$$\dot{D}(\mathbf{x}_r|\mathbf{x}_s) = \mathscr{F}(d(\mathbf{x}_r, t|\mathbf{x}_s)) = A(\mathbf{x}_r, \omega; \mathbf{x}_s) exp[i\theta(\mathbf{x}_r, \omega; \mathbf{x}_s)], \tag{5.3}$$

where A and θ represent the magnitude and phase spectra, respectively, and F( ) denotes the Fourier transform operator in the time domain. We replace the magnitude spectrum of the calculated data A with that of the observed data A_{obs} and then apply the inverse Fourier transform to get the modified data:

$$\tilde{d}(\mathbf{x}_r, t|\mathbf{x}_s) = \mathscr{F}^{-1}\{\tilde{D}(\mathbf{x}_r|\mathbf{x}_s)\} = \mathscr{F}^{-1}\{A_{obs}(\mathbf{x}_r, \omega; \mathbf{x}_s) exp[i\theta(\mathbf{x}_r, \omega; \mathbf{x}_s)]\}. \tag{5.4}$$

Therefore, there is only the phase mismatch between the observed and modified synthetic data. This phase mismatch is mostly related to the kinematic errors controlled by the low-wavenumber components of the velocity model.

As an example, Figures 5.2a, 5.2b, and 5.2c show the acoustic, viscoacoustic, and elastic data, respectively. It clearly shows that the acoustic data do not contain the amplitude and phase distortions seen in the viscoacoustic data or the converted waves in the elastic data. Figures 5.2e and 5.2f show the acoustic traces where the

amplitude spectra are replaced by the viscoacoustic and elastic data, respectively. The data residuals for the acoustic and viscoacoustic data with and without the amplitude replacement are compared in Figures 5.3a and 5.3b, respectively. They clearly show that the subtraction errors due to the viscoacoustic effects in the seismic data are largely reduced in Figure 5.3b. Similarly, the errors in the data residuals for the acoustic and elastic data are also greatly reduced after amplitude replacement when compared with Figure 5.3d with Figure 5.3c.

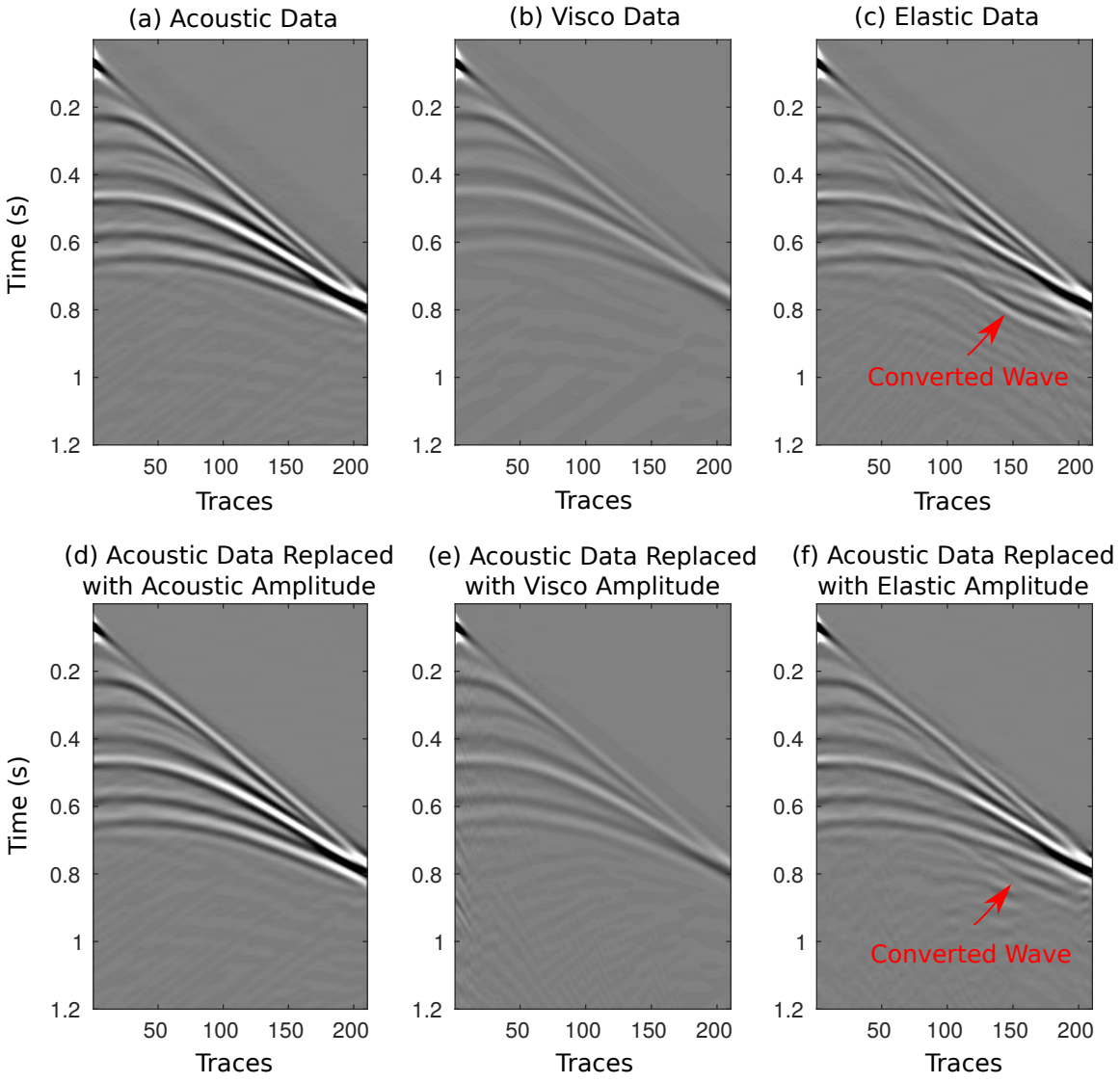


Figure 5.2: The (a) acoustic, (b) viscoacoustic, (c) elastic traces, where the acoustic magnitude spectra are, respectively, replaced with the amplitude spectra of (d) acoustic data, (e) viscoacoustic data, and (f) elastic data.

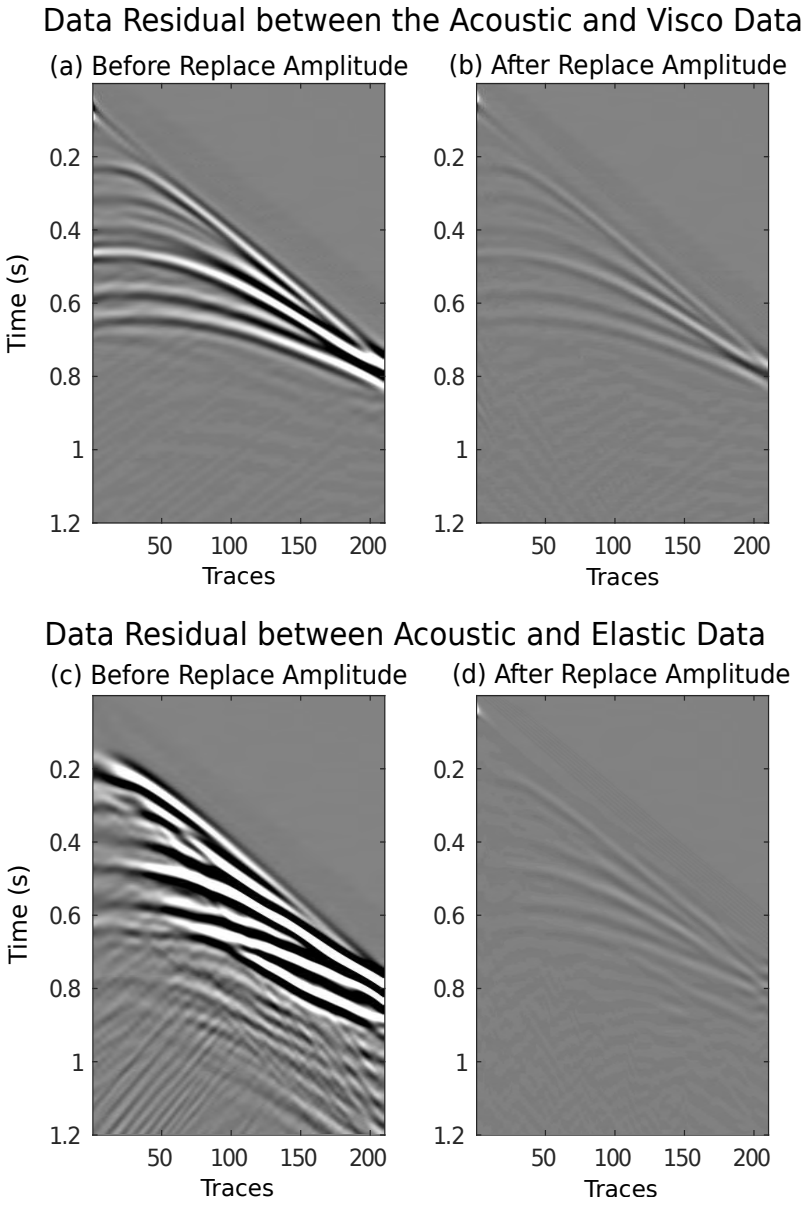


Figure 5.3: The data residual between the acoustic data and visco data (a) before and (b) after amplitude spectra replacement. The data residual between the acoustic data and elastic data (c) before and (d) after amplitude spectra replacement.

5.2.3 Time Integration

Time integration of a signal can be viewed as applying a low-boost filter to this signal (Tavares, 1966). Integrating a signal in the time domain is equivalent to dividing the

spectrum by $i2\pi f$ in the frequency domain, and so boosts the low-frequency components and suppresses the high-frequency components of a signal. Figures 5.4a, 5.4b and 5.4c show the traces with zero, one and two integrations, respectively, and Figure 5.5 shows their corresponding amplitude spectra. As the number of integrations increases, the high-frequency components are gradually attenuated and the low-frequency information becomes dominant. Therefore, the recorded traces are integrated several times to boost their low-frequency components so that the integrated residual traces are less prone to be cycle-skipped (Fu et al., 2017). This is very similar to the multiscale inversion approach using bandpass filters (Bunks et al., 1995b), except integration is in the time-domain method. However, the problem with a bandpass filter is that the bandpassed signal can be polluted by artifacts generated by the narrow bandwidth and steep boundary of the bandpass filter. In comparison, the time integration operator does not have such effects. One disadvantage of the time integration, however, is that it might also boost the direct current (DC) term of the signal. But this problem can be eliminated by demeaning the data. In practice, we only integrate the source wavelet instead of integrating the synthetic data (see Appendix D).

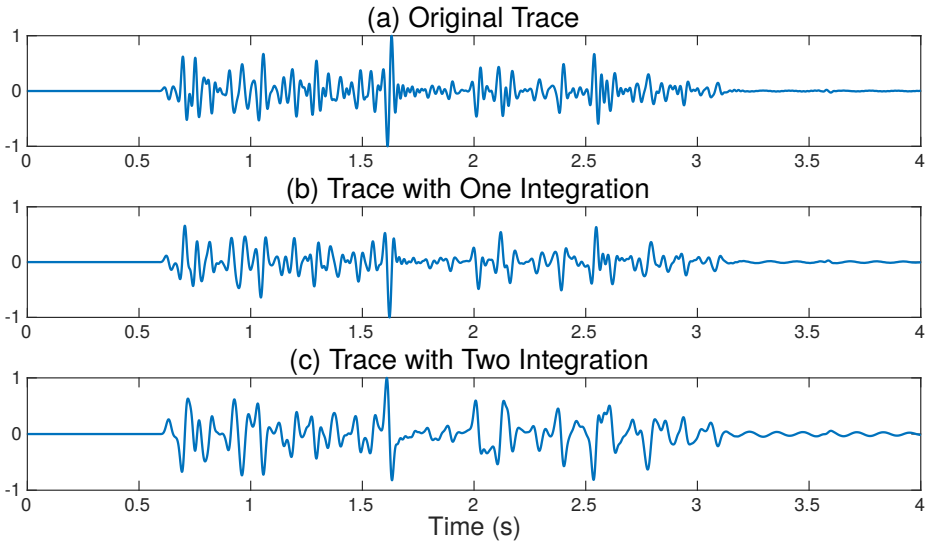


Figure 5.4: The original trace (a) without integration, (b) with one and (c) two integration.

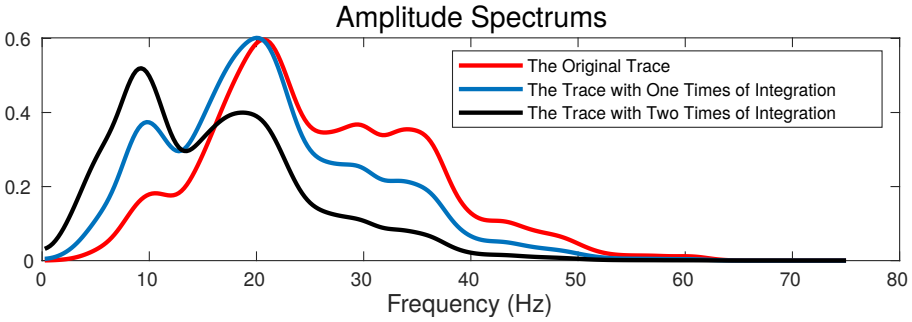


Figure 5.5: The frequency spectrum of the original trace (a) without integration, with (b) one and (c) two integrations, which are represented by the red, blue and black curves, respectively.

5.2.4 Rolling Offset

For a poor starting model, simultaneous inversion of the full-offset range of data can lead to two problems: (1) the far-offset data are more likely to be cycle-skipped compared with the near-offset data because of the longer wavepaths. (2) At the near-offset region, the amplitude and phase of the Born data do not agree with the observed data. The reason is that the stacked traces in the migration image becomes increasingly out-of-phase with larger offsets. This is because an incorrect migration

velocity will adjust the reflector depth so that the near-offset synthetic traces match those from the field data, but this will not happen with increasing source-receiver offsets. Therefore, the inversion strategy is to first adjust the velocity model to first explain the near-offset data and then gradually adjust the velocity model to explain the traces with increasing offset. In this way, we can mitigate getting stuck in a local minimum and, hopefully, converge to the global minimum.

To numerically demonstrate the importance and advantages of the rolling-offset strategy, we test it on the two-layer model shown in Figure 5.6, where the upper- and bottom-layer velocities are 1800 m/s and 2300 m/s respectively. The interface of this two-layer model is at a depth of 0.5 km. Shot gathers are compared for 100 shots spaced at 20 m intervals. Each shot is recorded using 200 receivers separated at an interval of 10 m on the surface and the migration velocity model is homogeneous with the velocity of 1400 m/s. Figures 5.7a to 5.7e show that the traces with different offset ranges are migrated to different depths, and the reflector images shift to shallower depths with increasing source-receiver offsets. Figure 5.7f shows the migration image using the full-offset data, which is quite similar to the migration image using only the near-offset data shown in Figure 5.7a. This similarity can be explained by the stationary phase theory (Schuster, 2009), except there are still significant artifacts due to incomplete cancellation (see Figure 5.7f).

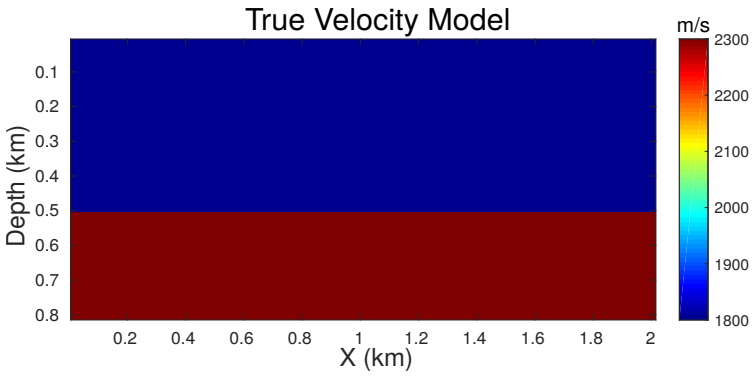


Figure 5.6: Two-layer velocity model.

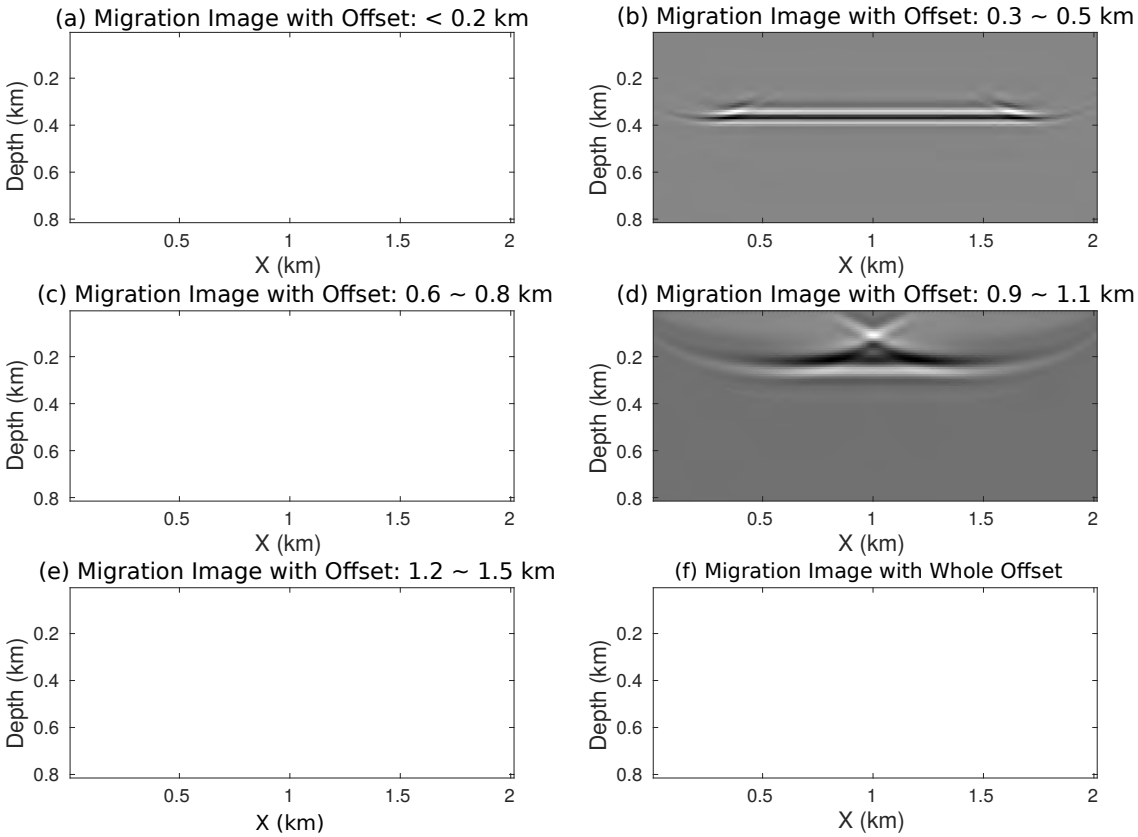


Figure 5.7: Migration image with offset ranges of (a) 0.0 to 0.2 km, (b) 0.3 to 0.5 km, (c) 0.6 to 0.8 km, (d) 0.9 to 1.1 km, (e) 1.2 to 1.5 km and (f) the entire offset.

Figure 5.8a shows the comparison between the observed traces (red wiggles) and the Born data (green wiggles) which are computed from the migration velocity model and the migration image in Figure 5.7f. Two phenomena can be noticed in Figures 5.8a: (1) The far-offset data suffers severely from cycle-skipping and (2) the seismic events in the Born modeled data arrive earlier in time at the near-offset region compared to the observed data. However, the Born modeled events gradually arrive later in time as the offset increases. These phase mismatches can be seen more clearly in Figure 5.8b, which are the zoomed views of the blue boxes in Figure 5.8a. Moreover, the amplitudes of the Born reflections are much weaker compared to the observed data as the out-of-phase summation severely weakens the amplitudes of the migration image. This phenomenon can be observed in Figure 5.8c which shows the data

residuals of Figure 5.8a. It clearly shows that the data residual is dominated by the energy of the observed data.

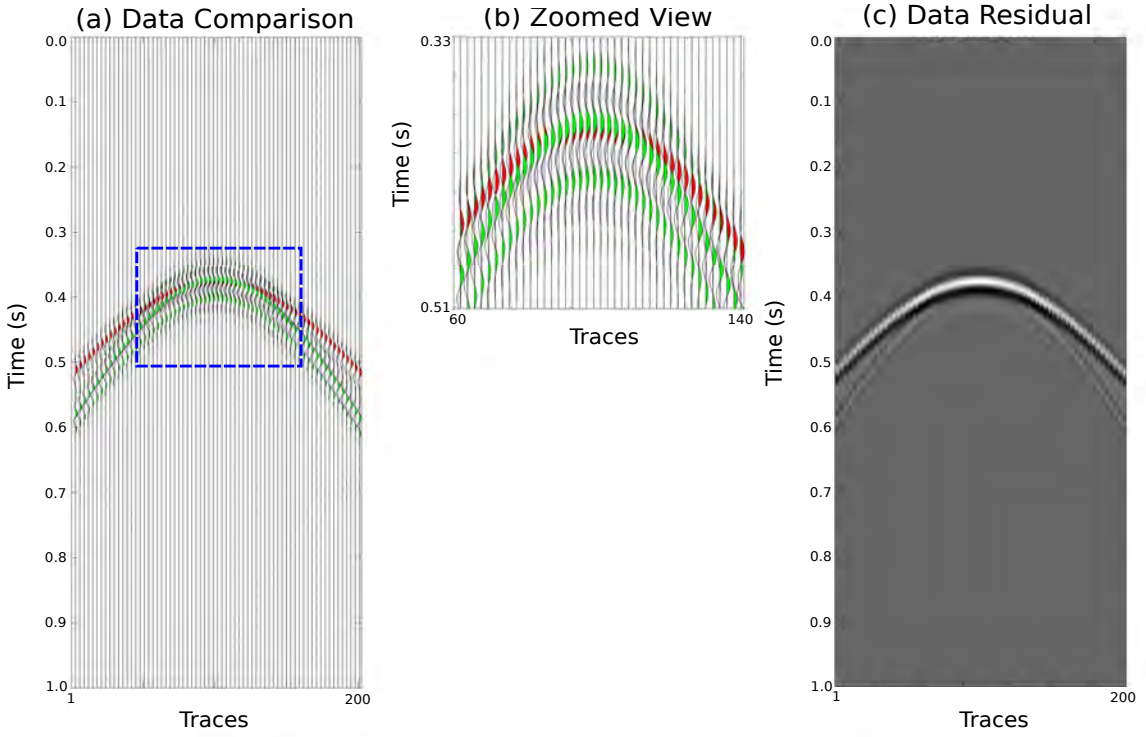


Figure 5.8: The (a) trace comparsions between the observed data and Born data at x = 1.0 km; (b) zoomed view of the blue box and (c) the data residuals.

Figure 5.7a shows the migration image computed from the near-offset data with source-receiver offsets smaller than 0.2 km. This image has fewer amplitude and phase errors compared to Figure 5.7f because the shorter wavepaths suffer less from velocity errors. The comparisons between the observed and Born data generated from Figure 5.7a are shown in Figure 5.9a. Even though there are still severe cycle-skipping problems at the far-offset region, their waveforms correlate very well at the near-offset region. This behavior is even more pronounced in the zoomed view shown in Figure 5.9b. Figure 5.9c shows the data residuals in Figure 5.9a, where the residual traces are almost negligible because the observed and recorded data mostly agree with one another in both their phases and amplitudes. The blue box indicates the offset region used to compute the migration image.

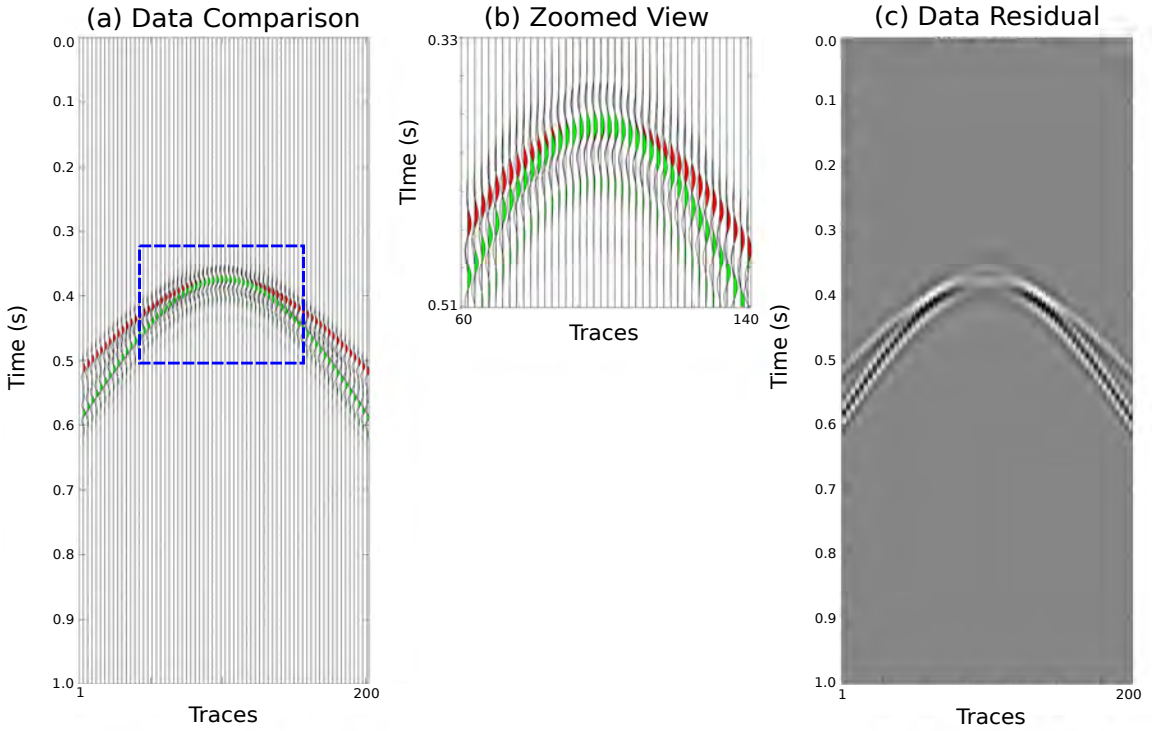


Figure 5.9: The (a) trace comparsion between the observed data and Born data at x = 1.0 km; (b) zoomed view of the blue box and (c) the data residuals.

The rolling-offset strategy is now tested on three different cases. In the first example, we use full-offset data for both migration and inversion. Therefore the RFWI gradient suffers from cycle skipping as well as amplitude and phase mismatches at both the near and far offsets. The corresponding RFWI gradient is shown in Figure 5.10a, which exhibits strong aliasing artifacts along the sides of the computational model. In the second example, we use full-offset data for migration, but we only select the uncycle-skipped data at the near-offsets for inversion. To mitigate the amplitude mismatches, we normalize both the computed Born and the observed data before subtraction. Therefore, the phase mismatch problem is the dominant one in the second case. Figure 5.10b shows the RFWI gradient for the second case, which has balanced low-wavenumber updates in the central part of the model. However, there is still aliasing along the side boundaries. In the third example, we use near-offset uncycle-skipped data for both migration and inversion. The corresponding

RFWI gradient is shown in Figure 5.10c, which generates strong and balanced low-wavenumber updates.

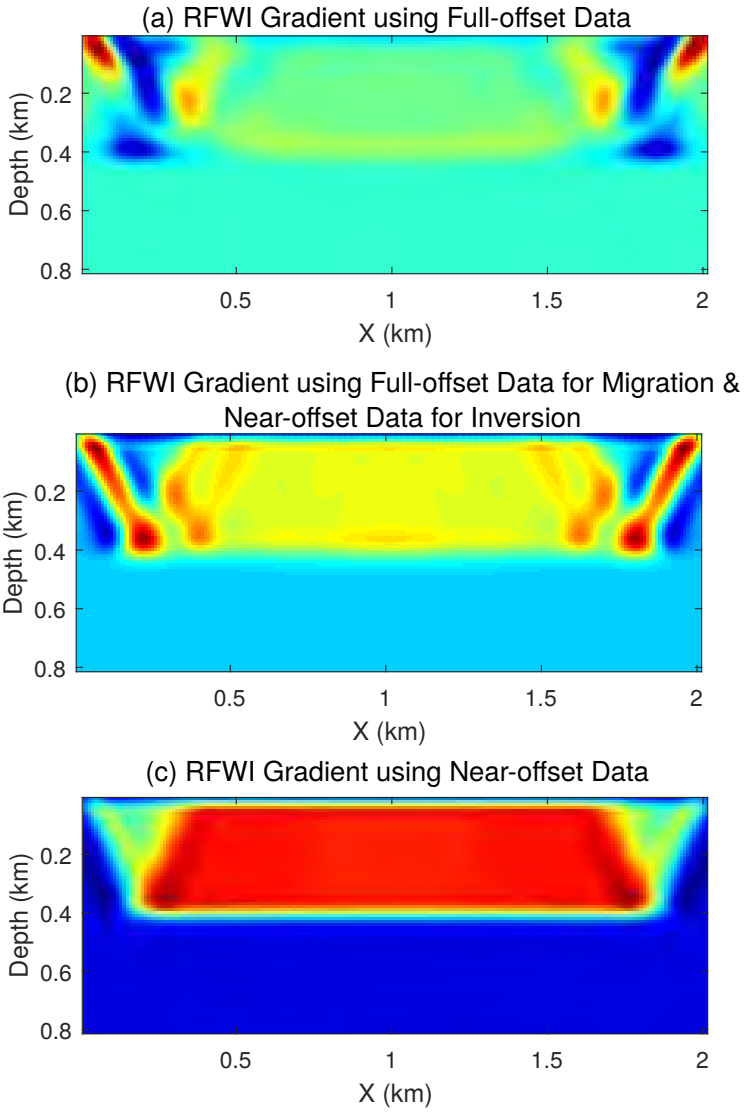


Figure 5.10: Demonstration of the influence of the cycle skipping, amplitude and phase problems associated with the RFWI gradient. (a) RFWI gradient using full-offset range of data for both migration and inversion. (b) RFWI gradient using near-offset data for both migration and inversion. (c) RFWI gradient using the full-offset data for migration and near-offset data for inversion.

5.2.5 Discussion of the Offset Selection

To remove the cycle-skipped traces, we can either calculate their traveltime or phase differences. Cycle skipping occurs when the traveltime or phase differences between the synthetic and observed traces are larger than $\frac{T}{2}$ or π, respectively (Shah et al., 2012). Here, T is the period of the seismic signal. The traveltime difference $\Delta\tau$ can be found by maximizing the crosscorrelation function $f(\tau)$ as

$$max\Big\{f(\tau)\Big\} = max\Big\{\int p(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} p(\mathbf{x}_r, t+\tau; \mathbf{x}_s)_{syn} dt\Big\}, \tag{5.5}$$

where $p(\mathbf{x}_r, t; \mathbf{x}_s)_{obs}$ represents the observed trace at the receiver location $\mathbf{x}_r$ for a shot at location $\mathbf{x}_s$. Similarly, $p(\mathbf{x}_r, t+\Delta\tau; \mathbf{x}_s)_{syn}$ indicates the synthetic trace at the same source and receiver location but with a time shift τ. This equation seeks a $\Delta\tau$ that shifts the synthetic seismogram so that it "best" matches the observed seismogram.

The phase difference $\Delta\phi$ between the synthetic and observed seismic traces can be computed by two steps. First, transform the observed and synthetic traces into the frequency domain by FFT

$$\begin{aligned} \tilde{p}_{obs} &= \mathscr{F}(p(\mathbf{x}_r, t; \mathbf{x}_s)_{obs}) = A_{obs} e^{i\phi_{obs}}, \\ \tilde{p}_{syn} &= \mathscr{F}(p(\mathbf{x}_r, t; \mathbf{x}_s)_{syn}) = A_{syn} e^{i\phi_{syn}}, \end{aligned} \tag{5.6}$$

where A_{obs} and A_{syn} represent the magnitude spectra of the observed and synthetic seismic traces, respectively. ϕ_{obs} and ϕ_{syn} indicate their corresponding phase spectrums. Second, compute their phase difference by

$$\begin{aligned} \Delta\phi &= phase\Big(\tilde{p}_{obs}\tilde{p}^*_{syn}\Big) = phase\Big(A_{obs}A_{syn}e^{i(\phi_{obs}-\phi_{syn})}\Big), \\ &= \phi_{obs} - \phi_{syn}, \end{aligned} \tag{5.7}$$

where $*$ is the conjugate symbol and *phase* indicates a phase extraction operator.

In this chapter, we prefer the phase difference approach rather than the traveltime difference approach. The reason is that the phase difference $\Delta\phi$ varies smoothly with increasing offset if the seismic traces are not cycle-skipped. However, a sudden 2π phase jump will occur at the cycle-skipped trace. This phase jump is easy to detect and can be used directly to remove the cycle-skipped traces.

Figure 5.11b shows a comparison between two shots where the nearest-offset trace is in the middle. The reflection events in the two shot gathers disagree more with increasing offset. Figure 5.11a shows the corresponding phase differences which clearly shows the phase-jump points. These phase-jump points indicate the locations where the cycle-skipping appears. We can simply remove the traces where its offset is larger than the phase jump locations to avoid the cycle-skipping problem.

The offset range associated with uncycle-skipped traces is easy to identify for a simple case where there is only one seismic event per trace. However, for complicated cases, the offset selection strategy is the following:

1. Migrate the observed data with the initial velocity model. Then manually pick the points of the reference reflector which is relatively continuous and has a good signal-to-noise ratio.

2. Generate the Born data based on the picked reflectivity model and window out the corresponding reflections in the observed data.

3. Measure the phase differences between the Born data and the windowed observe data, then identify the offset with the first phase jump point as a reference to remove the cycle-skipped seismic traces.

Theoretically, one can pick single or multiple reflectors as the reference to guide the offset selection procedure. The more the reflectors are used, the higher the accuracy can be achieved in removing cycle-skipped traces.

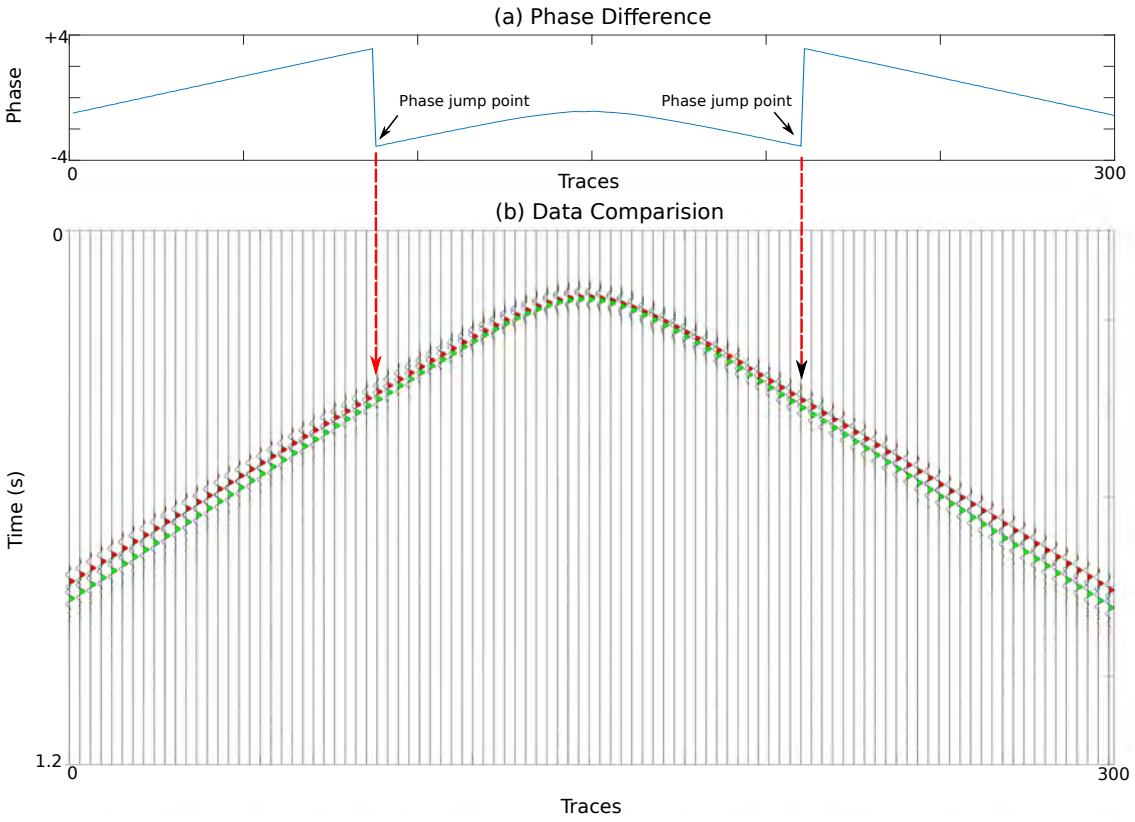


Figure 5.11: The (a) phase differences between (b) two shot gathers with trace mismatches.

5.2.6 Misfit Function

The misfit function of multiscale reflection phase inversion is defined as

$$\begin{aligned}\epsilon_i &= \sum_{s,r}\int dt\big[W_i M I^n(d(\mathbf{x}_r,t|\mathbf{x}_s)-d_{obs}(\mathbf{x}_r,t|\mathbf{x}_s))\big]^2, \\ &= \sum_{s,r}\int dt[I^n(\bar{d}_i-\bar{d}_{i\_obs})]^2\end{aligned} \tag{5.8}$$

where W_i is a weighting operator that zeros out the cycle-skipped events in the data at the ith stage. Here, M indicates the amplitude replacement procedure in equation 5.8, I^n is the integration operator $I \equiv \int dt$ and n indicates that the integration is performed n times. The observed data d_{obs} only contain the reflection energy, d represents the synthetic data generated from the Born modeling, and $\bar{d}$ and $\bar{d}_{obs}$ represent the modified traces. The gradient of the misfit function ϵ with respect to velocity v is given by

$$\begin{aligned}\frac{\partial\epsilon_i}{\partial v} &= \sum_{s,r}\int dt\Big(I^n\frac{\partial\bar{d}_i}{\partial v}\Big)^T(I^n\Delta\bar{d}_i), \\ &= -\sum_{s,r}\int_0^T 2\rho v(I^n\nabla\cdot\mathbf{v})(I^n\delta q)+2\rho v(I^n\nabla\cdot\delta\mathbf{v})(I^n q)dt,\end{aligned} \tag{5.9}$$

where $\mathbf{v}$ and $\delta\mathbf{v}$ represent the background and perturbed particle-velocity fields, respectively, and q and δq indicate the adjoint-state variables of the background and perturbed pressure wavefield p and δp, respectively. The density is represented as ρ. As an alternative, the integration operator can be applied to the source wavelet directly. Therefore, equation A-6 can be simplified as

$$\frac{\partial\epsilon_i}{\partial v} = -\sum_{s,r}\int_0^T 2\rho v(\nabla\cdot\tilde{\mathbf{v}})\delta\tilde{q}+2\rho v(\nabla\cdot\delta\tilde{\mathbf{v}})\tilde{q}dt, \tag{5.10}$$

where the tilde indicates that the wavefields are generated by an integrated source. The detailed derivation of the gradient is shown in Appendix E using the adjoint-state method.

The gradient in equation 5.10 has two terms. This first term $\int_0^T 2\rho v(\nabla \cdot \tilde{\mathbf{v}})\delta\tilde{q}dt$ corresponds to the source-side reflection wavepath, which is the dot product between the downward-propagated source-side wavefield $\nabla \cdot \tilde{\mathbf{v}}$ and the upward-propagated receiver-side Born wavefield $\delta\tilde{q}$. Similarly, the second term $\int_0^T 2\rho v(\nabla \cdot \delta\tilde{\mathbf{v}})\tilde{q}dt$ can be interpreted as the receiver-side reflection wavepath, which is the temporal dot product between the downward-propagated receiver-side wavefield $\tilde{q}$ and the upward-propagated source-side Born wavefield $\nabla \cdot \delta\tilde{\mathbf{v}}$. The background velocity is updated by smearing the energy of the data misfit along the wavepaths associated with the sources and receivers.

For a single source-receiver pair, Figure 5.12b shows the reflection wavepath of MRPI with the integration time $n = 3$, which is smoother, wider and more balanced in the spatial distribution of amplitudes compared to the reflection wavepath of conventional RFWI shown in Figure 5.12a. Given the gradient, the steepest-descent method can be used to iteratively update the background-velocity model until the data misfit is sufficiently small.

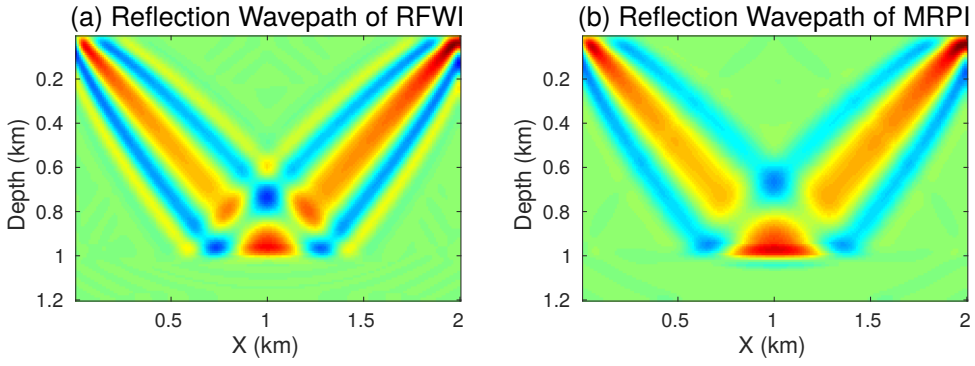


Figure 5.12: Comparison of reflection wavepaths between (a) conventional RFWI and (b) MRPI.

5.2.7 WorkFlow of MRPI

The workflows for conventional RFWI and MRPI + DF method are compared in Figure 5.13. For the conventional RFWI workflow shown in Figure 5.13a, the observed data are directly fed into the migration and inversion algorithms. In this case, the convergence of conventional RFWI can be hardly guaranteed. Figure 5.13b provides the workflow of the multiscale reflection phase inversion:

1. The observed traces are temporally integrated n times to attenuate the high-frequency information and boost the low-frequency information.

2. Select the near-offset uncycle-skipped data to compute the migration image using DF.

3. Compare the Born modeled data with the observed data, select the uncycle-skipped traces again to compute the reflection wavepaths and update the background velocity model.

4. Compute the new migration image based on the updated velocity model.

5. Repeat steps 3 - 4 until the residual falls below a specified tolerance. Reduce the integration number from n to $n-1$ and repeat steps 2 - 4 again. For all of the iterations, the synthetic magnitude spectra are replaced with the observed spectra.

This workflow mitigates the problem of getting stuck in a local minimum and is computationally less expensive than standard RFWI which uses many iterations for the LSM inner loop.

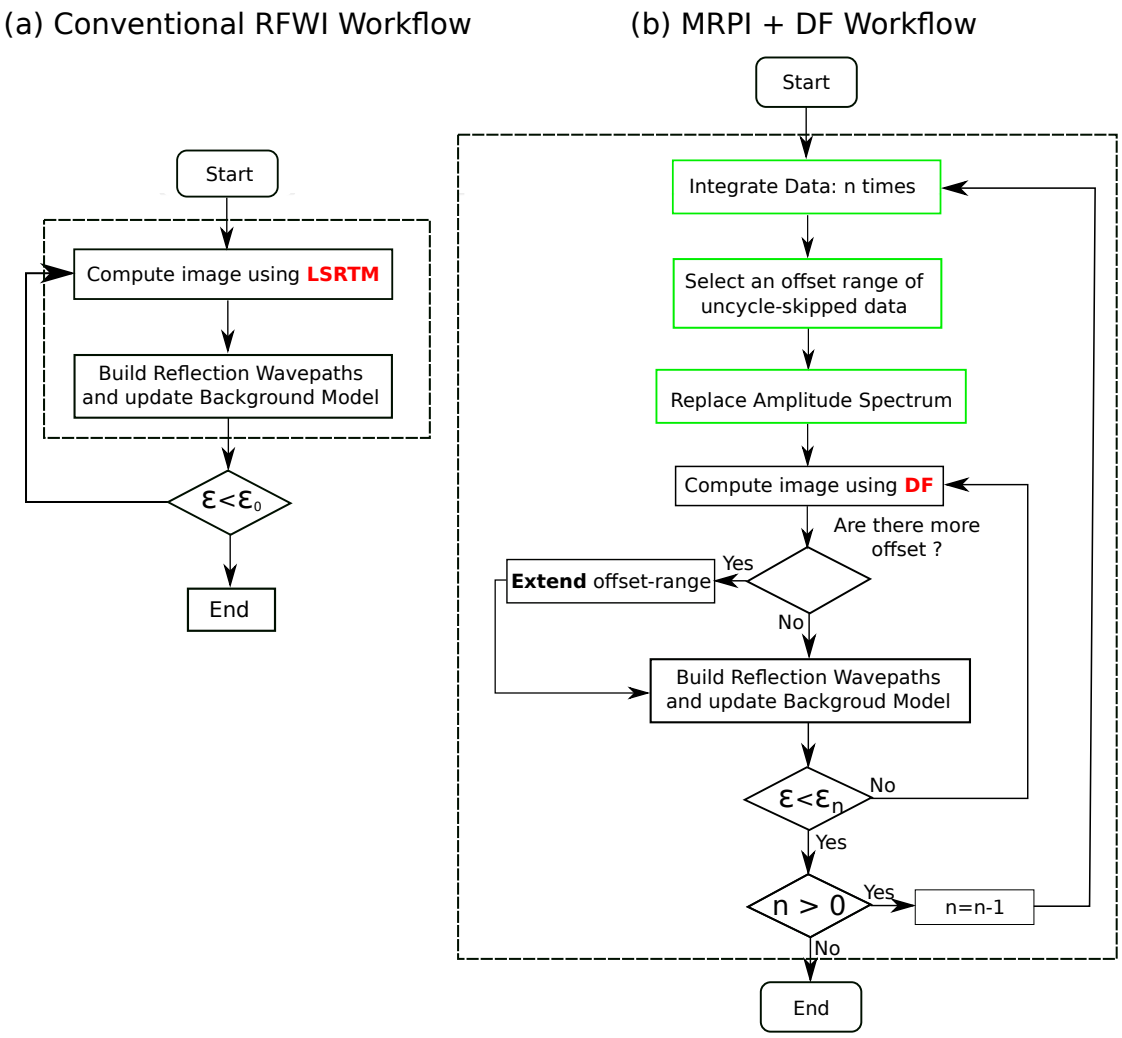


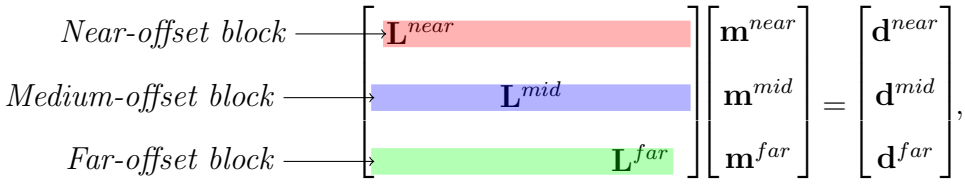
Figure 5.13: Comparison of the workflows between conventional RFWI and the MRPI + DF method.

5.2.8 Quantitative Analysis of MRPI + DF Method

The MRPI + DF method is similar to the iterative block Gauss-Seidel method for solving systems of equations. The goal of iterative block Gauss-Seidel (IBGS) is to iteratively find the solution to a large system of equations. The large matrix is broken up into smaller blocks, where there is no or very little overlap between blocks and an initial starting model is specified. The solution associated with the top block of equations is found first. This solution is then used as the starting solution for the next

block of equations. This procedure is repeated until the solution to the last block of equations is found. However, the equations for the MRPI procedure are non-linear with respect to the model parameters, so that we use a non-linear IBGS procedure where the coefficients in the matrix blocks are updated after a sufficient number of sweeps through the entire matrix.

The equations for MRPI + DF can be arranged from top to bottom in blocks of traces with different offset ranges (for simplicity, we use the linearized equations for demonstration and also ignore the integration times)



(5.11)

where $\mathbf{L}^{near}$, $\mathbf{L}^{mid}$ and $\mathbf{L}^{far}$ represent the block of modeling operators, each with a different offset-range. Each modeling operator can be mathematically represented as $WG(\mathbf{r}|\mathbf{x})G(\mathbf{x}|\mathbf{s})$. Here, W, $G(\mathbf{x}|\mathbf{s})$ and $G(\mathbf{r}|\mathbf{x})$ indicate the source wavelet, the Green's function from the source location $\mathbf{s}$ to the pertubation point $\mathbf{x}$ and from the pertubation point $\mathbf{x}$ to the receiver at $\mathbf{r}$, respectively. Here, $\mathbf{m}$ and $\mathbf{d}$ represent the corresponding model parameters and seismic traces with different offset ranges.

In equation 5.11, the solution of each block is found by the procedure listed in the MRPI + DF workflow. Once the model $\mathbf{m}^{near}$ is founded that explains the near-offset traces $\mathbf{d}^{near}$, then the model $\mathbf{m}^{near}$ is used as the starting model for the next block of traces. This procedure is repeated until the last block of equations is solved. In the practical case, to ensure that the inverted model from the later blocks is still able to explain the seismic traces in the previous block, we solve the system of equations in the following way:

1. Solve the first block in equation 5.11 to invert for $\mathbf{m}^{near}$.

2. Use $\mathbf{m}^{near}$ as the starting model to solve the first block in equation 5.12

$$\begin{matrix} \textit{Near to medium-offset block} \longrightarrow \\ \textit{Far-offset block} \longrightarrow \end{matrix} \begin{bmatrix} \mathbf{L}^{near\_mid} & \\ & \mathbf{L}^{far} \end{bmatrix} \begin{bmatrix} \mathbf{m}^{near\_mid} \\ \mathbf{m}^{far} \end{bmatrix} = \begin{bmatrix} \mathbf{d}^{near\_mid} \\ \mathbf{d}^{far} \end{bmatrix}, \tag{5.12}$$

where the near- to medium-offset block indicated by the blue color is the combination of the near- and medium- block in equation 5.11. The inverted model $\mathbf{m}^{near\_mid}$ is mostly inverted from the medium offset data $\mathbf{d}^{mid}$, while still satisfying the system of equations in the first block of equation 5.11.

3. Use $\mathbf{m}^{near\_mid}$ as the starting model to solve equation

$$\textit{Full-offset block} \longrightarrow \begin{bmatrix} \mathbf{L}^{near\_mid\_far} \end{bmatrix} \begin{bmatrix} \mathbf{m}^{near\_mid\_far} \end{bmatrix} = \begin{bmatrix} \mathbf{d}^{near\_mid\_far} \end{bmatrix}, \tag{5.13}$$

Similar to the IGBS method, this block-by-block method is repeated until convergence.

5.3 Numerical Results

We now use synthetic data and recorded marine data to test the effectiveness and advantages of MRPI + DF compared to conventional FWI and RFWI. The marine data are field OBC traces recorded in a North sea seismic survey. In all of the examples, the first arrivals are muted from the traces to ensure that the velocity updates are only contributed by the reflection wavepaths.

5.3.1 Two-Layer Model

We first use a two-layer model to illustrate the merits of MRPI + DF compared to conventional FWI and RFWI. Figure 5.15a depicts the velocity model used to

generate the data. There are 100 shots evenly spaced at a distance of 20 m and each shot is recorded using 200 receivers placed on the surface with an interval of 10 m. The source function is a Ricker wavelet with a peak frequency of 25 Hz. A time-domain first-order acoustic finite-difference modeling algorithm is used for both data simulations and inversion. The initial velocity model is homogeneous with $v = 1400\ m/s$, which is about 23% smaller than the true velocity.

Figures 5.14a and 5.15c show, respectively, the virtual source and the corresponding gradient at the 1st iteration of conventional FWI. This gradient is dominated by high-wavenumber energy where the low-wavenumber information is slightly updated. Figure 5.14b displays the virtual source from conventional RFWI which suffers from severe cycle-skipping problems. Therefore the gradient of conventional RFWI in Figure 5.15e contains strong aliasing artifacts which promote incorrect updates of the velocity model. The virtual source from the MRPI + DF method is shown in Figure 5.14c where the cycle-skipped data are muted out. Figure 5.15g shows the MRPI + DF gradient which is characterized by balanced amplitudes at low-wavenumbers. The final tomograms computed from these methods are shown in Figures 5.15d, 5.15f, and 5.15h, where both FWI and RFWI fail to update the velocity model. The area indicated by the black box gets less updated because of the imbalanced illumination caused by the acquisition geometry.

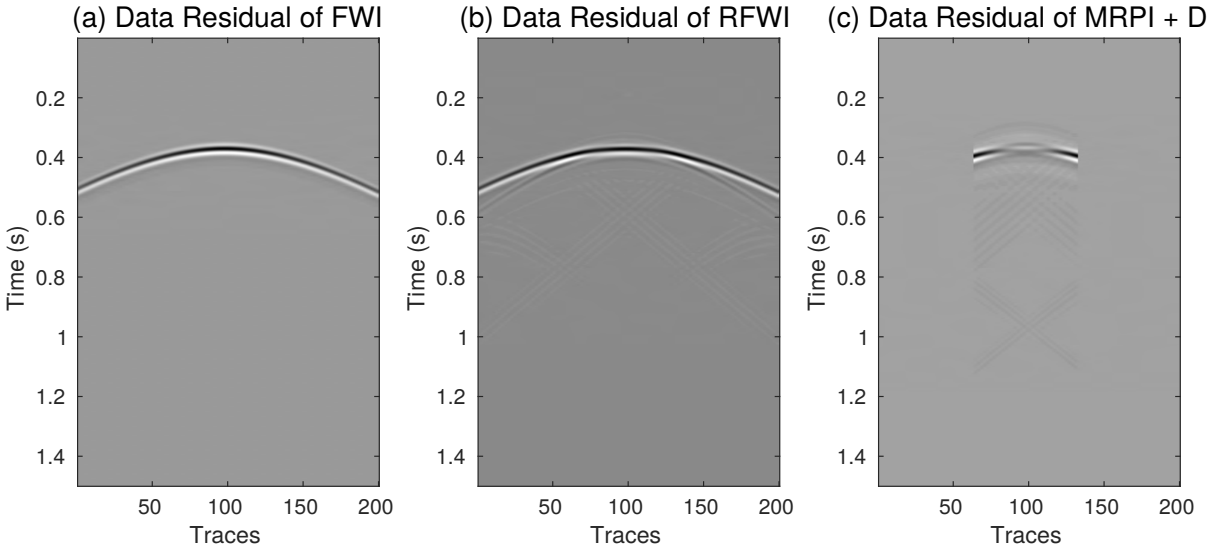


Figure 5.14: Comparison of the virtual source at the 1st iteration of (a) conventional FWI, (b) conventional RFWI and the MRPI + DF method.

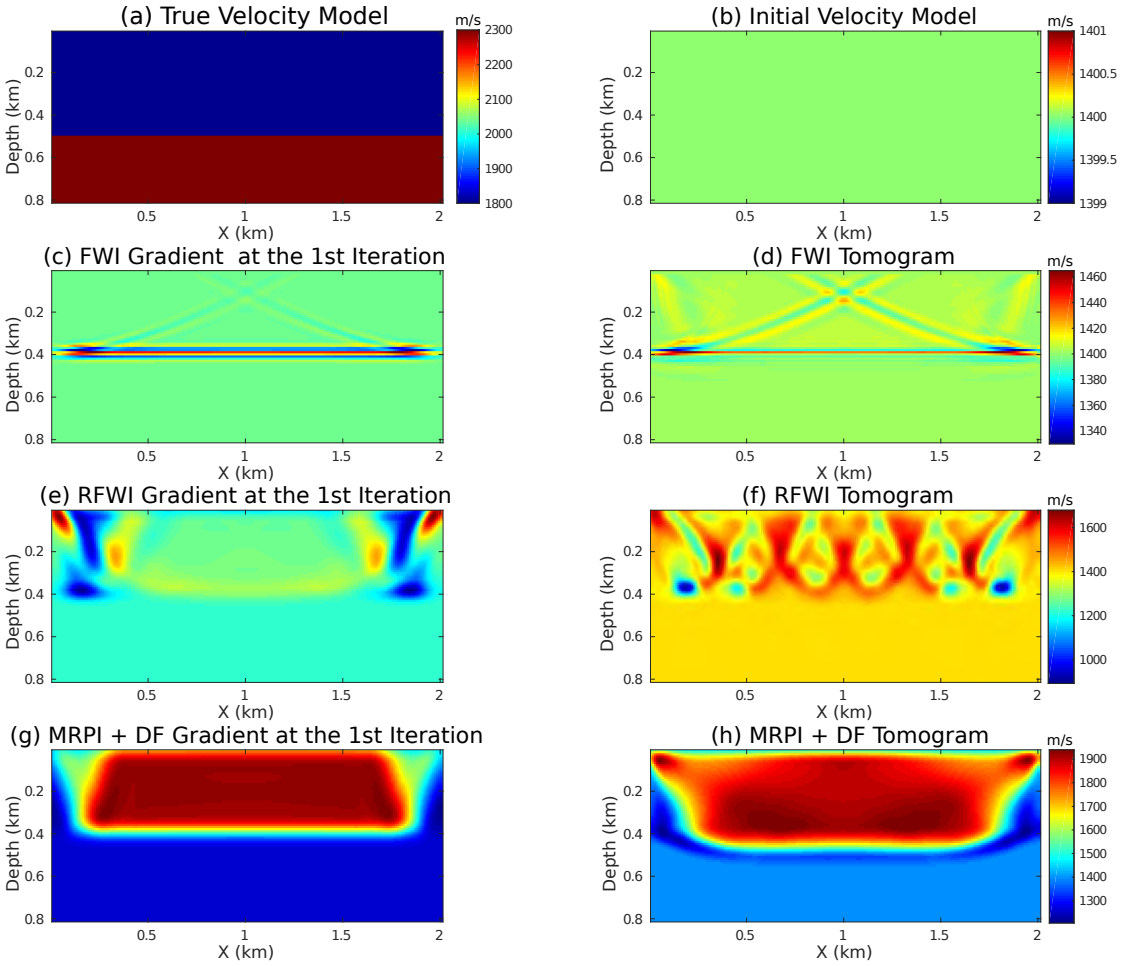


Figure 5.15: The (a) true velocity model, (b) initial velocity model, (c) FWI gradient at the 1st iteration for the, (d) FWI tomogram, (e) RFWI gradient at the 1st iteration of the (f) RFWI Tomogram, (g) MRPI + DF gradient at the 1st iteration of the (h) MRPI + DF tomogram.

5.3.2 Marmousi Model

Data computed from the Marmousi model are used to test the MRPI + DF method. We select the upper-right region of the Marmousi model with a size of 121 x 272 grid points. We then extend the model by 50 grid points on each side. The finite-difference method is used to compute 186 shot gathers on the surface of the Figure 5.16a model with 20 m shot intervals. Each shot is recorded with 800 receivers that are evenly distributed on the surface at a spacing of 10 m. A Ricker wavelet with a

peak frequency of 15 Hz is used as the source wavelet and the initial velocity model is shown in Figure 5.16b.

Figures 5.16c to 5.16g compares the gradients at the first iteration for the FWI, RFWI and MRPI + DF methods. The FWI gradient in Figure 5.16c is dominated by high-wavenumber components, where the low-wavenumber information is hardly updated. Figures 5.16e and 5.16g display the gradients for RFWI and MRPI + DF, which are mainly controlled by the low-wavenumber components. However, the RFWI gradient fails to update the right part of the model. The final inverted models are compared in Figures 5.16d to 5.16h, where the FWI tomogram in Figure 5.16b mainly updates the high-wavenumber components of the velocity model. The RFWI tomogram shown in Figure 5.16f mostly updates the low-wavenumber components in the shallow area and largely ignores the deeper portions of the velocity model. The weak updates in the deeper parts of the velocity model are largely a result of cycle-skipping between the observed and predicted traces. Figure 5.16h displays the tomogram inverted by MRPI + DF, where the low-wavenumber components of the velocity model are largely recovered. Figure 5.17 compares the vertical profiles between the inverted results and the true model, which suggests that MRPI + DF is more accurate than the conventional waveform inversion method.

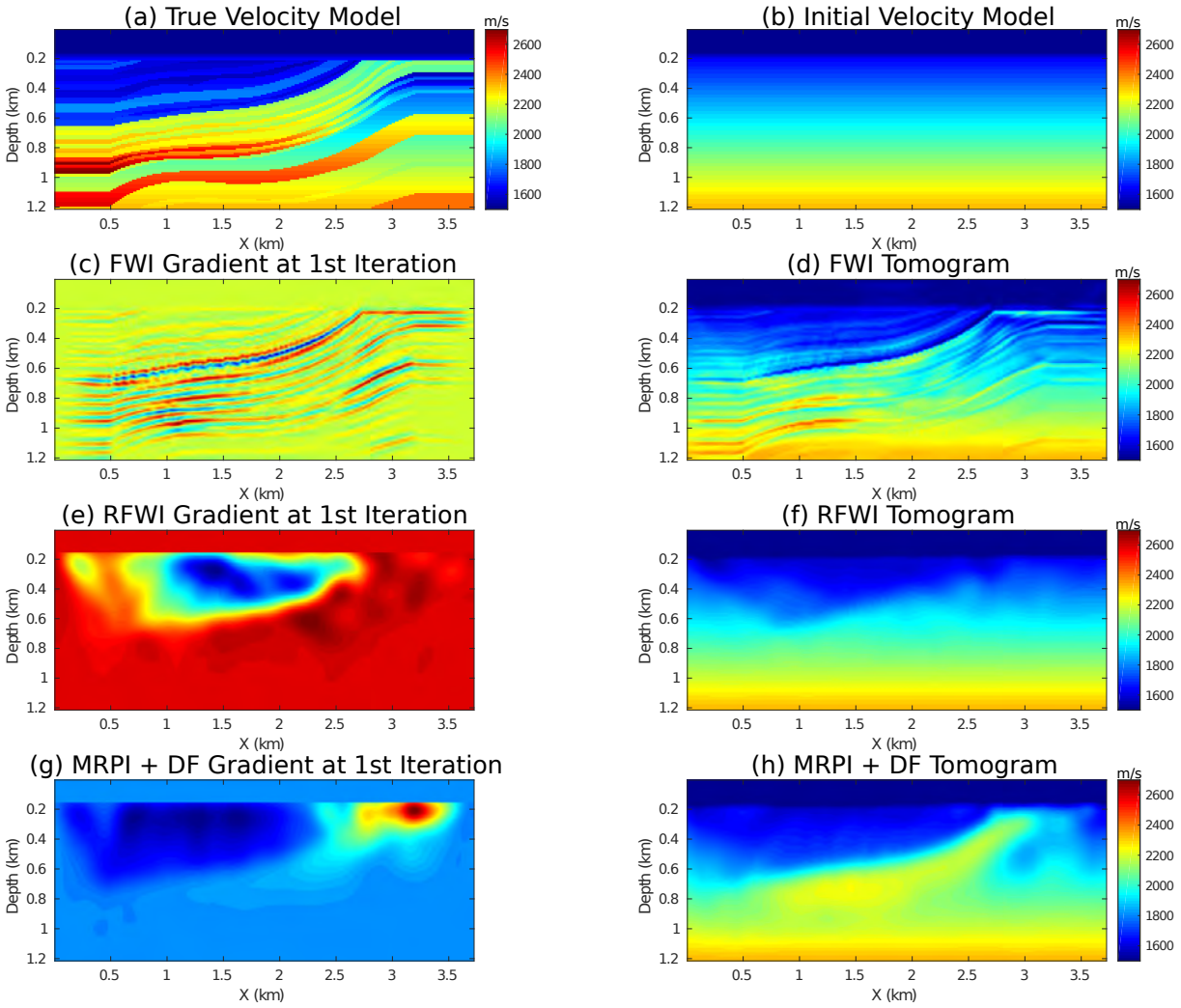


Figure 5.16: The (a) true velocity model, (b) initial velocity model, (c) FWI gradient at the 1st iteration, (d) FWI tomogram, (e) RFWI gradient at the 1st iteration, (f) RFWI Tomogram, (g) MRPI + DF gradient at the 1st iteration, and (h) MRPI + DF tomogram.

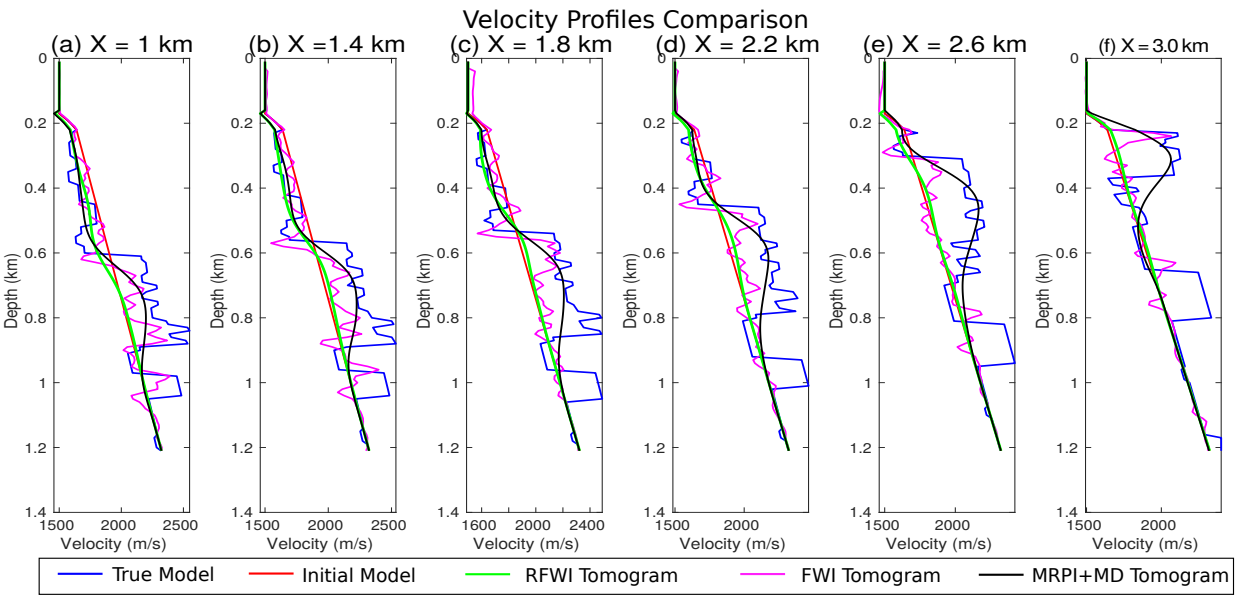


Figure 5.17: The comparisons of the velocity profiles in Figures 5.16(a), (b), (d), (f) and (h) at (a) x = 1 km, (b) x = 1.4 km, (c) x = 1.8 km, (d) x = 2.2 km, (e) x = 2.6 km and (f) x = 3.0 km.

The migration images computed from the tomograms in Figure 5.16 are shown in Figure 5.18. Here, the FWI and RFWI migration images fail to accurately image the deep reflectors. In comparison, the deep reflectors are accurately imaged by the MRPI + DF tomogram with more balanced amplitudes and better resolution.

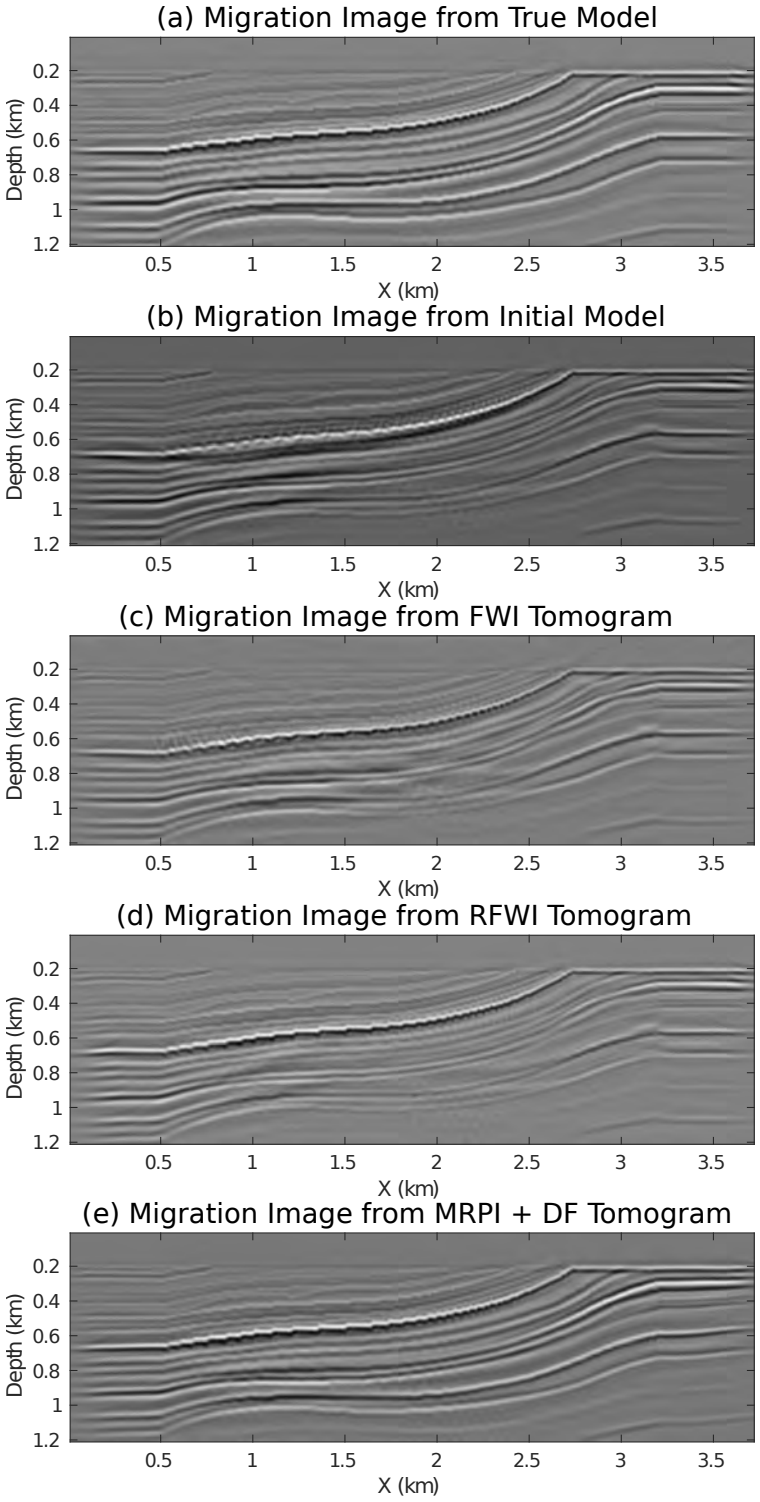


Figure 5.18: Comparisons of the migration images computed from (a) true velocity model, (b) initial velocity model, (c) FWI tomogram, (d) RFWI tomogram and (e) MRPI + DF tomogram.

5.3.3 Volve OBC Data

We test the effectiveness of MRPI + DF using a 2D slice of inline shot gathers extracted from the 3D Volve OBC data (Fu et al., 2017). The 3D dataset consists of 12 parallel cables with 400 m separations, and each cable contains 240 receivers. The 2D dataset is extracted from line 6 which is in the middle of the survey. The recorded data consist of 240 common-shot gathers with a shot interval of around 50 m, and the number of receivers for each shot ranges from 77 to 237 with a receiver interval of 50 m. The data were recorded with a sampling interval of 4 ms for the total recording time of 7.504 s. The lowest available frequency of the data is about 2 Hz. During processing, the data are low-pass filtered to 14 Hz and resampled to 1 ms. The traces are muted after 5 s as there are no reflections below 5 s. Figures 5.19a and 5.19b show the comparison between a raw and processed shot gather, where the reflections arriving before 2 s have a relatively low signal-to-noise (SNR) ratio; therefore, these low SNR reflections are also removed. We also apply a gain in time to compensate for the 3D effects. Finally, the traces are normalized to enhance the contribution of the far-offset traces and the source time history is a Ricker wavelet with the peak frequency of 7 Hz. We do not estimate the source from the data because the migrated image automatically contains the source signature information embedded in the observed data (Wang et al., 2013). Figures 5.19b and 5.19c show the comparison between the original processed data before and after three integrations. Their normalized amplitude spectra are shown in Figure 5.19d where the low-frequency components in the data are largely boosted from the integration.

The arrivals between 2.5-3.0 s are the reflections from the chalk layer (Duan et al., 2017). Applying FWI, RFWI and MRPI + DF to these data result in the inverted tomograms shown in Figure 5.20. The data are separated into 4 offset groups for migration and inversion, which are 1250/1875 m, 1875/2650 m, 2650/3250 m and 3250/5000 m. The first and second number in each group indicates the offset range

of data for migration and inversion, respectively. Here the data are separated into the four different offset groups by following the workflow described in the previous "Discussion of the Offset Selection" section. The chalk layer located at the depth around 2.5 km is selected as the reference reflector and equation 5.8 is used to decide the optimal data offset under the current velocity model. Once the offset is selected, we update the velocity model with the traces in this offset range. The velocity is updated until the residual decrease fall below a specified threshold. We repeat this procedure until the end of the offset range is reached. The migration images computed from the different velocity models in Figure 5.20 are shown in Figure 5.21. Figure 5.21d is the migration image calculated from the MRPI + DF tomogram, which has a wider imaging area and the reflectors are more focused and continuous when compared to the others. The red arrows point to the areas in which noticeable improvements in image quality can be seen.

To further test for the accuracy of the tomograms, the sub-offset gathers shown in Figures 5.22a and 5.22b are computed from the initial velocity model and MRPI tomograms. The energy in the sub-offset gather in Figure 5.22a is spread out over a large range of sub-offset values, which indicates that the initial model is far away from the true model. But the energy of the sub-offset gather in Figure 5.22b is mostly focused at zero sub-offset, which indicates that the MRPI + DF tomogram is a good representation of the true velocity model. Figures 5.23a and 5.23b show the angle-domain common image gathers (ADCIGs) calculated from the initial model and MRPI + DF tomogram, respectively. It is obvious that the ADCIGs associated with the MRPI + DF tomogram is more flattened than those from the initial model. The observed and Born data from the MRPI + DF tomogram are compared in Figure 5.24. The red and green wiggles represent the observed and Born data, respectively, where their phases are in good agreement.

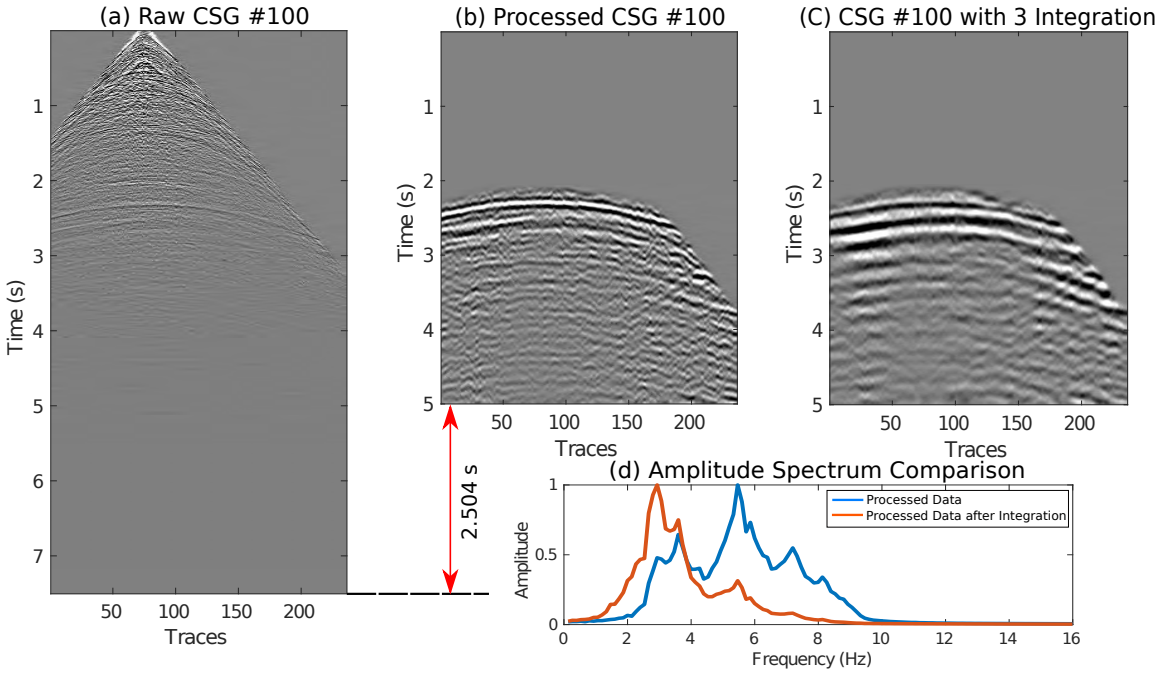


Figure 5.19: The comparisons between the (a) raw shot gather, (b) processed shot gather, and (c) processed shot gather with three integrations. (d) is the comparison of the normalized amplitude spectrum for (b) and (c).

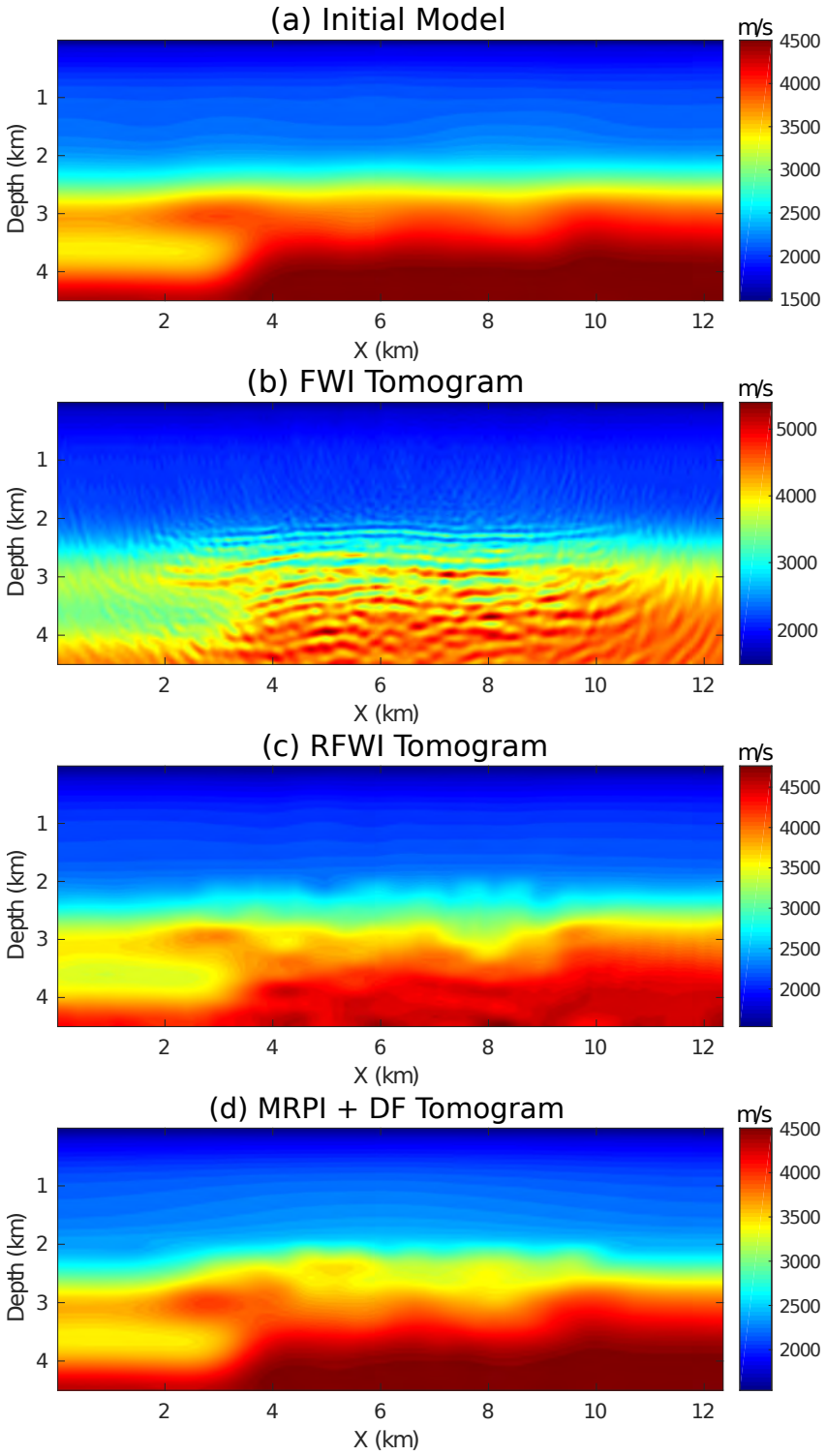


Figure 5.20: Comparsions of the velocity models for the (a) initial model, (b) FWI tomogram, (c) RFWI tomogram and (d) MRPI + DF tomogram.

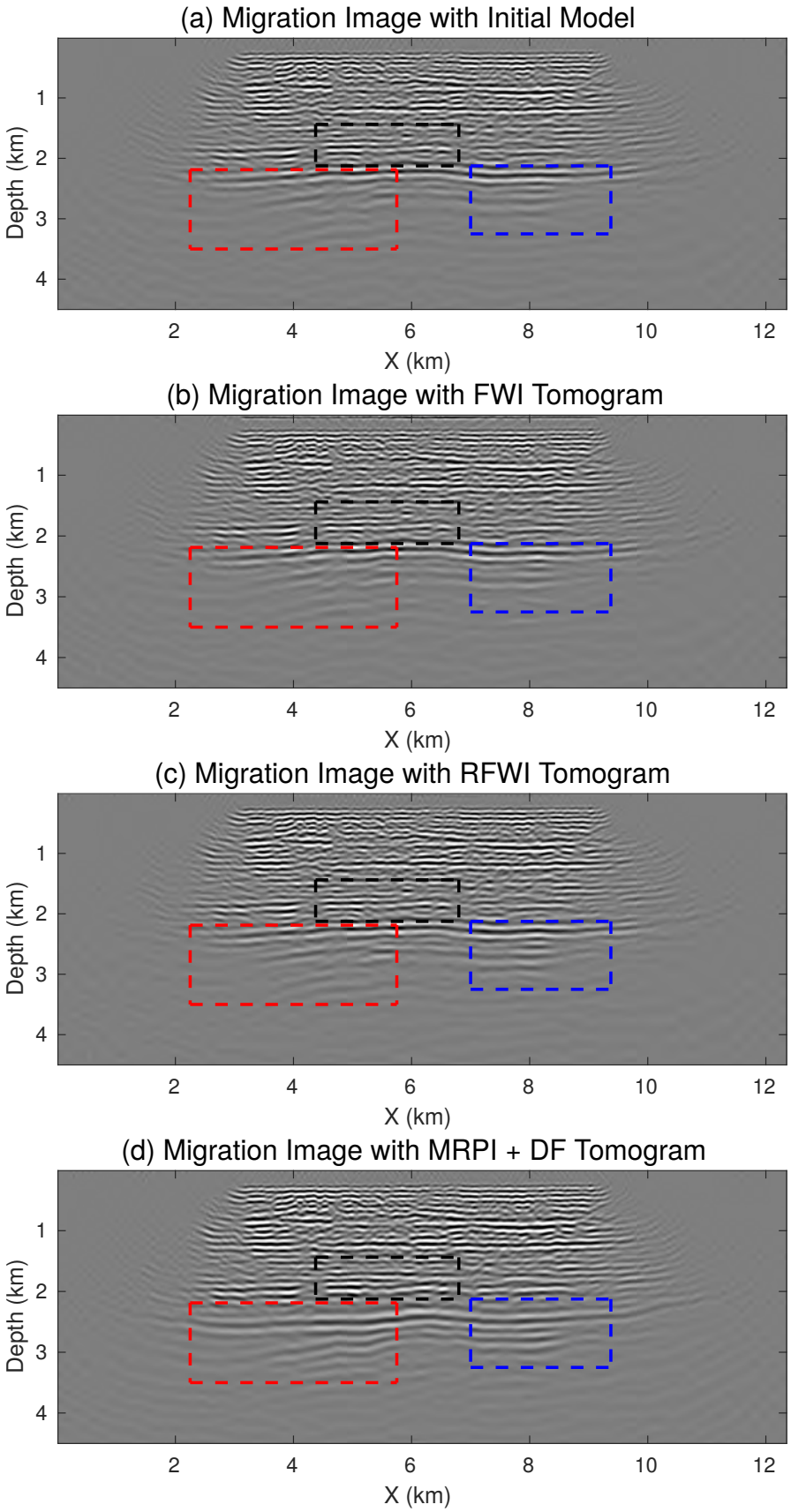


Figure 5.21: Comparsions of the migration images computed from the (a) initial model, (b) FWI tomogram, (c) RFWI tomogram and (d) MRPI + DF tomogram.

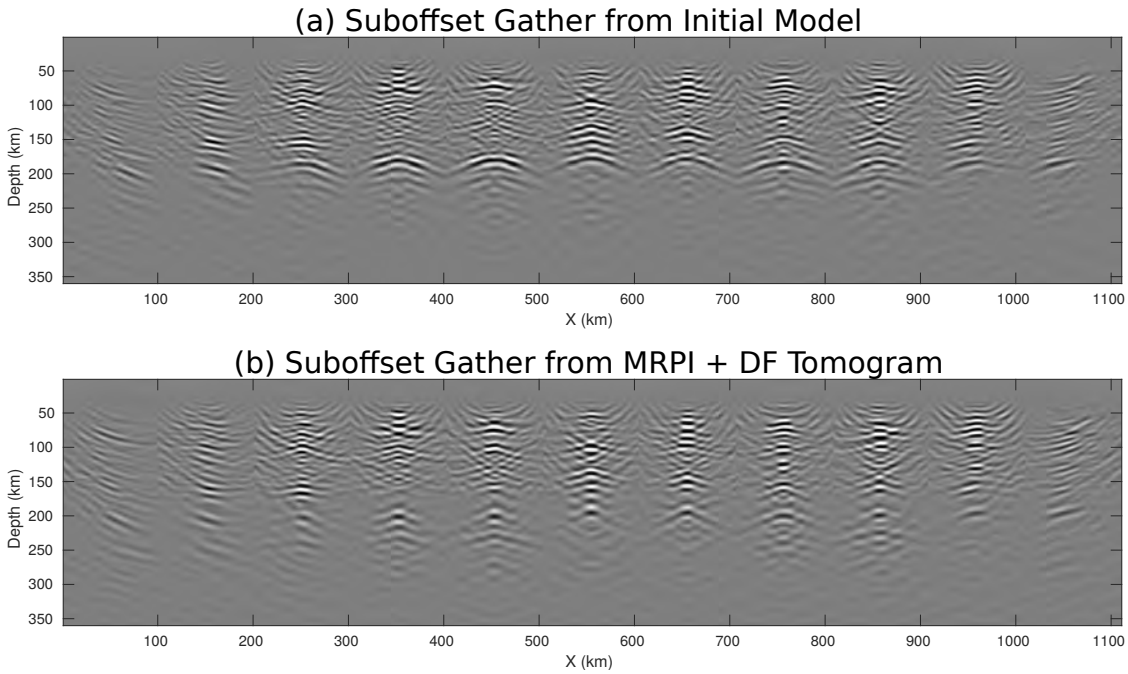


Figure 5.22: Comparsion of the sub-offset gathers computed from the (a) initial model and (b) MRPI + DF tomogram.

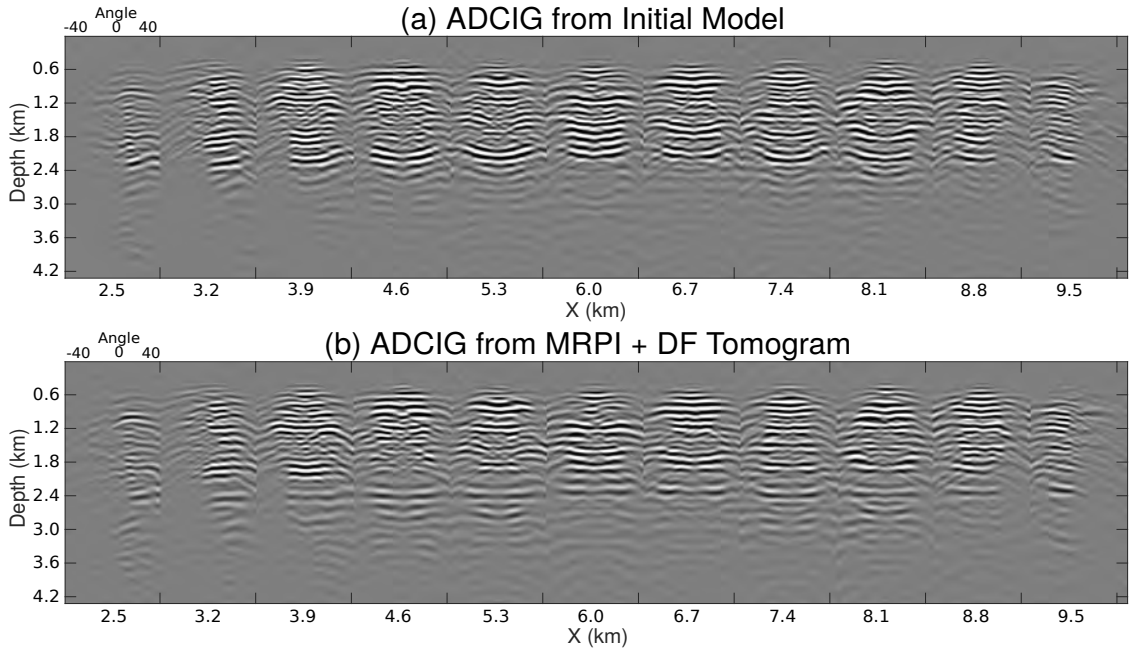


Figure 5.23: Comparsion of the angle gathers computed from the (a) initial model and (b) MRPI + DF tomogram.

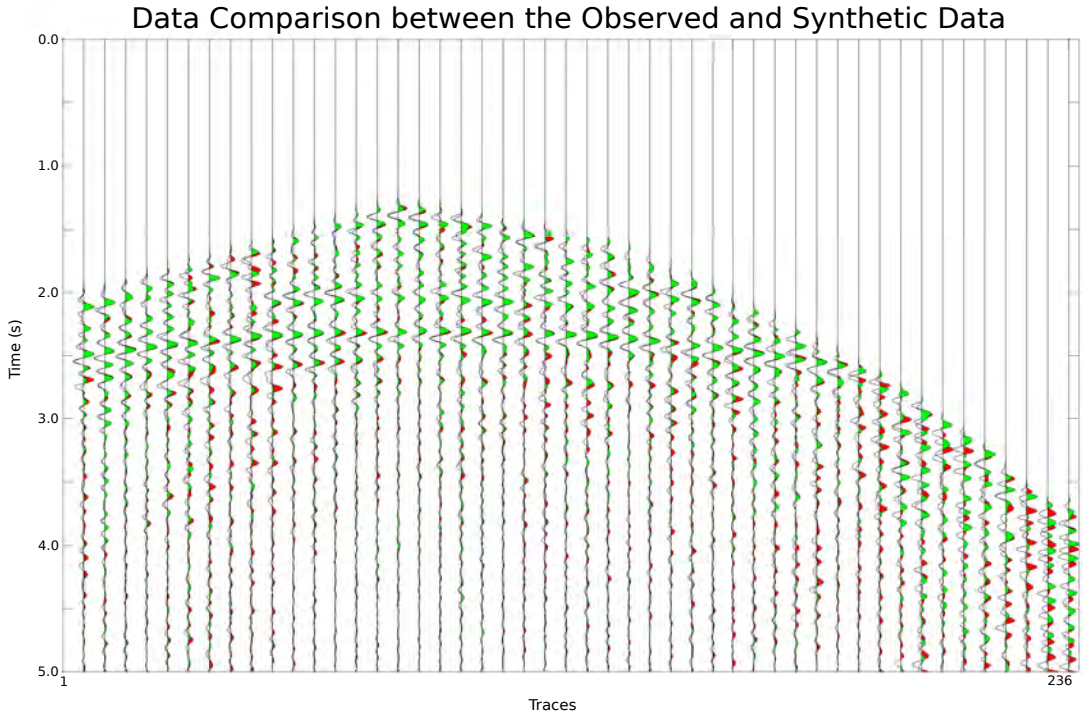


Figure 5.24: Comparison between the observed data and Born data, where the rad and green wiggles represent the observed and Born data, respectively.

Chapter 6

Seismic Inversion by Newtonian Machine Learning

We present a seismic inversion method by newtonian machine learning that inverts skeletonized data for the subsurface velocity model. The skeletonized representation of the seismic traces consists of the low-rank latent-space variables predicted by a well-trained autoencoder neural network. The input to the autoencoder consist of the recorded common shot gathers, and the implicit function theorem is used to determine the perturbation of the skeletonized data with respect to the velocity perturbation. The gradient is computed by migrating the observed traces weighted by the residuals of the skeletonized data, and the final velocity model is the one that best predicts the observed latent-space parameters. We denote this hybrid inversion method as inversion by Newtonian machine learning because it inverts for the model parameters by combining the deterministic laws of Newtonian physics with the statistical capabilities of machine learning. Empirical results suggest that the cycle-skipping problem is largely mitigated compared to the conventional full waveform inversion (FWI) method by replacing the waveform differences by the those of the latent-space parameters. Numerical tests on both the synthetic and real data demonstrate the success of this skeletonized inversion method in recovering a low-wavenumber approximation to the subsurface velocity model. The advantage of this method over other skeletonized data methods is that no manual picking of important features is required because the skeletal data are automatically selected by the autoencoder. The disadvantage is that the inverted velocity model has less resolution compared to the FWI result, but which can be a good initial model for FWI. We suggest that the

lowered resolution problem can be mitigated by using a multiscale method where the dimension of the latent space is gradually increased and more complexity is included into the input data.

The most significant contribution of this paper is that it provides a general framework for using solutions to the governing PDE to invert skeletal data generated by any type of a neural network. The governing equation can be that for gravity, seismic waves, electromagnetic fields, and magnetic fields. The input data can be the records from different types of data and their skeletal features, as long as the model parameters are sensitive to their perturbations. The skeletal data can be the latent space variables of an autoencoder, a variational autoencoder, or a feature map from a convolutional neural network (CNN), or principal component analysis (PCA) features. In other words, we have combined the best features of Newtonian physics and the pattern matching capabilities of machine learning to invert seismic data by Newtonian machine learning.

6.1 Introduction

Full waveform inversion (FWI) has been shown to accurately invert seismic data for high-resolution velocity models (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). However, the success of FWI heavily relies on a good initial model that is close to the true model, otherwise, cycle-skipping problems will trap the FWI in a local minimum(Bunks et al., 1995b). To mitigate the FWI cycle-skipping problem, Bunks et al. (1995b) proposed a multiscale inversion approach which initially inverts low-pass seismic data and then gradually admits higher frequencies as the iterations proceed. AlTheyab and Schuster (2015) removed the mid- and far-offset cycle-skipped seismic traces before inversion and gradually incorporates them into the iteration solution as the velocity model become closer to the true model. Wu et al. (2014) use the envelope of the seismic traces to invert for the subsurface model as they claim

that the envelope carries the ultra-low frequency information of the seismic data. Ha and Shin (2012) invert the data in the Laplace-domain which is less sensitive to the lack of low frequencies than conventional FWI. Sun and Schuster (1993) and Fu et al. (2017) use an amplitude replacement method to focus the inversion on reducing the phase mismatch instead of the waveform mismatch. In addition, they employ a multiscale approach by temporally integrating the traces to boost the low-frequencies and mitigate cycle-skipping problems, and then gradually introduce the higher frequencies as the iterations proceed.

The main reason non-linear inversion gets stuck in a local minimum is that the data are very complex (i.e, wiggly in time), which means that the objective function is very complex and characterized by many multiple minimums. To avoid this problem, Luo and Schuster (1991a) suggested a skeletonized inversion method which combines the skeletonized representation of seismic data with the implicit function theorem to accelerate convergence to the vicinity of the global minimum (Lu et al., 2017). Simplification of the data by skeletonization reduces the complexity of the misfit function and reduces the number of local minima. Examples of wave-equation inversion of skeletonized data include the following:

- Lu et al. (2017) uses the solutions to the wave equation to invert the first-arrival traveltimes for the low-to-intermediate wavenumber details of the background velocity model. Feng and Schuster (2019) uses the traveltime misfit function to invert for both the subsurface velocity and anisotropic parameters in a vertical transverse isotropic medium.

- Instead of minimizing the traveltime misfit function, Li et al. (2016) finds the optimal S-velocity model that minimizes the difference between the observed and predicted dispersion curves associated with surface waves. Liu et al. (2018) extend 2D dispersion inversion of surface waves to the 3D case.

- Instead of inverting for the velocity model, Dutta and Schuster (2016) developed a wave-equation inversion method that inverts for the subsurface Q_p distribution. Here, they find the optimal Q_p model by minimizing the misfit between the observed and the predicted peak/centroid-frequency shifts of the early arrivals. Similarly, Li et al. (2017) utilize the peak frequency shift of the surface waves to invert for the Q_s model.

- A good tutorial for skeletonized inversion is by Lu et al. (2017).

One of the key problems with skeletonized inversion is that the skeletonized data must be picked from the original data, which can be labor intensive for large data sets. To overcome this problem, we propose obtaining the skeletonized data from an autoencoder, and then use solutions to the wave equation to invert such data for the model of interest (Schuster, 2018). The skeletonized data correspond to the feature map in the latent space of the autoencoder, which has a reduced dimension and contains the significant parts of the input data related to the model. That is, we have combined the best features of Newtonian physics and the pattern matching capabilities of machine learning to invert seismic data by Newtonian machine learning.

The autoencoder neural network is an unsupervised deep learning method that is trained for dimensionality reduction (Schmidhuber, 2015). An autoencoder maps the data into a lower-dimensional space by extracting the data's most important features. It encodes the original data into a much more condensed representation, also denoted as the skeletonized representation, of the input data. The input data can be reconstructed by a decoder from the encoded value. In this paper, we first use the observed seismic traces as the training set to train the autoencoder neural network. Once the autoencoder is well trained, we feed both the observed and synthetic traces into the autoencoder to get their corresponding low-dimension representations. We build the misfit function as the sum of the squared differences between the observed and the predicted encoded value. To compute the gradient with respect to the model

parameters such as the velocity in each pixel, we use the implicit function theorem to compute the perturbation of the skeletonized information with respect to the velocity. The high-level strategy for inverting the skeletonized latent variables is summarized in Figure 6.1, where **L** corresponds to the forward modeling operator of the governing equations, such as the wave equation.

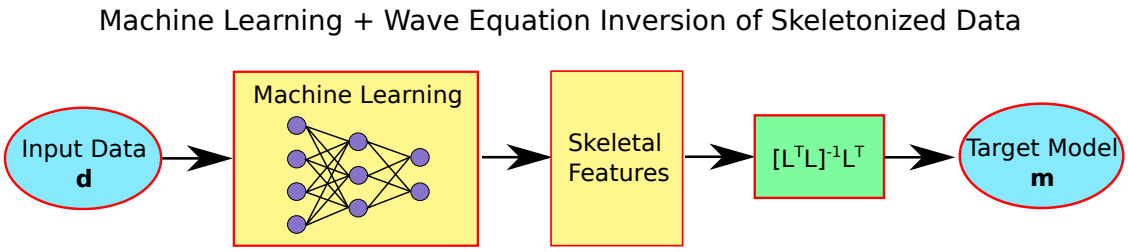


Figure 6.1: The strategy for inverting the skeletonized latent variables.

This paper is organized into four sections. After the introduction, we explain the theory of the wave equation inversion of seismic data skeletonized by an autoencoder. This theory includes the formulation first presented in Luo and Schuster (1991a,b) where the implicit function theorem is used to employ numerical solutions to the wave equation for generating the Fréchet derivative of the skeletal data. We then present the numerical results for both synthetic data and field data recorded by a crosswell experiment. The last section provides a discussion, a summary of our work and its significance.

6.2 Theory

Conventional full waveform inversion (FWI) inverts for the subsurface distribution by minimizing the l^2 norm of the waveform difference between the observed and synthetic data. However, this misfit function is highly nonlinear and the iterative solution often gets stuck into the local minima (Bunks et al., 1995b). To mitigate the problem, skeletonized inversion methods simplify the objective function by combining

the skeletonized representation of data, such as the traveltimes, with the implicit function theorem, to give a gradient optimization method that quickly converges to the vicinity of the glocal minimum. Instead of manually picking the skeletonized data, we allow the unsupervised autoencoder to generate it.

6.2.1 Theory of Autoencoder

An autoencoder is an unsupervised neural network in which the predicted output is the same as the input data, as illustrated in Figure 6.2. An autoencoder is trained to learn the extremely low-dimensional representation of the input data, also denoted as the skeletonized representation, in an unsupervised manner. It is similar to the principal component analysis (PCA), which is generally used to represent input data using a smaller dimensional space than originally present (Hotelling, 1933). However, PCA is restricted to finding the optimal rotation of the original data axes that maximizes its projections to the principal components axes. In comparison, the autoencoder with a sufficient number of layers can find almost any non-linear sparse mapping between the input and output images. A typical autoencoder architecture is shown in Figure 6.2 which generally includes three parts: the encoder, the latent space, and the decoder.

- Encoder: Unsupervised learning by an autoencoder uses a set of training data consisting of N training samples $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}\}$, where $\mathbf{x}^{(i)}$ is the i^{th} feature vector with dimension $D \times 1$ and D represent the number of features for each feature vector. The encoder neural network indicated by the pink box in Figure 6.2 encodes the high-dimension input data $\mathbf{x}^{(i)}$ into a low-dimension latent space with dimension $C \times 1$ using a series of neural layers with a decreasing number of neurons; here C is smaller than D. This encoding operations can be mathematically described as $\mathbf{z}^{(i)} = g\big(\mathbf{W}_1\mathbf{x}^{(i)} + \mathbf{b}_1\big)$, where $\mathbf{W}_1$ and $\mathbf{b}_1$ represent the model parameter and the vector of bias terms for the first layer, and $g()$

indicates the activation function such as a sigmoid, ReLU, Tanh and so on.

- Latent Space: The compressed data $\mathbf{z}^{(i)}$ with dimension $C \times 1$ in the latent space layer (emphasized by the green box) is the lowest dimension space in which the input data is reduced and the key information is preserved. The latent space usually has a few neurons which forces the autoencoder neural network to create effective low-dimension representations of the high-dimension input data. These low-dimension attributes can be used by the decoder to reconstruct the original input.

- Decoder: The decoder portion of the neural network represented by the purple box reconstructs the input data from the latent space representation $\mathbf{z}^{(i)}$ by a series of neural network layers with an increasing number of neurons. The reconstructed data $\tilde{\mathbf{x}}^{(\mathbf{i})}$ are calculated by $\tilde{\mathbf{x}}^{(i)} = \mathbf{W}_2\mathbf{z}^{(i)} + \mathbf{b}_2$, where $\mathbf{W}_2$ and $\mathbf{b}_2$ represent the model parameter and the bias term of the decoder neural network, respectively.

The parameters of the autoencoder neural network are determined by finding the values of $\mathbf{w}_i$ and $\mathbf{b}_i$ for $i = 1, 2$ that minimize the following objective function:

$$\begin{aligned} J(\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2) &= \sum_{i=1}^{N} (\tilde{\mathbf{x}}^{(i)} - \mathbf{x}^{(i)})^2, \\ &= \sum_{i=1}^{N} \left(\mathbf{W}_2\big(g(\mathbf{W}_1\mathbf{x}^{(i)} + \mathbf{b}_1)\big) + \mathbf{b}_2 - \mathbf{x}^{(i)} \right)^2. \end{aligned} \tag{6.1}$$

In practice a preconditioned steepest descent method is used for mini-batch inputs.

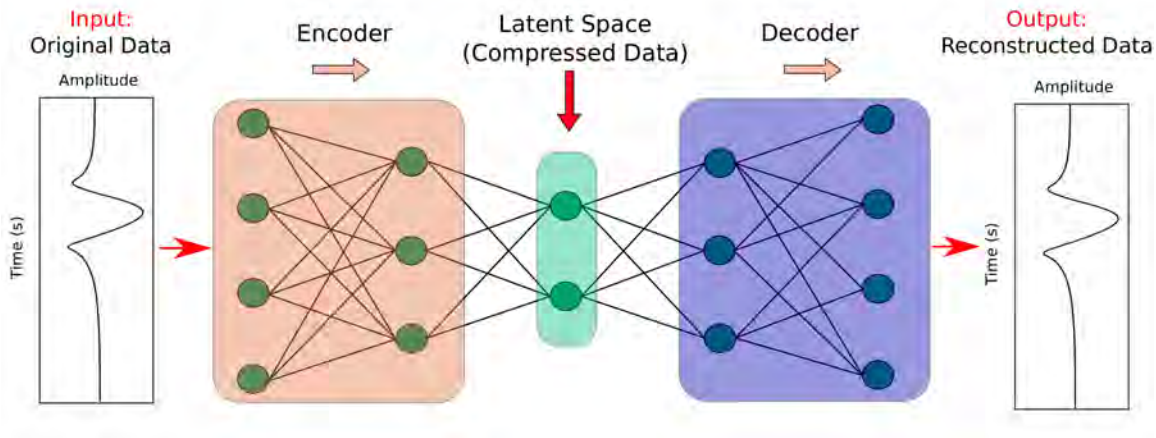


Figure 6.2: An example of an autoencoder architecture with two layers for encoder and two layer for decoder. The dimension of the latent space is two.

6.2.2 Skeletonized Representation of Seismic Data by Autoencoder

To get the low-dimension skeletonized representation of seismic data by the autoencoder, the input data consist of seismic traces, each with the dimension of $nt \times 1$. In this case, each seismic trace is defined as one training example in the training set generated by a crosswell seismic experiment. For the crosswell experiment, there are N_s sources in the source well and N_r receivers in the receiver well. We mainly focus on the inversion of the transmitted arrivals by windowing the input data around the early arrivals.

Figure 6.3a shows a homogeneous velocity model with a Gaussian anomaly in the center. Figure 6.3b is the corresponding initial velocity model which has the same background velocity as the true velocity model. A crosswell acquisition system with two 1570-m-deep cased wells separated by 1350 m is used as the source and receiver wells. The finite-difference method is used to compute 77 acoustic shot gathers for both the observed and synthetic data with 20 m shot intervals. Each shot is recorded with 156 receivers that are evenly distributed along the depth at a spacing of 10 m. To train the autoencoder network, we use the following workflow.

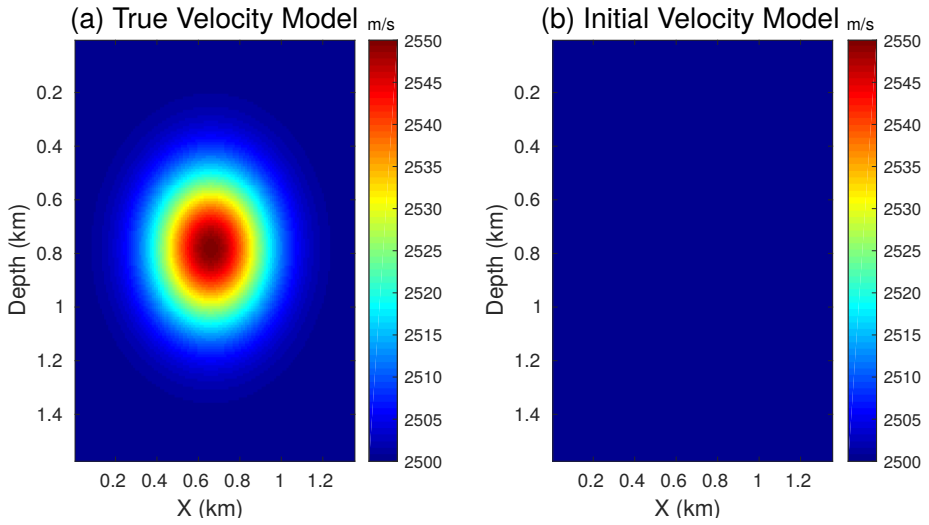


Figure 6.3: A homogeneous velocity model with a Gausian velocity anormaly in the center.

1. Build the training set. For every five observed shots, we randomly select one shot gather as part of the training set that consist of a total of 2496 training examples, or seismic traces. We didn't use all the shot gathers for training because of the increase in the computation cost.

2. Data processing. Each seismic trace is Hilbert transformed to get its envelope then subtracted by their mean and divided by their variance. Figure 6.4a and 6.4b show a seismic trace before and after processing, respectively. We use the signal envelope instead of the original seismic trace because it is less complicated than the original signal. And according to our tests, the signal envelope leads to faster convergence compared to the original seismic signal.

3. Training the autoencoder. We feed the processed training set into an autoencoder network where the dimension of its latent space is equal to 1. In other words, each training example with a dimension of $nt \times 1$ will be encoded as a smaller number of latent variables by the encoder. The autoencoder parameters are updated by iteratively minimizing equation 1. The Adam and mini-batch gradient descent methods are used to train this network. Figure 6.5a and 6.5b

show an input training example and its corresponding reconstructed signal by the autoencoder, respectively, and their difference is shown in Figure 6.5c.

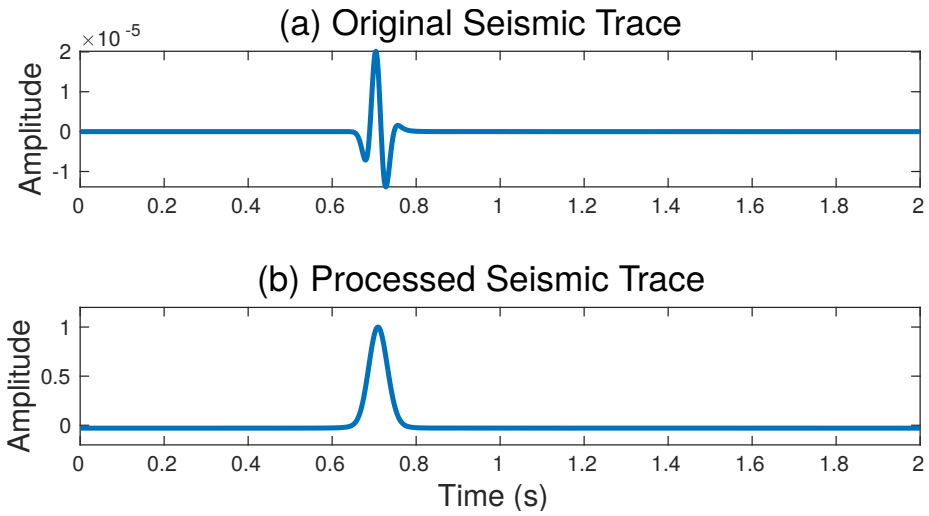


Figure 6.4: The (a) orginal seismic trace and the (b) processed seismic trace.

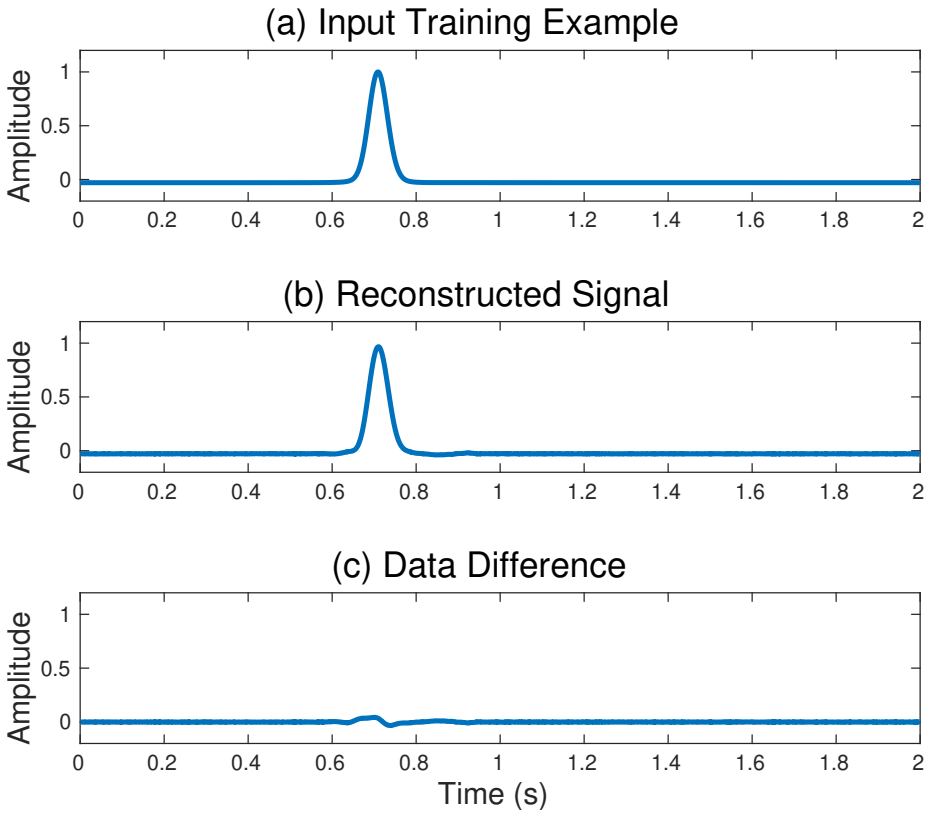


Figure 6.5: The (a) input training example, (b) reconstructed signal by autoencoder and their difference.

After training is finished, we input all the observed and predicted seismic traces into the well-trained autoencoder network to get their skeletonized low-dimensional representation. Of course, each input seismic trace requires the same data processing procedure as we did for the training set. Figure 6.6a, 6.6b and 6.6c shows three observed shot gathers which are not included in the training set, and their encoded values are shown in Figure 6.6d, 6.6e and 6.6f which are the skeletonized representations of the input seismic traces. The encoded values do not have any units and can be considered as a skeletonized attribute of the data. However, the autoencoder believes that these encoded values are the best low-dimension representation of the original input in the least-square sense.

We compare the traveltime differences and the encoded low-dimension representation differences for the observed and synthetic data in Figure 6.7. The black and red curves represent the observed and synthetic data, respectively. Figure 6.7b shows a larger traveltime difference than Figure 6.7a and 6.7c as its propagating waves are affected more by the Gaussian anomaly than the other two shots. However, the misfit function for the low-dimensional representation of the seismic data exhibits a pattern similar to that of the traveltime misfit function. Both reveal a large misfit at the traces affected by the velocity anomaly. Similar to the traveltime misfit values, the encoded values are also sensitive to the velocity changes. In this case, we can conclude that the (1) autoencoder network is able to estimate the effective low-dimension representation of the input data and (2) the encoded low-dimensional representation can be used as a skeletonized feature sensitive to changes in the velocity model.

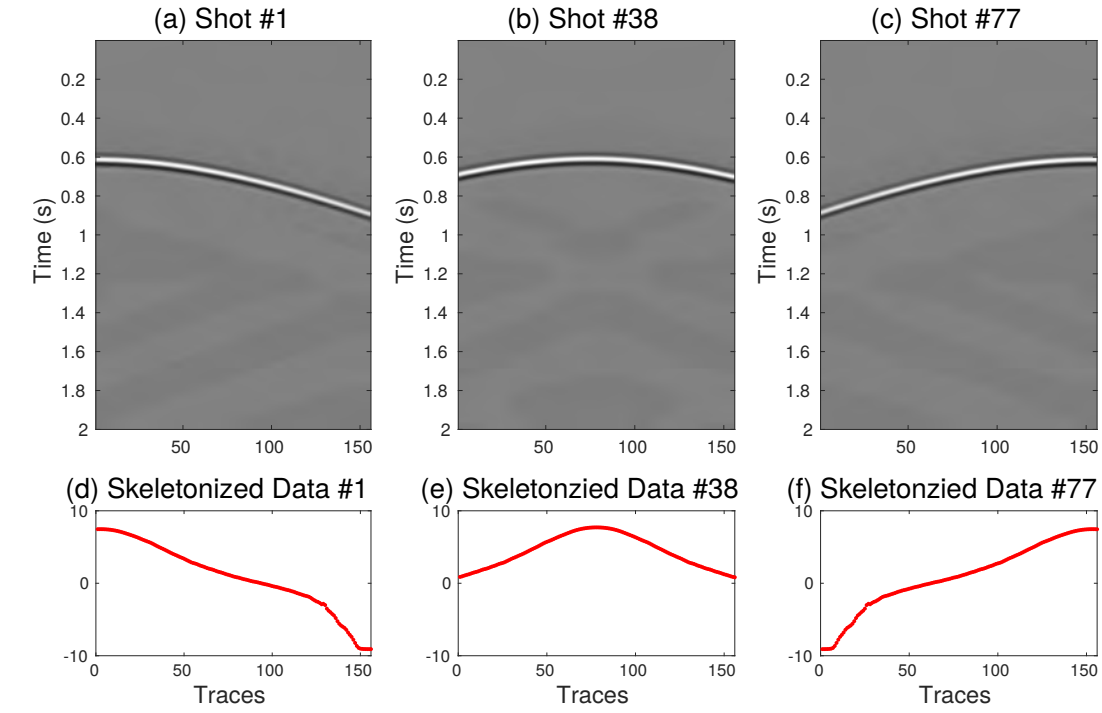


Figure 6.6: Three shot gathers with their corresponding encoded data.

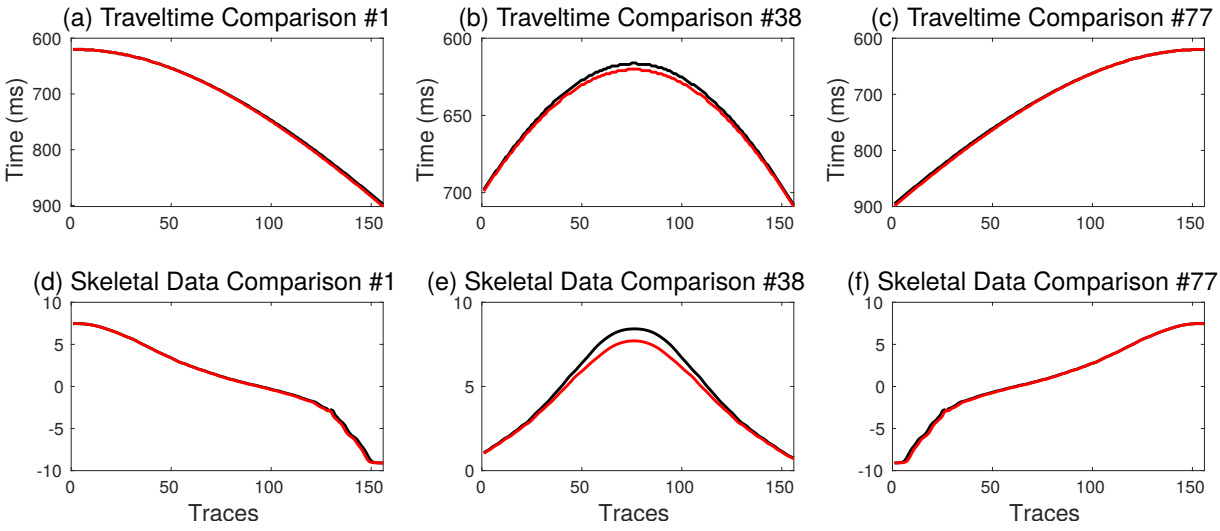


Figure 6.7: The comparison of the traveltime misfit functions and skeletal data misfit functions for different shot gathers. The black and red curves represent the observed and synthetic data, respectively.

6.2.3 Theory of the Skeletonized Inversion with Autoencoder

In order to invert for the velocity model from the skeletonized data, we use the implicit function theorem to compute the perturbation of the skeletonized data with respect to the velocity.

Connective Function

A cross-correlation function is defined as the connective function that connects the skeletonized data with the pressure field. This connective function measures the similarity between the observed and synthetic traces as

$$f_{z_1}(\mathbf{x}_r, t; \mathbf{x}_s) = \int dt p_{z-z_1}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} p_z(\mathbf{x}_r, t; \mathbf{x}_s)_{syn}, \tag{6.2}$$

where $p_z(\mathbf{x}_r, t; \mathbf{x}_s)_{syn}$ represents a synthetic trace for a given background velocity model recorded at the receiver location $\mathbf{x}_r$ due to a source excited at location $\mathbf{x}_s$. The subscript z is the skeletonized feature (low-dimension representaion of the seismic trace) that is encoded by a well-trained autoencoder network. Similarly, $p_{z-z_1}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs}$ denotes the observed trace with a encoded skeletonized feature equal to $z - z_1$ that has the same source and receiver location as $p_z(\mathbf{x}_r, t; \mathbf{x}_s)_{syn}$, and z_1 is the distance between the synthetic and observed skeletal data in the latent space.

For an accurate velocity model, the observed and synthetic traces will have the same encoded values in the latent space. Therefore, we seek to minimize the distance in the latent space between an observed and synthetic traces. This can be done by finding the shift value $z_1 = \Delta z$ that maximizes the crosscorrelation function in equation 6.2. If $\Delta z = 0$, it indicates that the correct velocity model has been found and the synthetic and observed traces have the same encoded values in the latent space. The Δz that maximizes the crosscorrelation function in equation 6.2 should satisfy the condition that the derivative of $f_{z_1}(\mathbf{x}_r, t; \mathbf{x}_s)$ with respect to z_1 is equal to

zero. Thus,

$$\dot{f}_{\Delta z} = \left[\frac{\partial f_{z_1(\mathbf{x}_r,t;\mathbf{x}_s)}}{\partial z_1}\right]_{z_1=\Delta z}, \tag{6.3}$$
$$= \int dt \dot{p}_{z-\Delta z}(\mathbf{x}_r,t;\mathbf{x}_s)_{obs} p_z(\mathbf{x}_r,t;\mathbf{x}_s)_{syn} = 0,$$

where $\dot{p}_{z-\Delta z}(\mathbf{x}_r,t;\mathbf{x}_s)_{obs} = \partial p_{z-z_1}(\mathbf{x}_r,t;\mathbf{x}_s)/\partial z_1$. Equation 3 is the connective function that acts as an intermediate equation to connect the seismogram with the skeletonized data, which are the encoded values of the seismograms (Luo and Schuster, 1991a,b). Such a connective function is necessary because there is no wave equation that relates the skeletonized data to a single type of model parameters (Dutta and Schuster, 2016). The connective function will be later used to derive the derivative of skeletonized data with respect to the velocity.

Misfit Function

The misfit function of the skeletonized inversion with the autoencoder method is defined as

$$\epsilon = \frac{1}{2}\sum_s \sum_r \Delta z(\mathbf{x}_r,\mathbf{x}_s)^2, \tag{6.4}$$

where Δz is the difference of the encoded value in the latent space between the observed and synthetic data. The gradient $\gamma(\mathbf{x})$ is given by

$$\gamma(\mathbf{x}) = -\frac{\partial \epsilon}{\partial v(\mathbf{x})} = -\sum_s \sum_r \frac{\partial \Delta z}{\partial v(\mathbf{x})} \Delta z(\mathbf{x}_r,\mathbf{x}_s). \tag{6.5}$$

Figure 6.8 shows the encoded value misfit versus different values of velocity, which clearly shows that the misfit monotonically decreases as the velocity value approaches to the correct velocity value ($v = 2200\ m/s$). Therefore, the skeletonized misfit

function in equation 6.5 is able to quickly converge to the global minimum when using the gradient optimization method. Using equation 3 and the implicit function theorem we can get

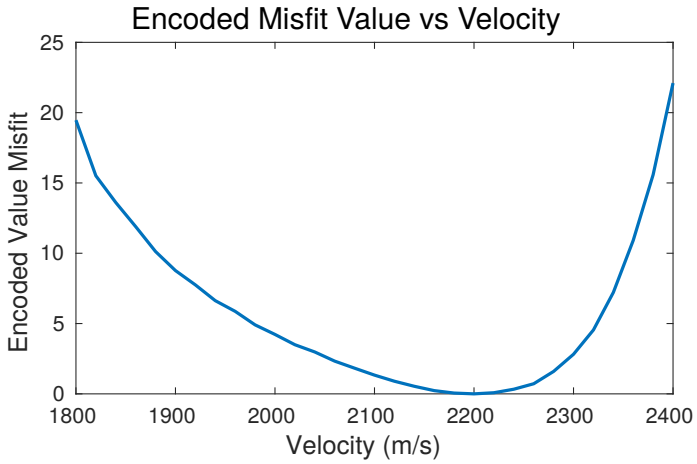


Figure 6.8: Plot of the encoded value misfit function versus hypothetical velocity values for the velocity model. The observed data is generated with $v = 2200\ m/s$.

$$\begin{aligned}\frac{\partial \Delta z}{\partial v(\mathbf{x})} &= \frac{\left[\frac{\partial \dot{f}_{\Delta z}}{\partial v(\mathbf{x})}\right]}{\left[\frac{\partial \dot{f}_{\Delta z}}{\partial \Delta z}\right]}, \\ &= \frac{1}{E}\int dt \dot{p}_{z-\Delta z}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} \frac{\partial p_z(\mathbf{x}_r, t; \mathbf{x}_s)_{syn}}{\partial v(\mathbf{x})},\end{aligned} \tag{6.6}$$

where

$$E = \int dt \ddot{p}_{z-\Delta z}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} p_z(\mathbf{x}_r, t; \mathbf{x}_s)_{syn}. \tag{6.7}$$

The Fréchet derivative $\partial p_z(\mathbf{x}_r, t; \mathbf{x}_s)/\partial v(\mathbf{x})$ is derived in the next section.

Fréchet Derivative

The first-order acoustic wave-equation can be written as

$$\frac{\partial p}{\partial t} + \rho v^2 \nabla \cdot \mathbf{v} = S(\mathbf{x}_s, t), \tag{6.8}$$
$$\frac{1}{\rho}\nabla p + \frac{\partial \mathbf{v}}{\partial t} = 0,$$

where p represents the pressure, $\mathbf{v}$ represents the particle velocity, and ρ and v indicate the density and velocity, respectively. $S(\mathbf{x}_s, t)$ denotes a source excited at location $\mathbf{x}_s$ and at the excitation time 0 and the listening time is t. To derive the formula for the Fréchet derivative of the pressure field with respect to the perturbation in velocity $v(\mathbf{x})$, we linearize the wave equation in equation 8. A perturbation of $v(\mathbf{x}) \rightarrow v(\mathbf{x}) + \delta v(\mathbf{x})$ will produce a perturbation in pressure $p(\mathbf{x}) \rightarrow p(\mathbf{x}) + \delta p(\mathbf{x})$ and particle velocity $\mathbf{v}(\mathbf{x}) \rightarrow \mathbf{v}(\mathbf{x}) + \delta \mathbf{v}(\mathbf{x})$, which satisfy the linearized acoustic equation given by

$$\frac{\partial \delta p}{\partial t} + \rho v^2 \nabla \cdot \delta \mathbf{v} = -2\rho v \delta v \nabla \cdot \mathbf{v}, \tag{6.9}$$
$$\frac{1}{\rho}\nabla \delta p + \frac{\partial \delta \mathbf{v}}{\partial t} = 0.$$

Using the Green's function $g_p(\mathbf{x}_r, t; \mathbf{x}, 0)$, the solution of equation 9 can be written as

$$\delta p(\mathbf{x}_r, t; \mathbf{x}_s) = -\big(2\rho v g_p(\mathbf{x}_r, t; \mathbf{x}, 0) * \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s))\delta c(\mathbf{x})\big), \tag{6.10}$$

where $*$ indicates convolution operator in time. Dividing by $\delta v(\mathbf{x})$ on both sides, we get

$$\frac{\delta p(\mathbf{x}_r, t; \mathbf{x}_s)}{\delta v(\mathbf{x})\big)} = -2\rho v g_p(\mathbf{x}_r, t; \mathbf{x}, 0) * \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s). \tag{6.11}$$

Substituting equation 6.11 into equation 6 we get

$$\frac{\partial \Delta z}{\partial v(\mathbf{x})} = -\frac{1}{E}\int dt(2\rho v g_p(\mathbf{x}_r, t; \mathbf{x}, 0) * \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s)) \times \dot{p}_{z-\Delta z}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs}. \quad (6.12)$$

Substituting equation 6.12 into equation 6.5, the gradient of $\gamma(\mathbf{x})$ can be expressed as

$$\begin{aligned}
\gamma(\mathbf{x}) &= -\sum_s \sum_r \frac{\partial \Delta z}{\partial v(\mathbf{x})} \Delta z(\mathbf{x}_r, \mathbf{x}_s), \quad (6.13)\\
&= \sum_s \sum_r \frac{1}{E}\int dt(2\rho v g_p(\mathbf{x}_r, t; \mathbf{x}, 0) * \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s)) \times \dot{p}_{z-\Delta z}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} \Delta z(\mathbf{x}_r, \mathbf{x}_s),\\
&= \sum_s \sum_r \frac{1}{E}\int dt(2\rho v g_p(\mathbf{x}_r, t; \mathbf{x}, 0) * \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s)) \times \Delta p_z(\mathbf{x}_r, t; \mathbf{x}_s),
\end{aligned}$$

where $\Delta p_z(\mathbf{x}_r, t; \mathbf{x}_s) = \dot{p}_{z-\Delta z}(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} \Delta z(\mathbf{x}_r, \mathbf{x}_s)$ denotes the data residual which is obtained by weighting the derivative of the observed trace with respect to the latent variable z. Then the difference of observed and predicted encoded values Δz are scaled by a factor of E. Using the identity

$$\int dt[f(t) * g(t)]h(t) = \int dt g(t)[f(-t) * h(t)], \quad (6.14)$$

equation 13 can be rewritten as

$$\begin{aligned}
\gamma(\mathbf{x}) &= -2\rho v \sum_s \sum_r \int dt \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s)\big(g_p(\mathbf{x}_r, -t; \mathbf{x}, 0) * \Delta p_z(\mathbf{x}_r, t; \mathbf{x}_s)\big), \quad (6.15)\\
&= -2\rho v \sum_s \int dt \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s) \sum_r \big(g_p(\mathbf{x}_r, -t; \mathbf{x}, 0) * \Delta p_z(\mathbf{x}_r, t; \mathbf{x}_s)\big),\\
&= -2\rho v \sum_s \int dt \nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s) q(\mathbf{x}, t; \mathbf{x}_s),
\end{aligned}$$

where q is the adjoint-state variables of p (Plessix, 2006). Equation 15 is the gradient of the skeletonized data which can be numerically calculated by a zero-lag crosscor-

relation of a forward-wavefield $\nabla \cdot \mathbf{v}(\mathbf{x}, t; \mathbf{x}_s)$ with the backward-propagated wavefield $q(\mathbf{x}, t; \mathbf{x}_s)$. The velocity model is updated by the steepest gradient descent method

$$v(\mathbf{x})_{k+1} = v(\mathbf{x})_k + \alpha_k \gamma(\mathbf{x})_k, \tag{6.16}$$

where k indicates the iteration number and α_k represents the step length.

6.3 NUMERICAL TEST

The effectiveness of wave equation inversion of data skeletonized by the autoencoder method is now demonstrated with two synthetic data and with crosswell data collected by Exxon in Texas (Chen et al., 1990). The synthetic data are generated for a crosswell acquisition system using a 2-8 finite-difference solutions to the acoustic wave equation.

6.3.1 Crosswell Layer Model

Skeletonized inversion with the autoencoder is now tested on a layered model and a crosswell acquisition geometry. Figure 6.9a shows the true velocity model which has three high-velocity horizontal layers and a linear increasing background velocity. A Ricker wavelet with a peak frequency of 15 Hz is used as the source wavelet. A fixed-spread crosswell acquisition geometry is deployed where 99 shots at a source interval of 20 m are evenly distributed along a vertical well located at x = 10. The data are recorded by 200 receivers for each shot, where the receivers are uniformly distributed every 10 m in depth along with a receiver well located 1000 m away from the source well. The simulation time of the seismic data is 2 s with a time interval of 1 ms.

The training set includes 4000 observed seismic traces because every five shots we take one shot gather as part of the training data. After data processing, we feed the training data into the autoencoder network shown in Figure 6.10. The number

below each layer indicates the dimension of that layer. The boxes with pink, green and blue colors represent the encoder network, latent space and decoder network, respectively. The autoencoder network is trained with mini-batches of 50 traces. We use Tanh activation function instead of ReLU because the input data have both positive and negative parts. The whole training progress only takes several minutes on a workstation with 56 cores and 1 GPU.

After the autoencoder neural network is well trained, we can simply input the synthetic traces generated at each iteration of the inversion to get their encoded values. Therefore the skeletonized misfit and gradient functions can be calculated in order to update the velocity model. Figure 6.9b shows a linear increasing initial model and the inverted result is in Figure 6.11a, which successfully recovers the three high-velocity layers. To further check the correctness of the inverted result, we compared the vertical velocity profiles between the initial, true and inverted velocity model at x = 0.4 km and x = 0.6 km. The blue, red and black lines in Figures 6.11b and 6.12c represent the velocity profiles from the initial, true and inverted velocity models, respectively. Figure 6.12 shows the normalized data residual plotted against the iteration number, which clearly shows a fast convergence to the global minimum.

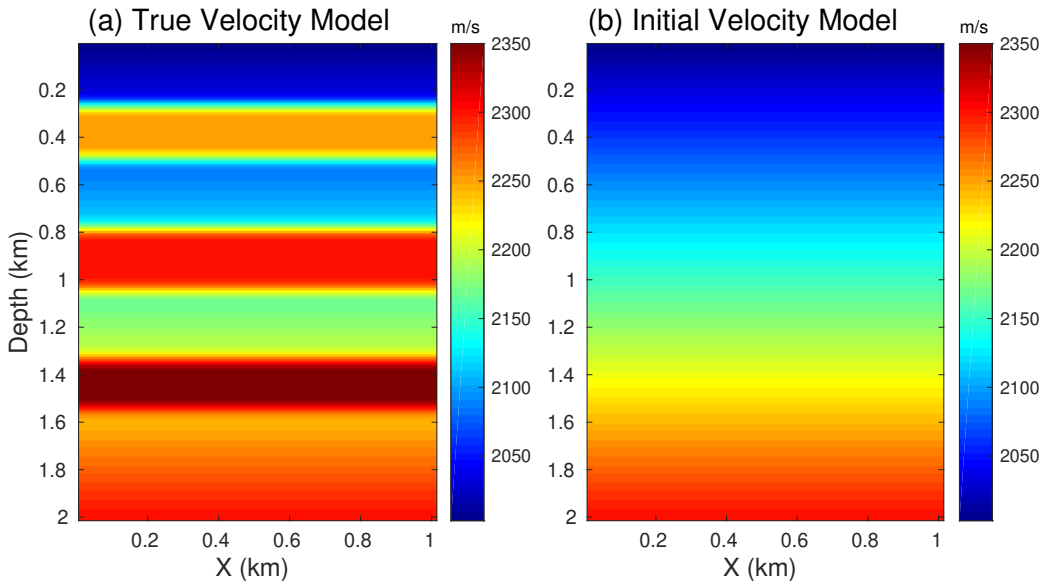


Figure 6.9: The (a) true velocity and (b) linear increasing initial models.

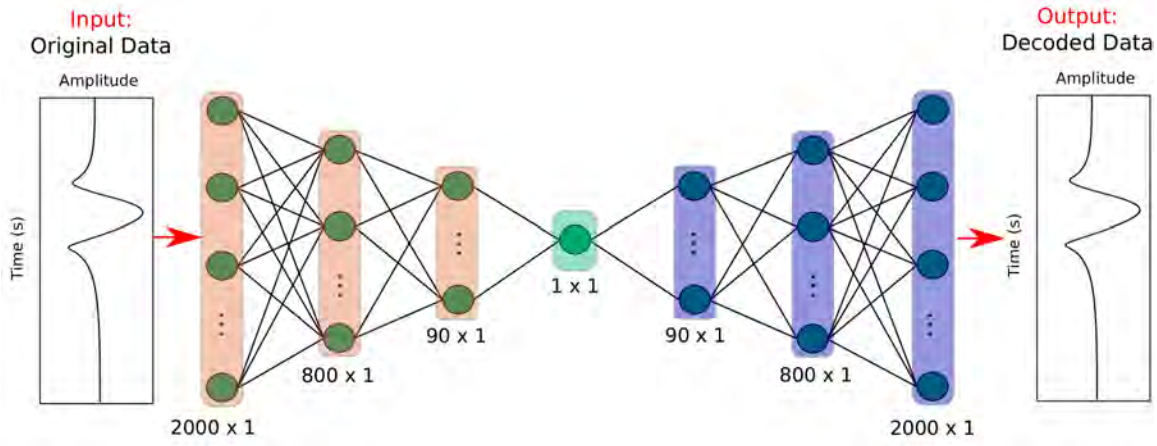


Figure 6.10: The architecture of the autoencoder neural network.

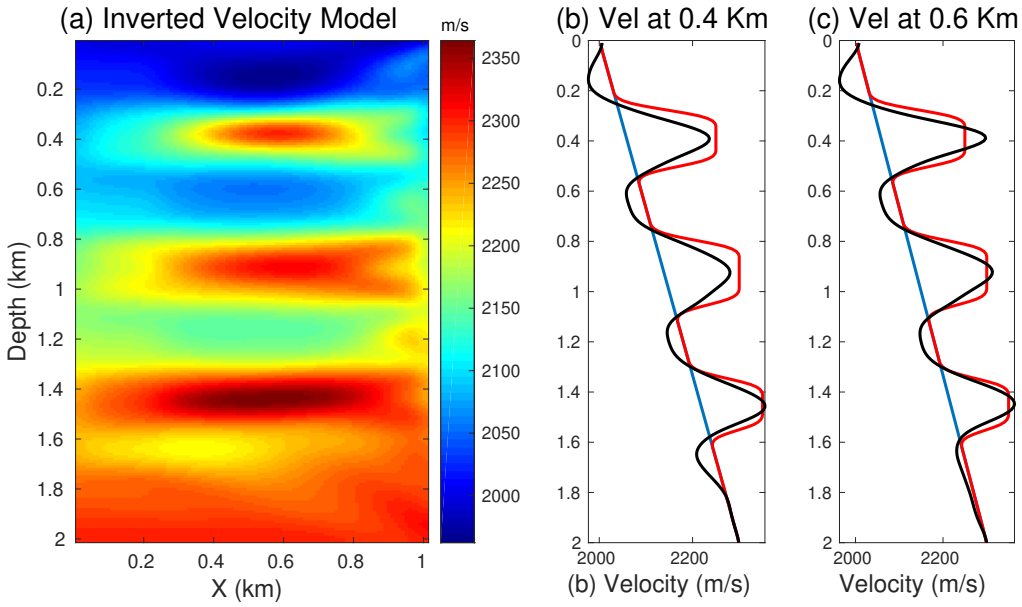


Figure 6.11: The (a) inverted velocity model and the comparison of the vertical velocity profiles at (b) x = 0.4 km and x = 0.6 km. The blue, red and black curve indicate the velocity profiles of the initial, true and inverted velocity model, respectively.

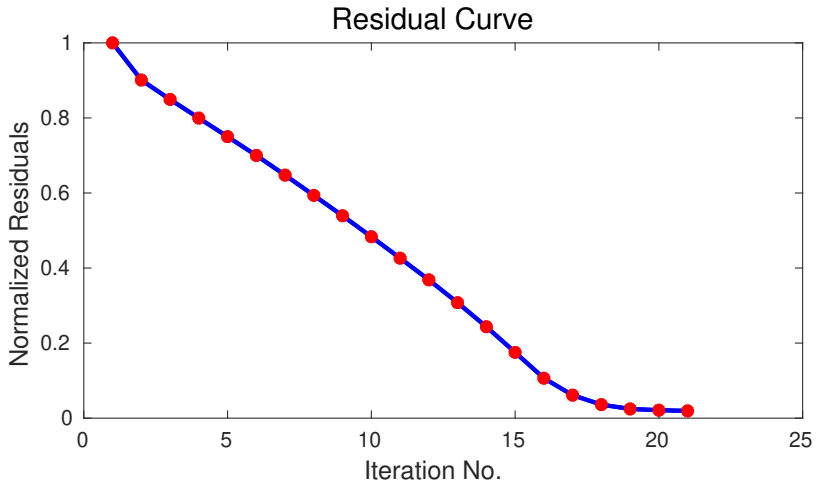


Figure 6.12: The normalized data residual versus iteration numbers.

6.3.2 Crosswell Marmousi Model

Data computed from a part of the Marmousi model are used to test the skeletonized inversion method with the autoencoder method. We select the upper-right region of the Marmousi model shown in Figure 6.13a with 157 x 135 grid points. The

finite-difference method is used to compute 77 acoustic shot gathers with 20 m source intervals along the depth of the well located at 10 m. Each shot contains 156 receivers that are evenly distributed at a spacing of 10 m along the vertical receiver well, which is located 1340 m away from the vertical source well. The data simulation time is 2 s with a time interval of 1 ms. The source wavelet is a 15 Hz Ricker wavelet and the initial model is shown in Figure 6.13b. Here we use the same autoencoder architecture and training strategy as was used in the previous numerical example. The inverted velocity model is shown in Figure 6.14a and the comparison of their vertical profiles at x = 0.5 and x = 0.8 are shown in Figure 6.14b and 6.14c, respectively. The blue, red and black curves represent the velocity profile of the initial, true and inverted velocity model, respectively. It shows that the inverted model is only able to reconstruct the low-wavenumber information in the true velocity model. To get a high-resolution inversion result, a hybrid approach such as the skeletonized inversion + full waveform inversion approach can be used (Luo and Schuster, 1991a,b). A plan for future research is to include a high-dimensional latent space.

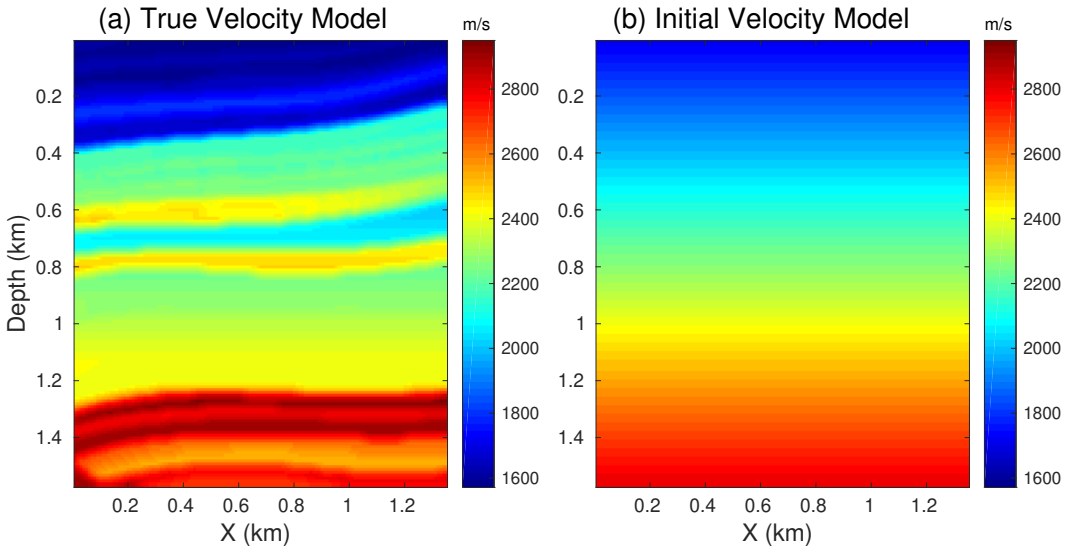


Figure 6.13: The (a) true velocity model and (b) linear increasing initial model.

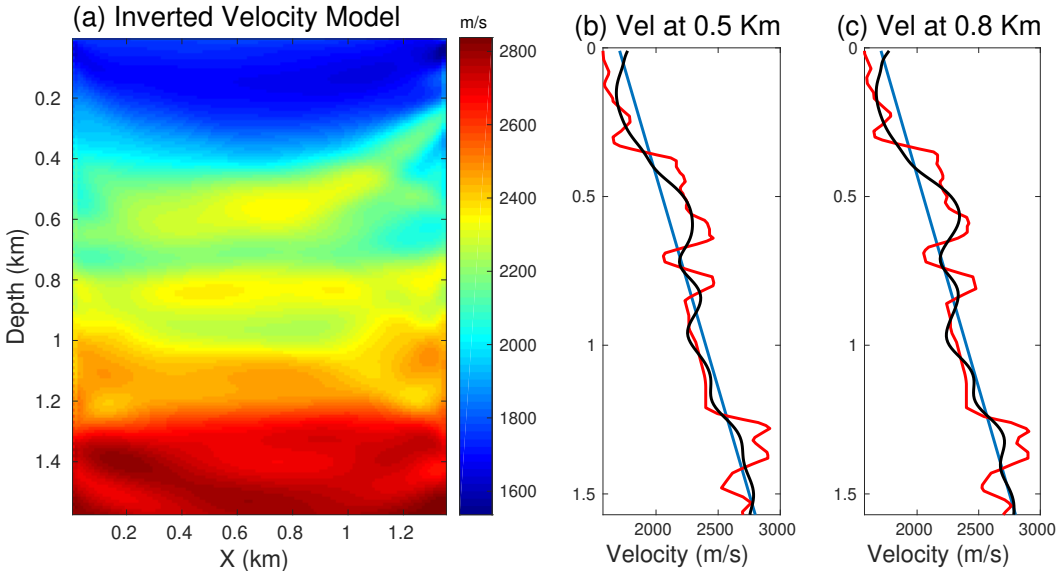


Figure 6.14: The (a) inverted velocity model and the comparison of the vertical velocity profiles at (b) x = 0.5 km and x = 0.8 km. The blue, red and black curves indicate the velocity profiles of the initial, true and inverted velocity model, respectively.

6.3.3 Friendswood Crosswell Field Data

We now test our method on the Friendswood crosswell field data set. Two 305-m-deep cased wells separated by 183 m were used as the source and receiver wells. Downhole explosive charges were fired at intervals of 3 m from 9 m to 305 m in the source well, and the receiver well had 96 receivers placed at depths ranging from 3 m to 293 m. The data are low-pass filtered to 100 Hz with a peak frequency of 58 Hz. The seismic data were recorded with a sampling interval of 0.25 ms for total recording time of 0.375 s. However, we interpolate the data to 0.1 ms time interval for the numerical stable. A processed shot gather is shown in 6.15. Here, we mainly focus the inversion on the transmitted arrivals by windowing the input data around the early arrivals.

The autoencoder architecture we used here is almost the same as the previous two cases, except the dimensions of the input and output layer are changed to 3750×1. Only a portion of the observed data is used for training (every fifth shot gather is used for training). We do not stop the training until the misfit falls below a certain

threshold. A linear increasing velocity model is used as the initial model which is shown in Figure 6.16a. Figure 6.16b shows the inverted velocity model with 15 iterations. Two high-velocity zones at the depth ranges between 85 to 115 m and 170 to 250 m appear in the inverted result. However, there are also some artifacts at the corners of the model that are due to statics and the geometry problems. Figure 6.17a shows the encoded value map of the observed data, where the vertical and horizontal axis represents the source and receivers indexes, respectively. It clearly shows that the near-offset traces have large positive values and the encoded values decrease as the offset increases.

Figure 6.17b and 6.17c show the encoded value map of the seismic data generated from the initial and inverted velocity models, respectively, where the latter one is much more similar to the encoded value map of the observed data. To measure the distance between the true model and the initial model, we plot the values of the encoded misfit function in Figure 6.17d. It shows that there is a relatively larger misfit values at the near-offset traces than at the far offset traces. However, these misfits are largely reduced in the inverted tomogram that is shown in Figure 6.17e. This clearly demonstrates that our inverted tomogram is much closer to the true velocity model compared to the initial model.

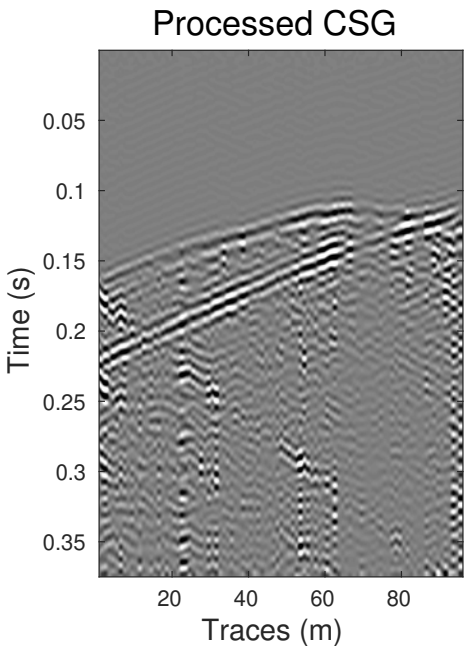


Figure 6.15: A processed shot gather of Friendswoords data.

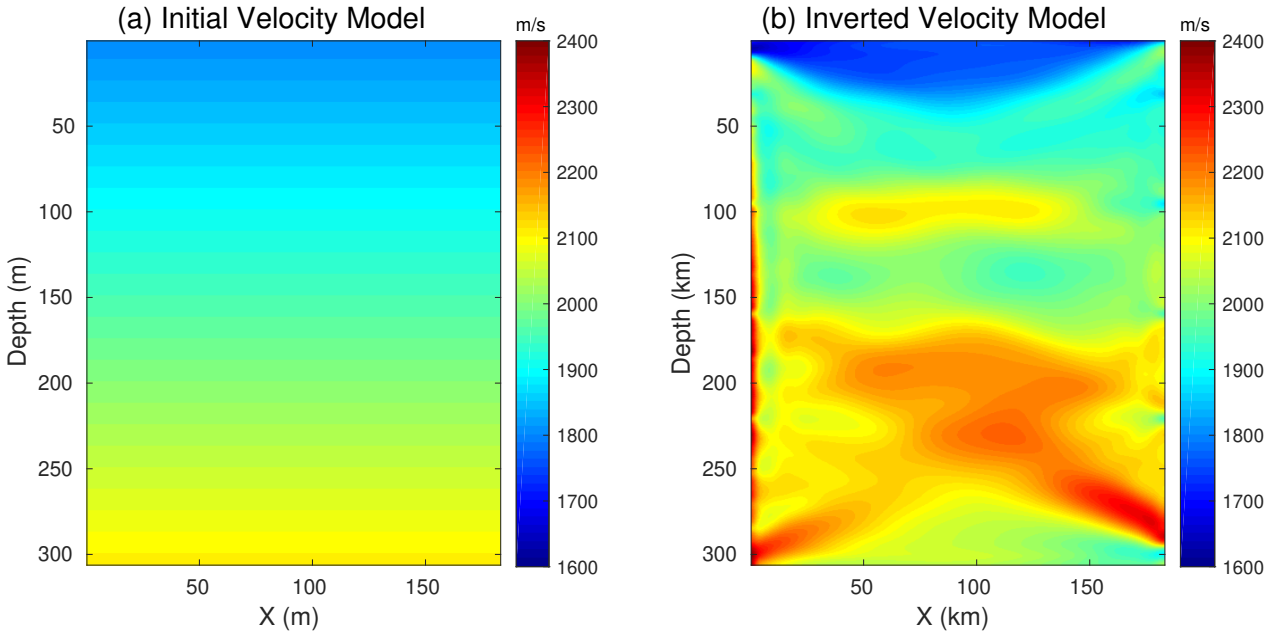


Figure 6.16: The (a) initial linear increasing velocity and (b) inverted velocity models.

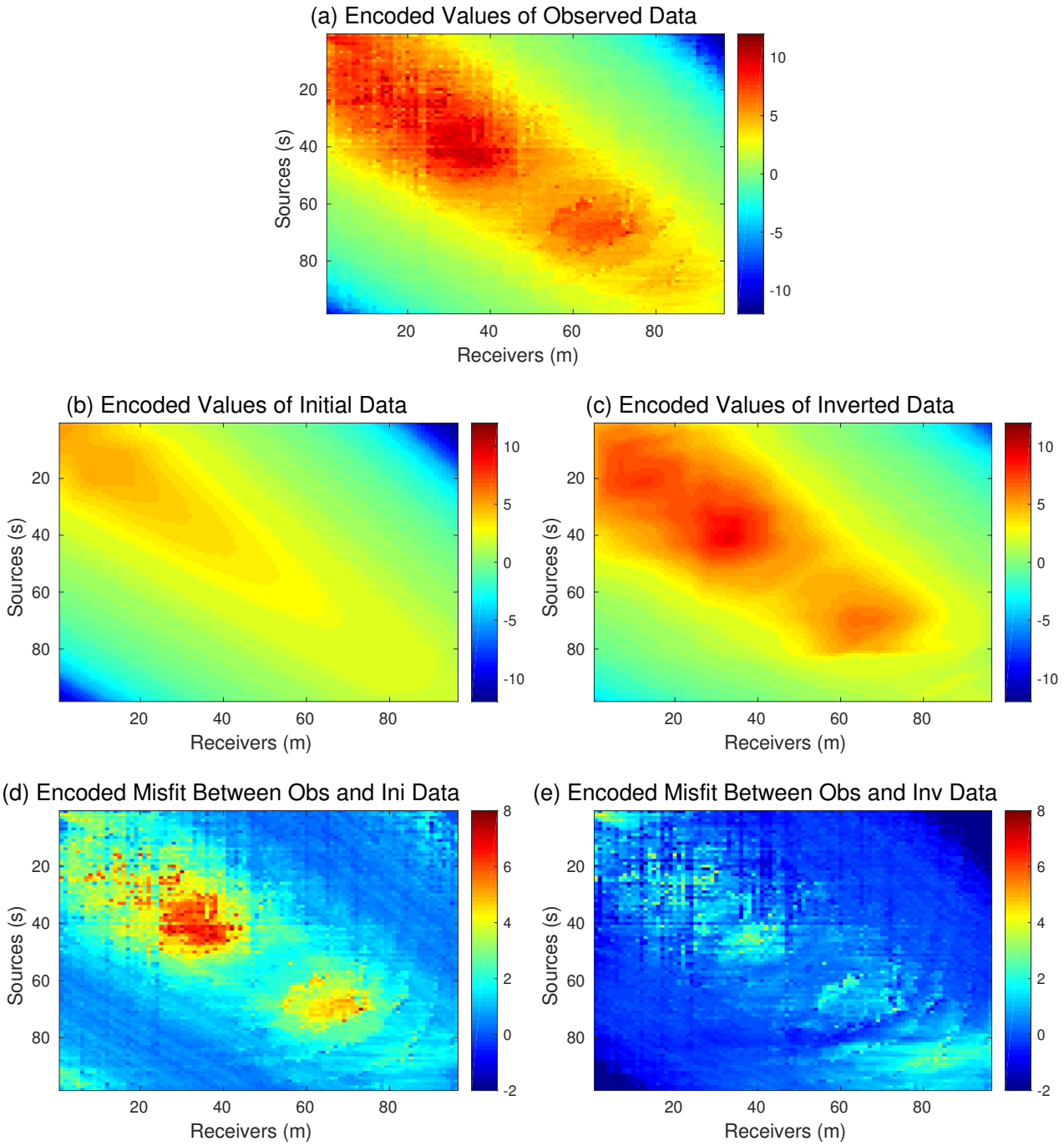


Figure 6.17: The encoded value map of the (a) observed data, the synthetic data generated from the (b) initial model and (c) inverted model. The encoded misfit between the (d) observed data and initial data, (e) observed data and inverted data, respectively.

6.4 DISCUSSION

Tests on both synthetic and observed data demonstrate that the wave equation inversion of seismic data skeletonized by an autoencoder can invert for the low-to-intermediate wavenumber details of the subsurface velocity model. To make this method practical we need to address the method's sensitivity to noisy data.

6.4.1 Noise Sensitivity Tests

In the previous synthetic tests we assumed that the seismic data is noise free. We now repeat the synthetic tests associated with Figure 6.7, except we add random noise to the input data. Different levels of noise are added to both the observed and synthetic data. Figure 6.18a, 6.18d, 6.18g and 6.18j show four shot gathers and their $80th$ traces are displayed in Figure 6.18c, 6.18f, 6.18i and 6.18l. Their encoded results are shown in Figure 6.18b, 6.18e, 6.18h and 6.18k, where the black and red curves represent the encoded values for the observed and synthetic data, respectively. It appears that the range of encoded values decreases as the noise level increases. Moreover, the encoded residual also decreases, which indicates that the encoded values becomes less sensitive to the velocity changes as the data noise level increase.

Figure 6.19 shows the zoomed views of the encoded values in Figure 6.18, where some oscillations appear in the noisy data. These oscillations could further affect the accuracy of the inverted result, especially if the small velocity perturbation are omitted. Therefore, good data quality with less noise is preferred for the autoencoder method in order to recover an accurate subsurface velocity model.

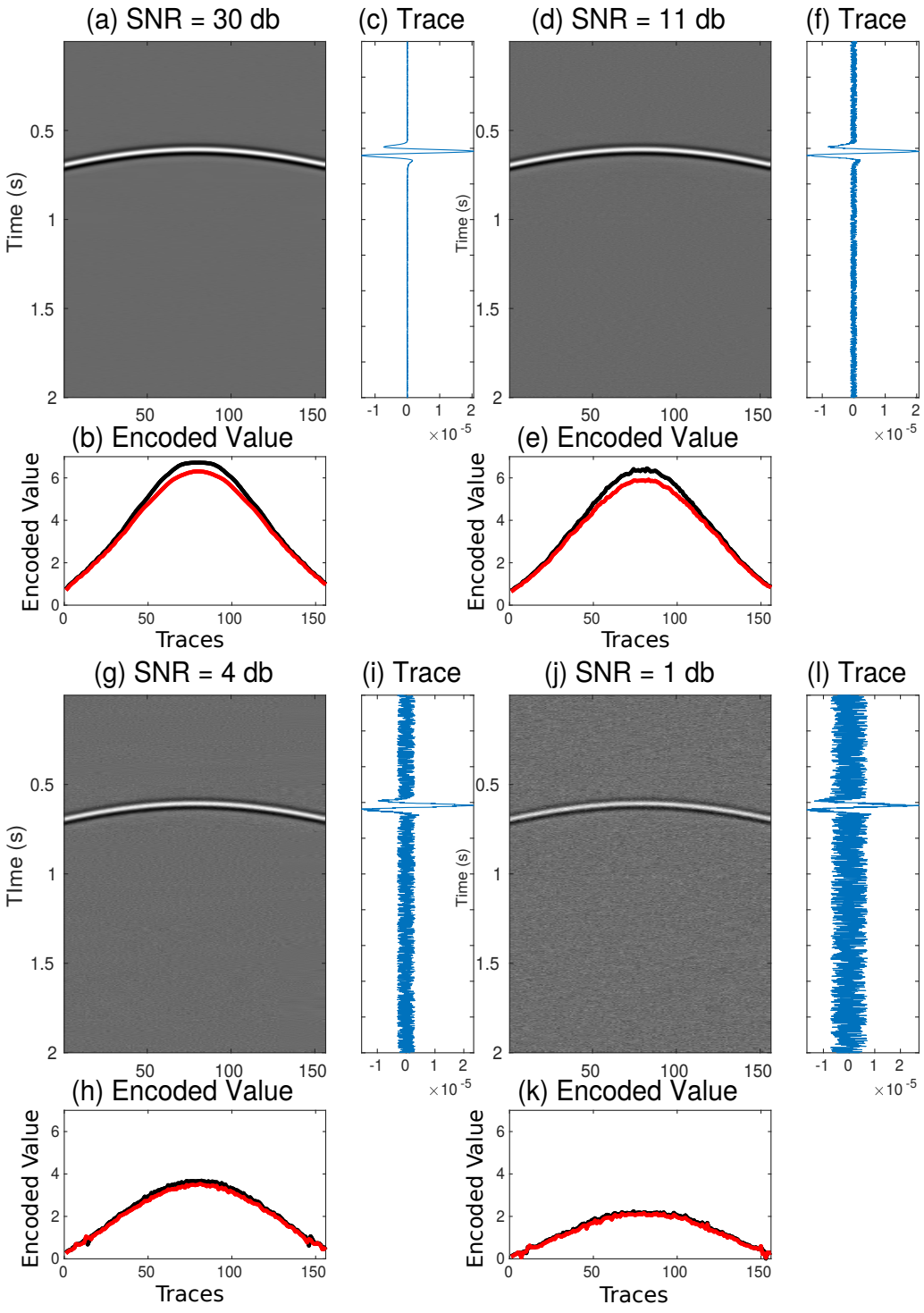


Figure 6.18: The encoded value graphs of the (a) observed data, the synthetic data generated from the (b) initial and (c) inverted models. The encoded misfit between the (d) observed data and initial data, and (e) observed data and inverted data, respectively.

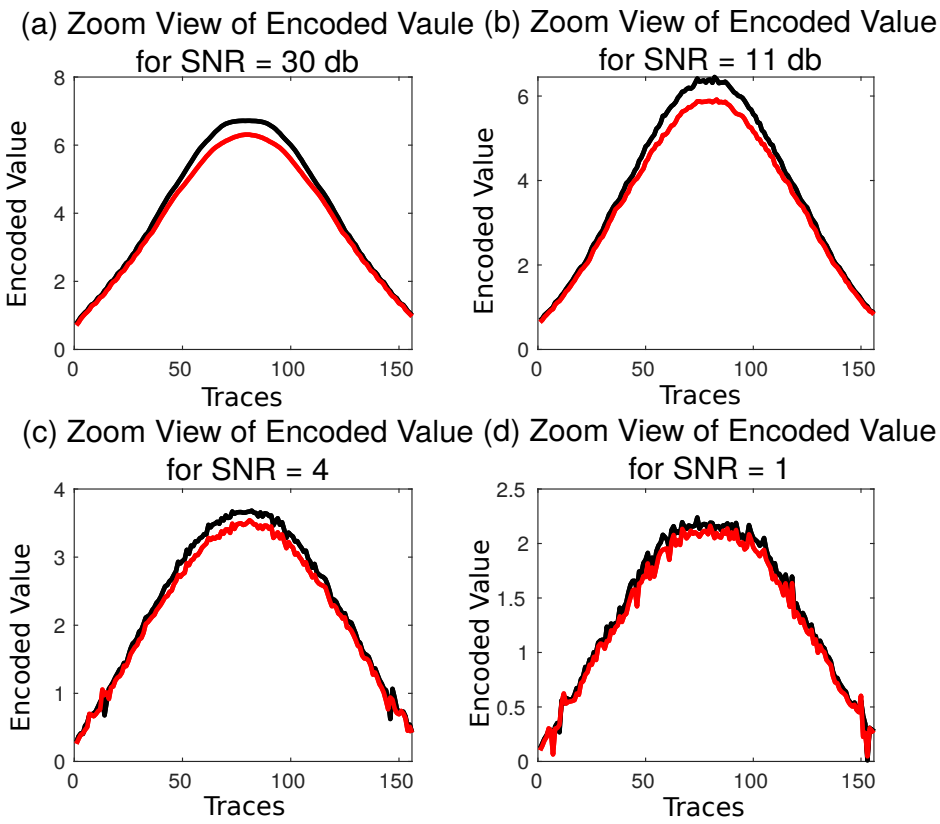


Figure 6.19: The encoded value map of the (a) observed data, the synthetic data generated from the (b) initial model and (c) inverted model. The encoded misfit between the (d) observed data and initial data, (e) observed data and inverted data, respectively.

6.4.2 Overfitting Problem

In our examples, the number of seismic traces in the training set is usually smaller than the number of unknowns in an autoencoder, which might result in an overfitting problem. If the data are overfitted, the network learns the intricacies of the training data set at the expense of its ability to represent unseen examples (e.g., in the test test) (Valentine and Trampert, 2012). In other words, at some point during training, the reconstruction error of the training set keeps decreasing while the reconstruction error of the testing set is either stable or becomes worse. Figures 6.20a and 6.20b show the reconstruction errors of the training set and testing set versus the iteration

number, respectively. It clearly shows that the reconstruction errors of both datasets decrease rapidly within the first 10 iterations, and then gradually become stable. A similar pattern with our training and testing sets demonstrates that we do not suffer from the overfitting problem during training.

Figure 6.20: The reconstruction error of the (a) training set and (b) testing set versus the iteration number.

6.4.3 The Connection Between the Encoded Value and Decoded Waveform

An ideal autoencoder neural network seeks to identify the common characteristic of all training examples in the training set and encapsulates these within the encoder and decoder functions. The information contained in the latent space only that necessary to distinguish between individual examples in the dataset (Valentine and Trampert, 2012). To illustrate this point, we perturb the encoded values in the latent space and see how the decoded waveform changes. Figure 6.21 shows the encoded value changes versus the decoded waveform changes. It clearly shows that the changes in the encoded value result in the waveform shift in time, but the shape of the waveform barely changes. Therefore, in this case, the latent space information is related to the traveltimes which are necessary to distinguish different examples in the dataset.

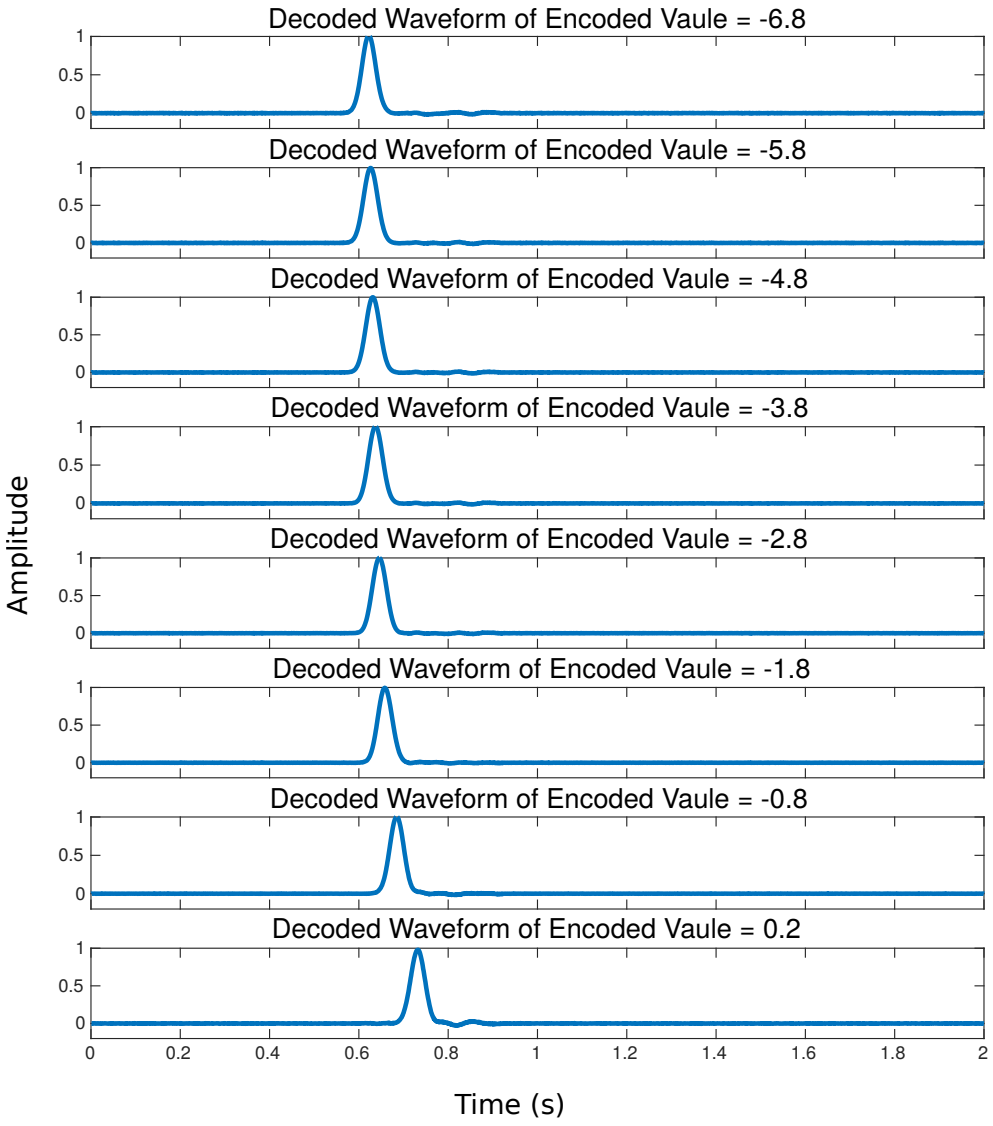


Figure 6.21: The decoded waveform changes with the increasing of encoded values.

6.4.4 Data Selection by Autoencoder

A well-trained autoencoder neural network can be used to separate waveforms that share the common characteristic from those that do not (Valentine and Trampert, 2012). We randomly select some examples from the testing set and add noise to them. We then input the testing set into a well-trained autoencoder. Figures 6.22a,

6.22b and 6.22c compare the noise-free input data examples and their corresponding decoded waveforms, which are indicated by the blue and red curves, respectively. It clearly shows that the reconstructed waveforms closely match with the input waveform. This can be further demonstrated by their reconstruction error which is very small.

Figures 6.22d, 6.22e and 6.22f show the comparison results between the noisy input data and the decoded waveform, which has a large reconstruction error. Therefore we can use the autoencoder to automatically select and remove the "bad" data examples by setting a threshold value for the reconstruction error. Figure 6.23 shows the relationship between the percentage of removed "bad" and "good" data examples with respect to the reconstruction error threshold, which represented by the black and red curves, respectively. It shows that almost 100% of the "bad" data examples are removed successfully and no "good" data examples are removed when the thresholding value is smaller than 2. However, with an increasing threshold value, fewer "bad" data examples are removed.

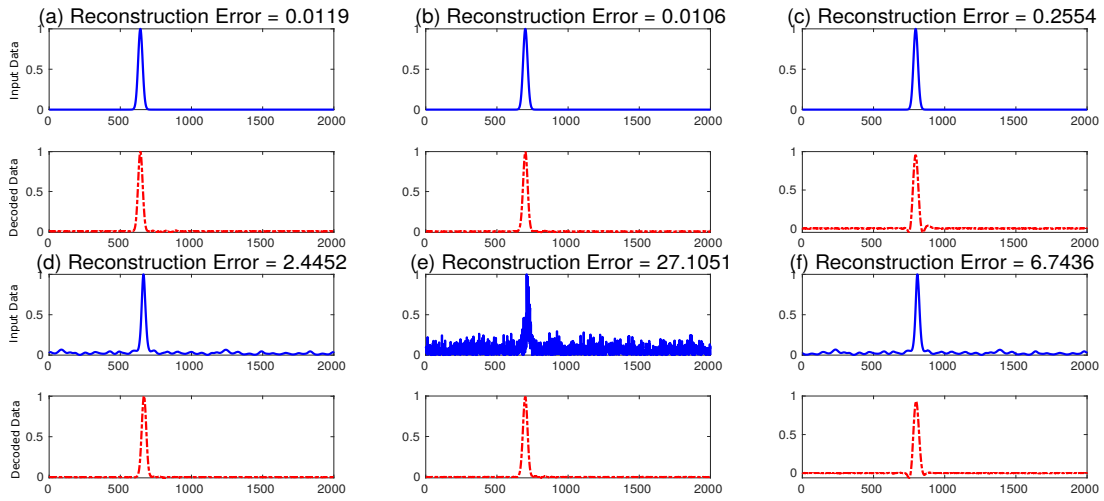


Figure 6.22: Comparisons between the input and decoded waveforms, which are indicated by the blue and red curves, respectively.

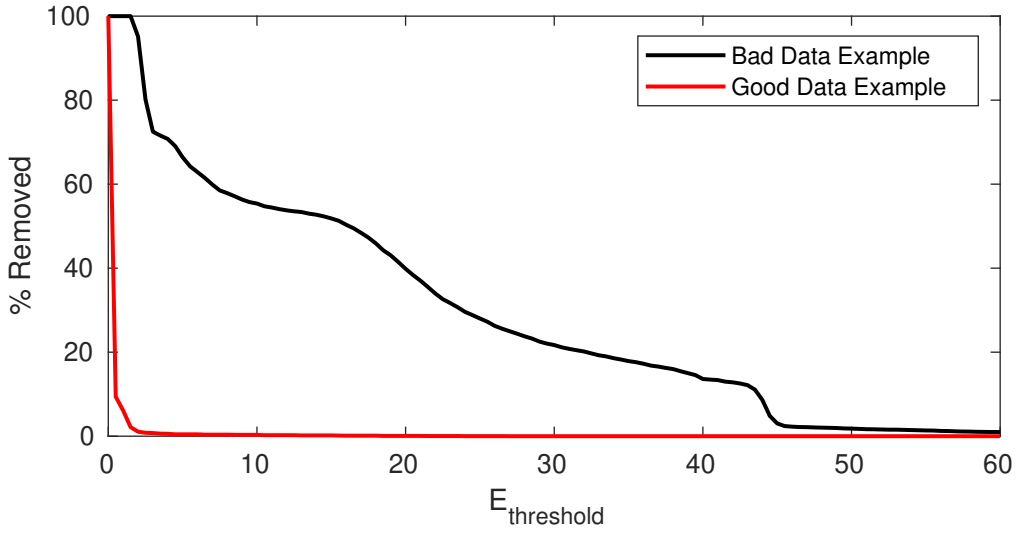


Figure 6.23: The encoded value changes versus the decoded waveform changes.

6.5 CONCLUSIONS

We introduce a wave equation method that finds the velocity model that minimizes the misfit function associated with the skeletonized data in the autoencoder's latent space. The autoencoder can compress a high-dimension seismic trace to a smaller dimension which best represents the original data in the latent space. In this case, measuring the encoded misift between the observed and synthetic data largely reduces the nonlinearity when compared with measuring their waveform differences. Therefore the inverted result will be less prone to getting stuck in a local minimum. The implicit function theorem is used to connect the perturbation of the encoded value with the velocity perturbation in order to calculate the gradient. Numerical results with both synthetic and field data demonstrate that skeletonized inversion with the autoencoder network can accurately estimate the background velocity model. The inverted result can be used as a good initial model for full waveform inversion.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, I develop novel seismic imaging and inversion method to improve image quality and estimate the low-wavenumber components of the subsurface velocity. The main results and conclusions of my thesis are summarized below:

7.1.1 Q-LSRTM with Viscoacoustic Deblurring Filter

In chapter 2, I introduce a preconditioned Q-LSRTM that uses viscoacoustic deblurring filters (DF) to compensate for the amplitude and resolution losses due to strong subsurface attenuation. The viscoacoustic DF can improve the image quality and accelerate the convergence rates of Q-LSRTM by approximating the true Hessian inverse in the viscoacoustic medium. Numerical tests on the synthetic and field data empirically validate that the proposed preconditioning method mitigates the problem of low-resolution associated with standard Q-LSRTM due to the attenuative property of adjoint Q propagator. The limitation of this method is that the localized deblurring filter will not reduce artifacts associated with strong migration artifacts far away from the scattering point. Moreover, a fairly accurate estimation of the background velocity and Q model is required to see noticeable improvements in the image quality with the preconditioned Q-LSRTM method.

7.1.2 Acoustic RTM with Hybrid Deblurring Filtering

In chapter 3, I migrate the viscoacoustic data using acoustic RTM instead of Q-LSRTM and compensate for the amplitude losses and phase shifts using a hybrid deblurring filter (DF). This method successfully avoids the usage of the attenuative adjoint Q propagator and replaces it with the adjoint acoustic propagator, and then corrects the attenuation effects by the hybrid DFs. Numerical tests on synthetic and field data show that the proposed method increases the spatial resolution in the migration image and can reduce the computational costs and storage requirements by a factor of $O(N-1)$ times compared to iterative Q-LSRTM. Here, N is the number of iterations in the least squares migration scheme.

The limitations of this method are that the hybrid DFs are more suitable for correcting high-to-intermediate wavenumber distortion errors in a migration image caused by inadequate compensation for attenuation distortion, but for the low-wavenumber case, it doesn't work well. Similar to viscoacoustic DFs, a fairly accurate background velocity and Q model is required to achieve a desired image quality uplift.

7.1.3 Reduce Migration Artifacts by SVM Filtering

In chapter 4, I develop a support vector machine-based method for suppressing migration artifacts resulting from aliased/sparse data. The SVM filtering method employs three skeletal features from selected dip-angle angle domain common image gather (ADCIG) to distinguish the signal points from artifact points. For each migration image, only a few dip-angle ADCIGs are needed for training and the trained model is used to predict the SVM weights for all dip-angle ADCIGs. Only 1% to 10% of the points in the selected dip-angle ADCIGs are used for training, which is cheap in computation. The numerical results on both synthetic and field data show that the SVM filtering method can effectively remove both random and coherent noises. The additional computational cost of SVM filtering is trivial compared with preconditioned

LSRTM.

7.1.4 Multiscale Reflection Phase Inversion with Deblurring Filter

In chapter 5, I use multiscale reflection phase inversion (MRPI) with a deblurring filter (DF) to invert for the low-wavenumber components of the subsurface velocity model. The benefit is that MRPI is less prone to getting stuck in a local minimum compared to conventional RFWI. To reduce the computation costs, DF is used as an inexpensive alternative to LSRTM, which can produce a migration image with image quality comparable to LSRTM without iterations. To alleviate cycle-skipping, temporally integration is firstly used to boost the low-frequency information of seismic traces. The amplitude replacement procedure is then applied to eliminate the need to explain the magnitude spectra in the recorded traces. Finally, the rolling offset strategy is used to eliminate cycle-skipping in the traces before migration and inversion. The combination of these strategies largely reduces the tendency of MRPI getting stuck in a local minimum.

The disadvantage of this method is that the amplitude replacement method is not able to mitigate the phase mismatch caused by not accounting for attenuation or anisotropy, which might lead to overestimation of the velocity model. To mitigate this problem, more realistic physics should be introduced to achieve a multi-parameter reflection phase inversion.

7.1.5 Seismic Inversion by Newtonian Machine Learning

In chapter 6, I develop a novel skeletonized inversion method that estimates the velocity model that minimizes the misfit function associated with the skeletonized data in the autoencoder's latent space. The autoencoder can compress a high-dimensional seismic trace into a low-dimensional space which best represents the original data. In

this case, measuring the differences between the encoded observed and synthetic data largely reduces the non-linearity with respect to the velocity model when compared with measuring their waveform differences. Therefore, the inverted tomogram is less prone to getting stuck in a local minimum. The most significant contribution of this method is that it provides a general framework for using solutions to the governing PDE to invert skeletal data generated by any type of neural network. The governing equation can be that for gravity, seismic waves, thermal fields, electromagnetic fields, and magnetic fields.

7.2 Future Work

I now list some future research directions from my thesis.

7.2.1 Wave-equation Multiscale Skeletonized Inversion by an Autoencoder

In chapter 6, I compress the high-dimensional seismic trace into one-dimension and invert these encoded values for the subsurface velocity model. However, the inverted result has less resolution compared to the FWI result. To mitigate this problem, I propose a multiscale skeletonized inversion approach where the dimension of the latent space is gradually increased and more complexity is included in the input data. In this case, the resolution of the inverted tomogram will be gradually increased as the dimension of the latent space increases. However, every time the dimension of the latent space increases, the autoencoder needs to be trained again which will increase the computation costs.

7.2.2 Wave-equation Skeletonized Reflection Inversion by an Autoencoder

In chapter 6, I applied the skeletonized inversion method to cross-well data. However, for the surface acquisition system, the early arrivals are limited in their penetration depths. I propose to use a well-trained autoencoder neural network to estimate the distance between the observed and synthetic reflection events in the latent space, and then smear the residuals of the encoded value along the reflection wavepaths to update the deep velocity model. In this case, a convolution neural network (CNN) might identify the reflection events before inputting the data into the autoencoder.

PAPERS PUBLISHED AND SUBMITTED

Journal Papers

• Chen, Y., and G. T. Schuster, 2019, Seismic inversion by Newtonian machine learning, arXiv:1904.10936

• Chen, Y., Z. Feng, L. Fu, A. AlTheyab, S. Feng, and G. T. Schuster, 2019, Multiscale reflection phase inversion with migration deconvolution: Geophysics (accepted subject to minor revision)

• Chen, Y., B. Guo, and G. T. Schuster, 2019, Migration of viscoacoustic data using acoustic reverse time migration with hybrid deblurring filters: Geophysics, 84, S127-S136.

• Chen, Y., G. Dutta, W. Dai, and G. T. Schuster, 2017, Q-least-squares reverse time migration with viscoacoustic deblurring filters: Geophysics, 82, no. 6, S425-S438.

Abstracts

• Chen, Y., G. Dutta, and G. T. Schuster, 2018, Image-domain Q inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, 4121-4125

• Chen, Y., 2018, Automatic Semblance Picking by a Bottom-up Clustering Method: Maximizing Asset Value Through Artificial Intelligence and Machine Learning Workshop, SEG, Expanded Abstracts, 44-48

• Chen, Y., and G. Schuster, 2018, Migration of viscoacoustic data using acoustic reverse time migration with hybrid deblurring filters: First EAGE/SBGf Workshop on Least-Squares Migration, doi 10.3997/2214-4609.201803070.

• Chen, Y., G. Dutta, W. Dai, and G. T. Schuster, 2017, Q-least-squares reverse time

migration with viscoacoustic deblurring filters: 87th Annual International Meeting, SEG, Expanded Abstracts, 4417-4421

REFERENCES

Agudo, Ò. C., N. V. da Silva, M. Warner, and J. Morgan, 2018, Acoustic full-waveform inversion in an elastic world: Geophysics, **83**, R257–R271.

Aki, K. and P. G. Richards, 1980, Quantitative Seismology: W. H. Freeman and Sons.

AlTheyab, A. and G. Schuster, 2015, Reflection full-waveform inversion for inaccurate starting models: Depth Model Building: Full-waveform inversion, SEG, Expanded Abstracts, 18–22.

Aoki, N. and G. T. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, **74**, WCA83–WCA93.

Asnaashari, A., R. Brossier, C. Castellanos, B. Dupuy, V. Etienne, Y. Gholami, G. Hu, L. Métivier, S. Operto, and D. Pageot, 2012, Hierarchical approach of seismic full waveform inversion: Numerical Analysis and Applications, **5**, 99–108.

Bai, J., D. Yingst, R. Bloor, and J. Leveille, 2014, Viscoacoustic waveform inversion of velocity structures in the time domainviscoacoustic waveform inversion: Geophysics, **79(3)**, R103–R119.

Bi, H. and T. Lin, 2014, Impact of adaptive data selection on full waveform inversion: 1094–1098.

Billette, F. and S. Brandsberg-Dahl, 2005, The 2004 bp velocity benchmark: 67th EAGE Conference & Exhibition.

Blanch, J. O., J. O. Robertsson, and W. W. Symes, 1995, Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique: Geophysics, **60**, 176–184.

Blanch, J. O. and W. W. Symes, 1994, Linear inversion in layered viscoacoustic media using a time-domain method: 64th Annual International Meeting, SEG, Expanded Abstracts, 1053–1056.

———, 1995, Efficient iterative viscoacoustic linearized inversion: 65th Annual International Meeting, SEG, Expanded Abstracts, 627–630.

Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Brossier, R., S. Operto, and J. Virieux, 2015, Velocity model building from seismic reflection data by full-waveform inversion: Geophysical Prospecting, **63**, 354–367.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995a, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

———, 1995b, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Cabrales-Vargas, A. and K. J. Marfurt, 2013, Amplitude-preserving imaging of aliased data using preconditioned kirchhoff least-squares depth migration.: 83th Annual International Meeting, SEG, Expanded Abstracts, 3726–3730.

Carcione, J. M., D. Kosloff, and R. Kosloff, 1988, Wave propagation simulation in a linear viscoacoustic medium: Geophysical Journal International, **93**, 393–401.

Cavalca, M., R. Fletcher, and M. Riedel, 2013, Q-compensation in complex media–ray-based and wavefield extrapolation approaches: 83th Annual International Meeting, SEG, Expanded Abstracts, 3831–3835.

Chen, G.-X., R.-S. Wu, and S.-C. Chen, 2018, Reflection multi-scale envelope inversion: Geophysical Prospecting, **66**, 1258–1271.

Chen, S., L. Zimmerman, and J. Tugnait, 1990, Subsurface imaging using reversed vertical seismic profiling and crosshole tomographic methods: Geophysics, **55**, 1478–1487.

Chen, Y., G. Dutta, W. Dai, and G. T. Schuster, 2017, Q-least-squares reverse time migration with viscoacoustic deblurring filters: Geophysics, **82(6)**, S425–S438.

Chen, Y., B. Guo, and G. T. Schuster, 2019, Migration of viscoacoustic data using acoustic reverse time migration with hybrid deblurring filters: Geophysics, **84**, S127–S136.

Cheng*, X., K. Jiao, D. Sun, and D. Vigh, 2015, A new approach of visco-acoustic waveform inversion in the time domain: 85th Annual International Meeting, SEG, Expanded Abstracts, 1183–1187.

Chi, B., L. Dong, and Y. Liu, 2015, Correlation-based reflection full-waveform inversion: Geophysics, **80**, R189–R202.

Christensen, R., 1982, Theory of viscoelasticity: Academic Press.

Cortes, C. and V. Vapnik, 1995, Support-vector networks: Machine Learning, **20**, 273–297.

Dafni, R. and W. W. Symes, 2016a, Kinematic artifacts in the subsurface-offset extended image and their elimination by a dip-domain specularity filter: Geophysics, **81**, S477–S495.

———, 2016b, Scattering and dip angle decomposition based on subsurface offset extended wave-equation migration: Geophysics, **81**, S119–S138.

Dai, N. and G. F. West, 1994, Inverse Q migration: 64th Annual International Meeting, SEG, Expanded Abstracts, 1418–1421.

Dai, W. and G. T. Schuster, 2009, Least-squares migration of simultaneous sources data with a deblurring filter: 79th Annual International Meeting, SEG, Expanded Abstracts, 2990–2994.

Dai, W., X. Wang, and G. T. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: Geophysics, **76**, R135–R146.

Dai, W., Z. Xu, and R. Coates, 2015a, Least-squares reverse-time migration for visco-acoustic media: SEG Technical Program Expanded Abstracts 2015, 3387–3391.

———, 2015b, Least-squares reverse-time migration for visco-acoustic media: 79th Annual International Meeting, SEG, Expanded Abstracts, 3387–3391.

Duan, Y., A. Guitton, and P. Sava, 2017, Elastic least-squares reverse time migration: Geophysics, **82**, S315–S325.

Dutta, G., 2016, Skeletonized wave-equation inversion for Q: 3618–3623.

———, 2017, Sparse least-squares reverse time migration using seislets: Journal of Applied Geophysics, **136**, 142–155.

Dutta, G., M. Giboli, C. Agut, P. Williamson, and G. T. Schuster, 2017, Least-squares reverse time migration with local radon-based preconditioning: Geophysics, **82**, S75–S84.

Dutta, G. and G. T. Schuster, 2014a, Attenuation compensation for least-squares reverse time migration using the viscoacoustic-wave equation: Geophysics, **79**, S251–S262.

———, 2014b, Attenuation compensation for least-squares reverse time migration using the viscoacoustic-wave equation: Geophysics, **79**, S251–S262.

———, 2015, Sparse least-squares reverse time migration using seislets: 85th Annual International Meeting, SEG, Expanded Abstracts, 4232–4237.

———, 2016, Wave-equation q tomography: Geophysics, **81**, R471–R484.

Feng, S. and G. T. Schuster, 2016, Anisotropic wave-equation traveltime and waveform inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 1196–1200.

———, 2019, Transmission+ reflection anisotropic wave-equation traveltime and waveform inversion: Geophysical Prospecting, **67**, 423–442.

Fletcher, R., D. Nichols, and M. Cavalca, 2012, Wavepath-consistent effective Q estimation for Q-compensated reverse-time migration: 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, 2214–4069–1487.

Fu, L., B. Guo, Y. Sun, and G. T. Schuster, 2017, Multiscale phase inversion of seismic data: Geophysics, **83**, 1–52.

Gamar, F., D. Carotti*, P. Guillaume, and A. Gacha, 2015, Success of high-resolution

volumetric q-tomography in the automatic detection of gas anomalies on offshore brunei data: 5184–5188.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017–1024.

Guo, Q. and T. Alkhalifah, 2017, Elastic reflection-based waveform inversion with a nonlinear approach: Geophysics, **82**, R309–R321.

Ha, W. and C. Shin, 2012, Laplace-domain full-waveform inversion of seismic data lacking low-frequency information: Geophysics, **77**, R199–R206.

Hotelling, H., 1933, Analysis of a complex of statistical variables into principal components.: Journal of educational psychology, **24**, 417.

Hu, J. and G. T. Schuster, 1998, Migration deconvolution: Proc. SPIE, **3453**, 118–124.

Hu, J., G. T. Schuster, and P. A. Valasek, 2001, Poststack migration deconvolution: Geophysics, **66**, 939–952.

Jansson, P. A. and M. Richardson, 1997, Deconvolution of images and spectra: Optical Engineering, **36**, 3224–3225.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Inverse Scattering–Theory and Application, 467–481.

Li, J., G. Dutta, and G. Schuster, 2017, Wave-equation Qs inversion of skeletonized surface waves: Geophysical Journal International, **209**, 979–991.

Li, J., Z. Feng, and G. Schuster, 2016, Wave-equation dispersion inversion: Geophysical Journal International, **208**, 1567–1578.

Lin, Y., L. Huang, J. Queen, J. Moore, and E. Majer, 2016, Least-squares reverse-time migration with compressive sensing for sparse seismic data: Workshop on Geothermal Engineering, 1–5.

Liu, Q. and J. Zhang, 2018, Efficient dip-angle angle-domain common-image gather estimation using poynting vector in acoustic reverse time migration and its application in noise suppression: Geophysical Prospecting, **66**, 1714–1725.

Liu, Z., J. Li, S. M. Hanafy, and G. Schuster, 2018, 3D wave-equation dispersion inversion of surface waves: 4733–4737.

Lu, K., J. Li, B. Guo, L. Fu, and G. Schuster, 2017, Tutorial for wave-equation inversion of skeletonized data: Interpretation, **5**, SO1–SO10.

Luo, Y. and G. T. Schuster, 1991a, Wave equation inversion of skeletalized geophysical data: Geophysical Journal International, **105**, 289–294.

———, 1991b, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Neidell, N. S. and M. T. Taner, 1971, Semblance and other coherency measures for

multichannel data: Geophysics, **36**, 482–497.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Quan, Y. and J. M. Harris, 1997, Seismic attenuation tomography using the frequency shift method: Geophysics, **62**, 895–905.

Ribodetti, A., J. Virieux, and S. Durand, 1995, Asymptotic theory for viscoacoustic seismic imaging: 65th Annual International Meeting, SEG, Expanded Abstracts, 631–634.

Schmidhuber, J., 2015, Deep learning in neural networks: An overview: Neural networks, **61**, 85–117.

Schuster, G., 2009, Seismic interferometry: Cambridge University Press.

———, 2017, Seismic inversion: SEG Publishing: Tulsa, Oklahoma.

———, 2018, Machine learning and wave equation inversion of skeletonized data: 80th EAGE Conference and Exhibition, WS01.

Schuster, G. T. and J. Hu, 2000, Green's function for migration: Continuous recording geometry: Geophysics, **65**, 167–175.

Shah, N., M. Warner, T. Nangoo, A. Umpleby, I. Stekl, J. Morgan, and L. Guasch, 2012, Quality assured full-waveform inversion: Ensuring starting model adequacy: 82th Annual International Meeting, SEG, Expanded Abstracts, 1–5.

Shen*, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Wave-equation migration Q analysis WEMQA: 84th Annual International Meeting, SEG, Expanded Abstracts, 3757–3762.

Suh, S., K. Yoon, J. Cai, and B. Wang, 2012, Compensating visco-acoustic effects in anisotropic resverse-time migration: 82th Annual International Meeting, SEG, Expanded Abstracts, 1–5.

Sun, J., S. Fomel, T. Zhu, and J. Hu, 2016, Q-compensated least-squares reverse time migration using low-rank one-step wave extrapolation: Geophysics, **81**, S271–S279.

Sun, Y. and G. T. Schuster, 1993, Time-domain phase inversion: 63th Annual International Meeting, SEG, Expanded Abstracts, 684–687.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Tavares, S., 1966, A comparison of integration and low-pass filtering: IEEE Transactions on Instrumentation and Measurement, **15**, 33–38.

Valenciano, A., N. Chemingui, D. Whitmore, and S. B. Dahl, 2011, Wave equation migration with attenuation and anisotropy compensation: 81th Annual International Meeting, SEG, Expanded Abstracts, 232–236.

Valentine, A. P. and J. Trampert, 2012, Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data: Geophysical Journal International, **189**, 1183–1202.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, G., S. Wang, Q. Du, and S. Yuan, 2017, Traveltime-based reflection full-waveform inversion for elastic medium: Journal of Applied Geophysics, **141**, 68–76.

Wang, S., F. Chen, H. Zhang, and Y. Shen, 2013, Reflection-based full waveform inversion (rfwi) in the frequency domain: 83th Annual International Meeting, SEG, Expanded Abstracts, 877–881.

Wang, Y., 2007, Inverse-Q filtered migration: Geophysics, **73**, S1–S6.

Wikipedia, 2019, Reflection seismology — Wikipedia, the free encyclopedia: `http://en.wikipedia.org/w/index.php?title=Reflection%20seismology&oldid=884431620`. ([Online; accessed 28-April-2019]).

Wold, S., K. Esbensen, and P. Geladi, 1987, Principal component analysis: Chemometrics and intelligent laboratory systems, **2**, 37–52.

Wright, S. and J. Nocedal, 1999, Numerical optimization: Springer Science, **35**, 7.

Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, WA13–WA24.

Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambare, 2012, Full waveform inversion for reflected seismic data: 74th EAGE Conference and Exhibition, W024.

Yu, J. and B. Hornby, 2008, A strategy for attenuating vsp migration artifacts: local beam migration: SEG Annual Meeting, 3385–3389.

Yu, J., J. Hu, G. T. Schuster, and R. Estill, 2006, Prestack migration deconvolution: Geophysics, **71**, S53–S62.

Yu, Y., R. S. Lu, and M. M. Deal, 2002, Compensation for the effects of shallow gas attenuation with viscoacoustic wave-equation migration: SEG Technical Program Expanded Abstracts, **21**, 2062–2065.

Zhang, Y., P. Zhang, and H. Zhang, 2010, Compensating for visco-acoustic effects in reverse-time migration: SEG Expanded Abstracts, 3160–3164.

Zhou, Y.-H., J.-H. Gao, B.-L. Wang, and Y.-Y. He, 2010, Migration scheme for imaging offset vsp data within local phase space: Applied Geophysics, **7**, 31–40.

Zhu, T. and J. M. Harris, 2015, Improved seismic image by q-compensated reverse

time migration: Application to crosswell field data, West Texas: Geophysics, **80**, B61–B67.

Zhu, T., J. M. Harris, and B. Biondi, 2014, Q-compensated reverse-time migration: Geophysics, **79**, S77–S87.

APPENDICES

A Acoustic Modeling and Blurring Operator

Under the Born approximation, the observed data $d(\mathbf{x}_r|\mathbf{x}_s,\omega)$, excited by a point source at $\mathbf{x}_s$ and recorded by a receiver at $\mathbf{x}_r$, can be represented as

$$d(\mathbf{x_r}|\mathbf{x_s},\omega) = \int_{V_0} w(\omega)G(\mathbf{x}_r|\mathbf{x}_0,\omega)G(\mathbf{x}_0|\mathbf{x}_s,\omega)m(\mathbf{x}_0), \tag{A.1}$$

where ω denotes the angular frequency, $w(\omega)$ denotes the spectrum of the second-order time derivative of the source wavelet, $m(\mathbf{x}_0)$ is the reflectivity distribution at the subsurface location $\mathbf{x}_0$ and V_0 is the 3D integration volume. $G(\mathbf{x}^{'}|\mathbf{x},\omega)$ denotes the background Green's function for a source at $\mathbf{x}$ and a receiver at $\mathbf{x}^{'}$. Using a matrix-vector notation, equation A.1 can also be written as

$$\mathbf{d} = \mathbf{L}\mathbf{m}_0, \tag{A.2}$$

where $\mathbf{L}$ represents a linear modeling operator, $\mathbf{d}$ is the data vector, and $\mathbf{m}_0$ is a vector that represents the subsurface reflectivity model. The migration image $\mathbf{m}_{mig}$ is computed by applying the migration operator $\mathbf{L}^T$ to the observed data to give

$$\mathbf{m}_{mig} = \mathbf{L}^T\mathbf{d} = \overbrace{\mathbf{L}^T\mathbf{L}}^{blurring\ operator}\ \mathbf{m}_0, \tag{A.3}$$

where $\mathbf{L^T}$ denotes the adjoint of the forward modeling operator $\mathbf{L}$. Using Green's function notation and equation A.1, equation 3 can also be expressed as

$$m_{mig}(\mathbf{x}) = \int_{V_0} \int_{-\infty}^{\infty} d\omega \sum_s \sum_r |w(\omega)|^2 G^*(\mathbf{x}_r|\mathbf{x},\omega) G^*(\mathbf{x}|\mathbf{x}_s,\omega) G(\mathbf{x}_r|\mathbf{x}_0,\omega) G(\mathbf{x}_0|\mathbf{x}_s,\omega) m(\mathbf{x}_0) dV_0,$$
$$= \int_{V_0} \Gamma(\mathbf{x}|\mathbf{x}_0) m(\mathbf{x}_0) dV_0, \tag{A.4}$$

where the migration Green's function (Schuster and Hu, 2000) is defined as

$$\Gamma(\mathbf{x}|\mathbf{x}_0) = \int_{-\infty}^{\infty} d\omega \sum_s \sum_r w(\omega) w(\omega)^* G^*(\mathbf{x}_r|\mathbf{x},\omega) G^*(\mathbf{x}|\mathbf{x}_s,\omega) G(\mathbf{x}_r|\mathbf{x}_0,\omega) G(\mathbf{x}_0|\mathbf{x}_s,\omega). \tag{A.5}$$

Here, $*$ indicates the complex conjugate and $\Gamma(\mathbf{x}|\mathbf{x}_0)$ denotes the migration Green's function response at $\mathbf{x} = (x, y, z)$ for a point scatterer at $\mathbf{x}_0 = (x_0, y_0, z_0)$.

B Adjoint Equations and Gradient for Q-LSRTM

Based on the SLS model, the equations of motion for a 2D viscoacoustic medium can be written as (Carcione et al., 1988; Blanch and Symes, 1994).

$$\begin{aligned}
&\frac{\partial P}{\partial t} + K(\tau+1)(\nabla\cdot\mathbf{v}) + r_p = S(\mathbf{x}_s, t), \\
&\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{\rho}\nabla P = 0, \\
&\frac{\partial r_p}{\partial t} + \frac{1}{\tau_\sigma}(r_p + \tau K(\nabla\cdot\mathbf{v})) = 0,
\end{aligned} \tag{B.1}$$

where ρ represents the density, P represents the pressure wavefield, K represent the bulk modulus, r represent the memory variables and $\mathbf{v}$ represents the particle velocity vector. Here, τ is related to the quality factor Q, and the reference angular frequency ω, which can be expressed as

$$\tau = \frac{\tau_\epsilon}{\tau_\sigma} - 1 = \frac{2}{Q}\left(\frac{1}{Q} + \sqrt{1 + \frac{1}{Q^2}}\right), \tag{B.2}$$

where τ_σ and τ_ϵ represents the stress and strain relaxation times, respectively. If we perturb K by an amount δK, the perturbed wavefields can be written as

$$\begin{aligned}
&\frac{\partial \delta P}{\partial t} + K(\tau+1)(\nabla\cdot\delta\mathbf{v}) + \delta r_p = -\delta K(\tau+1)(\nabla\cdot\mathbf{v}), \\
&\frac{\partial \delta\mathbf{v}}{\partial t} + \frac{1}{\rho}\nabla\delta P = 0, \\
&\frac{\partial \delta r_p}{\partial t} + \frac{1}{\tau_\sigma}(\delta r_p + \tau K(\nabla\cdot\delta\mathbf{v})) = -\frac{\tau}{\tau_\sigma}\delta K(\nabla\cdot\mathbf{v}).
\end{aligned} \tag{B.3}$$

Using the Green's function $g_P(\mathbf{x}_r, t; \mathbf{x}_0, 0)$ and $g_{r_p}(\mathbf{x}_r, t; \mathbf{x}_0, 0)$, equation B.3 can also be expressed as

$$\begin{aligned}\delta P(\mathbf{x}_r, t; \mathbf{x}_s) = \int_{V_0} -\delta K(\mathbf{x}_0)\Big((\tau(\mathbf{x}_0)+1)\big(g_P(\mathbf{x}_r, t; \mathbf{x}_0, 0) * \nabla\cdot \mathbf{v}(\mathbf{x}_0, t; \mathbf{x}_s)\big)\\ + \frac{\tau(\mathbf{x}_0)}{\tau_\sigma(\mathbf{x}_0)}\big(g_{r_p}(\mathbf{x}_r, t; \mathbf{x}_0, 0) * \nabla\cdot \mathbf{v}(\mathbf{x}_0, t; \mathbf{x}_s)\big)\Big)\mathrm{d}V_0,\end{aligned} \tag{B.4}$$

where K represent the bulk modulus, $\mathbf{v}$ represents the particle velocity vector, τ_σ represents the stress relaxation time, τ is related to the quality factor Q, $g_P(\mathbf{x}_r, t; \mathbf{x}_0, 0)$ and $g_{r_p}(\mathbf{x}_r, t; \mathbf{x}_0, 0)$ are the pressure and memory variable Green's functions, respectively. In the context of Q-LSRTM, the solution to equations B.4 is equivalent to the matrix-vector operation

$$\mathbf{d}_Q = \mathbf{L}_Q \mathbf{m}_0, \tag{B.5}$$

where $\mathbf{d}_Q$ denotes the Born-modeled data with attenuation, $\mathbf{L}_Q$ is a linear viscoacoustic modeling operator and $\mathbf{m}_0$ is related to the reflectivity of the medium. The adjoint equations for equation B.3 can be derived using the adjoint-state method (Lailly, 1983; Blanch and Symes, 1994; Blanch et al., 1995) and is given by

$$\begin{aligned}&\frac{\partial q}{\partial t} + \nabla\cdot\Big(\frac{1}{\rho}\mathbf{u}\Big) = -\Delta d(\mathbf{x}_g, t; \mathbf{x}_s),\\ &\frac{\partial \mathbf{u}}{\partial t} + \Big[\nabla(K(1+\tau)q) + \nabla\Big(\frac{1}{\tau_\sigma}K\tau s\Big)\Big] = 0,\\ &\frac{\partial s}{\partial t} - \frac{s}{\tau_\sigma} - q = 0.\end{aligned} \tag{B.6}$$

Here $(q, \mathbf{u}, s)$ are the adjoint-state variables of the state variables $(P, \mathbf{v}, r_p)$ and Δd represents the data residual for the predicted and the observed pressure data at every

iteration.

The perturbation in the image, δm, is related to the perturbation in the bulk modulus, δK, which in turn can be obtained by zero-lag cross-correlation of the adjoint fields with the background wavefields from equation B.3 as

$$\delta m \approx \delta K = \int_0^T (1+\tau)(\nabla\cdot\mathbf{v})q + \frac{\tau}{\tau_\sigma}(\nabla\cdot\mathbf{v})s\,dt, \tag{B.7}$$

Using Green's function notation, equation B.7 can also be expressed as

$$\begin{aligned}\delta m \approx \delta K &= \sum_r\sum_s \int_0^T dt\Big(\big(\tau(\mathbf{x})+1\big)\nabla\cdot\mathbf{v}(\mathbf{x},t;\mathbf{x}_s)\big(g_P(\mathbf{x}_r,-t;\mathbf{x},0)*(\Delta P(\mathbf{x}_r,t;\mathbf{x}_s)\big)\\ &\quad+\frac{\tau(\mathbf{x})}{\tau_\sigma(\mathbf{x})}\nabla\cdot\mathbf{v}(\mathbf{x},t;\mathbf{x}_s)\big(g_{r_p}(\mathbf{x}_r,-t;\mathbf{x},0)*\Delta P(\mathbf{x}_r,t;\mathbf{x}_s)\big)\Big),\\ &= \sum_s \int dt\Big(\big(\tau(\mathbf{x})+1\big)\big(\nabla\cdot\mathbf{v}(\mathbf{x},t;\mathbf{x}_s)q(\mathbf{x},t;\mathbf{x}_s)\big)+\big(\frac{\tau(\mathbf{x})}{\tau_\sigma(\mathbf{x})}\nabla\cdot\mathbf{v}(\mathbf{x},t;\mathbf{x}_s)s(\mathbf{x},t;\mathbf{x}_s)\big)\Big).\end{aligned} \tag{B.8}$$

Here, q and s are the adjoint-state variables of the state variable P and r_p, respectively. Equation B.8 is the complete migration Green's function in a viscoacoustic medium, and can also be represented by the matrix-vector notation

$$\mathbf{m}_{mig} = \mathbf{L}_Q^T\mathbf{d}_Q. \tag{B.9}$$

C Derivation of the Viscoacoustic Deblurring Filter

Similar to the estimation of the acoustic deblurring filter, viscoacoustic deblurring filters can be estimated using a reference model and its Q migration image. The reference model is constructed using a uniform distribution of point scatterers. The reference data $\mathbf{d}^Q_{ref}$ are generated from the reference reflectivity, background velocity model, and Q models, which are then migrated to get a reference Q migration image $\mathbf{m}^Q_{mig\_ref}$ as

$$\mathbf{m}^Q_{mig\_ref} = \mathbf{L}^T_Q \mathbf{L}_Q \mathbf{m}_{ref} = \mathbf{L}^T_Q \mathbf{d}^Q_{ref}. \tag{C.1}$$

Viscoacoustic deblurring filters for different subdomains of the Q migration image are then estimated by locally matching the reference Q migration image with the reference reflectivity model as

$$m_{ref}(x,y,z)_i = \int_{v_0} F_Q(x-x_0, y-y_0, z-z_0)_i m^Q_{ref\_mig} \times (x_0,y_0,z_0)_i dV_0, \tag{C.2}$$

which can be written in the matrix-vector notation:

$$[\mathbf{m}_{ref}]_i = [\mathbf{F}_Q]_i \otimes [\mathbf{m}^Q_{ref\_mig}]_i. \tag{C.3}$$

Here, $[\mathbf{F}_Q]_i$,$[\mathbf{m}_{ref}]_i$, and $[\mathbf{m}^Q_{ref_{mig}}]_i$ denote the viscoacoustic deblurring filter, the reference reflectivity model, and the reference Q migration image within the ith local window, respectively.

D Trace Integration

Assume a time-domain signal $g(t)$ is equal to $G(f)$ in the frequency domain. According to the derivative property of the Fourier transform, we have

$$F(\frac{dg(t)}{dt}) = i2\pi f G(f), \tag{D.1}$$

where F represents the Fourier transform operator. We can re-write equation D.1 as

$$F(\frac{dg(t)}{dt}) = i2\pi f F(g(t)), \tag{D.2}$$

and substitute $p(t) = \frac{dg(t)}{dt}$ into equation D.2 to get

$$F(p(t)) = i2\pi f F(\int_{-\infty}^{t} p(\tau)d\tau). \tag{D.3}$$

Dividing $i2\pi f$ on both side of the equation gives

$$F(\int_{-\infty}^{t} p(\tau)d\tau) = \frac{F(p(t))}{i2\pi f}. \tag{D.4}$$

We conclude that integrating a signal in the time domain is equal to divide by $i2\pi f$ in the frequency domain. In this case, high-frequency information will be attenuated and the low-frequency information will become dominant. However, if there is a DC component b in the signal $g(t)$ as $g(t) + b$, then the equation $p(t) = \frac{d(g(t)+b)}{dt}$ is also true. The value of the DC term increases quickly with an increase in the number of integrations. This DC term can be simply removed by demeaning the signal prior to inversion.

E Derivation of MRPI Gradient

The first-order acoustic wave equation is

$$\begin{bmatrix} \frac{\partial}{\partial t} & \rho v^2 \nabla \\ \frac{1}{\rho}\nabla & \frac{\partial}{\partial t} \end{bmatrix} \begin{bmatrix} P \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} F \\ 0 \end{bmatrix}, \tag{E.1}$$

where P and $\mathbf{v}$ represent the background pressure wavefield and particle-velocity field, respectively. ρ is the density and c indicates the velocity and the source term is represented by F. This equation can be rewritten more compactly as

$$\mathbf{S}(\mathbf{m_s})\mathbf{w}(\mathbf{m_s}) = \mathbf{F}, \tag{E.2}$$

where

$$\mathbf{S} = \begin{bmatrix} \frac{\partial}{\partial t} & \rho v^2 \nabla \\ \frac{1}{\rho}\nabla & \frac{\partial}{\partial t} \end{bmatrix}, \mathbf{w} = \begin{bmatrix} P \\ \mathbf{v} \end{bmatrix}, and\ \mathbf{F} = \begin{bmatrix} F \\ 0 \end{bmatrix}. \tag{E.3}$$

Here, $\mathbf{w}$ represent the background wavefield and $\mathbf{S}$ indicates as the forward modeling operator. The model parameter is defind as $\mathbf{m} = \begin{pmatrix} \mathbf{m_s} \\ \mathbf{m_r} \end{pmatrix}$, where $\mathbf{m_s}$ is the smooth background model and $\mathbf{m_r}$ is the reflectivity model. Similarly, the Born modeling equation can be written as

$$\mathbf{S}(\mathbf{m_s})\delta\mathbf{w}(\mathbf{m}) = \delta\mathbf{F}(\mathbf{w}, \mathbf{m_r}), \tag{E.4}$$

where $\delta\mathbf{w}$ represents the perturbed wavefield and $\delta\mathbf{F}$ is the virtual source calculated from the background wavefield $\mathbf{w}$ and the reflectivity model $\mathbf{m_r}$.

The least-squares misfit function $\epsilon(\mathbf{m})$ for the model parameter $\mathbf{m}$ can be written as

$$\epsilon(\mathbf{m}) = \frac{1}{2}\big\langle \delta\mathbf{w}(\mathbf{m}) - \mathbf{d}, \delta\mathbf{w}(\mathbf{m}) - \mathbf{d} \big\rangle, \tag{E.5}$$

where $\delta\mathbf{w}(\mathbf{m})$ and $\mathbf{d}$ represent the predicted and recorded data, respectively. The gradient of $\epsilon(\mathbf{m})$ with respect to the model parameter $\mathbf{m}$ is given by

$$\frac{\partial \epsilon(\mathbf{m})}{\partial \mathbf{m}} = \big\langle \frac{\partial \delta\mathbf{w}}{\partial \mathbf{m}}, \Delta\mathbf{d} \big\rangle. \tag{E.6}$$

Computing the derivative of equation E.4 with respect to the model parameter $\mathbf{m}$, we get

$$\begin{aligned} &\frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\delta\mathbf{w}(\mathbf{m}) + \mathbf{S}(\mathbf{m_s})\frac{\partial \delta\mathbf{w}(\mathbf{m})}{\partial \mathbf{m}} = \frac{\delta\mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{m}}, \\ \Rightarrow &\frac{\partial \delta\mathbf{w}(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{S}^{-1}\Big(\frac{\delta\mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{m}} - \frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\delta\mathbf{w}(\mathbf{m})\Big), \end{aligned} \tag{E.7}$$

where

$$\frac{\delta\mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{m}} = \frac{\partial \delta\mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{w}(\mathbf{m_s})}\frac{\partial \mathbf{w}(\mathbf{m_s})}{\partial \mathbf{m}} + \frac{\partial \delta\mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{m_r}}\frac{\partial \mathbf{m_r}}{\partial \mathbf{m}}. \tag{E.8}$$

Note that $\mathbf{m} = (\mathbf{m_s}, \mathbf{m_r})$, so that

$$\frac{\partial \mathbf{m_r}}{\partial \mathbf{m}} = \begin{cases} 1, & \text{if } \mathbf{m} = \mathbf{m_r} \\ 0, & \text{if } \mathbf{m} = \mathbf{m_s} \end{cases}, \quad \frac{\partial \mathbf{w}(\mathbf{m_s})}{\partial \mathbf{m}} = \begin{cases} \frac{\partial \mathbf{w}(\mathbf{m_s})}{\partial \mathbf{m}}, & \text{if } \mathbf{m} = \mathbf{m_s} \\ 0, & \text{if } \mathbf{m} = \mathbf{m_r} \end{cases}. \tag{E.9}$$

Similarly, taking the derivative of Equation E.2 with respect to the model parameter $\mathbf{m}$, we have

$$\frac{\partial \mathbf{w}(\mathbf{m_s})}{\partial \mathbf{m}} = -\mathbf{S}(\mathbf{m_s})^{-1}\frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\mathbf{w}. \tag{E.10}$$

Inserting Equation A-7, E.8 and E.10 into Equation E.6, gives

$$\begin{aligned}\frac{\epsilon(\mathbf{m})}{\partial \mathbf{m}} = \Big\langle & \frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{w}(\mathbf{m_s})}(-\mathbf{S}(\mathbf{m_s})^{-1})\frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\mathbf{w} + \\ & \frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{m_r}}\frac{\partial \mathbf{m_r}}{\partial \mathbf{m}} - \frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\delta \mathbf{w}(\mathbf{m}), (\mathbf{S}(\mathbf{m_s})^{-1})^* \Delta \mathbf{d} \Big\rangle \\ = & -\Big\langle \frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{w}(\mathbf{m_s})}\mathbf{S}(\mathbf{m_s})^{-1}\frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\mathbf{w}, \mathbf{w}^* \Big\rangle - \\ & \Big\langle \frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m}}\delta \mathbf{w}(\mathbf{m}), \mathbf{w}^* \Big\rangle + \Big\langle \frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{m_r}}\frac{\partial \mathbf{m_r}}{\partial \mathbf{m}}, \mathbf{w}^* \Big\rangle\end{aligned} \tag{E.11}$$

where $*$ indicates the adjoint. Here, we denote $\mathbf{w}(\mathbf{m})^* = (\mathbf{S}(\mathbf{m_s})^{-1})^* \Delta \mathbf{d}$ as the solution of the adjoint equation with the residual resismograms acting as virtual soures:

$$\mathbf{S}^*(\mathbf{m_s})\mathbf{w}^*(\mathbf{m}) = \Delta \mathbf{d}, \tag{E.12}$$

where $\mathbf{S}^*$ is the adjoint operator of $\mathbf{S}$. Therefore, according to Equation A-12, the gradient with respect to the background model $\mathbf{m_s}$ can be written as

$$\frac{\epsilon(\mathbf{m})}{\partial \mathbf{m_s}} = -\Big\langle \frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{w}(\mathbf{m_s})}\mathbf{S}(\mathbf{m_s})^{-1}\frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m_s}}\mathbf{w}, \mathbf{w}^* \Big\rangle - \Big\langle \frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m_s}}\delta \mathbf{w}(\mathbf{m}), \mathbf{w}^* \Big\rangle, \tag{E.13}$$

$$\begin{aligned} & = -\Big\langle \frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m_s}}\mathbf{w}, (\mathbf{S}(\mathbf{m_s})^{-1})^* \big(\frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{w}(\mathbf{m_s})}\big)^* \mathbf{w}^* \Big\rangle - \Big\langle \frac{\partial \mathbf{S}(\mathbf{m_s})}{\partial \mathbf{m_s}}\delta \mathbf{w}(\mathbf{m}), \mathbf{w}^* \Big\rangle, \\ & = -\Big\langle \frac{\partial \mathbf{S}}{\partial \mathbf{m_s}}\mathbf{w}, \delta \mathbf{w}^* \Big\rangle - \Big\langle \frac{\partial \mathbf{S}}{\partial \mathbf{m_s}}\delta \mathbf{w}, \mathbf{w}^* \Big\rangle, \end{aligned}$$

where we denote $\delta \mathbf{w}^* = (\mathbf{S}(\mathbf{m_s})^{-1})^* \big(\frac{\partial \delta \mathbf{F}(\mathbf{w}(\mathbf{m_s}), \mathbf{m_r})}{\partial \mathbf{w}(\mathbf{m_s})}\big)^* \mathbf{w}^*$ as the solution of the adjoint Born equation

$$\mathbf{S}^*(\mathbf{m_s})\delta \mathbf{w}^*(\mathbf{m}) = \delta \mathbf{F}^*(\mathbf{w}^*, \mathbf{m_r}), \tag{E.14}$$

and $\mathbf{w}^*$ is background adjoint wavefield which satisfies the adjoint equation

$$\mathbf{S}^*(\mathbf{m_s})\mathbf{w}^*(\mathbf{m}) = \mathbf{F}^*. \tag{E.15}$$

Here,

$$\mathbf{S}^* = \begin{bmatrix} -\frac{\partial}{\partial t} & -\nabla \cdot \frac{1}{\rho} \\ -\nabla \rho v^2 & -\frac{\partial}{\partial t} \end{bmatrix}, \mathbf{w}^* = \begin{bmatrix} q \\ \mathbf{u} \end{bmatrix}, and\ \mathbf{F}^* = \begin{bmatrix} \mathbf{\Delta d} \\ 0 \end{bmatrix}, \tag{E.16}$$

where q is the adjoint state variable of the pressure wavefield P and $\mathbf{u}$ is the adjoint of the particle velocity vector $\mathbf{v}$. For $\mathbf{m_s} = v$, the gradient in equation A-13 can be written as

$$-\Big\langle \frac{\partial \mathbf{S}}{\partial \mathbf{m_s}} \mathbf{w}, \delta \mathbf{w}^* \Big\rangle = -\Big\langle \begin{bmatrix} 0 & 2\rho v \nabla \cdot \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P \\ \mathbf{v} \end{bmatrix}, \begin{bmatrix} \delta q \\ \delta \mathbf{u} \end{bmatrix} \Big\rangle, \tag{E.17}$$

$$= -\int_0^T 2\rho v (\nabla \cdot \mathbf{v}) \delta q dt,$$

$$-\Big\langle \frac{\partial \mathbf{S}}{\partial \mathbf{m_s}} \delta \mathbf{w}, \mathbf{w}^* \Big\rangle = -\Big\langle \begin{bmatrix} 0 & 2\rho v \nabla \cdot \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta P \\ \delta \mathbf{v}, \end{bmatrix}, \begin{bmatrix} q \\ \mathbf{u} \end{bmatrix} \Big\rangle \tag{E.18}$$

$$= -\int_0^T 2\rho v (\nabla \cdot \delta \mathbf{v}) q dt, \tag{E.19}$$