


# THÈSE

**En vue de l'obtention du**

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :** *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

**Présentée et soutenue le** *09/02/2023* **par :**

**Li-Yu KAN**

---

# Développements méthodologiques pour l'inversion des formes d'ondes télésismiques complètes et applications à l'imagerie des zones de subduction des Cascades et du sud Pérou

---

**JURY**

| | | |
|---|---|---|
| Stephanie DUCHENE | Professeur d'Université | Présidente du Jury |
| Nobuaki FUJI | Maître de Conférences HDR | Rapporteur |
| Stephane RONDENAY | Professeur d'Université | Rapporteur |
| Joseph MARTINOD | Professeur d'Université | Examinateur |
| Shu-Huei HUNG | Professeur d'Université | Examinatrice |
| Sébastien CHEVROT | Directeur de Recherche CNRS | Directeur de thèse |
| Vadim MONTEILLER | Ingénieur de Recherche CNRS | Co-Directeur de thèse |

---

**École doctorale et spécialité :**
*SDU2E : Sciences de la Terre et des Planètes Solides*
**Unité de Recherche :**
*Géosciences Environnement Toulouse (UMR 5563)*
**Directeur(s) de Thèse :**
*Sébastien CHEVROT* et *Vadim MONTEILLER*
**Rapporteurs :**
*Nobuaki FUJI* et *Stephane RONDENAY*

# Résumé

La tomographie sismique nous permet d'imager des propriétés physiques telles que vitesses sismiques ou densité, fournissant ainsi des contraintes directes sur la composition et l'état thermique de l'intérieur de la Terre. La densification des réseaux sismologiques temporaires comme permanents, ainsi que les récentes avancées dans le domaine du calcul haute performance ont ouvert de nouvelles opportunités pour l'application des approches d'inversion de forme d'onde complète (FWI) aux données sismologiques. En effet, celles-ci permettent d'obtenir des modèles tomographiques multiparamétriques quantitatifs de la subsurface avec une limite de résolution allant jusqu'à la demie longueur d'onde. Les inversions de formes d'onde télésismiques à l'échelle régionale sont ainsi devenues réalisables sur des clusters CPU ou GPU de taille modérée et cette méthode a déjà été appliquée avec succès pour imager la structure lithosphérique des Pyrénées et des Alpes.

Dans cette thèse, nous apportons des améliorations à la méthode d'inversion de formes d'ondes qui permettent de prendre en compte les corrélations existant entre les paramètres du modèle. Pour cela, nous introduisons une matrice de covariance du modèle dont les termes non-diagonaux décrivent la corrélation entre chaque paire de paramètres du modèle. Des expériences d'inversion synthétique sont réalisées sur un modèle de subduction simple en prenant en compte les corrélations entre la densité, $V_P$ et $V_S$. Nous démontrons qu'une telle matrice de covariance du modèle non diagonale complète nous permet d'obtenir des résultats d'inversion cohérents, c'est-à-dire indépendants de la paramétrisation du modèle choisie. Avec cette nouvelle approche, nous obtenons de meilleurs modèles reconstruits avec moins d'artefacts numériques. En outre, les inversions sont plus rapidement que les inversions qui supposent que les paramètres du modèle sont indépendants. Nous appliquons ensuite la nouvelle méthode aux formes d'onde télésismiques P et SH enregistrées par l'expérience temporaire CASC93, un profil sismique déployé dans le centre de l'Oregon, aux États-Unis.

Les modèles finaux de densité, $V_P$, $V_S$, et $V_P/V_S$ révèlent la structure lithosphérique de la zone de subduction des Cascades avec une résolution sans précédent. En particulier, nous imageons la couche de faible vitesse associée à la croûte océanique de la plaque Juan de Fuca, le coin mantellique serpentinisé sous l'avant-arc, et la fusion partielle sous l'arc volcanique à environ 80 km de profondeur. Enfin, la nouvelle méthode FWI est appliquée aux formes d'ondes télésismiques enregistrées par plusieurs expériences temporaires déployées dans le sud du Pérou (PeruSE, CAUGHT, PULSE). Notre région d'étude couvre à la fois les segments de subduction normale et plate dans le nord des Andes centrales. Les modèles finaux de densité, $V_P$ et $V_S$ fournissent des informations nouvelles sur la structure de l'Altiplano et la géométrie de la plaque Nazca en subduction. Les faibles vitesses observées sous l'Altiplano dans la partie sud du modèle suggèrent que la lithosphère a disparu, ce qui pourrait expliquer le récent soulèvement de cette partie des Andes. Nous observons également une faible anomalie de vitesse sous le segment plat, qui pourrait expliquer le maintien d'un pendage faible pour la plaque subductée.

## Summary

Seismic tomography allows us to image physical properties such as seismic velocities or density, providing direct constraints on the composition and thermal state of the Earth's interior. The densification of seismic deployments, as well as the recent advances in high-performance computing, have opened new opportunities for applying full-waveform inversion (FWI) approaches to seismic data. Indeed, FWI can produce quantitative multi-parameter tomographic models of the subsurface with a resolution limit down to half a wavelength. Today, full waveform inversion of teleseismic waves at the regional scale has become feasible on moderate-size CPU or GPU clusters. This method has already been successfully applied to imaging the lithospheric structure of the Pyrenees and the Alps.

In this thesis, we make improvements to the FWI method that allows taking into account the correlation between the model parameters. For this purpose, we introduce a model covariance matrix whose non-diagonal terms describe the correlation between each pair of model parameters. Synthetic inversion experiments are performed on a simple subduction model taking into account the correlations between the density, $V_P$, and $V_S$. We show that such a complete model covariance matrix leads to consistent inversion results, i.e., independent of the chosen model parameterization. With this new approach, we obtain better-reconstructed models with fewer numerical artifacts. In addition, the inversions have a faster convergence rate compared to inversions that assume that the model parameters are independent. We then apply the new method to the teleseismic P and SH waveforms recorded by the CASC93 temporary experiment, a seismic profile deployed in central Oregon, USA. The final density, $V_P$, $V_S$, and $V_P/V_S$ models reveal the lithospheric structure of the Cascadia subduction zone with unprecedented resolution. In particular, we image the low-velocity layer associated with the oceanic crust of the Juan de Fuca plate, the serpentinized mantle wedge beneath the forearc, and the partial melting beneath the volcanic arc at around

80 km depth. Finally, the new FWI method is applied to teleseismic waveforms recorded by several temporary experiments deployed in southern Peru (PeruSE, CAUGHT, PULSE). Our study region covers both the normal and flat subduction segments in the northern Central Andes. The final density, $V_P$, and $V_S$ models provide new insights into the structure of the Altiplano and the geometry of the subducting Nazca plate. The low velocities observed beneath the Altiplano in the southern part of the model suggest that the lithosphere has been removed, which could explain the recent uplift in this part of the Andes. We also observe a low-velocity anomaly under the flat segment, which could be the maintenance of a shallow dip for the subducted plate.

# Acknowledgements

I am deeply grateful to everyone who has helped me to complete this thesis.

First and foremost, I would like to express my special thanks to my advisors, Sébastien Chevrot and Vadim Monteiller. Sébastien guided me with great care throughout this thesis, both in terms of research and life assistance, always answering my questions with patience and care, and guiding me to explore more details of my research. It is hard to express in words my gratitude for his help. Vadim provided comprehensive assistance with methods and programming, with a keen eye for identifying and resolving the issues I reported. Without the guidance and support of the two excellent advisors, my thesis would not have gone as smoothly as it did.

I would like to thank my jury members Stephanie Duchene, Nobuaki Fuji, Stéphane Rondenay, Joseph Martinod, and Shu-Huei Hung for their positive evaluations, thoughtful feedback, and meticulous review of my thesis. The insightful discussions we had during the defense were also invaluable.

I want to thank my colleagues Clément and Martinod. I felt like I belonged in France by chatting with you. I want to express my appreciation to the Taiwanese members of the KFC Addiction Group for your company since the lockdown. I thank Hao Kuo-Chen for his constant contact and encouragement. As this work required intense numerical calculations, I would like to also thank CALMIP (UT3) and Jean-Zay (IDRIS) that provide the high-performance computing resources.

Thanks to France, I had the opportunity to experience European life for an extended period, allowing me to understand the importance of work-life balance. The rich cultural activities and historical heritage were feasts for my eyes.

Last, I want to express my deep appreciation to my family for their unwavering support, encouragement, and love throughout my academic journey. Without their support, I could not have achieved this milestone in my life.

# Contents


**1. Introduction** ..... **1**
1.1 Objectives and content of the thesis ..... 11

**2. Method** ..... **15**
2.1 Solving the wave equation ..... 15
2.1.1 Regional 3D modeling with Spectral Element Method ..... 15
2.1.2 Modeling teleseismic wavefields with hybrid methods ..... 26
2.1.3 Practical implementation of the Spectral Element Method ..... 29
2.2 Inverse Problem ..... 34
2.2.1 Gradient based local optimization ..... 37
2.3 Application to the FWI problem ..... 41
2.3.1 Gradient computation based on adjoint state method ..... 43
2.3.2 Regularization ..... 48

**3. Preparation of FWI inversions** ..... **56**
3.1 Outline of the inversion workflow ..... 56
3.2 Data Selection ..... 57
3.3 Mesh design ..... 62
3.4 Source wavelet estimation ..... 64
3.5 Event normalization of adjoint sources ..... 67

**4. A consistent multi-parameter Bayesian full waveform inversion scheme for imaging heterogeneous isotropic elastic media** ..... **72**
4.1 Summary ..... 73
4.2 Introduction ..... 73
4.3 Formulation of a consistent Bayesian full waveform inversion problem ..... 76
4.3.1 Model parameterization and consistency of the inverse problem ..... 77
4.3.2 The model covariance matrix for a 3-D isotropic Earth model ..... 80
4.4 Numerical experiments ..... 84
4.4.1 Description of the model ..... 84
4.4.2 Computation of synthetic seismograms with the FK/SEM hybrid method 86
4.4.3 Results ..... 88
4.5 Application to real data: the southern Peru transect ..... 97
4.6 Discussion ..... 100
4.6.1 Tuning FWI hyperparameters: model parameterization, standard deviations, correlation coefficients, and smoothing length ..... 100

4.6.2 Joint inversions of P and SH events ..... 102
4.7 Conclusions ..... 105
Appendix 4.A Decomposition of the model covariance matrix ..... 108
Appendix 4.B Change of model parameterization and transformation of the model covariance matrix ..... 110

**5. Dehydration of the subducting Juan de Fuca plate and fluid pathways revealed by full waveform inversion of teleseismic P and SH waves in central Oregon ..... 113**
5.1 Abstract ..... 114
5.2 Introduction ..... 115
5.3 Data and methodology ..... 118
5.3.1 The CASC93 experiment ..... 118
5.3.2 Data selection and preparation ..... 118
5.3.3 Design of the tomographic mesh ..... 122
5.3.4 Definition of the cost function ..... 124
5.3.5 The full waveform inversion method ..... 124
5.3.6 The iterative inversion algorithm ..... 126
5.3.7 Event normalization ..... 127
5.4 Results ..... 128
5.4.1 Tomographic model obtained after hierarchical inversion ..... 128
5.4.2 Waveform fits ..... 131
5.5 Discussion ..... 133
5.5.1 The impact of adding SH waveforms and accounting for the correlation between model parameters ..... 133
5.5.2 Scaling relationships between model parameters and the effect of the initial smooth 1D model ..... 138
5.5.3 Fluid migrations, melting, and metamorphic reactions in the central Cascadia subduction revealed by FWI ..... 139
5.6 Conclusions ..... 144

**6. Application of FWI to image the normal and flat subductions beneath the northern Central Andes ..... 145**
6.1 Introduction ..... 146
6.2 Data and methodology ..... 149
6.2.1 Temporary deployments in southern Peru ..... 149
6.2.2 Data preparation and selection ..... 152
6.2.3 The tomographic meshes ..... 171
6.2.4 The full waveform inversion method ..... 174
6.3 FWI of PE profile data in the small mesh ..... 179
6.3.1 The model ..... 179
6.3.2 Waveform fits ..... 181
6.4 FWI of in the big mesh with all the permanent and temporary stations installed in southern Peru ..... 184
6.4.1 The model ..... 186

6.4.2 Waveform fits ..... 187
6.5 Discussion ..... 187
6.5.1 Architecture and composition of the northern Central Andes crust ..... 193
6.5.2 Slab geometry and comparison between FWI and travel time tomography 195
6.5.3 Lithosphere delamination ..... 197
6.5.4 Buoyancy of the flat slab beneath central Peru ..... 198
6.5.5 The fate of the Amazonian lithosphere ..... 200
6.6 Conclusions ..... 200
Appendix 6.A Waveform fits after hierarchical FWI inversions in small box inversion 202

**7. Conclusions and perspectives ..... 209**
7.1 Français ..... 209
7.2 English ..... 214

# List of Figures

2.1 Transformation of one element between the physical coordinate and the reference coordinate ..... 19
2.2 GLL points ..... 22
2.3 Acceptable step length with Wolfe's conditions ..... 42
2.4 Two-dimensional Gaussian intensity with different correlation coefficient ..... 54

3.1 FWI workflow ..... 57
3.2 Graphical interface for data selection ..... 59
3.3 Example of data selection ..... 61
3.4 Vertical cross-section of the Cascadia mesh ..... 63
3.5 Effects of event normalization in vertical component ..... 69
3.6 Effects of event normalization in radial component ..... 70
3.7 $V_S$ gradient for each teleseismic event without event normalization ..... 70
3.8 $V_S$ gradient for each teleseismic event with event normalization ..... 71
3.9 Comparison of total $V_S$ gradient with and without event normalization. (a) $V_S$ gradient obtained by summing all the individual $V_S$ gradients shown in Figure 3.7 (without event normalization). (b) $V_S$ gradient obtained by summing all the individual $V_S$ gradients shown in Figure 3.8 (with event normalization). .. 71

4.1 Synthetic model ..... 84
4.2 Scattering plots of joint variations of the selected three pairs of parameters .... 87
4.3 FWI results inverted with noisy data ..... 88
4.4 An example of L-curve ..... 91
4.5 FWI results inverted without noise ..... 91
4.6 Consistency FWI results obtained in different parameterization ..... 93

4.7 Hierarchical FWI inversion ..... 94
4.8 Scatter plots for inversion with and without correlations between model parameters $\rho$, $V_P$, and $V_S$ ..... 96
4.9 Evolution of the normalized cost reduction ..... 96
4.10 PERUSE deployment ..... 97
4.11 FWI image of the first profile of PERUSE deployment ..... 99
4.12 FWI models obtained with different correlation coefficients ..... 100
4.13 FWI models obtained with different standard deviation of $V_P$ ..... 103
4.14 FWI models obtained with or without including SH events ..... 104
4.15 Waveform fits ..... 106
4.16 The consistency of bivarariate normal distributions over different parameterizations ..... 112

5.1 Positions of the CASC93 broadband stations ..... 119
5.2 Example of data selection procedure ..... 120
5.3 Teleseismic events used in the study ..... 121
5.4 Procedure of source wavelet estimation ..... 122
5.5 Spectral-element mesh ..... 123
5.6 Evolution of relative cost reductions for the four hierarchical inversions ..... 127
5.7 Results of the hierarchical full waveform inversion ..... 129
5.8 Comparison between final $V_S$ model and interfaces from (Bostock et al., 2002) 131
5.9 Waveform fits for event 9 ..... 132
5.10 Waveform fits ..... 134
5.11 Residual misfits for the different types of inversions ..... 135
5.12 Results of the hierarchical FWI with different inversion strategies ..... 137
5.13 Scatter plot of $V_P$ and $\rho$ in the final model ..... 139
5.14 Interpretation of the final $V_S$ model ..... 143

6.1 Station location and regional map of southern Peru ..... 150
6.2 Data availability chart of the temporary experiments in southern Peru between 2008 and 2013 ..... 151
6.3 Four-stage evolution of the PeruSE deployment ..... 153
6.4 Example using the Matlab interface to visualize the waveform ..... 154
6.5 Zoom-in of the residual maps for the 2012/04/12 event shown in the GUI example ..... 156
6.6 Scatter plots for the travel time residuals and the amplitude residuals ..... 157
6.7 Location of teleseismic P events used for data correction ..... 161
6.8 Histogram of difference of time and amplitude measurements ..... 162
6.9 Scatter plots for the travel time residuals and the amplitude residuals after data correction ..... 162
6.10 Waveform example before and after correction ..... 163
6.11 Location of teleseismic events used for waveform inversion ..... 164
6.12 Travel time residual maps for the selected P events ..... 165
6.13 Travel time residual maps for the selected S events ..... 167
6.14 Amplitude residual maps for the selected P events ..... 168

6.15 Amplitude residual maps for the selected S events ..... 170
6.16 Illustration of the average source wavelet estimation procedure ..... 172
6.17 Range of the big and small mesh ..... 173
6.18 Evolution of cost reductions of the four hierarchical inversions ..... 175
6.19 Standard deviation of model parameters as the function of depth ..... 178
6.20 Results of the hierarchical FWI in small box inversion ..... 182
6.21 Results of the hierarchical FWI in small box inversion (zoom-in) ..... 183
6.22 Waveform fits of the PF transect for the 2010/01/12 event ..... 185
6.23 Waveform fits of the PF transect for the 2010/03/07 event ..... 185
6.24 Preliminary 10 s results of the FWI in big mesh along the PE transect. ..... 188
6.25 Preliminary 10 s results of the FWI in big mesh along the PF transect. ..... 189
6.26 Preliminary 10 s results of the FWI in big mesh along the PG transect. ..... 190
6.27 Preliminary 10 s results of the FWI in big mesh along the PG transect (zoom-in) 191
6.28 Waveform fits of the PG transect for the 2012/04/12 event ..... 192
6.29 Waveform fits of the PG transect for the 2013/01/05 event ..... 192
6.30 Comparison of the 10 s $V_P$ perturbation model from big-box inversion and the one from finite-frequency tomography ..... 196
6.31 Comparison of the 10 s $V_S$ perturbation model from big-box inversion and the one from finite-frequency tomography ..... 197
6.32 Waveform fits of the PE transect for the 2008/09/10 event ..... 202
6.33 Waveform fits of the PE transect for the 2010/01/05 event ..... 203
6.34 Waveform fits of the PE transect for the 2010/04/04 event ..... 203
6.35 Waveform fits of the PE transect for the 2010/12/08 event ..... 204
6.36 Waveform fits of the PE transect for the 2012/01/15 event ..... 204
6.37 Waveform fits of the PE transect for the 2012/01/22 event ..... 205
6.38 Waveform fits of the PE transect for the 2012/04/12 event ..... 205
6.39 Waveform fits of the PE transect for the 2013/01/05 event ..... 206
6.40 Waveform fits of the PE transect for the 2010/01/17 event ..... 206
6.41 Waveform fits of the PE transect for the 2011/03/01 event ..... 207
6.42 Waveform fits of the PE transect for the 2011/03/01 event ..... 207

# List of Tables

3.1 Parameters for the three different meshes used in this thesis. ..... 63

4.1 Model parameters used to build the synthetic model ..... 85

5.1 List of teleseismic events selected for waveform inversion. ..... 121
5.2 Hyperparameters used in the Cascadia inversions ..... 126

6.1 Selected station pairs for estimating the travel and amplitude shifts. ..... 161
6.2 List of P and SH teleseismic events selected for this study. ..... 164
6.3 Hyperparameters used in the small-grid inversions in Peru ..... 176
6.4 Approximate CPU hours required to perform forward and adjoint computations 179
6.5 Residual misfits for each event ..... 208

# Chapter 1. Introduction

## Contents

1.1 Objectives and content of the thesis ..... **11**

The main objective of geophysics is to study the structure and composition of the Earth's interior by quantitative physical methods. Geophysical researches also have many applications for mineral resource exploration, natural disaster prevention, and environmental protection. By using physical signals such as seismic waves, gravity, geomagnetism, electricity, or geothermics, we can indirectly estimate the properties of Earth's interior that we cannot directly observe. Among these disciplines, seismology is concerned with the study of earthquakes themselves and related phenomena such as the propagation of seismic waves, earthquake hazards, volcanic activity, and plate motions. Seismology is an observational science in which seismic waves are recorded at seismic stations after they have propagated through the Earth's interior from a source. The observed waveforms carry the imprint of the medium through which they have propagated in addition to the information of the source. By analyzing the propagation of seismic waves at different scales, it is possible to determine the internal structure, from the surface to the center of the Earth, which is known as seismic imaging.

Seismic imaging applications cover a wide range of observations that can extend from the first few meters of the subsurface to the entire Earth. Near-surface applications, such as monitoring geothermal sites, developing hydrocarbon reservoirs, assessing CO2 storage areas, or monitoring volcanic activity, often have important economic aspects. In contrast, deep Earth applications focus on long-term scientific questions, such as mantle structure and dynamics or earthquake hazards. The experimental design for these applications should take into account the required spatial resolutions, which can differ depending on the specific application.

The resolution of tomographic models depends on the type of source used, the geometry of the acquisitions, and the distance between the source and the receivers. There are two types of sources: active sources and passive sources, each providing different frequency content. In seismic exploration surveys, manually controlled active sources are used to generate high-frequency signals via explosions or ground shaking, so the location and signature of the source are known. The location of the receiver is also known whether it is at the surface or in the borehole. In this configuration, the acquisition is fully controlled, meaning that data of one source-receiver pair can be stacked with multiple measurements to enhance the signal or moved freely to increase ray coverage. This type of acquisition is ideal for imaging near-surface structures using reflected waves, as depth migration can produce high-resolution images of the reflectivity of the medium (Yilmaz, 2001). However, since the energy of the active source is relatively low, the ability to image deep lithospheric structures is limited.

Passive seismic methods, which use natural earthquakes as a source of energy to analyze the Earth's structure, are generally less expensive compared to seismic exploration which typically involves the use of specialized equipment and skilled personnel. This is because natural earthquakes tend to have higher energy, which can penetrate to great depths and effectively illuminate the subsurface. As a result, passive seismic imaging is widely used in geophysical investigations (e.g. Nolet, 2008).

However, there are some limitations to using natural earthquakes as a source of energy. The location and signature of the source are uncertain, and earthquakes are unevenly distributed across the globe. This results in variable resolution of seismic images, depending on the geographic region being studied. Large earthquakes are typically confined to plate boundaries, and seismological stations are predominantly located on continents, with insufficient station coverage in the ocean and southern hemisphere. Seismic events occurring at greater distances, typically over a few tens of degrees, must travel longer propagation paths. As a result, the high-frequency signal generated by the source attenuates as it propagates through the Earth, leaving only its low-frequency components to be recorded. Thus, the

use of natural earthquakes may not provide the same level of spatial resolution as seismic exploration. However, it can still provide valuable insights into large-scale structures (few kilometers).

Seismic imaging techniques have been applied to imaging subducting zones (e.g. Zhao et al., 2016), orogenic belts (e.g. Wu et al., 2014) or mantle plumes (e.g. Koppers et al., 2021). An area of interest is the lithosphere, the outermost rigid part of the Earth, consisting of the crust and a portion of the upper mantle, which is about 100 to 250 km thick beneath the continents. This outer shell is divided into several plates (Pichon, 1968). The study of the structure and evolution of the lithosphere is of great importance for understanding the mechanisms governing plate tectonics, geodynamic processes, and geologic history, or the generation of stable continental cratons (Artemieva, 2011). It is very difficult to obtain high-resolution 3D tomographic images at the lithospheric scale. Since the depth distribution of regional earthquakes is extremely uneven, even with dense arrays, there is a problem of insufficient ray coverage at depth.

Two widely used methods for regional-scale imaging are receiver functions and travel-time tomography. In the receiver function method, the diffracted fields on the discontinuities of the medium are exploited to estimate their depth and reflectivity. Inversion or migration of receiver functions allows reconstruction of the geometry of seismic discontinuities and has become an important tool for studying lithospheric discontinuities (e.g. Kind et al., 2002; Rondenay et al., 2008; Chevrot et al., 2015). The second approach, travel time tomography, is based on ray theory and provides velocity perturbation models by minimizing the difference between measured travel times and those predicted in a reference Earth model. It allows imaging long-wavelength structures (Aki et al., 1977). Recent developments in seismic tomography have focused on considering more complicated media and propagation physics, as well as on better exploitation of the information carried in seismograms. Of particular interest is full waveform inversion (FWI) because its resolution can potentially reach half the wavelength (Virieux and Operto, 2009), which would constitute a dramatic improve-

ment over conventional tomographic approaches. In the following, we briefly introduce the development of seismic tomography, with an emphasis on FWI.

The seismic tomography method using travel times was first proposed by Aki and Lee (1976). It was used to constrain 3-D crustal P-wave velocity anomalies in Bear Valley, California, by inverting the deviation between picked and predicted P-wave arrival travel times from local events. Early seismic tomographic studies were primarily based on asymptotic high-frequency ray theory and the use of P-wave or S-wave traveltime information. The ray paths were obtained by ray-tracing methods (Cerveny, 2001). Because of its simplicity, ray-based travel-time tomography has been widely used in regional and global applications (see Rawlinson et al., 2010; Liu and Gu, 2012). Earthquakes can be classified into two types of sources according to their distance from receivers. Local earthquakes are defined by epicentral distances of less than a few hundred kilometers, whereas teleseismic earthquakes are defined by epicentral distances of a few thousand kilometers. Local earthquake tomography involves updating model parameters and jointly estimating earthquake source parameters with the sources and receivers inside the regional tomographic grid. Improvements in local earthquake tomography include, for example, the use of three-dimensional ray tracing (Cerveny, 2001) or double-difference methods (Zhang and Thurber, 2003). Aki et al. (1977) extended regional travel time tomography to the lithospheric scale by considering teleseismic sources, an approach often referred to as ACH(77) in the literature, named after the three authors of this study. Using teleseismic P wave travel time residuals from the Norwegian NORSAR array, they obtained maps of lateral velocity anomalies within 5 layers, from the surface to 126 km depth. This method allows imaging of structures of 25 to 50 km in size.

In early traveltime tomography studies, arrival times were mostly hand-picked. To reduce artificial errors, seismologists began to measure the arrival time difference of a particular seismic phase using cross-correlation of observed seismic waveforms and synthetic seismograms. The measured arrival time difference is related to the frequency content of the signal, whereas ray theory is an asymptotic infinite frequency theory. In fact, seismic waves have a finite

frequency content, so structural anomalies away from the ray path will also contribute to the observed differential traveltimes. To describe this effect, Dahlen et al. (2000) proposed the concept of finite-frequency tomography, a theory that combines ray theory and the Born approximation of scattered waves. This study concluded that the arrival times measured by correlation are also affected by the velocity structure surrounding the ray, that the sensitivity is zero on the ray path itself, and that the extent of the effect is related to the width of the Fresnel zone (Hung et al., 2000; Zhao et al., 2000). Finite-frequency sensitivity kernels also correctly account for wavefront healing effects (Nolet and Dahlen, 2000). This approach has also found applications in exploration seismology (Luo and Schuster, 1991; Woodward, 1992). The computation of sensitivity kernels can be divided into two types: the adjoint method and the scattering integral method (Chen et al., 2007), with the former being more efficient and requiring less storage, therefore more popular.

Unlike the imaging methods presented above, full waveform inversion aims to use all the information contained in the observed seismograms to find a subsurface model that can explain the data. Since complete seismograms are considered, the information related to the arrival time, phase variations, and amplitude change of different signals, whether they are transmitted, reflected or converted waves, are inherently taken into account in the inversion. As a consequence, the resolution of the images obtained by FWI is superior to that of other conventional tomographic techniques, potentially approaching half the wavelength (Virieux and Operto, 2009). Thus, to resolve the structure within the lithosphere with a spatial resolution of several kilometers, assuming an average $V_S$ velocity of 3.5 km/s in the crust, the full waveform inversion needs to be inverted down to periods of 1-2 s.

FWI was initially introduced into the field of exploration seismology in the 1980s (Lailly, 1983; Tarantola, 1984a), in which the structure of the Earth's interior is obtained by inverting seismograms under the acoustic approximation. Tarantola (1984a) introduced the adjoint method to calculate the target gradient of the function, which avoids solving the Frechét derivative of the cost function directly. This approach is similar to the reverse time migration

used in exploration seismology (Claerbout, 1971). However, the adjoint method can adopt a variety of seismic phases, including direct, reflected, and surface waves, to construct a 3D structural model of the Earth, whereas reverse time migration uses reflected waves to obtain information on the location and amplitude of crustal reflectors. The FWI approach was later extended to elastic and viscoelastic media (Gauthier et al., 1986; Mora, 1987). Pratt (1990); Pratt et al. (1998) developed FWI in the frequency domain, but the complexity of solving the wave equations and the need for computational resources in large-scale 3-D problems still limited its development.

These pioneering works showed the potential of FWI to obtain better images of deep structures. However, they also indicated the potential challenges of FWI. For example Gauthier et al. (1986) pointed out that the numerical cost of the method, its strong non-linearity, and its ill-posed nature, made the application of FWI to real data challenging. The size of the inverse problem and the computational cost of solving it made it impossible to use global optimization methods to solve the inverse problem. For this reason, optimizations relied on local optimization methods to iteratively solve the FWI problem (Tarantola, 1984b). Several studies attempted to use global optimization methods such as simulated annealing, genetic algorithms, or MCMC (Markov chain Monte Carlo) (e.g. Sen and Stoffa, 1992, 1995; Zhao and Sen, 2021) in FWI applications. However, global optimization is slow to converge and computationally demanding. Another issue concerns the initial model. If the initial model is far from the true model, gradient-based waveform inversions can be trapped in local minima. To reduce the non-linearity of the problem, Bunks et al. (1995) proposed a multi-scale time-domain inversion method, which decomposes the problem into different frequency bands, i.e., starting from the low-frequency components and gradually moving to higher frequencies.

Tromp et al. (2005) proposes an adjoint imaging approach for tomographic imaging, which unifies the finite-frequency sensitivity kernels and adjoint methods. In this framework, various objective functions, such as full waveform residuals, finite-frequency arrival

times, or amplitudes, can be freely chosen. This method has been applied to various scales of structural imaging of the Earth's interior as it offers a superior resolution potential than traveltime tomography. With this approach, minimizing the phase misfits between observed and synthetic seismograms, crustal and upper mantle velocity structure of Southern California (Tape et al., 2009, 2010), Europe (Zhu et al., 2012, 2015), America (Ciardelli et al., 2022; Rodgers et al., 2022), Australia (Fichtner et al., 2009) as well as of the global Earth (Bozdağ et al., 2016; Lei et al., 2020) were successfully imaged. Compared to traveltime imaging, FWI using local earthquakes provides tomographic models with a finer resolution, allowing to image small-scale structures such as sedimentary basins or faults. (Tape et al., 2009) and Lee et al. (2014) first applied adjoint tomography to image the crust beneath Southern California. Their final model shows that the basin structure is better defined, with low-velocity perturbations that can reach 20% to 30% with respect to the initial model. Such large velocity perturbations are difficult to reconstruct with classical traveltime tomography, which are valid for small perturbations with respect to the reference Earth model. The 3-D model of Lee et al. (2014) clearly imaged large velocity variations related to fault zones.

In recent years, attempts have been made to improve the resolution of lithospheric models by applying FWI to teleseismic P wave records. Monteiller et al. (2015) demonstrated that FWI allows imaging simple crustal models with sharp Moho even if only four teleseismic events are used in the inversion. They also pointed out that FWI leads to better model reconstruction than adjoint tomography which only fits the phase of seismograms. The key step in FWI came from the application to data from a transect deployed in the western Pyrenees during the PYROPE experiment (Wang et al., 2016). The FWI of five teleseismic P waves recorded at 29 stations down to the period of 5 s provided P- and S-wave velocity models of the lithosphere with unprecedented resolution, in good agreement with a migrated receiver function section. Beller et al. (2018) later successfully imaged the crustal structure and the continental subduction beneath the western Alps.

Nowadays, it is still challenging to compute high-frequency (up to 1 Hz) teleseismic

seismograms in a fully 3-D medium, let alone to perform inversions. In the following, we introduce the practice and application of teleseismic FWI in detail.

**Toward efficient computation of teleseismic wavefields**

Because FWI requires multiple iterations to converge, a significant amount of computational resources is needed to perform a large number of 3-D forward modeling. In particular, for teleseismic FWI, we seek to invert short-period body waves propagating over distances of several thousand kilometers, which involves performing demanding simulations of wave propagation in the global Earth. The computational cost increases exponentially with the targeted period. For teleseismic applications, this limits the minimum period that can be processed with a reasonable numerical cost. So far, global scale simulations in a heterogeneous Earth have gone down to 17 s period (Lei et al., 2020). Such computations require a few petabytes of storage space and high-performance computing systems.

To make the calculations more efficient at the global scale while maintaining a 3-D modeling approach of the seismic wavefields, there are some approaches.

The first option is to simplify the numerical solver itself. Nissen-Meyer et al. (2007); Nissen-Meyer et al. (2014) introduced an axially symmetric modeling approach that can reduce the three-dimensional computations to a two-dimensional disk, allowing the period of the simulation of teleseismic wavefields to be as short as 1 s. Hosseini et al. (2019) used AxiSEM to model P, PP, and Pdiff waveforms down to a period of 3 s and imaged mantle structures with finite-frequency travel time tomography. The simplified numerical solver allows the 3-D wavefield simulation to be extended to higher frequencies at the expense of model structure accuracy. Subsequent theoretical and methodological developments of AxiSEM relaxed the axisymmetric assumption by describing the azimuthal dependence of 3-D wavefields in terms of Fourier series along the dimension perpendicular to the disk (Leng et al., 2016). With this approach, they were able to simulate the seismic scattering produced by small-scale heterogeneities (Leng et al., 2019).

Another option is to simplify wave propagation simulation outside the region of interest

by replacing it with simpler and more computationally efficient methods. For example, Capdeville et al. (2003) divided the entire Earth into two computational domains, coupling a 3-D SEM solver in an outer shell with a normal mode solution in an inner sphere. In this approach, the most demanding computations are limited to the outer shell. However, normal mode summation at periods below 8 s is challenging, which limits applications to rather long periods. This method has been successfully applied to model diffracted S waves at the core-mantle boundary, identifying a large ultralow-velocity zone at the base of the mantle (e.g. Yuan and Romanowicz, 2017).

The last option is to use hybrid methods to inject the pre-computed external wavefield into a regional domain, using virtual sources distributed on its boundaries. The key idea, originally proposed by Alterman and Karal Jr (1968), is to divide the wave propagation problem into two parts, in a global (external) domain, and in a regional (internal) domain. This approach is often referred to as a wavefield injection, hybrid, or domain reduction method. With this configuration, the short-period wavefields in the 3-D model are affordable from a computational point of view because the 3-D intensive computations are restricted to the (small) regional domain. In principle, we can couple any two solvers inside and outside the target region, as long as the global and regional wave fields match on the boundaries of the target domain (Adourian et al., 2022). In practice, most of the hybrid methods developed to date considered a target domain on the receiver side (e.g. Bielak and Christiano, 1984; Robertsson and Chapman, 2000; Chevrot et al., 2004; Roecker et al., 2010; Monteiller et al., 2013; Tong et al., 2014a; Wang et al., 2021; Monteiller et al., 2021). For example, Monteiller et al. (2013) proposed the first hybrid method that coupled the direct solution method (DSM) (e.g. Geller and Ohminato, 1994; Kawai et al., 2006) for computing the incident wavefield and SEM for the computation of regional wave propagation. However, DSM suffers from i/o and storage issues due to the large number of spherical harmonic coefficients that need to be computed and stored, which become quite severe for shallow sources. This motivated Monteiller et al. (2021) to provide two additional coupling methods with the frequency-

number method (FK, e.g. Zhu and Rivera, 2002) or AxiSEM (Nissen-Meyer et al., 2014) to SEM. When the regional grid is small (less than a few degrees), the curvature of the Earth and the teleseismic wavefront can be neglected, and the FK-SEM solution provides a reasonable approximation of the incident wavefield. On the other hand, when dealing with large regional domains such as in chapter 6 of this thesis where the curvature of the Earth cannot be neglected, the incident wavefields cannot be simply described as plane waves. In such cases, we resort to AxiSEM-SEM coupling. Recently, Wang et al. (2021) have extended the FK hybrid method to handle multiple plane-wave secondary arrivals. So far, these efficient hybrid methods have been applied to image the lithospheric structure of the Pyrenees (Wang et al., 2016) using DSM/SEM coupling and the Alps (Beller et al., 2018) using AxiSEM/SEM coupling.

The limitation of one-way hybrid methods is that they cannot model multiple scattered waves outside the target domain because the scattered waves produced by 3-D heterogeneities inside the regional domain are absorbed as they propagate out of the boundaries (e.g. Monteiller et al., 2021). Therefore all discrepancies between observed and synthetic waveforms are attributed to the 3D heterogeneities within the regional domain. Fortunately, these multiply scattered waves have much smaller amplitudes than the incident and the first-order scattered wavefields and can be safely neglected.

The target domain is not necessarily limited to the surface but to any location in the Earth (e.g. Masson et al., 2014; Yuan et al., 2021). Masson et al. (2014), Masson and Romanowicz (2017), and Adourian et al. (2022) developed a general framework called "box tomography" for any source-receiver configuration. This approach couples Specfem3D Globe (Komatitsch and Tromp, 2002) outside the target region and RegSEM inside the target region (Cupillard et al., 2012). Due to the nature of the external solver (Specfem3D Globe), the model can be 3-D both outside and inside the target region at the expense of higher computational costs. A similar concept is also presented by Pienkowska et al. (2020), allowing any source-receiver configuration but with AxiSEM/SEM coupling outside and inside the target region. Like

the AxiSEM/SEM coupling in Monteiller et al. (2021), the external wavefield is not 3-D, but the overall computational requirements will be lower.

## 1.1 Objectives and content of the thesis

This thesis aims to improve the existing full waveform inversion method and apply it to imaging the Cascadia and Southern Peru subduction systems. Previous applications of the FWI method have considered a parameterization of the model in terms of density, $V_P$, and $V_S$. In this thesis, we have introduced the correlation of these parameters as a priori knowledge in the inversion process. This a priori information is taken into account in the inversion via a complete covariance matrix of the model. This matrix is defined by the standard deviation and spatial correlation of each parameter in the whole model space, as well as the correlations between the different pairs of parameters. To validate the new inversion algorithm, we built a simple subduction model in which we performed various synthetic tests to characterize the improvement of the FWI results obtained using a full covariance matrix and to evaluate the feasibility and potential of the method. When the correlation between the parameters is considered, the resolution of the density and $V_P/V_S$ ratio is greatly improved. This is a significant improvement since these two parameters provide direct constraints on the thermal state and composition of the lithosphere, as well as on the presence of fluids or partial melting. Next, we apply the improved FWI algorithm to the waveform data recorded by the temporary CASC93 and PeruSE experiments. We developed a graphical interface to visualize the waveforms to facilitate selecting and preparing a high-quality data set. This interface allows us to interactively visualize auxiliary parameters, such as travel time and amplitude residuals, to identify possible problems in the waveform data. We have identified systematic errors in timing and amplitude at some stations belonging to the different profiles of the PERUSE experiment in southern Peru. After correcting the seismic waveforms, we obtained a high-resolution density, $V_P$, and $V_S$ models beneath central Cascadia and southern Peru. In both regions, we have successfully imaged a low-velocity layer associated with subduction of the oceanic crust and fine-scale structures in the crust and mantle wedge of the forearc

and backarc. These high-resolution tomographic models also provide new information about fluid migration within the subduction zones.

In Chapter 1, we review seismic imaging methods from classical traveltime tomography to full waveform inversion. We also introduce the recent development of full waveform inversion using hybrid methods.

In Chapter 2, we review the spectral-element and wavefield injection methods. We also present full waveform inversion algorithms, including the gradient computation using adjoint methods, the approximated Hessian computation using l-BFGS method, the regularization using a complete model covariance matrix, and the line search criterion using Wolfe conditions.

In Chapter 3, we describe the workflow of teleseismic full-waveform inversion. In particular, we detail the steps needed before inversion, for example, selecting a high-quality dataset, constructing a 3-D regional grid, estimating the source wavelets, and computing scaling factors for normalizing the earthquakes of different magnitudes.

In Chapter 4, we present new methodological developments to improve the teleseismic FWI method by introducing a full 3D non-diagonal model covariance matrix to take into account correlations between density, $V_P$, and $V_S$. This study has already been published in Geophysical Journal International. This new method is validated and tested on a synthetic subduction model. To construct the complete model covariance matrix, we consider the standard deviation and the spatial smoothing length for each type of parameter, as well as the correlation coefficients between each pair of model parameters. We analyze the effects of these different hyperparameters on FWI results. The standard deviations balance the contribution of the different physical parameters to the gradient. The introduction of correlations between model parameters leads to better-reconstructed models, in particular for density and the $V_P/V_S$ ratio. It also accelerates the convergence of the inversion since the correlations restrict the model space to be explored. The spatial correlation of each type of model parameter is described with an exponential covariance function. This kernel acts

as a spatial smoothing operator. We use moderate (∼10 km) spatial smoothing lengths to eliminate gradient singularities below the stations while preserving small-scale features of the gradient. Using such a complete model covariance matrix, the FWI can be consistent, i.e., independent of the chosen model parameterization. Using the model parameterization in terms of density, $V_P$, and $V_S$, we find that realistic Earth models are described by large (>0.9) correlation coefficients between the different pairs of parameters. In our inversions, we additionally include SH waves for the first time. The addition of SH waveforms improves the reconstruction of the $V_S$ model and allows the removal of artifacts in the final model, in particular those resulting from crustal multiples.

In Chapter 5, we apply the improved FWI algorithm to a seismic profile deployed in central Oregon, a small dataset that allows us to test the improved FWI algorithm on real data at relatively short periods. Our results lead to two main conclusions. First, despite the limited size of the dataset and the small number of stations available, aligned along a single profile, we are still able to retrieve high-resolution models of density, $V_P$, $V_S$ and $V_P/V_S$. After considering the correlation coefficients between model parameters, the resolution of density and $V_P/V_S$ models is greatly improved. Second, our FWI model reveals the structure at the lithospheric scale of the Cascadia subduction. In the final model, we can clearly see the subducting oceanic Juan de Fuca oceanic crust, which coincides with the distribution of tremor epicenters. The reduced density and velocity in the forearc mantle are associated with serpentinization. In addition, the different degree of serpentinization reveals the migration paths of the fluids released from the subducted JdF plate, in good agreement with the low resistivity anomalies from magnetotelluric studies. The velocity reduction at 75 km depth beneath the backarc provides evidence for partial melting.

In Chapter 6, our goal is to integrate and invert the waveform data collected by temporary experiments deployed in southern Peru. During data selection, we found a systematic shift in absolute time and amplitude between the different profiles of the PERUSE experiment. We analyzed these time and amplitude differences between stations in detail and concluded that

they can be corrected by a constant time and amplitude offset. We designed two inversion domains, a large box containing all the data and a small box inverting only the PE transect, for which we found internal consistencies in both phase and amplitude.

In Chapter 7, we summarize the main conclusion of this thesis and provide the perspectives for future development.

# Chapter 2. Method

## Contents


2.1 Solving the wave equation ..... **15**
2.1.1 Regional 3D modeling with Spectral Element Method ..... 15
2.1.2 Modeling teleseismic wavefields with hybrid methods ..... 26
2.1.3 Practical implementation of the Spectral Element Method ..... 29
2.2 Inverse Problem ..... **34**
2.2.1 Gradient based local optimization ..... 37
2.3 Application to the FWI problem ..... **41**
2.3.1 Gradient computation based on adjoint state method ..... 43
2.3.2 Regularization ..... 48


In this chapter we briefly introduce the formulation of full waveform inversion which involves both the forward problem tackled with the hybrid AxiSEM-SEM method and inverse problem, in addition to regularization strategies.

### 2.1 Solving the wave equation

The simulation of teleseismic wave propagation at the regional scale is done by using the hybrid method AxiSEM/SEM introduced in Monteiller et al. (2021). We will only give a brief overview of the method here and refer the reader to that paper for more details. The major advantage of the hybrid method is that it restricts expensive 3D computations to a limited regional domain size, which dramatically reduce the computational cost compared to full domain 3D methods.

#### 2.1.1 Regional 3D modeling with Spectral Element Method

To simulate 3D wave propagation in the regional domain we have chosen to use the open source SPECFEM3D code (https://specfem.org/). This method solves the elastodynamics equation in 3D models for elastic, isotropic or anisotropic media as well as attenuating media using spectral finite element method on unstructured hexahedral mesh. Moreover, the SPECFEM3D code is optimized and is very efficient on modern computers, whether

on CPU or GPU. Particular attention has been given to the parallelization to achieve high performance. Below we give a quick overview of the spectral element method, for further details we suggest the reader to refer to Komatitsch and Vilotte (1998) and Komatitsch and Tromp (1999)

#### *2.1.1.1 Weak form of elastodynamics equations*

The displacement field $\mathbf{u}$ produced by a source $\mathbf{f}$ is governed by the momentum equation

$$\rho(\mathbf{x})\frac{\partial^2\mathbf{u}(\mathbf{x},t)}{\partial t^2} = \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\mathbf{x},t) + \mathbf{f}(\mathbf{x},t), \tag{2.1}$$

where $\rho$ is the distribution of density and $\boldsymbol{\sigma}$ the stress tensor. In the linear elasticity framework, the stress tensor depends on the displacement gradient $\nabla\mathbf{u}$ by Hooke's law which may be written in the form $\boldsymbol{\sigma} = \mathbf{c} : \nabla\mathbf{u}$, where $\mathbf{c}$ represents the fourth-order elastic tensor. The formulation 2.1 is usually referred as strong form of the problem which is directly used in finite-difference and pseudospectral methods with boundary conditions in differential form. However the SEM seeks to solve the equation in a weak (or integral) formulation. The weak form of the problem is expressed by multiplying a test function $\mathbf{w}$ to equation 2.1, and then integrating over the computational domain $\Omega$.

$$\int_\Omega \rho(\mathbf{x})\mathbf{w}\cdot\frac{\partial^2\mathbf{u}}{\partial t^2}d^3\mathbf{x} = \int_\Omega \mathbf{w}\cdot(\boldsymbol{\nabla}\cdot\boldsymbol{\sigma})d^3\mathbf{x} + \int_\Omega \mathbf{w}\cdot\mathbf{f}d^3\mathbf{x}, \tag{2.2}$$

where the dependency of position and time are omitted for simplicity. The first term of the right-hand side of equation 2.2 contains the divergence of the stress tensor. By integrating by parts over the volume $\Omega$ and using the divergence theorem (Gauss-Ostrogradsky's theorem), we derive

$$\int_\Omega \rho(\mathbf{x})\mathbf{w}\cdot\frac{\partial^2\mathbf{u}}{\partial t^2}d^3\mathbf{x} = \int_\Gamma \mathbf{w}\cdot(\boldsymbol{\sigma}\cdot\hat{\mathbf{n}})d^2\mathbf{x} - \int_\Omega \boldsymbol{\nabla}\mathbf{w} : \boldsymbol{\sigma}d^3\mathbf{x} + \int_\Omega \mathbf{w}\cdot\mathbf{f}d^3\mathbf{x}, \tag{2.3}$$

where $\hat{\mathbf{n}}$ is the outward unit normal vector at the boundary $\Gamma$. The volume integral involving the stress tensor has been replaced by a surface integral of tractions on the edges of the domain $\mathbf{T} = \boldsymbol{\sigma} \cdot \hat{\mathbf{n}}$ and a volume integral of the gradient of the test function. The derivatives of the stress tensor are then transferred to the test functions $\mathbf{w}$ which is analytic, and then decreases the computational cost. The strong and the weak formulations are equivalent mathematically, however, in weak form we can see that the boundary condition is naturally included in the system through the surface integral. In particular, the free surface condition implies that the traction $\boldsymbol{\sigma}$ at the surface of the domain cancels out

$$\boldsymbol{\sigma} \cdot \hat{\mathbf{n}} = 0. \tag{2.4}$$

As a result, surface integral becomes zero at the surface of the domain. This simplicity is one of the key advantages of the SEM and allows to model accurately the surface waves, in particular when considering complex topographies.

In regional or local simulation, besides the free surface, we need to deal with internal boundaries that absorb the energy travelling out the volume. The most popular absorbing boundary condition is introduced by Clayton and Engquist (1977) and Stacey (1988) which is based on the paraxial approximations of the elastodynamics equations. The implementation of absorbing condition in SEM is presented by Komatitsch et al. (1998)

$$\boldsymbol{\sigma} \cdot \hat{\mathbf{n}} = \rho V_P (\hat{\mathbf{n}} \cdot \partial_t \mathbf{u}) \hat{\mathbf{n}} + \rho V_S (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \partial_t \mathbf{u}. \tag{2.5}$$

This absorbing condition works well as the wave propagating orthogonally to the edge, but less effective for oblique incidences. It then generate spurious reflections at grazing angles. Several alternative conditions are introduced, for example the sponge layer (Cerjan et al., 1985), the perfectly matched layer (PML) (Berenger, 1994; Komatitsch and Martin, 2007) or SMART layers (Tago et al., 2014). Nevertheless, the efficiency of these methods comes at the expense of computational cost and they are more difficult to implement.

*2.1.1.2 Spatial Discretization*

The SEM discretize the computational domain $\Omega$ into a finite number of elements $\Omega_e$ that indexed $e = 1, ..., N_e$. The physical properties or solutions are therefore represented by the union of these non-overlapping elements $\Omega = \cup_e^{N_e} \Omega_e$. In FEM several kinds of elements can be used such as tetrahedra, hexahedra, pyramids and prisms. In contrast, the SEM only favors the use of hexahedra (and quadrangles for the surface of the elements) which allow the 3D basis functions to be expressed as tensor product of the 1D basis functions (Komatitsch et al., 2000), resulting a diagonal mass matrix. These elements are deformed to fit the contours of the physical domain under consideration.

The shape of an element in physical Cartesian domain is formulated from mapping (deformation) of a reference cube. The mapping from the physical coordinate $\mathbf{x} = (x, y, z)$ inside a deformed element $\Omega_e$ to the reference local cube $\boldsymbol{\xi} = (\xi, \eta, \gamma)$ is expressed as

$$\mathbf{x}(\boldsymbol{\xi}) = \sum_{a=1}^{n_a} N_a(\boldsymbol{\xi}) \mathbf{x}_a, \tag{2.6}$$

where $n_a$ is the number of control points, $-1 \leq \xi \leq 1$, $-1 \leq \eta \leq 1$, and $-1 \leq \gamma \leq 1$. Each volume element is defined in terms of $n_a$ anchor points and shape functions $N_a(\boldsymbol{\xi})$. For a 3-D element, we need at least eight corner nodes to define a hexahedral element; if we consider additional mid anchor points, it becomes 27. The shape functions are calculated by triple products of Lagrange polynomials of degree 1 or 2 (Komatitsch and Tromp, 1999).

The volume of an element $dV = dxdydz$ can be written in terms of the coordinate of a reference cube by

$$dV = dxdydz = J_e d\xi d\eta d\gamma, \tag{2.7}$$

where $J_e$ the Jacobian of the mapping is given by

$$J_e = \left| \frac{\partial(x, y, z)}{\partial(\xi, \eta, \gamma)} \right|. \tag{2.8}$$

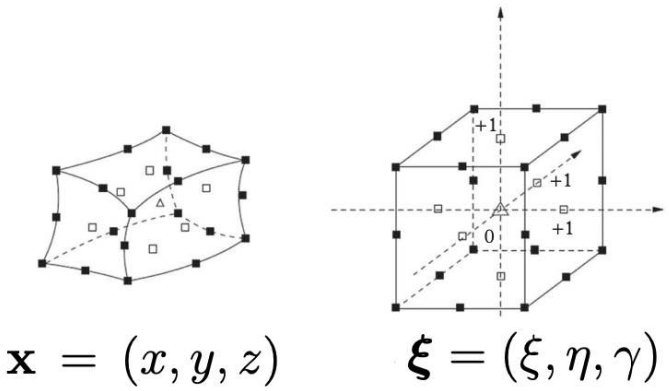


Figure 2.1: Transformation of one element between the physical coordinate $\mathbf{x}$ and the reference coordinate $\boldsymbol{\xi}$, which is related with a Jacobian matrix.

The Jacobian can be obtained by differentiating the mapping equation 2.6:

$$\frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} = \sum_{a=1}^{n_a} \frac{\partial N_a}{\partial \boldsymbol{\xi}} \mathbf{x}_a. \tag{2.9}$$

Since $N_a$ is expressed in terms of Lagrange polynomials, $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}}$ can be determined analytically.

Similar transformations mentioned above are also applied to the boundary elements. In this thesis, all the elements are built on hexahedrons with 27 anchor points and quadrangles with 9 points. The representation of the geometry of the element used in SEM is similar to that of FEM, by using low-degree polynomials.

#### *2.1.1.3 Polynomial representation and integration over one element*

As we are simulating wave propagation in heterogeneous Earth model with numerical methods, we need an appropriate discrete representations of our computational domain. The conversion from a continuous problem (usually in weak formulation) to a discrete problem can be achieved by the Galerkin method, which the exact solution $u(x,t)$ is approximated by a finite superposition of $n$ basis functions $\psi_i(x)$

$$u(x,t) \approx \overline{u}(x,t) = \sum_{i=1}^{n} u_i(t)\psi_i(x). \tag{2.10}$$

In SEM, the basis functions $\psi_i(x)$ are defined by a product of high-degree Lagrange polynomials. The Lagrange polynomials of degree $N$ are defined in terms of $N+1$ control points

$$\ell_i^{(N)}(\xi) = \prod_{j \neq i}^{N+1} \frac{\xi - \xi_j}{\xi_i - \xi_j} = \frac{\xi - \xi_0}{\xi_i - \xi_0} \cdot \frac{\xi - \xi_1}{\xi_i - \xi_1} \cdot \ldots \cdot \frac{\xi - \xi_N}{\xi_i - \xi_N} \cdot \frac{\xi - \xi_{N+1}}{\xi_i - \xi_{N+1}}, \tag{2.11}$$

where $\xi_i$ are defined in the closed interval $[-1, 1]$. Note that at specific points $\xi_j$, the value is zero; whereas at points $\xi_i$, the value is one. We can see that the Lagrange polynomials are equal to either zero or one at control points, which can be expressed as

$$\ell_i^{(N)}(\xi_j) = \delta_{ij}, \tag{2.12}$$

where $\delta_{ij}$ denotes the Kronecker symbol.

Now the problem becomes which collocation points to choose. The direct approach is to use equidistant points. However, by using polynomials of high degree over these equidistant points, the interpolation becomes oscillating and overshooting near the boundaries, which is the well-known Runge's phenomenon. To avoid this, it is natural to choose a set of points that contains more sample points towards the boundaries.

In the SEM, the control points $\xi_i$ are chosen to be the Gauss-Lobatto-Legendre (GLL) points, which are the roots of the first derivative of the Legendre polynomials of degree $N$. The use of GLL points and Lagrange polynomials not only minimizes the Runge phenomenon but is also a key factor in constructing a diagonal mass matrix. Figure 2.2a illustrates the 1-D spatial distribution of GLL points of degree $N = 2$ to 9 and Figure 2.2b shows 3-D sampling with order $N = 7$.

Now we express the function in terms of the Lagrange polynomials at the element level. In three dimensions, the function $f$ on one element are approximated (or interpolated) by triple products of Lagrange polynomials

$$f(\mathbf{x}(\xi, \eta, \gamma)) \approx \sum_{i=0}^{N_i} \sum_{j=0}^{N_j} \sum_{k=0}^{N_k} f^{ijk} \ell_i(\xi) \ell_j(\eta) \ell_k(\gamma), \tag{2.13}$$

where $f^{ijk} = f(\mathbf{x}(\xi_i, \eta_j, \gamma_k))$ denotes the value of $f$ at the GLL point $\mathbf{x}$. The superscript $(N)$ of the Lagrange polynomials is now omitted as the order is indicated by the summation limits. Note this expression implies that the method allows different degrees in each of the three directions. With equation 2.13, the gradient of the function $\boldsymbol{\nabla} f$ can be written in terms of the derivatives of the Lagrange polynomials

$$\begin{aligned}\boldsymbol{\nabla} f(\mathbf{x}(\xi,\eta,\gamma)) &\approx \sum_{\alpha=1}^{3} \hat{\mathbf{x}}_\alpha \partial_\alpha f(\mathbf{x}(\xi,\eta,\gamma)) \\ &= \sum_{\alpha=1}^{3} \hat{\mathbf{x}}_\alpha \sum_{i=0}^{N_i}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k} f^{ijk}[\ell_i'(\xi)\ell_j(\eta)\ell_k(\gamma)\partial_\alpha\xi \\ &\qquad + \ell_i(\xi)\ell_j'(\eta)\ell_k(\gamma)\partial_\alpha\eta \\ &\qquad + \ell_i(\xi)\ell_j(\eta)\ell_k'(\gamma)\partial_\alpha\gamma],\end{aligned} \tag{2.14}$$

where $\alpha$ represents the direction of $x, y, z$ respectively and the prime denotes the derivatives of Lagrange polynomials. The matrix $\frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}$ can be obtained by the inverse of Jacobian matrix (equation 2.9). In particular when the gradient is evaluated at any GLL points $\mathbf{x}(\xi_i^G, \eta_j^G, \gamma_k^G)$, with the relation $\partial_\xi \ell_i(\xi_i) = 0$ the expression reduces to

$$\boldsymbol{\nabla} f(\mathbf{x}(\xi,\eta,\gamma)) \approx \sum_{\alpha=1}^{3} \hat{\mathbf{x}}_\alpha \left[\sum_{i=0}^{N_i} f^{ij^G k^G}\ell_i'(\xi_i^G)\partial_\alpha\xi + \sum_{j=0}^{N_j} f^{i^G j k^G}\ell_j'(\eta_j^G)\partial_\alpha\eta + \sum_{k=0}^{N_k} f^{i^G j^G k}\ell_k'(\gamma_k^G)\partial_\alpha\gamma\right]. \tag{2.15}$$

Now we have expressed the approximated solution inside one element.

#### *2.1.1.4 Numerical integration of the weak form of the wave equation*

To solve the weak form of the equation of motion one needs to perform numerical integrations over the elements. In FEM context, one frequently uses Gauss quadrature. In SEM, one uses Gauss-Lobbatto-Legendre quadrature together with GLL points. Using GLL quadrature, the integration over a volume element $\Omega_e$ for an arbitrary function $f$ defined in

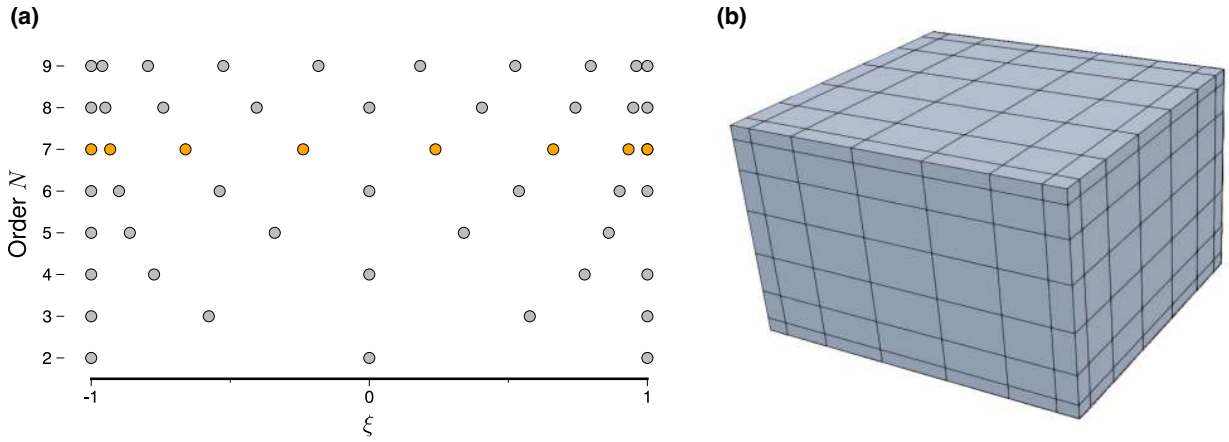


Figure 2.2: GLL points. (a) Illustration of their spatial distribution in the interval [-1,1] for polynomial order N=2 to N=9 corresponding to Np =N+1 collocation points. In this thesis we use order of 7, denoted with orange. (b) Discretization on one spectral element.

the interval $[-1, 1]$ can be expressed as

$$\begin{aligned} \int_{\Omega_e} f(\mathbf{x}) d\mathbf{x}^3 &= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} f(\mathbf{x}(\xi,\eta,\gamma)) J_e(\xi,\eta,\gamma) d\xi d\eta d\gamma \\ &\approx \sum_{i=0}^{N_i}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k} \omega_i\omega_j\omega_k f^{ijk} J_e^{ijk}, \end{aligned} \tag{2.16}$$

where $\omega_e$ denotes the weights of the GLL quadrature, $J_e^{ijk} = J_e(\xi_i, \eta_j, \gamma_k)$ the Jacobian at GLL point $\mathbf{x}(\xi_i, \eta_j, \gamma_k)$. The integration weights $\omega_i$ for the $i$-th root are calculated with

$$\omega_i = \frac{2}{(1 - x_i)[P_n'(x_i)]^2}, \tag{2.17}$$

where $P_n'$ denotes the derivative of the Legendre polynomial of degree $n$. From equation 2.16, we see that using polynomial representation of function, the numerical error in spatial domain only depends on the integration scheme we choose. One can see that the error of integral is reduced when high order of polynomial degrees is considered.

**Mass matrix** To solve the weak formulation of the elastodynamic equation 2.3, one need to evaluate the integrals numerically over each element. As shown in previous section, in SEM, the integration is performed using GLL quadrature in conjunction with GLL points.

Therefore, the displacement $\mathbf{u}$ and the test function $\mathbf{w}$ are expressed as Lagrange polynomials, using equation 2.13.

The left term of equation 2.3 corresponds to the mass matrix. At element level, using the GLL integration scheme (equation 2.16) the integration can be written

$$\begin{aligned}\int_{\Omega_e} \rho(\mathbf{x})\mathbf{w}\cdot\frac{\partial^2\mathbf{u}}{\partial t^2}d^3\mathbf{x} &= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}\rho(\mathbf{x}(\boldsymbol{\xi}))\mathbf{w}(\mathbf{x}(\boldsymbol{\xi}))\cdot\frac{\partial^2\mathbf{u}}{\partial t^2}J_e(\boldsymbol{\xi})d^3\boldsymbol{\xi} \\ &\approx \sum_{i=0}^{N_i}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k}\omega_i\omega_j\omega_k J_e^{ijk}\rho^{ijk}\sum_{\alpha=1}^{3}w_\alpha^{ijk}\ddot{u}_\alpha^{ijk}(t) \\ &\approx \mathbf{M}\ddot{\mathbf{u}},\end{aligned} \tag{2.18}$$

where $\mathbf{M}$ is the mass matrix and the double dot denotes double differentiation with respect to time.

We can see that the value of acceleration $\ddot{u}_i$ at grid point $\boldsymbol{\xi}$ is simply weighted by the factor $\omega_i\omega_j\omega_k J_e^{ijk}\rho^{ijk}$. This indicates that by construction the mass matrix $\mathbf{M}$ is diagonal because it only demands the corresponding value on the same sample point. This comes from the fact that the Gauss-Lobatto-Legendre quadrature is only evaluated at the GLL points.

**Stiffness matrix** The second term in the right-hand side of the elastodynamic function. For convinence we first determine the expression of $\boldsymbol{\nabla}\mathbf{w}$ and $\boldsymbol{\sigma}$ in the reference coordinate,

$$\boldsymbol{\nabla}\mathbf{w} : \boldsymbol{\sigma} = \sum_{a,b=1}^{3}\sigma_{ab}\partial_b w_a = \sum_{a,c=1}^{3}\left(\sum_{b=1}^{3}\sigma_{ab}\partial_b\boldsymbol{\xi}_c\right)\frac{\partial w_a}{\partial\boldsymbol{\xi}_c} = \sum_{a,c=1}^{3}F_{ac}\frac{\partial w_a}{\partial\boldsymbol{\xi}_c}, \tag{2.19}$$

where $F_{ac} = \sum_{b=1}^{3}\sigma_{ab}\partial_b\boldsymbol{\xi}_c$. To evaluate $F$ at integration point $\boldsymbol{\xi}$ we apply Hooke's law

$$\boldsymbol{\sigma}(\boldsymbol{\xi},t) = \mathbf{c}(\boldsymbol{\xi}) : \boldsymbol{\epsilon}(\boldsymbol{\xi},t) \tag{2.20}$$

where $\mathbf{c}$ is the stiffness tensor that describes the properties of the medium and $\boldsymbol{\epsilon}$ the strain tensor that can be derived by the gradient of displacement. Note that the structure of $\mathbf{c}$

has no particular formulation, therefore besides elastic cases, it can also handle either an anisotropic media (Komatitsch et al., 2000) or a poroelastic media (Boxberg et al., 2015).

We can now integrate numerically using Lagrange interpolation and GLL quadrature:

$$\begin{aligned}\int_{\Omega_e} \boldsymbol{\nabla}\mathbf{w} : \boldsymbol{\sigma} d\mathbf{x}^3 &= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}\left(\sum_{a,c=1}^{3} F_{ac}\frac{\partial w_a}{\partial \boldsymbol{\xi}_c}\right) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}^3 \\ &\approx \sum_{a=1}^{3} w_a^{ijk}[\omega_j\omega_k \sum_{i=0}^{N_i} \omega_i J^{i^G jk} F_{i1}^{i^G jk} \ell_i'(\xi_i^G) \\ &\qquad + \omega_i\omega_k \sum_{j=0}^{N_j} \omega_j J^{ij^G k} F_{i2}^{ij^G k} \ell_j'(\eta_j^G) \\ &\qquad + \omega_i\omega_j \sum_{k=0}^{N_k} \omega_k J^{ijk^G} F_{i3}^{ijk^G} \ell_k'(\gamma_k^G)] \\ &= \mathbf{K}\mathbf{u},\end{aligned} \tag{2.21}$$

where $\mathbf{K}$ is the stiffness matrix. Unlike the mass matrix, the stiffness matrix is full.

**Point force** The integration of the source term for a point force $\mathbf{f}(\mathbf{x}_s)$ on the element $\Omega_s$ is given directly by :

$$\begin{aligned}\int_{\Omega_s} \mathbf{w}\cdot\mathbf{f} d\mathbf{x}^3 &\approx \sum_{\alpha,\beta,\gamma=0}^{N} \ell_\alpha(\boldsymbol{\xi}(\mathbf{x}_s))\ell_\beta(\boldsymbol{\xi}(\mathbf{x}_s))\ell_\gamma(\boldsymbol{\xi}(\mathbf{x}_s)) f_i \\ &= \mathbf{f},\end{aligned} \tag{2.22}$$

where $\mathbf{f}$ is the force vector. The source term would be more complicated when considering a complete moment tensor (Komatitsch and Tromp, 1999).

**Boundary condition** The boundary conditions is evaluated by a surface integral

$$\begin{aligned}\int_{\Gamma} \mathbf{w}\cdot(\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}) &\approx \sum_{\alpha,\beta=0}^{N} \omega_\alpha\omega_\beta J^{\alpha\beta} t_i^{\alpha\beta} \\ &= \mathbf{C}\dot{\mathbf{u}},\end{aligned} \tag{2.23}$$

where $t_i^{\alpha\beta}$ is the traction on the edge which is zero in case of a free surface or the value given by the absorbing condition 2.5. The discretization of the absorbing condition results in the damping matrix $\mathbf{C}$.

**Time Marching** Using the previous matrix formulation of integration at the element level, we can extend it to the entire mesh and rewrite the weak form of the elastodynamic equation in a matrix discrete form

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u} = \mathbf{f}(t). \tag{2.24}$$

The Newmark scheme is used to solve the differential system 2.24, which is composed of 3 steps. The predictor consists in using a Taylor expansion to update the displacement and velocity. For a given displacement $\mathbf{u}^n$, velocity $\dot{\mathbf{u}}^n$ and acceleration $\ddot{\mathbf{u}}^n$ at step time $n\Delta t$, we can write the displacement and velocity for a new the time step,

$$\begin{aligned} \mathbf{u}^{n+1} &= \mathbf{u}^n + \Delta t\dot{\mathbf{u}}^n + \tfrac{\Delta t^2}{2}\ddot{\mathbf{u}}^n, \\ \dot{\mathbf{u}}^{n+1/2} &= \dot{\mathbf{u}}^n + \tfrac{\Delta t}{2}\ddot{\mathbf{u}}^n. \end{aligned} \tag{2.25}$$

The second steps involves the wave equation, where the acceleration is expressed as the sum of forces acting on the medium,

$$\ddot{\mathbf{u}}^{n+1} = \mathbf{M}^{-1}(\mathbf{f}^{n+1} - \mathbf{C}\dot{\mathbf{u}}^{n+1/2} - \mathbf{K}\mathbf{u}^{n+1}), \tag{2.26}$$

and then the corrector consists in velocity update at $n+1$ time step using a Taylor expansion,

$$\dot{\mathbf{u}}^{n+1} = \dot{\mathbf{u}}^{n+1/2} + \frac{\Delta t}{2}\ddot{\mathbf{u}}^{n+1}. \tag{2.27}$$

Repeating the above procedure, we can compute the fields at the next time step $(n+2)\Delta t$ and so on. Note that in the relation 2.26 it is necessary to invert the mass matrix $\mathbf{M}$. The choices made in SEM leads to a diagonal mass matrix, which greatly simplifies this step

by making the scheme explicit in time. But this time scheme is conditionally stable which constrains the choice of the time step $\Delta t$ by the Courant-Friedrich-Levy (CFL) condition,

$$\Delta t < c\frac{d_{min}}{v_{max}}, \tag{2.28}$$

where $d_{min}$ is the minimal distance between two discretization points (GLL) and $v_{max}$ is the maximum wave speed in the element. In the SPECFEM3D code the value of $c$ is set to 0.5.

**Accuracy** The spatial discretization requires a particular setting to solve accurately the wave equation. The size of the elements is controlled by the minimum wavelength and the polynomial order used. According to Lyu et al. (2020), 5 points per wavelength are necessary to reach an error less than 1% on the seismograms for degree equal to 4 and 4 points for a degree equal to 7. This means that one element per wavelength is needed for 5 discretization points and 2 wavelengths per element for 8 discretization points.

### 2.1.2 Modeling teleseismic wavefields with hybrid methods

A teleseismic wave is generated by an earthquake whose hypocenter is located at a distance between 30 and 120 degree of epicentral distance. The exploitation of these teleseismic waves recorded by a dense regional seismic network can be used to extract some information on the physical properties of the underlying medium. Teleseismic wavefields include the direct P and S waves but also many secondary phases that have interacted with the different internal discontinuities of the Earth. The interfaces that have the most significant effect on the waves are the free surface, the Moho, and the core-mantle boundary. In addition, other interfaces with lower impedance contrast such as the discontinuties at 410 and 660 km depth can also produce conversions and reflections. The reflections of seismic waves in the vicinity or the earthquake source have a significant impact on seismic records. For example, the reflections on the free surface produce seismic phases that arrive just after the direct P phase (pP, sP, ...). Depending on the hypocentral depth, epicentral distance, and focal mechanism, these secondary phases are more or less energetic. One of the challenges of the

modeling method is to correctly take into account these phases. Moreover, on the receiver side, the free surface and the Moho produce converted and/or reflected waves that arrive just after the direct P (Ps, PmP,...) or S (Sp, SmS, ...) waves. These conversions are used in the receiver function method to image lithospheric discontinuities such as the Moho. In travel time tomography, only the travel time of the direct waves are used to invert velocity perturbations in the medium. In Full Waveform Inversion method, the objective is to use all the information contained in seismic records. This will allow to image both the velocity perturbations as well as the different interfaces in the crust and the upper mantle under the array.

In order to implement a teleseismic imaging method based on the exploitation of waveforms, we must be able to model the complete seismic wavefields, taking into account all the complexities of the medium that can affect the propagation of seismic waves such as the topography of the free surface or internal discontinuities such as the Moho, the heterogeneities of velocity and density, etc.... We must thus use numerical methods solving the elastodynamic wave equation. The calculations must be performed with minimum periods of $\sim$1 s to ensure sufficient resolution of the inverted model, which makes waveform inversion a challenging numerical problem.

In principle, any 3D global-scale simulation code could solve the problem of modeling teleseismic waves. For example, the SPECFEM3D_GLOBE code is used for global scale imaging with a minimum period of 17 s (Bozdağ et al., 2016). With our goal being to go down to a few seconds of periods, it is clearly not feasible to use such a global solver. Indeed, the computation time of this type of 3D solver scales in power of 4 (space of dimension 3 plus time) of the minimum period. This means that going from 17 s to 5 s (still above from our minimum period target) requires 2 orders of magnitude more computing power. In addition, short period modeling tests have been performed with SPECFEM3D_GLOBE on the K computer in Japan. It took mobilizing 80% of the largest computer available in the world in 2015 to calculate a few minutes of seismograms at 1 s (Tsuboi et al., 2016).

Even with the increase in computing resources we are experiencing today, such global scale simulations are still exceptions.

To overcome this important computational bottleneck, several methods have been proposed. One method, developed in recent years, is increasingly used. It consists in injecting the teleseismic wave field from the boundaries of a regional domain. The modeling is done in 2 steps. The first is to calculate the global incident field and store it on the edges of the regional mesh. This step is performed once per earthquake. This allows to inject in a second time the incident wave field in a regional domain in which we model the 3D wave propagation. This general principle can be applied with different numerical methods to simulate the global and regional wave propagation. The first example of this method can be found in Chevrot et al. (2004), where a plane shear wave is injected into a local solver based on spectral elements. Later, Monteiller et al. (2013) developed a coupling between the DSM method at the global scale in a spherically symmetric earth and the regional SPECFEM3D_CARTESIAN code. The advantage here is to be able to model any teleseismic phase taking into account the depth, the epicentral distance and the focal mechanism of the earthquake. Subsequently, Masson et al. (2014) has modified this injection principle so that it can be extended to other numerical methods such as finite differences. In parallel, Tong et al. (2014a) have generalized plane wave injection using a frequency-wave number (FK) code. This approach approximates the injected incident wave field as a plane wave solution in a layered model. The different reflections and conversions of the waves at the plane interfaces are taken into account in the injected solution. This strategy of injecting a teleseismic field in a 3D regional model allows to drastically reduce the computation time compared to the 3D global approach. The size of the computational domain is reduced by 3 to 4 orders of magnitude (volume of the global earth compared to the volume of the regional domain). Moreover, the duration of the simulations is also reduced because we can just model the temporal window around the chosen phase and not systematically start the execution from the origin time of the earthquake.

However, this injection method is not exact, in the sense that the solution calculated in

the regional domain by injection is not equal to the global 3D solution. This is due to the fact that only heterogeneities located inside the regional domain are taken into account in the 3D computations. The diffracted field outside the regional domain is not taken into account. The contribution of double diffractions with a scatterer outside the regional domain are neglected. To overcome this difficulty, some authors have developed true coupling methods. For example, Capdeville et al. (2003) has proposed a coupling method between the normal mode summation and the spectral element method (SEM). However, this method is not suitable for telesimic waveform inversion because of the difficulty in computing synthetic seismograms for periods less than 8 s by normal mode summation. Another general method of code coupling based on the representation theorem and Kirchoff integrals (Broggini et al., 2017) has been developed in 2D. However, this approach remains too expensive in 3D due to an extrapolation step that allows the wave field to be propagated outside the local domain to allow it to interact with the medium in which it is embedded. At present, injection methods are thus still the most widely used and have largely proven their effectiveness for teleseismic imaging despite some simplifying assumptions. More recently, Pienkowska et al. (2020) compared 3D global solutions (SPECFEM3D_GLOBE) with the injection method and showed that teleseismic body waves (P and S) as well as their coda remain well modeled by the injection methods. However, this is not necessarily the case for surface waves.

### 2.1.3 Practical implementation of the Spectral Element Method

An important foundation of this thesis is the ability to efficiently compute short-period teleseismic wave fields, with minimum periods typically on the order of 1 to 4 s. As already stated, the SEM itself is capable of simulating wave propagation at a minimum period close to one second at the global scale (Komatitsch and Tromp, 2002), however the large number of simulations required for iterative FWI makes this approach unaffordable at this time. Limiting the computational domain to a regional box with a hybrid modeling method thus remains a good compromise between the computational cost and the targeted high frequencies.

When using the hybrid method to propagate the external wavefield into the regional domain $\Omega$, there is no internal force to be applied but the change of tractions over time on the boundary. The weak form of the elastodynamic equation can be written

$$\int_\Omega \rho(\mathbf{x})\mathbf{w}\cdot\frac{\partial^2\mathbf{u}}{\partial t^2}d^3\mathbf{x} = \int_\Gamma \mathbf{w}\cdot\mathbf{T}d^2\mathbf{x} - \int_\Omega \boldsymbol{\nabla}\mathbf{w} : \boldsymbol{\sigma} d^3\mathbf{x}. \tag{2.29}$$

With the injection of an external wavefield $\mathbf{u}^i$, the total displacement field $\mathbf{u}$ can be decomposed into the sum of the incident wavefield $\mathbf{u}^i$ in and the scattered wavefield $\mathbf{u}^s$ generated by the heterogeneities inside the medium

$$\mathbf{u} = \mathbf{u}^i + \mathbf{u}^s. \tag{2.30}$$

Similarly, the total tractions $\mathbf{T}$ on the boundary $\Gamma$ is expressed

$$\mathbf{T} = \mathbf{T}^i + \mathbf{T}^s. \tag{2.31}$$

At the boundaries, we want to suppress only the outward scattered field (Bielak and Christiano, 1984; Monteiller et al., 2013; Tong et al., 2014b)

$$\mathbf{T}^s = \rho V_P(\hat{\mathbf{n}}\cdot\partial_t(\mathbf{u}-\mathbf{u}^i))\hat{\mathbf{n}} + \rho V_S(\mathbf{I}-\hat{\mathbf{n}}\hat{\mathbf{n}})\cdot\partial_t(\mathbf{u}-\mathbf{u}^i). \tag{2.32}$$

From equation 2.31 and 2.32 we see that the incident wavefield is naturally included into the SEM given the knowledge of external displacement filed $\mathbf{u}^i$ and tractions $\mathbf{T}^i$ on the boundary $\Gamma$. Both fields can be pre-computed by any external method. Currently there are three implementation methods: coupling with the AxiSEM (Nissen-Meyer et al., 2014; Monteiller et al., 2021), the direct solution method (DSM) (e.g. Kawai et al., 2006; Monteiller et al., 2013), and the frequency-wavenumber method (FK) (e.g. Zhu and Rivera, 2002; Tong et al., 2014a,b) methods. We briefly introduce them below.

**FK** The frequency-wavenumber method computes the response of a layered 1D medium for planar P or S waves, following the Thompson-Haskell propagator matrix approach (e.g. Thomson, 1950; Haskell, 1953; Takeuchi and Saito, 1972). The plane wavefronts are injected into the half-space under a stack of $n$ homogeneous layers. The detailed description of the computation of the velocities and tractions with the SEM-FK hybrid method is given by Tong et al. (2014a,b).

The advantage of the FK method is its efficiency compared to other numerical methods. On the other hand, its limitation is that it only models the propagation of a single incident plane wave. This prevents from taking into account reflections in the vicinity of the teleseismic source (e.g. pP and sP waves) or secondary phases that could arrive after the main phase (e.g. PcP or PP waves). For a teleseismic event, the plane wave approximation remains valid if the regional domain is limited to a few hundred kilometers in size. In this case the curvature of the wavefront and the Earth have a negligible effect on the surface wave field (Monteiller et al., 2021). A recent study by Wang et al. (2021) generalized the coupling approach with FK to model secondary arrivals that arrive after the direct P or S wave. However, this approach has an application domain that remains limited by the size of the regional mesh.

**DSM** The direct solution method (DSM), as the name implies, directly solves the weak form of the equation of motion in the frequency domain (e.g. Geller and Ohminato, 1994; Kawai et al., 2006). The displacement field (in the frequency domain) is expressed as a series of basis functions which are spline functions in the vertical direction and spherical harmonics in the horizontal directions. Once the coefficients of the basis functions are calculated, the solution of the wave equation in the frequency domain is obtained. The solution in the time domain is then computed with an inverse Fast Fourier Transform (FFT). By carefully choosing the grid spacing, maximum angular order, and cutoff depth, the DSM calculations can be accurate and efficient down to a frequency of 2 Hz (Kawai et al., 2006). The calculation of velocities and tractions with the hybrid SEM-DSM method is described in Monteiller et al.

(2013).

Although DSM is capable of simulating synthetic seismograms for short-period teleseismic body waves, these calculations are very demanding for a shallow source (Kawai et al., 2006). Indeed, for shallow earthquakes, it is necessary to use a very high maximum angular order to account for near-field terms. Since most earthquakes occur at shallow depths, this method remains expensive and difficult to implement in practice.

**AxiSEM** The axisymmetric spectral element method (AxiSEM) is an SEM implementation that models global wave propagation in a spherically symmetric earth model. The efficiency of this method consists of solving the equation of motion in a disk, recombining different components to reconstruct the three-dimensional wavefield (Nissen-Meyer et al., 2014). Under this axisymmetric approximation, where the symmetry axis passes through the center of the Earth and the source, one can analytically decompose the full 3D responses by combining solutions for monopole, dipole, and quadrupole source radiation patterns. As a result, the three-dimensional volumetric integral of the elastodynamic equation reduces to a problem expressed on a two-dimensional disk, which significantly reduces the dimension of the problem. Numerical validation of the SEM-AxiSEM coupling has been presented in Monteiller et al. (2021).

The 3-D response of a moment tensor with azimuthal dependency $\phi$ is

$$\mathbf{u}_m(\mathbf{x}) = \begin{bmatrix} u_s(\tilde{\mathbf{x}} \cos m\phi) \\ u_\phi(\tilde{\mathbf{x}} \sin m\phi) \\ u_z(\tilde{\mathbf{x}} \cos m\phi) \end{bmatrix}, \tag{2.33}$$

where $m = 0$ is for monopole, 1 for dipole, 2 for quadrupole radiation patterns, respectively and the tilde represents quantities evaluated in two-dimensional cylindrical $(s, z)$ or spherical $(r, \phi)$ coordinates.

In this axisymmetric approach, four independent types of radiation patterns are required

to describe the full response of a moment tensor source

$$\begin{aligned}
\tilde{M}_1(\phi) &= M_{rr} \\
\tilde{M}_2(\phi) &= (M_{\theta\theta} + M_{\phi\phi})/2 \\
\tilde{M}_3(\phi) &= M_{r\theta}\cos\phi + M_{r\theta}\sin\phi \\
\tilde{M}_4(\phi) &= (M_{\theta\theta} - M_{\phi\phi})\cos 2\phi + M_{\phi\theta}\sin 2\phi,
\end{aligned} \tag{2.34}$$

where $\tilde{M}_1(\phi)$ and $\tilde{M}_2(\phi)$ are monopoles, $\tilde{M}_3(\phi)$ dipole and $\tilde{M}_4(\phi)$ quadrupole.

In the post-processing step, the full response of a complete moment tensor is expressed as the summation of four independent 2-D simulations

$$\mathbf{u}(\mathbf{x}, t) = \sum_{m=1}^{4} \tilde{M}_m(\phi, t)\mathbf{G}_m(\tilde{\mathbf{x}}, t)S(t), \tag{2.35}$$

where $\mathbf{G}_m$ is the 2-D Green's function of each simulation $m$ and $S$ is the source time function. Note that in AxiSEM the source wavelet is imposed at the source position, and that the far-field displacement is proportional to the derivative of the displacement at the source. Therefore, to obtain a Gaussian wavelet in the far-field, we need to apply the error function at the source

$$S(t) = \mathrm{erf}\left(\frac{\alpha(t - t_0)}{\tau_0}\right), \tag{2.36}$$

so that the derivative of $S$ is the Gaussian function

$$\dot{S} = \frac{2}{\sqrt{\pi}}\exp\left(-\frac{\alpha^2(t - t_0)^2}{\tau_0^2}\right), \tag{2.37}$$

where $\alpha$ is chosen to be 3.5 so that $\tau_0$ corresponds to the width of the Gaussian function.

#### *2.1.3.1 Sampling rates*

Since the incident wavefield is calculated with AxiSEM, in which the sampling rate $dt_{\mathrm{AxiSEM}}$ is not always the same as that used in the SEM $dt_{\mathrm{SEM}}$. This means that we

have to interpolate to resample from $dt_{\text{AxiSEM}}$ to $dt_{\text{SEM}}$ on the edges of the SEM domain. However, storing the incident field for all the nodes on the edges of the SEM domain and for each time step remains quite prohibitive. The time step required for the stability of an AxiSEM simulation is much smaller than that required by the Nyquist condition. Therefore, by downsampling the AxiSEM solutions at the mesh boundaries to the Nyquist frequency (Unser, 2000; Monteiller et al., 2021), it is possible to reduce the storage demand by one to two orders of magnitude. These solutions are then oversampled when injecting the wavefield into the SEM.

The previous implementations in Beller (2017) used a sinc interpolation which requires all the samples of the subsampled solutions for the reconstitution of the solutions at each interpolated step. Note that this operation is computationally expensive, not to mention convolving the apparent source wavelet at each edge grid point. Instead, it is possible to convolve the source wavelet with the displacement calculated at the station after the SEM simulation.

An alternative approach adopted in this thesis is to use cubic splines which are cheaper for the injection step. In practice, we first resample the source wavelet to $dt_{\text{AxiSEM}}$ and then convolve it with the tractions and velocities in the Fourier domain. Thus, the synthetic tractions and velocities already contain the source signature.

Finally, we note that the data recorded at the seismic stations have their own sample rate $dt_{\text{data}}$. One need to resample $dt_{\text{data}}$ to $dt_{\text{SEM}}$ so that the forward synthetics and the data have the same dimension for the construction of the adjoint source.

## 2.2 Inverse Problem

An inverse problem is a process in which one tries to determine the causes of a phenomenon from a set of observations. Solving the inverse problem can provide parameters that we cannot directly observe. An example is to locate the position of an earthquake from measurements of seismic stations spread over the surface of the earth. The link between the causes (parameters) and the effects (data) is described as a physical modeling $\mathbf{G}$ which in

general is written as the direct problem

$$\mathbf{d}_{\text{obs}} = \mathbf{g}(\mathbf{m}). \tag{2.38}$$

In the case of modeling elastic wave propagation as presented in Section 2.1, $\mathbf{g}$ represents the equations of elastodynamics, $\mathbf{d}_{\text{obs}}$ is the displacement and $\mathbf{m}$ the model parameters. Unlike the inverse problem, the direct problem predicts the outcome with knowledges of the causes.

Usually in forward problem, the model (cause) is well-posed so that the prediction is unique. Solving the forward problem 2.38 allows us to predict the displacement field in the entire domain from the knowledge of the elastic properties of the medium. In reality, we have no access to the entire seismic wavefield but only a limited number of traces recorded by stations located on the surface. This causes the inverse problem usually being ill-posed, and the existence, uniqueness and stabibility of the solution being not guaranteed. Moreover, the forward operator $\mathbf{g}$ is non-linear in context of wave propagation, which means there is no simple matrix manipulations to invert for the model.

Note the forward problem 2.38 is in the form of error-free, i.e. no measurement uncertainties and modelization imperfections are included. When explicitly considering these uncertainties, each term in equation 2.38 is rewritten in probabilistic formulation (Tarantola, 2005). By using the Bayes theorem, the conditional probability (or posterior probability density) of a model $\mathbf{m}$ given a dataset $\mathbf{d}$ is

$$\rho(\mathbf{m}|\mathbf{d}) = \frac{\rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})}{\rho(\mathbf{d})}, \tag{2.39}$$

where $\rho(\mathbf{d}|\mathbf{m})$ is the probability density of $\mathbf{d}$ given the model $\mathbf{m}$ and $\rho(\mathbf{m})$ is the probability density of model parameters. Given a particular realization of the observed data $\mathbf{d}^{\text{obs}}$, the probability density $\rho(\mathbf{d}_{\text{obs}})$ is constant and therefore

$$\rho(\mathbf{m}|\mathbf{d}_{\text{obs}}) \propto \rho(\mathbf{d}_{\text{obs}}|\mathbf{m})\rho(\mathbf{m}). \tag{2.40}$$

Assuming there is no modeling error (i.e. $\mathbf{d} = \mathbf{g}(\mathbf{m})$) and the data uncertainties are represented by a Gaussian probability density centered at $\mathbf{d}_{\text{obs}}$ and with covariance matrix $\mathbf{C_D}$:

$$\rho(\mathbf{d}_{\text{obs}}|\mathbf{m}) = \text{const.} \times \exp\Big(-\frac{1}{2}(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))\Big). \tag{2.41}$$

Note that actually $\mathbf{C}_\text{D}$ is the combination of the covariance of observational uncertainties and modelization uncertainties (Tarantola, 2005). In our case the latter ones from SEM is comparable smaller than the former ones, thus we safely assume the latter ones to be zero. Therefore $\mathbf{C}_\text{D}$ simply represents the data covariance.

Similarly, it is also possible to describe prior information using probability densities. Assuming the a priori information on model parameters is also Gaussian, then

$$\rho(\mathbf{m}) = \text{const.} \times \exp\Big(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_{\text{prior}})^t \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}})\Big). \tag{2.42}$$

Combining the three types of information from equation 2.40, 2.41 and 2.42, we seek to find the point in the model space that maximize the posterior probability density $\rho(\mathbf{m}|\mathbf{d})$

$$\rho(\mathbf{m}|\mathbf{d}) = \text{const.} \exp(-\chi(\mathbf{m})), \tag{2.43}$$

which is equivalent to minimizing the misfit function $\chi(\mathbf{m})$ that is defined

$$\begin{aligned}
\chi(\mathbf{m}) &= \frac{1}{2}||\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})||_\text{D}^2 + \frac{1}{2}||\mathbf{m} - \mathbf{m}_{\text{prior}}||_\text{M}^2 \\
&= \frac{1}{2}(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) + \frac{1}{2}(\mathbf{m} - \mathbf{m}_{\text{prior}})^t \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}) \\
&= \mathbf{J}(\mathbf{m}) + \mathbf{R}(\mathbf{m}).
\end{aligned} \tag{2.44}$$

The misfit function $\chi(\mathbf{m})$ is sum of squares, which justifies the use of terminology "least-squares" for the problems based on Gaussian assumption. The function is expressed as the sum of two terms. The first term represents the data residual, which measures the degree of similarity between the synthetics and the actual data. The second term is the regularization

term, which ensures that the model does not exhibit excessive, non-physical fluctuations.

When the dimensionality of the model space is low (less than a few thousand parameters), stochastic methods (e.g. Monte-Carlo, simulated annealing) can be used to explore $\rho(\mathbf{m}|\mathbf{d})$. However, for the waveform inversion problems, with over millions of free parameters, it is impractical to explore the entire model domain. Instead, we consider the optimization method that finds the most probable model that minimizes the misfit function $\chi(\mathbf{m})$ (Tarantola, 2005).

### 2.2.1 Gradient based local optimization

To minimize the function $\chi(\mathbf{m})$, we use local methods with descent direction based on the gradient of the cost function. Since it is a nonlinear problem, we need to solve it iteratively. If we perturb the model by $\delta\mathbf{m}$), the new cost value can be expanded into the Taylor series

$$\chi(\mathbf{m}+\delta\mathbf{m}) \approx \chi(\mathbf{m}) + \frac{\partial\chi(\mathbf{m})}{\partial\mathbf{m}} \cdot \delta\mathbf{m} + \frac{\partial^2\chi(\mathbf{m})}{\partial\mathbf{m}^2} \cdot \delta\mathbf{m}^2. \tag{2.45}$$

This expression can be reformulated as

$$\chi(\mathbf{m}+\delta\mathbf{m}) \approx \chi(\mathbf{m}) + \nabla\chi(\mathbf{m}) \cdot \delta\mathbf{m} + \mathbf{H}(\mathbf{m}) \cdot \delta\mathbf{m}^2, \tag{2.46}$$

where $\nabla\chi(\mathbf{m})$ is the gradient of the misfit function (the first model derivative of the misfit function), and $\mathbf{H}(\mathbf{m})$ is the Hessian (the second model derivative of the misfit function). The differentiation of equation 2.46, when we neglect the 3th derivative terms, is:

$$\nabla\chi(\mathbf{m}+\delta\mathbf{m}) = \nabla\chi(\mathbf{m}) + \mathbf{H}(\mathbf{m}) \cdot \delta\mathbf{m}. \tag{2.47}$$

When a critical point like minimun, maximun or saddle point of the functional $\chi(\mathbf{m}+\delta\mathbf{m})$ is reached, the left hand side of the first derivative 2.47 will become zero. Therefore we get

$$\delta\mathbf{m} = -\mathbf{H}^{-1}(\mathbf{m})\nabla\chi(\mathbf{m}). \tag{2.48}$$

This formulation corresponds to Newton's equation which has a quadratic convergence. The model perturbation $\delta\mathbf{m}$ is determined by the gradient and the inverse of the Hessian matrix of the cost function.

The dimensions of the Hessian operator being the number of parameters squared, for a 3D FWI problem it is not feasible to store and invert such a matrix. To circumvent the problem, Métivier et al. (2013) have proposed a truncated Newton method which consists in solving this linear system in an approximate way by performing some conjugate gradient iterations. So far, this method has been used in 2D and in the frequency domain. For the moment, it is not possible to use it in the time domain and in 3D, for reasons of computational cost. Instead of computing the full Hessian, one can use a quasi-Newton approximation, which requires less computational resources and converges faster than the simple steepest descent method.

The problem is then formulated as a two-step problem, first computing the pertubation with the approximate Hessian, and then computing the step length along the descent direction

$$\delta\mathbf{m} = \alpha\mathbf{p}, \tag{2.49}$$

where $\alpha$ is the step length and $\mathbf{p}$ the descent direction, given by

$$\mathbf{p} = -\tilde{\mathbf{H}}^{-1}\nabla\chi(\mathbf{m}). \tag{2.50}$$

This formula implies that the search direction is determined by the gradient of the misfit function, preconditioned by the inverse of approximate Hessian $\tilde{\mathbf{H}}$. Nocedal and Wright (2006) introduced the l-BFGS method which is a numerically efficient method for computing a Hessian approximation based on the previously computed and stored model and gradient history. However, it is necessary to calculate the step length at each iteration because the Hessian approximation, in general, does not give correct physical scales. The l-BFGS method is effective when the step length follows Wolfe's rules, which can be used in a linear search

method.

#### *2.2.1.1 Optimization method: l-BFGS algorithm*

The BFGS method is one of the quasi-Newton method that computes the approximate inverse Hessian. The descent direction is expressed similar to the one in Newton's method:

$$\mathbf{p}_k = -\mathbf{H}_k^{-1} \nabla \chi(\mathbf{m}_k). \tag{2.51}$$

The BFGS formula to update the inverse Hessian at next iteration $k+1$ is given by (Nocedal and Wright, 2006; Monteiller et al., 2015)

$$\mathbf{H}_{k+1}^{-1} = \left( \mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) \mathbf{H}_k^{-1} \left( \mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \tag{2.52}$$

where $\mathbf{s}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$ is the model difference between the model at current and previous iteration, and $\mathbf{y}_k = \nabla \chi(\mathbf{m}_{k+1}) - \nabla \chi(\mathbf{m}_k)$ the change of gradient. From this equation we see that the inverse Hessian can be approximated iteratively with the history of models and gradients. However with this formulation we still need to store and manipulate the big matrix $\mathbf{H}_{k+1}^{-1}$ which has a size of the square of the number of points.

Equation 2.51 indicates that to obtain the descent direction one only needs to compute a matrix-vector product between the inverse Hessian and the gradient. It is therefore not necessary to store the entire inverse Hessian but only to compute the product between the inverse Hessian approximate with the gradient vector. This is implemented in l-BFGS (limited-memory BFGS) method (Nocedal and Wright, 2006). The principle of l-BFGS is to use the relation 2.52 to express $\mathbf{H}_{k+1}^{-1}$ in terms of the model difference $\mathbf{s}_{1\ldots k}$ and gradient difference $\mathbf{y}_{1\ldots k}$. With this approximation, only $2(k+1)$ arrays of size number of points is required to be stored in the memory.

One needs an initial estimation of inverse Hessian $\mathbf{H}_0^{-1}$. Nocedal and Wright (2006) suggest to set $\mathbf{H}_0^{-1} = \gamma \mathbf{I}$, where the scaling factor $\gamma = \frac{\mathbf{s}_0^T \mathbf{y}_0}{\mathbf{y}_0^T \mathbf{y}_0}$ is to ensure that the search

direction $\mathbf{p}_k$ is properly scaled for the line search method.

*2.2.1.2 Line search: Wolfe condition to accept a step length*

In the previous section, we saw how to obtain the descent direction with the l-BFGS method. Using the equation 2.49, we still need to determine the step length $\alpha$. In this search, a small value of $\alpha$ will tend to slow down the convergence, while a large value of $\alpha$ will prevent the cost function from decreasing. Two strategies allow us to find an appropriate step: the thrust region search and the line search (Nocedal and Wright, 2006). We focus here on the line search method because it is easier to implement. The line search approach aims to determine the appropriate step length $\alpha$ based on a fixed direction. This can be formulated as an optimization problem to find the step length $\alpha$ that minimizes the univariate function

$$\phi(\alpha_k) = \chi(\mathbf{m}_k + \alpha_k \mathbf{p}_k). \tag{2.53}$$

For large problems, it is too expensive to evaluate this cost function to find the exact global minimum or even a local minimum of $\phi$. A more practical strategy is to perform an inexact line search to find a step length that satisfies certain criteria for the decay of the cost function. A practical and efficient criterion is the Wolfe's conditions, which we detail below.

Let us consider the second order Taylor expansion of $\phi(\alpha)$

$$\phi(\alpha_k) \approx \phi(0) + \alpha_k \phi^{'}(0) + \mathcal{O}(\alpha_k^2), \tag{2.54}$$

where $\phi^{'}(\alpha)$ is the derivative of $\phi$. The first Wolfe's condition is to get a sufficient decrease of the objective function from one iteration to the next. That is, we want

$$\phi(\alpha_k) \leq \phi(0) + c_1 \alpha_k \phi^{'}(0) \tag{2.55}$$

to be satisfied for a constant $c_1 \in (0, 1)$. Note that the second term is negative since the $\mathbf{p}_k$ is the direction of descent which means that $\phi^{'}(0) = \mathbf{p}_k^T \nabla \chi(\mathbf{m}_k)$ is negative. This condition

is illustrated in Figure 2.3. If $c_1$ is chosen close to 1 so that the decay rate is large, it is more difficult to find the step. In practice, we thus use a small value $c_1 = 0.1$.

The first condition is in general not sufficient to avoid finding too short steps, which leads to a slow convergence. To exclude these unacceptably short steps, the second (strong) Wolfe's condition

$$|\phi^{'}(\alpha)| \leq c_2 |\phi^{'}(0)| \tag{2.56}$$

is introduced, for a constant $c_2 \in (c_1, 1)$. This condition imposes that the slope of the cost function for the step $\alpha$ is less than that the slope at 0. This constraint on the slope makes it possible to exclude steps that are too short as well as points that are far from $\phi$ stationary points.

In practice (e.g. Monteiller et al., 2015), given an initial estimate of $\alpha$, we update the model and then examine whether the corresponding functional $\phi$ satisfies both Wolfe's conditions. If the step length is too large, the first inequality is not verified and a smaller step is tested. On the other hand, if the step length is too small, the second condition is not verified and a larger step is considered. If the number of line search exceeds a given prescribed threshold, typically 10, we assume that the convergence is reached and terminate the inversion.

## 2.3 Application to the FWI problem

We want to minimize a cost function that is composed of the l2-norm of the misfit function and a regularization term (2.44). The misfit function can be written

$$\mathbf{J}(\mathbf{m}) = \sum_{s=1}^{N} \sum_{r=1}^{M} \int_{t_1}^{t_2} \frac{1}{2} \|\mathbf{u}(\mathbf{x}_r, \mathbf{x}_s; t) - \mathbf{d}(\mathbf{x}_r, \mathbf{x}_s; t)\|^2 dt, \tag{2.57}$$

where $\mathbf{u}$ represents the synthetic seismograms and $\mathbf{d}$ the observed records for the source-receiver pair $(s, r)$, $\mathbf{x}_r$ and $\mathbf{x}_s$ the location of receiver and station, respectively, and $[t_1, t_2]$ the time window considered. Since we are using local descent direction algorithms to iteratively minimize the cost function, the descent direction, or model perturbation, is computed by

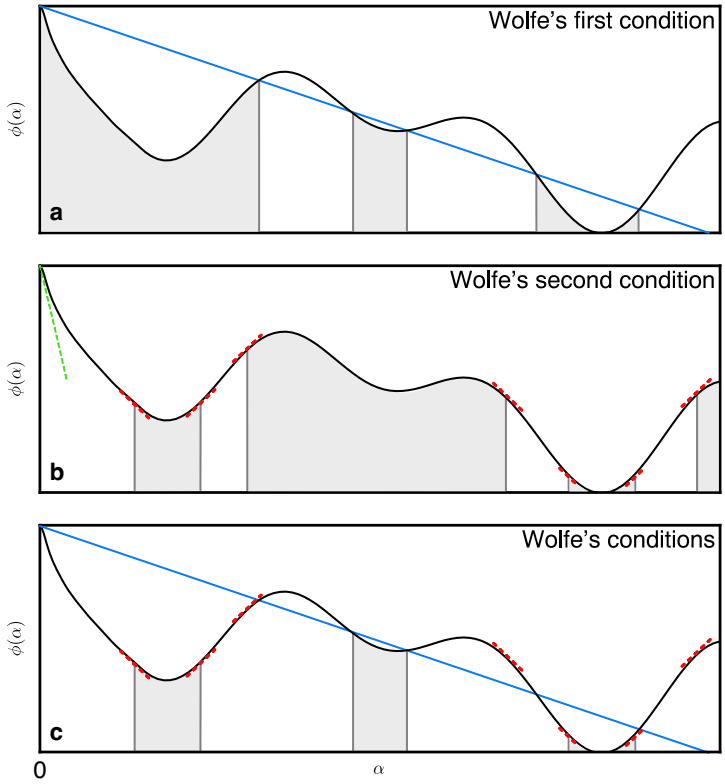


Figure 2.3: Acceptable step length with Wolfe's conditions. (a) Step length that satisfies the first Wolfe condition. The blue line shows the threshold defined by 2.55, all points below this line satisfy the first Wolfe condition (sufficient decrease). (b) Step length that satisfies the second Wolfe condition. The green line shows the slope threshold (2.56). The gray shaded area shows the points where the second Wolfe condition is satisfied (curvature condition). (c) Points where the both Wolfe conditions are satisfied. Note that the curvature condition prevent to use values which are to close to 0.

using the gradient of the misfit function and model regularization. The gradient of the misfit function is calculated using the adjoint state formulation (Tromp et al., 2005; Fichtner et al., 2009), because it is optimal for 3D numerical methods.

### 2.3.1 Gradient computation based on adjoint state method

The perturbation of the misfit function can be expressed as

$$\delta\mathbf{J}(\mathbf{m}) = \sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T [\mathbf{u}(\mathbf{x_r},\mathbf{x_s};\mathbf{t}) - \mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)]\cdot\delta\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)dt, \tag{2.58}$$

where $\delta\mathbf{u}$ represents the perturbation of displacement field $\mathbf{u}$ produced by a model perturbation $\delta\mathbf{m}$. The exact solution of $\delta\mathbf{u}$ can be described with a Lippmann-Schwinger equation, whose resolution is complicated. Since the heterogeneities we want to image represent small deviations from the local mean, the first order Born approximation is usually adopted. It has been applied in ray-theoretical (e.g. Hudson, 1977) and normal-mode approaches (e.g. Snieder and Nolet, 1987). The perturbation of displacement $\delta\mathbf{u}$ can be written in terms of perturbation of the density $\delta\rho$ and the fourth-order elasticity tensor $\delta\mathbf{c}$ as

$$\begin{aligned}\delta\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t) = -\int_V\int_0^T \Big[&\delta\rho(\mathbf{x})\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t')\partial_{t'}^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t') \\ &+ \nabla\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t') : \delta\mathbf{c}(\mathbf{x}) : \nabla\mathbf{u}(\mathbf{x},\mathbf{x}_s;t')\Big]d^3\mathbf{x}\;dt',\end{aligned} \tag{2.59}$$

where $V$ is the model volume, $\mathbf{G}$ the Green's tensor, and colon represents a double dot product of two tensors. Upon insertion of equation (2.59) in (2.58) we obtain

$$\begin{aligned}\delta\mathbf{J}(\mathbf{m}) = -\sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T [\mathbf{u}(\mathbf{x_r},\mathbf{x_s};\mathbf{t}) - \mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)]\cdot \\ \left[\int_V\int_0^T \Big[\delta\rho(\mathbf{x})\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t')\partial_{t'}^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t') + \nabla\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t') : \delta\mathbf{c}(\mathbf{x}) : \nabla\mathbf{u}(\mathbf{x},\mathbf{x}_s;t')\Big]d^3\mathbf{x}\;dt'\right]dt.\end{aligned} \tag{2.60}$$

We define the waveform adjoint source for each source $\mathbf{x}_s$ as

$$\mathbf{f}^\dagger(\mathbf{x}, \mathbf{x}_s; t) = \sum_{r=1}^{M} [\mathbf{u}(\mathbf{x}_r, \mathbf{x}_s; T-t) - \mathbf{d}(\mathbf{x}_r, \mathbf{x}_s; T-t)] \delta(\mathbf{x} - \mathbf{x_r}). \tag{2.61}$$

The corresponding adjoint wavefield is thus

$$\mathbf{u}^\dagger(\mathbf{x}, \mathbf{x}_s; t) = \int_V \int_0^{t'} [\mathbf{G}(\mathbf{x}, \mathbf{x}'; t'-t) \cdot \mathbf{f}^\dagger(\mathbf{x}', \mathbf{x}_s; t)] \; d^3\mathbf{x}' \; dt. \tag{2.62}$$

Equation (2.58) can thus be expressed as

$$\begin{aligned} \delta \mathbf{J}(\mathbf{m}) = -\sum_{s=1}^{N} \int_V \int_0^T [\delta\rho(\mathbf{x}) \mathbf{u}^\dagger(\mathbf{x}, \mathbf{x}_s; T-t) \partial_t^2 \mathbf{u}(\mathbf{x}, \mathbf{x}_s; t) \\ + \nabla \mathbf{u}^\dagger(\mathbf{x}, \mathbf{x}_s; T-t) : \delta\mathbf{c} : \nabla \mathbf{u}(\mathbf{x}, \mathbf{x}_s; t)] d^3\mathbf{x} dt. \end{aligned} \tag{2.63}$$

Following Tromp et al. (2005), we consider an isotropic elastic tensor $\mathbf{c}$ described by the bulk modulus $\kappa$ and the shear modulus $\mu$:

$$\mathbf{c} = c_{ijkl} = (\kappa - \frac{2}{3}\mu)\delta_{ij}\delta_{kl} + \mu(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl}). \tag{2.64}$$

The perturbation of the misfit function in the isotropic case thus writes

$$\delta \mathbf{J} = -\sum_{s=1}^{N} \int_V [K_\rho(\mathbf{x}, \mathbf{x}_s)\delta\rho(\mathbf{x}) + K_\kappa(\mathbf{x}, \mathbf{x}_s)\delta\kappa(\mathbf{x}) + K_\mu(\mathbf{x}, \mathbf{x}_s)\delta\mu(\mathbf{x})], \tag{2.65}$$

where

$$\begin{aligned} K_\rho(\mathbf{x}, \mathbf{x}_s) &= -\int_0^T \mathbf{u}^\dagger(\mathbf{x}, \mathbf{x}_s; T-t) \partial_t^2 \mathbf{u}(\mathbf{x}, \mathbf{x}_s; t) dt \\ K_\kappa(\mathbf{x}, \mathbf{x}_s) &= -\int_0^T \nabla \cdot \mathbf{u}^\dagger(\mathbf{x}, \mathbf{x}_s; T-t) \nabla \cdot \mathbf{u}(\mathbf{x}, \mathbf{x}_s; t) dt \\ K_\mu(\mathbf{x}, \mathbf{x}_s) &= -\int_0^T \mathbf{D}\mathbf{u}^\dagger(\mathbf{x}, \mathbf{x}_s; T-t) : \mathbf{D}\mathbf{u}(\mathbf{x}, \mathbf{x}_s; t) dt. \end{aligned} \tag{2.66}$$

In these expressions, $\mathbf{Du}$ and $\mathbf{Du}^\dagger$ denote the traceless strain deviator and its adjoint form,

respectively. The equations indicate that the gradient of the cost function can be calculated by the correlation between the incident field $\mathbf{u}$ and the adjoint field $\mathbf{u}^\dagger$. For the parameterization $(\rho, \kappa, \mu)$, the gradient with respect to the density is obtained by correlating the particle accelerations of the incident field and the particle displacements of the adjoint field, while the gradient with respect to the elastic parameters is constructed by correlating the incident and adjoint strain fields. Note that it is possible to compute directly the kernels for other parameterizations (e.g. $(\rho, \lambda, \mu)$ or $(\rho, v_p, v_s)$ ). We have chosen to follow the formulation of Tromp et al. (2005) since in the SPECFEM3D code, the displacements as well as the trace of the strain and the deviatoric strain are already stored in memory and are thus easily reusable for this purpose. This remains general as one can use another parameterization by applying a simple chain rule. Our approach is thus to post-process the gradient if one wants to use another parameterization. This approach allows to easily test different types of parameterization without changing the code.

#### *2.3.1.1 Parametrization change*

Given a reference parametization $m = (m_1, m_2, ..., m_n)$ let us define a new parametrization $p = (p_1, p_2, ..., p_n)$. For simplicity, we identify the functions that allow the change of parameters and the parameters themselves,

$$\begin{aligned} p_i &= p_i(m) = p_i(m_1, m_2, ..., m_n) \\ m_i &= m_i(p) = m_i(p_1, p_2, ..., p_n). \end{aligned} \tag{2.67}$$

We further assume that the functions are invertible. The chain rule formula ensure that

$$\nabla_{p_k} \mathbf{J}(p) = \sum_1^n \nabla_{m_i} \mathbf{J}(m) \frac{\partial m_i(p)}{\partial p_k}. \tag{2.68}$$

From a reference parameterization $m$ and the gradients $\nabla_m \mathbf{J}(m)$ associated to it, it is very simple to compute the gradient $\nabla_p \mathbf{J}(p)$ for a new parameterization $p$ by simply applying the Jacobian matrix of the parameter change $\frac{\partial m}{\partial p}$. This approach is flexible and allows us

to add on demand a new family based either on the reference parameterization or on a parameterization already defined. To do this, we have developed post-processing library for gradients. The advantage is that only a small part of the computation workflow is affected and the SPECFEM3D code remains unchanged. In addition, the gradient post-processing time is negligible compared to the gradient calculation itself, making our workflow efficient and flexible. For the SPECFEM3D code the reference parameterization used is $m = (\rho, \kappa, \mu)$, so we can define the gradients for different parameterizations such as $p = (\rho, v_p, v_s)$ or $p = (\rho, \lambda, \mu)$. We illustrate this with some examples in the following.

**Parameterization** $p = (\rho, v_p, v_s)$

$$p(m) = \begin{pmatrix} \rho \\ v_p \\ v_s \end{pmatrix} = \begin{pmatrix} \rho \\ \sqrt{\frac{\kappa - \frac{4}{3}\mu}{\rho}} \\ \sqrt{\frac{\mu}{\rho}} \end{pmatrix} \tag{2.69}$$

$$m(p) = \begin{pmatrix} \rho \\ \kappa \\ \mu \end{pmatrix} = \begin{pmatrix} \rho \\ \rho(v_p^2 + \frac{4}{3}v_s^2) \\ \rho v_s^2 \end{pmatrix} \tag{2.70}$$

$$\frac{\partial m}{\partial p}(p) = \begin{pmatrix} 1 & 0 & 0 \\ (v_p^2 + \frac{4}{3}v_s^2) & 2\rho v_p & \frac{8}{3}\rho v_s \\ v_s^2 & 0 & 2\rho v_s \end{pmatrix} \tag{2.71}$$

The gradient is given by

$$\begin{aligned} \nabla_\rho \mathbf{J}(p) &= \nabla_\rho \mathbf{J}(m) + (v_p^2 + \tfrac{4}{3}v_s^2)\nabla_\kappa \mathbf{J}(m) + v_s^2 \nabla_\mu \mathbf{J}(m) \\ \nabla_{v_p} \mathbf{J}(p) &= 2\rho v_p \nabla_\kappa \mathbf{J}(m) \\ \nabla_{v_s} \mathbf{J}(p) &= 2\rho v_s (\tfrac{4}{3}\nabla_\kappa \mathbf{J}(m) + \nabla_\mu \mathbf{J}(m)) \end{aligned} \tag{2.72}$$

**Parametrization** $p = (ln(\rho), ln(v_p), ln(v_s))$ It is also possible to make a change of parameterization concerning the metric, for example using the logarithm of the parameter. In that

case

$$p(m) = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} ln(\rho) \\ ln(v_p) \\ ln(v_s) \end{pmatrix}, \tag{2.73}$$

$$m(p) = \begin{pmatrix} \rho \\ v_p \\ v_s \end{pmatrix} = \begin{pmatrix} exp(p_1) \\ exp(p_2) \\ exp(p_3) \end{pmatrix}, \tag{2.74}$$

$$\frac{\partial m}{\partial p}(p) = \begin{pmatrix} \rho & 0 & 0 \\ 0 & v_p & 0 \\ 0 & 0 & v_s \end{pmatrix}, \tag{2.75}$$

and the gradient is given by

$$\begin{aligned} \nabla_{ln(\rho)}\mathbf{J}(p) &= \rho\nabla_{\rho}\mathbf{J}(m) \\ \nabla_{ln(v_p)}\mathbf{J}(p) &= v_p\nabla_{v_p}\mathbf{J}(m) \\ \nabla_{ln(v_s)}\mathbf{J}(p) &= v_s\nabla_{v_s}\mathbf{J}(m). \end{aligned} \tag{2.76}$$

Then changing the parameter to the logarithm just consists of scaling the gradient by the parameter itself, as it can also be seen in (Tromp et al., 2005).

**Parametrization** $p = (\rho, v_p, \nu = \frac{v_p}{v_s})$

$$p(m) = \begin{pmatrix} \rho \\ v_p \\ \nu \end{pmatrix} = \begin{pmatrix} \rho \\ v_p \\ \frac{v_p}{v_s} \end{pmatrix} \tag{2.77}$$

$$m(p) = \begin{pmatrix} \rho \\ v_p \\ v_s \end{pmatrix} = \begin{pmatrix} \rho \\ v_p \\ \frac{v_p}{\nu} \end{pmatrix} \tag{2.78}$$

$$\frac{\partial m}{\partial p}(p) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{v_s}{v_p} & -\frac{v_s^2}{v_p} \end{pmatrix} \tag{2.79}$$

and the gradients are given by

$$\begin{aligned} \nabla_\rho \mathbf{J}(p) &= \nabla_\rho \mathbf{J}(m) \\ \nabla_{v_p} \mathbf{J}(p) &= \nabla_{v_p} \mathbf{J}(m) + \tfrac{v_s}{v_p} \nabla_{v_p} \mathbf{J}(m) \\ \nabla_\nu \mathbf{J}(p) &= -\tfrac{v_s^2}{v_p} \nabla_{v_s} \mathbf{J}(m) \end{aligned} \tag{2.80}$$

### 2.3.2 Regularization

The FWI problem is ill-conditioned, which implies that the solution to the minimization problem is not unique. In other words, it is always possible to find several different models that are capable of explaining the observed data. To better constrain the inversion, a regularization (or penalty) term is thus usually added to the cost function. The regularized problem imposes conditions on the model, for example that it be smooth or that it deviate as little as possible from an a priori reference model. A popular regularization method is the Tikhonov regularization (Tikhonov and Arsenin, 1977) which seeks to minimize a cost function with a penalty term $\mathbf{L}(\mathbf{m})$ of the form

$$\chi(\mathbf{m}) = \mathbf{J}(\mathbf{m}) + \lambda ||\mathbf{L}\mathbf{m}||^2. \tag{2.81}$$

Different types of penalties can be considered, depending on the a priori information (or assumptions) that one wishes to add to the problem. It can be an identity matrix $\mathbf{I}$, the gradient of the model $\mathbf{L} = \boldsymbol{\nabla}$ to penalize the discontinuities, the roughness of the model $\mathbf{L} = \boldsymbol{\nabla}^2$ to favor smooth models, or a square root of inverse covaraince matrix $\mathbf{L} = \mathbf{C}_\mathbf{M}^{-\frac{1}{2}}$ for the a Gaussian prior on model (e.g. Tarantola, 2005).

The regularization parameter $\lambda$ allows to reach a good compromise between the mini-

mization of the waveform and model residuals. If the parameter $\lambda$ is too small, the second term has little impact on the inversion. The contributions of the data errors will dominate the inversion and lead to unrealistic parameter reconstructions. On the other hand, if the $\lambda$ parameter is too large, the first residual term on the data will have less impact on the inversion and the data will not be fit correctly. To get a good balance between the two terms, a popular method is to use the L-curve approach (Hansen, 1992). The idea is to plot the curve in a log-log scale of the norm of the data residuals versus the norm of the model residuals, for different $\lambda$. It is usually L-shaped and the optimal $\lambda$ is chosen near the bend or inflection point of the curve.

In previous implementations of teleseismic FWI, the Laplacian of the model has been used to regularize the the problem (Monteiller et al., 2015; Wang et al., 2016; Beller et al., 2018), with a regularization term $\mathbf{R}$ of the form

$$\mathbf{R}(\mathbf{m}) = \frac{1}{2}||\boldsymbol{\nabla}^2\mathbf{m}||^2. \tag{2.82}$$

In that case, the cost function to minimize is

$$\chi(\mathbf{m}) = \mathbf{J}(\mathbf{m}) + \frac{\lambda}{2}||\boldsymbol{\nabla}^2\mathbf{m}||^2, \tag{2.83}$$

and the corresponding gradient writes

$$\nabla\chi(\mathbf{m}) = \nabla\mathbf{J}(\mathbf{m}) + \lambda\frac{\partial\mathbf{R}(\mathbf{m})}{\partial\mathbf{m}}, \tag{2.84}$$

where the gradient of the norm of the penalty term, i.e. the Laplacian of the model, can be obtained by applying twice the discrete Laplacian to the model $\mathbf{m}$:

$$\frac{\partial\mathbf{R}(\mathbf{m})}{\partial\mathbf{m}} = \boldsymbol{\nabla}^2\boldsymbol{\nabla}^2\mathbf{m}. \tag{2.85}$$

In this thesis, we chose to follow Tarantola's approach which consists in treating the regularization term as a prior information on the model. This is a general approach that allows to control both the deviation from the prior model (damping) and the spatial regularity of the model (smoothing). Moreover, it is possible to consider a correlation between parameters (eg $\rho$ and $v_p$, $v_p$ and $v_s$, ... ) which provides an additional level of control to the inversion.

The regualrization term is then

$$\mathbf{R}(\mathbf{m}) = \frac{1}{2}||\mathbf{C}_\mathbf{M}^{-\frac{1}{2}}(\mathbf{m} - \mathbf{m}_{\text{prior}})||^2, \tag{2.86}$$

with the gradient which simply writes

$$\frac{\partial \mathbf{R}(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}). \tag{2.87}$$

#### *2.3.2.1 Model covariance matrix*

**Diagonal $\mathbf{C_M}$** The choice of the covariance matrix needs to be carrefull in order to be numerically tractable and efficient. The simplest way is to assume that there is no correlation between the parameters which leads to a diagonal model covariance matrix. Let us assume that the model is described by $n$ physical parameters. These parameters can be density, wave speed, bulk modulus, Lamé coefficients, Young's modulus... (any physical parameter that intervenes in the elastodynamic equation). Usually we consider 3 parameters, the density $\rho$, the P-wave velocity $v_p$ and S-wave velocity $v_s$. In that case the covariance matrix writes

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{v_p}^2 & 0 \\ 0 & 0 & \sigma_{v_s}^2 \end{bmatrix}. \tag{2.88}$$

The inverse or square root of the model covariance matrix is trivial to compute.

At this point we can see that previous implementations of the regularization constraints

did not explicitly take into account the non-diagonal terms. In practice, the model covariance matrix was only a scaling to balance the contributions of the different parameters.

**Non-diagonal $\mathbf{C_M}$** In the context of describing the Earth's structure, some physical parameters are not independent, there is a degree of correlation between different model parameters. For example, the Birch's law that establishes a positive linear relation between compressional wave velocity $v_p$ and density $\rho$ of rocks (Birch, 1960). The common ratio of the compressional wave velocity and the shear wave velocity $v_p/v_s$ is also reported to lie between 1.7 and 1.9 (e.g. Christensen, 1996; Peacock et al., 2011), which is an implicit correlation of $v_p$ and $v_s$. Another kind of correlation is the spatial correlation, which assumes the smooth variation of a value from one point to its vicinity. This is usually done by applying an $\ell_2$ norm constraint on the solution. Now we integrate both information of the correlations to the a priori covarivance matrix.

First, the correlation between family of parameters $(\rho, v_p, v_s)$ writes (e.g. Nataf et al., 1986; Tarantola, 2005)

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & r(\rho, v_p)\mathbf{I} & r(\rho, v_s)\mathbf{I} \\ r(\rho, v_p)\mathbf{I} & \mathbf{I} & r(v_p, v_s)\mathbf{I} \\ r(\rho, v_s)\mathbf{I} & r(v_p, v_s)\mathbf{I} & \mathbf{I} \end{bmatrix}, \tag{2.89}$$

where $r(\alpha, \beta)$ is the correlation coefficient of the two parameters $\alpha$ and $\beta$ that has the value in $-1 \leq r(\alpha, \beta) \leq 1$. Figure 2.4 shows the plotting of 2-D Gaussian functions with different correlation coefficients. Clearly, when $r$ is close to 1 or -1, the Gaussian distributions of the two variables are close to positive or negative correlation, respectively. When $r = 0$ the correlation matrix becomes the diagonal matrix as in equation 2.88, the two variables are uncorrelated.

The correlation coefficients can also be used for spatial regularization. In this case, a correlation function that depends on the distance between two points is used to constrain the smoothing of the model. To correlate the value at different nodes $\mathbf{x}_i$ and $\mathbf{x}_j$ inside the

mesh, we define a spatial kernel function $\phi(|\mathbf{x}_i - \mathbf{x}_j|)$ which depends on the distance of two nodes. The choice of $\phi$ is important to be able to efficiently evaluate in SEM discretization. Following Potin (2016) and Araujo et al. (2021) who choose $\phi$ as an exponential kernel, the 3-D exponential kernel is defined (Oliver, 1998; Trinh et al., 2017)

$$C(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\sqrt{\frac{(x_i - x_j)^2}{L_x^2} + \frac{(y_i - y_j)^2}{L_y^2} + \frac{(z_i - z_j)^2}{L_z^2}}\right), \tag{2.90}$$

where $L_x$, $L_y$ and $L_z$ are the correlation lengths (or smoothing lengths) in each cartesian direction. The covariance matrix containing both the correlations between parameters and the smoothing is written,

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2 \mathbf{C} & r(\rho, v_p)\sigma_\rho\sigma_{v_p}\mathbf{C} & r(\rho, v_s)\sigma_\rho\sigma_{v_s}\mathbf{C} \\ r(\rho, v_p)\sigma_\rho\sigma_{v_p}\mathbf{C} & \sigma_{v_p}^2\mathbf{C} & r(v_p, v_s)\sigma_{v_p}\sigma_{v_s}\mathbf{C} \\ r(\rho, v_s)\sigma_\rho\sigma_{v_s}\mathbf{C} & r(v_p, v_s)\sigma_{v_p}\sigma_{v_s}\mathbf{C} & \sigma_{v_s}^2\mathbf{C}, \end{bmatrix}, \tag{2.91}$$

where $\mathbf{C}$ is the spatial covariance matrix which contains the smoothing values as in2.90. It can be decomposed into

$$\mathbf{C_M} = \mathbf{\Sigma S P S \Sigma}, \tag{2.92}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\rho \mathbf{I} & 0 & 0 \\ 0 & \sigma_{V_P}\mathbf{I} & 0 \\ 0 & 0 & \sigma_{V_S}\mathbf{I} \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \mathbf{C}^{\frac{1}{2}} & 0 & 0 \\ 0 & \mathbf{C}^{\frac{1}{2}} & 0 \\ 0 & 0 & \mathbf{C}^{\frac{1}{2}} \end{bmatrix}. \tag{2.93}$$

$\mathbf{\Sigma}$ is the scaling matrix, $\mathbf{S}$ is the smoothing matrix and $\mathbf{P}$ is the correlation matrix defined in equation 2.89.

Note that in the equations 2.86 and 2.87 the matrices $\mathbf{C_M^{-1}}$ and $\mathbf{C_M^{-\frac{1}{2}}}$ are needed to be evaluated. The decomposition 2.92 is helpful to compute those matrices since,

$$\mathbf{C_M^{-1}} = \mathbf{\Sigma^{-1} S^{-1} P^{-1} S^{-1} \Sigma^{-1}} \tag{2.94}$$

and

$$\mathbf{C_M^{-\frac{1}{2}}} = \mathbf{\Sigma}^{-1}\mathbf{S}^{-1}\mathbf{P}^{-\frac{1}{2}}. \tag{2.95}$$

The effective computation of each of these matrices is therefore necessary to define an efficient inversion workflow. One of the difficulties comes from the calculation of the smoothing matrix $\mathbf{S}^{-1}$. But, it is now well established (Oliver, 1998; Potin, 2016; Trinh et al., 2017; Araujo et al., 2021) that in the case of an exponential smoothing kernel, its inverse operator writes,

$$C^{-1}(\mathbf{r}, \mathbf{r}') = \frac{1}{8\pi L_x L_y L_z} \left(\mathbf{I} - \Delta_L\right)^2, \tag{2.96}$$

with

$$\Delta_L = L_x^2\partial_x^2 + L_y^2\partial_y^2 + L_z^2\partial_z^2. \tag{2.97}$$

Thus the matrix $\mathbf{S}^{-1}$ is easily computed using the Laplacian operator 2.96. (Trinh et al., 2017) showed how to efficiently apply this matrix on SEM mesh using a weak formulation. It is a matter of reusing the basis functions on the SEM mesh and the different partial derivative operators that are already used to solve the wave equation. It is thus very easy to implement this smoothing method in SPECFEM3D because all the arrays and matrix-vector products are already defined, a simple addition of a new function is enough to achieve this goal. The last step is the computation of $\mathbf{P}^{-1}$ or $\mathbf{P}^{-\frac{1}{2}}$ which can be done with a singular value decomposition,

$$\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t, \tag{2.98}$$

with $\mathbf{V}$ the matrix containing the orthogonal eigenvectors of $\mathbf{P}$. We can thus simply write,

$$\mathbf{P}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^t, \tag{2.99}$$

$$\mathbf{P}^{-\frac{1}{2}} = \mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^t. \tag{2.100}$$

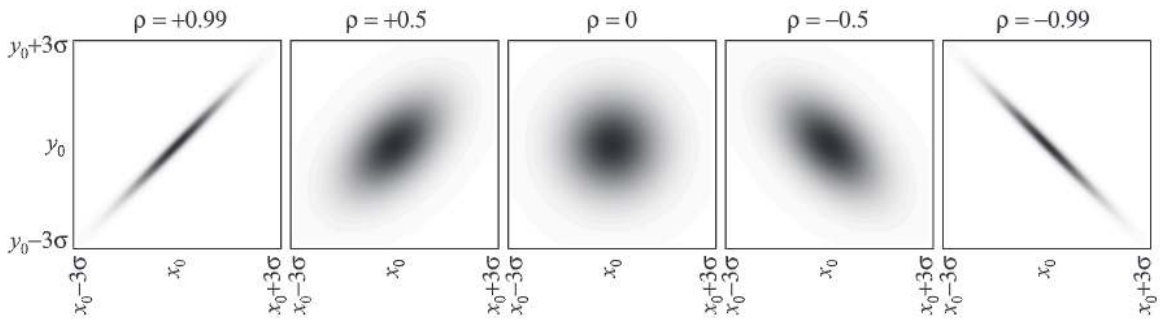


Figure 2.4: Two-dimensional Gaussian intensity map plotted using three standard deviations in each of the two axes for different values of the coefficient of correlation. (from Nocedal and Wright, 2006)

#### *2.3.2.2 Smoothing as variable change*

In practice the gradient of the cost function is smoothed to avoid small scale fluctuations (Bozdağ et al., 2016; Zhu et al., 2015). However this can introduce inconsistencies in the application of Wolfe's rule (2.55) where the magnitude of the smooth gradient $\phi'(0)$ can lead to wrong step length estimation. To overcome this issue, one possibility is to introduce a variable change to make the smooth gradient valid in Wolfe's rules. We then define an auxiliary model parameter

$$\hat{\mathbf{m}}_k = \mathbf{C}_{\mathbf{M}}^{-\frac{1}{2}}(\mathbf{m}_k - \mathbf{m}_{\text{prior}}), \tag{2.101}$$

where the reciprocal formula for $\mathbf{m}_k$ is

$$\mathbf{m}_k = \mathbf{C}_{\mathbf{M}}^{\frac{1}{2}}\hat{\mathbf{m}}_k + \mathbf{m}_{\text{prior}}. \tag{2.102}$$

Considering this auxiliary model parameter, the cost function is written as

$$\chi(\hat{\mathbf{m}}_k) = \frac{1}{2}||\mathbf{C}_{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}_k))||^2 + \frac{\lambda}{2}||\hat{\mathbf{m}}_k||^2. \tag{2.103}$$

The regularization then becomes a simple damping which is more numerically stable than

computing the Laplacian of the model. The corresponding gradient of $\chi$ with respect to this new model can be obatined with the chain rule

$$\frac{\partial \chi(\hat{\mathbf{m}})}{\partial \hat{\mathbf{m}}} = \frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}} \frac{\partial \mathbf{m}}{\partial \hat{\mathbf{m}}}, \tag{2.104}$$

where

$$\frac{\partial \mathbf{m}}{\partial \hat{\mathbf{m}}} = \mathbf{C}_{\mathbf{M}}^{\frac{1}{2}}. \tag{2.105}$$

Substituting equation 2.104 with equations 2.84 and 2.105, we have

$$\frac{\partial \chi(\hat{\mathbf{m}})}{\partial \hat{\mathbf{m}}} = \mathbf{C}_{\mathbf{M}}^{\frac{1}{2}} \nabla \mathbf{J}(\mathbf{m}_k) + \lambda \hat{\mathbf{m}}_k. \tag{2.106}$$

It consists to apply the square root of the covaraince matrix on regular gradient and add the model weighted by the damping factor.

After the new physical model is updated by

$$\mathbf{m}_{k+1} = \mathbf{C}_{\mathbf{M}}^{\frac{1}{2}} \hat{\mathbf{m}}_{k+1} + \mathbf{m}_{\text{prior}}. \tag{2.107}$$

We see that it is the same procedure as applied for the gradient, applying the square root of covariance matrix on auxiliary model and add the prior model. An advantage of this formulation is that it allows to use only the smoothing procedure which is more numerically stable than the differentiation procedures (equations 2.86, 2.87, and 2.96).

# Chapter 3. Preparation of FWI inversions

## Contents


3.1 Outline of the inversion workflow ..... **56**
3.2 Data Selection ..... **57**
3.3 Mesh design ..... **62**
3.4 Source wavelet estimation ..... **64**
3.5 Event normalization of adjoint sources ..... **67**


### 3.1 Outline of the inversion workflow

In this section, we describe our FWI workflow, which is illustrated in Fig. 3.1. We first select the waveform data that will be used in the inversion and build the regional mesh that will also be our tomographic grid. The forward and inverse problems are performed in this regional mesh. For each teleseismic event, we compute the incident wavefield on the edges of our regional grid produced by an impulsive source at the position of the source, using the source parameters taken from the GCMT catalog. The synthetic seismograms computed at the station positions give us Green's functions. By deconvolving these Green's functions from the observed data, we obtain an apparent source wavelet at each station. The average source wavelet is then obtained by aligning and averaging all the source wavelets. By convolving this average source wavelet to the tractions and velocities produced by the incident field on the edges of the regional grid, synthetic seismograms can then be computed which can be directly compared to the observed data. The optimization process that aims to reduce the root mean square deviation between the observed and synthetic traces is then performed in a prescribed frequency band. The convergence of the inversion relies on a line search criterion. The inversions progressively move toward higher frequencies, initiating the new inversions with the final model obtained at the previous lower frequency inversion. Before each inversion, we reestimate a new source wavelet using Green's functions computed in the

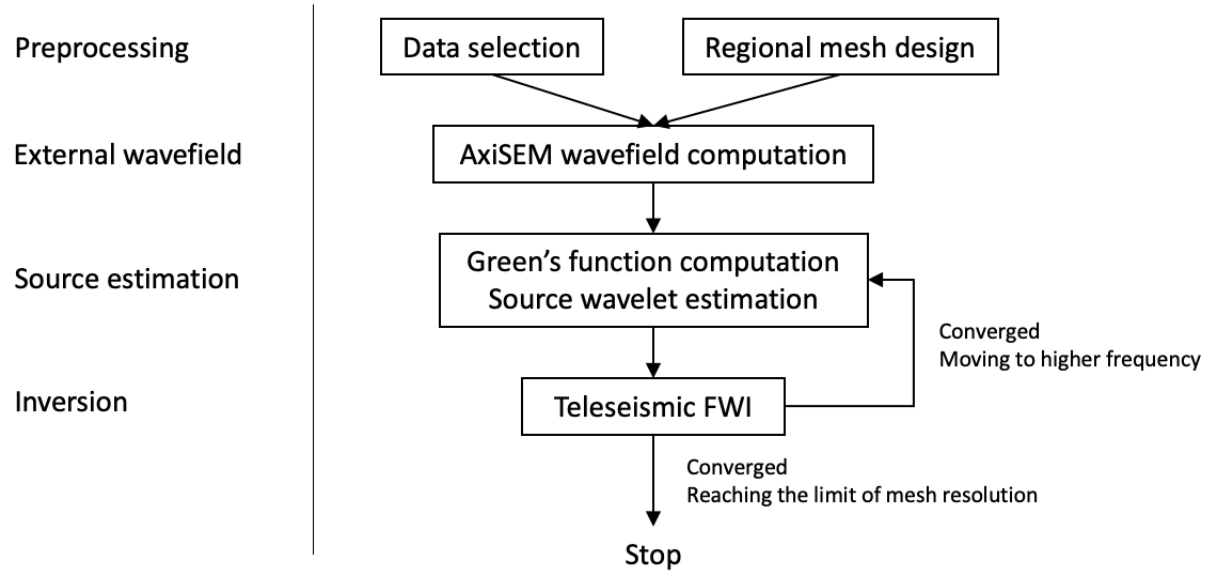


Figure 3.1: Workflow of the teleseismic full waveform inversion with coupling with external wavefield.

current 3D model.

## 3.2 Data Selection

Data selection and preparation are a key step of full waveform inversion. Indeed, since we aim to fit each wiggle in the observed seismic traces, the quality of the dataset is crucial.

The event selection is guided by a few simple rules. First, teleseismic events are selected in order to get the best possible azimuthal coverage possible. As a general rule, only earthquakes of magnitude larger than 6 have a good signal-to-noise ratio unless the event is close to the array or the array is in a direction of maximum radiation from the source, in which case a lower magnitude event may be considered. On the other hand, the higher the magnitude of the earthquake, the longer the source duration. An earthquake of magnitude greater than 7.5 has a rupture length of up to several hundred kilometers and a rupture time function of several tens of seconds. For such large events, it is thus necessary to consider longer time windows, which may be problematic if later secondary arrivals arrive in that same time window. In general, deep earthquakes have more impulsive sources and will therefore always

be favored over shallow earthquakes.

Once a list of high-quality events has been obtained, we refine the data selection to discard the bad records. Our trace selection procedure relies on the visual examination of travel time and amplitude residual maps in order to detect and discard records that show very different travel time and amplitude residuals compared with those measured at nearby stations that could result, for example, from a clock problem or from a wrong station response.

The relative time residuals are defined as the time difference between the trace and the average wavelet in a selected window taken around the main phase arrival. The travel time residual is measured by taking the maximum of the cross-correlation function

$$C(u_i(t), \overline{u}(t))(\tau) = \frac{1}{2\pi} \int_{t_1}^{t_2} u_i(t)\overline{u}(t+\tau)dt. \tag{3.1}$$

where $\overline{u}$ is the average wavelet obtained by aligning and stacking the main P or S wave records, and $u_i$ is the trace recorded by station $i$. We then remove the average time residual in order to obtain $\Delta T_i$, the travel time residual at each station $i$.

The relative amplitude anomaly $\Delta A_i$ for station $i$ is defined as the ratio of the logarithm of the energy of $u_i$ and of the energy of the average wavelet $\overline{u}$

$$\Delta A_i = \log_{10}\left(\frac{\sqrt{\int_{t_1}^{t_2} |u_i(t)|^2 dt}}{\sqrt{\int_{t_1}^{t_2} |\overline{u}(t)|^2 dt}}\right). \tag{3.2}$$

For two stations spaced a few kilometers apart, their travel time and amplitude residuals should be similar. When displayed in map view, amplitude and travel time residuals are thus expected to exhibit smooth and coherent spatial variations.

We define three criteria in order to evaluate trace quality: 1) the values of time and amplitude residuals, 2) their spatial coherence, and 3) the correlation coefficient between the trace and the average wavelet. Only the traces that have low residuals (less than 2 s for travel times and less than 0.5 for log amplitudes), consistent with those observed at adjacent stations, as well as a high correlation coefficient with the average trace, are retained.

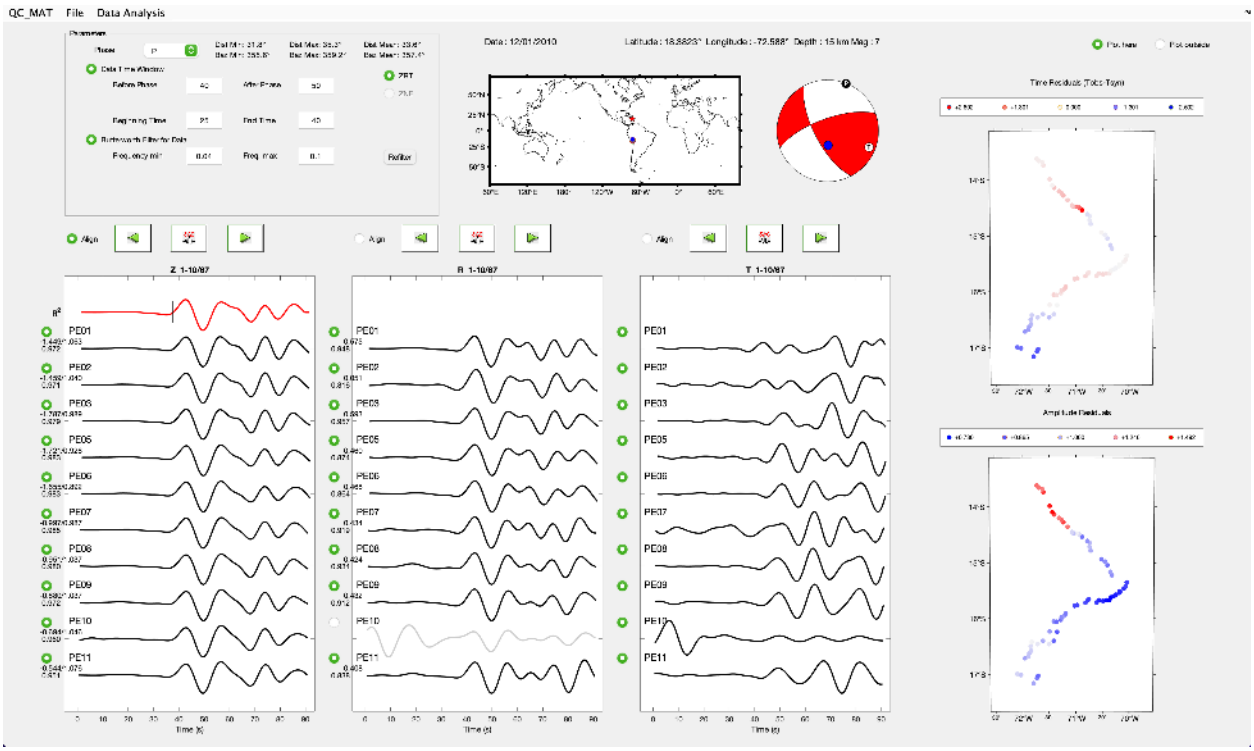


Figure 3.2: Graphical interface to perform data selection and quality control.

To better visualize the consistency and coherence of travel time and amplitude residuals, we developed a Matlab-based graphical interface[1]. This interface reads the seismic data in SAC format, displays the waveforms, and calculates and plots the relative time and amplitude residuals. (Fig. 3.2).

In hierarchical inversions, which consist in first inverting long period waveforms and progressively including higher frequency signal, the data selection, and quality control procedures need to be repeated to ensure that all the traces are of sufficient quality in the frequency range considered in the inversion. For example, when periods shorter than ∼8 s are included, microseismic noise may become an issue at some stations, which provided good S/N traces at longer periods.

[1] `https://gitlab.com/esleokan/teleseis`

#### *3.2.0.1 An example of data selection for teleseismic FWI*

We illustrate our quality control procedure on the October 11th, 1993 event recorded by the CASC93 experiment deployed in central Oregon (Fig. 3.3). After downloading the raw data from IRIS, we preprocess the data by removing the average and trend on each trace. We then remove the instrument response and apply a lowpass Butterworth filter with a corner frequency of 0.1 Hz. Note that we work with velocity records to minimize the problems that could result from inaccurate station responses. For P phases, we inspect both Z and R components, whereas for SH waves, we inspect the transverse components. The travel time and amplitude residuals are plotted in map views to visualize their spatial coherence. The correlation coefficients are computed for each trace and written below the station names in the plot showing the vertical component traces.

We can clearly see that the vertical component of stations A28 and A30 show abnormal travel time and amplitude measurements (Fig. 3.3a-c). The corresponding traces are flat (Fig. 3.3d) and this should be discarded. Another obvious problem is detected with station A37, which shows a considerable travel time residual of about -15 s, most likely resulting from a clock problem. A more subtle problem is also detected at station A16, which shows a small but significant amplitude discrepancy with adjacent stations, leading us to discard this station. A similar inspection of radial components (Fig. 3.3e-g) shows that stations A16, A28, A30 and A37 have similar problems on the radial component. However, we can detect a new problem on station A15 which is too noisy to be kept for the inversion. Note that the correlation coefficients of the radial components display larger variations along the profile because of a stronger waveform variability resulting from conversions and reverberations in the coda of the direct P wave. The traces selected for the inversion are plotted in black, whereas the rejected traces are plotted in red. Coherent and smooth waveform variations are observed along the transect, with a strong signature of the subducting oceanic crust.

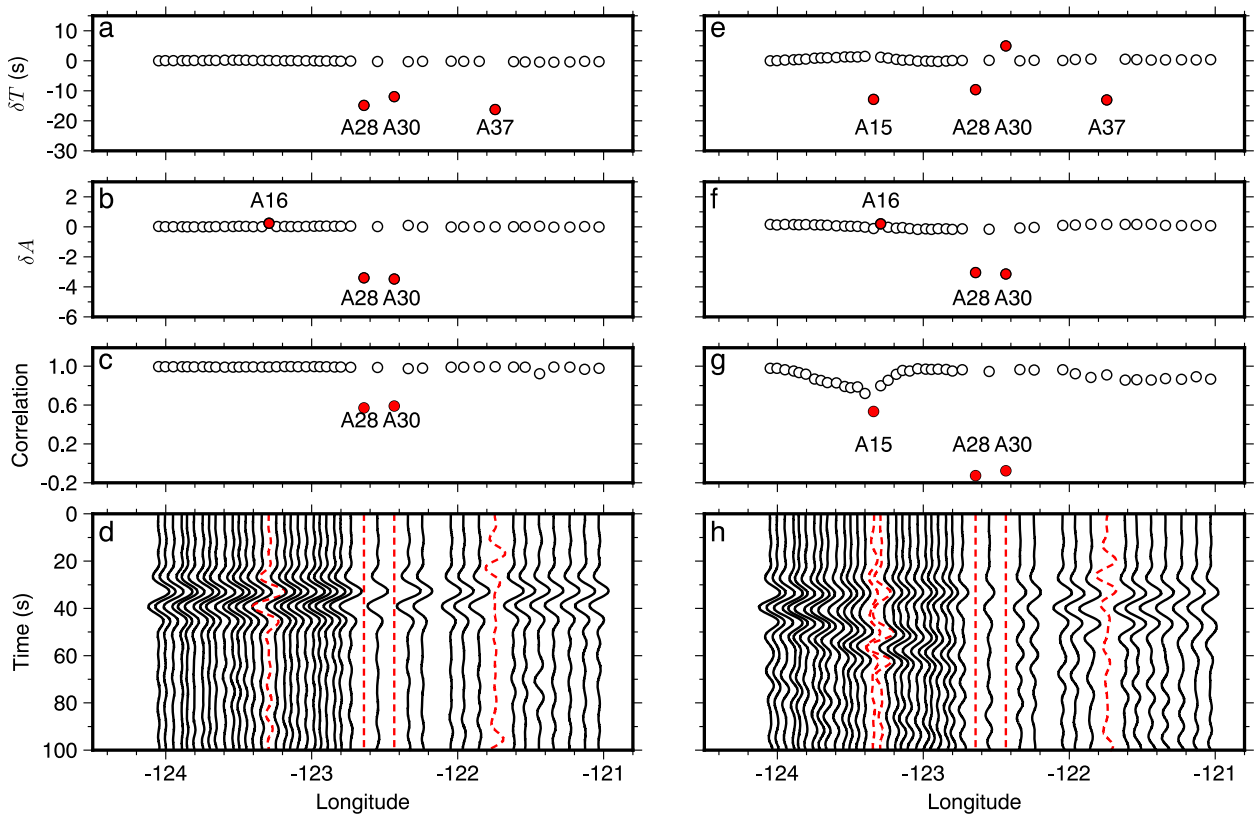


Figure 3.3: Data selection for the 1993 October 11th event in Cascadia dataset. (a-d) Time residual, amplitude residual, correlation coefficient and aligned waveform filtered in 4 to 10 s for vertical component. (e-h) Same as (a-d) but for the radial component. The traces that have been discarded are shown in red.

## 3.3 Mesh design

Once the waveform dataset has been prepared, we need to build the mesh for the spectral-element computations. The design of this mesh follows a few simple rules. Its center is positioned in the middle of the station array. The mesh extent in latitude and longitude is minimized to make the computations as efficient as possible. A tapered zone of at least 50 km from the four lateral edges is applied to ensure a smooth transition from the 1-D reference Earth model imposed on the edges of the spectral-element grid to compute the incident wavefield and the 3-D model inside. For real data applications, surface topography from the ETOPO1 model is considered. Fig. 3.4) shows a vertical cross-section in the mesh generated for the Cascadia inversion that will be presented later. Table 3.1 describes the properties of the three different meshes that we have used in this thesis.

The shortest period is determined by the number of grid points per wavelength. For each element $\Omega_e$, the shortest period $T_s$ is given by

$$T_s = N_e \cdot \frac{\overline{h}}{v}, \tag{3.3}$$

where $N_e$ is the number of points per wavelength, $\overline{h}$ is the average distance of GLL points, and $v$ is the $V_P$ or $V_S$ velocity. From the empirical study of Seriani and Priolo (1994), to correctly sample the wavefield, the SEM needs roughly 4.5 points per wavelength. In all our computations, $N_e$ is fixed to 5 to satisfy this condition.

To get a finer mesh resolution, we can reduce $T_S$ or $\overline{h}$. However, with the same degree of Lagrange polynomial, reducing $\overline{h}$ increases the number of grid points. This increases the computational cost, especially for large grid sizes (larger than ten degrees). Instead, it is advisable to increase the degree of Lagrange polynomials (Lyu et al., 2020).

#### *3.3.0.1 Computational requirements*

Forward modelings and inversions are run on the CALMIP and Jean-Zay supercomputers. Both provide CPU and GPU computing nodes. CALMIP has 12 GPU nodes equipped with 4

Table 3.1: Parameters for the three different meshes used in this thesis.

| | Synthetic | Cascadia | Peru (Big) | Peru (Small) |
|---|---|---|---|---|
| Shape of mesh | Flat | Chunk of Earth | Chunk of Earth | Chunk of Earth |
| Topography | - | ETOPO1 | ETOPO1 | ETOPO1 |
| Center of chunk (lat, lon) | - | 44.41, -122.23 | -14.0, -70.8 | -16, -71.0 |
| Length of chunk in X, Y, Z (km) | 340, 280, 150 | 660, 220, 300 | 1668, 1334, 1000 | 330, 888, 600 |
| Azimuth of chunk (°) | 0 | 0 | 50 | 50 |
| Number of elements in X, Y, Z | 40, 31, 16 | 56, 19, 33 | 84, 67, 54 | 28, 75, 52 |
| Minimum resolved period (s) | 2.5 | 2.3 | 3.4 | 2.1 |
| Domain decomposition | 1 | 2 | 8 | 4 |
| Total elements | 19.840 | 35,112 | 303,912 | 109,200 |

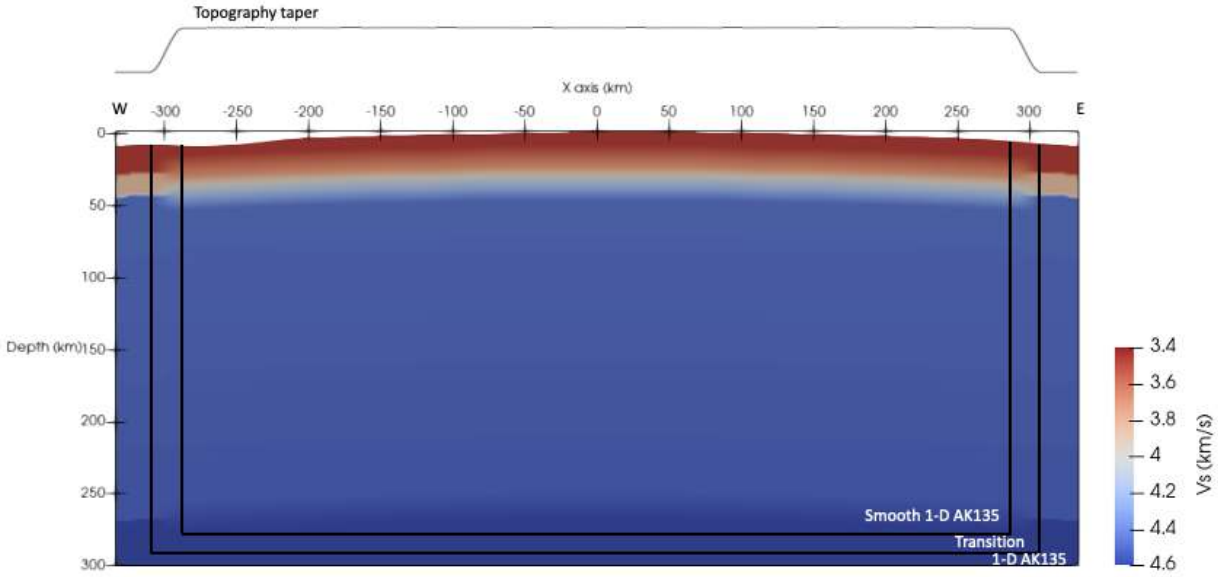


Figure 3.4: Vertical cross-section of the Cascadia mesh.

Nvidia V100 GPU and 384 GB of memory, whereas Jean-Zay offers more powerful calculation resources, with 612 GPU nodes with four Nvidia V100 and 192 GB of memory.

One inversion run scales with the size of the grid, the numbe r of events, and the length of the simulations. For applications on small meshes, as in our synthetic continental subduction and Cascadia inversions, the gradient computation for one event, which involves one forward run and one adjoint run, takes about 1 minute. For the Peru inversion, performed in a much larger grid, the computation time for one gradient takes about 2 to 3 minutes, depending on the event.

The external wavefield computation with AxiSEM requires significant disk space, which depends on the size of the mesh, the number of GLL points on the edges of the mesh, and the duration of the simulation. For example, for the Cascadia inversion, the disk space is about 5 GB for each event, whereas it is about 25 GB per event for the Peru mesh. A 700-second long trace in SAC binary format occupies 92 KB. The size of the database itself is much small in size. It scales with the number of events and stations. For example, the Cascadia dataset contains 12 events with 35 stations which consumes about 120 MB of disk storage.

### 3.4 Source wavelet estimation

In real data FWI application, we need the source parameters (hypocenter location, focal mechanism, and source-time function) to compute synthetic seismograms. In regional adjoint tomography applications (e.g. Fichtner et al., 2009), the source parameters are re-estimated as the model upgrades. In our teleseismic applications, however, since the events are located far from the target domain, re-estimating the location and the focal mechanism of a teleseismic event is unnecessary. However, we still need to estimate the apparent source wavelet, which captures all the complexities on the source side that may affect the teleseismic records, such as depth phases.

The displacement waveform $u$ at time $t$ generated by a distant source at $\mathbf{x}_s$ and recorded

at station position $\mathbf{x}_r$ can be expressed as the convolution product,

$$u(\mathbf{x}_r, t) = S(\mathbf{x_s}, t) * S_E(t) * S_R(\mathbf{x}_r, t) * I(\mathbf{r}_r, t), \tag{3.4}$$

where $S$ is the source-time function, $S_E$ the structural response outside the regional box, $S_R$ the structural response inside the box, and $I$ the instrumental response. For now, the unknown term in equation 3.4 is the exact source time function $S$ and structural response outside and inside the box $S_R$ and $S_E$. For a teleseismic event recorded by a regional dense array, $S_E$ is approximately the same at all the stations because of similar propagation paths outside the box. Therefore, $S$ and $S_E$ can be combined to define the apparent source wavelet $W$ (where $\mathbf{x_s}$ has been omitted for simplicity)

$$W(t) = S(t) * S_E(t). \tag{3.5}$$

Usually, the instrumental response is known and can be easily removed from the seismic records. Using (3.5), equation 3.4 thus simplifies to

$$u(\mathbf{x}_r, t) \approx W(t) * S_R(\mathbf{x}_r, t). \tag{3.6}$$

Note that the regional box response $S_R$ term is not known precisely, and it is exactly the objective of FWI to determine the 3-D regional model able to predict the observed waveforms. We thus need to make the additional assumption that the average structure response seen by all the stations is close to the average Green's function $\mathbf{G}_R$ in a spherically symmetric reference Earth model

$$S_R(\mathbf{x}_r, t) \approx \mathbf{G}_R(\mathbf{x}_r, t). \tag{3.7}$$

This is in general a very reasonable assumption as testified by the extreme similarity of vertical component P waveforms recorded by a regional array (see for example Figure 3.3d). In that case, we just need to compute Green's functions. For the first long-period inversion,

the Green's functions are computed in the ak135 reference Earth model with AxiSEM for the Dirac source impulse function. In later inversions, the Green's functions will be computed using the current 3-D model obtained at the previous iteration of the hierarchical FWI.

Using (3.6), we can obtain the apparent source wavelet $W_i$ at station $i$ with a simple water-level deconvolution in the frequency domain, according to

$$W_i(\mathbf{x}_r, \omega) = \frac{u_i(\mathbf{x}_r, \omega) \cdot \mathbf{G}_R(\mathbf{x}_r, \omega)'}{\mathbf{G}_R(\mathbf{x}_r, \omega) \cdot \mathbf{G}_R(\mathbf{x}_r, \omega)' + \epsilon}, \tag{3.8}$$

where $\epsilon$ is the water level, defined as a small fraction of the maximum value of the power spectrum of the windowed Green's function and the prime represents the conjugate of the matrix. This water level avoids divisions by zero and stabilizes the deconvolution. In practice, for each station, we cut data and Green's function in a time window that contains the main teleseismic arrival and compute their Fourier transform. The windowing avoids including secondary arrivals (e.g., PP, PcP, etc...) which followed different propagation paths before reaching the regional domain. To stabilize the deconvolutions, the window needs to be slightly longer than the one considered in the inversion, and that the Green's functions need to have a higher frequency content than the observed filtered waveforms. Note that deconvolution approaches in the time domain, such as the iterative-time deconvolution method (Ligorria and Ammon, 1999), could also be used for the average wavelet estimation.

Finally, we define the average apparent wavelet by averaging the wavelets obtained at all the stations:

$$W(t) = \frac{1}{n} \sum_{i=1}^{n} W_i(\mathbf{x}_r, t - \tau_i), \tag{3.9}$$

with

$$W_i(\mathbf{x}_r, t) = \mathcal{F}^{-1}(W_i(\mathbf{x}_r, \omega)). \tag{3.10}$$

Note that the individual wavelets are aligned by cross-correlation before stacking, in order to preserve all the details of the average source wavelet. To determine the time shifts $\tau_i$ that align the apparent station wavelets, we first align all the traces to the first one and compute

a first average wavelet $W_1$. In the second step, we align all the traces with respect to $W_1$ and compute the average wavelet to get a new average wavelet $W_2$. Finally, we determine the time shift that we need to apply to $W_2$ to have a zero average travel time residual between the average source wavelet and the station wavelets. The shifted average wavelet is our source wavelet $W$ that will be used to compute synthetic seismograms. The average source wavelet $W$ absorbs all the propagation effects that happen outside the regional box, for example, the depth phases and source-side scattering. It also removes the average travel time residual accumulated between the source and the regional domain, a critical step in regional travel time tomography (Aki et al., 1977).

### 3.5 Event normalization of adjoint sources

The gradient of the cost function is obtained by summing up the gradients computed in different individual events. Since the amplitude of the main teleseismic phase (here the P or SH waves) depends on the epicentral distance, the focal mechanism, and the magnitude of the event, a large amplitude variability is thus expected within a selected waveform dataset. Large amplitude differences will make the most energetic earthquakes dominate the inversion, which can have a detrimental effect on the spatial resolution in the final model. To balance the contributions of the different events, we need to normalize the event gradients. Mathematically, the gradient is proportional to the product of the amplitudes of the incoming phase and of data residuals.

The root mean square (rms) amplitude of the main phase $A_s$ is defined by

$$A_s(k) = \sqrt{\sum_{i=1}^{N} \frac{1}{N(t_2 - t_1)} \int_{t_1}^{t_2} u_i^2(t)\, dt}, \tag{3.11}$$

where $N$ is the number of selected stations for event $k$, and $[t_1, t_2]$ the time window which contains the main seismic arrival.

Similarly, we define $A_r$, the rms amplitude of data residuals as

$$A_r(k) = \sqrt{\sum_{i=1}^{N} \frac{1}{N(t_2 - t_1)} \int_{t_1}^{t_2} \left[s_i(t) - u_i(t)\right]^2 \, dt}, \tag{3.12}$$

where $s_i(t)$ and $u_i(t)$ are respectively the synthetic and observed seismograms at station $i$.

Combining (3.11) and (3.12), the normalization factor $N(k)$ for event $k$ is thus given by

$$N(k) = A_s(k) \cdot A_r(k) = \frac{1}{N(t_2 - t_1)} \sqrt{\sum_{i=1}^{N} \int_{t_1}^{t_2} s_i^2(t) \, dt} \cdot \sqrt{\sum_{i=1}^{N} \int_{t_1}^{t_2} \left[s_i(t) - u_i(t)\right]^2 \, dt}. \tag{3.13}$$

In practice, for P waveforms, we use the vertical and radial components to compute the normalization factors, whereas for SH waveforms, we use the transverse component. The normalization factors are computed at the beginning of each inversion and remain the same during iterations.

Figure 3.5 illustrates the effect of event normalization on the vertical component of P waveforms recorded by station PE32 (southern Peru), filtered between 0.04 Hz and 0.10 Hz. The first column shows the observed and synthetic waveforms corresponding to the different teleseismic sources. The absolute amplitudes differ widely from one event to another. For example, the 2010/04/04 event has the largest amplitude and the 2010/12/08 event the smallest (Fig. 3.5a). Their amplitudes differ by about a factor of 10. These large amplitude differences are also reflected in large amplitude variations in the corresponding data residuals (Fig. 3.5b). After normalization, the data residuals now have similar amplitudes (Fig. 3.5c). From section 2.3.1 and equation 2.66, the gradients are proportional to the convolution of the forward and adjoint wavefields. To check the effectiveness of event normalization, we thus compute the convolution of the synthetic trace with the corresponding data residual. Without normalization, this convolution scales roughly with the amplitude of the traces (Fig. 3.5d) while after normalization, the convolutions now have similar amplitudes (Fig. 3.5e). Fig. 3.6 illustrates the effect of event normalization on the radial component (Fig. 3.6), with similar behavior to the vertical component.

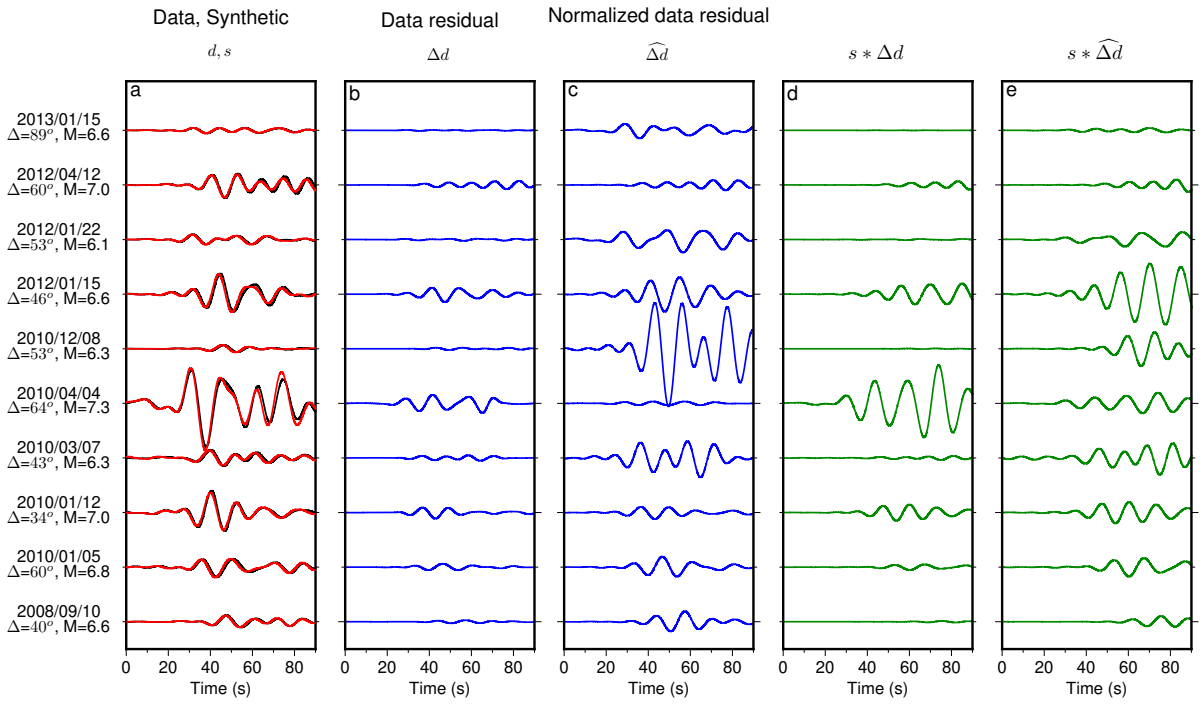


Figure 3.5: Example that demonstrates the effects of event normalization for P waveforms. Each trace represents one event recorded by the station PE32 in Peru dataset. All the traces are filtered between 0.04 and 0.10 Hz. (a) Data and synthetic in black and red, respectively. (b) Data residual. (c) Normalized data residual after applying the scaling factor 3.13. (d) Convolution of the synthetics and the data residuals. (e) Convolution of the synthetics and the normalized data residuals.

We further illustrate the impact of source normalization by showing the $V_S$ gradients of all the events in the Peruvian dataset (the detailed inversion procedure is described in Chapter 6) computed in the initial smooth 1-D model. Figure 3.7 and 3.8 shows these gradients without and with normalization, respectively. The unnormalized gradients show significant variations in amplitude. Note that the same color table has been used for these two figures. For example, P and SH gradients for the 2010/01/12 event are very large and completely dominate the total gradient (obtained by summing all the individual event gradients) shown in Fig. 3.9a). On the other hand, after applying the source normalization, all the gradients now have comparable amplitudes (Fig. 3.8). The total $V_S$ gradient looks more regular and less contaminated by artifacts (Fig. 3.9b).

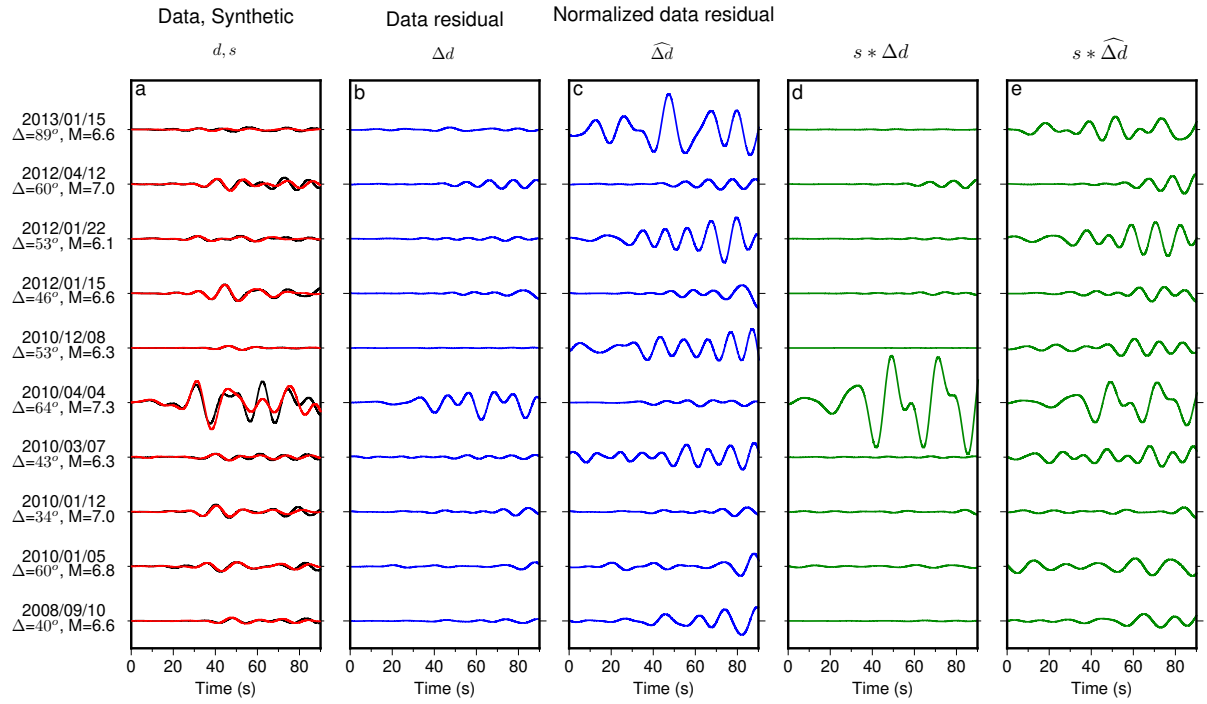


Figure 3.6: Same for Fig. 3.5 but for radial component.

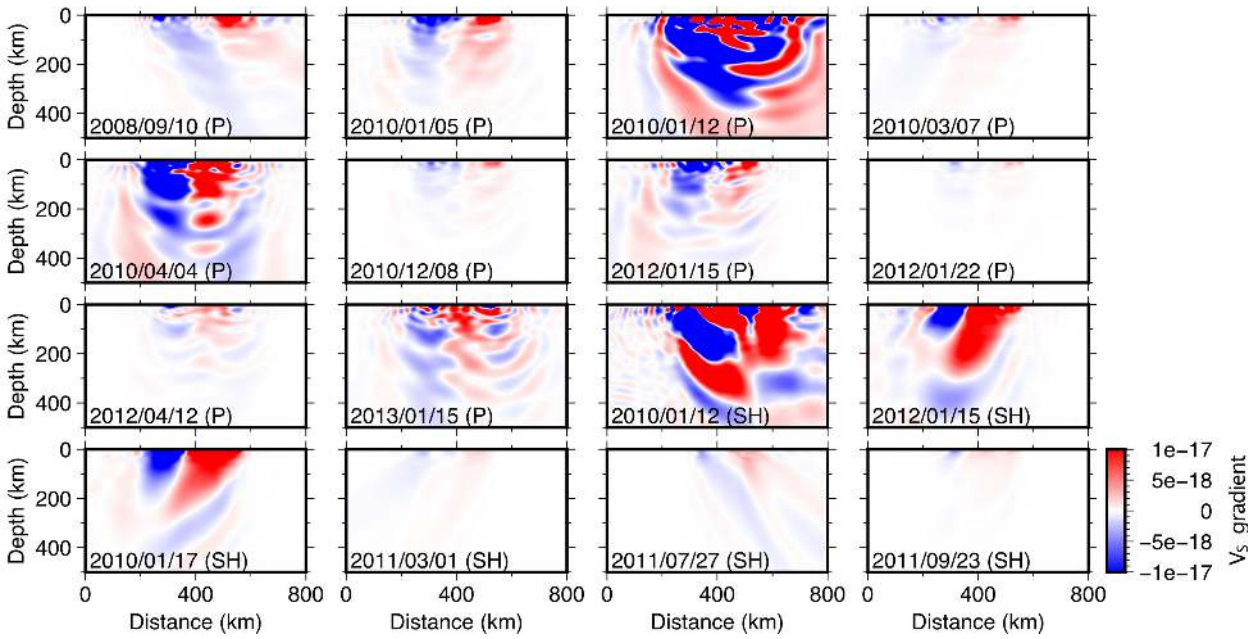


Figure 3.7: $V_S$ gradient for each teleseismic event computed in the initial smooth 1-D model. Some gradients have particularly large amplitudes, for example, the 2010/01/12 event.

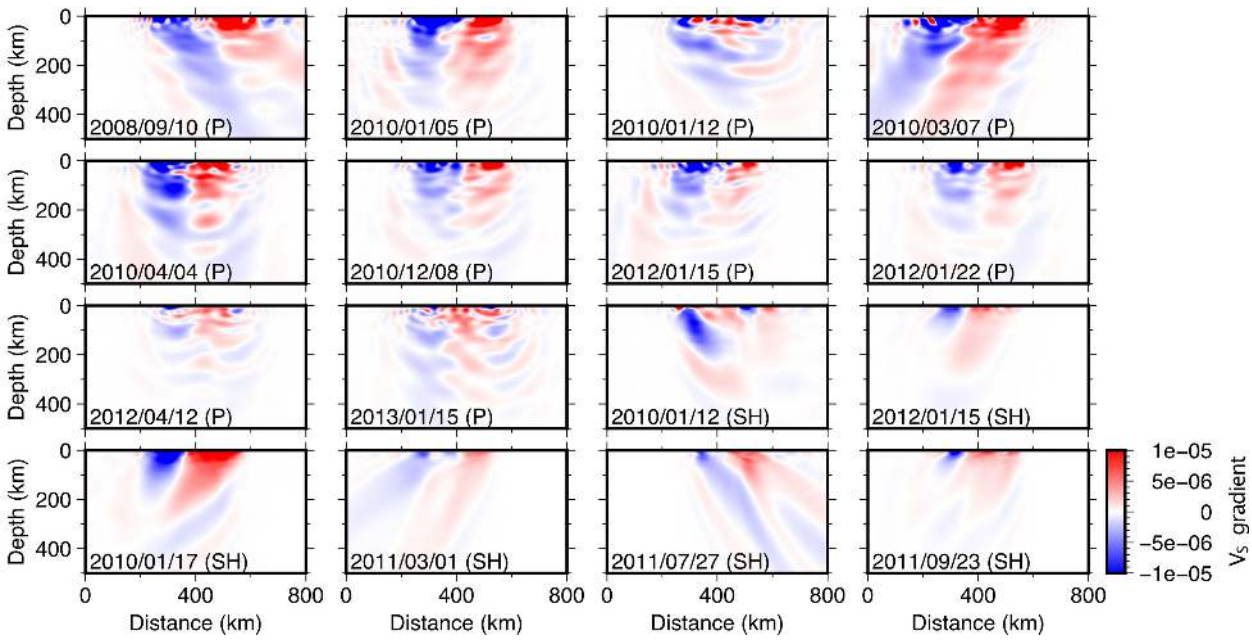


Figure 3.8: Same as Fig. 3.7 but with event normalization. The different gradients now have similar amplitudes.

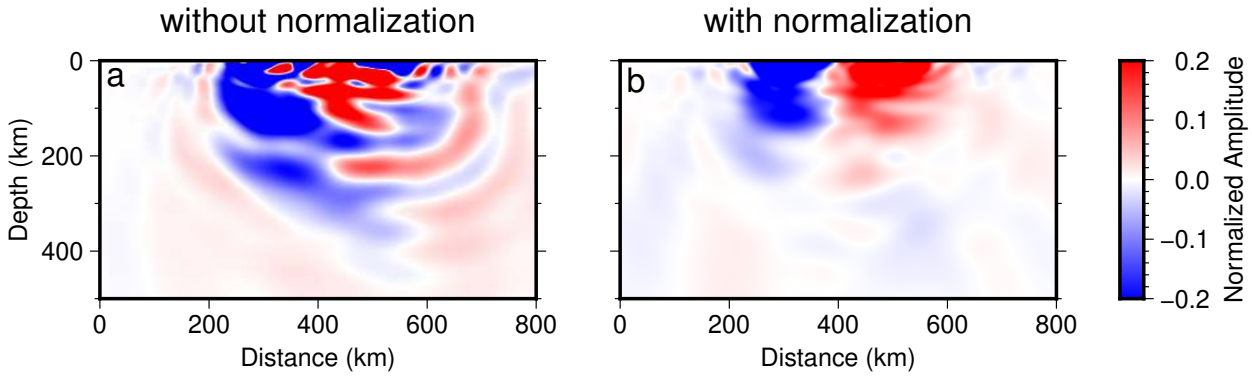


Figure 3.9: Comparison of total $V_S$ gradient with and without event normalization. (a) $V_S$ gradient obtained by summing all the individual $V_S$ gradients shown in Figure 3.7 (without event normalization). (b) $V_S$ gradient obtained by summing all the individual $V_S$ gradients shown in Figure 3.8 (with event normalization).

# Chapter 4. A consistent multi-parameter Bayesian full waveform inversion scheme for imaging heterogeneous isotropic elastic media

This chapter has been already published in Geophysical Journal International as:

Li-Yu Kan, Sébastien Chevrot and Vadim Monteiller, A consistent multi-parameter Bayesian full waveform inversion scheme for imaging heterogeneous isotropic elastic media, *Geophysical Journal International*, 232(2): 864-883, 2023. doi: 10.1093/gji/ggac363

## Contents

4.1 Summary ..... **73**
4.2 Introduction ..... **73**
4.3 Formulation of a consistent Bayesian full waveform inversion problem ..... **76**
4.3.1 Model parameterization and consistency of the inverse problem .... 77
4.3.2 The model covariance matrix for a 3-D isotropic Earth model ..... 80
4.4 Numerical experiments ..... **84**
4.4.1 Description of the model ..... 84
4.4.2 Computation of synthetic seismograms with the FK/SEM hybrid method ..... 86
4.4.3 Results ..... 88
4.5 Application to real data: the southern Peru transect ..... **97**
4.6 Discussion ..... **100**
4.6.1 Tuning FWI hyperparameters: model parameterization, standard deviations, correlation coefficients, and smoothing length ..... 100
4.6.2 Joint inversions of P and SH events ..... 102
4.7 Conclusions ..... **105**
Appendix 4.A Decomposition of the model covariance matrix ..... **108**
Appendix 4.B Change of model parameterization and transformation of the model covariance matrix ..... **110**

## 4.1 Summary

The inversion of complete seismic waveforms can potentially fully constrain the elastic properties of Earth's interior. However, the sensitivity to density decreases with depth whereas the spatial resolution of compressional velocity models is usually coarser than shear velocity models, owing to the longer wavelength of P waves compared with S waves. As a result, reconstructed models of density and seismic velocities are usually characterized by markedly different and uneven spatial resolutions. Here we introduce the well-documented correlations between density and seismic velocities by considering a non-diagonal 3-D model covariance matrix, in which the spatial correlations of elastic properties are described with an exponential covariance function. The inverse of such a model covariance matrix is easy to compute, and we thus have all the ingredients to build a consistent Bayesian full waveform inversion scheme. We show that accounting for the correlations between density and compressional velocities, and between compressional and shear velocities, leads to dramatic improvements on the reconstructed models of density and seismic velocities, and of the $V_P/V_S$ ratio. This new imaging approach opens new perspectives for refining tomographic images of density and seismic velocities in the lithosphere and upper mantle at the regional scale by full waveform inversion of teleseismic body waves.

## 4.2 Introduction

Because seismic tomography still heavily relies on the exploitation of the phase of body and/or surface waves it thus mostly constrains the shear and/or compressional velocities in Earth's interior. With the continuous improvement of computer power, full waveform inversions (FWI) have now become feasible, at least over regional domains of a few degrees in size, opening new perspectives for obtaining finely resolved models of seismic velocities (Wang et al., 2016), density (Beller et al., 2018) and even seismic anisotropy (Beller and Chevrot, 2020). When inverting complete teleseismic wavefields, the different seismic phases have very different contributions to the model reconstruction. Direct transmitted waves mainly

constrain the long wavelengths of the velocity model, whereas later arrivals such as reflected and converted phases are sensitive to sharp velocity and density gradients. Consequently, if full waveform inversion of teleseismic P waves allows us to simultaneously reconstruct 3-D models of $V_P$, $V_S$, and density, which are key parameters to get crucial insight into the compositional and thermal state of the lithosphere, these models are characterized by very contrasted spatial resolutions.

This contribution aims at investigating the beneficial effects for FWI of a new regularization approach in the model space that imposes a degree of correlation between density, $V_P$, and $V_S$, the three parameters that describe isotropic elastic Earth models. Whereas these correlations are well known and already found important applications in various branches of geophysics, they were so far largely ignored in tomographic studies.

The correlation between compressional wave velocity and density has been particularly well studied, and several simple laws relating $V_P$ to density have been proposed (e.g. Nafe and Drake, 1957; Birch, 1961; Brocher, 2005b). Such relations were important to build the first reference Earth model, because normal mode and dispersion data are insufficient to fully resolve the density structure. For example, the PREM model (Dziewonski and Anderson, 1981) was required to satisfy Birch's law in the upper mantle. These relationships were also successfully applied for modeling the gravity field, by predicting a density model from a tomographic model (e.g. Wang et al., 2016; Martin et al., 2021), or for joint/cooperative inversions of seismic and gravity data (e.g. Lines et al., 1988; Lees and VanDecar, 1991). Mantle circulation models, which provide key insights into plate tectonics and the evolution of the Earth's interior, also rely on density models derived from tomographic models (e.g. Becker and O'Connell, 2001; Simmons et al., 2006). The strong correlation of P and S station corrections (e.g Robertson and Woodhouse, 1997) provides a direct observational evidence for the correlation between lateral variations of $V_P$ and $V_S$. The global 2.85 average ratio of S to P corrections leads to $\nu = \partial \ln V_S / \partial \ln V_P = 1.6$ in the upper mantle, assuming a constant $V_P/V_S$ of 1.8. This result is in rather good agreement with the value obtained at

the top of the lower mantle in a later global scale statistical analysis of P and S wave travel times (Bolton and Masters, 2001). Experimental mineral physics data predict variations of $\nu$ in the upper mantle from 1.3 at low temperature to 2.2 at high temperature, accounting for the strong effect of anelasticity (Cammarano et al., 2003). In contrast, the $V_P/V_S$ ratio is more variable. For crustal rocks, it is mainly controlled by the silica content (Christensen, 1996). For mantle rocks, this ratio varies with the Mg# (e.g. Lee, 2003) or with the fraction of partial melt (e.g. Watanabe, 1993).

To summarize this brief overview of the literature, the strong correlations between density and seismic velocities are now clearly established and widely exploited. However, tomographic studies still often rely on the simplifying assumption that model parameters are not correlated. This is a major source of inconsistencies and biases in multi-parameter tomographic imaging.

The purpose of this study is to demonstrate that accounting for the well established correlations between density and $V_P$ and between $V_P$ and $V_S$ improves recovered models of $V_S$, density, as well as $V_P/V_S$ ratio with FWI. The paper is organized as follows. We first recall in section 2 the Bayesian formulation of a waveform inversion problem for imaging isotropic elastic media. We then discuss the issue of how to make the inversion results consistent, i.e. independent of the choice of model parameterization. The key is to consider non-zero off-diagonal terms in the model covariance matrix, which describe the correlation between model parameters. Another important practical ingredient is to consider an exponential kernel to describe the spatial correlation functions because such a kernel leads to an inverse covariance matrix that is a simple differential operator. The impact of accounting for the correlations between model parameters is then investigated through a number of synthetic inversion experiments performed in a simplified continental subduction model in section 3. These tests are performed with a hybrid numerical method that couples the frequency-wavenumber method to model the incident wavefields with the spectral-element method to compute complete synthetic seismograms in a 3-D regional model. The results demonstrate

that the correlations between model parameters compensate for the poor sensitivity to deep density and shear velocity anomalies when only teleseismic P waves are inverted. We also demonstrate that the final shear velocity model is comparable to the one that would have been obtained after adding teleseismic SH waveforms to the dataset. This suggests that the complete model covariance matrix can also compensate at least to some extent the lack of shear waveforms and that finely resolved models of density, $V_P$, and $V_S$ can be obtained with the sole inversion of teleseismic P waveforms. A preliminary application to long period teleseismic P waveforms recorded along a transect deployed in southern Peru is then presented and discussed in section 4.

### 4.3 Formulation of a consistent Bayesian full waveform inversion problem

The problem of full waveform inversion (FWI) is to find the model $\mathbf{m}$ that minimizes the objective function (Tarantola, 2005)

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{obs})^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{obs}) + \frac{\lambda}{2}(\mathbf{m} - \mathbf{m}_{prior})^t \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m} - \mathbf{m}_{prior}), \qquad (4.1)$$

where $\mathbf{g}$ is the forward wave equation operator, $\mathbf{C_D}$ the data covariance matrix, $\mathbf{C_M}$ the a priori model covariance matrix, and $\lambda$ the regularization parameter introduced to balance the two contributions of the objective function. The model covariance matrix $\mathbf{C_M}$ describes the a priori information that we have on the model parameters. In this study, we will illustrate the methodology by focusing on the inversion of complete teleseismic waveforms to obtain 3-D models of the isotropic elastic structure of the Earth. However, its domain of application is more general and it can be generalized to any type of multi-parameter inversion.

The minimum of function (4.1) can be found with an iterative algorithm (e.g. Tarantola and Valette, 1982; Tarantola, 2005), in which the model $\mathbf{m_k}$ at iteration $k$ is perturbed by the quantity

$$\Delta\mathbf{m}_k = -\alpha_k \mathbf{H}_k^{-1} \cdot \gamma_\mathbf{k}, \qquad (4.2)$$

where $\alpha$ is the step length, $\mathbf{H}^{-1}$ the inverse Hessian, and $\gamma_\mathbf{k}$ the gradient of cost function

(4.1) at iteration $k$. This gradient is given by

$$\gamma_{\mathbf{k}} = \frac{\partial \chi_k}{\partial \mathbf{m}} = \mathbf{G}_k^t \mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m_k}) - \mathbf{u}_{obs}) + \lambda \mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m_k} - \mathbf{m}_{prior}), \tag{4.3}$$

with $\mathbf{G_k}$ the derivative or Jacobian operator given by

$$\mathbf{G}_k = \frac{\partial \mathbf{g}}{\partial \mathbf{m}}(\mathbf{m_k}). \tag{4.4}$$

In tomographic applications based upon complete seismic waveforms, the first contribution to the gradient can be computed with the adjoint method (Tromp et al., 2005) whereas the second term is usually discarded.

The L-BFGS method (Nocedal and Wright, 2006) is a quasi-Newton method that has rapidly gained in popularity to solve FWI problems. Its main advantages are that it behaves like a preconditioned steepest descent during the first iteration and has a fast convergence comparable to a Newton method. In addition, the L-BFGS method does not require computing and storing the inverse Hessian. Instead, the L-BFGS method iteratively builds a finite difference approximation of the inverse Hessian from the stored gradients and models obtained at previous iterations. For a more detailed description of the L-BFGS method and its application to FWI the reader is referred to Monteiller et al. (2015).

### 4.3.1 Model parameterization and consistency of the inverse problem

Isotropic elastic media are described by three parameters, the density $\rho$, the compressional velocity $V_P$ and the shear velocity $V_S$. Alternatively, one may also consider other parameterizations such as $(\rho, I_P, I_S)$, with $I_P$ and $I_S$ the P and S impedances respectively, $(\rho, \lambda, \mu)$, with $\lambda$ and $\mu$ the two Lamé parameters, $(\rho, V_P, V_P/V_S)$, or $(\rho, \kappa, \mu)$, with $\kappa$ the bulk modulus. In principle, these different parameterizations are equivalent, because any set of parameters can be recombined to reconstruct say the $(\rho, V_P, V_S)$ model. However, in practice, owing to the overlap of scattering diagrams for the different parameters (e.g. Wu

and Aki, 1985; Operto et al., 2013; Gholami et al., 2013), the strong non-linearity of full waveform inversion, and the presence of noise contamination in the data, the reconstructed tomographic models strongly depend on the chosen parameterization and regularization.

Another issue stems from the physical units chosen for the model parameters. To define a norm in the model space, model parameters need to be normalized by their standard deviation, which are stored on the diagonal of the model covariance matrix. The use of reduced centered variables allows us to work in an adimensional model space, which makes the inverse problem better conditioned. In that case, the regularization term is a simple damping of the norm of the model. Let us define the auxiliary model

$$\hat{\mathbf{m}}_\mathbf{k} = \mathbf{C}_\mathbf{M}^{-\frac{1}{2}}(\mathbf{m}_k - \mathbf{m}_{prior}). \tag{4.5}$$

With this new metric, the normalized gradient becomes

$$\frac{\partial \chi_k}{\partial \hat{\mathbf{m}}} = \mathbf{C}_\mathbf{M}^{\frac{1}{2}}\mathbf{G}_k^t\mathbf{C}_\mathbf{D}^{-1}(\mathbf{g}(\mathbf{m_k}) - \mathbf{u}_{obs}) + \lambda\hat{\mathbf{m}}_\mathbf{k}. \tag{4.6}$$

This formulation leads to a stable regularization of the inverse problem by smoothing the gradient with the model covariance matrix. Another advantage of this approach is that we can determine directly the model perturbations at each node of the spectral-element mesh, thereby avoiding back-and-forth projections in a Cartesian tomographic grid, as in Wang et al. (2016). This results in a simplified and versatile inversion algorithm which can be applied in arbitrarily complex 3-D spectral-element meshes. For example, in order to account for the effects of surface topography we just need to deform the upper edge of the mesh accordingly, which greatly simplifies the inversion algorithm.

For convenience, it is often assumed that the model covariance matrix is purely diagonal. This is a source of inconsistencies in multi-parameter inversions. This problem has been clearly illustrated by Babuška and Cara (1991) with the inversion of the dispersion of the fundamental and higher modes of both Love and Rayleigh waves to determine a vertical

transversely isotropic model of the upper mantle. They performed the first inversion using a transversely isotropic model parametrization in terms of $\beta_V$ and $\xi = (\beta_H/\beta_V)^2$, with $\beta_V$ and $\beta_H$ the velocity of horizontally propagating S waves polarized respectively horizontally and vertically. However, a new inversion using parameters $\beta_H$ and $\beta_V$ lead to a drastically different model of the $\beta_H/\beta_V$ ratio. To make the inversion results consistent, they pointed out that it is necessary to introduce the matrix $\mathbf{T}$ that transforms model perturbations $(\delta\beta_V, \delta\xi)$ to $(\delta\beta_V, \delta\beta_H)$, defined by

$$\begin{bmatrix} \delta\beta_V \\ \delta\beta_H \end{bmatrix} = \mathbf{T} \cdot \begin{bmatrix} \delta\beta_V \\ \delta\xi \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \sqrt{\xi} & \frac{\beta_V}{2\sqrt{\xi}} \end{bmatrix} \begin{bmatrix} \delta\beta_V \\ \delta\xi \end{bmatrix}. \tag{4.7}$$

If the model covariance matrix $\mathbf{C_M}$ is diagonal for the set of parameters $(\beta_V, \xi)$ then we need to use the covariance matrix $\mathbf{C'_M}$ for parameters $(\beta_V, \beta_H)$ given by

$$\mathbf{C'_M} = \mathbf{T} \cdot \mathbf{C_M} \cdot \mathbf{T}^t \tag{4.8}$$

to ensure the consistency of the inversion.

For isotropic FWI, this means that if we want to transform model parametrization $(\rho, V_P, V_S)$ with the model covariance matrix

$$\mathbf{C}_{(\rho,V_P,V_S)} = \begin{bmatrix} \sigma^2_\rho & 0 & 0 \\ 0 & \sigma^2_{V_P} & 0 \\ 0 & 0 & \sigma^2_{V_S} \end{bmatrix} \tag{4.9}$$

to parameterization $(\rho, V_P, V_P/V_S)$, according to

$$\begin{bmatrix} \delta\rho \\ \delta V_P \\ \delta\frac{V_P}{V_S} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{V_S} & -\frac{V_P}{V_S^2} \end{bmatrix} \begin{bmatrix} \delta\rho \\ \delta V_P \\ \delta V_S \end{bmatrix} \tag{4.10}$$

then we need to define the new model covariance matrix as

$$\mathbf{C}_{(\rho,V_P,\frac{V_P}{V_S})} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_P}^2 & \frac{1}{V_S}\sigma_{V_P}^2 \\ 0 & \frac{1}{V_S}\sigma_{V_P}^2 & \frac{1}{V_S^2}\sigma_{V_P}^2 + \frac{V_P^2}{V_S^4}\sigma_{V_S}^2 \end{bmatrix} \tag{4.11}$$

in order to get consistent inversion results. From these simple considerations, we thus see that the non-diagonal terms of the model covariance matrix, which describe the correlations between model parameters, are key ingredients for the formulation of a consistent inverse problem. Whereas model parametrization can be chosen at will, the assumptions that we make (implicitly or not) on the degree of correlation between model parameters can potentially strongly impact the results of inversions.

### 4.3.2 The model covariance matrix for a 3-D isotropic Earth model

The main impediment to introduce a complete non-diagonal model covariance matrix into an inversion algorithm stems from much increased computational cost and algorithmic complexity, which can indeed become prohibitive for very large 3-D problems.

Let us consider a model covariance matrix of the form

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2\mathbf{C} & r_{\rho,V_P}\sigma_\rho\sigma_{V_P}\mathbf{C} & r_{\rho,V_S}\sigma_\rho\sigma_{V_S}\mathbf{C} \\ r_{\rho,V_P}\sigma_\rho\sigma_{V_P}\mathbf{C} & \sigma_{V_P}^2\mathbf{C} & r_{V_P,V_S}\sigma_{V_P}\sigma_{V_S}\mathbf{C} \\ r_{\rho,V_S}\sigma_\rho\sigma_{V_S}\mathbf{C} & r_{V_P,V_S}\sigma_{V_P}\sigma_{V_S}\mathbf{C} & \sigma_{V_S}^2\mathbf{C} \end{bmatrix}, \tag{4.12}$$

where the $r_{i,j}$ are the correlation coefficients between parameters $i$ and $j$, $\mathbf{C}$ the spatial covariance kernel, $\sigma_\rho$ the standard deviation of density, $\sigma_{V_P}$ the standard deviation of $V_P$, and $\sigma_{V_S}$ the standard deviation of $V_S$.

This model covariance matrix can be rewritten

$$\mathbf{C_M} = \mathbf{\Sigma S R S \Sigma}, \tag{4.13}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\rho \mathbf{I} & 0 & 0 \\ 0 & \sigma_{V_P} \mathbf{I} & 0 \\ 0 & 0 & \sigma_{V_S} \mathbf{I} \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \mathbf{C}^{\frac{1}{2}} & 0 & 0 \\ 0 & \mathbf{C}^{\frac{1}{2}} & 0 \\ 0 & 0 & \mathbf{C}^{\frac{1}{2}} \end{bmatrix}, \tag{4.14}$$

and

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & r_{\rho,V_P} \mathbf{I} & r_{\rho,V_S} \mathbf{I} \\ r_{\rho,V_P} \mathbf{I} & \mathbf{I} & r_{V_P,V_S} \mathbf{I} \\ r_{\rho,V_S} \mathbf{I} & r_{V_P,V_S} \mathbf{I} & \mathbf{I} \end{bmatrix}. \tag{4.15}$$

The matrix $\mathbf{R}$ encodes the correlations between model parameters. The decomposition of this matrix when the correlations between the different pairs are equal is detailed in Appendix A. In the general case, i.e. when the correlation coefficients are not equal, the singular value decomposition of $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^t, \tag{4.16}$$

with $\mathbf{V}$ the matrix containing the orthogonal eigenvectors of $\mathbf{R}$. Using the decomposition of $\mathbf{C_M}$, we can derive the expression of the inverse covariance matrix

$$\mathbf{C}_\mathbf{M}^{-1} = \mathbf{\Sigma}^{-1} \mathbf{S}^{-1} \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{V}^t \mathbf{S}^{-1} \mathbf{\Sigma}^{-1}, \tag{4.17}$$

from which we can deduce that

$$\mathbf{C}_\mathbf{M}^{\frac{1}{2}} = \mathbf{V} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{V}^t \mathbf{S} \mathbf{\Sigma} = \mathbf{\Sigma} \mathbf{S} \mathbf{V} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{V}^t, \tag{4.18}$$

and

$$\mathbf{C}_\mathbf{M}^{-\frac{1}{2}} = \mathbf{\Sigma}^{-1} \mathbf{S}^{-1} \mathbf{V} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{V}^t = \mathbf{V} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{V}^t \mathbf{S}^{-1} \mathbf{\Sigma}^{-1}. \tag{4.19}$$

The operator $C(\mathbf{r}, \mathbf{r}')$ describes the correlation of model parameters at positions $\mathbf{r}$ and $\mathbf{r}'$. Correlation kernels can be seen as smoothing operators and their inverse as roughening

operators (Oliver, 1998). As can be deduced from Equations (4.5) and (4.6), both the $\mathbf{C_M^{\frac{1}{2}}}$ and $\mathbf{C_M^{-\frac{1}{2}}}$ operators are needed for the implementation of a consistent full waveform inversion algorithm. The application of the model covariance matrix to a model vector thus involves three steps: First, normalization by the standard deviation with the diagonal matrix $\mathbf{\Sigma}$, as in (4.5), followed by smoothing with the spatial filter $\mathbf{S}$, and finally recombination of model parameters to account for their physical correlation with the matrix $\mathbf{R}$.

In tomographic inversions, Gaussian kernels are often considered (e.g. Nataf et al., 1986) but they strongly taper the short wavelengths of the model. Since the main motivation of full waveform inversion is to refine the spatial resolution of tomographic images, we thus need to use correlation kernels that decay less rapidly at high wavenumbers, such as exponential (or Laplacian) kernels (e.g. Oliver, 1998; Trinh et al., 2017; Araujo et al., 2021). Because Gaussian or exponential kernels have a broad support, convolving these filters over a large 3-D grid can be time consuming and also difficult to implement when the computational grid is distributed with domain decomposition. In addition, as pointed out by Oliver (1998), computation (and storage) of the full inverse covariance matrix is a challenge. However, inverse exponential kernels can be defined analytically with finite difference operators (e.g. Oliver, 1998; Tarantola, 2005; Trinh et al., 2017), an interesting property that has been already exploited to regularize tomographic inversions (e.g. Potin, 2016; Araujo et al., 2021).

The normalized 3-D exponential filter given by

$$C(\mathbf{r}, \mathbf{r}') = e^{-\sqrt{\frac{(x-x')^2}{L_x} + \frac{(y-y')^2}{L_y} + \frac{(z-z')^2}{L_z}}}, \tag{4.20}$$

with $L_x$, $L_y$, and $L_z$ the correlation (or smoothing) lengths along respectively the $x$, $y$, and $z$ directions, admits the simple inverse operator

$$C^{-1}(\mathbf{r}, \mathbf{r}') = \frac{1}{8\pi L_x L_y L_z} \left(\mathbf{I} - \Delta_L\right)^2, \tag{4.21}$$

with

$$\Delta_L = L_x^2 \partial_x^2 + L_y^2 \partial_y^2 + L_z^2 \partial_z^2 \tag{4.22}$$

the Laplacian operator (Oliver, 1998; Trinh et al., 2017). From Eq. (4.21), we deduce that

$$C^{-\frac{1}{2}} = \frac{1}{\sqrt{8\pi L_x L_y L_z}} \left(\mathbf{I} - \Delta_L\right). \tag{4.23}$$

The $C^{-\frac{1}{2}}$ is a local differential operator that only involves the identity and Laplacian operators. It is thus easy to compute the application of this operator to a model vector, for example with a spectral-element method (SEM), with the advantage that this can be implemented on the same mesh that is used for the resolution of the weak form of the wave equation. By doing so, we also avoid having to project back and forth the model to a regular tomographic Cartesian grid, as in our earlier implementations of the inversion algorithm (Monteiller et al., 2015; Wang et al., 2016). Using exponential kernels, it is thus straightforward to introduce a complete non-diagonal inverse model covariance operator in full waveform inversion. As emphasized by Trinh et al. (2017), in the case of large problems involving domain decomposition, it may be easier and more efficient to compute the filtered model $\mathbf{m_f} = C^{\frac{1}{2}}\mathbf{m}$ by resolving the linear system

$$C^{-\frac{1}{2}}\mathbf{m_f} = \mathbf{m}. \tag{4.24}$$

Because $C^{-\frac{1}{2}}$ is sparse, only the non-zero elements need to be stored, and this system can be resolved iteratively with a conjugate gradient method (Trinh et al., 2017). However, as mentioned by Trinh et al. (2017), a simpler matrix-free conjugate gradient algorithm can be implemented with the SEM. In that case, we only need to compute matrix-vector products involving the weak form of the Laplacian, which is already stored by the SEM. Therefore, the vector $\mathbf{m_f}$ is easy to compute, for negligible additional memory and computational costs.

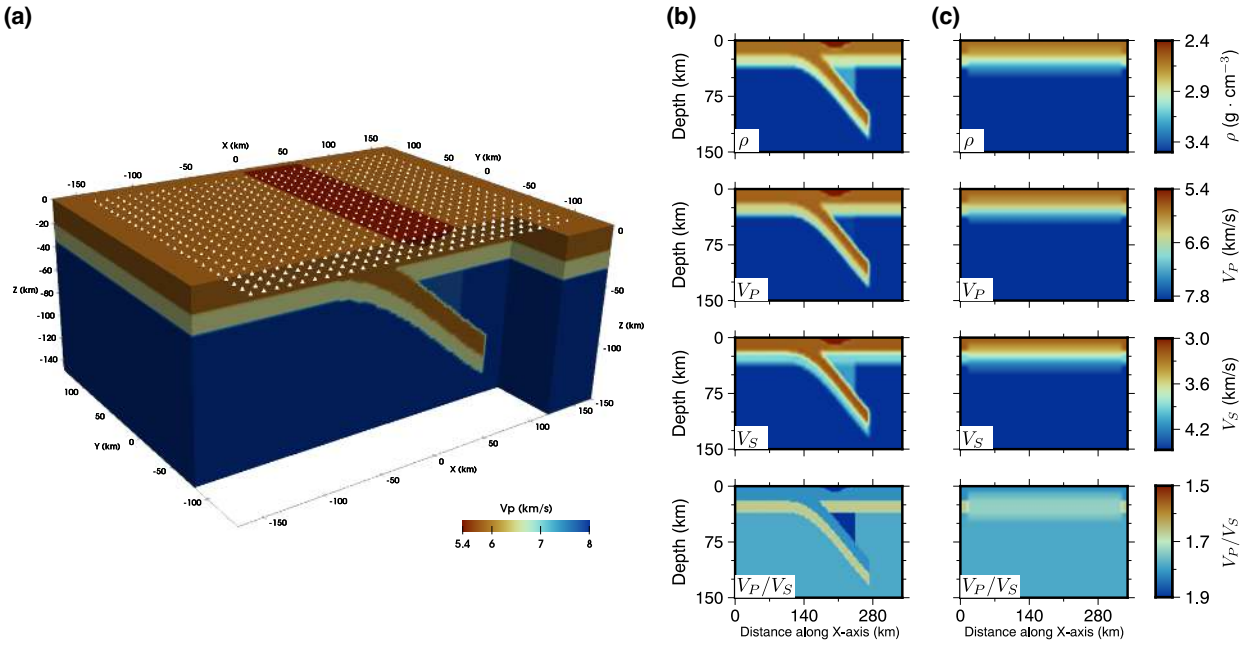


Figure 4.1: The continental subduction model used in the synthetic tests. The model is parameterized by the values of density, $V_P$, and $V_S$ at each node of the spectral-element mesh. We show here the $V_P$ model, expressed in km/s, and the positions of the receivers located on the surface (white triangles). The slab extends down to 130 km depth, and the width of the fore-arc basin is 50 km.

## 4.4 Numerical experiments

We now perform several numerical tests to assess the influence of using non-diagonal model covariance matrices in full waveform inversions.

### 4.4.1 Description of the model

We build and mesh a 2.5-D model of continental subduction (Fig. 4.1) composed of five distinct units: upper crust, lower crust, upper mantle, fore-arc basin, and mantle wedge. Inside each domain the values of density, $V_P$, and $V_S$ are constant (see table 4.1). These values were chosen such as to be representative of the different geological units and standard continental crust and mantle. We impose a large $V_P/V_S$ ratio in the fore-arc basin and in the mantle wedge, where hydration of mantle rocks is expected to strongly decrease the shear velocity. This model is similar to the one considered in Monteiller et al. (2021).

Table 4.1: Values of model parameters inside each structural unit of the model shown in Figure 4.1. Density $\rho$ is in g·cm$^{-3}$, $V_P$ and $V_S$ in km·s$^{-1}$, P and S impedances $I_P$ and $I_S$ in g·cm$^{-3}$·km·s$^{-1}$, and the Lamé parameters $\lambda$ and $\mu$ in GPa.

| Unit | $\rho$ | $V_P$ | $V_S$ | $V_P/V_S$ | $I_P$ | $I_S$ | $\lambda$ | $\mu$ |
|---|---|---|---|---|---|---|---|---|
| Forearc basin | 1.80 | 4.80 | 2.40 | 2.00 | 8.64 | 4.32 | 20.74 | 10.37 |
| Upper crust | 2.60 | 5.80 | 3.20 | 1.81 | 15.08 | 8.32 | 34.22 | 26.62 |
| Lower crust | 2.90 | 6.50 | 3.90 | 1.67 | 18.85 | 11.31 | 34.31 | 44.11 |
| Mantle wedge | 3.20 | 7.80 | 4.10 | 1.90 | 24.96 | 13.12 | 87.10 | 53.79 |
| Upper mantle | 3.50 | 8.00 | 4.48 | 1.79 | 28.00 | 15.68 | 83.51 | 70.25 |

Table 4.1 also lists the values of alternative parameters that could be considered to describe this elastic model. These parameters can be derived from the values of density and seismic velocities: $I_P = \rho V_P$, $I_S = \rho V_S$, $\lambda = \rho(V_P^2 - 2V_S^2)$, and $\mu = \rho V_S^2$. Obviously, because we consider a model composed of only five distinct homogeneous units, the variability of each parameter is rather limited. Nevertheless, examining the joint variations of the different pairs of parameters provides some insight into their correlations and thus into the a priori information that could be integrated into the inversions.

We consider the four parameterizations $(\rho, V_P, V_S)$, $(\rho, I_P, I_S)$, $(\rho, \lambda, \mu)$, and $(\rho, V_P, V_P/V_S)$. Figure 4.2 shows the plots of the joint variations of the three pairs of parameters corresponding to these parameterizations, with the corresponding correlation coefficients given at the bottom right of each diagram. The first three parameterizations are characterized by large ($> 0.86$) correlation coefficients. The correlations between model parameters are particularly strong ($> 0.96$) for the $(\rho, V_P, V_S)$ and $(\rho, I_P, I_S)$ parameterizations. In general, these plots suggest that imposing a diagonal model covariance matrix to regularize the inversion, i.e. assuming that model parameters are uncorrelated, is clearly a poor assumption that can potentially bias the results of full waveform inversions. The correlation coefficients for the three pairs of parameters are very similar for the first three parameterizations, which suggests that building the model covariance matrix with equal correlation coefficients is a reasonable assumption. In that case, the decomposition of the model covariance matrix is very simple (see

Appendix A). The last parameterization ($\rho, V_P, V_P/V_S$) displays a very different behavior. Whereas the correlation between $\rho$ and $V_P$ is strong, with a correlation coefficient of 0.97, the $V_P/V_S$ ratio is poorly correlated to either $\rho$ or $V_P$, with a correlation coefficient that is even negative. For this parameterization, assuming a diagonal model covariance matrix is thus probably less prejudicial.

### 4.4.2 Computation of synthetic seismograms with the FK/SEM hybrid method

The computational domain is discretized in a regular Cartesian grid with dimensions 340x280x150 km along the $x$, $y$, and $z$ directions, respectively. The model is meshed with 10 km hexahedral elements and Lagrange polynomials of degree 8 along the three spatial dimensions. Since the minimum shear wave velocity is 2.4 km$\cdot$ s$^{-1}$ (Table 4.1), the minimum period resolved is around 2 s.

We use a numerical hybrid method to compute the complete regional wavefield produced by an incoming teleseismic body wave (Monteiller et al., 2013, 2021). Owing to the small size of our regional grid, the curvature of both the wavefronts and of the Earth can be neglected (Monteiller et al., 2021). We can thus approximate the incident wavefields with plane waves, and use the FK/SEM hybrid modelling method as described in Monteiller et al. (2021). The FK method is particularly efficient because for teleseismic applications we only need to compute the incident wavefield corresponding to a single wavenumber. Another advantage of the FK method is that, in contrast to AxiSEM, the amount of computations scales linearly with the maximum frequency, and thus it is possible to reach high frequencies with modest computational resources. By limiting the time consuming 3-D computations inside the regional domain, hybrid methods make full waveform inversions feasible on moderate-size clusters.

We compute the synthetic seismograms of teleseismic P waves coming from four different backazimuths: 0°, 70°, 180° and 300°. The incident ray parameters of the four incoming plane waves correspond to epicentral distances of 60° for the first two, and 30° for the last two. The wavefields are recorded by a 2-D array of 720 receivers located on the surface, with a regular 10 km spacing. We consider a Gaussian source wavelet with a dominant period of

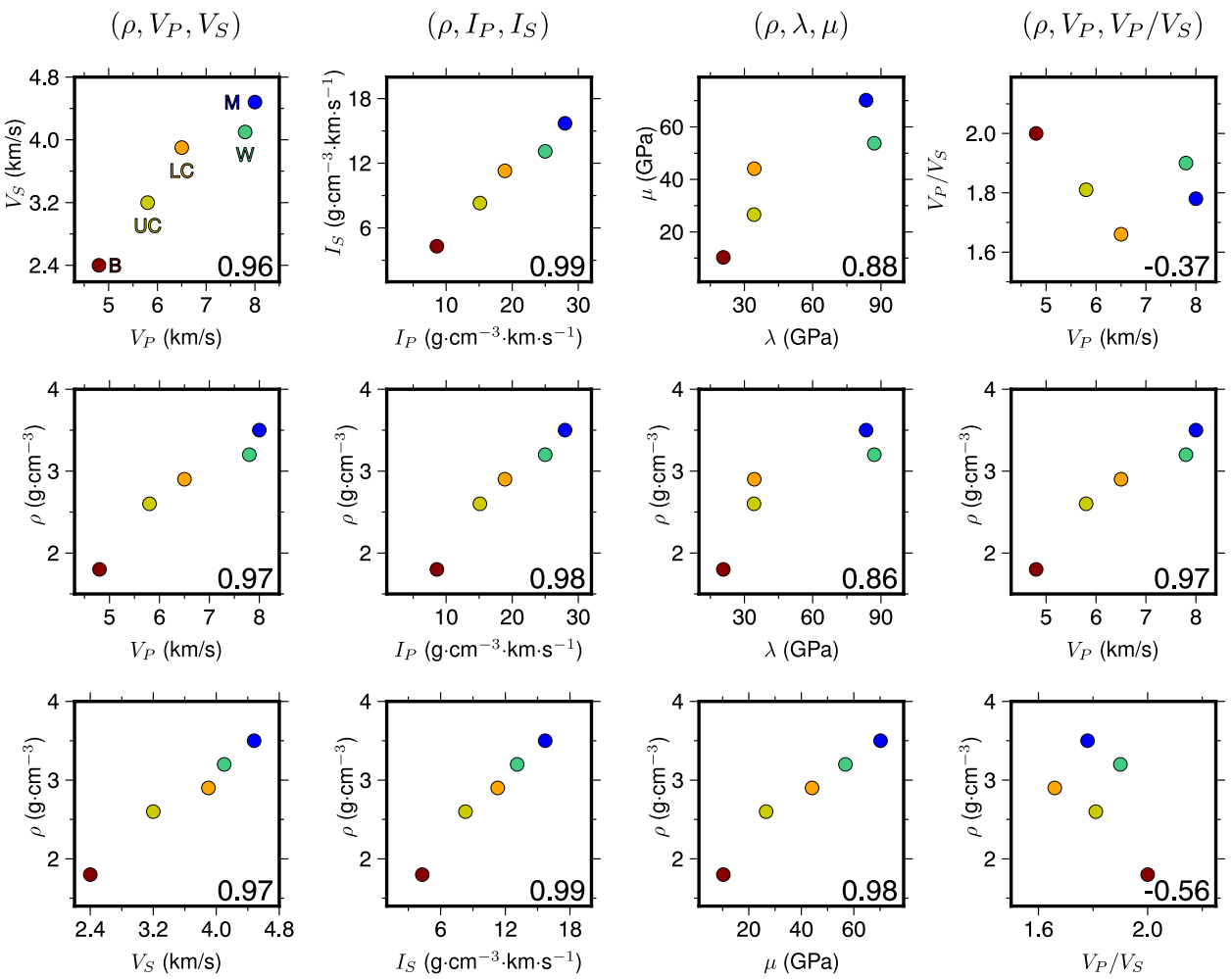


Figure 4.2: Scattering plots of joint variations of the three pairs of parameters for the four parameterizations $(\rho, V_P, V_S)$, $(\rho, I_P, I_S)$, $(\rho, \lambda, \mu)$, and $(\rho, V_P, V_P/V_S)$. The values of the parameters for the five geological units are indicated by colored circles: forearc basin (B: Brown circles), upper crust (UC: yellow circles), lower crust (LC: orange circles), mantle wedge (W: green circles), and upper mantle (M: blue circles). The parameters units are the same as those given in the caption of Table 4.1. The correlation coefficient for each pair of parameters is indicated in the lower right corner of the diagrams.

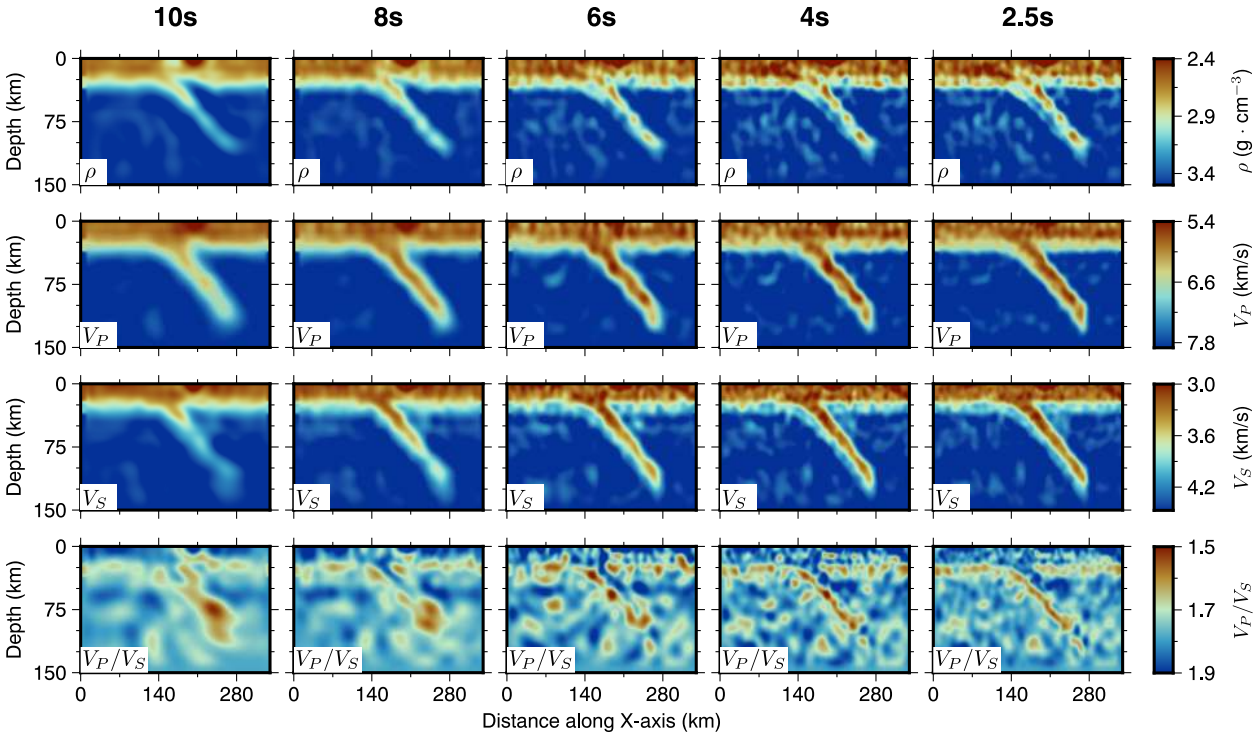


Figure 4.3: Results of multiscale FWI on noisy data with $(\rho, V_P, V_S)$ parametrization, assuming a diagonal model covariance matrix and a Laplacian smoothing length of 5 km. The first inversion starts from long period filtered in 25 s to 10 s, from which we take the model as the input of next run and decrease the cut-off period progressively to 8 s, 6 s, 4 s and finally 2.5 s.

2.5 s, which is slightly longer than the resolution of the mesh to avoid numerical noise. To investigate the impact of noise in the inversion, we also generate noisy synthetic seismograms by adding real noise time series recorded by station MLS from the French RLBP network. The amplitude of noise is tuned to get an S/N ratio of 6, a value that may seem high but that reflects the strict criteria applied to data selection in FWI applications.

### 4.4.3 Results

#### *4.4.3.1 Reference case: $(\rho, V_P, V_S)$ inversion with a diagonal model covariance matrix*

We perform a first FWI of the noisy dataset using model parameterization $(\rho, V_P, V_S)$, assuming that the parameters are uncorrelated. We thus basically reproduce the approach followed in our previous implementations of FWI on synthetic (Monteiller et al., 2015) or real

data (Wang et al., 2016) but with a regularization scheme relying on the model covariance matrix, as detailed in the previous section. This first inversion will constitute our reference test case. We briefly describe the inversion strategy here and refer the reader to Monteiller et al. (2015) for more details. We use a hierarchical iterative L-BFGS inversion, starting with a long period inversion of data filtered with a Butterworth bandpass filter between 0.04 and 0.1 Hz in order to obtain a long wavelength velocity model. For this first inversion we use a smooth 1-D model derived from the ak135 reference Earth model (Kennett et al., 1995). The value of the regularization parameter $\lambda$ is chosen at the kink of the L-curve obtained by plotting the joint variations of the data and model misfit contributions to the cost function (4.1) (Fig. 4.4). We then perform successive inversions in which we progressively decrease the cut-off period to 8 s, 6 s, 4 s, and finally 2.5 s, using at each iteration the model obtained at the previous iteration as the starting model. The correlation (or smoothing) length of the exponential correlation operator is set to 5 km along the three spatial dimensions for all the periods. At each iteration, we compute synthetic seismograms in the current model which are compared to the data to be fitted, here the synthetic seismograms are computed in the subduction model. The time windows considered for the measurements of waveform misfits start 5 s before and end 70 s after the P wave arrival. The waveform residuals are injected at the position of the receivers to compute the adjoint wavefields. The gradients corresponding to each parameter are then computed by correlating the forward wavefield with the adjoint wavefield, following the method described in Tromp et al. (2005). The standard deviations for density, $V_P$ and $V_S$ are set to 0.27 g·cm$^{-3}$, 0.65 km/s and 0.37 km/s, respectively. These standard deviations were simply estimated from the distributions of model parameters in the input model. The inversion algorithm is iterated until the inversion fails to meet the Wolfe conditions during the line search, or when the reduction of the cost function goes below 0.1%.

Figure 4.3 shows cross-sections along the $x$ axis in the $\rho$, $V_P$, and $V_S$ models obtained at each stage of the hierarchical inversions. We can observe that the spatial resolution

improves and that the interfaces get sharper when the cut-off period decreases. The $V_P$ model seems to be the best resolved, especially at depth. The deep part of the slab is well reconstructed only in the $V_P$ model. The negative anomaly related to the forearc basin has a clear signature on the density and $V_S$ models. In general, the $\rho$ and $V_S$ models are characterized by a finer resolution, but only in the shallow part of the model. For a given frequency content, the wavelengths of compressional waves are about twice that of shear waves. In addition, because we only consider teleseismic P waves here, it is the phase (and amplitude) of direct transmitted P waves that contribute to the $V_P$ model and thus mainly constrain the long wavelength $V_P$ model. In contrast, the $V_S$ model is mainly controlled by P-to-S reflections and/or conversions which are produced on sharp gradients or discontinuities in the underlying medium. This basically explains the main differences observed between the $V_P$ and $V_S$ models. Not surprisingly, the $V_P/V_S$ model derived from the $V_P$ and $V_S$ models shows many artifacts that simply result from the different spatial resolutions in these two models. The sensitivity to density mainly comes from the reverberations between the free surface and internal discontinuities such as the Moho or the slab edges. This sensitivity decreases with depth, and almost vanishes beneath 75 km depth. The density model is also contaminated by significant artifacts in the upper crust. Therefore, whereas seismic waveforms do have some sensitivity to density, in particular at shallow depth, seismic data alone are not sufficient to fully recover the density structure.

For comparison, the results of the hierarchical inversion obtained on noise-free and noisy data are shown in Figure 4.5. The degradation of spatial resolution and the contamination by small-scale artifacts are especially visible on the density and $V_S$ models.

#### *4.4.3.2 Consistent inversion results with different model parameterizations*

As noted by Nataf et al. (1986) and Babuška and Cara (1991), changing the parameterization without changing the covariance matrix accordingly leads to inconsistent inversion results. We now perform the same inversion as in the reference case, but using this time the $(\rho, V_P, V_P/V_S)$ parameterization. The results of the hierarchical inversion for this parame-

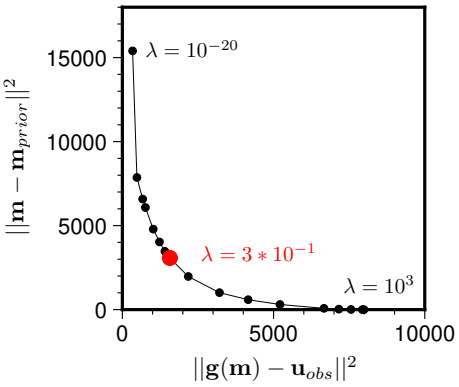


Figure 4.4: L-curve showing the variations of data and model terms of the cost function for different values of the regularization parameter $\lambda$. The optimal regularization coefficient ($\lambda = 0.3$) is chosen at the kink of the L-curve.

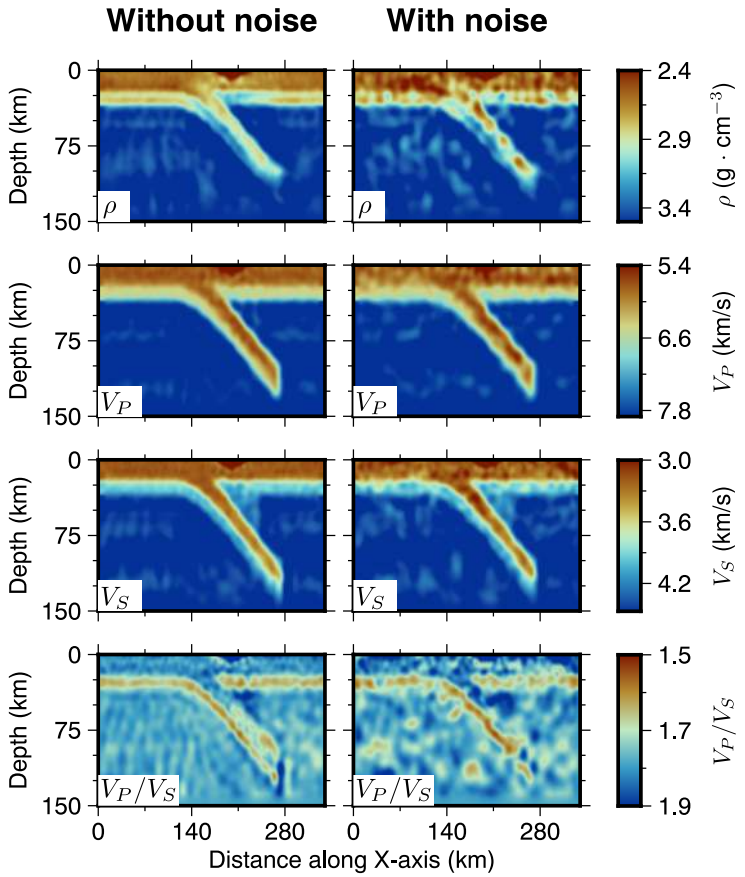


Figure 4.5: Results of FWI obtained on noise-free (Left column) and noisy (Right column) data using a diagonal model covariance matrix.

terization are shown in the third column of Figure 4.6 whereas the results obtained with the $(\rho, V_P, V_S)$ parameterization are shown in the first column of Figure 4.6. The reconstructed models obtained with the two parameterizations and diagonal model covariance matrices clearly differ. With the new parameterization $(\rho, V_P, V_P/V_S)$, the $V_P$ and $V_S$ models seem to be slightly better reconstructed, especially in the deep part of the subduction. The general improvement in the $V_P$ and $V_S$ models results from the more realistic a priori assumption that $V_P$ and $V_P/V_S$ are independent, as suggested by the last column of Figure 4.2.

We now consider the model covariance matrix given by (4.11), i.e. we assume that $V_P$ and $V_S$ are independent, but keep the inversion parameters $(\rho, V_P, V_P/V_S)$. The results, shown in the middle column of Figure 4.6, are very similar to the ones obtained with the initial $(\rho, V_P, V_S)$ parameterization (shown in the first column of Figure 4.6). In particular, the similarity of the $V_P/V_S$ models emphasizes the consistency of both $V_P$ and $V_S$ models. From these simple tests, we may thus conclude that the parameterization choice is not so important provided the model covariance matrix is transformed accordingly in order to make the inversion consistent.

#### *4.4.3.3 $(\rho, V_P, V_S)$ inversion with a non-diagonal model covariance matrix*

We now consider the $(\rho, V_P, V_S)$ parameterization and a non-diagonal model covariance matrix in the form of (4.12). We perform a hierarchical inversion in which we impose $r = 0.97$ for all the pairs of parameters for cut-off periods from 10 s to 4 s, and $r = 0.80$ in the final inversion at 2.5 s. The resulting models at each stage of this hierarchical inversion are shown in Figure 4.7. At 10 s, the $\rho$, $V_P$, and $V_S$ structures of the subducted slab and of the basin are generally well retrieved, with a noticeable blurring effect that comes from the limited spatial resolution of this low frequency inversion. The final $\rho$, $V_P$, and $V_S$ models are all nicely recovered with a sharp definition. The remaining small-scale artifacts could perhaps be reduced by additional smoothing, but this would also likely degrade the spatial resolution. Comparing these models to those previously obtained with a diagonal model covariance matrix (Fig. 4.5 right), we can thus conclude that introducing strong correlations

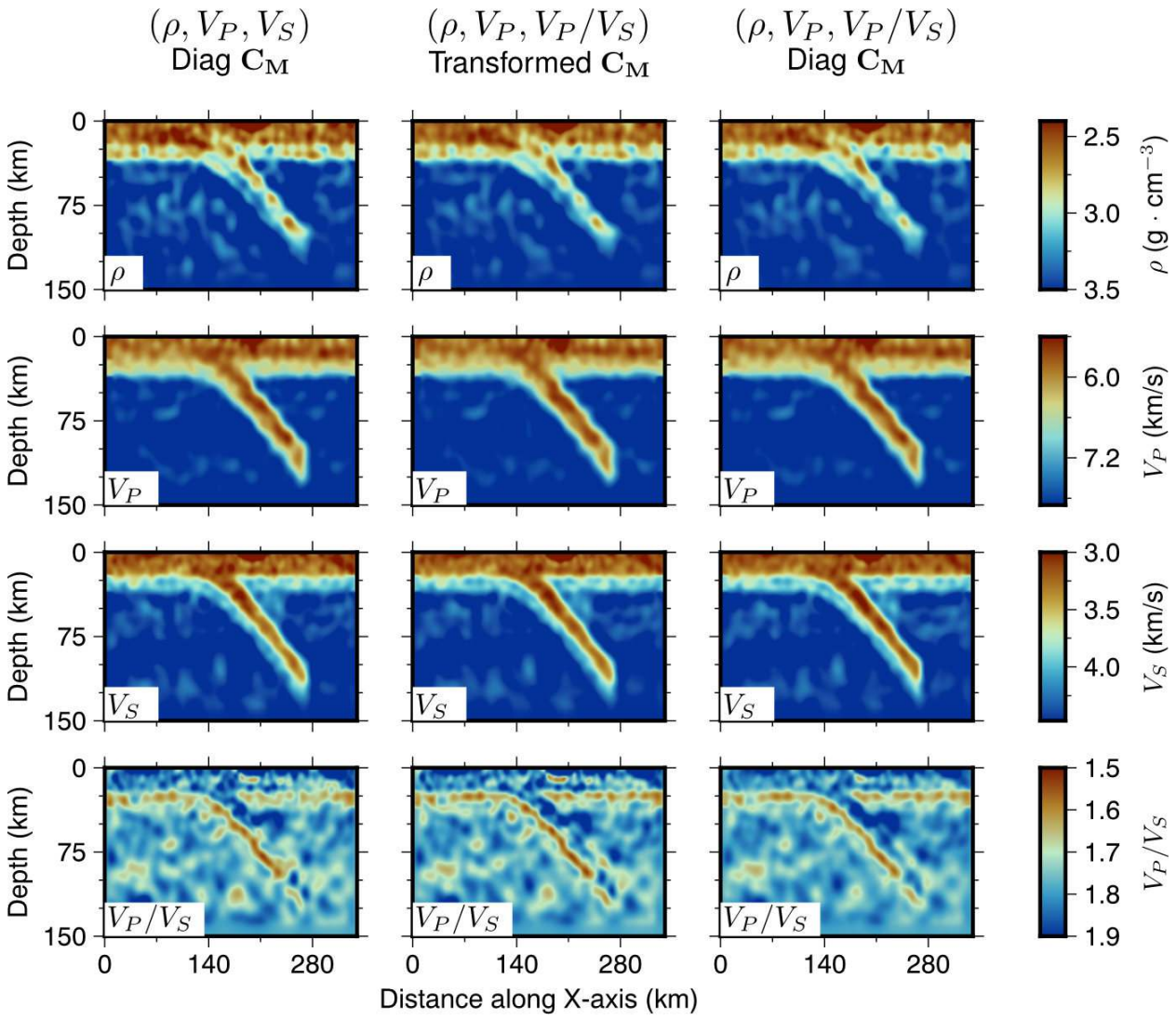


Figure 4.6: FWI results obtained with (a) the $(\rho, V_P, V_S)$ parameterization and a diagonal model covariance matrix, (b) the $(\rho, V_P, V_P/V_S)$ parameterization and a transformed $\mathbf{C_M}$ defined in Equation 4.11, and (c) the $(\rho, V_P, V_P/V_S)$ and a diagonal $\mathbf{C_M}$. In all cases the data are band-pass filtered between 0.04 and 0.40 Hz. Note the similarity of models (a) and (b) obtained with different parameterizations.

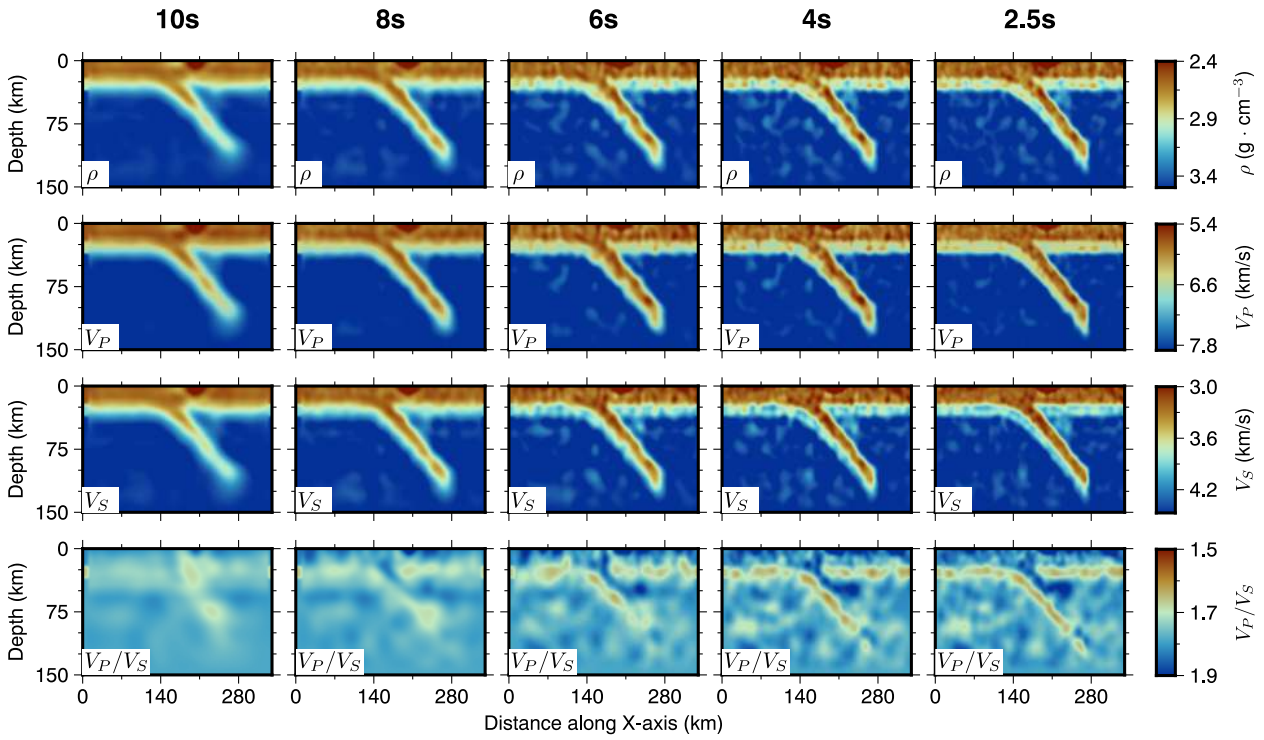


Figure 4.7: Results of five stages of multiscale FWI with $(\rho, V_P, V_S)$ parameterization, assuming a non-diagonal model covariance. The first inversion starts from long period filtered in 25 s to 10 s, from which we take the model as the input of next run and decrease the cut-off period progressively to 8 s, 6 s, 4 s and finally 2.5 s.

between model parameters improves significantly the results of FWI.

In Fig. 4.8, we display 2-D plots of model parameters corresponding to the first and last stages of the hierarchical inversions with (Figure 4.7) and without correlations (Figure 4.3). Note that after the long period (10 s) inversions, in all cases, neither the basin (brown circle) nor the mantle wedge (green circle) are recovered. The spatial resolution is simply insufficient to properly resolve these small-scale structures. When the model parameters are assumed to be uncorrelated ($r = 0$), model parameters can deviate significantly from the average trends (Figure 4.8b). When large correlation coefficients are imposed ($r = 0.97$), we observe a very different behavior. The model parameters are almost linearly correlated, with very small deviations from the average trends (Fig. 4.8c). The average trend is quasi linear in several distinct intervals which means that the inversion algorithm still has some degrees of freedom to reproduce the variability of model parameters. To summarize, the inversion has successfully retrieved the averaged scaling relationships that exist between model parameters. Future applications of FWI on real data should thus bring important insights into these scaling relationships between density and seismic velocities in different tectonic environments. This versatility would not be possible if linear a priori constraints were imposed between model parameters, for example between density and compressional velocities. Note that this would also imply defining prior values for the proportionality between these two parameters, which could also bias the inversion results.

Figure 4.9 compares the evolution of the residual variance with the number of iterations in the long period inversions with diagonal and non-diagonal model covariance matrices. When the model parameters are assumed to be uncorrelated, the convergence is slower in the first 10 iterations, and a much larger number of iterations are necessary to reach convergence. Intuitively, when strong correlations are imposed between model parameters, this restricts the model space explored by the iterative inversion algorithm, which both accelerates the convergence and restricts the occurrence of artifacts in the reconstructed models.

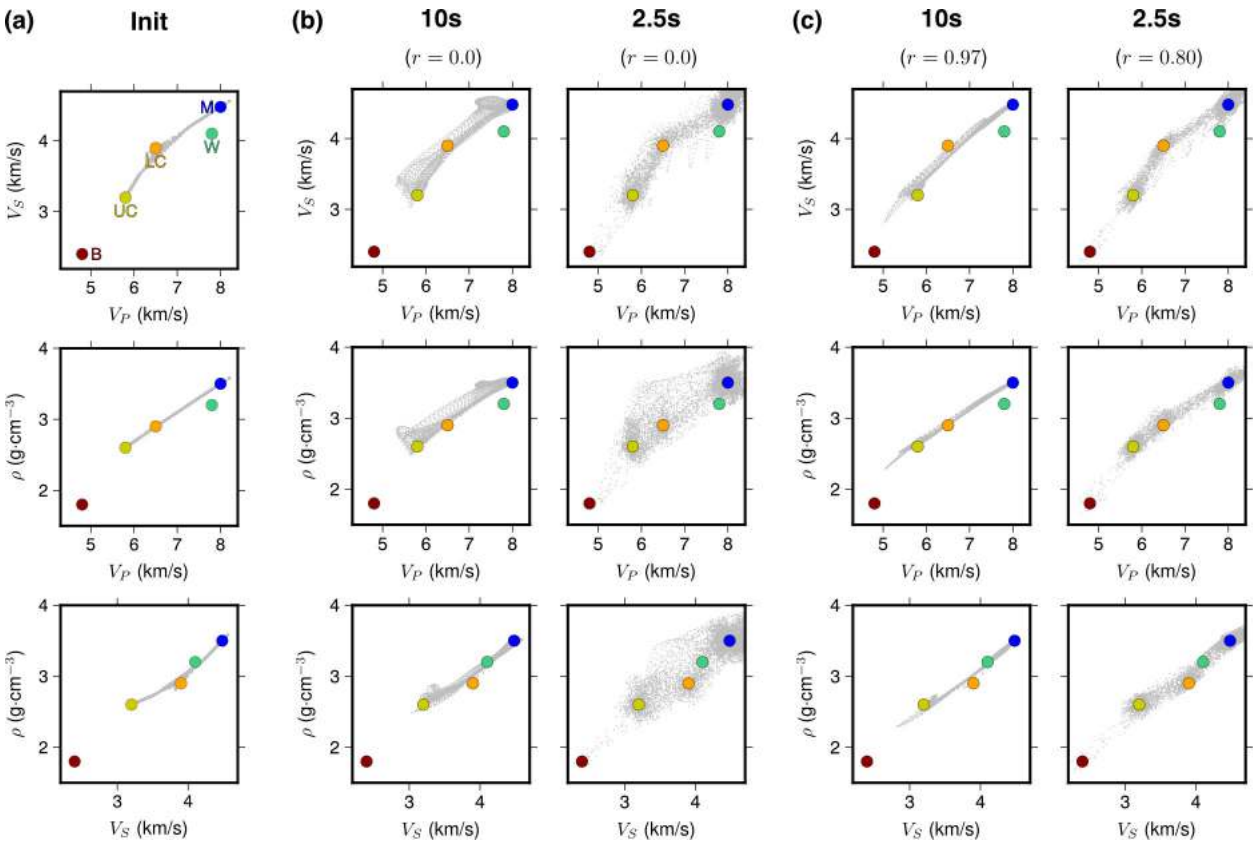


Figure 4.8: Scatter plots for inversion with and without correlations between model parameters $\rho$, $V_P$, and $V_S$. (a) Scatter plots for initial smooth 1D model. (b) Scattering plots for models with zero correlations between model parameters for the 10 s and 2.5 s inversions. (c) Scattering plots for models with strong correlations (0.97 and 0.80) between model parameters parameters corresponding to the 10 s and 2.5 s inversions. The gray dots represents the value in the model. The color dots are the same layer as detailed in Figure 4.2b.

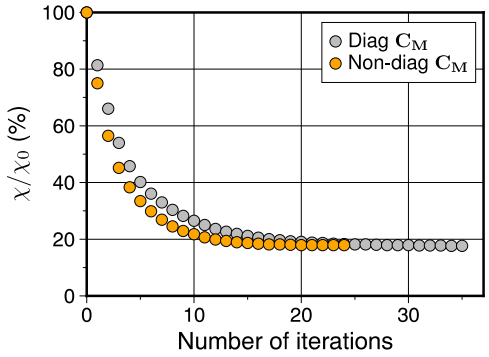


Figure 4.9: Evolution of the normalized cost reduction in the long period inversions (data filtered between 0.04 and 0.10 Hz) with (orange dots) and without (grey) correlation imposed between model parameters. The inversion with the non-diagonal covariance matrix converges faster.

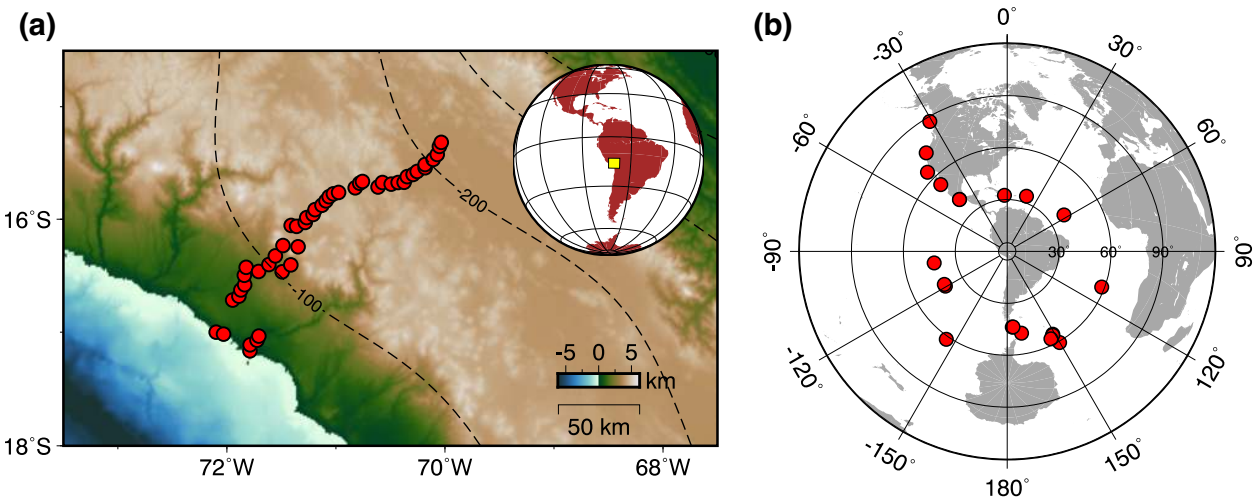


Figure 4.10: Geometry of the PERUSE deployment (a) and teleseismic sources (b) used in the real data inversion. The dashed lines show the depth contours of the Nazca plate taken from the Slab2 model (Hayes et al., 2018).

### 4.5 Application to real data: the southern Peru transect

We now illustrate the impact of using a complete non-diagonal model covariance matrix on a real dataset, the PE profile of the Peru Subduction Experiment (Clayton, 2013) (Figure 4.10). The motivation of this experiment was to image the subduction of the Nazca plate beneath South America. Under this profile, the subduction of the Nazca plate is normal, with a dip of the Nazca plate around 30°. Receiver functions have shown that the crustal thickness is increasing from 25 km close to the coast to 75 km beneath the Altiplano (Phillips et al., 2012).

We only provide here a brief outline of the data selection and preparation. The results of this tomographic study will be presented and discussed in full detail elsewhere. We retrieved all the data corresponding to teleseismic events with magnitude larger than 6 from IRIS-DMC. For this study, we selected the records of 16 events with large S/N ratios. We invert both the radial and vertical component records of the P waves, over time windows that start 5 s before the P arrival and end from 50 s to 70 s after, depending on the event. The data are band-pass filtered between 0.04 and 0.08 Hz. We first compute synthetic

impulse responses using a Gaussian source wavelet with a dominant period of 4 s with the hybrid AxiSEM/SEM method (Monteiller et al., 2021). For each event, we estimate the source wavelet by deconvolving the vertical component records from their corresponding synthetic impulse responses, aligning the resulting traces, and finally stacking them. The complete synthetic seismograms are obtained by convolving the impulse responses with the corresponding source wavelet. The inversions are started from a smooth 1-D reference model derived from the ak135 reference Earth model (Kennett et al., 1995). We perform two types of inversions. In the first inversion, we assume that model parameters are independent and thus consider a diagonal model covariance matrix. In the second inversion, we impose a correlation coefficient of 0.9 between all the pairs of parameters. The smoothing length in both inversions is set to 30 km.

The convergence is achieved after 19 and 17 iterations in inversions 1 and 2 respectively, with a variance reduction of about 80% in both cases. The final models obtained are shown in Figure 4.11a, which displays the absolute values of density and seismic velocities whereas Figure 4.11b shows the deviations of density and seismic velocities with respect to the depth averaged 1-D models. The $V_P$ and $V_S$ models are relatively similar in the two inversions, with a pronounced low velocity anomaly down to about 80 km depth that corresponds to the thickened crust beneath the Altiplano. The $V_P$ model is slightly better resolved at depth, as in the synthetic experiments shown previously. In particular, the fast velocity anomaly related to the subducting Nazca plate is imaged in the $V_P$ model, but with a much finer resolution in the inversion that accounts for the correlation between model parameters. However, the benefits of accounting for the correlations between model parameters are more pronounced on the density and $V_S$ models. From this very simple inversion experiment on real data, we can thus already conclude that accounting for the correlation between density and seismic velocities looks very promising to improve FWI results.

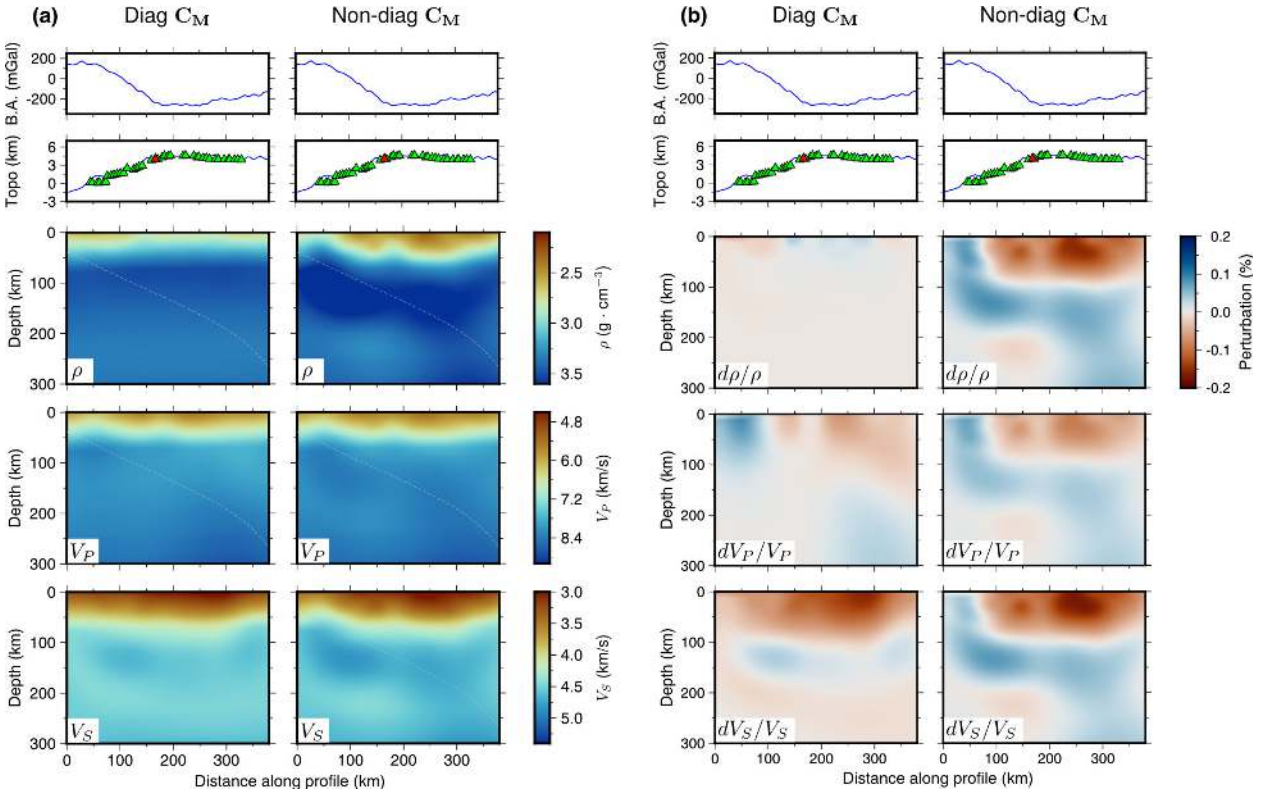


Figure 4.11: Absolute $\rho$, $V_P$, and $V_S$ models (a) obtained by FWI of 16 teleseismic events recorded by the profile PE of PeruSE array, with data filtered between 0.04 and 0.08 Hz. (b) Deviations of $\rho$, $V_P$, and $V_S$ from their average 1-D models. The green triangles represent the seismic stations whereas the red triangle indicates the location of the active volcanic arc. B.A. represents Bouguer anomaly.

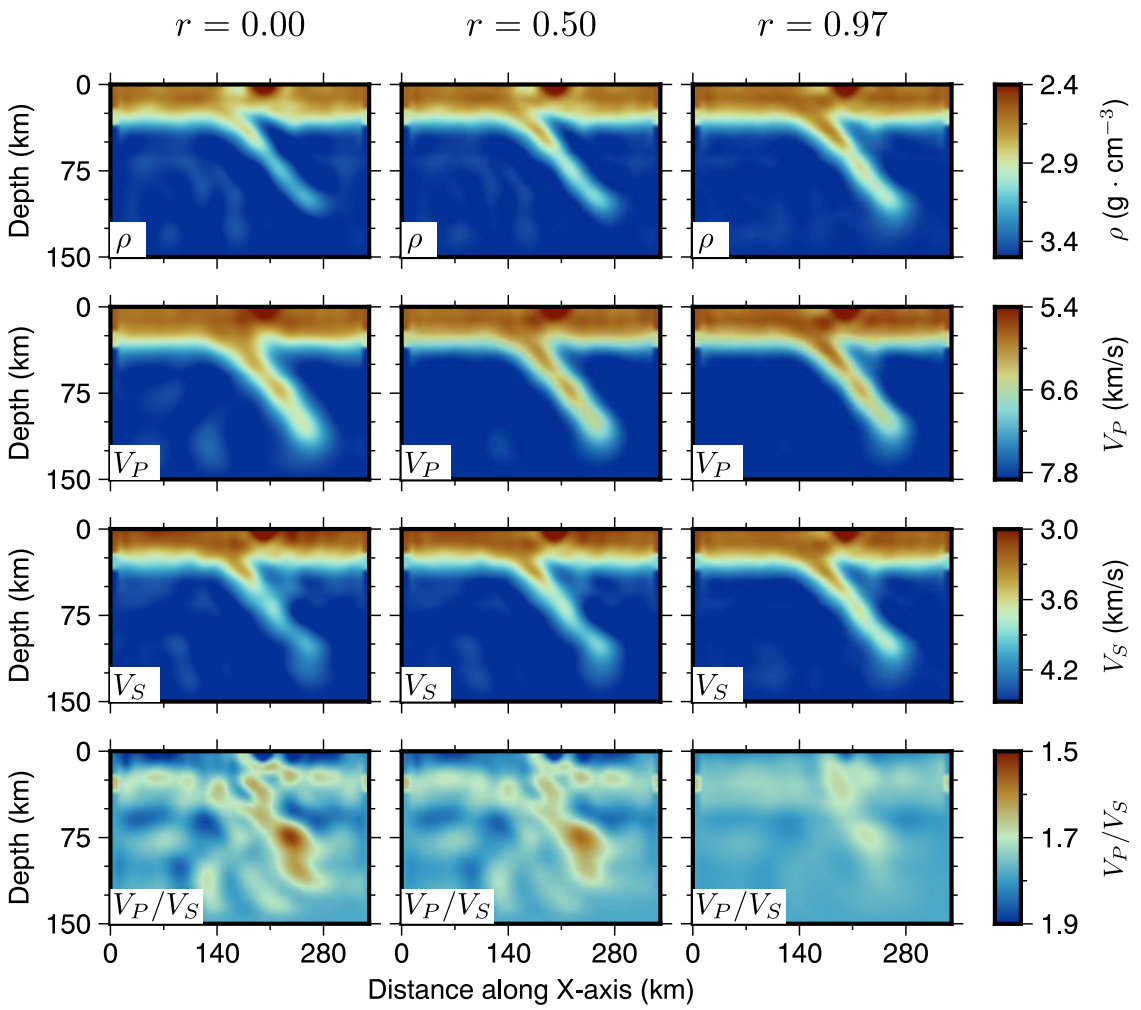


Figure 4.12: FWI models obtained with data filtered between 0.04 and 0.1 Hz, using correlation coefficients between model parameters varying from 0 (no correlation) to 0.97 (strong correlation).

## 4.6 Discussion

### 4.6.1 Tuning FWI hyperparameters: model parameterization, standard deviations, correlation coefficients, and smoothing length

If we assume that the model parameters are not correlated, then the $(\rho, V_P, V_P/V_S)$ parameterization is expected to outperform the $(\rho, V_P, V_S)$ parameterization. This can be simply understood by looking at Figure 4.2, which suggests that to first order $V_P/V_S$ and $V_P$ are indeed poorly correlated ($r = -0.37$), whereas assuming that $V_P$ and $V_S$ are independent

is clearly a poor assumption ($r = 0.96$). The scatter plots shown in Figure 4.2 show that parameters ($\rho$, $I_P$, $I_S$) and ($\rho, V_P, V_S$) are strongly correlated. This suggests that choosing any of these two sets of parameters, combined with strong ($r > 0.95$) correlation coefficients, is probably the best strategy for imaging the Earth with multi-parameter full waveform inversion.

We now explore the influence of parameter correlations by performing low frequency (0.04-0.1 Hz) inversions with $r$ equals to 0.00, 0.50 and 0.97, and a constant smoothing length of 5 km. The results of these 3 inversions are shown in Figure 4.12. The most noticeable differences are observed in the reconstructed $\rho$ models, for which the imposed correlations with $V_P$ and $V_S$ compensate the limited P wave sensitivity to density structure. The best density model is obtained for very large correlation coefficients ($r = 0.97$) between density and seismic velocities. The sedimentary basin is more sharply defined, the slab is better imaged, especially in its deeper part, and the spurious small-scale artifacts that are clearly visible in the inversion with no correlation have almost completely disappeared. A similar albeit less spectacular improvement is observed on the reconstructed $V_S$ models. As for the density reconstruction, the imposed correlation between $V_P$ and $V_S$ compensates for the lack of sensitivity to the deep, large scale, $V_S$ heterogeneities. Not surprisingly, the correlation coefficient imposed between $V_P$ and $V_S$ has also a large impact on the $V_P/V_S$ ratio. For small $r$ values, the $V_P/V_S$ model is contaminated by strong artifacts resulting from the marked difference in resolution between the $V_P$ and $V_S$ models, the latter being characterized by a very poor resolution at depth. This long period inversion test suggests that complete non-diagonal model covariance matrices are key ingredients to image lateral variations of density in Earth's interior.

Other synthetic tests (not shown) have demonstrated that the models are better reconstructed with moderate smoothing lengths (5 to 10 km) and that significant artifacts emerge when smoothing lengths larger than about 30 km are used. Intuitively, we can understand this by the fact that in the hierarchical inversion algorithm, we first invert long period

($T > 10$ s) signals to obtain smooth long wavelength models. At this stage, the gradients are smooth and do not need additional smoothing. On the other hand, when the frequency cut-off is progressively increased, the models start to resolve smaller structural details that could potentially be erased by imposing a too large amount of smoothing. Imposing a minimum amount of smoothing in the inversions is thus advisable.

The last hyperparameters that need to be tuned are the standard deviations of each class of model parameters. In all the inversions that we have presented so far, we have used $\sigma_\rho = 0.27$ g·cm$^{-3}$, $\sigma_{V_P} = 0.65$ km/s and $\sigma_{V_S} = 0.37$ km/s. These values were simply derived by the inspection of the scattering plots shown in Figure 4.2. Figure 4.13 shows the results of the long period inversion obtained for different values of $\sigma_{V_P}$. When $\sigma_{V_P}$ is reduced, the amplitude of the anomalies in the $V_P$ models are smaller and smoother. By contrast, the density model shows sharper variations, in particular in the shallow par of the slab. The $V_S$ is basically unaffected. When $\sigma_{V_P}$ is increased, we observe opposite effects on the $\rho$ and $V_P$. The $V_P$ model shows stronger and sharper contrasts whereas the $\rho$ model is smoother and reduced amplitude anomalies. Unbalanced relative standard deviations of model parameters can thus lead to trade-offs between model parameters, in particular between $\rho$ and $V_P$.

### 4.6.2 Joint inversions of P and SH events

As we have seen in the previous sections, accounting for the correlation between $V_P$ and $V_S$ in the inversions allowed us to improve the resolution in the deep parts of the $V_S$ models. Figure 4.14 compares the reconstructed models obtained by adding 4 SH events to the dataset, with or without accounting for the correlations between model parameters. Figure 4.15 shows the synthetic seismograms computed in the final models (black lines), and the seismograms computed in the initial (grey lines) and final (red lines for P waves and blue lines for SH waves) models. This comparison reveals that the SH waveforms are already quite nicely matched by the model derived from the inversion of P waveforms only, provided a strong correlation between $V_P$ and $V_S$ is assumed. When SH waves are included in the inversion, the fits of SH waveforms are significantly improved, in particular for the

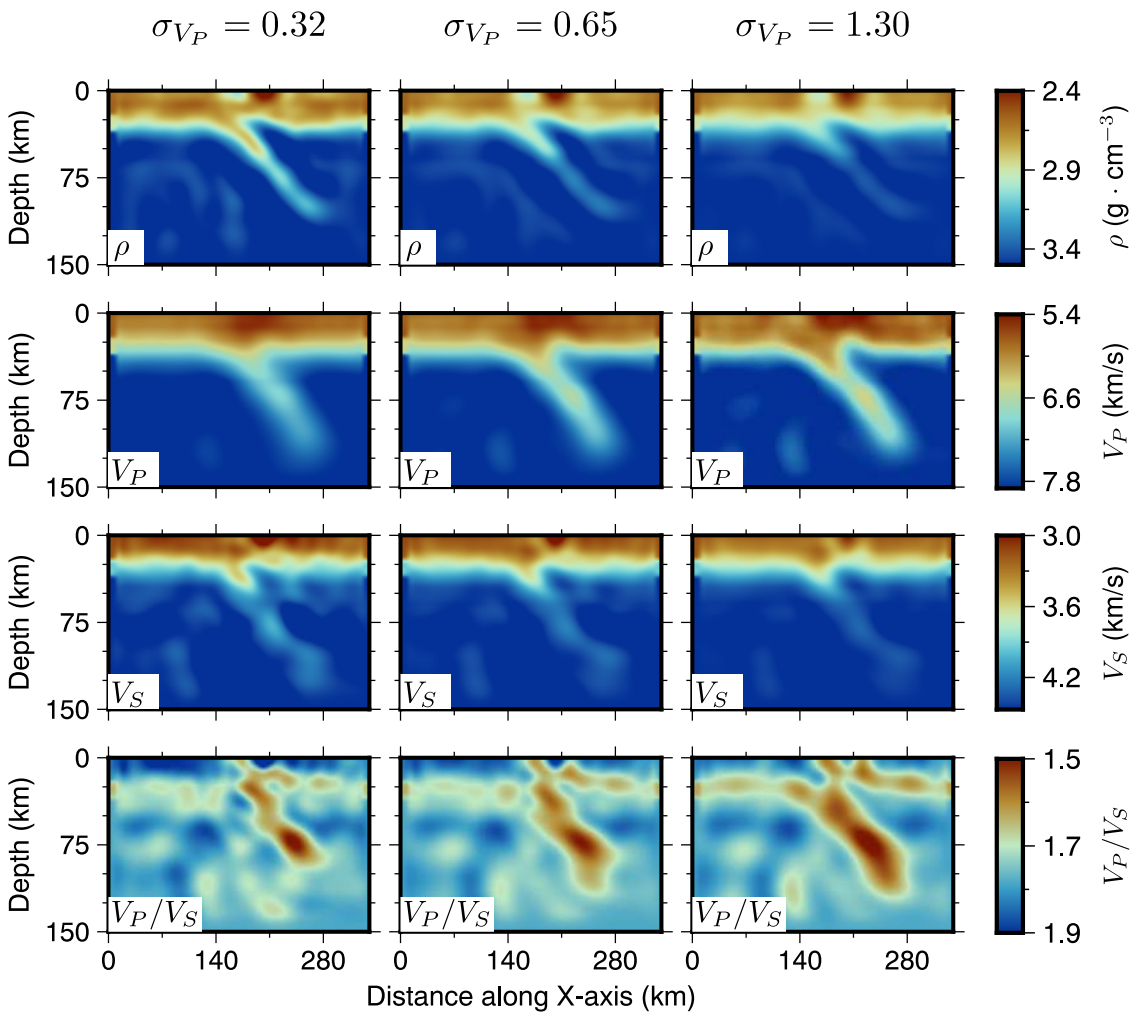


Figure 4.13: Comparison of long period FWI results for different $\sigma_{V_P}$. All the other hyper-parameters are the same as in the reference inversion shown in Figure 4.3.

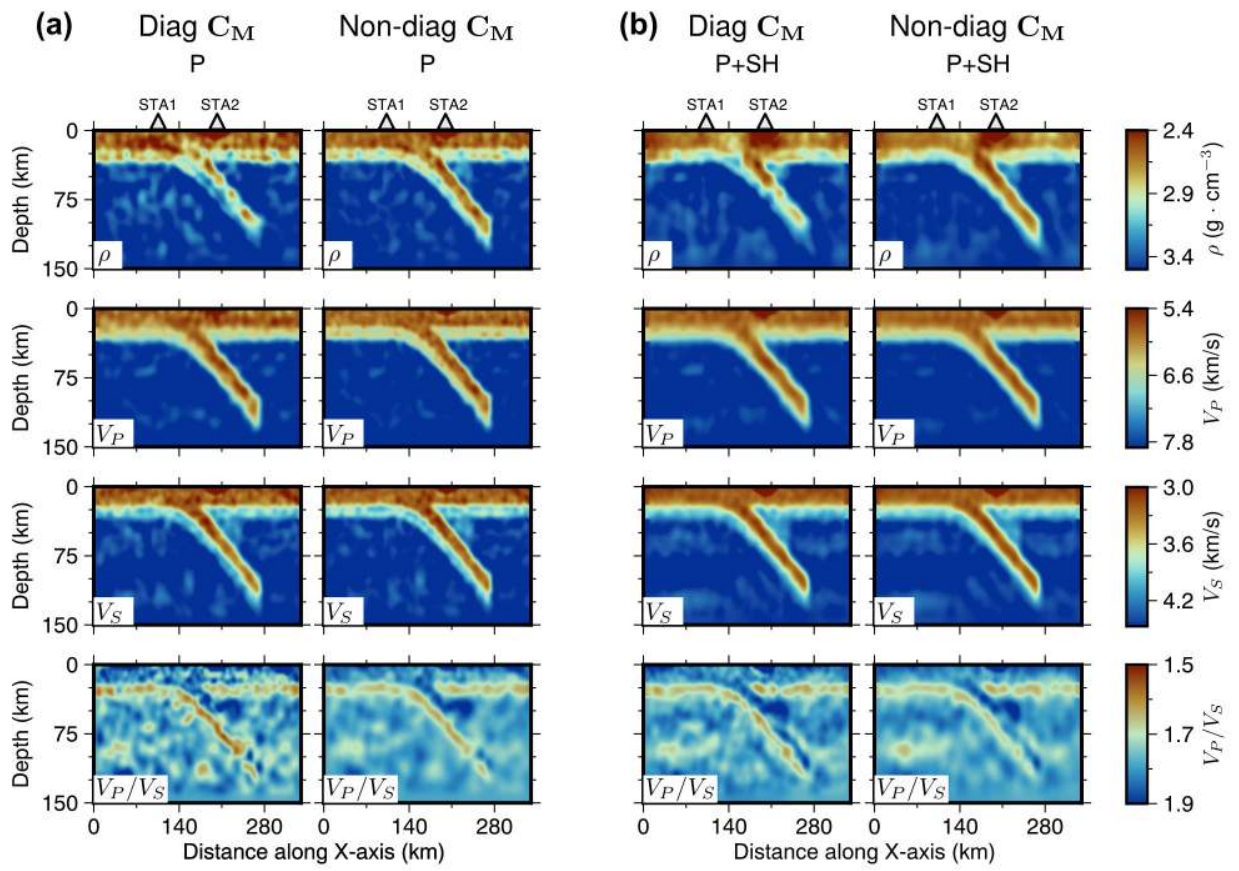


Figure 4.14: Comparison of the results obtained after multiscale (0.04 to 0.40 Hz frequency bandwidth) FWI considering P waveforms only (a) and both P and SH waveforms (b), with a diagonal model covariance matrix or with a non-diagonal model covariance matrix. The multiscale inversion strategies are the same as in Figure 4.3 and 4.7. The corresponding waveforms for STA1 and STA2 are shown in Figure 4.15.

multiples that arrive ∼15 s after the direct SH waves. We can first notice that in the joint inversion of P and S waveforms, the reconstructed $V_P$ and $V_S$ models assuming a diagonal or a non-diagonal model covariance matrix are very similar. In addition, these models are also remarkably similar to the models obtained previously from the sole inversion of P waveforms but using the full covariance matrix. These results thus suggest that accounting for the correlation between $V_P$ and $V_S$ models indeed allows us to reconstruct the $V_S$ model even if we only invert P waveforms. However, the $V_P/V_S$ ratio model is better recovered when both P and S waveforms are inverted and with a non-diagonal model covariance matrix.

Although these results seem to suggest that it may not be necessary to include shear waves in the inversion of waveform to image $V_S$ heterogeneities, the contribution of shear waves in waveform inversion is in fact far from negligible. First, shear waves provide an independent source of data that can be used to assess the quality of the model obtained by the inversion of teleseismic P waveforms. Second, adding shear waves in the inversion will allow us to get a more robust constraint on the $V_V/V_S$ ratio, which would otherwise suffer from the contrast in the spatial resolution of the $V_P$ and $V_S$ models. In addition to improving spatial coverage, in particular of deep and long wavelengths $V_S$ heterogeneities, combining S waveforms with P waveforms is also essential to properly constrain seismic anisotropy, as demonstrated by Beller and Chevrot (2020).

## 4.7 Conclusions

We have presented a new implementation of Bayesian full waveform inversion that introduces a priori information through a full 3-D model covariance matrix. The spatial correlation function that describes the statistical spatial correlations of model parameters is an exponential kernel. Its inverse admits a simple semi-analytical expression and it can thus be easily applied to any vector model within the spectral-element framework that is used to simulate the wave propagation. With this formalism, it is also possible to account for the physical correlation between density and compressional and shear wave velocities by introducing off-diagonal terms in the model covariance matrix.

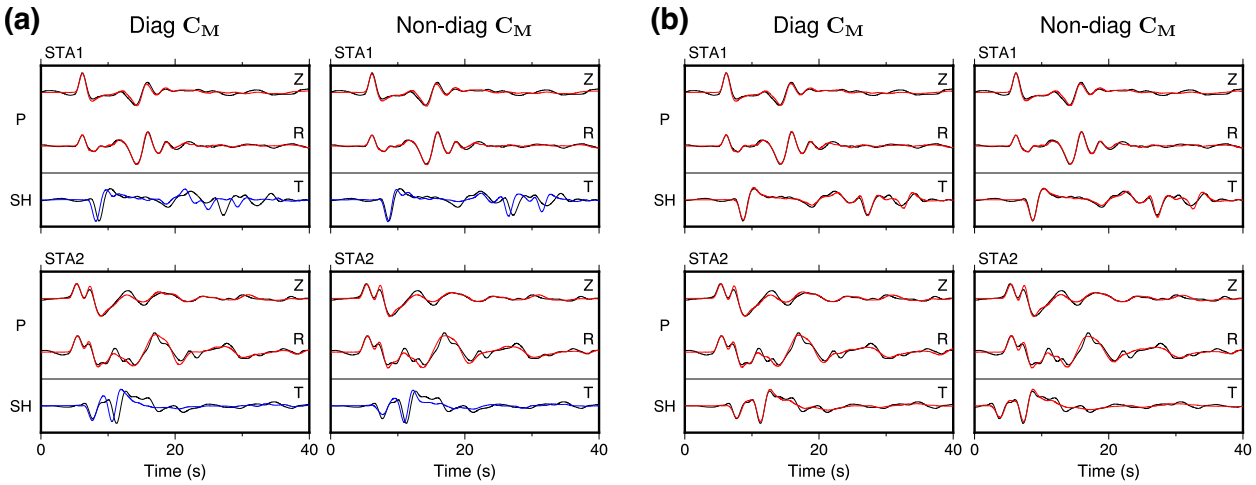


Figure 4.15: Waveform fits for the event with a backazimuth of 70 degree. (a) Waveform fits obtained by inverting P waveforms considering a diagonal and a non-diagonal model covariance matrix. (b) Waveform fits obtained by inverting both P and SH waveforms considering a diagonal and a non-diagonal model covariance matrix. All the traces are bandpass filtered between 0.04 and 0.40 Hz. Black lines show the noisy data. Red traces represent the corresponding waveforms computed in the final model shown in Figure 4.14. Blue lines represent the predicted SH waveforms when only P waves are included in the inversion. The position of stations STA1 and STA2 are marked in Figure 4.14.

Synthetic inversion experiments demonstrate that with this new inversion method the parameter correlations compensate for the lower sensitivity of P waveforms to shear velocities and the final shear velocity model is on par with the one that would have been obtained by jointly inverting P and S waveforms. The improvement on the reconstruction of the density is even more spectacular. The reconstruction of the $V_P/V_S$ ratio, a key parameter to constrain the composition and thermal state of the lithosphere, is also greatly improved. Whereas the models obtained by assuming the model parameters are uncorrelated, i.e. assuming a diagonal model covariance matrix, are contaminated by significant artifacts, most of these artifacts disappear when the correlations between model parameters are accounted for during the inversion. In addition, complete non-diagonal model covariance matrices reduce the number of degrees of freedom of the inverse problem and thus the size of the model space to explore. The convergence of the inversion algorithm is consequently faster.

Future work will be devoted to apply this new approach on real teleseismic waveforms

recorded by dense regional arrays. The extension of the Bayesian FWI approach described here to general anisotropic media will be presented in a future contribution

**Acknowledgements**

V. Monteiller was supported by the European Union's Horizon 2020 research and innovation programme under the ChEESE project, grant agreement No. 823844. This work was performed using HPC resources from GENCI–IDRIS (Grant 2021-101085) and CALMIP (Grant P20012). We also acknowledge PRACE for awarding us access to Marconi100 at CINECA, Italy.

## DATA AVAILABILITY

The FWI code is available upon request to the corresponding author, and it will be released to the public in the near future. The waveform data of station MLS from the French RLBP network and of the profile PE in PERUSE are available from IRIS-DMC.

## 4.A Decomposition of the model covariance matrix

We consider a model covariance matrix of the form

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2 \mathbf{C} & r\sigma_\rho\sigma_{V_P}\mathbf{C} & r\sigma_\rho\sigma_{V_S}\mathbf{C} \\ r\sigma_\rho\sigma_{V_P}\mathbf{C} & \sigma_{V_P}^2\mathbf{C} & r\sigma_{V_P}\sigma_{V_S}\mathbf{C} \\ r\sigma_\rho\sigma_{V_S}\mathbf{C} & r\sigma_{V_P}\sigma_{V_S}\mathbf{C} & \sigma_{V_S}^2\mathbf{C} \end{bmatrix}, \tag{4.25}$$

where $r$ is the correlation coefficient, $\mathbf{C}$ the spatial covariance kernel, $\sigma_\rho$ the standard deviation of density, $\sigma_{V_P}$ the standard deviation of $V_P$, and $\sigma_{V_S}$ the standard deviation of $V_S$. Note that we assume here for simplicity that the correlation coefficients between each pair of parameters are equal.

The model covariance matrix can be rewritten

$$\mathbf{C_M} = \mathbf{\Sigma S R S \Sigma}, \tag{4.26}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\rho & 0 & 0 \\ 0 & \sigma_{V_P} & 0 \\ 0 & 0 & \sigma_{V_S} \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \mathbf{C}^{\frac{1}{2}} & 0 & 0 \\ 0 & \mathbf{C}^{\frac{1}{2}} & 0 \\ 0 & 0 & \mathbf{C}^{\frac{1}{2}} \end{bmatrix}, \tag{4.27}$$

and

$$\mathbf{R} = \begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}. \tag{4.28}$$

The eigenvalues $\lambda$ of $\mathbf{R}$ are solutions of

$$|\mathbf{\Theta} - \lambda\mathbf{I}| = (1-\lambda)^3 - 3r^2(1-\lambda) + 2r^3 = 0. \tag{4.29}$$

The third-order polynomial in $x = 1 - \lambda$ can be factorized as

$$x^3 - 3r^2x + 2r^3 = (x + 2r)(x - r)^2. \tag{4.30}$$

The first eigenvalue of $\mathbf{R}$ is thus $\lambda_1 = 1 + 2r$ and the first eigenvector

$$\mathbf{v_1} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}. \tag{4.31}$$

The second and third eigenvalues are $\lambda_2 = \lambda_3 = 1 - r$ and

$$\mathbf{v_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \mathbf{v_3} = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}. \tag{4.32}$$

The singular value decomposition of $\mathbf{R}$ is

$$\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t, \tag{4.33}$$

with

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \text{ and } \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \tag{4.34}$$

The model covariance matrix can therefore be decomposed into the product

$$\mathbf{C_M} = \mathbf{\Sigma S V \Lambda V}^t\mathbf{S\Sigma}. \tag{4.35}$$

This decomposition provides some insights into the anatomy of the physical constraints that are imposed on the model parameters. The matrix $\Sigma$ defines a metric that normalizes the

perturbations of model parameters. It makes the inversion insensitive to the units chosen and equalizes the contribution of each class of parameter to the gradient of the misfit function. The matrix $\mathbf{V}$ can be seen as a transformation matrix that rotates the covariance matrix along its principal axes. The first axis, oriented along $\mathbf{v_1}$, describes the principal linear relationship between the different parameters. The second and third eigenvectors describe the deviations of model perturbations with respect to this linear relationship. When $r \to 1$, $\lambda_1 \to 3$ and $\lambda_{2,3} \to 0$. The three parameters are perfectly correlated and described by their position along the line oriented along $\mathbf{v_1}$. When $r \to 0$, $\lambda_{1,2,3} \to 1$, variations of the parameters are completely random and the model covariance matrix is diagonal. This is the implicit assumption that has been made so far in most multiparameter tomographic inversions. The matrix $\mathbf{S}$ describes how model parameters vary as a function of the distance between two points and can thus be seen as a smoothing operator.

## 4.B Change of model parameterization and transformation of the model covariance matrix

We illustrate the transformation of the model covariance matrix after a change of model parameterization. We first consider a diagonal covariance matrix for model parameterization $(\rho, V_P, V_S)$ with mean $\overline{V}_P = 7$ km/s and $\overline{V}_S = 4$ km/s, $\sigma_{V_P} = 0.65$ km/s and $\sigma_{V_S} = 0.37$ km/s. We draw 2000 samples from the bivariate probability density function

$$f(V_P, V_S) = \frac{1}{2\pi\sigma_{V_P}\sigma_{V_S}\sqrt{1-r^2}} \cdot \exp\left(-\frac{1}{2(1-r^2)}\left[\left(\frac{V_P-\overline{V}_P}{\sigma_{V_P}}\right)^2 + \left(\frac{V_S-\overline{V}_S}{\sigma_{V_S}}\right)^2 - 2r\left(\frac{V_P-\overline{V}_P}{\sigma_{V_P}}\right)\left(\frac{V_S-\overline{V}_S}{\sigma_{V_S}}\right)\right]\right), \tag{4.36}$$

where $r$ is the correlation coefficient between $V_P$ and $V_S$. The scattering plot showing the distribution of these samples is displayed in Figure 4.16a. Since $\rho$ is independent of either $V_P$, $V_S$ and $V_P/V_S$, the scattering plots with $\rho$ are not shown. The same distribution after conversion to the $(V_P, V_P/V_S)$ parameterization is plotted in Figure 4.16b. Using equation (4.11), we can derive the transformed covariance matrix for the $(V_P, V_P/V_S)$ parameteri-

zation. We draw a new set of 2000 samples that follow the distribution described by this transformed covariance matrix and plot them in Figure 4.16c. As expected, the distributions shown in Figure 4.16b and Figure 4.16c are very similar.

We now consider the diagonal covariance matrix

$$\mathbf{C}_{(\rho,V_P,\frac{V_P}{V_S})} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_P}^2 & 0 \\ 0 & 0 & \sigma_{\frac{V_P}{V_S}}^2 \end{bmatrix} \tag{4.37}$$

for the $(\rho, V_P, V_P/V_S)$ parameterization, which transforms to

$$\mathbf{C}_{(\rho,V_P,V_S)} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_P}^2 & \frac{V_S}{V_P}\sigma_{V_P}^2 \\ 0 & \frac{V_S}{V_P}\sigma_{V_P}^2 & \frac{V_S^2}{V_P^2}\sigma_{V_P}^2 + \frac{V_S^4}{V_P^2}\sigma_{\frac{V_P}{V_S}}^2 \end{bmatrix} \tag{4.38}$$

for the $(\rho, V_P, V_S)$ parameterization.

We draw 2000 samples from the bivariate normal distribution (4.37) using $\overline{V}_P = 7$ km/s, $\overline{\frac{V_P}{V_S}} = 1.75$, $\sigma_{V_P} = 0.65$ km/s and $\sigma_{\frac{V_P}{V_S}} = 0.1$. The resulting scattering plot is shown in Figure 4.16d. The same distribution after conversion to the $(V_P, V_S)$ parameterization is plotted in Figure 4.16e. Finally, the distribution obtained by sampling the bivariate distribution (4.38) is shown in Figure 4.16f. Again, the two distributions in Figure 4.16e and Figure 4.16f are similar. These two examples emphasize the importance of the non-diagonal terms of the model covariance matrix when the model parameterization is changed.

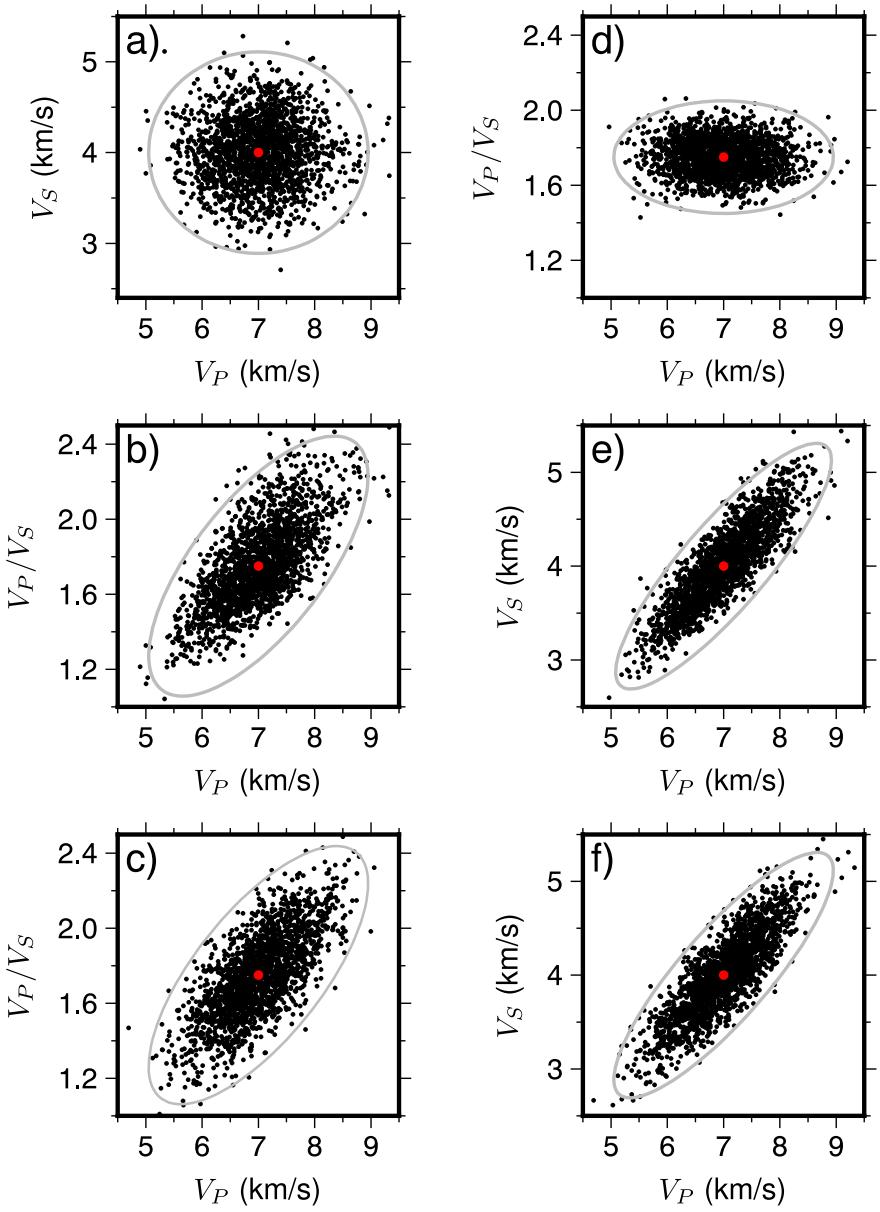


Figure 4.16: The consistency of bivarariate normal distributions over different parameterizations, with the covariance matrix properly transformed. (a) Bivariate normal distribution of $V_P$ and $V_S$, generated with a diagonal covariance matrix. (b) Same distribution of (a) but plotted in $V_P$ and $V_P/V_S$ (c) Bivariate normal distribution of $V_P$ and $V_P/V_S$, generated with the transformed covariance matrix. (d) Bivariate normal distribution of $V_P$ and $V_P/V_S$, generated with a diagonal covariance matrix. (e) Same distribution of (d) but plotted in $V_P$ and $V_S$ (f) Bivariate normal distribution of $V_P$ and $V_S$, generated with the transformed covariance matrix. The grey lines correspond to 3 standard deviations.

# Chapter 5. Dehydration of the subducting Juan de Fuca plate and fluid pathways revealed by full waveform inversion of teleseismic P and SH waves in central Oregon

This chapter has been submitted to *Journal of Geophysical Research: Solid Earth* as a research paper entitled « Dehydration of the subducting Juan de Fuca plate and fluid pathways revealed by full waveform inversion of teleseismic P and SH waves in central Oregon » by Li-Yu Kan, Sébastien Chevrot and Vadim Monteiller.

## Contents


5.1 Abstract ..... **114**
5.2 Introduction ..... **115**
5.3 Data and methodology ..... **118**
5.3.1 The CASC93 experiment ..... 118
5.3.2 Data selection and preparation ..... 118
5.3.3 Design of the tomographic mesh ..... 122
5.3.4 Definition of the cost function ..... 124
5.3.5 The full waveform inversion method ..... 124
5.3.6 The iterative inversion algorithm ..... 126
5.3.7 Event normalization ..... 127
5.4 Results ..... **128**
5.4.1 Tomographic model obtained after hierarchical inversion ..... 128
5.4.2 Waveform fits ..... 131
5.5 Discussion ..... **133**
5.5.1 The impact of adding SH waveforms and accounting for the correlation between model parameters ..... 133
5.5.2 Scaling relationships between model parameters and the effect of the initial smooth 1D model ..... 138
5.5.3 Fluid migrations, melting, and metamorphic reactions in the central Cascadia subduction revealed by FWI ..... 139
5.6 Conclusions ..... **144**

## 5.1 Abstract

The dehydration of subducting slabs expels a massive amount of water into the forearc and backarc mantle which is responsible for the serpentinization of the mantle wedge, as well as the production of melt and arc magmatism. These processes are expected to have characteristic signatures in density and seismic velocity models, which remain largely elusive to date due to the limited spatial resolution of classical passive tomographic approaches. Here we present a tomographic model of density, $V_P$, $V_S$, and $V_P/V_S$ beneath central Oregon, obtained by inverting complete teleseismic P and SH waveforms recorded by the CASC93 temporary experiment. The final model shows an east-dipping low-velocity layer less than 10 km thick that can be associated with the fluid-saturated Juan de Fuca oceanic crust. The distribution of tremors at the surface closely coincides with the horizontal extent of this low-velocity layer. Below 40 km depth, seismic velocities and density increase progressively to the values of a typical mantle. This transitional domain corresponds to the eclogitization of the oceanic crust. Silica-saturated fluids released by pore collapse migrate upward, producing serpentinization reactions in the forearc mantle which lower the density and seismic velocities. The very low $V_P/V_S$ ratio documented in the Cascadia forearc crust is evidence that these silica-saturated fluids reach the crust, where they produce massive quartz mineralization. At greater depth, a low seismic velocity and high $V_P/V_S$ ratio anomaly provides evidence for partial melting at around 75 km depth beneath the volcanic arc.

### Plain Language Summary

The subduction of the hydrated oceanic lithosphere injects large amounts of water into the Earth's interior. However, owing to the limited spatial resolution of classical tomographic approaches, the fluid pathways and related metamorphic reactions remain elusive. In this study, we image the Cascadia subduction beneath central Oregon by inverting complete P and SH teleseismic waveforms. The resulting tomographic models of density and seismic velocities reveal the eclogitization reactions in the subducting Juan de Fuca oceanic crust

and the associated release of fluids into the forearc mantle. These silica-saturated fluids migrate upward, producing serpentinization of the forearc mantle. When these fluids reach the crust, the temperature decrease produces quartz mineralization, as evidenced by the very low $V_P/V_S$ ratio observed in the Cascadia forearc crust. At greater depth, we find evidence of partial melting around 75 km depth below the volcanic arc, resulting from dehydration of the slab that lowers the solidus temperature in the lithospheric mantle of the upper plate.

## 5.2 Introduction

Subduction of hydrated oceanic lithosphere introduces massive amounts of water into the deep Earth (e.g. Peacock, 1990). This water controls the mechanical properties of the subduction interface and the depth distribution of earthquakes and episodic tremor and slip (Peacock and Hyndman, 1999). The temperature-dependent transition from blueschist to eclogite facies governs the dehydration of the subducting plate (Peacock and Wang, 1999). For example, for the cold and old Pacific plate subducting beneath NE Japan, this transition is around 110 km depth whereas for the hot and young Philippine Sea plate subducting beneath SW Japan, the downgoing oceanic crust enters the eclogite facies at around 50 km depth (Peacock and Wang, 1999). These estimates are in good agreement with the seismologically constrained depth extent of a low seismic velocity layer often interpreted as the fluid-saturated oceanic crust (e.g. Bostock, 2013). The eclogitization of the downgoing oceanic crust and the related collapse of porosity release large amounts of fluids in the upper plate. These silica-saturated fluids will percolate through and transforms the forearc mantle into weak assemblages of serpentine, talc, and brucite (Peacock and Hyndman, 1999). In warm subduction zones, the downdip limit of subduction earthquakes is thus controlled by the depth at which the subducting slab intersects the forearc mantle (Peacock and Hyndman, 1999; Oleskevich et al., 1999; Hyndman, 2013; Hyndman et al., 2015). At deeper levels, water released by the dehydrating subducting plate will lower the solidus temperature, triggering the partial melt of mantle material that will migrate upward to feed the volcanic arc (Peacock, 1990).

Despite the important role played by fluids in arc magmatism and more generally in the processes occurring in subduction zones, fluid transfers and pathways in the mantle wedge and forearc crust remain largely elusive. The motivation of this study is to investigate the potential of full waveform inversion of teleseismic body waves to provide new tomographic constraints on deep subduction processes. In particular, recent methodological developments have shown that accounting for the correlation between density and seismic velocities improve very significantly the retrieved density and $V_P/V_S$ models (Kan et al., 2022). These two parameters are of special interest because they are reliable indicators of the presence of water, partial melt (e.g. Watanabe, 1993; Nakajima et al., 2001), or serpentinized mantle (Christensen, 2004).

For the first application of our multi-parameter full waveform inversion approach, we focus on Cascadia, which stands among the most well-studied subduction zone worldwide and represents the end member of hot and young subduction, where extensive shallow dehydration of the subducting plate is expected to occur. Fluid processes are thus expected to be strongly expressed in the structural variations of seismic velocities, and in particular of the $V_P/V_S$ ratio. Along the Cascadia subduction margin, which extends from northern California to southern British Columbia, the Juan de Fuca (JdF) plate underthrusts the northern America plate with a convergence rate that ranges from 30 mm/yr at its southern end to 45 mm/yr in its northern end (Wilson, 1993). In contrast to Northern Cascadia where the abundant seismicity constrains the geometry of the subducting slab (McCrory et al., 2012), earthquakes in central Cascadia are scarce. Therefore, studies of subduction processes in that region mostly rely on seismic recordings of distant teleseismic sources. Regional travel time tomography has found that the fast anomaly associated with the subducting Juan de Fuca plate can be traced down to the transition zone except beneath central Oregon where it disappears around 200 km depth (Schmandt and Humphreys, 2010; Obrebski et al., 2010, 2011; James et al., 2011). The Juan de Fuca is thus segmented, and the limited depth extension of the subduction beneath central Oregon which implies a small slab pull force, may

explain the lack of deep seismicity in that region. In Cascadia, as in many other subduction zones, receiver function studies have revealed the presence of a thin (less than 10 km thick) low seismic velocity and high $V_P/V_S$ ratio layer (Bostock et al., 2002; Audet et al., 2009, 2010; Hansen et al., 2012; Bostock, 2013; Tauzin et al., 2016). Episodic tremor and slip have been well studied in Cascadia (Rogers and Dragert, 2003). Whereas tremor locations are not accurate, especially for their depths (Wech and Creager, 2008; Wech, 2021), tremors seem to be related to the overpressured oceanic crust, i.e. to the thin east-dipping low-velocity layer evidenced by receiver function studies (Audet et al., 2009). The serpentinization of the forearc mantle in Cascadia is now well documented by both geological and geophysical observations (Peacock, 1993; Hyndman and Peacock, 2003; Bostock et al., 2002; Blakely et al., 2005). Local earthquake tomography studies in northern Cascadia have documented very low $V_P/V_S$ (lower than 1.7) in the forearc crust (Ramachandran and Hyndman, 2012; Peacock et al., 2011; Audet and Bürgmann, 2014; Hyndman et al., 2015; Savard et al., 2018; Guo et al., 2021). Such low values can only be explained by a significant quartz enrichment (Chevrot and van der Hilst, 2000). This would suggest that part of the silica-saturated fluids expelled by dehydration of the subduction slab reaches the crust where the temperature decrease will produce massive quartz mineralizations. This process may have important consequences for the formation and evolution of a silica-enriched continental crust.

Here, we invert the waveforms of 9 teleseismic events recorded by the stations of the CASC93 experiment in order to obtain models of density, $V_P$, $V_S$, and $V_P/V_S$ beneath central Oregon, from the surface down to 300 km depth. We first describe the data selection and preparation, and the principles of the inversion method. We then present the results of the inversion and discuss the main features of the final model. Finally, we provide preliminary interpretations of the structural anomalies in density, seismic velocities, and $V_P/V_S$ in terms of the different subduction processes that occur at depth as well as comparisons with previous geophysical studies.

## 5.3 Data and methodology

### 5.3.1 The CASC93 experiment

We use P and SH waveform records from the 42 broadband seismometers of the CASC93 temporary experiment, deployed along an E-W transect in central Oregon from March 1993 to December 1993 (Figure 5.1). The waveform inversion is performed on a selection of teleseismic events (epicentral distance between 30° and 90°) with magnitude larger than 6 and large signal-to-noise ratio.

### 5.3.2 Data selection and preparation

Because in FWI we fit complete waveforms, i.e. every wiggle observed on the seismic traces, the records that enter the inversion need to be carefully selected. Waveform data are first downloaded from the IRIS datacenter (`https://www.iris.edu`). We then apply basic signal processing (e.g. remove mean and trend) before applying a taper and removing the station response on each record. Next, we inspect each event and determine travel time and amplitude residuals at each station. We first determine a reference wavelet by aligning and averaging all the records. The travel time residuals are obtained from the time shifts measured by cross-correlation between the reference wavelet and each individual trace after removing their average. The relative amplitude residual $\delta A_i$ for station $i$ is defined as

$$\delta A_i = \log_{10}\left(\frac{\int_{t_1}^{t_2} |u_i(t)|^2 dt}{\int_{t_1}^{t_2} |\overline{u}(t)|^2 dt}\right), \tag{5.1}$$

where $u_i(t)$ is the observed trace at station $i$ and $\overline{u}(t)$ is the reference average wavelet. For P waves, these residuals are measured on the vertical component, whereas for SH waves, these residuals are measured on the transverse component. When visualized in map view, travel time and amplitude residuals allow us to quickly identify stations with abnormal time or amplitude, which are removed from the dataset.

The data selection procedure is illustrated with event 9 (Figure 5.2). The waveforms

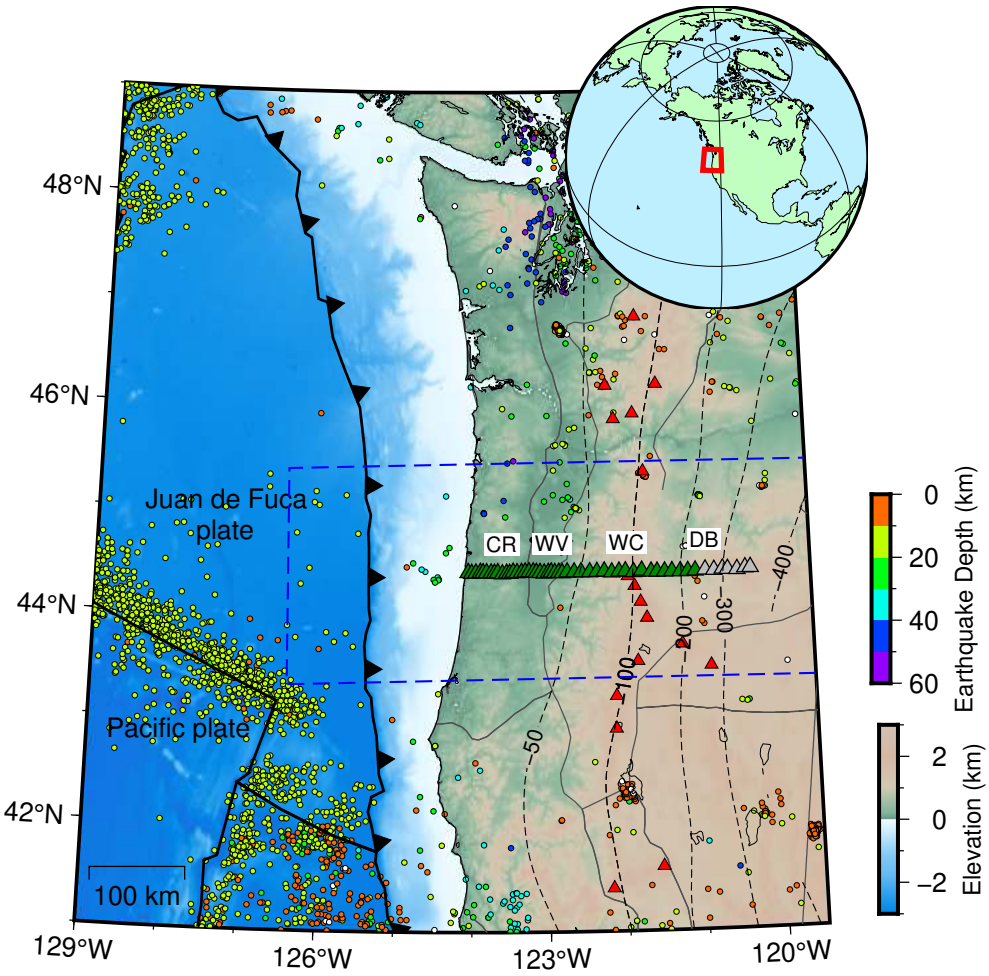


Figure 5.1: Positions of the CASC93 broadband stations (green triangles) and physiographic provinces of Oregon. The grey triangles show the stations that were discarded owing to the absence of well-recorded teleseismic events. The blue dashed line outlines the inversion domain (Fig. 5.5). The color circles correspond to the M $\geq$ 3.0 earthquakes in the USGS catalog from 1990 to 2022. The red triangles show the positions of Holocene volcanoes. The black dashed lines represent the isovalues of slab depth, taken from the Slab2 model (Hayes et al., 2018). CR: Coastal Range, WV: Willamette Valley, WC: Western Cascades, DB: Deschutes Basin.

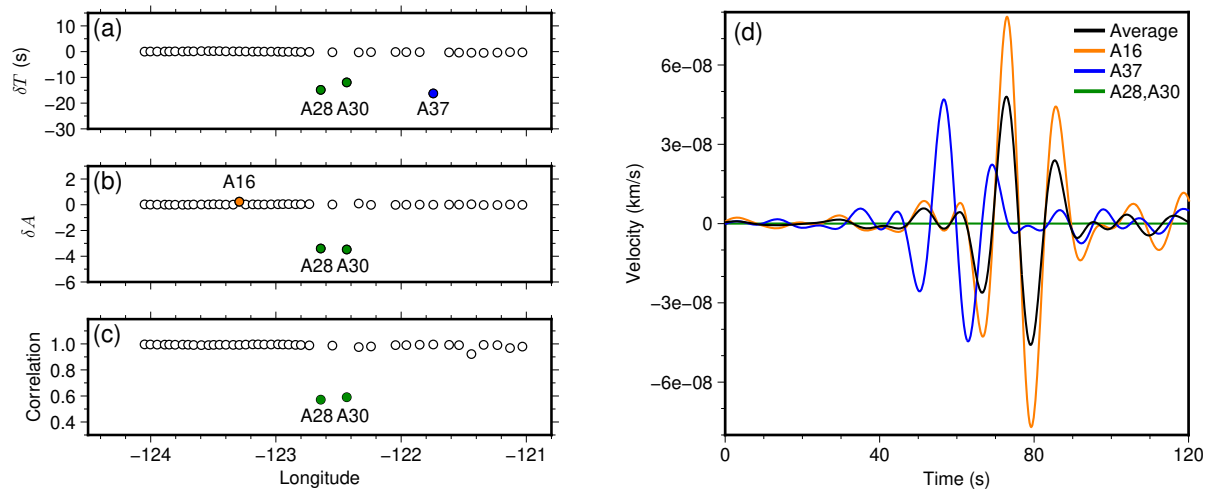


Figure 5.2: Example of data selection procedure for vertical component of event 9 in Table 5.1. (a) Travel time residuals. Negative residuals represent phases arriving earlier than the average trace. (b) Amplitude residuals. (c) Correlation coefficient of each station and the average trace. (d) Observed vertical data filtered between 10 and 25 s period.

are filtered between 10 and 25 s period, which corresponds to the frequency band used in the first inversion. The vertical component of station A28 and A30 is flat, with very small amplitudes compared with those at the other stations. A large negative travel time anomaly is observed at station A37, presumably due to a clock problem (Figure 5.2a). Finally, the relative amplitude anomaly at station A16 is 0.24, which means the amplitude at this station is 1.6 times larger than at adjacent stations. We thus decided to also remove the records of this station for this particular event.

Owing to the limited duration of the acquisition, we only found 7 P and 4 SH teleseismic arrivals of sufficient quality for FWI at periods down to 4 s (Figure 5.3). The event information is given in Table 5.1.

#### *5.3.2.1 Source wavelet estimation*

To compute synthetic seismograms of teleseismic body waves we need the apparent incident source wavelet for each selected event. To determine these source wavelets, we first compute the Green's functions for each event that corresponds to a short duration (here 2 s)

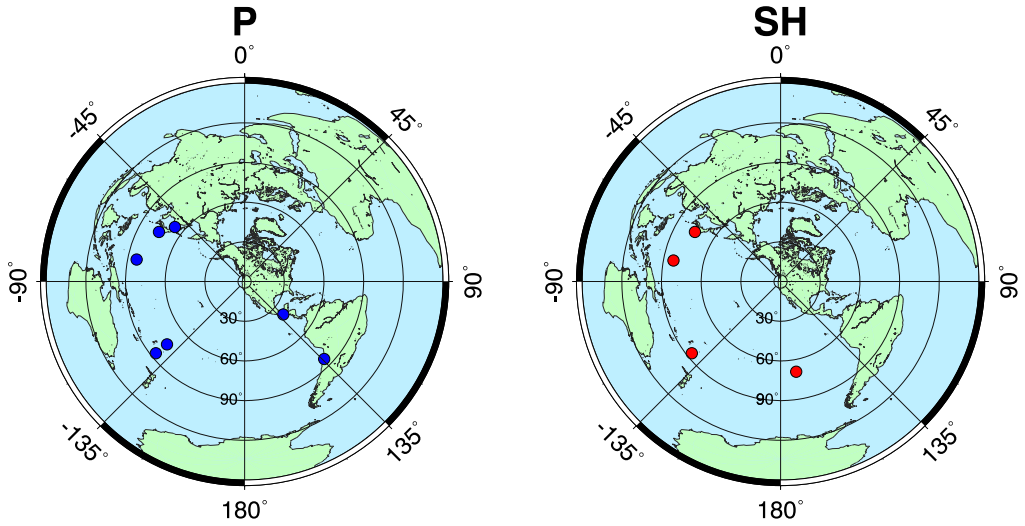


Figure 5.3: Teleseismic events for (a) P waves (b) and SH used in this study. Detailed event informations are given in Table 5.1.

Table 5.1: List of teleseismic events selected for waveform inversion.

| **Event** | **Phases** | **Origin Time** | **Location** (lat, lon) | **Depth** (km) | **Distance** (°) | **Baz** (°) | $\mathbf{M_w}$ |
|---|---|---|---|---|---|---|---|
| 1 | P | 1993/05/16 21:44:50.0 | 15.34S, 173.35W | 33 | 75.4 | 231 | 6.6 |
| 2 | SH | 1993/07/06 02:53:03.9 | 24.56S, 111.86W | 10 | 69.3 | 170 | 6.3 |
| 3 | P | 1993/07/11 13:36:19.2 | 25.35S, 70.18W | 33 | 84.2 | 134 | 6.6 |
| 4 | P | 1993/07/12 13:17:11.9 | 42.84N, 139.25E | 17 | 66.7 | 308 | 7.3 |
| 5 | P+SH | 1993/08/07 17:53:27.0 | 23.87S, 179.82E | 555 | 86.1 | 231 | 6.7 |
| 6 | P | 1993/08/08 08:34:25.0 | 12.96N, 144.78E | 61 | 83.1 | 282 | 7.5 |
| 7 | SH | 1993/08/11 14:17:41.7 | 13.13N, 145.62E | 56 | 82.3 | 281 | 6.2 |
| 8 | P | 1993/09/10 19:12:54.8 | 14.73S, 92.68W | 34 | 38.9 | 130 | 7.2 |
| 9 | P+SH | 1993/10/11 15:54:22.4 | 32.00N, 137.85E | 365 | 74.6 | 300 | 6.8 |

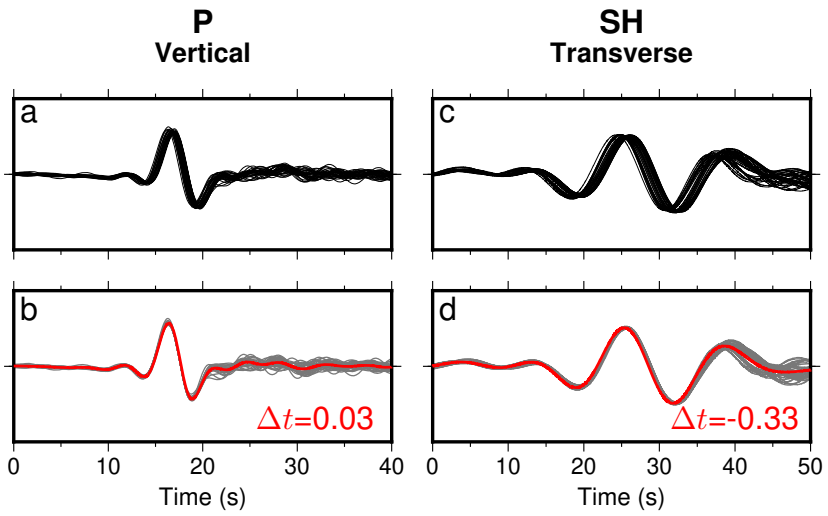


Figure 5.4: Illustration of the source wavelet estimation procedure on event 9 (see Table 5.1). The vertical and transverse components are used for P and SH waveforms, respectively. (a) Observed vertical components filtered in the period band 10-25 s. (b) Time-aligned traces (grey lines) after being deconvolved from the corresponding Green's functions and the average source wavelet (red line). (c,d) Same as (a,b) but for the transverse component SH waveforms. The average time shift $\Delta t$ is used to correct the timing of the source wavelet so that for each event the average travel time residual is zero.

Gaussian source wavelet (Monteiller et al., 2021). These computations are performed with AxiSEM in the ak135 reference Earth model (Kennett et al., 1995). The source wavelet is then obtained by first deconvolving the Green's function from the observed vertical component at each station, using a simple water-level deconvolution in the frequency domain. The deconvolved traces are then aligned with respect to the first trace and averaged. Finally, the source wavelet is shifted so that the average time residual between synthetic and observed traces is zero. This time correction is necessary to remove the contribution of heterogeneities located outside the tomographic grid, similar to the practices in classical regional travel time tomography (Aki et al., 1977).

### 5.3.3 Design of the tomographic mesh

We build a spectral-element mesh that represents a chunk of the spherical Earth, centered at -122.2° longitude and 44.4° latitude (Figure 5.5). The size of the mesh is 6 degrees in

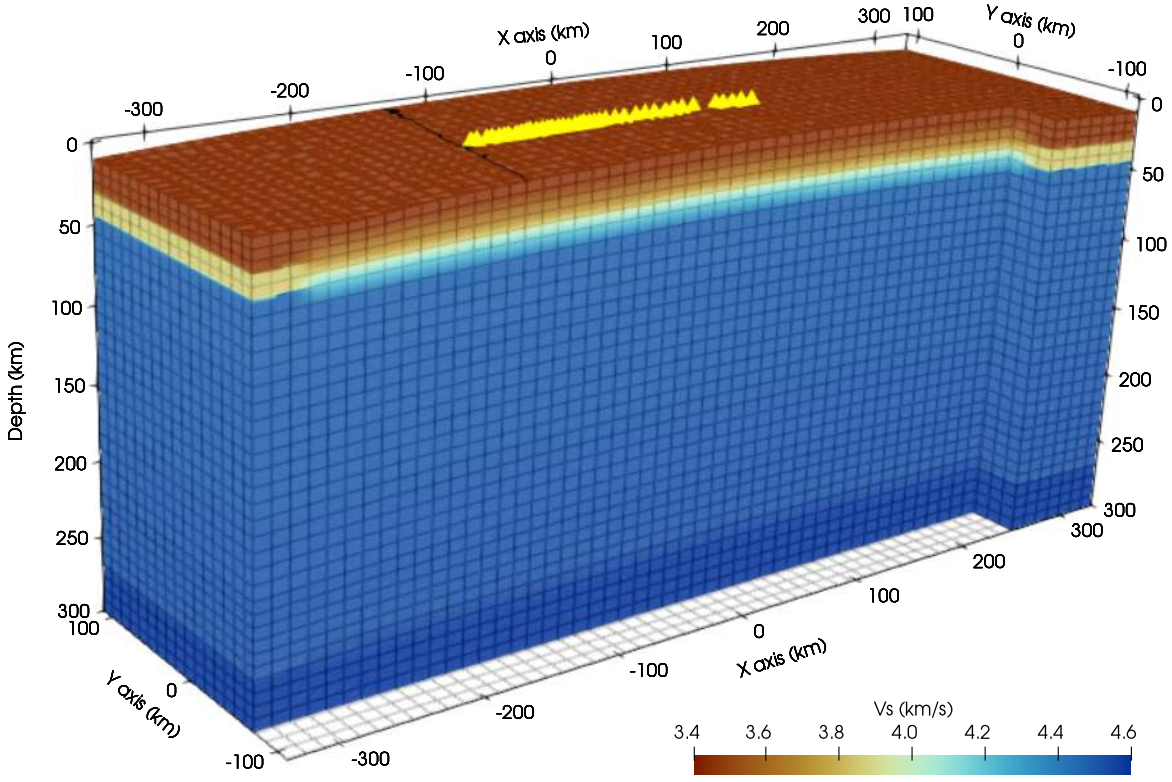


Figure 5.5: Spectral-element mesh. The number of elements along the x, y, and z directions are 56, 19, and 33, respectively. The yellow triangles show the positions of the stations from the CASC93 experiment that were used in this study. The black contour shows the coastline. The background colors in the grid show the initial $V_S$ model, a smoothed version of the ak135 reference Earth model. Note the velocity model near the boundary of the mesh is specified for the layered ak135 model, to be consistent with the one used to compute the external wavefield.

longitude and 2 degrees in latitude, with hexahedral elements of about 6 km in size. Surface topography from the ETOPO1 Global Relief Model is applied on the upper edge of the spectral-element mesh. The mesh goes from the surface down to 300 km depth, the depth extension of the subducting slab in regional tomographic models beneath central Oregon (e.g. Obrebski et al., 2010; James et al., 2011). We use degree-7 Lagrange polynomials in the three orthogonal directions so that the geometry of each element is defined by 8x8x8=512 nodes. The grid is composed of 35,112 spectral elements and 12,622,584 nodes or grid points. The minimum resolved period of this mesh is about 2 s. Our goal is to invert the values

of density, $V_P$, and $V_S$ at each node of the spectral-element mesh which will also be our tomographic grid.

### 5.3.4 Definition of the cost function

We invert the vertical and radial components of P waveforms, and the transverse component of SH waveforms. The time windows used in the inversions start 5 s before the main phase arrival and end 30 to 60 s after the phase (P or SH), depending on the event. These time windows are sufficiently long to include not only the direct phases but also all the reverberations and conversions on the main lithospheric interfaces.

The waveform inversion problem consists in finding the model $\mathbf{m}$ that minimizes the cost function (Tarantola, 2005)

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\text{obs}})^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\text{obs}}) + \frac{\lambda}{2}(\mathbf{m} - \mathbf{m}_{\text{prior}})^t \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}), \qquad (5.2)$$

where $\mathbf{g}$ is the forward wave equation operator, $\mathbf{C_D}$ the data covariance matrix, and $\mathbf{C_M}$ the a priori model covariance matrix. The first term in (5.2) quantifies the waveform misfits, where integration over selected time windows and summation over events and seismic phases is implicit. The second term quantifies the model deviation from the a priori model. The regularization parameter $\lambda$ is introduced so that these two terms have balanced contributions to the misfit function.

### 5.3.5 The full waveform inversion method

The full waveform inversion method follows the general workflow described in Monteiller et al. (2015) and Kan et al. (2022), to which the reader is referred for further details. The forward wave-equation problem is solved with a hybrid numerical method (Monteiller et al., 2013, 2021) that couples AxiSEM (Nissen-Meyer et al., 2014) with a spectral-element method (SEM) in a regional domain (Komatitsch and Tromp, 1999). The principle is to compute and inject the incident wavefield on the edges of the SEM grid in order to limit the time-consuming 3D computations inside a small regional domain, thereby reducing the

computational costs by several orders of magnitude.

Whereas it is a common practice to consider a diagonal model covariance matrix, i.e. to assume that model parameters are uncorrelated, Kan et al. (2022) have demonstrated that such assumption leads to inconsistent and biased model reconstructions, and that accounting for the well-documented correlations between density and $V_P$, and between $V_P$ and $V_S$ improved substantially the reconstructed density and velocity models. Following their approach, we consider a model covariance matrix of the form

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2 \mathbf{C} & r\sigma_\rho\sigma_{V_P}\mathbf{C} & r\sigma_\rho\sigma_{V_S}\mathbf{C} \\ r\sigma_\rho\sigma_{V_P}\mathbf{C} & \sigma_{V_P}^2\mathbf{C} & r\sigma_{V_P}\sigma_{V_S}\mathbf{C} \\ r\sigma_\rho\sigma_{V_S}\mathbf{C} & r\sigma_{V_P}\sigma_{V_S}\mathbf{C} & \sigma_{V_S}^2\mathbf{C} \end{bmatrix}, \tag{5.3}$$

where $r$ is the correlation coefficient between each pair of parameters, $\mathbf{C}$ the spatial covariance kernel, $\sigma_\rho$ the standard deviation of density, $\sigma_{V_P}$ the standard deviation of $V_P$, and $\sigma_{V_S}$ the standard deviation of $V_S$. For the spatial covariance of model parameters between two points located at $\mathbf{r}$ and $\mathbf{r}'$, we consider an exponential kernel of the form

$$C(\mathbf{r}, \mathbf{r}') = e^{-\sqrt{\frac{|\mathbf{r}-\mathbf{r}'|^2}{\ell}}}, \tag{5.4}$$

where $\ell$ is the correlation (or smoothing) length. Its main advantage is that it admits the simple inverse operator

$$C^{-1}(\mathbf{r}, \mathbf{r}') = \frac{1}{8\pi\ell^3}\left(\mathbf{I} - \ell^2\Delta\right)^2, \tag{5.5}$$

where $\Delta$ is the Laplacian operator (Oliver, 1998; Trinh et al., 2017). The values of the hyperparameters used in the inversions are given in Table 5.2.

Table 5.2: Hyperparameters used in the inversions.

| Periods | $\sigma_{V_P}$ (km/s) | $\sigma_{V_S}$ (km/s) | $\sigma_\rho$ (g/cm$^3$) | $r$ | $\ell$ (km) |
|---|---|---|---|---|---|
| 25-10, 8, 6 s | 0.65 | 0.37 | 0.27 | 0.97 | 5 |
| 25-4 s | 0.65 | 0.37 | 0.27 | 0.80 | 5 |

### 5.3.6 The iterative inversion algorithm

We search for the minimum of (5.2) with an iterative algorithm (e.g. Tarantola, 2005), in which the model $\mathbf{m_k}$ at iteration $k$ is perturbed by the quantity

$$\Delta \mathbf{m}_k = -\alpha_k \mathbf{H}_k^{-1} \cdot \gamma_k, \tag{5.6}$$

where $\alpha$ is the step length, $\mathbf{H}^{-1}$ the inverse Hessian, and $\gamma_\mathbf{k}$ the gradient of cost function (5.2) at iteration $k$. This gradient is given by

$$\gamma_k = \frac{\partial \chi_k}{\partial \mathbf{m}} = \mathbf{G}_k^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{g}(\mathbf{m}_k) - \mathbf{u}_{\text{obs}}) + \lambda \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m}_k - \mathbf{m}_{\text{prior}}), \tag{5.7}$$

with $\mathbf{G_k}$ the derivative or Jacobian operator given by

$$\mathbf{G}_k = \frac{\partial \mathbf{g}}{\partial \mathbf{m}}(\mathbf{m}_k). \tag{5.8}$$

The first contribution to the gradient is computed with the adjoint method (Tromp et al., 2005). Instead of computing the inverse Hessian, we use its finite-difference approximation given by the L-BFGS method (Nocedal and Wright, 2006) from the model and gradients obtained at the previous iterations. We refer the reader to Monteiller et al. (2015) for a more detailed description of the L-BFGS method and its application to waveform inversion.

The hierarchical inversion starts at long period with data bandpass filtered between 25 s and 10 s. We then run new inversions on data filtered with a lower corner period that is progressively decreased from 10 s to 8 s, 6 s, and finally 4 s, using as a starting model the

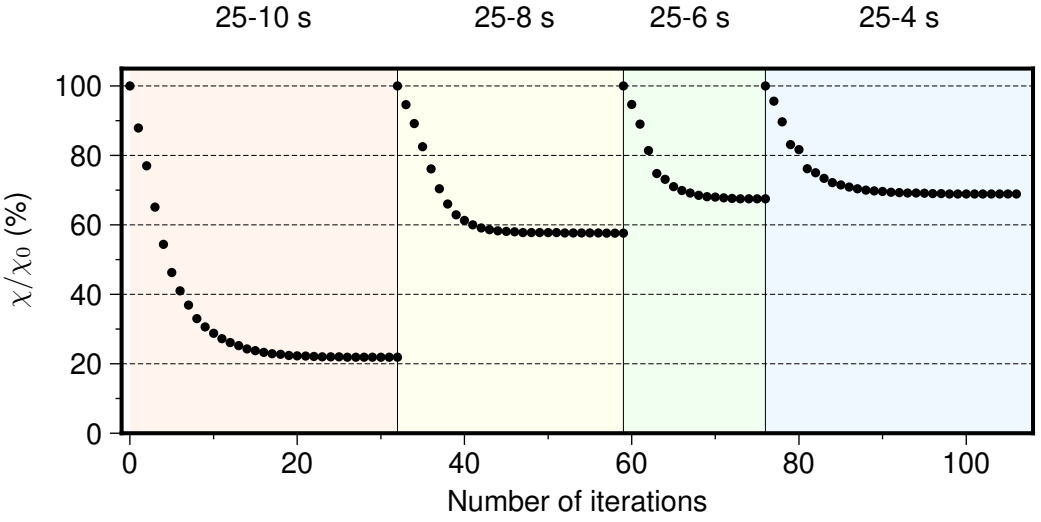


Figure 5.6: Evolution of cost reductions for the four hierarchical inversions with progressively reduced lower corner period of 10 s, 8 s, 6 s, and 4 s. The misfits are normalized to 100% when we start a new inversion at shorter period.

model obtained at the previous longer period inversion.

### 5.3.7 Event normalization

Since our teleseismic dataset contains events with different magnitudes, ranging between 6.2 and 7.5 (Table 5.1), we need to normalize the gradients of each event so that all the events contribute equally to model reconstruction. Since each event gradient is obtained by correlating the direct and adjoint wavefields (Tromp et al., 2005), we define a normalization that scales with the product of the characteristic amplitudes of the direct and adjoint wavefields. Since the adjoint wavefield is the solution of the wave equation with the adjoint sources that are the time-reversed waveform residuals at each station, it scales with the root mean square waveform residual. The normalization $N_k$ for event $k$ is thus defined by

$$N_k = \frac{1}{N(t_2 - t_1)} \sqrt{\sum_{i=1}^{N} \int_{t_1}^{t_2} s_i^2(t)\, dt} \cdot \sqrt{\sum_{i=1}^{N} \int_{t_1}^{t_2} \left[ s_i(t) - u_{obs,i}(t) \right]^2 \, dt}, \tag{5.9}$$

where $s_i(t)$ and $u_{obs,i}(t)$ are respectively the synthetic and observed seismograms at station $i$. For P waveforms, we use the vertical component to compute the normalization factors whereas for SH waveforms we use the transverse component. These normalization factors are computed at the beginning of each inversion and remain the same during iterations.

## 5.4 Results

Figure 5.6 shows the normalized cost function in the different frequency bands. The cost reduction can reach about 80 % in long period inversion. Note that the misfit reductions in the inversions are not directly comparable because the scaling factors applied at the beginning of each FWI run differ.

### 5.4.1 Tomographic model obtained after hierarchical inversion

The final density, $V_P$, and $V_S$ models obtained after hierarchical inversion are shown in Figure 5.7, using both absolute and relative scales. The most salient feature is the eastward dipping low-velocity zone (LVZ) on the western side of the profile, with compressional velocities between 5.6 and 6.2 km/s and shear velocities between 3.2 and 3.5 km/s. The top of this LVZ, observed between 20 and 40 km depth, coincides with the contour of the subducted JdF slab (Hayes et al., 2018). Since the minimum period in the inversions is 4 s, its thickness is not well resolved, but we can estimate it to be less than 10 km. In Figure 5.8, we reproduce our $V_S$ model with the main seismic interfaces reconstructed from the migration of P-to-S conversions and multiples (Bostock et al., 2002). A remarkable agreement is observed with our velocity model, and in particular for the geometry of the LVZ. The LVZ is also characterized by a rather large $V_P/V_S$ ratio $\sim$1.9. Even though this feature is on the margin of what is confidently resolved by our inversion, such a large $V_P/V_S$ ratio is in rather good agreement with estimation coming from the modeling of receiver function which suggests even larger values, that can reach up to 2.2 (Audet et al., 2009; Peacock et al., 2011). The LVZ has a shallower dipping angle in its eastern termination, which is consistent with the RF migrated section (Bostock et al., 2002). We note that the horizontal extension of

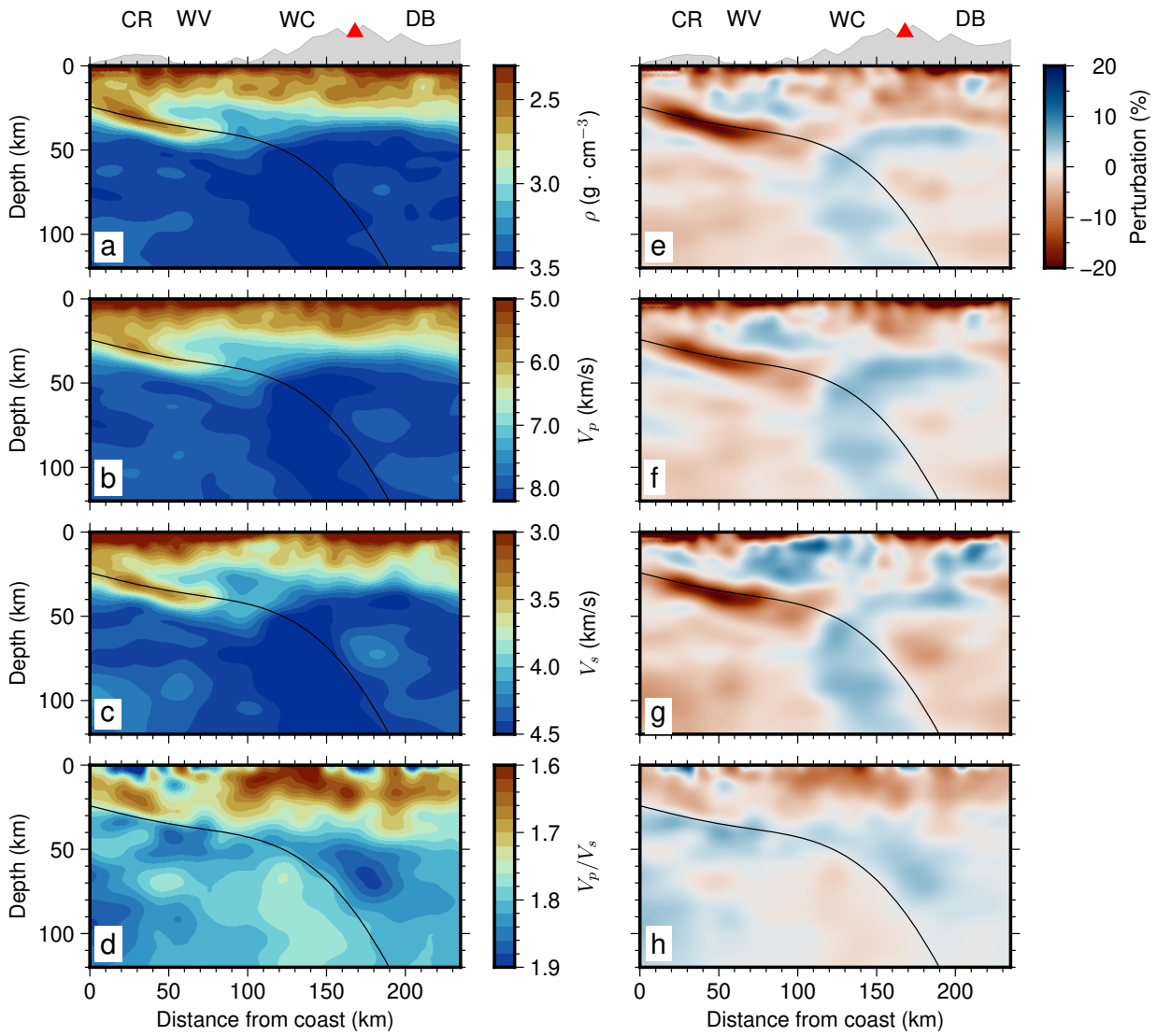


Figure 5.7: Results of the hierarchical full waveform inversion. Left: models of absolute (a) density, (b) $V_p$, (c) $V_s$ and (d) $V_p/V_s$. Right: perturbations of (e) density, (f) $V_p$, (g) $V_s$ and (h) $V_p/V_s$ with respect to the initial 1-D smooth ak135 model. The black line shows the slab contour of the JdF plate taken from Hayes et al. (2018), and the red triangle marks the position of Mount Jefferey in the volcanic arc. Physiographic provinces are the same as specified in Figure 5.1.

the LVZ is strongly correlated with the epicentral location of tremors given by Wech (2021) (Figure 5.8). Below 40 km depth, seismic velocities as well as density progressively increase

downdip and reach values of about 8 km/s for $V_P$, 3.5 km/s for $V_S$, and 3.5 g $\cdot$ cm$^{-3}$ for density. A similar though less pronounced velocity increase is observed in the subducting oceanic mantle, which has a clear signature beneath 50 km depth. The upper limit of this high-velocity body again closely follows the contour of the JdF slab.

To the east of the section, i.e. beneath the High Cascades and Deschutes basin, the crust-mantle boundary is sharp, especially in the density and $V_P$ models. In contrast, beneath the Willamette Valley and the western part of the western Cascades, it is more gradual, with a smooth gradient of density and seismic velocities from 20 to about 40 km depth. Note that the density as well as the seismic velocities in the mantle are much lower than those observed beneath the western Cascades. With a density around 3.0 g $\cdot$ cm$^{-3}$, $V_P$ around 7.0 km/s, and $V_S$ around 3.8 km/s, the reduction is about 15 % compared to the values of a typical continental lithospheric mantle. This suggests a high degree of serpentinization which can be estimated to be around 30 %, considering the experimental data from Christensen (1966) and Christensen (2004). This value is about half of the indirect estimate obtained from the analysis of scattered teleseismic waves, which is of the order of 50-60 % (Bostock et al., 2002). In the shallow crust, we image three low-velocity regions related to the accretionary prism, the forearc basin (Willamette Valley) and the backarc basin (Deschutes basin), which have a particularly clear signature in the $V_S$ model.

The $V_P/V_S$ ratio model displays several interesting features. The forearc crust is characterized by a $V_P/V_S$ ratio lower than 1.65. Such low values, which are well below those of typical crustal rocks (e.g. Christensen, 1996; Brocher, 2005a), have been previously documented in local earthquake tomography studies of the Cascadia forearc crust (Ramachandran and Hyndman, 2012; Savard et al., 2018; Guo et al., 2021). Near the surface and below the coastal range, $V_P/V_S$ is larger than 1.9, as expected for sediments saturated in fluids in the accretionary wedge. The sediments in the Willamette Valley have also a clear signature in the $V_P/V_S$ model, in good agreement with the local earthquake tomography performed by Dunham and Kiser (2020). Beneath the volcanic arc, a shallow prominent positive $V_P/V_S$

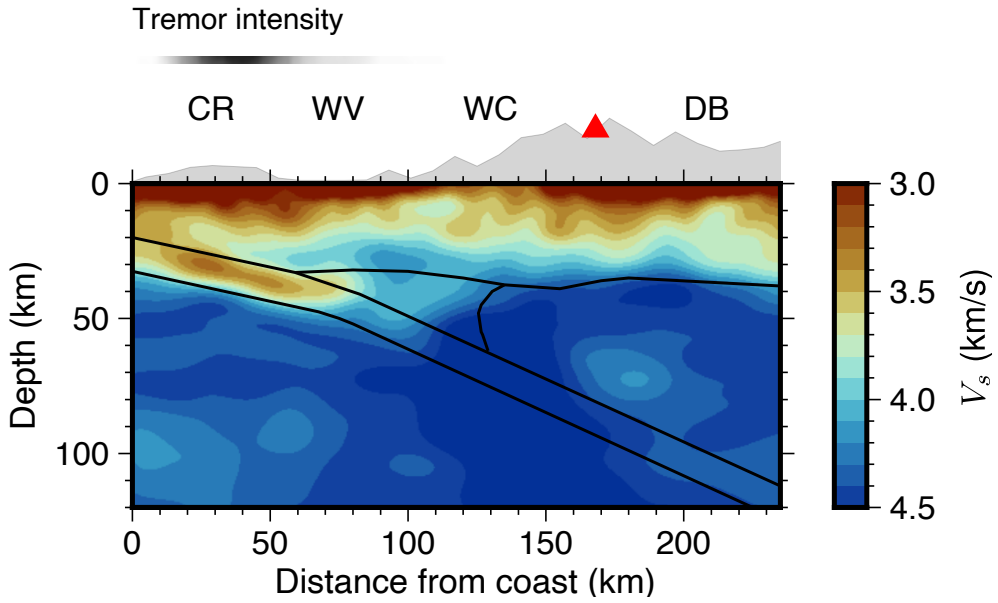


Figure 5.8: Comparison between our final $V_S$ model and the main interfaces reconstructed by migration of receiver functions (Bostock et al., 2002). The tremor intensity represents the number of tremors in every 2 km bin with a horizontal search radius of 20 km. Note the coincidence between the horizontal extent of the low-velocity zone and the occurrence of tremors.

anomaly is imaged, that could be related to a magma reservoir. Finally, a prominent high $V_P/V_S$ anomaly larger than 1.9 is observed, around 75 km beneath the volcanic arc. This region is also characterized by reduced density and seismic velocities, which could suggest the presence of melt.

### 5.4.2 Waveform fits

The P and S waveform fits for event 9 are shown in Figure 5.9. We compare the observed (black lines) and synthetic (red lines) seismograms computed in the initial smooth 1-D model and in the final 3-D model (Figure 5.7). For P waves, the fit of the vertical component before inversion is much better than the radial component. This results from the fact that the vertical component of teleseismic P waves is dominated by the direct P wave, which is highly coherent at the regional scale. It is this very property that allows us to make robust estimate of the average incoming P wavelet. The P reverberations, such as the PmP

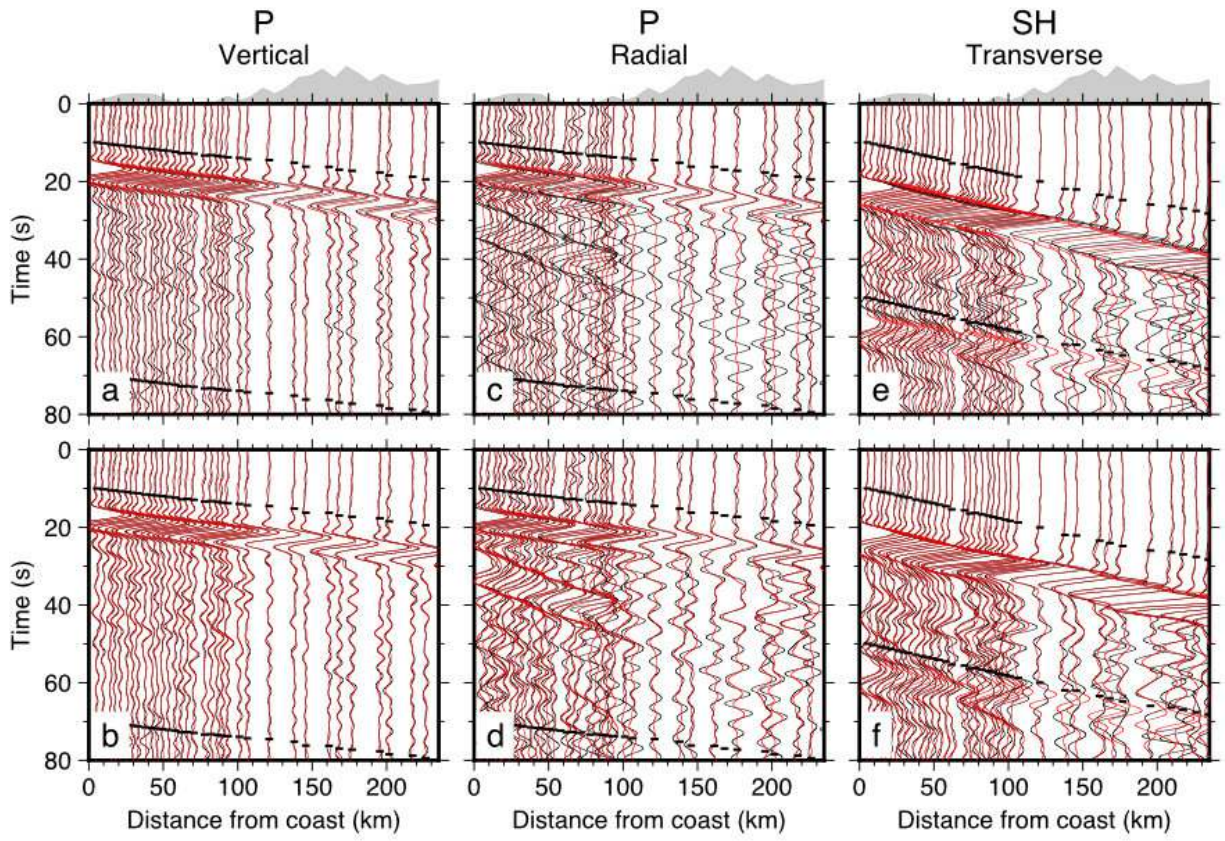


Figure 5.9: Waveform fits for event 9 at filtered between 4 and 25 s period (see Table 5.1). Observed (black lines) and synthetic (red lines) vertical component P waveforms computed in the initial (a) and final (b) models. Observed (black lines) and synthetic (red lines) radial component P waveforms computed in the initial (c) and final (d) models. Observed (black lines) and synthetic (red lines) transverse component SH waveforms computed in the initial (e) and final (f) models. The black ticks on each trace indicate the beginning and end of the time windows used in the inversion. The reductions of cost functions for the vertical, radial, and transverse components for this event are 85 %, 81 %, and 89 %, respectively. The secondary arrivals observed ∼40 s after the S wave in (e,f) are the SKS and ScS phases.

phase reflected on the Moho, generally have small amplitudes, of the order of a few percent compared to the directly transmitted P wave (Li and Nábělek, 1999). The vertical component misfits thus mainly reflect the small amplitude and travel time residuals of the P wave at each station. The radial component, on the other hand, is characterized by a more evenly distributed energy in the selected time window. Secondary arrivals in the P wave coda, such as P-to-S conversions and reflections on seismic discontinuities, have amplitudes comparable to the direct P wave on the radial component. On the western part of the transect, the signature of eastward dipping interfaces related to the LVZ, is clearly observed. Because the initial model is a smooth 1-D model, the radial component synthetic seismograms computed in this model provide a poor fit of observed data. The transverse component of SH waves is also dominated by the direct SH wave, with strong later arriving reverberations. Note that the time window selected for the inversion of SH waveforms is shorter than the one for P waveforms, in order to exclude the ScS and SKS waves that arrive ∼40 s after the S wave. After inversion, the improvement of P and SH waveforms fits is spectacular ($> 80\%$), not only for the energetic P and SH waves but also for the later arriving scattered waves.

## 5.5 Discussion

### 5.5.1 The impact of adding SH waveforms and accounting for the correlation between model parameters

To investigate the impact of adding SH waveforms and of the a priori assumption that model parameters are correlated, we performed three additional hierarchical inversions:

- Inversion of only P waveforms with a diagonal model covariance matrix, i.e. assuming that model parameters are uncorrelated
- Inversion of both P and SH waveforms with a diagonal model covariance matrix
- Inversion of only P waveforms with a complete model covariance matrix

The waveform fits for event 9 obtained after these three hierarchical inversions and the

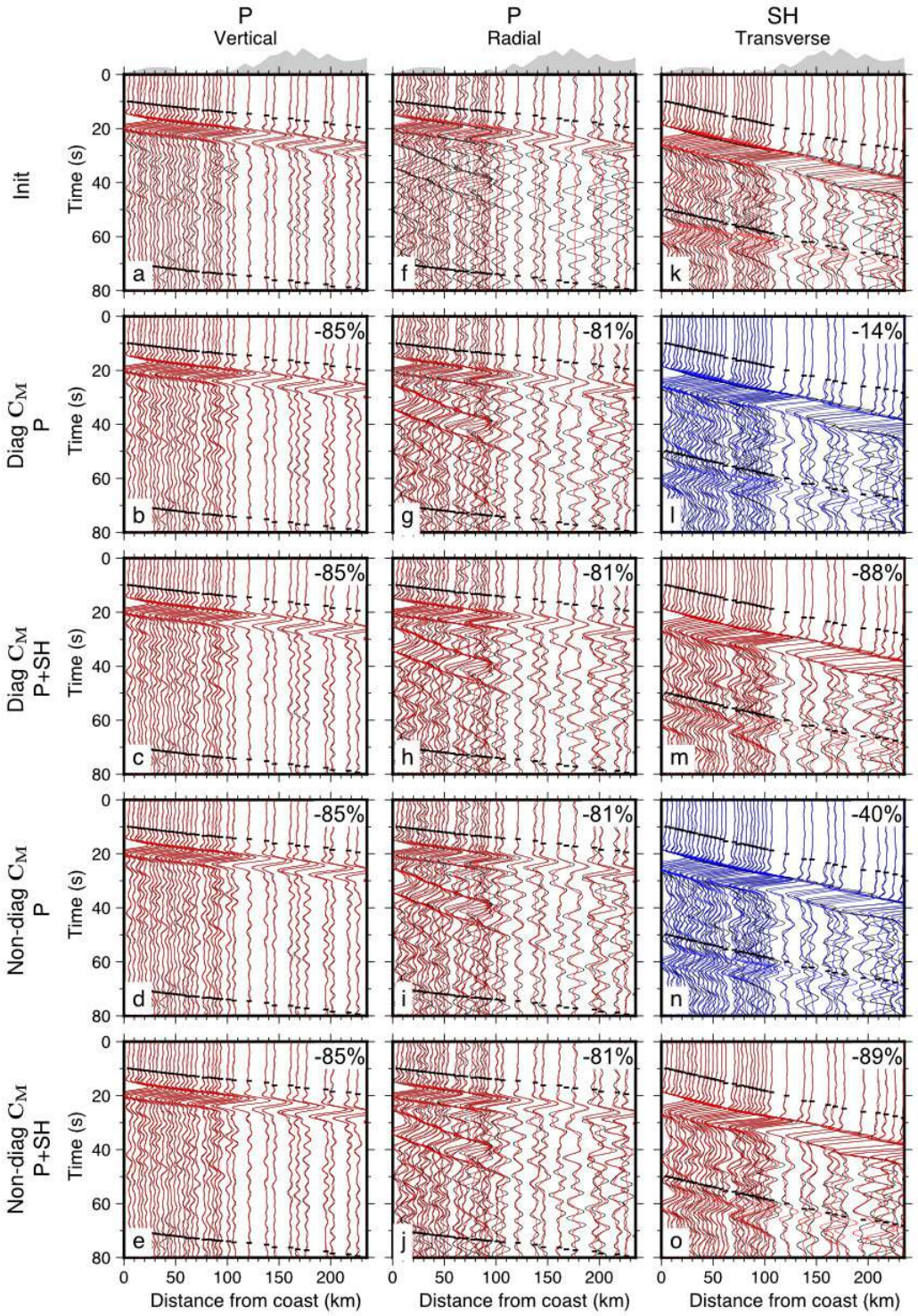


Figure 5.10: Waveform fits obtained after the four hierarchical inversions for event 9. The solid lines show the observed seismograms and the red lines show the synthetic seismograms. The blue lines show the SH synthetic seismograms that were obtained in the two inversions in which only P waveforms were inverted. The cost reduction in percentage after hierarchical inversion is specified at the upper-left of each panel.

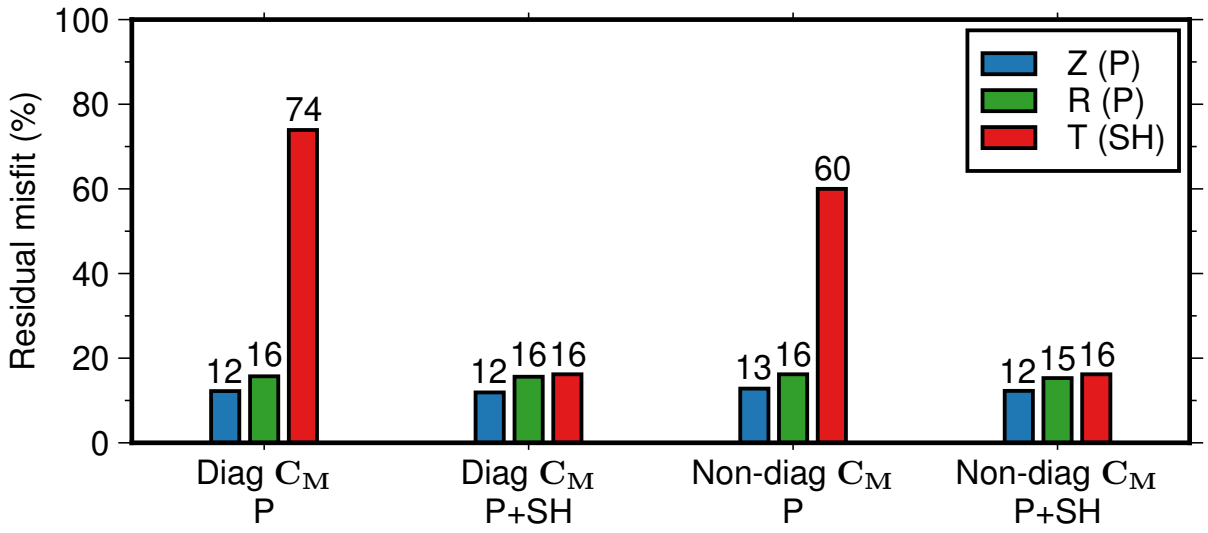


Figure 5.11: Residual misfits for the different types of inversions, defined as the ratio between the misfit in the final and the misfit in the initial smooth 1-D model (in percents).

ones already presented in the previous section are shown in Figure 5.10. For this event, the reduction of the vertical and radial P waveform misfits are similar in the four inversions. The main difference is observed for the SH waveform fits, which are largely improved when SH waveforms are included in the inversions. The inversion of only the P waveforms with a complete covariance matrix present intermediate SH waveforms fits (40% misfit reduction) which suggests that accounting for the correlation between $V_S$ and density and between $V_S$ and $V_P$ can at least partly compensate for the absence of SH waveform for the reconstruction of the $V_S$ model.

Figure 5.11 gives the waveform misfits computed on the complete dataset. We again observe similar P waveform fits in the four inversions, with a slighter larger misfit reduction on the vertical component compared to the radial component. A good SH waveform fit is only obtained when SH waves are included in the inversion.

Whereas the four inversions provided quasi-identical P waveform fits, the final $V_P$ models differ substantially (Figure 5.12). The models obtained with a diagonal model covariance matrix, i.e. assuming that density and seismic velocities are not correlated, only show long

wavelength structures such as the crust and its lateral thickness variations, as well as the subducting slab. Structural details such as the LVZ, which has a strong signature on the radial component of P waves (Figure 5.7c), is not retrieved at all. On the other hand, it is seen in the density models, as well as in the $V_S$ models, but with a reduced amplitude of the low-velocity anomaly. Intuitively, this can be understood by the fact that for a given frequency content, the wavelength of P waves is about twice that of S waves, and since in FWI the resolution approximately scales with the wavelength (e.g. Virieux and Operto, 2009) a sharper resolution is indeed expected in the $V_S$ model. Another reason is that the $V_P$ model is mainly reconstructed from transmitted P waves, whereas for the density and $V_S$ models the dominant contribution comes from P-to-S conversions and reflections. The former are mainly sensitive to long wavelength structures whereas the latter are mainly sensitive to sharp gradients and discontinuities. A consequence of the uneven resolution in the $V_P$ and $V_S$ models is that the $V_P/V_S$ model is poorly constrained, and with strong artifacts, notably in the forearc, where extreme values lower than 1.50 are observed (Figure 5.12d, 5.12h).

These results underscore the non-unicity of the inversion and the intrinsic trade-offs between density, $V_P$, and $V_S$. Small waveform misfits therefore cannot be interpreted as a reliable indicator of the quality of the inversion results. Other informed criteria describing our degree of knowledge on model parameters need to be considered. Our results confirm that accounting for the well-documented correlations between $V_P$ and $V_S$, and between density and $V_P$ (or $V_S$) lead to better-resolved models, especially of $V_P$ and $V_S$ models at depth. The common assumption (often made implicitly) that model parameters are not correlated, has a strong detrimental effect on the results of FWI.

The addition of SH waveforms, which were not considered in previous applications of FWI on real data (Wang et al., 2016; Beller et al., 2018; Wang et al., 2021; Chevrot et al., 2022), has also a beneficial though less noticeable effect. Whereas large-scale structures in the $V_S$ models shown in Figures 5.12k and 5.12o are extremely similar, small-scale structures such as the LVZ, the serpentinized mantle wedge, or the low-velocity anomaly around 75 km

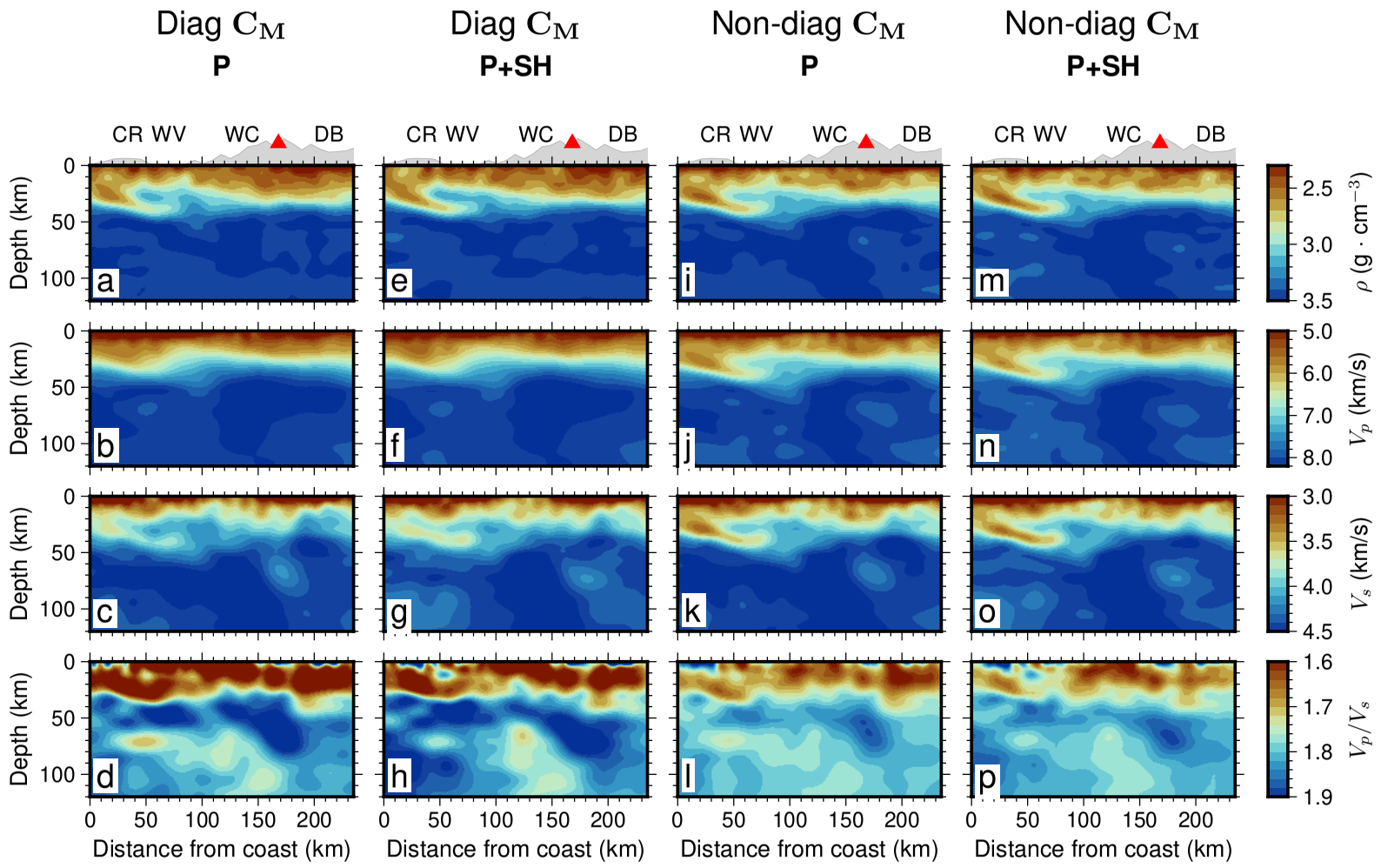


Figure 5.12: Results of the hierarchical FWI with different inversion strategies. Each column shows models of absolute density, $V_p$, $V_s$, and $V_p/V_s$ from the first row to the fourth row. (a-d) Inversion with P events and a diagonal covariance matrix. (e-h) Inversion with P and SH events and a diagonal covariance matrix. (i-l) Inversion with P events and a non-diagonal covariance matrix. (m-p) Same as Figure 5.7(a-d). Inversion with P and SH events and a non-diagonal covariance matrix.

depth beneath the volcanic arc, are more sharply defined in the inversion that includes SH waveforms. An important limitation of FWI of teleseismic P waves comes from reverberations in sedimentary basins. In the earlier applications of FWI in the Pyrenees region, strong artifacts were indeed observed in the crust and shallow mantle beneath the Aquitaine Basin (Wang et al., 2016, 2021; Chevrot et al., 2022). Similar artifacts are observed in the eastern part of the section, beneath the Deschutes Basin, where values of $V_S$ larger than 4.0 are observed in the lower crust (Figure 5.12k). Delph et al. (2018) also reported very high shear velocities in the crust from their joint inversion of surface wave dispersion curves and receiver functions. In contrast, values more typical of a continental lower crust (between 3.6 and 3.8 km/s) are obtained after joint inversion of P and SH waveforms (Figure 5.12o). Therefore, transmitted shear waves seem to provide sufficient independent information to correct the velocity model from the detrimental effects of reverberations on shallow interfaces such as the basement of sedimentary basins, which leads us to advocate the exploitation of shear waves in future applications of FWI. Beller and Chevrot (2020) came to a similar conclusion regarding the importance of jointly inverting compressional and shear waveforms for proper 3-D imaging of seismic anisotropy. We can thus foresee a growing interest of performing such joint P and S inversions in the near future.

### 5.5.2 Scaling relationships between model parameters and the effect of the initial smooth 1D model

Figure 5.13 shows a scatter plot of the $V_P$ versus density models, where each point represents a spectral-element grid, colored as a function of its depth position. The average linear trend between $\rho$ and $V_P$ in the final model (solid black line) falls close to the colored circles, which represent the parameter values in the initial model. This can simply be explained by the fact that the average travel time and amplitude residuals for each event has been removed and thus the inversion is blind to the average absolute structural model, which remains unchanged after inversion. This implies that to constrain the absolute 1D model, it will be necessary to add independent information, which could come for example from the

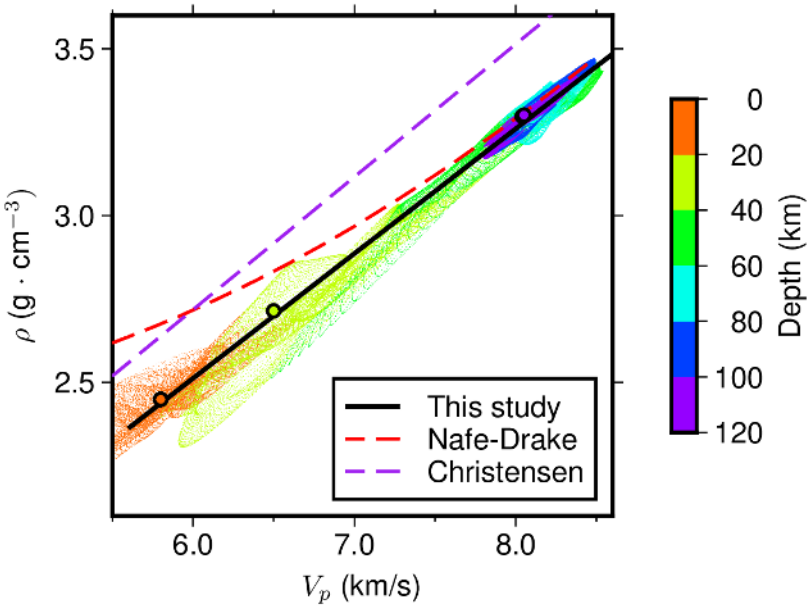


Figure 5.13: Scatter plot of $V_P$ and $\rho$ in the final model. Solid colored circles represent values in 1-D ak135 model. The solid black line represents the average linear trend of the final model. Red and purple dashed lines represent the $V_P$ versus density relationships from Nafe-Drake relationship (Brocher, 2005a) and Christensen and Mooney (1995) respectively.

dispersion of surface waves or from gravity data.

### 5.5.3 Fluid migrations, melting, and metamorphic reactions in the central Cascadia subduction revealed by FWI

Our tomographic image (Figure 5.7) reveals the 3-D structures of the Cascadia subduction in central Oregon with a wealth of details that provide crucial insight into fluid migration and related metamorphic mineral reactions in the subduction factory. Our interpretations of the model are summarized in Figure 5.14.

The nature of the ubiquitous low-velocity layer beneath the Cascadia forearc, which is a prominent feature of our tomographic model, has been discussed by Bostock (2013). It is thought to be composed of hydrated sediments and pillow basalts. Dehydration reactions release fluids that remain trapped between a low permeability seal above the subduction

interface and an underlying layer of low-porosity gabbroic rocks. Tremors in central Cascadia are weaker and less frequent than in northern or southern Cascadia (Wech, 2021). Episodic tremor and slip (Rogers and Dragert, 2003) locations from the Pacific Northwest Seismic Network (https://pnsn.org/tremor) (Wech, 2021) are closely related to the LVZ (Gosselin et al., 2020), which suggests a hydrologic control of slow slip events and tremors. The depth of tremors is not well constrained, especially in central Oregon where the station distribution is rather sparse, but they probably occur at the subduction interface (black stars in Figure 5.14), as the low-frequency events to which they are strongly related (Thomas and Bostock, 2015).

In our model, we observe a strong contrast between the backarc mantle and the forearc mantle, where density and seismic velocities are reduced by about 15% compared to the backarc mantle. The limit between these two domains is quite sharp and approximately follows the geometry of isotherm 600° in the thermal structure calculated for a hot continental subduction (Hyndman and Peacock, 2003, Figure 4a). The transition from a low-permeability layer interface to a high-permeability layer interface corresponds to the onset of eclogitization of the oceanic crust (Audet et al., 2009). The eclogitization domain is clearly observed in our model, between the LVZ and the eclogitic crust, where the density and seismic velocities are close to those of the surrounding mantle. The width of this eclogitization domain, which is about 40 km, allows us to estimate the kinetics of the reaction. The velocity of the JdF subducting plate is about 4 cm/year (DeMets et al., 1990) and thus a complete eclogitization of the oceanic crust takes $\sim$1 Myr.

The collapse of porosity releases a large amount of water into the overlying mantle wedge. This water will hydrate the olivine minerals in the peridotites, producing serpentine minerals (Hyndman and Peacock, 2003). Serpentine and hydrous minerals significantly alter the physical and mechanical properties of the forearc mantle, with density and seismic velocities reduced in proportion to the degree of serpentinization (Christensen, 2004). Back-arc mantle is too hot for hydrous minerals to be stable (Hyndman and Peacock, 2003). Serpentiniza-

tion thus leads to a weak and buoyant mantle wedge, isolated from the subduction-related circulation flow. Serpentinization reactions also produce magnetite. According to thermal modeling (e.g. Hyndman and Peacock, 2003), temperatures in the mantle wedge of a warm continental subduction zone such as Cascadia are expected to be cooler than 580°, the Curie temperature of magnetite. A long wavelength magnetic anomaly is thus expected along the Cascadia forearc. Such a large anomaly is indeed observed, associated with a gravity low (Blakely et al., 2005) and thus the hypothesis of a highly hydrated mantle wedge is also supported by potential field anomalies.

Fluid pathways in the mantle wedge are still unclear, but tomographic studies suggest that fluids do not migrate vertically (e.g. Schurr et al., 2003; Halpaap et al., 2018). Instead, they follow oblique diverging paths oriented toward the volcanic arc and/or upward the subduction interface toward the accretionary wedge, separated by a zone in the mantle wedge with low-fluid content. These intriguing observations were recently explained by numerical modeling of water migration in the mantle wedge with a two-phase flow approach (Wang et al., 2019). The water migration models in this study are characterized by the development of two water pathways with a low-porosity divide in between, which results from the bending of the overlying lithosphere that produces a zone of overpressure (which can reach up to 50% of the lithostatic pressure) and low-porosity in the mantle wedge. If we relate the decrease of density and seismic velocities observed in the serpentinized mantle wedge to the amount of mantle hydration, our model clearly suggests two diverging fluid pathways with a core that has higher density and seismic velocities, located beneath the eastern margin of the Willamette Valley. Magnetotelluric studies imaged conductive features that approximately correspond to these two hypothesized fluid pathways (Evans et al., 2014; Wannamaker et al., 2014). A first prominent conductive anomaly (resistivity $< 1\ \Omega\cdot$m) which diverges eastward, can be traced from the top of the slab to the Moho beneath the western Cascades, from where it extends horizontally further to the east. Interestingly, Vidale et al. (2014) reported deep long-period earthquakes (red circles in Figure 5.14) located $\sim$40 km west of the volcanic

arc, near the Moho, i.e. coinciding with the horizontal part of the first conductivity anomaly. Deep long-period earthquakes are commonly observed beneath volcanoes, associated with the movement of fluids (Chouet and Matoza, 2013). The location beneath the western Cascades near the Moho thus gives further support to the hypothesis of fluids rising from the slab, percolating through the serpentinized mantle wedge, and reaching the base of the forearc crust. The second conductive anomaly follows the plate interface updip and emerges offshore, at the front of the accretionary wedge.

Around 100 km depth, fluids released by dehydration of the slab reduce the solidus temperature and produce partial melting. The presence of melt will lower the shear velocity and, to a lower extent, the compressional velocity, and thus a positive $V_P/V_S$ anomaly is a good indicator of partial melt. Our model indeed reveals an anomaly that possesses all the attributes of partial melt, beneath the volcanic arc at around 70 km depth, which could indeed be the deep source of the magma plumbing system. Magnetotelluric data are equivocal on whether there is a conductor that could be related to partial melt at this depth, because of the mask produced by the strong shallow conductive anomaly beneath the arc (Evans et al., 2014).

The western Cascadia crust is characterized by an anomalously low $V_P/V_S$ ratio, which can only by explained by large-scale silica enrichment (Christensen, 1996; Brocher, 2005a; Chevrot and van der Hilst, 2000). This enrichment requires massive transport of fluids saturated in silica that migrate upward through the forearc and backarc crust and precipitate when the solubility of silica drops because of the temperature decrease (Manning, 1994). Massive and pervasive mineralized quartz veins are indeed a well-documented ubiquitous feature of forearc crusts (e.g. Breeding and Ague, 2002; Ramachandran and Hyndman, 2012; Halpaap et al., 2018), which may have contributed to the silica enrichment of continents. These quartz veins often host gold deposits which result from extensive leaching of the crust by hot fluids (Kerrich, 1999).

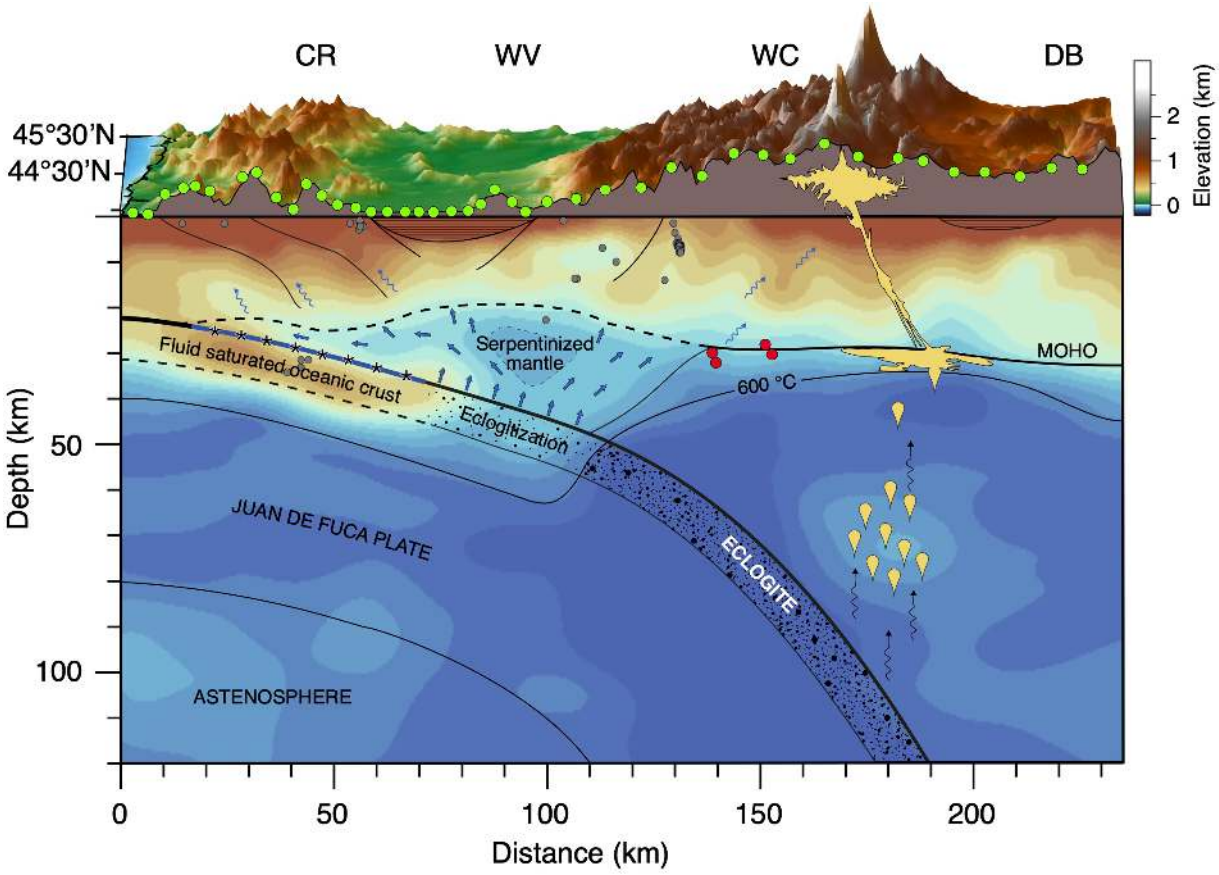


Figure 5.14: Interpretation of the $V_S$ model shown in Figure 5.7, with (vertical exaggerated) topography plotted on top. The green circles show the positions of the stations of the CASC93 temporary experiment. The gray and red circles indicate the locations of crust earthquakes (Dunham and Kiser, 2020) and deep long-period events (Vidale et al., 2014), respectively. The thick black line draws the subduction interface that corresponds to the downgoing Juan de Fuca plate. Straight thick blue arrows indicate the path of ascending fluids originating from the slab that react with the olivine minerals of the mantle wedge to produce serpentinized mantle with reduced density and seismic velocities. Wiggled thin blue arrows indicate the hypothesized paths followed by fluids in the forearc crust. The Moho beneath the forearc and the volcanic arc is defined from the 3.8 km/s isocontour in the $V_S$ model. The 600 °C isotherm taken from Hyndman and Peacock (2003) is indicated by a thin black line.

## 5.6 Conclusions

We have imaged the deep architecture of the subduction beneath central Cascadia, from the surface down to 300 km depth, by inverting complete P and SH teleseismic waveforms. Compared to classical tomographic studies, the teleseismic FWI approach simultaneously constrains density, $V_P$, $V_S$, and the $V_P/V_S$ ratio with a fine spatial resolution. The final tomographic models provide new insight into the dehydration of the subducting Juan de Fuca slab and related eclogitization metamorphic reactions. They also revealed the fluid pathways through the forearc mantle and a zone of partial melting around 75 km depth beneath the volcanic arc.

The main limitations of this study stem from the linear geometry of the acquisition, and from the short duration of the CASC93 experiment. We thus expect that more detailed and robust tomographic images could be obtained in northern Cascadia, where the stations are both more numerous and better distributed. Another important perspective will be to extend full waveform inversion to general elastic media (Beller and Chevrot, 2020) to constrain the 3D distribution of seismic anisotropy in the Cascadia subduction zone.

## Open Research

The teleseismic waveform is publicly available from Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC). The FDSN code of the CASC93 experiment is XZ (1993-1994). The FWI code is available upon request to the corresponding author, and it will be released to the public soon.

## Acknowledgments

V. Monteiller was supported by the European Union's Horizon 2020 research and innovation programme under the ChEESE project, grant agreement No. 823844. This work was performed using HPC resources from GENCI–IDRIS (Grant 2021-101085) and CALMIP (Grant P20012). We also acknowledge PRACE for awarding us access to Marconi100 at CINECA, Italy.

# Chapter 6. Application of FWI to image the normal and flat subductions beneath the northern Central Andes

We now apply our new FWI algorithm to the teleseismic P and SH waveforms recorded by several temporary deployments in southern Peru (PeruSE, CAUGHT and PULSE) during 2008-2013.

## Contents


6.1 Introduction ..... **146**

6.2 Data and methodology ..... **149**

6.2.1 Temporary deployments in southern Peru ..... 149

6.2.2 Data preparation and selection ..... 152

6.2.3 The tomographic meshes ..... 171

6.2.4 The full waveform inversion method ..... 174

6.3 FWI of PE profile data in the small mesh ..... **179**

6.3.1 The model ..... 179

6.3.2 Waveform fits ..... 181

6.4 FWI of in the big mesh with all the permanent and temporary stations installed in southern Peru ..... **184**

6.4.1 The model ..... 186

6.4.2 Waveform fits ..... 187

6.5 Discussion ..... **187**

6.5.1 Architecture and composition of the northern Central Andes crust 193

6.5.2 Slab geometry and comparison between FWI and travel time tomography ..... 195

6.5.3 Lithosphere delamination ..... 197

6.5.4 Buoyancy of the flat slab beneath central Peru ..... 198

6.5.5 The fate of the Amazonian lithosphere ..... 200

6.6 Conclusions ..... **200**

Appendix 6.A Waveform fits after hierarchical FWI inversions in small box inversion ..... **202**

## 6.1 Introduction

The Altiplano, with a width of about 400 km, a length of about 1800 km, and an average altitude of about 4 km, is the largest and highest plateau along an active subduction zone. It results from the convergence of the Nazca and South America plates, at a rate of 6-7 cm/year (DeMets et al., 1990). The high topography of the Altiplano is isostatically supported by thick crustal roots, locally up to 75 km thick. Most, if not all, of this crustal thickening can be attributed to shortening, with a marginal contribution from magmatism (Isacks, 1988; Allmendinger et al., 1997; Martinod et al., 2020). While crustal shortening and thickening are thought to have occurred over protracted periods of time, the uplift of the Andean plateau is relatively recent. It began around 30 Ma, reaching its peak around 10-7 Ma (Garzione et al., 2006, 2008). The geodynamic processes responsible for the formation of the Altiplano are still debated, but a consensus seems to emerge to link it to a change in subduction geometry, with a sequence of flat-slab subduction followed by subduction steepening (Ramos and Folguera, 2009; James and Sacks, 1999; Martinod et al., 2020). Indeed, an episode of flat-slab subduction has been documented in southern Peru between 14°S and 20°S, from 35 to 25 Ma, during which the volcanic arc disappeared (Ramos and Folguera, 2009). This episode of flat subduction was followed by a southward retreat of the volcanic arc, between 29 and 15 Ma, accompanied by magmatism and eruption of large volumes of ignimbrites (Mamani et al., 2009). This widespread volcanism has been attributed to asthenospheric flow resulting in the decompression and melting of the mantle in the space opened by the steepening of the subduction (James and Sacks, 1999). The origin of the flattening of the slab beneath central Peru is still controversial. It has been explained by the subduction of the Nazca ridge (Gutscher et al., 2000). However, numerical modeling suggests that subduction of a ridge is not sufficient in itself to produce slab flattening (Gerya et al., 2009) and the other mechanisms, such as the overriding plate velocity (Espurt et al., 2008), slab suction (Manea et al., 2012), or slab retreat (Guillaume et al., 2009; Faccenna et al., 2021) must be invoked to maintain a shallow slab dip.

Seismic imaging can provide first-order constraints on the modern lithospheric structure of the Central Andean Plateau but the lack of seismic stations, especially in the south Peru region, has hampered tomographic studies. To compensate for the lack of permanent stations in southern Peru, several temporary experiments (PeruSE, CAUGHT, PULSE) were deployed in that regions over the last two decades. The data of these experiments were used in several tomographic studies that revealed the architecture of the Altiplano beneath southern Peru and the transition from the normal subduction beneath southern Peru to the flat-slab subduction beneath central Peru.

Receivers functions studies have shown that the forearc crust is thin, with a Moho depth around 25 km near the coast, whereas it is considerably thicker beneath the Altiplano, where the Moho depth can locally reach values up to 75 km (Phillips et al., 2012; Phillips and Clayton, 2014). These studies also evidenced P-to-S conversions on a mid-crust interface at about 40 km depth beneath the Altiplano, corresponding to a positive downward $V_S$ contrast, which was interpreted as the top of the underthrusted Brazilian shield by Phillips et al. (2012). Ma and Clayton (2014) imaged the lithosphere beneath southern Peru with surface waves from ambient noise correlations and earthquakes. Considering a period range of 6-67 s, they were able to image structures down to 160 km depth. They documented a layered crust beneath the Altiplano, with low shear velocities, between 3.2 and 3.5 km/s from the surface down to 30-40 km depth, and a lower crust with much faster velocity, between 3.6 and 4.0 km/s. Following a similar approach, Ward et al. (2016) confirmed this layered structure of the crust but reported even lower shear velocities in the upper layer, between 2.8 and 3.5 km/s. A local P-wave travel time tomography revealed a similar layering of the crust in northern Chile (Graeber and Asch, 1999) whereas Yuan et al. (2000) also documented a pervasive intracrustal low velocity layer, 10 to 20 km thick, beneath the Altiplano and Puna Plateaux, from the analysis of P receiver functions.

Beneath the Moho, between 80 and 120 km depth, a low-velocity anomaly located east of the volcanic arc has been reported by Ward et al. (2016), which was interpreted as a

concentration of fluids released by the Nazca slab. They also found a localized sub-Moho high-velocity anomaly beneath the central part of the Altiplano, with $V_S > 4.5$ km/s, thought to represent a dense piece of lithosphere in the process of delaminating.

The geometry of the Nazca slab and the transition from normal to flat subduction have been characterized with various tomographic approaches. The analysis of P-to-S conversions revealed a normal ($\sim$30°) dipping subduction interface beneath southern Peru and a flat segment marked by a positive velocity contrast at around 100 km depth beneath central Peru, with a smooth and gradual transition between the two segments (Phillips et al., 2012; Phillips and Clayton, 2014). These results on the geometry of the Nazca slab beneath Peru are in good agreement with the distribution of seismicity (Dougherty and Clayton, 2014; Phillips and Clayton, 2014) and with the results of regional travel time tomography (Scire et al., 2016, 2017; Portner et al., 2020) or surface wave tomography (Ma and Clayton, 2014; Ward et al., 2016).

This brief overview of previous tomographic studies in the northern Central Andes clearly highlights that many aspects regarding the structure of the lithosphere and of the underlying subduction, which hold the key to understanding the geodynamical processes responsible for the crustal thickening and uplift of the Altiplano, remain controversial. The objective of this study is therefore, to refine the tomographic images in this region, by exploiting teleseismic P and SH waveforms with a full waveform inversion approach. We follow the methodology described in Kan et al. (2022), which introduced a new formalism involving a full non-diagonal model covariance matrix that takes into account the correlation between density, $V_P$, and $V_S$, as well as their spatial correlation, described by an exponential kernel that can be seen as a smoothing operator. This approach allowed us to obtain finely resolved models of density, $V_P$, $V_S$, and $V_P/V_S$, from the surface down to the top of the transition zone beneath southern Peru.

The manuscript is organized as follows. We first describe the dataset and the waveform selection procedure, during which various inconsistencies were discovered between the dif-

ferent deployment phases of the PeruSE experiment. From a careful analysis of the travel time and amplitude residuals measured on teleseismic P and SH waves recorded by the stations of the PeruSE experiment, we were able to document constant and systematic shifts in phase and amplitude between the PE and PF transects. We also find that the stations that were removed from these two profiles to construct the PG transect have retained the same phase and amplitude shifts. We then perform two hierarchical waveform inversions on the selected P and SH waveforms. The first inversion is performed in a small tomographic grid surrounding the PE profile, which shows the internal coherence of the travel time and amplitude residuals. The second inversion is performed in a large tomographic grid that contains all permanent and temporary stations installed in southern Peru, after applying phase and amplitude corrections to the stations belonging to the different PeruSE transects. We then present the results of these two inversions and discuss the main features of the final models. Finally, we provide preliminary interpretations of the density, $V_P$, $V_S$, and $V_P/V_S$ models in terms of geodynamic processes responsible for crustal thickening and uplift of the Central Andes.

## 6.2 Data and methodology

### 6.2.1 Temporary deployments in southern Peru

Three temporary arrays were deployed between 2008 and 2013 in our study region, which covers the northern part of the central Andes: the Peru lithosphere and Slab Experiment (PeruSE), Central Andes Uplift and Geodynamics of High Topography (CAUGHT) and PerU Lithosphere and Slab Experiment (PULSE) (Fig. 6.1).

The PeruSE project (TO) aimed to study the transition from normal to flat subduction beneath southern Peru (Clayton, 2013). The core of the PeruSE experiment consisted of three dense transects deployed from July 2008 to January 2013, in four consecutive stages (Figs. 6.2 and 6.3). All the stations were equipped with Guralp CMG-3T broadband sensors. The first 300 km long SW-NE PE transect was deployed from Mollendo, near the coastline,

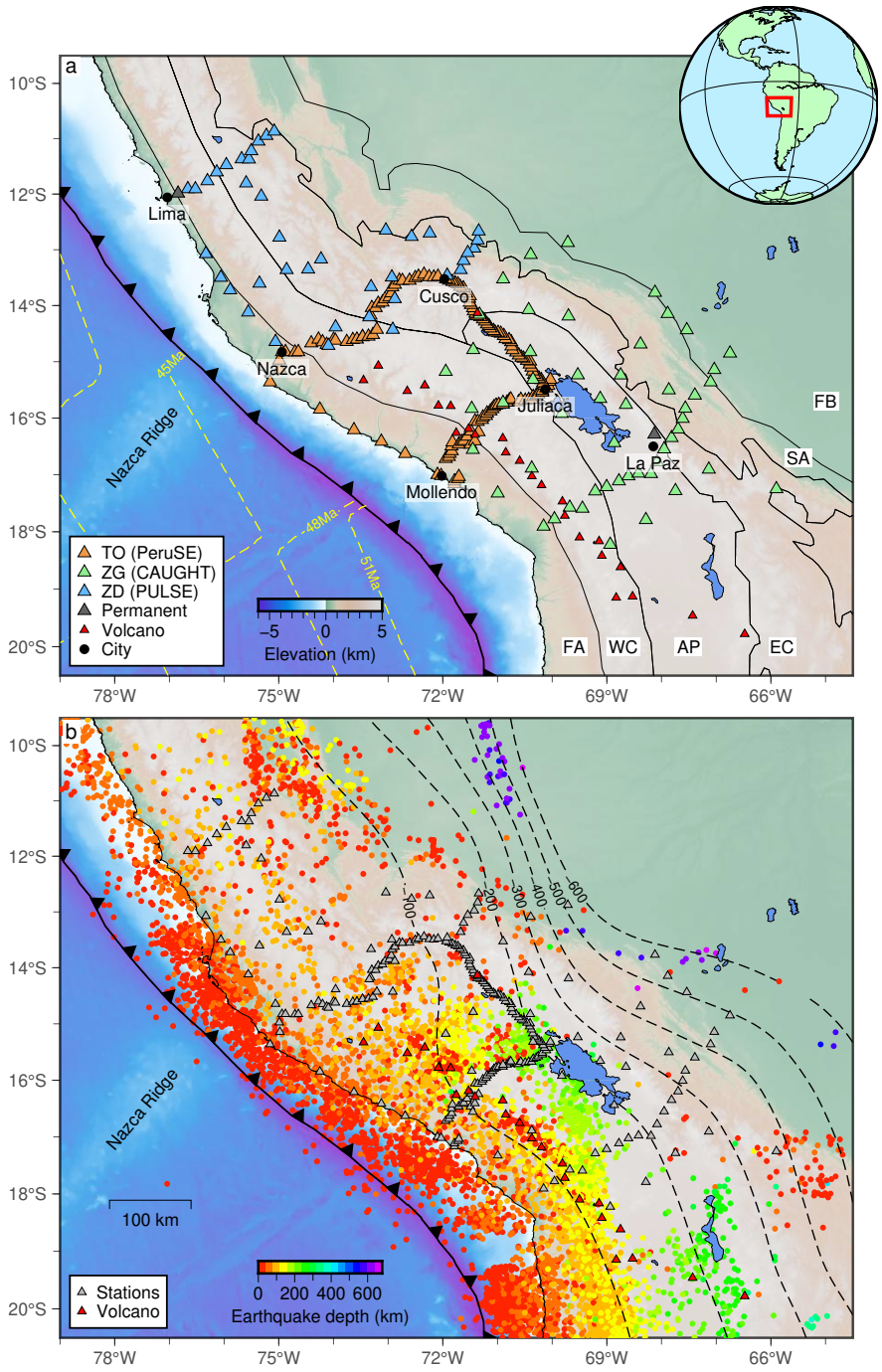


Figure 6.1: (a) Stations of the PeruSE (brown triangles), CAUGHT (green triangles), and PULSE (blue triangles) temporary deployments and physiographic provinces. Yellow dashed lines indicate the ocean floor age. FA: Forearc, WC: Western Cordillera, AP: Altiplano, EC: Eastern Cordillera, SA: Subandean zone, FB: Foreland Basin. (b) Map of local earthquakes with a magnitude larger than 3 extracted from the USGS catalog. The black dashed lines represent the contours of the Nazca slab from the Slab 2.0 model (Hayes et al., 2018).

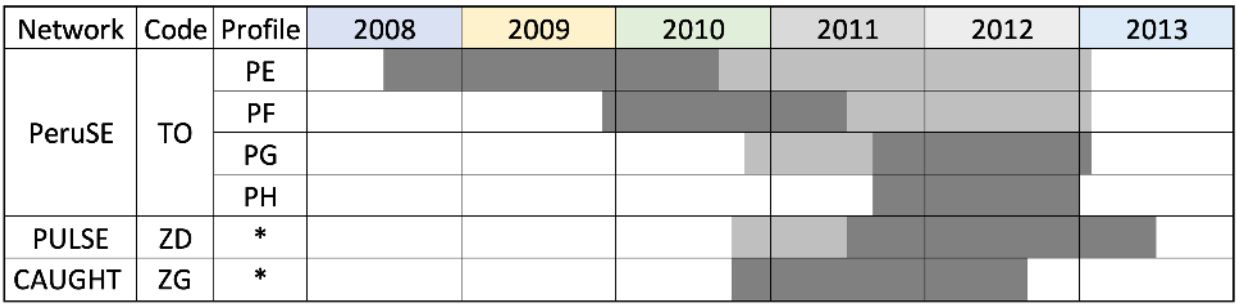


Figure 6.2: Data availability chart of the three temporary experiments in southern Peru between 2008 and 2013. Deep gray represents the complete deployment configuration, while light gray shows the configuration with half the density.

to Juliaca, near Lake Titicaca, with an average station spacing of about 6 km. It operated from June 2008 to August 2010, when half of the stations were moved to the third PG transect. The remaining 25 stations on the PE transect were operational till January 2013. The second 250 km long NW-SE transect (PF) was deployed from Juliaca to Cusco to cover the transition from normal subduction to the south to flat subduction to the north, with 49 stations spaced about 6 km. This second transect is approximately perpendicular to the first and third transects. The complete PF transect ran from December 2009 to June 2011, and in August 2011 half of the sensors were moved to the third transect PG. The third 400 km long PG transect was deployed from Nazca to Cusco, to cover the flat subduction with 50 stations spaced by about 8 km. The orientation of this third transect is approximately parallel to the first PE transect. From November 2010 to June 2011, it operated with the 25 stations taken from the PE transect [1], and after June 2011, after the addition of stations from the PF transect, it was fully operational with its 50 stations until the end of the deployment in January 2013. The fourth PH transect had only four stations deployed from August 2011 to December 2012 along the coast between the southern terminations of the PE and PG transects. These additional stations were mainly intended to supplement the acquisition

[1] According to the data availability of PeruSE (http://web.gps.caltech.edu/ clay/PeruWeb/PeruSE.html), we speculate that the following PG stations were taken from PE transect: PG01, PG05, PG07, PG10, PG11, PG15, PG17, PG18, PG20, PG24, PG26, PG28, PG30, PG32, PG34, PG36, PG38, PG40, PG42,PG46, PG48 and PG50.

of ambient noise and surface wave tomography applications (e.g. Ma et al., 2013; Ma and Clayton, 2014).

The main objective of the PULSE (ZD) experiment was to image the flat subduction beneath south-central Peru with 40 broadband stations deployed between Lima and Cusco, primarily along three SW-NE profiles (Wagner et al., 2010). Ten stations were first deployed in October 2010, followed by an additional 30 stations in May 2011. The entire network was dismantled in June 2013.

The CAUGHT experiment (ZG) was designed to discriminate the end-member geodynamic models explaining the recent uplift of the central Andes, by providing new tomographic images of the crust and mantle in southern Peru and Bolivia (Beck et al., 2010). It consists of a backbone and of an SW-NE trending transect of broadband instruments deployed in southern Peru and Bolivia, from November 2010 to August 2012.

In addition, we have also included the data of stations LPAZ and NNA, the only permanent stations installed in this area.

### 6.2.2 Data preparation and selection

The catalog of teleseismic events and waveform data were both retrieved from the IRIS datacenter. We downloaded the P and SH waveforms corresponding to the events with the magnitude larger than 5.8 and in an epicentral distance range between 30° to 90° from southern Peru. The raw waveforms data are preprocessed by removing the mean, trend, and station response from each trace. We follow the same procedure as described in Section 3.2 to select the P and SH waveform data. The travel time and amplitude residuals for each event are scrutinized to identify and discard problematic stations. To visualize the waveforms and maps of travel time and amplitude residuals, we use a MATLAB graphical interface that we developed for this purpose. Figure 6.4 illustrates the GUI interface that we have developed for data selection on the P waveforms of the 2012/04/12 event. The interface allows us to visualize simultaneously the three components of the teleseismic waveforms, as well as the maps of travel time and amplitude residuals.

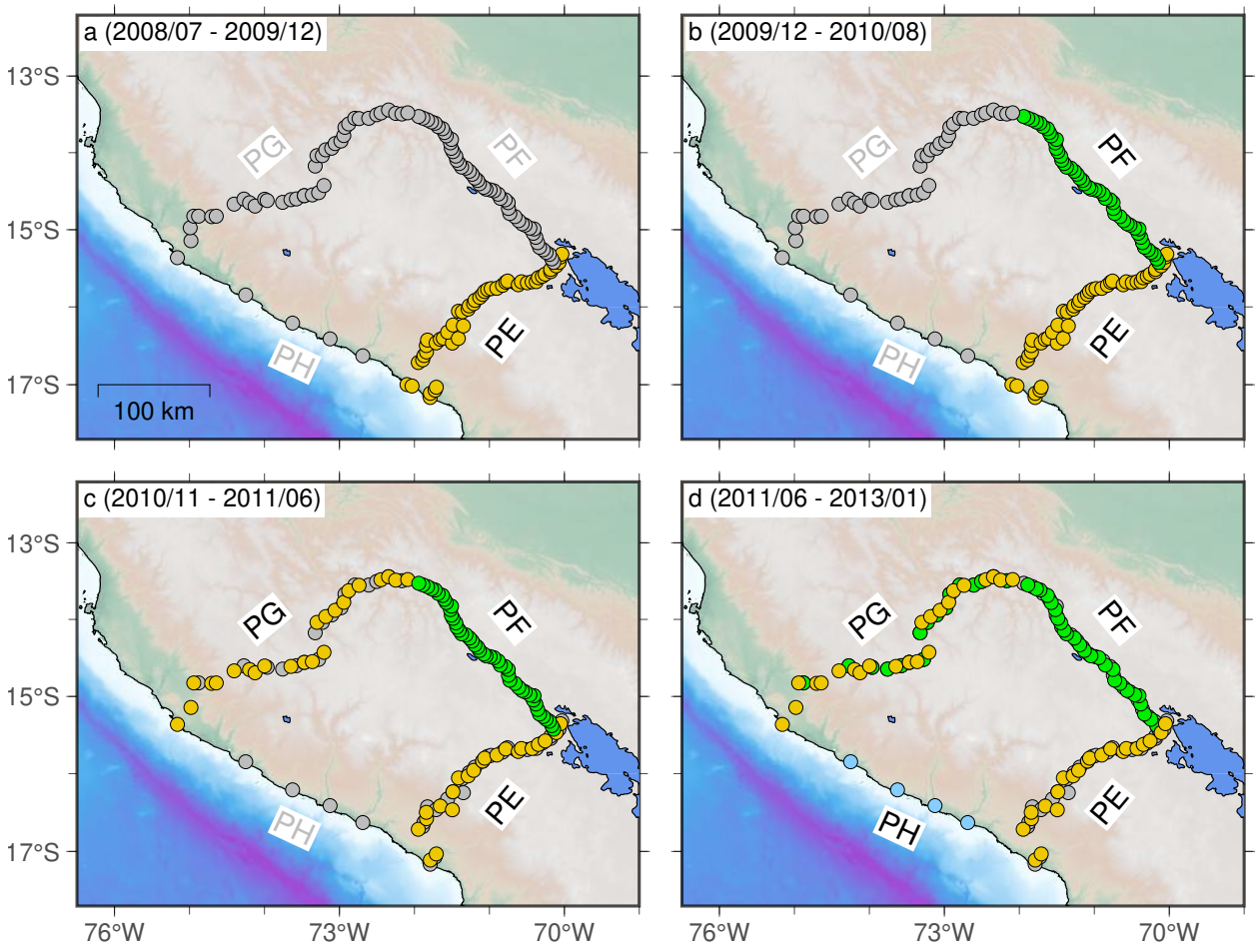


Figure 6.3: Four-stage evolution of the PeruSE deployment (Clayton, 2013). (a) Full configuration of PE. (b) Full configuration of PE and PF. (c) Half configuration of PG with stations from PE. (d) Full configuration of PG and PH with stations from PF. Note that to emphasize that the PG line is composed of stations from both PE and PF, we intentionally used the same color for it as we did for the PE and PF lines.

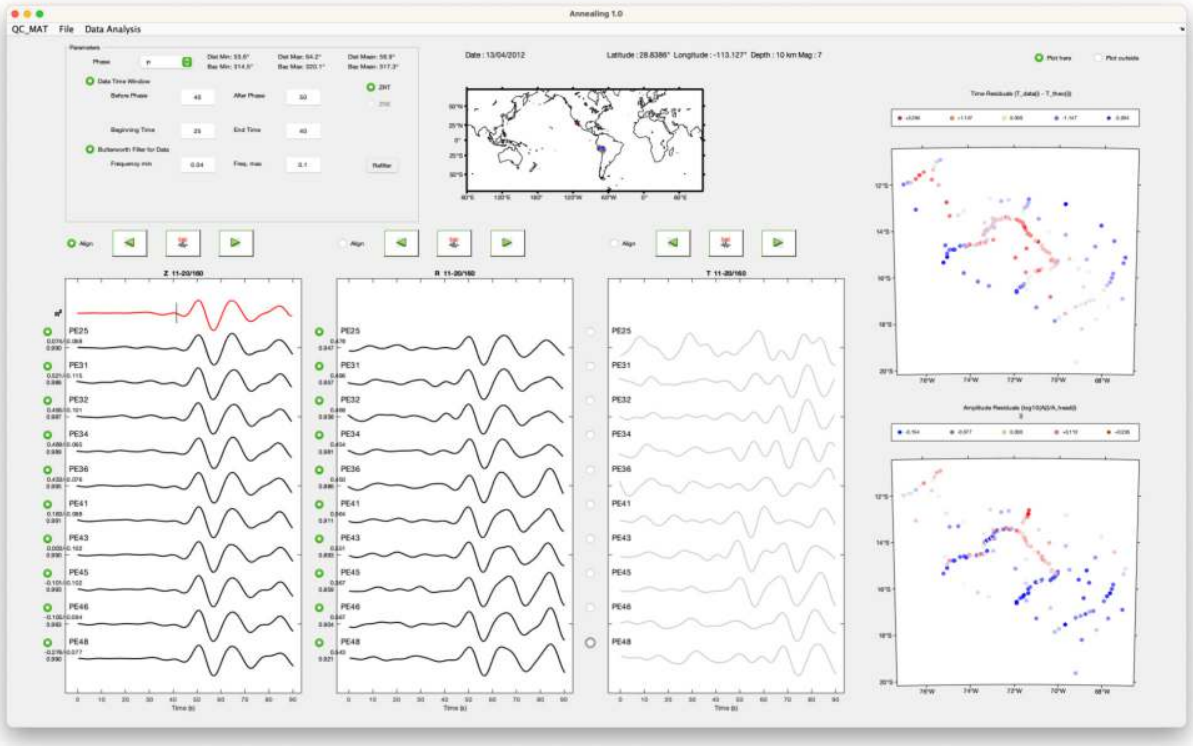


Figure 6.4: Example using the Matlab interface to visualize the waveform, the time, and the amplitude residuals for the 2012/04/12 P event.

#### *6.2.2.1 Problem in integrating waveform data from the different temporary deployments*

A close look at the travel time and amplitude maps (Fig. 6.5) reveals various important anomalies. For example, an abrupt change in travel times and amplitudes is observed between the northernmost stations of the PE transect and the southernmost stations of the PF transect (black circle in Fig. 6.5). Inconsistent variations in travel times and amplitudes are also observed along the PG transect as shown by the color changes from blue to red. In fact, this problem is also present for all other events recorded by the different transects. These incoherent and erratic variations of travel time and amplitude suggest potential problems in the data.

Fig. 6.6 shows scatter plots of travel time and amplitude residuals measured on the stations of PE, PF, and PF profiles for the 12/04/2012 event. We first focus on the PE and PF transects. The variations of the travel time residuals along the two profiles are smooth (Fig. 6.6a). However, we can clearly observe a significant gap between the two profiles, particularly clear at their crossing point, where for example stations PE50 and PF02, which are only about 14 km apart, show a very large difference in travel time, of the order of 1 s. The amplitude residuals on the PE and PF profiles also show significant differences (Fig. 6.6b), the amplitudes on the PF profile are on average larger than those on the PE profile. If we consider again stations PE50 and PF02, which are very close to each other, we observe a difference in amplitude of about 1.4, a very important (and unrealistic) variation over such a small distance (less than 15 km).

We now look at the third PG profile, which was deployed using instruments taken on the PE and PF lines. Travel time and amplitude problems are clearly present with steep and erratic spatial variations in travel time and amplitude residuals on the full PG profile (Figs. 6.6a and 6.6b). The travel time and amplitude residuals allow us to distinguish two groups of stations, which correspond to stations from either the PE or the PF profiles. There appears to be a positive offset of 1 s between stations from the PF and PE profiles. The amplitude residuals are also systematically larger at stations from the PF profile, by about

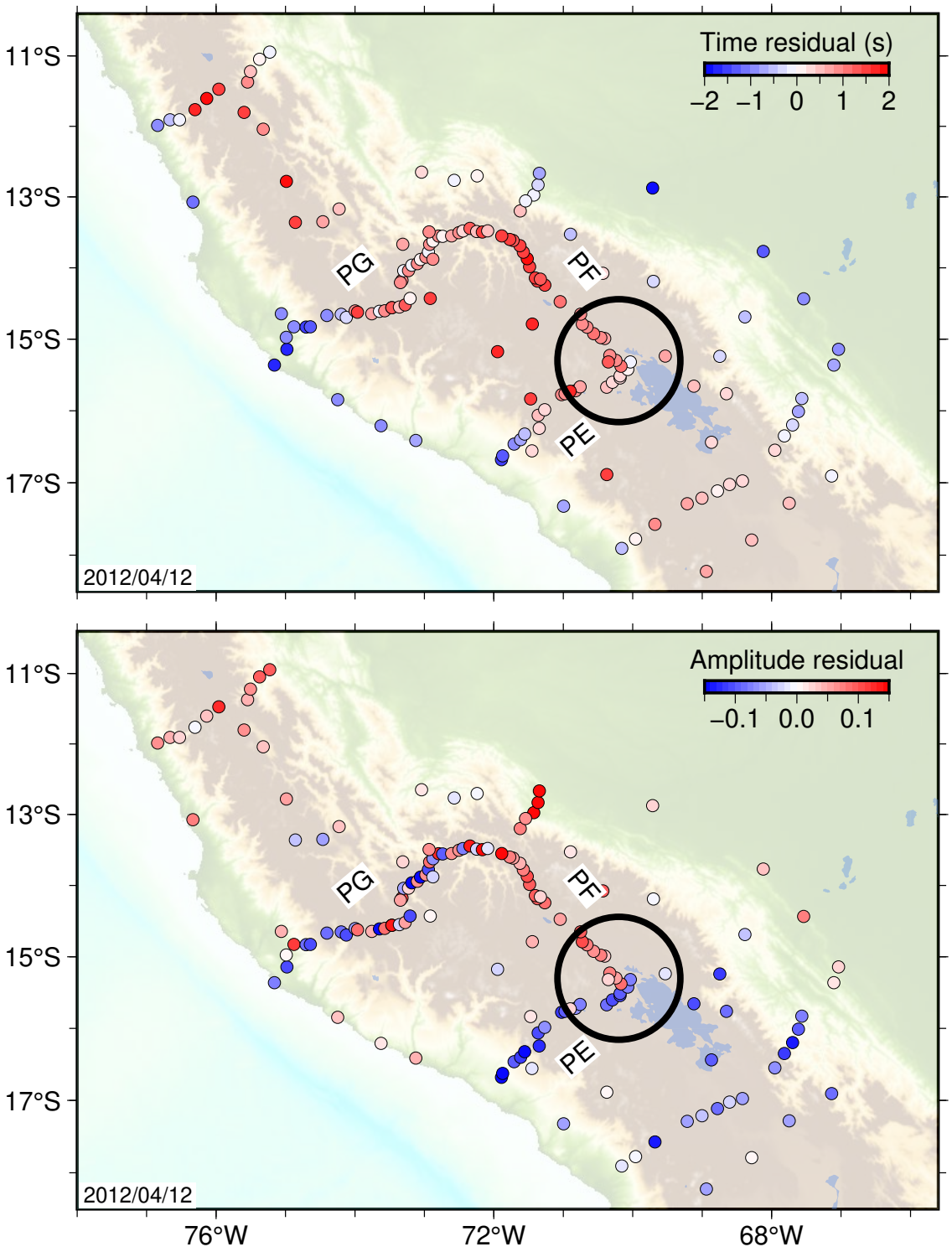


Figure 6.5: Zoom-in of the residual maps for the 2012/04/12 event shown in the GUI example (Fig.6.4). The black circle emphasizes the region where an abrupt shift of travel time and amplitude residuals is observed between profiles PE and PF. Note also the abrupt color changes of the PG transect.

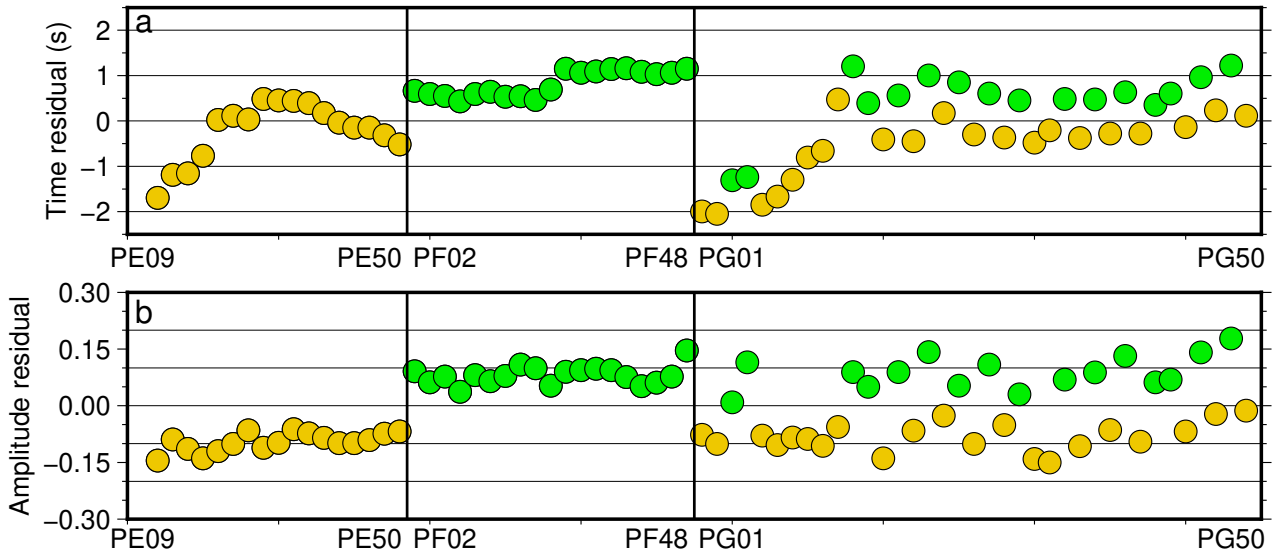


Figure 6.6: Scatter plots for the travel time residuals (a) and the amplitude residuals (b) of the 2012/04/12 event (showing the PeruSE stations only for simplicity). Note that the orange dots represent the instruments in/from the PE transect, while the green ones represent the instruments in/from the PF transect.

0.15 (in logarithmic scale).

#### *6.2.2.2 Data corrections*

At this point, we have documented inconsistent variations in travel times and amplitudes between stations in the PE, PF, and PG profiles. We will now attempt to correct the waveform data for the various problems identified earlier. To do this, we need independent information, for example from nearby stations belonging to other networks, to calibrate the absolute times and amplitudes. We can assume that if two stations are close enough (say a few kilometers apart), the arrival times and amplitudes of the P wave at these two stations should be similar. The differences of travel time residuals between two nearby stations can thus be used to estimate the clock corrections to be applied to the data. Similarly, the amplitude residual difference between two nearby stations, quantified by the logarithm of

the energy ratio of the two traces

$$\Delta A_1 - \Delta A_2 = \log_{10}\left(\sqrt{\frac{\int_{t_1}^{t_2}|u_1(t)|^2dt}{\int_{t_1}^{t_2}|\overline{u}(t)|^2dt}}\right) - \log_{10}\left(\sqrt{\frac{\int_{t_1}^{t_2}|u_2(t)|^2dt}{\int_{t_1}^{t_2}|\overline{u}(t)|^2dt}}\right) = \log_{10}\left(\sqrt{\frac{\int_{t_1}^{t_2}|u_1(t)|^2dt}{\int_{t_1}^{t_2}|u_2(t)|^2dt}}\right). \tag{6.1}$$

provides an estimate of the amplitude corrections.

To quantify the average time and amplitude difference between PE and PF transect, we first consider stations from the PULSE and CAUGHT deployments. Let's take a look at Fig. 6.5 again. The spatial residuals measured at stations of the PULSE and CAUGHT experiments are consistent and smooth. This suggests that these two networks do not suffer from clock or station response problems and that they can be used as a reference to calibrate the stations of the PeruSE experiment.

In fact, several stations from the CAUGHT and PULSE experiments are close enough to PeruSE stations to be used for comparison. We select the station pairs with distances smaller than 20 km. For example, along the PE profile, stations PE32 and PE34 are close to station CP05 from the CAUGHT experiment. Along the PF profile, stations PF04 and PF06 are close to station CP11, station PF32 and PF34 are close to station CP15, and station PF48 is close to station FS07 from the PULSE experiment. Along the PG profile, station PG31 is close to station FS04, stations PG36 PG37 and PG38 are close to station FS05, stations PG40 and PG4s are close to station FS6A, and stations PG18 and PG20 are close to station FS01. Table 6.1 summarizes the station pairs that are selected to estimate time and amplitude corrections. For this purpose, we have selected 34 events with a magnitude larger than 6 and with a good SNR (Fig. 6.7).

We performed time and amplitude residual analysis for each station pair. The measurements are presented as histograms (Fig. 6.8). Fig. 6.8a shows the distribution of arrival time differences between PE and other neighboring stations. The 113 measurements approximately follow a Gaussian distribution with an average of -0.95. For the PF profile, we obtained 140 measurements of time differences with an average of around 0 s. This suggests

that the timing of PF stations is correct, whereas the stations of the PE profile have a time offset of about 1 s. Fig. 6.8b gives the difference in measured amplitude residuals between all the station pairs. The average amplitude difference between the PE stations and the neighboring stations in other networks is about -0.12 (on a decimal logarithmic scale), i.e. the amplitudes of the PE stations are on average 0.75 times smaller. The stations of the PF profile, on the other hand, show an average amplitude difference of $\sim$0.06, which means that, on average, they are amplified by a factor of $\sim$1.07. The distribution of amplitude differences suggests that the amplitudes of PE or PF stations are neither consistent with each other nor with neighboring stations from other deployments. This is a serious problem because FWI exploits the complete wavefield and therefore, any inconsistency in amplitude will result in significant artifacts in the inverted models, especially near the surface.

We will now determine whether there are internal time and/or amplitude offsets in the PE and PF profiles. To do this, we measure the average time and amplitude differences between each station and two stations that flank it along the profile, separated by about 15 km, an average distance close to those of nearby stations from other networks considered previously. For example, we select stations PE01, PE03, and PE05 instead of PE01, PE02, and PE03. We then calculate the time and amplitude differences between station PE03 and the average of stations PE01 and PE05. Fig. 6.8c and 6.8d show the distributions of the obtained time and amplitude residual differences, both of which are centered on zero, indicating that there is no time and amplitude problem in the PE or the PF profiles. Note also that the distributions of residual differences, which reflect the statistical variations of travel times and amplitudes for stations $\sim$15 km apart, are very similar in shape to those obtained from the comparisons with nearby stations from other networks. This suggests that indeed, to first order, the phase and amplitude of the stations in the PE and PF profiles can be corrected by applying a constant offset. Since the amplitude residuals are expressed in a logarithmic scale, a constant amplitude shift is equivalent to a constant normalization factor.

We now consider the PG profile that was constructed with the stations previously installed along and removed from the PE and PF profiles. We separate the stations in this profile into two groups, based on whether they previously belonged to the PE or PF profile. Since the stations were installed along the profile by alternating stations from either PE or PF, the time and amplitude residual differences are determined by calculating the difference between the residuals observed at a particular station and the two adjacent stations, which previously belonged to another profile. Fig. 6.8e and 6.8f show the resulting time and amplitude residual differences. The time difference between the stations from the PE and PF profiles is similar to that determined previously, around 1 s. The amplitude differences between the stations from the PE and PF profiles is ∼-0.18, which is also consistent with the amplitude offset between PE and PF stations determined previously (Fig. 6.8b and 6.8d).

In summary, the detailed analysis of the travel time and amplitude residuals between the different temporary arrays suggests systematic phase and amplitude offset of the stations in the PE and PF profiles, identical for all the stations. For stations in the PE profile, we need to apply a positive time shift of 1 s and scale the amplitudes by multiplying the data by a factor of 1.3. For stations in the PF profile, we need only multiply the data by a factor of 0.87. Fig. 6.9 shows the same scatter plot of the same data as in Fig. 6.6 but after travel time and amplitude corrections. We can clearly see that after correction, the systematic travel time and amplitude offsets have been removed. Fig. 6.10 shows a P wave recorded by the neighboring stations PE46 and PF02 before and after correction. Before the correction, the P wave arrives about 1 s earlier at station PE46 than at station PF02, and with a significantly smaller amplitude. After correction, the waveforms at the two stations are almost identical.

#### *6.2.2.3 Final selection of teleseismic events*

In southern Peru, the most energetic teleseismic waveforms are mostly produced by earthquakes in California (NW backazimuth) and in the South Sandwich Islands (SE backazimuth). In case we find several events with similar backazimuth and epicentral distance, we

Table 6.1: Selected station pairs for estimating the travel and amplitude shifts.

| | | | |
|---|---|---|---|
| Instruments from PE | PE32-CP05 | PE34-CP05 | PG18-FS01 |
| | PG36-FS05 | PG38-FS05 | PG40-FS6A |
| Instruments from PF | PF04-CP11 | PF06-CP11 | PF32-CP15 |
| | PF33-CP15 | PF34-CP15 | PG20-FS01 |
| | PG31-FS04 | PG37-FS05 | PG41-FS6A |

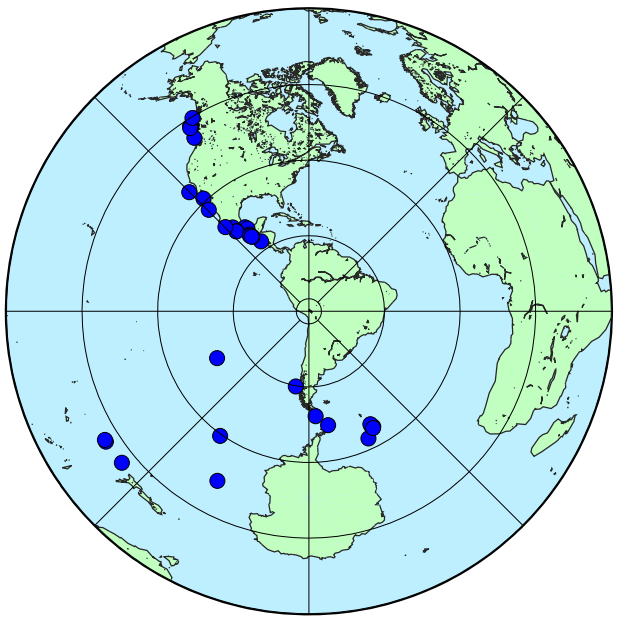

Figure 6.7: Location of teleseismic P events used for estimating time and amplitude differences between the PE and PF profiles.

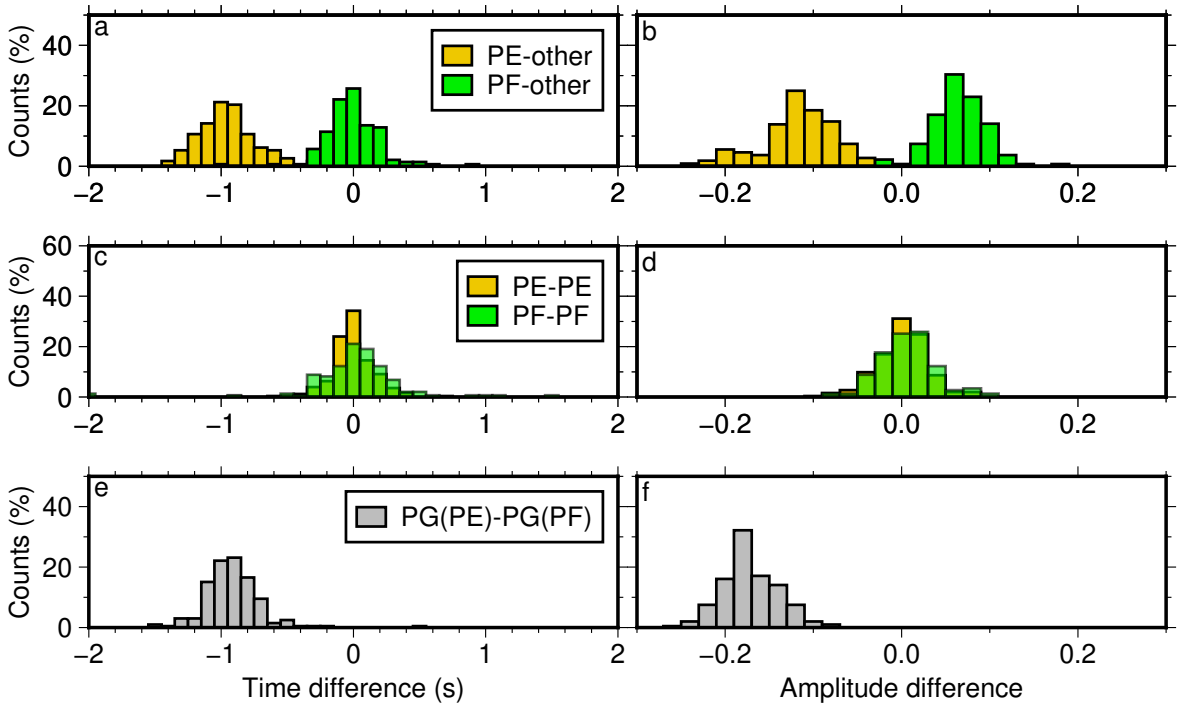


Figure 6.8: Histogram of difference of time and amplitude measurements. (a) Time differences were measured for stations in the PE (yellow) and PF (green) profiles. (b) Amplitude differences from stations in the PE (yellow) and PF (green) profiles. (c) Time differences measured with nearby stations. (d)

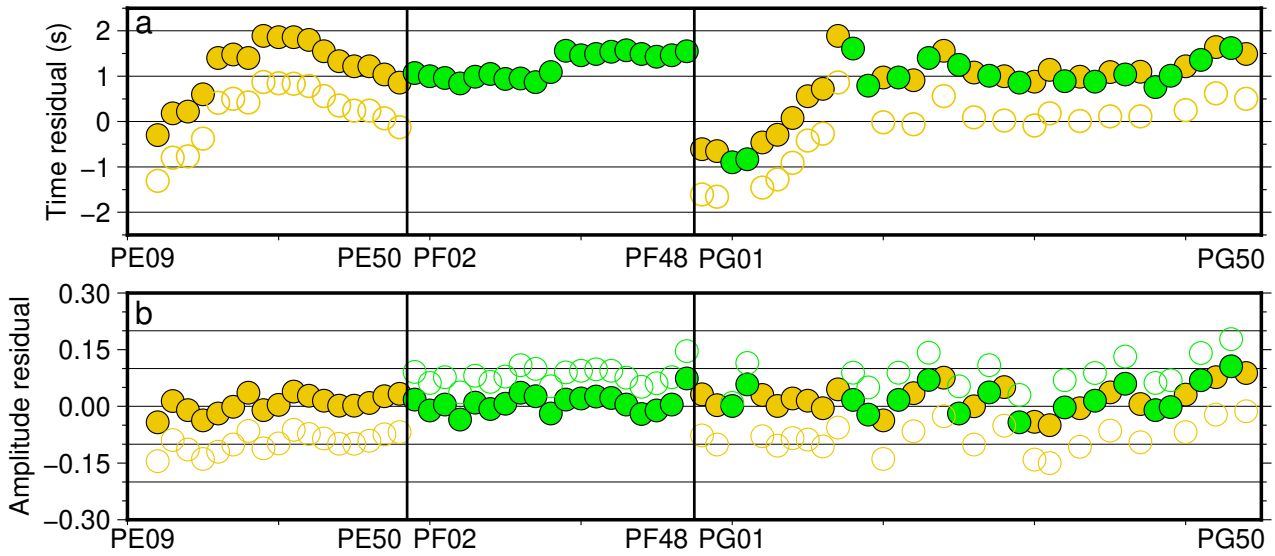


Figure 6.9: Scatter plots for the travel time residuals (a) and the amplitude residuals (b) of the 2012/04/12 event after the time and the amplitude corrections (see text). The dots without filling represent the value before the corrections (the same as Fig. 6.6).

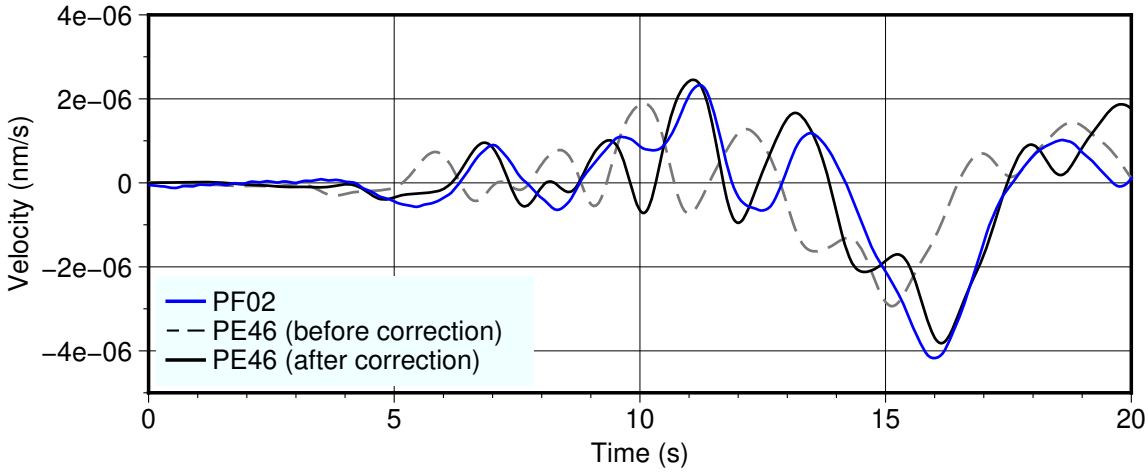


Figure 6.10: Vertical component P wave of the 2010/03/07 event recorded by stations PF02 (blue line) and PE46 before (grey dashed line) and after (black solid line) phase and amplitude corrections. All the traces are filtered between 0.04 and 1.0 Hz.

keep the one with the highest number of high-quality records. After applying the correction on PeruSE stations and re-perform the data selection, we selected 10 P and 6 SH teleseismic events with good signal-to-noise ratio at periods down to 4 s for P and 6 s for S that can be used for full waveform inversion (Fig. 6.11). Detailed event information is given in Table 6.2.

Figures 6.12 and 6.13 show the travel time residual maps before and after correction, respectively. Figures 6.14 and 6.15 show the amplitude residual maps before and after correction, respectively. After correction, the travel time and amplitude residual maps both now show smooth and spatially coherent variations, especially along the entire PG line and at the intersection point of the PE and PF transects.

#### *6.2.2.4 Source wavelet estimation*

The method to estimate the source wavelets has been described in section 3.4 so we here only give a brief overview and illustrate its application on the 2008/09/10 event (Figure 6.16). The data traces are filtered in the same frequency band as the one used for inversion (Fig. 6.16a), here 0.04 - 0.10 Hz. The Green's functions are computed in the smooth 1-D ak135 model (Fig. 6.16b). After water-level deconvolution (equation 3.8), we retrieve an

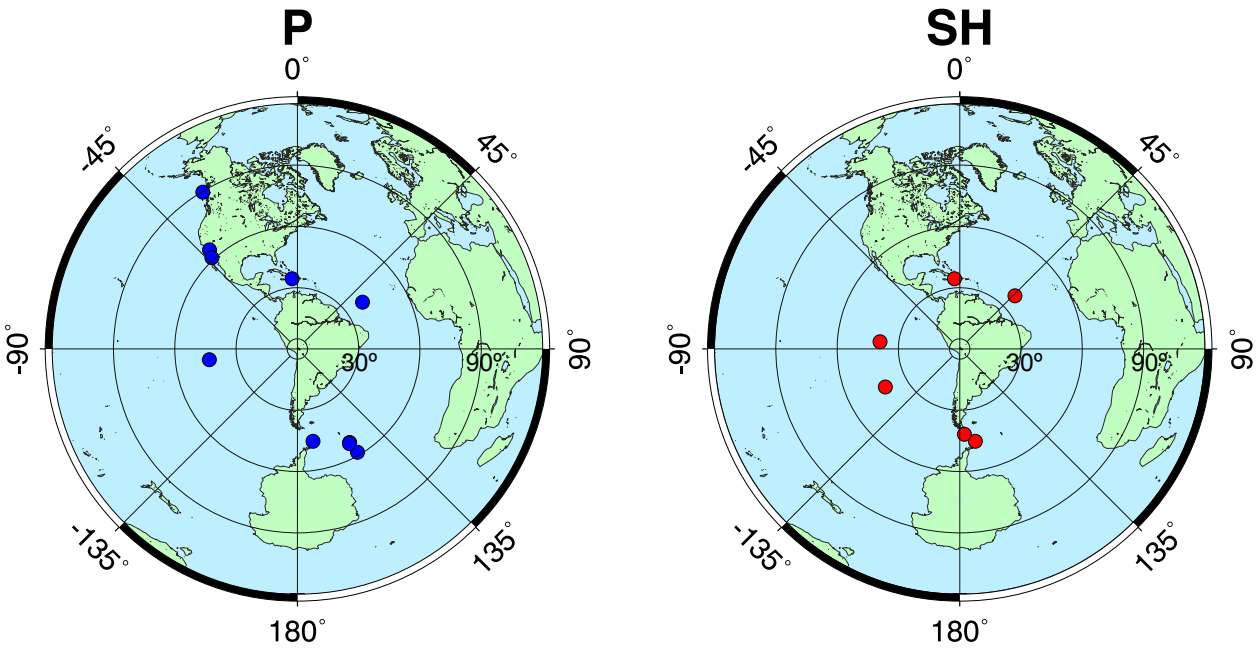


Figure 6.11: Location of teleseismic events used for waveform inversion. Detailed event informations are given in Table 6.2

Table 6.2: List of P and SH teleseismic events selected for this study.

| **Event** | **Phases** | **Origin Time** | **Location** (lat, lon) | **Depth** (km) | **Distance** (°) | **Baz** (°) | $\mathbf{M_w}$ |
|---|---|---|---|---|---|---|---|
| 1 | P | 2008/09/10 13:08:14.9 | 8.18N, 38.54W | 15 | 40 | 55 | 6.6 |
| 2 | P | 2010/01/05 04:55:38.9 | 58.5S, 14.85W | 15 | 60 | 149 | 6.8 |
| 3 | P+SH | 2010/01/12 21:53:10.2 | 18.61N, 72.62W | 12 | 34 | 357 | 7.0 |
| 4 | P | 2010/03/07 07:05:23.3 | 16.34S, 115.45W | 13 | 43 | 263 | 6.3 |
| 5 | P | 2010/04/04 22:40:43.0 | 32.31N, 115.39W | 13 | 64 | 319 | 7.3 |
| 6 | P | 2010/12/08 05:25:35.3 | 56.49S, 25.48W | 18 | 53 | 150 | 6.3 |
| 7 | P+SH | 2012/01/15 13:40:19.5 | 60.62S, 56.47W | 14 | 46 | 169 | 6.6 |
| 8 | P | 2012/01/22 05:53:42.1 | 56.8S, 24.52W | 12 | 53 | 150 | 6.1 |
| 9 | P | 2012/04/12 07:15:48.5 | 28.57N, 112.76W | 16 | 60 | 317 | 7.0 |
| 10 | P | 2013/01/05 08:58:19.3 | 55.69N, 134.97W | 14 | 89 | 329 | 7.7 |
| 11 | SH | 2010/01/17 12:00:01.1 | 57.94S, 66.16W | 19 | 42 | 176 | 6.3 |
| 12 | SH | 2011/03/01 00:53:46.3 | 29.68S, 112.26W | 20 | 40 | 243 | 6.1 |
| 13 | SH | 2011/07/27 23:00:30.3 | 10.9N, 43.34W | 15 | 38 | 48 | 6.0 |
| 14 | SH | 2011/09/23 19:02:48.2 | 9.07S, 109.66W | 16 | 38 | 275 | 6.6 |

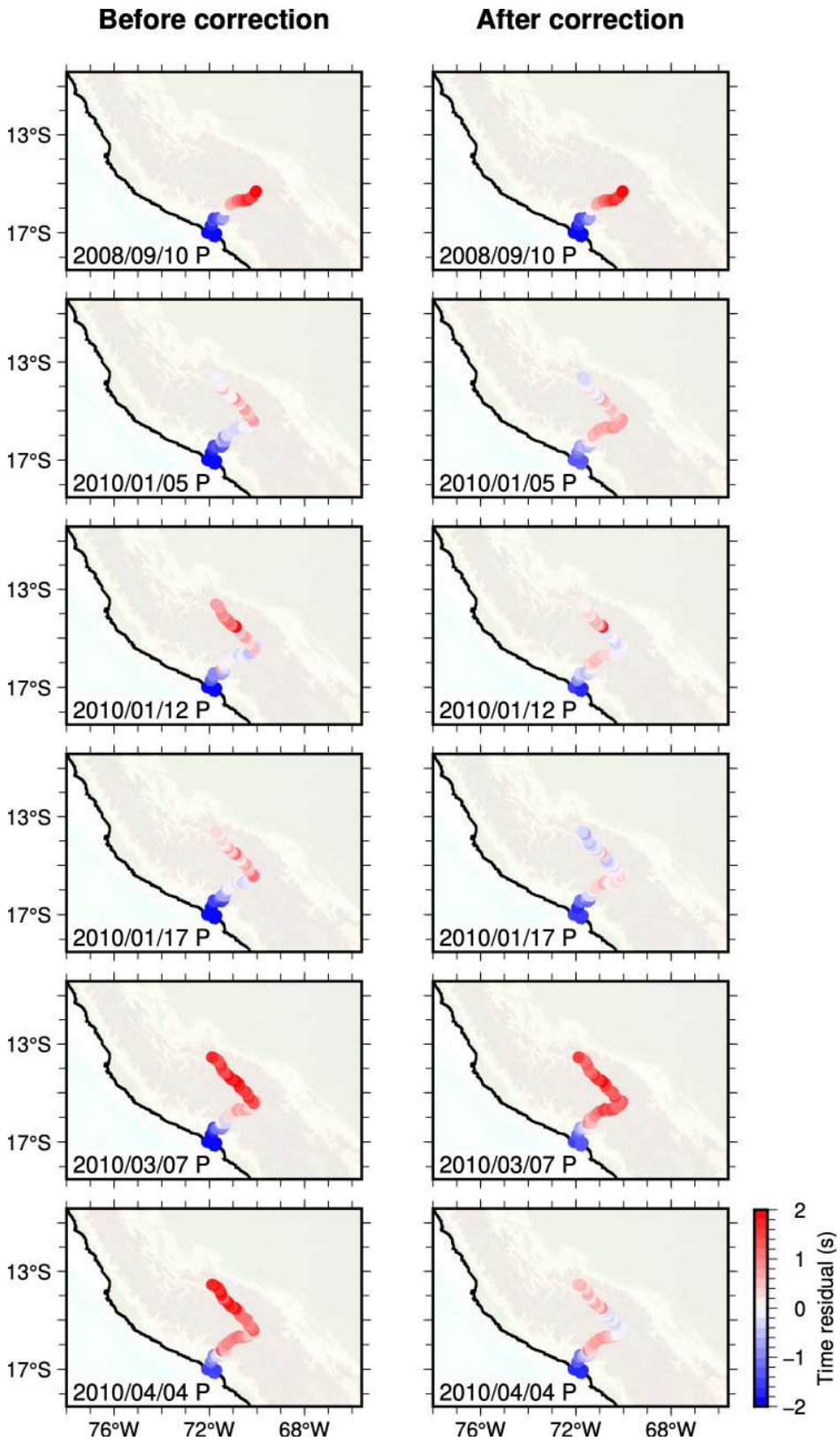


Figure 6.12: Travel time residual maps for the selected P events listed in Table 6.2. Left column: before the time corrections. Right column: after the time corrections.

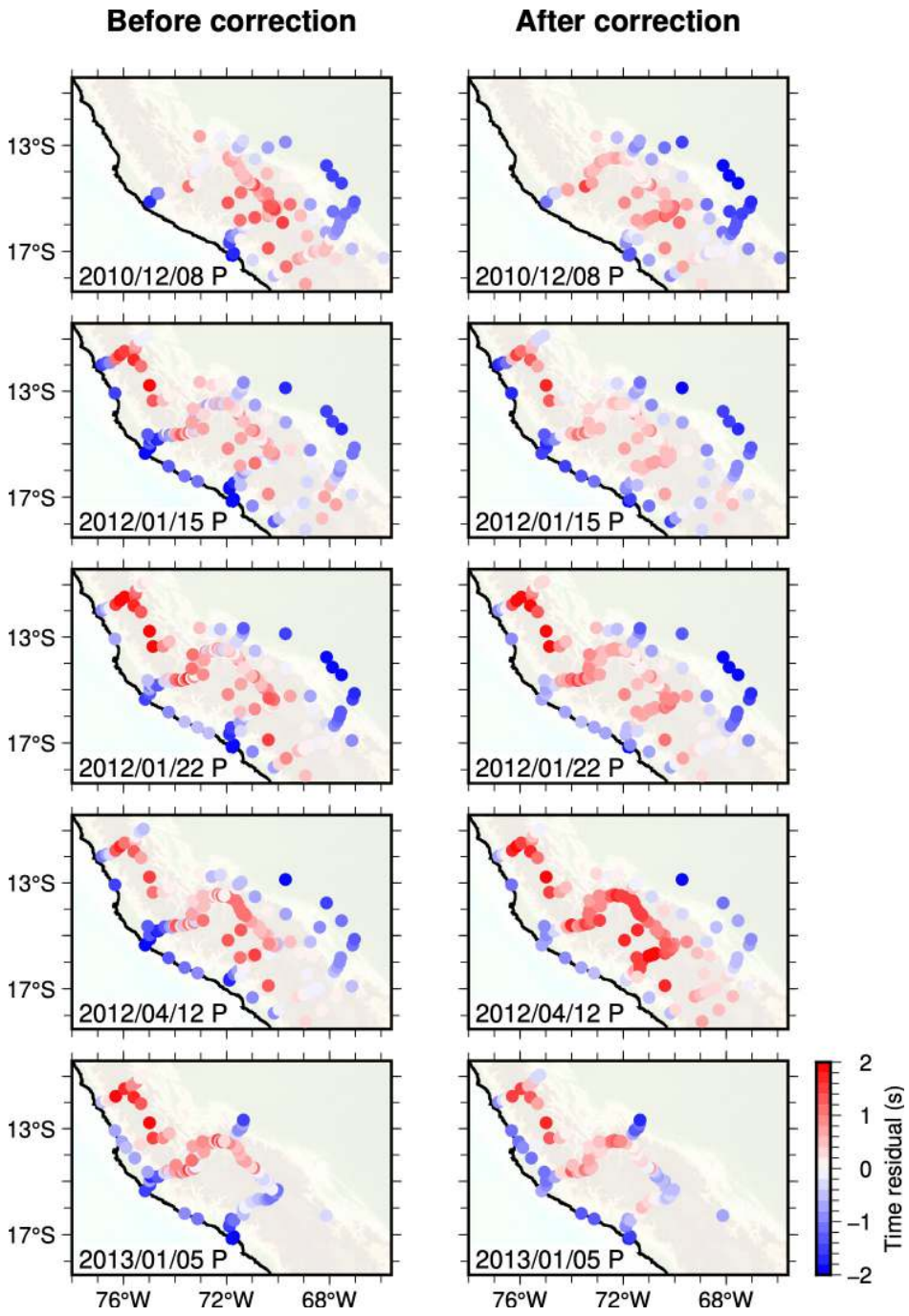


Figure 6.12: (continued)

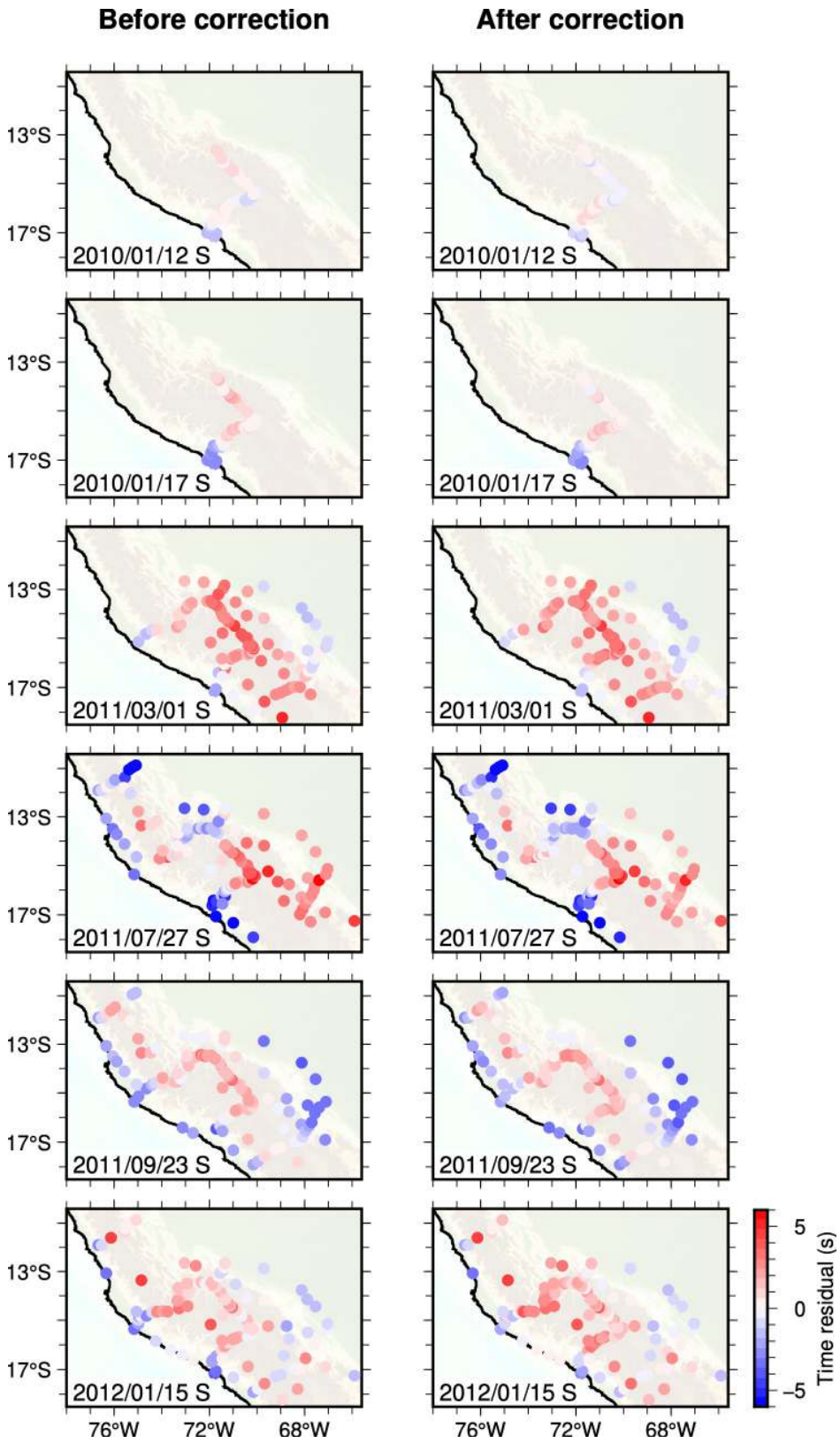


Figure 6.13: Same travel time residual maps as Fig. 6.12 but for S events.

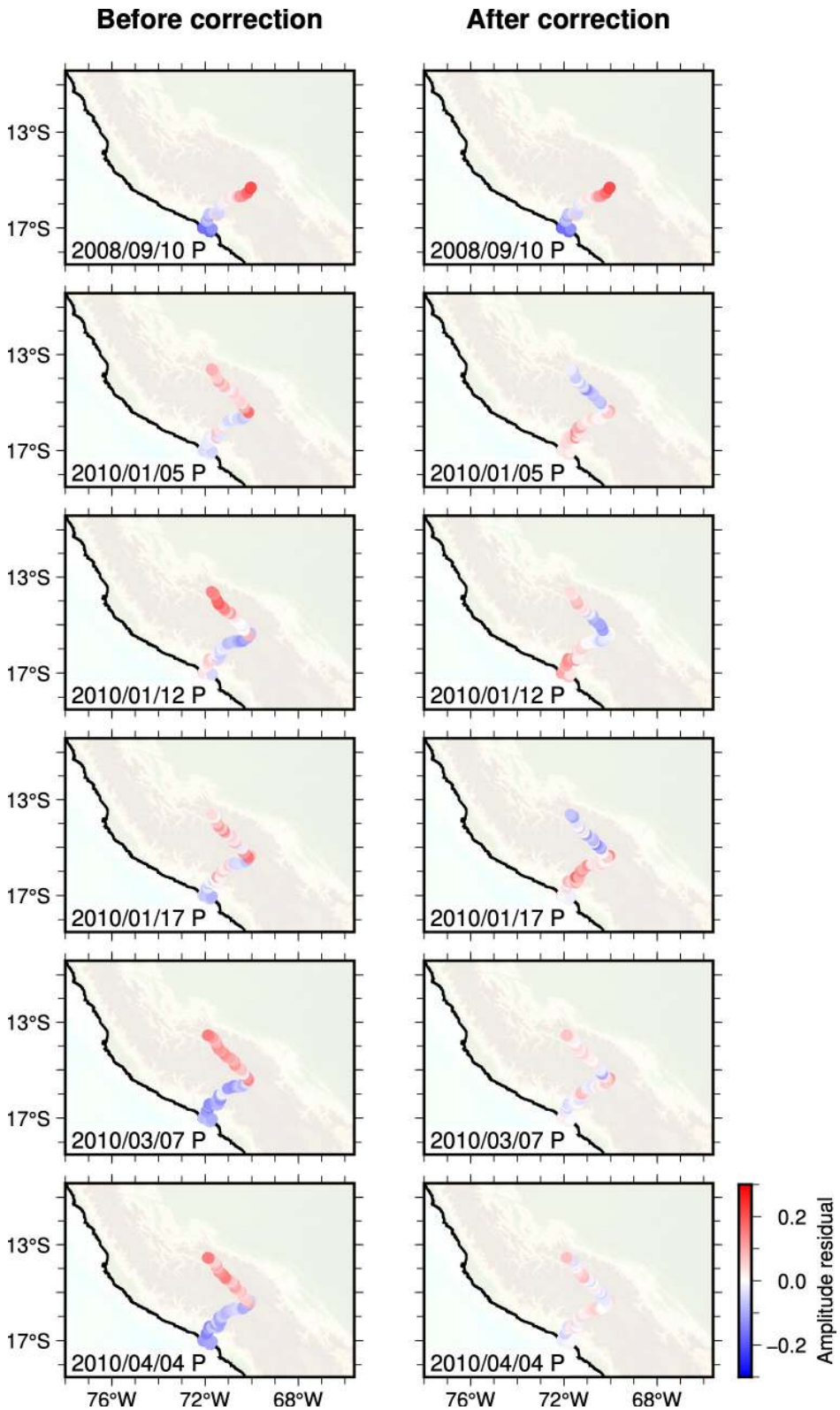


Figure 6.14: Amplitude residual maps for the selected P events listed in Table 6.2. Left column: before the amplitude corrections. Right column: after amplitude the corrections.

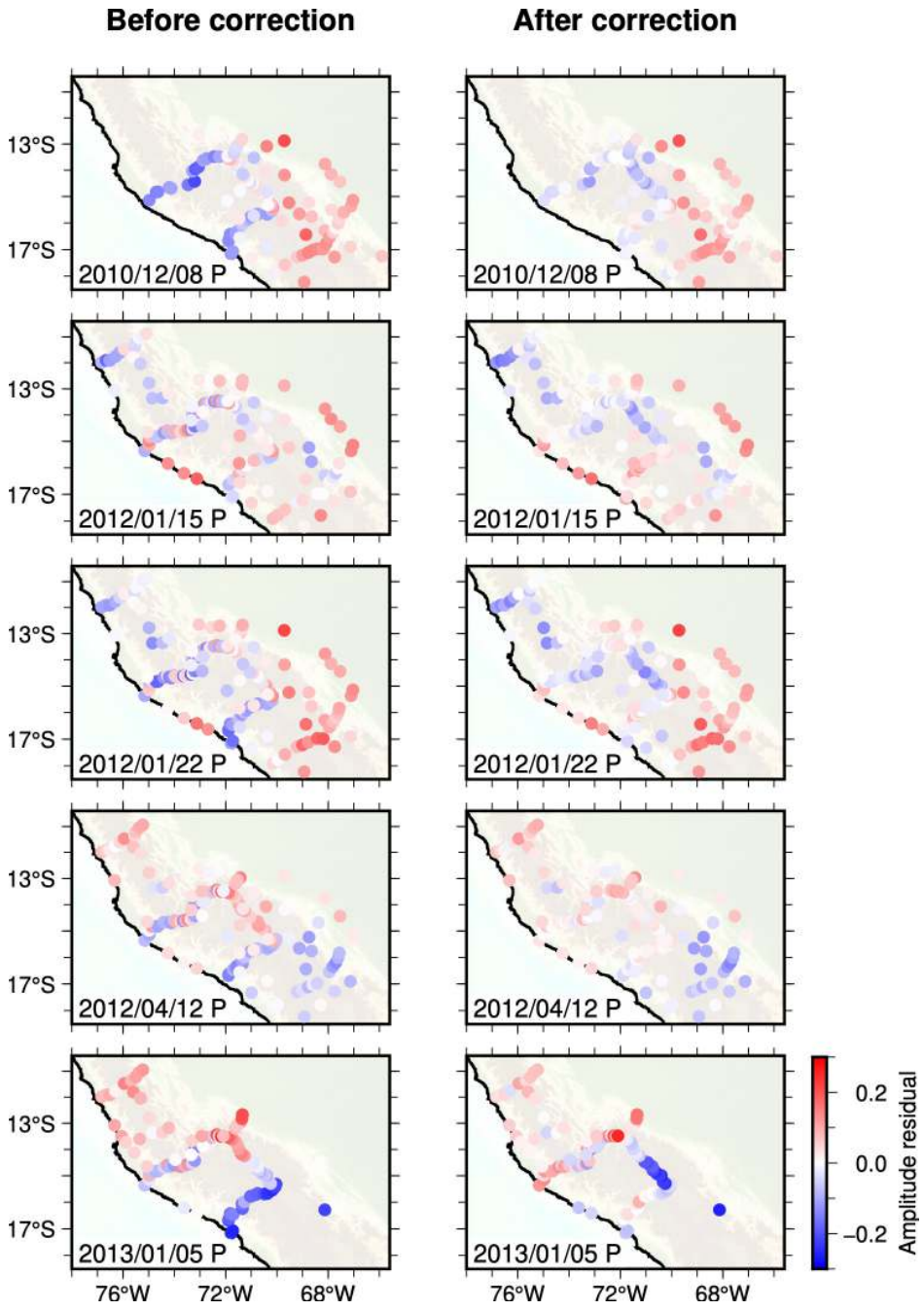


Figure 6.14: (continued)

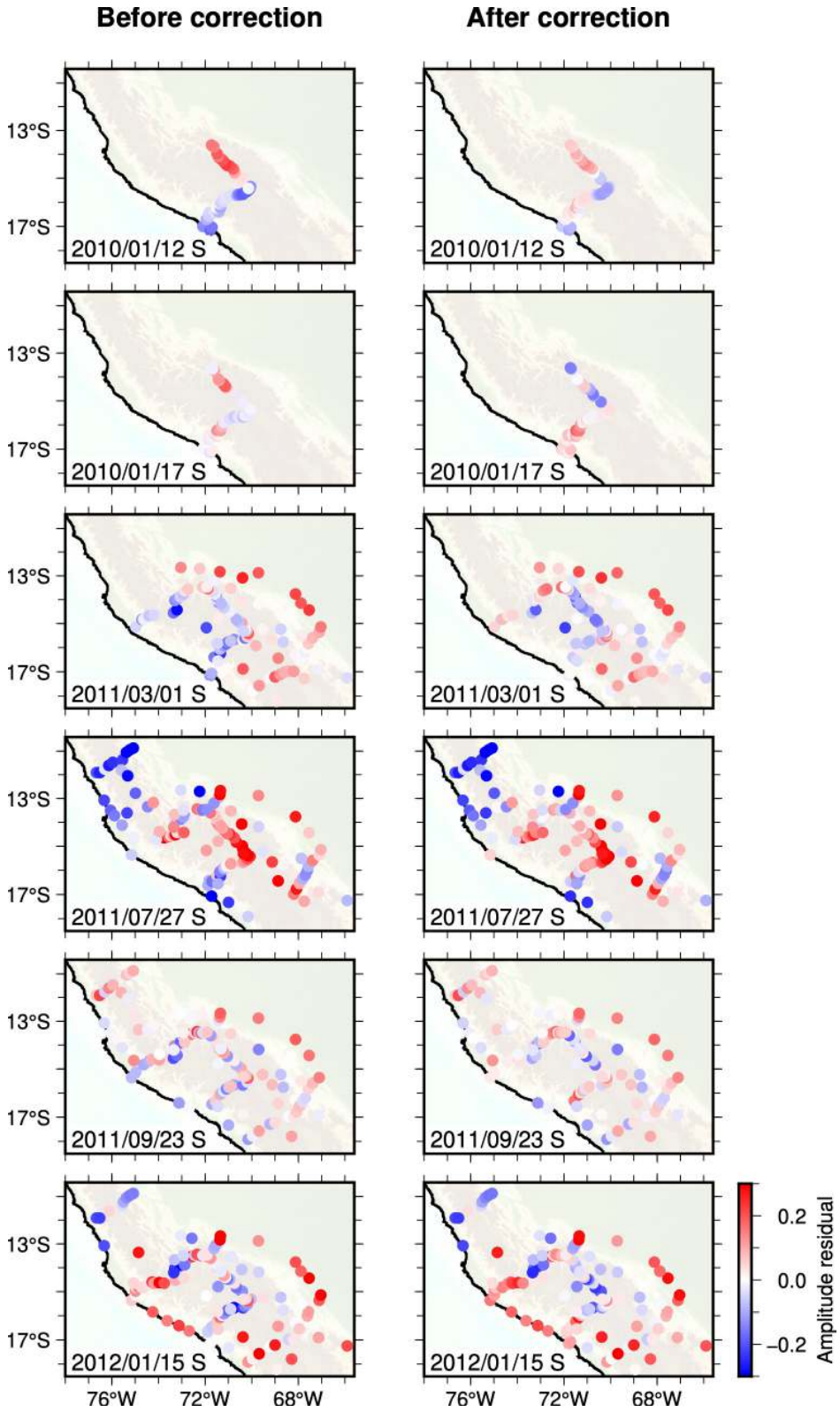


Figure 6.15: Same amplitude residual maps as Fig. 6.14 but for S events.

apparent source wavelet for each station (Fig. 6.16c). We then align and stack stacking these wavelets to obtain the average source wavelet recorded by the array (orange line in Fig. 6.16d). After aligning with equation 3.9 we get the average source wavelet (Fig. 6.16e). This average wavelet is then shifted so that the average travel time residuals is zero (Fig. 6.16f). This travel time shift is important because it removes the average travel time between the source and the regional tomographic grid, as in classical regional travel time tomography (Aki et al., 1977). Finally, we apply zero padding on both sides of the wavelet so that the average wavelet has the same length as the SEM simulation (Fig. 6.16g). As already explained in section 2.1.2, the convolution of the external wavefield and the source wavelet is performed by FFT, which implies a circular convolution. We apply to the average wavelet a time shift equal to half the length of the SEM simulation (Fig. 6.16h). The Fourier transform of this final wavelet can then be used directly to compute the convolutions with the tractions on the edges of the grid in the spectral domain.

### 6.2.3 The tomographic meshes

The big spectral-element mesh is a chunk of the spherical Earth (Fig. 6.17), designed to include all available stations in our study region (Fig. 6.1). The center of the mesh is positioned at longitude -70.8°and latitude -14°. The size of the mesh is 15°, 12°and 1000 km along the x, y, z directions, respectively. It is composed of ∼20 km hexahedral elements. Within each element, the discretization involves degree-7 Lagrange polynomials, i.e. 8 nodes along each spatial dimension. This mesh consists of 303,912 elements and 12,744,019 grid points, for a minimum resolved period of about 3.4 s.

We also designed a smaller mesh, in order to invert only the first PE profile of the PeruSE experiment. The motivation for performing full waveform inversion only on the PE profile data is that, except for a systematic time shift with stations from other deployments the time and amplitude of PE stations appear to be internally consistent. Note that any average time or amplitude offset will be absorbed by the average wavelet and its timing, which imposes a zero average time shift between the observed and synthetic seismograms. The

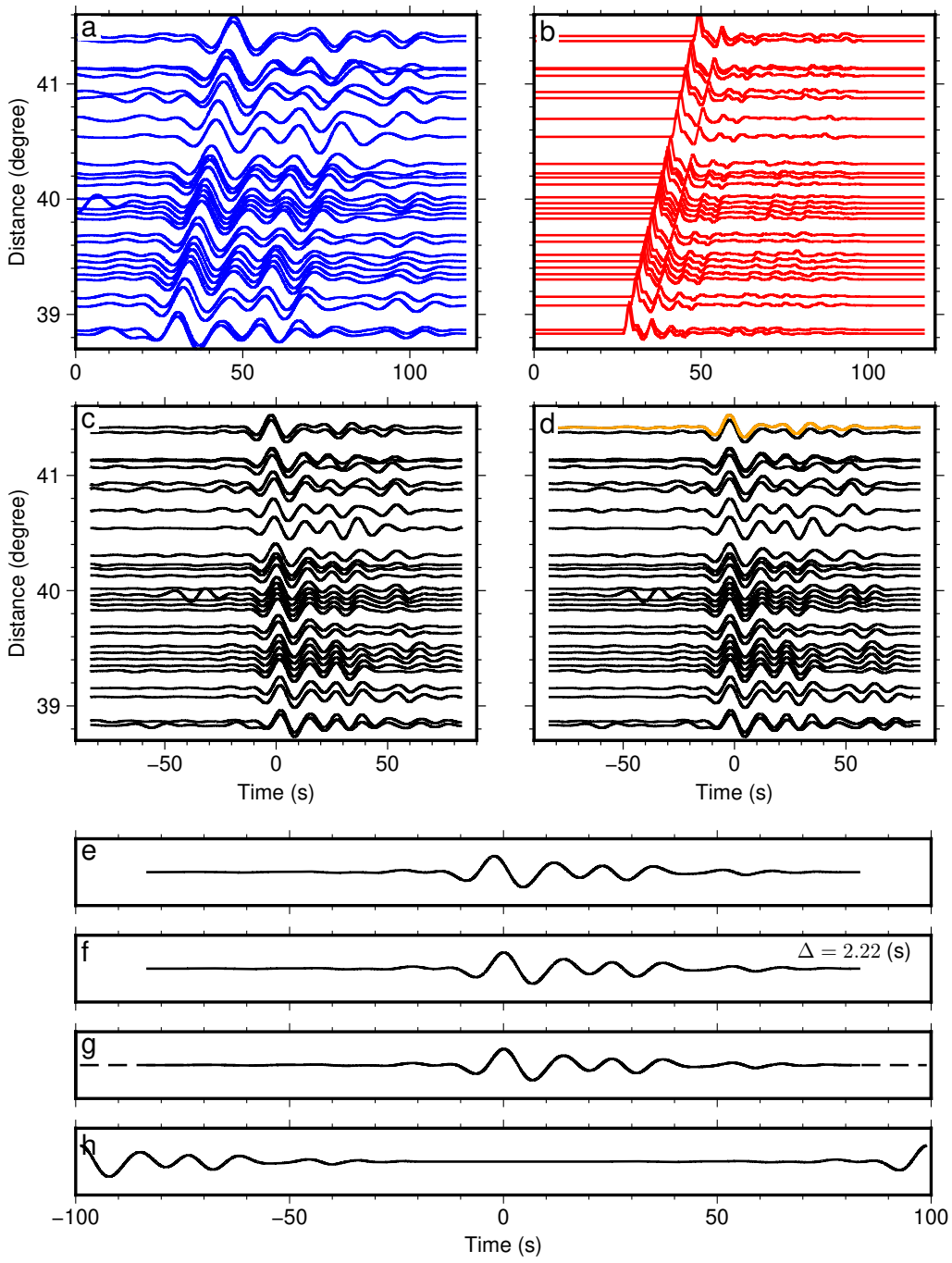


Figure 6.16: Illustration of the average source wavelet estimation procedure. (a) Vertical component P waveforms for the 2008/09/10 event filtered between 0.04 and 0.10 Hz. (b) Green's function. (c) Apparent source wavelets obtained by deconvolving the Green's functions in (b) from the observed traces in (a). (d) Aligned wavelets, using the first trace as a reference (orange line). (e) The average of the aligned wavelets is shown in (d). (f) Time shift to the zero mean delay, with the time of 2.22s. (g) Padding the wavelet with zeros to the length of the SEM simulation. (h) Final wavelet for the circular convolution is performed on the edge of the box.

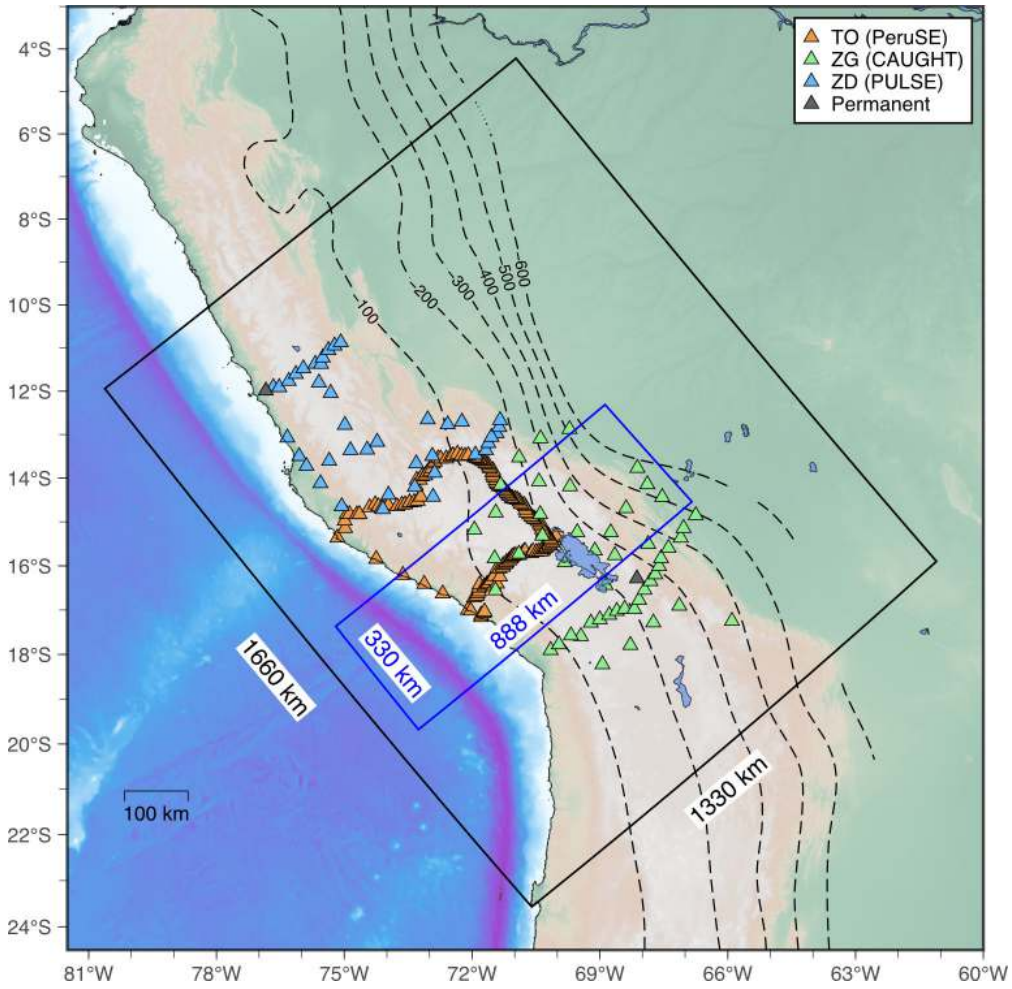


Figure 6.17: Map of the big mesh (black) centered at longitude -70.8 and latitude -14.0 with the size of 1660 km by 1330 km; the small mesh (blue) centered at longitude -71.0 and latitude -16.0 with the size of 330 km by 888 km. The azimuth of both meshes is 50°clockwise from north.

small mesh is centered on longitude -70°and latitude -16°, with a size of 3°, 8°and 600 km in x, y, z direction respectively. The size of the hexahedral elements is about 12 km. This grid consists of 109,200 elements and 9,480,073 grid points and the minimum period that can be resolved is about 2.1 s.

Both meshes are rotated by 50 degrees so that the $y$ axis approximately corresponds to the strike of the PE and PG profiles. The topography extracted from the ETOPO1 model (Amante, 2009) is imposed on the free surface of both meshes.

### 6.2.4 The full waveform inversion method

We invert the vertical and radial components of the P waveforms as well as the transverse component of the SH waveforms in our data selection (see Table 6.2). The time windows used in the inversion start 10 s before the main phase arrival (either P or SH) and end 40 to 70 s after, depending on the event. The end of the time windows is selected before the arrival of secondary phases (e.g. PP, PcP, etc...). The selected time windows are sufficiently long to include conversions and reverberations from the main lithospheric discontinuities, such as the Moho, in addition to the main direct transmitted phase.

The inverse problem consists in finding the model $\mathbf{m}$ that minimizes the cost function (Tarantola, 2005)

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\text{obs}})^t \mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\text{obs}}) + \frac{\lambda}{2}(\mathbf{m} - \mathbf{m}_{\text{prior}})^t \mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}), \qquad (6.2)$$

where $\mathbf{g}$ is the forward wave equation operator, $\mathbf{C_D}$ the data covariance matrix, $\lambda$ the regularization parameter, and $\mathbf{C_M}$ the a priori model covariance matrix. The first term in the cost function 6.2 quantifies the waveform misfits between the data and the synthetics whereas the second term quantifies the deviation from the a priori model. The regularization parameter $\lambda$ determined is introduced to balance the contributions of the first and the second terms to the cost function (Kan et al., 2022).

We follow the FWI workflow described in Chapter 3. The source wavelet for each event is estimated by a simple water-level deconvolution. The Green's function for deconvolution is computed in the input model. The forward problem is solved with the hybrid method described in Monteiller et al. (2021) that couples AxiSEM (Nissen-Meyer et al., 2014) with a regional SEM (Komatitsch and Tromp, 1999). The minimum of the cost function 6.2 is searched with an iterative algorithm, in which the model $\mathbf{m}_k$ at iteration $k$ is perturbed by

$$\Delta\mathbf{m}_k = -\alpha_k \mathbf{H}_k^{-1} \cdot \gamma_k \qquad (6.3)$$

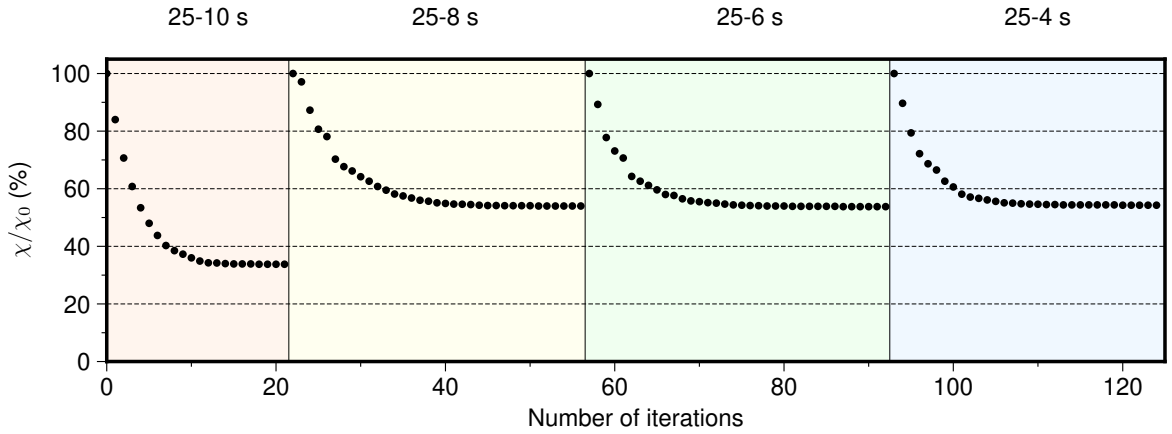


Figure 6.18: Evolution of cost reductions of the four hierarchical inversions with progressively reduced lower corner period of 10 s, 8 s, 6 s, and 4 s. The misfits are normalized to 100% when we start the new inversion in a shorter period.

where $\alpha$ is the step length, $\mathbf{H}^{-1}$ the inverse Hessian, and $\gamma_k$ the gradient of the cost function at iteration $k$. The gradients of the cost function are computed with the adjoint method (Tromp et al., 2005) and are preconditioned with the approximated Hessian computed by the l-BFGS method (Nocedal and Wright, 2006). Event normalization is applied to balance the contribution of each event to the gradient. The selected step length $\alpha$ is defined according to the Wolfe line search criteria (Nocedal and Wright, 2006). The convergence of the inversion is defined when the cost reduction between two successive iterations is too small (say $<$ 0.1%) or the number of line searches at the same iteration exceeds 5.

For the small grid inversions, we performed hierarchical inversions starting at long-period with data filtered between 25 s and 10 s. After convergence of the first long period inversion, we start a new inversion initiated with the model obtained after the previous inversion, with data filtered with a lower corner period that is progressively decreased from 10 s to 8 s, 6 s, and finally 4 s. Fig. 6.18 shows the normalized cost functions after the four stages of the hierarchical inversion. The waveform misfit is reduced by about 60% after the first long-period inversion and about 50% after the shorter period inversions.

Table 6.3: Hyperparameters used in the small-grid inversions.

| Periods | $\sigma_{V_P}$ (km/s) | $\sigma_{V_S}$ (km/s) | $\sigma_\rho$ (g/cm$^3$) | $r$ | $\ell$ (km) |
|---|---|---|---|---|---|
| 25-10, 8 s | 0.65 | 0.37 | 0.27 | 0.97 | 15 |
| 25-6, 4 s | 0.65 | 0.37 | 0.27 | 0.80 | 8 |

#### *6.2.4.1 Hyperparameters for the inversion*

The complete model covariance matrix in (6.2) is written of the form (Kan et al., 2022)

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2\mathbf{C} & r\sigma_\rho\sigma_{V_P}\mathbf{C} & r\sigma_\rho\sigma_{V_S}\mathbf{C} \\ r\sigma_\rho\sigma_{V_P}\mathbf{C} & \sigma_{V_P}^2\mathbf{C} & r\sigma_{V_P}\sigma_{V_S}\mathbf{C} \\ r\sigma_\rho\sigma_{V_S}\mathbf{C} & r\sigma_{V_P}\sigma_{V_S}\mathbf{C} & \sigma_{V_S}^2\mathbf{C} \end{bmatrix}, \tag{6.4}$$

where $r$ is the correlation coefficient between each pair of parameters, $\mathbf{C}$ the spatial covariance kernel, $\sigma_\rho$, $\sigma_{V_P}$ and $\sigma_{V_S}$ the standard deviation of density, $V_P$ and $V_S$, respectively. We use a smoothing length of 15 km in the 10 s and 8 s inversions, reduced to 8 km in the 6 s and 4 s inversion. These moderate values of the smoothing length are chosen in order to preserve the sensitivity to small-scale structures, as discussed in Kan et al. (2022). In long-period inversions, we impose large correlation coefficients ($r = 0.97$) and reduce them to 0.80 in the final short-period inversion at 4 s. The value of the hyperparameters used in the hierarchical inversions is listed in Table 6.3.

#### *6.2.4.2 Model parameterization*

As discussed in Kan et al. (2022), with a complete model covariance matrix the choice of model parameterization becomes less critical. Here we use the more intuitive parameterization in terms ($\rho$, $V_P$, $V_S$) which presents the advantage that the correlations between model parameters are well-documented, for example, the density-$V_P$ relationship (Christensen and Mooney, 1995) or the $V_P/V_S$ ratio (Christensen, 1996) and the a priori values of the standard deviations of each type of parameters are easy to define.

#### *6.2.4.3 Depth variation of variance of model parameters*

In previous application of FWI (Wang et al., 2016; Beller et al., 2018) (see also the Chapter on Cascadia), the hyperparameters of the inversions, such as the standard deviation $\sigma$ on model parameters or the correlation length $\ell$ of exponential kernels, were assumed constant, i.e. with no depth variation. However, as seen in global velocity models (e.g. Lei et al., 2020), the model perturbations with respect to the reference model are larger in particular the upper 200 km and less perturbed at the deep mantle. The difference of perturbations at 100 km and 300 km can reach 2 to 3 times, which suggests that a fixed $\sigma$ as the a priori information is not the optimal assumption and may also bias the results. In terms of a priori knowledge, the $\sigma$ would be larger above the upper 200 km depth. With the complete model covariance matrix introduced in Kan et al. (2022), the high flexibility of the covariance matrix enables us to investigate the effect of depth variation of $\sigma$.

We use a sigmoid function that scales the $\sigma$ with respect to the depth

$$\sigma(z) = \sigma \cdot \left( \frac{1}{3} \cdot \frac{1}{1 + e^{-a(z-b)}} + \frac{2}{3} \right), \tag{6.5}$$

where $a$ the scaling parameter and the shift $b$ are set to -0.03 and 200 respectively (the unit of depth is in km). Fig. 6.19 shows the variation of $\sigma$ as a function of depth with a uniform and the sigmoid function. Compared with a uniform function that is fixed for all the depths, the designed sigmoid function starts to decrease at 100 km depth and progressively reduces to reach $0.66\sigma$ beneath 300 km. We apply this depth variations in the small-grid inversion.

#### *6.2.4.4 Computational resource*

For the computations, the small mesh is decomposed into 4 domains, whereas the big mesh is decomposed into 8 domains. Independent forward modeling and gradient computations are performed for each event. These computations thus require 4 GPUs per event for the small mesh, and 8 GPUs for the big mesh. For the small mesh, the time step is 0.03 s, and the number of time steps is set to 6600 and 9000 for P and S events, respectively, so

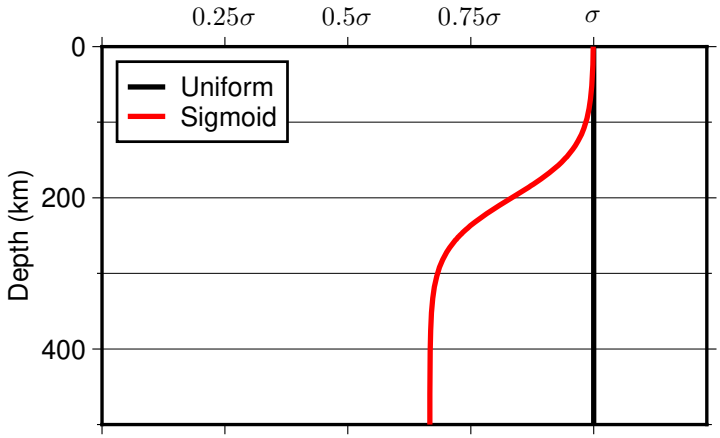


Figure 6.19: Standard deviation of model parameters as the function of depth with a uniform function (black line) and a sigmoid function (red line).

that the simulation lengths are 198 s and 270 s for P and S events, respectively. For the big mesh, we need longer time windows before the main phase arrival (P or S) to ensure that the phase has not reached the regional box at the beginning of the wavefield injection. The number of time steps is set to 10000 and 14000 for P and S events, respectively. The time interval is still set to 0.03 s, and thus the simulation lengths are 300 s and 420 s for P and S events, respectively.

In practice, on CALMIP, it takes about 2.5 and 3 minutes to compute the gradient for a P and S event, respectively. On that same machine, it takes about 5 and 7 minutes to compute the gradient for a P and S event, respectively. The computations of all the gradients in the small mesh using only 4 GPUs take about 40 minutes, while the computation in the big mesh using 8 GPUs takes about 70 minutes. This is the time required to perform one iteration for all the periods. An inversion requires typically about 25 iterations to converge. Note that the time required to compute all the gradients can be reduced if we use more GPU since the different events can be computed in parallel.

The CPU time required to perform one forward and adjoint computation in both meshes is listed in 6.4, taking into account the mesh size, mesh decomposition, and simulation length.

Table 6.4: Approximate CPU hours required to perform forward and adjoint computations in small and big mesh on CALMIP supercomputer.

| | Small mesh | | Big mesh | |
|---|---|---|---|---|
| Phase | P | S | P | S |
| Forward | 2 | 3 | 8 | 13 |
| Adjoint | 3 | 4 | 13 | 18 |

## 6.3 FWI of PE profile data in the small mesh

Since we identified data phase and amplitude inconsistencies between the different PeruSE transects, we first focused on the data of the PE transect, which showed internal consistency in both phase and amplitude.

### 6.3.1 The model

The final density, $V_P$ and $V_S$ models obtained after hierarchical full waveform inversion of the P and SH waveforms recorded by the PE profile are shown in Fig. 6.20, both in absolute and relative scales. The most prominent feature is the NE-dipping positive anomaly in both seismic velocities and density, which roughly follows the contour of the subducting Nazca slab extracted from the slab2.0 model (Hayes et al., 2018). Three distinct morphological segments can be identified in the 0-80 km, 80-200 km, and 200-500 km depth ranges. The dipping angle of the first two segments is $\sim 30$ degrees but the second segment shows significant irregularity of the subduction interface, whereas that of the first segment appears much more regular and smooth. Below 200 km depth, the subducting slab dips almost vertically. This change in the dynamics of subduction is probably accompanied by significant deformations, which is in good agreement with the concentration of seismicity observed at the level of the bend made by the subducting plate.

We will now focus on the shallow part of the subduction, which is shown in Fig. 6.21. The first segment of the subducting Nazca plate is characterized by a thin low-velocity anomaly $\sim 10$ km thick which lies on the contour of the top of the Nazca plate. Seismic velocities in this thin layer are comprised between 6.0 and 7.0 km/s for $V_P$, and between 3.5 and 4.3 for

$V_S$, while the density is between 2.7 and 3.0. These values are well below those of typical mantle rocks and are probably related to the oceanic crust of the Nazca plate. Since the minimum period in our inversions is 4 s, the thickness of the oceanic crust is smaller than the shear wave wavelength, and the reduction in seismic velocities, as well as density in this layer, is probably underestimated. In the second segment, the thin low-velocity anomaly has disappeared, and the top of the high-velocity and density anomaly has an irregular shape. The irregular surface of this fast anomaly could correspond to the subduction interface, which would thus result from the subduction of a thickened and now eclogitized crust or more simply from a small-scale undulation of the subducting plate geometry. In either case, these hypotheses could explain the concentration of seismicity localized in the shallow part of the downgoing plate. An alternative would be that the subduction interface is regular and the high velocities located above the slab correspond to a cold mantle wedge. However, this would imply that the seismicity is located in the mantle wedge, which is unlikely because the mantle wedge is composed of hot and ductile material that is less prone to brittle failure and sudden release of energy that causes earthquakes.

Seismic velocities ($V_P$ between 7.6 and 8.0 km/s, and $V_S$ between 4.2 and 4.4 km/s) in the forearc mantle wedge are slightly lower than in a normal mantle, which may result from a moderate degree of serpentinization, on the order of 10 to 20 % (Christensen, 2004). Seismic velocities are most strongly reduced within a near-vertical corridor located at distance 340 km, which extends vertically from the top of the slab to the base of the forearc crust. This region is also characterized by a high $V_P/V_S$ ratio of about 1.86, suggesting a locally stronger degree of serpentinization, which may be as high as 30 %. East of the volcanic arc, a prominent large low-velocity anomaly is observed between the Altiplano crust and the Nazca plate, with a reduction of $V_P$ and $V_S$ of the order of 6 % and up 8 % beneath the central part of the Altiplano. This low-velocity anomaly extends beneath the Eastern Cordillera (EC) and subandean region, where it can be tracked to 300 km depth. However, owing to the coarser distribution of seismic stations NE of the PE profile, the depth resolution in that

part of the model is likely degraded and probably suffers from vertical smearing.

In the upper plate, the continental crust is relatively thin, with a thickness between 20 and 30 km. This domain is located west of the volcanic arc and therefore corresponds to the forearc crust. The transition between the forearc and the volcanic arc corresponds to a sharp increase of crustal thickness which is about 70 km beneath the volcanic arc. The compressional velocities beneath the volcanic crust are between 6.4 and 6.8 km/s. We thus find no evidence for a mafic lower crust or for eclogitization of the lower crust. Beneath the Altiplano, the structure of the crust appears stratified, with an upper and a lower layer with very contrasted signatures. A sharp transition is observed between these two layers, at about 35 km depth, with very slow compressional velocities between 5.6 and 6.2 km/s in the upper layer, which increase to 6.8-7.8 km/s in the lower layer. A similar stratified crustal structure has been reported in the southern Central Andes (Graeber and Asch, 1999; Yuan et al., 2000).

### 6.3.2 Waveform fits

We show examples of P and SH waveform fits for two events with different back-azimuth, before and after hierarchical inversion (Fig. 6.22 and 6.23). The waveform fits of the other events are shown in Appendix 6.A. We compare the observed (black lines) and synthetic (red lines) seismograms computed in the initial smooth 1-D model and in the final 3-D model. The synthetics computed in the final model provide better fits in phase and amplitude, especially for the radial component. Fig. 6.22 shows the waveform fit of the 2010/01/12 event (coming from North) when both P and S waves are included in the inversion. The cost reduction for this event is about 80% for P waves and 60% for S waves. The improvement can be observed not only for the energetic P and SH waves but also for the later arriving scattered waves. The poorer fit of SH waves may result from seismic anisotropy, which is not yet taken into account.

In general, the fit of vertical component P waves before inversion is much better than that of the radial component. This results from the fact that the vertical component of

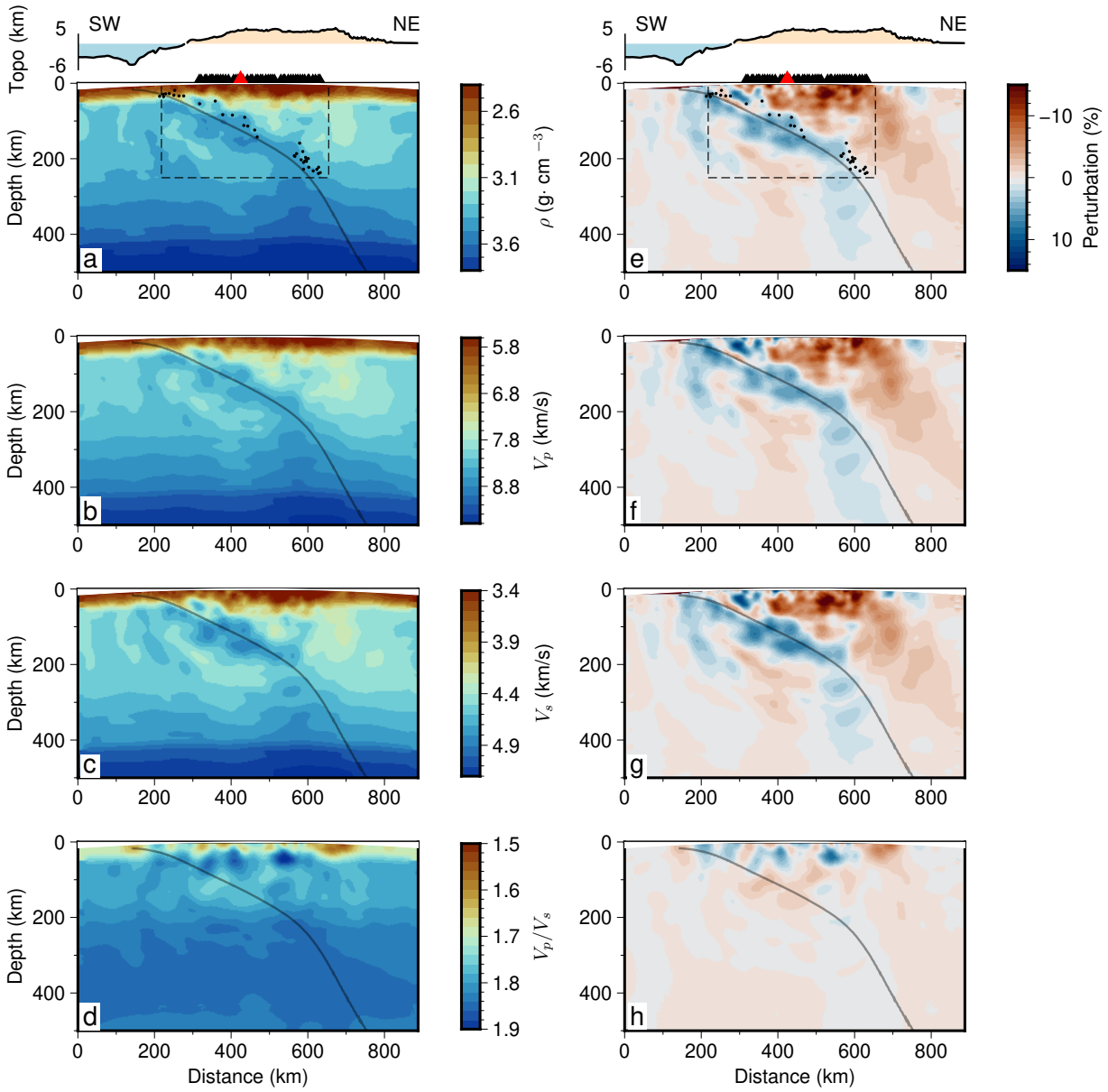


Figure 6.20: Results of the hierarchical full waveform inversion. (a-d) Models of absolute density, $V_P$, $V_S$, and $V_S/V_S$ respectively. (e-f) Perturbations of density, $V_P$, $V_S$, and $V_S/V_S$ with respect to the initial 1-D smooth AK135 model. The black and red triangles denote the stations and the volcano along the slice, respectively. The black line shows the Slab2 Nazca contour. The black dots in (a) and (e) denote local earthquakes from the ISC-EHB catalog (1980-2020). The dashed lines mark the region of the zoomed plots in Fig. 6.21.

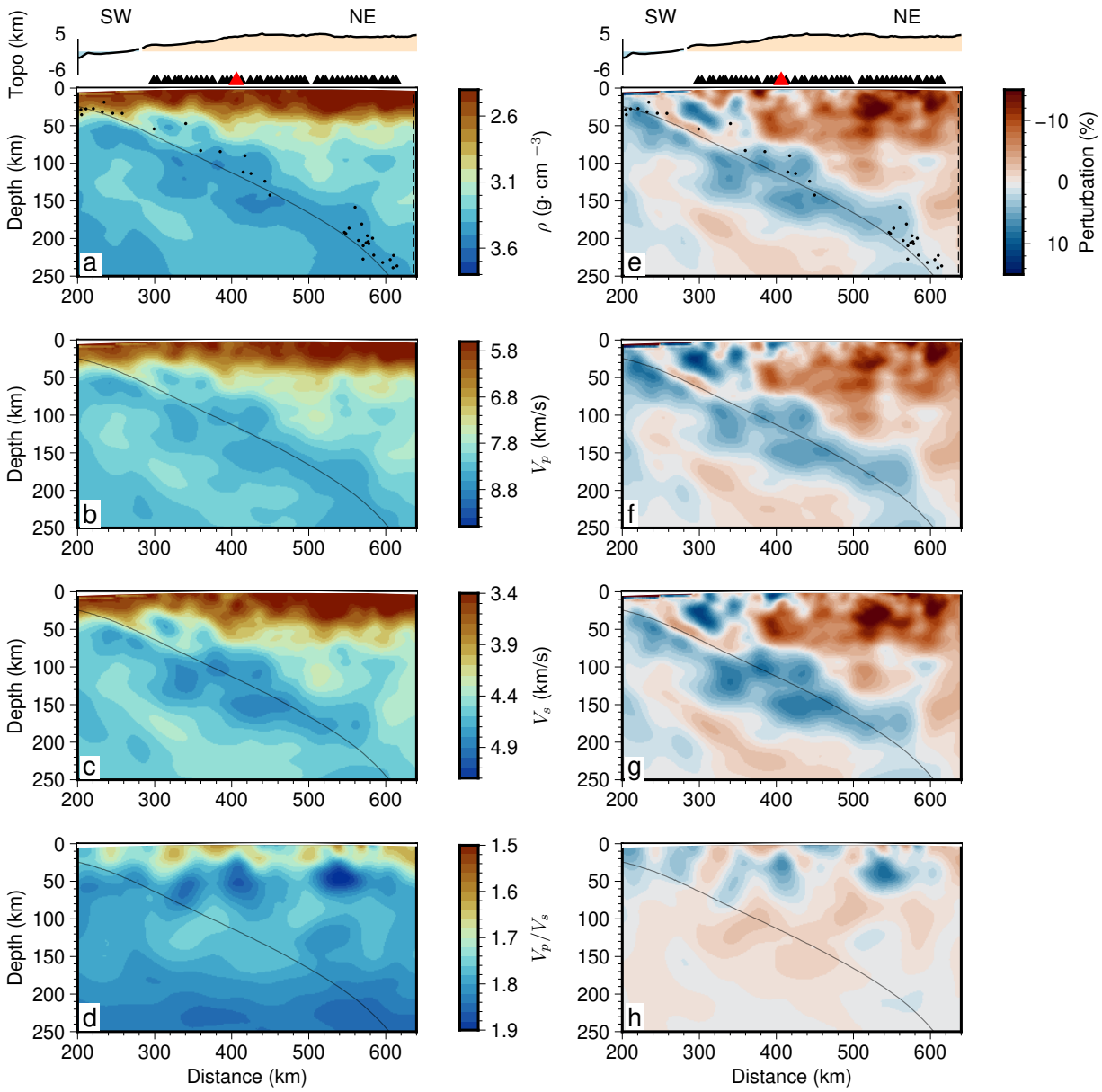


Figure 6.21: Same for Figure 6.20 but zoomed to the top 250 km of the model.

teleseismic P-waves is dominated by the direct P-wave, which is highly coherent on a regional scale. The radial component, on the other hand, contains secondary arrivals in the coda of the P-wave that have comparable amplitude to the direct P-wave, which are not honored in the initial smooth model which explains why the initial fit of the synthetic radial component seismograms is poorer. For S-waves, we use the transverse component because it is dominated by the direct SH wave. The inversion windows for SH waves are generally shorter than for P waves to avoid including secondary arrivals, e.g., ScS or SKS. Nevertheless, while P-to-S conversions observed on the radial component provide an indirect constraint on $V_S$ from P waveform fits, additional constraints on the $V_S$ model come from the phase of direct SH waves.

Figure 6.23 shows another waveform fit of the 2010/03/07 event (only P) from the west of the PE transect. The waveforms are filtered between 0.04 and 0.15 Hz due to the low signal-to-noise ratio of the radial component when filtered between 0.04 and 0.25 Hz. Near the southwestern part of the transect, we can clearly observe coherent secondary phases after the P phase on the radial component, which is most likely related to the LVZ seen in the inverted image (Fig. 6.20). The cost reduction for this event reaches about 85%.

## 6.4 FWI of in the big mesh with all the permanent and temporary stations installed in southern Peru

We also attempted to perform a hierarchical FWI on the complete Peru dataset, after applying the phase and amplitude correction described above. This inversion is much more computationally demanding and we were not able to finalize it before the end of this thesis. We will therefore only present here temporary results, obtained by inverting the P and SH waveforms down to a period of 10 s. The hyperparameters used are the same as the small grid inversions (Table 6.3). This work will be finalized and submitted for publication in the coming months.

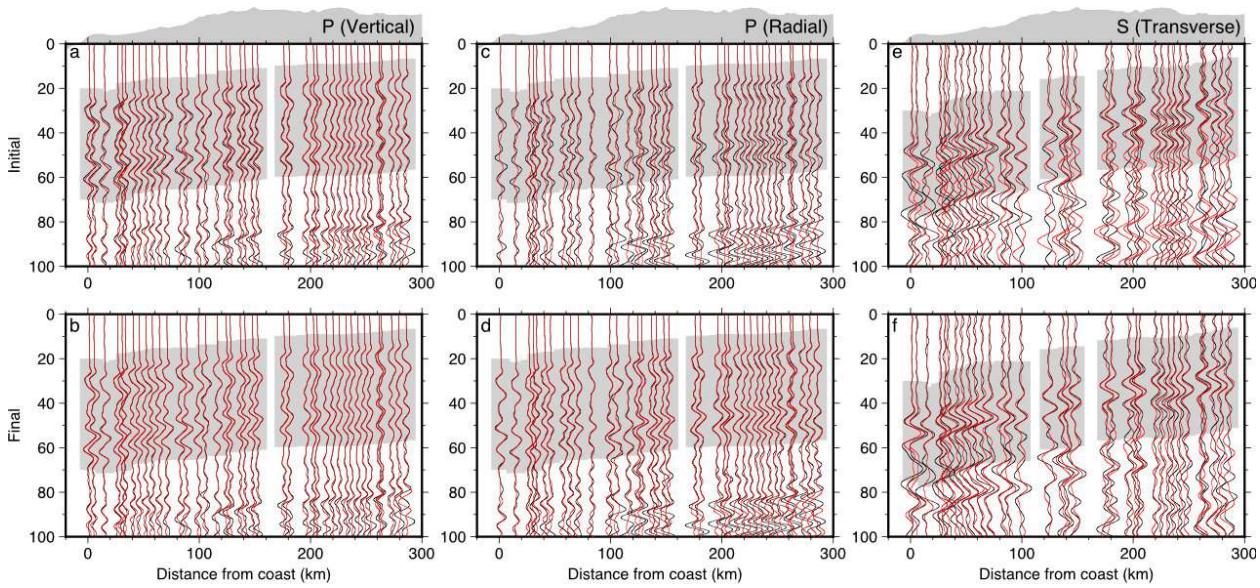


Figure 6.22: Waveform fits of the PE transect for the 2010/01/12 event filtered between 0.04 to 0.25 Hz. Observed (black lines) and synthetic (red lines) vertical component P waveforms computed in the initial (a) and final (b) models. (c,d) Same as (a,b) but for radial component P waveforms. (e,f) Same as (a,b) but for transverse component S waveforms. The grey backgrounds indicate the beginning and end of the time windows used in the inversion.

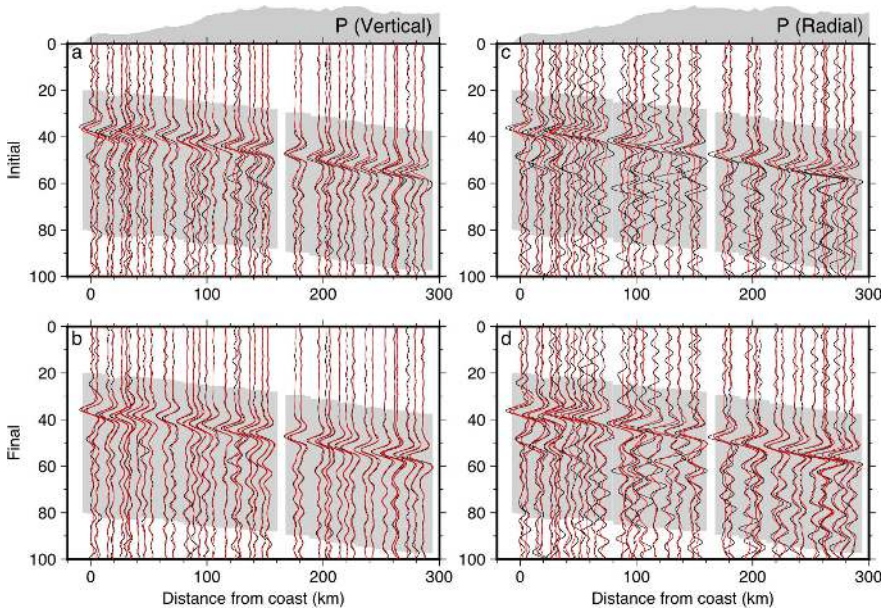


Figure 6.23: Waveform fits of the PE transect for the 2010/03/07 event filtered between 0.04 to 0.15 Hz. Observed (black lines) and synthetic (red lines) vertical component P waveforms computed in the initial (a) and final (b) models. (c,d) Same as (a,b) but for radial component P waveforms. (e,f) Same as (a,b) but for transverse component S waveforms. The grey backgrounds indicate the beginning and end of the time windows used in the inversion.

### 6.4.1 The model

We present the preliminary inversion results obtained from waveform data filtered between 25 and 10 s for density, $V_P$ and $V_S$ along the PE, PF and PG transects in Fig. 6.24, 6.25 and 6.26 respectively.

The cross-section along the PE profile (Fig. 6.24) is in good agreement with the model obtained in the small mesh inversion, but with a coarser spatial resolution, owing to the lower frequency content of seismic waveforms. This good agreement suggests that the travel time and amplitude corrections applied on the complete dataset are adequate and thus that the main features in this new model are robust.

The PF profile, which connects the PE and PG profiles and samples the transition from normal to flat subduction, clearly shows the thick crustal roots beneath the Altiplano (Fig. 6.25). We find no evidence of lateral variations in crustal thickness along this transect. In contrast, the geometry of the Nazca plate is poorly resolved. We hope that future inversions performed at shorter periods will provide additional structural details about the transition between the two subduction domains, and in particular, reveal whether it is continuous or whether there is a slab tear.

The section along the PG transect (Fig. 6.26) samples the flat subduction beneath central Peru. The shallow part of the same section is shown in Fig. 6.27. The geometry of the subducting Nazca plate is well reconstructed and in fairly good agreement with the slab2 model (Hayes et al., 2018) with the exception of the deepest segment, which begins to dip almost vertically around distance 600 km. Due to the low-frequency content of the inverted seismic waveforms, the resolution is not sufficient to document the possible presence of subducting oceanic crust. The flat segment shows a long-wavelength bulge of its surface, also highlighted by the distribution of seismicity. Interestingly, the Moho of the upper plate, as imaged from receiver functions in Ma and Clayton (2015), follows a similar curved geometry. These observations indeed suggest a strong coupling between the two plates (Ma and Clayton, 2015), with a continental lithosphere that has been almost completely removed

(Bishop et al., 2017). Another striking feature is the pronounced low-velocity anomaly beneath the flat slab. We image very thick crustal roots beneath the Altiplano where the Moho is irregular, with depths ranging from 60 to 70 km. The maximum crustal thickness is beneath the current extinct volcanic arc, where we also observe, as in the southern profile, lower seismic velocities in the lower crust. We also observe a strong stratification of the Altiplano crust, with low velocities in the upper crust that is about ∼40 km thick. However, some of the observed contrast between the upper and lower crust may still reflect the footprint of the smooth 1D starting model, as suggested by the stratification observed in the $V_P/V_S$ model. For this reason, it would be wise to await the results of the final short-period inversion results before drawing definite conclusions about the stratification of the crust in that part of the northern Central Andes. Finally, a westward dipping positive velocity anomaly is observed beneath the northern end of the profile, which may result from the underthrusting of the Brazilian lithosphere beneath the Altiplano.

### 6.4.2 Waveform fits

We present two waveform fit examples along the PG transect (including the 5 stations from the PULSE experiment) before and after the 10 s inversion in Fig 6.28 and 6.29. In both examples, the initial waveform fits of the vertical component are already quite coherent with only small differences in arrival time. On the other hand, the fits of the radial component are very poor in the coda of the P wave, in particular for the stations located in the Western and Eastern Cordillera (150 and 400 km distance). The model obtained after the 10 s inversion is able to explain the complex coda wave signals on the radial component.

## 6.5 Discussion

While we are aware that some of the models presented in this thesis are not definitive and that the inversion results are expected to be improved in the near future, we will now present preliminary interpretations of what we believe are the robust features of these tomographic models. Tomographic studies can potentially provide new insight on the removal

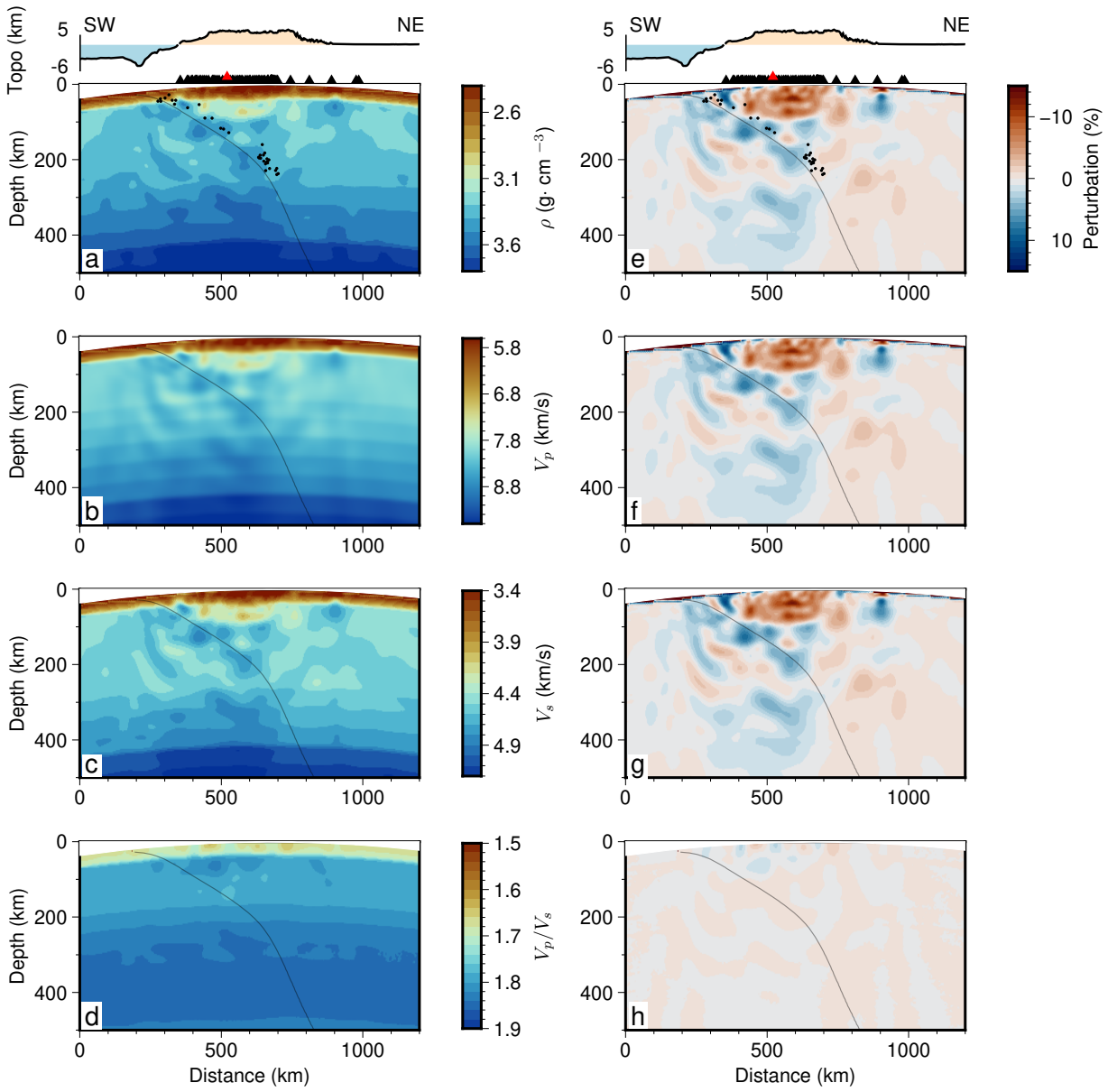


Figure 6.24: Preliminary 10 s results of the FWI in big mesh along the PE transect. (a-d) Models of absolute density, $V_P$, $V_S$, and $V_S/V_S$ respectively. (e-f) Perturbations of density, $V_P$, $V_S$, and $V_S/V_S$ with respect to the initial 1-D smooth AK135 model. The black line shows the Slab2 Nazca contour. The black and red triangles denote the stations and the volcano along the slice, respectively.

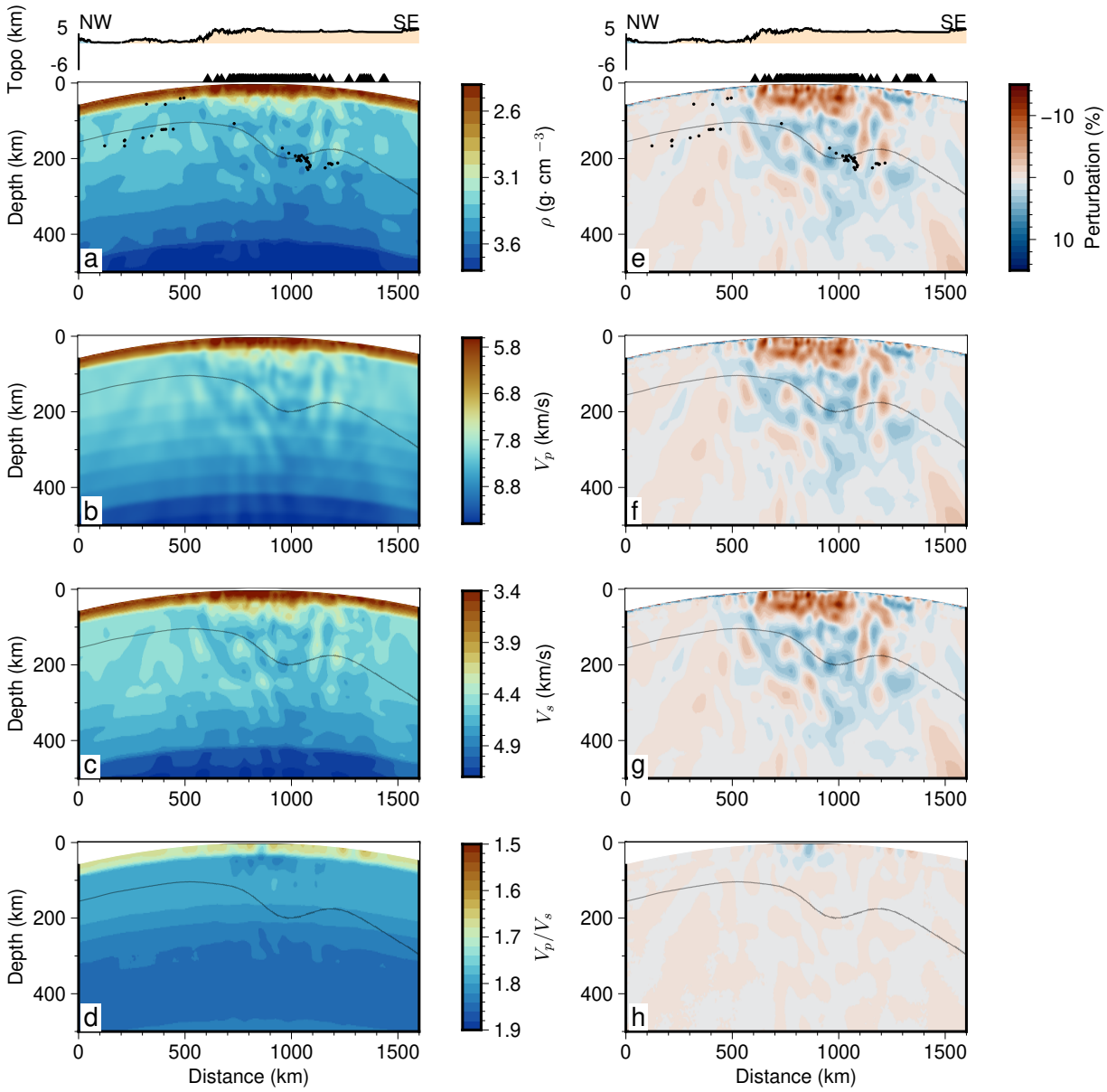


Figure 6.25: Same for Figure 6.24 but for the PF transect.

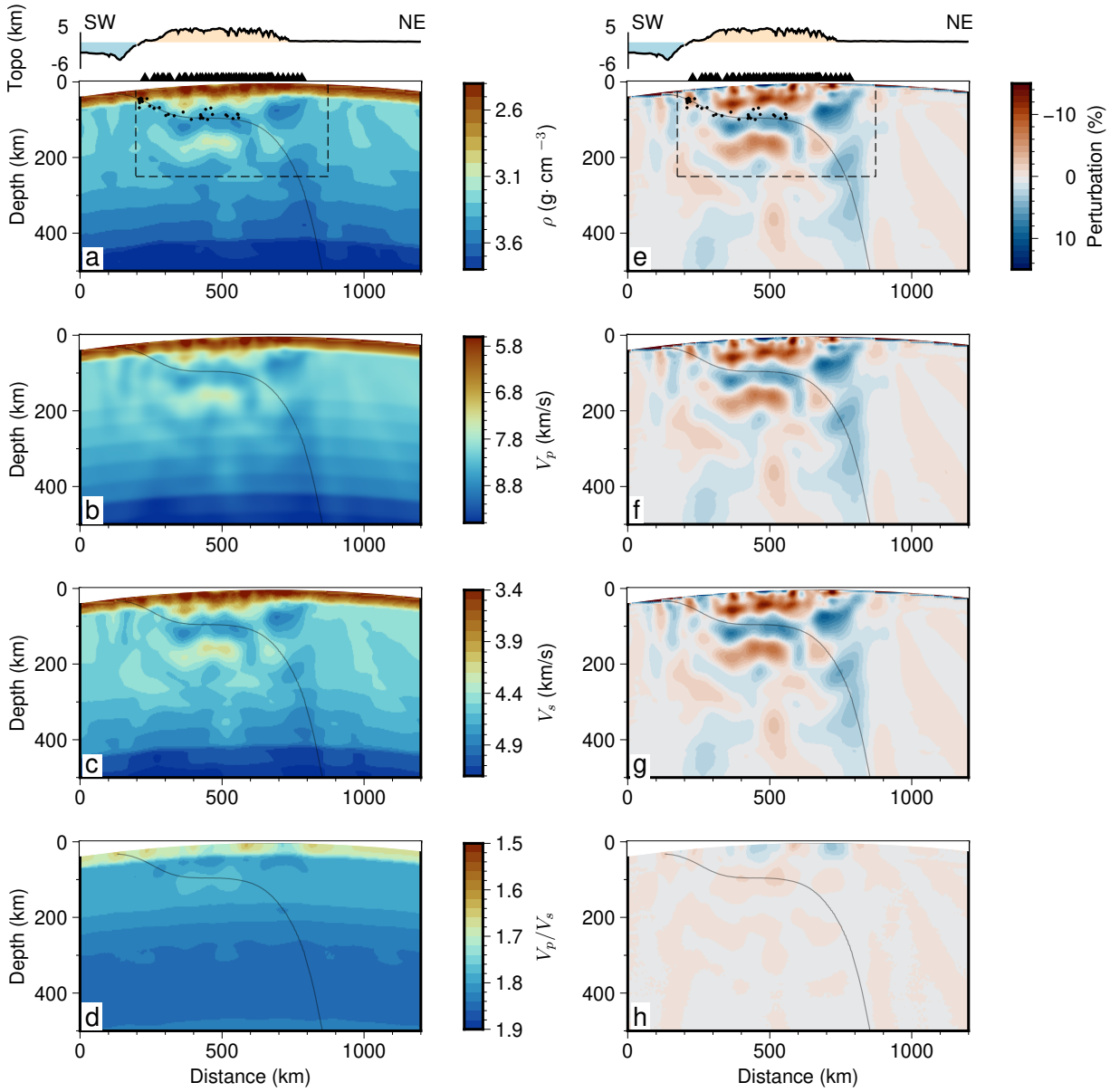


Figure 6.26: Same for Figure 6.24 but for the PG transect.

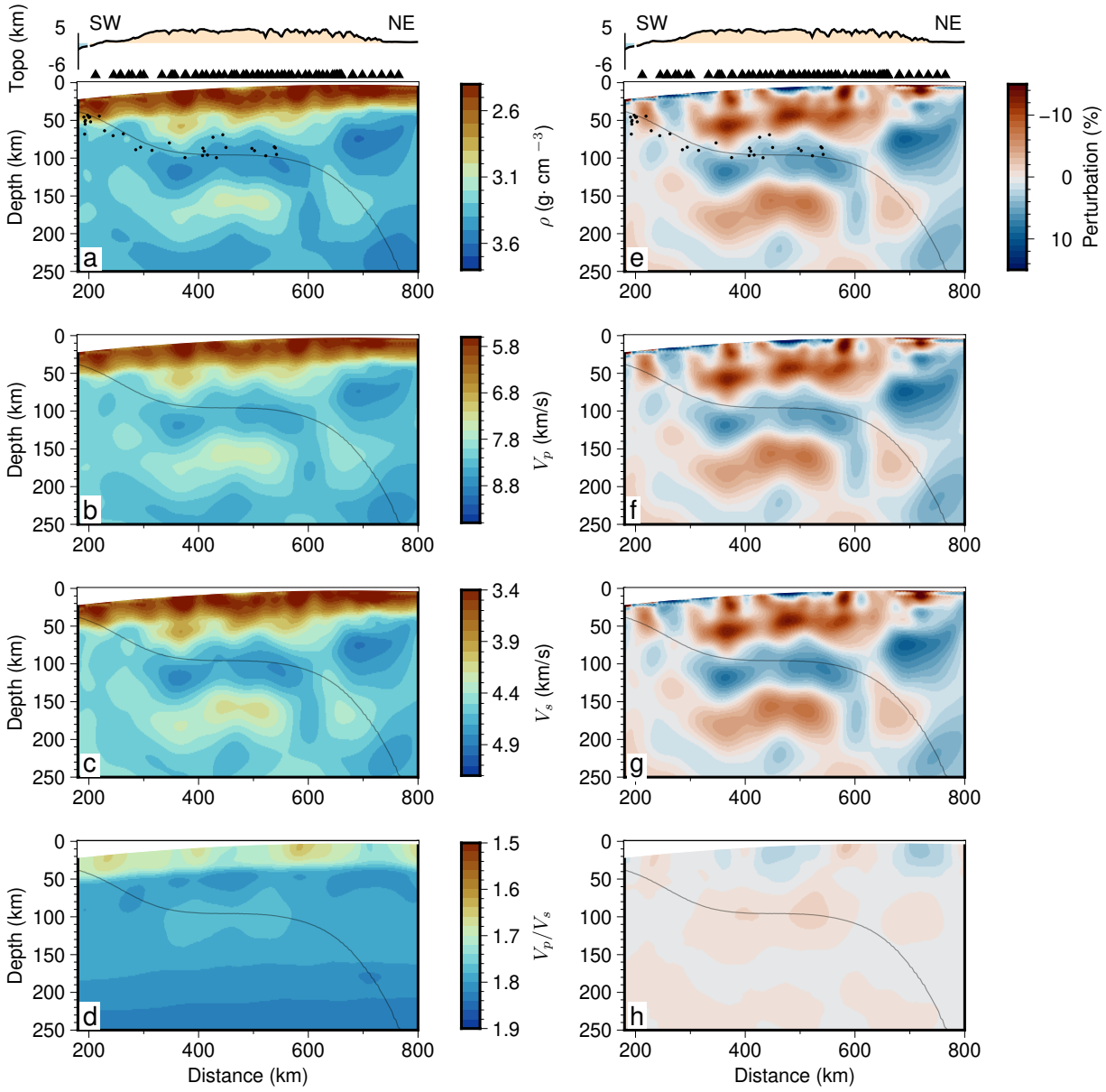


Figure 6.27: Zoom-in of the PG transect shown in Fig. 6.26.

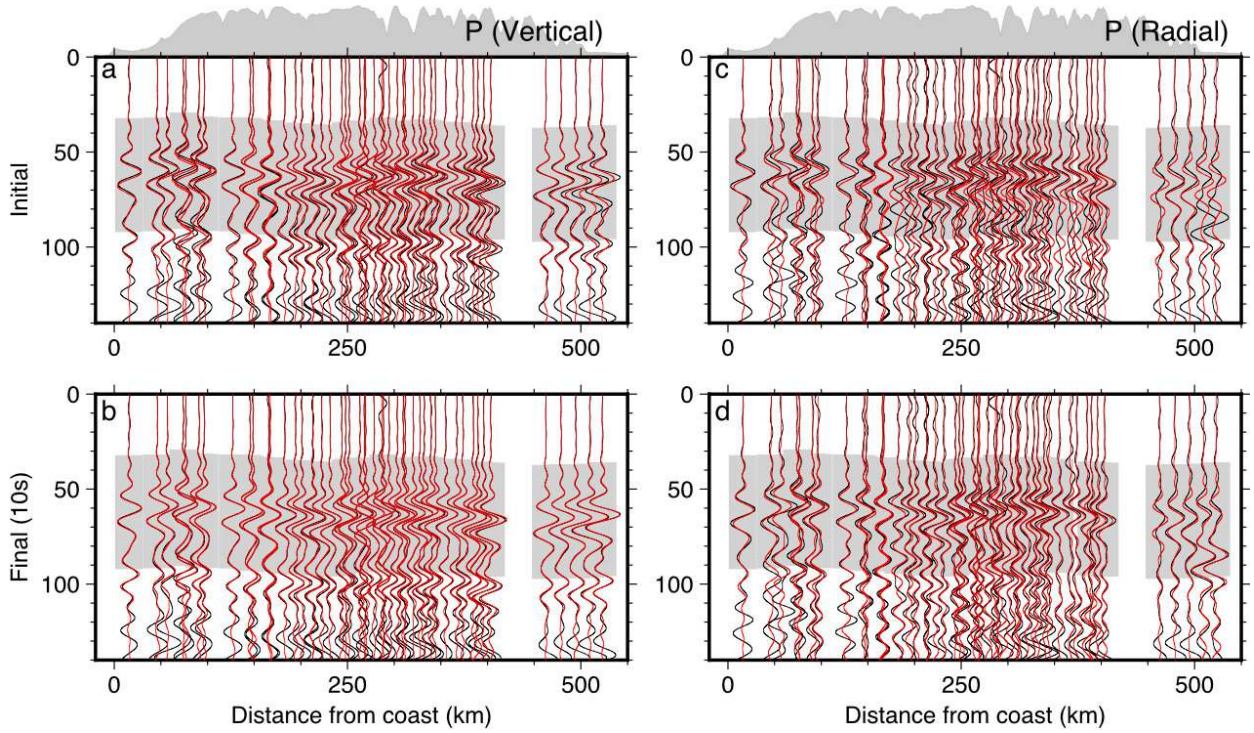


Figure 6.28: Waveform fits of the PG transect for the 2012/04/12 event filtered between 0.04 to 0.10 Hz.

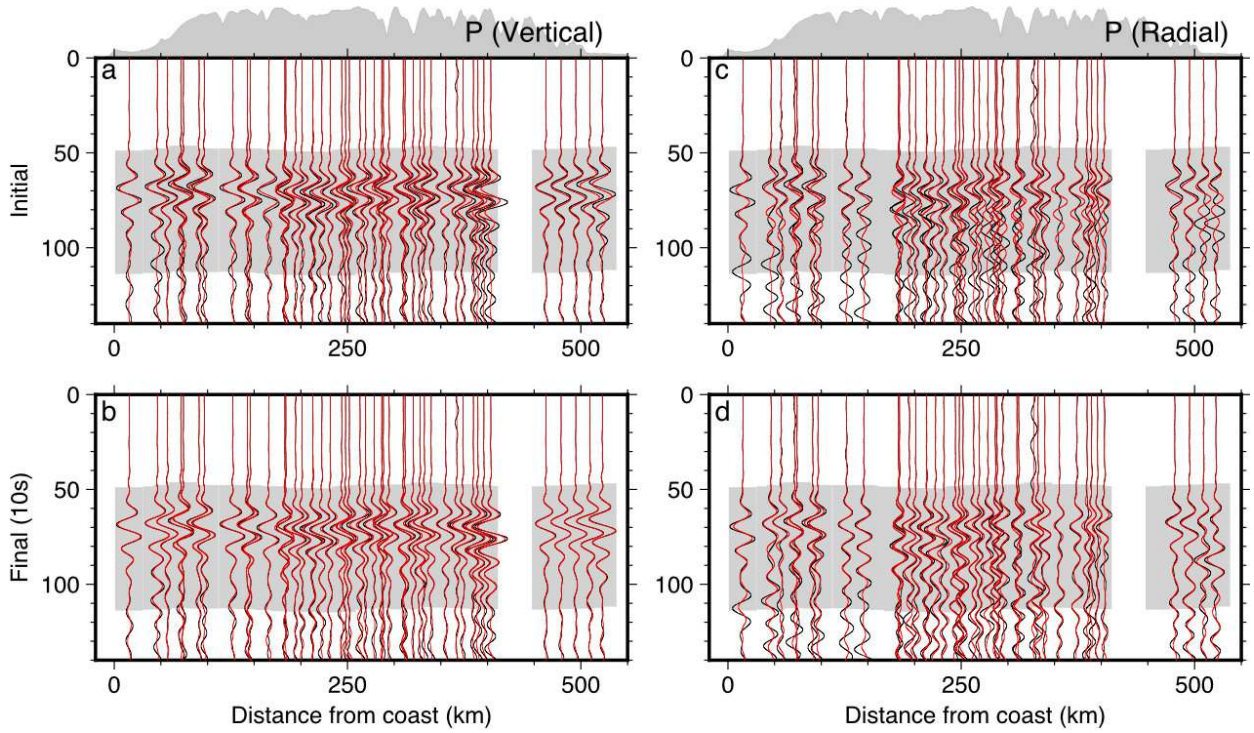


Figure 6.29: Waveform fits of the PG transect for the 2013/01/05 event filtered between 0.04 to 0.10 Hz.

of lithosphere removal and uplift of the Altiplano, on the role of lower crustal flow insight into the redistribution of crustal material (Garzione et al., 2017), as well as on the fate of the lithosphere of the Brazilian craton. In the following, we will confront our tomographic models to these various geodynamical processes that have shaped the Central Andes.

### 6.5.1 Architecture and composition of the northern Central Andes crust

Contrasting results have been reported regarding the velocity structure of the thickened crust and in particular the lower crust, beneath the Altiplano. From the analysis of receiver functions and a compilation of published Rayleigh wave dispersion curves in the Central Andes, Beck and Zandt (2002) concluded that the crust supporting the Altiplano is characterized by an anomalous thickness of about 65 km, which can locally reach up to 75 km beneath the Western and Eastern Cordilleras. This thick crust has also anomalously low seismic velocities suggesting a felsic bulk composition with a mafic lower crust limited at most to a very thin layer at the base of the crust. These authors proposed that this felsic crust would result from crustal thickening and delamination of the eclogitized mafic lower crust driven by density instabilities. This crustal delamination, presumably combined with at least a partial delamination of the lithosphere, could also explain the recent uplift of the Altiplano (Garzione et al., 2006).

Because of the trade-off between discontinuity depths and mean crustal velocity, additional information is needed to constrain absolute depth velocity profiles. In the case of Beck and Zandt (2002), these mean crustal velocities were derived from earlier studies of regional shear-coupled P waves (Zandt et al., 1996), regional P waveforms (Swenson et al., 2000). Our tomographic models as well as more recent tomographic studies based on denser deployments and on 3D imaging approaches (e.g. Graeber and Asch, 1999; Baumont et al., 2002; Ma and Clayton, 2014; Ward et al., 2016; Gao et al., 2021) instead clearly show a layered crust with a high-velocity lower crust, which challenges these interpretations. This discrepancy could be explained by the limitations of the 1D approaches implemented in the early tomographic studies of the Central Andes, as well as the low density of sensors de-

ployed during earlier seismic experiments. Another explanation could be that the crustal architecture is different beneath the northern and southern Altiplano, since the study by Beck and Zandt (2002) focused in the southern Central Andes, whereas the recent dense experiments were mainly deployed in the northern Central Andes. Indeed, the surface wave dispersion results from Baumont et al. (2002) suggest that the average crustal structure is different in the southern and northern parts of the Altiplano. These authors reported low shear velocities in the lower crust (between 3.4 and 3.6 km/s) in the southern Altiplano, whereas beneath the northern Altiplano they found evidence of layered crust, with fast velocities in the lower crust ($>$ 3.6 km/s). In all cases, a strong mid-crustal discontinuity has been documented in most receiver functions studies, both in northern (Phillips et al., 2012; Phillips and Clayton, 2014) and southern (Yuan et al., 2000), including in the study by Beck and Zandt (2002). In our models, the 1D starting model, a smoothed version of the ak135 reference Earth model (Kennett et al., 1995) may still have an imprint in the final 3D models. However, the different sections do show significantly lowered velocities in the top 30 km of the crust, which leads us to conclude that the layered structure of the crust beneath the Altiplano is a robust feature.

Several studies have pointed out that tectonic shortening, on the order of 300 to 350 km, is sufficient to explain the crustal thickening of the Altiplano (Isacks, 1988; Allmendinger et al., 1997). However, the deformation is mainly concentrated in the Eastern Cordillera, with a significant deficit beneath the Altiplano, which suggests that the crustal thickening cannot be explained by a homogeneous crustal deformation (Baby et al., 1997; Husson and Sempere, 2003; Martinod et al., 2020). Different mechanisms have thus been proposed to explain the crustal thickening of the Altiplano, such as the addition of magmatic material, crustal flow (Husson and Sempere, 2003), or the transformation of the eclogite into granulite of the lower crust (Pichon et al., 1997). These different mechanisms are not necessarily mutually exclusive and in any case, discriminating them by considering only the modern lithospheric architecture seen by seismic tomography is probably not an easy task. However,

we note that our models show a thicker crust beneath the volcanic arc, with low seismic velocities down to the Moho. This suggests a crustal thickening resulting from magmatic accumulations which could have driven eastward lateral crustal flow from this gravitational high toward the lower part of the crust beneath the Altiplano, in good agreement with the model proposed by Husson and Sempere (2003).

### 6.5.2 Slab geometry and comparison between FWI and travel time tomography

Figure 6.30 compares the vertical cross-sections along profiles PE and PG in our $V_P$ model with that obtained by Portner et al. (2020) using travel time tomography. The solid black line represents the top of the Nazca plate from the Slab2.0 model (Hayes et al., 2018). We note a good match between the high $V_P$ anomaly in our model with the Slab2.0 model, down to about 250 km depth in the PE section (top), and 400 km beneath the PG section (bottom). At greater depths, the Nazca plate disappears in our model. In contrast, in the travel time tomography model obtained by (Portner et al., 2020), the subducting slab is imaged down to at least 1000 km depth, with the geometry in fair agreement with the Slab2.0 model.

The comparison of these two tomographic models highlights the advantages and shortcomings of FWI over conventional tomographic approaches. FWI mostly constrains the shallow part of the model with a fine spatial resolution but degrades below 300 km depth. The improvement in spatial resolution comes from the full exploitation of teleseismic waveforms, as opposed to only exploiting the arrival time of the direct P wave in traveltime tomography. Because in FWI we search for the model capable of providing a good fit of all the wiggles on three-component seismograms, we must apply very restrictive criteria in our data selection. As a result, we often end up with a very small number of usable teleseismic events, 14 in our case for southern Peru, whereas Portner et al. (2020) inverted 83675 P travel time residuals from 2084 earthquakes. They also considered a much larger number of stations, 1113 in total, which covers the whole of South America. This configuration, with a much larger spatial aperture, is obviously much more favorable for imaging deep

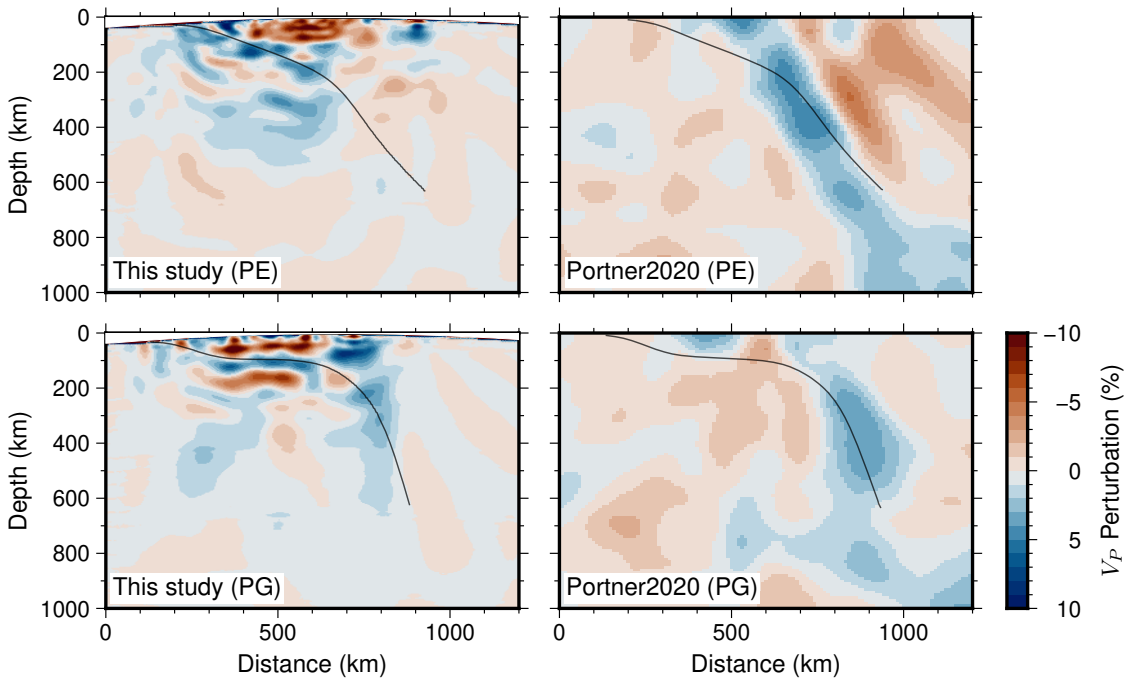


Figure 6.30: Comparison of 10 s $V_P$ perturbation model from big-box inversion along the PE and PG transect and the $V_P$ perturbation model from the finite-frequency teleseismic P wave tomography of South America (Portner et al., 2020).

and large-scale velocity anomalies, such as subducting slabs in the transition zone and lower mantle. However, regional travel time tomography suffers from poor vertical resolution down to 200 km depth, as can be seen in the tomographic section. Since performing FWI at the continental scale with a few thousand of teleseismic events is still beyond the reach of modern supercomputers, combining these two approaches in joint inversions is a promising solution for improving seismic tomography at the regional scale.

Figure 6.31 compares the vertical cross-sections along profiles PE and PG in our $V_S$ model with that obtained by Rodríguez et al. (2021) using travel time tomography. In the latter model, a fast east-dipping anomaly is indeed imaged but offset more than 100 km eastward from the anomaly in the $V_P$ model. A low velocity is also imaged at about 200 km depth, but it is smeared vertically. In general, the results obtained with these two types of tomographic approaches thus appear to be less consistent for shear waves.

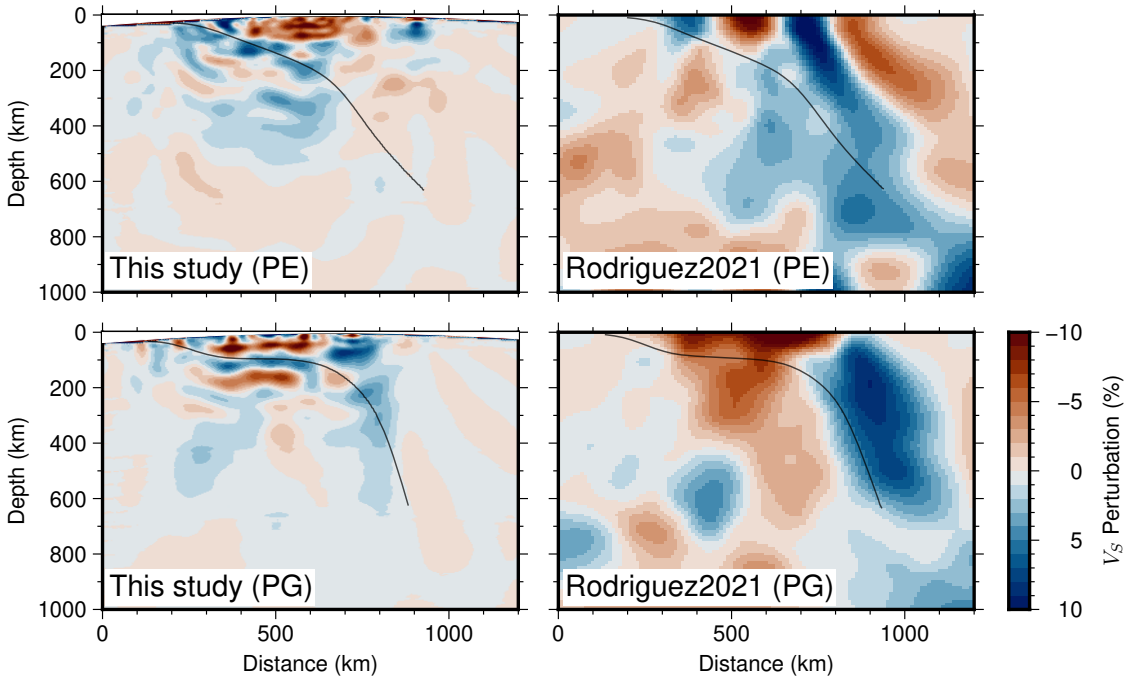


Figure 6.31: Comparison of 10 s $V_S$ perturbation model from big-box inversion along the PE and PG transect and the $V_S$ perturbation model from the finite-frequency teleseismic S wave tomography of South America (Rodríguez et al., 2021).

### 6.5.3 Lithosphere delamination

Our tomographic images beneath the PE (Fig. 6.24) and PG (Fig. 6.26) profiles suggest that the continental lithosphere is reduced in size or even absent. The contact between the two plates in the flat slab segment is near the continental Moho, in good agreement with previous tomographic studies (Phillips and Clayton, 2014; Ma and Clayton, 2015; Scire et al., 2017; Gao et al., 2021). The limited space between the two plates reduces mantle flow and enhances the suction force that favors flat subduction. This raises the question of the timing and mechanism of lithosphere removal. One possibility is that the lithosphere has been progressively eroded after shallowing of the subduction angle. This erosion could have been enhanced by the subduction of a thick oceanic plateau, such as the Inca Plateau (Gutscher et al., 1999; Skinner and Clayton, 2013). An alternative would be that the lithosphere was already eroded before the incipient of the slab subduction, which could have favored the

transition to a flat subduction regime.

The signature of the transition from flat to normal subduction in geological records has been clearly identified, with the rapid eastward migration of the volcanic arc, widespread volcanic activity, and uplift of the Altiplano. However, the processes able to explain these drastic phenomena remain controversial. The steepening of the subducting plate is often associated with the upward flow of the asthenosphere, resulting in widespread melting of the lithosphere (e.g. James and Sacks, 1999; Ramos and Folguera, 2009), which could explain the large recent ignimbrite fields found in the northern Andes (Ramos and Folguera, 2009; Mamani et al., 2009). Our tomographic images are generally in good agreement with this model even though we found no evidence of lithosphere delamination beneath southern Peru. If the flat subduction beneath southern Peru was identical to the one currently observed further north, then this would imply that the lithosphere was already removed before the transition from flat to normal subduction.

### 6.5.4 Buoyancy of the flat slab beneath central Peru

The flat subduction beneath central Peru, first identified by Barazangi and Isacks (1976), has been attributed to the subduction of overthickened oceanic crust (Gutscher et al., 2000). From plate reconstruction, Gutscher et al. (1999) proposed that the subduction of the conjugate of the Marquesas Plateau in the western Pacific, coined the Inca Plateau, could explain the Peruvian flat subduction. However, later reconstructions (Skinner and Clayton, 2013) found a position shifted by about 600 km for the Inca Plateau, therefore unable to explain the present-day flat slab support. Since the positive buoyancy of the subducted oceanic crust is suppressed by eclogitization reactions, this mechanism also requires a metastability of the basalts and gabbros of the subducting oceanic crust (Gutscher et al., 2000). Our detailed tomographic images of the subduction beneath Cascadia (see Chapter 5) suggest that eclogitization is complete after about 1 Ma, a very fast reaction rate. This is confirmed by the analysis of receiver functions in the flat slab region which shows a rapid decrease of P-to-S conversion amplitudes beneath 75 km depth, which suggests that eclogitization of the

oceanic crust is completed well ahead of the termination of the flat slab segment (Bishop et al., 2017). This indicates that additional forces must be acting to maintain the flat geometry. Analogue modelling (Martinod et al., 2005) as well as numerical thermo-mechanical modeling of flat subduction has also shown that the subduction of the oceanic plateau is insufficient to sustain flat subduction (van Hunen et al., 2002; Gerya et al., 2009) and that other mechanisms need to be invoked, such as the active overthrusting of the overriding continental plate or plate suction (Manea et al., 2012). This latter mechanism is influenced by the shape of the mantle wedge. For example, a 250 km thick cratonic root 300 km from the trench results in a suction force that is twice that of a rootless model (O'Driscoll et al., 2009). Manea et al. (2012) have shown that indeed the presence of a thick craton at a distance of up to 900 km from the trench favors the development of a flat subduction zone.

A prominent feature in our model is the low-velocity anomaly beneath the flat slab segment, between 120 and 170 km depth (Fig. 6.26). This anomaly is in remarkable agreement with the Rayleigh wave tomography study by Antonijevic et al. (2015), which also documented a pronounced low-velocity anomaly at the exact same location. In their finite-frequency travel time tomography of south Peru, Scire et al. (2016) also imaged a slow velocity anomaly beneath the flat segment, which extends vertically down to about 250 km depth. However, this vertical smearing may simply result from the limited depth resolution of regional body wave travel time tomography. These authors noticed that this low-velocity anomaly is confined to the region where the Nazca ridge has been subducted. This observation led Portner et al. (2017) to relate it to a thermal anomaly generated at the Easter hotspot that has been entrained with the Nazca plate. According to Anderson (1989), this low-velocity anomaly could be explained by a temperature anomaly of 200°C or the presence of at least 1 % of partial melt. If this hypothesis is correct, then this anomalous mantle could contribute to the buoyancy of the flat slab segment.

### 6.5.5 The fate of the Amazonian lithosphere

Tectonic shortening in the northern Central Andes is not homogeneously distributed. Balanced cross sections have shown that it is concentrated in the subandean and Eastern Cordillera, with a total shortening of the order of 200 km (Baby et al., 1997). While the thin-skin deformation in the subandean domain is rather well constrained, the fate of the Brazilian craton lithosphere is still controversial. The analysis of gravity anomalies has shown that the subandean domain is overcompensated whereas the Chaco Plain further east is undercompensated (Lyon-Caen et al., 1985). These observations can be explained by the flexure of the Brazilian shield plate, but this requires a westward extension of at least 150 km, reaching the Eastern Cordillera. Thermo-mechanical modeling is in good agreement with these estimates and predicts a delamination of the Brazilian lithosphere beneath the EC (Sobolev et al., 2006).

The southern PE transect, sampling the normal subduction beneath southern Peru, ends near Arequipa, west of the EC. The deep structures beneath the EC are thus poorly constrained. In contrast, farther north, above the flat subduction segment, we have a complete transect of the northern Central Andes from the coastline to the subandean domain, combining stations installed along the PG profile, which has been extended to the NE with the addition of 5 stations from the PULSE experiment. The cross-section along the PG transect in our final models (Fig. 6.27) clearly show a high-velocity anomaly dipping to the west, which appears to be connected to the subandean crust. Similar high-velocity anomalies beneath the EC have been documented by Scire et al. (2016, 2017) beneath south Peru.

## 6.6 Conclusions

We have imaged the transition from normal to flat subduction beneath South Peru, from the surface down to 500 km depth, by inverting complete P and SH teleseismic waveforms. The careful analysis of P and SH waveforms recorded by the PE, PF, and PG transects from the PeruSE experiment has revealed inconsistent phase and amplitude residuals between

the different deployment phases. We quantified the phase and amplitude corrections that need to be applied in order to obtain a consistent waveform dataset. The final density, $V_P$, and $V_S$ tomographic models provided new insight into the architecture of the Altiplano, the geometry of the Nazca plate, in both the normal and flat segments. The low seismic velocities observed beneath the southern part of the model suggest that the lithosphere has been removed, which could explain the recent uplift of the Altiplano. We also observe a low-velocity anomaly beneath the flat segment of the Nazca plate, which may contribute to sustaining the shallow dip of the slab.

## 6.A Waveform fits after hierarchical FWI inversions in small box inversion

Besides the waveform fits shown in the main text, we present the waveform fits for the other events that are included into the inversion (Table 6.2).

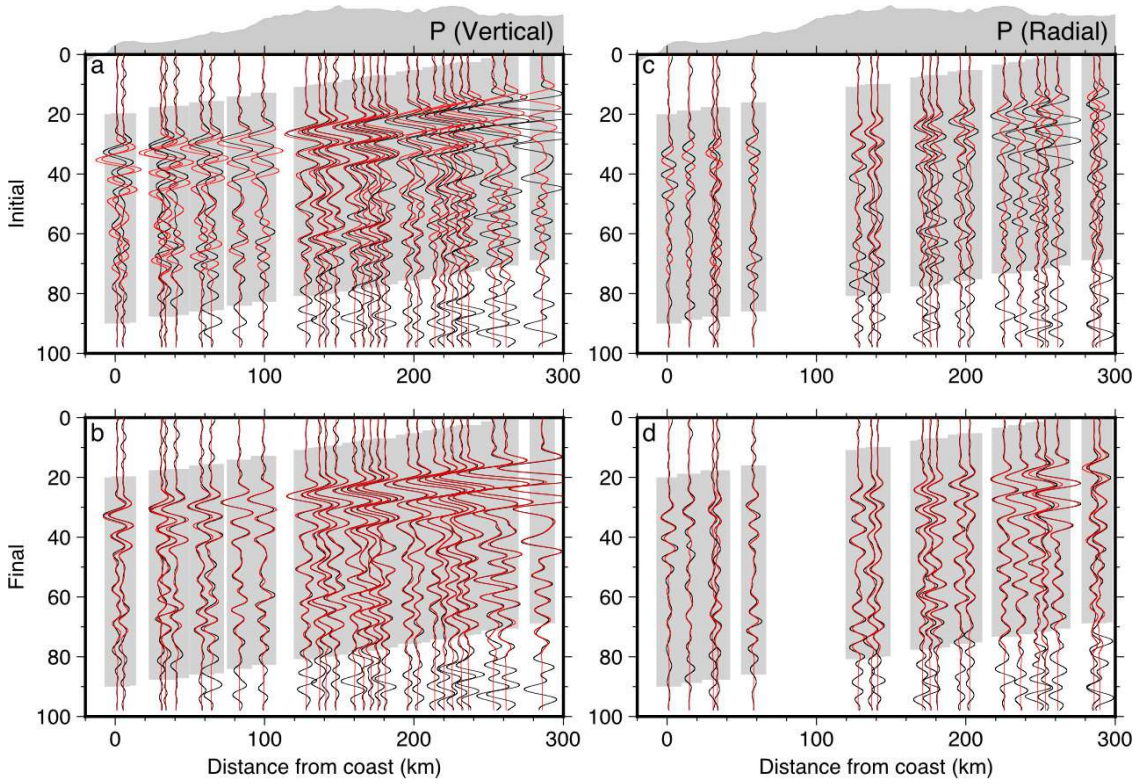


Figure 6.32: Waveform fits of the PE transect for the event 1 (2008/09/10) filtered between 0.04 to 0.25 Hz. Observed (black lines) and synthetic (red lines) vertical component P waveforms computed in the initial (a) and final (b) models. (c,d) Same as (a,b) but for radial component P waveforms. (e,f) Same as (a,b) but for transverse component S waveforms. The grey backgrounds indicate the beginning and end of the time windows used in the inversion.

Table 6.5 shows the percentage of cost reduction for each event. The cost reductions between the final model and the initial smooth 1-D mode are about 70% to 90%.

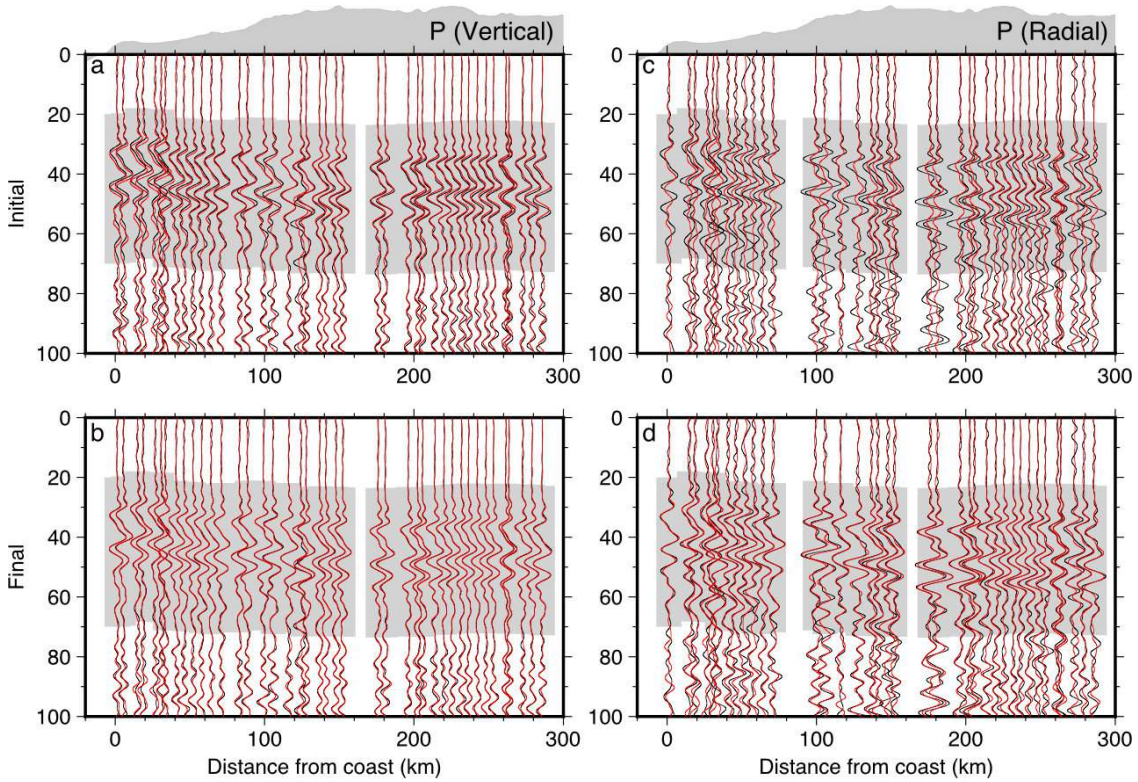


Figure 6.33: Same as Fig. 6.32 but for the waveform fits of event 2 (2010/01/05).

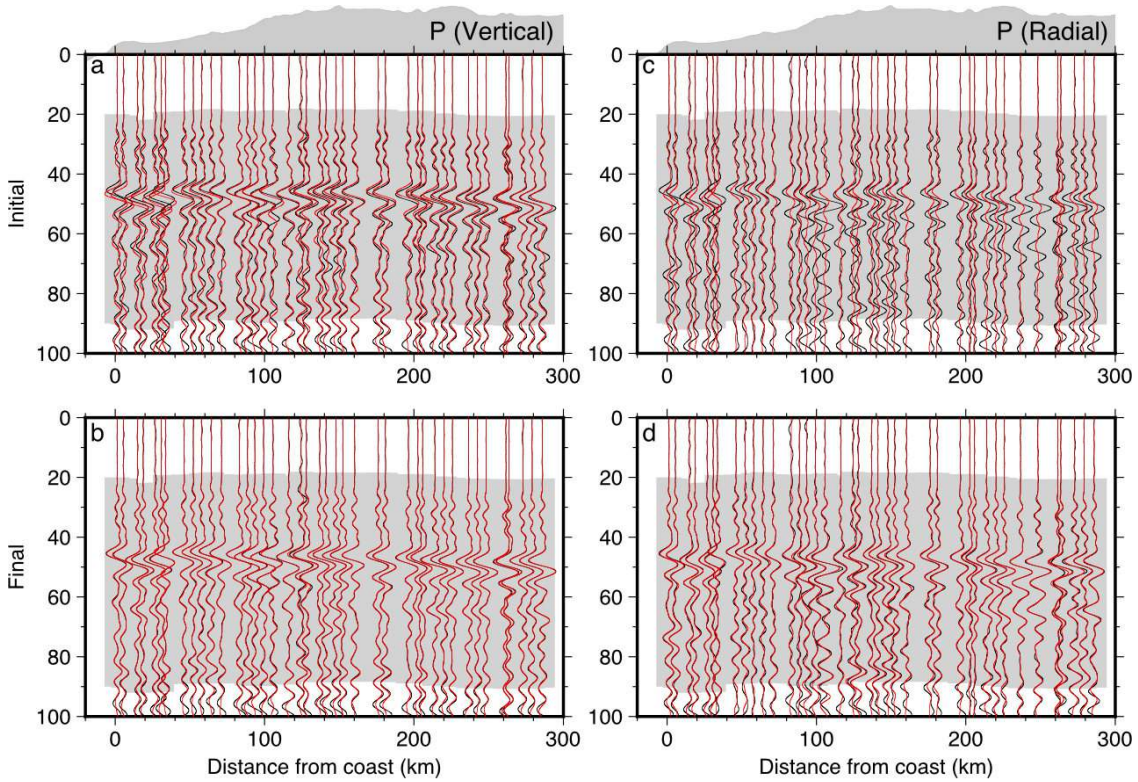


Figure 6.34: Same as Fig. 6.32 but for the waveform fits of event 5 (2010/04/04).

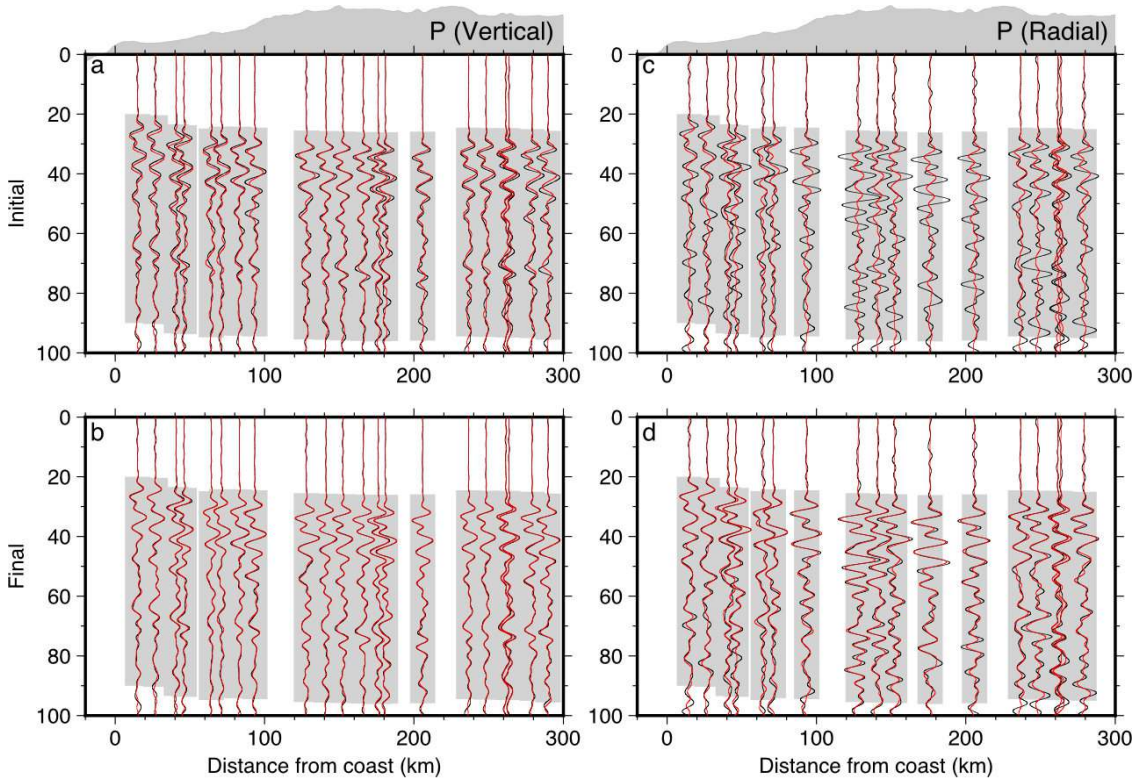


Figure 6.35: Same as Fig. 6.32 but for the waveform fits of event 6 (2010/12/08).

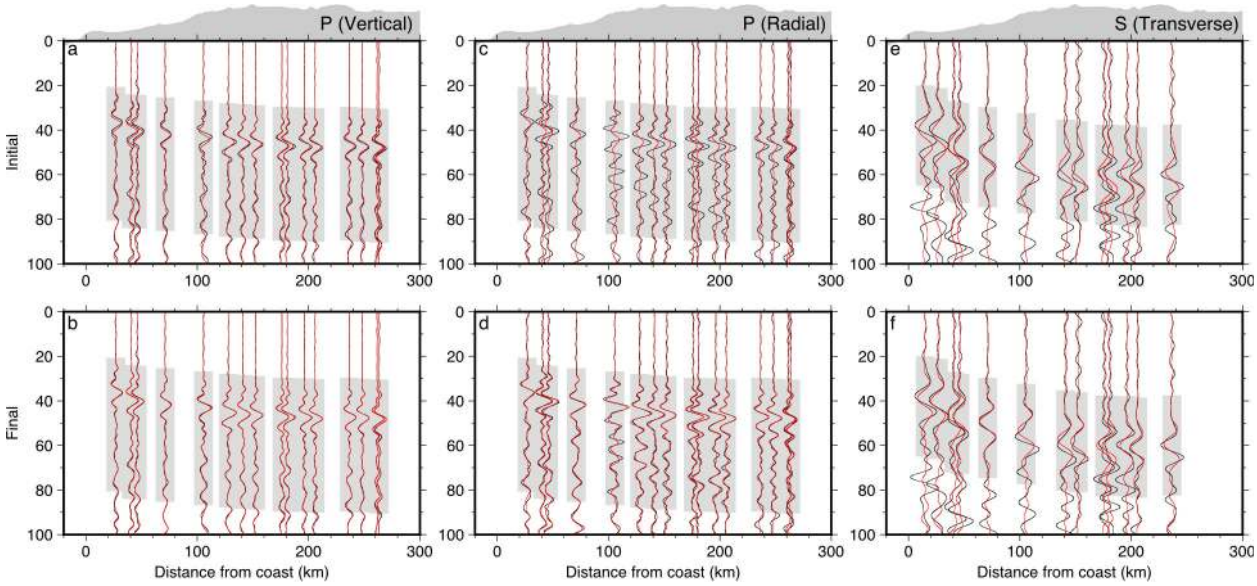


Figure 6.36: Same as Fig. 6.32 but for the waveform fits of event 7 (2012/01/15).

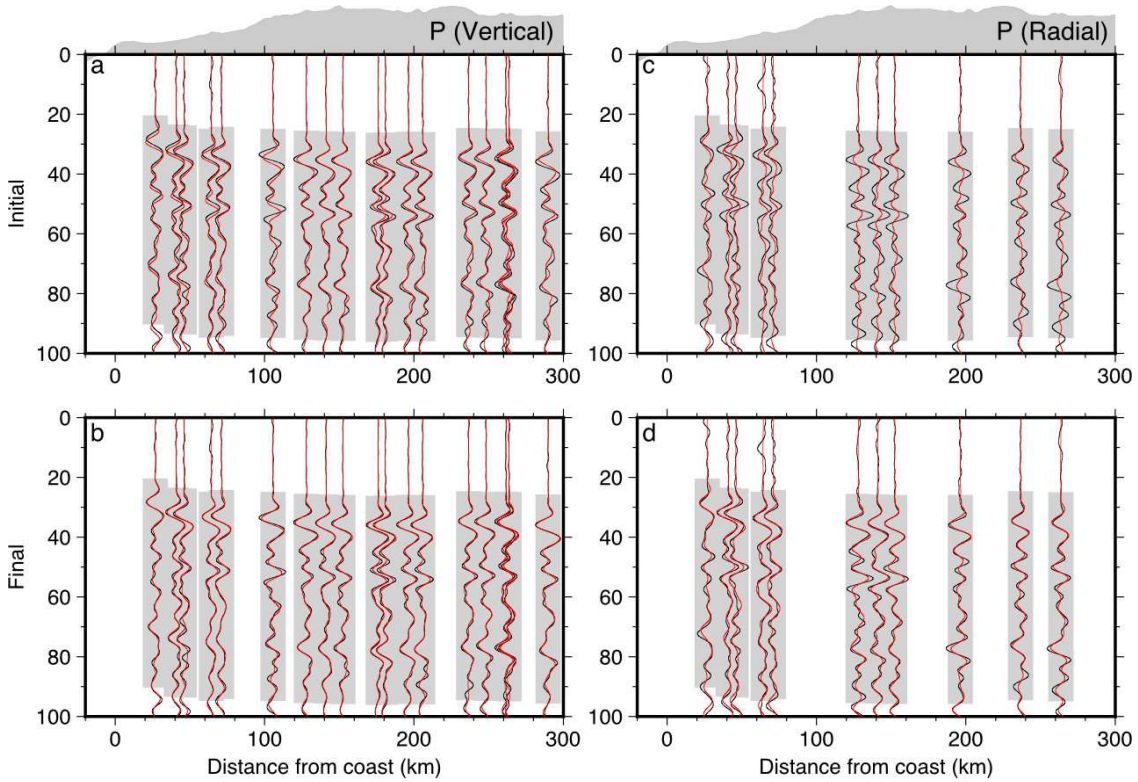


Figure 6.37: Same as Fig. 6.32 but for the waveform fits of event 8 (2012/01/22).

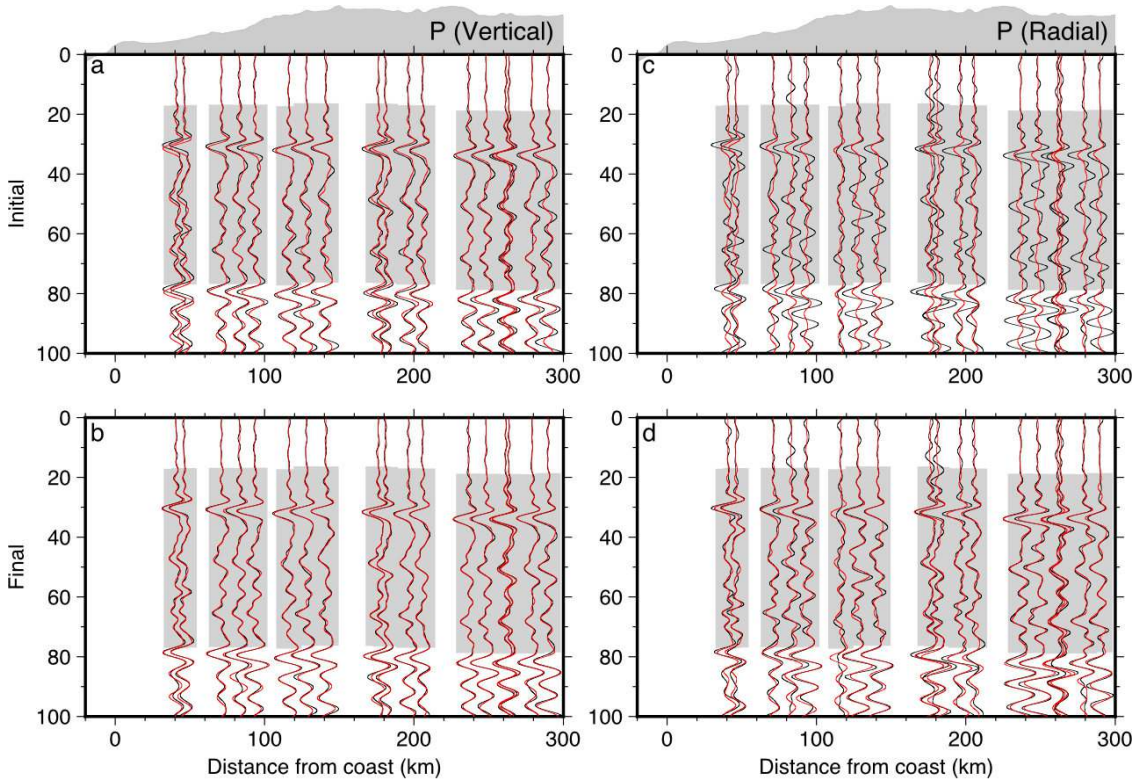


Figure 6.38: Same as Fig. 6.32 but for the waveform fits of event 9 (2012/04/12).

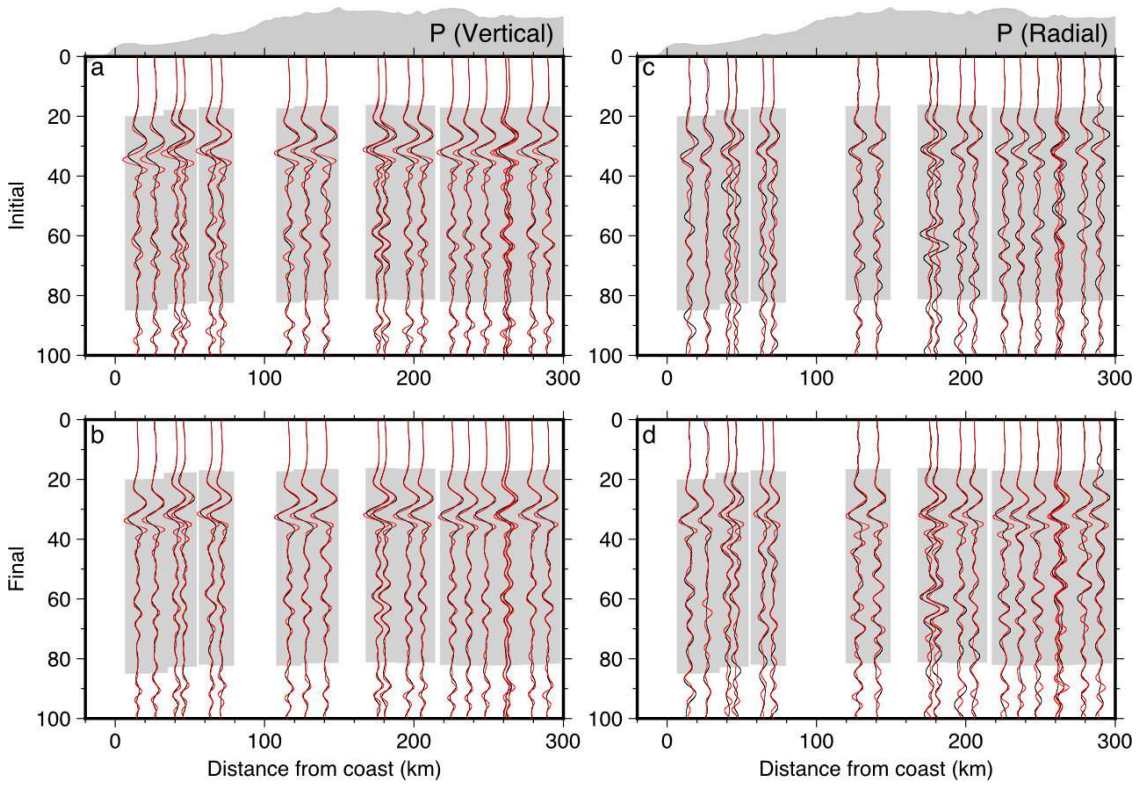


Figure 6.39: Same as Fig. 6.32 but for the waveform fits of event 10 (2013/01/05).

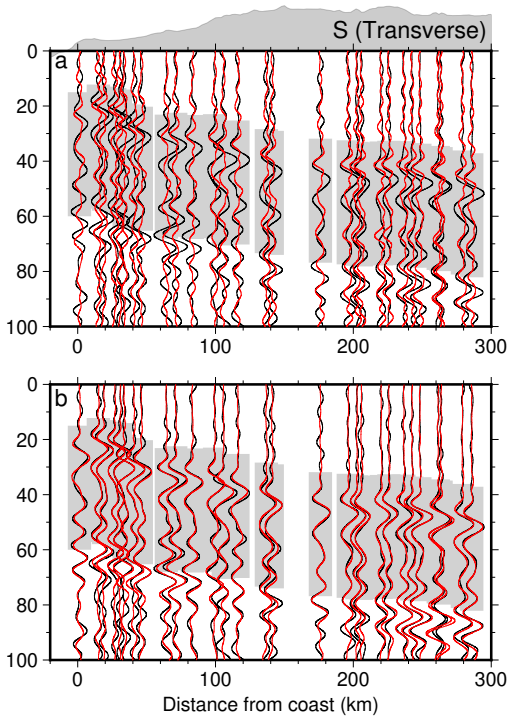


Figure 6.40: Same as Fig. 6.32 but for the waveform fits of event 11 (2010/01/17).

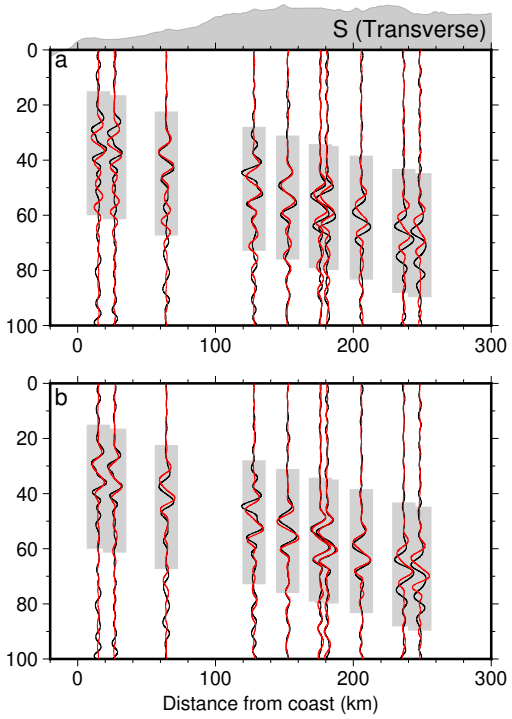


Figure 6.41: Same as Fig. 6.32 but for the waveform fits of event 12 (2011/03/01).

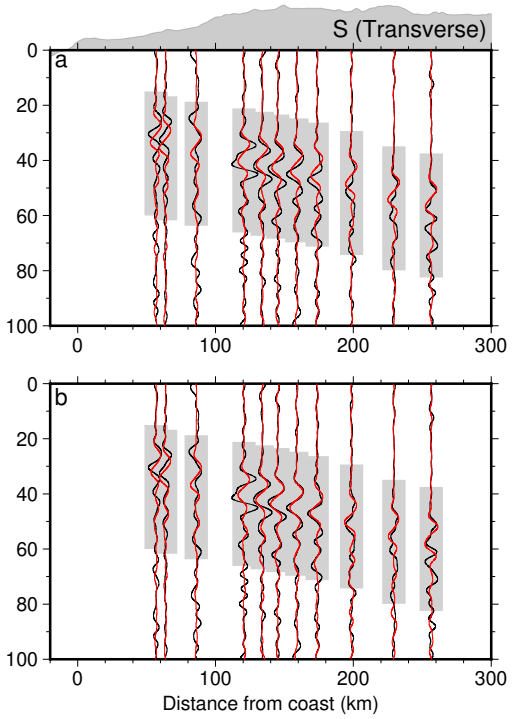


Figure 6.42: Same as Fig. 6.32 but for the waveform fits of event 14 (2011/03/01).

Table 6.5: Residual misfits for each event, defined as the ratio between the misfit in the final and the misfit in the initial smooth 1-D model (in percents).

| Event | Phases | Origin Time | Vertical (%) | Radial (%) | Transverse (%) | Total (%) |
|---|---|---|---|---|---|---|
| 1 | P | 2008/09/10 13:08:14.9 | 96 | 88 | - | 94 |
| 2 | P | 2010/01/05 04:55:38.9 | 90 | 80 | - | 86 |
| 3 | P+SH | 2010/01/12 21:53:10.2 | 84 | 81 | 71 | 80 |
| 4 | P | 2010/03/07 07:05:23.3 | 91 | 85 | - | 88 |
| 5 | P | 2010/04/04 22:40:43.0 | 93 | 87 | - | 91 |
| 6 | P | 2010/12/08 05:25:35.3 | 89 | 77 | - | 84 |
| 7 | P+SH | 2012/01/15 13:40:19.5 | 82 | 75 | 52 | 69 |
| 8 | P | 2012/01/22 05:53:42.1 | 94 | 88 | - | 92 |
| 9 | P | 2012/04/12 07:15:48.5 | 81 | 73 | - | 77 |
| 10 | P | 2013/01/05 08:58:19.3 | 84 | 70 | - | 82 |
| 11 | SH | 2010/01/17 12:00:01.1 | - | - | 91 | 91 |
| 12 | SH | 2011/03/01 00:53:46.3 | - | - | 93 | 93 |
| 13 | SH | 2011/07/27 23:00:30.3 | - | - | 91 | 91 |
| 14 | SH | 2011/09/23 19:02:48.2 | - | - | 93 | 93 |

# Chapter 7. Conclusions and perspectives

## Contents

7.1 Français ..... **209**
7.2 English ..... **214**

### 7.1 Français

L'objectif principal de cette thèse était d'améliorer l'algorithme existant d'inversion des formes d'ondes complète télésismiques. Le potentiel de la nouvelle approche a été illustré sur le problème de l'imagerie des zones de subduction, avec deux applications dans les régions des Cascadia et du sud Pérou.

Les applications précédentes de l'inversion de formes d'ondes télésismiques ont considéré une paramétrisation du modèle en termes de densité, $V_P$ et $V_S$, mais les corrélations existant entre ces paramètres n'ont pas été prises en compte comme une information a priori dans l'inversion. Dans la première partie de cette thèse, nous avons introduit une matrice de covariance du modèle a priori complète, qui décrit l'écart-type et la corrélation spatiale de chaque paramètre dans le modèle, ainsi que les corrélations entre les paramètres. Nous avons également montré comment une telle matrice de covariance du modèle complète peut être incluse dans une nouvelle formulation bayésienne d'inversion de formes d'onde complètes. Cette nouvelle approche a été validée sur un modèle simple de subduction dans lequel nous avons effectué différents tests synthétiques d'inversion afin de quantifier l'amélioration des résultats de l'inversion de formes d'onde lorsque l'on considère une matrice de covariance complète du modèle. Lorsque les corrélations entre les paramètres du modèle sont prises en compte, la résolution spatiale dans les modèles reconstruits est grandement améliorée, en particulier pour la densité et le rapport $V_P/V_S$. Ceci constitue une amélioration significative par rapport aux approches classiques qui négligent les corrélations entre les paramètres du

modèle car ces deux paramètres fournissent des contraintes directes sur l'état thermique, la composition et la présence de fluides dans la lithosphère. Nous avons ensuite appliqué avec succès le nouvel algorithme d'inversion de formes d'ondes à des données provenant de l'expérience CASC93 dans les Cascades et de différentes expériences temporaires déployées dans le sud du Pérou. Pour améliorer la procédure de sélection des données, nous avons développé une interface graphique pour visualiser les formes d'onde ainsi que des observations auxiliaires telles que les résidus de temps de trajet et d'amplitude pour détecter d'éventuels problèmes dans les formes d'onde. Nous avons ainsi identifié des décalages systématiques de temps et d'amplitude sur des stations appartenant à différents profils de l'expérience PERUSE dans le sud du Pérou, et nous les avons corrigés. L'ajout de formes d'onde SH dans les inversions de formes d'onde complètes est un autre développement nouveau réalisé dans cette thèse. Au final, nous avons obtenu des modèles à haute résolution de la densité, $V_P$, $V_S$ et $V_P/V_S$ pour les zones de subduction des Cascadia Centrales et du sud Pérou. Dans les deux régions, nous pouvons clairement visualiser une couche de faible vitesse associée à la croûte océanique subduite et la structure détaillée sous l'avant-arc et l'arrière-arc. Nous avons également obtenu des images des plaques plongeantes avec des niveaux de détails structurels sans précédent. Ces images tomographiques à haute résolution ont fourni de nouvelles contraintes sur les migrations de fluides dans les zones de subduction.

**Amélioration de l'algorithme et du flux de travail de la FWI**

Malgré les développements méthodologiques de l'algorithme d'inversion de formes d'onde et les résultats tomographiques plus qu'encourageants obtenus au cours de cette thèse, nous avons identifié plusieurs aspects qui pourraient être améliorés.

Tout d'abord, nous n'avons considéré dans cette thèse que des milieux élastiques isotropes, et nous avons ignoré les effets de l'anisotropie sismique sur les formes d'onde P et S. Cela peut conduire à des artefacts de vitesse qui peuvent éventuellement être mal interprétés comme des hétérogénéités de composition, en particulier dans les zones de subduction (Bezada et al., 2016; Beller and Chevrot, 2020; VanderBeek and Faccenda, 2021). En principe, il est simple

d'étendre la modélisation de la propagation des ondes dans les milieux anisotropes avec des solveurs SEM (e.g. Chevrot et al., 2004). L'étude récente de Beller and Chevrot (2020) a étendu la FWI isotrope à la FWI anisotrope avec des modèles paramétrés en termes de densité et des 21 coefficients élastiques du tenseur d'élasticité d'ordre 4. Avec cette approche, ils ont pu récupérer l'orientation de l'axe de symétrie (y compris le pendage) ou la stratification de l'anisotropie sismique, deux problèmes notoirement difficiles. Pour généraliser la méthode d'inversion présentée dans le chapitre 4, le défi sera de définir des matrices de covariance du modèle a priori impliquant la densité et les 21 composantes du tenseur élastique. Dans tous les cas, les futures applications de la FWI pour l'imagerie des milieux anisotropes 3D impliqueront l'inversion conjointe des phases télésismiques P et S afin de caractériser complètement l'anisotropie sismique (Beller and Chevrot, 2020). Nous pensons que ces nouveaux modèles tomographiques conduiront également à des modèles tomographiques isotropes moins biaisés.

Deuxièmement, le modèle initial utilisé dans cette étude est fixé au modèle AK135 lissé. Nous avons vu dans le chapitre sur les Cascades que le modèle inversé reste proche du modèle initial. Ceci est dû au fait que les résidus de temps de parcours moyen et d'amplitude pour chaque événement ont été supprimés. Par conséquent, nous pouvons seulement imager les perturbations relatives au modèle de départ et les modèles de vitesses absolus sont moins bien contraints. Ce problème est particulièrement critique dans des contextes tectoniques tels que les Andes centrales où le modèle 1-D moyen est très éloigné du modèle moyen global. Utiliser un modèle de vitesse 1D adapté à la région étudiée comme modèle initial au lieu du modèle moyen global devrait améliorer les résultats de l'inversion. Du point de vue de l'espace des modèles, cela devrait également accélérer la convergence de l'inversion. Pour mieux contraindre le modèle absolu, nous pourrions également exploiter des données supplémentaires et indépendantes, telles que la dispersion des ondes de surface (e.g. Rawlinson and Fishwick, 2012; Schmandt and Lin, 2014) ou utiliser des moyennes locales de modèles tomographiques globaux (e.g. Bastow, 2012).

Troisièmement, nous avons montré que la reconstruction du modèle de densité en profondeur est grandement améliorée lorsque les corrélations entre la densité et $V_P$ et entre la densité et $V_S$ sont prises en compte dans les inversions. Lorsqu'aucune corrélation n'est introduite, la reconstruction de la densité est principalement limitée à la partie peu profonde du modèle (e.g. Beller et al., 2018), ce qui implique que les formes d'ondes télésismiques ne sont pas très sensibles aux anomalies de densité profondes. Les données gravimétriques (anomalie de Bouguer ou ondulations du géoïde) fournissent des contraintes directes sur la densité, et nous pouvons qu'en incluant ces données dans l'inversion de formes d'onde télésismiques devrait permettre de mieux contraindre le modèle de densité 3D. L'inversion conjointe des données classiques de temps de trajet et des données gravimétriques est déjà un domaine actif (e.g. Lelièvre et al., 2012) mais l'application à l'inversion de formes d'onde télésismiques restent à faire.

Quatrièmement, lorsque des données sur les tremblements de terre locaux sont disponibles, en plus des données télésismiques, elles devraient permettre d'améliorer la couverture et donc potentiellement la résolution spatiale de la FWI. L'étude pionnière de FWI réalisée par Gao et al. (2021) dans les Andes centrales a démontré le potentiel des sources locales pour l'imagerie à haute résolution des structures lithosphériques. En outre, la fréquence des séismes locaux est plus élevée (jusqu'à 10 Hz) que celle des événements télésismiques. Ainsi, les séismes locaux peuvent fournir des contraintes sur les modèles structurels et d'atténuation à petite échelle, bien que nous ayons besoin de grilles plus fines pour modéliser la propagation des ondes à des fréquences plus élevées, avec un coût de calcul accru. Un compromis pourrait être de contraindre d'abord les structures à grande échelle en utilisant la FWI télésismique, comme nous l'avons fait dans cette thèse, puis de limiter le domaine de calcul à la profondeur des séismes locaux pour affiner le modèle dans sa partie superficielle. L'inversion conjointe des temps de trajet des événements locaux et télésismiques (e.g. Kuo-Chen et al., 2012) est une autre voie prometteuse pour améliorer la résolution de la FWI télésismique.

Enfin, la caractérisation de la résolution des modèles tomographiques reste une question

importante. En effet, être capable de calculer la matrice de covariance du modèle a posteriori pourrait fournir les incertitudes du modèle ainsi que la résolution du modèle inversé.

**Applications aux données réelles**

Dans cette thèse, nous avons introduit pour la première fois les ondes SH dans la FWI télésismique et nous avons montré que cela améliorait la résolution et la robustesse des modèles inversés. D'autres types d'ondes de cisaillement, telles que les ondes SV ou les ondes SKS, devraient fournir des contraintes supplémentaires sur les modèles $V_S$ (ou l'anisotropie sismique). En général, l'inclusion de phases sismiques supplémentaires augmentera l'illumination angulaire des cibles lithosphériques, ce qui devrait améliorer la résolution spatiale.

En plus des données provenant d'expériences temporaires, il existe des réseaux denses de stations large bande avec des données disponibles sur plusieurs années, par exemple dans la région de Vancouver, au Japon, ou dans le nord de Taïwan, installés au dessus de zones de subduction actives. Les futures applications de la FWI sur ces réseaux régionaux permanents et denses devraient fournir des images à très haute résolution et apporter de nouvelles connaissances sur les systèmes de subduction.

## 7.2 English

The main objective of this thesis was to improve the existing teleseismic full waveform inversion algorithm. The potential of the new approach has been illustrated in the problem of imaging subduction zones, with two applications in Cascadia and Southern Peru.

Previous FWI applications have mostly parameterized the model in terms of density, $V_P$, and $V_S$, but the correlations among these parameters were not considered as a priori information in the inversion. In the first part of this thesis, we have introduced a complete a priori model covariance matrix, which describes the standard deviation and spatial correlation of each parameter in the whole model space, as well as the correlations between parameters. We have also shown how such a complete model covariance matrix can be included in a new formulation of a Bayesian full waveform inversion approach. This new approach has been validated on a simple subduction model in which we have performed different inversion synthetic tests to quantify the improvement of FWI results when considering a complete model covariance matrix. When the correlations between model parameters are taken into account, the spatial resolution in the reconstructed models is greatly improved, especially for density and $V_P/V_S$. This constitutes a significant improvement over classical approaches, which neglect the correlations between model parameters as these two parameters provide direct constraints on the thermal state, composition, and fluid content of the lithosphere. We then successfully applied the improved FWI algorithm to real waveform datasets from the CASC93 experiment in the Cascades and different temporary experiments deployed in South Peru. To improve and make the selection of high-quality datasets, we developed a graphical interface to visualize the waveforms and determine auxiliary observations such as travel time and amplitude residuals to detect problems in the waveforms. We identified systematic time and amplitude shifts of the stations across different profiles in the PERUSE experiment in southern Peru and corrected them. The addition of SH waveforms in full waveform inversions is another new development accomplished in this thesis. In the end, we obtained high-resolution models of density, $V_P$, $V_S$, and $V_P/V_S$ for central Cascadia and

South Peru. In both regions, we can clearly image a low-velocity layer associated with the subducted oceanic crust and the detailed structure beneath the forearc and backarc. We also imaged the high-velocity subducted slabs with unprecedented levels of structural details. These high-resolution tomographic images provided new constraints on fluid migrations in subduction zones.

**Improving the FWI algorithm and workflow**

Despite methodological developments of the FWI algorithm and the more than encouraging tomographic results obtained during the course of this thesis, we have identified several aspects of FWI that could lead to further improvements.

First, we have considered only isotropic elastic media in this thesis and have ignored the effects of seismic anisotropy on the P and S waveforms. This can lead to velocity artifacts that can eventually be misinterpreted as compositional heterogeneities, especially in subduction zones (Bezada et al., 2016; Beller and Chevrot, 2020; VanderBeek and Faccenda, 2021). In principle, it is straightforward to extend the modeling of wave propagation in general anisotropic media with SEM solvers (e.g. Chevrot et al., 2004). The recent study of Beller and Chevrot (2020) extended isotropic FWI to anisotropic FWI with models parameterized in terms of density and the 21 elastic coefficients of the fourth-order elasticity tensor. With this approach, they were able to recover the orientation of the symmetry axis (including the dip) or stratification of seismic anisotropy, two notoriously difficult problems. To generalize the inversion method presented in Chapter 4, the challenge will be to define a priori model covariance matrices for the density and the 21 components of the elasticity tensors. In any case, future applications of FWI for imaging 3D anisotropic media will involve jointly inverting P and S teleseismic phases to fully recover seismic anisotropy (Beller and Chevrot, 2020). We expect these new tomographic models will also lead to less biased isotropic tomographic models.

Second, the initial model used in this study is set to the smoothed AK135 model. We saw in the Cascadia Chapter that the inverted model remains close to the initial model.

This is because the average travel time and amplitude residuals for each event have been removed. As a result, we can only image the relative perturbations to the input model, and the absolute velocity model is less well-constrained. This problem is particularly critical in tectonic settings such as the Central Andes, where the average 1-D model is far removed from the global average model. Using a 1D velocity model tailored to the study region as the initial model instead of the global average one should improve the inversion results. From a model space perspective, this should also speed up the convergence of the inversion. To better constrain the absolute model, we could also exploit additional and independent data, such as the dispersion of surface waves (e.g. Rawlinson and Fishwick, 2012; Schmandt and Lin, 2014) or use local averages of global tomographic models (e.g. Bastow, 2012).

Third, we have shown that the depth reconstruction of the density model is greatly improved when the correlations between density and $V_P$ and between density and $V_S$ are taken into account in the inversions. When no correlations are introduced, the density reconstruction is mainly restricted to the shallow part of the model (e.g. Beller et al., 2018), which implies that teleseismic waveforms are not very sensitive to deep density anomalies. Gravity data (Bouguer anomaly or geoid undulations) provide direct constraints on density, so we can consider including them into teleseismic FWI to better constrain the 3D density model. The joint inversion of classical traveltime data and gravity data is already an active field (e.g. Lelièvre et al., 2012), but applications to teleseismic FWI are yet to be made.

Fourth, when local earthquake data are available, in addition to teleseismic data, they should improve the coverage and, thus, potentially the spatial resolution of FWI. The pioneering FWI study by Gao et al. (2021) in the central Andes demonstrated the potential of local sources for high-resolution imaging of lithospheric structures. Furthermore, the frequency content of local earthquakes is higher (up to 10 Hz) than that of teleseismic events. Thus, local earthquakes can provide constraints on small-scale structural and attenuation models, although we need finer grids to model wave propagation at higher frequencies, with increased computational cost. A compromise might be first to constrain the large-scale

structures using teleseismic FWI, as we have done in this thesis, and then limit the computational domain to the depth of local earthquakes to refine the model. Joint inversion of local and teleseismic event travel times (e.g. Kuo-Chen et al., 2012) is another promising way to improve the resolution of the teleseismic FWI.

Finally, characterizing the resolution of tomographic models remains an important issue. Indeed, being able to calculate the a posteriori model covariance matrix could provide model uncertainties as well as the resolution of the inverted model.

**Real data applications**

In this thesis, we introduced for the first time SH waves in the teleseismic FWI and showed that this improves the resolution and robustness of the inverse models. Other types of shear waves, such as SV waves or SKS waves, should provide additional constraints on $V_S$ models (or seismic anisotropy). In general, the inclusion of additional seismic phases will increase the angular illumination of lithospheric targets which should further improve spatial resolution.

In addition to data from temporary experiments, there are dense permanent broadband stations with years of data available, for example, in the Vancouver region, Japan, or northern Taiwan, installed on top of active subduction zones. Future FWI applications on these dense permanent regional arrays are expected to provide high-resolution images and new insights into subduction systems.

# Bibliography


Adourian, S., Lyu, C., Masson, Y., Munch, F., and Romanowicz, B. Combining different 3-D global and regional seismic wave propagation solvers towards box tomography in the deep Earth. *Geophysical Journal International*, 232(2):1340–1356, oct 2022. doi: 10.1093/gji/ggac394.

Aki, K. and Lee, W. H. K. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model. *Journal of Geophysical Research*, 81(23):4381–4399, 1976. ISSN 0148-0227. doi: 10.1029/jb081i023p04381.

Aki, K., Christofferson, A., and Husebye, E. S. Determination of the three-dimensional seismic structures of the lithosphere. *J. Geophys. Res.*, 82:277–296, 1977.

Allmendinger, R. W., Jordan, T. E., Kay, S. M., and Isacks, B. L. THE EVOLUTION OF THE ALTIPLANO-PUNA PLATEAU OF THE CENTRAL ANDES. *Annual Review of Earth and Planetary Sciences*, 25(1):139–174, may 1997. doi: 10.1146/annurev.earth.25.1.139.

Alterman, Z. and Karal Jr, F. Propagation of Elastic Waves in layered media by finite difference methods. *Bulletin of the Seismological Society of America*, 59(1):471–471, 1968. ISSN 0037-1106. doi: 10.1785/bssa0590010471.

Amante, C. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, 2009.

Anderson, D. L. *Theory of the Earth*. Blachwell Sci. Publ., Osney Mead, Oxford, UK, 1989.

Antonijevic, S. K., Wagner, L. S., Kumar, A., Beck, S. L., Long, M. D., Zandt, G., Tavera, H., and Condori, C. The role of ridges in the formation and longevity of flat slabs. *Nature*, 524(7564):212–215, aug 2015. doi: 10.1038/nature14648.

Araujo, S., Valette, B., Potin, B., and Ruiz, M. A preliminary seismic travel time tomography beneath Ecuador from data of the national network. *Journal of South American Earth Sciences*, 111:103486, 2021. ISSN 0895-9811. doi: 10.1016/j.jsames.2021.103486.

Artemieva, I. *The Lithosphere*. Cambridge University Press, 2011. ISBN 9780521605120. doi: 10.1017/cbo9780511975417.

Audet, P. and Bürgmann, R. Possible control of subduction zone slow-earthquake periodicity by silica enrichment. *Nature*, 510(7505):389–392, jun 2014. doi: 10.1038/nature13391.

Audet, P., Bostock, M. G., Christensen, N. I., and Peacock, S. M. Seismic evidence for overpressured subducted oceanic crust and megathrust fault sealing. *Nature*, 457:76–78, 2009.

Audet, P., Bostock, M. G., Boyarko, D. C., Brudzinski, M. R., and Allen, R. M. Slab morphology in the Cascadia fore arc and its relation to episodic tremor and slip. *Journal of Geophysical Research*, 115, apr 2010. doi: 10.1029/2008jb006053.

Babuška, V. and Cara, M. *Seismic Anisotropy of the Earth*. Kluwer Acad., Norwell, Mass., 1991.

Baby, P., Rochat, P., Mascle, G., and Hérail, G. Neogene shortening contribution to crustal thickening in the back arc of the Central Andes. *Geology*, 25(10):883, 1997. doi: 10.1130/0091-7613(1997)025<0883:nsctct>2.3.co;2.

Barazangi, M. and Isacks, B. L. Spatial distribution of earthquakes and subduction of the Nazca plate beneath South America. *Geology*, 4(11):686, 1976. doi: 10.1130/0091-7613(1976)4<686:sdoeas>2.0.co;2.

Bastow, I. D. Relative arrival-time upper-mantle tomography and the elusive background mean. *Geophysical Journal International*, 190(2):1271–1278, 2012. ISSN 0956-540X. doi: 10.1111/j.1365-246x.2012.05559.x.

Baumont, D., Paul, A., Zandt, G., Beck, S. L., and Pedersen, H. Lithospheric structure of the central Andes based on surface wave dispersion. *Journal of Geophysical Research: Solid Earth*, 107(B12):ESE 18–1–ESE 18–13, dec 2002. doi: 10.1029/2001jb000345.

Beck, S., Zandt, G., and Wagner, L. Central andean uplift and the geodynamics of the high topography, 2010. URL `https://www.fdsn.org/networks/detail/ZG_2010/`.

Beck, S. L. and Zandt, G. The nature of orogenic crust in the central Andes. *Journal of Geophysical Research: Solid Earth*, 107(B10):ESE 7–1–ESE 7–16, oct 2002. doi: 10.1029/2000jb000124.

Becker, T. W. and O'Connell, R. J. Predicting plate velocities with mantle circulation models. *Geochemistry, Geophysics, Geosystems*, 2(12):2001GC000171, dec 2001. doi: 10.1029/2001gc000171.

Beller, S. *Imagerie lithosphérique par inversion de formes d'ondes télésismiques – Application aux Alpes Occidentales*. PhD thesis, Université Côte d'Azur, 2017.

Beller, S. and Chevrot, S. Probing depth and lateral variations of upper-mantle seismic anisotropy from full-waveform inversion of teleseismic body-waves. *Geophys. J. Int.*, 222: 352–387, 2020.

Beller, S., Monteiller, V., Operto, S., Nolet, G., Paul, A., and Zhao, L. Lithospheric architecture of the South-Western Alps revealed by multiparameter teleseismic full-waveform inversion. *Geophys. J. Int.*, 212(2):1369–1388, 2018.

Berenger, J.-P. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994. ISSN 0021-9991. doi: 10.1006/jcph.1994.1159.

Bezada, M. J., Faccenda, M., and Toomey, D. R. Representing anisotropic subduction zones with isotropic velocity models: A characterization of the problem and some steps on a possible path forward. *Geochemistry, Geophysics, Geosystems*, 17(8):3164–3189, 2016. ISSN 1525-2027. doi: 10.1002/2016gc006507.

Bielak, J. and Christiano, P. On the effective seismic input for non-linear soil-structure interaction systems. *Earthquake Engineering & Structural Dynamics*, 12(1):107–119, 1984. ISSN 1096-9845. doi: 10.1002/eqe.4290120108.

Birch, F. The velocity of compressional waves in rocks to 10 kilobars: 1. *Journal of Geophysical Research*, 65(4):1083–1102, 1960. ISSN 0148-0227. doi: 10.1029/jz065i004p01083.

Birch, F. The velocity of compressional waves in rocks to 10 kilobars, part 2. *J. Geophys. Res.*, 66(7):2199–2224, 1961.

Bishop, B. T., Beck, S. L., Zandt, G., Wagner, L., Long, M., Antonijevic, S. K., Kumar, A., and Tavera, H. Causes and consequences of flat-slab subduction in southern Peru. *Geosphere*, 13(5):1392–1407, jul 2017. doi: 10.1130/ges01440.1.

Blakely, R. J., Brocher, T. M., and Wells, R. E. Subduction-zone magnetic anomalies and implications for hydrated forearc mantle. *Geology*, 33(6):445, 2005. doi: 10.1130/g21447.1.

Bolton, H. and Masters, G. Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle. *Journal of Geophysical Research: Solid Earth*, 106(B7):13527–13540, jul 2001. doi: 10.1029/2000jb900378.

Bostock, M. The Moho in subduction zones. *Tectonophysics*, 609:547–557, dec 2013. doi: 10.1016/j.tecto.2012.07.007.

Bostock, M. G., Hyndman, R. D., Rondenay, S., and Peacock, S. M. An inverted continental Moho and serpentinization of the forearc mantle. *Nature*, 417:536–538, 2002.

Boxberg, M. S., Prévost, J. H., and Tromp, J. Wave Propagation in Porous Media Saturated with Two Fluids. *Transport in Porous Media*, 107(1):49–63, 2015. ISSN 0169-3913. doi: 10.1007/s11242-014-0424-2.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D. Global adjoint tomography: first-generation model. *Geophysical Journal International*, 207(3):1739–1766, sep 2016. doi: 10.1093/gji/ggw356.

Breeding, C. M. and Ague, J. J. Slab-derived fluids and quartz-vein formation in an accretionary prism, Otago Schist, New Zealand. *Geology*, 30(6):499, 2002. doi: 10.1130/0091-7613(2002)030<0499:sdfaqv>2.0.co;2.

Brocher, T. Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bull. Seismol. Soc. Am.*, 95(6):2081–2092, 2005a.

Brocher, T. M. Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America*, 95(6):2081–2092, dec 2005b. doi: 10.1785/0120050077.

Broggini, F., Vasmel, M., Robertsson, J. O. A., and van Manen, D.-J. Immersive boundary conditions: Theory, implementation, and examples. *GEOPHYSICS*, 82(3):T97–T110, may 2017. doi: 10.1190/geo2016-0458.1.

Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G. Multiscale seismic waveform inversion. *GEOPHYSICS*, 60(5):1457–1473, 9 1995. ISSN 0016-8033. doi: 10.1190/1.1443880.

Cammarano, F., Goes, S., Vacher, P., and Giardini, D. Inferring upper-mantle temperatures from seismic velocities. *Physics of the Earth and Planetary Interiors*, 138(3-4):197–222, aug 2003. doi: 10.1016/s0031-9201(03)00156-0.

Capdeville, Y., Chaljub, E., Vilotte, J. P., and Montagner, J. P. Coupling the spectral element method with a modal solution for elastic wave propgation in global Earth models. *Geophysical Journal International*, 152:34–66, 2003.

Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *GEOPHYSICS*, 50(4):705–708, 1985. ISSN 0016-8033. doi: 10.1190/1.1441945.

Cerveny, V. *Seismic Ray Theory*. Cambridge University Press, 2001. ISBN 9780521018227. doi: 10.1017/cbo9780511529399.

Chen, P., Jordan, T. H., and Zhao, L. Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, 170(1):175–181, 2007. ISSN 0956-540X. doi: 10.1111/j.1365-246x.2007.03429.x.

Chevrot, S. and van der Hilst, R. The poisson ratio of the Australian crust: Geological and geophysical implications. *Earth Planet. Sci. Lett.*, 183:121–132, 2000.

Chevrot, S., Favier, N., and Komatitsch, D. Shear wave splitting in three-dimensional anisotropic media. *Geophysical Journal International*, 159(2):711–720, nov 2004. doi: 10.1111/j.1365-246x.2004.02432.x.

Chevrot, S., Sylvander, M., Diaz, J., Ruiz, M., and Paul, A. The Pyrenean architecture as revealed by teleseismic P-to-S converted waves recorded along two dense transects. *Geophysical Journal International*, 200(2):1094–1105, 2015. ISSN 0956-540X. doi: 10.1093/gji/ggu400.

Chevrot, S., Sylvander, M., Villaseñor, A., Díaz, J., Stehly, L., Boué, P., Monteiller, V., Martin, R., Lehujeur, M., Beller, S., Brives, J., Bitri, A., Calassou, S., Collin, M., Ford, M., Jolivet, L., Manatschal, G., Masini, E., Mouthereau, F., and Vidal, O. Passive imaging of collisional orogens: a review of a decade of geophysical studies in the Pyrénées. *BSGF - Earth Sciences Bulletin*, 193:1, 2022. doi: 10.1051/bsgf/2021049.

Chouet, B. A. and Matoza, R. S. A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *Journal of Volcanology and Geothermal Research*, 252:108–175, feb 2013. doi: 10.1016/j.jvolgeores.2012.11.013.

Christensen, N. I. Elasticity of ultrabasic rocks. *Journal of Geophysical Research*, 71(24): 5921–5931, dec 1966. doi: 10.1029/jz071i024p05921.

Christensen, N. I. Poisson's ratio and crustal seismology. *Journal of Geophysical Research: Solid Earth*, 101(B2):3139–3156, 1996. ISSN 0148-0227. doi: 10.1029/95jb03446.

Christensen, N. I. Serpentinites, peridotites, and seismology. *Int. Geol. Rev.*, 49(9):795–816, 2004.

Christensen, N. I. and Mooney, W. D. Seismic velocity structure and composition of the continental crust : A global view. *J. Geophys. Res.*, 100:9761–9788, 1995.

Ciardelli, C., Assumpção, M., Bozdağ, E., and Lee, S. Adjoint Waveform Tomography of South America. *Journal of Geophysical Research: Solid Earth*, 127(2), 2022. ISSN 2169-9313. doi: 10.1029/2021jb022575.

Claerbout, J. F. TOWARD A UNIFIED THEORY OF REFLECTOR MAPPING. *GEOPHYSICS*, 36(3):467–481, 1971. ISSN 0016-8033. doi: 10.1190/1.1440185.

Clayton, R. Peru subduction experiment. caltech. dataset., 2013.

Clayton, R. and Engquist, B. Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the Seismological Society of America*, 67(6), 1977. ISSN 0037-1106. doi: 10.1785/bssa0670061529.

Cupillard, P., Delavaud, E., Burgos, G., Festa, G., Vilotte, J., Capdeville, Y., and Montagner, J. RegSEM: a versatile code based on the spectral element method to compute seismic wave propagation at the regional scale. *Geophysical Journal International*, 188(3): 1203–1220, 2012. ISSN 0956-540X. doi: 10.1111/j.1365-246x.2011.05311.x.

Dahlen, F. A., Hung, S. H., and Nolet, G. Fréchet kernels for finite-frequency traveltimes-I. Theory. *Geophysical Journal International*, 141(1):157–174, 2000. ISSN 0956-540X. doi: 10.1046/j.1365-246X.2000.00070.x.

Delph, J. R., Levander, A., and Niu, F. Fluid Controls on the Heterogeneous Seismic Characteristics of the Cascadia Margin. *Geophysical Research Letters*, 45(20):11,021–11,029, 2018. ISSN 0094-8276. doi: 10.1029/2018gl079518.

DeMets, C., Gordon, R. G., Argus, D. F., and Stein, S. Current plate motions. *Geophys. J. Int.*, 101:425–478, 1990.

Dougherty, S. L. and Clayton, R. W. Seismic structure in southern peru: evidence for a smooth contortion between flat and normal subduction of the Nazca plate. *Geophysical Journal International*, 200(1):534–555, nov 2014. doi: 10.1093/gji/ggu415.

Dunham, A. and Kiser, E. Local earthquake tomography of the central Oregon forearc using a large-N, short duration, nodal array. *Earth and Planetary Science Letters*, 551:116559, dec 2020. doi: 10.1016/j.epsl.2020.116559.

Dziewonski, A. M. and Anderson, D. L. Preliminary reference Earth model. *Phys. Earth. Planet. Inter.*, 25:297–356, 1981.

Espurt, N., Funiciello, F., Martinod, J., Guillaume, B., Regard, V., Faccenna, C., and Brusset, S. Flat subduction dynamics and deformation of the South American plate: Insights from analog modeling. *Tectonics*, 27(3):n/a–n/a, jun 2008. doi: 10.1029/2007tc002175.

Evans, R. L., Wannamaker, P. E., McGary, R. S., and Elsenbeck, J. Electrical structure of the central Cascadia subduction zone: The EMSLAB Lincoln Line revisited. *Earth and Planetary Science Letters*, 402:265–274, sep 2014. doi: 10.1016/j.epsl.2013.04.021.

Faccenna, C., Becker, T. W., Holt, A. F., and Brun, J. P. Mountain building, mantle convection, and supercontinents: revisited. *Earth and Planetary Science Letters*, 564: 116905, jun 2021. doi: 10.1016/j.epsl.2021.116905.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725, 2009. ISSN 1365-246X. doi: 10.1111/j.1365-246x.2009.04368.x.

Gao, Y., Tilmann, F., Herwaarden, D.-P., Thrastarson, S., Fichtner, A., Heit, B., Yuan, X., and Schurr, B. Full Waveform Inversion Beneath the Central Andes: Insight Into the Dehydration of the Nazca Slab and Delamination of the Back-Arc Lithosphere. *Journal of Geophysical Research: Solid Earth*, 126(7), jun 2021. doi: 10.1029/2021jb021984.

Garzione, C. N., Molnar, P., Libarkin, J. C., and MacFadden, B. J. Rapid late Miocene rise of the Bolivian Altiplano: Evidence for removal of mantle lithosphere. *Earth and Planetary Science Letters*, 241(3-4):543–556, jan 2006. doi: 10.1016/j.epsl.2005.11.026.

Garzione, C. N., Hoke, G. D., Libarkin, J. C., Withers, S., MacFadden, B., Eiler, J., Ghosh, P., and Mulch, A. Rise of the Andes. *Science*, 320(5881):1304–1307, jun 2008. doi: 10.1126/science.1148615.

Garzione, C. N., McQuarrie, N., Perez, N. D., Ehlers, T. A., Beck, S. L., Kar, N., Eichelberger, N., Chapman, A. D., Ward, K. M., Ducea, M. N., Lease, R. O., Poulsen, C. J., Wagner, L. S., Saylor, J. E., Zandt, G., and Horton, B. K. Tectonic Evolution of the Central Andean Plateau and Implications for the Growth of Plateaus. *Annual Review of Earth and Planetary Sciences*, 45(1):529–559, aug 2017. doi: 10.1146/annurev-earth-063016-020612.

Gauthier, O., Virieux, J., and Tarantola, A. Two-dimensional nonlinear inversion of seismic waveforms; numerical results. *Geophysics*, 51(7):1387–1403, 1986. ISSN 0016-8033. doi: 10.1190/1.1442188.

Geller, R. J. and Ohminato, T. 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, 1994. ISSN 1365-246X. doi: 10.1111/j.1365-246x.1994.tb01807.x.

Gerya, T. V., Fossati, D., Cantieni, C., and Seward, D. Dynamic effects of aseismic ridge subduction: Numerical modelling. *European Journal of Mineralogy*, 21(3):649–661, jun 2009.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, J. Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? Part 1: Sensitivity and trade-off analysis. *Geophysics*, 78, 2013. doi: 10.1190/geo2012-0204.1.

Gosselin, J. M., Audet, P., Estève, C., McLellan, M., Mosher, S. G., and Schaeffer, A. J. Seismic evidence for megathrust fault-valve behavior during episodic tremor and slip. *Science Advances*, 6(4), jan 2020. doi: 10.1126/sciadv.aay5174.

Graeber, F. M. and Asch, G. Three-dimensional models of p wave velocity and p-to-s velocity ratio in the southern central andes by simultaneous inversion of local earthquake data. *Journal of Geophysical Research: Solid Earth*, 104(B9):20237–20256, sep 1999. doi: 10.1029/1999jb900037.

Guillaume, B., Martinod, J., and Espurt, N. Variations of slab dip and overriding plate tectonics during subduction: Insights from analogue modelling. *Tectonophysics*, 463(1-4): 167–174, jan 2009. doi: 10.1016/j.tecto.2008.09.043.

Guo, H., McGuire, J. J., and Zhang, H. Correlation of porosity variations and rheological transitions on the southern Cascadia megathrust. *Nature Geoscience*, 14(5):341–348, may 2021. doi: 10.1038/s41561-021-00740-1.

Gutscher, M.-A., Olivet, J.-L., Aslanian, D., Eissen, J.-P., and Maury, R. The "lost Inca plateau": cause of flat subduction beneath Peru? *Earth and Planetary Science Letters*, 171(3):335–341, sep 1999. doi: 10.1016/s0012-821x(99)00153-3.

Gutscher, M.-A., Spakman, W., Bijwaard, H., and Engdahl, E. R. Geodynamics of flat subduction: Seismicity and tomographic constraints from the Andean margin. *Tectonics*, 19(5):814–833, oct 2000. doi: 10.1029/1999tc001152.

Halpaap, F., Rondenay, S., and Ottemöller, L. Seismicity, deformation, and metamorphism in the western Hellenic subduction zone: New constraints from tomography. *Journal of Geophysical Research: Solid Earth*, 123(4):3000–3026, apr 2018. doi: 10.1002/2017jb015154.

Hansen, P. C. Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*, 34(4):561–580, 1992. ISSN 0036-1445. doi: 10.1137/1034115.

Hansen, R. T., Bostock, M. G., and Christensen, N. I. Nature of the low velocity zone in Cascadia from receiver function waveform inversion. *Earth and Planetary Science Letters*, 337-338:25–38, jul 2012. doi: 10.1016/j.epsl.2012.05.031.

Haskell, N. A. The dispersion of surface waves on multilayered media. *Bulletin of the Seismological Society of America*, 43(1):17–34, 1953. ISSN 0037-1106. doi: 10.1785/bssa0430010017.

Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., and Smoczyk, G. M. Slab2, a comprehensive subduction zone geometry model. *Science*, 362 (6410):58–61, oct 2018. doi: 10.1126/science.aat4723.

Hosseini, K., Sigloch, K., Tsekhmistrenko, M., Zaheri, A., Nissen-Meyer, T., and Igel, H. Global mantle structure from multifrequency tomography using P, PP and P-diffracted waves. *Geophysical Journal International*, 220(1):96–141, 09 2019. ISSN 0956-540X. doi: 10.1093/gji/ggz394.

Hudson, J. A. Scattered waves in the coda of P. *Journal of Geophysics*, 1(43):359–374, 1977.

Hung, S., Dahlen, F., and Nolet, G. Fréchet kernels for finite-frequency traveltimes—II. Examples. *Geophysical Journal International*, 141(1):175–203, 2000. ISSN 0956-540X. doi: 10.1046/j.1365-246x.2000.00072.x.

Husson, L. and Sempere, T. Thickening the Altiplano crust by gravity-driven crustal channel flow. *Geophysical Research Letters*, 30(5):n/a–n/a, mar 2003. doi: 10.1029/2002gl016877.

Hyndman, R. D. Downdip landward limit of Cascadia great earthquake rupture. *Journal of Geophysical Research: Solid Earth*, 118(10):5530–5549, oct 2013. doi: 10.1002/jgrb.50390.

Hyndman, R. D. and Peacock, S. M. Serpentinization of the forearc mantle. *Earth and Planetary Science Letters*, 212(3-4):417–432, jul 2003. doi: 10.1016/s0012-821x(03)00263-2.

Hyndman, R. D., McCrory, P. A., Wech, A., Kao, H., and Ague, J. Cascadia subducting plate fluids channelled to fore-arc mantle corner: ETS and silica deposition. *Journal of Geophysical Research: Solid Earth*, 120(6):4344–4358, jun 2015. doi: 10.1002/2015jb011920.

Isacks, B. L. Uplift of the Central Andean Plateau and bending of the Bolivian Orocline. *Journal of Geophysical Research*, 93(B4):3211, 1988. doi: 10.1029/jb093ib04p03211.

James, D. E. and Sacks, I. S. Cenozoic formation of the central andes: A geophysical perspective. In *Geology and Ore Deposits of the Central Andes*. Society of Economic Geologists, 1999. doi: 10.5382/sp.07.01.

James, D. E., Fouch, M. J., Carlson, R. W., and Roth, J. B. Slab fragmentation, edge flow and the origin of the Yellowstone hotspot track. *Earth and Planetary Science Letters*, 311 (1-2):124–135, nov 2011. doi: 10.1016/j.epsl.2011.09.007.

Kan, L.-Y., Chevrot, S., and Monteiller, V. A consistent multiparameter Bayesian full waveform inversion scheme for imaging heterogeneous isotropic elastic media. *Geophysical Journal International*, 232(2):864–883, 2022. ISSN 0956-540X. doi: 10.1093/gji/ggac363.

Kawai, K., Takeuchi, N., and Geller, R. J. Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media. *Geophysical Journal International*, 164(2):411–424, 2006. ISSN 1365-246X. doi: 10.1111/j.1365-246x.2005.02829.x.

Kennett, B. L. N., Engdahl, E. R., and Buland, R. Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.*, 122:108–124, 1995.

Kerrich, R. Nature's gold factory. *Science*, 284(5423):2101–2102, jun 1999. doi: 10.1126/science.284.5423.2101.

Kind, R., Yuan, X., Saul, J., Nelson, D., Sobolev, S. V., Mechie, J., Zhao, W., Kosarev, G., Ni, J., Achauer, U., and Jiang, M. Seismic Images of Crust and Upper Mantle Beneath Tibet: Evidence for Eurasian Plate Subduction. *Science*, 298(5596):1219–1221, 2002. ISSN 0036-8075. doi: 10.1126/science.1078115.

Komatitsch, D. and Martin, R. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *GEOPHYSICS*, 72(5):SM155–SM167, 2007. ISSN 0016-8033. doi: 10.1190/1.2757586.

Komatitsch, D. and Tromp, J. Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, 139(3):806–822, 1999. ISSN 1365-246X. doi: 10.1046/j.1365-246x.1999.00967.x.

Komatitsch, D. and Tromp, J. Spectral-element simulations of global seismic wave propagation—I. Validation. *Geophysical Journal International*, 149(2):390–412, 2002. ISSN 0956-540X. doi: 10.1046/j.1365-246x.2002.01653.x.

Komatitsch, D. and Vilotte, J. P. 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, 1998.

Komatitsch, D., Tromp, J., and Vilotte, J. The spectral element method for elastic wave equations: Application to 2D and 3D seismic problems. *SEG Technical Program Expanded Abstracts 1998*, pages 1460–1463, 1998. doi: 10.1190/1.1820185.

Komatitsch, D., Barnes, C., and Tromp, J. Simulation of anisotropic wave propagation based upon a spectral element methodSEM for Anisotropic Wave Propagation. *Geophysics*, 65 (4):1251–1260, 2000. ISSN 0016-8033. doi: 10.1190/1.1444816.

Koppers, A. A. P., Becker, T. W., Jackson, M. G., Konrad, K., Müller, R. D., Romanowicz, B., Steinberger, B., and Whittaker, J. M. Mantle plumes and their role in Earth processes. *Nature Reviews Earth & Environment*, 2(6):382–401, 2021. doi: 10.1038/s43017-021-00168-6.

Kuo-Chen, H., Wu, F. T., and Roecker, S. W. Three-dimensional P velocity structures of the lithosphere beneath Taiwan from the analysis of TAIGER and related seismic data sets. *Journal of Geophysical Research: Solid Earth*, 117(B6):n/a–n/a, 2012. ISSN 0148-0227. doi: 10.1029/2011jb009108.

Lailly, P. The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering. *Conference on Inverse Scattering : Theory and Applications*, page 206–220, 1983.

Lee, C. T. A. Compositional variation of density and seismic velocities in natural peridotites at STP conditions: Implications for seismic imaging of compositional heterogeneities in the upper mantle. *J. Geophys. Res.*, 108 (B9)(B9):Art No. 2441, doi:10.1029/2003JB002413, 2003.

Lee, E., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A. M., and Beroza, G. C. Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research: Solid Earth*, 119(8):6421–6451, 2014. ISSN 2169-9313. doi: 10.1002/2014jb011346.

Lees, J. M. and VanDecar, J. C. Seismic tomography constrained by Bouguer gravity anomalies: Applications in western Washington. *Pure and Applied Geophysics PAGEOPH*, 135 (1):31–52, jan 1991. doi: 10.1007/bf00877007.

Lei, W., Ruan, Y., Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D. Global adjoint tomography—model GLAD-M25. *Geophysical Journal International*, 223(1):1–21, 2020. ISSN 0956-540X. doi: 10.1093/gji/ggaa253.

Lelièvre, P. G., Farquharson, C. G., and Hurich, C. A. Joint inversion of seismic traveltimes and gravity data on unstructured grids with application to mineral explorationJoint inversion of traveltimes and gravity. *Geophysics*, 77(1):K1–K15, 2012. ISSN 0016-8033. doi: 10.1190/geo2011-0154.1.

Leng, K., Nissen-Meyer, T., and van Driel, M. Efficient global wave propagation adapted to 3-d structural complexity: a pseudospectral/spectral-element approach. *Geophysical Journal International*, 207(3):1700–1721, sep 2016. doi: 10.1093/gji/ggw363.

Leng, K., Nissen-Meyer, T., van Driel, M., Hosseini, K., and Al-Attar, D. AxiSEM3d: broad-band seismic wavefields in 3-d global earth models with undulating discontinuities. *Geophysical Journal International*, 217(3):2125–2146, feb 2019. doi: 10.1093/gji/ggz092.

Li, X.-Q. and Nábělek, J. L. Deconvolution of teleseismic body waves for enhancing structure beneath a seismometer array. *Bulletin of the Seismological Society of America*, 89(1):190–201, 1999. ISSN 0037-1106. doi: 10.1785/bssa0890010190.

Ligorria, J. P. and Ammon, C. J. Iterative deconvolution and receiver-function estimation. *Bull. Seism. Soc. Am.*, 89:1395–1400, 1999.

Lines, L. R., Schultz, A. K., and Treitel, S. Cooperative inversion of geophysical data. *GEOPHYSICS*, 53(1):8–20, jan 1988. doi: 10.1190/1.1442403.

Liu, Q. and Gu, Y. Seismic imaging: From classical to adjoint tomography. *Tectonophysics*, 566:31–66, 2012. ISSN 0040-1951. doi: 10.1016/j.tecto.2012.07.006.

Luo, Y. and Schuster, G. T. Wave-equation traveltime inversion. *Geophysics*, 56(5):645–653, 1991. ISSN 0016-8033. doi: 10.1190/1.1443081.

Lyon-Caen, H., Molnar, P., and Suárez, G. Gravity anomalies and flexure of the Brazilian Shield beneath the Bolivian Andes. *Earth and Planetary Science Letters*, 75(1):81–92, sep 1985. doi: 10.1016/0012-821x(85)90053-6.

Lyu, C., Capdeville, Y., and Zhao, L. Efficiency of the spectral element method with very high polynomial degree to solve the elastic wave equation. *Geophysics*, 85(1):T33–T43, 2020.

Ma, Y. and Clayton, R. W. The crust and uppermost mantle structure of southern peru from ambient noise and earthquake surface wave analysis. *Earth and Planetary Science Letters*, 395:61–70, jun 2014. doi: 10.1016/j.epsl.2014.03.013.

Ma, Y. and Clayton, R. W. Flat slab deformation caused by interplate suction force. *Geophysical Research Letters*, 42(17):7064–7072, sep 2015. doi: 10.1002/2015gl065195.

Ma, Y., Clayton, R. W., Tsai, V. C., and Zhan, Z. Locating a scatterer in the active volcanic area of Southern Peru from ambient noise cross-correlation. *Geophysical Journal International*, 192(3):1332–1341, 2013. ISSN 0956-540X. doi: 10.1093/gji/ggs103.

Mamani, M., Worner, G., and Sempere, T. Geochemical variations in igneous rocks of the central andean orocline (13 s to 18 s): Tracing crustal thickening and magma generation through time and space. *Geological Society of America Bulletin*, 122(1-2):162–182, sep 2009. doi: 10.1130/b26538.1.

Manea, V. C., Pérez-Gussinyé, M., and Manea, M. Chilean flat slab subduction controlled by overriding plate thickness and trench rollback. *Geology*, 40(1):35–38, jan 2012. doi: 10.1130/g32543.1.

Manning, C. E. The solubility of quartz in H2O in the lower crust and upper mantle. *Geochimica et Cosmochimica Acta*, 58(22):4831–4839, nov 1994. doi: 10.1016/0016-7037(94)90214-3.

Martin, R., Giraud, J., Ogarko, V., Chevrot, S., Beller, S., Gégout, P., and Jessell, M. Three-dimensional gravity anomaly data inversion in the Pyrenees using compressional seismic velocity model as structural similarity constraints. *Geophysical Journal International*, 225 (2):1063–1085, sep 2021. doi: 10.1093/gji/ggaa414.

Martinod, J., Funiciello, F., Faccenna, C., Labanieh, S., and Regard, V. Dynamical effects of subducting ridges: insights from 3-D laboratory models. *Geophysical Journal International*, 163(3):1137–1150, dec 2005. doi: 10.1111/j.1365-246x.2005.02797.x.

Martinod, J., Gérault, M., Husson, L., and Regard, V. Widening of the andes: An interplay between subduction dynamics and crustal wedge tectonics. *Earth-Science Reviews*, 204: 103170, may 2020. doi: 10.1016/j.earscirev.2020.103170.

Masson, Y. and Romanowicz, B. Box tomography: localized imaging of remote targets buried in an unknown medium, a step forward for understanding key structures in the deep Earth. *Geophysical Journal International*, 211(1):141–163, 2017. ISSN 0956-540X. doi: 10.1093/gji/ggx141. URL `https://dx.doi.org/10.1093/gji/ggx141`.

Masson, Y., Cupillard, P., Capdeville, Y., and Romanowicz, B. On the numerical implementation of time-reversal mirrors for tomographic imaging. *Geophysical Journal International*, 196(3):1580–1599, 2014. ISSN 0956-540X. doi: 10.1093/gji/ggt459. URL `https://hal.archives-ouvertes.fr/hal-01303023`.

McCrory, P. A., Blair, J. L., Waldhauser, F., and Oppenheimer, D. H. Juan de Fuca slab geometry and its relation to Wadati-Benioff zone seismicity. *Journal of Geophysical Research: Solid Earth*, 117(B9), sep 2012. doi: 10.1029/2012jb009407.

Métivier, L., Brossier, R., Virieux, J., and Operto, S. Full Waveform Inversion and the Truncated Newton Method. *SIAM Journal on Scientific Computing*, 35(2):B401–B437, jan 2013. doi: 10.1137/120877854.

Monteiller, V., Chevrot, S., Komatitsch, D., and Fuji, N. A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-D regional model. *Geophys. J. Int.*, 192:230–247, 2013.

Monteiller, V., Chevrot, S., Komatitsch, D., and Wang, Y. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM-DSM hybrid method. *Geophys. J. Int.*, 202:811–827, 2015.

Monteiller, V., Beller, S., Plazolles, B., and Chevrot, S. On the validity of the planar wave approximation to compute synthetic seismograms of teleseismic body waves in a 3-D regional model. *Geophys. J. Int.*, 224:2060–2076, 2021. doi: 10.1093/gji/ggaa570.

Mora, P. Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics*, 52(9):1211–1228, 1987. ISSN 0016-8033. doi: 10.1190/1.1442384.

Nafe, J. E. and Drake, C. L. Variation with depth in shallow and deep water marine sediments of porosity, density and the velocities of compressional and shear waves. *GEOPHYSICS*, 22(3):523–552, jul 1957. doi: 10.1190/1.1438386.

Nakajima, J., Matsuzawa, T., Hasegawa, A., and Zhao, D. Three-dimensional structure of $v_p$, $v_s$, and $v_p/v_s$ beneath northeastern Japan: Implications for arc magmatism and fluids. *Journal of Geophysical Research: Solid Earth*, 106(B10):21843–21857, oct 2001. doi: 10.1029/2000jb000008.

Nataf, H., Nakanishi, I., and Anderson, D. L. Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy: 3. Inversion. *Journal of Geophysical Research: Solid Earth*, 91(B7):7261–7307, 1986. ISSN 0148-0227. doi: 10.1029/jb091ib07p07261.

Nissen-Meyer, T., Driel, M. v., Stähler, S. C., Hosseini, K., Hempel, S., Auer, L., Colombi, A., and Fournier, A. AxiSEM: broadband 3-D seismic wavefields in axisymmetric media. *Solid Earth*, 5(1):425–445, 2014. doi: 10.5194/se-5-425-2014.

Nissen-Meyer, T., Fournier, A., and Dahlen, F. A. A two-dimensional spectral-element method for computing spherical-earth seismograms – I. Moment-tensor source. *Geophysical Journal International*, 168(3):1067–1092, 2007. ISSN 0956-540X. doi: 10.1111/j.1365-246x.2006.03121.x.

Nocedal, J. and Wright, S. J. *Numerical Optimization*. Springer, 2nd edition edition, 2006.

Nolet, G. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*. Cambridge University Press, 2008. doi: 10.1017/CBO9780511984709.

Nolet, G. and Dahlen, F. A. Wave front healing and the evolution of seismic delay times. *Journal of Geophysical Research: Solid Earth*, 105(B8):19043–19054, 8 2000. ISSN 0148-0227. doi: 10.1029/2000JB900161.

Obrebski, M., Allen, R. M., Xue, M., and Hung, S.-H. Slab-plume interaction beneath the Pacific Northwest. *Geophysical Research Letters*, 37(14):n/a–n/a, jul 2010. doi: 10.1029/2010gl043489.

Obrebski, M., Allen, R. M., Pollitz, F., and Hung, S.-H. Lithosphere-asthenosphere interaction beneath the western united states from the joint inversion of body-wave traveltimes and surface-wave phase velocities. *Geophysical Journal International*, 185(2):1003–1021, mar 2011. doi: 10.1111/j.1365-246x.2011.04990.x.

O'Driscoll, L. J., Humphreys, E. D., and Saucier, F. Subduction adjacent to deep continental roots: Enhanced negative pressure in the mantle wedge, mountain building and continental motion. *Earth and Planetary Science Letters*, 280(1-4):61–70, apr 2009. doi: 10.1016/j.epsl.2009.01.020.

Oleskevich, D. A., Hyndman, R. D., and Wang, K. The updip and downdip limits to great subduction earthquakes: Thermal and structural models of Cascadia, south Alaska, SW Japan, and Chile. *Journal of Geophysical Research: Solid Earth*, 104(B7):14965–14991, jul 1999. doi: 10.1029/1999jb900060.

Oliver, D. S. Calculation of the Inverse of the Covariance. *Mathematical Geology*, 30(7):911–933, 1998. ISSN 0882-8121. doi: 10.1023/a:1021734811230.

Operto, S., Gholami, S., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice. *The Leading Edge*, 32:1040–1054, 2013.

Peacock, S. A. Fluid processes in subduction zones. *Science*, 248(4953):329–337, apr 1990. doi: 10.1126/science.248.4953.329.

Peacock, S. M. Large-scale hydration of the lithosphere above subducting slabs. *Chemical Geology*, 108(1-4):49–59, aug 1993. doi: 10.1016/0009-2541(93)90317-c.

Peacock, S. M. and Hyndman, R. D. Hydrous minerals in the mantle wedge and the maximum depth of subduction thrust earthquakes. *Geophysical Research Letters*, 26(16):2517–2520, aug 1999. doi: 10.1029/1999gl900558.

Peacock, S. M. and Wang, K. Seismic consequences of warm versus cool subduction metamorphism: Examples from southwest and northeast japan. *Science*, 286(5441):937–939, oct 1999. doi: 10.1126/science.286.5441.937.

Peacock, S. M., Christensen, N. I., Bostock, M. G., and Audet, P. High pore pressures and porosity at 35 km depth in the Cascadia subduction zone. *Geology*, 39(5):471–474, 2011. ISSN 0091-7613. doi: 10.1130/g31649.1.

Phillips, K. and Clayton, R. W. Structure of the subduction transition region from seismic array data in southern Peru. *Geophys. J. Int.*, 196(3):1889–1905, 2014.

Phillips, K., Clayton, R. W., Davis, P., Tavera, H., Guy, R., Skinner, S., Stubailo, I., Audin, L., and Aguilar, V. Structure of the subduction system in southern Peru from seismic array data. *J. Geophys. Res.*, 117:doi:10.1029/2012JB009540, 2012.

Pichon, X. L. Sea-floor spreading and continental drift. *Journal of Geophysical Research*, 73 (12):3661–3697, 1968. ISSN 0148-0227. doi: 10.1029/jb073i012p03661.

Pichon, X. L., Henry, P., and Goffé, B. Uplift of Tibet: from eclogites to granulites — implications for the Andean Plateau and the Variscan belt. *Tectonophysics*, 273(1-2): 57–76, may 1997. doi: 10.1016/s0040-1951(96)00288-0.

Pienkowska, M., Monteiller, V., and Nissen-Meyer, T. High-frequency global wavefields for local 3-D structures by wavefield injection and extrapolation. *Geophysical Journal International*, 225(3):1782–1798, 11 2020. ISSN 0956-540X. doi: 10.1093/gji/ggaa563. URL `https://doi.org/10.1093/gji/ggaa563`.

Portner, D. E., Beck, S., Zandt, G., and Scire, A. The nature of subslab slow velocity anomalies beneath South America. *Geophysical Research Letters*, 44(10):4747–4755, may 2017. doi: 10.1002/2017gl073106.

Portner, D. E., Rodríguez, E. E., Beck, S., Zandt, G., Scire, A., Rocha, M. P., Bianchi, M. B., Ruiz, M., França, G. S., Condori, C., and Alvarado, P. Detailed Structure of the Subducted Nazca Slab into the Lower Mantle Derived From Continent-Scale Teleseismic P Wave Tomography. *Journal of Geophysical Research: Solid Earth*, 125(5), may 2020. doi: 10.1029/2019jb017884.

Potin, B. *Les Alpes occidentales : tomographie, localisation de séismes et topographie du Moho*. PhD thesis, Université Grenoble Alpes, 2016.

Pratt, G., Shin, C., and Hicks. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998. ISSN 1365-246X. doi: 10.1046/j.1365-246x.1998.00498.x.

Pratt, R. G. Frequency-domain elastic wave modeling by finite differences; a tool for crosshole seismic imaging. *Geophysics*, 55(5):626–632, 1990. ISSN 0016-8033. doi: 10.1190/1.1442874.

Ramachandran, K. and Hyndman, R. D. The fate of fluids released from subducting slab in northern Cascadia. *Solid Earth*, 3(1):121–129, apr 2012. doi: 10.5194/se-3-121-2012.

Ramos, V. A. and Folguera, A. Andean flat-slab subduction through time. *Geological Society, London, Special Publications*, 327(1):31–54, jan 2009. doi: 10.1144/sp327.3.

Rawlinson, N. and Fishwick, S. Seismic structure of the southeast Australian lithosphere from surface and body wave tomography. *Tectonophysics*, 572:111–122, 2012. ISSN 0040-1951. doi: 10.1016/j.tecto.2011.11.016.

Rawlinson, N., Pozgay, S., and Fishwick, S. Seismic tomography: A window into deep Earth. *Physics of the Earth and Planetary Interiors*, 178(3-4):101–135, 2010. ISSN 0031-9201. doi: 10.1016/j.pepi.2009.10.002.

Robertson, G. S. and Woodhouse, J. H. Comparison of $P$ and $S$ station corrections and their relationship to upper mantle structure. *J. Geophys. Res.*, 102:27355–27366, 1997.

Robertsson, J. O. A. and Chapman, C. H. An efficient method for calculating finite-difference seismograms after model alterations. *SEG Technical Program Expanded Abstracts 1999*, pages 1848–1851, 2000. doi: 10.1190/1.1820903.

Rodgers, A., Krischer, L., Afanasiev, M., Boehm, C., Doody, C., Chiang, A., and Simmons, N. WUS256: An Adjoint Waveform Tomography Model of the Crust and Upper Mantle of the Western United States for Improved Waveform Simulations. *Journal of Geophysical Research: Solid Earth*, 127(7), 2022. ISSN 2169-9313. doi: 10.1029/2022jb024549.

Rodríguez, E. E., Portner, D. E., Beck, S. L., Rocha, M. P., Bianchi, M. B., Assumpção, M., Ruiz, M., Alvarado, P., Condori, C., and Lynner, C. Mantle dynamics of the Andean Subduction Zone from continent-scale teleseismic S -wave tomography. *Geophysical Journal International*, 224(3):1553–1571, 2021. ISSN 0956-540X. doi: 10.1093/gji/ggaa536. SAM5 S 2020 SAM model.

Roecker, S., Baker, B., and McLaughlin, J. A finite-difference algorithm for full waveform teleseismic tomography. *Geophysical Journal International*, 181(2):1017–1040, 2010. ISSN 1365-246X. doi: 10.1111/j.1365-246x.2010.04553.x.

Rogers, G. and Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science*, 300(5627):1942–1943, jun 2003. doi: 10.1126/science.1084783.

Rondenay, S., Abers, G. A., and van Keken, P. E. Seismic imaging of subduction zone metamorphism. *Geology*, 36(4):275, 2008. doi: 10.1130/g24112a.1.

Savard, G., Bostock, M. G., and Christensen, N. I. Seismicity, metamorphism, and fluid evolution across the northern Cascadia fore arc. *Geochemistry, Geophysics, Geosystems*, 19(6):1881–1897, jun 2018. doi: 10.1029/2017gc007417.

Schmandt, B. and Humphreys, E. Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth and Planetary Science Letters*, 297(3-4):435–445, sep 2010. doi: 10.1016/j.epsl.2010.06.047.

Schmandt, B. and Lin, F. P and S wave tomography of the mantle beneath the United States. *Geophysical Research Letters*, 41(18):6342–6349, 2014. ISSN 0094-8276. doi: 10.1002/2014gl061231.

Schurr, B., Asch, G., Rietbrock, A., Trumbull, R., and Haberland, C. Complex patterns of fluid and melt transport in the central Andean subduction zone revealed by attenuation tomography. *Earth and Planetary Science Letters*, 215(1-2):105–119, oct 2003. doi: 10.1016/s0012-821x(03)00441-2.

Scire, A., Zandt, G., Beck, S., Long, M., Wagner, L., Minaya, E., and Tavera, H. Imaging the transition from flat to normal subduction: variations in the structure of the Nazca slab and upper mantle under southern Peru and northwestern Bolivia. *Geophysical Journal International*, 204(1):457–479, 2016. doi: 10.1093/gji/ggv452.

Scire, A., Zandt, G., Beck, S., Long, M. D., and Wagner, L. The deforming Nazca slab in the mantle transition zone and lower mantle: Constraints from teleseismic tomography on the deeply subducted slab between 6 degrees S and 32 degrees S. *Geosphere*, 13(3):665–680, 2017.

Sen, M. K. and Stoffa, P. L. Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion. *Geophysical Journal International*, 108(1):281–292, 1 1992. ISSN 0956-540X. doi: 10.1111/j.1365-246X.1992.tb00857.x.

Sen, M. K. and Stoffa, P. L. *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press, 1995. doi: 10.1017/cbo9780511997570.

Seriani, G. and Priolo, E. A spectral element method for acoustic wave simulation in heterogeneous media. *Finite Elements in Analysis and Design*, 16:337–348, 1994.

Simmons, N., Forte, A., and Grand, S. Constraining mantle flow with seismic and geodynamic data: A joint approach. *Earth and Planetary Science Letters*, 246(1-2):109–124, jun 2006. doi: 10.1016/j.epsl.2006.04.003.

Skinner, S. M. and Clayton, R. W. The lack of correlation between flat slabs and bathymetric impactors in South America. *Earth and Planetary Science Letters*, 371-372:1–5, jun 2013. doi: 10.1016/j.epsl.2013.04.013.

Snieder, R. and Nolet, G. Linearized scattering of surface waves on a spherical Earth. *Journal of Geophysics*, 61(1):55–63, 1987.

Sobolev, S. V., Babeyko, A. Y., Koulakov, I., and Oncken, O. Mechanism of the andean orogeny: Insight from numerical modeling. In *The Andes*, pages 513–535. Springer Berlin Heidelberg, 2006. doi: 10.1007/978-3-540-48684-8_25.

Stacey, R. Improved transparent boundary formulations for the elastic-wave equation. *Bulletin of the Seismological Society of America*, 78(6):2089–2097, 1988. ISSN 0037-1106. doi: 10.1785/bssa0780062089.

Swenson, J. L., Beck, S. L., and Zandt, G. Crustal structure of the Altiplano from broadband regional waveform modeling: Implications for the composition of thick continental crust. *Journal of Geophysical Research: Solid Earth*, 105(B1):607–621, jan 2000. doi: 10.1029/1999jb900327.

Tago, J., Métivier, L., and Virieux, J. SMART layers: a simple and robust alternative to PML approaches for elastodynamics. *Geophysical Journal International*, 199(2):700–706, 2014. ISSN 0956-540X. doi: 10.1093/gji/ggu298.

Takeuchi, H. and Saito, M. Seismic Surface Waves. *Methods in Computational Physics: Advances in Research and Applications*, 11:217–295, 1972. ISSN 0076-6860. doi: 10.1016/b978-0-12-460811-5.50010-6.

Tape, C., Liu, Q., Maggi, A., and Tromp, J. Adjoint Tomography of the Southern California Crust. *Science*, 325(5943):988–992, 8 2009. ISSN 0036-8075. doi: 10.1126/science.1175298.

Tape, C., Liu, Q., Maggi, A., and Tromp, J. Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophys. J. Int.*, 180:433–462, 2010.

Tarantola, A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984a. ISSN 0016-8033. doi: 10.1190/1.1441754.

Tarantola, A. Linearized inversion of seismic reflection data . *Geophysical Prospecting*, 32 (6):998–1015, 1984b. ISSN 0016-8025. doi: 10.1111/j.1365-2478.1984.tb00751.x.

Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics., Philadelphia, 2005.

Tarantola, A. and Valette, B. Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys. Space Phys.*, 20(2):219–232, 1982.

Tauzin, B., Bodin, T., Debayle, E., Perrillat, J.-P., and Reynard, B. Multi-mode conversion imaging of the subducted Gorda and Juan de Fuca plates below the North American continent. *Earth and Planetary Science Letters*, 440:135–146, apr 2016. doi: 10.1016/j.epsl.2016.01.036.

Thomas, A. M. and Bostock, M. G. Identifying low-frequency earthquakes in central Cascadia using cross-station correlation. *Tectonophysics*, 658:111–116, sep 2015. doi: 10.1016/j.tecto.2015.07.013.

Thomson, W. T. Transmission of Elastic Waves through a Stratified Solid Medium. *Journal of Applied Physics*, 21(2):89–93, 1950. ISSN 0021-8979. doi: 10.1063/1.1699629.

Tikhonov, A. N. and Arsenin, V. Y. *Solutions of Ill-posed problems*. W.H. Winston, 1977.

Tong, P., Chen, C.-w., Komatitsch, D., Basini, P., and Liu, Q. High-resolution seismic array imaging based on an SEM-FK hybrid method. *Geophysical Journal International*, 197(1): 369–395, 2014a. ISSN 0956-540X. doi: 10.1093/gji/ggt508.

Tong, P., Komatitsch, D., Tseng, T., Hung, S., Chen, C., Basini, P., and Liu, Q. A 3-D spectral-element and frequency-wave number hybrid method for high-resolution seismic array imaging. *Geophysical Research Letters*, 41(20):7025–7034, 2014b. ISSN 0094-8276. doi: 10.1002/2014gl061644.

Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., and Wellington, P. Bessel smoothing filter for spectral-element mesh. *Geophysical Journal International*, 209(3):1489–1512, 2017. ISSN 0956-540X. doi: 10.1093/gji/ggx103.

Tromp, J., Tape, C., and Liu, Q. Y. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys. J. Int.*, 160(1):195–216, 2005.

Tsuboi, S., Ando, K., Miyoshi, T., Peter, D., Komatitsch, D., and Tromp, J. A 1.8 trillion degrees-of-freedom, 1.24 petaflops global seismic wave simulation on the K computer. *The International Journal of High Performance Computing Applications*, 30(4):411–422, jul 2016. doi: 10.1177/1094342016632596.

Unser, M. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, 2000. ISSN 0018-9219. doi: 10.1109/5.843002.

van Hunen, J., van den Berg, A. P., and Vlaar, N. J. On the role of subducting oceanic plateaus in the development of shallow flat subduction. *Tectonophysics*, 352(3-4):317–333, aug 2002. doi: 10.1016/s0040-1951(02)00263-9.

VanderBeek, B. P. and Faccenda, M. Imaging upper mantle anisotropy with teleseismic P-wave delays: Insights from tomographic reconstructions of subduction simulations. *Geophysical Journal International*, 225(3):ggab081–, 2021. ISSN 0956-540X. doi: 10.1093/gji/ggab081.

Vidale, J. E., Schmidt, D. A., Malone, S. D., Hotovec-Ellis, A. J., Moran, S. C., Creager, K. C., and Houston, H. Deep long-period earthquakes west of the volcanic arc in Oregon: Evidence of serpentine dehydration in the fore-arc mantle wedge. *Geophysical Research Letters*, 41(2):370–376, jan 2014. doi: 10.1002/2013gl059118.

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

Wagner, L., Beck, S., and Long, M. Peru lithosphere and slab experiment, 2010. URL `https://www.fdsn.org/networks/detail/ZD_2010/`.

Wang, H., Huismans, R. S., and Rondenay, S. Water migration in the subduction mantle wedge: A two-phase flow approach. *Journal of Geophysical Research: Solid Earth*, 124(8): 9208–9225, aug 2019. doi: 10.1029/2018jb017097.

Wang, K., Wang, Y., Song, X., Tong, P., Liu, Q., and Yang, Y. Full-Waveform Inversion of High-Frequency Teleseismic Body Waves Based on Multiple Plane-Wave Incidence: Methods and Practical Applications. *Bulletin of the Seismological Society of America*, 112 (1):118–132, 2021. ISSN 0037-1106. doi: 10.1785/0120210094.

Wang, Y., Chevrot, S., Monteiller, V., Komatitsch, D., Mouthereau, F., Manatschal, G., Sylvander, M., Diaz, J., Ruiz, M., Grimaud, F., Benahmed, S., Pauchet, H., and Martin, R. The deep roots of the western Pyrenees revealed by full waveform inversion of teleseismic P waves. *Geology*, 44(6):475–478, 2016. ISSN 0091-7613. doi: 10.1130/g37812.1.

Wannamaker, P. E., Evans, R. L., Bedrosian, P. A., Unsworth, M. J., Maris, V., and McGary, R. S. Segmentation of plate coupling, fate of subduction fluids, and modes of arc magmatism in Cascadia, inferred from magnetotelluric resistivity. *Geochemistry, Geophysics, Geosystems*, 15(11):4230–4253, nov 2014. doi: 10.1002/2014gc005509.

Ward, K. M., Zandt, G., Beck, S. L., Wagner, L. S., and Tavera, H. Lithospheric structure beneath the northern Central Andean Plateau from the joint inversion of ambient noise and earthquake-generated surface waves. *Journal of Geophysical Research: Solid Earth*, 121(11):8217–8238, nov 2016. doi: 10.1002/2016jb013237.

Watanabe, T. Effects of water and melt on seismic velocities and their application to characterization of seismic reflectors. *Geophys. Res. Lett.*, 20:2933–2936, 1993.

Wech, A. G. Cataloging tectonic tremor energy radiation in the Cascadia subduction zone. *Journal of Geophysical Research: Solid Earth*, 126(10), oct 2021. doi: 10.1029/2021jb022523.

Wech, A. G. and Creager, K. C. Automated detection and location of Cascadia tremor. *Geophysical Research Letters*, 35(20), oct 2008. doi: 10.1029/2008gl035458.

Wilson, D. S. Confidence intervals for motion and deformation of the Juan de Fuca Plate. *Journal of Geophysical Research*, 98(B9):16053, 1993. doi: 10.1029/93jb01227.

Woodward, M. J. Wave-equation tomography. *GEOPHYSICS*, 57(1):15–26, 1992. ISSN 0016-8033. doi: 10.1190/1.1443179.

Wu, F. T., Kuo-Chen, H., and McIntosh, K. Subsurface imaging, TAIGER experiments and tectonic models of Taiwan. *Journal of Asian Earth Sciences*, 90(Geology71979):173–208, 2014. ISSN 1367-9120. doi: 10.1016/j.jseaes.2014.03.024.

Wu, R. S. and Aki, K. Scattering characteristics of elastic waves by an elastic heterogeneity. *Geophysics*, 50:582–595, 1985.

Yilmaz, O. *Seismic Data Analysis*, volume 10 of *Investigations in Geophysics*. Society Of Exploration Geophysicists, Tulsa, USA, 2 ed edition, 2001. ISBN 1560800941. doi: 10.1190/1.9781560801580. URL `http://dx.doi.org/10.1190/1.9781560801580`.

Yuan, K. and Romanowicz, B. Seismic evidence for partial melting at the root of major hot spot plumes. *Science*, 357(6349):393–397, 2017. ISSN 0036-8075. doi: 10.1126/science.aan0760.

Yuan, S., Fuji, N., and Singh, S. C. High-frequency localized elastic full-waveform inversion for time-lapse seismic surveys. *GEOPHYSICS*, 86(3):R277–R292, 2021. ISSN 0016-8033. doi: 10.1190/geo2020-0286.1.

Yuan, X., Sobolev, S. V., Kind, R., Oncken, O., Bock, G., Asch, G., B. Schurr, B., Graeber, F., Rudloff, A., Hanka, W., Wylegalla, K., Tibi, R., Haberland, C., Rietbrock, A., Giese, P., Wigger, P., Rower, P., Zandt, G., Beck, S., Wallace, T., Pardo, M., and Comte, D. Subduction and collision processes in the Central Andes constrained by converted seismic phases. *Nature*, 408:958–961, 2000.

Zandt, G., Beck, S. L., Ruppert, S. R., Ammon, C. J., Rock, D., Minaya, E., Wallace, T. C., and Silver, P. G. Anomalous crust of the Bolivian Altiplano, central andes: Constraints from broadband regional seismic waveforms. *Geophysical Research Letters*, 23(10):1159–1162, may 1996. doi: 10.1029/96gl00967.

Zhang, H. and Thurber, C. H. Double-Difference Tomography: The Method and Its Application to the Hayward Fault, California. *Bulletin of the Seismological Society of America*, 93(5):1875–1889, 2003. ISSN 0037-1106. doi: 10.1785/0120020190.

Zhao, D., Yu, S., and Liu, X. Seismic anisotropy tomography: New insight into subduction dynamics. *Gondwana Research*, 33:24–43, 2016. ISSN 1342-937X. doi: 10.1016/j.gr.2015.05.008.

Zhao, L., Jordan, T. H., and Chapman, C. H. Three-dimensional Fréchet differential kernels for seismicdelay times. *Geophysical Journal International*, 141(3):558–576, 2000. ISSN 0956-540X. doi: 10.1046/j.1365-246x.2000.00085.x.

Zhao, Z. and Sen, M. K. A gradient-based Markov chain Monte Carlo method for full-waveform inversion and uncertainty analysis. *Geophysics*, 86(1):R15–R30, 1 2021. ISSN 0016-8033. doi: 10.1190/geo2019-0585.1.

Zhu, H., Bozdağ, E., Peter, D., and Tromp, J. Structure of the European upper mantle revealed by adjoint tomography. *Nature Geoscience*, 5(7):493–498, 7 2012. ISSN 1752-0894. doi: 10.1038/ngeo1501.

Zhu, H., Bozdağ, E., and Tromp, J. Seismic structure of the european upper mantle based on adjoint tomography. *Geophysical Journal International*, 201(1):18–52, feb 2015. doi: 10.1093/gji/ggu492.

Zhu, L. and Rivera, L. A. A note on the dynamic and static displacements from a point source in multilayered media. *Geophysical Journal International*, 148(3):619–627, 2002. ISSN 1365-246X. doi: 10.1046/j.1365-246x.2002.01610.x.