High resolution imaging of lithospheric structure by full waveform inversion of short period teleseismic P waves

Yi Wang

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)
Cotutelle internationale avec :
University of Science and Technology of China

Présentée et soutenue par :

Yi WANG

le *lundi 6 mars 2017*

Titre :

Imagerie haute résolution des structures lithosphériques par inversion de formes d'ondes P télésismiques courte période

---

École doctorale et discipline ou spécialité :
SDU2E : Sciences de la Terre et des Planètes Solides

Unité de Recherche :
Géosciences Environnement Toulouse (UMR 5563)

Directeur/trice(s) de Thèse :
Sébastien CHEVROT
Sidao NI

Jury :

Anne PAUL	DR,CNRS-ISTerre, Grenoble	Rapporteur
Li ZHAO	Senior scientist,IES, Taipei	Rapporteur
Frédéric MOUTHEREAU	Professeur UPS, Toulouse	Examinateur
György HETÉNYI	Professeur UNIL, Lausanne	Examinateur
Sébastien CHEVROT	DR,CNRS-GET, Toulouse	Directeur de Thèse
Sidao NI	Professeur USTC, Hefei	Codirecteur de Thèse
Dimitri KOMATITSCH	DR,CNRS-LMA, Marseille	Invité
Jean-Luc BOELLE	Ingénieur expert,TOTAL, Pau	Invité

Résumé

La tomographie sismique permet d'imager l'intérieur de la Terre à partir de l'observation des ondes sismiques faite à la surface. L'inversion de forme d'ondes complètes est une méthode tomographique qui permet d'imager les structures lithosphériques de petite échelle. Cette approche demande des méthodes numériques efficaces et précises pour résoudre l'équation des ondes dans des milieux hétérogènes complexes. En théorie, la limite de résolution que l'on peut atteindre avec cette technique est de l'ordre de grandeur de la plus petite longueur d'onde présente dans le champ d'onde utilisé. Du fait de son coût élevé en temps de calcul, l'inversion de formes d'ondes complètes constituait encore récemment un formidable défi pour le sismologue. Cependant, cette situation est en train d'évoluer rapidement du fait des progrès récents à la fois des moyens de calcul haute performance ainsi que des méthodes numériques, mais aussi des déploiements de réseaux denses à l'échelle régionale.

Dans cette thèse, nous nous sommes intéressés à l'imagerie haute résolution des structures lithosphériques sous la chaine des Pyrénées par inversion de formes d'ondes P télésismiques courte période. L'objectif principal est d'apporter des contraintes nouvelles sur le taux de raccourcissement subi par cette chaine de montagnes pendant la convergence alpine. Nous utilisons une méthode de modélisation hybride qui couple une méthode de propagation d'onde globale 1D à une méthode d'éléments spectraux 3D à l'échelle régionale. Cette méthode hybride permet de coupler les champs globaux et régionaux sur les bords du domaine régional 3D. Elle limite les calculs 3D qui sont extrêmement couteux à l'intérieur du domaine régional, ce qui permet de réduire considérablement le temps de calcul. La méthode hybride permet ainsi de modéliser des sismogrammes synthétiques jusqu'à des périodes de l'ordre de la seconde, en prenant en compte toutes les complexités qui peuvent affecter la propagation des ondes dans le domaine régional 3D. A l'aide de cette méthode, il est également possible de calculer par la méthode de l'adjoint les dérivées de Fréchet qui relient les perturbations des formes d'onde observées aux perturbations des paramètres élastiques et de la densité dans le milieu. Ces noyaux de sensibilité sont utilisés pour formuler un problème inverse résolu par un algorithme itératif de type L-BFGS. Nous inversons les données de 5 sources télésismiques enregistrées par deux transects denses déployés au niveau des Pyrénées occidentales et centrales pendant l'expérience PYROPE. Nous avons ainsi obtenu les premières sections haute résolution de vitesses des ondes P et S au travers d'une chaine de montagnes. Les modèles tomographiques apportent des évidences claires en faveur d'un sous charriage de la plaque ibérique sous la plaque européenne. Ils montrent également l'importance de l'héritage et en particulier des structures liées à l'épisode de rifting crétacé dans la structuration de la chaine.

Les mots-clés: L'inversion de forme d'ondes, La tomographie sismique, Imagerie par réseau régional, Ondes P télésismiques, Orogène

Abstract

Seismic tomography allows us to image the Earth's interior based on surface observations of seismic waves. The full waveform inversion (FWI) method has the potential to improve tomographic images for the fine scale structures of the lithosphere. For this reason it receives a lot of attention of seismologists. FWI requires an efficient and precise numerical techniques to solve the elastic wave equation in 3D heterogeneous media. Its resolution potential is limited by the shortest wavelength in the seismic wavefield and the wavefield sampling density. Because of the high computational cost of modeling the propagation of seismic waves in heterogeneous media, FWI remains challenging. However, owing to the progress in high performance computational resources and numerical simulation techniques, as well as the deployment of permanent and temporary broadband arrays in the last two decades, this situation has changed dramatically.

In this thesis, we focus on the high resolution imaging of lithospheric structure beneath the Pyrenean range by FWI, to quantify the highly controversial amount of convergence that occurred during the formation of this mountain range. In order to obtain finely resolved tomographic images, we exploit short period teleseismic P waves recorded by dense transects. We use a hybrid method that couples a global wave propagation method in a 1D Earth model to a 3D spectral-element method in a regional domain. A boundary coupling approach is used to match the global and regional wavefields on the boundaries of the regional domain. This hybrid method restricts the costly 3D computations inside the regional domain, which dramatically decreases the computational cost. The hybrid method can model teleseismic wavefields down to 1s period, accounting for all the complexities that may affect the propagation of seismic waves in the 3D regional domain. By using this hybrid method, the sensitivity kernels of the least square waveform misfit function with respect to elastic and density perturbations in the regional domain are computed with the adjoint state method. These waveform sensitivity kernels are used in an iterative L-BFGS algorithm to invert broad-band waveform data recorded by two dense transects deployed during the temporary PYROPE experiment across the Pyrenees mountains. We obtain the first high resolution lithospheric sections of compressional and shear velocities across the Pyrenean orogenic belt. The tomographic models provide clear evidence for the underthrust of the thinned Iberian crust beneath the European plate and for the important role of rift-inherited mantle structures during the formation of the Pyrenees.

Keywords: Waveform inversion, Seismic tomography, Regional array imaging, Teleseismic P waves, Orogen

Introduction générale

Exploiter les données sismologiques pour caractériser l'intérieur de notre planète reste un enjeu majeur de la sismologie moderne. Bien que des progrès importants ont été accomplis, de nombreuses questions restent ouvertes, en particulier à l'échelle régionale. Dans cette thèse, nous nous sommes intéressés à l'imagerie des structures lithosphériques sous la chaine des Pyrénées.

Les Pyrénées, qui constituent la partie la plus jeune de la chaine Alpine, résultent de la convergence des plaque Ibérie et Europe pendant le Cénozoïque. C'est un orogène orienté E-O qui est bordé par deux bassins flexuraux: le bassin de l'Ebre au sud et la bassin aquitain au nord. Les structures lithosphériques sous les Pyrénées, qui sont la clé pour comprendre le support dynamique de la topographie et pour la reconstruction cinématique de cette marge convergente, restent très mal connues. Deux grands profils de sismique active ont été acquis pendant le programme ECORS au cours des années 80 afin d'imager l'architecture profonde de cette chaine de montagnes: les profils ECORS-Pyrenees et ECORS-Arzacq. Différentes interprétations des sections obtenues ont été proposées dans la littérature. Cependant, du fait de la difficulté d'illuminer les réflecteurs profonds et de l'ambigüité de leur migration en profondeur, de larges pans de l'architecture profonde des Pyrénées restent controversés. Par exemple les structures dans la croûte supérieure ou la géométrie du Moho restent très incertains.

Deux expériences temporaires, PYROPE et IBERARRAY, ont ainsi été déployées pour tenter d'apporter des contraintes nouvelles sur l'architecture profonde des Pyrénées. Ces expériences ont ouvert de nouvelles perspectives pour imager les structures lithosphériques avec des approches d'imagerie passive. Pour compléter ce dispositif, deux transects denses ont également été déployés au travers des Pyrénées centrales et occidentales d'octobre 2011 à octobre 2013. Ces deux profils suivaient approximativement le tracé des deux lignes ECORS.

Cependant, l'imagerie fine des structures lithosphériques sous cette chaine de montagnes restait un défi, qui a demandé de faire évoluer les méthodes d'imagerie classiques reposant encore largement sur des théories asymptotiques vers des méthodes exploitant les formes d'ondes complètes. Ces méthodes d'inversion de formes d'ondes ont une limite potentielle de résolution spatiale de l'ordre de la moitié de la plus petite longueur d'onde présente dans le champ d'onde. Cependant, ces méthodes nécessitent d'utiliser des méthodes numériques précises qui sont très couteuses en temps de calcul. Grâce aux progrès à la fois théoriques et méthodologiques récents, ces calculs sont dorénavant à la portée des sismologues. Dans cette thèse, nous avons utilisé une méthode hybride qui couple une méthode de propagation d'onde dans une Terre globale de symétrie sphérique, à une méthode d'éléments spectraux 3D dans un domaine régional. Cette méthode hybride permet de calculer des sismogrammes synthétiques jusqu'à des périodes de l'ordre de la seconde, en prenant en compte toutes les complexités qui pourraient affecter la propagation des ondes sismiques dans le domaine régional: topographie de la surface libre et des discontinuités internes, variations latérales de vitesses isotropes et anisotropes, atténuation,

etc. . .

Cette méthode hybride a été utilisée pour calculer les noyaux de sensibilié des formes d'ondes aux perturbations des paramètres élastiques dans le domaine régional à l'aide de la méthode de l'adjoint. Ces noyaux de sensibilité ou dérivées de Fréchet ont permis d'inverser de façon itérative les composantes radiales et verticales des formes d'ondes P télésismiques à courte période enregistrées par les deux transects pyrénéens. L'inversion repose sur un algorithme itératif L-BFGS de type Gauss-Newton. Les images tomographiques obtenues nous ont permis d'apporter de nouvelles contraines sur le taux de raccourcissement dans les Pyrénées centrales et occidentales.

Dans le premier chapitre, nous passons en revue les principales questions encore ouvertes portant sur la formation et l'évolution des Pyrénées. L'inversion de forme d'ondes complètes, la méthode de l'adjoint, et les méthodes hybrides sont également revues.

Dans le chapitre 2 nous présentons les trois ingrédients nécessaires pour résoudre le problème direct: la méthode hybride, la construction du maillage 3D des 2 domaines pyrénéens, ainsi que l'estimation des ondelettes source.

Dans le chapitre 3 nous présentons la méthode de calcul des dérivées de Fréchet ainsi que l'algorithme itératif d'inversion de formes d'ondes complètes.

Dans le chapitre 4, nous présentons la méthode de traitement et de sélection des données sismologiques.

Dans le chapitre 5, nous montrons les résultats de la première application de notre méthode d'inversion de formes d'ondes aux données récoltées le long du transect déployé à l'ouest des Pyrénées. Le modèle tomographique obtenu est décrit et interprété.

Dans le chapitre 6, nous appliquons la même méthode aux données du transect central. Du fait d'un niveau de bruit plus élévé, nous avons du envisager différentes stratégies d'inversion, qui ont été testées sur des jeux de données synthétiques calculés dans des modèles en damiers. Nous présentons un modèle et quelques interprétations pour ce transect, qui restent encore préliminaires.

Dans le chapitre 7, nous présentons les conclusions et les perspectives de ce travail.

General introduction

Using seismological data to characterize the internal structure of our planet is one of the key tasks of modern seismology. Even though many progresses in our understanding of the Earth's interior have been accomplished, there are still many open questions, especially at the regional scale. In this thesis, we will focus on the imaging of lithospheric structures beneath the Pyrenean orogenic belt.

The Pyrenees mountains, which are the youngest part of the Alpine-Himalayan collisional orogenic belt, was produced by the convergence between the Iberian and European plates during the Cenozoic. It is an east-west, bivergent orogen between two foreland basins, the Ebro basin in the south and the Aquitaine basin in the north. The lithospheric structures beneath the Pyrenees, which are key to understanding the support of topography and for kinematic reconstructions at this convergent plate boundary, are still poorly known. In order to image the architecture of the Pyrenees, two active source seismic experiments were carried out during the ECORS program: the ECORS-Pyrenees and ECORS-Arzacq profiles. Different geological interpretations of these deep reflection surveys have been proposed. However, because of the difficulty in illuminating the deep structures with active source experiments and the inherent ambiguity of deep reflector migration, many parts of the Pyrenees architecture remained controversial. For instance, the structures of the upper crust and the geometry of the Moho interface remain uncertain.

To get further insight into the deep architecture of the Pyrenees, the temporary PYROPE and IBERARRAY experiments have been deployed in the Pyrenees, opening new opportunities to image lithospheric structures with passive imaging approaches. In addition, two dense transects were deployed across the Pyrenees between 2011 and 2013, approximately following the previous ECORS profiles, in order to better constrain the geometry of this collisional orogen.

However, obtaining finely resolved tomographic images of lithospheric structures beneath this orogen belt remained a challenge, which involved moving from simplified tomographic methods relying on asymptotic descriptions of the wavefield to waveform inversion methods exploiting the complete wavefield. Full waveform inversion has a theoretical resolution limit of the order of half of the shortest wavelength in the seismic wavefield, but it requires accurate numerical techniques to simulate the propagation of seismic waves in 3D heterogeneous media, which is still computationally demanding. Owing to recent theoretical and numerical developments, this problem can now be tackled by seismologists. In this thesis, we have used a hybrid method that couples a global wave propagation method in a spherically symmetric Earth model to a 3D spectral-element method in a regional domain. This allows us to compute synthetic seismograms down to 1 s period, accounting for all the complexities that may affect the propagation of seismic waves in the regional domain: the topography of free surface and of internal discontinuities, lateral variations of isotropic and anisotropic elastic properties,

attenuation, and so on.

We use this efficient hybrid method to compute waveform sensitivity kernels with respect to elastic and density perturbations in the regional domain based on the adjoint state method. These sensitivity kernels are then used to invert the vertical and radial short period teleseismic P waveforms recorded by dense regional arrays. The inversion relies on an iterative quasi-Newton L-BFGS algorithm. Compared to general gradient-based algorithms, the L-BFGS algorithm which has a faster convergence rate significantly reduce the computational cost for the non-linear waveform inversion. The obtained tomographic images allow us to quantify the highly controversial amount of convergence in the western and central Pyrenees.

In Chapter 1 we review the main open questions regarding the formation and evolution of the Pyrenees. The full waveform inversion, the adjoint state method, and the hybrid approaches are also reviewed.

In Chapter 2 we give an overview of the three main ingredients involved in the forward problem: the DSM/SEM hybrid method, the construction of the 3D regional mesh for the Pyrenean domains, and the source wavelet estimation.

In Chapter 3 we present the computation of the sensitivity kernels using the adjoint state method and the full waveform inversion algorithm.

In Chapter 4 we discuss our data processing method and perform the checkerboard resolution tests for the western Pyrenees.

In Chapter 5 we apply our full waveform inversion algorithm to the real data from the western Pyrenees. The new tomographic model for the western Pyrenees is described and interpreted.

In Chapter 6 we invert the real data from the central Pyrenees. Since the data of the central Pyrenees have lower S/N ratio compared to the western Pyrenees, we discuss the inversion strategies based on checkerboard resolution tests. A preliminary model and interpretation for the central Pyrenees transect are also presented.

In Chapter 7 we give the conclusions and perspectives of this thesis.

Acknowledgments

As time goes by, I have spent more than a quarter o my adult life in France. My PhD study is coming to an end. Many people have contributed in a direct or an indirect way to this thesis.

First, I owe a special thanks to my Chinese supervisor, Sidao Ni, who brought me into the wonderful field of seismology, offered me this precious opportunity to study in Toulouse, and gives me constant support over these years.

My deepest gratitude goes to my French supervisor, Sébastien Chevrot, for his constant patience and guidance. He teaches me how to explore the seismic waveforms as a seismologist and provide full support for my work. He has walked me through all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could never have been finished.

It is my honour to be supervised by these two excellent scholars. The most important thing they taught me is not the knowledge of seismology, but methodology and scientific attitude that are the true wealth of the lifetime.

I have also collaborated with many experts in the lab, in particular Dr. Dimitri Komatitsch, Vadim Monteiller and Roland Martin. Many results in this thesis are based on their remarkable works. I would like to give my sincerity appreciation and great respect to them.

The friends I met in Toulouse are warm, funny and helpful. I would like to thank Yu Chen, Xiaojun Feng, Rui Xue, Yu Zhao, Yingning He, Jing Ye, Hongmei Xu, Chuxian Li, Youcun Liu, Xing Huang, Le Yao, Buyun Du, Wei Yuan and Qingyuan Qu for enduring my weird personality. I appreciate collaborations and friendships with my classmates in USTC, Xiaohui He, Xiaoli Wang, Wenbo Wu, Jun Xie and others, who have been always kind to me. I also would like to thank IRAP/GET/USTC staff members, who make my work environment well and my life easier.

Finally, my deep thanks would go to my parents for their unconditional support and love. To my parents-in-law, for their patience and waiting. To my dear Penny for her unselfish love and endless support, especially during the stressful and tough times. This thesis is also dedicated to my lovely naughty son, Ruiqian. Dad really wish you can sleep tight.

Contents

Chapter 1

Introduction

Contents

1.1 The open problems in the Pyrenees

The development of Earth science has experienced a long history. In 1912, German geologist Wegener proposed the hypothesis of continental drift, opening the prelude to the modern revolution of plate tectonics. In the beginning, Wegener's theory was widely questioned on both geophysical and geological grounds. After the World War II, this situation changed dramatically. The notion of seafloor spreading was first proposed by American marine geologist Hess (1962) and marine geophysicist Dietz (1961), who put forward that new seafloor is created at mid-ocean ridges, spreads away, and is subducted in subduction zones.

Building on Wegener's hypothesis and Hess and Dietz's contributions, Canadian geophysicist Wilson (1965) developed the theory of plate tectonics and transform faults. Plate tectonics show that the rigid outer layers of the Earth (the crust and part of the upper mantle), also called the lithosphere, decompose into several pieces or "plates". These plates float on the top of an underlying weaker rock layer called the asthenosphere. Rocks are at high temperatures and pressures so that they behave similar to the viscous liquid in the asthenosphere. Wilson conjectured there are three types of plate boundaries: mid-ocean ridges (where ocean crust is created), trenches (where the ocean plates are subducted) and large fractures in the seafloor

called transform faults (where two plates move horizontally from each other). This periodic cycle theory of ocean basins' opening and closing, is named Wilson cycle. The oceanographic evidence for seafloor spreading (Vine & Wilson, 1965) and development of plate tectonics have validated Wegener's basic hypothesis that modern continents are drifted fragments of Pangaea, a supercontinent that incorporated almost all the landmasses on Earth and existed during the late Paleozoic and early Mesozoic.

The most important point of view in plate tectonics theory is that the Earth's surface consists of oceanic and continental lithospheres which are divided into plates of varying sizes. The relative motions of these plates occur in three types of plate boundaries(convergent, divergent, or transform). Earthquakes, volcanoes, mountain belts and rifts can develop along these boundaries. Plate tectonics provide a unifying theory to explain the fundamental processes that shape the Earth's surface.

However, since the late 20th century, geologists started to recognize that the concepts of plate tectonics often could not apply to describe the deformation of continents which can occur on the much broader region than simply along the plate boundaries. Among the problems of continental dynamics, the formation of continental orogenic belts is of particular importance. Orogenic belts are generally regarded as resulting from the convergent motion of tectonic plates. In general, continental orogenic belts are mainly caused by the thickening of lithosphere after shortening and bending.

Continental orogens are key for understanding the formation and evolution of continental lithosphere. However, the formation of many complex orogenic belts is still poorly understood. Cawood et al. (2009) classified continental orogens into three modes: (1) accretionary orogens with oceanic-type subduction; (2) collisional orogens with continental-type subduction; (3) intra-continental orogens, away from an active plate margin. According to the definitions of Cawood, the oceanic circum-Pacific subduction system is an accretionary orogen, which involves subduction of oceanic lithosphere. A representative example of a collisional orogen is the Alpine-Himalayan system, which formed by subduction and collision between the African, Indian and Eurasian plates. It is characterized by a deep subduction. The intra-continental orogens, broadly distributed in continent interiors, are usually considered as the distant effects of the forces acting on the plate boundaries.

1.1.1 Overview of the study region: the Pyrenees

The Pyrenees are the youngest part of the Alpine-Himalayan collisional orogenic belt, the formation of which is closely related to the Tethyan evolution. The Pyrenees are a 453 km long (from west to east), but relatively narrow (150 km wide from north to south) continental collisional orogen located between France in the north and Spain in the south which extend from the northern Iberian margin in the west to the Mediterranean Sea in the east. The average

elevation of the Pyrenees is about 1800 m.

The general structure of the Pyrenees is an asymmetrical double-wedge with a larger extension to the South than to the North (Souriau et al., 2008). The Pyrenees can be divided into five structural units that are bounded by some major faults (Figure 1.1), which are from south to north (Sibuet et al., 2004; Vergés & Fernàndez, 2012):

1. The Aquitaine foreland basin, filled with early Cretaceous and early Neogene clastic rocks.

2. The north-directed North Pyrenean thrust system, in which Cretaceous flysch deposits are locally highly strained and metamorphosed.

3. The Axial Zone (AZ) made of Hercynian metamorphic and Mesozoic sedimentary rocks. It includes the highest summit, with an elevation higher than 3000 m.

4. The larger South Pyrenean thrust system (Mesozoic and paleozoic series) which is translated southward and overthrust.

5. the Ebro foreland basin in the south, linked to the southern Pyrenean wedge.

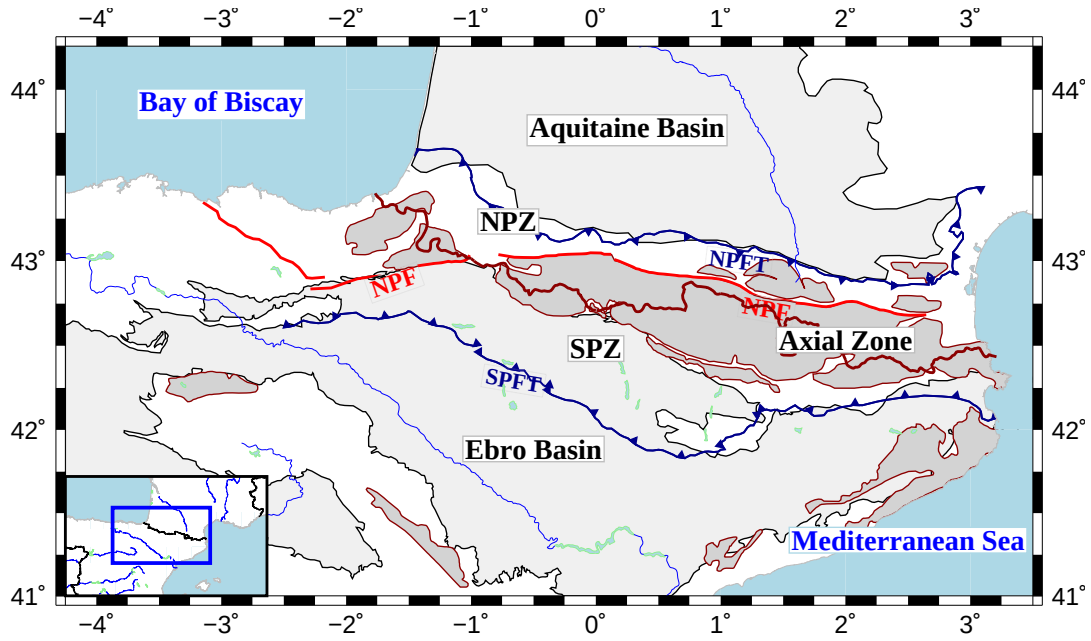


Figure 1.1: Main geological and structural units of the Pyrenees region. NPF: North Pyrenean Fault, NPFT: North Pyrenean Front Thrust, SPFT: South Pyrenean Front Thrust, NPZ: North Pyrenean Zone, and SPZ: South Pyrenean Zone.

During the Pyrenean orogeny, the flexure of the Iberian and the European plates led to the development of the foreland basins on both sides of the Pyrenees (Brunet, 1986). The North Pyrenean Fault (NPF) is regarded as the suture between the two plates (Souriau et al., 2008). Recent GPS studies indicate that there is no deformation detected across and within the Pyrenees. The relative motion across the Pyrenees has an upper bound of only 0.2 mm/year (Nocquet, 2012). This implies that the convergence between the Iberian and the European

Plates is nowadays inactive.

1.1.2 The open problems regarding the formation and evolution of the Pyrenees

Many important aspects of the creation of the Pyrenees still remain controversial. In the following, we will expose briefly the main open questions regarding the Pyrenees.

1.1.2.1 The problem of plate kinematic reconstruction

The opening of the Bay of Biscay and the formation of the Pyrenees are continuing events tightly linked to the relative displacement of Iberia with respect to Eurasia (Choukroune, 1992). The opening of the Bay of Biscay corresponds to the preorogenic stages of the Pyrenean domain. Three contradictory models for the kinematics of Iberia have been proposed for this opening process (Figure 1.2):

(1) A scissors-type opening of the bay with a pole of rotation located in the southeastern Bay of Biscay corner (Sibuet et al., 2004; Vissers & Meijer, 2012).

(2) An opening following a left-lateral strike-slip motion along the NPF, with a pole of rotation located in northern France (Pichon & Sibuet, 1971).

(3) An opening along a transtensional type Iberia-Eurasia plate boundary until the earliest Albian, followed by the beginning of orthogonal extension between Iberia and Eurasia until the onset of compression at C_{34} (85 Ma) (Jammes et al., 2009; Choukroune & Mattauer, 1978).

The total rotation of the Iberian plate with respect to the European plate in the first two models is about 35°, in good agreement with the paleomagnetic data (van der Voo, 1969). The second kinematics model (Sibuet & Pichon, 1971; Pichon & Sibuet, 1971) infers that the NPF was a former transform plate boundary. The transtensional pull-apart basins developed along the NPF during the preorogenic period. But according to recent study of plate kinematic reconstructions, it is found that the scissors-type opening of the Bay of Biscay shows a substantially better consistency with the magnetic anomaly M_0 (120 Ma) (Sibuet et al., 2004). This reconstruction implies that there was an opening of a broad oceanic domain (up to 300 km) between Iberia and Eurasia before the Aptian (113-126 Ma). During the Aptian, It gradually closed as a result of the convergence between Iberia and Eurasia. In the meantime, the Bay of Biscay was opening. This inference means that the closure of this oceanic domain should have left a suture located beneath the northernmost part of the Ebro basin, but none has ever been recognized (Sibuet et al., 2004).

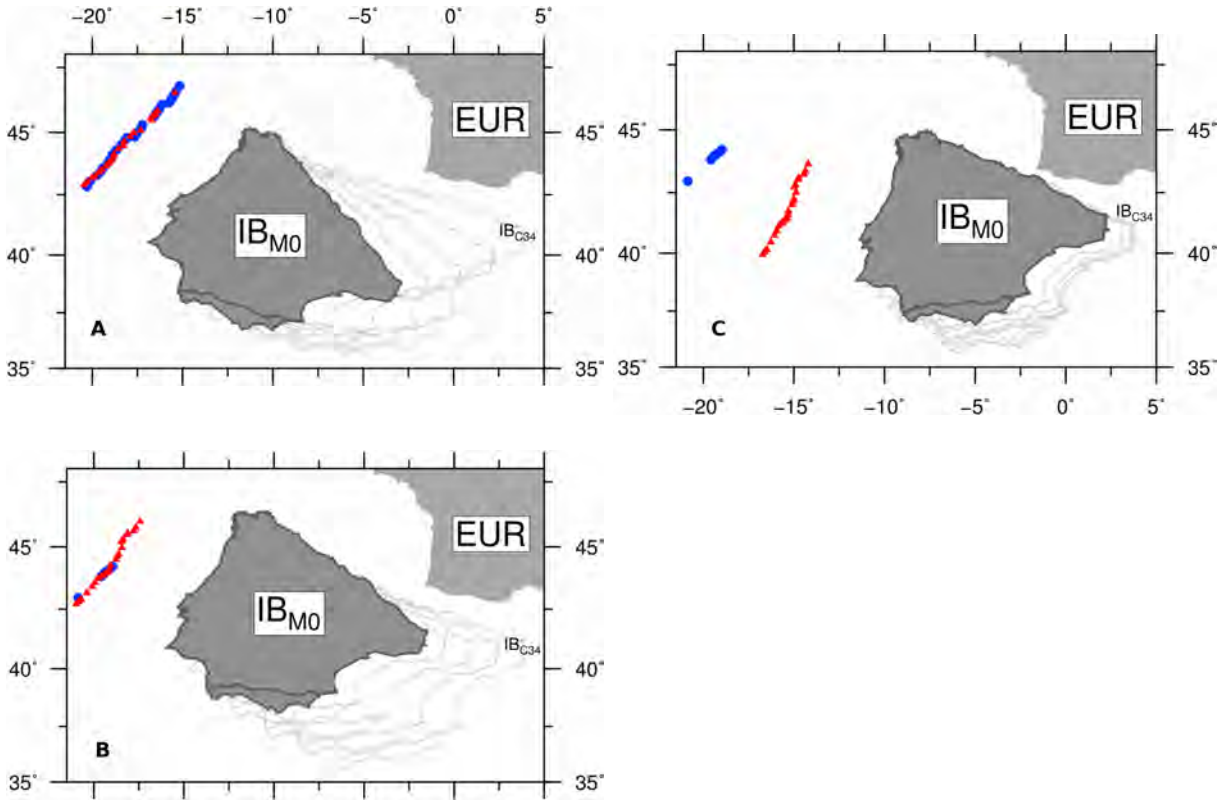


Figure 1.2: Three different kinematic reconstruction models of Iberia with respect to a fixed Eurasia at chron M0 (from Barnett-Moore et al., 2016). The light grey contours represent relative plate motions of Iberia in 10 Myr increments until C_{34}. The red triangles represent M_0 along the west Iberian margin; the blue circles represent M_0 along the Newfoundland margin. The model A presents a scissor-type opening of the Bay of Biscay. The model B argues a dominant left-lateral strike-slip motion along the NPF. The model C proposes a transtensional motion along the Iberia-Eurasia plate boundary by the earliest Albian. Since this model does not consider the seafloor magnetic anomalies as important kinematic constraints, it shows a large mismatch between M_0 along the west Iberian margin and the Newfoundland margin.

The third plate model is supported by three constraints: (1) the analysis of preserved inverted rift structures along the Pyrenees (Barnett-Moore et al., 2016); (2) a major plate reorganization of Iberia occurred during the Aptian; (3) the transtensional motion should accommodate between 300 and 400 km of continental breakup along the Iberia-Newfoundland margin in late Aptian to early Albian (Tucholke et al., 2007). However, if the magnetic anomaly M_0 is considered as an isochron, this model shows a large mismatch between the west Iberian margin and the Newfoundland margin, because it does not take into account the seafloor magnetic anomalies which are important kinematic constraints (Barnett-Moore et al., 2016). To reconcile the kinematic model with geological data, Vissers and Meijer (2012) proposed a new model with the subduction of a Neotethys oceanic domain beneath the Pyrenees. During the Aptian, this subducted slab became gravitationally unstable, detached and sank into the mantle. As a result, asthenospheric mantle rose to form the high temperature metamorphism exposed in the NPZ nowadays. However, because of the absence of intermediate stages between the seafloor magnetic anomaly M_0 and C_{34}, the proposed plate reconstruction models are debated (Bronner

et al., 2011, 2012; Tucholke & Sibuet, 2012). Therefore, the state of the relative displacement between Iberia and Eurasia is still highly questionable.

1.1.2.2 The nature of the Pyrenean domain and the amount of convergence

The nature of the Pyrenean domain during the Mesozoic has direct relation to the Iberian plate kinematic problem previously mentioned, which is still controversial. Early studies proposed that during the Cretaceous several asymmetrical pull-apart basins opened along the plate boundary between Iberia and Eurasia. This opening responded to the eastward sinistral strike-slip motion of Iberia along the NPF (Daignières, 1978; Choukroune & Mattauer, 1978). In this model, the NPF played an important role and was regarded as the former plate boundary between Iberia and Eurasia. This model has been challenged, because of the geological structures is continuous across the Pyrenees. This structural continuity excludes any evident left-lateral movement of Iberia with respect to Eurasia after the Early Cretaceous (Souquet & Mediavilla, 1976).

On the other hand, the amount of north-south convergence that occurred during the formation of the Pyrenees is a key indicator to quantify the relative movement between the Iberian and the European plates and to understand the evolution of the Pyrenean orogenic belt. This convergence gave rise to compressional deformation during the Late Cretaceous and Pyrenean orogeny. However, the precise amount of convergence accommodated in the Pyrenees is still highly debated. Shortenings between 50 km and 400 km, varying from east to west, have been proposed (Teixell, 1998; Roure et al., 1989; Muñoz, 1992). Accordingly, different mechanisms have been proposed for this shortening: homogeneous crustal thickening (Teixell, 1998), crustal stacking (Roure et al., 1989), or continental subduction (Muñoz, 1992).

1.1.3 Previous geophysical studies

1.1.3.1 The seismic exploration results

In order to better constrain the Pyrenean deep structure, some active source seismic experiments are performed in the Pyrenees. In 1978, an extensive seismic refraction survey was carried out along two east-west lines, deployed in the axial zone and the NPZ (Pyrenees, 1980). The experiment involved several shots at sea and on land, recorded by two dense profiles. This study showed that the Moho is offset vertically across the NPF by approximately 15 km in the central Pyrenees (Daignières et al., 1982). This Moho step reduces east towards the Mediterranean Sea. The magnitude of this Moho step reaches about 5 km beneath the eastern Pyrenees until it is not observed beneath the Mediterranean Sea any more (Gallart et al., 1980). This Moho step is approximately coincident with the trace of the NPF at the surface. The NPF was thus considered as the plate boundary between Iberia and Eurasia. The thick crustal roots beneath the axial zone evidenced by the refraction survey also provides an interpretation for the evident

east-west negative Bouguer anomaly observed along most of the Pyrenees, towards its eastern termination close to the Mediterranean Sea (Daignières et al., 1982). On the other hand, the strong positive Bouguer anomalies observed in the Labourd and Saint Gaudens regions could be ascribed to the existence of dense materials at shallow depth (Daignieres et al., 1989), but their origin remains uncertain.

From 1985 to 1986, a 250 km long deep seismic reflection profile was performed across the central Pyrenees by the French-Spanish ECORS (Etude Continentale et Océanique par Réflexion et Réfraction Sismique) group. It was the first deep reflection survey covering the whole orogenic belt (ECORS Pyrenees Team, 1988). The profile was deployed across the central Pyrenees from the Aquitaine basin in the north to the Ebro basin in the south (Figure 1.3). Figure 1.4a shows the line drawing of the ECORS-Pyrenees profile, extracted from unmigrated reflection data. The clear reflectors in the upper crust obtained from the reflection survey confirmed the fan-like cross-section geometry in the central part of the belt (Choukroune, 1989). This survey also showed that the Iberian crust seems to be 2 s two-way travel time (TWT) thicker than the European crust, with a thickness of approximately 50 km to the south of the NPF. Both the Iberian and European lower crusts show a strong layering above the Moho, with a series of north and south-dipping reflectors explained as Hercynian thrusts (Choukroune et al., 1990). The deep reflectors beneath the NPF show that the Iberian crust lies beneath the southern edge of the European crust beneath the axial zone (Choukroune, 1989). In the external domains, the reflectors near the surface precisely define the locations of major thrusts and crustal structures which have an important influence on the Mesozoic and Cenozoic cover of the Pyrenees (Choukroune, 1992).

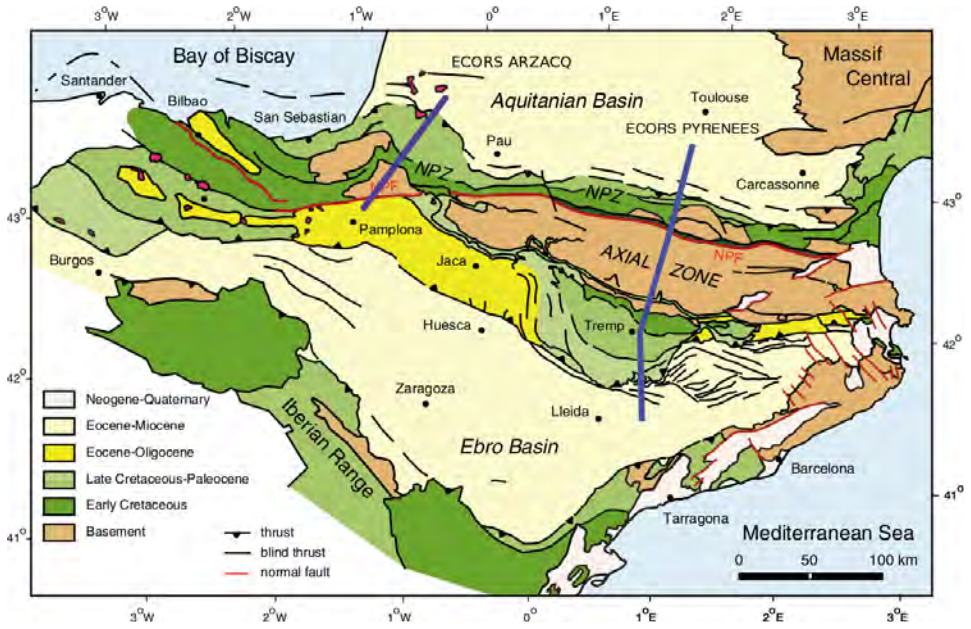
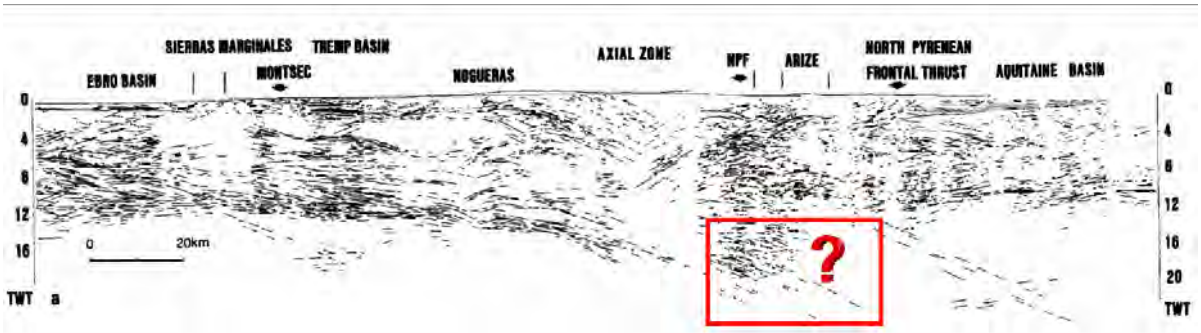
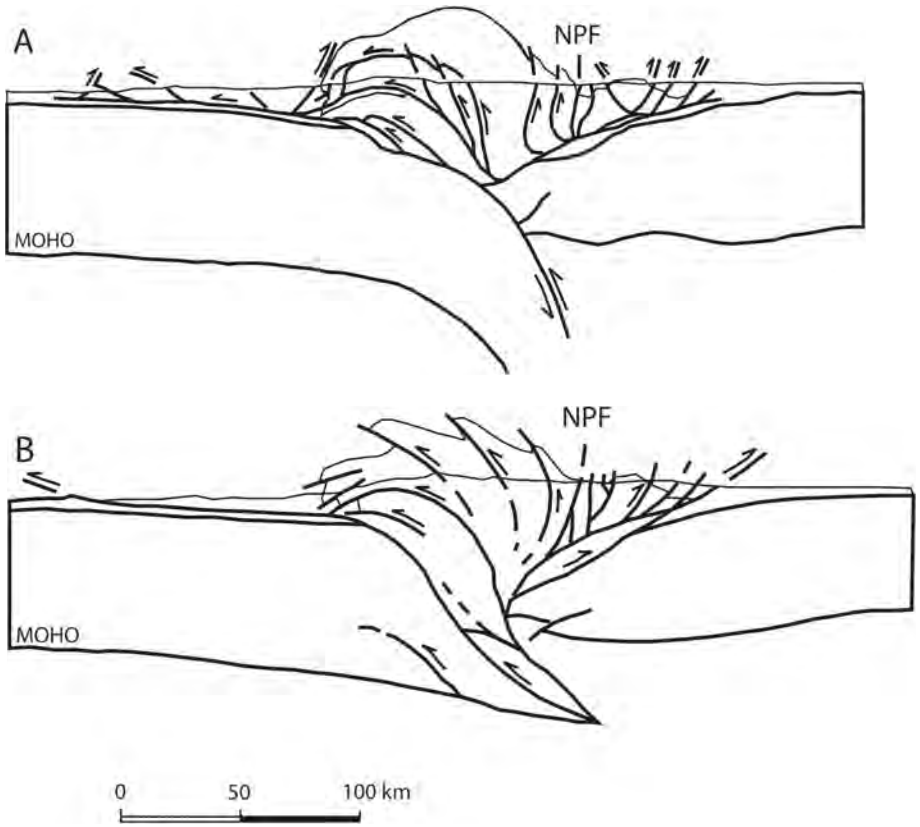


Figure 1.3: Previous active source seismic experiments deployed across the central and western Pyrenees: two ECORS deep seismic reflection profiles are shown with darkblue line (modified from Vissers & Meijer, 2012).

As a part of the ECORS project, a wide-angle experiment along the vertical reflection profile was also performed. The wide-angle reflectivity pattern shows the widespread occurrence of north-dipping impedance contrasts embedded in the middle and lower crust (Suriñach et al., 1993). Reconstructions of balanced cross-sections for the central Pyrenees along the ECORS-Pyrenees profile presented an evident discrepancy between the top of the Iberian Palaeozoic basement and the layered Iberian lower crust (Roure et al., 1989). Although the reflectors in the upper crust are well-defined, the deep structure of the central Pyrenees were still poorly constrained. The precise amount of convergence indicated by different geological interpretation is rather controversial, with values ranging from 100 km (Roure et al., 1989) to 147 km (Muñoz, 1992).These differences are mainly derived from the different hypotheses about the location of the missing Iberian lower crust. In the first hypothesis suggested by Muñoz (1992), the Iberian lower crust subducts beneath the European Plate. But the modeling of gravity anomalies and wide-angle seismic experiments suggest that the extension of the Iberian lower crust is less than 20 km beneath the NPZ (Daignieres et al., 1989; Torné et al., 1989). However, this light subducted crustal material has probably been eclogitized, thus possesses density and seismic velocities similar to those of the surrounding mantle material. Therefore, its detection by geophysical survey is rather difficult (Roure & Choukroune, 1998). The second hypothesis is that the missing Iberian lower crust is currently stacked beneath the axial zone, leading to the thick crustal root confirmed by ECORS reflection survey (Roure et al., 1989). In this model, the sharp Moho step beneath the NPF suggests that the NPF was a subvertical transform plate boundary. This model implies that the Albian basins were formed as pull-apart basins along the North Pyrenean fault system. These two extreme models are both compatible with the observations of ECORS profile beneath the axial zone since the deeper seismic reflectors in the lithosphere are not resolved well. Therefore, although the ECORS-Pyrenees profile put considerable first order constraints on the crustal structures in the central Pyrenees, it was still insufficient to end the controversies on the deep architecture of the Pyrenees. The illustrations of cross-sections for these two models are shown in Figure 1.4.



(a) Unmigrated reflection data from ECORS-Pyrenees profile.

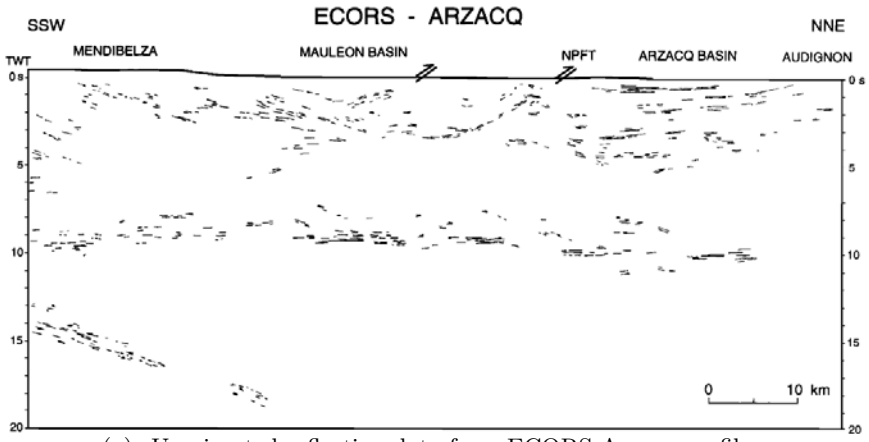


(c) Interpretations of the ECORS-Pyrenees profile.

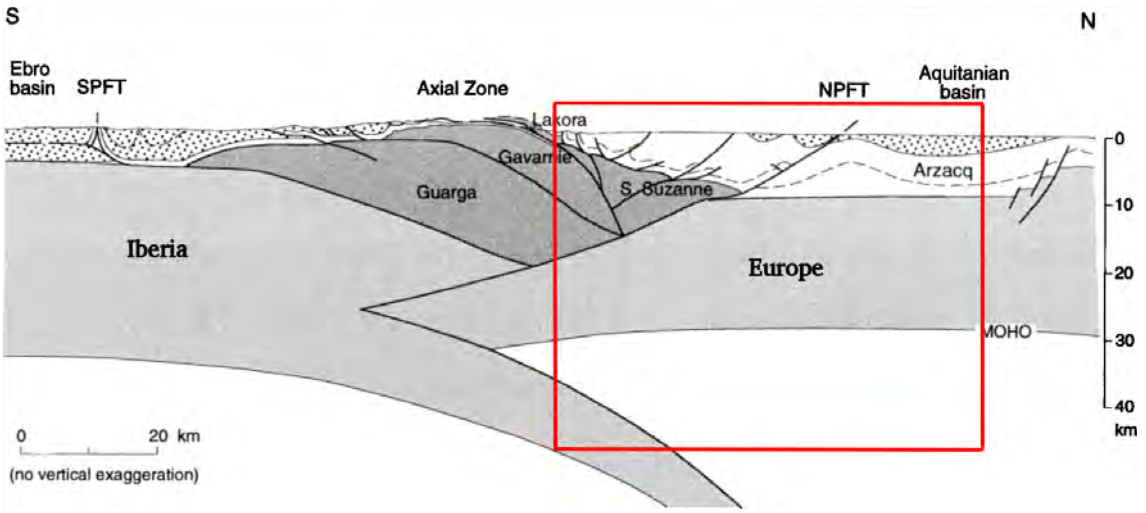
Figure 1.4: Line drawing of the unmigrated reflection data (from Roure et al., 1989) and two main interpretations of the ECORS-Pyrenees profile (from Chevrot et al., 2015). (A) Model with subduction of Iberian lower crust, as proposed in Muñoz (1992). (B) Model with stacking of the Iberian crust beneath the axial Zone, as proposed in Roure et al. (1989).

In the late eighties, another 100 km long seismic reflection profile was deployed in the western Pyrenees (Figure 1.3), at the limit between the Arzacq Basin and the Mendibelza Massif (Daignières et al., 1994). In the western Pyrenees, the middle and lower crust of the NPZ appears to be highly heterogeneous with high compressional wave velocity (Daignières et al., 1994). The previous refraction surveys show that the thickness of the crust is 45 km to the south of the Mendibelza Massif, but less than 30 km beneath the Mauléon basin (Daignières et al., 1994). Based on unmigrated seismic reflection data from the ECORS-Arzacq deployment

(Daignières et al., 1994), figure 1.5a shows the line drawing of this vertical reflection profile.



(a) Unmigrated reflection data from ECORS-Arzacq profile.



(b) Crustal section of the west Pyrenees.

Figure 1.5: Reflection data and interpretation of the ECORS-Arzacq profile (from Teixell, 1998). (a) Line drawing of the unmigrated ECORS-Arzacq vertical reflection profile, from the published data given by Daignières et al. (1994). The evident subhorizontal reflections at 9 to 10 s two-way time along the entire profile are identified as the European Moho. The dipping reflections at 14 to 18 s TWT in the southern part of the profile display the underthrusting of the Iberian lower crust. (b) Teixell's crustal section of the western Pyrenees. The cross-section shows the stacked tectonic wedges at different crustal levels, formed by deep indentation of the Iberian crust by the European crust. The structures of the crustal root and the European crust are imaged by the ECORS-Arzacq profile shown with the red frame. The dotted pattern indicates Tertiary rocks. The dashed line in the northern Pyrenees displays the top of Jurassic.

From the northern Axial Zone to the southern edge of the Mauléon basin, the ECORS-Arzacq profile imaged a series of north-dipping reflections at 14 to 18 s two-way time, which is identified as the Iberian Moho (Daignières et al., 1994). This dipping reflection Moho is more clearly visible beneath the Mendibelza Massif. From the reflection profile, we can see that a set of shallow reflections at 0 to 4 s TWT towards the north of the Mendibelza Massif put some constraints on the geometry of the North Pyrenean thrust belt and the Aquitanian foreland basin. Daignières et al. (1994) identified this TWT of 4 to 4.5 s (indicating the depths from 9 to 10 km) as the top of the Jurassic basement beneath the Arzacq basin. From the Arzacq

basin southward, these reflectors showing the top of basement shallow gradually. In addition to the north-dipping Iberian Moho, the European Moho is identified by a series of relatively flat reflections at an approximately constant two-way time of 9 to 10 s (corresponding to a depth of about 28 to 30 km) along the entire profile. The European Moho that is identified towards the southern end of the profile shows the implication of a duplication of the lower crust, along with the north-dipping Iberian Moho underthrusting the European plate. Based on the observations from ECORS-Arzacq reflection profile and other available geophysical data as well as the geometric inferences from the near-surface structures, Teixell (1998) later proposed a model for the deep indentation of the Iberian crust by the European crust which produces a double wedge beneath the axial zone (Figure 1.5b). However, the ECORS-Arzacq profile did not extend to the south sufficiently to image this possible indentation structure. In order to balance the budget of crustal material, Teixell (1998) estimated that the Iberian lower crust must be underthrusted to a depth of no less than 85 km. The shortening along the ECORS-Arzacq profile is only about 75 to 80 km, less than that of the central Pyrenees.

1.1.3.2 Regional body wave tomography

The regional body wave tomography is a powerful and conceptually simple approach to image the crust and mantle structure. Following the development of the instrumental and increasing number of deployed seismographic stations, the resolution of tomographic images has improved dramatically over the last decades, revealing more details of the lithospheric structures. An earlier tomographic study of the Pyrenees was performed by Souriau and Granet (1995) by analyzing teleseismic P waves records at Pyrenean stations. Their P wave model in the crust showed two high P wave velocity and high density bodies in the central and western parts of the NPZ. The S wave model in the crust and P wave model in the upper mantle were also derived. Similar to the P wave model, the S wave model also detected these high velocity anomalies in the crust. These anomalies were later interpreted as lower crustal material uplifted through the upper crust during the extension period before the collision (Vacher & Souriau, 2001). However, this early tomographic study suffered from the great heterogeneity, insufficient coverage of seismographic stations, a lack of stations in the center and at the western end of the Pyrenees, the fairly small number of records, and the poor quality of phase readings at some old analogue stations.

Owing to the redeployment of permanent digital stations in 1996-97 with a more even distribution on both sides of the Pyrenees, Souriau et al. (2008) conducted a new tomographic study for P wave velocity based on this improvement in ray coverage. In this study, introducing crustal corrections computed in a model derived from previous refraction and reflection experiments removed the strong contamination of the crust down to 100 km depth. Indeed, with crustal corrections, the large low velocity anomaly that was observed down to about 100

km depth in Souriau and Granet (1995), and interpreted as subducted Iberian crust (Souriau & Granet, 1995; Vacher & Souriau, 2001), completely disappeared. However, because of the overall poor quality of the manual phase readings and the very limited north-south aperture of the seismological array, the resolution in the deep parts of the tomographic model is still limited, which led to uncertain results for the deep structures beneath the Pyrenees. Another remarkable result of this study was the absence of any continuous deep high velocity anomaly all along the range, which suggested that the opening of seafloor between Iberia and Eurasia was rather limited if not absent before the collision.

1.1.4 The PYROPE and IBERARRAY experiments

Classical imaging methods are limited by either insufficient coverage (e.g. regional body wave tomography), or by a relatively poor illumination of deep reflectors (e.g. active seismic exploration experiment). The temporary PYROPE and IBERARRAY experiments have recently improved the distribution and density of seismological stations in the Pyrenees, providing a new and special opportunity to image the deep structures beneath the Pyrenean orogenic belt with very high resolution.

Following the spirit of the USArray from Earthscope, the IBERARRAY network was designed to cover the entire Iberian Peninsula in three successive footprints from south to north, each lasting for about 18 months. It officially started in 2007 with an interstation spacing of approximately 60 km. After the starting of IBERARRAY project, the French seismological community decided to grasp this particular opportunity to deploy a temporary array consisting of broad-band stations in the south of France, and around the Bay of Biscay, that would be synchronized with the launching of the third IBERARRAY deployment (Chevrot et al., 2014). This project, named PYROPE, officially set off in September 2009 and lasted for 4 years. The 2D array of the PYROPE and IBERARRAY experiments comprises 130 broad-band stations, providing dense and uniform coverage over a broad area around the Pyrenees.

In addition, in order to obtain finely resolved images for the deep architecture of the Pyrenees beneath the two ECORS profiles, two dense transects of temporary stations with broad-band seismometers were deployed across the Pyrenees from 2011 to 2013, during the PYROPE experiment (Chevrot et al., 2015). The first transect for the central Pyrenees, including 37 stations installed from October 2011 to October 2012, approximately followed strike of the previous ECORS-Pyrenees profile. The second transect for the western Pyrenees, including 29 stations installed from October 2012 to October 2013, followed a line ranging from Pamplona to Mont-de-Marsan. The interstation spacings of each transect are between 4 and 7 km. Both transects have a denser coverage in the axial zone to better constrain the complicated crustal structures there. The western transect is further extended towards the south to cover the region where the deep indentation of the Iberian crust by the European crust may occur. The map of sta-

tions of these two dense transects is shown in Figure 1.6. Except for precisely constraining the deep structures of the Pyrenean orogenic belt, the other motivation of the deployment of dense transects along previous ECORS profiles was to compare the potential of passive and active imaging approaches and discuss their complementarity (Chevrot et al., 2014, 2015).

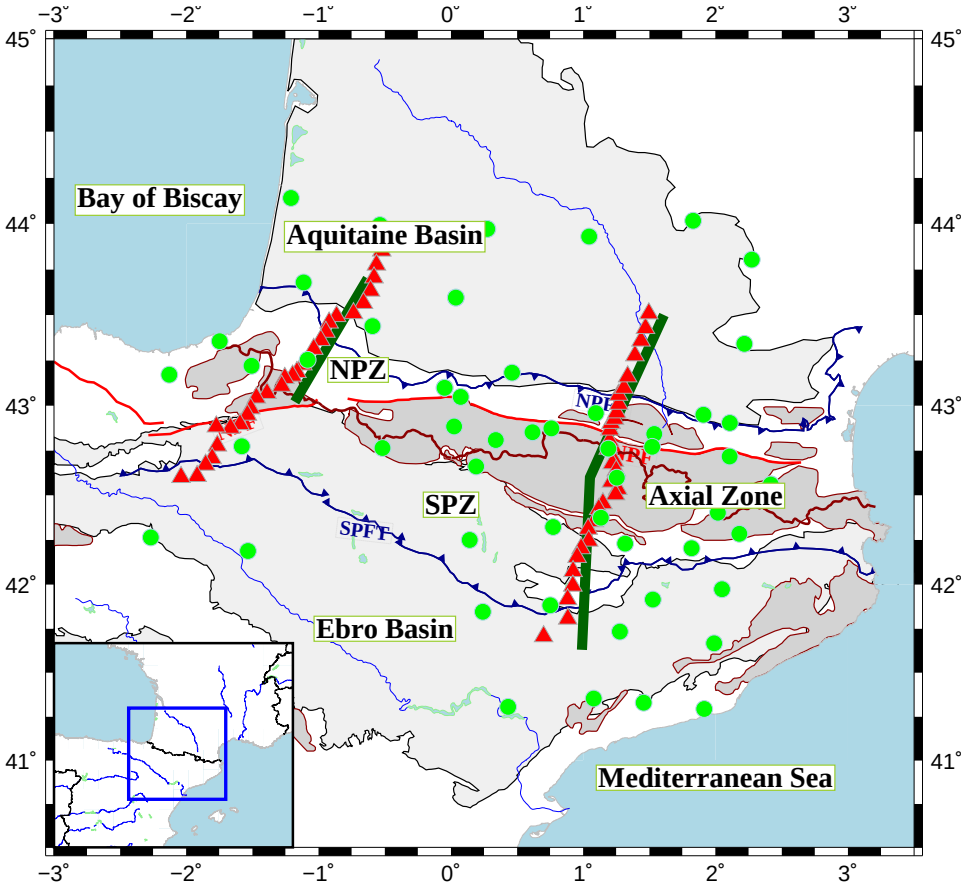


Figure 1.6: Map of the seismological stations in the Pyrenees. The temporary stations deployed along the two dense transects (red triangles) were complemented by the temporary IBERARRAY and PYROPE broad-band stations and permanent stations (green circles). The thick darkgreen lines show the strikes of the ECORS-Pyrenees and ECORS-Arzacq profiles.

1.1.5 New results from the data of the PYROPE and IBERARRAY experiments

The first tomographic study exploiting the data of the PYROPE and IBERARRAY experiments imaged the deep structures beneath the Pyrenees (Chevrot et al., 2014). Compared to previous tomographic studies (Souriau & Granet, 1995; Souriau et al., 2008), this study relied on absolute and relative travel time measurements obtained by nonlinear simulated annealing waveform fit (Chevrot, 2002) and on accurate crustal corrections computed in a detailed crustal model of the Pyrenean region. Figure 1.7 shows the map views and vertical cross-sections of this regional tomographic model. The map views of the tomographic model show the clear segmented

lithospheric structures beneath the Pyrenees. Although the P wave velocities of European lithosphere are generally faster than those of Iberian lithosphere at lithospheric depth (125 to 150 km), the limit separating the fast and slow lithospheres does not exactly follow the NPF. This limit is towards the north of the NPF in the western Pyrenees and towards the south in the eastern Pyrenees, showing a NE-SW segmentation of lithospheric structures mainly by two major faults (the Toulouse fault (TF) and the Pamplona fault (PF) drawn with black dashed lines in top right panel of figure 1.7). These inherited Hercynian faults are identified as the main segmentation of a rift that affected the formation of the Pyrenees, which would suggest that the E-W left-lateral movement of Iberia relative to Europe before the rifting episode (Jammes et al., 2009).

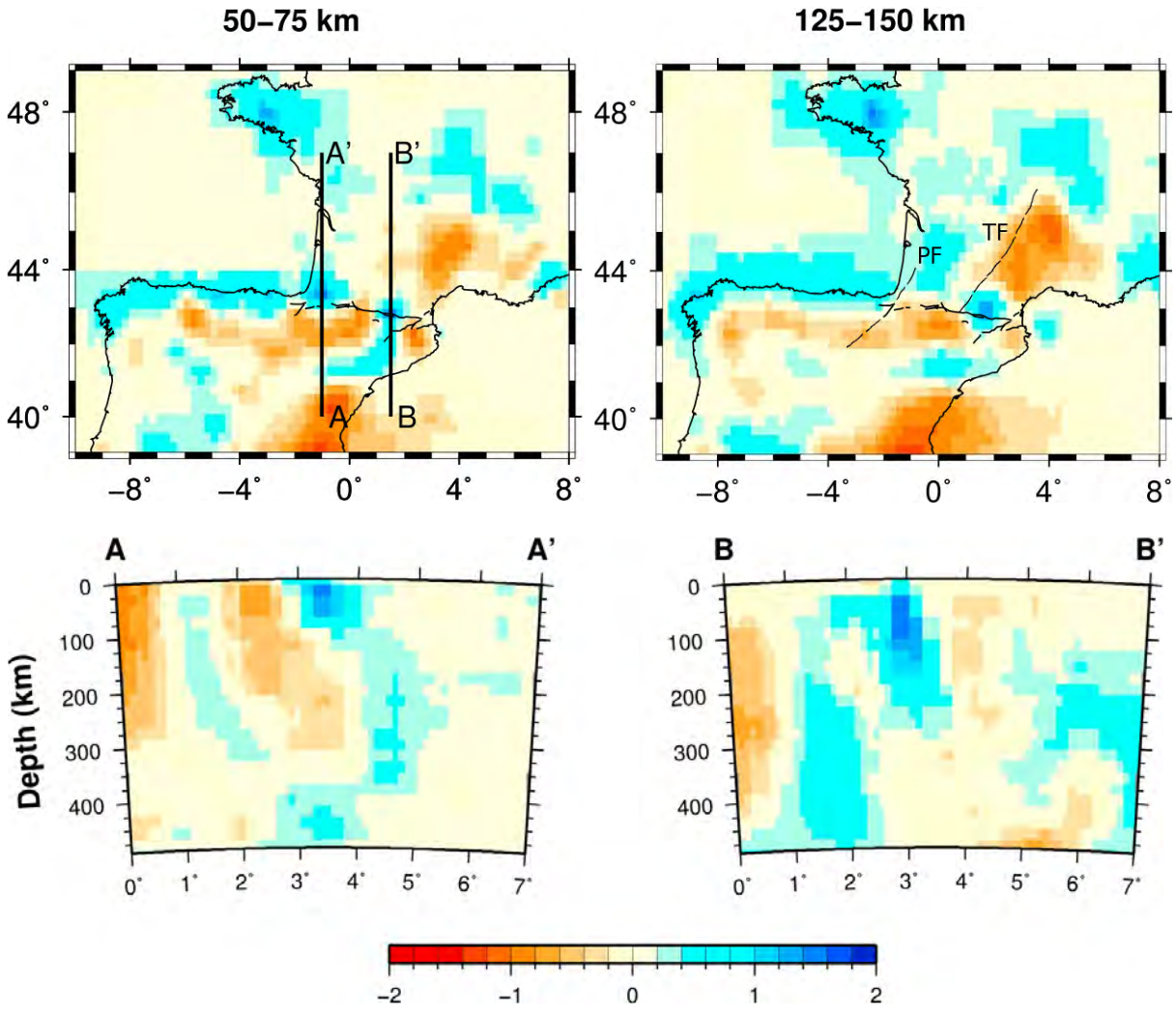


Figure 1.7: Top panels: Map views of the P velocity tomographic model obtained after crustal corrections at 50 and 125 km depth. Bottom panels: N-S vertical cross-sections along longitudes (left) 1.0° W and (right) 1.5° E. The surface strikes of the two cross-sections are shown in top left panel. The color scale is in percents with respect to the ak135 reference Earth model for all the plots (from Chevrot et al., 2014).

A pronounced fast velocity anomaly is observed at shallow depth in the Labourd region

(vertical section A-A' of figure 1.7). This fast anomaly almost disappears below 80 km depth and spatially coincides with the pronounced Labourd positive Bouguer anomaly shown in figure 1.8. It can also be explained as a shallow isolated block of mantle material (Casas et al., 1997) due to the strong velocity gradients observed in the tomographic model. The more significant fast velocity anomaly is beneath the eastern Pyrenees between 50 and 200 km depth shown in vertical section B-B' of figure 1.7, which shows a good consistency with the previous tomographic result (Souriau et al., 2008). This deep fast anomaly may be attributed to a thermal origin located in the deeper European lithospheric mantle. In addition to these prominent velocity anomalies, the absence of a deep evident fast velocity anomaly in the upper mantle and transition zone along the entire Pyrenean domain seems to exclude the presence of a detached oceanic lithosphere beneath the European plate, which precludes the subduction of oceanic lithosphere beneath the Pyrenees during the convergence.

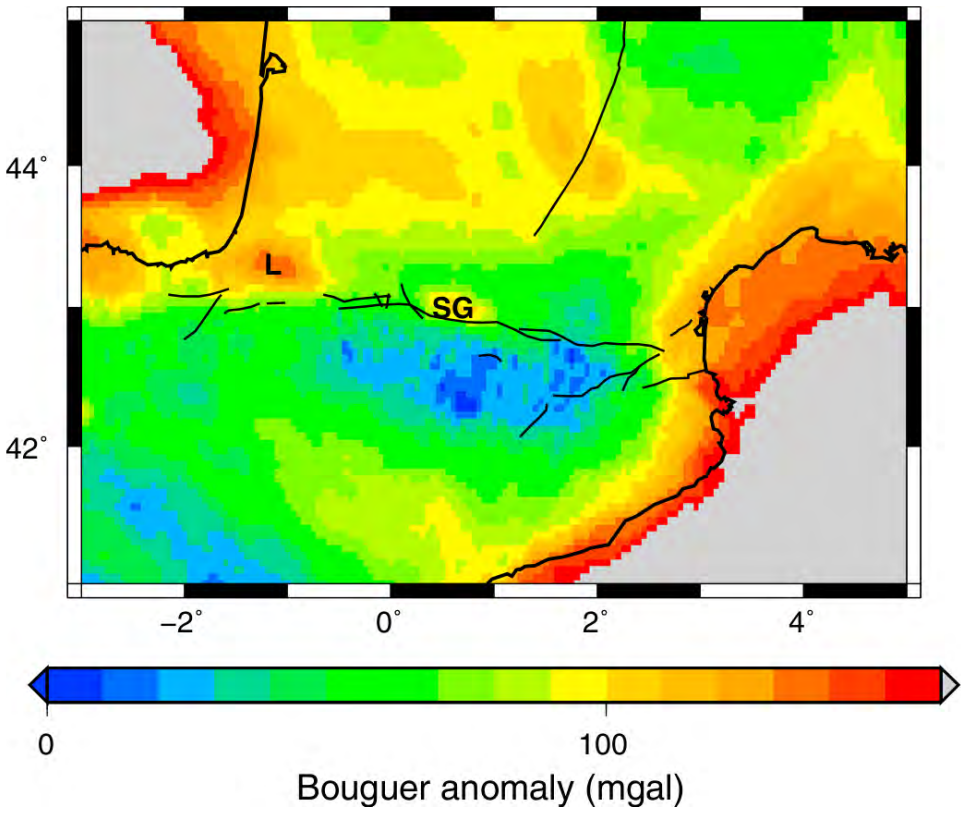


Figure 1.8: Map of Bouguer anomaly (in mGal) in the Pyrenean domain, showing the Labourd (L) and Saint Gaudens (SG) positive anomalies (from Chevrot et al., 2014).

To conclude, the new tomographic model suggested the subduction of a segmented Iberian lithosphere beneath the European plate, further supporting the inference that the formation of the Pyrenees is caused by the tectonic inversion of a segmented rift that was buried by subduction beneath the European Plate. This interpretation indicates that some fragments of the Iberian lithosphere should have subducted beneath the European Plate, which would prefer

the model proposed by Muñoz (1992) for the ECORS-Pyrenees profile and the kinematic model of Jammes et al. (2009).

Thanks to the broader frequency content and larger amplitude of teleseismic waves that can better illuminate lithospheric structures from below, passive teleseismic imaging extracting more information from the waveform data has a greater potential to locate deep seismic interfaces than active source imaging. Chevrot et al. (2015) exploited the teleseismic P waves recorded by the two dense PYROPE transects to obtain a detailed 2D cross-section of seismic interfaces from the migration of P-to-S conversions. Abundant new details of the deep architectures of Iberian and European Moho are revealed in the new images obtained by migrating the radial receiver functions with a common conversion point (CCP) stacking technique (Figure 1.9).

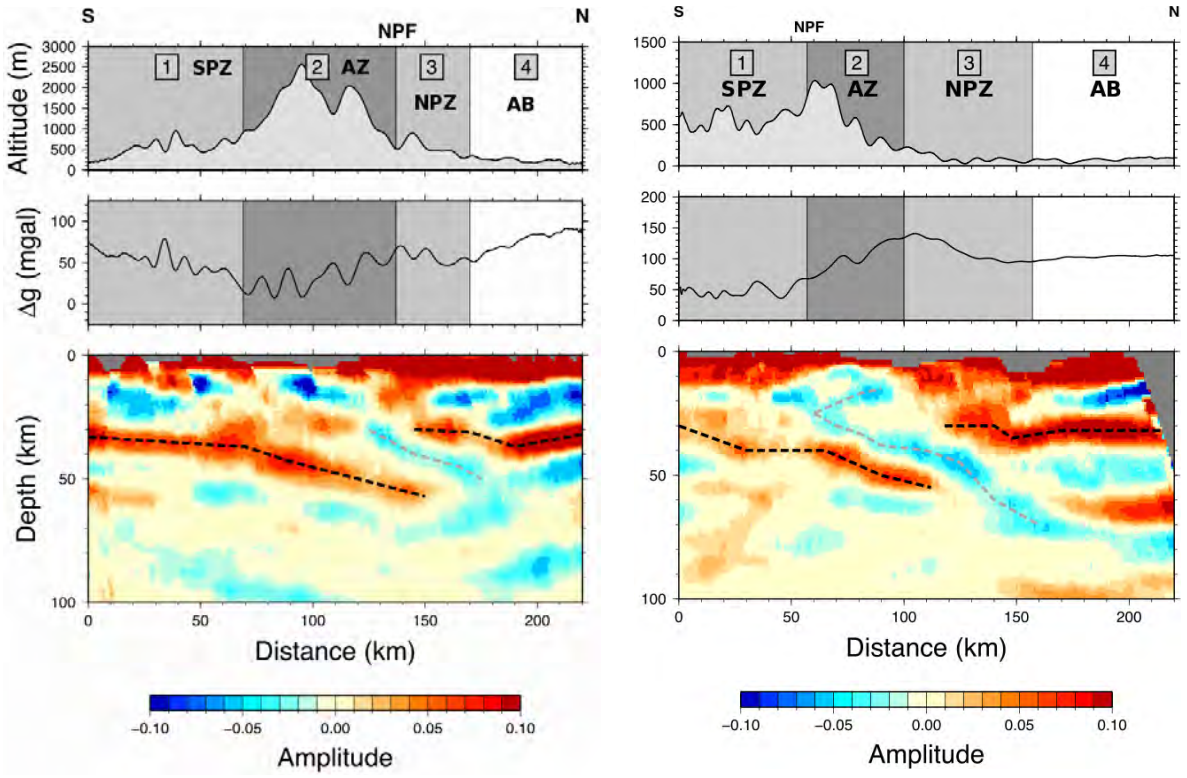


Figure 1.9: Topography (top panels) and Bouguer anomalies (middle panels) along the central (left panels) and western (right panels) transects. Bottom panels: CCP stack section for the two transects. The Moho, which corresponds to a positive downward velocity jump, appears in red while the top of the subducted Iberian crust appears in blue. The Iberian and European Moho are represented as black dashed lines and the top of the subducting Iberian crust as a grey dashed line (from Chevrot et al., 2015).

The migrated sections of both profiles show strong evidence for the Iberian crust subducting down to at least 65 km depth beneath both the central and western Pyrenees. This subduction of the Iberian lithosphere results in reconsidering the amount of convergence between Iberia and Eurasia during the Cenozoic. The images also show that the subducted Iberian crust is much thinner than a typical continental crust, from 15 km in the western Pyrenees to 25 km in the central Pyrenees. The different and more remarkable western crustal thinning suggests that

the Albian extension in the Pyrenean domain was more pronounced to the west. The image for the central profile shows a clearly disconnection between the Iberian and European Moho. This disconnection is related to a strong Moho offset, which approximately corresponds to the position of the North Pyrenean Fault. The image for the western profile also detects a similar Moho offset beneath the NPZ. A 'crocodile' pattern can be observed at crustal level beneath the western Pyrenees clearly support the interpretation of the ECORS-Arzacq reflection profile proposed by Teixell (1998), which shows the Iberian lower crust plunging into the European mantle down to a depth of 60 km. All these results have significant implications for the present architecture of the Pyrenean lithosphere.

Except for the seismic observations, the Bouguer gravity anomalies also show important indications of lithospheric structures. The Labourd and Saint Gaudens positive anomalies observed in the Pyrenees (figure 1.8) could be attributed to the presence of dense materials at shallow depth. For both transects, the variations of Bouguer anomalies show simple patterns that can directly correspond to the structures shown in the migrated sections. In the central Pyrenees, the Bouguer anomalies in South Pyrenean Zone decrease towards the north may reflect the gradually deepening of the Iberian lower crust. Beneath the Axial Zone, the Bouguer anomalies increase to reach a peak value around the NPF, which may be interpreted as uplifted European lower crust or mantle material. For the western transect, the evident Labourd positive anomaly coincides spatially with the crustal structure showing the indentation of the Iberian crust by the European crust. In addition to this, we can observe a 50 km shift between possible deep crustal root and topographic highs which is departure from local (Airy) isostasy. At this point, the further geological interpretations of these observations will require 3D images of density and velocity structures with a much finer resolution than in current images.

1.1.6 Discussion and directions for further improvements

1.1.6.1 The discussion of the model coming from regional body wave tomography

The first tomographic study brought new insights into deep lithospheric structures beneath the Pyrenees, but also revealed the limitations of classical passive imaging approaches.

The crustal model used to compute crustal corrections in the regional tomography was obtained from analysis of receiver functions (RF). In general, it is extremely difficult to locate the Moho depth below the stations deployed in sedimentary basin, owing to strong reverberations inside shallow unconsolidated sedimentary layers that masked conversions on deeper seismic interfaces. In addition, constraining the crustal thickness can be also problematic in stations deployed in the axial zone, due to their receiver functions usually showing very complex signatures. Except for the considerations mentioned above, the crustal model constructed based on RF had a much coarser resolution in the Pyrenean domain, especially for the mountain ranges where the highest resolution is generally required. The density of PYROPE and IBERARRAY

deployments is still insufficient in the Pyrenees, where crustal structure is rather heterogeneous and can vary significantly over very short distances.

One critical and general conclusion from this regional P wave travel time tomography study is that the lateral resolution at 100 km depth is about 25 km, but the vertical resolution is much poorer due to the effect of vertical smearing in teleseismic tomography. Chevrot et al. (2014) also showed that the vertical resolution is of the order of 80 to 100 km at lithospheric depths, with a quick deterioration at greater depth. Such a resolution is enough to map major structural units, but it is insufficient to provide sharp interfaces and details of lithospheric architecture. One reason for this poor vertical resolution in regional tomography is the subvertical incidence of teleseismic body waves and the fact that the sensitivity of travel time with respect to seismic velocities is broadly distributed around the geometrical ray path. These fundamental limitations of travel time tomography are inevitable, even if finite frequency effects are considered.

1.1.6.2 The comparison between active and passive imaging

Regarding the results of RF migration, the comparison of active and passive internal discontinuity imaging is insightful. The advantages of passive imaging technique over active imaging techniques are:

1. Passive imaging brings additional important information that could not be inverted by active imaging, especially for the deep seismic interfaces that are hard to illuminate by using active experiments due to power limitations of man-made seismic sources.

2. Passive imaging approach based on phase conversions other than detection of deep reflectors in seismic reflection experiment can characterize the sign of the velocity jumps on internal discontinuities, which provides significant guidelines for geological interpretation.

On the other hand, the disadvantages of passive imaging are:

1. Passive approach is difficult to image the structure of upper crust in details, because it is hard to isolate the P-to-S conversions on shallow seismic discontinuities from the direct P wave. By contrast, the higher frequency content of seismic reflection data allows accurate characterization of reflectors at shallow depth, which are usually correlated to well logs data for precisely determining the nature and composition of the main crustal units.

2. The strong reverberations appearing in shallow sedimentary layers, which often mask conversions on deep seismic interfaces, is a fundamental limitation of passive imaging performed in the region including basins. Because the foreland basins usually surround the mountain ranges, this problem often happens in practice. A possible remedy is to exploit S-to-P (Sp) conversions, which arrive earlier than the direct S phase and are not contaminated by such later arriving reverberations (Hansen et al., 2009).

The comparisons show that passive and active imaging technique have both merits and limitations. Sometimes they can complement each other. The ideal situation is that we can

utilize both sources of information, as in the study of the Pyrenees, which offers us a unique opportunity to improve our understanding of the formation of the orogenic belt.

1.1.6.3 The limitations of receiver function migration and directions for further improvemens

In addition to important results already obtained (Chevrot et al., 2014, 2015), a large part of PYROPE and IBERARRAY data has not been fully explored yet. CCP stack technique suffers from several limitations:

1. The migration is performed in a 1D background model. Neglecting the strong lateral variations of seismic velocities will more or less result in distortions in the recovered depth and geometry of seismic interfaces.

2. An implicit assumption of CCP stacks is that the phase conversions occur on the subhorizontal discontinuities. Although this algorithm can still detect dipping interfaces, the apparent dip will be biased. Schneider et al. (2013) performed synthetic experiments to test the validity of CCP stack of receiver functions by assuming horizontal and dipping converters for target models with different dipping interfaces. Their results showed that the migrated dipping interfaces tends to be shifted to a shallower position when horizontal converters are assumed. The bias between migrated and target inclined interface increases with the dip angle. These artifacts could be reduced by 3D Kirchhoff migration technique (Bostock et al., 2001).

To summarize, migration approach is very efficient in imaging small-scale structural heterogeneities, such as velocity discontinuities. However, compared to classical travel time tomography, it can not retrieve the velocity structures. Considering the shortcomings of above imaging approaches and the requirement of 3D high resolution images of density and velocities structures for geologic interpretations, the next stage will be to move from simplified migration and tomographic methods relying on asymptotic descriptions of the wavefield towards full waveform inversion methods exploiting the complete wavefield (Liu & Gu, 2012).

1.2 Towards full waveform inversion of short period teleseismic body waves

Obtaining 3D tomographic images of the internal structures of orogenic belts with a sufficiently fine resolution to characterize their detailed lithospheric architecture is very challenging. It requires a method that can take full advantage of the information contained in seismic records. The full waveform inversion (FWI) which exploits as much information contained in seismograms as is physically reasonable, is a very promising selection. Indeed, It has a theoretical resolution limit of the order of half of the shortest wavelength in the seismic wavefield (Virieux & Operto, 2009), superior to any other traditional tomographic techniques. Therefore, imaging

the lithosphere with a resolution of the order of a few kilometres will require inverting waveform records in the period range 1 to 10 s.

In general, the incident seismic waves excited by either distant or local earthquakes can be used to image lithospheric structure. However, considering the purpose of our study and the size of our target regions, the teleseismic waves originate from distant sources are chosen here mainly by the following reasons. First, the direct primary phase and its coda are well isolated from other primary phases since the temporal spacing between specific phases tends to increases with distance. Second, the ray coverage of teleseismic waves can provide more constraints on the deep structure of the target region. Third, the teleseismic waves are less affected by errors of sources location than incident waves excited by local earthquakes. Finally, if the size of the regional domain is of the order of several hundreds of kilometers, the incident teleseismic waves recorded by regional array have almost the same incidence angles. They are thus insensitive to possible heterogeneities outside the regional domain.

Based on these considerations, we will perform full waveform inversion on short period body waves records excited by distant earthquakes. This is a natural extension of previous studies in the Pyrenees (Chevrot et al., 2014, 2015).

FWI searches for a model that can predict all the wiggles in observed seismograms. It is particularly advantageous because it allows for the direct analysis of strongly interfering signals caused by heterogeneity in the model, without the need to identify complex seismic phases. FWI updates the model in an iterative way that accounts for the non-linear relationship between model parameters and synthetic waveforms. The resolution and precision of the FWI critically rely on the forward modeling technique. A good forward modeling technique is a key ingredient for a successful and efficient FWI. In the following, we review the development of FWI and of forward modeling techniques to identify the methods that are best suitable for our goals.

1.2.1 A review of FWI approaches

In this section, we review the different aspects involved for the implementation of full waveform inversion. The first part gives an overview of the historical development of full waveform inversion, including its applications both in the time and frequency domains. In the second part, the computation of sensitivity kernel in heterogeneous model by the beautiful adjoint state method and the recent developments of adjoint tomography are introduced.

1.2.1.1 Historical development of full waveform inversions

Full waveform inversion methods have been introduced in exploration seismology at the beginning of the eighties (Tarantola & Valette, 1982; Lailly, 1983; Tarantola, 1984). Early studies focused on 2D acoustic problems. They formulated an inverse problem which consists in searching for the best model that minimize the sum of generalized least square differences between

observed seismograms $\mathbf{d}(\mathbf{x}_r, t)$ and synthetic seismograms $\mathbf{s}(\mathbf{x}_r, \mathbf{m}, t)$, at the receiver positions $\mathbf{x}_r$. The scattered wavefield is generated by the interaction of the incident wavefield with the heterogeneous perturbations of the background velocity model. Based on a first order scattering approximation, the gradient of the waveform misfit function with respect to model parameters is computed by correlating the forward propagating wavefield from the actual source with the backward propagating secondary wavefield generated by weighted time-reversed data residuals at the receiver, which represent as fictitious sources in the current model.

Tarantola & Valette (1982) showed that the first iteration of the time domain waveform inversion is equivalent to pre-stack migration techniques based on Claerbout's imaging principle (Claerbout, 1971). Pre-stack migration can be treated as a partial inversion, but obtains the reflectivity images without visual physical interpretation. In contrast to traditional migration techniques, the nonlinear waveform inversions update the model parameters by successive iterations.

Following this pioneering work by Tarantola, many studies investigated nonlinear waveform inversion approaches in the time domain. Gauthier et al. (1986) carried out the first 2D numerical nonlinear waveform inversion in the acoustical case. Their test proved that the nonlinear inversion of seismic waveforms is feasible with a steepest descent gradient based inversion algorithm, combined with accurate numerical waveform modeling technique. This numerical test also pointed out that if the initial model is far enough from the true model (errors of up to 10%), the gradient based waveform inversion will be trapped inside a secondary minimum.

The acoustic case was later generalized to the elastic (Tarantola, 1986) and anelastic (Tarantola, 1988) cases. Tarantola indicated that an adequate choice of model parameterization is critical as long period waveforms are mainly sensitive to seismic velocities while short period waveforms are significantly affected by model impedance contrasts in exploration scale problems.

Further numerical synthetic tests of elastic waveform inversion for multi-component reflection data (Mora, 1987) and transmission data (Mora, 1988) showed that surface reflected data (shot gathers) mainly resolve the high frequency content of the model. Adding transmitted VSP data improves the resolution of the low frequency content of the model. Mora also derived a matrix formulation of the inversion problem and suggested that the best preconditioner is the inverse Hessian matrix of the waveform misfit function with respect to model parameters. He also concluded that an iterative elastic waveform inversion algorithm could resolve all the frequency contents of the model parameters when the following two conditions are satisfied: 1) The starting model is close to the true model in the sense that the first order Born approximation holds and 2) the heterogeneities are illuminated from various directions either by transmitted or reflected waves. The waveform inversion can be regarded as a combination of iterative migration and reflection tomography approaches.

Pica et al. (1990) performed iterative procedures equivalent to individual reverse time mi-

grations for waveform inversion, implemented on a real marine data set. The final synthetics computed in the inverted model matched the input data very well, and the inversion results were in good agreement with the adjusted and smoothed real sonic logs.

Although moderate success has been achieved when applying FWI to real industrial data (Crase et al., 1990; Igel et al., 1996), the challenges regarding the nonlinearity of the waveform misfit function and dependence on the starting model remain unsolved issues. Gradient-based algorithms for FWI can not mitigate these problems, need to be further explored.

In order to address the pitfalls of local minima, global optimization methods were introduced. Nonlinear global optimization methods that originate from thermodynamic and cellular systems, such as simulated annealing (Sen et al., 1997), genetic algorithm (Sambridge & Drijkoningen, 1992) can be used for FWI. With proper parameter tuning, these global optimization methods are suitable for inversion problems with high degrees of nonlinearity or multiple extrema of the misfit function. Probabilistic inversion theory can also be used to find the solution of the inversion problem, introducing a marginal probability density assigning in model space. In general, this probability density distribution is not simple, which can only be characterized by exploring the model space, usually by Monte Carlo methods (Tarantola, 2005). Although very general, these approaches suffer from a heavy computational cost owing to the very large number of forward modelings required to explore the model space.

FWI involves fitting arrival times and amplitudes of the different seismic phases in a chosen time window. The arrival times are generally quasi-linearly related to the structure along the ray path, unlike the amplitude of the seismic waves. Luo & Schuster (1991) tried to address this nonlinear problem by introducing cross-correlation misfit functions that quantify the travel time difference between observed and synthetic seismograms. They noted that the cross-correlation travel time misfit functions are much more linearly related to model parameters compared to full waveform misfit functions. However, the inverted models obtained with this misfit function have a much lower resolution. This would suggest to invert cross-correlation travel times to first determine the long wavelength structure of the model. A new full waveform inversion can be performed starting from this inverted model, avoiding being trapped inside a local minimum.

A different idea to address the nonlinearity of FWI is multiscale decomposition (Pratt et al., 1998; Pratt, 1999). Performing inversion successively from long to short wavelengths is a powerful strategy to approach the global minimum (Nolet, 2012). For long wavelength measurements, the number of the local minimum is greatly reduced, and the inversion problem is more likely to converge towards the global minimum. Bunks et al. (1995) presented a multigrid method for FWI. This multigrid method improves the performance of iterative inversion by decomposing the problem at different scales. At large scales, only few local minima exist and are further apart from each other, so that it becomes easy to locate the neighborhood of the global minimum. The inverted model at large scale can serve as a starting model for subsequent inversions at smaller scales. This method was tested on a subsampled, low-frequency version

of the 2D Marmousi data set. The results showed that multiscale iterative waveform inversion performs very well, and that it can significantly reduce the computational cost of the inversion. It will thus be more efficient for 3D applications.

As a natural extension of this idea, multiscale methods using wavelet bases and frames can be formulated in the data space (Wu et al., 2007), in model space (Chiao & Kuo, 2001), or in the space of the sensitivity kernels (Wu & Yang, 1997; Chevrot & Zhao, 2007). Wavelets methods offer a great flexibility to select the basis functions or frames to describe the model.

Another successful development is the full waveform inversion method based on numerical solvers in the frequency domain (Pratt et al., 1996; Pratt, 1999). According to Marfurt (1984), the most efficient numerical discretization technique to treat the forward modeling of the multi-source problems is the finite-difference or finite-element method in the frequency domain. In the frequency domain, the wave equation reduces to a sparse linear system:

$$\mathbf{B}(\mathbf{x},\omega)\mathbf{u}(\mathbf{x},\omega) = \mathbf{s}(\mathbf{x},\omega) \tag{1.1}$$

where the right hand side $\mathbf{s}$ is the source, the solution $\mathbf{u}$ is the seismic wavefield, $\mathbf{B}$ is the impedance matrix. It can be solved with a parallel LU decomposition. The LU decomposition of the impedance matrix $\mathbf{B}$ obtained by the discretization of the forward problem is also reused to solve the adjoint problem. The solution of the adjoint problem is the backpropagated wavefield which is used to compute the gradient and Hessian along with the forward wavefield. This is particularly important when the number of sources is large. In that case, the same LU decomposition is used to solve the forward and adjoint problems related to each source at a given frequency. This feature is one of the critical reason for FWI efficiency in the frequency domain: if the LU decomposition of the impedance matrix can be stored, this method has a substantially lower computational cost, compared to time domain FWI. Model are first inverted at a low frequency to mitigate the nonlinear effects, and then successively inverted at higher frequencies using the inversion results obtained in the previous frequency range. The direct LU solver approach in frequency domain FWI is efficient for 2D problems. However, the huge computational costs and memory requirements of LU decomposition are a strong limitation for large-scale 3D problems by using frequency domain FWI(Virieux & Operto, 2009).

The application of full waveform inversion has so far been mainly limited to exploration scale (Pecher et al., 1996; Ravaut et al., 2004; Plessix & Perkins, 2010; Plessix et al., 2012; Sirgue et al., 2010). In theory, although FWI can be directly applied to crustal or lithospheric scale imaging, several problems need to be addressed first.

Firstly, as pointed out by Fichtner et al. (2008), obtaining a sufficiently accurate starting model is difficult. This is due to the limited spatial coverage when using natural earthquakes. Even the most popular 1D Earth models (Dziewonski & anderson, 1981; Kennett et al., 1995) can apparently differ by several percents, especially in the upper mantle and transition zone.

In addition, the 1D density model of the Earth is rather poorly constrained (Kennett, 1998). To summarize, it is difficult to find a proper initial background model for FWI at regional or continental scale.

Secondly, the data coverage is fairly uneven in the earth. Most regions of the Earth's surface are practically inaccessible (like the oceans, which cover more than 70% of the surface of the Earth), and sufficiently strong earthquake sources are found only in a few tectonically active regions. For all these reasons, FWI at continental or global scales is difficult, unless very long period data are used (Capdeville et al., 2005).

Thirdly, regional scale imaging of the lithosphere mainly relies on teleseismic body waves. The computational cost of precise 3D numerical forward modeling at continental or global scales are still very high. Consequently, most tomographic studies still rely on the exploitation of travel time data to image the internal structures of the Earth, discarding the amplitude information which needs accurate numerical simulations and very high quality data.

However, with the recent progress of imaging algorithm as well as the rapid development of dense temporary deployments at regional scale, FWI applications are getting more and more attention.

1.2.1.2 The adjoint state method

The computation of sensitivity kernel (also known as Fréchet derivative) which relates the data space to the model space is a key issue in the nonlinear optimization problems. The Fréchet derivative of a misfit function χ with respect to model parameters $\mathbf{m}$, denoted by $\nabla_m \chi$ is defined as:

$$\nabla_m \chi \delta \mathbf{m} = \lim_{\varepsilon \to 0} \frac{\chi(\mathbf{m} + \varepsilon \delta \mathbf{m}) - \chi(\mathbf{m})}{\varepsilon} \tag{1.2}$$

It is the key component to building a gradient-based inversion method and for quantitative resolution analysis. The main difficulty is that computing $\nabla_m \chi$ requires the calculation of $\nabla_m \mathbf{u}$, which represents the Fréchet derivative of the wavefield $\mathbf{u}$ with respect to the model parameters $\mathbf{m}$. Because of the huge size of the model space, it is impractical to compute this quantity by classical finite difference methods for all possible perturbation directions $\delta \mathbf{m}$.

The elegant and physically insightful adjoint state method has been introduced to compute the Fréchet derivative with optimal efficiency. The adjoint state of an hyperbolic differential equation, such as the wave equation, can be found as early as 1968 in the book of Jacques Louis Lions « optimal control of systems governed by partial differential equations ». Lions's student, G. Chavent, was the first to use the adjoint state method to determine the distributed parameters, in his PhD study. Following the theoretical mathematical studies of 1D inverse problems in reflection seismic exploration (Bamberger et al., 1979), Bamberger et al. (1982) presented probably the first application of the adjoint state method to seismic imaging. The

inversion of normal incidence seismograms was formulated as an optimal control problem. In that case, the adjoint state method can be introduced naturally for iterative inversion of the 1D wave equation, together with a stabilizing constraint on the sum of absolute values of the jumps of the desired impedance distribution for stabilizing the inversion problem.

The general approach of inverse problems developed by Tarantola (1984) also shows that the adjoint state method is an effective way to explicitly compute the sensitivity kernel of a misfit function. Each iteration in structure and source parameter inversion can be accomplished by only two simulations for a specific source: one forward simulation excited by the regular seismic sources and one adjoint simulation excited by the adjoint fictitious sources located at the receiver positions which depend on the chosen misfit function.

Tarantola showed that the perturbation of a misfit function defined as squared waveform misfit residual integrated over a time window can be rewritten as a spatial integral involving the product of a sensitivity kernel with the perturbations of the model. The adjoint state method determines the adjoint wavefields, which are convolved with regular forward wavefields to construct the 3D kernels. The construction of the sensitivity kernel is relatively intuitive in physics for seismic inversion problem because of the linearity and self-adjoint property of elastic wave equation. The self-adjoint elastic wave operator leads to the corresponding adjoint equation. It has the same form as the original regular wave equation, only the adjoint source terms differ. This property implies that the adjoint equation is able to be solved by using the same numerical technique as for solving the forward regular wave equation. Using the same numerical solver for both regular and adjoint equations allows us to focus on developing highly efficient and accurate numerical methods.

Following the influential work by Tarantola in 1980s, the adjoint state method has been widely used to tackle a variety of problems. Nowadays, seismologists generally name 'adjoint tomography' the tomographic approaches based on 3D numerical modeling of seismic waves and the sensitivity kernel computed by adjoint state method. In the following, we will briefly review some recent adjoint tomography applications in seismology, which provide some meaningful insights to our work.

The seismic inversion problem can be solved as a problem of minimization of a misfit function that measures the difference of certain seismic observable between data and synthetic seismograms. The efficiency of the inversion algorithm is determined by the precision of the computation of gradient of the misfit function. In adjoint tomography, it requires two wavefield simulations and a properly chosen gradient based optimization algorithm. The model parameters are updated iteratively during the inversion.

Geller & Hara (1993) present two equivalent algorithms for iterative linearized waveform inversion for 3D Earth structure starting from a 3D model. One relies on a matrix formulation, and the other on a wavefield formulation. Both algorithms require the accurate modeling of synthetic seismograms, but neither specifies a particular method. In spite of the small difference

in computational costs, the wavefield algorithm which is equivalent to the adjoint state method for inversion of elastic wavefield data by Tarantola (1986) and Mora (1987), appears to be more efficient. After that, a series of works have been devoted to waveform inversion focusing on the global structure of the upper mantle, computing the sensitivity kernels with the Direct Solution Method (DSM) (Geller & Hara, 1993; Hara & Geller, 2000; Hara, 2004).

Akcelik et al. (2002, 2003) present a high accuracy finite-element method for simulating high frequency seismic wave propagating in complex 3D basins (LA Basin, maximum frequency for earthquake simulations up to 2 Hz, lowest shear wave velocity down to 100 m/s) on the terascale supercomputers. Based on this precise numerical solver and adjoint state method, they showed results for media and source parameters inversion of high resolution models of 2D sedimentary basins undergoing antiplane motion, by performing parallel multiscale Gauss-Newton conjugate gradient inversion algorithms.

Tromp et al. (2005) discussed the theoretical connections between adjoint state method, finite frequency tomography, and time reversal imaging in acoustic (Fink, 1992; Fink & Prada, 2001). This study illustrated the ability of the adjoint state method to isolate the region where the seismic wavefield is sensitive to. The physical idea is similar to time reversal imaging: Green's function of the current model can be used to compute the residuals between data and synthetics. Using these time reversed residuals as sources at receiver positions leads to illuminating those parts of the model that need to be perturbed, similar to a time-reversal mirror refocusing on the source of an acoustic signal. The characteristics and distribution patterns of sensitivity kernels associated with different misfit function are illustrated for a variety of waveform sections based on a 2D spectral-element method. The definition of the adjoint source is determined by the considered misfit function.

The derivation of the adjoint state method has been formulated by many researchers, either based on the single scattering Born approximation combined with the reciprocity of the Green's function (Tarantola, 1984; Tromp et al., 2005); or on the Lagrange multiplier method (Akcelik et al., 2003; Liu & Tromp, 2006; Tromp et al., 2008) where the adjoint wavefield is the time reversal of the Lagrange multiplier. Fichtner et al. (2006) made a more general functional analysis to formulate the adjoint state method. Their derivation shows that the adjoint state method can compute the exact derivatives of a misfit function without relying on the existence or reciprocity of the Green's functions. Only the transposes of partial derivatives of the differential operator are needed. The adjoint state method can be generalized to complex nonlinear operators. The self-adjointness of the elastic wave equation is directly linked to energy conservation and spatial reciprocity of elastic wave propagation.

The universality of the adjoint state method fills the gaps between traditional body wave and surface wave tomography by fully exploiting the information in seismic records. Adjoint tomography has been successfully applied to image crustal structures in southern California (Tape et al., 2009, 2010), the upper mantle structures beneath Australia (Fichtner et al., 2009,

2010), European continent (Zhu et al., 2012, 2015), Turkey and the eastern Mediterranean (Fichtner et al., 2013), the North Atlantic region (Rickers et al., 2013), as well as to invert the centroid-moment tensors (Kim et al., 2011).

1.2.2 The hybrid forward modeling technique

The modern FWI involves replacing simple forward modeling techniques (such as ray theory or mode summations) by complete 3D numerical solutions of the wave equation in heterogeneous media and computing the sensitivity kernel with adjoint state method. For our applications, we need to model the propagation of short period teleseismic body waves in a global 3D heterogeneous media, which is still challenging. Huge efforts have been made to develop suitable numerical techniques for this problem. Owing to recent theoretical and computational developments, this problem can now be tackled by seismologists. Various efficient numerical methods can be used to carry out full waveform modeling of elastic waves propagating in a 3D model, such as the finite-difference method (FD; Virieux & Operto, 2009), the pseudospectral method (PS; Fornberg, 1998), the regular finite-element method (FEM; Kallivokas et al., 2013), or the spectral-element method (SEM; Komatitsch & Vilotte, 1998; Komatitsch & Tromp, 1999, 2002). However, the need to iterate FWI problems requires performing 3D forward modeling a large number of times, leading to a rather high computational cost. In addition, because of the large size of the simulation domain, the application to complete waveforms at periods of a few seconds in the full 3D Earth requires enormous computational resources, which limits the applicability of FWI, even on the largest supercomputers (Tsuboi et al., 2003; Peter et al., 2011).

For the purpose of reducing the computational costs for inverting short period teleseismic body waveforms, some simplifications for this problem are proposed. Still keeping the three-dimensionality of the problem and precise forward solver for modeling full waveforms, a possible alternative is to use a smaller domain for 3D wavefield simulation (Monteiller et al., 2013; Tong et al., 2014b). That is, constructing a hybrid approach to model the propagation of short period teleseismic waves in two steps. The first step is to simulate the seismic waves from the source to the boundaries of the target 3D regional domain in a 1D or 2D global earth model. The second step is to model the wavefield in the regional domain by injecting the incident global wavefield. Each method in this hybrid approach can accurately solve the wave equation in its simulation domain. The key problem is then to make the global and regional wavefields match on the boundaries of the target domain. This boundary coupling approach is general for any combination of 3D and 1D methods. At the moment, most studies choose a simple one-way coupling strategy (Roecker et al., 2010; Pageot et al., 2013; Monteiller et al., 2013; Tong et al., 2014a,b). The scattered waves produced by local 3D heterogeneities may propagate outwards through the boundaries of the regional domain. They will be absorbed based on approximate absorbing condition imposed on the boundaries. The traveling out scattered waves

generated by regional heterogeneities can not be further modeled by hybrid approach any more. Such that the coda waves of teleseismic phases are only attributed to 3D heterogeneities inside the regional domain, while the heterogeneities outside the domain are ignored. This one-way coupling strategy is an important limitation in most hybrid approaches, because the outward scattered waves travel off the boundaries of the regional domain may be scattered back again by outer heterogeneities. However, the contributions of such multi-scattered waves are of second order compared to the incident wavefield, and in most applications can be neglected safely. A complete but also more complicated, fully-coupled strategy is proposed by Capdeville et al. (2003b,a).

Among the various hybrid methods, Monteiller et al. (2013) introduced the first high frequency 3D hybrid modeling method by matching an incident wavefield computed in a spherical 1D Earth model based on the DSM (Geller & Ohminato, 1994; Geller & Takeuchi, 1995; Takeuchi et al., 1996) with a regional wave propagation simulated based on the SEM by executing the SPECFEM3D open source software package. The paraxial absorbing boundary condition (Stacey, 1988) is used for one-way coupling strategy of this hybrid, with satisfactory efficiency. DSM provides precise synthetic full seismograms up to frequencies as high as 2 Hz (Kawai et al., 2006) but with a computational cost much larger than that of the FK method based on horizontally layered 1D model used for the hybrid approach developed by (Tong et al., 2014a,b). Modeling the complete incident wavefield in a spherical 1D Earth allows us to consider large and deep regional domains, which more than compensates for the additional computational cost (Monteiller et al., 2015). Another option for computing the 1D global Green's function by 2D axisymmetric numerical simulations for elementary sources (AxiSEM; Nissen-Meyer et al., 2014) is available at the moment. Other choices of 1D and 3D simulators are also feasible by carefully considering there domain of validity and frequency range of simulation. Eventually, the full 3D simulation will become a viable alternative. Even so, the hybrid approach would still have the advantage of drastically mitigating the computational cost of full waveform inversion for the regional 3D domain.

Chapter 2

Forward modeling: The Hybrid method and wavelet estimation

Contents

2.1 The 3D forward hybrid modeling technique

To simulate the propagation of teleseismic waves at the regional scale, we use the hybrid DSM/SEM method introduced in Monteiller et al. (2013). Here we will only give a brief overview of the method and refer the reader to that article for further details. The main advantage of the hybrid method is to restrict the costly 3D computations to a regional domain of limited size, which is a crucial point in terms of computational cost for iterative waveform inversion. Each of the DSM and SEM methods provides very accurate solutions of the wave equation in its own domain, and the problem is thus to match the global and regional wavefields on the boundaries of the regional mesh.

DSM is a numerical method that solves the weak form of the equation of motion for a set of frequencies required for inversion to the time domain by Fast Fourier Transformation

(FFT) (Cummins et al., 1994a,b; Geller & Ohminato, 1994; Geller & Takeuchi, 1995; Takeuchi et al., 1996). In DSM, The displacements are expanded in the frequency domain by a series of basis functions that are constructed by the product of low order polynomials in the vertical direction and spherical harmonics in the angular direction. The coefficients for these basis functions are then solved by a Galerkin method. Because of the discretization treatment in the vertical direction, the point source will usually not located between the vertical nodes. In these cases, the source representation method discussed in Takeuchi & Geller (2003) can be used to compute synthetics with the identical accuracy as for the source located at a node. Bye laborately adjusting the vertical grid spacing, maximum angular order and cut-off depth, the DSM can be executed very efficiently and accurately, even at frequencies as high as 2 Hz (Kawai et al., 2006). Therefore, this method is particularly suitable for modeling complete short period teleseismic wavefields.

In the regional domain, we use the spectral element method (SEM; Komatitsch & Vilotte, 1998; Komatitsch & Tromp, 1999; Vai et al., 1999; Komatitsch & Tromp, 2002; Komatitsch et al., 2005; Tromp et al., 2008), which is a highly accurate continuous Galerkin technique to model seismic wave propagation in elastic or anelastic (viscoelastic) media. The SEM is based upon the weak form of the seismic wave equation solved in the time domain. Because it uses high-degree polynomial basis functions, can handle very distorted meshes and does not necessitate interpolation of material properties, it is highly accurate and allows one to include all the complexity that may affect the seismic wavefields: topography of the free surface and of internal discontinuities, anelasticity, anisotropy and lateral variations of elastic parameters and density. Monteiller et al. (2013), for instance, have shown that topography produces very significant body-to-surface wave conversions, which are predominant in the coda of teleseismic P waves.

Using the DSM code, we compute the tractions and velocities produced by each source at all the SEM grid points located on the edges of the regional SEM mesh and store them to disk. We then read them back at the beginning of each SEM simulation. In the regional mesh, it is not necessary to use a discontinuous (i.e. geometrically non-conforming) mesh and thus a discontinuous Galerkin formulation because material property contrasts are not drastic, therefore, resorting to a continuous Galerkin formulation is sufficient.

Our hybrid method implements a simple one-way coupling approach suggested by Bielak & Christiano (1984), in which absorbing boundary conditions are only implemented to the diffracted wavefield. The total wavefield $\mathbf{u}$ can be written as the sum of the incident wavefield $\mathbf{u}_0$ and the diffracted wavefield $\mathbf{u}_d$. The total wavefield in the regional 3D domain is computed by SEM. On the boundaries, the diffracted wavefield $\mathbf{u}_d$ can be absorbed based on simple paraxial absorbing boundary conditions (Stacey, 1988):

$$(\mathbf{T} - \mathbf{T}_0) \cdot \hat{\mathbf{n}} = \rho\alpha[\hat{\mathbf{n}} \cdot \partial_t(\mathbf{u} - \mathbf{u}_0)]\hat{\mathbf{n}} + \rho\beta[\hat{\mathbf{t}} \cdot \partial_t(\mathbf{u} - \mathbf{u}_0)]\hat{\mathbf{t}} \tag{2.1}$$

where $\mathbf{T}$ denotes the stress tensor, $\hat{\mathbf{n}}$ is the unit outward normal of the boundary, and $\hat{\mathbf{t}}$ is the unit vector tangential to the boundary. ρ, α and β denote the density, compressional wave velocity and shear wave velocity on the boundary, respectively. In order to accurately interface the 1D DSM and 3D SEM simulation on the boundary, the boundary region of the 3D domain is required to smoothly transit from the outer 1D layered background model to the 3D heterogeneous structures. The simple and approximate Stacey absorbing condition is quite effective for our applications. Figure 2.1 shows the DSM/SEM snapshots of the vertical component of total wavefields propagating in a model with the free surface topography of the Pyrenees, between 355 s and 405 s after the origin time of a teleseismic event. This event is located north of the center of the regional mesh, at an epicentral distance of around 30°. The hypocentral depth of the event is 9.75 km. We can clearly see that the incident teleseismic wavefront is curved on the snapshots, while many applications usually assume a planar teleseismic wavefront (Tong et al., 2014a,b).

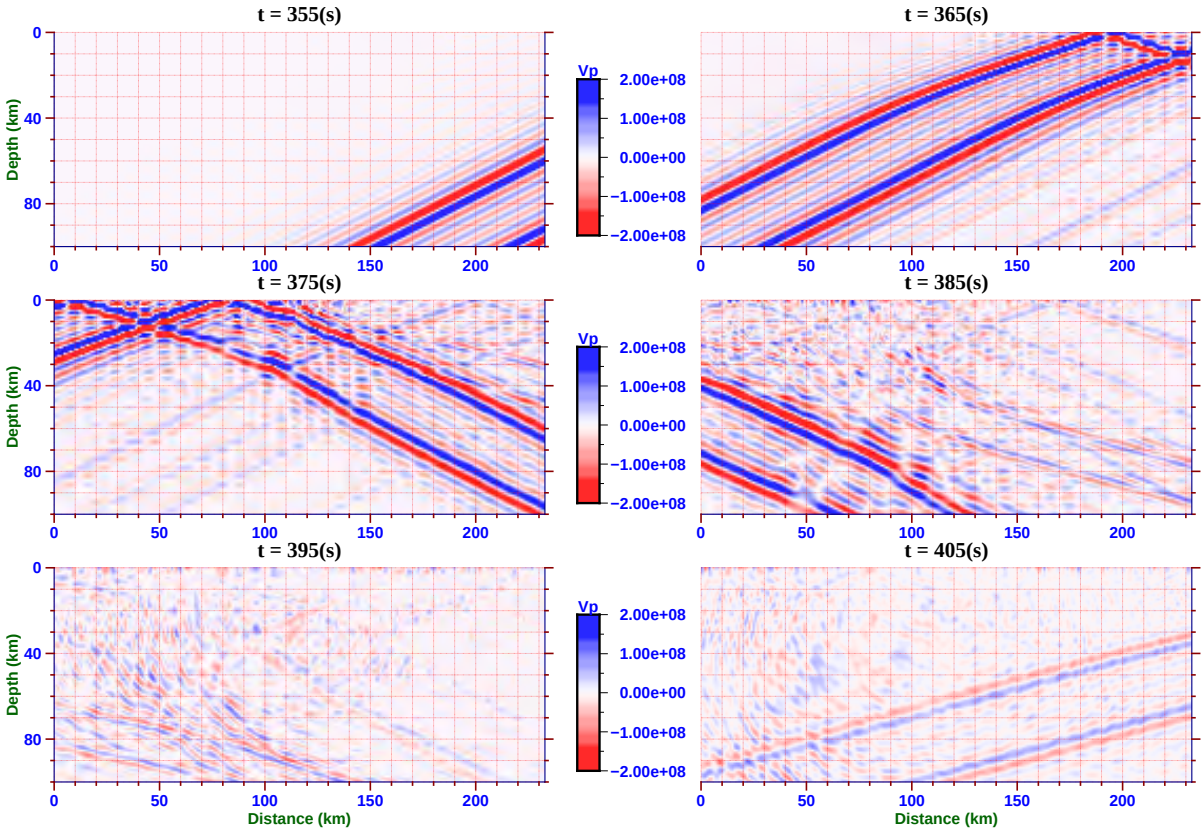


Figure 2.1: The DSM/SEM snapshots of the vertical component of the total wavefield in a model with the free surface topography of the Pyrenees, observed in the vertical section along the ECORS-Pyrenees profile. A teleseismic event is located north of the center of the regional domain, at an epicentral distance of around 30°. The hypocentral depth of the event is 9.75 km.

2.2 Construction of the 3D regional mesh and computational demands

The inversion in the regional domain based on the hybrid method requires a well designed regional 3D mesh, which can capture the complexities of real geological structures.

2.2.1 Overview of SPECFEM3D Cartesian package for regional mesh generation

Our study relies on the open source SPECFEM3D Cartesian package which can handle both forward and adjoint spectral-element simulations of wave propagation on fully unstructured hexahedral meshes inside an arbitrarily shaped regional domains. With this well-developed method, it is possible to introduce the topography of the free surface, or of irregular internal discontinuities for modeling the propagation of seismic waves. Figure 2.2 presents the schematic work flow of the SPECFEM3D package from meshing and partitioning to performing spectral-element simulations and corresponding applications. The SPECFEM3D package generally divides the simulation into two separate parts: first, constructing the hexahedral mesh and node partitioning for target computational domain; second, resolving the wave equation. This separation avoids recomputing the mesh, which can be time consuming when multiple simulations are run inside the same grid, as is the case in iterative waveform inversion (Peter et al., 2011). In this subsection, we focus on the issues of designing meshes adapted to waveform inversion for deep structures beneath a mountain range.

A significant superiority of the SEM is that it can treat very distorted mesh elements in a flexible way (Oliveira & Seriani, 2011). In addition, it is possible to introduce transition layers consisting of conforming unstructured mesh doubling bricks to efficiently accommodate mesh size variations if necessary (Komatitsch & Tromp, 2002; Komatitsch et al., 2004). This mesh doubling allows us to refine the size of the elements close to the surface, a key point to accommodate the sharp and strong variations of elastic properties in the shallow crust and describe the free surface topography accurately. In the unstructured mesh, the corners of the elements can be shared with any number of adjacent elements, relying on the topology of the mesh (Komatitsch et al., 2008). As an open source package developed by many researchers, the new version of SPECFEM3D benefits from recent advances in flexible meshing technique, load balancing, and software optimization (Peter et al., 2011). Meshing may be completed by using an external mesh generation toolkit such as CUBIT (Blacker et al., 1994), or the internal mesher *xmeshfem3D* inside the SPECFEM3D package. Load balanced parallel simulations in SPECFEM3D are obtained thanks to the graph partitioning software package SCOTCH (Chevalier & Pellegrini, 2008). This new package can be used for applications in a wide range of scales, such as the global Earth, seismic exploration, ocean acoustics, or medical tomography.

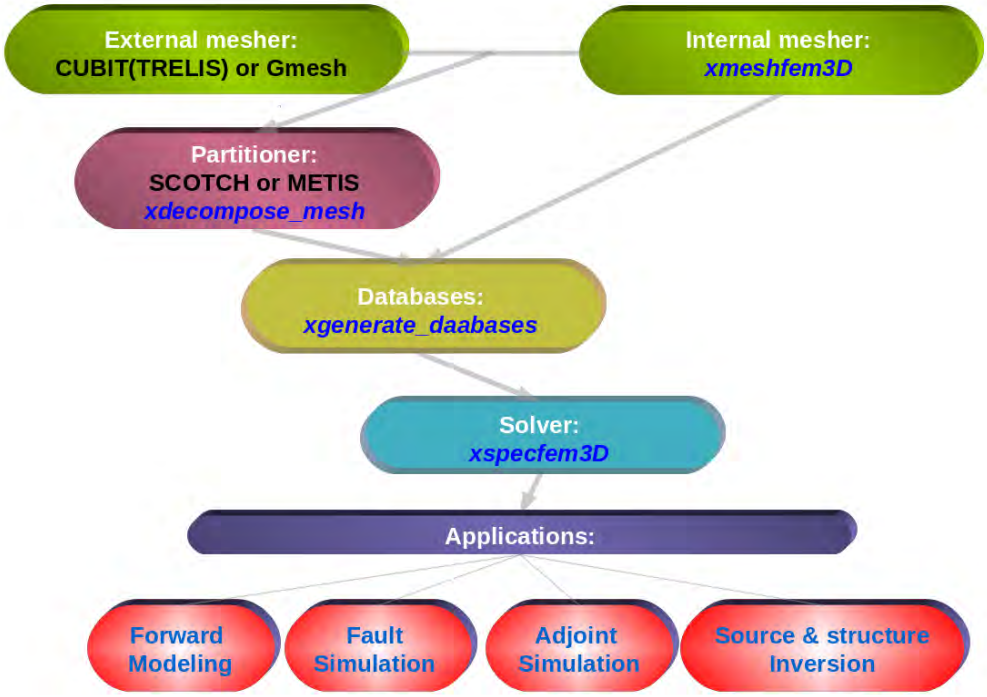


Figure 2.2: Work flow of a spectral-element simulation with SPECFEM3D. After the mesh construction and partition, the executable routine *xgenerate_ databases* creates the GLL mesh grids and assigns specific model parameters to each grid point. Once the databases are generated, we can compile the solver *xspecfem3D* and run the main spectral-element simulation.

2.2.2 The consideration in regional SEM mesh generation

The first step in regional SEM mesh generation is that the target region Ω is subdivided into a set of non-overlapping hexahedral elements: $\Omega = \bigcup_{e=1}^{N} \Omega_e$, where N is the total amount of elements and Ω_e is a hexahedral element. This discretization constructs a conforming mesh, which means that all the elements match on their edges, and that the mesh cannot be discontinuous. Hexahedral meshing is attractive because it benefits from reduced errors and a smaller number of elements compared with other types of meshing. Each hexahedral element Ω_e can be mapped onto an unit reference cube $\Lambda = [-1, 1]^3$ based on an invertible mapping (Komatitsch & Tromp, 2002). This mapping is usually not analytical, especially for complex meshes describing geologic structures. It needs to be dealt with approximately. A set of specific polynomial functions which are called 'shape functions' and their corresponding so-called 'control nodes' form the basis of this approximation. The Cartesian coordinates of every point $\mathbf{x} = (x, y, z)$ within the physical hexahedral element Ω_e can be defined by n_a control nodes $\mathbf{x}_a$ and their corresponding shape functions N_a by:

$$\mathbf{x}(\boldsymbol{\xi}) = \sum_{a=1}^{n_a} N_a(\boldsymbol{\xi}) \mathbf{x}_a \tag{2.2}$$

n_a can take the values 8 or 27. The Cartesian point $\mathbf{x}$ is related to position vector $\boldsymbol{\xi} = (\xi, \eta, \zeta)$ within the reference unit Λ. The eight corners of each hexahedral element are used as control

nodes for all cases. For a physical element Ω_e with straight edges, eight corners are sufficient to precisely represent its geometry, in which case the n_a takes the value 8. The corresponding shape functions are the triple products of linear interpolation functions:

$$\begin{aligned} N_a(\boldsymbol{\xi}) = N_{n_\xi n_\eta n_\zeta}(\xi,\eta,\zeta) = h_{n_\xi}(\xi)\cdot h_{n_\eta}(\eta)\cdot h_{n_\zeta}(\zeta) \\ \Big(n_\xi = \pm 1, n_\eta = \pm 1, n_\zeta = \pm 1\Big) \end{aligned} \tag{2.3}$$

Here $h_{-1}(\xi)$ and $h_{+1}(\xi)$ represent Lagrange linear interpolation functions corresponding to reference vertex points $\xi = -1$ and $\xi = +1$:

$$h_{-1}(\xi) = \frac{1}{2}(1-\xi) \;\;,\;\; h_{+1}(\xi) = \frac{1}{2}(1+\xi) \tag{2.4}$$

On the other hand, an element with curved edges is approximately defined by eight corners, plus twelve nodes on the mid-edges, six nodes on the mid-faces and one central node. This means that $n_a = 27$ for this case. The corresponding shape function is a multivariate quadratic function, which are triple products of degree 2 Lagrange interpolation polynomials:

$$\begin{aligned} N_a(\boldsymbol{\xi}) = N_{n_\xi n_\eta n_\zeta}(\xi,\eta,\zeta) = h_{n_\xi}(\xi)\cdot h_{n_\eta}(\eta)\cdot h_{n_\zeta}(\zeta) \\ \Big(n_\xi = 0,\pm 1,\;\; n_\eta = 0,\pm 1,\;\; n_\zeta = 0,\pm 1\Big) \end{aligned} \tag{2.5}$$

The Lagrange polynomials of degree 2 with three reference points $\xi = -1$, $\xi = -0$ and $\xi = +1$ are :

$$h_{-1}(\xi) = \frac{1}{2}\xi(\xi-1) \;\;,\;\; h_0(\xi) = (1-\xi^2),\;\; h_{+1}(\xi) = \frac{1}{2}\xi(\xi+1) \tag{2.6}$$

Given the n_a coordinates of control nodes $\mathbf{x}_a$ and corresponding reference points $\boldsymbol{\xi}_a$, the weighted sum of shape functions $N_a(\boldsymbol{\xi})(a = 1, 2, ..., n_a)$ completely determine the geometry of the element for SEM simulation. The shape function and the mapping $\xi \in \Lambda \mapsto \mathbf{x} \in \Omega_e$ satisfy the conditions:

$$\begin{aligned} N_a(\boldsymbol{\xi}_b) &= \delta_{ab} \\ \mathbf{x}_a = \mathbf{x}(\boldsymbol{\xi}_a) &= \sum_{a=1}^{n_a} N_a(\boldsymbol{\xi}_a)\mathbf{x}_a \end{aligned} \tag{2.7}$$

An infinitesimal volume $dxdydz$ within given physical element Ω_e relates to the corresponding infinitesimal volume $d\xi d\eta d\zeta$ in the reference cube Λ by:

$$dxdydz = J(\xi,\eta,\zeta)d\xi d\eta d\zeta \tag{2.8}$$

where $J(\xi,\eta,\zeta)$ is the Jacobian of the mapping (2.2) given by:

$$J(\xi,\eta,\zeta) = \Big|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\Big| \tag{2.9}$$

The partial derivative matrix $\partial(x,y,z)/\partial(\xi,\eta,\zeta)$ can be obtained by analytically differentiating the explicit expression (2.2). We need to ensure that the mapping (2.2) is unique and

invertible, that is, $\mathbf{x}(\boldsymbol{\xi})$ should be well-defined and the Jacobian $J(\xi, \eta, \zeta)$ should never vanish. As in the finite element method, the geometry of the mesh controls the behavior of the Jacobian. The relative positions of n_a control nodes $\mathbf{x}_a$ in physical space with the corresponding reference points $\boldsymbol{\xi}_a$ in the reference cube will totally determine the quality of the SEM mesh. The inappropriate choice of control nodes for the mesh will lead to elements with poor quality.

Generally speaking, the design of a mesh can not be accomplished fully automatically. The main difficulty is to design a mesh that honors the complexities of the structural model with a reasonable number of elements. During the mesh generation, we need to honor the free surface and the main geological interfaces across which the material properties of geological structures change discontinuously. In fact, the material properties are approximately represented by smooth interpolation polynomials inside each element, meaning that sharp local variations of geological structure such as a seismic interface between two layers can only be imposed on element boundaries. Note that since the interfaces will be approximated by continuous piecewise polynomials defined by shape functions of interconnected elements, large number of elements are required to correctly describe the topography of the free surface and inner discontinuities. Of course, using very small hexahedra elements or geometrically more flexible tetrahedral elements (which is popular in classical finite element methods) can better honor the complexity and correctly capture the rapid changes of material properties. However, certain conditions need to be met to generate a good quality mesh.

Firstly, a proper mesh should be relatively regular. We should sample the seismic wavelengths relatively uniformly throughout the investigated region. Using undersized hexahedra results in prohibitively large amounts of elements. As a result, the mesh will oversample the shorter wavelengths which will result in inefficient computations and big memory storage requirements. In addition, using small elements will impose to use a small time step in the explicit time-marching scheme in SEM due to the Courant-Friedrichs-Lewy (CFL) stability criterion:

$$\Delta t \leq C \cdot \min_{\Omega} \left(\frac{h}{v} \right), \tag{2.10}$$

where h is the distance between adjacent grid points, v is model's compressional (or shear) wave velocity, Ω denotes the total mesh volume and C is the so-called Courant number that depends on the geometry of the mesh and on the spatial dimension of the problem (typically ranges between 0.3 and 0.5, Chaljub et al. (2007)). The CFL criterion requires that the speed at which information travels can not exceed one grid cell per time step.

Secondly, the SEM solves the weak form of the equation of motion. A 3D mesh can be discretized with hexahedra or tetrahedra. However, the integrals appearing in the weak form of the wave equation are numerically evaluated by Gauss-Lobatto-Legendre quadratures in SEM. The quadrature is obtained by a triple tensorization process that will be explained next, which is the main reason why it is more convenient to use hexahedra rather than tetrahedra in 3D simulations.

In SEM, the field quantities such as displacement, stress or model parameters inside each spectral element Ω_e, are described by interpolation polynomials and interpolation points along each of the three spatial directions, compared with the low degree (1 or 2) polynomials $N_a(\boldsymbol{\xi})$ used to describe the shape of the element in (2.2). Typically, SEM uses Lagrange polynomials of degree 4 to 10 as the interpolation of functions. The $n+1$ Lagrange polynomials of degree n are defined by $n+1$ interpolation points $-1 \leq \xi_\alpha \leq 1 (\alpha = 0, 1, ..., n)$ as:

$$\begin{aligned} l_\alpha(\xi) = \frac{(\xi-\xi_0)\cdots(\xi-\xi_{\alpha-1})(\xi-\xi_\alpha)(\xi-\xi_{\alpha+1})\cdots(\xi-\xi_n)}{(\xi_\alpha-\xi_0)\cdots(\xi_\alpha-\xi_{\alpha-1})(\xi_\alpha-\xi_{\alpha+1})\cdots(\xi_\alpha-\xi_n)} \\ l_\alpha(\xi_\beta) = \delta_{\alpha\beta} (\alpha, \beta = 0, 1, ..., n) \end{aligned} \tag{2.11}$$

where l_α is the Lagrange polynomial for the single variable defined along one dimension of the reference domain. The Lagrange polynomials for the 3D reference cube Λ are obtained by tensorization of these 1D functions. Accordingly, interpolation points in Λ are obtained by tensorization of the 1D sets of interpolation points $-1 \leq \xi_\alpha \leq 1 (\alpha = 0, 1, ..., n)$ so that each element Ω_e contains $(n+1)^3$ interpolation points, if the numbers of interpolation points along three directions are the same. Any function f inside an element can now be interpolated with these Lagrange polynomials defined in reference cube Λ:

$$f(\mathbf{x}(\xi, \eta, \zeta)) \approx \sum_{\alpha,\beta,\gamma=0}^{n_\alpha, n_\beta, n_\gamma} f_{\alpha\beta\gamma} l_\alpha(\xi) l_\beta(\eta) l_\gamma(\zeta) \tag{2.12}$$

where $f_{\alpha\beta\gamma} = f(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma))$ denotes the value of function f at the interpolation point $\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma)$. In traditional finite element method, the same low order polynomials (order 1 or 2) are used to define the shape functions and the functions of field quantities defined inside each element, such that the accuracy of traditional FEM is mainly controlled by the characteristic size of the elements. In SEM, high order Lagrange polynomials are used to represent the functions of field quantity inside each element. The polynomial degree used to interpolate the wavefield inside the SEM element is higher than that defining the shape function, the mapping (2.2) is said to be sub-parametric (Chaljub et al., 2007). We can choose the polynomial degree n to interpolate functions in an element as an additional parameter to control the accuracy of the SEM. Because there is no definite rule to estimate the accuracy of SEM simulation for elastic media, some heuristic rules for choosing the proper polynomial degree used in elastic SEM simulation are explored (Seriani & Oliveira, 2008). These rules have been developed over the years based on many numerical tests for which a quasi-analytical reference solution was known. Here we refer to the main conclusions of Seriani & Priolo (1994) made in the acoustic case. It has been shown that these conclusions can extend reasonably well to the elastic case. In our simulations, we use a polynomial degree $n_\alpha = n_\beta = n_\gamma = 4$. This choice is an excellent compromise between accuracy and time integration stability.

Since the weak form of the equation only involves gradients operation of the field quantities, the gradient of a function $f(\mathbf{x}(\xi, \eta, \zeta))$ evaluated at the interpolation point $\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma)$ may be

written in the form:

$$\nabla f(\mathbf{x}(\xi,\eta,\zeta)) \approx \sum_{i=1}^{3} \hat{\mathbf{x}}_i \sum_{\alpha=0}^{n_\alpha} \sum_{\beta=0}^{n_\beta} \sum_{\gamma=0}^{n_\gamma} f_{\alpha\beta\gamma} \Big[l'_\alpha(\xi) l_\beta(\eta) l_\gamma(\zeta) \partial_i \xi + l_\alpha(\xi) l'_\beta(\eta) l_\gamma(\zeta) \partial_i \eta + l_\alpha(\xi) l_\beta(\eta) l'_\gamma(\zeta) \partial_i \gamma \Big] \tag{2.13}$$

where the prime denotes the derivative of a single variable Lagrange polynomial. We use the index notation $\partial_i = \partial_{x_i}$, and $x_1 = x, x_2 = y, x_3 = z$. Equation (2.13) contains the elements of matrix $\partial\boldsymbol{\xi}/\partial\boldsymbol{x}$, which are obtained by inverting the partial derivative matrix $\partial(x,y,z)/\partial(\xi,\eta,\zeta)$.

In SEM, the interpolation points required in the definition of Lagrange polynomials are chosen to be the $n+1$ Gauss-Lobatto-Legendre (GLL) points for the numerical integration required in the weak form of the wave equation. The reason for this collocation is that the resulting mass matrix is exactly diagonal when using the Gauss-Lobatto-Legendre integration rule (Komatitsch & Tromp, 2002). This diagonal mass matrix leads to fully explicit time-marching schemes, which is a key property of the SEM for wave propagation problems.

In addition to the above considerations, there is another important issue for the mesh generation. When we use the elements with the same size throughout the discretized regional domain, wavelength oversampling occurs because the seismic velocities usually increase with depth. A similar problem arises when investigating the response of sedimentary basins (Komatitsch et al., 2004), because there is typically at least a factor of 2 increase of seismic velocities in the bedrock compared to that of sediments. To overcome this issue and keep the geometrical conformity of the mesh, unstructured meshes can be used. For more details see Komatitsch et al. (2008) which shows examples of conforming unstructured mesh.

To summarize, the mesh generation is the most critical and time-consuming task before running an accurate simulation in many cases of practical interest, even if the mesh generation has received much attention and significantly evolved in recent years (Peter et al., 2011). This task is still challenging because the use of hexahedra mesh elements is not always convenient to capture geological structures. For instance, when we mesh regions with strong contrasts in material properties on both sides, it can be challenging to define elements whose edges follow the interface and whose shapes are less distorted. A compromise must be found by designing such kind of elements (Komatitsch et al., 2004). Consequently, SEM mesh design is always a difficult problem that requires much experience and trial-and-error. However, it can also be advantageous to pay less attention to this problem. In practice, the precise location and the sharpness of structural discontinuities of the Earth's interior still remain a topic of much uncertainty because our data usually have limited coverage and spatial resolution. Imposing sharp discontinuities in SEM mesh can thus result in strong interface phases that are not observed or not as sharp as predicted. In most cases, it is sufficient to design a mesh that is as regular and simple as possible and yet able to describe a large panel of realistic geological models.

2.2.3 Designing regional SEM meshes for the Pyrenees

2.2.3.1 Basic parameter to evaluate the quality of SEM mesh

In addition to various external meshers, the new version of SPECFEM3D includes a more flexible internal mesher *xmeshfem3D* that can be used for generating layered meshes with complex topography and inner interfaces based on pure Cartesian coordinate system or a spherical layered 'chunk' of the Earth but without topography. The mesh is a 'chunk' of the Earth, defined as a portion of the so-called 'cubed sphere' that is used to mesh part of the spherical earth based on hexahedra (Chaljub et al., 2003). We use the conventions given in Chevrot et al. (2012) to define an invertible mapping between the geographical coordinates and the local cubed sphere coordinates. The chunk is defined by the position of the center of its upper surface and two reference axes along the E-W and N-S directions. In our study, we developed an internal mesher to generate a chunk of the earth with the real topography on the free surface. With this modified internal mesher, we can include the surface topography of the Pyrenees region. Since we want to model the propagation of seismic waves at short period, we will need a fine mesh. There are two parameters that reflect the quality of the mesh: 1. the number of grid points per wavelength, which determines how the wavefield is sampled; 2. the CFL stability criterion which determines the time step of the explicit time integration scheme in SEM necessary for a stable simulation (Komatitsch & Tromp, 2002).

The number of grid points per wavelength determines the shortest period at which the simulations are accurate. We use another rule of thumb from the study of Seriani & Priolo (1994) which shows that the SEM needs roughly 4.5 grid points per minimum wavelength of a local element to be accurate. The shortest period should be chosen such that:

$$N_e = p_{s_e} \cdot \min_{\Omega_e} \left(\frac{v}{\bar{h}_e} \right) \approx N_{emp} \tag{2.14}$$

where N_e is the number of grid points per minimum wavelength in element Ω_e, p_{s_e} is the shortest period that can be resolved in Ω_e, v is model's compressional (or shear) wave velocity, $\bar{h}_e$ denotes the average distance between GLL points within the element and N_{emp} is the empirical number of grid points per minimum wavelength. N_{emp} is fixed value at 5 in SPECFEM3D package, which is slightly larger than the rough estimation value 4.5 from Seriani & Priolo (1994). By using the approximate expression (2.14), the shortest period can be resolved in element Ω_e is:

$$p_{s_e} \approx N_{emp} \cdot \max_{\Omega_e} \left(\frac{\bar{h}_e}{v} \right) \tag{2.15}$$

Noticing that the shortest period defined in (2.15) is just an estimation and there is no such sharp cut-off period for validation of SEM simulation. The synthetic seismograms just become more and more inaccurate for periods shorter than p_{s_e}. In any case, the shortest period that can be resolved in the whole mesh will determine the resolution of the SEM computations.

The second stability criterion fixes the upper limit for the time step Δt used in the simulation. Note that the distance h in (2.10) depends on the size and geometry of the mesh elements as well as the number of GLL interpolation points specified in the SPECFEM3D routines. The SPECFEM3D internal mesher *xmeshfem3D* tries to evaluate the value of Δt for an empirically chosen Courant number $C \backsim 0.5$. Both quantities can be checked in the *'output_ mesher.txt'* file generated after running the internal mesher as a verification before launching the simulation. Note that the Newmark time scheme for SEM solver uses the same time step Δt globally so that the simulation becomes more expensive for small mesh elements in high wave speed regions. The SPECFEM3D routine can also output the resolved minimum period and a stability condition parameter defined as $\Delta t \cdot \max_{\Omega_e}(\frac{v}{h})$ for each element in the VTK format which can be visualized with Paraview. By visually inspecting these two quantities, we can check and adjust our mesh easily.

The mesh generation and the modeling of seismic wave propagation in the Pyrenees region are challenging mainly for two reasons: First, the Pyrenees have strong topographic variations of the free surface and complex subsurface geological architecture. Second, the Pyrenees is surrounded by two foreland basins, the Ebro basin to the south and the Aquitanian basin to the north. Several stations used in our FWI located inside these two basins are shown in Figure 2.3. Due to the difficulties in determining the precise structure of the sedimentary basins and to mesh the large impedance contrasts between bedrock and sediments, we will ignore the modeling of sedimentary basins for the first SEM mesh generation and simulation for the time being and leave this to future work.

2.2.3.2 Topography tapering for the mesh

The surface topography of the Pyrenees is taken from the Shuttle Radar Topographic Mission (SRTM) data. The SRTM is an international research effort to obtain digital elevation models on a near-global scale from latitude 56°S to 60°N. It is an international project lead by the U.S. National Geospatial Intelligence Agency (NGA) and the U.S. National Aeronautics and Space Administration (NASA). SRTM maps cover approximately 80 percent of the Earth's surface. The SRTM can offer maps with different resolution. For example, SRTM1 maps have the highest spatial resolution of 1 arc-second (approximately 30 meters) for the USA, while SRTM30_PLUS maps have a resolution of 30 arc-second (approximately 900 meters) resolution for both global topography and bathymetry worldwide.

Since we will invert the data collected by two transects separately, we will generate two distinct regional meshes for the inversions using the internal mesher *xmeshfem3D*. The size of each regional domain is chosen such as to contain all the selected stations. In order to reduce the computational cost, we rotate the chunk so that it approximately follows the strike of the transect. In addition, since we will use a cosine taper along the edges of the mesh to suppress

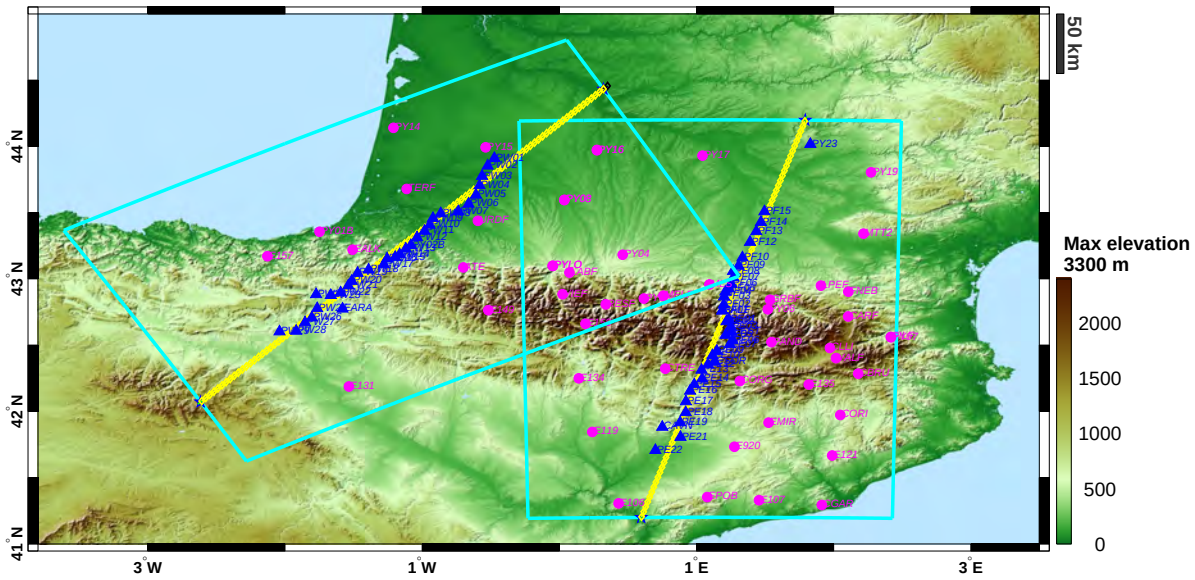


Figure 2.3: The map view of the limits of the regional domains for the two temporary transects. The blue triangles show the temporary stations of dense transects, and pink circles show the adjacent permanent stations and temporary PYROPE/IBERARRAY stations. Projections of the boundaries of the regional domains on the surface are shown with cyan lines. Yellow lines show the projection of the vertical sections along the two transects. The topography is extracted from Shuttle Radar Topographic Mission (SRTM) data.

structural discontinuity between the regional model inside the chunk and the spherically symmetric 1D background model outside, we should slightly extend the size of the regional domain to include these taper elements (Monteiller et al., 2013).

Table 2.4 shows the setting parameters for the SPECFEM3D internal mesher for the two simulation domains in the Pyrenees. Both chunks have a size of 2° by 2° along the horizontal X and Y directions and 150 km along the vertical Z direction. The chunk for the western transect is rotated by 39.63 degrees, but no rotation is applied to the chunk of the central transect. For the moment, we only use regular hexahedral elements, without introducing irregular unstructured layers for mesh doubling. The total number of elements for each mesh is $90 \times 135 \times 60 = 729000$. The mesher automatically determines the appropriate number of elements inside each layer based on a reference 1D model, here the AK135 model. Both chunks are discretized with hexahedral elements with size from $2.413\,km \times 2.413\,km \times 2.5\,km$ in the bottom layer to $2.471\,km \times 2.471\,km \times 2.5\,km$ in the top layer. Since we use polynomial degrees of order 4 along three reference directions (that is, $(n_\alpha + 1) \times (n_\beta + 1) \times (n_\gamma + 1) = 125$ GLL points in each hexahedral element) in the SEM meshes, which yields a total of 50 207 162 grid points and 150 621 486 degrees of freedom.

Figure 2.3 shows the boundaries of the two regional domains and the stations of the two PYROPE transects. The topography is extracted and interpolated from SRTM30_PLUS map because part of the simulation domain comprises the ocean (for both regional domains as can

Inversion domain	Central transect	Western transect
X range of chunk(°)	2.0	2.0
Y range of chunk(°)	3.0	3.0
Longitude of chunk center	1.1	-1.1444
Latitude of chunk center	42.7	43.219
Rotation angle of chunk(°)	0.0	39.63
Depth of chunk(km)	150.0	150.0
Number of elements along x	90	90
Number of elements along y	135	135
Number of elements along z	60	60
Reference model	AK135	AK135

Table 2.4: The setting parameters of the SPECFEM3D internal mesher for the two regional domains.

be seen in Figure 2.3). The typical distance between two grid points inside our SEM elements of the surface is from 393 m to 1012 m, and thus the resolution of SRTM30_PLUS is sufficient for our applications.

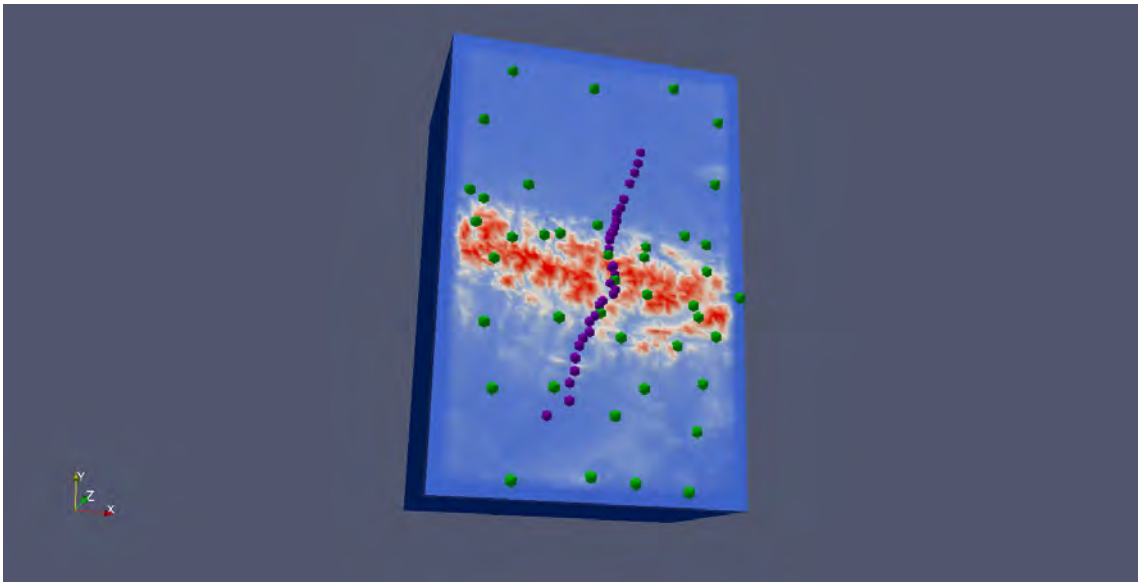
The SEM meshes with surface topography and bathymetry are shown in Figure 2.5. The temporary transect stations and adjacent permanent and temporary PYROPE/IBERARRAY stations used in FWI are shown with violet and green cones, respectively. The rotated chunk for the western transect comprises a larger area of ocean. The effect of the oceanic water layer is not incorporated in the regional model since there is no station located offshore and since we mainly consider the modeling of direct P waves recorded by stations on land.

The topography is included by the internal mesher as a small variation with respect to a spherical reference surface. The mesher uses a cosine taper function along the depth direction to determine the varying height of the elements from the free surface down to the first inner interface, in our case the spherical local Conrad discontinuity. The topography modulations with respect to a spherical reference free surface are distributed to each GLL grid point at a given radius. We illustrate this with the mesh for the central transect.

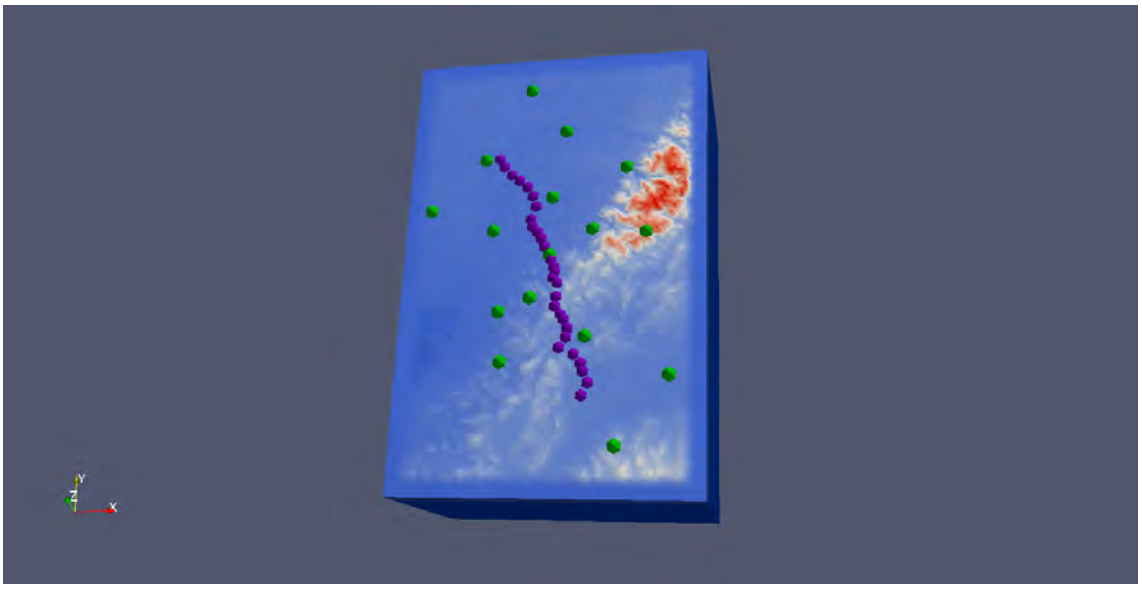
Figure 2.6a shows the surface view of the SEM mesh for the central transect. The color scale indicates the local Z coordinates. The mesh contains 90×135 elements on the surface layer. The thick black contour shows the part of the mesh used for illustrating the depth tapering of topography in Figure 2.6b. The blue star located at the center of the thick black contour shows the surface projection of the elements presented in Figure 2.6b.

Figure 2.6b shows the tapered topography at different depth levels along the depth direction. The hexahedra with different colors are elements with tapered topography surface at their upper and lower faces. The bottom deep blue mesh shows the spherical local Conrad interface, while the top deep red mesh shows the topography of the free surface. There are 8 layers of SEM elements that are used to fill the space from the local Conrad interface to the free surface.

In order to get an intuitive understanding for the tapering of the topography, Figure 2.7

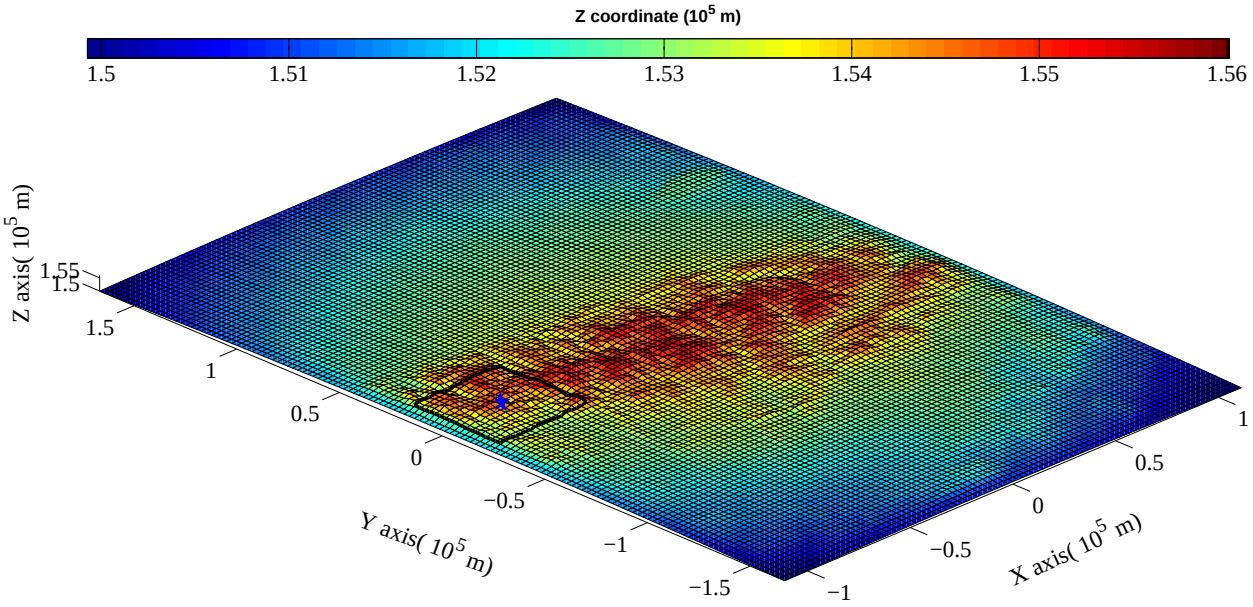


(a) Simulation domain for the central transect

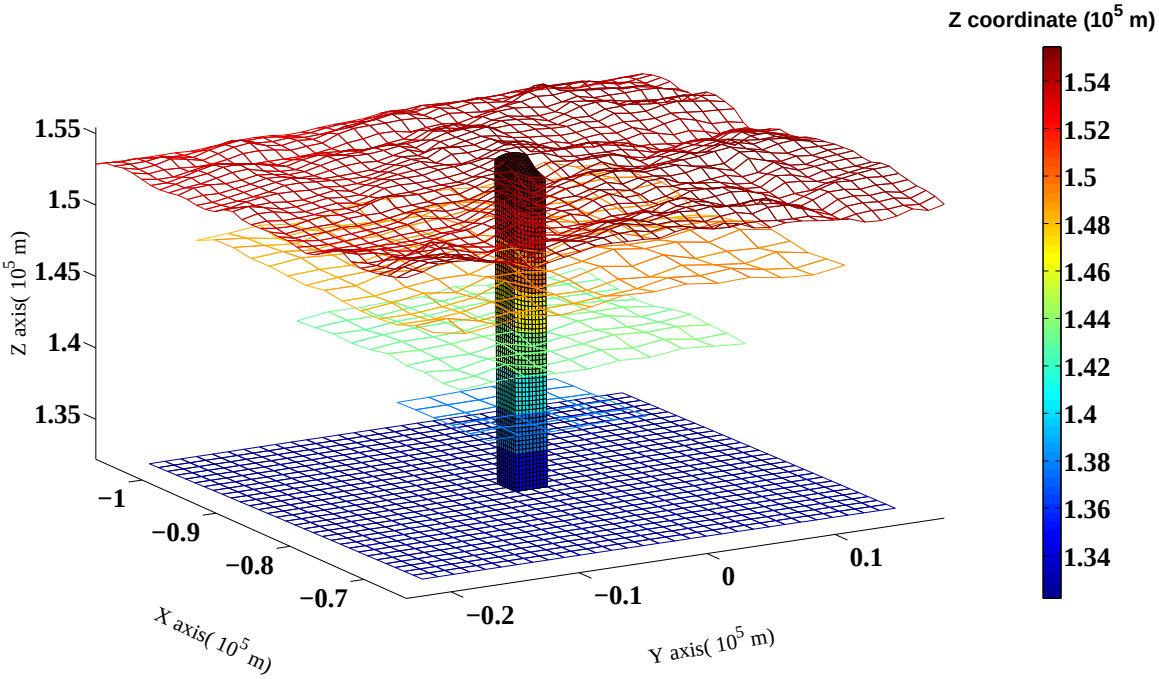


(b) Rotated simulation domain for the western transect

Figure 2.5: (a) Simulation domain for the central transect with topography of the free surface. The temporary transect stations and adjacent permanent and temporary PYROPE/IBERARRAY station used in FWI are shown with violet and green cones, respectively. The color scale ranging from deep blue to deep red indicates the elevation of free surface and bathymetry. (b) Rotated simulation domain for the western transect with topography of the free surface.



(a) Surface view of the SEM mesh for the central transect.



(b) Tapered topography for different elements along the depth direction.

Figure 2.6: (a) Surface view of the spectral element mesh for the central transect. The color scale indicates local Z coordinates of the surface mesh. The mesh contains 90×135 surface elements. The thick black contour near the Y axis shows a small part of the mesh used to illustrate the topography depth tapering of SEM elements in (b). The blue star located at the center of the thick black contour shows the surface location of representative elements in (b). (b) Tapered topography for elements along the depth direction. The topography variation with respect to a spherical reference free surface is distributed to each GLL grid in SEM element by a depth taper function. The hollow meshes show the tapered topography surface for elements in each layer. The different colored hexahedra show the elements with tapered topography surface at their upper and lower faces. The bottom deep blue mesh shows the spherical local Conrad interface, the top deep red mesh shows the topography of the free surface. Eight layers of SEM elements are used from the local Conrad interface to the free surface.

shows the 8 elements from the layer just above the spherical local Conrad interface to the layer just beneath the free surface, which are shown by colored hexahedra in Figure 2.6b. The tapered topography forms the top and bottom faces of corresponding elements from 2.7a to 2.7h.

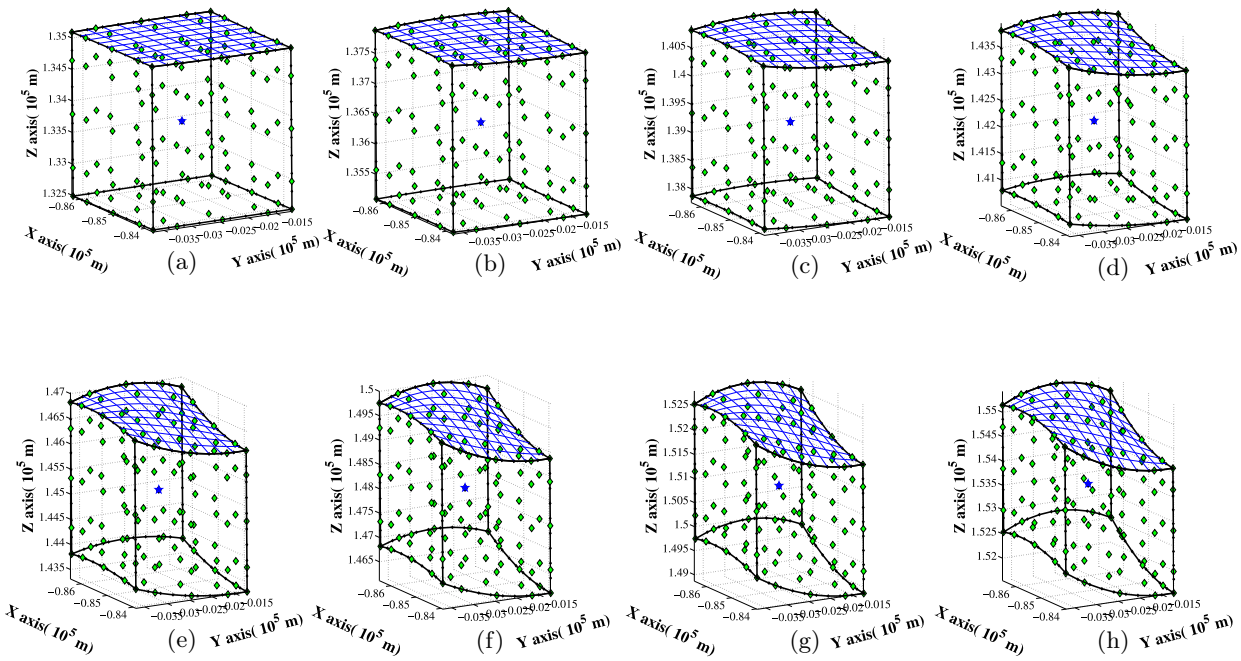
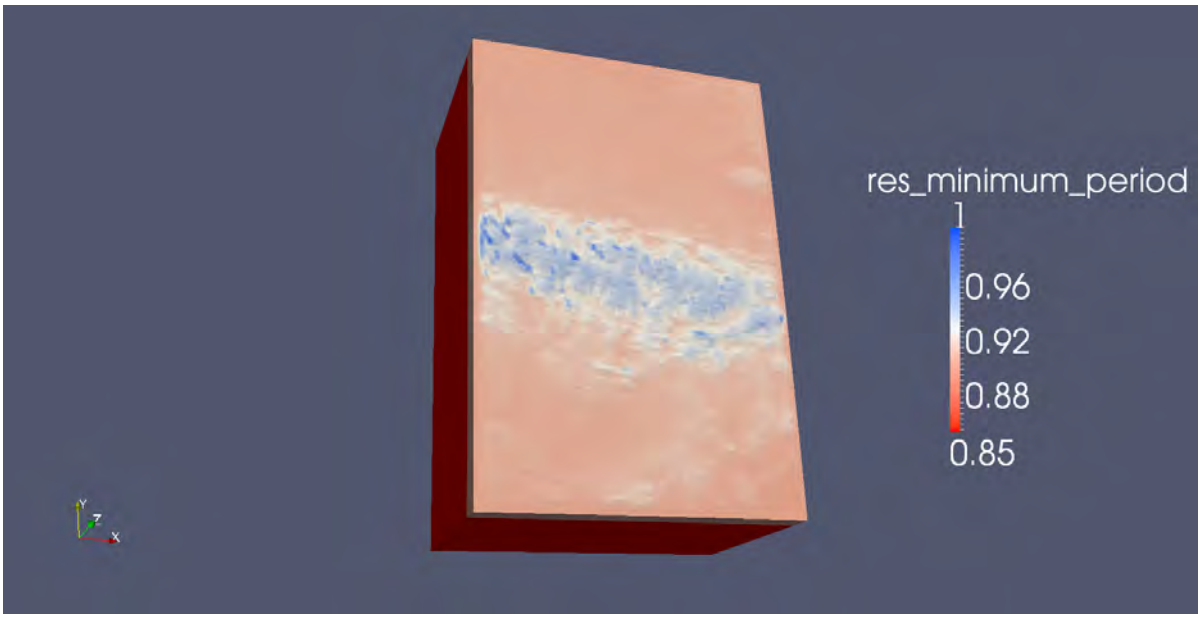
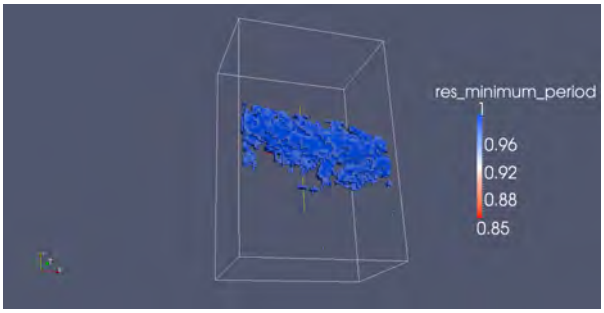


Figure 2.7: Eight hexahedral elements from the layer just above the spherical local Conrad interface to the layer just beneath the free surface. The tapered topography for each layer forms the top and bottom face of corresponding element from figure (a) to figure (h). The green diamonds inside each element represent the corresponding GLL points.

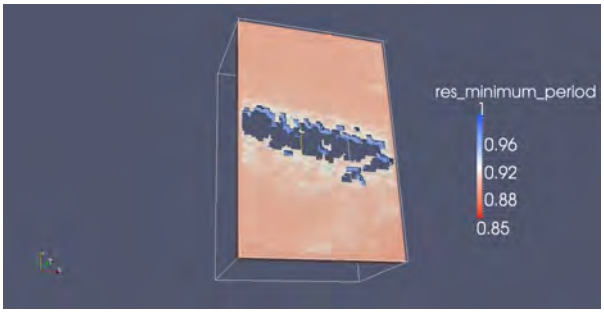
Figure 2.8 illustrates the shortest period p_{s_e} that can be resolved by each element according to definition (2.15) in the final mesh with topography. The approximate shortest period p_{s_e} depends on the geometry and size of the elements as well as the velocity model assigned to the grid. In our mesh design, we use a smoothed ak135 model to assign the velocity at the GLL grid points in the regional SEM mesh. Figure 2.8a shows the shortest period p_{s_e} that can be resolved in the SEM mesh shown in Figure 2.5a. The minimum and maximum value of the shortest resolved period for the whole mesh are 0.69 s and 1.12 s, respectively. Figure 2.8b and 2.8c show the elements with the shortest resolved period from 0.975 s to max value 1.12 s and from 0.875 s to 0.975 s in our mesh, respectively. The elements discretizing the region with higher topographic relief show the larger values of shortest resolved period, such as the elements located in the mountain range that can be seen in Figure 2.8b. This value is sufficient for our current waveform inversion, which explores the frequency content of the seismic wavefield that is generally no less than 2 s.



(a) Shortest period can be resolved in the SEM mesh for the central transect.



(b) Elements with the shortest resolved period from 0.975 s to max value.



(c) Elements with the shortest resolved period from 0.875 to 0.975 s.

Figure 2.8: (a) The shortest period p_{s_e} that can be resolved in the SEM mesh shown in Figure 2.5a. The color scale indicates the value of p_{s_e} from minimum value 0.69 to maximum value 1.12 s.(b) Elements with the shortest resolved period from 0.975 s to max value. (c) Elements with the shortest resolved period from 0.875 to 0.975 s.

2.2.4 Parallel implementation and partitioning

When the mesh is successfully generated, the next step is to solve the wave equation. The significant feature of SEM is that it is very well adapted to parallel machine, which has a distributed memory architecture. The SEM uses a fully explicit time-marching scheme. This scheme makes that SEM only needs to perform small local matrix vector products inside each element, such that the processors spend most of their work time on performing these calculations, and only a small fraction of time on exchanging data, which means that the algorithm is insensitive to the speed of communication (Chaljub et al., 2007). In order to optimally exploit the distributed systems, SEM is implemented on the parallel computer by using the library of pure Message Passing Interface (MPI). For the parallel implementation of SEM in large scale applications, the load balancing is an important computational issue. The mesh and corresponding computational load need to be evenly distributed on a large number of processors (Martin et al., 2008). The powerful external partitioner SCOTCH (Chevalier & Pellegrini, 2008) is available for this purpose. With SCOTCH, the mesh is decomposed into as many slices as the number of processors for which we apply on the parallel machine. The computations are performed locally by each processor on the SEM elements that belong to its mesh slice. Once the local computations are completed, the processors must communicate to exchange the information required by neighboring slices. In the time-marching scheme, one global communication is required at each time step to sum the internal forces computed at the common faces, edges, and corners shared by the mesh slices that belong to distinct processors (Chaljub et al., 2003). The MPI communications are implemented based on a fixed communication topology produced once and for all by the mesh partitioner. This communication topology contains the target addresses of messages that need to be exchanged between all the slices. The topology is usually fixed because there is no need for a dynamic remeshing procedure for wave propagation problems (Tromp et al., 2008).

Figure 2.9 presents an example of partitioning and load balancing for the mesh of the central transect. We illustrate the decomposition process of the mesh onto eight processors by using the SCOTCH partitioner, where each mesh slice is shown by different colors. The total number of SEM elements in the chunk is 729000 so that each partition has about 90000 elements after decomposition. SCOTCH can load balance the complex mesh by minimizing the number of edge cuts which appears when two contiguous elements are distributed to different processors. This minimization procedure aims at reducing the amount of MPI communications between adjacent processors.

2.2.5 The computational requirements for forward modeling

Our forward hybrid method is implemented on the supercomputer 'Curie', owned by Grand Équipement National de Calcul Intensif (GENCI) and operated into the Très Grand Centre

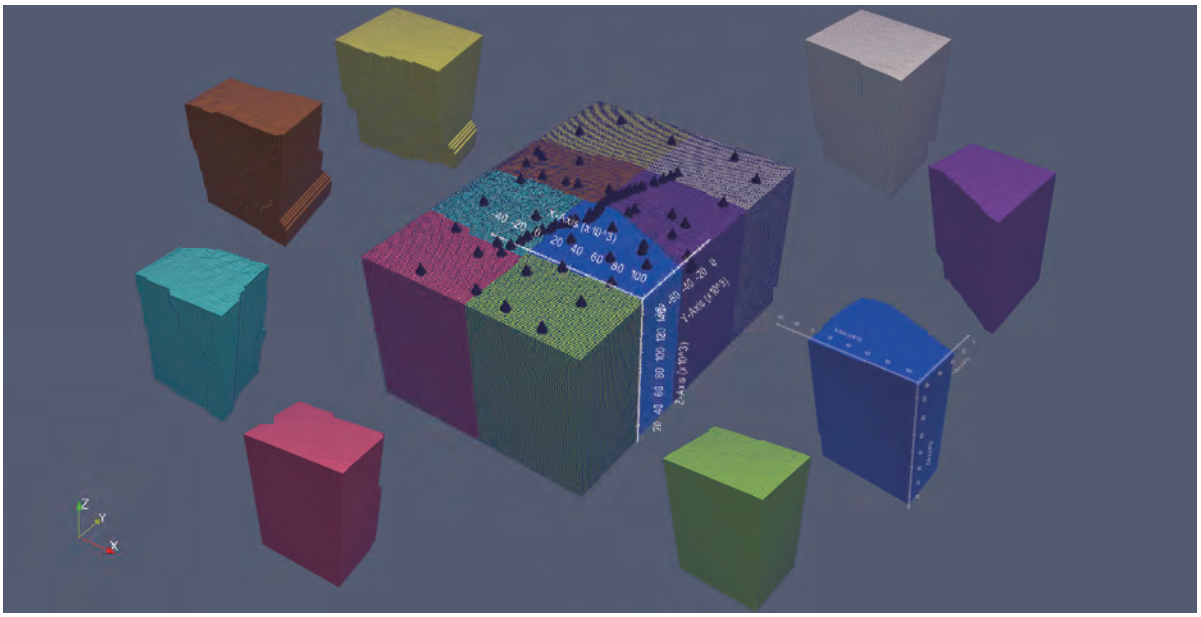


Figure 2.9: The mesh for the central transect is partitioned and load balanced to run on eight processors in parallel. The eight mesh slices are indicated by different colors and have nearly the same number of elements.

de calcul (TGCC) by Commissariat à l'énergie atomique et aux énergies alternatives (CEA), which is the first French petascale supercomputers open to scientists for research purposes (http://www-hpc.cea.fr/). The Curie supercomputer can offer 3 different fractions of x86-64 computing resources for dealing with a variety of scientific problems and can reach an aggregate theoretical peak performance of 1.7 PetaFlops. The three different computing resources offered by Curie are Curie thin nodes, Curie fat nodes and Curie hybrid nodes with different specifications. For our current applications, we submit the parallel task on Curie thin nodes. For each thin node, it contains 2 eight-cores Intel E5-2680 (code-named Sandy Bridge EP) 2.7 GHz processors, 64 GB memory and one local SSD disk. The Curie thin nodes have 10080 eight-core processors, with 80640 cores. These nodes are targeted for MPI parallel coding and can be accessible through the standard queue.

In order to take advantage of the parallel machine architecture, our SEM mesh is partitioned into 512 slices that are distributed on 512 cores with SCOTCH. The computations thus require 2048Gb of distributed memory. Considering our mesh and corresponding CFL stability criterion, we set a time step of $\Delta t = 0.0125s$, which is smaller than the upper limit $\Delta t_{\max} = 0.0256s$ estimated from (2.10). We will use the 2012 May 24 event as an example. For this event, a 85 s long seismogram including the direct P phase and its coda requires 6800 time steps. The duration of a simulation depends on the event azimuth, the distribution of stations and the specific time window required for FWI. Therefore, the number of CPU-hours required to make all the simulations for a teleseismic event is the product of the following parameters: number of cores per simulation, duration of the seismogram and the average CPU work time for one time

step Δt per core. For our simulation it takes about 34 CPU-hours to compute the seismograms for this event. This means that just 4 minutes are required for computing the seismograms for all the stations when we use 512 Curie thin nodes, which is an acceptable computational cost for iterative waveform inversion.

2.3 The source excitation and wavelet estimation

2.3.1 The effects of source excitation and propagation on teleseismic waveforms

In FWI, it is important to model the broad-band waveforms as precisely as possible. To do so, we need to know the broad-band source time functions for each teleseismic event.

We consider teleseismic P waves recorded at epicentral distances between 30 and 80 degrees in order to avoid the complexities coming from transition zone triplications, caused by the strong velocity gradient in that part of the mantle, and also ensure that core reflected PcP phases are well separated from the primary P and depth phases (pP and sP). We will focus on distant earthquakes with magnitudes larger than 6.0 to get signals with sufficiently high S/N ratio. While large earthquakes usually have complex source time functions and complex slip distribution on finite faults, the simple point source approximation in the spatial domain is usually suitable for characterizing the source wavelet since we consider a very narrow azimuthal band from the source location. Shallow events, which are far more abundant than deep events have surface reflected phases (pP and sP) that complicate the determination of the source excitation. Teleseismic waves are affected by the structures along their propagation path from the source to the regional domain. Since the different paths for regional stations are very close from each other, these effects can be appropriately absorbed and described by an average incident source wavelet. Thus, an important step toward systematically exploiting teleseismic waveform data is to develop a method to precisely determine the source wavelet function.

2.3.2 The principle of wavelet estimation

The displacement/velocity produced by a distant source at $\mathbf{x}_s$ and recorded at surface position $\mathbf{x}_r$, is given by:

$$u_i(\mathbf{x_r}, t) = \mathbf{G}(\mathbf{x}_r, t; \mathbf{x}_s, t') * S(\mathbf{x_s}, t') * E_s(\mathbf{x}) * E_g(\mathbf{x}) * E_r(\mathbf{x}_r) * I(\mathbf{x_r}, t) \tag{2.16}$$

In this expression, $\mathbf{G}(\mathbf{x}_r, t; \mathbf{x}_s, t')$ is the Green's function for the global 1D reference model corresponding to an impulsive point source located at $\mathbf{x}_s$. $S(\mathbf{x_s}, t)$, $E_s(\mathbf{x})$, $E_g(\mathbf{x})$, $E_r(\mathbf{x}_r)$ and $I(\mathbf{x_r}, t)$ are source time function, near source 3D structure effects, global wave propagation effects, receiver-side 3D effects and instrumental response, respectively.

Assuming that the instrumental response and Green's function are known, the other four terms in (2.16) still intricately contribute to the observed wavefield. The main objective of

regional scale FWI is to invert the receiver-side 3D structure $E_r(\mathbf{x}_r)$. This requires a good method to estimate and separate the effects of the other three terms.

At teleseismic distances, not only the source time function term $S(\mathbf{x_s}, t)$ but also the near source 3D structure term $E_s(\mathbf{x})$ and the 3D global wave propagation term $E_g(\mathbf{x})$ are similar for teleseismic waveforms recorded by a regional array which is very distant from the source region. We can thus combine these three terms into one, which will be called wavelet function thereafter:

$$W(t) = S(\mathbf{x_s}, t) * E_s(\mathbf{x}) * E_g(\mathbf{x}) \tag{2.17}$$

The wavelet function $W(t)$ ignores the space dependence $\mathbf{x}$ since it is assumed to be the same for all the receivers. Assuming that the instrumental response $I(\mathbf{x_r}, t)$ has been removed from every record, then (2.16) can now be approximated as:

$$u_i(\mathbf{x_r}, t) = \mathbf{G}(\mathbf{x}_r, t) * W(t) * E_r(\mathbf{x}_r) \tag{2.18}$$

Here we again omit the dependence of source coordinates $\mathbf{x}_s$ in the Green's function expression and use the same expression $u_i(\mathbf{x_r}, t)$ for the record with instrumental response removed. Since the objective of FWI is to invert $E_r(\mathbf{x}_r)$ in the regional domain, it is important to isolate the wavelet term $W(t)$. At this stage, the problem can be reformulated as the simultaneous deconvolution of terms $W(t)$ and $E_r(\mathbf{x}_r)$ from the receiver wavefield $u_i(\mathbf{x_r}, t)$, given the Green's functions $\mathbf{G}(\mathbf{x}_r, t)$. Since both $W(t)$ and $E_r(\mathbf{x}_r)$ are unknown, this operation requires to make additional assumptions.

In this study, the regional 3D heterogeneous structure near the receiver-side will be inverted by updating the regional model to fit the differences in observed waveforms recorded by close receivers. In other words, the regional heterogeneous model will explain the waveform variations, while the average waveform will be closely related to the average Green's function and incident wavelet function. Therefore, we can make the approximation that:

$$\begin{aligned} \mathbf{G}(\mathbf{x}_r, t) * E_r(\mathbf{x}_r) &\approx \mathbf{G}(\mathbf{x}_r, t) \\ \text{and } u_i(\mathbf{x_r}, t) &\approx \mathbf{G}(\mathbf{x}_r, t) * W(t) \end{aligned} \tag{2.19}$$

Equation (2.19) temporarily ignores the effects of regional 3D structures and use the Green's function computed in a global 1D model. Similar approximations are usual in regional scale tomography, such as in the classical regional travel time tomography (Aki & Lee, 1976). By using this assumption, we can now estimate the wavelet function from the observed waveforms. In this section, we only use the vertical components of the P waveforms to estimate the wavelet function $W(t)$ because they are the first arrival and they have a relatively high S/N ratio of the vertical component.

The estimated wavelet function $W(t)$ contains all the wave propagation effects which occur outside of the regional domain. For instance, source-side scattering and depth phases will

be absorbed in the source signature. On the other hand, the coherent receiver-side wavefield interactions will also be absorbed in the wavelet function estimation according to (2.19), such as P crustal multiples if crustal structures are simple and close to tabular layers. In any case, since we will also consider the radial components in the inversion, the method developed here will still be able to image crustal discontinuities.

2.3.3 Data preparation and Green's functions computation

The wavelet estimation procedure is illustrated on an event recorded by the dense transect deployed across the central part of the Pyrenees during the PYROPE experiment. We consider digital broad-band recordings of the vertical displacement produced by the Mw 6.1, 10 km deep Norwegian sea earthquake, located about 30 degrees to the north of the transect, sampled at 20 Hz. Waveform data are deconvolved from their instrumental responses and band-pass filtered between 0.03 and 1.05 Hz. The USGS source information of this event is listed in Table 2.10. In the following, we will detail the procedure of data preparation and Green's functions computation.

T0	depth(km)	latitude	longitude	Magnitude	Mrr (N·m)	Mtt	Mpp	Mrt	Mrp	Mtp
2012/ 05/ 24 22:47:46	11.0	72.99°	5.651°	6.2	-1.89 $\times 10^{18}$	0.75 $\times 10^{18}$	1.13 $\times 10^{18}$	0.55 $\times 10^{18}$	-0.18 $\times 10^{18}$	1.22 $\times 10^{18}$

Table 2.10: The CMT information for the 2012 May 24 event, from the USGS

2.3.3.1 Data alignment

The first step of the wavelet estimation is the temporal alignment of all the vertical component seismograms. This alignment relies on accurate estimations of relative time shifts. Here we follow the method introduced by Luo & Schuster (1991). The relative time shift ΔT between two traces $u(\mathbf{x}_1, t)$ and $u(\mathbf{x}_2, t)$ is obtained by searching for the maximum of the cross-correlation function defined by:

$$C(u_i(\mathbf{x}_1, t), u_i(\mathbf{x}_2, t))(\tau) = \frac{1}{2\pi} \int_R w_r(t) u_i(\mathbf{x}_1, t) u_i(\mathbf{x}_2, t+\tau) dt \tag{2.20}$$

where $w_r(t)$ denotes a window function. This implies that ΔT satisfies:

$$\frac{\partial C(u_i(\mathbf{x}_1, t), u_i(\mathbf{x}_2, t))(\tau)}{\partial \tau}|_{\tau=\Delta T} = \frac{1}{2\pi} \int_R w_r(t) u_i(\mathbf{x}_1, t) \dot{u}_i(\mathbf{x}_2, t+\tau) dt = 0 \tag{2.21}$$

where $\dot{u}_i$ denotes the time derivative of record u_i. We align the traces for the 2012 May 24 event by choosing a 60 s long time window starting 15 s before the theoretical P arrival time. Figure 2.11 shows the location of the receivers that recorded this event in the investigated

regional domain. The red triangles show the temporary stations of the dense transect and the green circles show the permanent and temporary broad-band stations of the IBERARRAY and PYROPE experiments.

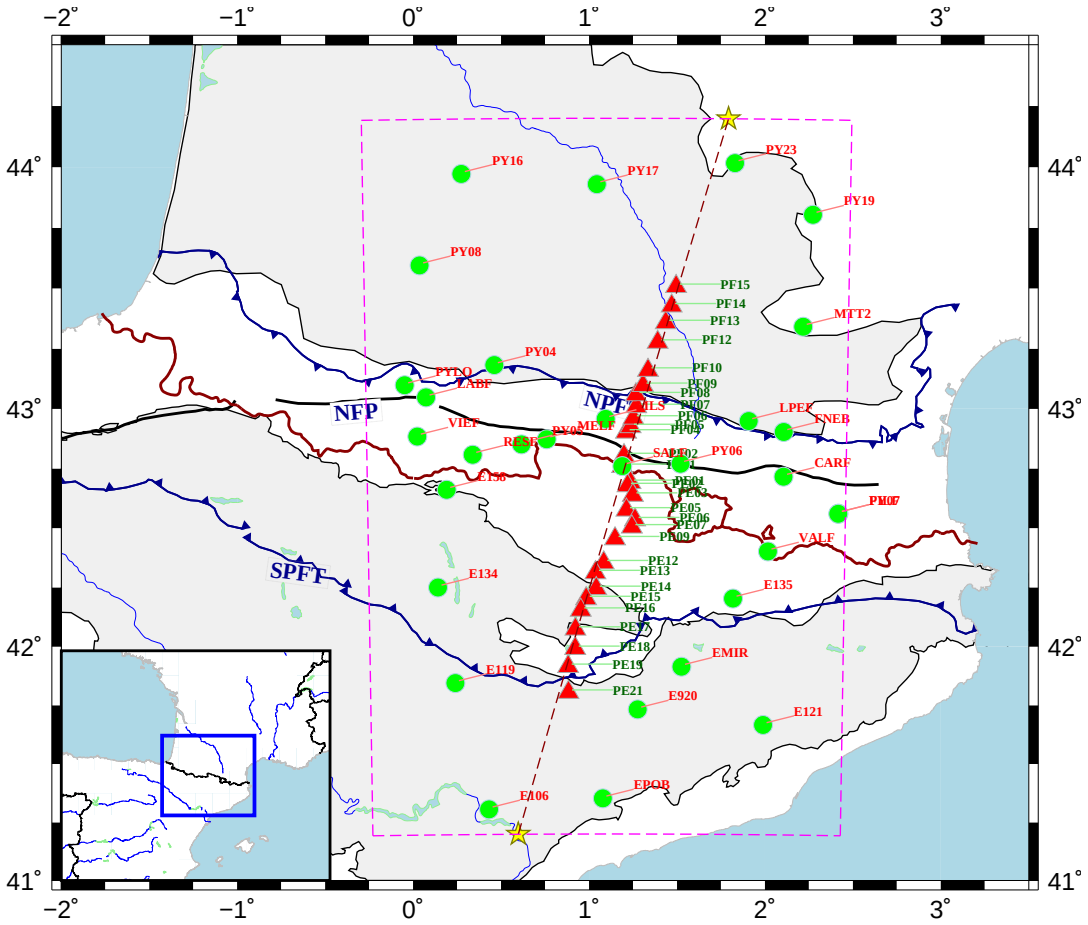


Figure 2.11: The locations of stations that recorded the 2012 May 24 event.

2.3.3.2 Data selection

Once the data traces are aligned, we check for the quality of aligned waveforms. Here we use a singular value decomposition (SVD) technique to determine an average reference trace. We then select the data traces by comparing every trace with this reference trace.

The SVD is a technique based on coherence between the different traces. It can simultaneously enhance the S/N ratio and suppress background noise (Bekara & der Baan, 2007). The aligned data in a chosen time window can be represented by a data matrix $\mathbf{D}$ (dimension is $m \times n$), with m traces and n time samples per trace (generally $m < n$). Assuming that data matrix $\mathbf{D}$ has rank $r < m$, the SVD of $\mathbf{D}$ leads to a linear orthogonal expansion of the data space given by:

$$\mathbf{D} = \sum_{p=1}^{r} \sigma_p \mathbf{u}_p \mathbf{v}_p^T \tag{2.22}$$

where the $\mathbf{u}_p$ and $\mathbf{v}_p$ are left and right singular eigenvector of dimensions m and n respectively, and σ_p is the $p-th$ singular value. The product $\mathbf{u}_p\mathbf{v}_p^T$ is a $(m \times n)$ matrix of rank one, called the $p-th$ eigenimage of data matrix $\mathbf{D}$.

If the traces in the chosen time window show a high degree of trace-to-trace coherence, then the coherent signals can be approximately reconstructed from the largest eigenimage with the largest eigenvalue. The SVD will act as a data-dependent, low-pass filter by rejecting highly incoherent noise. In practice, we keep the eigenimage corresponding to the largest singular value σ_1 and define the right singular eigenvector $\mathbf{v}_1$ as the reference waveform. The criteria used to select waveforms is that the normalized correlation coefficient between the reference waveform $\bar{u}_z(t)$ and each observed waveform $u_z(\mathbf{x_r}, t)$

$$CC(\mathbf{x_r}) = \frac{u_z(\mathbf{x_r}, t) \cdot \bar{u}_z(t)^T}{\parallel u_z(\mathbf{x_r}, t) \parallel_2 \parallel u_z(\mathbf{x_r}, t) \parallel_2} \tag{2.23}$$

is larger than 0.7. Figure 2.12a shows the comparison of a few representative traces. For permanent short period station RESF, the normalized correlation coefficient is below the threshold value, which leads us to reject this station. The waveform recorded by station PY07 has the highest normalized correlation coefficient among all the data traces. Figure 2.12b shows all the aligned vertical data traces that satisfy our correlation coefficient criterion.

2.3.3.3 Green's functions computation

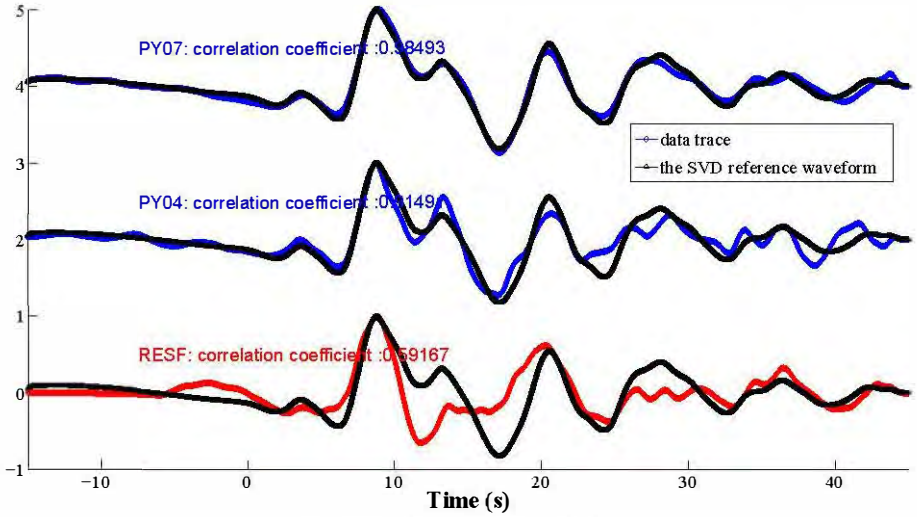
Once the data traces are aligned and selected, we also align their corresponding Green's function for the next step, which consists in obtaining wavelet functions by deconvolving these Green's functions from the observed waveforms.

The Green's functions can be computed in the approximate 1D starting model (global 1D layered AK135 model outside the regional mesh and 1D smooth model inside) with the hybrid method or with the global 1D method DSM. Note that the data and Green's functions are all normalized according to their maximum absolute amplitude since we mainly focus on the estimation of the source wavelet and ignore the absolute amplitude information for the time being, the amplitude correction will be introduced in the next section. Figure 2.13 shows the aligned vertical component Green's functions, which exhibit a high degree of trace-to-trace coherence.

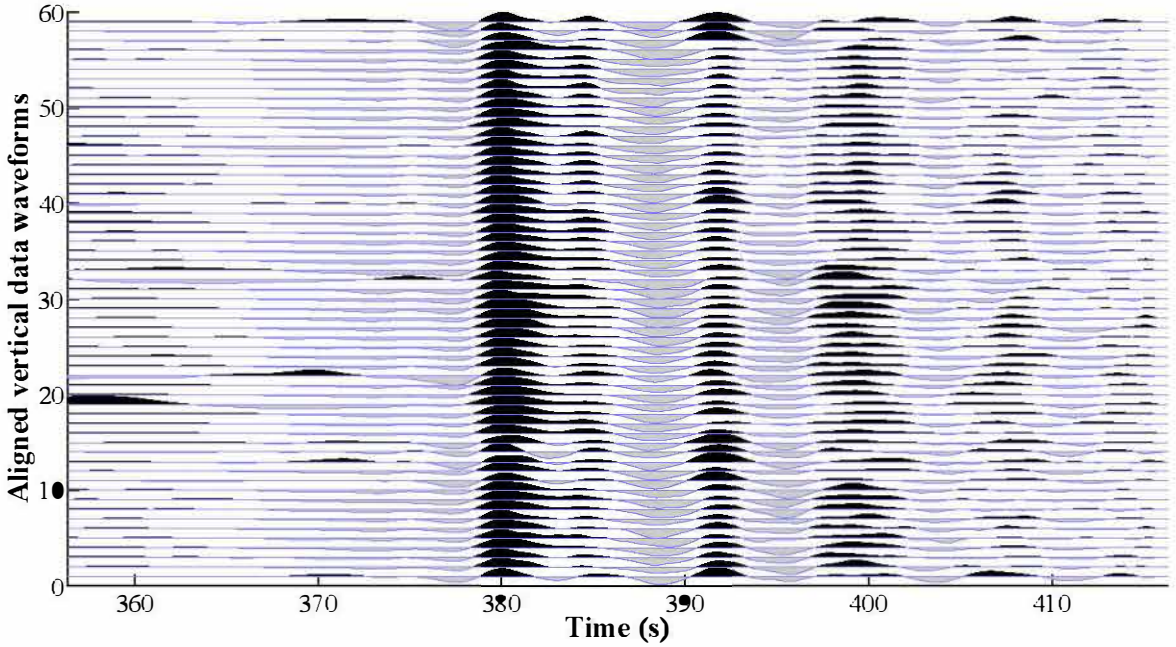
2.3.4 Wavelet estimation by deconvolution

The wavelet estimation problem (2.19) is a standard deconvolution problem. There are various methods that can be used to address this ill-posed problem. Most of them were introduced and applied in receiver function studies or source inversions. We use the time domain iterative deconvolution approach introduced by Ligorria & Ammon (1999).

The principle of this iterative deconvolution approach is a least-squares minimization of the difference between the observed signal and its estimate obtained by convolving iteratively a train



(a) Waveform quality check



(b) The aligned broad-band vertical component waveforms that satisfy the correlation coefficient criterion.

Figure 2.12: Illustration of the trace selection procedure based on the 2012 May 24 event.

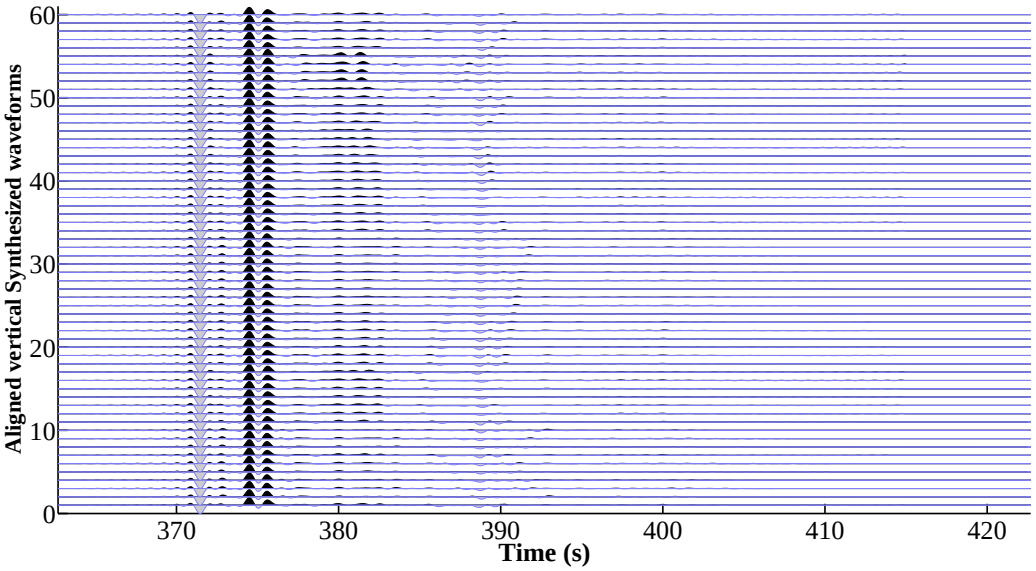


Figure 2.13: Aligned broad-band synthetic seismograms of the 2012 May 24 event.

of pulses with the simulated signal. The pulse train is progressively constructed by iterative cross-correlation between the observed signal and the simulated signal. At each iteration, the most highly correlated signal is extracted from the observed signal (Kikuchi & Kanamori, 1982). Ligorria and Ammon introduced this method for receiver function analysis. The idea is to iteratively construct the receiver function by adding a Gaussian pulse with the amplitude proportional to the highest cross-correlation coefficient. The time offset of the Gaussian pulse is obtained by the corresponding time lag. At each iteration, the contribution of the pulse introduced in the previous steps is stripped away from the observed signal. The iteration is executed until the misfit between the observed and modeled signal is below a preset threshold. A detailed description of this algorithm can be found in Kikuchi & Kanamori (1982).

Compared to other deconvolution approaches, the iterative deconvolution method has several significant advantages. Firstly, it provides a good estimate of the source wavelet at long period. The long period stability of the iterative time domain deconvolution approach stems from the reconstruction of the source wavelet by a sum of weighted Gaussian pulses. Secondly, it keeps the causality since it does not produce acausal troughs surrounding the P arrival as in the water-level deconvolution approach for example. Finally, it uses an intuitive way to progressively strip off the information from the objective signal, by extracting the largest and most important features first and then capturing the details.

Using the time domain iterative deconvolution approach, we obtain an estimate of the wavelet function at each station $W_r(t)$. Let us illustrate this algorithm on the vertical component waveform recorded at station PF01. In the first step, the lag time and height of the pulse are estimated from the absolute value of the cross-correlation between the vertical component of the Green's function $\mathbf{G}(\mathbf{x}_r, t)$ and the vertical component of the observed waveform $u_z(\mathbf{x_r}, t)$.

Then the wavelet function $W_r(t)$ for one station is constructed by adding a series of weighted and time lagged Gaussian pulses:

$$G(\omega) = \exp\left(\frac{-\omega^2}{4a^2}\right) \tag{2.24}$$

in the frequency domain, where a is the standard deviation of the Gaussian distribution. The parameter a can be considered as the bandwidth parameter of the spectrum of the time-domain Gaussian pulse. This Gaussian pulse is shifted at the estimated lag time and scaled by an amplitude equal to the value of the largest peak in the cross-correlation signal.

For the next step, the convolution of the current estimate of the wavelet function (denoted by $\hat{W}_r(t)$) with $\mathbf{G}(\mathbf{x}_r, t)$ is subtracted from the vertical data waveform $u_z(\mathbf{x_r}, t)$, and the procedure is repeated until the change of relative residue becomes smaller than a predetermined threshold. The relative residue at the $j-th$ iteration ε_j is defined as the ratio of the L_2 norm of the residue to the vertical component waveform $u_z(\mathbf{x_r}, t)$:

$$\varepsilon_j = \frac{\parallel u_z(\mathbf{x_r}, t) - \hat{u}_z(\mathbf{x_r}, t) \parallel_2^2}{\parallel u_z(\mathbf{x_r}, t) \parallel_2^2} \tag{2.25}$$

with:

$$\hat{u}_z(\mathbf{x_r}, t) = \mathbf{G}(\mathbf{x}_r, t) * \hat{W}_r(t) \tag{2.26}$$

This iterative deconvolution procedure is easy to implement. The choice of the Gaussian bandwidth parameter a will control the width of the time-domain Gaussian pulses. We test different a value and choose the parameter which leads to the minimum relative residue.

The aligned broad-band data waveform and corresponding Green's function are shown in Figure 2.14. The relative residues after deconvolution obtained with different Gaussian bandwidths a are shown in Figure 2.15. The inset panels along the curve showing the relative residues illustrate the shape of normalized time-domain Gaussian pulses for corresponding Gaussian bandwidth a. We find that $a = 38.8$ (rad/s) is the optimal Gaussian bandwidth leading to the minimum relative residue.

Figure 2.16 shows the iterative deconvolution process to construct the wavelet for station PF01. The comparison of data waveform and Green's function convolved with corresponding wavelet function at different iterations is shown in the right panel of Figure 2.16. After 254 iterations, the agreement between predicted vertical component waveform (the obtained wavelet function convolved with Green's function) and the vertical data waveform is quite good, with a power misfit less than 5%. Long period noise is sometimes present at some stations, but it can usually be filtered out by a high pass filter.

2.3.5 Determination and further applications of source wavelets

2.3.5.1 Determination of the average wavelet function

When the wavelet functions for all selected traces are obtained (shown in Figure 2.17), we determine an average wavelet function by using another singular value decomposition. The

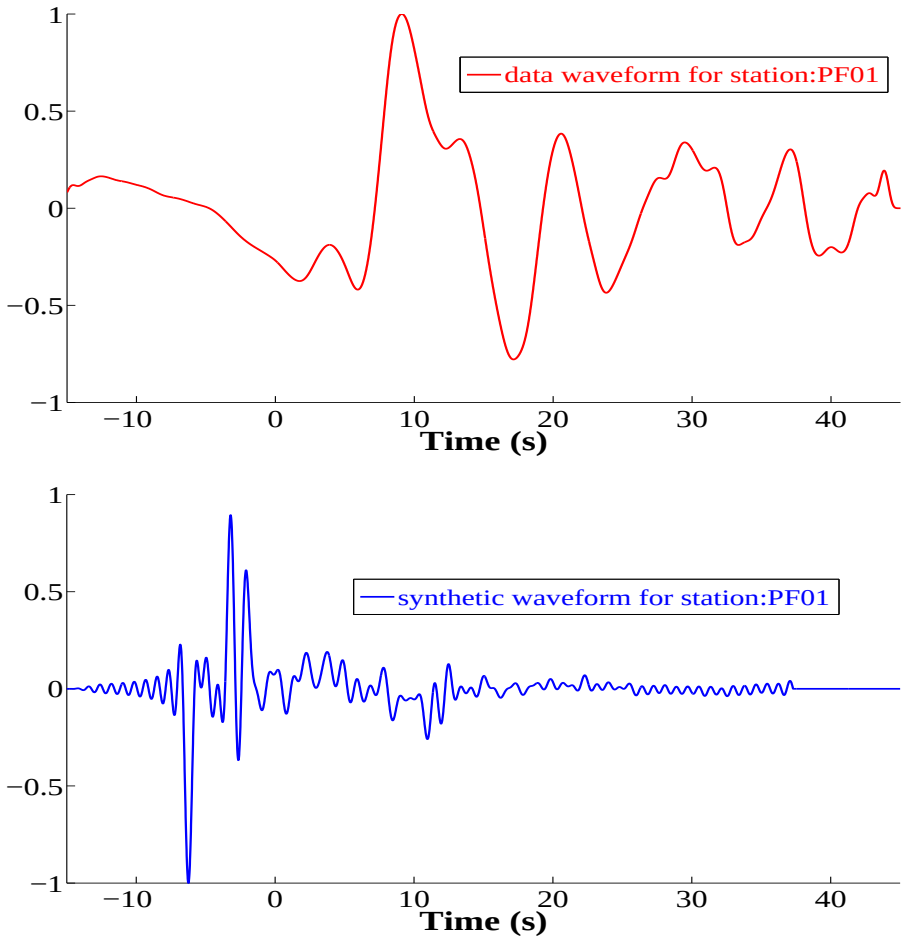


Figure 2.14: Aligned observed (top) and synthetic (bottom) waveforms for the 2012 May 24 event, recorded at station PF01.

blue curve in Figure 2.18a shows this SVD average wavelet function. We can see many high frequency oscillations appearing in this average wavelet. In order to eliminate these unwanted oscillations and capture the main features of the wavelet function, we smooth the SVD average wavelet function by convolving it with a Gaussian filter. This Gaussian filter is also defined in (2.24). The Gaussian bandwidth a controls the corner frequency of this low pass filter. In order to determine a proper Gaussian bandwidth a used for smoothing the SVD average wavelet function, the variations of the sum of correlation coefficients between observed vertical component waveforms and the corresponding Green's functions convolved with SVD average wavelet function versus Gaussian bandwidth are shown in Figure 2.18b. This curve drops monotonically with increasing a but asymptotically tends to convergence after an inflection point. This inflection point gives us the optimal value of a. (here we have $a = 5$). The smoothed wavelet function is shown in Figure 2.18a. This smoothed wavelet function is the one that we will retain for teleseismic wavefield modeling and FWI in the following.

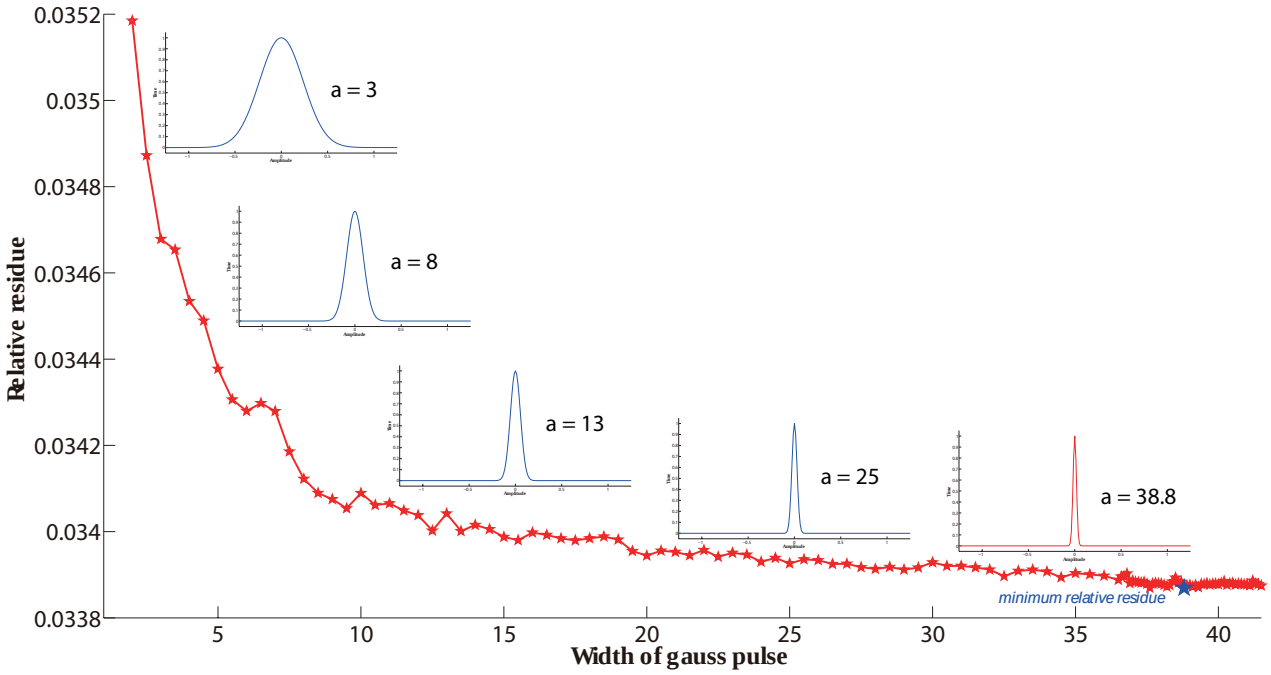


Figure 2.15: Wavelet estimation procedure for station PF01

2.3.5.2 Source depth estimation for wavelet estimation

In the wavelet estimation problem, there is a trade-off between an inaccurate depth in the CMT solution and apparent (deconvolved) source wavelet function. In this subsection, we will discuss and investigate this problem in detail.

In general, for shallow events the teleseismic P wave train results from a complicated interference between the direct P wave and the depth phases, such as P, pP, and sP. Teleseismic waveforms are thus strongly sensitive to the crustal structure in both the source and receiver sides and to the focal mechanism (source depth, source time function and radiation pattern) (Bostock, 2007).

Focal depth estimates coming from source inversions often come with large uncertainties. Compared to the radiation pattern, the uncertainty in focal depth has an evident influence on the shape of the Green's function since it directly determines the temporal spacing of the P, pP, and sP pulses. The shape of the deconvolved source wavelet thus strongly depends on the hypocentral depth, which results in a trade-off between focal depth and source time function. In the following, we will discuss different approaches to deal with this problem.

The positivity of the source wavelet is a physical constraint that can be used (Sigloch & Nolet, 2006). Positivity of the source wavelet comes from the physical assumption that the direction of rupture does not reverse during the earthquake event. However, the wavelet deconvolution process will incorporate the contribution of structures on the source and receiver

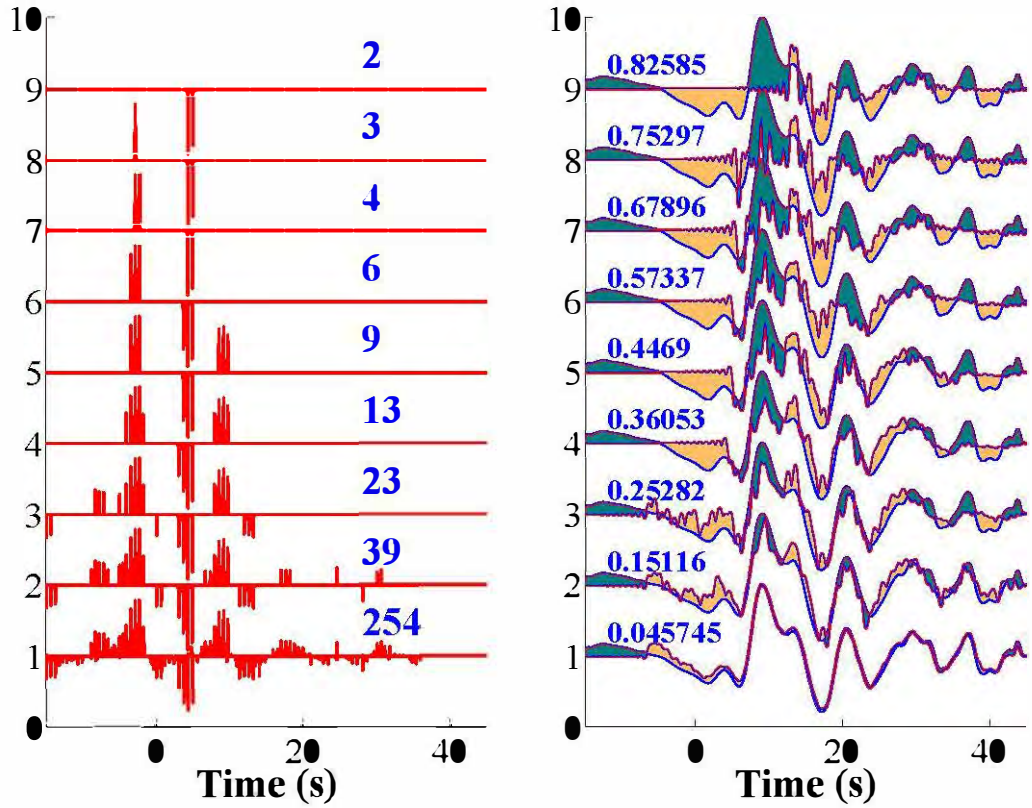


Figure 2.16: The right panel shows the comparison between the data waveform and corresponding Green's function convolved with wavelet estimation at different iteration. The iterative construction of the wavelet function is shown on the left. The numbers to the right of each wavelet function refer to the number of Gaussian pulses in the iterative deconvolution procedure. The ratio of the L_2 norm of the residue to the vertical component waveform is also shown in the right panel.

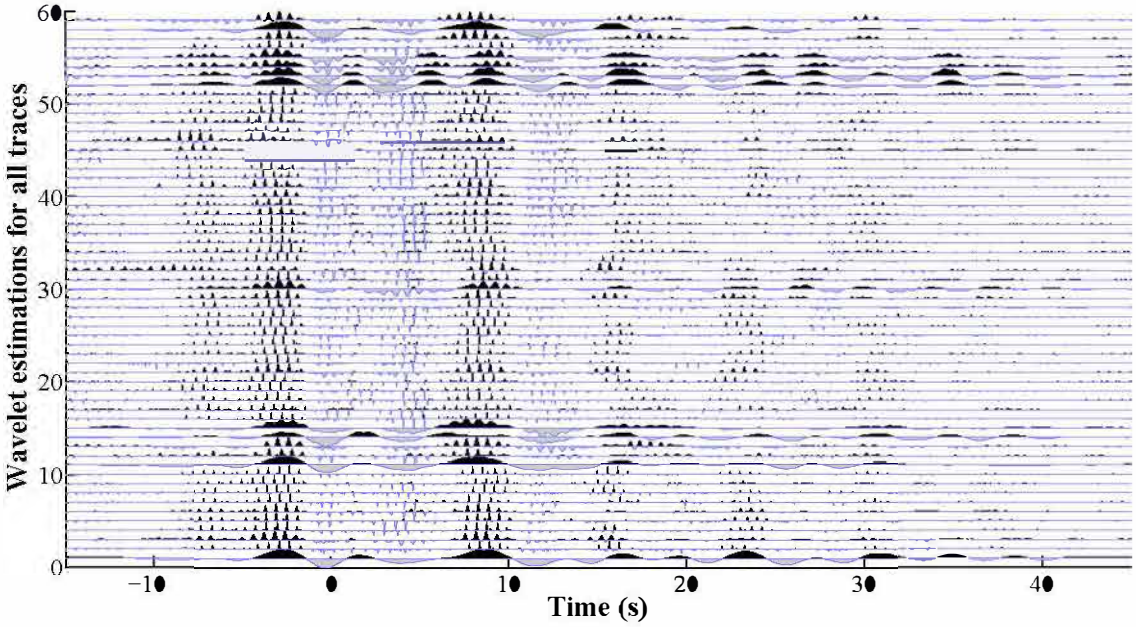
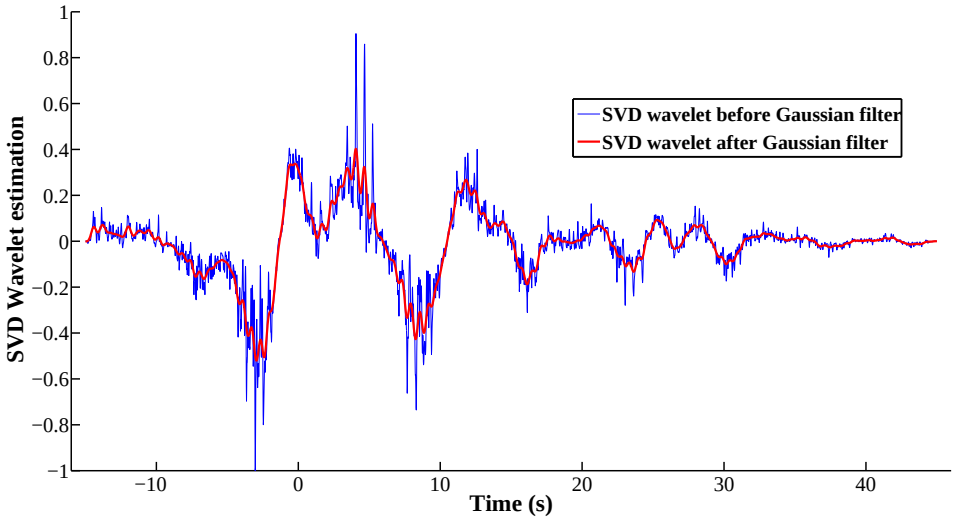
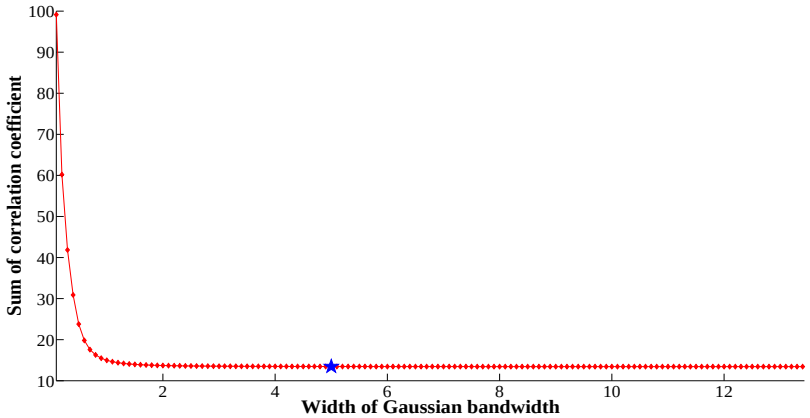


Figure 2.17: Aligned source wavelet functions estimated at each station.



(a) SVD Average wavelet function without and with Gaussian filter.



(b) Variations of correlation coefficient as a function of the band-pass width of Gaussian filter.

Figure 2.18: SVD Wavelet estimation.

sides into the source time function. In addition, since we apply different filters to the data, the positivity of the source time function will be lost, especially at short period. For this reason, we do not impose a strict positivity condition when we determine the source wavelet.

On the other hand, Christensen & Ruff (1985) have shown that using an inappropriate Green's function excited by an inaccurate source depth will introduce extra spectral peaks and holes to the spectrum of the source time function. In the time domain, these spectral peaks and holes translate into quasi-periodic oscillations. In other words, extra pulses will appear in the source time function. Therefore, an incorrect focal depth often leads to more complicated deconvolved source functions. This property leads them to propose a new procedure to determine the focal depth. After observing deconvolved source wavelets for a range of trial depths, the most plausible depth corresponds to the one that produces the simplest source wavelet. They defined different simplicity criteria that can be applied to the deconvolved source wavelet functions to estimate the hypocentral depth: 1) the varimax norm V_{max} and 2) the half-absolute moment time $T_{\frac{1}{2}}$. Firstly, the varimax norm V_{max} is expressed as:

$$V_{max} = \frac{\int_{T_s}^{T_e} W(t)^4 dt}{(\int_{T_s}^{T_e} W(t)^2 dt)^2} \tag{2.27}$$

where T_s and T_e are the start and end time for the deconvolved source wavelet $W(t)$. The varimax norm V_{max} reaches a value of one for a single nonzero spike (which would correspond to the maximum simplicity) and decreases when additional spikes are present in spite of their position in time. The highest value of V_{max} provide an estimate of the hypocentral depth.

Secondly, the half-absolute moment time $T_{\frac{1}{2}}$ of the source time function can be used for measuring the moment concentration property. It is defined as the half-absolute moment time of the temporal spacing between the onset time of the source wavelet function and the time at which half of the integral absolute value of the moment is achieved:

$$\int_{T_s}^{T_{\frac{1}{2}}} |W(t)| dt = \frac{1}{2} \int_{T_s}^{T_e} |W(t)| dt \tag{2.28}$$

In practice, the effect of an inappropriate depth assumption will increase the scatter of the moment release over the allowed duration of the source wavelet. The best depth corresponds to the source time function with the moment release most concentrated toward the beginning of the source wavelet. In other words, the optimal depth tends to predict a short source time function. Smaller values of $T_{\frac{1}{2}}$ indicate that the moment release is more concentrated toward the beginning of the source time function. Larger values of this parameter reflect the dispersed nature of the moment release. Both criteria successfully characterize the complexities that appear in the deconvolved source wavelet for incorrect assumed depths. The optimal depth of an event can be estimated from the maximum of the V_{max} versus depth curve and minimum of the $T_{\frac{1}{2}}$ versus depth curve.

Below we show an example to illustrate the procedure with these two simplicity criteria. On the 2012 May 24 event. Like the procedure introduced in the previous subsection, we first implement the wavelet estimation by using the Green's functions for a range of source depths (going from 5.75 and 13.75 km, with increments of 2 km). We use the DSM package to compute the Green's functions for different source depths in the 1D AK135 model, because the huge boundary traction database used in the Hybrid method simulation needs to be redone for different source depth which requires much more computation time than DSM simulations.

Figure 2.19 shows the aligned vertical broad-band synthetic seismograms for station CARF for different source depths. The Green's functions corresponding to the different source depths are aligned on the direct P wave. We estimate the source wavelet assuming these different source depths. Figure 2.20 shows the normalized SVD average wavelet functions obtained for these different source depths. The wavelets filtered with a 5 Hz Gaussian filter are also shown in the same figures.

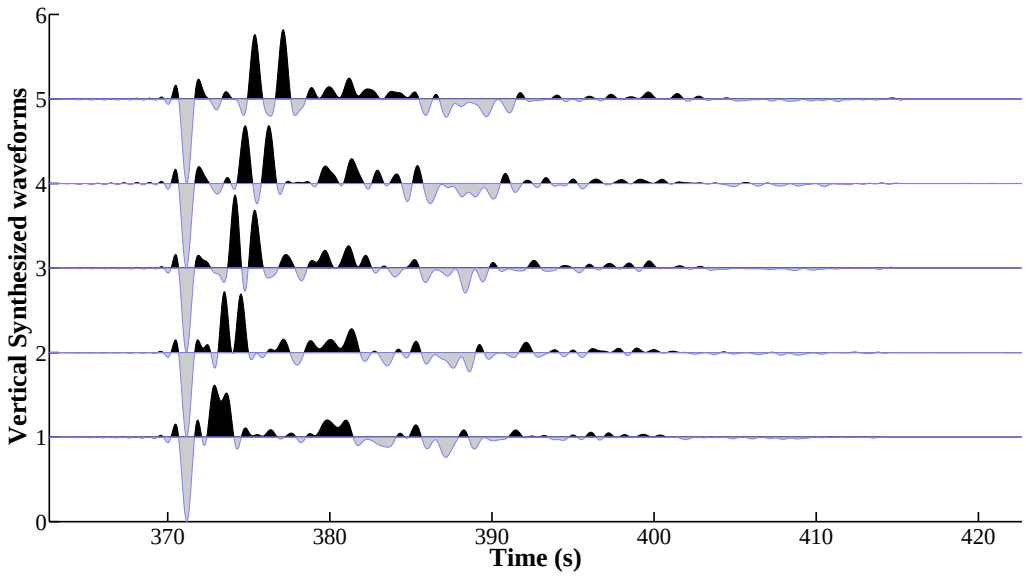
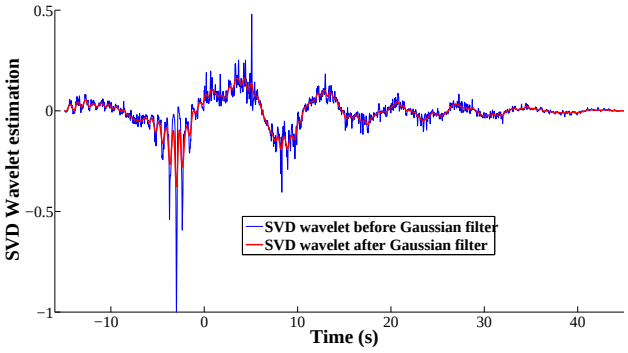
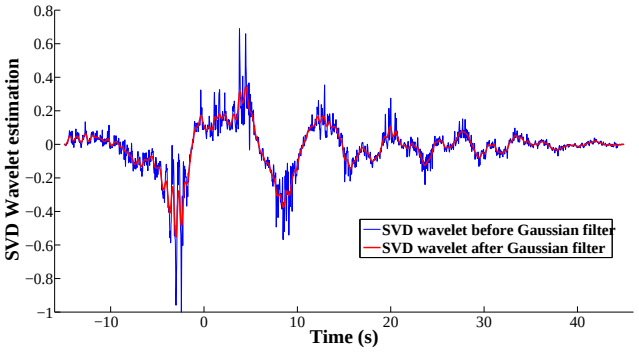


Figure 2.19: Aligned vertical broad-band synthetic seismograms for station CARF for different source depths. From the bottom to the top, the vertical component Green's functions correspond to source depth of 5.75, 7.75, 9.75, 11.75 and 13.75 km, respectively.

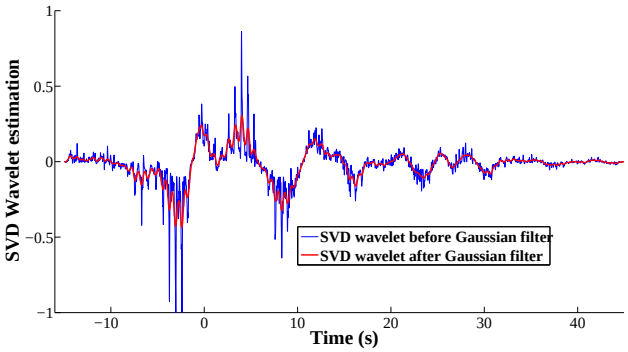
Table 2.21 lists the varimax norm V_{max}, half-absolute moment time $T_{\frac{1}{2}}$ and also the mean value of normalized correlation coefficient for all traces for the different source depth. The optimal source depth is given by the lowest values of the $T_{\frac{1}{2}}$ parameter or the largest values of the V_{max} parameter in Table 2.21. Christensen & Ruff (1985) showed that the upper limit for the optimal source depth is not very well constrained. Consequently, we infer that the apparent focal depth lies between 9.75 and 11.75 km for this event. The mean value of normalized correlation coefficients for all traces is a parameter to check the total waveform fit between



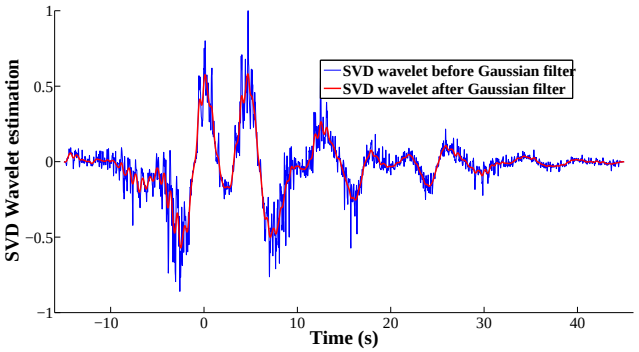
(a) Wavelet for hypocentral depth 5.75 km.



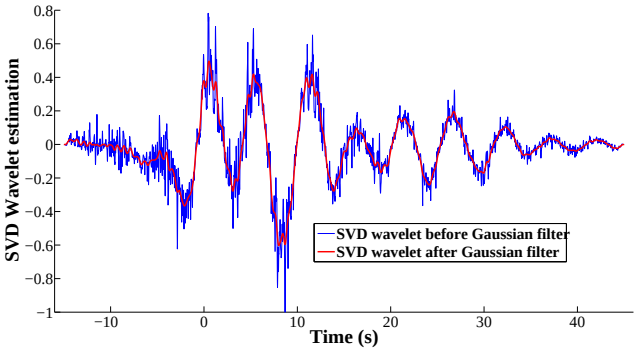
(b) Wavelet for hypocentral depth 7.75 km.



(c) Wavelet for hypocentral depth 9.75 km.



(d) Wavelet for hypocentral depth 11.75 km.



(e) Wavelet for hypocentral depth 13.75 km.

Figure 2.20: Wavelet estimation for different hypocentral depth, from 5.75 km (a) to 13.75 km (e), with increments of 2 km. The blue curve shows the unprocessed SVD average wavelet, red curve shows the Gaussian filtered SVD average wavelet, with Gaussian bandwidth $a = 5Hz$.

data and convolved synthetic seismograms. However, it does not show any significant variation in this source depth interval, which emphasizes the trade-off between hypocentral depth and source wavelet. This trade-off means that we had better use the $T_{\frac{1}{2}}$ and V_{max} as the indications of inferring the appropriate hypocentral depth.

Focal depth (km)	5.75	7.75	9.75	11.75	13.75
V_{max}	0.0013	0.0011	0.0012	0.0011	$9.2202e^{-4}$
$T_{\frac{1}{2}}$ (second)	15.8875	12.9375	12.0375	11.5	13.65
Average of normalized correlation coefficients	0.886	0.8854	0.8861	0.8865	0.8862

Table 2.21: Simplicity criteria for source wavelet function at different depth.

2.3.5.3 Teleseismic wavefield modeling by using average wavelet function

After determining the smoothed average SVD wavelet function $\bar{W}(t)$ with appropriate source depth, we can model the teleseismic wavefield by convolving the Green's function at each station $\mathbf{G}(\mathbf{x}_r, t)$ with this average wavelet function. The resulting synthetic seismogram at each station corresponding to the 'average' incident teleseismic wavefield. As pointed out in section 2.3.2, the waveform variations between different stations will allow us to invert the 3D regional heterogeneous structures. The differences between observed and synthetic waveforms give the misfit that we want to minimize in FWI. Before moving to the waveform inversion, we need to carefully take into account the amplitude of the modeled teleseismic wavefield.

The amplitude of recorded waveforms vary from one station to another and can be affected by many different effects: attenuation, local amplification, focusing/defocusing effects, etc. We describe the amplitude correction procedure as follows. Firstly, We define the amplitude factor $A(\mathbf{x_r})$ between the observed waveform $u_z(\mathbf{x_r}, t)$ and the modeled waveform $\tilde{u}_z(\mathbf{x_r}, t) = \mathbf{G}(\mathbf{x}_r, t) * \bar{W}(t)$ as:

$$A(\mathbf{x_r}) = \frac{u_z(\mathbf{x_r}, t) \cdot \tilde{u}_z(\mathbf{x_r}, t)^T}{\| \tilde{u}_z(\mathbf{x_r}, t) \|_2^2} \tag{2.29}$$

This definition is derived from the least squares minimization of the differences between observed and normalized synthetic waveforms:

$$\max_{A(\mathbf{x_r})} \| u_z(\mathbf{x_r}, t) - A(\mathbf{x_r}) \cdot \tilde{u}_z(\mathbf{x_r}, t) \|_2^2 \tag{2.30}$$

When the amplitude factors for all traces are obtained, we average them to get an initial average amplitude factor:

$$\bar{A}_0 = \sum_{r=1}^{N} A(\mathbf{x_r}) \tag{2.31}$$

By using this initial average amplitude factor $\bar{A}_0$, we can select those stations with moderate amplitude factor with the condition:

$$| A(\mathbf{x_r}) - \bar{A}_0 | < b \tag{2.32}$$

where b is a threshold value set to 0.2 in our study. After this initial selection, the final average amplitude factor $\bar{A}$ is the average amplitude of all the chosen stations. We use this $\bar{A}$ value to scale all the modeled teleseismic wavefield.

2.3.5.4 Definition of travel time and amplitude anomalies of teleseismic P waves

The most popular definition of amplitude anomaly is the relative RMS amplitude misfit between observed and modeled waveforms, introduced by Dahlen & Baig (2002). It is defined by:

$$\Delta A(\mathbf{x_r}) = \log_{10}\left(\frac{A_m(\mathbf{x_r})}{A_d(\mathbf{x_r})}\right) \tag{2.33}$$

where:

$$A_m(\mathbf{x_r}) = \left[\int_R w_r(t)\tilde{u}_z^2(\mathbf{x}_r, t)dt\right]^{1/2} \tag{2.34}$$

and:

$$A_d(\mathbf{x_r}) = \left[\int_R w_r(t)u_z^2(\mathbf{x}_r, t)dt\right]^{1/2} \tag{2.35}$$

In this study, we define the amplitude anomaly as the ratio of the amplitude factor at each station (defined by (2.29)) to the average amplitude factor $\bar{A}$:

$$\Delta A(\mathbf{x_r}) = \log_{10}\left[\frac{A(\mathbf{x_r})}{\bar{A}}\right] \tag{2.36}$$

The travel time anomaly is defined as the cross-correlation time shift $\Delta T(\mathbf{x_r})$ between observed and modeled waveforms measured over a chosen time window in (2.21).

The comparison between observed and synthetic teleseismic waveforms at station PF01 is shown in Figure 2.22. The observed and corresponding synthetic waveforms are shown by green and red curves, respectively. The vertical black lines show the start time and end time of the window used to measure the amplitude and travel time anomalies. In Figure 2.22, we also show the amplitude normalized waveform (blue curve) with the amplification factor $\Delta\tilde{A} = \frac{A(\mathbf{x_r})}{\bar{A}}$ shifted by the time delay ΔT. These corrections minimize the least-square misfit in the chosen time window, which has a length of 30 s and start 7.5 s before the theoretical P arrival time. Therefore, the shifted and normalized waveform has the form : $\Delta\tilde{A}\tilde{u}_z(\mathbf{x_r}, t - \Delta T)$. In this example, $\Delta\tilde{A} = 1.129$ and $\Delta T = -0.3s$.

2.3.5.5 Travel time and amplitude anomalies measured on the vertical component of the 2012 May 24 event

We can now compute the travel time and amplitude anomalies at all the selected stations for the 2012 May 24 event. Figure 2.23 shows the travel time anomalies measured on the vertical component. These anomalies are computed in a time window starting 7.5 s before the direct P

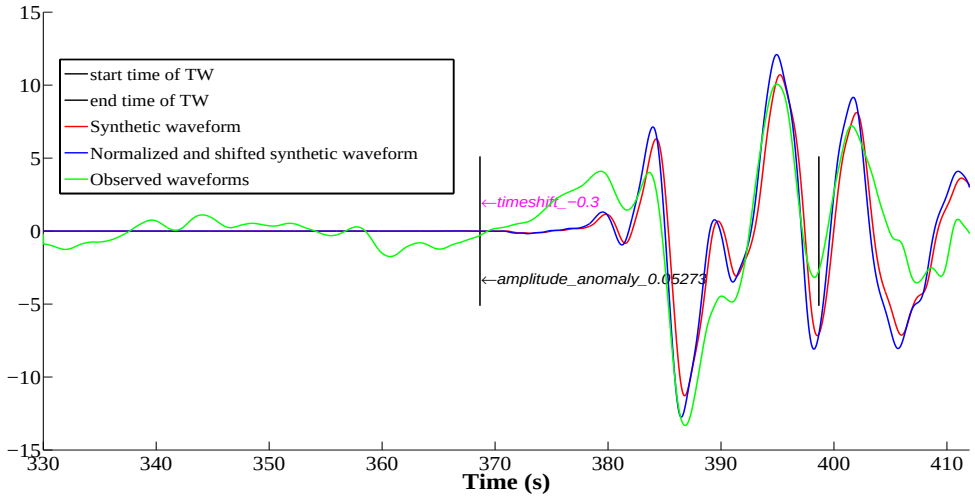


Figure 2.22: Comparison between observed and synthetic vertical component waveforms for station PF01. The thick vertical lines denote the start and end times of a chosen time window.

wave and ending 30 s after. The travel time anomalies show a spatially coherent pattern with small anomalies observed on the Spanish side of the Pyrenees and positive/negative anomalies on the French side.

Figure 2.24 shows the amplitude anomalies for the vertical component of the 2012 May 24 event. It is worth noting that the stations with the largest amplitude anomalies are the short period stations of the RSSP (Réseau de Surveillance Sismique des Pyrénées) permanent network.

Interestingly, a temporary broad-band station PF01 of the PYROPE deployment was installed at the same location as the short period SALF station. Figure 2.25 shows the comparison of the vertical component records at stations PF01 and SALF. We can see that the permanent short period station SALF apparently underestimate the amplitude by more than 20% compared to the broad-band station PF01. The reason for this underestimation is mainly coming from the instrumental response of short period seismometer which results in a loss of the low frequency content that has relative large amplitudes. Therefore, if we want to invert the waveform recorded by these short period stations, we need to carefully select the frequency content that will be exploited in FWI.

Figure 2.26a shows the observed and synthetic vertical component waveforms along the central transect (red and blue curve, respectively). We can see that some stations have large amplitude long period noises which need to be filtered out before used in waveform inversion. Figure 2.26b shows the waveform differences for stations located on the northern French part of the transect (stations names starting with 'PF'). The dark cyan region shows the part of the modeled waveform that is larger than the observed waveform, while the sandy brown region shows the part of the observed waveform that is larger in amplitude than the synthetic waveform.

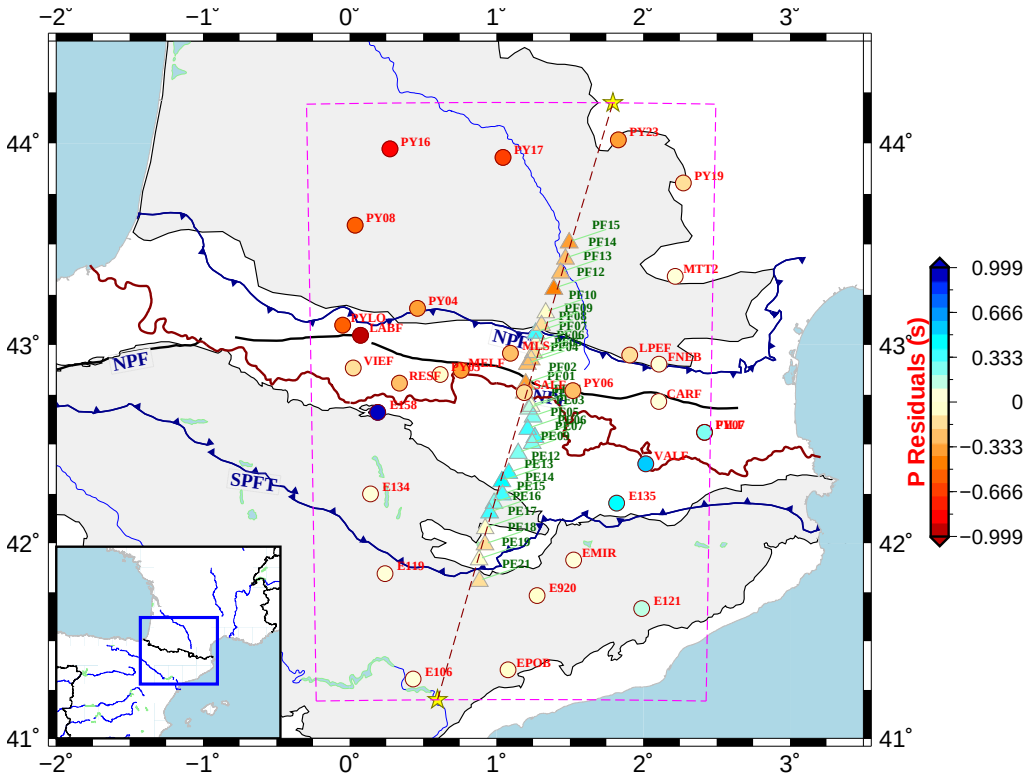


Figure 2.23: Broad-band travel time anomalies measured on the vertical components of the 2012 May 24 event.

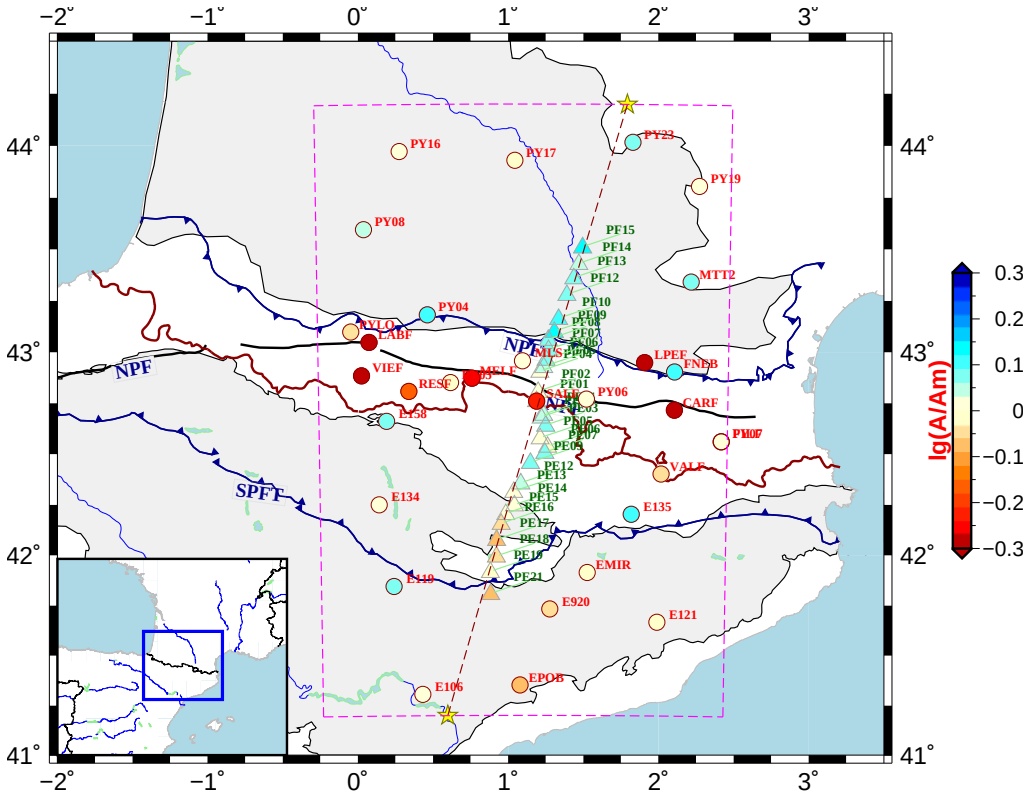


Figure 2.24: Broad-band amplitude anomalies measured on the vertical components of the 2012 May 24 event.

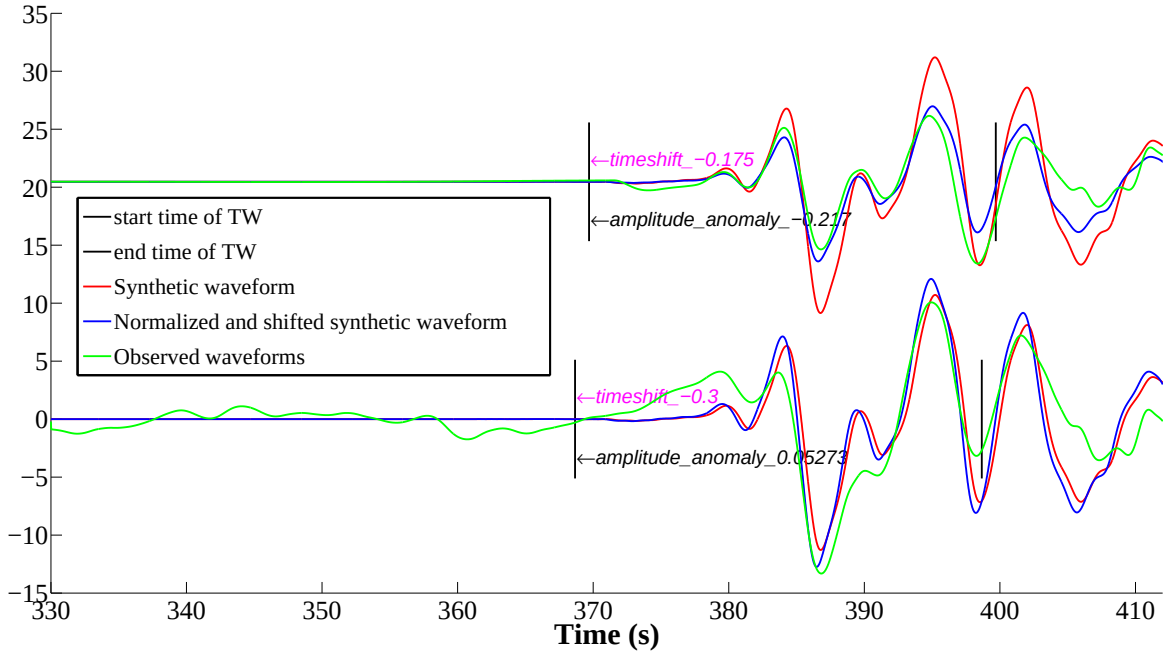


Figure 2.25: Comparison between observed and synthetic vertical component waveforms for stations SALF and PF01. The thick vertical lines mark the beginning and end of the chosen time window.

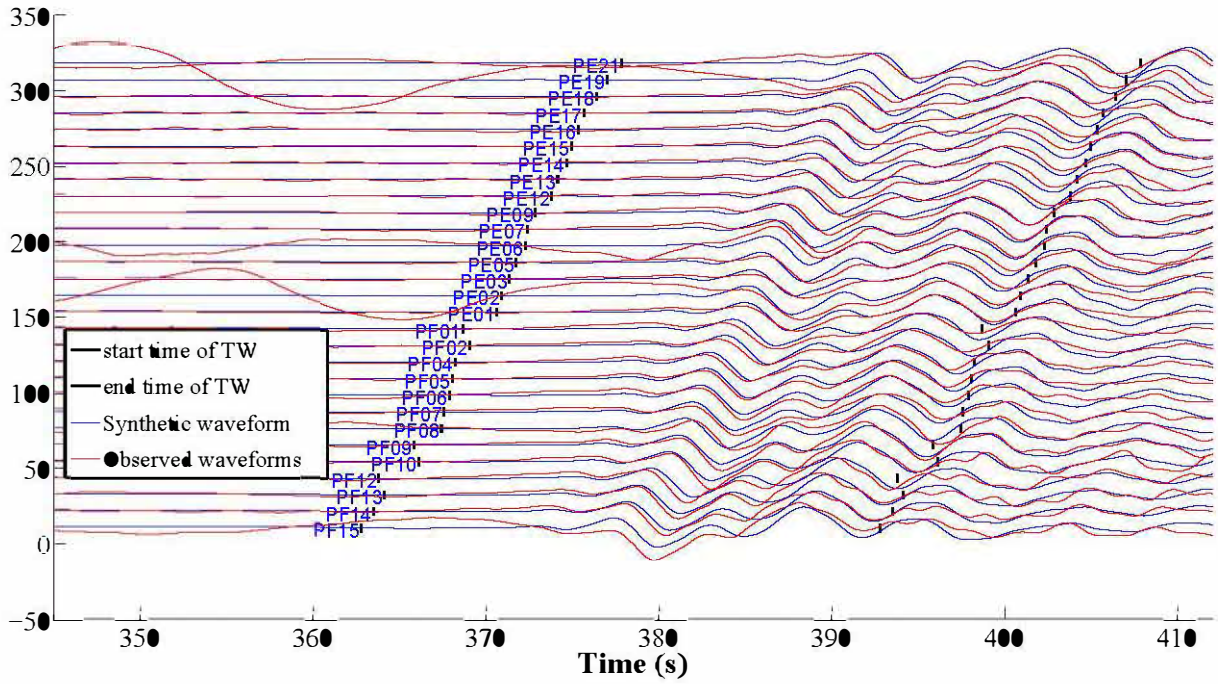
From Figure 2.23, most stations of the central transect show negative travel time anomalies, except for station PF07 and PF08. The negative anomaly means that the P wave of corresponding station arrives with a delay, which suggests the presence of a low velocity structure beneath those stations.

2.3.5.6 Travel time and amplitude anomalies measured on the radial component of the 2012 May 24 event

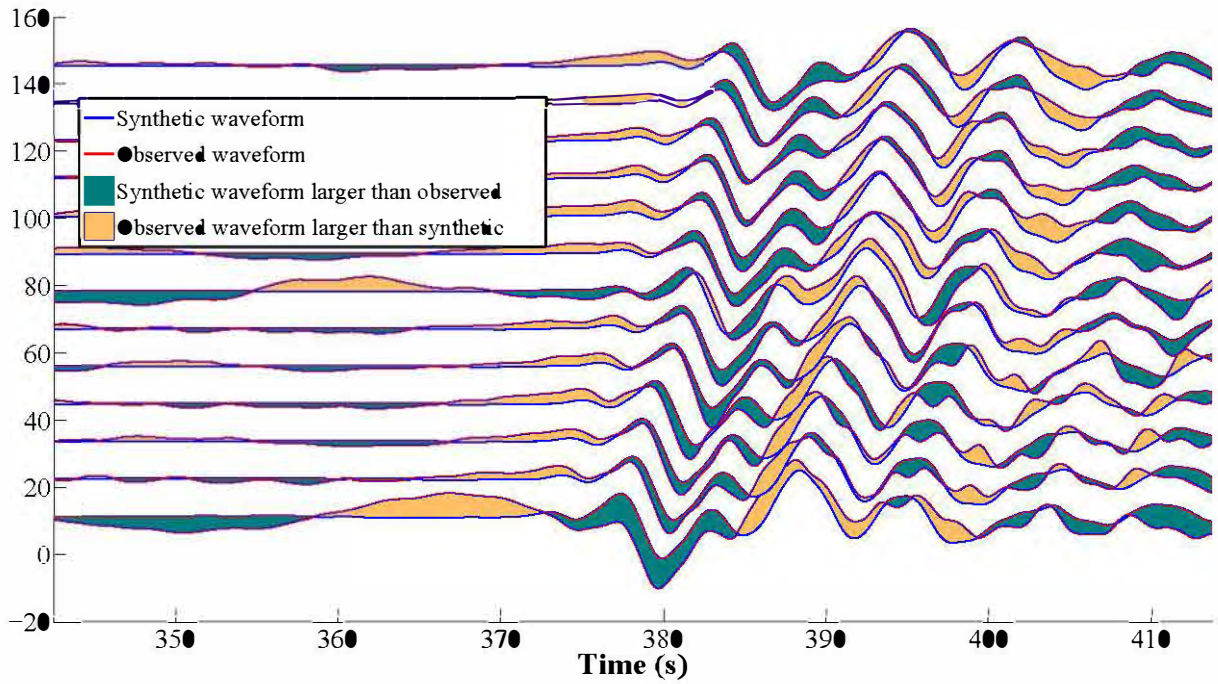
To compute radial component synthetic seismograms, the average wavelet $\bar{W}(t)$ is convolved with the radial component Green's function. By inspecting the normalized coefficient of the radial component waveform between data and modeled wavefield, we can select the stations with good quality radial component records for inversion.

We compute the travel time and amplitude anomalies following the same procedure as for the vertical component. Figures 2.27 and 2.28 show the travel time and amplitude anomalies of the radial component of the 2012 May 24 event. Compared to Figure 2.24, we observe larger amplitude anomalies on the radial component.

Figure 2.29 shows the waveform differences of radial components for stations located on the northern French part of the central transect. Again the waveform differences are much larger on the radial component than on the vertical component. There are several reasons for this. Firstly, the radial component waveforms have relatively lower S/N ratio compared to the



(a) Vertical components of observed and synthetic waveforms



(b) Zoom in of waveforms of selected traces

Figure 2.26: (a) Observed and synthetic vertical component waveforms. The start and end times of the time window used for computating anomalies are shown with black lines. (b) Zoom showing the waveforms of selected traces.

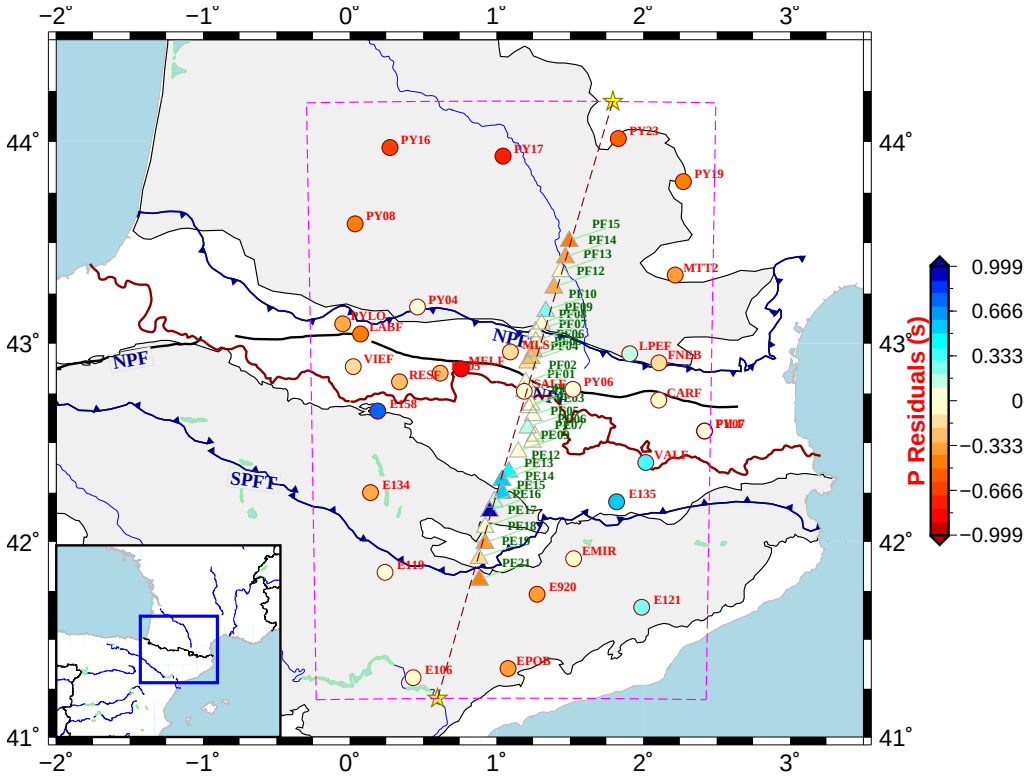


Figure 2.27: Broad-band travel time anomalies measured on the radial components of the 2012 May 24 event.

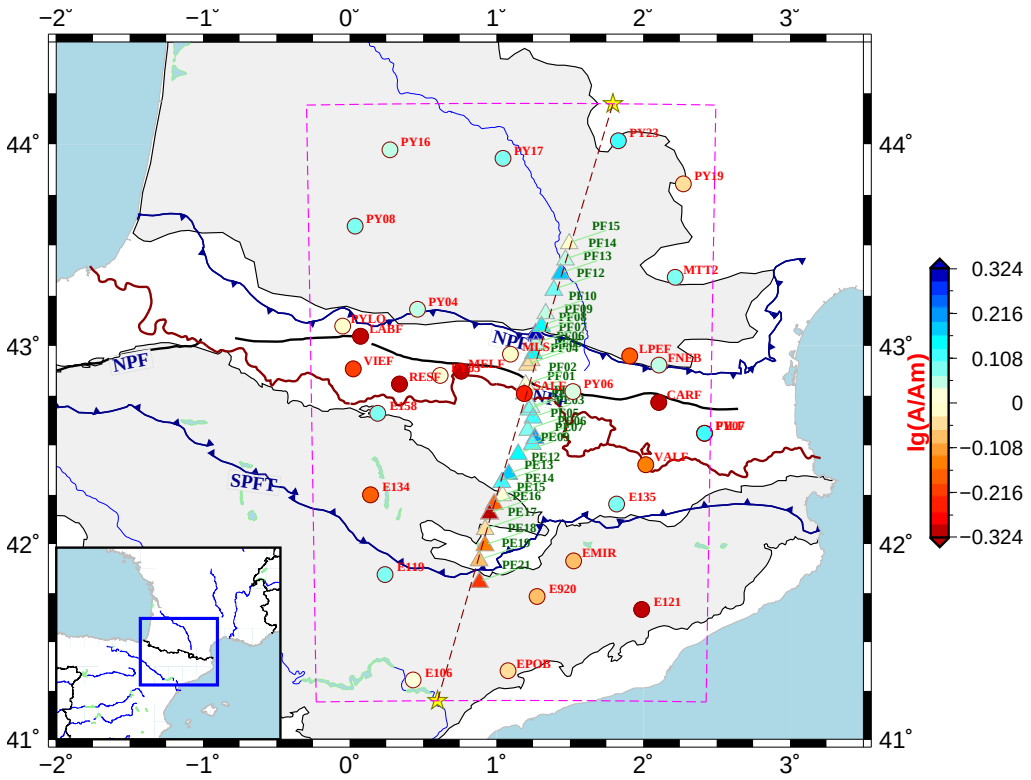


Figure 2.28: Broad-band amplitude anomalies measured on the radial components of the 2012 May 24 event.

vertical components. Secondly, the P-to-S conversions and crustal multiples are not modeled by the simple 1D model. Thirdly, complex surface topography will affect the radial component waveforms significantly. Finally, reverberations in the sedimentary basins on both sides of the Pyrenees may also have a strong influence on the radial component waveforms.

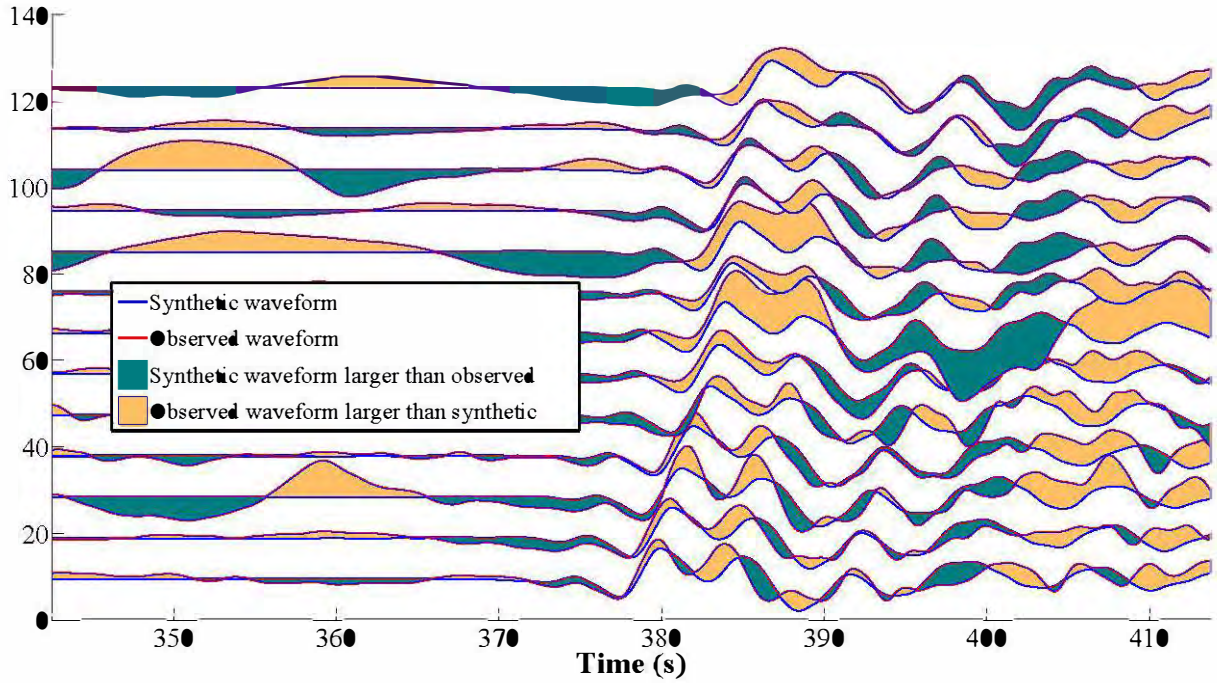


Figure 2.29: Observed and synthetic radial component waveforms. The waveform differences of radial components for stations located on the northern French part of the transect are shown by dark cyan and sandy brown regions.

Chapter 3

The strategy of Full waveform inversion

Contents

Part of this chapter was published as a paper entitled "Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method" by Vadim Monteiller, Sébastien Chevrot and Dimitri Komatitsch and Yi Wang in Geophysical Journal International, 202: 811-827, 2015.

3.1 General formulation of the FWI problem

3.1.1 The gradient and Hessian of the cost function

As reviewed in the first chapter, the objective of the seismic inversion problem is to find an optimal Earth model, $\mathbf{m}$, that minimizes a functional $\chi(\mathbf{m})$, usually defined as the discrepancies between a set of observed data, for example waveforms or travel times, and a corresponding set of synthetics, generated by the model $\mathbf{m}$. The model $\mathbf{m}$ may comprise, for instance, the spatial distributions of both structural and earthquake source parameters such as compressional wave velocity $\alpha(\mathbf{x})$, the shear wave velocity $\beta(\mathbf{x})$, and the density $\rho(\mathbf{x})$, the centroid-moment tensor $\mathbf{M}$, centroid location $\mathbf{x}_s$, source time function S(t), such that dS(t)/dt is the moment-rate function:

$$\mathbf{m}(\mathbf{x}) = [m_1(\mathbf{x}), m_2(\mathbf{x}), m_3(\mathbf{x}), ...] = [\alpha(\mathbf{x}), \beta(\mathbf{x}), \rho(\mathbf{x}), \mathbf{M}, \mathbf{x}_s, S(t), ...] \tag{3.1}$$

Our aim is to find perturbations of the model $\delta\mathbf{m}$ that reduce the misfit between observed and synthetic data. The functional $\chi(\mathbf{m})$ is called misfit or cost function. For example, χ can represent the waveform or travel time square differences. Since χ is generally a complex nonlinear functional of the model function, we approximate the optimum model $\tilde{\mathbf{m}}$ with an iterative opimization algorithm, starting from a reasonable initial guess model $\mathbf{m}_0$, and successively updating the current model $\mathbf{m}_i$ to a new model $\mathbf{m}_{i+1}$ according to:

$$\mathbf{m}_{i+1} = \mathbf{m}_i + s_i \mathbf{p}_i \qquad \text{with} \qquad \chi(\mathbf{m}_i) < \chi(\mathbf{m}_{i+1}) \tag{3.2}$$

where $\mathbf{p}_i$ is the steepest descent direction, and s_i is the step length of iteration i. The iterative procedure is repeated until a good fit of the data is obtained. In the vicinity of the current model $\mathbf{m}$, the misfit function can be expanded into a Taylor series:

$$\chi(\mathbf{m} + \delta\mathbf{m}) \approx \chi(\mathbf{m}) + \mathbf{g}(\mathbf{m}) \cdot \delta\mathbf{m} + \frac{1}{2}\delta\mathbf{m} \cdot \mathbf{H}(\mathbf{m}) \cdot \delta\mathbf{m} \tag{3.3}$$

where $\mathbf{g}(\mathbf{m})$ is the first derivative of the misfit function with respect to $\mathbf{m}$ (also known as the Fréchet derivative) given by:

$$\mathbf{g}(\mathbf{m}) = \nabla_m \chi(\mathbf{m}) \tag{3.4}$$

It is a linear functional that operates on the model $\mathbf{m}$. The Hessian operator $\mathbf{H}(\mathbf{m})$ is defined as the second derivative of the misfit function:

$$\mathbf{H}(\mathbf{m}) = \nabla_m \nabla_m \chi(\mathbf{m}) \tag{3.5}$$

The Hessian is a linear operator that maps the model space to its dual space. The gradient and Hessian depend on model $\mathbf{m}$. In the following, this dependence will be implicitly assumed and omitted in the notations for simplicity. The gradient of (3.3) with respect to model perturbation $\delta\mathbf{m}$ is given by:

$$\mathbf{g}(\mathbf{m} + \delta\mathbf{m}) \approx \mathbf{g}(\mathbf{m}) + \mathbf{H}(\mathbf{m}) \cdot \delta\mathbf{m} \tag{3.6}$$

Once a minimum of χ is reached, the perturbation direction is given by:

$$\begin{aligned} \mathbf{H}(\mathbf{m}) \cdot \delta\mathbf{m} &= -\mathbf{g}(\mathbf{m}) \\ \delta\mathbf{m} &= -\mathbf{H}^{-1} \cdot \mathbf{g} \end{aligned} \tag{3.7}$$

If the gradient and Hessian (or approximate Hessian) are both available, then the inversion algorithm is known as a Newton method. The Newton method has a quadratic convergence. If only the gradient is available, then one has to rely on a gradient based method (such as the steepest descent, conjugate gradient method).

In classical travel time tomography, the gradient $\mathbf{g}$ and Hessian $\mathbf{H}$ of the misfit function can be both computed easily, such that the model update $\delta\mathbf{m}$ may be obtained from (3.7). When we consider the inversion problem based on full waveform modeling in complex 3D model, computation of the gradient is generally still possible. However, the direct computation of the Hessian is usually prohibitive.

3.1.2 The choice of misfit function

The choice of the misfit function is determinant in seismic tomography. The misfit function should exploit as much information as possible and should be adapted to the restrictions imposed by the collected data and physics of the problem. In our study, we perform full waveform inversion and compare the results from FWI with adjoint travel time inversion.

Full waveform inversion means that one considers the observed seismograms (possibly filtered) as the basic observables to fit. One thus searches for the model that minimizes the mean squared difference between observed and synthetic seismograms. In other words, the goal is to find a model that can explain a larger portion of seismological records, and not simply the phase of a few seismic arrivals. The waveform misfit function is defined as:

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T w_r(t) \parallel \mathbf{u}(\mathbf{x}_r, \mathbf{x}_s; t) - \mathbf{d}(\mathbf{x}_r, \mathbf{x}_s; t)\|_2^2 dt \tag{3.8}$$

Here $w_r(t)$ denotes the temporal window function for the specific section of waveform measurement. This functional quantifies the L_2 difference between the observed waveforms $\mathbf{d}(\mathbf{x}_r, \mathbf{x}_s; t)$ at receivers $\mathbf{x}_r$, $r = 1, ..., M$ produced by sources at $\mathbf{x}_s$, $s = 1, ..., N$, and the corresponding synthetic seismograms $\mathbf{u}(\mathbf{x}_r, \mathbf{x}_s; t)$ computed in model $\mathbf{m}$. While this misfit function is indeed classical, it is worth mentioning that in the case of noisy real data other norms could be used, since in the oil industry for instance it is known that the L_1 norm (Crase et al., 1990; Brossier et al., 2010), hybrid L_1-L_2 norms (Bube & Langan, 1997), Hubert norm (Ha et al., 2009), Student-t distribution (Aravkin et al., 2011; Jeong et al., 2015), etc., can be more robust than the L_2 norm used here in the context of synthetic data with no noise.

With the cross-correlation time shift ΔT introduced in section 2.3.3.1, the corresponding misfit function is defined as:

$$\chi_T(\mathbf{m}) = \frac{1}{2}\sum_{s=1}^{N}\sum_{r=1}^{M}\Delta T_{sr}^2 \tag{3.9}$$

where ΔT_{sr} denotes the travel time of a specific phase measured at receiver $\mathbf{x}_r$ excited by source $\mathbf{x}_s$. The relative RMS amplitude misfit (Ritsema et al., 2002) is defined as (slightly different from the section 2.3.5.4):

$$\begin{aligned} \chi_A(\mathbf{m}) &= \sum_{s=1}^{N}\sum_{r=1}^{M} E_{sr} \\ E_{sr} &= \frac{1}{2}\frac{(A^0 - A)^2}{A^2} = \frac{1}{2}\frac{(\Delta A)^2}{A^2} \end{aligned} \tag{3.10}$$

where:

$$A^0 = [\int_0^T w_r(t) d_i(\mathbf{x}_r, \mathbf{x}_s; t)^2 dt]^{1/2} \tag{3.11}$$

and:

$$A = [\int_0^T w_r(t) u_i(\mathbf{x}_r, \mathbf{x}_s; t)^2 dt]^{1/2} \tag{3.12}$$

We can filter the seismograms to measure travel time and amplitude in multiple frequency bands.

3.2 Computation of the gradient based on the adjoint method

In our study, we mainly focus on the waveform inversion by using the waveform misfit function. A direct method to compute the gradient is to take the derivative of (3.8) with respect to model parameters:

$$\frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}} = \sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T \frac{\partial \mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)}{\partial \mathbf{m}} \cdot w_r(t)\left[\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)-\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)\right] dt\,. \tag{3.13}$$

This equation can be reformulated as the matrix-vector product:

$$\mathbf{g} = -\mathbf{J}^* \cdot \delta\mathbf{d}\,, \tag{3.14}$$

where $\mathbf{J}^*$ is the adjoint of the Jacobian matrix of the forward problem that contains the Fréchet derivatives of the data with respect to model parameters, and $\delta\mathbf{d}$ is the vector that contains the data residuals. The determination of $\mathbf{J}$ would require computing the Fréchet derivatives for each time step in the time window considered and for all the source-station pairs, which is prohibitively expensive on current computers and supercomputers (let us note that this situation may change one day). However, it is possible to obtain this gradient without computing the Jacobian matrix explicitly. The approach to determine the gradient without computing the Fréchet derivatives has been introduced in non-linear optimization by Chavent (1974) working with J. L. Lions, and later applied to seismic exploration problems by Bamberger et al. (1982); Lailly (1983) and Tarantola (1984). The idea is to resort to the adjoint state, which corresponds to the wavefield emitted and back-propagated from the receivers (e.g. Tromp et al., 2005; Fichtner et al., 2006; Plessix, 2006; Tromp et al., 2008).

Let us give an outline of the theory to compute the gradient with the adjoint method, and refer the reader to, for example, Tromp et al. (2005, 2008) for further details. The perturbation of the misfit function can be expressed as:

$$\delta\chi(\mathbf{m}) = \sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T w_r(t)\left[\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)-\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)\right]\cdot\delta\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)\,dt\,, \tag{3.15}$$

where $\delta\mathbf{u}$ is the perturbation of displacement given by the first order Born approximation (e.g. Hudson, 1977):

$$\begin{aligned}\delta\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t) &= -\int_0^t\int_V \left[\delta\rho(\mathbf{x})\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t')\cdot\partial_{t'}^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t')\right.\\ &+ \left.\nabla\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t'):\delta\mathbf{c}(\mathbf{x}):\nabla\mathbf{u}(\mathbf{x};t')\right]\,d^3\mathbf{x}\,dt'\,.\end{aligned} \tag{3.16}$$

In this expression, $\mathbf{G}$ is the Green's tensor, $\delta\rho$ the perturbation of density, and δc the perturbation of the fourth-order elasticity tensor. Inserting (3.16) into (3.15) we obtain:

$$\begin{aligned}\delta\chi(\mathbf{m}) &= -\sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T w_r(t)\left[\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)-\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)\right]\int_0^t\int_V\big[\delta\rho(\mathbf{x})\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t')\cdot\partial_{t'}^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t')\\ &+ \nabla\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t'):\delta\mathbf{c}(\mathbf{x}):\nabla\mathbf{u}(\mathbf{x},t')\big]\,d^3\mathbf{x}\,dt'\,dt\,. \qquad (3.17)\end{aligned}$$

Defining the waveform adjoint source for each source at $\mathbf{x}_r$

$$\mathbf{f}^\dagger(\mathbf{x},\mathbf{x}_r;t)=\sum_{r=1}^{M}w_r(t)\left[\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;T-t)-\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;T-t)\right]\delta(\mathbf{x}-\mathbf{x}_r)\,, \qquad (3.18)$$

and the corresponding adjoint wavefield

$$\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_r;t)=\int_0^{t'}\int_V\mathbf{G}(\mathbf{x},\mathbf{x}';t'-t)\cdot\mathbf{f}^\dagger(\mathbf{x}',\mathbf{x}_r;t)\,d^3\mathbf{x}'\,dt\,, \qquad (3.19)$$

the perturbation of the misfit function may be expressed as:

$$\begin{aligned}\delta\chi(\mathbf{m}) &= -\sum_{s=1}^{N}\int_V\int_0^T\Big[\delta\rho\,\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_r;T-t)\cdot\partial_t^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\\ &+ \nabla\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_r;T-t):\delta\mathbf{c}:\nabla\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\Big]\,d^3\mathbf{x}\,dt\,. \qquad (3.20)\end{aligned}$$

At this point, we make some assumptions on the nature of the elasticity tensor. A general fourth-order elasticity tensor is described by 21 elastic parameters, a very large number that makes its complete characterization way beyond the reach of any tomographic approach. For the time being, let us thus consider isotropic elasticity tensors, described by the two Lamé parameters λ and μ:

$$c_{ijkl}=\lambda\delta_{ij}\delta_{kl}+\mu(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})\,. \qquad (3.21)$$

In this case, (3.20) can be written as:

$$\begin{aligned}\delta\chi(\mathbf{m}) &= -\sum_{s=1}^{N}\int_V\left[K_\rho(\mathbf{x},\mathbf{x}_s)\delta\ln\rho(\mathbf{x})+K_\lambda(\mathbf{x},\mathbf{x}_s)\delta\ln\lambda(\mathbf{x})\right.\\ &+ \left.K_\mu(\mathbf{x},\mathbf{x}_s)\delta\ln\mu(\mathbf{x})\right]\,d^3\mathbf{x}\,, \qquad (3.22)\end{aligned}$$

where ln() is the natural logarithm and where the Fréchet derivatives with respect to the density and Lamé parameters are given by:

$$K_\rho(\mathbf{x},\mathbf{x}_s) = -\int_0^T\rho(\mathbf{x})\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_r;T-t)\cdot\partial_t^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\,dt \qquad (3.23)$$

$$K_\lambda(\mathbf{x},\mathbf{x}_s) = -\int_0^T\lambda(\mathbf{x})\nabla\cdot\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_r;T-t)\nabla\cdot\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\,dt \qquad (3.24)$$

$$K_\mu(\mathbf{x},\mathbf{x}_s) = -2\int_0^T\mu(\mathbf{x})(\mathbf{x})\nabla\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_r;T-t):\nabla\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\,dt \qquad (3.25)$$

Since the propagation of seismic waves mainly depends on the variations of compressional wave velocity α and shear wave velocity β, but also because these seismic velocities are easier to interpret, tomographic models are usually described based on these two parameters. With this new parameterization, the perturbation of the misfit function may be written as:

$$\begin{aligned} \delta\chi(\mathbf{m}) &= -\sum_{s=1}^{N}\int_V \left[K'_\rho(\mathbf{x},\mathbf{x}_s)\delta\ln\rho(\mathbf{x}) + K'_\alpha(\mathbf{x},\mathbf{x}_s)\delta\ln\alpha(\mathbf{x})\right. \\ &+ \left. K'_\beta(\mathbf{x},\mathbf{x}_s)\delta\ln\beta(\mathbf{x})\right]\, d^3\mathbf{x}\,, \end{aligned} \tag{3.26}$$

where

$$K'_\rho(\mathbf{x},\mathbf{x}_s) = K_\rho(\mathbf{x},\mathbf{x}_s) + K_\lambda(\mathbf{x},\mathbf{x}_s) + K_\mu(\mathbf{x},\mathbf{x}_s) \tag{3.27}$$

$$K'_\alpha(\mathbf{x},\mathbf{x}_s) = 2\left(\frac{\lambda+2\mu}{\lambda}\right)K_\lambda(\mathbf{x},\mathbf{x}_s) \tag{3.28}$$

$$K'_\beta(\mathbf{x},\mathbf{x}_s) = 2K_\mu - \frac{4\mu}{\lambda}K_\lambda(\mathbf{x},\mathbf{x}_s) \tag{3.29}$$

As can be seen from these expressions, the principle of the adjoint-state method is to correlate two wavefields: the direct (i.e. forward) field that propagates from the source to the receivers, and the adjoint field that propagates from all the receivers backward in time. The same approach can be followed for any type of seismic observable (phase, amplitude, envelope, time series, etc.), provided the appropriate adjoint source is used (Tromp et al., 2005, 2008).For example, for the cross-correlated travel time of a seismic phase,the adjoint source is defined as the velocity of that synthetic phase weighted by the travel time residual.

Computing the gradient based on the adjoint state method requires performing two simulations per source (forward and adjoint fields) regardless of the type of observable. However, to define the adjoint field one must know the adjoint source, and that source is computed from the results of the forward simulation. One must therefore perform the forward simulation before the adjoint simulation. A straightforward solution for time-domain methods would be to store the whole forward field to disk at each time step during the forward run and then read it back during the adjoint simulation to calculate the interaction of these two fields. In 2D this is feasible but in the 3D case for very short seismic periods and without lossy compression, down sampling or a large amount of disk or memory check-pointing (Fichtner et al., 2009; Dalmau et al., 2014; Cyr et al., 2015) the amount of disk storage required would currently be too large. However, let us note again that this situation will change in the future. In the mean time, a standard possible solution is to perform three simulations per source (Tromp et al., 2008; Peter et al., 2011), that is, perform the forward calculation twice, once to compute the adjoint sources and once again at the same time as the adjoint simulation to correlate the two fields and sum their interaction on the fly over all the time steps. Doing so for an elastic Earth, one only needs

a small amount of disk storage to store the last time step of the forward run, which is then used as an initial condition to redo the forward run backwards, as well as the field on the outer edges of the mesh for each time steps in order to be able to undo the absorbing boundary conditions.

3.3 The regularization and grids projection

To stabilize the inversion it is necessary to regularize the problem. To do so, we add a classical penalty condition on the norm of the Laplacian of the model to the waveform misfit function and define a new misfit function:

$$\chi'(\mathbf{m}) = \chi(\mathbf{m}) + \frac{\lambda}{2} \parallel \Delta m \parallel^2 \tag{3.30}$$

where λ is a smoothing coefficient that weights the contribution of the norm of the Laplacian with respect to the waveform misfit. Denoting $L(\mathbf{m})$ the norm of the Laplacian:

$$L(\mathbf{m}) = \frac{1}{2} \parallel \Delta m \parallel^2 \tag{3.31}$$

the gradient of the new misfit function is then given by:

$$\nabla\chi'(\mathbf{m}) = \mathbf{g}(\mathbf{m}) + \lambda\nabla L(\mathbf{m}) \tag{3.32}$$

We have recalled in the previous section how to compute $\mathbf{g}(\mathbf{m})$ based on the adjoint method. The Laplacian of the model can be computed using a finite-difference operator in the regular 3-D Cartesian inversion grid:

$$\Delta m_{(i,j,k)} = -8m_{(i,j,k)} + m_{(i-1,j,k)} + m_{(i+1,j,k)} + \tag{3.33}$$

$$m_{(i,j-1,k)} + m_{(i,j+1,k)} + m_{(i,j,k-1)} + m_{(i,j,k+1)} \tag{3.34}$$

This formula can easily be adapted to the case of different smoothing values along the horizontal and vertical directions:

$$\begin{aligned} \Delta m_{(i,j,k)} = (-4l_h - 2l_v)m_{(i,j,k)} + l_h(m_{(i-1,j,k)} + m_{(i+1,j,k)} + \\ m_{(i,j-1,k)} + m_{(i,j+1,k)}) + l_v(m_{(i,j,k-1)} + m_{(i,j,k+1)}) \end{aligned} \tag{3.35}$$

where l_h and l_v are the smoothing coefficients along the horizontal and vertical directions, respectively. Let us denote Δ_d the discrete Laplacian operator in the inversion grid, which is described by a symmetric band-diagonal matrix. The norm of the discrete Laplacian can then be written as:

$$L(\mathbf{m}) = \frac{1}{2} \parallel \Delta_d\mathbf{m} \parallel^2 = \frac{1}{2} < \Delta_d\mathbf{m}|\Delta_d\mathbf{m} > \tag{3.36}$$

where <> denotes the scalar product. Using these definitions, the product:

$$\nabla L(\mathbf{m}).\delta m =< \Delta_d\mathbf{m}|\Delta_d\delta\mathbf{m} > \tag{3.37}$$

can be written as:

$$\nabla L(\mathbf{m}).\delta m =< \Delta_d^* \Delta_d \mathbf{m} | \delta \mathbf{m} >=< \Delta_d \Delta_d \mathbf{m} | \delta \mathbf{m} > \tag{3.38}$$

since the discrete Laplacian is self-adjoint. From this last expression we thus see that the gradient of the penalty function $L(\mathbf{m})$ is simply obtained by applying the discrete Laplacian to the model vector $\mathbf{m}$ twice:

$$\nabla L(\mathbf{m}) = \Delta_d \Delta_d \mathbf{m} \tag{3.39}$$

When the complex free surface topography is included in the SEM mesh, the SEM elements close to the free surface are relatively distorted (Figure 2.7). If we want to perform high-order spatial derivatives operation of the field quantities (such as the double Laplacian operation in regularization introduced above), it is more convenient to use a topologically regular tomographic grid than a unstructured SEM mesh. We distinguish the grid used to solve the wave propagation problem (the SEM grid) and the tomographic grid, which we choose to be regular.

To project the model defined on the SEM mesh onto the topologically regular tomographic grid, we search for the SEM elements that contain the vertices of each given tomographic cell. We then use the products of Lagrange polynomials (i.e. the SEM interpolation matrix) to interpolate the model parameters at the eight corners of this tomographic cell and assign to it the average of the values at these eight corners. We similarly compute the gradient from the correlation of the forward and adjoint wavefields, and for accuracy reasons we compute the volume integrals needed based on the SEM Gauss-Lobatto-Legendre quadrature formula. This requires interpolating the values of the correlations between the forward and adjoint wavefields not only at the vertices of each tomographic cell but also at all the collocation points located inside that cell. Once the model and the gradients for all the sources are projected onto the tomographic grid, we use the summed gradient to update the current tomographic model, which we then project back to the SEM mesh before proceeding to the next iteration. For efficiency, we once and for all define a mapping function that matches the SEM elements and their collocation point identifiers with the identifiers of the cell in the tomographic grid. Based on this mapping, the inverse projection becomes straightforward.

3.4 FWI with L-BFGS algorithm

Different classes of nonlinear optimization methods can be used to find the minimum of a cost function. Stochastic methods randomly explore the model space with no other information than the value of the cost function (Mosegaard & Tarantola, 1995; Sambridge, 1999a,b). Their main advantage is that they converge towards the global minimum of the cost function, but at the price of a much larger computational cost. In practice, these methods are thus efficient only

if the dimension of the parameter space to explore is small, which is not the case in 3D full waveform inversion problems.

Steepest descent methods search for a perturbation of the model along the direction defined by the gradient of the misfit function. Their main two drawbacks are that the inversion may converge towards a local minimum and remain trapped there, and that it may converge very slowly. They are therefore not recommended for very large problems in which the cost to estimate the misfit function and the gradient is important.

The Gauss-Newton method is the most efficient technique in terms of convergence speed. In this method, the search direction is obtained by preconditioning the gradient by the inverse Hessian (e.g. Pratt et al., 1998; Pratt, 1999). However, this method requires to first compute and store the Hessian, which requires a huge amount of computation and storage space, and then to solve a very large system of linear equations. In the case of 3D full waveform inversion, such an approach is currently completely prohibitive. For all these reasons, we thus choose to use the L-BFGS method (Nocedal & Wright, 2006) to handle the full waveform inversion problem. The main appeal of this method is that while it is affiliated to Gauss-Newton methods, it does not require computing and storing the Hessian (or its inverse) explicitly.

3.4.1 The L-BFGS algorithm

Using iterative methods, it is possible to compute an estimate of the inverse Hessian based only on the knowledge of the gradient at the previous iterations, the quality of the approximation improving with the number of previous iterations used. The method generates a series of models that gradually converge towards a minimum of the misfit function (which may be local) and a series of matrices that converge towards the inverse Hessian.

The BFGS formula to compute $\mathbf{H}_k^{-1}$, the approximate inverse Hessian at iteration k, is given by Nocedal & Wright (2006):

$$\mathbf{H}_k^{-1} \simeq \mathbf{H}_{k-1}^{-1} - \frac{\mathbf{H}_{k-1}^{-1} \cdot \mathbf{s}_{k-1} \otimes \mathbf{s}_{k-1} \cdot \mathbf{H}_{k-1}^{-1}}{\mathbf{s}_{k-1} \cdot \mathbf{H}_{k-1} \cdot \mathbf{s}_{k-1}} + \frac{\mathbf{y}_{k-1} \otimes \mathbf{y}_{k-1}}{\mathbf{y}_{k-1} \cdot \mathbf{s}_{k-1}} \tag{3.40}$$

where $\otimes$ is the tensor product, $\mathbf{s}_k = \mathbf{m}_k - \mathbf{m}_{k-1}$ is the difference between the current model and the model at the previous iteration, and $\mathbf{y}_k = \nabla\chi_k - \nabla\chi_{k-1}$ is the gradient change. Using (3.40) one can iteratively calculate an estimate of the inverse Hessian $\mathbf{H}^{-1}$ based on the knowledge of the approximation of the inverse Hessian after previous iteration, the difference $\mathbf{s}$ between the current model and its value at the previous iteration, and the difference $\mathbf{y}$ between the current gradient and the gradient at the previous iteration. Compared to the gradient method, the convergence of BFGS is much faster (for a mathematical proof see, e.g. Nocedal & Wright (2006)) for the same numerical cost. Convergence is about a factor of four faster in the examples that we will show below. Compared to the classical Gauss-Newton method, BFGS is also easier to implement because it does not require to compute and store the inverse Hessian.

The reader is referred to chapter 7 of Nocedal & Wright (2006) for a more detailed presentation of the BFGS algorithm.

To compute the search direction at iteration k

$$\mathbf{p}_k = -\mathbf{H}_k^{-1} \cdot \nabla\chi_k \,, \tag{3.41}$$

One only needs to perform a matrix vector multiplication. However, in the case of large problems it is currently impossible to compute and store even the approximate inverse Hessian matrix. Since in (3.41) one does not need to explicitly store it but only be able to compute its effect on a vector (the gradient), a modified method called the L-BFGS algorithm (for 'limited-memory BFGS') has been developed in order to compute the matrix vector product in (3.41) without having to store the inverse Hessian. The principle of L-BFGS is to use (3.40) iteratively to compute the product of the inverse Hessian, using the gradient from the initial inverse Hessian and the history of models and gradients accumulated in the iterations of the algorithm. In this case, one only needs to store a set of models and of gradients, which represents only a fraction of the storage space required to store the complete inverse Hessian. The number of previous models and gradients that are kept in memory is a parameter chosen by the user, called memory value l. In the following, we show the main algorithm of L-BFGS two-loop recursion.

Given the initial model $\mathbf{m}_0$, memory value l and convergence condition, the L-BFGS algorithm can be stated formally as follows:

Initialization: Evaluate $\chi_0 = \chi(\mathbf{m}_0)$, $\mathbf{y}_0 = \nabla\chi(\mathbf{m}_0)$.

repeat:

Choose initial inverse Hessian matrix $\mathbf{H}_{\mathbf{k_0}}^{-\mathbf{1}}$.

Compute current descent direction $\mathbf{p}_k$ (3.41) from the recursion algorithm shown below.

Update new model $\mathbf{m}_{k+1} = \mathbf{m}_k + s\mathbf{p}_k$, where step length s is chosen to satisfy proper line search condition.

If $k > l$, discard the vector pairs prior to $(\mathbf{m}_k, \mathbf{y}_k)$ from storage or memory.

Set $k = k + 1$, compute and save $\mathbf{s}_{k+1} = \mathbf{m}_{k+1} - \mathbf{m}_k$ and $\mathbf{y}_{k+1} = \nabla\chi_{k+1} - \nabla\chi_k$.

until convergence.

Recursion:

Set $\mathbf{q} = \nabla\chi_k$, $j = min(k, l)$.

for $i = k-1, k-2, ..., k-j$

$\lambda_i = \frac{\mathbf{s}_i^T \mathbf{q}}{\mathbf{s}_i^T \mathbf{y}_i}$

$\mathbf{q} = \mathbf{q} - \lambda_i \mathbf{y}_i$

end

set $\mathbf{r} = \mathbf{H}_{\mathbf{k_0}}^{-\mathbf{1}}\mathbf{q}$

for $i = k-j, k-j+1, ..., k-1$

$\beta_i = \frac{\mathbf{y}_i^T \mathbf{r}}{\mathbf{s}_i^T \mathbf{y}_i}$

$\mathbf{r} = \mathbf{r} + (\lambda_i - \beta)\mathbf{s}_i$

end

stop with result $\mathbf{p}_k = -\mathbf{H}_\mathbf{k}^{-1}\nabla\chi_\mathbf{k}$.

Except for being efficient, this recursion algorithm has an advantage that the multiplication by the initial inverse Hessian $\mathbf{H}_{\mathbf{k_0}}^{-1}\mathbf{q}$ can be isolated from the rest of the computations, allowing this matrix to be chosen freely and be able to vary between iterations. During the first $l-1$ iterations, above L-BFGS algorithm is equivalent to the standard BFGS algorithm if the initial inverse Hessian matrix $\mathbf{H_0^{-1}}$ is the same in both methods, and if L-BFGS chooses $\mathbf{H}_{\mathbf{k_0}}^{-1} = \mathbf{H_0^{-1}}$ at each iteration.

A method for choosing $\mathbf{H}_{\mathbf{k_0}}^{-1}$ that has been proved effective in practice is to set $\mathbf{H}_{\mathbf{k_0}}^{-1} = \gamma_\mathbf{k}\mathbf{I}$, where

$$\gamma_k = \frac{\mathbf{y}_k^T\mathbf{s}_k}{\mathbf{y}_k^T\mathbf{y}_k} \tag{3.42}$$

The scaling factor γ_k, which accounts for differences between the true Hessian and the approximation Hessian. This choice helps to ensure that the search direction $\mathbf{p}_k$ is well scaled, and is essential to the good performance of the algorithm. As a generalization, this initial inverse Hessian can also be chosen as the multiplication of γ_k and a diagonal scaling matrix $\mathbf{D}$, in connection with the preconditioning conception in the nonlinear algorithm. We will present below some approaches to defining it for our FWI applications.

3.4.2 Calculation of the step length

Once the descent direction $\mathbf{p}_k$ at iteration k has been obtained, it is necessary to determine the step length, or in other words to decide how far to move along that direction. This problem can be formulated as finding the step s that minimizes

$$\phi(s) = \chi(\mathbf{m}_k + s\mathbf{p}_k)\,. \tag{3.43}$$

In practice, determining that optimal step precisely may require to test a large number of step lengths, which can thus be very expensive. However, one should keep in mind that $\chi(\mathbf{m})$ rather than $\phi(s)$ is the quantity that we need to minimize. It is thus sufficient to find an approximate step at minimal cost that honors certain conditions in order to make the optimization method converge. In practice, the step length variations between two iterations must be sufficiently large so that the algorithm requires a moderate number of iterations to converge, and sufficiently small to avoid the divergence of the algorithm. A good compromise is to use the so-called Wolfe conditions to select the step length (Nocedal & Wright, 2006). These rules test if the current step provides a sufficient decrease of both the cost function and the gradient. Introducing parameters $0 < c_1 < c_2 < 1$, and $\phi'(s)$ the derivative of ϕ with respect to s, the step length is kept if:

$$\phi(s) \leq \phi(0) + c_1 s\phi'(0) \tag{3.44}$$

$$|\phi'(s)| \leq c_2|\phi'(0)| \tag{3.45}$$

If these two conditions are not satisfied, a new step is tested. If

$$\phi(s) > \phi(0) + c_1 s\phi'(0) \tag{3.46}$$

the step is too long, and one then tests a smaller step length. On the other hand, if

$$\phi(s) \leq \phi(0) + c_1 s\phi'(0) \tag{3.47}$$

$$|\phi'(s)| > c_2|\phi'(0)| \tag{3.48}$$

the step is too short, and one then tests a longer step. When it is no longer possible to find a step that satisfies these relations, convergence is reached and one then stops the algorithm. Tuning parameters c_1 and c_2 makes the selection rules more or less restrictive in terms of accepting the step length. For example, if c_1 is chosen close to 0, it is easier to honour the first inequality. In our implementation, we select $c_1 = 0.1$ and $c_2 = 0.9$, the standard values recommended by Nocedal & Wright (2006).

3.5 Test of the method on a checkerboard model

In order to test the inversion method described in the previous sections, we consider a simple ideal (in terms of ray coverage) synthetic checkerboard test. The checkerboard model is composed of 20 km cube-shaped velocity anomalies of ± 0.5 km/s embedded in a homogeneous model in which $V_p = 8$ km/s , $V_s = 4.5$ km/s and $\rho = 3000$ kg/m^3 . We perform the computations in a 100 km $\times$ 100 km $\times$ 60 km regular Cartesian grid discretized with 2 km cubic elements. The model is illuminated by 20 different plane-wave sources coming from the south, east, north and west directions and with incidence angles of 20, 30, 45, 60 and 80°. These plane waves are recorded by an array of 2116 stations uniformly distributed at the free surface with an inter-station spacing of 2 km. The source wavelet is a Gaussian with a dominant period of 1.25 s, and we invert the vertical component of P waveforms in time windows that start 5 s before the P-wave arrival and end 10 s after. The starting model in the iterative algorithm is the homogeneous background model. In the beginning of the inversion, we invert the waveforms low-pass filtered at 0.2 Hz and gradually increase the corner frequency of the filter up to 2 Hz. Moving from low to high frequencies helps the inversion to converge towards a model close to the true model (Figure 3.1) without getting trapped in a secondary minimum of the misfit function (e.g. Pratt, 1999; Pageot et al., 2013).

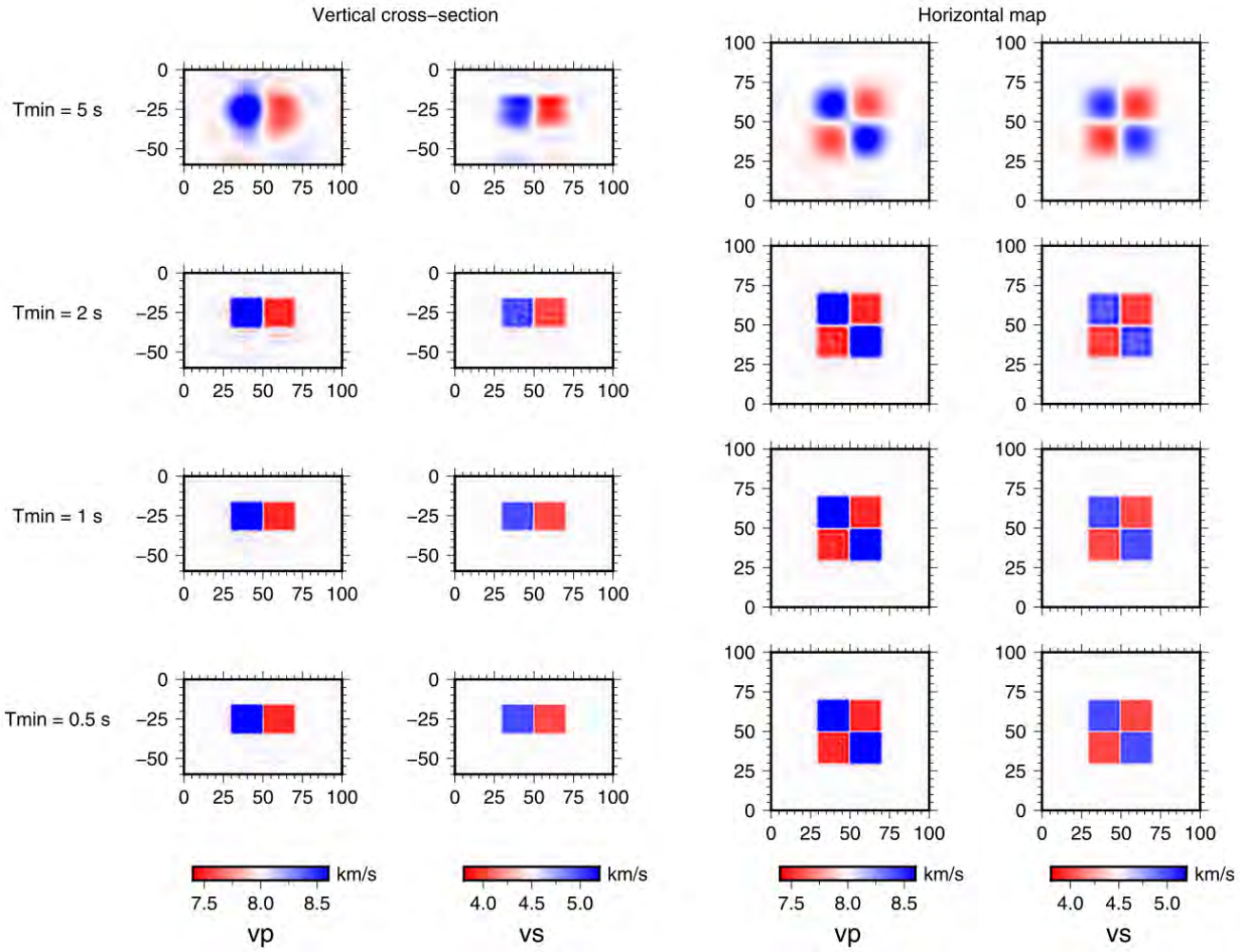


Figure 3.1: Results of full waveform inversion for a checkerboard test. The plots show vertical cross-sections (left) and map views (right) of the Vp and Vs models obtained with a hierarchical algorithm starting with data low-pass filtered at 5 s (top) and then gradually decreasing the corner period of the filter to 0.5 s (bottom). We used 20 plane waves with 5 coming from the north, 5 from the south, 5 from the east and 5 from the west, with incidence angles of 20, 30, 45 and 80° for each of the four waves in each case. The misfit function decreases from 13.16 to 0.14 in 18 iterations for the data filtered at 5 s, from 6.45 to 0.17 in 62 iterations for the data filtered at 2 s, from 1.99 to 0.54 in 76 iterations for the data filtered at 1 s and from 1.8 to 1.05 in 63 iterations for the data filtered at 0.5 s.

Chapter 4

Applications of Full waveform inversion to the western Pyrenees transect

Contents

In this chapter, we will present the results of our FWI approach applied on the data of the western PYROPE transect.

4.1 Data selection and preparation

4.1.1 Initial data inspection

We assume that our teleseismic sources can be represented by a point source. Each earthquake source is described by 10 parameters: the origin time, coordinates of the hypocenter (longitude, latitude and focal depth), and moment tensor (6 parameters, describing the second order symmetric moment tensor). Some studies inverted and updated the source model during the iterative waveform inversion process (Tape et al., 2010). However, in our study, since the construction of the database for hybrid simulation containing the tractions and velocities produced by each teleseismic source suffers from a relatively high computational cost, the source parameters are fixed throughout the regional inversion process. When 3D regional model is updated, only the source wavelet function can be updated.

Table 4.1 shows the source parameters of the five events which are used in our inversion, taken from USGS Centroid Moment Tensor. These events were recorded by the total number

of 45 stations (29 temporary broad-band stations along the western PYROPE transect, 7 PYROPE broad-band stations and 9 permanent French and Spanish stations). The five events are in a distance range between 50° and 89° and occurred during the deployment of the western transect in 2013, with magnitudes between Mw 6.1 and Mw 8.3. The focal depths are located between 10 and 610 km. The azimuthal coverage for the western transect is shown in Figure 4.2.

Source parameters	2013 May 11	2013 May 24	2013 Aug 13	2013 Aug 30	2013 Sep 25
Longitude(°)	57.77	153.28	-78.20	-175.23	-74.51
Latitude(°)	26.56	54.87	5.77	51.54	-15.84
focal depth(km)	15.0	609.0	12.0	29.0	40.0
Magnitude (M_w)	6.1	8.3	6.7	7.0	7.1
exponent($N{\cdot}M$)	18	21	18	19	19
M_{rr}	-0.03	-1.11	-1.94	1.70	3.61
M_{tt}	-1.70	0.25	7.25	-1.66	-2.25
M_{pp}	1.73	0.87	-5.31	-0.04	-1.36
M_{rt}	0.33	-0.91	-0.98	1.62	1.54
M_{rp}	-0.06	-2.91	2.85	0.63	-1.54
M_{tp}	3.29	0.12	5.25	-0.32	2.65

Table 4.1: The CMT solutions of the five events used for the western transect

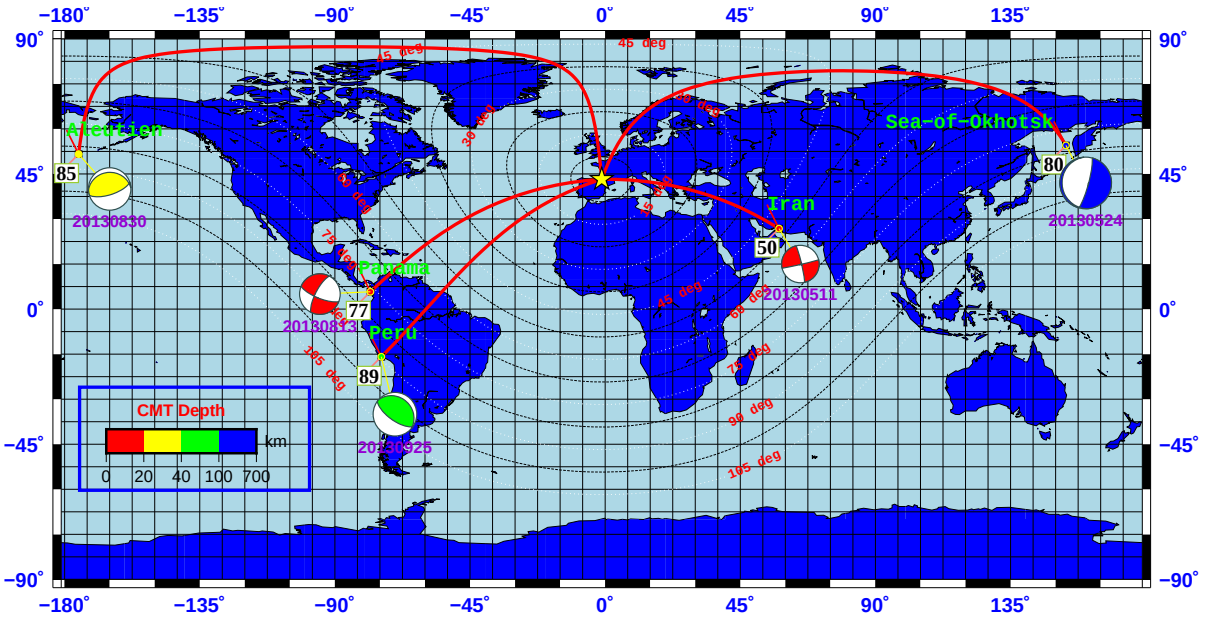


Figure 4.2: The Azimuthal coverage of 5 events used in FWI of the western transect.

The initial three component velocity waveform data are cut from 40 s before to 110 s after the theoretical P arrival time and rotated to vertical, radial and transverse components. The records are integrated to obtain displacement waveforms. The source wavelets are then estimated

from the vertical component displacements. Hereafter, all the figures will show displacement waveforms. After the iterative wavelet estimation, we have visually selected high quality traces based on the correlation coefficient between observed and synthetic vertical components. We only keep for inversion the traces which have a correlation coefficient larger than 0.85 to 0.9 with the synthetic seismograms, depending on the specific event. A trace with a poor correlation coefficient usually reflects a low S/N ratio on the record, and it should be abandoned. After this initial selection, 186 traces collected from 5 teleseismic events are kept for the western transect.

4.1.2 Selection of the frequency band

Our full waveform inversion exploits simultaneously the vertical and radial component data waveforms. The exploitation of transverse component will be considered later when seismic anisotropy will be introduced in the inversion on real data.

The fit between observed and synthetic waveforms significantly relies on the frequency content of the seismic wavefields. The resolution and quality of tomographic images are thus dependent on the frequency band of inverted waveforms. In the case of low S/N ratio, the useful seismic signals are masked by seismic noise. Therefore, the critical problems in full waveform inversion are how to choose the appropriate frequency range to improve the S/N ratio. When the frequency spectrum of the useful seismic signal differs from the frequency spectrum of the background seismic noise, performing a simple band-pass filtering can improve the S/N ratio.

In general, the transient seismic signals excited by localized earthquake sources are coherent. In contrast, the ambient noise is a more or less stationary stochastic process, caused by spatially randomly distributed incoherent noise sources. Because of this stochastic nature of seismic noise, we can compute the coherence between synthetic and observed seismograms at each station. We use the frequency-dependent coherence C_{xy} which takes values between 0 and 1. This function measures how well a time series $x(t)$ correlates to another time series $y(t)$ at each frequency. It is defined by using the individual power spectral densities $P_{xx}(f)$ and $P_{yy}(f)$ of x and y and their cross power spectral density $P_{xy}(f)$:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f) \cdot P_{yy}(f)} \tag{4.1}$$

In practice, the magnitude squared coherence should be computed by using some kinds of time-frequency ensemble averaging methods. In our study, we choose to use the multitaper method (MTM) to compute $C_{xy}(f)$. The MTM (Thomson, 1982) attempts to reduce the variance of spectral estimates by using a small set of tapers rather than the single data taper or spectral window introduced in general Blackman-Tukey methods. A set of independent estimates of the power spectrum is obtained by pre-multiplying the data by orthogonal tapers which are designed to minimize the spectral leakage as a result of the finite length of the data set. The only control parameter for MTM is the time-half-bandwidth product (NW, where

N is the length of the time series and $2W$ is a chosen bandwidth). It is used to compute the number of windows $K = 2NW - 1$ for spectral estimation. As NW increases, the resolution in the time domain gets higher but the spectral leakage also increases. The choice of NW thus results from a compromise between bias and resolution of the spectrum estimate.

Here we just briefly review the main steps in MTM coherence estimate. For more details we refer the reader to Thomson (1982) and Park et al. (1987). First, a set of orthogonal eigentapers $w_k(t)$ with associated eigenvalues λ_k are constructed for a chosen time-half-bandwidth product NW (Park et al., 1987). These eigentapers which have the same length as the time series are called discrete prolate spheroidal sequences (DPSS), or Slepian sequences (Slepian, 1978). Second, the k_{th} eigenspectrum estimate $X_k(f)$ is obtained by computing the DFT of the windowed time series $x(t)w_k(t)$. The power spectrum $P_{xx}(f)$ is then estimated by a weighted average of the individual spectra $X_k(f)$:

$$P_{xx}(f) = \frac{\sum_{k=1}^{K} \mu_k |X_k(f)|^2}{\sum_{k=1}^{K} \mu_k} \tag{4.2}$$

where the weights μ_k are based on the eigenvalues λ_k of the DPSS solution (Thomson, 1982). Finally, $C_{xy}(f)$ is computed by using the individual DFT spectrum estimates $X_k(f)$ and $Y_k(f)$ of time series $x(t)$ and $y(t)$ tapered by k_{th} taper $w_k(t)$ according to:

$$C_{xy}(f) = \frac{|\sum_{k=1}^{K} \mu_k X_k(f) \cdot Y_k^*(f)|^2}{\sum_{k=1}^{K} \mu_k |X_k(f)|^2 \cdot \sum_{k=1}^{K} \mu_k |Y_k(f)|^2} \tag{4.3}$$

The time-half-bandwidth product NW in our coherence computation is set to 3 after testing different values 2.5, 3, 3.5, 4, 4.5 and 5, which provides a good trade-off between the required frequency resolution and estimation bias. Based on the MTM coherence, we can define a data-adaptive coherence filter to remove the ambient noise.

Even though the background noise level was generally high during the western transect deployment, the S/N ratio was significantly higher than that of the central transect. This may result from the deeper focal depths and larger amplitudes of recorded earthquakes. We describe below our strategy to determine the optimum corner frequency and data selection for western transect events.

First of all, we search for a corner frequency to low pass filter the data at around 5 s. We set trial initial corner frequencies at 0.01 and 0.1 Hz for all the vertical component traces and change the lower corner frequency to 0.035 Hz for radial component because of its relative lower S/N ratio. An eight-order Butterworth filter is used to filter the waveforms. After filtering, we compute the correlation coefficient between filtered data and synthetic waveforms in a time window starting about 10 s before the theoretical arrival time of P wave and ending about 40 s after. We preset the threshold value of correlation coefficient to 0.88 for vertical component and 0.7 for radial component. We also compute the MTM coherence between observed and synthetic waveforms for all traces.

The first event for the western transect is the 2013 May 11 Southern Iran earthquake. There are 6 traces with vertical component correlation coefficients lower than the threshold value with a filter using the preset corner frequencies 0.01 and 0.1 Hz (Figure 4.3). The light grey shaded area shown in the left panel is the time window used for computing the correlation coefficient for this event. The observed waveform amplitudes are not normalized at the moment. The amplitude has no effect on the estimates of coherence and correlation since these two computed quantities are normalized. The right panel of Figure 4.3 shows the corresponding coherence for each trace. The pink area corresponds to coherence larger than 0.7 for vertical component. There is a common valley at around 0.1 Hz where coherence is low (shown in light blue color). Figure 4.4 shows the vertical component waveforms and corresponding coherence for the other traces with higher correlation coefficients. This valley of low coherence region is also observed. Not only the vertical components but also the coherence functions for radial components show the same features, which will be shown next. This low coherence mainly reflects the strong influence of secondary microseisms in the Atlantic Ocean and the Mediterranean Sea (Chevrot et al., 2007).

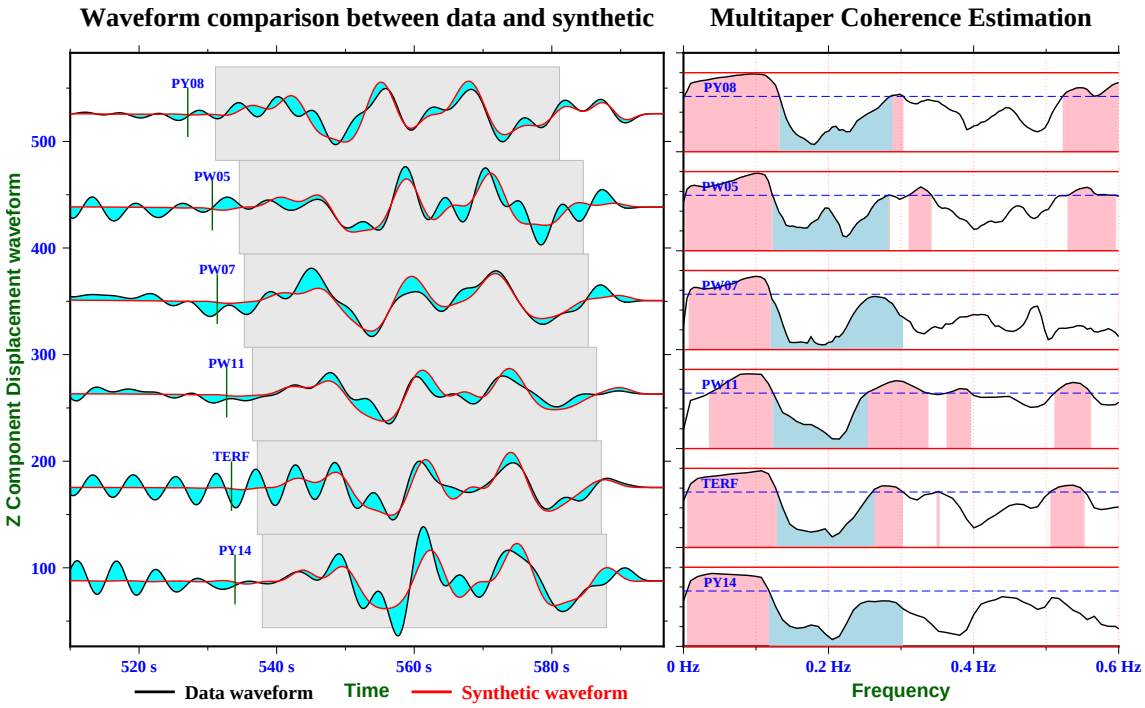


Figure 4.3: Vertical component waveforms and corresponding coherence estimate for the traces of the 2013 May 11 Southern Iran earthquake with low correlation coefficients.

The microseisms are typically caused by ocean waves on coasts. In general, noise spectrum show a smaller amplitude primary ocean microseismic peak at periods around $14 \pm 2s$ and a secondary ocean microseismic peak around 6 s. The inland stations may record both of them. The primary ocean microseisms are generated in the shallow waters close to the coast, where

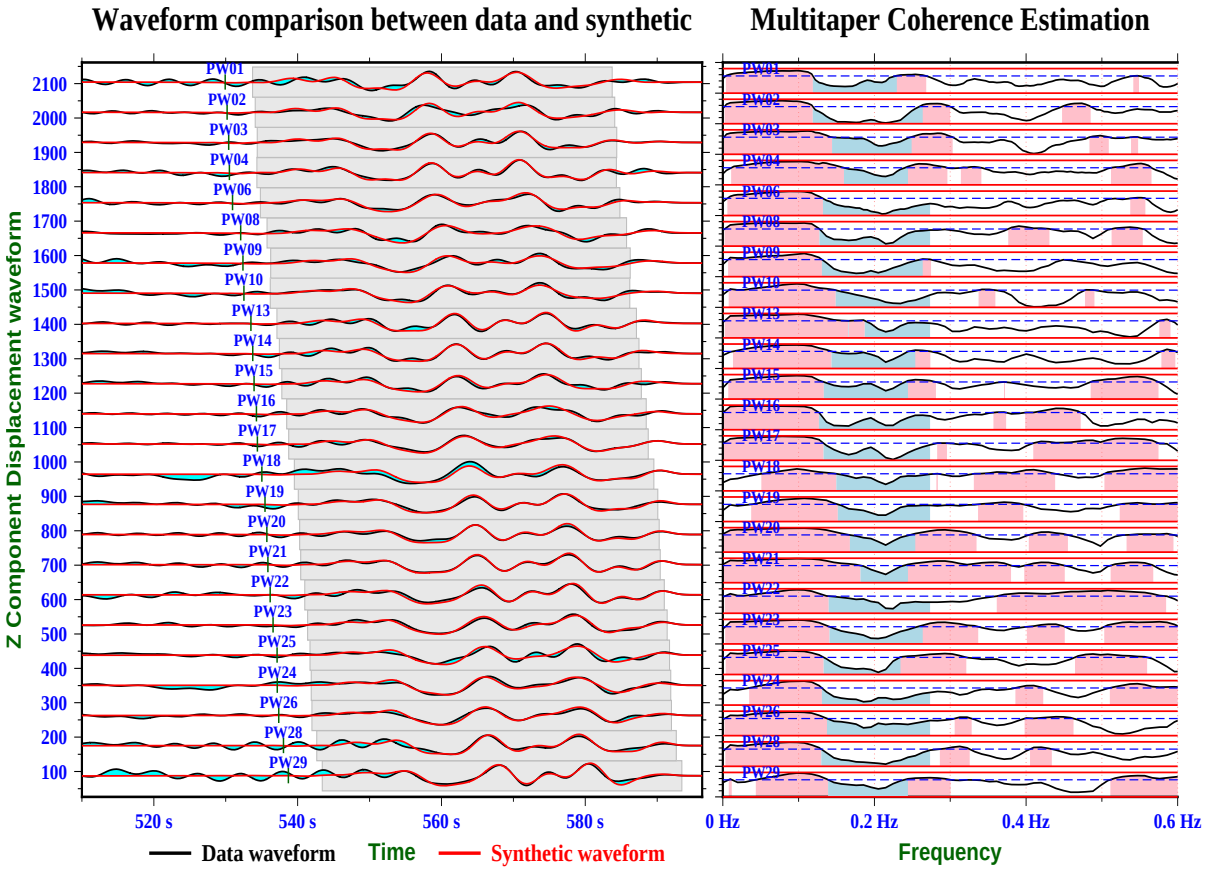


Figure 4.4: Vertical component waveforms and corresponding coherence estimate for the traces of the 2013 May 11 event with good correlation coefficients.

the ocean gravity wave energy can be converted directly into seismic energy either by vertical pressure variations or by the breaking of waves on the shore. The primary ocean microseisms generally have the same characteristic period as the water waves (0.05 to 0.1 Hz). The secondary or double frequency microseisms (DFM) produce a much stronger peak in a frequency range between 0.1 and 0.3 Hz, usually at a frequency which is twice the peak frequency of primary microseisms. The mechanisms for DFM are complex. The main explanation is that DFM is produced by standing waves in the ocean. Because the western transect was closer to the Bay of Biscay, near the major sources of microseismic noise located in the northern Galicia margin and the western coast of Ireland (Chevrot et al., 2007), DFM controlled the background noise level in the seismological records.

The coherence can guide us to find the proper corner frequencies for filtering. We search the lower and upper corner frequencies fs and fe based on coherence function. For example, we set the search range for fs at 0.01 and 0.07 Hz and the range for fe at between 0.15 and 0.23 Hz for station PY14. The objective of this 2D search is to maximize the correlation coefficient over a selected time window. Figure 4.5 shows the results for this station. Panel (a) shows the comparison for the unfiltered waveform, and panel (b) shows the comparison for the waveform

filtered at the initial preset corner frequencies 0.01 Hz and 0.2 Hz. We can see apparent noise ripples with a dominant period around 6 s in the filtered data since the upper corner frequency is set to 0.2 Hz. This quasi-monochromatic noise is the DFM that remains unfiltered with this frequency range. Panel (d) shows the correlation coefficient versus fs and fe. The dark red shows the maximum of the correlation coefficient obtained with $fs = 0.05$ Hz and $fe = 0.11$ Hz. The frequencies fs and fe that lead to a maximum correlation coefficient give the optimal frequency range for filtering the data. However, it is desirable to keep the larger frequency range to retain as much information as possible for the inversion. We set a new slightly smaller bound (0.9 of the maximum correlation coefficient) and search for the possible widest frequency bandwidth defined by $\Delta f = fe - fs$. By executing the 2D search once again, the new corner frequencies fs and fe which keep a relatively high correlation coefficient are 0.042 and 0.131 Hz, respectively. Panel (c) shows the comparison for the waveform filtered at the new corner frequencies 0.042 and 0.131 Hz. In this panel, the amplitudes of waveforms are normalized for comparing the similarity between observed and synthetic waveforms. The better waveform similarity after new filtering is obtained. Panel (e) and (f) show the coherence with pink shaded area bounded by the two sets of corner frequencies.

The radial components are always noisier than vertical components. A lower threshold of 0.7 is used for radial components, and there are 19 traces with radial component correlation coefficients lower than this threshold filtered between the preset frequencies 0.035 and 0.1 Hz shown in Figure 4.6. The radial component waveforms and corresponding coherence for traces with good correlation coefficients are also shown in Figure 4.7. Since the amplitudes of observed waveforms are not normalized, we can see that a large amplitude long period noise still needs to be filtered out. The coherence functions plotted in the right panels of Figure 4.6 and 4.7 show the presence of strong DFM noise around 6 s. As a result, many stations have very bad coherence for their radial component (for example stations PW03, PW07, PW10 etc). We thus have to discard these traces. The station PW08 has the lowest correlation coefficient (0.583) but it is nevertheless kept for inversion after optimal filtering based on coherence function. Figure 4.8 shows the 2D search procedure for this station.

The vertical components for the other 4 events selected for the western transect all show a very good coherence between data and synthetics, with the correlation coefficients above the threshold value of 0.88. For these events, using the simple uniform corner frequencies $fs = 0.01$ Hz and $fs = 0.1$ Hz, the inversion is expected to perform well. One possible reason for this larger coherence is the larger amplitudes of these 4 earthquakes (all with $M_w > 6.7$, compared to the M_w 6.1 for the 2013 May 11 Iran earthquake) which have P waves well above the background noise level in spite of a relatively large DFM microseismic noise level.

The radial component waveforms for events May 24, Aug 30 and Sep 25 are clean and most traces for these 3 events are kept. The 2013 Aug 13 Panama earthquake has fewer good quality radial component traces. Figure 4.9 shows the comparison between observed and synthetic

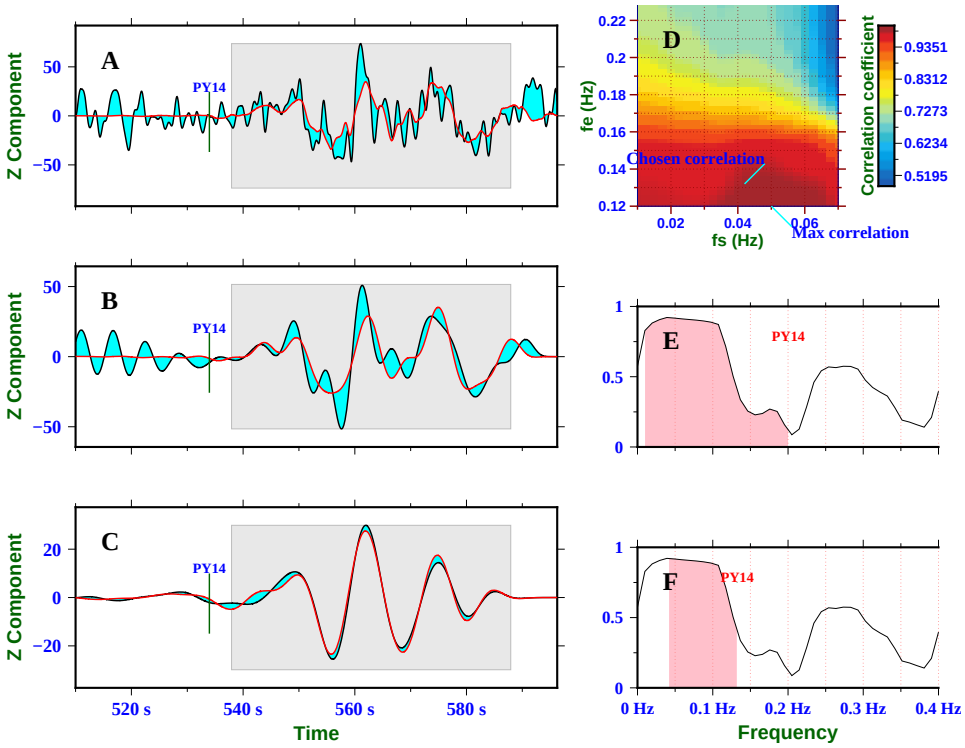


Figure 4.5: 2D corner frequency search for vertical component trace of station PY14 for the 2013 May 11 earthquake. Panel (a) shows the unfiltered waveform, (b) for the waveform filtered at the initial preset corner frequencies 0.01 Hz and 0.2 Hz, and (c) for the waveform filtered at the new corner frequencies 0.042 and 0.131 Hz. Panel (d) shows the 2D objective function Vs. fs and fe. The dark red shows the maximum of correlation coefficient obtained with $fs = 0.05$ Hz and $fe = 0.12$ Hz. This panel also shows that the optimal corner frequencies fs and fe which keep a relatively high correlation coefficient are 0.042 and 0.131 Hz, respectively. Panel (e) and (f) show the coherence for this trace, with pink shaded area bounded by the two sets of corner frequencies corresponding to (b) and (c).

waveforms for traces with lower correlation coefficients, as well as their coherence. From Figure 4.9, it seems that the common incoherent frequency band is located around 0.05 Hz, near the frequency range of primary ocean microseisms. Station PW16 is kept for inversion after the 2D optimal corner frequency search. Figure 4.10 shows the 2D search procedure for this station. The lower corner frequency is increased to 0.066 Hz, which is used for filtering out this large amplitude 20 s long period noise.

Once the optimal frequency band has been determined for all the traces, we can obtain travel time and amplitude anomaly maps for each event. The amplitude anomaly is defined in section 2.3.5.4. The Figure 4.11 and 4.12 show the maps of travel time and amplitude anomalies for the vertical and radial components of the 2013 May 11 event. These anomalies are measured on waveforms filtered in the corresponding optimal corner frequency.

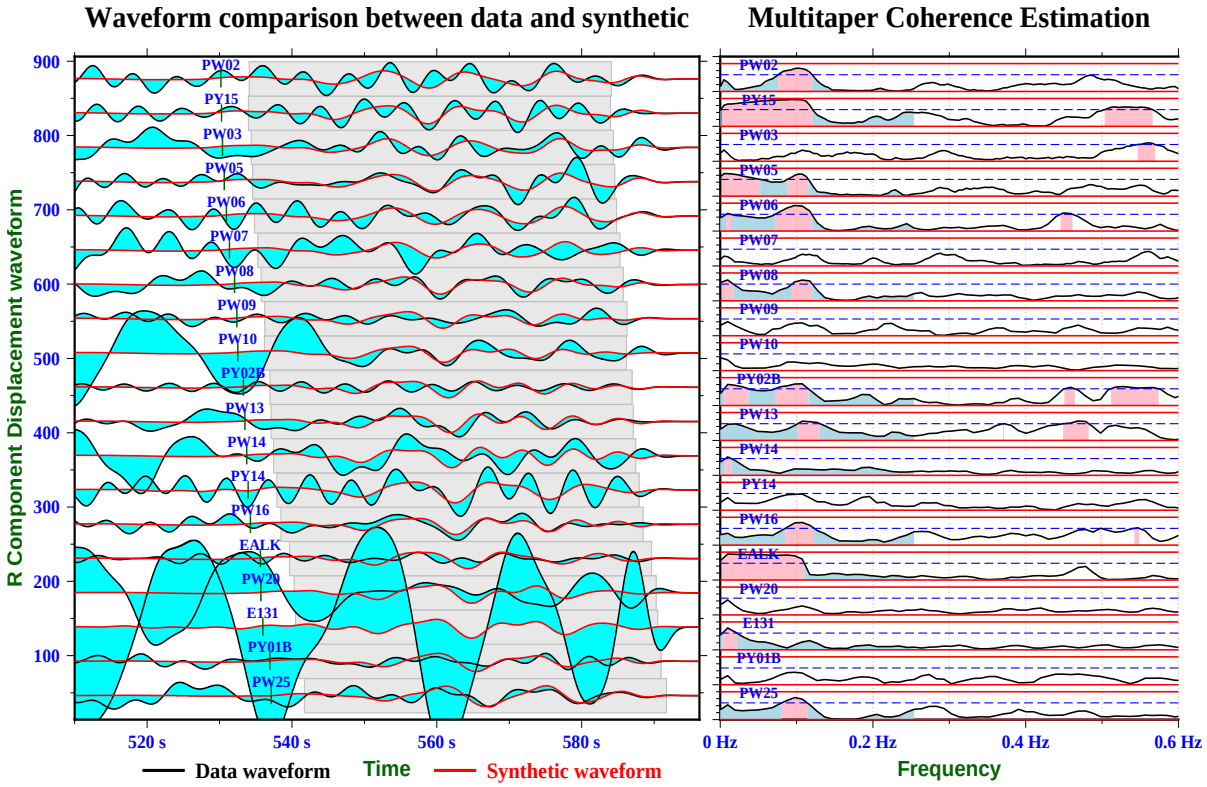


Figure 4.6: Radial component waveforms and corresponding coherence estimate for the traces of the 2013 May 11 Southern Iran earthquake with low correlation coefficients.

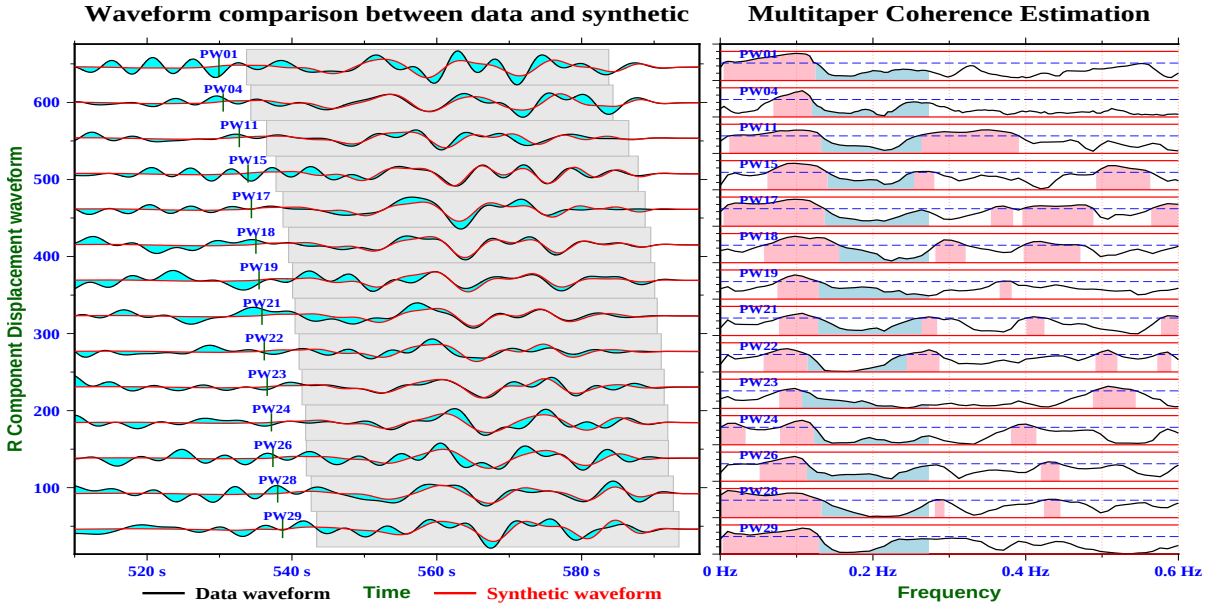


Figure 4.7: Radial component waveforms and corresponding coherence estimate for the traces of the 2013 May 11 event with good correlation coefficients.

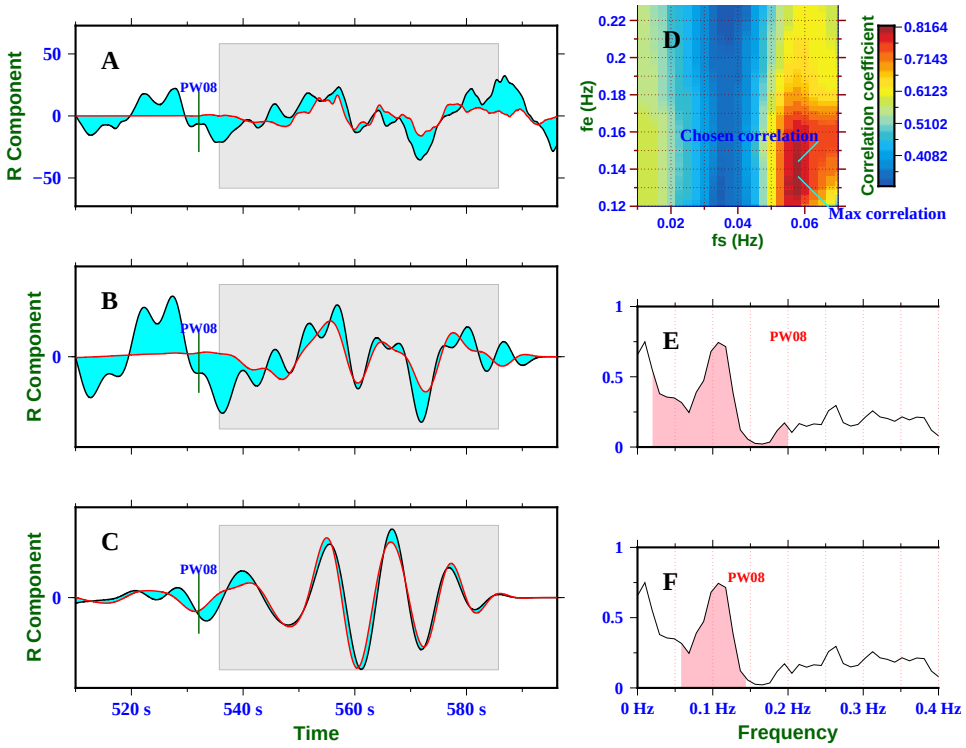


Figure 4.8: 2D corner frequency search for the radial component of station PW08 for the 2013 May 11 earthquake. Panel (a) shows the unfiltered waveform, (b) for the waveform filtered at the initial preset corner frequencies 0.035 Hz and 0.2 Hz, and (c) for the waveform filtered at the new corner frequencies 0.058 and 0.144 Hz. Panel (d) shows the 2D objective function Vs. fs and fe. The maximum of correlation coefficient 0.81 is obtained with $fs = 0.058$ Hz and $fe = 0.136$ Hz. This panel also shows that the optimal corner frequencies fs and fe which keep a relatively high correlation coefficient 0.8 are 0.058 and 0.144 Hz, respectively. Panel (e) and (f) show the coherence for this trace, with pink shaded area bounded by the two sets of corner frequencies corresponding to (b) and (c).

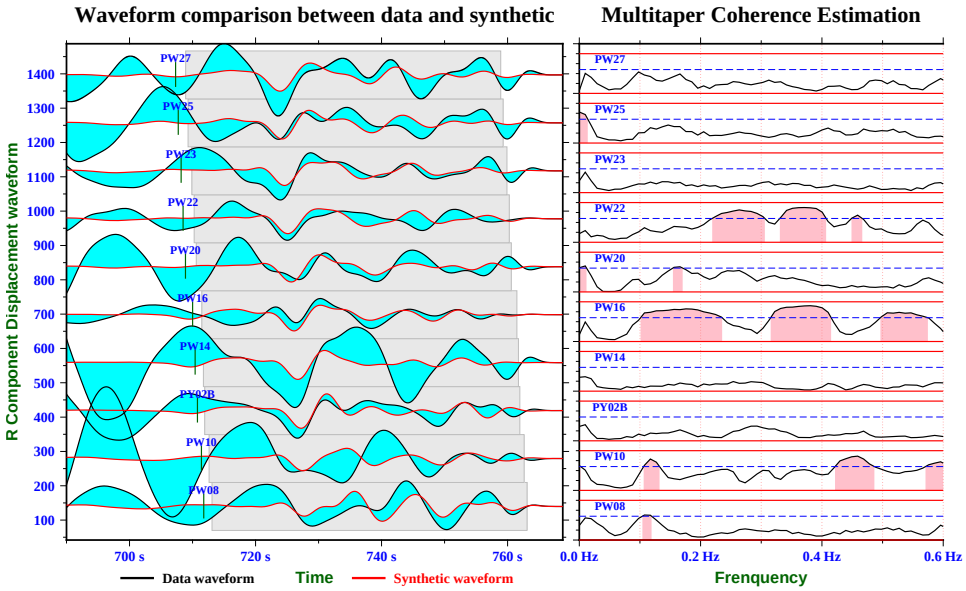


Figure 4.9: Radial component waveforms and corresponding coherence estimate for the traces of the 2013 Aug 13 Panama earthquake with low correlation coefficients.

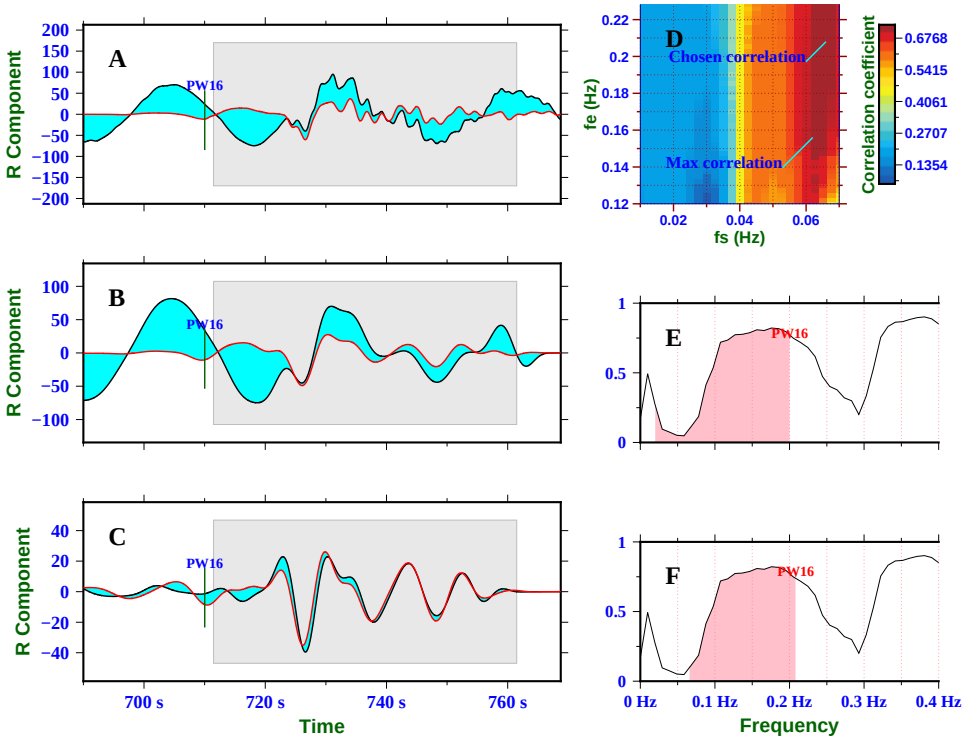
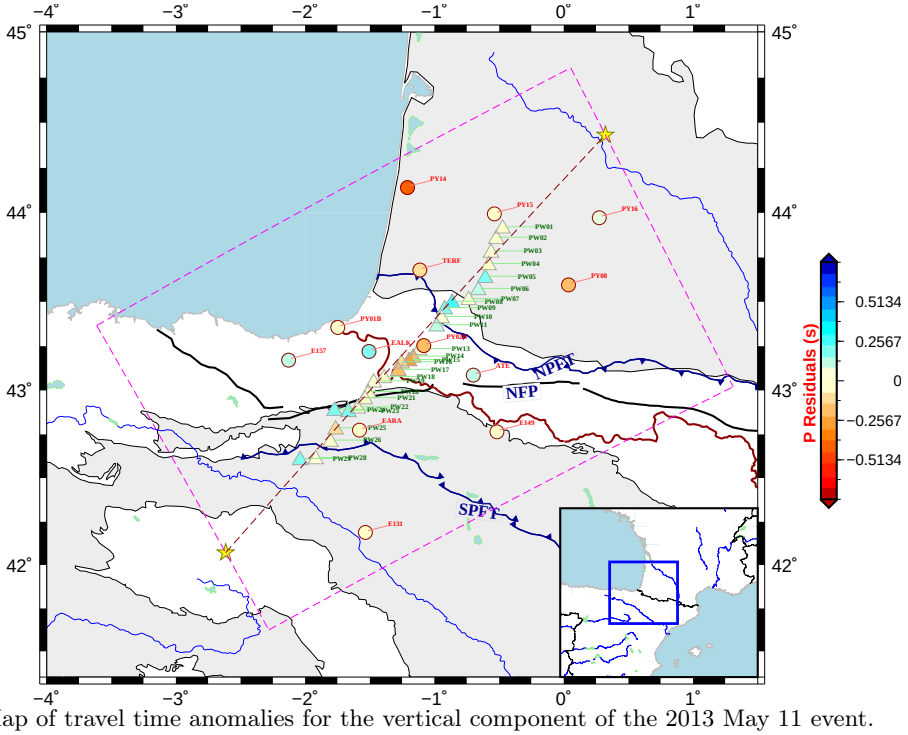
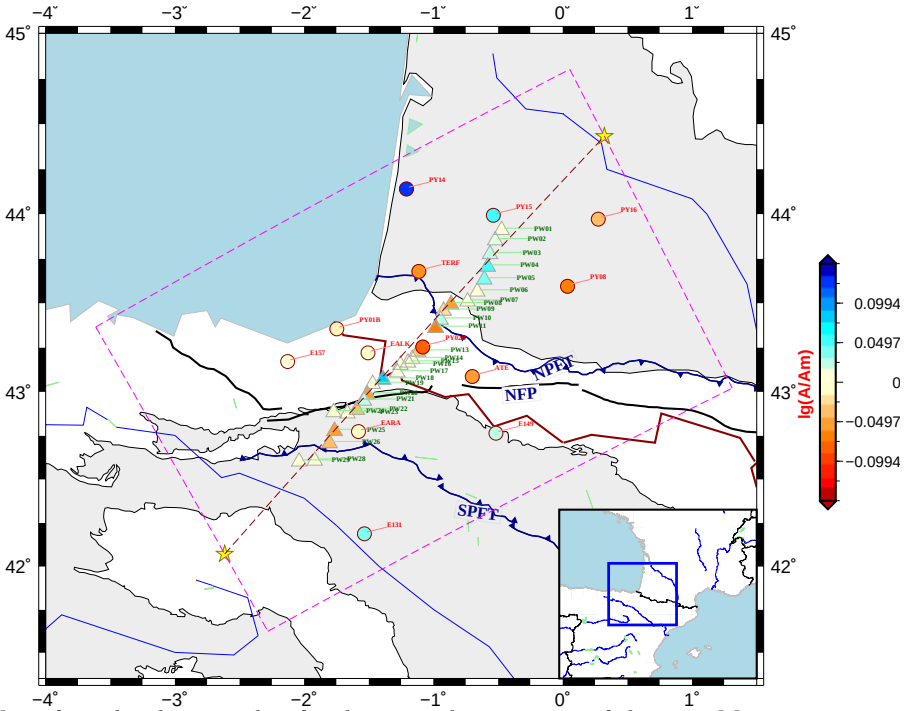


Figure 4.10: 2D corner frequency search for the radial component of station PW16 for the 2013 Aug 13 Panama earthquake. Panel (a) shows the unfiltered waveform, (b) for the waveform filtered at the initial preset corner frequencies 0.035 Hz and 0.2 Hz, and (c) for the waveform filtered at the new corner frequencies 0.066 and 0.21 Hz. Panel (d) shows the 2D objective function Vs. fs and fe. The maximum of correlation coefficient 0.72 is obtained with $fs = 0.061$ Hz and $fe = 0.156$ Hz. This panel also shows that the optimal corner frequencies fs and fe which keep a relatively high correlation coefficient 0.71 are 0.066 and 0.21 Hz, respectively.

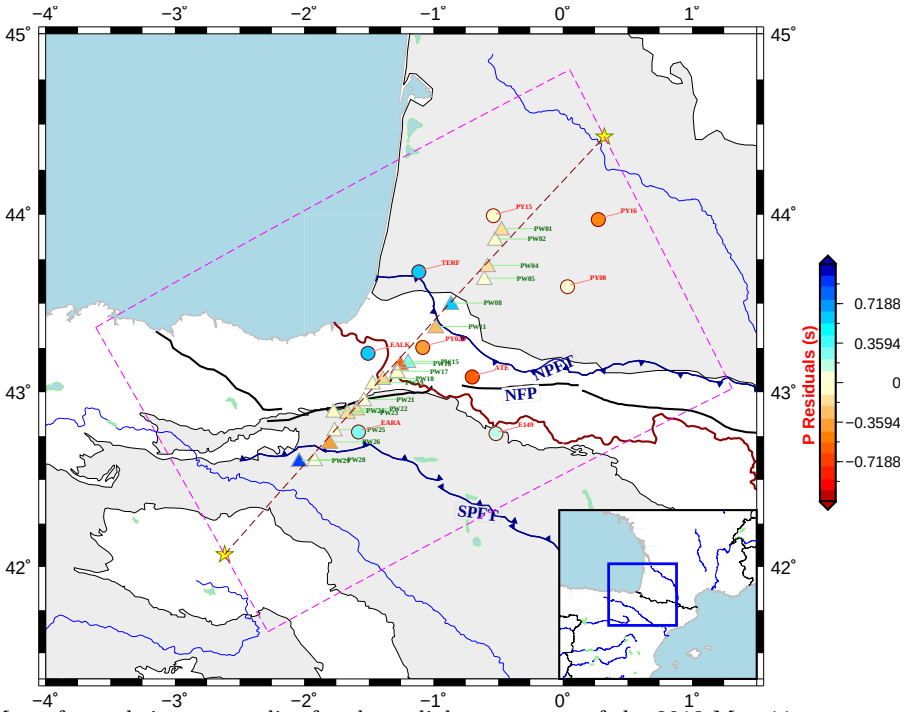


(a) Map of travel time anomalies for the vertical component of the 2013 May 11 event.

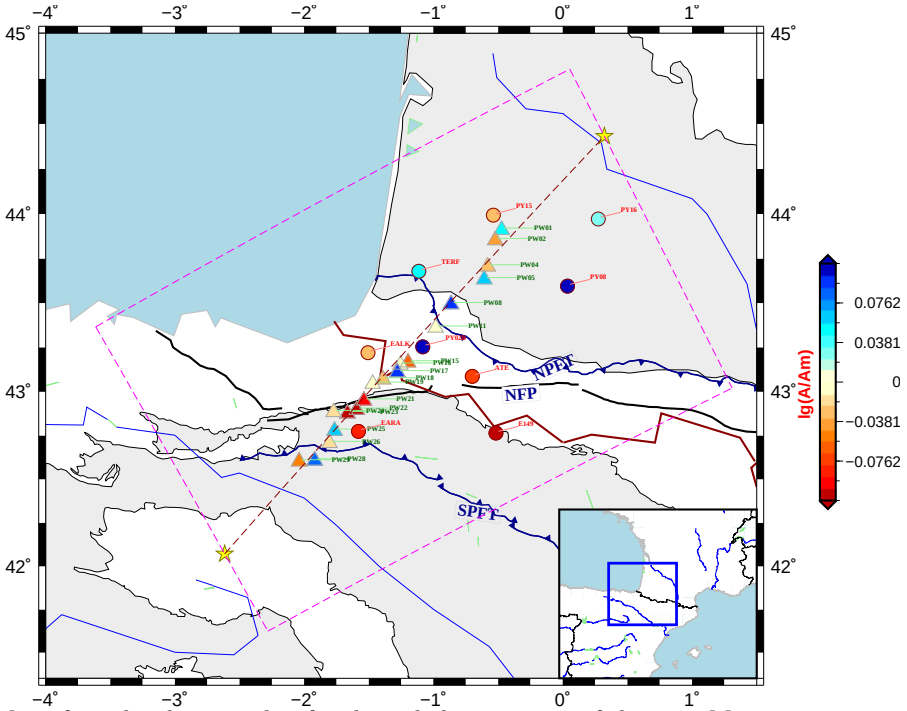


(b) Map of amplitude anomalies for the vertical component of the 2013 May 11 event.

Figure 4.11: Map of travel time (a) and amplitude (b) anomalies for the vertical component of the 2013 May 11 event. The western transect (triangles) was complemented with permanent and temporary IBERARRAY and PYROPE broad-band stations (circles). Gray area in the North: Aquitaine Basin; Gray aera in the South: Ebro Basin; NPF, North Pyrenean Fault (black solid line); NPFT, North Pyrenean Front Thrust; SPFT, South Pyrenean Front Thrust (blue solid line with triangles).



(a) Map of travel time anomalies for the radial component of the 2013 May 11 event.



(b) Map of amplitude anomalies for the radial component of the 2013 May 11 event.

Figure 4.12: Map of travel time (a) and amplitude (b) anomalies for the radial component of the 2013 May 11 event.

4.2 Resolution analysis of FWI for the western Pyrenees: checkerboard test

Assessing the resolution of FWI before moving to the inversion of real data is critical. Theoretically, the Hessian, the second derivative of the misfit function $\chi(\mathbf{m})$, plays an important role in local resolution analysis. After convergence of one inversion, a minimum of the misfit function is reached, which is denoted by zero Fréchet derivative of $\chi(\mathbf{m})$:

$$\nabla_m \chi(\widetilde{\mathbf{m}}) = 0 \tag{4.4}$$

In the vicinity of this minimum one can estimate the local resolution of the inversion and the trade-offs between model parameters. These quantities can be estimated from the Hessian via a second order Taylor expansion of $\chi(\mathbf{m})$ around the convergent model $\widetilde{\mathbf{m}}$:

$$\begin{aligned} \chi(\mathbf{m}) \approx \chi(\widetilde{\mathbf{m}}) + \frac{1}{2}(\mathbf{m}-\widetilde{\mathbf{m}})^T \mathbf{H}(\widetilde{\mathbf{m}})\,(\mathbf{m}-\widetilde{\mathbf{m}}) \\ \mathbf{H}(\widetilde{\mathbf{m}}) = \nabla_m \nabla_m \chi(\widetilde{\mathbf{m}}) \end{aligned} \tag{4.5}$$

The Hessian $\mathbf{H}(\widetilde{\mathbf{m}})$ describes the curvature or convexity of the misfit function $\chi(\mathbf{m})$ in the neighborhood of convergent model $\widetilde{\mathbf{m}}$. From the second order approximation (4.5), the Hessian gives the change of the misfit for any small model pertubation $\delta\mathbf{m} = \mathbf{m} - \widetilde{\mathbf{m}}$. When a diagonal element of the Hessian $\mathbf{H}(\widetilde{\mathbf{m}})$ is larger than other Hessian matrix element, the change of the corresponding single model parameter results in a significant increase of the misfit function. It means that the diagonal elements of $\mathbf{H}(\widetilde{\mathbf{m}})$ define the local resolution of the model parameters. In this sense, the nonzero off-diagonal elements of $\mathbf{H}(\widetilde{\mathbf{m}})$ measure the trade-offs between the different model parameters. Large off-diagonal elements indicates that there exists strong trade-offs between model parameters (Fichtner et al., 2011).

For the least squares misfit function used in our FWI study:

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{s}(\mathbf{m}) - \mathbf{d})^T \cdot (\mathbf{s}(\mathbf{m}) - \mathbf{d}) \tag{4.6}$$

where $\mathbf{s}(\mathbf{m})$ and $\mathbf{d}$ represent the synthetic and observed waveforms, respectively, the Hessian $\mathbf{H}(\mathbf{m})$ is given by :

$$\mathbf{H}(\mathbf{m}) = \mathbf{G}(\mathbf{m})^T \mathbf{G}(\mathbf{m}) + (\mathbf{s}(\mathbf{m}) - \mathbf{d})^T \nabla_m \mathbf{G}(\mathbf{m}) \tag{4.7}$$

with:

$$\mathbf{G}(\mathbf{m}) = \nabla_m \mathbf{s}(\mathbf{m}) \tag{4.8}$$

The full Hessian contains two terms. When the inversion problem is approximately linear ($\nabla_m \nabla_m \mathbf{s}(\mathbf{m}) \approx \mathbf{0}$) or the wavefom misfit is small ($\mathbf{s}(\mathbf{m}) \approx \mathbf{d}$), the second term can be ignored and the full Hessian $\mathbf{H}(\mathbf{m})$ reduces to the approximate Hessian $\widetilde{\mathbf{H}}(\mathbf{m})$:

$$\widetilde{\mathbf{H}}(\mathbf{m}) = \mathbf{G}(\mathbf{m})^T \mathbf{G}(\mathbf{m}) \tag{4.9}$$

In many nonlinear inversion when a model $\mathbf{m}$ approaches to the model $\tilde{\mathbf{m}}$ which corresponds to the minimum of χ, the Fréchet derivative of misfit function tends to zero:

$$\nabla_m \chi(\mathbf{m}) = (\mathbf{s}(\mathbf{m}) - \mathbf{d})^T \mathbf{G}(\mathbf{m}) \approx 0 \tag{4.10}$$

This condition may be obtained by $\mathbf{G}(\mathbf{m}) \approx \mathbf{0}$, while $(\mathbf{s}(\mathbf{m}) - \mathbf{d}) \neq \mathbf{0}$ in general. Therefore, the full Hessian $\mathbf{H}(\mathbf{m})$ may differ substantially from the approximate Hessian $\widetilde{\mathbf{H}}(\mathbf{m})$ when $\mathbf{m} \approx \tilde{\mathbf{m}}$.

Unlike the approximate Hessian, the full Hessian for our FWI problem is generally too expensive to calculate. Also, it requires a huge space to be stored because of its gigantic dimensions. Given these difficulties, assessing the resolution and robustness of FWI images is usually performed through synthetic inversion tests and on the visual inspection of the tomographic images or on the analysis of the data fit (Tape et al., 2007; Fichtner et al., 2009; Luo et al., 2013). In our study, the hybrid approach allows us to efficiently carry out resolution analysis using 3D synthetic models. In that case, no approximation is required to construct the synthetic dataset. We use the classical checkerboard test for resolution analysis (Lévêque et al., 1993). The reconstructions of the input models provide a reasonable estimate of the resolution capability of our FWI method. It will also allow us to assess and compare the benefit of using different preconditioners.

The input checkerboard model consists of a mosaic of alternatively positive and negative 16% anomalies with respect to a smooth 1D background velocity and density model, shown in the top right panel of Figure 4.13. The 20 km checkerboard pattern anomalies are imposed between 20 km and 80 km depth. The smooth 1D background model is defined by fitting the arrival times of the direct P and S waves computed in the reference AK135 Earth model. We match this smooth model, shown in the middle right panel and bottom right panel of Figure 4.13, to the 1D AK135 model on the edges of the SEM grid by using a cosine taper function with a width of 15 km. The anomalies have an infinite extension along the x axis, shown in the left panel of Figure 4.13.

The synthetic data are computed in the input model, using the same paths and source wavelets as in the real dataset. No noise is added to the synthetic data. We use the smooth 1D model as the starting model for the inversions. The tomographic models are parameterized with three parameters: **density, and isotropic compressional and shear velocity** values defined on each SEM grid point. All the following inversions use the regional mesh with the Pyrenees free surface topography. The stations used for inversion are located at their elevation on the free surface topography. The gradient of the waveform misfit function with respect to Vp, Vs, and density are computed by using the adjoint method. The descent direction or directional derivative is usually obtained by preconditioning the gradient. The term preconditioning is generally used in conjugate gradient methods. Its synonym 'rescaling' is usually used in quasi-Newton algorithms. The preconditioning can be seen as a change of variables,

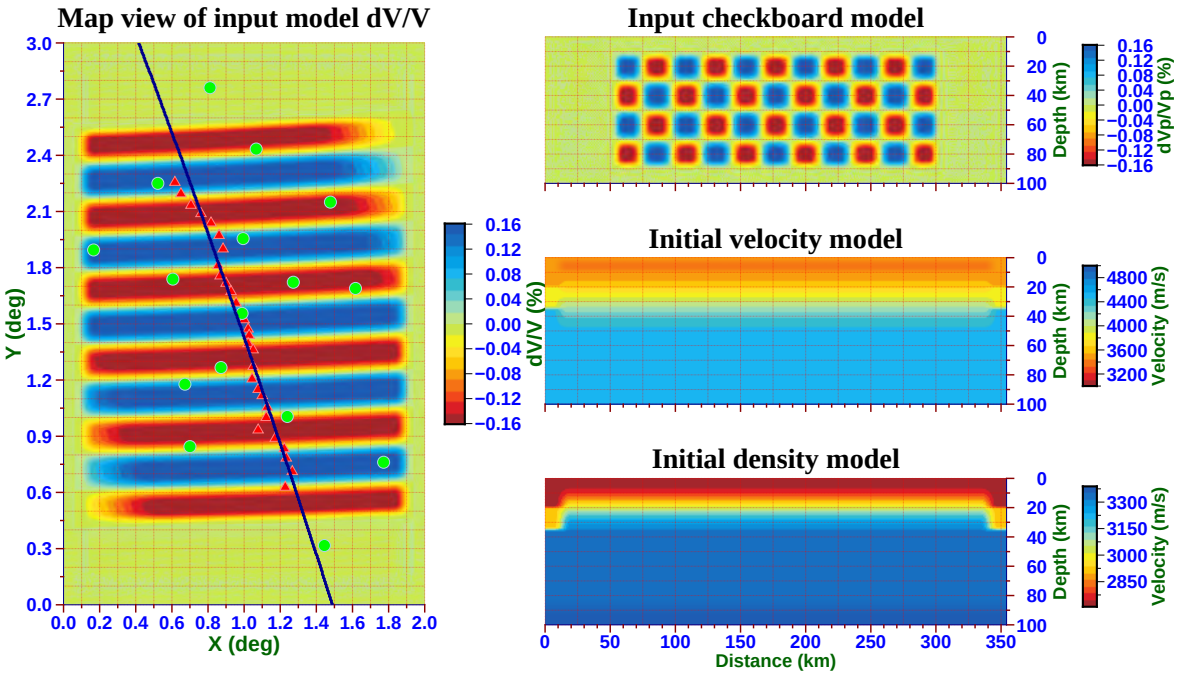


Figure 4.13: Map and vertical section views of checkerboard model.

aimed at providing accelerated convergence of the algorithm; rescaling is more often thought of an initial approximation to the Hessian. The Newton method uses the inverse of the Hessian as the preconditioner which is usually prohibitively expensive to compute and store. In our study, we solve the FWI problem with the iterative L-BFGS algorithm which computes an approximate inverse Hessian refined during the different iterations of the algorithm. An initial inverse Hessian or initial preconditioner is needed for this algorithm. We define the initial inverse Hessian by the square root of depth at each SEM grid point. Physically, this can be seen as weighting the gradient by a factor that is inversely proportional to the geometrical spreading of the adjoint wavefield. It is worth noting that each resolution test requires the same number of forward and adjoint simulations as the FWI for real the dataset.

We perform various resolution tests based on different datasets and different data selections. In the first resolution test, we invert the vertical and radial component waveforms filtered in the optimal frequency bands determined previously. Table 4.14 shows the number of vertical and radial traces before and after selection. The average optimal filtering corner frequencies for these selected traces are shown in Table 4.15. We do not update the density model in this first resolution test.

The convergence of the inversion is defined as the moment at which the L-BFGS algorithm can not honour the Wolfe conditions or the differences of updated model between two successive iterations are small enough. After 9 iterations of our iterative waveform inversion algorithm, we obtain the convergent Vp and Vs models shown in Figure 4.16a and 4.16b, respectively.

Number of components	Vertical trace after selection	Vertical trace before selection	Radial trace after selection	Radial trace before selection
2013 May 11	40	40	28	40
2013 May 24	44	44	44	44
2013 Aug 13	37	37	26	37
2013 Aug 30	35	35	29	35
2013 Sep 25	30	30	29	30

Table 4.14: The number of traces before and after selection.

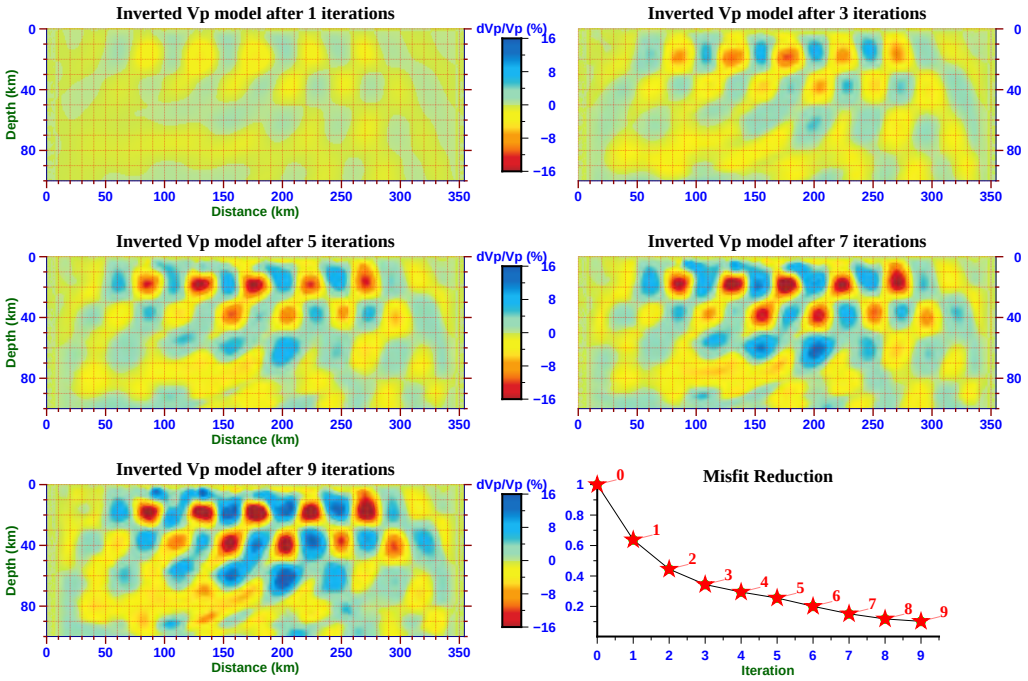
Number of components	fs for vertical component	fe for vertical component	fs for radial component	fe for radial component
2013 May 11	0.03055	0.16452	0.044402	0.1463
2013 May 24	0.0096591	0.2	0.014133	0.18023
2013 Aug 13	0.011804	0.20019	0.047544	0.18677
2013 Aug 30	0.01	0.2	0.021661	0.17614
2013 Sep 25	0.01	0.2	0.032076	0.22196

Table 4.15: The average corner frequencies for each event.

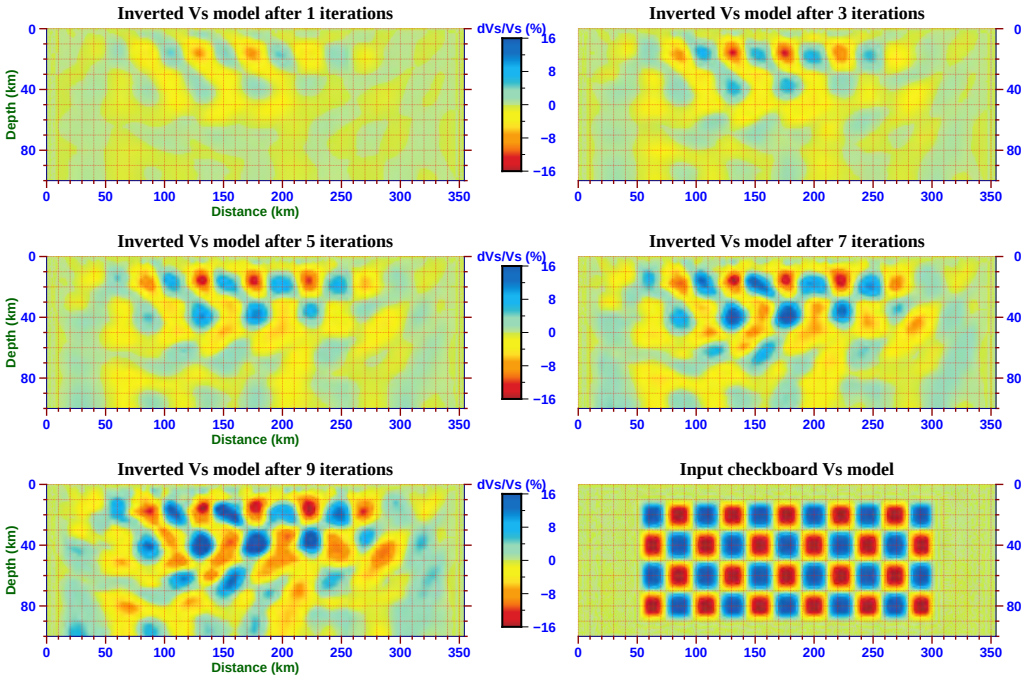
The waveform relative misfit residual, which is the ratio of the waveform misfit before FWI to the misfit after FWI, is 0.117676 after 9 iterations. Table 4.17 shows the vertical and radial component waveform misfits of each event before FWI and the waveform relative misfit residuals after FWI. The waveform misfit of each event is normalized by its average trace energy. Among the five events, the relative misfit residuals are the largest for the 2013 May 11 and Aug 30 earthquakes.

Figure 4.18a and 4.18b show the travel time anomalies for the 2013 May 11 event before and after FWI, respectively. Figure 4.18c and 4.18d show the amplitude anomalies for this event before and after FWI, respectively. We can see that before inversion, the travel time anomalies measured on the stations of the western transect (shown by triangles) exhibit the spatial distribution of subsurface checkerboard pattern. The amplitude anomalies also show regular spatial variations along the western transect. Our input model has large perturbations of ±16%. For such large perturbations, the linearized assumption made in asymptotic ray theory is no longer valid. The large velocity perturbations will produce strong variations of both phase and amplitude anomalies as well as strong waveform distortions. The information carried by full waveform allowed us to retrieve these strong anomalies. After FWI, we can see that the travel time and amplitude anomalies have decreased very significantly. The sum of the square of travel time anomalies decreases from 4.8646 to 0.4267. The amplitude relative misfit residual after inversion is 0.1151 for this event.

Figure 4.19a and 4.19b show the misfit of vertical component waveforms for the 2013 May 11 event before and after FWI. The waveform misfit reduction is significant on most traces. If



(a) Vp model for the first checkerboard inversion test.

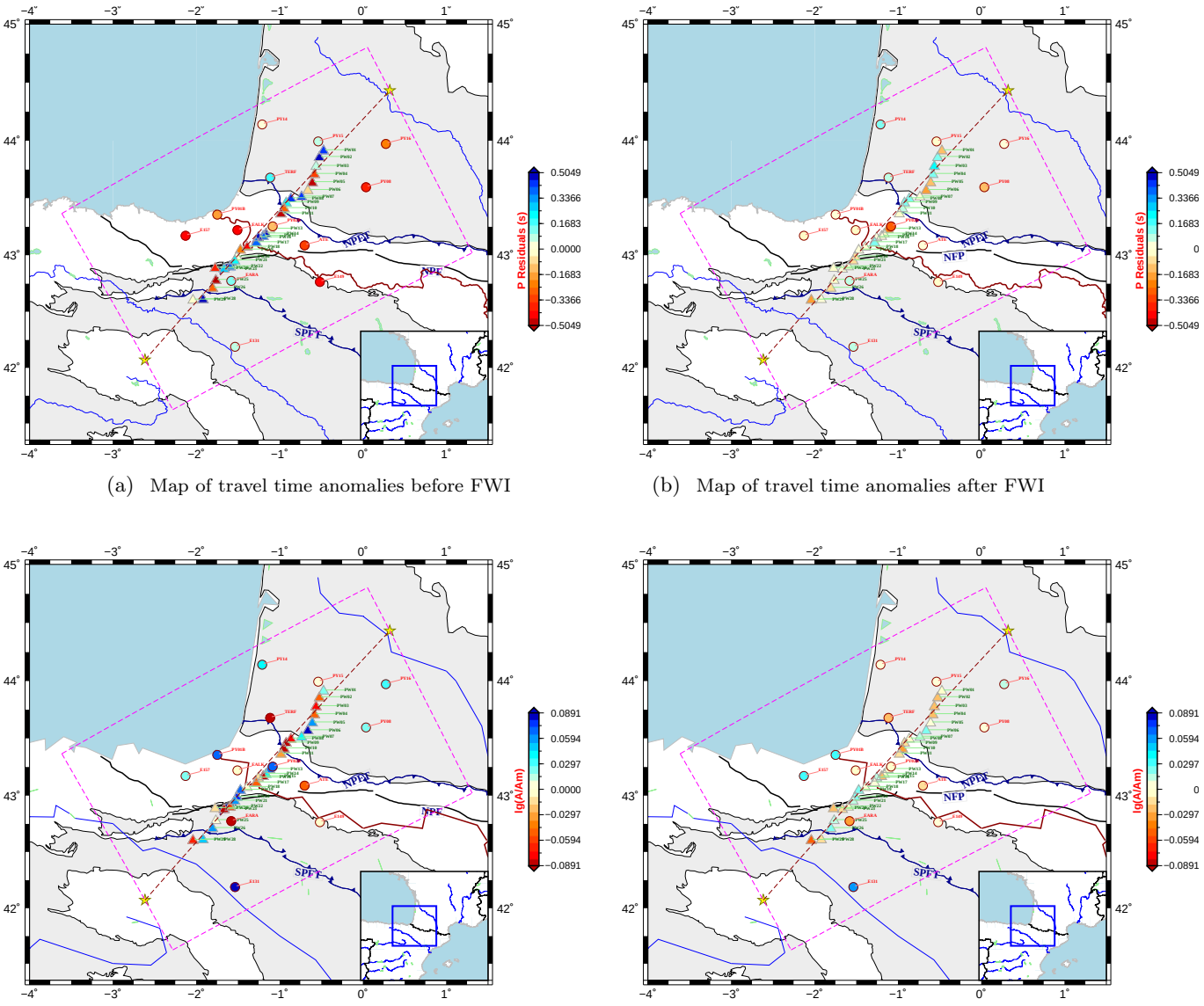


(b) Vs model for the first checkerboard inversion test.

Figure 4.16: Results of the first synthetic checkerboard inversion test. (a) inverted Vp model and relative residual reduction. (b) inverted Vs model and input checkerboard model.

waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
2013 May 11	75.516	0.1245	14.0	0.247
2013 May 24	139.237	0.096	48.9	0.26
2013 Aug 13	182.266	0.053	30.26	0.065
2013 Aug 30	19.9687	0.124	4.55	0.525
2013 Sep 25	84.257	0.066	30.68	0.113

Table 4.17: Waveform misfits and relative misfit residuals for each event.



(a) Map of travel time anomalies before FWI

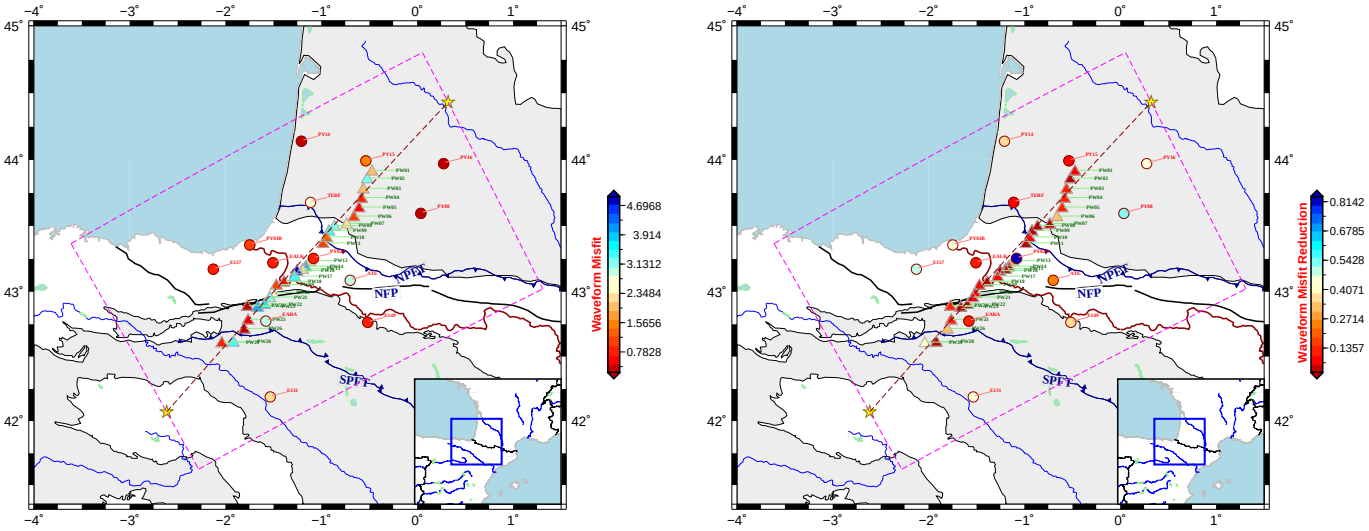
(b) Map of travel time anomalies after FWI

(c) Map of amplitude anomalies before FWI

(d) Map of amplitude anomalies after FWI

Figure 4.18: The map of travel time anomalies for the vertical component of the 2013 May 11 event before FWI (a) and after FWI (b). The map of amplitude anomalies for the vertical component of the 2013 May 11 event before FWI (c) and after FWI (d).

we check the vertical component synthetic waveforms after FWI in Figure 4.20a and 4.20b, we can see that all the traces have a very low misfit residual.



(a) Waveform misfit before FWI

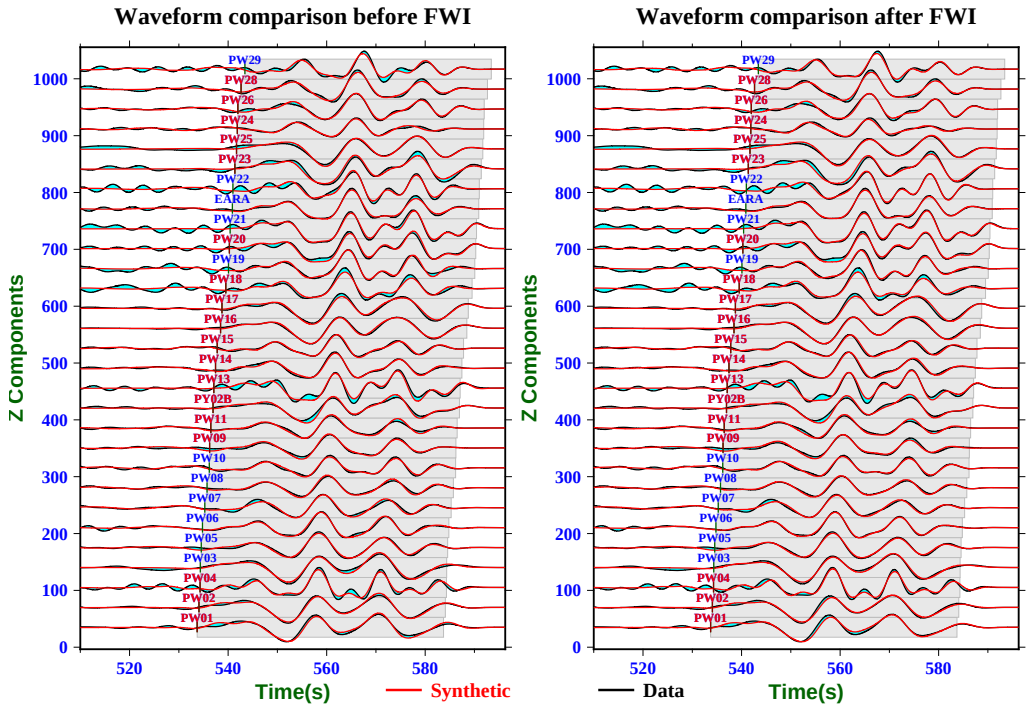
(b) Waveform misfit reduction after FWI

Figure 4.19: (a) Map of waveform misfit for the vertical component of the 2013 May 11 event before FWI. (b) Map of waveform relative misfit residual for the vertical component of the 2013 May 11 event after FWI.

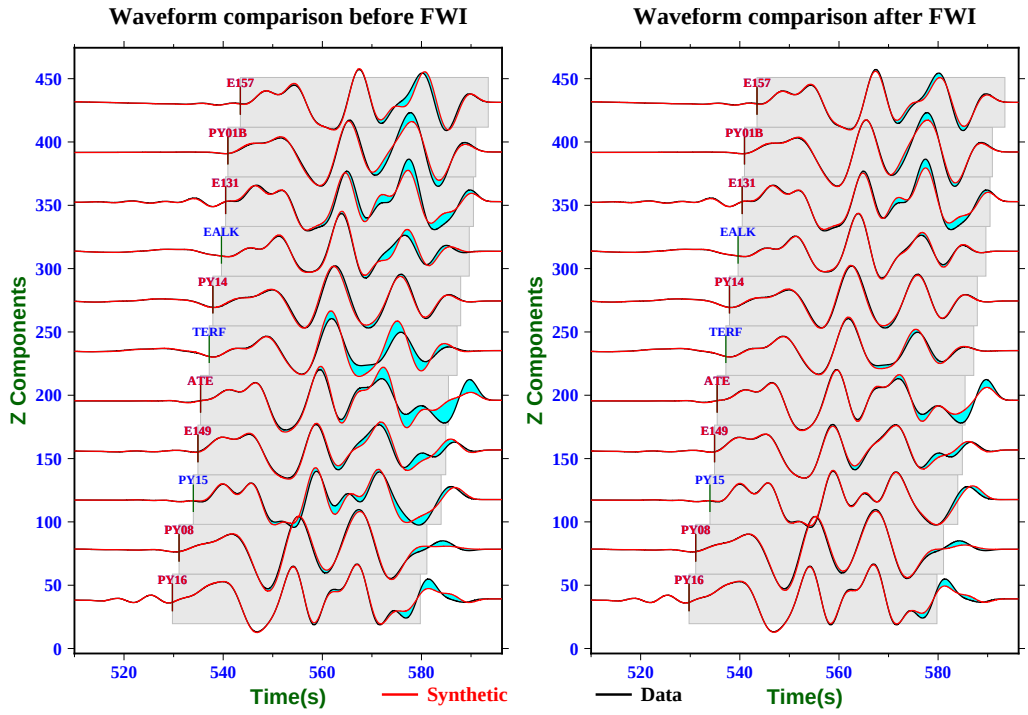
In general, the relative misfit residual for the vertical component waveforms is lower than that of radial component. Figure 4.21a and 4.21b show the misfit of radial component waveforms for the 2013 May 11 event before and after FWI. Figure 4.22a and 4.22b show the radial component waveform variation after FWI. Usually, the waveform fit degrades as time lapse increase. This probably comes from the increasing contribution of multiply scattered waves in the late coda of the P wave.

For the 2013 Aug 30 earthquake, we only show the misfit of radial component waveforms before FWI and the relative misfit residual after FWI in Figure 4.23a and 4.23b, since its vertical component waveform misfit reduction is significant. Figure 4.24a and 4.24b show the comparison of radial component waveforms before and after FWI. The traces with lowest misfit reductions are mainly located outside the transect. In Figure 4.24a, there are some traces with small amplitude (for example, PW13, PW21 and PW27). This is due to their relatively narrow frequency bandwidth for filtering, determined by optimal frequency range selection.

From Figure 4.16, the principal features of both the Vp and Vs checkerboard patterns are well retrieved down to 60 km depth. Deep structures are less well resolved than shallower ones. This may be due to the relative sparsity of ray coverage in deep parts but also to vertical smearing. However, in spite of the decrease of spatial resolution with depth, our method is capable of retrieving velocity anomalies that are smaller than both P and S wavelengths. Especially, the FWI is able to invert the velocity gradient with depth, which is usually impossible with

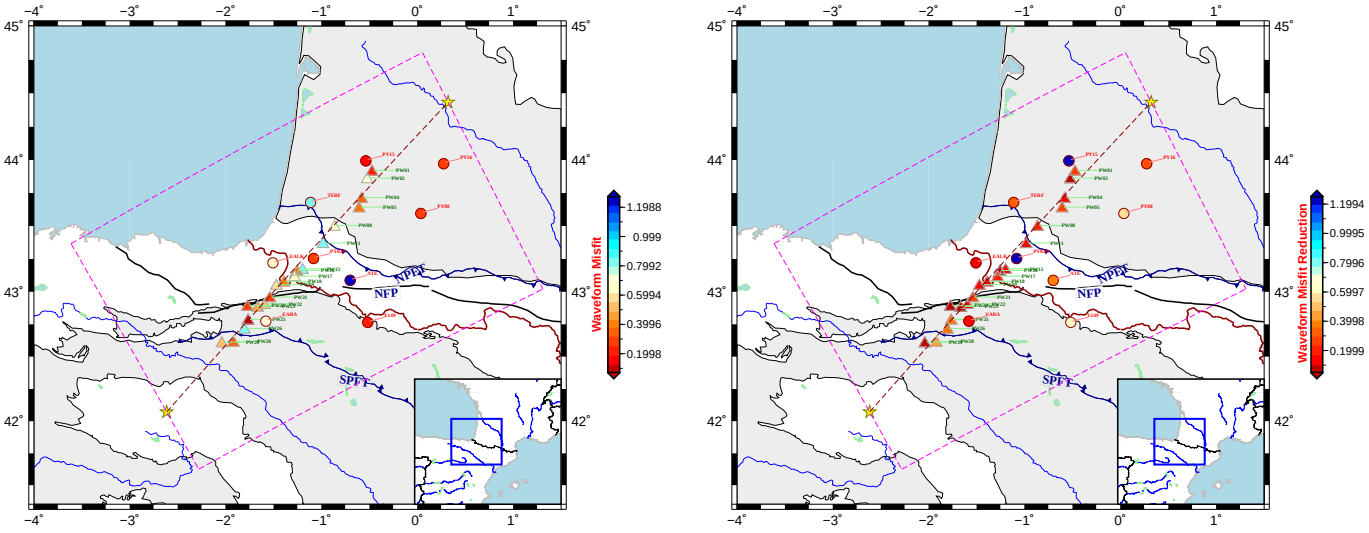


(a) Vertical component waveforms along the transect.



(b) Vertical component waveforms outside the transect.

Figure 4.20: Observed and synthetic vertical component waveforms along the transect (a) and outside the transect (b) for the 2013 May 11 event.



(a) Waveform misfit before FWI

(b) Waveform misfit reduction after FWI

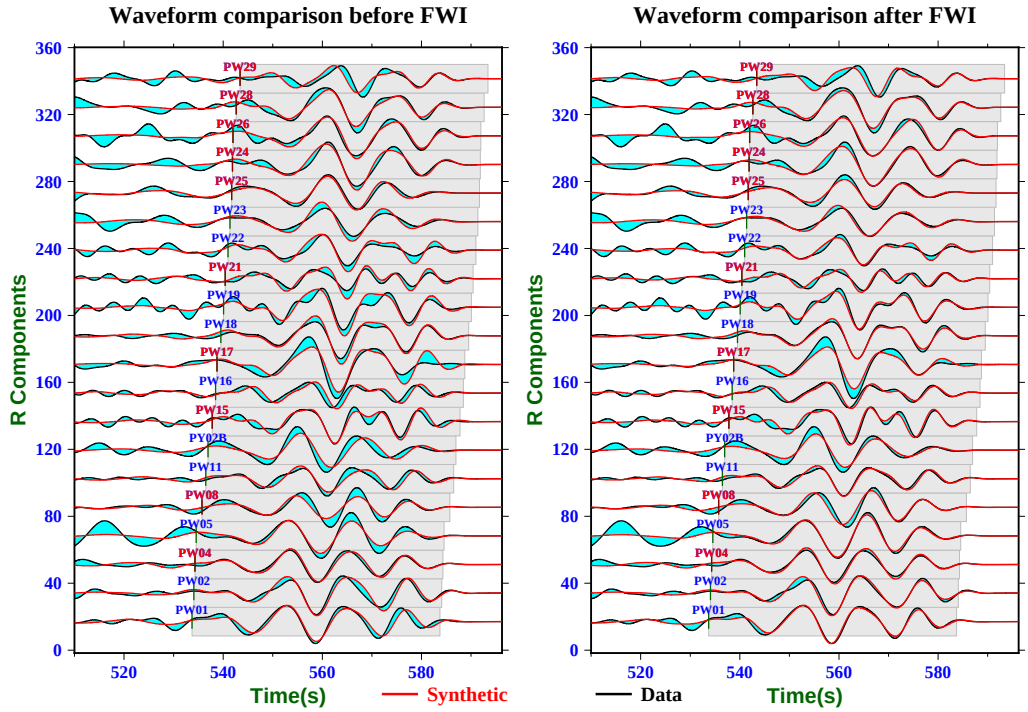
Figure 4.21: (a) Map of waveform misfit for the radial component of the 2013 May 11 event before FWI. (b) Map of waveform relative misfit residual for the radial component of the 2013 May 11 event after FWI.

classical travel time tomography. Another significant conclusion of this first resolution test is that we can get a fairly good resolution with only five events, which is quite remarkable. This suggests that this method is well adapted to handle data from short duration temporary deployments. Exploiting the waveform information seems to greatly improve the resolution of seismic tomography.

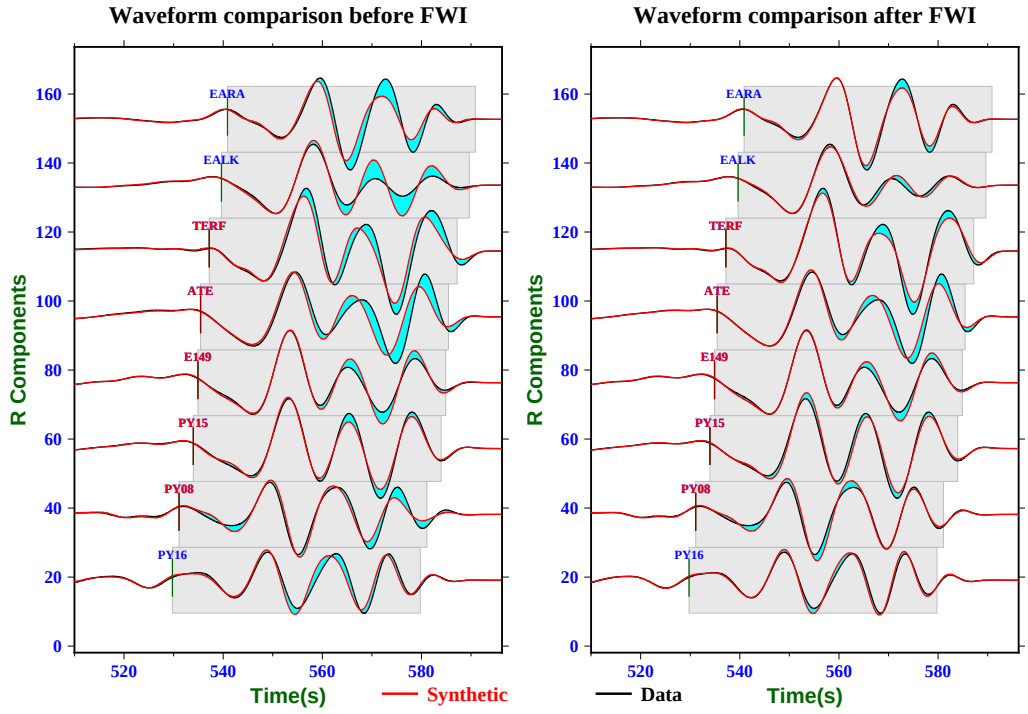
The frequency range selection based on the coherence between observed and synthetic waveforms can improve the S/N ratio. However, many traces are discarded after selection. This may produce some spatial sampling 'hole' when the final event kernel is constructed. In a second resolution test, we use all available vertical and radial component waveforms filtered at uniform corner frequencies for FWI (Since we generate the synthetic data without adding noise, the lower corner frequency is simply set to 0.01 Hz for the vertical component and 0.01 Hz for radial component; the upper corner frequency is fixed to 0.1 Hz in this test). At the same time, the density model is now updated in this new resolution test. We try to investigate the possible difference in resolution between the previous resolution test and this more general case.

After 9 iterations, we obtain the convergent Vp, Vs and ρ models shown in Figure 4.25a. The relative misfit residual after 9 iterations is 0.1221. The total number of traces and frequency contents of waveform set with uniform corner frequency filter used for this resolution test increase. Figure 4.25b shows the final Vp and Vs models obtained when the seismograms are filtered with the optimal corner frequencies, as a comparison.

In general, travel times of seismic waves are much more sensitive to velocities than density. Therefore, the density is usually not inverted for travel time tomography. However, our resolution tests show that waveform inversion can put some constraints on density, especially at

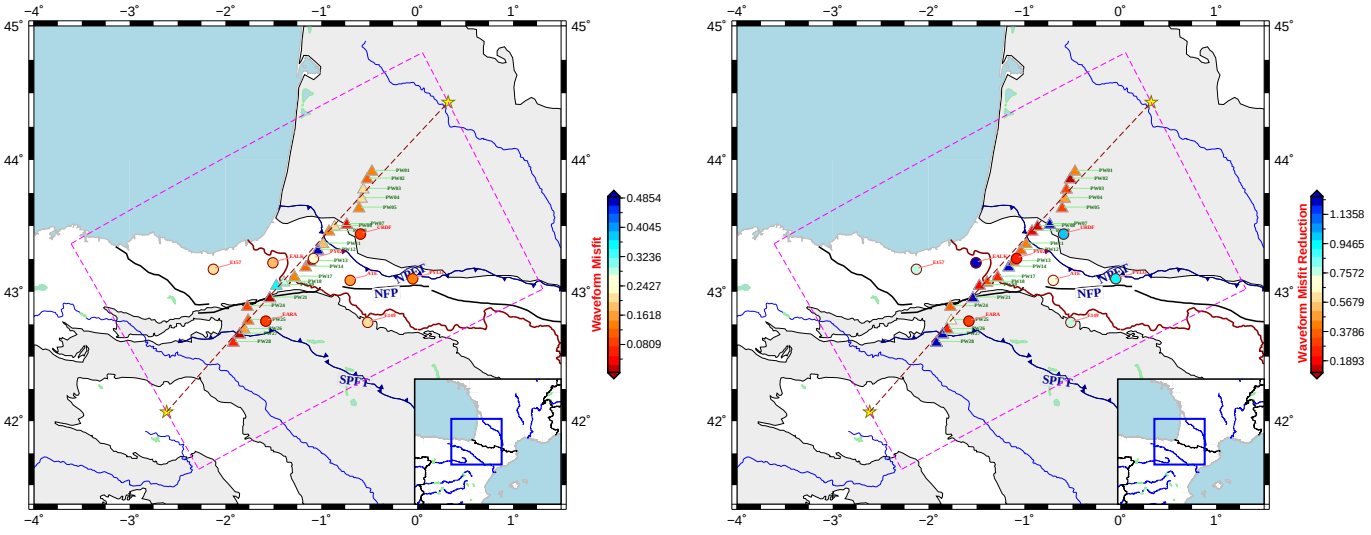


(a) Radial component waveforms along the transect.



(b) Radial component waveforms outside the transect.

Figure 4.22: Observed and synthetic radial component waveforms along the transect (a) and outside the transect (b) for the 2013 May 11 event.



(a) Waveform misfit before FWI

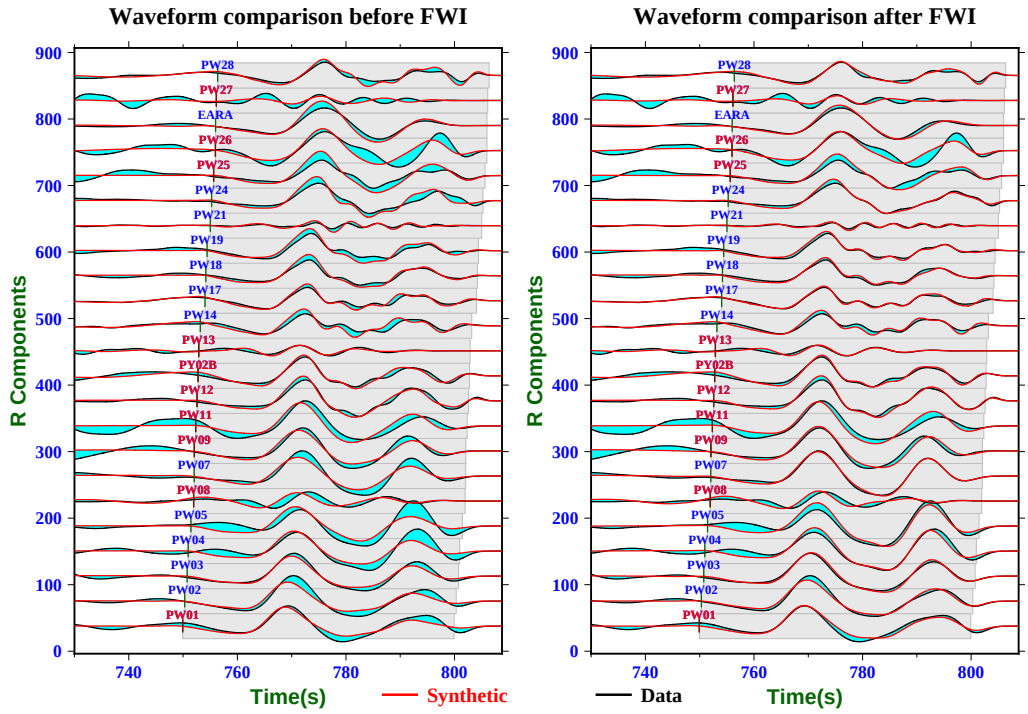
(b) Waveform misfit reduction after FWI

Figure 4.23: (a) Map of waveform misfit for the radial component of the 2013 Aug 30 event before FWI. (b) Map of waveform relative misfit residual for the radial component.

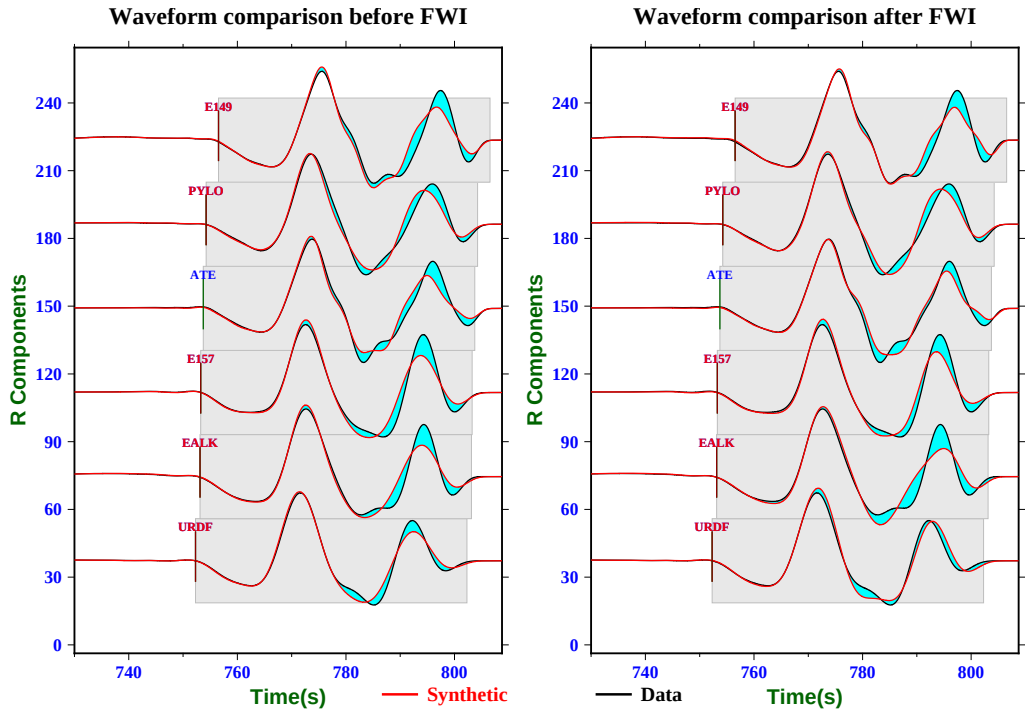
short wavelength. This sensitivity comes from the contribution of back-scattered waves such as $PpPp$ and $PpPs$ phases (Tong et al., 2014a), which are included within the time window used in this inversion test. The shapes of the inverted density anomalies are well recovered, down to 60 km depth. But the amplitudes of anomalies are underestimated. The waveform misfit reduction in these two resolution tests are comparable, and the final inverted velocity models are very similar.

So far, we have shown the inverted models in vertical sections along the strike of the western transect. It is also insightful to inspect the inverted model at different depths in the map view. This allows us to check the horizontal resolution in our FWI resolution tests. Figures 4.26 and 4.27 show the input and inverted Vp and Vs models at depths between 10 to 80 km, respectively. The left, center and right columns in each figure show input velocity model, inverted velocity model with uniform corner frequency filter, and inverted velocity model with optimal frequency filter, respectively.

The amplitudes of recovered checkerboard velocity anomalies by FWI using optimal frequency filter are slightly better than that of using uniform corner frequency filter. The amplitudes of positive anomalies are generally less well recovered than their adjacent negative anomalies. This may result from a better resolution in the regions with lower velocities. The width of the well resolved region is about 15 km perpendicular to the strike of the western transect above 60 km depth in the Vp model. In contrast, the width of the well resolved region in the Vs model is about 10 km perpendicular to the strike of the transect above 50 km depth. This difference in horizontal resolution may be due to the shorter wavelength of shear waves compared to compressional waves, in a given frequency range. Another possible way to improve

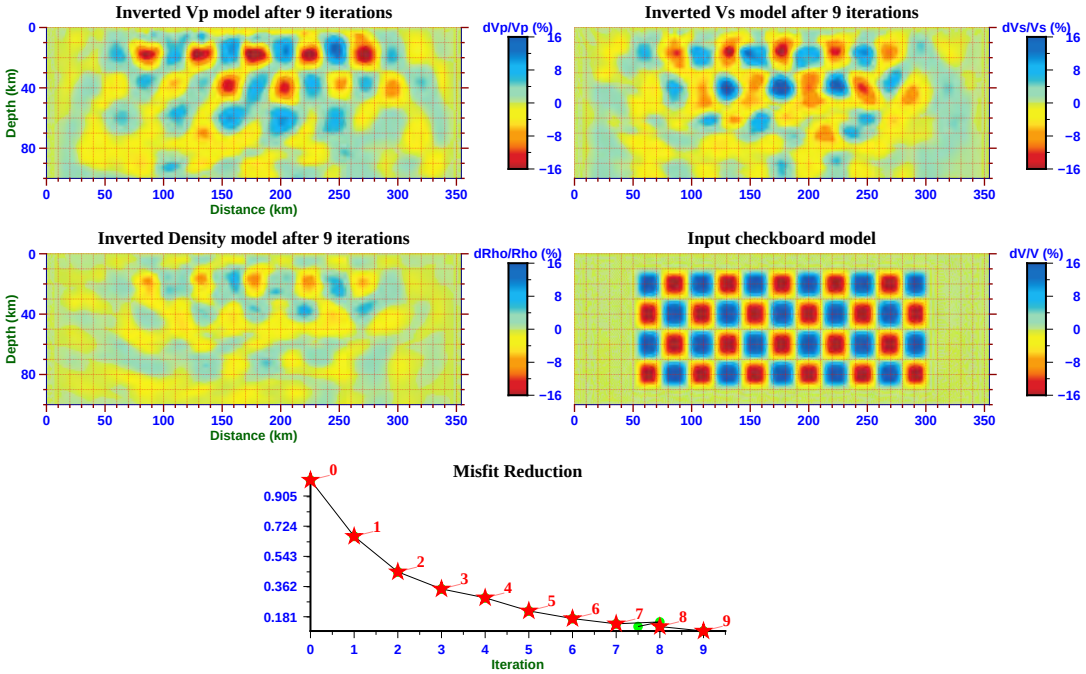


(a) Radial component waveforms along the transect.

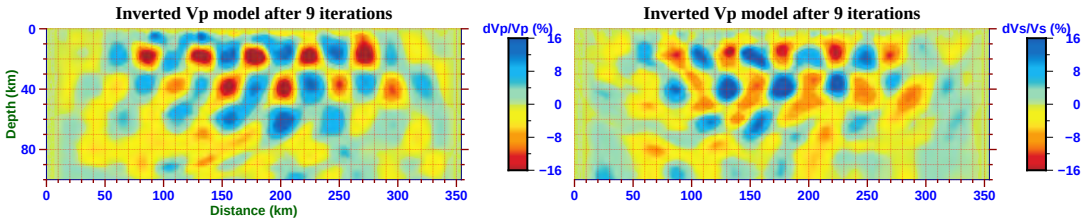


(b) Radial component waveforms outside the transect.

Figure 4.24: (a) Radial component waveforms along the transect (a) and outside the transect (b) for the 2013 Aug 30 event.

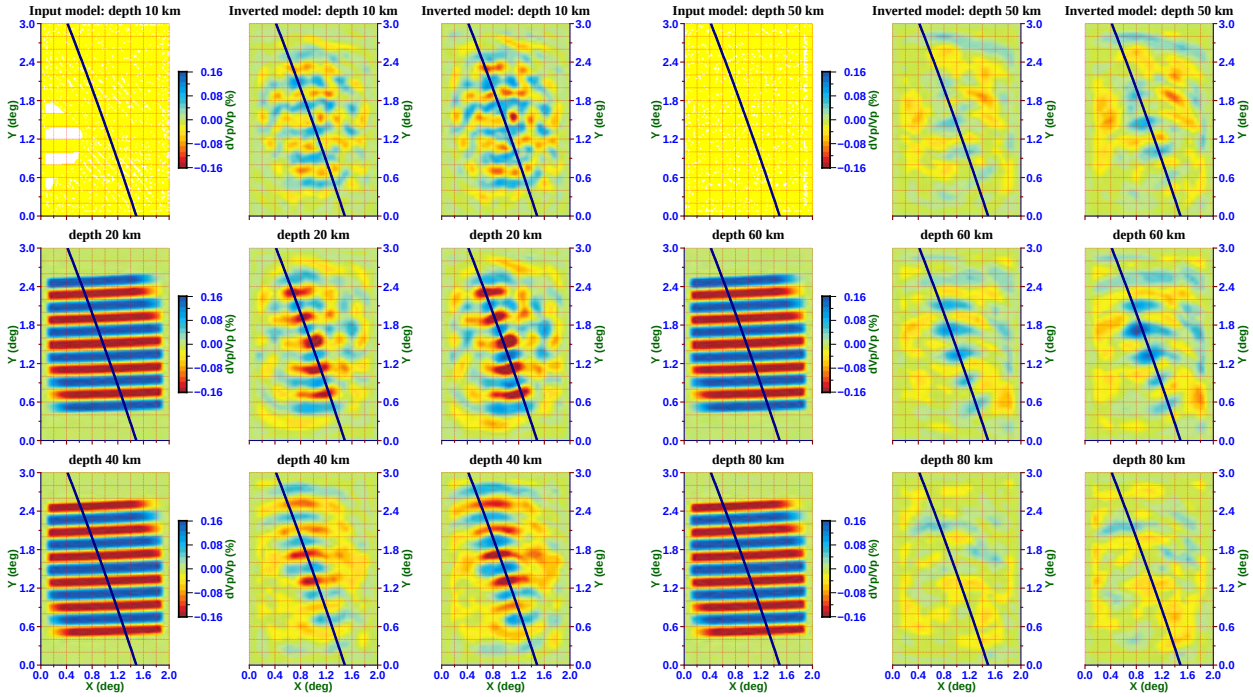


(a) Inverted velocity and density models using a uniform corner frequency filter for all the available traces.



(b) Inverted velocity models with optimal frequency filter for each trace.

Figure 4.25: (a) Inverted velocity and density models and waveform misfit reduction for second checkerboard test. (b) Inverted velocity models by using optimal frequency filter for each trace.



(a) Map views of Vp models at shallow depth.

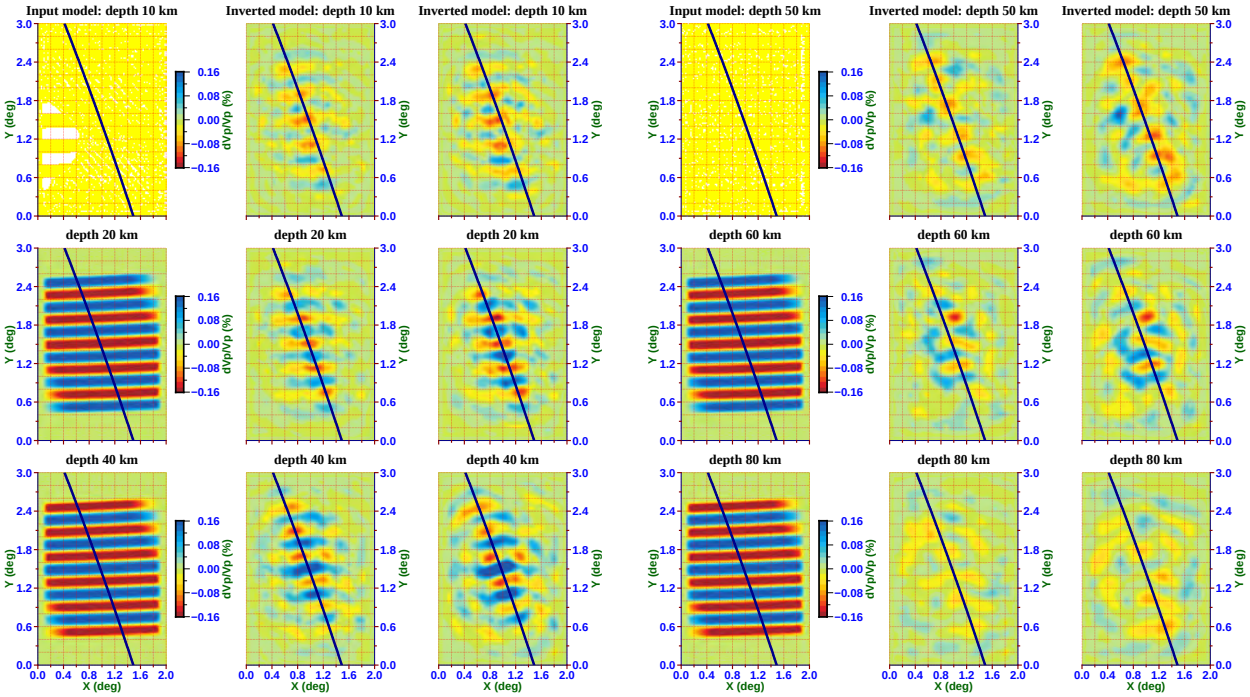
(b) Map views of Vp models at deep depth.

Figure 4.26: Map views of input and inverted Vp models. (a) The left, center and right columns are input Vp model, inverted Vp model with uniform corner frequency filter and inverted Vp model with optimal frequency filter at 10(top), 20(middle) and 40(bottom) km depth, respectively. (b) The same as (a), but at 50(top), 60(middle) and 80(bottom) km depth, respectively.

the Vs model will come from the exploitation of the transverse component waveform for FWI which will be shown next.

From these comparisons, we can conclude that discarding the traces which have worse coherence between synthetic and observed waveforms by performing our optimal frequency range selection algorithm from the dataset used in waveform inversion does not have much impact on inversion results. Another possible inference is that the velocity inversion is not very sensitive to the density model. In practice, 3D variations of density have little effect on seismic waveforms. Therefore, we can either update the density model from the inverted Vp model using empirical scaling laws or keep it unchanged during the waveform inversion.

We perform a third resolution test by using all the available three component waveforms (vertical, radial and transverse) filtered at uniform corner frequencies for FWI. The lower corner frequency is set to 0.01 Hz for the vertical component and 0.01 Hz for the radial and transverse components; the upper corner frequency is fixed at 0.1 Hz for all components. The density model is also updated. After 9 iterations, we obtain the final Vp, Vs and ρ models shown in Figure 4.28a. The waveform relative misfit residual after 9 iterations is 0.094. Figure 4.28b shows the final Vp and Vs models of the first resolution test obtained with the optimal frequency filters for comparison.



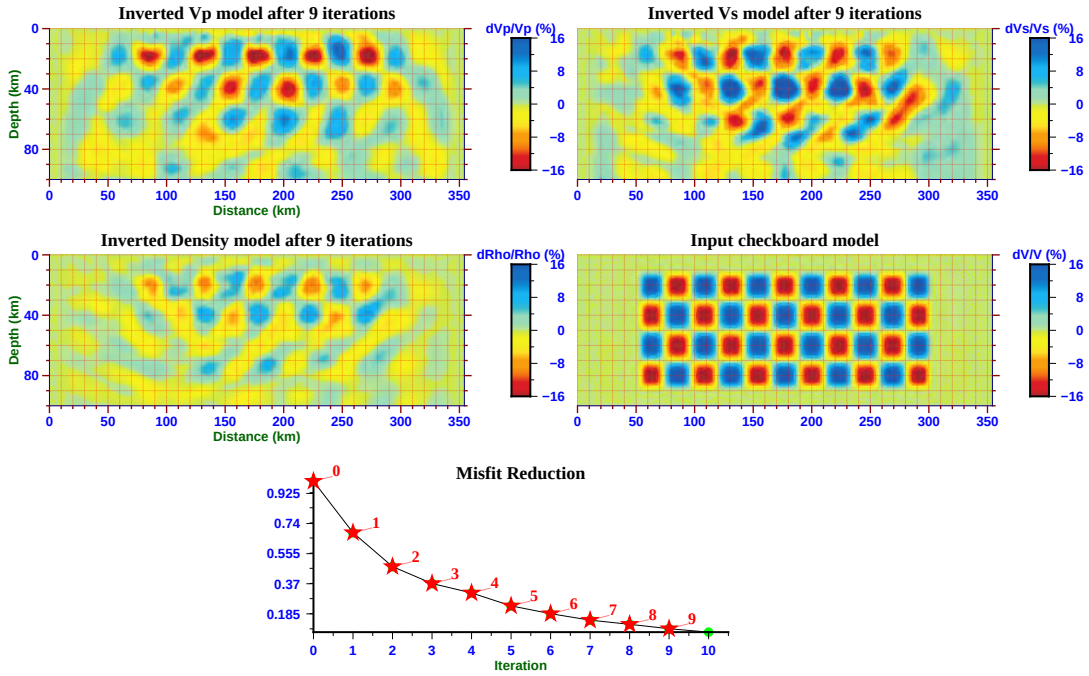
(a) Map views of Vs models at shallow depth.

(b) Map views of Vs models at deep depth.

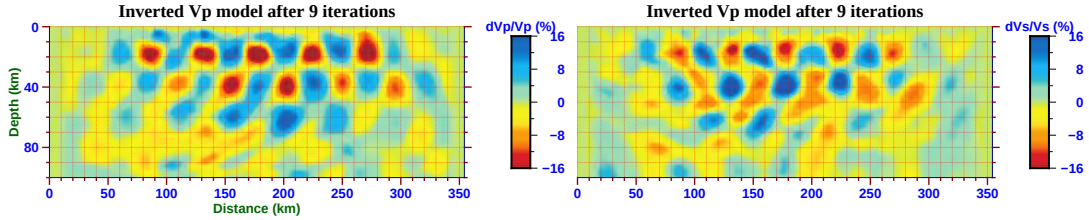
Figure 4.27: Map views of input and inverted Vs models. (a) The left, center and right columns are input Vs model, inverted Vs model with uniform corner frequency filter and inverted Vs model with optimal frequency filter at 10(top), 20(middle) and 40(bottom) km depth, respectively. (b) The same as (a), but at 50(top), 60(middle) and 80(bottom) km depth, respectively.

Compared to the second test, the amplitudes and shape of checkerboard anomalies in the inverted density model are better recovered in the third test, down to 80 km depth. Note that the relative misfit residual for this resolution test is similar to the previous two tests. The final inverted Vp models in these three tests are very similar. However, the inverted Vs model in the third test is better resolved than in the previous two tests. Both the amplitude and shape of Vs anomalies are better recovered, especially at deeper depth. Figure 4.29 and 4.30 show the map view of input and inverted Vp and Vs models, respectively. The left, center and right columns in each figure are input velocity model, inverted velocity model with uniform corner frequency filter for all three component waveforms, and inverted velocity model with optimal frequency filter for vertical and radial component waveforms, respectively.

The inverted Vp models in these three resolution tests are similar. However, the amplitudes and shape of Vs anomalies are better resolved when all three components are exploited. The horizontal width of the well resolved region for Vs is comparable to that of Vp in the last resolution test. The most evident improvement appears in the deep part of the model, from 40 to 60 km depth. Therefore, including the transverse component waveform in the inversion seems to improve the resolution in the Vs model. We should try to do this on real data in future work.

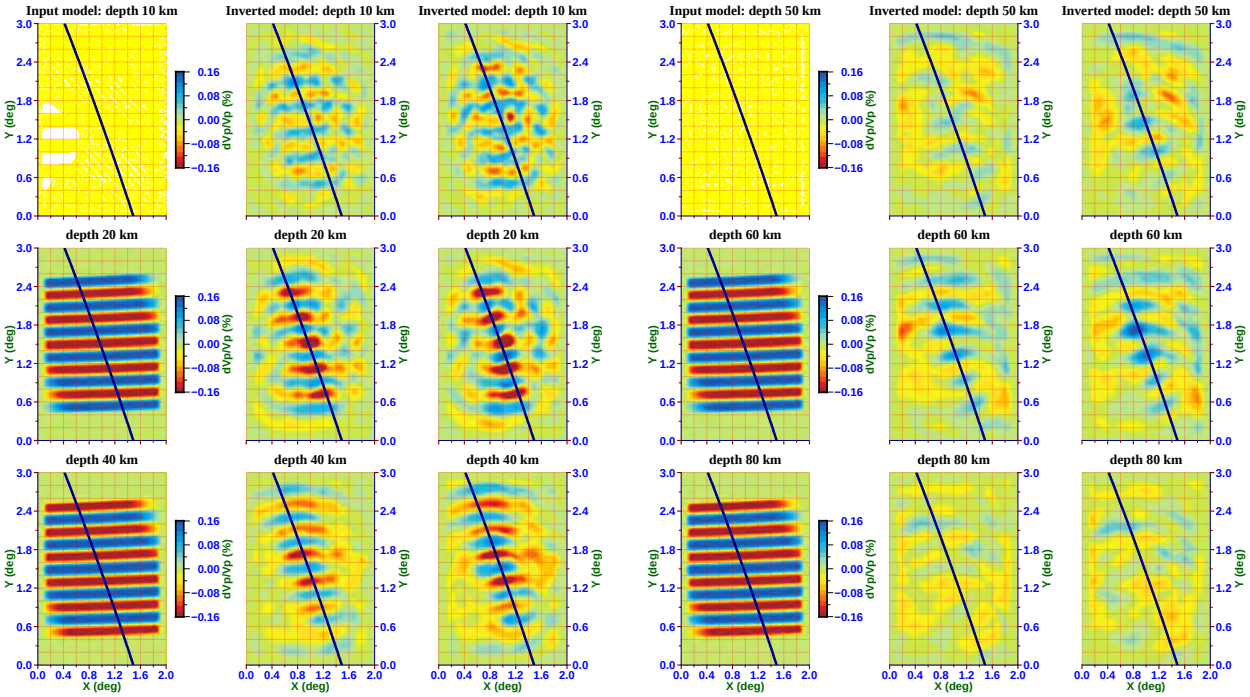


(a) Inverted velocity and density models using a uniform corner frequency filter on all the available three component waveforms.



(b) Inverted velocity models with optimal frequency filter for each trace.

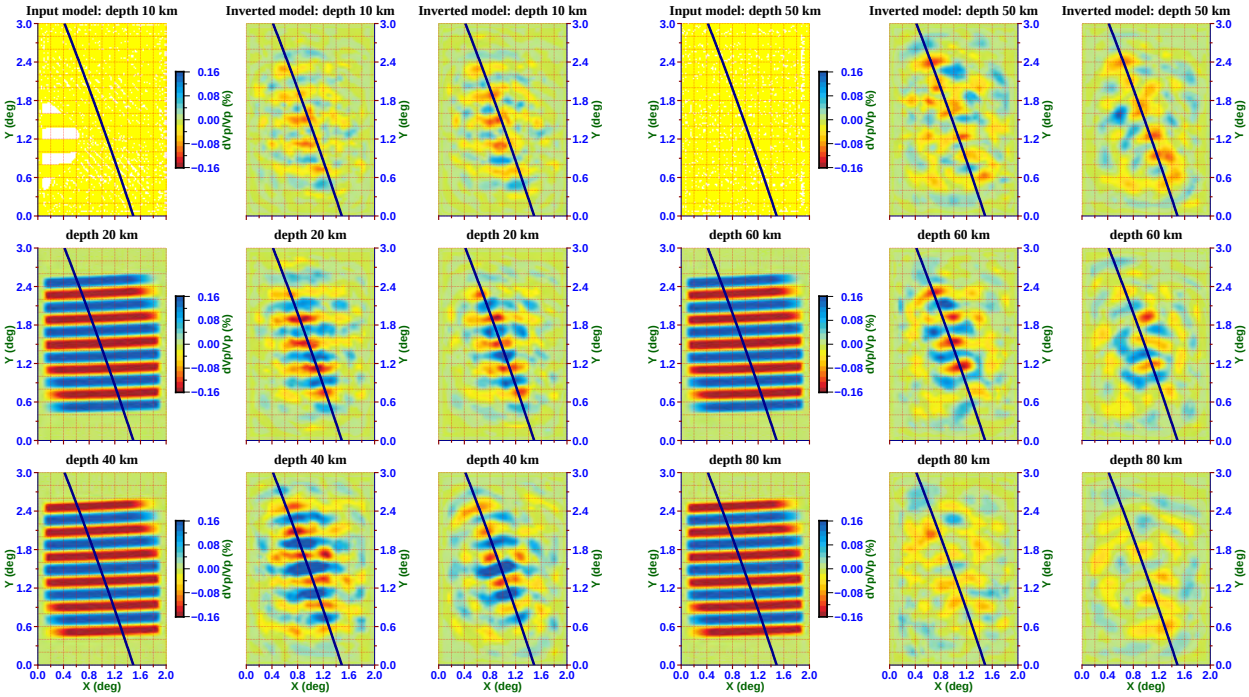
Figure 4.28: (a) Inverted velocity and density models and waveform misfit reduction for the third checkerboard test. (b) Inverted velocity models by using optimal frequency filter on each trace.



(a) Map views of Vp models at shallow depth.

(b) Map views of Vp models at deep depth.

Figure 4.29: Map views of input and inverted Vp models. (a) The left, center and right columns are input Vp model, inverted Vp model with uniform corner frequency filter for all three component waveforms and inverted Vp model with optimal frequency filter for vertical and radial component waveforms at 10(top), 20(middle) and 40(bottom) km depth, respectively. (b) The same as (a), but at 50(top), 60(middle) and 80(bottom) km depth, respectively.



(a) Map views of Vs models at shallow depth.

(b) Map views of Vs models at deep depth.

Figure 4.30: Map views of input and inverted Vs models. (a) The left, center and right columns are input Vs model, inverted Vs model with uniform corner frequency filter for all three component waveforms and inverted Vs model with optimal frequency filter for vertical and radial component waveforms at 10(top), 20(middle) and 40(bottom) km depth, respectively. (b) The same as (a), but at 50(top), 60(middle) and 80(bottom) km depth, respectively.

4.3 FWI results for the western transect

4.3.1 Waveform inversion

In the first application of our FWI approach to real data recorded by western transect, we invert the vertical and radial component waveform and neglect the effect of attenuation. We define a time window that starts 10 s before the arrival time of the P wave and ends 40 s after. The reason why we choose this length of the time window results from the relatively long durations of source wavelets (of the order of 30 to 40 s, depending on the specific event) and the requirement for including the contributions of P-to-S conversions along with the first multiples into the inversion. The geometry of crustal roots are better inverted with longer time windows including the reverberations on the Moho which are later arrivals of the P waves, as shown in Monteiller et al. (2015). These arrivals are key to constrain the sharp velocity gradients related to the main crustal interfaces such as the Moho. The main difference between our inversion method and the one described in Tong et al. (2014a) is that we invert broadband waveforms instead of receiver functions. The idea is to exploit the long-period components of the seismic wavefield to better constrain long wavelength heterogeneities while utilizing the short-period waveforms to invert the seismological interfaces.

We compute the gradient of the waveform misfit function with respect to Vp, Vs and density using the adjoint method. The resolution tests have shown that the amplitudes of density gradients are usually negligible compared to those for Vp and Vs. Therefore, the density model will be kept constant during the inversion. The optimal frequency range selection reduces the size of the available dataset. However, the synthetic tests have shown that the anomalies are similarly recovered when using the ideal larger dataset. This encourages us to use the selection algorithm in waveform inversion of real data. As a comparison, we also use a uniform corner frequency to filter the data. After filtering, we can select the traces with correlation coefficients larger than 0.88 for vertical component and 0.75 for radial component. The number of selected traces for each event with uniform corner frequency filtering is different from that of the optimal frequency range selection algorithm. Comparing the results of waveform inversion obtained with these two different filtering operation is insightful. We perform full waveform inversion of both vertical and radial teleseismic P waveforms, low-pass filtered at 5 s. The uniform lower corner frequency is set to 0.01 Hz for vertical component and 0.01 Hz for radial component, while the uniform upper corner frequency is fixed at 0.1 Hz. Table 4.31 shows the number of vertical and radial component waveforms after selection following these two different procedures. The inverted 3D models obtained with these two slightly different datasets and frequency ranges will be shown next.

We perform the FWI with the iterative L-BFGS algorithm, starting from the initial smooth 1D model as in the resolution tests. For FWI performed on real dataset, regularization is introduced to stabilize the inverse problem. We add a penalty condition on the L_2 norm of the

Number of traces	Vertical after selection 1	Vertical after selection 2	Radial after selection 1	Radial after selection 2
2013 May 11	40	39	28	22
2013 May 24	44	43	44	41
2013 Aug 13	37	37	26	27
2013 Aug 30	35	35	29	32
2013 Sep 25	30	30	29	26
total number	186	184	156	148

Table 4.31: The number of available traces after different selection procedures. Selection procedure 1 is the optimal frequency range selection algorithm, selection procedure 2 is based on uniform corner frequency filtering.

Laplacian of the model to the waveform misfit function and define the new misfit function by (3.30). The penalty on the Laplacian of the model aims at smoothing the model update. It can also be seen as preconditioning the descent direction in an attempt to guide the inversion towards smoother models. Regularization will have a significant impact on the inversion results by adjusting the smoothing coefficient λ. The choice of λ is based on trial-and-error. Executing both forward and adjoint simulations for the five events only needs 12 minutes with 2560 processor cores. It is worth mentioning that the whole inversion can be run on a moderate size cluster with the total number of processors typically smaller than 500. In general, the convergence is achieved after 10 to 20 iterations which requires less than one day on a cluster with 512 processors. Therefore, we can test FWI with different smoothing coefficients and datasets at a moderate computational cost.

Figure 4.32 shows the surface projections of five parallel vertical sections with a length of 220 km along the trend of the western transect. The spacing between the five sections is 6 km. All the temporary stations on the profile are bounded by the outermost two lines. In the following, the inverted model will be shown along these vertical sections. The incoming azimuths of the five teleseismic P waves used in this study are shown by thick red lines and labeled according to the event dates.

The central vertical sections of the gradient for each event computed in the initial smooth 1D model are shown in Figures 4.33 and 4.34. The total gradients are obtained by summing all the event gradients.

We compare the gradients computed from datasets filtered at uniform corner frequencies and optimal corner frequencies. Note that the color scale for each event is different in Figure 4.33 and 4.34. The strength of the gradients for the different events depends on their waveform misfit and on the S/N ratio, and thus to first order on the magnitude of the event. The initial gradients for the same event but generated with different filters only show minor differences. The initial Vp gradients show more common features than the initial Vs gradients. After summing the event gradients, the total gradient will enhance these common features and cancel out the

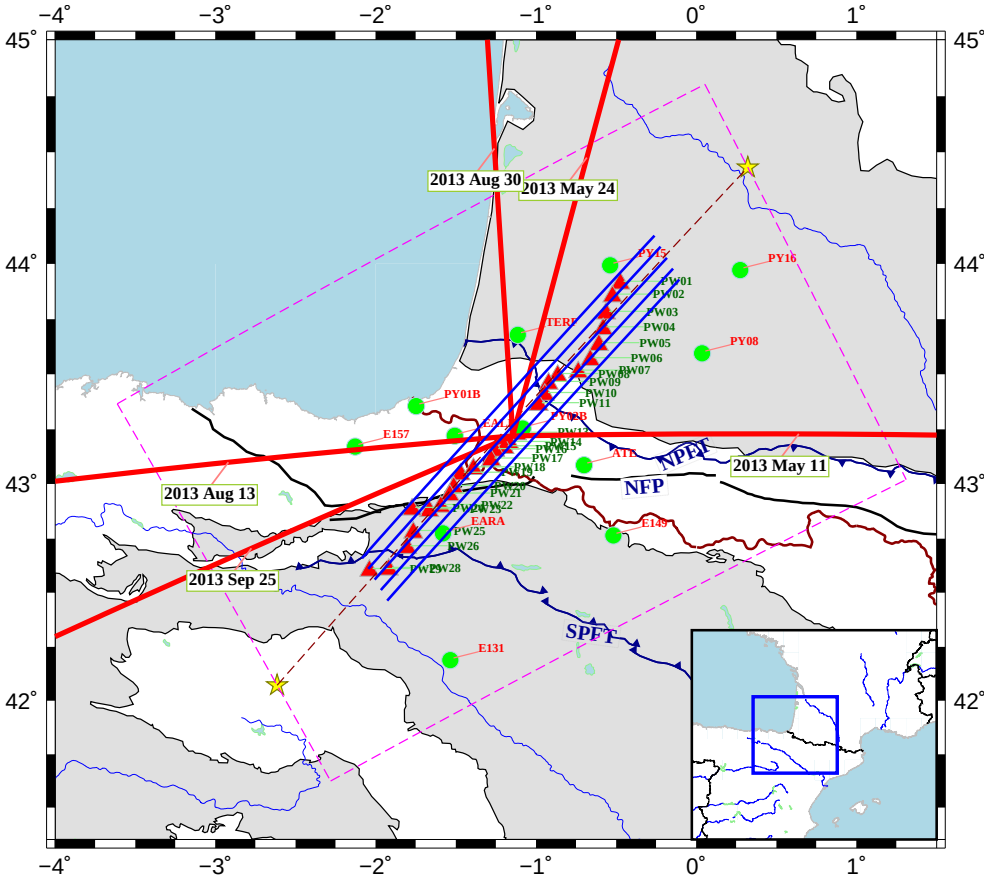
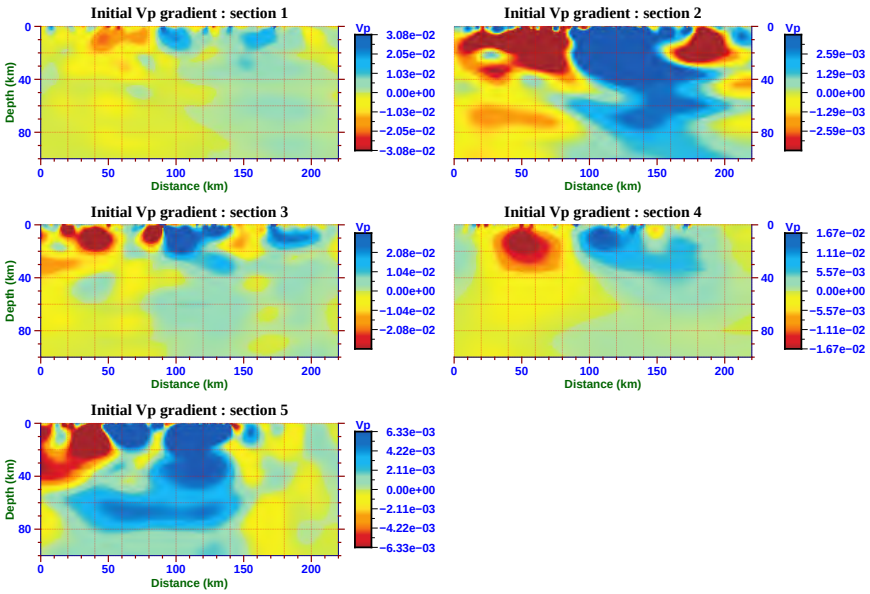


Figure 4.32: Map view of the parallel vertical sections along the trend of the western transect, shown with thick blue lines. The azimuths of the five earthquakes are shown with thick red lines.

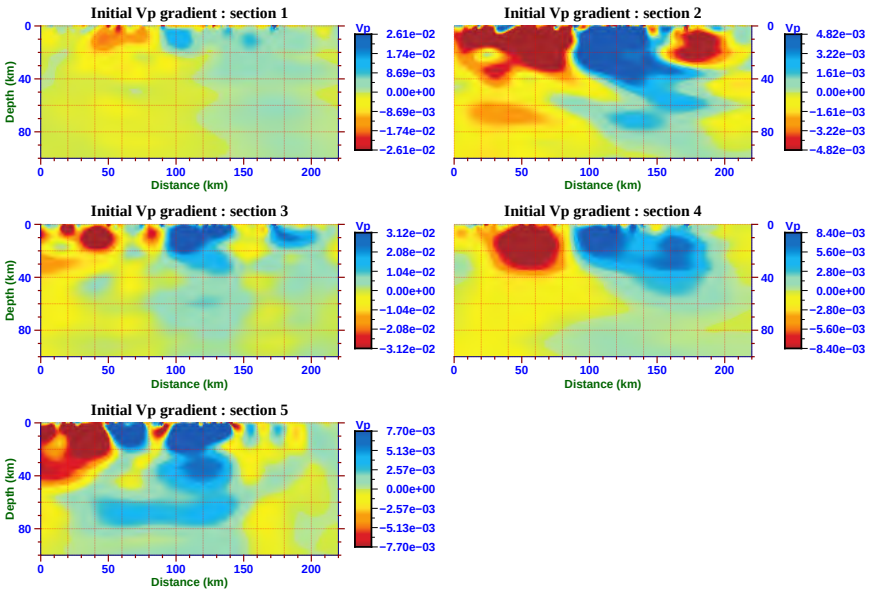
irrelevant details. The stronger values are mainly observed above 70 km depth, suggesting strong heterogeneities in the crust and the top of the upper mantle.

The total Vp and Vs gradients show different features. The Vp gradient is mainly dominated by the contribution of the direct P wave and thus mainly reflect the long wavelength structures of the velocity model. In contrast, the Vs gradient shows more sensitivity to seismic velocity jumps at seismological interfaces. The P-to-S conversion on the Moho, in particular, has a strong contribution to the gradient. Later arriving multiples also contribute to the retrieval of seismic interfaces. Including these back-scattered waves in the inversion significantly improves the resolution. When we consider longer periods, the Vs gradient carries information on large scale anomalies.

Another important feature in the Vp and Vs gradients is the strong amplification close to the surface. Because of the maximum amplitude cut-off in the color scales, these excessively strong gradients are not apparent in the plots. These very large values of the gradients close to the receivers mainly result from the geometrical spreading of the adjoint wavefields. These large values are balanced by the preconditioner which has small values close to the receivers. In our case, we recall that we use an inverse Hessian whose initial values contain the depth

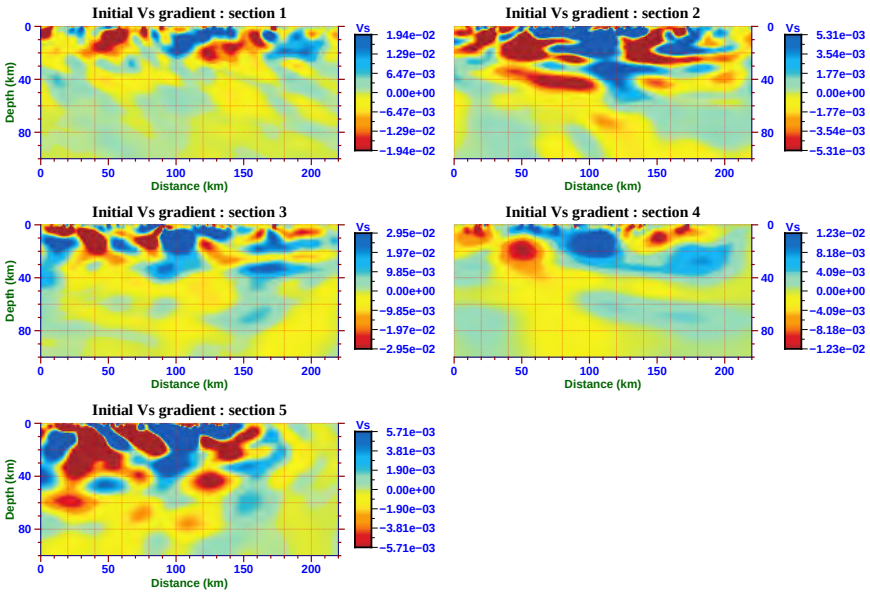


(a) Vp gradient of each event obtained with uniform corner frequency filter.

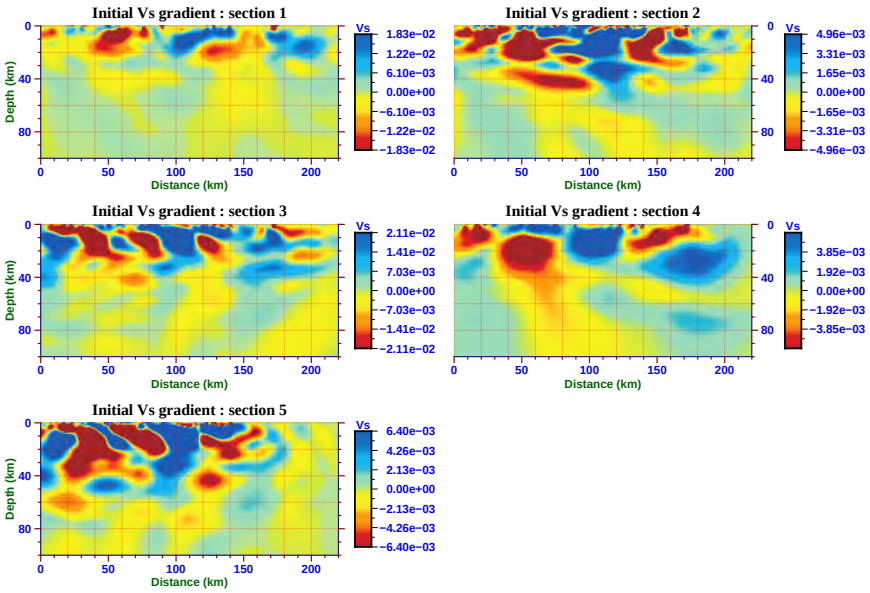


(b) Vp gradient of each event obtained with optimal frequency filter.

Figure 4.33: Vp gradient for each event computed in the initial smooth 1D model: (a) Using the dataset filtered at uniform corner frequencies. (b) Using the dataset filtered at optimal corner frequencies.

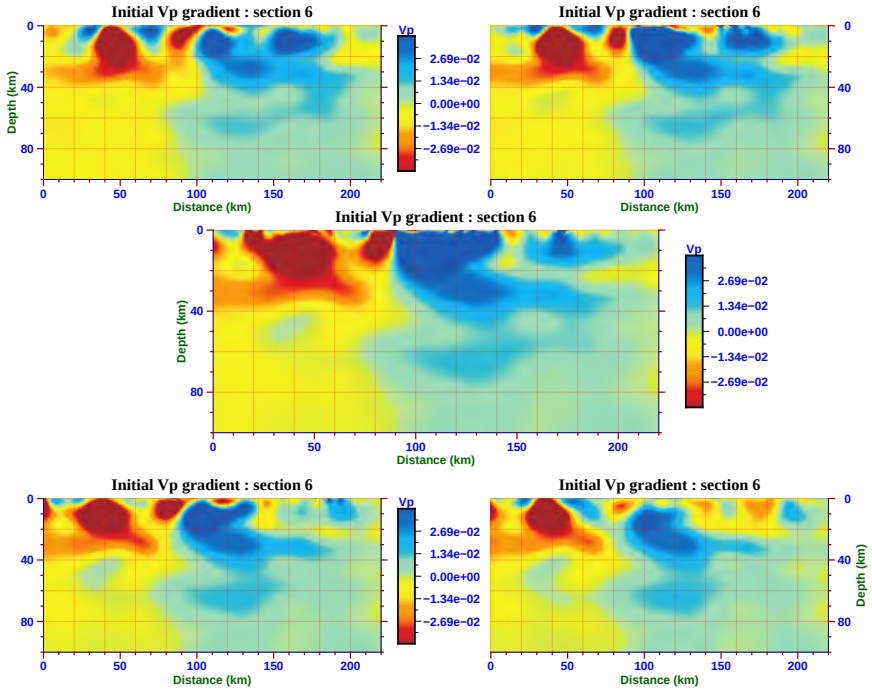


(a) Vs gradient of each event obtained with uniform corner frequency filter.

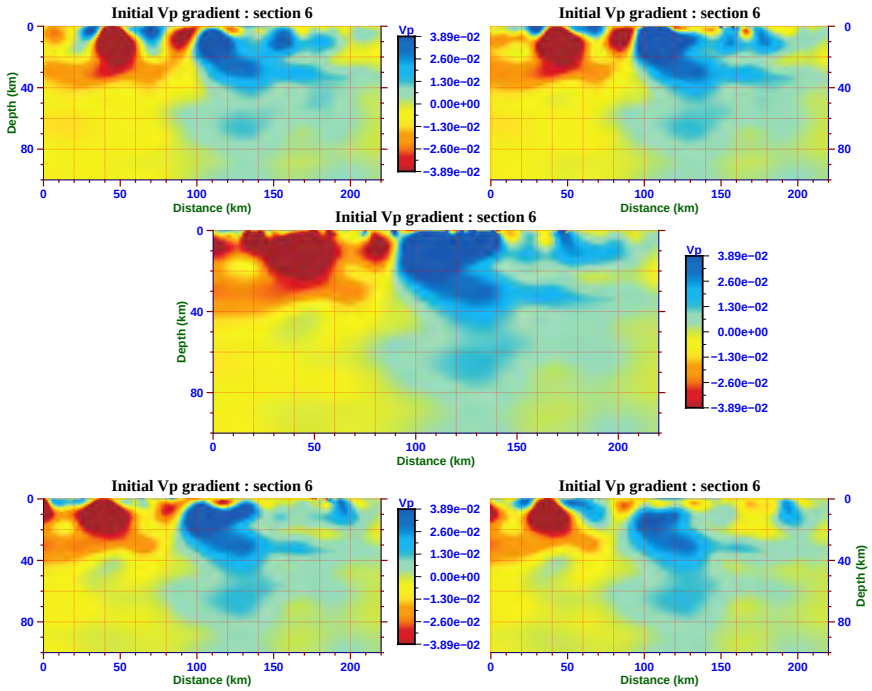


(b) Vs gradient of each event obtained with optimal frequency filter.

Figure 4.34: Vs gradient for each event computed in the initial smooth 1D model: (a) Using the dataset filtered at uniform corner frequencies. (b) Using the dataset filtered at optimal corner frequencies.

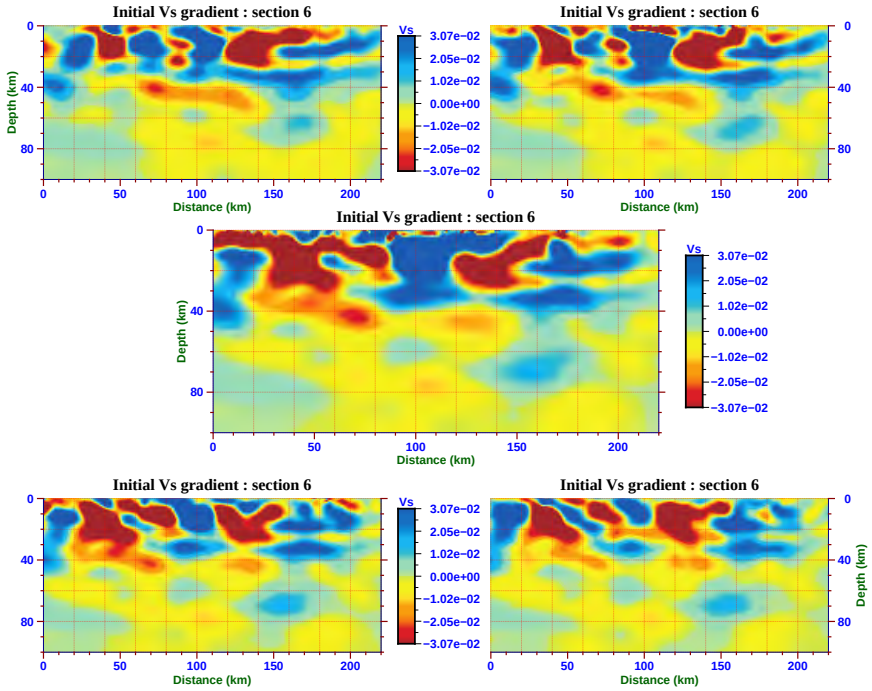


(a) Total Vp gradient with uniform corner frequency filter.

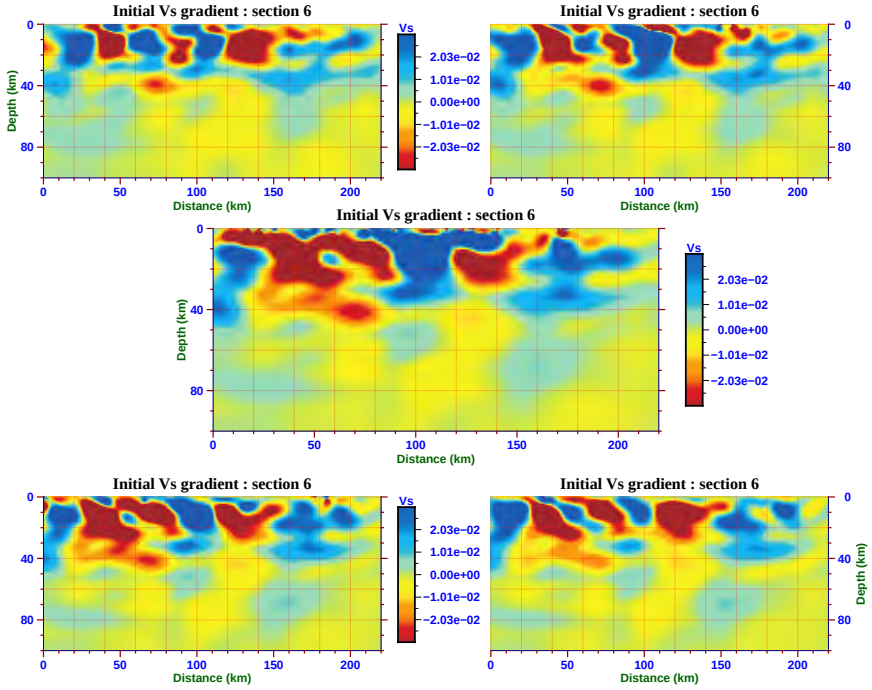


(b) Total Vp gradient with optimal frequency filter.

Figure 4.35: Vp gradients computed in the initial smooth 1D model shown in five parallel vertical sections: (a) Using the dataset filtered at uniform corner frequencies. (b) Using the dataset filtered at optimal corner frequencies.



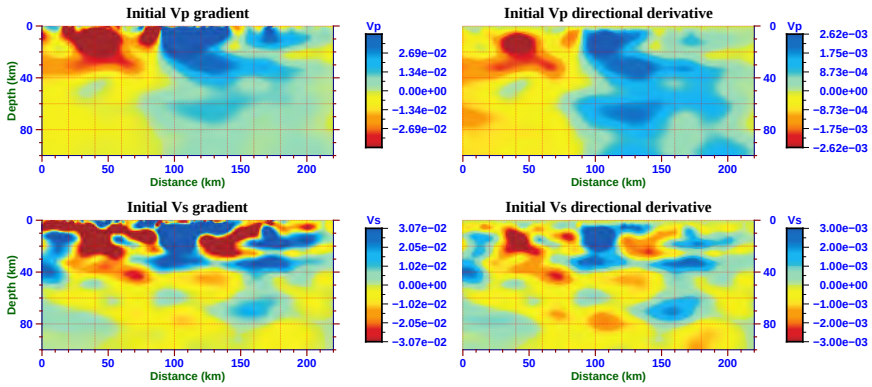
(a) Total Vs gradient with uniform corner frequency filter.



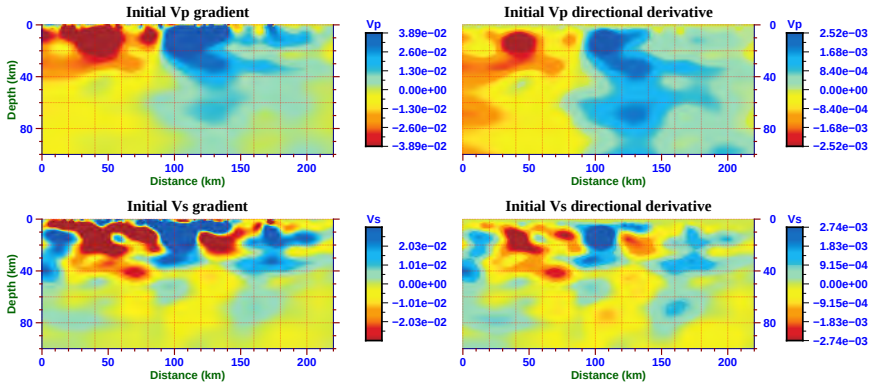
(b) Total Vs gradient with optimal frequency filter.

Figure 4.36: Vs gradients computed in the initial smooth 1D model shown in five parallel vertical sections: (a) Using the dataset filtered at uniform corner frequencies. (b) Using the dataset filtered at optimal corner frequencies.

of each grid point in the initial 1D model. After preconditioning, the directional derivatives are defined. Their values in the deep parts of the model are therefore enhanced. The central vertical sections of total gradients and directional derivatives computed in the initial 1D model are shown in Figure 4.37.



(a) Comparison between total gradient and directional derivative computed with uniform corner frequency filter.



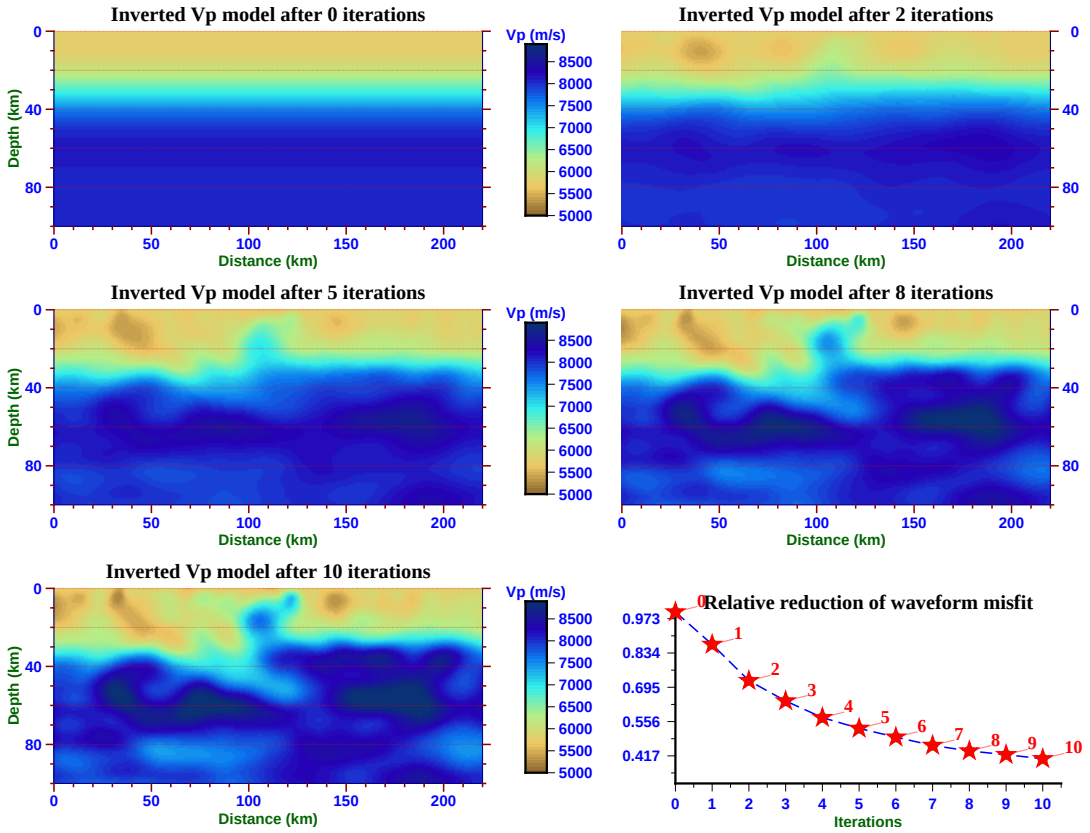
(b) Comparison between total gradient and directional derivative computed with optimal corner frequency filter.

Figure 4.37: Comparison between total gradient and directional derivative: (a) Using the dataset filtered at uniform corner frequencies. (b) Using the dataset filtered at optimal corner frequencies.

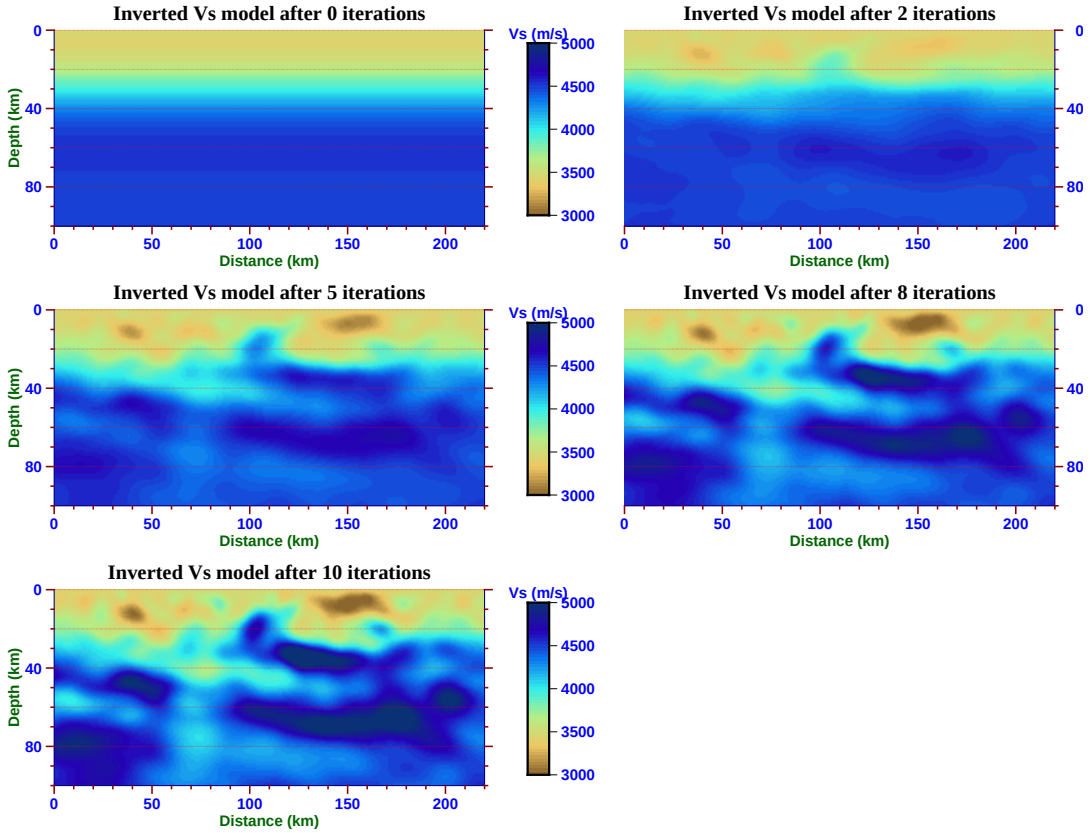
We first perform the FWI on the dataset filtered at uniform corner frequencies and using different smoothing coefficients λ. Since the wavelength of compressional waves is almost twice the wavelength of shear waves, we set the smoothing coefficient λ_s for Vs model at twice the value of the smoothing coefficient λ_p for Vp model.

When a small smoothing coefficient $\lambda_p = 0.1$ is used, we obtain the final Vp and Vs models after 10 iterations of our waveform inversion algorithm. Figure 4.38 shows the convergence history for Vp and Vs models with this small smoothing coefficient.

The waveform relative misfit residual for FWI with $\lambda_p = 0.1$ after 10 iterations is 0.41. We test two other smoothing coefficient $\lambda_p = 0.35$ and $\lambda_p = 0.5$. Both inversions converged



(a) Vp model update.



(b) Vs model update.

Figure 4.38: The Vp (a) and Vs (b) models obtained after 0,2,5,8,10 iterations of the L-BFGS algorithm.

after 9 iterations. The relative misfit residual for FWI with intermediate smoothing coefficient $\lambda_p = 0.35$ and FWI with large smoothing coefficient $\lambda_p = 0.5$ are 0.44 and 0.5, respectively. Figure 4.39 shows the final Vp and Vs models for FWI with these three different smoothing coefficient λ_p.

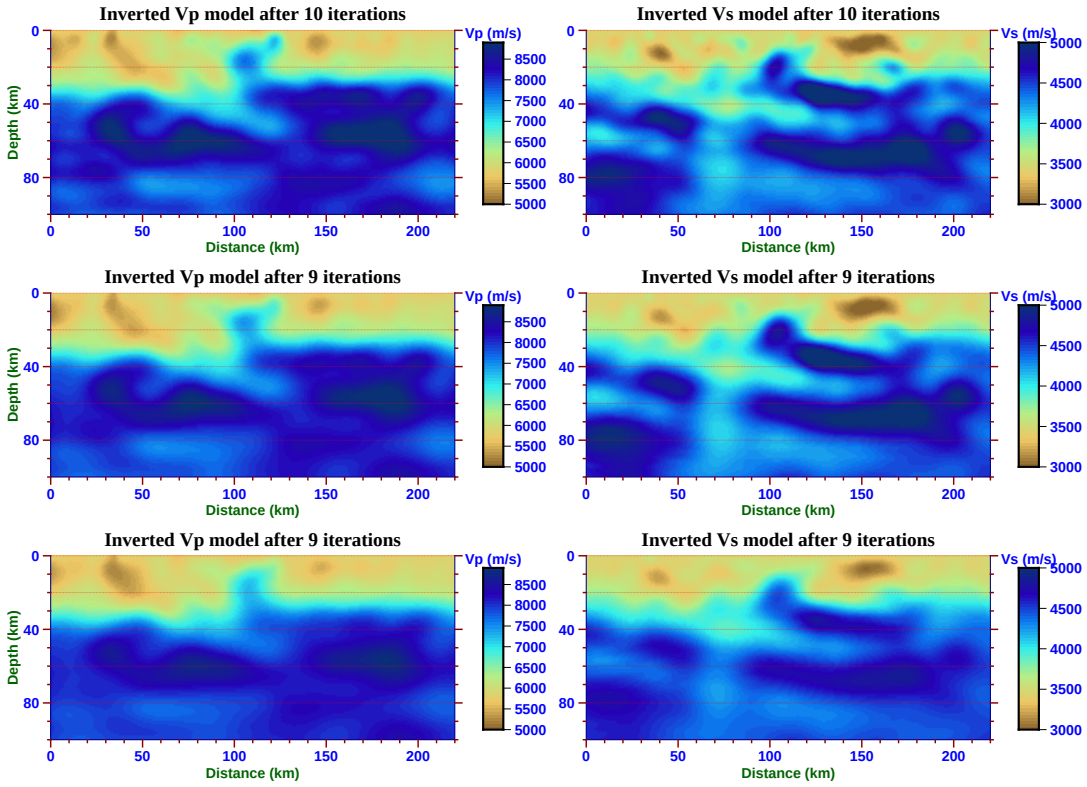


Figure 4.39: Final Vp and Vs models obtained by Full waveform inversion using uniform corner frequency filter with smoothing coefficient λ_p equals to 0.1 (top panels), 0.35 (middle panels) and 0.5 (bottom panels), respectively.

Since the final velocity models with these different smoothing coefficients possess similar features and their waveform misfit reductions are similar, we prefer to use the smoothest Vp and Vs model from FWI obtained with the largest smoothing coefficient $\lambda_p = 0.5$ as our best inverted models when using the dataset filtered at uniform corner frequencies.

In the following, we perform the FWI based on the dataset with optimal corner frequency filter and testing different smoothing coefficient λ. We show the FWI results for this dataset with three different smoothing coefficient λ_p equal to 0.1, 0.2 and 0.5. Because the dataset is changed, even if we use the same smoothing coefficient to regularize the inversion as before we will obtain different results. The inversions with the three different smoothing coefficients converged after 10, 9 and 7 iterations, respectively. The relative misfit residual for FWI obtained with the three different λ_p are 0.31, 0.34 and 0.52, respectively. Figure 4.40 shows the final Vp and Vs model for FWI with these different smoothing coefficients.

The models with large smoothing coefficient $\lambda_p = 0.5$ are obviously too smooth. On the

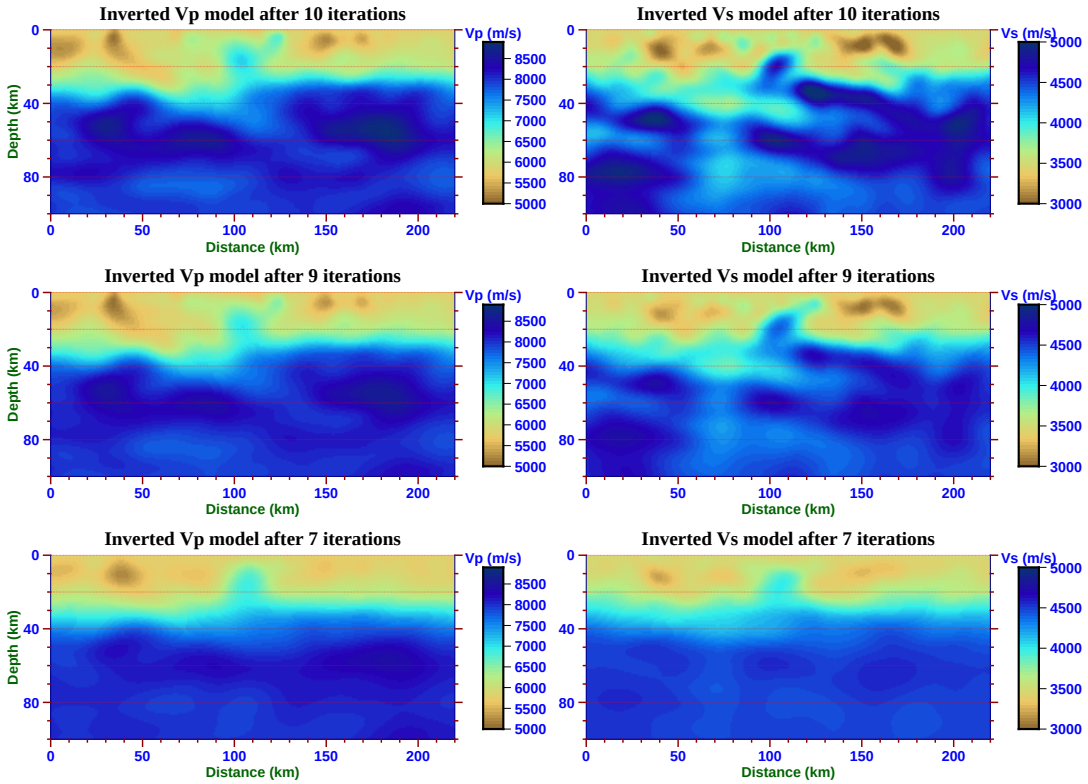


Figure 4.40: Final Vp and Vs models obtained by Full waveform inversion using optimal corner frequency filer and using smoothing coefficient λ_p equal to 0.1 (top panels), 0.2 (middle panels) and 0.5 (bottom panels), respectively.

other hand, the models obtained with a small smoothing coefficient $\lambda_p = 0.1$ contain too many details. Therefore, we retain the model obtained with the intermediate smoothing coefficient $\lambda_p = 0.2$. The resolution tests and inverted models all show that the resolved anomalies are mainly located above 70 km depth. Consequently, we suspect that the waveform inversion may be limited to a smaller region corresponding to the top 70 km depth. We use the dataset with optimal corner frequency filter and keep the intermediate smoothing coefficient $\lambda_p = 0.2$ for this small region inversion. The waveform inversion converged after 10 iterations. The relative misfit residual for this special case is 0.337, which is close to the value obtained with the FWI in the full region. The comparison between final models from FWI in the full region and that in the small region is shown in top and middle panels of Figure 4.41. The models show common features in the crust and upper mantle, confirming our conjecture. We also perform the inversion in full region starting from the inverted small region 3D models. The inversion also converged after 10 iterations. The final models are also shown in the bottom panels of Figure 4.41. The final waveform relative misfit residual for this special hierarchical inversion is 0.24, even lower than the relative misfit residual in the full region inversion performed with a small smoothing coefficient $\lambda_p = 0.1$. This larger waveform misfit reduction by hierarchical inversion presents the nonlinear nature of the waveform inversion problem. Different convergent

models may provide the similar level of waveforms matching.

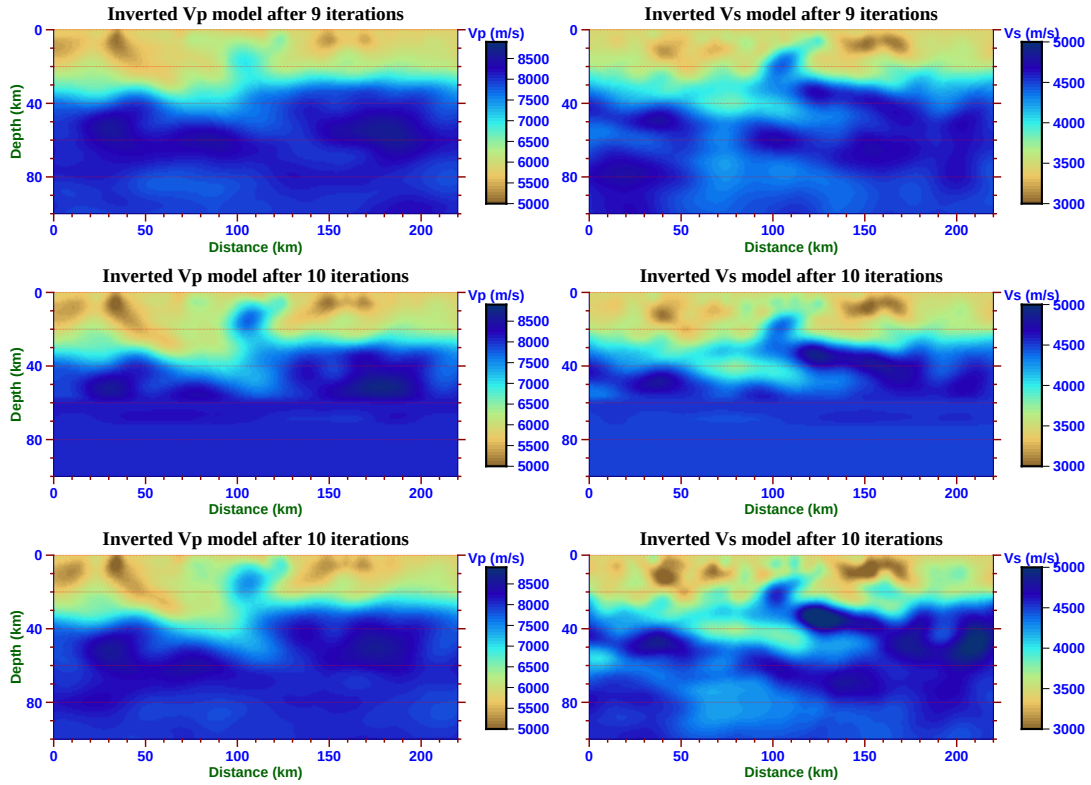
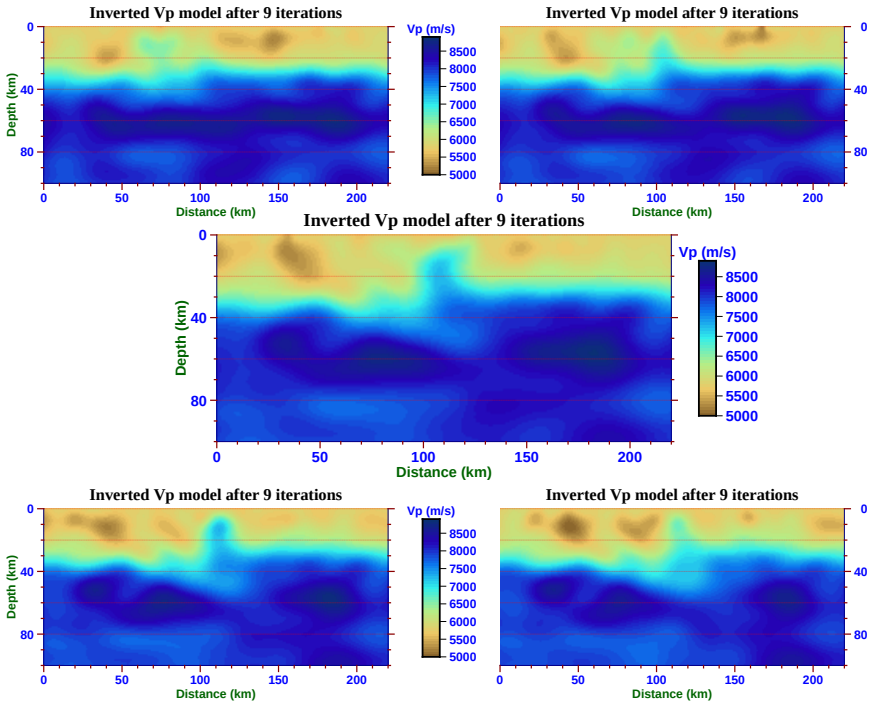


Figure 4.41: Final Vp and Vs models obtained by Full waveform inversion using optimal frequency filter and using intermediate smoothing coefficient $\lambda_p = 0.2$ in full region (top panels), in region above 70 km depth (middle panels) and in full region starting from 3D models shown in middle panels (bottom panels), respectively.

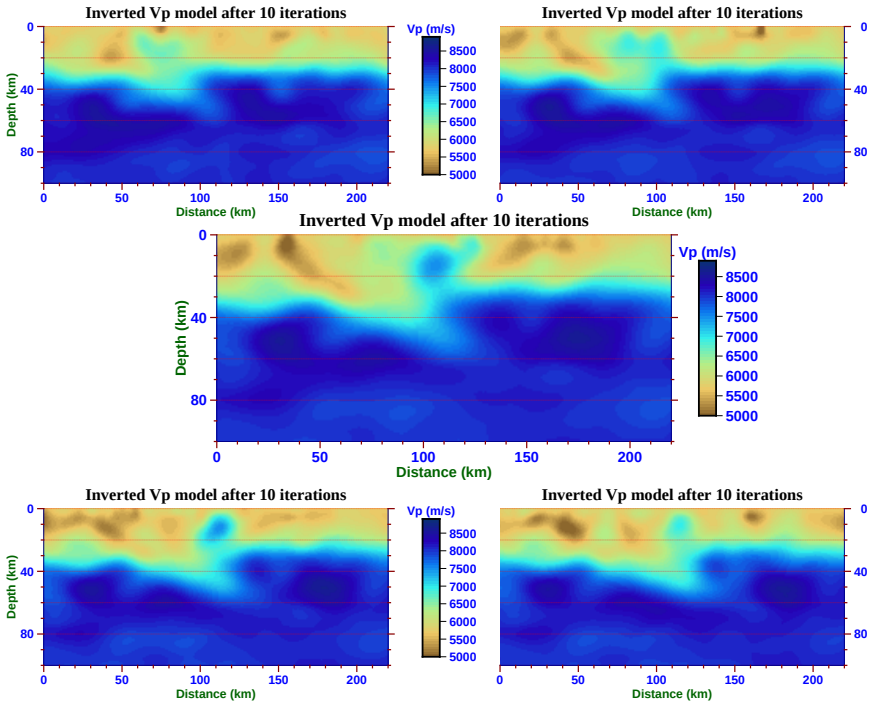
Comparing the models directly obtained from FWI in the full region with intermediate smoothing coefficient and the models from hierarchical FWI, the latter shows more distinct crustal features. Figure 4.42 and 4.43 compare the 5 parallel vertical sections in the best Vp and Vs models obtained by filtering with uniform corner frequency and the models obtained from hierarchical FWI with optimal corner frequency filter. Both models show similar geological structures. But the preferable models are obtained with uniform corner frequency filter. These models are smoother and simpler. We will retain these models for geological interpretation.

Table 4.44 shows the vertical and radial component waveform misfits of each event before FWI and the waveform relative misfit residual after FWI with uniform corner frequency and with optimal frequency filter. The comparisons between observed and synthetic seismograms along the transect computed in the starting 1D model and the two sets of final 3D models are shown in Figures 4.45 to 4.54.

Despite a higher noise level on the horizontal component records, the improvement of waveform fits is more important on the radial component because the synthetic radial components computed in the starting smooth 1D model only contain the contribution of the direct P wave. The final model better explains not only the phase and amplitude of the direct P waves but

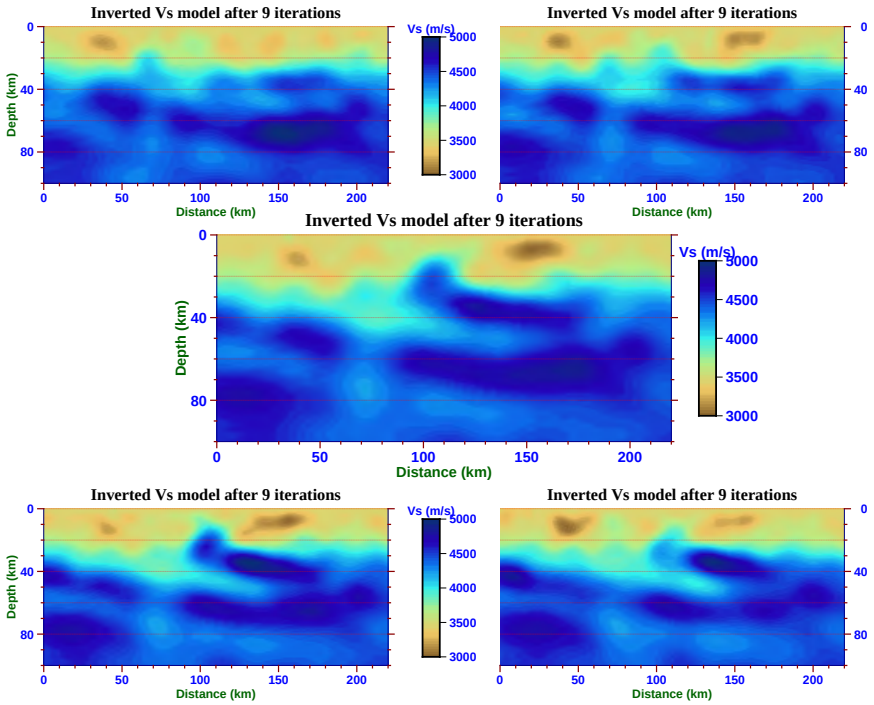


(a) Best Vp model obtained with uniform corner frequency filter.

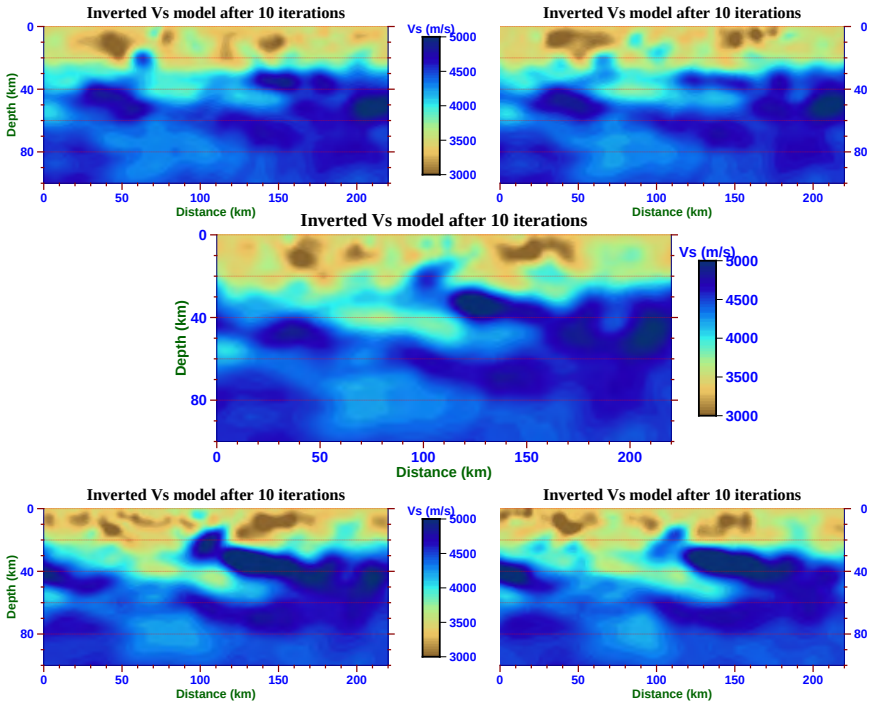


(b) Vp model from hierarchical FWI obtained with optimal frequency filter.

Figure 4.42: Vp model obtained with uniform corner frequency filter (a) and from hierarchical FWI with optimal frequency filter(b).



(a) Best Vs model obtained with uniform corner frequency filter.



(b) Vs model from hierarchical FWI obtained with optimal frequency filter.

Figure 4.43: Vs model obtained with uniform corner frequency filter (a) and from hierarchical FWI with optimal frequency filter(b).

waveform misfit	initial misfit: dataset 1	relative misfit residual: dataset 1	initial misfit: dataset 2	relative misfit residual: dataset 2
2013 May 11	198.7615	0.5694	88.0939	0.2614
2013 May 24	74.1804	0.2896	64.7238	0.0838
2013 Aug 13	53.2831	0.3620	48.6452	0.1619
2013 Aug 30	24.5602	0.4006	25.1188	0.3482
2013 Sep 25	51.7806	0.5834	49.9159	0.4602

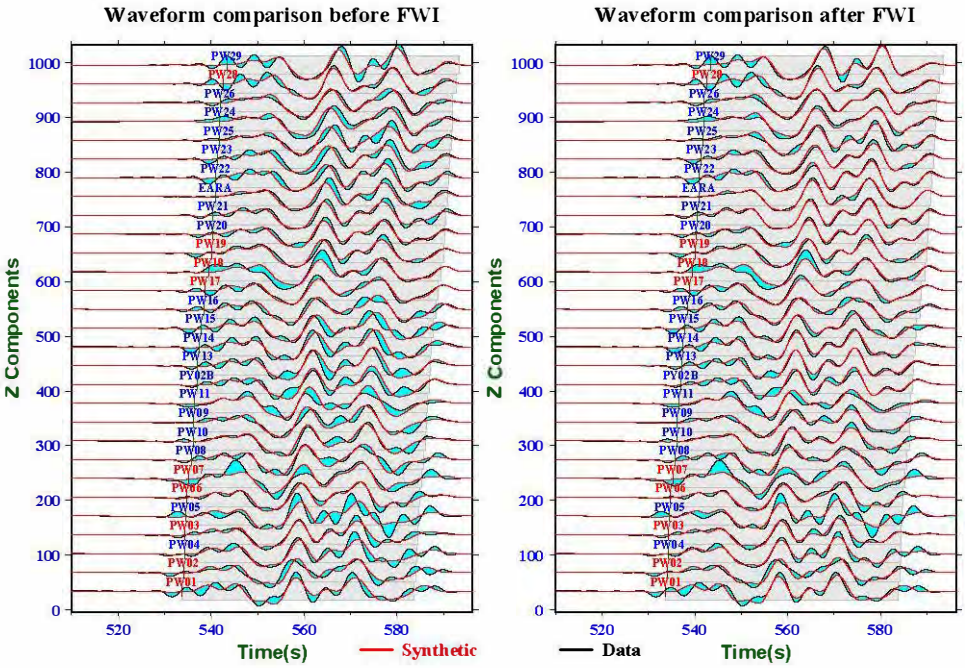
(a) Vertical component waveform misfit and relative misfit residual

waveform misfit	initial misfit: dataset 1	relative misfit residual: dataset 1	initial misfit: dataset 2	relative misfit residual: dataset 2
2013 May 11	96.1262	0.8039	15.5614	0.4460
2013 May 24	62.5674	0.2762	37.9310	0.1406
2013 Aug 13	58.0054	0.4581	17.3595	0.1404
2013 Aug 30	45.4038	0.4664	14.7589	0.3998
2013 Sep 25	22.8885	0.3872	19.4884	0.3145

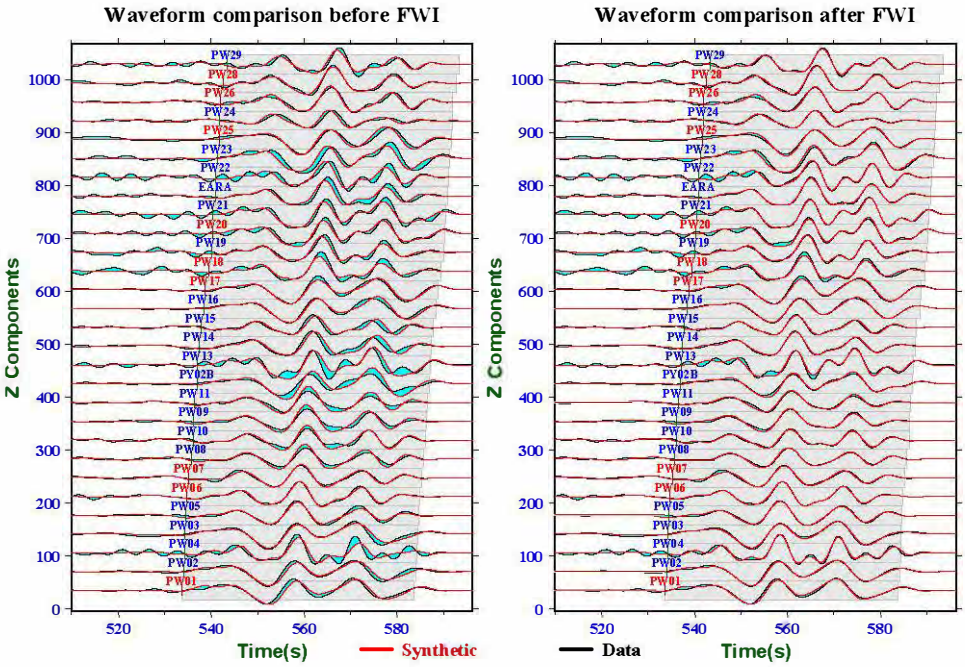
(b) Radial component waveform misfit and relative misfit residual

Table 4.44: Waveform misfits and relative misfit residuals for each event. Dataset 1: dataset with uniform corner frequency filter, dataset 2: dataset with optimal corner frequency filter. (a) Vertical component (b) Radial component.

also of all the converted and multiply reflected waves that arrive in their coda, mainly observed on the radial components. As we predict, the quality of the fit changes from event to event. For example, it is worst for event 1, which has the smallest magnitude (6.1), while it is best for event 2, which has the largest magnitude (8.3). Another significant feature is that after optimal frequency range selection, the fit of filtered waveforms for each event is improved, especially on the radial component as can be seen in Figure 4.46. The relative misfit residual with optimal frequency range selection is nearly half the one observed in the case of uniform frequency range selection.

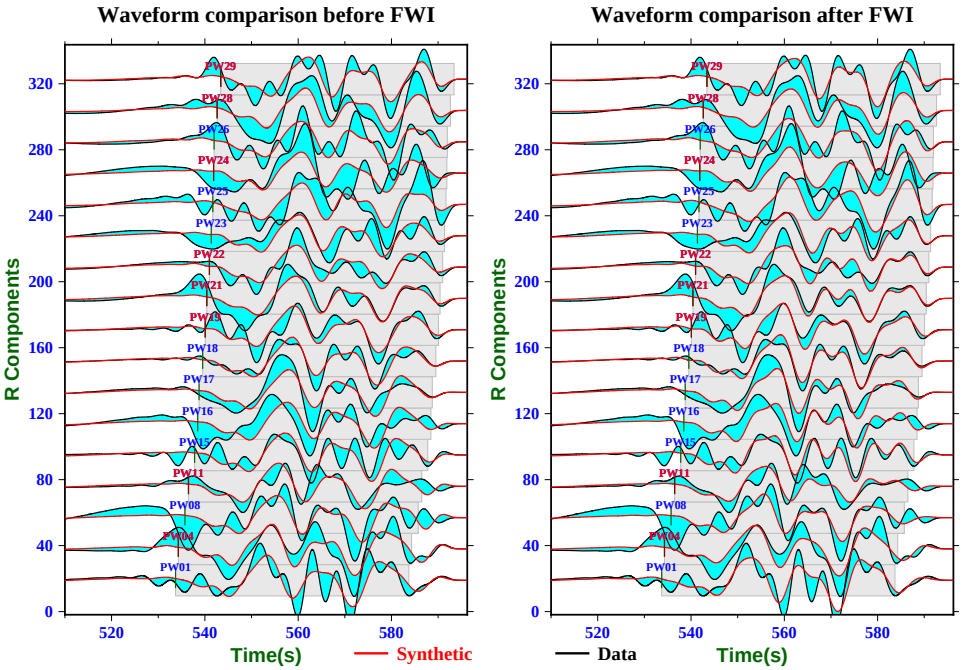


(a) vertical component waveforms along the transect for FWI case 1

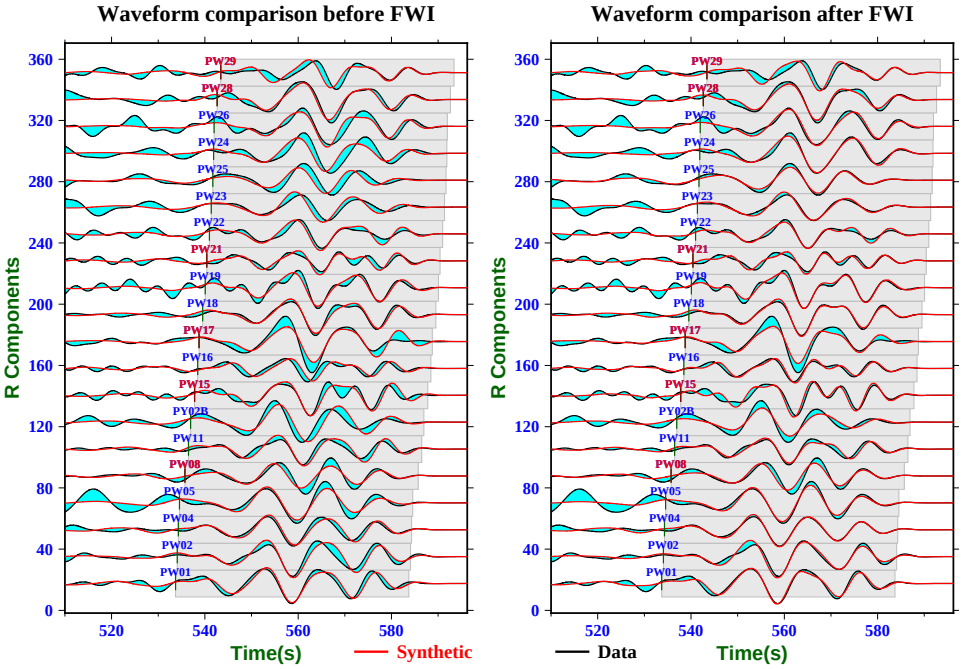


(b) vertical component waveforms along the transect for FWI case 2

Figure 4.45: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 May 11 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter. Left panel: Synthetic seismograms computed in the smooth initial 1D model. Right panel: Synthetic seismograms computed in the final 3D model.

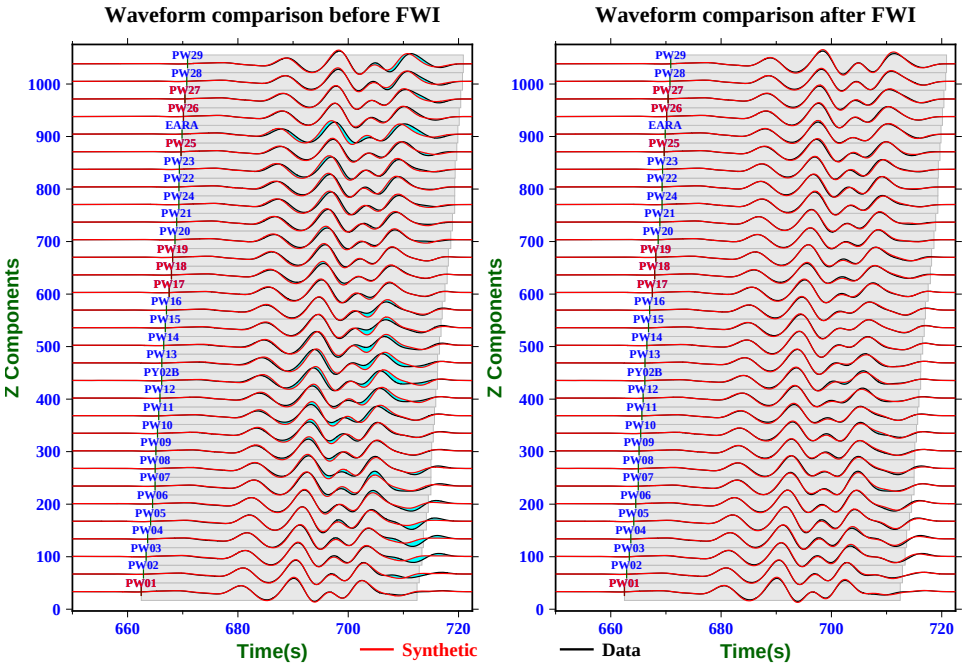


(a) radial component waveforms along the transect for FWI case 1

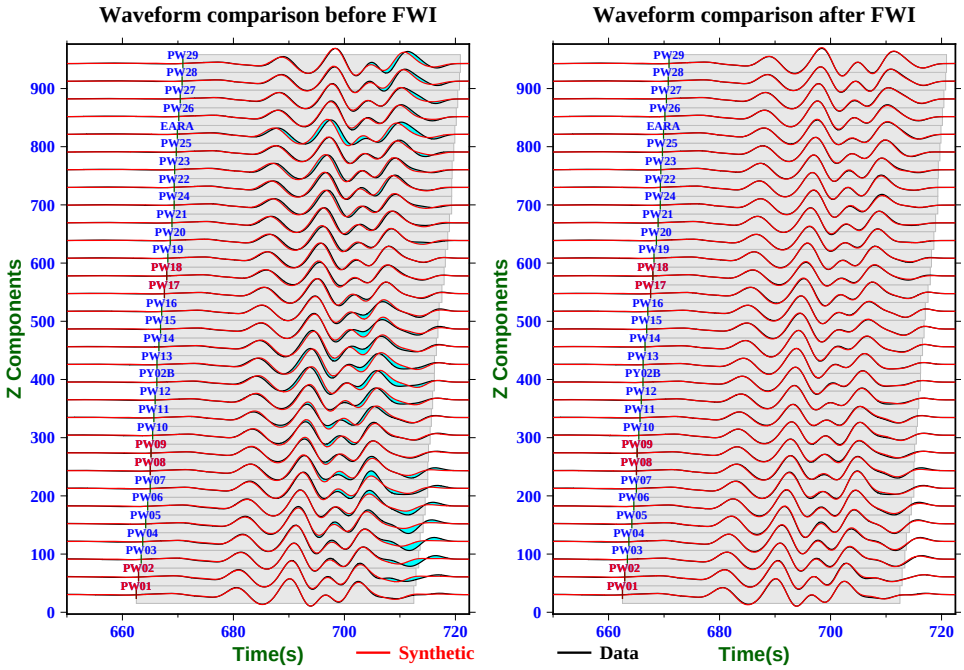


(b) radial component waveforms along the transect for FWI case 2

Figure 4.46: Observed (black lines) and synthetic (red lines) radial component seismograms for the 2013 May 11 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

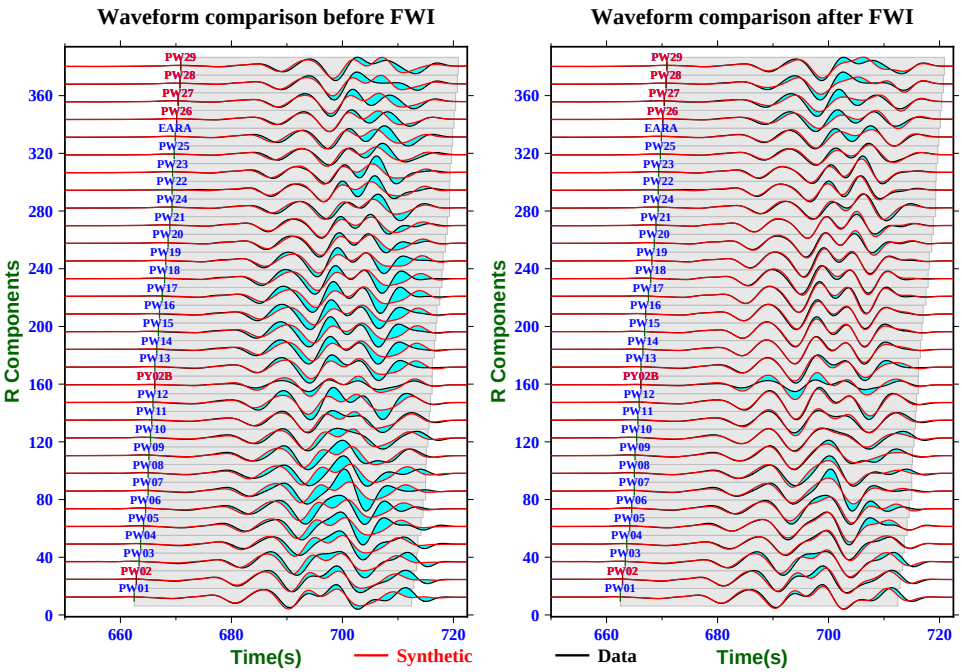


(a) vertical component waveforms along the transect for FWI case 1

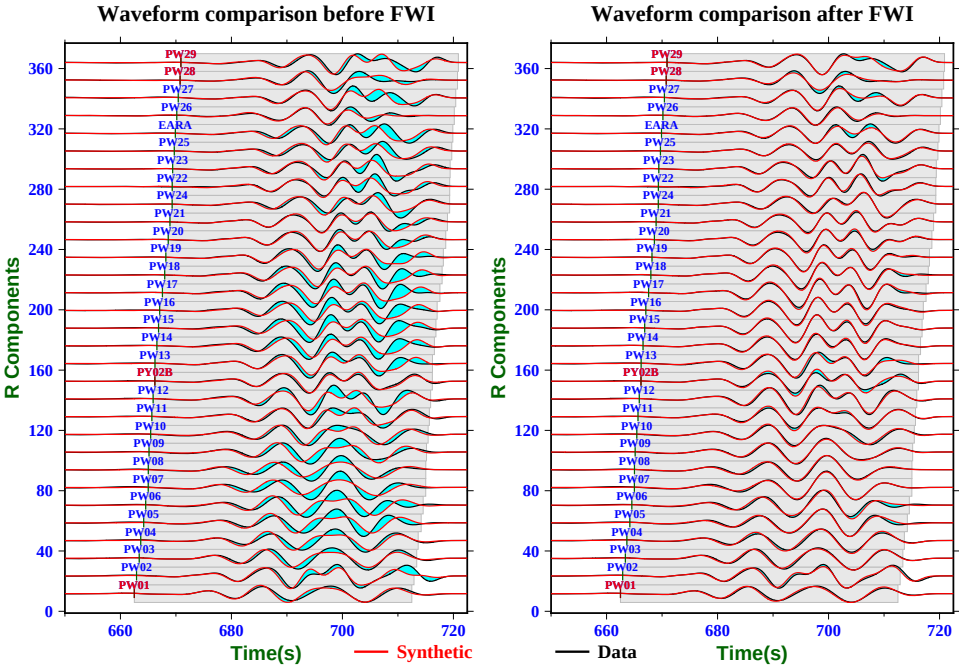


(b) vertical component waveforms along the transect for FWI case 2

Figure 4.47: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 May 24 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

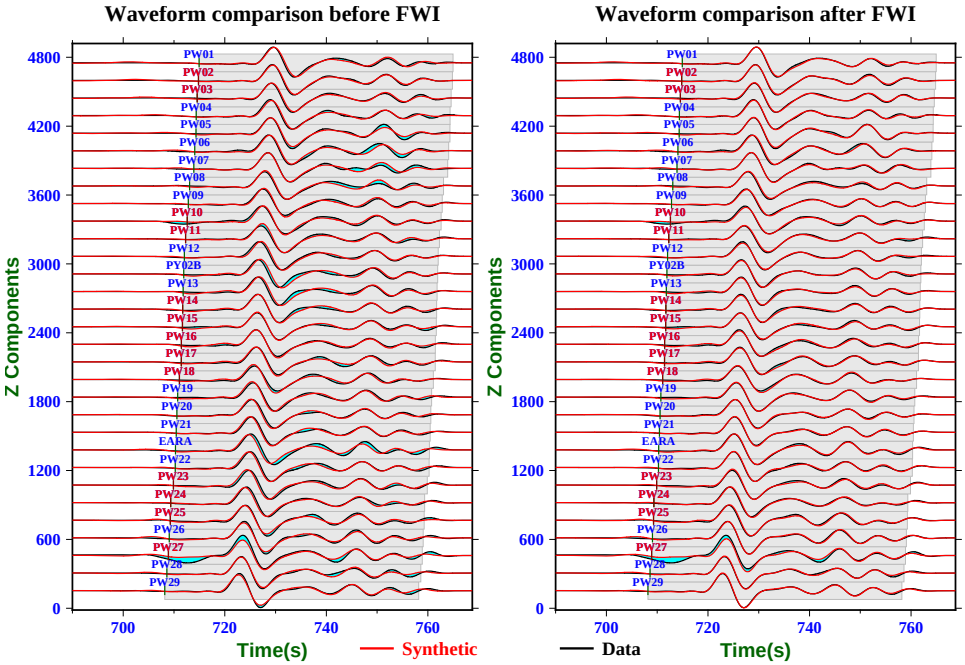


(a) radial component waveforms along the transect for FWI case 1

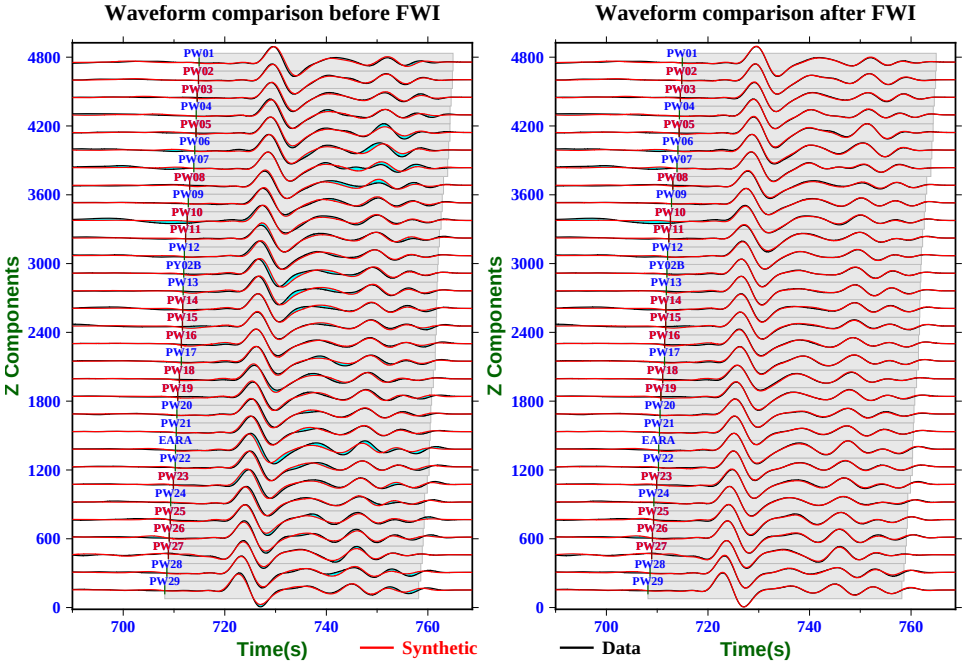


(b) radial component waveforms along the transect for FWI case 2

Figure 4.48: Observed (black lines) and synthetic (red lines) radial component seismograms for the 2013 May 24 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

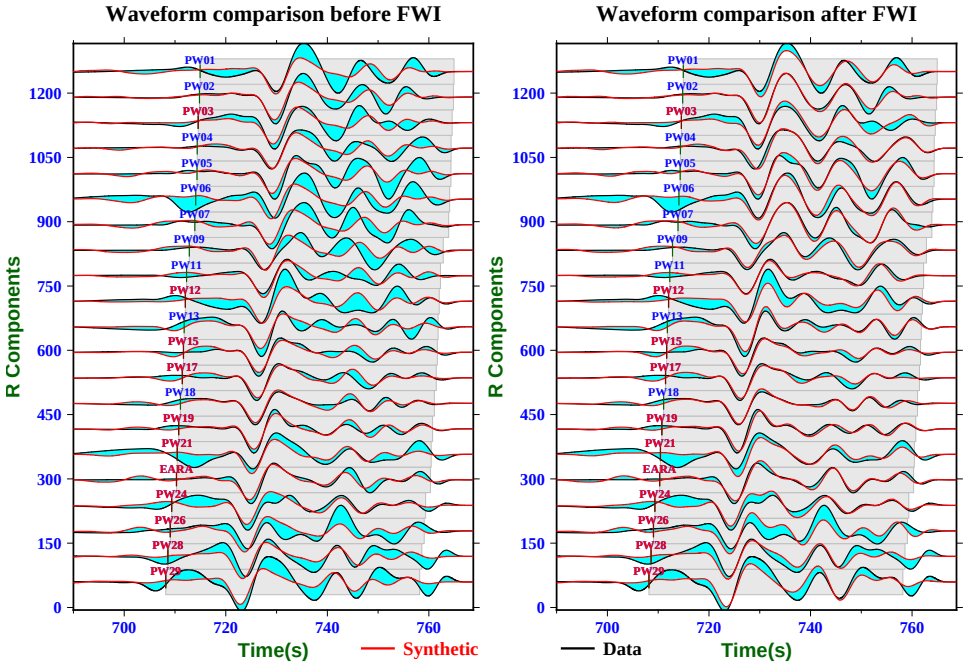


(a) vertical component waveforms along the transect for FWI case 1

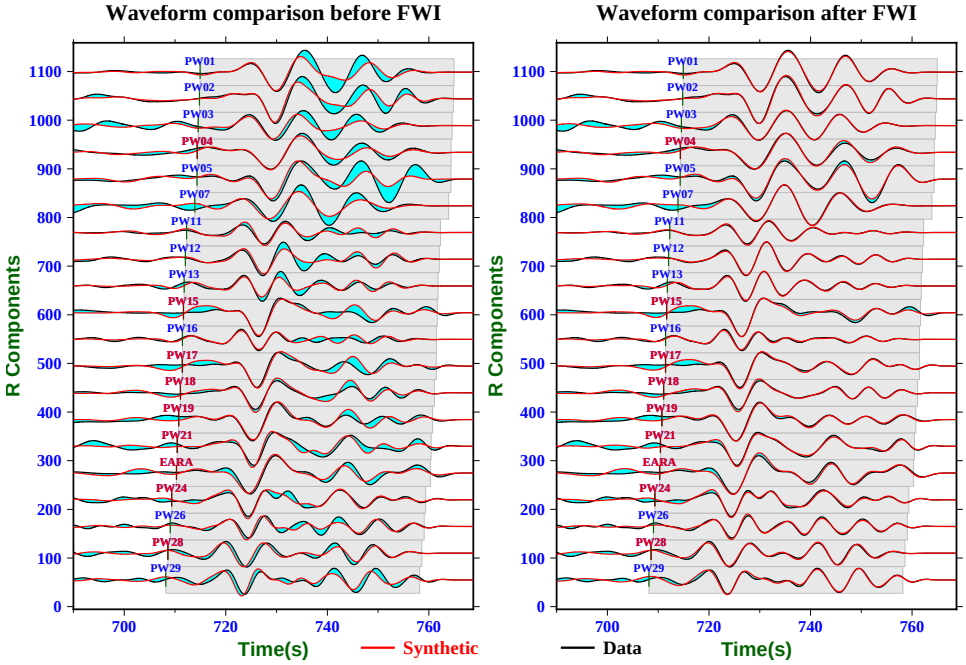


(b) vertical component waveforms along the transect for FWI case 2

Figure 4.49: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 Aug 13 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

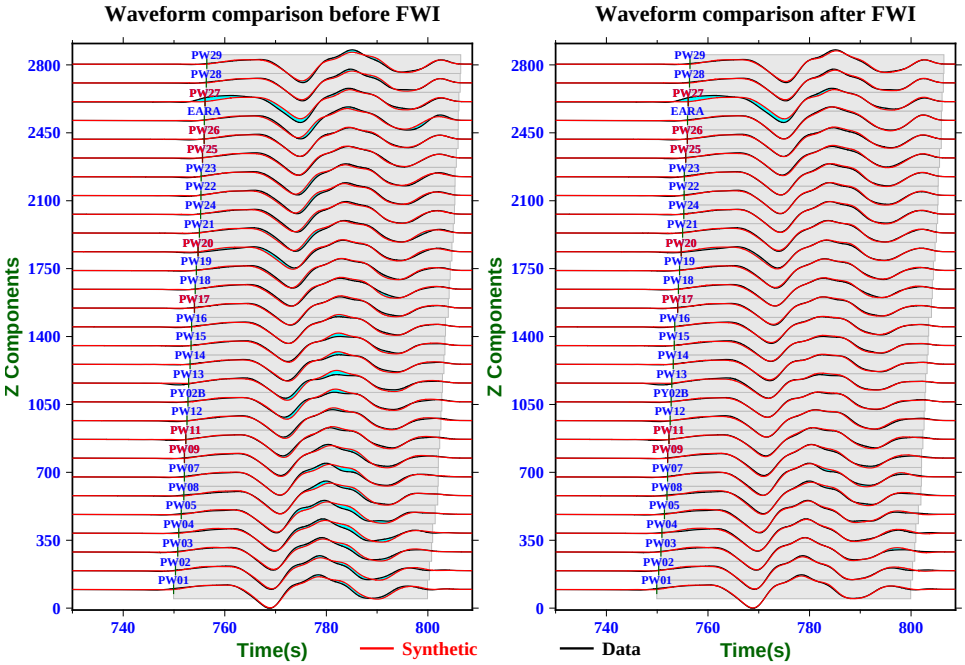


(a) radial component waveforms along the transect for FWI case 1

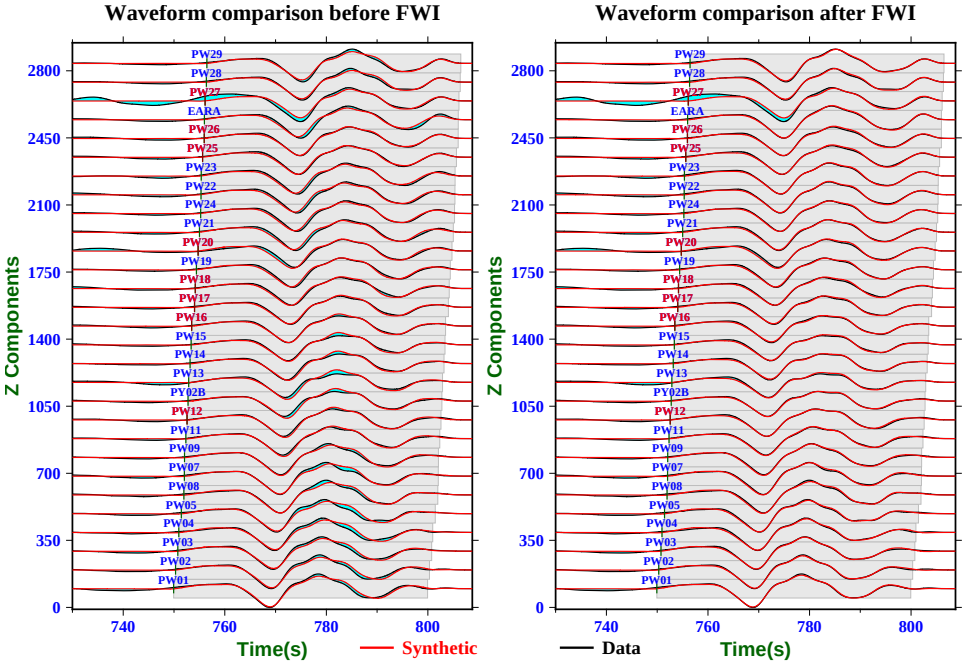


(b) radial component waveforms along the transect for FWI case 2

Figure 4.50: Observed (black lines) and synthetic (red lines) radial component seismograms for the 2013 Aug 13 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

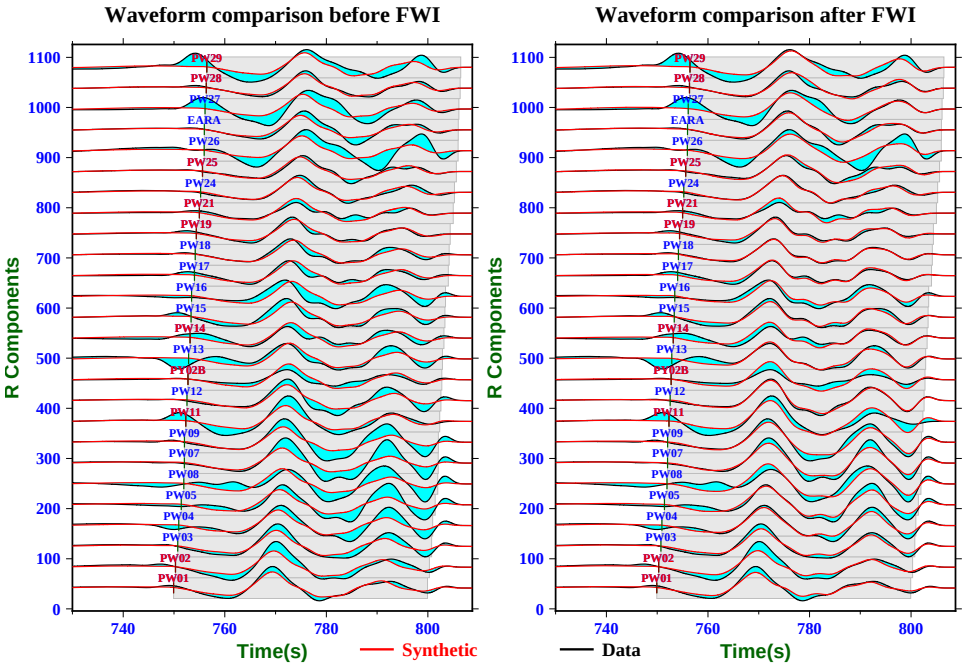


(a) vertical component waveforms along the transect for FWI case 1

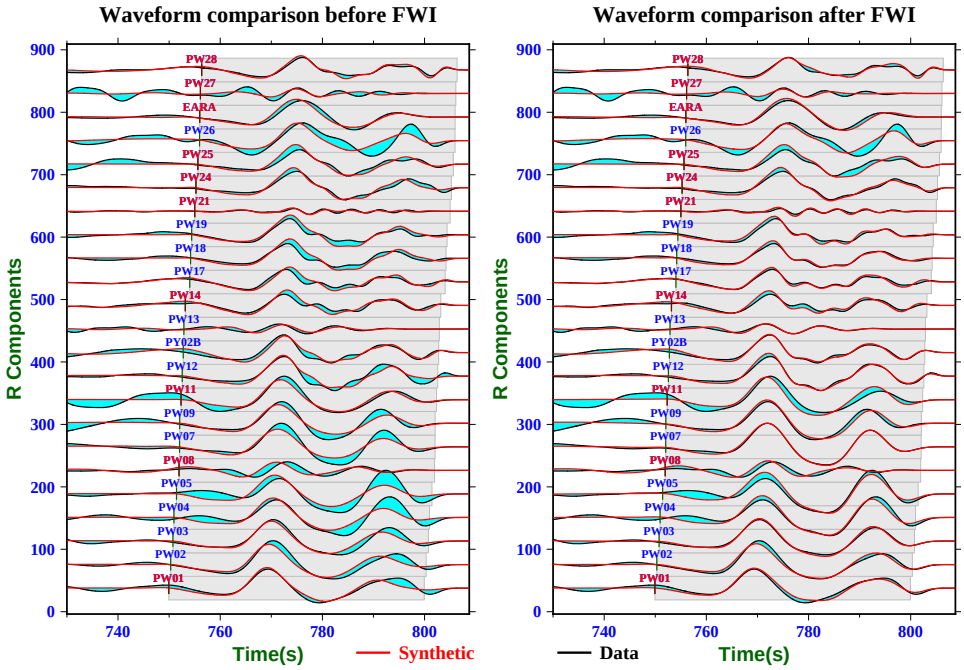


(b) vertical component waveforms along the transect for FWI case 2

Figure 4.51: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 Aug 30 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

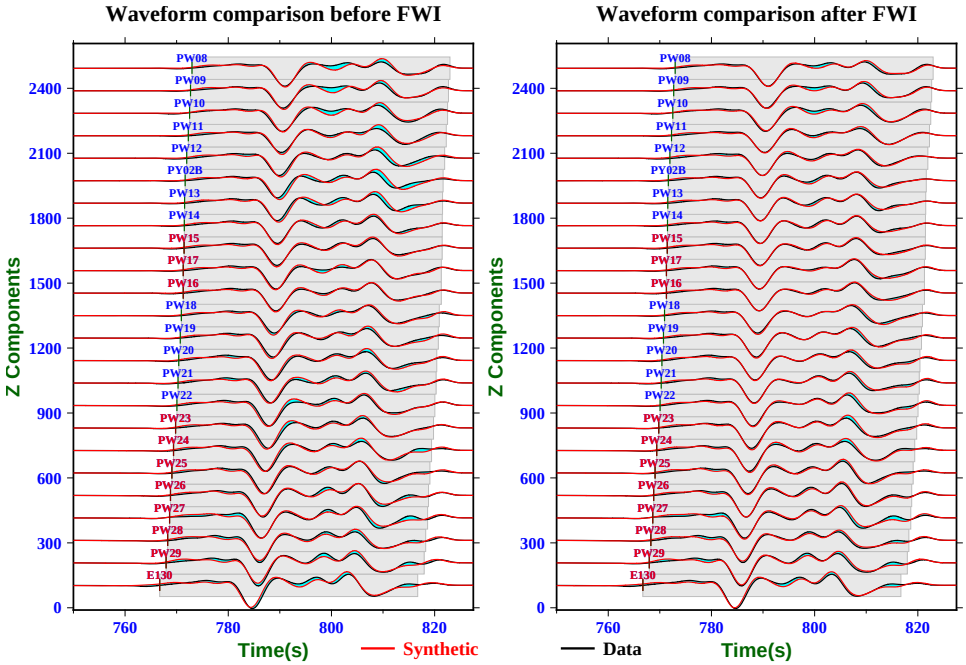


(a) radial component waveforms along the transect for FWI case 1

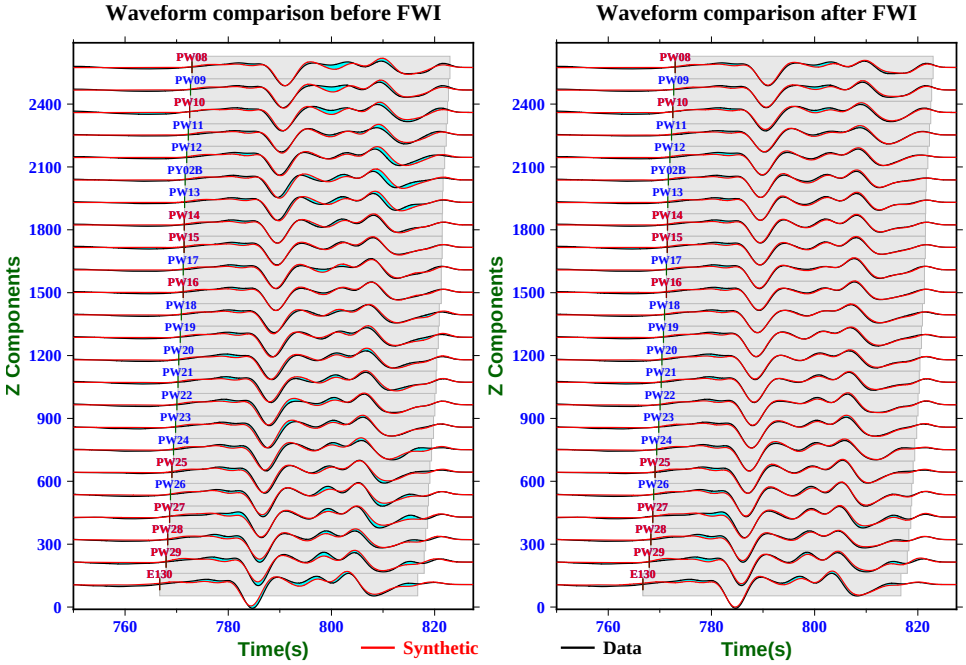


(b) radial component waveforms along the transect for FWI case 2

Figure 4.52: Observed (black lines) and synthetic (red lines) radial component seismograms for the 2013 Aug 30 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

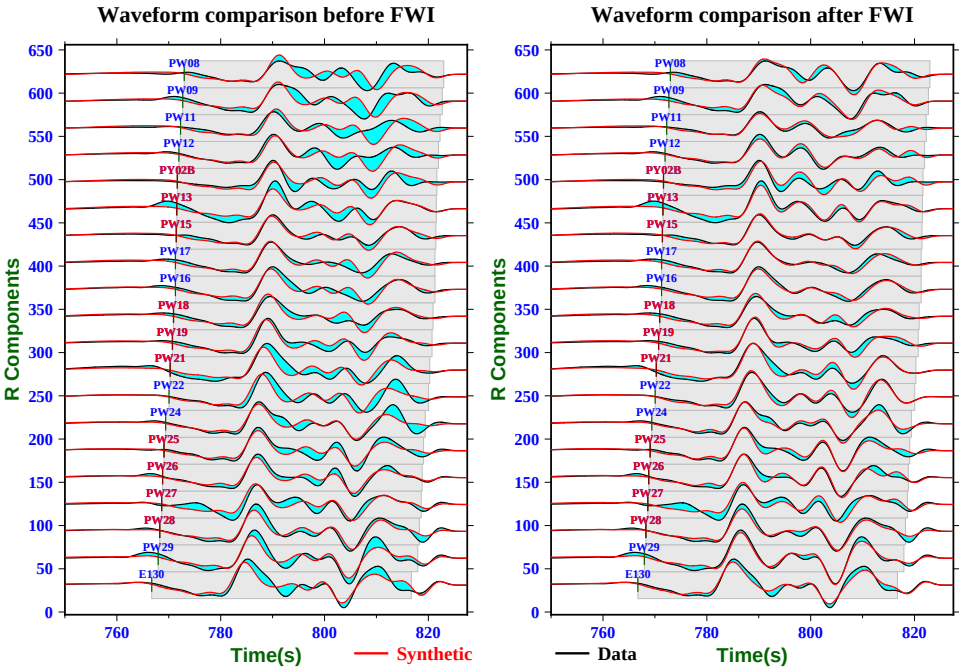


(a) vertical component waveforms along the transect for FWI case 1

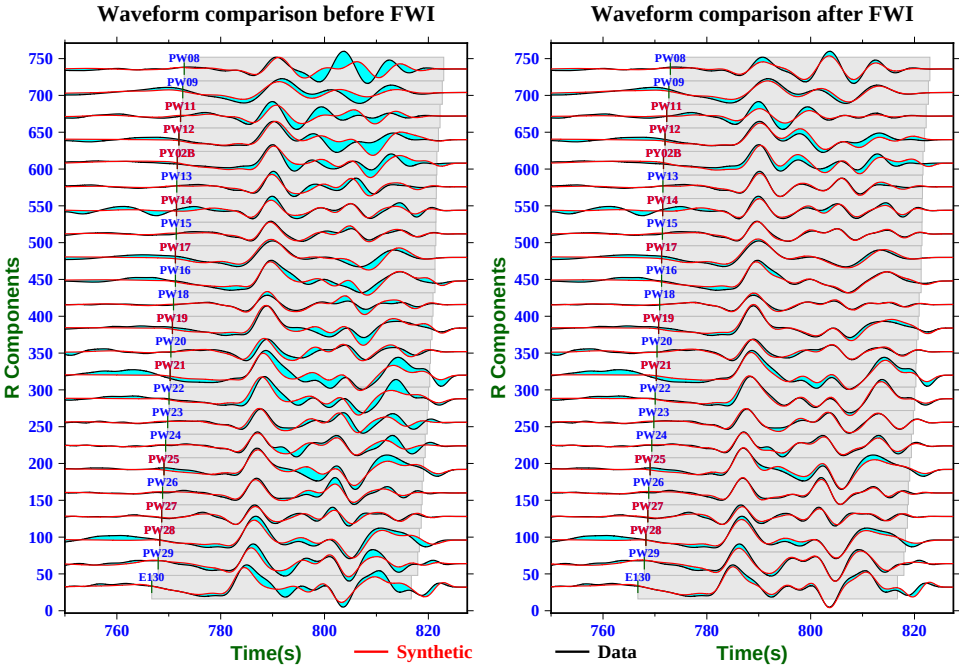


(b) vertical component waveforms along the transect for FWI case 2

Figure 4.53: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 Sep 25 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.



(a) radial component waveforms along the transect for FWI case 1



(b) radial component waveforms along the transect for FWI case 2

Figure 4.54: Observed (black lines) and synthetic (red lines) radial component seismograms for the 2013 Sep 25 event. The shaded region represent the limits of the time windows that are used in the waveform inversion. (a): FWI case 1: best models obtained with uniform corner frequency filter; (b): FWI case 2: models from hierarchical FWI obtained with optimal frequency filter.

4.3.2 Comparison between different inversion strategies

In this section, we compare the results obtained from different inversion strategies. Two types of frequency range selection and filtering (mentioned in Table 4.31) are used in these tests. We first use the optimal frequency filter and corresponding dataset to implement the phase adjoint tomography and full waveform inversions with different time windows. Then the dataset selected from filtering with uniform corner frequency is used to compare the waveform inversion results with and without including the topography of the free surface.

4.3.2.1 Comparison between adjoint tomography and full waveform inversion

Over the past two decades, the phase (travel time) adjoint tomography (PAT) has gained popularity owing to its better resolution potential compared to classical ray tomography. It exploits the information of phase arrival time which is robust and quasi-linearly related to the perturbations of seismic velocities inside the Earth. Since both travel time tomography and full waveform inversion use the adjoint state method to compute the gradient, it is meaningful to investigate whether full waveform inversion can offer significant improvement over phase adjoint tomography.

We implemented PAT to minimize the phase misfit between the direct P wave in the observed and synthetic vertical component waveforms with a time window starting 10 s before the theoretical P arrival time and ending 20 s after. We filter the seismograms with optimal corner frequencies and start the iterative inversion from the same initial smooth 1D model used in previous full waveform inversion. In PAT, we simultaneously update the Vp and Vs models, because vertical component waveforms are also sensitive to the Vs model. After 11 iterations, the inversion converged towards a model that provides a travel time misfit reduction of 92%. The final velocity models are shown in the top panels of Figure 4.55. We then performed FWI on the same dataset as PAT. The algorithm converged after 9 iterations, with a waveform misfit reduction of 61%. The model obtained by full waveform inversion of only vertical components is shown in the middle panels of Figure 4.55. For comparison, we also perform a waveform inversion on both the vertical and radial components with the same 30 s long time window. The inversion converged after 9 iterations. Its waveform misfit reduction for the vertical and radial component are 61 and 46%, respectively. The inverted models are shown in the bottom panels of Figure 4.55.

We inspect the waveform fits for the vertical components within the short 30 s long time window between observed and synthetic seismograms before and after inversions. The waveform misfit reduction for PAT is only 3% compared to 61% for waveform inversion. Table 4.56 shows the vertical component waveform misfits and relative misfit residuals of each event for PAT and FWI using only the vertical components in this short time window. In FWI, the 2013 May 24 and Aug 13 event have the largest waveform misfit reduction value of 75 and 74%, respectively.

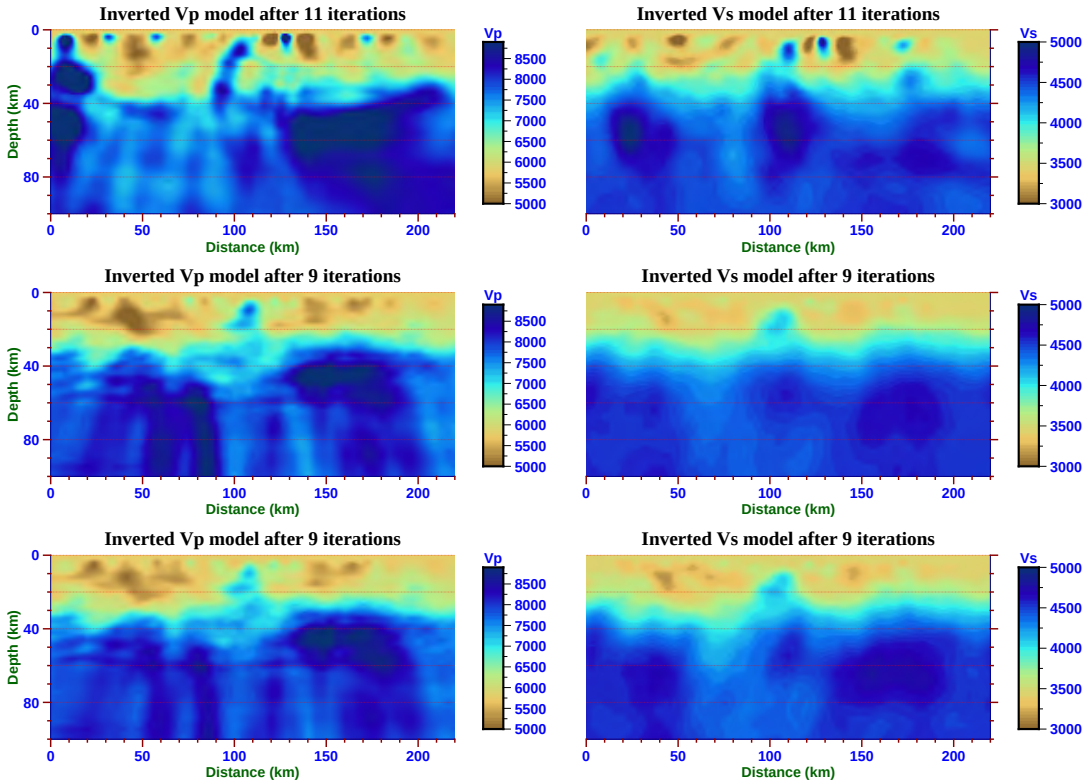


Figure 4.55: Final Vp and Vs models obtained by adjoint tomography (filtered at 5 s in the top panels), full waveform inversion using the same dataset as PAT tomography (middle panels) and full waveform inversion using both the vertical and radial components with the same 30 s long time window (bottom panels), respectively.

For PAT, the waveform misfit reductions for these two events are 35 and -50%. The negative value of the misfit reduction means that the waveform misfit even increases after inversion.

waveform misfit	initial misfit: PAT inversion 1	relative misfit residual: PAT inversion 1	initial misfit: FWI short TW	relative misfit residual: FWI short TW
2013 May 11	30.9301	0.747283	30.9301	0.561526
2013 May 24	24.555	0.654152	24.555	0.2508
2013 Aug 13	20.1124	1.50813	20.1124	0.240564
2013 Aug 30	15.1571	0.831851	15.1571	0.342744
2013 Sep 25	18.6252	1.33147	18.6252	0.590184

Table 4.56: Waveform misfits and relative misfit residuals for each event for PAT and FWI using only the vertical components in the short time window.

The comparisons of observed and synthetic seismograms along the transect computed in the starting 1D model and in two sets of final models for these two events are shown in Figures 4.57 and 4.58. The seismograms computed in the final FWI models clearly offer a better match of the waveforms than PAT. In contrast, the waveform misfits for some traces after PAT even increase beyond the range of the time window used in the inversion (for example for station

PW01, PW03, etc.). Since PAT does not utilize the information contained in the amplitude of waveforms, it fails to predict the later secondary arrivals after the primary P phase.

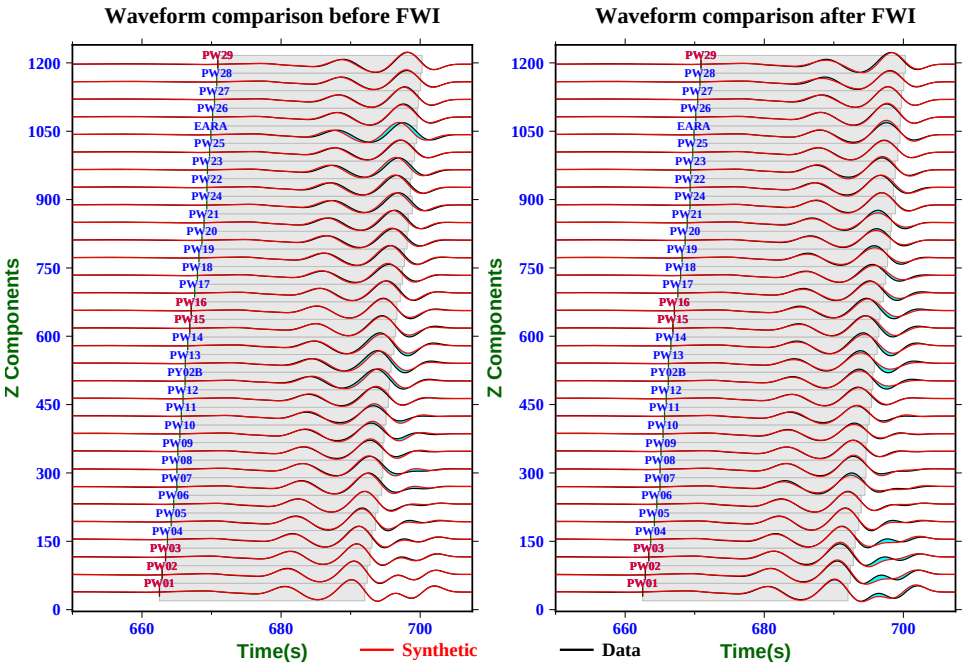
Compared to the models obtained in the previous section (Figure 4.42 and 4.43), both PAT and FWI have a poor spatial resolution. They only detect a slow structure in the shallow crust beneath the central part of the transect.

When the forward modeling can be precisely performed, the resolution of PAT mainly relies on data coverage, data quality, and the frequency content of the inverted wavefield. In our application, we have optimized the data selection procedure to determine the optimal frequency content of the wavefield for inversion based on coherence. Therefore, the poor resolution for lithospheric structures probably results from the insufficient data coverage since only a small number of teleseismic P wave records are used. Indeed, the checkerboard tests for assessing the resolution of tomographic models obtained by adjoint tomography, which have been performed successfully in 2D (Tape et al., 2007) and 3D (Fichtner et al., 2009), showed that a successful adjoint tomography is critically controlled by data coverage. The central shallow region in our images where the resolution is acceptable has the denser ray coverage.

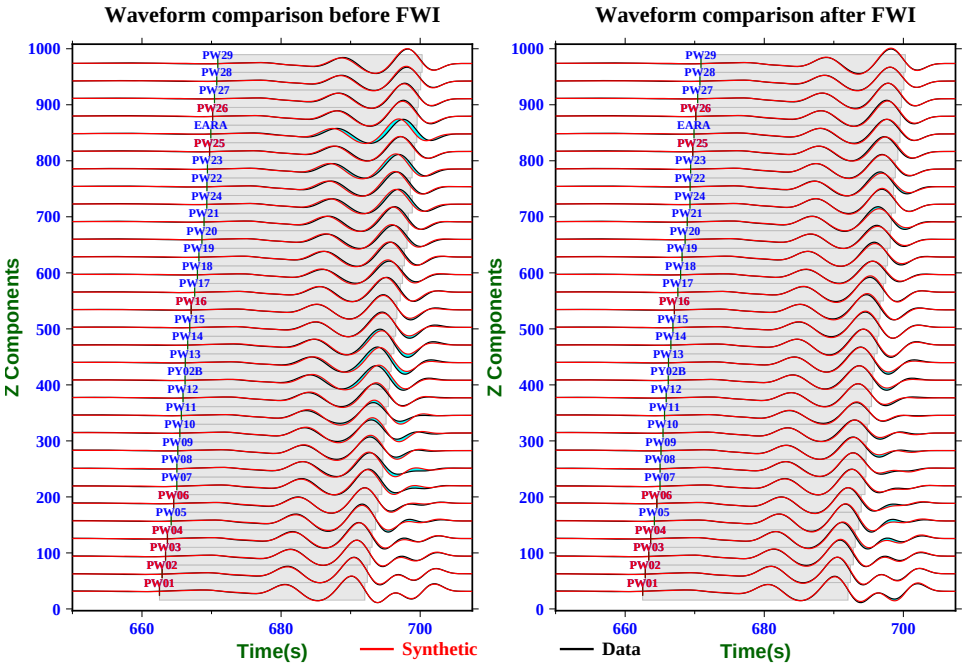
Since we use a short time window that only contains the direct P wave and the depth phases for inversion, the absence of P-to-S conversion on the Moho and later arriving multiples makes the inverted waveform only sensitive to the long wavelength structures of the velocity model. This is why FWI is almost equivalent to PAT in this scenario. The advantages of FWI over PAT will become more significant when the frequency band of the inverted wavefield broadens and more information in seismograms are exploited.

4.3.2.2 Comparison of waveform inversions using different sets of components

In order to include more arrivals in the inversion, we extend the length of the time window to the one used in the previous inversion (50 s) but perform a FWI only on the vertical components. After 8 iterations, the FWI algorithm converged with a waveform misfit reduction of 65%. The inverted models are shown in the top panels of Figure 4.59. Compared to the models obtained with short time windows shown in Figure 4.56, the shallow structures in both Vp and Vs model are now better retrieved. This demonstrates the importance of extending the time window to include the P wave coda in the inversion. These vertical coda waveforms contain key information to constrain both Vp and Vs models. For primary phase such as the direct P wave, the Fresnel zone of the sensitivity kernel is a simple cigar-shaped region surrounding its geometrical ray. The sensitivity kernel for the waveform misfit within a chosen time window depends on many factors: the frequency band of the waveforms, the length of the time window, and the current model. When more converted phases and multiples are included in the FWI, the corresponding sensitivity kernel becomes more complex and contains remarkable sensitivities far from the geometrical ray of the primary phase (Tong et al., 2014a). In general, individual

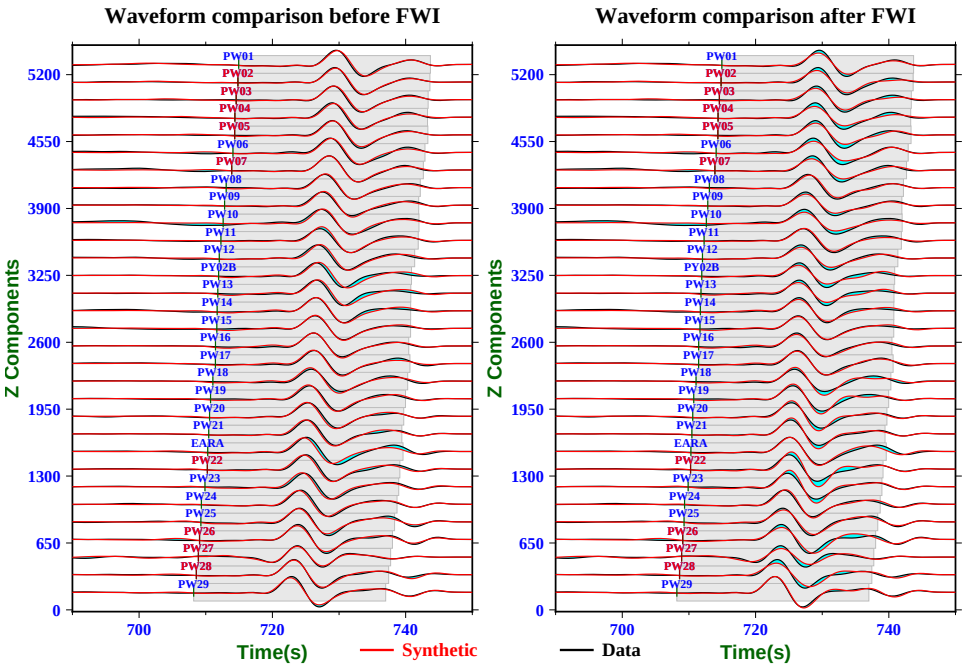


(a) vertical component waveforms along the transect for PAT adjoint inversion.

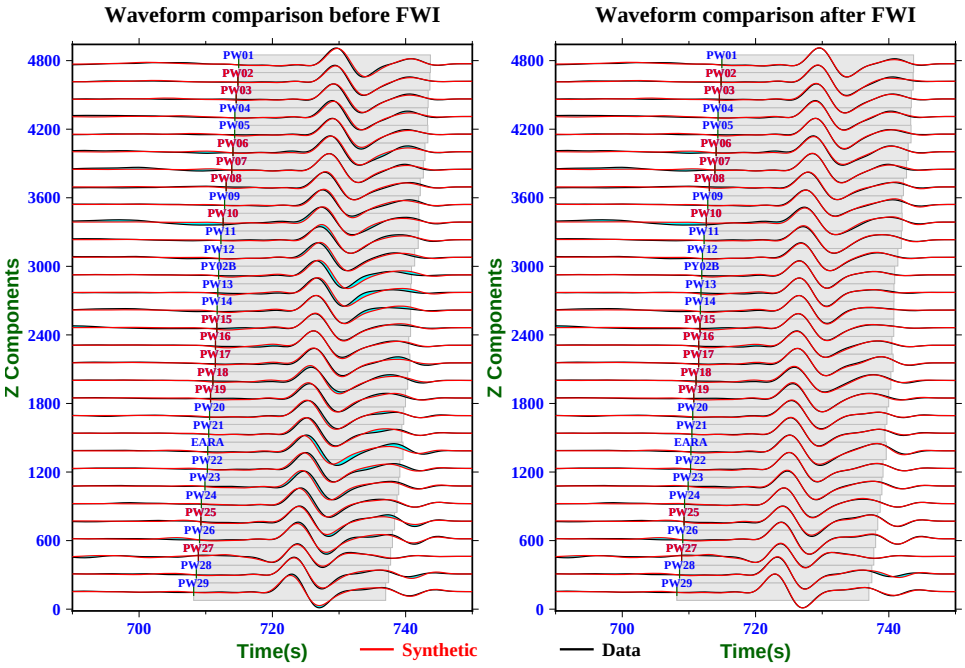


(b) vertical component waveforms along the transect for FWI with short time window.

Figure 4.57: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 May 24 event. The shaded region represent the limits of the time windows that are used in the inversion. (a): phase adjoint inversion for the direct P waves; (b): full waveform inversion for only vertical components with 30 s long time window. Left panel: Synthetic seismograms computed in the smooth initial 1D model. Right panel: Synthetic seismograms computed in the final 3D model.



(a) vertical component waveforms along the transect for PAT adjoint inversion.



(b) vertical component waveforms along the transect for FWI with short time window.

Figure 4.58: Observed (black lines) and synthetic (red lines) vertical component seismograms for the 2013 Aug 13 event. (a): phase adjoint inversion for the direct P waves; (b): full waveform inversion for only vertical components with 25 s long time window.

sensitivity kernels may not constrain the velocity anomalies very well, in contrast to the total event kernel obtained by summing up all the individual kernels for all the traces and events. Since the sensitivity kernels are computed for the vertical components of the direct P waves and their coda, we mainly use the free surface (back-scattered) multiples (such as the $PpPmp$, $PsPmp$, and so on) to invert the sharp velocity gradients of the Vp model. As a result of the lack of converted phases in the inversions, we only retrieve some shallow structures in the Vs model by inverting some reverberations on the Moho (for example the $PpPms$, $PsSms$, etc). The waveform misfit of the radial components after this FWI is nearly unchanged, which suggests that the information in the radial component is still needed. The recovery of deep structures of Vp and Vs model requires the inclusion of radial component waveforms in FWI.

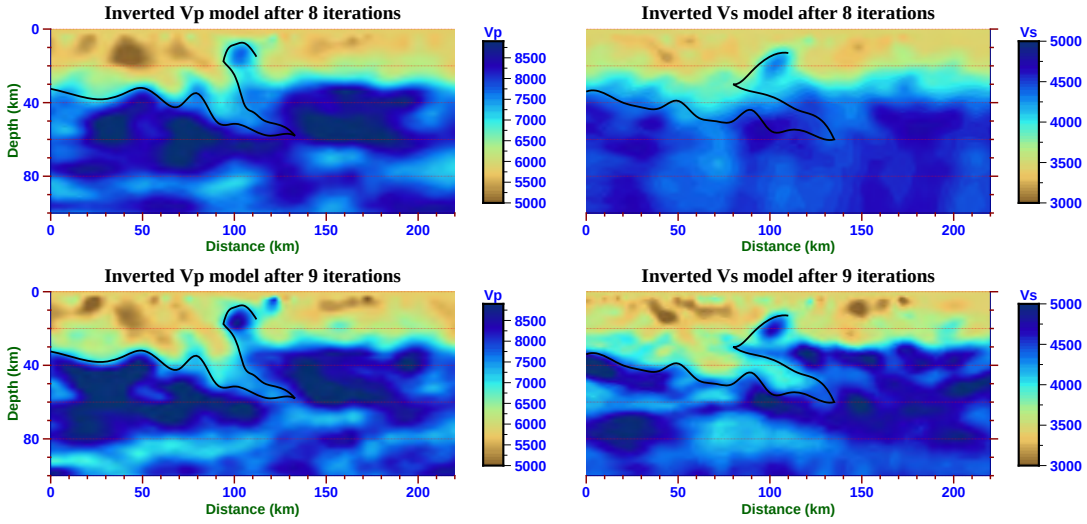


Figure 4.59: Final Vp and Vs models obtained by full waveform inversion using only the vertical components with a 50 s long time window (top panels) and full waveform inversion using both the vertical and radial components starting from 3D models (bottom panels), respectively.

We perform a new FWI by including the radial component waveforms with a 50 s long time window, starting from the final 3D models obtained from FWI on the vertical components. The new inversion converged after 9 iterations. The waveform misfit reduction for the vertical and radial components are now 15 and 67%, respectively. The final models are also shown in the bottom panels of Figure 4.59. Not surprisingly, the total waveform misfit reduction mainly comes from the radial components. Compared to the final models obtained from FWI using only the vertical components, the new inverted models mainly improve the resolution of Vs models in the deeper part. In the Vs model, a 'crocodile' pattern is now apparent, which was not present in the final model obtained by inverting only the vertical components. We draw the black contour line of the crust-mantle boundaries for the final Vp and Vs models in both panels of Figure 4.59 to compare the model update. The new models reveal the subduction of the Iberian crust beneath the European plate. Table 4.60 shows the vertical and radial component

waveform misfits for each event before and after these two inversions.

waveform misfit	initial misfit: (vertical)	relative misfit residual: (vertical)	initial misfit: (radial)	relative misfit residual: (radial)
2013 May 11	88.0939	0.390464	15.5614	0.819291
2013 May 24	64.7238	0.240729	37.931	1.04699
2013 Aug 13	48.6452	0.266656	17.3595	1.04769
2013 Aug 30	25.1188	0.435147	14.7589	0.922454
2013 Sep 25	49.9159	0.544409	19.4884	1.03197

(a) Waveform misfits and relative misfit residuals: FWI P only

waveform misfit	initial misfit: (vertical)	relative misfit residual: (vertical)	initial misfit: (radial)	relative misfit residual: (radial)
2013 May 11	32.6695	0.832551	12.895	0.537022
2013 May 24	13.8666	0.811911	39.8943	0.233834
2013 Aug 13	12.2507	0.94794	18.9574	0.246101
2013 Aug 30	11.2164	0.821739	13.7152	0.474548
2013 Sep 25	25.6323	0.948883	20.725	0.366308

(b) Waveform misfits and relative misfit residuals: FWI P and SV

Table 4.60: Waveform misfits and relative misfit residuals for each event. (a) FWI P only: FWI using only the vertical components. (b) FWI P and SV: FWI using both the vertical and radial components, starting from the 3D final model of FWI only on the vertical components.

Including the radial components in FWI provides little improvement to the Vp model. This is due to the relatively long period (5 s) and the small sensitivity of radial component waveform with respect to Vp model perturbation. In contrast, the significant improvement of the Vs model is mainly concentrated in the lower jaw of the 'crocodile' pattern. Figure 4.61 shows the total Vs event kernels (left panels) for FWI using both the vertical and radial components and the updated Vs models (right panels) after different iterations. The black contour line that delimits the crust-mantle boundary in the new final Vs model is also shown in each panel of Figure 4.61. We can see the strong negative Vs sensitivity inside the lower jaw of the 'crocodile' pattern in the first 2 iterations, which means that the model parameters in this region of the current model should be reduced. Another significant feature is the change of polarity of the Vs kernel through the top and bottom surfaces of the lower jaw of the 'crocodile' pattern. The recovery of the Vs model in this region may come from the P-to-S conversions in the radial components.

To summarize, it is necessary to include as much information as possible in the inversion. The resolution and quality of tomographic models are controlled by many factors, such as the choice of the length of the time windows and the number of components used in the inversion. FWI leads to better results than PAT, especially in the case of sparse data coverage.

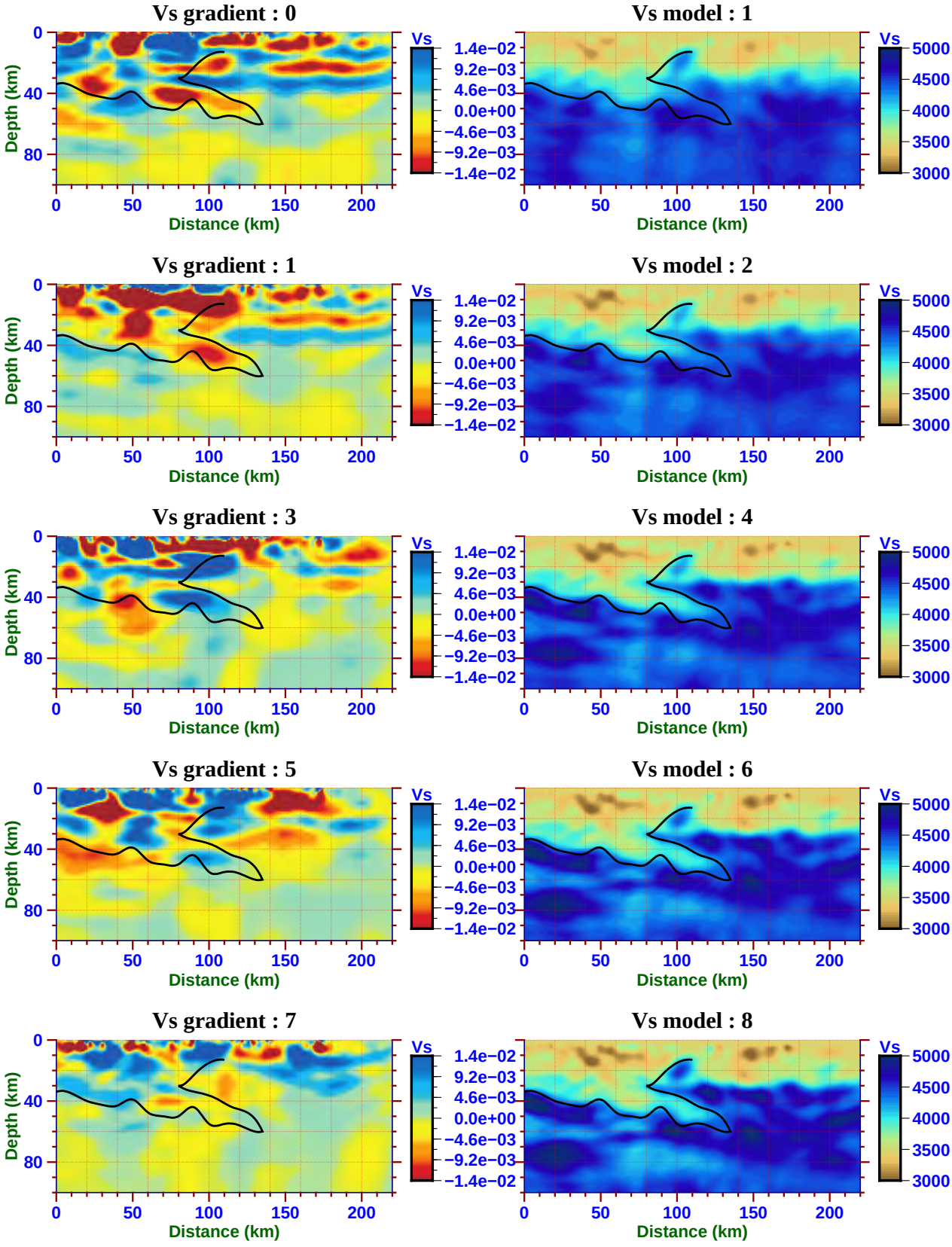


Figure 4.61: Total Vs event kernels (left) for FWI using both the vertical and radial components and updated Vs models (right) after different iterations.

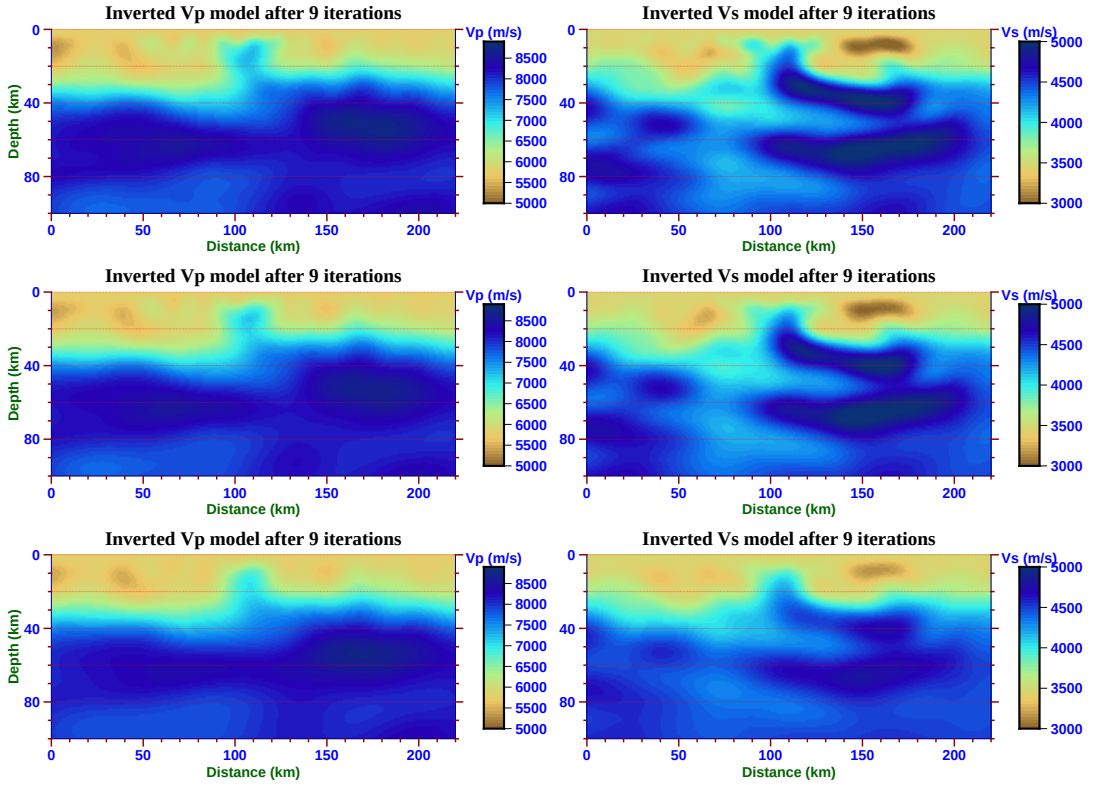
4.3.2.3 Comparison between full waveform inversion with and without including the topography of the free surface

Finally, we make a comparison between full waveform inversion with and without including the topography of the free surface, using the dataset selected after applying a uniform corner frequency filter. Models obtained by FWI without including the topography of the free surface with different smoothing coefficients are shown in Figure 4.62a. As a comparison, we also show the best models obtained by FWI using the same dataset with the topography in Figure 4.62b. The main effect of considering the free surface topography is on the resolution of shallow structures. Comparing the Vs models shown in the middle right panel of Figure 4.62a with the models with the topography in the right panel of Figure 4.62b, we observe different low velocity anomalies in the southern part of two Vs models, between 10 and 20 km depth.

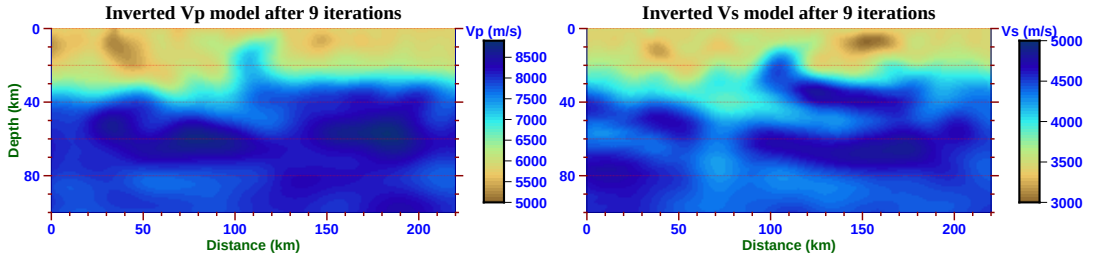
The complex topography of the free surface has a significant effect on the seismic wavefield not only by perturbing the travel time and amplitude of the direct P waves but also by producing strong body-to-surface wave conversions (Monteiller et al., 2013). Since we filter the data at 5 s for the current inversion (which is a relatively long period), the topography still has a moderate influence on the waveform inversion results. In the following, we make a brief analysis of the influence of free surface topography on teleseismic waveforms and on waveform inversion.

We impose a simple 1D surface topography for our regional domain as shown in Figure 4.63a. The 65 virtual receivers (blue circles) are positioned along an N-S array with an interstation spacing of 4 km which is similar to that in the deployment of PYROPE temporary transect. Figure 4.63b shows the elevation of receivers (red triangles) on the free surface topography (blue line). We compute synthetic seismograms excited by a distant earthquake at 10 km depth, located 30° north of the center of the regional domain. We compute the synthetic seismograms in the regional smooth 1D model with and without the free surface topography shown in Figure 4.64. The broad-band vertical (Figure 4.64a) and radial component (Figure 4.64b) waveform differences between synthetic seismograms of the model with (red lines) and without (black lines) the free surface topography are filled with cyan. The synthetic seismograms are aligned on the epicentral distance of each receiver.

The free surface topography results in a general loss of waveform coherency and has a substantial influence on both the vertical and radial components. We consider a 15 s long time window (blue lines) starting 5 s before the theoretical P arrival shown in Figure 4.64. For broad-band waveforms, the most evident waveform differences are in this time window containing direct P and the depth phases. We filtered the seismograms with a 5 s (Figure 4.65) and 2.5 s (Figure 4.66) low-pass filter. For seismograms filtered at 5 s, the amplitudes of waveform differences for the radial components are larger than for the vertical components. For seismograms filtered at shorter period 2.5 s, we observe the significant body-to-surface mode conversions propagating both forward and backward. The high relief produces a strong radiation

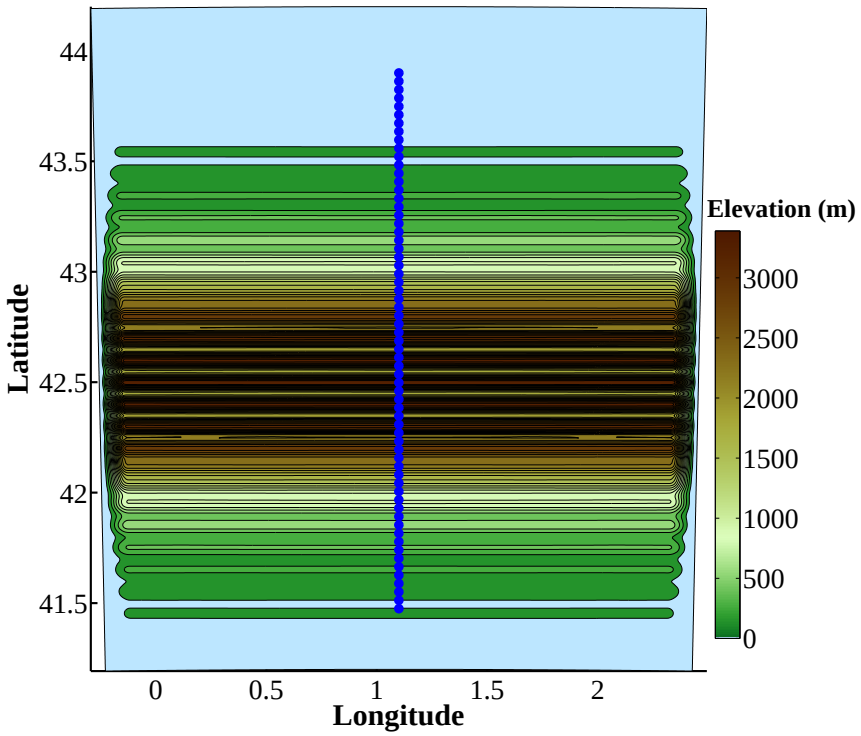


(a) FWI of 5 s without including the topography.

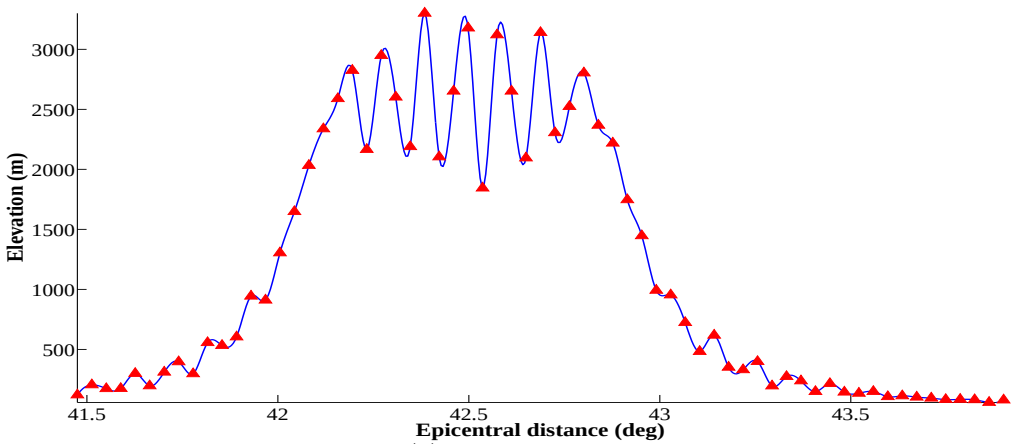


(b) FWI of 5 s with topography of the free surface

Figure 4.62: The FWI results filtered at 5 s: (a) FWI without including the topography of the free surface, with the smoothing coefficient 0.1(top), 0.25(middle) and 0.4(bottom panels). (b) FWI with topography and using smoothing coefficient 0.5.

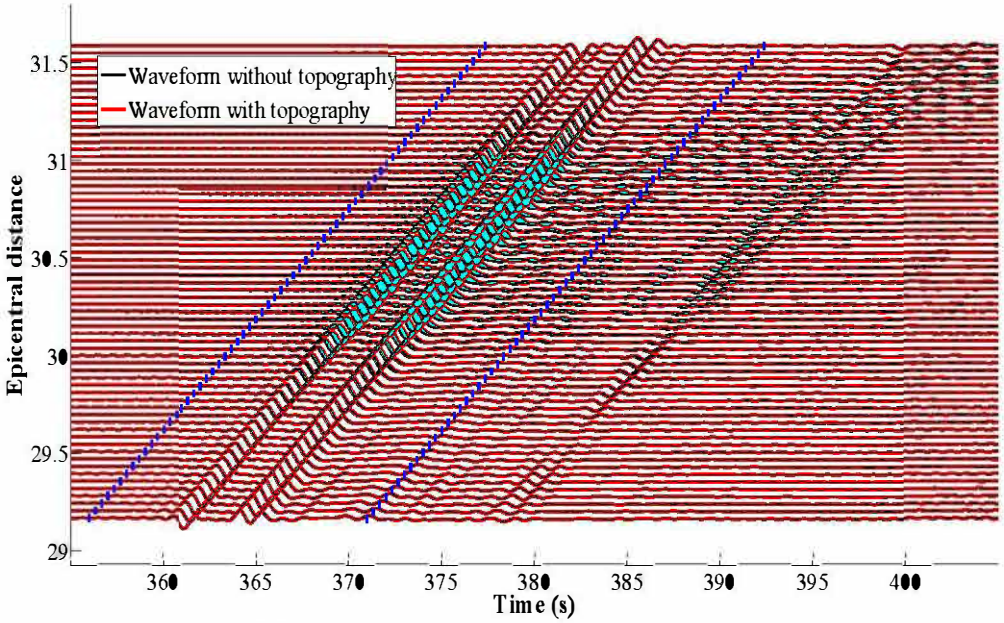


(a) Map of the simple 1D free surface topography.

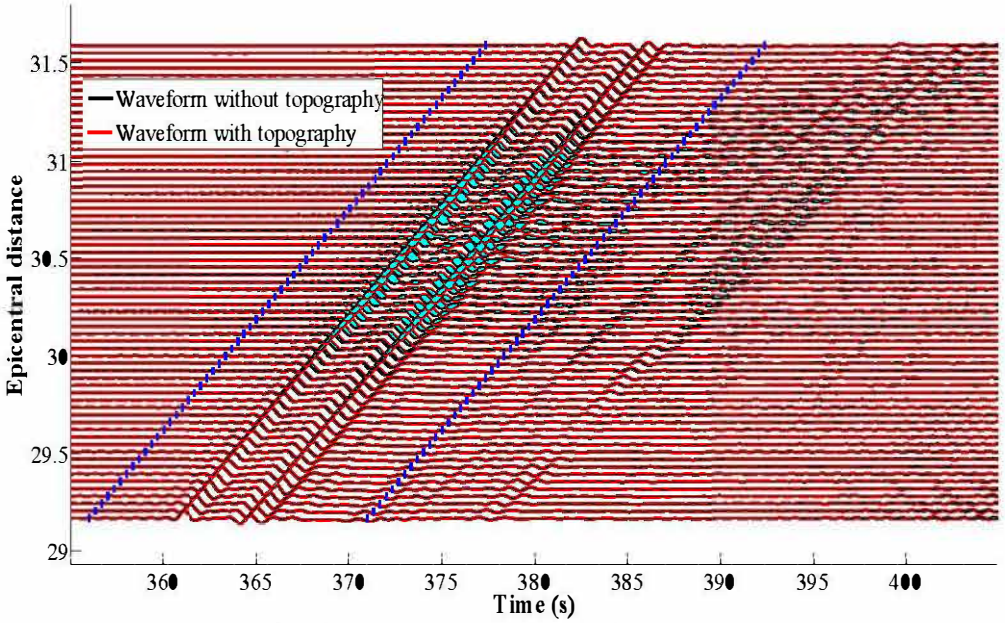


(b) The elevations of receiver array.

Figure 4.63: (a) Map of the simple 1D free surface topography and 65 virtual receivers. (b) The elevations of receiver array (red triangles). The red line shows the curve of static correctionss usually used in body wave tomography.



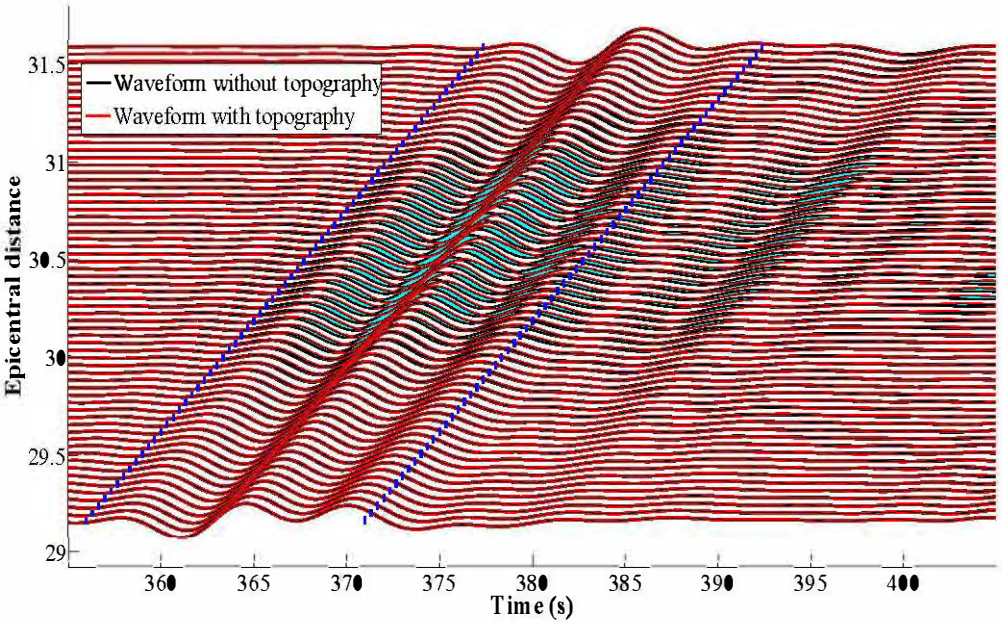
(a) Vertical component waveforms along the virtual transect.



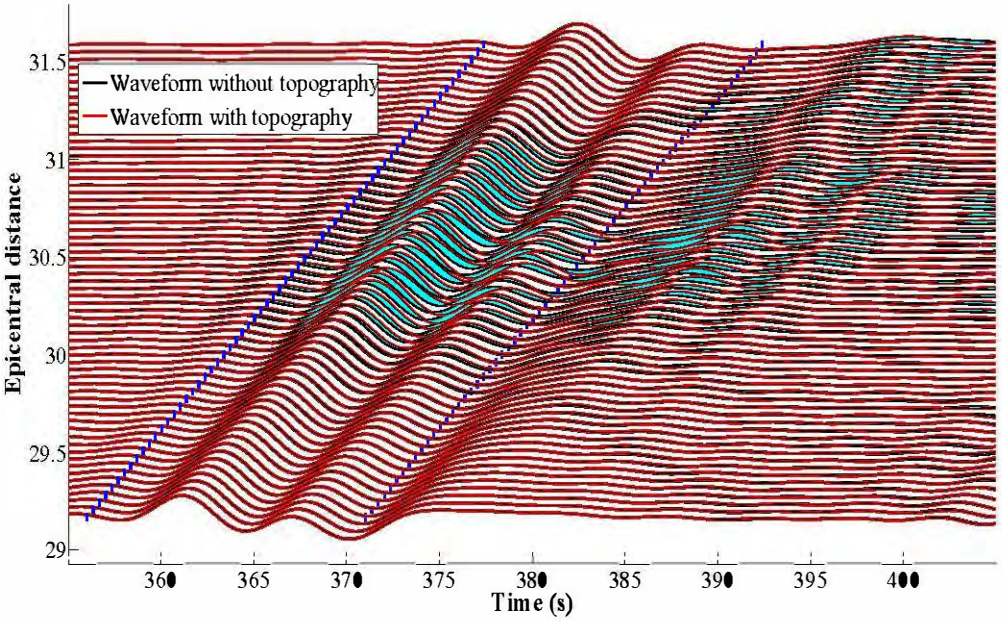
(b) Radial component waveforms along the virtual transect.

Figure 4.64: Broad-band vertical (a) and radial component (b) synthetic seismograms computed in the model with (red lines) and without (black lines) the free surface topography. The blue lines show the 15 s long time windows used in the kernel computations for waveform misift.

of surface waves which is stronger in the forward propagation direction. These wave trains have a small apparent velocity (around 3 km/s) on both the vertical and radial components, which



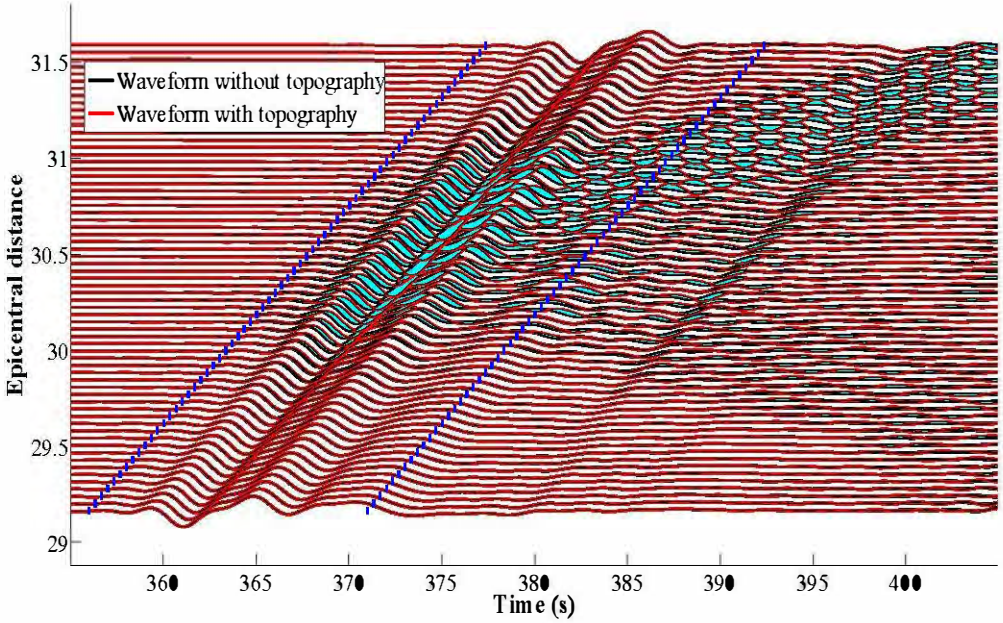
(a) Vertical component waveforms along the virtual transect.



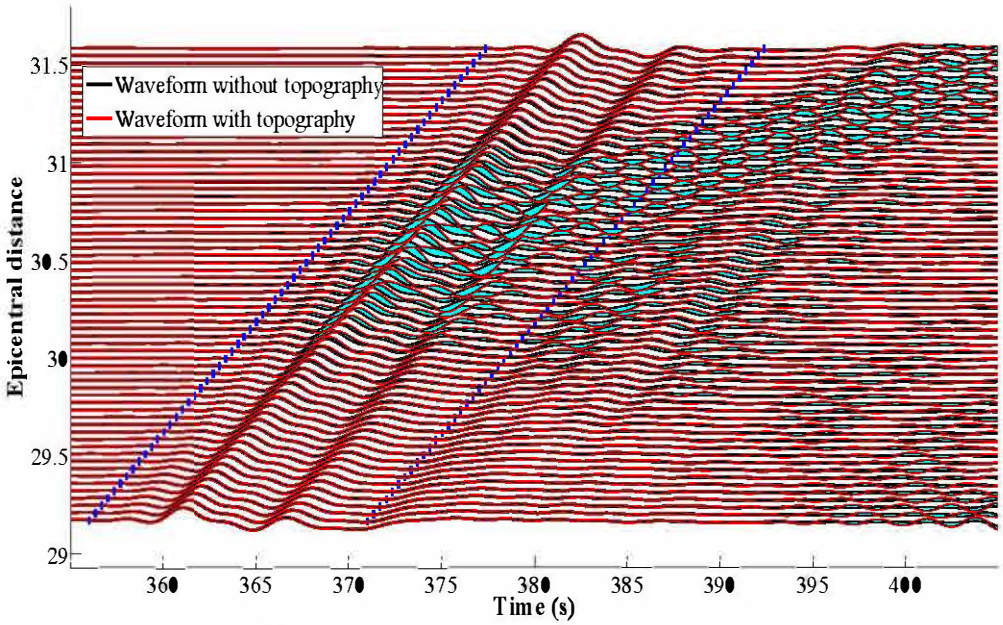
(b) Radial component waveforms along the virtual transect.

Figure 4.65: The vertical (a) and radial component (b) synthetic seismograms computed in the model with (red lines) and without (black lines) the free surface topography, filtered at the corner frequencies 0.02 and 0.2 Hz.

The perturbations of waveforms by the free surface topography mainly concern the direct *P* wave arrival time. These perturbations can be eliminated by applying static corrections (Yilmaz, 1990) with our regional smooth 1D model. Figures 4.67 and 4.68 show the corrected



(a) Vertical component waveforms along the virtual transect.

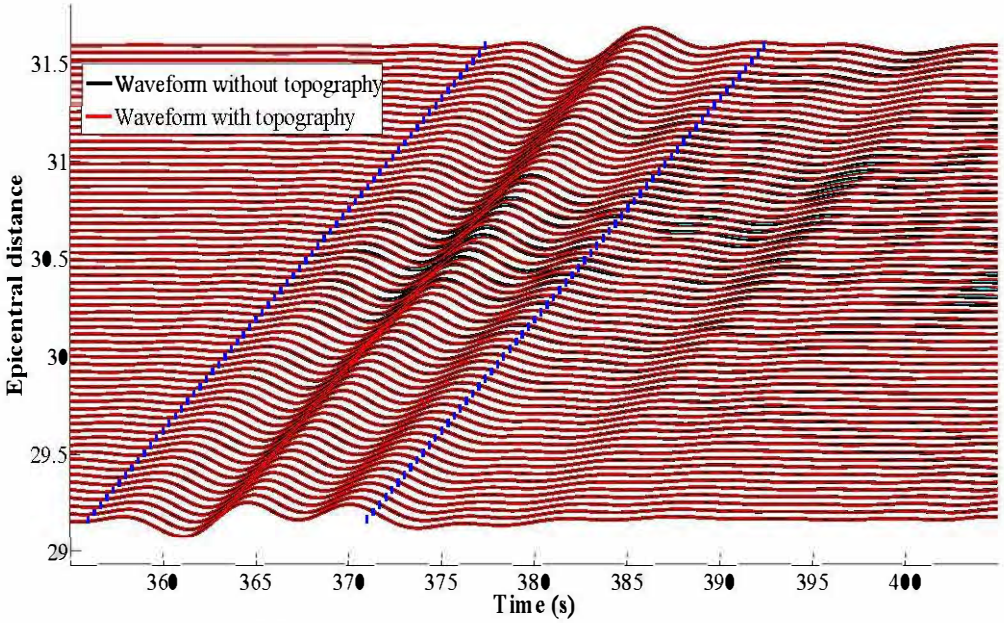


(b) Radial component waveforms along the virtual transect.

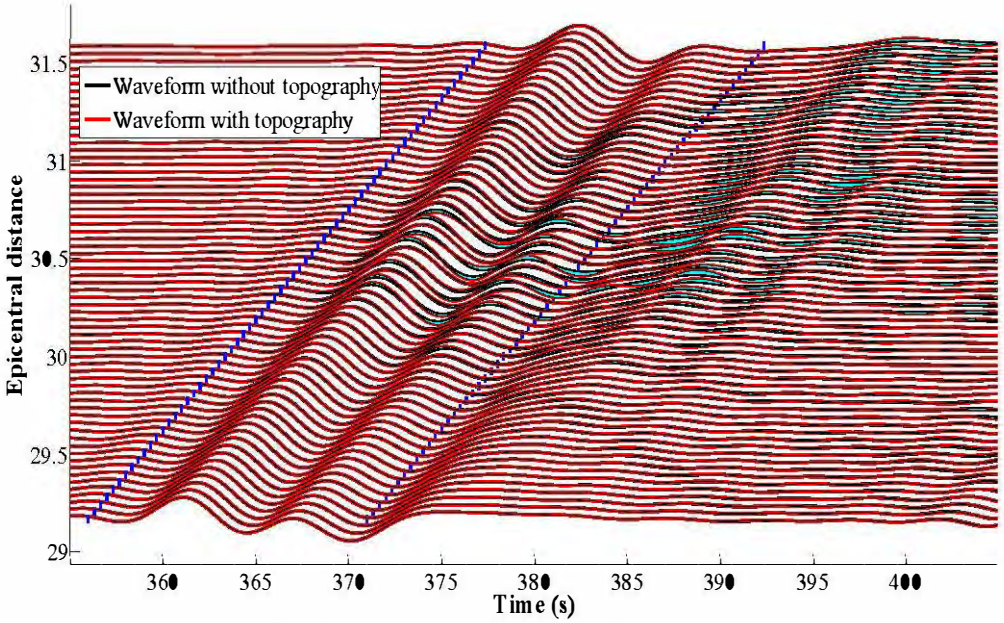
Figure 4.66: The vertical (a) and radial component (b) synthetic seismograms computed in the model with (red lines) and without (black lines) the free surface topography, filtered at the corner frequencies 0.02 and 0.4 Hz.

synthetic seismograms computed in the model with topography and the residual waveform differences that are attributed to the amplitude perturbations of the direct P waves and the mode conversions. Compared to the waveform differences without static corrections filtered at 5 s, the residual waveform differences for the vertical components are rather small. On the other hand, the residual waveform differences for the radial components are much larger. However, they are not simply correlated to the topography. These large residual waveform differences may explain the differences in FWI results with and without considering the free surface topography, and why these differences are mainly observed in the Vs model. Compared to seismograms filtered at 5 s, we find that the residual waveform differences filtered at 2.5 s are still large after static corrections for both components. The strong body-to-surface wave conversions contaminate the coda waves of the direct P waves (Monteiller et al., 2013) at the shorter period. This implies that the effect of the free surface topography on waveforms will be larger when the FWI gradually moves towards higher frequencies.

We next compute the sensitivity kernel of both vertical and radial components waveform misfits between filtered seismograms of the model with and without the static corrections. The 15 s long time window shown in Figure 4.64 is used to define the waveform misfit function. In order to eliminate the large sensitivities in the vicinity of the receivers, we precondition the sensitivity kernel by the square root of depth at each SEM grid point. The preconditioned kernels of the waveform misfits filtered at 5 s with and without static corrections are shown in Figure 4.69. The negative Vp and Vs sensitivity kernels for waveform misfit without the static corrections (top panels) are concentrated close to the free surface, down to a depth of 20 km. The horizontal range of these negative Vp and Vs sensitivity kernels corresponds to the distribution of the receivers located in high relief which have larger waveform perturbations due to topography. The horizontal range of this negative sensitivity to Vs perturbation is larger than to Vp perturbation. This is due to the perturbation range for the radial component waveforms which is larger in the chosen time window. Bottom panels of Figure 4.69 show the preconditioned kernels of the waveform misfits with the static corrections. The discontinuously distributed negative Vp and Vs sensitivity kernels have disappeared, replaced by the alternative variation of their polarities along the free surface. The amplitude and range of the Vs sensitivity are larger than Vp sensitivity, corresponding to the significantly larger residual waveform differences for the radial components. The patterns of the sensitivity kernels may be related to the forward propagation path of the scattered Rayleigh waves. Although this topography model is simpler than the real Pyrenees topography, the strong body-to-surface wave conversions are similar to the one observed in Monteiller et al. (2013). Our FWI without considering topography (without static corrections) shifts the observed waveforms with an average value of the measured travel time anomalies on all the traces. This operation has a similar effect as the static corrections, which eliminates the overall shallow negative sensitivity in waveform kernels. Thus we suspect that the inferior resolution of shallow structures in the Vs model obtained by

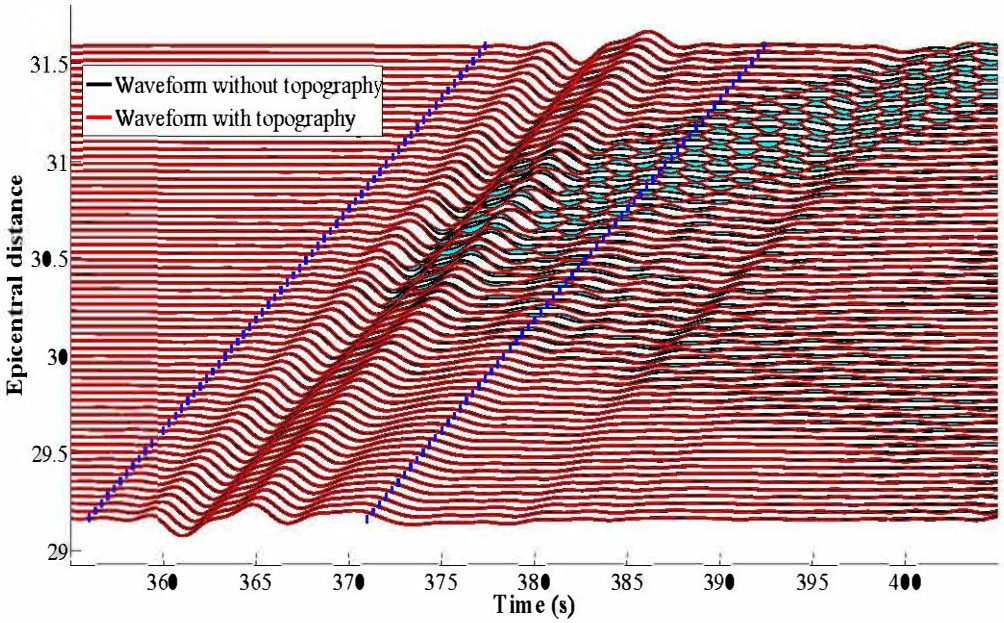


(a) Vertical component waveforms along the virtual transect.

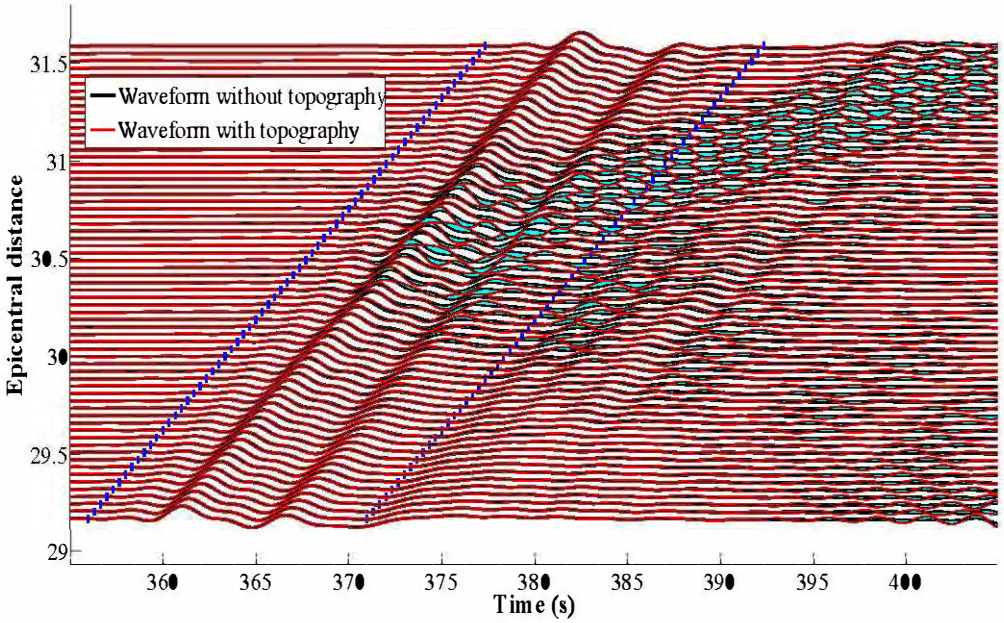


(b) Radial component waveforms along the virtual transect.

Figure 4.67: The vertical (a) and radial component (b) synthetic seismograms computed in the model with (red lines) and without (black lines) the free surface topography after static corrections, filtered at the corner frequencies 0.02 and 0.2 Hz.



(a) Vertical component waveforms along the virtual transect.



(b) Radial component waveforms along the virtual transect.

Figure 4.68: The vertical (a) and radial component (b) synthetic seismograms computed in the model with (red lines) and without (black lines) the free surface topography after static corrections, filtered at the corner frequencies 0.02 and 0.4 Hz.

FWI without including the topography mainly comes from the sensitivities generated by the imprint of topography on the waveforms.

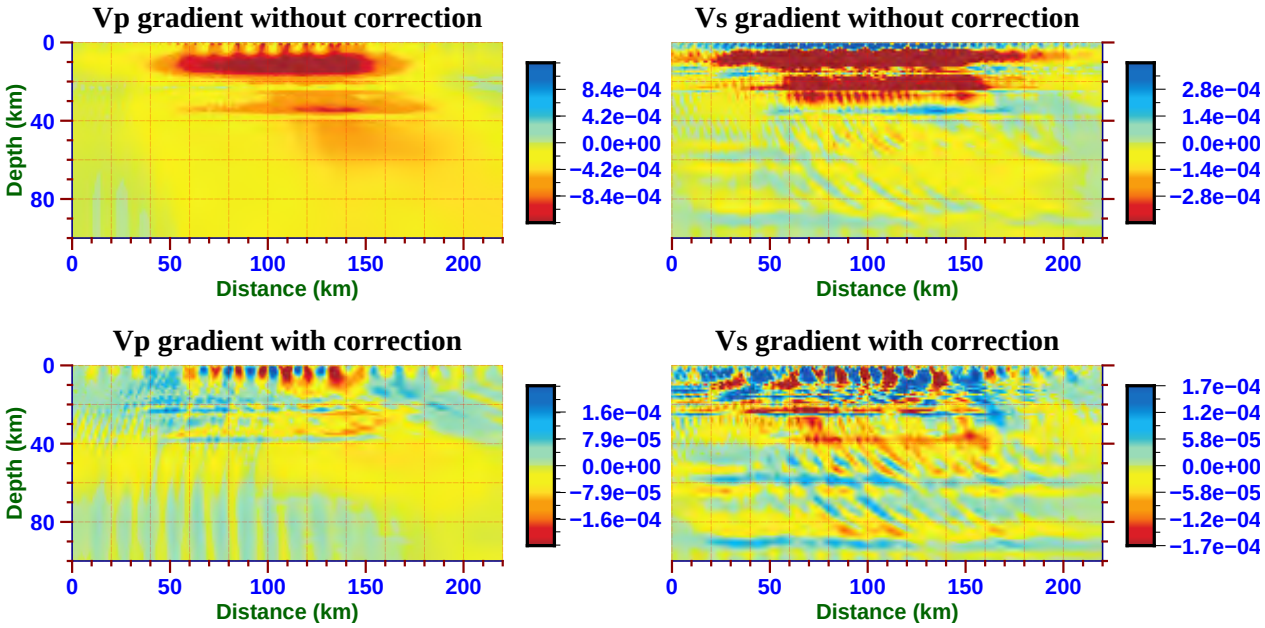


Figure 4.69: The preconditioned Vp (left) and Vs (right) sensitivity kernels of the waveform misfit computed in a 15 s time window with (bottom) and without (top) static corrections. We filtered the waveforms with a 5 s low-pass Butterworth filter.

The preconditioned kernels of the waveform misfits filtered at 2.5 s with and without static corrections are shown in Figure 4.70. The sensitivity is not only concentrated at shallow depth but also distributed down to 50 km depth with complex patterns. Note that we only use a short time window containing the direct P waves to compute the sensitivity kernel of waveform misfit. Our FWI for real data usually uses a longer time window that contains more later arrivals after the direct P waves and of course includes the effects of topographic conversions. These topographic conversions may reduce the S/N ratio of the useful later arrivals such as the P-to-S conversions and the crustal multiples which are key to retrieve the main seismic interfaces. Ignoring the free surface topography at shorter period will thus lead to a loss of spatial resolution and more artifacts in the tomographic model. Therefore, it is necessary to separate the influences of topographic effects from the focusing/defocusing effects of seismic wavefield propagating through a heterogeneous medium. The systematic study of the impact of the free surface topography on waveforms filtered at different frequency range and FWI results, which requires a mass of simulations and well-designed resolution test, is under way.

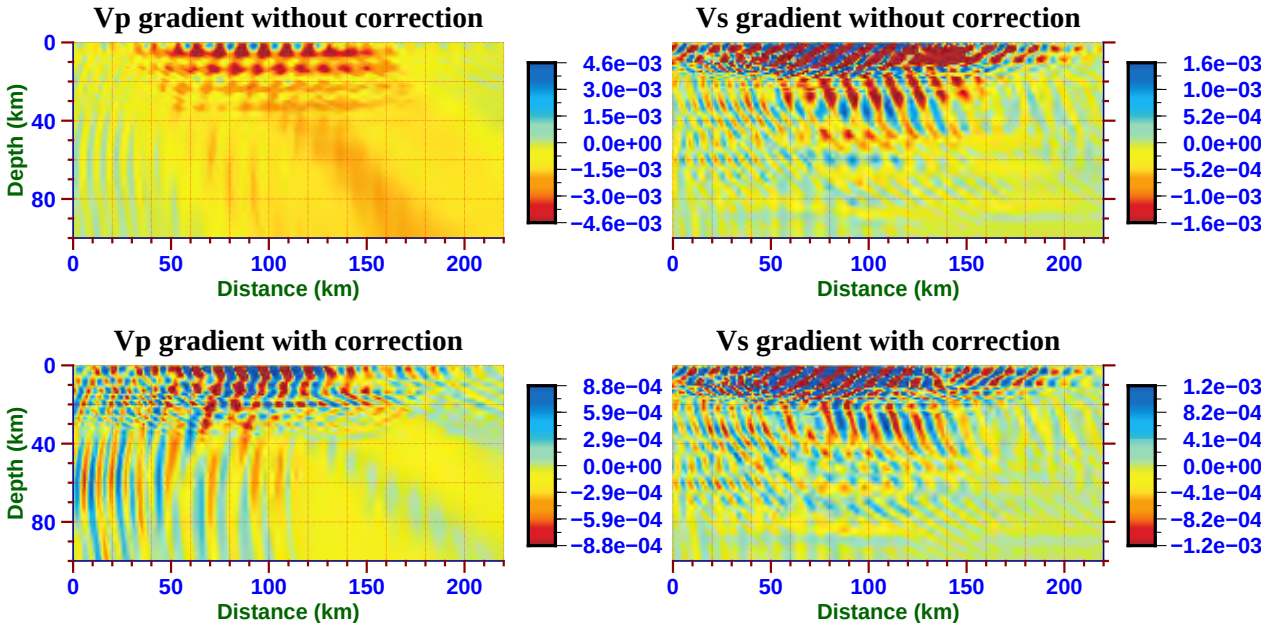


Figure 4.70: The preconditioned Vp (left) and Vs (right) sensitivity kernels of the waveform misfit computed in a 15 s time window with (bottom) and without (top) static corrections. We filtered the waveforms with a 2.5 s low-pass Butterworth filter.

Chapter 5

The deep roots of the western Pyrenees revealed by full waveform inversion of teleseismic P waves

Contents

This chapter was published as a paper entitled 'The deep roots of the western Pyrenees revealed by full waveform inversion of teleseismic P waves' by Yi Wang, Sébastien Chevrot, Vadim Monteiller, Dimitri Komatitsch, Frédéric Mouthereau, Gianreto Manatschal, Matthieu Sylvander, Jordi Diaz, Mario Ruiz, Franck Grimaud, Sébastien Benahmed, Hélène Pauchet, and Roland Martin in GEOLOGY, 44(6): 475-478, 2016.

5.1 Abstract

Imaging the architecture of mountain roots is required to understand the support of topography and for kinematic reconstructions at convergent plate boundaries, but is still challenging with

conventional seismic imaging approaches. Here we present a three-dimensional model of both compressional and shear velocities in the lithosphere beneath the western Pyrenees (southwest Europe), obtained by full waveform inversion of teleseismic P waves. This tomographic model reveals the subduction of the Iberian crust beneath the European plate, and the European serpentinized subcontinental mantle emplaced at shallow crustal levels beneath the Mauléon basin. The rift-inherited mantle wedge acted as an indenter during the Pyrenean convergence. These new results provide compelling evidence for the role of rift-inherited structures during mountain building in Alpine-type orogens.

5.2 Introduction

Collisional orogens require deep roots to support their topography over millions of years (Watts, 2001). However, the deep internal structure and nature of these orogenic roots have long been the focus of debates and contrasting interpretations. According to local (Airy) isostasy, high topographic reliefs of mountain ranges are compensated by deep crustal roots. However, the crust in orogens does not respond to surface loads locally but rather by flexure over a broad region (Karner & Watts, 1983). Departure from the Airy model is often observed, with a shift of maximum crustal thickness with respect to topographic highs, the gravity field displaying a positive-negative anomaly couple. The positive anomalies have been classically ascribed to buried loads, whereas the broad gravity lows reflect the downward flexure of the underthrust crust produced by the combined effects of surface (i.e., topography) and internal loads. This simple conceptual model has been successfully applied to reproduce the pattern of Bouguer anomalies in various orogenic belts (e.g. Karner & Watts, 1983; Royden & Karner, 1984). However, owing to the insufficient spatial resolution of classical seismic tomography, the nature and origin of these buried loads have so far remained elusive.

The Pyrenees (southwest Europe) are an intracontinental orogen that result from the tectonic inversion of an Early Cretaceous rift system formed between the Iberia and European plates (e.g. Choukroune, 1989). Tectonic restorations and kinematic models of plate convergence indicate a moderate shortening of $<$200 km since the Late Cretaceous (Roure et al., 1989; Muñoz, 1992; Teixell, 1998; Mouthereau et al., 2014) that ended $\backsim$20 m.y. ago. The Pyrenees can thus be considered as a fossilized plate boundary. This has been confirmed by recent GPS studies that did not find any measurable relative motion between Iberia and Europe (Nocquet & Calais, 2004). Precollisional rift-related structures are particularly well preserved in the western Pyrenees, where extension was greatest and collision reached a less advanced stage (Masini et al., 2014). This region thus offers a unique opportunity to study a crustal section across an embryonic stage of a collisional orogen. Here we present a lithospheric section of the western Pyrenees constructed from full waveform inversion of vertical and radial component records of teleseismic P waves that enables us to decipher the enigmatic nature and structure of buried

loads.

5.3 Data and waveform inversion method

Seismological data are from a dense transect deployed during the temporary PYROPE (Pyrenean Observational Portable Experiment) project (Chevrot et al., 2014) in 2012–2013 (Fig. 5.1). We selected the vertical and radial components of 5 earthquakes recorded by the 29 stations deployed along the transect. The full waveform inversion (FWI) method (Monteiller et al., 2015); for details on the method, see the GSA Data Repository) searches for three-dimensional compressional and shear velocity models that minimize the misfit between observed and synthetic seismograms computed with the hybrid direct solution method–spectral-element technique (Monteiller et al., 2013). We use 50-s-long time windows from the vertical and radial component records around the P wave arrivals, low-pass filtered at 5 s. We thus include in the inversion all the crustal reverberations that arrive in the coda of the P waves. These arrivals are crucial to constrain the sharp velocity gradients associated with the main crustal interfaces such as the crust-mantle boundary (Moho). The final model obtained after nine iterations provides an excellent fit of both vertical and radial component waveforms (Figs. 4.45a–4.54a in Chapter 4). A synthetic resolution test performed on a checkerboard model (Fig. 5.4) demonstrates that our imaging technique is able to resolve lateral and vertical variations of seismic velocities in the crust with a spatial resolution of a few kilometers, even with a limited number of teleseismic sources. The FWI inversion approach reveals structural details that could not be seen with conventional regional travel time (Chevrot et al., 2014) or ambient noise (Macquet et al., 2014) tomography.

5.4 Tomographic model

The vertical cross sections through our Vp and Vs models along the transect (Figs. 5.2C and 5.2D) show striking similarities, even though Vp and Vs were allowed to vary freely and independently during the inversion. This is remarkable because the Vp model is mostly constrained by transmitted P waves on the vertical component, while the main contribution to the Vs model comes from the P to S conversions and multiple reflections on crustal interfaces recorded on the radial components. However, the structures are more sharply defined in the Vs model. Because in FWI, as in any tomographic method, the spatial resolution scales with the seismic wavelength, this simply results from the shorter wavelengths of shear waves compared to compressional waves.

The crust-mantle boundary, expressed as a sharp velocity gradient in both the Vp and Vs models, exhibits a very complex geometry, which is in remarkable agreement with the results of receiver function migration (Chevrot et al., 2015) shown in Fig. 5.2B. We observe two distinct

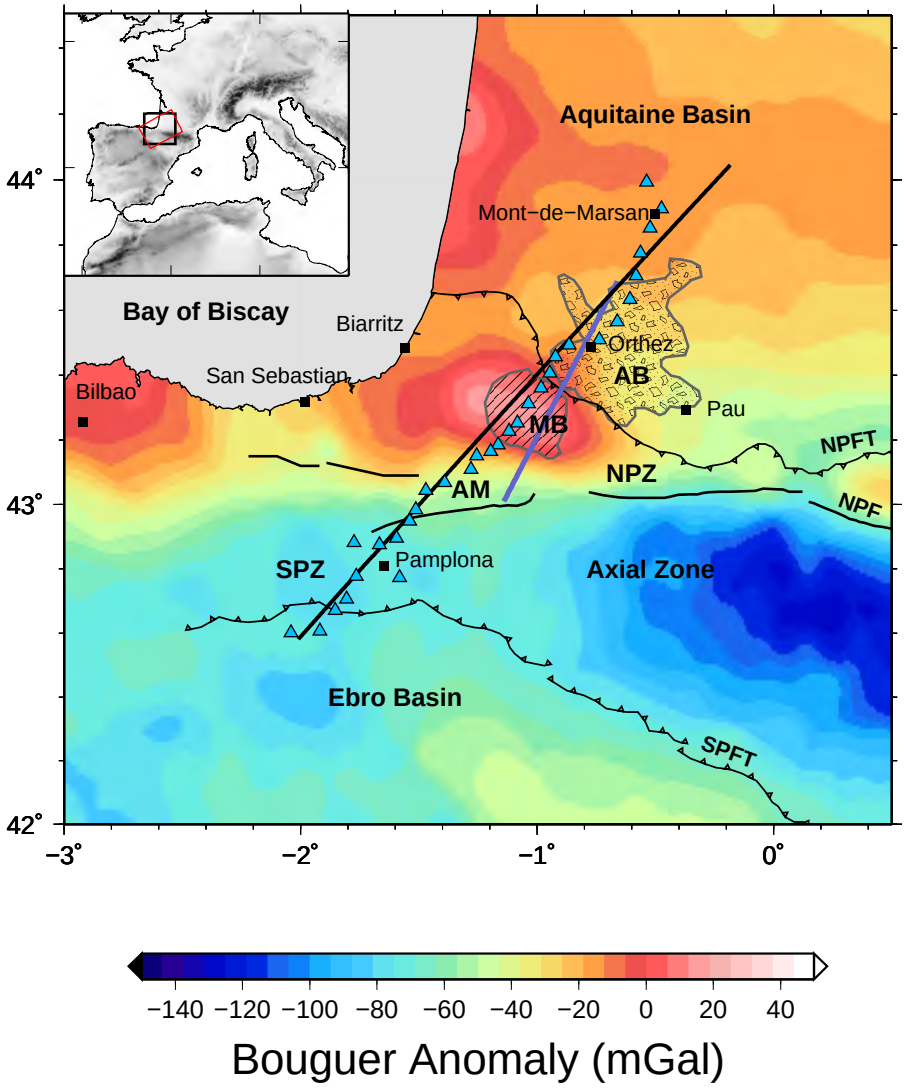


Figure 5.1: Map of Bouguer gravity anomalies with the locations of seismic stations (blue triangles) in the Pyrenees. The black and purple solid lines show the positions of the western PYROPE and ECORS-Arzacq transects, respectively. NPF–North Pyrenean fault, NPZ–North Pyrenean zone, SPZ–South Pyrenean zone, SPFT–South Pyrenean frontal thrust, NPFT–North Pyrenean frontal thrust, MB–Mauléon Basin, AB–Arzacq Basin, AM–Aldudes Massif. Inset shows location of the study region (black square) and limits of the spectral-element grid (red square).

Mohos, belonging to the Iberia and European plates, that are superposed beneath the North Pyrenean zone. The Iberian Moho dips gently from a standard depth of 30 km at the southern end of the profile to a depth of 40 km. Further north, it deepens and flattens to reach a depth of 50 km, delimiting a slice of Iberian material that underthrusts the European mantle. The Moho of the European plate is shallower and has much stronger topography. However, beneath the Arzacq basin, the different images may be contaminated by the reverberations inside shallow unconsolidated sedimentary layers.

A salient feature in the tomographic model is the north-dipping low-velocity anomaly observed beneath the European plate, located at ∽50 km depth. The top of this anomaly coincides with a negative polarity interface observed in the migration section. This strongly suggests the underthrusting of a fragment of the Iberian crust beneath the European plate, topped by the European subcontinental mantle, as proposed in a receiver function migration study (Chevrot et al., 2015). Seismic velocities in that subducted body (Vp ∽ 7.2 km/s, Vs ∽ 4 km/s) are typical of a mafic lower crust (Rudnick & Fountain, 1995), but could also be compatible with a serpentinized mantle (Christensen, 2004).

Another prominent anomaly is observed beneath the Mauléon basin, between 10 and 30 km depth, expressed in both the Vp and Vs models, that coincides with the strong positive Labourd-Mauléon Bouguer gravity anomaly (Fig. 5.1). The top of this fast velocity anomaly also corresponds to a strong Vs contrast observed in the migrated section. We have built a density model from the Vp model using a standard Birch law (see the Data Repository). The Bouguer anomalies predicted by this density model are in excellent agreement with the observations (Fig. 5.2B; Figs. 5.5 and 5.6). This suggests that the details of the deep architecture revealed with our new imaging technique are robust and can be exploited to propose a new geological model for the western Pyrenees.

5.5 Geological interpretation

Discriminating the nature of rocks from seismic velocities is a difficult problem. The seismic velocities beneath the Mauléon basin (Vp ∽ 7.3 km/s and Vs ∽ 4.2 km/s), although not incompatible with a mafic lower crust, would be close to the extreme values reported for this type of material (Rudnick & Fountain, 1995). However, the compressional velocities observed at the base of the European crust are significantly lower, ∽6.9 km/s, and in excellent agreement with those typically found in the lower crust of Cenozoic convergent margins by seismic reflection or refraction surveys (Rudnick & Fountain, 1995). We thus think that it is very unlikely that the velocity anomaly beneath the Mauléon basin reflects a thick accumulation of mafic rocks in the lower crust.

The alternative is that this fast velocity body is made of serpentinized mantle. This hypothesis is supported by many recent geological studies in the western Pyrenees that describe

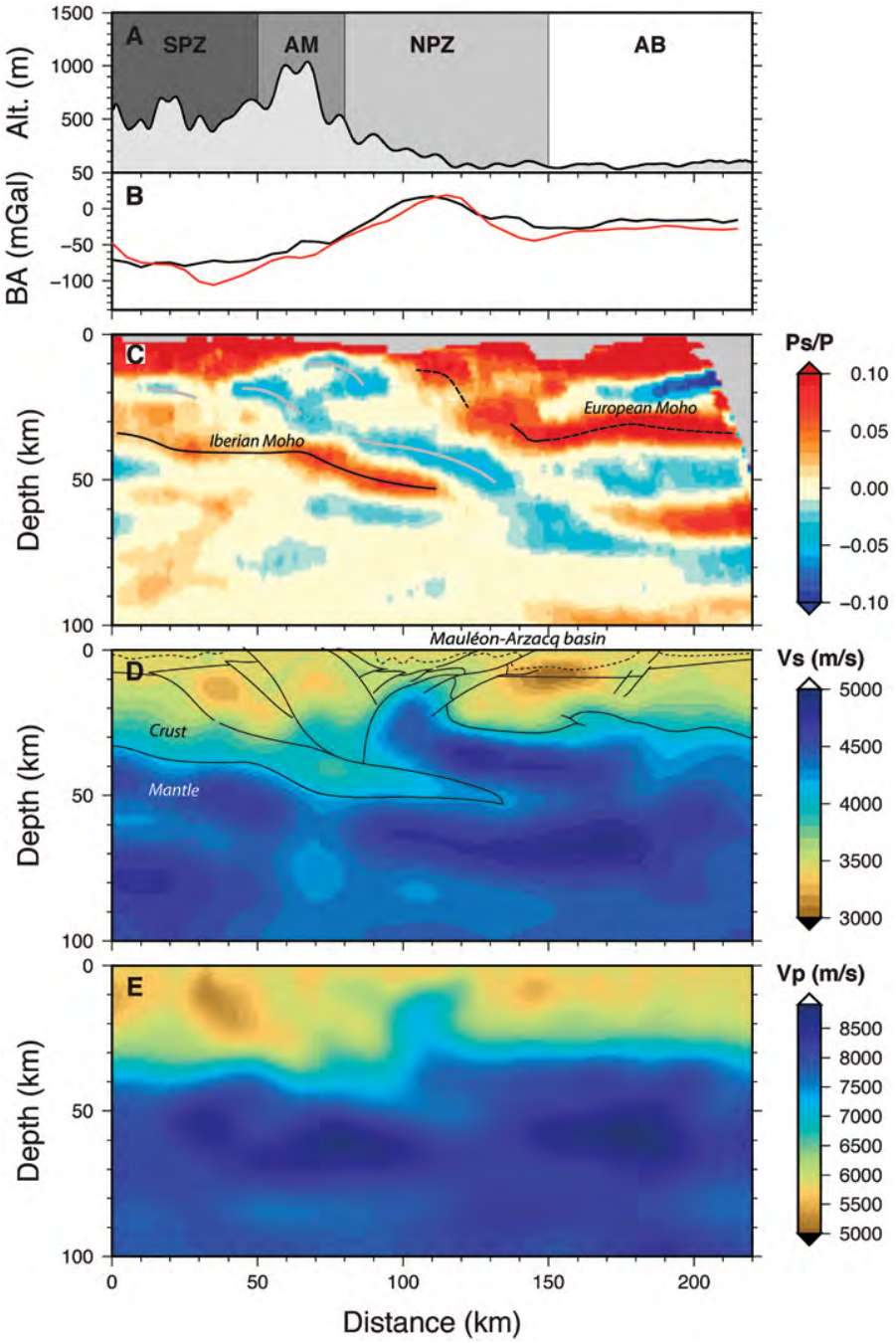


Figure 5.2: A: Topography (Alt–altitude) along the western PYROPE seismic transect. The gray areas delimitate the South Pyrenean zone (SPZ), the Aldudes Massif (AM), the North Pyrenean zone (NPZ), and the Aquitaine Basin (AB). The North Pyrenean frontal thrust (NPFT) marks the limit between NPZ and AB. B: Profiles of observed (black line) and modeled (red line) Bouguer gravity anomalies (BA). C: Common conversion point stack of receiver functions for the western transect (from Chevrot et al. (2015)). D: Vs model obtained by full waveform inversion. E: Vp model obtained by full waveform inversion.

remnants of a hyperextended rifted margin with the presence of an exhumed mantle locally exposed within small outcrops along the southern reactivated border of the Mauléon basin as well as reworked in the Albian–Cenomanian sediments filling the Mauléon basin (e.g. Jammes et al., 2009; Lagabrielle et al., 2010).

The top of the serpentinized mantle body, which corresponds to the European petrological Moho beneath the North Pyrenean zone, is very close to the surface, at ∽10 km depth. Recent studies estimate that as much as 8 km of sediments accumulated in the Mauléon basin since the Triassic (Vacherat et al., 2014), while drilling has shown that the depth of the basement is now found at ∽6 km depth, suggesting that ∽2 km of sediments were eroded during the Pyrenean convergence. This would imply that the crust beneath the Mauléon basin is <4 km thick, and may correspond to the continuation of a hyperextended crust of the European rifted margin (Tugend et al., 2014). The tomographic model also suggests that shortening in the North Pyrenean zone involved deep-seated folding and thrusting of the European subcontinental mantle of the previously thinned European lithosphere. Our geological interpretations of the tomographic model are summarized in Fig. 5.3.

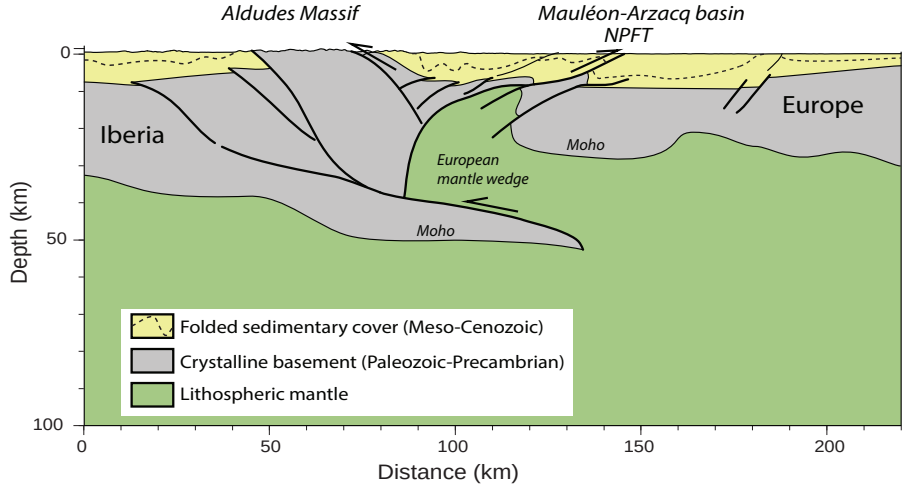


Figure 5.3: Geological interpretation of the tomographic model shown in Figure 2. NPFT–North Pyrenean frontal thrust.

5.6 Discussion and Conclusions

Our geological model differs notably from published crustal sections built from the interpretation of the Étude Continentale et Océanique par Réflexion et Réfraction Sisimique (ECORS) Arzacq profile (Daignières et al., 1994) and surface geology (Teixell, 1998; Masini et al., 2014; Daignières et al., 1994; Jammes et al., 2009). The main reason for this discrepancy stems from the difficulty to detect the Moho on the migrated section presented in Daignières et al. (1994). Beneath the Arzacq basin there is a clear deep reflector at ∽10 s, which probably corresponds to the Moho. Sporadic reflectors at ∽9 s are also detected beneath the northern part of the Mauléon basin,

but not further south. This means that the previous interpretations of the ECORS-Arzacq section were not constrained by the seismic reflection data in their central part. However, our geological model is consistent with the ECORS Arzacq section, with a rather flat European Moho at$\backsim$30 km depth beneath the Arzacq basin. It also explains why deep reflectors are not observed beneath the southern Mauléon basin and the Arbailles massif. One key observation in serpentinized mantle domains is that the Moho reflections are usually absent (e.g. Minshull, 2009); thus the lack of a well-defined Moho beneath the Mauléon basin may be additional evidence for serpentinized mantle.

Our new structural model could also explain why Lg seismic waves are strongly attenuated when they cross the western Pyrenees (Chazalon et al., 1993; Sens-Schönfelder et al., 2009), a puzzling observation that numerical modeling has so far been unable to reproduce. It is interesting that similar observations have been made in the Alps, where the attenuation of Lg waves has been related to the dense Ivrea body (Campillo et al., 1993). It is well known that crustal thickness has an important effect on the propagation of Lg waves by limiting the number of overtones in a given frequency range (Zhang & Lay, 1995). For this reason, Lg waves are almost never present in oceanic paths. The crust that we image beneath the Mauléon basin is extremely thin, perhaps even locally absent, which should strongly impede the propagation of Lg waves.

To explain the strong positive Bouguer anomaly of the Mauléon basin, former studies invoked a block of European mantle (Casas et al., 1997) or lower crust (Grandjean, 1992; Vacher & Souriau, 2001; Pedreira et al., 2007; Jammes et al., 2010). Our model would rather suggest that this anomaly is the signature of an exhumed mantle, inherited from the pre-compressional hyperextended Pyrenean rift system. The mantle wedge beneath the Mauléon basin loads and causes flexure of the underlying Iberian plate, which explains why the Pyrenees appear isostatically over-compensated, and why the deep crustal roots are shifted 50 km northward with respect to the topographic highs.

5.7 Acknowledgments

We thank Carl Tape, Andreas Fichtner, and four anonymous reviewers for comments that improved the manuscript. We gratefully acknowledge support from the Centre National de la Recherche Scientifique (CNRS), the Bureau de Recherches Géologiques et Minières (BRGM) and TOTAL through the OROGEN project. The PYROPE (Pyrenean Observational Portable Experiment) project was supported by the Agence Nationale de la Recherche (ANR) Blanc Programme (project PYROPE, ANR-09- BLAN-0229). We also acknowledge Sistema de Monitoramento de Obras (SISMOB), the French seismic mobile pool (a component of the RESIF consortium), for providing us with the seismological instrumentation for the temporary deployments. RESIF (http://portal.resif.fr) is a national research infrastructure, recognized as such by

the Ministère de l'Enseignement Supérieur et de la Recherche and managed by the RESIF Consortium, composed of 18 research institutions and universities in France. RESIF is additionally supported by a public grant overseen by the ANR as part of the Investissements d'Avenir program (reference: ANR-11-EQPX-0040) and the French Ministère de l'Écologie, du Développement durable et de l'Énergie. Calculations were performed on the French high-performance computer resources of the Très Grand Centre de Calcul du Commissariat à l'Énergie Atomique et aux Énergies Alternatives under allocations 2014-gen7165 and 2015-gen7165 made by GENCI (Grand Équipement National de Calcul Intensif).

5.8 GSA Data Repository

5.8.1 Dataset

We use teleseismic events (Table 4.1) recorded by the dense PYROPE transect deployed across the western Pyrenees between October 2012 and October 2013. The transect is composed of 29 CMG40 recording stations, evenly spaced along a 250 km-long profile with a typical inter station spacing around 8 km. We select events with a magnitude larger than 6.0 at epicentral distances comprised between 30° and 92° and keep those with the largest signal-to-noise coming from the different azimuths covered. These criteria led to the selection of five teleseismic events. We estimate the source wavelet of each event by first deconvolving the vertical-component seismograms from the Green's functions computed with the Direct Solution Method (DSM) (Geller & Takeuchi, 1995). The source wavelet is then obtained by computing the first eigenvector of the aligned deconvolved vertical traces.

5.8.2 Forward modeling

Forward and adjoint wave propagations are performed with the DSM/spectral-element method (Komatitsch & Tromp, 1999) hybrid numerical technique described in Monteiller et al. (2013). The principle of this method is to first compute the tractions and displacements produced by the distant teleseismic sources on the edges of a regional 3-D spectral-element grid. We consider impulsive sources with moment tensors taken from the GCMT catalogue (Dziewonski & anderson, 1981). We then solve regional wave propagation problems with the spectral-element method, imposing the tractions computed in the previous step convolved with the source wavelet as an input boundary condition. The regional spectral-element domain is a chunk of the spherical Earth, with a free surface that includes the topographic relief of the Pyrenees.

5.8.3 Full waveform inversion method

We perform full waveform inversion of both vertical and radial components of teleseismic P wave records, low-pass filtered at 5 s, following the method described in Monteiller et al. (2015). In this first application of our method to real data, we chose not to include the transverse components into the inversion in order to avoid the complexities that may result from the presence of seismic anisotropy. In the same way, since the shear waves that arrive in the coda of the P wave are generated by P-to-S conversions on crustal discontinuities, their propagation distances are extremely short (of the order of a few wavelengths). We can thus safely neglect the effect of attenuation. We consider time windows that start 10 s before the onset of the P wave and end 40 s after. This choice for the length of the time window is dictated by the rather long durations of source wavelets (of the order of 30-40 s, depending on the event) and the necessity to include the contributions of P-to-S conversion and of the first multiples into the inversion. The main difference with the method described in Tong et al. (2014a) is that we invert broadband waveforms instead of receiver functions (i.e. components deconvolved from the vertical component). The idea is to retain the long-period components of the seismic wavefield to better constrain long wavelength heterogeneities but also to exploit the vertical-component waveforms.

The tomographic model is parameterized in terms of Vp and Vs values in a regular Cartesian grid, with 2 km cubic cells. The starting models are smooth 1-D density, Vp and Vs profiles derived from the ak135 reference Earth model (Kennett et al., 1995). For each source we compute the gradient of the waveform misfit function with respect to Vp and Vs using the adjoint method (Tromp et al., 2005). We also compute gradients for density, but since their amplitude are negligible compared to those for Vp and Vs, the density model is kept constant. We have checked that adding density perturbations in the inversion does not improve the waveform misfit. This suggests that seismic waveforms alone cannot constrain the density model, at least when considering periods longer than 5 s. We solve the non-linear inverse problem with an iterative L-BFGS method (Nocedal & Wright, 2006), starting from the initial smooth 1-D models. The selection of step lengths is performed using the Wolfe conditions, which also provide a stopping criterion (see Monteiller et al. (2015) for more details). The algorithm converges in 9 iterations, reaching a misfit reduction of about 50%.

Comparisons between real and synthetic seismograms computed in the starting 1-D model and in the final 3-D model are shown in Figs 4.45a–4.54a. Note that to produce these figures the real seismograms have been normalized to a unit maximum amplitude, and that the same normalization has been applied to the corresponding synthetic seismograms. Since the amplitudes on the radial components are 2-3 times smaller than on the vertical component, the radial component is thus amplified compared to the vertical component. In spite of a higher noise level on the horizontal component, the improvement of waveform fits is more important on the

radial component because the synthetic radial component computed in the starting smooth 1-D model only contain the contribution of the direct P wave. The final model better explains not only the phase and amplitude of direct P waves but also all the converted and multiply-reflected waves that arrive in the coda of P waves, mainly observed on the radial components. However, not surprisingly, the quality of the fit varies from one event to the other. For example, it is poorest for event 1, which has the smallest magnitude (6.1), while it is best for event 2, which has the largest magnitude (8.3).

5.8.4 Resolution test

In order to infer the spatial resolution given by our waveform dataset, we have performed a synthetic inversion experiment on a checkerboard model. This model consists of a mosaic of alternatively positive and negative 16% anomalies with respect to a smooth 1-D background medium. The anomalies have an infinite extension along the direction perpendicular to the strike of the seismic profile. The input Vp and Vs models are shown in Figs. 5.4C and 5.4A, respectively. The synthetic seismograms are computed in the input model, using the source wavelets that were determined for each teleseismic source. This means that both the frequency content of the signals and the path distribution are the same as those in the real data inversion. After 12 iterations of our iterative waveform inversion algorithm, we obtain the final Vp and Vs models shown in Figs. 5.4D and 5.4B, respectively. The checkerboard pattern is well retrieved down to 60 km depth in both the Vp and Vs models. In spite of a decrease of spatial resolution with depth, our method is capable of retrieving velocity anomalies that are smaller than both P and S wavelengths. In particular, the inversion is able to capture the inversions of the velocity gradient with depth, which would be impossible with classical travel time tomography.

5.8.5 Modeling of Bouguer gravity anomalies

For the modeling of Bouguer gravity anomalies, the Vp model is mapped to density using the relation $\rho = Vp/3 + 0.60$ (Birch, 1961), where density is in g·cm^{-3} and P velocity in km.s-1. For mantle rocks we use a constant density of 3.3 g·cm^{-3}. Density anomalies (Fig. 5.5) are calculated with respect to a reference homogeneous crust with a density of 2.7 g·cm^{-3} and a thickness of 35 km. Bouguer gravity anomalies (Fig. 5.6) are computed by integrating the contributions of density anomalies along finite-width horizontal line elements (Talwani, 1973). In our computations we consider that the width of the density anomalies in the direction perpendicular to the seismic profile is 20 km. This width is on par with the size of the Bouguer anomaly observed in the Mauléon basin.

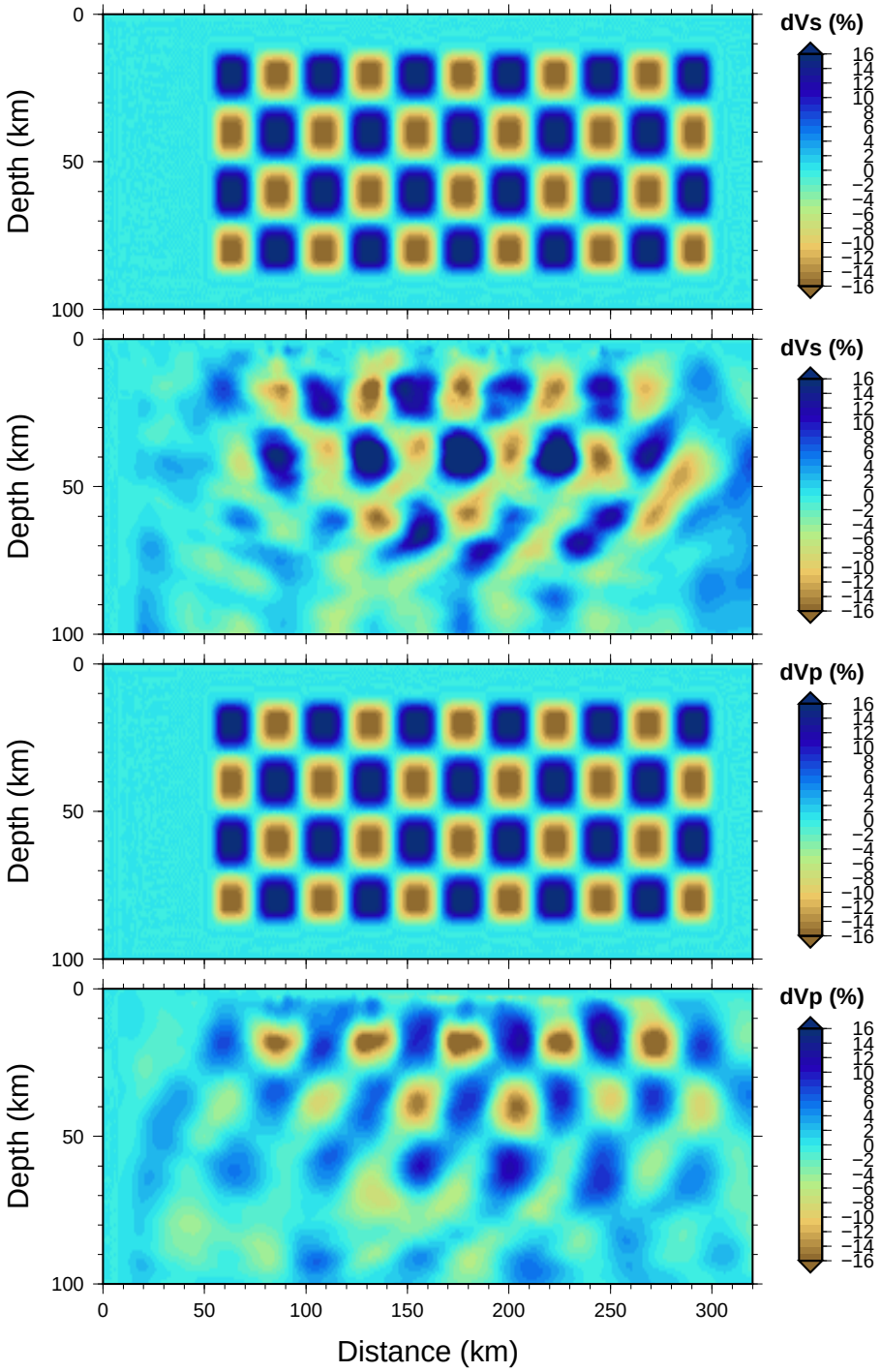


Figure 5.4: Results of the synthetic checkerboard inversion test. The figure shows the input Vp (C) and Vs (A) models, and the output Vp (D) and Vs (B) models.

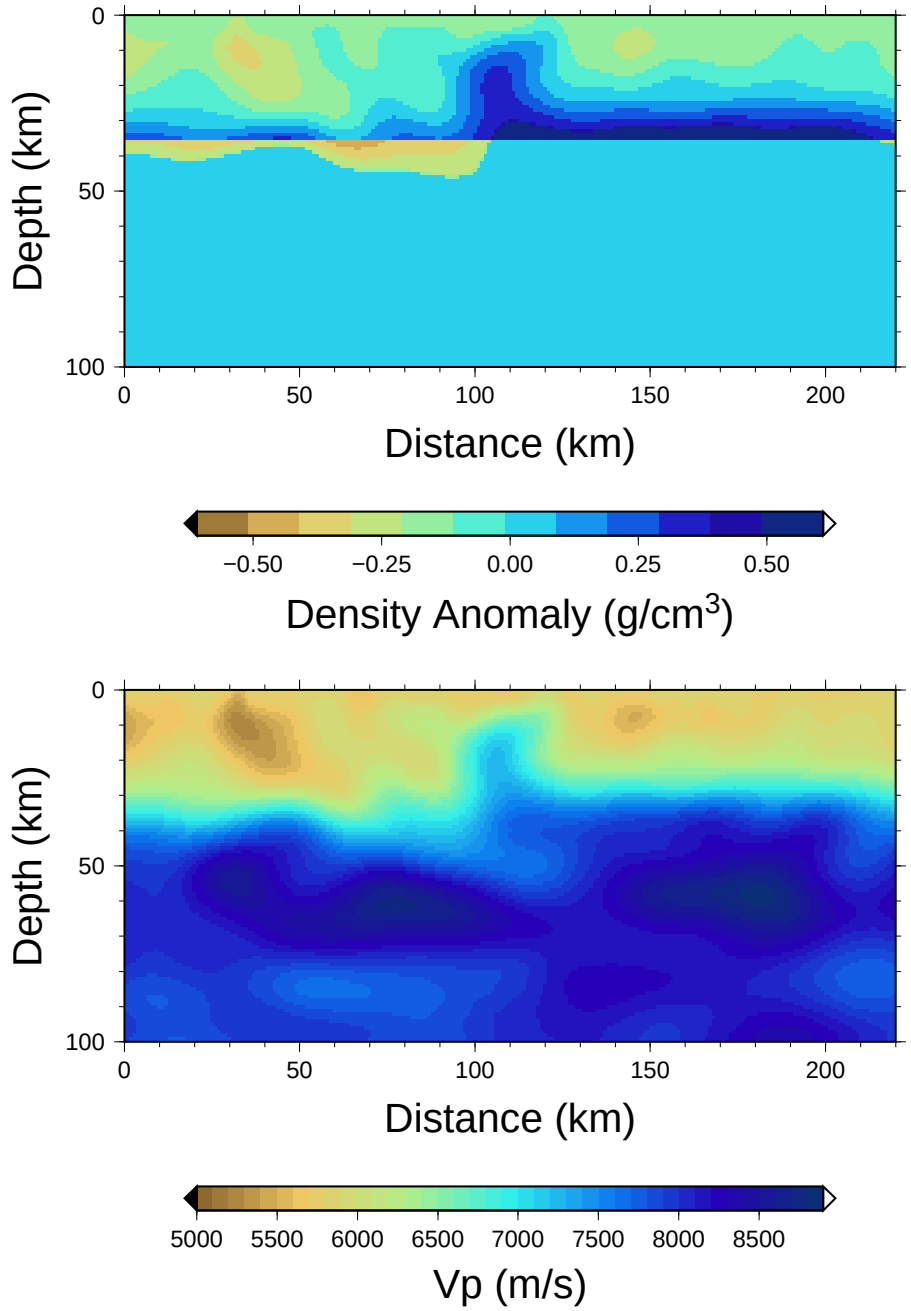


Figure 5.5: Model of density anomalies (top) derived from the P wave speed model (bottom).

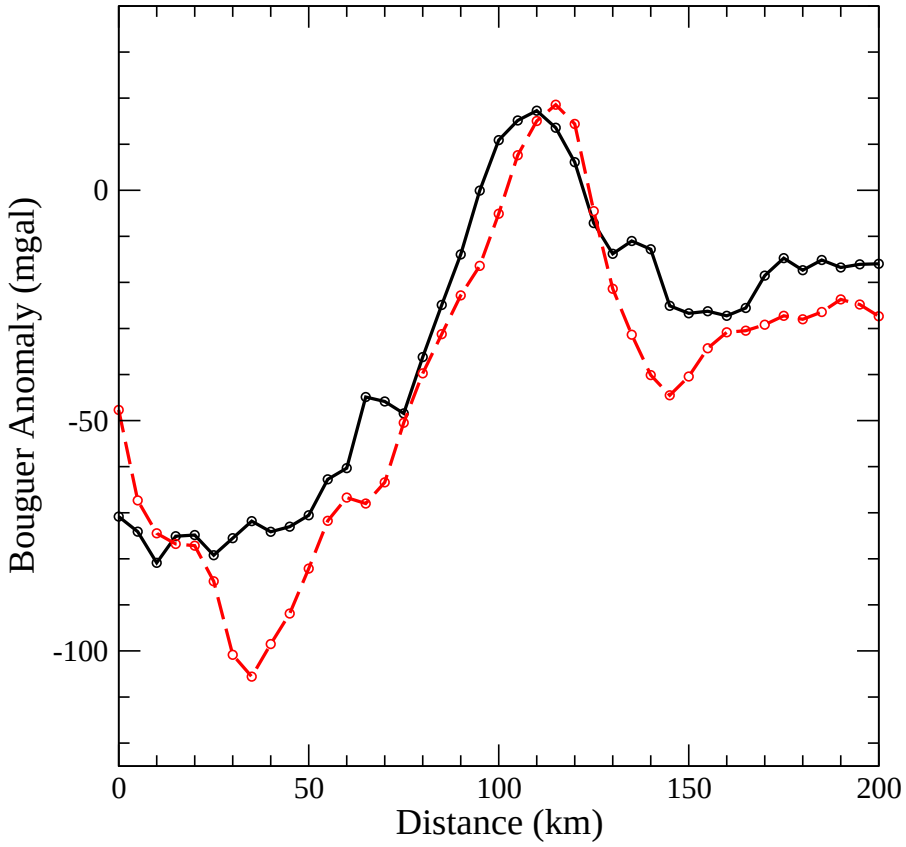


Figure 5.6: Computed (red dashed line) and observed (black solid line) Bouguer gravity anomalies.

Chapter 6

Applications of Full waveform inversion to the central Pyrenees transect

Contents

In this chapter, we will present the preliminary results of our FWI approach applied on the data of the central PYROPE transect.

6.1 Data selection

Table 6.1 shows the USGS source parameters of the five events recorded by the central transect which are used in our full waveform inversion. These events with Mw magnitudes between 6.0 and 6.7 were recorded by a total number of 75 stations (14 temporary broad-band transect stations on the French side, 18 temporary broad-band transect stations in the Spain side, 10 PYROPE broadband stations and 33 permanent stations from the French and Spanish networks). The five events are in a distance range between 30° and 85°. All the events occurred during the deployment of the central transect from December 2011 to September 2012. The focal depths are all located around 10 km depth. The azimuthal coverage for the central transect is quite good (in Figure 6.2), but is limited by a lack of events coming from the Southern Hemisphere.

We followed the same preprocessing of waveform data as for the western transect. After estimating the wavelets from the vertical component displacements, we visually select high quality traces based on the correlation coefficient between observed and synthetic vertical components, and only keep for inversion the traces which have a correlation coefficient larger than 0.7 to 0.85 with the synthetic seismograms, depending on the specific event. We also check the traveltime residuals of the direct P waves measured on the vertical components. Some stations with good correlation coefficients have anomalously large travel time residuals, which may result from clock drift in the recording system. We corrected clock problems by time corrections deter-

Source parameters	26/02/2012	02/04/2012	24/05/2012	10/08/2012	11/08/2012
Longitude(°)	5.71	95.436	-98.182	-167.468	47.128
Latitude(°)	72.96	51.831	16.576	52.095	38.378
focal depth(km)	9.75	11.0	11.0	11.0	11.0
Magnitude (M_w)	6.0	6.7	6.2	6.2	6.4
exponent($N{\cdot}M$)	18	19	18	18	18
M_{rr}	-2.890	0.60	-1.02	1.54	0.19
M_{tt}	1.250	-0.45	0.98	-0.96	0.26
M_{pp}	1.630	-0.15	0.04	-0.58	-0.45
M_{rt}	-0.256	-1.08	0.66	1.37	2.15
M_{rp}	-0.604	0.11	0.25	0.85	-1.16
M_{tp}	1.690	0.51	0.05	-0.82	-4.57

Table 6.1: The CMT solutions of the five events used for the central transect

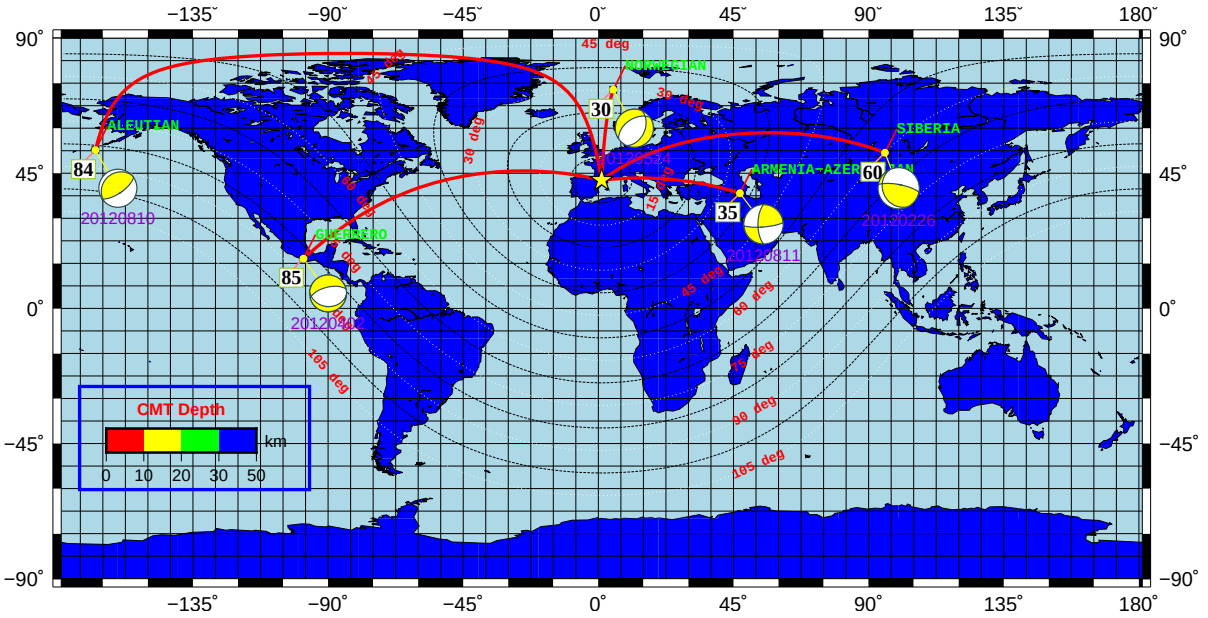


Figure 6.2: The Azimuthal coverage of the 5 events used in FWI.

mined from the temporal variations of the travel time of surface waves observed in empirical Green's functions obtained by ambient noise cross-correlations (Stehly et al., 2007). For a pair of stations 1 and 2, the phase of the Rayleigh waves traveling between 1 and 2 can be measured by cross-correlating ambient noise. If the noise sources are distributed evenly, the surface wave travel time should be the same on the positive (corresponding to the surface wave propagating from station 1 to 2) and on the negative cross-correlation time (corresponding to the wave propagating from 2 to 1), and should remain unchanged as time goes on. However, in practice, the fluctuations of the phase of surface waves for both positive and negative correlation times is often observed. There are three main factors that can give rise to these fluctuations (Stehly et al., 2007). First, any physical change in the media will lead to the same travel time fluctuation measured in both positive and negative sides of the cross-correlation. Second, a relative drift of the two station clocks results in an apparent smaller travel time in the positive correlation time, and larger travel time in the negative correlation time or vise versa. Third, variations in the spatial distribution of noise sources affect the positive and negative cross-correlation times independently. This is due to the sensitivity of cross-correlation to the distribution of noise sources.

Based on these considerations, we can simultaneously analyze the temporal evolution of surface wave travel time obtained from positive and negative cross-correlation times to separate the fluctuations related to physical change in the media, from those associated with a clock drift or other instrumental errors, or resulting from a change in the distribution of noise sources. This separation does not require any assumption of the velocity model. It only depends on the principle of time symmetry, which can be exploited for effective detection of clock drift problems.

We use the 10 months of records during the deployment of the central transect to extract the surface wave part of the Green's function between each pair of PYROPE stations in the central Pyrenees by performing cross-correlations of ambient noise. The cross-correlations are computed with the time length of one day and 5 days, respectively. According to the method of measuring the travel time variations introduced in Stehly et al. (2007), the apparent travel time fluctuations of surface waves for both positive and negative correlation times are then evaluated. The cross-correlation with such small time window (1 day or 5 days) can estimate the clock drift with a very fine temporal resolution. However, using short time windows is only possible to detect large instrumental errors (Stehly et al., 2007), because the travel time variations due to changes in the distribution of noise sources are relatively large.

In our study, we observed some variations with the same polarity appearing in the travel time fluctuations for both positive and negative correlation times. This observation can not be related to the change of the physical property of the media which would cause variations with opposite polarity for the travel time from positive and negative correlation times (Stehly et al., 2007). Therefore, these large amplitude travel time variations substantially suggest a relative

clock drift at one of the two stations.

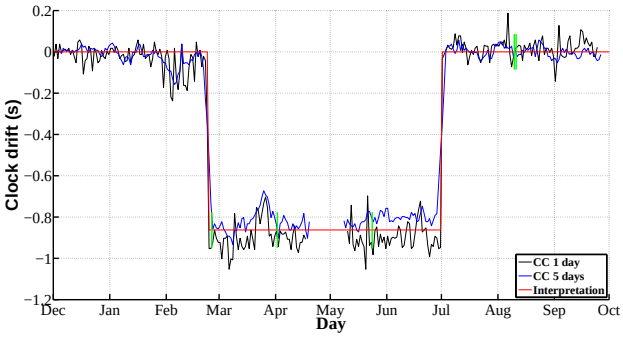
For each station, we determined the absolute clock drift by the median of all the relative clock drift values measured between itself and the other PYROPE stations. We found that stations PF01, PF03, PF06, PF09, PF12 and PF15 had a clock problem during the year 2012. This clock problem was already identified in the tomographic study by Chevrot et al. (2014). It comes from a software problem on several specific data loggers that were only used on the central transect. The absolute clock drifts of these stations for every day during the deployment of the central transect are shown in Figure 6.3. Estimates performed on the one day cross-correlations are shown with thin black lines, and those on the 5 days cross-correlations are shown with thin blue lines. In spite of small differences in detail, the two time series show very consistent changes. Figure 6.3 shows that the time series are discontinuous on some days. This is due to the instrumental failure or maintenance. We fit a step function to the apparent clock drift curves and filled the gaps due to lacking of data, as shown with thick red lines. The small oscillations of clock drifts around the step functions mainly result from the variation of background noise. We also label the values of clock drift of the six stations at the dates of the 5 earthquakes with thick green bars. The phases of the seismic records are corrected with these clock drifts.

After this initial preprocessing, 334 traces from 5 teleseismic events are kept for the central transect.

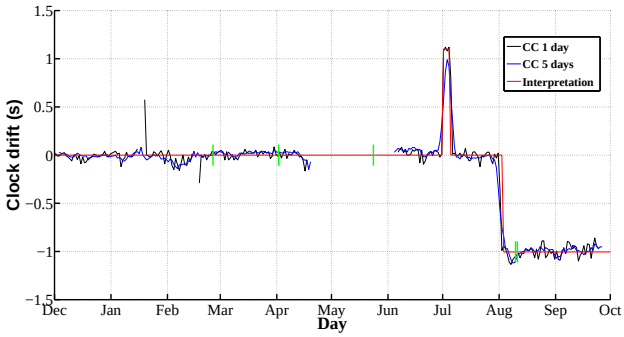
We use the optimal frequency range selection algorithm described in the previous chapter for the vertical and radial components waveform data. The S/N ratio of the data from the central transect is significantly lower than that of the western transect. After the optimal frequency range selection, Table 6.4 lists the number of traces of vertical and radial component waveforms before and after selection for the central transect.

The vertical components of the 5 events all show a good coherence between data and synthetics. As expected, the radial components are noisier than the vertical components. Compared to other events, the Apr 02 and Aug 10 earthquakes have fewer good quality radial component traces. Figure 6.5 and 6.6 show the comparisons between observed and synthetic waveforms for these events, as well as their coherence functions. From the two figures, it seems that the common incoherent frequency band is located between 0.05 to 0.1 Hz, in the frequency band of primary ocean microseisms. Some radial component waveforms of short period stations with reversed polarity, such as station FNEB and GRBF for Aug 10 event, were discarded. Few radial component traces for these two events are kept for inversion after our optimal corner frequency search procedure.

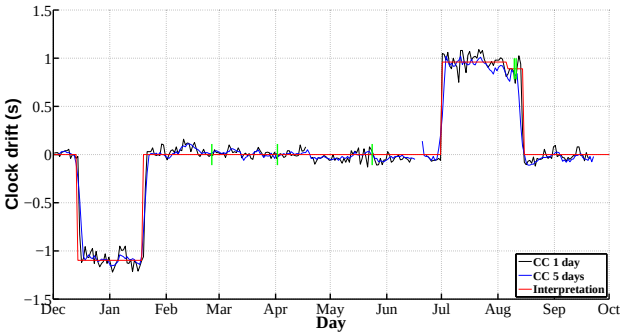
Figure 6.7, 6.8 and 6.9 show the 2D optimal corner frequency search procedure for radial components of the 2012 Apr 02 event recorded by transect stations PE09, PE17 and PF09, respectively. Their optimal lower corner frequencies used in FWI are set to more than 0.086 Hz, in order to filter out the large amplitude 20 s long period noise. After optimal selection, the numbers of available radial component traces and frequency range for Apr 02 and Aug 10 events



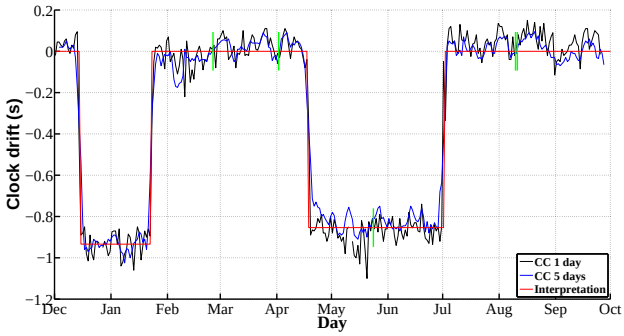
(a) The clock drift curve for PF01



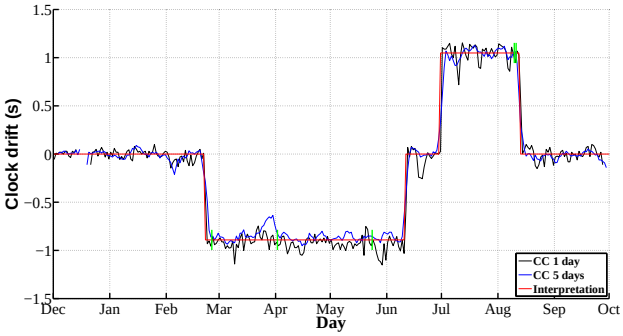
(b) The clock drift curve for PF03



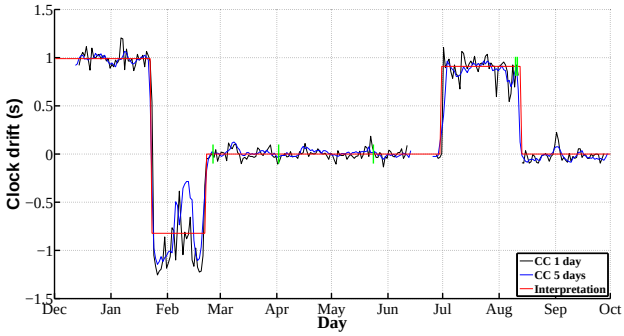
(c) The clock drift curve for PF06



(d) The clock drift curve for PF09



(e) The clock drift curve for PF012



(f) The clock drift curve for PF15

Figure 6.3: Clock drift curves estimated by one day and five days cross-correlation for six problematic PYROPE stations: (a) PF01, (b) PF03, (c) PF06, (d) PF09, (e) PF12, (f) PF15. Estimates performed on the one day cross-correlation are shown with thin black lines, and those on the 5 days correlation are shown with thin blue lines for each panel. In red thick line we show a step function to fit the clock drift curves. The values of clock drift of these stations at the dates of the 5 earthquakes are labeled with the thick green bars.

Number of component	Vertical trace after selection	Vertical trace before selection	Radial trace after selection	Radial trace before selection
26/02/2012	67	68	64	68
02/04/2012	70	70	39	70
24/05/2012	60	60	56	60
10/08/2012	65	68	45	68
11/08/2012	68	68	61	68

Table 6.4: The number of traces before and after selection for the central transect.

remarkably decrease, leading to lower radial component waveform misfits. Finally, we correct the amplitude of the stations with anomalous amplitudes (the absolute value of the amplitude anomaly is larger than 0.2) by an average amplitude of all the stations with moderate amplitude anomalies, to prevent them from dominating the waveform misfit.

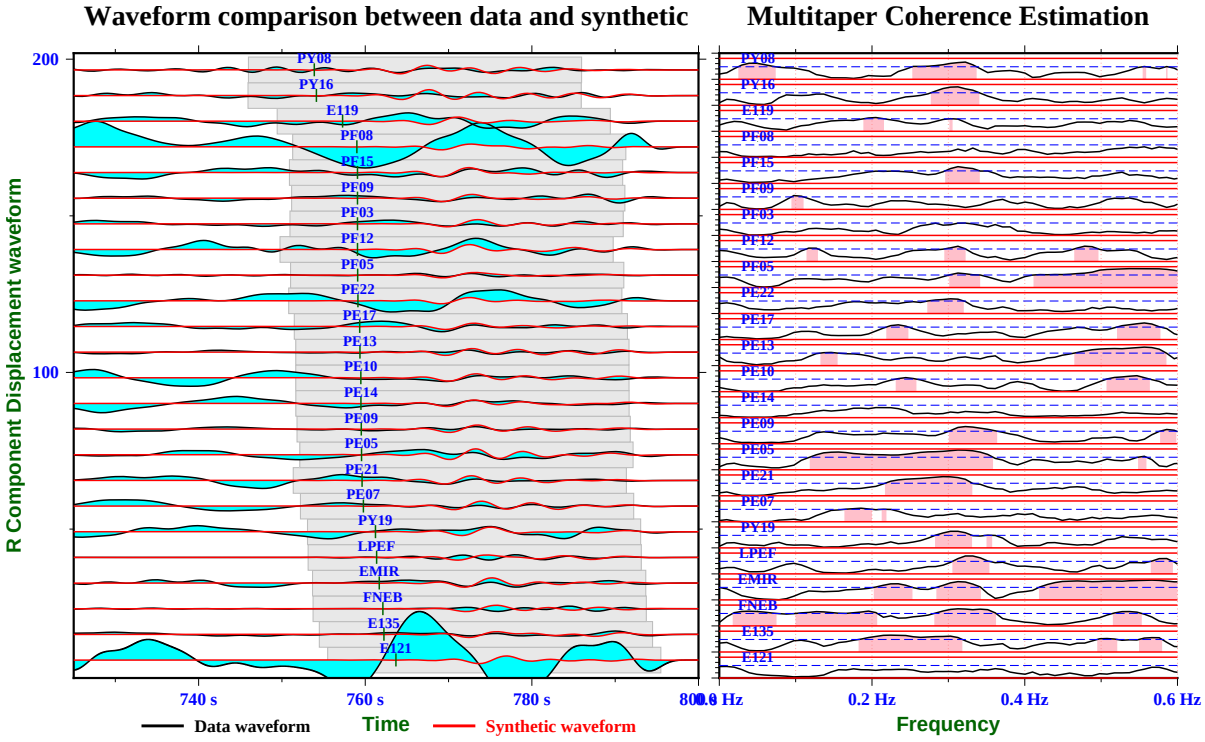


Figure 6.5: Radial component waveforms and corresponding coherence estimate for the traces of the 2012 Apr 02 Mexico Guerrero earthquake with low correlation coefficients.

6.2 Resolution analysis

We perform a resolution analysis for the central transect by using a simpler checkerboard test. Resolution of FWI is controlled by data coverage, the frequency range used in the inversion, and the estimated source wavelet functions. We assess the resolution of our FWI algorithm by

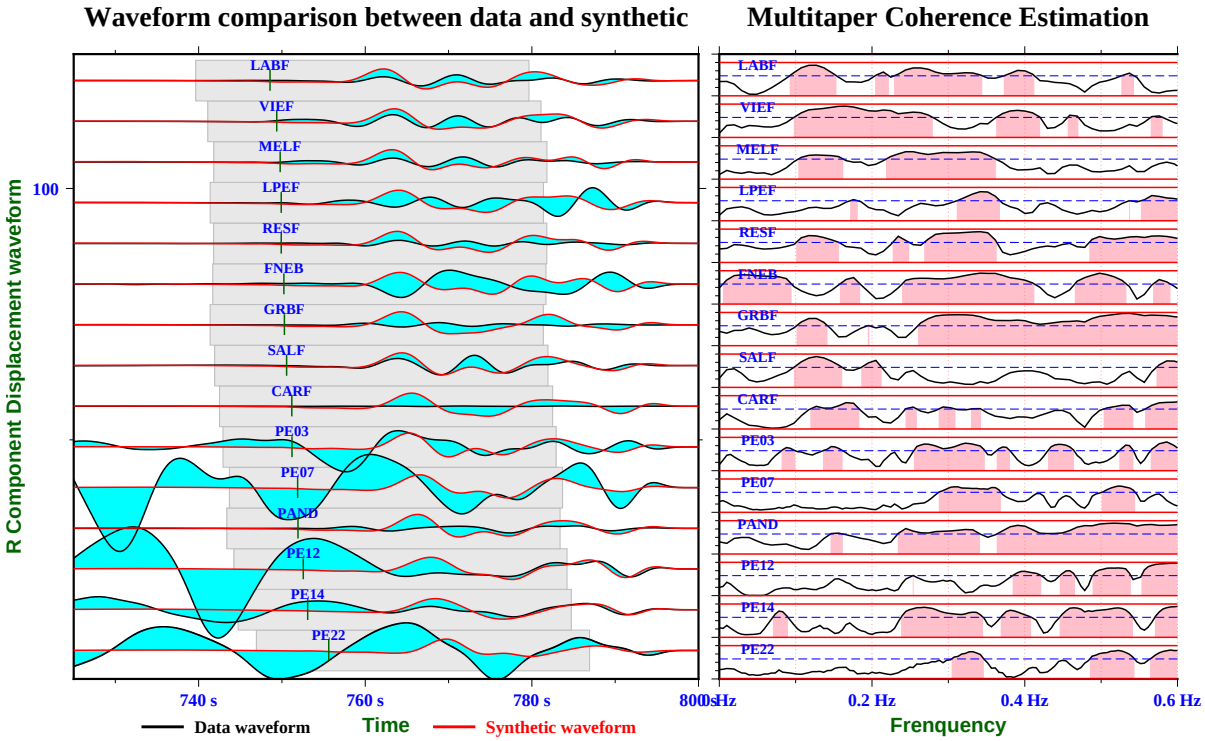


Figure 6.6: Radial component waveforms and corresponding coherence estimate for the traces of the 2012 Aug 10 Aleutian Islands earthquake with low correlation coefficients.

visual inspection of the recovered models and analysis of the data fit.

The input checkerboard model is a mosaic of alternatively positive and negative 16% anomalies with respect to a smooth 1D background velocity and density media, shown in the top right panel of Figure 6.10. Only two layers of checkerboard pattern anomalies with a size of 26×20 km are imposed between 20 km to 60 km depth. The inversion starts in the smooth 1D background model, shown in the middle right and bottom right panel of Figure 6.10. The anomalies have an infinite extension nearly perpendicular to the trend of the central transect as shown in the left panel of Figure 6.10.

The synthetic data are computed in the input checkerboard model, using the same source wavelet and receivers as in the real dataset. The synthetic data do not contain noise. In the following resolution tests, we always use the smooth 1D model as the starting model for the current waveform inversions. We perform various resolution tests based on different datasets and different strategies. We first define a time window that starts 15 s before the arrival time of the P wave and ends 30 s after. Based on the experience of the resolution analysis for the western transect, we do not update the density model in these new resolution tests. In the beginning, we perform two basic tests which are listed below:

1. We invert the vertical and radial component waveforms filtered in the optimal frequency bands determined before. The average filtering corner frequencies for these selected traces are shown in Table 6.11.

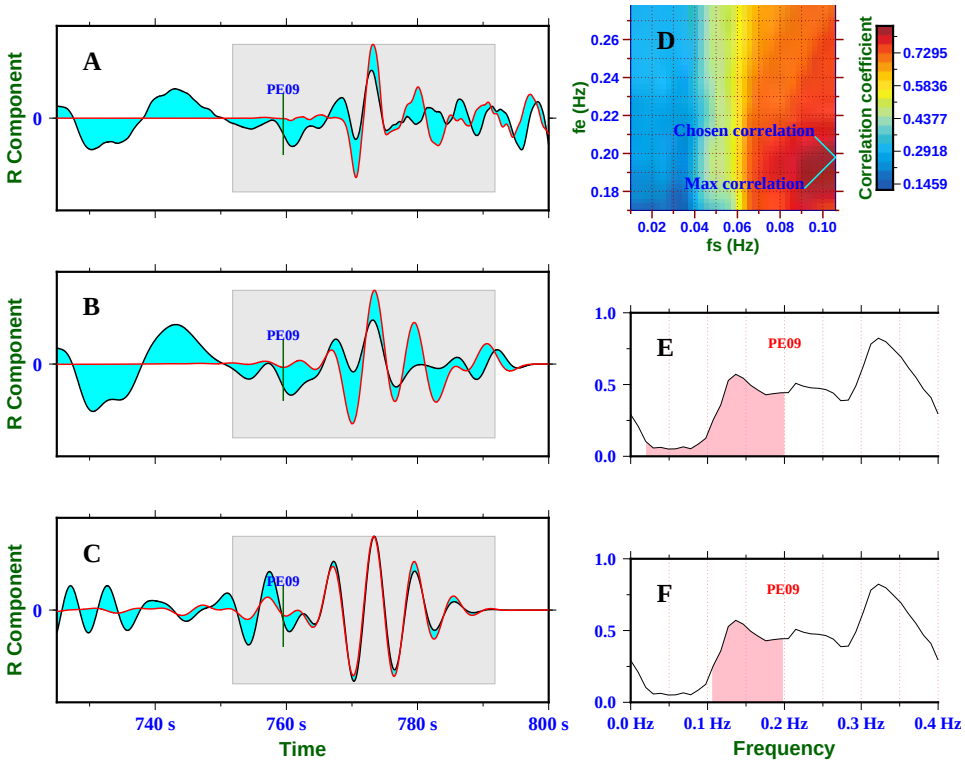


Figure 6.7: 2D corner frequency search for the radial component of station PE09 for the 2012 April 02 event. Panel (a) shows the unfiltered waveform, (b) the waveform filtered at the initial preset corner frequencies 0.035 Hz and 0.2 Hz, and (c) the waveform filtered at the new corner frequencies 0.106 and 0.2 Hz. Panel (d) shows the 2D objective function Vs. fs and fe. The maximum correlation coefficient 0.84 is obtained with these new corner frequencies: $fs = 0.106$ Hz and $fe = 0.2$ Hz. Panel (e) and (f) show the coherence for this trace, with pink shaded area bounded by the two sets of corner frequencies corresponding to (b) and (c).

2. We use all available vertical and radial component waveforms filtered at optimal corner frequencies. The filtered vertical component waveforms are the same as in test 1, but we use all the corresponding radial component waveforms in this new test. For the radial component traces that are also used in test 1, we filtered them at their optimal corner frequencies. For the radial component traces that are newly added in this test, we filter them at corner frequencies corresponding to their vertical counterparts.

Figure 6.12 shows the surface projection of 220 km long vertical section along the central transect with blue line. The southern termination of the transect is station PE22. The incoming angle of the five teleseismic P waves used here are shown with thick red lines and labeled according to their dates of occurrence. Figure 6.12 also shows the surface projections of the three sets of vertical sections parallel to the central vertical section at a spacing of 20 (sandy lines), 15 (dark green lines) and 10 km (cyan lines), respectively. All the temporary stations on the central transect are bounded by the three sets of lines.

In the following, the central vertical section of the inverted model will be shown in any case. It is directly extracted from our 3D inverted models. We also computed the integration of the model in the horizontal direction perpendicular to the central vertical section. The range of

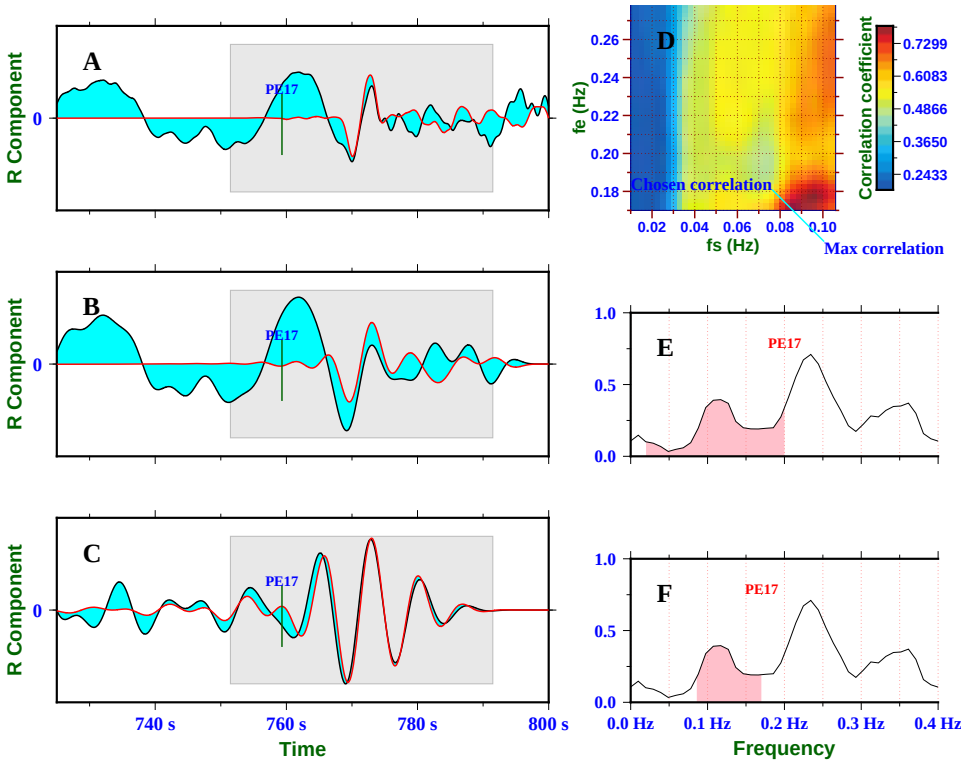


Figure 6.8: 2D corner frequency search for the radial component of station PE17 for the 2012 Apr 02 earthquake. Panel (a) shows the unfiltered waveform, (b) the waveform filtered at the initial preset corner frequencies 0.035 Hz and 0.2 Hz, and (c) the waveform filtered at the new corner frequencies 0.086 and 0.17 Hz. Panel (d) shows the 2D objective function Vs. fs and fe. The maximum correlation coefficient 0.78 is obtained with these new corner frequencies: $fs = 0.086$ Hz and $fe = 0.17$ Hz. Panel (e) and (f) show the coherence for this trace, with pink shaded area bounded by the two sets of corner frequencies corresponding to (b) and (c).

integration is limited to the three sets of parallel vertical sections shown in Figure 6.12. After normalization of these integrals, we obtain a new 2D vertical section. This average 2D vertical section enhances the major structural features beneath the central transect. One reason for this averaging is that the incident teleseismic wavefields sample the structure beneath the transect from different azimuths and dip angles. Even though we can sometimes record earthquakes almost ideally located along the strike of the transect (such as the 2012 May 24 Norwegian Sea earthquake), most earthquakes never occur along the azimuth of the acquisition. When a small number of events are used in waveform inversion, the 3D spatial coverage of the incident wavefields is generally insufficient to constrain properly the deep structures. The average 2D vertical section of the model includes the contributions from off-line propagations. Note that we use the records from the 2012 May 24 Norwegian Sea earthquake, which is located almost along the azimuth of the central transect at the smallest epicentral distance 30° among the five events. All the transect stations recorded this event with a relatively high S/N ratio. Coherent Pms phase is clearly identified along the whole transect (Chevrot et al., 2015). The amplitude of the event kernel for this event is significantly larger than the other 4 events, which may result in the waveform inversion being dominated by this event, especially for the structures beneath

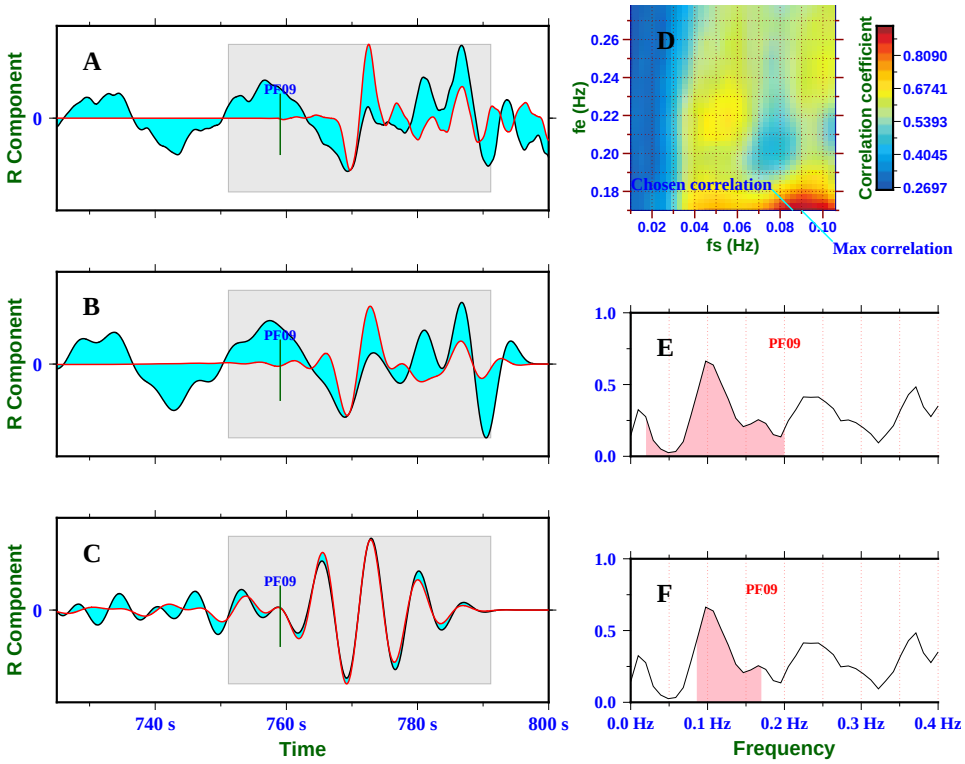


Figure 6.9: 2D corner frequency search for the radial component of station PF09 for the 2012 Apr 02 earthquake. Panel (a) shows the unfiltered waveform, (b) the waveform filtered at the initial preset corner frequencies 0.035 Hz and 0.2 Hz, and (c) the waveform filtered at the new corner frequencies 0.086 and 0.17 Hz. Panel (d) shows the 2D objective function Vs. fs and fe. The maximum correlation coefficient 0.93 is obtained with $fs = 0.09$ Hz and $fe = 0.17$ Hz. This panel also shows that the optimal corner frequencies fs and fe which keep a relatively high correlation coefficient 0.9 are 0.086 and 0.17 Hz, respectively. Panel (e) and (f) show the coherence for this trace, with pink shaded area bounded by the two sets of corner frequencies corresponding to (b) and (c).

the transect. For this reason, the spatial averaging allows us to balance the sensitivity and enhance the recovery of deep structures.

After 7 and 9 iterations of the L-BFGS algorithm, we obtain the final Vp and Vs models shown in Figure 6.13 and 6.14 for resolution tests 1 and 2, respectively. The relative misfit residuals for these two tests are 0.18 and 0.27, respectively. We compare the vertical sections in the resulting models for these two tests, with different average range. The target input checkerboard model is also shown in Figure 6.13b as a reference.

Both resolution tests lead to comparable results, which proves again that our optimal frequency range selection procedure is well suited for waveform inversion. Comparing the inverted Vp model with and without horizontal average, we can see that the artifacts close to the stations are reduced by the horizontal average. When the horizontal average range increases from 10 km to 20 km, the recovery of the anomalies located at depths between 20 and 60 km is slightly improved. The strong oblique smearing appearing reflects the dominant contribution of the shallow 2012 May 24 event. Even with horizontal average, these smearing effects are still

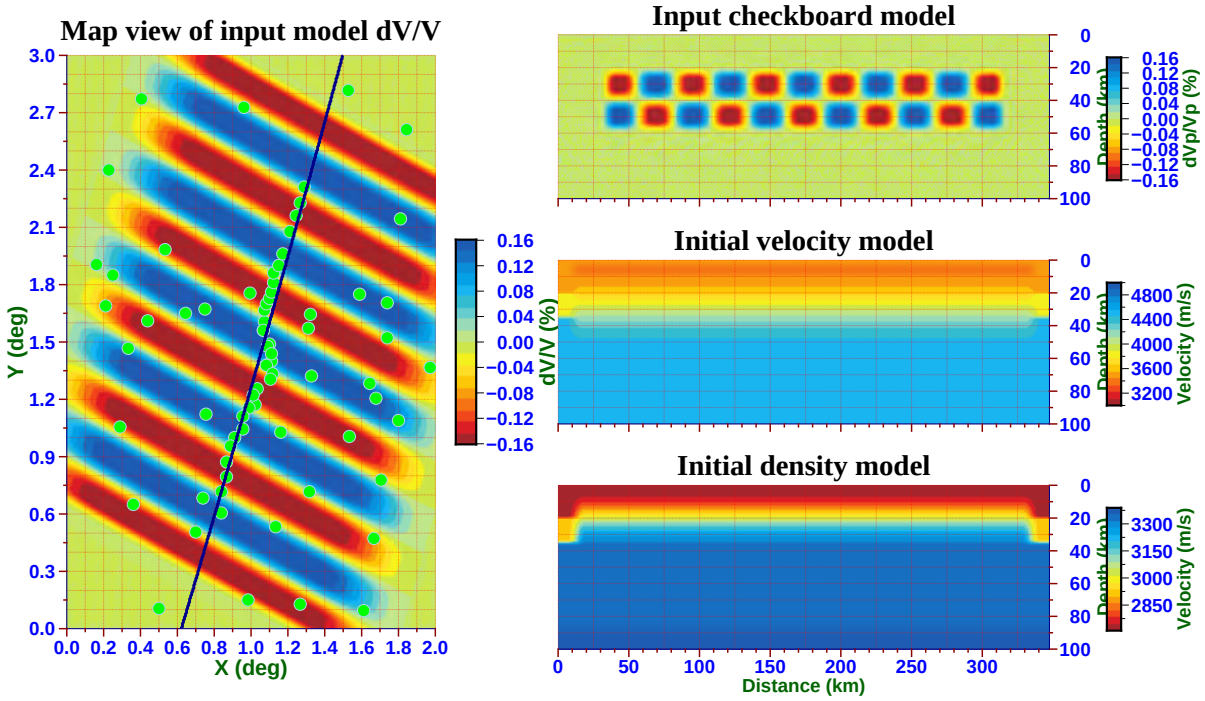


Figure 6.10: Map and vertical section views of checkerboard model for the central transect.

Number of component	fs for vertical component	fe for vertical component	fs for radial component	fe for radial component
26/02/2012	0.013	0.2	0.0255	0.192
02/04/2012	0.0385	0.2	0.0454	0.161
24/05/2012	0.0298	0.202	0.0395	0.186
10/08/2012	0.0263	0.189	0.0404	0.1782
11/08/2012	0.031	0.2	0.0515	0.199

Table 6.11: The average corner frequencies for each event of the central transect after the optimal frequency range selection.

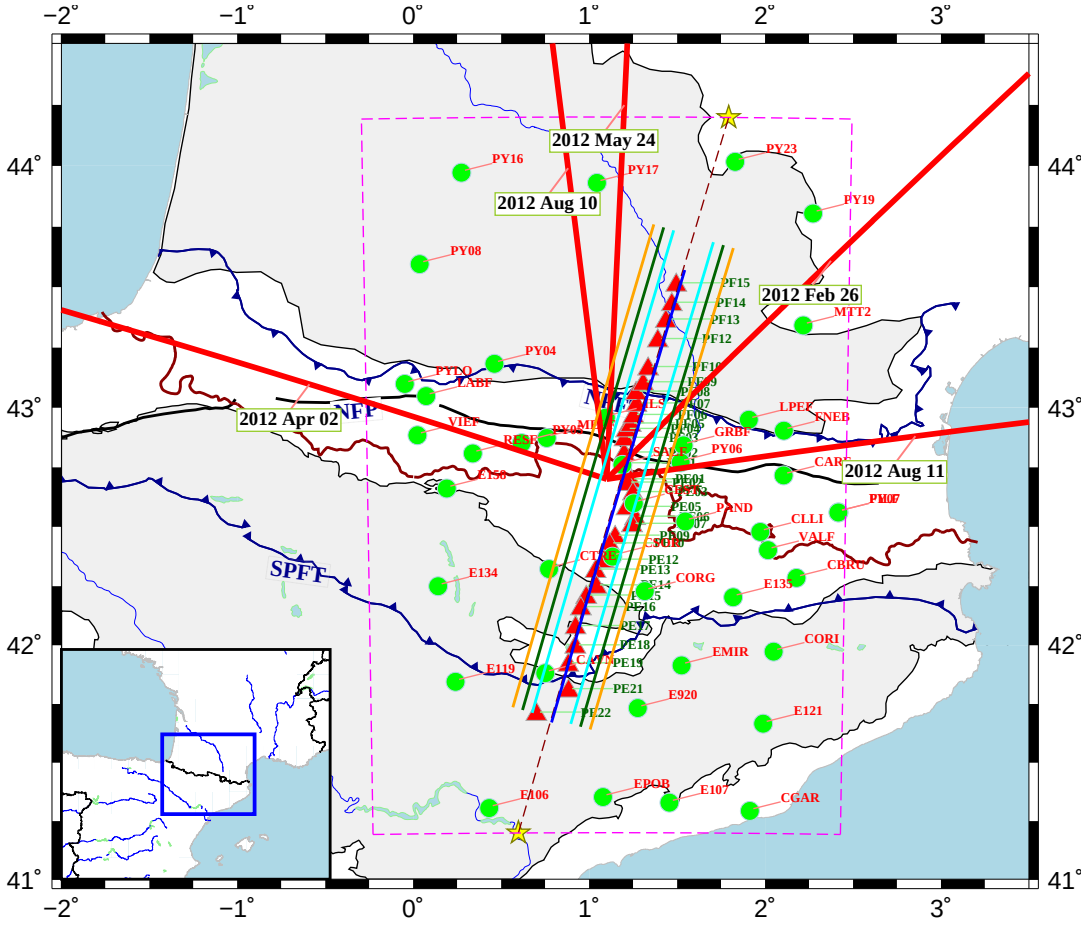
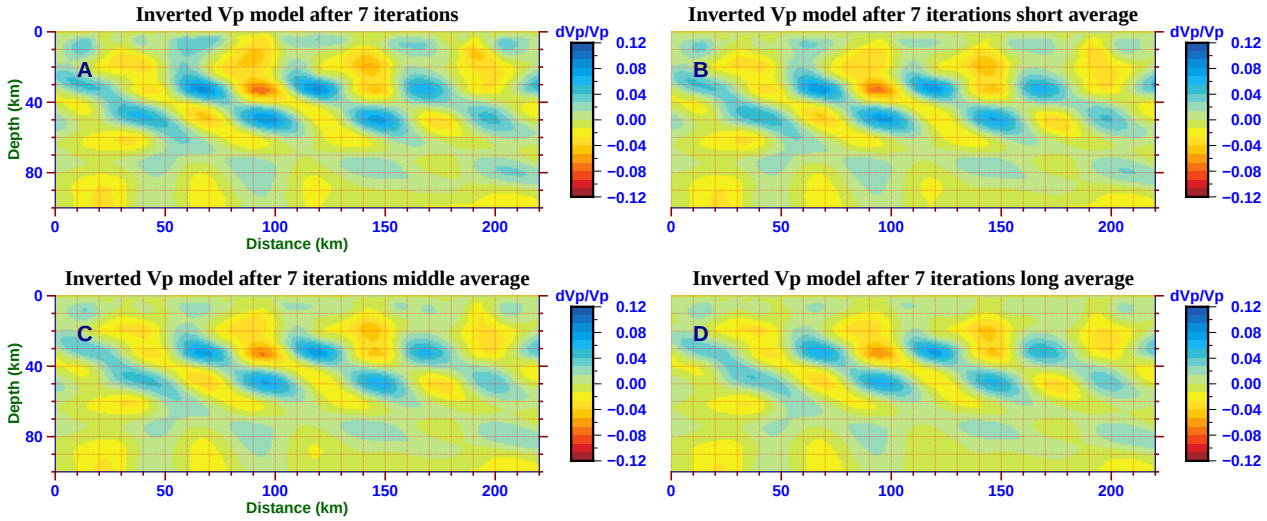


Figure 6.12: Map view of projections of the parallel vertical sections along the strike of the central transect, shown with thick blue lines. The azimuths of the five earthquakes are shown with thick red lines.

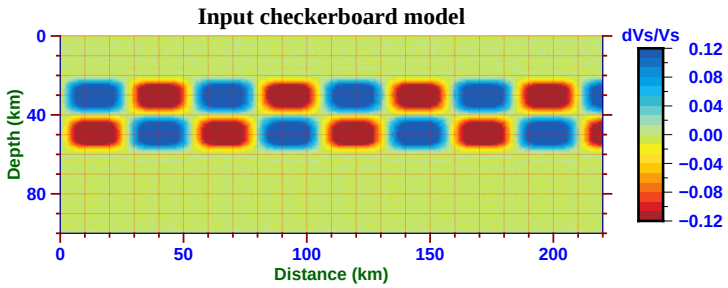
present in the inverted models. The recovered Vp models in both tests are acceptable, but the Vs models are not well recovered. Both tests failed to recover the target anomalies below 40 km depth.

The resolution tests for the western transect showed that the more complex four-layer target anomalies could be resolved down to a depth of 60 km. We speculate that there are other factors that may contribute to the lower depth resolution for the central transect. First, the time window used here ends 30 s after the arrival time of the P wave, which is 10 s earlier than the end time of the window used in FWI for western transect. Second, the shape and timing of source wavelets will affect the waveform misfits significantly. Third, the five events used in the central transect inversion are coming from the north, with no source coming from the south. In contrast to the western transect, this asymmetric ray coverage may degrade the resolution. In the following, we will look for appropriate strategies to mitigate these problems.

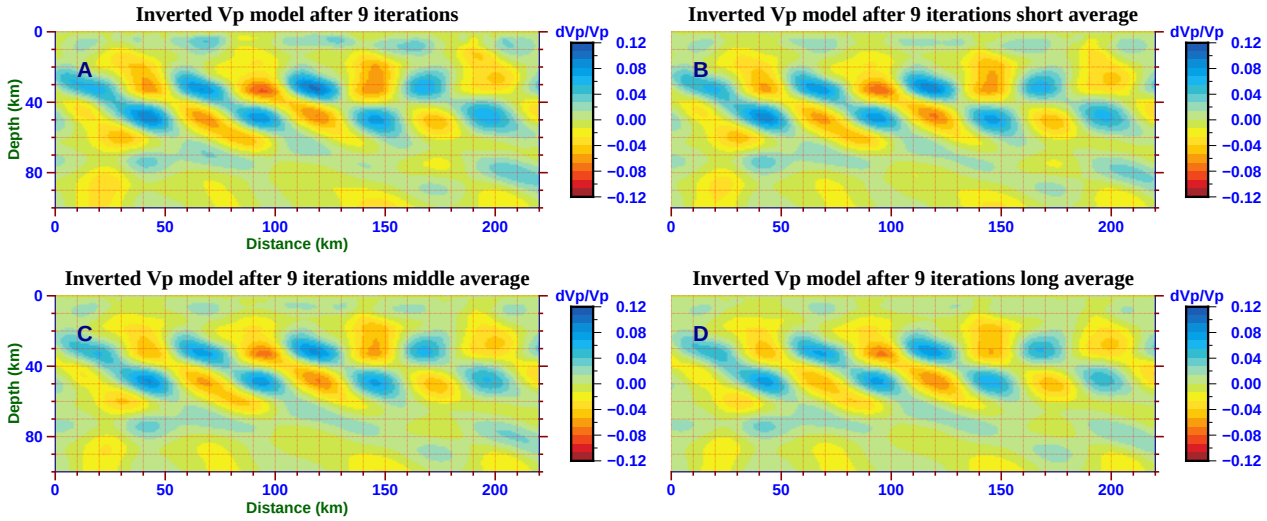
We carefully examine the waveform misfit reduction, the variation of waveform fit and the gradient for each event. Table 6.15 shows the vertical and radial component waveform misfits of each event before FWI and the relative misfit residuals after FWI for resolution test 1. Among



(a) Inverted Vp model for resolution test 1

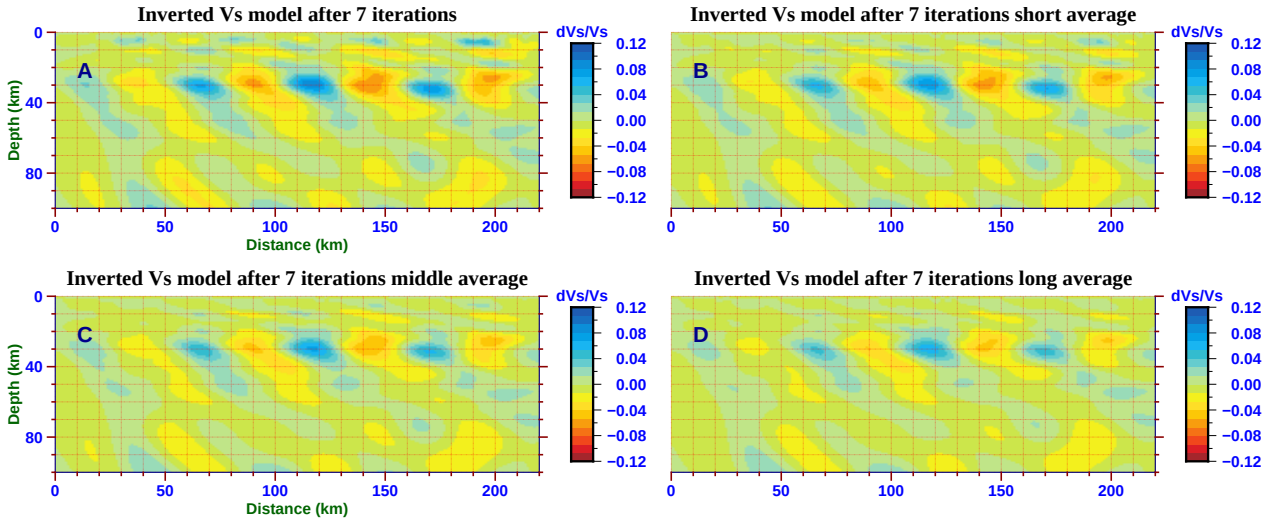


(b) Input checkerboard model

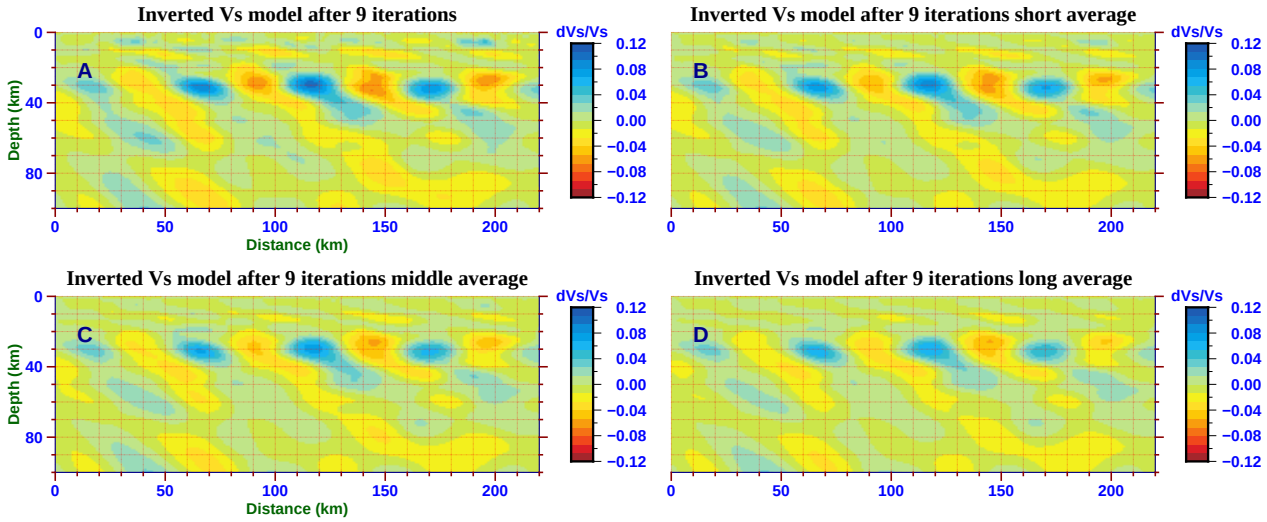


(c) Inverted Vp model for resolution test 2

Figure 6.13: Final Vp model for resolution test 1 (a) and 2 (b). In each figure, panels A, B, C, D show the vertical section extracted from the 3D model, averaged section over a horizontal range of 10 km, 15 km and 20 km, respectively.



(a) Inverted Vs model for resolution test 1



(b) Inverted Vs model for resolution test 2

Figure 6.14: Final Vs model for resolution test 1 (a) and 2 (b). In each figure, panels A, B, C, D show the vertical section extracted from the 3D model, averaged section over a horizontal range of 10 km, 15 km and 20 km, respectively.

the five events, the relative misfit residuals are the largest for the 2012 Apr 02 event in the radial component and Aug 10 event in the vertical component. The 2012 May 24 earthquake has the smallest relative misfit residual for both components. Here we show the map of radial component waveform misfit reduction and the comparisons between observed and synthetic radial component seismograms along the transect for the 2012 Apr 02 event, along with the map of vertical component waveform misfit reduction and the vertical component waveform comparisons for Aug 10 event.

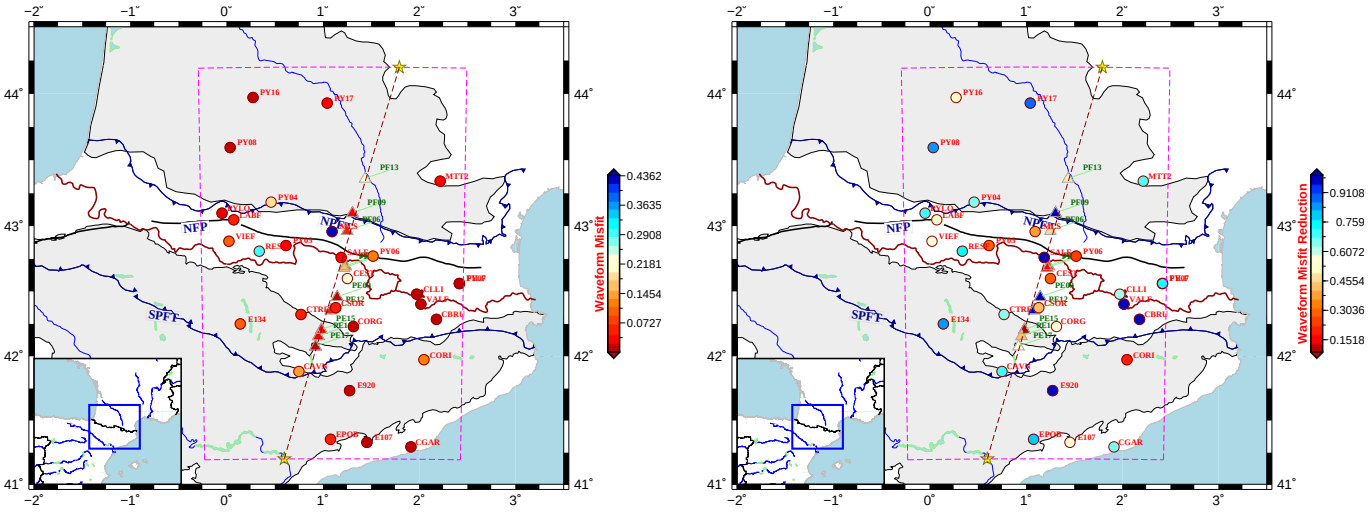
waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
26/02/2012	25.475	0.1832	18.4078	0.2222
02/04/2012	44.147	0.1406	15.5099	0.5250
24/05/2012	18.1229	0.1296	33.1522	0.0668
10/08/2012	27.4925	0.3657	13.2558	0.2684
11/08/2012	21.9103	0.1957	48.9967	0.1307

Table 6.15: Waveform misfits and relative misfit residuals for each event in resolution test 1.

As can be seen in Figure 6.16 and 6.17, the 2012 Apr 02 event retains the least amount of radial component waveform traces along the transect after optimal frequency range selection, resulting in the lowest waveform misfit reduction. In addition, the azimuth of this event is far from the other four events and is nearly perpendicular to the strike of the transect. This also leads to uneven coverage. One obvious reason for the low misfit reduction of vertical component waveforms for the August 10 event is the small vertical component waveform misfit value before inversion (Figures 6.18 and 6.19), which means that the fit was already quite a good event before starting the inversion.

Vertical sections of the event kernels for the five events computed in the initial smooth 1D model are shown in Figure 6.20. Although the waveform misfit of each event is normalized by its average trace energy, the amplitudes of both vertical and radial component waveform event kernels for the May 24 event are larger than for the other events. The oblique distribution of the event kernel patterns along the wave propagation path can be seen clearly, especially for the Vs kernel, which is responsible for the oblique smearing in the final model. These 'footprints' of ray paths are particularly evident for the Feb 26, May 24 and Aug 10 event since their azimuths are all approximately along the strike of the transect. For the other two events, this problem is less severe.

We try to find some ways to improve our inversion results based on this simple checkerboard test. An intuitive solution is to weight each event kernel to balance the total gradient as well as possible. Moreover, in order to better invert the target anomalies in the deep part, we select different preconditioners to weight the total gradient. Below we will set up two new resolution tests to explore the feasibility of these two ideas.



(a) Waveform misfit before FWI

(b) Waveform misfit reduction after FWI

Figure 6.16: (a) Map of waveform misfit for the radial component of the 2012 April 02 event before FWI. (b) Map of waveform relative misfit residual for the radial component of the 2012 April 02 event after FWI.

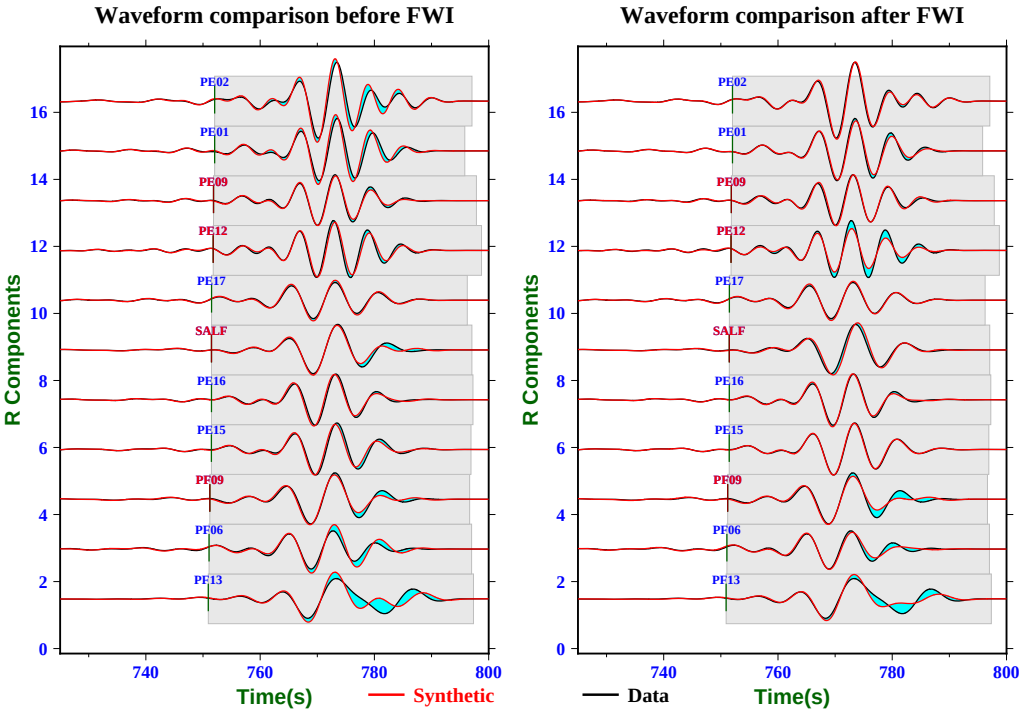
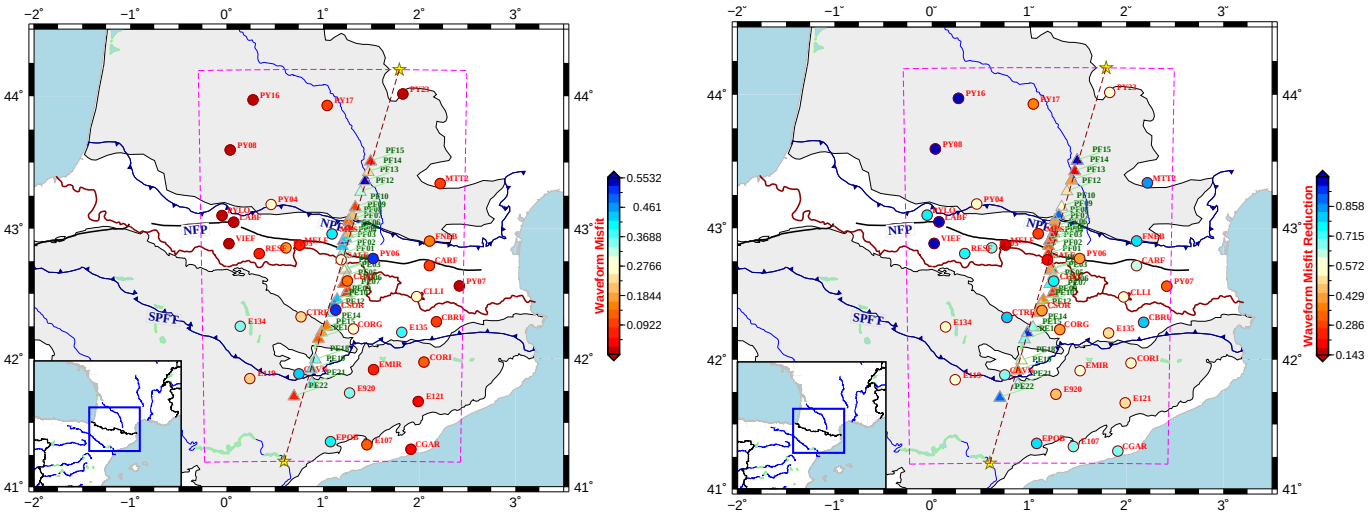


Figure 6.17: Observed and synthetic radial component waveforms along the transect for the 2012 April 02 event.



(a) Waveform misfit before FWI

(b) Waveform misfit reduction after FWI

Figure 6.18: (a) Map of waveform misfit for the vertical component of the 2012 August 10 event before FWI. (b) Map of waveform relative misfit residual for the radial component of the 2012 August 10 event after FWI.

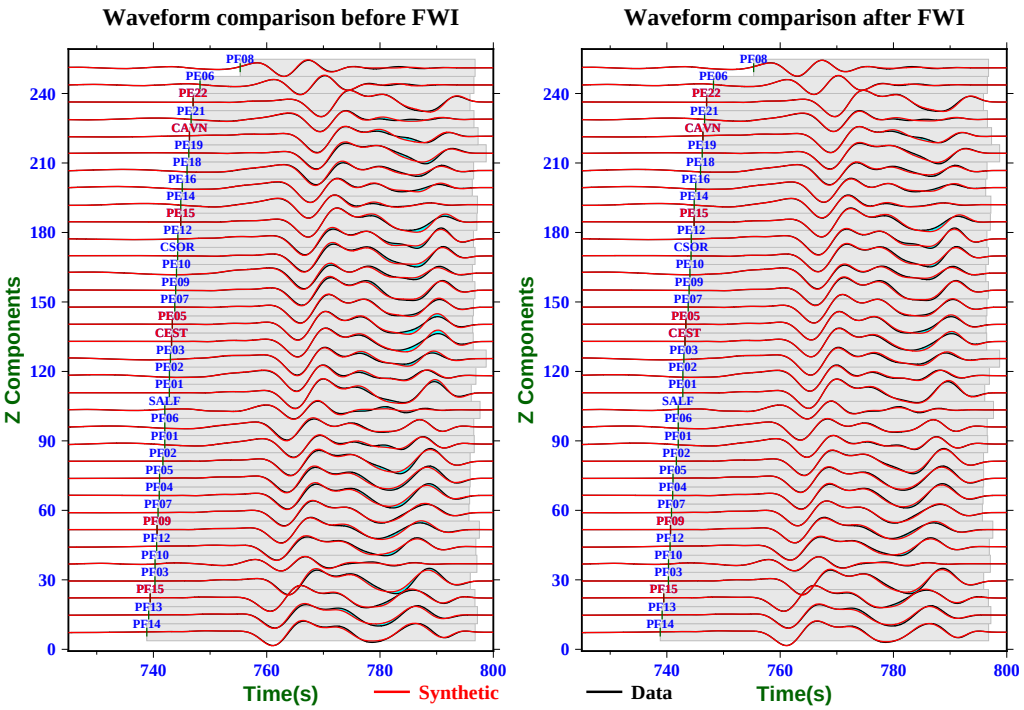


Figure 6.19: Observed and synthetic vertical component waveforms along the transect for the 2012 August 10 event.

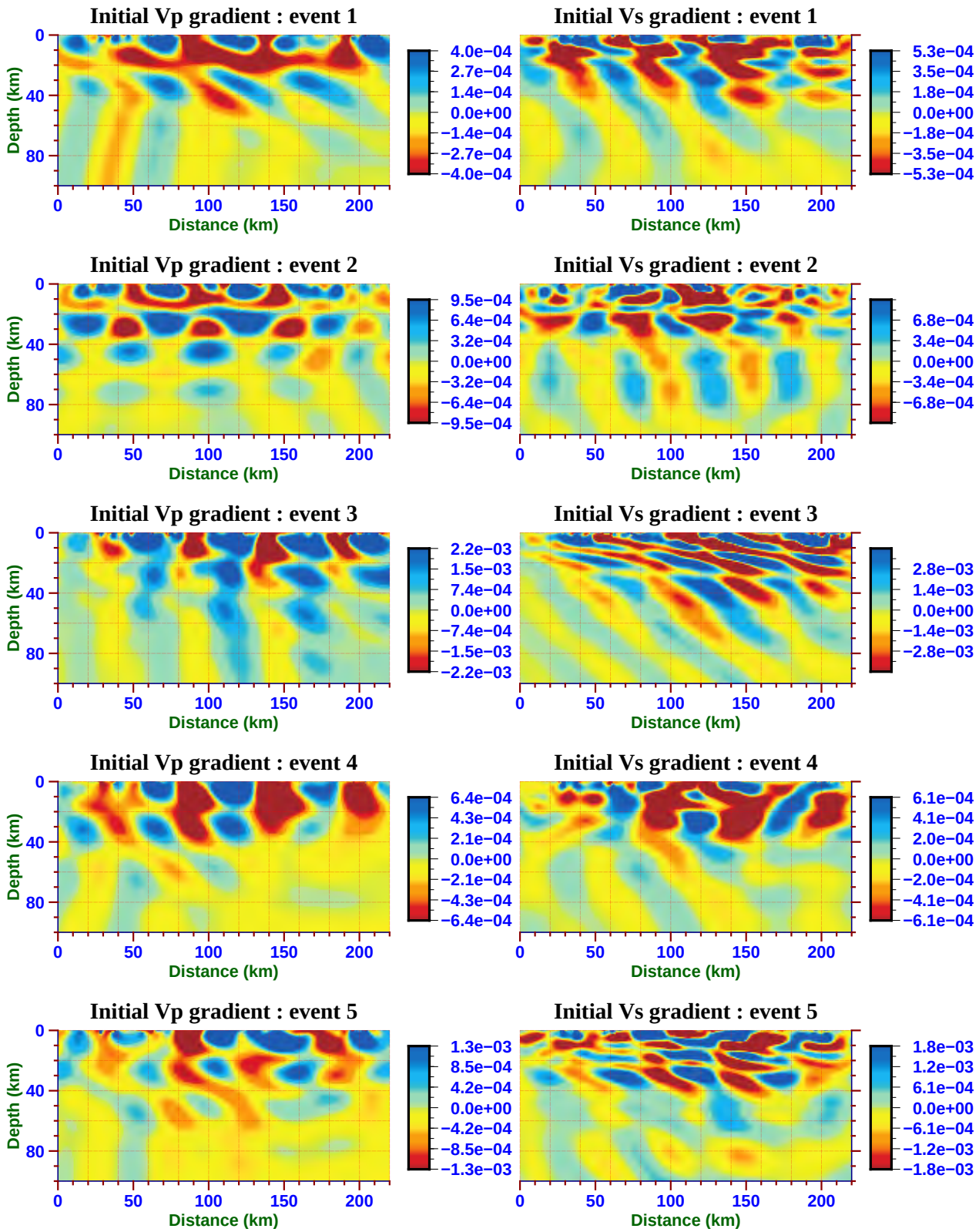


Figure 6.20: Vp (left) and Vs (right) event kernels (gradients) for each event computed in the initial smooth 1D model for resolution test 1.

3. We use the reciprocal of the L1 norm of the individual event kernel along the vertical section to balance their contributions before summing them. We simultaneously use the depth value at every grid point as the preconditioner to precondition the total gradient. The dataset and filters are the same as in the second test.

4. The same as in test 3, but keeping the previously used square root of depth preconditioner.

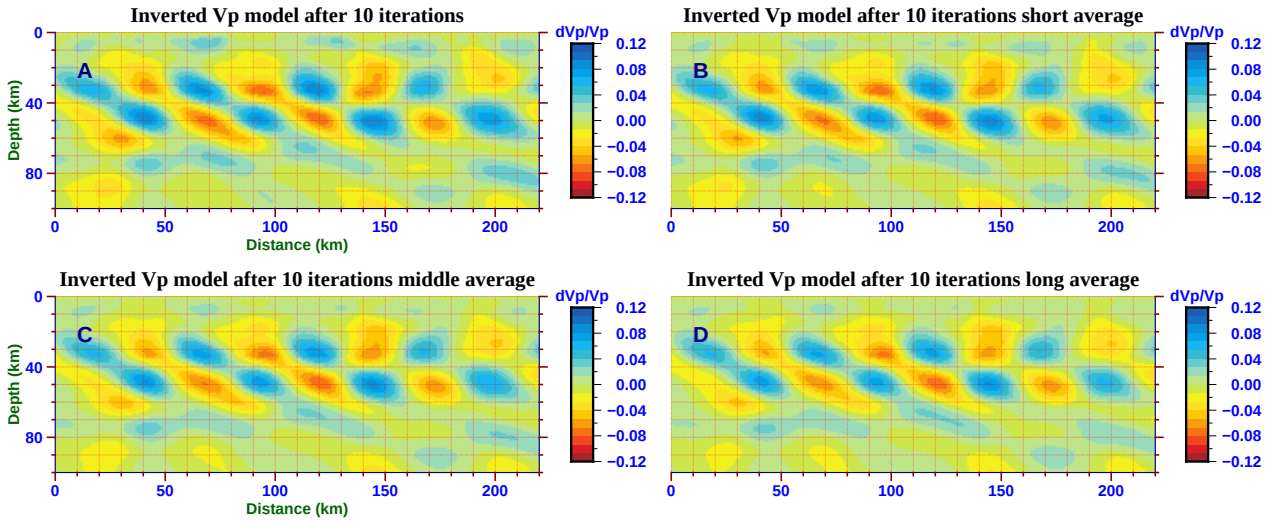
After 10 and 11 iterations, we get the final Vp and Vs models shown in Figure 6.21 and 6.22 for resolution tests 3 and 4, respectively. The relative misfit residuals in the final models of these two tests are 0.285 and 0.161, respectively. These two tests recover the target anomalies in the shallow part of the Vs model better than the previous two tests and mitigate the oblique smearings. The differences between inverted models of tests 3 and 4 are not significant. However, test 4 has a larger waveform misfit reduction. The problem of poorly resolved Vs model in the deep part remains unresolved, even after adjusting the weights for the event kernels whatever preconditioner used.

A possible direction is the multiscale strategy or hierarchical algorithm, which progressively moves the inversion towards the shorter scales on shorter periods. When the inversion is hierarchically performed towards high frequencies, shorter wavelengths are injected in the lithospheric model. Since the poor resolution of Vs model in the deep part may be limited by relatively higher velocity there and larger Fresnel volumes of the sensitivity kernel, using higher frequencies can decrease the wavelength of the seismic wave to image this part with higher resolving power. Therefore, we perform the inversion from the final model obtained in test 4, by increasing the upper corner frequency to 0.4 Hz and keeping the same time window and dataset. The inversion converged after 8 iterations. The final Vp and Vs models are shown in Figure 6.23. The relative misfit residual after inversion is 0.34. In this test, we finally begin to recover the target anomalies in the deep part of Vs model. Compared to the previous tests, the amplitudes of the target anomalies in the Vp and Vs models are better recovered. Unfortunately, the shapes of the deep Vs anomalies are still not well retrieved.

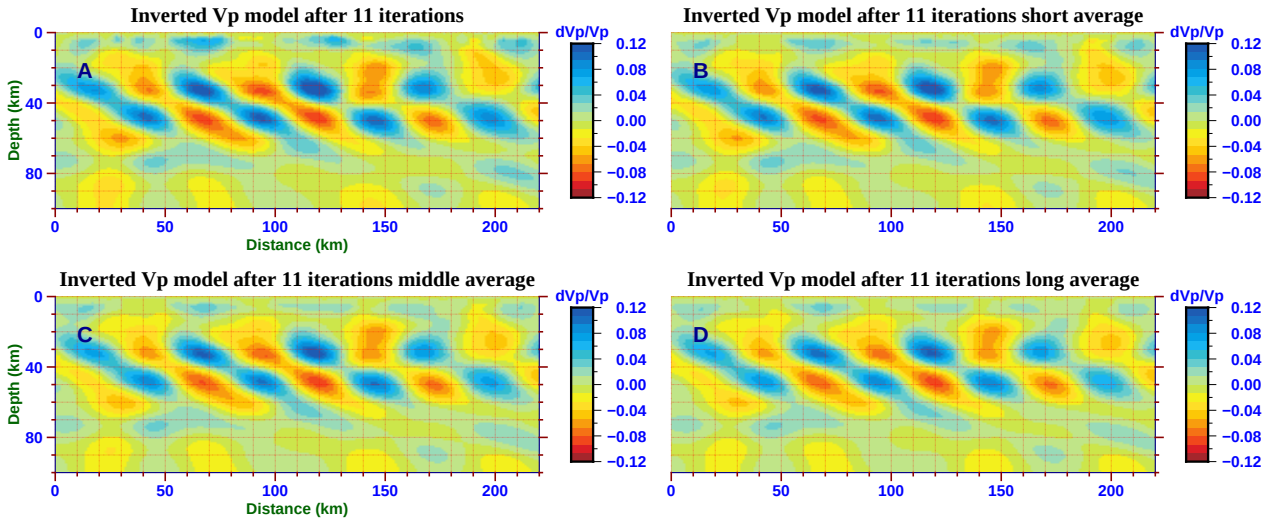
So far, except for considering the inversion strategies themselves, the main factors that may control the resolution in the deep part of the Vs model are the length of the time window, data coverage, and the source wavelets. Adding new events to improve the data coverage could improve the results. However, due to the large amount of computations and storage required to add new events, this will have to wait for the development of a new generation of codes which are more efficient. These developments are under way. Therefore, we only investigated the effect of the time window and source wavelet on the waveform inversion.

We perform a sixth resolution test by extending the previously used time window to 40 s after the direct P arrival time.

6. We use the reciprocal of the L1 norm of the individual event kernel along the vertical section to balance their contributions before summing them. The square root of depth at each grid point is used as the preconditioner. The dataset is the same as in test 1.

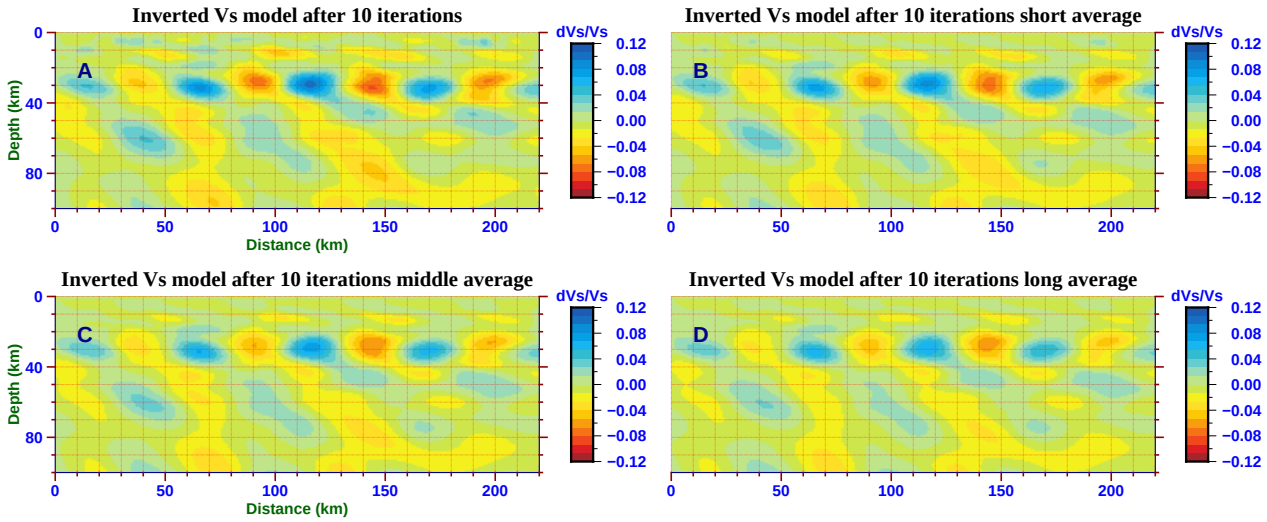


(a) Inverted Vp model for resolution test 3

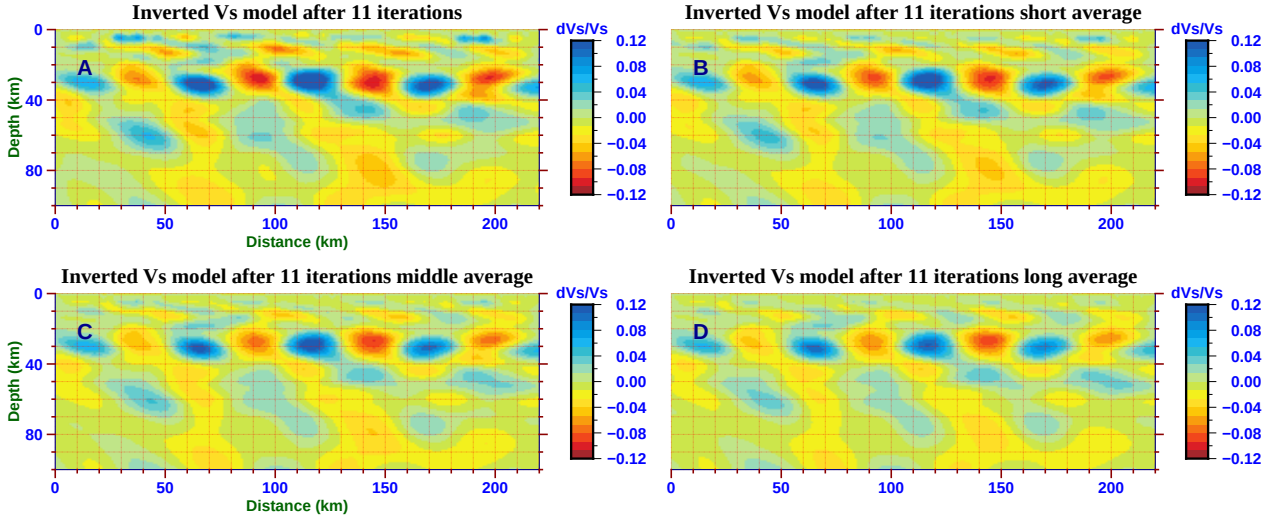


(b) Inverted Vp model for resolution test 4

Figure 6.21: Final Vp model for resolution test 3 (a) and 4 (b).

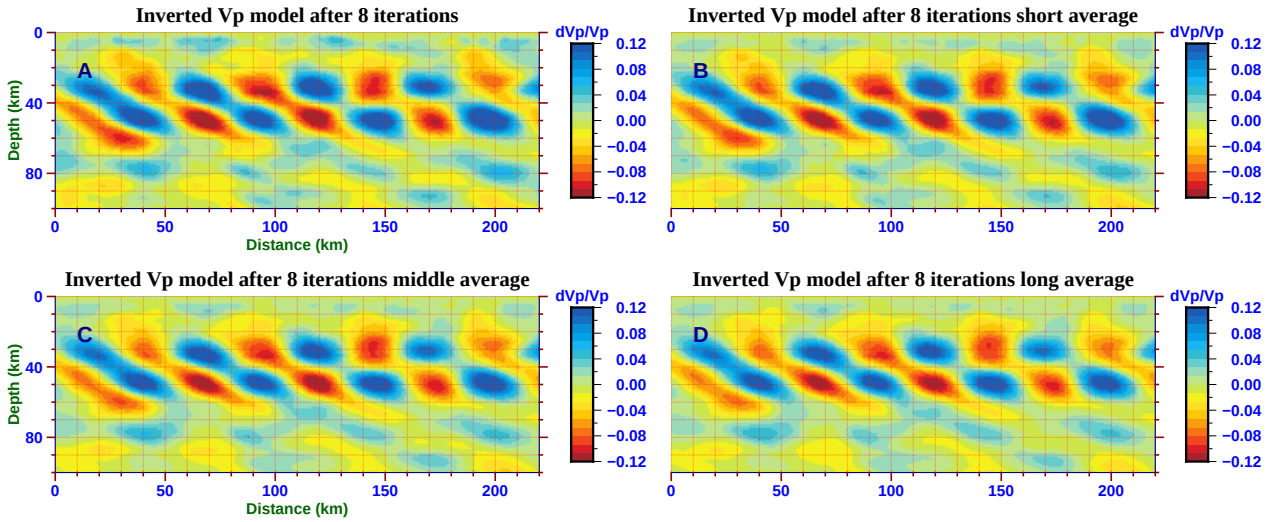


(a) Inverted Vs model for resolution test 3

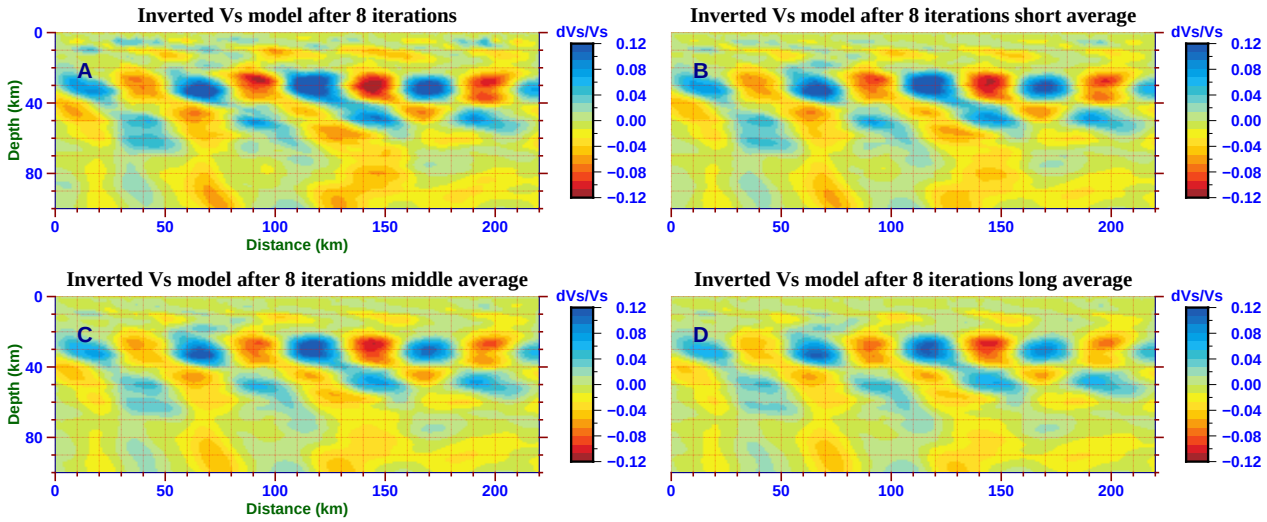


(b) Inverted Vs model for resolution test 4

Figure 6.22: Final Vs model for resolution test 3 (a) and 4 (b).



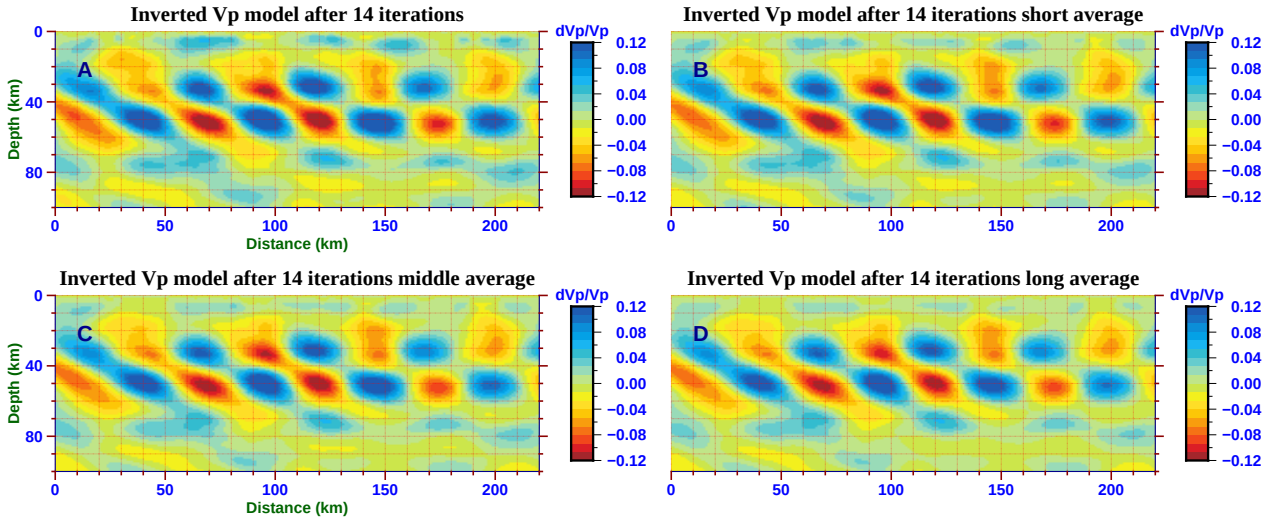
(a) Inverted Vp model for resolution test 5



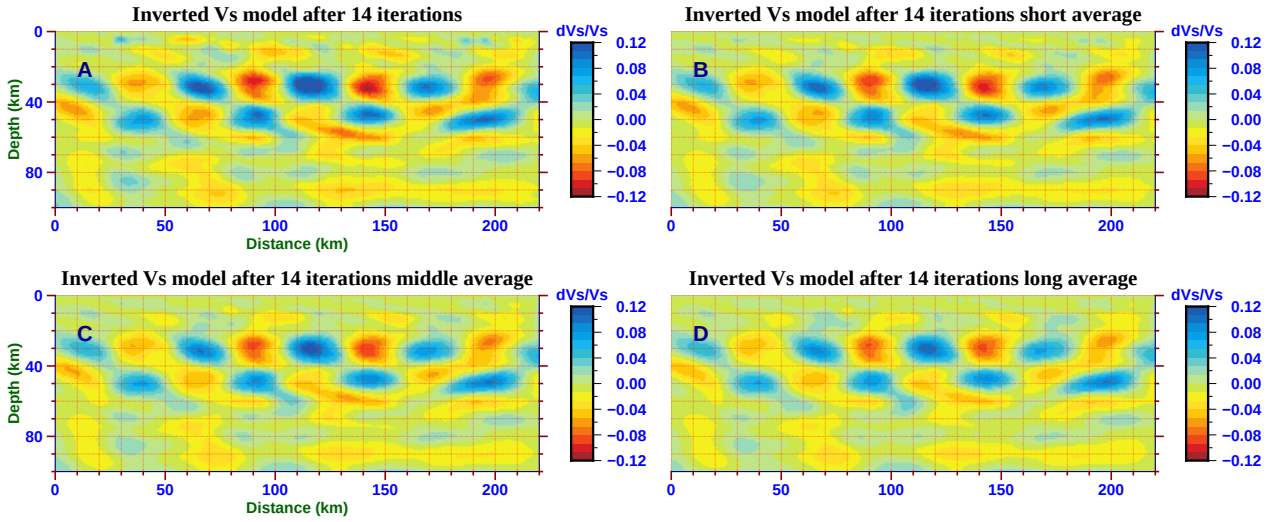
(b) Inverted Vs model for resolution test 5

Figure 6.23: Final Vp (a) and Vs (b) models for resolution test 5.

The inversion converged after 14 iterations. The final Vp and Vs models are shown in Figure 6.24. The relative misfit residual after inversion is 0.23. Finally, we successfully retrieve the target anomalies in the deep part of the Vs model, with moderate amplitude recovery. The oblique smearing effects in both Vp and Vs models are reduced significantly. The improvement of the tomographic models in this resolution test again demonstrates the significance of utilizing as much information as possible in the inversion. The later arriving coda waves contain critical information to constrain the deep structures of the velocity models, which should be carefully considered. The good resolution obtained in test 6 also suggests that the data coverage for the central transect is reasonable.



(a) Inverted Vp model for resolution test 6



(b) Inverted Vs model for resolution test 6

Figure 6.24: Final Vp (a) and Vs (b) models for resolution test 6. We extend the previously used time window to 40 s after the direct P arrival time.

Table 6.25 shows the vertical and radial component waveform misfits for each event before FWI and the relative misfit residuals after FWI for resolution test 6. For the vertical components, the relative misfit residuals for all the events are comparable. For the radial components, the 2012 Apr 02 has the largest relative misfit residuals.

waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
26/02/2012	113.604	0.158489	46.8267	0.255203
02/04/2012	114.364	0.207339	19.7586	0.570457
24/05/2012	57.177	0.156949	44.187	0.169216
10/08/2012	39.1141	0.210813	10.3787	0.300333
11/08/2012	53.3053	0.224507	91.1291	0.368197

Table 6.25: Waveform misfits and relative misfit residuals for each event in resolution test 6.

The vertical sections of the five event kernels for the waveform misfit defined in a longer time window, which are computed in the initial smooth 1D model, are shown in Figure 6.26. Compared to the event kernel for the inversion using convolved seismograms (Figure 6.20), the Vs kernels for this test show a better consistency with the of input checkerboard patterns.

Finally, we discuss the impact of the source wavelet on the waveform inversion. The synthetic seismograms $\tilde{u}(\mathbf{x_r}, t)$ used in the previous resolution tests are point source Green's functions $G(\mathbf{x}_r, t)$ convolved with the estimated wavelet function $\bar{W}(t)$. The spectrum of synthetic seismogram $\tilde{U}(\mathbf{x_r}, \omega)$, defined by their Fourier transformation $\mathcal{F}\{\tilde{u}(\mathbf{x_r}, t)\}$, is the product of the spectrum of the point source Green's function $G(\mathbf{x_r}, \omega)$ and the spectrum of the source wavelet $\bar{W}(\omega)$:

$$\tilde{U}(\mathbf{x_r}, \omega) = G(\mathbf{x_r}, \omega) \cdot \bar{W}(\omega) \tag{6.1}$$

The source wavelet acts on the point source Green's function as a band-pass filter. Sometimes the useful frequency content contained in $G(\mathbf{x_r}, \omega)$ is limited by the frequency content of the source wavelet $\bar{W}(\omega)$. We use the amplitude spectrum of the vertical component synthetic data (generated from checkerboard model) for three events: April 02, May 24 and Aug 11, which have big differences of azimuths between each other, as illustrations.

The left panels of Figures 6.27 to 6.29 show the normalized amplitude spectra of source wavelets $|\bar{W}(\omega)|$ (blue shaded area) and the normalized amplitude spectra of the point source Green's function $|G(\mathbf{x_r}, \omega)|$ (red shaded area) for the transect station on the French side. The right panels show the normalized amplitude spectra of synthetic seismograms $|\tilde{U}(\mathbf{x_r}, \omega)|$ (violet shaded area). The amplitude spectra of source wavelets show larger value below 0.3 Hz for these three events. The distribution of the source wavelet amplitude spectra depends on the specific event. In any case, the spectra are not flat. Since the waveform inversion is a nonlinear problem, the least square waveform misfit function depends on the frequency content of seismograms in a nonlinear way. As a result, the inversion using the convolved synthetics $\tilde{u}(\mathbf{x_r}, t)$ is expected

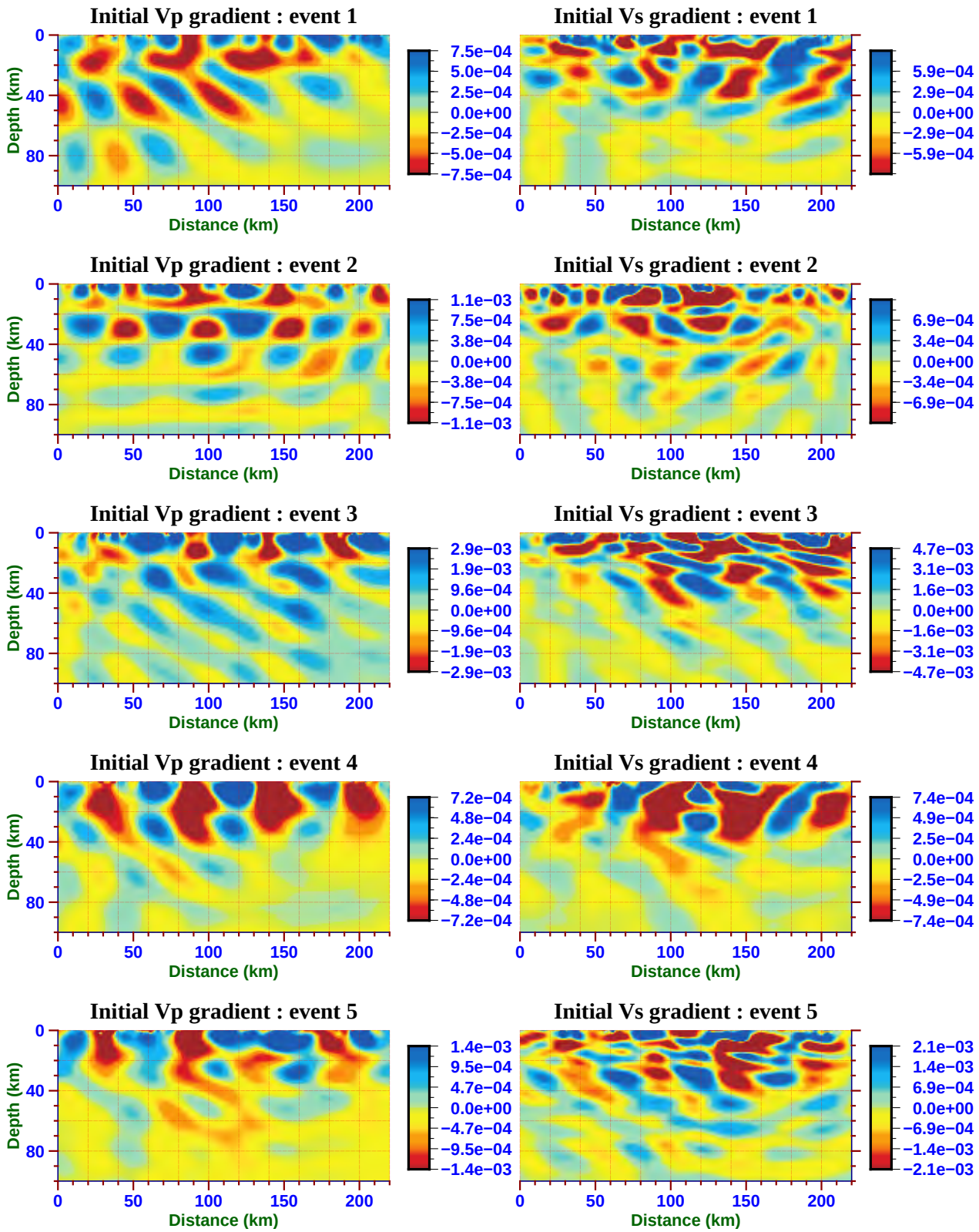


Figure 6.26: Vp (left) and Vs (right) event kernels (gradients) for each event computed in the initial smooth 1D model for resolution test 6.

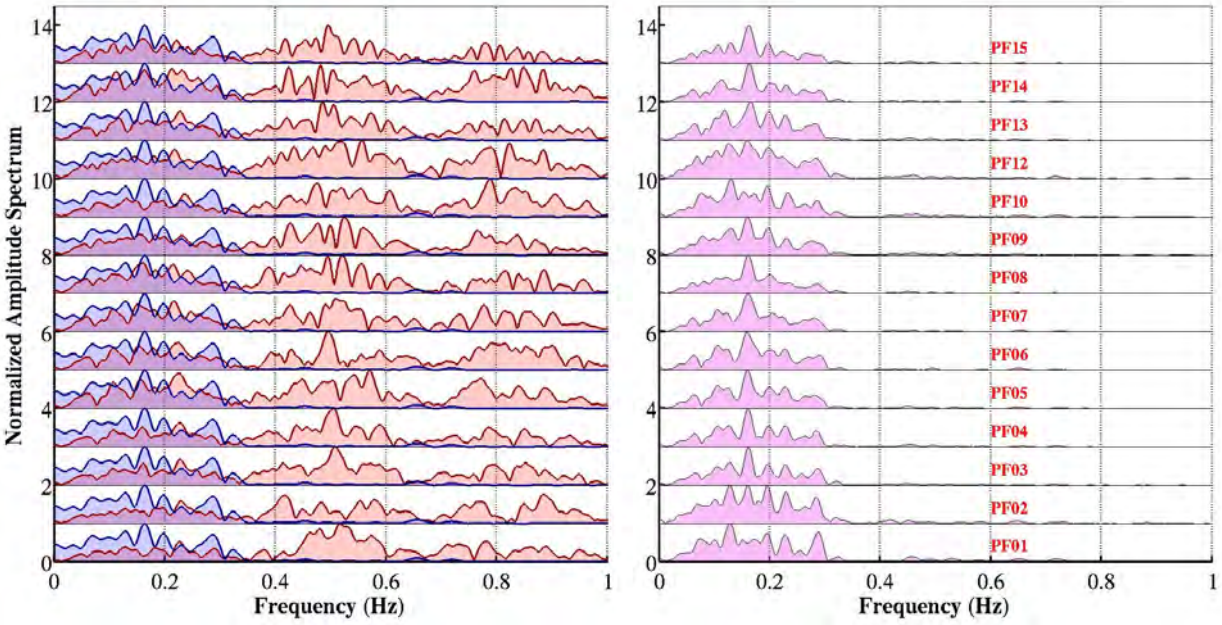


Figure 6.27: Amplitude spectra of the vertical component of the 2012 April 02 event. The left panel shows the normalized amplitude spectrum of the source wavelet for the event $|\bar{W}(\omega)|$ (blue shaded area) and the normalized amplitude spectra of point source Green's function $|G(\mathbf{x_r}, \omega)|$ (red shaded area) for the transect stations on the French side. The right panel shows the normalized amplitude spectra of the corresponding synthetic seismograms $|\tilde{U}(\mathbf{x_r}, \omega)|$(violet shaded area).

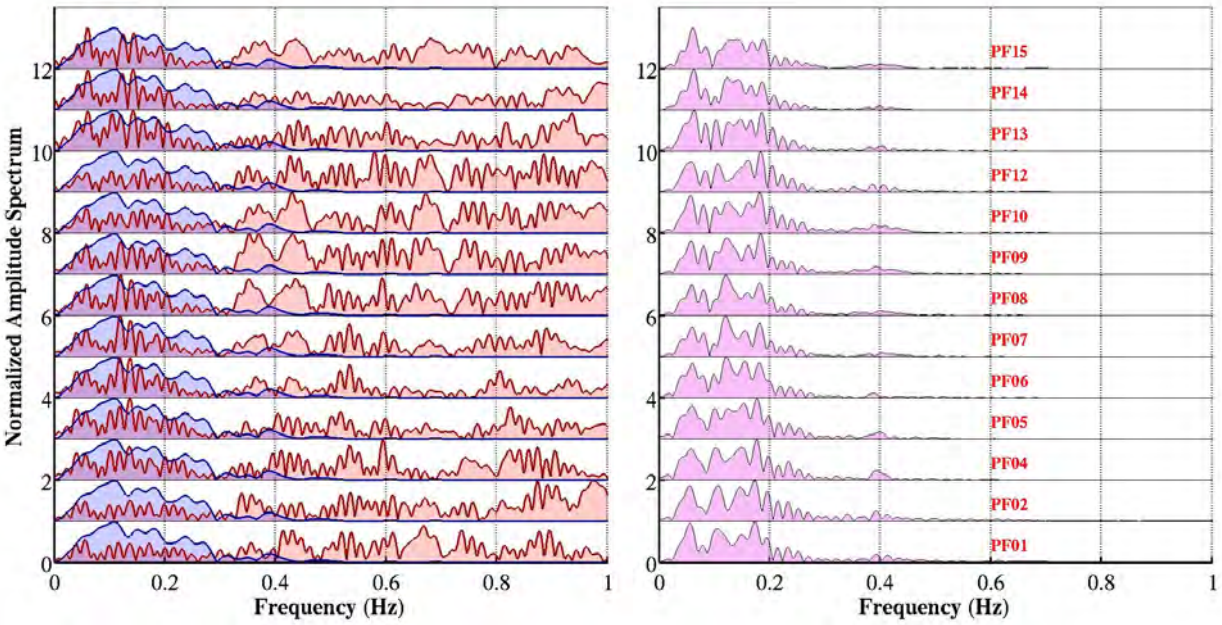


Figure 6.28: Same as Figure 6.27 but for the 2012 May 24 event.

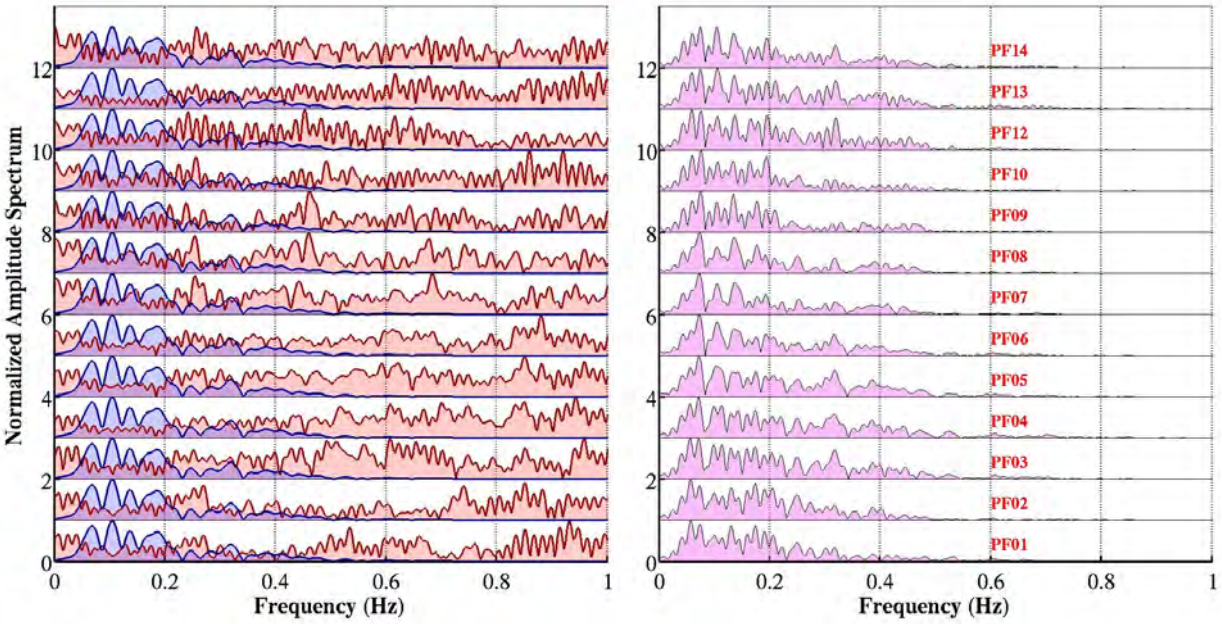


Figure 6.29: Same as Figure 6.27 but for the 2012 August 11 event.

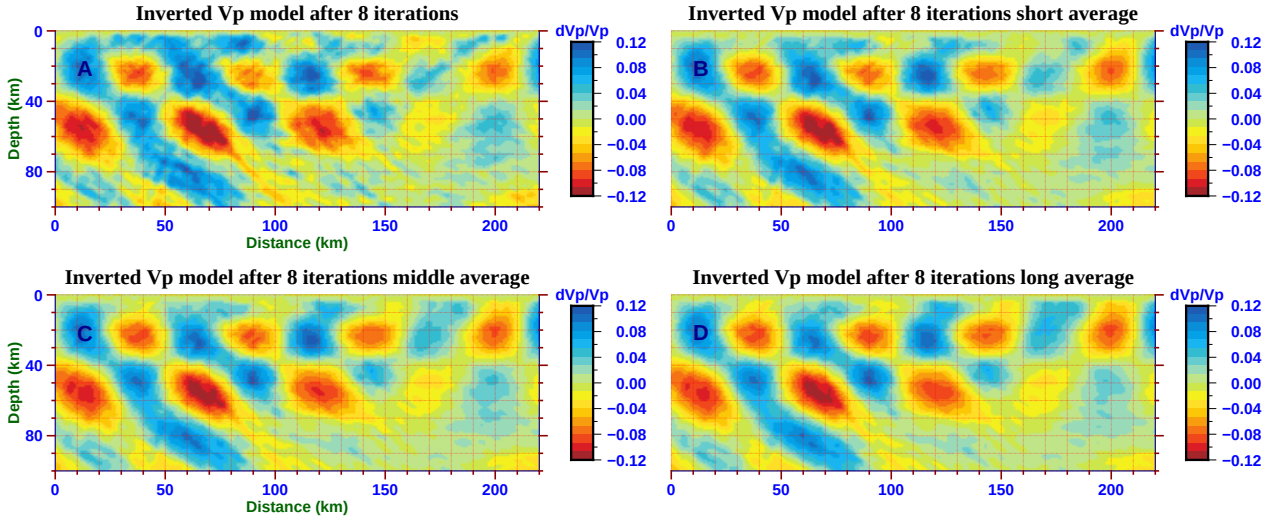
to obtain different results from the inversion using the point source Green's functions.

We perform a seventh resolution test by using the point source Green's function directly. The synthetic data are simply the point source Green's function propagating through the input checkerboard model.

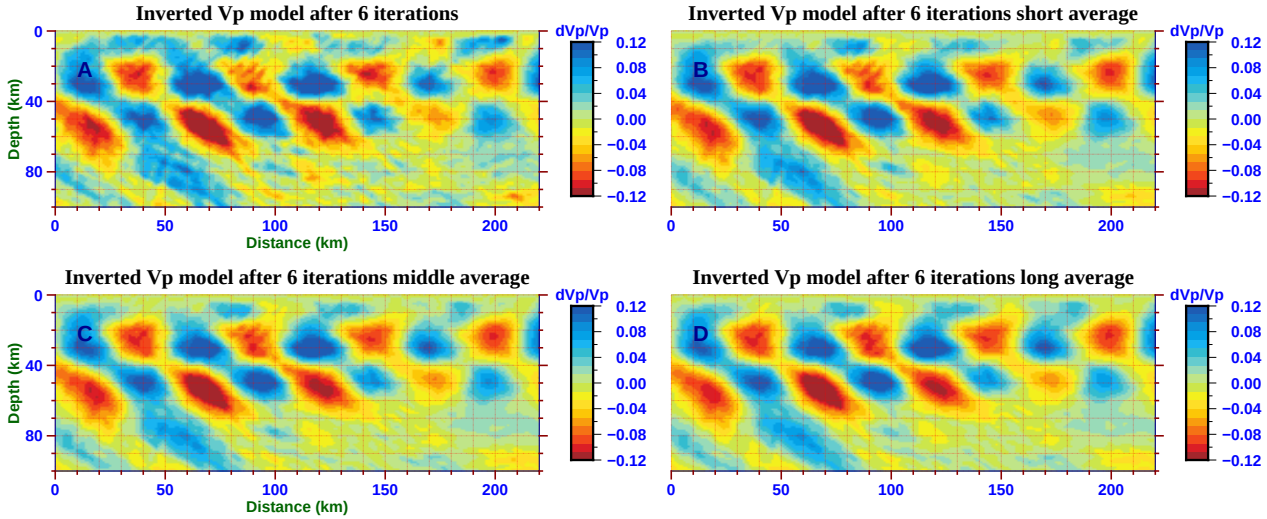
7. This test is conducted with a hierarchical algorithm starting with data low-pass filtered at 0.1 Hz and then gradually increasing the upper corner frequency of the filter to 0.2 Hz. Each event kernel is weighted by the reciprocal of its L1 norm before summing them. The time window, dataset, and the preconditioner are the same as in test 6.

After 8 and 6 iterations, the two-stage hierarchical inversion converged. The relative misfit residuals for the initial 10 s inversion and subsequent 5 s inversion are 0.12 and 0.3, respectively. Figure 6.30 and 6.31 show the final Vp and Vs models for this resolution test. We successfully retrieve the target anomalies in the deep part of the Vs model, with very good amplitude recovery. The shapes of the deep Vs anomalies are recovered better than resolution test 6. The artifacts in the northern part of the final Vs model between 60 and 80 km depth, may result from the asymmetric data coverage of the five events. The target anomalies in the Vp model are gradually recovered from the inversion starting at 10 s and then going down to 5 s. Surprisingly, for both shallow and deep parts of the Vs model are well recovered after the first inversion at 10 s. Increasing the frequency content of waveform inversion to 0.2 Hz does not significantly improve the Vs model. These results suggest that convolving with the source wavelet may suppress the low frequency content of the point source Green's function, which is important to recover the Vs model.

As a conclusion, the key issues to improve the resolution and quality of tomographic models

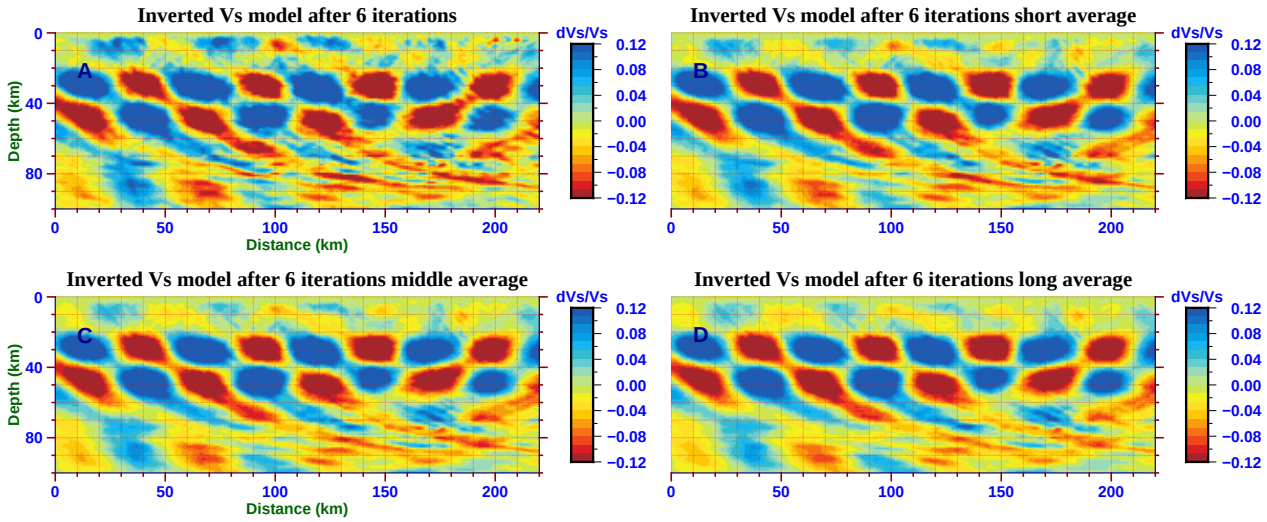


(a) Inverted Vp model for inversion filtered at 0.1 Hz

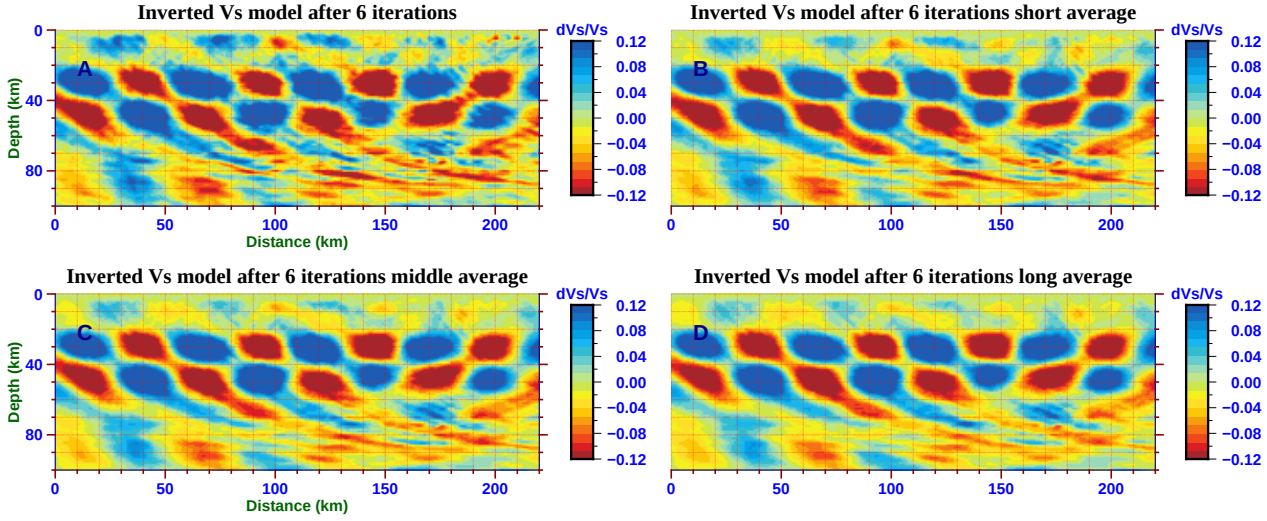


(b) Inverted Vp model for subsequent inversion filtered at 0.2 Hz

Figure 6.30: Final Vp model obtained in resolution test 7, (a) for the initial inversion filtered at 0.1 Hz, (b) and for the subsequent inversion filtered at 0.2 Hz.



(a) Inverted Vs model for inversion filtered at 0.1 Hz



(b) Inverted Vs model for subsequent inversion filtered at 0.2 Hz

Figure 6.31: Same as Figure 6.30 but for the Vs model.

are the length of the time window as well as the source wavelet used in teleseismic wavefield modeling and inversion. These resolution tests lead us to reconsider our waveform inversion strategy on real dataset.

6.3 FWI results and the preliminary interpretations

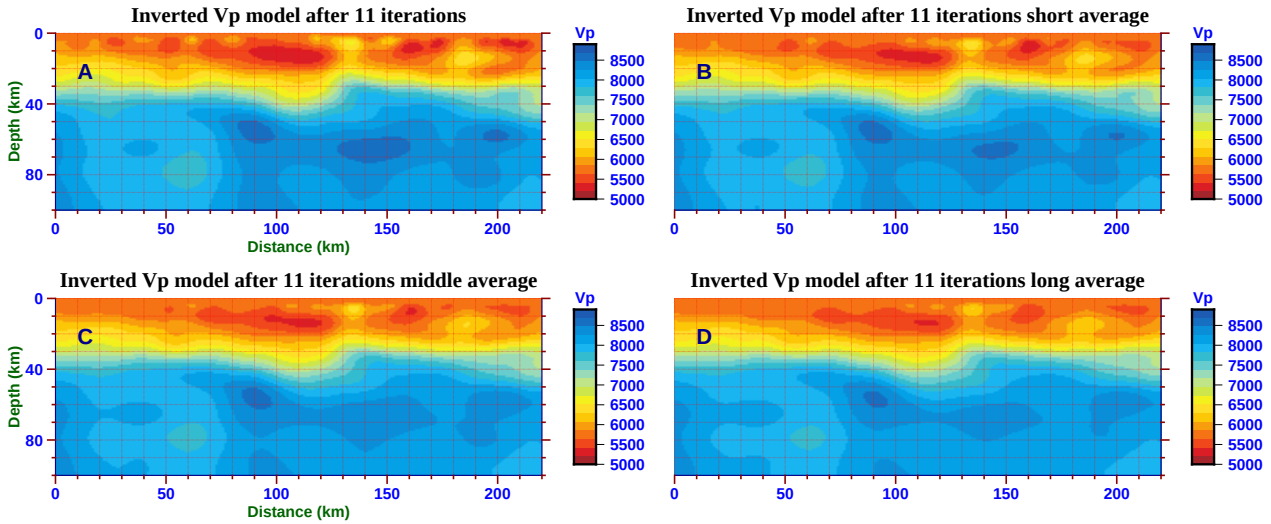
For the first application of FWI to the real data recorded by the central transect, we define a shorter time window that starts 15 s before the arrival time of the P wave and ends 30 s after, which is the same as in resolution tests 1 to 5 . This time window covers the long durations of the estimated source wavelets (of the order of around 40 s) and includes the contributions of P-to-S conversions along with some earlier multiples into the inversion. The vertical and radial component waveforms are filtered in the optimal frequency bands determined before. We perform the inversion starting from the same initial smooth 1D model as before. The reciprocals of the L1 norm of the event kernels along the vertical section are used to balance their contributions before summing them. We use the square root of depth at each grid point as the preconditioner for the gradient. Since the real data are affected by background noise, regularization is introduced by adding a penalty condition on the L_2 norm of the Laplacian of the model. The choice of the smoothing coefficient λ is based on trial-and-error. Convergence of FWI is reached after 10 to 15 iterations which required less than 5 hours on supercomputer 'Curie' with 2560 processors.

Figure 6.32 shows the final Vp and Vs models obtained with a moderate smoothing coefficient after 11 iterations. The relative misfit residual after inversion is 0.54. The inverted Vp model seems very smooth. The Moho interface seen in the Vp model dips gently from the south to the north. No sharp Moho step beneath the NPF is observed in the central Pyrenees. In contrast, the Vs model shows more interesting features. A low velocity anomaly is dipping down to almost 100 km depth.

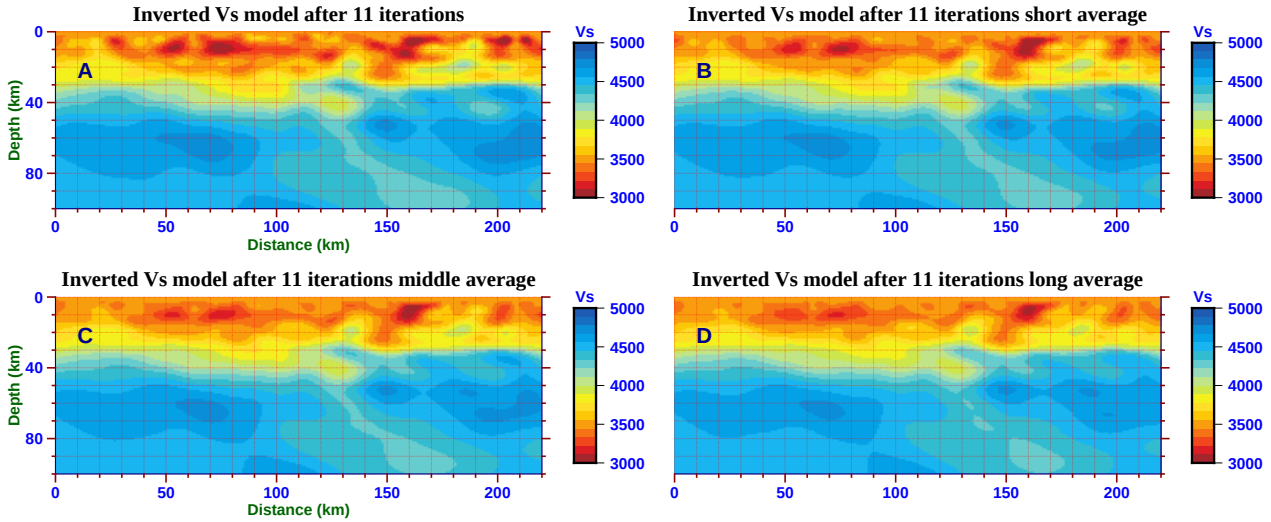
Table 6.33 shows the vertical and radial component waveform misfits for each event before FWI and the relative misfit residuals after FWI with a shorter time window on the real dataset. For the vertical components, the relative misfit residuals for all the events are comparable. For the radial components, the 2012 Apr 02 and May 24 event have larger relative misfit residuals.

We perform a new FWI with a longer time window that starts 15 s before the arrival time of the P wave and ends 40 s after. The inversion strategy, dataset, filters, and the preconditioner are the same as in the previous FWI. The inversion with a moderate smoothing coefficient converged after 12 iterations. The final Vp and Vs models are shown in Figure 6.34. The relative misfit residual after inversion is 0.48. Table 6.35 shows the vertical and radial component waveform misfits for each event before and after FWI with a 55 s long time window on the real dataset. The relative misfit residuals for both components of all the events are comparable.

The Vp model obtained by this FWI now shows a better consistency with the Vs model. The subducting Iberian crust is observed beneath the European Plate in both Vp and Vs models. Compared to the Vs model obtained by FWI with a shorter time window, the low velocity anomaly dipping at deep depth has been interrupted in the new Vs model. The low velocity



(a) Inverted Vp model for middle smooth

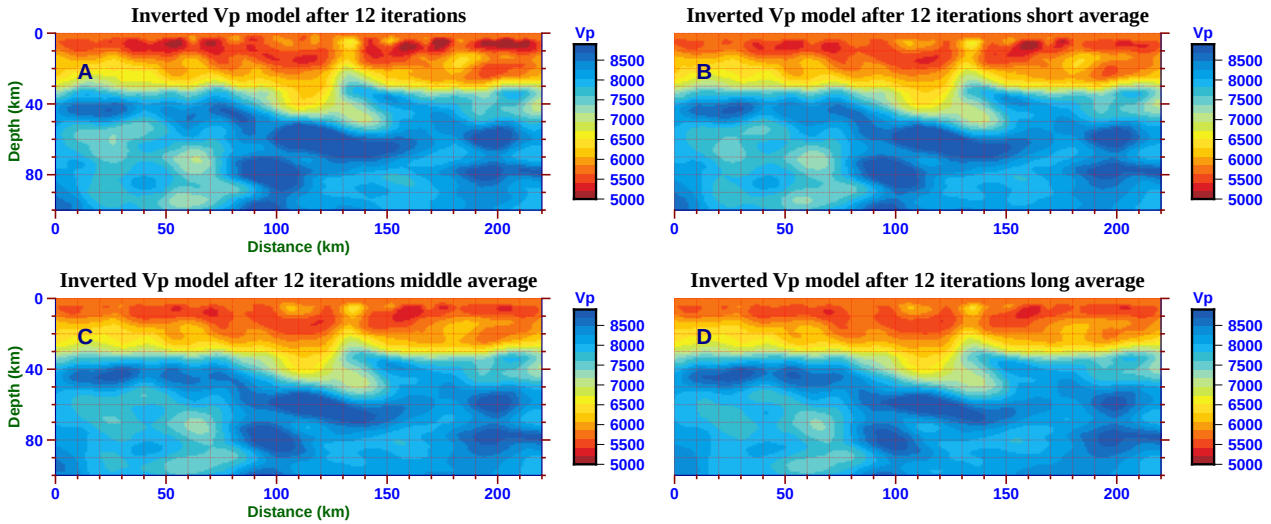


(b) Inverted Vs model for middle smooth

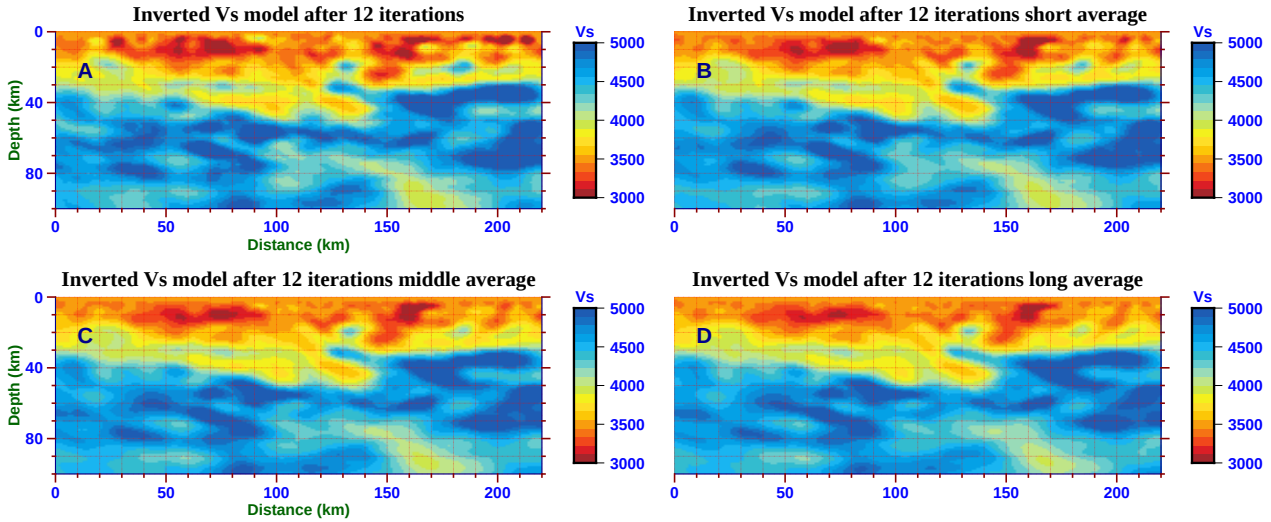
Figure 6.32: Final Vs (a) and Vp (b) models obtained by FWI using a shorter time window (45 s) with a moderate smoothing coefficient after 11 iterations.

waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
26/02/2012	193.36	0.543649	164.699	0.46982
02/04/2012	161.827	0.467627	21.2923	0.787622
24/05/2012	310.139	0.572512	171.812	0.706581
10/08/2012	109.317	0.548226	28.1915	0.563476
11/08/2012	130.958	0.495164	125.125	0.416389

Table 6.33: Waveform misfits and relative misfit residuals for each event for the inversion on the real dataset with a shorter time window (45 s).



(a) Inverted Vp model for middle smooth



(b) Inverted Vs model for middle smooth

Figure 6.34: Final Vs (a) and Vp (b) models obtained by FWI using a longer time window (55 s) with a moderate smoothing coefficient after 12 iterations.

waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
26/02/2012	303.162	0.488617	214.704	0.36729
02/04/2012	155.388	0.342625	32.6136	0.470042
24/05/2012	355.222	0.585123	197.88	0.597771
10/08/2012	145.543	0.477443	33.5843	0.520769
11/08/2012	164.988	0.456723	106.529	0.382749

Table 6.35: Waveform misfits and relative misfit residuals for each event for the inversion on the real dataset with a longer time window (55 s).

anomaly in the northern part of the final Vs model, between 60 and 100 km depth, is probably due to the asymmetric data coverage in this part.

We compare the initial event kernels for both inversions, shown in Figure 6.36 and 6.37. The major differences of which, are observed in the crust beneath the Axial Zone. The initial kernels mainly locate the strong material contrasts and the long wavelength heterogeneities which are related to the direct P waveform differences. Because of the nonlinearity of iterative FWI, the initial kernels can not simply explain the differences in the final inverted models. We notice that the amplitudes of the initial Vp and Vs kernels for the May 24 event are significantly larger than the other 4 events. Although the event kernels are balanced during the FWI, we still suspect that the coda waves of this event will put more constraints on the Vp model in the possible subduction zone beneath the axial zone because of its special azimuth. Figure 6.38 shows the Vp event kernels of this event for FWI using the longer time window (left panels), Vp kernels for FWI using the shorter time window (center panels), and the Vp models updated by FWI using the longer time window (right panels) after different iterations. The black contour line that delimits the crust-mantle boundary in the new final Vp model (panel C of Figure 6.34a) is also shown in each panel of Figure 6.38. We can see the negative Vp sensitivity appearing inside the subduction zone in the iterations 2 to 5 of FWI using the longer time window, which however is not observed in the Vp kernel for FWI using the shorter the time window. The subduction structure of the Vp model beneath the axial zone is gradually inverted after corresponding iterations.

Based on the discussions about the event kernel for the May 24 event above, we perform a FWI on the real data by using the shorter time window only for the May 24 event and using the longer time window for the other 4 events. The inversion strategy, dataset, filters, and the preconditioner keep the same as in the previous FWI. The inversion with a moderate smoothing coefficient converged after 8 iterations. The relative misfit residual after inversion is 0.66. The average vertical sections of the final Vp and Vs models with a 15 km horizontal averaging range are shown in the middle panels of Figure 6.39. As a comparison, we also show the average vertical sections of the final models obtained by FWI using the shorter time window (top panels) and those by FWI using the longer time window (bottom panels) in Figure 6.39. By comparing the previously inverted models, we find that the absence of the information contained in the coda waves from 30 s to 40 s after the P wave arrival time of the May 24 event leads to the poor resolution of the Vp model in the subduction zone as expected. Table 6.40 shows the vertical and radial component waveform misfits for each event before and after this FWI. The relative misfit residuals for both components of all the events are substantially lower than the relative misfit residuals after FWI using the longer time window. Therefore, we infer that the later arrivals from 30 s to 40 s after the P wave arrival time of the May 24 event contribute to the resolution for the subduction zone. When a small number of earthquakes are used in FWI, each event may provide constraints on different locations of seismic interfaces and

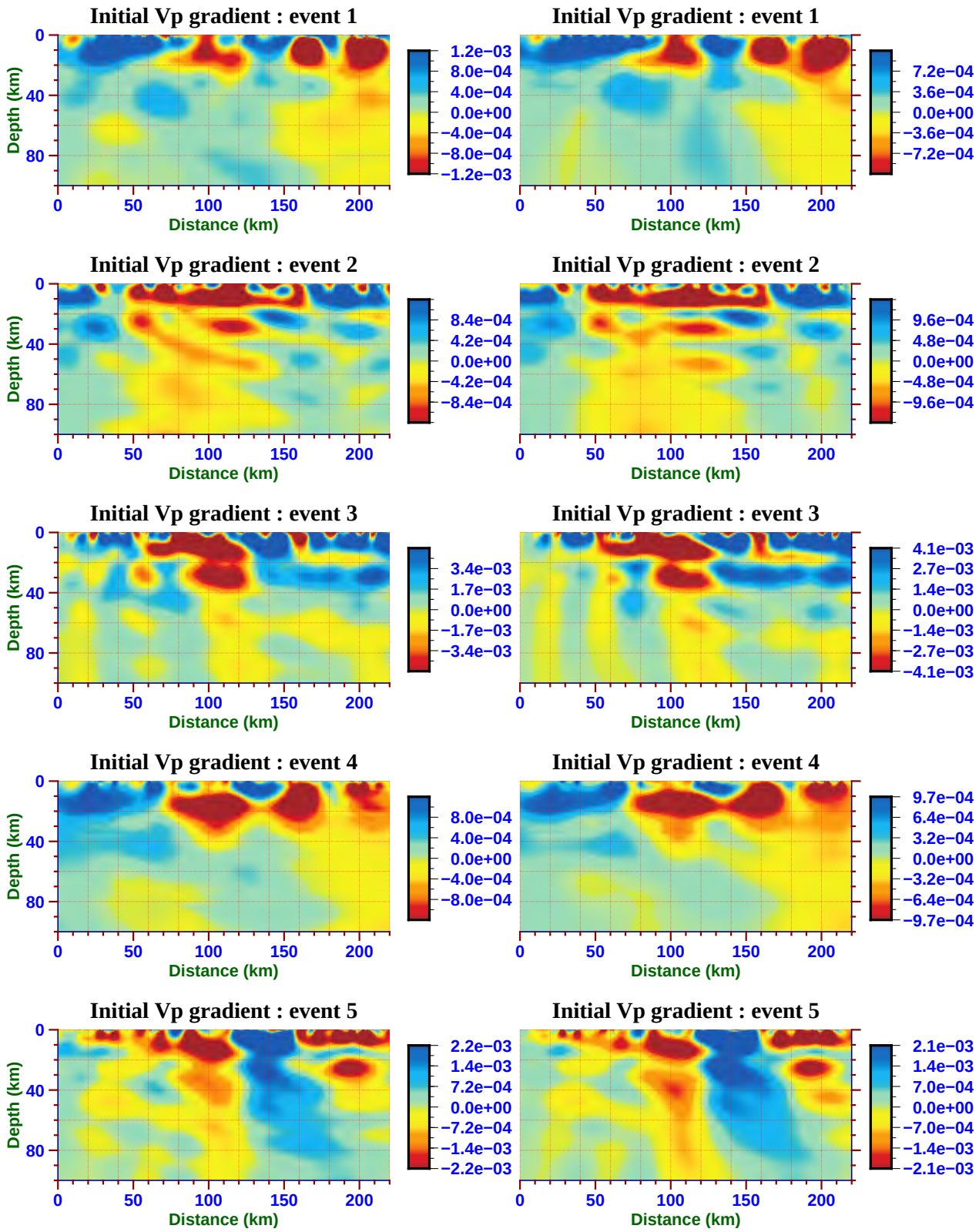


Figure 6.36: Vp kernels for each event computed in the initial smooth 1D model for FWI using a longer time window (left) and a shorter time window (right).

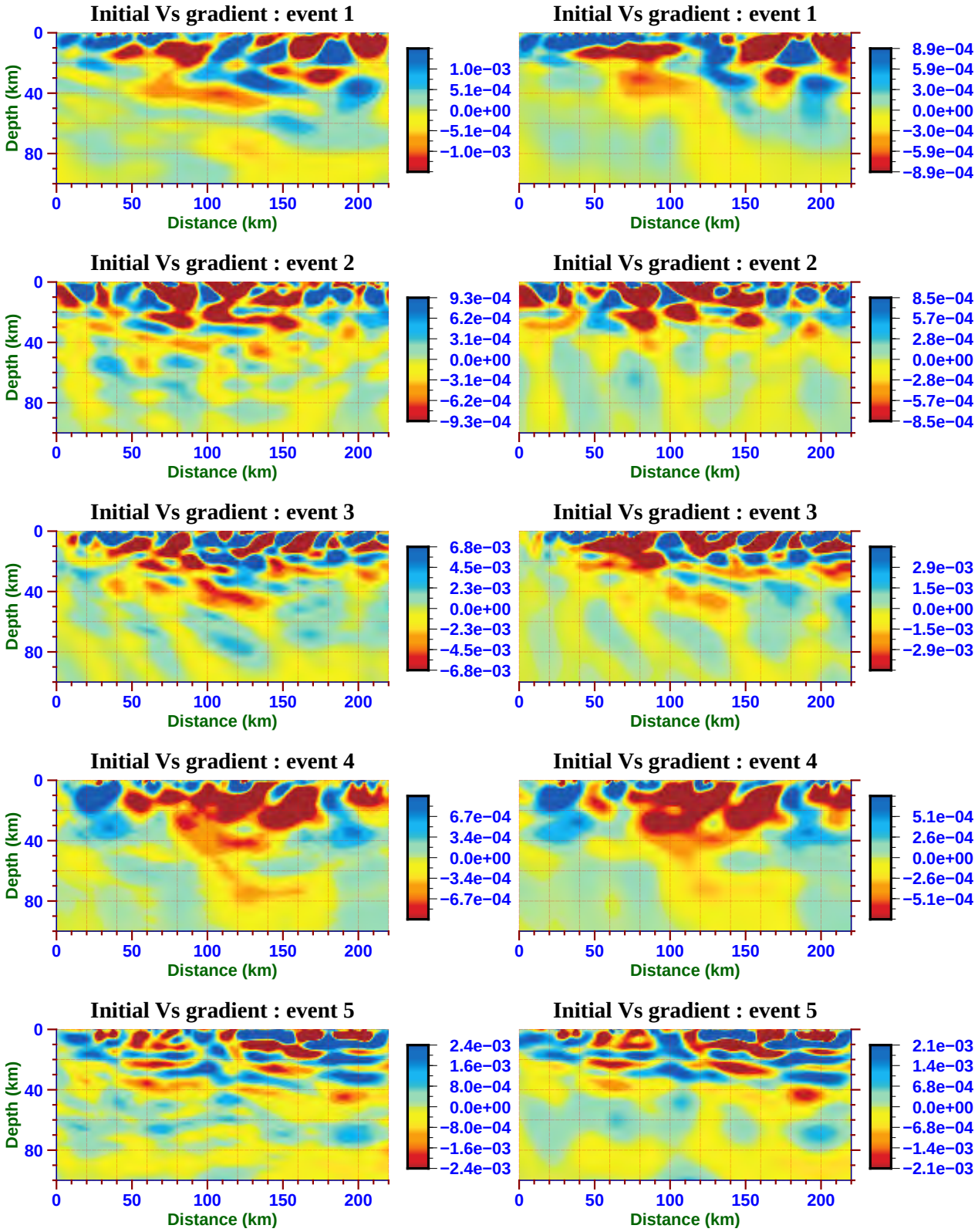


Figure 6.37: Vs kernels for each event computed in the initial smooth 1D model for FWI using a longer time window (left) and a shorter time window (right).

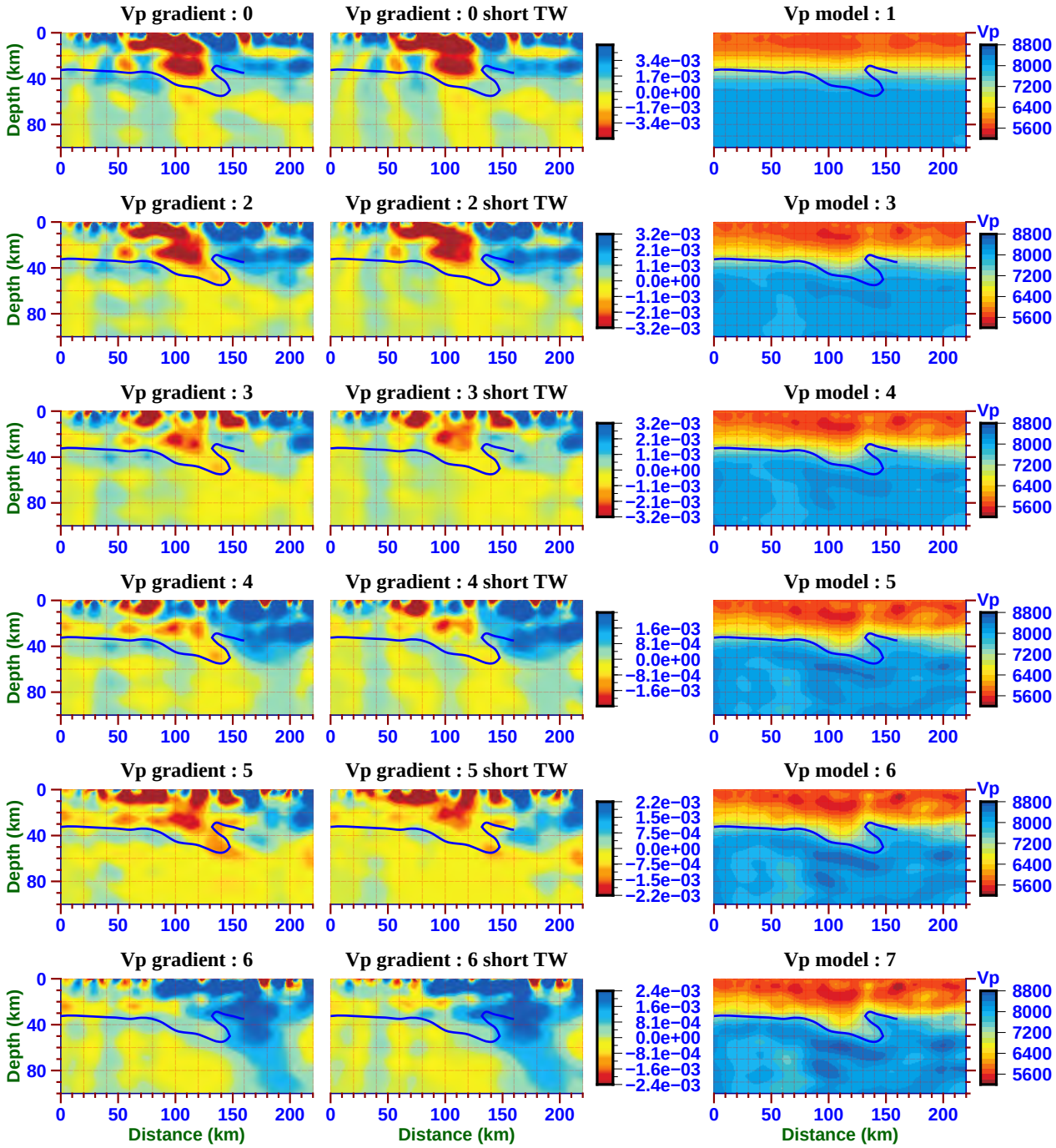


Figure 6.38: Vp event kernels of the 2012 May 24 event for FWI using the longer time window (left), Vp kernels for FWI using the shorter time window (center), and Vp models updated by FWI using the longer time window (right) after different iterations.

velocity structures. Hence, we should carefully deal with the data selection, length and starting of the time window and other data processing issues for every event.

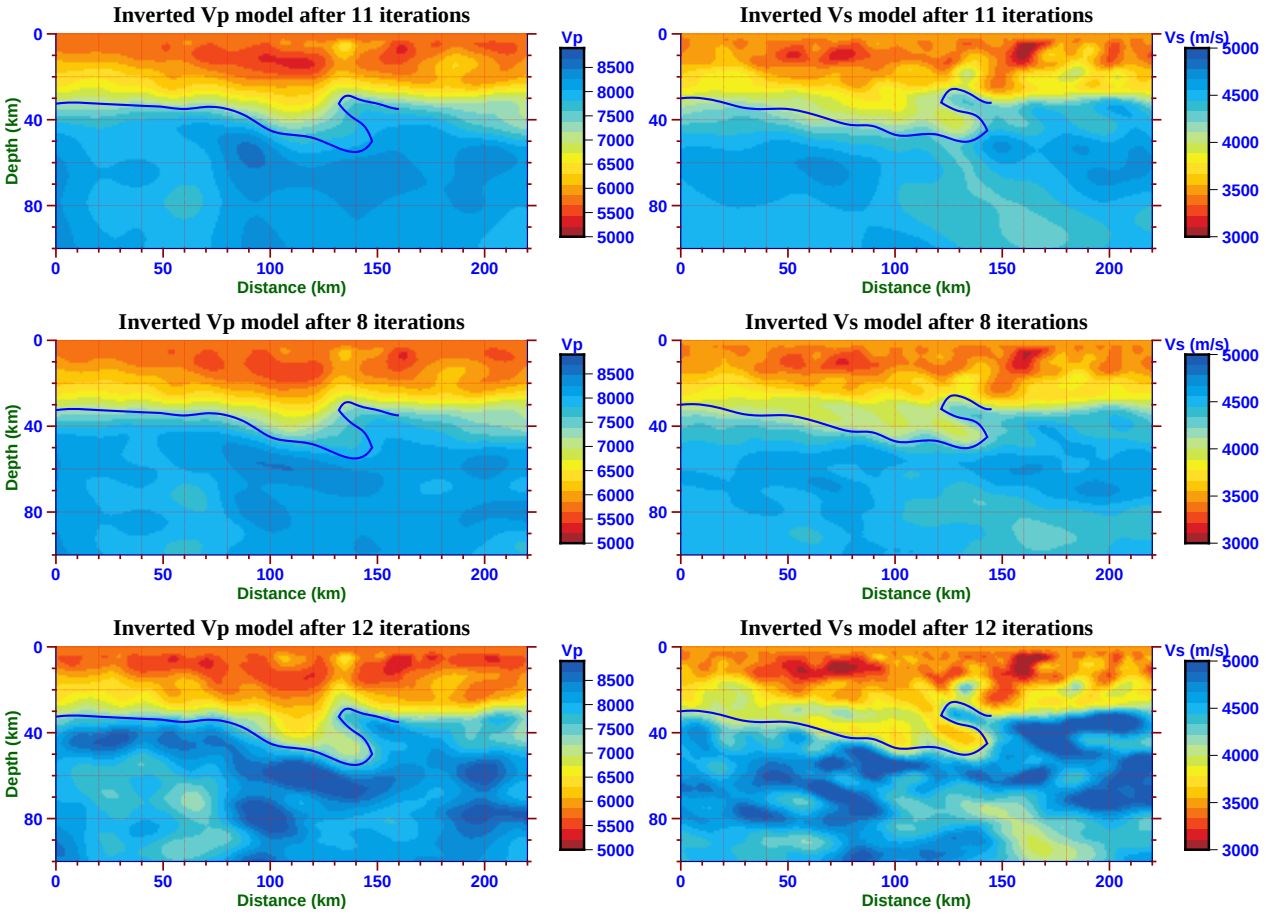


Figure 6.39: Top panel: Average vertical sections of the final Vp and Vs models obtained by FWI using the shorter time window, with a 15 km horizontal averaging range. Middle panel: Average vertical sections of the final models obtained by FWI using the shorter time window for the May 24 event and using the longer time window for the other 4 events. Bottom panel: Average vertical sections of the final models obtained by FWI using the longer time window. The black contour line delimits the crust-mantle boundary in the Vp and Vs models obtained by FWI using the longer time window.

At last following the idea introduced in resolution test 7, we deconvolve the estimated source wavelet from the real dataset to obtain a deconvolved dataset. After that, we use this deconvolved dataset for FWI to see whether we can improve the results of the inversion. We use the water-level deconvolution approach proposed by Clayton & Wiggins (1976), because the spectrum of our estimated source wavelet is relatively simple. The new deconvolved data waveform $\tilde{d}(\mathbf{x_r}, t)$ is given by:

$$\tilde{d}(\mathbf{x_r}, t) = \mathcal{F}^{-1}\left\{ \frac{\bar{W}^*(\omega) D(\mathbf{x_r}, \omega)}{\max\left(\bar{W}(\omega)\bar{W}^*(\omega), c \cdot \bar{W}^*_{max}\bar{W}_{max}\right)} \right\} \tag{6.2}$$

where $\mathcal{F}^{-1}\{\}$ denotes the inverse Fourier transformation, $D(\mathbf{x_r}, \omega)$ is the spectrum of the original trace $d(\mathbf{x_r}, t)$, $\bar{W}_{max}$ is the spectral value for which $|\bar{W}(\omega)|$ achieves its maximum

waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
26/02/2012	303.162	0.632741	214.704	0.569817
02/04/2012	155.388	0.438819	32.6136	0.652223
24/05/2012	310.282	0.702333	171.726	0.722865
10/08/2012	145.543	0.634763	33.5843	0.650036
11/08/2012	164.988	0.581025	106.529	0.537829

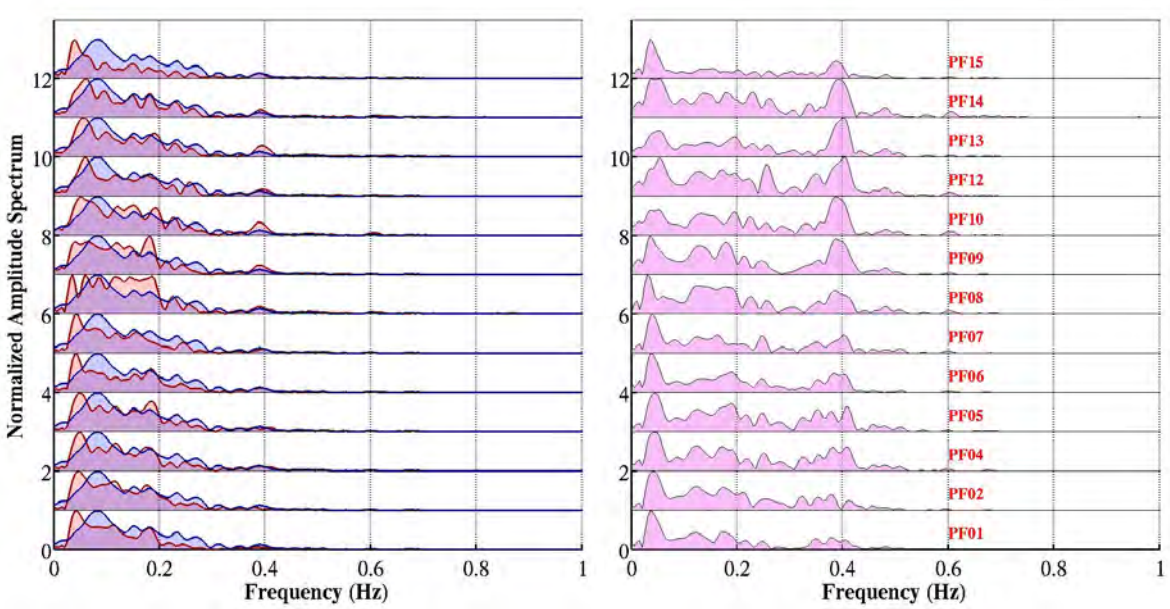
Table 6.40: Waveform misfits and relative misfit residuals for each event for the FWI on the real dataset by using the shorter time window only for the May 24 event and using the longer time window for the other 4 events.

value, and c is a user specified water-level parameter that regularizes the deconvolution. For our application, we use $c = 0.005$.

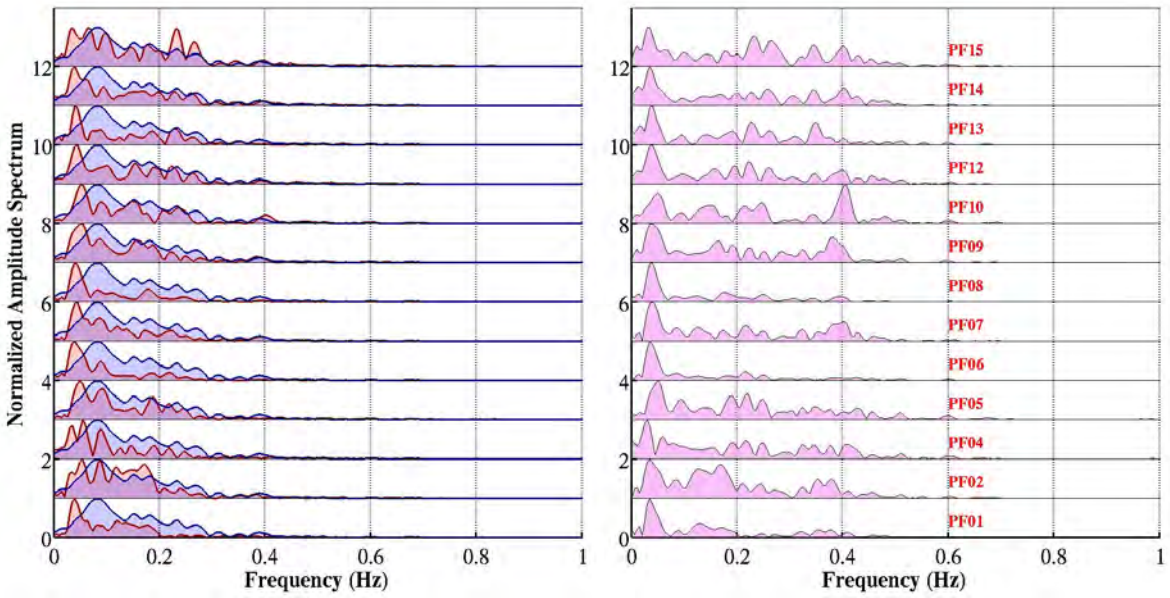
We show the normalized amplitude spectra of the May 24 source wavelet $|\bar{W}(\omega)|$ (blue shaded area) and data traces $|D(\mathbf{x_r}, \omega)|$ (red shaded area) for some stations on the transect in the left panels of Figure 6.41a (for vertical component) and 6.41b (for radial component). The right panels of Figure 6.41 show the normalized amplitude spectra of the deconvolved data traces $|\tilde{d}(\mathbf{x_r}, \omega)|$ (violet shaded area). Since the amplitude spectrum of source wavelet for this event is relatively low, between 0.15 and 0.3 Hz, the spectra of deconvolved traces are enhanced in this frequency range after deconvolution.

We perform a FWI on the deconvolved dataset with the longer time window. The inversion strategy, filters, and the preconditioner for this FWI is the same as before. After 11 iterations, the new FWI with a small smoothing coefficient converged. The relative misfit residual is 0.55, almost the same as in the previous FWI on the original dataset with the longer time window. Figure 6.42 shows the final Vp and Vs models by this FWI.

The Vp and Vs models now show a high consistency with the models obtained by FWI on the original dataset using the longer time window. The low velocity anomaly dipping at deep depth has disappeared in the new Vs model. However, some artifacts are still present, which may result from insufficient smoothing during the inversion. The deconvolution procedure is equivalent to filtering the waveforms. In our application, high frequencies were enhanced, and we were then able to retrieve more detailed tomographic images. Table 6.43 shows the vertical and radial component waveform misfits of each event before and after FWI performed on deconvolved dataset. The waveform misfit reductions are similar to the previous results. Although the results of this FWI do not show enough improvements compared to the results obtained by FWI on the original dataset using the longer time window, still, we mention an important advantage of FWI performed on deconvolved dataset over original dataset. In future inversions, we will include a source wavelet update procedure in iterative FWI, to better separate the wave propagation effects inside and outside the 3D regional domain. This iterative source wavelet estimation is equivalent to an earthquake source parameter inversion introduced in the current

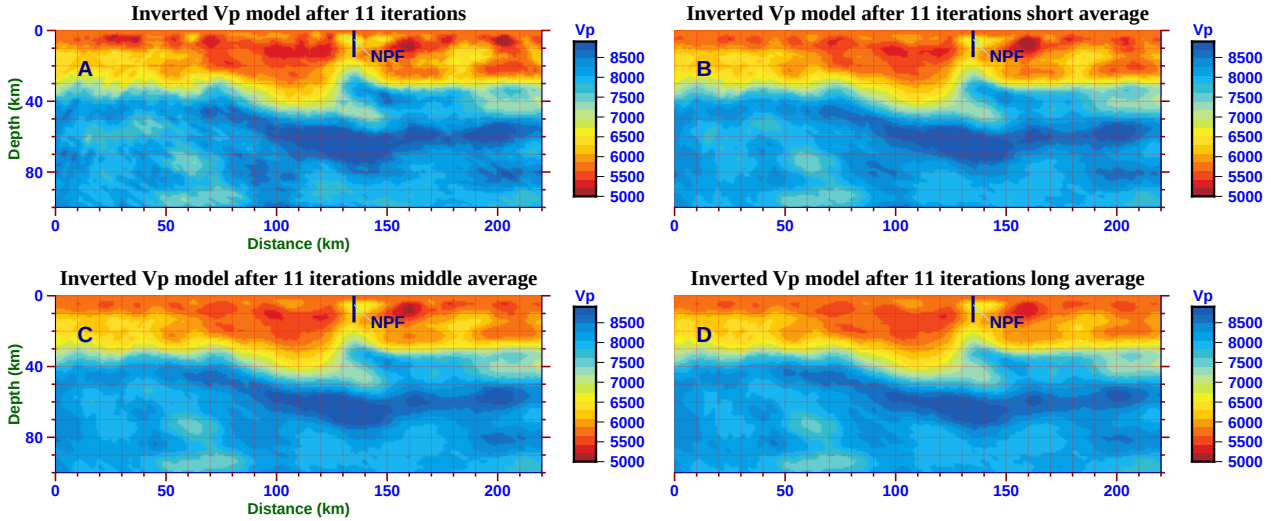


(a) Amplitude of Z components

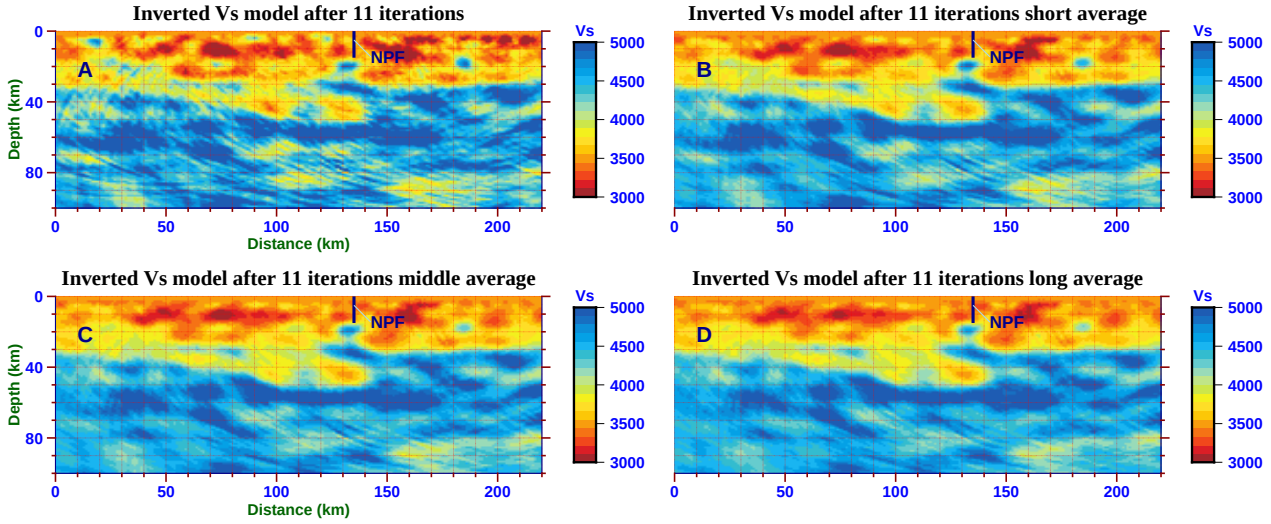


(b) Amplitude of R components

Figure 6.41: Amplitude spectra of the vertical component (a) and the radial component (b). Left panel: Normalized amplitude spectra of the May 24 source wavelet $|\bar{W}(\omega)|$ (blue shaded area) and data traces $|D(\mathbf{x_r}, \omega)|$ (red shaded area) of some transect station. Right panel: Normalized amplitude spectra of the deconvolved data traces $|\tilde{d}(\mathbf{x_r}, \omega)|$ (violet shaded area).



(a) Inverted Vp model for deconvolved dataset



(b) Inverted Vs model for deconvolved dataset

Figure 6.42: Final Vs (a) and Vp (b) models obtained by FWI on deconvolved dataset with a small smoothing coefficient.

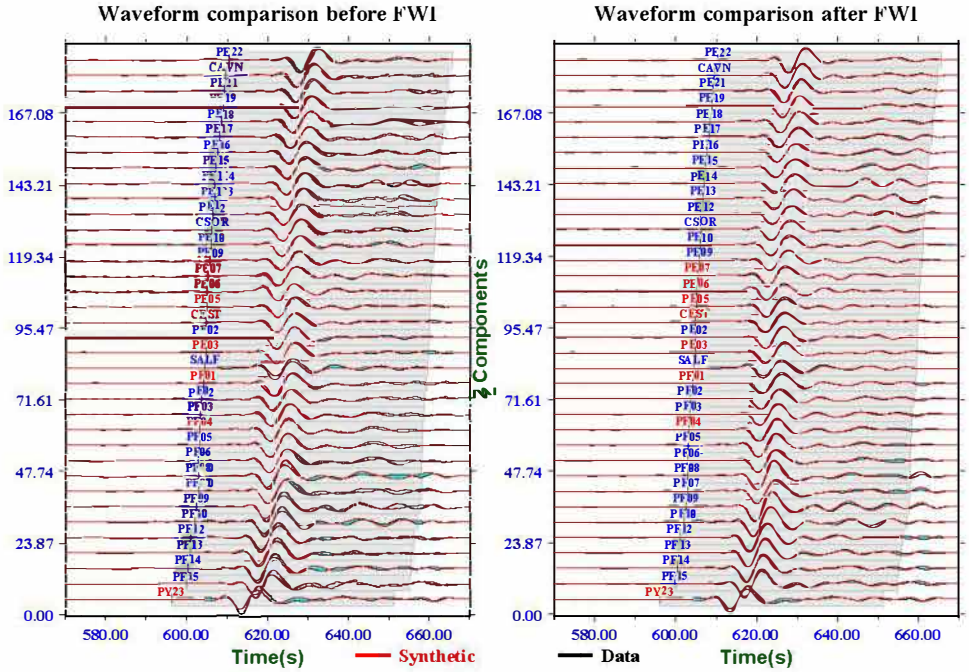
iteration, for the purpose of improving the FWI results in the regional domain (Tape et al., 2010; Fichtner et al., 2009). If we perform FWI on the original dataset, each new source wavelet estimated from the synthetic seismograms computed in the current regional 3D model should be convolved with huge boundary traction database for performing the subsequent iterations. This convolution operation requires a relatively high computational cost. In contrast, FWI performed on deconvolved dataset do not need this additional convolution operation, which makes the inversion more efficient.

waveform misfit	misfit (vertical)	relative misfit residual (vertical)	misfit (radial)	relative misfit residual (radial)
26/02/2012	287.656	0.544717	173.132	0.457699
02/04/2012	142.096	0.483475	30.3544	0.710331
24/05/2012	193.733	0.560605	157.479	0.658262
10/08/2012	152.303	0.523528	67.6322	0.577172
11/08/2012	48.1177	0.641139	23.9097	0.53548

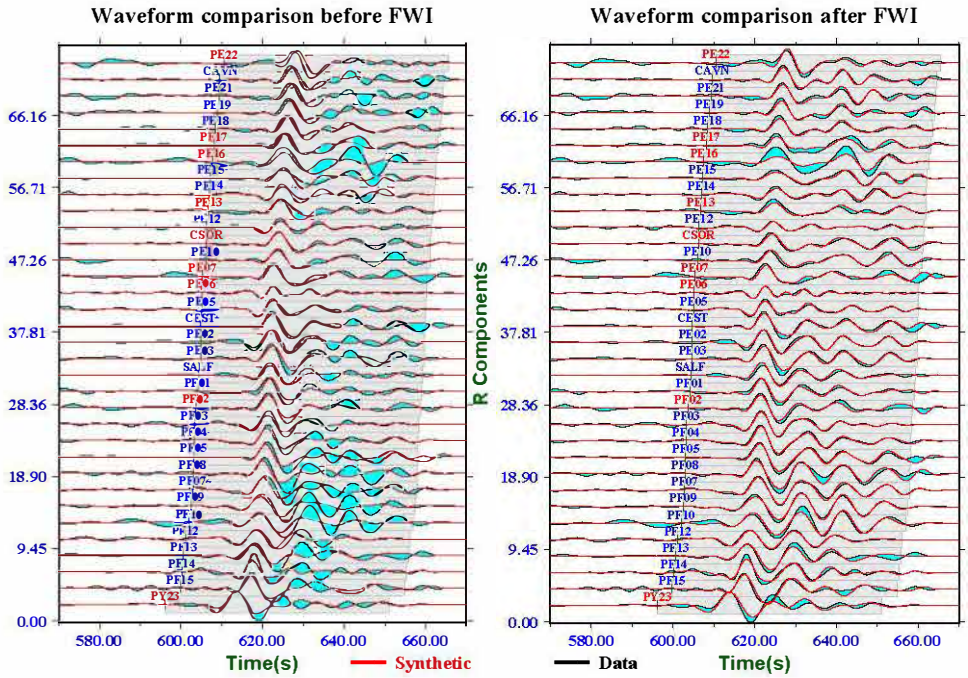
Table 6.43: Waveform misfits and relative misfit residuals for each event for the inversion of the deconvolved data.

Comparing the above four FWI results, we consider the final models obtained by FWI on the original dataset using the longer time window for geological interpretation since they are smoother and contain main geological structures. The comparisons between observed and synthetic seismograms along the transect computed in the initial 1D model and in the final 3D model obtained by FWI on the original dataset using the longer time window are shown in Figures 6.44 to 6.48. The vertical component waveform fits for all the events along the transect are good. Because several radial component waveform traces along the central transect were discarded after our optimal frequency range selection, the waveform fits for the 2012 Apr 02 and Aug 10 event are not very good. Traces for these two events have lower S/N ratio than the other three events. We need to develop a better data processing method to reduce their noise level.

Figure 6.49 shows the average vertical sections of our Vp and Vs models along the central transect in the middle and bottom panels. The horizontal averaging range is 15 km. The vertical sections of our Vp and Vs models show remarkable similarities, especially beneath the Axial Zone, but the interfaces of main geological units are more sharply resolved in the Vs model. The crust-mantle boundary in our Vp and Vs models presents a complex geometry, which is in good agreement with the images obtained by receiver function migration (Chevrot et al., 2015) shown in the top panel of Figure 6.49. We label the North Pyrenean Fault (NPF) with a thick blue bar in all panels. The Iberian and European Moho represented as black dashed lines and the top of the subducting Iberian crust represented as a grey dashed line are also shown in all panels. The positions of the two black dashed lines are in agreement with the

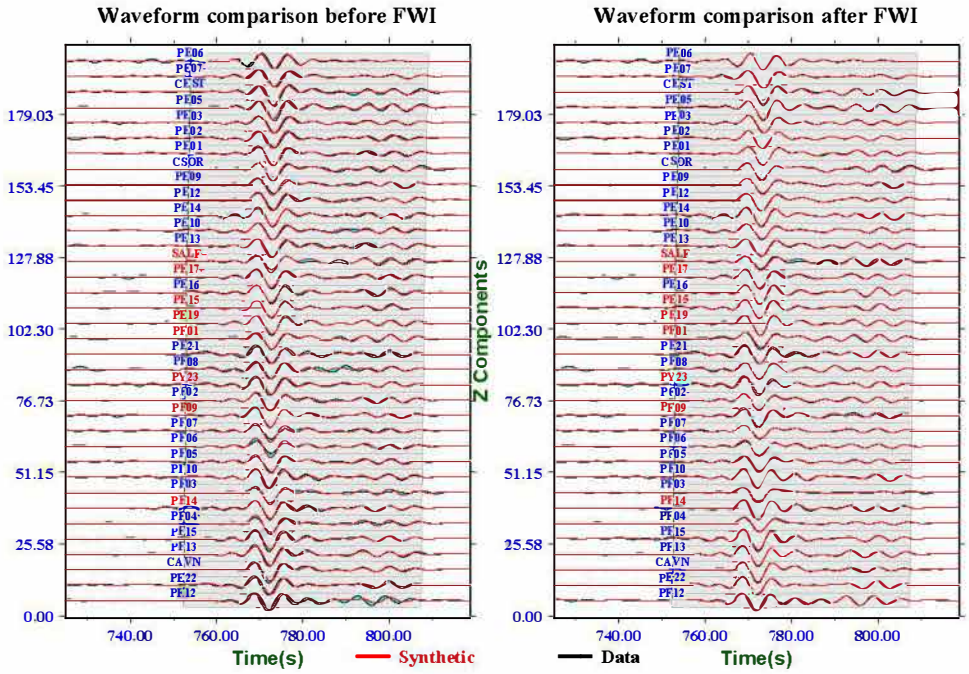


(a) vertical component waveforms along the transect.

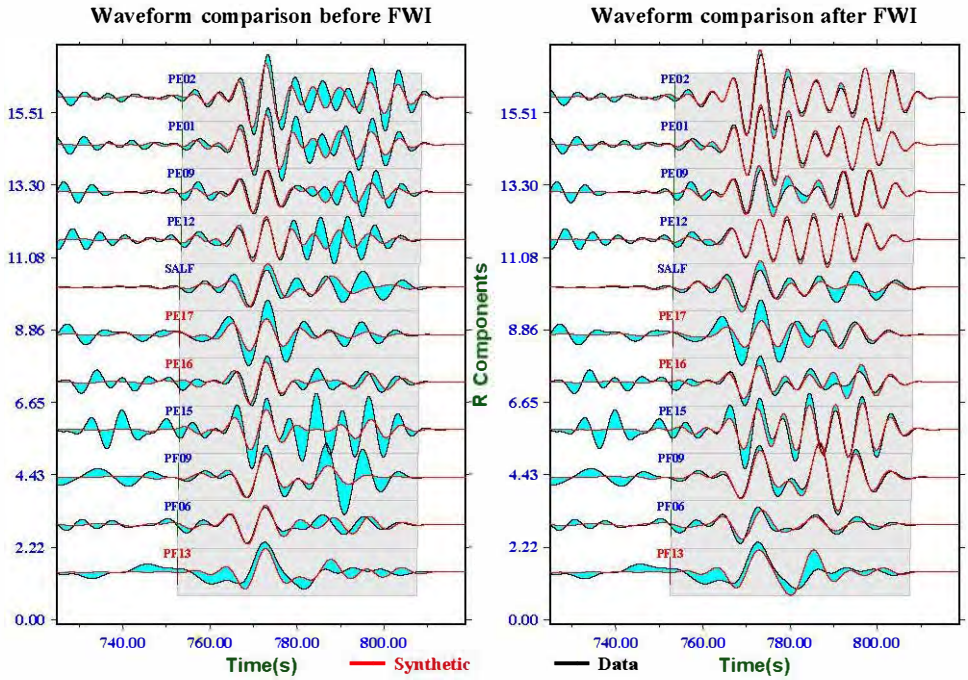


(b) radial component waveforms along the transect.

Figure 6.44: Observed and synthetic vertical (a) and radial (b) component waveforms along the transect for the 2012 Feb 26 event.

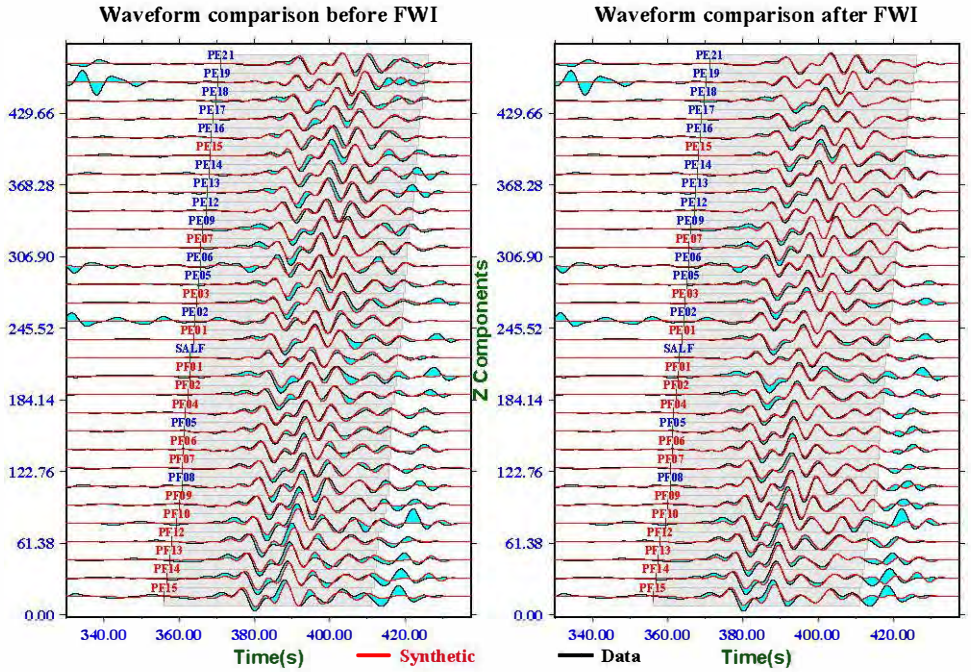


(a) vertical component waveforms along the transect.

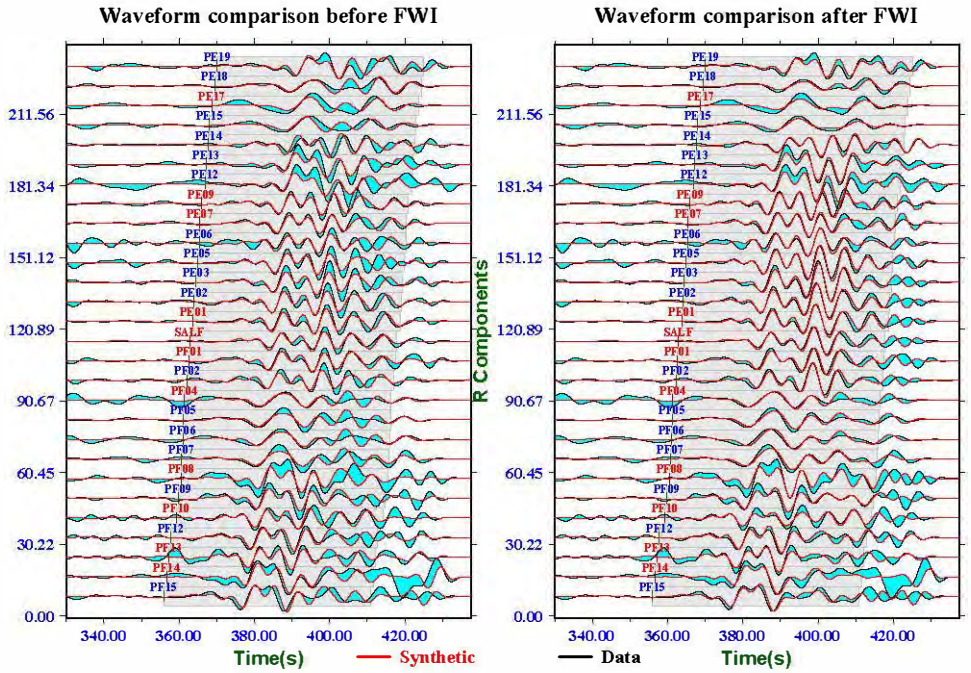


(b) radial component waveforms along the transect.

Figure 6.45: Observed and synthetic vertical (a) and radial (b) component waveforms along the transect for the 2012 Apr 02 event.

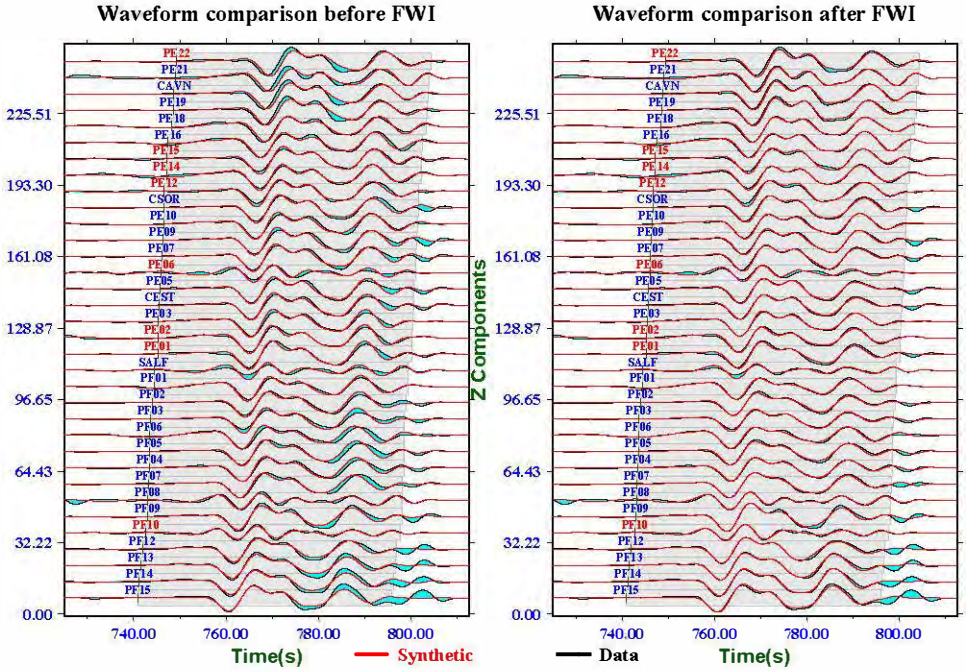


(a) vertical component waveforms along the transect.

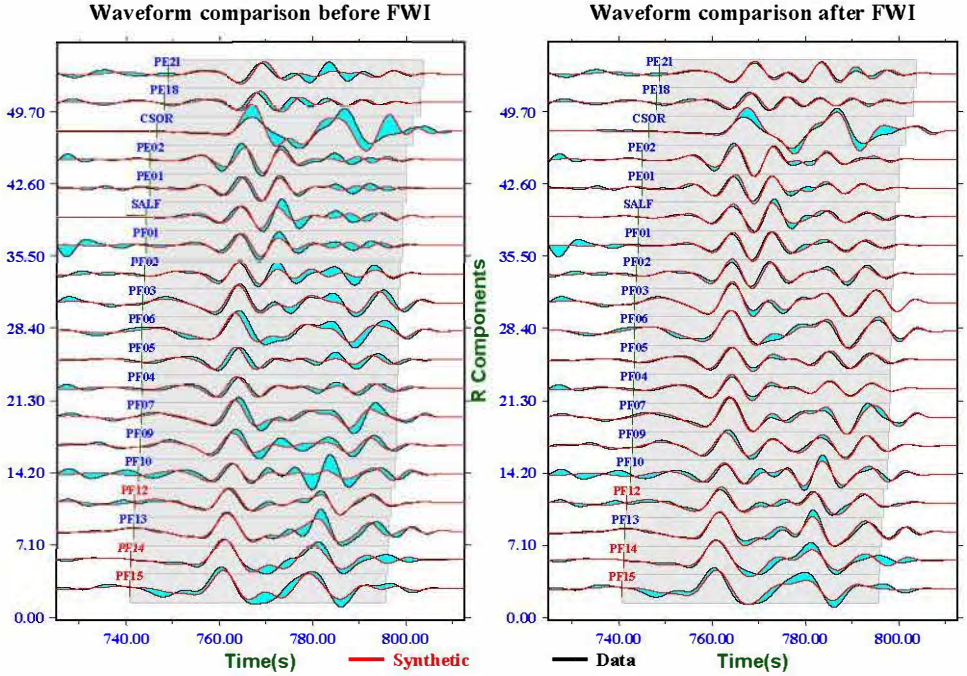


(b) radial component waveforms along the transect.

Figure 6.46: Observed and synthetic vertical (a) and radial (b) component waveforms along the transect for the 2012 May 24 event.

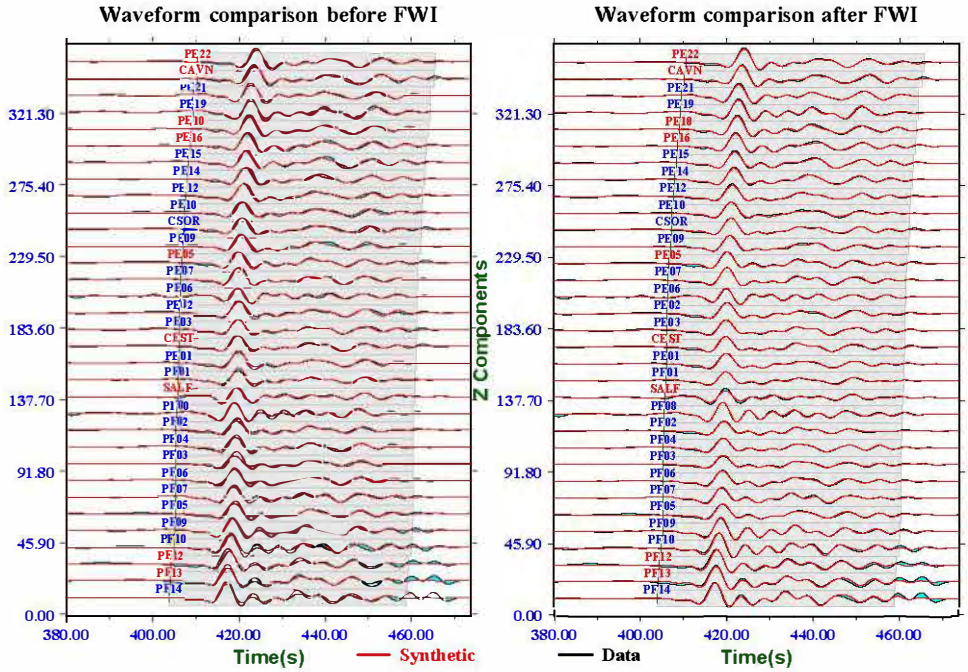


(a) vertical component waveforms along the transect.

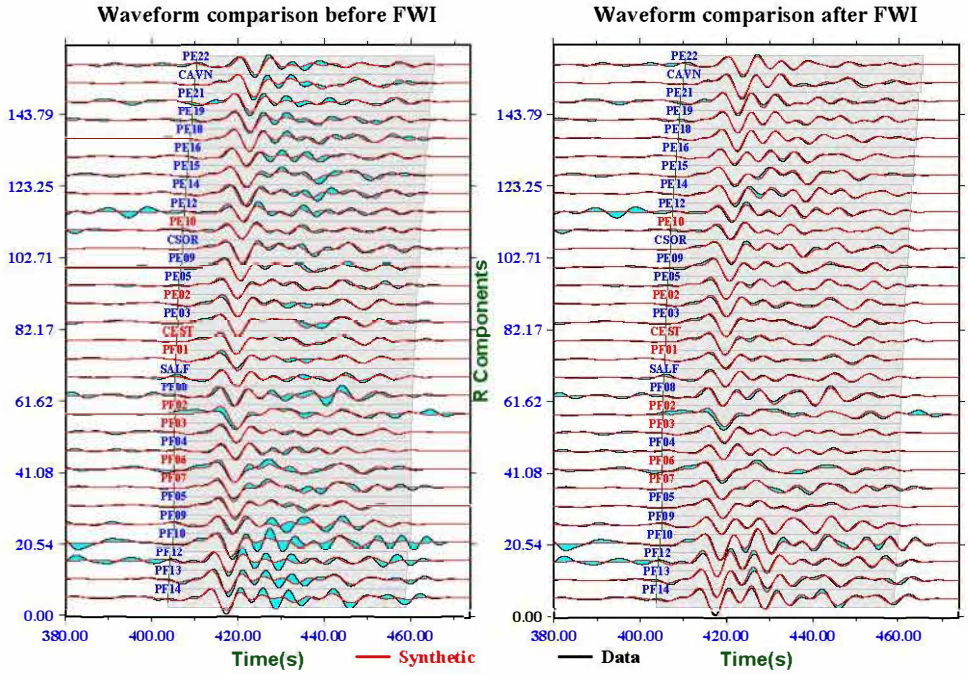


(b) radial component waveforms along the transect.

Figure 6.47: Observed and synthetic vertical (a) and radial (b) component waveforms along the transect for the 2012 Aug 10 event.



(a) vertical component waveforms along the transect.



(b) radial component waveforms along the transect.

Figure 6.48: Observed and synthetic vertical (a) and radial (b) component waveforms along the transect for the 2012 Aug 11 event.

Iberian and European Moho seen in the FWI model. The Iberian Moho dips down to about 55 km depth beneath the Axial Zone. However, determining the exact position of the northern extent of the subducting Iberian crust is still controversial because of the limited resolution in this region. We expect to obtain get a better resolved models when we perform FWI at the higher frequency or include more seismic phases and events with the different incident angle in inversion. The depth of the European Moho is around 30 km from the NPF to the northern end of the transect. We observe a high velocity anomaly in the northern part of our Vp and Vs models, between 10 and 25 km depth, beneath the Aquitaine Basin. This anomaly is likely due to the strong reverberations in the shallow sedimentary layers of the Aquitaine Basin. For this reason, the geometry of the European Moho in this region is not very well constrained. Another significant feature is that the European mantle is located on top of the subducting Iberian plate from 120 to 150 km distance along the transect. This is also consistent with the results of migration, which represents the seismic interface that produces negative conversions. To conclude, our FWI models give further support to the model of subduction of the Iberian crust beneath the European plate in the central Pyrenees (Muñoz, 1992).

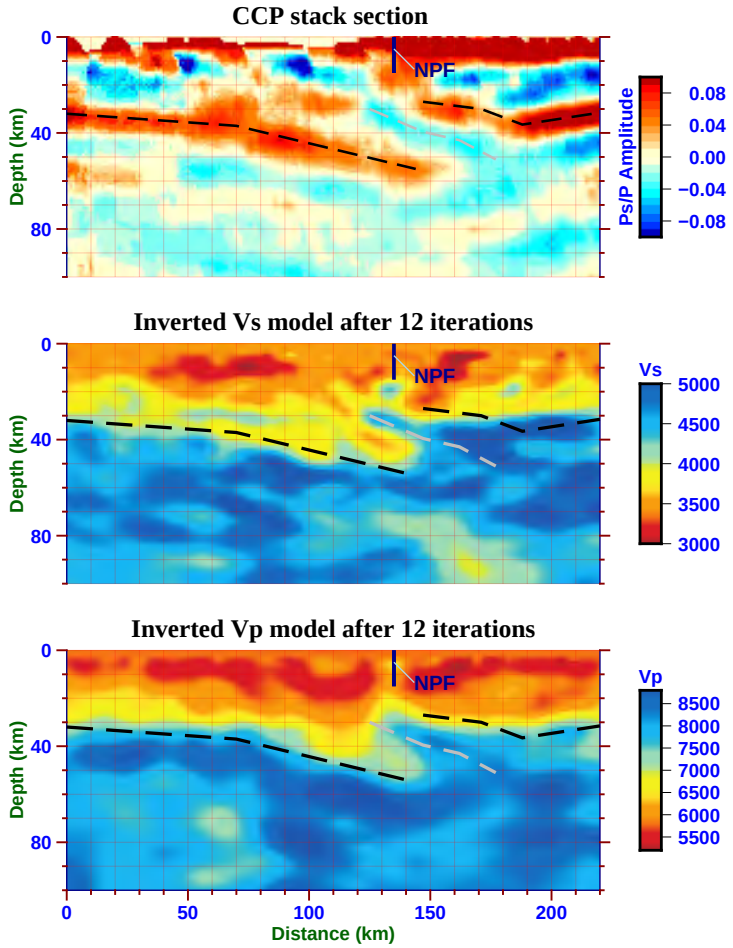


Figure 6.49: Top panel: Common conversion point stack of receiver functions for the central transect (from Chevrot et al., 2015). Middle panel: Average Vs model obtained by FWI on the original dataset using the longer time window. Bottom panel: Average Vp model obtained by FWI. The North Pyrenean Fault (NPF) is labeled by a thick blue bar in all panels. The Iberian and European Moho are represented as black dashed lines and the top of the subducting Iberian crust as a grey dashed line, obtained from the CCP stack section.

Chapter 7

Discussion and Conclusion

Contents

Thanks to the recently developed DSM/SEM hybrid method, it is now possible to compute synthetic seismograms of short period teleseismic wavefields, accounting for all the complexities that may affect the propagation of seismic waves in the regional domain of interest. This hybrid method restricts the time consuming 3D full waveform modeling inside the regional domain, and is thus very suitable for iterative waveform inversion at regional scale compared to other numerical techniques.

We applied this FWI technique on the data of the western and central PYROPE transects. The construction of the 3D regional mesh for the Pyrenees with free surface topography is described, as well as the estimation of source wavelets. The adjoint state method is used to compute the sensitivity kernel of misfit function with respect to elastic perturbations in the regional domain with the aid of the efficient hybrid method. Considering its efficiency and stability, we choose the L-BFGS algorithm for our full waveform inversion applications. The main merit of L-BFGS algorithm is that while it has nearly the same level of convergence speed as a Gauss-Newton method, it does not require to compute and store the huge inverse Hessian of the misfit function explicitly. We also describe the main issues for our full waveform inversion algorithm, such as the regularization and grids projection.

Since the real data always contain background noise, we develop an optimal frequency band selection based on the coherence between synthetic and observed seismogram. After these preparatory steps, we perform FWI on checkerboard resolution tests, and on the real data recorded by two dense transects deployed during the IBERARRAY and PYROPE experiment. We obtain high resolution Vp and Vs tomographic models beneath the western and central Pyrenees transects. The geologic interpretations of the western Pyrenees transect suggest the subduction of the Iberian crust beneath the European plate, and the European subcontinental mantle emplaced at shallow crustal levels beneath the Mauléon basin. This rift-inherited mantle wedge acted as an indenter during the Pyrenean convergence. The preliminary tomographic model for the central Pyrenees also reveals the subduction of Iberian crust under the European plate beneath the axial zone. These new models provide new and key constraints on the present

architecture of the Pyrenean lithosphere.

An important conclusion of our study is that our FWI approach has the potential to image lithospheric structures with a resolution comparable to the wavelength of inverted waves, a dramatic improvement in comparison to classical passive imaging approach. Another remarkable conclusion is that we were able to obtain detailed images with a small number of events (5), which suggests that our method is well adapted to handle data from short duration temporary deployments.

Although our FWI are encouraging, there are still several important issues that need to be investigated and require further developments:

1. The high velocity anomaly beneath the Aquitaine basin in the central Pyrenees is probably an artifact produced by reverberations in the shallow sedimentary layers of the Aquitaine Basin. These effects should be accounted for in future inversions. Alternatively, we could also try to remove these multiples before inversion.

2. Because of the low S/N ratio of some traces, we meed to improve and simplify the selection and filtering procedures to better extract the useful information contained in the waveform records.

3. Integrate all the data processing and inversion scripts into a completed program to improve the computational efficiency.

4. Change the simple Stacey absorbing boundary condition to a better type of absorbing boundary conditions such as PML.

5. Include a source wavelet update procedure in the iterative inversion, to better separate the wave propagation effects inside and outside the regional domain.

6. Exploit shear waveforms. Including the transverse component waveform in the inversion can put more constraints on the Vs model.

7. Invert for seismic anisotropy.

8. Add gravity data in a joint inversion with seismic waveforms. This should improve the long wavelength structures and the density model.

9. Apply our FWI algorithm to other continental orogens: Taiwan, Himalayas, Gibraltar, Japan, Peru, ...

7.1 Discussion et conclusion (FR)

Grâce à la méthode hybride DSM/SEM, il est désormais possible de calculer des sismogrammes synthétiques à courte période, en prenant en compte toutes les complexités qui peuvent affecter la propagation des ondes dans le domaine régional que l'on veut étudier. Cette méthode limite la modélisation 3D gourmande en temps de calcul à l'intérieur du domaine régional, et est donc bien adaptée pour l'inversion de forme d'ondes itérative.

Nous avons appliqué une nouvelle méthode d'inversion de forme d'ondes complètes aux

données de deux transects déployés à l'est et au centre des Pyrénées. Cette étude a tout d'abord nécessité de construire des maillages régionaux 3D ainsi que de déterminer les ondelettes émises par la région source pour les téléséismes utilisés. Nous avons utilisé la méthode de l'adjoint pour calculer le gradient de la fonction cout qui mesure l'écart quadratique moyen entre les formes d'ondes observées et modélisées. Les dérivées de Fréchet obtenues relient les perturbations de formes d'ondes aux perturbations des paramètres élastiques et de la densité dans le milieu. Ces gradients ont été utilisés dans une méthode d'inversion itérative reposant sur un algorithme de type L-BFGS. L'intérêt de la méthode L-BFGS est qu'elle offre une convergence aussi rapide qu'une méthode de type Gauss-Newton, sans nécessiter le calcul ni le stockage de l'inverse du Hessien. Nous décrivons également les principaux problèmes de notre algorithme d'inversion de forme d'ondes complètes, tels que la régularisation et la projection des grilles.

Les données réelles sont contaminées le bruit de fond ambiant, ce qui nous a conduit à développer une méthode de sélection optimale de la bande de fréquences utile, reposant sur la cohérence entre données synthétiques et données observées. Nous avons d'abord réalisé des inversions sur des jeux de données synthétiques dans des modèles de type damiers afin d'estimer la résolution de notre méthode tomographique. Nous avons ensuite inversé les données enregistrées par deux transects dense déployés pendant l'expérience PYROPE. Nous avons ainsi obtenu des modèles tomographiques haute résolution à la fois des vitesses P et S sous les Pyrénées occidentales et centrales. L'interprétation de la section ouest suggère le sous charriage de la plaque ibérique sous la plaque européenne, ainsi que l'exhumation du manteau sous le bassin de Mauléon, responsable de l'anomalie de Bouguer positive observée. Ce coin de manteau hérité de l'épisode de rifting crétacé a identé la plaque ibérique pour former des chevauchements de vergence sud, à l'origine des topographies dans la partie haute de la chaine. Le modèle préliminaire pour les Pyrénées centrales révèle également la subduction continentale de l'Ibérie sous la plaque Europe. Ces nouveaux modèles tomographiques apportent des contraintes très importantes concernant l'architecture profonde des Pyrénées.

Une conclusion importante de cette étude est que l'inversion de formes d'ondes est capable d'imager les structures lithosphériques avec une résolution comparable à la longueur d'onde la plus courte. Ceci constitue une avancée considérable par rapport aux approches d'imagerie passive classiques. Une autre conclusion remarquable est que cette approche donne de bons résultats même avec un nombre limité de sources sismiques, ce qui démontre qu'elle est bien adaptée à l'exploitation des données des expériences temporaires de courte durée.

Même si les résultats obtenus sont extrêmement encourageants, un certain nombre de pistes restent à explorer pour améliorer notre méthode:

1. Les images sont encore polluées par les ondes multiples piégées dans les couches de sédiments non consolidés proches de la surface. Dans le futur, il faudra raffiner le maillage proche de la surface pour mieux décrire ces couches dans lesquelles les vitesses sismiques sont très faibles. On pourra également chercher à supprimer ces multiples des données avant inversion.

2. Les données réelles sont contaminées par le bruit ambiant et il sera important de continuer à améliorer et simplifier les étapes de filtrage et de sélection des traces.

3. Intégrer tous les scripts utilisés dans la préparation et l'inversion des données dans un même programme de façon à simplifier la procédure d'inversion et la rendre plus efficace.

4. Implémenter des conditions de bords absorbants plus performantes du type PML.

5. Mettre à jour les ondelettes source de façon itérative de façon à mieux isoler les effets de propagation à l'intérieur et à l'extérieur du domaine régional 3D.

6. Inclure les ondes de cisaillement dans l'inversion.

7. Etendre l'inversion au cas anisotrope.

8. Ajouter les données gravimétriques et résoudre un problème conjoint de façon à mieux résoudre le modèle de densité.

9. Appliquer notre approche d'inversion de formes d'ondes complètes à d'autres orogènes continentaux: Himalaya, Alpes, Taiwan, etc...

Bibliography

Akcelik, V., Biros, G., & Ghattas, O., 2002. Parallel multiscale gauss-newton-krylov methods for inverse wave propagation, in *Supercomputing. ACM/IEEE 2002 Conference*, pp. 41–41, IEEE.

Akcelik, V., Bielak, J., Biros, G., Epanomeritakis, I., Fernandez, A., Ghattas, O., Kim, E., Lopez, J., O'Hallaron, D., Tu, T., & Urbanic, J., 2003. High resolution forward and inverse earthquake modeling on terascale computers, in *Supercomputing. 2003 ACM/IEEE Conference*, pp. 52–52, IEEE.

Aki, K. & Lee, W., 1976. Array using first p arrival times from local earthquakes, *Journal of Geophysical research*.

Aravkin, A., Leeuwen, T. V., & Herrmann, F., 2011. Robust full-waveform inversion using the student's t-distribution, in *2011 SEG Annual Meeting*, Society of Exploration Geophysicists.

Bamberger, A., Chavent, G., & Lailly, P., 1979. About the stability of the inverse problem in 1-d wave equations—application to the interpretation of seismic profiles, *Applied Mathematics and Optimization*, **5**(1), 1–47.

Bamberger, A., Chavent, G., Hemon, C., & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics*, **47**(5), 757–770.

Barnett-Moore, N., Hosseinpour, M., & Maus, S., 2016. Assessing discrepancies between previous plate kinematic models of mesozoic iberia and their constraints, *Tectonics*, **35**(8), 1843–1862.

Bekara, M. & der Baan, M. V., 2007. Local singular value decomposition for signal enhancement of seismic data, *Geophysics*, **72**(2), V59–V65.

Bielak, J. & Christiano, P., 1984. On the effective seismic input for non-linear soil-structure interaction systems, *Earthquake engineering & structural dynamics*, **12**(1), 107–119.

Birch, F., 1961. The velocity of compressional waves in rocks to 10 kilobars: Part 2, *Journal of Geophysical Research*, **66**(7), 2199–2224.

Blacker, T., Bohnhoff, W., & Edwards, T., 1994. Cubit mesh generation environment users manual, *Sandia National Labs. Albuquerque. NM*, **1**.

Bostock, M., 2007. Teleseismic body-wave scattering and receiver-side structure, *Treatise on Geophysics*, **1**, 219–246.

Bostock, M. G., Rondenay, S., & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 1. theory for oblique incidence, *Journal of Geophysical Research: Solid Earth*, **106**(B12), 30771–30782.

Bronner, A., Sauter, D., Manatschal, G., Péron-Pinvidic, G., & Munschy, M., 2011. Magmatic breakup as an explanation for magnetic anomalies at magma-poor rifted margins, *Nature Geoscience*, **4**(8), 549–553.

Bronner, A., Sauter, D., Manatschal., G., Péron-Pinvidic, G., & Munschy, M., 2012. Reply to'problematic plate reconstruction', *Nature Geoscience*, **5**(10), 677–677.

Brossier, R., Operto, S., & Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion?, *Geophysics*, **75**(3), R37–R46.

Brunet, M. F., 1986. The influence of the evolution of the pyrenees on adjacent basins, *Tectonophysics*, **129**(1-4), 343–354.

Bube, K. P. & Langan, R. T., 1997. Hybrid l1/l2 minimization with applications to tomography, *Geophysics*, **62**(4), 1183–1195.

Bunks, C., Saleck, F. M., Zaleski, S., & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.

Campillo, M., Feignier, B., Bouchon, M., & BÉthoux, N., 1993. Attenuation of crustal waves across the alpine range, *Journal of Geophysical Research: Solid Earth*, **98**, 1987–1996.

Capdeville, Y., Chaljub, E., & Montagner, J. P., 2003a. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophysical Journal International*, **152**(1), 34–67.

Capdeville, Y., To, A., & Romanowicz, B., 2003b. Coupling spectral elements and modes in a spherical earth: an extension to the 'sandwich'case, *Geophysical Journal International*, **154**(1), 44–57.

Capdeville, Y., Gung, Y., & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophysical Journal International*, **162**(2), 541–554.

Casas, A., Kearey, P., Rivero, L., & Adam, C. R., 1997. Gravity anomaly map of the pyrenean region and a comparison of the deep geological structure of the western and eastern pyrenees, *Earth and Planetary Science Letters*, **150**(1), 65–78.

Cawood, P. A., Kröner, A., Collins, W. J., Kusky, T. M., Mooney, W. D., & Windley, B. F., 2009. Accretionary orogens through earth history, *Geological Society. London. Special Publications*, **318**(1), 1–36.

Chaljub, E., Capdeville, Y., & Vilotte, J. P., 2003. Solving elastodynamics in a fluid-solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids, *Journal of Computational Physics*, **187**(2), 457–491.

Chaljub, E., Komatitsch, D., Vilotte, J. P., Capdeville, Y., Valette, B., & Festa, G., 2007. Spectral-element analysis in seismology, *Advances in Geophysics*, **48**, 365–419.

Chazalon, A., Campillo, M., Gibson, R., & Carreno, E., 1993. Crustal wave propagation anomaly across the pyrenean range. comparison between observations and numerical simulations, *Geophysical Journal International*, **115**(3), 829–838.

Chevalier, C. & Pellegrini, F., 2008. Cubit mesh generation environment users manual, *Parallel Computing*, **34**(6-8), 318–331.

Chevrot, S., 2002. Optimal measurement of relative and absolute delay times by simulated annealing, *Geophysical Journal International*, **151**(1), 164–171.

Chevrot, S. & Zhao, L., 2007. Multiscale finite-frequency rayleigh wave tomography of the kaapvaal craton, *Geophysical Journal International*, **169**(1), 201–215.

Chevrot, S., Sylvander, M., S., B., Ponsolles, C., Lefevre, J. M., & Paradis, D., 2007. Source locations of secondary microseisms in western europe: Evidence for both coastal and pelagic sources, *Journal of Geophysical Research: Solid Earth*, **112**(B11).

Chevrot, S., Martin, R., & Komatitsch, D., 2012. Optimized discrete wavelet transforms in the cubed sphere with the lifting scheme. implications for global finite-frequency tomography, *Geophysical Journal International*, **191**(3), 1391–1402.

Chevrot, S., Villaseñor, A., Sylvander, M., Benahmed, S., Beucler, E., Cougoulat, G., Delmas, P., Blanquat, M., Diaz, J., Gallart, J., Grimaud, F., Lagabrielle, Y., Manatschal, G., Mocquet, A., Pauchet, H., Paul, A., Péquegnat, C., Quillard, O., Roussel, S., Ruiz, M., & Wolyniec, D., 2014. High-resolution imaging of the pyrenees and massif central from the data of the pyrope and iberarray portable array deployments, *Journal of Geophysical Research: Solid Earth*, **119**(8), 6399–6420.

Chevrot, S., Sylvander, M., Diaz, J., Ruiz, M., Paul, A., & the PYROPE Working Group, 2015. The pyrenean architecture as revealed by teleseismic p-to-s converted waves recorded along two dense transects, *Geophysical Journal International*, **200**(2), 1096–1107.

Chiao, L. & Kuo, B., 2001. Multiscale seismic tomography, *Geophysical Journal International*, **145**(2), 517–527.

Choukroune, P., 1989. The ecors pyrenean deep seismic profile reflection data and the overall structure of an orogenic belt, *Tectonics*, **8**(1), 23–39.

Choukroune, P., 1992. Tectonic evolution of the pyrenees, *Annual Review of Earth and Planetary Sciences*, **20**, 143.

Choukroune, P. & Mattauer, M., 1978. Tectonique des plaques et pyrenees; sur le fonctionnement de la faille transformante nord-pyreneenne; comparaisons avec des modeles actuels, *Bulletin de la Société géologique de France*, **7**(5), 689–700.

Choukroune, P., Pinet, B., Roure, F., & Cazes, M., 1990. Major hercynian thrusts along the ecors pyrenees and biscay lines, *Bull. Soc. Geol. Fr*, **8**, 313–320.

Christensen, D. & Ruff, L., 1985. Analysis of the trade-off between hypocentral depth and source time function, *Bulletin of the Seismological Society of America*, **75**(6), 1637–1656.

Christensen, N. I., 2004. Serpentinites. peridotites and seismology, *International Geology Review*, **46**(9), 795–816.

Claerbout, J. F., 1971. Toward a unified theory of reflector mapping, *Geophysics*, **36**(3), 467–481.

Clayton, R. W. & Wiggins, R. A., 1976. Source shape estimation and deconvolution of teleseismic bodywaves, *Geophysical Journal International*, **47**(1), 151–177.

Crase, E., Pica, A., Noble, M., McDonald, J., & Tarantola, A., 1990. Robust elastic nonlinear waveform inversion: Application to real data, *Geophysics*, **55**(5), 527–538.

Cummins, P. R., Geller, R. J., Hatori, T., & Takeuchi, N., 1994a. Dsm complete synthetic seismograms: Sh spherically symmetric case, *Geophysical Research Letters*, **21**(7), 533–536.

Cummins, P. R., Geller, R. J., & Takeuchi, N., 1994b. Dsm complete synthetic seismograms: P-sv spherically symmetric case, *Geophysical research letters*, **21**(15), 1663–1666.

Cyr, E. C., Shadid, J. N., & Wildey, T., 2015. Towards efficient backward-in-time adjoint computations using data compression techniques, *Computer Methods in Applied Mechanics and Engineering*, **288**, 24–44.

Dahlen, F. A. & Baig, A. M., 2002. Fréchet kernels for body-wave amplitudes, *Geophysical Journal International*, **150**(2), 440–466.

Daignières, M., 1978. Géophysique et faille nord-pyrénéenne, *Bulletin de la Société Géologique de France*, **7**(5), 677–680.

Daignières, M., Gallart, J., Banda, E., & Hirn, A., 1982. Implications of the seismic structure for the orogenic evolution of the pyrenean range, *Earth and Planetary Science Letters*, **57**(1), 88–100.

Daignieres, M., Cabissole, B. D., Gallart, J., Hirn, A., Surinach, E., & Torné, M., 1989. Geophysical constraints on the deep structure along the ecors pyrenees line, *Tectonics*, **8**(5), 1051–1058.

Daignières, M., Séguret, M., Specht, M., & Team, E., 1994. The arzacq-western pyrenees ecors deep seismic profile, in *Hydrocarbon and petroleum geology of France*, pp. 199–208, Springer.

Dalmau, F. R., Hanzich, M., de la Puente, J., & Gutiérrez, N., 2014. Lossy data compression with dct transforms, in *EAGE Workshop on High Performance Computing for Upstream*.

Dietz, R. S., 1961. Continent and ocean basin evolution by spreading of the sea floor, *Nature*, **190**(4779), 854–857.

Dziewonski, A. & anderson, D. L., 1981. Preliminary reference earth model, *Physics of the earth and planetary interiors*, **25**(4), 297–356.

Fichtner, A., Bunge, H. P., , & Igel, H., 2006. The adjoint method in seismology i. theory, *Physics of the Earth and Planetary Interiors*, **157**, 86–104.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H. P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophysical Journal International*, **175**, 665–685.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H. P., 2009. Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods, *Geophysical Journal International*, **179**, 1703–1725.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H. P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the australasian upper mantle, *Earth and Planetary Science Letters*, **290**(3-4), 270–280.

Fichtner, A., , & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophysical Journal International*, **187**(3), 1604–1624.

Fichtner, A., Saygin, E., Taymaz, T., Cupillard, P., Capdeville, Y., , & Trampert, J., 2013. The deep structure of the north anatolian fault zone, *Earth and Planetary Science Letters*, **373**, 109–117.

Fink, M., 1992. Time reversal of ultrasonic fields. i. basic principles, *IEEE transactions on Ultrasonics Ferroelectrics and Frequency Control*, **39**(5), 555–566.

Fink, M. & Prada, C., 2001. Acoustic time-reversal mirrors, *Inverse Problems*, **17**(1), R1.

Fornberg, B., 1998. *A practical guide to pseudospectral methods*, vol. 1, Cambridge university press.

Gallart, J., Daignières, M., Banda, E., Surinach, E., & Hirn, A., 1980. The eastern pyrenean domain-lateral variations at crust-mantle level, in *Annales de Géophysique*, vol. 36, pp. 141–157, EDITIONS CNRS 20/22 RUE ST. AMAND. 75015 PARIS. FRANCE.

Gauthier, O., Virieux, J., & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results, *Geophysics*, **51**(7), 1387–1403.

Geller, R. J. & Hara, T., 1993. Two efficient algorithms for iterative linearized inversion of seismic waveform data, *Geophysical Journal International*, **115**(3), 699–710.

Geller, R. J. & Ohminato, T., 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**(2), 421–446.

Geller, R. J. & Takeuchi, N., 1995. A new method for computing highly accurate dsm synthetic seismograms, *Geophysical Journal International*, **123**(2), 449–470.

Grandjean, G., 1992. *Etudes des structures crustales dans une portion de la chaîne et de leur relation avec les bassins sédimentaires. Application auc Pyrénées Occidentales au travers du projet ECORS-Arzacq-Pyrénées*, Ph.D. thesis, PhD thesis. Université de Montpellier III. Montpellier. France.

Ha, T., Chung, W., & Shin, C., 2009. Waveform inversion using a back-propagation algorithm and a huber function norm, *Geophysics*, **74**(3), R15–R24.

Hansen, S. E., Julia, J., A., A. N., Pyle, M. L., Wiens, D. A., & Anandakrishnan, S., 2009. Using s wave receiver functions to estimate crustal structure beneath ice sheets: An application to the transantarctic mountains and east antarctic craton, *Geochemistry. Geophysics. Geosystems*, **10**(8).

Hara, T., 2004. Waveform inversion for 3-d earth structure using the direct solution method implemented on vector-parallel supercomputer, *Physics of the Earth and Planetary Interiors*, **146**(1), 65–74.

Hara, T. & Geller, R. J., 2000. Simultaneous waveform inversion for three-dimensional earth structure and earthquake source parameters considering a wide range of modal coupling, *Geophysical Journal International*, **142**(2), 539–550.

Hess, H. H., 1962. History of ocean basins, *Petrologic studies*, **4**, 599–620.

Hudson, J. A., 1977. Scattered waves in coda of p, *Journal of Geophysical research*, **43**(1-2), 359–374.

Igel, H., Djikpéssé, H., & Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for p impedance and poisson's ratio, *Geophysical Journal International*, **124**(2), 363–371.

Jammes, S., Manatschal, G., Lavier, L., & Masini, E., 2009. Tectonosedimentary evolution related to extreme crustal thinning ahead of a propagating ocean: Example of the western pyrenees, *Tectonics*, **28**(4).

Jammes, S., Tiberi, C., & Manatschal, G., 2010. 3d architecture of a complex transcurrent rift system: the example of the bay of biscay–western pyrenees, *Tectonophysics*, **489**(1), 210–226.

Jeong, W., Kang, M., Kim, S., Min, D., & Kim, W., 2015. Full waveform inversion using student'st distribution: a numerical study for elastic waveform inversion and simultaneous-source method, *Pure and Applied Geophysics*, **172**(6), 1491–1509.

Kallivokas, L. F., Fathi, A., Kucukcoban, S., Stokoe, K. H., Bielak, J., & Ghattas, O., 2013. Site characterization using full waveform inversion, *Soil Dynamics and Earthquake Engineering*, **47**, 62–82.

Karner, G. D. & Watts, A. B., 1983. Gravity anomalies and flexure of the lithosphere at mountain ranges, *Journal of Geophysical Research: Solid Earth*, **88**(B12), 10449–10477.

Kawai, K., Takeuchi, N., & Geller, R. J., 2006. Complete synthetic seismograms up to 2 hz for transversely isotropic spherically symmetric media, *Geophysical Journal International*, **164**(2), 411–424.

Kennett, B. L. N., 1998. On the density distribution within the earth, *Geophysical Journal International*, **132**(2), 374–382.

Kennett, B. L. N., R., E. R. E., & Buland, R., 1995. Constraints on seismic velocities in the earth from traveltimes, *Geophysical Journal International*, **122**(1), 108–124.

Kikuchi, M. & Kanamori, H., 1982. Inversion of complex body waves, *Bulletin of the Seismological Society of America*, **72**(2), 491–506.

Kim, Y., Liu, Q., & Tromp, J., 2011. Adjoint centroid-moment tensor inversions, *Geophysical Journal International*, **186**(1), 264–278.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophysical journal international*, **139**(3), 806–822.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation-i. validation, *Geophysical Journal International*, **149**(2), 390–412.

Komatitsch, D. & Vilotte, J. P., 1998. The spectral element method: An efficient tool to simulate the seismic response of 2d and 3d geological structures, *Bulletin of the seismological society of America*, **88**(2), 368–392.

Komatitsch, D., Liu, Q., Tromp, J., Süss, P., , Stidham, C., & Shaw, J. H., 2004. Simulations of ground motion in the los angeles basin based upon the spectral-element method, *Bull. Seismol. Soc. Am.*, **94**(1), 187–206.

Komatitsch, D., Tsuboi, S., & Tromp, J., 2005. The spectral-element method in seismology, *Seismic Earth: Array Analysis of Broadband Seismograms*, pp. 205–227.

Komatitsch, D., Labarta, J., & Michéa, D. L., 2008. A simulation of seismic wave propagation at high resolution in the inner core of the earth on 2166 processors of marenostrum, in *International Conference on High Performance Computing for Computational Science*, pp. 364–377, Springer.

Lagabrielle, Y., Labaume, P., & de Saint Blanquat, M., 2010. Mantle exhumation. crustal denudation and gravity tectonics during cretaceous rifting in the pyrenean realm (sw europe): Insights from the geological setting of the lherzolite bodies, *Tectonics*, **29**(4).

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Proceedings of the Conference on Inverse Scattering, Theory and Application Expanded Abstracts*, pp. 206–220, A. Society of Industrial and Applied Mathematics, Philadelphia.

Lévêque, J. J., Rivera, L., & Wittlinger, G., 1993. On the use of the checker-board test to assess the resolution of tomographic inversions, *Geophysical Journal International*, **115**(1), 313–318.

Ligorria, J. P. & Ammon, C. J., 1999. Iterative deconvolution and receiver-function estimation, *Bulletin of the seismological Society of America*, **89**(5), 1395–1400.

Liu, Q. & Gu, Y., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics*, **566**, 31–66.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bulletin of the Seismological Society of America*, **96**(6), 2383–2397.

Luo, Y. & Schuster, G. T., 1991. Wave-equation traveltime inversion, *Geophysics*, **56**(5), 645–653.

Luo, Y., Tromp, J., Denel, B., & Calandra, H., 2013. 3d coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods, *Geophysics*, **78**(4), S193–S202.

Macquet, M., Paul, A., Pedersen, H. A., Villaseñor, A., Chevrot, S., Sylvander, M., Wolyniec, D., & Group, P. W., 2014. Ambient noise tomography of the pyrenees and the surrounding regions: inversion for a 3-d vs model in the presence of a very heterogeneous crust, *Geophysical Journal International*, **199**(1), 402–415.

Marfurt, K. J., 1984. Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, *Geophysics*, **49**(5), 533–549.

Martin, R., Komatitsch, D., Blitz, C., & Goff, N. L., 2008. Simulation of seismic wave propagation in an asteroid based upon an unstructured mpi spectral-element method: blocking and non-blocking communication strategies, in *International Conference on High Performance Computing for Computational Science*, pp. 350–363, Springer.

Masini, E., Manatschal, G., Tugend, J., Mohn, G., & Flament, J. M., 2014. The tectono-sedimentary evolution of a hyper-extended rift basin: the example of the arzacq–mauléon rift system (western pyrenees. sw france), *International Journal of Earth Sciences*, **103**(6), 1569–1596.

Minshull, T. A., 2009. Geophysical characterisation of the ocean–continent transition at magma-poor rifted margins, *Comptes Rendus Geoscience*, **341**(5), 382–393.

Monteiller, V., Chevrot, S., Komatitsch, D., & Fuji, N., 2013. A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-d regional model, *Geophysical Journal International*, **192**(1), 230–247.

Monteiller, V., Chevrot, S., Komatitsch, D., & Wang, Y., 2015. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the sem–dsm hybrid method, *Geophysical Journal International*, **202**(2), 811–827.

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, **52**(9), 1211–1228.

Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics*, **53**(6), 750–759.

Mosegaard, K. & Tarantola, A., 1995. Monte carlo sampling of solutions to inverse problems, *Journal of Geophysical Research: Solid Earth*, **100**(B7), 12431–12447.

Mouthereau, F., Filleaudeau, P. Y., Vacherat, A., Pik, R., Lacombe, O., Fellin, M. G., Castelltort, S., Christophoul, F., & Masini, E., 2014. Placing limits to shortening evolution in the pyrenees: Role of margin architecture and implications for the iberia/europe convergence, *Tectonics*, **33**(12), 2283–2314.

Muñoz, J. A., 1992. Evolution of a continental collision belt: Ecors-pyrenees crustal balanced cross-section, in *Thrust tectonics*, pp. 235–246, Springer.

Nissen-Meyer, T., van Driel, M., Stähler, S. C., Hosseini, K., Hempel, S., Auer, L., Colombi, A., & Fournier, A., 2014. Axisem: broadband 3-d seismic wavefields in axisymmetric media, *Solid Earth*, **5**(1), 425.

Nocedal, J. & Wright, S., 2006. *Numerical optimization*, Springer Science & Business Media.

Nocquet, J. M., 2012. Present-day kinematics of the mediterranean: A comprehensive overview of gps results, *Tectonophysics*, **579**, 220–242.

Nocquet, J. M. & Calais, E., 2004. Geodetic measurements of crustal deformation in the western mediterranean and europe, *Pure and Applied Geophysics*, **161**(3), 661–681.

Nolet, G., 2012. *Seismic tomography: with applications in global seismology and exploration geophysics*, vol. 5, Springer Science & Business Media.

Oliveira, S. & Seriani, G., 2011. Effect of element distortion on the numerical dispersion of spectral element methods, *Communications in Computational Physics*, **9**(4), 937.

Pageot, D., Operto, S., Vallée, M., Brossier, R., & Virieux, J., 2013. A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging, *Geophysical Journal International*, **193**(3), 1479–1505.

Park, J., Lindberg, C. R., & Vernon, F. L., 1987. Multitaper spectral analysis of high-frequency seismograms, *Journal of Geophysical Research: Solid Earth*, **92**(B12), 12675–12684.

Pecher, I. A., Minshull, T. A., Singh, S. C., & von Huene, R., 1996. Velocity structure of a bottom simulating reflector offshore peru: Results from full waveform inversion, *Earth and Planetary Science Letters*, **139**(3), 459–469.

Pedreira, D., Pulgar, J. A., Gallart, J., & Torné, M., 2007. Three-dimensional gravity and magnetic modeling of crustal indentation and wedging in the western pyrenees-cantabrian mountains, *Journal of Geophysical Research: Solid Earth*, **112**(B12).

Peter, D., Komatitsch, D., Luo, Y., Martin, R., Goff, N. L., Casarotti, E., Loher, P. L., Magnoni, F., Liu, Q., Blitz, C., Nissen-Meyer, T., Basini, P., & Tromp, J., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophysical Journal International*, **186**(2), 721–739.

Pica, A., Diet, J. P., & Tarantola, A., 1990. Nonlinear inversion of seismic reflection data in a laterally invariant medium, *Geophysics*, **55**(3), 284–292.

Pichon, X. L. & Sibuet, J. C., 1971. Western extension of boundary between european and iberian plates during the pyrenean orogeny, *Earth and Planetary Science Letters*, **12**(1), 83–88.

Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

Plessix, R. E. & Perkins, C., 2010. Thematic set: Full waveform inversion of a deep water ocean bottom seismometer dataset, *First Break*, **28**(4), 71–78.

Plessix, R. E., Baeten, G., de Maag, J. W., ten F. Kroode, & Rujie, Z., 2012. Full waveform inversion and distance separated simultaneous sweeping: A study with a land seismic data set, *Geophysical Prospecting*, **60**(4), 733–747.

Pratt, R. G., 1999. Seismic waveform inversion in the frequency domain. part 1: Theory and verification in a physical scale model, *Geophysics*, **64**(3), 888–901.

Pratt, R. G., Song, Z., Williamson, P., & Warner, M., 1996. Two-dimensional velocity models from wide-angle seismic data by wavefield inversion, *Geophysical Journal International*, **124**(2), 323–340.

Pratt, R. G., Shin, C., & Hick, G. J., 1998. Gauss–newton and full newton methods in frequency–space seismic waveform inversion, *Geophysical Journal International*, **133**(2), 341–362.

Pyrenees, E. S. G., 1980. Seismic reconnaissance of the structure of the pyrénées, *Annales de Géophysique*, **36**(2), 135–140.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., & Dell'Aversana, P., 2004. Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: Application to a thrust belt, *Geophysical Journal International*, **159**(3), 1032–1056.

Rickers, F., Fichtner, A., & Trampert, J., 2013. The iceland–jan mayen plume system and its impact on mantle dynamics in the north atlantic region: evidence from full-waveform inversion, *Earth and Planetary Science Letters*, **367**, 39–51.

Ritsema, J., Rivera, L. A., Komatitsch, D., Tromp, J., & van Heijst, H. J., 2002. Effects of crust and mantle heterogeneity on pp/p and ss/s amplitude ratios, *Geophysical research letters*, **29**(10).

Roecker, S., Baker, B., & McLaughlin, J., 2010. A finite-difference algorithm for full waveform teleseismic tomography, *Geophysical Journal International*, **181**(2), 1017–1040.

Roure, F. & Choukroune, P., 1998. Contribution of the ecors seismic data to the pyrenean geology: Crustal architecture and geodynamic evolution of the pyrenees, *Mémoires de la Société géologique de France*, **173**, 37–52.

Roure, F., Choukroune, P., Berastegui, X., Muñoz, J. A., Villien, A., Matheron, P., Bareyt, M., Seguret, M., amara, P., & Deramond, J., 1989. Ecors deep seismic data and balanced cross sections: Geometric constraints on the evolution of the pyrenees, *Tectonics*, **8**(1), 41–50.

Royden, L. L. & Karner, G. D., 1984. Flexure of lithosphere beneath apennine and carpathian foredeep basins: evidence for an insufficient topographic load, *AAPG Bulletin*, **68**(6), 704–712.

Rudnick, R. L. & Fountain, D. M., 1995. Nature and composition of the continental crust: a lower crustal perspective, *Reviews of geophysics*, **33**(3), 267–309.

Sambridge, M., 1999a. Geophysical inversion with a neighbourhood algorithm-i. searching a parameter space, *Geophysical Journal International*, **138**(2), 479–494.

Sambridge, M., 1999b. Geophysical inversion with a neighbourhood algorithm-ii. appraising the ensemble, *Geophysical Journal International*, **138**(3), 727–746.

Sambridge, M. & Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion, *Geophysical Journal International*, **109**(2), 323–342.

Schneider, F. M., Yuan, X., Schurr, B., Mechie, J., Sippl, C., Haberland, C., Minaev, V., Oimahmadov, I., Gadoev, M., Radjabov, N., Abdybachaevc, U., Orunbaevc, S., & Negmatullaevd, S., 2013. Seismic imaging of subducting continental lower crust beneath the pamir, *Earth and Planetary Science Letters*, **375**, 101–112.

Sen, M., Stoffa, P. L., & Pardalos, P. M., 1997. Global optimization methods in geophysical inversion, *Journal of Global Optimization*, **10**(4), 467–468.

Sens-Schönfelder, C., Margerin, L., & Campillo, M., 2009. Laterally heterogeneous scattering explains lg blockage in the pyrenees, *Journal of Geophysical Research: Solid Earth*, **114**(B7).

Seriani, G. & Oliveira, S., 2008. Dispersion analysis of spectral element methods for elastic wave propagation, *Wave Motion*, **45**(6), 729–744.

Seriani, G. & Priolo, E., 1994. Spectral element method for acoustic wave simulation in heterogeneous media, *Finite elements in analysis and design*, **16**(3), 337–348.

Sibuet, J. C. & Pichon, X. L., 1971. Structure gravimétrique du golfe de gascogne et le fossé marginal nord-espagnol, *Histoire structurale du Golfe de Gascogne*, **2**.

Sibuet, J. C., Srivastava, S. P., & Spakman, W., 2004. Pyrenean orogeny and plate kinematics, *Journal of Geophysical Research: Solid Earth*, **109**(B8).

Sigloch, K. & Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes, *Geophysical Journal International*, **167**(1), 271–287.

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., & Kommedal, J. H., 2010. Thematic set: Full waveform inversion: The next leap forward in imaging at valhall, *First Break*, **28**(4), 65–70.

Slepian, D., 1978. Prolate spheroidal wave functions. fourier analysis and uncertainty—v: The discrete case, *Bell System Technical Journal*, **57**(5), 1371–1430.

Souquet, P. & Mediavilla, F., 1976. Nouvelle hypothèse sur la formation des pyrénées, *Comptes rendus de l'Académie des sciences*, **282**, 2139–214.

Souriau, A. & Granet, M., 1995. A tomographic study of the lithosphere beneath the pyrenees from local and teleseismic data, *Journal of Geophysical Research: Solid Earth*, **100**(B9), 18117–18134.

Souriau, A., Chevrot, S., & Olivera, C., 2008. A new tomographic image of the pyrenean lithosphere from teleseismic data, *Tectonophysics*, **460**(1), 206–214.

Stacey, R., 1988. Improved transparent boundary formulations for the elastic-wave equation, *Bulletin of the Seismological Society of America*, **78**(6), 2089–2097.

Stehly, L., Campillo, M., & Shapiro, N. M., 2007. Traveltime measurements from noise correlation: stability and detection of instrumental time-shifts, *Geophysical Journal International*, **171**(1), 223–230.

Suriñach, E., Marthelot, J. M., Gallart, J., Daignières, M., & Hirn, A., 1993. Seismic images and evolution of the iberian crust in the pyrenees, *Tectonophysics*, **221**(1), 67–80.

Takeuchi, N. & Geller, R. J., 2003. Accurate numerical methods for solving the elastic equation of motion for arbitrary source locations, *Geophysical Journal International*, **154**(3), 852–866.

Takeuchi, N., Geller, R. J., & Cummins, P. R., 1996. Highly accurate p-sv complete synthetic seismograms using modified dsm operators, *Geophysical research letters*, **23**(10), 1175–1178.

Talwani, M., 1973. Computer usage in the computation of gravity anomalies, *Geophysics*, **38**, 343–389.

Tape, C., Liu, Q., & Tromp, J., 2007. Finite-frequency tomography using adjoint methods—methodology and examples using membrane surface waves, *Geophysical Journal International*, **168**(3), 1105–1129.

Tape, C., Liu, Q., Maggi, A., & Tromp, J., 2009. Adjoint tomography of the southern california crust, *Science*, **325**(5943), 988–992.

Tape, C., Liu, Q., Maggi, A., & Tromp, J., 2010. Seismic tomography of the southern california crust based on spectral-element and adjoint methods, *Geophysical Journal International*, **180**(1), 433–462.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms including elasticity and attenuation, *Pure and Applied Geophysics*, **128**(1-2), 365–399.

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, siam.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Reviews of Geophysics*, **20**(2), 219–232.

Team, E. P., 1988. The ecors deep reflection seismic survey across the pyrenees, *Nature*, **331**(6156), 508–511.

Teixell, A., 1998. Crustal structure and orogenic material budget in the west central pyrenees, *Tectonics*, **17**(3), 395–406.

Thomson, D. J., 1982. Spectrum estimation and harmonic analysis, *Proceedings of the IEEE*, **70**(9), 1055–1096.

Tong, P., Chen, C., Komatitsch, D., Basini, P., & Liu, Q., 2014a. High-resolution seismic array imaging based on an sem-fk hybrid method, *Geophysical Journal International*, **197**(1), 369–395.

Tong, P., Komatitsch, D., Tseng, T., Hung, S., Chen, C., Basini, P., & Liu, Q., 2014b. A 3-d spectral-element and frequency-wave number hybrid method for high-resolution seismic array imaging, *Geophysical Research Letters*, **41**(20), 7025–7034.

Torné, M., Cabissole, B. D., Bayer, R., Casas, A., Daignières, M., & Rivero, A., 1989. Gravity constraints on the deep structure of the pyrenean belt along the ecors profile, *Tectonophysics*, **165**(1-4), 105–116.

Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography. adjoint methods. time reversal and banana-doughnut kernels, *Geophysical Journal International*, **160**(1), 195–216.

Tromp, J., Komatitsch, D., & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Communications in Computational Physics*, **3**(1), 1–32.

Tsuboi, S., Komatitsch, D., Ji, C., & Tromp, J., 2003. Broadband modeling of the 2002 denali fault earthquake on the earth simulator, *Physics of the Earth and Planetary Interiors*, **139**(3), 305–313.

Tucholke, B. E. & Sibuet, J. C., 2012. Problematic plate reconstruction, *Nature Geoscience*, **5**(10), 676–677.

Tucholke, B. E., Sawyer, D. S., & Sibuet, J. C., 2007. Breakup of the newfoundland–iberia rift, *Geological Society. London. Special Publications*, **282**(1), 9–46.

Tugend, J., Manatschal, G., Kusznir, N. J., Masini, E., Mohn, G., & Thinon, I., 2014. Formation and deformation of hyperextended rift systems: Insights from rift domain mapping in the bay of biscay-pyrenees, *Tectonics*, **33**(7), 1239–1276.

Vacher, P. & Souriau, A., 2001. A three-dimensional model of the pyrenean deep structure based on gravity modelling. seismic images and petrological constraints, *Geophysical Journal International*, **145**(2), 460–470.

Vacherat, A., Mouthereau, F., Pik, R., Bernet, M., Gautheron, C., Masini, E., Pourhiet, L. L., Tibari, B., & Lahfid, A., 2014. Thermal imprint of rift-related processes in orogens as recorded in the pyrenees, *Earth and Planetary Science Letters*, **408**, 296–306.

Vai, R., Castillo-Covarrubias, J. M., Sánchez-Sesma, F., Komatitsch, D., & Vilotte, J. P., 1999. Elastic wave propagation in an irregularly layered medium, *Soil Dynamics and Earthquake Engineering*, **18**(1), 11–18.

van der Voo, R., 1969. Paleomagnetic evidence for the rotation of the iberian peninsula, *Tectonophysics*, **7**(1), 5–56.

Vergés, J. & Fernàndez, M., 2012. Tethys–atlantic interaction along the iberia–africa plate boundary: The betic–rif orogenic system, *Tectonophysics*, **579**, 144–172.

Vine, F. J. & Wilson, J. T., 1965. Magnetic anomalies over a young oceanic ridge off vancouver island, *Science*, **150**(3695), 485–489.

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

Vissers, R. L. M. & Meijer, P. T., 2012. Mesozoic rotation of iberia: Subduction in the pyrenees?, *Earth-Science Reviews*, **110**(1), 93–110.

Watts, A. B., 2001. *Isostasy and Flexure of the Lithosphere*, Cambridge University Press.

Wilson, J. T., 1965. A new class of faults and their bearing on continental drift, *Nature*, **207**, 343–347.

Wu, Q., Li, Y., Zhang, R., & Zeng, R., 2007. Wavelet modelling of broad-band receiver functions, *Geophysical Journal International*, **170**(2), 534–544.

Wu, R. & Yang, F., 1997. Seismic imaging in wavelet domain: Decomposition and compression of imaging operator, in *Optical Science. Engineering and Instrumentation'97*, pp. 148–162, International Society for Optics and Photonics.

Yilmaz, Ö., 1990. *Seismic data processing*, Video Studios of Western Atlas International.

Zhang, T. & Lay, T., 1995. Why the lg phase does not traverse oceanic crust, *Bulletin of the Seismological Society of America*, **85**(6), 1665–1678.

Zhu, H., Bozdağ, E., Peter, D., & Tromp, J., 2012. Structure of the european upper mantle revealed by adjoint tomography, *Nature Geoscience*, **5**(7), 493–498.

Zhu, H., Bozdağ, E., & Tromp, J., 2015. Seismic structure of the european upper mantle based on adjoint tomography, *Geophysical Journal International*, **201**(1), 18–52.