



HAL
open science

Inversion of surface waves in an oil and gas exploration context

Isabella Masoni

► **To cite this version:**

Isabella Masoni. Inversion of surface waves in an oil and gas exploration context. Earth Sciences. Université Grenoble Alpes, 2016. English. NNT : 2016GREAU029 . tel-01567641

HAL Id: tel-01567641

<https://theses.hal.science/tel-01567641>

Submitted on 24 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité: Sciences de la Terre, de l'Univers et de l'Environnement

Arrêté ministériel : 7 Août 2006

Présentée par

ISABELLA MASONI

Thèse dirigée par

Jean Virieux, Romain Brossier et Jean-Luc Boelle

préparée au sein de Institut des Sciences de la Terre
et de l'Ecole Doctorale Terre Univers Environnement

Inversion of surface waves in an oil & gas exploration context

23 Septembre 2016 ,
devant le jury composé de:

Hervé Chauris

Professeur à MINES ParisTech, France, Rapporteur

René-Édouard Plessix

Chercheur à Shell Global Solutions, Pays-Bas, Rapporteur

Hansruedi Maurer

Professeur à ETH Zurich, Suisse, Examineur

Jean-Luc Boelle

Chercheur à TOTAL E&P, France, Co-Directeur de thèse

Romain Brossier

Maître de Conférences à Université Grenoble Alpes, France, Co-Directeur
de thèse

Jean Virieux

Professeur à Université Grenoble Alpes, France, Directeur de thèse



Acknowledgements

To begin with I would like to thank each of my supervisors for guiding and advising me these past few years. Thank you to Romain Brossier for his pedagogical approach to teaching and for his help with all things code related. Thank you to Jean Virieux for his insight, suggestions and persistence. And thank you to Jean-Luc Boelle for his continuous support and his assistance with the real data application.

I would like to thank my PhD committee members Rene-Edouard Plessix, Hervé Chauris and Hansreudi Maurer for taking the time to read my manuscript and for their insightful comments and suggestions.

I would like to thank the people I worked with, both at the ISTerre laboratory and at Total, who made significant contributions to the work presented in this thesis, and without whom this study would not have been as successful.

Thank you to the past and current members of the SEISCOPE Consortium for their advice and support, and especially for providing the codes, toolboxes and constructive working environment I was able to profit from. Thank you to the past and current SEISCOPE students as well as the other "Ondes" team students for the helpful discussions on both professional and personal levels. Thank you also to the technicians and the CIMENT high performance computing facilities.

Thank you to the past and current members of the Earth Imaging research team at Total. I especially thank the members of the Near Surface group, both from Total and CGG, for their very significant contribution to the real data application presented in this thesis. A special thank you to the computer scientists of our team for helping me solve any computing problem I encountered.

I would like to thank my friends, both those in France and those continents apart, for their unceasing encouragement and pep-talks. Thank you for the traveling adventures, the mountain activities, and the frisbee throwing. I especially thank my partner for his patience, his helpful corrections and suggestions, and for his invaluable advice.

Last but not least, I would like to thank my family for giving me the opportunity to reach my goals, and for being there for me every step of the way.

Thank you. Merci. Grazie. Dankeschön.

Abstract

The characterization of the near surface is an important topic for the oil and gas industry. For land and Ocean Bottom Cable (OBC) acquisitions, weathered or unconsolidated top layers, prominent topography and complex shallow structures may make imaging at target depth very difficult. Energetic and complex surface waves often dominate such recordings, masking the signal and challenging conventional seismic processing. Static corrections and the painstaking removal of surface waves are required to obtain viable exploration information.

Yet surface waves, which sample the near surface region, are considered as signal on both the engineering and geotechnical scale as well as the global seismology scale. Their dispersive property is conventionally used in surface wave analysis techniques to obtain local shear velocity depth profiles. But limitations such as the picking of dispersion curves and poor lateral resolution have led to the proposal of Full Waveform Inversion (FWI) as an alternative high resolution technique. FWI can theoretically be used to explain the complete waveforms recorded in seismograms, but FWI with surface waves has its own set of challenges. A sufficiently accurate initial velocity model is required or otherwise cycle-skipping problems will prevent the inversion to converge.

This study investigates alternative misfit functions, that can overcome cycle-skipping and decrease the dependence on the initial model required. Computing the data-fitting in different domains such as the frequency-wavenumber ($\omega - k$) and frequency-slowness ($\omega - p$) domains is proposed for robust FWI, and successful results are achieved with a synthetic dataset, in retrieving lateral shear velocity variations.

In the second part of this study a FWI layer stripping strategy, specifically adapted to the physics of surface waves is proposed. The penetration of surface waves is dependent on their wavelength, and therefore on their frequency. High-to-low frequency data is therefore sequentially inverted to update top-to-bottom layer depths of the shear velocity model. In addition, near-to-far offsets are considered to avoid cycle-skipping issues. Results with a synthetic dataset show that this strategy is more successful than conventional multiscale FWI in using surface waves to update the shear velocity model.

Finally inversion of surface waves for near surface characterization is attempted on a real dataset at the oil and gas exploration scale. The construction of initial models and the difficulties encountered during FWI with real data are discussed.

Keywords: surface waves, full waveform inversion, near surface, inverse problems

Résumé

La caractérisation de la proche surface est un enjeu majeur pour l'industrie pétrolière. Lors des acquisitions terrestres et Ocean Bottom Cable (OBC), les couches superficielles généralement altérées ou peu consolidées, présentent des structures géologiques complexes et ont éventuellement des variations topographiques importantes. Les ondes de surface, énergétiques, se propagent dans ce milieu complexe et dominent les sismogrammes, ce qui masque le signal utile pour le traitement sismique classique et rend difficile l'imagerie à la profondeur du réservoir.

Il est donc important de pouvoir atténuer ces ondes, éventuellement d'appliquer des corrections statiques et/ou d'amplitude. Ceci qui nécessite une connaissance précise du modèle de vitesse de la proche surface. L'étude de la dispersion des ondes de surface est couramment utilisée en sismologie globale et à l'échelle géotechnique pour évaluer les propriétés des milieux terrestres. Il existe néanmoins des limitations: la mesure de cette dispersion est souvent difficile et les profils de vitesses obtenus sont 1D. A l'échelle pétrolière, l'hypothèse 1D n'est pas toujours adaptée, ce qui motive l'utilisation d'une méthode alternative d'imagerie plus haute résolution, la méthode d'inversion de la forme d'onde (FWI). Cependant, le modèle de vitesse initial doit être assez précis pour éviter le "cycle-skipping" et permettre la convergence vers la solution optimale.

Cette étude explore différentes alternatives de fonctions coûts pour résoudre le "cycle-skipping" et diminuer la dépendance de l'inversion à la qualité du modèle initial. En exprimant les fonctions coûts dans le domaine $\omega - k$ (fréquence-nombre d'onde) et le domaine $\omega - p$ (fréquence-lenteur), la FWI est plus robuste. A l'aide d'exemples synthétiques, nous démontrons l'efficacité de ces nouvelles approches qui permettent bien de retrouver les variations latérales de vitesses d'onde S.

Dans une seconde partie, nous développons une inversion FWI en "layer stripping", adaptée spécifiquement à la physique des ondes de surface. Comme la profondeur de pénétration de ces ondes dépend de leur longueur d'onde, et donc, de leur contenu fréquentiel, nous proposons d'inverser séquentiellement des plus hautes aux plus basses fréquences de ces ondes pour contraindre successivement les couches superficielles jusqu'aux plus profondes. Un fenêtrage selon la distance source-station est également appliqué. Dans un premier temps seules les courtes distances sont inversées, au fur à mesure les données associées à des plus grandes distances sont rajoutées, plus fortement impactées par le "cycle-skipping". Nous démontrons à l'aide d'exemples synthétiques l'avantage de cette méthode par rapport aux méthodes multi-échelles conventionnelles inversant des basses

vers le hautes fréquences.

Enfin, l'inversion des ondes de surface pour la caractérisation de la proche surface est confrontée à un cas réel. Nous discutons la construction et la pertinence du modèle initial et les difficultés rencontrées lors de l'inversion.

Mots clés: ondes de surface, inversion de formes d'onde, proche surface, problèmes inverse

Contents

General Introduction	13
1 Inversion of Surface Waves	19
1.1 Surface waves and their properties	20
1.1.1 Theory of elasticity	21
1.1.2 Resolution of the elastodynamic equation in a semi-infinite half-space: Rayleigh wave definition	23
1.1.3 Amplitude of Rayleigh waves with depth	28
1.1.4 Dispersion of surface waves	32
1.2 Inversion of dispersion curves	35
1.2.1 Surface wave analysis methods	35
1.2.2 Picking dispersion curves	38
1.2.3 Dealing with lateral variations	39
1.3 Full Waveform Inversion	40
1.3.1 Introduction to Full Waveform Inversion	41
1.3.1.1 Elastic forward modelling	41
1.3.1.2 Inverse problem	44
1.3.1.3 Initial models	47
1.3.2 Local minima and the cycle-skipping issue	48
1.3.2.1 The multiscale strategy	50
1.3.2.2 Other strategies to avoid cycle-skipping	51
1.3.3 Full Waveform Inversion with surface waves	53
1.3.3.1 From the global scale to the near surface scale	53
1.3.3.2 At the oil and gas exploration scale	56
2 Alternative Misfit Functions for Robust FWI	61
2.1 Introduction	62
2.2 Robust Full Waveform Inversion with surface waves	63
2.2.1 Abstract	63
2.2.2 Introduction	64
2.2.3 Forward modelling	66
2.2.4 Misfit function analysis	68
2.2.4.1 Robust misfit design	69
2.2.4.2 Sample-by-sample strategy	71

2.2.5	Optimization approach	73
2.2.5.1	Misfit in the domain (t, h)	73
2.2.5.2	Misfit in the domain (ω, p)	75
2.2.5.3	Misfit in the domain (ω, k)	77
2.2.6	Illustrative synthetic example	79
2.2.7	Conclusion and Perspectives	83
2.2.8	Acknowledgements	84
2.3	Other proposed misfit functions	85
3	Strategy for FWI with Surface Waves	87
3.1	Introduction	88
3.1.1	Depth preconditioning	91
3.2	Layer stripping approach	93
3.2.1	Motivation	95
3.2.2	Layer stripping workflow	96
3.2.3	Illustrative synthetic example	100
3.2.4	Parameter selection analysis	112
3.2.5	Parameter sensitivity and misfit function selection	117
3.2.6	Quality control	121
3.3	Far initial models	123
3.3.1	Far initial shear velocity model	123
3.3.2	Far initial P-wave velocity model	127
3.4	Conclusion and perspectives	131
4	Real Data Case Study	133
4.1	Case study background & acquisition	134
4.1.1	Analysis of common shot gathers	136
4.1.2	Data pre-processing for FWI	141
4.2	Surface wave inversion	143
4.2.1	Initial models for FWI	143
4.2.1.1	Surface wave 1D inversion	144
4.2.1.2	Quality control of dispersion curves	148
4.2.1.3	Selection of initial models	151
4.2.2	Impact of receiver array	153
4.2.3	Source estimation	155
4.2.4	Forward modelling and cycle-skipping analysis	157
4.2.4.1	Analysis of low frequency data ($4 - 15 Hz$)	158
4.2.4.2	Muting and windowing in offset	160
4.2.4.3	Masking in the (ω, k) domain	161
4.2.5	Full Waveform Inversion	163
4.3	Conclusion and evaluation	168
4.4	Perspectives	169
4.4.1	Dealing with topography	169

Conclusions and Perspectives	175
A Complementary Publications	181
A.1 Alternative misfit functions for FWI applied to surface waves	182
Bibliography	186

General Introduction

Imaging techniques, such as for seismic or medical imaging, aim to obtain information on the properties of a medium that is otherwise inaccessible. Waves that travel through such mediums and that can be measured on the surface, can provide diagnostic information that is fundamental to these investigations. In the case of seismic prospecting, the medium we want to image is the subsurface of the Earth. Knowledge about subsurface characteristics is critical for the oil and gas industry. In conventional exploration acquisitions, the reflections of body waves recorded in the seismograms can provide information on target reservoirs present underneath the Earth's surface.

However, for land and Ocean Bottom Cable (OBC) acquisitions, topography and weathered or unconsolidated top layers that correspond to a very complex near surface, may lead to highly variable seismic signals difficult to interpret. In such environments, energetic and complex surface waves may dominate these seismograms, masking the reflections, and making the extraction of information challenging for conventional seismic processing (see Figure 0.1). Furthermore, due to their characteristic exponential decay with depth, these surface waves are not influenced by deep targets of interest, but only by the near surface properties.

These surface waves, also known as ground roll, are often considered as noise, or more precisely, as an undesired signal. As stated in [Dobrin \(1950\)](#), *"the geophysicist engaged in seismic prospecting ordinarily looks upon surface waves as a perverse creation designed to interfere with his reflections. He adjusts his equipment and field procedure to eliminate them from his records"*.

Over the years, sophisticated techniques have been developed to remove or attenuate this coherent noise in seismic data, both during the acquisition and processing steps. Dip- or velocity-based filters such as FK filters, Radon methods or other tools such as static correction methods, stacking or modelling are used to filter out surface waves and enhance the useful signal, of smaller amplitude, related to reflections ([Yilmaz, 1987](#);

Ernst and Herman, 1998; Herman and Verschuur, 2004). Yet this remains a difficult and tedious task.

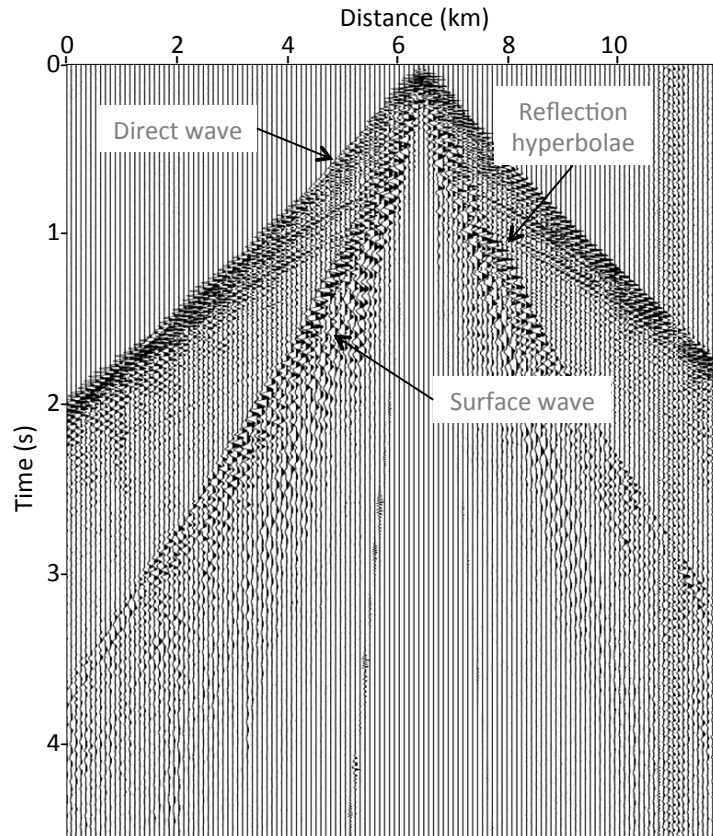


Figure 0.1: Example of a seismic record acquired on land. This common shot gather contains traces from receivers spread out for a few kilometers on either side of the source position. The surface waves travel the slowest through the subsurface and have the largest amplitudes as they are confined near the free surface. They can often mask reflections of lower amplitude, which contain information on properties of the medium at the target depth. Note that an amplitude gain has been applied for better visualization of the data.

The aim of this study is to build on our understanding of how surface waves may contribute to seismic velocity model building, and enhance our ability to deal with them, so that *"surface waves can be made a useful by-product rather than a waste product of seismic exploration"* as proposed by [Dobrin \(1950\)](#).

Motivation

There is an increasing interest by the oil and gas industry in the large amount of information contained in surface waves. Their dispersive property and their characteristic high amplitudes, have historically made them a useful signal for global seismology investigations as well as near surface applications on the engineering/geotechnical scale. Classical wave theory indicates that surface waves will contain information on the near surface region. Surface waves have been expressly recorded and analyzed to characterize the Earth's crust, as the near surface region on the global scale, as well as the near surface region up to 10 *m* depth on the engineering scale. On the intermediate oil and gas exploration scale, surface waves could provide information on the low-velocity weathered zone of the near surface, which could allow to better image at depth.

Successful imaging of target reservoirs at depth, for land acquisitions, depends on sufficiently reliable static corrections, and knowledge on the geological structures in the near surface region. Therefore, the possibility of high resolution imaging of this region, with the use of surface waves, can be a significant advantage. Ultimately, an improved, reliable image of the subsurface, can reduce the financial and environmental risk involved in oil and gas exploration, by enabling to drill in exactly the right location.

Research objectives

This study is driven by the question: "Can surface waves be used to characterize the near surface at the oil and gas exploration scale using Full Waveform Inversion?" To answer this question, the main aim of this thesis is to develop an alternative formulation or strategy to perform FWI with surface waves, and overcome the known problems of initial model dependence and cycle-skipping issues.

The first objective is to perform a sensitivity analysis on alternative misfit functions to determine which may be more robust when considering surface waves in FWI. The second objective deals with investigating an appropriate inversion strategy that both targets surface waves, and results in successful velocity model building. For both of these objectives FWI results on synthetic data examples need to be provided.

Beyond developing the methodology, the third objective of this study is to perform FWI with surface waves on a real oil and gas exploration dataset. This experience will

allow to highlight the difficulties encountered when dealing with real data and which directions to move forward in for future studies.

Outline

The work presented in this thesis is organized into four main chapters, followed by the conclusions and perspectives.

Chapter 1 is an introductory chapter. Surface waves are first defined, and some of their useful properties and characteristics described. The inversion of surface waves is then reviewed, starting with the surface wave analysis method, and investigating some of the limitations of assumed stratified medium (1D) inversion using group velocity or phase velocity measurements from dispersion curves. The Full Waveform Inversion (FWI) method is then described in detail, and a comprehensive explanation of cycle-skipping limitations for FWI is provided. A review of the literature available on applications of FWI using surface waves on the global seismology scale, the engineering/geotechnical scale, and the oil & gas reservoir scale follows.

In Chapter 2 I investigate the influence of different, more robust misfit functions suitable for FWI with surface waves, and I compare them using a grid analysis. Misfit functions where the data-fitting is computed in different data domains such as the frequency-wavenumber ($\omega - k$) and frequency-slowness ($\omega - p$) are proposed. The gradient formulation for these misfit functions is also developed. A simple 2D synthetic example is tested to illustrate the success of these misfit functions proposed.

In Chapter 3 I investigate a realistic configuration and develop an inversion strategy to implement for FWI using surface waves. A layer stripping approach specifically adapted to the physics of surface waves is proposed. For this approach, high-to-low frequency bands of data are sequentially inverted, to update the model in a top-to-bottom manner. Several tests using a more complex 2D synthetic model taken from Pérez Solano et al. (2014) investigate the advantages of this approach compared to the conventional multi-scale low-to-high frequency continuation approach. Due to the strongly varying sampling in depth of the surface waves, layer stripping techniques in the time-offset ($t - x$) domain can provide a sustainable workflow, which can compete efficiently with the previously proposed frequency-wavenumber ($\omega - k$) domain misfit function. Different methods for handling the complex problem of a real data application are therefore developed.

In Chapter 4, a real data application of FWI with surface waves is presented. The

dataset, provided by Total, is acquired at the oil and gas exploration scale. The real seismograms are analyzed and presented. Surface wave 1D inversion (SWODI) is used to provide an initial study of the near surface features as well as an initial model for FWI. The difficulties of FWI with real data are highlighted and discussed. Potential areas of improvement and research directions for future studies are also identified.

Finally the conclusions and perspectives highlight the important aspects for the use of surface waves in FWI and possible future research steps.

Chapter 1

Inversion of Surface Waves

Contents

1.1	Surface waves and their properties	20
1.1.1	Theory of elasticity	21
1.1.2	Resolution of the elastodynamic equation in a semi-infinite half-space: Rayleigh wave definition	23
1.1.3	Amplitude of Rayleigh waves with depth	28
1.1.4	Dispersion of surface waves	32
1.2	Inversion of dispersion curves	35
1.2.1	Surface wave analysis methods	35
1.2.2	Picking dispersion curves	38
1.2.3	Dealing with lateral variations	39
1.3	Full Waveform Inversion	40
1.3.1	Introduction to Full Waveform Inversion	41
1.3.2	Local minima and the cycle-skipping issue	48
1.3.3	Full Waveform Inversion with surface waves	53

1.1 Surface waves and their properties

Surface waves are generated in the presence of a free surface boundary condition, as is the case of the solid/air interface on the surface of the Earth. Unlike body waves, surface waves do not radiate energy towards the Earth's interior, but instead propagate parallel to the Earth's surface, confined to the shallow zone of the near surface. The most common type of surface waves are Rayleigh waves, the resonant mode of the elastic halfspace, first predicted by [Rayleigh \(1887\)](#).

Rayleigh waves are generated whenever a free surface is present, and propagate along it, earning the name *groundroll* in the oil and gas industry. There is a coupling between the propagation of P-waves and vertically polarized SV shear waves, which gives the motion of Rayleigh waves. The particle motion generated by Rayleigh waves, at the free surface of a homogeneous medium, takes the shape of a retrograde ellipse at the free surface (see [Figure 1.1a](#)).

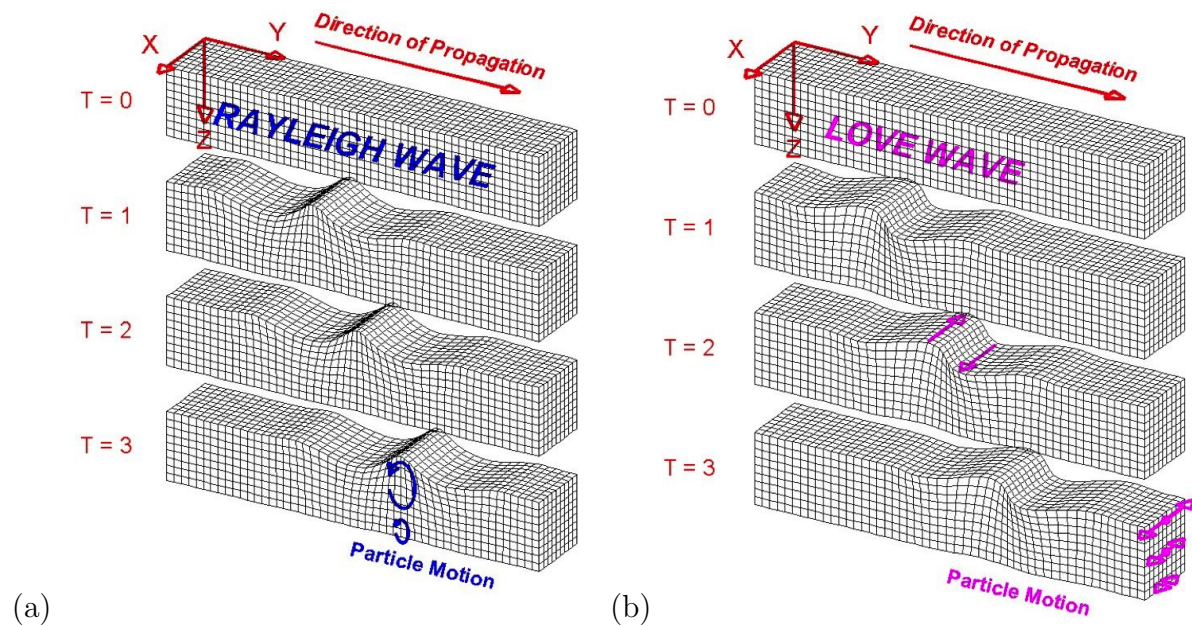


Figure 1.1: Perspective view of surface waves propagating through a grid representing an elastic homogeneous medium. Rayleigh wave propagation generates a retrograde elliptical particle motion (a). Love waves generate a horizontal particle motion, perpendicular to the direction of wave propagation, and are considered as a layer mode (b). The amplitude of the surface waves decreases with depth. Illustration from [Braile \(2010\)](#).

Another example of surface waves are Love waves, a layer-mode propagating wave, first discovered by [Love \(1911\)](#). Love waves, occur when a guided wave is generated at the

interface between a medium and the layer above, of lower velocity and limited thickness, which itself is bounded by a free surface. They are produced by the propagation of horizontally polarized SH shear waves, and have a shear displacement confined to the horizontal plane. The motion of Love waves is perpendicular to the wave propagation direction and parallel to the free surface (see Figure 1.1b). They are very sensitive to geometrical horizontal variations of low velocity.

Rayleigh and Love waves are recorded during land seismic acquisitions. In the case of a water layer overlaying an elastic solid medium, another type of surface wave, Scholte waves, are created at the fluid/solid interface (Scholte, 1947). Their particle motion is similar to Rayleigh waves, and they are generally recorded in shallow Ocean Bottom Cable (OBC) acquisitions.

To better understand surface waves and their characteristics, one can begin from their mathematical definition. From the expressions obtained, their physical description and distinctive properties can then be determined and analyzed.

1.1.1 Theory of elasticity

The definition of surface waves can be derived from Newton's law of motion. When considering that all internal and external forces are in equilibrium, the equation of motion is given as

$$\partial_j \sigma_{ij} + f_i = \rho \frac{\partial^2 u_i}{\partial t^2}, \quad (1.1)$$

illustrating the relation linking body forces f_i and forces originating from stresses σ_{ij} within the body, with the displacement u_i in direction i . The terms u_i , σ_{ij} and f_i , written in index notation for a cartesian coordinate system, are each a function of position (x, y, z) and time t . The term ρ is the bulk density.

The Earth can be considered as an elastic medium, where the stress can be expressed as a function of strain ϵ_{kl} by Hooke's law given as

$$\sigma_{ij} = c_{ijkl} \epsilon_{kl}, \quad (1.2)$$

where c_{ijkl} is the stiffness tensor. Equation 1.2 shows how the strain (or deformation) of a linearly elastic medium is proportional to the stress applied.

When considering the propagation of a wavefield in a homogeneous isotropic medium, with a constant stiffness tensor and no preferred direction, Equation 1.2 can be expressed as

$$\sigma_{ij} = \lambda \epsilon_{kk} \delta_{ij} + 2\mu \epsilon_{ij} , \quad (1.3)$$

where λ and μ are the Lamé parameters and δ_{ij} is the Kronecker delta (equal to 1 for $i = j$, and otherwise 0).

The strain tensor is related to the displacement as

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) . \quad (1.4)$$

Using Equations 1.3 and 1.4, the stress in Equation 1.1 can be expressed in terms of displacement. Considering the case where no external forces are applied, and all forces are derived from stresses, this leads to the second-order elastodynamic wave equation given as

$$(\lambda + 2\mu) \nabla \nabla \cdot \mathbf{u} - \mu \nabla \times \nabla \times \mathbf{u} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} , \quad (1.5)$$

where \mathbf{u} is the displacement vector in vector notation. The solution to Equation 1.5 can be obtained either by Helmholtz's decomposition, giving

$$\mathbf{u} = \nabla \Phi + \nabla \times \Psi , \quad (1.6)$$

where Φ is the scalar potential and Ψ is the vector potential, or following Kennett's approach (Kennett, 1983), as is done in the following section 1.1.2. The potentials Φ and Ψ may be shown to be directly linked to compressional P-waves and shear S-waves respectively, obeying wave equations:

$$\nabla^2 \Phi = \frac{1}{V_P^2} \frac{\partial^2 \Phi}{\partial t^2} \quad (1.7)$$

$$\nabla^2 \Psi = \frac{1}{V_S^2} \frac{\partial^2 \Psi}{\partial t^2} , \quad (1.8)$$

where $V_P = \sqrt{\frac{\lambda+2\mu}{\rho}}$, and $V_S = \sqrt{\frac{\mu}{\rho}}$. In these expressions V_P and V_S are respectively

the velocities of P-waves and S-waves. One can observe that they only depend on the properties of the material through which they propagate.

1.1.2 Resolution of the elastodynamic equation in a semi-infinite half-space: Rayleigh wave definition

For a semi-infinite elastic half-space, the solution of the elastodynamic equation requires to solve Equation 1.5. Alternatively one can solve the elastodynamic hyperbolic equations of order one, where the derivatives with z are considered, where z is chosen as the vertical axis (perpendicular to stratification in the medium). This technique will then be used for the more general case of horizontally stratified media. The system of equations obtained from integrating Equation 1.4 in Equation 1.3 is

$$\begin{aligned}
 \partial_z u_z &= \frac{1}{\lambda + 2\mu} \sigma_{zz} - \frac{\lambda}{\lambda + 2\mu} \partial_x u_x \\
 \partial_z u_x &= \frac{1}{\mu} \sigma_{xz} - \partial_x u_z \\
 \partial_z \sigma_{zz} &= \rho \partial_t^2 u_z - \partial_x \sigma_{xz} \\
 \partial_z \sigma_{xz} &= \rho \partial_t^2 u_x - \partial_x \sigma_{xx} ,
 \end{aligned} \tag{1.9}$$

with

$$\partial_x \sigma_{xx} = 4(\lambda + 2\mu) \partial_x^2 u_x + \frac{\lambda}{\lambda + 2\mu} \partial_x \sigma_{zz} . \tag{1.10}$$

In the following, the focus will remain on the 2D propagation of Rayleigh waves, although a similar approach may be derived for Love waves. The 2D (x, z) domain is chosen for simplicity but an equivalent formulation may be derived in 3D either with a 3D cartesian (x, y, z) grid, or a 3D cylindrical (r, θ, z) approach (Kennett, 1983).

The following boundary conditions also need to be applied:

- the free surface boundary condition for null stress at the free surface, giving $\sigma_{xz} = \sigma_{zz} = 0$ at $z = 0$;
- the radiation condition at $z \rightarrow \infty$, for which amplitudes diminish to zero at large distances from the surface, and for which no waves propagate from infinity back into the medium (Eringen and Suhubi, 1975).

The system of equations obtained in 1.9 may be advantageously transformed into the frequency-wavenumber (ω, k) domain (Aki and Richards, 1980; Kennett, 1983) with a 2D Fourier transform applied to both the time t and distance x axis. The transform operator is defined as

$$\hat{f}(k, z, \omega) = \int \int_{-\infty}^{+\infty} f(x, z, t) e^{i(\omega t - kx)} dx dt, \quad (1.11)$$

and gives the following propagation equation in vector notation, which needs to be solved:

$$\frac{\partial}{\partial z} \mathbf{B} = -i \mathbf{A} \mathbf{B}, \quad (1.12)$$

where the stress-displacement vector \mathbf{B} is defined as

$$\mathbf{B} = \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \hat{u}_z & \hat{u}_x & i\hat{\sigma}_{zz} & i\hat{\sigma}_{xz} \end{bmatrix}^T, \quad (1.13)$$

containing displacement (\hat{u}) and stress ($\hat{\sigma}$) terms, and where the coefficient matrix \mathbf{A} is given as

$$\mathbf{A} = \begin{bmatrix} 0 & k(1 - \frac{2V_S^2}{V_P^2}) & \frac{1}{\rho V_P^2} & 0 \\ k & 0 & 0 & \frac{1}{\rho V_S^2} \\ \rho\omega & 0 & 0 & k \\ 0 & \rho[\omega^2 - 4k^2 V_S^2(1 - \frac{V_S^2}{V_P^2})] & k(1 - \frac{2V_S^2}{V_P^2}) & 0 \end{bmatrix}, \quad (1.14)$$

and can be obtained directly from the system of equations in 1.9. Diagonalizing matrix \mathbf{A} yields

$$\mathbf{A} = \mathbf{D} \mathbf{\Lambda} \mathbf{D}^{-1}, \quad (1.15)$$

where the diagonal matrix $\mathbf{\Lambda}$, containing the eigenvalues of \mathbf{A} , is given as

$$\mathbf{\Lambda} = \text{diag} \left[q_P, q_S, -q_P, -q_S \right]. \quad (1.16)$$

The terms q_P and q_S are the vertical wavenumbers for P-waves and S-waves respectively, and are defined as

$$q_P = \left(\frac{\omega^2}{V_P^2} - k^2 \right)^{\frac{1}{2}}, \quad (1.17)$$

and

$$q_S = \left(\frac{\omega^2}{V_S^2} - k^2 \right)^{\frac{1}{2}}. \quad (1.18)$$

The matrix \mathbf{D} is chosen to be the local eigenvector matrix for \mathbf{A} , given as

$$\mathbf{D} = \begin{bmatrix} q_P & -k & -q_P & -k \\ k & q_S & k & -q_S \\ 2\mu k^2 - \rho\omega^2 & 2\mu k q_S & 2\mu k^2 - \rho\omega^2 & -2\mu k q_S \\ -2\mu k q_P & 2\mu k^2 - \rho\omega^2 & 2\mu k q_P & 2\mu k^2 - \rho\omega^2 \end{bmatrix}. \quad (1.19)$$

From integrating Equation 1.15 in Equation 1.12 one obtains

$$\frac{\partial}{\partial z} \mathbf{B} = -i \mathbf{D} \mathbf{A} \mathbf{D}^{-1} \mathbf{B}. \quad (1.20)$$

Since a homogeneous medium is considered, the coefficient matrix \mathbf{A} is constant, and so the eigenvector matrix \mathbf{D} is independent of z . One can introduce matrix \mathbf{V} such as

$$\mathbf{B}(z) = \mathbf{D} \mathbf{V}(z), \quad (1.21)$$

therefore obtaining a system of independent equations that can be solved, given as

$$\frac{\partial}{\partial z} \mathbf{V}(z) = -i \mathbf{A} \mathbf{V}(z). \quad (1.22)$$

The solution to Equation 1.22 is given as

$$\mathbf{V}(z) = \begin{bmatrix} e^{-iq_P(z-z_0)} & 0 & 0 & 0 \\ 0 & e^{-iq_S(z-z_0)} & 0 & 0 \\ 0 & 0 & e^{+iq_P(z-z_0)} & 0 \\ 0 & 0 & 0 & e^{+iq_S(z-z_0)} \end{bmatrix} \mathbf{V}(z_0), \quad (1.23)$$

where the terms depend on the difference between the current depth z and the reference level z_0 . The matrix \mathbf{V} can be summarized in terms of upgoing (U) and downgoing

(D) waves, characterized by their dependence on z , using Kennett's notation (Kennett, 1983), and giving

$$\mathbf{V}(z) = \begin{bmatrix} \mathbf{V}_U(z) \\ \mathbf{V}_D(z) \end{bmatrix}. \quad (1.24)$$

where \mathbf{V}_D is the downgoing wavefield for evanescent waves whose amplitudes decay with depth, and \mathbf{V}_U is the upgoing wavefield. The components of $\mathbf{V}(z)$ are very similar to the scalar and vector potentials Φ and Ψ , previously mentioned in Equation 1.6.

One can introduce the upward and downward transmission operators from z_0 to z , notated by T_U and T_D respectively, as

$$\begin{bmatrix} \mathbf{V}_U(z) \\ \mathbf{V}_D(z) \end{bmatrix} = \begin{bmatrix} \mathbf{T}_U & 0 \\ 0 & \mathbf{T}_D \end{bmatrix} \begin{bmatrix} \mathbf{V}_U(z_0) \\ \mathbf{V}_D(z_0) \end{bmatrix}, \quad (1.25)$$

giving

$$\mathbf{T}_D = \mathbf{T}_U = \begin{bmatrix} e^{+iq_P(z-z_0)} & 0 \\ 0 & e^{+iq_S(z-z_0)} \end{bmatrix}. \quad (1.26)$$

From the radiation boundary condition at $z \rightarrow \infty$, the determination of terms q_P and q_S in the complex plane are chosen as follows

$$\text{Im}(q_P) > 0 \quad (1.27)$$

$$\text{Im}(q_S) > 0 \quad , \quad (1.28)$$

to ensure the decaying of the downgoing wavefield \mathbf{V}_D , for imaginary q_P and q_S values of $e^{+iq_P z}$ and $e^{+iq_S z}$ respectively. Following the radiation condition in a semi infinite homogeneous half-space,

$$\mathbf{V}_U(z) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (1.29)$$

Following the free surface boundary condition required, for which $\hat{\sigma}_{xz} = \hat{\sigma}_{zz} = 0$ when $z = 0$, Equation 1.21 results in

$$\mathbf{B}(0) = \begin{bmatrix} \hat{\mathbf{u}} \\ 0 \end{bmatrix} = \mathbf{D}\mathbf{V}(0) = \begin{bmatrix} \mathbf{m}_U & \mathbf{m}_D \\ \mathbf{n}_U & \mathbf{n}_D \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{V}_D(0) \end{bmatrix}, \quad (1.30)$$

where \mathbf{m}_U , \mathbf{m}_D , \mathbf{n}_U and \mathbf{n}_D are the sub-matrices, in partitioned form, derived from matrix \mathbf{D} and the term $\hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_z & \hat{u}_x \end{bmatrix}^{\mathbf{T}}$. From Equation 1.30 one obtains

$$\mathbf{n}_D \mathbf{V}_D(z=0) = 0, \quad (1.31)$$

which has two solutions. Either $\mathbf{V}_D = 0$, which does not lead one any further, or

$$\det(\mathbf{n}_D) = 0. \quad (1.32)$$

It is important to notice that this solution may occur without any external source applied to the system. This explains why Rayleigh waves may be understood as a resonant mode of an elastic halfspace. Equation 1.32 leads to the dispersion relation for Rayleigh waves (Rayleigh, 1887), relating ω and k , such as

$$\left[2\mu k^2 - \rho\omega^2\right]^2 + 4\mu^2 k^2 \left(\frac{\omega^2}{V_P^2} - k^2\right)^{\frac{1}{2}} \left(\frac{\omega^2}{V_S^2} - k^2\right)^{\frac{1}{2}} = 0. \quad (1.33)$$

Equation 1.33 has a solution, only if $k^2 > \frac{\omega^2}{V_P^2}$ and $k^2 > \frac{\omega^2}{V_S^2}$.

One can define

$$k = \frac{\omega}{c_\phi}, \quad (1.34)$$

where c_ϕ is the phase velocity. This enables to simplify Equation 1.33 to yield

$$\left[2 - \left(\frac{V_R}{V_S}\right)^2\right]^2 - 4\left[1 - \left(\frac{V_R}{V_P}\right)^2\right]^{\frac{1}{2}} \left[1 - \left(\frac{V_R}{V_S}\right)^2\right]^{\frac{1}{2}} = 0, \quad (1.35)$$

where the phase velocity of Rayleigh waves ($c_\phi = V_R$) is independent of angular frequency ω , illustrating how Rayleigh waves in an infinite homogeneous half space are not dispersive.

The velocity of Rayleigh waves (V_R) is smaller than the velocity of S-waves and P-waves ($V_R < V_S < V_P$). Equation 1.35 can be solved numerically to determine V_R , if the relation V_S/V_P is known, for example from the Poisson ratio ν . A good estimation of the relationship between V_R and V_S is given by Viktorov (1967), which shows how V_R is

very much dependent on V_S and to a lesser extent dependent on the Poisson ratio ν and therefore also V_P . A gross and quite often used approximation of V_R is given by

$$V_R \approx 0.9V_S . \quad (1.36)$$

Surface waves in general have lower velocities compared to body waves. Travelling through the near surface, the generally lower velocities of this weathered region further slows down the surface waves. On the oil and gas exploration scale, surface waves are often recorded at the same time as reflected body waves that have travelled much deeper and covered longer distances. This is why removing surface waves is crucial for analyzing these reflected waves in conventional seismic image processing.

1.1.3 Amplitude of Rayleigh waves with depth

Now that the boundary conditions have been applied to constrain the system of equations, one can reconsider the evolution of stress-displacement vector \mathbf{B} with depth z , such as

$$\mathbf{B}(z) = \mathbf{D}\mathbf{V}(z) , \quad (1.37)$$

which is equivalent to

$$\begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_U & \mathbf{m}_D \\ \mathbf{n}_U & \mathbf{n}_D \end{bmatrix} \begin{bmatrix} \mathbf{V}_U(z) \\ \mathbf{V}_D(z) \end{bmatrix} . \quad (1.38)$$

The assumption that there is no upgoing wavefield $\mathbf{V}_U(z) = 0$, following the radiation condition (see Equation 1.29), leads to

$$\hat{\mathbf{u}}(z) = \begin{bmatrix} \hat{u}_z \\ \hat{u}_x \end{bmatrix} = \mathbf{m}_D \mathbf{V}_D(z) . \quad (1.39)$$

From Equation 1.25, $\mathbf{V}_D(z)$ can be computed by

$$\mathbf{V}_D(z) = \mathbf{T}_D \mathbf{V}_D(0) , \quad (1.40)$$

where the transmission operator \mathbf{T}_D from Equation 1.26, from 0 to depth z is

$$\mathbf{T}_D(0 \rightarrow z) = \begin{bmatrix} e^{-|q_P|z} & 0 \\ 0 & e^{-|q_S|z} \end{bmatrix}. \quad (1.41)$$

From Equations 1.39 and 1.40 one can conclude that

$$\hat{\mathbf{u}} = \mathbf{m}_D \mathbf{T}_D(z) \mathbf{V}_D(0). \quad (1.42)$$

As the terms in the expression can be evaluated, one finally obtains an expression for the displacement vector \mathbf{u} for Rayleigh waves, for a given depth z , giving

$$\hat{\mathbf{u}} = \begin{bmatrix} -q_P e^{-|q_P|z} & -k e^{-|q_S|z} \\ k e^{-|q_P|z} & -q_S e^{-|q_S|z} \end{bmatrix} \mathbf{V}_D(0). \quad (1.43)$$

The term $\mathbf{V}_D(0)$ is an amplitude factor which becomes explicit when a source term is introduced in Equation 1.12, making it possible to evaluate $\hat{\mathbf{u}}(0)$. The solution $\mathbf{u}(x, z, t)$ in the time-offset domain (t, x) can then be obtained by applying an inverse Fourier transformation on $\hat{\mathbf{u}}(k, z, \omega)$.

To obtain an approximate quantitative indication of the change in amplitude with depth for Rayleigh waves, one can analyse the term u_z , giving the displacement of the particle motion with depth. In the following the exponential decrease of the amplitude, an important characteristic of surface waves is derived.

As the term $\mathbf{V}_D(0)$ is a constant that does not vary with depth, one can simplify Equation 1.43 to

$$\hat{u}_z = q_P e^{-|q_P|z} + k e^{-|q_S|z}. \quad (1.44)$$

For Rayleigh waves, $k = k_R$, where $k_R = \frac{\omega}{V_R}$. The values of the vertical wavenumbers $q_P(k_R)$ and $q_S(k_R)$ are therefore

$$q_P(k_R) = i \sqrt{k_R^2 - \frac{\omega^2}{V_P^2}}, \quad (1.45)$$

and

$$q_S(k_R) = i \sqrt{k_R^2 - \frac{\omega^2}{V_S^2}}. \quad (1.46)$$

This gives the following for the displacement with depth

$$\hat{u}_z = q_P(k_R)e^{-z\left|i\sqrt{k_R^2-\frac{\omega^2}{V_P^2}}\right|} + k_Re^{-z\left|i\sqrt{k_R^2-\frac{\omega^2}{V_S^2}}\right|}, \quad (1.47)$$

$$\hat{u}_z = q_P(k_R)e^{-z|i|\left|\omega\sqrt{\frac{1}{V_R^2}-\frac{1}{V_P^2}}\right|} + k_Re^{-z|i|\left|\omega\sqrt{\frac{1}{V_R^2}-\frac{1}{V_S^2}}\right|}. \quad (1.48)$$

Assuming that $V_P \gg V_S$, the value of the first term on the right-hand side of Equation 1.48 can be assumed to be significantly smaller than the value of the second term. One can therefore focus on the second term only, to obtain an approximative solution. Using the relation in Equation 1.36 one obtains

$$\hat{u}_z \simeq k_Re^{-z\left|\frac{\omega}{V_S}\sqrt{\frac{1}{0.92^2}-1}\right|}, \quad (1.49)$$

$$\hat{u}_z \simeq k_Re^{-z\left|\frac{0.48\omega}{V_S}\right|}, \quad (1.50)$$

$$\hat{u}_z \simeq k_Re^{\frac{-\pi z}{\lambda_S}}, \quad (1.51)$$

where $\omega = \frac{V_S 2\pi}{\lambda_S}$.

A rapidly decaying amplitude with depth is a significant characteristic for both Rayleigh and Love waves, as illustrated in Figure 1.2. As surface waves propagate through the near surface, they will only contain information on the properties of the medium in this shallow region, and not be influenced by changes at depth. Considering Equation 1.51, the displacement of surface waves at different frequencies will decrease with depth z but will also depend on the wavelength λ . Therefore surface waves with longer wavelengths, and consequently lower frequencies, will be less attenuated at a certain depth than surface waves with shorter wavelengths, and of higher frequency content. As such, when using the information contained in surface waves to image the subsurface, the maximum depth of the investigation, is limited by the penetration depth of the surface waves equivalent to about one wavelength (Grant and West, 1965; Foti, 2000; Gedge and Hill, 2012), and can be determined by the frequency content of the signal. The surface waves will not be influenced by properties of the layers below this depth.

Figure 1.2a illustrates how the amplitudes of both vertical and horizontal components of Rayleigh waves decrease with depth. As the vertical and horizontal motion of Rayleigh waves are exactly 90° out of phase, vertical component is greater than the horizontal one. The horizontal displacement decreases rapidly with depth, and at a depth of about $z/\lambda =$

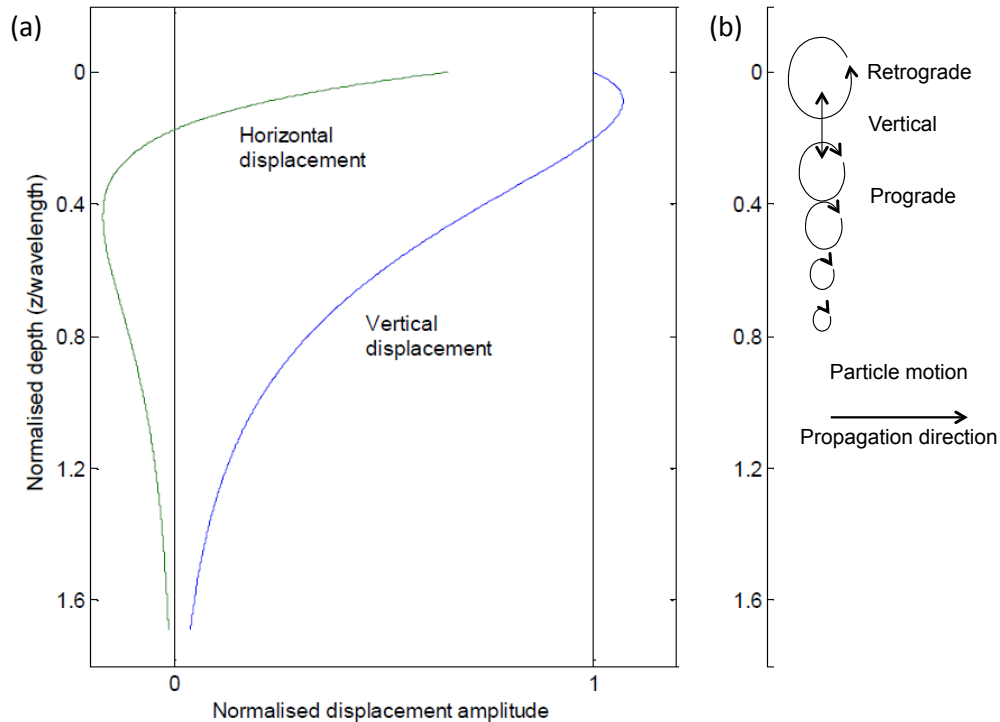


Figure 1.2: The graph in (a) depicts the vertical (blue) and horizontal (green) displacement amplitudes for a Rayleigh wave in an isotropic, homogeneous, elastic half-space, as a function of depth z , given in terms of the wavelength. The amplitudes are normalized against the amplitude of the motion perpendicular to the propagation direction at the surface. Illustrations of the partial motion of a Rayleigh wave corresponding to the respective changes with depth, starting as a retrograde ellipse, to linear vertical motion and finally becoming a prograde ellipse are shown in (b). Figure modified from [Gedge and Hill \(2012\)](#).

0.2, the displacement changes polarity. The vertical displacement increases slightly before reaching a maximum at a depth of about 0.076λ and then decreases but does not change polarity.

Rayleigh waves have a *skin depth* limited to approximately 0.94λ , at which the vertical component will have decayed to about 20% of its surface value and the horizontal component to about 10% ([Strobbia, 2002](#); [Foti et al., 2014](#)). The larger the wavelength of the Rayleigh wave, the deeper the skin depth and the greater the penetration depth of the wave within the medium.

The relation between the vertical and horizontal displacement components changes

with time and depth, as well as the particle motion, both shown in Figure 1.2b. Initially the motion is that of a retrograde ellipse, but as the horizontal component of the Rayleigh wave goes to zero, at a depth of $\lambda/5$, the particle motion becomes vertical linear. As the horizontal displacement becomes negative the particle motion becomes elliptical again but with a prograde direction, as the horizontal displacement is now in the opposite direction.

The amplitude of surface waves will also decrease due to geometrical spreading, caused by an increase in the surface of the wavefront, which attenuates the 3D wave propagation. However, different to body waves which propagate energy both horizontally and vertically, and have a geometric attenuation factor of $1/r$ where r is the distance from the source, the geometrical spreading of surface waves is reduced. Since the energy of surface waves only decays horizontally, the amplitudes decrease by $1/\sqrt{r}$ during propagation (Woods, 1968). Due to this difference, surface waves often have much higher amplitudes compared to body waves. Their high energy generally dominates seismic recordings, especially at distances greater than on the order of one wavelength away from the source (Foti et al., 2014). Furthermore, as they remain restricted to the near surface region, they are generally the main cause of the shaking felt during an earthquake event.

It should be noted that the derivation to estimate \hat{u}_z in this section, does not consider a dependence of the Rayleigh wave velocity V_R on the frequency ω . However in the case that the velocity is not homogeneous in the half-space considered, Rayleigh waves will become dispersive. The velocity of Rayleigh waves will therefore also depend on the frequency giving $V_R(\omega)$, though this will still result in an overall exponential decay with depth.

1.1.4 Dispersion of surface waves

One important and useful characteristic of surface waves is their dispersive property (Lamb, 1904; Love, 1911). This occurs when the medium considered is no longer homogeneous, or when the topography of the free surface is no longer flat. In such cases the phase velocity of the surface waves depends on their frequency (Aki and Richards, 2002).

Figure 1.3 shows how waves of different wavelengths propagate with different phase velocities due to the changing properties of the medium. Since the wavelengths are proportional to the inverse of the frequency, the phase velocity is shown to be frequency dependent. Waves of shorter wavelength and high frequency sample the first layer of the medium, while waves with longer wavelengths and lower frequencies sample the deeper

structures of the medium (Babuska and Cara, 1991). As velocities generally increase with depth, the longer wavelengths, which penetrate deeper, will generally also travel faster. The dispersion of the surface waves can therefore provide information on velocity changes in the subsurface (Rix, 1988).

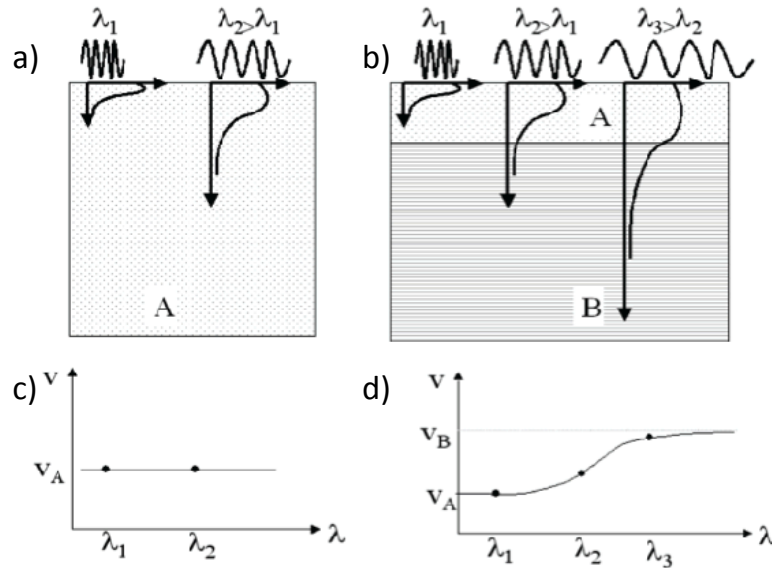


Figure 1.3: Schematic of how the phase velocity is dependent on the wavelength of the propagating wave. For the homogeneous medium depicted in (a) all wavelengths, each associated to one frequency, propagate with velocity A as shown on graph (c). For the layered medium shown in (b), the phase velocity, which now has a more complex frequency relation, is different for each wavelength on the graph (d). Figure taken from Strobbia (2002).

By plotting the variation of the phase velocity with frequency, one can observe the dispersion behaviour of surface waves. In stratified media, surface waves can contain several discrete modes for a single given frequency (see Figure 1.4), as they all satisfy the wave equation and the free surface condition. The slowest propagating mode is called the fundamental mode (Sheriff and Geldart, 1995). All higher modes are harmonics of the fundamental mode, of which an infinite number exist. Surface wave phases define dispersion curves and the fundamental mode dispersion curve is the most often used phase measurement.

For a homogeneous halfspace medium, only the fundamental mode is present. For a weakly heterogeneous medium, such as stratified layers, higher modes propagate as well, but the fundamental mode remains the most energetic. For a complex medium, higher modes may become very energetic, sometimes more than the fundamental mode, making

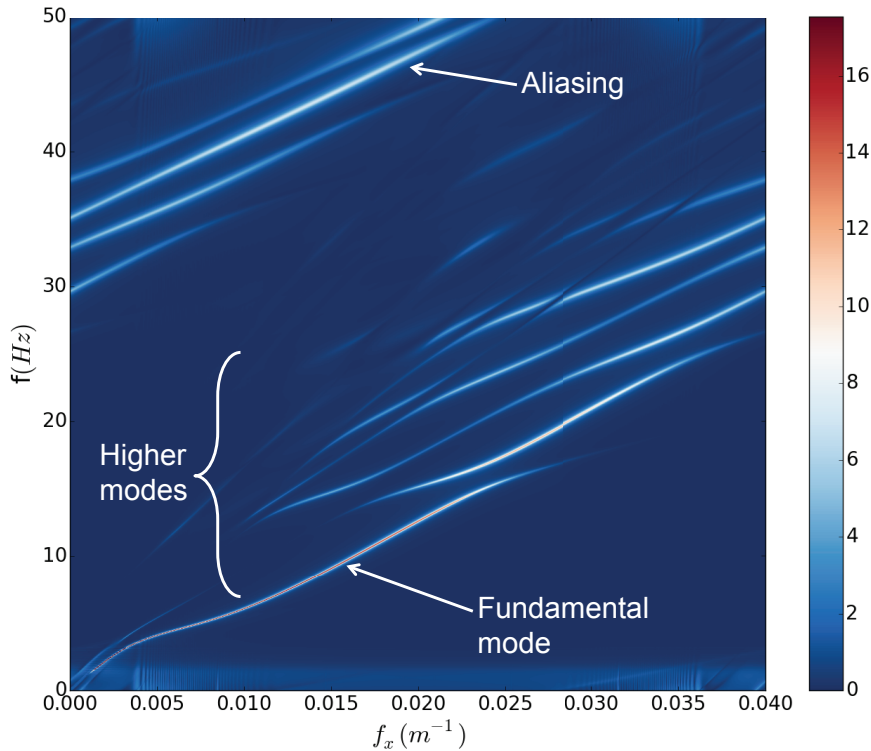


Figure 1.4: Synthetic example of dispersion curves, shown in the frequency-wavenumber (ω, k) domain, for a simple layered medium. The fundamental mode is the most energetic and can be distinguished relatively clearly. The higher modes appear to merge one with another, although they do not intersect. For noisy datasets or when considering a more complex medium, the higher modes, and sometimes even the fundamental mode, could be easily misinterpreted. Aliasing of the data may further contribute to problems with automated picking algorithms. Courtesy of Jean-Baptiste Laffitte, 2015.

the distinction between them difficult (Gucunski and Woods, 1992). At far offsets, the separation between the different modes in the time domain becomes more evident as the modes propagate at different velocities through the medium (Ernst and Herman, 2000) and the *wave train* of the surface waves is longer.

Picked dispersion curves can provide valuable information on the velocity structure of the subsurface, which is of interest to geophysics applications from small scale engineering purposes to large scale global seismology investigations. Seismic acquisitions on such scales have been deployed with the aim of recording surface waves travelling through the subsurface. From phase measurements of the dispersive characteristics of a site, it is possible to use inversion to estimate the subsurface soil properties. Inversion techniques are used to obtain a best fit between the actual data and modeled data, to

retrieve an estimation of the velocity of the medium with depth. In the following two sections of this chapter, two different inversion techniques, which are applied to surface waves, are discussed. The first one takes advantage of surface wave dispersion curves, by picking and inverting phases of the fundamental mode, and sometimes also those of higher modes, to obtain a local vertical (1D) velocity profile. The second technique aims to invert the whole recorded seismogram to obtain a high resolution, multiparameter, multidimensional image of the subsurface.

1.2 Inversion of dispersion curves

1.2.1 Surface wave analysis methods

Using surface wave analysis methods to obtain information on the subsurface is not a new idea. Such methods have been promoted for both near surface geotechnical applications (Socco et al., 2010) and global seismology applications (Romanowicz, 2002). Surface waves, with their high amplitudes, have a good signal-to-noise ratio (SNR) compared to other wave types, and are therefore an attractive source of information for the extraction of subsurface properties.

Surface waves have been used by seismologists to characterize the Earth's interior since the 1920's, obtaining information on subsurface properties along the source-receiver path, as well as on the earthquake source. Romanowicz (2002) provides an extensive review. For lithospheric imaging, the group velocity can be measured and inverted for. Dispersion curves are obtained by picking surface wave phases, in the time-offset (t, x) domain, for narrow frequency band of data. Since larger distances are considered, surface waves modes can be more easily distinguished and phases are better separated. Numerous applications of surface wave tomography have been performed for crustal and mantle imaging (Nishimura and Forsyth, 1989; Stutzmann, 1993; Macquet, 2014).

With the introduction of the spectral analysis of surface waves (SASW) method (Heisey et al., 1982; Nazarian and Stokoe, 1984; Stokoe et al., 1994), there was an increased momentum in the development of techniques to exploit surface waves, especially for applications on the near surface geotechnical and engineering scale. The method allowed a faster and theoretically grounded approach, that took advantage of the growth in computer technology at the time. Furthermore, the use of seismic imaging as a non-invasive and low cost technique, can be a critical advantage for certain applications,

avoiding the need for boreholes or probes. In the SASW method, a configuration of two receivers is required, with several active sources exerted on the ground surface. Generally only vertical sources and vertical component receivers are used in such experiments, since the vertical displacement of surface waves with depth is greater than the horizontal one (Figure 1.2). Often, only Rayleigh waves are exploited, being the easiest to generate and record with readily available equipment (Socco et al., 2010). The fundamental dispersion curve of the surface waves can be derived from the phase delay between two seismic trains recorded at the two receivers, for each frequency component. Yet problems may occur when strong higher modes exist in the data. Because the method suffers from a limitation on the frequency band, the test is repeated with several different receiver configurations to better estimate the dispersion curves. For global seismology applications, the use of noise (generated from passive sources) was developed, exploiting low-frequency Rayleigh waves in the always existing ambient wavefield (Shapiro and Campillo, 2004).

The SASW method was later extended to a multiple receiver approach with the development of the multichannel analysis of surface waves (MASW) method (McMechan and Yedlin, 1981; Gabriels et al., 1987; Park et al., 1999; Xia et al., 1999). With multiple receivers, the calculation of the dispersion curves is more robust and accurate, and the data processing faster. The data traces are normalized in the frequency domain, before being transformed by the application of a slant-stack $\tau - p$ transform and subsequent Fourier transform on the time axis, to visualise the dispersion curves in the frequency-slowness (ω, p) domain (McMechan and Yedlin, 1981; Luo et al., 2008) or frequency-velocity (ω, c) domain (Xia et al., 1999; Strobbia, 2002).

Surface waves are most sensitive to the V_s parameter of the subsurface, although several other properties influence the surface wave velocities, such as V_p , Poisson ratio, density, and layer thickness (Nazarian, 1984; Xia et al., 1999; Foti and Strobbia, 2002). Information on these parameters, can therefore be obtained from the inversion of surface wave dispersion curves, as illustrated in Figure 1.5. Often only a vertical profile of V_s with depth is obtained, while V_p and density are fixed a priori, to reduce the number of unknowns in the ill-posed inversion problem. The medium is frequently assumed to be layered with depth for the inversion, with number of layers given as an input parameter for the inversion. The number of layers can be overestimated to obtain a better resolution of the V_s profile, although this can also be computationally expensive. Otherwise the number of layers may also be a model parameter which is inverted for (Bodin and Sambridge, 2009). The V_s property is an important lithological and geotechnical parameter to retrieve, useful for engineering applications as it does not have a water-masking

effect in saturated media.

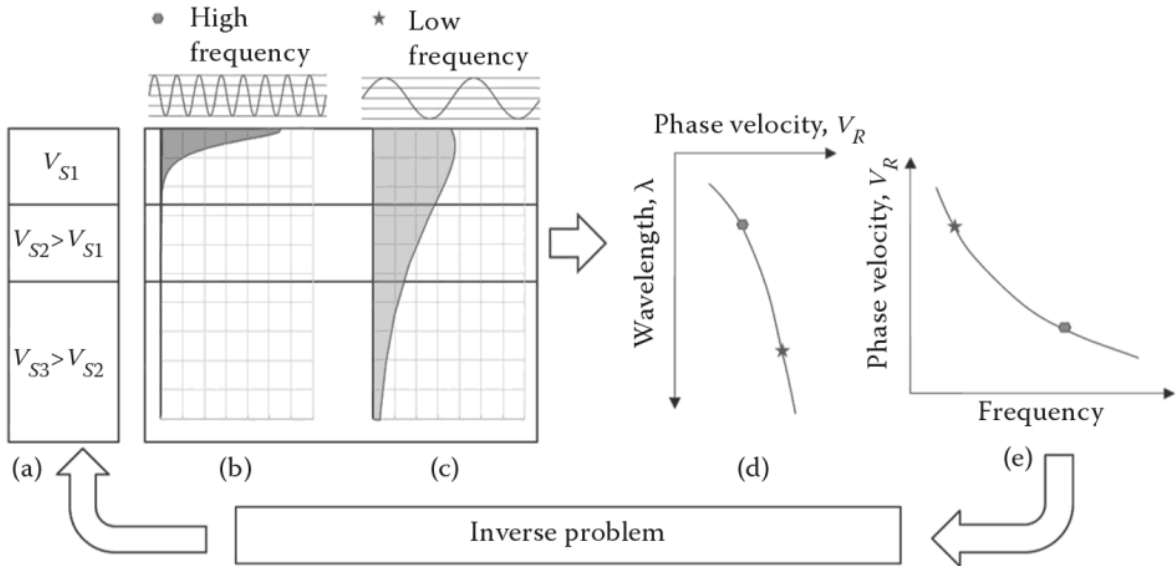


Figure 1.5: Schematic for dispersion curve inversion. As waves of different wavelengths and frequencies sample the medium at different depths in (b) and (c), dispersion curves can be obtained such as (e), and these can be inverted to obtain information on the velocity of the medium with depth given in (a). Figure taken from [Foti et al. \(2014\)](#).

The dispersion curves need to be accurately identified and picked, and are then compared to simulated dispersion curves of an estimated model, with the aim of minimizing the difference between the two. Usually the least-squares misfit is computed and minimized. Several inversion algorithms have been implemented for surface wave analysis. Local search methods are often implemented ([Xia et al., 1999](#)). Brute global methods such as Monte Carlo, although effective ([Maraschini and Foti, 2010](#)), are computationally expensive, as the whole model domain is sampled to find the global minimum. Semi-global methods such as simulated annealing ([Beatty et al., 2002](#)), genetic or neighborhood algorithms ([Wathelet et al., 2004](#); [Wathelet, 2008](#)) are often used to optimize the problem and reduce the model space investigated.

[Socco et al. \(2010\)](#) provides a comprehensive review on the surface wave analysis method. Although the method is now conventionally used to obtain near subsurface properties, it also has several limitations, such as the picking of dispersion curves and an intrinsic locally laterally invariant assumption.

1.2.2 Picking dispersion curves

To invert the dispersion curves, the phase velocity values associated with different frequencies first need to be picked. This requires someone to visually identify and manually pick points along the dispersion curves. Although dispersion curves of different modes will never cross each other, they can often appear to merge with each other (Panza et al., 1972), and are non-continuous, only appearing at certain frequencies, as shown in Figure 1.4 for a synthetically computed example. This makes it difficult to identify the fundamental and higher modes (Stutzmann and Montagner, 1994).

Often the data are visualized in the frequency-wavenumber (ω, k) domain to better enable the distinction of different modes (see Figure 1.6). For data acquired using a linear array of receivers, they can be transformed by the application of a Fourier transform on both the time and offset axis (Nolet and Panza, 1976). The stacking of the data in the (ω, k) domain also improves the signal-to-noise ratio (SNR). Prior to the transformation the traces are often normalized to reduce the contribution of large amplitude signals in proximity to the source. The (ω, k) domain is a natural approach for the identification of surface waves, as the energy of the wavefield is located in eigenvalues, which are represented by lines in the (ω, k) domain. Of course the energy of noise will also be present, and the energy density maxima need to be picked. The resulting dispersion curves are then transformed to the frequency-velocity (ω, c) domain through the relation $c(\omega) = \omega/k(\omega)$.

Yet difficulties with mode identification may still occur, especially for noisy datasets or when the data are not properly sampled and aliasing is present (see Figure 1.4). For complex targets, which are often also of higher interest to investigate, features such as the topography, lateral heterogeneities and velocity inversions, may lead to complex dispersion curves. In many studies only the fundamental mode is inverted, being easier and cheaper to implement. For cases where it is the most energetic mode, such an implementation can give robust and reliable results (Gucunski and Woods, 1992). Yet the inversion of higher modes can play an important role, especially in cases of more complex geology, and they sometimes need to be considered to avoid misinterpretation (Beatty et al., 2002; Ryden et al., 2004; Maraschini et al., 2010). Extracting higher modes during the processing often increases the sensitivity and resolution obtained during inversion (Stutzmann, 1993). Dense acquisitions with large offsets, such as those for oil and gas exploration are sometimes more suited for inversion with higher modes (Socco and Strobbia, 2004).

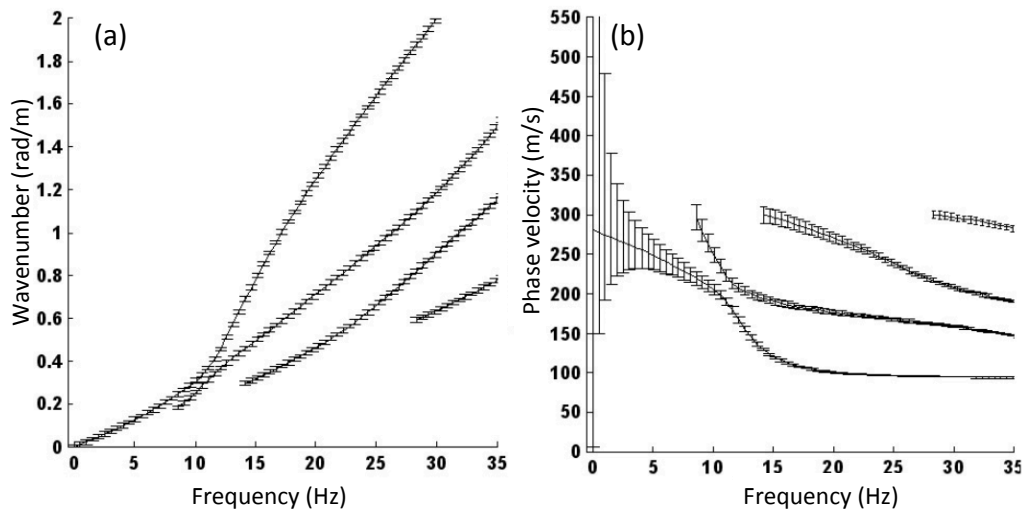


Figure 1.6: Dispersion curves for a synthetic three-layered model. The uncertainty distribution for picking the spectral maxima of the fundamental and higher modes is shown for the frequency-wavenumber ($\omega - k$) domain (a), and for the frequency-phase velocity ($\omega - c$) domain (b). Although the different modes appear to merge, they do not intersect. Image modified from [Socco and Strobbia \(2004\)](#).

The human interpretation and the labor involved in manual picking has led to (semi-) automatic strategies for a more objective identification and selection, where the spectral maxima for the fundamental (or higher modes) is computed and traced along the dispersion curve. Such automated processes include algorithms based on semblance ([Neidell and Taner, 1971](#); [Parker and Hawman, 2012](#)) used for near surface studies, as well as a *moving window analysis* and the *multiple filter technique*, generally used for global seismology studies ([Dziewonski et al., 1969](#); [Pedersen et al., 2003](#)). However these methods often fail for noisy real data obtained for complex near-surface targets, where searching for the location of the spectral maxima is not sufficient to successfully distinguish the different modes.

An alternative approach is to invert for the full waveform of the signal, avoiding human interpretation and picking issues, and naturally inverting implicitly for all modes present in the data. This approach is discussed in the next section [1.3](#).

1.2.3 Dealing with lateral variations

Surface wave analysis methods based on dispersion curve inversion, presume a layered medium with depth, with no significant lateral variations within each layer, which are

intrinsically assumed to be homogeneous within the range of the experiment. When lateral variations are present in the subsurface, the fundamental mode may be more difficult to pick, as it is strongly weakened by heterogeneities. Lateral heterogeneities have been shown to impact surface wave phase velocities, leading to false depth-related dispersion with a 1D assumption during surface wave analysis (Kennett and Yoshizawa, 2002; Strobbia and Foti, 2006). For example, two fundamental modes may be present in the dispersion curve spectrum, corresponding to two different areas of a laterally varying region. Multioffset phase analysis methods have been developed to help detect and locate strong lateral variations in velocity, while aiming to maintain the data quality and robustness from using multiple sources and receivers (Vignoli et al., 2011).

A multiple laterally invariant approach inverting local dispersion curves can be applied to obtain better results. This technique often relies on a spatial windowing of the data. The local dispersion curves are inverted for local Vs profiles with several disjointed inversions, to then generate a larger interpolated 2D Vs section and retrieve smooth lateral variations (Bohlen et al., 2004). Boiero and Socco (2010) propose a sliding window approach for a laterally constrained algorithm, to better image lateral variations than can be obtained by smoothing over individual layered medium (1D) inversions. The technique relies on the minimization of a common misfit function for the whole dataset. On the global seismology scale, a modal approach that consists in splitting the wavefield in lateral and vertical components, has been proposed to deal with smooth lateral variations (Maupin and Kennett, 1987; Maupin, 1988; Ernst et al., 2002).

For strong lateral contrasts however, surface wave analysis methods based on phase measurements generally break down. There is a need for an alternative approach that inverts for the 2D or 3D subsurface directly from seismograms without phase picking. Full waveform inversion (FWI) is one such method that allows to use surface waves to go beyond a layered medium assumption and will be detailed in the next section 1.3.

1.3 Full Waveform Inversion

With the increase in computer power available, and as our ability to understand complex non-linear inversion has developed, Full Waveform Inversion (FWI) has become an increasingly feasible and attractive method for the oil and gas industry. It has the ability to outperform other velocity model building techniques in providing high resolution and dependable models for migration, taking advantage of the full waveforms of the acquired

data. Furthermore, being a data-driven method it has a quantitative characteristic, and can potentially provide multiparameter models, of V_p , V_s , density, anisotropy and attenuation, that can help with interpreting and understanding the subsurface properties.

The oil and gas industry has the resources to acquire increasingly dense datasets, with long offsets and low frequencies, that are required for successful FWI. However, although FWI is now widely implemented in seismic exploration workflows, it is less robust than many conventional velocity model building techniques. Due to the strong non-linearities of the problem, convergence towards a sufficiently good result is not always easy.

Yet the advantages of moving from dispersion curve inversion techniques explained in previous section 1.2, towards FWI to image the near surface are evident. FWI allows to go beyond the (multi) layered medium limitations and does not involve any picking that would require human interpretation, even though difficulties related to surface waves characteristics may need to be overcome, such as the essential forward scattering property of surface waves.

1.3.1 Introduction to Full Waveform Inversion

FWI is a high resolution technique, used to obtain quantitative images of model parameters in the subsurface by modelling the physics of the measurement. Since the early 80's (Lailly, 1983; Tarantola, 1984b; Mora, 1988) the FWI method has developed with the computer resources available and has been used in a variety of applications. A schematic detailing the conventional FWI workflow can be found in Figure 1.7. In this method the full waveform of an observed seismogram is matched to that of a corresponding synthetic seismogram. The synthetic seismogram is computed by propagating the wave equation through estimated model parameters of the subsurface, and recording the propagated wavefield at receiver points, simulating the real acquisition scheme. All phases in the observed seismic traces are considered, and therefore different types of propagation need to be modeled to reproduce them in the synthetic dataset. The following two sections will describe the forward modelling and the inverse problem in more detail.

1.3.1.1 Elastic forward modelling

To generate synthetic datasets, an accurate forward modelling engine is required. Several options exist for discretising the wave equation for FWI schemes. They include finite

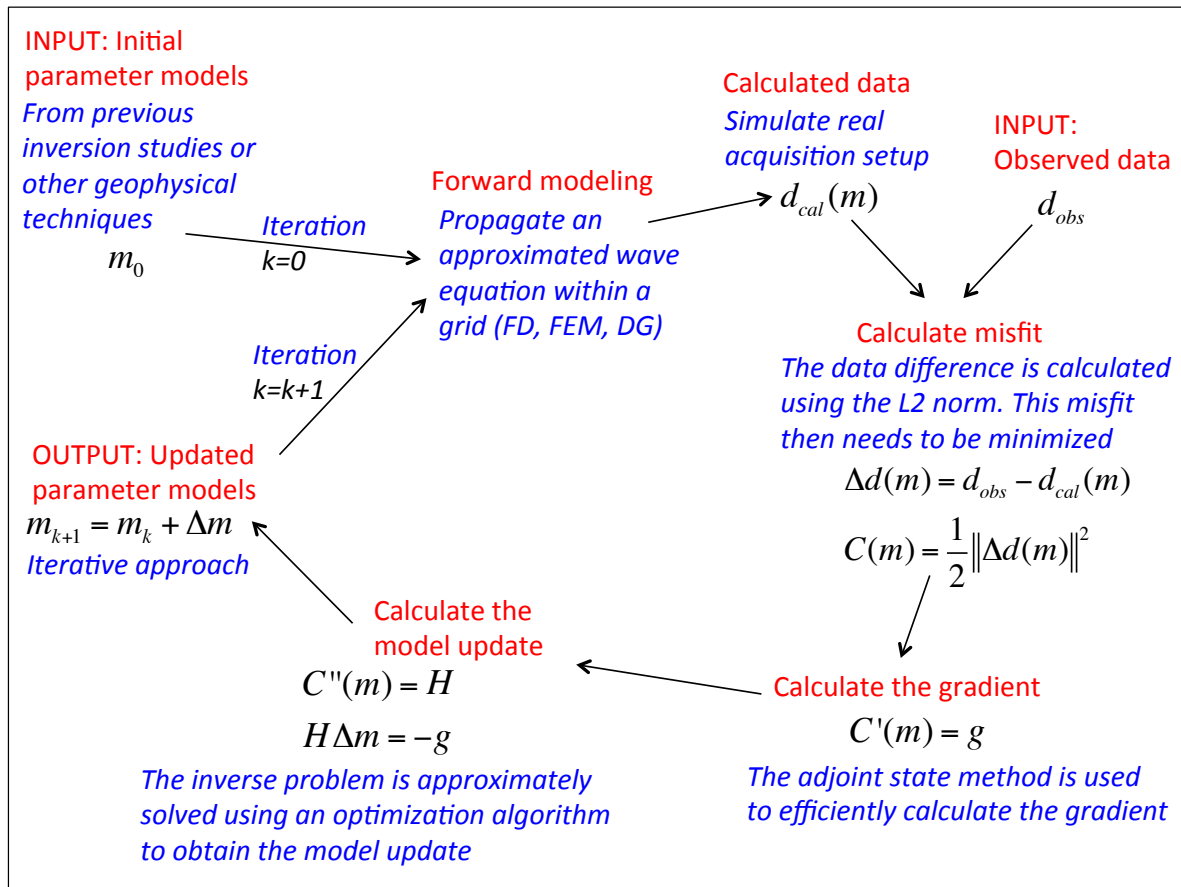


Figure 1.7: A generalized workflow for conventional FWI. The method requires a priori information in the form of initial models, to compute synthetic data which are then compared to the observed data. The data difference is minimized by a gradient-based optimization algorithm, for which the gradient is computed, and the inverse problem to obtain the model update is approximately solved. The model parameters are updated in an iterative approach until a minimum in the data difference is obtained.

differences (FD), finite element method (FEM), spectral element method, and discontinuous Galerkin (DG), as well as boundary integral methods when considering stratified media. Although discontinuous Galerkin would be the most adapted when considering surface waves, suitable for simulating a complex topography and strong contrasts in the medium, it is also the most costly and difficult to implement. The work presented in this thesis employs a finite differences method, which can be simply implemented, to discretize the elastodynamics equation for 2D wave propagation in the time domain. More details can be found in section 2.2.3. The current excessive computational demands for 3D elastic FWI makes 2D forward modelling the preferred choice for gaining experience

and code development, as it allows repeated runs.

The finite differences scheme is implemented with a Virieux-Levander staggered grid stencil (Virieux, 1986; Levander, 1988) with a fourth-order accuracy in space and a second-order accuracy in time $\mathcal{O}(\Delta x^4, \Delta t^2)$. The scheme is illustrated in Figure 1.8.

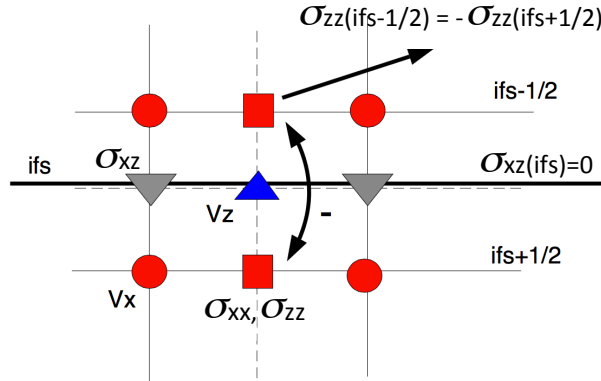


Figure 1.8: Implementation of the free surface boundary condition for the finite difference Virieux-Levander staggered grid stencil, using image theory. The free surface is along the $\sigma_{xz} = 0$ grid. Modified from (Operto et al., 2007).

Numerical modelling of the free surface is often difficult, especially when considering a complex topography as is often the case for land acquisitions. With finite differences, a staircase approximation can be used to discretise changes in the topography, but the number of points required for this is often very high, and demands a large computational cost, on the order of 60 points per wavelength (Bohlen and Saenger, 2006). The work in this thesis will only consider cases where a flat topography can be approximated. The free surface, which requires the normal stress to be null at the boundary, is thus implemented using image theory (Virieux, 1986; Levander, 1988). The free surface boundary is implemented along the $\sigma_{xz} = 0$ grid of the staggered grid stencil considered, as illustrated in Figure 1.8. Due to the staggered characteristic, the V_z grid is along the boundary, while the V_x grid is half-a-grid interval below. Knowledge of the wavefield above the boundary is therefore used to interpolate the missing values. The remaining sides of the model have absorbing boundary conditions, implemented with perfectly matched layers (PML) procedure, to imitate an infinite medium (Bérenger, 1994; Hastings et al., 1996; Hustedt et al., 2004). The PML are therefore augmented to the right, left and bottom of the finite difference grid.

For numerical stability, the time discretization must satisfy the Courant-Friedrichs-

Lewy (CFL) stability condition (Courant et al., 1967; Richtmyer and Morton, 1967). In the case of 2D propagation with an accuracy of $\mathcal{O}(\Delta x^4, \Delta t^2)$, the time step Δt is required to satisfy the following equation

$$\Delta t \leq 0.606 \frac{\Delta x}{c_{max}}, \quad (1.52)$$

where Δx is the step-size of the finite difference grid, and c_{max} is the maximum velocity present in the model parameters (Levander, 1988). Regarding the spatial discretization, at least 10 grid points per minimum wavelength should be considered for accurate modelling of body and surface waves.

To improve the efficiency of the forward modelling, the algorithm used for the work presented in this thesis, is implemented with a parallel decomposition on the sources. Each source relies on its own Message Passing Interface (MPI) process, running on a separate core of the machine. The number of required MPI communications is small, limited to only the global misfit summation and the gradient calculation, at each iteration.

1.3.1.2 Inverse problem

The inverse problem generally comprises of a misfit functional which needs to be minimized by solving an optimization problem. The data residual quantifies the differences between the observed and synthetic data giving the vector $\Delta \mathbf{d}$, defined as

$$\Delta \mathbf{d} = (\mathbf{d}_{obs} - \mathbf{d}_{cal}(\mathbf{m})), \quad (1.53)$$

where \mathbf{d}_{obs} is the observed data and \mathbf{d}_{cal} is the synthetic data calculated using the estimated model \mathbf{m} . There is an implicit summation over all sources and receivers. The data may be for either or both vertical and horizontal receiver components.

The misfit value is conventionally given by the L2 norm of the data residual (Tarantola, 1987) defined as

$$C(\mathbf{m}) = \frac{1}{2} \int_0^T \Delta \mathbf{d}^\dagger \Delta \mathbf{d} dt, \quad (1.54)$$

where T is the recorded trace length, and the symbol † stands for the transpose operator. The misfit function is formulated here in the time domain, with an implicit summation over all sources and receivers, and is twice differentiable (see Figure 1.7). Preconditioning

or regularization terms may be included in the misfit to favor certain directions of search. Prior information from previous inversion studies of other geophysical methods may also be included to help drive the inversion in a certain direction. When considering data which includes surface waves, because of their generally higher amplitudes compared to other wave types (see section 1.1.3), they will dominate the misfit, and drive the inversion.

In the FWI scheme, the inverse problem is created by the non-linear relation between the observed data and the true model space. It is possible to set up an acquisition on the Earth's surface, and record seismic waves passing through the Earth, but not to directly obtain the physical parameters of the Earth from the recorded seismograms. Due to the large number of model parameters, global or semi-local stochastic methods such as neighbourhood or genetic algorithms are not yet feasible, especially if considering 3D geometries.

The solution of the inversion can be computed with linearized iterative methods that minimize the misfit. The method requires to start with an initial estimated model of the subsurface coming from other geophysical investigations. At each new iteration $k + 1$, the model update $\Delta \mathbf{m}^{k+1}$ is added to the previously estimated model \mathbf{m}^k , in the search for the local minimisation of the misfit function $C(\mathbf{m}^{k+1})$. This gives

$$C(\mathbf{m}^{k+1}) = C(\mathbf{m}^k + \Delta \mathbf{m}^{k+1}) . \quad (1.55)$$

When considering FWI with surface waves the model parameter that is the most sensitive, as seen for the inversion of dispersion curves in section 1.2, is the velocity V_s . The V_p and density model parameters may also possibly be inverted for and updated with FWI.

The minimization proceeds iteratively, through repeated calculation of a local gradient. The gradient is obtained by the derivative of the misfit function with respect to the model parameters. Considering the case of a small model perturbation, with the Born approximation (Born and Wolf, 1980; Hudson and Heritage, 1981; Coates and Chapman, 1990), the second-order Taylor expansion of the misfit function gives

$$\begin{aligned} C(\mathbf{m}^k + \Delta \mathbf{m}^{k+1}) &= C(\mathbf{m}^k) + \sum_{i=1}^n \frac{\partial C(\mathbf{m}^k)}{\partial m_i} \Delta m_i^{k+1} \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 C(\mathbf{m}^k)}{\partial m_i \partial m_j} \Delta m_i^{k+1} \Delta m_j^{k+1} + \mathcal{O}(\mathbf{m}^3) , \end{aligned} \quad (1.56)$$

where n is the size of the model parameter vector. The derivative of Equation 1.56 with respect to the model parameter m_l gives

$$\frac{\partial \mathcal{C}(\mathbf{m}^{k+1})}{\partial m_l} = \frac{\partial \mathcal{C}(\mathbf{m}^k)}{\partial m_l} + \sum_{i=1}^n \frac{\partial^2 \mathcal{C}(\mathbf{m}^k)}{\partial m_i \partial m_l} \Delta \mathbf{m}_i^{k+1} . \quad (1.57)$$

The solution is minimized, to ideally reach $\partial \mathcal{C}(\mathbf{m}^{k+1})/\partial m_l = 0$, at which the global minimum is attained. The estimated model is iteratively updated during the inversion to reach the true model. The model update is obtained from

$$\frac{\partial^2 \mathcal{C}(\mathbf{m}^k)}{\partial m^2} \Delta \mathbf{m}^{k+1} = - \frac{\partial \mathcal{C}(\mathbf{m}^k)}{\partial m} . \quad (1.58)$$

The first term of the left hand side in Equation 1.58 is the second derivative of the misfit function, also known as the Hessian matrix \mathbf{H} , the second term of the left hand side is the model update, and the right hand side is the gradient \mathcal{G} of the misfit function. The gradient provides the direction of descent towards the local minimum, while the Hessian contains the curvature information of the misfit function. Equation 1.58 can be written more simply as

$$\mathbf{H}^k(\mathbf{m}^k) \Delta \mathbf{m}^{k+1} = -\mathcal{G}^k(\mathbf{m}^k) , \quad (1.59)$$

also known as the normal equation, and illustrated in the FWI workflow in Figure 1.7. The algorithm used in this study implements the adjoint state method (Chavent, 1974; Lailly, 1983; Tromp et al., 2005; Plessix, 2006) for optimal efficiency in the calculation of the gradient. In the time domain, the gradient can be described by zero-lag correlation of the real forward wavefield with the back-propagated adjoint wavefield. A more detailed explanation on the formulation of the gradient with the adjoint state method can be found in section 2.2.5. The cost of computing the gradient is independent of the data complexity and only depends on the number of sources. It therefore has a computational cost on the same order as the cost of the forward modelling.

At each iteration, the locally linearized problem expressed by the Newton equation (1.59) needs to be solved to obtain the model update. This requires to recompute the

gradient vector and Hessian matrix at the new position for each new iteration, until convergence is reached. Due to the high computation costs involved, gradient-based optimization algorithms which does not require to explicitly compute and store the inverse Hessian are used. The *steepest descent* algorithm only considers the gradient, and is the simplest to implement, but an approximation of the influence of the Hessian is often helpful for FWI convergence (Pratt et al., 1998). The *truncated Newton* method computes an approximate solution of Equation 1.59 using a conjugate gradient algorithm and computing Hessian-vector products (Métivier et al., 2013). Although Hessian effects are considered, the method is more costly, and may not yet be appropriate for an application of elastic FWI. For this study all inversion results, unless otherwise specified, are computed using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) algorithm for optimization (Nocedal, 1980; Byrd et al., 1995; Nocedal and Wright, 1999). This quasi-Newton method has a reduced computational cost, and allows to speed up the inversion by computing an approximation of the product $-(\mathbf{H}^k)^{-1}\mathcal{G}^k$.

1.3.1.3 Initial models

As can be observed from Equation 1.55, an initial estimation of the parameter models is required for inversion. The availability of sufficiently good initial models is a key limitation for the FWI method. Furthermore, at the start of the inversion, the initial models need to explain the observed data to within half a wavelength or cycle-skipping occurs and the inversion will get stuck (see section 1.3.2).

Prior information on the model parameters to be inverted, needs to be obtained from other geophysical methods. First-arrival tomography and reflection tomography are conventionally used as velocity model building tools to provide a smooth initial V_p model for FWI when inverting only body waves.

When considering surface waves in an elastic FWI framework, lower resolution results from the inversion of dispersion curves may perhaps provide a smooth initial V_s model for the near surface region. The issue of a sufficiently good initial model is especially important for surface waves, as the inversion problem becomes highly non-linear, and the problem of cycle-skipping is greater than for body waves.

1.3.2 Local minima and the cycle-skipping issue

Full waveform inversion is a non-linear problem, generally formulated as a local descent method, to minimize the data difference. This can be viewed in Figure 1.9 where the ball will descend the slope of the local valley, by iterative calculation of the gradient, to minimize the data misfit. However the non-linearities cause the presence of local minima in the topography of the misfit function, and the location of the initial model is crucial for a successful inversion (Gauthier et al., 1986). The initial model must be close enough to the true model so that a descent path leading to the global minimum exists (such as for point B in Figure 1.9), otherwise the inversion will get stuck in a local minimum (such as for point A in Figure 1.9).

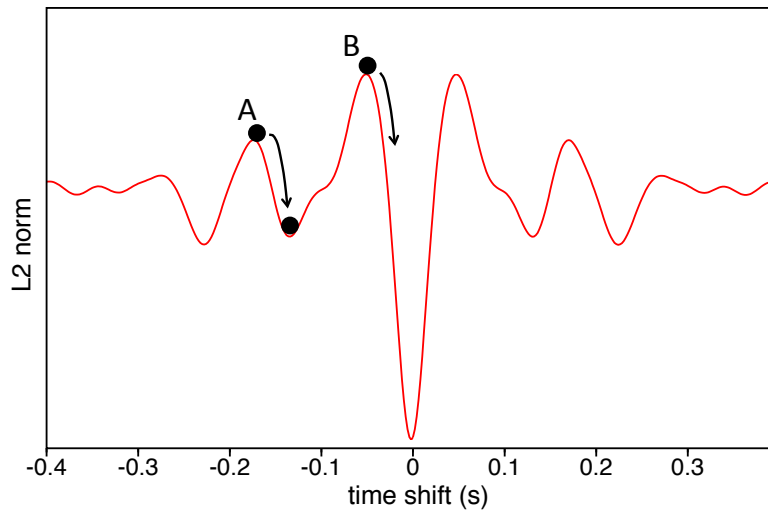


Figure 1.9: The shape of the classical difference-based L2 norm misfit function for a 1D example, produced from comparing two wavelets that are time shifted compared to each other, simulating a velocity difference between the true and estimated velocity models. When the initial model is at point A, the inversion will converge towards a local minimum and remain stuck there. However when the initial model is closer to the true model at point B, the inversion will converge towards the global minimum.

The presence of local minima may be explained by *cycle-skipping*. Figure 1.10 illustrates how these two issues are related. When two traces are compared they need to be within half a wavelength to avoid cycle-skipping, otherwise the wrong phases are matched up and a local minimum occurs. For slow surface waves propagating in the low velocity near surface, the problem of cycle-skipping is even greater due to their small wavelengths.

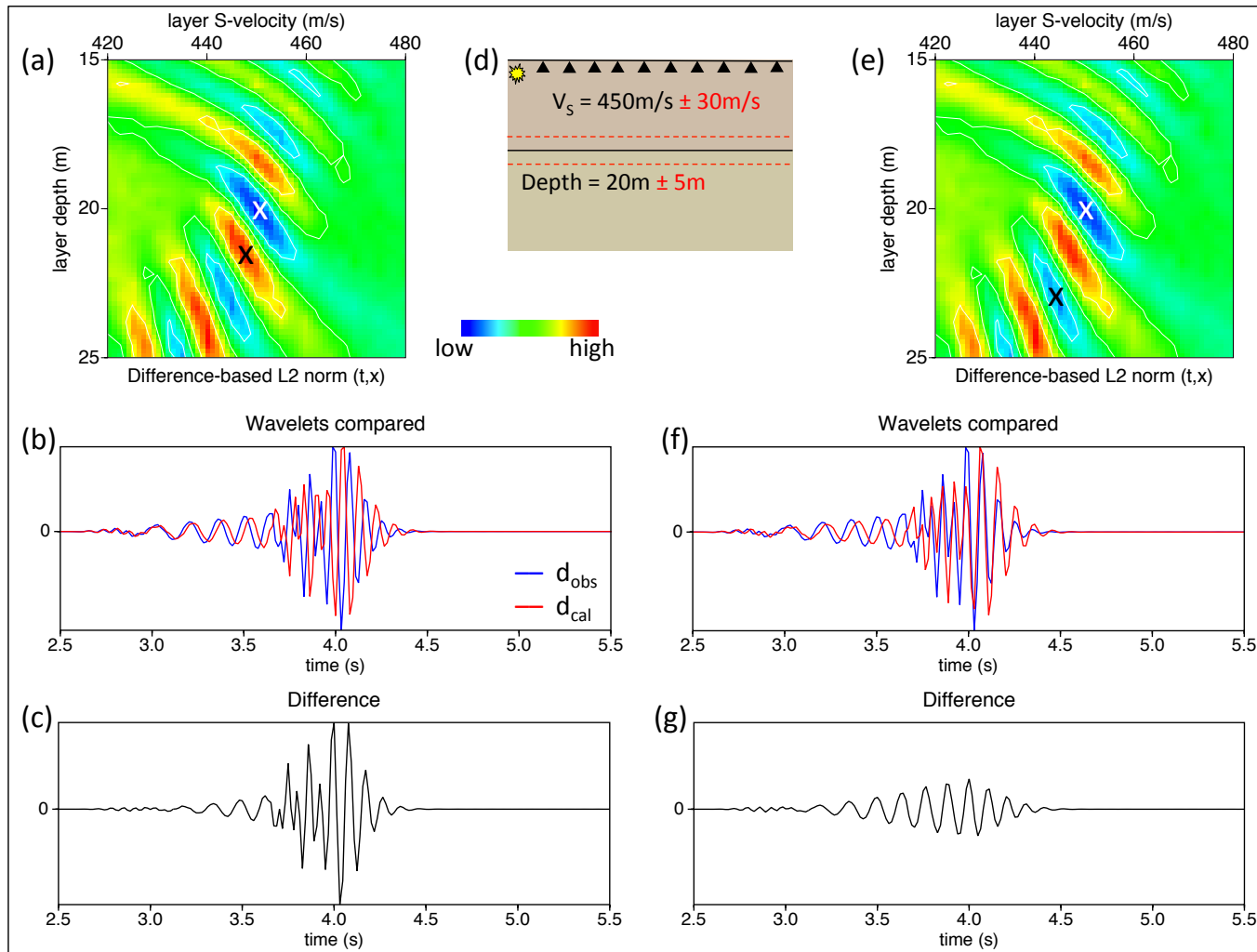


Figure 1.10: 2D misfit maps for the classical difference-based misfit are shown in (a) and (e), as two model parameters are perturbed in (d). The white cross marks the global minimum. The perturbed (d_{cal}) and reference (d_{obs}) data traces are shown in (b) and (f) for the two cases marked by the black cross in (a) and (e) respectively. The data difference is shown in (c) and (g) respectively. The difference in (g) is smaller because the wrong phases match up in (f) causing a local minima.

In Figure 1.10 the perturbed model parameters have not yet caused the first arrivals to be very much shifted compared to the reference dataset, yet the surface waves in the trace are clearly cycle-skipped. When performing FWI with surface waves, because the surface waves often have a higher amplitude than other wave types, they will naturally dominate the misfit and limit any chance of convergence. It is for this reason that including surface waves in FWI is so challenging.

Even when only considering acoustic cases, non-linearities are present and limit the success of FWI. Several strategies have therefore been proposed, and shown to help mitigate this issue. In the absence of implementing a strategy to avoid cycle-skipping, the only other possibility for a successful FWI is to start from a sufficiently good initial model that can explain the data to within half a wavelength: a very challenging issue for near surface imaging.

1.3.2.1 The multiscale strategy

The multiscale strategy (Bunks et al., 1995), is one way to mitigate the problem of cycle-skipping. Described by Sirgue and Pratt (2004) in the frequency domain, it is also called the frequency continuation approach, and is widely implemented in FWI applications, for both acoustic and elastic, on both synthetic and real datasets. The few examples of FWI with surface waves, on the crustal exploration scale, also implement this strategy or a variation of it (see section 1.3.3.2).

The aim of this method is to reduce the non-linearities by first introducing only the low frequency content of the data in the inversion, and subsequently progressing towards the use of higher frequency data content. The resulting model obtained with each frequency band is used as an initial model for the subsequent inversion.

Figure 1.11 illustrates the shape of the classical difference-based misfit function for data at different frequencies. It is evident that for lower frequencies the shape of the valley of attraction of the global minimum is more convex, allowing for convergence starting from an initial model which is further away (Mulder and Plessix, 2008). Contrarily, the misfit function for the high frequency data has a less convex shape, but the global minimum is better constrained, and may lead to a higher resolution result with inversion.

By first inverting for the low frequencies, the initial update of the model is limited to low-resolution large-scale features. As the estimated model approaches the global minimum, convergence is possible with higher frequency data, which are inverted to

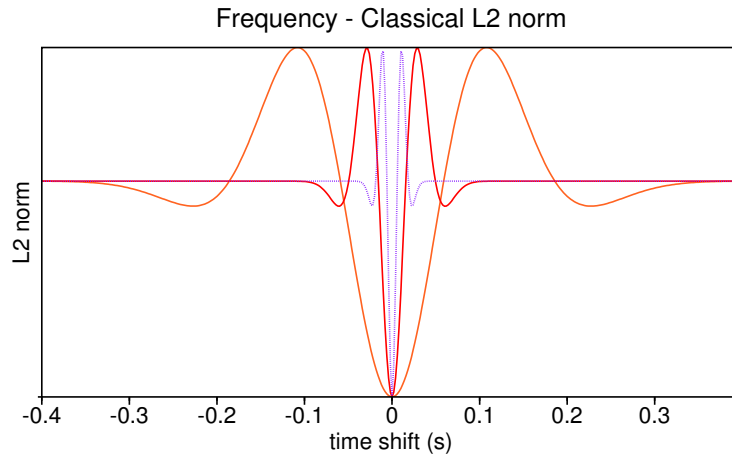


Figure 1.11: Classical difference-based L2 norm misfit of two Ricker wavelets, both with a peak frequency of 4 Hz (orange), 15 Hz (red) and 40 Hz (dashed purple). The wavelets are time shifted compared to each other, simulating a velocity difference between the true and estimated velocity models. The convexity of the misfit shape depends on the frequency content of the data.

update the fine scale details of the model. This multi-resolution approach permits to incorporate increasingly short wavelengths into the model.

The choice of the single frequencies or frequency bands used in this multiscale method is critical and needs to allow for a continuous sampling of vertical wavenumbers for no loss of information (Pratt et al., 1996; Sirgue and Pratt, 2004; Operto et al., 2004; Mulder and Plessix, 2008). This selection also permits to invert only a coarse sampling of frequencies, to exploit redundancy in wavenumber coverage, and reduce the computational time required to obtain a satisfactory result. However, the success of this method is limited by the requirement of low frequency data, which are often absent or very noisy in acquisitions for oil and gas exploration.

1.3.2.2 Other strategies to avoid cycle-skipping

Many strategies have been proposed to avoid cycle-skipping. One approach is based on phase unwrapping in the source-receiver plane for frequency domain FWI, which aims to eliminate phase ambiguity, yet this implementation is often difficult (Choi and Alkhalifah, 2012; Alkhalifah, 2013). Further strategies include dynamic warping (Ma and Hale, 2013), deconvolution-based functions (Luo and Sava, 2011; Warner and Guasch, 2014),

Laplace domain approaches (Shin and Cha, 2008; Ha and Shin, 2012), and semblance-based optimization schemes (Pratt and Symes, 2002; Gao et al., 2014) to name a few. Certain such methods have been targeted for initial model building rather than providing a final result, as large scale features can sometimes be recovered even in the absence of low frequencies. Yet these methods have not yet been applied to elastic FWI using surface waves. As they were designed for other applications, they also do not try to take advantage of the information specifically contained in surface waves.

Other strategies propose to use alternative misfit functions that are more robust, though still considering a measure of the data difference. FWI applications that only consider body waves, show these approaches to be less sensitive to model perturbations in the data, or more adapted to deal with noise. Alternative norms to measure the misfit include the L1 norm (Crase et al., 1990; Djikpéssé and Tarantola, 1999), the hybrid L1/L2 norm (Brossier et al., 2010), the Huber norm (Guitton and Symes, 2003), and the student's t-distribution (Aravkin et al., 2011; van Leeuwen et al., 2013). These make the FWI algorithm more robust, and less sensitive to outliers. However with a strict data-matching of both amplitude and phase, they all suffer from cycle skipping and local minima and are probably not more suitable with regard to FWI of surface waves.

Others have proposed to use a zero-lag cross-correlation misfit function (Routh et al., 2011), where the similarity of the two datasets is maximized in the optimization. Further solutions come from FWI applications on global seismology scales, where surface waves are considered. Envelope and phase only misfit functions appear to mitigate cycle-skipping, and help recover the large-scale features of the model (Bednar et al., 2007; Fichtner et al., 2008; Böldag et al., 2011; Luo and Wu, 2015). However all of these alternative formulations may still suffer from cycle-skipping in the absence of low frequencies.

The option of alternative misfit functions specially designed for FWI with surface waves at the crustal exploration scale, is discussed and investigated in detail in Chapter 2. Alternative formulations such as the cross-correlation based misfit (van Leeuwen and Mulder, 2010), a Singular Value Decomposition (SVD) based misfit (Moghaddam and Mulder, 2012), and a multiscale wavelet decomposition formulation (Yuan et al., 2014), as well as taking advantage of different data domains (Masoni et al., 2013; Pérez Solano et al., 2014) when comparing datasets are addressed.

1.3.3 Full Waveform Inversion with surface waves

Full waveform inversion can be used to image the subsurface of the Earth for various applications and on different scales. Each comes with their own set of difficulties and problematics. Including surface waves within an elastic FWI scheme, as signal to be inverted, has already been achieved on the global and on the near surface scales, while the applications at the oil and gas exploration scale remain limited.

1.3.3.1 From the global scale to the near surface scale

On the global and continental scale, seismologists perform FWI using surface waves to image areas of interest, such as tectonic plate boundary zones, deducing properties of structures on the crustal or mantle scale [Lévêque et al. \(1991\)](#); [Fichtner et al. \(2009\)](#); [Tape et al. \(2010\)](#); [Lekić and Romanowicz \(2011\)](#). These observed surface waves are often generated during large earthquakes, and travel very long distances along the surface of the Earth, making data availability and sparse spatial distributions one of the main limitations at this scale. The lack of dense acquisitions or sufficiently large ranges of source-receiver distances, as well as the computational cost of modelling such large scale structures are the major obstacles that need to be confronted.

Yet at this larger scale, where the heterogeneities of the Earth's crust are weaker, the inversion process is generally also more linear. Surface waves are slowed down by propagating for a longer time through the medium. It is relatively easier to distinguish between different phases at such long offsets, than for seismic recordings at the oil and gas exploration scale. Furthermore the recorded surface waves are of much lower frequency than the signal produced by active sources for exploration, and the problem of cycle-skipping when the analyzing waveform differences is greatly reduced. Still, adapted and more robust misfit functions, based on phase and envelope formulations, are often implemented to reduce the non-linearity of the inversion problem and to account for biased or uncalibrated amplitudes ([Fichtner et al., 2008](#); [Bözdag et al., 2011](#)). For the low resolution required at such large scales, and for the recovery of weak heterogeneities, such robust misfit functions are generally sufficient for convergence.

On the opposite end of the spectrum, on the near surface scale, there is also a great need to image the subsurface, especially for geotechnical engineering purposes. Information on the characteristics of the soil in the first few meters below the surface, obtained from non-invasive techniques, is valuable for the construction of infrastructure, evaluation

of natural risk hazards, investigations related to archeological sites, and environmental issues.

The near surface is often very complex. Weathered surfaces, difficult topographies, low velocity layers, and shallow geologic structures with strong heterogeneities are all characteristics of this region. This makes seismic methods for imaging the subsurface at this scale a challenging problem to tackle. Furthermore the active sources available for such investigations sometimes lack the high frequencies needed to image with sufficient detail. Yet surface waves have high amplitudes compared to other seismic waves, and therefore a high signal-to-noise ratio. This makes them a valuable source of information at this scale.

Most of the classical seismic methods are based on a simple velocity distribution assumption, such as a (locally) layered Earth for dispersion analysis of surface waves using MASW, or an increasing velocity with depth for refraction seismic. In such cases, lateral variations, velocity inversion or strong heterogeneities, push classical seismic methods towards their limits of applicability and reliability. As an alternative to such methods, FWI allows to avoid any assumptions on subsurface complexities. Using FWI for near surface characterization can be tackled by early waveform tomography using either P or SH waves to improve resolution compared to classical ray-based tomography (Ellefsen, 2009; Smithyman et al., 2009). More recently there have also been attempts to exploit the full (visco-)elastic waveform including surface waves (Gélis et al., 2007; Romdhane et al., 2008, 2011; Bretaudeau et al., 2013; Schäfer et al., 2013; Groos et al., 2014).

The first to consider elastic models with a free surface, and try to use surface waves in the context of FWI, was Gélis (2005), who aimed to detect an underground cavity in the near surface. Although the investigation was successful when using elastic body waves, the study resulted in the conclusion that including surface waves in FWI makes the problem highly non-linear. The work of Gélis (2005) highlights some of the difficulties and limitations of the method. As surface waves are more energetic than other waves, they control the misfit and drive the inversion, aggravating the problem of cycle-skipping for FWI, since they travel more slowly. Inversion for a Vs model only was shown to be most successful, with multiparameter inversion, or inverting for Vp only, being more difficult. The implementation of a multiscale strategy going from low to high frequency content was shown to be very helpful, though ultimately the choice of a sufficiently close initial model appeared to be critical for successful convergence.

A second application of FWI with surface waves at the near surface scale, considers a

simple two layer model with the aim of recovering a slightly dipping interface (Romdhane et al., 2008). The study illustrates the limitations of the MASW method for recovering lateral velocity variations, compared to the more successful results with FWI. Again a multiscale strategy from low-to-high frequencies is implemented to help avoid cycle skipping. However the success of the inversion is once again linked to the prior knowledge of a kinematically correct initial model.

A rather successful application that uses surface waves in an FWI scheme was performed by Bretaudeau et al. (2013) with a lab-scale experiment designed to mimic a real case scenario. Although scaled down to the millimeter scale, the diffraction of the surface waves was exploited to detect an anomaly in the medium for a "real" dataset. Shown in Figure 1.12 are the initial and final Vs models. No dispersion is present as the background velocity is constant and known. Yet the example illustrates our capacity to extract information contained in the surface waves by FWI, at least when the initial kinematics are correct.

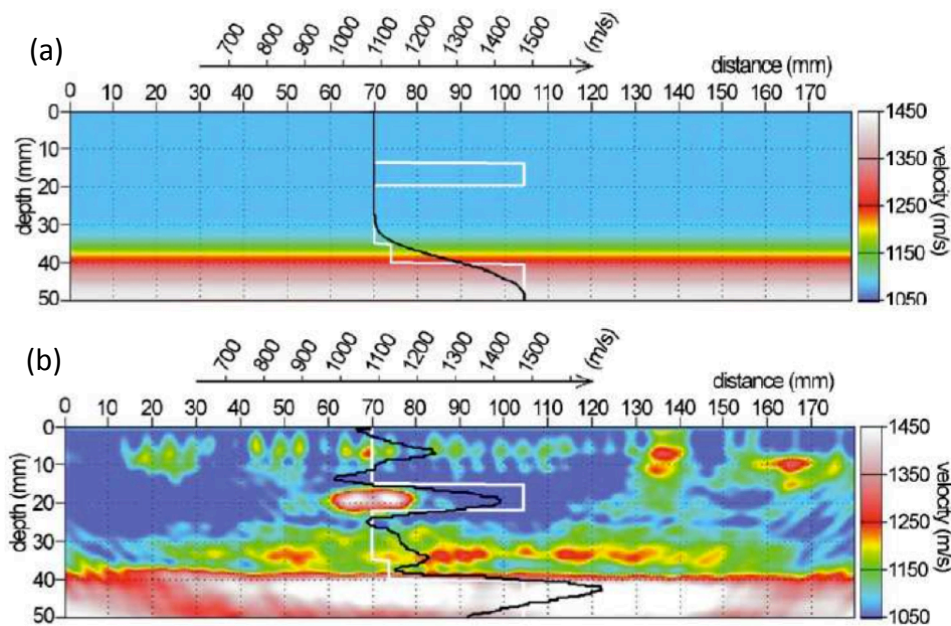


Figure 1.12: Image of the initial Vs model (a), and the recovered Vs model using FWI (b). The white velocity profile is for the true lab-scale medium, and the black velocity profile is shown for the initial and final estimated velocity models. Artifacts in the result are related to the imperfect source wavelet inversion and noise during the "real" experiment. Figure modified from Bretaudeau et al. (2013).

The previous examples all consider the recovery of a heterogeneity of contrasting properties within a relatively homogeneous and well known background model. Recently,

Schäfer et al. (2013) presented a real dataset application, which illustrates some of the complexities that arise when the kinematics are not well known. The study considers the prospection of a fault in the first 15 m below the surface, and uses a linear velocity gradient as a starting V_s model for FWI. Only the low frequency content is successfully inverted. The addition of higher frequencies causes local minima to occur, even when smoothing preconditioning strategies are implemented. The study concludes that a more accurate initial model is required to obtain a good result. Furthermore it notes the importance of taking into account attenuation for elastic cases (Groos et al., 2014).

1.3.3.2 At the oil and gas exploration scale

Bridging the gap between the global scale and the near surface scale, applications of FWI with surface waves performed at the crustal exploration scale are few. There are several problematics at this scale which have limited success. Acquisitions for the oil and gas industry generally lack the low frequencies attained at the global scale, and there is often a limited aperture and/or illumination. More critically, the targets of interest are often very complex and highly heterogeneous.

Almost all successful FWI applications at the oil and gas exploration scale assume an acoustic approximation. One successful attempt at elastic FWI using only body waves, implements a hierarchical multiparameter approach (Prieux et al., 2013a,b). In a first step only a V_p model is inverted for with acoustic FWI. In a second step a multiparameter FWI is performed to update parameters such as density and attenuation, again assuming an acoustic approximation. At this stage, elastic FWI using only body waves is finally performed to improve the V_s model. However applications of elastic FWI remain limited to simple environments and are generally only used for interpretation (Mordret et al., 2011).

There have been even fewer attempts of elastic FWI with the inclusion of surface waves. One approach to help mitigate the non-linearities caused by surface waves has been to apply a time damping to the data for frequency domain FWI (Brossier et al., 2009). This preconditioning of the data, helps the convergence of the inversion problem by permitting to window specific arrivals (Sheng et al., 2006; Sears et al., 2008). The shape of the misfit function can be shown to be more convex when computing the data differences for damped datasets rather than full datasets (see Figure 1.13).

The implementation by Brossier et al. (2009) in the frequency domain, by means of complex frequencies ($\omega + i\gamma$), allows to initially select P-wave events at wide-apertures

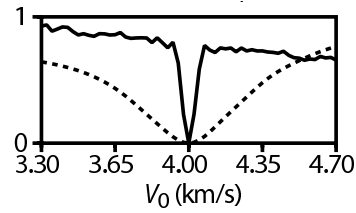


Figure 1.13: The misfit for a full dataset (solid) and a damped dataset (dashed) as a function of one model parameter. Figure modified from [Brossier et al. \(2009\)](#).

while damping shorter aperture P-wave events and converted waves, as well as free-surface multiples and surface waves. As the time-damping is relaxed during the inversion, more strongly non-linear signal is added. This is illustrated in Figure 1.14, showing a synthetic dataset computed for a marine Ocean-Bottom-Cable (OBC) acquisition, to which different damping values have been applied. Looking at the figure, one can observe how surface waves (in this case Scholte waves) at progressively longer offsets are included in the selected data. By the time the surface waves at mid to long offsets are inverted, the kinematics of the estimated velocity models have already been largely corrected by body waves.

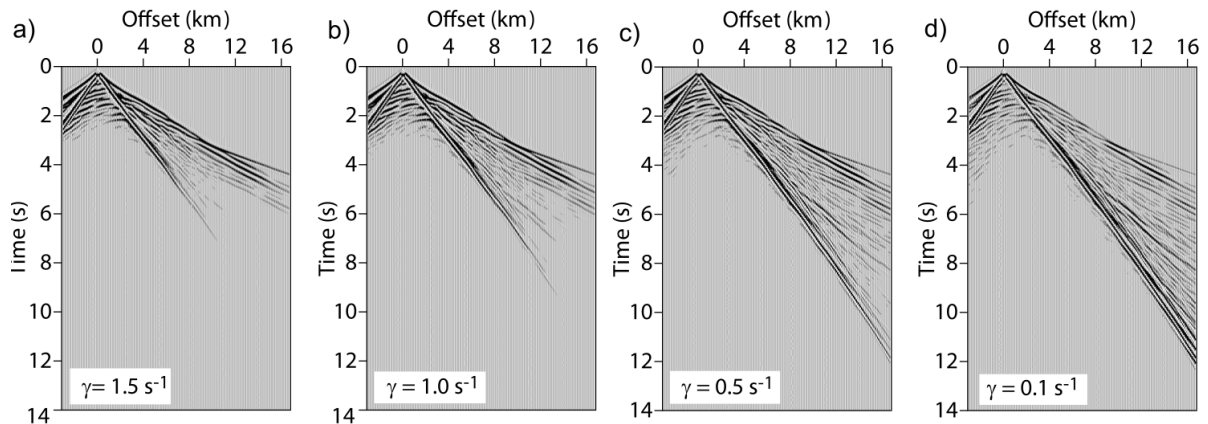


Figure 1.14: Synthetic data with a time-damping applied from the first-arrival time. The time-damping is progressively relaxed in (a) through (d); the damping term is given by the value of the imaginary frequency γ . Figure taken from [Brossier et al. \(2009\)](#).

The results of an application of this method by [Brossier et al. \(2009\)](#) on a dip section of the SEG/EAGE Overthrust model are given in Figure 1.15. The V_p model is greatly improved with the time-damping strategy, validating the possibility to include surface waves in a FWI framework at this scale. The V_s model is also inverted and improved,

suggesting that the elastic wave propagation considered has helped the inversion. Yet the real value that the surface waves bring to the inversion is not evident. Compared to the same synthetic test without a free surface boundary condition, one can only observe a substantial increase in the non-linearity when surface waves are included ([Brossier et al., 2009](#)).

Further work by [Romdhane et al. \(2011\)](#) successfully tested this strategy on a more near surface application considering a complex topography. The implementation of this strategy was critical for convergence to occur. The study also highlighted how using both vertical and horizontal receiver components produced a better result, while with only vertical component receivers the final solution was degraded by 24%.

For both case studies, it is not fully clear if the estimated velocity model is corrected by the information provided by surface waves, or rather that the surface waves are only sufficiently well modelled to allow the inversion to be essentially driven by body waves in an elastic case. There is a need for new strategies, better adapted to exploit surface waves, and specialized for near surface exploration at the oil and gas scale. In the following, [Chapters 2 and 3](#) will deal with finding an alternative, more robust, misfit function, and a strategy, specifically adapted to surface waves, to successfully invert them within an FWI scheme.

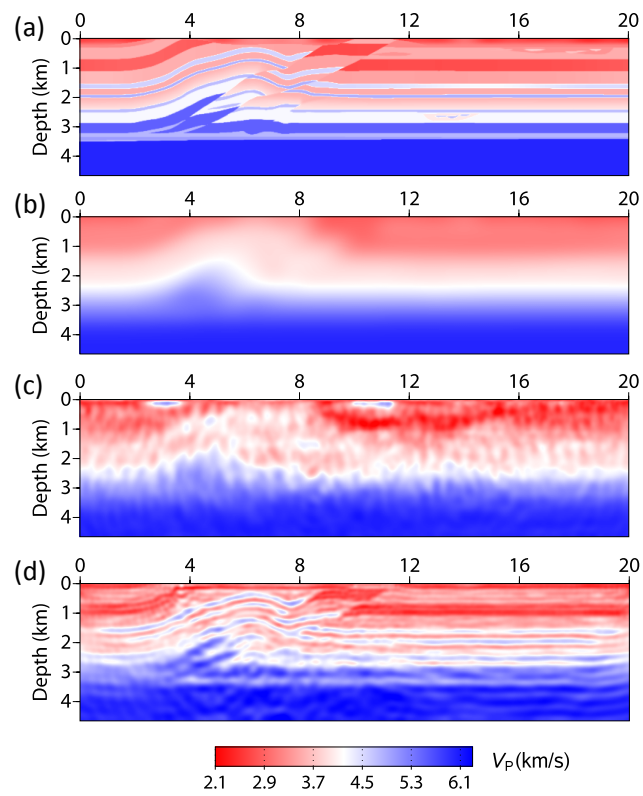


Figure 1.15: True model (a), initial model (b), and the resulting models from 2D frequency domain FWI without (c) and with (d) the implementation of a time-damping approach. Figure modified from [Brossier et al. \(2009\)](#).

Chapter 2

Alternative Misfit Functions for Robust FWI

Contents

2.1	Introduction	62
2.2	Robust Full Waveform Inversion with surface waves	63
2.2.1	Abstract	63
2.2.2	Introduction	64
2.2.3	Forward modelling	66
2.2.4	Misfit function analysis	68
2.2.5	Optimization approach	73
2.2.6	Illustrative synthetic example	79
2.2.7	Conclusion and Perspectives	83
2.2.8	Acknowledgements	84
2.3	Other proposed misfit functions	85

2.1 Introduction

There is still a need to investigate strategies that can specifically exploit the potential information contained in surface waves. Finding a more suitable and robust misfit function is an important ingredient for such strategies. Ideally one would like to obtain a solution such as the one shown in Figure 2.1.

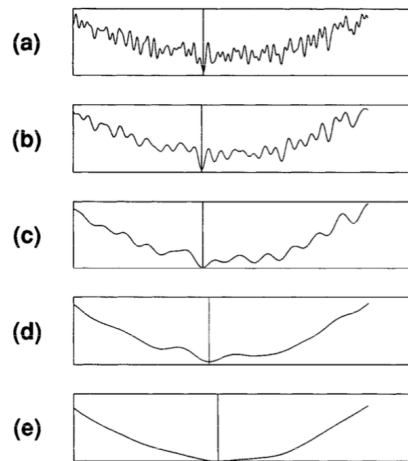


Figure 2.1: With an increased convexity in the shape of the misfit function, as shown progressively from (a) to (e), one can avoid the problem of local minima and cycleskipping, and allow the possibility to converge even when starting with an initial model which is far from the global minimum. Figure taken from [Bunks et al. \(1995\)](#).

A convex shape and a lack in local minima are two key features that a suitable misfit function should possess. This will allow the possibility to converge even when starting with an initial model which is far from the global minimum. However, with more robust misfit functions, the region of the global minimum, at the center of a convex valley, may become very flat, and may lead to a lower resolution of the final model obtained. When this occurs, the resulting model can be used as an initial model for further inversions with less convex misfit functions, as it is now closer to the true model and has a higher chance to converge, as proposed by [Bunks et al. \(1995\)](#) in a multigrid approach.

Some robust misfit functions that have been used to mitigate cycle-skipping issues are discussed in section 1.3.2.2, yet these do not take advantage of the physical characteristics of surface waves. The following article investigates alternative more robust misfit functions, adapted to the inversion of surface waves, and their integration in a FWI scheme. Tests with synthetic data are illustrated to evaluate their performance.

2.2 Robust Full Waveform Inversion with surface waves

Robust Full Waveform Inversion with surface waves

I. Masoni, R. Brossier, J-L. Boelle, M. Macquet, and J. Virieux

2015, *Seismic Technology*, 11(4), p. 48-59,

doi: 10.3997/2405-7495.2015114

2.2.1 Abstract

We have investigated a local optimization procedure for seismic imaging by full waveform inversion of surface waves. These waves are highly energetic when considering near-surface imaging and present dispersive effects as well as significant forward scattering. These propagation behaviours have led us to consider alternative data domains where the data fitting should be performed. The frequency-slowness domain as well as the frequency-wavenumber domain have been considered. We have shown how to build the gradient of the misfit function in these two domains based on the adjoint approach. The adjoint source term has been estimated with a systematic procedure based on Lagrangian multipliers, making the cost of the procedure identical to the one formulated in the standard time-offset domain.

A simple 2D synthetic example shows that the reconstruction of the shear velocity is possible, although real applications could require a more extensive investigation including windowing and filtering, as a hierarchical strategy appears to be necessary when considering surface waves with high variations in amplitude.

A linearised approach is therefore possible and quite efficient. This avoids the picking of dispersion curves and attempts to fill in the velocity spectrum, extending our search domain to models with lateral velocity variations as done previously using global model search.

Keywords: Surface waves, Inverse theory, Dispersion curve analysis, Velocity spectrum, Full waveform inversion

2.2.2 Introduction

The recorded seismic signals at the free surface provide a complex signature we may need to analyze and decipher. Of the different recorded phases in seismograms or traces, surface waves are the most energetic ones and can provide interesting information on targeted zones from near surface to deep structures. In exploration seismics, these phases are considered as noise, although they have been intensively used in seismology for lithospheric reconstruction ([van der Lee, 2002](#)), mantle imaging ([Nolet, 1977](#); [Romanowicz, 2003](#)) and for global Earth imaging ([Lerner-Lam and Jordan, 1983](#)). Shear wave velocity (and bulk velocity and density to a lesser extent) influences the propagation of surface waves through the dispersion relation. Seismic imaging with these surface waves proceeds in a rather different way compared to using body waves as they are dispersive in heterogeneous media ([Aki and Richards, 2002](#)) and provide forward scattering interaction with the free surface when propagating almost horizontally.

Through time-frequency analysis based on moving Gaussian filter ([Cara, 1973](#)), dispersion curve analysis is built at each station: the fundamental mode is picked manually or automatically for different frequencies. Phase differences are obtained between two stations when considering earthquakes. From these 2D mapping of apparent Rayleigh wave phase velocities between two stations at different frequencies, a 3D mapping of the S-wave phase velocities is reconstructed, using the different vertical variations of normal modes with frequency. Although difficult, the detection and the interpretation of the fundamental mode is easier at large scale than at short scale as for subsurface imaging. This has led to consider surface waves as ground rolls and, therefore, noise, although they carry the most important part of the energy in subsurface imaging.

When considering active sources for the emission of seismic waves at shallow depths from tens of meters to few hundreds of meters, the interpretation of surface waves has been promoted as the spectral analysis of surface waves (SASW) using two stations and many active sources as hammers. The picking of dispersion curves done through frequency-wavenumber or slant analysis provides phase velocity diagrams enabling the reconstruction of a local 1D shear-wave model ([McMechan and Yedlin, 1981](#); [Heisey et al., 1982](#); [Nazarian and Stokoe, 1984](#); [Stokoe et al., 1994](#)). This method has been efficiently used in many geotechnical engineering investigations leading to 1D S-wave profiles. Later, an extension to many stations as for refraction seismic investigations, through the multichannel analysis of surface waves (MASW) ([Park et al., 1999](#)), has allowed improved reconstructions of the 1D S-wave velocity profile. Phase analysis on

multiple stations enables a better detection of unwanted energy bursts and a better detection/delineation of the dispersion curve which is used for the profile reconstruction. This approach, intrinsically valid for laterally invariant media, has been extended to smoothly varying medium but is still strongly limited by the requirement to pick and interpret dispersion curves as well as the local constant-velocity layer-based representation of the subsurface (Socco and Strobbia, 2004; Boiero and Socco, 2010; Bergamo et al., 2012). Real applications have been performed in different subsurface environments, but detection of low velocity zones is still difficult for the dispersion-curves interpretation, especially when considering higher modes (Gabriels et al., 1987; Louie, 2001; O’Neil, 2004; Park et al., 2005).

Therefore, mainly the fundamental mode of Rayleigh waves has been used in most applications and efficient inversion schemes have been designed for recovering the S-wave velocity profile. Linearised approaches (Herrmann, 1987) are used and complemented by pure Monte-Carlo perturbation (Shapiro et al., 1997). Semi-global search such as the neighbourhood algorithm has been promoted by Wathelet (2008); Wathelet et al. (2008) for more efficient exploration of the model space with constraints. Very few studies include waveform inversion of surface waves using non-local investigations (Ryden and Park, 2006; Maraschini and Foti, 2010) where one escapes the problem of detecting and interpreting dispersion curves. In these investigations, the extensive exploration of the model space is limited to a rather simple model description. Other studies based on linearised approaches do not focus on surface waves but on the entire seismogram: the so-called full waveform inversion includes both body and surface waves and turns out to be difficult even on synthetic data (Gélis et al., 2007; Romdhane et al., 2011; Schäfer et al., 2013) for near-surface targets. Brossier et al. (2009) have shown that preconditioning the data by time damping allows successful reconstruction of acceptable models while considering the entire waveform on a section of the SEG/EAGE overthrust crustal model using elastic wave propagation with a free surface. The contribution of the surface waves is unclear as one wonders if the body waves are mainly driving the imaging process.

Specific developments of local optimization of surface waveforms deserve an investigation starting from the definition of the misfit function between observed seismograms and synthetic ones at the free surface. In order to avoid local minima, an analysis of different misfit functions has been performed by Masoni et al. (2013) in the framework of a linearised formulation which can be considered as an extension from the dispersion curve analysis towards an interpretation of the full velocity spectrum. Other investigations

have also shown potential interests (Pérez Solano et al., 2014; Yuan et al., 2014).

In this paper, we shall detail the different misfit functions to be used when considering the waveforms of surface waves. After reviewing the elastic wave equation solved with a finite difference scheme to allow lateral variations of material properties, we formulate the optimization procedure using the adjoint formulation for the minimisation of the three misfit functions we have considered. We show how to solve the related adjoint wave equation with a specific source term. We finally proceed with a simple synthetic example to show how we can reconstruct the velocity spectrum starting from an initial model we construct. A complete workflow starting from a MASW analysis to a FWI procedure is beyond the scope of this paper devoted to the design of the optimization scheme in the particular case of surface waves.

2.2.3 Forward modelling

In a 2D medium, seismic waves follow the elastodynamics equation which can be recast into first-order hyperbolic partial differential equations with particle velocity vector (v_x, v_z) and symmetrical stress tensor, recasted as the vector $(\sigma_{xx}, \sigma_{zz}, \sigma_{xz})$, as spatially varying unknowns to be constructed. We consider a time formulation where these equations can be written explicitly as

$$\begin{aligned}
 \rho \frac{\partial v_x}{\partial t} &= \frac{1}{\rho} \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z} + f_{v_x} \\
 \rho \frac{\partial v_z}{\partial t} &= \frac{1}{\rho} \frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z} + f_{v_z} \\
 \frac{\partial \sigma_{xx}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_x}{\partial x} + \lambda \frac{\partial v_z}{\partial z} + g_{\sigma_{xx}} \\
 \frac{\partial \sigma_{zz}}{\partial t} &= (\lambda + 2\mu) \frac{\partial v_z}{\partial z} + \lambda \frac{\partial v_x}{\partial x} + g_{\sigma_{zz}} \\
 \frac{\partial \sigma_{xz}}{\partial t} &= \mu \left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x} \right) + g_{\sigma_{xz}},
 \end{aligned} \tag{2.1}$$

where Lamé coefficients λ and μ as well as the density ρ characterise the medium properties and can spatially change. We have made the spatial dependence implicit in the writing of these equations. External sources could be applied as forces (f_{v_x}, f_{v_z}) or as an internal time-derivative of stress tensor $(g_{\sigma_{xx}}, g_{\sigma_{zz}}, g_{\sigma_{xz}})$. We have discretised these equations at nodes of a staggered finite-difference grid. The spatial derivatives are ap-

proximated by a fourth-order central finite-difference stencil (Levander, 1988) and the explicit time integration relies on a second-order leap-frog scheme.

This system could be given to the vector $u^t = (v_x, v_z, \sigma_{xx}, \sigma_{zz}, \sigma_{xz})^t$ with a compact expression

$$\partial_t u(t, \mathbf{x}) = A_x(\mathbf{x}, m) \partial_x u(t, \mathbf{x}) + A_z(\mathbf{x}, m) \partial_z u(t, \mathbf{x}) + s(t, \mathbf{x}), \quad (2.3)$$

where differential operators A_x and A_z depend on the model parameters m at the position \mathbf{x} in the model. The transpose of a vector is denoted by the symbol t . A conservative form could be designed by considering symmetrical compliance operator $S(m, \mathbf{x})$ leading to an implicit differential system

$$S(\mathbf{x}, m) \partial_t u(t, \mathbf{x}) = H_x(\mathbf{x}) \partial_x u(t, \mathbf{x}) + H_z(\mathbf{x}) \partial_z u(t, \mathbf{x}) + S(m, \mathbf{x}) s(t, \mathbf{x}), \quad (2.4)$$

where symmetrical differential operators H_x and H_z are composed of values one or zero and do not depend on the model properties (Burrige, 1996). Please note that this transformation modifies the source term in this conservative expression. Expression (2.4) is not very useful for modelling as it turns out to be implicit in time, but is quite interesting for the formulation of the inversion because it is self-adjoint and terms depending on both model properties and spatial derivatives can be split.

Boundary conditions are crucial for proper surface wave simulations as well as absorbing boundary conditions for mimicking the extended infinite medium. For flat free surfaces, we have considered the stress imaging condition (Levander, 1988; Robertsson, 1996) which presents a good compromise between computational cost and accuracy, although a limited number of propagated wavelengths can be considered to avoid accumulation of numerical dispersion. The free surface interface goes through nodes sampling stress components $(\sigma_{xx}, \sigma_{zz})$ and velocity component v_x . Finally, the three other edges of the simulation box are implemented with Perfectly Matched Layers (Bérenger, 1994) to mimic a non-reflecting infinite medium.

This finite difference modelling tool will be the forward modelling kernel in the optimization scheme we shall consider, whichever data domain is used for comparing observed and synthetic data.

2.2.4 Misfit function analysis

FWI is a data-fitting imaging technique used to derive quantitative models of the subsurface considering the entire set of waves, including surface waves (Virieux and Operto, 2009). It aims to solve the ill-posed inverse problem of deriving the model parameters that describe the subsurface from observed seismic data. At the oil & gas exploration scale, FWI is up to now mainly used to image at depth considering an acoustic approximation to delineate high resolution velocity models for migration, only using body waves.

The method aims to match the full observed seismogram with a corresponding synthetic seismogram computed from an initial velocity model, solving a local optimization problem. To measure the residual between the observed and the calculated seismograms, the L2 norm of the difference is conventionally used, fitting both amplitude and phase information of the waveforms (Tarantola, 1984a; Pratt et al., 1998). This L2 norm generally focuses on the sample-to-sample difference between the observed and the computed data in a given domain, generally the time-offset domain (t, h) , or frequency-offset (ω, h) for frequency-domain FWI, leading to the following misfit function

$$C_{t,h} = \frac{1}{2} (d_{obs}(t, h) - d_{cal}(t, h))^2, \quad (2.5)$$

where we assume an implicit summation over sources and receivers. By exploiting the full information content of the data and employing a strict data-matching approach, this method is able to produce very high resolution results, but may therefore not be very robust or stable. FWI is indeed susceptible to non-linearity problems such as cycle skipping and local minima, which reduce the convexity of the misfit function (Mulder and Plessix, 2008). As a result, the initial predicted velocity model needs to be accurate enough so that it lies within the small basin of attraction of the global minimum, meaning that it has to be kinematically correct. Otherwise the minimisation easily ends up in the nearest minimum without ever reaching the global minimum, and converges to a wrong velocity model. It is believed that the initial model has to explain the data to within half a wavelength to avoid cycle skipping and converge.

In order to relax this constraint, we may either design a more robust comparison criterion between observed and synthetic data or we may consider comparing observed and synthetic data on a sample-by-sample strategy in different domains. In the following, we shall initially analyze different misfit functions before going more in detail on alternative domains for sample-to-sample comparisons of data with a main focus on surface waves.

2.2.4.1 Robust misfit design

Designing a more robust misfit comparison between observed data and synthetic data could prevent cycle skipping problems and local minima issues related to the essentially oscillating nature of seismic waves. Extracting time-shifts to be minimised by the cross-correlation of windowed waveforms has been promoted by [Luo and Schuster \(1991\)](#); [Tromp et al. \(2005\)](#). A maximisation of the cross-correlation function between observed and calculated data to focus energy close to zero-time lag is used by [van Leeuwen and Mulder \(2010\)](#); [Brossier et al. \(2015\)](#). These comparisons based on cross-correlation between two signals are robust with respect to cycle-skipping but face ambiguity when considering several events. Zero-lag cross-correlation to minimise the effect of amplitude in the misfit ([Routh et al., 2011](#)) or a phase-only misfit function ([Bednar et al., 2007](#)) could mitigate amplitude effects but are sensitive to cycle-skipping unless unwrapping is performed adequately. Integration of the signal preventing the oscillation appears to be sensitive to noise and amplitude prediction ([Donno et al., 2013](#)). The use of the envelope of the signal also appears to relax the initial model accuracy but cycle-skipping is still present ([Bözdag et al., 2011](#)). Deconvolution is also an alternative and has been proposed by [Luo and Sava \(2011\)](#).

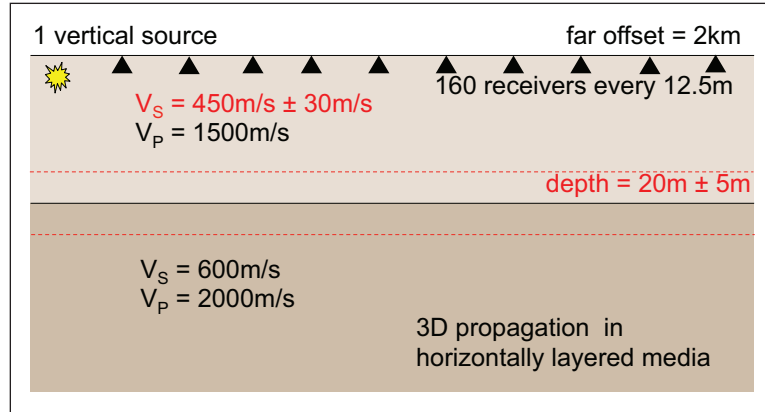


Figure 2.2: Design of a simple model with two parameters defined as the depth of the interface and the S-wave velocity in the upper layer

In order to illustrate differences between different misfit functions, we consider the cross-correlation $Xcorr(\tau)$ between the observed and synthetic data

$$Xcorr(\tau) = \sum_{t=1}^{nt} d_{obs}(t + \tau) d_{cal}(t) \quad (2.6)$$

and the normalised cross-correlation $Xcorr_N(\tau)$

$$Xcorr_N(\tau) = \frac{Xcorr(\tau)}{\|d_{obs}\| \|d_{cal}\|}. \quad (2.7)$$

This cross-correlation function (2.7) is insensitive to the amplitudes of the wavelet and allows to access the phase of the data. This function is used by [van Leeuwen and Mulder \(2010\)](#); [Brossier et al. \(2015\)](#) to define a weighted norm of the normalised cross-correlation through the expression

$$C_{W_i} = \frac{1}{2} \sum_{\tau=1}^{n\tau} (W_i(\tau) Xcorr_N(\tau))^2, \quad (2.8)$$

where the weight $W_i(\tau)$ is applied at each time sample. The width and the convexity of these misfit functions are modified depending on the applied weighting. Two types of weighting are considered through the following expressions

$$\begin{aligned} W_1(\tau) &= (\tau/\tau_{max})^2 \\ W_2(\tau) &= -e^{-\alpha\tau^2}. \end{aligned} \quad (2.9)$$

The first weighting W_1 linearly penalises the values further away from zero-lag value while the second one concentrates on values near by the zero-lag ([van Leeuwen and Mulder, 2010](#)). If the observed and the synthetic data match, the energy in the cross-correlation is maximised at zero-lag, and the energy at zero-lag is subsequently annihilated by the penalty function. The weighted norm can only be zero if the wavelet is a delta pulse and all cross-talk is excluded by choosing an appropriate maximal shift. However, the cross-correlation function suffers from cross-talk between different events in traces which leads to significant energy away from zero-lag even for perfectly fitting data (this is also true for the auto-correlation of the signal). This effect limits the application of such a misfit to windowed data or when one arrival dominates, which could be the case for surface waves.

In order to illustrate the impact of the three different misfit functions $C_{(t,h)}$, C_{W_1} , C_{W_2} , we consider a simple model with two parameters which are perturbed in order to extensively sample the misfit function: the velocity of the first layer and the thickness of this layer (Figure 2.2). Using finite-difference modelling the true data values are computed, as well as the synthetic data values for each model sampling the two-parameter space. The range of the perturbed layer depth is of 15 m to 25 m, while the true depth is 20 m, and the perturbed layer velocity ranges from to 480 m/s, while the true value is

450 m/s . Figure 2.3 illustrates the difficulties of the standard sample-by-sample misfit function and the advantages of considering normalised cross-correlations at the expense of an additional parameter coming from the weight.

These cross-correlation based misfit functions can be quite interesting alternatives when starting from crude initial models although the expected resolution characterised by the shape of the attraction basin may be quite poor. Furthermore, cross-talk effects may require to window the data and the inaccuracy of the source wavelet could mitigate this robustness as the cross-correlation is quite sensitive to the spectrum of the data. Further research would be required to tackle these problems. In the following, we focus on the sample-to-sample strategy, considering alternative domains.

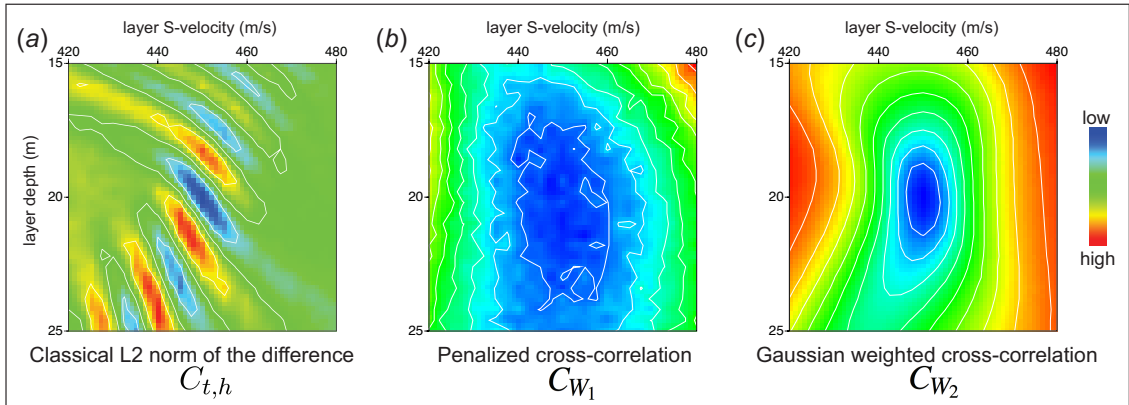


Figure 2.3: Misfit values for the three different misfit functions we have investigated. On the left panel (a), one can see that the classical misfit function has many local minima while the middle panel (b), penalising values far from zero-lag, is more noisy. The right panel (c) presents a nice smooth valley, given by the Gaussian windowing which needs to be carefully designed.

2.2.4.2 Sample-by-sample strategy

We may benefit from dense acquisition sampling for an improved analysis of the sample-by-sample comparison. Alternative misfit functions can be used following similar strategies as those used in dispersion curve analysis. McMechan and Yedlin (1981) have transformed data in the frequency-slowness domain while Gabriels et al. (1987) have considered data transformed to the frequency-wavenumber domain. This leads us to

propose the following misfit functions for FWI, first in the $\omega - p$ domain

$$C_{\omega,p} = \frac{1}{2} (|d_{obs_N}(\omega, p)| - |d_{cal_N}(\omega, p)|)^2, \quad (2.10)$$

and then in the $\omega - k$ domain after two Fourier transforms of the data

$$C_{\omega,k} = \frac{1}{2} (|d_{obs_N}(\omega, k)| - |d_{cal_N}(\omega, k)|)^2. \quad (2.11)$$

The data is normalized before being transformed to the alternative domain. The phase has also been removed from these two new misfit functions, and only the amplitude of the velocity spectrum is considered. This strategy is related to dispersion curve analysis where maximum-energy picks allow the extraction of the apparent dispersion velocity at each frequency (Pérez Solano et al., 2014). A related misfit design has been promoted by Maraschini et al. (2010) for the particular case of layered media in order to avoid detection and picking of dispersion curves.

Applying a linear moveout (LMO) to obtain data in the $\omega - p$ domain or applying a wavenumber Fourier transform to obtain data in the $\omega - k$ domain separates events through slopes or wavenumbers. This might allow the extraction of kinematic information in a more robust manner. Furthermore the stacking involved in the transformation may also make the misfit function more efficient in the presence of noise. In these domains, we may have a better identification of frequency-dependent dispersion effects. Moreover, because we consider the modulus of the data, the misfit function is no more sensitive to the phase of the source wavelet, but should not limit the capacity of fitting the kinematic properties contained in the data. The simple example we have considered may help illustrate the effects of these transformations. Figure 2.4 displays the misfit function structure where one can observe major differences between the (t, h) domain and both the transformed (ω, p) and (ω, k) domains, which have a broad attraction basin. Of course, this structure may be due to the simple description of the model space.

We may now consider how to update the model parameters while considering the comparison of data in these different domains, but still doing the forward modelling in the (t, h) domain using the finite-difference engine we have selected through time integration of the differential system (Equation 2.3).

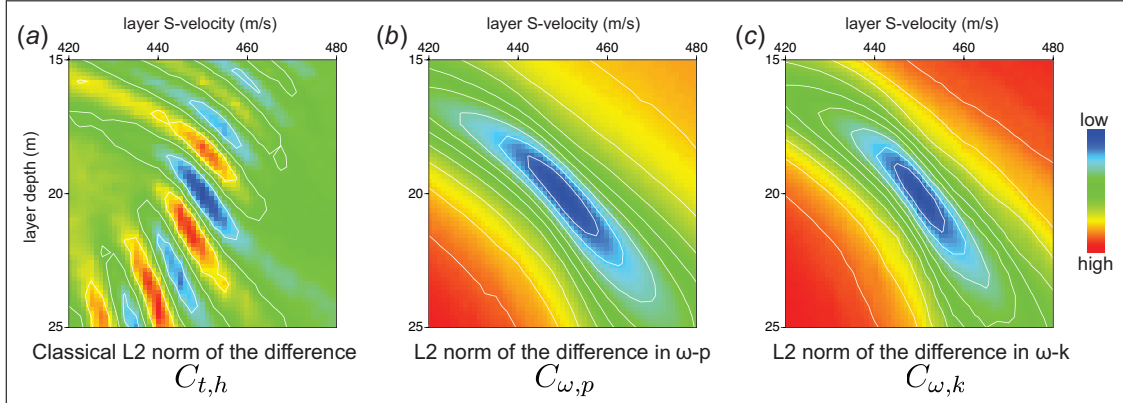


Figure 2.4: Sample-by-sample differential misfit values for the three different domains we consider. The left panel (a) is the same as the one in Figure 2.3, while the middle panel (b) is for the (ω, k) domain and the right panel for the (ω, p) domain. Relation $k = \omega p$ between slowness and wavenumber leads to a similar structure of the misfit functions.

2.2.5 Optimization approach

We shall proceed by a linearized approach for updating the model, and we may therefore estimate the gradient operator of each misfit function and eventually the effect of the Hessian operator on the model perturbation vector. In this work, we shall concentrate on the estimation of the gradient and we shall consider quasi-Newton approaches for updating the model (Byrd et al., 1995). We shall proceed following Plessix (2006) by constructing a Lagrangian function for each domain in which the seismic data are expressed.

2.2.5.1 Misfit in the domain (t, h)

We shall first consider the generic scalar product

$$\langle f(t, \mathbf{x}) | g(t, \mathbf{x}) \rangle_{T, \Omega^2} = \int_T \iint_{\Omega^2} dt d\mathbf{x} f(t, \mathbf{x})^\dagger g(t, \mathbf{x}), \quad (2.12)$$

over the time window $[0, T]$ and the 2D space Ω^2 . Note that symbol \dagger denotes the transpose complex conjugate of a complex number vector if $f(t, \mathbf{x})$ is defined in \mathcal{C}^n .

The misfit function will be defined still assuming an implicit summation over sources

through the expression

$$\begin{aligned} C_{t,h}(m, d_{cal}) &= \frac{1}{2} \|d_{obs}(t, h) - d_{cal}(t, h)\|_{T, \Omega_r}^2 \\ &= \frac{1}{2} \langle d_{obs}(t, h) - d_{cal}(t, h) | d_{obs}(t, h) - d_{cal}(t, h) \rangle_{T, \Omega_r}, \end{aligned} \quad (2.13)$$

where Ω_r is the receiver space.

We shall consider the following Lagrangian

$$\begin{aligned} \mathcal{L}(m, d_{cal}, u, \xi, \zeta) &= C_{t,h}(m, d_{cal}) + \langle \xi(t, h) | d_{cal}(t, h) - R_h u(t, \mathbf{x}) \rangle_{T, \Omega_r} \\ &+ \langle \zeta(t, \mathbf{x}) | S \partial_t u(t, \mathbf{x}) - H_x \partial_x u(t, \mathbf{x}) - H_z \partial_z u(t, \mathbf{x}) - S s(t, \mathbf{x}) \rangle_T \\ &+ \langle \zeta(0, \mathbf{x}) | u(0, \mathbf{x}) \rangle, \end{aligned} \quad (2.14)$$

which must have derivatives equal to zero for the different arguments which are considered as independent (Chavent, 2009). Please note that we have considered here the differential system (Equation 2.4) with an initial value condition, and that spatial dependency on operators S , H_x and H_z have been removed for compactness.

Zeroing the derivative with respect to the vector ζ gives the state equation of wave propagation at all spatial positions and all times, while zeroing the derivative with respect to the vector ξ will define the state equation which projects the wavefield $u(t, \mathbf{x})$ at receiver positions through the operators R_h .

Zeroing the derivative with respect to the state variable $u(t, \mathbf{x})$ and $d_{cal}(t, h)$ leads to the definition of two adjoint variables

$$\xi(t, h) = d_{obs}(t, h) - d_{cal}(t, h) = \Delta d(t, h) \quad (2.15)$$

and

$$\begin{aligned} S \partial_t \zeta(t, \mathbf{x}) - H_x \partial_x \zeta(t, \mathbf{x}) - H_z \partial_z \zeta(t, \mathbf{x}) &= \sum_{\Omega_r} R_h^t \xi(t, h) \\ \text{subject to } \zeta(T, \mathbf{x}) &= 0 \end{aligned} \quad (2.16)$$

Note that the prolongation operator R^t is the transpose of the operator R . Note also that because of the self-adjoint expression of the wave-equation (2.4), the ζ adjoint equation is a wave-equation problem similar to the state equation, but subjected to a final condition.

This adjoint-equation is classically solved using the change of variable $\tau = T - t$, giving

$$\begin{aligned}
 -S\partial_\tau\zeta(T - \tau, \mathbf{x}) - H_x\partial_x\zeta(T - \tau, \mathbf{x}) - H_z\partial_z\zeta(T - \tau, \mathbf{x}) &= \sum_{\Omega_r} R_h^t(\Delta d(T - \tau, h)) \\
 \text{subject to } \zeta(0, \mathbf{x}) &= 0.
 \end{aligned} \tag{2.17}$$

Finally, one can observe that the partial differential equation of the adjoint system is very close to the one of the forward wavefield with a specific source term, gathering the residues at receivers positions.

One can deduce the gradient by taking the derivative of the Lagrangian with respect to each model parameter m_i through the expression

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial m_i} &= \frac{\partial \mathcal{C}}{\partial m_i} = \langle \zeta(t, \mathbf{x}) | \frac{\partial S(m)}{\partial m_i} \partial_t u(t, \mathbf{x}) \rangle_T \\
 &= \int_T dt \zeta(t, \mathbf{x})^t \frac{\partial S(m)}{\partial m_i} \partial_t u(t, \mathbf{x})
 \end{aligned} \tag{2.18}$$

away from active sources. The expression shows that the gradient is the zero-lag cross-correlation between the adjoint wavefield and the derivative of the incident wavefield multiplied by a local scattering operator expressed with parameter derivatives of the compliance matrix. The scattering operator expresses both the illumination of the scattering point by the incident wavefield and the re-emission capacity towards the receiver through the adjoint wavefield.

2.2.5.2 Misfit in the domain (ω, p)

The misfit function is expressed in the (ω, p) transformed domain through the expression

$$\begin{aligned}
 C_{\omega,p}(m, \bar{d}_{cal_N}) &= || |\bar{d}_{obs_N}(\omega, p)| - |\bar{d}_{cal_N}(\omega, p)| ||_{F,P}^2 \\
 &= \langle |\bar{d}_{obs_N}(\omega, p)| - |\bar{d}_{cal_N}(\omega, p)| | |\bar{d}_{obs_N}(\omega, p)| - |\bar{d}_{cal_N}(\omega, p)| \rangle_{F,P},
 \end{aligned} \tag{2.19}$$

where F is the space of considered frequencies and P the space of considered slownesses.

We note the linear move-out (LMO) transform of the wavefield at receivers $u(t, h)$, known also as the slant stack, as

$$\tilde{u}(t, p) = \int_{\Omega_r} dh u(t + ph, h), \tag{2.20}$$

and the Fourier transform of $\tilde{u}(t, p)$ with the compact notation

$$\bar{u}(\omega, p) = \int_T dt \tilde{u}(t, p) e^{-i\omega t}. \quad (2.21)$$

We construct the related Lagrangian $\mathcal{L}(m, u, d_{cal}, d_{cal_N}, \tilde{d}_{cal_N}, \bar{d}_{cal_N}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ with five adjoint wavefields related to applied transformations, normalisation and differentiation as the following expression

$$\begin{aligned} \mathcal{L} = & C_{\omega,p}(m, \bar{d}_{cal_N}) + \langle \lambda_1(\omega, p) | \bar{d}_{cal_N}(\omega, p) - \int_T dt \tilde{d}_{cal_N}(t, p) e^{-i\omega t} \rangle_{F,P} \\ & + \langle \lambda_2(t, p) | \tilde{d}_{cal_N}(t, p) - \int_{\Omega_r} dh d_{cal_N}(t + ph, h) \rangle_{T,P} \\ & + \langle \lambda_3(t, h) | d_{cal_N}(t, h) - \frac{d_{cal}(t, h)}{\|d_{obs}(h)\|} \rangle_{T, \Omega_r} \\ & + \langle \lambda_4(t, h) | d_{cal}(t, h) - R_h u(t, \mathbf{x}) \rangle_{T, \Omega_r} \\ & + \langle \lambda_5(t, \mathbf{x}) | S \partial_t u(t, \mathbf{x}) - H_x \partial_x u(t, \mathbf{x}) - H_z \partial_z u(t, \mathbf{x}) - S s(t, \mathbf{x}) \rangle_T \\ & + \langle \lambda_5(0, \mathbf{x}) | u(0, \mathbf{x}) \rangle, \end{aligned} \quad (2.22)$$

We may proceed in the same way as for the (t, h) domain. The derivative with respect to each adjoint variable λ_i will provide states equations of computation and transformations requested on data. The derivatives with respect to the variable \bar{d}_{cal_N} provides the definition of the adjoint variable λ_1 through the expression

$$\lambda_1(\omega, p) = (\bar{d}_{obs_N}(\omega, p) - \bar{d}_{cal_N}(\omega, p)) \frac{\bar{d}_{cal_N}(\omega, p)^\dagger}{\|\bar{d}_{cal_N}(\omega, p)\|}. \quad (2.23)$$

Note that $\bar{d}_{cal_N}(\omega, p)^\dagger / \|\bar{d}_{cal_N}(\omega, p)\|$ represents the opposite of the phase of $\bar{d}_{cal_N}(\omega, p)$, meaning that if $\bar{d}_{cal_N}(\omega, p) = R e^{i\phi}$, then $\bar{d}_{cal_N}(\omega, p)^\dagger / \|\bar{d}_{cal_N}(\omega, p)\| = e^{-i\phi}$.

The derivative with respect to the variable \tilde{d}_{cal_N} provides the adjoint variables λ_2 through the expression

$$\lambda_2(t, p) = \int_F d\omega \lambda_1(\omega, p) e^{i\omega t}, \quad (2.24)$$

giving the adjoint of the Fourier transform of the λ_1 variable, an inverse Fourier transform if F is considering the whole frequency range. The derivative with respect to the variable d_{cal_N} requires a change of variable $t' = t + ph$ and we assume that boundaries have no impact on the interval defined for variable t and for t' and we can obtain the following

expression

$$\lambda_3(t, h) = \int_P dp \lambda_2(t - ph, p), \quad (2.25)$$

giving the adjoint of the LMO transform of λ_2 , an inverse LMO transform if P is considering the whole range of slownesses. The derivative with respect to the variable d_{cal} gives us

$$\lambda_4(t, h) = \frac{\lambda_3(t, h)}{\|d_{obs}(h)\|}. \quad (2.26)$$

Finally the derivative with respect to the variable $u(t, \mathbf{x})$ gives the partial differential equation of the adjoint variable λ_5

$$\begin{aligned} S\partial_t\lambda_5(t, \mathbf{x}) - H_x\partial_x\lambda_5(t, \mathbf{x}) - H_z\partial_z\lambda_5(t, \mathbf{x}) &= \sum_{\Omega_r} R_h^t\lambda_4(t, h) \\ \text{subject to } \lambda_5(T, \mathbf{x}) &= 0. \end{aligned} \quad (2.27)$$

One can see that the source of the adjoint wavefield can be recursively deduced from the different transformations/normalisation (Equations 2.23, 2.24 2.25, 2.26, 2.27), making the gradient computation (2.18) feasible in the (ω, p) domain.

2.2.5.3 Misfit in the domain (ω, k)

We can proceed in exactly the same way for the (ω, k) domain. The misfit function is expressed in the (ω, k) transformed domain through the expression

$$\begin{aligned} C_{\omega, k}(m, \bar{d}_{cal_N}) &= \| |\bar{d}_{obs_N}(\omega, k)| - |\bar{d}_{cal_N}(\omega, k)| \|_{F, K}^2 \\ &= \langle |\bar{d}_{obs_N}(\omega, k)| - |\bar{d}_{cal_N}(\omega, k)| \mid |\bar{d}_{obs_N}(\omega, k)| - |\bar{d}_{cal_N}(\omega, k)| \rangle_{F, K}, \end{aligned} \quad (2.28)$$

where F is the space of considered frequencies and K the space of considered wavenumbers. We note the wavenumber Fourier transform of the wavefield at receivers $u(t, h)$

$$\tilde{u}(t, k) = \int_{\Omega_r} dh u(t, h) e^{-ikh} \quad (2.29)$$

and the Fourier transform of $\tilde{u}(t, k)$ with the compact notation

$$\bar{u}(\omega, p) = \int_T dt \tilde{u}(t, k) e^{-i\omega t}. \quad (2.30)$$

We construct the related Lagrangian $\mathcal{L}(m, u, d_{cal}, d_{cal_N}, \tilde{d}_{cal_N}, \bar{d}_{cal_N}, \mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$ with five adjoint wavefields related to applied transformations, normalisation and differentiation as the following expression

$$\begin{aligned}
 \mathcal{L} = & C_{\omega,k}(m, \bar{d}_{cal_N}) + \langle \mu_1(\omega, k) | \bar{d}_{cal_N}(\omega, k) - \int_T dt \tilde{d}_{cal_N}(t, k) e^{-i\omega t} \rangle_{F,K} \\
 & + \langle \mu_2(t, k) | \tilde{d}_{cal_N}(t, k) - \int_{\Omega_r} dh d_{cal_N}(t, h) e^{-ikh} \rangle_{T,K} \\
 & + \langle \mu_3(t, h) | d_{cal_N}(t, h) - \frac{d_{cal}(t, h)}{\|d_{obs}(h)\|} \rangle_{T,\Omega_r} \\
 & + \langle \mu_4(t, h) | d_{cal}(t, h) - R_h u(t, \mathbf{x}) \rangle_{T,\Omega_r} \\
 & + \langle \mu_5(t, \mathbf{x}) | S \partial_t u(t, \mathbf{x}) - H_x \partial_x u(t, \mathbf{x}) - H_z \partial_z u(t, \mathbf{x}) - S s(t, \mathbf{x}) \rangle_T \\
 & + \langle \mu_5(0, \mathbf{x}) | u(0, \mathbf{x}) \rangle,
 \end{aligned} \tag{2.31}$$

Following the same strategy, the adjoint expressions are successively obtained

$$\mu_1(\omega, k) = \bar{d}_{obs_N}(\omega, k) - \bar{d}_{cal_N}(\omega, k) \frac{\bar{d}_{cal_N}(\omega, k)^\dagger}{\|\bar{d}_{cal_N}(\omega, k)\|}, \tag{2.32}$$

giving the residues,

$$\mu_2(t, k) = \int_F d\omega \mu_1(\omega, k) e^{i\omega t}, \tag{2.33}$$

giving an adjoint temporal Fourier transform,

$$\mu_3(t, h) = \int_K dk \mu_2(t, k) e^{ikh}, \tag{2.34}$$

giving an adjoint spatial Fourier transform,

$$\mu_4(t, h) = \frac{\mu_3(t, h)}{\|d_{obs}(h)\|}, \tag{2.35}$$

giving the normalisation

$$\begin{aligned}
 S \partial_t \mu_5(t, \mathbf{x}) - H_x \partial_x \mu_5(t, \mathbf{x}) - H_z \partial_z \mu_5(t, \mathbf{x}) &= \sum_{\Omega_r} R_h^t \mu_4(t, h) \\
 \text{subject to } \mu_5(T, \mathbf{x}) &= 0
 \end{aligned} \tag{2.36}$$

giving the partial differential equation of the μ_5 adjoint variable.

Again, one can see that the source of the adjoint wavefield μ_5 can be recursively

deduced from the different transformations and normalization (Equations 2.32, 2.33, 2.34, 2.35, 2.36), making the gradient computation (2.18) feasible in the (ω, k) domain.

2.2.6 Illustrative synthetic example

We shall consider a simple synthetic example where an oblique velocity gradient is assumed in the upper layer as illustrated in Figure 2.5a. The shear velocity increases laterally by 0.5 s^{-1} , and the vertically by 5 s^{-1} . The P-wave velocity is determined by a constant poisson ratio, and the density is homogeneous. The acquisition simulated is a fixed-spread consisting of 24 vertical sources and 48 multi-component receivers evenly spaced on the surface of the 600 m long model. The resulting data in the $t - x$ domain is shown in Figure 2.6. We can see quite dispersive surface waves coming from the gradient in the true model. We consider an initial model with two layers of constant velocities: we have eliminated the gradient in the upper layer as can be seen in Figure 2.5b. Differences with the initial data are quite strong even for the phase evolution of surface waves and one may wonder how the FWI will perform on these datasets. One may also look at the data in the phase velocity spectrum as shown in Figure 2.7 which is the standard domain when analysing the dispersion of surface waves.

We perform the FWI reconstruction of the shear velocity both in the $\omega - p$ domain and in the $\omega - k$ domain. Both approaches extract information from the data in the frequency range we consider. Figures 2.5c-2.5d show the reconstructed models, while the fit of the data is shown in Figure 2.6 viewed in the domain where the data is collected, and in Figure 2.7 which highlights the dispersion of the surface waves. The depth of penetration of the surface waves controls the depth at which we can reconstruct the velocity which reaches the depth of the first layer at 20 m depth.

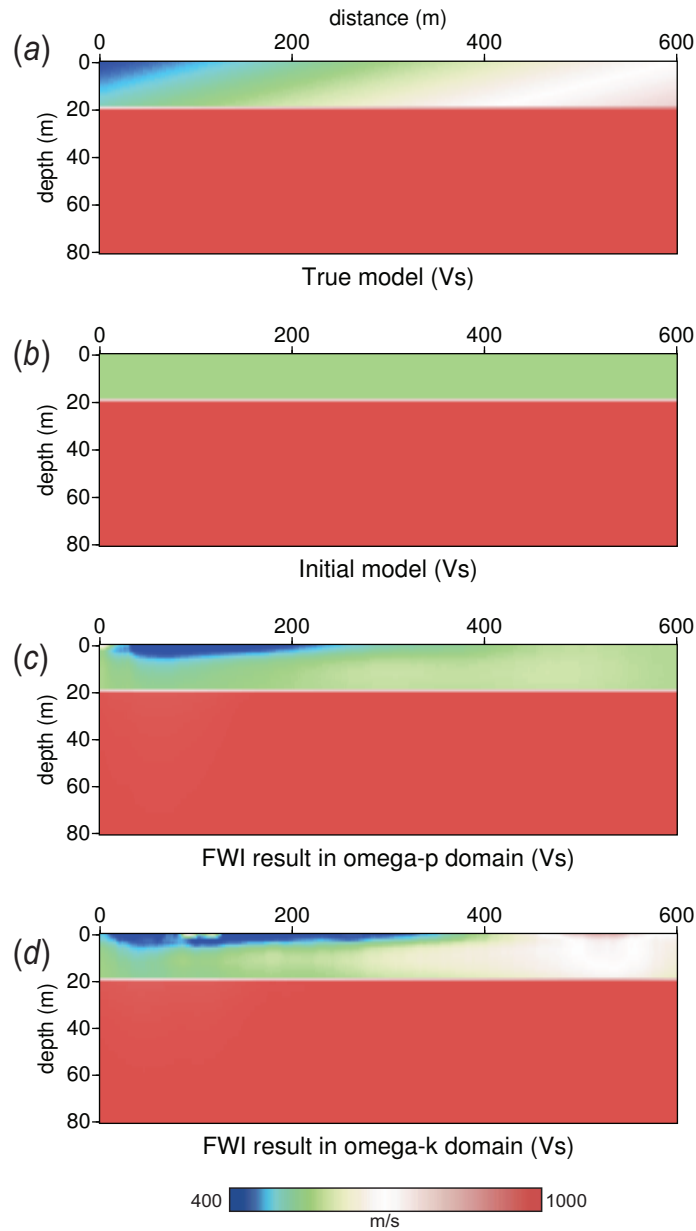


Figure 2.5: Model description and reconstruction: the top two models represent the true model with an oblique gradient (a) and the initial two-layered model (b). The bottom two models represent the reconstructed model when performing FWI in the $\omega - p$ domain (c) and the reconstructed model when performing FWI in the $\omega - k$ domain (d). Please note that in this example the horizontal gradient reconstruction is more successful for FWI in the $\omega - k$ domain.

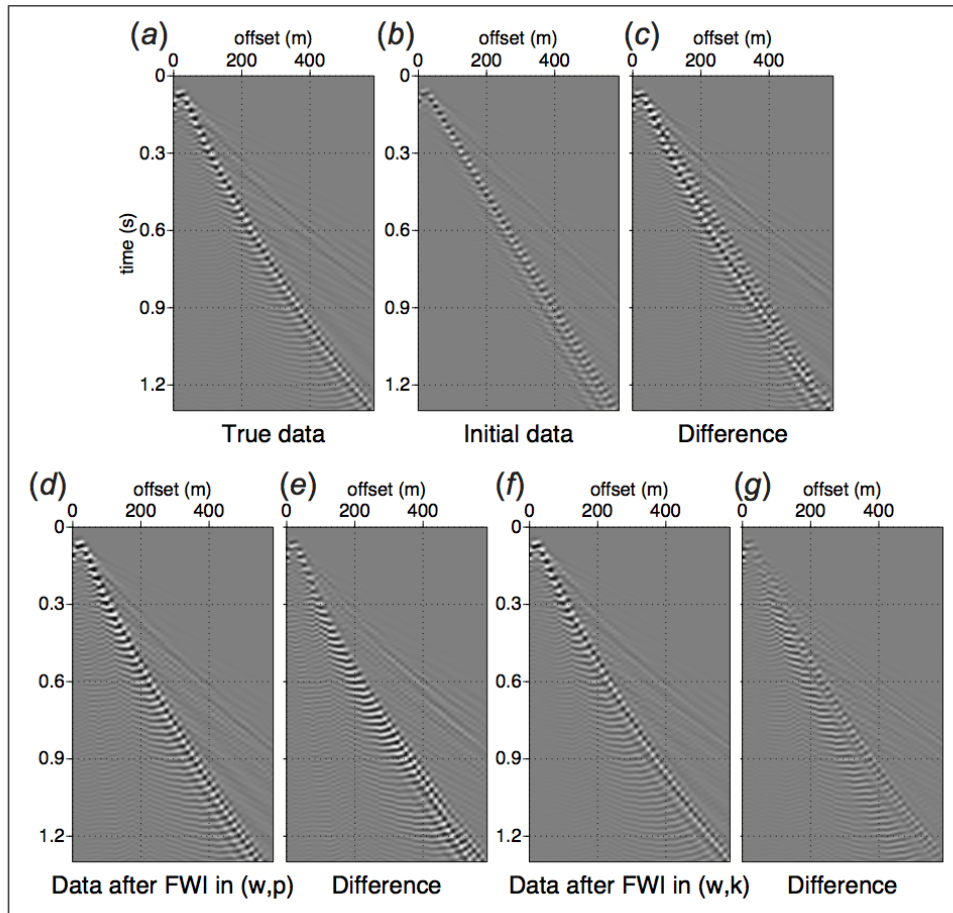


Figure 2.6: The true data with strong energetic dispersive surface waves (a); the initial data for the two-layered model (b) and the difference to the true data (c); the final data after FWI in the $\omega - p$ domain (d) and the difference to the true data (e); and the final data after FWI in the $\omega - k$ domain (f) and the difference to the true data (g). In this example, the FWI performs better in the $\omega - k$ domain than in the $\omega - p$ domain.

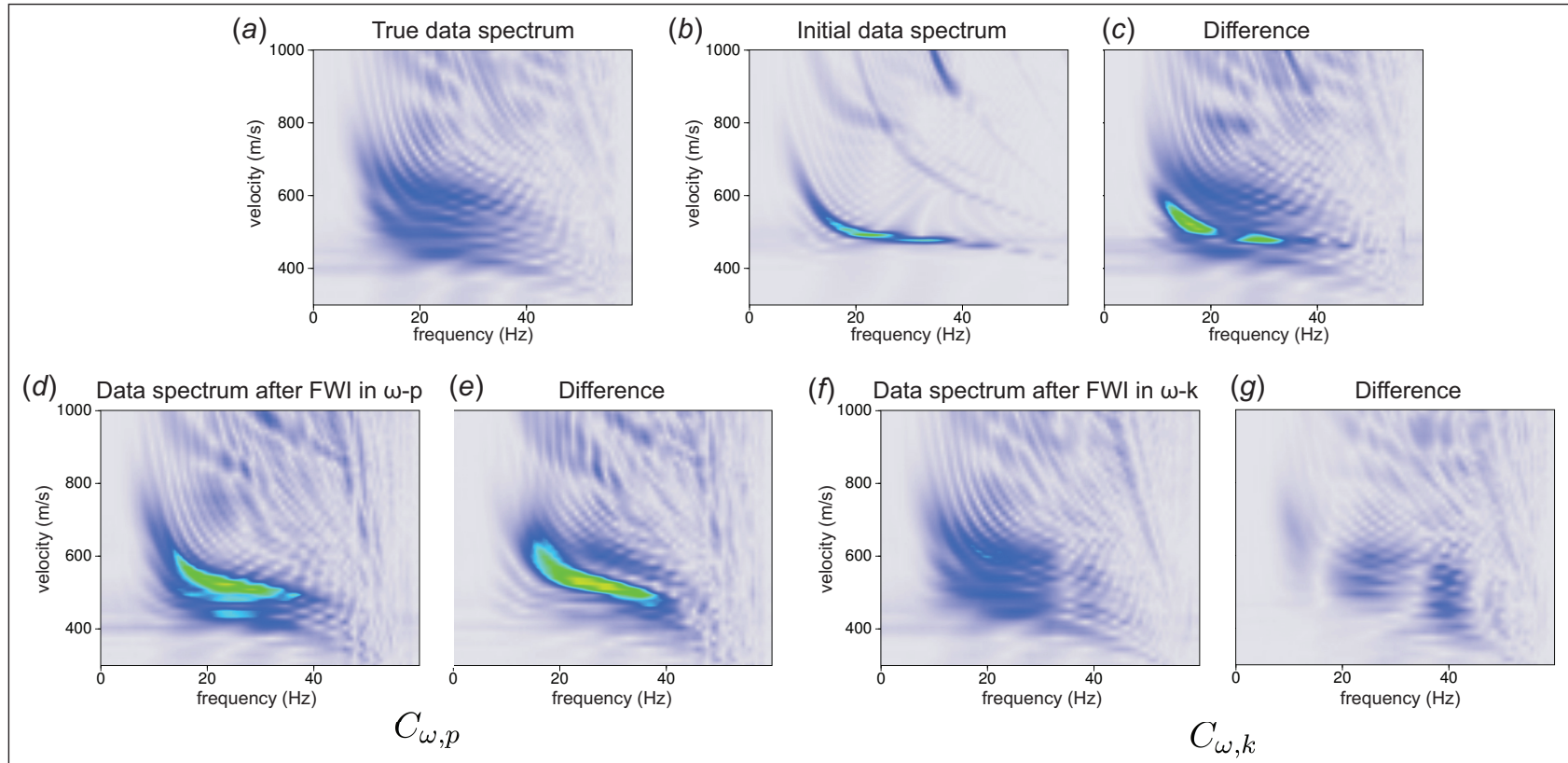


Figure 2.7: Represented in the top row are the true velocity spectrum (a), the initial spectrum for the simple two-layer model (b) and the velocity spectrum differences (c). The bottom row represents the final velocity spectrum for the FWI result in the $\omega - p$ domain (d), and the difference with the true velocity spectrum (e); and the final velocity spectrum for the FWI result in the $\omega - k$ domain (f), and the difference (g). In this example, the FWI performs better in the $\omega - k$ domain than in the $\omega - p$ domain.

The inversion result of the shear velocity structure shown in Figure 2.5c obtained with the expression of the data in the (ω, p) domain leads to the data displayed in Figures 2.6d-2.6e and in Figures 2.7d-2.7e. One can observe that the data misfit is still quite strong. It is also interesting to examine the data in the domain related to the optimization procedure. Figure 2.8 shows the data in the (ω, p) domain and how the inversion attempts to fit some particularly energetic packets leading to some strong misfits in specific parts of the domain.

The inversion result of the shear velocity structure obtained using the expression of the data in the (ω, k) domain leads to Figures 2.6f-2.6g, where one can note that the data misfit is much better than for the (ω, p) domain. The data is also represented in the data domain where the optimization is performed in Figure 2.9. The inversion performs much better than in other domains as can be seen from the difference between the true data and the recovered data (Figure 2.9e). The pattern one can see in the final data structure has been successfully updated from the simple initial dispersion curves, and the velocity spectrum has been filled-in.

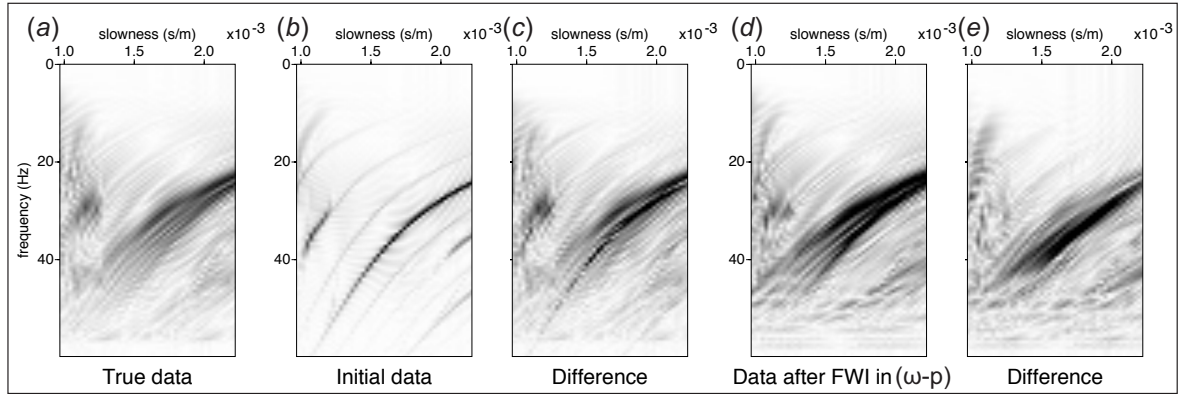


Figure 2.8: The true data in the (ω, p) domain (a); the initial data and the difference to the true data (b,c); the recovered data after FWI in the (ω, p) domain and the difference to the true data (d,e). Some particularly energetic packets can be observed.

2.2.7 Conclusion and Perspectives

We have presented a systematic framework to exploit the full velocity spectrum of surface waves, while explicitly taking into account their dispersive behaviour, as part of a FWI scheme. We have formulated the inversion through a local optimization procedure,

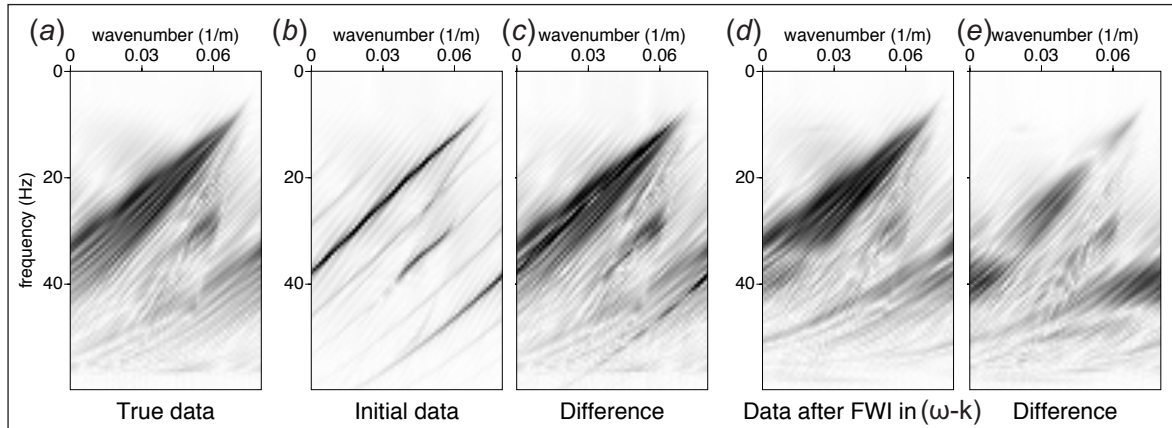


Figure 2.9: The true data in the (ω, k) domain (a); the initial data and the difference to the true data (b,c); the recovered data and the difference to the true data (d,e). One can see a significantly reduced misfit between the initial data difference and the final data difference with less energetic patterns than for the (ω, p) domain.

considering data transformed to alternative data domains such as the (ω, p) domain and the (ω, k) domain. We have shown that we are able to extract useful information from surface waves in an automatic way, and without relying on dispersion curve picking or assuming a layered medium. Furthermore, the workflow is still affordable as it relies on a local optimization procedure and avoids a systematic exploration of the model space, which would require a more intensive exploration workflow.

The synthetic example we consider has a relatively dispersive behaviour as the model contains combined vertical and lateral velocity gradients. A systematic investigation of the inversion procedure when considering more complex models will be the purpose of further work. Surface waves may have strong energetic pulses and can generate complex forward and backward scattering in complex subsurface environments. We therefore expect that specific windowing and filtering approaches may be necessary to perform inversion in more complex cases, although such operations should be relaxed in the final steps of the inversion.

2.2.8 Acknowledgements

The authors thank TOTAL E&P for support and for permission to show these results. This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO,

SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d’Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS under the allocation 046091 made by GENCI. Authors appreciate fruitful discussions with Ludovic Métivier (LJK-CNRS, Université Grenoble Alpes) and Stéphane Operto (GEOAZUR, Université de Nice-Sophia Antipolis).

2.3 Other proposed misfit functions

A complete comparison of alternative misfit functions is presented in [Masoni et al. \(2013\)](#) (see appendix A), which analyzes difference based and cross-correlation based misfit functions, for various data domains, as well as a Singular Value Decomposition (SVD) based misfit function. The analysis validates the work by [Pérez Solano \(2013\)](#) and [Pérez Solano et al. \(2014\)](#), which propose the same robust misfit function in the frequency-wavenumber (ω, k) domain, but also implement a windowing in offset. In the time-offset (t, x) domain, this windowing in offset helps with cycle-skipping effects, as the long offsets which are most cycle-skipped are removed. In the (ω, k) domain, the offset windowing helps dealing with lateral velocity variations and 2D effects. An optimal maximum offset value is determined by the equilibrium between the smearing that occurs with long offsets and amplitude instabilities for short offsets. The added value of the surface waves included in the FWI remains however slightly ambiguous, since for the synthetic tests shown by [Pérez Solano \(2013\)](#), the P-wave velocity model is equally well recovered by the inversion as the shear wave velocity model. However the inversion of the Lamé parameters rather than the P-wave and shear wave velocities directly may have played an important role in the results.

Work by [Yuan et al. \(2014\)](#) and [Yuan et al. \(2015\)](#), follows a different approach for FWI with surface waves. An envelope misfit function is implemented as a more robust alternative to the classical misfit function. Furthermore a particular multiscale approach based on a wavelet decomposition is applied during inversion to help with cycle-skipping issues. The wavelet decomposition, allows to invert low to high frequency data, as shown by the frequency spectra for wavelet functions of different scales in [Figure 2.10](#). This approach helps mitigate cycle skipping by updating the model from large scale to more

detailed features, similar to conventional frequency continuation multiscale approaches. The presence of low frequency data is however critical for successful convergence. [Yuan et al. \(2015\)](#) separately inverts surface waves and body waves, showing the significant contribution of both wave types in the synthetic example they present.

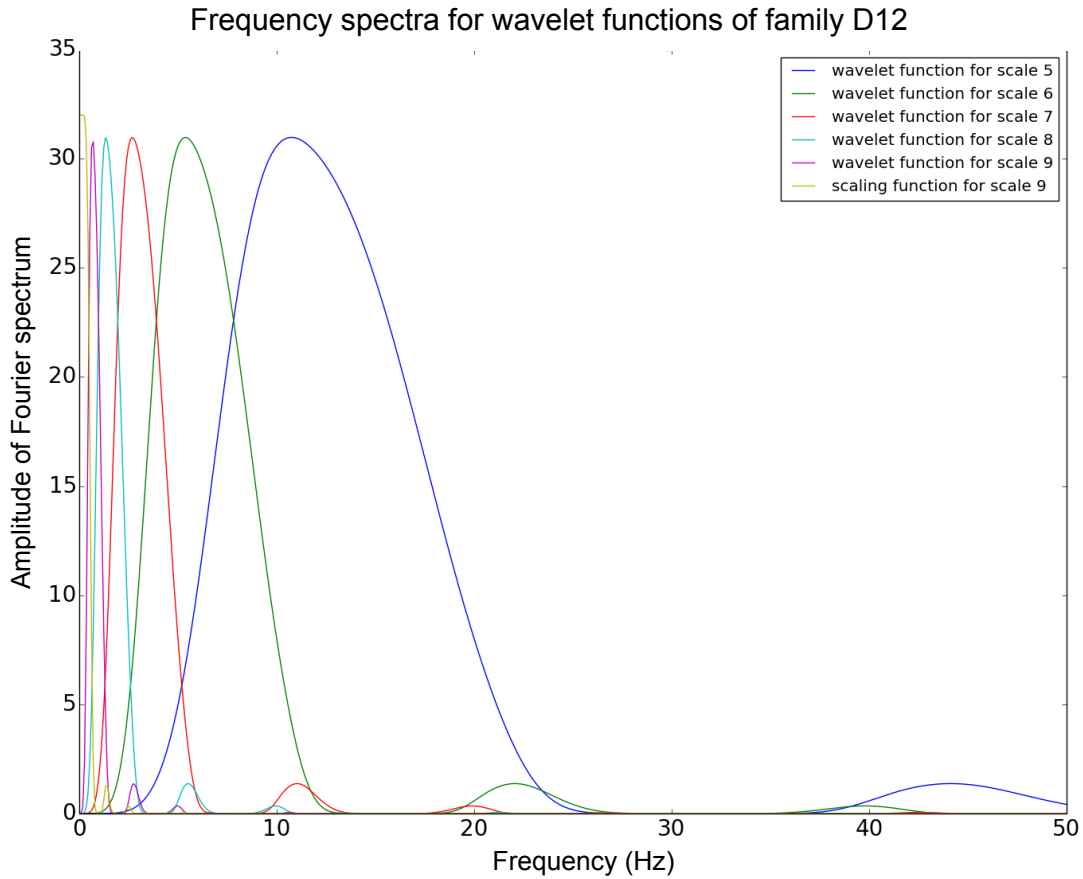


Figure 2.10: Frequency spectra for wavelet functions of family D12 as used by [Yuan et al. \(2015\)](#). Courtesy of Jean-Baptiste Laffitte, 2015

[Silvestrov et al. \(2015\)](#) performs a numerical sensitivity analysis of the Hessian matrix, considering FWI with surface waves. The study further validates that an (ω, k) domain misfit function, as proposed by [Masoni et al. \(2013\)](#) and ([Pérez Solano et al., 2014](#)), is more robust for the inversion of surface waves.

Chapter 3

Strategy for FWI with Surface Waves

Contents

3.1	Introduction	88
3.1.1	Depth preconditioning	91
3.2	Layer stripping approach	93
3.2.1	Motivation	95
3.2.2	Layer stripping workflow	96
3.2.3	Illustrative synthetic example	100
3.2.4	Parameter selection analysis	112
3.2.5	Parameter sensitivity and misfit function selection	117
3.2.6	Quality control	121
3.3	Far initial models	123
3.3.1	Far initial shear velocity model	123
3.3.2	Far initial P-wave velocity model	127
3.4	Conclusion and perspectives	131

3.1 Introduction

In the previous Chapter 2, several alternative more robust misfit functions for Full Waveform Inversion (FWI) with surface waves, are investigated and compared. The selection of a suitable misfit function is a key ingredient that needs to be considered. Yet other steps in the inversion process may also play important and critical roles in the development of a strategy to enable successful FWI using surface waves. This chapter investigates and discusses an inversion strategy specifically adapted to the physics of surface waves. The synthetic model used for the tests performed in this chapter, is taken from Pérez Solano et al. (2014), and illustrated in Figure 3.1.

This model was constructed to explore near surface problematics, considering a maximum depth of 12 *m* over a distance of 58 *m*. The S-wave and P-wave velocity models are related by a constant poisson ratio ($V_P/V_S = 2.0$), with velocities that generally increase with depth. Two high velocity anomalies at the center of the model, are the targets to be recovered during inversion. The velocities of the models are quite low compared to typical values found for oil and gas reservoir models, and are consistent with the near surface region that is considered. A homogeneous density model of $\rho = 1000 \text{ kgm}^{-3}$ is used.

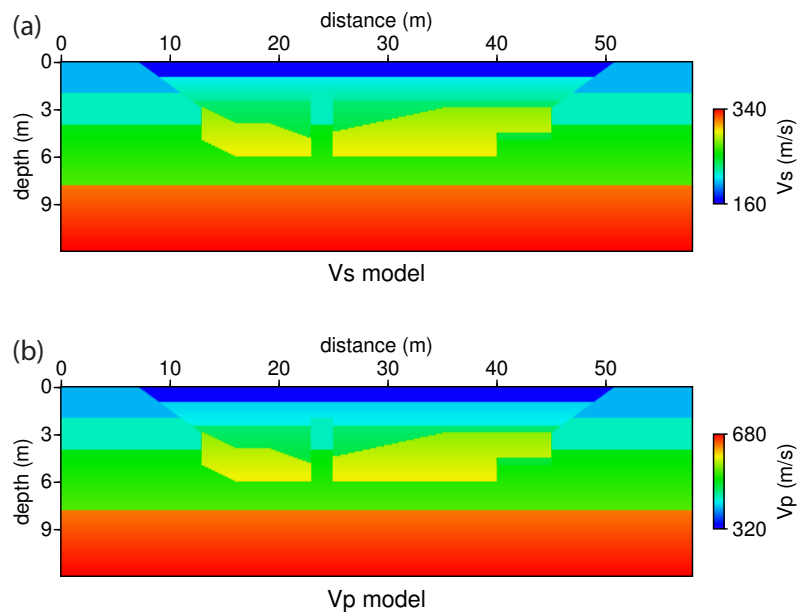


Figure 3.1: The true S-wave (a) and P-wave (b) velocity models used to create the synthetic dataset for this investigation. A homogeneous density of 1000 kgm^{-3} is considered. Models used with permission from Pérez Solano et al. (2014).

Similar testing conditions are used as in Pérez Solano (2013). Regarding the acquisition, the synthetic data is recorded by 145 vertical and horizontal component receivers, placed 0.2 m below the surface, with an equidistant spacing of 0.4 m between receivers. For all inversion tests, both vertical and horizontal receiver components are inverted together. To create the dataset, 20 vertical sources are simulated. These are also positioned at 0.2 m below the surface, with an equidistant spacing of 2.5 m between each source. The source wavelet used is the primitive of a 40 Hz Ricker wavelet. The forward modelling scheme implemented is the same as described in section 2.2.3, using a staggered finite-difference grid. The spatial discretization step of the synthetic models is set to 0.1 m , to guarantee at least 10 discretization points per wavelength. The time discretization step for the simulated receivers is set to 0.04 ms to satisfy the Courant-Friedrichs-Lewy (CFL) stability condition given in Equation 1.52 of section 1.3.1.1. An example shot of the resulting true dataset is shown in Figure 3.2. One can observe that the surface waves are of a dispersive nature, and are the signal with the most dominant amplitude.

First the conclusions on misfit function robustness from Chapter 2 are validated, by testing the proposed alternative misfit functions described in section 2.2.4. Inversion is performed using both the conventional L2 norm of the difference in the time-offset (t, x) domain, and the more robust misfit in the frequency-wavenumber (ω, k) domain. For these tests, only the shear velocity was inverted for, using true and known P-wave velocity and density models. The inversion was performed using an l -BFGS optimization (Byrd et al., 1995; Nocedal and Wright, 2006), with a small wavelength adaptive Gaussian smoothing applied to the gradient after each iteration (see section 1.3.1.2 for more details). The SEISCOPE optimization toolbox (Métivier and Brossier, 2016) was used to implement this minimization scheme. The initial shear velocity model (see Figure 3.3a) consists of a linear gradient, with correct minimum and maximum velocities as found in the true model, and similar to the one used in Pérez Solano (2013). The aim is to minimize the difference between the true and the calculated data and to recover the two high velocity anomalies in the true model.

The initial difference between true data and the one corresponding to the initial model used (Figure 3.2c), illustrates that almost all of the dispersed surface waves still need to be explained by the inversion. The strong cycle-skipping that can be observed from the initial residual, does not allow convergence when using the conventional (t, x) domain misfit. When FWI is attempted, the inverted shear velocity model diverges at the first iteration, and remains stuck in a local minimum, as illustrated in Figure 3.3b. By

analyzing the resulting data (Figure 3.2d) that correspond to the shear velocity model obtained, one can observe that the phases have been wrongly matched, not allowing the inversion to converge towards the global minimum.

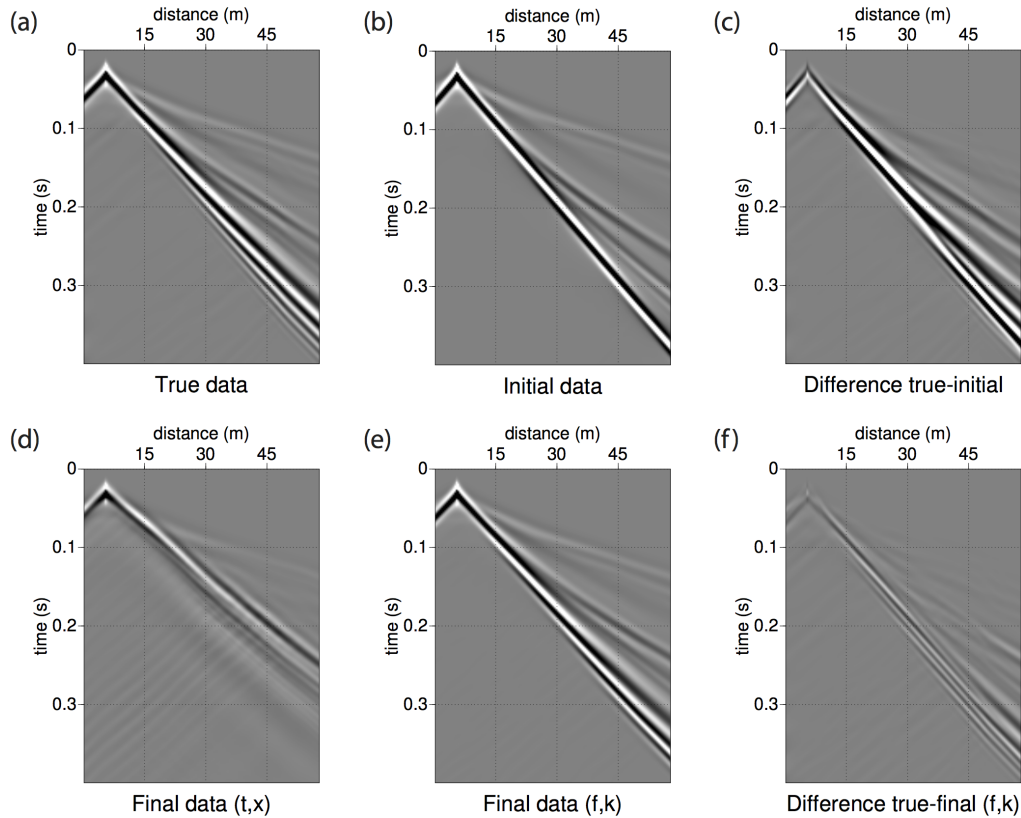


Figure 3.2: Example of a common shot gather for the true data (a); the initial data (b) and the difference with the true data (c); the final data after FWI using the conventional (t, x) domain misfit (d); the final data after FWI using the (ω, k) domain misfit and depth preconditioning (e) and the difference with the true data (f). A small gain is applied to the true data (a) for visualization purposes only, all other figures (b) to (f) have the same color scale as (a).

When the more robust (ω, k) domain misfit is implemented, the shear velocity is updated in the top of the model, but little or no update occurs at depths below 5 m, as can be observed in Figure 3.3c. This is most likely due to the characteristic exponential decrease in amplitude with depth of surface waves (detailed in section 1.1.3). Applying a depth preconditioning to the model update can be shown to greatly improve the inversion result.

3.1.1 Depth preconditioning

A preconditioned l -BFGS optimization (Byrd et al., 1995; Nocedal and Wright, 2006) is investigated and implemented. From an initial guess m_0 , each new iteration is given as

$$m_{k+1} = m_k + \alpha_k \Delta m_k, \quad (3.1)$$

where α_k is the steplength computed from a linesearch, and Δm_k is the model update for iteration k . When a preconditioning is implemented, the descent direction is given as

$$\Delta m_k = -\tilde{Q}_k \nabla f(m_k), \quad (3.2)$$

where $\nabla f(m_k)$ is the gradient of the function $f(m_k)$ at point m_k , and \tilde{Q}_k is the l -BFGS approximation of the inverse Hessian operator $H(m_k)^{-1} = \nabla^2 f(m_k)^{-1}$ computed from an initial estimation P_k of $H(m_k)^{-1}$. The preconditioning P_k applied in this investigation is a vertical scaling linear with depth z , and is applied on the inverse Hessian approximation. This weighting increases the importance of the model update at depth compared to the near surface, and partially compensates for the decrease of surface wave amplitudes with depth.

Such depth preconditioning strategies have already been used in applications with electromagnetic waves as well as with surface waves (Plessix and Mulder, 2008; Pérez Solano, 2013), and are essential to balance the model update and retrieve velocity anomalies at depth.

The final shear velocity model obtained after FWI using the (ω, k) domain misfit function and a depth preconditioned optimization, is illustrated in Figure 3.3d. The two high velocity anomalies, present in the true model, can be identified. However the model remains of low resolution and details are not recovered. The final data and the difference to the true data set, for the resulting shear velocity model, is illustrated in Figures 3.2e and 3.2f. The final data reveals that the dispersion observed in the true data has been at least partially explained. When comparing the final difference to the difference between the true and the initial data, illustrated in Figure 3.2c, one can observe that the data is relatively well recovered. The data misfit is minimised to less than 5 percent of the initial value. Yet both high and low frequency content still remains in the final data difference, and the far offsets are less well explained (see Figure 3.2f). The depth preconditioning does not however improve the result obtained from FWI when using the (x, t) domain misfit function, as the data remains cycle-skipped, and the inversion does not converge.

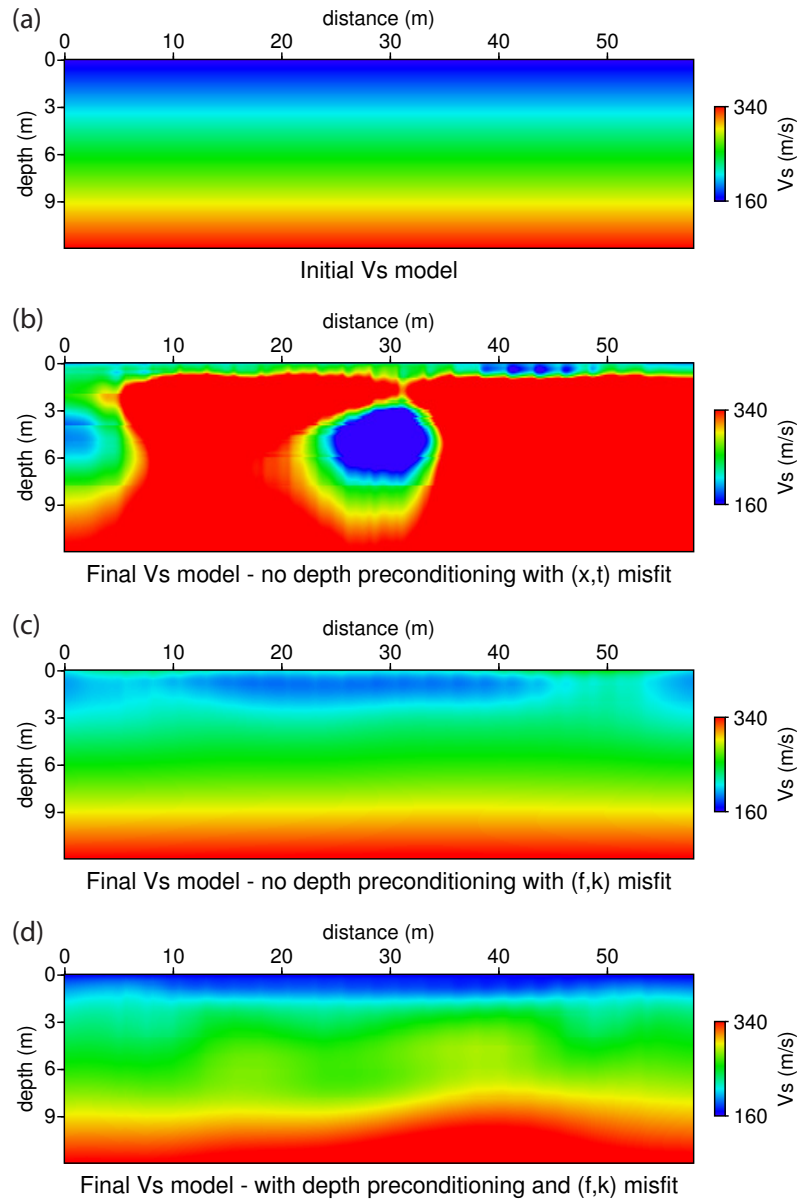


Figure 3.3: Shear velocity model inversion results, using the initial model (a); using the conventional (x, t) misfit (b); using the more robust (ω, k) domain misfit without a depth preconditioning (c); and using the more robust (ω, k) domain misfit with a depth preconditioning implemented (d).

This result further validates the conclusions of section 2.2.7: the (ω, k) domain misfit function is found to be more robust to cycle-skipping than the conventional misfit function computed in the (t, x) domain, although a lower resolution in the resulting velocity model can be expected. A sequential inversion with the conventional (t, x) domain misfit function, using the final model obtained with the (ω, k) domain misfit function as an

initial model, unfortunately does not converge, as the cycle-skipping is not completely overcome.

This investigation also highlights the critical importance of implementing a depth preconditioning for FWI using surface waves. For these tests, no frequency filtering strategy was implemented, and all frequencies were therefore inverted for simultaneously. However it is important to note that the presence of low frequency content in the data is also an essential factor that will allow, or limit, convergence. For example, similar inversion tests using a source wavelet with a lower peak frequency were found to converge better, with lower final data misfit values, and successful final shear velocity models. The following section investigates which frequency filtering strategy is most adapted for a successful inversion of surface waves.

3.2 Layer stripping approach

Layer stripping is a well known approach used in inversion methods ([Gibson et al., 2009](#); [Shi et al., 2015](#)). Initially, only the top layer of a model is recovered, then the underlying layers are sequentially updated, as observable data corresponding to the upper layers will have been previously explained. For this approach, different parts of the data need to explain properties at different layer depths of the model, as the model is recovered layer by layer, in a top-to-bottom manner.

Such layer stripping approaches have already been proposed for FWI applications using reflection data, where a complex near surface, or overburden region, prevents successful imaging of deeper targets ([Wang and Rao, 2009](#); [Bian and Yu, 2011](#)). In such cases layer stripping approaches can remove the effect of upper layers and help the inversion. [Wang and Rao \(2009\)](#) propose a procedure in which the top layer of the velocity model is fixed, and the data of the corresponding time window frozen. Gradually the model is allowed to update layer by layer, until only the bottom layer is retrieved, using data recorded during later and later time intervals. One of the pitfalls of this method is the need to correctly relate the chosen time interval to the associated layer depth explained by the data. [Bian and Yu \(2011\)](#) argument that, due to the relation between model velocity, depth of model and time taken, the method is not adapted for lateral heterogeneities. They propose a similar method, where a damping is applied to both observed and calculated datasets in the misfit function. The relevant sensitivity kernels and Born wave-paths suggest that larger damping constants will result in a reduced

contribution from deeper targets to the wavefield response. Therefore, as the damping constants applied are gradually decreased, successively deeper parts of the model appear to be "automatically" allowed to update.

A similar progressive time damping FWI approach is used by [Shipp and Singh \(2002\)](#), which also implement a progressive offset windowing. Initially only the shallow part of the model is inverted for, by computing the model update only up to a certain depth. The following stages however, first aim to recover the deeper large scale structures with undamped, long offset data, and afterwards the smaller scale features are progressively recovered with damped, short offset data. A good relation between the offset selection and the model depth recovered with FWI is illustrated for reflection data. The reverse strategy is used by [Métivier et al. \(2016\)](#) which apply a time windowing with progressively increasing offsets for FWI in a layer stripping, top-to-bottom approach, to help recover salt structures present in the model.

Beyond the interest in reducing the effects of the top layers on the recovery of deeper structures, applying a layer stripping approach to FWI with surface waves, also allows one to take advantage of a special physical property. Surface waves of shorter wavelength and higher frequency will sample the top layers of a medium, while waves with longer wavelengths and lower frequencies will sample deeper parts of a medium (see section [1.1.4](#) and [Figure 1.3](#)). As such, when applying a frequency filtering strategy on the data, it will contain information related to a specific depth of the model, which can then be recovered. In the following, a high-to-low frequency filtering strategy, complemented by a progressive offset tapering, is proposed for FWI with surface waves, to update first the upper part of the model and then the deeper, in a layer stripping approach.

A high-to-low frequency filtering strategy goes against all literature on FWI. To avoid cycle-skipping issues, a multiscale frequency continuation approach is conventionally applied ([Bunks et al., 1995](#); [Sirgue and Pratt, 2004](#)), inverting low-to-high frequency content of the data. The progressive increase in frequency corresponds to updating first the large-scale and then the more detailed features of the velocity model. Further details on this strategy can be found in section [1.3.2.1](#). Such a strategy is implemented in [Pérez Solano \(2013\)](#) for FWI with surface waves. The wavelet-multiscale approach employed by [Yuan et al. \(2014\)](#) for FWI with surface waves, is also equivalent to filtering first the low frequency content and moving to higher frequencies as the inversion progresses (see [Figure 2.10](#)).

3.2.1 Motivation

A high-to-low frequency filtering strategy can be shown to be a natural way to deal with the inversion of surface waves, when observing the relation between the frequency content of the data and the penetration depth of the model update. Sensitivity kernels (or Frechet derivatives) play an important role in understanding and analyzing an inversion method. They provide an insight on both the sensitivity and the resolution of the model update that can be obtained from the information contained in the data. Sensitivity kernels are computed as the first gradient calculation for one source and one receiver.

Figure 3.4 shows two sensitivity kernels, obtained from summing over all sources and receivers, using data filtered by two different frequency bands. The effect of the frequency content of the data on the model update is therefore analyzed, using as a test case the model and data from the previous section 3.1. When a high frequency bandpass filter of $70 - 110 \text{ Hz}$ is applied, the update appears restricted to the upper layer of the model, as can be observed in Figure 3.4a. When a lower frequency bandpass filter of $35 - 55 \text{ Hz}$ is applied, the update has significant amplitudes down to a depth of about 6 m as shown in Figure 3.4b. These results reflect the relationship between frequency and depth penetration illustrated in Figure 1.3 and further discussed in section 3.2.4.

One disadvantage of a high-to-low frequency filtering strategy however, is that the resolution of the model update will decrease progressively with each frequency band used. One can observe from Figures 3.4a and 3.4b that when data of low frequency content is used, the resolution of the model update is lower. As such, the resolution will also decrease with the depth of the model.

In the following layer stripping strategy proposed, data filtered by sequential frequency bands, going from high to low frequencies, are inverted. The resulting model at each frequency step is used as an initial model for the next frequency band. The gradient is frozen below an estimated penetration depth of the surface waves. To avoid a deterioration of the upper layers, after they have been obtained using high frequency data, and therefore with a high resolution, the top of the gradient is also frozen for later updates with frequency band steps of lower frequency content. Such windowing specifically targets the use of surface waves to update the model. Body waves are not explicitly removed from the data, and may still be used to update the model. However the combined filtering and windowing of the gradient favours the information coming from surface waves to update the model. This strategy is therefore specifically adapted to near surface applications of FWI using surface waves.

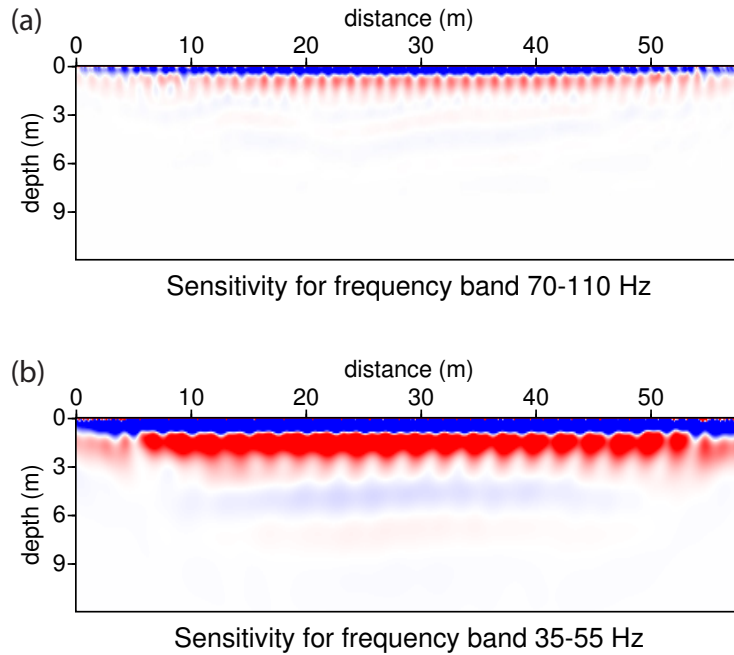


Figure 3.4: Sensitivity kernels, summed over all sources and receivers, for data of frequency band 70 – 110 Hz (a) and 35 – 55 Hz (b). One can observe that with the lower frequency band the model is updated at greater depths, but is also of lower resolution. The color scale for the amplitude is the same for both sensitivity kernels. Note that a wider frequency range is chosen for the high frequency window, due to a lower amplitude of the data spectrum at these frequencies, to compensate for an equal amount of energy (see section 3.2.4).

3.2.2 Layer stripping workflow

The workflow proposed for the layer stripping strategy, integrated within a FWI scheme, is illustrated in Figure 3.5. Similar to a conventional FWI scheme, starting models for each model parameter (V_p, V_s, ρ) are required, as well as the acquisition geometry to simulate the true acquisition of the observed data, and an estimation of the source wavelet to use for the forward modelling.

The input parameters related to the layer stripping strategy include the minimum and maximum frequencies to define the frequency band that will be used, and the minimum and maximum depth of the layer to be updated. The choice of these parameters, the width of the frequency bands, and the depth of the layers, are discussed in more detail in section 3.2.4.

The velocity of the surface waves is assumed to depend mainly on the shear wave velocity as shown in Equation 1.36. The average frequency, and the average shear velocity

bands are used. The progressive windowing enforces the layer stripping approach, focusing only on surface waves that contain information on the layer to be updated. Inspired by techniques applied to conventional surface wave analysis using dispersion curves, the offset windowing allows a valid 1D assumption, and therefore to obtain a better image of the local fundamental surface wave mode (see section 1.2). This can be an advantage when the data misfit is calculated in the (ω, k) domain. In addition, the limited offsets reduce the dependence on the initial model, helping to avoid cycle-skipping problems and the presence of local minima. The windowing is implemented so that the whole dataset is considered when the lowest frequencies are reached.

The maximum offset length is determined as a function of the penetration depth for each frequency band step. The relation between the penetration depth and a suitable offset range, was investigated by analyzing the model update for a single frequency and varying offset ranges, using a homogeneous initial model. For the model and acquisition described in section 3.1.1, the point where an increase in offset range no longer brought a significant amount of information to the model update was found to be close to

$$x_{max} \simeq 5 \left(\frac{\bar{V}_S}{\bar{f}} \right) \simeq 5 \bar{\lambda}_S, \quad (3.3)$$

where x_{max} is the maximum offset allowed, \bar{V}_S the average shear velocity (in m/s), \bar{f} the average frequency (in Hz), and $\bar{\lambda}_S$ the average wavelength of surface waves (in m). A windowing, determined by the maximum offset (x_{max}) allowed between a receiver position and the source, is applied to the data in the time-offset (t, x) domain. A Hanning function, given as $H(i) = \left(0.5 \left(1 - \cos\left(\frac{\pi i}{N-1}\right)\right)\right)^2$, where N is the length of the taper, is applied on either side of the offset window, to avoid ringing artifacts due to Gibb's phenomenon. During inversion tests, the x_{max} parameter was not found to be very critical, and is therefore calculated automatically within the algorithm following Equation 3.3, using the initial model for each frequency step to compute an average shear velocity. It should be noted however, that the relation described in Equation 3.3 may be case dependent, and may therefore have to be modified for datasets at different scales or with different acquisition parameters.

This offset windowing step is similar in some aspects to the offset windowing in the w-AWI workflow presented by Pérez Solano et al. (2014). Both propositions aim to help with cycle-skipping issues, and increase the lateral resolution during inversion. Yet differently to work by Pérez Solano et al. (2014), the offset windowing in this workflow follows a layer stripping approach, since it is directly driven by the penetration depth of

the surface waves used to update the shear velocity model.

After being windowed in offset, observed and calculated data are transformed to the frequency-offset (ω, x) domain by the application of a Fourier transform on the time axis. For a given source and frequency, a normalization by the root mean square (*rms*) of the data is applied. This allows to whiten the data, giving each frequency the same weight. The very low and very high frequency content is therefore boosted, making their contribution to the data misfit important. Although this may be problematic in cases of low signal-to-noise ratio, the information from a wide frequency range is required for inversion of surface waves.

A band-pass filter is subsequently applied to the data, with cutoff frequencies determined by the input minimum and maximum frequency parameters. As with the offset windowing, a Hanning taper on either side of the filter is implemented in the frequency domain, to avoid ringing artifacts due to Gibb's phenomenon. An inverse Fourier transform on the frequency axis is then applied to the data to return to the time-offset (t, x) domain.

The misfit between the observed and calculated data is then obtained using either the conventional L2 norm of the difference in the (t, x) domain, or the more robust misfit function in the (ω, k) domain discussed in section 2.2.4. The misfit formulations, are slightly modified from section 2.2.4 to include the offset windowing applied during the layer stripping strategy. The modified formulations of the two misfit functions are given as

$$C_{t,x} = \frac{1}{2} \sum_S \sum_R \mathbf{w}_R (\mathbf{d}_{obs}(t, x) - \mathbf{d}_{cal}(t, x))^2, \quad (3.4)$$

and

$$C_{\omega,k} = \frac{1}{2} \sum_S \sum_R \mathbf{w}_R (|\mathbf{d}_{obs_N}(\omega, k)| - |\mathbf{d}_{cal_N}(\omega, k)|)^2, \quad (3.5)$$

where \mathbf{w}_R is the weighting (between 0 and 1) applied to the receivers R during the windowing in offset, and determined by the maximum offset (x_{max}) and a Hanning function as described above.

The gradient is then calculated for the relevant misfit function used, as described in section 2.2.5 using the adjoint method. The adjoint source is slightly modified to include the windowing in offset.

Only the gradient values within a depth layer defined by the minimum and maximum depth parameters are considered. The minimum depth is introduced to avoid reducing

the resolution of the top part of the model, already retrieved with high frequency data, and the maximum depth is defined by the penetration limit of surface waves. A Hanning function is used to taper both sides of the depth window, to avoid abrupt changes in the model update. The tapering applied also reduces the sensitivity to the choice of minimum and maximum depth parameters, and is therefore of significant length. The choice of tapering is further discussed in section 3.2.4.

A wavelength adaptive Gaussian smoothing with a small correlation length is then applied to the gradient to remove high frequency artifacts, such as an imprint of the acquisition. The smoothing parameters are kept constant for all frequency band steps. All tests presented in this chapter implement correlation lengths, given as a fraction of the local wavelength considering a 40 Hz source, of 0.4 in the x direction and 0.2 in the z direction. The smoothing applied may also smoothen the edges of the depth layer update, and a leakage of the gradient to above or below the layer may occur. A depth preconditioned l -BFGS optimization, as described in section 3.1.1, is used for the inversion. This accounts for the exponential decrease of surface wave amplitudes with depth. A convergence criterion as well as a maximum number of iterations are used as stopping criteria to define a final model, which is then used as an initial model for the following frequency band. For all tests in this chapter, the inversion is allowed to continue until the update is minimal, and the relative misfit function value, normalized by the initial value, is less than 0.0001. Although only a maximum number of 50 iterations are allowed, convergence always occurs before this limitation is reached.

The workflow is repeated for each frequency band. To reduce the computing time the algorithm is parallelized over each shot gather with a Message Passing Interface (MPI) communicator to perform the global misfit summation and gradient calculation at each iteration. The final model obtained with the last frequency band is the resulting model obtained with layer stripping FWI.

3.2.3 Illustrative synthetic example

The model and acquisition geometry presented in section 3.1 are used to test and compare the proposed layer stripping high-to-low frequency approach to the conventional multiscale low-to-high frequency strategy. For the inversion only the shear velocity parameter is inverted for, while P-wave velocity and density models are considered as true and known. The true source wavelet is also considered as known for the inversion. The initial shear velocity model is a linear velocity gradient with true minimum and maximum

velocities as illustrated in Figure 3.3a.

The final models after FWI using the layer stripping approach, are obtained after 8 frequency band steps, following the workflow presented in previous section 3.2.2. The frequency bands used are 70 – 110 Hz, 60 – 90 Hz, 50 – 80 Hz, 45 – 70 Hz, 40 – 60 Hz, 35 – 55 Hz, 25 – 45 Hz and 20 – 38 Hz. These correspond to overlapping layer depths, on which a significant depth tapering of 1.2 m is applied. The choice of the minimum and maximum frequency and depth window parameters (see Figure 3.16), as well as the choice in taper length, are discussed in the following section 3.2.4.

Results for both the more robust misfit function in the (ω, k) domain, as well as the conventional L2 norm of the difference in the (t, x) domain are presented. For both misfit functions, the final model after each frequency step is illustrated in Figures 3.6 and 3.7 respectively. It can be observed that the model is updated in a top-to-bottom manner, as lower frequency bands are used. One can observe that the first four frequency band steps, recover the top low velocity layer. The model update during the frequency band step of 35 – 55 Hz, recovers the two high velocity anomalies. Only the last frequency band step of 20 – 38 Hz images the interface with the high velocity deeper layer.

Comparing the two tests for the different misfit functions, one can observe that when the misfit function in the (ω, k) domain is implemented (Figure 3.6), the model update contains a few, small, high and low velocity artifacts. While when the conventional L2 norm of the difference in the (t, x) domain is implemented (Figure 3.7), the results at each frequency step appear smoother. A comparison of the two different misfit functions for layer stripping FWI is discussed in section 3.2.5.

The layer stripping approach appears very effective against cycle-skipping issues, as the inversion for each frequency band step can be observed to converge in the right direction. For both misfit functions tested, the final models, pictured again in Figures 3.8b and 3.8c, are of much higher resolution than obtained when no frequency strategy is implemented (such as in Figure 3.3d). The final models resemble the true model (Figure 3.8d) very well, as can be observed from the data misfit (see Figure 3.14). The two high velocity anomalies are recovered as well as their geometry. In addition, the smaller low velocity anomaly at 3 m depth, as well as the high velocity layer at 7.8 m depth are also partially seen and recovered by the data during the inversion. The inversion does not appear to have difficulties in recovering lateral variations in velocity. The small artifacts present on the extreme ends of the model can be explained by a lack of illumination, as the acquisition does not extend beyond the model.

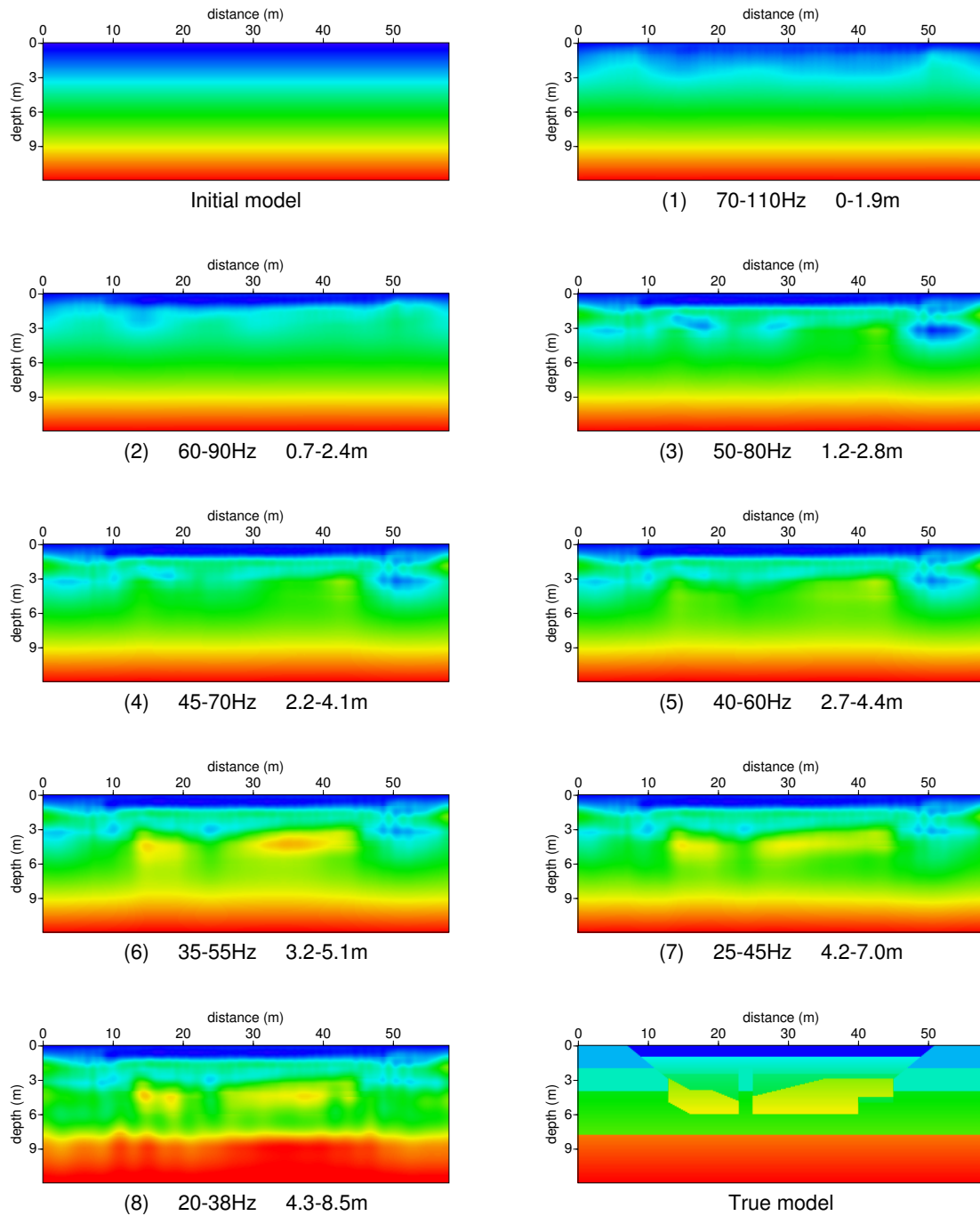


Figure 3.6: Evolution of the resulting shear velocity model at each frequency band step (1) to (8) when using the misfit function in the (ω, k) domain. For each step a frequency band and depth layer are given as parameters to filter the data and window the model update. The depth values mark the mid-point of the 1.2 m taper applied during the depth windowing. This layer stripping strategy progressively updates the model in a top-to-bottom manner, using a high-to-low frequency approach. The initial and true shear velocity models are also shown for reference.

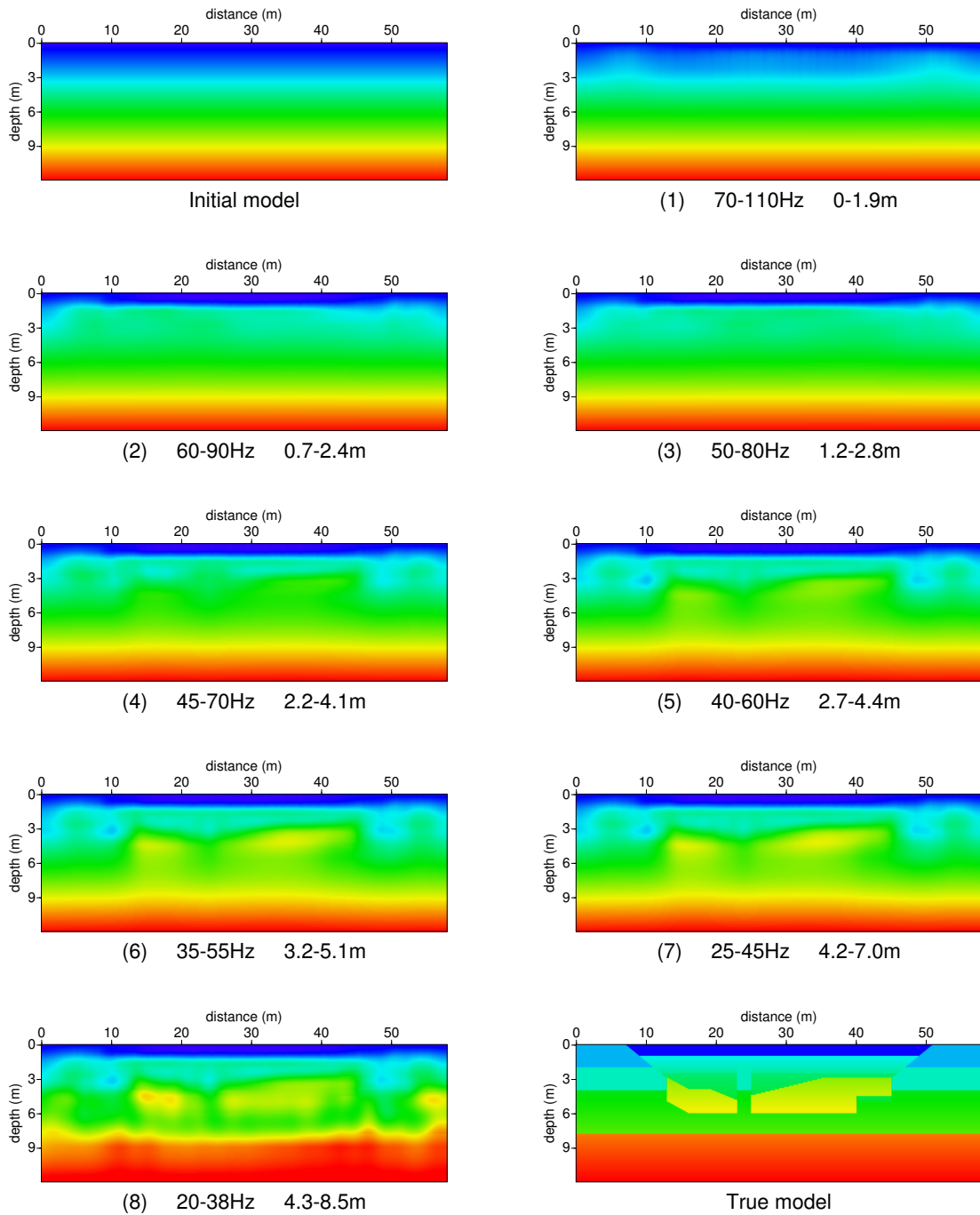


Figure 3.7: Evolution of the resulting shear velocity model at each frequency band step (1) to (8) when using the conventional L2 norm of the difference in the (t, x) domain. For each step a frequency band and depth layer are given as parameters to filter the data and window the model update. This layer stripping strategy progressively updates the model in a top-to-bottom manner, using a high-to-low frequency approach. The initial and true shear velocity models are also shown for reference.

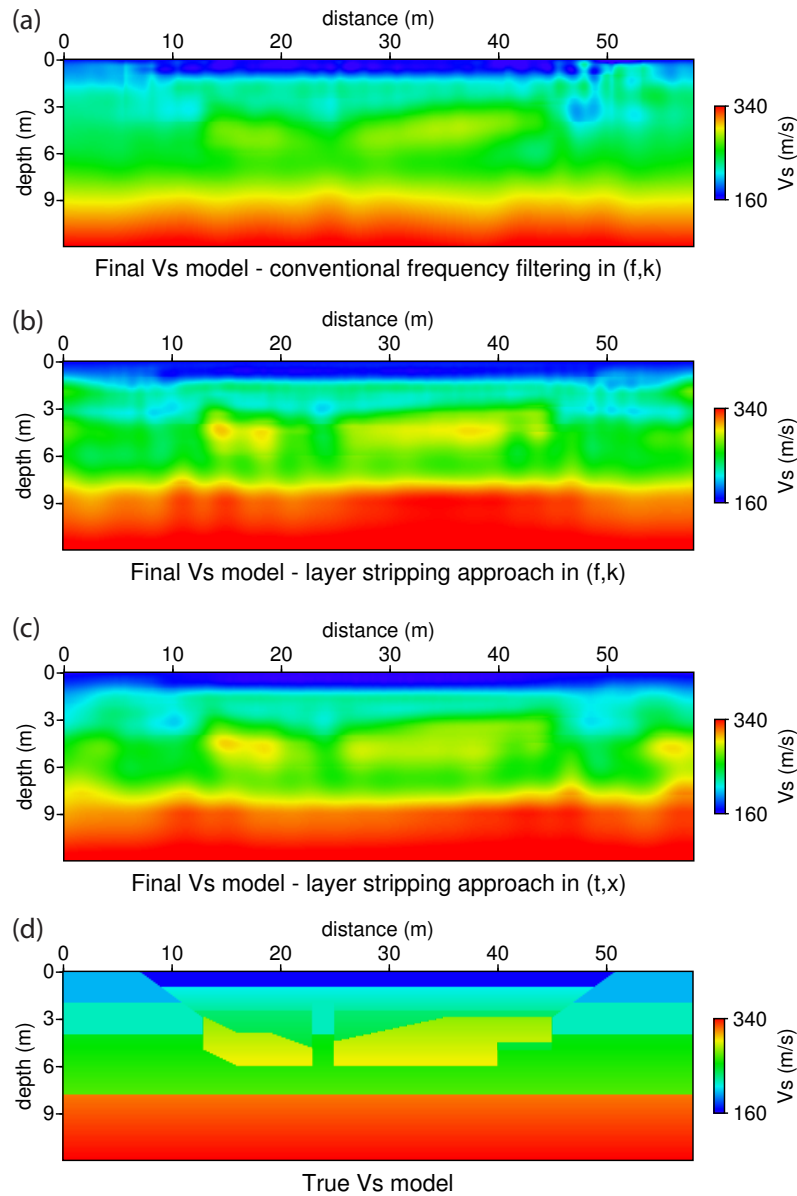


Figure 3.8: Comparison of final shear velocity models obtained after FWI using a low-to-high frequency approach and a misfit function in the (ω, k) domain (a); using a layer stripping approach and a misfit function in the (ω, k) domain (b); using a layer stripping approach and the conventional misfit function in the (t, x) domain (c). The true shear velocity model (d) is shown below for reference.

To better evaluate the results from the layer stripping approach, they are compared to those obtained from FWI using a more conventional low-to-high frequency strategy. For this approach the observed and calculated data are band-pass filtered using 6 successive frequency bands, with a low-cut frequency of 10 Hz and high-cut frequencies of 18 Hz ,

25 *Hz*, 36 *Hz*, 50 *Hz*, 80 *Hz* and 110 *Hz*, progressively allowing high frequency data in the inversion. The final model from each frequency band step is used as the initial model for the next frequency band, as the model update becomes finer at each step with higher frequency detail. No offset windowing is applied and the whole gradient is considered. The misfit function in the (ω, k) domain is implemented. All other FWI parameters are kept consistent to the previous tests.

Starting with low frequencies is conventionally used to overcome cycle-skipping issues, and the final shear velocity model after FWI using this approach (see Figure 3.8a) successfully converges towards the true model. Yet the final models obtained with the layer stripping approach appear to better recover the true model, especially with a higher resolution at depth. The layer stripping allows to previously explain the top layers, so that the deeper part can then be more accurately recovered. Furthermore the depth windowing allows to focus the model update layer by layer, avoiding cycle-skipping issues and local minima in the inversion.

When applying the layer stripping approach, it is essential to implement a frequency window rather than gradually add low frequency content, as otherwise the inversion may get stuck in a local minima, due to cycle-skipping of the high frequency data, when longer offsets are introduced. The layer stripping strategy does not inherently overcome the cycle-skipping limitations during inversion, rather, it is the data selection and gradient windowing which reduce the dependence on the initial model and are critical to avoid inverting cycle-skipped data, allowing convergence.

One of the important advantages of the layer stripping approach is that it does not depend on the low frequency content of the data, which is often missing from real acquisitions, for successful convergence. For conventional low-to-high frequency FWI, the low frequencies are critical for overcoming the cycle-skipping problem and allowing convergence. In both layer stripping tests presented, the lowest frequency used is 20 *Hz*, which is only useful for recovering the deeper part of the model. Instead for the conventional low-to-high frequency FWI test, a low-cut frequency of 10 *Hz* is required, since a deteriorated shear velocity model is obtained when using a low-cut frequency of 20 *Hz*.

To evaluate the results from the two different frequency strategies, the final data and the final difference to the true data are plotted in Figure 3.9 for each test. Although all final datasets recover the true data relatively well (see Figure 3.10), the illustrations of the data difference reveal that the FWI results obtained with the layer stripping approach are slightly better, especially for longer offsets. It is interesting to note that the data

residual after conventional low-to-high frequency FWI has a rather low frequency content compared to the data residual after layer stripping FWI (see Figures 3.9 and 3.10). This suggests that large scale features of the model are not as well recovered with conventional low-to-high frequency FWI compared to when the layer stripping method is implemented.

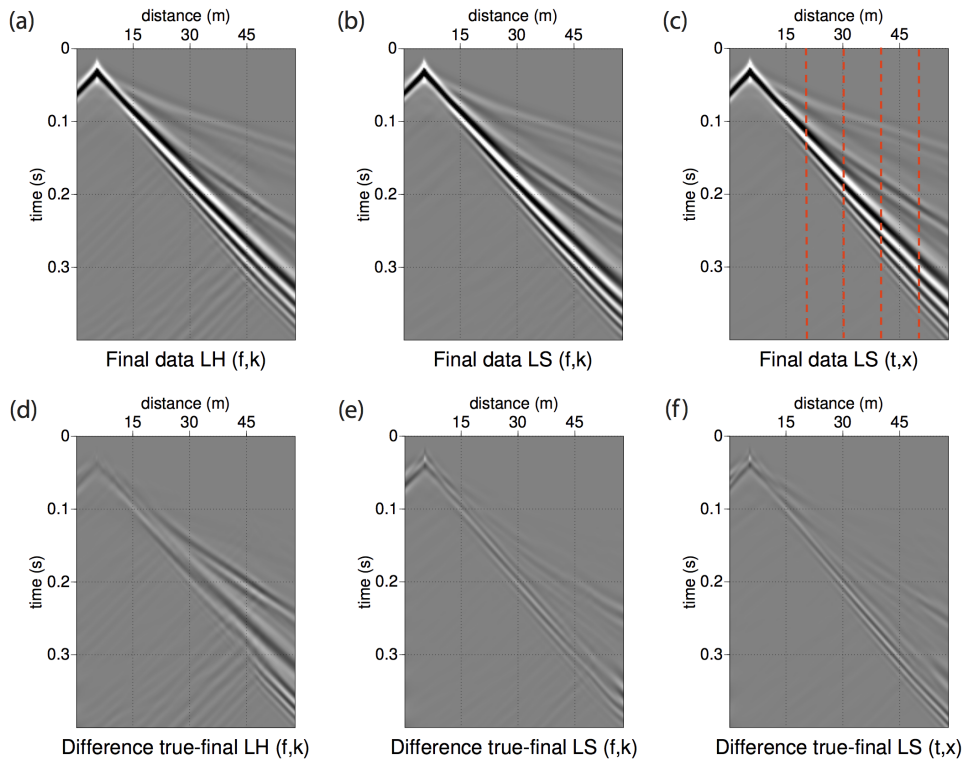


Figure 3.9: Example of a common shot gather for final data after conventional low-to-high (LH) frequency FWI with an (ω, k) misfit function (a) and the difference to the true data (d); final data after layer stripping (LS) FWI with an (ω, k) misfit (b) and the difference to the true data (e); final data after layer stripping (LS) FWI with an (t, x) misfit (c) and the difference to the true data (f). All figures have the same color scale. The traces marked by the red dotted lines are compared in Figure 3.11.

A comparison of individual data traces is shown in Figure 3.11. By comparing the true data (black) with the initial data (grey), one can observe that at far offsets, the phases of the initial data trace match the wrong phases of the true data trace. For misfit minimization in the (t, x) domain, the data are strongly cycle-skipped at far offsets, and may even be cycle-skipped at shorter offsets, although this is less evident.

The comparison with the final data reveals that at short offsets, both the conventional low-to-high frequency FWI result (blue), and the layer stripping FWI result (red), fit the true data well. As further offsets are considered, there is a clear improvement in the

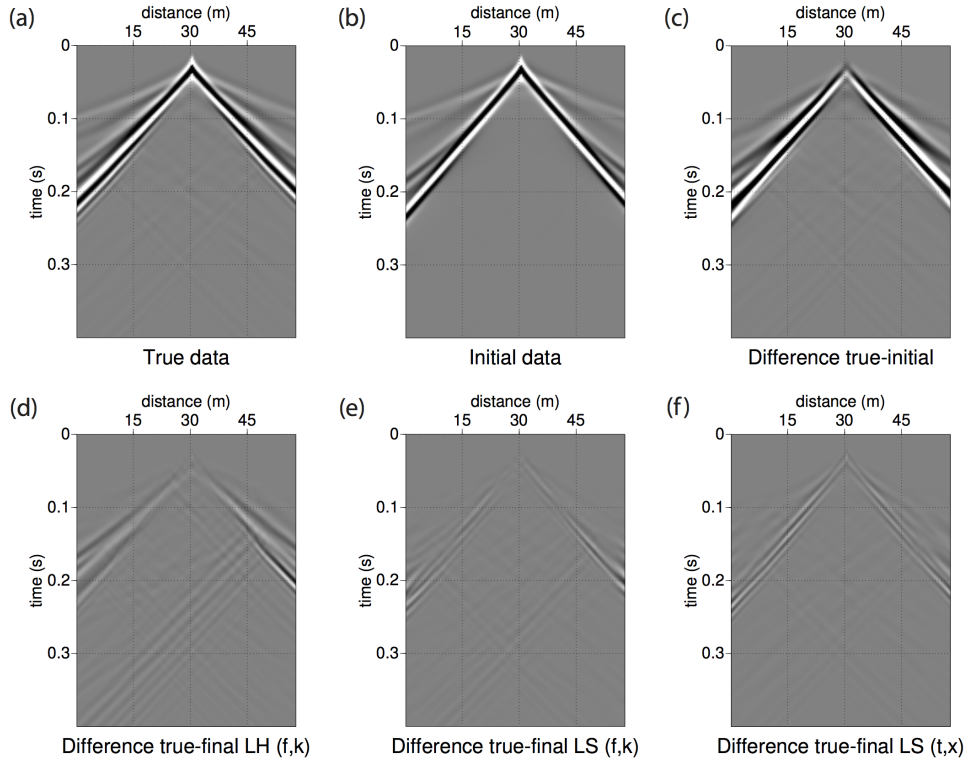


Figure 3.10: Example of a common shot gather for the true data (a); initial data (b); difference between initial and true data (c); difference between final data after conventional low-to-high frequency FWI with an (ω, k) misfit function and the true data (d); difference between final data after layer stripping FWI with an (ω, k) misfit function and the true data (e); difference between final data after layer stripping FWI with an (t, x) misfit function and the true data (f). A small gain is applied to the true data (a) only for visualization purposes, all other figures (b) to (f) have the same color scale as (a).

data fitting when the layer stripping approach is implemented, and the final data trace is no longer cycle-skipped. Instead, the final data trace for the conventional low-to-high frequency approach does not match all the phases correctly. This trace analysis explains the successful results obtained by Pérez Solano et al. (2014) for a similar inversion test (see section 2.3), who use a conventional low-to-high frequency approach, but implement a constant offset window, therefore removing the cycle-skipped data at far offsets, and allowing convergence.

To further compare these results for the different FWI strategies tested, the final data and difference to the true data is also analyzed in the frequency-wavenumber (ω, k) and frequency-velocity (ω, c) domains. As can be observed in Figure 3.12c, the fundamental mode of the initial data is shifted compared to the fundamental mode found for the

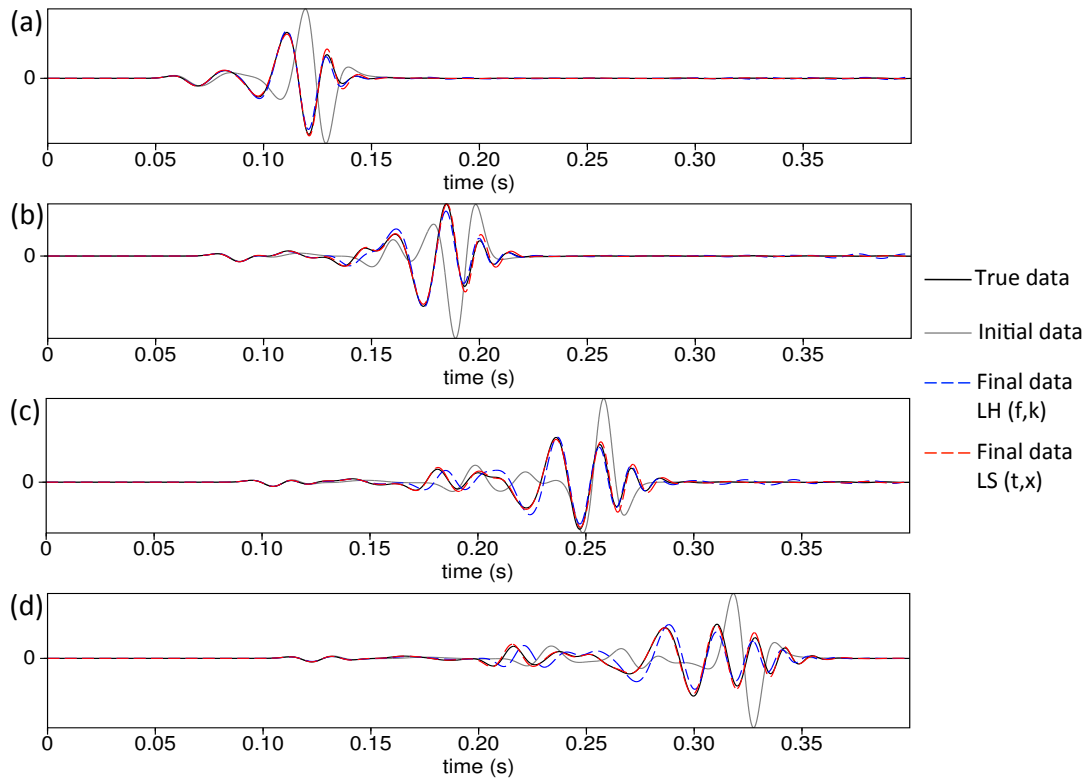


Figure 3.11: Comparison of a single trace at varying offsets of 15 *m* (a), 25 *m* (b), 35 *m* (c), and 45 *m* (d), located by the red dotted lines on the common shot gather in Figure 3.9. The compared traces are: the true data (black), the initial data (grey), the final data after conventional low-to-high (LH) frequency FWI with an (ω, k) domain misfit function (blue dashed), and the final data after layer stripping (LS) FWI with a (t, x) domain misfit function (red dashed).

true data. All three FWI results are able to correct for this. However, as can be seen in the final differences (Figures 3.12d, 3.12e and 3.12f) the results with layer stripping FWI are better able to reconstruct the higher modes and the low frequency part of the fundamental mode. It is interesting to note that even though the very low frequencies are not inverted in the layer stripping FWI, implementing frequency bands only down to 20 *Hz*, they have been corrected by the inversion.

When comparing the global misfit in the (ω, k) domain for the complete final datasets obtained from FWI, the smallest value is obtained for the result from layer stripping FWI using the (ω, k) domain misfit function, minimizing the misfit to less than 1 percent of the initial value. The result from layer stripping FWI using the (t, x) domain misfit function reduces the misfit to 1.3 percent of the initial value, and the result from conventional low-to-high frequency FWI reduces the misfit to 3.7 percent of the initial value.

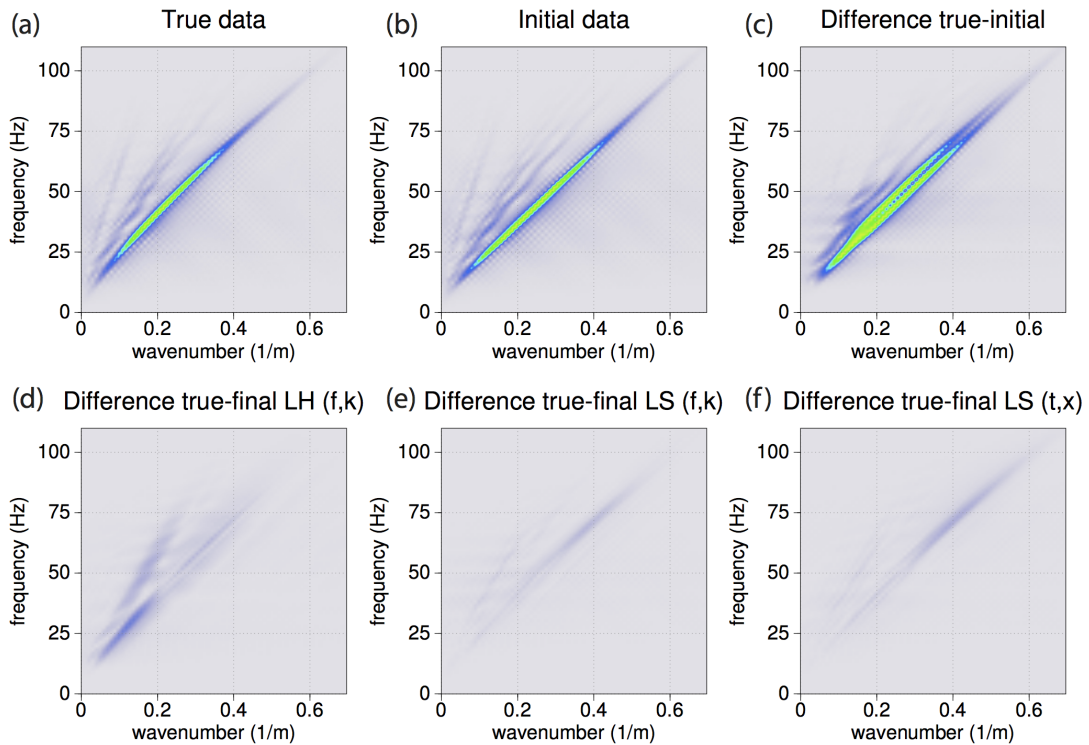


Figure 3.12: Example of the (ω, k) spectrum for a single shot, for the true data (a); initial data (b); difference between initial and true data (c); difference between final data after conventional low-to-high frequency FWI with an (ω, k) misfit function and the true data (d); difference between final data after layer stripping FWI with an (ω, k) misfit function and the true data (e); difference between final data after layer stripping FWI with an (t, x) misfit function and the true data (f). A small gain is applied to the true data (a) only for visualization purposes, all other figures (b) to (f) have the same color scale as (a).

The same conclusions are reached when analyzing data in the frequency-velocity (ω, c) domain, which is conventionally used for dispersion curve inversion. The final differences for the results from layer stripping FWI (Figures 3.13e and 3.13f) are much smaller than when the conventional low-to-high frequency FWI is implemented, suggesting that the velocity model was better recovered. It is also interesting to note that the final difference for both layer stripping FWI results is more important for high frequencies. This suggests that details are still missing in the final shear velocity model, or that perhaps the parameters chosen for the first few frequency steps in the layer stripping approach may not have been optimal.

The analysis of the data in the different domains supports the success of the layer stripping approach for FWI. A low frequency band is the last to be inverted with layer

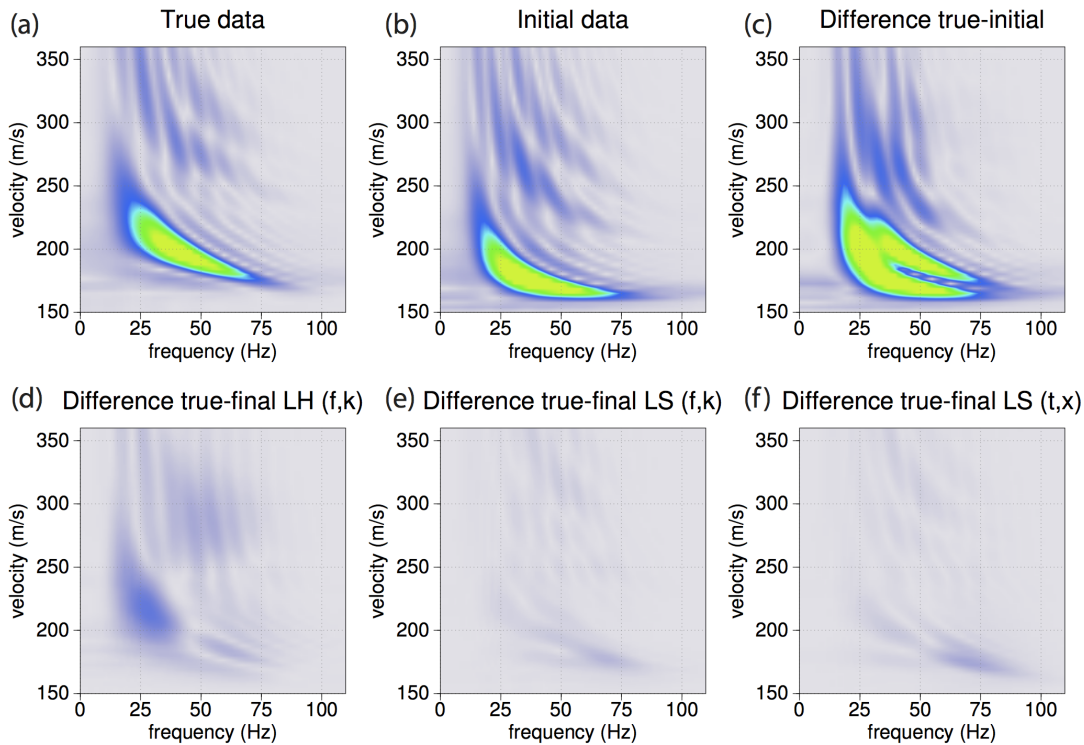


Figure 3.13: Example of the dispersion curves for a single shot, for the true data (a); initial data (b); difference between initial and true data (c); difference between final data after conventional low-to-high frequency FWI with an (ω, k) misfit function and the true data (d); difference between final data after layer stripping FWI with an (ω, k) domain misfit function and the true data (e); difference between final data after layer stripping FWI with an (t, x) domain misfit function and the true data (f). A small gain is applied to the true data (a) for visualization purposes, all other figures (b) to (f) have the same color scale.

stripping FWI, allowing to fully correct the model at depth as the top layers have already been dealt with. Instead, for the conventional low-to-high frequency approach, the low frequencies are the first to be inverted, and although they avoid cycle skipping issues, their full potential is limited for recovery at depth because the top layers of the shear velocity model are not yet corrected.

To perform an objective quality control on the inversion results, the L2 norm of the difference in the (t, x) domain between the whole, unfiltered, true data, and the final data obtained after each frequency band step during FWI, is plotted in Figure 3.14. The plot enables an analysis of the raw misfit evolution during inversion. One can observe that for FWI using the conventional low-to-high frequency approach, and an (ω, k) domain misfit function, the raw misfit reaches a local minima during the fourth frequency step.

The continued convergence during the following frequency step is further validation that the (ω, k) domain misfit function is more robust to cycle-skipping than the conventional (t, x) domain misfit function. For both layer stripping FWI tests, the raw misfit decreases regularly for each successive frequency band step, as if finding a path within the global valley of attraction that avoids the locations of local minima. The choice of this path allows the less robust classical (t, x) domain misfit function to be equally successful as the (ω, k) domain misfit function during inversion.

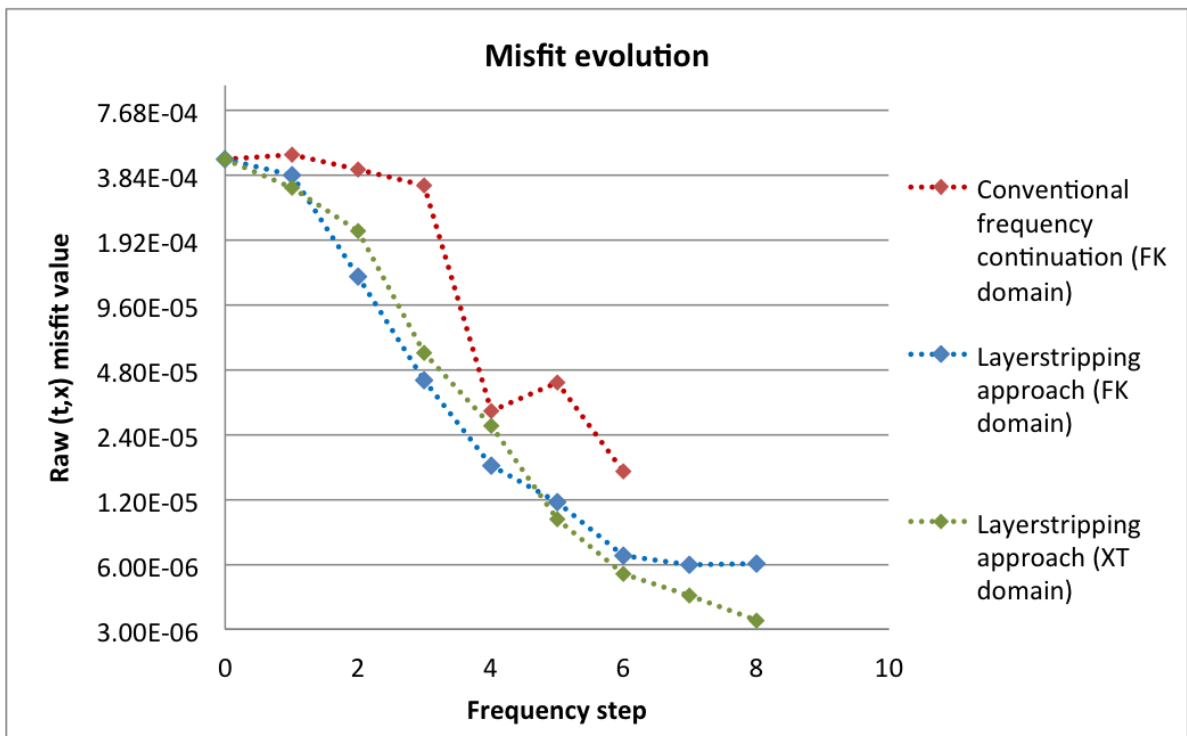


Figure 3.14: The conventional (t, x) domain misfit is used to calculate the difference between the true data, and the data corresponding to the model obtained after each frequency band step, to provide a misfit evolution. The misfit evolution for both the conventional low-to-high frequency strategy and the layer stripping strategy are plotted. This objective comparison shows how the layer stripping approach has a smoother misfit evolution, and gives a greater final reduction of the misfit value. Note that the y-axis is plotted on a log scale.

The final shear velocity model, obtained with the conventional low-to-high frequency approach and a misfit in the (ω, k) domain, reduces the data misfit by less than 5 percent of the initial value. A better performance is achieved when the layer stripping approach is applied, and using a misfit in either the (ω, k) domain or the (t, x) domain. Both final shear velocity models reduce the misfit by less than 1 percent of the initial value.

When the final model obtained from layerstripping FWI with the (t, x) domain misfit is used as an initial model for FWI using all frequencies together (no filtering), to update the complete model (no depth windowing), using the (t, x) domain misfit, the inversion appears to converge, as the data are no longer cycle-skipped. Yet the small number of iterations, only produce a very minimal update of the shear velocity model.

To conclude, this synthetic example has illustrated that employing a layer stripping approach for FWI with surface waves is more successful than the conventional low-to-high frequency approach. The layer stripping FWI allows to overcome the cycle-skipping issue so that very successful results can be obtained with a conventional misfit function in the (t, x) domain, and there is no apparent need of an alternative, more robust, misfit function for convergence to occur. Furthermore the strategy does not rely on low frequency content in the data for convergence, as opposed to conventional low-to-high frequency FWI.

3.2.4 Parameter selection analysis

The main pitfall for layer stripping FWI, as mentioned by [Bian and Yu \(2011\)](#), is the sensitivity to the choice of parameters. The data which is selected and inverted needs to fit to the correct depth which is updated. The parameter x_{max} , which describes the offset windowing, is found to be less sensitive than the relation between frequency and depth parameters, and is therefore automatically calculated within the algorithm following Equation 3.3. Ultimately, the correct relation between the choice of the frequency band and the depth window, will depend on the velocity of the layer to be recovered in the model, and is therefore not well known. Yet there are several factors that can lead one to make educated inversion parameter choices.

The natural frequency content of the source is used to determine the choice of frequency bands used during layer stripping. Slightly larger frequency windows are selected for the initial high frequency bands, for which the frequency spectrum of the data has low amplitudes (see Figure 3.15). Where the spectrum of the data is of higher amplitude, and where one can expect a better illumination of the velocity model, the frequency bands are more tightly sampled.

Choosing frequency windows that gradually decrease in size for layer stripping FWI, also corresponds to the inverse of the frequency ranges used in the conventional low-to-high frequency approach ([Bunks et al., 1995](#); [Sirgue and Pratt, 2004](#)). This general guideline can be further explained by lower frequencies having larger wavelengths, and

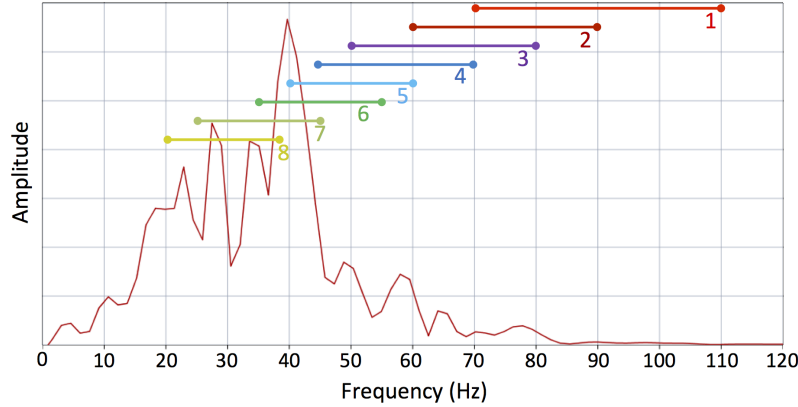


Figure 3.15: Frequency spectrum of the whole true observed data. The energy peaks can be related to the source wavelet with peak frequency of 40 Hz , and also correspond to frequencies which contain information on the main velocity anomalies in the parameter models. The frequency band ranges for layer stripping FWI steps 1-8 (see Figure 3.16) are superimposed on the graph.

therefore a smaller range of frequencies are required to recover a layer of a certain size. The low frequency bands are however kept sufficiently wide to allow the matching of distinct surface wave modes in the (ω, k) domain.

A first good indication for choosing the depth layer to be recovered for a specific frequency band is to analyze the sensitivity kernels, such as illustrated in Figure 3.4. The depth at which the amplitudes of the sensitivity kernel for a given frequency band are no longer significant, will be equivalent to the penetration depth, and should be the maximum depth considered for the depth layer selection. This is how the depth parameters for the tests presented in Figures 3.6 and 3.7 are chosen.

However the penetration depths obtained from the sensitivity kernels are dependent on the initial model provided for FWI. An efficient way to evaluate the chosen layer depths for each frequency band is to compare them to the estimated wavelength of the surface waves such as illustrated in Figure 3.16. The wavelength is estimated as

$$\lambda \simeq \frac{\bar{V}_S}{\bar{f}}, \quad (3.6)$$

using the average frequency \bar{f} of the frequency band and the average shear velocity \bar{V}_S of the depth layer of the initial model for each frequency band step. The average shear velocity value, used to compute the wavelength, is therefore updated at each frequency band step, as the chosen depth layers overlap eachother.

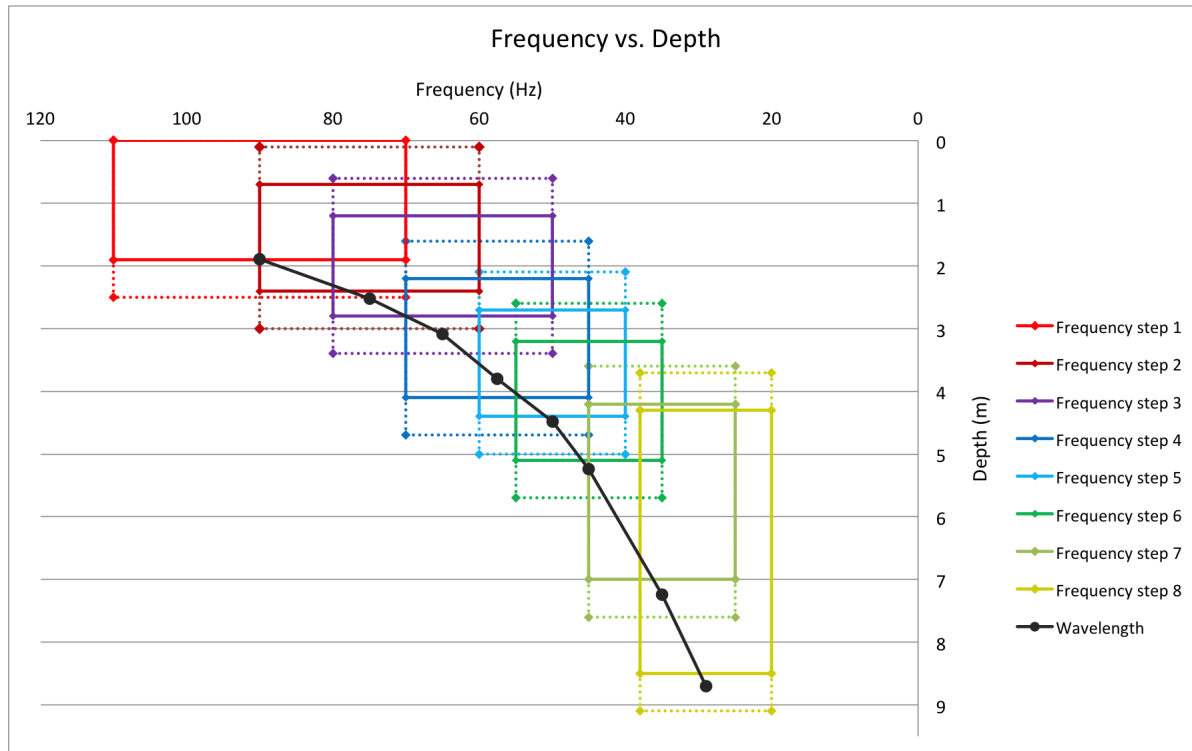


Figure 3.16: Plot of frequency band against depth window parameters used for FWI with a layer stripping approach to produce the results in Figures 3.6 and 3.7. Each color corresponds to a frequency band inversion step. The dotted lines correspond to the extent of the taper outside of the depth value chosen; for simplicity the extent of the inside taper, equivalent in length, is not plotted. The tapering implemented for the frequency windowing is also not illustrated. An estimation of the average wavelength is plotted in black, obtained from average frequency and shear velocity values using Equation 3.6.

The plotted estimated wavelength correlates well with the maximum depth chosen for each layer. This result is in line with literature describing the penetration depth of surface waves to be roughly equivalent to one wavelength (Grant and West, 1965; Foti, 2000; Gedge and Hill, 2012), further details can be found in section 1.1.3. The plot confirms that the penetration depths used for layer stripping FWI, and estimated using the sensitivity kernels, are well chosen. Furthermore the maximum depth estimated from sensitivity kernels can be automatically corrected by the algorithm during inversion, to at least reach the estimated wavelength of the layer, as calculated using Equation 3.6. This correction corrects for the initial dependency on the initial model, and the estimated wavelength is computed using the average shear velocity updated during each new frequency band step.

The special relationship between frequency content and penetration depth is further investigated in Figure 3.17. The initial shear velocity model used for the FWI tests has a linear velocity gradient that can be defined as

$$\bar{V}_S(s) = V_0 + \alpha z, \quad (3.7)$$

where V_0 is the shear velocity at the surface (160 m/s), α the shear velocity gradient (15 s^{-1}) and z the depth. If the penetration depth is considered equivalent to the wavelength ($z = \lambda$), as observed in Figure 3.16, one can obtain a new expression for Equation 3.6 given as

$$\bar{f} = \alpha + \frac{V_0}{z}, \quad (3.8)$$

directly relating the average frequency to use, for each penetration depth considered.

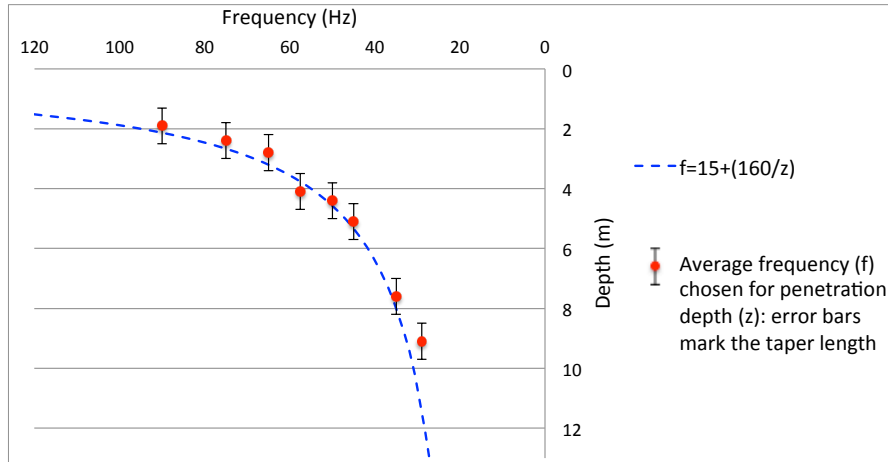


Figure 3.17: The relation given in Equation 3.8, and considering the initial shear velocity model used for FWI, is plotted (blue dashed). The average frequency and relative penetration depth parameter values used for layer stripping FWI are also plotted (red) with depth error bars representing the taper length.

Equation 3.8 is plotted in Figure 3.17 considering the initial shear velocity model used for the FWI tests. For the penetration depth considered during each of the eight frequency band steps of layer stripping FWI (as given in Figure 3.16), the average frequency used is plotted on Figure 3.17 as well. The figure shows that the parameters chosen for layer stripping FWI correspond very well to the theoretical relation given in Equation 3.8. Deviations from the theoretical curve may be linked to the low and high velocity anomalies present in the true model. These observations further validate the choice of parameters, and may be used as a good indicator when the initial model

available is reliable.

To reduce the sensitivity to the minimum and maximum layer depths chosen a tapering is applied to the gradient. The taper implemented is a Hanning function given as $H(i) = \left(0.5\left(1 - \cos\left(\frac{\pi i}{N-1}\right)\right)\right)^2$, where N is the length of the taper. FWI tests with depth window tapering are shown to produce better results compared to tests with no depth tapering (see Figure 3.18). Furthermore large values of N are found to be the most successful, though the taper length should evidently not exceed the initial width of the depth window layer.

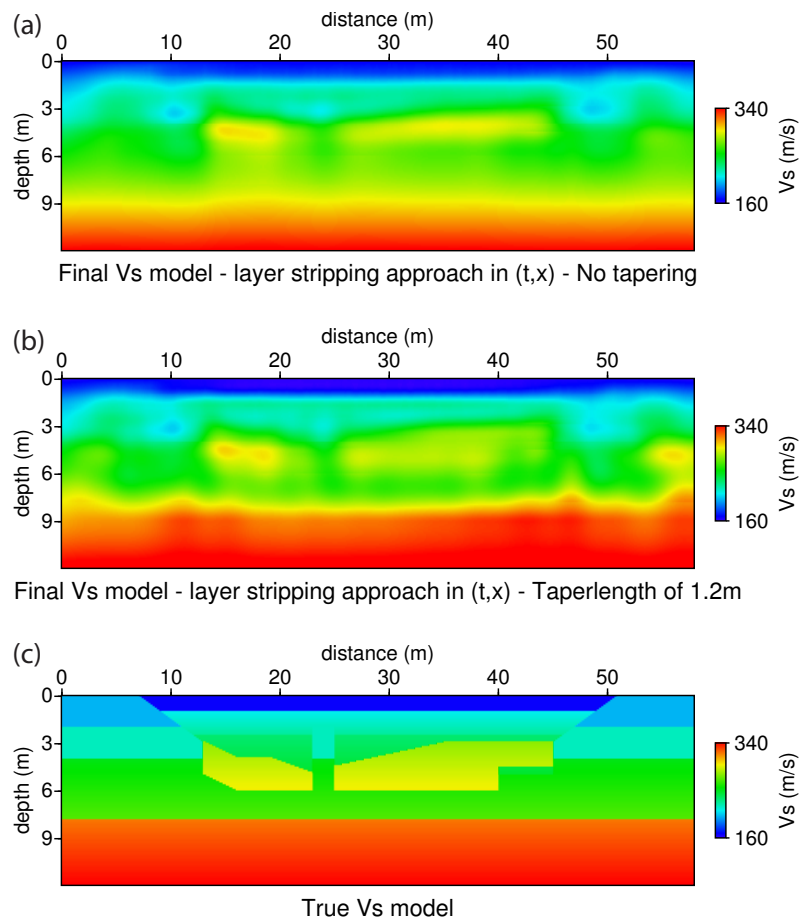


Figure 3.18: Final shear velocity models using the layer stripping approach, and a misfit in the (t, x) domain: when no depth window tapering is applied (a), and when a depth window tapering of 1.2 m is applied (b). The true shear velocity model is shown for reference (c). The tapering reduces the sensitivity to the chosen depth parameters, and improves the inversion result. Similar results are obtained when the misfit in the (ω, k) domain is used.

Figure 3.18 illustrates the result with layer stripping FWI and the (t, x) domain

misfit, without depth tapering (a) and with a depth tapering of 1.2 m (b). All other FWI parameters are kept constant throughout these two tests. A wide depth tapering is shown to improve the quality of the inverted shear velocity model.

For an objective quality control, the raw (t, x) domain difference between the true data and the data corresponding to the final model was calculated. When no tapering is implemented, the raw misfit value of 9.499×10^{-6} remains larger than the value obtained with tapering, although still an order of magnitude smaller than the value that is obtained when a conventional low-to-high frequency strategy is applied (see Figure 3.14).

To conclude, the plot in Figure 3.16 illustrates two general guidelines for parameter choices. The first is that depth windows should gradually increase in size when the frequency content is lowered. This can be explained physically since lower frequencies have larger wavelengths, and will therefore update a larger area. The second is that frequency windows should gradually decrease in size, as previously discussed in this section. Sensitivity kernels as well as wavelength estimations can guide the selection of the layer depth parameter for each frequency band, and the sensitivity to the parameters chosen is reduced when a depth window tapering is implemented. As a quality control, the final data difference and the misfit of the final whole dataset in the (t, x) domain can be used to determine the successfulness of the inversion.

3.2.5 Parameter sensitivity and misfit function selection

In certain cases, frequency and depth parameters may be inadequately chosen during layer stripping FWI. The sensitivity of the FWI to the choice of parameters is therefore investigated by testing layer stripping FWI with linearly increasing parameter values. Although a general guideline of increasing depth window size, and decreasing frequency band size for each successive step is maintained, the chosen parameters are not directly related to the sensitivity kernels or an estimated wavelength. All other FWI parameters are kept constant to the previous layer stripping test, while both misfits in the (ω, k) domain and the (t, x) domain are tested. Again 8 frequency band steps are successively inverted for, the frequency and depth parameters for each step are pictured in Figure 3.19. This regular windowing allows to test the sensitivity to parameters which may not be well chosen.

The final shear velocity models obtained from layer stripping FWI using the misfit function in the (ω, k) domain and the conventional misfit in the (t, x) domain are shown in Figures 3.20a and 3.20b respectively. The inadequate choice of frequency and depth

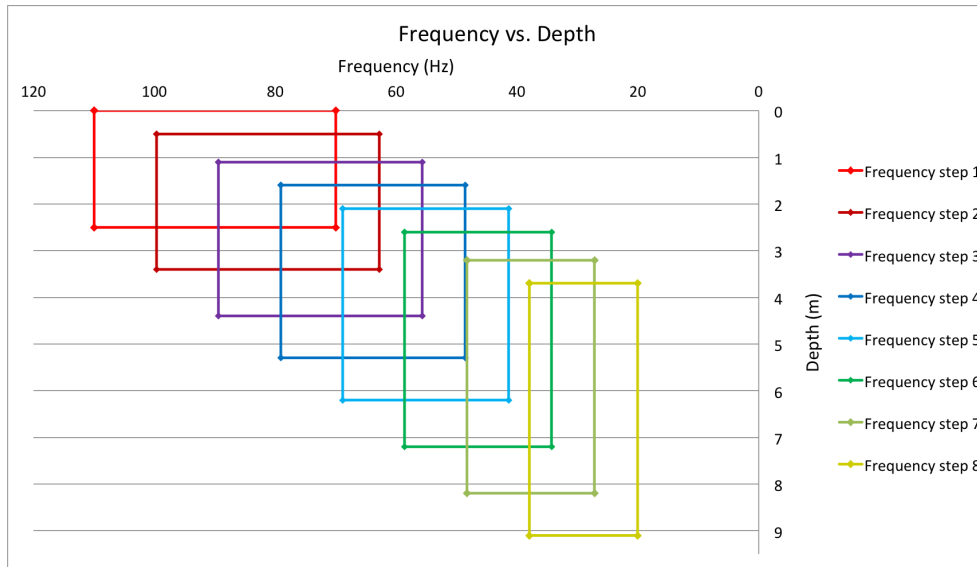


Figure 3.19: Plot of frequency band against depth window parameters used for FWI with a layer stripping approach to produce the results in Figure 3.20. The frequency and depth windows increase linearly during each successive step, corresponding to a new color. Note that for simplicity the depth taper is included within the depth window value.

window parameters, results in a reduced quality of the final model obtained with the conventional (t, x) domain misfit. However the resulting shear velocity model is clearly less successful when the (ω, k) domain misfit is used. The inversion has more difficulty in recovering the two high velocity anomalies, and several artifacts are present in the final model. This suggests that the (ω, k) domain misfit function has a higher sensitivity to the layer stripping parameters chosen, and appears less robust when the layer stripping approach is implemented.

When the conventional (t, x) domain misfit is used, the resulting shear velocity model remains smooth and no significant artifacts are introduced. Furthermore the FWI is still able to reduce the (t, x) domain misfit for the complete final dataset down to 3.993×10^{-6} , equivalent to less than 1 percent of the initial value. This raw misfit value is of the same order of values obtained by layer stripping FWI with adequately chosen parameters (see Figure 3.14). This suggests that when a (t, x) domain misfit is implemented in the layer stripping approach, the sensitivity to the frequency and depth parameters is not significantly high to prevent convergence.

The instability when implementing the misfit function in the (ω, k) domain, is due to cycle-skipping. The layer stripping approach is not well adapted to data minimization

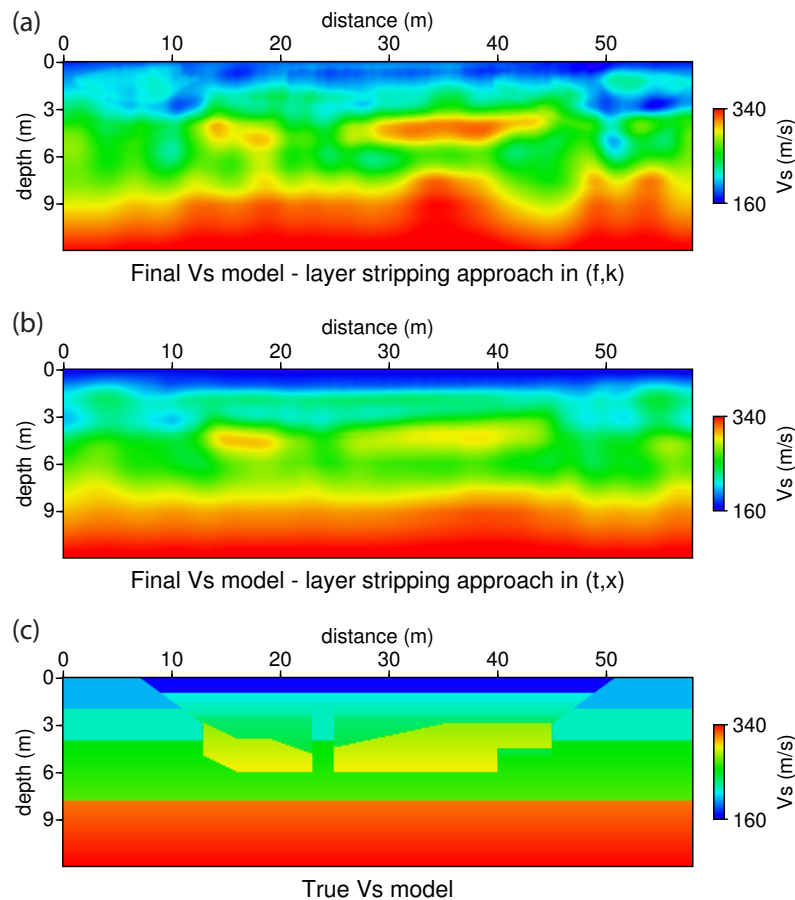


Figure 3.20: Final shear velocity models using the layer stripping approach with a misfit function in the (ω, k) domain (a) and a misfit function in the (t, x) domain (b). The true shear velocity model is shown for reference (c). Frequency and depth windowing parameters are chosen to linearly increase during each inversion step.

in the (ω, k) domain. The selection of high frequencies at the beginning of the inversion, and the whitening of the frequency spectrum (see section 3.2.2 for more details), leads to complex higher modes having a significant importance in the calculation of the misfit.

Figure 3.21 illustrates the true, initial and final data traces for data at the moment that the misfit is calculated, for both the (ω, k) and the (t, x) domain misfit functions. The data shown are filtered in frequency and windowed in offset, according to the first frequency band step implemented for layer stripping FWI. By analyzing the traces in the (ω, k) domain (see Figure 3.21c), one can observe that cycle skipping of the higher modes occurs, which will cause local minima during the inversion. The cycle-skipping occurs due to the high frequency window and the whitening of the frequency spectrum required for the layer stripping approach.

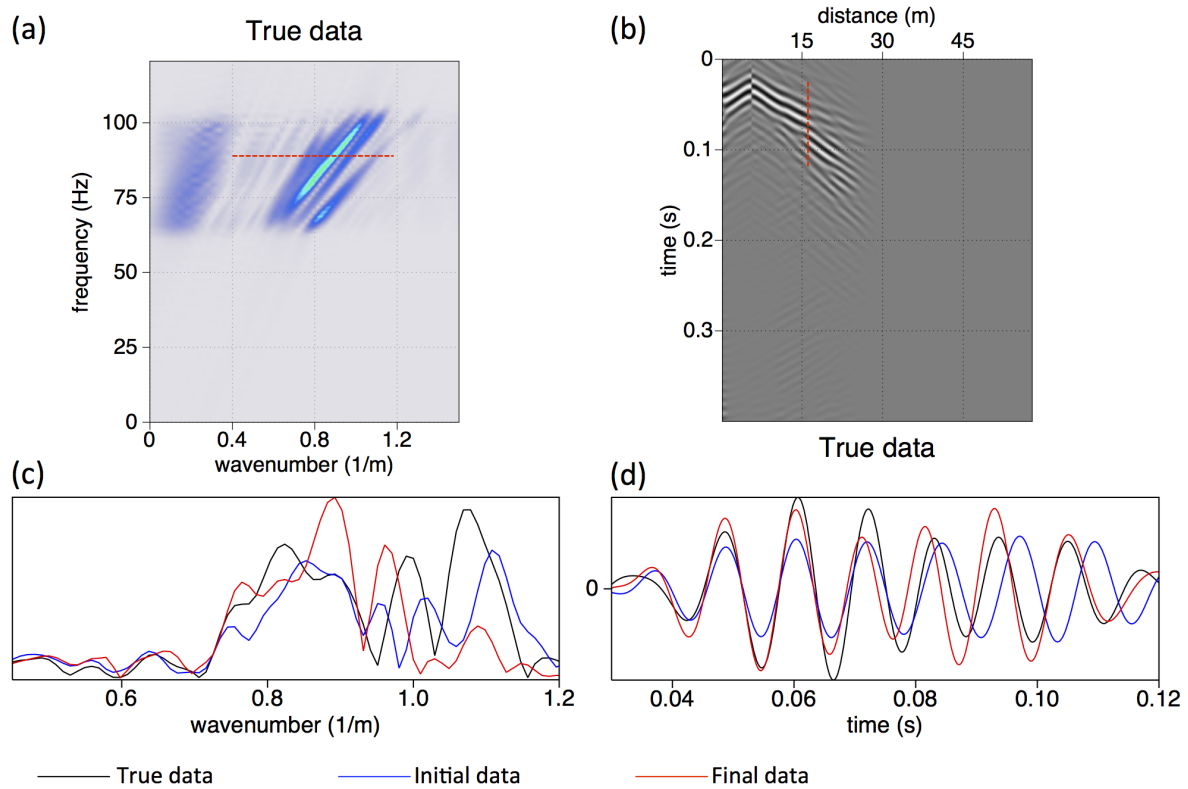


Figure 3.21: Comparison of the true data, after offset windowing and frequency filtering according to the first frequency band step, at the moment when the misfit is computed, for the misfit function in the (ω, k) domain (a) and the misfit function in the (t, x) domain (b). An example trace, is plotted for the true data (black), the initial data (blue) and the final data after the first frequency band step has been inverted (red), for the (ω, k) domain misfit function (c) and the (t, x) domain misfit function (d). The chosen traces are marked in red on (a) and (b) respectively. One can observe that cycle skipping occurs in (c) in the (ω, k) domain but not in (d) in the (t, x) domain.

In addition, the selection of smaller windows for the low frequencies, required for the layer stripping approach, may also make the (ω, k) domain unsuitable, as the robustness of this domain results from being able to match the fundamental mode over a range of frequencies.

By comparison, the layer stripping approach reduces cycle-skipping in the (t, x) domain. By analyzing the traces in Figure 3.21d, one can observe that no cycle skipping occurs. The initially strong windowing in offset removes cycle-skipped data at far offsets from the misfit calculation. Furthermore, the high frequencies which are initially selected, will also be those which will be the most attenuated, and will therefore rely less on the initial model, which may also reduce cycle skipping issues.

3.2.6 Quality control

A final quality control is performed on the successful result obtained in section 3.2.3, using layer stripping FWI and the (t, x) domain misfit function (see Figure 3.7). To determine if the cycle skipping observed in Figure 3.21 has been overcome, the misfit function in the (ω, k) domain is used to compare the data corresponding to the final model obtained in Figure 3.7, with the true data, filtered for the high frequency band of 70 – 100 Hz.

Figure 3.22 shows the results of the quality control exercise. The difference between true and final data in the (ω, k) domain is shown to be very small in Figure 3.22c, even for the high frequency band considered, suggesting that the data features observable in the (ω, k) domain have been recovered. Four example traces are compared in more detail in Figures 3.22d-3.22g. One can observe that the phases and amplitudes of the final data match the true data relatively well. The cycle-skipping that is present for the initial data has been at least partially overcome for the final data, validating the success of the layer stripping method.

When a sequential inversion using the (ω, k) domain misfit function is performed, and using the final model from Figure 3.7 as an initial model, only a small number of iterations occur and the shear velocity model is not significantly modified. The inversion result remains close to the initial model, and does not converge further, suggesting a strong smoothing of the initial shear velocity model may be required to restart convergence for inversion.

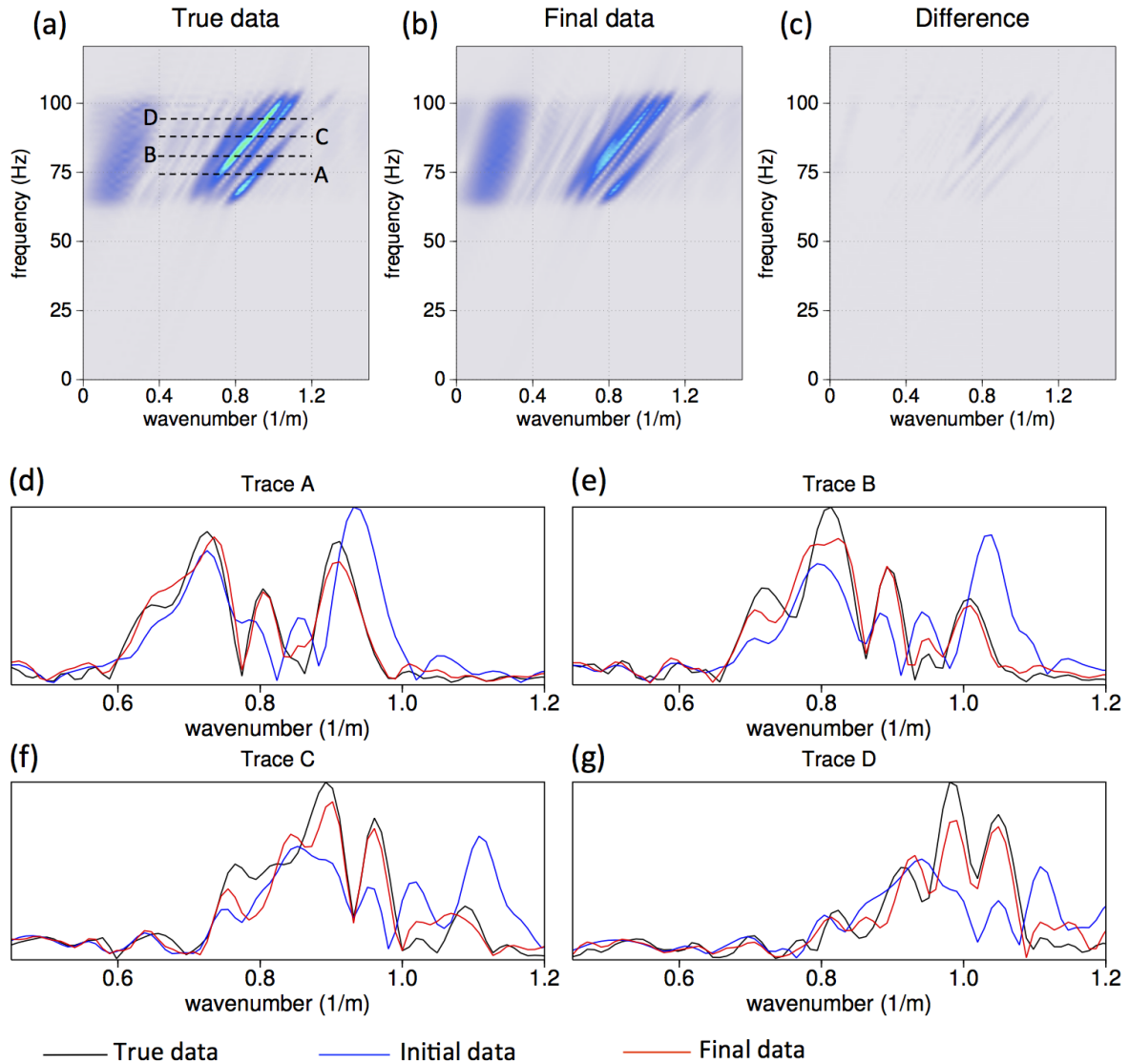


Figure 3.22: Comparison of the true data (a), the final data from layer stripping FWI using the (t, x) domain misfit function (b), and the difference (c), after offset windowing and frequency filtering according to the high frequency band of $70 - 100 \text{ Hz}$, at the moment when the misfit is computed, for the misfit function in the (ω, k) domain. Four example traces are plotted in (d) to (g), for the true data (black), the initial data (blue) and the final data obtained from layer stripping FWI with the (t, x) domain misfit function (red). The position of the selected traces are marked by the black dashed lines on (a).

3.3 Far initial models

The FWI method inherently relies on the availability of an initial model. Only a sufficiently accurate initial model will allow convergence to occur. In this section the layer stripping FWI workflow is tested using initial models which are far from the true model, to investigate the sensitivity to the initial model when implementing this approach.

3.3.1 Far initial shear velocity model

For all previous FWI tests presented in this Chapter, the initial shear velocity model used is relatively close to the true model. Being a vertical velocity gradient (see Figure 3.3a), the main updates by FWI consist in lateral variations and imaging of the velocity anomalies. In this section an initial shear velocity model with velocities far away from the true velocity values is considered. This far initial model, pictured in Figure 3.23a, contains a small linear gradient, but does not have correct velocity values for the top layers or the layer at depth, appearing almost homogeneous.

Layer stripping FWI tests, implementing either the conventional (t, x) domain misfit function, or the more robust (ω, k) domain misfit function are investigated. Conventional low-to-high frequency FWI with the (ω, k) domain misfit function is also tested. Again only the shear velocity is inverted for while the P-wave velocity and density are considered true and known. The same acquisition as in the previous section is simulated, and the same frequency band and depth window parameters as illustrated in Figure 3.16 are applied for the layer stripping approach.

As can be observed from the data difference between the true and the initial data (see Figures 3.24a - 3.24c), cycle-skipping is strongly present. This is especially the case for far offsets (see Figure 3.25b), but even at close offsets, arrivals present in the true data are non existent in the initial data (see Figure 3.25a).

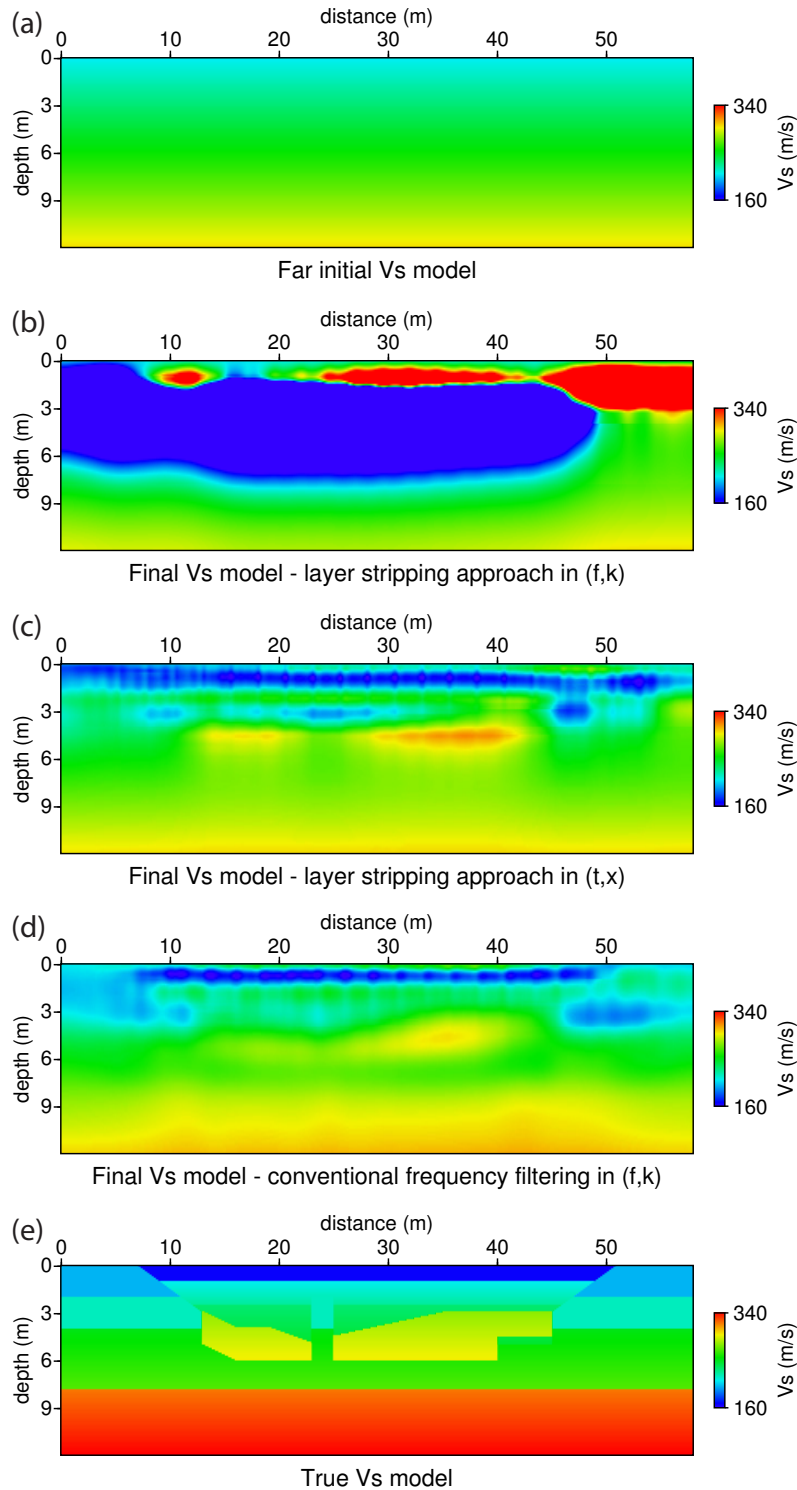


Figure 3.23: Far initial shear velocity model (a); final shear velocity model obtained with layer stripping FWI and the (ω, k) domain misfit function (b); final shear velocity model obtained with layer stripping FWI and the (t, x) domain misfit function (c); final shear velocity model obtained with conventional low-to-high frequency FWI and the (ω, k) domain misfit function (d); and true shear velocity model (e).

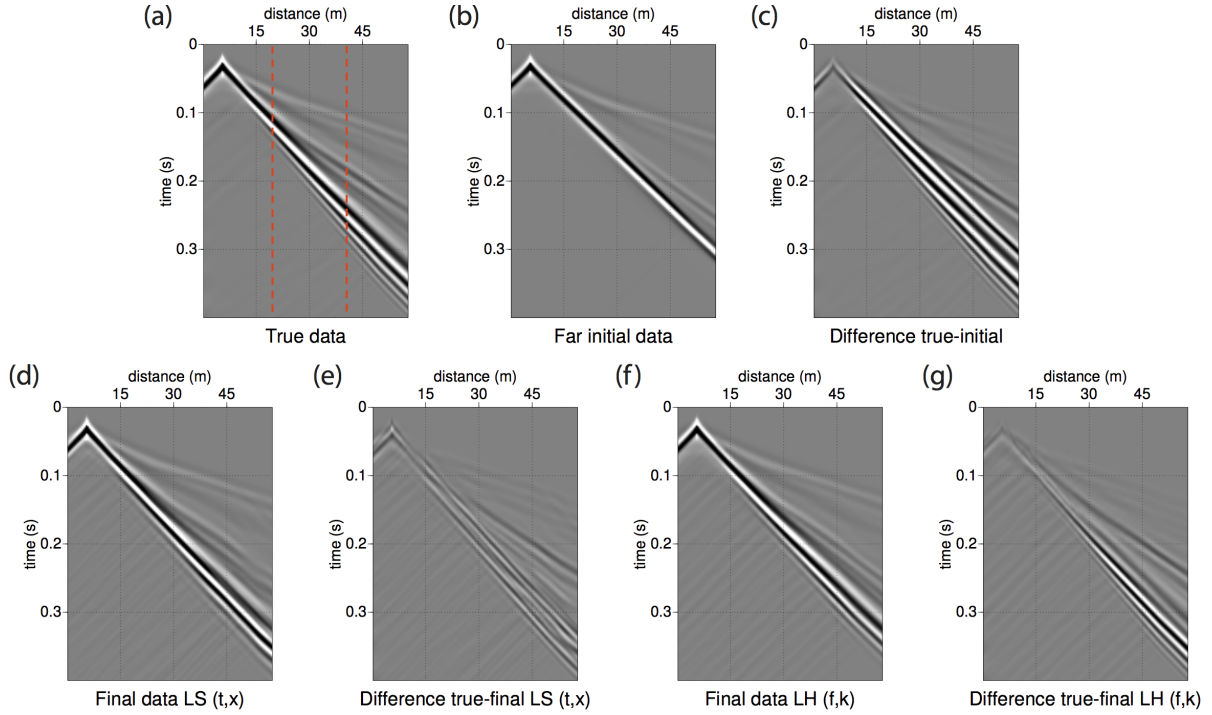


Figure 3.24: Example of a common shot gather for the true data (a); the far initial data (b) and the difference to the true data (c); the final data after a layer stripping FWI with a (t, x) domain misfit function (d) and the difference to the true data (e); the final data after conventional low-to-high frequency FWI with a (ω, k) domain misfit function (f) and the difference to the true data (g). A small gain is applied to the true data (a) for visualization purposes, all other figures (b) to (f) have the same color scale.

The final shear velocity models obtained from layer stripping FWI with the (ω, k) domain and the (t, x) domain misfit functions are shown in Figures 3.23b and 3.23c respectively. When the (ω, k) domain misfit function is implemented, the inversion explodes and does not converge. The high frequency windowing and spectral whitening applied in the layer stripping approach does not seem adapted to comparing data in this domain (see Figure 3.21). Instead, inversion with the conventional (t, x) domain misfit function converges to an interesting result. The top low velocity layer missing in the initial model is recovered, as well as an indication of the two high velocity anomalies in the middle of the model. The high velocity layer at depth is however not recovered.

For conventional low-to-high (LH) frequency FWI with the (ω, k) domain misfit function, the resulting shear velocity model appears of very low resolution, but recovers the top low velocity layer as well as the two high velocity anomalies to some extent, and even somewhat detects the deep high velocity layer. The model obtained appears similar, if

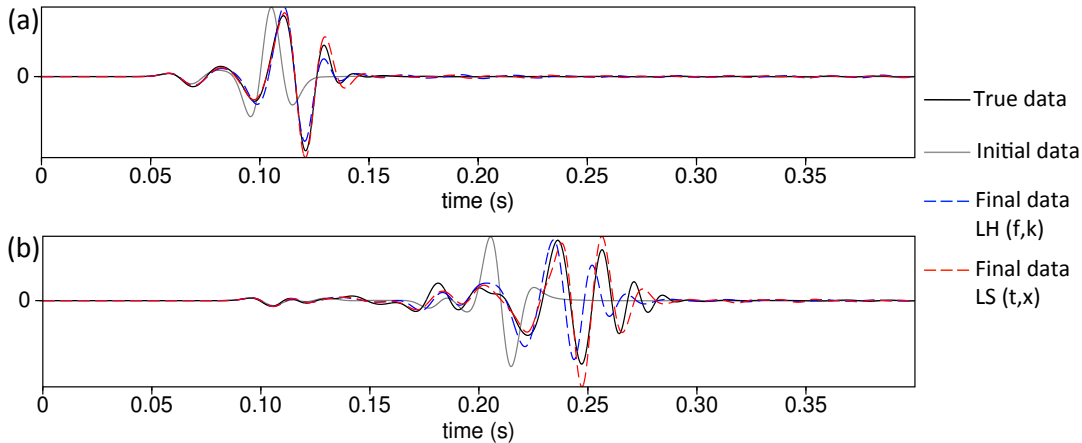


Figure 3.25: Comparison of a single trace at short offset (a) and at far offset (b), located by the red dotted lines on the common shot gather in previous Figure 3.24. The traces compared are: the true data (black), the initial data (grey), the final data after conventional low-to-high (LH) frequency FWI with an (ω, k) domain misfit function (blue dashed), and the final data after layer stripping (LS) FWI with a (t, x) domain misfit function (red dashed).

not slightly better to the one obtained with layer stripping FWI and the (t, x) domain misfit function.

When comparing individual data traces, as illustrated in Figure 3.25, it is not clear which method produces the better result. The final data traces generally follow the traces of the true data, both at short and far offsets, but not perfectly. Both methods also greatly improve the data difference as shown in Figures 3.24e and 3.24g, though a considerable data residue remains for both cases. For conventional low-to-high frequency FWI the data residue can be observed to be especially significant at far offsets. For layerstripping FWI the data residue is more spread out, but the raw misfit of the complete final data is slightly larger than for conventional low-to-high frequency FWI. For both methods, the inversion is able to reduce the raw misfit for the complete final dataset to less than 2 percent of the initial misfit value.

These results suggest that when starting from a far initial model, layer stripping FWI will still allow the velocity model to be updated and for inversion to converge in the right direction. However the layer stripping method no longer has many advantages compared to the conventional low-to-high frequency FWI when using the robust (ω, k) domain misfit function. For both methods the quality of the velocity model recovered is reduced. It is important to note however that the layer stripping FWI does not rely on

the presence of low frequencies in the data, while the conventional low-to-high frequency FWI requires these low frequencies to converge, when starting from a far initial model. On the other hand, if low frequency content is available for the inversion, conventional low-to-high frequency FWI may possibly provide a better initial model for layer stripping FWI, when the starting model is very far from the true one.

3.3.2 Far initial P-wave velocity model

For real data cases, when considering FWI with surface waves to recover the near surface, the initial shear velocity model can often be obtained from the inversion of dispersion curves. Dispersion curve inversion can produce $1D$ shear velocity profiles (see section 1.2). Although they provide only limited information on lateral variations, they can be relatively accurate. For the same near surface region, it is more difficult to obtain a good P-wave initial model, since first arrival tomography can only provide a very low resolution result. In this section, the case of a far initial P-wave velocity model is therefore investigated.

All previous tests in this Chapter consider a correct P-wave and density model and only invert for the shear velocity model. In this section, an initial P-wave velocity model which is far from the true model will be considered and inverted for simultaneously with the shear velocity. This far initial model has a small linear velocity gradient, but appears almost homogeneous around 480 m/s (see Figure 3.26b). The initial shear velocity model consists of a linear gradient with correct minimum and maximum velocities, as used in previous tests (see Figure 3.26a). The density is considered true and known during the inversion. The same acquisition parameters are used as in section 3.2.3, and the same frequency and depth windows for layer stripping are chosen as in Figure 3.16.

The final shear wave and P-wave velocity models obtained from inversion are illustrated in Figure 3.27. When layer stripping FWI is implemented the inversion appears insensitive to the P-wave velocity model, resulting in only a minimal update to the model, not visually evident from Figure 3.27d. The shear wave velocity is successfully updated, and the two high velocity anomalies are recovered. This result suggests that layer stripping FWI is mainly driven by the misfit minimization of surface waves, as is expected with this approach, and is therefore more sensitive to the shear velocity model. The layer stripping approach is therefore promoted for cases when the P-wave velocity of the medium is not well known.

For conventional low-to-high frequency FWI, the inversion appears limited by the far

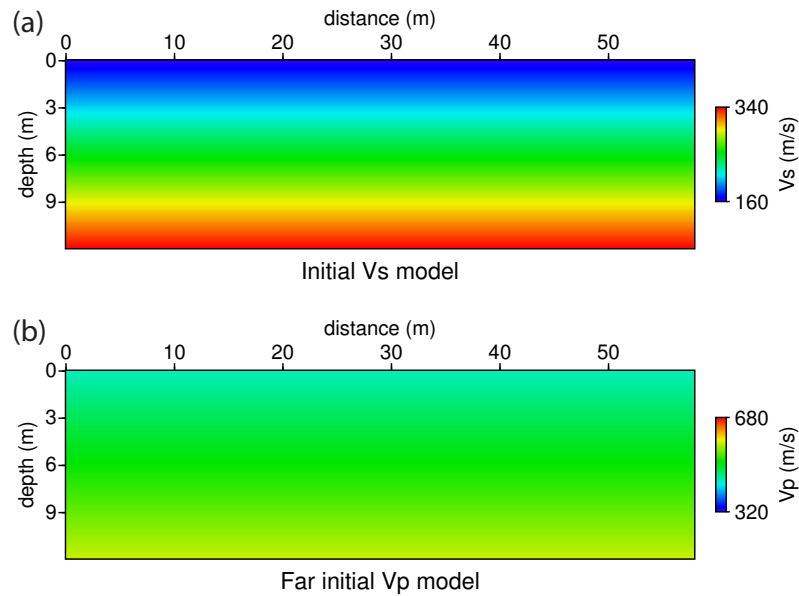


Figure 3.26: Initial shear wave velocity model (a) and initial far P-wave velocity model (b) used for inversion.

initial P-wave velocity model, which is wrongly updated, and causes the inversion to fail to converge. As a result the shear velocity model is only weakly updated, and the two high velocity anomalies are not retrieved.

The same test for conventional low-to-high frequency FWI was repeated, for inversion of the shear velocity only, keeping the far initial P-wave velocity model fixed. However the resulting shear velocity model was not improved, and the inversion did not successfully converge. This result suggests that surface waves are not the only arrival used to drive the inversion, and convergence remains stuck in a local minima due to cycle-skipping occurring for other phases.

Figure 3.28 illustrates an example of the data for these inversion results. The difference between the final data after layer stripping FWI and the true data (Figure 3.28e) appears reduced for the dispersed surface waves, compared to the initial difference (Figure 3.28c), but remains present for the first arrival and other phases. The relative success of the layer stripping method compared to conventional low-to-high frequency FWI can be observed when individual traces are analyzed in Figure 3.29.

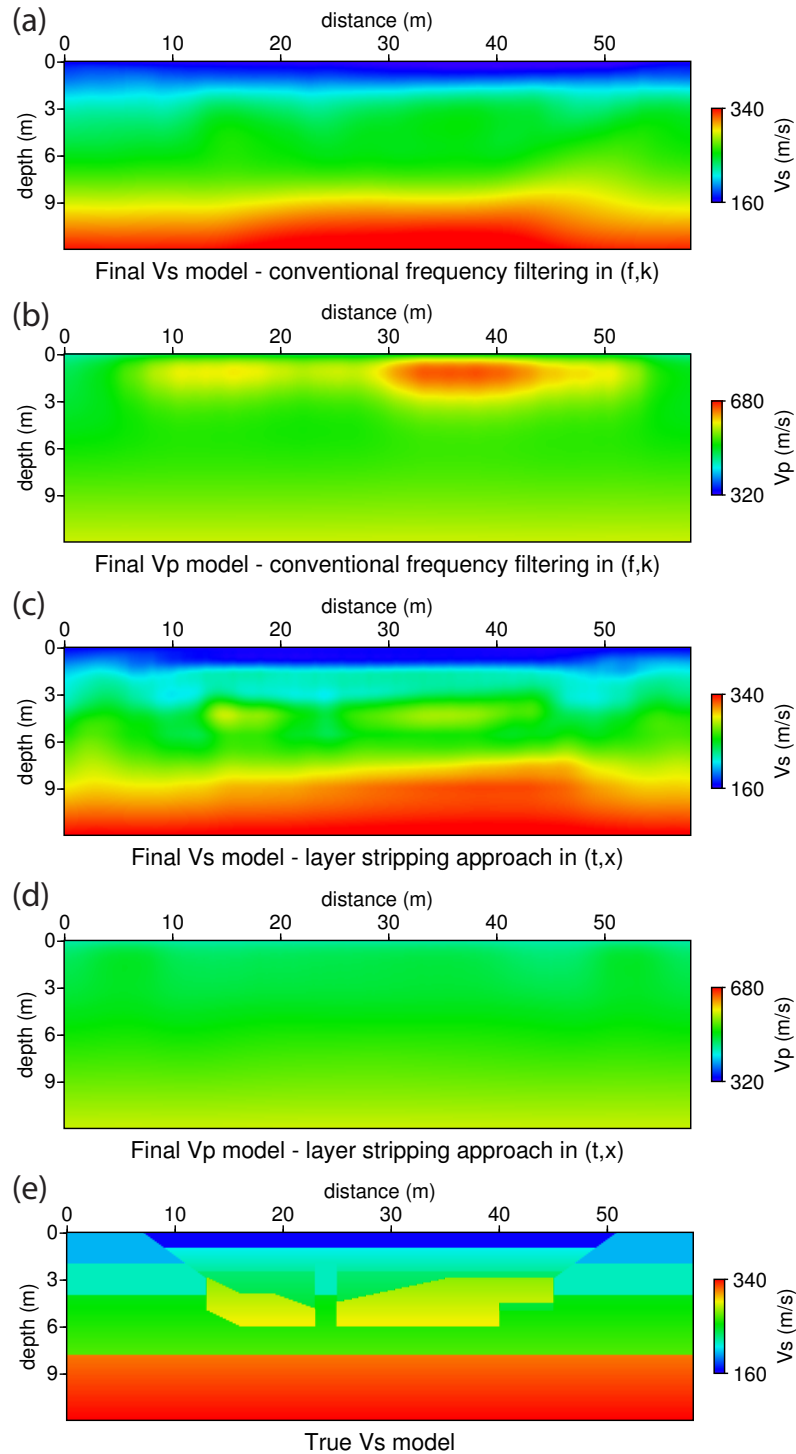


Figure 3.27: Final shear wave and P-wave velocity models obtained for the conventional low-to-high frequency FWI and a (ω, k) domain misfit function (a-b); and for layer stripping FWI with the (t, x) domain misfit function (c-d). The true shear velocity model (e) is shown for comparison.

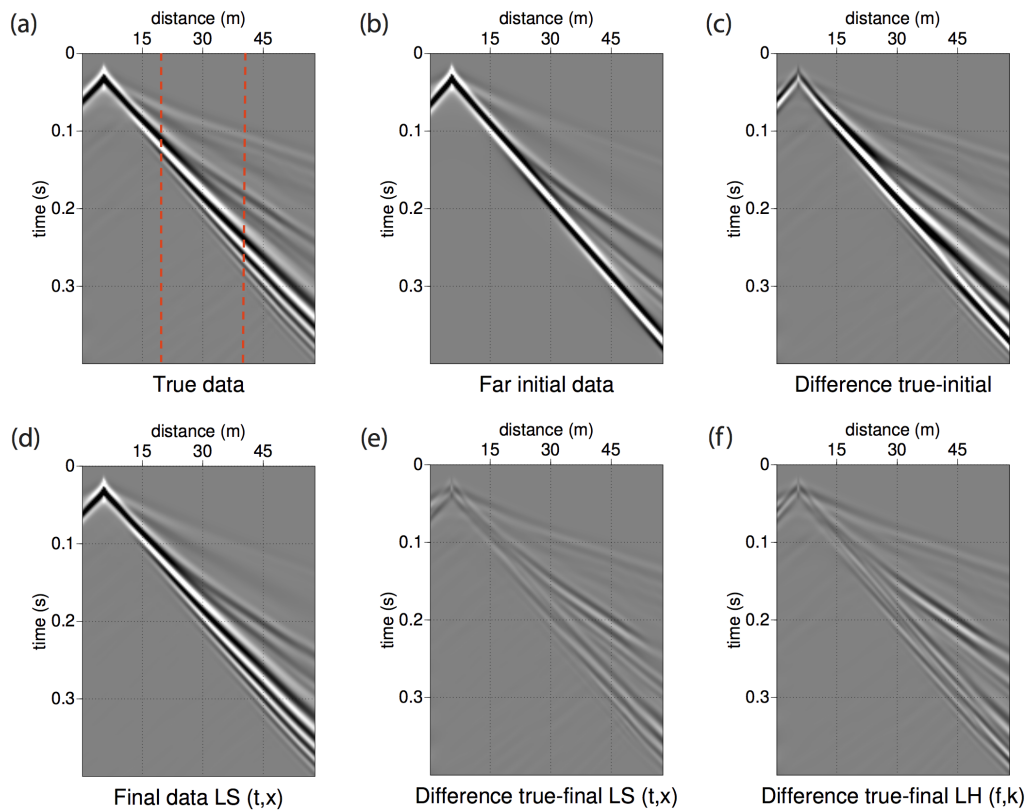


Figure 3.28: Example of a common shot gather for the true data (a); the far initial data (b) and the difference to the true data (c); the final data after a layer stripping FWI with a (t, x) domain misfit function (d) and the difference to the true data (e); and the difference between the true data and the final data after conventional low-to-high frequency FWI with a (ω, k) domain misfit function (f). A small gain is applied to the true data (a) for visualization purposes, all other figures (b) to (f) have the same color scale.

One can observe that at short offsets (see Figure 3.29a), the final data after both the layer stripping FWI and the conventional low-to-high frequency FWI, is able to recover the true surface waves present in the data. The other phases present in the trace are slightly less well recovered by both methods. At far offsets (see Figure 3.29b), the final data from layer stripping FWI recovers the surface wave phases better than the final data after conventional low-to-high frequency FWI. Neither of the FWI methods are however able to fit the first arrival or other guided waves.

This inversion test presents another advantage of the layer stripping approach: the lack of dependency on the P-wave model. Due to the offset ranges, frequency ranges, and gradient depths selected, surface waves are clearly privileged in driving the inversion. As

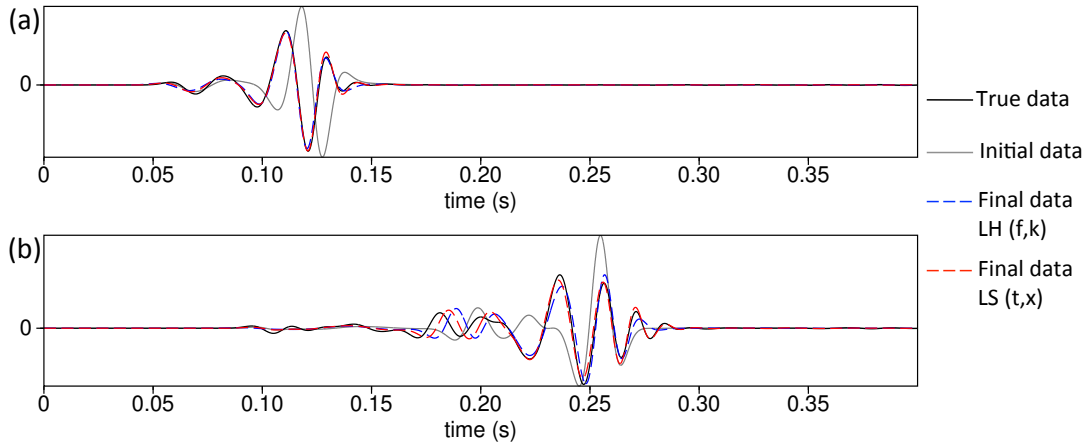


Figure 3.29: Comparison of a single trace at short offset (a) and at far offset (b), located by the red dotted lines on the common shot gather in previous Figure 3.28a. The traces compared are: the true data (black), the initial data (grey), the final data after conventional low-to-high (LH) frequency FWI with an (ω, k) domain misfit function (blue dashed), and the final data after layer stripping (LS) FWI with a (t, x) domain misfit function (red dashed).

surface waves are more sensitive to the shear velocity, this parameter is recovered, while the P-wave velocity is only minimally updated. The inversion result is a clear indication that surface waves are driving the FWI, which is not proven in other applications of FWI with surface waves (Pérez Solano, 2013; Pérez Solano et al., 2014; Yuan et al., 2014).

3.4 Conclusion and perspectives

This chapter investigates a strategy to deal with surface waves for FWI. The inversion of sequential frequency bands of data, such as multiscale approaches or for layer stripping, is shown to improve the inversion. Offset windowing such as used for layer stripping or by Pérez Solano et al. (2014) is also essential to reduce cycle-skipping problems. Finally, the depth variation of the model update, either by gradient preconditioning or layer stripping, is found to be another key ingredient for successful inversion. However it is the combination of all of these features which is required to overcome the difficulties of surface wave FWI, in addition to the alternative misfit functions discussed in chapter 2.

For the layer stripping approach presented in this chapter, high-to-low frequencies are inverted, combined with short to long offsets and top to bottom depth layers. The frequency filtering is the opposite of multiscale approaches that are conventionally used

for FWI, but specifically adapted to the physics of surface waves. The relation between frequency and depth windows, one of the pitfalls for layer stripping approaches, can be determined by sensitivity kernels, as well as average wavelength calculations.

Synthetic tests are performed to compare layer stripping FWI with conventional multiscale FWI, testing different misfit functions. While the (ω, k) domain may lead to cycle-skipping during the initial selection of high frequencies, the (t, x) domain misfit function produces very successful results. These results are better than those obtained with conventional multiscale FWI, and have lower misfit values, validating the layer stripping strategy for FWI.

Layer stripping approaches as well as depth preconditioning are shown to help mitigate difficulties related to forward scattering features of surface waves, where locating velocity anomalies considering long offsets is difficult. Different to [Pérez Solano et al. \(2014\)](#), long offset data is therefore allowed to be inverted.

One of the advantages of layer stripping FWI is that it does not rely on low frequency data, often missing from real data acquisitions, to converge. Tests with far initial models show that layer stripping FWI is able to at least partially overcome cycle skipping, although an initial inversion using the more robust (ω, k) domain misfit function and a multiscale approach may be more favorable when low frequency data is present. Layer stripping FWI is also shown to be insensitive to the P-wave velocity parameter, validating the theory that surface waves are targeted to drive the FWI and update the shear velocity model.

As a perspective one could envision to use a combination of both layer stripping FWI (with the (t, x) domain misfit function) as well as multiscale FWI (with the (ω, k) domain misfit function), in a V-cycle workflow, switching from one strategy to the other, to slowly but consistently build a velocity model. Thus taking advantage of both retrieving the top layers of the medium and improving the overall resolution moving from low-to-high frequencies.

Although successful results have been achieved using synthetic examples, difficulties may be encountered when considering a real data application. For example an accurate enough or kinematically compatible initial model may not be available, or a significant presence of noise may be present in the data, or the acquisition design may not be well adapted. The methodology developed in this chapter is applied on a real data case study in the following Chapter 4, which highlights the problems that may occur when performing FWI with surface waves on real data.

Chapter 4

Real Data Case Study

Contents

4.1 Case study background & acquisition	134
4.1.1 Analysis of common shot gathers	136
4.1.2 Data pre-processing for FWI	141
4.2 Surface wave inversion	143
4.2.1 Initial models for FWI	143
4.2.2 Impact of receiver array	153
4.2.3 Source estimation	155
4.2.4 Forward modelling and cycle-skipping analysis	157
4.2.5 Full Waveform Inversion	163
4.3 Conclusion and evaluation	168
4.4 Perspectives	169
4.4.1 Dealing with topography	169

4.1 Case study background & acquisition

In this chapter, possibilities and challenges for performing Full Waveform Inversion (FWI) on a real dataset including surface waves are investigated. The dataset consists of a single line of 12 *km*, recently acquired on land, to perform 2D imaging of the subsurface. The location of this dataset, provided by Total, will not be detailed for confidentiality reasons. The methodology developed in Chapters 2 and 3 is tested on this real case scenario. This is one of the first applications of FWI with surface waves on a real dataset at the oil and gas exploration scale. The results obtained may provide a better understanding of the near surface with direct implications for exploration studies of deeper targets. Moreover this study will explore the challenges to be solved when considering surface wave inversion of real data.

The real dataset was acquired towards the bottom of a mountainside, a schematic of the location is given in Figure 4.1. Several superimposed alluvial fans originating from the top of the mountain have deposited layers of alluvial debris on top of a sedimentary bedrock, creating a low velocity near surface layer. On the surface, a small amount of vegetation is present, consisting mainly of shrubs.

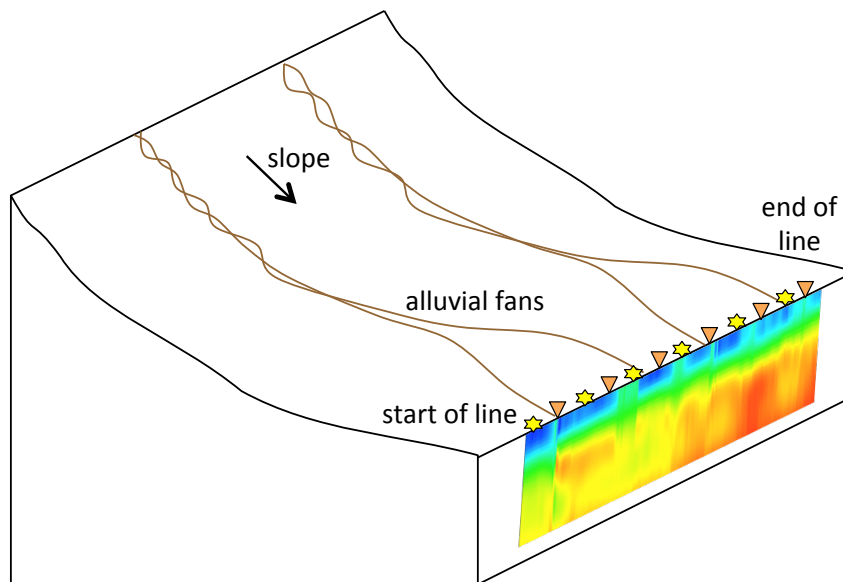


Figure 4.1: Schematic showing the location of the acquired line and subsurface imaged within the environmental setting.

The geology of the bedrock is mainly defined by flat layers of deposited sedimentary rock. Previous exploration studies have shown that subsurface P-wave velocities are in

the range of 2500 m/s to 3500 m/s in the top 200 m of the subsurface, much higher than what is generally considered in civil engineering case studies, as considered in Chapter 3. The characteristic high velocities of the medium suggest that it is very compact and hard.

The line is located perpendicular to the slope direction, and therefore has a relatively flat topography, only interrupted by two canyons creating a topography variation of maximum 30 m over a width of 500 m , containing small streams flowing in the direction of the slope.

The dataset was obtained using two vibroseis energy sources, with a distance between each source station of 20 m , giving a total of 600 shots. For each source position, four sweeps were recorded separately by all receivers and then stacked. The sweep profile has a starting frequency of 4 Hz and a stopping frequency of 90 Hz . A linear sweep rate was applied, with a sweep length of 16 s . The data was recorded by 601 receiver stations spaced 20 m apart, each consisting of an array of 12 inline geophones, spaced 1.66 m apart. The geophones are vertical component and have a 10 Hz natural frequency. A typical amplitude response of the geophones is shown in Figure 4.2, illustrating the flat response between a frequency of 10 Hz and $> 240\text{ Hz}$, while the output is divided by 100 for a frequency of 1 Hz . The data acquired was preprocessed on the field, to create the provided dataset, which has a record length of 6 s with a sample interval of 0.002 s .

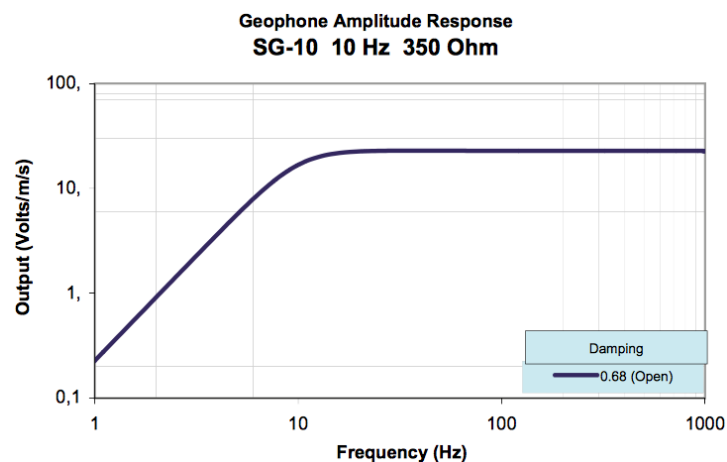


Figure 4.2: Typical amplitude response for a geophone with a 10 Hz natural frequency, as used during the real data acquisition. Figure taken from [Sercel \(2016\)](#).

4.1.1 Analysis of common shot gathers

Two examples of common shot gathers at the beginning and in the middle of the acquired line are shown in Figure 4.3. One can observe the direct wave pointed out by (A), which can be roughly estimated to travel at about 3300 m/s . The refracted wave marked by (B) is possibly created by an interface that is roughly calculated to be below 900 m , and therefore outside of the region that will be considered for the inversion of surface waves. These waves will therefore not be recovered, and may have to be muted. The main surface wave package is marked by (D), and is the slowest travelling waveform of interest. Velocities may be roughly calculated at 1100 m/s and appear to increase for far offsets. The surface wave amplitudes appear to strongly decrease with offset, due to attenuation and possible lateral heterogeneities. Slightly faster waves, marked by (C), may be interpreted as either refracted shear waves, converted waves or guided waves. These could be considered as higher modes for surface waves. Finally the slow signal marked by (E) is assumed to be the airwave and considered as noise.

To better comprehend the different waveforms and where they origin, simple forward modelling using a layered velocity model is performed (see Figure 4.4). Velocities and interface layers of the shear velocity, P-wave velocity and density models are approximated, using information from the initial model building in section 4.2.1, and well information which is only available for the deeper part. A discrete wavenumber summation method (Bouchon and Aki, 1977) is used to model horizontally layered media with a free surface, and simulating 3D elastic wave propagation with a Ricker wavelet source of 40 Hz peak frequency.

As one can observe from Figure 4.4, the surface wave package does not significantly change when depths below 120 m are considered. The surface waves are only impacted by the near surface region, and therefore only propagation within this region is required to model these waves. It is interesting to note that the guided waves or higher modes marked by (C) in Figure 4.3 only appear when a strong velocity contrast is present due to the presence of a high velocity layer below the first 100 m of the shallow low velocity zone (see Figure 4.4b). The reflections present in the observed dataset are shown to origin from very deep P-wave velocity contrasts, as they are retrieved in Figure 4.4c using velocity models up to 3.2 km in depth.

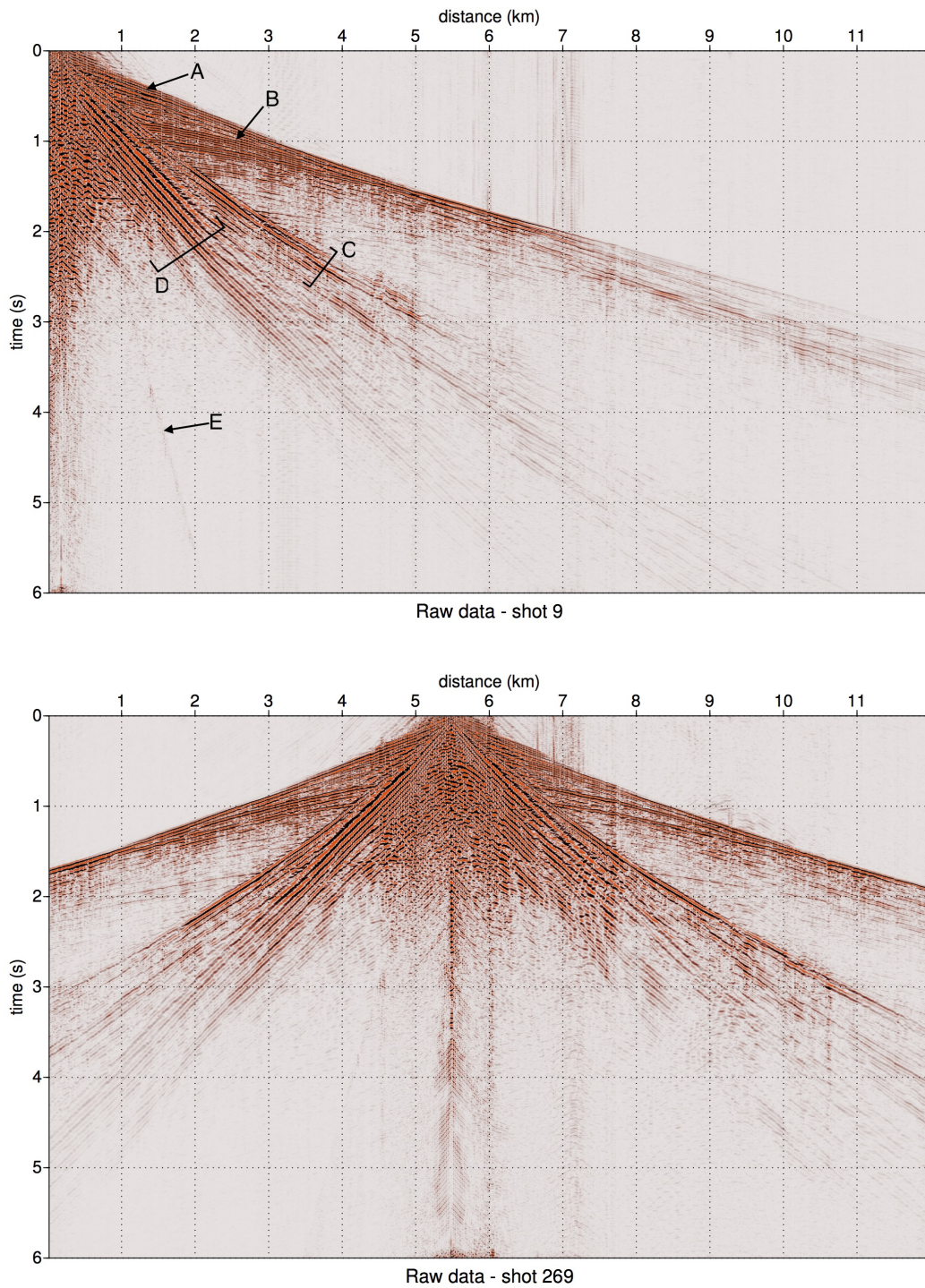


Figure 4.3: Two examples of common shot gathers, located at the beginning and in the middle of the acquired line respectively. Highlighted are the direct wave (A), a refracted wave (B), the surface waves (D), other guided waves (C), and the airwave (E).

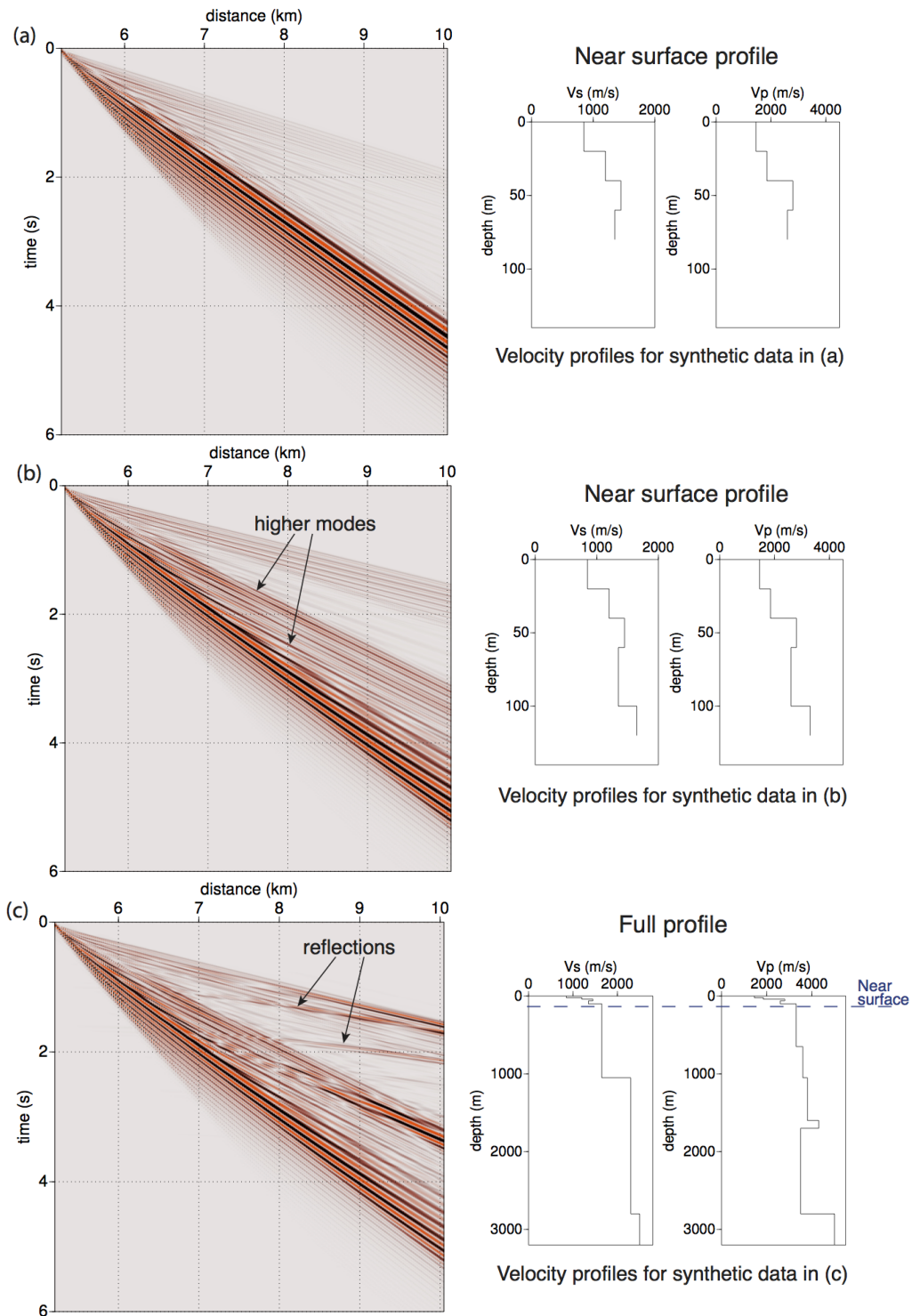


Figure 4.4: Synthetic shot gathers computed with a layered velocity model, considering depths up to 80 m in (a), 120 m in (b) and 3.2 km in (c). The surface wave package does not significantly change when depths below 120 m are considered.

A close up of the surface waves in the observed dataset is shown in Figure 4.5. One can easily track the continuity of the surface wave phases in the wiggle traces. The quality of the surface waves, and the strong signal to noise ratio makes this dataset ideal for an investigation of FWI with surface waves.

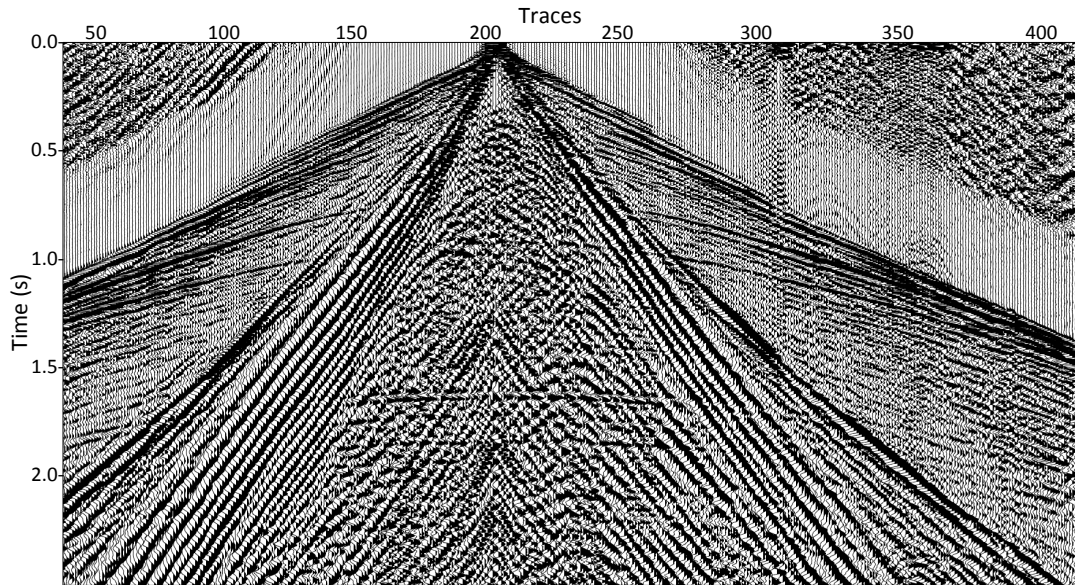


Figure 4.5: Detail of a common shot gather (shot 198) illustrated in wiggle form with an AGC applied.

The spectrum of the whole dataset is visualized in Figure 4.6. Some noise is present at frequencies below 3 Hz . The lowest frequency content of the data occurs at $4 - 5\text{ Hz}$, which is the starting frequency of the vibroseis sweep. Peaks in the spectrum occur at 15 Hz and 18 Hz . Only minimal data is present with frequencies above 50 Hz . When the data is muted to focus on the surface waves, the frequency content is shown to be in the range of 5 Hz to 30 Hz , with significant peaks at frequencies between $15 - 20\text{ Hz}$ as shown in Figure 4.7.

The frequencies considered for this study are lower than the peak frequency of 40 Hz used for the synthetic example presented in Chapter 3. However the size of the subsurface to be imaged, and wavelengths required, are also much larger than those previously considered.

Considering an average surface wave velocity of 1100 m/s and a minimum frequency content of 5 Hz , the penetration depth of surface waves, and therefore the maximum depth one can hope to recover is 220 m (see Equation 3.6). Similarly if the maximum

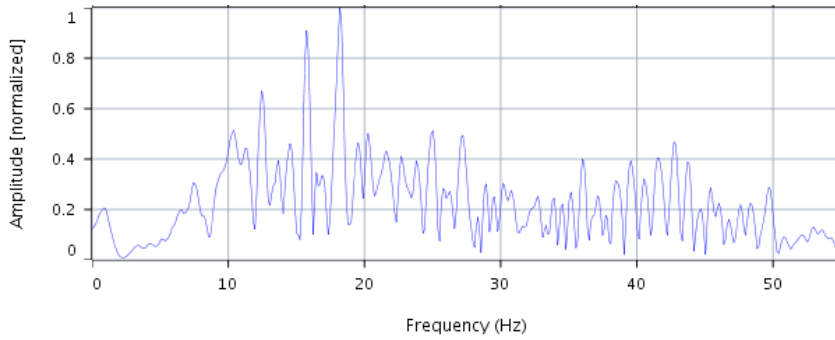


Figure 4.6: Frequency spectrum of the whole dataset: the frequency content of the data starts at 4 – 5 Hz and dies out above 50 Hz .

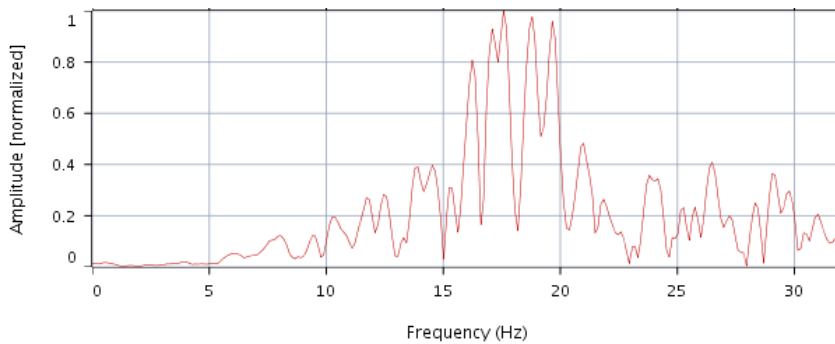


Figure 4.7: Frequency spectrum of a muted version of the data focused on surface waves: the frequency content of the surface waves starts at 5 Hz and dies out above 30 Hz .

frequency content of surface waves considered is of 30 Hz , the resolution achievable is on the order of 35 m .

The common shot gathers in Figure 4.3 are shown in the frequency-wavenumber (ω, k) domain in Figure 4.8. In Figure 4.8a, the fundamental mode is visible between 6 – 18 Hz , but appears relatively weak compared to other signal. As the shot is at the beginning of the line, the mode is only visible on the forward side of the spectrum. In Figure 4.8b, the fundamental mode is clearly visible between frequencies 5 – 25 Hz . As the shot is in the middle of the line, the mode appears on both forward and reverse sides. These appear to overlap relatively well, suggesting that subsurface velocities do not vary significantly on either side of this source position, and that laterally invariant surface wave analysis methods may be used to successfully retrieve a shear velocity model.

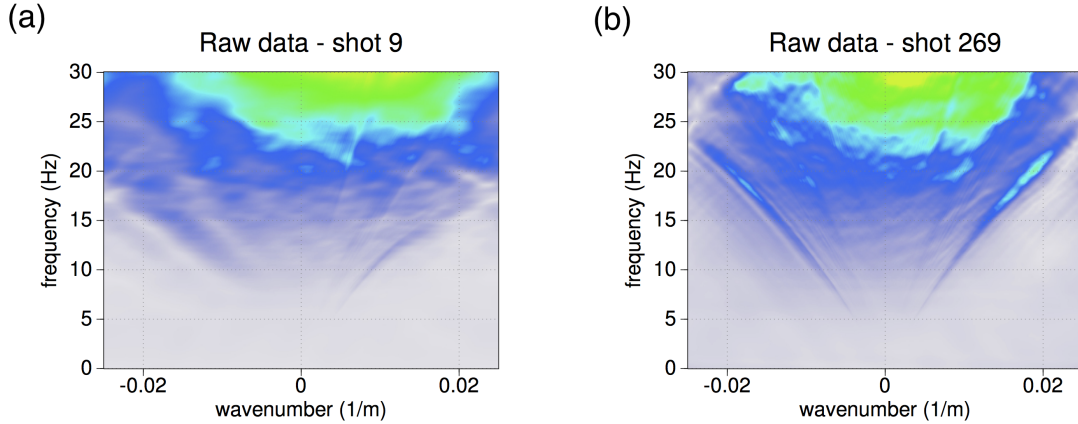


Figure 4.8: Common shot gathers from Figure 4.3, located at the beginning (a) and in the middle (b) of the acquired line viewed in the (ω, k) domain.

4.1.2 Data pre-processing for FWI

Several pre-processing steps were applied to the raw dataset by CGG, to help remove some of the noise present, as illustrated for an example common shot gather in Figure 4.9a. A reverse polarity correction was applied to certain data traces, which had been recorded with a wrong polarity. Certain very noisy traces were muted and interpolated from neighboring traces. Furthermore a de-spiking process was applied to the data. The de-spiking, applied in the frequency-space $(\omega - x)$ domain, is used to zero or rescale spikey spectral gates of the data. An FX-decon filter was applied to optimally extract linear features and suppress random noise (Gulunay, 1986).

After the noise removal, a spatial regularization was applied to the whole dataset to obtain a fixed grid spacing between sources and receivers, as is assumed in the FWI algorithm used for this study. Missing source positions, such as at the location of the canyons, were interpolated using neighboring shots. An example common shot gather of the data before and after pre-processing is illustrated in Figure 4.9. The data were then resampled in time to obtain a sample interval of 0.4 ms to accommodate the Courant-Friedrichs-Lewy (CFL) stability condition of the FWI algorithm used (see Equation 1.52), and allow a spatial discretization step of 3 m for the velocity models.

The real data was obtained using point sources that produce wavefields that have a 3D geometrical spreading in the subsurface media. A 3D forward modelling algorithm could appropriately simulate such point source wave propagation. However 3D elastic FWI is difficult due to the excessive computational requirements. The forward modelling

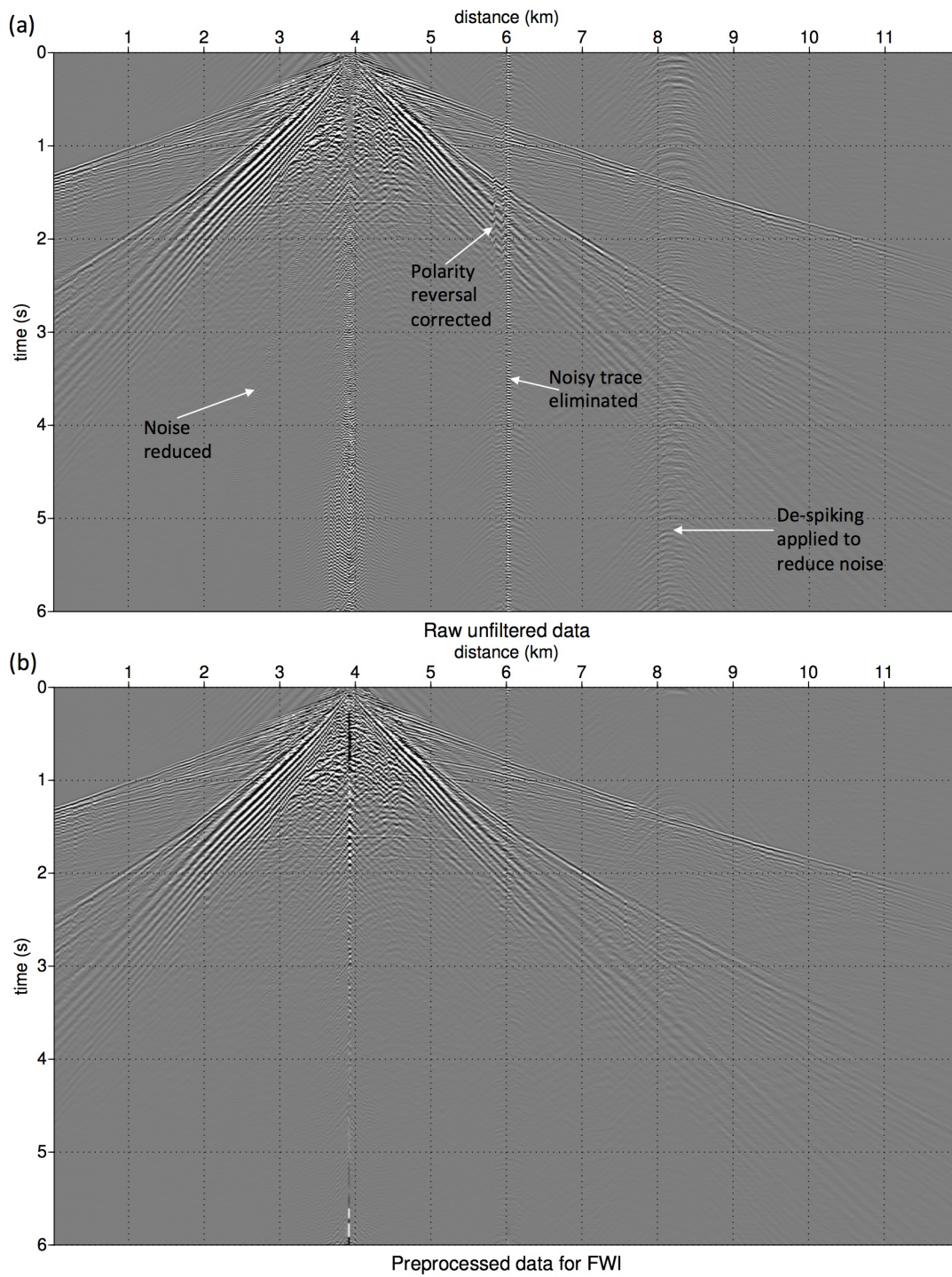


Figure 4.9: An example common shot gather of the raw data (a) and of the pre-processed data (b) after noise removal and regularisation processing steps. Different sources of noise that are corrected for during the pre-processing step are pointed out in (a).

algorithm used in this study is 2D, and therefore implicitly propagates a transverse line source, assuming the invariance of the 3D velocity model along the transverse direction. Spreading transformations for 3D-to-2D media therefore need to be applied to the observed data, to simulate the response to a transverse line source, and allow an appropriate comparison between observed and calculated data (Schäfer et al., 2014; Forbriger et al., 2014).

In this study the observed data are multiplied by \sqrt{t} , where t is the recorded time, as supported by the specific amplitude decay of surface waves in 2D and 3D. This pre-processing 3D-to-2D geometrical spreading correction has been used for previous similar studies (Bretaudeau et al., 2013; Pérez Solano, 2013), although others argue that it may be insufficient for FWI, when applied to waves propagating with different phase velocities, such as P, S and Rayleigh waves (Schäfer et al., 2014).

4.2 Surface wave inversion

4.2.1 Initial models for FWI

First break picking and inversion was performed on the dataset to provide an initial P-wave velocity model for FWI. This model was also used as an input for Surface Wave 1D Inversion (SWODI), for which dispersion curves were picked to obtain a shear velocity model of the near surface region.

Figure 4.11a illustrates the first 228 m of the model obtained by the first break travel time inversion. For the near surface region considered, the P-wave velocities retrieved do not vary greatly, with velocities ranging between 2800 – 3250 m/s . The velocities retrieved generally increase with depth, and are slightly higher on the right hand side of the model. The resolution is very low due to the method not being well adapted to retrieve the shallow near surface region. The ray coverage for the tomography is illustrated in Figure 4.10. For the first 300 m the fold appears high, although slightly reduced on the edges of the acquisition.

The initial density model (see Figure 4.11c), was chosen to be constant at 2200 kgm^{-3} . This value was arbitrarily chosen based on average values of the sedimentary rock types present in the region, because no in situ values or well information were available for this near surface region.

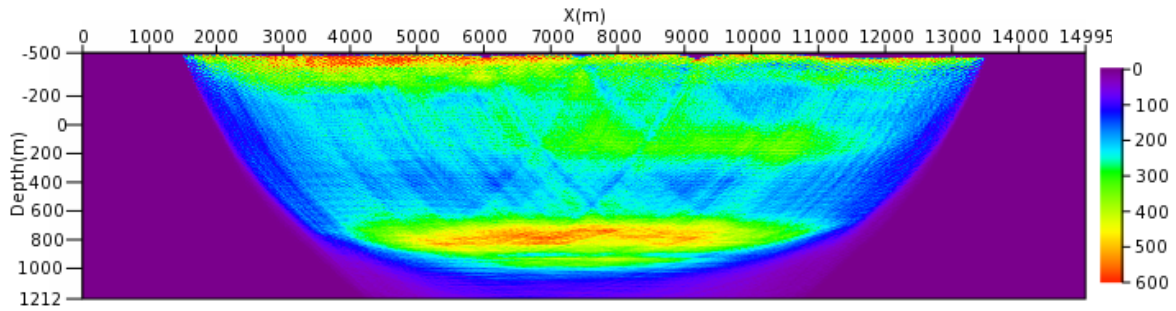


Figure 4.10: Ray coverage for first break travel time inversion of the real data to obtain an initial P-wave velocity model. Note that the distance coordinates are local, but data is only present for the 12 km length of the acquisition line. The depth of the inversion (also in local coordinates) extends beyond the near surface region considered for inversion of surface waves.

4.2.1.1 Surface wave 1D inversion

Two separate studies of conventional surface wave 1D inversion (SWODI) were performed on the dataset. This surface wave analysis method, described in detail in section 1.2, assumes a stratified medium and obtains a shear velocity profile with depth from the inversion of dispersion curves. Multiple laterally invariant inversions allow to produce an image of the lateral variations in the shear velocity model. Although the same methodology is used for both studies, different algorithms and tuning parameters were independently applied, leading to two separate shear-velocity models. The comparison allows a quality control and validation of the models, and to evaluate epistemic uncertainties.

For the first study, common midpoint (CMP) gathers every 10 m along the acquired line, were first preprocessed. A compensation for the natural attenuation of surface waves was applied to the data. The data were then muted to focus on the surface wave energy and reduce the ambient noise. Note that the guided waves marked by (C) in Figure 4.3 were also removed. The mute facilitates the building of good spectral estimations on which the fundamental mode can be picked. Several tests with varying offset windows were performed to reach an equilibrium between the presence of low frequency content with a high penetration depth, obtained by far offset data, and minimizing the presence of 2D effects that are caused by such far offset data.

The fundamental mode was picked in the (ω, k) domain, computed using the Capon spectral estimator (Capon, 1969). The picking was aided by an automatic search for the local maximum of the spectrum, and an extrapolation of the dispersion curve to

higher and lower frequencies. Furthermore an interpolation strategy allowed to reduce the number of dispersion curves manually picked. Examples of the (ω, k) spectra and picks are shown in Figure 4.14. Although some higher modes may have been identified, only the fundamental mode was inverted.

The initial shear velocity model for SWODI (see Figure 4.11b), was obtained from applying a constant Poisson ratio of $V_P/V_S = 2.0$ to the P-wave model obtained from first break tomography. Alternative initial shear velocity models such a homogeneous model, or velocities based on a V30 model (Brown et al., 2000; Martin and Diehl, 2004) did not suffice for convergence to occur. The topography of the selected initial models was smoothed over 300m for the inversion.

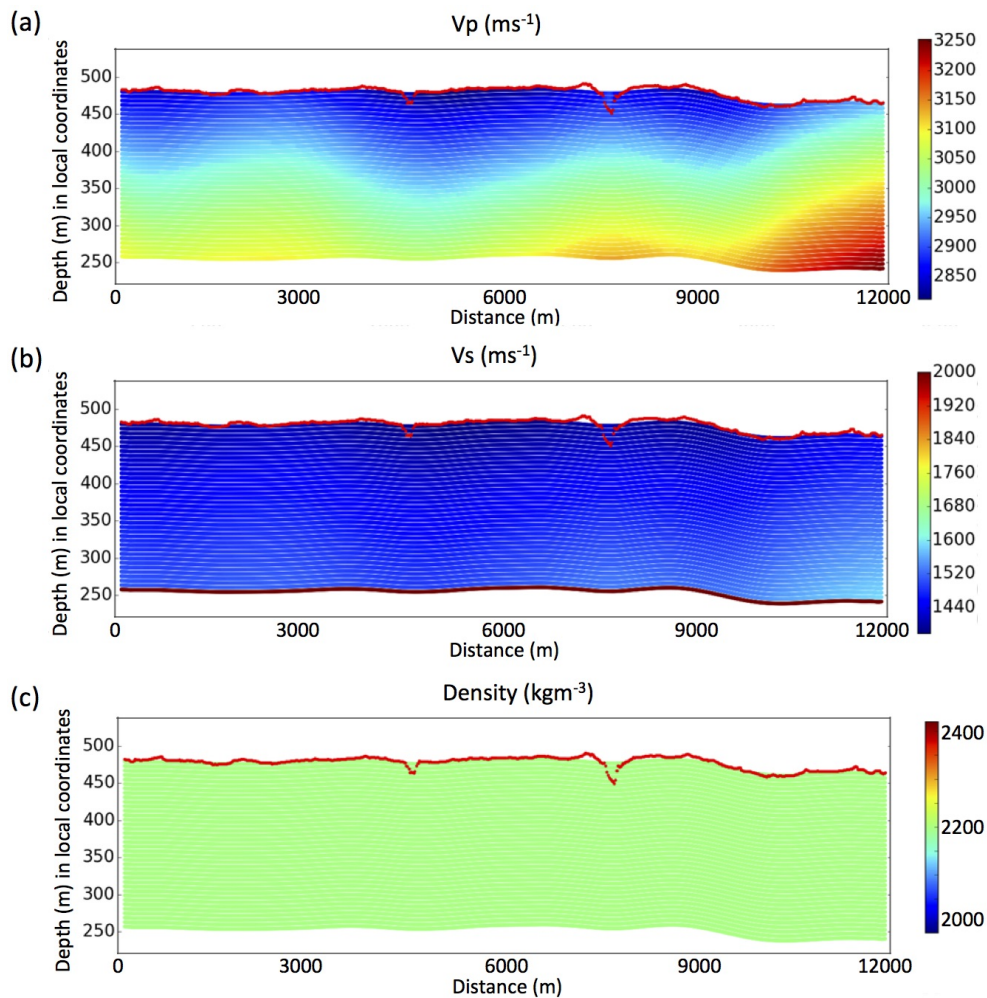


Figure 4.11: Initial P-wave (a), shear wave (b) and density (c) models used for SWODI. The real topography is plotted in red. Note that the color scale is different for each model.

Only the shear wave velocity was inverted for, while the P-wave velocity and density were kept fixed. The forward modelling kernel is based on the hypothesis of a 1D horizontally layered model. An assumed 37 layers, each with a thickness of 6 m , were inverted for, with a final layer at depth representing the infinite half space, fixed with a shear velocity of 2000 m/s (see Figure 4.11b). The depth of the model is therefore 222 m . Within each layer the velocity was assumed to be constant. Velocity inversions with depth were allowed by the inversion. The resulting 2D model of the subsurface therefore consists of several 1D inversions.

The L2 norm of the difference of dispersion curves was used as a misfit function to compare picked dispersion curves with the ones obtained by the current velocity model. The Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) was then used as a local search method to navigate in the parameters domain and minimize the misfit function. For the cases where the inversion did not converge, the shear velocity was interpolated.

The final shear wave velocity model obtained by SWODI is illustrated in Figure 4.12 with the smoothed topography. One can observe that very strong velocity anomalies are present. A low velocity top layer of about 20 m can be identified along the whole line, with velocities of 800–900 m/s . A relatively consistent velocity inversion can be observed at around 60 m depth relative to the surface. Strong velocity contrasts sometimes appear at a depth of around 110 m . This may be coherent with the results shown in Figure 4.4b, where the presence of higher modes observed in the data are found to be linked to such a feature. However the model appears poorly constrained at depths greater than 125 m , suggesting that a lower confidence level should be given to lower depths, as this may be the maximum penetration depth for the frequency content of the data used for inversion.

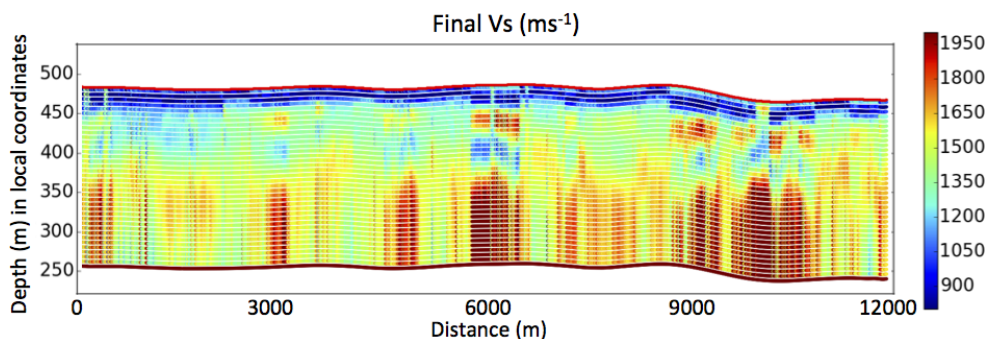


Figure 4.12: Final shear wave velocity model obtained by SWODI. The smoothed topography is plotted in orange. Note that velocities above 2000 m/s were clipped.

To be used for FWI, the shear-velocity model will have to be heavily smoothed, as the strong lateral variations and contrasts will create instabilities in the forward modelling (discussed in section 4.2.1.3), and produce significant backscattering which is not as evident in the observed data.

A separate SWODI was performed on the same dataset by CGG. This second inversion was computed using the same initial P-wave velocity, shear velocity and density models and similar inversion parameters, but using CGG in-house picking and inversion tools. The comparison of the two inversion results can be used as a quality control. The resulting shear velocity model obtained by CGG is compared to the one obtained by the first study in Figure 4.13.

One of the main differences is that the picking of the dispersion curves by CGG was done semi-automatically, as each picking was guided by neighboring picks, leading to a smoother result. The fundamental mode is assumed as the most energetic and predominant signal, and is the only mode picked. The picked frequencies are not as high as those for the first study. Certain artifacts present at the surface level in the final model obtained by CGG may be due to the lack of high frequency information in the picked dispersion curves, leading to velocities being extrapolated up to the surface. One can observe anomalous velocity values in the top surface layer, which appears less continuous than in the model obtained from the first study (see Figure 4.13).

Again, only the shear velocity was inverted in the inversion by CGG. For the inversion, 25 layers of 5 *m* thickness, each of a constant shear velocity, were assumed. The resulting model is therefore only 125 *m* deep, and appears relatively well constrained up to this depth, as opposed to the model from the first study. For this second inversion, a smoothed topography over 1 *km* was applied. This is also the length of the array that was used for the computation of the dispersion curves.

The SWODI results compared in Figure 4.13 are a clear example of how manual picking and inversion parameters can lead to very different models. One can observe that the model obtained by CGG (Figure 4.13b) is much smoother and displays a smaller range of velocities than the model obtained in the first study shown in Figure 4.13a. Yet certain similarities between the two models allow the validation of certain subsurface features. The low velocity top layer about 20 *m* deep, is recovered by both studies. The velocity inversion at about 60 *m* from the surface is also present in both results. Furthermore a strong lateral velocity discontinuity is imaged in both models, just after the first 2 *km* of the acquired line (marked by the red dashed line in Figure 4.13).

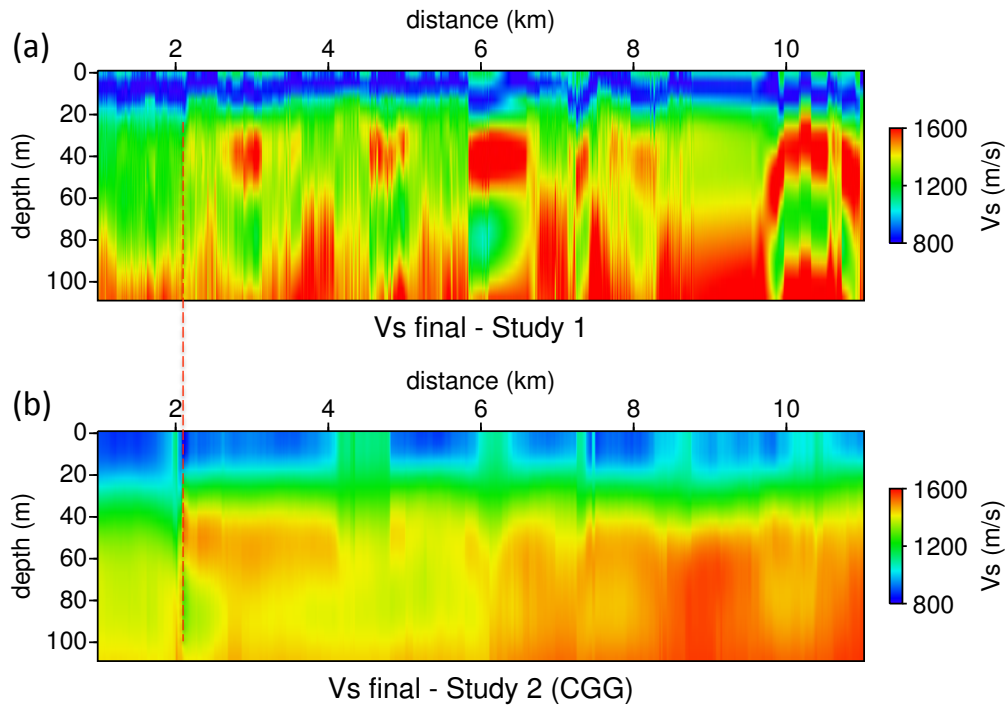


Figure 4.13: Final shear wave velocity models obtained from two separate SWODI studies. Note that only the first 110 *m* are shown of the velocity model in (a), for a fair comparison to the velocity model in (b) provided by CGG. The topography is also flattened for both models. The red dashed line marks a strong lateral velocity discontinuity.

The left-hand side of the discontinuity appears to have lower shear velocity values than the right-hand side. From observations of satellite imagery, the discontinuity in the subsurface appears to match with the edge of an alluvial fan, visible on the surface of the terrain, and located on the faster right hand side of the discontinuity.

4.2.1.2 Quality control of dispersion curves

A more detailed quality control can be achieved by analysis of the dispersion curves. Selected (ω, k) domain spectra of the data are illustrated in Figure 4.14. The picked fundamental mode (green dashed line) and fundamental mode corresponding to the final model obtained in the first study by the SWODI approach (blue solid line) are plotted on the top of the spectra. The position of the gathers in relation to the final shear velocity model of the first study is given in Figure 4.14e for reference.

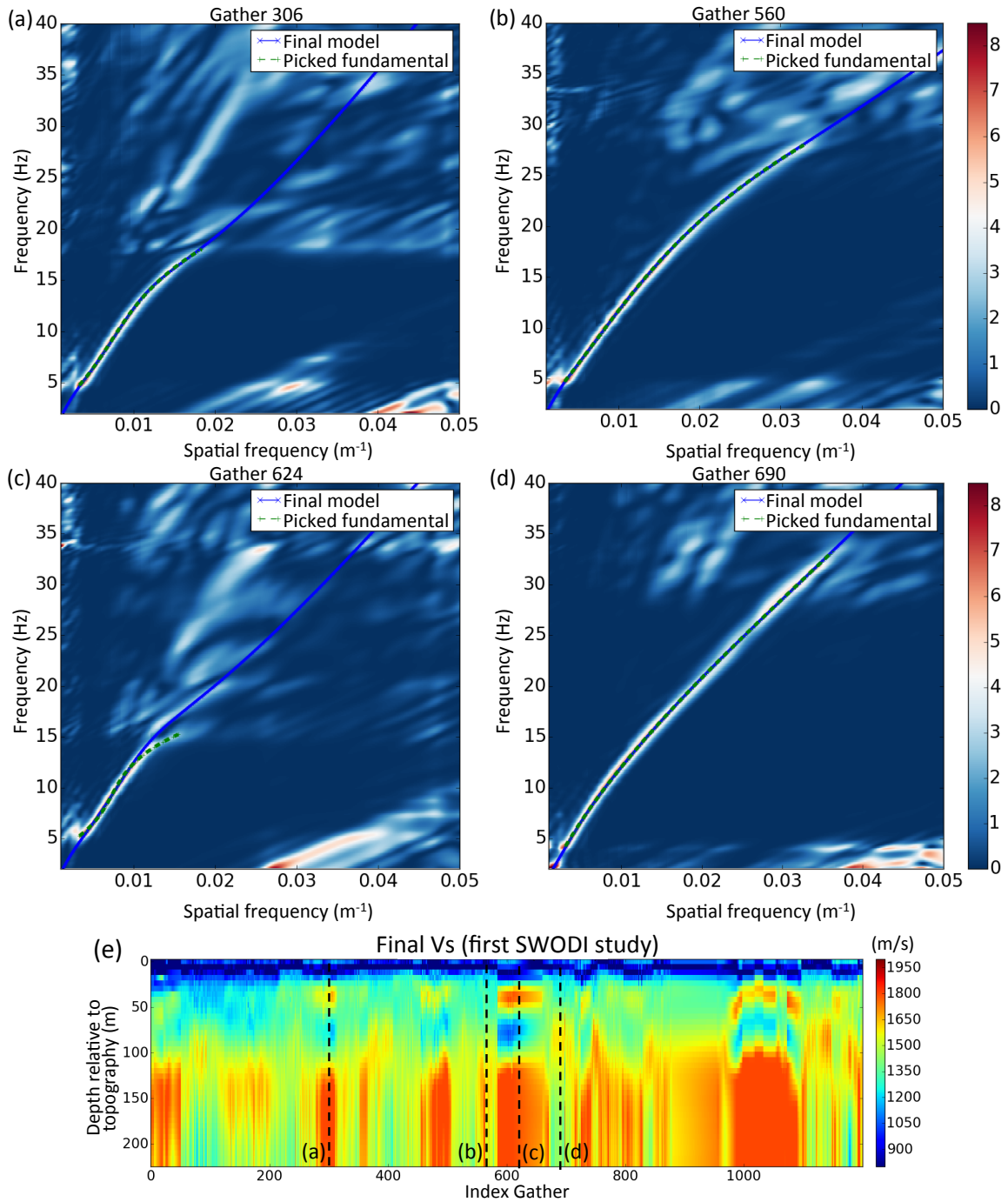


Figure 4.14: Selected CMP gathers in the (ω, k) domain (a-d), on which the fundamental mode was picked (green dashed line). The fundamental mode of the final SWODI model from the first study, and shown for reference in (e) is also plotted for a quality control (blue solid line). Note that a normalisation is applied on both frequency and spatial frequency axis for better visualisation of the spectrum and all spectra have the same color scale. The final shear wave velocity model is flattened, as required by the FWI algorithm, and velocities above 1800 m/s where clipped. The location of the CMP gathers is marked by black dashed lines.

One can observe that, for certain regions, the fundamental mode is strongly visible in the (ω, k) spectrum, and is correctly retrieved within a frequency range of 4 – 30 Hz (see Figures 4.14b and 4.14d). These results indicate a good quality of the final shear velocity model, perhaps due to a homogeneous subsurface.

However regions exist where the (ω, k) spectrum is more complex and energetic higher modes can be identified at higher frequencies. For these spectra, the picking of the fundamental mode is limited to 15 Hz (see Figure 4.14a). For these areas the resulting shear velocity model often yields a high velocity contrast at depth. This is coherent with the results in Figure 4.4b, where the appearance of higher modes is shown to be linked to the presence of a high velocity contrast. Yet this high velocity contrast does not appear to be laterally consistent in the model, and the lateral variations retrieved may not be reliable.

An interesting area to analyze is the location of the lateral velocity discontinuity, retrieved by both models, just after the first 2 km of the line. An example of the (ω, k) spectra before, at the discontinuity, and after is shown in Figure 4.15. Comparing the spectra before and after, one can observe a change in the slope of the fundamental mode, indicating an increase in velocity from before to after the discontinuity. The spectrum located at the discontinuity appears smeared, as if two dispersion curves of different velocities, coming from the forward and reverse sides of the spectrum, were superimposed. The analysis further validates the presence of a strong lateral velocity discontinuity at this location.

There are several other regions along the 12 km line, where significant 2D effects may be observed, and where only a limited convergence of SWODI occurred for the first study. An example spectrum is shown in Figure 4.14c, in which the fundamental mode appears to split at a frequency of about 15 Hz . The spectra may be explained by an occurrence of two fundamental modes due to 2D effects, or due to a significant portion of energy transferred to higher modes. The subsurface geology may be more complex for this region, and its reconstruction beyond the limitations imposed by 1D inversion.

Local 2D effects can also be observed in Figure 4.14c for gather 624. It is interesting to note that on either side of this region, the fundamental mode is clearly visible, as illustrated in gathers 560 and 690 (Figures 4.14b and 4.14d). This shows that the strong variations in the shear velocity encountered are local. Dealing with such 2D effects is one of the limitations of conventional surface wave inversion that may be overcome by FWI.

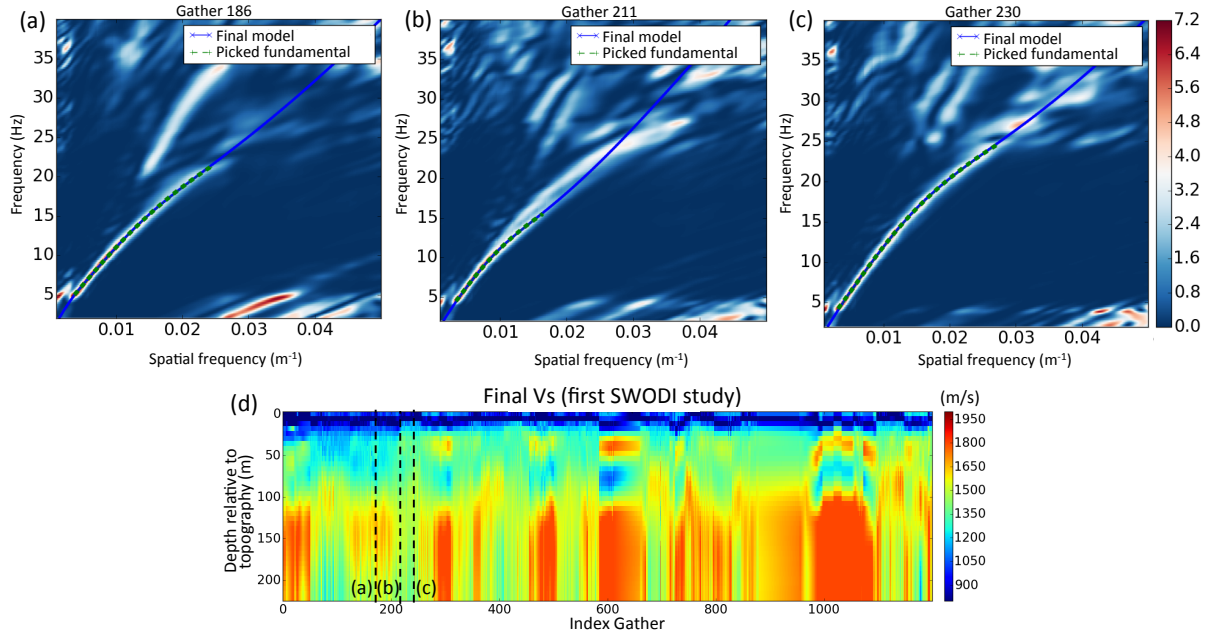


Figure 4.15: Selected CMP gathers in the (ω, k) domain located before (a), during (b) and after (c) a lateral velocity discontinuity retrieved by both shear velocity models just after the first 2 km. The final model of the first study is shown for reference, on which the CMP locations are marked by the dashed lines. The picked fundamental mode (green dashed line) and the fundamental mode of the final model obtained by the first study (blue solid line) are plotted. Note that a normalisation is applied on both frequency and spatial frequency axis for better visualisation of the spectrum.

4.2.1.3 Selection of initial models

One can interpret the regions of strong velocity contrasts at depth, present in the SWODI results of the first study, as caused by the presence of higher modes. Although a high velocity contrast appears to be validated by the test shown in Figure 4.4, a lower confidence level is given to the lateral variations retrieved, especially since the geology of the area is assumed to be flat. The semi-automatic picking by CGG may have helped to avoid picking instabilities due to higher modes, providing a smoother, shallower and simpler result, with velocity variations that appear as more realistic. The quality control analysis privileges the shear velocity model obtained by CGG to be used as an initial model for FWI.

Another significant reason to prefer the model obtained by CGG compared to the one from the first study, is related to the numerical instabilities that may occur during forward modelling for FWI at the oil and gas exploration scale. Even when the shear

velocity model from the first study is heavily smoothed, the strong velocity contrasts cause the absorbing boundary conditions (implemented with perfectly matched layers procedure), to explode, leading to the presence of significant artifacts in the data.

Such instabilities are characteristic for strong Poisson ratio variations located close to the model boundaries during elastic modelling, especially on the real exploration data scale. Implementing absorbing boundary conditions as proposed by [Cerjan et al. \(1985\)](#) may be a possible solution to mitigate artifacts. Otherwise one may rely on smoothing strategies for the zones close to the boundaries, or a tapering of the shear velocity to reach an acoustic propagation at the model boundary to help reduce such instabilities, but they remain difficult to deal with. The initial model for FWI is therefore required to be smooth, and the smoother shear velocity model obtained by CGG does not create these numerical instabilities.

An additional Gaussian smoothing over 14 *m* vertically and 150 *m* horizontally was applied on the shear velocity model obtained by CGG to finalize it as an input for FWI. Both this shear velocity model and the P-wave velocity model obtained from first break travel time inversion, were then resampled to a spatial discretization step of 3 *m*, to guarantee at least 10 discretization points per wavelength. Furthermore they were flattened as required by the assumptions of the FWI algorithm used.

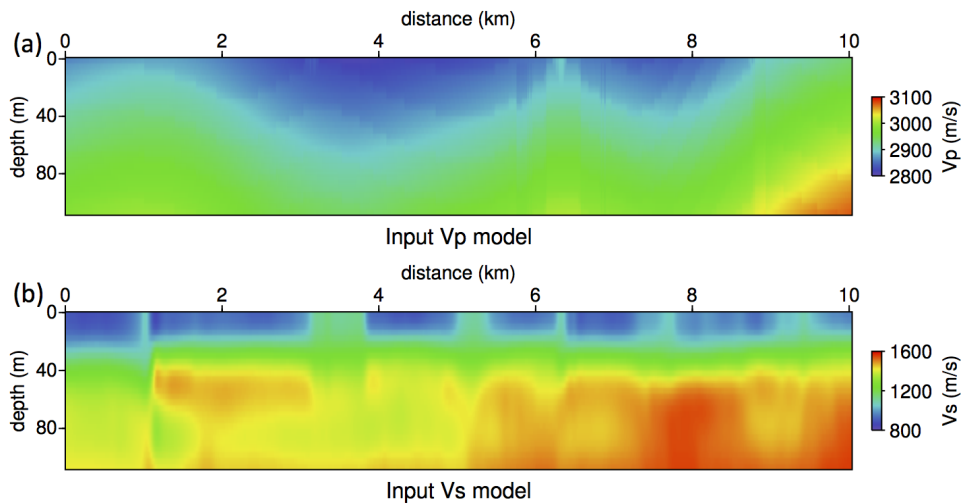


Figure 4.16: P-wave velocity (a) and shear wave velocity (b) input models for FWI. Note that the two models are obtained at different resolutions.

Figure 4.16 illustrates the finalized P-wave velocity and shear wave velocity models used as inputs for FWI. It is important to note that the two models have been obtained at

different resolutions. One could assume that the weak velocity variation in the initial P-wave model, obtained at very low resolution for the shallow near surface scale considered, has very little impact on the surface waves simulated during the forward modelling.

Yet how good are these initial models and are they sufficiently accurate for FWI? It is not evident to answer this question. No well information is available and no other non-seismic geophysical prospection was performed. The only way to analyze and validate them is through forward modelling and the comparison of calculated data with the observed dataset.

4.2.2 Impact of receiver array

A validation test for the impact of the receiver arrays used in the real acquisition is investigated here. Receiver arrays are implemented to help reduce the surface noise recorded during real acquisitions, as required for conventional seismic processing. The aim of this study however, relies on the presence of surface waves, which may have been weakened at high frequencies due to the arrays. Furthermore, side effects such as spatial averaging and loss of lateral resolution may also occur.

The observed data was acquired with receiver arrays consisting of 12 inline vertical geophones, spaced 1.66 *m* apart, to create receiver stations every 20 *m*. This acts as a spatial filtering (see Figure 4.17) when the traces of the receiver stations are summed together. The response of such a filter is usually analyzed in the wavenumber domain, where a simple calculation shows that

$$\int_{-a}^a e^{ikx} dx = 2 \int_0^a \cos(kx) dx = \frac{2}{k} \left[\frac{\sin(kx)}{k} \right]_0^a = \frac{2 \sin(ka)}{k} = 2a \left(\frac{\sin(ka)}{ka} \right). \quad (4.1)$$

The maximum distance of the stations to the center of the array is considered to be $a = 10$ *m*. The impulse response of the filter is illustrated in Figure 4.17, showing that the first spatial notch occurs at $\lambda = 2a$. Surface wave amplitudes of wavelengths $\lambda = 20$ *m* will therefore be filtered, as destructive interference of the phase e^{ikx} will occur. The first spatial notch corresponds to a frequency of 50 *Hz* when an average surface wave velocity of 1000 *m/s* is considered.

During FWI, the forward modelling is computed considering single vertical receivers spaced 20 *m* apart (see Figure 4.18). To investigate the impact of this acquisition

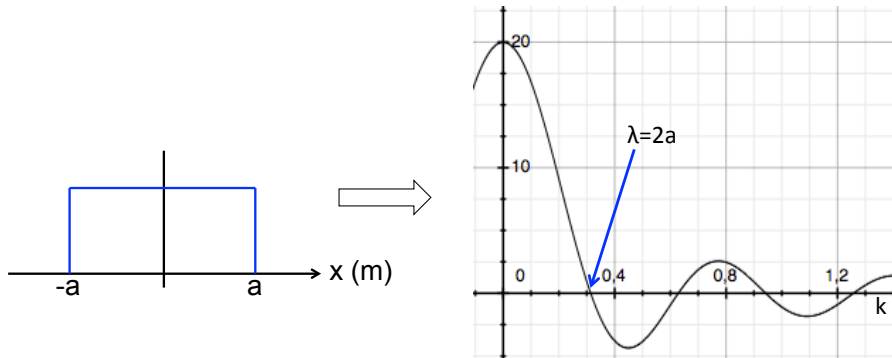


Figure 4.17: Schematic of spatial filter and impulse response applied to the data due to the array acquisition.

difference on the data, a simple forward modelling test is performed to compute data with simulated receiver arrays and with simulated single receiver positions, keeping all other modelling parameters constant. A staggered finite-difference grid is used for the forward modelling such as described in section 2.2.3. The initial shear velocity and P-wave velocity models illustrated in Figure 4.16, as well as a homogeneous density model of $\rho = 2200 \text{ kgm}^{-3}$ are used as input models. A Ricker of 20 Hz peak frequency is used as the source wavelet. The simulated receiver arrays are computed as the mean of 12 individual receivers spaced 1.66 m apart, to limit differences in amplitude.

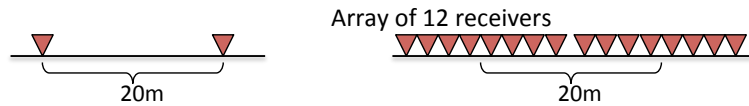


Figure 4.18: Schematic of single receiver acquisition (left) used to obtain the calculated data and receiver arrays (right) actually used for the acquisition of the observed data.

An example common shot gather for both modelling computations, bandpass filtered between 4 – 25 Hz, as well as the difference between the two is shown in Figure 4.19. One can observe that at low frequencies the data are almost equivalent, but for higher frequencies differences in the data occur, as is expected. From a closer analysis of an example trace in Figure 4.19d, it appears that the difference is mainly in amplitude, and that the phases are not significantly impacted. The signal-to-noise ratio for surface waves will therefore be poor at medium to high frequencies ($> 25 \text{ Hz}$), suggesting that layer stripping inversion strategies investigated in Chapter 3 may not be well adapted for this dataset. To conclude one can assume that the difference in acquisition between the real observed data and the data computed with single receiver positions is minimal

and may not significantly impact FWI when only low frequency data is considered.

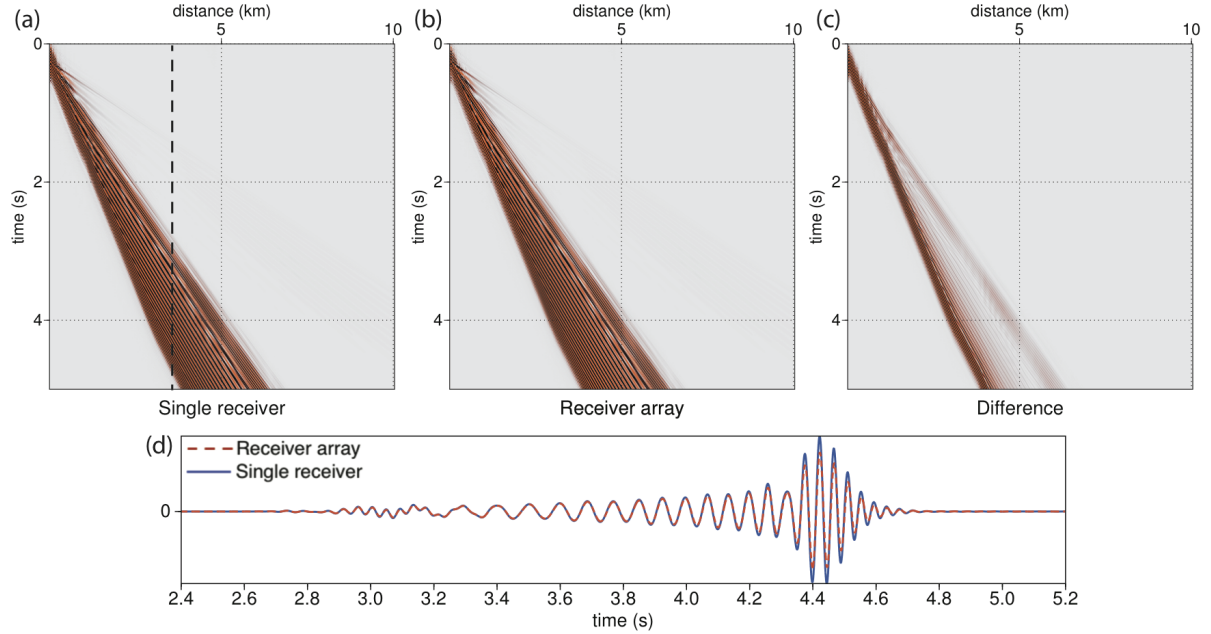


Figure 4.19: Example common shot gather obtained from modelling tests with simulated single receiver positions (a) and with simulated receiver arrays (b), bandpass filtered between 4 – 25 Hz , and the data difference (c). An example trace is compared in (d), located by the black dashed line in (a).

4.2.3 Source estimation

The source signature is estimated following the strategy proposed by Pratt (1999). The estimation problem is posed in the frequency domain, where the source terms in the modelling of the synthetic data are assumed to be multiplied by an unknown source signature $s(\omega)$, a complex scalar. When assuming that the other model parameters, such as the shear or P-wave velocity models, are known, the source signature can be obtained by minimizing the L2 norm of the data difference, where the minimum misfit occurs when

$$s(\omega) = \frac{\mathbf{d}_{cal}^* \mathbf{d}_{obs}}{\mathbf{d}_{cal}^* \mathbf{d}_{cal}}, \quad (4.2)$$

where \mathbf{d}_{obs} is the observed real data, \mathbf{d}_{cal} is the calculated data with a Dirac impulse as a source, and the symbol * represents the complex conjugate. Although the initial velocity models are only approximate for this inversion, Pratt (1999) claims that significantly good estimation of the source wavelet may be achieved.

Source inversion is performed on the real dataset. The shear wave velocity and P-wave velocity initial models illustrated in Figure 4.16 are used, as well as a homogeneous density model of $\rho = 2200 \text{ kgm}^{-3}$ as also assumed by SWODI. The forward modelling scheme implemented is the same as described in section 2.2.3, using a staggered finite-difference grid. Only the subsurface covered by the initial models, surrounded by absorbing boundary conditions and the free surface condition at the top, is considered for the propagation of waves. Therefore only the real data with receivers and sources located within the lateral boundaries of the shear velocity model is selected and used. The real acquisition of the observed data is simulated to compute the calculated data.

The topography is assumed as flat for the forward modelling. This assumption is constrained by the finite-difference scheme and free surface condition implemented in the algorithm. Considering an average surface wave velocity of 1100 m/s and a maximum frequency for surface waves of 30 Hz , giving a minimum wavelength of over 30 m , one could evaluate the true variation in topography as near negligible, and validate the assumption. Backscattering due to the canyons cutting through the acquisition line, is only weakly visible in the observed data, and may not have a strong impact on the misfit calculation. The limitations of a flat topography assumption are further discussed in section 4.4.1 of the perspectives, showing that this assumption appears valid for frequencies between $4 - 15 \text{ Hz}$, but may cause cycle-skipping for higher frequency data content.

The observed data was obtained from the correlation of consecutive vibroseis sweeps. As expected, the source function obtained from inversion is therefore non-causal, but rather zero-phase, with energy arriving before time zero. To simplify the manipulation of the source wavelet, a small time shift was applied to the real dataset, to obtain a causal source wavelet during inversion. Such a manipulation can be theoretically justified to be correct, as the time shift applied to the observed data was kept constant for all further manipulations and for later inversions.

Several mutes were also tested and applied to the observed and calculated data, to focus on certain waveforms such as surface waves or the direct wave, as well as muting the near offsets or far offsets. The sources obtained from inversion are then filtered between $4 - 90 \text{ Hz}$, which is the frequency spectrum of the vibroseis used for acquisition, and tapered to zero at the ends. Examples of two different source functions obtained are shown in Figure 4.20.

Very similar wavelets are obtained when using a mute that allows only surface waves,

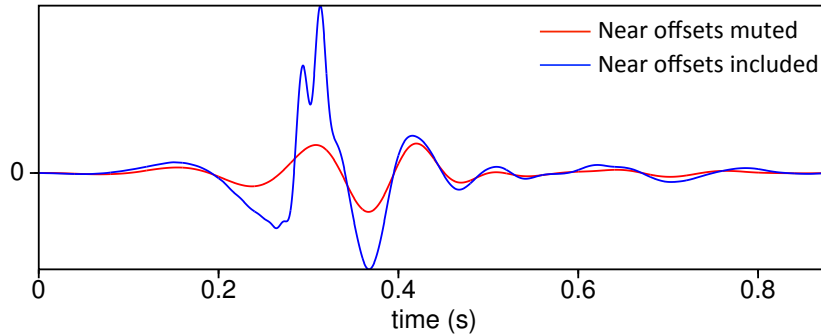


Figure 4.20: Source function obtained with a wide mute allowing surface waves and the direct wave: with near offsets removed (red); and near offsets included (blue).

and a mute that includes also other waveforms such as the direct wave. This is due to the surface waves having superior amplitudes, and therefore driving the inversion. The removal of near offset data has a bigger impact on the source function obtained. However the phases of the different sources obtained match well (see Figure 4.20), suggesting that the source inversion is relatively stable.

The source function obtained with a wide mute allowing surface waves and the direct wave, and with near offsets included, was chosen as the source used to compute the calculated data, as it provided the best match when calculated data and real data were compared.

4.2.4 Forward modelling and cycle-skipping analysis

The same initial models and forward modelling used for the source estimation are used to compute the calculated data using the source wavelet obtained in section 4.2.3. An initial comparison of the calculated and observed data allows to evaluate the differences, whether or not cycle-skipping is present, and if convergence is possible with FWI.

One of the significant factors that will create a data difference, and which is not considered in the forward modelling, is attenuation. Attenuation may create differences in amplitude especially for far offsets, and is a parameter that should be considered for perspective studies. Scaling coefficients may be used to consider attenuation effects during source estimation (Groos et al., 2014), although using appropriate viscoelastic modelling with a priori values for the quality factor Q is more accurate. On the global seismology scale, attenuation can even be retrieved from the inversion of surface waves (Romanowicz, 1995), suggesting a potential for future studies at the oil and gas scale.

In the following investigation, phase differences of surface waves will be analyzed, since the aim of the inversion is to retrieve the kinematics rather than the amplitudes. The phase differences between observed and calculated data will lead to an initial guess for the presence of cycle-skipped data and local minima for FWI. Low frequency data is investigated, for both the time-offset (t, x) domain and the frequency-wavenumber (ω, k) domain, to determine the possible use of multiscale strategies for FWI.

4.2.4.1 Analysis of low frequency data (4 – 15 Hz)

A selected common shot gather located at the beginning of the line, and bandpass filtered between 4 – 15 Hz, is shown for observed and calculated data in Figure 4.21. One can observe that several of the waveforms present in the observed data are not present in the calculated data. Refracted and reflected waves coming from interfaces below the depth of the near velocity model used to compute the calculated data will not be retrieved by the forward modelling. The guided waves indicated in Figure 4.3 are not retrieved either, at least by the initial models available.

At an offset of about 2 km, the initial phases of the surface wave package, filtered between 4 – 15 Hz, appear to roughly fit calculated data. Cycle-skipping occurs however for surface wave phases within the package. The same result is obtained at an offset of about 4 km, although a significant difference in amplitude between observed and calculated data can be observed. This first comparison is very positive, suggesting that the fundamental mode was correctly picked and inverted. The result validates the shear velocity values provided by the initial model.

Conventional FWI in the (t, x) domain would not be able to converge due to the cycle-skipping effect present for the surface waves. Multiscale approaches would also not be helpful in the (t, x) data domain. Only time damping strategies, or layer stripping may help overcome the cycle-skipping problem. However layer stripping strategies are not an optimal solution as previous analysis has shown that the medium to high frequency surface waves have a low signal-to-noise ratio (see section 4.2.2). Computing the data misfit in alternative domains, such as the (ω, k) domain, can be a more robust approach to mitigate the cycle-skipping feature present.

Another selected common shot gather, viewed in the (ω, k) domain, and bandpass filtered between 4 – 15 Hz is shown for observed and calculated data in Figure 4.22. One can observe that the fundamental mode appears as the most energetic signal in both datasets, Observed and calculated traces are compared at a frequency of 7.6 Hz,

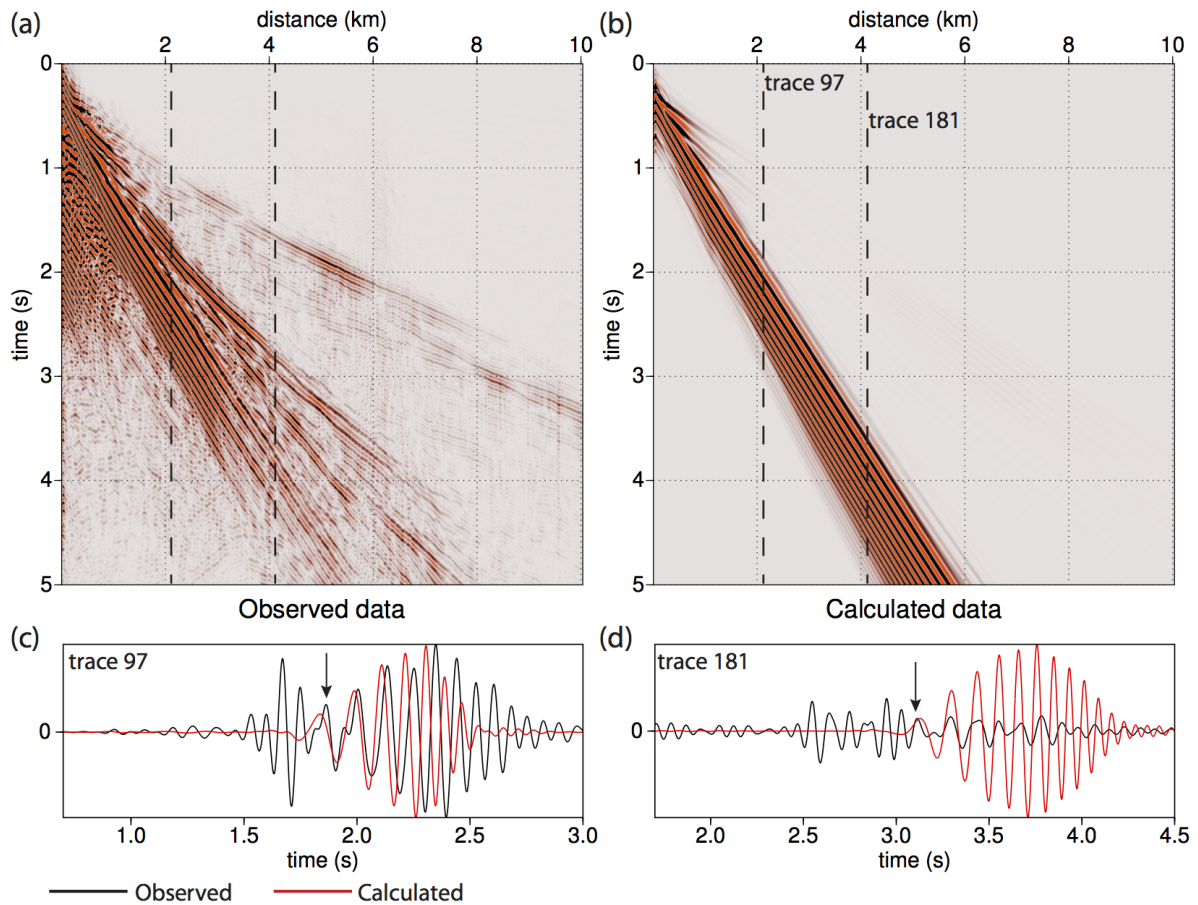


Figure 4.21: A selected common shot gather at the beginning of the line, filtered between $4 - 15 \text{ Hz}$, shown for the observed data (a) and the calculated data (b) in the time-offset (t, x) domain. Two traces, located by the black dashed lines, are visualized and compared in (c) and (d). The arrow marks the start of the surface wave package. Although the initial phases of the surface waves appear to roughly match, cycle-skipping of surface wave phases is also present at both offsets. Note that the amplitudes of the observed and calculated data are shown at the same scale.

showing that the peak of the fundamental mode is well matched. At such low frequencies, cycle-skipping effects do not appear to occur in the (ω, k) domain. At a slightly higher frequency of 12.2 Hz , the energetic peak of the fundamental is still sufficiently well matched between observed and calculated, although the data appears more vulnerable to cycle-skipping problems, due to the presence of secondary energy peaks in the observed data, and differences in amplitude between observed and calculated traces. The well matching peaks in energy corresponding to the fundamental mode, validate the initial shear velocity model obtained by Surface Wave 1D Inversion (SWODI) in section 4.2.1, which only uses information from picked fundamental modes for inversion.

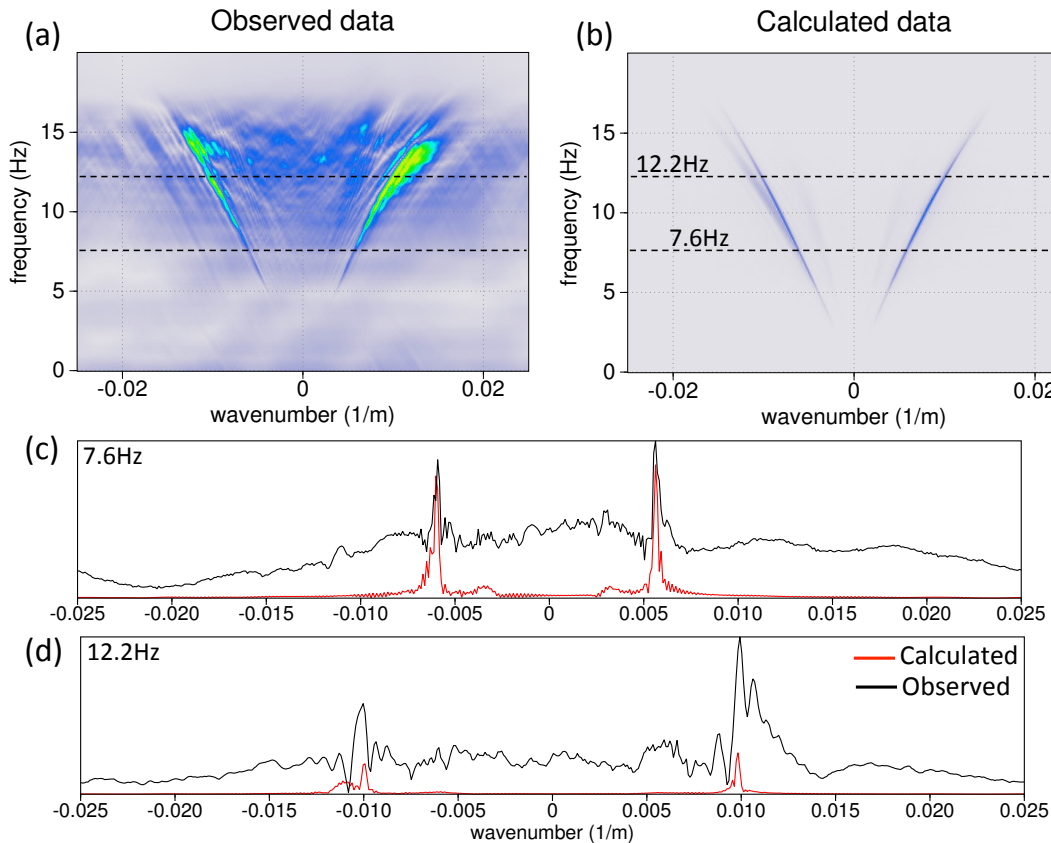


Figure 4.22: A selected common shot gather in the middle of the line, filtered between $4 - 15 \text{ Hz}$, shown for the observed data (a) and the calculated data (b) in the frequency-wavenumber (ω, k) domain. Two traces, located by the black dashed lines, at a frequency of 7.6 Hz (c) and 12.2 Hz (d) are selected and compared. The peak of the fundamental mode appears to correlate well between observed and calculated data. Note that the amplitudes of the observed and calculated data are shown at the same scale.

This analysis suggests that FWI with a misfit function in the (ω, k) domain and a multiscale approach inverting low-to-high frequencies may possibly converge. The (ω, k) domain is shown to be more robust than the (t, x) domain when considering low frequency data for multiscale strategies.

4.2.4.2 Muting and windowing in offset

To further analyze the data difference between observed and modelled data, a mute is applied in the time-offset (t, x) domain. This mute allows to focus on the surface waves, and tries to remove the reflected and guided waves (marked by B and C in Figure 4.3), which are not retrieved by the forward modelling as only the shallow near surface region

is considered for wave propagation, and these phases originate from much deeper. As a consequence the direct arrival which is retrieved by the forward modelling is also muted. A windowing in offset is then also applied to remove both the strong amplitudes at short offsets, and the far offsets that may be more strongly cycle-skipped. Examples of common shot gathers before and after the mute and windowing, are shown in Figure 4.23.

The receiver information at offsets less than 450 m is therefore removed. For the real dataset considered, it is clear that Equation 3.3 obtained in Chapter 3 to calculate a maximum offset, is no longer valid and much larger offsets need to be considered such as $x_{max} \simeq 10\bar{\lambda}_S$ up to $x_{max} \simeq 20\bar{\lambda}_S$.

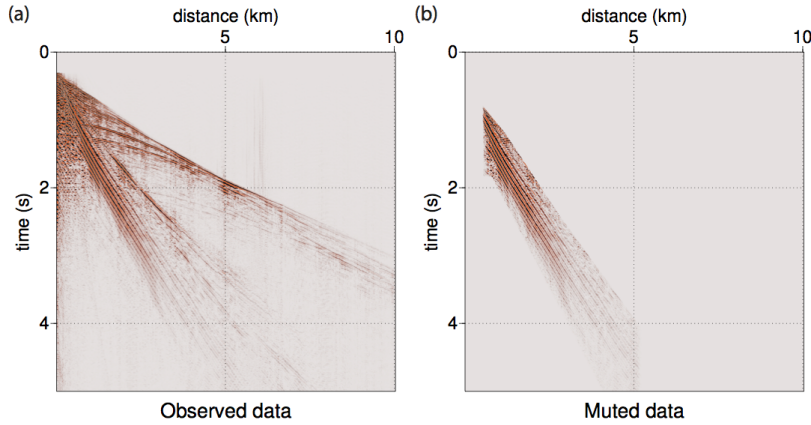


Figure 4.23: A selected common shot gather of the observed data before (a) and after (b) muting in the time-offset (t, x) domain, and windowing in offset.

4.2.4.3 Masking in the (ω, k) domain

To further focus on the correctly modelled fundamental mode observed in the frequency-wavenumber (ω, k) domain, as illustrated in Figure 4.22, a masking is applied while in the (ω, k) domain to remove higher modes and other signals. Masking in the (ω, k) domain has already been mentioned by Pérez Solano (2013) as a possible perspective for FWI with surface waves, to allow a progressive inversion of first the fundamental mode and then higher mode information.

In the (ω, k) domain, a straight line passing through the origin corresponds to a constant phase velocity. As such, prior information on the velocity of surface waves may help define a section of the (ω, k) spectrum to keep and a section to remove. In this study,

a weighting is created, tapered on either side with a Hanning function, and applied to the (ω, k) spectrum to prioritize the fundamental mode. The weighting is designed using the modelled data, which is cleaner and has a clear fundamental mode. The data are previously muted, tapered, and filtered with a low frequency bandpass ($4 - 16 \text{ Hz}$).

An example of a common shot gather in the (ω, k) domain, with the weighting applied, is shown in Figure 4.24 for both observed and calculated data. One can observe that for the low frequencies considered, the fundamental mode is correctly selected by the applied weighting. Furthermore the energy peaks of the fundamental fit well between observed and calculated data, although secondary peaks are visible at higher frequencies as seen in the trace analysis in Figures 4.24c - 4.24d.

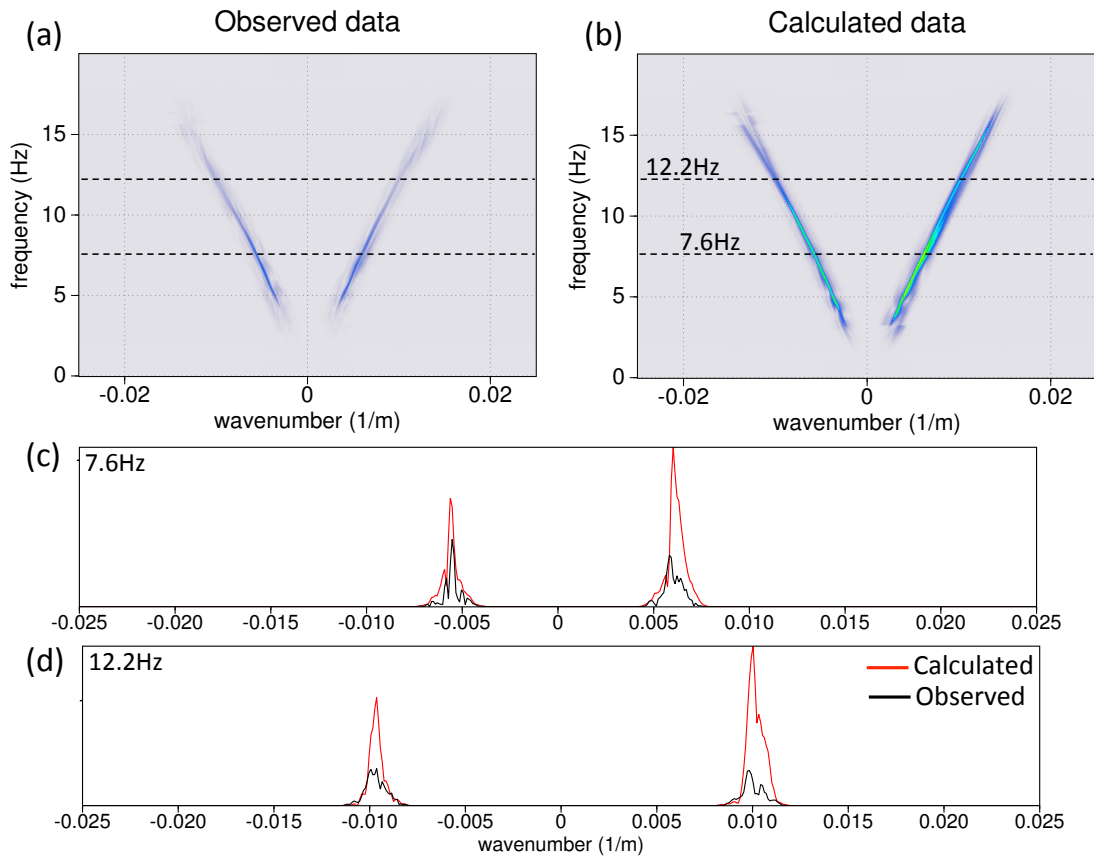


Figure 4.24: A selected common shot gather, muted in the (t, x) domain, filtered between $4 - 15 \text{ Hz}$, is shown in the frequency-wavenumber (ω, k) domain for the observed data (a) and the calculated data (b). A weighting has been applied in the (ω, k) domain to focus on the fundamental mode. Two example traces, at a frequency of 7.6 Hz (c) and 12.2 Hz (d) are selected and compared. The peak of the fundamental mode appears to correlate well between observed and calculated data. Note that the amplitudes of the observed and calculated data are at the same scale.

The data which has been weighted in the (ω, k) domain, can be transformed back to the time-offset (t, x) domain to observe what signal is contained in the fundamental mode. The shot gather previously pictured in Figure 4.24, is illustrated in Figure 4.25 transformed back to the (t, x) domain. As expected, the signal is almost of constant slope and strongly oscillating.

Selected traces between observed and calculated data are compared in Figures 4.25c - 4.25e. At near offsets (see Figure 4.25c), the initial two modelled phases, at relatively low frequency, appear to arrive too slow compared to the true phases. The next phase however, appears to fit well with the true phase. The following modelled phases, at relatively higher frequencies arrive too fast compared to the observed ones, and cycle-skipping clearly occurs. At medium offsets (see Figure 4.25d), the first four modelled phases, at relatively low frequency, appear to arrive too slow compared to the true phases, possibly causing cycle-skipping. However the following phases, at slightly higher frequencies, fit well once again. The final phases, at even higher frequencies, illustrate again a phase shift, as the modelled phases arrive too fast compared to the true ones. At far offsets, beyond 3000 – 3500 *m* distance from the source position, the previously observed trend is no longer valid. All modelled phases appear too slow compared to the true ones, clearly causing cycle-skipping to occur (see Figure 4.25e). The differences in amplitude visible for medium to far offsets, and especially for the higher frequencies is due to attenuation effects not being considered during the propagation of the modelled data.

This analysis on the signal contained by the fundamental mode of the surface waves suggests that overall the initial velocity model used to create the modelled data may be a bit too slow, especially considering the long wavelength scale, and therefore the deeper part of the model. The velocity model appears to work well for medium frequency data at near to medium offsets. However the cycle-skipping issue for signal at frequencies higher than 10 – 15 *Hz*, suggests that areas in the top half of the velocity model may be too fast.

4.2.5 Full Waveform Inversion

Although the forward modelling and cycle-skipping analysis presented in the previous section highlights many difficulties, FWI is attempted on the real dataset. Layer stripping strategies for FWI are not investigated because of the low signal-to-noise ratio for high frequency surface waves, due to the array acquisition design (see section 4.2.2), and lack

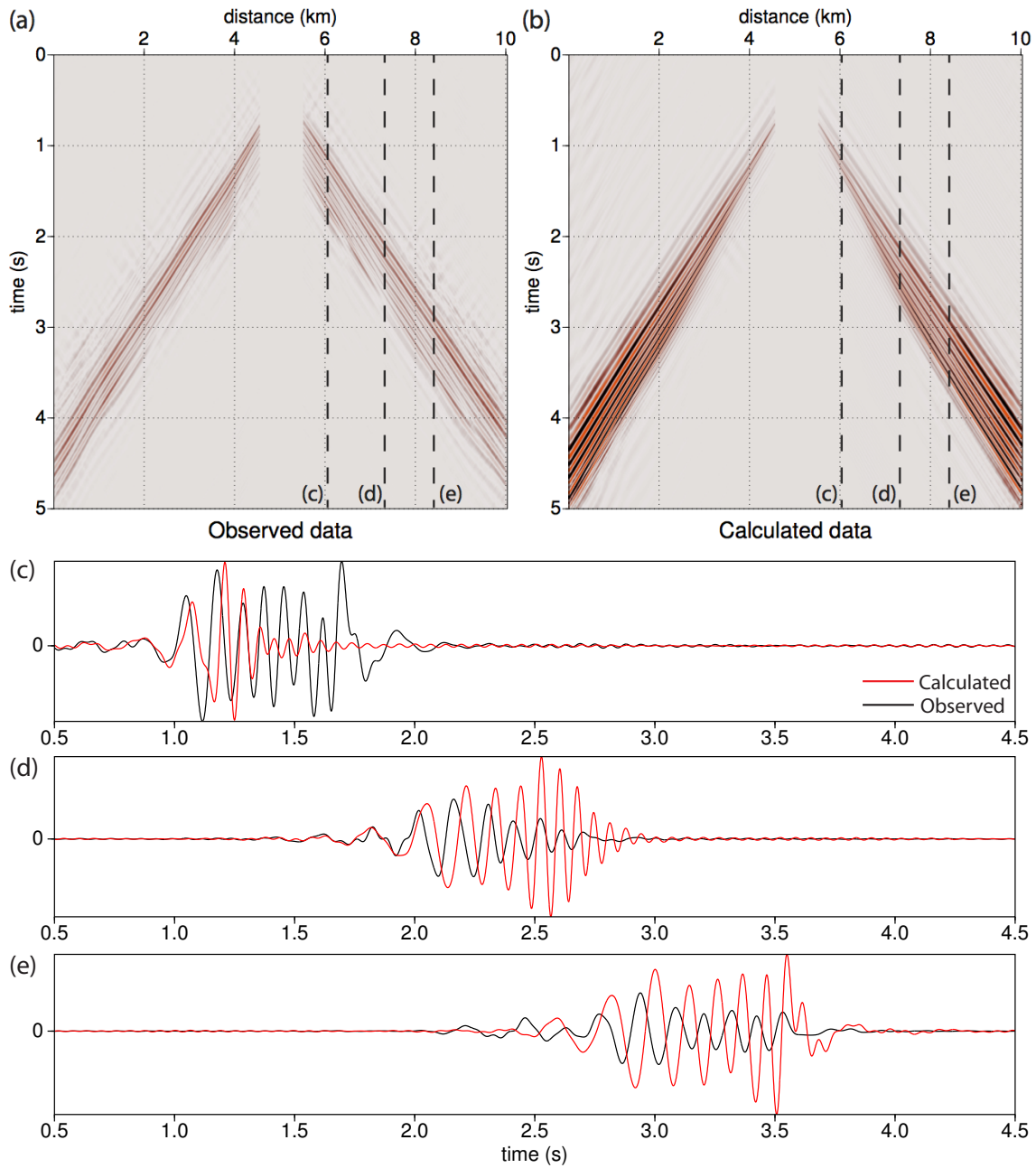


Figure 4.25: The same common shot gather as in Figure 4.24 is shown for observed (a) and calculated (b) data, transformed back to the time-offset (t, x) domain, after a weighting has been applied in the (ω, k) domain. Example traces comparing observed and calculated data, located by the black dashed lines, are shown in (c) to (e). Note that the amplitudes of the observed and calculated data are shown at the same scale.

of topography modelisation (see section 4.4.1).

Multiscale FWI with a misfit function in the frequency-wavenumber (ω, k) domain

is therefore performed. The (ω, k) domain misfit function, presented in section 2.2.4.2 by Equation 2.11 is considered as the most robust option for this real dataset. For low frequency data, the energetic peak in the (ω, k) domain, created by the fundamental mode of the surface waves, is already approximately retrieved by the initial model, suggesting that convergence with FWI could be possible (see Figure 4.22). Different initial frequency bands are tested for inversion, such as 4 – 6 Hz or 4 – 10 Hz.

Another factor that is taken into account for FWI with a real dataset, which has previously not been discussed for inversion of synthetic data examples, is the muting of near offsets. Offsets less than 450 m are removed due to their proximity to the source, and their strong amplitudes (see Figure 4.23). Similarly to the layer stripping strategy presented in section 3.2.2, a maximum offset is also considered, and offsets larger than 4200 m are removed.

A mute in the time-offset (t, x) domain is also applied to focus only on the surface waves and remove the wavetypes that will not be modelled in the synthetic data, such as shown in Figure 4.23.

Only the shear velocity model is inverted during FWI, while the P-wave velocity and density models are kept constant. Differently to layer stripping approaches, the whole depth of the gradient is considered for the model update.

Due to limited computer resources, only the first 200 shot gathers were inverted. As the FWI algorithm used is not intended for production purposes, large memory requirements make repeated runs difficult. This factor is further amplified due to the size of the data and the small discretization steps required for the stability conditions of the algorithm. The reduction of common shot gathers, easily removed due to the parallel nature of the code, allows the algorithm to be run repeatedly for testing.

Unfortunately several FWI tests with different parameter variations resulted in a linesearch failure. This occurs after 20 misfit function and gradient evaluations, when no proper scaling of the descent direction is found. This failure indicates that the computed descent direction is not accurate enough, and underlines the nonlinearity of the misfit function. This result is most likely due to the occurrence of cycle-skipping, and significant differences in amplitude between observed and calculated data created by attenuation effects.

However when a masking in the frequency-wavenumber (ω, k) domain is applied, the FWI is able to update the model at the first iteration. The masking applied is the same as discussed in section 4.2.4.3 and focuses on the fundamental mode information. The

frequency window considered for FWI is 4 – 10 Hz . The updated shear velocity model after FWI is shown in Figure 4.26b.

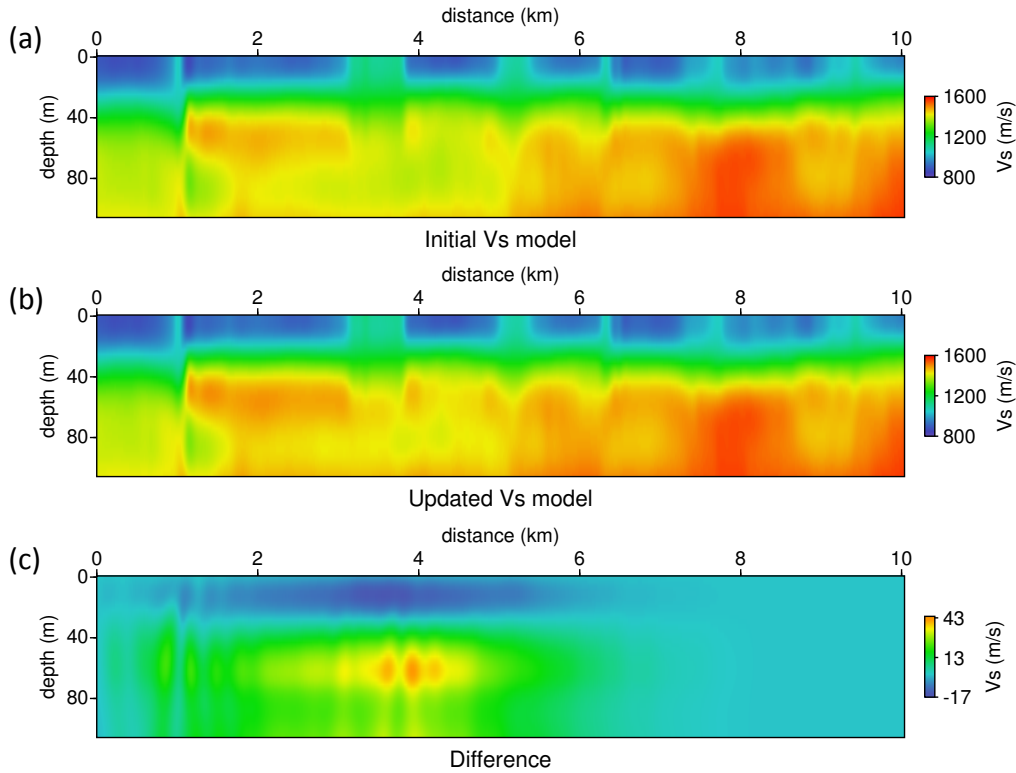


Figure 4.26: Initial shear velocity model (a), updated shear velocity model after FWI (b), and the difference between the two (c).

When compared to the starting model, one can observe that the velocities in the top layer have been reduced, while the velocities at depth have generally been increased (see Figure 4.26c). The difference between the models is zero beyond 6 km from the start of the line, because only the first 200 common shot gathers are considered, and therefore very few data are available to update the right hand side of the model.

The misfit of the data is decreased by 30 percent of the initial misfit value by the FWI. The data modelled using the updated shear velocity model, and low bandpass filtered, is compared to the initial and true data, in both the (t, x) and the (ω, k) domains in Figures 4.27 and 4.28 respectively. Because only the left hand side of the model has been recovered the first common shot gather is analyzed. In the (t, x) domain, one can observe that the FWI has not converged in the right direction, as the updated data trace is further away from the observed than the initial one (see Figures 4.27d and 4.27e).

By looking at the same common shot gather in the (ω, k) domain, which was used

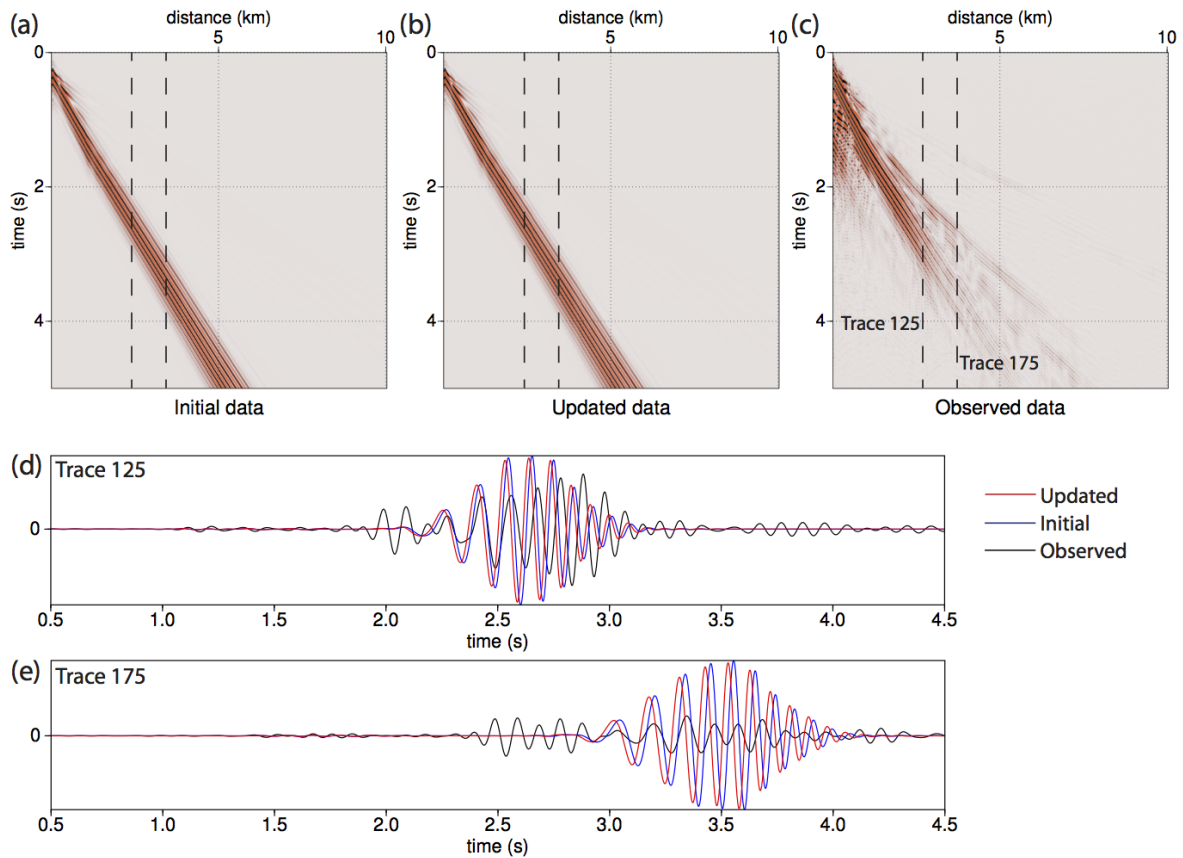


Figure 4.27: Common shot gather filtered between 4 – 15 Hz for the initial data (a), the updated data (b) and the observed data (c). Two example traces, located by the black dashed lines, are illustrated in (d) and (e).

to compute the data misfit, one can understand that the FWI has converged towards a local minimum due to cycle-skipping effects. Figure 4.28d shows how the FWI has shifted the energy peak of the synthetic data to fit a secondary energy peak of the observed data. The main energy peaks of synthetic and observed data, corresponding to their fundamental modes, are now further apart.

Unfortunately, this result is inevitable when the initial velocity models provided to the FWI are not sufficiently accurate. Although the (ω, k) domain misfit function is more robust than the conventional (t, x) domain misfit function, it is not immune to cycle-skipping issues. This is especially evident for this real case scenario where secondary energy peaks corresponding to higher modes are present.

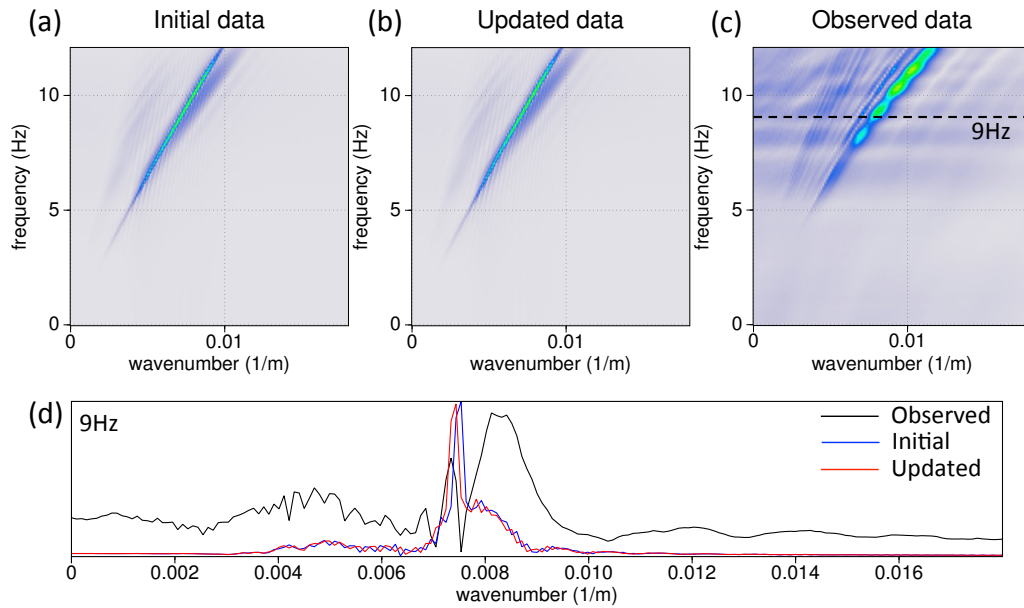


Figure 4.28: The same common shot gather as in Figure 4.27, filtered between 4–15 Hz , is shown in the (ω, k) domain for the initial data (a), the updated data (b) and the observed data (c). An example trace, located by the black dashed line, shows the effects of cycle-skipping in (d).

4.3 Conclusion and evaluation

To conclude, this chapter examines the possibilities and difficulties that may occur when applying FWI with surface waves on a real data application. This case study contains each step required for FWI, starting from the initial model building, and identifies some of the challenges that need to be tackled.

First of all, the dataset used for this study may not have been the ideal example, due to the array acquisition design, the lack of multicomponent receivers, and the topography that could not be taken into consideration by the modelling algorithm used. Furthermore no well data or other non-seismic geophysical prospection information was available for the near surface region to help with the initial model building or data validation.

The step of initial model building illustrates typical problems with dispersion curve picking due to lateral variations, providing the motivation to move towards FWI of surface waves. The results show that very different shear velocity models can be obtained from surface wave 1D inversion (SWODI), for the same dataset. The parameters chosen and the different algorithms available can have a significant impact on the result. It would be interesting to try FWI with modified initial models, for example by integrating

a high velocity layer at the bottom of the model, whose presence is suggested by tests shown in Figure 4.4.

The initial model building investigation also identifies the difficulties that may occur with the absorbing boundary conditions during elastic forward modelling, when high shear velocity contrasts are present in the model close to the boundaries. Solutions for this modelling issue may need to be further investigated for future studies.

The source wavelet was only estimated once for FWI. Ideally it could be further improved once the initial velocity models have been updated with FWI, to provide a more accurate wavelet.

Several mutes in both the time-offset (t, x) domain and in the (ω, k) domain were applied to the data for both analysis purposes and during FWI. More work could be done to better tailor these manipulations on the real data, and perhaps improve the results. Yet the analysis of the model update obtained clearly shows the presence of cycle-skipping effects and convergence towards a local minima. As pointed out from the beginning of this thesis, it is the key issue that needs to be solved for successful FWI with surface waves.

Finally, several aspects of the forward modelling could be improved and perhaps lead to more successful results, as further discussed in the perspectives.

4.4 Perspectives

Several aspects of the forward modelling could be improved and perhaps lead to more successful results. The added value of 3D wave propagation for this dataset, compared to the added computational cost required may be debatable, but features such as topography, attenuation and even improved 3D-to-2D conversion could significantly impact the results at little additional cost.

4.4.1 Dealing with topography

One of the limitations of the forward modelling in the FWI algorithm used in this study, is the assumption of a flat topography. This assumption is constrained by the finite-difference scheme and free surface condition implemented in the algorithm. Complex topographies are difficult to handle due to the regular pattern of finite difference schemes,

and they require a greater computational cost for implementation of free surface boundaries (Gao et al., 2015). The real data investigated in this chapter were however acquired on a line crossing two small canyon features, giving a variation in elevation of up to 30 m (see Figure 4.29).

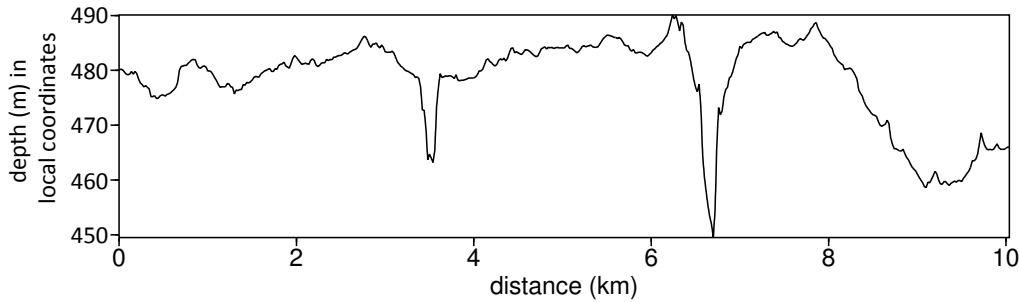


Figure 4.29: Topography profile of the acquired line for the real data considered in Chapter 4, showing the location of the two canyons. Note that the vertical axis is exaggerated.

To evaluate the flat topography assumption, forward modelling tests using a different, more complete algorithm, capable of handling topography, are performed. This new propagator is implemented with a curvilinear scheme, to model elastic wave propagation in the presence of curved topography and a free surface condition (Tarrass et al., 2011), and can be incorporated within an FWI algorithm for future studies. The curvilinear transformation, based on a 1D transformation along the depth axis, allows to propagate the wave equation through the domain (η, ξ) , with the relation between depth z and η following a coordinate transformation given as

$$z = z_{topo}(x) + (\eta - z_0) \frac{(z_{max} - z_{topo}(x))}{(z_{max} - z_0)}, \quad (4.3)$$

$$\eta = z_0 + (z - z_{topo}(x)) \frac{(z_{max} - z_0)}{(z_{max} - z_{topo}(x))}. \quad (4.4)$$

where z_0 is a reference topographic value (Tarrass et al., 2011). This domain transformation is represented in Figure 4.30, where the initial medium is stretched or compressed to obtain a flat surface in the (η, ξ) domain. This curvilinear domain allows to implement an efficient finite differences scheme for the wave propagation and avoid staircase approximations when considering topography.

Using a Ricker of 20 Hz peak frequency as a source wavelet, and the initial models

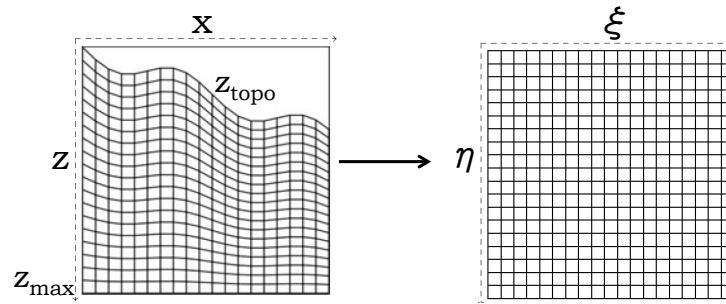


Figure 4.30: Initial Cartesian models with topography are transformed into models with curvilinear coordinates and having a flat topography. Modified from Tarrass et al. (2011).

obtained in section 4.2.1, synthetic data is modeled using the curvilinear propagator, with and without the true topography applied. The shear velocity models used are shown in Figure 4.31 with and without the true topography applied.

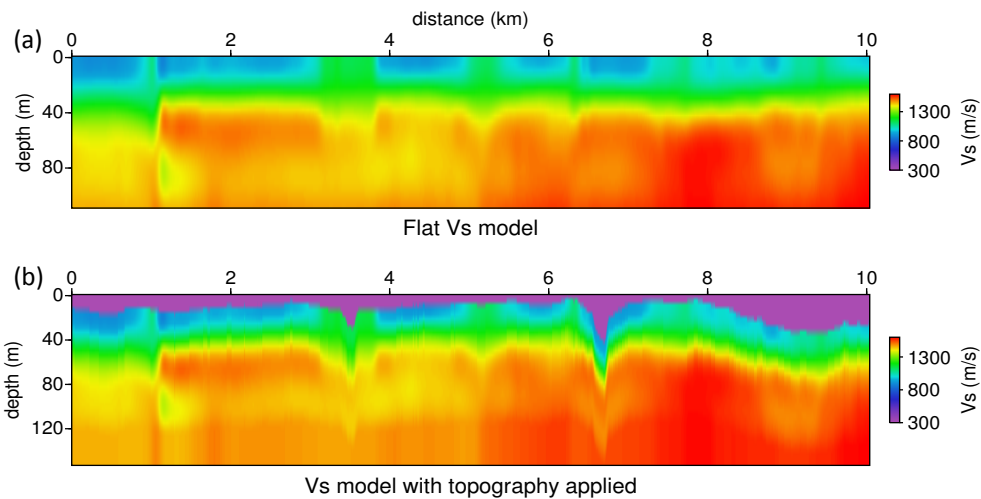


Figure 4.31: Shear velocity model without (a) and with (b) topography, used for forward modelling with the curvilinear propagator. Note that the vertical axis is exaggerated.

Example common shot gathers for one source at the beginning of the line and for another in the middle are shown in Figures 4.32a-4.32d. A low frequency bandpass of 4 – 15 Hz is applied to the data for the comparison. In Figure 4.32c one can observe backscattering occurring at a distance of about 3.5 km , at the location of the first canyon in the topography profile, and at other locations to a lesser extent. Backscattered surface waves can be clearly observed to originate at both canyon locations in Figure 4.32d.

For the observed data considered in this chapter, backscattering due to the presence of the canyons is only weakly visible. An example common shot gather where it can

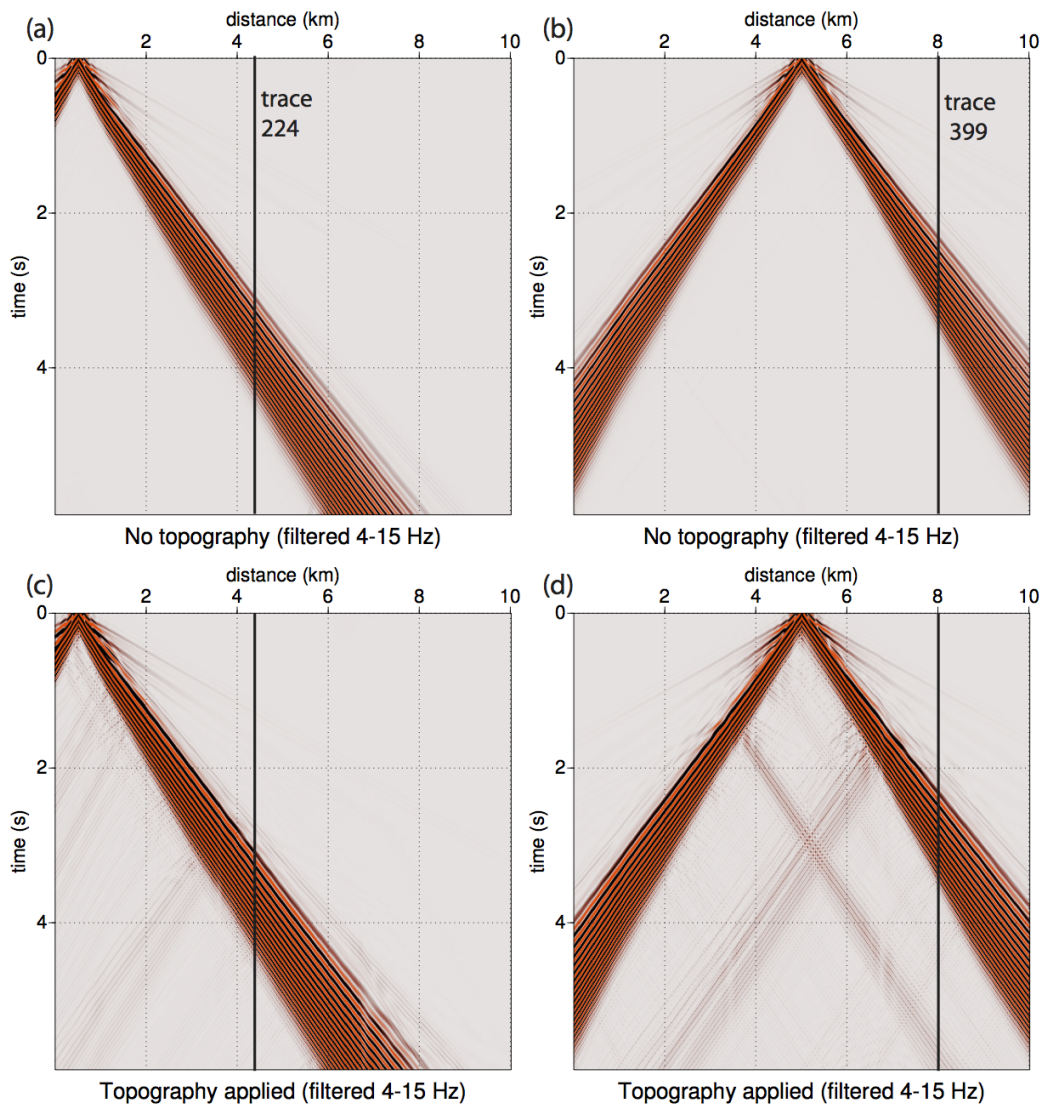


Figure 4.32: Example common shot gathers for data modeled without topography (a-b) and with topography (c-d) using the curvilinear propagator. The data are low bandpass filtered between 4 – 15 Hz . A comparison of individual traces, located by the solid black line is shown in Figure 4.34.

be observed is shown in Figure 4.33. Although often a weak signal in real data, correct modelling of backscattering effects may help to better localize the position of velocity anomalies, and improve inversion results (Schwenk et al., 2014). The ability to take into account topography variations in the forward modelling scheme used by the FWI algorithm is therefore a very interesting perspective for future studies.

Individual traces of the modeled data for the low frequency band of 4 – 15 Hz are compared in Figures 4.34a-4.34b. The chosen traces are located directly after where the

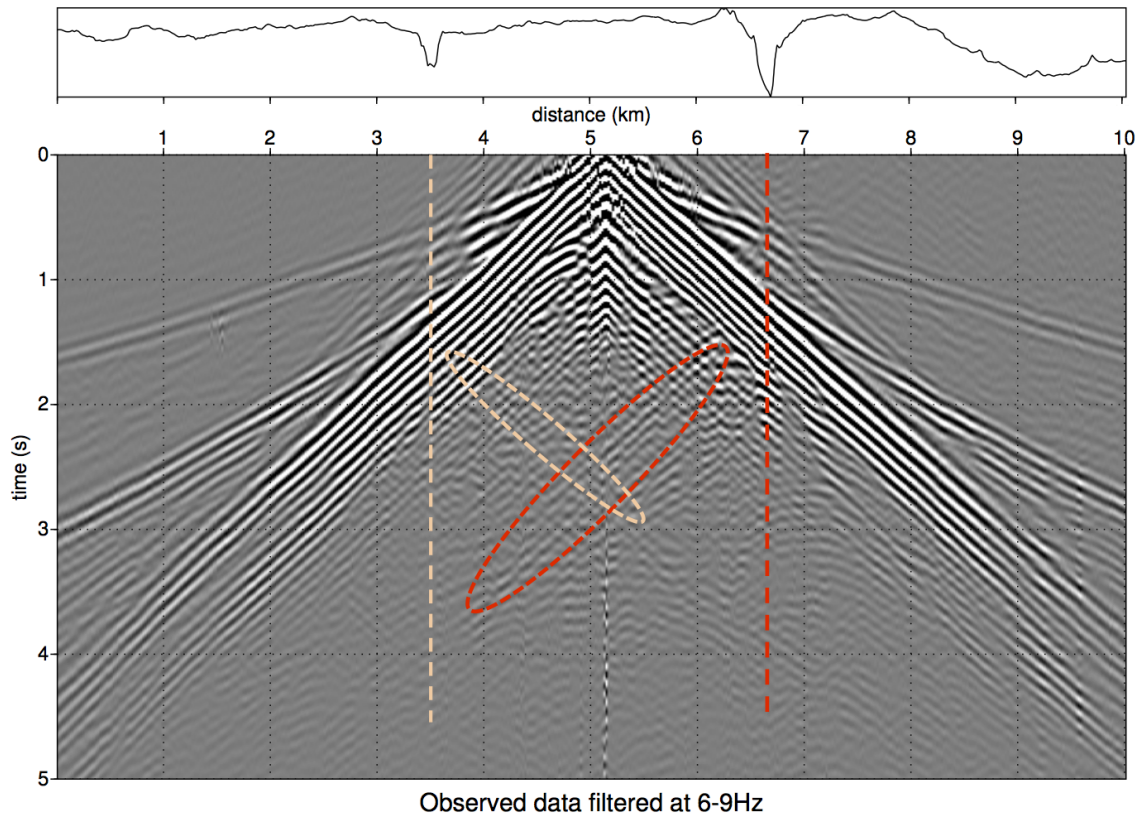


Figure 4.33: Example common shot gather of the observed data with a bandpass filter of $6 - 9 \text{ Hz}$ applied, where a weak backscattered signal (circled) created by the presence of two canyons (indicated by the dashed lines) can be observed. The topography profile is shown above for reference.

surface waves pass through the two canyon features. Although differences in amplitude are evident, the phases appear to match well, suggesting that the present topography does not create cycle-skipping effects at this frequency scale. However when the low bandpass filter is widened to $4 - 27 \text{ Hz}$, and higher frequencies are considered, the trace comparison matches less well as shown in Figures 4.34c-4.34d. Although the main phases are not cycle-skipped, small phase differences are present in the traces, as highlighted by the green circles, and the assumption of a flat topography may not be valid. To apply layer stripping FWI strategies that start with the inversion of high frequency data, topography would need to be taken into account in the forward modelling for a successful inversion.

This analysis mainly validates the flat topography assumption for the case study considered in this chapter. However, considering a forward modelling propagator capable of handling topography effects may be much more significant for other real data cases,

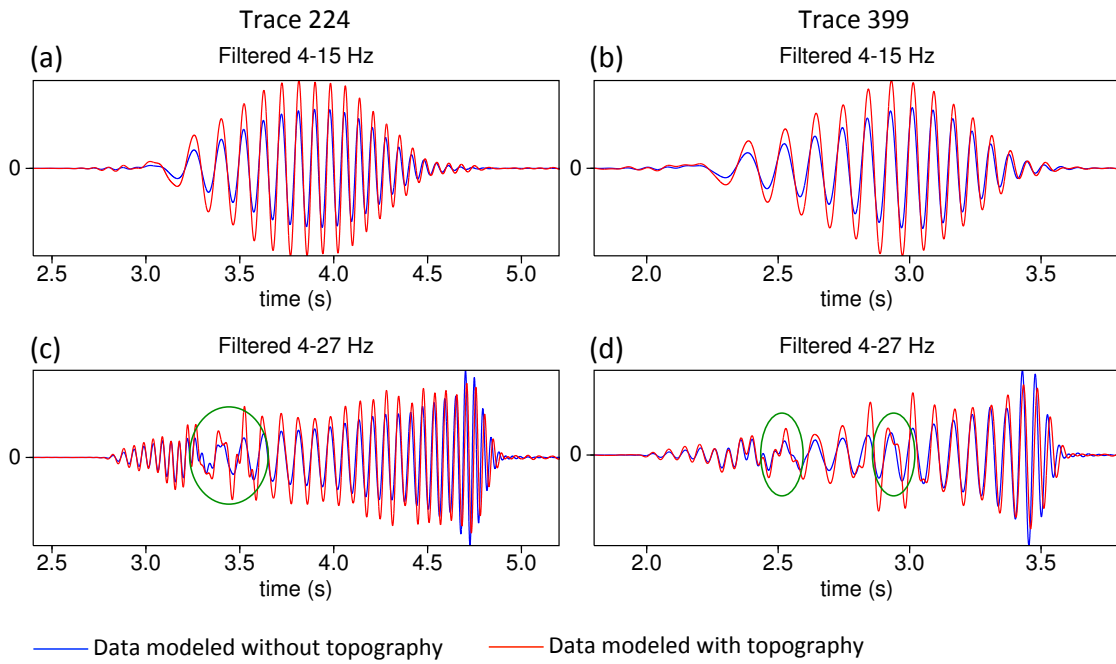


Figure 4.34: Comparison of individual traces, located by the solid black lines in Figure 4.32, for data filtered between 4 – 15 Hz in (a-b) and for data filtered between 4 – 27 Hz in (c-d). The blue line represents data modelled without topography, and the red line represents data modelled with topography. The green circles highlight phase differences visible in the higher frequency band.

and is an important area of research for future studies.

Conclusions and Perspectives

Conclusions

This study assumes that surface waves, conventionally treated as noise by the oil and gas community, can potentially be used to characterize the near surface on the exploration scale. A better understanding of this shallow zone, and a potential improvement for imaging and interpretation of deeper targets, at no extra exploration cost, is an important resource for the oil and gas industry.

The investigation presented in this study, goes beyond the limitations of conventional surface wave inversion, based on a layered model assumption, and tackles the challenges of Full Waveform Inversion (FWI) with surface waves. Although more commonly applied to body waves, FWI allows to recover lateral velocity variations with high resolution capabilities.

The first part of this study uses synthetic examples to develop and adapt the conventional FWI method to include surface wave data. In Chapter 2 it is shown that misfit functions where the data-fitting is computed in different data domains, such as the frequency-wavenumber ($\omega - k$) and frequency-slowness ($\omega - p$) domains, are more robust alternatives to conventional FWI. These robust misfit functions can overcome cycle-skipping limitations that often occur when dealing with surface waves. These misfit functions are shown to have more convex valleys of attraction and reduced local minima when considering the topography of the data misfit for two model parameters in a grid analysis compared to the conventional misfit function in the time-offset ($t - x$) domain. Successful inversion results are obtained with these robust misfit functions on a synthetic dataset to retrieve lateral variations in shear velocity.

Different strategies for FWI with surface waves are investigated and proposed in Chapter 3. Synthetic data examples are used to promote a layer stripping approach specifically adapted to the physics of surface waves for FWI. Different to previous literature on FWI,

high-to-low frequencies are inverted, combined with short to long offsets, to sequentially update top to bottom depth layers. This innovative layer stripping strategy for FWI is shown to provide better results than those obtained with conventional multiscale FWI. Furthermore, unlike multiscale approaches, layer stripping FWI does not rely on the presence of low frequency data, often missing from real data acquisitions, to converge. The layer stripping method is able to target the surface waves to drive the FWI, giving evidence that surface waves can update the shear velocity model and help characterize the near surface region.

In the final part of this study, a real data application is presented, illustrating the difficulties involved and the possible near surface characterization achieved with surface waves. Initial velocity models are obtained from first break travel time inversion and Surface Wave 1D Inversion (SWODI). The analysis of the dispersion curve picking clearly shows the laterally invariant limitations of SWODI, a clear motivation for moving to FWI with surface waves. The analysis comparing the synthetic data and the observed data highlights the potential cycle-skipping problems, and the significant propagation differences, such as the lack of attenuation in the synthetic data. The misfit function in the $(\omega - k)$ domain is shown as a more robust option compared to conventional FWI, but cycle-skipping features prevent the FWI from converging in the right direction. This final chapter leads to several suggestions for improvements in both forward modelling and FWI strategies, as discussed in the perspectives.

Although the real data case study suggests that further research is still necessary, the work presented in this study contributes to the truth behind the words of [Dobrin \(1950\)](#): *“surface waves can be made a useful by-product rather than a waste product of seismic exploration”*.

Perspectives

Full Waveform Inversion (FWI) is currently a subject of strong interest and extensive research for the oil and gas industry. Moving from acoustic to elastic and to visco-elastic modelling, from 2D to 3D and 4D acquisitions, from synthetic case studies to real industry datasets, are some examples of challenges that are currently tackled.

Forward modelling

As larger, more powerful computers become available, the capabilities of forward modelling techniques may become more enhanced. Sufficiently accurate modelling of an elastic 3D medium, considering anisotropy and attenuation effects, and with a free surface condition following a complex topography profile, is currently a very important and active subject of research and investigation.

One perspective to extend the results presented in this study, is to improve the forward modelling of surface waves. Propagators other than finite differences (implemented in this study) may be more adapted to dealing with the simulation of a complex topography and a curved free surface condition. Of course regions with topography are also those where an understanding of the near surface would be most beneficial to depth imaging exploration projects. The added value of being able to model topography, for example by implementing a curvilinear propagator (Tarrass et al., 2011), are discussed in section 4.4.1 for the real data case study analyzed in Chapter 4. Other options such as the discontinuous Galerkin method may also be a potential approach.

Beyond the different propagators available, the implementation of different absorbing boundary conditions or free surface conditions may also be investigated. For example numerical artifacts are often encountered due to the PML (Perfectly Matched Layers) boundary condition for elastic propagation at the oil and gas exploration scale, when significant variations in the Poisson ratio are present close to the edges of the velocity models used. Such artifacts, as encountered during the real data case study in Chapter 4, are difficult to resolve.

From the difficulties that occur when dealing with a real dataset in Chapter 4, it is also very clear that attenuation is an important parameter that should be included in further studies of surface wave inversion. Significant differences in amplitude between observed and calculated data, especially a longer offsets, are clearly visible. Both intrinsic attenuation and the attenuation due to scattering should be taken into account. The importance of attenuation has already been underlined by studies of surface wave inversion at the geotechnical and engineering scale (Lai, 1998; Groos et al., 2014). Groos et al. (2014) compares data modelled using elastic and viscoelastic propagators, showing significant distance dependent and frequency dependent damping differences that impact FWI results. Retrieving an attenuation parameter from inversion may be difficult, but implementing viscoelastic propagation with a fixed damping parameter, could be critical for future studies of surface waves.

An extension to 3D may also be a future perspective, as computational availability make 3D elastic FWI a reality. The methodology developed by this study in 2D may be extended to 3D without major difficulties. Furthermore 3D wave propagation would result in better modelling of amplitudes, avoiding 3D-to-2D conversion issues.

Surface wave inversion strategy

The results presented in Chapter 3 using a synthetic data example promote a layer stripping strategy for FWI with surface waves. A direct extension of this work could investigate the combined use of both layer stripping strategies with conventional multi-scale approaches, for example in a V-cycle workflow, taking advantage of the different strengths of both strategies to slowly but consistently build a velocity model. Layer stripping FWI would help recover the top layers of the velocity model, while the multi-scale FWI would improve the overall resolution when moving from one frequency band to the next.

The synthetic results obtained in Chapter 3 using layer stripping FWI highlight that surface wave information can help retrieve the shear wave velocity model much better than the P-wave velocity model. While this allows a certain insensitivity to the initial P-wave velocity model, this parameter is only weakly updated with each iteration. A possible extension of this study would be to update the P-wave velocity model at each iteration based on the shear-velocity model and assuming a known Poisson ratio. To implement this, the gradient term would need to be modified as the P-wave velocity parameter would depend on the shear wave velocity, so that after each update of the shear-wave velocity model during inversion, the P-wave velocity model can also be updated.

Another possible extension to the work presented in Chapter 3 is to further adapt the depth preconditioning implemented to the physics of surface waves. Instead of applying a vertical scaling linear with depth, an exponential depth preconditioning could be tested to compensate for the exponential decrease of surface wave amplitudes with depth. The preconditioning could also be made dependent on the frequency band considered during each consecutive step in layer stripping FWI, as surface wave amplitudes decrease differently at different frequencies. This could further improve the recovery of deep layers, imaged in the last stages of layer stripping FWI.

Finally, for the real data application in Chapter 4, a masking in the frequency-wavenumber (ω, k) domain is tested when performing an analysis on the data. Masking

data in (ω, k) domain could be used during FWI with surface waves, in a multiscale approach, allowing first to focus only on the fundamental mode, and gradually trying to invert higher mode information. Investigating different ways to create such masks could also be very interesting, and could be inspired by available semi-automatic picking tools used for picking dispersion curves for conventional surface wave analysis.

As a final remark for future studies on FWI with surface waves at the oil and gas exploration scale, it is also important that real datasets are acquired for the purpose of near surface characterization, and that the surface waves recorded are not destroyed by pre-processing procedures on the field, such as with the use of arrays. Multicomponent acquisitions may also be beneficial for surface wave inversion. Furthermore well information for the shallow near surface zone would also be useful to constrain and validate results. It is also interesting to note that ideally non-seismic methods that also sample the near surface could provide additional information and potentially be used in a joint inversion to better characterize the near surface region. In the future, adequate integration of near surface characterization studies within exploration projects, for example through statics, will also have to be investigated.

Appendix A

Complementary Publications

Alternative misfit functions for FWI applied to surface waves

I. Masoni, R. Brossier, J. Virieux, and J-L. Boelle

2013, *EAGE Technical Program Expanded Abstracts*, Th P10 13,

Summary

The aim of this study is to determine the advantages and limitations of different misfit functions for an application of Full Waveform Inversion (FWI) to surface waves. The difference-based L2 norm, classically used in FWI and sensitive to both amplitude and phase information, suffers from cycle-skipping and local minima. For slow surface waves propagating in the low velocity near surface, the problem of cycle-skipping is even greater due to their small wavelengths. In the absence of low frequencies, convergence may not be possible when starting from a smooth initial mode. Alternative misfit functions applied in various data domains are therefore investigated with the aim of overcoming this issue. Taking the difference-based L2 norm as a basis for comparison, simple synthetic tests are conducted to evaluate a weighted cross-correlation and a singular value decomposition approach as alternative misfit functions, as well as investigating the effect of calculating the residual in different data domains such as the $(\omega - k)$, $(\tau - p)$ and $(\omega - p)$ domains.

Introduction

The construction of subsurface velocity models is a central problem for oil & gas exploration. As well known prolific hydrocarbon basins have been explored and produced, it is now necessary to investigate more complex regions, such as on-shore foothills. In such cases, the heterogeneous near-surface has a great impact on the complexity of seismic propagation and can obstruct the imaging of deeper targets.

Although conventionally viewed as coherent noise or "ground roll", surface waves sample the near subsurface and can be used for imaging. In civil engineering, dispersion curve inversion allows imaging the first tens of meters. However this approach relies on a 1D assumption and only smooth lateral heterogeneities are tolerated. An alternative method may be Full Waveform Inversion (FWI), that extends beyond 1D limitations and avoids time picking or dispersion analysis.

Classical FWI

FWI is a high resolution technique used to derive quantitative models of the subsurface by matching the full observed seismogram with a corresponding synthetic seismogram calculated from a velocity model, and solving a local optimization problem. The L_2 norm of the difference is conventionally used to calculate the misfit (Tarantola, 1984), fitting both the amplitude and the phase of the waveforms:

$$C_{diff} = \sum_t \sum_x \frac{1}{2} (\mathbf{d}_{obs}(t,x) - \mathbf{d}_{cal}(t,x))^2, \quad (1)$$

where $\mathbf{d}_{obs}(t,x)$ is the measured data and $\mathbf{d}_{cal}(t,x)$ is the calculated data recorded at time t and offset x . As the misfit function is minimized in a least-squares sense, the model is iteratively updated with a gradient-based descent method until a minimum is reached (Virieux and Operto, 2009).

By exploiting the full data content and using a strict data-matching approach, this method appears to be very sensitive and may not be robust. Non-linearities, such as cycle skipping, can reduce the convexity of the misfit function (Bunks et al., 1995; Mulder and Plessix, 2008) and the minimization may get stuck in a local minimum. In the absence of very low-frequency data, the initial velocity model needs to explain the data to within half a wavelength, so that it lies within the small basin of attraction of the global minimum and can converge. For slow surface waves propagating in the low velocity near surface, the problem of cycle-skipping is even greater due to their small wavelengths.

Synthetic datasets were created using a discrete wavenumber summation method (Bouchon and Aki, 1977), for horizontally layered media with a free surface, and simulating 3D elastic wave propagation with a Ricker wavelet source of 10 Hz peak frequency. Figure 1a shows the two-layer model used to create the "observed" dataset to which random Gaussian noise is added (Figure 1b). A grid analysis is performed on the S-velocity and the depth of the first layer to investigate the accuracy required for the initial model. Even for this simple framework and only small shifts in the model parameters (example in Figure 1c), the grid analysis result for the classical difference-based L_2 norm approach (Figure 2a) contains many local minima due to the high amplitude of the surface waves that dominate the misfit.

In this study, alternative, more robust, misfit functions applied in various data domains are investigated to improve the convexity of the valley of attraction and reduce the presence of local minima. To evaluate the misfit functions, grid analysis results for the same synthetic test are compared.

Alternative misfit functions

Current solutions to calculating the misfit more robustly are based on other norms such as the hybrid L_1/L_2 or Huber norm (Brossier et al., 2010; Guitton and Symes, 2003) or on zero-lag cross-correlation (Routh et al., 2011), but these also suffer from cycle-skipping in the absence of low frequencies.

A weighted cross-correlation proposed by Van Leeuwen and Mulder (2008) is investigated here as a more robust alternative to the classical difference-based L_2 norm. The misfit is given by a cross correlation on the time axis of the observed and calculated data where events are separated by arrival times.

$$C_{W_i} = \sum_{\Delta t} \sum_x \frac{1}{2} \left(W_i(\Delta t) \sum_t \mathbf{d}_{obs}(t + \Delta t, x) \mathbf{d}_{cal}(t, x) \right)^2. \quad (2)$$

The weighting $W_i(\Delta t)$ is applied to each time sample. Two weightings are tested in this study. The first $W_1(\Delta t) = \Delta t / \Delta t_{max}$, linearly penalizes with distance away from zero lag. The second $W_2(\Delta t) = e^{-\alpha \Delta t^2}$ is a Gaussian weighting to maximize zero-lag energy and with a width controlled by the α parameter, giving a misfit function whose negative is minimized. An appropriate width needs to be chosen to at least be in the order of the length of the wavelet, since it can greatly influence the convexity of the misfit function (Van Leeuwen and Mulder, 2008).

The grid analysis result for the penalized version of the cross-correlation (Figure 2b) illustrates how it is highly sensitive to noise at large lags and not robust enough. Instead the sensitivity of the Gaussian weighted cross-correlation can be better tuned to obtain a convex result with no local minima (Figure 2c) allowing convergence from an initial model further away from the true one. The weighting applied to the cross-correlation is therefore critical for a stable misfit function.

The weighted cross-correlation is however not sensitive to the frequency and phase rotation of an event. Due to the dispersive property of surface waves, the frequency may contain key information on the depth of the signal and may need to be identified. Furthermore "cross-talk" may occur for multiple arrivals. Therefore coupling this misfit function with a strategy to separate arrivals, such as comparing data in a different domain, needs to be considered.

For comparison, a recently proposed misfit function based on a singular-value decomposition (SVD) approach (Moghaddam and Mulder, 2012) is also tested. An SVD is applied to data matrix $\mathbf{A}(\omega)$, of size number of receivers by number of sources, for each frequency so that $\mathbf{A}_{obs}(\omega) = \mathbf{U}_{obs} \mathbf{S}_{obs} \mathbf{V}_{obs}^H$. When the calculated data is equal to the observed data then the matrix $\mathbf{S}(\omega) = \mathbf{U}_{obs}^H \mathbf{A}_{cal}(\omega) \mathbf{V}_{obs}$, will be diagonal. A weighting W_{ij} is applied to the misfit function to linearly penalize the off-diagonal values:

$$C_{SVD} = \sum_{\omega} \sum_i \sum_j \frac{1}{2} (W_{ij} S_{ij}(\omega))^2. \quad (3)$$

The grid analysis result (Figure 2d) shows a larger basin of attraction than the classical difference-based L_2 norm, but sensitivity is lacking to ensure convergence for high S-velocities and small layer depths.

Alternative data domains

The data domain in which observed and calculated datasets are compared also affects the sensitivity of the misfit function. Perez Solano et al. (2012) proposes the (ω, k) domain, to reduce the presence of local minima for the difference-based L_2 norm (Eq 4). A weighted cross-correlation of the modulus of the (ω, k) data is applied on the wavenumber k -axis (Eq 5):

$$C_{diff} = \sum_{\omega} \sum_k \frac{1}{2} (|\mathbf{d}_{obs}(\omega, k)| - |\mathbf{d}_{cal}(\omega, k)|)^2, \quad (4)$$

$$C_{W_i} = \sum_{\omega} \sum_{\Delta k} \frac{1}{2} \left(W_i(\Delta k) \sum_k |\mathbf{d}_{obs}(\omega, k + \Delta k)| |\mathbf{d}_{cal}(\omega, k)| \right)^2. \quad (5)$$

The (τ, p) domain or "slant-stack" is also investigated. In this domain, data have undergone a linear move-out (LMO) correction, and are summed over the offset axis. This is done for a range of slowness p values. Separating events by their slowness may reduce the cycle-skipping problem, and stacking may also make the misfit function more robust in the presence of noise. Both the difference-based L_2 norm (Eq 6) and a weighted cross-correlation applied on the slowness p axis (Eq 7) are considered:

$$C_{diff} = \sum_{\tau} \sum_p \frac{1}{2} (\mathbf{d}_{obs}(\tau, p) - \mathbf{d}_{cal}(\tau, p))^2, \quad (6)$$

$$C_{W_i} = \sum_{\tau} \sum_{\Delta p} \frac{1}{2} \left(W_i(\Delta p) \sum_p \mathbf{d}_{obs}(\tau, p + \Delta p) \mathbf{d}_{cal}(\tau, p) \right)^2. \quad (7)$$

Finally, the (ω, p) domain, equivalent to the (ω, k) domain but with a different sampling, may also help separate events by their slowness p as well as identify their frequency ω , which can be helpful to use the dispersive property of surface waves. Again the difference-based L_2 norm (Eq 8) and the weighted cross-correlation (Eq 9) misfit functions are tested:

$$C_{diff} = \sum_{\omega} \sum_p \frac{1}{2} (|\mathbf{d}_{obs}(\omega, p)| - |\mathbf{d}_{cal}(\omega, p)|)^2, \quad (8)$$

$$C_{w_i} = \sum_{\omega} \sum_{\Delta p} \frac{1}{2} \left(W_i(\Delta p) \sum_p |\mathbf{d}_{obs}(\omega, p + \Delta p)| |\mathbf{d}_{cal}(\omega, p)| \right)^2. \quad (9)$$

The difference-based L_2 norm becomes more convex in all alternative data domains tested as shown by Figures 2(e,h,k). Convergence is especially successful in the (ω, k) and (ω, p) domains. The ω domain appears to efficiently mitigate non-linearities related to dispersive effects. Where local minima are no longer present, it may be possible to start with an initial model far from the true one.

On the other hand, the penalized cross-correlation is not very successful. Noise dominates the misfit in the (ω, k) domain (Figure 2f), and the global minimum is no longer at the true S-velocity and depth of the layer for the other tested domains (Figures 2i, 2l). Yet when a gaussian weighting is applied, the global minimum is correctly centered and the convexity can be tuned by the α parameter (Figures 2j, 2m). Only in the (ω, k) domain (Figure 2g), some local minima remain present even with an optimized weighting, limiting convergence.

Conclusions

We have used numerical tests to compare alternative FWI misfit functions for surface wave applications. Both previously proposed SVD (Moghaddam and Mulder, 2012), and difference-based (ω, k) domain (Perez Solano et al., 2012) misfit functions are validated as improvements to classical difference-based FWI. Furthermore the difference-based (ω, p) domain approach as well as a Gaussian weighted cross-correlation in the (t, x) , (τ, p) and (ω, p) domains are also shown to be promising alternatives. Cross-correlations on other axis or double cross-correlations (Van Leeuwen and Mulder, 2008) could also be tested. Our future work will investigate the robustness of these misfit functions on laterally varying models.

Acknowledgements

The authors would like to thank TOTAL E&P for permission to show these results. This work was produced using the CIMENT high-performance computing facilities (Université Joseph Fourier, Grenoble).

References

- Bouchon, M. and Aki, K. [1977] Discrete wave-number representation of seismic-source wave fields. *Bulletin of the Seismological Society of America*, **67**(2), 259–277.
- Brossier, R., Operto, S. and Virieux, J. [2010] Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, **75**(3), R37–R46, doi:10.1190/1.3379323.
- Bunks, C., Salek, F.M., Zaleski, S. and Chavent, G. [1995] Multiscale seismic waveform inversion. *Geophysics*, **60**(5), 1457–1473.
- Guitton, A. and Symes, W.W. [2003] Robust inversion of seismic data using the Huber norm. *Geophysics*, **68**(4), 1310–1319, doi:10.1190/1.1598124.
- Moghaddam, P. and Mulder, W. [2012] The Diagonalator, an alternative cost functional for wave-equation inversion. *Expanded Abstracts, 74th Annual meeting, EAGE, W022*.
- Mulder, W. and Plessix, R.E. [2008] Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, **56**(6), 827–841.
- Perez Solano, C., Donno, D. and Chauris, H. [2012] Alternative objective function for inversion of surface waves in 2D media. *Expanded Abstracts, 18th European Meeting of Environmental and Engineering Geophysics, Near Surface Geoscience, A21*.
- Routh, P. et al. [2011] Encoded simultaneous source full-wavefield inversion for spectrally shaped marine streamer data. *SEG Technical Program Expanded Abstracts*, **30**(1), 2433–2438, doi:10.1190/1.3627697.
- Tarantola, A. [1984] Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, **49**(8), 1259–1266.
- Van Leeuwen, T. and Mulder, W. [2008] Velocity analysis based on data correlation. *Geophysical Prospecting*, **56**(6), 791–803, ISSN 1365-2478, doi:10.1111/j.1365-2478.2008.00704.x.
- Virieux, J. and Operto, S. [2009] An overview of full waveform inversion in exploration geophysics. *Geophysics*, **74**(6), WCC127–WCC152.

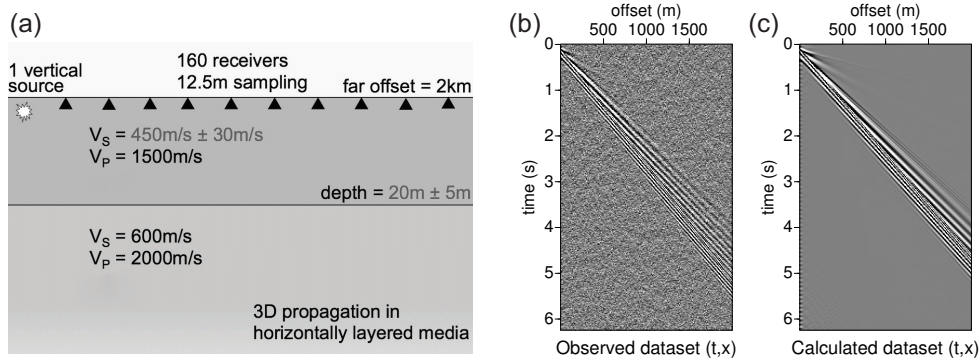


Figure 1 Two-layer model (a) used to create the observed dataset with added random Gaussian noise (b), and an example of the calculated dataset for a layer S-velocity at 480 m/s and depth of 15 m (c).

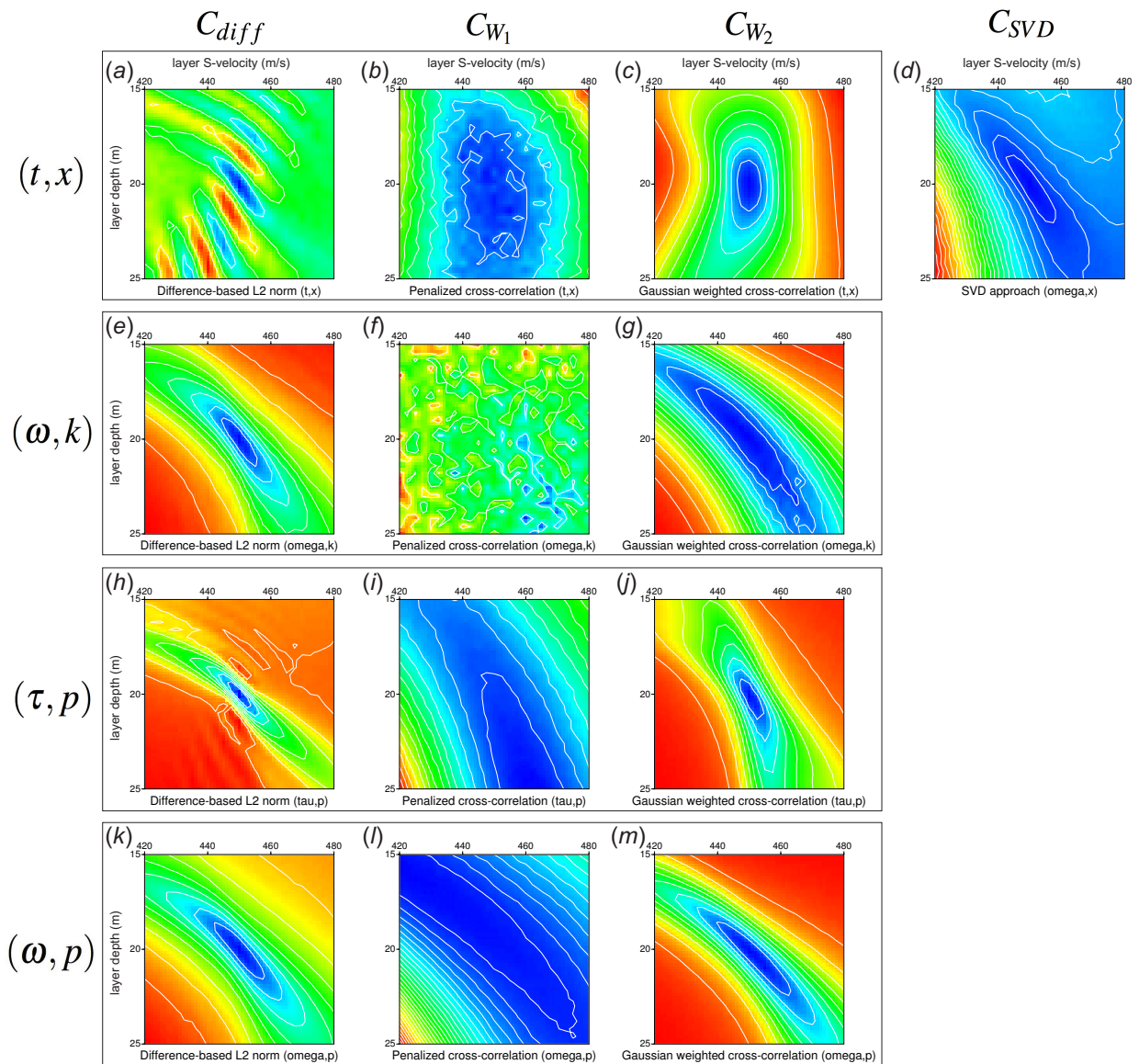


Figure 2 Two-parameter grid analysis for all misfit functions tested in this study (a-m), for an observed dataset with the true global minimum at 450 m/s layer S-velocity and 20 m layer depth.

Bibliography

- Aki, K. and Richards, P. (1980). *Quantitative Seismology: Theory and Methods*. W. H. Freeman & Co, San Francisco.
- Aki, K. and Richards, P. G. (2002). *Quantitative seismology, theory and methods, second edition*. University Science Books, Sausalito, California.
- Alkhalifah, T. (2013). Research note: Full-waveform inversion of the unwrapped phase of a model. *Geophysical Prospecting*, 62(2):397–403.
- Aravkin, A., van Leeuwen, T., and Herrmann, F. (2011). Robust full-waveform inversion using the Student's t-distribution. In *SEG Technical Program Expanded Abstracts*, volume 30, pages 2669–2673. SEG.
- Babuska, V. and Cara, M. (1991). *Seismic Anisotropy in the Earth*. Boston: Kluwer Academic Publishers.
- Beaty, K., Schmitt, D., and Sacchi, M. (2002). Simulated annealing inversion of multimodal Rayleigh wave dispersion curves for geological structure. *Geophysical Journal International*, 151:622–631.
- Bednar, J. B., Shin, C., and Pyun, S. (2007). Comparison of waveform inversion, part 2: phase approach. *Geophysical Prospecting*, 55(4):465–475.
- Bérenger, J.-P. (1994). A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.
- Bergamo, P., Boiero, D., and Socco, L. V. (2012). Retrieving 2d structures from surface-wave data by means of space-varying spatial windowing. *Geophysics*, 77(4):EN39–EN51.
- Bian, A. and Yu, W. (2011). Layer-stripping full waveform inversion with damped seismic reflection data. *Journal of Earth Science*, 22(2):241–249.

BIBLIOGRAPHY

- Bodin, T. and Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, 178(3):1411–1436.
- Bohlen, T., Kugler, S., Klein, G., and Theilen, F. (2004). 1.5D inversion of lateral variation of Sholte-wave dispersion. *Geophysics*, 69:330–344.
- Bohlen, T. and Saenger, E. H. (2006). Accuracy of heterogeneous staggered-grid finite-difference modeling of Rayleigh waves. *Geophysics*, 71:109–115.
- Boiero, D. and Socco, L. V. (2010). Retrieving lateral variations from surface wave dispersion curves. *Geophysical Prospecting*, 58:977–996.
- Born, M. and Wolf, E. (1980). *Principles of optics*. Pergamon Press, 6th edition.
- Bouchon, M. and Aki, K. (1977). Discrete wave-number representation of seismic-source wave fields. *Bulletin of the Seismological Society of America*, 67(2):259–277.
- Braile, L. (2010). *Seismic waves and the slinky: a guide for teachers*.
- Bretaudeau, F., Brossier, R., Leparoux, D., Abraham, O., and Virieux, J. (2013). 2D elastic full waveform imaging of the near surface: Application to synthetic and a physical modelling data sets. *Near Surface Geophysics*, 11:307–316.
- Brossier, R., Operto, S., and Virieux, J. (2009). Seismic imaging of complex on-shore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118.
- Brossier, R., Operto, S., and Virieux, J. (2010). Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75(3):R37–R46.
- Brossier, R., Operto, S., and Virieux, J. (2015). Velocity model building from seismic reflection data by full waveform inversion. *Geophysical Prospecting*, 63:354–367.
- Brown, L., Diehl, J., and Nigbor, R. (2000). A simplified procedure to measure average shear-wave velocity to a depth of 30 meters (V_{s30}). In *Proceedings of 12th World Conference on Earthquake Engineering*.
- Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.

- Bözdag, E., Trampert, J., and Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2):845–870.
- Burridge, R. (1996). Elastic waves in anisotropic media. Schlumberger-Doll Research.
- Byrd, R. H., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208.
- Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418.
- Cara, M. (1973). Filtering of dispersed wave trains. *Geophysical Journal of Royal Astronomy Society*, 33:65–80.
- Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):2117–2131.
- Chavent, G. (1974). Identification of parameter distributed systems. In Goodson, R. and Polis, M., editors, *Identification of function parameters in partial differential equations*, pages 31–48. American Society of Mechanical Engineers, New York.
- Chavent, G. (2009). *Nonlinear least squares for inverse problems*. Springer Dordrecht Heidelberg London New York.
- Choi and Alkhalifah, T. (2012). Multi-source waveform inversion of marine streamer data using the normalized wavefield. In *Expanded Abstracts*. EAGE.
- Coates, R. T. and Chapman, C. H. (1990). Ray perturbation theory and the Born approximation. *Geophysical Journal International*, 100:379–392.
- Courant, R., Friedrichs, O., and Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal (translated from Courant et al, 1928)*, 11:215–234.
- Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic non-linear waveform inversion: application to real data. *Geophysics*, 55:527–538.
- Djikpéssé, H. A. and Tarantola, A. (1999). Multiparameter l_1 norm waveform fitting: Interpretation of gulf of mexico reflection seismograms. *Geophysics*, 64(4):1023–1035.

BIBLIOGRAPHY

- Dobrin, M. (1950). Dispersion in seismic surface waves. In *20th Annual SEG Meeting (Chicago)*.
- Donno, D., Chauris, H., and Calandra, H. (2013). Estimating the background velocity model with the normalized integration method. In *EAGE Technical Program Expanded Abstracts 2013*, page Tu0704.
- Dziewonski, A., Bloch, S., and Landisman, M. (1969). A technique for the analysis of transient seismic signals. *Bulletin of the Seismological Society of America*, 59(1):427–444.
- Ellefsen, K. J. (2009). A comparison of phase inversion and travelttime tomography for processing near-surface refraction traveltimes. *Geophysics*, 74(6):WCB11–WCB24.
- Eringen, A. and Suhubi, E. (1975). *Elastodynamics : Linear Theory Volume II*. Academic Press.
- Ernst, F. and Herman, G. (2000). Tomography of dispersive media. *Journal of Acoustic Society of America*, 108(1):105–116.
- Ernst, F. E., Herman, G., and Ditzel, A. (2002). Removal of scattered guided waves from seismic data. *Geophysics*, 67:1240–1248.
- Ernst, F. E. and Herman, G. C. (1998). Removal of guided waves from seismic data in laterally varying media. *Wave Motion*, 28:173–189.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophysical Journal International*, 175:665–685.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725.
- Forbriger, T., Groos, L., and Schäfer, M. (2014). Line-source simulation for shallow seismic data. part 1: theoretical background. *Geophysical Journal International*, 198(3):1387–1404.
- Foti, S. (2000). *Multistation methods for geotechnical characterisation using surface waves*. PhD thesis, Politecnico di Torino.

- Foti, S., Lai, C. G., Rix, G., and Strobbia, C. (2014). *Surface Wave Methods for Near-Surface Site Characterization*. CRC Press.
- Foti, S. and Strobbia, C. (2002). Some notes on model parameters for surface wave data inversion. In *Expanded Abstracts, Symposium on the Application of Geophysics to Engineering and Environmental Problems*. SEG.
- Gabriels, P., Snieder, R., and Nolet, G. (1987). In situ measurements of shear-wave velocity in sediments with high-mode Rayleigh waves. *Geophysical Prospecting*, 35:187–196.
- Gao, F., Williamson, P., and Pratt, R. G. (2014). A new objective function for full waveform inversion: Differential semblance in the data domain. In *SEG Technical Program Expanded Abstracts 2014*, pages 1178–1183.
- Gao, L., Brossier, R., Pajot, B., Tago, J., and Virieux, J. (2015). An immersed free surface boundary treatment for seismic wave simulation. *Geophysics*, 80(5):T193–T209.
- Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics*, 51(7):1387–1403.
- Gedge, M. and Hill, M. (2012). Acoustofluidics 17: Theory and applications of surface acoustic wave devices for particle manipulation. *Lab on a Chip*, 12(17):2998–3007.
- Gélis, C. (2005). *Inversion de formes d’onde élastique dans le domaine espace-fréquence en deux dimensions. Application à la caractérisation de la subsurface dans le cadre de la détection de cavités souterraines*. PhD thesis, Université Nice-Sophia Antipolis.
- Gélis, C., Virieux, J., and Grandjean, G. (2007). 2D elastic waveform inversion using Born and Rytov approximations in the frequency domain. *Geophysical Journal International*, 168:605–633.
- Gibson, B., Odegard, M., and Sutton, G. (2009). Nonlinear least-squares inversion of traveltimes data for a linear velocity-depth relationship. *Geophysics*, 44(2):185–194.
- Grant, F. and West, G. (1965). *Interpretation theory in applied geophysics*. McGraw-Hill Book.

BIBLIOGRAPHY

- Groos, L., Schäfer, M., Forbriger, T., and Bohlen, T. (2014). The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves. *Geophysics*, 79(6):R247–R261.
- Gucunski, N. and Woods, R. (1992). Numerical simulation of the sasw test. *Soil Dynamics and Earthquake Engineering*, 11:213–227.
- Guitton, A. and Symes, W. W. (2003). Robust inversion of seismic data using the Huber norm. *Geophysics*, 68(4):1310–1319.
- Gulunay, N. (1986). FXDECON and complex Wiener prediction filter. In *1986 Annual SEG Meeting*.
- Ha, W. and Shin, C. (2012). Proof of the existence of both zero- and low-frequency information in a damped wavefield. *Journal of Applied Geophysics*, 83:96–99.
- Hastings, F. D., Schneider, J. B., and Brochat, S. L. (1996). Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation. *Journal of Acoustical Society of America*, 100:3061–3069.
- Heisey, J., Stokoe, K., and Meyer, A. (1982). Moduli of pavement systems from spectral analysis of surface waves. *Transportation Research Record*, 852:22–31.
- Herman, F. and Verschuur, E. (2004). Curvelet-domain multiple elimination with sparseness constraints. In *SEG Technical Program Expanded Abstracts*, pages 1333–1336. SEG.
- Herrmann, R. (1987). Computer programs in seismology, vol. iv: Surface waves inversion. Technical report, St. Louis University Press.
- Hudson, J. A. and Heritage, J. R. (1981). *Use of the Born approximation in seismic scattering problems*. Cambridge University Press.
- Hustedt, B., Operto, S., and Virieux, J. (2004). Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. *Geophysical Journal International*, 157:1269–1296.
- Kennett, B. and Yoshizawa, K. (2002). A reappraisal of regional surface wave tomography. *Geophysical Journal International*, 150:37–44.
- Kennett, B. L. N. (1983). *Seismic wave propagation in stratified media*. Cambridge University Press, Cambridge.

- Lai, C. (1998). *Simultaneous inversion of Rayleigh phase velocity and attenuation for near-surface site characterization*. PhD thesis, Georgia Institute of Technology.
- Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia*, pages 206–220.
- Lamb, H. (1904). On the propagation of tremors over the surface of an elastic solid. *Philosophical Transactions of the Royal Society A*, A 203:1–42.
- Lekić, V. and Romanowicz, B. (2011). Inferring upper-mantle structure by full waveform tomography with the spectral element method. *Geophysical Journal International*, 185:799–831.
- Lerner-Lam, A. and Jordan, T. (1983). Earth structure from fundamental and higher-mode waveform analysis. *Geophysical Journal of Royal Astronomy Society*, 75:759–797.
- Levander, A. R. (1988). Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53(11):1425–1436.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least-squares. *Quarterly of Applied Mathematics*, 2:162–168.
- Lévêque, J., Cara, M., and Rouland, D. (1991). Waveform inversion of surface wave data: test of a new tool for systematic investigation of upper mantle structures. *Geophysical Journal International*, 104:565–581.
- Louie, J. (2001). Faster, better: Shear-wave velocity to 100 meters depth from refraction microtremor arrays. *Bulletin of Seismological Society of America*, 91:347–364.
- Love, A. E. H. (1911). Theory of the propagation of seismic waves. *Some problemes of Geodynamics (in Geophysics reprint series (vol 24): Classics of Elastic Wave Theory)*, pages 144–179.
- Luo, J. and Wu, R.-S. (2015). Seismic envelope inversion: reduction of local minima and noise resistance. *Geophysical Prospecting*, 63:597–614.
- Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1):2788–2792.

BIBLIOGRAPHY

- Luo, Y. and Schuster, G. T. (1991). Wave-equation travelttime inversion. *Geophysics*, 56(5):645–653.
- Luo, Y., Xia, J., Miller, R., Xu, Y., Liu, J., and Liu, Q. (2008). Rayleigh-wave dispersive energy imaging using a high-resolution linear radon transform. *Pure and Applied Geophysics*, 165:903–922.
- Ma, Y. and Hale, D. (2013). Wave-equation reflection travelttime inversion with dynamic warping and full waveform inversion. *Geophysics*, 78(6):R223–R233.
- Macquet, M. (2014). *Tomographie crustale des Pyrénées et régions avoisinantes par corrélation de bruit*. PhD thesis, Université de Grenoble.
- Maraschini, M., Ernst, F., Foti, S., and Socco, L. V. (2010). A new misfit function for multimodal inversion of surface waves. *Geophysics*, 75(4):G31–G43.
- Maraschini, M. and Foti, S. (2010). A Monte Carlo multimodal inversion of surface waves. *Geophysical Journal International*, 182:1557–1566.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.
- Martin, A. and Diehl, J. (2004). Practical experience using a simplified procedure to measure average shear-wave velocity to a depth of 30 meters (Vs30). In *Proceedings of 13th World Conference on Earthquake Engineering, Vancouver, Canada*.
- Masoni, I., Brossier, R., Virieux, J., and Boelle, J. (2013). Alternative misfit functions for FWI applied to surface waves. In *EAGE Technical Program Expanded Abstracts 2013*, page Th P10 13. EAGE.
- Maupin, V. (1988). Surface waves across 2-D structures: a method based on coupled local modes. *Geophysical Journal*, 93:173–185.
- Maupin, V. and Kennett, B. (1987). On the use of truncated modal expansions in laterally varying media. *Geophysical Journal International*, 91(3):837–851.
- McMechan, G. A. and Yedlin, M. (1981). Analysis of dispersive wave by wavefield transformation. *Geophysics*, 46:869–874.
- Métivier, L. and Brossier, R. (2016). The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. *Geophysics*, 81(2):F11–F25.

- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, 205:345–377.
- Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full Waveform Inversion and the truncated Newton method. *SIAM Journal On Scientific Computing*, 35(2):B401–B437.
- Moghaddam, P. and Mulder, W. (2012). The Diagonalator, an alternative cost functional for wave-equation inversion. In *Expanded Abstracts, 74th Annual meeting*, page W022. EAGE.
- Mora, P. R. (1988). Elastic wavefield inversion of reflection and transmission data. *Geophysics*, 53:750–759.
- Mordret, A., Shapiro, N., Singh, S., P.Roux, and Barkved, O. (2011). Low-frequency passive seismic tomography using valhall lofs. In *Expanded Abstracts*.
- Mulder, W. and Plessix, R. E. (2008). Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56(6):827–841.
- Nazarian, S. (1984). *In situ determination of elastic moduli of soil deposits and pavement systems by spectral-analysis-of-surface waves method*. PhD thesis, University of Texas at Austin.
- Nazarian, S. and Stokoe, K. H. (1984). In situ shear wave velocities from spectral analysis of surface waves. *Proc. 8th Conf. on Earthquake Eng., San Francisco*, 3:31–38.
- Neidell, N. S. and Taner, M. T. (1971). Semblance and other coherency measures for multichannel data. *Geophysics*, 36(3):482–497.
- Nishimura, C. and Forsyth, D. (1989). The anisotropic structure of the upper mantle in the pacific. *Geophysical Journal International*, 96:203–229.
- Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151):773–782.
- Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. New York, US : Springer.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

BIBLIOGRAPHY

- Nolet, G. (1977). The upper mantle under Western Europe inferred from the dispersion of Rayleigh modes. *Journal of Geophysics*, 43:265–286.
- Nolet, G. and Panza, G. (1976). Array analysis of seismic surface waves: Limits and possibilities. *Pure and Applied Geophysics*, 114:775–790.
- O’Neil, A. (2004). Full waveform reflectivity for inversion of surface wave dispersion in shallow site investigations. In *17th EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems*.
- Operto, S., Brossier, R., and Virieux, J. (2007). Documentation of FWM2DPSV programs: 2D P-SV finite-difference time-domain modelling of elastic wave propagation. Technical Report TR 07, SEISCOPE Project.
- Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A., and Dell’Aversana, P. (2004). Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study. *Geophysical Prospecting*, 52:625–651.
- Panza, G., Schwab, F., and Knopoff, L. (1972). Channel and crustal rayleigh waves. *Geophysical Journal International*, 30(3):273–280.
- Park, C., Miller, R., Ryden, N., Xia, J., and Ivanov, J. (2005). Combined use of active and passive surface waves. *JEEG*, 10(3):323–334.
- Park, C., Miller, R., and Xia, J. (1999). Multichannel analysis of surface waves. *Geophysics*, 64:800–808.
- Parker, E. and Hawman, R. (2012). Multi-channel analysis of surface waves (MASW) in karst terrain, Southwest Georgia: Implications for detecting anomalous features and fracture zones. *Journal of Environmental and Engineering Geophysics, SEG*, 17:129–150.
- Pedersen, H., Mars, J., and Amblard, P.-O. (2003). Improving surface-wave group velocity measurements by energy reassignment. *Geophysics*, 68(2):1–8.
- Pérez Solano, C. (2013). *Two-dimensional near-surface seismic imaging with surface waves: alternative methodology for waveform inversion*. PhD thesis, École Nationale Supérieure des Mines de Paris.

- Pérez Solano, C., Donno, D., and Chauris, H. (2014). Alternative waveform inversion for surface wave analysis in 2-d media. *Geophysical Journal international*, 198:1359–1372.
- 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(2):495–503.
- Plessix, R. E. and Mulder, W. A. (2008). Resistivity imaging with controlled-source electromagnetic data: depth and data weighting. *Inverse Problems*, 24:034012.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I : theory and verification in a physical scale model. *Geophysics*, 64:888–901.
- Pratt, R. G., Shin, C., and Hicks, G. J. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133:341–362.
- Pratt, R. G., Song, Z. M., Williamson, P. R., and Warner, M. (1996). Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International*, 124:323–340.
- Pratt, R. G. and Symes, W. (2002). Semblance and differential semblance optimisation for waveform tomography: a frequency domain implementation. In *Journal of Conference Abstracts*, volume 7(2), pages 183–184. Cambridge publications.
- Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013a). Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 1: imaging compressional wavespeed, density and attenuation. *Geophysical Journal International*, 194(3):1640–1664.
- Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013b). Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 2: imaging compressional and shear-wave velocities. *Geophysical Journal International*, 194(3):1665–1681.
- Rayleigh, L. (1887). On waves propagated along the plane surface of an elastic solid. *Proceedings of the London Mathematical Society (in Geophysics reprint series (vol 24): Classics of Elastic Wave Theory)*, 17:4–11.
- Richtmyer, R. D. and Morton, K. W. (1967). *Difference methods for initial value problems*. Wiley-Interscience, Kreiger, New York; réimpression (1994).

BIBLIOGRAPHY

- Rix, G. (1988). *Experimental study of factors affecting the Spectral-Analysis-of-Surface-Waves method*. PhD thesis, University of Texas at Austin.
- Robertsson, J. O. A. (1996). A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography. *Geophysics*, 61:1921–1934.
- Romanowicz, B. (1995). A global tomographic model of shear attenuation in the upper mantle. *Journal of Geophysical Research*, 100:12375–12394.
- Romanowicz, B. (2002). Inversion of surface waves: a review. *International Geophysics Series*, 81.A:149–174.
- Romanowicz, B. (2003). Global mantle tomography: progress status in the past 10 years. *Annual Review of Earth and Planetary Sciences*, 31:303–328.
- Romdhane, A., Grandjean, G., Brossier, R., Réjiba, F., Operto, S., and Virieux, J. (2011). Shallow structures characterization by 2d elastic waveform inversion. *Geophysics*, 76(3):R81.
- Romdhane, A., Grandjean, G., Rejiba, F., and Bitri, A. (2008). Inversion of surface waves in complex structures. In *21st EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems*.
- Routh, P., Krebs, J., Lazaratos, S., Baumstein, A., Lee, S., Cha, Y. H., Chikichev, I., Downey, N., Hinkley, D., and Anderson, J. (2011). Encoded simultaneous source full-wavefield inversion for spectrally shaped marine streamer data. *SEG Technical Program Expanded Abstracts 2011*, 30(1):2433–2438.
- Ryden, N., Park, C., Ulriksen, P., and Miller, R. (2004). Multimodal approach to seismic pavement testing. *Journal of Geotechnical and Geoenvironmental Engineering*, 130(6):636–645.
- Ryden, N. and Park, C. B. (2006). Fast simulated annealing inversion of surface waves on pavement using phase-velocity spectra. *Geophysics*, 71.
- Schäfer, M., Groos, L., Forbriger, T., and Bohlen, T. (2013). 2D full waveform inversion of recorded shallow seismic Rayleigh waves on a significantly 2D structure. In *in Proceedings of 19th European Meeting of Environmental and Engineering Geophysics, Expanded Abstracts, Bochum, Germany*.

- Schäfer, M., Groos, L., Forbriger, T., and Bohlen, T. (2014). Line-source simulation for shallow seismic data. part 2: full-waveform inversion - a synthetic 2-D case study. *Geophysical Journal International*, 198(3):1405–1418.
- Scholte, J. (1947). The range of existence of Rayleigh and Stoneley waves. *Geophysical Journal International*, 5(s5):120–126.
- Schwenk, J., Sloan, S., Miller, R., and Ivanov, J. (2014). Correlation of the backscatter analysis of surface waves method (BASW) for anomaly detection. In *84th Annual SEG Meeting (Denver)*.
- Sears, T., Singh, S., and Barton, P. (2008). Elastic full waveform inversion of multi-component OBC seismic data. *Geophysical Prospecting*, 56(6):843–862.
- Sercel (2016). *Geophones Specifications (April, 2016)*. <http://www.sercel.com>.
- Shapiro, N. and Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31.
- Shapiro, N., Campillo, M., Paul, A., Singh, S., Jongmans, D., and Sanchez-Sesma, F. (1997). Surface wave propagation across the mexican volcanic belt and origin of the long-period seismic-wave amplification in the valley of mexico. *Geophysical Journal International*, 128:151–166.
- Sheng, J., Leeds, A., Buddensiek, M., and Schuster, G. T. (2006). Early arrival waveform tomography on near-surface refraction data. *Geophysics*, 71(4):U47–U57.
- Sheriff, R. E. and Geldart, L. P. (1995). *Exploration seismology, second edition*. Cambridge University Press.
- Shi, T., Zhang, J., Huang, Z., and Jin, C. (2015). A layer-stripping method for 3D near-surface velocity model building using seismic first-arrival times. *Journal of Earth Science*, 26(4):502–507.
- Shin, C. and Cha, Y. H. (2008). Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173(3):922–931.
- Shipp, R. M. and Singh, S. C. (2002). Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. *Geophysical Journal International*, 151:325–344.

BIBLIOGRAPHY

- Silvestrov, I., Bakulin, A., Dmitriev, M., Gadylshin, K., Golikov, P., Neklyudov, D., and Tcheverda, V. (2015). Full-waveform Inversion of Surface Waves - Numerical Sensitivity Analysis. In *Extended Abstracts, 77th Annual EAGE Conference & Exhibition, Madrid*. EAGE.
- Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.
- Smithyman, B., Pratt, R. G., Hayles, J., and Wittebolle, R. (2009). Detecting near-surface objects with seismic waveform tomography. *Geophysics*, 74(6):WCC119–WCC127.
- Socco, L. and Strobbia, C. (2004). Surface-wave method for near-surface characterization: a tutorial. *Near Surface Geophysics*, 2:165–185.
- Socco, L. V., Foti, S., and Boiero, D. (2010). Surface-wave analysis for building near-surface velocity models — established approaches and new perspectives. *Geophysics*, 75(5):75A83–75A102.
- Stokoe, K. H., Wright, S. G., Bay, J. A., and Roesset, J. M. (1994). Characterization of geotechnical sites by sasw method. *Geophysical characterization of sites*, pages 15–25.
- Strobbia, C. (2002). *Surface wave method: acquisition, processing, and inversion*. PhD thesis, Politecnico di Torino.
- Strobbia, C. and Foti, S. (2006). Multi-offset phase analysis of surface wave data (MOPA). *Journal of Applied Geophysics*, 59:300–313.
- Stutzmann, E. (1993). *Tomographie du manteau à partir des modes harmoniques des ondes de surface*. PhD thesis, Université Paris VII.
- Stutzmann, E. and Montagner, J. (1994). Tomography of the transition zone from the inversion of higher-mode surface waves. *Physics of the Earth and Planetary Interiors*, 86:99–116.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180:433–462.
- Tarantola, A. (1984a). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

- Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.
- Tarantola, A. (1987). *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, New York.
- Tarrass, I., Giraud, L., and Thore, P. (2011). New curvilinear scheme for elastic wave propagation in presence of curved topography. *Geophysical Prospecting*, 59(5):889–906.
- Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160:195–216.
- van der Lee, S. (2002). High-resolution estimates of lithospheric thickness from Missouri to Massachusetts, usa. *Earth and Planetary Science Letters*, 203(1):15–23.
- van Leeuwen, T., Aravkin, A., Calandra, H., and Herrmann, F. J. (2013). In which domain should we measure the misfit for robust full waveform inversion. In *Expanded Abstracts*. EAGE.
- van Leeuwen, T. and Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation travelttime tomography. *Geophysical Journal International*, 182(3):1383–1394.
- Vignoli, G., Strobbia, C., Cassiani, G., and Vermeer, P. (2011). Statistical multioffset phase analysis for surface-wave processing in laterally varying media. *Geophysics*, 76(2):U1–U11.
- Viktorov, I. A. (1967). *Rayleigh and Lamb Waves: physical theory and applications*. Plenum Press, New York.
- Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite difference method. *Geophysics*, 51:889–901.
- Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.
- Wang, Y. and Rao, Y. (2009). Reflection seismic waveform tomography. *Journal of Geophysical Research*, 114(B3):1978–2012.
- Warner, M. and Guasch, L. (2014). Adaptive waveform inversion - fwi without cycle skipping - theory. In *76th EAGE Conference and Exhibition 2014*.

BIBLIOGRAPHY

- Wathelet, M. (2008). An improved neighborhood algorithm: parameter conditions and dynamic scaling. *Geophysical Research Letters*, 35:L09301.
- Wathelet, M., Jongmans, D., and Ohrnberger, M. (2004). Surface-wave inversion using a direct search algorithm and its application to ambient vibration measurements. *Near Surface Geophysics*, 2(4):211–221.
- Wathelet, M., Jongmans, D., Ohrnberger, M., and Bonnefoy-Claudet, S. (2008). Array performances for ambient vibrations on a shallow structure and consequences over Vs inversion. *Journal of Seismology*, 12:1–19.
- Woods, R. (1968). Screening of surface waves in soils. *Journal of the Soil Mechanics and Foundations Division*, 94(4):951–979.
- Xia, J., Miller, R., and Park, C. (1999). Estimation of near-surface shearwave velocity by inversion of Rayleigh wave. *Geophysics*, 64:691–700.
- Yilmaz, Ö. (1987). *Seismic data processing*. Society of Exploration Geophysicists.
- Yuan, Y., Simons, F., and Bozdag, E. (2015). Multiscale adjoint waveform tomography for surface and body waves. *Geophysics*, 80:R281–R302.
- Yuan, Y. O., Simons, F. J., and Bözdag, E. (2014). Full-waveform adjoint tomography in a multiscale perspective. In *SEG Technical Program Expanded Abstracts 2014*, pages 1194–1199.

