Università degli Studi di Torino
Doctoral School in Earth Sciences – cycle 32nd

Doctoral Thesis

in

Applied Geophysics

Initial S - wave and P - wave velocity models for full-waveform inversion from surface wave dispersion curves analysis

PhD candidate:

Daniela TEODOR

Supervisor: **Prof. Cesare COMINA** - Università degli Studi di Torino, DST

Co-Supervisor: **Prof. Laura Valentina SOCCO** - Politecnico di Torino, DIATI

Co-Supervisor: **Prof. Romain BROSSIER** - Université Grenoble Alpes, ISTERRE

Co-Supervisor: **Prof. Jean VIRIEUX** - Université Grenoble Alpes, ISTERRE

Referee and Examiner: **Prof. Thomas BOHLEN** - Geophysikalisches Institut, Karlsruher Institut für Technologie

Referee and Examiner: **Prof. Michele CERCATO** - Università di Roma La Sapienza, DICEA

Examiner: **Prof. Sergio Carmelo VINCIGUERRA** - Università degli Studi di Torino, DST

PhD Coordinator: **Prof. Anna Maria Ferrero** - Università degli Studi di Torino, DST

Torino, April 2020

ACKNOWLEDGEMENTS

I would like to thank the CNR group (of Turin) for allowing us to perform the seismic acquisition in the experimental site.

This study has granted access to the HPC resources of the Froggy platform of the UGA CIMENT infrastructure (https://ciment.ujf-grenoble.fr).

Many thanks to the Erasmus program and Italian Ministry of Research for financing me to work for six months in the Seiscope team at IsTerre, Grenoble.

Many thanks to the Italian Ministry of Research for financing my PhD grant.

I would also like to thank all the persons who contributed to my personal and professional evolution.

Abstract

Retrieving the physical properties of the shallow subsurface has various uses, from the site characterization for seismic hazard analysis (based on the S-wave velocity model), to the static corrections of reflection seismic data (that require the knowledge of the P-wave velocity model). An inaccurate reconstruction of the shallow heterogeneities may affect significantly the imaging results at greater depth and the related seismic attributes. In the presence of complex near-surface geometries, the use of specific seismic methods may be necessary for guaranteeing the targeted resolution.

Elastic full-waveform inversion (FWI) is an effective tool for high-resolution subsurface multi-parameter characterization. However, FWI is rarely employed for near-surface applications, since the seismic data are dominated by highly energetic, dispersive and scattered surface waves (SWs). In these conditions, a successful deterministic FWI scheme requires a very accurate initial model. Besides, various inversion strategies may be necessary to guarantee adequate convergence. Particularly, the initial model becomes very important when the SWs data domination reduces the algorithm's sensitivity with respect to specific waveform components and related parameters (e.g. first or reflected arrivals and P-wave velocity). Contextually, this study aims at enhancing the resolution of complex shallow targets imaging, by integrating devoted SW analysis techniques (for the initial model building) with a spectral-element based elastic 3D FWI.

Different initial S-wave (V_S) and P-wave (V_P) velocity models (laterally homogeneous and laterally variable) are retrieved from dispersion curves (DCs), which are extracted through a moving Gaussian windowing technique and *f-k* domain processing. The workflow for initial model building is based on a clustering algorithm, followed by a data-transform procedure. The workflow is applied to both synthetic and field seismic data, recorded in a shallow geological context characterized by a known low-velocity target of a particular shape.

On these models, various FWI tests are performed, using a simple scheme in the first place, to evaluate the influence of the different initial models (both laterally homogeneous and laterally variable), retrieved using SW analysis, on the results. The adopted FWI workflow enriched the overall content of the initial models, allowing a reliable reconstruction over the shallow target. Particularly, more accurate results were obtained after FWI when using laterally variable initial models.

The integration of different model and data-oriented strategies into the main optimization scheme allowed obtaining an even better convergence. Among others, a preliminary monoparametric FWI, which exploits the higher sensitivity of SW concerning the shear properties, contributed to the improvement of the initial V_S model. Moreover, data-based strategies, such as offset weighting, time windowing and multiscale inversion led to the achievement of better data fitting and more accurate model reconstruction after FWI.

A 3D acquisition layout, characterized by different source positions and variable source number, was also tested, ensuring a better lateral resolution of the reconstructed target after FWI.

Riassunto

Conoscere le proprietà fisiche degli orizzonti geologici superficiali ha varie utilità, dalla caratterizzazione sismica dei siti (basata sul modello di velocità delle onde S) alle correzioni statiche dei dati sismici a riflessione (che richiedono la conoscenza del modello di velocità delle onde P). Una ricostruzione inaccurata della variazione delle proprietà fisiche negli orizzonti superficiali può influire in modo significativo sui risultati di *imaging* a profondità più elevate e sui relativi attributi sismici. Di conseguenza, quando gli orizzonti superficiali sono caratterizzati da geometrie complesse, l'integrazione di più metodi sismici può essere importante per garantire la risoluzione ricercata.

La *full-waveform inversion* (FWI) in approssimazione elastica è un potente strumento per la caratterizzazione multi-parametrica e ad alta risoluzione del sottosuolo. Tuttavia, la FWI viene raramente applicata ai dati sismici provenienti da acquisizioni a bassa profondità, in quanto i sismogrammi relativi ad esse sono dominati da onde superficiali altamente energetiche e dispersive. In queste condizioni, uno schema funzionale di FWI deterministico richiede un modello iniziale molto accurato; inoltre, varie strategie di inversione sono necessarie per garantire un'adeguata convergenza. In particolare, il modello iniziale diventa molto importante quando la dominazione delle onde superficiali nei dati riduce la sensibilità dell'algoritmo rispetto a specifici componenti della forma d'onda e parametri correlati (ad esempio: i primi arrivi, gli arrivi riflessi e la velocità delle onde P). Contestualmente, l'obiettivo del presente studio è la ricostruzione ad alta risoluzione di target superficiali complessi attraverso l'integrazione di tecniche particolari di analisi delle onde superficiali (per la definizione del modello iniziale) con un algoritmo di FWI elastica 3D, basato sul metodo degli elementi spettrali.

A tal fine, diversi modelli iniziali di velocità delle onde S (V_S) e delle onde P (V_P) (lateralmente omogenei e lateralmente variabili) sono ricavati dalle curve di dispersione. Tali modelli sono estratti dai dati sismici utilizzando una finestra gaussiana mobile e tecniche di processing nel dominio *fk*. Il *workflow* completo per la costruzione del modello iniziale prevede un algoritmo di *clustering*, seguito da una procedura di *data transform*. Tale *workflow* viene applicato sia ai dati sismici sintetici, che ai dati reali, acquisiti in un contesto geologico poco profondo, caratterizzato da un target a bassa velocità sismica.

I primi test di FWI sono stati eseguiti utilizzando uno schema semplice, ai fini di valutare l'efficacia dei diversi modelli iniziali ricavati dalle curve di dispersione. I risultati appartenenti a questi esperimenti hanno dimostrato come la FWI consenta il miglioramento complessivo dei modelli iniziali e una ricostruzione affidabile del target. In particolare, i modelli iniziali variabili lateralmente hanno garantito risultati più accurati dopo la FWI.

L'integrazione di diverse strategie (orientate ai parametri del modello oppure ai dati sismici) nello schema di ottimizzazione principale ha consentito di ottenere una migliore convergenza dopo la FWI. Ad esempio, una preliminare FWI mono-parametrica, che sfrutta la maggiore sensibilità delle onde superficiali rispetto al modulo di taglio, ha contribuito al miglioramento del modello iniziale di V_S. In più, le strategie di FWI orientate ai dati, tra cui l'applicazione di una funzione di peso variabile con l'offset, la finestratura sull'asse dei tempi e l'inversione *multiscale*, hanno portato al raggiungimento di un migliore *data-fitting* e ad una ricostruzione del modello più accurata.

Inoltre, è stata testata anche una geometria di acquisizione sismica 3D, caratterizzata da una posizione variabile delle sorgenti e un numero variabile dei punti di scoppio stessi. Tale configurazione ha garantito una migliore risoluzione laterale del target dopo FWI.

Thesis structure

The Thesis consists of two main bodies, following the *Introduction* section:

Part I (*Chapters* 2 to 5), which provides an overview of the theoretical background and describes the applied methodologies.

Part II (*Chapters* 6 to 10), which contains the main experiments and connected results.

Several Appendices are attached at the end of the corresponding Chapters for fluency reading purposes.

Chapter 1 gives a general introduction on the investigated problem, with emphasis on the applications and uses of surface wave analysis and full-waveform inversion. The overview regards the state of art and the importance of the initial model's accuracy when full-waveform inversion aims to reconstruct shallow targets characterized by high impedance contrasts. In this general framework, the objectives of the Thesis are presented, as well as the connected scientific question.

Chapter 2 presents basic theoretical concepts on the seismic wave propagation and the analytical and numerical solution of the wave equation. Some space (in *Appendix* 1) is allocated to the spectral element method, as it represents the architecture of the full-waveform inversion tool used for this study.

Chapter 3 offers a synthetic presentation of the inverse problem theory, with conceptual insights on the deterministic and probabilistic approaches for the solution of non-linear inverse problems.

Chapter 4 gives an overview of the surface wave analysis through the exploitation of the dispersive behavior occurring in vertically heterogeneous media. After some theoretical insights on the surface wave methods, the workflow for the initial model building is presented.

Chapter 5 is dedicated to the full-waveform inversion method. The limitations of the gradient-based techniques are introduced and possible strategies that can be employed to constrain the optimization are presented as well. Some examples from the literature, related to FWI application to shallow wavefields, are also presented. The final part of this chapter contains a description of the spectral-element based full-waveform inversion code used for this study, followed by some insights on the source estimation from field data.

N.B. **Chapter 2, Chapter 3** and **part of Chapter 4** and **Chapter 5 are not intended for expert readers.**

In the above-mentioned Chapters, the key elements necessary for a fluent reading are presented in **Section 4.3** of Chapter 4 and **Section 5.4** of Chapter 5.

Chapter 6 is dedicated to forward modeling simulations and presentation of the real and synthetic data sets analyzed in this work. The first part gathers some synthetic experiments related to 3D viscoelastic wave propagation, conducted over models with different geometries. Some of these models reflect the geological asset of the test site. The second part describes the parameters and characteristics of the seismic acquisition carried out in the experimental site, as well as the corresponding field data set. The third part presents the synthetic data set used for validating the methodology applied in this study.

Chapter 7 presents the various initial models retrieved from surface wave dispersion curves analysis application on both synthetic and field data. 3D numerical simulations are conducted over some of these models and a novel set of dispersion curves is extracted from the modeled data set. A QC in terms of data-fitting and dispersion curve comparison is then presented.

Chapter 8 contains various full-waveform inversion tests, performed using the initial models retrieved from surface wave analysis over both synthetic and field data. The first full-waveform inversion tests involve a relatively simple workflow and a 2D acquisition layout, to focus the investigation on the dispersion curves based initial model's effectiveness. Further, different model-oriented and data-based strategies are integrated into the main full-waveform inversion workflow. A 3D acquisition geometry is also tested. Data-fitting comparisons are presented for all experiments.

Chapter 9 offers a quantitative interpretation of the results obtained after full-waveform inversion, as well as some comparisons and discussion related to the effectiveness of the seismic methods used for geotechnical scale exploration. The results obtained in this work are also compared with other published results.

Chapter 10 summarizes the main conclusions related to the addressed scientific question and the connected experiments while presenting also some perspectives of the study.

Frequently used terminology

SW	*Surface Wave*
BW	*Body Wave*
DC	*Dispersion Curve of the fundamental mode of Rayleigh waves*
V_P	*P-wave velocity model*
V_S	*S-wave velocity model*
FWI	*Full Waveform Inversion*
EAWI	*Early Arrival Waveform Inversion*
SEM	*Spectral Element Method*
f. a.	*first-arrival*
f-k	*frequency-wavenumber*
MASW	*Multichannel Analysis of Surface Waves*
Initial model	*Velocity model obtained via surface wave dispersion curves analysis*
Single-DC analysis	*The analysis of one dispersion curve (i.e. 1D analysis)*
Single-DC based initial model	*Velocity model obtained from the analysis of one dispersion curve (1.e. 1D analysis)*
Full-DC analysis	*The analysis of all dispersion curves along a seismic line (i.e. 2D analysis)*
Full-DC based initial model	*Velocity model obtained from the analysis of all dispersion curves along a seismic line (1.e. 2D analysis)*
Reference/True model	*The synthetic velocity model used to generate the synthetic reference (i.e. true) data*
Final model	*Velocity model obtained after full-waveform inversion*
Initial data	*Seismic data belonging to the initial model retrieved from dispersion curves analysis*
Inverted data	*Seismic data belonging to the final model obtained after full-waveform inversion*
Real data	*Seismic data belonging to the field acquisition*
Synthetic data	*Seismic data obtained from numerical simulation over the reference, initial or final model*

Contents

PART I – Theory and Methodology

PART II – Experiments

Chapter 1

Introduction

Contents

1.1 – Generic introduction and historical overview

The seismic method aims at estimating the internal properties of the Earth by recording, with an array of receivers, the mechanical vibrations caused by artificial (man-maid) or natural sources. The recorded data, called seismograms or traces, are oscillatory wriggles sampled in time. Depending on the instrument response, these vibrations are essentially either particle velocities or particle accelerations. Most of the sensors are based on inertial properties: they can record horizontal and/or vertical motions. For geotechnical applications, the vibrations emitted by each source are recorded on different geophones with often regular offsets.

For many decades until the 1960s, the advances in seismic wave theory, the instrumentation's precision and the computation barriers have controlled the seismograms quantitative interpretation's feasibility. The main developments between the 1960s and 1980s regarding the wave propagation theory, digital processing and computing architectures, supported the quantitative seismic method's effectiveness. Moreover, the continuous evolution of the data inversion theory provided accurate mathematical tools, allowing for the recorded data conversion into high-resolution quantitative models of the subsurface physical parameters. Until the introduction of full-waveform inversion (FWI) in the early 80s, only limited information, such as travel times, was used for subsurface imaging. Although the first FWI experiments for crustal imaging were focused on deep marine environments, FWI gradually became a widely used tool for land data sets as well. Yet, using FWI for the reconstruction of near-surface land targets is not a routine procedure.

Knowing the physical properties of the shallow geological layers is useful for various purposes, from the site characterization to the static corrections of data belonging to reflection seismic surveys. The individuation of localized low-velocity shallow heterogeneities is important for the accurate imaging of the structures at greater depth. However, the accurate reconstruction of sharp-interface shallow targets, from land acquisition data, is controlled by the complex behavior of the wavefield when interacting with them. In the presence of high lateral impedance contrasts, the effectiveness of the commonly employed tools for near-surface characterization is partially limited and the integration of various seismic methods may be required to obtain reliable results.

The first-arrival traveltime tomography often used to retrieve the P-wave velocity (V_P) variation is not sensitive to embedded low-velocity horizons, since the first-arrival phases tend to propagate through higher-velocity layers. Moreover, for small-offset acquisitions, the time picking process may be complicated due to the difficulty to distinguish the separation of various phases. Several surface wave (SW) analysis and inversion techniques can be employed to retrieve the S-wave velocity (V_S) variation, using processing workflows based on windowing and wavefield transform to extract and invert local dispersion curves (DCs) – i.e. the SW phase velocity variation with the frequency/wavelength. However, the reconstructed V_S may not be accurate enough in the presence of sharp lateral transitions, given the locally layered-model assumption of SW methods. Moreover, due to the plane-

wave approximation commonly adopted by the SW methods, the localized lateral heterogeneities cannot be easily identified and picked in the f-k domain, since they do not exhibit a continuous dispersive energy trend (Pan et al., 2019b).

Compared with the near-surface phase-oriented techniques, FWI, based on waveform analysis, overcomes the above-mentioned limitations while ensuring a theoretical resolution equal to half of the local propagating wavelength (Vireux and Operto, 2009). FWI also allows for the simultaneous reconstruction of multiple parameters (e.g., V_P, V_S, mass density, attenuation factor) providing a more detailed site characterization.

FWI is defined as a non-linear iterative data fitting procedure, based on the numerical solution of the seismic wave equation (Virieux and Operto, 2009; Fichtner, 2011). After the introduction of FWI by Lailly (1983) and Tarantola (1984), different successful applications have been suddenly performed both in the time domain (Tarantola, 1986; Mora, 1987a,b; Bunks et al., 1995; Zhou et al., 1995 and 1997; Fichtner et al., 2013) and in the frequency domain (Liao and McMechan, 1996; Pratt et al., 1998; Pratt, 1999; Ravaut et al., 2004; Ben-Hadj-Ali et al., 2008; Brossier et al., 2009 and 2010; Plessix, 2009). FWI is under continuous development, supported by the simultaneous evolution of the parallel computing architectures. Nevertheless, one of its drawbacks is still the high computational cost. For this reason, most of the FWI applications, especially for marine data, are performed using the acoustic approximation (e.g. Plessix and Perkins, 2010; Warner et al., 2013; Operto et al., 2015). The 2D elastic approximation is also used, both for marine and land applications (e.g. Sheen et al., 2006; Sears et al., 2008; Bretaudeau et al., 2013; Vigh et al., 2014), but the time-domain 3D elastic or viscoelastic FWI examples are still rare in the literature (e.g. Epanomeritakis et al., 2008; Butzer et al., 2013; Borisov and Singh, 2015; Fathi et al., 2015; Trinh et al., 2019).

Few studies used FWI for shallow structures imaging, but most of them are based on the 2D approximation of the wave equation (e.g. Romdhane et al., 2011; Tran and McVay, 2012; Bretaudeau et al., 2013; Tran et al., 2013; Dou and Ajo-Franklin, 2014; Gross et al., 2014; Masoni et al., 2014; Schäfer et al., 2014; Amrouche and Yamanaka, 2015; Bohlen et al., 2015; Nuber et al., 2016; Athanasopoulos and Bohlen, 2017; Gross et al., 2017; Köhn et al., 2016; Pan et al., 2016; Chen et al., 2017; Gross et al., 2017; Nuber et al., 2017; Athanasopoulos et al., 2018a,b; Köhn et al., 2018; Krampe et al., 2018; Athanasopoulos and Bohlen, 2019; Krampe et al., 2019; Pan et al., 2019a,b; Wang et al., 2019; Xing and Mazzotti, 2019a,b; Borisov et al., 2020). The full-3D FWI applications for near-surface characterization are nowadays less common (Butzer et al., 2013; Fathi et al., 2016; Nguyen and Tran, 2018; Borisov et al., 2018; Smith et al., 2019; Tran et al., 2019; Irnaka et al., 2019). Some of these publications are summarized in Section 5.3.

While FWI is mainly driven by body waves for upper-crust imaging (Sirgue et al., 2010; Warner et al., 2013), data from shallow acquisitions are dominated by SWs. For near-surface reconstruction, elastic or viscoelastic propagation need to be considered. The (visco)elastic multi-parameter FWI is facing more challenges. One of them is related to the different sensitivity of the algorithm with respect to each parameter class. Another one is the increased computational cost and inverse problem's dimensionality when

implementing the viscoelastic propagation. A linearized formulation based on quasi-Newton methods is the most widespread technique, given its computational efficiency. Nevertheless, the commonly used least squares (L_2) objective function contains many local minima where the solution could be trapped. Consequently, an accurate initial model, which predicts the observed data with an error smaller than half-period, is required for a proper convergence (Pratt, 1999; Mulder and Plessix, 2008; Virieux and Operto, 2009).

One strategy that aims at mitigating the FWI's non-linearity follows a multi-scale and multi-resolution hierarchical approach (Bunks et al., 1995). Accordingly, the inversion starts from low frequencies (long wavelengths) and gradually incorporates the higher frequencies (short wavelengths): The lower frequencies are less likely to favorize delays of more than a half-cycle in the predicted waveform (Sirgue and Pratt, 2004; Pratt, 1999). An opposite alternative is the layer-stripping procedure (high-to-low frequency filtering). In this case, the model reconstruction proceeds hierarchically from the shallow part to the bottom, while the offset range is gradually increased as a function of penetration depth (Virieux and Operto, 2009; Shi et al., 2015; Masoni et al., 2016): This is especially true for SWs with a specific medium probing related the depth dependence of eigenmodes.

The FWI applications for shallow target reconstruction are very sensitive to the initial model configuration. This is mainly due to the complex structure of the wavefield, with often-combined highly energetic and scattered phases. Therefore, devoted strategies are required for the correct definition of the initial model. An accurate initial model can be retrieved with global inversion strategies, essentially used for 1D or 2D geometries (Tran and Hiltunen, 2012a,b; Xing and Mazzotti, 2019a,b). Nevertheless, for 3D applications, the high computation resources required by the probabilistic techniques and the computationally oversized character of the inverse problem limit their employment to a few cases. Instead, SW analysis may be a more efficient tool for building accurate V_S and V_P initial models. However, only the V_S initial model is commonly retrieved, while the V_P model is usually inferred from V_S by assuming a constant Poisson's ratio. An alternative for the initial V_P model building is the first-arrival traveltime tomography, although this technique is not very efficient for the reconstruction of localized low-velocity shallow layers.

SW analysis commonly provides only the V_S model, but some studies (Socco et al., 2017; Socco and Comina, 2017) introduced a method based on the concept of SW skin depth, to estimate also the V_P model. Further, Khosro Anjom et al. (2019) proposed a data-clustering algorithm, which efficiently organizes data subsets for mitigating the impacts of strong lateral variations. Moreover, a specific data transform allows building both V_S and V_P models over laterally variable shallow environments. Following these approaches, the method used in this study for the initial model building allows considering high lateral contrasts and provides both V_S and V_P models.

1.2 –Thesis objectives

This study aims at enhancing the resolution of complex shallow targets imaging, by integrating 2D SW analysis techniques (for the initial model building) with a spectral element based elastic 3D FWI. Specifically, the purpose of the work is investigating whether the initial models retrieved from SW analysis can efficiently handle the additional challenges, arising when FWI is applied to shallow land data sets, recorded over targets characterized by high impedance contrasts and lateral heterogeneities.

In this framework, the study addresses the following questions:

I) Is an initial model derived from SW analysis accurate enough for FWI?

II) Given the initial model derived from SW analysis, which is the best FWI workflow and seismic acquisition layout (2D or 3D)?

To answer the first question, initial V_S and V_P models, retrieved using SW analysis, are tested with a relatively simple FWI scheme. These initial models are characterized by a different lateral resolution, according to the adopted SW analysis workflow:

A) "single-DC analysis (i.e. 1D analysis)", which uses only one DC, characteristic of the background medium, and provides laterally homogeneous models, or

B) "full-DC analysis (i.e. 2D analysis)", which uses the entire set of DCs extracted from the seismic data and provides laterally heterogeneous initial models.

To answer the second question, the FWI workflow, which is relatively simple during the first experiments, becomes gradually more sophisticated. Specifically, model-oriented and data-based strategies are incorporated into the main FWI scheme, to evaluate their potential efficiency for better data fitting and more accurate model reconstruction.

Finally, a 3D acquisition design, with flexible source positions and variable source number, is also tested, to evaluate which configuration guarantees the best resolution of the 3D target-model after FWI.

Some publications are connected to Thesis research. Among them:

- Teodor et al. (2017) presented a preliminary QC analysis for seismic data belonging to V_P and V_S models built from SW analysis.
- Teodor et al. (2018a) showed an example of an initial model building from DC analysis using a synthetic data set, while
- Khosro Anjom et al. (2019) applied the same workflow (DC clustering and data transform) to a real data set, acquired over the complex-shaped shallow target.

  For both synthetic and real tests, a good waveform matching was noticed when comparing the reference (i.e. true) or real data with the corresponding synthetic data belonging to the initial models from DC analysis.
- Teodor et al. (2018b) and Teodor et al. (2019) presented some preliminary results of multi-parameter 3D FWI on synthetic data, starting from initial models retrieved by DC analysis.
- Teodor et al. (2020 – submitted) investigated the accuracy of DC analysis based initial models by carrying out various 3D FWI tests on both synthetic and field data. Some data-based and model-oriented strategies, as well as different acquisition geometries, are experimented in this study.
- Teodor et al. (2020 – submitted) questioned about the best acquisition layout for imaging near-surface heterogeneities with high accuracy (when using 2D SW analysis and 3D FWI).

PART I
Theory and Methodology

Chapter 2

Forward problem overview

Contents

2.1 – Introduction to seismic waves propagation

Describing the waves propagation in highly heterogeneous media often involves the use of some assumptions (i.e. the elastic particle displacement, the harmonic repetition of the wave in space and time, the small-strain assumption). These assumptions allow using the elastic theory (e.g. Telford et al., 1991; Aki & Richards, 2002) to describe the medium's particles variation in shape and volume as a reaction to the seismic wave propagation.

Considering a harmonic signal crossing a certain element of the space, the amount of time it takes for one wave cycle to complete is called *period* T, while the *wavelength* λ is the distance between two adjacent peaks. The number of times the wave cycle repeats in each period (i.e. one second) is defined as *frequency f*, whereas the *wavenumber k* is the spatial frequency of the wave, measured in cycle or radians per unit distance.

When the energy is released from the source, part of it propagates through the medium as body waves (*BWs*: pressure waves and shear waves), while another part distributes along the surface (*SWs*: Rayleigh waves and Love waves). The *seismic wave velocity* represents the velocity of the seismic energy propagating in a medium. It is different from the *particle velocity*, which is the velocity of the medium's particles caused by the seismic wave propagation.

The pressure (P) wave propagation inside a medium causes the vibration of the particles according to compressive and relaxing movements, in the same direction of propagation (Figure **2.1a**). The action of a shear (S) wave causes the oscillation of the particles transversally to the direction of propagation (Figure **2.1b**), in the vertical plane (SV-wave, *vertical shear*) and the horizontal plane (SH-wave, *horizontal shear*). The P-wave velocity is defined by the relation:

$$V_p = \sqrt{\frac{\lambda+2\mu}{\rho}} = \sqrt{\frac{k+\frac{4}{3}\mu}{\rho}}, \tag{2.1}$$

while the S-wave velocity is defined by the relation:

$$V_s = \sqrt{\frac{\mu}{\rho}}, \tag{2.2}$$

where:

- $k = \lambda + \frac{2}{3}\mu$ is the *Bulk* or *compressibility modulus* (which describes the change in volume of a material that follows the action of hydrostatic stress); k is always positive.
- $\mu = \frac{E}{2(1+\nu)}$ is the *shear* or *stiffness modulus* (which describes the change in the shape of a material related to the action of tangential stress). In a fluid $\mu = 0$, then S-waves do not propagate in fluids.
- $E = \frac{\sigma_{xx}}{\varepsilon_{xx}}$ is the *Young modulus* (which describes the deformation of the material when normal stress is applied).

- $\lambda = \frac{(\nu E)}{(1+\nu)(1-2\nu)}$ is the *Lamé* coefficient.
- $\nu = \frac{\lambda}{2(\lambda+\mu)}$ is the Poisson's coefficient, which connects k and E according to the relation $k = \frac{E}{3(1-2\nu)}$.
- $\left(\frac{V_P}{V_S}\right)^2 = \frac{\lambda+2\mu}{\mu} = \frac{2(1-\nu)}{(1-2\nu)} > 1$. For ν=0.25, $V_P = \sqrt{3}V_S$.

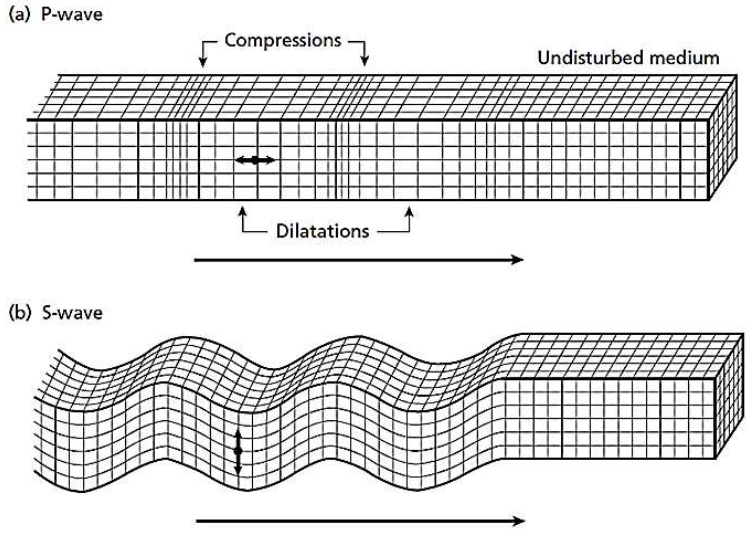

*Figure **2.1** – Elastic deformation of the medium's particles related to the body waves propagation. (**a**) P-wave, (**b**) S-wave – from Kearey et al. (2002).*

The surface on which all particles are moving with the same phase, in each instant, represents the *wave front*. At a short distance from the source, in a homogeneous medium, BWs have a spherical wave front. According to the Huygens principle, all the points belonging to a spherical wave front become sources of secondary spherical waves; the new wave front is defined by the envelope of these secondary waves. As the distance from the source increases, the observed curvature of the wave front decreases, up to the point where it can be considered as flat. At such distance, the seismic wave can be assimilated to a *plane wave*. For isotropic media, the direction perpendicular to the wave front represents the *ray.*

SWs are mainly divided into two categories: *Rayleigh waves* (L_R) and *Love waves* (L_Q). Rayleigh waves are generated from the interaction between evanescent pressure waves (P) and vertically polarized shear-waves (SV). They propagate along the shallow part of a medium, bearing about 2/3 of the waveform energy. In a homogeneous medium, the amplitude displacement of the particles decreases exponentially with depth. The particles of the L_R wave front vibrate in a vertical plane, with an elliptical retrograde motion with respect to the direction of propagation; the major (vertical) axis and the minor (horizontal) axis are oriented toward the direction of propagation (Figure **2.2a**).

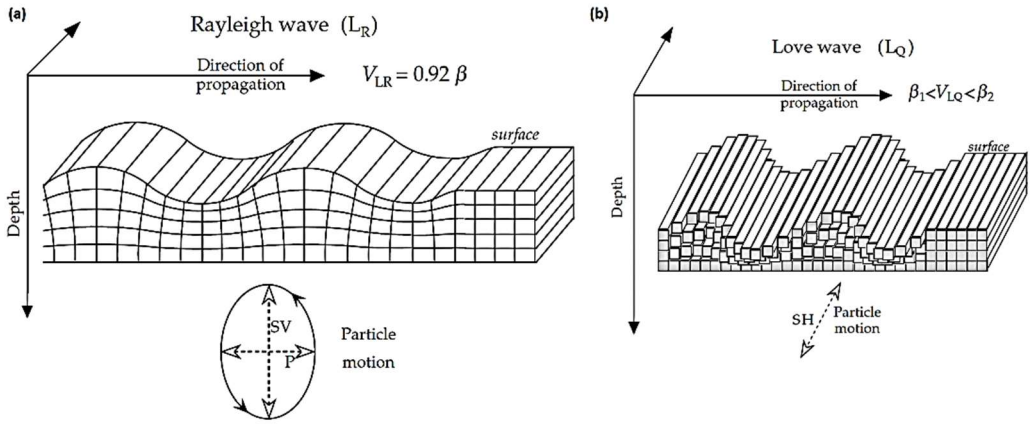

*Figure **2.2** – Elastic deformation of the medium particles related to the surface waves propagation. (**a**) Rayleigh wave, (**b**) Love wave – from Lowrie (2007).*

In the presence of a homogeneous medium, the Rayleigh wave velocity does not depend on the frequency. Contrariwise, for a heterogeneous medium (where the stiffness varies with depth), the Rayleigh wave phase velocity is frequency dependent. In particular, the larger wavelengths (lower frequencies) travel deeper in the subsurface, reaching geological layers that are characterized by different mechanical proprieties, and then by different velocities. This phenomenon is known as *surface waves dispersion*, while a *dispersion curve* defines the relationship between the phase velocity of the Rayleigh waves V_R and the frequency *f.*

Love waves (L_Q) are generated in heterogeneous media because of constructive interference among S waves trapped under the free surface. They are mainly observed in the presence of a thin shallow horizon, overlying a stiff half-space, or in the presence of a low-velocity shallow layer trapped between two stiffer layers. The L_Q waves engine is represented by a supercritical angle of incidence, the consequent upward reflection, followed by the downward reflection of the energy in correspondence of the free surface. They cause the vibration of the medium particles orthogonally to the direction of propagation, with a shear motion polarized in a horizontal plane (SH), parallel to the surface, with rectilinear ray path (Figure **2.2b**). Love wave velocity always depends on frequency (wavenumber) and therefore LQ are dispersive.

As a seismic signal propagates away from the source, its energy decreases. In part, the signal attenuation is due to the geometrical spreading: the energy is distributed on spherical surfaces (BWs) or cylindrical surfaces (SWs) that become progressively greater due to the increase of the radius *r*. Consequently, while BWs energy attenuation is proportional to $\frac{1}{r^2}$: $\left[I_V(r) = \frac{E_V}{4\,\pi\,r^2}\right]$, SWs attenuation is proportional to $\frac{1}{r}$: $\left[I_S(r) = \frac{E_S}{4\,\pi\,r\,d}\right]$. The corresponding amplitude attenuation is $1/r$ and $1/\sqrt{r}$, respectively (as the intensity is proportional to the square of the amplitude). Therefore, BWs amplitude decreases more with the distance than SWs amplitude.

The seismic wave amplitude attenuation is also caused by the inelastic properties of the medium (anelastic damping), described by a parameter called quality factor: $Q=2\pi E/\Delta E$. In the literature, the *Q value* also refers to a field measurement (i.e. an observable), which is different from the so-called *Q parameter* from rheology description. The parameter Q expresses the energy loss per cycle, where E represents the total amount of energy and ΔE the energy lost in one cycle:

$$A = A_0 \, exp\left(-\frac{\pi r f}{Q v}\right) = A_0 \, exp\left(-\frac{r}{D}\right), \tag{2.3}$$

where f is the frequency and D^{-1} is the absorption coefficient. The absorption coefficient is inversely proportional to the wavelength and directly proportional to the frequency: $\lambda=v/f$, where v is the velocity. As a result, the higher frequencies (smaller wavelengths) attenuate more during the wave propagation than the lower frequencies (larger wavelengths). Accordingly, the seismic signal's spectrum varies with the distance from the source as the geological medium acts as a low pass filter (Lowrie, 2007).

For a given wavelength, D is proportional to Q. The greater Q value is the longer seismic energy travels with small absorption. Generally, the Q parameter for P waves (Q_P) is greater than the corresponding value for the S waves (Q_S) – Lowrie (2007). Therefore, the anelastic attenuation is mainly due to the shear component of the strain tensor, thus shear attenuation is much more important than bulk attenuation (Dahlen and Tromp, 1998). In solid materials with low rigidity moduli, the shear deformation can reach high values and, therefore, the attenuation is greater than the attenuation in materials with a high value of shear modulus. The Q factor exhibits high values in fluids, where the attenuation is low because the shear deformation is zero and only the compression waves propagate (Lowrie, 2007).

2.2 – Seismic wave equation: analytical background

The following description of the elastic wave equation and the numerical solution is based on the characterization given by Fichtner (2011). To maintain a coherency over the entire manuscript, some notations inside the formulas were modified.

2.2.1 – Displacement-stress formulation in elastic approximation

The seismic wave equation is the linearized version of Newton's second law, balancing the momentum of particle displacement, the internal stress and the external forces, corresponding to the source of seismic wave propagation (Aki and Richard, 2002; Fichtner, 2011, modified):

$$\rho(\mathbf{x})\frac{\partial^2}{\partial t^2}\mathrm{u}(\mathbf{x},t) - \nabla\cdot\sigma(\boldsymbol{x},t) = \mathrm{f}(\boldsymbol{x},t), \tag{2.4}$$

with $\boldsymbol{x} \in G \subset R^3$, $t \in [t_0, t_1] \subset R$, where u is the displacement field, ρ is the mass density, σ is the stress tensor and f is an external force. Some conditions need to be imposed (Fichtner, 2011):

- *The free boundary condition*: the normal component of the stress tensor σ vanishes at the Earth's surface: $\sigma \cdot n|_{x \,\epsilon\, \partial\, G} = 0$.
- *The initial condition*: the displacement and velocity fields need to be zero before the action of the external source: $\mathrm{u}|_{t<t_0} = \mathrm{v}|_{t<t_0} = 0$.

The complementary equation defines the viscoelastic rheology and connects the stress tensor σ with the displacement field u. Following the assumption that σ depends linearly on the strain tensor $\varepsilon = \frac{1}{2}(\nabla \mathrm{u} + \nabla \mathrm{u}^2)$, leads to (from Fichtner, 2011, modified):

$$\sigma_{ij}(\mathbf{x}, t) = \int_{-\infty}^{\infty} \frac{\partial}{\partial t} \mathrm{C}_{ijkl}\,(\mathbf{x}, t - t') * \varepsilon_{kl}\,(\mathbf{x}, t')dt' \tag{2.5}$$

where C is the (unrelaxed) elastic tensor (i.e. stiffness coefficient) - 4th order, which honors the causality condition $\mathrm{C}(t)|_{t<t_0} = 0$ and ε is the strain tensor – 2nd order. The components of C are characterized by the symmetry relation $\mathrm{C}_{ijkl} = \mathrm{C}_{klij} = \mathrm{C}_{jikl}$, according to the symmetry of ε, the conservation of angular momentum and the relation of the C with the internal energy (Aki and Richards, 2002; Fichtner, 2011). The symmetries of the C_{ijkl} coefficients reduce the number of its independent components to 21 for a fully anisotropic medium, allowing the equation (2.5) to be written in terms of displacement gradient ∇u (from Fichtner, 2011, modified):

$$\sigma_{ij}(\mathbf{x}, t) = \int_{-\infty}^{\infty} \frac{\partial}{\partial t} \mathrm{C}_{ijkl}(\mathbf{x}, t - t') * \nabla \mathrm{u}\,(\mathbf{x}, t')dt', \tag{2.6}$$

where " $*$ " is the convolution operator.

For an isotropic medium, the components of C can be expressed as linear combinations of two elastic moduli, the *Lamé parameter* λ and the *shear modulus* μ (from Fichtner, 2011): $\mathrm{C}_{ijkl} = \lambda\, \delta_{ij}\delta_{kl} + \mu\, \delta_{ik}\, \delta_{jl} + \mu\, \delta_{il}\, \delta_{jk}$. Often, λ is replaced by the *bulk modulus* $k = \lambda + 2/3$, which relates the strain tensor ε with the scalar pressure $p = -k\, \nabla \cdot \mathrm{u}$.

In elastic approximation (non-dissipative medium), the stress tensor formulation simplifies to (from Fichtner, 2011):

$$\sigma(\mathbf{x}, t) = \mathrm{C}(\mathbf{x}) : \nabla \mathrm{u}(\mathbf{x}, t), \tag{2.7}$$

where the double dots denote a contraction over two adjacent indices (e.g. $\mathrm{A} : \mathrm{B} = \sum_{i,j=1}^{n} \mathrm{A}_{ij}\mathrm{B}_{ij}$, $\mathrm{B}\, \epsilon\, \mathrm{R}^{\mathrm{nxn}}$).

The equations (2.4) and (2.7) represent the *displacement-stress formulation* in elastic approximation. To avoid the convolution present in the stress tensor's expression, many numerical implementations are based on the elastic approximation, while incorporating the dissipation into a less computation-demanding memory variable.

2.2.2 – Viscoelastic dissipation

The time dependency of the elastic tensor C in equations (2.5) and (2.6) describes the viscoelastic dissipation. The components of the elastic tensor can be also expressed as a superposition of standard linear solids (SLS) – i.e. the Zener model, controlled by relaxation parameters, related to frequency-dependent Q factors: $\omega_\nu \in [\omega_{min}, \omega_{max}]$. The inverse of the quality factor tensor $Q_{ijkl}(\mathbf{x})^{-1}$describes the seismic attenuation (Emmerich and Korn, 1987), which can be related to a memory variable acting on the stress field. The expression for the stress tensor becomes (e.g. Komatitsch and Tromp 1999; Trinh et al., 2019 – modified):

$$\sigma_{ij}(\mathbf{x}, t) = \mathrm{C}_{ijkl}(\mathbf{x}) * \varepsilon_{kl}(\mathbf{x}, t) - \mathrm{C}^{R}_{ijkl}(\mathbf{x}) \sum_{s=1}^{n} \psi_{\nu;\, k,l}(\mathbf{x}, t), \tag{2.8}$$

$$\text{with } \mathrm{C}^{R}_{ijkl} = \mathrm{C}_{ijkl}(\mathbf{x})\, Q^{-1}_{ijkl}(\mathbf{x}), \tag{2.9}$$

$$\text{and } \frac{\partial}{\partial t}\, \psi_{\nu;\, k,l}(\mathbf{x}, t) + \omega_\nu \psi_{\nu;\, k,l}(\mathbf{x}, t) = \omega_\nu y_\nu \varepsilon_{kl}(\mathbf{x}, t), \tag{2.10}$$

where C_{ijkl} are the unrelaxed coefficients of the stiffness tensor, C^{R}_{ijkl} are the relaxed stiffness coefficients, $\psi_{\nu;\, k,l}$ is the memory variable, responsible for the attenuation and y_ν (for $\nu = 1, \dots, n$) is a dimensionless anelastic coefficient, with n scalars connected to n quality factors Q_{ijkl} (Yang et al., 2016a); it is important to mention that here Q_{ijkl} does not refer to the Q field measurement.

2.3 – Numerical solution of the elastic wave equation

Analytical solutions for the seismic wave equation exist only for simple models, which however do not reflect the real structural complexity of the subsoil. For moderately heterogeneous media, perturbation methods can be used to find an approximate solution. However, FWI is focused above all on highly heterogeneous media, for which analytical or perturbation methods cannot be easily applicable (Fichtner, 2011).

The difficulty of finding analytical solutions for the wave propagation in heterogeneous media led to the design of numerical solutions, based on the partial derivatives discretization in space and time. The spatial discretization leads to a system of ordinary partial-derivatives non-homogeneous differential equations, which can be numerically solved both in the time domain and frequency domain.

2.3.1 – Wave equation discretization in space

All numerical methods have in common the discrete approximation in space of the continuous wavefield u(x, t), through a finite number of time-dependent coefficients $\bar{\mathrm{u}}_1(t), \dots\ \bar{\mathrm{u}}_N(t)$ that can be synthesized in an N-dimensional vector (Fichtner, 2011). Using the space discretization, the elastic wave equation is transformed into a differential equation that can be written in a form that allows eliminating the gradient (from Fichtner, 2011, modified):

$$\mathrm{M} \cdot \frac{\partial^2}{\partial t^2} \bar{\mathrm{u}}(t) + \mathrm{K} \cdot \bar{\mathrm{u}}(t) = \bar{\mathrm{f}}(t), \tag{2.11}$$

where M is the mass matrix, K is the stiffness matrix, $\bar{f}$ is the discrete version of the external source vector and $\bar{u}$ is the discrete version of the particle displacement. The mass and stiffness matrices are sparse. The expression for the space-discretized displacement-stress formulation is (from Fichtner, 2011, modified):

$$M \cdot \frac{\partial^2}{\partial t^2}\bar{u}(t) + K_1 \cdot \bar{s}(t) = \bar{f}(t)\,, \text{with } \bar{s}(t) = K_2 \cdot \bar{u}(t), \tag{2.12}$$

where: K_1 e K_2 are the stiffness matrices and $\bar{s}$ is the discrete approximation of the stress-tensor.

Viscoelastic equation

Using the Voight indexing to describe the strain and memory-variable vectors, the matrix $C = (C_{ij})_{6x6}$ contains the elastic-stiffness coefficients. The viscoelastic wave equation becomes (from Trinh et al., 2019, modified):

$$M \cdot \frac{\partial^2}{\partial t^2}\bar{u}(t) + K \cdot \bar{u}(t) - D\, C^R \sum_{s=1}^{L} \psi_s = \bar{f}(t), \tag{2.13}$$

$$\frac{\partial}{\partial t}\psi_s + \omega_s \psi_s = \omega_s\, y_s \varepsilon, \tag{2.14}$$

where D is a spatial derivative operator. In the spectral element discretization (described in *Appendix 1*), the spatial derivative operator is weighted by GLL weights D^w.

2.3.2 – Wave equation discretization in time or frequency

The system of differential equations in time, resulting from the spatial discretization of the wave equation, can be solved both in the time domain and in the frequency domain.

Displacement formulation in time-domain

The time-domain modeling is based on the substitution of the time derivatives with finite-difference approximations. This procedure allows representing the wavefield propagation for discrete time intervals Δt. The time-discretized expression for the displacement-stress formulation is (from Fichtner, 2011, modified):

$$\frac{\partial^2}{\partial t^2}\bar{u}(t) = M^{-1} \cdot \left[\bar{f}(t) - K \cdot \bar{u}(t)\right], \tag{2.15}$$

$$\frac{\partial^2}{\partial t^2}\bar{u}(t) \approx \frac{1}{\Delta t^2}\left[\bar{u}(t + \Delta t) - 2\,\bar{u}(t) + \bar{u}(t - \Delta t)\right], \tag{2.16}$$

where the equation (2.16) is the expression for the second-order finite difference approximation. Therefore, the relation that allows computing the displacement at time the $t + \Delta t$ from the displacement at time the $t - \Delta t$ is:

$$\bar{u}(t + \Delta t) \approx 2\,\bar{u}(t) + \bar{u}(t - \Delta t) + \Delta t^2\, M^{-1} \cdot \left[\bar{f}\,(t) - K \cdot \bar{u}(t)\right]. \tag{2.17}$$

An alternative to the relation (2.17) is the *Newmark scheme* (Newmark, 1959; Chaljub et al., 2007; Fichtner, 2011, modified):

$$\bar{u}(t + \Delta t) = \bar{u}(t) + \Delta t\,\bar{v}(t) + \Delta t^2 \left[\left(\tfrac{1}{2} - \beta\right)\frac{\partial^2}{\partial t^2}\bar{u}(t) + \beta \frac{\partial^2}{\partial t^2}\bar{u}(t + \Delta t)\right], \tag{2.18}$$

$$\bar{v}(t+\Delta t) = \bar{v}(t) + \Delta t \left[(1-\gamma)\frac{\partial^2}{\partial t^2}\bar{u}(t) + \gamma\frac{\partial^2}{\partial t^2}\bar{u}(t+\Delta t)\right], \quad (2.19)$$

where $\gamma \in [0,1]$ and $\beta \in [0,1/2]$. If $\gamma = 1/2$ and $\beta = 0$, the explicit central-difference scheme is obtained (Fichtner, 2011, modified):

$$\bar{u}(t+\Delta t) = \bar{u}(t) + \Delta t\,\bar{v}(t) + \tfrac{1}{2}\Delta t^2\frac{\partial^2}{\partial t^2}\bar{u}(t) \quad (2.20)$$

$$\bar{v}(t+\Delta t) = \bar{v}(t) + \tfrac{1}{2}\Delta t\left[\frac{\partial^2}{\partial t^2}\bar{u}(t) + \frac{\partial^2}{\partial t^2}\bar{u}(t+\Delta t)\right] \quad (2.21)$$

All time-discretization algorithms are explicit (Marfurt, 1984), in the sense that the dynamic field at time t depends only on the dynamic field at the previous time. These algorithms are stable only under certain conditions. The numerical stability criterion, called *CFL condition* (from the names of the authors who defined it: R. Courant, K. Friedrichs and H. Levy), limits the maximum time-increment and therefore the efficiency of each time sampling (Fichtner, 2011):

$$\Delta t \leq C\,\frac{min\,(h)}{max\,(v)}, \quad (2.22)$$

where h is the size of a grid's cell, v is the P-wave velocity and C is a constant depending on the specific discretization method in time and space.

Discretization in frequency domain

The Fourier transform of the expression for the space-discretization of the displacement is (Fichtner, 2011):

$$-\omega^2\,\mathrm{M}\cdot\bar{u}(\omega) + \mathrm{K}\cdot\bar{u}(\omega) = \bar{f}(\omega), \quad (2.23)$$

where $\mathrm{L}(\omega) = -\omega^2\mathrm{M} + \mathrm{K}$ is the impedance matrix, leading to a linear system $\mathrm{L}(\omega)\cdot\bar{u}(\omega) = \bar{f}(\omega)$, which can be solved using direct matrix factorization methods, such as LU decomposition (Pratt et al., 1998; Pratt, 1999; Press et al., 2007). Nevertheless, since the direct matrix factorization is memory demanding, the use of iterative solvers is often preferred (Quarteroni et al., 2000).

One important advantage of the frequency domain modeling is the easier and less memory-demanding implementation of the viscoelastic dissipation since the convolution in the stress tensor formulation becomes a product (Fichtner, 2011).

2.3.3 – Numerical methods overview

Several numerical methods are available for the solution of the seismic wave equation, such as:

- *Finite difference* - e.g. Alford et al. (1974); Kelly et al. (1976); Virieux (1984, 1986); Igel et al. (1995); Graves (1996).
- *Optimal operators* - e.g. Geller and Takeuchi (1995); Geller and Takeuchi (1998); Takeuchi and Geller (2000).
- *Pseudospectral method* - e.g. Furumura et al. (1988); Tessmer and Kosloff (1994); Igel, (1999).
- *Finite elements method* - e.g. Lysmer and Drake (1972); Bao et al. (1998); Moczo et al. (2007).
- *Spectral element method* - e.g. Patera (1984) - application for fluid mechanics; Seriani (1998); Komatitsch and Vilotte (1998); Komatitsch and Tromp (1999); Komatitsch and Tromp (2002a,b); Chaljub and Valette (2004); Chaljub et al. (2007); Nissen-Meyer et al. (2007, 2008); Tape et al. (2010); Peter et al. (2011).
- *Direct solution method* - e.g. Geller and Ohmiato (1994); Cummings et al., (1994a,b); Kawai and Geller (2010).
- *Discontinuous Galerkin method* – e.g. Dumbser and Käser (2006); de la Puente et al. (2007, 2008, 2009); Brossier et al. (2009b).

Among them, the finite difference (FD) method is the most frequently used in FWI. It is based on the approximation of the partial derivatives with incremental ratios, which may lead to numerical dispersion of the signal (Alford et al., 1974; Fichtner, 2011). In the presence of numerical dispersion, the various frequencies of a signal propagate at different velocities and, therefore, the velocity of every single phase is different from the group velocity (Telford et al., 2009). Consequently, the numerical solution may be "slower" than the analytical one, but it can also be faster at some frequencies due to the numerical dispersion introduced by time discretization.

The numerical dispersion depends, among other factors, on the number of grid points covered by the smallest wavelength (corresponding to the highest frequency). The choice of the optimum grid points' number per wavelength is highly dependent on the approximation order. Increasing the order of the finite difference operator generally leads to more accurate results. The error decreases rapidly as the dominant wavelength increases, or the grid spacing decreases. The first implementations of the finite difference method were based on the conventional grid, where all the field variables (stress, strain, displacement) were defined in the same positions. Once the finite difference method has been developed in 3D, a new approach, based on the use of a staggered grid, was defined (e.g. Virieux, 1984; 1986; Hustedt et al., 2004); in this case, the dynamic fields are evaluated in different positions. This geometry leads to a smaller spacing between the samples, reducing the numerical dispersion (Fichtner, 2011).

The spectral element method (SEM) is also used in many FWI tools, especially for the applications involving irregular geometries and complex structures. Based on the weak form of the wave equation, it involves the decomposition of the computational domain into non-overlapping elements that are mapped to a reference space. Within each element in the reference space, the dynamic fields, represented by Lagrange polynomials, are collocated at the Gauss-Lobatto-Legendre (GLL) points.

The SEM solution's accuracy is controlled by the size of the elements and by the degree of the Lagrange polynomials. Since the GLL quadrature is exact only for polynomials of degree (*2N-1*), the SE discretization, where the integrands are products of polynomials of degree *N,* is not exact. Increasing the polynomial degree helps for a better representation of complex geometries, although it is limited by the CFL condition. In practice, polynomial degrees of 4 to 7 and at least 5 grid points per minimum wavelength are used to ensure a reasonable numerical accuracy, when the number of propagating wavelengths varies from 10 to 50 (Fichtner, 2011).

Many successful applications of the spectral element method (SEM) exist for heterogeneous media with complex geometries (e.g. Komatitsch and Vilotte, 1998; Seriani, 1998; Komatitsch et al., 2004; Fichtner et al., 2009; Trinh et al., 2019). The SEM is computationally advantageous because of the weak solution of the wave equation (which implicitly contains, in the stiffness matrix, the free-surface condition) and of the possibility of adapting the mesh's shape and dimension to irregular topographies and variable wavelengths inside the medium. A review of the SEM can be found, for example, in Komatitsch et al. (2005) and Chaljub et al. (2007).

Overall, the most important advantages of the SEM are:

- The implicit implementation of the free-surface boundary condition in the stiffness matrix of the wave equation weak form.
- The mesh's adaptability to irregular topographies and complex structures (variable wavelengths) and the facilitation for parallel implementation.
- The mass matrix is diagonal in the spectral element discretization: This property allows computing the explicit formula for the second-order time derivatives of the wavefield coefficients, with limited computational resources.
- The direct computation of the stiffness matrix is not necessary: only the computation of its product with the expansion coefficients vector is normally required.

Since the FWI tool used in this study is based on the spectral element method, a brief description of this method is given in *Appendix 1*, following Fichtner (2011). A more detailed presentation of the mathematical background of SEM, including the orthogonal polynomials, the function interpolation (Lagrange, Lobatto) and the numerical integration (Gauss-Lobatto-Legendre quadrature) can be found in Fichtner (2011), Appendix A, pp 301-315. Other detailed treatments can be also found, for example, in Quarteroni et al. (2000), Karniadakis and Sherwin (2005), Zienkiewicz et al. (2005).

2.4 – Highlights

- When an external force is activated in a certain volume, some energy propagates inside the volume (as BWs: pressure waves and shear waves), while some energy propagates near the surface (as SWs: Rayleigh waves and Love waves).
- The velocity of seismic energy propagation inside a medium is known as seismic wave velocity. The latter is different from the medium's particles displacement (particle velocity) that are moving as a consequence of the seismic wave propagation.
- In a heterogeneous medium, SWs exhibit prominent dispersive behavior: the phase velocity varies with the frequency (wavelength).
- The seismic energy decreases with the distance from the source, due to various factors as geometrical spreading and anelastic damping. The BWs amplitude attenuation due to geometrical spreading varies more quickly with the distance ($1/r$) than the SWs amplitude ($1/\sqrt{r}$). The attenuation caused by anelastic damping is described by a parameter, called quality factor Q. The greater Q value is the less seismic energy is subject to attenuation.
- The seismic wave equation is the linearized version of Newton's second law. It is formulated as a second-order partial-derivatives nonhomogeneous differential equation, balancing the momentum of particle displacement, the internal stress and the external forces, corresponding to the seismic wave source.
- The complementary equation in the stress-displacement formulation defines the viscoelastic rheology and connects the stress tensor with the displacement field. The time dependency of the elastic tensor (convolution term) describes the viscoelastic dissipation. The components of the elastic tensor can be expressed as a superposition of standard linear solids, controlled by relaxation parameters, related to frequency-dependent Q factors. To avoid convolution, many FWI implementations are based on the elastic approximation for the wave equation, while incorporating the dissipation into a memory variable.
- For the real data applications, involving highly heterogeneous models, the analytical solutions for the seismic wave equation are substituted by numerical solutions, based on the spatial discretization of the partial derivatives. Such discretization over the space leads to a system of ordinary partial-derivatives non-homogeneous differential equations, which can be solved numerically both time domain and frequency domain.
- There are several numerical methods for the solution of the seismic wave equation, such as finite difference, optimal operators, pseudospectral method, finite elements, spectral element, direct solution, discontinuous Galerkin, etc. Among them, the finite difference method is the most frequently used, due to the staggered grid efficient implementation and the consequent relatively low necessary computer memory. Nevertheless, in the presence of complex geometries and variable topography, the spectral element method (SEM) is often preferred.

- In the SEM, the continuous domain is decomposed into a set of non-overlapping elements. The equations governing each element are then assembled into a global system of equations through a continuity condition imposed to neighboring elements. All elements are mapped to a unitary 3D space, using shape functions and collocation nodes. Tensorized Lagrange polynomials are often used as basis functions in the 3D SEM, leading to the elements' representation through hexahedra.

Appendix 1

Spectral element method insights

The following presentation is based on the description provided by Fichtner (2011).

A 1.1 - The weak form of the seismic wave equation

The strong form of the wave equation in 1D is (from Fichtner, 2011):

$$(\rho)\frac{\partial^2}{\partial t^t}\mathrm{u}(\mathrm{x},t) - \frac{\partial}{\partial x}\left[\mu(\mathrm{x})\frac{\partial}{\partial x}\mathrm{u}(\mathrm{x},t)\right] = \mathrm{f}(\mathrm{x},t), \tag{A1.1}$$

where the space variable $\mathrm{x}\,\epsilon\, G = [0, L]$, the time variable $t\,\epsilon\,[0, T]$, and the displacement field u is subject to the Neumann boundary conditions:

$$\frac{\partial}{\partial x}\mathrm{u}(\mathrm{x},t)|_{x=0} = \frac{\partial}{\partial x}\mathrm{u}(\mathrm{x},t)|_{\mathrm{x}=L} = 0, \tag{A1.2}$$

and to the initial conditions: $\mathrm{u}|_{t=0} = \frac{\partial \mathrm{u}}{\partial t}|_{t=0} = 0.$ (A1.3)

The weak form of the wave equation can be derived multiplying the strong form by an arbitrary *test function* w ($\epsilon\, G \rightarrow R$), which is time-dependent, and integrating over the space G (from Fichtner, 2011):

$$\int_0^G \rho\, w \frac{\partial^2}{\partial t^t}\mathrm{u}\, dx - \int_0^G w \frac{\partial}{\partial x}\left(\mu \frac{\partial}{\partial x}\mathrm{u}\right) dx = \int_0^G w\, \mathrm{f}\, dx. \tag{A1.4}$$

Integrating the left-hand side by parts and honoring the boundary condition (A1.2), leads to (from Fichtner, 2011):

$$\int_0^G \rho\, w \frac{\partial^2}{\partial t^t}\mathrm{u}\, dx + \int_o^G \mu \frac{\partial}{\partial x} w \frac{\partial}{\partial x}\mathrm{u}\, dx = \int_0^G w\, \mathrm{f}\, dx. \tag{A1.5}$$

The boundary condition (A1.2) corresponds to the free surface condition and it is implicitly satisfied. The solution of the weak form of the wave equation is the wavefield u that satisfies the above equation, for any test function w, under the initial conditions (from Fichtner, 2011):

$$\int_0^G \rho\, w\, \mathrm{u}\,|_{t=0}\, dx = \int_o^g \rho\, w \frac{\partial}{\partial t}\mathrm{u}\,|_{t=0}\, dx = 0. \tag{A1.6}$$

When both density ρ and the elastic parameter μ are variable in space, finding analytical solutions for the (strong and weak form of the) wave equation is almost impossible. Therefore, also for the weak form of the wave equation, numerical solutions, based on wavefield discretization, are designed.

A 1.2 – Discretization in space and the Galerkin method

The *Galerkin method* is commonly used for the space discretization. It is based on the approximation of the exact solution $\mathrm{u}(x,t)$ by a finite superposition of n so-called space-dependent *basis functions* ψ_i $(i = 1, \ldots, n)$ (from Fichtner, 2011):

$$\mathrm{u}(\mathrm{x},t) \approx \bar{\mathrm{u}}(\mathrm{x},t) = \sum_{i=1}^{n} \mathrm{u_i}(t)\, \psi_i(\mathrm{x}), \tag{A1.7}$$

where $u_i(t)$ are the time-dependent expansion coefficients. The basis functions depend only on space and they are used as *test functions* in the weak form of the wave equation. Therefore, the solution of the wave equation requires the wavefield $\bar{u}$ to satisfy the approximate weak form (from Fichtner, 2011):

$$\int_0^G \rho\, \psi_i\, \frac{\partial^2}{\partial t^2}\bar{u}\, dx + \int_o^G \mu \frac{\partial}{\partial x}\psi_i \frac{\partial}{\partial x}\bar{u}\, dx = \int_0^G \psi_i\, f\, dx, \tag{A1.8}$$

for all basis functions ψ_i under the initial conditions:

$$\int_0^G \rho\, \psi_i\, \bar{u}|_{t=0}\, dx = \int_o^g \rho\, \psi_i \frac{\partial}{\partial t}\bar{u}\,|_{t=0}\, dx = 0. \tag{A1.9}$$

Using the equations (A1.7) and (A1.8), the *Galerkin projection* transforms the weak form into a linear equation for the coefficients $u_i(t)$ (from Fichtner, 2011):

$$\sum_{i=1}^n \left[\frac{\partial^2}{\partial t^2} u_i(t) \int_o^G \rho(x)\, \psi_j(x)\, \psi_i(x)\, dx\right] + \sum_{i=1}^n \left[u_i(t) \int_o^G \mu(x) \frac{\partial}{\partial x}\psi_j(x)\, \frac{\partial}{\partial x}\psi_i(x)\, dx\right] = \int_0^G \psi_j(x)\, f(x,t)\, dx, \tag{A1.10}$$

for $j = 1, \ldots, n$. The differential equation (A1.10) can be written in matrix notation as (from Fichtner, 2011):

$$M \cdot \frac{\partial^2}{\partial t^2}\, u(t) + K \cdot u(t) = f(t), \tag{A1.11}$$

where the vector u contains the expansion coefficients u_i,

the mass matrix is: $M_{i,j} = \int_0^G \rho(x)\, \psi_j(x)\, \psi_i(x)\, dx$, (A1.12)

the stiffness matrix is: $K_{j,i} = \int_0^G \mu(x)\, \frac{\partial}{\partial x}\psi_j(x) \frac{\partial}{\partial x}\psi_i(x)\, dx$, (A1.13)

and the source term is: $f_i(t) = \int_0^G \psi_j(x)\, f(x,t)\, dx$. (A1.14)

The stiffness matrix K implicitly contains the free-surface boundary condition. This fact renders the SEM numerically more convenient than other methods (i.e. finite difference) that require additional implementation for the free-surface condition.

In the SEM, the continuous G domain is decomposed into a set of n_e non-overlapping *elements* G_e. The size of each element is usually chosen proportional to the wave velocity, to ensure a relatively uniform sampling of the wavelength. The element decomposition transforms the equation (A1.10) into (from Fichtner, 2011):

$$\sum_{i=1}^n \left[\frac{\partial^2}{\partial t^2} u_i(t) \sum_{e=1}^{n_e} \int_o^{G_e} \rho(x)\, \psi_j(x)\, \psi_i(x)\, dx\right] + \sum_{i=1}^n \left[u_i(t) \sum_{e=1}^{n_e} \int_o^{G_e} \mu(x) \frac{\partial}{\partial x}\psi_j(x) \frac{\partial}{\partial x}\psi_i(x)\, dx\right] = \sum_{e=1}^{n_e} \int_0^{G_e} \psi_j(x)\, f(x,t)\, dx, \tag{A1.15}$$

where each expansion coefficient u_i depends on the integrals over all elements. To overcome this dependency related inconvenient, *N+1* local *basis functions* ψ_i^e ($e = 1, \ldots, N+1$), each one related one element n_e of the G_e space, are defined, and the discrete equation is solved for each element individually. The displacement field for the G_e elements is approximated by (from Fichtner, 2011):

$$\bar{u}(x,t)|_{x \in G_e} = \sum_{i=1}^{N+1} u_i^e(t)\, \psi_i^e(x), \tag{A1.16}$$

and the equation (A1.15) becomes (from Fichtner, 2011, modified):

$$\sum_{i=1}^{N+1} \frac{\partial^2}{\partial t^t} \mathrm{u}_i^e(t) \int_o^{G_e} \rho(\mathrm{x})\, \psi_j^e(\mathrm{x})\, \psi_i^e(\mathrm{x})\, dx + \sum_{i=1}^{N+1} \mathrm{u}_i^e(t) \int_o^{G_e} \mu(\mathrm{x}) \frac{\partial}{\partial x} \psi_j^e(\mathrm{x}) \frac{\partial}{\partial x} \psi_i^e(\mathrm{x})\, dx = \int_0^{G_e} \psi_j^e(\mathrm{x})\, f(\mathrm{x}, t)\, dx, \quad \text{(A1.17)}$$

whose corresponding matrix notation is:

$$\mathrm{M^e} \cdot \frac{\partial^2}{\partial t^t} \mathrm{u^e}(t) + \mathrm{K^e} \cdot \mathrm{u}^e(t) = \mathrm{f}^e(t), \quad e = 1, \dots, n_e, \quad \text{(A1.18)}$$

where $\mathrm{u^e}$, $\mathrm{M^e}$ and $\mathrm{K^e}$ are the local coefficient vector, mass matrix and stiffness matrix, respectively.

To satisfy the discrete approximation of the wavefield $\bar{\mathrm{u}}$, the implementation of a continuity condition between all basis functions inside each element n_e is required. To allow all integrals to be managed in the same way, each element G_e is mapped into a *standard reference interval* $[-1, 1]$, via an element transformation (from Fichtner, 2011):

$$F_e : [-1, 1] \rightarrow G_e, \quad \mathrm{x} = F_e(\xi), \quad \xi = \xi(\mathrm{x}) = F_e^{-1}(x), \quad e = 1, \dots, n_e \quad \text{(A1.19)}$$

Introducing the above transformation into the equation (A1.17), gives (from Fichtner, 2011):

$$\sum_{i=1}^{N+1} \frac{\partial^2}{\partial t^t} \mathrm{u}_i^e(t) \int_{-1}^{1} \rho[\mathrm{x}(\xi)]\ \psi_j^e[\mathrm{x}(\xi)]\ \psi_i^e[\mathrm{x}(\xi)]\ \frac{dx}{d\xi} d\xi + \sum_{i=1}^{N+1} \mathrm{u_i^e}(t) \int_{-1}^{1} \mu[\mathrm{x}(\xi)]\ \frac{d}{d\xi} \psi_j^e[\mathrm{x}(\xi)]\ \frac{d}{\partial \xi} \psi_i^e[\mathrm{x}(\xi)] \left(\frac{dx}{d\xi}\right)^2 \left(\frac{dx}{d\xi}\right) d\xi = \int_{-1}^{1} \psi_j^e[\mathrm{x}(\xi)]\ \mathrm{f}\{[\mathrm{x}(\xi)], t\}\ \frac{dx}{d\xi}\ d\xi. \quad \text{(A1.20)}$$

The basis functions definition is based on the choice of *N+1* Lagrange polynomials of degree *N*, having the corresponding Gauss-Lobatto-Legendre (GLL) points collocated as $\psi_i^e[\mathrm{x}(\xi)] = l_i^{(N)}(\xi)$, $\xi \in [-1, 1]$. Since the GLL points are the collocation points of the GLL quadrature, the GLL quadrature formulas allow obtaining accurate approximations of the integrals in equation (A1.20). Substituting $l_i(\xi)$ for *ψ[x(ξ)]* and omitting the superscript $^{(N)}$ in $l_i^{(N)}$, leads to (from Fichtner, 2011):

$$\sum_{i=1}^{N+1} \frac{\partial^2}{\partial t^t} \mathrm{u_i^e}(t) \int_{-1}^{1} \rho'(\xi)\, l_j(\xi)\, l_i(\xi) \frac{dx}{d\xi} d\xi + \sum_{i=1}^{N+1} \mu'(\xi) \frac{\partial}{\partial \xi} l_j(\xi)\ \frac{\partial}{\partial \xi} l_i(\xi) \left(\frac{dx}{d\xi}\right)^2 \left(\frac{dx}{d\xi}\right) d\xi = \int_{-1}^{1} l_j(\xi)\, \mathrm{f}'(\xi, t) \frac{dx}{d\xi}\, d\xi, \quad \text{(A1.21)}$$

where the transformed density ρ', the elastic moduli μ' and the external force f' are defined by: $\rho'(\xi) = \rho\,[\mathrm{x}(\xi)]$, $\mu'(\xi) = \mu[\mathrm{x}(\xi)]$, $\mathrm{f}'(\xi) = f[\mathrm{x}(\xi)]$. (A1.22)

Using the GLL quadrature formula: $\int_{-1}^{1} \mathrm{f(x)} dx \approx \int_{-1}^{1} P_N(\mathrm{x}) dx = \sum_{i=1}^{N+1} w_i\, \mathrm{f}(\mathrm{x}_i)$, (A1.23)

where f(x) is and arbitrary function, $P_N(\mathrm{x})$ is an interpolating polynomial and $w_i = \int_{-1}^{1} l_i^{(N)}(\mathrm{x})\, dx$ are the integration weights, the integrals in (A1.21) can be approximated as (from Fichtner, 2011):

$$\sum_{i,k=1}^{N+1} \frac{\partial^2}{\partial t^t} \mathrm{u}_{\mathrm{i}}^{\mathrm{e}}(t)\, w_k\, \rho'(\xi)\, l_j(\xi)\, l_i(\xi)\, \frac{dx}{d\xi}|_{\xi=\xi_k} +$$
$$\sum_{i,k=1}^{N+1} w_k\, \mathrm{u}_{\mathrm{i}}^{\mathrm{e}}(t)\, \mu'(\xi) \frac{\partial}{\partial \xi} l_j(\xi) \frac{\partial}{\partial \xi} l_i(\xi) \left(\frac{dx}{d\xi}\right)^2 \frac{dx}{d\xi}|_{\xi=\xi_k} \approx \sum_{k=1}^{N+1} w_k\, l_j(\xi)\, f'(\xi, t) \frac{dx}{d\xi}|_{\xi=\xi_k}. \quad \text{(A1.24)}$$

The symbols ξ_k and w_k are the GLL points and their corresponding integration weights. The numerical integration in the above equation is not exact because the integrands are not polynomials of degree *2N – 1* or lower (Fichtner, 2011). Considering the interpolation property of the Lagrange polynomials $l_i(\xi_k) = \delta_{ik}$, the equation (A1.24) can be written in a more compact form as (from Fichtner, 2011):

$$\sum_{i=1}^{N+1} \mathrm{M}_{\mathrm{ji}}^{\mathrm{e}} \frac{\partial^2}{\partial t^t} \mathrm{u}_{\mathrm{i}}^{\mathrm{e}}(t) + \sum_{i=1}^{N+1} \mathrm{K}_{\mathrm{ji}}^{\mathrm{e}}\, \mathrm{u}_{\mathrm{i}}^{\mathrm{e}}(t) = \mathrm{f}_{\mathrm{i}}^{\mathrm{e}}(t)\,,\ \mathrm{e} = 1, \ldots . \mathrm{n}_{\mathrm{e}},\ \text{where} \quad \text{(A1.25)}$$

$$\mathrm{M}_{\mathrm{ji}}^{\mathrm{e}} = w_j\, \rho'(\xi)\, \frac{dx}{d\xi}\, \delta_{ik}|_{\xi=\xi_j}, \quad \text{(A1.26)}$$

$$\mathrm{K}_{\mathrm{ji}}^{\mathrm{e}} = \sum_{k=1}^{N+1} w_k\, \mu'(\xi)\, \frac{\partial}{\partial \xi} l_j(\xi)\, \frac{\partial}{\partial \xi} l_i(\xi)\, \left(\frac{dx}{d\xi}\right)^2 \frac{dx}{d\xi}|_{\xi=\xi_k}, \quad \text{(A1.27)}$$

$$\mathrm{f}_{\mathrm{i}}^{\mathrm{e}}(\mathrm{t}) = w_j\, f'(\xi, t)\, \frac{dx}{d\xi}|_{\xi=\xi_j}. \quad \text{(A1.28)}$$

An important numerical advantage of the spectral element discretization is the property of the mass matrix M^{e} of being diagonal, which renders the inversion of $\mathrm{M}_{\mathrm{ji}}^{\mathrm{e}}$ computationally convenient and allows obtaining explicitly the formula for the second-order time derivative of $\mathrm{u}_{\mathrm{i}}(t)$. Another advantage is the non-necessity of computing the stiffness matrix: only the computation of its product with the expansion coefficients vector is normally required.

Equation (A1.24) refers to the local numerical integration, and the resulting linear system in equation (A1.25) is also related to a local element. The GLL points of an element can be addressed by exploiting a local numbering scheme, in which the points shared between neighboring elements are counted twice (because they are considered in the discretized version of two integrals). However, the continuity condition of the approximation $\bar{\mathrm{u}}$ across the element boundaries needs to be satisfied. This process is guaranteed by assembling a global system of equations, where a global numbering scheme counts each GLL point only once, including the point shared between neighboring elements (from Fichtner, 2011, modified):

$$\mathrm{M}^{gl} \cdot \frac{\partial^2}{\partial t^2} \mathrm{u}^{gl}(t) + \mathrm{K}^{gl} \cdot \mathrm{u}^{gl}(t) = \mathrm{f}^{gl}(t), \quad \text{(A1.29)}$$

where M^{gl} is the global stiffness matrix (obtained by associating the local matrices M^{e}), K^{gl} is the global stiffness matrix (resulted from the association of the local stiffness matrices K^{e}), and u^{gl} is the global displacement vector. The global values are obtained by summing the local elements of the matrices at coincident node points.

A 1.3 – Mesh generation

The mesh generation is related to the subdivision of the domain G into n_e non-overlapping elements G_e proportional to the seismic wave velocity and subject to the condition that $G = \cup_{i=1}^{n_e} G_e$. In a 3D implementation, all elements G_e are mapped to the unit cube $X = [-1, 1]^3$, using *shape functions* and *anchor nodes* (Fichtner, 2011).

In practice, each element G_e is defined by n_a anchor nodes x^a (normally coincident with the 8 corners of each element but depending on the geometry and the approximation order) and their corresponding shape functions N^a. A position vector x in G_e is related to a position vector ξ in the reference cube X, via a transformation that relates each anchor node x^a of the physical element G_e to a corresponding anchor node ξ^a in the reference cube X (Fichtner, 2011):

$$\mathrm{x}(\xi^a) = x^a, \text{where } \mathrm{x}(\xi) = \mathrm{F}_e(\xi) = \sum_{a=1}^{n_a} N^a(\xi)\,\mathrm{x}^a \text{ and } N^a(\xi^b) = \delta_{a,b}. \quad \text{(A1.30)}$$

The shape functions N^a are commonly defined as a product of three Lagrange polynomials, with variable degree according to the element's complexity. Their collocation points are the coordinates ξ_i^a of the anchor nodes ξ^a in the reference cube X (Fichtner, 2011):

$$N^a(\xi) = l_{1,a}(\xi_1)\, l_{2,a}(\xi_2)\, l_{3,a}(\xi_3), \text{ with } l_{i,a}(\xi_i) = \delta_{a,b}. \quad \text{(A1.31)}$$

When tensorized Lagrange polynomials are used as basis functions in 3D spectral elements, the mesh's elements are represented by hexahedra (Fichtner, 2011).

A 1.4 – Spectral element method for 3D geometries

Weak solution of the elastic wave equation

The weak form for the equation of motion in 3D is obtained by multiplying the strong displacement-stress formulation $\rho(x)\frac{\partial^2}{\partial t^2}\mathrm{u}(x,t) - \nabla\cdot\sigma(x,t) = \mathrm{f}(x,t)$ and $\sigma(\mathrm{x},t) = \mathrm{C}(\mathrm{x}):\nabla\mathrm{u}(\mathrm{x},t)$ by a time-dependent test function w, and integrating over the 3D space G (from Fichtner, 2011):

$$\int_0^G \rho\,\mathrm{w}\cdot\frac{\partial^2}{\partial t^2}\,d^3\mathrm{x} - \int_0^G w\cdot(\nabla\cdot\sigma)\,d^3\mathrm{x} = \int_0^G w\cdot\mathrm{f}\,d^3\mathrm{x}. \quad \text{(A1.32)}$$

Using the identity $\mathrm{w}\cdot(\nabla\cdot\sigma) = \nabla\cdot(\mathrm{w}\cdot\sigma) - \nabla\mathrm{w}:\sigma$, together with the Gauss theorem, equation A1.32 becomes (from Fichtner, 2011):

$$\int_0^G \rho\,\mathrm{w}\cdot\frac{\partial^2}{\partial t}\mathrm{u}\,d^3\mathrm{x} - \int_0^{\partial G}\mathrm{w}\cdot\sigma\cdot\mathrm{n}\,d^2\mathrm{x} + \int_0^G \nabla\mathrm{w}:\sigma\,d^3\mathrm{x} = \int_0^G \mathrm{w}\cdot\mathrm{f}\,d^3\mathrm{x} \quad \text{(A1.33)}$$

Inserting the free surface boundary condition, equation (A1.33) becomes (from Fichtner, 2011):

$$\int_0^G \rho\,\mathrm{w}\cdot\frac{\partial^2}{\partial t^2}\mathrm{u}\,d^3\mathrm{x} + \int_0^G \nabla\mathrm{w}:\sigma\,d^3\mathrm{x} = \int_0^G \mathrm{w}\cdot\mathrm{f}\,d^3\mathrm{x}\,. \quad \text{(A1.34)}$$

The weak solution for the equation (A1.34) is the displacement field u that satisfies the above integral relation and the initial conditions (from Fichtner, 2011):

$$\int_0^G \rho\,\mathrm{w}\cdot\mathrm{u}|_{t=0}\,d^3\mathrm{x} = \int_0^G \rho\,\mathrm{w}\cdot\frac{\partial}{\partial t}\mathrm{u}|_{t=0}\,d^3\mathrm{x} = 0. \quad \text{(A1.35)}$$

Galerkin discretization

Approximating the p-components of the $\mathrm{u_p}$ displacement field u by a superposition of $(N+1)^3$ basis functions inside an element $G_e \in R^3$, gives (Fichtner, 2011):

$$\psi_{ijk}(\mathrm{x}) = \psi_{ijk}(x_1, x_2, x_3), \tag{A1.36}$$

weighted by the expansion coefficients $\mathrm{u_p^{ijk}}$:

$$\mathrm{u_p}(\mathrm{x}, t) \approx \bar{\mathrm{u}}_\mathrm{p}(\mathrm{x}, t) = \sum_{i,j,k=1}^{N+1} \mathrm{u_p^{ijk}}(t)\, \psi_{ijk}(\mathrm{x}). \tag{A1.37}$$

The approximation of the stress tensor components σ_{pq} is (Fichtner, 2011):

$$\sigma_{\mathrm{pq}}(\mathrm{x}, t) \approx \bar{\sigma}_{\mathrm{pq}}(\mathrm{x}, t) = \sum_{i,j,k=1}^{N+1} \sigma_{\mathrm{pq}}^{\mathrm{ijk}}(t)\, \psi_{ijk}(\mathrm{x}). \tag{A1.38}$$

The discretized formulation of the weak solution $\bar{\mathrm{u}}$ needs to satisfy the relations (from Fichtner, 2011):

$$\int_0^{G_e} \rho\, \psi_{ijk}\, \mathrm{e}_p \cdot \frac{\partial^2}{\partial t^2} \bar{\mathrm{u}}\, d^3\mathrm{x} + \int_0^{G_e} \nabla(\psi_{ijk} \mathrm{e}_p) : \bar{\sigma}\, d^3\mathrm{x} = \int_0^{G_e} \psi_{ijk} \mathrm{e}_p \cdot \mathrm{f}\, d^3\mathrm{x}\,, \text{ and} \tag{A1.39}$$

$$\int_0^{G_e} \psi_{ijk}\, \mathrm{e}_p \cdot \bar{\sigma}\, d^3\mathrm{x} = \int_0^{G_e} \psi_{ijk} \mathrm{e}_p \cdot \mathrm{C} : \nabla \bar{\mathrm{u}}\, d^3\mathrm{x}\,, \tag{A1.40}$$

and the initial conditions for the weak formulation:

$$\int_0^{G_e} \rho\, \psi_{ijk}\, \mathrm{e}_p \cdot \bar{\mathrm{u}}\,|_{t=0}\, d^3\mathrm{x} = \int_0^{G_e} \rho\, \psi_{ijk} \mathrm{e}_p \frac{\partial}{\partial t} \bar{\mathrm{u}}|_{t=0}\, d^3\mathrm{x} = 0, \tag{A1.41}$$

for all basis functions ψ_{ijk} and all unit vectors e_p, with $p = 1,2,3$.

a) The <u>first term</u> in the left-hand side of (A1.39) is transformed as follows (from Fichtner, 2011):

$$F_{qrs}\left[\rho \frac{\partial^2}{\partial t^2} \mathrm{u_p}\right] = \int_0^{G_e} \rho\, \psi_{ijk}\, \mathrm{e}_p \cdot \frac{\partial^2}{\partial t^2} \bar{\mathrm{u}}\, d^3\mathrm{x} =$$
$$\sum_{i,j,k=1}^{N+1} \int_0^{G_e} \rho\,(\mathrm{x}) \frac{\partial^2}{\partial t^2} \mathrm{u_p^{ijk}}(t)\, \psi_{ijk}(\mathrm{x})\, \psi_{qrs}(\mathrm{x}) d^3\mathrm{x}\,, \tag{A1.42}$$

where $\mathrm{F_{qrs}}\left[\rho \frac{\partial^2}{\partial t^2} \mathrm{u_p}\right]$ is a discrete local force, averaged over the element G_e.

Further, the element G_e $(e = 1, \ldots, n_e)$ is mapped to the reference cube $X = [-1, 1]^3$ by the specific transformation (Fichtner, 2011):

$$\mathrm{F}_e : [-1, 1]^3 = X \rightarrow G_e\,, \mathrm{x} = \mathrm{F}_e(\xi),\ \xi = \xi(\mathrm{x}) = \mathrm{F}_e^{-1}(\mathrm{x}), e = 1, \ldots, n_e, \tag{A1.43}$$

leading to (from Fichtner, 2011):

$$\mathrm{F_{qrs}}\left[\rho \frac{\partial^2}{\partial t^2} \mathrm{u_p}\right] = \sum_{i,j,k=1}^{N+1} \int_0^X \rho[\mathrm{x}(\xi)]\ \frac{\partial^2}{\partial t^2} \mathrm{u_p^{ijk}}(t)\, \psi_{ijk}[\mathrm{x}(\xi)]\, \psi_{qrs}[\mathrm{x}(\xi)]\, \mathrm{J}(\xi)\, d^3\xi\,, \tag{A1.44}$$

where the symbol J denoted the Jacobian of F_e. The elements are chosen such that $\mathrm{J} > 0$.

All basis functions $\psi_{ijk}[\mathrm{x}(\xi)]$ are related to the product of three Lagrange polynomials, collocated at the GLL points (Fichtner, 2011):

$\psi_{ijk}[\mathrm{x}(\xi)] = l_i(\xi_1)\, l_j(\xi_2)\, l_k(\xi_3)$, leading to (Fichtner, 2011): (A1.45)

$$\mathrm{F_{qrs}}\left[\rho \frac{\partial^2}{\partial t^2} \mathrm{u_p}\right] = \sum_{i,j,k=1}^{N+1} \int_0^X \rho'(\xi)\ \frac{\partial^2}{\partial t^2} \mathrm{u_p^{ijk}}(t)\ \ l_i(\xi_1)\, l_j(\xi_2) l_k(\xi_3)\ \ l_q(\xi_1)\, l_r(\xi_2)\, l_s(\xi_3)\, \mathrm{J}(\xi)\, d^3\xi\ , \tag{A1.46}$$

where $\rho'(\xi) = \rho[\mathrm{x}(\xi)]$ is the transformed density. Applying the GLL quadrature rule to equation (A1.46) yields to (from Fichtner, 2011):

$$\mathrm{F_{qrs}}\left[\rho\frac{\partial^2}{\partial t^2}\mathrm{u_p}\right]= w_q\, w_r\, w_s\, \rho'(\xi^{qrs})\frac{\partial^2}{\partial t^2}\mathrm{u_p^{qrs}}\,\mathrm{J}(\xi^{qrs}), \tag{A1.47}$$

where the transformed density and the Jacobian are evaluated at the GLL points:

$$\xi^{qrs} = (\xi_1^q, \xi_2^r\, \xi_3^s). \tag{A1.48}$$

The equation (A1.47) is the 3D equivalent of the mass matrix for the 1D case. The combination of GLL quadrature as Lagrange polynomials collocated at GLL points ensures the diagonality of the mass matrix in 3D (Fichtner, 2011).

b) The second term of the left-hand side of (A1.39) is transformed as follows (Fichtner, 2011):

$$F_{qrs}\left[(\nabla\cdot\sigma)_p\right]= \int_0^{G_e}\nabla(\psi_{ijk}\mathrm{e}_p):\bar{\sigma}\, d^3\mathrm{x}. \tag{A1.49}$$

Transforming the reference cube *X* and substituting the Lagrange polynomials for the basis functions, gives (Fichtner, 2011):

$$F_{qrs}\left[(\nabla\cdot\sigma)_p\right]=\sum_{n,m=1}^{3}\int_0^X \frac{\partial\xi_m}{\partial x_n}\frac{\partial}{\partial\xi_m}\left[l_q(\xi_1)\, l_r(\xi_2)\, l_s(\xi_3)\right]\bar{\sigma}'_{np}(\xi)\,\mathrm{J}(\xi)\, d^3\xi\,, \tag{A1.50}$$

where $\bar{\sigma}'(\xi) = \bar{\sigma}\,[\mathrm{x}(\xi)]$ is the transformed stress tensor.

The approximation of the above integral via the GLL quadrature leads to a summation over *9(N+1)* elements (from Fichtner, 2011):

$$\begin{aligned}\mathrm{F_{qrs}}\left[(\nabla\cdot\sigma)_p\right] = {} & \sum_{n=1}^{3}\sum_{i=1}^{N+1} w_i\, w_r\, w_s\,\frac{\partial}{\partial\xi}\, l_q\left(\xi_1^i\right)\xi_1^i\sigma_{np}^{irs}\, J(\xi^{irs})\frac{\partial\xi_1}{\partial x_n}(\xi^{irs}) \\ & + \sum_{n=1}^{3}\sum_{i=1}^{N+1} w_q\, w_i\, w_s\,\frac{\partial}{\partial\xi}\, l_r\left(\xi_2^i\right)\xi_2^i\sigma_{np}^{qis}\, J(\xi^{qis})\frac{\partial\xi_2}{\partial x_n}(\xi^{gis}) \\ & + \sum_{n=1}^{3}\sum_{i=1}^{N+1} w_q\, w_r\, w_i\,\frac{\partial}{\partial\xi}\, l_s\left(\xi_3^i\right)\xi_3^i\sigma_{np}^{qri}\, J(\xi^{qri})\frac{\partial\xi_3}{\partial x_n}(\xi^{gri}).\end{aligned} \tag{A1.51}$$

c) After the transformation, the source term f in equation (A1.39) becomes (Fichtner, 2011): $\mathrm{F_{qrs}}\,(\mathrm{f_p})= \int_0^{G_e}\psi_{qrs}\mathrm{e}_p\cdot\mathrm{f}\, d^3\mathrm{x} = w_q\, w_r\, w_s\,\mathrm{f'_p}(\xi^{qrs})\, J(\xi^{qrs})$, (A1.52)

where $\mathrm{f'_p}(\xi) = \mathrm{f_p}[\mathrm{x}(\xi)]$ is the transformed density.

Since the equation (A1.52) is obtained by applying the GLL quadrature, which is exact only for polynomials of maximum degree equal to $2N-1$, the expression of the source term is an approximation.

The Galerkin projection of the (*mn*)-component of the stress tensor σ in the approximate weak form of the displacement-stress formulation (A1.40) follows (from Fichtner, 2011):

$$\mathrm{F_{qrs}}(\sigma_{\mathrm{mn}}) = \int_0^{G_e}(\psi_{qrs}\,\mathrm{e}_m\cdot\bar{\sigma})_n\, d^3\mathrm{x} = w_q w_r w_s\,\sigma_{\mathrm{mn}}^{\mathrm{qrs}}\, J(\xi^{qrs}), \tag{A1.53}$$

and

$$\begin{aligned}\mathrm{F_{qrs}}\ [(\mathrm{C}:\nabla\mathrm{u})]_{mn} = {} & \int_0^{G_e}(\psi_{qrs}\mathrm{e}_m\cdot\mathrm{C}:\nabla\bar{\mathrm{u}})_n\, d^3\mathrm{x} = \\ & \sum_{a,b=1}^{3}\sum_{i,j,k=1}^{N+1}\psi_{qrs}(\mathrm{x})\ C_{mnab}\frac{\partial}{\partial x_a}\left[u_b^{ijk}\psi_{qrs}(\mathrm{x})\right]d^3\mathrm{x}\end{aligned} \tag{A1.54}$$

for the left-hand side and right-hand side, respectively. Inserting the Lagrange polynomials as basis functions into the equation (A1.53), the relation for mapping the elements to the reference volume *X* is (from Fichtner, 2011):

$$\mathrm{F_{qrs}}\,[(\mathrm{C}:\nabla \mathrm{u})]_{mn} = \sum_{a,b,p=1}^{3} \sum_{i,j,k=1}^{N+1} \int_{o}^{X} u_b^{ijk}\; l_q(\xi_1)\, l_r(\xi_2)\, l_s(\xi_3) \cdot C'_{mnab}(\xi) \frac{\partial\, \xi_p}{\partial x_a} \frac{\partial}{\partial \xi_p} [l_i(\xi_1)\, l_j(\xi_2)\, l_k(\xi_3)]\, \mathrm{J}(\xi)\, d^3\xi\,, \tag{A1.55}$$

with $\mathrm{C}'(\xi) = \mathrm{C}[\mathrm{x}(\xi)]$.

Finally, the Galerkin projection for the stress tensor in the approximate weak form leads to (from Fichtner, 2011):

$$F_{qrs}\;[(\mathrm{C}:\nabla \mathrm{u})]_{mn} = w_q w_r w_s\, J(\xi^{qrs}) \sum_{a,b=1}^{3} C'_{mnab}(\xi^{qrs}) \cdot \sum_{i=1}^{N} \left[\mathrm{u_n^{irs}} \frac{\partial}{\partial \xi} l_i(\xi_1^q) \frac{\partial \xi_1}{\partial x_m} + \mathrm{u_n^{qis}} \frac{\partial}{\partial \xi} l_i(\xi_2^r) \frac{\partial \xi_2}{\partial x_m} + \mathrm{u_n^{qri}} \frac{\partial}{\partial \xi} l_i(\xi_1^s) \frac{\partial \xi_3}{\partial x_m}\right]. \tag{A1.56}$$

The discrete equations of the displacement-stress formulation become (from Fichtner, 2011):

$$[w_q\, w_r\, w_s\; \rho'(\xi^{qrs})\, J(\xi^{qrs})] \frac{\partial^2}{\partial t^2} \mathrm{u_p^{qrs}} + \mathrm{F_{qrs}}\left[(\nabla \cdot \sigma)_p\right] = \mathrm{F_{qrs}}\,(\mathrm{f_p}), \tag{A1.57}$$

$$[w_q\, w_r\, w_s\, J\,(\xi^{qrs})]\; \sigma_{\mathrm{mn}}^{\mathrm{qrs}} = \mathrm{F_{qrs}}\;\; [(\mathrm{C}:\nabla \mathrm{u})]_{mn} \tag{A1.58}$$

Equation (A1.58) provides the expansion coefficients $\sigma_{\mathrm{mn}}^{\mathrm{qrs}}$ of σ in terms of expansion coefficients $\mathrm{u_n^{qrs}}$of u, and it can be insert into the equation (A1.57) to compute the local acceleration $\frac{\partial^2}{\partial t^2}\mathrm{u_p^{qrs}}$. Due to the use of GLL node points and GLL quadrature, the explicit computation of the mass and stiffness matrices in unnecessary (Fichtner, 2011).

Point source implementation

The approximation of the projection integral in equation (A1.52) by GLL quadrature is inappropriate when the integrand is not a polynomial of degree *2N-1*. Usually, the seismic source is approximately represented by a single-force point source (Fichtner, 2011):

$$\mathrm{f}(x,t) = \mathrm{s}(t)\delta(\mathrm{x} - \mathrm{x}^s), \tag{A1.59}$$

or by a moment tensor point-source:

$$\mathrm{f}(x,t) = -\nabla \cdot [\mathrm{M}(t)\delta(\mathrm{x} - \mathrm{x}^s)], \tag{A1.60}$$

where $(\mathrm{x} - \mathrm{x}^s)$ indicates the source location, *s* denotes a vectorial source time function and M denotes a moment tensor.

There are two common methods for the point source implementation: (1) the exact integration and (2) the polynomial approximation of the delta function (Fichtner, 2011).

(1) In *the exact integration,* the single-force point source is (Fichtner, 2011):

$$F_{qrs}(f_p) = \int_0^{G_e} \psi_{qrs} \mathrm{e}_p \cdot \mathrm{f}\, d^3\mathrm{x} = \int_0^{G_e} s_p(t)\, \psi_{qrs}(\mathrm{x})\, (\mathrm{x} - \mathrm{x}^s)\, d^3\mathrm{x} = s_p(t)\, \psi_{qrs}(\mathrm{x}^s), \qquad \text{(A1.61)}$$

while the moment tensor source is:

$$F_{qrs}(f_p) = -\int_0^{G_e} \psi_{qrs}(\mathrm{x}) \mathrm{e}_p \cdot \nabla \cdot [\mathrm{M}(t)\, \delta(\mathrm{x} - \mathrm{x}^s)]\, d^3\mathrm{x} = \mathrm{e}_p \cdot \mathrm{M}(t) \cdot \nabla\, \psi_{qrs}(\mathrm{x}^s). \qquad \text{(A1.62)}$$

Since in the reference cube X the basis functions ψ_{qrs} are Lagrange polynomials, with the GLL points as collocation points, more than one coefficient $F_{qrs}(f_p)$ will be non-zero, unless the source location x^s coincides with a grid point. The numerical point source is therefore potentially non-local (Fichtner, 2011).

(2) The alternative to the source-point implementation is the *approximation of the delta function by Lagrange polynomials*, leading to a low-passed filtered version of the point source (Faccioli et al., 1997; Fichtner, 2011). Assuming that $\delta(\mathrm{x} - \mathrm{x}^s)$ in the source-bearing element G_e can be approximated in terms of basis functions $\psi_{ijk}(\mathrm{x})$ (Fichtner, 2011):

$$\delta(\mathrm{x} - \mathrm{x}^s) \approx \bar{\delta}(\mathrm{x} - \mathrm{x}^s) = \sum_{i,j,k=1}^{N+1} \delta^{ijk}\, \psi_{ijk}(\mathrm{x}). \qquad \text{(A1.63)}$$

The polynomial coefficients δ^{ijk} for the approximated function $\bar{\delta}$ are determined such that (Fichtner, 2011):

$$\psi(\mathrm{x}^s) = \int_0^{G_e} \bar{\delta}(\mathrm{x} - \mathrm{x}^s)\, \psi(x)\, d^3\mathrm{x}\,, \qquad \text{(A1.64)}$$

for any test function ψ that can be represented by $\psi_{ijk}(\mathrm{x})$.

Transforming the above equation to a reference cube X and substituting the equation (A1.63), gives (Fichtner, 2011):

$$\psi(\mathrm{x}^s) = \sum_{i,j,k=1}^{N+1} \int_0^X \delta^{ijk} \psi_{ijk}\, [\mathrm{x}(\xi)]\, \psi[\mathrm{x}(\xi)]\, \mathrm{J}(\xi)\, d^3\xi. \qquad \text{(A1.65)}$$

Inserting the Lagrange polynomials as basis functions, leads to (Fichtner, 2011):

$$\psi(\mathrm{x}^s) = \sum_{i,j,k=1}^{N+1} \int_0^X \delta^{ijk}\, l_i(\xi_1)\, l_j(\xi_2)\, l_k(\xi_3)\, \psi[\mathrm{x}(\xi)]\, \mathrm{J}(\xi)\, d^3\xi. \qquad \text{(A1.66)}$$

Choosing $\psi[\mathrm{x}(\xi)] = l_q(\xi_1)\, l_r(\xi_2)\, l_s(\xi_3)$ as test function in equation (A1.66) and approximating the integral with GLL quadrature, gives (Fichtner, 2011):

$$l_q(\xi_1^s)\, l_r(\xi_2^s)\, l_s(\xi_3^s) = w_q w_r\, w_s\, \delta^{qrs}\, J\, (\xi^{qrs}), \qquad \text{(A1.67)}$$

where $(\xi_1^s, \xi_2^s, \xi_3^s) = \xi^s = \mathrm{F}_e\, (\mathrm{x}^s), \qquad \text{(A1.68)}$

denotes the source position in the reference coordinate system. Finally, the polynomial coefficients δ^{ijk} are then given by (Fichtner, 2011):

$$\delta^{ijk} = \frac{l_i(\xi_1^s)\, l_j(\xi_2^s)\, k(\xi_3^s)}{w_i w_j\, w_k\, J\, (\xi^{ijk})}. \qquad \text{(A1.69)}$$

For a tensor point source, in most applications, the polynomial approximation of $[\mathrm{M}(t)\delta(\mathrm{x} - \mathrm{x}^s)]$ is added directly to the stress tensor σ, before the computation of the Galerkin projection of $\nabla \cdot \sigma$ (Fichtner, 2011).

Chapter 3

Inverse problem overview

Contents

3.1 – Introduction

Solving an inverse problem consists in inferring the parameters of a model (system), for which the predicted data (measurements) fit the observed (experimental) ones. Contrariwise, the forward problem solution consists in predicting a data set (results of an experiment) starting from a set of parameters that characterize a model *m* (Figure **3.1**). While the solution of a forward problem is unique (in deterministic physics), this is not the case for an inverse problem, which may have multiple solutions. To mitigate this issue, available a priori information and data uncertainties may be useful key elements (Tarantola, 2005).

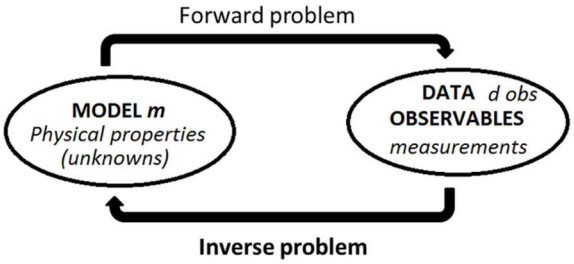

*Figure **3.1** – Sketch of the forward problem and inverse problem concept.*

Generally, the solution of an inverse problem consists of two phases:

1. Solving the direct problem (forward modeling) to obtain the predicted (synthetic) data (d_{pred}) belonging to an initial model (first guess m_0) $d_{pred} = \mathrm{G}\, m_o$, where G is the direct (or prediction) operator.
2. Using the difference between the observed data and the predicted ones to infer the so-called *error function* or *objective function*. In this phase, the inverse problem is solved: $m = \mathrm{G}^{-1}\, d_{obs}$, where G^{-1} is the inverse operator.

The complexity of an inverse problem is generally determined by the characteristics of the matrix G. If the relationship between the model parameters *m* and the observed data (d_{obs}) is linear, the inverse problem solution can be obtained by minimizing the so-called objective function: $\chi = d_{obs} - d_{pred} = d_{obs} - \mathrm{G}\, m$, where d_{obs} are the observed data, d_{pred} are the data computed in the predicted model and $\mathrm{G}\, m$ is the solution of the forward problem, which provides the predicted data. If G is a function of the unknown parameters m as well, the inverse problem is *nonlinear:* $d_{pred} = \mathrm{G}(m)$ and the solution requires the application of different techniques.

There are two main classes of methods for the solution of nonlinear inverse problems: local/deterministic methods (*Local Search*), based on the computation of derivatives, and global/probabilistic methods (*Global Search*), which do not use derivatives.

3.2 – Probabilistic versus deterministic approach

The amount of independent information contained in the seismic data is limited; this aspect leads to the intrinsic non-uniqueness of the reconstructed models. Therefore, solving an inverse problem often means evaluating multiple solutions (models) that can be compatible with the data. This issue can be mitigated by using a probabilistic approach (Tarantola, 2005), which assigns to each model m a probability $p(m)$ to represent the observed data. The definition of this probability is based on two factors: the degree of consistency of the model with a priori knowledge and the misfit between the observed data and the predicted ones (Fichtner, 2011). Nevertheless, the reliable evaluation of the probability requires many forward simulations (Rothman, 1985) while the number of necessary samples increases almost exponentially with the number of model parameters. Therefore, solving an inverse problem with a probabilistic approach becomes very expensive when the number of unknowns is large.

Differently, the computation cost can be efficiently reduced by using the deterministic approach, based on the definition of a misfit functional $\chi(m)$ that quantifies the discrepancy between the observed and predicted data. Solving an inverse problem with a deterministic approach means finding an optimal model m that minimizes $\chi(m)$. The advantage of this approach is the possibility of managing a greater number of parameters, at the expense of having only limited information on the solution non-uniqueness (Fichtner, 2011).

The misfit functional (i.e. objective function) can be defined using various norms (Menke, 1989): $\|\chi_n\| = (\sum_{i=1}^{N} \lceil e_i \rceil^n)^{\frac{1}{n}}$, where n is the norm's magnitude order. As the exponent n grows, the weight of the outliers (anomalous values, far from the characteristic distribution of the data) becomes more important. For $n = 2$, the previous relation stays for the least-squares norm (L_2). Solving an inverse problem formulated in the L_2 norm $\|\chi\|_2 = (\sum_{i=1}^{N} |\chi^2|)^{\frac{1}{2}} \rightarrow \|\chi\|_2^2 = \sum_{i=1}^{N} |\chi^2|$ means obtaining the minimum of the squared differences between the observed data and the data predicted for a certain model.

The matrix formulation for the L_2 norm-based objective function is:

$$\chi^2 = \chi^T \chi = (d_{obs} - d_{pred})^T \cdot (d_{obs} - d_{pred}), \tag{3.1}$$

where $d_{pred} = \mathbf{G}\, m$, *"T"* is the transpose operator, m is a column vector containing the model parameters, d_{pred} is a column vector which contains the data computed for the model m, and $\mathbf{G}$ is the *N x M* matrix relating the data with the model parameters m.

In this work, both a probabilistic and a deterministic approach were used. The probabilistic one was employed for dispersion curves inversion, while the adopted FWI tool is based on a deterministic technique.

The following sections provide a brief introduction to the global search and local search methods. A detailed quantitative description can be found in Tarantola et al. (2005).

3.3 – Global optimization: a conceptual overview

The global optimization, related to the probabilistic theory, can be adopted for the inverse problem solution. In this framework, the Bayesian inference consists in the combination of a priori information with observed data, leading to a posteriori probability distribution function (e.g. Tarantola, 1987 e 2005), which represents the complete solution of the inverse problem (Sambridge and Mosegaard, 2002).

Very popular global methods are the so-called Monte Carlo (MC) (Rothman, 1985; Tarantola, 1987) that allow an efficient extension of the Bayesian theory to nonlinear problems (Sambridge and Mosegaard, 2002). MC methods are based on the discrete simulation of the probability distribution and direct sampling in the parameter space, through a random (uniform) or pseudo-random (non-uniform) search. For each model's parameter, a probabilistic distribution is defined. Then, in the range of this distribution, a random parameters extraction is carried out. The extracted parameters are used to solve the forward problem. The inverse problem solution is based on statistical tests, in which the consistency of each predicted data with the observed data is verified. The result of the inversion is no longer a single value, but a probability distribution function.

When the number of unknowns is high, the sampling of the parameter space can be executed in a non-uniform manner: e.g. *Simulated Annealing*, *Neighborhood Algorithm.* Further, Genetic Algorithms (GA), introduced by Holland (1975) and developed in various works, i.e. Whitley (1994), are heuristic procedures inspired by an analogy with the natural selection and genetic evolution, defined as a form of optimization (Sambridge and Mosegaard, 2002). They are based on conceptually similar mechanisms to some processes from biology and genetics, in particular the selection, recombination and mutation. The models are therefore considered as individuals of a population, whose evolution is based on natural selection, recombination and mutation.

Firstly, a population of individuals with an initial random distribution (within a predetermined range) is generated, representing all the possible solutions for the inverse problem. The starting models are evaluated, by modifying them (using selection, recombination and mutation operators) during a series of successive generations, until the convergence towards an optimal model is obtained. To this aim, forward modeling is performed, and the associated errors are calculated for each of the selected initial model (relatives) and a new model (the so-called descendants of the parent models, resulted from the first recombination) of a given generation. After this operation, the models are evaluated in terms of data misfit. The process continues with the selection and recombination of the best individuals (relatives and descendants) of each generation, recombination and mutation, calculation of forward modeling and associated errors, new selection and so on until a certain convergence criterion is satisfied. As the selection acts to remove the diversity of the relatives, the purpose of the mutation operator is maintaining a certain degree of diversity within the population (Sambridge and Mosegaard, 2002). The individuals can also be gathered into various sub-populations that evolve separately and are recombined after a certain number of generations. The final goal of the GAs is finding

the model with the lowest misfit, within the possible solutions generated and the associated uncertainties (defined as the probability of a posteriori distribution).

3.4 – Local search methods and the linearized technique

The local inversion techniques (i.e. gradient-based) aim at minimizing a certain objective function, starting from a model that should guide the convergence towards the nearest minimum of such function. During this iterative minimization process, the initial model is subsequently updated, following the negative gradient, until a certain convergence criterion or a certain number of iterations is reached. The minimum of the objective function is found by iteratively zeroing the partial derivatives, computed with respect to each parameter of the model m, and exploiting some matrix transposition properties, like $(\mathrm{AB})^T = \mathrm{B}^T\mathrm{A}^T$ and $\mathrm{A}^T\mathrm{B} = \mathrm{B}^T\mathrm{A}$:

$$\nabla_m \left[(d_{obs} - \mathrm{G}m)^T \cdot (d_{obs} - \mathrm{G}\,m)\right] = 0 \rightarrow$$

$$\nabla_m \left[{d_{obs}}^T d_{obs} - 2\,{d_{obs}}^T (\mathrm{G}\,m) + m^T \mathrm{G}^T (\mathrm{G}\,m)\right] = 0 \rightarrow$$

$$-2\,{d_{obs}}^T \mathrm{G} + 2\,m^T \mathrm{G}^T\mathrm{G} = 0 \rightarrow m^T \mathrm{G}^T\mathrm{G} = \mathrm{G}^T d_{obs} \rightarrow$$

$$m = (\mathrm{G}^T \mathrm{G})^{-1} \mathrm{G}^T d_{obs} \tag{3.2}$$

Regarding the seismic data inversion, the nonlinear relationship between the seismic wavefield and the model parameters can be written in compact form as: $\mathrm{u}(\mathrm{x}, t) = \mathrm{G}(m)$ (Virieux and Operto, 2009). The solution of such a nonlinear inverse problem can be iteratively retrieved through the gradient descent techniques. They are usually (but not necessarily) based on the least-squares norm: $\chi(m) = \frac{1}{2}\Delta d^T \Delta d$, where T is the transpose conjugate and $\Delta d = d_{obs} - d_{calc}(m_i)$.

For nonlinear inverse problems, a widely used approach for the misfit function computation is the quasi-Newton method. Accordingly, for small variations inside the matrix G, the higher-order terms can be ignored, and the inverse problem can be "linearized" in a small neighborhood of the model parameters m (i.e. the Born approximation). This fact implies the assumption of a linear relationship between the model parameters and wavefield perturbation (Woodward, 1992; Virieux and Operto, 2009). Consequently, the updated model is expressed through the summation of the starting model m_o plus a small perturbation Δm, in the opposite direction of the gradient: $m_1 = m_o + \Delta m$. Exploiting this relation, the objective function's formulation is based on the analogy with the Taylor-Lagrange development in the starting model's m_0 neighborhood Δm. The series can be truncated at the first order (Jacobian matrix) or the second order (Hessian matrix):

$$\chi(m_0 \pm \Delta m)\,\chi(m_o) \pm \frac{\partial \chi(m_0)}{\partial m}\Delta m + \frac{1}{2}\frac{\partial^2 \chi(m_0)}{\partial m^2}\Delta m^2 \pm \frac{1}{3!}\frac{\partial^3 \chi(m_0)}{\partial m^3}\Delta m^3 + \frac{1}{n!}\frac{\partial^n \chi(m_0)}{\partial m^n}\Delta m^n + \mathcal{O}(\Delta m)^{n+1}. \tag{3.3}$$

The Jacobian matrix contains the first derivatives of the objective function with respect to each parameter belonging to the model space, while the Hessian matrix contains the corresponding second-order derivatives:

$$\chi(m_0 + \Delta m) = \chi(m_0) + \sum_{i=1}^{M} \frac{\partial \chi(m_o)}{\partial m_i} \Delta m_i + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{\partial^2 \chi(m_o)}{\partial m_i \partial m_j} \Delta m_i \Delta m_j + \mathcal{O}(\Delta m)^3 . \quad (3.4)$$

The error $\mathcal{O}(\Delta m)^3$ is zero when the misfit functional is a quadratic function of the variable m (Virieux and Operto, 2009).

The general scheme of the gradient descent methods gathers the following steps:

1. The choice of an initial model m_o and the forward problem solution: $d_{pred} = \mathrm{G}(m_0)$.
2. The computation of the residuals: $\Delta d = (d_{obs} - d_{pred})$.
3. The definition of the objective function: $\chi(m_0) = \frac{1}{2} \Delta d^T \Delta d$.
4. The computation of the model update amount and direction, through the minimization of the objective function.
5. The model update: $m_{i+1} = m_i - G^{-1} \nabla_m \chi(m_i)$, using a step that guarantees the condition $\chi(m_{i+1}) < \chi(m_i)$ to be true.
6. Setting $i = i + 1$ and returning to step 2.

In matrix notation, the linearized relationship providing the predicted data can be expressed as: $\Delta d_i = \frac{\partial \mathbf{G}_i}{\partial m_j} \Delta m_j$. In detail:

$$\Delta d_1 = \frac{\partial \mathbf{G}_1}{\partial m_1} \Delta m_1 + \frac{\partial \mathbf{G}_1}{\partial m_2} \Delta m_2 + \cdots + \frac{\partial \mathbf{G}_1}{\partial m_M} \Delta m_M,$$

$$\Delta d_2 = \frac{\partial \mathbf{G}_2}{\partial m_1} \Delta m_1 + \frac{\partial \mathbf{G}_2}{\partial m_2} \Delta m_2 + \cdots + \frac{\partial \mathbf{G}_2}{\partial m_M} \Delta m_M,$$

...

$$\Delta d_N = \frac{\partial \mathbf{G}_N}{\partial m_1} \Delta m_1 + \frac{\partial \mathbf{G}_N}{\partial m_2} \Delta m_2 + \cdots + \frac{\partial \mathbf{G}_N}{\partial m_M} \Delta m_M, \quad (3.5)$$

where $i = 1 \ldots N$ is the number of rows and $j = 1 \ldots M$ is the number of columns of the matrix G. The above relation can be written as: $\Delta d = \mathrm{J}\, \Delta m$, where Δd is the increment data vector, Δm is the increment model parameters vector and J is the Jacobian matrix.

The above formulation recalls the configuration of a linear inverse problem and, therefore, similar techniques can be adopted for inferring the solution. For example, according to the L_2 minimization of the objective function, the inverse problem solution can be computed as:

$$\Delta m = (\mathrm{J}^{\mathrm{T}}\, \mathrm{J})^{-1}\, \mathrm{J}^{\mathrm{T}}\, \Delta d_{obs}. \quad (3.6)$$

Nevertheless, in this case, Δm is not the final solution, but the correction to be applied to the initial model m_o (or to the predicted model of a certain iteration m_i), before proceeding towards the next iteration.

When adopting the second-order approximation, the model update is obtained by minimizing the objective function χ in the proximity of the starting model m_o as follows:

$$\frac{\partial \chi(m_0 + \Delta m)}{\partial m} = \frac{\partial \chi(m_0)}{\partial m} + \sum_{j=1}^{M} \frac{\partial^2 \chi(m_o)}{\partial m_i \partial m} \Delta m_i, \quad (3.7)$$

$$\Delta m_i = -\left[\frac{\partial^2 \chi(m_o)}{\partial m_i \partial m}\right]^{-1} \frac{\partial \chi(m_0)}{\partial m} = -H^{-1} \nabla_m \chi(m_0). \quad (3.8)$$

where H is the Hessian matrix, defining the curvature of the objective function, and the term $\partial\chi\,(m_0)/\partial m = \nabla_m\,\chi(m_0)$ is the gradient of the objective function, providing the *steepest descent* direction (Virieux and Operto, 2009). The direction and amount of model perturbation are not provided by the gradient of the misfit function itself, but by the action of the Hessian's inverse to the gradient (Mora, 1987b; Virieux and Operto, 2009). The perturbation model Δm is searched in the opposite direction of the gradient.

As the computation of the Hessian matrix inverse is usually expensive for an elastic problem, a scalar, called *step length,* is commonly used instead (Virieux and Operto, 2009). The relations for the model perturbation and model update become:

$$\Delta m = -\alpha\,\frac{\partial\chi\,(m_0)}{\partial m} = -\,\alpha\,\nabla_m\,\chi(m_0). \tag{3.9}$$

$$m_1 = m_0 - \alpha\,\nabla_m\,\chi(m_0). \tag{3.10}$$

Several techniques for the step length estimation have been developed. Among these, the most common one is the *line search* (Gauthier et al., 1986; Tarantola 1984b and 1987; Sambridge et al., 1991). An inaccurate estimation of the step length can guide the convergence towards a wrong model.

According to Shin et al. (2001) and Warner et al. (2013), the convergence of the inversion algorithm can be significantly improved when dividing the gradient by the diagonal terms of the Hessian matrix, which may act as scale property of the gradient (Pratt et al., 1998):

$$\Delta m_i \approx -\frac{\alpha}{h_{ii}}\,\frac{\partial\chi\,(m_0)}{\partial m} = -\frac{\alpha}{h_{ii}}\,\nabla_m\,\chi(m_0) \tag{3.11}$$

Even though the linearized techniques are computationally efficient, the inverse problem solution is reliable only if the initial model m_o is close to the objective function's global minimum. Thus, for a non-linear problem, the risk of ending up in a local minimum is high and the choice of the initial model imperatively constrains the proper convergence of the algorithm.

3.5 – Highlights

- An inverse problem may have multiple solutions; a priori information, or a very accurate first guess, are necessary elements for mitigating this issue.
- If the relation connecting the data with the model parameters (unknowns) depends on the unknowns, the inverse problem is defined as *nonlinear*.
- There are two main methods for the solution of a nonlinear inverse problem: local/deterministic (based on the computation of the gradient) and global/probabilistic (based on the global search in the parameters space).
- The deterministic methods are computationally less demanding than the probabilistic ones, but they require a very accurate initial model, "located" in the proximity of the objective function's global minimum. Otherwise, the convergence can reach a local minimum, related to a wrong solution.

Chapter 4

Surface wave analysis

Contents

4.1 – Some concepts related to surface waves propagation

This section presents some concepts related to SW propagation in vertically heterogeneous media. A detailed quantitative description can be found, for example, in Aki and Richard (2002) and Foti et al. (2015).

The basic theory is referred to both Rayleigh waves and Love waves, although more attention is given to the Rayleigh waves because they are preferentially employed in most of the practical applications for near-surface characterization. Few applications use Love waves (generated from horizontally polarized SWs) since they suffer from stratigraphy related limitations. For marine applications, Scholte waves (propagating along the fluid-solid interface) are used, although the geometric dispersion is influenced by the thickness of the fluid layer.

4.1.1 – Skin depth, critical depth and geometric dispersion

SWs are generated at the boundary of a domain (i.e. the free surface of the Earth) where the internal stress vanishes (Foti et al., 2015). They generally exhibit a 2D radiation pattern and a lower geometrical attenuation than body waves (BWs), dominating the wavefield after a certain number of wavelengths from the source (Lamb, 1904). Since only a small amount of energy propagates inside the medium, the displacement field generated by SW decays exponentially with depth and it is confined within about one wavelength from the surface (Achenbach, 1984). In this framework, the concept of *skin depth* is related to the depth at which the SW amplitude decreases by a factor of 1/e (which is about 0.94λ for Rayleigh waves in a homogeneous medium – Foti et al., 2015).

The horizontal and vertical particle displacements associated with the propagation of a Rayleigh wave are out of phase by $\pi/2$ (in an elastic homogeneous half-space), leading to elliptical particle motion on the $\{x_1, x_2\}$ plane, with a retrograde orbit in correspondence of the free surface and prograde orbit at a depth of about $x_2^c = 0.2\lambda$. The ratio between vertical and horizontal displacement is about 1.5 at the free surface, while in correspondence of the so-called *critical depth* x_2^c, the horizontal displacement vanishes and the motion is only vertical (Foti et al., 2015) – Figure **4.1**.

In a viscoelastic medium, the phase difference between the horizontal and vertical components of the wavefield is no longer $\pi/2$, because the Rayleigh eigenfunctions are complex-valued. This fact leads to a backward or forward rotation of the medium particles with respect to the free boundary (Foti et al., 2015).

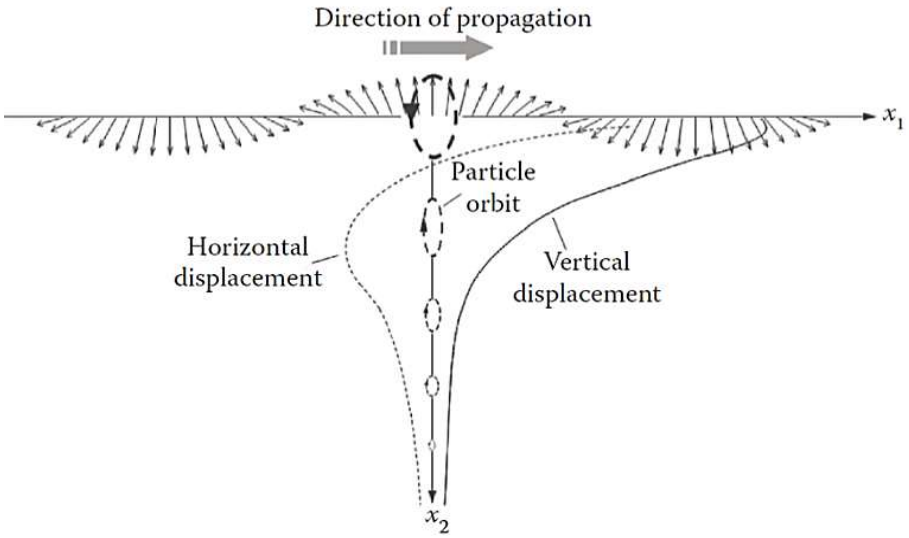

*Figure **4.1** - Particle displacement associated with the Rayleigh wave propagation; here the medium is considered as homogeneous (from Foti et al., 2015).*

In a heterogeneous medium where velocity of the seismic waves increases with depth, the lower-frequency components (larger wavelengths), which penetrate deeper in the subsurface than the higher-frequency components (shorter wavelengths), "propagate" with higher velocities (controlled by the rheological properties of the deeper geological layers). This phenomenon is known as *geometric dispersion* (i.e. the phase velocity of Rayleigh waves is frequency dependent) - Figure **4.2**.

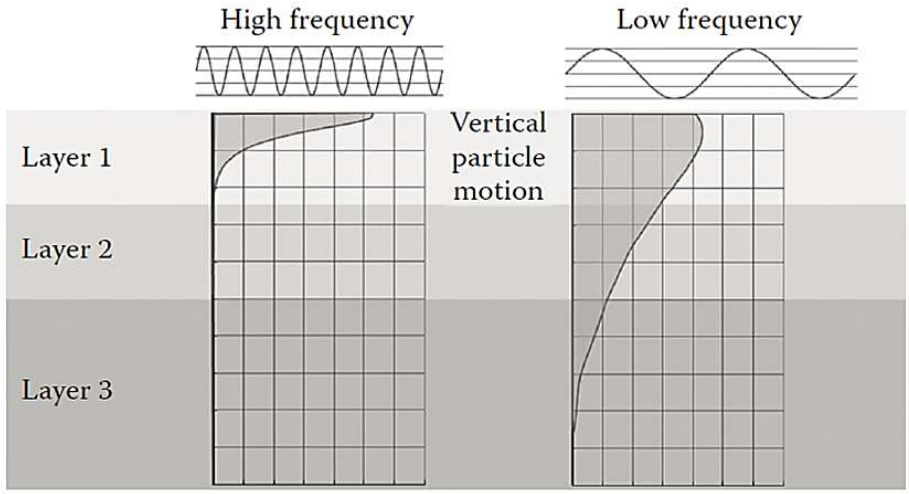

*Figure **4.2** – Geometric dispersion of Rayleigh waves: the variation of vertical particle displacement related to the propagation of two signals having different frequencies (from Foti et al., 2015).*

The solution of a linear dispersive wave equation can be written as:

$$f(x,t) = A\, e^{i[kx-\omega(k)t]}, \tag{4.1}$$

where A is the wave's amplitude, $\omega(k)$ is the wavenumber-dependent circular frequency, and $[kx-\omega(k)t]$ is the (constant) phase of the wave front. The wavenumber/frequency-dependent phase velocity is: $C_o = \frac{dx}{dt} = \frac{\omega(k)}{k}$. (4.2)

The solution of the partial differential wave equation for a dispersive medium can be retrieved by using the Fourier integral:

$$f(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k) e^{i[kx-\omega(k)t]}\, dk, \tag{4.3}$$

which describes a superposition of waves characterized by different wavenumbers and phase velocities.

The geometric dispersion is used to retrieve some characteristic parameters of the subsurface (commonly the S-waves velocity profile) via *dispersion curves* (i.e. the SW phase velocity variation as a function of wavelength/frequency) inversion (Figure **4.3**).

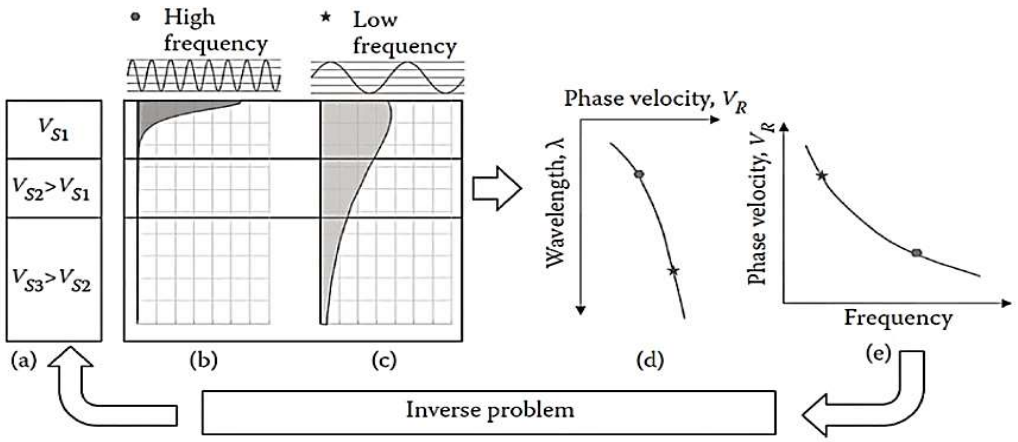

*Figure **4.3** – Sketch of the dispersion curves inversion process (from Foti et al., 2015).*

4.1.2- Near-field effects

Near the source, BWs and SWs component are not separated, biasing the phase velocity estimation, especially in the low-frequency band. Since BWs attenuation (spherical front) is faster than SWs attenuation (cylindrical front), a different spreading of energy in space is induced. At a certain distance from the source, as BWs energy reduces, their contribution can be neglected. Therefore, the concept of *near field* indicates the distance from the source where the influence of BWs field is significant (Aki and Richard, 2002; Foti et al., 2015).

Most of the techniques used for extracting the DCs from seismic data are based on the plane wave front assumption. Nevertheless, in the near field, the cylindrical shape of the wave front (Figure **4.4**) cannot be neglected and the use of cylindrical coordinates may be necessary (Zywicki and Rix, 2005).

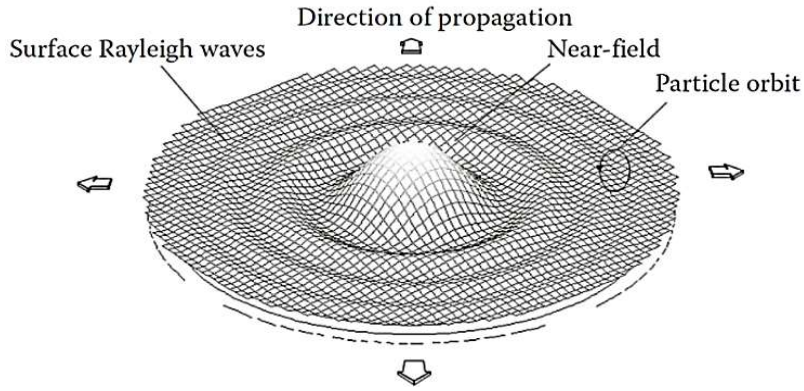

*Figure **4.4** - Radiation pattern of Rayleigh waves associated with a vertical point source (from Foti et al., 2015).*

A way of mitigating the near-field effects is, for example, increasing the offset of the first receiver (Yoon and Rix, 2009). The drawback of this approach is the loss of high-frequency components and the consequent loss in resolution in the shallow layers.

4.1.3- The eigenvalue problem

In heterogeneous media, the interference among waves scattered by different interfaces generates different modes of propagation. Considering a SW with a frequency ω propagating in the horizontal direction x (Aki and Richards, 2002):

$$\mathbf{u}(x, y, z, t) = \mathbf{Z}(z)\, e^{[i\,(kx-\omega t)]}, \tag{4.4}$$

under the free-surface condition $z = 0$, depth boundary condition $z = \infty$, and the condition to satisfy the equation of motion, such solution can be retrieved by giving an arbitrary value to ω, and considering k as dependent on ω: $k_n(\omega)$. This is an *eigenvalue* problem, where the symbol n indicates that more than one value of k may provide SWs with a given frequency ω. In other words, different modes may correspond to a given frequency ω (Aki and Richards, 2002).

For a given frequency ω, the SW wavenumbers can be expresses as $k_0(\omega), k_1(\omega), \dots, k_n(\omega)$, where $k_n(\omega)$ is an *eigenvalue* corresponding to the *eigenfunction* $\mathbf{u}_n(z)$; among all eigenvalues, $k_0(\omega)$ is the largest one ($k_0 > k_1 > \cdots > k_n$). The phase velocities $C_n = \omega/k_n$ have discrete values for any frequency ω; among all phase velocity values, C_0 is the lowest one ($C_0 < C_1 < \cdots < C_n$).

In a medium composed of a finite number of homogeneous layers overlying a homogeneous half-space, the number of modes is limited (Ewing et al., 1975). The fundamental mode $n = 0$ is generally dominant in long-period waves generated by shallow sources. The separation of various modes can be identified at great propagation distances, where they arrive at different times, as propagating with different group velocities.

In some cases, the energy associated with the higher modes is negligible, while the fundamental mode assumes a predominant role in the propagation of Rayleigh waves.

Differently, when the higher modes play an important role (e.g. in inversely dispersive sites or in presence of high impedance contrasts), the phase velocity retrieved from a relatively short acquisition spread is an apparent one, related to an apparent DC, determined by the superposition of various modes. In such conditions, the modes superposition must be accounted for in the forward problem solution and inversion, or various processing techniques must filtrate the various modal components before extracting the experimental DC (Foti et al., 2015).

When Rayleigh waves exhibit multimodal propagation, a greater penetration depth is reached. Therefore, the reconstruction of the deeper layers through DCs inversion also depends on the correct identification of the higher modes of propagation (Foti et al., 2015) – Figure **4.5**.

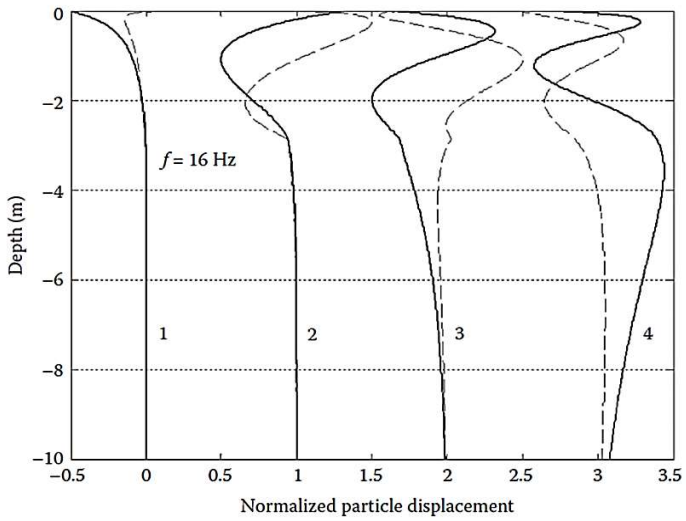

*Figure **4.5** – Various modes of horizontal component (dashed lines) and vertical component (bold lines) of Rayleigh displacement eigenfunctions for a vertically layered medium (from Foti et al., 2015). It can be noticed how the particle displacement associated with the mode n=4 reaches the highest depth.*

4.1.4 - Phase velocity versus Group velocity

In a homogeneous medium, SWs are not dispersive, and the waveform does not change during the propagation since the phase velocity C_n is not frequency dependent. Conversely, in a dispersive medium, SW velocity varies as a function of frequency and the aspect of the waveform varies with increasing offset.

The phase velocity is the propagation velocity of a single phase of the waveform. The group velocity is related to a wave packet consisting of contribution from a frequency range around a given frequency ω_0. The peaks and troughs of the wave packet propagate at the same phase velocity, generally different from the group velocity (Aki and Richards, 2002). Given a wave packet with a spectral density $|F(\omega)|$ and initial phase $\phi(\omega)$, containing a single mode (Aki and Richards, 2002):

$$f(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}|F(\omega)|\, e^{[-i\omega t + ik_n x + i\phi(\omega)]}\, d\omega, \tag{4.5}$$

in a non-dispersive medium (with invariable $C_n = \omega/k_n$) the waveform remains constant over the propagation path:

$$f(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}|F(\omega)|\, e^{[-i\omega(t\, -\, x/C_n)\,]}\, d\omega = f(t - x/C_n). \tag{4.6}$$

For a strongly dispersive SW, the waveform can be approximated, for example, by using the stationary phase method (Aki and Richards, 2002). This involves the integral estimation in the vicinity of the saddle point, where the phase is stationary or vary slowly with the frequency ω:

$$\frac{d}{d\omega}(-\omega t + k_n x) = 0 \text{ or } \frac{x}{t} = \frac{d\omega}{dk_n}. \tag{4.7}$$

The above equation provides the frequency $\omega_s(x,t)$ dominating at time *t* and distance *x* for which the group velocity is equal to *x/t*. The character of the SW seismogram at the position *x* is determined by how many ω_s frequencies exist for the fix position *x* and variable time *t* (Aki and Richards, 2002).

The second-order approximation of Taylor's series expansion of the phase near the saddle point (of stationary phase) gives:

$$-\omega t + k_n x \approx -\omega_s t + k_n(\omega_s)x + \frac{x}{2}\frac{d^2k_n}{d\omega^2}(\omega - \omega_s)^2. \tag{4.8}$$

Considering twice the contribution of positive frequencies, the formula (4.6) becomes (Aki and Richards, 2002):

$$f(x,t) = \frac{1}{2\pi} 2Re\left\{e^{[-i\omega_s t + ik_n(\omega_s)x]}|F(\omega)| \int_{-\infty}^{\infty} e^{\left[\frac{x}{2}\frac{d^2k_n}{d\omega^2}(\omega-\omega_s)^2\right]}\, d\omega\right\}. \tag{4.9}$$

Considering the property $\int_0^{\infty}\sin\left(\frac{1}{2}a\omega^2\right)d\omega = \int_0^{\infty}\cos\left(\frac{1}{2}a\omega^2\right)d\omega = \frac{1}{2}\sqrt{\frac{\pi}{a}}$, the formula (4.9) can be written as (Aki and Richards, 2002):

$$f(x,t) \approx \frac{|F(\omega_s)|}{\pi}\sqrt{\frac{2\pi}{x\left|\frac{d^2k_n}{d\omega^2}\right|}}\cos\left[-\omega_s t + k_n(\omega_s)x \pm \frac{\pi}{4}\right]. \tag{4.10}$$

The formula (4.10) describes the normal dispersion of SW, for a given phase velocity $C_n = \omega/k_n$ and group velocity $C_g = d\omega/dk$. The sign of the phase shift by $\pi/4$ (time shift of 1/8 T) marks a delay when the group velocity increases with the period. When the group velocity is stationary for various frequencies, the term $d^2k_n/d\omega^2$ is zero and the above formula is not valid anymore (Aki and Richards, 2002).

The expression of the phase, when using a higher-order approximation in the Taylor expansion, provides results in terms of Airy function, while the arrivals associated with the maxima and minima of the group velocity are called Airy phases (e.g. Savage, 1969).

The group velocity is simpler defined as (Stokes, 1880):

$$C_g = \frac{d\omega(k)}{dk} = C_0 + k\frac{dC_0}{dk} = C_0\left(1 - k\frac{dC_0}{d\omega}\right)^{-1}. \tag{4.11}$$

The above expression can be retrieved by considering the propagation, in a dispersive medium, of a signal composed by the superposition of two monochromatic waves with identical amplitude and frequency:

$$f(x,t) = A\sin(k_1 x - \omega_1 t) + A\sin(k_1 x - \omega_1 t). \quad (4.12)$$

Considering $k_m = \frac{k_1+k_2}{2}, k = \frac{k_1-k_2}{2}, \omega_m = \frac{\omega_1+\omega_2}{2}, \omega = \frac{\omega_1-\omega_2}{2}$,

the equation (4.12) can be written as (Foti et al., 2015):

$$f(x,t) = 2A\cos(\Delta k x - \Delta\omega t)\sin(k_m x - \omega_m t), \quad (4.13)$$

describing a signal propagating with phase velocity $C_0 = \omega_m/k_m$, and an envelope velocity $C_g = d\omega/dk$. The equation for the group velocity is obtained for $dk \to 0$. The phase velocity may be higher or lower than the group velocity. In normally dispersive conditions, the phase velocity is usually higher than the group velocity. When the derivative of the phase velocity is equal to zero (i.e. in a homogeneous medium), the phase velocity is equal to the group velocity.

Figure **4.6** illustrates schematically the concept of phase velocity and group velocity.

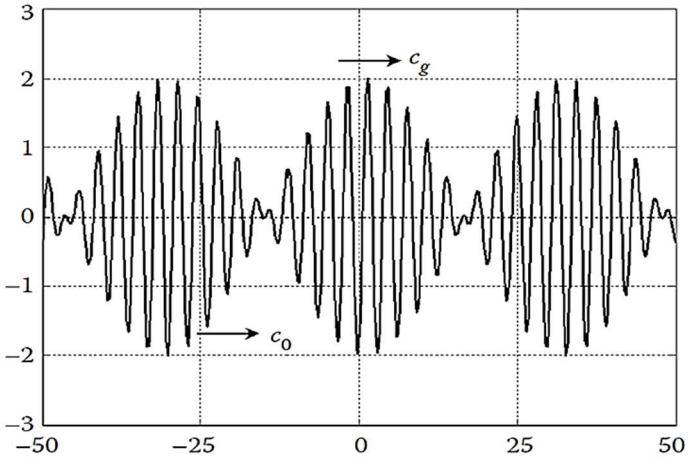

*Figure **4.6** - Phase velocity C_0 vs Group velocity C_g for a signal composed by a superposition of several monochromatic waves with similar frequencies (from Foti et al., 2015).*

In the presence of higher modes of propagation, each mode is described by a frequency-dependent phase and group velocity (Foti et al., 2015) – Figure **4.7**.

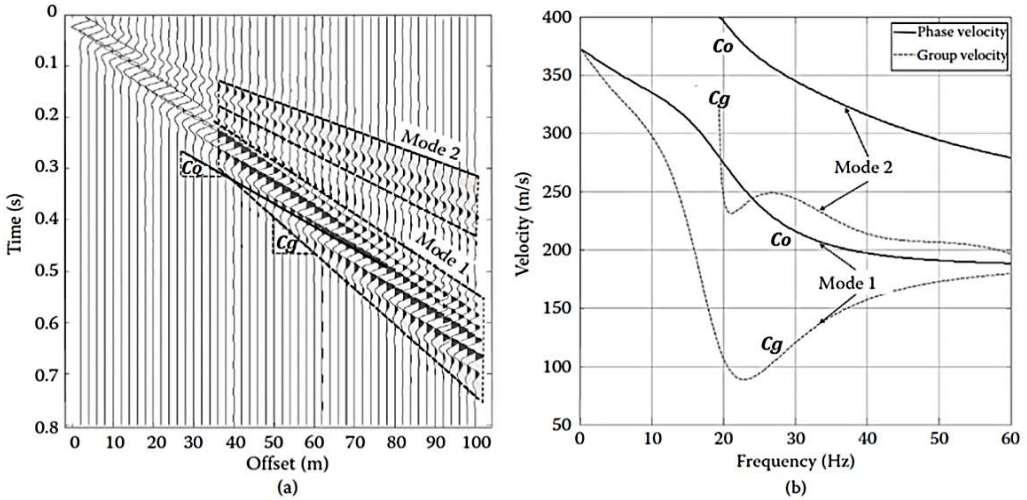

*Figure **4.7** – Example of dispersive behavior of SWs for a normally dispersive site: (a) common-shot gather, (b) dispersion curves for the phase velocity (c_0) and group velocity (c_g) of each mode (from Foti et al., 2015 - modified).*

The group velocity analysis is commonly used in seismology for the Earth crust characterization, or it is applied to microtremors for basin characterization; however, it is not a commonly used method for near-surface characterization, because of the limited resolution and interference between the various modes of propagation (Foti et al., 2015).

4.2 – Applications and developments of the surface wave method

Most of the surface waves (SW) based applications exploit the physical phenomenon of geometric dispersion to retrieve the values of some depth-dependent parameters of the subsurface.

There are two main categories of SW methods: passive methods, exploiting natural sources that generate low-frequency signals (such as ambient noise) and active methods, based on sources that produce signals characterized by higher frequencies. The first ones allow exploring deeper geological formations (at the expense of the shallower layers resolution), while the second ones explore the shallower horizons with relatively high resolution.

There are various applications based on the joint use of active and passive methods (e.g. Foti, 2005; Foti et al., 2007). They allow reconstructing the dispersion curve DC (i.e. the SW phase-velocity associated with different frequencies/wavelengths) for a wider band of frequencies (and then for various depths), while still preserving a high resolution for the shallower layers - Figure **4.8**.

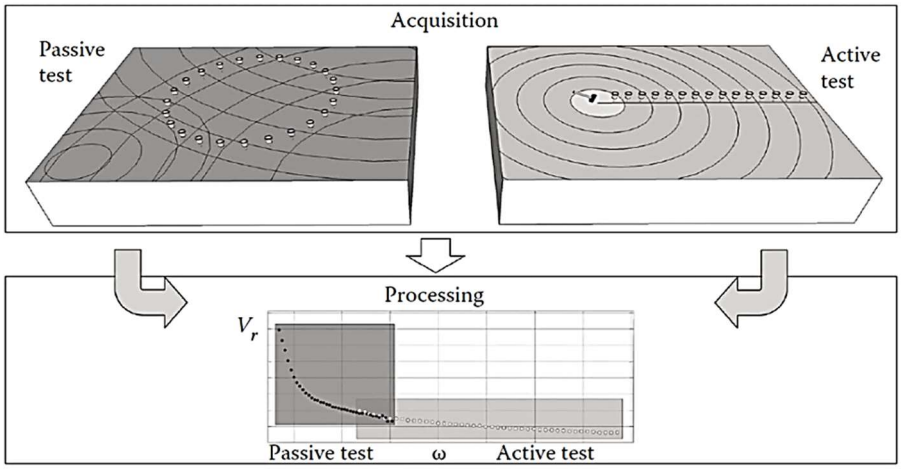

*Figure **4.8** – Schematic illustration of the combination of passive methods (lower frequencies) with active methods (higher frequencies), providing the reconstruction of the dispersion curve for a wider frequency range (from Foti et al., 2015).*

DCs are functions of the physical and mechanical proprieties of the subsurface and they can be used for the solution of an inverse problem, aimed at retrieving the model parameters. The inversion of SW DCs classically provides the local S-wave velocity model (e.g. Foti et al., 2000, 2001, Socco and Strobia, 2004; Socco et al., 2010b; Haney and Miller, 2013). The retrieved models are used for many applications. One of them is defining the weighted average value of the S-wave velocity in the upper 30 meters of the soil profile (Vs,30) for the site seismic response characterization (e.g. Brown et al., 2000; Moss, 2008; Comina et al., 2011). Another one is retrieving the one-way travel-times, used to compute the static corrections of the 3C seismic reflection data (e.g. Papadopolou et al., 2020).

In heterogeneous media, SW propagation may exhibit different modes (one single frequency can "travel" with different velocities). In this case, the DC is an apparent one, formed by the superposition of various modes and related to an apparent phase velocity. For a normally dispersive vertical profile (whose stiffness increases with depth), the fundamental mode may be dominant. This fact does not occur in the presence of high impedance contrasts and velocity inversions, where the higher modes may exhibit preferential sensitivity to some parameters. However, they could be used to increase the inversion's robustness (Gabriels et al., 1987; Socco and Strobbia, 2004; Boiero et al., 2009; Maraschini and Foti, 2010; Bergano et al., 2011).

A longer array of receivers allows distinguishing better the various modes of propagation than a shorter array, which usually limits the analysis to the fundamental mode. The necessity of considering the higher modes triggered the development of various filtering techniques based, for example, on group velocity separation (e.g. Dziewonski et al., 1969; Levshin et al., 1994). The main advantage of inverting the higher modes is extending the investigation to higher depths than the ones interested by the particle displacement of the fundamental mode (Gabriels et al., 1987; Socco et al., 2010a; Foti et al., 2015). They may also enhance the resolution of the inverted model (Xia et al., 2003).

Regarding the DC inversion approach, a Monte Carlo method proved to be more suitable than a deterministic one, especially when the multi-modal Rayleigh waves propagation is considered; this occurs because the multi-modal objective function is characterized by many local minima. For the classical inversion of the SW DCs, a vertically layered model, which neglects the lateral variations, is commonly hypothesized in the forward modeling step. This simplification allows using different algorithms for the forward problem solution, i.e. propagator matrix (Thomson, 1950; Haskell, 1953; Gilbert and Backus, 1966) or stiffness matrix (Kausel and Roesset, 1981). Despite the 1D assumption of SW methods, there are some successful examples in literature for the reconstruction of lateral heterogeneities and anomaly detections (e.g. Socco et al.,2009; Boiero and Socco, 2010; Schwenk et al., 2016; Liu et al., 2019).

Commonly, to simplify the inverse problem's dimensionality and trade-off and to reduce the computation time, some parameters (such as the density and the Poisson's ratio) are considered a priori known, and only the S-wave velocity and the layers' thickness vary during inversion. The choice of considering the density and the Poisson's ratio as known parameters during inversion is also based on a sensitivity analysis performed by Nazarian (1984), who showed how this simplification reduces the number of unknown parameters from $4n-1$ to $2n-1$, where n refers to the number of layers overlaying a half-space, half-space included. A common issue of DC inversion is the solution's non-uniqueness (several velocity profiles can provide different DCs positioned at the same distance from the experimental DC). This renders the evaluation of the experimental errors, together with the uncertainties' assessment, non-negligible aspects for defining the solution's reliability (e.g. Cercato, 2009; Foti et al., 2009; Foti et al., 2015).

***

SWs have been intensively used in seismology for crustal layers characterization from earthquake recordings (Aki & Richards, 2002). Since the 1950s, the SW method was also applied for engineering site characterization (e.g. Jones 1958, 1962). After the 1970s, SWs were widely used for the reconstruction of the Earth's interior at various depths (including the upper mantle), supported by the continuous development of seismic networks and numerical analysis theory.

The use of passive SW tests (microtremor surveys – e.g. Louie, 2001; Strobbia and Cassiani, 2011, recording the ambient noise with a 2D array of receivers) has been supported by the development of various processing techniques, such as the frequency-domain beamforming (Lacoss et al., 1969) and the spatial autocorrelation (Aki, 1957). Passive data sets are usually acquired using 2D arrays, but there are few examples of passive data recorded with linear arrays (ReMi – e.g. Louie, 2001).

Another SW analysis approach, which uses the ratio between the vertical and horizontal radial component, helps to deconvolve the effects of the source while inverting the polarization/ellipticity of SW to estimate the velocity profile. The horizontal-to-vertical spectral ratio (HVSR or H/V) is commonly used for the estimation of the natural frequency of the site (e.g. Nakamura, 1989, 1996, 2000; Fäh et al., 2001; Fäh et al., 2003).

SWs techniques are widely employed in the geotechnical and engineering field, especially after the introduction of the SASW (Spectral Analysis of Surface Waves; e.g. Nazarian and Stokoe, 1983; Stokoe et al., 1994) and MASW method (Multichannel Analysis of Surface Waves - Park et al., 1999; Xia et al., 1999). Most of the applications are based on the analysis of Rayleigh waves since they can be more easily identified in the vertical component of the seismograms than the transverse motion of Love waves.

The first SW applications for geotechnical and engineering purposes were based on the laterally homogeneous medium assumption, but further strategies have been proposed to account for lateral heterogeneities as well (e.g. Bohlen et al., 2004; Bergamo et al., 2012; Ikeda et al., 2013). Socco et al. (2009) and Boiero and Socco (2010) proposed a filtering strategy for retrieving smooth lateral variations instead of a 1D model. Their method regards the DCs extraction along the seismic line trough moving receiver array or moving spatial windows, and the simultaneous inversion of evenly spaced DCs under certain lateral constrains. A pseudo-2D or a pseudo-3D model represents the output. Boiero and Socco (2014) experimented the joint inversion of the Rayleigh waves DCs and refracted P-waves travel-times, highlighting the advantage of eliminating the ambiguities related to velocity inversions and hidden horizons.

Socco et al. (2017) developed a method for the direct estimation from DCs of an average S-wave velocity profile. This method is based on an approximative linear relationship between the SW wavelength and the DC investigation depth. Specifically, the authors proved the existence of a link between the depth where the shear wave velocity registers a certain value (Vs,z – i.e. the time average S-wave velocity) and the wavelength for which the Rayleigh wave phase velocity has the same value. The so-called *wavelength-depth* relationship, retrieved from one DC and the corresponding $V_{S,z}$ profile (known from the inversion of from the borehole measurements), can be used to calculate the time average S-waves velocities corresponding to a set of similar DCs belonging to the same area. The authors have shown that the use of the SW inversion for the average $V_{S,Z}$ estimation is only partially affected by the solution non-uniqueness problem; this happens because the DC is not particularly sensitive to one single model parameter, while it is strongly sensitive to the RMS proprieties of the velocity model.

Even though in classical applications the Poisson's ratio is kept as constant (as well as the density) during the DC curve inversion to a V_S profile, many authors highlighted the Poisson's ratio influence on the SW investigation depth (e.g. Karray and Lefebvre, 2008; Pelekis and Athanasopoulos, 2011). In this framework, Socco and Comina (2017) introduced a novel approach for the Poisson ratio estimation from DCs, which exploits the sensitivity of the wavelength-depth relationship to the Poisson's ratio variation. This method allows retrieving an average P-waves velocity model (V_P,z), broadening the SW usefulness to the statics correction computation. Khosro Anjom et al. (2019) extended the application of the above method to the reconstruction of the S-wave and P-wave interval velocity by exploiting a Dix-type formula and a total variation regularization approach.

4.3 - Workflow for initial model building from surface wave analysis

In this study, the initial model building process follows the method proposed by Socco et al. (2017) and Socco and Comina (2017). The procedure (Figure **4.9**) gathers the following steps:

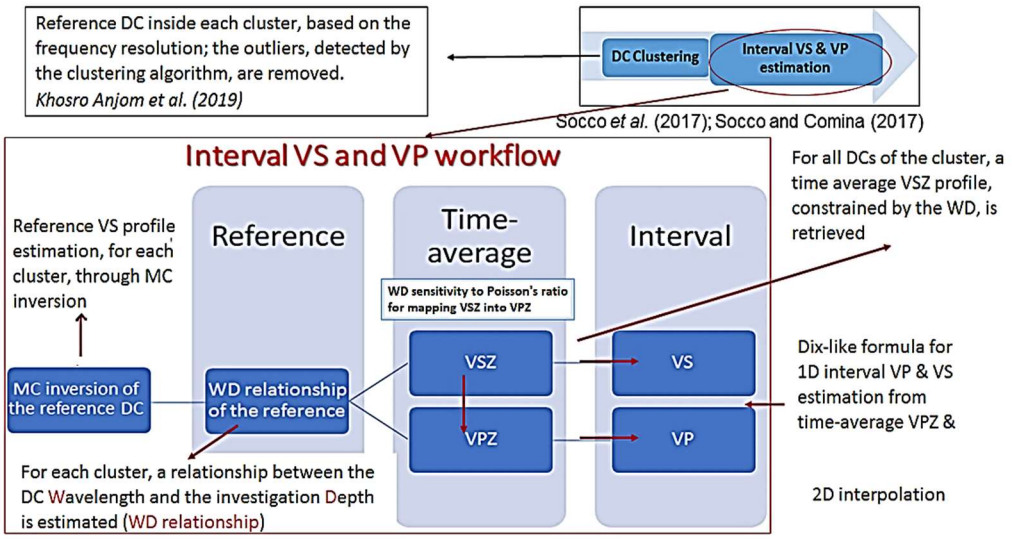

*Figure **4.9** – Synthetic sketch of the initial V_P and V_S model building workflow.*

a) DCs extraction

The DCs are extracted from seismic data using a Gaussian windowing approach. Inside each Gaussian window, the amplitude of the signal decreases as the distance from the window's center increases. The Gaussian window represents a good compromise between the optimum spectral (but low lateral) resolution of a Hanning window (that considers the entire receivers line and provides one DC) and the high lateral (but low spectral) resolution of a box window (that divides the seismic record in sub-records and provides a set of DC) – Figure **4.10**, Bergamo et al (2012).

This Gaussian windowing procedure allows extracting DC related to local physical properties along a seismic line. The Gaussian window is defined as (Bergamo et al., 2012):

$$w_{(k)} = e^{-\frac{1}{2}\left(\alpha \frac{k-(N-1)\beta}{(N-1)/2}\right)} \quad (4.14)$$

where $w_{(k)}$ is the weight assigned to the kth trace, N stays for the number of receivers, β is a parameter related to the position of the Gaussian maximum on the receiver array and α is a parameter inversely proportional to the window's standard deviation: $\sigma = N / (2\ \alpha)$. The value of the parameter α controls the lateral resolution: for great values of α, the window standard deviation decreases and, therefore, the lateral resolution increases. This path in offset domain is related to a decrease in spectral resolution, which leads to the necessity of

finding an optimum trade-off between lateral resolution and wavenumber resolution. The parameter α value needs to guarantee a certain minimum wavenumber resolution, which allows for the various SW modes identification and events separation in the *f-k* domain, while still providing us with DCs as local as possible (Socco et al., 2009; Bergamo et al., 2012).

In the windowing process, only the external shots, placed outside the receiver line, are considered. Once the windowing is completed, for each Gaussian window, all shots are converted to the *f-k* domain and stacked. For each stacked spectrum, the picking of the spectral maxima provides the DC (together with the experimental uncertainties), corresponding to the related Gaussian window. Each DC is associated with the spatial coordinate of the corresponding Gaussian window's center. At the end of this process, several DCs, distributed along the seismic line, are obtained.

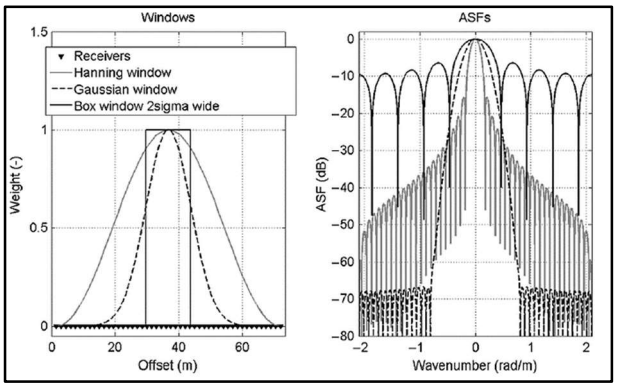

*Figure **4.10** – Lateral and corresponding spectral resolution of different windows: Hanning, Gaussian, Box window (from Bergamo et al., 2012). ASF= Array Smoothing Function, related to the spectral resolution. According to the Rayleigh resolution criteria, the wavenumber resolution can be quantified as half the main lobe width (HW) at – 6 dB (from Bergamo et al., 2012).*

Figure **4.11** shows some details related to the Gaussian window design. In Figure **4.11a**, it can be noticed how increasing the parameter α causes the decrease of the Gaussian window's standard deviation (and therefore the increase in spatial resolution). In Figure **4.11b** we can see how only the external shots, with respect to the receivers belonging to a certain Gaussian window, are considered for DC extraction.

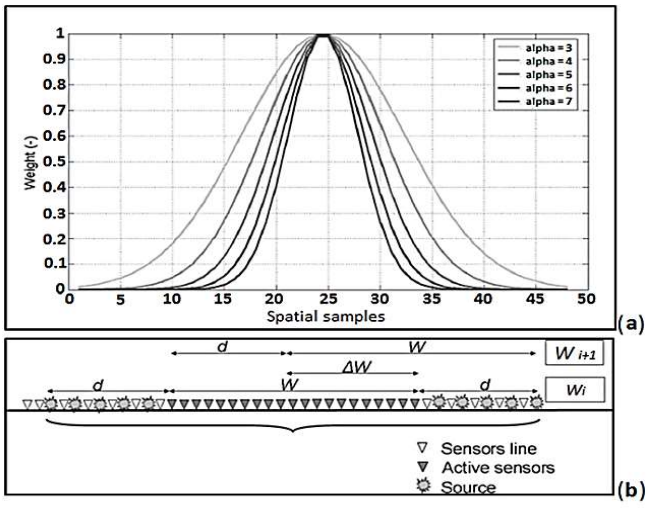

*Figure **4.11** – **a)** The relation between the Gaussian window's standard deviation and the value of the parameter a (from Bergamo et al., 2012).*

***b)** Sketch of the selected receivers and shots for a certain Gaussian window: d is the offset range in which the shots are selected, W is the length of the Gaussian window, ΔW indicates the amount of window shifting (from Socco et al., 2009).*

Figure **4.12** shows an example of *f-k* spectra and DC extraction belonging to two different Gaussian windows, placed along a seismic line that crosses a low-velocity (sand) target.

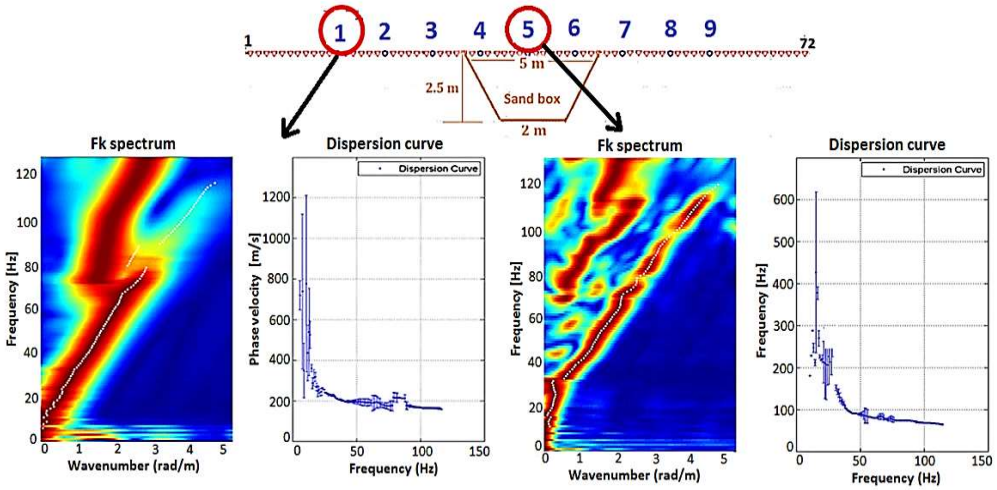

*Figure **4.12** – Example of f-k spectra and the related DC for two Gaussian windows along a seismic line crossing a sand body of low velocity. It can be noticed how the DC related to the Gaussian window 1, placed far from the sand target, registers higher phase velocities compared with the DC related to the Gaussian window 5, placed in the centre of the sand box.*

c) DCs clustering

A clustering algorithm divides the DCs into homogeneous sets, according to metric criteria based on the Euclidean distance between adjacent DCs (Khosro Anjom et al., 2019):

$$d\,(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}, \tag{4.15}$$

where d is the Euclidean distance and x and y are two adjacent dispersion curves represented by phase velocity as a function of frequency. Different clusters are separated by a distance-related linkage criterion.

Figure **4.13** shows an example of DCs extracted along a seismic line crossing a sand target of low seismic velocity. The extracted DCs are divided into two different groups: a group characterized by higher phase velocities, which gathers the DCs placed in external positions with respect to the sand body (plotted in blue) and a group characterized by lower phase velocities, which gathers the DC placed in correspondence of the sand body (plotted in yellow). The clustering algorithm efficiently distinguishes these two groups of DCs according to different phase velocity trends.

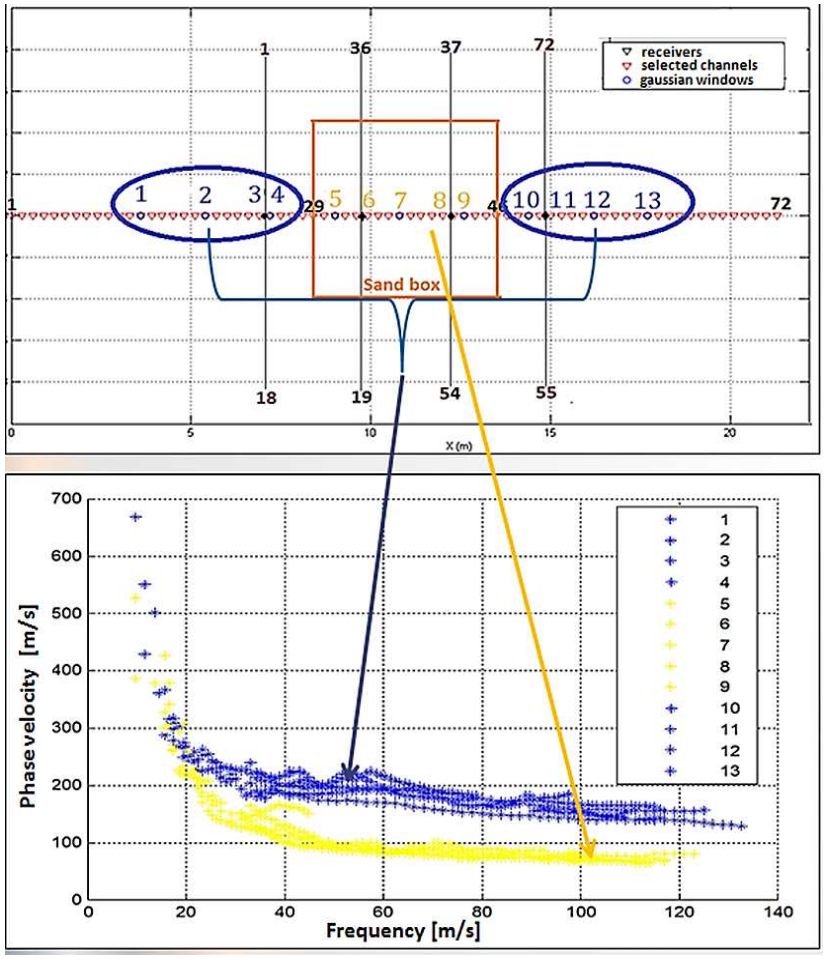

*Figure **4.13** – Example of dispersion curves extracted along a seismic line crossing a low-velocity target. The clustering algorithm can separate the extracted DCs into different groups according to the phase velocity value: i.e. a group of DCs with higher phase velocities (in blue) and a group of DCs with lower phase velocities (in yellow).*

d) Reference DC selection and DC inversion to a reference V_S profile

A reference DC (having a broad frequency band – e.g. Figure **4.14a**) is selected inside each cluster and inverted to a reference V_S profile (e.g. Figure **4.14b**) using an optimized Monte Carlo approach, based on the application of scale properties between the various parameters (Socco and Boiero, 2008; Maraschini et al., 2011). The model parameters considered for the inversion are the *V*s, the thickness of each layer and the corresponding Poisson's ratio. The forward modeling scheme is based on the Haskell and Thompson algorithm (Haskell, 1953; Thompson, 1950; Maraskini, 2008). The results of the MCI are a set of accepted V_S models, inside the imposed level of confidence in a statistical one-tailed Fisher test. The reference V_S model is computed by averaging these accepted models at each depth (Khosro Anjom et al., 2019).

e) Reference V_S transformation into a time-average V_S profile

The reference V_S is transformed into a depth-continuous time-average V_S profile, according to equation 4.16 (Socco et al., 2017):

$$V_{Sz} = \frac{\sum_n h_i}{\sum_n \frac{h_i}{v_{Si}}} \tag{4.16}$$

where h_i and V_{Si} are the thickness and the shear velocity of the ith layer, respectively

The time-average velocity at a given depth allows for the direct computation of the one-way time at a selected depth z (Socco et al., 2017). Socco et al. (2017) have proven that using the time-average velocity instead of the layered velocity for DC computation efficiently reduces the solution non-uniqueness problem while enforcing the robustness of the results.

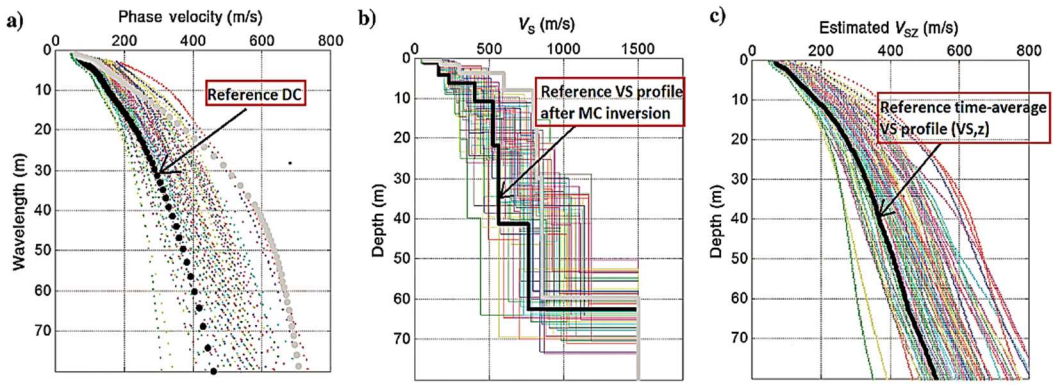

*Figure **4.14** – Example of DC inversion to a reference V_S profile and V_S profile transformation to a time-average V_S profile (Vs,z) **a)** Reference DC of the fundamental mode of Rayleigh waves, plotted as phase velocity function of wavelength (in black). **b)** Example of reference V_S profile (plotted in black) obtained from MC inversion of the reference DC. **c)** Example of time-average V_S profile (Vs,z), corresponding to the reference V_S profile in b) – from Socco et al. (2017), modified.*

f) Wavelength-depth relationship definition

Starting from the reference DC and the corresponding time-average V_S profile, a characteristic relation between the SW Wavelength and the investigation Depth is inferred (W/D relationship) – Figure **4.15**. The W/D relationship is retrieved by finding, for each $V_{S,Z}$ value, the wavelength at which the phase velocity of the DC is equal to $V_{S,Z}$. The W/D pairs obtained in this way are plotted as wavelength function of depth and interpolated with a piece-wise polynomial fit (Figure **4.15b**). From this relation, it is possible to directly retrieve the $V_{S,Z}$ corresponding to each wavelength related to all DCs belonging to a certain cluster (Socco et al., 2017).

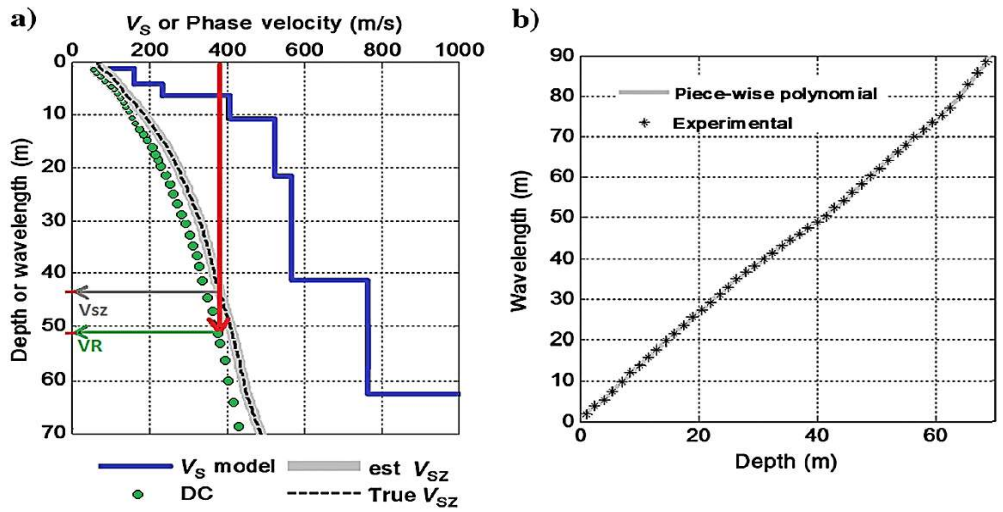

*Figure **4.15** - Example of wavelength-depth relationship construction from the reference DC (plotted in green) and the corresponding reference time-average V_S profile (plotted with grey). **a)** For each identical values of time-average V_S velocity and SW phase velocity (e.g. 390 m/s), the SW wavelength (e.g. about 51 m) is related with the corresponding $V_{S,Z}$ investigated depth (e.g. about 44 m). **b)** The wavelength-depth relationship is built by plotting all the wavelength-depth couples having identical velocity values and interpolating them with a piece-wise polynomial – from Socco et al. (2017), modified.*

g) Transformation of all DCs inside each cluster to time-average V_S profiles

All DCs inside each cluster are directly transformed into time-average V_S profiles by using the retrieved W/D relationship. The obtained V_S profiles provide the one-way time values at each depth within the investigated limit (Socco et al., 2017).

h) Apparent Poisson's ratio estimation for each cluster

For each reference model (DC-V_{SZ}), inside each cluster, the W/D relationship's sensitivity to Poisson's ratio variation is exploited to estimate the apparent Poisson ratio (i.e. the Poison's ratio value that relates the time-average V_S to the time-average V_P; Socco and Comina, 2017a). For example (Figure **4.16**), the synthetic DCs corresponding to each Poisson's ratio value in the range from 0.1 to 0.45 are inferred from the reference V_{SZ}. Further, from each synthetic DC (corresponding to different Poisson's ratio values, which are the same for all the layers of the related synthetic DC) and using the reference $V_{S,Z}$, the related W/Ds are computed. Analysing Figure **4.16**, it can be noted how the Poisson ratio acts on the slope of the W/D relationship: The W/D slope decreases with increasing Poisson's ratio. The Poisson's ratio value that matches the experimental data is retrieved by comparing the experimental W/D with the simulated ones (Figure **4.17**): For each depth, the precise value for the experimental Poisson ratio is inferred, from the experimental W/D (in red), by linear interpolation between the nearest upper and lower values of simulated W/Ds for the same reference V_{SZ} at different Poisson's ratio values (in blue) - Socco and Comina (2017a).

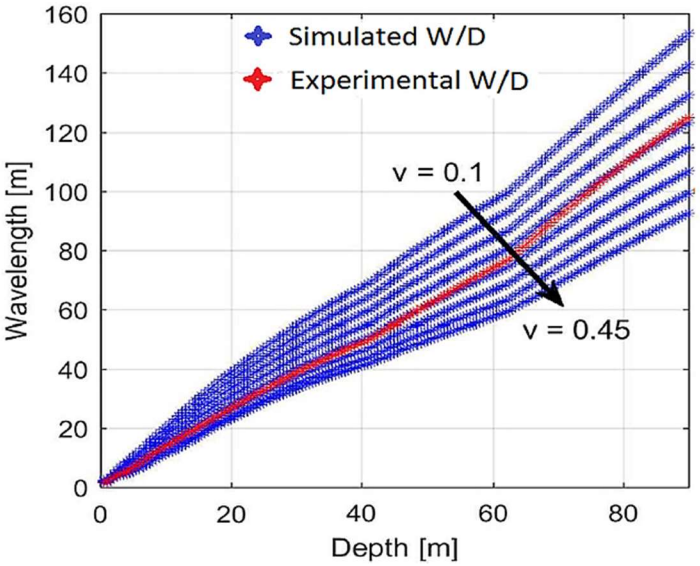

*Figure **4.16** - Example of W/S relationship for the reference model (red), compared with the estimated W/D for the same model and different values of Poisson's ratio (blue) – from Socco et al. (2017b), modified.*

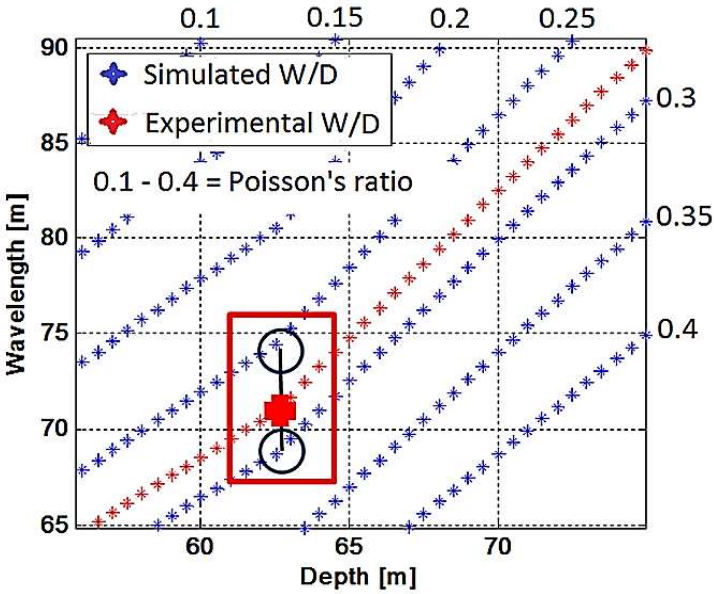

*Figure **4.17** - Example of experimental Poisson's ratio estimation from the experimental W/D and the upper and lower simulated W/Ds. The simulated W/Ds are inferred from different DCs and the reference V_{SZ}. The different DCs are computed using different Poisson's ratio values. Each DC is associated with a constant Poisson's ratio value, from 0.1 to 0.45 (courtesy of Prof. Valentina Socco).*

i) Transformation of the time-average V_S profiles into time-average V_P profiles using the Poisson's ratio

The apparent Poisson's ratio is employed to transform the time-average *V*s profiles, for each cluster, into time-average V_P profiles (Socco and Comina, 2017)

$$\nu_z = \frac{1}{2}\frac{\left(\frac{V_{PZ}}{V_{SZ}}\right)^2 - 2}{\left(\frac{V_{PZ}}{V_{SZ}}\right)^2 - 1} \quad \rightarrow \quad V_P z = V_S z \sqrt{\frac{2\,(\nu_z - 1)}{2\,\nu_z - 1}}, \tag{4.17}$$

where ν is the apparent Poisson's ratio.

j) Time-average velocity profiles conversion to interval velocity profiles

The time-average V_S and V_P velocities are converted to interval velocities by using a total variation regularization on a Dix-type formula (Khosro Anjom et al., 2019):

$$V_{S_i,P_i} = \frac{z_i - z_{(i-1)}}{\frac{z_i}{\overline{V}_{S_i,P_i}} - \frac{z_{(i-1)}}{\overline{V}_{S_{(i-1)},P_{S_{(i-1)}}}}}$$, where z is the thickness of the jth layer. (4.18)

k) *1D to 2D velocity profiles interpolation and 2D to 3D velocity models extension*

The depth-dependent continuous interval-velocity profiles, retrieved along the seismic line, are interpolated to 2D models and then converted to 3D volumes. The continuous velocity profile is well suited for the FWI procedure, avoiding sharp interfaces. The purpose of the 3D extension is honouring the input requirements of the spectral-element FWI code, described in the next chapter.

Limitations and particularities

- Since the results obtained using the above method are data-driven, the experimental uncertainties, as well as the data coverage, directly affect the uncertainties of the results.
- Only the fundamental mode of SW propagation is considered in this work and, therefore, the W/D relationship is related to the fundamental mode's skin depth. In laterally variable sites, the challenges in higher modes identification may introduce additional experimental uncertainties in the DC extraction step. An example related to the "W/D method" application to higher modes of DC can be found in Bamarouf et al. (2017).
- Since the method is based on the direct transformation of local DCs, precaution is required in the presence of complex topography.
- In this study, the estimated Poisson ratio is variable with depth. The highest uncertainty for the Poisson estimation occurs at shallow depth (Khosro Anjom et al., 2019). This result is coherent with the experimental W/Ds relationships' feature, showing an increase in Poisson's ratio sensitivity with depth. Indeed, it can be noted in Figure **4.16** how the W/D relationships related to different Poisson ratios diverge for large wavelengths, allowing thus for a more consistent estimation.
- The first-arrival traveltime tomography is an alternative technique regularly adopted to retrieve the V_P model. Some experiments and results related to this approach are presented in Chapter 9.

4.4 – Highlights

- Most of the SW methods use the geometrical dispersion of Rayleigh wave in vertically heterogeneous media to retrieve the depth-dependent variation of some physical parameters.
- The SW technique is based on the analysis of the dispersion curve. The DC represents the variation of the phase velocity with frequency/wavelength and it is commonly retrieved by spectral analysis in the *f-k* domain ($V_R = 2\,\pi\,f\,/k$).
- SW methods are divided into two main categories: passive methods, characterized by natural sources, generating signals of relatively low frequencies, and active methods, generating signals characterized by higher frequencies.
- In heterogeneous media, the SW propagation is characterized by a *phase velocity* (which is the velocity of propagation of a single phase of the waveform) and *group velocity* (which is the velocity of propagation of a pack of phases around a certain frequency). In a normally dispersive medium, the group velocity is lower than the phase velocity. The group velocity is not a commonly used method for near-surface characterization because of the limited resolution and the interference between the various modes of propagation.
- In the presence of heterogeneities or velocity inversions, SW propagation may exhibit different modes (i.e. *the eigenvalue problem*: different phase velocities can correspond to the same frequency). Accordingly, the DC is an apparent one, composed by the various modes superposition and related to an apparent phase velocity.
- Generally, considering the higher modes in the DC inversion process helps to reconstruct the physical properties of the deeper layers. This happens because the particle displacement associated with the higher modes of Rayleigh wave propagation affects a deeper part of the medium.
- Despite the 1D layered model assumption of SW methods, there are some examples in the literature showing successful reconstruction of smooth lateral variations of the site.
- SW methods are mainly used to retrieve the S-wave velocity, but some studies proposed a method for the estimation of the P-wave velocity as well. This technique exploits the sensitivity, with respect to Poisson's ratio variation, of a certain relationship that exists between the surface wave wavelength and the investigation depth (i.e. the *wavelength-depth relationship*).

- The method used in this study for building velocity models from SW DCs analysis allows for the estimation of both V_P and V_S model. The workflow embeds the following steps:
  - ✓ DC extraction through a Gaussian moving window followed by DC clustering.
  - ✓ The inversion of one DC (representative for each cluster) through a Monte Carlo technique.
  - ✓ The construction of a relationship between SW wavelength and investigation depth (the wavelength-depth relationship) using the representative DC and the corresponding time-average Vs profile (which is referred to as the one-way time at a selected depth z).
  - ✓ Data transform: all DCs of each cluster are directly converted to time-average V_S profiles, using the wavelength-depth relationship.
  - ✓ The estimation of the (apparent) Poisson's ratio from the wavelength-depth relationship and the time-average V_P model computation.
  - ✓ A total variation regularization on a Dix-type formula and the interval velocities computation.
  - ✓ 1D to 2D interpolation (and 3D extension).
- The Poisson's ratio estimated in this study is variable with depth.

Chapter 5

Full-waveform inversion

Contents

5.1 – Overview

FWI is a quantitative seismic imaging technique, based on data-fitting and characterized by the non-linearity of the relationship between the model parameters and the calculated data, the numerical solution of the seismic wave equation, the consideration of the full information in the seismograms and the iterative optimization of the tomographic image (Virieux and Operto, 2009; Fichtner, 2011). FWI aims at obtaining accurate models of the sub-surface physical properties, with a theoretical resolution up to half the local propagating wavelength (Virieux and Operto, 2009). Nevertheless, the limited bandwidth of the source and the partial illumination of the subsurface (related to the acquisition design), renders FWI an ill-posed problem (i.e. many models can explain the data equally well) (Virieux and Operto, 2009). For this reason, the application of different strategies and constrains over the data and model parameter space is required in most of the cases, especially for near-surface investigations, where the SW domination introduces additional non-linearities in the inverse problem formulation.

According to Mora (1988 and 1989), FWI can be considered as a combination of the prestack reverse time migration RTM (which "solves" the high wavenumbers, corresponding to the reflected components) with the tomography (which "solves" the small wavenumbers, corresponding to the transmitted components). Nevertheless, even though both RTM and FWI are based on the full-wavefield modeling (e.g. Gardner et al., 1974; Virieux and Operto, 2009), there are some substantial differences between them. For instance, while the former aims at obtaining an image of the reflectivity field, which is not necessarily a quantitative measure of the subsoil impedance properties, FWI aims at obtaining a quantitative model of the Earth's velocity field (Pratt et al., 1998; Sirgue & Pratt, 2004). Moreover, in RTM the reflected and diffracted seismic wavefield is propagated from the receiver position, whereas in FWI only the data residuals are back-propagated (Virieux and Operto, 2009).

Before the 1980s, most of the seismic deductions concerning the Earth's internal structure were based on the simplified assumption, related to the geometric optics, of representing the seismic waves as rays. The ray theory's widespread use was mainly due to the difficulty of finding analytical solutions for the equation that describes the wave propagation in real heterogeneous media. Nevertheless, tray theory is based on many simplifications and is applied only to the high-frequency arrivals (Fichtner et al., 2006). The wave propagation is described through successive positions of the wave front and the arrival times of seismic waves are related to the velocity distribution only along a curve that connects the source with the receiver (Kennett et al., 1995; Fichtner, 2011). Also, the spatial resolution of the tomographic images is locally limited by the size of the Fresnel zone (Williamson, 1991). FWI overcomes the ray theory's limitations, thanks to the numerical solution of the wave equation, which guarantees the accurate modeling of seismic wave propagation in heterogeneous media. Besides, the numerical solution provides complete information on the seismic waveform, which can be used to improve the tomographic resolution without the need to identify the separation of specific phases (Fichtner, 2011).

5.1.1 – Time domain or frequency domain?

FWI was first implemented in the time domain (Tarantola, 1984a and 1984b; Gauthier et al., 1986; Mora, 1987a; Crase et al., 1990) and subsequently in the frequency domain (Pratt, 1990; Pratt and Worthington, 1990; Pratt and Goulty, 1991; Pratt, 1996; Pratt, 1999; Operto et al., 2006). According to Sirgue and Pratt (2004), in the frequency domain, it is possible to use a larger sampling step and still obtain images not affected by aliasing. More, the frequency domain formulation allows managing a more compact volume of data, with less checkpointing related core storage requests (Virieux and Operto, 2009).

The frequency domain is more appropriate for the selection of certain frequencies of interest, while the time domain allows for the selection of a particular type of arrival, i.e. the P-waves first arrivals, the reflected arrivals, the arrivals corresponding to the surface waves. This aspect led to the development of some FWI strategies in the frequency domain, based on time-domain modeling (e.g. Sirgue et al., 2008). Even if the frequency domain guarantees some advantages, in some cases related to the lower computational cost, most of the FWI applications are nowadays implemented in time-domain, especially when considering dispersive wavefields and the viscoelastic equation.

5.1.2 – Acoustic or elastic approximation?

FWI is a widely used tool in exploration seismology, even though the numerical solution of the (visco)elastic wave equation is nowadays computationally expensive and oversized. For this reason, various FWI experiments are still performed in acoustic approximation. Even if this approach requires fewer computation resources, it introduces amplitude errors in the P-wavefield (except the near-vertical angle reflections; Warner et al., 2012), due to the incorrect modeling of the amplitude variation with offset (AVO = Amplitude Versus Offset) (Virieux and Operto, 2009). In particular, the reflection coefficient in the elastic approximation is lower than its acoustic equivalent, since the incident energy is also distributed between converted phases.

The errors of the acoustic approximation may be also related to the incorrect modeling of the sources and receivers' directivity (Pratt, 1999; Ravaut et al., 2004; Operto et al., 2004; Brenders and Pratt, 2007a,b; Barnes and Charara, 2008). Accordingly, some corrections may be necessary for the sources and receivers' directivity (if the sources and receivers are directional) and for the offset dependent amplitude decay (Virieux and Operto, 2009). The amplitude corrections aim at making the predicted and observed seismic data comparable. In this way, the data misfit depends only on the differences in velocity and not on the inappropriate representation of the amplitudes in the initial model (Isaac and Magrave, 2012).

A proposed solution, aimed at limiting the disadvantages of the acoustic approximation, is combining the acoustic inversion with the classical traveltime tomography (Plessix et al., 2013). However, in some cases, the obtainment of accurate results is highly conditioned by considering the entire information from the recorded seismic data. This means using (at least) the elastic approximation for the wave equation, where the pressure is replaced with

stress and particle displacement velocities. For example, for conventional marine data, the information related to pressure waves is greater than the information related to the shear waves and, therefore, an accurate estimation of the P-wave velocity is possible in acoustic approximation. For an OBC (Ocean Bottom Cable) acquisition, an accurate reconstruction of the S-wave velocity field requires the use of the elastic wave equation, since S waves and converted arrivals are present as well in the recorded seismograms (Raknes et al., 2013). Furthermore, for land acquisitions, where the seismic data are dominated by SWs, using the (visco)elastic equation for modeling the waves propagation is imperative.

Nevertheless, in the elastic FWI experiments, the introduction of new parameters (i.e. the S-wave velocity, SWs) increases the inverse problem's non-uniqueness, due to the increase in freedom degrees and because of different FWI sensitivity to each of these parameters (Virieux and Operto, 2009; Operto et al., 2013; Raknes et al., 2013).

5.2 – Gradient methods insights

The use of global methods for FWI is nowadays computationally expensive, especially for 3D applications. For this reason, most of the FWI codes, as the one used for this work, are based on the deterministic method. There are various gradient-based methods, such as steepest descent (e.g. Fichter, 2011), Newton and Gauss-Newton (Pratt et al., 1998; Askan et al., 2007; Askan and Bielak., 2008; Epanomeritakis et al., 2008), conjugate gradients (Luenberg, 1984; Mora, 1987a, 1987b and 1988; Tarantola, 1984a, 1984b and 1987; Crase et al., 1990; Gilbert and Nocedal, 1992). In this work, the gradient is computed with the adjoint-state technique.

5.2.1 – Conceptual introduction to the adjoint-state technique

In many applications, the FWI gradient is computed using the adjoint state technique (e.g. Tarantola, 1986, Mora, 1987, Tarantola, 1988, Pratt et al., 1998, Plessix, 2006; Fichtner et al., 2006, Virieux and Operto, 2009, Fichtner, 2011).

The adjoint state method is also called "the retro-propagation technique", since it involves the propagation of residuals over time, from the terminal point towards their origin. In particular, the gradient is computed by time-domain zero-lag cross-correlation between the direct wavefield (propagating from the source to the receivers) and the residual wavefield (retro-propagated from the receivers' position to the source). Two forward problems per source need to be solved: one for the incident field computation and one for the residuals back-propagation (Virieux and Operto, 2009). The individual gradients are then superimposed to form a global gradient (Warner et al., 2013).

The result of this operation contributes to the approximate Hessian matrix, whose diagonal terms contain the zero-lag values of the cross-correlation, representing the squared amplitude of the partial-derivative wavefields (Virieux and Operto, 2009). Scaling the gradient by the diagonal terms removes the partial-derivative wavefield amplitude and the residuals (Virieux and Operto, 2009). The off-diagonal terms of the Hessian are computed

through the cross-correlation of the partial derivative wavefields associated with different model parameters (Virieux and Operto, 2009). The application of the off-diagonal Hessian's inverse to the gradient can be thought as a gradient deconvolution from the limited bandwidth effects (described by the off-diagonal terms), caused by partial illumination, parameterization and trade-off (Virieux and Operto, 2009).

According to Tarantola (1984b), the retro-propagated residuals from receivers' position can be thought of as a missing diffraction field, i.e. the part of the signal that is not described by the initial model. Therefore, for retrieving the correct model, the cross-correlation of the incident wavefield with the one not considered in the initial model (corresponding to the perturbation) is required. In the model areas with "non-zero" cross-correlation, the values of the incident field must be modified, considering the perturbed field. The retro-propagation technique has some similarities with the Reverse Time Migration (Lailly, 1983; Pratt, 1999; Warner et al., 2013; Liu et al., 2013; Raknes et al., 2013). However, only the residual field is retro-propagated in FWI, while in RTM the reflected and diffracted wavefield is retro-propagated; besides, as already mentioned, the imaging condition is different: while in FWI the imaging is done in the velocity space, in RTM an image of reflectivity is targeted.

For the gradient computation, a simultaneous availability of the direct and adjoint wavefield is necessary. This problem can be overcome by transforming the initial condition into a terminal one, i.e. solving the wave equation in the forward direction and storing the final state at the boundaries. Then, the adjoint equation can be solved by propagating the regular wavefield back in time, starting from the previously stored final stage, together with the forward propagation of the adjoint wavefield. However, the application of this strategy is possible if the wave equation remains reversible (conservation of wave energy). This allows using the same modeling scheme for the incident and the back-propagated wavefield, which facilitates the gradient computation implementation through domain decomposition in parallel environments (Virieux and Operto, 2009).

Nevertheless, in the presence of anisotropies, dissipations and absorbing boundaries, the elastic energy is attenuated during the wave propagation and, therefore, the early stages of the wavefield cannot be reconstructed from the terminal point. In these conditions, the wavefield value could be memorized at intermediate steps (checkpoints) and used during the application of the adjoint algorithm.

5.2.2 – Strategies for the local minima issue mitigation

For nonlinear ill-posed inverse problems, many local minima are present in the gradient and, therefore, different models can match the same data set. This issue requires the use of various strategies, aimed at guiding the convergence of the iterative algorithms towards the global minimum of the objective function. Among others, the use of alternative norms, data-gradient preconditioning, introduction of weighting operators, regularizations and the multiscale approach can be adopted.

Which misfit functional?

The commonly used L_2 norm assumes a Gaussian distribution of the misfit function (Tarantola, 1987). This assumption can be problematic when outliers are present in the data (Claerbout and Muir, 1973). Moreover, this functional emphasizes the nonlinearity of the inverse problem, due to the presence of many local minima in the gradient. An alternative to the L_2 norm is the L_1 functional (e.g. Brossier et al., 2009a), which is less sensitive to outliers, but still does not help to bypass the local minima issue.

To overcome the limitations imposed by the functional's choice, some strategies, based on the combinations of various norms, have been proposed: e.g. the Cauchy criterion (Crase et al., 1990; Amundsen, 1991; Amundsen and Ursin, 1991), the Huber's norm (Bube and Nemeth, 2007; Ha et al. 2009). Moreover, an envelope misfit function (Bozdag et al., 2011; Luo and Wu, 2015; Yuan et al., 2015; Borisov et al, 2018) can be used when the emphasis of the low-frequency content is required.

Shin and Cha (2008), Shin and Ha (2008), Ha et al. (2012), Ha and Shin (2012a and 2012b) and Shin et al. (2013) proposed the Laplace domain inversion as this algorithm may be more robust when low frequencies are absent in the data. Indeed, the logarithmic objective function has fewer local minima, compared with the objective function in the frequency or time domain. However, only smooth tomographic images can be obtained. Moreover, this strategy is efficient only if the first arrivals can be easily detected and no noise precedes them.

Further, the optimal transport based objective function (Ferradans et al., 2014; Lellmann et al., 2014; Engquist et al., 2016; Métivier et al., 2016a,b; Yang et al., 2018) balances the amplitude of various seismic events. He et al. (2019) proposed the combination of an optimal transport misfit function with a Gaussian time-windowing strategy, to guarantee a more balanced contribution between BWs and SWs while focusing the analysis on the phase shift. Nevertheless, a strong assumption of the optimal transport method is the positivity of the data, while the seismic trace is oscillatory (He et al., 2019). A possible solution for this issue may be the Kantorovich – Rubinstein norm, which can consider negative values (Métivier et al., 2016a and 2016b).

For near-surface applications, Masoni et al. (2013) and Solano et al. (2014) proposed the use of alternative domains for the misfit function computation, such as the ω-p (frequency-slowness) and the ω-k (frequency-wavenumber) domain. In these cases, the event separation through slopes and wavenumbers allows for easier extraction of the kinematic information and better identification of frequency-dependent dispersion effects.

Gradient pre-conditioning

The data-gradient pre-conditioning (Fichtner, 2011) commonly consists in a smoothing operation, which involves the application of a low-pass filter. The smoothing operation removes the small wavelengths that cannot be solved with a certain frequency range.

The introduction of weighting operators modifies the formulation of the objective function (i.e. Tarantola, 1987; Scales and Smith, 1994; Virieux and Operto, 2009):

$$\chi(m) = \frac{1}{2}(d_{obs} - d_{calc})^T W_d\ (d_{obs} - d_{calc}) + \frac{1}{2}\varepsilon(m - m_{prior})^T W_m \left(m - m_{prior}\right), \quad (5.1)$$

where $W_d = S_d^T S_d$ e $W_m = S_m^T S_m$ are weighting operators, representing the inverse of the data covariance matrix and, respectively, model covariance matrix. The term S_d can be implemented as a diagonal weighting operator, which controls the weights of each element of the vector containing the residuals. Operto et al. (2006) implemented the value of S_d as a power of the source-receiver distance, intending to strengthen the far-offset data contribution in the frame of crustal-scale imaging (Virieux and Operto, 2009). The term ${W_m}^{-1}$ (smoothing operator) can be used, for example, to penalize the model's roughness (W_m), leading to smooth final models after inversion (Virieux and Operto, 2009).

Moreover, the model-gradient regularization aims at guiding the model update towards solutions that are coherent with some a priori knowledge. For instance, the Tikhonov regularization (Hansen, 1998) aims at eliminating the small-scale oscillations through minimization. In the general algorithm, the objective function is expressed as the summation of the misfit functional with a quadratic term:

$$\chi_{reg}(m) = \chi_m + R\ \gamma \left\|m - m_{ref}\right\|_2^2 \quad (5.2)$$

where $R \left\|(m - m_{ref})\right\|_2^2$ is the regularization term, R is the Tikhonov matrix, γ is a positive constant that has the role of balancing the contribution of the "pure" misfit term $\chi(m)$, and m_{ref} is a reference model, usually simple and homogeneous.

When $R = I$ (the identity matrix), the regularization acts through the minimization of the norm (norm damping), imposing small differences $(m - m_{ref})$ and leading to homogeneous models close to m_{ref}. This type of regularization (*first-order Tikhonov)* is commonly used for those model areas where the number of unknowns is greater than the observed data. In the *second-order Tikhonov* regularization, the matrix R becomes a finite differences operator D, which acts in the model space, emphasizing the small wavelengths. The algorithm involves the computation and damping of the discrete derivatives, which is equivalent to the implementation of the minimum model variation condition. This type of regularization privileges the presence of small wavelength components, which are the features FWI tries to reconstruct.

A method combining the norm minimization with the derivatives is the *total variation regularization*. This approach aims at obtaining models that are simultaneously characterized by smooth areas and sinuous interfaces (Vogel & Oman, 1996; Vogel, 2002).

The algorithm involves the minimization of the *L1* norm of the differential discrete operator:

$$\chi_{reg}\ (m) = \chi(m) + \gamma\ \|\bar{\nabla}\ m\|_1, \quad (5.3)$$

where γ has the same role as the Tikhonov regularizations.

Which frequency content and offset range?

For most of the exploration scale applications, the success of FWI depends on the availability of large offsets in the recorded data and on the presence of refracted arrivals (Mora 1988 and 1989; Pratt et al., 1996; Ravaut et al., 2004; Warner et al., 2012; Guitton and Alkhalifah, 2013), essential for the correct reconstruction of the medium's large wavelengths (Sirgue and Pratt, 2004). This requirement becomes even more important in complex geological environments (for example to improve the salt imaging: Vigh et al., 2013). This happens because the high-frequency data correspond to an objective function that contains many local minima; the lack of low frequencies could drive the inversion towards one of them. This requirement aims to avoid cycle-skipping, which leads to the association of the predicted data with the wrong event in the recorded data (Virieux and Operto, 2009; Ratcliffe et al., 2013; Morgan et al., 2013).

A correct FWI configuration (which avoids cycle-skipping) requires the difference between the observed and the modelled (predicted) wavefield to be smaller than half-period (Beydoun and Tarantola, 1988; Mulder and Plessix, 2008; Warner et al., 2013). This condition is honored if the initial model is accurately collocated in the basin of attraction of the objective function's global minimum (Virieux and Operto, 2009). Otherwise, the predicted wavefield would be related to a delayed (wrong) cycle of the observed wavefield, leading the optimization towards a so-called cycle-skipped local minimum (Figure **5.1**). The dimension of the attraction basin and the half-cycle of the wave period increases with decreasing frequency.

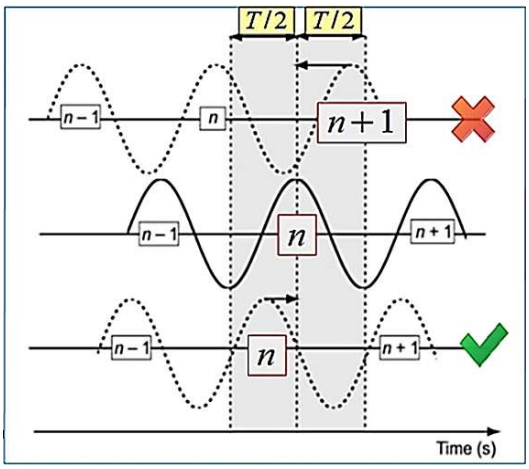

*Figure **5.1** – Illustration of cycle-skipping. The observed data (shown in the center) are schematized through a monochromatic wave of period T as a function of time. Predicted data are schematically shown at the top and at the bottom. Between the predicted data shown at the top and the observed ones, there is a delay greater than T/2; therefore, the cycle n of the observed data is incorrectly related to the cycle n+1 of the predicted ones. Contrariwise, between the observed data and the predicted ones shown at the bottom, there is a phase shift lower than T/2; consequently, the cycle n of the observed data is correctly related to the cycle n of the predicted data (from Virieux and Operto, 2009, modified).*

An alternative strategy for the lack of large offsets compensation is combining the waveform inversion with the classical first arrival tomography (Pratt and Goulty, 1991). The latter could be used for the initial model building (Shin and Cha, 2008). However, picking the first arrivals in the presence of low-velocity areas may be tricky.

Multiscale or layer-stripping approach?

The multiscale approach (Bunks et al., 1995; Sirgue and Pratt, 2004) aims at guiding the iterative inversion towards the global minimum of the objective function, by progressively increasing the high-frequency content (small wavelengths) – Figure **5.2**. The maximum frequency of each group must be sufficiently low to prevent cycle-skipping (Virieux and Operto, 2009). FWI is more tolerant to velocity errors in the initial model when only low frequencies are present since they are less likely to lead to errors of more than have a wavelength between the observed data and the predicted ones (Pratt, 1999; Virieux and Operto, 2009).

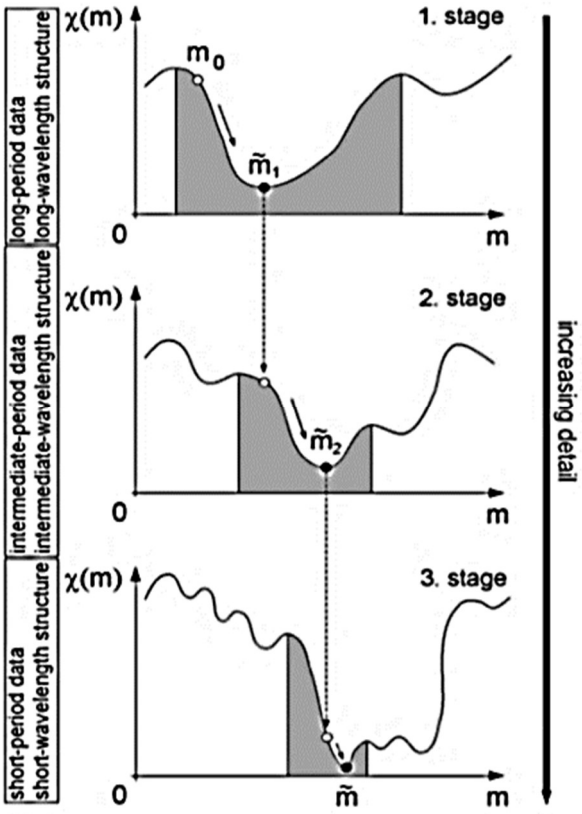

*Figure **5.2** - Illustration of the multi-scale approach for an inversion performed in three stages. The grey area represents the attraction basin of the objective function's global minimum. The first stage of the inversion uses only the large wavelengths. The inversion starts from the initial model m_0 and proceeds towards a better model $\widetilde{m_1}$. In the second stage, the best model of the first iteration is used as a starting model for the second iteration. In the third stage, the model $\widetilde{m_2}$ (obtained in the second iteration) is used as a starting point for the third iteration. Passing from the first stage to the third one, the objective function becomes increasingly wrinkled and the attraction basin of the global minimum becomes narrower (from Fichtner, 2011).*

However, the robust application of FWI on data containing large offsets still represents a numerical challenge, due to the nonlinearities introduced by the presence of a wide range of wavenumbers and variable incidence angles (Sirgue, 2006). If low frequencies are not present in the recorded data, the layer-stripping approach (high-to-low frequency) can be used (Kolb and Canadas, 1986; Shipp and Shing, 2002; Wang and Rao, 2009; Virieux and Operto 2009; Masoni et al., 2016). This approach is based on the progressive introduction into the inversion scheme of the later arrivals and higher offsets. The model is gradually updated from the shallow parts to the deeper ones, according to an empirically established relation between a time window and a spatial window (which acts as a function of offset).

Tapering and time windowing

The above techniques may not be enough to guarantee reliable results for FWI in the presence of complex structures. A further applicable strategy could be muting the arrivals that cannot be predicted with a given approximation of the wave equation, i.e. the PS converted arrivals in the case of acoustic approximation. This process involves the application of a window in the time domain, aimed at selecting the first arrivals. This is equivalent to the selection of high-angle data components, which allows reconstructing the large wavelengths of the medium (Pratt, 1999; Virieux and Operto, 2009).

Moreover, a time window could isolate a specific phase or a specific parameter of interest, such as the converted PS phases (Shipp and Singh, 2002; Sears et al., 2008; Brossier et al., 2009b), body waves from surface waves (Trinh et al., 2019) in the case of the FWI carried out in elastic approximation.

***

Most of the above-presented strategies are commonly applied for FWI tests performed for crustal-scale exploitation data. For the geotechnical scale exploration, characterized by a different acquisition geometry (e.g. in terms of source and group interval), the consequent presence of higher frequencies in the recorded data, the presence of energetic SWs and strong attenuation, additional FWI strategies may be needful, as well as the use of the viscoelastic equation for simulating the wave propagation. Besides, a very accurate initial model may be a key element for properly guiding and speeding up the convergence of the inversion algorithm.

Some journal papers published over the last decade, related to elastic FWI applications for near-surface imaging, are presented in the next section. This review does not try to be exhaustive while it aims at providing a more extended introduction regarding the FWI application to near-surface data sets.

5.3 – Examples of FWI applications to shallow seismic wavefields

Amrouche and Yamanaka (2015) developed a time-domain waveform inversion method for 2D shallow soil characterization. The forward modeling for P-SV wavefield simulation is based on a 2.5D staggered-grid FD wave propagation solver. A hybrid heuristic inversion, combining the Genetic Algorithm and Simulating Annealing, is adopted (Yamanaka, 2007). For soil modeling, the soil layers are divided into several blocks characterized by discontinuous interfaces. The results proved the effectiveness of the method to reconstruct the soil structures with a blind layer, in noisy environments, using a single source. The application on field data (using two shots on a survey line of 30 m) provided 2D soil models that nicely agree with borehole data (except for the shallow layer that was subject to interventions for construction purposes after the borehole drill).

Borisov et al. (2018) applied a 3D elastic FWI of SWs on synthetic data (simulated from the SEAM phase II foothills model – Oristaglio, 2012), considering irregular topography. Their approach relies on the use of an envelope-based objective function to invert for shallow large-scale heterogeneities in the first stages. It proved to be well suited for recovering near-surface structures and providing a V_S initial model for a conventional BWs FWI scheme. In the later stages of the inversion, the authors use a waveform difference objective function to obtain a higher resolution model. Since irregular topography is concerned, the forward modeling is based on a spectral element wave propagation solver (SPECFEM3D). No multiscale approach is used for FWI. The authors invert simultaneously for V_P and V_S and keep the density as constant. The quasi-Newton method l-BFGS is used for optimization while the step length is computed using the safeguarded backtracking line search (Modrak and Tromp, 2016). Generally, the initial models are smooth versions of the target models. The authors concluded that the envelope misfit function is effective in minimizing the cycle-skipping issues during SWs inversion and that SWs are useful for retrieving complex shallow structures.

Borisov et al. (2020) applied a 2D FWI on exploration land data, characterized by a maximum offset of 12 km and the lowest available frequency of 5 Hz. They invert for SWs first, to constrain the shallow part of the V_S model, and incorporate the BWs in the later stages of the inversion. At the final stage, which involves the entire offset range and frequency band between 5 Hz and 15 Hz, SWs and BWs are inverted simultaneously. The forward modeling relies on a spectral element-based wave propagation solver (SPECFEM2D). The initial V_P model is derived from first break traveltime tomography while the initial V_S model is derived from the initial V_P using a V_P/V_S ratio of 1.7. Frequency and offset-continuation approaches are used for mitigating cycle-skipping effects. To account for viscoelastic effects present in the field data, the authors use a normalized cross-correlation objective function and a source wavelet correction. However, they state that including a model for attenuation in the forward simulation guarantees better results.

The overall results, compared with well-log information, indicate a better V_P/V_S and data fitting after FWI compared to the initial model configuration.

Bretaudeau et al. (2013) proposed a 2D FWI approach, based on frequency-domain viscoelastic modeling, for building 2D images of V_P and V_S models from land seismic data. The reference synthetic model chosen for investigation is a vertically layered one. The authors applied the method on both synthetic data and laboratory data obtained by small-scale physical modeling (on a laser-ultrasonic bench). Strong SWs and weak diffractions are present in the data. The method proved to be successful in reconstructing the 2D near-surface model, even in the presence of noisy data. The authors explained the presence of some artefacts on the base of the strong nonlinearity introduced by SWs, the offset and frequency dependence of the S/N, the source signature inversion and the gradient features close to the sources.

Butzer et al. (2013) applied a 3D elastic FWI to synthetic data belonging to random medium velocity structures that resembles a realistic crystalline rock environment. The FWI implementation is based on the adjoint state method, with an optimization regarding the runtime and storage cost based on a time-frequency approach. The gradient is computed from monochromatic frequency-domain particle-velocity wavefields, which are calculated with a time-domain FD velocity-stress wave propagation solver. The authors applied a multiparameter (V_P and V_S) inversion for two different acquisition geometries: 3D and 2D. They run 3D FWI for both acquisitions and 2D FWI for the 2D acquisition data set for comparison purposes. The results proved that the 3D FWI applied to the 3D data set is capable to reconstruct V_P and V_S 3D structures with a resolution of about one wavelength. 3D structures in an area around the acquisition plan can be identified when 3D FWI is applied to multicomponent cross-well data. The authors stated that, if homogeneous initial models are used, sufficiently low frequencies must be included in the inversion. When a 2D acquisition layout is used for 3D FWI, the resolution of the results is lower compared to the resolution of the 3D acquisition geometry results. The results from the 2D FWI contain artefacts caused by 3D scattering, especially when inverting at high frequencies for reconstructing small-scale heterogeneities. The authors stated that a 3D FWI would be preferred to a 2D FWI in presence of 3D small-scale heterogeneities.

Chen et al. (2017) applied a combined workflow based on frequency-dependent traveltime tomography (FDTT) and FWI to 2D near-surface P- and SH-wave seismic data for detecting a buried tunnel characterized by concrete walls. The FDTT inverts for P- and SH-waves picked traveltime at 250 HZ and provides wavelength-background initial models for FWI. FWI, conducted in the frequency band of 18-54 Hz for P-wave data and 16-50 Hz for SH-wave data, improved the resolution of the initial models. The top of the tunnel was imaged at the correct location (1.6 m depth) as a high-velocity anomaly while the void space inside the tunnel was imaged as a low-velocity anomaly in the V_P model. The authors also applied the conventional ray-theory infinite-frequency traveltime tomography (IFTT) for comparison purposes. The experiments based on FDTT provided a better initial model for FWI compared with the IFTT. This led to a more accurate velocity estimation after FWI,

with better detection of the buried anomaly's amplitude and position. The authors stated that FDTT may outperform IFTT for the cases where the ray theory is valid.

Dokter et al. (2017) developed a 2D FWI for simultaneously reconstructing the V_S and density model from SH and Love waves. The purpose of the study was mapping historical building foundations buried in a highly heterogeneous medium. The authors inverted for both synthetic and real data sets acquired in a near-surface environment. The field data were recorded along 2 parallel SH profiles with a maximum offset of 24 m. The frequency content is in the range between 5 Hz and 80 Hz. The objective function is computed using the global correlation norm. The initial velocity model is characterized by a vertical gradient and was determined using the Wiechert-Herglotz inversion of the first arrival picks. The results proved the small influence of the viscoelastic effect on the waveform, as well as the small influence of the density on the final data fitting. Overall, the results proved the effectiveness of the 2D SH waveform inversion to image small-scale soil structures (with a size of 1 m or 2 m and down to a depth of 6 m), although the successful reconstruction of the foundation walls was not achieved.

Dou and Ajo-Franklin (2014) applied full-wavefield inversion of SW to an inversely dispersive SW data set acquired in a permafrost site (Barrow-Alaska) for mapping embedded low-velocity zones. Such a data set provides unusual dispersion spectra characterized by dominant higher models and leaky modes, difficult to be picked. To overcome this difficulty, their full-wavefield method does not rely on individual modal curves. It is based on the inversion for velocity models that can best fit the dispersion spectra instead of dispersion curves. To constrain the velocity structures, their method uses the entire signal content of the wavefield, including higher modes and leaky models. The results are consistent with results from electric resistivity surveys. They revealed the presence of embedded low-velocity zones underlying the ice-rich permafrost site. The full-wavefield inversion of surface waves was efficient if identifying these unfrozen zones (that may correspond to saline-pore waters that prevent the soil from freezing), characterized by velocity values of 70%-80% lower than the overlying frozen material. However, since a parsimonious model parameterization must be used to contain the inverse-problem high-dimensionality, the obtained velocity models are smooth, without small-scale features. The authors concluded that the models obtained from their full-wavefield method could be good candidates as starting models for a classical time- or frequency-domain waveform inversion when a model refinement is required.

Gross et al. (2014) investigated the role of attenuation for 2D FWI application to shallow wavefields. They proposed a source wavelet correction strategy to partially compensate for the residuals between elastically and viscoelastically simulated data. The source wavelet correction filter is retrieved by a stabilized deconvolution technique including an offset dependent weighting data for the high-frequency band; such correction reduces the high-frequency ringing present at far offset in the elastically modeled data. The authors showed that the source wavelet correction can act only as a frequency-dependent filter and not as a distance-dependent filter. They also stated that this strategy is not fully

efficient in producing the frequency-dependent amplitude decay with offset present in viscoelastic wavefields. The field data set used in the study was acquired over a predominantly depth-dependent structure. The estimated Q factors vary between 10 and 50, characteristic for shallow environments. The authors used a time-domain 2D viscoelastic FWI, with a priori fixed Q parameters and correct V_P and density. A reference model was inferred from the data using joint inversion of Fourier-Bessel expansion coefficients (for the V_S model) and P-waves traveltime (for the V_P model) – Forbriger et al. (2003 a,b). They use a direct wave 3D to 2D transformation (Forbriger et al., 2014). The forward modeling is based on 2D FD time-domain (Bohlen, 2002) stress-velocity formulation on the standard staggered grid (Virieux, 1986; Levander, 1998).

The authors obtained equally well reconstructed V_S models when using Q_P=Q_S=10 and Q_P=Q_S=50. The V_S model obtained by FWI was reconstructed with less accuracy when using elastic modeling and source wavelet correction than when using appropriate viscoelastic modeling. The authors proved that negative effects due to the wrong choice of the Q factors can be reduced by adopting the source wavelet correction, aimed at adjusting the bandwidth of the modeled data with respect to the bandwidth of the observed data. Moreover, the authors stated that too-high wrong Q values can be compensated better by source wavelet correction than too-low wrong Q values, as the correction filter can act also as a low-pass filter.

Gross et al. (2017) proposed a 2D elastic FWI workflow and tested it on a field data set acquired over a predominantly depth-dependent velocity structure, with a discontinuity at 6.8 m (probably related to the water table). The obtained velocity models are compared with the results of a 1D inversion based on wavefield spectra (Fourier-Bessel expansion coefficients) – for the V_S model and first arrival P-wave traveltime – for the V_P model. The initial V_P model for 1D inversion is retrieved by refracted P-waves inversion, while the V_S model is obtained dividing the V_P by $\sqrt{3}$. The authors used the 2D time-domain FWI code developed by Köhn (2011) and adopted viscoelastic forward modeling based on stress-velocity FD formulation (Bohlen, 2002). The viscoelastic damping implementation is based on a generalized SLS with 3 relaxation mechanisms, relatively constant over the frequency band of the field data (5 Hz – 70 Hz). The boundary reflections are reduced using convolutional perfectly matched layers (CPMLs) – Komatitsch and Martin (2007); Martin and Komatitsch (2009). The free surface is implemented by image technique (Levander, 1988; Robertsson et al., 1995).

The misfit function is based on the least-squares of the normalized amplitude signals proposed by Choi and Alkhalifah (2012). The authors motivated such a choice with the normalized misfit's insensitivity to amplitude decay with offset caused by geometric spreading, which guarantees a uniform contribution of the near and far-offset traces. On the other side, the relative amplitude differences, caused by the frequency-dependent amplitude decay with offset due to anelastic damping, remain effective. The authors inverted simultaneously for V_P and V_S and kept the density (1700 kg/m^3 – 2000 kg/m^3) and the quality factors (Q_P=Q_S=20; Q_P=Q_S=50) as constant. The quality factors are

estimated from the field data by comparison with viscoelastically modelled data for different quality factors. As an initial model for FWI, the authors used a V_P model retrieved from first arrival traveltime and a V_S model derived by taking a vertical gradient down to 9 m depth. The choice of the vertical gradient highly influenced the FWI results. The FWI workflow is based on different strategies, such as multiscale inversion according to the frequency intervals of 5 Hz – 10 Hz – 15 Hz – 20 Hz – 30 Hz – 40 – Hz – 55 Hz and 70 Hz, a source wavelet correction filter estimated by stabilized deconvolution, for each quality factor and each shot (Gross et al. ,2014), smoothing the inferred subsurface models, a gradient preconditioning taper by semicircular windows in the sources' vicinity, gradient smoothing with a 2D median filter for each iteration step and forcing the V_P/V_S ratio to be greater than 1.56.

Krampe et al (2019) investigated the effects of anisotropy on surface waves and FWI applied to shallow seismic wavefields in vertically transversely isotropic (VTI) media. As synthetic experiments (performed using an FD scheme – Bohlen, 2002) proved that Love waves are more affected by vertically transverse anisotropy than Rayleigh waves, the authors performed 2D anisotropic FWI tests for shallow seismic reconstruction using Love waves and horizontally polarized S waves. The gradient of the objective function is derived using the Lagrange multipliers. The inversion is performed for the vertical velocity, horizontal velocity and density, following a multiscale approach from 5 Hz to 95 Hz with an increasing step of 10 Hz. A Gaussian taper with a radius of 0.5 m is applied at the source locations. As an initial model, a linear gradient velocity and density model is used for one of the reconstructing tests, while one isotropic model is used for another test.

Experiments showed how, when using an isotropic model or a VTI anisotropic model, there is a difference between the corresponding phases up to three times the dominant period. This means that multiple cycle-skips may occur if isotropic FWI is applied to anisotropic data. More, when isotropic FWI is applied to VTI data, the anisotropic effects on the waveform that cannot be explained by the V_S model are projected into the density model. The errors in the initial density model also lead to a worse reconstruction of the velocity models because anisotropic effects are partially explained by the velocity model. Differently, for isotropic data, the influence of a wrong density model is minor than anisotropic effects; the potential inaccuracy of the initial density model does not influence significantly the accuracy of the VTI velocity reconstruction. To avoid overcompensation of wrong densities in the velocity models or the projection of anisotropic effects into the density model, the authors proposed to keep the density model fixed during the first stages of the inversion.

In one reconstruction test, the targeted anomaly is underestimated by 50 m/s after FWI, probably because of the limited bandwidth and limited Love waves penetration depth. Overall, the results on synthetic data proved the effectiveness of anisotropic FWI of shallow seismic Love waves for reliable near-surface characterization. The reconstruction is accurate for the horizontal and vertical S-wave velocities, but the density cannot be retrieved accurately with a multi-parameter FWI. When applying isotropic FWI on VTI

data, the density inversion is unreliable as it is influenced by unresolved structures in the velocity model. The authors proposed to exploit this behavior for detecting the presence of anisotropy through a preliminary isotropic FWI with fixed density, followed by VTI FWI. If artefacts are present in the isotropic FWI results, the medium is anisotropic. In such a case, the VTI FWI can correct these artefacts.

Masoni et al. (2014) investigated the efficiency of using alternative data domains (i.e. the frequency-slowness and the frequency-wavenumber domain) for the misfit function minimization when considering local optimization procedures for near-surface seismic imaging by 2D FWI of SW. The frequency-dependent dispersive effects of SWs are explicitly considered in the FWI scheme. Data normalization is applied before the domain transformation (linear moveout – for the frequency-slowness domain and Fourier transform for the wavenumber domain). In these alternative domains, the separation of various seismic events through slopes or wavenumbers allows for more efficient extraction of the kinematic information, while the stacking increases the misfit function's robustness in the presence of noise. Since the absolute value of the data is considered, the misfit function is less sensitive to the phase of the source wavelet. The elastic wave equation is solved with a FD scheme. The gradient computation is based on the adjoint approach while the adjoint source term is computed using Lagrangian multiplicators. The space derivatives are approximated by central finite-differences while a second-order leap-frog scheme is used for the explicit time integration. An example of 2D synthetic data application proved a good reconstruction of the V_S model. Such reconstruction seems to be more accurate when using the frequency-wavenumber based objective function than when using the frequency-slowness one.

Nguyen and Tran (2018) proposed a 3D elastic FWI workflow for site characterization, based on the adjoint state method. The forward modeling relies on the velocity-stress staggered-grid FD formulation, perfectly matched layers for boundary truncation and free-surface condition. The geometric spreading attenuation is accounted for by the 3D forward modeling scheme. An offset dependent correction factor acts on the estimated data and (partially) compensates for attenuation due to material damping. A gradient preconditioning and scaling with dept, together with a Tikhonov regularization are implemented to reduce the inversion artefacts and acquisition footprints. For the field data application, the estimated source signature is updated at the beginning of each inversion iteration. A layered V_S initial model is estimated through spectral analysis of the data while the V_P model is inferred from the V_S model using a constant Poisson's ratio of 1/3. The authors apply a constraint on the V_P to maintain a Poisson's ratio of more than 0 and less than 0.45. They performed two inversion runs, in the frequency band of 5 – 20 Hz and 5 – 30 Hz, respectively. The results for synthetic data proved the algorithm's effectiveness to characterize laterally variable layers. The results for real data are consistent with the values from standard penetration tests and proved the method's effectiveness to identify embedded low-velocity layers.

Nuber et al. (2016) investigated the effects of neglecting the topography in elastic 2D FWI. They performed several tests on synthetic data characterized by high contrasts in velocity and density. The experiments proved that, if variations in topography (although small) are not accounted for correctly, artefacts are introduced in the velocity model. This fact produces a decrease in spatial resolution for greater depth features. The artefacts appearance at greater depths is explained on the base of decreasing sensitivity: The effects of systematic errors tend to be more pronounced in the low-sensitivity zones. The authors state that the above effects are very pronounced when the topography fluctuations have a wavelength comparable with the minimum seismic wavelength. Moreover, they state that neglecting topography with amplitude fluctuations greater than the minimum seismic wavelength also produces artefacts in the velocity models. Experiments performed with data containing random noise proved that artefacts introduced by the random noise are smaller than those caused by neglecting the topography. Overall, the authors concluded that incorporating the topography is essential when successful results are expected from elastic FWI applied to land data. They also stated that, when elastic 2D FWI is applied to data recorded in areas with complex 3D topography, minor topographic variations perpendicular to the 2D profile may cause artefacts.

Nuber et al. (2017) demonstrated, employing optimized experimental design (ED) techniques, that a very dense spatial sampling and a great number of sources is not mandatory for obtaining a detailed image from acoustic or elastic FWI. Using a noise-free acoustic example, the authors showed that only a few carefully selected source positions are required for computing high-quality images. However, in the presence of noise and complex structures, additional sources may help to increase the inversion's stability. More, using elastic FWI with noise-contaminated data and free-surface boundary condition, they proved that a receiver spacing with the dimension of the expected minimum shear wavelength is enough for obtaining accurate results. They also stated that horizontally oriented sources and multicomponent receivers are preferred for elastic 2D FWI. When only single component sources and receivers are available, the authors recommend using z-directed sources and x-directed receivers. Their ED procedure requires in input one a priori smooth model. The procedure also assumes that a suitable inversion strategy that accounts for the FWI nonlinearity is available. The authors stated that their approach is suitable for analyzing large-scale data while extracting for the inversion only a small data subset representative for the entire data set.

Pan et al. (2016) used the Love-wave waveform inversion in the time-domain to estimate the shallow 2D V_S model. The forward modeling scheme is based on the FD method. The source effect is removed from the data by the deconvolution technique. The authors divide the model into blocks of different sizes that depend on the Love waves resolution. The velocities of each block are updated using a conjugate gradient algorithm. The effectiveness of the method is proved by two synthetic applications (noise-free and noise-contaminated) and a real case.

Pan et al. (2019b) provided a state of the art of the MASW and FWI application to shallow seismic wavefields as well as a comparison of the efficiency of these methods in reconstructing near-surface structures. The authors tested various FWI approaches (i.e. the classical, the multiscale, the envelope based, and the amplitude-spectrum based) and compared the resolution of MASW and FWI using a checkboard method. A numerical example is used to compare the shape of the objective functions of MASW, classical FWI and modified FWI approaches. The authors proved how the classical FWI has the highest resolution (and nonlinearity), MASW has the lowest resolution (and nonlinearity), while the other FWI approaches have an intermediate resolution between the classical FWI and MASW. They also proved that, among the tested methods, the classical FWI shows the highest dependency on the initial model's accuracy. A sequential MASW (for 1D velocity model building) and multiscale FWI strategy is applied to two different field data sets. The implementation relies on the staggered-grid FD method for the 2D elastic wave propagation simulation (Virieux, 1986). A conjugate gradient method is adopted as an optimization algorithm. The initial V_S model is built by applying multiple 1D MASW to adjacent shot gathers (and smoothing the results) while the initial V_P model is derived from f. a. traveltime tomography. Comparisons with borehole data and GPR profiles validated the high-resolution character of the obtained S-wave velocity images.

Romdhane et al. (2011) applied 2D elastic FWI for reconstructing shallow structures in the presence of complex topography. The forward modeling is performed in the frequency domain and is based on a mixed finite-element P_0-P_1 discontinuous Galerkin method aimed at efficiently handling the presence of complex topography. The inversion relies on the limited-memory quasi-Newton method. The numerical experiments are based on a realistic model which reproduces an earthflow structure. The model is characterized by strong velocity contrasts in correspondence of the layers and strong lateral variations. The authors experimented a hierarchical preconditioning strategy, based on a simultaneous multifrequency inversion of damped data, aimed at mitigating the strong nonlinearities introduced by SWs while alleviating the near-surface related effects. The results confirmed that the simultaneous inversion of damped data, based on the progressive introduction of converted and free-surface waves, outperforms the single-frequency inversion approach. Moreover, the experiments proved that using the vertical component only for the inversion may not lead to a significant loss in resolution of the reconstructed model (the degradation of the model is only 24%). Besides, experiments with increased source sampling proved that severe aliasing may occur if the source sampling is less than three times the maximum of one half-wavelength.

Schäfer et al. (2014) proposed a single-trace hybrid transformation correction filter, adapted for direct BWs and Rayleigh waves, for retrieving the equivalent line-source response for point-source data. The authors recommend scaling the seismograms with $\sqrt{2\,r\,phase\ velocity}$ for small receiver offsets r and then gradually applying the classical $\sqrt{t^{-1}}$ time-domain taper and scaling the waveform with $r\sqrt{2}$ for larger receiver offsets. They demonstrate the performance of the hybrid transformation for 2D heterogeneous structures. The amplitude and phases for all shots and components (vertical and radial)

nicely agree with the reference line-source data (except for back-scattered waves) while the classical Fourier-Bessel transformation produces significant artefacts. The authors applied 2D FWI to point-source and line-source synthetic data to evaluate the various performances in reconstructing the reference model. The results obtained from FWI applied to the point-source waveform with no explicit correction show moderate artefacts. The authors stated that better overall reconstructed models and inverted data are obtained when applying the hybrid transformation to the field data prior to 2D FWI.

Smith et al. (2019) applied a time-domain 3D elastic FWI to reconstruct a known tunnel placed at 10 m below the surface. The implementation relies on the spectral element wave propagation solver SPECFEM3D (Komatitsch and Tromp, 1999) and the SeisFlow tool for inversion (Modrak et al., 2018). The inversion relays on the quasi-Newton BFGS method; the step-length is computed using a safeguard backtracking line search (Modrak and Tramp., 2016). During inversion, each trace is normalized by its L2 norm and gradient smoothing is applied after each model update; a Gaussian mask is also applied around each source. The inversion is performed simultaneously for V_P and V_S while keeping the density fixed. The initial model is estimated from SW methods. The workflow is also based on the use of V_P/V_S ratio for removing artefacts from the inverted model and highlighting the localization of the tunnel.

The conducted experiments proved that the use of shear sources and horizontal receivers, aside with vertical sources and vertical receivers, can improve the overall velocity structure. The authors proved that better results are obtained after inversion when removing the instrument response of the horizontal sensors. The vertical sensors' instrument response is not removed because their resonance frequency is lower (4.5 Hz) than the horizontal sensors' resonance frequency (14 Hz), and the synthetic and observed data look similar in the bandpass chosen for inversion (8 – 20 Hz). The tunnel cannot be identified in the V_P results (neither in the vertical not in the horizontal slice), but the high V_P/V_S values identified in the results from combined V_P – V_S inversion highlight the tunnel's position in both vertical and horizontal slices. Accordingly, the 3D contour of the high V_P to V_S ratio would be an efficient way to identify the tunnel's location. The authors stated that the results from 3D FWI do not bring significant improvement in imaging the tunnel compared with the results from 2D FWI presented in Wang et al. (2019), although the depth of the tunnel is more accurate in the 3D results.

Sun et al. (2017) developed a new strategy for near-surface imaging, consisting of alternatively applying first-arrival traveltime tomography and waveform inversion. The first arrival traveltime tomography is based on wave front ray tracing and nonlinear inversion. A multiscale approach is used for waveform inversion and a wavelet transform is applied to the data for mitigating cycle-skipping issues. The method is applied to one synthetic and two field data sets. The results proved that the alternation of inversions that minimize two separate objective functions leads to a better constraint of near-surface structures than the application of waveform inversion alone. The results obtained from the field data application are artefact-less and agree with known geological structures present

in the investigated sites. The authors stated that the wavelet-based multiscale waveform inversion may handle the cycle-skipping problem better than the frequency-based multiscale inversion of Bunks et al. (1995). They also stated that, in their approach, the integration of traveltime tomography plays an important role in avoiding the solution trapping into local minima.

Tran et al. (2019) proposed a 3D time-domain FWI approach for near-surface site characterization, based on elastic modeling and Gauss-Newton inversion, using the complete inverse Hessian. The numerical implementation relies on the classical staggered-grid FD method (Virieux, 1986). The free-surface boundary condition is implemented using the image technique (Robertsson, 1996) while perfectly matched layers are used for the other boundaries (Komatitsch and Martin, 2007). Some constraints are applied for regularization to increase the invertibility of the approximate Hessian matrix, which acts as a weighting function, avoiding the use of vector gradient scaling. The algorithm is tested on both synthetic and real data and the results are compared with data from standard penetration tests. The medium is a layered one with a low-velocity layer embedded at about 5 m depth. The initial V_S model is estimated via spectral analysis, the initial V_P is calculated from V_S (using a constant Poison ratio of 1/3) and the density is 1800 kg/m^3. The authors invert for V_P and V_S while keeping the density fixed. An offset-dependent correction factor acts during inversion to account for viscoelasticity. Two inversion runs are performed, using the frequency intervals of 5 – 20 Hz and 5 – 30 Hz. The embedded low-velocity zone is recovered alter FWI and confirmed by the results from standard penetration tests. The authors also compare the results from 3D Gauss-Newton FWI with results from cross-adjoint 3D FWI (reported in Nguyen and Tran, 2018), in which they use the same initial model and frequency content. Even if both FWI implementations produced accurate V_S models, the authors stated that the V_S results from Gauss-Newton FWI are more consistent with the N-values from standard penetration tests. Accordingly, the cross-adjoint 3D FWI would be less effective for detecting low-velocity zones compared with the Gauss-Newton 3D FWI.

Wang et al. (2019) applied a 2D time-domain FWI scheme, based on the FD method and l-BFGS algorithm, for detecting a 10 m deep (1.5 m width x 1.76 m high) known tunnel. The optimization is based on the normalized L2 norm (Choi and Alkhalifah, 2012). The acquisition survey is perpendicular to the tunnel axis. The initial V_S model is estimated by merging results from MASW and joint analysis of refractions with surface wave (JASW – Ivanov et al., 2013). The initial V_P model is inferred multiplying the V_S model by a constant V_P/V_S ratio. The value of this ratio (2) was estimated from previous studies conducted over the investigated area, including traveltime tomography (Schwenk et al., 2012) and diffraction research (Peterie et al., 2016). The tunnel-imaging FWI workflow is based on the use of a V_P/V_S quotient masking method for detecting the geometry and orientation of anomalies characterized by high V_P to V_S ratios: Since the V_S value in the tunnel is close to zero, setting a threshold for the V_P/V_S quotient generates a spatial mask that highlights voids. A multiscale FWI is performed in the frequency intervals of 1- 20 Hz, 1 – 25 Hz, 1 – 30 Hz, 1 – 35 Hz and 1 – 40 Hz. The results show how FWI can detect the tunnel as a low-

velocity anomaly at the correct depth. The authors stated that, when using an initial model from SW analysis, an intermediate 2D FWI could perform an efficient model refinement before implementing a 3D FWI.

Xing and Mazzotti (2019a) proposed a two-grid genetic-algorithm based 2D FWI approach for predicting the V_S model from Rayleigh waves. In the two-grid parameterization, the forward modeling is performed on a fine grid while the inversion (for V_S and density) is performed on a coarser grid. The coarse-grid inversion allows for the reduction of the computing time required by a genetic-algorithm based FWI. The experiments on different synthetic data sets (belonging to models characterized by strong velocity contrasts, velocity inversion, strong lateral variations or irregular topography) proved an accurate prediction of the long-wavelength structures of the model obtained after FWI, which can be used as an initial model in a gradient-based Rayleigh wave FWI for further refinement.

Xing and Mazzotti (2019b) applied a two-grid genetic-algorithm Rayleigh wave FWI to two different field data sets. As the forward modeling scheme is based on a 2D FD wave propagation solver, 3D to 2D corrections were applied to the data before inversion. An envelope-based objective function and offset marching strategies were used during inversion. The method succeeded in guaranteeing an overall accurate matching between observed and inverted data, within half-wavelength, even if limiting the maximum frequency to 30 Hz. Only in one case, the low resolution did not allow identifying a velocity decrease trend (observed in the borehole logs). The authors stated that the models obtained from the coarse-grid global FWI are accurate enough to be used as initial models for local FWI.

Yuan et al. (2015) developed a wavelet-multiscale adjoint approach for elastic FWI of BWs and SWs. The inversion starts from SWs and minimizes, in the first stages, an envelope based objective function and, in the late stages, a waveform-difference one. The waveform-difference objective function is also used at the end of the sequence when inverting simultaneously for V_P, V_S and density. A multiscale approach is adopted through all the sequence. The authors proved the effectiveness of their method using a toy method and the synthetic Marmousi model. The authors also investigated the effects of an incorrect density model in the elastic FWI.

5.4 – The SEM46 code

The 3D FWI code SEM46 (*Spectral Element for Seismic Imaging in eXploration*) (Trinh et al., 2019) was used for all numerical tests in this study. The code is written in Fortran 90 and it has two levels of MPI parallelization: the domain decomposition and the simultaneous managing of multiple shots, mitigating the I/O disk resources.

Numerical implementation

The implementation relies on non-overlapping hexahedra spectral-element architecture, where a deformed Cartesian-based mesh is coupled with high-order shape functions, allowing to model complex topography variations (Trinh et al., 2019). Each hexahedral element can be mapped to a reference cube [-1,1]x[-1,1]x[-1,1] and projected to a 3D Gauss-Lobatto-Legendre (GLL) domain, discretized into a set of $(n+1)^3$ points, where n is the interpolation order (Trinh et al., 2019). An interpolation order equal to 4 or 5 allows to accurately model the viscoelastic wave propagation with approximately five or six GLL nodes per shortest wavelength (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999; Trinh et al., 2019). The GLL points are used to define $(n+1)^3$ basis functions, each function being a triple product of Lagrange polynomials of degree n.

A compressed boundary implementation is used for memory mitigation (Yang et al., 2016c; Trinh et al., 2019). Moreover, an efficient strategy for the stable recovery of the decimated incident field, needed for the construction of the misfit gradient, is implemented (Yang et al., 2016b; Trinh et al., 2019). Specifically, the viscoelastic wave equation (which is not self-adjoint), containing the memory variables, may introduce instability during the incident field reconstruction. To restore its reversibility, a viscoelastic version of the checkpointing strategy is implemented in the code (Anderson et al., 2012; Komatitsch et al., 2016; Yang et al 2016b). The optimal number of checkpoints (in terms of displacement, velocity and memory variables) is related to the total number of time steps nt by $log_2\ (nt)$ - Trinh et al. (2019). The checkpoint-assisted reverse-forward simulation (CARFS) is based on the strain energy measure and a value of tolerance (Yang et al., 2016b). The latter is related to the reference global energy, stored at the boundaries during the forward propagation of the incident filed, to quantify the stability of the reverse simulation (Yang et al 2016b; Trinh et al., 2019).

Viscoelastic forward modeling

The viscoelastic wave propagation is modeled using second-order partial differential equations – from Fichtner (2011) and Trinh et al. (2019), modified:

$$\rho(\mathbf{x})\frac{\partial^2}{\partial t^2}\mathrm{u_i}(\mathbf{x},t) - \nabla\cdot\sigma_{\mathrm{ij}}(\mathbf{x},t) = \mathrm{f_i}(\mathbf{x},t)\,, \tag{5.4}$$

$$\sigma_{ij}(\mathbf{x},t) = \mathrm{C}_{ijkl} * \varepsilon_{kl}(\mathbf{x},t) + \mathrm{T}_{ij}\,(\mathbf{x},t), \tag{5.5}$$

where C_{ijkl} are the unrelaxed coefficients of the stiffness (i.e. elastic) tensor and the tensor T is the possible stress failure.

The attenuation mechanism works under the approximation of constant quality factors (Q_P and Q_S), represented by standard-linear-solids (SLS – Robertsson et al., 1994; Blanch et al., 1995), each one effective over limited frequency bands (Yang et al., 2016a). The attenuation coefficients (Q_{ijkl}) are incorporated into the wave equation and a memory-variable vector is associated with each SLS (Trinh et al., 2019).

In the frequency domain, the inverse of the attenuation coefficients is described as the ratio between the imaginary and the real part of the complex relaxation rate $\hat{C}_{ijkl}(\mathbf{x}, \omega)$, where '^' indicates the Fourier transform (Trinh et al., 2019). Given n SLS systems chosen to fit a constant Q value over a certain frequency interval $\omega_{\nu} \in [\omega_{min}, \omega_{max}]$, the anelastic coefficients Y_{ν}^{ijkl} are introduced in the definition of the complex relaxation rate $\hat{C}_{ijkl}(\mathbf{x}, \omega)$ as follows (Emmerich and Korn, 1987; Blanch et al., 1995; Moczo and Kristek, 2005; van Driel and Nissen-Mayer, 2014; Trinh et al., 2019):

$$\hat{C}_{ijkl}(\mathbf{x}, \omega) = C_{ijkl}(\mathbf{x}) \left(1 - \sum_{\nu=1}^{n} Y_{\nu}^{ijkl}(\mathbf{x}) \frac{\omega_{\nu}}{\omega_{\nu} + i\omega}\right). \tag{5.6}$$

Following the approximation proposed by Yang et al. (2016a), regarding the use of $Y_{\nu}^{ijkl} \approx y_{\nu} Q_{ijkl}^{-1}(\mathbf{x})$ for the entire space, with y_{ν} dimensionless anelastic coefficients, the complex relaxation rate becomes (Trinh et al., 2019):

$$\hat{C}_{ijkl}(\mathbf{x}, \omega) = C_{ijkl}(\mathbf{x}) - C_{ijkl}(\mathbf{x})\, Q_{ijkl}^{-1}(\mathbf{x}) \sum_{\nu=1}^{n} y_{\nu} \frac{\omega_{\nu}}{\omega_{\nu} + i} + \Upsilon_{ij}(\mathbf{x}, t), \tag{5.7}$$

where the second term of the right-hand side describes the attenuation mechanism, $C_{ijkl}(\mathbf{x})\, Q_{ijkl}^{-1}(\mathbf{x})$ is a term-by-term product and the coefficients y_{ν} are estimated as a least-squares minimization problem (Trinh et al., 2019):

$$min_{y_{\nu}} \left\{ \int_{\omega_{min}}^{\omega_{max}} \left[Q_{ref}^{-1} \left(y_{\nu} \frac{\omega_{\nu}\, \omega}{\omega^2 + \omega_{\nu}^2} - 1 \right) \right]^2 \right\}, \tag{5.8}$$

which minimizes the distance between a constant value Q_{ref}^{-1} and its approximation over a constant frequency band $[\omega_{min}, \omega_{max}]$, with $Q_{ref}\ [min_{ijkl.x}(Q_{ijkl}(\mathbf{x}), max_{ijkl.x}(Q_{ijkl}(\mathbf{x}))]$ (Trinh et al., 2019). This method is associated with a cheap-memory storage strategy: instead of storing n x 21 anelastic coefficients $Y_{\nu}^{ijkl}(\mathbf{x})$ associated with 21 coefficients $C_{ijkl}(\mathbf{x})$ at each space location (Komatitsch and Tromp, 1999), only n scalars y_{ν} and n reference frequencies ω_{ν} (associated with n SLSs) are stored for the entire medium (Trinh et al., 2019).

The unrelaxed stiffness coefficients $C_{ijkl}(\mathbf{x})$ are representative for the elastic rheology. Attenuative stiffness coefficients C_{ijkl}^{R} are introduced to describe the frequency-dependent attenuation mechanism (Trinh et al., 2019):

$$C_{ijkl}^{R}(\mathbf{x}) = C_{ijkl}(\mathbf{x})\, Q_{ijkl}^{-1}(\mathbf{x}). \tag{5.9}$$

The relation for the stress tensor becomes (Trinh et al., 2019):

$$\sigma_{ij}(\mathbf{x}, t) = C_{ijkl}(\mathbf{x}) * \varepsilon_{kl}(\mathbf{x}, t) - C_{ijkl}^{R}(\mathbf{x}) \sum_{\nu=1}^{n} \psi_{\nu;\, k,l}(\mathbf{x}, t) + \Upsilon_{ij}(\mathbf{x}, t)\,,$$

where $\psi_{\nu;\, k,l}(\mathbf{x}, t)$ is the memory variable (associated with each SLS) satisfying the equation (Trinh et al., 2019):

$$\frac{\partial}{\partial t}\,\psi_{v;\,k,l}(\mathbf{x},t)+\omega_v\psi_{v;\,k,l}(\mathbf{x},t)=\omega_v y_v\varepsilon_{kl}(\mathbf{x},t). \tag{5.10}$$

The introduction of the memory variable into the stress field expression helps to optimize the computation cost.

When the attenuation mechanism is isotropic, the relaxed stiffness tensor becomes (e.g. Moczo et al., 1997; Trinh et al., 2019):

$$\mathrm{C}^R=\begin{bmatrix}\lambda^R+2\mu^R & \lambda^R & \lambda^R & 0 & 0 & 0\\ \lambda^R & \lambda^R+2\mu^R & \lambda^R & 0 & 0 & 0\\ \lambda^R & \lambda^R & \lambda^R+2\mu^R & 0 & 0 & 0\\ 0 & 0 & 0 & \mu^R & 0 & 0\\ 0 & 0 & 0 & 0 & \mu^R & 0\\ 0 & 0 & 0 & 0 & 0 & \mu^R\end{bmatrix}, \tag{5.11}$$

where the relaxed (attenuative) coefficients λ^R and μ^R can be estimated as (Trinh et al., 2019):

$$\lambda^R+2\mu^R=\frac{1}{3}\,{Q_\mathrm{P}}^{-1}\,(C_{11}+\,C_{22}+\,C_{33})\text{ and }\,\mu^R=\,\frac{1}{3}\,{Q_\mathrm{S}}^{-1}\,(C_{44}+\,C_{55}+\,C_{66}). \tag{5.12}$$

Discretization in space

Considering the SEM implementation based on Lagrange polynomials as basis functions and GLL quadrature for numerical integration, the weak formulation of the viscoelastic wave equation for the displacement field is (Trinh et al., 2019):

$$\mathrm{M}\,\frac{\partial^2}{\partial t^2}\mathrm{u}+\mathrm{Ku}-\,D^wC^R\sum_{v=1}^{n}\psi_v=\mathrm{F},$$

$$\frac{\partial}{\partial t}\psi_v+\,\omega_v\psi_v=\,\omega_v y_v\,D\mathrm{u}, \tag{5.13}$$

where M is the global mass matrix (diagonal), K is the global stiffness matrix, F is the external source, D is the spatial derivative operator of a vector in the Cartesian space and D^w is the spatial derivative operator weighted by GLL weights (Trinh et al., 2019).

As the product of the displacement field vector u by the stiffness matrix K is computationally intensive, it is factorized as K=D^wCD (Trinh et al., 2019), which is a three-step factorization: 1) the strain component computation via estimation of the displacement field's spatial derivatives, 2) the stress component computation by the product with the stiffness coefficients and 3) the last estimation of the spatial derivatives following the implementation proposed by Deville et al. (2002), which takes benefit from the tensorial properties of hexahedral elements and the efficient loop vectorization combined with manual unrolling (Trinh et al., 2019).

Discretization in time

A second-order explicit Newmark scheme is implemented for time integration, to compute the displacement field at each time step (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999; Trinh et al., 2019). The first-order ordinary differential equation related to the memory variable (equation 5.13) is efficiently incorporated into the time scheme. The memory variables are updated at half time steps compared to the displacement fields, to maintain the second-order accuracy (Trinh et al., 2019):

$$\psi_{\nu}^{it+1/2} = \frac{2-\omega_{\nu}\Delta t}{2+\omega_{\nu}\Delta} \psi_{\nu}^{it-1/2} + \omega_{\nu} y_{\nu} \frac{2\Delta t}{2+\omega_{\nu}\Delta t} \mathrm{Du}^{it}, \quad \text{for } \nu = 1,\ldots,n. \tag{5.14}$$

The expression above allows computing the memory variable at each step explicitly, from the memory variable at the previous step and the displacement field (Trinh et al., 2019).

***

The memory requirements of the viscoelastic simulation are efficiently reduced by the above strategy (storing only a single ψ_{ν} per step) and by storing only *n* scalar anelastic coefficients y_{ν} (equations 5.7 and 5.8) related to the SLS attenuation mechanism (Trinh et al., 2019). For each checkpoint needed for the incident field reconstruction, the displacement u^{it}, velocity v^{it} and memory variable $\psi^{it-\frac{1}{2}}$ fields are stored. During the forward propagation of the incident field, a reference global energy is also recorded at each time step E_{ref}^{it}=1/2 $(\sigma^{it}\varepsilon^{it})_{\Omega}$, to monitor the stability of the incident wavefield's reverse simulation, according to a certain energy tolerance (Trinh et al., 2019).

Nonlinear optimization

The linearized inversion relies on various local optimization methods, gathered by the Seiscope optimization toolbox (Métivier and Brossier, 2016), which is embedded in the SEM46 code through a reverse-communication interface. All Newton-like strategies are based on the same line-search algorithm, that satisfies the Wolfe conditions (Nocedal and Wright, 2006; Trinh et al., 2019).

The gradient of the least-squares misfit function $\chi(\mathbf{m}) = 1/2\|d_{obs} - d_{calc}\|^2$ with respect to the model parameters $\mathbf{m}$ (e.g. elastic tensor, density, attenuation factor, etc.) is computed in time-domain using the adjoint state approach (Plessix, 2006). For example, the gradient with respect to the elastic tensor's coefficients C_{ij} is the zero-lag cross-correlation between the incident and the adjoint wavefield (Plessix, 2006; Vigh et al., 2014):

$$\frac{\partial\chi(\mathrm{m})}{\partial \mathrm{C}_{ij}} = \left(\bar{\varepsilon}, \frac{\partial \mathrm{C}}{\partial \mathrm{C}_{ij}}\varepsilon\right)_{\Omega,t} \tag{5.15}$$

where C_{ij} is a matrix (6 x 6) containing the coefficients of the elastic tensor (i.e. the unrelaxed stiffness components), $\bar{\varepsilon}$ is the adjoint strain field and ε is the incident strain field. In viscoelastic approximation, a memory variable term ψ_s is also present in the gradient (Fichther and van Driel, 2014; Trinh et al., 2019):

$$\frac{\partial \chi(\mathrm{m})}{\partial \mathrm{C}_{ij}} = \left(\bar{\varepsilon}, \frac{\partial \mathrm{C}}{\partial \mathrm{C}_{ij}} \varepsilon\right)_{\Omega,t} - \left(\bar{\varepsilon}, \sum_{v=1}^{n} \frac{\partial \mathrm{C}^R}{\partial \mathrm{C}_{ij}} \psi_v\right)_{\Omega,t}, \tag{5.16}$$

where C^R is the relaxed stiffness tensor.

The gradient for the density is the always-computed misfit gradient during one iteration of the FWI workflow. Its computation is done through the zero-lag cross-correlation between the adjoint displacement field and the incident acceleration field (Trinh et al., 2019):

$$\frac{\partial \chi(\mathbf{m})}{\partial \rho} = \left(\bar{\mathrm{u}}, \frac{\partial^2}{\partial t^2} \mathrm{u}\right)_{\Omega,t}. \tag{5.17}$$

The gradient of the attenuation parameters can be estimated as the zero-lag cross-correlation between the adjoint strain field $\bar{\varepsilon}$ and the incident memory-variable field ψ_v (Trinh et al., 2019):

$$\frac{\partial \chi(\mathbf{m})}{\partial Q^{-1}} = -\left(\bar{\varepsilon}, \sum_{v=1}^{n} \frac{\partial C^R}{\partial Q^{-1}} \psi_v\right)_{\Omega,t}.$$

The gradient with respect to any desirable parameter α (e.g. seismic velocity, anisotropic parameter, impedance, etc.) can be retrieved by the chain rule, using the density ρ, the stiffness coefficient C_{ij} and the attenuation parameters Q_P^{-1}, Q_S^{-1} (Trinh et al., 2019):

$$\frac{\partial \chi(\mathbf{m})}{\partial \alpha} = \sum_{i=1}^{6} \sum_{i=1}^{6} \frac{\partial \chi}{\partial \mathrm{C}_{ij}} \frac{\partial \mathrm{C}_{ij}}{\partial \alpha} + \frac{\partial \chi}{\partial \rho} \frac{\partial \rho}{\partial \alpha} + \frac{\partial \chi}{\partial Q_P^{-1}} \frac{\partial Q_P^{-1}}{\partial \alpha} + \frac{\partial \chi}{\partial Q_S^{-1}} \frac{\partial Q_S^{-1}}{\partial \alpha}. \tag{5.18}$$

Gradient smoothing

A gradient smoothing through a 3D anisotropic Bessel filter is efficiently performed on the SEM modeling mesh (Trinh et al., 2017). This filter has an arbitrary local orientation using 3 Euler angles and spatially variable coherent lengths along these 3 specific directions (*z*, *x* and *y*).

The gradient smoothing is implemented through the application of an inverse operator. Instead of obtaining the smoothed gradient by an explicit convolution of the gradient vector *g* with the coefficients of the Bessel filter B_{3D} (Trinh et al., 2017):

$$\mathrm{sg}(z,x,y) \approx B_{3D}\,(z,x,y) * \mathrm{g}(z,x,y), \tag{5.19}$$

where sg is the smoothed gradient, g is the raw gradient vector and the symbol "*" is the convolution operator, the Bessel filter acts as an inverse operator ${B_{3D}}^{-1}$ (Trinh et al., 2017):

$${B_{3D}}^{-1}\,(z,x,y) * \mathrm{sg}(z,x,y) = \mathrm{g}(z,x,y). \tag{5.20}$$

Under the assumption of small spatial variation for the filter's parameters (aimed to preserve the self-adjoint property), their spatial derivatives can be neglected, and the coefficients of the Bessel filter can be obtained as the unique solution of a Helmholtz-like PDE (Trinh et al., 2017):

$$\left[1 - \left(Lz^2 \frac{\partial^2}{\partial z^2} + Lx^2 \frac{\partial^2}{\partial x^2} + Ly^2 \frac{\partial^2}{\partial y^2}\right)\right] \mathrm{sg}(z,x,y) = \mathrm{g}(z,x,y), \tag{5.21}$$

where *Lz*, *Lx* and *Ly* are the coherent lengths in the *z*, *x* and *y* directions. The above system is discretized using the FE method and expressed as a sparse operator (Thinh et al., 2017):

$$[1 - (\nabla_{z,x,y}{}^{t}\, \mathrm{P}(z,x,y)\, \mathrm{P}(z,x,y)^{t}\, \nabla_{z,x,y})]\, \mathrm{sg}(z,x,y) = \mathrm{g}(z,x,y) \tag{5.22}$$

where ∇ gathers the spatial derivatives ($\delta/\delta x$, $\delta/\delta y$, $\delta/\delta z$), t stands for the transposed operator and P is a matrix, containing the filter's parameters and representing the projection between the Cartesian space and the locally rotated coordinate system (Thinh et al., 2017):

$$\mathrm{P}(z,x,y) = \begin{bmatrix} Lv\cos\varphi & Lu\sin\varphi & 0 \\ -Lv\cos\theta\sin\varphi & Lu\cos\theta\cos\varphi & Lw\sin\theta \\ Lv\sin\theta\sin\varphi & -Lu\sin\theta\cos\varphi & Lw\cos\theta \end{bmatrix} \tag{5.23}$$

The linear system $(\mathrm{M_b} + \mathrm{K_b})\,\mathrm{sg} = \mathrm{M_b}\,\mathrm{g}$ (with $\mathrm{M_b}$= the mass matrix - diagonal and $\mathrm{K_b}$ stiffness matrix - symmetric) associated with the Bessel filter and related to the PDE's discretization through the weak formulation of SEM, is solved using a parallel conjugate-gradient iterative solver (Trinh et al., 2017). Since the product $\mathrm{K_b}$ sg is computationally demanding, the sparse-stiffness matrix $\mathrm{K_b}$ is factorized as $\mathrm{K_b}=\mathrm{D^w}\,(\mathrm{PP^t})\,\mathrm{D}$ (Deville et al., 2002; Trinh et al., 2019).

The code allows performing multiple sequential gradient smoothing runs. For example, a double application of the Bessel filter proved to be an accurate approximation of a Laplace filter (Trinh et al., 2017).

A more detailed presentation of the SEM46 numerical implementation and computation performance can be found in Trinh et al. (2017) and Trinh et al. (2019).

5.5 – Highlights

- Based on the numerical solution of the seismic wave equation and exploiting the entire information in the seismogram, FWI guarantees a theoretical resolution up to half of the propagated wavelength. However, such achievement is not guaranteed for real data applications, due to resolution-restrictive factors, such as the partial illumination of the subsoil (related to the acquisition design) and the trade-off between different parameters.
- FWI is a mathematically ill-posed problem (many models can fit equally-well the same data set), reason for which different strategies are required to guide and constrain the convergence.
- The first FWI applications were based on the acoustic approximation, but the continuous evolution of parallel computing architectures renders the elastic approximation increasingly feasible and efficient. This opens a gate for successful FWI applications to land data recorded in environments characterized by complex geological assets.

- A probabilistic approach for the inversion efficiently accounts for the solution non-uniqueness while providing information on the uncertainties. Nevertheless, for 3D applications using the elastic or viscoelastic equation, the computing architectures nowadays still cannot handle conveniently such implementation. Thus, the deterministic approach is adopted for most of the 3D FWI tools. This fact renders the initial model's accuracy very important for FWI's success.
- There are various gradient methods, such as steepest descent, conjugate gradients, Newton and Gauss-Newton. The adjoint state technique, based on the gradient computation through time-domain cross-correlation between the incident and residual wavefield, is widely implemented in many FWI codes, as the one used in this work.
- Some sensible aspects of the adjoint-state technique are the necessity of storing the wavefield at the boundaries and of satisfying the condition of wave energy conservation (to guarantee the reversibility of the wave equation). Different strategies (e.g. check-pointing) are implemented for bypassing these issues.
- The gradient of the misfit function contains many local minima. To avoid the algorithm's convergence towards one of them, different strategies are required, like the use of alternative misfit functions (to the classical one based on L_2 norm), data-gradient preconditioning, introduction of weighting operators, regularizations, multiscale inversion, time windowing, etc.
- Although FWI was designed for deep exploration purposes, various examples are available in the literature for FWI application on shallow seismic wavefields. Most of them use 2D solvers for wave propagation simulation, but there are also some successful examples of FWI based on the use of 3D wave propagation equation.
- The FWI code used for this study (SEM46) is based on 3D hexahedra-spectral elements architecture and is characterized by a double level of MPI parallelization: domain decomposition and parallel managing of multiple shots.
  - ✓ The code relies on viscoelastic forward modeling and elastic inversion. The attenuation is implemented as a superposition of SLS, related to Q factors, which are constant over limited frequency bands and are associated with specific memory variables.
  - ✓ A compressed boundary implementation is used for memory mitigation while an efficient checkpointing strategy is adopted for the stable recovery of the incident wavefield.
  - ✓ A reverse communication interface embeds different optimization tools, based on linearized inversion methods.
  - ✓ A gradient smoothing by a 3D anisotropic and non-stationary Bessel filter is efficiently performed on the SEM modeling mesh, under the assumption of small spatial variation of filter's parameters (i.e. the coherent lengths in the z-x- and y-direction).

Appendix 2

Source estimation

The estimation of the source wavelet can be obtained as a solution of a linear inverse problem. This fact is possible because the relationship between the seismic wavefield and the source is linear (Virieux and Operto, 2009).

The source signature estimation requires the Green's function computation in the initial model (m_o). For this operation, a known wavelet is used, with the same frequency content as the one configured for the inversion.

The source signature is recovered as a matching Wiener filter that minimizes the quadratic difference between the observed seismic data and the synthetic ones $E = \frac{1}{2}(d_{obs}-d_{calc})^t(d_{obs}-d_{calc})$. There are two main assumptions in this process: the source signature is the only unknown and the initial velocity model (m_o) is relatively correct (Pratt, 1999).

In the frame of the optimum Wiener filter, the field data d_{obs} is the desired output, the simulated data d_{calc} (m_0) is the output obtained by propagating the known Richer source over the initial model (m_0) and w is the inverse operator that minimizes the quadratic difference between the two datasets.

Considering the seismic traces as the result of the convolution between the Green's function and the source wavelet, the coefficients of the unknown filter are obtained by solving the following inverse linear problem:

$$\chi = \sum_{n=1}^{traces\ number}[d_{obs} - w * d_{calc}]^2$$

$$\frac{\partial \chi}{\partial w} = 2\,[d_{obs} - w * d_{calc}]\frac{\partial}{\partial w}\,[d_{obs} - w * d_{calc}] = 0,$$

$$\frac{\partial \chi}{\partial w} = 2\,[d_{obs} - w * d_{calc}]\,(-\,d_{calc}) = 0,$$

$$d_{obs} * d_{calc} = w * d_{calc} * d_{calc}, \qquad \text{(A2.1)}$$

where "*" indicates a convolution operator in the time domain (corresponding to a product in the frequency domain). The final expression for the source signature estimation in the frequency domain is (Pratt, 1999):

$$w(f) = \sum_{n=1}^{nb.\ of\ traces} \frac{d_{obs}(f)\,.\,d_{calc}(f)}{d_{calc}(f)\,.\,d_{calc}(f)} \qquad \text{(A2.2)}$$

The above expression is a complex-valued one, providing both the amplitude and phase of the desired source signature in the frequency domain (Pratt, 1999).

After Fourier frequency-time transform, the higher-time oscillations of the source function are filtered. This operation aims at honoring the deconvolution's assumption related to the stationary character of the wavelet over the propagation path. Besides, a minimum phase source is the one that guarantees high resolution.

The source estimation can be iteratively improved in terms of amplitude and phase matching, following the model updated. Nevertheless, phase and amplitude discrepancies may occur between the estimated wavelet and the real (desired) one (Pratt, 1999). For example, if the real wavelet is not of minimum phase, the optimum Wiener filter is not causal, leading to an estimated wavelet that is more advanced in phase than the minimum-phase desired output. To mitigate this effect, phase correction may be applied to the estimated source (Pratt, 1999).

Moreover, the amplitude of the estimated source may be higher than that of the real source due to geometrical dispersion and mode conversion to S-waves, which affects the real amplitudes (Pratt, 1999). The matching between observed and synthetic amplitudes can be forced, for example, by reducing the value of the Q parameter in the synthetic simulation, which is equivalent of increasing the intrinsic attenuation (Pratt, 1999).

PART II
Experiments

Chapter 6

Data sets

Contents

6.1– Introduction

This chapter is dedicated to the analysis of various data sets, obtained from numerical simulations or belonging to a seismic acquisition conducted in an experimental site. All numerical simulations have been performed using the SEM46 code, based on a 3D wave propagation solver. Some of these results are published in Teodor et al. (2017), Khosro Anjom et al. (2019) and Teodor et al. (2020 - submitted).

Chapter 6 contains three main sections:

- The first one (6.2) is dedicated to the analysis of various shot gathers, obtained from numerical simulations performed over different synthetic models characterized by various geometries. The aim of this section is providing some examples related to the wave propagation behavior in a shallow geological context, as this study is focused on geotechnical scale applications. The dimensions of these synthetic models are similar to the ones of the test site.
- The second one (6.3) is dedicated to the presentation of a seismic acquisition carried out on purpose for this study, as well as the analysis of the corresponding field data set.
- Tre third one (6.4) is dedicated to the presentation of a synthetic model (that mimics the characteristics of the real test site) and the connected data set used in this study to validate the adopted methodology.

6.2– Numerical simulations over synthetic models

Data related to near-surface land acquisitions are dominated by SWs, showing high amplitude and dispersive behavior. These peculiar features render the accurate reconstruction of shallow targets a challenging task, with an important impact on the imaging accuracy at greater depths. Therefore, understanding the physics of wave propagation in shallow environments (geotechnical scale) requires devoted attention.

The aim of this section is illustrating the above behavior through some synthetic examples. The results of a numerical simulation conducted over a synthetic model, having high lateral and vertical impedance contrasts, are schematically shown at the beginning of the Chapter, with the aim of providing a global picture of the wave propagation behavior in this kind of environment. Successively, other examples, with simple-to-complex increasing geometry, are presented. In the examples presented below, all the models are characterized by flat topography, as the real geological context investigated in this study. It is known that a flat topography facilitates a forward scattering regime of SW propagation with the consequent fundamental mode's domination.

Table 1 shows a summary of the numerical simulations performed in this section and the related parameters. A particularly high frequency was used for the source during the first simulation, while realistic frequency values were used for the other simulations. The choice of a particularly high frequency for the first simulation was aimed to guarantee a

relatively high number of propagating wavelengths over the maximum offset and allow distinguishing the separation of various arrivals in the simulated data.

Synthetic models	Central frequency of the source [Hz]	Time sampling [s]	Mesh dimensions (m)
Simulation 1: Multilayered model with a low-velocity target embedded	200	4.8e-6 s	0.1 m
Simulation 2: Homogeneous medium	48	1.4e-5 s	0.3 m
Simulation 3: Medium composed of two layers	48	1.4e-5 s	0.3 m
Simulation 4: Homogeneous medium with a low-velocity target embedded	48	1.4e-5 s	0.3 m
Simulation 5: Medium composed of two layers and a low-velocity target embedded	48	1.4e-5 s	0.3 m
Simulation 6: Multilayered medium with a low-velocity target embedded	48	1.4e-5 s	0.3 m

Table 1 – List of the numerical simulations of Section 6.2 and corresponding parameters

6.2.1 – A quick look at the "global picture"

Figure **6.1** summarizes the parameters and the results of the wave propagation simulation conducted over a multilayered model, containing a low-velocity target of a particular shape (Figure **6.1a**) – *Simulation 1*. The source time function is a Ricker wavelet, with a central frequency of 200 Hz. A 0.1 m element size (in *z*-, *x*- and *y*-direction) was used for the 3D mesh, according to the condition of including at least 5 GLL points per shortest wavelength. To honor the CFL (Courant-Friedrichs-Lewy) time stability condition, the time sampling was 4.8e-06 s. A free surface condition was applied at the top of the model while absorbing sponges were used for the other boundaries.

The analysis of a shot placed in a transition area between the target and the vertically layered medium (Figure **6.1b**) reveals the SWs domination and the weak BWs amplitude. Various forward-scattered and back-scattered phases are also present in correspondence of the target's boundaries (Figure **6.1c**).

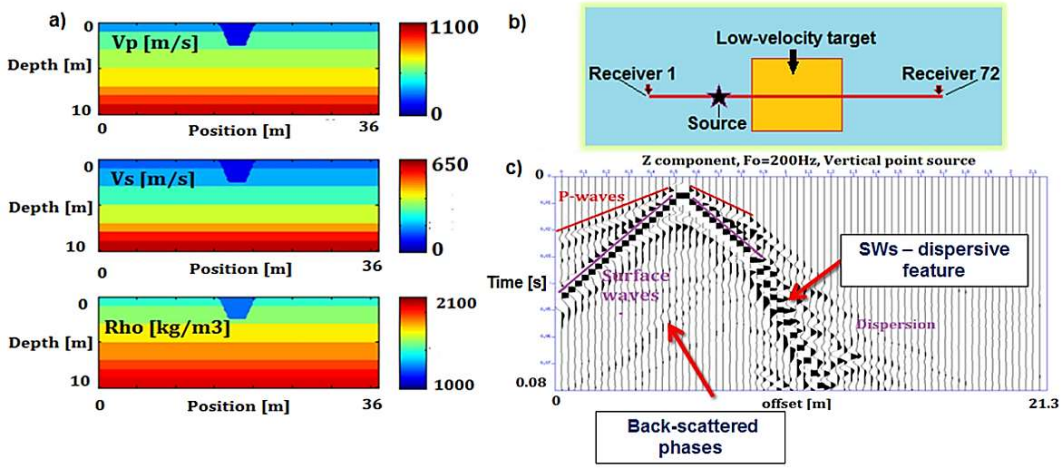

*Figure **6.1** – Example of seismic wave propagation in a shallow synthetic model. **a**) 2D vertical section of the 3D P-wave velocity, S-wave velocity and density model. **b**) Sketch of the acquisition geometry used for the synthetic experiment. **c**) The vertical component of the recorded data.*

6.2.2 – From simple to complex

Numerical simulations were performed over different 3D models. The acquisition design gathers 72 vertical receivers that cross the models center. A Ricker wavelet, with a central frequency of 48 Hz (characteristic for geotechnical scale exploration) was the vertical point source. The source was located towards the left extremity of the seismic line, outside the low-velocity target (Figure 6.1b). A 0.3 m element size (in *z*-, *x*- and *y*-direction) was used for 3D mesh. To honor the CFL time stability condition, the time sampling was 1.4e-05 s. The maximum duration of the simulations was 0.27 s.

Simulation 2: Homogeneous medium

This model has homogeneous properties (Figure **6.2**), with a V_P of 1050 m/s and a V_S of 600 m/s. The density (kg/m^3) was retrieved using the relation $\rho = 0.5 \cdot V_S^{0.22} \cdot 10^3$, characteristic for shallow environments.

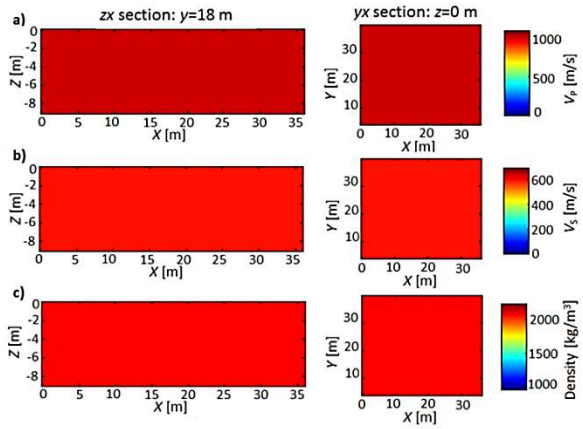

*Figure **6.2** – 2D sections of the 3D model of the Simulation 2: **a**) P-wave velocity, **b**) S-wave velocity, **c**) density. Vertical sections (zx) are shown on the left-hand side, while horizontal sections (yx), in correspondence of the surface, are shown on the right-hand side.*

Analyzing the common shot gather obtained from this simulation (Figure **6.3**), one can notice dominant SWs as the main seismic phase. Since the medium is homogeneous, no dispersive behavior occurs.

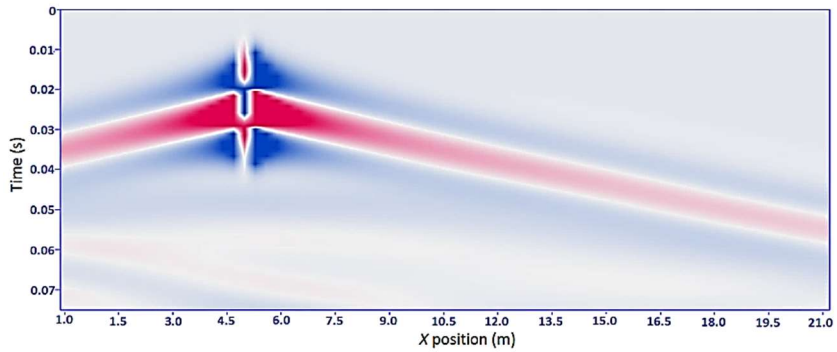

***Figure 6.3** – Shot gather related to wave propagation through the model in Figure 6.2.*

Simulation 3: Medium composed of two layers

This model (Figure **6.4**) is composed of two layers, characterized by the velocity values shown in Table 2. For the density model computation, the same relation of the previous case was used.

Model of the simulation 3	V_P *[m/s]*	V_S *[m/s]*	ρ *[kg/m^3]*
Layer 1	250	150	1505
Layer 2	1050	600	2043

Table 2. Characteristics of Model 2.

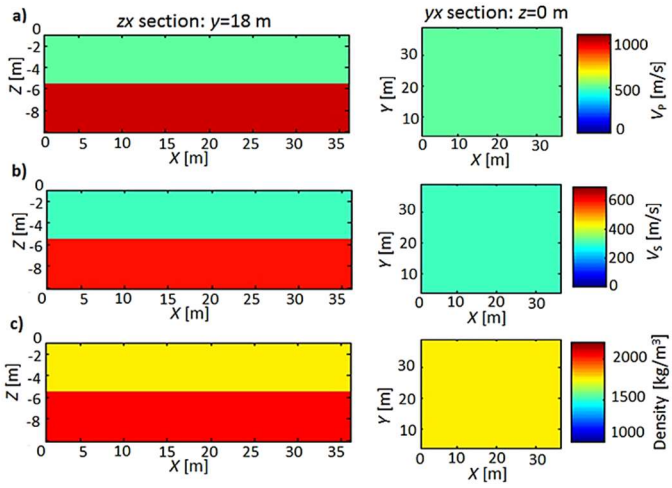

*Figure **6.4** – 2D sections of the 3D model of the simulation 3: **a)** P-wave velocity, **b)** S-wave velocity, **c)** density. Vertical sections (zx) are shown on the left-hand side, while horizontal sections (yx), in correspondence of the surface, are shown on the right-hand side.*

In the synthetic data obtained from the simulation (Figure **6.5**), SWs are dominant, while BWs are very weak (almost indistinguishable). A weak reflected phase, corresponding to the interface between the two layers, can be noticed as well at about 0.075 s.

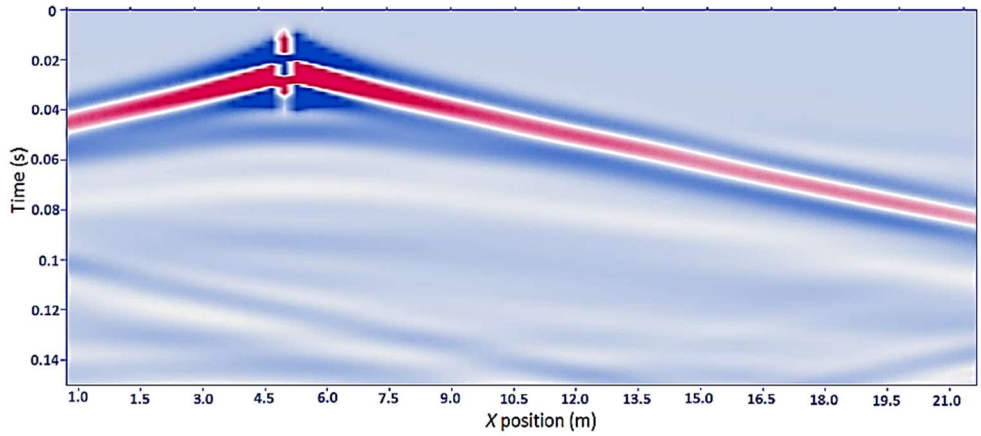

***Figure 6.5** – Shot gather related to wave propagation through the model in Figure 6.4.*

Simulation 4: Homogeneous medium with a low-velocity target embedded

The model (Figure **6.6**) is characterized by a low-velocity target (V_P=120 m/s, V_S=70 m/s) embedded in a homogeneous background (V_P=1050 m/s; a V_S=600 m/s).

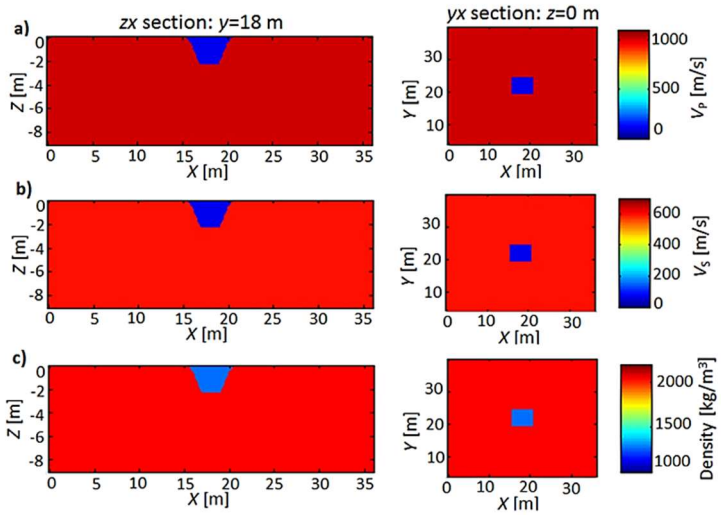

*Figure **6.6** – 2D sections of the 3D model of the simulation 4: **a)** P-wave velocity, **b)** S-wave velocity, **c)** density. Vertical sections (zx) are shown on the left-hand side, while horizontal sections (yx), in correspondence of the surface, are shown on the right-hand side.*

As in the previous cases, the seismic data are dominated by SWs, while the BWs amplitude is very weak (Figure **6.7**). Continuous phases confined inside the low-velocity target can be noticed, while the laterally leaked phases, in correspondence of the target's boundaries, are barely distinguishable: Almost all seismic energy remains trapped inside the low-velocity target (the position from 8 m to 13 m in the seismogram).

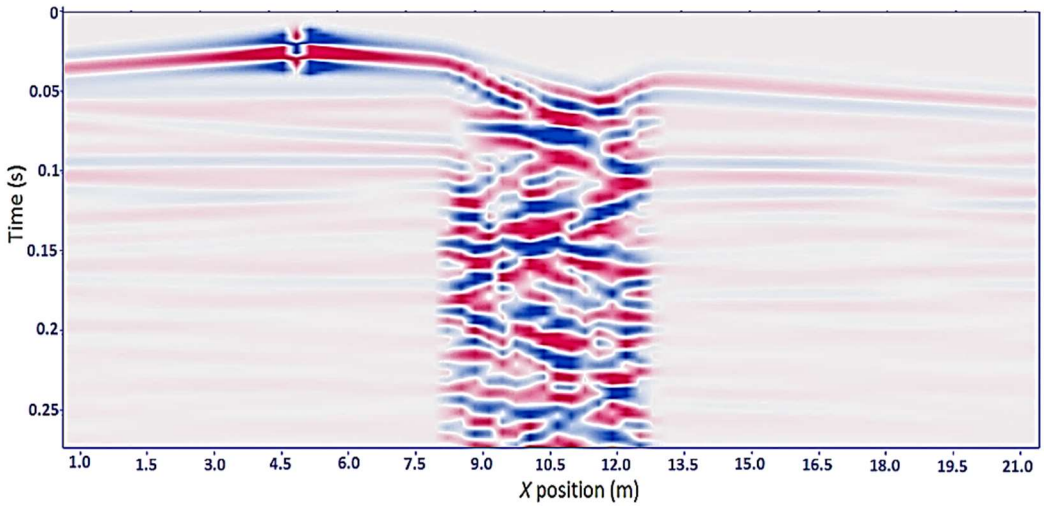

***Figure 6.7** – Shot gather related to wave propagation through the model in Figure 6.6.*

Simulation 5: Medium composed of two layers and a low-velocity target embedded

This model (Figure **6.8**) is characterized by the same values of density and velocity as the model of the *Simulation 3*, while the embedded low-velocity target has the same characteristics of the model used in the *Simulation 4*.

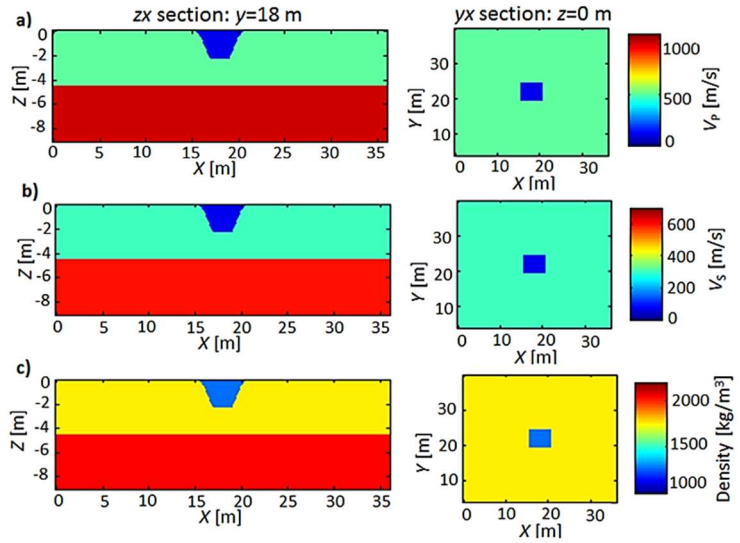

*Figure **6.8** – 2D sections of the 3D model of the simulation 5: **a**) P-wave velocity, **b**) S-wave velocity, **c**) density. Vertical sections (zx) are shown on the left-hand side, while horizontal sections (yx), in correspondence of the surface, are shown on the right-hand side.*

One can notice dominant SWs, with some dispersive features (Figure **6.9**). Again, BWs are very weak. Apart from the phases trapped inside the sand body, some forward and backward scattered phases can be noticed in correspondence of the target's boundaries. The lateral leakage outside the target (offset from 8 m to 13 m in the seismogram) is higher than in the previous case.

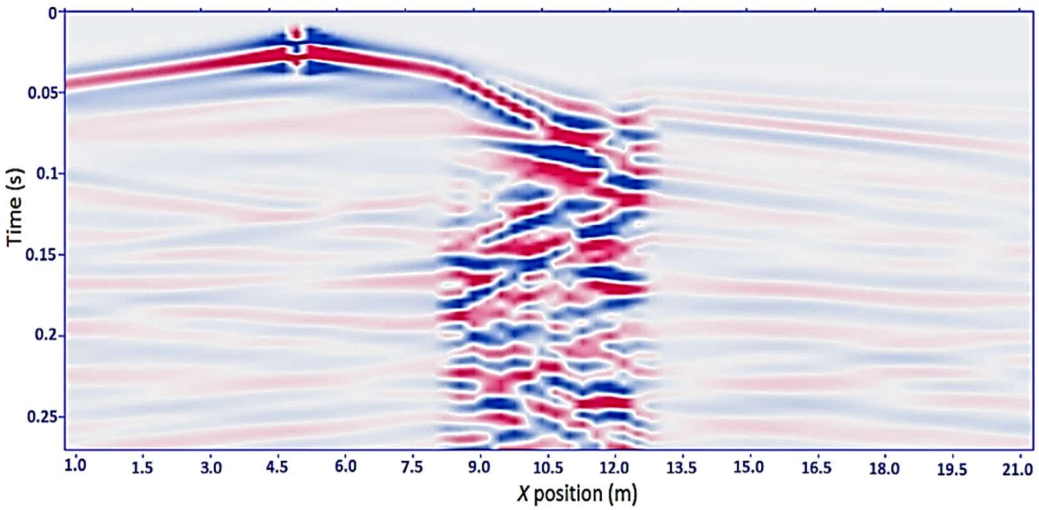

***Figure 6.9** – Shot gather related to wave propagation through the model in Figure 6.8.*

Simulation 6: Multilayered medium with a low-velocity target embedded

The object of this simulation is a multi-layered model bearing a low-velocity target (V_P=120 m/s, V_S=70 m/s) - Figure **6.10**. This model is similar to the one presented in Section 6.2.1, but here the frequency content of the source is much lower. The various layers are characterized by the velocity values reported in Table 3, while the density follows the path described by the relation used in the previous cases.

Model of the simulation 6	*V_P [m/s]*	*V_S [m/s]*	*ρ [kg/m^3]*
Layer 1	300	150	1505
Layer 2	500	200	1604
Layer 3	600	300	1754
Layer 4	700	400	1868
Layer 5	800	500	1962
Layer 6	900	600	2043
Layer 7	1050	650	2079

Table 3. Characteristics of Model 5.

The seismic data belonging to this experiment (Figure **6.11**) contain slower and weaker SWs. Moreover, the phases trapped inside the target decrease, while the leakage outside the target's boundaries increases. The scattered phases are less energetic compared with the previous case (Figure **6.9**) since additional energy dissipates in correspondence of the various interfaces between the layers. Some boundary artefacts are also present, probably due to the use of a too-small number of absorbing sponges (4).

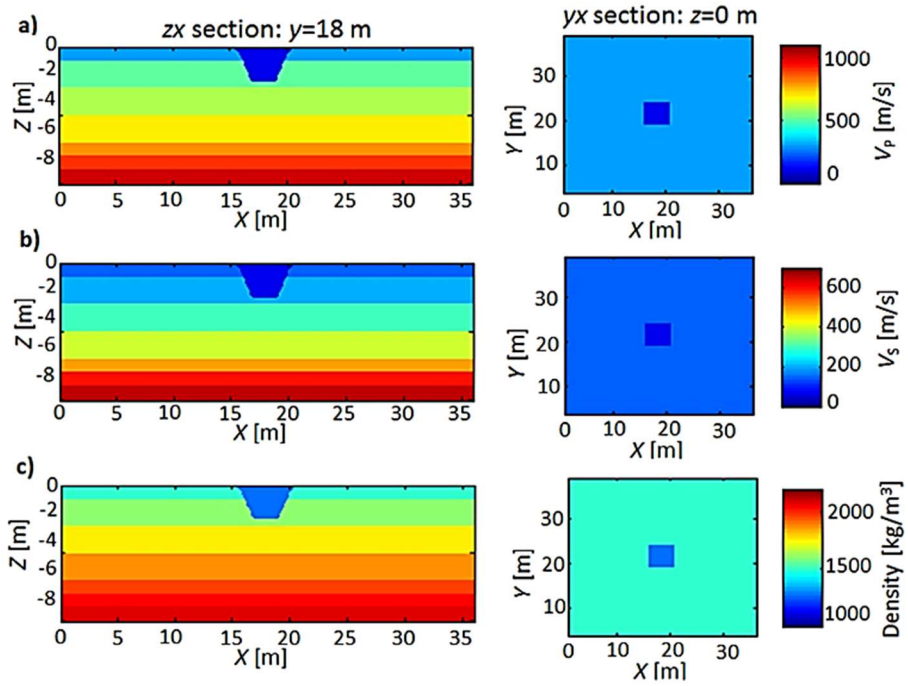

*Figure **6.10** – 2D sections of the 3D model of the simulation 6: **a**) P-wave velocity, **b**) S-wave velocity, **c**) density. Vertical sections (zx) are shown on the left-hand side, while horizontal sections (yx), in correspondence of the surface, are shown on the right-hand side.*

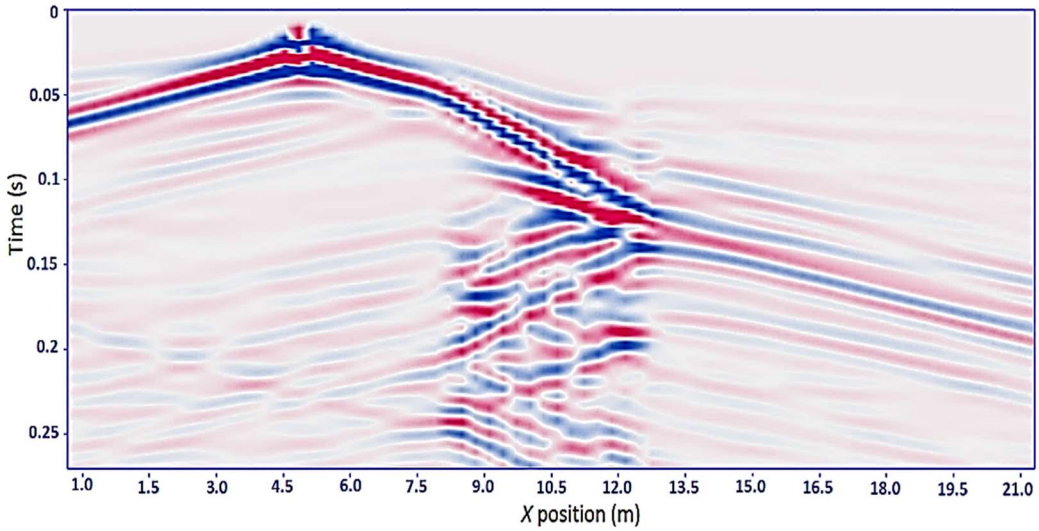

***Figure 6.11** – Shot gather related to wave propagation through the model in Figure 6.10.*

6.3 – Seismic acquisition and real data set

The real data set analysed in this study belongs to a dedicated seismic acquisition carried out in the experimental test site of CNR (National Research Council), located in Turin-Italy. The site is characterized by an on purpose buried loose-sand body (Figure **6.12a**), surrounded by more compact sediments (gravelly-sands and gravels). The maximum width of the sand target is 5 m x 5 m in the shallower part, and 2 m x 2 m in correspondence of the maximum depth (2.5 m).

Two acquisitions have been performed following a 2D and a 3D pattern, respectively. For both acquisitions, an 8 kg sledgehammer was the adopted vertical point source, generating a signal with a central frequency of about 60 Hz. The S/N has been improved by stacking from 8 to 10 shots for each source-point. The time sampling was set to 0.125 ms, for an acquisition duration of 0.512 s (with a pre-trig of 0.1 s).

For the 2D acquisition, 11 seismograms were recorded along the seismic line R1 – R72 (Figure **6.12b**), which gathers 72 vertical receivers (4.5 Hz) evenly spaced every 0.3 m (Figure **6.12c**), connected to 3 Geode - seismic modules. The maximum offset for the 2D acquisition (S 1 – R 72) is 25.3 m.

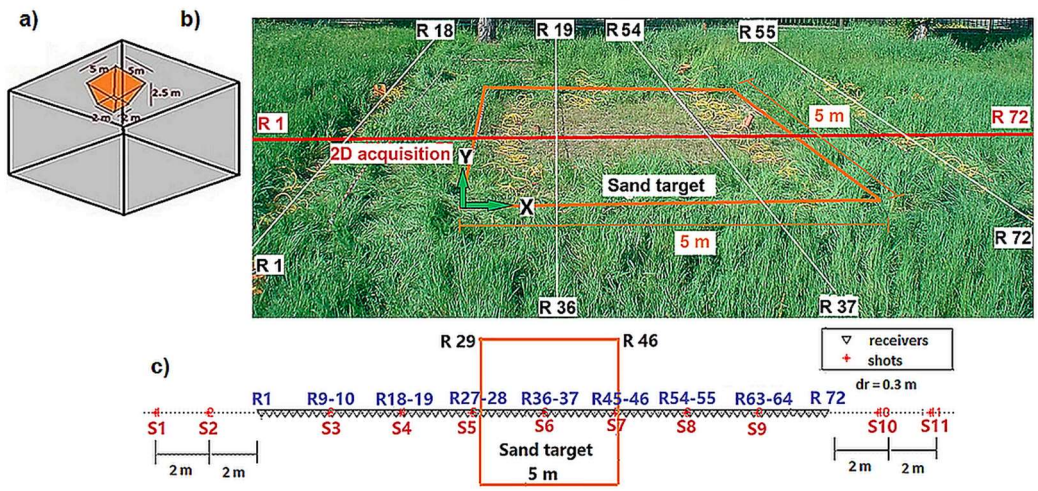

*Figure **6.12** – Field acquisition and site characteristics. **a**) Simplified sketch of the site's geometry. **b**) Image of the site: the line R 1 – R 72 gathers the receivers of the 2D acquisition, while the lines R 1 – R 18, R 19 – R 36, R 37 – R 54, and R 55 – R 72 contain the receivers of the 3D acquisition. **c**) Details of the 2D acquisition: the receivers (R) located near the source positions are indicated with numbers from 1 to 72. The shots (S) are signed with numbers from 1 to 11. The orange rectangle marks the position of the sand target.*

The 3D acquisition contains 4 seismic lines perpendicular to the 2D acquisition direction (Figure **6.12b**). Each seismic line of the 3D acquisition (Figure **6.13**) gathers 18 vertical receivers (4.5 Hz), evenly spaced every 0.5 m. The in-line distance between the sources is 0.75 m. The shot points are perpendicular to the receiver lines. The distance between two adjacent receiver lines is 2.5 m, while the distance between the shot "lines" is 2 m.

Additional 8 common shot gathers (highlighted with green circles in Figure **6.13**) have been also recorded outside both extremities of each seismic line. The in-line distance between these shots and the corresponding first receiver is 2 m.

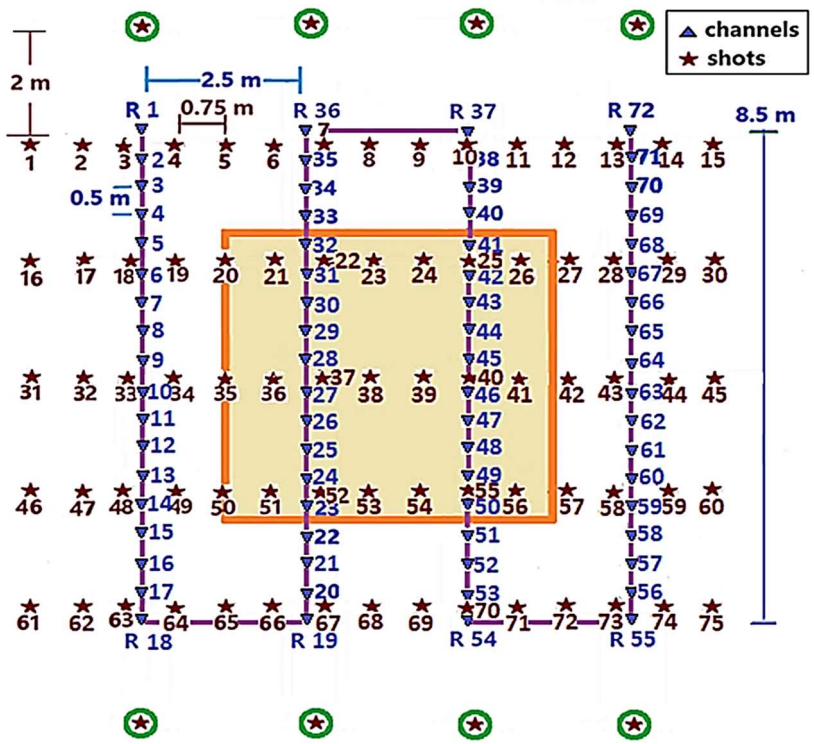

*Figure **6.13** – Scheme of the 3D acquisition. The channels are indicated with blue triangles, while the shots are signed with dark-red stars. The orange rectangle marks the position of the low-velocity target. The shots highlighted with green circles have been also considered (together with the 2D acquisition's shot) in the DC analysis step presented in the next Chapter.*

The 2D acquisition contains 18 receivers placed in correspondence of the low-velocity target, while 20 receivers of the 3D acquisition are located inside the target. The signal recorded by these channels may be affected by a different instrument response, compared with the signal registered by the receivers placed outside the sand target. Thus, the particle motion amplitude's measurement may exhibit some distortions of the recorded wavefield over the acquisition area, caused by the different coupling of the receivers with the ground and the potential geophone tilting. The same observation regards the different coupling of the sources with the ground over the acquisition area.

Figure **6.14** shows an example of two common shot gathers, belonging to the 2D acquisition: Shot 1 (Figure **6.14a**), placed at the left extremity of the receiver array, and Shot 4 (Figure **6.14c**), located in a transition area between the background medium and the low-velocity target. Figures **6.14b** e **6.14d** show some details of the same shots for a narrower time window (from 0 to 0.15 s).

The wave propagation along the seismic line markedly evidences the presence of the low-velocity target. In both figures, the target's position is marked by scattered and dispersive SWs. The BWs amplitude, together with the SWs amplitude after crossing the low-velocity target (far-offset), is very low (of about three orders of magnitude lower compared with the SWs amplitude before crossing the target). The same patterns can be observed as well in Figure **6.15**, which shows examples of common shot gathers belonging to the 3D acquisition: Shot 8 (Figures **6.15a** and **6.15b**) and Shot 68 (Figures **6.15c** and **6.15d**).

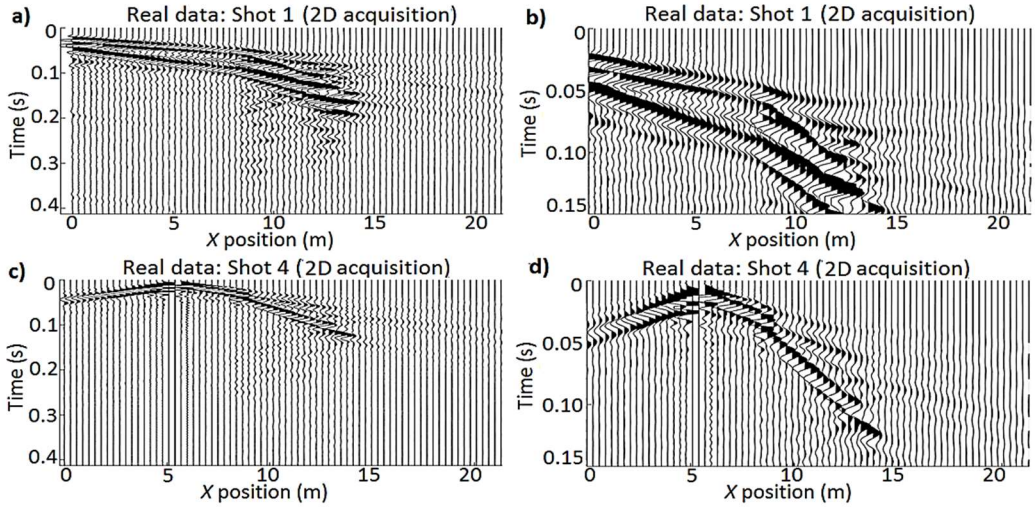

*Figure **6.14** – Example of data recorded in the CNR area (2D acquisition). **a)** Shot 1, **b)** detail of an early tine-window for the Shot 1, **c)** Shot 4, **d)** detail of an early tine-window for the Shot 4. The seismic traces placed in correspondence of the source have been removed because of saturation related effects. The entire seismograms are normalized by the corresponding maximum value.*

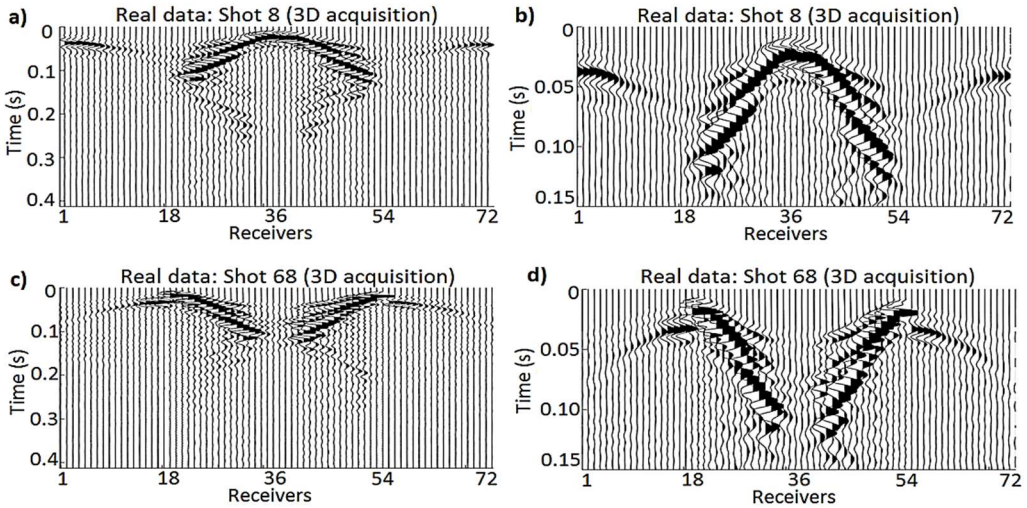

*Figure **6.15** – Example of data recorded in the CNR area (3D acquisition). **a)** Shot 8, **b)** detail of an early tine-window for the shot 8, **c)** Shot 68, **d)** detail of an early tine-window for the shot 68. The entire seismograms are normalized by the corresponding maximum value.*

Figure **6.16** shows an example of frequency spectra for all traces corresponding to some common shot gathers belonging to the 2D acquisition, in particular: Shot 1 (Figure **6.16a**), placed far from the low-velocity target and Shot 6 (Figure **6.16b**), placed in the middle of the sand target. It can be noticed that the dominant frequency is about 20-40 Hz, the central frequency is about 50-60 Hz, while the maximum frequency is about 150 Hz.

Of course, the amplitude of the recorded signal is lower for the shot activated at a longer distance from the sand target (about 3e+08 - Figure **6.16a**) than for the shot activated in correspondence of the low-velocity target (about 1e+09 - Figure **6.16b**).

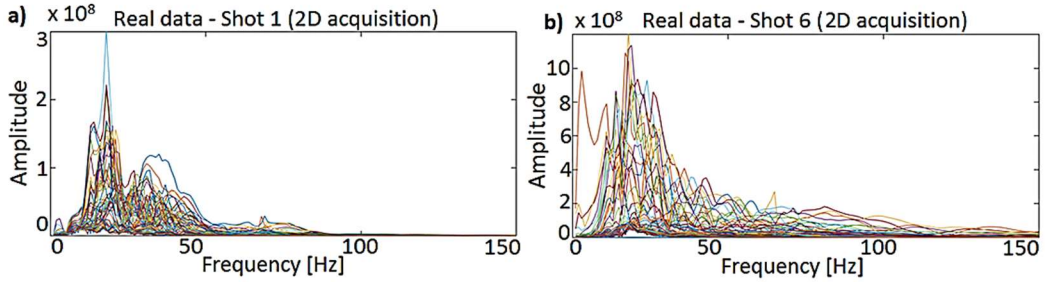

*Figure **6.16** – Example frequency spectra of data recorded in the CNR area (2D acquisition). **a**) Shot 1, **b**) Shot 6.*

6.4 – Synthetic data set

A synthetic reference 3D model (Figure **6.17**), which mimics the real site's geometry and dimensions, was built for generating the synthetic data set (i.e. reference/true data). The model is characterized by a 3D layered structure with a low-velocity target embedded. To avoid boundary artefacts during numerical simulations, a model extension beyond the acquisition area was done.

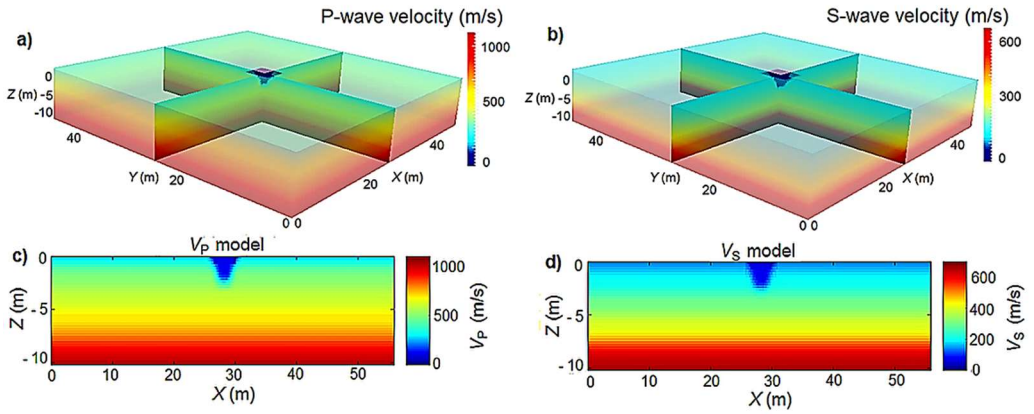

*Figure **6.17** – Synthetic 3D model used to generate the reference data set. **a**) V_P model, **b**) V_S model, **c**) and **d**) 2D vertical sections (crossing the target) of the 3D V_P and V_S models, respectively.*

Forward simulations of wave propagation on this synthetic 3D model were performed using the SEM46 code. To obtain a realistic data set, the viscoelastic wave equation was used. Accordingly, together with the *P*-wave velocity, *S*-wave velocity and density model, the attenuation, in terms of quality factors (Q_P and Q_S), was also considered.

The density value is constant over the 3D domain (1800 kg/m^3, characteristic for unconsolidated sediments), while the 3D variation of the attenuation parameters was retrieved using rheological relations (Hauksson & Shearer, 2006, modified to fit the characteristics of a shallow environment): $Q_S = 0.15 \cdot V_S$; $Q_P = 1.5 \cdot Q_S$. Particularly, the Q_S varies between 12 (in correspondence of the sand target) and 74 (at the deepest part of the medium), while the corresponding minimum and maximum Q_P is 18 and 110, respectively.

The simulation scheme was designed considering the numerical stability criteria in space and time. For the computation mesh, the condition of including at least five GLL points per shortest wavelength was honored, leading to an element size of 0.3 m for the 3D grid, equal in the *z*, *x* and *y* directions. According to the CFL (Courant-Friedrichs-Lewy) time stability criterion, the sampling time step was set to 1.4e-05 s, for a total duration of the simulation of 0.412 s, containing 29400 samples (i.e. time iterations).

A Ricker wavelet with a central frequency of 60 Hz (similar to the one registered in the real data) was the source time function. A free surface condition was applied at the top of the model while absorbing sponges were used for the other boundaries.

Figure **6.18** shows an example of two common shot gathers, belonging to the 2D acquisition performed over the reference (i.e. true) model: Shot 1 (Figure **6.18a**), placed at the left extremity of the seismic line and Shot 4 (Figure **6.18c**), located in a transition area between the background medium and the low-velocity target. Figures **6.18b** e **6.18d** show some details of the above-mentioned shots for a narrower time window (from 0 to 0.15 s).

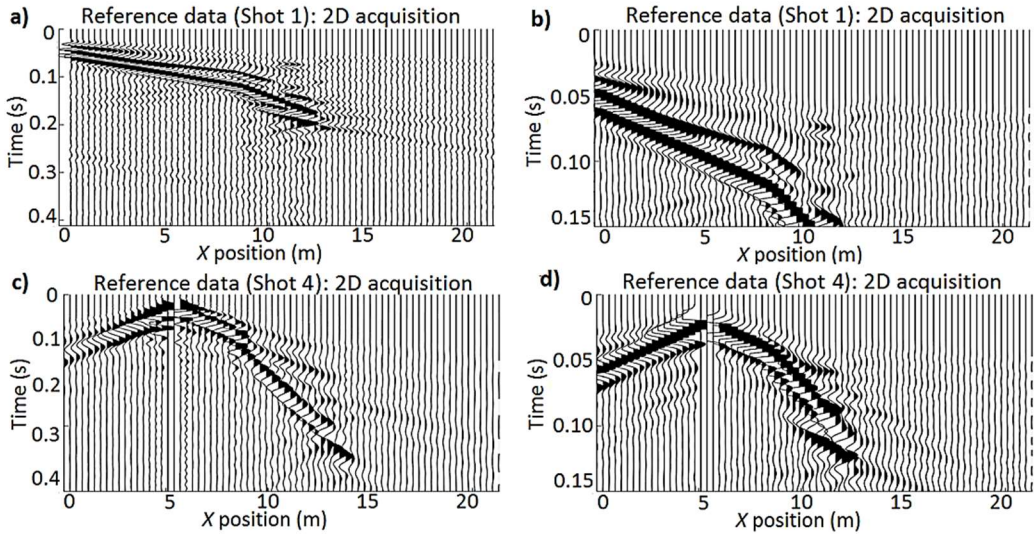

*Figure **6.18** – Example of synthetic data belonging to the reference model. **a**) Shot 1, **b**) detail of an early tine-window for the Shot 1, **c**) Shot 4, **d**) detail of an early tine-window for the Shot 4. The seismic traces placed in correspondence of the source have been removed because of saturation related effects. The entire seismograms are normalized by the corresponding maximum value.*

Analyzing Figure **6.18**, many similarities can be noticed between the wavefield pattern of the real data and the wavefield features of the reference (i.e. true) data. The presence of the sand-target is highlighted by high-amplitude SWs, exhibiting scattering behavior in correspondence of the target boundaries. The BWs amplitude is about three orders of magnitude lower than the SWs amplitude. Most of the seismic energy remains trapped inside the target boundaries: the signal is lower (of about 4 orders of magnitudes) after crossing the low-velocity target.

6.5 – Highlights

Numerical simulations

- The numerical simulations results proved how seismic data recorded in shallow land environments are dominated by SWs, while BWs are very weak. The presence of a flat topography facilitates the SW forward scattering propagation regime, as well as the fundamental mode's dominance.
- Data recorded over a low-velocity target, embedded into a homogeneous medium, are dominated by scattered SWs, almost completely trapped inside the target's boundaries.
- Passing from a homogeneous background to a vertically variable one, the energy trapped inside the target boundaries decreases as the lateral leakage increases.
- For a multilayered background, the forward and back-scattered phases (in correspondence of the target boundaries) are less energetic (but still significant); an important amount of SWs energy is lost in the dispersion across the vertically layered structure.

Real and synthetic data set

- The real target of this study reaches a depth of 2.5 m and is characterized by low velocity and a particular shape.
- Two different seismic acquisitions follow a 2D and a 3D pattern, respectively.
- The acquisition parameters correspond to the geotechnical scale exploration: The maximum offset is 25.3 m, while the minimum source-receiver distance is 0.3 m.
- The source direction is vertical. The receivers recorded only the vertical component of the particle displacement. Possible coupling related effects may be present.
- To increase the S/N, for each source position, a stack of 8 to 10 shots was done.
- The real and synthetic data set present similar characteristics.
- The data sets are dominated by high-amplitude SWs, showing dispersive behaviour and complex scattered phases, especially in correspondence of the low-velocity target.
- The BWs amplitude is about three orders of magnitude lower than SWs amplitude. The SWs amplitude is higher in correspondence of the low-velocity target and weaker after crossing the target.
- Small-wavelength phases, scattered from the target's boundaries, are present in the data sets.

Chapter 7

S-wave and P-wave velocity models from surface wave analysis

Contents

7.1 – Introduction

This Chapter presents some practical applications of the velocity model building procedure from SW analysis (described in Section 4.3). The SW analysis workflow provides initial V_S and V_P models for FWI. These models are subsequently used as a first guess in a spectral element based elastic 3D FWI workflow, presented in the next Chapter.

In this chapter, the results from SW analysis are evaluated in terms of data fitting comparison and DCs comparison. Two different sets of V_P - V_S models, mentioned in Section 1.2, were retrieved from DC analysis: laterally homogeneous and laterally variable. All experiments were conducted over both a synthetic and a field data set. Some results are presented in Teodor et al. (2017), Teodor et al. (2018a) and Khosro Anjom et al. (2019).

7.2 – Synthetic example

Several tests were first carried out for the synthetic data set presented in Section 6.4, which realistically reproduces the CNR site characteristics and acquisition design. Further, the same workflow was applied to the field data set presented in Section 6.3.

7.2.1 – Initial model from dispersion curves analysis

From the reference synthetic data set (i.e. true data), 9 DCs have been extracted along the seismic line (2D acquisition), corresponding to 9 Gaussian windows (Figure **7.1a** and **7.1b**). For the Gaussian windowing algorithm (Bergamo et al., 2012), the value α = 6 (number of receivers considered for one Gaussian window, related to the window's standard deviation $\sigma=N/2\alpha$) was used, leading to a lateral resolution of the local DC of about 3.6 m (considered as twice the standard deviation). Since high lateral heterogeneities are expected along the seismic line, the value of the parameter β (related to the distance between adjacent Gaussian picks) was the same as that of the standard deviation (which is 6). This means that one DC every 1.8 m was extracted. In practice, if no significant lateral heterogeneities are expected, the value of the parameter β can be equal to the conventional lateral resolution 2σ (Bergamo et al., 2012). Regarding the spectral resolution, considering the maximum reliable DC investigation depth (for the tested case) of about 8-10 m, a minimum wavenumber of about 0.3 rad/m was achieved ($k_{min} \approx 2\pi/2.5\ z_{max}$). The choice of the parameters α and β should account for the trade-off between lateral resolution of DC and spectral resolution (please see Section 4.3): the wider the Gaussian window (the lower the lateral resolution) is, the higher the spectral resolution is.

The DC(s) were converted into velocity profiles by inverting or using the data transform on a single DC curve (single-DC analysis) and a clustering algorithm together with the data transform on all DCs along the receiver array (full-DC analysis). For the latter case, the DCs corresponding to the receivers 25 (DC 3) and 49 (DC 7), placed in the transition zone between the sand target and the vertically layered structure, have been identified as outliers by the clustering algorithm (Figure **7.1c**); for this reason, they were not considered for building the velocity model. Further, by replicating laterally the 1D velocity

profile retrieved from the single-DC analysis, and by interpolating the V_P and V_S profiles retrieved from the full-DC analysis, 2D models were obtained. The models have been extended laterally and in depth to avoid boundary artefacts during numerical simulations. The lateral extension was based on invariance, while for the depth extension the low-frequency information provided by DCs was exploited. Therefore, no information coming from the reference models was used for the extension process. The obtained 2D models are shown in Figures **7.2a** and **7.2b** for the single-DC analysis, and Figures **7.2c** and **7.2d** for the full-DC analysis.

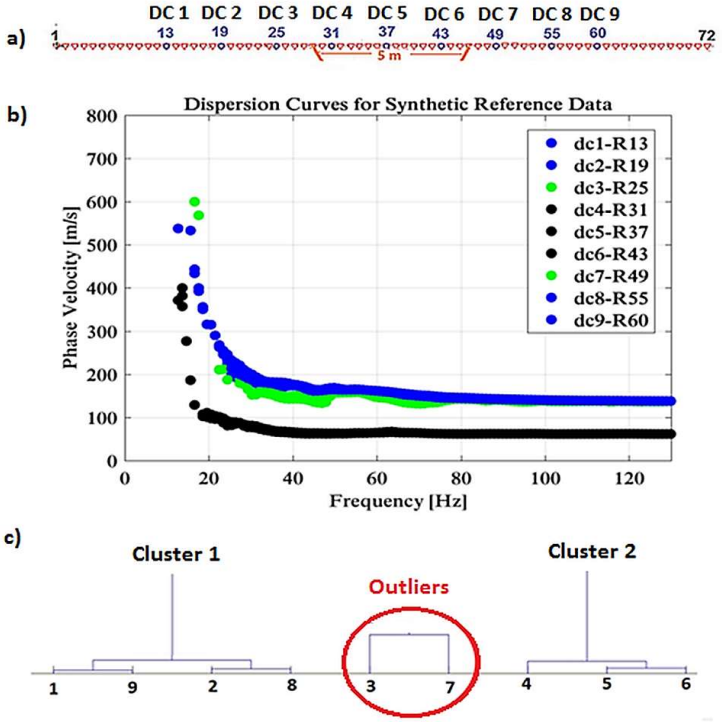

*Figure **7.1** –**a**) DCs position: The numbers 13, 19, 25, 31, 37, 43, 49, 55 and 60 indicate the receivers coincident with a DC. **b**) DCs extracted from reference synthetic data set. **c**) Example of DC clustering (from Teodor et al., 2018 - modified).*

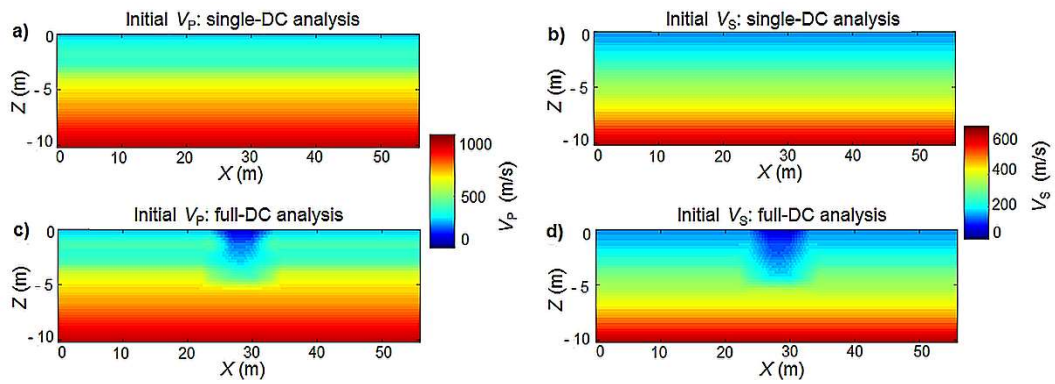

*Figure **7.2** – **a**) V_P model and **b**) V_S model retrieved from DC 1 – R 13 (single-DC analysis). **c**) V_P model and **d**) V_S model retrieved from the analysis of the entire set of DCs along the seismic line (full-DC analysis) – from Teodor et al. (2019), modified.*

The initial model from the single-DC analysis does not contain any information related to the low-velocity target, while the initial model related to the full-DC analysis contains the low-velocity target, although the resolution is low in correspondence of the boundaries.

As required by the FWI code, the 2D models were converted to 3D volumes, by symmetrically replicating the *zx* vertical sections in the *y*-direction and reproducing the 3D target's extension (Figure **7.3**).

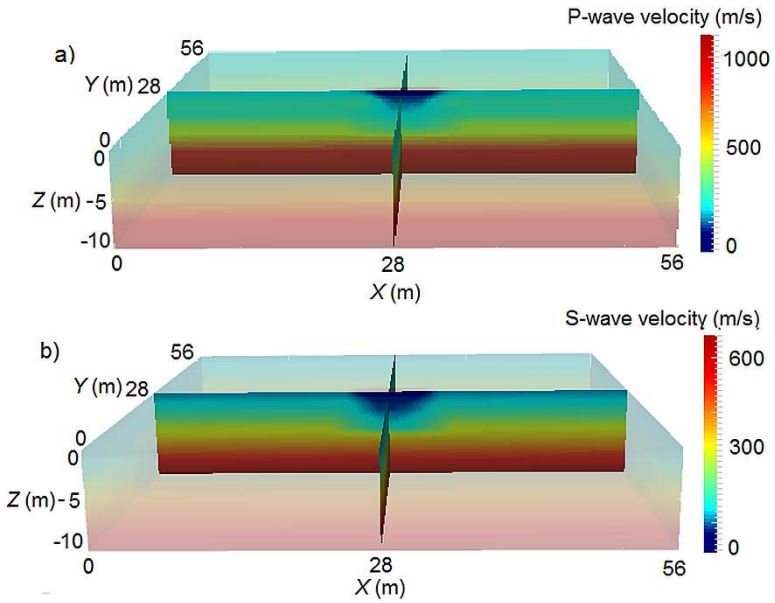

*Figure **7.3** – Example of 3D-extended initial models for the synthetic case:*
***a)** V_P model, **b)** V_S model.*

In the *y*-direction, beyond the horizontal extension of the low-velocity target (5 m), the model is characterized by a vertically layered and laterally homogeneous structure (Figure **7.4**), as in the *x*-direction, far from the sand target.

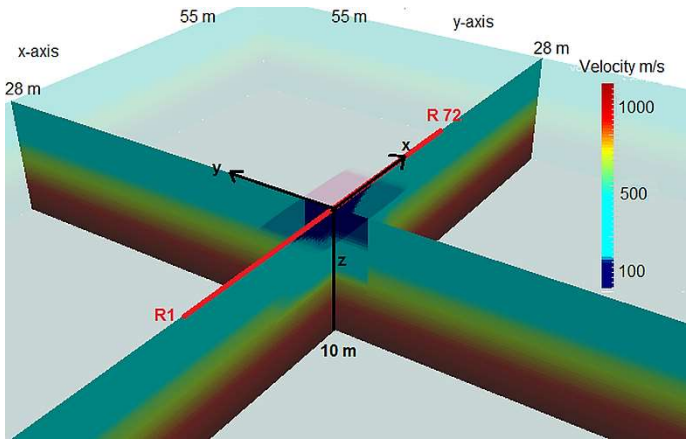

*Figure **7.4** – Details on the model extension in the y-direction.*

7.2.2 – Data fitting comparison

For a preliminary evaluation of the initial models' accuracy, forward simulations were performed on the estimated models (characterized by higher resolution), keeping unchanged all the parameters used to generate the reference (i.e. true) data set.

Figure **7.5a** shows an example of reference data while Figure **7.5b** shows an example of seismic data belonging to the initial model built from synthetic DCs (called, hereafter, "initial" data). Similar features of the waveforms can be distinguished. The SWs domination is evident, while the BWs amplitude is very low. Better continuity of the phases leaked outside the target can be noticed in the initial data (Figure **7.5b**), due to the smooth character of the initial model retrieved from DC analysis, where the impedance contrast in correspondence of the target's boundary is lower than in the reference model.

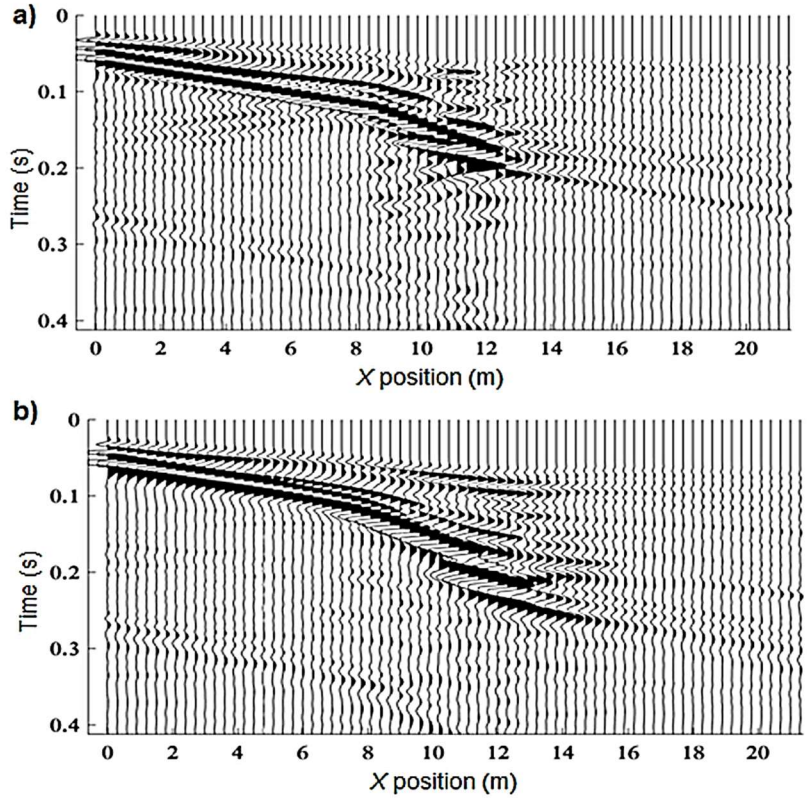

*Figure **7.5**. Example of synthetic data (vertical component) for Shot 1. Seismic data belonging to the **a)** reference model and **b)** initial model. The seismograms are normalized by the corresponding maximum value.*

Figure **7.6** displays the trace-by-trace comparison between reference and initial data. The initial data are not cycle-skipped, in the near-offset positions, compared with the reference data. Contrariwise, the matching between reference and initial data is less accurate in some far-offset positions, after crossing the sand target.

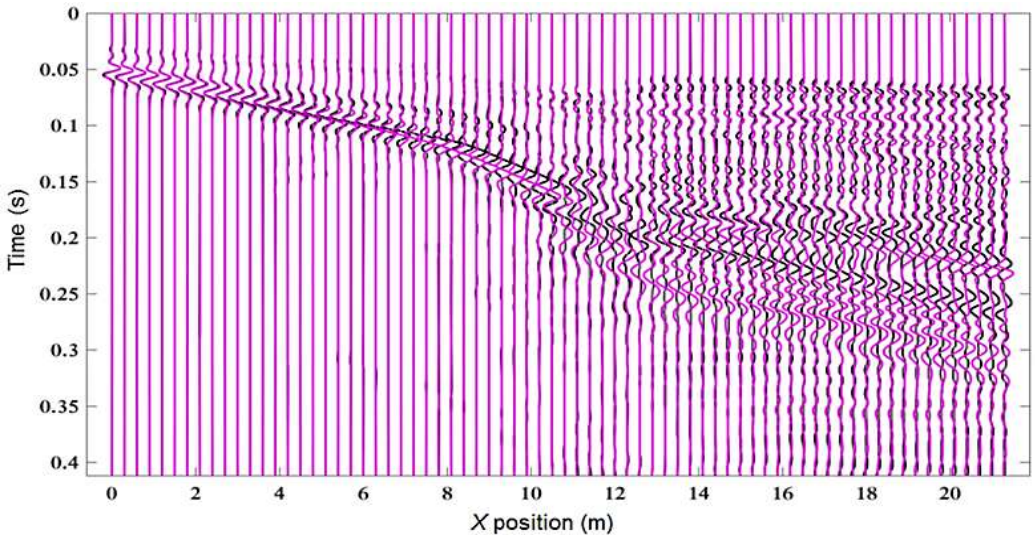

*Figure **7.6** – Trace by trace data fitting comparison between the reference data (in black) and initial data (in magenta). The seismograms are trace-by-trace normalized – from Teodor et al. (2018).*

7.2.3 – Dispersion curves comparison

Figure **7.7** shows an example of *f-k* spectra corresponding to a Gaussian window centred on the sand target, while Figure **7.8** displays all the DCs retrieved from the reference data (Figure **7.8a**) and the data belonging to the initial model built from synthetic DCs (Figure **7.8b**).

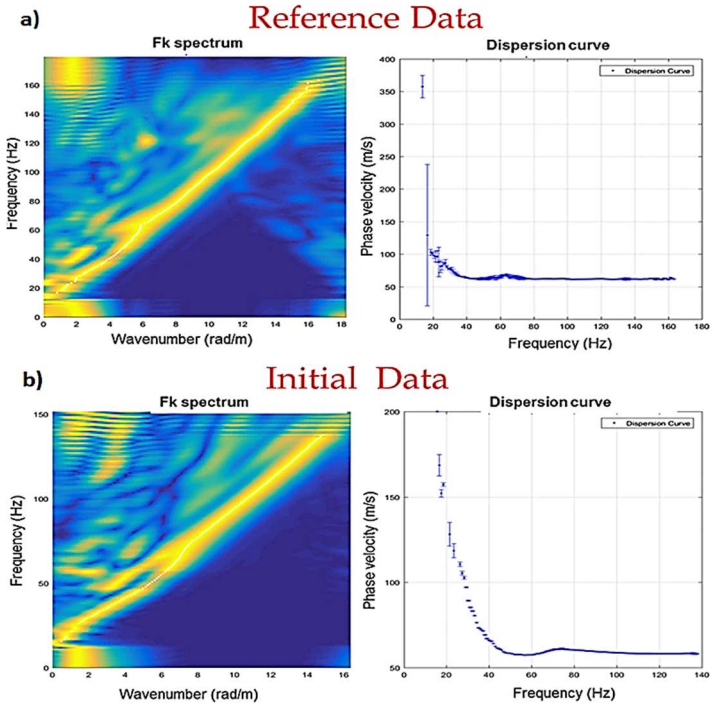

*Figure **7.7** – Example of f-k spectra and the corresponding DCs for a Gaussian window located in the middle of the low-velocity target: **a)** DC belonging to the reference data, **b)** DC belonging to the data computed in the estimated initial model – from Teodor et al. (2018), modified.*

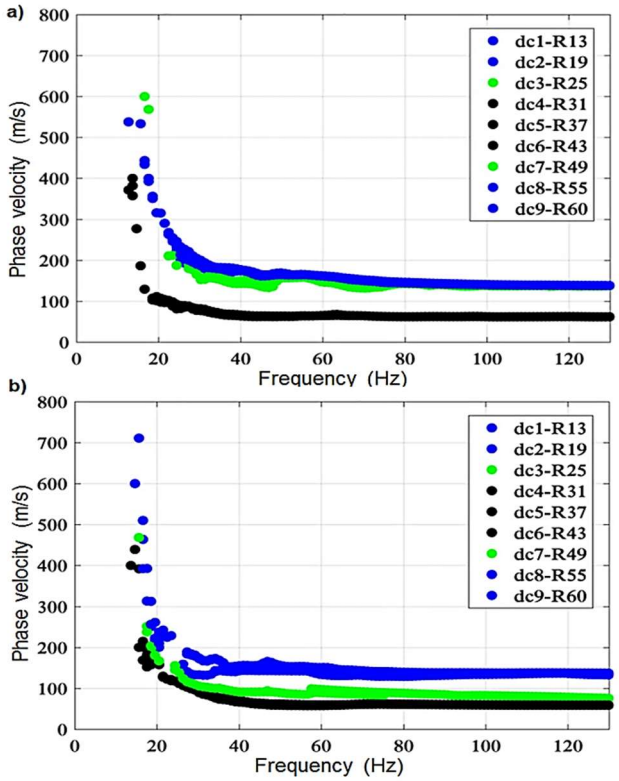

*Figure **7.8** – The DCs belonging to all Gaussian windows. **a**) DCs belonging to the reference data, **b**) DCs belonging to the data computed for the initial model – from Teodor et al. (2018), modified.*

Similar phase-velocity trends can be noticed when comparing the reference DCs (i.e. the DCs extracted from the reference/true data) with the DCs extracted from synthetic initial data. This observation regards particularly the DCs belonging to the Gaussian windows placed outside the low-velocity target (signed with blue), and the DCs belonging to the windows located inside the anomaly (DCs plotted in black). Differently, there are few DCs (represented with green) that exhibit a different phase-velocity trend in the reference and initial data, respectively. These DCs (identified as outliers by the clustering algorithm – Figure 7.1c) belong to a transition zone between the low-velocity target and the vertically layered medium, where the picking of the fundamental mode is tricky. This pattern can be noticed as well in Figure **7.9**, related to the normalized phase-velocity difference between each reference DC and the corresponding DC belonging to the initial data.

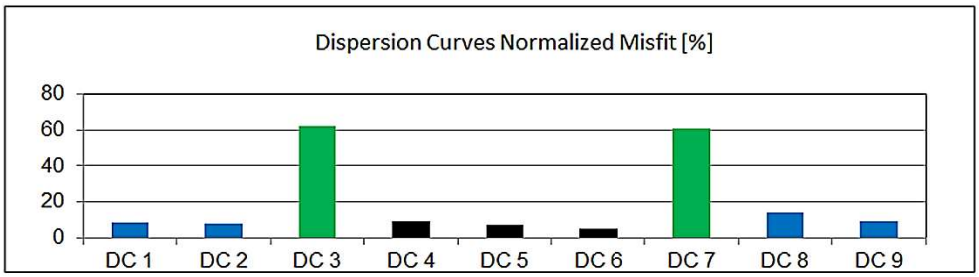

*Figure **7.9** – Normalized misfit (computed as frequency-by-frequency velocity difference) between the DCs belonging to the reference data and the DC extracted from the initial data.*

7.3 – Real data application

For the real data application, all parameters used for the synthetic case were kept unchanged. The additional 4 lines, belonging to a 3D acquisition (Figure 6.13), were also considered to enrich the DCs information content.

Figure **7.10** shows a detail of the DCs positions, including also the DCs belonging to the 3D acquisition, while Figure **7.11** shows all the DCs extracted from real data. Two different phase-velocity trends can be noticed in Figure **7.11**: a lower-velocity one, gathering the DC placed in correspondence of the target or nearby, and a higher velocity one, gathering the DCs placed far away from the target.

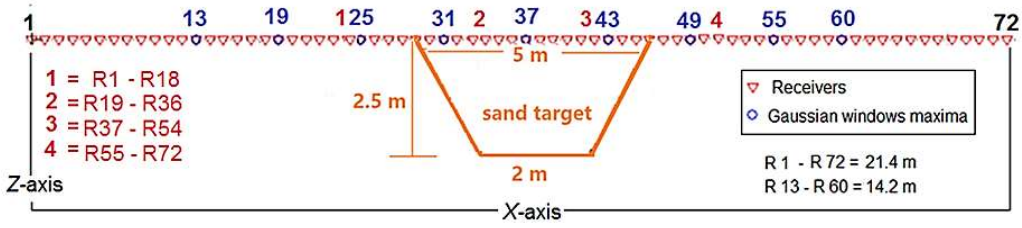

*Figure **7.10** – Scheme of the DCs position. In blue, with numbers from 13 to 60, are indicated the receivers coincident with a DC extracted from the real data belonging to the 2D acquisition. The positions of the 4 DCs related to the 3D acquisition's data are indicated in red, with numbers from 1 to 4 – from Teodor et al. (2018), modified.*

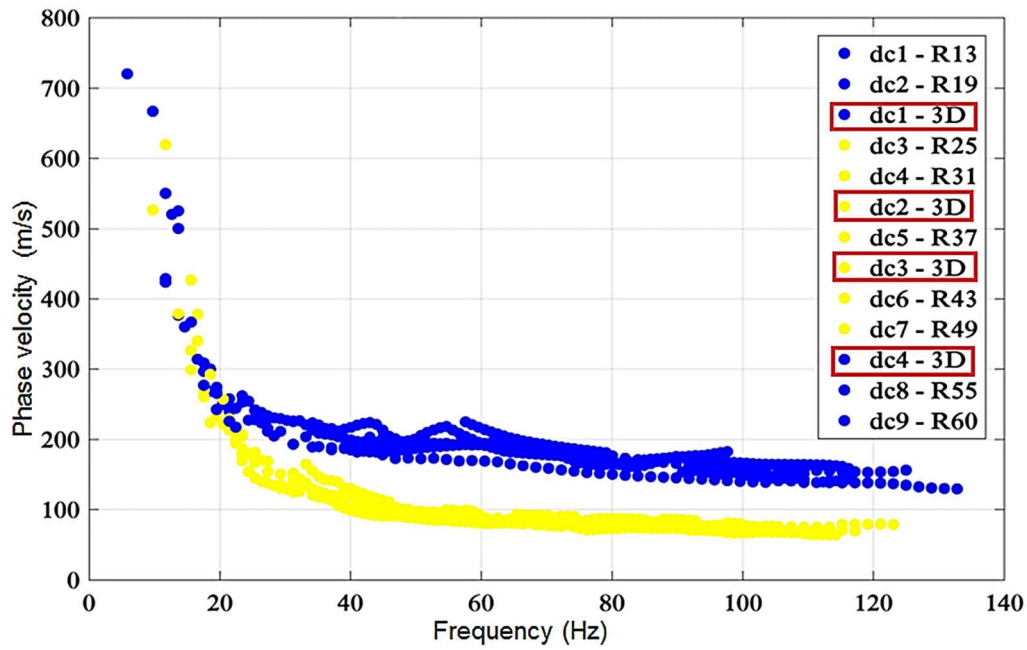

*Figure **7.11** – All the DCs extracted from real data. The red rectangles indicate the DCs extracted from the data belonging to the 3D acquisition.*

7.3.1 – Initial model from dispersion curves analysis

Figures **7.12a** and **7.12b** show the velocity models obtained from the single-DC analysis, while Figures **7.12c** and **7.12d** show the velocity models obtained through the full-DC analysis (Khosro Anjom et al., 2019). In the last two figures, a nice reconstruction of the low-velocity target can be noticed, where the maximum depth and the lateral boundaries are correctly recovered. Again, the 2D models were extended to 3D following the same procedure as for the synthetic case.

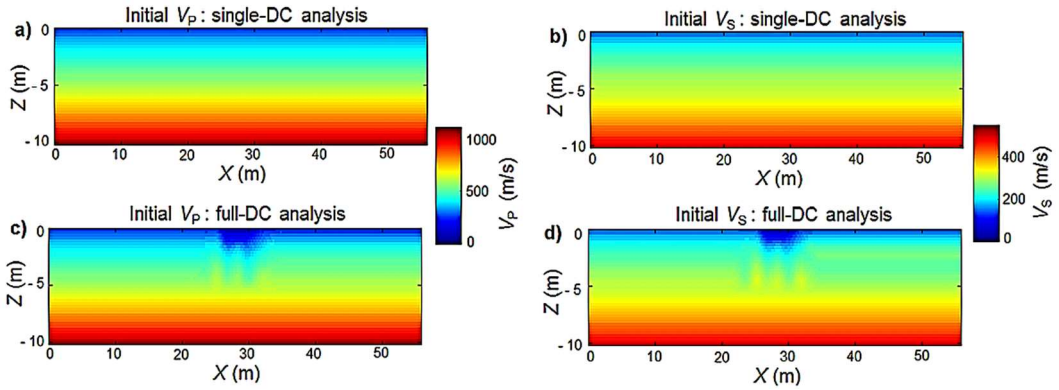

*Figure **7.12** – The 2D models retrieved from the analysis of real-data related DCs: **a)** V_P model and **b)** V_S model obtained from the single-DC analysis of real data. **c)** V_P model and **d)** V_S model obtained from the full-DC analysis of real data.*

7.3.2 – Data fitting comparison

Some 3D forward simulations were performed on the initial models obtained from full-DC analysis, to retrieve the corresponding initial data and compare them with the real ones. During the simulations, all parameters were maintained unchanged, except for the source time function. In this case, the real source was estimated from the field data by deconvolution (Pratt, 1999). The Green's function was computed in the starting model using a known Ricker source, with a central frequency of 60 Hz and unitary amplitude.

Figure **7.13a** shows an example of real data, while Figure **7.13b** shows an example of data belonging to the elastic model retrieved from DCs. Waveforms characterized by similar features can be noticed, whereas some differences at the target's position are also perceptible.

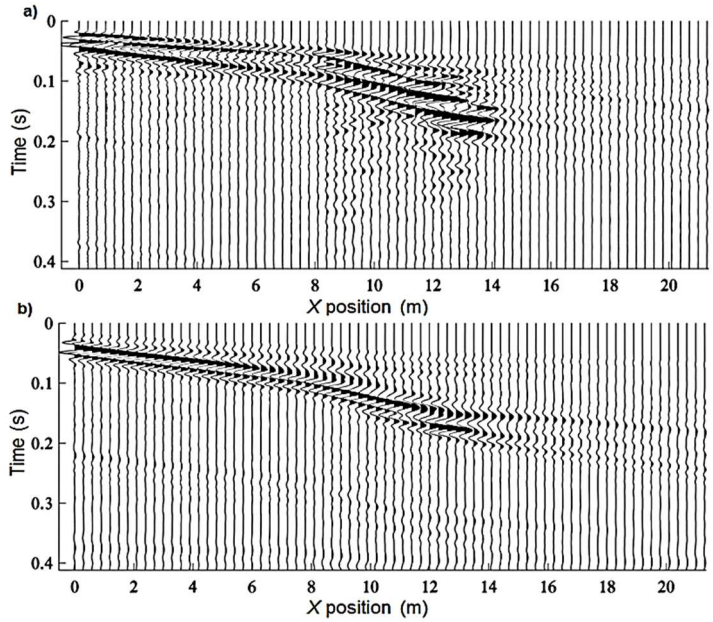

*Figure **7.13** – Example of seismic data for Shot 1. **a**) Real data. **b**) Data belonging to the initial model retrieved from the real dispersion curves. The entire seismograms are normalized by the maximum value.*

Figure **7.14** shows a trace-by-trace comparison between real and simulated data, for the same shot position (far-offset configuration). The data corresponding to the estimated models are in good agreement with the field data, especially outside the low-velocity target. There is also a good agreement between the two data sets inside the target, especially for near-offset positions. The fitting is less accurate at far offset and for the BWs.

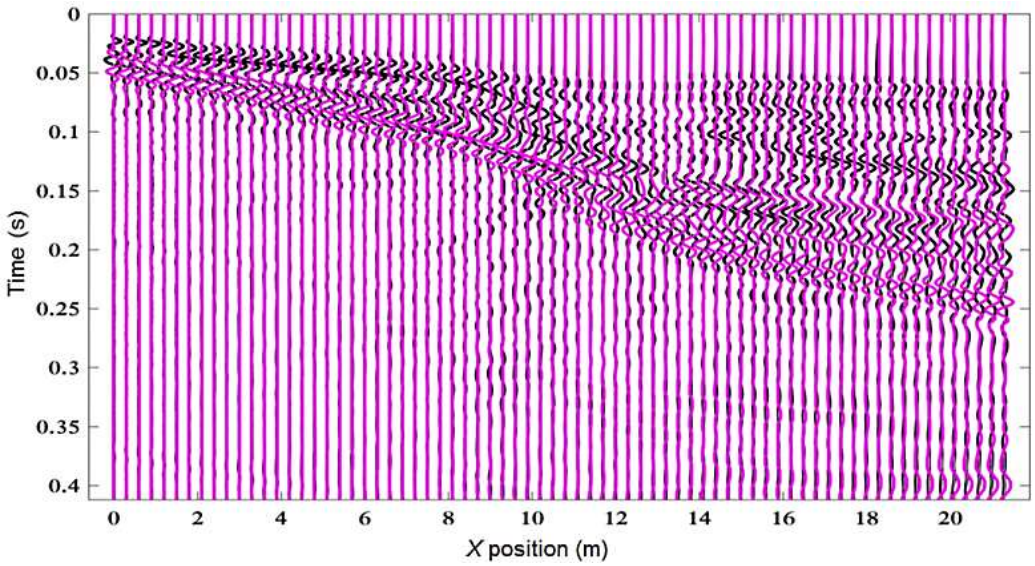

*Figure **7.14** – Trace by trace comparison between field data (in black) and initial data (in magenta). The seismograms are trace-by-trace normalized.*

After a cycle-skipping analysis, a data misfit lower than half of the dominated wavelength was quantified for most parts of the seismograms (Khosro Anjom et al., 2019: the half period for the dominant frequency is about 8.33 ms). This led to the classification of the initial elastic model retrieved from DC analysis as a potentially good candidate for FWI.

A more detailed illustration of the trace-by-trace data-fitting comparison is presented in Appendix 3.

7.3.3 – Dispersion curves comparison

In Figure **7.15**, an example of *f-k* spectra of a Gaussian window centred on the sand target is shown, both for the real data and for the data computed for the estimated initial model.

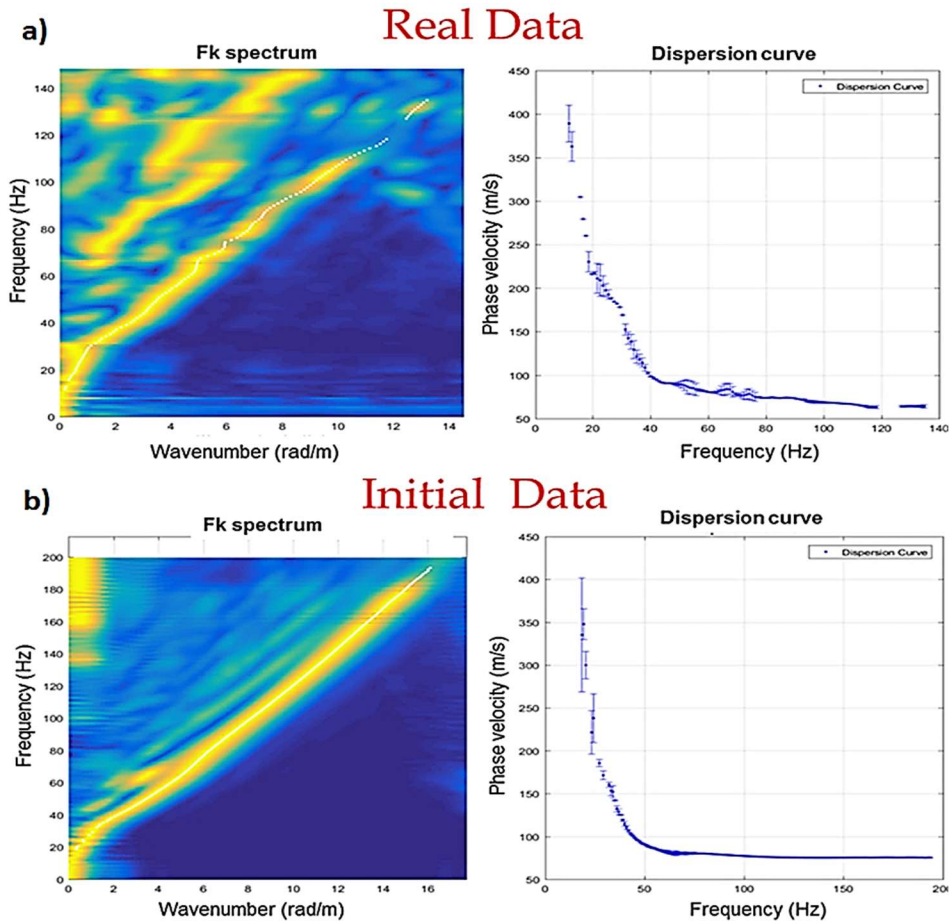

*Figure **7.15** – Example of f-k spectra and the corresponding DCs for a Gaussian window located in the middle of the low-velocity anomaly: **a**) DC belonging to the real data, **b**) DC belonging to the data computed in the estimated initial model.*

Figure **7.16** displays all the DCs retrieved from real data (Figure **7.16a**) and the data belonging to the initial model (Figure **7.16b**), respectively. As for the synthetic application, one can notice similar phase-velocity trends when comparing the real DCs with the DCs belonging to the initial data, especially for DCs placed outside the low-velocity target (signed with blue), and for the ones located inside the anomaly (plotted in black).

Again, as for the synthetic application, there are some DCs (represented with green) that exhibit a different phase-velocity trend in the real and initial data. These DCs are located in a transition zone between the low-velocity target and the vertically layered medium, where the individuation of the fundamental mode is difficult.

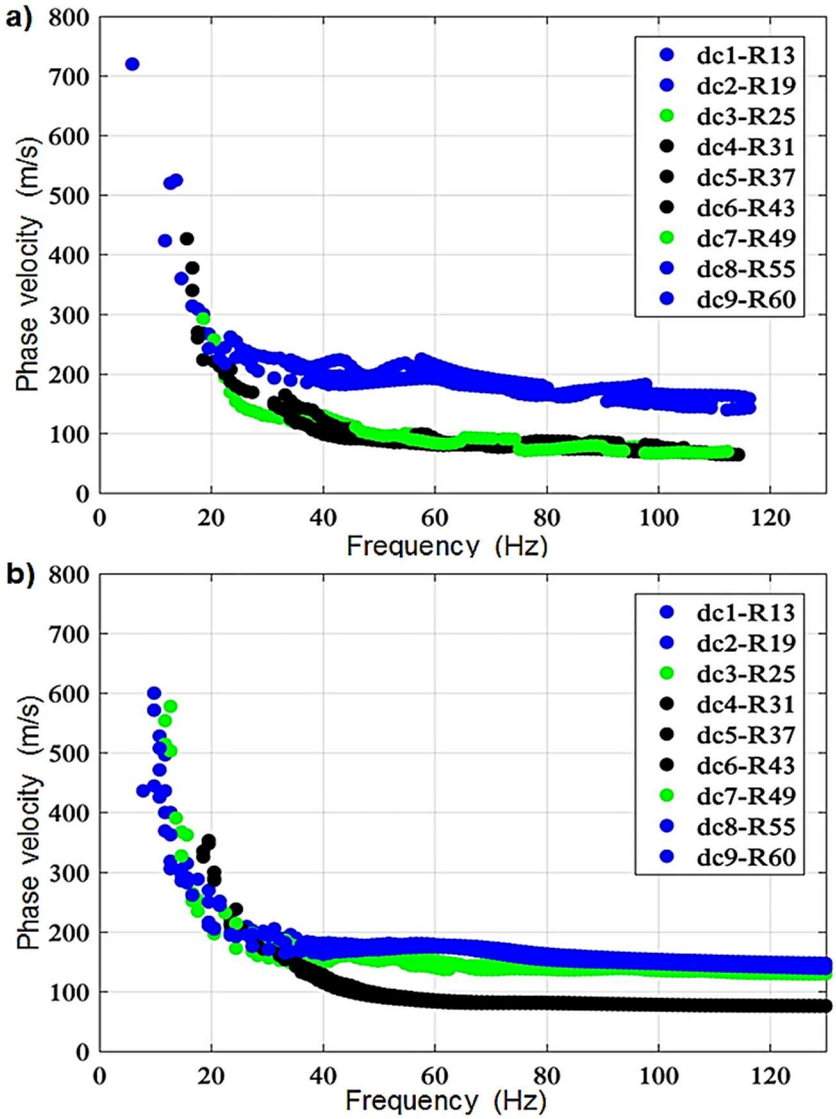

*Figure **7.16** – The DC belonging to all Gaussian windows. **a**) DCs belonging to the real data (2D acquisition), **b**) DCs belonging to the data predicted in the initial model.*

The overall comparison between the reference/real DCs and the ones extracted from the corresponding initial data also show a good degree of matching and similar trends (Figure **7.17**).

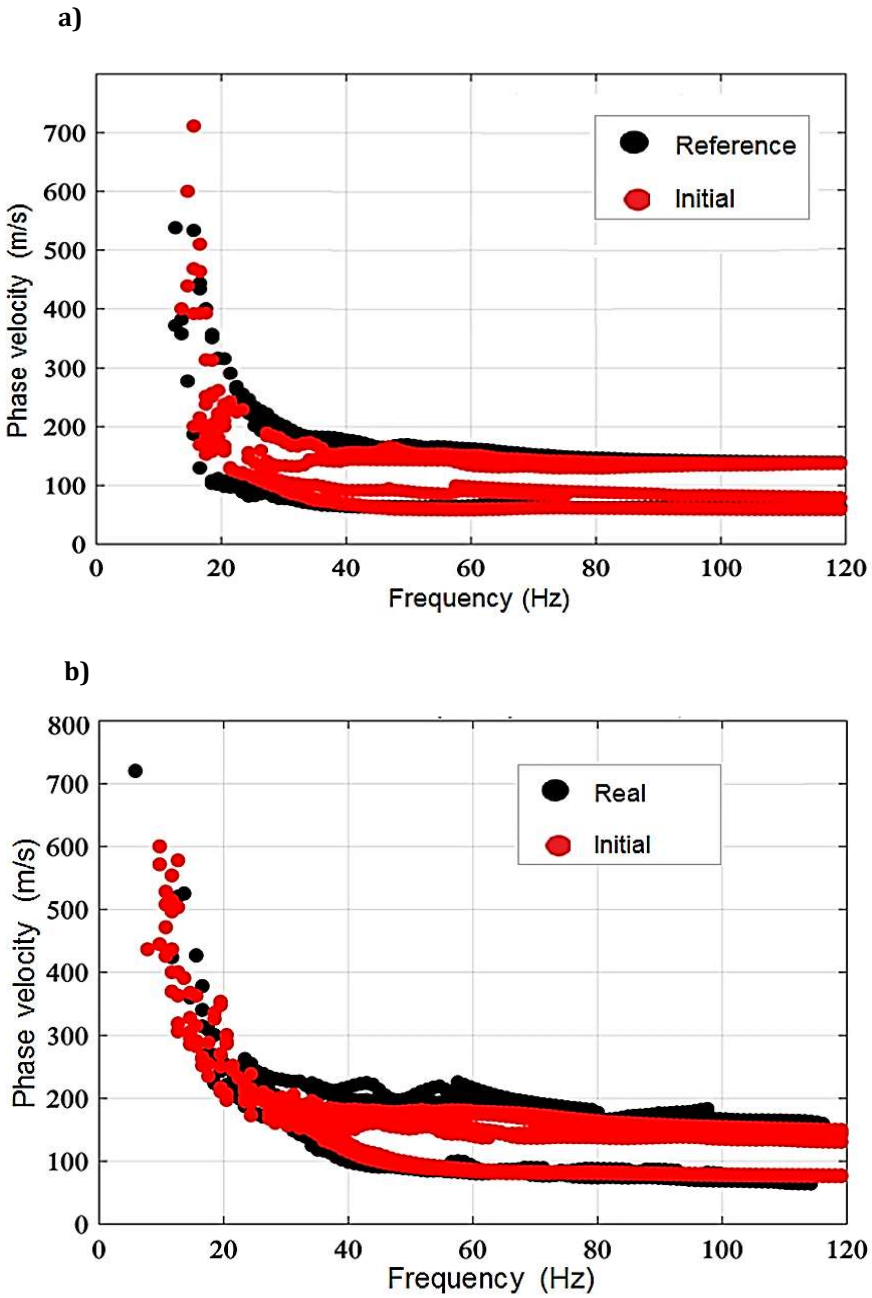

*Figure **7.17** – **a**) Comparison between the DCs belonging to the reference data (in black) and the DCs belonging to the initial data for the synthetic application (in red). **b**) Comparison between the real DCs (in black) and the DCs belonging to the initial data for the real case (in red).*

7.4 – Highlights

- The DC extraction from the data set belonging to the 2D acquisition was performed using a Gaussian moving window along the seismic line. Only the shots placed outside the receiver line have been considered for the stacking in the *f-k* domain. The Gaussian window's lateral resolution is about 3.6 m, corresponding to a wavenumber resolution of about 0.3 rad/m.
- For the real data application, additional 4 DCs were extracted from the shots placed in line with the receivers of the 3D acquisition.
- The method used for the initial model building is based on a DC clustering algorithm and a data transform procedure. It was applied here to both synthetic and field data.
- The estimated Poisson's ratio is variable with depth. For the synthetic case, the minimum estimated Poisson's ratio value is 0.14 while the maximum value is 0.40. For the field case, the minimum and maximum Poisson ratio's values are 0.20 and 0.36, respectively.
- Two different sets of initial models were retrieved: laterally homogeneous and laterally variable. The laterally homogeneous models were obtained by the analysis of one single representative DC (single-DC analysis = 1D), while for building the laterally variable models, a clustering algorithm and data transform was applied to all the DCs along the seismic line (full-DC analysis = 2D).
- In the full-DC analysis, the DCs placed in the transition zone between the low-velocity anomaly and the vertically layered medium were identified as outliers by the clustering algorithm and they were not considered for the velocity model building.
- The trace-by-trace comparison between reference/real data and data corresponding to the initial models retrieved from full-DC analysis shows a relatively good agreement, without cycle-skipping, especially for the near-offset positions. This may qualify the model built from DCs analysis as a potentially good candidate as an initial model for FWI, speeding up the FWI workflow.
- Because of resolution-limitation at greater depth and the consideration of the fundamental mode only, in this study, the DCs are not well-separated in the frequency range lower than 20-30 Hz, while exhibiting clear differences among clusters at higher frequencies. Therefore, while a high-frequency piece of information is somehow well-solved in the initial model from SW analysis, we expect that FWI will better solve also the low-frequency information content. The clear separation of the DCs at high frequencies (that usually does not happen for unconsolidated layers) is granted in this study by the small receiver-interval used in the seismic acquisition, the high signal-to-noise ratio of the data obtained after the stacking process, as well as the good compromise achieved between lateral and spectral resolution in the DCs extraction step through Gaussian windowing.

Appendix 3

Examples of data fitting comparison for initial models retrieved from surface wave dispersion curves analysis

Figures **A 3.1** - **A 3.3** show a data fitting comparison (synthetic case) for a shot position placed at the left extremity of the seismic line (Shot 1). A good degree of matching between reference and initial data can be noticed for the near-offset traces (Figure **A 3.1**), and some traces placed in correspondence of the low-velocity target (Figure **A 3.2**). Conversely, the fitting is less accurate for the far-offset traces (Figure **A 3.3**).

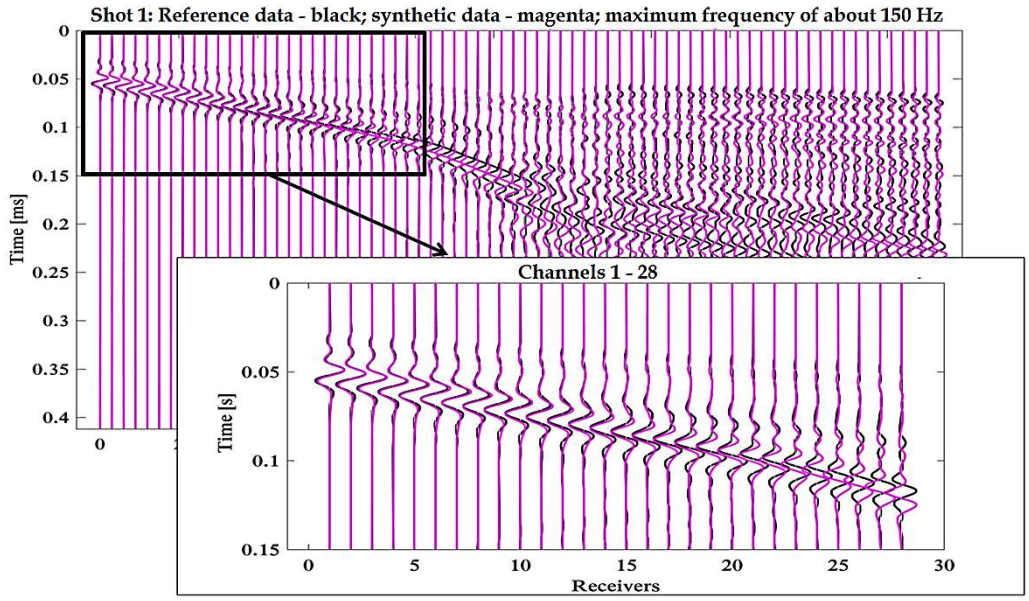

*Figure **A 3.1** – Trace-by-trace data fitting comparison between reference data and data belonging to the initial model from DC analysis. The maximum frequency is about 150 Hz: Shot 1, near-offset traces.*

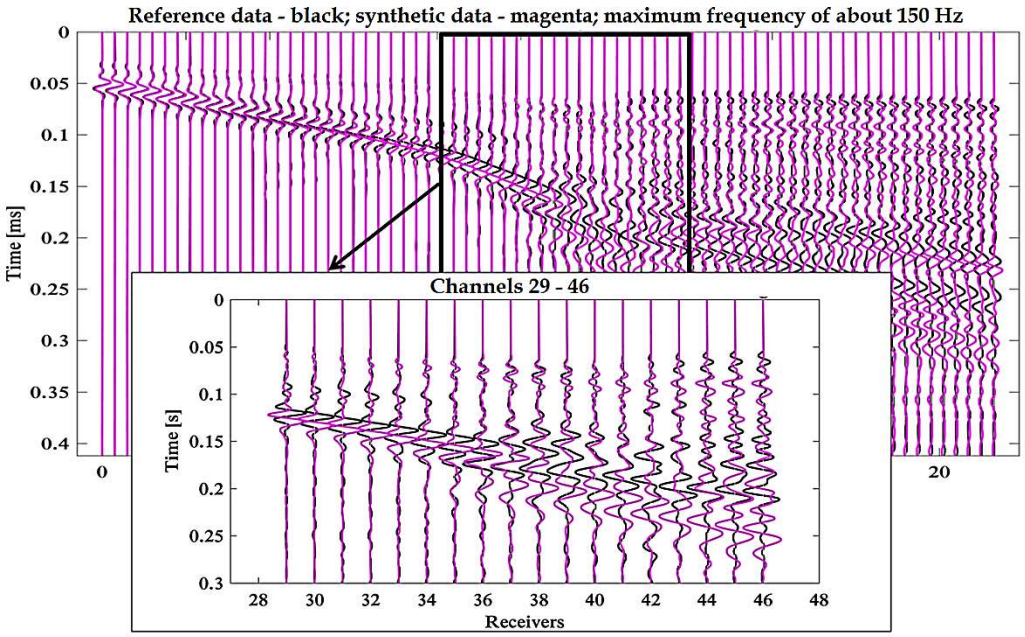

*Figure **A 3.2** – Trace-by-trace data fitting comparison between reference data and data belonging to the initial model from DC analysis: Shot 1, intermediate-offset traces.*

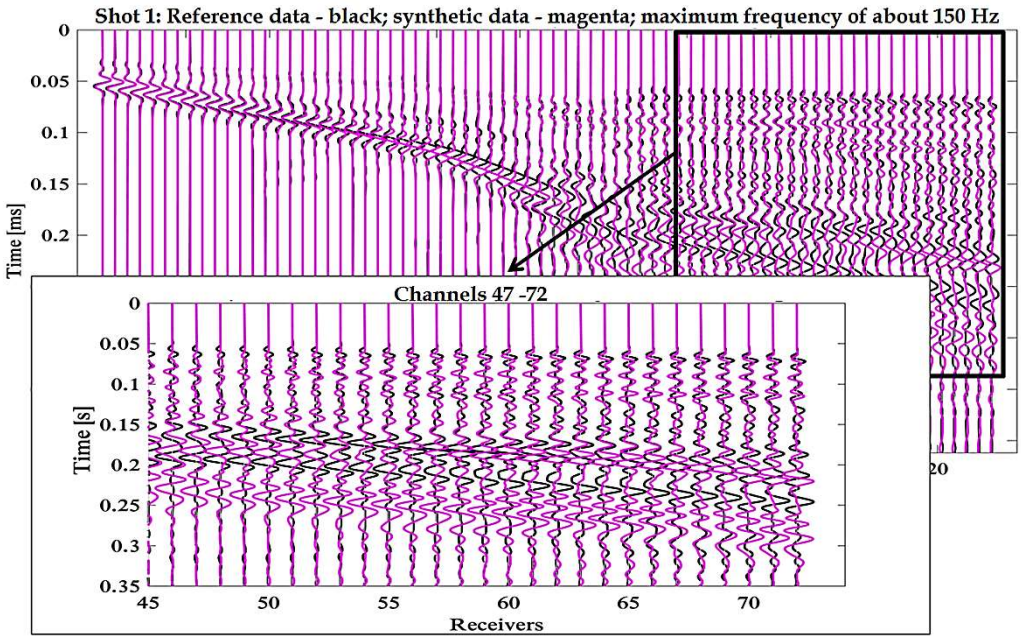

*Figure **A 3.3** – Trace-by-trace data fitting comparison between reference data and data belonging to the initial model from DC analysis: Shot 1, far-offset traces.*

Figures **A 3.4** - **A 3.6** show a data fitting comparison (real case) for a shot position placed at the left extremity of the seismic line (Shot 1). As for the synthetic case, a relatively good degree of matching between real and initial data can be noticed for the near-offset traces (Figure **A 3.4**) and some traces placed in correspondence of the low-velocity target (Figure **A 3.5**). Conversely, the fitting is less accurate for the far-offset traces (Figure **A 3.6**).

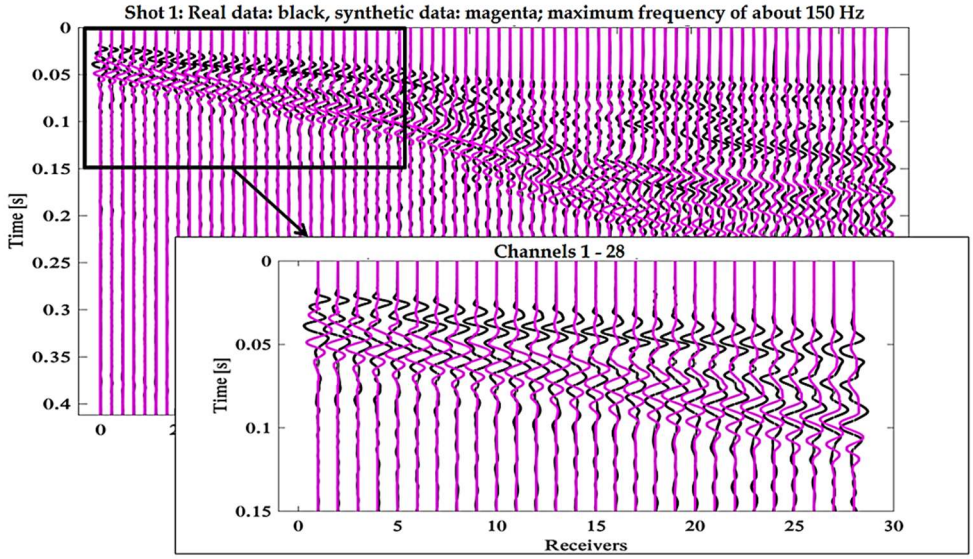

*Figure **A 3.4** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from DC analysis: Shot 1, near-offset traces.*

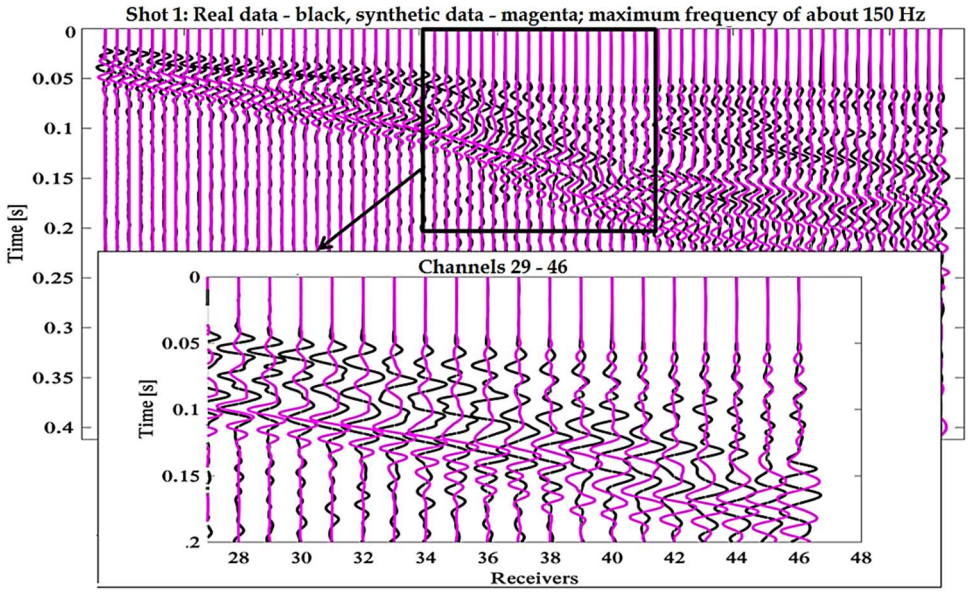

*Figure **A 3.5** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from DC analysis: Shot 1, intermediate-offset traces.*

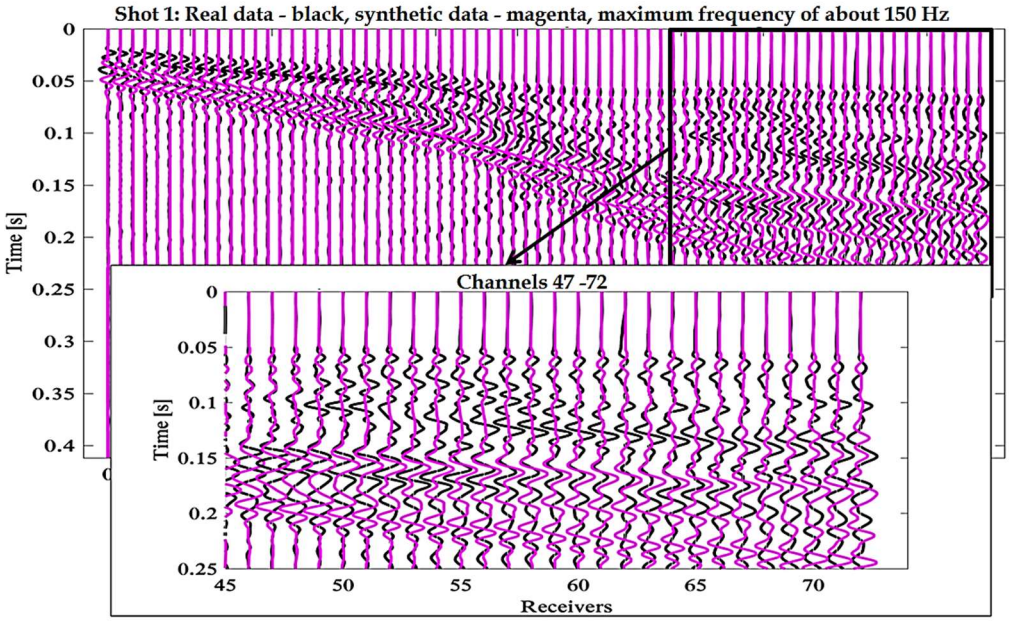

*Figure **A 3.6** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from DC analysis: Shot 1, far-offset traces.*

Figures **A 3.7** - **A 3.9** show a data fitting comparison (real case), for a shot position placed in a transition area between the low-velocity target and the homogeneous background (Shot 4). Again, the fitting is a relatively accurate for the near-offset traces (Figure **A 3.7**), while the degree of matching between real and initial data is less accurate at far offset (Figure **A 3.8** and **A 3.9**).

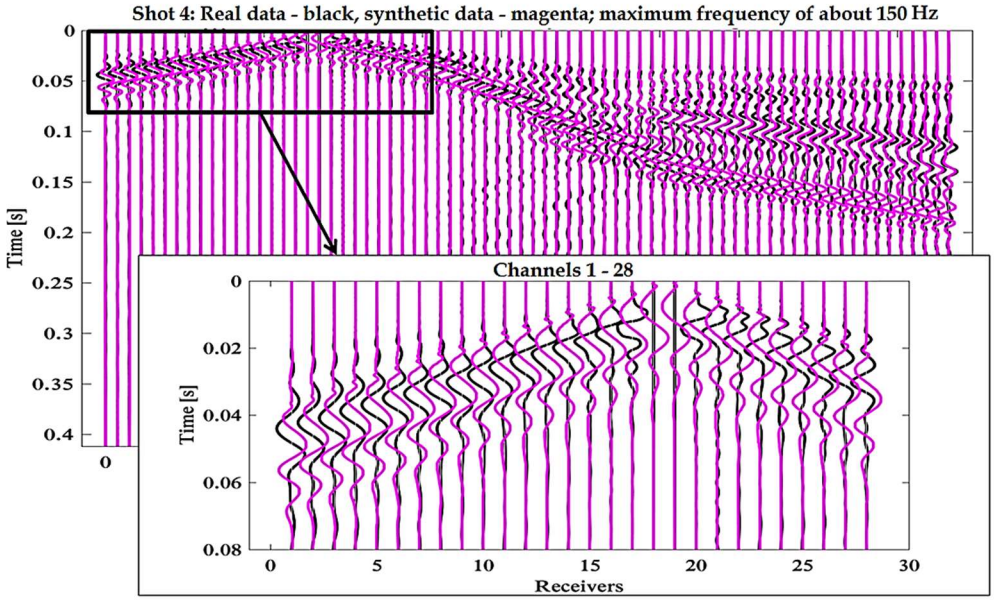

*Figure **A 3.7** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from DC analysis, Shot 4- near-offset traces.*

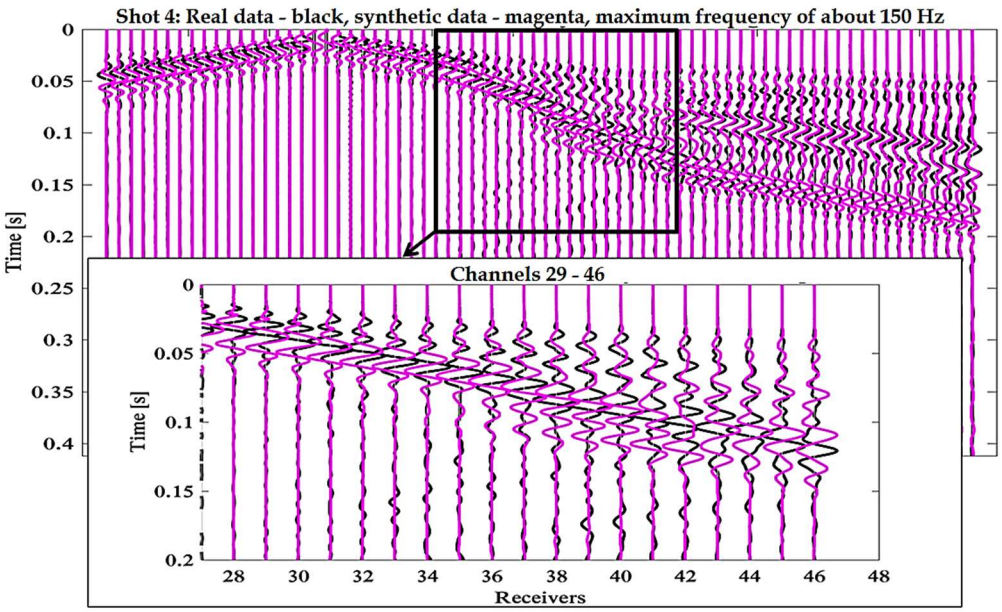

*Figure **A 3.8** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from DC analysis: Shot 4, intermediate-offset traces.*

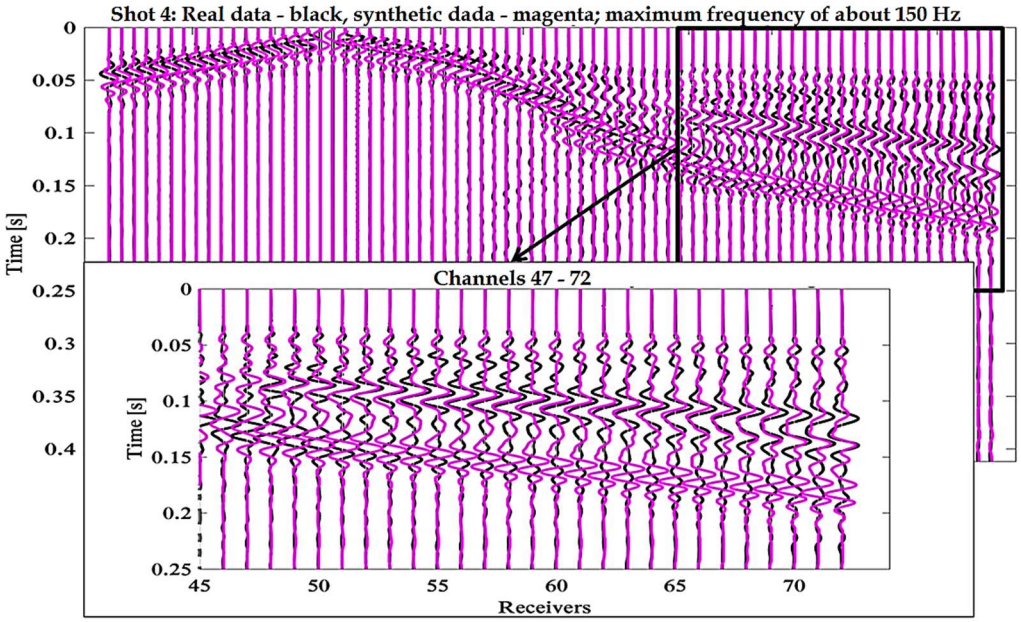

*Figure **A 3.9** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model estimated from DCs: Shot 4, far-offset traces.*

Chapter 8

Multi-parameter 3D elastic full-waveform inversion

Contents

8.1 – Introduction

In this Chapter, some FWI applications are presented, both on synthetic and real data. The inversions are conducted over two different initial V_P and V_S models, extracted from DCs: laterally homogeneous and laterally variable. Hereafter, the first configuration is called "single-DC case (i.e. 1D case)", while the second configuration is called "full-DC case (i.e. 2D case)". The models obtained after FWI are referred to as "final" models, while the data belonging to them are called "inverted" data. Some results of this chapter have been presented in Teodor et al. (2018b), Teodor et al. (2019) and Teodor et al. (2020 - submitted).

8.2 – The general FWI workflow

A simple FWI workflow was adopted in the first place, and its complexity was gradually increased when necessary. As data processing before inversion, a 5th order Butterworth filter, with corner frequencies of 3 Hz and 40 Hz, was applied (e.g. Figure **8.1**).

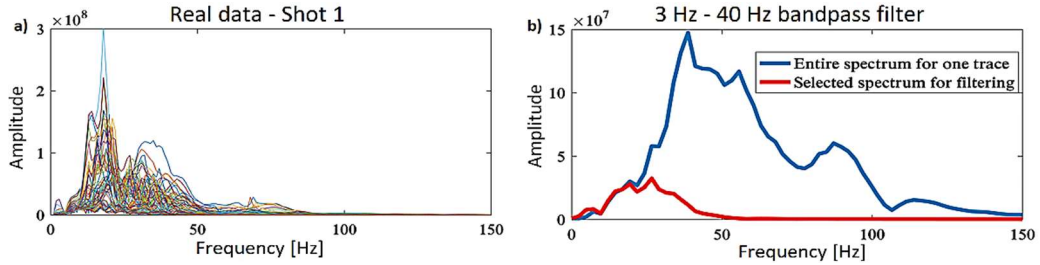

*Figure **8.1** – Amplitude spectra for real data **a**) Example of amplitude spectra for all traces belonging to one shot gather. **b**) Example of the amplitude spectra of a trace before (blue) and after (red) filtering.*

Some noisy traces, close to the shot positions (offset from 0.3 m to 0.6 m), have been muted (Figure **8.2**) to avoid their contribution in the inversion process.

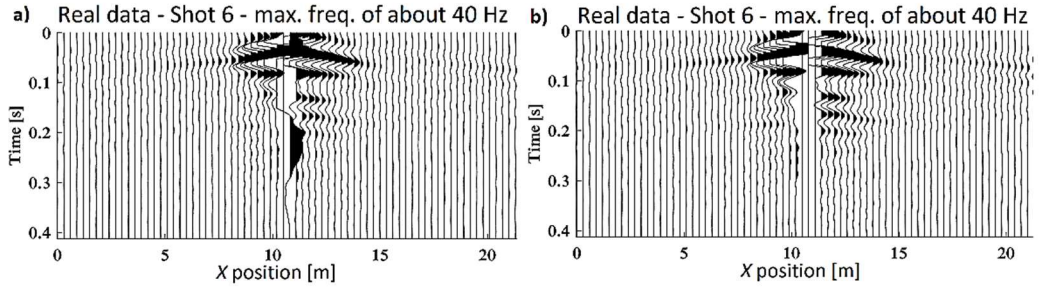

*Figure **8.2** – Example of filtered common shot gather, before **a**) and after **b**) muting.*

8.2.1 – Modeling parameters

According to the space-sampling criterion for SEM modeling and considering the maximum frequency of 40 Hz, a cell size of 1 m was used, equal in all directions of the 3D mesh. The time stability criterion required a sampling step of 4.8e-05 s, leading to 8600 samples in

time for a total duration of the simulation of 0.4128 s. 8 shots were considered for the gradient computation and inversion (1, 2, 4, 5, 6, 9, 10, and 11 in Figure **8.3**).

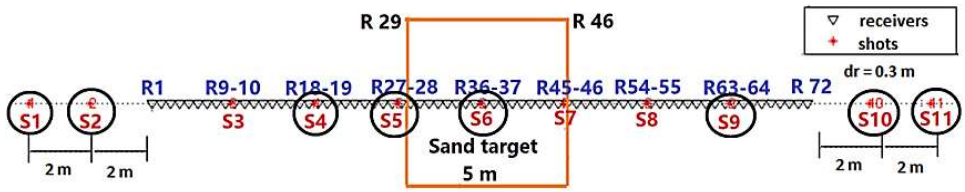

*Figure **8.3** - Selected shots (indicated with black circles) for FWI on the 2D data set.*

As no significant cycle-skipping is present in the initial model, all FWI tests were performed in the full frequency band of 3 Hz – 40 Hz and the same source function was used for all iterations. The value of the upper-frequency band (40 Hz) chosen for inversion is due to the available computational resources.

The source time function for the synthetic tests is a Ricker wavelet, with a central frequency of 16 Hz (maximum frequency of about 40 Hz) and a maximum amplitude of 1e+06. For the field data application, the real source was estimated by deconvolution (Pratt, 1999), after applying the 3 Hz – 40 Hz band-pass filter. The main propagating wavelength inside the investigated models is about 16 m for V_P and 10 m for V_S. Consequently, the expected main FWI resolution is about 8 m for V_P and 5 m for V_S. The average resolution for V_S is similar to the dimensions of the investigated target. Of course, a higher resolution is expected in the shallow part of the model (which has lower velocities) than in the deeper part (where higher velocities are registered).

8.2.2 – Offset weighting, boundary constraints and gradient smoothing

No data normalization was applied, except for an offset-variable data weighting. The aim of this data strategy is guaranteeing a similar contribution for each receiver over the entire seismogram. Since almost all the energy tends to remain trapped inside the low-velocity target (Figure **8.4a**), the amplitude of the far-offset traces was enhanced, up to the same magnitude order of the near-offset traces (Figure **8.4b**), to increase the contribution of the far-offset traces in the optimization process.

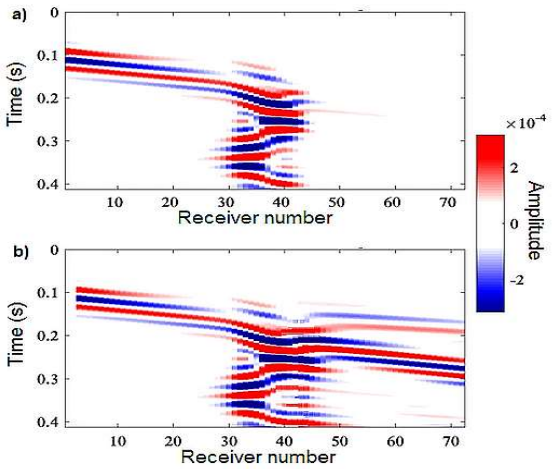

*Figure **8.4**. Reference synthetic data (Shot 1). **a**) Not normalized seismogram before applying the offset-variable weighting function. **b**) Not normalized seismogram after the application of the weighting function to the far-offset traces.*

Some experiments related to the effectiveness of the data weighting function are presented in Appendix 4.

To avoid a possible non-physical model update during inversion, boundary constraints for the V_P and V_S parameters have been used. In particular, the maximum V_P and minimum V_S was fixed to 1000 m/s and 60 m/s, respectively, while depth-variable constraints were imposed for the minimum V_P and maximum V_S variation (e.g. Figure **8.5**). The depth-variable constraints have been inferred from the initial models, by applying a vertical gradient to a scaling factor based on the depth-by-depth difference between V_P and V_S.

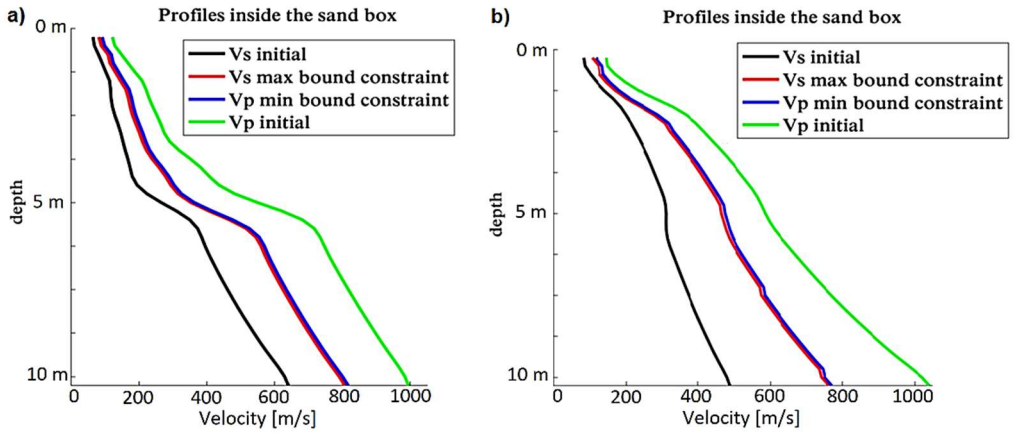

*Figure **8.5** – Example of 1D profiles belonging to the 3D variable boundary constraints on model parameters (maximum V_S and minimum V_P): **a)** synthetic case, **b)** field data application.*

No preconditioning or regularization was used, other than gradient smoothing. Analyzing Figure **8.6**, small-wavelength artefacts can be noticed in the gradient, not in agreement with the FWI resolution allowed by the frequency content of the data (Figure **8.6a** and **8.6b**). To control the wavenumber content of the updated models according to the resolution limits, the gradient was smoothed (Figure **8.6c** and **8.6d**) through a double application of the anisotropic Bessel filter described in Section 5.4 (Trinh et al., 2017).

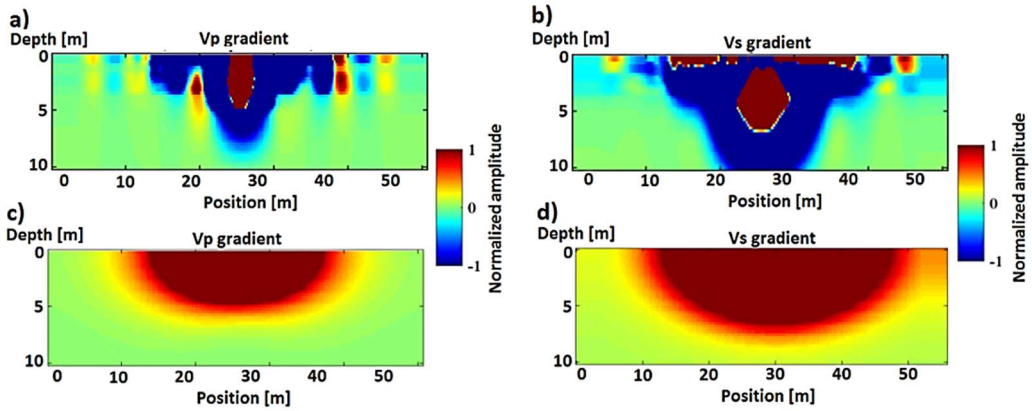

*Figure **8.6** – Example of 2D section (crossing the target's center) of the 3D FWI gradients: **a)** V_P gradient before smoothing, **b)** V_S gradient before smoothing, **c)** V_P gradient after smoothing **d)** V_S gradient after smoothing.*

The choice for the filter's shape was based on the mesh dimension (1 m) for the vertical direction, wavelength resolution (λ/2) for the in-line direction, and acquisition design (2D) for the crossline direction. The filter has the following lengths in the *z*, *x* and *y* directions: 1.5 m, 5 m and 12 m, respectively. No rotation was applied here, although one may potentially design specific Bessel filters when considering FWI first attempts. Such a strategy could be investigated in the future. Some details on the experiments conducted for the choice of the Bessel filter parameters are shown in Appendix 5.

8.2.3 – Inversion parameters

During FWI, the V_P and V_S were considered simultaneously, while the density was kept constant (1800 kg/m^3). To focus the analysis on the velocity model update, constant values were also used for the attenuation coefficients in the forward modeling step of the FWI (Q_P = Q_S = 40), and these parameters were kept fixed during inversion.

The inversion scheme was based on the quasi-Newton l-BFGS (Broyden-Fletcher-Goldfarb-Shanno) method (Nocedal and Wright, 2006; Métivier and Brossier, 2016). The double level of MPI parallelization of the SEM code was exploited, by decomposing the 3D volumes into 2 x 4 x 2 sub-domains (in *z*-*x*-*y*-direction, respectively).

All shots and decomposed domains have been distributed over 128 cores (8 nodes with 16 cores each) of the HPC UGA Froggy architecture. The inversion stopped when the value of the misfit function no longer decreased for more than two consecutive iterations. The necessary time for the first gradient computation was about 11 minutes, while each FWI test took about 12 hours. The convergence criterion was reached after 10 – 25 iterations. Each FWI test required an available computer memory of about 9.8 GB.

Table 8.1 shows a synthesis of the modeling and inversion parameters:

Butterworth filter (5th order)	**3** Hz – **40** Hz
Trace muting	offset from **0.3 m** to **0.6 m**
Mesh dimension	**1 m**: the same in the x-y and –z direction
Time sampling	**4.8e-05 s**
Time iterations	**8600**
Acquisition duration	**0.4128 s**
Number of channels and shots	**72** (vertical) channels and **8** shots
Data weighting	offset variable

V_P upper boundary	**1000 m/s**
V_P lower boundary	variable
V_S upper boundary	variable
V_S lower boundary	**60 m/s**
Gradient smoothing: coherent lengths for the Bessel filter in the z – x – y direction	**1.5 m – 5 m – 12 m**
Density (constant)	**1800 kg/m^3**
Quality factors (Q_P=Q_S)	**40**
Inversion method	quasi-Newton **l-BFGS**
Decimation ratio for boundary wavefield saving	**86**
CARFS – nunber of saved checkpoints	**13**
Instability tolerance	**0.001**
Domain decomposition in the z – x –y direction	**2 – 4 – 2** subdomains
Parallel computing: number of nodes and cores	**8** nodes **x 16** cores **= 128 cores**
Necessary time for gradient computation	about **11 minutes**
Necessary time for one FWI test	about **12 hours**
Necessary computer memory for one FWI test	**9.8 GB**

Table 8.1 – Parameters used during FWI for the 2D acquisition layout.

8.2.4 – Synthetic example

Figures **8.7a** and **8.7b** show 2D sections of the 3D V_P and the V_S models obtained after FWI for the single-DC case. The target is reconstructed with low-resolution in the V_S model, while the V_P model is not significantly modified after inversion. Figures **8.7c** and **8.7d** show the final models, obtained when starting the FWI from the full-DC initial models. The results show a better resolution compared with the results of the single-DC case and again, V_S is updated better than V_P.

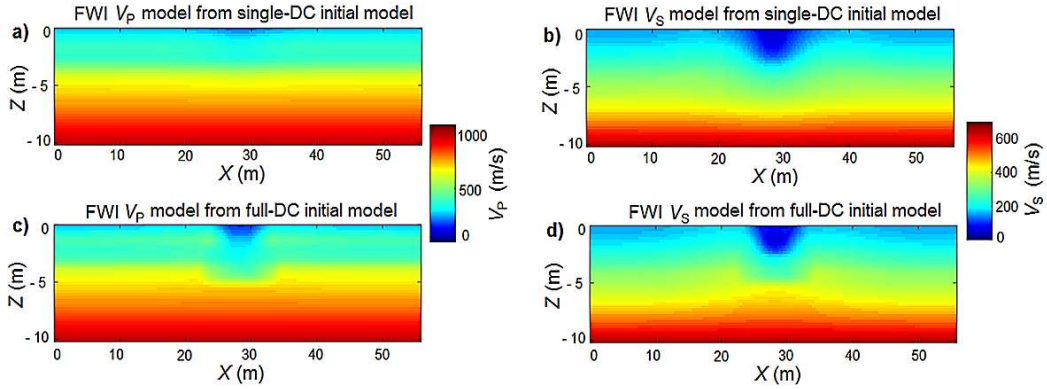

*Figure **8.7** – Vertical 2D sections, crossing the low-velocity target, of the 3D final models for the synthetic application. **a**) and **b**): V_P and V_S models obtained after FWI, using the initial models from the single-DC analysis. **c**) and **d**): V_P and V_S models obtained after FWI, using the initial models from the full-DC analysis.*

Figure **8.8** shows the normalized model misfit (i.e. normalized difference) before and after FWI, for the single-DC case. While V_S improves (Figure **8.8d**), the V_P improvement is limited (Figure **8.8c**). Figure **8.8d** reveals an overall low misfit value at the target's position after FWI, except for some boundary areas, where V_S is underestimated.

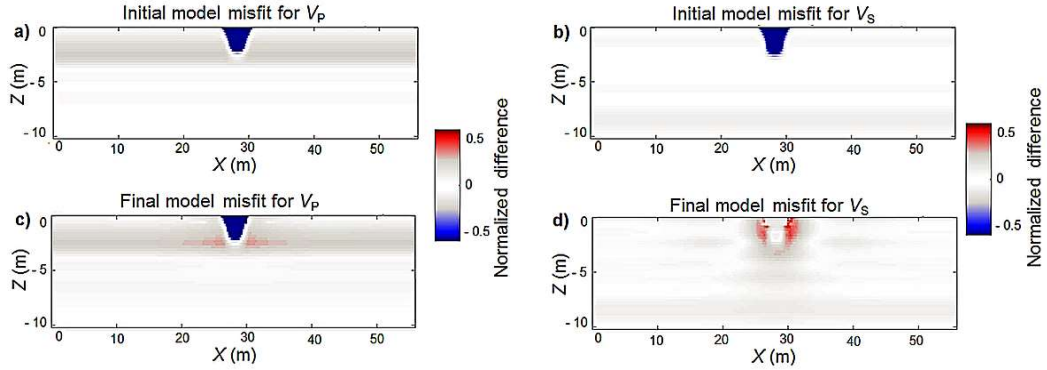

*Figure **8.8** – Model misfit for the single-DC synthetic case, computed as normalized difference: [(Reference model - Initial or Final model) / Reference model]. **a**) Initial model misfit for V_P. **b**) Initial model misfit for V_S. **c**) Model misfit for V_P after inversion. **d**) Model misfit for V_S after inversion.*

Figure **8.9** shows the normalized model misfit (i.e. the normalized difference), before and after FWI, for the full-DC case. V_S is better reconstructed (Figure **8.9d**) compared with the final V_S of the single-DC case. The misfit value in correspondence of the target is very low. The target's boundaries are better defined, although a very slight underestimation in velocity is still present. V_P is better recovered (Figure **8.9c**) than in the single-DC case, although the V_P model misfit in correspondence of the target is still high after inversion.

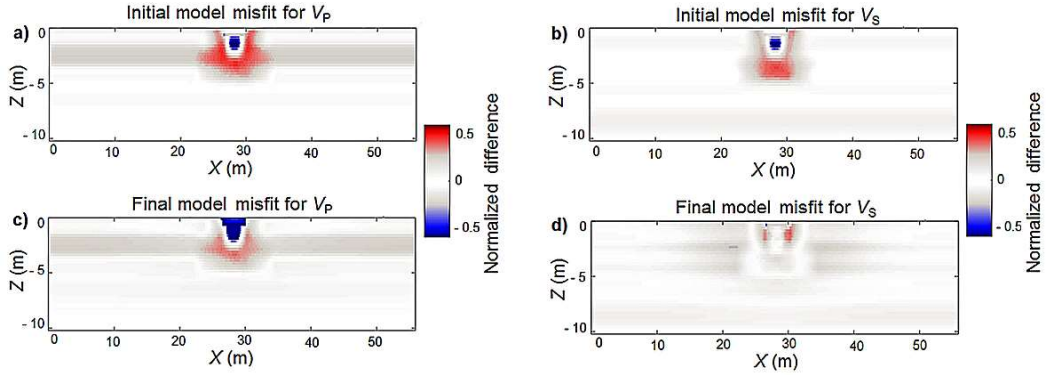

*Figure **8.9** – Model misfit for the full-DC synthetic case, computed as normalized difference: [(Reference model - Initial or Final model) / Reference model]. **a)** Initial model misfit for V_P. **b)** Initial model misfit for V_S. **c)** Model misfit for V_P after inversion. **d)** Model misfit for V_S after inversion.*

Figure **8.10** shows an example of the trace-by-trace data fitting comparison, for a far offset, before (**8.10a**) and after FWI (**8.10b**) on the full-DC case, while Figure **8.11** shows an example of data fitting comparison, before (**8.11a**) and after FWI (**8.11b**), for a near-offset shot on the full-DC case. In the far-offset configuration, the fitting is already accurate before inversion for all traces close to the shot positions. One can notice a data fitting improvement after FWI, especially in correspondence of the target. Still, the back-scattered phases, close to the target's boundaries, and some far-offset arrivals, are not properly fitted yet.

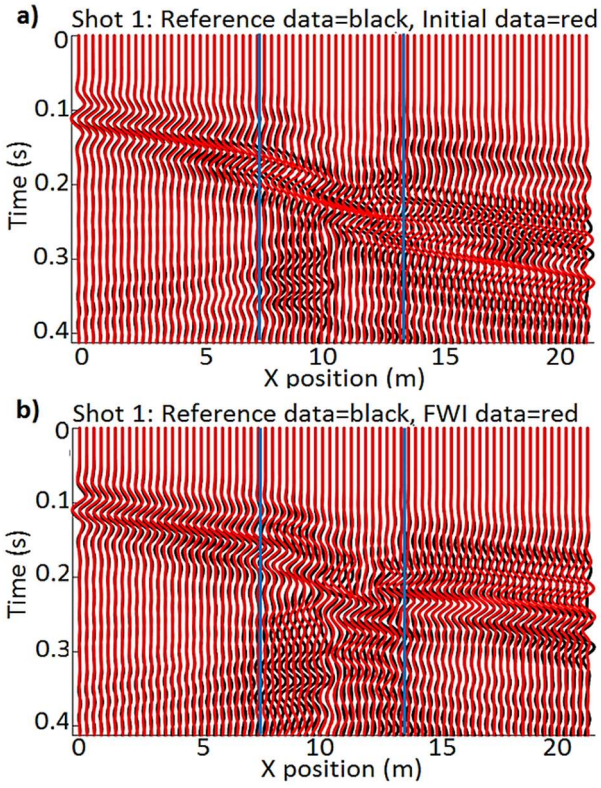

*Figure **8.10** – Example of data fitting comparison for the full-DC case, far offset, synthetic application. **a)** Comparison between reference data (in black) and initial data (in red). **b)** Comparison between the reference data (in black) and inverted data (in red). The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

For the near-offset configuration, all arrivals are properly fitted after FWI, both in correspondence of the target and far from the source position (Figure **8.11b**).

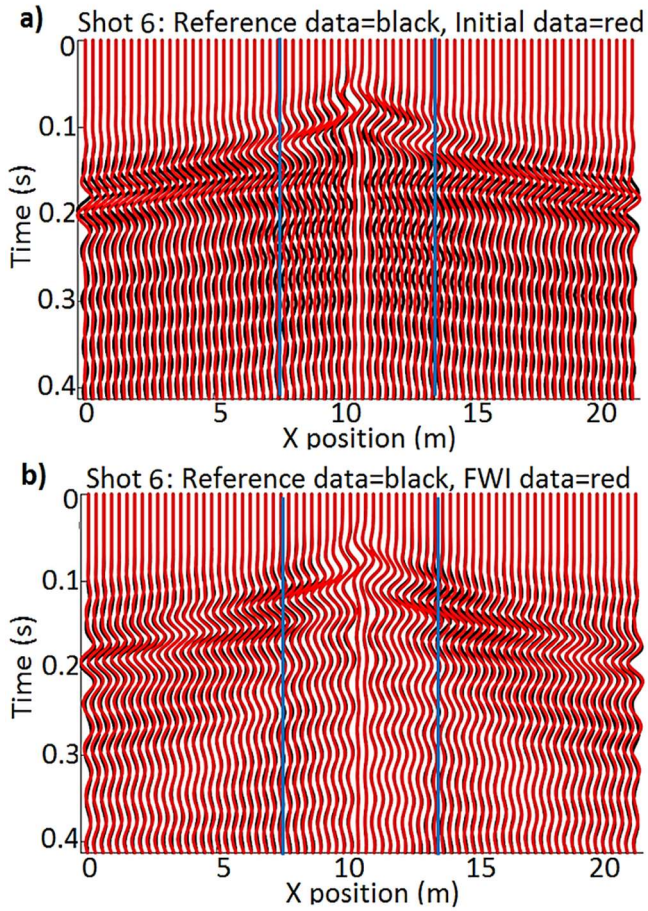

*Figure **8.11** – Example of data fitting comparison for the full-DC case, near offset, synthetic application. **a)** Comparison between reference data (in black) and initial data (in red). **b)** Comparison between reference data (in black) and inverted data (in red). The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

8.2.5 – Real data application

The same FWI workflow, designed on synthetic data, was applied to the field data set. Figures **8.12a** and **8.12b** display 2D sections of the 3D V_P and the V_S models after FWI for the single-DC case. The target becomes distinguishable both in the V_P and V_S models, although V_S is reconstructed better than V_P. Figures **8.12c** and **8.12d** show the final V_P and the final V_S model for the full-DC case. The target is already well defined in the initial models and the resolution related to a maximum 40 Hz FWI configuration does not allow for significant further improvement.

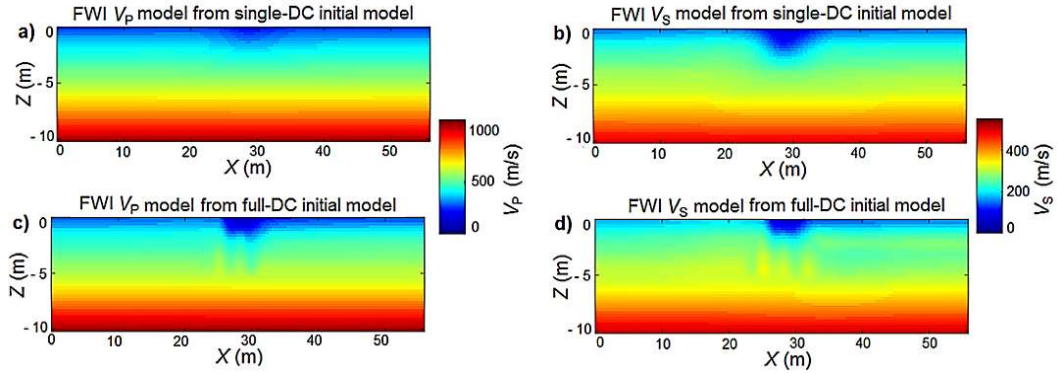

*Figure **8.12** – Vertical 2D sections, crossing the low-velocity target, of the 3D models obtained after FWI on real data: **a**) Final V_P model for the single-DC case, **b**) final V_S model for the single-DC case, **c**) final V_P model for the full-DC case and **d**) final V_S model for the full-DC case.*

The normalized difference between the final and initial models related to the single-DC case indicates a velocity decrease in correspondence of the target, both for V_P (Figure **8.13a**) and V_S (Figure **8.13b**), although the V_S resolution is higher than the resolution of V_P.

The normalized difference between the final and initial model for the full-DC case reveals a very little velocity decrease in correspondence of the target, both for V_P (Figure **8.13c**) and V_S (Figure **8.13d**). The model is not modified outside the target after FWI since the layered structure of the initial model from DC analysis already reproduces correctly the velocity variation of the site.

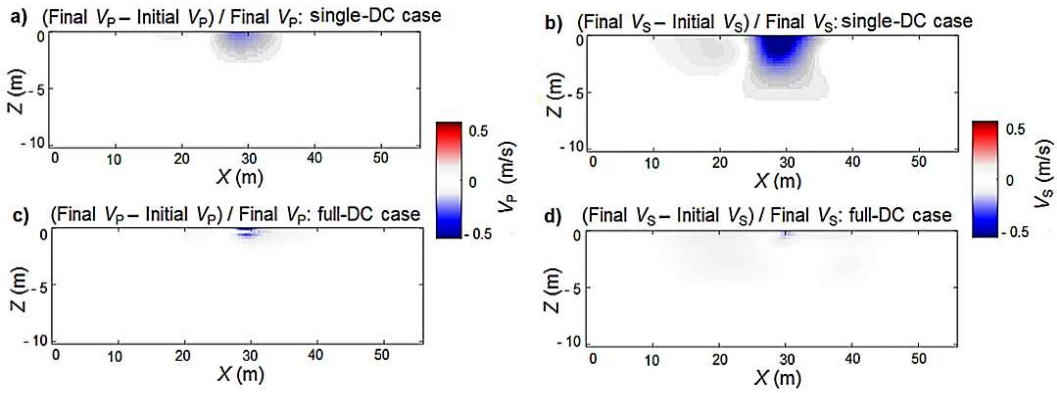

*Figure **8.13** – Normalized difference between the final and initial models for the real data application: **a**) V_P difference, single-DC case, **b**) V_S difference, single-DC case, **c**) V_P difference, full-DC case, **d**) V_S difference, full-DC case.*

Figures **8.14a** and **8.14b** show a data fitting comparison, before and after FWI, for the single-DC case, while Figures **8.15a** and **8.15b** show an example of data fitting comparison, before and after FWI, for the full-DC case. For the single-DC case, the data-fitting improvement is evident at the sand target position. Nevertheless, some far-offset arrivals are still not properly fitted.

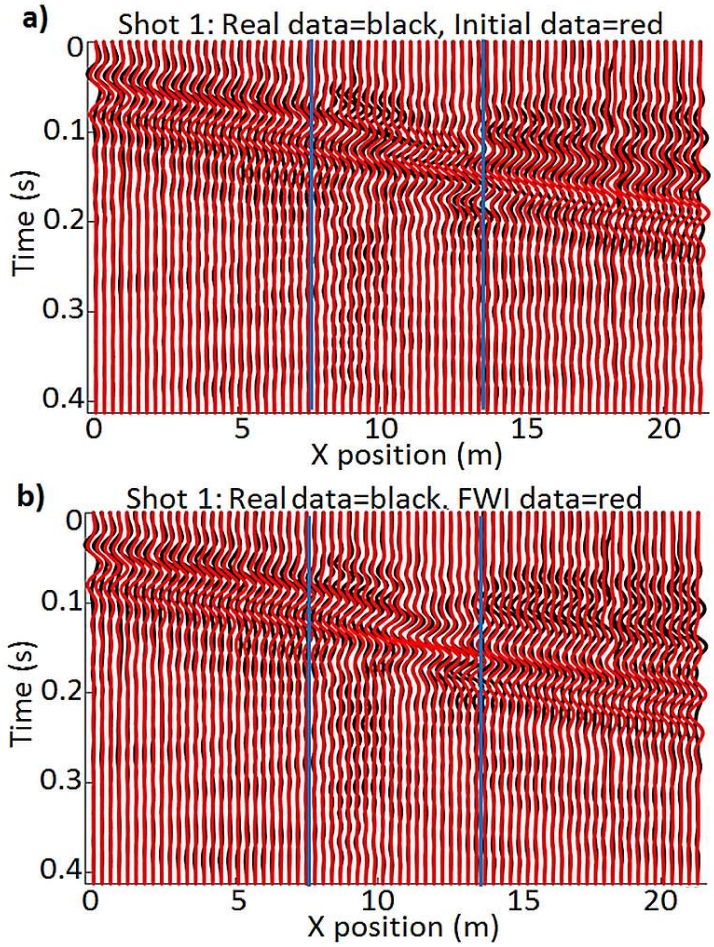

*Figure **8.14** – Example of data fitting comparison for far offset, field data application, single-DC case: **a**) Real data (in black), initial data (in red), **b**) real data (in black), inverted data (in red). The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

For the full-DC case, one can notice an already accurate data fitting in the initial model (Figure **8.15a**) at the target's position. Consequently, FWI does not bring significant changes (Figure **8.15b**). Besides, the proper fitting of some far-offset arrivals and back-scattered phases may require more elaborated FWI strategies or higher frequencies. Howsoever, the smooth character of the initial model straitens the consented fitting of the small-wavelength phases scattered from the target's boundaries.

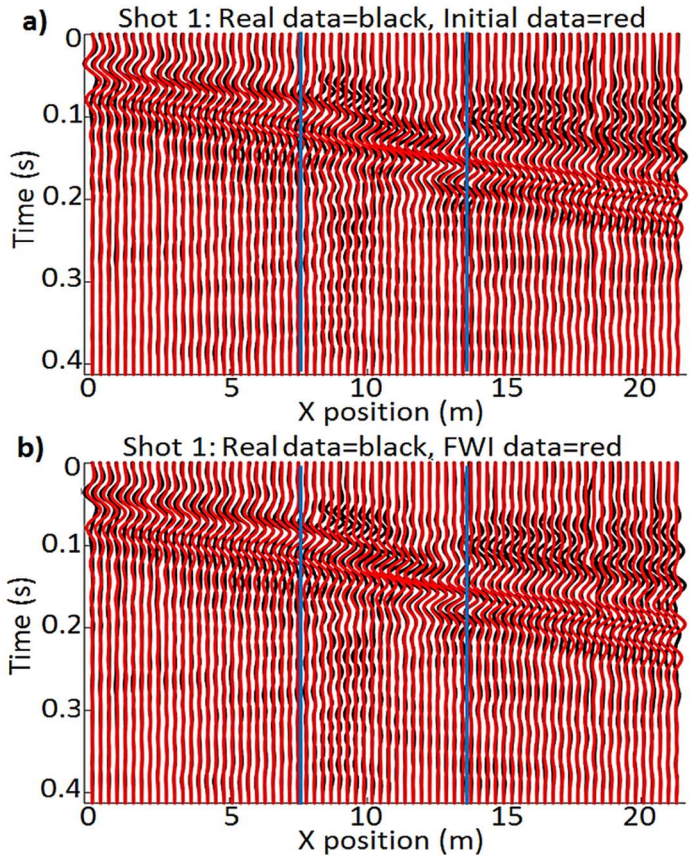

*Figure **8.15** – Example of data fitting comparison for far offset, field data application, full-DC case: **a**) Real data (in black), initial data (in red), **b**) real data (in black), inverted data (in red). The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

Figure **8.16** shows the difference between the final models of the single-DC case and the final models of the full-DC case. Figure **8.16a** reveals an overestimation of V_P in correspondence of the target for the single-DC case compared with the full-DC case, while the difference between the final single-DC V_S model and the final full-DC V_S model is small (Figure **8.16b**).

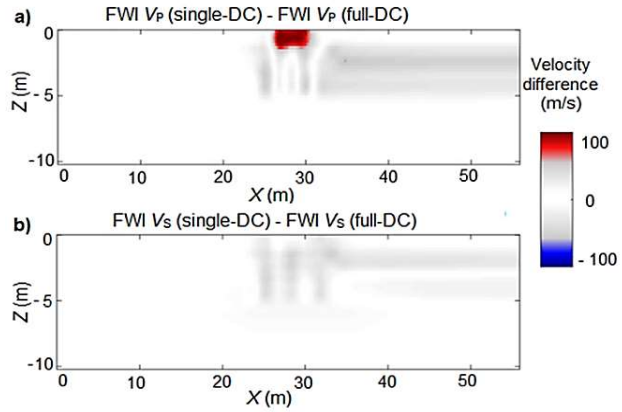

*Figure **8.16** – Vertical 2D sections, crossing the low-velocity target, of the 3D velocity difference between the final models for the single-DC case and the final models for the full-DC case, real data application: **a**) V_P difference **b**) V_S difference.*

8.3 – A more elaborated FWI workflow

The improvement of the model reconstruction and data fitting at far-offset may require the use of some model-oriented and data-based strategies. In this section, some examples of such strategies (i.e. parameter selection, data windowing and multiscale FWI) are presented for the synthetic data set. The FWI tests are based on the initial models retrieved from the full-DC analysis.

8.3.1 – Model strategy: preliminary mono-parametric FWI

This model strategy aims at exploiting the higher SW sensitivity to the shear properties, through a preliminary monoparametric FWI (Teodor et al., 2018b; Teodor et al., 2019; Teodor et al., 2020 - submitted). The strategy consists in two sequential steps: monoparametric FWI (with respect to V_S), followed by multi-parameter (V_P+V_S) FWI. In the preliminary step, the V_P from SW analysis is kept invariable.

Figure **8.17a** displays the initial V_P model from full-DC analysis, while Figure **8.17b** shows the V_S model obtained after the monoparametric FWI step. Figures **8.17c** and **8.17d** show the V_P and V_S models obtained after the multi-parameter FWI. While the V_P model is not significantly improved (as expected), the resolution of the V_S model improves significantly, especially in correspondence of the target boundaries.

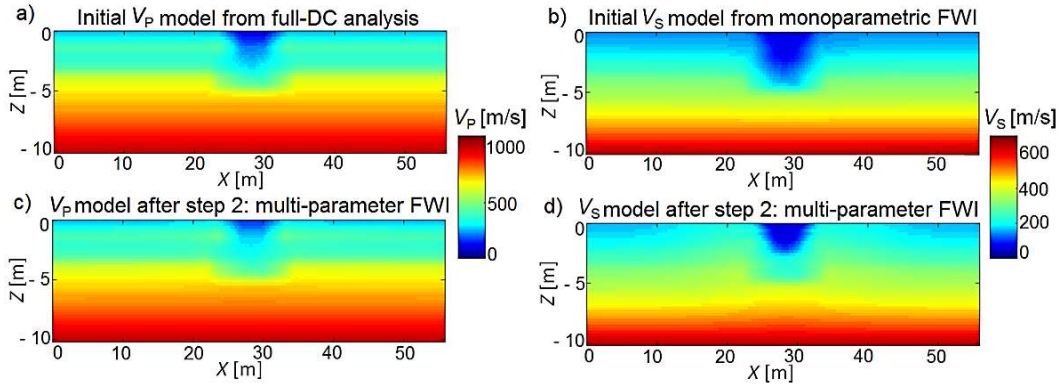

*Figure **8.17** – 2D sections (crossing the target's center) of the 3D velocity models. **a**) Initial V_P model from the full-DC analysis. **b**) Initial V_S model from monoparametric FWI. **c**) Final V_P model after multi-parameter FWI. **d**) Final V_S model after multi-parameter FWI. An offset variable weighting function was also applied.*

This aspect can be noticed as well when analyzing the normalized model misfit reported in Figure **8.18**. Specifically, the V_S model reconstructed after the monoparametric FWI is accurate only down to 2 m (Figure **8.18c**). Better results, for the deeper reconstruction of the V_S model, are obtained only after performing the multi-parameter FWI (Figure **8.18d**). Therefore, a relatively correct initial V_P model (although smooth), retrieved from SW analysis, and an accurate shallow V_S model, retrieved from monoparametric FWI, efficiently integrate each other in the multi-parameter FWI scheme.

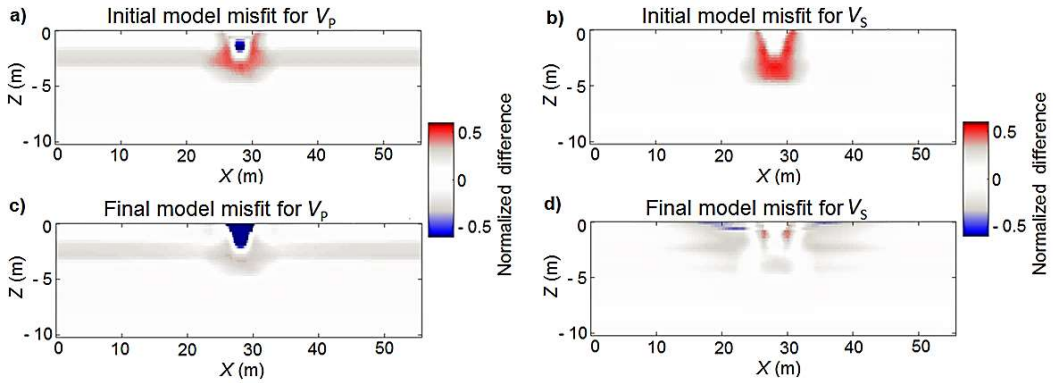

*Figure **8.18** – 2D sections (crossing the target's center) of the 3D model misfit: **a)** (Reference V_P model – Initial V_P model from DC)/Reference V_P model. **b)** (Reference V_S model – Initial V_S model from monoparametric FWI)/Reference V_S model. **c)** (Reference V_P model – Final V_P model after multi-parameter FWI)/Reference V_P model. **d)** (Reference V_S model – Final V_S model after multi-parameter FWI)/Reference V_S model.*

Figure **8.19** presents a comparison between the V_S model-update when no model strategy is employed, and the V_S model-update when using a preliminary monoparametric FWI. One can notice, in Figure **8.19a**, an overestimation in velocity after FWI (black curve), down to 0.5 m, whereas the shallow velocity is accurately estimated when using the model strategy (figure **8.19b**, black curve).

Moreover, one can notice, in Figure **8.19b**, how the monoparametric FWI provides accurate velocity estimation only down to 2 m (red curve). The velocity estimation below 2 m improves only after the multi-parameter FWI step (black curve).

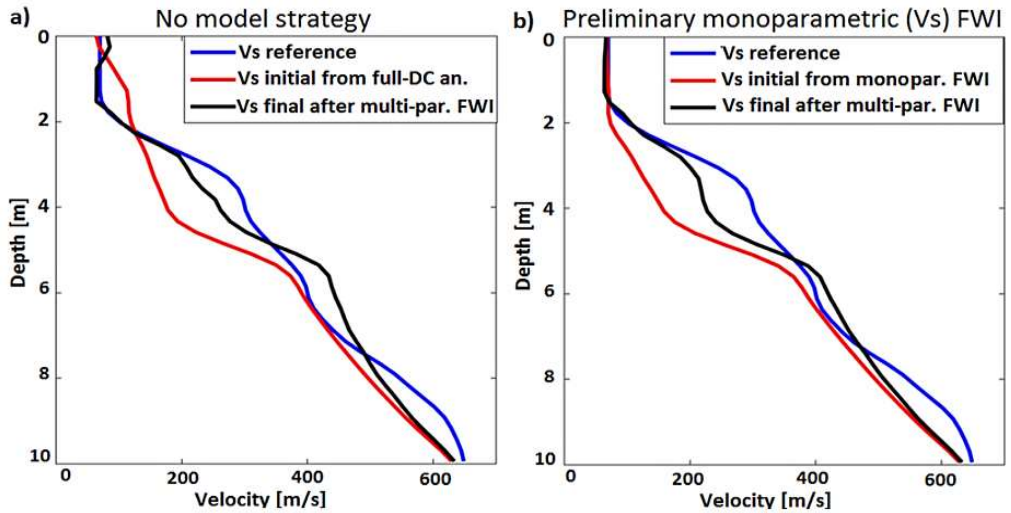

*Figure **8.19** – 1D profiles (in correspondence of the target) of the 3D S-wave velocity models. **a)** In this case, the initial V_S model from the full-DC analysis was used for the multi-parameter FWI. **b)** In this case, the initial V_S model from the mono-parameter FWI was used for the multi-parameter FWI.*

8.3.2 – Data strategy: time-windowing

The time-windowing strategy aims at ensuring a better fitting of the far-offset arrivals and potential better reconstruction of the V_P model, as far as FWI can efficiently exploit the BWs in a preliminary stage (EAWI=Early Arrival Waveform Inversion). The procedure consists in starting the inversion from a time window that contains only early arrivals (bearing more P-wave signature) and introducing the entire wavefield in a second step. Such a strategy was experimented in other studies (e.g., Trinh et al., 2018; He et al., 2018; Trinh et al., 2019). Previously, Brossier et al. (2009) proposed to incorporate gradually the entire seismogram in the inversion process by applying a data weighting strategy consisting of an exponentially decay time windowing around the first break.

Figure **8.20a** shows an example of reference data used in the first step (trace muting after 0.06 s in correspondence of the receiver 1 at offset = 4 m and after 0.10 s in correspondence of the receiver 72 at offset = 25.3 m). Figure **8.20b** shows the entire time window of the reference data used in the second step: BWs are indistinguishable because of the high amplitude of SWs.

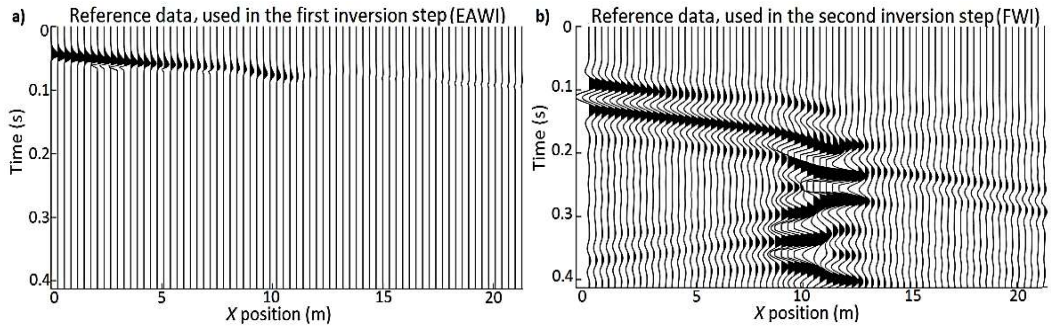

*Figure **8.20** – Example of reference data belonging to the Shot 1: **a**) Data used during the first FWI step (only BWs are present), **b**) Data used during the second FWI step (SWs are very energetic and "cover" the BWs amplitude). A weighting function was applied to the far-offset traces during FWI.*

Figures **8.21a** and **8.21b** show the V_P and V_S model, respectively, obtained after the first inversion step (EAWI), while Figures **8.21c** and **8.21d** show the corresponding models obtained after EAWI+FWI. The reader can notice an improvement of the target's resolution in the V_S model after the second inversion step, ensured by the introduction of SWs information, while the V_P model improvement is still not substantial.

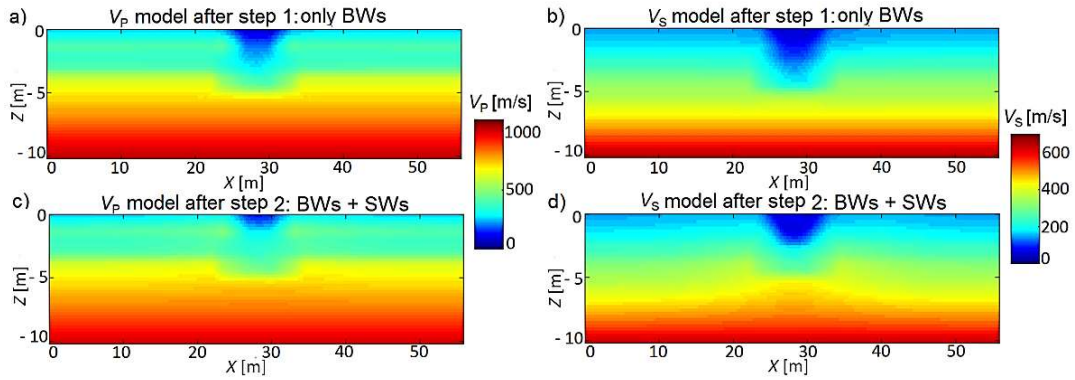

*Figure **8.21** – 2D sections (crossing the target's center) of the 3D velocity models. **a**) Initial V_P model after the first time-windowing step. **b**) Initial V_S after the first time-windowing step. **c**) Final V_P model after two-steps FWI. **d**) Final V_S model after two-steps FWI.*

Figure **8.22** shows the initial and final model misfit, computed as the normalized difference between the reference model and the initial or final (FWI) model, respectively. A general improvement of the V_S model can be noticed after FWI (Figure **8.22d**), except for some localized areas in correspondence of the target boundaries, where there is a slight underestimation in velocity. Differently, the target's reconstruction improvement is not so evident in the V_P model. Anyhow, the final V_P model misfit is, in this case, lower (close to zero) in correspondence of the very shallow part of the target, compared with the case when no trace windowing is applied (Figure **8.9c**). However, the overestimation in velocity still occurs in the deeper region. The V_P model update is still relatively inaccurate since the BWs have very weak amplitude.

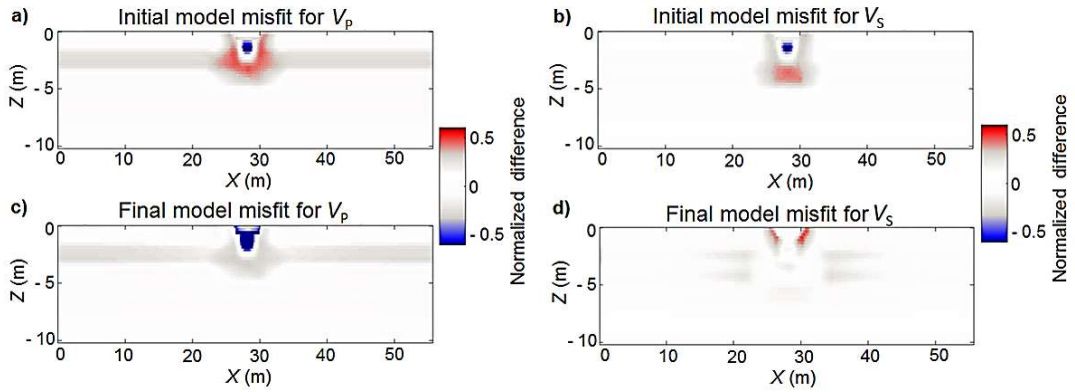

*Figure **8.22** – 2D sections (crossing the target's center) of the 3D model misfit: **a**) (Reference V_P model – Initial V_P model)/Reference V_P model. **b**) (Reference V_S model – Initial V_S model)/Reference V_S model. **c**) (Reference V_P model – Final V_P model)/Reference V_P model. **d**) (Reference V_S model – Final V_S model)/Reference V_S model.*

Anyhow, the EAWI+FWI strategy efficiently ensures better data fitting for the far-offset arrivals, since BWs investigate the deeper part of the model. This fact can be noticed when comparing the data fitting for FWI performed directly for the entire time-window (Figure **8.23a**), with the data fitting obtained when an EAWI step precedes FWI (Figure **8.23b**).

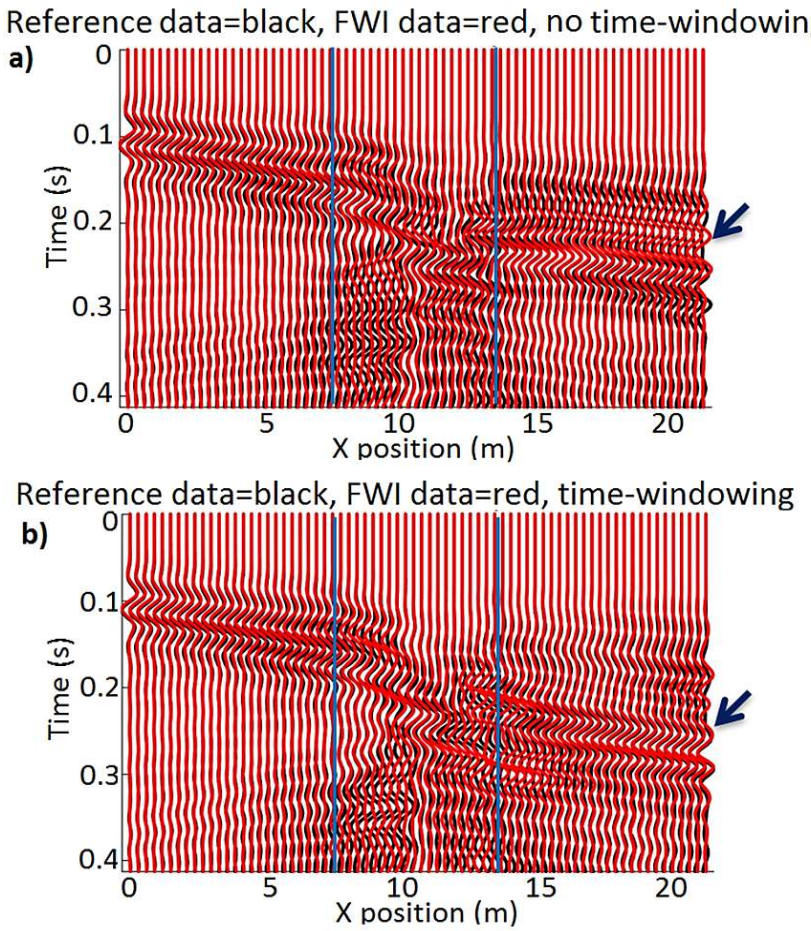

*Figure **8.23** – Trace-by-trace data fitting comparison between reference data and data belonging to the model obtained after FWI, for the full-DC case, 2D acquisition, synthetic example (Shot 1). **a**) FWI is conducted directly over the entire time-window. **b**) A two-step time windowing FWI strategy is applied. The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

8.3.3 – Data strategy: multiscale FWI

The multiscale FWI is a routine procedure (e.g. Bunks et al., 1995), based on the reconstruction of the large wavelengths of the medium in the first stages, followed by the progressive reconstruction of the small wavelengths, in the later stages. Therefore, only low frequencies are considered in the first inversion runs, while the higher frequencies are introduced in the later inversion runs.

All FWI tests presented in Section 8.2 were performed in the full-frequency band because no significant cycle-skipping was identified in the initial model. Here, a multiscale approach was also experimented, on synthetic data, to investigating whether it provides better results.

In particular, 4 inversion runs were conducted: 3 Hz – 10 Hz, 3 Hz – 20 Hz, 3 Hz - 30 Hz and 3 Hz – 40 Hz. Figure **8.24** synthetizes the characteristics of the source wavelet and Bessel's filter coherent lengths in the z-x and y direction, for each inversion run.

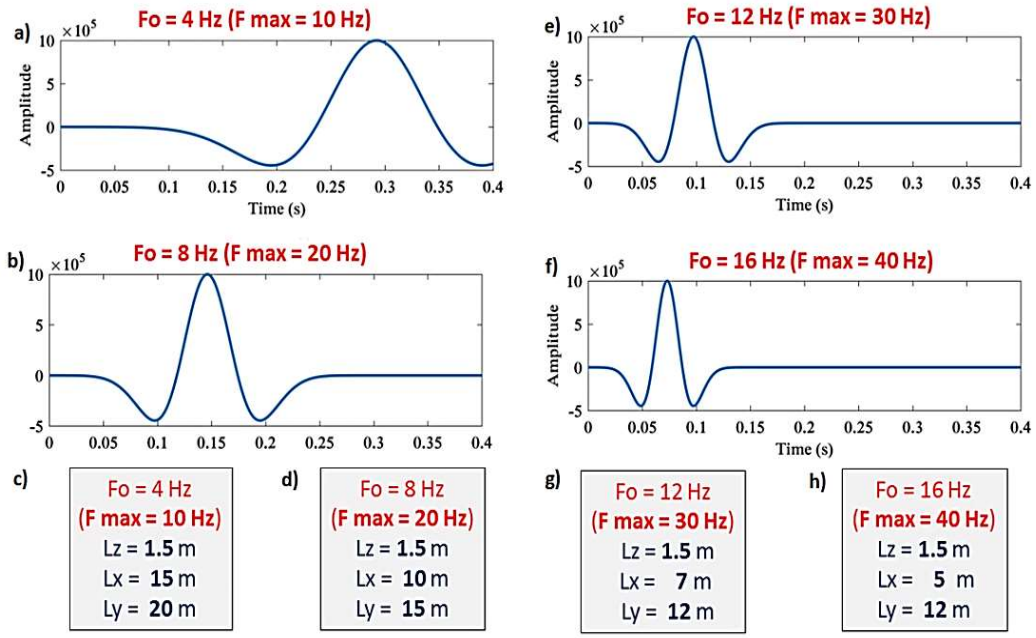

*Figure **8.24** – Source time functions and parameters for the Bessel smoothing filter of each inversion run. **a**) Wavelet used for the first inversion run (maximum 10 Hz). **b**) Wavelet used for the second inversion run (maximum 20 Hz). **c**) Gradient smoothing parameters for the first inversion run. **d**) Gradient smoothing parameters for the second inversion run. **e**) Wavelet used for the third inversion run (maximum 30 Hz). **f**) Wavelet used for the fourth inversion run (maximum 40 Hz). **g**) Gradient smoothing parameters for the third inversion run. **h**) Gradient smoothing parameters for the fourth inversion run.*

The first inversion run started from the initial models obtained by full-DC analysis. Successively, the models obtained at the end of each inversion run were used as initial models for the next iteration.

The results are shown in Figure **8.25**. As expected, the models reconstructed after the last frequency run are not very different from the ones obtained when performing FWI in the full-frequency band (Figures **8.7c** and **8.7d**). Only a slightly better reconstruction is noticed, both for V_P (in the very shallow part, down to 1m) and V_S.

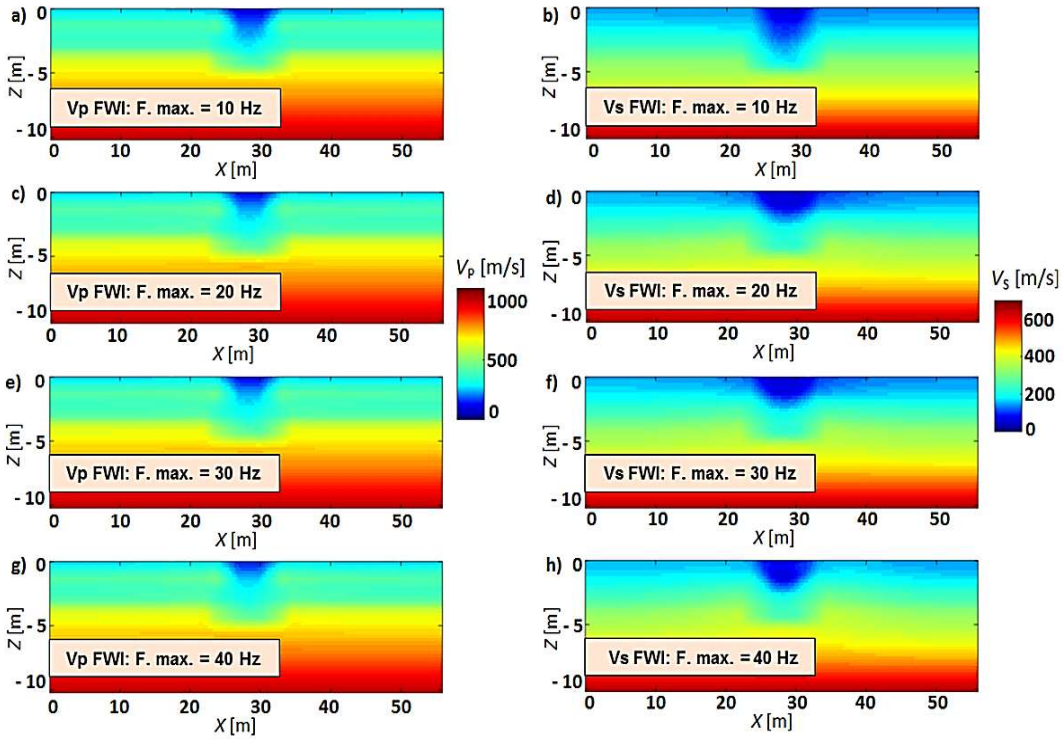

*Figure **8.25** – Velocity models obtained after FWI: **a)** V_P – maximum frequency = 10 Hz, **b)** V_S – maximum frequency = 10 Hz, **c)** V_P – maximum frequency = 20 Hz, **d)** V_S – maximum frequency = 20 Hz, **e)** V_P – maximum frequency = 30 Hz, **f)** V_S – maximum frequency = 30 Hz, **g)** V_P – maximum frequency = 40 Hz, **h)** V_S – maximum frequency = 40 Hz.*

Figure **8.26** shows the model misfit, computed as the normalized difference between the reference model and the model obtained after each inversion run. It can be noticed how the main improvements of the reconstructed models occur after the second inversion run (3 Hz – 20 Hz) and after the last inversion run (3 Hz – 40 Hz).

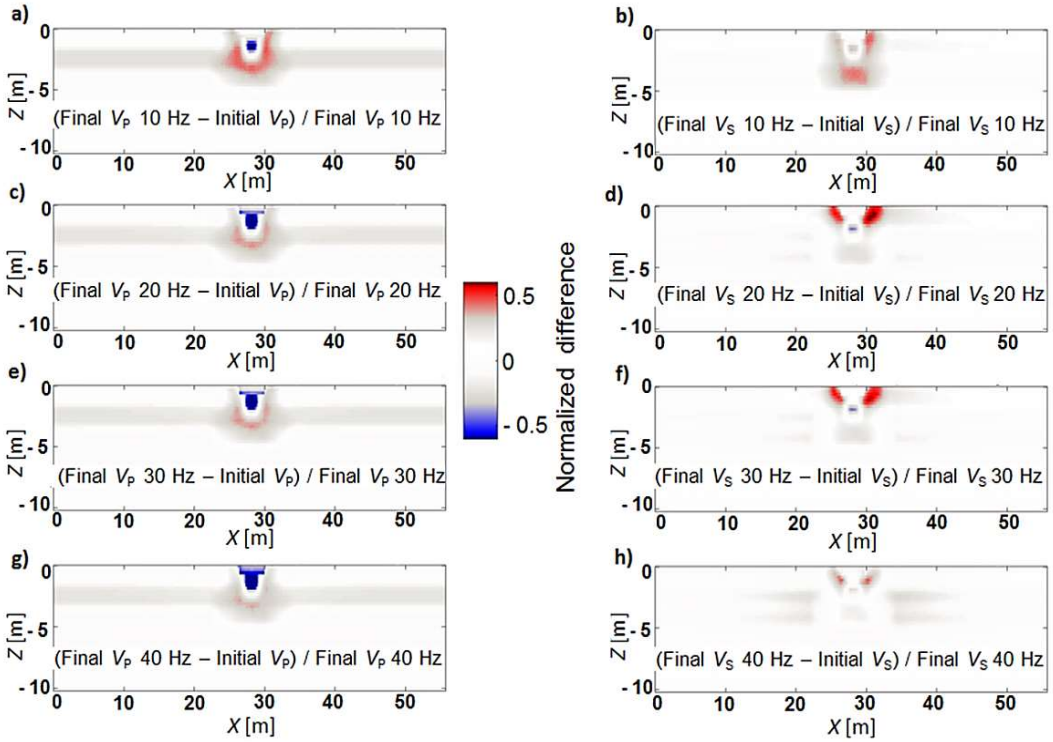

*Figure **8.26** – 2D sections, crossing the target, of the 3D normalized model misfit [(Final model-Initial model)/Final model], for each inversion run.*

In figure **8.27**, showing a trace-by-trace data fitting comparison before and after the first inversion run (3 Hz – 10 Hz), an accurate data fitting can be noticed.

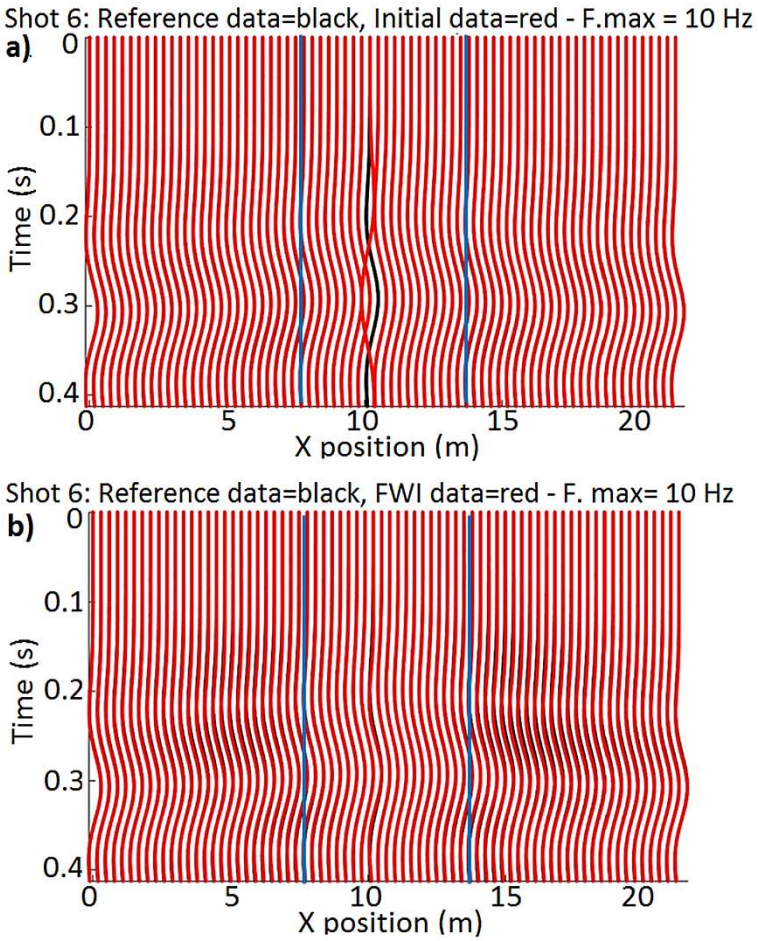

*Figure **8.27** – Trace-by-trace data fitting comparison before **a)** and after **b)** FWI, synthetic example, full-DC case, shot 6 (near-offset), multiscale approach, maximum **10 Hz** inversion run. The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

In figure **8.28**, showing a trace-by-trace data fitting comparison before and after the second inversion run (3 Hz – 20 Hz), a significant improvement in data fitting can be noticed after FWI.

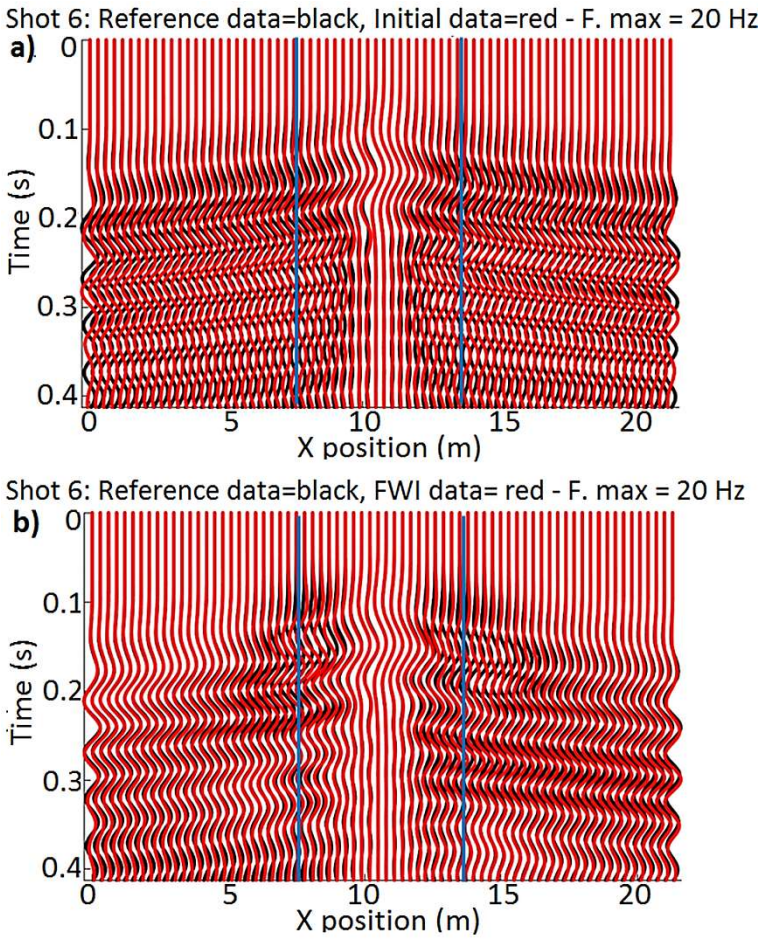

*Figure **8.28** – Trace-by-trace data fitting comparison before **a**) and after **b**) FWI, synthetic example, full-DC case, shot 6 (near-offset), multiscale approach, maximum **20 Hz** inversion run. The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

In figure **8.29**, showing a trace-by-trace data fitting comparison before and after the third inversion run (3 Hz – 30 Hz), only a little improvement in data fitting can be noticed after this stage.

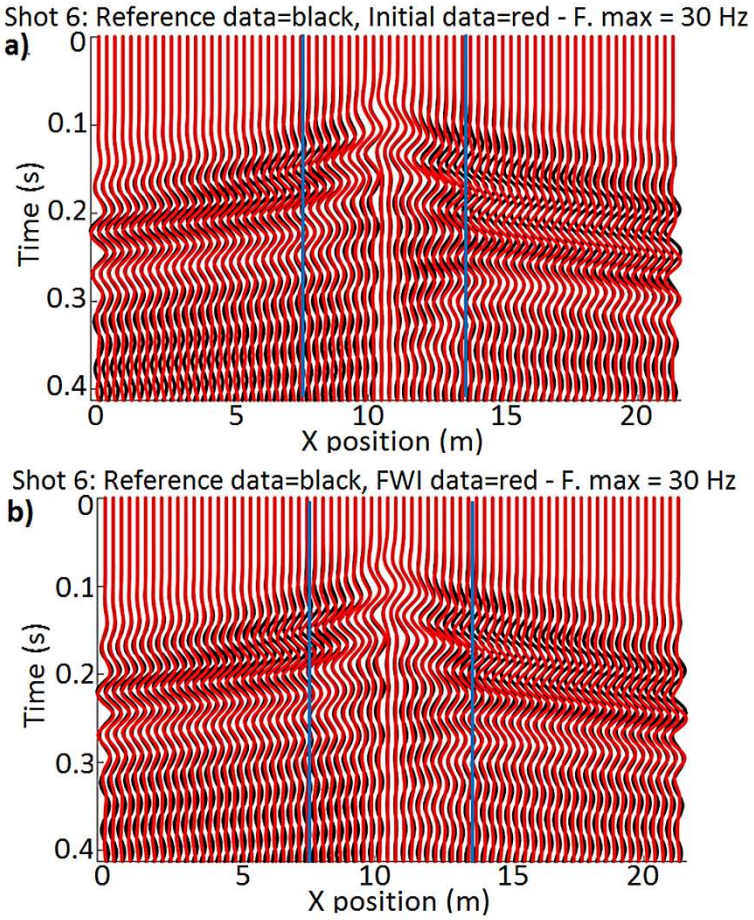

*Figure **8.29** – Trace-by-trace data fitting comparison before **a)** and after **b)** FWI, synthetic example, full-DC case, shot 6 (near-offset), multiscale approach: maximum **30 Hz** inversion run. The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

In figure **8.30**, showing a trace-by-trace data fitting comparison before and after the fourth inversion run (3 Hz – 40 Hz), an accurate data fitting can be noticed after FWI.

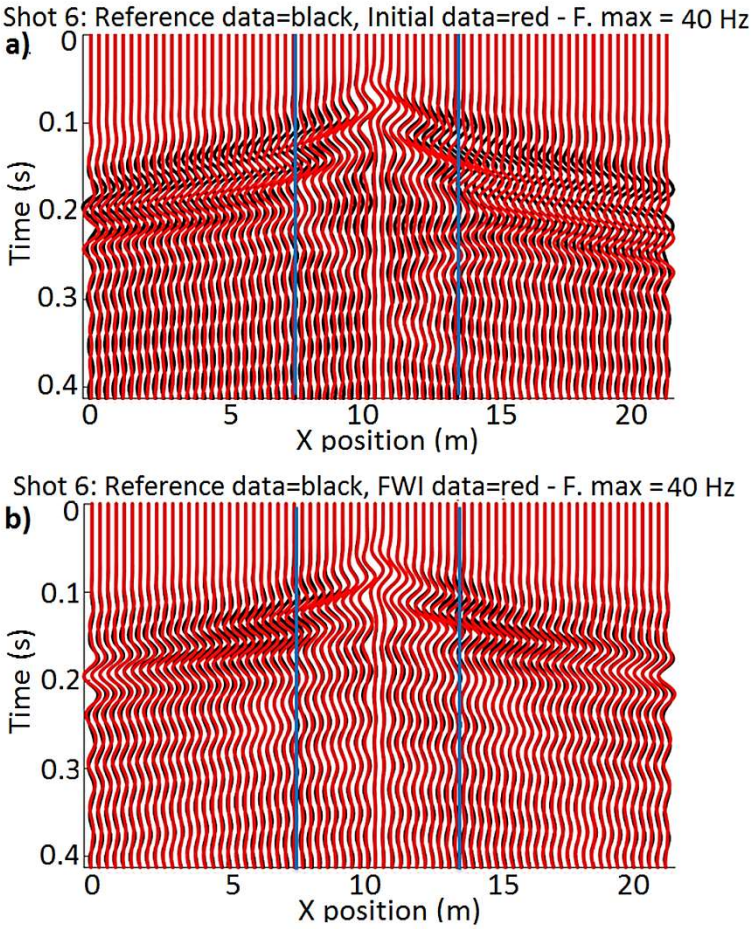

*Figure **8.30** – Trace-by-trace data fitting comparison before **a**) and after **b**) FWI, synthetic example, full-DC case, shot 6 (near-offset), multiscale approach, maximum **40 Hz** inversion run. The blue vertical lines indicate the approximate position of the low-velocity target. The seismograms are trace-by-trace normalized.*

8.4 – FWI experiments using a 3D acquisition layout

These tests are based on the 3D acquisition design presented in Section 6.3. The FWI workflow is kept relatively simple, to focus the analysis on the source positions and source number influence. Similar inversion parameters, as for the tests related to the 2D acquisition data set, are used, except for the coherent length of the Bessel smoothing filter in the *y*-direction, which here is 5 m. The offset weighting function, that proved to be very important for the 2D acquisition design, is less effective here (but still functional). The reason for this is the smaller maximum offset of the 3D acquisition (about 10 m) than the one of the 2D acquisition (about 25 m).

Different experiments have been conducted over both synthetic and real data, using a variable number of shots (from 8 to 16) and variable locations for the sources. The real source was estimated by deconvolution, for any single experiment, and the same source was used for all iterations of that FWI test. For instance, Figure **8.31** shows two examples of the estimated source signature for the 3D data set. In the first example, the sources are placed outside the low-velocity target (Figure **8.31a**). In the second example, some sources are activated also inside the target (Figure **8.31b**). The source signature is computed as an average of the individual estimated sources for each acquisition configuration (Figure **8.32**).

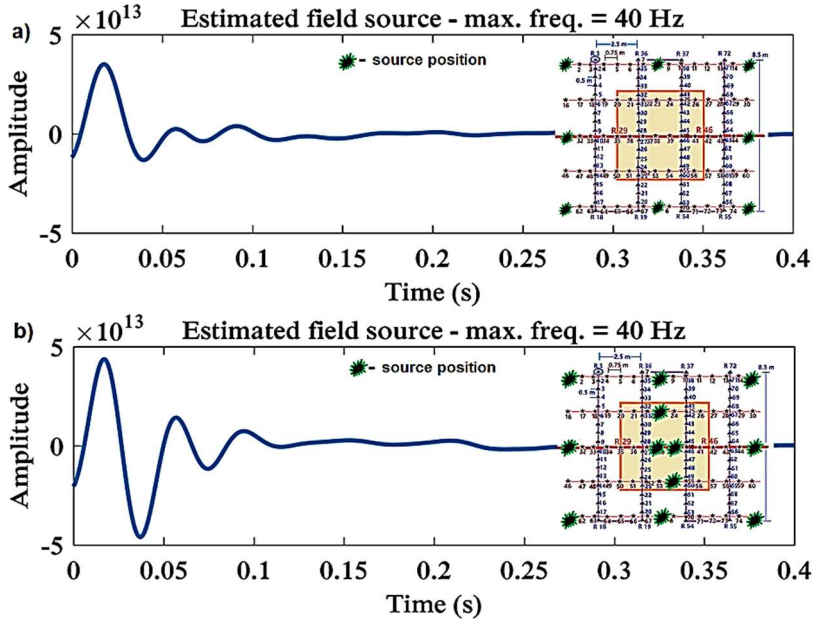

*Figure **8.31** – Example of estimated source for the 3D data set. **a**) Estimated source when considering 8 shots placed outside the low-velocity anomaly. **b**) The source estimated when considering 12 shots, placed both inside and outside the low-velocity anomaly. The green and black stars indicate the shot positions.*

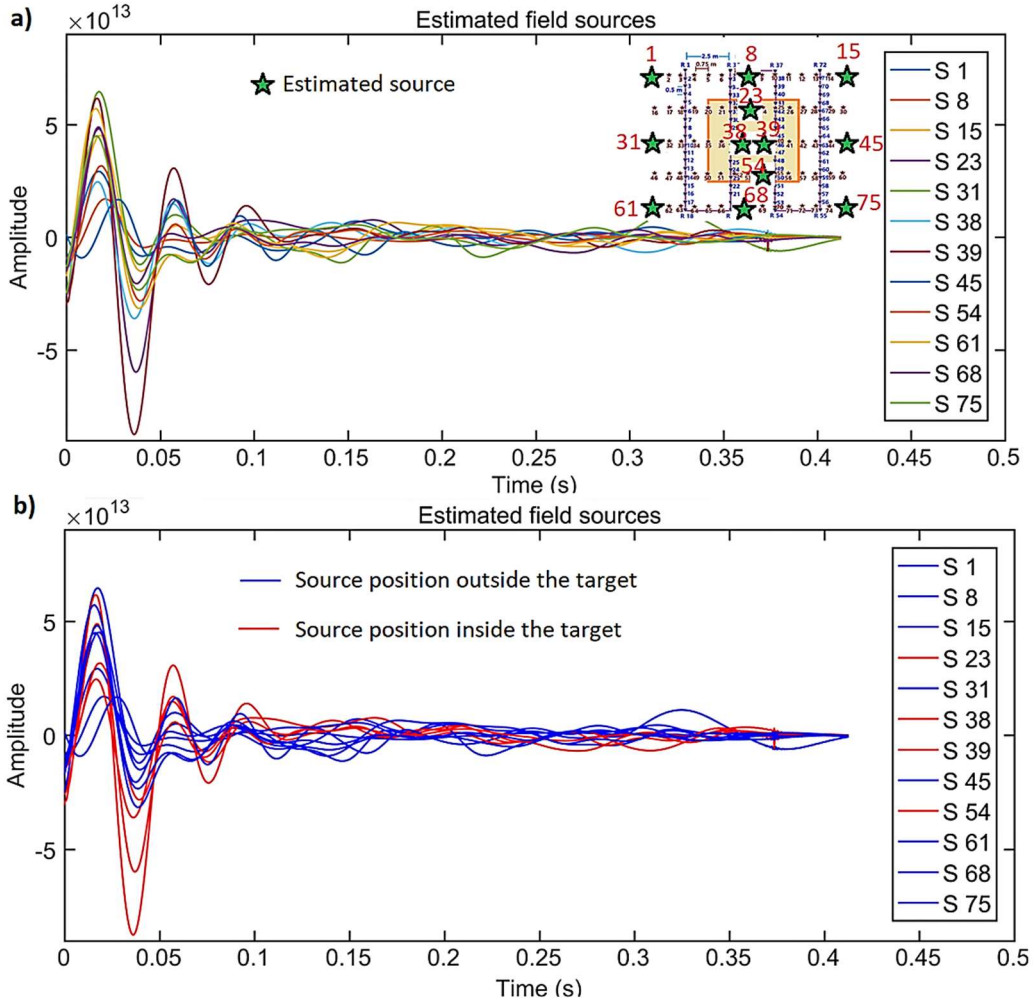

*Figure **8.32** – Example of sources estimated for the 3D data set. **a**) The sources estimated in different positions over the acquisition area are plotted with different colors. **b**) The sources estimated inside the low-velocity target are plotted in red while the sources estimated outside the target are plotted in blue.*

When using 8 sources for FWI, the computation mesh was decomposed into 2 x 4 x 2 sub-domains in the in *z*-*x*-*y*-direction, respectively. When using more than 8 sources, the domain decomposition followed the 2 x 2 x 2 scheme. All shots and decomposed domains have been distributed over 128 cores (8 nodes with 16 cores each - when considering 8 or 16 shots) or over 96 cores (6 nodes with 16 cores each - when considering 12 shots). Each FWI test required a computer memory of about 9.8 GB (when considering 8 shots) or 19.6 GB (when using 16 shots).

Table 8.2 gathers some inversion parameters used for the FWI experiments based on a 3D acquisition scheme.

Number of shots	**8 – 12 – 16**
Bessel smoothing filter: coherent lengths in the z – x – y direction	**1.5 m - 5 m - 5 m**
Domain decomposition when using 8 shots	**2-4-2** in the **z – x – y** direction
Domain decomposition when using 12 or 16 shots	**2-2-2** in the **z – x – y** direction
Computation resources involved when using 12 shots	**6** nodes **x 16** cores **= 96 cores**
Computation resources involved when using 16 shots	**8** nodes **x 16** cores = **128 cores**
Necessary computer memory for one FWI test when using 16 shots	**19.6 GB**

Table 8.2 – Parameters used during FWI for the 3D acquisition layout.

8.4.1 – Synthetic example

In the following section, the experiments for the 3D data set are first presented for the full-DC case (i.e. laterally variable initial models), which previously ensured the best FWI results.

Full-DC case

Figure **8.33** summarizes all acquisition layouts tested for the 3D data set. Similar results were obtained for the acquisition geometry of the cases A - D - E on one side, and B - C on the other side. The main observation arising from these tests is that varying the number of shots has less influence on the FWI results' accuracy than varying the shot positions.

Another observation is related to the achievement of better lateral resolution in correspondence of the target when positioning the shots outside the target (cased C) or nearby its borders (case B). Differently, when positioning some sources inside the target (case A, D and E), the estimated shallow velocity is more accurate, but the resolution of the reconstructed boundaries is lower.

Among the cases A, D and E, the worse results were obtained for the case A, while the best results were obtained for the case E. Among the cases B and C, the best results were obtained for the case C.

The best overall model reconstruction and data-fitting are provided by the configuration of the case C.

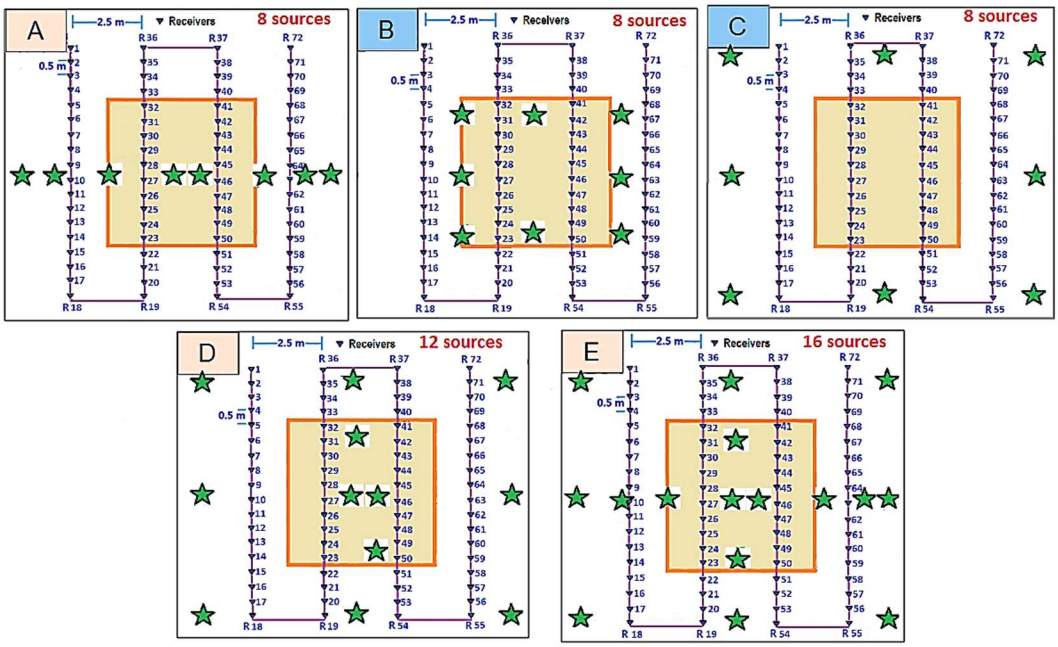

*Figure **8.33** – Acquisition geometries used for the FWI tests. The receivers are indicated with blue triangles, while the active shots, for each experiment, are signed with black and green stars. Capital letters of identical colors are used for the cases that provided similar results after FWI (A-D-E versus B-C).*

Some examples of reconstructed models and data-fitting are presented below for the best case of each group (A-D-E vs B-C), while the results for the other cases are reported in Appendix 6. Figure **8.34** shows 2D sections of the reconstructed 3D V_P and V_S models for the case E. The results show a good reconstruction of the shallow velocity, both for V_P and V_S while the resolution of target's shape and lateral extension can be further improved.

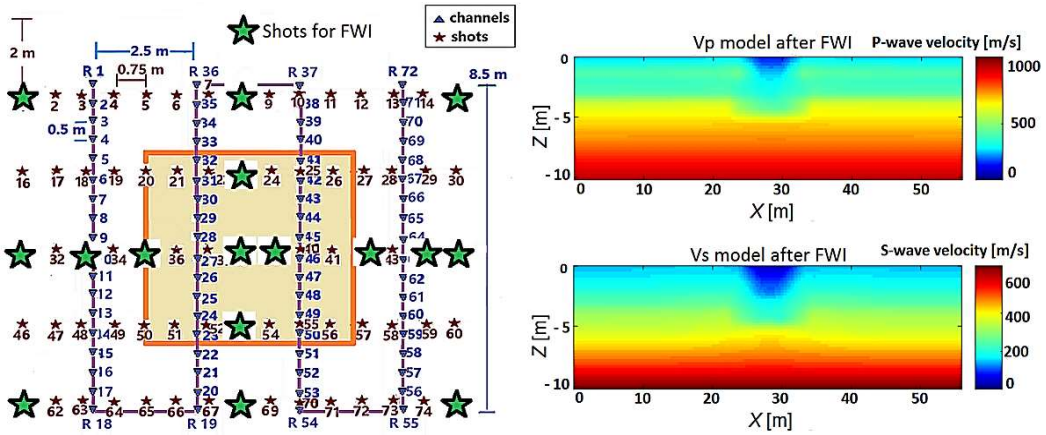

*Figure **8.34** – FWI results for case E, synthetic data. **Left**) Acquisition scheme (the receivers are indicated with blue triangles, while the active shots are signed in black and green stars). **Right**) 2D section (crossing the target's center) of the 3D P-wave (up) and S-wave (down) velocity models after FWI.*

Figure **8.35** shows the normalized model misfit, before and after FWI, for the case E: a low model misfit can be noticed in the final V_S model, as well as in the very shallow part of the V_P model (down to 1 m).

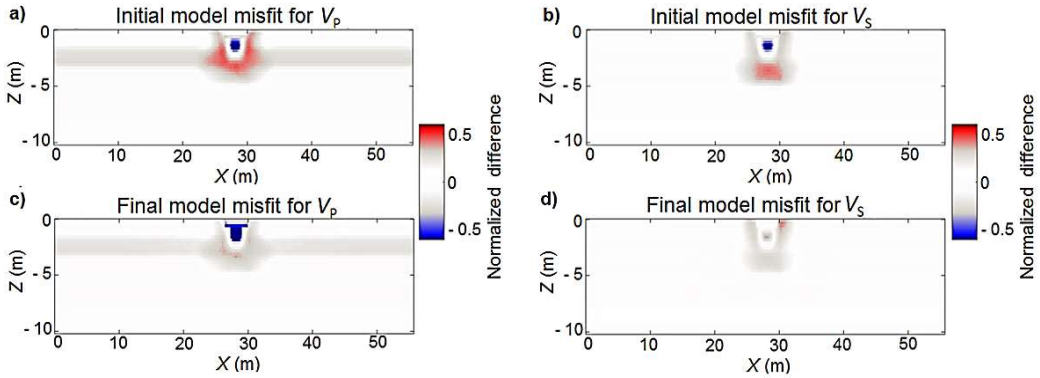

*Figure **8.35** – 2D sections (crossing the target's center) of the 3D model misfit: **a)** (Reference V_P model – Initial V_P model)/Reference V_P model. **b)** (Reference V_S model – Initial V_S model)/Reference V_S model. **c)** (Reference V_P model – Final V_P model)/Reference V_P model. **d)** (Reference V_S model – Final V_S model)/Reference V_S model.*

Figure **8.36** shows 2D sections of the reconstructed 3D models for case C, where all shots are placed outside the low-velocity target. The reader can notice how the target's shape is reconstructed with a higher resolution than in the previous case; nevertheless, there is a slight overestimation in velocity in the shallow part of the V_S model. This aspect can be noticed better in Figure **8.37b**, related to the normalized model misfit. Besides, the V_P final model misfit is very low in this case (Figure **8.37a**): The shots configuration of case C is the one that guarantees the best reconstruction of the V_P model.

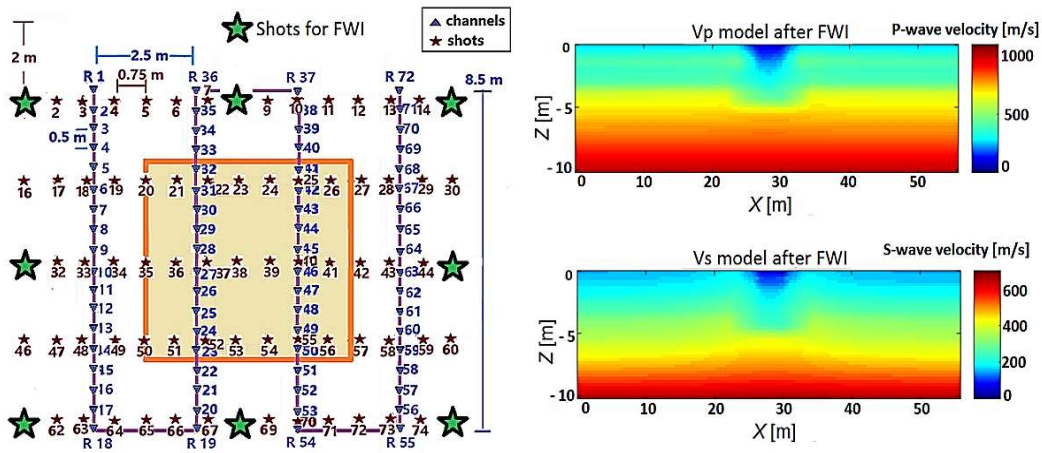

*Figure **8.36** – FWI results for case C, synthetic data. **Left**) Acquisition scheme (the receivers are indicated with blue triangles, while the active shots are signed in black and green stars). **Right**) 2D section (crossing the target's center) of the 3D P-wave (up) and S-wave (down) velocity models after FWI.*

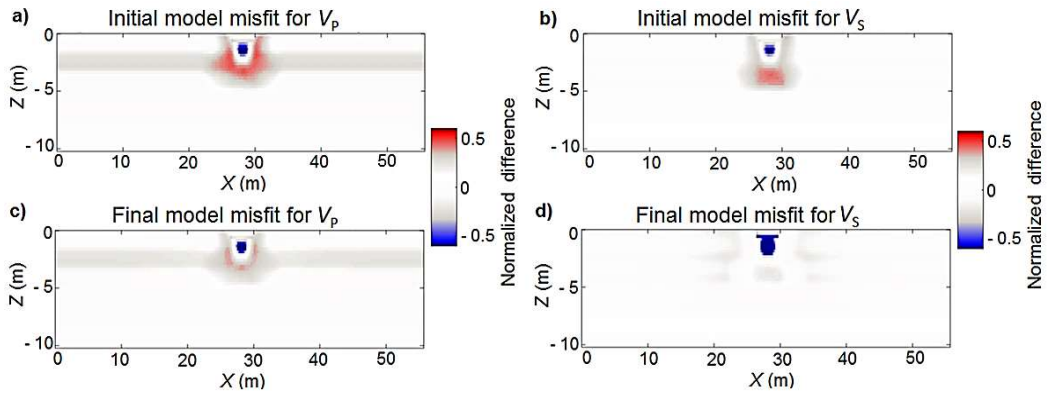

*Figure **8.37** – 2D sections (crossing the target's center) of the 3D model misfit: **a)** (Reference V_P model – Initial V_P model)/Reference V_P model. **b)** (Reference V_S model – Initial V_S model)/Reference V_S model. **c)** (Reference V_P model – Final V_P model)/Reference V_P model. **d)** (Reference V_S model – Final V_S model)/Reference V_S model.*

Figure **8.38** shows details of the 3D V_S models for z = 0 m, in particular: the initial model from full-DC analysis (**8.38a**), the reconstructed model of case E (**8.38b**) and the reconstructed model of case C (**8.38c**). One can notice a better lateral resolution of the horizontal target's extension for the case C (Figure **8.38c**), in the x-direction (the lateral resolution in the y-direction is already correct, after the 2D to 3D initial model extension). The shallow lateral extension of the final model for case C (Figure **8.37c**) is 5 meters in the x and y direction, as the extension of the reference (i.e. true) model.

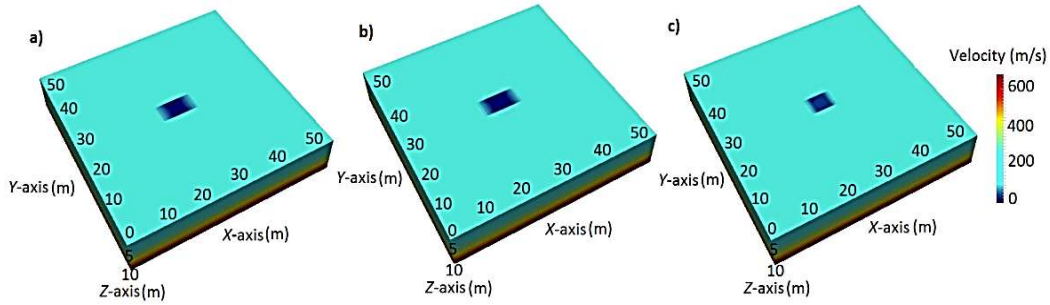

*Figure **8.38** – S-wave velocity models: **a)** Initial model obtained from the full-DC analysis. **b)** The model obtained after FWI when placing some sources in correspondence of the sand-target (Case E). **c)** The model obtained after FWI when using a 3D pattern for the source distribution, and no source is placed in correspondence of the sand-target (Case C).*

A trace-by-trace data fitting analysis, for a shot belonging to the case E, before and after FWI, is presented in Figure **8.39**. It can be noticed that the waveform matching of the initial configuration (Figure **8.39a**) improves alter FWI (Figure **8.39b**).

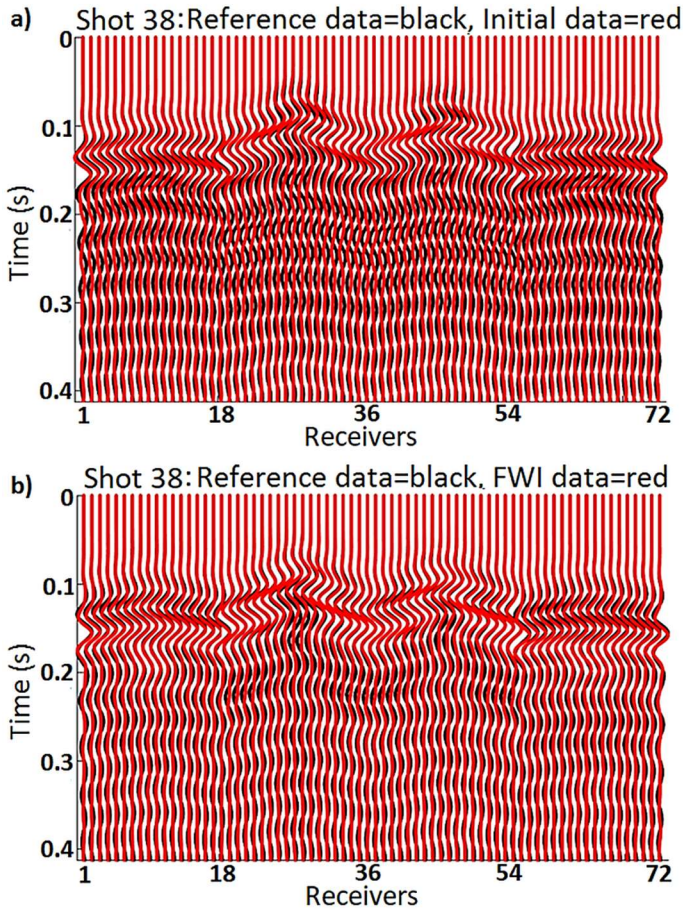

*Figure **8.39** – Trace-by-trace data fitting comparison before **a**) and after **b**) FWI, synthetic example, full-DC case, Shot 38 (placed inside the sand target), case E.*

Figure **8.40** displays a trace-by-trace data fitting comparison, for a shot belonging to the case C, before and after FWI. One can notice a data fitting improvement after FWI, especially for near-offset positions and in correspondence of the target' boundaries.

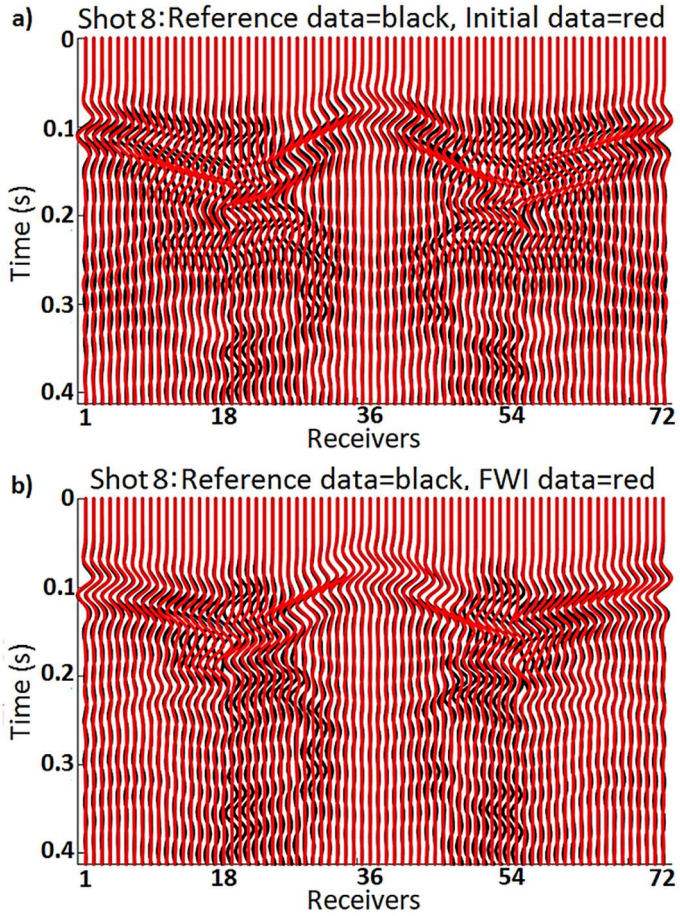

*Figure **8.40** – Trace-by-trace data fitting comparison before **a)** and after **b)** FWI, synthetic example, full-DC case, Shot 8 (placed outside the sand target), case C.*

Single-DC case

We have seen in Section 8.2 that FWI results characterized by lower accuracy are obtained when using laterally homogeneous initial models. Howsoever, some FWI tests were also performed using the initial models retrieved from single-DC analysis, to investigate if the low-velocity target can be reconstructed from a horizontally homogeneous background when using a 3D acquisition geometry.

Figure **8.41** shows 2D sections of the 3D models obtained after FWI on synthetic data, using the acquisition configuration of the case C and the single-DC based initial models. As expected, the target is not recovered in the V_P model (Figure **8.41a**), while its reconstruction is relatively accurate in the V_S model (Figure **8.41b**).

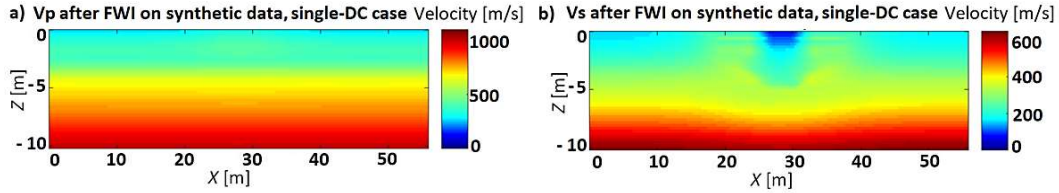

*Figure **8.41** – FWI results for the 3D acquisition, synthetic data (case C), single-DC case. 2D section (crossing the target's center) of the 3D **a)** P-wave velocity model and **b)** S-wave velocity model.*

This aspect can be also noticed when analyzing the Figure **8.42**, which displays the normalized difference between the reference and initial/final models, respectively: After FWI, the V_P is overestimated in correspondence of the target, while the V_S is accurately reconstructed in the shallow part of the sand body. Nevertheless, there are some artefacts in the V_S reconstructed model, far from the low-velocity target, consisting of velocity overestimation.

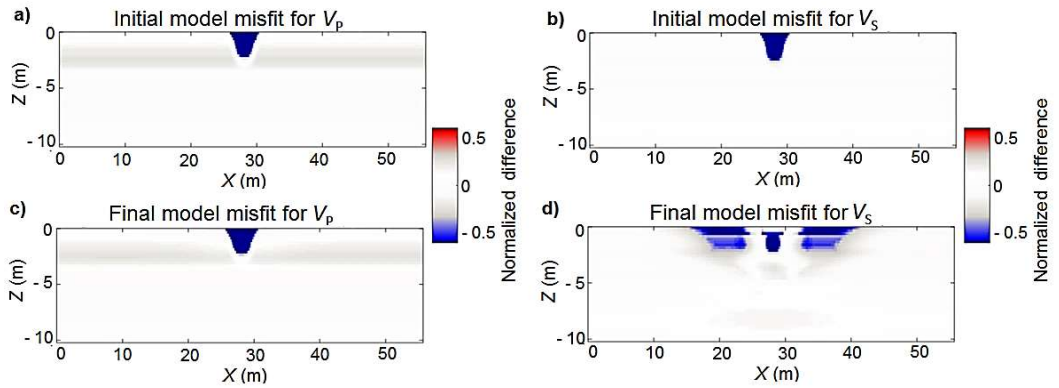

*Figure **8.42** – 2D sections (crossing the target's center) of the 3D model misfit: **a)** (Reference V_P model – Initial V_P model)/Reference V_P model. **b)** (Reference V_S model – Initial V_S model)/Reference V_S model. **c)** (Reference V_P model – Final V_P model)/Reference V_P model. **d)** (Reference V_S model – Final V_S model)/Reference V_S model.*

In Figure **8.43**, showing a detail of the 3D V_S model in correspondence of the surface, before (**8.43a**) and after (**8.43b**) FWI, the reader can notice how the target's horizontal extension is accurately reconstructed, even if the inversion starts from a laterally homogeneous background.

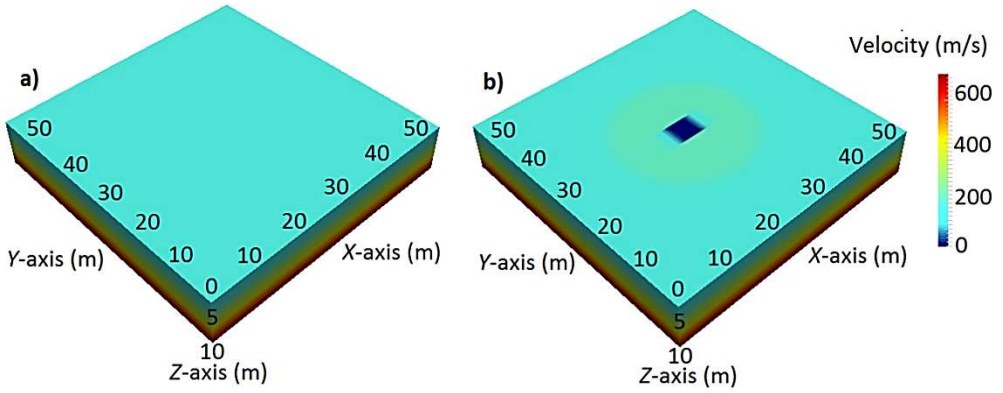

*Figure **8.43** – S-wave velocity models, synthetic case: **a)** Initial model obtained from the single-DC analysis. **c)** V_S model obtained after FWI when using a 3D pattern for the source distributions, and no shot is placed in correspondence of the sand-target (Case C).*

8.4.2 – Real data application

Single-DC case

Figure **8.44** shows 2D sections of the 3D models obtained after FWI on real data, using the configuration in C and the single-DC based initial model (i.e. laterally homogeneous). Again, the target is not recovered in the V_P model (Figure **8.44a**) while its reconstruction is relatively accurate in the V_S model (Figure **8.44b**).

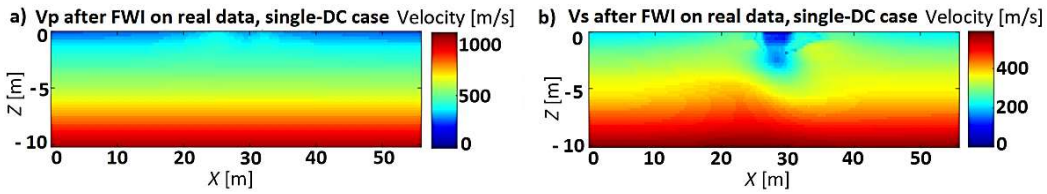

*Figure **8.44** – FWI results for the 3D acquisition, real data application (case C), single-DC case. 2D section (crossing the target's center) of the 3D **a)** P-wave velocity model and **b)** S-wave velocity model.*

Still, some artefacts nearby the low-velocity anomaly and at greater depth are present in the reconstructed V_S model. These elements can be also noticed in Figure **8.45**, which shows the normalized difference between the final and initial models. In particular, one can observe how the V_S value in correspondence of the target is lower (and therefore more accurate) in the final model, compared with the initial one.

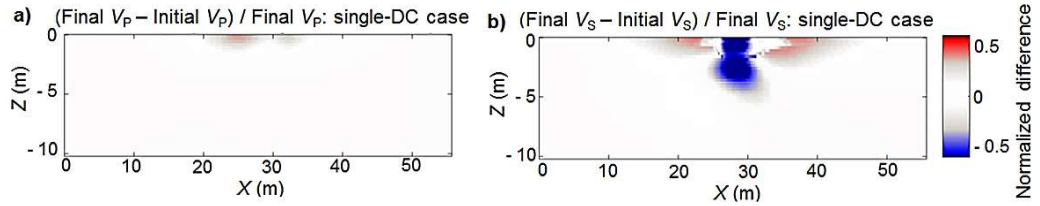

*Figure **8.45** – 2D sections (crossing the target's center) of the 3D model misfit: **a)** (Final V_P model – Initial V_P model)/Final V_P model. **b)** (Final V_S model – Initial V_S model)/Final V_S model.*

In Figure **8.46**, showing a detail of the 3D V_S model, in correspondence of the surface, before (**8.46a**) and after (**8.46b**) FWI, one can notice how the target's horizontal extension is properly reconstructed, even when starting from a laterally homogeneous background. However, some artefacts are present here around the low-velocity target.

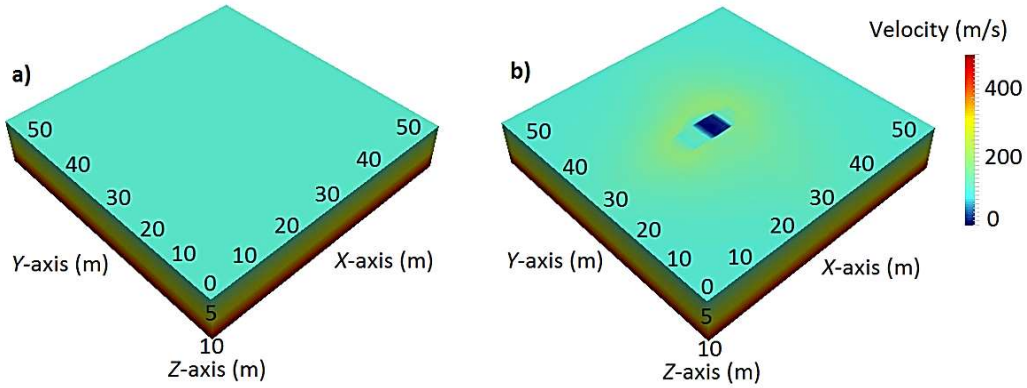

*Figure **8.46** – S-wave velocity models, real data application: **a)** Initial model obtained from the single-DC analysis. **b)** The model obtained after FWI when using a 3D pattern for the source distribution, and no source is placed in correspondence of the sand-target (Case C).*

Figure **8.47** presents a trace-by-trace data fittings comparison, before (**8.47a**) and after (**8.47b**) inversion, for the real data FWI application on the single-DC initial model. Accurate reconstruction of all phases can be noticed after inversion. The fitting improves especially in correspondence of the target since the correct velocity field is recovered after FWI.

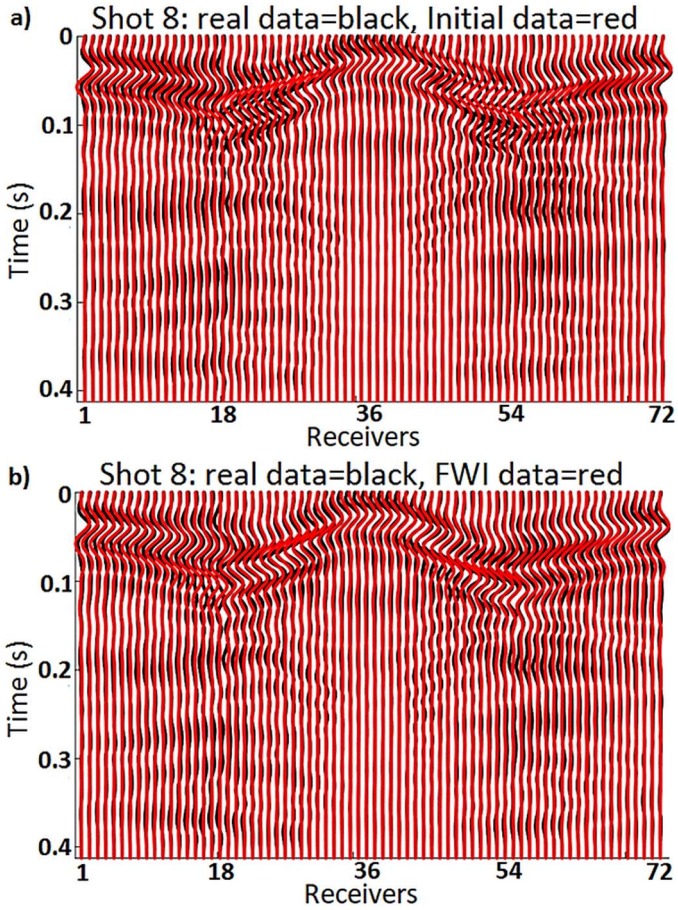

*Figure **8.47** – Trace-by-trace data fitting comparison before **a**) and after **b**) FWI, real data application, single-DC case, Shot 8, case C.*

The considerations of the previous paragraph regard also Figure **8.48**, showing a data fitting comparison for another shot gather belonging to the same FWI experiment. One can notice a significant improvement in data fitting, in correspondence of the target and for the far-offset arrivals, when passing from the initial configuration (Figure **8.48a**) to the final one (**8.48b**).

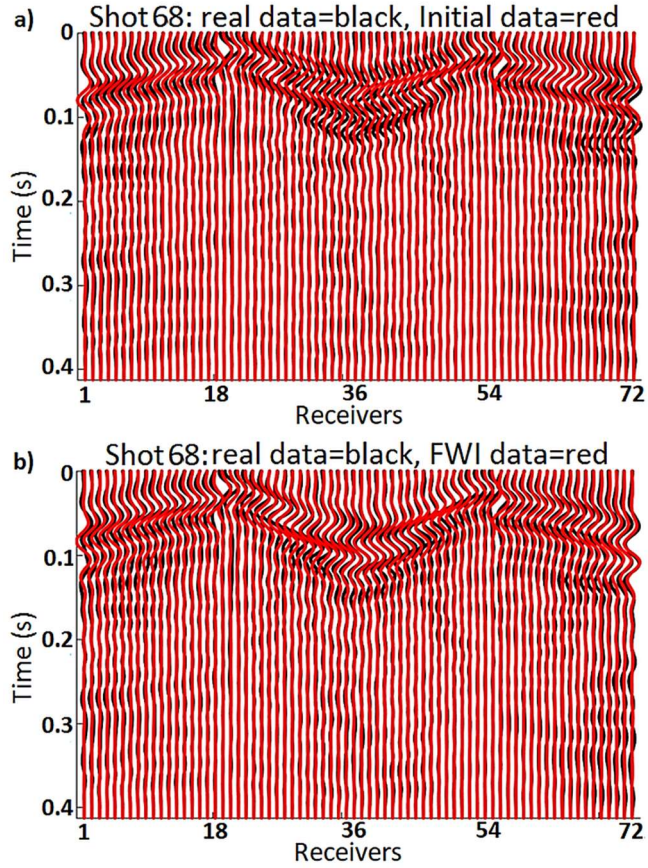

*Figure **8.48** – Trace-by-trace data fitting comparison before **a**) and after **b**) FWI, real data application, single-DC case, Shot 68, case C.*

Full-DC case

Figure **8.49** shows 2D sections of the 3D reconstructed models when using the acquisition configuration of the case C and the initial model from full-DC analysis (i.e. laterally heterogeneous), for FWI applied to real data. The resolution of the reconstructed model improves, especially for V_S. This feature can be observed also in Figure **8.50**, which displays the normalized difference between the final models and the initial ones. In the final V_S model, there is a slight increase in velocity in correspondence of the target after FWI (Figure **8.50b**).

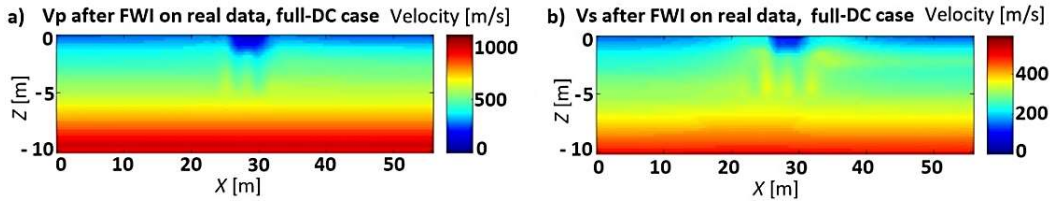

*Figure **8.49** – FWI results for the 3D acquisition, real data (case C), full-DC case. 2D section (crossing the target's center) of the 3D **a**) P-wave velocity model and **b**) S-wave velocity model.*

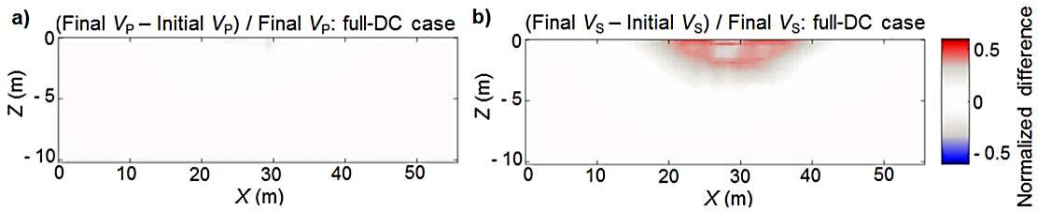

*Figure **8.50** – 2D sections (crossing the target's center) of the 3D model misfit: **a**) (Final V_P model – Initial V_P model)/Final V_P model. **b**) (Final V_S model – Initial V_S model)/Final V_S model.*

Figure **8.51** displays a detail of the 3D V_S model for $z = 0$ m, in particular: the initial model from full-DC analysis (Figure **8.51a**) and the model obtained after FWI for the acquisition configuration in C (figure **8.51b**). One can notice a better lateral resolution of the target in the *x*-direction after FWI; the lateral resolution in the y-direction is already accurate after the 2D to 3D initial model's extension.

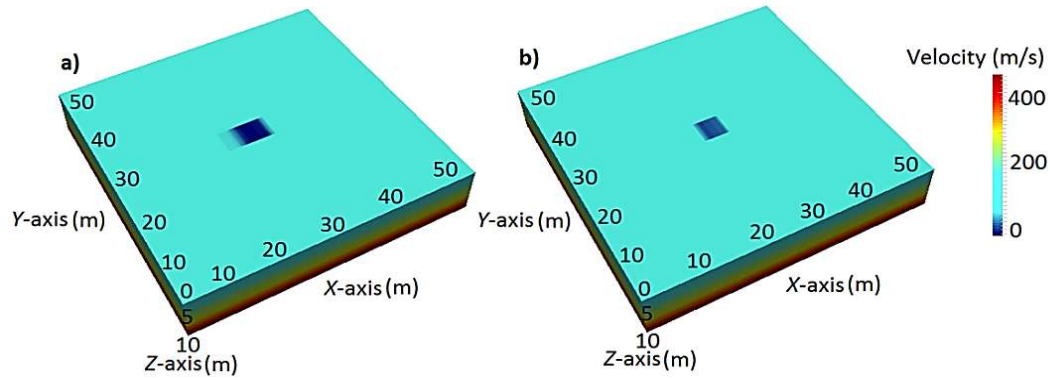

*Figure **8.51** – S-wave velocity models, real data application: **a**) Initial model obtained from the full-DC analysis. **b**) The model obtained after FWI when using a 3D pattern for the shot distributions, and no source is placed in correspondence of the sand-target (Case C).*

In Figure **8.52**, a trace-by-trace data fitting comparison is reported for a shot gather belonging to case C, before (**8.52a**) and after FWI (**8.52b**). An accurate data fitting can be noticed for the entire offset and all phases. Since the fitting is already accurate in the initial configuration, there are no significant changes after FWI. Still, during a careful examination, a better fitting of the phases scattered from the target's boundaries can be observed after FWI.

Other studies (e.g. Butzer et al., 2013) proved that the use of a 3D acquisition geometry for the 3D FWI allows reconstructing different sized 3D structures with a higher resolution compared with the 2D acquisition geometry results.

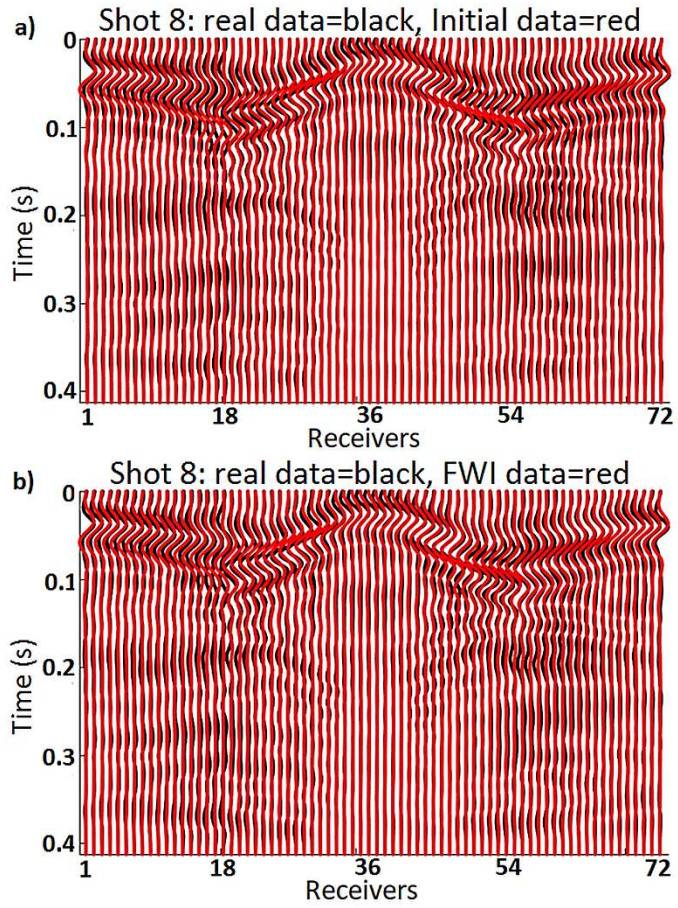

*Figure **8.52** – Trace-by-trace data fitting comparison before **a**) and after **b**) FWI, real data application, full-DC case, Shot 8, case C.*

8.5 – Highlights

- An integrated workflow, based on 2D SWs analysis and 3D FWI, was tested, with the aim of enhancing the resolution in the imaging of complex-shaped shallow targets. The tests are based on a well-known real target, represented by a loose-sand body buried among more compact sediments, and on synthetic data simulating this environment.
- Both real and synthetic seismic data are dominated by complex-scattered and highly energetic SWs that make the application of FWI a challenging task.
- Two different series of initial models (laterally homogeneous and laterally variable), built using SW analysis procedures, were tested. A simple 3D elastic FWI workflow, applied to synthetic and real data, improved the initial models' resolution. Better results were obtained after FWI when using laterally variable initial models.
- The final data fitting is accurate for the near-offset traces and in correspondence of the target, while the fitting of some far-offset arrivals and back-scattered phases can still be improved.
- Further investigations were focused on the integration, into the main FWI workflow, of different model-oriented and data-oriented strategies.
  - ✓ A preliminary monoparametric FWI was helpful for the initial V_S model improvement.
  - ✓ A two-step time windowing FWI, which firstly exploits the BWs content, helped for a better V_P model reconstruction. Even if the changes of the reconstructed model are not substantial, significant data fitting improvement still occurs for the far-offset traces. More successful examples can be found in the literature.
- A (commonly adopted) multiscale FWI provided slightly improved results after FWI. The difference between these results and the results obtained when applying FWI directly in the full frequency-band is small because the initial model from DC analysis is already accurate.
- A 3D acquisition geometry ensured a more accurate reconstruction of the target boundaries, especially when placing the sources outside the low-velocity target.
  - ✓ When activating some sources inside the sand target, the shallow velocity is more accurately reconstructed, but the resolution in correspondence of the target boundaries is lower. Since no data normalization acts during inversion, the sources placed inside the target contribute with high amplitude and control the optimization. Increasing the shots number does not influence the result in a significant manner.
  - ✓ Even though better results are obtained when starting the FWI from the full-DC analysis based initial model, the initial model obtained by single-DC analysis also guarantees the premise for an accurate reconstruction of the target shape and horizontal extension. Nevertheless, in this case, some additional artefacts appear in the reconstructed model.

Appendix 4

The effectiveness of data weighting strategy

In Figure **A 4.1**, showing the V_S models reconstructed after FWI for the case when no offset weighting function is used (**A 4.1a** and **A 4.1b**) and the case when such data strategy is integrated into the FWI workflow (**A 4.1c** and **A 4.1d**), a more accurate reconstruction of the target can be noticed when using the weighting function.

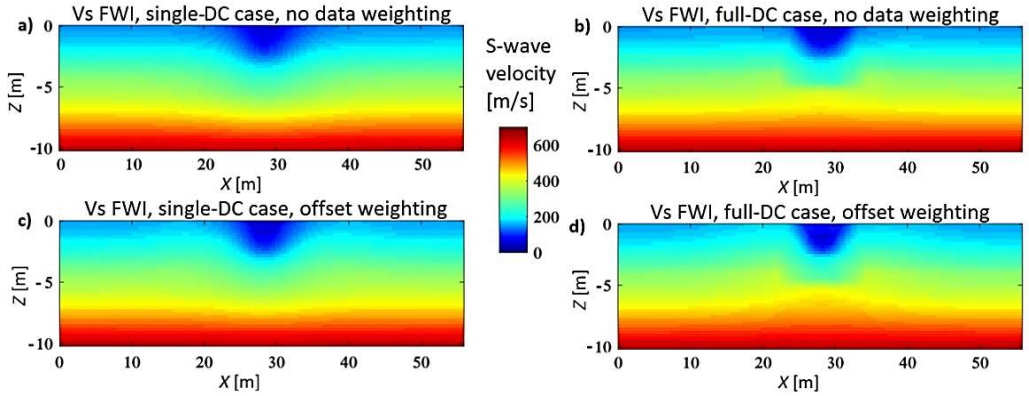

*Figure **A 4.1** – 2D sections (crossing the target's center) of the 3D S-wave velocity models after FWI on synthetic data (2D acquisition): **a**) Single-DC case (i.e. 1D), without using the weighting function, **b**) full-DC case (i.e. 2D), without using the weighting function, **c**) single-DC case (i.e. 1D), using the weighting function, **d**) full-DC case (i.e. 2D), using the weighting function.*

The effectiveness of the weighting function is also proven by the normalized model misfit plotted in Figure **A 4.2**. A lower model misfit after FWI, in correspondence of the target's boundaries, is achieved when using the weighting function.

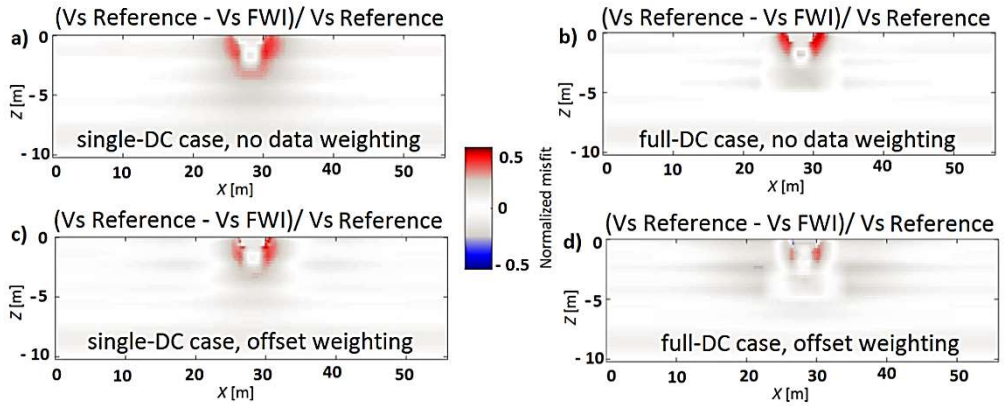

*Figure **A 4.2** – 2D sections (crossing the target's center) of the 3D S-wave velocity model misfit, regarding the FWI application on synthetic data (2D acquisition): **a**) Single-DC case, without the weighting function, **b**) full-DC case, without the weighting function, **c**) single-DC case, using the weighting function, **d**) full-DC case, using the weighting function.*

In Figures **A 4.3** and **A 4.4**, showing a trace-by-trace data fitting comparison for the case when no offset weighting is used and the case when data weighting is applied, more accurate results can be noticed in the second case (**A 4.3b** and **A 4.4b**).

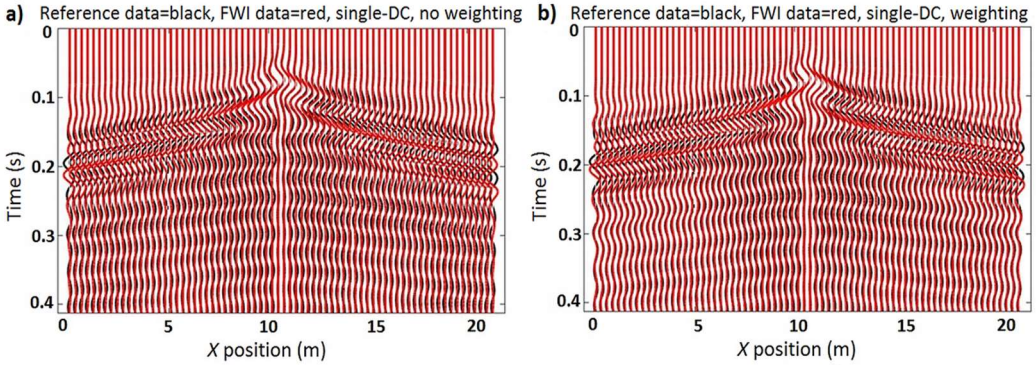

*Figure **A 4.3** – Trace-by-trace data fitting comparison between reference data and data belonging to the model obtained after FWI, for the single-DC case (i.e. 1D), 2D acquisition, synthetic example (Shot 6): **a**) No weighting function was used during FWI; **b**) a weighting function was used during FWI for the far-offset traces.*

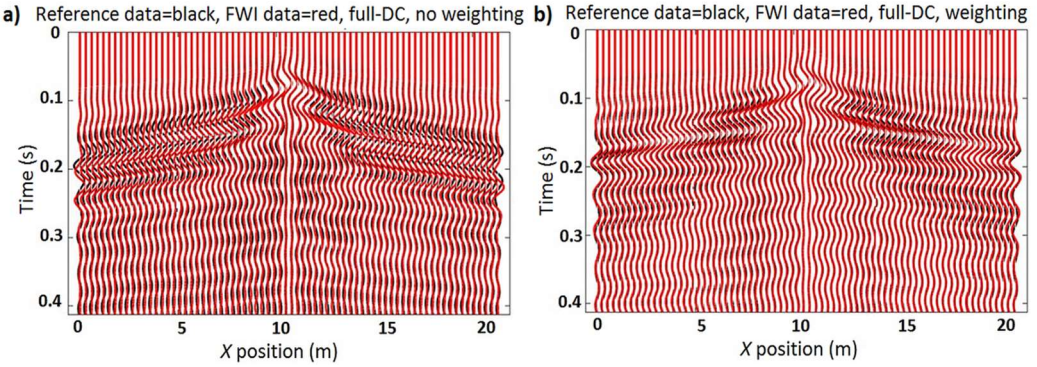

*Figure **A 4.4** – Trace-by-trace data fitting comparison between reference data and data belonging to the model obtained after FWI, for the full-DC case (i.e. 2D), 2D acquisition, synthetic example (Shot 6): **a**) No weighting function was used during FWI; **b**) a weighting function was used during FWI for the far-offset traces: A more accurate data fitting can be noticed.*

In Figure **A 4.5**, which shows the FWI results obtained when no data weighting is used (**A 4.5b**) and when an offset weighting is applied (**A 4.5c**), a more accurate reconstruction of the target can be noticed in the second case (Figure **A 4.5c**). The same aspect can be observed in Figure **A 4.6**, where a more accurate reconstruction of the V_S after FWI can be noticed in the shallow part of the model (Figure **A 4.6b**).

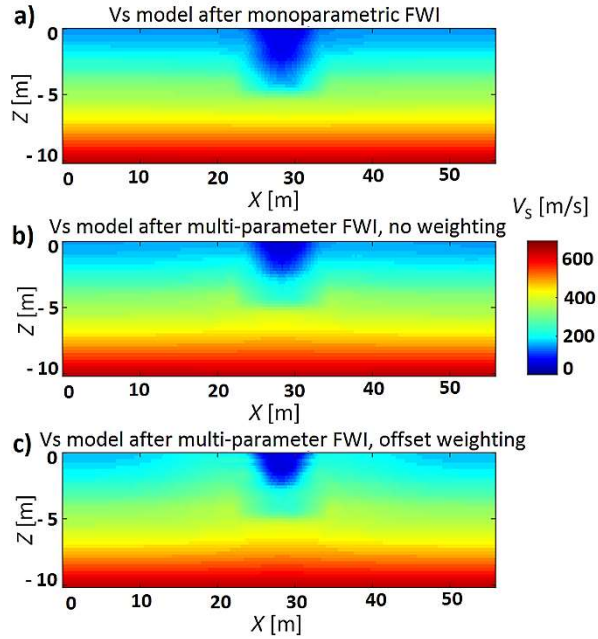

*Figure **A 4.5** – 2D sections of the 3D S-wave velocity models, synthetic example. **a)** Improved V_S model through a preliminary monoparametric FWI. **b)** V_S model after multi-parameter FWI (starting from the improved initial Vs model and the previous V_P model, retrieved from full-DC analysis): No offset weighting function was applied. **c)** V_S model after multi-parameter FWI (starting from the improved initial V_S model and the previous V_P model, derived from full-DC analysis): A weighting function was applied to the far-offset traces.*

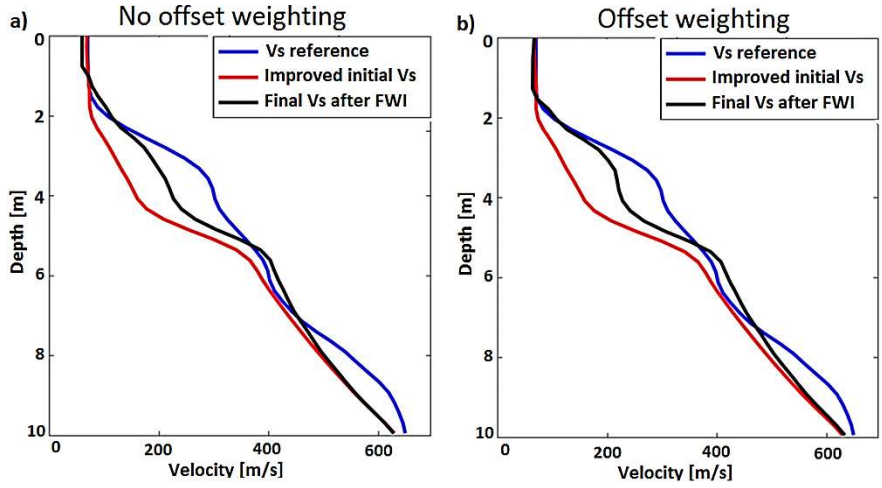

*Figure **A 4.6** – 1D profiles (in correspondence of the target) of the 3D S-wave velocity models, synthetic example. **a)** No weighting function was used. **b)** An offset variable weighting function was used.*

Again, a more accurate data-fitting can be noticed, in Figure **A 4.7b,** when an offset weighting function is used during FWI.

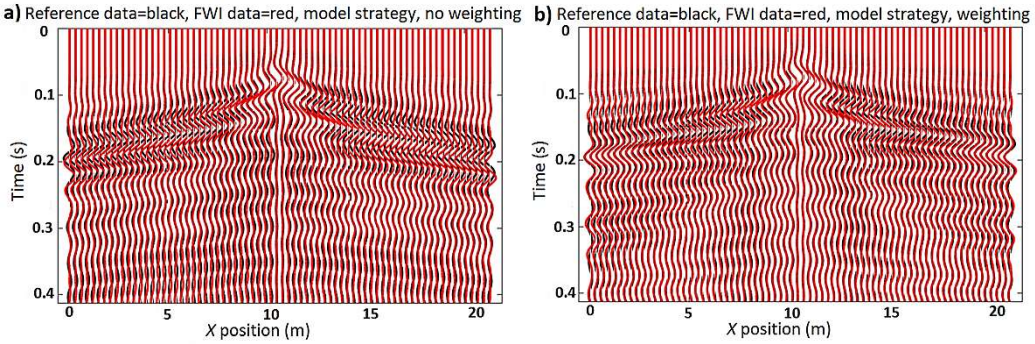

*Figure **A 4.7** – Trace-by-trace data fitting comparison between reference data and data belonging to the model obtained after FWI (using a preliminary monoparametric FWI), for the full-DC case (i.e. 2D), 2D acquisition (Shot 6): (a) No weighting function was used during FWI; (b) a weighting function was used during FWI for the far-offset traces.*

Furthermore, for the multiscale FWI strategy, one can notice better results when an offset weighting strategy is integrated into the FWI workflow (Figure **A 4.8b** and Figure **4.9b**).

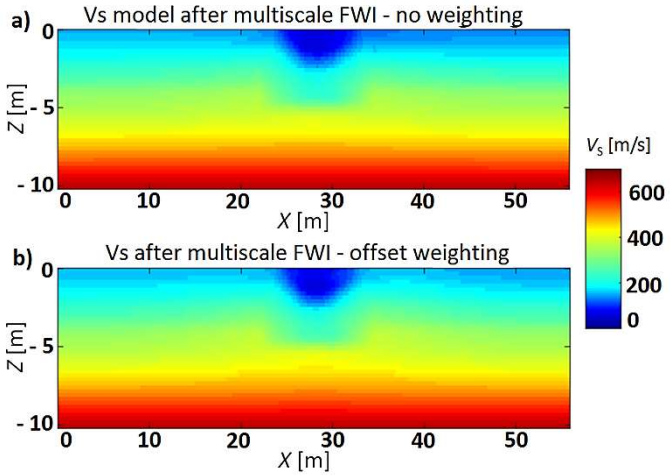

*Figure **A 4.8** – 2D sections of the 3D S-wave velocity models: **a)** V_S model after multi-parameter FWI (full-DC case): No offset weighting function has been applied. **b)** V_S model after multi-parameter FWI: A weighting function was applied to the far-offset traces: A better resolution can be noticed in correspondence of the target.*

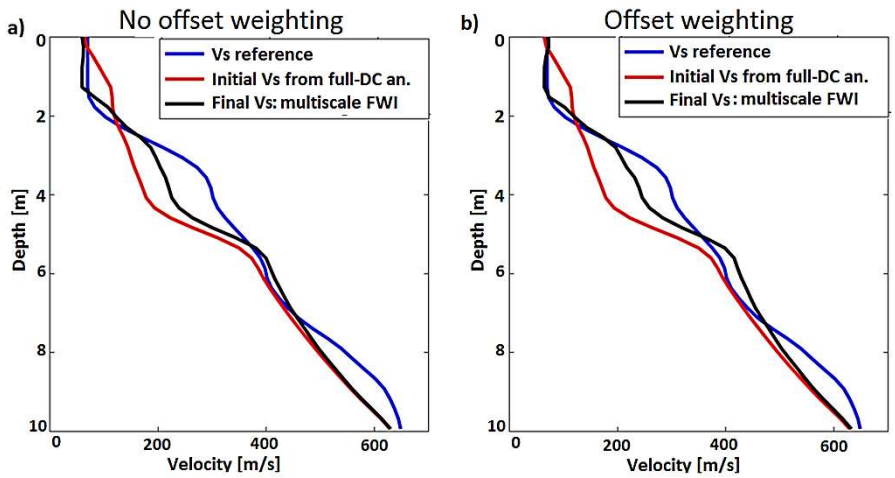

*Figure **A 4.9** – 1D profiles (in correspondence of the target) of the 3D S-wave velocity models. **a**) No weighting function was used during the multiscale FWI. **b**) An offset variable weighting function was used during FWI: A more accurate reconstruction of the V_S model after FWI can be noticed in the shallow part of the model.*

Appendix 5

Bessel smoothing filter for the FWI gradient: experiments

Figure **A 5.1** shows the effects of the Bessel filter's application to the first FWI gradient, when different values of the coherent lengths, in the z-x- and y-direction, are used for the filter parameters. The values for the smoothing filter of the cases **(e)** and **(f)**, used for the FWI tests of the 2D acquisition data set, are in agreement with the FWI wavelength resolution, acquisition design and mesh size.

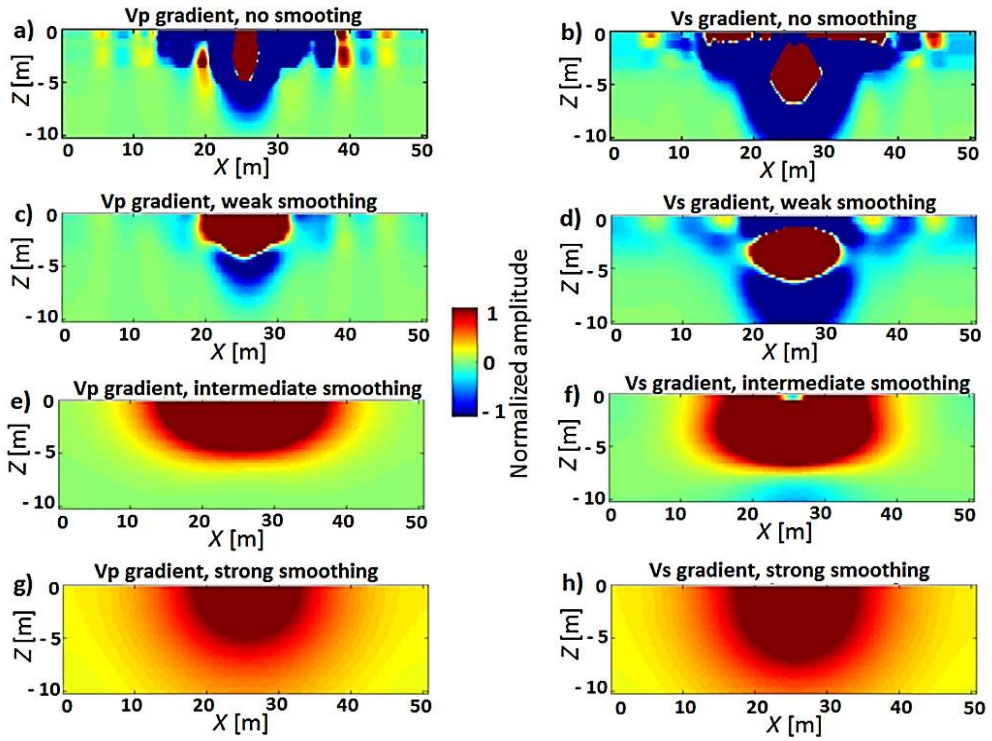

*Figure **A 5.1** – 2D sections (crossing the target's center) of the 3D FWI gradients.*

***a)** Gradient of the objective function with respect to V_P: no smoothing.*	***b)** Gradient of the objective function with respect to V_S: no smoothing.*
***c)** Gradient of the objective function with respect to V_P: Bessel smoothing filter's values in the z – x - y direction = 0.5 m - 1.5 m – 5 m.*	***d)** Gradient of the objective function with respect to V_S: Bessel smoothing filter's values in the z – x - y direction = 0.5 m - 1.5 m – 5 m.*
***e)** Gradient of the objective function with respect to V_P: Bessel smoothing filter's values in the z – x - y direction= 1.5 m – 5 m – 12 m.*	***f)** Gradient of the objective function with respect to V_S: Bessel smoothing filter's values in the z – x - y direction= 1.5 m – 5 m – 12 m.*
***g)** Gradient of the objective function with respect to V_P: Bessel smoothing filter's values in the z – x - y direction = 5 m – 10 m – 15 m.*	***h)** Gradient of the objective function with respect to V_S: Bessel smoothing filter's values in the z – x - y direction= 5 m – 10 m – 15 m.*

Appendix 6

FWI results for the 3D acquisition geometry: some details

In Figure **A 6.1**, showing the FWI results related to the case A of the 3D acquisition design, one can notice how the shape of the target is recovered with lower resolution compared with the next example (Figure **A 6.2**). Indeed, lower data misfit is obtained when placing the shots in these positions (Figure **A 6.2 –** case B).

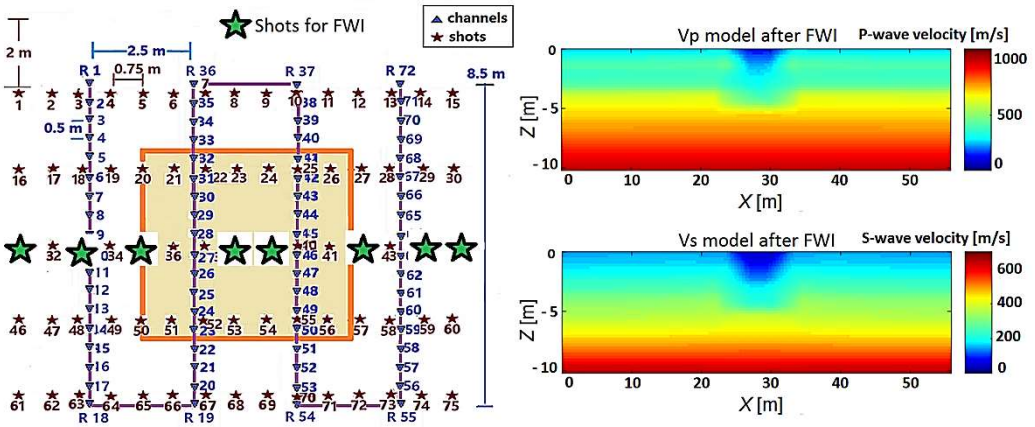

*Figure **A 6.1** – FWI results for the 3D acquisition, synthetic data (case A). **Left**) Acquisition scheme (the receivers are indicated with blue triangles, while the active shots are signed in black and green). **Right**) 2D section (crossing the target's center) of the 3D P-wave (up) and S-wave (down) velocity model after FWI.*

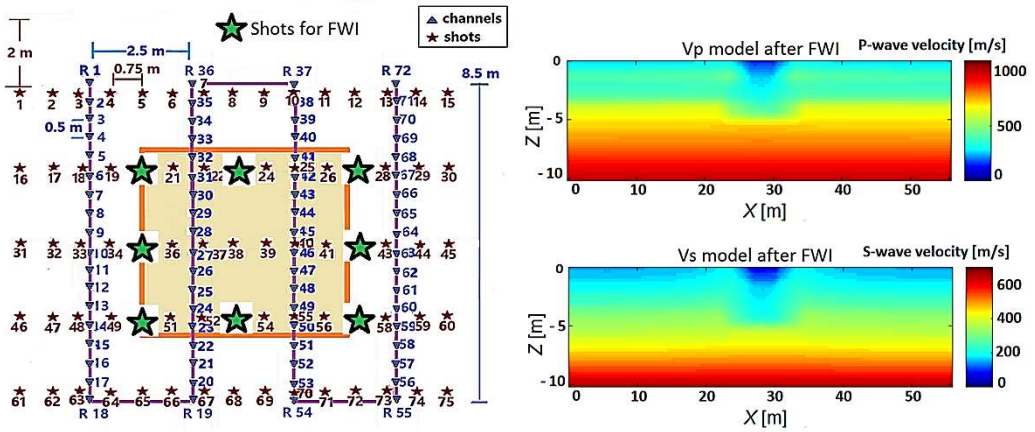

*Figure **A 6.2** – FWI results for the 3D acquisition, synthetic data (case B). **Left**) Acquisition scheme (the receivers are indicated with blue triangles, while the active shots are signed in black and green). **Right**) 2D sections (crossing the target's center) of the 3D P-wave (up) and S-wave (down) velocity model after FWI.*

When analyzing the corresponding model misfits for these two cases, one can notice how the V_S estimation, in correspondence of the target, is more accurate for the shots configuration of the Case A (Figure **A 6.3**) than for the shots distribution of the Case B (Figure **A 6.4**). Conversely, the target's lateral extension is more accurately reconstructed when using the acquisition geometry of the case B (Figure **A 6.4**), compared with the case A (Figure **A 6.3**).

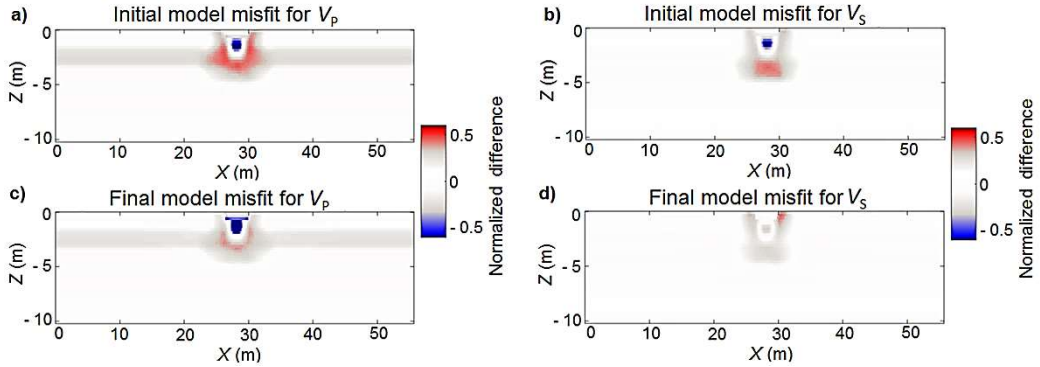

*Figure **A 6.3** – 2D sections (crossing the target's center) of the 3D model misfit for case A: **a)** (Reference V_P model – Initial V_P model)/Reference V_P model. **b)** (Reference V_S model – Initial V_S model)/Reference V_S model. **c)** (Reference V_P model – Final V_P model)/Reference V_P model. **d)** (Reference V_S model – Final V_S model)/Reference V_S model.*

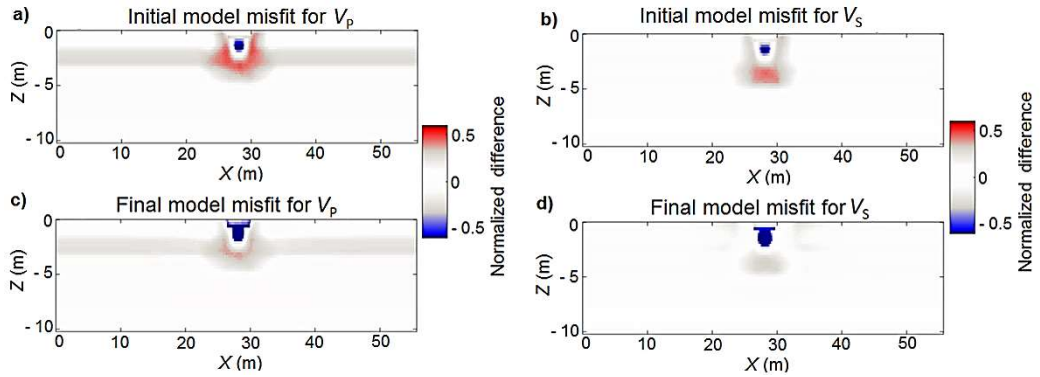

*Figure **A 6.4** – 2D sections (crossing the target's center) of the 3D model misfit for case B: **a)** (Reference V_P model – Initial V_P model)/Reference V_P model. **b)** (Reference V_S model – Initial V_S model)/Reference V_S model. **c)** (Reference V_P model – Final V_P model)/Reference V_P model. **d)** (Reference V_S model – Final V_S model)/Reference V_S model*

When using the shots configuration of the case D (Figure **A 6.5** and Figure **A 6.6**), the FWI results are similar to the ones obtained when using the acquisition design of the case E (presented in Section 8.4). Nevertheless, lower data misfit is obtained here.

For the experiments synthetized in Figures **A 6.5** and **A 6.6**, the estimation of the shallow velocity is more accurate in correspondence of the target than in the previous case (Figures **A 6.2** and **A 6.4**), especially for the *Vs* model, at the expense of less accurate reconstruction of the target's shape and lateral extension.

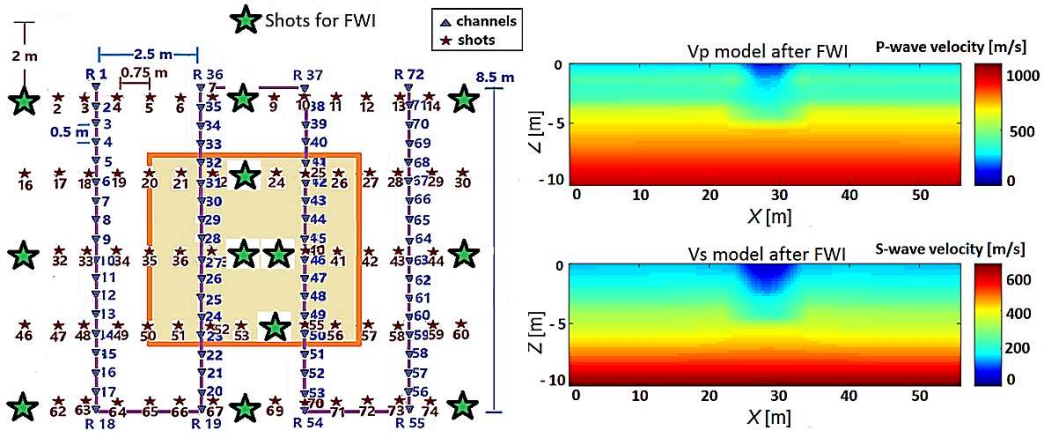

*Figure **A 6.5** – FWI results for the 3D acquisition, synthetic data (case D): **Left**) Acquisition scheme (the receivers are indicated with blue triangles, while the active shots are signed in black and green). **Right**) 2D sections (crossing the target's center) of the 3D P-wave (up) and S-wave (down) velocity models after FWI.*

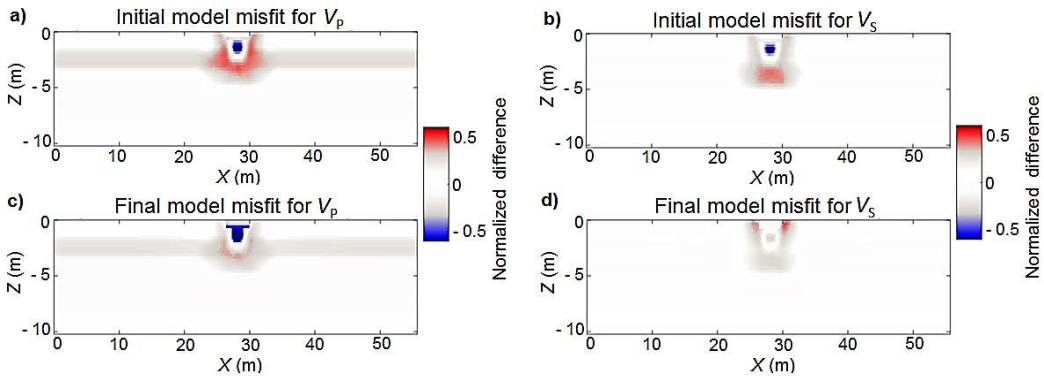

*Figure **A 6.6** – 2D sections (crossing the target's center) of the 3D model misfit for the case D: **a**) (Reference V_P model – Initial V_P model)/Reference V_P model. **b**) (Reference V_S model – Initial V_S model)/Reference V_S model. **c**) (Reference V_P model – Final V_P model)/Reference V_P model. **d**) (Reference V_S model – Final V_S model)/Reference V_S model.*

Figure **A 6.7** shows the data fitting comparison, for the same shot position and different acquisition configurations: case D, case E and case C. One can observe a better data fitting for the case E (Figure **A 6.7b**) than for the case D (Figure **A 6.7a**), and for the case C (Figure **A 6.7c**), compared with the case D (Figure **A 6.7b**).

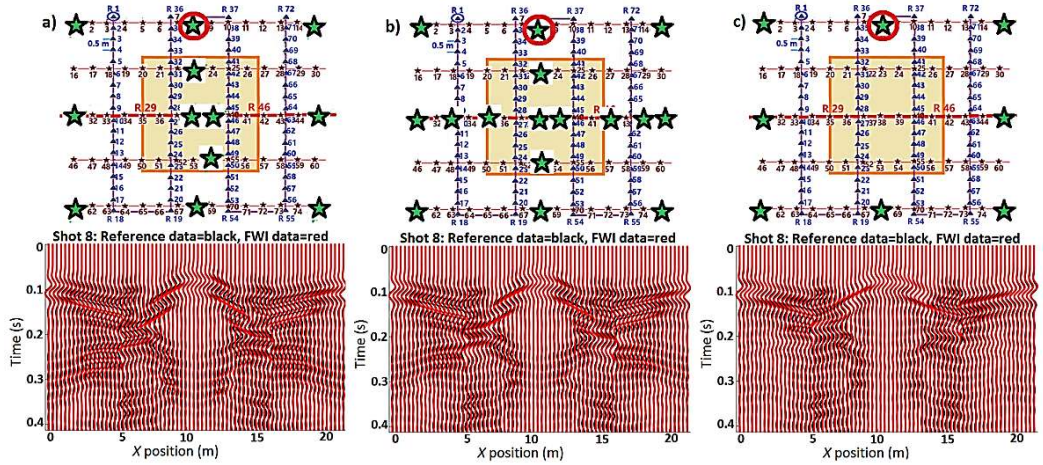

*Figure **A 6.7** – Trace-by-trace data fitting comparison, after FWI, for the same shot position (Shot 8) over different acquisition geometries, synthetic example: a) Case D, b) Case E, c) Case C. The reference data are plotted in black, while the final data, for each experiment, are represented in red. The best data-fitting after FW is obtained for case C (Figure A 6.7c).*

Figure **A 6.8** shows the data fitting comparison, for the same shot position and different acquisition configurations: case A, case E and case B. In can be noticed that better data fitting is obtained for the case E (Figure **A 6.8b**) than for the case A (Figure **A 6.8a**), and for the case B (Figure **A 6.8c**), compared with the case E (Figure **A 6.8b**).

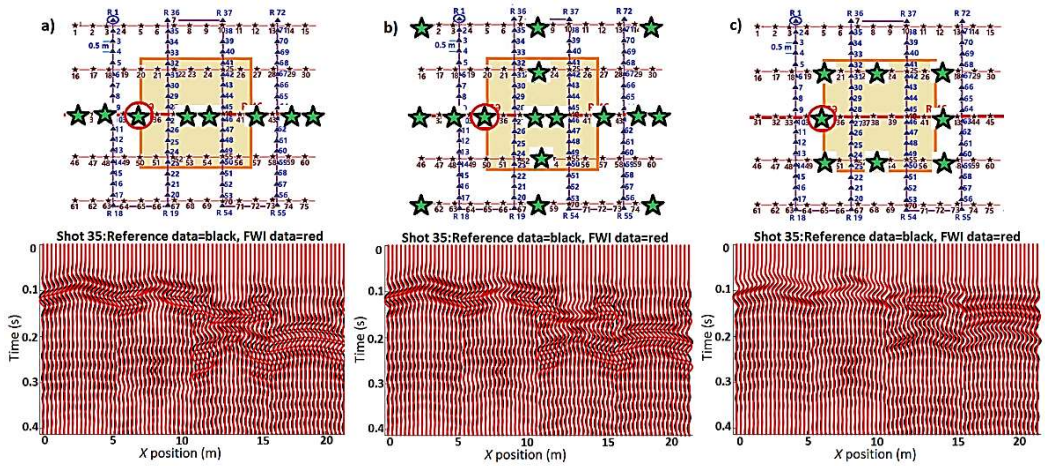

*Figure **A 6.8** – Trace-by-trace data fitting comparison, after FWI, for the same shot position (Shot 35) over different acquisition designs, synthetic example: **a**) Case A, **b**) Case E, **c**) Case B. The reference data are plotted in black, while the final data, for each experiment, are represented in red. The best data-fitting after FW is obtained for case B (Figure A 6.8c).*

Figure **A 6.9** shows the data fitting comparison, for the same shot position (Shot 38, belonging to the 3D acquisition) and different acquisition configurations: case A, case D and case E. It can be observed that better data fitting is obtained for the case D (Figure **A 6.9b**) than for the case A (Figure **A 6.9a**), and for the case E (Figure **A 6.9c**), compared with the case D (Figure **A 6.9b**).

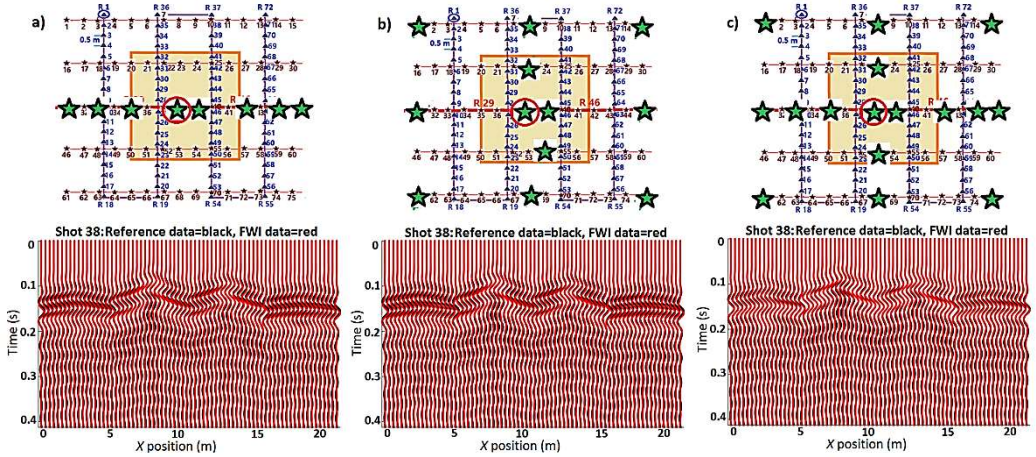

*Figure **A 6.9** – Trace-by-trace data fitting comparison, after FWI, for the same shot position (Shot 38), over different acquisition designs, synthetic example: **a**) Case A, **b**) Case D, **c**) Case E. The reference data are plotted in black, while the final data, for each experiment, are represented in red. The best data-fitting after FW is obtained for case E (Figure A 7.6c).*

Figure **A 6.10** shows the final models obtained when using the 3D acquisition configuration of the case E on real data. These results are less accurate than the ones related to the case C (shown in Section 8.4)

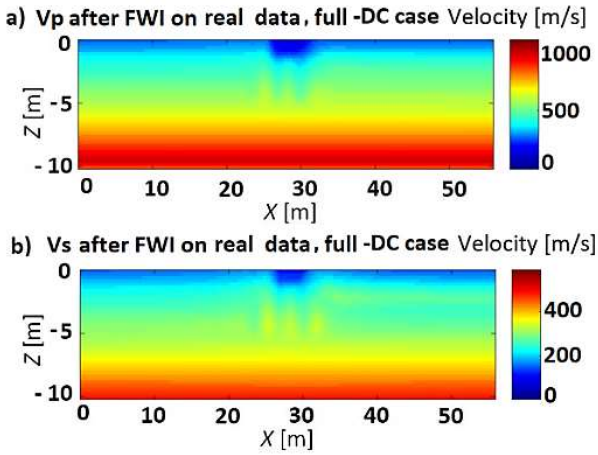

*Figure **A 6.10** – FWI results for the 3D acquisition, real data (case D): **a**) 2D section (crossing the target's center) of the 3D P-wave velocity, **b**) 2D section (crossing the target's center) of the 3D S-wave velocity.*

Analyzing the normalized difference between the final and the initial models (Figure **6.11**), it can be noticed, in the final V_S model (Figure **A 6.11b**), a slight increase in velocity around the target. This feature is not observed in the final V_P model (Figure **A 6.11a**).

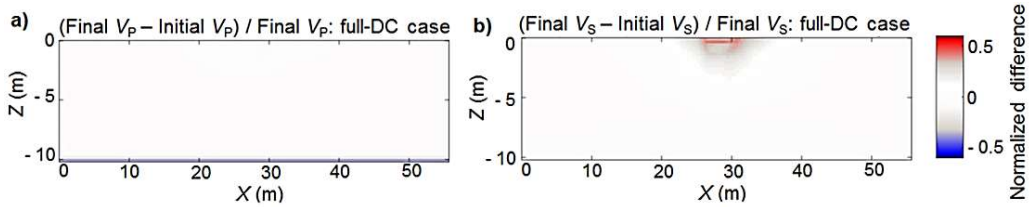

*Figure **A 6.11** – 2D sections (crossing the target's center) of the 3D model misfit: a) (Final V_P model – Initial V_P model)/Final V_P model. b) (Final V_S model – Initial V_S model)/Final V_S model.*

Figure **A 6.12** shows the data fitting comparison, for the same shot position (Shot 8, belonging to the 3D acquisition) and different acquisition configurations: case D and case C. In can be observed that better data fitting is obtained for the acquisition configuration of the case C (Figure **A 6.12b**) than for the acquisition layout used for case D (Figure **A 6.12a**).

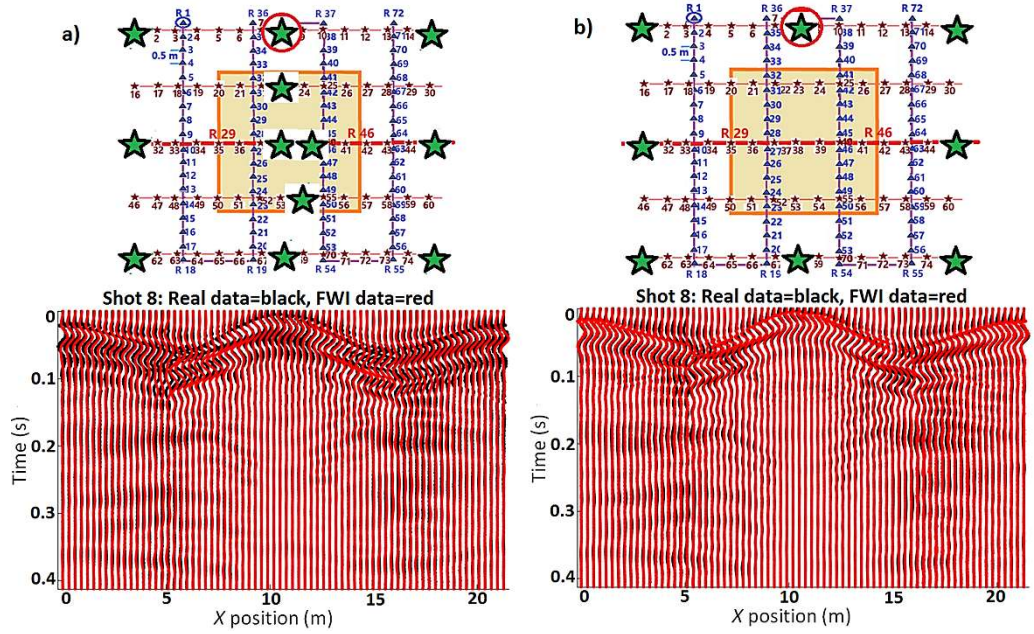

*Figure **A 6.12** – Trace-by-trace data fitting comparison after FWI, for the same shot position (Shot 8), over different acquisition geometries for the real data application: **a)** Case D, **b)** Case C. The best data fitting after FWI is obtained for the Case C.*

Chapter 9

Discussion

As an overall observation, when performing multi-parameter FWI on seismic data from near-surface acquisitions, V_S is updated more than V_P and this fact was proven by many FWI experiments (e.g. Butzer et al., 2013; Gross et al., 2017; Borisov et al., 2018; Smith et al., 2019; Wang et al., 2019; Trinh et al., 2019). This behavior is justified by the physics of SW propagation, more sensitive to the shear properties, and by the intrinsic lower resolution of V_P (higher velocity, larger wavelengths). A strategy aimed at better updating the V_P model may be the application of a constraint on the V_P to V_S ratio, presented in Trinh et al. (2018).

Another overall observation regards the predominantly shallow model update. On one side, the SW method is affected by the loss in resolution with depth. On the other side, since for near-surface applications the FWI gradient is highly dominated by SWs, its sensitivity at a depth greater than the SW propagating wavelengths is low. Therefore, detecting anomalies buried at depths greater than the SW penetration may be affected by loss in resolution. Some recent studies proved the effectiveness of FWI, based of SW analysis derived initial models, to detect anomalies buried at about 10 m depth (e.g. Wang et al., 2019; Smith et al., 2019).

As far as deep anomalies are concerned, other approaches can be adopted to increase the resolution at greater depths. Firstly, since the SW maximum investigation depth is controlled by the low-frequency band of the DC, data from passive seismic measurements can be integrated to extend the low-frequency band. Moreover, including higher modes in DC analysis may also improve the vertical resolution (Socco et al., 2010b). Another approach could be the f. a. traveltime tomography, although it potentially fails in detecting low-velocity zones. Regarding this last observation, Chen et al. (2017) showed a successful example on the use of the frequency-dependent traveltime tomography for building initial models for FWI, aimed at detecting a tunnel (made of concrete walls) buried at 1.6 m depth. Alternatively, other strategies can be integrated into the FWI workflow for enabling a deeper model reconstruction: i.e. a depth preconditioning of the gradient (e.g. He et al., 2018) or scaling the approximated Hessian matrix (e.g. Nuber et al., 2015).

All in all, for a data set dominated by SWs, since the V_P model is not significantly updated when using a simple multi-parameter FWI scheme, a very accurate initial model is necessary. A possible approach to obtain the V_P model is the first-arrival tomography. Preliminary investigations made in this study on synthetic data (Figure **9.1** and **9.2**) provided accurate results for the higher-velocity layers, although a slight overestimation in velocity still exists in the deepest region (Figure **9.2c**). However, the velocity values are considerably underestimated in the shallow part of the model, down to 2 – 2.5 m depth (Figure **9.2b** and **9.2c**). These results suggest that, in the presence of low-velocity areas, the DC analysis may provide a better initial V_P model for FWI compared with the f. a. traveltime tomography.

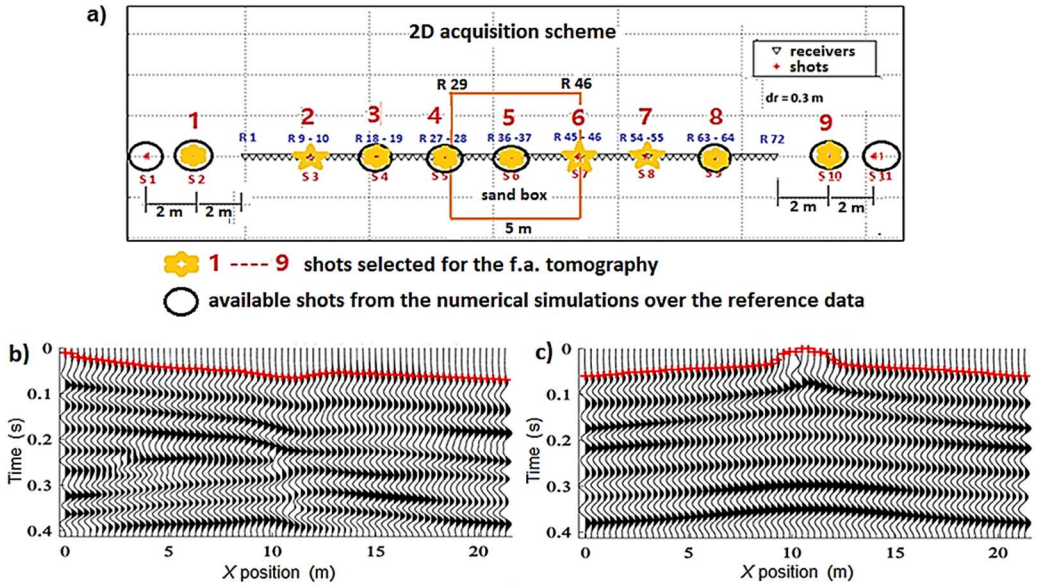

*Figure **9.1** – Example of the first arrivals picking on the synthetic reference data: **a**) Scheme of the shot positions, **b**) Shot 2, **c**) Shot 5. A 0.03s AGC window was applied to the seismograms to emphasize BWs visualization.*

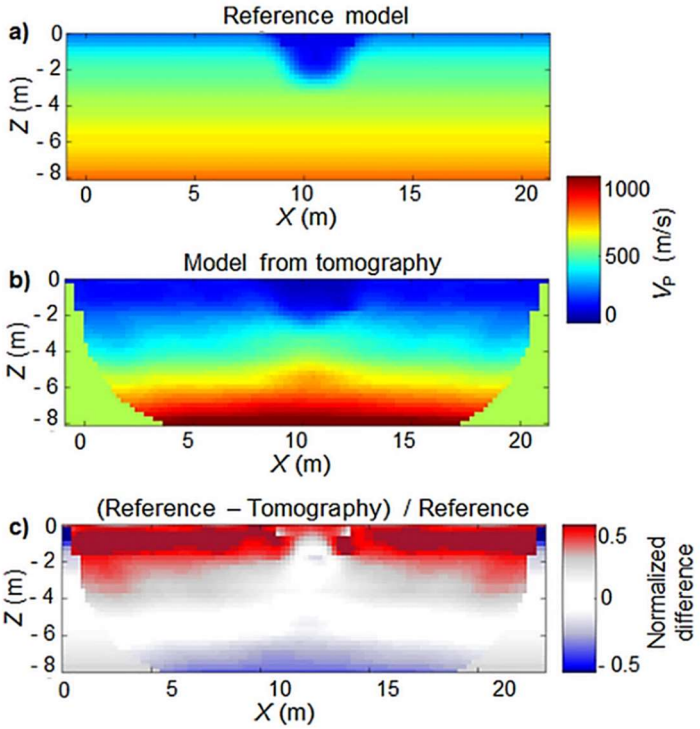

*Figure **9.2**. **a**) Vertical section of the 3D reference synthetic V_P model. **b**) V_P model obtained from the f. a. traveltime tomography on synthetic data. **c**) Model misfit computed as the normalized difference between the reference model and the model obtained from f. a. traveltime tomography.*

The same observation arises also when analyzing the P-wave velocity model obtained by applying the f. a. traveltime topography on real data (Figure **9.3a**). The minimum P-wave velocity of the model retrieved from f. a. traveltime tomography on real data is about 107 m/s, while the corresponding minimum P-wave velocity of the model retrieved from SW analysis is about 142 m/s. For the synthetic case, we noticed that the shallow velocity is underestimated when using the f. a. traveltime tomography. This observation leads to the conclusion that, as for the synthetic case, when applying the f. a. traveltime tomography on real data the velocity of the shallow layer is underestimated as well.

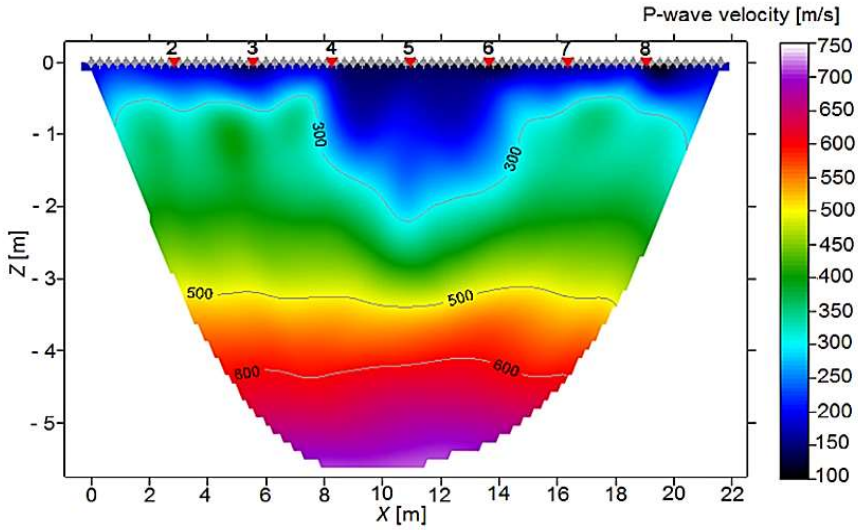

*Figure **9.3**. V_P model obtained from f. a. traveltime tomography application on real data (courtesy of Prof. Cesare Comina).*

***

An observation particularly related to this study is the achievement of FWI results characterized by different resolution when starting the inversion from a laterally variable or laterally homogeneous initial model. Figure **9.4a** shows the evolution of the L2 misfit function as the iterations proceed, while in Figure **9.4b**, the error related to each iteration is normalized by the error of the first iteration. The criterion used to stop each FWI experiment is related to the misfit function evolution: When the misfit function no longer decreases for more than two consecutive iterations, the iterative process stops.

Figure **9.4a** indicates lower overall misfit values (of one order of magnitude) for the full-DC cases (green and red curve) compared with the single-DC cases (blue and black curves) on both synthetic and real data. Therefore, a laterally variable initial model, derived from the analysis of the entire set of DCs, guarantees a lower initial data misfit and better convergence (Figure **9.4b**, green curve). The final values of the normalized misfit for the full-DC case (0.64 for real data and 0.12 for the synthetic data) are relatively similar to the ones obtained by other 3D FWI tests, performed for shallow environments at similar scale (e.g. Tran et al., 2019).

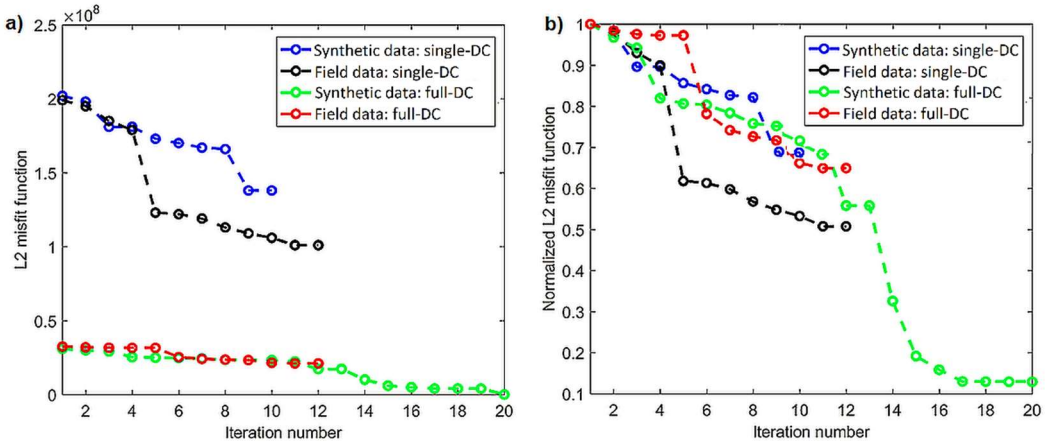

*Figure **9.4** – Data misfit versus iterations number for the 2D acquisition related cases: **a)** L_2 misfit function. **b)** Normalized L_2 misfit function.*

The L2 misfit related to the real data for the full-DC case (Figure **9.4b**, red curve) decreases less because the initial model is already accurate and the 2D acquisition geometry does not allow for further improvement. In the previous section, we noticed how the initial models from the real-data full-DC analysis are not significantly updated after FWI, while the initial models (particularly the V_S) from the single-DC analysis are correctly updated. These observations may suggest that an acceptable reconstruction of V_S could be obtained both from FWI (starting with a single-DC based initial model) and from the proposed full-DC analysis, without passing through the inversion step (e.g. Figure **8.16b**). However, results characterized by a higher resolution are expected from an FWI workflow based on higher frequencies.

***

A more elaborated FWI workflow, based on parameter selection and data-oriented strategies, potentially leads to better convergence. This study proved how a preliminary monoparametric (V_S) FWI helps to obtain a lower data misfit after the multi-parameter FWI (Figure **9.5**, black curve). Nevertheless, the success of this strategy might be case dependent, and it may not work properly in the presence of complex topography (although some applications related to preferential inversion of SWs to update the shallow V_S model showed promising results: Borisov et al., 2020). Anyhow, when irregular topography is present, the scattering engine behind FWI may be better solved, since the lateral variations of velocity may cause back-scattered SWs and the conversion to higher modes. Moreover, SWs propagation under a complex topography, generating secondary BWs, may help for the illumination of the deeper regions of the model.

The application of a two-step FWI (experimented also in other studies: e.g. Trinh et al., 2018, 2019), which exploits the BWs content during a preliminary early arrival inversion, guaranteed a better data fitting at far-offset. Nevertheless, the overall normalized data misfit is higher (Figure **9.5** red curve) than in the previous case (model strategy - black curve), since the data set is dominated by SWs, while the BWs amplitude is very weak. However, these results may be case dependent as well, and the above data strategy may work properly for another data set, characterized by BWs with higher amplitudes.

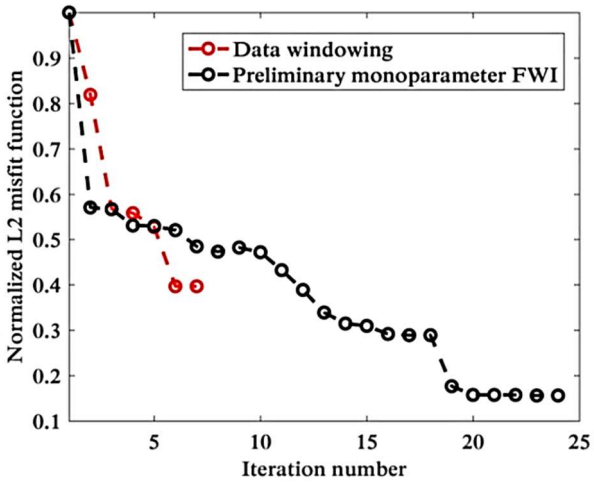

*Figure **9.5** – Normalized L_2 misfit function versus iteration number for the synthetic test, full-DC initial model, when using a model-strategy and a data-strategy, respectively.*

***

Further, the use of a 3D acquisition layout has led to different results after FWI, according to the variation of the source number and source positions. In Figure **9.6**, the blue, magenta and green curves represent the data-misfit evolution during FWI, when 8 sources, 12 sources and 16 sources, respectively, are placed in correspondence of the low-velocity anomaly. On the other side, the black and red curves are representative for the case in which the sources are activated near the target's boundaries (black) or at a greater distance from the target (red).

A preliminary observation related to this experiment is the higher influence of the source position than the influence of the source number. In particular, increasing the source number from 8 to 12 or 16 (Figure **9.6**) did not ensure a significantly lower data misfit after FWI, even if a slight decrease of this value occurred. Similar behavior was also described by Smith et al (2019), who concluded (after testing an acquisition scheme based on 24 vertical sources) that 6-10 shots are enough for detecting a tunnel buried at 10 m depth. Wang et al. (2019) also noticed that increasing the number of shots does not lead to significant improvement in the FWI results.

Differently, the experiments conducted for this study proved that the source position is important. There are few examples in the literature related to the importance of the source positions (e.g. Nuber et al., 2017). In the present study, we can notice that activating the sources outside the low-velocity target (or close to the boundaries) guarantees better data fitting after FWI (Figure **9.6**, black and red curves). Overall, better results are obtained when positioning the sources at a longer distance from the target (red curve). The reason for this pattern may be related to the small-wavelength phases scattered from the target's boundaries that play, in this case, a minor role in the misfit function computation than in the case when positioning the sources near the target (black curve).

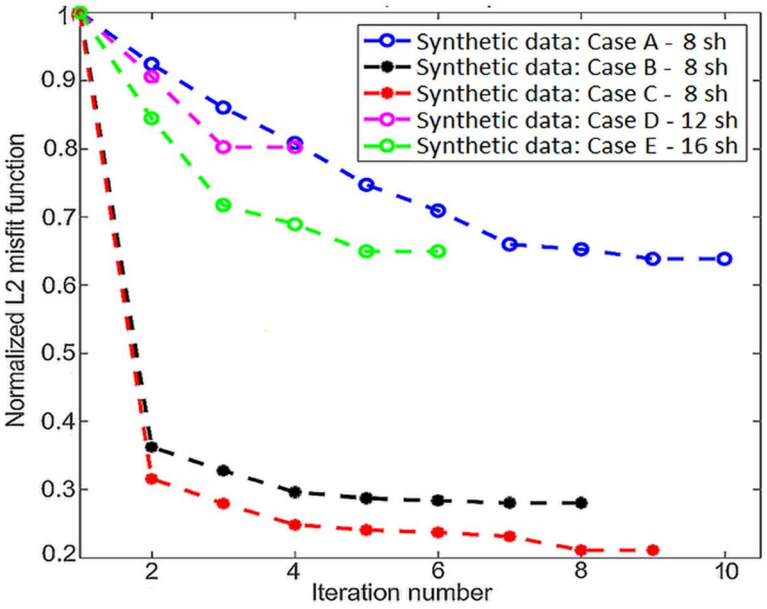

*Figure **9.6** – Normalized L_2 misfit function for the various synthetic applications related to the 3D acquisition design.*

As far as the initial model's effectiveness is concerned, when using a 3D acquisition geometry better FWI results were obtained, again, for the full-DC analysis based initial model (green and red curves in Figure **9.7**), both for the synthetic experiment and field data application. The final normalized data misfit is, in these cases, 0.21 and 0.26, respectively. The final data misfit is almost one magnitude order higher when staring the FWI from the laterally homogeneous models (blue and black curves) than when starting the FWI from the laterally variable ones (green and red curves).

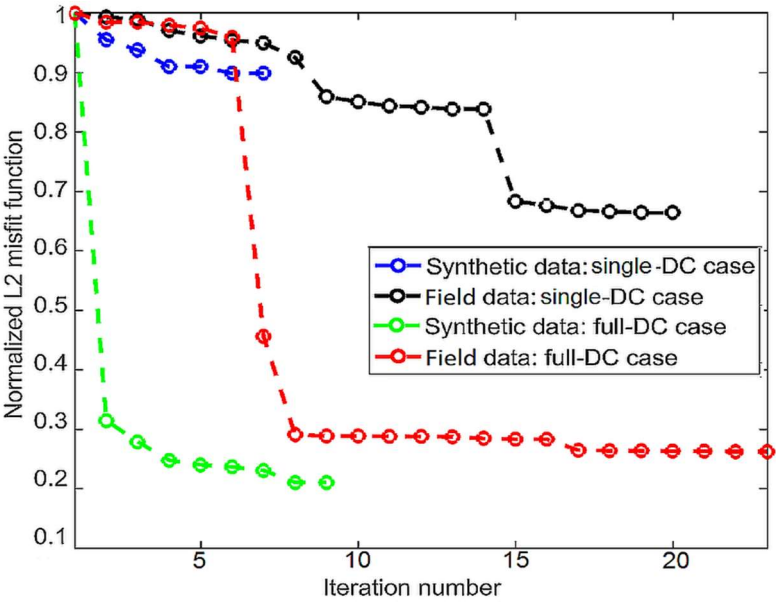

*Figure **9.7** – Normalized L_2 misfit function for the single-DC and full-DC cases, real and synthetic applications.*

***

Another aspect that requires analysis is the engine used in this study for simulating the wave propagation. As mentioned in Section 8.2, a viscoelastic modeling engine was used, while the values of the density and quality factors were kept invariable during inversion, to focusing the analysis on the velocity models reconstruction while avoiding eventual trade-off between the velocity and secondary parameters (i.e. quality factor and density).

However, the Q parameter is important for properly modeling the frequency-dependent amplitude decay with offset. Therefore, using a wrong Q value may lead to a wrong amplitude of the modeled data at far-offset. For the real data application, an inaccurate Q value may also affect the accuracy of the source wavelet estimation through deconvolution, leading to the necessity of introducing additional corrections (e.g. Pratt, 1999; Gross et al., 2014).

As presented in Section 5.4, the attenuation is implemented in the SEM46 tool through a constant SLS mechanism over a limited frequency band. The use of a constant Q value should not influence significantly the results since the frequency range involved in the experiments is also relatively limited (3 Hz – 40 Hz) and the number of propagated wavelengths is low (approximately 4 wavelengths propagated over a maximum offset of about 25 m). The presence of higher frequencies may require a more correct representation of the Q-values. In any case, more accurate results are expected when using variable (and accurate) Q models in the forward modeling scheme (e.g. Gross et al., 2014; Borisov et al., 2020). Moreover, better results would be obtained if applying amplitude and phase filters to the simulated data (Baumstein et al., 2011) or, better, inverting iteratively also for Q in addition to elastic parameters (Malinowski et al., 2011; Operto and Miniussi, 2018; Pan and Innanen, 2019).

As far as the density model is concerned, assuming a constant density during FWI is reasonable when the dominant energy propagates as SWs, which have low sensitivity to this parameter (Nazarian, 1983). Many studies proved that FWI guided by SWs is not particularly sensitive to density (e.g. Dokter et al., 2017; Borisov et al., 2018; Smith et al., 2019). Moreover, an additional trade-off is introduced when FWI aims at reconstructing simultaneously the velocity model and the density (e.g. Krampe et al., 2019). Anyhow, considering that V_P and density have the same radiation pattern for short aperture angles, an independent reconstruction of these parameters is challenging for short-offset data (e.g. Virieux and Operto, 2009).

Some results of preliminary forward modeling investigations conducted using variable values for density and quality factors are presented in *Appendix* 7.

Chapter 10

Conclusions and Perspectives

An integrated workflow, based on 2D surface wave analysis and 3D elastic full-waveform inversion, was tested, to enhance the resolution in the imaging of complex-shaped shallow targets.

The tests were based on a well-known real target, represented by a loose-sand body buried among more compact sediments, and on synthetic data simulating this geological asset. The seismic data are dominated by complex-scattered and highly energetic SWs, making the application of FWI a challenging task. Additional challenges for FWI are created by the flat topography of the investigated area, which facilitates the SWs forward scattering regime while FWI is mainly based on a backscattering one (BWs). The presence of irregular topography would have fulfilled better the engine behind FWI since it would have generated secondary BWs and SWs with higher modes of propagation, but it would have rendered the SWs fundamental mode's individuation more challenging.

Two different series of initial V_P and V_S models, laterally homogeneous and laterally variable, have been retrieved using innovative 2D SW analysis procedures. The laterally homogeneous models were obtained by inverting or transforming one single DC (representative for the background medium), while for building the laterally variable models, both a clustering algorithm and data transform have been applied to all the DCs extracted from the data.

A simple 3D elastic FWI workflow, tested on synthetic and real data, improved the initial models' resolution. In particular, the laterally variable initial model has granted better FWI results than the laterally homogeneous one. When using a simple FWI workflow, the final data fitting is accurate for the near-offset traces and in correspondence of the low-velocity target, while the fitting of some far-offset arrivals and back-scattered phases can still be improved. One way to achieve such improvement is integrating, into the basic FWI workflow, different model-based and data-oriented strategies. Thus, some of these strategies have been tested.

Firstly, the SWs preferential sensitivity to the shear properties was exploited, through a preliminary monoparametric (V_S) inversion step, to improve the initial V_S model's resolution in the shallow part. This strategy led to an overall better model reconstruction after the multi-parameter (V_S & V_P) FWI. However, the accuracy of the results may be case dependent, and it may vary in the presence of complex topography.

Apart from the offset weighting strategy, which proved to be very important for the 2D acquisition geometry (characterized by a longer offset), additional data-oriented strategies have been tested. Among them, a time-windowing strategy, aimed at exploiting the early arrivals content during a preliminary inversion step, ensured a better data fitting at far offset after FWI. Nevertheless, the V_P model did not improve significantly, due to the very

weak amplitude of BWs. However, the above strategy may work efficiently for a different data set, characterized by BWs with higher amplitudes. Successful examples regarding the windowing strategy can be found in the literature.

Another experimented data strategy was the ordinarily employed multiscale inversion, which constraints and guides the model update according to the wavelength resolution. Although it provided better data-fitting and model reconstruction after FWI, the difference between the results of the multiscale FWI and the results of the FWI conducted directly in the full frequency band is not significant in this study. This happens because an accurate data fitting, without cycle-skipping for the most part of the seismograms, is already present in the starting configuration related to the full-DC based model. However, many published results proved the effectiveness of this strategy.

Moreover, FWI experiments conducted using a 3D acquisition scheme led to different results, depending mainly on the source position. In particular, the position of the sources with respect to the low-velocity target resulted more important than the number of sources. Indeed, increasing the source number from 8 to 16 did not influence substantially the reconstructed models' resolution, even if results characterized by a slightly lower data misfit were obtained. On the contrary, moving the sources outside the low-velocity anomaly ensured a better reconstruction of the target's boundaries as well as a significantly lower data misfit after FWI. While there are many examples in the literature related to the influence of the number of sources used during FWI, the influence of the source position when using a 3D acquisition layout was not investigated very often.

In this work, considering only the experiments with sources activated outside the target or close to the boundaries, the configuration based on sources placed at a greater distance from the low-velocity anomaly provided better results. A possible explanation for this path could be related to the small-wavelength phases scattered from the target's boundaries, difficult to be fitted with a maximum 40 Hz FWI scheme. Indeed, these phases play an important role in the misfit function computation when the sources are activated near the target, leading to a higher data misfit after FWI. Moreover, the sources located near the low-velocity target contribute with high amplitudes and no data normalization acts during the misfit function computation.

Differently, when positioning some sources inside the low-velocity target, a lower lateral resolution is achieved in correspondence of the target's boundaries. However, in this case, better recovery of the shallow S-wave velocity value is obtained. On the contrary, when placing the sources outside the target, the better lateral resolution is obtained at the expense of a slight overestimation of the shallow velocity. The reason for this fact may be the lower information related to the shallow part of the target since no source is probing it appositely.

Overall, it may be concluded that the integration of 2D SW analysis (providing smooth laterally variable initial models) with a relatively simple 3D FWI scheme allows reconstructing complex-shaped shallow targets with high accuracy. This work proved that using laterally variable initial models from SWs analysis allows obtaining more accurate results after FWI while efficiently assisting the convergence. The integration of model-oriented and data-based strategies into the main FWI workflow enables an additional improvement of the results. Using a 3D acquisition layout (with a particular position of the sources) allows for precise reconstruction of the target's shape and lateral extension.

***

Future investigations could be focused on:

- The consideration of higher modes of SWs propagation in the initial model building step, to guarantee a greater investigation depth and improve the vertical resolution.
- The analysis of horizontal particle displacement components in the seismogram.
- The use of constraints on the V_P to V_S ratio during FWI, to guide the V_P model's update.
- The use of higher frequencies for FWI to increase the overall wavelength resolution.
- The use of variable values for the Q parameter and density over the 3D model. Some preliminary results are presented in *Appendix* 7.

Appendix 7

3D viscoelastic modeling using variable values for Q parameters and density: preliminary results

Preliminary 3D modeling investigations have been conducted on real data, in the full-frequency band (up to about 150 Hz), on the use of variable versus constant values for density and quality factors (Figure **10.1**). The experiments followed a 2D acquisition pattern (Figure **A 10.1**) and regarded 4 different configurations (Figure **A 10.2**):

A. constant quality factors and constant density (results plotted in red)
B. variable quality factors and variable density (results plotted in green)
C. variable quality factors and constant density (results plotted in cyan)
D. constant quality factors and variable density (results plotted in magenta)

The variable density model was computed using the relation $\rho = 0.5 \cdot V_s^{0.22} \cdot 10^3$ while the attenuation parameters were retrieved using the relations: $Q_S = 0.09 \cdot V_S$; $Q_P = 1.5 \cdot Q_S$. For the experiments based on constant values for the density and quality factors, the former is 1800 kg/m^3 while the Q factors are equal to 40. The source time function was estimated from the field data, in the full-frequency band, using the deconvolution technique (Pratt, 1999). The Green's function was computed in the initial model using a Ricker source with a central frequency of 60 Hz. All simulations were performed using the SEM46 code and exploiting the domain composition (2 x 6 x 6 in the z – x and y direction) on the parallel computing environment of the UGA Froggy HPC architecture (using 72 cores). A forward simulation for one shot took about 7 hours and required a computer memory of 25.6 GB.

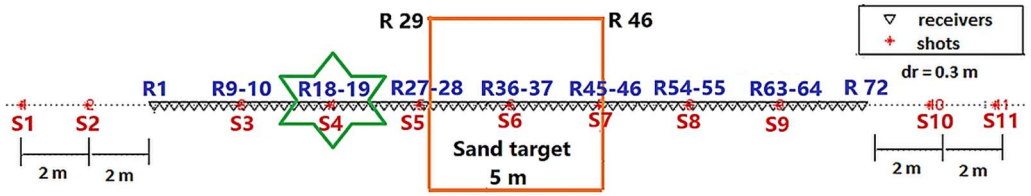

*Figure **A 10.1** – Scheck of the acquisition design. The shot positions are signed with sumbols and numbers from S 1 to S 11. The green star marks the position of the shot gathers presented hereafter.*

In Figure **A 10.2** the reader can notice how the lowest data misfit was obtained when using variable values for the Q parameters and constant values for the density, while the highest data misfit occurred when using variable density and constant values for Q parameters. Therefore, the use of constant density may be preferred to the use of a probably wrong variable density value that, in our case, may not fulfil properly the variation inside and outside the sand-target. Besides, the use of variable Q is very important for appropriate amplitude modeling.

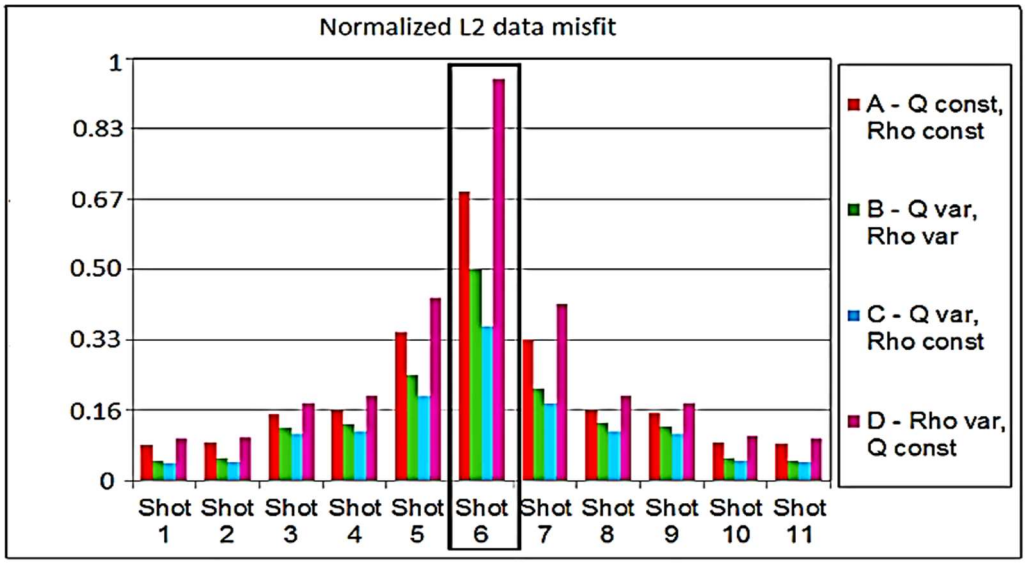

*Figure **A 10.2** - Normalized L2 misfit function for all tested cases and different shots position along the 2D acquisition direction: **A**) Constant Q and constant density, **B**) variable Q and variable density, **C**) variable Q and constant density, **D**) variable density and constant Q.*

Henceforth, some examples of trace-by-trace data fitting comparison are presented for all tasted cases, in a frequency band up to about 150 Hz. The chosen shot gather is placed in an intermediate position between the homogeneous background and the low-velocity target (Figure **A 10.1**). All seismograms are trace-by-trace normalized.

Figures **A 10.3, A 10.4** and **A 10.5** show some examples of trace-by-trace data fitting comparison for case A, related to the use of constant values for both the Q parameters and density.

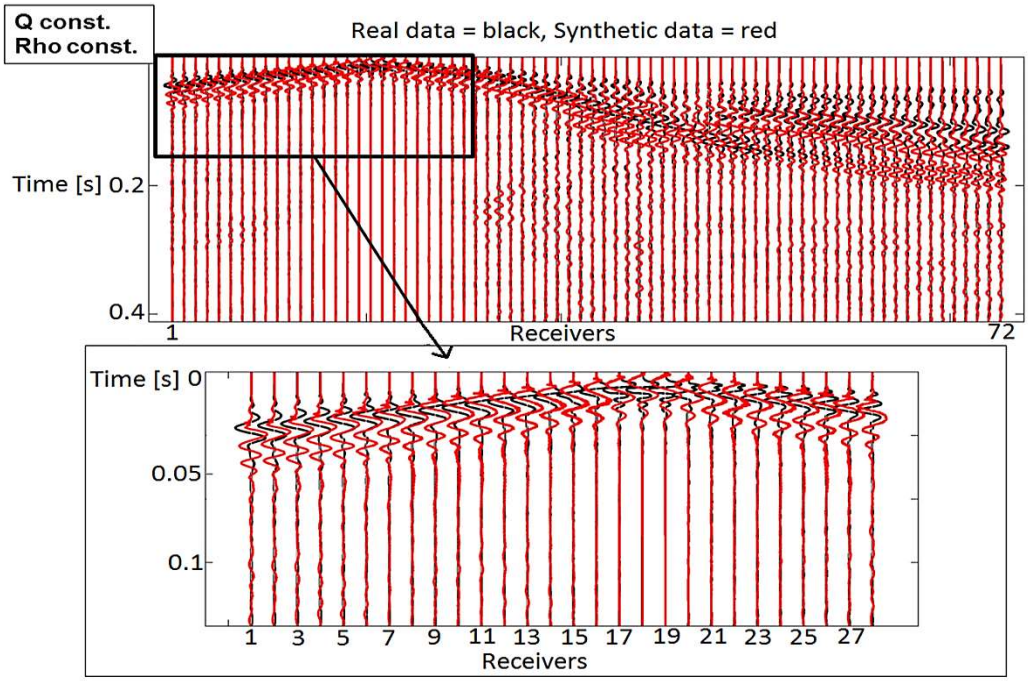

*Figure **A 10.3** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using constant values for Q parameters and density: Shot 4, near-offset traces.*

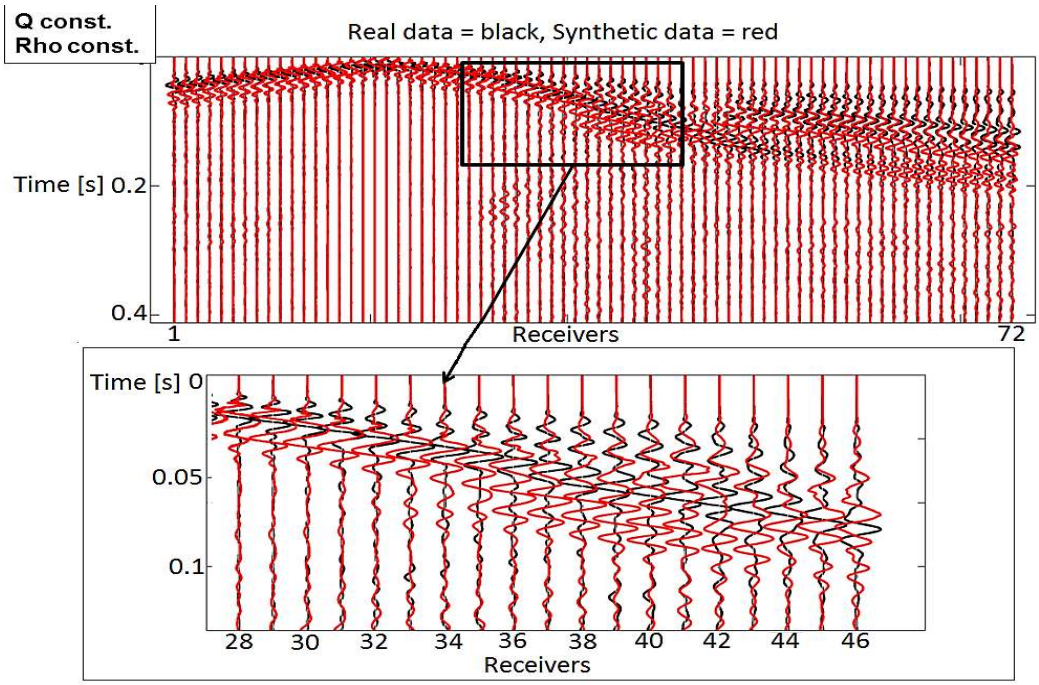

*Figure **A 10.4** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using constant values for Q parameters and density: Shot 4, intermediate-offset traces.*

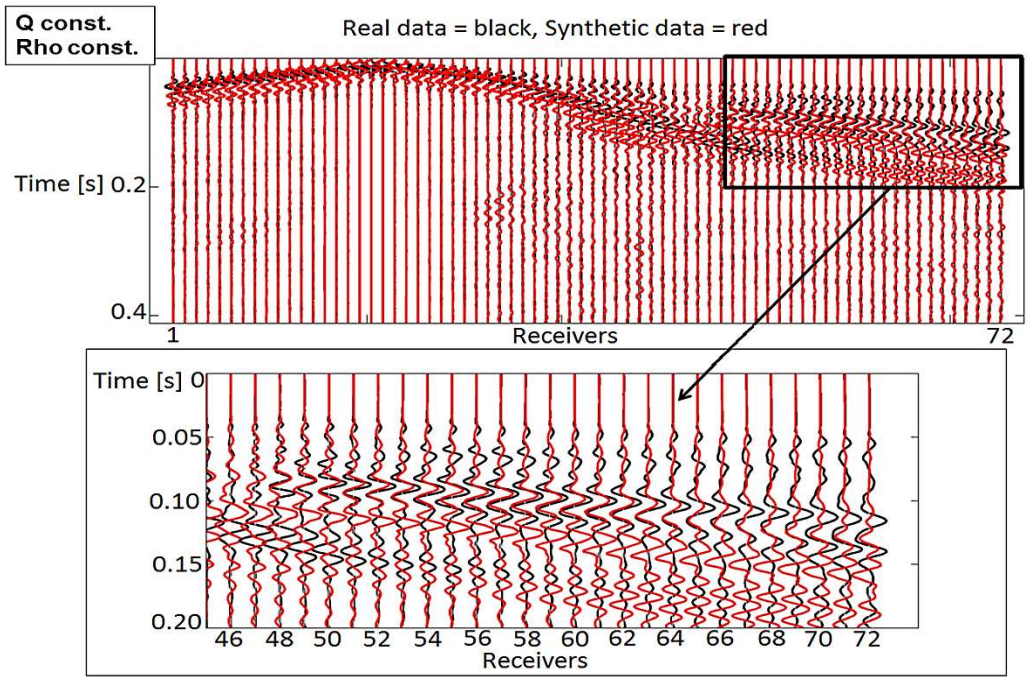

*Figure **A 10.5** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using constant values for Q parameters and density: Shot 4, far-offset traces.*

Figures **A 10.6, A 10.7** and **A 10.8** show some examples of trace-by-trace data fitting comparison for case B, related to the use of variable values for Q parameters and density.

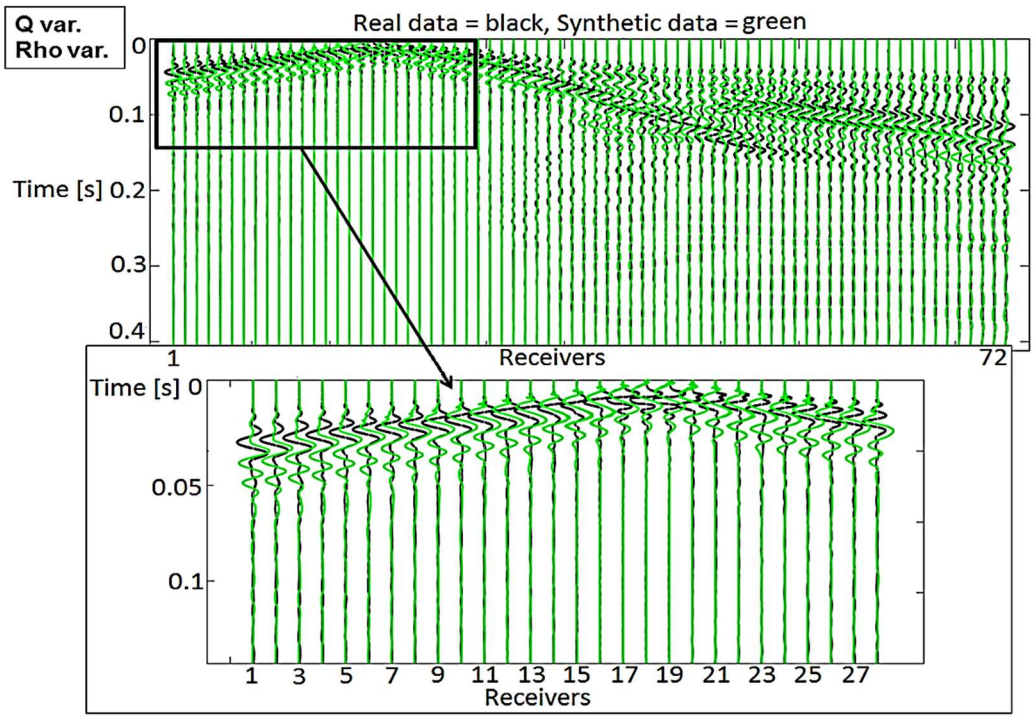

*Figure **A 10.6** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using variable values for Q parameters and density: Shot 4, near-offset traces.*

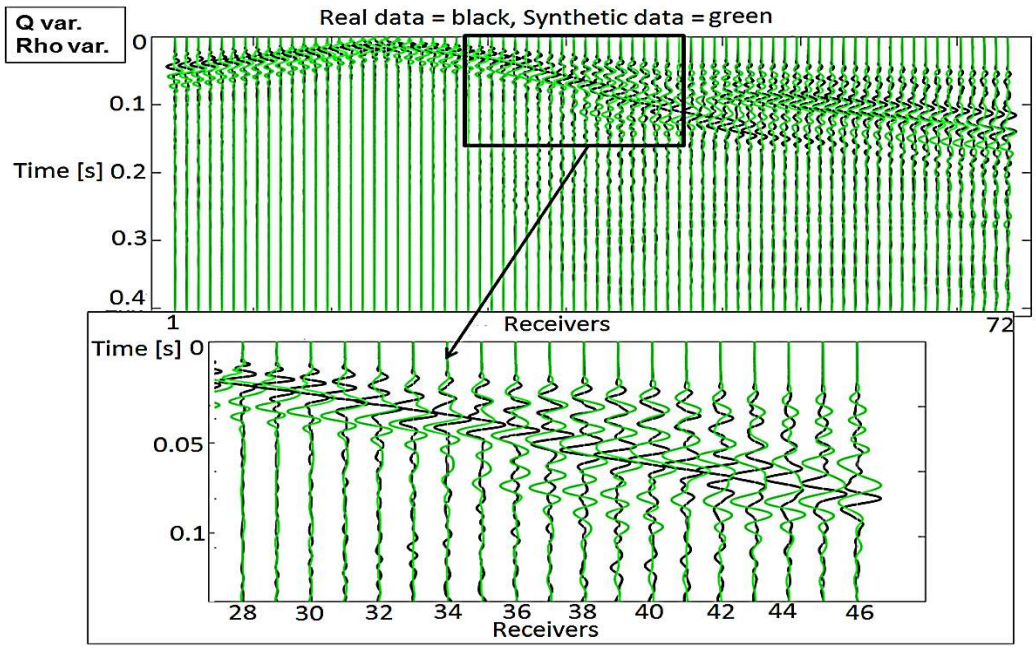

*Figure **A 10.7** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using variable values for Q parameters and density: Shot 4, intermediate-offset traces.*

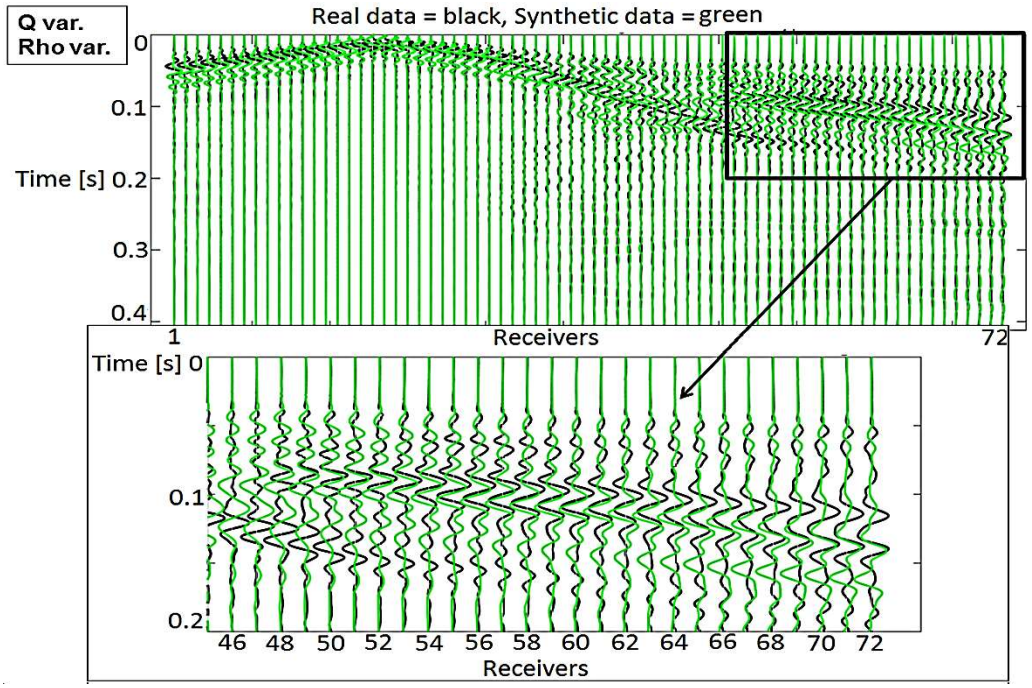

*Figure **A 10.8** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using variable values for Q parameters and density: Shot 4, far-offset traces.*

Figures **A 10.9, A 10.10** and **A 10.11** show some examples of trace-by-trace data fitting comparison for case C, related to the use of variable values for Q parameters and a constant value for density.

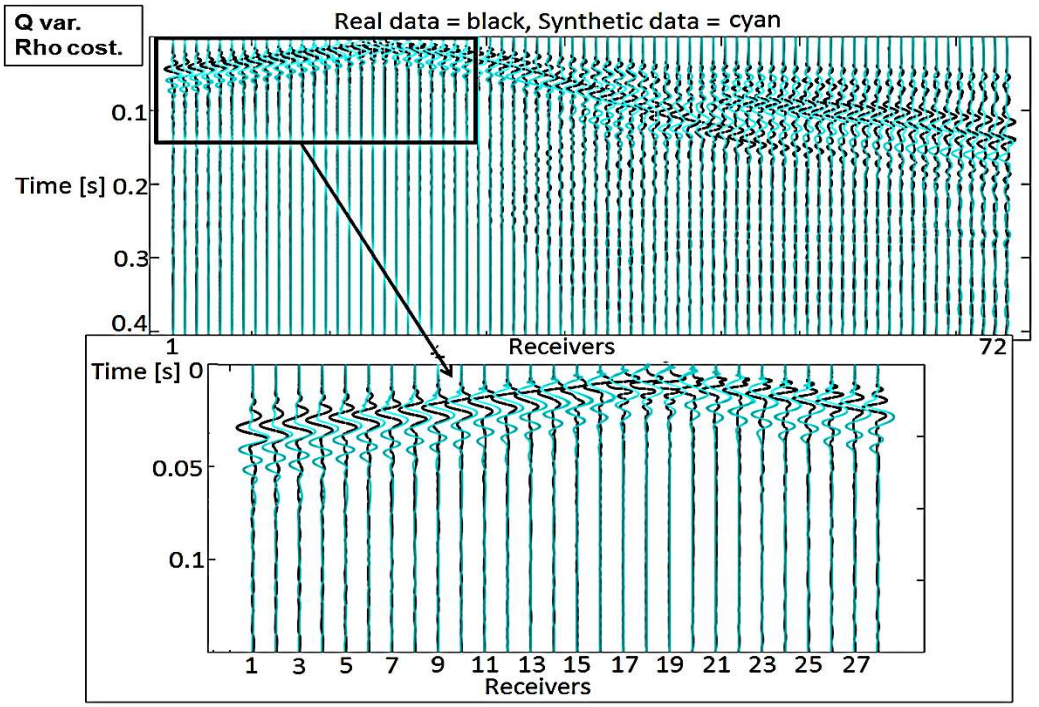

*Figure **A 10.9** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using variable values for Q parameter and a constant value for density: Shot 4, near-offset traces.*

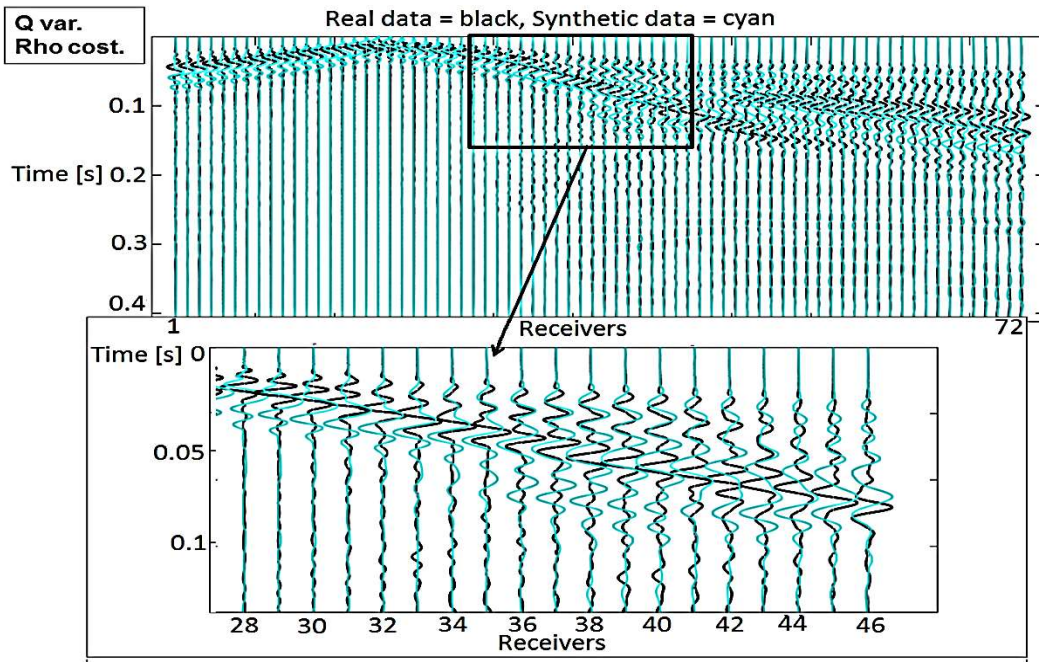

*Figure **A 10.10** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using variable values for Q parameters and a constant value for density: Shot 4, intermediate-offset traces.*

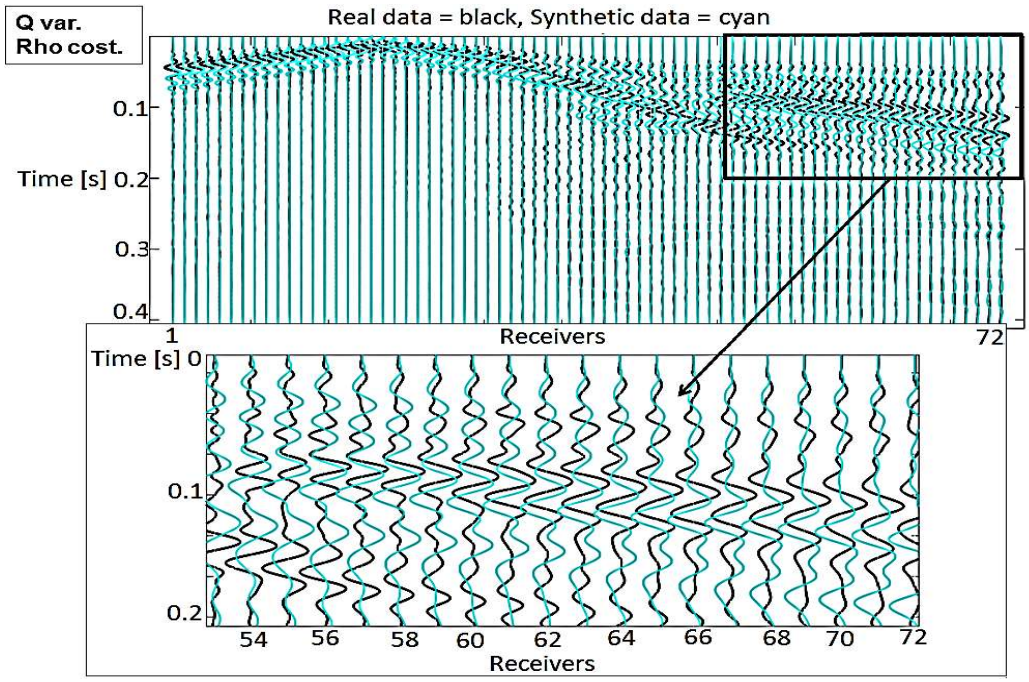

*Figure **A 10.11** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using variable values for Q parameters and a constant value for density: Shot 4, far-offset traces.*

Figures **A 10.12, A 10.13** and **A 10.14** show some examples of trace-by-trace data fitting comparison for case D, related to the use of a constant value for Q parameters and variable values for density.

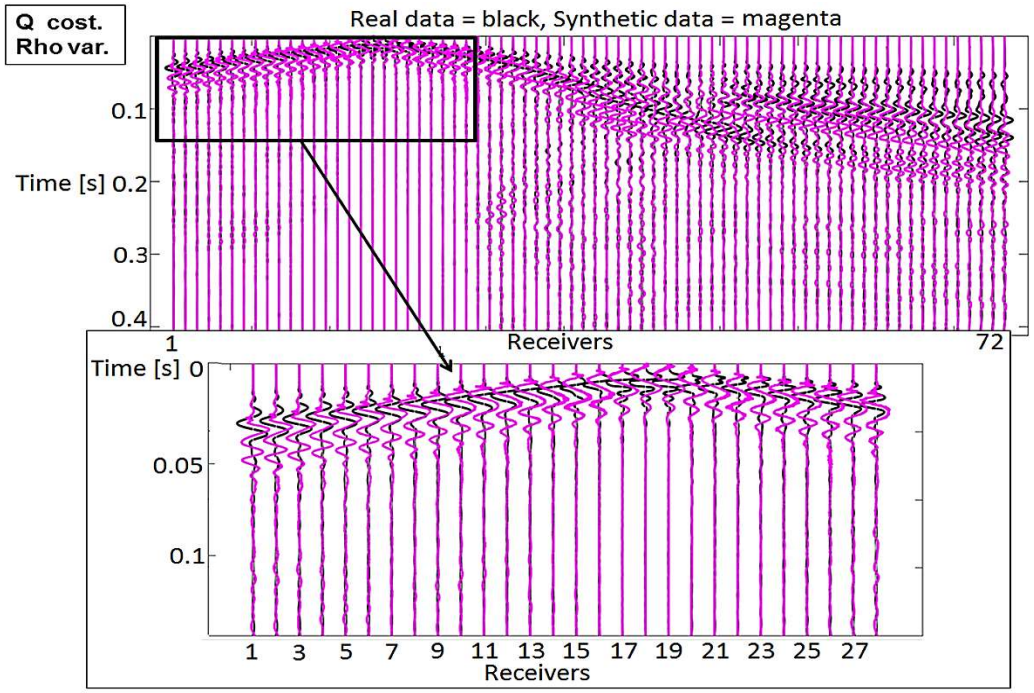

*Figure **A 10.12** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using a constant value for Q parameter and a variable value for density: Shot 4, near-offset traces.*

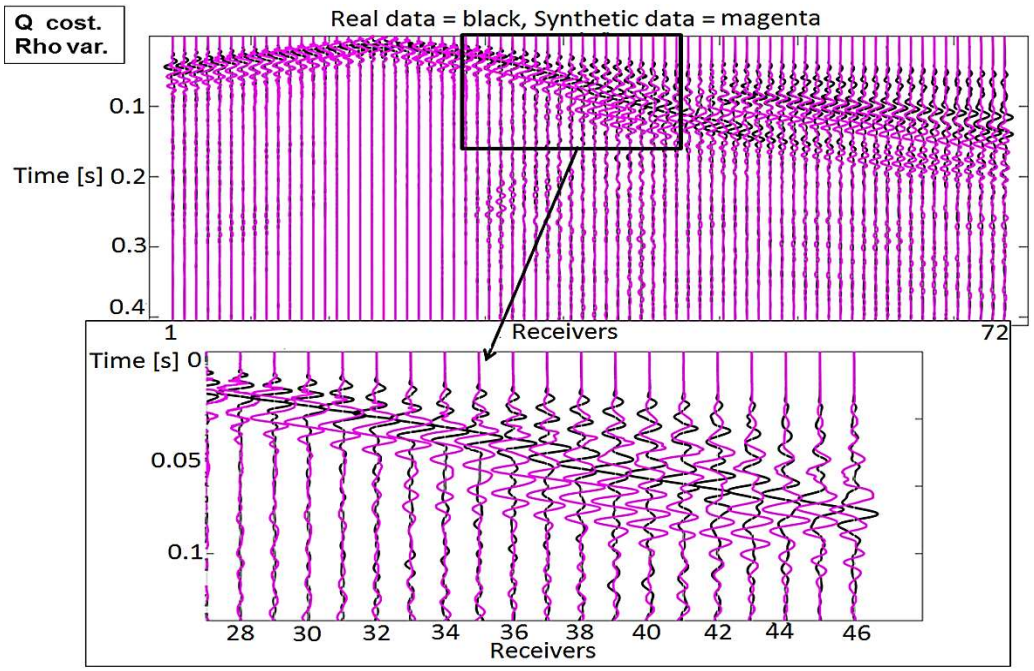

*Figure **A 10.13** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using a constant value for Q parameter and a variable value for density: Shot 4, intermediate-offset traces.*

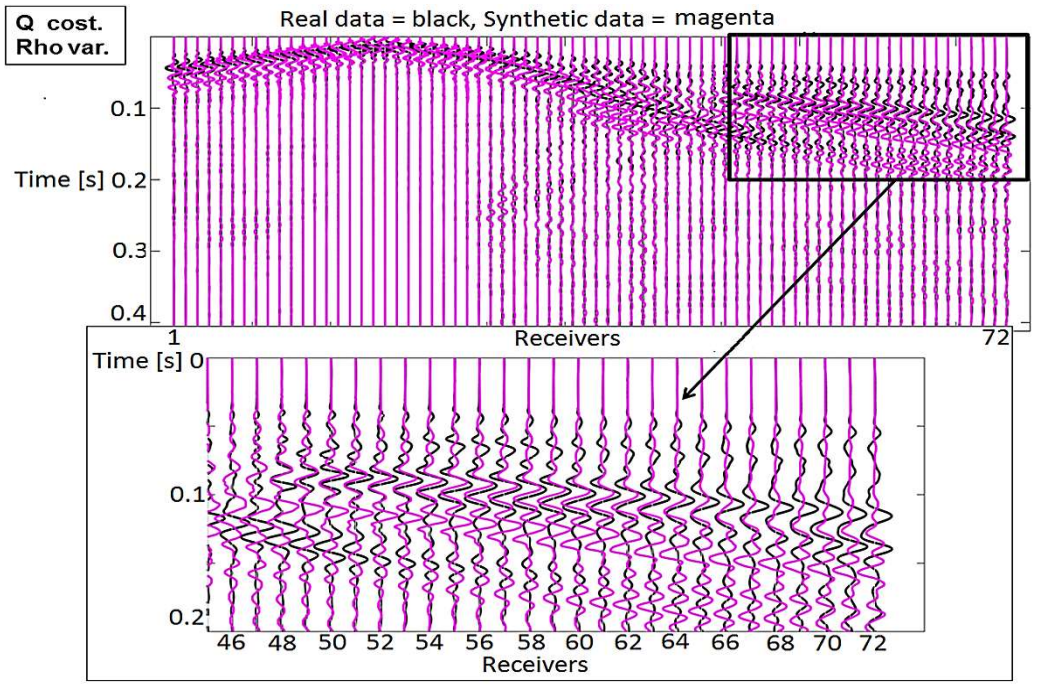

*Figure **A 10.14** – Trace-by-trace data fitting comparison between real data and data belonging to the initial model from full-DC analysis, when using a constant value for Q parameter and a variable value for density: Shot 4, far-offset traces.*

Appendix 8 - Numerical details

The relations used for computing the element size, time sampling and the number of time iterations (considering 5 GLL nodes per shortest wavelength) are (Trinh et al., 2019):

$$element\ size \leq \frac{\lambda_{min}}{1.3} \leq \frac{minumum\ velocity / maximum\ frequency}{1.3}\ (\text{m})$$

$$time\ sampling \leq \frac{0.3 \cdot 0.17 \cdot element\ size}{maximum\ velocity}\ (\text{s})$$

The number of GLL points is = (mesh + absorbing sponges a.s.) x 4 + 1.

Table A 8.1 gathers values of the element size, time sampling, time iterations and number of GLL points related to different frequency intervals used in this study.

Min. V_S ~ 60 m/s Max. V_P ~ 1050 m/s Acquisition ~ 0.412 s	Element size	Time sampling	Time iterations
Max. freq. = 40 Hz	1.0 m	4.8e-05	8 600
Max. freq. = 50 Hz *	0.9 m	4.3e-05	9 600
Max. freq. = 150 Hz	0.3 m	1.4e-05	29 400
Max. Freq. = 500 Hz	0.1 m	4.8e-06	85 800
GLL points (40 Hz) z-x-y direction 4 a.s. & a free surface	GLL points (50 Hz) z-x-y direction 4 a.s. & a free surface	GLL points (150 Hz) z-x-y direction 4 a.s. & a free surface	GLL points (500 Hz) z-x-y direction 4 a.s. & a free surface
57 – 253 – 253	61 – 281 – 281	149 – 777 – 777	421 – 2265 – 2365

Table A 8.1 – Vales for the element size, time sampling, time iterations and GLL points.

The necessary time for the simulations, conducted on the parallel computing architecture using 8 nodes with 16 cores each, is reported in Table A 8.2.

8 shots 128 cores	Forward modeling	First gradient computation	FWI – 10 iterations
Max. freq. = 40 Hz	4 minutes	11 minutes	~ 12 hours
Max. freq. = 50 Hz *	5 minutes	18 minutes	~ 20 hours
Max. freq. = 150 Hz	40 hours	↗	↗

Table A 8.2 – The time registered for forward modeling, 1st gradient computation and FWI.

The necessary computer memory for the simulations is reported in Table A 8.3.

8 shots 128 cores	Forward modeling	FWI – 10 iterations
40 Hz	4.1 GB	9.8 GB
50 Hz *	5.3 GB	13.0 GB
150 Hz	204.8 GB	↗

Table A 8.3 – Estimated computer memory

* Preliminary FWI experiments have been conducted also in the frequency band 3 Hz – 50 Hz but the results were not presented in this manuscript.

References

Achenbach, J. D. (1984). Wave Propagation in Elastic Solids. Amsterdam, The Netherlands: North-Holland Publishing.

Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors: Bulletin of the Earthquake Research Institute, **35**, 416–456.

Aki, K. and P. G. Richards (2002). Quantitative seismology, 2nd edition: University Science Books, Mill Valley, California.

Alford, R. M., K. R. Kelly and M. Boore (1974). Accuracy of finite-difference modeling of the acoustic wave equation: Geophysics, **39**, 834-842.

Amrouche, M. and H. Yamanaka (2015). Two-dimensional shallow soil profiling using time-domain waveform inversion: Geophysics, **80**, EN27–EN41, doi: 10.1190/geo2014-0027.1

Amundsen, L. (1991). Comparison of the last-squares criterion and the Cauchy criterion in the frequency-wavenumber inversion: Geophysics, **56,** 1027-1039.

Amundsen, L. and B. Ursin (1991). Frequency-wavenumber inversion of acoustic data: Geophysics, **56,** 1027-1039.

Anderson, J. E., L. Tan and D. Wang (2012). Time-reversal checkpointing methods for RTM and FWI: Geophysics, **77**, no. 4, S93–S103, doi: 10.1190/geo2011-0114.1

Askan, A., V. Akcelik, J. Bielak and O. Ghattas (2007). Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Bulletin of the Seismological Society of America, **97**, 1990–2008, doi: 10.1785/0120070079

Askan, A. and J. Bielak (2008). Full anelastic waveform tomography including model uncertainty: Bulletin of the Seismological Society of America, **98**, 2975-2989.

Athanasopoulos, N. and T. Bohlen (2017). Field data application of sequential full-waveform inversion of refracted and Rayleigh waves: 79th EAGE Annual Conference and Exhibition, Extended Abstracts.

Athanasopoulos, N., E. Manukyan, T. Bohlen and H. Maurer (2018a). Accurate reconstruction of shallow P-wave velocity model with time-windowed elastic full-waveform Inversion: 80th EAGE Annual Conference and Exhibition, Extended Abstracts.

Athanasopoulos, N., E. Manukyan, T. Bohlen and H. Maurer (2018b). Application of structurally similarity constraints on time-domain elastic full-waveform inversion: Near Surface Geoscience Conference and Exhibition, Extended Abstracts.

Athanasopoulos and Bohlen (2019). Shallow seismic wavefield scattering and implications for viscoelastic FWI: Near Surface Geoscience Conference and Exhibition, Extended Abstracts.

Bao, H., J. Bielak, O. Ghattas, L. F. Kallivokas, D. R. Shewchuk and J. Xu (1998). Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers: Computer Methods Applied Mechanics and Engineering, **152**, 85-102.

Bamarouf, T., L. V. Socco and C. Comina (2017). Direct Statics estimation from ground roll data—the role of higher modes, 79th EAGE Annual Conference and Exhibition, Extended Abstracts, doi:10.3997/2214-4609.201700958

Barnes, C. and M. Charara (2008). Full-waveform inversion results when using acoustic approximation instead of elastic medium: 78th SEG International Annual Meeting, Expended Abstracts, 1895-1899.

Baumstein, A., W. Ross and S. Lee (2011). Simultaneous source elastic inversion of surface waves: 73rd Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20149055

Ben-Hadj-Ali, H., S. Operto and J. Virieux (2008). Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data: Geophysics, **73**, VE101–VE117, doi: 10.1190/1.2957948.

Bergamo P., C. Comina, S. Foti and M. Maraschini (2011). Seismic characterization of shallow bedrock sites with multi-modal Monte Carlo inversion of surface wave data: Soil Dynamics and Earthquake Engineering, **31**, 530-554.

Bergamo, P., D. Boiero and L. V. Socco (2012). Retrieving 2D structures from surface-waves data by means of space-varying windowing: Geophysics, **77**, EN39-EN51.

Beydoun, W. B. and A. Tarantola (1988). First Born and Rytov approximation: modelling and inversion conditions in a canonical example: Journal of the Acoustical Society of America, **83,** 1045-1055.

Blanch, J. O., J. O. A. Robertsson and W. W. Symes (1995). Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique: Geophysics, **60**, 176–184, doi: 10.1190/1.1443744

Bohlen, T. (2002). Parallel 3D viscoelastic finite difference seismic modeling: Computers and Geosciences, **28**, 887–899.

Bohlen, T., S. Kugler, G. Klein and F. Theilen (2004). 1.5D inversion of lateral variation of Scholte wave dispersion: Geophysics, **69**, 330–344.

Bohlen, T., T. Forbriger, L. Groos, M. Schäfer, C. Wolf and M. Binnig (2015). Applications of 2-D elastic full waveform inversion to shallow seismic Rayleigh waves: 77th Conference & Exhibition, EAGE, Extended Abstracts.

Boiero, D., M. Maraschini and L. V. Socco (2009). P and S wave velocity model retrieved by multi modal surface wave analysis*:* 71st Conference & Exhibition, EAGE, Extended Abstracts, T010.

Boiero, D. and L. V. Socco (2010). Retrieving lateral variations from surface wave dispersion curves analysis: Geophysical Prospecting, **58**, doi: 10.1111/j.1365-2470.2010.00877. x

Boiero, D. and L. V. Socco (2014). Joint inversion of Rayleigh-wave dispersion and P-wave refraction data for laterally varying layered models: Geophysics, **79**, EN49-EN59, doi: 10.1190/geo2013-0212.1

Borisov, D. and S. C. Singh (2015). Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: A synthetic study: Geophysical Journal International, **201**, 1215-1234, doi: 10.1093/gji/ggv048

Borisov, D., R. Modrak, F. Gao and J. Tromp (2018). 3D elastic full-waveform inversion of surface waves in presence of irregular topography using an envelope-based misfit function: Geophysics, **83**, R 1-R 11, doi: 10.1190/GEO2017-0081.1.

Borisov, D., F. Gao, P. Williamson and J. Tromp (2020). Application of 2D full-waveform inversion on exploration land data: Geophysics, **85**, R75-R86, doi: 10.1190/GEO2019-0082.1

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

Brenders A. J. and R. G. Pratt (2007a). Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133-151.

Brenders, A. J. and R. G. Pratt (2007b). Efficient waveform tomography for stratospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152-170.

Bretaudeau, F., R. Brossier, D. Leparoux, O. Abraham and J. Virieux (2013). 2D elastic full-waveform imaging of the near surface: Application to synthetic and a physical modeling data set: Near Surface Geophysics, **11**, 307-316, doi: 10.3997/1873-0604.2012067

Brossier, R., S. Oterto and J. Vrieux (2009a). Robust elastic frequency-domain full-waveform inversion using the L1 norm: Geophysical Research Letters, **36**, doi: 10.1029/2009GL039458

Brossier, R., S. Operto and J. Virieux (2009b). Seismic imaging of complex onshore structures by 2D elastic frequency domain full-waveform inversion: Geophysics, **74**, WCC63-WCC76.

Brossier, R., S. Operto and J. Virieux (2010). Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, R37-R46, doi: 10.1190/1.3379323

Brown, L. T., J. G. Diehl and R. L. Nigbor (2000). A simplified procedure to measure average shear-wave velocity to a depth of 30 m (Vs30): 12th World Conference of Earthquake Engineering.

Bube, K. P. and T. Nemeth (2007). Fast line searches for the robust solution of linear systems in the hybrid l_1 /l_2 and Huber norms: Geophysics, **72,** A13-A17.

Bunks, C., F. M. Salek, S. Zaleski and G. Chavent (1995). Multiscale seismic waveform inversion: Geophysics, **60,** 1457-1473, doi: 10.1190/1.1443880

Butzer, S., A. Kurzmann and T. Bohlen (2013). 3D elastic full-waveform inversion of small-scale heterogeneities in transmission geometry: Geophysical Prospecting, **61**, 1238–1251, doi: 10.1111/1365-2478.12065

Cercato, M. (2009). Addressing non-uniqueness in linearized multichannel surface wave inversion: Geophysical Prospecting, **57**, 27–47.

Chaljub, E. and B. Valette (2004). Spectral element modelling of three-dimensional wave propagation in a self-gravitating Earth with an arbitrarily stratified outer core: Geophysical Journal International, **158**, 131–141.

Chaljub, E., D. Komatitsch, Y. Capdeville, J. P. Vilotte, B. Valette and G. Festa (2007). Spectral element analysis in seismology. In: Advances in Geophysics, **48**, 365-419. Editors: Wu, R.-S., Maupin, V. Elsevier, Amsterdam.

Chen, J., C. A. Zelt and P. Jaiswal (2017). Detecting a known near-surface target through application of frequency-dependent traveltime tomography and full-waveform inversion to P- and SH-wave seismic refraction data: Geophysics, **82**, R1-R17, doi: 10.1190/GEO2016-0085.1

Choi, Y. and T. Alkhalifah (2012). Application of multi-source waveform inversion to marine streamer data using the global correlation norm: Geophysical Prospecting, **60**, 748–758, doi: 10.1111/j.1365-2478.2012.01079.x

Claerbout, J. F. and F. Muir (1973). Robust modelling with erratic data: Geophysics, **46,** 1559-1567.

Comina, C., S. Foti, D. Boiero and L. V. Socco (2011). Reliability of $V_{S,30}$ evaluation from Surface Waves Tests: Journal of Geotechnical and Geoenvironmental Engineering, **137**, 579-586, doi: 10.1061/(ASCE)GT.1943-5606.0000452

Crase, E., A. Pica, M. Noble, J. McDonald and A. Tatantola (1990). Robust elastic non-linear waveform inversion: Application to real data: Geophysics, **55**, 527-538.

Cummins, P., R. Geller, T. Hatori and N. Takeuchi (1994a). DSM synthetic seismograms: SH, spherically symmetric case: Geophysical Research Letters, **21**, 533–536.

Cummins, P., R. Geller and N. Takeuchi (1994b). DSM complete synthetic seismograms: P-SV, spherically symmetric case: Geophysical Research Letters, **21**, 1663–1666.

Dahlen, F. and J. Tromp (1998). Theoretical global seismology: Princeton University Press.

de la Puente, J., M. Käser, M. Dumbser and H. Igel (2007). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes – IV. Anisotropy: Geophysical Journal International, **169**, 1210–1228.

de la Puente, J., M. Dumbser, M. Käser and H. Igel (2008). Discontinuous Galerkin methods for wave propagation in poroelastic media: Geophysics, **73**, T77–T97.

de la Puente, J., J.-P. Ampuero and M. Käser (2009). Dynamic rupture modeling on unstructured meshes using a discontinuous Galerkin method: Journal of Geophysical Research, **114**, doi: 10.1029/2008JB00627

Deville, M., P. Fischer and E. Mund (2002). High order methods for incompressible fluid flow: Cambridge University Press.

Dokter, E., D. Köhn, D. Wilken, D. De Nil and W. Rabbel (2017). Full-waveform inversion of SH- and Love wave data in near-surface prospecting: Geophysical Prospecting, **65**, 216-236, doi: 10.1111/1365-2478.12549

Dou, S. and J. B. Ajo-Franklin (2014). Full-wavefield inversion of surface waves for mapping embedded low-velocity zones in permafrost: Geophysics, **79**, EN107–EN124, doi: 10.1190/geo2013-0427.1

Dumbser, M. and M. Käser (2006). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes – II. The three-dimensional isotropic case: Geophysical Journal International, **167**, 319–336.

Dziewonski, A., S. Bloch and M. Landisman (1969). A technique for the analysis of transient seismic signals: Bulletin of Seismological Society of America, **59**, 427–444.

Emmerich, H. and M. Korn (1987). Incorporation of attenuation into time domain computation of seismic wavefield: Geophysics, **52**, 1252–1264, doi: 10.1190/1.144238

Engquist, B., B. D. Froese and Y. Yang (2016). Optimal transport for seismic full waveform inversion: Communications in Mathematical Sciences, **14**, 2309–2330.

Epanomeritakis, I., V. Akçelik, O. Ghattas and J. Bielak (2008). A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion: Inverse Problems, **24,** 1-26, doi: 10.1088/0266-5611/24/3/034015

Ewing, W. M., W. S. Jardetzky and F. Press (1957). Elastic Waves in Layered Media. New York: McGraw-Hill.

Faccioli, E., F. Maggio, R. Paolucci and A. Quarteroni (1997). 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method: Journal of Seismology, **1**, 237–251.

Fäh, D., F. Kind and D. Giardini (2001). A theoretical investigation of average H/V ratios: Geophysical Journal International, **145**, 535–549.

Fäh, D., F. Kind and D. Giardini (2003). Inversion of local S-wave velocity structures from average H/V ratios, and their use for the estimation of site-effects: Journal of Seismology, **7**, 449–467.

Fathi, A., L. Kallivokas and B. Poursartip (2015). Full waveform inversion in three-dimensional PML-truncated elastic media: Computer Methods in Applied Mechanics and Engineering, **296**, 39–72, doi: 10.1016/j.cma.2015.07.008.

Fathi, A., B. Poursartip, K. H. Stokoe and L. Kallivokas (2016). Three-dimensional P-and S-wave velocity profiling of geotechnical sites using full-waveform inversion driven by field data: Soil Dynamics and Earthquake Engineering, **87**, 63–81, doi: 10.1016/j.soildyn.2016.04.010

Ferradans, S., N. Papadakis, G. Peyré and J.-F. Aujol (2014). Regularized Discrete Optimal Transport: SIAM Journal on Imaging Sciences, **7**, 1853–1882.

Fichtner, A., H.-P. Bunge and H. Igel (2006). The adjoint method in seismology. I. Theory: Physics of Earth and Planetary Interiors, **157**, 84-104.

Fichtner, A., B. Kennett, H. Igel and H. P. Bunge (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods: Geophysical Journal International, **179**, 1703– 1725, doi: 10.1111/j.1365-246X.2009.04368.x

Fichtner, A. (2011). Full Seismic Waveform modeling and Inversion: Springer Science & Business Media.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville and A.V. Nor (2013). Multiscale full waveform inversion: Geophysical Journal International, **194**, 534-556, doi: 10.1093/gji/ggt118**.**

Fichtner, A. and M. van Driel (2014). Models and Fréchet kernels for frequency-(in)dependent Q: Geophysical Journal International, **198**, 1878–1889, doi: 10.1093/gji/ggu22

Forbriger, T. (2003a). Inversion of shallow-seismic wavefields. Part 1: Wavefield transformation: Geophysical Journal International, **153**, 719–734, doi: 10.1046/j.1365-246X.2003.01929.x

Forbriger, T. (2003b). Inversion of shallow-seismic wavefields. Part 2: Inferring subsurface properties from wavefield transforms: Geophysical Journal International, **153**, 735–752, doi: 10.1046/j.1365-246X.2003.01985.x

Forbriger, T., L. Groos and M. Schäfer (2014). Line-source simulation for shallow-seismic data. Part 1: Theoretical background: Geophysical Journal International, **198**, 1387–1404, doi: 10.1093/gji/ggu199

Foti, S., R. Lancelotta, L. Sambuelli and L. V. Socco (2000). Notes of fk analysis of surface waves: Annali di Geofisica, **43**, 1199-1209.

Foti, S., R. Lancelotta, L. Sambuelli and L. V. Socco (2001). Application of fk Analysis of Surface Waves for Geotechnical Characterization: International Conferences on Recent Advances in Geothecnical Earthquake Engineering and Soil Dynamics, 1-6.

Foti, S. (2005). Surface Wave Testing for Geotechnical Characterization. In: Surface Waves in Geomechanics: Direct and Inverse Modelling for Soil and Rocks, CISM Series, **481**, Springer, Wien, pp. 47-71.

Foti, S., C. Comina and D. Boiero (2007). Reliability of combined active and passive surface wave methods: Rivista Italiana di Geotecnica, **2**, 39-47.

Foti, S., C. Comina, D. Boiero D. and L. V. Socco (2009). Non-uniqueness in surface-wave inversion and consequences on seismic site response analyses: Soil Dynamics and Earthquake Engineering, **29**, 982-993, doi: 10.1016/j.soildym.2008.11.004

Foti, S., C. G. Lay, G. J. Rix and C. Strobbia (2015). Surface Wave Methods for Near-Surface Site Characterization, CRC Press, Taylor & Francis Group.

Furumura, T., B. L. N. Kennet and M. Furumura (1998). Seismic wavefield calculation for laterally heterogeneous whole Earth models using the pseudospectral method: Geophysical Journal International, **135**, 845-860.

Gabriels, P., R. Snieder and G. Nolet (1987). In situ measurement of shear wave velocity in sediments with higher-mode Rayleigh waves: Geophysical Prospecting, **35**, 187–196.

Gardner, G. H. F., W. S. French and T. Matzuk (1974). Elements of migration and velocity analysis: Geophysics, **39**, 811-825.

Gauthier, O., J. Virieux and A. Tarantola (1986). Two-dimensional nonlinear inversion of seismic waveforms - Numerical results: Geophysics, **51,** 1387-1403.

Geller, R. J. and T. Ohminato (1994). Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary conditions using the direct solution method: Geophysical Journal International, **116**, 421–446.

Geller, R. J. and N. Takeuchi (1995). A new method for computing highly accurate DSM synthetic seismograms: Geophysical Journal International, **123**, 449-470.

Geller, R. J. and N. Takeuchi (1998). Optimally accurate second-order time-domain finite difference schemes for elastic equation of motion: one dimensional case: Geophysical Journal International, **135**, 48-62.

Gilbert, F. and G. E. Backus (1966). Propagator matrices in elastic wave and vibration problems: Geophysics, **31**, 326–332.

Gilbert J. C. and J. Nocedal (1992). Global convergence properties of conjugate gradient methods for optimization: SIAM Journal of Optimization, **2**, 21–42.

Graves, R. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite difference: Bulletin of the Seismological Society of America, **86**, 1091-1106.

Groos, L., M. Schäfer, T. Forbriger and T. Bohlen (2014). The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves: Geophysics, **79**, R247–R261, doi: 10.1190/geo2013-0462.1.

Groos, L., M. Schäfer, T. Forbriger and T. Bohlen (2017). Application of a complete workflow for 2D elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves: Geophysics, **82**, R109 – R117, doi: 10.1190/GEO2016-0284.1

Guitton, A. and T. Alkhalifah (2013). An introduction to this special section: Full-waveform inversion and the way forward: The Leading Edge, **32,** 1026-1028.

Ha, T., W. Chung and C. Shin (2009). Waveform inversion using a back-propagation algorithm: Geophysics, **74,** R15-R24.

Ha, W., W. Chung, E. Park and C. Shin (2012). 2D acoustic Laplace-domain waveform inversion of marine field data: Geophysical Journal International, **190,** 421-428.

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

Ha, W. and C. Shin (2012b). Efficient Laplace domain modelling and inversion using an axis transformation technique: Geophysics, **77**, R141-R148.

Haskell, N. A. (1953). The dispersion of surface waves on multilayered media: Bulletin of Seismological Society of America, **43**, 17–34.

Hauksson, E., and P. M. Shearer (2006). Attenuation models (*Qp* and *Qs*) in three dimensions of the southern California crust: Inferred fluid saturation at seismogenic depths: Journal of Geophysical Research, **111**, B05302, doi: 10.1029/2005JB00394

Haney, M. and R. Miller (2013). Introduction to this special section: Non reflection seismic and inversion of surface waves and guides waves: The Leading Edge, **32**, 610-611.

Hansen, C. (1998). Rank-deficient and ill-posed problems-Numerical aspects of linear inversion. Society of Industrial and Applied Mathematics, Philadelphia, PA, USA.

He, W., R. Brossier, L. Metivier and J. Virieux (2018). A practical workflow of full waveform inversion to process land seismic data: Synthetic study: 80th EAGE Annual Conference and Exhibition, Extended Abstracts.

He, W., R. Brossier and L. Métivier (2019). 3D elastic FWI for land seismic data: a graph space OT approach: 89th Annual Meeting, SEG, Expanded Abstracts, 1320-1324, doi: 10.1190/segam2019-3216485.1.

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems; An introductory analysis with applications to biology, control and artificial intelligence, The University of Michigan Press.

Hustedt, B., S. Operto and J. Virieux (2004). Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modeling: Geophysical Journal International, **157**, 1269-1296.

Igel, H. and M. Weber (1995). SH-wave propagation in the whole mantel using high-order finite difference: Geophysical Research Letter, **22,** 731-734.

Igel, H. (1999). Wave propagation in three-dimensional spherical sections by Chebyshev spectral method: Geophysical Journal International, **136**, 559-566.

Ikeda, T, T. Tsuji and T. Matsuoka (2013). Window-controlled CMP crosscorrelation analysis for surface waves in laterally heterogeneous media: Geophysics, **78**, EN95–EN105.

Irnaka, T. M., R. Brossier, L. Métivier, T. Bohlen and Y. Pan (2019). Towards 3D 9C Elastic Full Waveform Inversion of Shallow Seismic Wavefields - Case Study Ettlingen Line: 81st EAGE Annual Conference and Exhibition, Extended Abstracts, doi: 10.3997/2214-4609.201900994

Isaac, J. H. and G. F. Margrave (2012). Towards full waveform inversion: A torturous path: CREWES Research Report, **24**, 1-8.

Ivanov, J., J. T. Schwenk, S. L. Peterie and J. Xia (2013) The joint analysis of refractions with surface waves (JARS) method for finding solutions to the inverse refraction problem: The Leading Edge, **32**, 692–697, doi: 10.1190/tle32060692.1

Jones, R. B. (1958). In-situ measurement of the dynamic properties of soil by vibration methods: Geotechnique, **8**, 1–21.

Jones, R. B. (1962). Surface wave technique for measuring the elastic properties and thickness of roads: Theoretical development: British Journal of Applied Physics, **13**, 21–29.

Karniadakis, G. M. and S. Sherwin (2005). Spectral/hp Element Methods for Computational Fluid Dynamics, 2nd edn. Oxford University Press, Oxford.

Karray, M. and G. Lefebvre (2008). Significance and Evaluation of Poisson's ratio: Canadian Geotechnical Journal, **45**, pp. 624-635.

Kausel, E. and J. M. Roesset (1981). Stiffness matrices for layered soils: Bulletin of Seismological Society of America, **71**, 1743–1761.

Kawai, K., and R. J. Geller (2010). Waveform inversion for localized seismic structure and an application to D" structure beneath the Pacific: Journal of Geophysical Research, **115**, doi: 10.1029/2009JB006503

Kearey, P., M. Brooks and J. Hill (2002). An Introduction to Geophysical Exploration, 3rd Edition. Blackwell Publishing, Oxford, Chapters: 2-5.

Kelly, K. R., R. W. Ward, S. Treitel, R. M. Alford (1976). Synthetic seismograms: a finite-difference approach: Geophysics, **41**, 2-27.

Kennett, B. L. N, E. R. Engdahl and R. Buland (1995). Constraints on seismic velocities in the Earth from traveltimes: Geophysical Journal International, **105**, 429-465.

Khosro Anjom, F., D. Teodor, C. Comina, R. Brossier, J. Virieux and L. V. Socco (2019). Full waveform matching of Vp and Vs models from surface waves: Geophysical Journal International, **218**, 1873-1891, doi: 10.1093/gji/ggz279

Köhn, D. (2011). Time domain 2D elastic full waveform tomography: Ph.D. thesis, Christian-Albrechts-Universität zu Kiel.

Köhn, D., T. Meier, M. Fehr, D. De Nil and M. Auras (2016). Application of 2D elastic Rayleigh waveform inversion to ultrasonic laboratory and field data: Near Surface Geophysics, **14**, 461–476.

Köhn, D., D. Wilken, D. De Nil, T. Wunderlich, W. Rabbel, L. Werther (2018). 2D full waveform inversion applied to a strongly- dispersive Love wave field dataset: 80th EAGE Annual Conference and Exhibition, Extended Abstracts.

Kolb, P. and G. Canadas (1986). Least-squares inversion of prestack data: Simultaneous identification of density and velocity. Presented at the 16th conference of mathematical geophysics, Oosterbeek, The Netherlands.

Komatitsch, D. and J. P. Vilotte (1998). The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of Seismological Society of America, **88**, 368–392.

Komatitsch, D., and J. Tromp, (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation: Geophysical Journal International, **139**, 806–822, doi: 10.1046/j.1365-246x.1999.00967.x

Komatisch, D. and J. Tromp (2002a). Spectral-element simulations of global seismic wave propagation – I. Validation: Geophysical Journal International, **149**, 390–412.

Komatitsch, D. and J. Tromp (2002b). Spectral-element simulations of global seismic wave propagation-II. 3-D models, oceans, rotation and self-gravitation: Geophysical Journal International, **150**, 303–318.

Komatitsch, D., Q. Liu, J. Tromp, P. Süss, C. Stidham and J. H. Shaw (2004): Simulation of ground motion in the Los Angeles basin based upon the spectral-element method: Bulletin of Seismological Society of America, **94**, 187–206.

Komatitsch, D., S. Tsuboi and J. Tromp (2005). The spectral-element method in seismology. In: Seismic Earth: Array Analysis of Broadband Seismograms, Eds. Levander, A. Nolet, G., AGU Monogr.157, 205–228.

Komatitsch, D. and R. Martin (2007). An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation: Geophysics, **72**, SM155–SM167, doi: 10.1190/1.2757586

Komatitsch, D., Z. Xie, E. Bozdag, E. S. de Andrade, D. Peter, Q. Liu and J. Tromp (2016). Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion: Geophysical Journal International, **206**, 1467–1478, doi: 10.1093/gji/ggw224

Krampe, V., Y. Pan and T. Bohlen (2018). Full-waveform inversion of near-surface seismic data in anisotropic media: Near Surface Geoscience Conference and Exhibition, Extended Abstracts.

Krampe, V., Y. Pan and T. Bohlen (2019). Two-dimensional elastic full-waveform inversion of Love wavesin shallow vertically transversely isotropic media: synthetic reconstruction tests: Near Surface Geophysics, **17**, 449-461, doi:10.1002/nsg.12061

Lacoss, R. T., E. J. Kelly and M. N. Toksoz. (1969). Estimation of seismic noise structure using arrays: Geophysics, **34**, 21–38.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. Conference of Inverse Scattering: Theory and Application, Society of Industrial and Applied Mathematics (SIAM), Philadelphia.

Lamb, H. (1904). On the propagation of tremors over the surface of an elastic solid: Philosophical Transactions of the Royal Society of London A, **203**, 1–42.

Lellmann, J., D. Lorenz, C. Schönlieb and T. Valkonen (2014). Imaging with kantorovich–rubinstein discrepancy: SIAM Journal on Imaging Sciences, **7**, 2833–2859.

Levander, A. (1988). Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436, doi: 10.1190/1.1442422

Levshin, A., M. Ritzwoller and L. Ratnikova (1994). The nature and cause of polarization anomalies of surface waves crossing northern and central Eurasia: Geophysical Journal International, **117**, 577–591.

Liao, Q. and G. A. McMechan (1996). Multifrequency viscoacoustic modeling and inversion: Geophysics, **61**, 1371–1378.

Liu, F., A. Morton, Z. Ma and S. Checkles (2013). Some key factors for the successful application of full-waveform inversion: The Leading Edge, **32,** 1124-1129.

Liu, Z., J. Li, S. M. Hanafy and G. Schuster (2019). 3D wave-equation dispersion inversion of Rayleigh waves: Geophysics, **84**, R673-R691, doi: 10.1190/geo2018-0543.1

Louie, J. N. (2001). Faster, better: Shear-wave velocity to 100 meters depth from refraction microtremor arrays: Bulletin of Seismological Society of America, **91**, 347–364.

Lowrie, W. (2007). Fundamentals of Geophysics, 2nd Edition: Cambridge University Press, Chapter 3,121-206.

Luenberger, D. (1984). Linear and nonlinear programming, 2nd Edition. Stanford University, Kluwer Academic Publishers.

Luo, J. and R.-S. Wu (2015). Seismic envelope inversion: reduction of local minima and noise resistance: Geophysical Prospecting, **63**, 597–614.

Lysmer, J. and L. A. Drake (1972). A finite-element method for seismology: Methods of Computational Physics, **11**, Academic Press, New York.

Malinowski, M., S. Operto and A. Ribodetti (2011). High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion: Geophysical Journal International, **186**, 1179–1204, doi: 10.1111/j.1365-246X.2011.05098.x

Maraschini, M. (2008). A new approach for the inversion of Rayleigh and Scholte waves in sit characterization, *PhD Thesis*, Politecnico di Torino.

Maraschini, M and S. Foti (2010). A Monte Carlo multimodal inversion of surface waves: Geophysical Journal International, **182**, 1557-1566, doi: 10.1111/j.1365-246X.2010.04703.x

Maraschini, M., D. Boiero, S. Foti and L. V. Socco (2011). Scale properties of the seismic wavefield: Perspectives for full waveform matching: Geophysics, **76**, A37–A44, doi: 10.1190/geo2010-0213.1.

Marfurt, K. (1984). Accuracy of finite difference and finite-element modeling of the scalar and elastic wave equations: Geophysics, **49**, 533-549, doi: 10.1190/1.1441689

Masoni, I., R. Brossier, J. Virieux and J. Boelle (2013). Alternative misfit functions for FWI applied to surface waves: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20130297

Masoni, I., R. Brossier, J.-L. Boelle, M. Macquet and J. Virieux (2014). Robust full-waveform inversion of surface waves: Seismic Technology, **11**, doi: 10.3997/2405-7495.2015114

Masoni, I., J.-L. Boelle, R. Brossier and J. Virieux (2016). Layer stripping FWI for surface waves: 86th Annual International Meeting, SEG, Expanded Abstracts, 1369–1373.

Martin, R. and D. Komatitsch (2009). An unsplit convolutional perfectly matched layer technique improved at grazing incidence for the viscoelastic wave equation: Geophysical Journal International, **179**, 333–344, doi:10.1111/j.1365-246X.2009.04278.

Menke, W. (1989). Geophysical data analysis: Discrete Inverse Theory. Rev ed., Academic Press, San Diego, California.

Métivier, L. and R. Brossier (2016). The seiscope optimization toolbox: A large-scale non-linear optimization library based on reverse communication: Geophysics, **81**, F11-F25.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet and J. Virieux (2016a). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: Geophysical Journal International, **205**, 345–377.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet and J. Virieux (2016b). An optimal transport approach for seismic tomography: Application to 3D full waveform inversion: Inverse Problems, **32**, 115008.

Moczo, P., E. Bystricky, J. M. Carcione and M. Bouchon (1997). Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures: Bulletin of the Seismological Society of America, **87**, 1305–1323.

Moczo, P. and J. Kristek (2005). On the rheological models used for time domain methods of seismic wave propagation: Geophysical Research Letters, **32**, doi: 10.1029/2004GL021598

Moczo, P., J. Kristek, M. Galis, P. Pazak and M. Balazovjech (2007). The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion: Acta Physica Slovaca, **52**, 177-406.

Modrak, R. and J. Tromp (2016). Seismic waveform inversion best practices: Regional, global and exploration test cases: Geophysical Journal International, **206**, 1864–1889, doi: 10.1093/gji/ggw202

Modrak, R., D. Borisov, M. Lefebvre and J. Tromp (2018). The SeisFlows framework for waveform inversion: Rapid prototyping and HPC applications: Computers and Geosciences, **115**, 88–95, doi: https://doi.org/10.1016/j.cageo.2018.02.004

Morgan, J., M. Warner, R. Bell, J. Ashley, D. Barnes, R. Little, R. Roelle, R. and C. Jones (2013). Next generation seismic experiments: wide-angle, multi-azimuth, three-dimensional, full-waveform inversion: Geophysical Journal International, **195,** 1657-1678.

Mora, P. (1987a). Elastic inversion of multioffset seismic amplitude data for P-wave velocity, S-wave velocity and density: in Danbom, S. & Domenico, M., Shear wave exploration, SEG.

Mora, P. (1987b). Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52,** 1211-1228.

Mora, P. (1988). Elastic wave-field inversion of reflection and transmission data: Geophysics, **53**, 750-759.

Mora, P. (1989). Inversion = migration+tomography: Geophysics, **54**, 1575-1596.

Moss, R. E. S. (2008). Quantifying measurement uncertainty of thirty-meter shear-wave velocity: Bulletin of the Seismological Society of America, **98**, 1399–1411.

Mulder, W. and R. E. Plessix (2008). Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, 827-841.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface: Q Rep Railway Tech Res Inst, **30**, 25–33.

Nakamura, Y. (1996). Real time information systems for seismic hazards mitigation UrEDAS, HERAS and PIC. Q Rep Railway Tech Res Inst, **37**, 112–127.

Nakamura, Y. (2000). Clear identification of fundamental idea of Nakamura's technique and its applications. In: Proceedings of the 12th World Conference on Earthquake Engineering, Auckland, New Zealand.

Nazarian, S. and K. M. Stokoe (1983). Use of Spectral Analysis of Surface Waves for Determination of Moduli and Thickness of Pavement System: Transportation Research Record, **954**, TRB, Transportation Research Board, Washington, D. C.

Nazarian, S. (1984). In situ determination of elastic moduli of soil deposits and pavement systems by spectra-analysis-of-surface-waves Method. PhD Dissertation, University of Texas, Austin.

Newmark, N. M. (1959). A method of computational structural dynamics. ASCE Journal of Engineering Mechanics Division, **85,** 67-94.

Nguyen, T. D. and K. T. Tran (2018). Site characterization with 3D elastic full-waveform tomography: Geophysics, **83**, R389 – R400, doi: 10.1190/GEO2017-0571.1

Nissen-Meyer, T., A. Fournier and F. A. Dahlen (2007). A two-dimensional spectral-element method for computing spherical-Earth seismograms – I. Moment-tensor source: Geophysical Journal International, **168**, 1067–1092.

Nissen-Meyer, T., A. Fournier and F. A. Dahlen (2008). A two-dimensional spectral-element method for computing spherical-Earth seismograms – II. Waves in solid-fluid media: Geophysical Journal International, **174**, 873–888.

Nocedal, J., and S. J. Wright (2006). Numerical optimization, 2nd edition: Springer.

Nuber, A., E. Manukyan and H. Maurer (2016). Enhancement of near surface elastic full waveform inversion results in regions of low sensitivities: Journal of Applied Geophysics, **122**, 192–201, doi: 10.1016/j.jappgeo.2015.09.020

Nuber, A. E. Manukyan and H. Maurer (2017). Optimizing measurement geometry for seismic near-surface full-waveform inversion: Geophysical Journal International, **210**, 1909-1921, doi: 0.1093/gji/ggx267

Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero and P. dell'Aversana (2004). Quantitative imaging of multifold wide aperture seismic data; a case study: Geophysical Prospecting, **52**, 625-651.

Operto, S., J. Virieux, J. X. Dessa and G. Pascal (2006). Crustal Imaging form multifold ocean bottom seismometers data by frequency-domain full-waveform tomography. Application to the eastern Nankai trough: Journal of Geophysical Research, **111**, doi: 10.1029/2005JB003835

Operto, S., Y. Gholami, V. Prieux, V. Ribodetti, R. Brossier, L. Metivier and J. Virieux (2013). A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, 1040-1054, doi: 10.1190/tle32091040.1

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti and J. Virieux (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: Application to Valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, **202**, 1362-1391, doi: 10.1093/gji/ggv226

Operto, S. and A. Miniussi (2018). On the role of density and attenuation in three-dimensional multiparameter viscoacoustic VTI frequency-domain FWI: An OBC case study from the North Sea: Geophysical Journal International, **213**, 2037–2059, doi: 10.1093/gji/ggy103

Oristaglio, M. (2012). SEAM update: SEAM Phase II-surface waves in land seismic exploration: The Leading Edge, **31**, 1130–1132, doi: 10.1190/tle31101130.1

Pan, Y., J. Xia, Y. Xu, L. Gao and Z. Xu (2016). Love-wave waveform inversion in time domain for shallow shear-wave velocity: Geophysics, **81**, R1–R14, doi: 10.1190/geo2014-0225.1

Pan, W. and K. A. Innanen (2019). Amplitude-based misfit functions in viscoelastic full-waveform inversion applied to walk-away vertical seismic profile data: Geophysics, **84**, B335–B351, doi: 10.1190/geo2018-0765.1

Pan, Y., L. Gao and T. Bohlen (2019a). 2D viscoelastic full waveform inversion on a field shallow-seismic Data: 81st EAGE Annual Conference and Exhibition, Extended Abstracts.

Pan, Y., L. Gao and T. Bohlen (2019b). High-Resolution Characterization of Near-Surface Structures by Surface Wave Inversions: From Dispersion Curve to Full Waveform: Survey in Geophysics, **40**, 167-195, doi: 10.1007/s10712-019-09508-0

Papadopoulou, M., F. Da Col, B. Mi, E. Bäckström, P. Marsden, B. Brodic, A. Malehmir and L. V. Socco (2020). Surface-wave analysis for static corrections in mineral exploration: A case study from central Sweden: Geophysical Prospecting, **68**, 214-231, doi: 10.1111/1365-2478.1289

Park C.B, R.D. Miller and J. Xia (1999). Multichannel analysis of surface waves: Geophysics, **64**, 800-808.

Patera, A. T. (1984). A spectral element method for fluid dynamics: laminar flow in a channel expansion: Journal of Computational Physics, **54**, 468-488.

Pelekis, P. C. and G. A. Athanasopoulo (2011). An overview of surface waves methods and a reliability study of a simplified inversion technique*:* Soil Dynamics and Earthquake Engineering, **31**, 1654-1668.

Peter, D., D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini and J. Tromp (2011). Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes: Geophysical Journal International, **186**, 721-739, doi: 10.1111/j.1365-246X.2011.05044.x

Peterie, S., R. Miller, S. Morton, Y. Wang, S. Sloan, M. Moran and H. Cudney (2016). Tunnel detection using SH-wave diffraction imaging: 86th Annual International Meeting, SEG, Expanded Abstracts, 5006–5010, doi: 10.1190/segam2016-13962354.1

Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495-503, doi: 10.1111/j.1365-246X.2006.02978.x

Plessix, R. E. (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver: Geophysics, **74**, WCC149–WCC157, doi: 10.1190/1.3211198

Plessix, R. E., and C. Perkins (2010). Full-waveform inversion of a deep water ocean bottom seismometer dataset: First Break, **28**, 71-78, doi: 10.3997/1365-2397.2010013

Plessix, R. E., P. Milcik, H. Rynja, A. Stopin and K. Matson (2013). Multiparameter full-waveform inversion: Marine and land examples: The Leading Edge, **32**, 1030-1039.

Pratt, R. G. and N. R. Goulty (1991). Combining wave-equation imaging with traveltime tomography to form high-resolution images from cross-hole data: Geophysics, **56,** 204-224.

Pratt, R. G., C. Shin and G. J. Hicks (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133,** 341-362. doi: 10.1046/j.1365-246X.1998.00498.x

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888-901.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, B.P. Flannery (2007). Numerical Recipes, 3rd edition. Cambridge University Press, Cambridge.

Quarteroni, A., R. Sacco, F. Saleri (2000). Numerical mathematics. Springer, New York, NY.

Raknes, E. B., W. Weibull and B. Arntsen (2013). Elastic velocity analysis and time-lapse full-waveform inversion: The Leading Edge, **32**, 1000-1006.

Ratcliffe, A., R. Jupp, R. Wombell, G. Body, V. Drussel, A. Fernandes, B. Glosing and M. Lombardi (2013). Full-Waveform inversion of variable-depth streamer data: An application to shallow channel modelling in the North Sea: The Leading Edge, **32,** 1110-1115.

Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero and P. dell'Aversana (2004). Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield tomography: Application to a thrust belt: Geophysical Journal International, **159**, 1032-1056, doi: 10.1111/j.1365-246X.2004.02442.x

Robertsson, J. O. A., J. O. Blanch, and W. W. Symes (1994). Viscoelastic finite difference modeling: Geophysics, **59**, 1444–1456, doi: 10.1190/1.1443701

Robertsson, J. O. A., A. Levander, W. W. Symes and K. Holliger (1995). A comparative study of free-surface boundary conditions for finite-difference simulation of elastic/viscoelastic wave propagation: 65th Annual International Meeting, SEG, Expanded Abstracts, 1277–1280.

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.

Romdhane, A., G. Grandjean, R. Brossier, F. Réjiba, S. Operto and J. Virieux (2011). Shallow structures characterization by 2D elastic waveform inversion: Geophysics, **76**: R81-R93, doi: 10.1190/1.3569798

Rothman, D. H. (1985). Nonlinear inversion, statistical mechanics and residual statics estimations: Geophysics, **50,** 2784-2796.

Sambridge, M. S., A. Tarantola and B. L. N. Kennett (1991). An alternative strategy for non linear inversion of seismic waveforms: Geophysical Prospecting, **39**, 457-491.

Sambridge, M and K. Mosegaard (2002). Monte Carlo Methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1-29.

Savage, J. C. (1959). A new method for analyzing the dispersion of oceanic Rayleigh waves: Journal of Geophysical Research, **74**, 2608-2617.

Scales, J. A. and M. L. Smith (1994). Introductory geophysical inverse theory. Samizdat Press.

Schäfer, M., L. Groos, T. Forbriger and T. Bohlen (2012). On the effects of geometrical spreading corrections for a 2D full waveform inversion of recorded shallow seismic surface waves: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20148327

Schäfer, M., L. Groos, T. Forbriger and T. Bohlen (2014). Line-source simulation for shallow-seismic data. Part 2: full-waveform inversion – a synthetic 2-D case study: Geophysical Journal International, **198**, 1405-1418, doi: 10.1093/gji/ggu171

Schwenk, J. T., R. D. Miller, J. Ivanov, S. D. Sloan and J. R. McKenna (2012a). Joint shear-wave analysis using MASW and refraction traveltime tomography: Symposium on the Application of Geophysics to Engineering and Environmental Problems, SEG, 197–206.

Schwenk, J. T., S. D. Sloan, J. Ivanov and R. D. Miller (2016), Surface-wave methods for anomaly detection: Geophysics, **81**, EN29–EN42, doi: 10.1190/geo2015-0356.1

Sears, T. J., S. C. Singh and P. J. Barton (2008). Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56,** 843-862, doi: 10.1111/j.1365-2478.2008.00692.x

Seriani, G. (1998). 3-D large-scale wave propagation modeling by a spectral element method on a Cray T3E multiprocessor: Computer Methods in Applied Mechanics and Engineering, **164**, 235–247

Sheen, D. H., K. Tuncay, C. E. Baag and P. J. Ortoleva (2006). Time domain Gauss-Newton seismic waveform inversion in elastic media: Geophysical Journal International, **167**, 1373–1384, doi: 10.1111/j.1365-246X.2006.03162.x

Shi, T., J. Zhang, Z. Huang and C. Jin (2015). A layer-stripping method for 3D near-surface velocity model building using seismic first-arrival times: Journal of Earth Sciences, **26**, 502507.

Shipp, R. M. and S. C. Singh (2002). Two dimensional full waveform inversion of wide-aperture marine seismic streamer data: Geophysical Journal International, **151**, 325-344, doi: 10.1046/j.1365-246X.2002.01645.

Shin, C., S. Jang and D. J. Min (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49,** 592-606.

Shin, C. and Y. Cha (2008). Waveform inversion in the Laplace domain: Geophysical Journal International, 173, 922-931.

Shin, C. and W. Ha (2008). A comparison between the behaviour of objective functions for waveform inversion in the frequency and Laplace domains: Geophysics, **73,** VE119-VE133.

Shin, C., W. Ha and Y. Kim (2013). Subsurface model estimation using Laplace-domain inversion methods: The Leading Edge, **32,** 1094-1099.

Sirgue, L. and R. G. Pratt (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69,** 231-248.

Sirgue, L. (2006). The importance of low frequency and large offsets in waveform inversion. 68th EAGE Conference & Exhibition Extended Abstracts, A037.

Sirgue, L., J. Etgen and U. Ulbertin (2008). 3D frequency domain waveform inversion using time domain finite difference methods. 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F022.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Abertin and J. H. Kommedal (2010). Full-waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, 65-70, doi: 10.3997/1365-2397.2010012.

Smith, J. A., D. Borisov, H. Cudney, R. D. Miller, R. Modrak, M. Moran, S. L. Peterie, S. D. Sloan, J. Tromp and Y. Wang (2019). Tunnel detection at Yuma Proving Ground, Arizona, USA — Part 2: 3D full-waveform inversion experiments: Geophysics, **84**, B101 – B120, doi: 10.1190/GEO2018-0599.1

Stokoe, K. H., S. G. Wright, J. Bay and J. M. Roesset (1994). Characterization of geotechnical sites by SASW method. In: Geophysical Characterization of Sites (R. D. Woods, ed.). ISSMFE Technical Committee #10. Oxford: IBH, pp. 15–25.

Stokes, G. G. (1880). Mathematical and Physical Papers (J. Larmor, ed.). Reprinted from the original journals and transactions, with additional notes by the author. 5 vol. Cambridge: Cambridge University Press.

Strobbia, C., and G. Cassiani. (2011). Refraction microtremors: Data analysis and diagnostics of key hypotheses: Geophysics, **76**, MA11–MA20.

Socco, L. V. and C. Strobbia (2004). Surface Waves Methods for near-surface characterization: A tutorial: Near Surface Geophysics, **2**, 165-185.

Socco, L. V. and D. Boiero (2008). Improved Monte Carlo inversion of surface wave data: Geophysical Prospecting, **56**, 357–371, doi: 10.1111/j.1365-2478.2007.00678.x

Socco, L. V., D. Boiero, S. Foti and R. Wisén (2009). Laterally constrained inversion of ground roll from seismic reflection records: Geophysics, **74**, G35-G45.

Socco, L.V., D. Jongmans, D. Boiero, S. Stocco, M. Maraschini, K. Tokeshi and D. Hantz (2010a). Geophysical investigation of the Sandalp rock avalanche deposits: Journal of Applied Geophysics, **70**, 277–291.

Socco, L. V., S. Foti and D. Boiero (2010b). Surface-wave analysis for building near-surface velocity models-Established approaches and new perspectives: Geophysics, **75**, 75A83-75A102, doi: 10.11901/1.3479491.

Socco, L. V., C. Comina and F. Khosro Anjom (2017a). Time-average velocity estimation through surface-wave analysis: Part 1 — S-wave velocity: Geophysics, **82**, U49-U59, doi: 10.1190/geo2016-0367.1

Socco, L. V., F. Khosro Anjom, C. Comina and D. Teodor (2017b). P- and S-wave velocity models from surface wave dispersion curves data transform: 87th Annual Meeting, SEG, Expanded Abstracts, 5197-5201, doi: 10.1190/segam2017-17734559.1

Socco, L. V. and C. Comina (2017). Time-average velocity estimation through surface-wave analysis: Part 2 – P-wave velocity: Geophysics, **82**, U61-U73, doi: 10.1190/geo2016-0368.1

Solano, C. P., D. Donno and H. Chauris (2014). Alternative waveform inversion for surface wave analysis in 2-D media: Geophysical Journal International, **198**, 1359–1372, doi: 10.1093/gji/ggu211

Sun, M., J. Zhang and W. Zhang (2017). Alternating first-arrival traveltime tomography and waveform inversion for near-surface imaging: Geophysics, **82**, R245-R257, doi: 10.1190/GEO2016-0576.1

Takeuchi, N. and R. J Geller (2000). Optimally accurate second-order time-domain finite difference scheme for computing synthetic seismograms in 2-D and 3-D media: Physics of the Earth and Planetary Interior, **119**, 99– 131.

Tape, C., Q. Liu, A. Maggi and J. Tromp (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods: Geophysical Journal International, **180**, 433-462, doi: 10.1111/j.1365-246X.2009.04429.x

Tarantola, A. (1984a). Linearized inversion of seismic reflection data: Geophysics Prospecting, **32,** 998-1015.

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

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51,** 1983-1903.

Tarantola, A. (1987). Inverse Problem Theory: Methods for data fitting and model parameter estimation. Elsevier Science Publ. Co., Inc, New York.

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveform, including anelasticity and attenuation: Pure Applied Geophysics, **128,** 365-399, doi: 10.1007/BF01772605.

Tarantola, A. (2005). Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.

Telford, W. M., L. P. Geldart and R. E. Sheriff (1991). Applied Geophysics, Second Edition. Cambridge University Press, Chapter 4, 136-282.

Teodor, D., C. Comina, L. V. Socco, R. Brossier, P. -T. Trinh and J. Virieux (2017). Initial model design for Full-Waveform Inversion - Preliminary elastic modeling from surface waves data analysis, 36th GNGTS Annual Meeting, Extended Abstracts, 733-756.

Teodor, D., C. Comina, F. Khosro Anjom, L.V. Socco, J. Virieux, P.-T. Trinh and R. Brossier (2018a). Building initial models for full-waveform inversion of shallow targets by surface waves dispersion curves clustering and data transform: 88th Annual Meeting, SEG, Expanded Abstracts, 4738-4742, doi: 10.1190/segam2018-2997848.1.

Teodor, D., C. Comina, L. V. Socco, R. Brossier, F. Khosro Anjom, P.-T. Trinh and J. Virieux (2018b). Multi-parameter full-waveform Inversion of shallow targets: preliminary results and critical aspects. 37th GNGTS Annual Meeting, Extended Abstracts, 220-223.

Teodor, D., C. Comina, L-V. Socco, F. Khosro Anjom, J. Virieux, R. Brossier and P.-T. Trinh (2019). Elastic full waveform inversion tests for shallow targets reconstruction from surface waves analysis based initial models: 81st EAGE Annual Conference and Exhibition, Extended Abstracts, doi: 10.3997/2214-4609.201900976

Teodor, D., C. Comina, F. Khosro Anjom, R. Brossier, L. V. Socco and J. Virieux (2020). Challenges in shallow targets reconstruction by 3D elastic full-waveform inversion – Which initial model?: submitted.

Teodor, D., C. Comina, L. V. Socco, R. Brossier and J. Virieux (2020). Imaging near-surface heterogeneities with high accuracy by dispersion curves analysis and 3D elastic full-waveform inversion: which acquisition layout? submitted.

Tessmer, E. and D. D. Kosloff (1994). 3D elastic modeling with surface topography by Chebyshev spectral method: Geophysics, **59**, 464-473.

Thomson, W. T. (1950). Transmission of elastic waves through a stratified solid medium: Journal of Applied Physics, **21**, 89–93

Tran, K. T. and D. R. Hiltunen (2012a). One-dimensional inversion of full waveform using genetic algorithm: Journal of Environmental and Engineering Geophysics, **17**, 197–213.

Tran, K. T. and D. R. Hiltunen, (2012b). Two-dimensional inversion of full waveform using simulated annealing: Journal of Geotechnical and Geoenvironmental Engineering, **138**, 1075–1090, doi: 10.1061/(ASCE)GT.1943-5606.0000685

Tran, K. T. and M. McVay (2012). Site characterization using Gauss-Newton inversion of 2-D full seismic waveform in time domain: Soil Dynamics and Earthquake Engineering, **43**, 16–24, doi: 10.1016/j.soildyn.2012.07.004

Tran, K. T., M. McVay, M. Faraone and D. Horhota (2013). Sinkhole detection using 2-D full seismic waveform tomography: Geophysics, **78**, R175–R183, doi: 10.1190/geo2013-0063.1

Tran, K. T., M. Mirzanejad, M. McVay and D. Horhota (2019). 3-D time-domain Gauss-Newton full waveform inversion for near-surface site characterization: Geophysical Journal International, **217**, 206-218, doi: 10.1093/gji/ggz020

Trinh, P.-T., R. Brossier, L. Métivier, J. Virieux and P. Wellington (2017). Bessel smoothing filter for spectral element mesh: Geophysical Journal International, **209**, 1489-1512, doi: 10.1093/gji/ggx103

Trinh, P. T., R. Brossier, L. Métivier, L. Tavard and J. Virieux (2018). Data-windowing hierarchy in multi-parameter elastic FWI: 3D synthetic foothills case study: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: https://doi.org/10.3997/2214-4609.201801570

Trinh, P.-T., R. Brossier, L. Metivier, L. Tavard and J. Virieux (2019). Efficient time-domain 3D elastic and visco-elastic FWI using a spectral-element method on flexible Cartesian-based mesh: Geophysics, **84**, R75-R97, doi: org/10.1190/geo2018-0059.1

van Driel, M. and T. Nissen-Meyer (2014). Optimized viscoelastic wave propagation for weakly dissipative media: Geophysical Journal International, **199**, 1078–1093, doi: 10.1093/gji/ggu314

Vigh, D., N. Moldoveanu, K. Jiao, W. Huang and J. Kapoor (2013). Ultralong-offsets data acquisition can complement full-waveform inversion and lead to improved subsalt imaging: The Leading Edge, **32,** 1116-1122, doi: 10.1190/tle32091116.1

Vigh, D., K. Jiao, D. Watts and D. Sun (2014). Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, R63-R77, doi: 10.1190/geo2013-0055.1

Virieux, J. (1984). SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **49**, 1933-1957.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51,** 889-901, doi: 10.1190/1 .1442147

Virieux, J. and S. Operto (2009). An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC1-WCC26, doi: 10.1190/1.3238367

Vogel, C. R. and M. E. Oman (1996). Iterative methods for total variation denoising: SIAM Journal on Scientific Computing, **17,** 277-238.

Vogel, C. (2002). Computational methods for inverse problems. Society of Industrial and Applied Mathematics.

Wang, Y. and Y. Rao (2009). Reflection seismic waveform tomography: Journal of Geophysical Research Solid Earth, **114** (B3) doi: 10.1029/2008JB003916

Wang, Y., R. D. Miller, S. L. Peterie, S. D. Sloan, M. L. Moran, H. H Cudney, J. A. Smith, D. Borisov, R. Modrak and J. Tromp (2019). Tunnel detection at Yuma Proving Ground, Arizona, USA — Part 1: 2D full-waveform inversion experiment: Geophysics, **84**, B95-B105, doi: 10.1190/GEO2018-0598.1

Warner, M., J. Morgan, A. Umpleby, I. Stekl and L. Guasch (2012). Which Physics for Full-wavefield Seismic Inversion? In 74th EAGE Conference & Exhibition, Extended Abstracts.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shan, V. Vinje, I. Stekl, L. Guash, C. Win, G. Conroy and A. Bertrand (2013). Anisotropic 3D full-waveform inversion: Geophysics, **78**, R59-R80, doi: 10.1190/geo2012-0338.1

Whitley, D. L. (1994). A genetic algorithm tutorial: Statistics and Computing, **4**, 65-85, doi: 10.1007/BF00175354

Williamson, P. R. (1991). A guide to the limits of resolution imposed by scattering in ray tomography: Geophysics, **56**, 202-207.

Woodward, M. J. (1992). Wave-equation tomography: Geophysics, **57,** 15-26.

Xia, J, R. Miller and C. Park (1999). Estimation of near-surface shear-wave velocity by inversion of Rayleigh wave: Geophysics, **64**, 691–700.

Xia, J., R. D. Miller, C. B. Park and G. Tian (2003). Inversion of high frequency surface waves with fundamental and higher modes: Journal of Applied Geophysics, **52**, 45–57.

Xing, Z. and A. Mazzotti (2019a). Two-grid full-waveform Rayleigh-wave inversion via a genetic algorithm: part 1 – method and synthetic examples: Geophysics, **84**, P. R805–R814, doi: 10.1190/geo2018-0799.1

Xing, Z. and A. Mazzotti (2019b). Two-grid full-waveform Rayleigh-wave inversion via a genetic algorithm: part 2 – application to two actual data sets: Geophysics, **84**, R815–R825, doi: 10.1190/geo2018-0800.1

Yamanaka, H. (2007). Inversion of surface-wave phase velocity using hybrid heuristic search method: Butsuri-Tansa/Geophysical Exploration, 60, 265–275 (in Japanese), doi: 10.3124/segj.60.265

Yang, P., R. Brossier, L. Métivier and J. Virieux (2016a). A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal International, **207**, 129-149, doi: 10.1093/gji/ggw262

Yang, P., R. Brossier, L. Métivier and J. Virieux (2016b). Wavefield reconstruction in attenuating media: A checkpoint-assisted reverse-forward simulation method: Geophysics, **81**: R349–R362, doi: 10.1190/geo2016-0082.1

Yang, P., R. Brossier and J. Virieux, (2016c). Wavefield reconstruction from significantly decimated boundaries: Geophysics, **80**, T197-T209, doi: 10.1190/geo2015-0711.1

Yang, Y., B. Engquist, J. Sun and B. F. Hamfeldt (2018). Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, **83**, R43–R62.

Yuan, Y. O., F. J. Simons and E. Bozdağ (2015), Multiscale adjoint waveform tomography for surface and body waves: Geophysics, **80**, R281–R302, doi: 10.1190/geo2014-0461.1

Yoon, S., and G. Rix. (2009). Near-field effects on array-based surface wave methods with active sources: Journal of Geotechnical Geoenvironmental Engineering, **135**, 399–406.

Zhou, C., W. Cai, Y. Luo, G. T. Schuster and S. Hassanzadeh (1995). Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data: Geophysics, **60**, 765–773.

Zhou, C. G. T. Schuster, S. Hassanzadeh and J. M. Harris (1997). Elastic wave equation traveltime and waveform inversion of crosswell data: Geophysics, **62**, 853-868.

Zienkiewicz, O. C., R. L. Taylor and J. Z. Zhu (2005). The Finite Element Method: Its Basis and Fundamentals, 6th Edition, Elsevier Butterworth-Heinemann.

Zywicki, D. J., and G. J. Rix. (2005). Mitigation of near-field effects for seismic surface wave velocity estimation with cylindrical beamformers: Journal of Geotechnical Geoenvironmental Engineering, **131**, 970–977.