



Alternative misfit functions in Full Waveform Inversion : From synthetic to field data

Arnaud Pladys

► To cite this version:

Arnaud Pladys. Alternative misfit functions in Full Waveform Inversion: From synthetic to field data. Earth Sciences. Université Grenoble Alpes [2020-..], 2021. English. NNT : 2021GRALU016 . tel-03293646

HAL Id: tel-03293646

<https://theses.hal.science/tel-03293646>

Submitted on 21 Jul 2021

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences de la Terre et de l'Univers et de l'Environnement
(CESTUE)**

Arrêté ministériel : 01 Septembre 2019

Présentée par

Arnaud Pladys

Thèse dirigée par **Romain Brossier**
et codirigée par **Ludovic Métivier**

préparée au sein de **l'institut des Sciences de la Terre**
dans **l'École doctorale Terre, Univers, Environnement**

Alternative misfit functions in Full Waveform Inversion From synthetic to field data

Thèse soutenue publiquement le **14 Juin 2021**,
devant le jury composé de :

Hervé Chauris

Professeur Université Mines ParisTech, Rapporteur

René-Edouard Plessix

Chercheur expert Shell, Rapporteur

Philippe Roux

Directeur de Recherche CNRS Université Grenoble Alpes, Président

Jean-Pierre Vilotte

Physicien Institut de Physique du Globe de Paris, Examinateur

Jérémie Messud

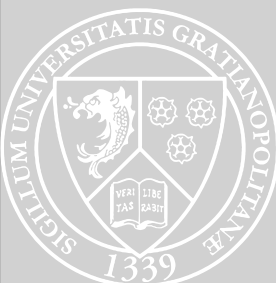
Research Manager CGG, Invité

Romain Brossier

Maître de conférences Université Grenoble Alpes, Directeur de thèse

Ludovic Métivier

Chargé de Recherche CNRS Université Grenoble Alpes, Co-Directeur de thèse



Remerciements

La thèse est une longue aventure, semblable à la descente d'un long fleuve, des sources perdues dans les montagnes jusqu'à son immense embouchure. C'est un voyage duquel on ressort grandi, une aventure qui nous forme, nous façonne et modifie profondément notre perception du monde. Ce fleuve, aux eaux parfois paisibles, mais aussi tumultueuses par moment, requiert toute notre force mais aussi celle de notre entourage. Je n'aurais pas facilement navigué jusqu'à l'embouchure que représente la soutenance sans l'aide et le soutien de nombreuses personnes que je tiens à remercier ici.

Tout d'abord, je voudrais chaleureusement remercier Romain et Ludovic, mes deux directeurs de thèse. Vous qui m'avez découvert et suivi depuis le master, qui m'avez fait découvrir le monde de la recherche en profondeur durant mon stage de M2 et m'avez donné le goût de continuer à approfondir un petit peu le savoir humain au travers de cette thèse. Vous avez été d'un grand soutien, toujours présents, même lors des coup durs, et toujours prêts à trouver des solutions. Travailler avec vous a été une immense chance, et faire partie du groupe SEISCOPE fut une expérience forte que peu de doctorants ont la chance de vivre. Ce sentiment d'appartenir à une équipe, de pouvoir collaborer avec d'autres personnes activement, de partir ensemble en conférence à l'autre bout du monde. Ce sont des moments que je n'oublierai pas.

C'est donc tout naturellement que je voudrais remercier les collègues de SEISCOPE avec qui j'ai vécu de formidables aventures. Merci Philippe Le Bouteiller pour les incroyables moments passés ensemble et pour notre amour des « pots » ! Un grand merci au reste de l'équipe pour tous les bons moments passés ensemble et les souvenirs partagés !

Merci aux membres de mon Jury, Hervé Chauris, René-Edouard Plessix, Philippe Roux, Jean-Pierre Vilotte et Jérémie Messud. Pour une grande partie d'entre vous, j'avais déjà eu le plaisir de vous rencontrer lors de conférences ou du meeting annuel de SEISCOPE. C'était un grand honneur de vous avoir dans mon jury et l'intérêt que vous avez porté à mon travail me touche particulièrement. Je vous remercie aussi pour la très enrichissante discussion que nous avons pu avoir ensemble lors de ma soutenance.

Merci aux collègues d'ISTerre, avec une pensée toute particulière pour Jacques Brives, toujours là pour discuter, rigoler, et avec qui j'ai passé de supers moments. Merci pour ton soutien. Merci aussi aux amis de licence, Jonathan Sekkat et William Moraglia. Vous avez tous les deux vécu des moments difficiles, et j'espère que le futur sera radieux pour vous. Vous avez été de magnifiques compagnons de route durant ces années universitaires et je ne vous oublierai pas.

Merci à Guillaume et Anne, rencontrés lors de la licence à Nice, et qui sont devenus au fil de temps des amis très proches. Merci pour les bons moments passés ensemble, d'être venus en Norvège me voir lors de mon séjour chez Equinor. Tout cela restera d'incroyables souvenirs pour la vie, et il y en aura encore plein d'autres à écrire j'espère !

Un merci tout particulier pour Grégoire Guillet, rencontré lors du master. Tu as été incroyable, et les moments passés ensemble, à créer les pires playlists de mauvais goût sur Spotify, à réussir à me motiver pour aller skier (et par la même occasion avoir les pires courbatures de ma vie !), mais aussi à me proposer un masque de ski pour couper des oignons puis de passer à deux doigts de mettre le feu au chalet. Même si on n'est souvent « pas rassuré », je te souhaite le meilleur pour le futur et tu resteras toujours un ami exceptionnel.

Enfin, et j'ai envie de dire « évidemment », un immense merci à Julien Thurin. On s'est rencontré en L1 à Nice, mais c'est à partir de la L2 et de notre parcours géophysique que nous avons commencé à être si proches. On s'est suivi durant toutes ces années, de la licence, du parcours géophysique, du M1/M2, jusqu'à la thèse. C'est complètement incroyable d'avoir fini par faire tous les deux une thèse dans la même équipe avec les mêmes directeurs de thèse, après toutes ces années partagées ! Du voyage au Bénin ensemble pour une mission de terrain à planter des géophones par 40 degrés dans du sol dur comme du béton, aux vacances à la montagne ensemble avec Greg, et tous les congrès qu'on a fait ensemble durant la thèse. Ton côté hipster incroyable, et ta capacité à être instantanément à 400% sur un projet me fascineront toujours ! Tu as vécu, comme moi, des moments pas faciles durant tes années de thèse et tu as aussi réussi à affronter tout cela et revenir toujours plus fort. Tu es maintenant marié et papa d'un magnifique petit Arthur qui est tellement mignon. J'espère que ce bonheur ne cessera jamais. Toi qui es devenu au fil du temps un de mes amis les plus proches, je te souhaite le meilleur dans ta vie et j'espère qu'on gardera toujours le contact !

Un grand merci à toute ma famille, et particulièrement à ma sœur, qui ne sait toujours pas si je suis plus scientifique que cuisinier ! C'était un immense plaisir d'avoir pu vous avoir tous ensemble présent dans l'amphi après une année 2020 plus que chaotique et d'avoir réussi à partager un petit mais magnifiquement réussi pot de thèse ensemble ! Vous m'avez toujours soutenu durant ces années de fac, du premier jour de licence jusqu'à ce 14 Juin 2021 que vous, comme moi, n'êtes pas prêts d'oublier.

Un grand merci à mon papa, lui aussi docteur, mais en biologie. Merci d'avoir toujours aiguïser ma curiosité scientifique, des premiers microscopes et loupe binoculaire que tu m'avais rapportés alors que j'étais encore au collège, mais aussi de cette incroyable visite d'ATLAS au LHC qui m'avait tellement fasciné et encore plus motivé dans mon amour pour les sciences. Merci pour tes conseils et de m'avoir aiguillé vers la géologie alors que j'étais parti pour m'inscrire en biologie pour marcher dans tes pas. C'était un grand conseil, et sans toi je ne pense pas que j'aurais réussi comme ça. Mes années de fac en géophysique furent une grande révélation et j'ai vraiment adoré cette liberté et ce sentiment d'être si proche de la science. C'est cette motivation qui m'a permis d'aller jusqu'au bout et je ne sais pas si j'aurais si bien réussi dans une autre branche ! J'imagine ta fierté aujourd'hui de savoir ton fiston docteur comme toi.

Un merci tout particulier à ma maman, qui souvent était ma « bouée » de secours lorsque les rapides se faisaient trop intenses. Tu as toujours été derrière moi, prête à m'aider à n'importe quel moment. Tu as toujours été là, durant tout ma scolarité, à me suivre, me motiver, alors que je n'étais pas forcément le meilleur des élèves au lycée. Tu m'as toujours poussé de l'avant, donné la confiance qu'il me manquait quand je n'étais pas sûr de moi. Merci pour ton soutien indéfectible durant mes années de fac. Je sais à quel point je t'ai rendu fière à chaque fois que je réussissais mes années avec brio. Merci aussi pour tes heures de relecture méticuleuse de textes scientifiques relativement hermétiques pour les non-initiés, sacrée fin carrière à la meilleure des profs d'anglais ! Je ne te remercierai jamais assez pour tout ça. Je sais à quel point tu es fière de me savoir aujourd'hui docteur et ce diplôme sera toujours intimement lié à toi !

Et évidemment, un immense et incommensurable merci à Sophie, ma moitié qui m'a supporté (dans tous les sens du terme) pendant toute ces années ! On a traversé de nombreuses épreuves ensemble, et bien que, si par moment le fleuve semblait nous avoir coulés, nous avons réussi à ne jamais nous noyer et à toujours revenir à flots. Je ne pense pas que j'aurais réussi à arriver jusque-là sans toi, et pour cela je te serai à toujours éternellement reconnaissant.

J'espère que le futur sera radieux, et qu'on trouvera notre équilibre et que tu réussiras enfin à t'épanouir complètement, que ça soit en devenant une botaniste aguerrie collectionnant les graines ou tout autre chose qui te passionne. C'est à moi maintenant de te soutenir après tout ce que tu m'as donné.

And now, a part in English.

I would like to thank good family friends from the US, Tony and Judy Schumacher, who were always kind to me and who even woke up at 3 am to watch me live on YouTube during my defense. You were in my heart and still are, and I hope to come soon to the US to visit you.

I would also thank all the SEISCOPE team, Hugo Sanchez, Marco Salcedo, Wei Zhou, Yang Li, Paul Wellington, and all the others. You were some fantastic teammates. I would like to give some special thanks to Pengliang Yang. It was a pleasure working with you a bit on the code. I know life hasn't been that easy with you in the past years, but I hope you are now finding a nice balance. You deserve it!

Another special thanks to Yubing Li, for all the good moments spent together, speaking of science but also of culture, society, and many other subjects. Your open mind was a delight, and your motivation was terrific, always ready to grab the best food and take a 30min Uber to get some good craft beer!

Thanks, Andrzej Górszczyk, for the excellent discussion together, for the beers, and for your spot on scientific feedback. You are someone amazing, and I wish you all the best for the future!

A special thanks to Nishant Kamath for all the time spend together, from barbecues to discussions on good craft beers. Thanks for the scientific works we did together working on the Valhall data! You also are one of a kind, and I hope the future will be bright for you and that we will keep in touch!

Finally, an extraordinary thanks to an extraordinary friend. Marwan Irnaka (first of his name!), I met you during my Ph.D. as you Joined SEISCOPE for your Ph.D. Since then, we got closer and closer. You quickly join the gang in our office, and from then we became close friends. Thanks for your support in the good but also hard times, thanks for all the chit-chat in the office, thanks for the fantastic memories during conferences with you. You became one of my best friends, and I will never forget you. You are one of the nicest people I know (and maybe on Earth?), with consistently good intentions. Even if I tried to give you some French course on "how to bitch" you were always too kind for this. You are amazing and unique, and I hope the future will be bright for you. Now that you are back in Indonesia, I definitely need to come and visit you there (with maybe one master code cracked?).

Armand Pladys

Abstract

Full-waveform inversion is a seismic imaging method known to yield high-resolution results that go towards directly interpretable reconstructed models. Unfortunately, it suffers from a flaw: the need for accurate initial models to converge toward meaningful reconstruction of subsurface parameters. This limitation is due to the non-convexity of the least-squares distance conventionally used as the distance measurement function, which translates into the cycle-skipping issue that has been documented since the introduction of full-waveform inversion itself. Thus, finding solutions to improve full-waveform inversion robustness to cycle-skipping has been the subject of a large number of studies. From all the propositions made to improve this method, one of the most documented is replacing the classical least-squares norm with alternative misfit functions.

We first propose a review of multiple propositions of alternative misfit functions. We explain which principles these methods are based on and illustrate how they are designed to better handle kinematic mismatch than the least-squares misfit function. A set of carefully designed synthetic benchmarking tests is then introduced to assess the behavior of a selection of alternative misfit functions containing two optimal transport-based misfit functions, a wiener filter based, an instantaneous enveloped based, and finally, a normalized integration based misfit function. On these several canonical synthetic tests, each formulation is pushed to its limits, allowing us to establish the pros and cons of each formulation.

This work can also be seen as an attempt to promote a more systematic cross-comparison of alternative misfit functions on fair benchmarking setups, as we observed that alternative misfit functions are often proposed on carefully designed synthetic setups that maximize their benefits. This observation can be related, in our view, to the discrepancy between the number of alternative formulations made and assessed on synthetic cases in the literature and actual case studies performed with them on field data.

Finally, after selecting what appears as the most promising candidates - a graph-space optimal transport-based misfit function - we compare it to the least-squares distance in a case study based on the three-dimensional ocean bottom cable data from the Valhall field. The motivation behind using this data set comes from the wide variety of studies already performed with it. This allows us to work in a relatively controlled framework while facing field data difficulties. The comparison is first performed starting from a reflection traveltime tomography initial model used in previous studies. Then, a second comparison is performed starting from a crude, linearly varying in-depth one-dimensional initial velocity model. This last setup illustrates the robustness to cycle-skipping obtained through a change of the misfit function inside the full-waveform inversion formalism. The encouraging and meaningful results obtained from this case study demonstrate that cycle-skipping robustness of full-waveform inversion can be drastically improved using the proposed graph-space optimal transport-based misfit function.

Résumé

L'inversion de forme d'onde complète est une méthode d'imagerie sismique permettant d'obtenir des résultats à haute résolution pouvant même parfois être directement interprétés. Malheureusement, cette méthode souffre d'un défaut : la nécessité d'avoir un modèle initial suffisamment précis pour assurer la convergence de l'inversion vers une reconstruction réaliste du sous-sol. Cette limitation est due à la non-convexité de la norme euclidienne qui est traditionnellement utilisée comme fonction de mesure de distance. Cette non-convexité se traduit par le problème du saut ou ambiguïté de phase, qui a été documenté depuis l'introduction même de l'inversion de forme d'onde complète. C'est donc naturellement qu'un grand nombre d'études ont été réalisées afin de trouver des solutions pour améliorer la robustesse au problème du saut de phase. Parmi toutes les propositions faites pour améliorer cette méthode, une des plus documentées est le remplacement de la fonction de mesure de distance par d'autres fonctions alternatives.

Nous proposons d'abord une analyse de différentes alternatives pour remplacer la fonction de mesure de distance traditionnellement utilisée. Les principes fondamentaux sur lesquels sont basées ces alternatives sont introduits, en association avec une analyse de leurs avantages en terme de robustesse au saut de phase. Une sélection de certaines de ces fonctions coût alternatives est ensuite testée sur différents tests synthétiques soigneusement conçus pour évaluer leurs comportements, poussant chacune des différentes formulations dans ses retranchements. Les avantages et inconvénients de chacune des propositions sont établis grâce à l'analyse effectuée sur ces différents tests. Une évaluation finale est effectuée afin de définir les formulations les plus prometteuses pour une application sur données réelles.

Ce travail sur les fonctions coût alternatives s'inscrit dans la continuité d'une observation que nous avons réalisée : la disparité entre le grand nombre de fonctions coût alternatives proposées dans la littérature et le nombre réduit d'applications sur données réelles avec de telles alternatives. En effet, trop souvent les fonctions coût alternatives sont proposées sur des tests synthétiques soigneusement conçus pour mettre en avant leurs bénéfices de façon "fictive". Ce travail prône une comparaison juste et reproductible des fonctions coût alternatives.

Finalement, après avoir sélectionné la fonction coût qui nous paraissait la plus prometteuse – le transport optimal du graphe – nous la comparons à la norme euclidienne dans une application sur le jeu de données réelles de fond de mer du champ de Valhall. Le choix de ce jeu de données est motivé par l'abondante littérature disponible, avec de nombreuses inversions de forme d'onde complète déjà réalisées. Cela nous permet d'avoir plus de contrôle dans notre analyse. La comparaison entre les deux fonctions coût est d'abord effectuée en partant du modèle initial provenant d'une tomographie en réflexion, qui est traditionnellement utilisé avec ces données. Puis, une deuxième comparaison est effectuée à partir d'un modèle initial mono-dimensionnel variant linéairement avec la profondeur. Ce deuxième cas sert à illustrer l'amélioration de la robustesse au saut de phase obtenue grâce au transport optimal du graphe. Les résultats présentés dans cette étude sont très encourageants et démontrent bien que l'inversion de forme d'onde complète peut être rendue plus résistante au saut de phase grâce à un changement de fonction coût.

Contents

General Introduction	1
An overview of seismic imaging	1
Going high resolution with full-waveform inversion	3
Non convexity of full waveform inversion	6
Outline of the manuscript	7
1 Full-waveform inversion	8
1.1 A rapid overview of full-waveform inversion development	9
1.2 Formulation of FWI	11
1.2.1 The forward problem	12
1.2.2 Solution to the FWI problem	17
1.3 Non-convexity in FWI	23
1.3.1 Non-convexity and cycle skipping limitations	23
1.3.2 Classical approaches	26
1.3.3 Reformulating the FWI problem	28
2 On cycle-skipping and misfit functions modification for full-wave inversion: comparison of five recent approaches	30
2.1 Introduction	31
2.2 General FWI framework and misfit function formulation	35
2.2.1 General framework	35
2.2.2 Adaptive waveform inversion	36
2.2.3 Instantaneous envelope	37
2.2.4 Normalized Integration Method	38
2.2.5 Kantorovich-Rubinstein optimal transport distance	39
2.2.6 Graph space optimal transport distance	40
2.3 A simple convexity analysis based on time-shifted Ricker wavelets	41
2.4 FWI tests on two canonical examples	46
2.4.1 FWI Test 1: transmission configuration	46
2.4.2 FWI Test 2: reflection configuration	49
2.5 Marmousi case study: toward a more realistic case study	58
2.5.1 Common framework	58
2.5.2 Inverse crime inversion	58
2.5.3 A more realistic inversion	62
2.6 Discussion	73
2.7 Conclusion	73

3	Robust FWI with graph space optimal transport: application to 3D OBC	74
	Valhall data	74
3.1	Introduction	75
3.2	Methodology	78
3.2.1	Modeling	78
3.2.2	Inversion	79
3.3	Field data presentation and full-waveform inversion workflow	81
3.3.1	Geological situation, dataset and initial models	81
3.3.2	Full-waveform inversion workflow	85
3.4	Full-waveform inversion results	94
3.4.1	Starting from the reflection tomography model	94
3.4.2	Starting from the 1D initial model	96
3.4.3	Computational costs	100
3.5	Discussion	107
3.6	Conclusion	107
4	Exploring new ideas and overcoming practical limitations	108
4.1	Overcoming practical issue for 3D KROT applications	109
4.2	Exploring new ideas of alternative misfit functions	112
4.2.1	Instantaneous phase and unwrapped phase misfit functions	112
4.2.2	Spline based envelope misfit function	117
4.2.3	Diffusion based misfit function	121
4.2.4	Beat-tone misfit function	125
4.3	Investigation published during my thesis	129
	Conclusion and perspectives	130
	Conclusion	130
	Perspectives	131
	References	142

CONTENTS

General Introduction

Studying the Earth's interior is a complex task, as only the surface is directly observable. For a long time, the only information geologists could gather came from surface observations, where only geological outcrop (*visible exposure of bedrock or ancient surface deposits*) can be directly observed. Using these observations and some fundamental principles of geology (uniformitarianism, original horizontality, superposition, cross-cutting relationships and Walther's law), reconstruction of geological structures has been made possible, such as mountain chain formation. Adding to these surface observations, the first few meters to kilometers could be sparsely explored through drilling. This sparse information can help constrain the reconstruction of subsurface structures made from surface observations. Still, complex areas, or locations with no outcrop, stayed out of reach, and understanding of such locations remains poor.

To gain knowledge where the eyes cannot see, we rely on one significant branch of physics: geophysics - the physics of the Earth - a vast domain of study that includes the physics of space, atmosphere, oceans, and the interior of the planet. Here, we narrow down our focus on the Earth's interior, to what the layperson thinks when he heard "geophysics".

This variety of applications is also related to various geophysical methods that can be distinguished: seismic, gravity, magnetic, electromagnetic, and electric. All these different methods would correspond to different observables and are linked to physical properties of the medium. Generally speaking, the idea is to rely on understanding and interpreting physical observables that can be measured directly at or near the Earth's surface. These observables are influenced by the internal distribution of physical properties of the subsurface. By analyzing these measurements, the physical properties of the Earth's interior can be obtained, defining the subsurface through the prism of some physical properties (be it wave velocities, density, conductivity...). Then, through rocks physics, rocks or deposits and structures can be retrieved (Telford et al., 1990).

An overview of seismic imaging

From these geophysical methods, one of the oldest is seismology. It has been at the origin of major discovery on Earth's interior structures. Using seismic waves traveling through the Earth's structure (Figure 1a), information of the physical properties is gathered along their travel and can be interpreted (Figure 1b). One of the most famous first descriptions of an internal structure of the Earth has been made by Mohorovičić in 1909 and is the Mohorovičić discontinuity, known as *Moho* for short. The core of the Earth was discovered by Oldham in 1906 and correctly delineated by Gutenberg in 1912 from studying earthquake data.

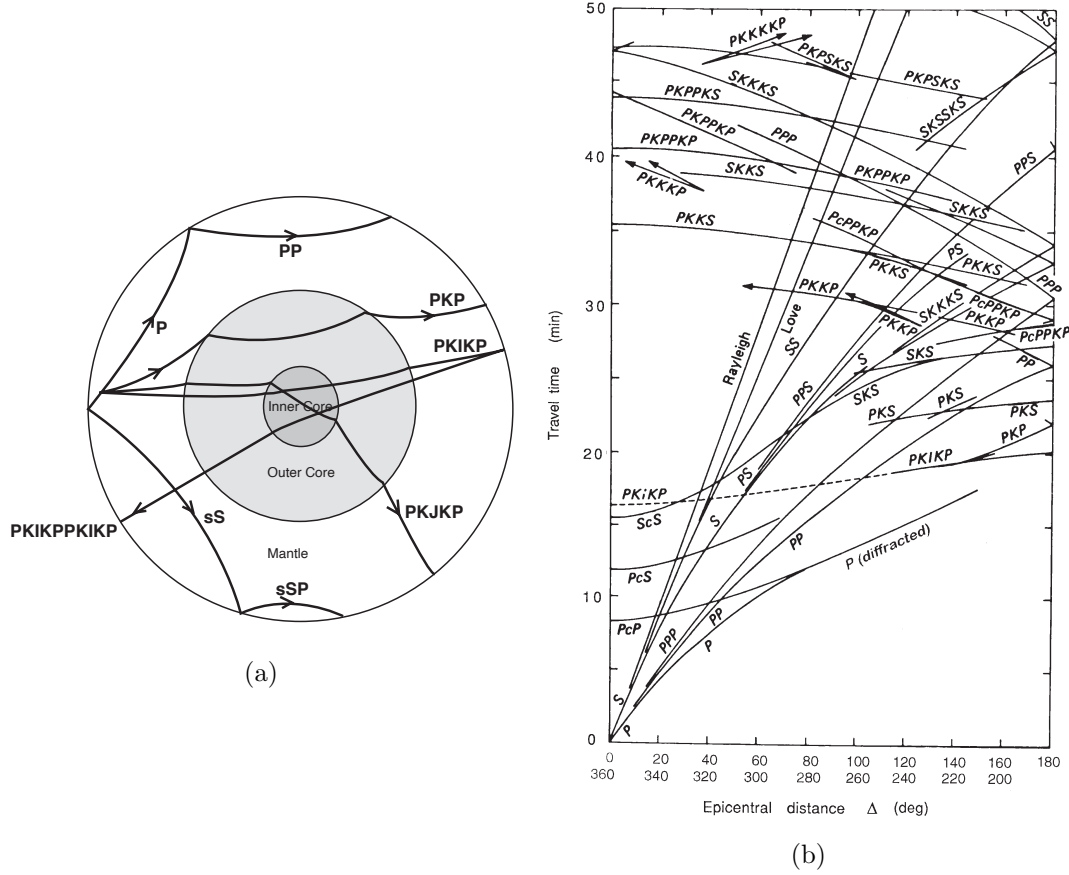


Figure 1: (a) Some of the possible ray paths for seismic waves passing through the Earth (from Fowler et al., 1990). (b) Jeffreys-Bullen (J-B) travel-times curves for an earthquake focus at the surface. Travel-times are represented as a function of epicentral distance. Each line represents a given travel path through the Earth's interior (after Bullen and Bolt, 1985).

The use of seismic waves or mechanical waves presents multiple advantages. First, they are accurately observable and recorded at the surface of the Earth. Second, they can be generated from the largest and most powerful events on Earth: earthquakes, and third, they contain one of the shortest wavelengths of “geophysical waves” (Lay and Wallace, 1995), apart from the electromagnetic waves, which are shorter but also present way shorter penetration depth. This last point could potentially lead to high-resolution reconstruction.

Inside the family of seismic imaging methods, two main groups can be observed. The first one relies on earthquakes as wave sources in a field known as *passive seismic tomography* and is often referred to as *passive seismic tomography*. This method has been applied from global to regional scales and have brought significant insights on the Earth's interior deep structures (Aki et al., 1977; Romanowicz, 2003; Fichtner et al., 2009; Bedle and Lee, 2009; Tape et al., 2010; Panning et al., 2010; French and Romanowicz, 2015; Bozdağ et al., 2016; Lei et al., 2020).

The second one uses active and controlled sources, such as explosions, air guns, vibroseis, or hammer sources. It also differs by its acquisition geometries and targets, which are at the local scales, from tens of meters to tens of kilometers, larger-scale applications being limited by the energy that active sources can generate.

Going high resolution with full-waveform inversion

Since seismic waves traveling through the Earth have relatively small wavelengths, the idea to extract the maximum resolution out of them is tempting. In the '80s, a high-resolution seismic imaging method with the potential resolution of up to half the propagated wavelength was proposed by Lailly (1983) and Tarantola (1984a). This method, referred to as full-waveform inversion (FWI), has been proposed as an attempt to “bridge the gap” between high wavenumber migration imaging principle (Claerbout, 1971) and low wavenumber approach of traveltime tomography (Aki et al., 1974).

This new method is based on an iterative data fitting procedure that aims at minimizing a residual between observed seismic data and synthetic seismograms obtained by solving the wave equation in a *first guess* model of the target area (known as the *starting model* or *initial model*). Unfortunately, the computational cost required in the early development years, even for modest 2D models, was often problematic for the available computer resources (Gauthier et al., 1986; Cary and Chapman, 1988; Crase et al., 1990; Jin et al., 1991; Lambaré et al., 1992). Besides, the datasets available at that time were mainly short offset acquisitions for active seismic, and good initial models were not often available. This made field data applications of FWI challenging. In recent years, thanks to the drastic improvement in computational power from large high-performance computing (HPC) facilities, combined with the availability of large offset wide angle wide azimuth seismic data, has made FWI an exciting method that yields high-resolution reconstruction of the subsurface.

This method has been first applied in geophysical exploration, using active sources, with either onshore (Figure 2a) or offshore acquisition (Figure 2b). The first application has been mainly performed by the oil & gas industry to precisely characterize the subsurface for resources exploitation on the scale of kilometers to tens of kilometers (Plessix and Perkins, 2009; Sirgue et al., 2010; Plessix et al., 2012; Etienne et al., 2012; Warner et al., 2013; Vigh et al., 2014; Operto et al., 2015; He et al., 2019c; Kamath et al., 2020). In the past years, impressive results were obtained as HPC facilities get more powerful, unlocking higher frequency inversion resulting in high-resolution results, as shown in Figure 3. Here, the improvement of resolution that FWI procures over traveltime tomography or reflection tomography is impressive. By yielding such high-resolution imaging, FWI starts to be directly interpretable without the need for migrated images.

Smaller-scale applications also exist, for civil-engineering or archeological targets (Köhn et al., 2018; Wittkamp et al., 2018; Smith et al., 2019; Irnaka et al., 2020). Again, FWI allows retrieving a high-resolution model as shown in Figure 4 where a low-velocity anomaly highlights the shape of fill-in materials used to level to the ground an historical trench.

Alternatively, FWI has been successfully applied to regional or global scales from the past decade (Bedle and Lee, 2009; Fichtner et al., 2009; Panning et al., 2010; Tape et al., 2010; French and Romanowicz, 2015; Bozdağ et al., 2016; Górszczyk et al., 2017; Tromp, 2019). Using active sources at this scale is often unrealistic due to the amount of energy required to propagate seismic waves over long distances. One can remember the *peaceful nuclear explosion* “scientific experiment” realized during the cold war that yield active seismic data at large scale using nuclear bomb as source (Sultanov et al., 1999; Pavlenkova and Pavlenkova, 2006). As this could be less acceptable today, using earthquakes or ambient noise as passive sources are often preferred. However, this requires precise knowledge of the source signature and position, which can be difficult to obtain from small amplitude earthquakes or if the recording network

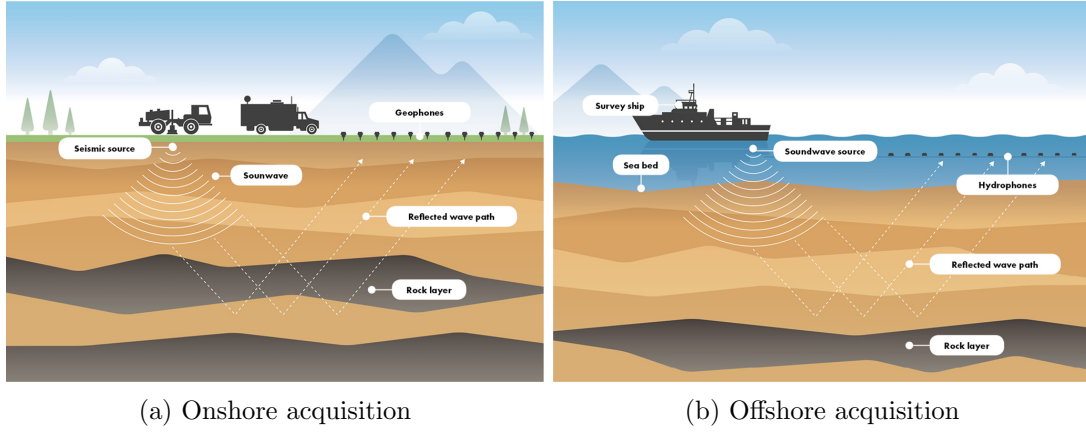


Figure 2: Schematic of an onshore and offshore setup in seismic exploration surveys (Energy Information Australia).

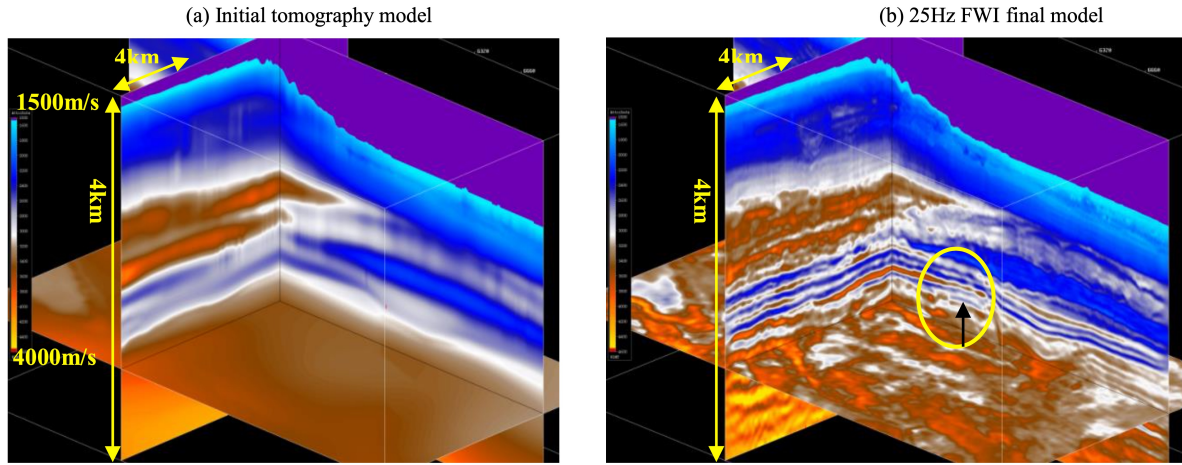


Figure 3: Refraction/reflection FWI results at exploration scale. (a) correspond to the initial model used to start FWI workflow. (b) final 25 Hz FWI results. From Dickinson et al. (2017).

is not dense enough. Despite this, excellent results are obtained using FWI at global scale, as shown in Figure 5 with shear-wave-speed reconstruction results on the last GLAD-M25 global FWI model. This model displays several notable features such as hotspots and subduction slabs. In recent years, strong development of passive seismic has been made (Stehly et al., 2006, 2009; Lu et al., 2018, 2020). These methods are based on ambient noise correlations which use recordings of noise generated by wide spread multiple sources (such as ocean waves, wind, human activity, micro-seismicity) to retrieve Green's functions between receivers couples. These Green's functions can be translated into "virtual sources" that can be used for FWI.

All of these examples confirm that FWI is now a mature seismic technic that is able, when it can achieve convergence, to yield breathtaking results that could be interpreted directly or results in drastically improved migrated images.

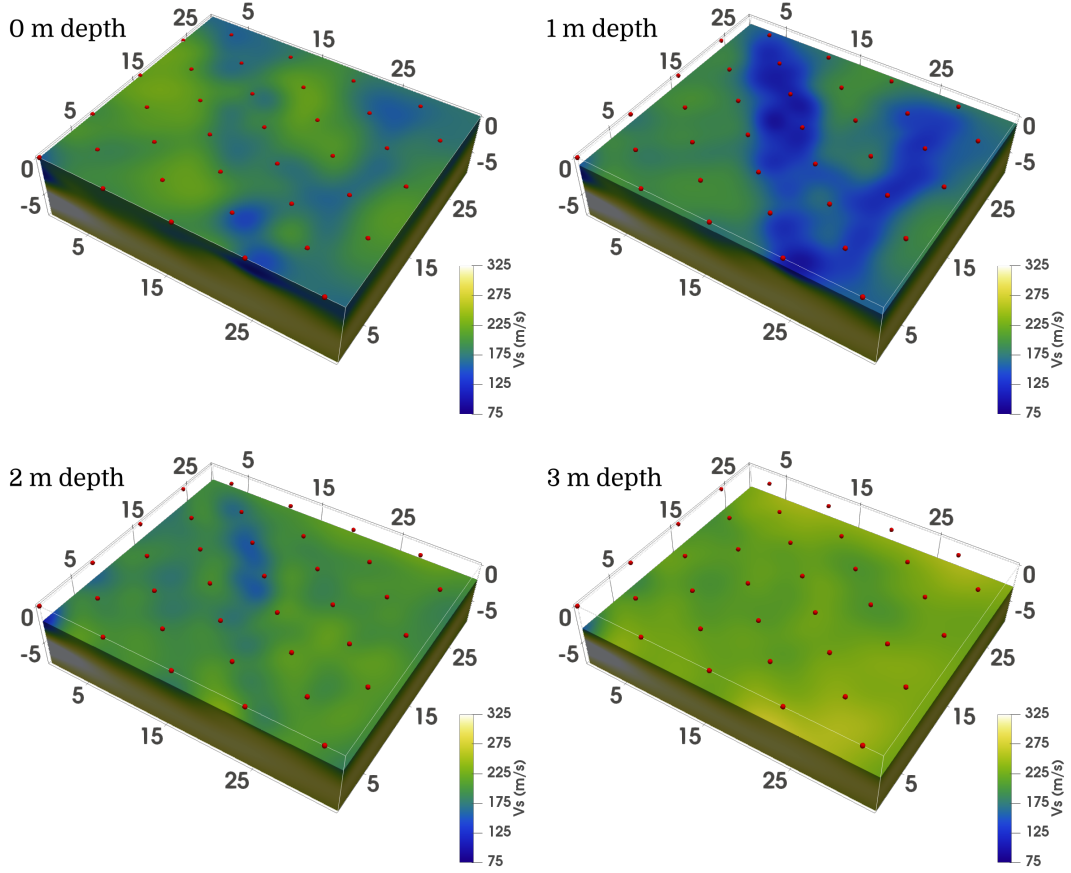


Figure 4: Small scale FWI results at 65 Hz. Here the target is the Ettlingen trench near Karlsruhe, Germany. The low-velocity anomaly in the center corresponds to fill-in materials used to level the ancient trench. From Irnaka, 2021.

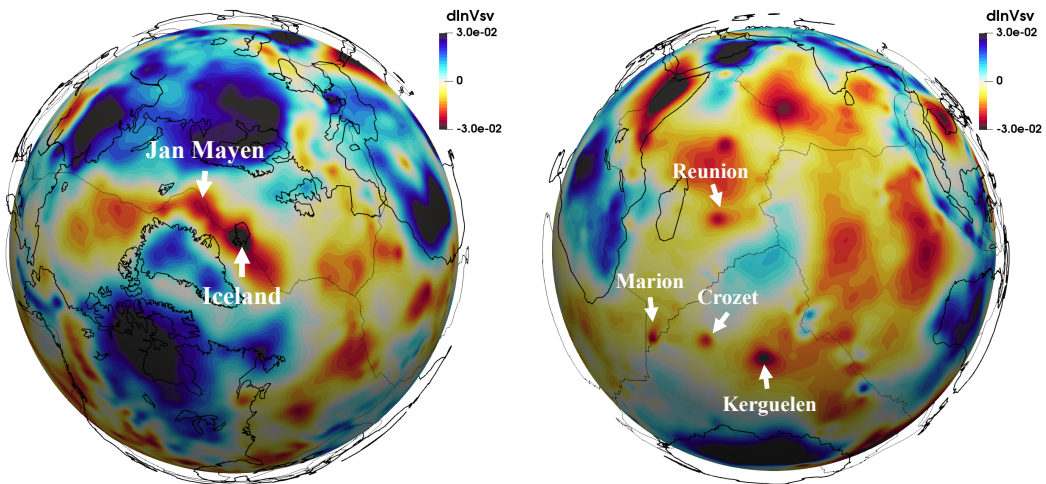


Figure 5: Example of global FWI for vertically polarized shear-wave-speed perturbations in GLAD-M25 at 250 km depth (Lei et al., 2020). Remarkable hotspots are highlighted with white arrow.

Non convexity of full waveform inversion

Still, FWI suffers from a critical flaw in its conventional formulation: the non-convexity of the least-squares norm used as a distance measurement function. This limitation has been documented since the beginning of the FWI formalism and results in the so-called cycle-skipping issue. In a simple way, cycle-skipping occurs when calculated data are shifted by more than half a period compared to the observed data. This results in a condition to apply FWI: start from a good enough initial model that predicts the data within half a period. If this requirement is satisfied, FWI should converge toward the correct solution. If not, FWI suffers from cycle-skipping and is led toward a local minimum.

Practically, solutions have been proposed to mitigate the cycle-skipping issue, and the vast majority of FWI applications rely on three preliminary steps to ensure successful results at exploration scales. The first step is to pick travel time from the data. This is a labor-intensive task requiring human supervision, which is very time-consuming. From these picked travel times, the second step is to build an initial model using traveltimes tomography. Ideally, the model should be as high-resolution as possible, for example, using stereotomography methods (Billette and Lambaré, 1998). The last step is to build a multi-scale workflow for FWI, which follows a frequency-continuation principle that results in inverting first for the lowest frequencies available and then progressively introducing higher frequencies.

The requirement for initial models generated by tomography clearly illustrates the famous “gap” in resolution between tomography and migration. A possible solution would be to make travel-time tomography methods high-resolution. Unfortunately, getting higher wavenumber content from traveltimes tomography is not an easy task, and the methods suffer from intrinsic limitation. Ideally, the solution would be to drag FWI toward low wavenumber content or even entirely remove the need for initial models generated with traveltimes tomography and only relies on FWI.

This motivates the idea to reformulate FWI to make it more flexible, and “automatic”, thus making it less dependent on the initial model quality. By changing its formulation and reducing its non-convexity, the range of applications tackled with FWI can be increased while reducing the uncertainties associated with complex workflows.

In an attempt to make FWI more flexible, two main families can be identified. The first one relies on extension strategies. Here the goal is to artificially extend the research space to mitigate the non-convexity, then progressively constrain the artificial degrees of freedom to converge toward a non-relaxed formulation of FWI. The second one relies on replacing the non-convex least-squares-based misfit function with alternative misfit functions that present increased convexity. In the second family, many propositions have been introduced by the FWI community.

In this thesis, the focus is made on the second family. We first want to systematically compare alternative misfit functions coming from already published methods or newly proposed solutions. It is crucial to keep in mind that replacing the least-squares norm is not an easy task as it often translates into a loss of resolution power, increased complexity, and increased computational cost. This has made the least-squares norm the most used misfit function even if it suffers from a fatal flaw. To test if alternative misfit functions can improve FWI robustness, we have selected a panel of five methods representing different ways of reformulating the problem. One of the scientific contributions of this thesis is the benchmarking of alternative

misfit functions in a systematic way, allowing us to exhibit their pros and cons.

Finally, we perform in-depth testing on one of the most promising solutions, a graph-space optimal transport-based misfit function, on a field dataset starting from a 1D starting model that is not based on prior inversion results. The second scientific contribution of this thesis showed that robust FWI by misfit function replacement is possible and provides encouraging results and that reformulating the original FWI problem is the way forward.

Outline of the manuscript

Chapter 1 introduces FWI and the history behind it. An overview of the formulation of FWI is performed with details on both the forward problem and the inverse problem. Details are given on how to solve the inverse problem of FWI, with an overview of the different approaches available and the introduction of the limitation of the most commonly used method.

Chapter 2 comes from a published paper in *Geophysics* journal (Pladys et al., 2021b). It focuses on a subset of alternative misfit functions and presents several synthetic benchmarks that are carefully designed to assess the capacity of a given misfit function. An in-depth analysis of the pros and cons of each formulation is performed. The conclusion is drawn on the most promising methods.

Chapter 3 comes from a submitted paper in *Geophysics* journal (Pladys et al., 2021a). We take one of the best candidates of the previously tested misfit function, graph-space optimal transport, and use it on a field data application. This application is performed in two parts. First, using the initial model already used in the literature to assess graph-space optimal transport capacity to perform similarly to the conventional L^2 norm in a controlled environment. Second, starting from a crude 1D initial model to introduce strong cycle-skipping and assess the capacity of graph-space optimal transport in one of the most challenging cases.

Chapter 4 presents some explorations I performed during my years of research, with some ideas of new alternative misfit functions and some workarounds developed for practical application.

Chapter 5 finally gives conclusion and perspectives of this manuscript.

Chapter 1

Full-waveform inversion

Contents

1.1	A rapid overview of full-waveform inversion development	9
1.2	Formulation of FWI	11
1.2.1	The forward problem	12
1.2.2	Solution to the FWI problem	17
1.3	Non-convexity in FWI	23
1.3.1	Non-convexity and cycle skipping limitations	23
1.3.2	Classical approaches	26
1.3.3	Reformulating the FWI problem	28

1.1 A rapid overview of full-waveform inversion development

In 1985, Claerbout’s analysis on seismic imaging methods shown there was a “gap” between migration imaging principle and travel-time tomographic imaging principle (Figure 1.1). The first family, relying on the migration principle, provides high-wavenumber content of the subsurface (which translates into small sharp details) but requires a good model of the Earth’s interior to provide accurate results. The second family, travel-time tomography, focuses on reconstructing a low-wavenumber model containing the overall kinematic of the Earth’s interior that is not easily interpretable. As an attempt to bridge the “gap” between these two families, full-waveform inversion (FWI) was first formulated by Lailly (1983) and Tarantola (1984b).

FWI is a seismic imaging method aiming at reconstructing high-resolution models (up to half the shortest wavelength) of the mechanical properties of the subsurface (Devaney, 1984; Pratt and Shipp, 1999; Plessix and Perkins, 2010). FWI relies on a fundamental principle: wavefields traveling through the Earth’s interior gather pieces of information along their path on medium physical parameters (such as P-wave velocity, S-wave velocity, density, attenuation or anisotropy parameters). Therefore, the idea behind FWI is simple: match calculated seismic data (computed by solving the wave equation numerically) to observed seismic data by iteratively updating a model of the Earth’s subsurface. This type of problem is called an “inverse problem” as it boils down to estimate the parameters of partial derivative equations instead of estimating its solution. This problem can also be seen as a data-fitting procedure.

As mentioned, the main interest of FWI is its high-resolution potential that could help filling the “gap” in seismic imaging methods. We can analyze the wavenumber content of the gradient of the misfit function used to update the model, for a given couple of source/receiver of a fixed diffraction point in a homogeneous medium, in the plane wave approximation (Figure 1.2):

$$k = \frac{2\omega}{c} \cos\left(\frac{\theta}{2}\right), \quad (1.1)$$

where k is the wavenumber modulus, ω the angular frequency of the monochromatic plane waves, c the velocity, and θ the angle between the source and receiver rays to the diffraction point. The dependency between the wavenumber content and the illumination angle is made clear, with maximum wavenumber for zero degrees angle, meaning a reflection with zero offsets between source and receiver rays (from Wu and Toksöz, 1987). This illustrates that FWI resolution is constrained by three main factors. First, the most straightforward one is the frequency content of the data used. Second, the wavenumber content at depth is lower as velocity is generally higher (which translates into larger wavelengths). Third, structures at depth are often constrained by small scattering angles. This is why it is important to have various propagation regimes and various arrivals types originating at the same structure to broaden the range of scattering angles (transmission, reflections and wide-angle reflections). Ideally, with omnidirectional illumination of a structure, the resolution can theoretically be half wavelength. Practically, optimal illumination conditions are never reached. It also exhibits the need for broad frequency data with good quality low frequency to easily start inversion and avoid cycle-skipping and high frequency to obtain high-resolution reconstruction.

Interestingly, when the FWI problem was formulated, the community missed two main ingredients to ensure successful results. The first one is that it was introduced in the early age of computational science. That led to serious difficulty solving the wave equation even in limited simple 2D cases as computational power was fairly limited. The second one was the

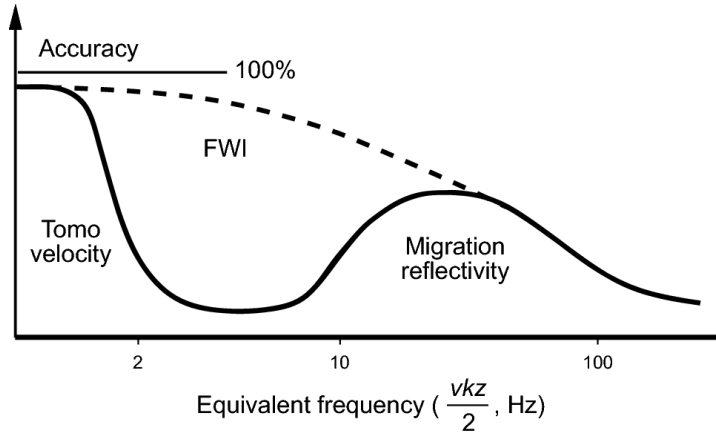


Figure 1.1: Schematic illustration of the accuracy and resolution of tomography, full-waveform inversion and migration methods. The solid black line corresponds to the resolution gap between tomography and migration technics (as proposed by Claerbout, 1985). This gap is supposed to be filled by full-waveform inversion (figure from Yao et al., 2020).

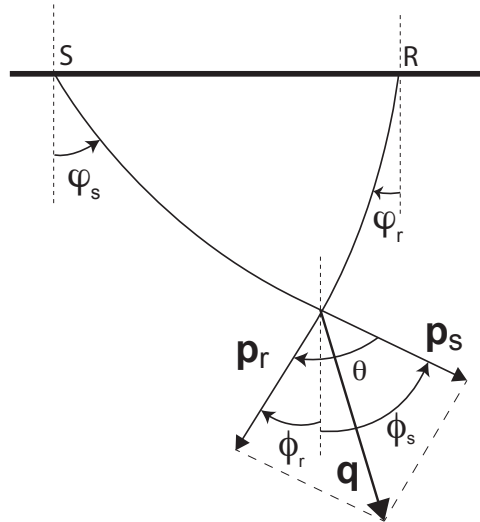


Figure 1.2: Couple of source/receiver of a fixed diffraction point in a homogeneous medium.

availability of appropriate seismic data, and a better understanding of the type of data needed to sufficiently constrain the problem (seismic data were mainly short-offset seismic reflection surveys). That explained why FWI was difficult to apply (Gauthier et al., 1986; Cary and Chapman, 1988; Crase et al., 1990; Jin et al., 1992; Lambaré et al., 1992). These limitations (inappropriate data and computational power) slowed the development of the method, but, in the '90s, Pratt's work brought the essential bricks that unleashed FWI understanding: the formulation of a 2D frequency-domain solver that considerably lighten the computational cost (Pratt, 1990; Pratt and Worthington, 1990; Pratt and Goult, 1991; Pratt et al., 1996, 1998; Pratt, 1999). By seeing FWI in terms of wavenumber covering and acquisition offset, he showed the potential of FWI for crosshole acquisition with transmitted waves. This work was essential to lay the foundation for the first field data applications in the years 2000 when surface acquisition presenting sufficiently large offset coverage started to show up and finally unlock FWI applications with meaningful updates at more than a few hundred meters.

Historically, the method was born in the context of exploration (related to the oil & gas industry). Since then, FWI has been continuously developed and improved, from 2D acoustic mono-parameter (Ravaut et al., 2002, 2004; Sourbier et al., 2009), 3D visco-acoustic mono or multi-parameters (Plessix, 2009; Malinowski and Operto, 2006; Sirgue et al., 2010; Plessix et al., 2012; Warner et al., 2013; Operto et al., 2015; Górszczyk et al., 2017, 2021; Kamath et al., 2020; Pladys et al., 2020) to 3D visco-elastic inversion (Zhu et al., 2015; Raknes et al., 2015; Trinh et al., 2019b,c,a)

Meanwhile, rapidly after the introduction of FWI at exploration scale, global to regional scale FWI from the academic community emerged, with groups from Caltech/Princeton (Tromp et al., 2005; Liu and Archuleta, 2006; Tape et al., 2007; Bedle and Lee, 2009; Tape et al., 2010; Bozdağ et al., 2016; Tromp, 2019), Berkeley (Panning et al., 2010; French and Romanowicz, 2015) and Munich/Utrecht/ETH (Fichtner et al., 2009; Fichtner and Trampert, 2011; Fichtner et al., 2013).

More recently, smaller-scale approaches have emerged and have made drastic improvements in the past years (Köhn et al., 2018; Wittkamp et al., 2018; Smith et al., 2019; Irnaka, 2021). This remains a challenging topic as many difficulties are faced at this scale, such as propagation of surface waves, poorly consolidated area, strong heterogeneities of mechanical properties in the first tens to hundreds of meters.

Finally, the interest of FWI is now going beyond the scope of seismic waves, with successful application on ground penetrating radar (GPR) (Klotzsche et al., 2010; Minet et al., 2010; Klotzsche et al., 2012, 2013; Lavoué et al., 2014; Klotzsche et al., 2019; Giannakis et al., 2019), or even in the medical imaging field that starts to use FWI (Guasch et al., 2020; Marty et al., 2021). This proves that now FWI is a mature method used in a wide variety of scales, targeting multiple imaging problems.

In the following, I introduce the physical and mathematical concepts related to FWI before developing on the actual limitations of the conventional FWI formulation.

1.2 Formulation of FWI

Full-waveform inversion is formulated as a minimization of a function F measuring the distance between observed (recorded) data d_{obs} and synthetics data $d_{cal}[m]$ calculated in a discrete physical model of the subsurface parameters m . These data, d_{obs} and $d_{cal}[m]$, are extracted from the wavefield at receivers' locations. FWI aims to retrieve the model m from the observed data, categorizing FWI into the inverse problem family. By taking into account the complete signal information (all the phases and all the amplitudes), this method is supposed to provide a high-resolution reconstruction of the subsurface model.

Mathematically, FWI can be formulated as finding the minimum of a misfit function f defined as

$$\min_m f(m) = F(d_{cal}[m], d_{obs}), \quad (1.2)$$

where F is a generic function measuring the distance between d_{obs} and $d_{cal}[m]$.

Under general notation, $d_{cal}[m]$ is obtained through the extraction of the values of wavefield at the receivers location such that

$$d_{cal}[m] = Ru[m], \quad (1.3)$$

where R is an extraction operator and $u[m]$ is the solution of the wave propagation problem

$$A[m]u = b, \quad (1.4)$$

with $A[m]$ a generic wave propagation operator (from acoustic to visco-elastic).

Several strategies can be used to solve eq. 1.2, but first, it is required to detail the *forward problem* $A[m]$ which is compulsory to obtain $d_{cal}[m]$

1.2.1 The forward problem

One way of seeing the seismic waves is as optic rays based on a high-frequency approximation (Červený, 2001; Chapman, 2004; Virieux and Lambaré, 2007). This would give access to the traveltime, which leads to raypath connecting a source-receiver couple, that can be solved by different methods (Zelt and Smith, 1992; Bishop et al., 1985; Le Meur et al., 1997; Popovici and Sethian, 1998; Lelièvre et al., 2011; Vidale, 1990; Bretaudeau et al., 2014; Le Bouteiller et al., 2019).

The second way, which is of interest here, is modeling the full seismic wavefield, referred to as *forward modeling*. To do so, we consider the wave equation to approximate as close as possible the real physics of seismic wave propagation in a given medium.

I will detail the elastodynamic equations in the following parts and derive the acoustic and the anisotropic visco-acoustic modeling. These two modeling choices are motivated by their usage in the FWI application performed in this thesis. The first one - acoustic - is the most “basic” modeling we can think of. It only accounts for the pressure waves. The second one introduces anisotropic effects and attenuation, while still only modeling for the pressure waves. This approximation of the physics is important to keep in mind for field data application as the Earth is elastic, and contains more complex anisotropy than the VTI approximation used here. This, by definition, will never allow obtaining a perfect fit of the modeling to the field data as we will miss a part of the physics. This could also lead to modification of physical properties such as P-waves velocity to take into account other physical effects that is not taken into account in the modeling, leading to the creation of *non-physical* updates. This issue is commonly referred to as cross-talk (Operto et al., 2013), but that will not be the focus of this work.

1.2.1.1 The elastodynamic equations

In the following, we follow Einstein convention (summation over repeated indices) for these equations. Generally, to describe the elastodynamic system, we rely on the following Newton’s law

$$\rho \partial_t v_i = \partial_j \sigma_{ij}, \quad (1.5)$$

and on Cauchy’s generalization to the Hooke’s law in tensorial form

$$\sigma_{ij} = c_{ijkl} \varepsilon_{kl}, \quad (1.6)$$

where ρ is density, σ is the stress tensor, c_{ijkl} is the stiffness tensor, ε is the strain tensor with $i, j, k, l \in [x, y, z]$. Strains are related to displacement vector $u = (u_x, u_y, u_z)$ through the expression

$$\varepsilon_{ij} = \frac{1}{2} (\partial_i u_j + \partial_j u_i), \quad (1.7)$$

whose first time derivative is

$$\dot{\varepsilon}_{ij} = \frac{1}{2}(\partial_j \dot{u}_i + \partial_i \dot{u}_j) = \frac{1}{2}(\partial_j v_i + \partial_i v_j). \quad (1.8)$$

Simplification of the stiffness tensor can be obtained due to the symmetry as well as the unique energy definition (Carcione, 2001, p. 55):

$$c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}, \quad (1.9)$$

which reduces the components to only 21.

We introduce the particle velocity vector as $\mathbf{v} = (v_x, v_y, v_z)^T$ and the stress vector as $\boldsymbol{\sigma} = (\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{yz}, \sigma_{xz}, \sigma_{xy})^T$. We can now define the first time derivative for the Hooke's law as the following matrix expression:

$$\partial_t \underbrace{\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{bmatrix}}_{\boldsymbol{\sigma}} = \underbrace{\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ \cdot & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ \cdot & \cdot & c_{33} & c_{34} & c_{35} & c_{36} \\ \cdot & \cdot & \cdot & c_{44} & c_{45} & c_{46} \\ \cdot & SYM & \cdot & \cdot & c_{55} & c_{56} \\ \cdot & \cdot & \cdot & \cdot & \cdot & c_{66} \end{bmatrix}}_C \underbrace{\begin{bmatrix} \dot{\varepsilon}_{xx} \\ \dot{\varepsilon}_{yy} \\ \dot{\varepsilon}_{zz} \\ 2\dot{\varepsilon}_{yz} \\ 2\dot{\varepsilon}_{xz} \\ 2\dot{\varepsilon}_{xy} \end{bmatrix}}_{\boldsymbol{\varepsilon}} = \underbrace{\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ \cdot & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ \cdot & \cdot & c_{33} & c_{34} & c_{35} & c_{36} \\ \cdot & \cdot & \cdot & c_{44} & c_{45} & c_{46} \\ \cdot & SYM & \cdot & \cdot & c_{55} & c_{56} \\ \cdot & \cdot & \cdot & \cdot & \cdot & c_{66} \end{bmatrix}}_C \underbrace{\begin{bmatrix} \partial_x & 0 & 0 \\ 0 & \partial_y & 0 \\ 0 & 0 & \partial_z \\ 0 & \partial_z & \partial_y \\ \partial_z & 0 & \partial_x \\ \partial_y & \partial_x & 0 \end{bmatrix}}_D \underbrace{\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}}_{\mathbf{v}}, \quad (1.10)$$

where we have introduced a differential operator D and the symmetric stiffness matrix C using Voigt indexing, (11) \rightarrow 1, (22) \rightarrow 2, (33) \rightarrow 3, (23) = (32) \rightarrow 4, (13) = (31) \rightarrow 5, (12) = (21) \rightarrow 6.

1.2.1.2 Anisotropic VTI acoustic wave approximation

We can now simplify the stiffness matrix C depending on which *type* of anisotropy we want. In the case of orthorhombic anisotropy, the stiffness matrix C reduces to

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ \cdot & c_{22} & c_{23} & 0 & 0 & 0 \\ \cdot & \cdot & c_{33} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & c_{44} & 0 & 0 \\ \cdot & SYM & \cdot & \cdot & c_{55} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & c_{66} \end{bmatrix}. \quad (1.11)$$

The stiffness matrix C can be further simplified with a special case of orthorhombic anisotropy known as VTI anisotropy. The simplification is obtained by assuming a symmetric (x, y) -plane and setting $\sigma_{xx} = \sigma_{yy}$, $\sigma_{xz} = \sigma_{yz}$, $\varepsilon_{xx} = \varepsilon_{yy}$, $\varepsilon_{xz} = \varepsilon_{yz}$, yielding $c_{11} = c_{22}$, $c_{13} = c_{23}$, $c_{44} = c_{55}$. The resulting stiffness matrix becomes (Duveneck et al., 2008)

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ \cdot & c_{11} & c_{13} & 0 & 0 & 0 \\ \cdot & \cdot & c_{33} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & c_{55} & 0 & 0 \\ \cdot & SYM & \cdot & \cdot & c_{55} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & c_{66} \end{bmatrix}, \quad (1.12)$$

which now involves only 5 independent coefficients due to the relation $c_{12} = c_{11} - 2c_{66}$.

Finally, to obtain the acoustic wave approximation in the presence of VTI anisotropy, we force shear wave velocity on the symmetry axis to be zero in the elastic case in the sense that $c_{55} = c_{66} = 0$, $c_{12} = c_{11}$. Therefore only first 3 rows of C remains:

$$C = \begin{bmatrix} c_{11} & c_{11} & c_{13} & 0 & 0 & 0 \\ \cdot & c_{11} & c_{13} & 0 & 0 & 0 \\ \cdot & \cdot & c_{33} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & SYM & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}. \quad (1.13)$$

To be rigorous, we can note that anisotropy does not exist in a pure acoustic regime. While forcing shear velocity to be zero, it does not prevent the propagation of fictitious shear waves (Alkhalifah, 1998).

The resulting VTI acoustic system is written as

$$\begin{cases} \rho \partial_t v_x = \partial_x \sigma_{xx} \\ \rho \partial_t v_y = \partial_y \sigma_{yy} \\ \rho \partial_t v_z = \partial_z \sigma_{zz} \\ \partial_t \sigma_{xx} = c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z \\ \partial_t \sigma_{yy} = c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z \\ \partial_t \sigma_{zz} = c_{13}(\partial_x v_x + \partial_y v_y) + c_{33} \partial_z v_z. \end{cases} \quad (1.14)$$

Let us denote $\sigma_{xx} = \sigma_{yy} := g$, and $\sigma_{zz} := q$. The pressure is the averaged trace of diagonal stress matrix:

$$p = -\frac{1}{3} \sum_i \sigma_{ii} = -\frac{1}{3}(2g + q). \quad (1.15)$$

Finally, we can introduce Thomsen's anisotropy parameters δ and ϵ as (Duveneck and Bakker, 2011; Zhang et al., 2011):

$$\begin{cases} c_{11} = \rho V_p^2(1 + 2\epsilon) = \rho V_h^2 \\ c_{13} = c_{33} \sqrt{1 + 2\delta} = \rho V_p^2 \sqrt{1 + 2\delta} \\ c_{33} = \rho V_p^2 = \kappa, \end{cases} \quad (1.16)$$

where κ is the bulk modulus while $V_h = V_p \sqrt{1 + 2\epsilon}$ denotes the horizontal velocity. Allowing for the external sources, the VTI acoustic system 1.14 becomes

$$\begin{cases} \rho \partial_t v_x = \partial_x g + f_{v_x} \\ \rho \partial_t v_y = \partial_y g + f_{v_y} \\ \rho \partial_t v_z = \partial_z q + f_{v_z} \\ \partial_t g = \kappa[(1 + 2\epsilon)(\partial_x v_x + \partial_y v_y) + \sqrt{1 + 2\delta} \partial_z v_z] + f_{\sigma_{xx/yy}} \\ \partial_t q = \kappa[\sqrt{1 + 2\delta}(\partial_x v_x + \partial_y v_y) + \partial_z v_z] + f_{\sigma_{zz}}. \end{cases} \quad (1.17)$$

We can further simplify the system by setting ϵ and δ to zero in eq 1.17. This yields the isotropic acoustic wave equation:

$$\begin{cases} \rho \partial_t v_x = -\partial_x p + f_{v_x} \\ \rho \partial_t v_y = -\partial_y p + f_{v_y} \\ \rho \partial_t v_z = -\partial_z p + f_{v_z} \\ \partial_t p = -\kappa(\partial_x v_x + \partial_y v_y + \partial_z v_z) + f_p, \end{cases} \quad (1.18)$$

where we have considered the pressure as defined in eq 1.15.

1.2.1.3 Anisotropic VTI viscoacoustic wave approximation

We considered before the propagation of waves in purely elastic solid. However, as waves propagate through the Earth, the wavefront loses energy (through thermal loss, for instance). This is called attenuation, and it is an important feature that should be taken into account. This effect (reduction of amplitude over distance) has been observed and confirmed from a wide range of experimental tests and field observations. Depending on the target, accounting for attenuation can be crucial to obtain good results of FWI when applied to field data; see Kamath et al. 2020 for instance.

In a first attempt to account for attenuation, constant- Q models were developed to approximate seismic attenuation in seismic exploration and seismology. It is mathematically simple to implement Q in the frequency domain. However, it is more complex in the time domain.

In order to simulate attenuation for time-domain modeling, the generalized Maxwell body (GMB), which is equivalent to the superposition of several standard linear solid (SLS) mechanisms, is used to approximate the constant Q over a specified frequency range.

The SLS mechanisms correspond to attenuation processes (e.g. interstitial atom relaxation, grain boundary relaxation, thermoelasticity, the diffusional motion of dislocations and point defects) that affect the relaxation time following the wave propagation. The general SLS rheology explains these processes very well. Some of them can be modeled with one mechanism, and others using several relaxation mechanisms.

A practical problem in superposing SLS mechanisms is determining the appropriate number of mechanisms, providing the desired constant- Q behavior, and saving computational costs for solving wave equations, particularly in 3D modeling. In general, three SLS mechanisms are considered to be accurate enough for 3D simulations in geophysical prospecting and global seismology study. Less there are SLS mechanisms, and fewer memory-variable equations have to be solved, resulting in time-saving.

Now, let us introduce, for a 1D case, the rheology of generalized Maxwell body (GMB) (Emmerich and Korn, 1987; Moczo and Kristek, 2005):

$$\begin{cases} \partial_t \sigma = M_u(\dot{\epsilon} - \sum_{\ell=1}^L Y_\ell \xi_\ell) \\ \partial_t \xi_\ell + \omega_\ell \xi_\ell = \omega_\ell \dot{\epsilon}, \ell = 1, 2, \dots, L, \end{cases} \quad (1.19)$$

where M_u is the unrelaxed modulus, ξ_ℓ is the memory variable independent of the model properties to delineate the historical energy dissipation associated with ℓ -th circular frequency ω_ℓ and the dimensionless anelastic coefficients Y_ℓ . From eq 1.19, we may easily find that moving

from elastics to viscoelastics is simply subtracting the weighted memory variables ξ_ℓ with the strain term $\dot{\epsilon}$ in lossless elastic wave equation. By analogy with the VTI acoustic wave equation, the VTI viscoacoustic system in 3D reads

$$\left\{ \begin{array}{l} \partial_t \underbrace{\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{xz} \\ \sigma_{yz} \end{bmatrix}}_{\boldsymbol{\sigma}} = \underbrace{\begin{bmatrix} c_{11} & c_{11} & c_{13} & 0 & 0 & 0 \\ \cdot & c_{11} & c_{13} & 0 & 0 & 0 \\ \cdot & \cdot & c_{33} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & SYM & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}}_{\boldsymbol{C}} \underbrace{\begin{bmatrix} \dot{\epsilon}_{xx} - \sum_{\ell=1}^L Y_\ell \xi_\ell^{xx} \\ \dot{\epsilon}_{yy} - \sum_{\ell=1}^L Y_\ell \xi_\ell^{yy} \\ \dot{\epsilon}_{zz} - \sum_{\ell=1}^L Y_\ell \xi_\ell^{zz} \\ 2\dot{\epsilon}_{xy} - \sum_{\ell=1}^L Y_\ell \xi_\ell^{xy} \\ 2\dot{\epsilon}_{xz} - \sum_{\ell=1}^L Y_\ell \xi_\ell^{xz} \\ 2\dot{\epsilon}_{yz} - \sum_{\ell=1}^L Y_\ell \xi_\ell^{yz} \end{bmatrix}}_{\dot{\epsilon} - \sum_{\ell=1}^L Y_\ell \xi_\ell} \\ \partial_t \underbrace{\begin{bmatrix} \xi_\ell^{xx} \\ \xi_\ell^{yy} \\ \xi_\ell^{zz} \end{bmatrix}}_{\boldsymbol{\xi}} + \omega_\ell \underbrace{\begin{bmatrix} \xi_\ell^{xx} \\ \xi_\ell^{yy} \\ \xi_\ell^{zz} \end{bmatrix}}_{\boldsymbol{\xi}} = \omega_\ell \underbrace{\begin{bmatrix} \dot{\epsilon}_{xx} \\ \dot{\epsilon}_{yy} \\ \dot{\epsilon}_{zz} \end{bmatrix}}_{\dot{\epsilon}} \quad \ell = 1, 2, \dots, L, \end{array} \right. \quad (1.20)$$

where $Y_\ell := Y_\ell^\kappa$ the shear modulus μ (and the corresponding anelastic coefficients) are forced to be zero in the elastic case in order to perform the acoustic simplification. Allowing for the Newton's law in eq 1.5, the VTI viscoacoustic wave equation is now defined as

$$\left\{ \begin{array}{l} \rho \partial_t v_x = \partial_x \sigma_{xx} \\ \rho \partial_t v_y = \partial_y \sigma_{yy} \\ \rho \partial_t v_z = \partial_z \sigma_{zz} \\ \partial_t \sigma_{xx} = c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{11}(\xi_\ell^{xx} + \xi_\ell^{yy}) + c_{13} \xi_\ell^{zz}] \\ \partial_t \sigma_{yy} = c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{11}(\xi_\ell^{xx} + \xi_\ell^{yy}) + c_{13} \xi_\ell^{zz}] \\ \partial_t \sigma_{zz} = c_{13}(\partial_x v_x + \partial_y v_y) + c_{33} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{13}(\xi_\ell^{xx} + \xi_\ell^{yy}) + c_{33} \xi_\ell^{zz}] \\ \partial_t \xi_\ell^{xx} + \omega_\ell \xi_\ell^{xx} = \omega_\ell \partial_x v_x, \ell = 1, 2, \dots, L \\ \partial_t \xi_\ell^{yy} + \omega_\ell \xi_\ell^{yy} = \omega_\ell \partial_y v_y, \ell = 1, 2, \dots, L \\ \partial_t \xi_\ell^{zz} + \omega_\ell \xi_\ell^{zz} = \omega_\ell \partial_z v_z, \ell = 1, 2, \dots, L. \end{array} \right. \quad (1.21)$$

Similar to VTI acoustic case, based upon the notation in eq 1.15, the VTI viscoacoustic wave equation including the external sources can then be expressed as

$$\left\{ \begin{array}{l} \rho \partial_t v_x = \partial_x g + f_{v_x} \\ \rho \partial_t v_y = \partial_y g + f_{v_y} \\ \rho \partial_t v_z = \partial_z q + f_{v_z} \\ \partial_t g = c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{11}(\xi_\ell^{xx} + \xi_\ell^{yy}) + c_{13} \xi_\ell^{zz}] + f_g \\ \partial_t q = c_{13}(\partial_x v_x + \partial_y v_y) + c_{33} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{13}(\xi_\ell^{xx} + \xi_\ell^{yy}) + c_{33} \xi_\ell^{zz}] + f_q \\ \partial_t (\xi_\ell^{xx} + \xi_\ell^{yy}) + \omega_\ell (\xi_\ell^{xx} + \xi_\ell^{yy}) = \omega_\ell (\partial_x v_x + \partial_y v_y), \ell = 1, 2, \dots, L \\ \partial_t \xi_\ell^{zz} + \omega_\ell \xi_\ell^{zz} = \omega_\ell \partial_z v_z, \ell = 1, 2, \dots, L. \end{array} \right. \quad (1.22)$$

Note that the memory variables in the symmetric plane (x, y) , namely ξ_ℓ^{xx} and ξ_ℓ^{yy} , are summed together as one quantity to reduce the volume of memory allocation in practical implementa-

tion. With the notation $\xi_\ell^{xx} + \xi_\ell^{yy} := \xi_\ell^g$, $\xi_\ell^{zz} := \xi_\ell^q$, system 1.22 becomes

$$\begin{cases} \rho \partial_t v_x = \partial_x g + f_{v_x} \\ \rho \partial_t v_y = \partial_y g + f_{v_y} \\ \rho \partial_t v_z = \partial_z q + f_{v_z} \\ \partial_t g = c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{11} \xi_\ell^g + c_{13} \xi_\ell^q] + f_g \\ \partial_t q = c_{13}(\partial_x v_x + \partial_y v_y) + c_{33} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{13} \xi_\ell^g + c_{33} \xi_\ell^q] + f_q \\ \partial_t \xi_\ell^g = -\omega_\ell \xi_\ell^g + \omega_\ell (\partial_x v_x + \partial_y v_y), \ell = 1, 2, \dots, L \\ \partial_t \xi_\ell^q = -\omega_\ell \xi_\ell^q + \omega_\ell \partial_z v_z, \ell = 1, 2, \dots, L. \end{cases} \quad (1.23)$$

1.2.1.4 Discretization and absorbing boundary conditions

The wave equation is a linear partial differential equation and can be described and solved with various schemes (Kelly et al., 1976; Marfurt, 1984; Virieux, 1984; Dablain, 1986; Levander, 1988; Brossier et al., 2008). The choice of the schemes and the choice of discretization will influence the numerical resolution in which the wavefield is resolved.

One of the most popular methods used to solve the wave equation relies on finite-difference schemes. The main advantages of this method are the good scalability, effectiveness, and simplicity of implementation.

As this manuscript focuses on more robust FWI, and as the applications will be performed on marine field data, the excellent scalability of finite-difference approaches will be needed to tackle large 3D field data. Therefore, this work relies on using a code based on finite-difference, based on staggered-grid methods (Virieux, 1986; Levander, 1988). The discretization is performed with fourth-order in space and second-order in time. To localize sources and receivers that do not directly fall on the finite-difference grid, we rely on Hicks interpolation (Hicks, 2002). This allows recasting the position on the grid using windowed sinc functions.

Finally, artificial reflections generated at the boundary are mitigated by modifying the equations at a continuous level. In this work, we rely on absorbing boundary conditions, more precisely, perfectly matched layers (Bérenger, 1994) for 2D acoustic FWI. For 3D visco-acoustic, we rely on “sponge” absorbing boundaries as PML would be challenging to implement in this case (Cerjan et al., 1985).

1.2.2 Solution to the FWI problem

Solving the minimization problem of FWI can be done in two ways. As developed later on in this manuscript, the misfit function conventionally used for FWI - the least-squares norm - is non-convex and presents local minima. Therefore, one of the first ideas that would come to the mind would be relying on global optimization to solve the FWI problem. This would ensure convergence toward the global minimum. The second idea that can be used is to rely on local optimization. Details of the two different approaches will be introduced in the following, with their advantages and disadvantages.

1.2.2.1 Global solution to the FWI problem

This solution was explored in the early years of FWI formulation (Sen and Stoffa, 1991; Stoffa and Sen, 1991) but also more recently (Tran and Hiltunen, 2011; Datta and Sen, 2016; Aleardi and Mazzotti, 2016; Mazzotti et al., 2016; Sajeve et al., 2017).

Global optimization method aims at finding the global minima or maxima of a continuous function (here a misfit function). The function can possibly be non-convex.

$$f[m] : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}, \quad (1.24)$$

with Ω is the (not necessarily convex) solution space, discretized over n parameters.

By handling potentially non-convex misfit function, global optimization methods can naturally deal with local minima in Ω . Therefore these methods naturally allow tackling the ill-posed inverse problem that FWI represents. Global solutions aim to explore the solution space Ω to evaluate all possible minima (or maxima) and find the global one. This could be illustrated in the example shown in Figure 1.3, global optimization method would allow exploring the solution space Ω entirely and find the possible solutions (here, the red dots) while avoiding the local minimum represented by the blue dot.

The main disadvantage of global search approaches is the difficulty faced when tackling problems with a large number of degrees of freedom. Recent applications using Hamiltonian Markov-Chain Monte-Carlo (H-MCMC) tackle problems with up to thousands of degrees of freedom (Fichtner et al., 2018a,b; Gebraad and Fichtner, 2018; Gebraad et al., 2020). The issue is that even small 2D synthetic studies correspond to tens to hundreds of thousands of degrees of freedom. This is known as the “curse of dimensionality”. Indeed, the computational cost of this method is directly linked with the solution space Ω , which is directly related to the number of degrees of freedom representing the model. Realistically, and considering the trend of moving toward high-frequency 3D FWI application, global optimization seems too computationally intensive to be used. Despite this issue, the method remains of interest for FWI as it could be used for uncertainty estimation (Martin et al., 2012; Bardsley et al., 2014; Biswas and Sen, 2017; Sajeve et al., 2017; Thurin, 2020) and low-frequency inversion.

1.2.2.2 Local solution to the FWI problem

For problem space larger than 10^1 to 10^3 , the solution used to solve the FWI inverse problem relies on local optimization techniques. This method aims at finding the closest minimum for a given starting point, using local information provided by the misfit function (such as the gradient and curvature).

This highlights a crucial parameter: the choice of initial model m_0 used to start the inversion. This initial model m_0 defines a subspace $\mathcal{A} \subset \Omega$, in which the misfit function is minimized. This drastically reduces the number of potential solutions but also implies that only the closest minimum is reached. Therefore a *good enough* initial model is needed to allow convergence of the scheme toward the global minimum if the misfit function used is not globally convex.

The formalism consists in iteratively minimizing the discrepancy between observed and synthetic data by building the following sequence of models

$$m_{k+1} = m_k + \alpha_k \Delta m_k, \quad (1.25)$$

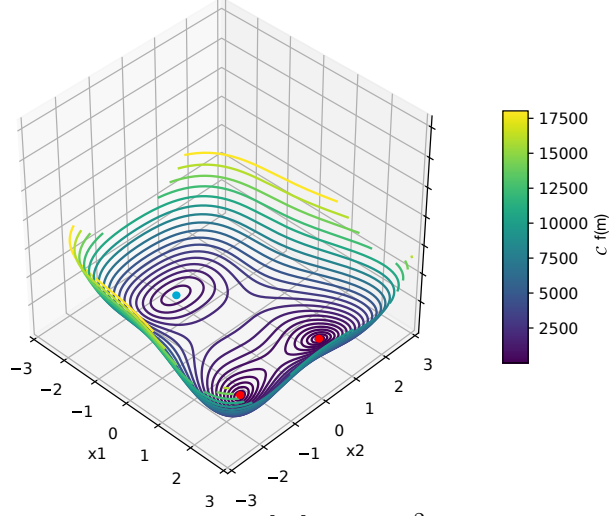


Figure 1.3: Representation of misfit function $f[m] : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ presenting three local minima, two of which are global and denoted by a red dot and a third one denoted by a blue dot. From Thurin (2020).

where m_0 the initial model, α is the step length, Δm is the model update (which corresponds to the descent direction) and k the index of the current iteration.

At each iteration k , both the step length α and the direction of descent Δm are updated to provide a better data fit

$$f[m_{k+1}] < f[m_k], \quad (1.26)$$

The steplength α should satisfy the Wolfe criterion (Wolfe, 1969) which are

1. Sufficient decrease

$$f[m + \alpha \Delta m] \leq f[m] + c_1 \alpha \nabla f[m]^T \Delta m, \quad (1.27)$$

where $\nabla f[m]$ the gradient. Practically c_1 is set to 10^{-4} in practice.

2. Curvature condition

$$\nabla f[m + \alpha \Delta m]^T \Delta m \geq c_2 \nabla f[m]^T \Delta m, \quad (1.28)$$

with $c_2 = 0.9$ in practice.

Satisfying the Wolfe conditions ensures the convergence towards the nearest local minimum (Nocedal and Wright, 2006).

To build the model update Δm_k , we usually rely on an approximate solution of the Newton equation

$$H[m] \Delta m = -\nabla f[m], \quad (1.29)$$

where $H[m]$ is the Hessian operator, the matrix of second-order derivatives of the function $f[m]$. As the computational cost of inverting for the Hessian operator is too high, we rely on an approximation P :

$$P[m] \simeq H[m]^{-1}. \quad (1.30)$$

If the approximation P is the identity, we have what is referred to as *steepest-descent* direction, which simply corresponds to taking the opposite of the gradient to as model update

$$\Delta m = -\nabla f[m]. \quad (1.31)$$

This steepest-descent algorithm is known to be slow and exhibits convergence difficulties even on elementary problems.

To avoid this issue, we rely on quasi-Newton methods that build an approximation of the inverse Hessian for P . The most famous of this method is called *l*-BFGS and build the approximation through the stored gradient and model values (Nocedal, 1980; Nocedal and Wright, 2006). This method will be the one used in this thesis.

1.2.2.3 Gradient computation

From the description of the quasi-Newton scheme, we see that in practice the critical quantity to compute is the gradient of $f[m]$. In this section, we explain how we can compute it. We define the number of sources (shots) as N_s (with s a subscript referring to shot number) and the number of receivers as N_r . We assume that the number of receivers is the same for each shot.

First, we consider the case of the least-squares norm:

$$\min_m f[m] = \sum_{s=1}^{N_s} F(d_{cal,s}[m], d_{obs,s}), \quad (1.32)$$

with

$$F(d_{cal,s}, d_{obs,s}) = \sum_{r=1}^{N_r} \int_0^T |d_{cal,s}[m](x_r, t) - d_{obs,s}(x_r, t)|^2 dt \quad (1.33)$$

$$\stackrel{\text{def}}{=} \|d_{cal,s} - d_{obs,s}\|^2. \quad (1.34)$$

We remind that

$$d_{cal,s}[m] = Ru_s[m], \quad (1.35)$$

and $u_s[m]$ solution of

$$A(m)u_s = b_s. \quad (1.36)$$

For convenience, we now consider a simple case with $N_s = 1$. For a perturbation dm we have

$$\begin{aligned} f[m + dm] &= \frac{1}{2} \|d_{cal}(m + dm) - d_{obs}\|^2 \\ &= \frac{1}{2} \|d_{cal}(m) - d_{obs} + J(m)dm + o(\|dm\|^2)\|^2, \end{aligned} \quad (1.37)$$

where

$$J(m) = \frac{\partial d_{cal}}{\partial m} = \begin{pmatrix} \frac{\partial(d_{cal})_1}{\partial m_1} & \frac{\partial(d_{cal})_1}{\partial m_2} & \cdots \\ \frac{\partial(d_{cal})_2}{\partial m_1} & \frac{\partial(d_{cal})_2}{\partial m_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (1.38)$$

is the Jacobian matrix. It can be used to re-write $f[m + dm]$ as

$$\begin{aligned} f[m + dm] &= \frac{1}{2} \|d_{cal}(m) - d_{obs}\|^2 + \langle d_{cal} - d_{obs}, J(m)dm \rangle + o(\|dm\|^2) \\ &\quad \frac{1}{2} \|d_{cal}(m) - d_{obs}\|^2 + \langle J(m)^T (d_{cal} - d_{obs}), dm \rangle + o(\|dm\|^2) \end{aligned} \quad (1.39)$$

Therefore

$$f(m + dm) - f(m) = \langle J(m)^T (d_{cal} - d_{obs}), dm \rangle + o(\|dm\|^2), \quad (1.40)$$

and

$$\nabla f(m) = J(m)^T (d_{cal} - d_{obs}). \quad (1.41)$$

This corresponds to the “direct” method to obtain the gradient. Unfortunately, this method relies on a Jacobian matrix which is computationally too intensive to compute or too large to be stored (I/O storage is too slow).

Therefore, we rely on the adjoint-state approach (Plessix, 2006). This method relies on using the Lagrangian functional. The main goal of the adjoint state method is to compute the gradient of a functional, $f[m]$, where f depends on $u(m)$, without explicitly computing the Jacobian matrix J .

We introduce the Lagrangian operator in the case of $N_s = 1$ with an arbitrary distance measurement function F :

$$L(m, u, d_{cal}, \lambda, \mu) = F(d_{cal}, d_{obs}) + \langle A(m)u - b, \lambda \rangle + \langle d_{cal} - Ru, \mu \rangle. \quad (1.42)$$

Let $\bar{u}(m)$ be the solution of the forward problem for a given m (eq 1.36). Similarly, if d_{cal} is extracted from $\bar{u}(m)$, we denote it by $\bar{d}_{cal}(m)$ (eq 1.35). We thus have

$$\mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu) = F(\bar{d}_{cal}(m), d_{obs}) = f[m]. \quad (1.43)$$

Therefore, by deriving the previous equation with respect to the model parameter m we obtain an expression of the gradient of $f[m]$ as:

$$\frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial m} = \nabla f[m]. \quad (1.44)$$

Using the chain rule on the Lagrangian, we obtain

$$\left(\frac{\partial A(m)}{\partial m} \bar{u}(m), \lambda \right) + \frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial u} \frac{\partial \bar{u}(m)}{\partial m} + \frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial d_{cal}} \frac{\partial \bar{d}_{cal}(m)}{\partial m} = \nabla f[m]. \quad (1.45)$$

We can identify in $\frac{\partial \bar{u}(m)}{\partial m}$ and $\frac{\partial \bar{d}_{cal}(m)}{\partial m}$ operators similar to the Jacobian one. As we do not want to compute such operators (it was to avoid them that we introduce the Lagrangian), we define λ and μ such as

$$\frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial u} = 0, \quad (1.46)$$

$$\frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial d_{cal}} = 0. \quad (1.47)$$

To obtain λ and μ that satisfies the two latter equations, we express the derivatives of the Lagrangian with respect to u and d_{cal} using first order Taylor developments. We have

$$\frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial u} = A(m)^T \lambda - R^T \mu, \quad (1.48)$$

and similarly

$$\frac{\partial \mathcal{L}(m, \bar{u}(m), \bar{d}_{cal}(m), \lambda, \mu)}{\partial d_{cal}} = \frac{\partial F}{\partial d_{cal}} - \mu. \quad (1.49)$$

This provides an equivalent to the system 1.46 as:

$$A(m)^T \lambda - R^T \mu = 0 \quad (1.50)$$

$$\frac{\partial F}{\partial d_{cal}} - \mu = 0, \quad (1.51)$$

which can be rewritten as

$$A(m)^T \lambda = R^T \mu \quad (1.52)$$

$$\mu = \frac{\partial F}{\partial d_{cal}}, \quad (1.53)$$

with the index T representing transpose. From the two latter equations, we can now extract conditions to cancel the terms in the gradient expression 1.45 which are:

1. μ is the adjoint source, which is just the derivative with respect to d_{cal} of the distance measurement function F
2. λ is the solution of an adjoint equation associated with the operator $A(m)^T$ with a source term equal to $R^T \mu$

This finally allows to have the expression of the gradient relying only on the correlation part as

$$\left\langle \frac{\partial A}{\partial m} u, \lambda \right\rangle = \nabla f[m], \quad (1.54)$$

with λ the adjoint field

$$A(m)^T \lambda = -R^T \frac{\partial F}{\partial d_{cal}}. \quad (1.55)$$

A very interesting property of this method is that a change in the misfit function *only* translates into a modification of the source term used for the adjoint field back-propagated from the receivers.

The gradient computation requires solving two wave propagation problems: one incident field and one adjoint field. This translates into a reduction of the computational cost compared to forming or storing the Jacobian matrix.

1.3 Non-convexity in FWI

1.3.1 Non-convexity and cycle skipping limitations

1.3.1.1 The least-squares norm misfit function

Since the introduction of the FWI formalism in the '80s, the least-squares norm - referred to as the L^2 norm - has been traditionally used as the “default” norm to measure the distances between the observed data and the calculated data (Tarantola, 1984a, 1988; Nolet, 1987).

Let us remind the formulation of the L^2 norm already introduced in eq 1.33 for an isolated traces such as $N_r = N_s = 1$

$$F_{L^2}(d_{cal}, d_{obs}) = \frac{1}{2} \int_0^T |d_{cal}(t) - d_{obs}(t)|^2 dt. \quad (1.56)$$

As demonstrated earlier, FWI formalism is based on the adjoint theory, requiring an adjoint source associated with the misfit function. For the L^2 norm, the adjoint source S is defined as

$$S_{L^2}(t) = \frac{\partial F_{L^2}}{\partial d_{cal}(t)} \quad (1.57)$$

$$= \frac{\partial}{\partial d_{cal}(t)} \left[0.5 \times (d_{cal}(t) - d_{obs}(t))^2 \right]. \quad (1.58)$$

By using the chain rule, we obtain the adjoint source

$$S_{L^2}(t) = 0.5 \times 2 \times (d_{cal}(t) - d_{obs}(t)) \times \frac{\partial(d_{cal}(t) - d_{obs}(t))}{\partial d_{cal}(t)} \quad (1.59)$$

$$= d_{cal}(t) - d_{obs}(t). \quad (1.60)$$

In the case of the L^2 norm, the obtained adjoint source is simply the difference between d_{cal} and d_{obs} . This is obviously straightforward to implement, it is not affected by any stability condition and associated computational cost is close to zero.

The L^2 misfit function presents several advantages, such as noise robustness and “usage” of the complete signal information. This explains why it is still widely used, and it is still the “default” choice for field data application. The main limitation of the L^2 norm is its lack of convexity for time-delays, leading to cycle-skipping being one of its most substantial weaknesses.

1.3.1.2 The cycle-skipping limitation

The L^2 distance measure is inherently affected by the phase ambiguity, known as cycle-skipping, which appears when the synthetic data are shifted by more than half a period compared to the observed ones (Figure 1.4). Unfortunately, finding *perfectly* convex misfit functions is challenging, resulting in FWI being strongly non-linear when based on a non-convex misfit function (such as the least-square norm conventionally used). This leads to the common issue of converging toward a non geologically informative local minimum, an issue that has been

faced since the beginning of FWI applications (Gauthier et al., 1986), and which translates into an erroneous reconstruction of the velocity model (Virieux and Operto, 2009). This non-convexity with respect to time-shifts link with Claerbout’s “gap” between seismic imaging methods. Indeed, if the misfit function could be made *perfectly* convex, it would allow starting from a low-wavenumber initial model and iteratively build high-wavenumber updates, thereby *bridging the gap*.

The non-convexity of the L^2 misfit function with respect to time-shifts can be illustrated easily in a straightforward and schematic example based on time-shifted Ricker wavelets. Let us introduce a reference signal composed of one Ricker wavelet in the center, seen as the observed data. Then, a second signal is created from a shifted in time version of the reference signal (with a time-shift going from -1.5 to 1.5 s). The distance between the reference signal and the calculated signal is then obtained using the L^2 norm. Results are presented in Figure 1.5. In this example, it can be observed that from time-shift superiors to ± 0.12 s, the L^2 norm is not convex and presents local minimum and a flat part for time-shift superior to ± 0.4 s. This non-convex behavior of the L^2 norm for time shift superiors to half the signal period corresponds to the so-called cycle-skipping issue. As FWI relies on local optimization schemes to be solved, converging toward a local minimum implies an erroneous reconstructed V_P model, with no possibilities to improve it.

Practically, the origin of time shifts between observed and calculated data can be easily tracked down to erroneous low wavenumber content in the model, which generally relates to the quality of the initial model. What could be referred to as a “good” initial model will predict the data within less than half a period, allowing the L^2 -based FWI scheme to converge toward the global minimum. A degradation in the initial model’s quality with erroneous low-wavenumber background velocities will likely translate into cycle-skipped arrivals, leading to artifacts during the reconstruction of the V_P model. This illustrates the need for a good initial model already

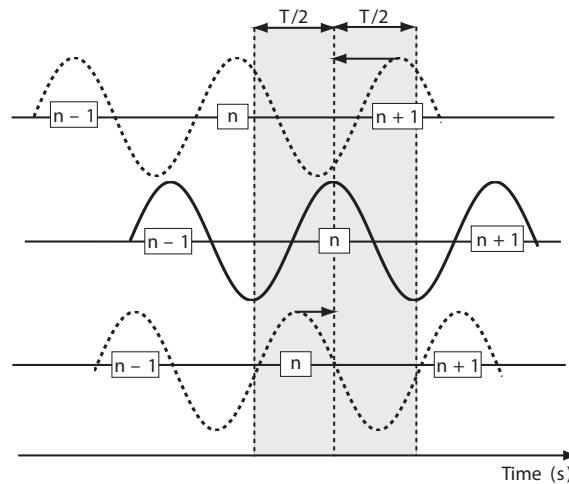


Figure 1.4: Schematic example of the cycle-skipping/phase-ambiguity issue on sinusoidal signals. As soon as the initial shift is larger than half a period, the signal’s fit using a least-squares distance is performed up to one or several phase shifts. One may try to fit the $n + 1$ dashed wriggle of the top signal with the n continuous wriggle of the middle signal moving in the wrong direction. The bottom dashed signal predicts the n wriggle in less than a half-period leading to a correct updating direction (from Virieux and Operto, 2009).

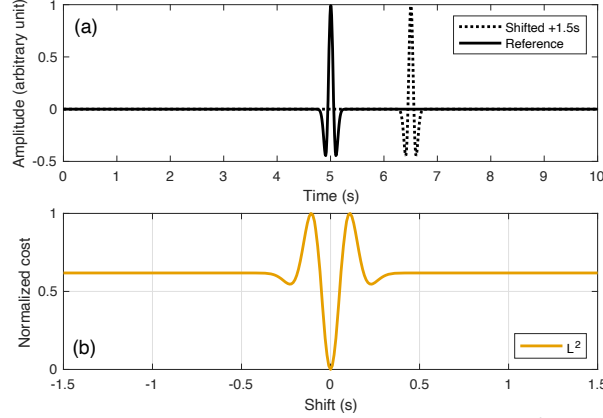


Figure 1.5: Simple time-shifted Ricker wavelet (4 Hz) test for L^2 misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the L^2 norm with respect to time-shift.

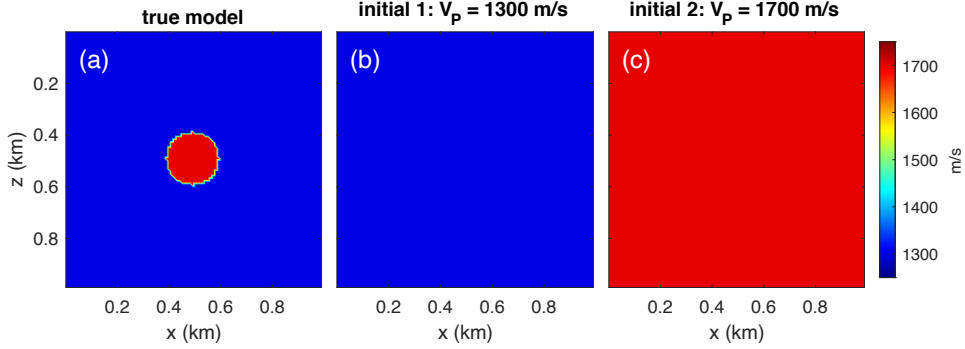


Figure 1.6: FWI schematic example for cycle skipping: (a) true model with spherical inclusion, (b) homogeneous initial model 1, (c) homogeneous initial model 2.

inside the global valley when starting the FWI process.

The link between the initial model “quality” and correct FWI reconstruction can be easily illustrated, as shown in a canonical test based on a spherical inclusion with a cross-hole acquisition (sources on the left, receivers on the right). The target is a simple square model of 1 by 1 km, of homogeneous $V_P = 1300 \text{ m.s}^{-1}$, with a spherical inclusion in the center of 100 m radius with $V_P = 1700 \text{ m.s}^{-1}$ (Figure 1.6 (a)). Two initial models are used, both being completely homogenous V_P , with one at the correct background velocity of the true model (1300 m.s^{-1}), while the second one is at 1700 m.s^{-1} (respectively (b) and (c) in Figure 1.6). The data are generated with acoustic modeling using a Ricker wavelet with a central frequency of 3 Hz. The difference between d_{obs} and the two d_{cal} (one for each initial model) is presented in Figure 1.7. It clearly shows that the 2nd initial model, containing a faster V_P , makes the signal arrive “too soon” compared to the observed data, creating out of phases arrivals by more than half a period. The results obtained with FWI starting from each initial model are presented in Figure 1.8. With no surprise, the reconstruction is correct for initial model 1, while results from initial model 2 present strong high-velocity artifacts, with no reconstruction of the spherical inclusion and no correct background velocity. This exhibits the limitation of L^2 based FWI for inaccurate initial model.

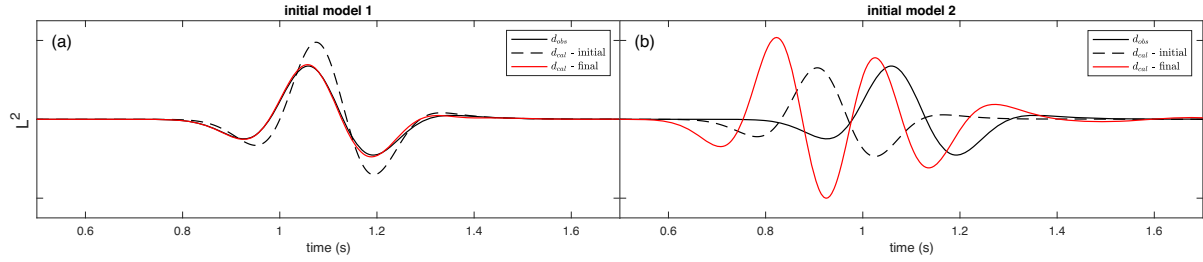


Figure 1.7: FWI schematic example for cycle skipping: traces for a source-receiver couple at $z = 500$ m for d_{obs} (solid black), d_{cal} in the initial model (dashed black) and d_{cal} in final reconstructed V_P model (solid red). (a) correspond to initial model 1, (b) to initial model 2.

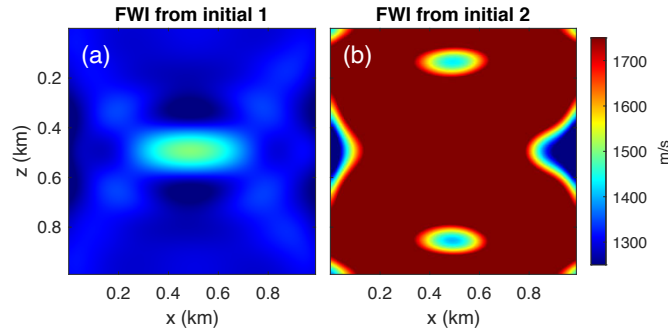


Figure 1.8: FWI schematic example for cycle skipping: reconstructed V_P models using FWI starting from (a) homogeneous initial model 1, (b) homogeneous initial model 2.

1.3.2 Classical approaches

The ideas developed here are solutions to successfully apply L^2 -based FWI, relying on data and initial models available. This is often the combination of the following elements.

1.3.2.1 Starting from a good initial model

The first condition to avoid convergence toward a local minimum is to start FWI from an initial model m_0 , which is inside the global minimum basin of attraction. Traditionally, FWI relies on using a *good enough* initial model that contains a correct approximation of the low wavenumber content of the model. Usually, such models come from reflection tomography or stereotomography (Lambaré, 2008). Unfortunately, these methods are based on travel-time or reflected events picking, which is a time-consuming task that requires extensive human expertise. Moreover, picking can introduce errors and uncertainties which will be propagated through FWI. This again links to the famous “gap” as we attempt to bring higher wavenumber content from tomography methods.

1.3.2.2 Data hierarchy

Complementary to the requirement of a good initial model, it is often required to select “smartly” the data for inversion, following an approach known as data hierarchy (Bunks et al., 1995; Pratt, 1999; Shipp and Singh, 2002; Wang and Rao, 2009; Brossier et al., 2009). As low-frequency data are less subject to cycle-skipping (as illustrated in Figure 1.9), the idea is

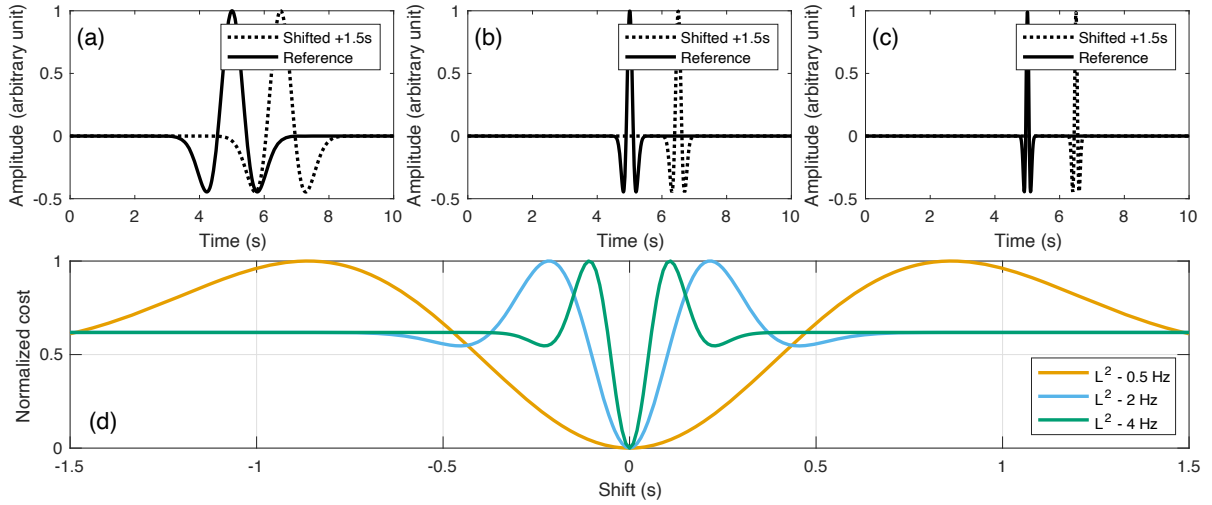


Figure 1.9: Same 1D setup as Figure 1.5, but this time comparing the influence of frequency on the convexity. (a) 0.5 Hz Ricker wavelet, (b) 2 Hz Ricker wavelet, (c) 4 Hz Ricker wavelet. (d) L^2 misfit function values with respect to time-shifts for the three different frequency cases. The width of the basin of attraction of the misfit function is enlarged for lower frequencies.

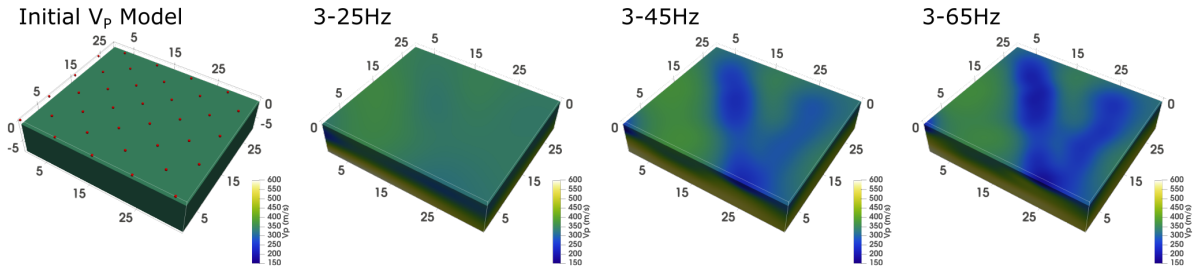


Figure 1.10: Illustration of frequency-continuation approach used in FWI; here, on a shallow near-surface target of 30 m (from Irnaka, 2021). FWI is started on a 3 to 25 Hz frequency band, then higher frequency are progressively introduced with an upper limit of 45 Hz then finally 65 Hz. A clear improvement in resolution (smaller wavenumber content) is observed for higher frequency data.

to start the inversion using the lowest available frequencies while restricting higher frequency. Then, higher frequencies are progressively introduced following a multi-scale approach as shown in Figure 1.10 (Sirgue and Pratt, 2004).

Another level of data hierarchy consists of restricting the temporal/offset selection of the data to reduce the number of propagated wavelengths interpreted. By doing so, the shallow part of the reconstructed models can be recovered first, while later on, deeper updates are progressively introduced (Shipp and Singh, 2002; Brossier et al., 2009; Sears et al., 2010; Kamei et al., 2013; Górszczyk et al., 2017). We use this strategy for the field-data application in this thesis, and an illustration can be found in Figure 3.11. Practically, these two levels of data hierarchy are often used together to obtained successful FWI applications on field data, no matter what other methods are used to tackle the non-linearity of FWI.

1.3.3 Reformulating the FWI problem

Nonetheless, even if the conditions details previously should allow successful FWI applications, they are not easy to gather or can even be insufficient to tackle some specific cases (such as sub-salt imaging). Therefore, finding other solutions to mitigate the non-linearity of FWI has been the motivation for a large number of studies in the past decades. Two main lines of investigation can be identified, which both boil down to reformulating the FWI problem.

1.3.3.1 Extension strategies

One way of seeing the non-linearity of FWI is that FWI is based on a reduced-space approach used to solve the PDE-constrained optimization problem, which is non-linear. Therefore, to mitigate this non-linearity, adding artificial degrees of freedom to the problem has been proposed. This corresponds to the first line of investigation and regroups methods often referred to as “extension strategies”. The philosophy behind this is to *artificially* introduce supplementary degrees of freedom to the FWI problem in the hope to match data in early iterations (and therefore reduce cycle-skipping sensitivity). To ensure convergence, these artificial degrees of freedom are constrained to converge to zero to recover a correct and physical velocity reconstruction.

These methods have been historically developed with model-based extension, known as migration velocity analysis (MVA) methods (Symes, 2008). MVA is based on scale separation assumption: subsurface parameters can be decomposed into two parts, a smooth macro-velocity background model and a sharp high-wavenumber reflectivity model. Using this scale separation, MVA introduces artificial degrees of freedom in the reflectivity model. The goal is to update the macro-velocity model to focus the energy of the *extended* reflectivity model at zero in the artificial dimension. Mathematical analysis shows that in a transmission regime, the new extended problem of FWI asymptotically converges toward a travel-time tomography problem known to be convex. Unfortunately, MVA suffers from two main drawbacks. The first one relates to the high computational cost of building extended reflectivity hypercubes. The second one corresponds to a more fundamental difficulty and is related to the difficulty faced by MVA to handle properly complex data with multi-arrival and multiple reflections (Cocher et al., 2017).

More recently, as an attempt to solves the issue affecting MVA, a new class of extension strategies has emerged, with a different way of introducing artificial degrees of freedom, for example, at the source level with source extension strategy. These methods are referred to as matched source waveform inversion (MSWI) (Huang et al., 2018a,b). These technics have shown promising results in 2D synthetic cases using frequency-domain FWI. Still, their implementation for time-domain FWI is under development (Wang et al., 2016; Aghamiry et al., 2020). The reason behind the preference for frequency-domain FWI for MSWI boils down to the fact that MSWI requires the solution of a square wave propagation problem that is easily obtained through the harmonic wave equation based on a direct solver, which is much more difficult to obtain for time-domain time-stepping algorithms. This represents a disincentive to 3D field data application as time-domain FWI remains the methods of choice for this kind of targets (frequency-domain FWI is for now limited to moderate-size targets as there is a lack of scalability of the direct solvers used to solve harmonic equations, see Li et al. (2020) for a recent review on the status of direct solvers to solve large scale harmonic wave equation problems).

MSWI methods can be seen as an equivalent of a class of methods that have been previously introduced to FWI: wavefield reconstruction inversion (WRI) method (van Leeuwen and Herrmann, 2013, 2016; Aghamiry et al., 2019). A parallel can also be drawn with deconvolution approaches, with one of the most documented ones being adaptive waveform inversion (AWI) by Warner and Guasch (2016), which can be recast as an MSWI technique. This is interesting as AWI can also be seen as an alternative misfit function (which will be the second line of investigation for reformulation of the FWI problem), indicating that the distinction between the two families is not as straightforward as one could think.

Finally, we can note that another type of extension strategy can be made with the extension being made at the receiver location (Métivier and Brossier, 2020). This new formulation bypasses the needs for square wave propagation operators and is therefore directly applicable to time-domain FWI with reasonable computational cost. This method has, however, not yet been tested on field data so far.

1.3.3.2 Alternative misfit functions

Another way of seeing the non-linearity of FWI relates to how the data is interpreted. As low-wavenumber content of the initial model *mainly* generates kinematic shifts in the data (time-shifts), the non-convexity of the least-squares (L^2) misfit function with respect to time-shifts partially explains the non-linearity of FWI (Jannane et al., 1989). This non-convexity with respect to time-shifts of the L^2 norm is the cause of the well-documented cycle-skipping issue.

Hence, the second line of investigation to mitigate the non-linearity of FWI is to reformulate FWI by replacing the L^2 misfit function used to compare the observed and calculated data with a more convex alternative. This idea is at the origin of a large variety of more-convex alternative misfit functions that have been proposed in the literature. This will be the subject of the analysis performed in the next chapter.

Chapter 2

On cycle-skipping and misfit functions modification for full-wave inversion: comparison of five recent approaches

Contents

2.1	Introduction	31
2.2	General FWI framework and misfit function formulation	35
2.2.1	General framework	35
2.2.2	Adaptive waveform inversion	36
2.2.3	Instantaneous envelope	37
2.2.4	Normalized Integration Method	38
2.2.5	Kantorovich-Rubinstein optimal transport distance	39
2.2.6	Graph space optimal transport distance	40
2.3	A simple convexity analysis based on time-shifted Ricker wavelets	41
2.4	FWI tests on two canonical examples	46
2.4.1	FWI Test 1: transmission configuration	46
2.4.2	FWI Test 2: reflection configuration	49
2.5	Marmousi case study: toward a more realistic case study	58
2.5.1	Common framework	58
2.5.2	Inverse crime inversion	58
2.5.3	A more realistic inversion	62
2.6	Discussion	73
2.7	Conclusion	73

This chapter directly comes from the published paper Pladys et al. (2021b).

abstract

Full waveform inversion, a high-resolution seismic imaging method, is known to require sufficiently accurate initial models to converge toward meaningful estimations of the subsurface mechanical properties. This limitation is due to the non-convexity of the least-squares distance with respect to kinematic mismatch. We propose a comparison of five misfit functions promoted recently to mitigate this issue: adaptive waveform inversion, instantaneous envelope, normalized integration, and two methods based on optimal transport. We explain which principles these methods are based on and illustrate how they are designed to better handle kinematic mismatch than a least-squares misfit function. By doing so, we can exhibit specific limitations of these methods in canonical cases. We further assess the interest of these five approaches for application to field data based on a synthetic Marmousi case study. We illustrate how adaptive waveform inversion and the two methods based on optimal transport possess interesting properties, making them appealing strategies applicable to field data. Another outcome is the definition of generic tools to compare misfit functions for full-waveform inversion.

2.1 Introduction

Full waveform inversion (FWI) is a high-resolution seismic imaging method dedicated to reconstructing the mechanical properties of the subsurface (Devaney, 1984; Pratt and Shipp, 1999; Plessix and Perkins, 2010; Raknes et al., 2015; Górszczyk et al., 2017). It is formulated as an iterative process based on minimizing a function measuring the misfit between observed and calculated data over a space of model parameters describing the subsurface. The resolution improvement FWI can procure, compared with standard tomography methods, is used to significantly improve depth-migration images or even produce directly interpretable quantitative estimates of the subsurface mechanical properties (Shen et al., 2018). FWI is applied at multiple scales, from global and regional scales in seismology to exploration scale for the oil & gas industry, and even, more recently, at near-surface scale for geotechnical applications. A thorough review of FWI and its applications can be found in Virieux et al. (2017).

FWI suffers from a significant shortcoming in its classical formulation: the non-convexity of the least-squares (L^2) misfit function on which it is conventionally based.

This non-convexity of the misfit function is an issue because the iterative process on which is based FWI is a local optimization algorithm. Standard size for realistic applications makes global optimization strategies beyond modern high-performance computing platforms current and predictable capabilities. Therefore, if the initial model used is too far away from the global minimum, FWI converges toward a potentially non geologically informative local minimum. This constraint leads to the need for an accurate enough initial model to ensure convergence toward the global minimum of the misfit function.

In a physical sense, the non-convexity of the L^2 misfit function is associated with a phenomenon known as cycle-skipping. It appears when the calculated data are shifted (in time) from more than half a period (corresponding to the signal dominant frequency) compared to the observed data. If the time-shift between observed and calculated data is larger than half a period, the minimization of the L^2 norm between the two signals will “skip” a phase and align the two signals on the closest phase (hence the name, cycle-skipping). This ambiguity translates into an erroneous reconstruction of the velocity model (Virieux and Operto, 2009).

This limitation of FWI has been documented since its origin (Gauthier et al., 1986). To address this limitation in practical cases, the workflow generally relies on data hierarchy (Bunks et al., 1995; Pratt, 1999; Shipp and Singh, 2002; Wang and Rao, 2009; Brossier et al., 2009). The historical approach consists in interpreting first the lowest frequency available (around 2 to 4 Hz for seismic exploration targets), then progressively introducing higher frequency data, following a multi-scale approach (Sirgue and Pratt, 2004). The lowest frequencies are, by definition, less subject to cycle-skipping. The second level of data hierarchy can then be defined by playing on temporal and/or offset selection of the data. The idea is to reduce the number of propagated wavelengths that are interpreted simultaneously, hence reducing the risk of cycle-skipping. In practice, this second level corresponds to first reconstructing the near-surface and progressively introducing deeper updates referred to as layer stripping approach.

Successful practical applications at the exploration scale often rely on the conjunction of these approaches as well as the design of an accurate initial starting model, obtained, for instance, through reflection tomography or stereotomography (Lambaré, 2008). Nonetheless, the conditions detailed previously to obtain a satisfactory FWI result are not always gathered. For instance, low-frequency data around 2 to 4 Hz are not always available or of sufficient quality. Moreover, obtaining low-frequency can increase the cost of acquisition, or can sometimes not be physically possible, or can even compromise the quality of the high frequency needed to obtain a very high resolution. Accurate initial model building can also be a time-consuming and challenging task requiring strong human expertise as it generally relies on tomography methods based on travel-time or reflected event picking. It also relies on prior information coming from geology or well logs; all of these require human expertise. This makes FWI less robust and reduces its range in terms of applications.

Mitigating the sensitivity to initial model quality has been the motivation for a large number of studies in the past decades. Two main lines of investigations can be identified, both leading to the reformulation of the conventional least-squares FWI problem.

Considering the first line, we regroup methods that can be cast under the frame of “extension strategies”. It is not our purpose to give an extensive overview of these methods here, but we try to sketch their main ingredients. The philosophy of extension strategies consists in introducing supplementary degrees of freedom to the FWI problem, which can match the data in the early iterations of the FWI process to avoid cycle-skipping. Relaxing iteratively the use of these artificial degrees of freedom should lead to a correct subsurface model estimation.

Historically, these methods derive from migration velocity analysis (MVA) (Symes, 2008). MVA relies on the scale separation assumption. The subsurface parameters to recover are decomposed as a smooth macro-velocity model and a high wavenumber content reflectivity model. Artificial degrees of freedom are introduced at the reflectivity level by introducing an extra dimension on offset, subsurface offset, or time-lag. The MVA problem is formulated as the iterative update of the macro-velocity model to focus the energy of the “extended” reflectivity model at zero in the artificial dimension. These methods have benefited from in-depth mathematical research work, leading to a clear understanding of their foundations, thanks to the theory of pseudo-differential operators. However, their application to field data is still limited, mainly because of two issues. First, the repeated construction of high-dimensional reflectivity cubes is computationally demanding. Second, the macro-velocity model construction through MVA is complicated as soon as complex data with multi-pathing and multiple reflections are considered.

More recently, another class of extension strategies has emerged. As opposed to model space extension, the artificial degrees of freedom are introduced at the source level, following a source extension strategy (Huang et al., 2018a; van Leeuwen and Herrmann, 2013). These methods have shown interesting promises in 2D synthetic case studies. However, their application to 3D field data seems still limited, mainly because of the difficulty of applying these methods in the time-domain. Current solutions either rely on relatively crude approximations (Wang et al., 2016) or on a sophisticated iterative solution, which increases the computational cost of the approach significantly (Aghamiry et al., 2020).

The second investigation line relies on reformulating the FWI problem using an alternative measure of the distance between observed and calculated data, namely a different misfit function. A large variety of approaches have been proposed on this framework. The first proposed along this line is to use cross-correlation measurements (Luo and Schuster, 1991), a strategy later revisited by van Leeuwen and Mulder (2010). The idea behind this is that cross-correlation should give access to the time-shifts between synthetic and observed traces. A misfit function based on the minimization of these time-shifts, resembling a tomography misfit function, should thus be less prone to cycle-skipping. The original approach of Luo and Schuster (1991) was labeled as “wave equation tomography” strategy.

However, when seismic traces contain multiple seismic events, the cross-correlation measurement might fail to give a correct estimation of a potential time-shift. This is why deconvolution based approaches have been later promoted, first by Luo and Sava (2011), then improved by Warner and Guasch (2016). The latter approach has been labeled as “adaptive waveform inversion” (AWI) and is based on a normalized deconvolution of the synthetic and observed seismic traces. It has shown very interesting properties both on synthetic and field data. The deconvolution of the traces yields a Wiener filter, which is then normalized and serves as an input for the misfit function. The misfit function penalizes the energy of the filter away from a bandpass Dirac filter, which would have been obtained in the correct subsurface model. Note that AWI shares some similarities with the extended source approach and can indeed be recast in the frame of these methods (Huang et al., 2018a). This indicates that the separation between extended methods and misfit function reformulation methods is not as watertight as one could think. Nevertheless, it is useful to draw a landscape of the investigations around the cycle-skipping issue in FWI.

Another family of misfit function modifications relies on transforming the signal itself prior to comparison through a least-squares distance. Extracting the instantaneous phase and envelope (Fichtner et al., 2008; Bozdağ et al., 2011) has been successfully used in seismology. The goal of the instantaneous phase is to avoid amplitude prediction issues, as earthquake source and receiver calibration are significant challenges in seismology. The use of the envelope to mitigate the cycle-skipping issue has also been developed in the framework of seismic exploration (Wu et al., 2014). An interesting alternative consists of using a normalized integration of the signal, namely the cumulative distribution of the traces. This approach has been promoted by Donno et al. (2013).

Finally, optimal transport distances have also been promoted to derive alternative misfit functions for FWI. The motivation is to benefit from the convexity of the optimal transport distance with respect to translation and dilation, which provides a misfit function convex with respect to time-shifts, this being a good proxy for convexity with respect to seismic velocities (Engquist and Froese, 2014; Métivier et al., 2018). The main difficulty in applying optimal transport in the framework of FWI is that the optimal transport theory is developed to compare

probability distributions, therefore positive functions with the same total integral. Seismic data do not fulfill this assumption.

To overcome this difficulty, different options have been promoted. For instance, one can rely on a prior transformation of the signal, such as extraction of positive and negative parts, squaring the data, affine scaling, exponential transform, softmax transform (Engquist and Froese, 2014; Qiu et al., 2017; Yang et al., 2018b; Yang and Engquist, 2018). This has been shown effective in some synthetic cases. However, relevant seismic information might be lost in the process of these transformations.

One solution is to rely on a specific optimal transport distance, which can be extended to comparing non-positive data. This is the Kantorovich-Rubinstein optimal transport (KROT) approach, which has been promoted in Métivier et al. (2016c,a,b), and which has been successfully applied to 3D synthetic elastic data (He et al., 2019b) as well as to field data (Poncet et al., 2018; Messud and Sedova, 2019; Sedova et al., 2019). One interest of this approach is its ability to account for lateral coherency in 2D or 3D shot gathers. One shortcoming is that, even if the valley of attraction is wider, compared with the L^2 approach, the convexity property of the optimal transport distance with respect to time-shifts is lost.

Another option has been promoted more recently. Considering each discrete seismic traces as point clouds and computing the optimal transport distance between synthetic and observed points clouds provide a new distance measurement. This specific optimal transport problem can be cast as a linear assignment problem, for which efficient solvers exist, for point clouds containing a few hundred to thousands of points, a situation we encounter for realistic scale exploration case studies (Métivier et al., 2018, 2019). The benefit of this graph-space optimal transport (GSOT) strategy is its ability to recover the convexity with respect to time-shifts. Compared with the KROT approach, GSOT is a trace-by-trace strategy that does not make it possible to account for lateral coherency. GSOT has been successfully applied to 3D synthetic and field data (He et al., 2019a; Pladys et al., 2019; Li et al., 2019; Górszczyk et al., 2019).

As can be seen, numerous investigations motivated by the inherent ill-posedness of the FWI problem have been lead in parallel. To our knowledge, no cross-comparison has been proposed so far, which is undoubtedly a lack. The first motivation of this study is to start developing tools that could be used to benchmark different FWI strategies. However, beyond a simple comparison of FWI strategies, we would like to highlight specific characteristics that an ideal misfit function should satisfy to render the FWI problem less ill-posed. Cycle-skipping is certainly an issue, but we also show that other criteria than robustness with respect to cycle-skipping should be considered, such as:

- sensitivity to the signal polarity;
- applicability in the framework of complex/multi-arrival data;
- number of tuning parameters and sensitivity to these parameters;
- sensitivity to wrong amplitude prediction and inaccurate wavelet estimation.

To illustrate these properties, we select a series of synthetic case studies of increasing complexity, from time-shifted Ricker traces to a realistic Marmousi II case study (not in inverse crime settings). We restrict our attention to five misfit functions, which have been promoted recently and have shown promising results: adaptive waveform inversion (AWI), instantaneous envelope

(IE), normalized integration method (NIM), KROT, and GSOT. We consider extended space strategies out of the scope of this study to keep it reasonably simple, and also because, as stated before, we consider that alternative misfit strategies have shown more promising results than extended space strategies so far in terms of practical applications. The tests that we develop here could, however, be used to benchmark extended space strategies also.

2.2 General FWI framework and misfit function formulation

The comparison between misfit functions is made simple by the FWI formalism (reviewed in the following section), more precisely by the adjoint state strategy used to compute the gradient at each iteration of the minimization loop. However, let us recall the main result: a modification of the misfit function results only in modifying the adjoint source. Therefore, implementing different misfit functions in the same FWI code can be done directly by isolating misfit function evaluation and adjoint source computation in different subroutines.

2.2.1 General framework

The FWI problem can be written as

$$\min_m f[m] = F(d_{cal}[m], d_{obs}), \quad (2.1)$$

where the subsurface parameters are denoted by m , d_{obs} is the observed data, $d_{cal}[m]$ is the synthetic data, and F is a generic function measuring the misfit between d_{obs} and d_{cal} . Under general notation, $d_{cal}[m]$ is obtained through the extraction of the values of wavefield at the receivers location such that

$$d_{cal}[m] = Ru[m], \quad (2.2)$$

where R is an extraction operator and $u[m]$ is the solution of the wave propagation problem

$$A[m]u = b, \quad (2.3)$$

with $A[m]$ a generic wave propagation operator (from acoustic to visco-elastic).

The solution of the minimization problem 2.1 is computed through local optimization following the iteration

$$m_{k+1} = m_k + \alpha_k \Delta m_k \quad (2.4)$$

starting from an initial guess m_0 . In eq. 2.4, α_k is the steplength, which should satisfy the Wolfe criterion (Nocedal and Wright, 2006), and Δm_k is the descent direction, given by

$$\Delta m_k = -P[m_k] \nabla f[m_k], \quad (2.5)$$

where $\nabla f(m_k)$ is the gradient of the misfit function $f[m]$ and $P[m_k]$ a preconditioner approximating the inverse Hessian operator

$$P[m_k] \simeq H[m_k]^{-1}, \quad H[m_k] = \nabla^2 f[m_k]. \quad (2.6)$$

Following the adjoint state strategy (Plessix, 2006), the gradient is given by

$$\nabla f[m] = \left(\frac{\partial A}{\partial m} u, \lambda \right), \quad (2.7)$$

where $(.,.)$ is the Euclidean scalar product in the wavefield space, and λ is the adjoint field, solution of the adjoint equation

$$A(m)^T \lambda = s, \quad (2.8)$$

where s is the generic adjoint source, given by

$$s = -R^T \left(\frac{\partial F}{\partial d_{cal}} \right). \quad (2.9)$$

Note that in the case of the L^2 norm, we recover immediately that

$$s = -R^T (Ru[m] - d_{obs}), \quad (2.10)$$

i.e. the adjoint source is equal to the residual (difference between observed and calculated data).

Next, we review the formulas for the five misfit functions selected in this study, as well as their corresponding adjoint sources. For convenience, we will introduce the distance measurement function associated with each strategy for a single source/receiver couple, except for the KROT strategy. The calculated and observed data will be denoted by $d_{cal}(t)$ and $d_{obs}(t)$ unless stated otherwise. Except for KROT, the final misfit function is built as a sum over each source/receiver couple of this distance measurement function, and by linearity, the resulting adjoint source is also obtained by summation.

2.2.2 Adaptive waveform inversion

We give here the AWI formalism. We have

$$F_{AWI}(d_{cal}, d_{obs}) = \frac{\int_0^T |\mathcal{P}(\tau)w(\tau)|^2 d\tau}{\int_0^T |w(\tau)|^2 d\tau}, \quad (2.11)$$

where $w(t)$ is the Wiener filter which either transforms the calculated $d_{cal}(t)$ into the observed data $d_{obs}(t)$ (forward AWI) or the opposite way around (reverse AWI). Both implementations are discussed in Warner and Guasch (2016). Also, the computation of $w(t)$ can be implemented either in the time-domain or the frequency-domain. In both cases, a water level ε is required to stabilize the deconvolution operation.

The role of the function $\mathcal{P}(\tau)$ is to penalize energy at non-zero time lag. There are several possibilities to define this penalty function. Here we focus only on a Gaussian formulation defined as

$$\mathcal{P}(\tau) = e^{-\tau^2/\sigma^2}, \quad (2.12)$$

where σ is a tuning parameter controlling the width of the Gaussian function away from 0 time-lag. This σ tuning parameter is defined in seconds and corresponds to the maximum expected time-shift between the observed and calculated data.

In the case of a frequency-domain reverse AWI implementation, the adjoint source for a single-trace reads

$$\frac{\partial F_{AWI}}{\partial d_{cal}} = \frac{\int d\tau (\mathcal{P}(\tau) - 2F(d_{cal}, d_{obs})) w(\tau)p(t + \tau)}{\int d\tau w^2(\tau)}, \quad (2.13)$$

where

$$p(t) \approx \int d\omega \frac{\hat{d}_{obs}(\omega)e^{i\omega t}}{\hat{d}_{obs}^*(\omega)\hat{d}_{obs}(\omega) + \varepsilon}, \quad (2.14)$$

with ε defined as

$$\varepsilon = (\max_{\omega} |\hat{d}_{obs}(\omega)|)\zeta. \quad (2.15)$$

In eq. 2.15, ζ is a user-defined damping ratio, ranging from 10^{-2} to 10^{-5} in our experiment. A large ζ will help when trying to tackle large time-shift, with a “smoothing/regularizing” effect. Large ζ is also required if there is noise on the data. A smaller ζ will help preserve small features present in the signal. In terms of computational cost, the overhead associated with the computation of the Wiener filter is negligible, and the AWI strategy can be easily implemented.

2.2.3 Instantaneous envelope

The separation of the phase and envelope information of the signal relies on the use of the analytical function defined as follows. For a given time signal $d(t)$, the analytical signal $\tilde{d}(t)$ is defined as

$$\tilde{d}(t) = d(t) + i\mathcal{H}[d(t)], \quad (2.16)$$

where \mathcal{H} is the Hilbert function which can be defined in the time domain as

$$\mathcal{H}[d(t)] = \frac{1}{\pi}P \int_{-\infty}^{+\infty} \frac{d(\tau)}{t - \tau} d\tau, \quad (2.17)$$

where P stands for the Cauchy principal value. Practically, we do not use the time formulation of the Hilbert function, but rather a frequency domain formulation that gives us the analytical signal in a three-step approach (Marple, 1999):

- Compute the Fourier transform of $d(t)$ using an FFT
- Change the negative frequency to zeros
- Compute the inverse Fourier transform

This directly gives us access to the analytical signal and, by extension, to the Hilbert transform by taking its imaginary part

$$\mathcal{H}[d(t)] = \mathcal{I}[\tilde{d}(t)]. \quad (2.18)$$

The analytical signal allows to separate the signal as the combinaison of the instantaneous phase $\phi(t)$ and the instantaneous envelope $E(t)$:

$$\tilde{d}(t) = E(t)e^{i\phi(t)}. \quad (2.19)$$

Thus, the instantaneous envelope $E(t)$ can be simply defined as:

$$E(t) = \sqrt{\mathcal{R}[\tilde{d}(t)]^2 + \mathcal{I}[\tilde{d}(t)]^2}. \quad (2.20)$$

We can define a new distance-measurement function using instantaneous envelope as

$$F_{IE}(d_{cal}, d_{obs}) = \frac{1}{2} \int_0^T |E_{cal}(t) - E_{obs}(t)|^2 dt, \quad (2.21)$$

where E_{cal} and E_{obs} are instantaneous envelopes of the calculated and observed data respectively. Following Bozdağ et al. (2011), the adjoint source is defined as:

$$\begin{aligned} \frac{\partial F_{IE}}{\partial d_{cal}} = & \frac{(E_{cal}(t) - E_{obs})d_{cal}(t)}{E_{cal}(t) + \varepsilon} \\ & - \mathcal{H} \left(\frac{(E_{cal}(t) - E_{obs})\mathcal{H}(d_{cal}(t))}{E_{cal}(t) + \varepsilon} \right), \end{aligned} \quad (2.22)$$

with ε a water level defined as

$$\varepsilon = (\max_t E_{obs}(t))\zeta. \quad (2.23)$$

Contrary to AWI, in the following experiments, ζ is fixed and taken at $\zeta = 10^{-5}$ for IE. We have verified that the results with IE are not sensitive to this choice.

The instantaneous envelope misfit formulation is straightforward to implement thanks to the algorithm from Marple (1999). No tuning parameter is required, and the computation cost overhead is negligible.

2.2.4 Normalized Integration Method

Donno et al. (2013) consider the least-squares difference between the cumulative distributions Q_{obs} and Q_{cal} . For a given time signal $d(t)$, its normalized cumulative distribution $Q(t)$ is defined by

$$Q(t) = \frac{\int_0^t d(\tau)^2 d\tau}{\int_0^T d(\tau)^2 d\tau}. \quad (2.24)$$

The NIM misfit function thus relies on the distance measurement

$$F_{NIM}(d_{cal}, d_{obs}) = \frac{1}{2} \int_0^T |Q_{cal}(\tau) - Q_{obs}(\tau)|^2 d\tau, \quad (2.25)$$

where $Q_{cal}(t)$ and $Q_{obs}(t)$ are the cumulative distributions associated with $d_{cal}(t)$ and $d_{obs}(t)$ respectively.

The corresponding adjoint source is

$$\begin{aligned} \frac{\partial F_{NIM}}{\partial d_{cal}}(t) = & \frac{2d_{cal}(t)}{\int_0^T Q_{cal}(t)} \left(\int_t^T (Q_{cal}(\tau) - Q_{obs}(\tau)) d\tau \right. \\ & \left. - \int_0^T Q_{cal}(\tau)(Q_{cal}(\tau) - Q_{obs}(\tau)) d\tau \right). \end{aligned} \quad (2.26)$$

The NIM implementation is straightforward and does not require any tuning parameters.

2.2.5 Kantorovich-Rubinstein optimal transport distance

In the frame of the KROT approach, we consider the data as a function of both time and receiver position, such that we denote the calculated and observed data as $d_{cal}(x_r, t)$ and $d_{obs}(x_r, t)$ respectively.

The KROT is based on a particular instance of optimal transport distance, namely the 1-Wasserstein distance. It can be applied to non-positive data, provided mass conservation is satisfied *i.e.*

$$\int_{x_r} \int_0^T d_{cal}(x_r, t) dx_r dt = \int_{x_r} \int_0^T d_{obs}(x_r, t) dx_r dt. \quad (2.27)$$

For a given shot in seismic data, this corresponds to the summation over each trace of the mean value in time of the trace. We consider this mean value is equal to 0 (this is the zero-frequency noise, which is usually removed from the data prior to inversion). Therefore the mass conservation assumption is satisfied for seismic data.

On this basis, the KROT distance can be written as

$$F_{KROT}(d_{cal}, d_{obs}) = \max_{\varphi \in \text{Lip}_1} \int_{x_r} \int_0^T \varphi(x_r, t) (d_{cal}(x_r, t) - d_{obs}(x_r, t)) dx_r dt, \quad (2.28)$$

where Lip_1 is the set of 1-Lipschitz functions for the ℓ_1 distance

$$\text{Lip}_1 = \{ \varphi(x_r, t), \quad |\varphi(x_r, t) - \varphi(x'_r, t')| < |x_r - x'_r| + |t - t'| \}. \quad (2.29)$$

The adjoint source is then given by

$$\frac{\partial F_{KROT}}{\partial d_{cal}} = \bar{\varphi}(x_r, t), \quad (2.30)$$

where

$$\bar{\varphi}(x_r, t) = \arg \max_{\varphi \in \text{Lip}_1} \int_{x_r} \int_0^T \varphi(x_r, t) (d_{cal}(x_r, t) - d_{obs}(x_r, t)) dx_r dt. \quad (2.31)$$

Compared with previous misfit functions, the final misfit is obtained here by summation over shot gather, and not a summation over source/receiver couples (not a trace-by-trace approach).

From the above equations, we see that the computation of the KROT misfit function and its corresponding adjoint source requires solving a constrained maximization problem per shot gather. Details on how to solve this problem are given in Métivier et al. (2016c). The proximal splitting algorithm ADMM is used (Combettes and Pesquet, 2011) and the resulting algorithm has complexity in $O(N \log N)$, where $N = N_r \times N_t$ with N_r the number of receivers and N_t the number of time samples. Compared with the previous misfit functions, the computational cost overhead is non-negligible. Tuning parameters will be associated with a prior scaling of the data to make its maximum amplitude close to 1, and the number of iterations required to solve the constrained maximization problem.

2.2.6 Graph space optimal transport distance

Let $(t_i, d(t_i))$, $i = 1, \dots, N$ be the discrete graph of the time function $d(t)$. This discrete graph is a point cloud containing N points. The GSOT distance measurement is formulated as

$$F_{GSOT}(d_{cal}, d_{obs}) = \min_{\sigma \in S(N)} \sum_{i=1}^N c_{i\sigma(i)}, \quad (2.32)$$

where c_{ij} is the L^2 distance between the points of the discrete graph of d_{cal} and d_{obs} , namely

$$c_{ij} = |t_i - t_j|^2 + \eta^2 |d_{cal}(t_i) - d_{obs}(t_j)|^2, \quad (2.33)$$

and $S(N)$ is the ensemble of permutations of $(1 \dots N)$. The function F_{GSOT} corresponds to the 2-Wasserstein distance between the discrete graph of the calculated trace $d_{cal}(t)$ and the observed trace $d_{obs}(t)$.

The scaling parameter η in eq. 2.33 controls the convexity of the misfit function f_{GSOT} with respect to time-shifts. In practice, we define it as

$$\eta = \frac{\tau}{A}, \quad (2.34)$$

where τ is a user-defined parameter corresponding to the maximum expected time-shift between observed and calculated data in the initial model, and A is the maximum amplitude discrepancy between observed and calculated data.

The adjoint source of the misfit function $f_{GSOT}[m]$ is computed from $\frac{\partial f_{GSOT}}{\partial_{cal}}$ using the adjoint-state strategy. We prove in Métivier et al. (2019) the following equality: denoting σ^* the minimizer in eq. 2.32, we have

$$\frac{\partial f_{GSOT}}{\partial_{cal}} = 2 \left(d_{cal} - d_{obs}^{\sigma^*} \right), \quad (2.35)$$

where

$$d_{obs}^{\sigma^*}(t_i) = d_{obs}(t_{\sigma^*(i)}). \quad (2.36)$$

In this sense, the GSOT approach can be viewed as a generalization of the L^2 distance: the adjoint source is equal to the difference between calculated and observed data at time samples connected by the optimal assignment σ^* . Similarly, as the KROT approach, the solution of the problem 2.32 provides the information to compute both the misfit function and the adjoint source.

The numerical algorithm used to solve the linear assignment problem 2.32 is the auction algorithm (Bertsekas and Castanon, 1989). For problems involving less than 1000 points, the auction algorithm is very efficient. In seismic exploration, Nyquist sampling yields traces containing a number of points within this order of magnitude. Consequently, we have designed an efficient numerical strategy, yielding lower computational overhead than the KROT approach. On 3D field data application, we observe 15 to 20% computation time increase for gradient computation on the lowest frequency bands compared with classical L^2 . This computational cost overhead decreases when the frequency band increases as the total complexity of the GSOT problem is $O(\omega^3)$, while the complexity of the wave propagation solver is in $O(\omega^4)$. For more details, the reader can refer to Métivier et al. (2019).

Compared with previous approaches, the computational cost overhead is comparable with AWI, IE, and NIM while being lower than KROT. In terms of implementation, as for KROT, the solution of the assignment problem requires specific solvers, which makes the GSOT implementation less trivial than for AWI, IE, or NIM. In terms of tuning parameters, the more important parameter is the parameter τ , which controls the convexity of GSOT misfit function with respect to time-shifts.

2.3 A simple convexity analysis based on time-shifted Ricker wavelets

We start by investigating the convexity of the proposed misfit functions with respect to time-shifts. We fix a reference signal composed of one Ricker wavelet in the center, seen as the observed data. The calculated data is the same Ricker wavelet, shifted in time with a time-shift going from -1.5 s to 1.5 s. We compute the distance between the reference signal and the calculated signal using the five selected misfit functions, depending on the input time-shift. Results are presented in Figure 2.1.

The results obtained here with alternative misfit functions might not reflect the performance of the algorithms with total accuracy, both in terms of computational efficiency and inversion results. Algorithms might not have been implemented in the most optimal way or in the way the original authors intended. Subtle choices of tuning parameters might improve the inversion results in some cases. However, the primary purpose of this comparison is to seek to understand how the data is interpreted within each of these strategies and how this affects the inversion results in each case. We intend to provide the reader with sufficient material to infer the main properties and philosophy behind the compared methods.

Let us first analyze the results obtained with L^2 waveform misfit, the reference for FWI. As expected, L^2 misfit displays a narrow basin of attraction, with local minima and a flat part for time-shift superior to 0.4 s. The local minimum appears when the time-shift is larger than 0.12 s, which corresponds to half the Ricker wavelet period. This validates that the L^2 misfit function presents low robustness for shifted-patterns, leading to cycle-skipping when signals are shifted by more than half a period. In such cases, L^2 misfit function does not guarantee convergence toward the global minimum.

We can now compare the selected alternative misfit functions to the L^2 misfit. From the obtained results, we can define two groups. The first one contains GSOT, AWI, and NIM, characterized by a large basin of attraction. The second group contains IE and KROT, characterized by a “slightly” larger basin of attraction than L^2 , but not as wide as the first group members.

Understanding why the first group members exhibit the convexity property is essential. Starting with GSOT, if the input parameters τ is correctly set to the maximum expected time-shift of 1.5 s, the convexity to shifted-patterns is expected as there is a direct link between the τ parameters and the width of the basin of attraction as shown in Métivier et al. (2019).

The same convexity property is observed with AWI. With σ set to 1.5 s, the results are satisfying with a large basin of attraction. Similarly as the τ parameter from GSOT, σ directly controls the convexity to shifted-patterns. Note that we use $\zeta = 10^{-5}$ in this analysis as we predict signal with machine precision.

Finally, to understand the robustness of the NIM approach, we display in Figure 2.2 the quantities Q_{obs} and Q_{cal} (for three time-shifts, -1.5 s, -0.1 s and in-phase). This makes visible the drastic modification of the signal shape induced by NIM. The NIM cost function boils down to be the area under the curve delimited by $Q_{obs} - Q_{cal}$. We see clearly that this area increases with time-shifts, illustrating the convexity to shifted patterns observed with NIM.

Moving to the second group, to understand why the IE misfit only slightly increases the width of the valley of attraction compared with L^2 , we display in Figure 2.3 the quantities E_{obs} and E_{cal} . Here we can observe the increase of temporal support of the signal induced by the envelope. This “broader” temporal support of the instantaneous envelope directly translates into the increase of the width of the valley of attraction as IE relies on a L^2 norm between E_{obs} and E_{cal} .

Finally, we present in Figure 2.4 the function $\bar{\varphi}(t)$ solution of the maximization problem defined in eq. 2.31, which defines the KROT distance, together with the residuals $d_{obs}(t) - d_{cal}(t)$. We can observe that when Ricker wavelets start to overlap at -0.3 s, we obtain a convexity that classical L^2 cannot achieve. This can be understood by looking at the function $\bar{\varphi}(t) [d_{obs}(t) - d_{cal}(t)]$. The area below the curve defined by this function corresponds to the KROT misfit function. This area remains constant as long as the two signals do not overlap and monotonically decrease as soon as the two signals overlap, reaching 0 at 0 time-shift.

On a second test, presented in Figure 2.5, we introduce a second Ricker wavelet that remains in phase. This test aims at validating the robustness to cycle-skipping when multiple arrivals are considered. From the results obtained, we observe that all misfit functions behave similarly as on the previous test except for AWI. In this case, the shape of the misfit function seems affected by oscillations near 0 time-shift, reducing the effective convexity to the one of classical L^2 formulation. This seems to be related to one of the potential issues of deconvolution based misfit function: the sensitivity to cross-talks between multiple events. To analyze this sensitivity of AWI to multi-arrivals, we display the Wiener filters together with the penalty function and the combination of both (Figure 2.6). In test B (where one wavelet is always in-phase), the Wiener filter presents a strong peak at 0 time-lag due to the in-phase arrivals. Because of finite frequency effect, it is not a Dirac delta function but a bandpass Dirac delta function. The oscillations of the bandpass delta function combine in a destructive/constructive manner when the two time-shift peaks (one for each Ricker wavelet) get closer to each other. These interferences are at the origin of the local minima observed.

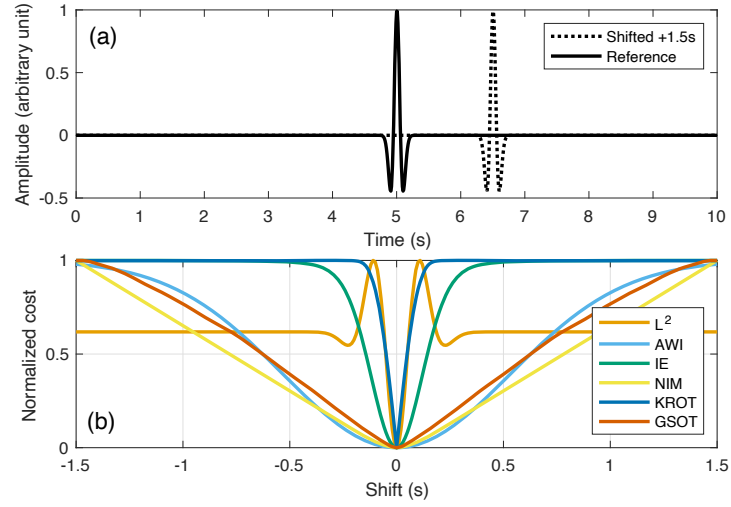


Figure 2.1: Comparison of several misfit functions in a simple 1D case for one shifted arrival. The arrival is set to be a Ricker wavelet with a central frequency of 4 Hz. (a) represents the signal used for the test (with only one arrival at the center). The fixed reference signal is displayed in continuous black. The shifted signal is displayed in dotted black (here for +1.5 s). (b) represents the normalized misfit function values with respect to the time-shift (from -1.5 s to 1.5 s).

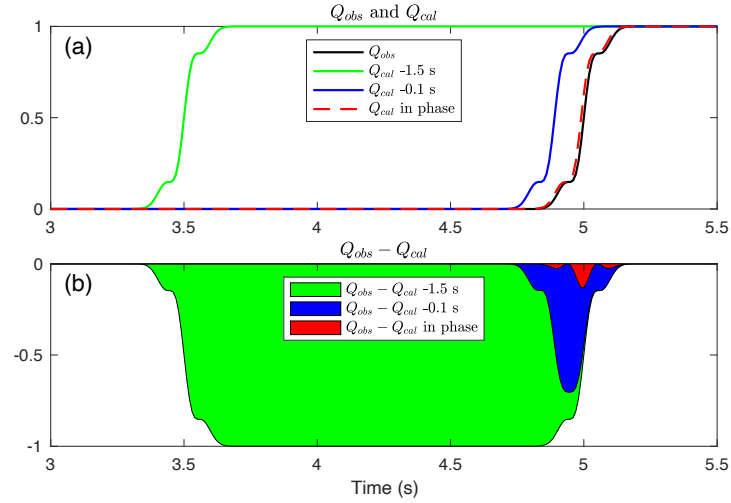


Figure 2.2: (a) quantities Q_{obs} and Q_{cal} for three time-shifts (-1.5 s in green, -0.1 s in blue and “in phase” in dashed red). (b) the area under the curve for $Q_{obs} - Q_{cal}$ for the three time-shifts.

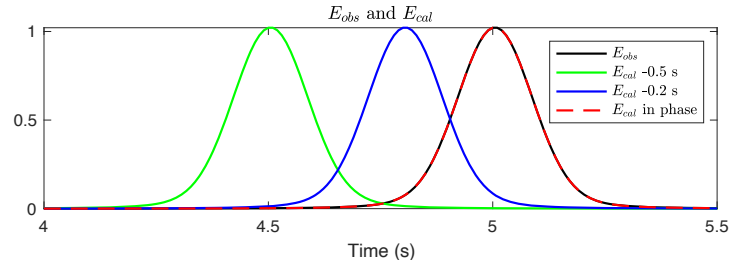


Figure 2.3: E_{obs} and E_{cal} for three time-shifts (-0.5 s, -0.2 s and in phase).

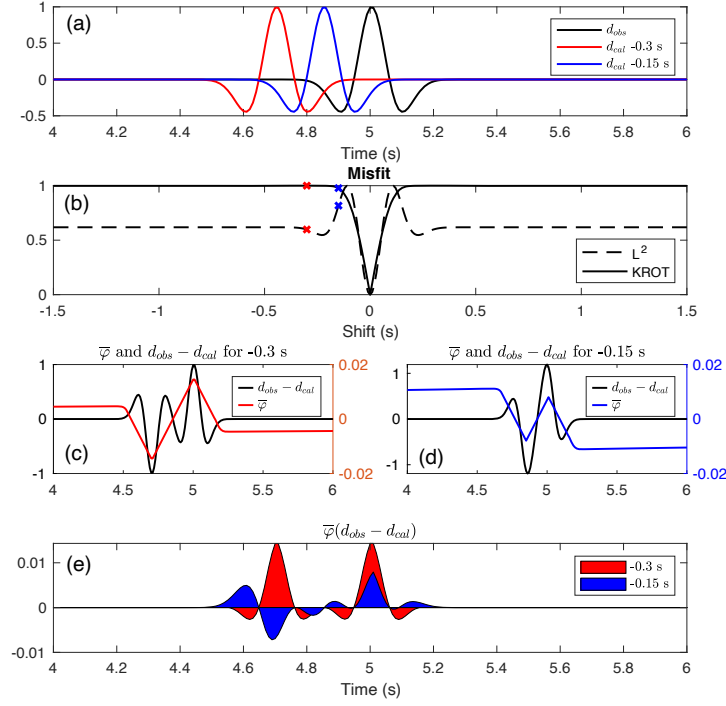


Figure 2.4: Detail for $\bar{\varphi}(t)$ from KROT. (a) the setup with d_{obs} (in black) and d_{cal} for two time-shifts (-0.3 s in red and -0.15 s in blue). (b) shape of L^2 and KROT misfit function with respect to time-shifts, red and blue cross represent the positions of the two time-shifts selected. (c) and (d) respectively display $\bar{\varphi}(t)$ and $d_{obs} - d_{cal}$ for the two time-shifts of -0.3 s and -0.15 s. (e) the area under the curve for $\bar{\varphi}(t)(d_{obs} - d_{cal})$ quantity for the two time-shifts. This last quantity is used to get the misfit function value after time integration.

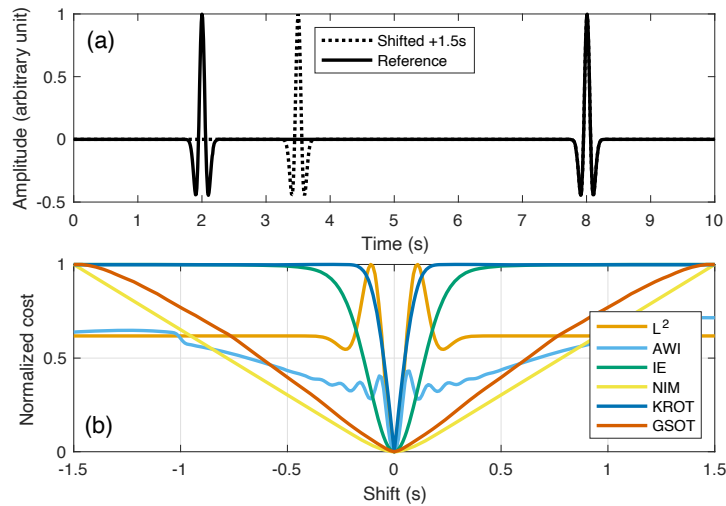


Figure 2.5: Same as Figure 2.1 but with two Ricker wavelets with one shifted (left) and one in phase (right).

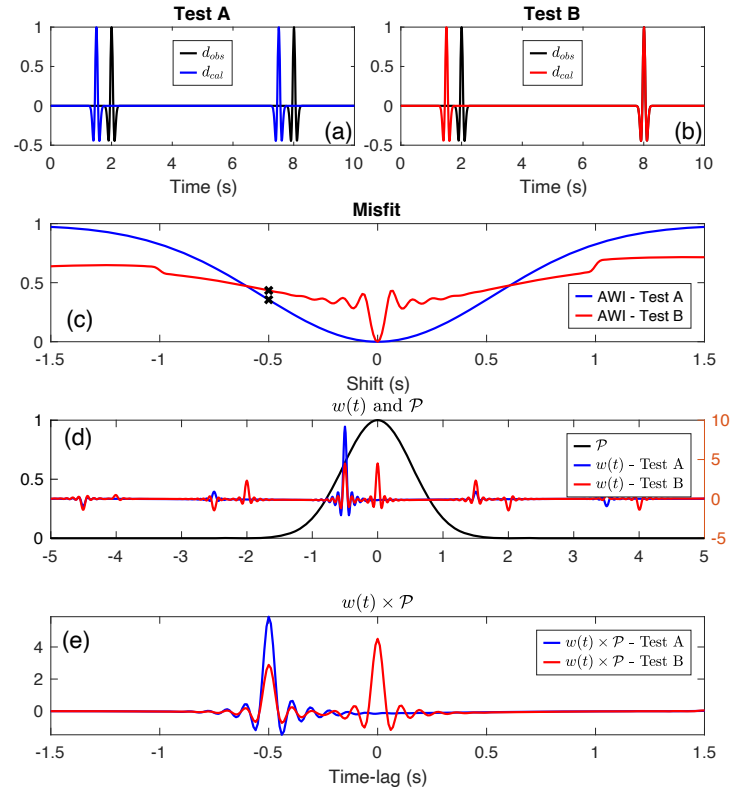


Figure 2.6: AWI analysis with two setups: test A and test B. (a) test A both wavelets shift, (b) test B only the left wavelet is shifted, the right one being always in phase (similarly to Figure 2.5). (c) shape of AWI misfit function with respect to time-shift in both cases. The Wiener filters presented under are shown for a time-shift of -0.5 s (black cross on the misfit). (d) Wiener filters ($w(t)$) and the Gaussian penalty function $\mathcal{P}(t)$. (e) the Wiener filters multiplied by the penalty function.

2.4 FWI tests on two canonical examples

This section attempts to assess the pros and cons of the selected alternative misfit functions on two schematic FWI tests, focusing on a different aspect of the information contained in a dataset. The first test focuses only on transmission with a cross-hole acquisition. The second test focuses mainly on reflection information. These two tests can be seen as a way of assessing if the proposed misfit function can improve the FWI robustness (cycle-skipping in transmission in the first test) while preserving the ability to correctly interpret reflection information (reflector positioning and imaging in the second test)

Both tests are performed in 2D using our 2D/3D time-domain acoustic modeling and inversion code in inverse crime settings (observed and calculated data are computed on the same grid, without noise introduced in the data). Besides, we use a constant density model and invert only for the P-wave velocity model. In both cases, the *l*-BFGS algorithm is used to minimize the misfit function, with FWI stopping criterion being a line search failure. The source wavelet is a Ricker wavelet with a central frequency $f_{ref} = 3$ Hz. The gradient is smoothed using a Gaussian filter with horizontal and vertical correlation lengths equal to 0.3 times the local wavelength

$$\lambda_{loc}(x, z) = 0.3 \frac{v_P(x, z)}{f_{ref}}. \quad (2.37)$$

2.4.1 FWI Test 1: transmission configuration

2.4.1.1 Case study presentation

This first case study focuses on transmitted energy. The exact model is defined as a square of 1000 m sides with homogeneous $V_P = 1300 \text{ m.s}^{-1}$ containing a spherical inclusion of 100 m radius in the center with $V_P = 1700 \text{ m.s}^{-1}$ (Figure 2.7). The acquisition mimics a crosshole setting, with 96 sources on the left side of the model and 256 receivers on the right side. The spacing is 10 m between sources and 3.8 m between receivers. The boundaries are all set to absorbing layers (Bérenger, 1994) to avoid reflections and only focus on transmitted events. The relatively strong contrast between the background and the anomaly generates an identifiable diffraction pattern in the data. In this experiment, no preconditioning is applied to the gradient. The lower and upper V_P bound constraints are respectively set to 1000 and 2500 m.s^{-1} .

We introduce three starting homogeneous models (Figure 2.7). The first is at the true model background velocity (1300 m.s^{-1}). The second is at $V_P = 1700 \text{ m.s}^{-1}$, setting a challenging FWI problem as the starting model is as fast as the inclusion. The third case is even more challenging, with a starting homogeneous V_P model at 1900 m.s^{-1} .

FWI results are presented in Figure 2.8 with reconstructed V_P at the final iteration. Figure 2.9 presents traces for a single source-receiver couple representing the shortest path through the spherical inclusion (straight horizontal path at 500 m depth). Traces are extracted from data generated in the true model, initial model, and final reconstructed model for all misfit functions.

2.4.1.2 Results from initial model 1

We start the analysis with the “reference” initial model. As shown in Figure 2.9, this model does not generate cycle-skipping (arrivals in the true model are less than half a period away from the arrivals in the initial model). The objective is to retrieve the high-velocity spherical inclusion in the center of the model. As expected, the L^2 misfit function produces a correct result: the inclusion is retrieved correctly, and the final data are in phase with the true data. The vertical resolution is higher than the horizontal resolution as expected from the cross-hole configuration. This has a lateral smoothing effect on the reconstructed anomaly, which explains why its peak amplitude (around 1500 m.s^{-1}) is lower than the amplitude of the true anomaly. The five selected misfit functions produce equivalently good results in this configuration. In all cases, the spherical anomaly is reconstructed with a similar resolution, and the data fit is equivalent. In terms of parameter settings, we choose here $\tau = 0.2 \text{ s}$ for GSOT and $\sigma = 0.2 \text{ s}$ for AWI, a choice motivated by the absence of cycle-skipping. For AWI, we use $\zeta = 10^{-2}$ in this transmission test (for all three models) to maximize the kinematic effects of AWI that work better when ζ is relatively high, which acts as a regularization effect.

2.4.1.3 Results from initial model 2

As can be observed in Figure 2.9, the second initial model generates clear cycle-skipping in the data. In this case, we expect the L^2 misfit function to fail in reconstructing the anomaly. Indeed, the L^2 fails to converge and reaches the boundary set for the inversion. The final synthetic trace does not match the observed trace. It is interesting to observe that four of the five selected misfit functions succeed in reconstructing the background and the anomaly and produce final synthetic traces in phase with the observed trace in this already quite challenging test. The only alternative misfit function that fails is KROT, which could be expected from the previous section (weak increase of robustness to cycle-skipping). AWI, IE, NIM and GSOT show that the increase in convexity procured by these formulations is enough here to make convergence achievable. The data-fit obtained with these methods is good in this case. In terms of tuning parameters, τ and σ are increased to 0.35 s for GSOT and AWI, according to the time-shift between the reference and the initial traces in the initial model.

2.4.1.4 Results from initial model 3

Finally, the initial model 3 generates an even more substantial cycle-skipping effect than model 2 (Figure 2.9). L^2 and KROT still fail to converge to the correct model, as it was already the case starting from model 2.

IE starts to exhibit diagonal cycle-skipping artifacts associated with the longest source/receiver paths in this more challenging setting. This is expected from the time-shift convexity analysis performed before: IE robustness to cycle-skipping is limited. AWI also starts to exhibit artifacts close from the acquisition, while central anomaly is correctly reconstructed (with $\sigma = 0.6 \text{ s}$). NIM and GSOT (with τ increased to 0.6 s) achieve a relatively satisfactory reconstruction of the background and anomaly, similar to the results obtained from the previous background models.

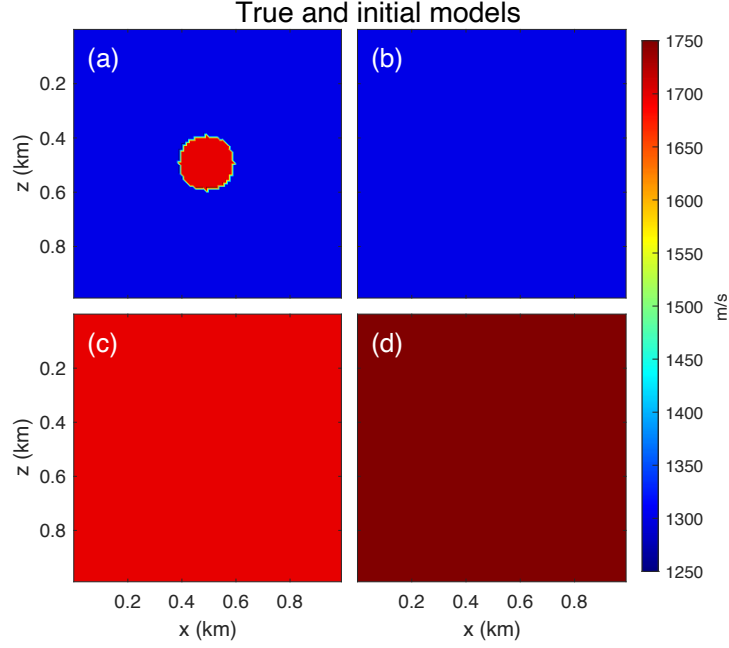


Figure 2.7: FWI Test 1: (a) true model, (b) initial model 1 with $V_P = 1300 \text{ m.s}^{-1}$, (c) initial model 2 with $V_P = 1700 \text{ m.s}^{-1}$ and (d) initial model 3 with $V_P = 1900 \text{ m.s}^{-1}$.

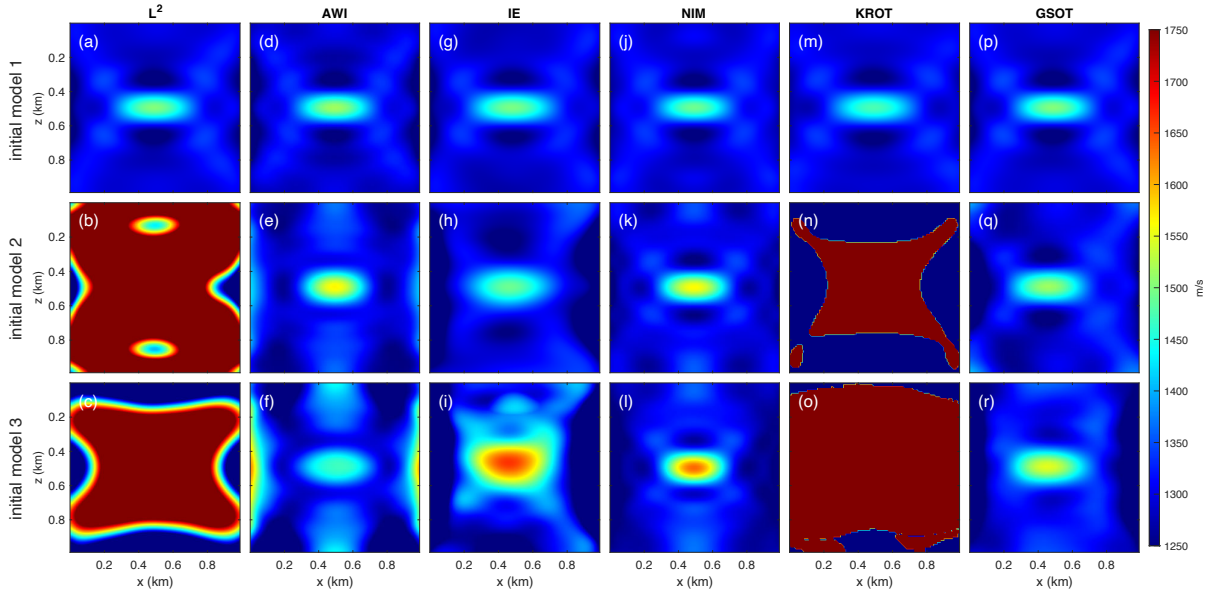


Figure 2.8: FWI Test 1: V_P results from FWI. First line corresponds to initial model 1, second line to initial model 2 and third line to initial model 3. Each column corresponds to reconstructed V_P model from FWI using respectively L^2 (a-c), AWI (d-f), IE (g-i), NIM (j-l), KROT (m-o) and GSOT (p-r) misfit functions.

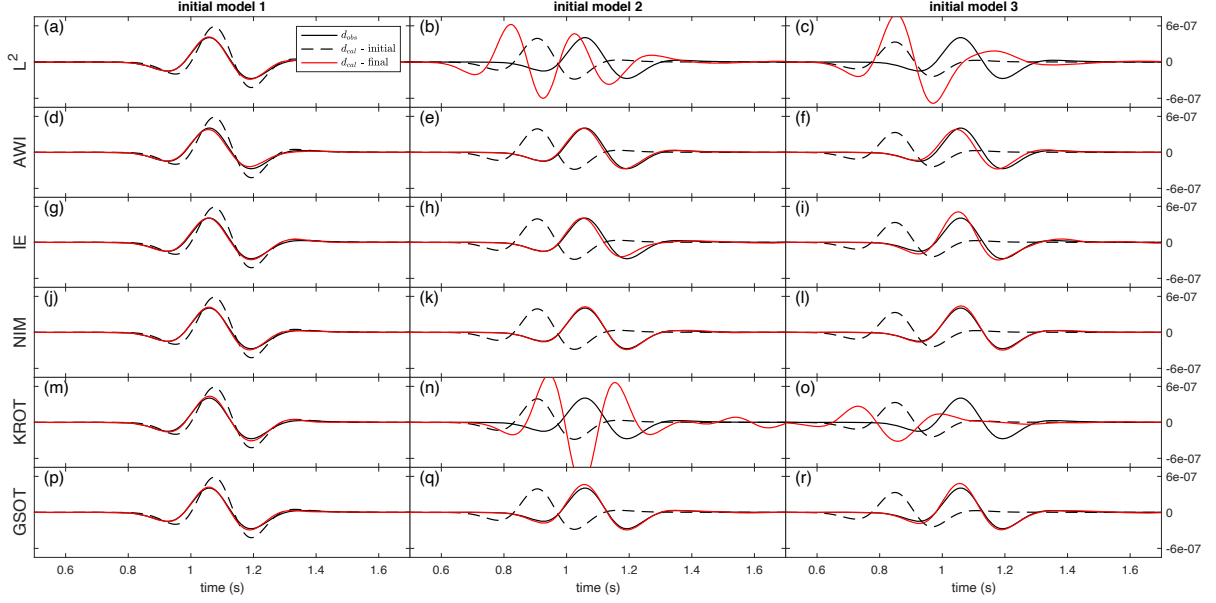


Figure 2.9: FWI Test 1: Extracted traces along the shortest path (horizontal straight line at 500 m depth passing through the spherical inclusion). Observed data are in solid black, synthetic data in the initial model in dashed black and final reconstructed synthetic data in solid red. First column corresponds to initial model 1, second column to initial model 2 and third column to initial model 3. Each line corresponds to reconstructed V_P model from FWI using respectively L^2 (a-c), AWI (d-f), IE (g-i), NIM (j-l), KROT (m-o) and GSOT (p-r) misfit functions.

2.4.2 FWI Test 2: reflection configuration

2.4.2.1 Case study presentation

This second case study focuses on reflected energy. We consider two different true models, composed of a homogeneous background at 1500 m.s^{-1} and a velocity layer 100 m thick at 300 m depth (Figure 2.10). In the first case, the velocity of the layer is set to 1600 m.s^{-1} , while in the second case, the velocity of the layer is set to 1400 m.s^{-1} . The starting model is homogeneous at the correct background velocity of 1500 m.s^{-1} (Figure 2.10). The surface acquisition comprises 96 sources and 512 receivers located close to the surface at 42 m depth. The spacing is 20 m between sources and 3.8 m between receivers. We implement PML absorbing conditions on the bottom and lateral sides of the medium to mimic a medium of infinite extension in these directions and a free surface condition on the top of the model. A simple linear in-depth preconditioner is also applied to compensate for geometrical spreading effects and accelerate the convergence. The lower and upper V_P boundaries for the inversion are respectively set to 1200 m.s^{-1} and 1800 m.s^{-1} .

This test analyzes how the reflected data is interpreted by FWI depending on the choice of misfit function. The difference between the two exact models is only the sign of the velocity change at the layer level: in one case, velocity increases; in the second case, it decreases. This induces a change of polarity of the reflected wave, as clearly visible in Figure 2.11. We want to identify how the different misfit functions are sensitive to this change of polarity.

2.4.2.2 Results analysis

The reconstructed models are presented in Figure 2.12. Traces from the observed and synthetic data in the initial and final models for zero offset couple (source and receiver at the same position) located in the middle of the acquisition are presented in Figure 2.13. For visualization purposes, we cropped over the reflection after the first arrival.

The L^2 results are coherent with the expectation, with a correct reconstruction of the layer in both cases. The L^2 norm is sensitive to amplitude variation and polarity and is expected to interpret reflected events correctly. As the background velocity is known, there is no cycle-skipping in the initial model for the two target models. The data fit in both cases is perfect.

Together with L^2 misfit function, results obtained with IE, KROT, and GSOT are equivalently correct. This is expected from KROT and GSOT, which should behave similarly as L^2 when cycle-skipping does not occur. GSOT relies on $\tau = 0.2$ s in this experiment. This is somehow more surprising from IE, as one could think that the polarity of reflected events might be lost in the envelope extraction process. However, this is not true in this case but might be due to the inverse crime settings we are using. There is indeed a subtle change in the envelope of the observed data between model 1 (positive layer) and model 2 (negative layer) (Figure 2.14), which is enough to guide the inversion in the right direction. However, this is probably possible only because the first arrival is correctly predicted. Small inaccuracies in predicting the first arrival might be enough to impede a correct reconstruction using IE.

The results obtained with NIM are less satisfactory. In particular, the reconstruction of the negative layer is altered by strong positive artifacts beneath the layer. In the opposite case, negative artifacts also pollute the reconstruction of the positive layer, although the strength of these artifacts seems weaker. Analyzing the data fit shows that NIM has difficulties reproducing the reflection pattern in both cases (Figure 2.13). Spurious oscillations appear, which can be associated with the positive artifacts observed on the model reconstruction. This lack of sensitivity to the polarity is somehow expected. From NIM formulation, Q_{obs} should be more or less the same independently of model 1 or model 2 being used. This is illustrated in Figure 2.14, where $Q_{obs}(t)$ is presented for both models (positive and negative layers). We can observe that the difference between the two true models leads to a very marginal modification of Q_{obs} compared to Q_{cal} . This is likely the explanation of the difficulties faced by NIM in interpreting the reflected waves correctly.

Finally, the results obtained with AWI are incorrect for both the negative and positive V_P anomaly. From observing the data, we can see that the direct waves exhibit a clear dominance in amplitude over the reflected events. Therefore, we expect the Wiener filter to be dominated by the direct waves and only show a small imprint of the reflected waves that are of small amplitude ($\approx 1\%$ of peak amplitude). This is illustrated by the Wiener filter shown in Figure 2.15 for both positive and negative layer models. They indeed present a main event around 0 lag, corresponding to the in-phase direct wave. Around 0.3 s, the imprint of reflected events is very weak but still visible in the Wiener filters. This motivates us to use a large $\sigma = 1$ s to maximize the information coming from the small reflected waves, together with a small $\zeta = 10^{-5}$. However, these settings do not make it possible to obtain satisfactory results with AWI. To have a deeper understanding of why AWI fails to reconstruct a proper V_P model in this case, we perform a sensitivity analysis of the misfit function with respect to the value of V_P in the layer. We compute the AWI misfit value (and L^2 misfit value for reference) between d_{obs} and $d_{cal}(V_P)$. Here d_{obs} corresponds to a shot gather in the center of the acquisition generated

in the true model. $d_{cal}(V_P)$ corresponds to data generated in different models similar to the true model, with as only varying parameter the layer velocity (ranging from $\pm 100 \text{ m.s}^{-1}$ around the layer velocity of the true model). The results of this analysis are presented in Figure 2.16. The L^2 results are coherent with the expectation: the misfit function is convex with respect to the variation of the layer velocity and presents a minimum when the velocity of the layer used to generate $d_{cal}(V_P)$ is similar to the one of the true model used to generate d_{obs} , so respectively 1400 and 1600 m.s^{-1} . For AWI, we observed that the minimum is not aligned with the correct velocity (1440 m.s^{-1} in the first case, 1610 m.s^{-1} in the second). This exhibits the loss of sensitivity of AWI in this case, explaining the failure of convergence of the FWI. Only reducing the σ below 0.04 s would make AWI behaves more like L^2 and converge to a result similar to the one of NIM. We do not think such a parameterization is interesting as it prevents the advantages introduced by AWI, which is improved convexity, and still introduces artifacts in the reconstructed model and computational overhead compared to L^2 .

This second FWI test is a good illustration of the potential limitation that an alternative misfit function mainly focused on resolving time-shift could introduce. Here AWI and NIM have difficulties providing satisfactory results when the main arrival is correctly predicted. The point to point approaches that classical L^2 procures is here the “reference”, making possible to fit the small perturbation properly following the main arrivals. AWI and NIM being more kinematic oriented, it is not surprising that this setup is challenging for such formulations.

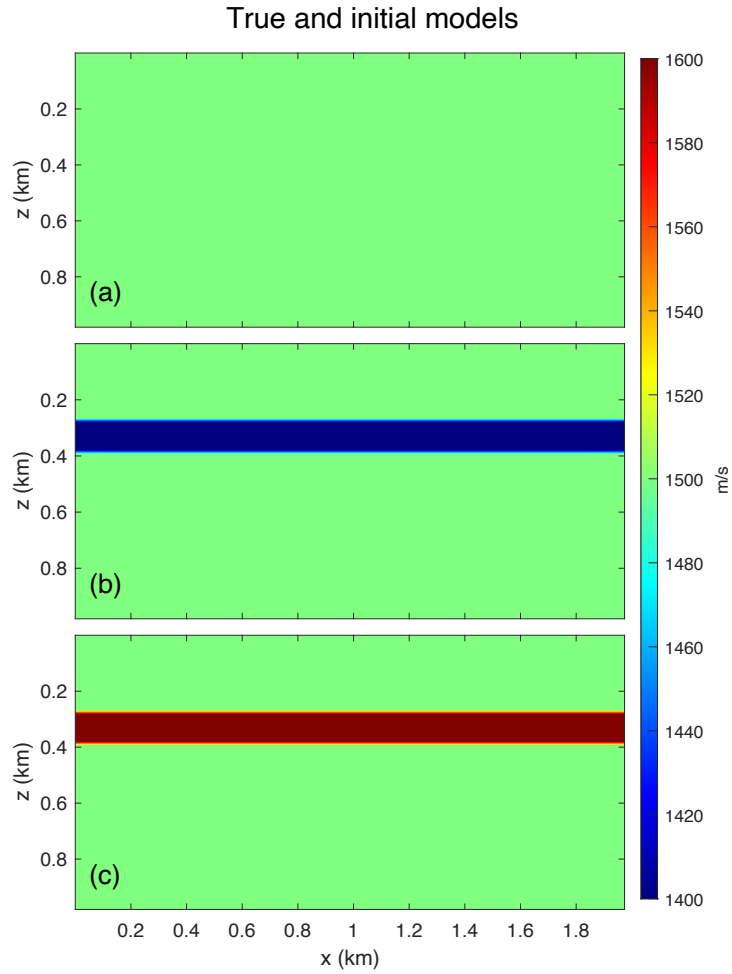


Figure 2.10: FWI Test 2: (a) Homogeneous initial model ($V_P = 1500 \text{ m.s}^{-1}$). (b) True model with a negative V_P anomaly, (c) true model with a positive V_P anomaly.

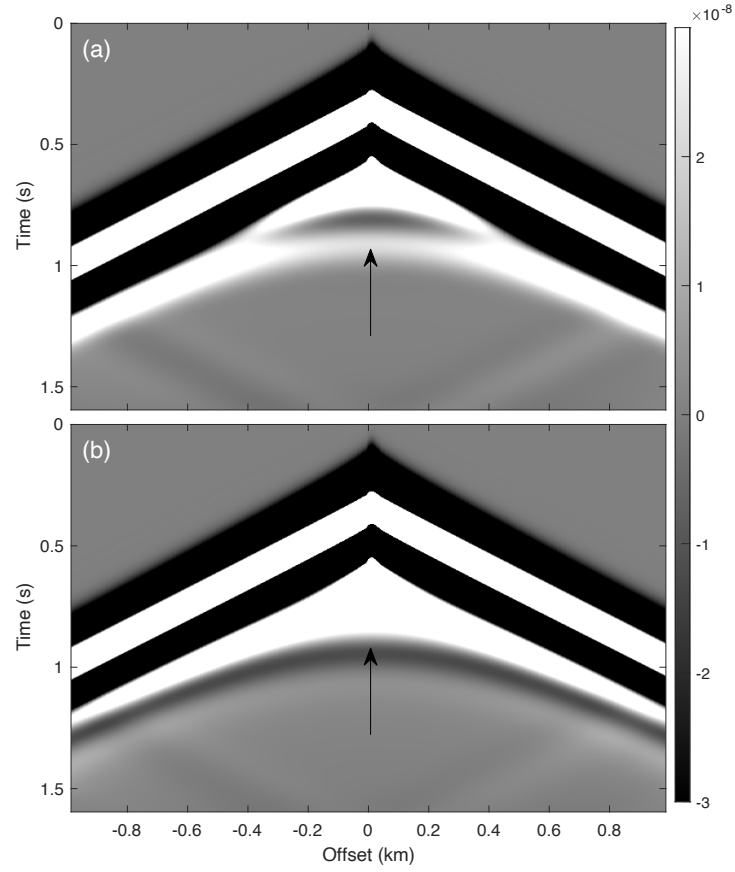


Figure 2.11: FWI Test 2: 2D common shot gathers for (a) observed data for negative V_P anomaly and (b) positive V_P anomaly. Strong clipping is applied to enhanced the reflected waves which are approximately 100 times smaller in amplitude than the transmitted waves. Polarity reversal of the reflected waves is pointed with black arrows.

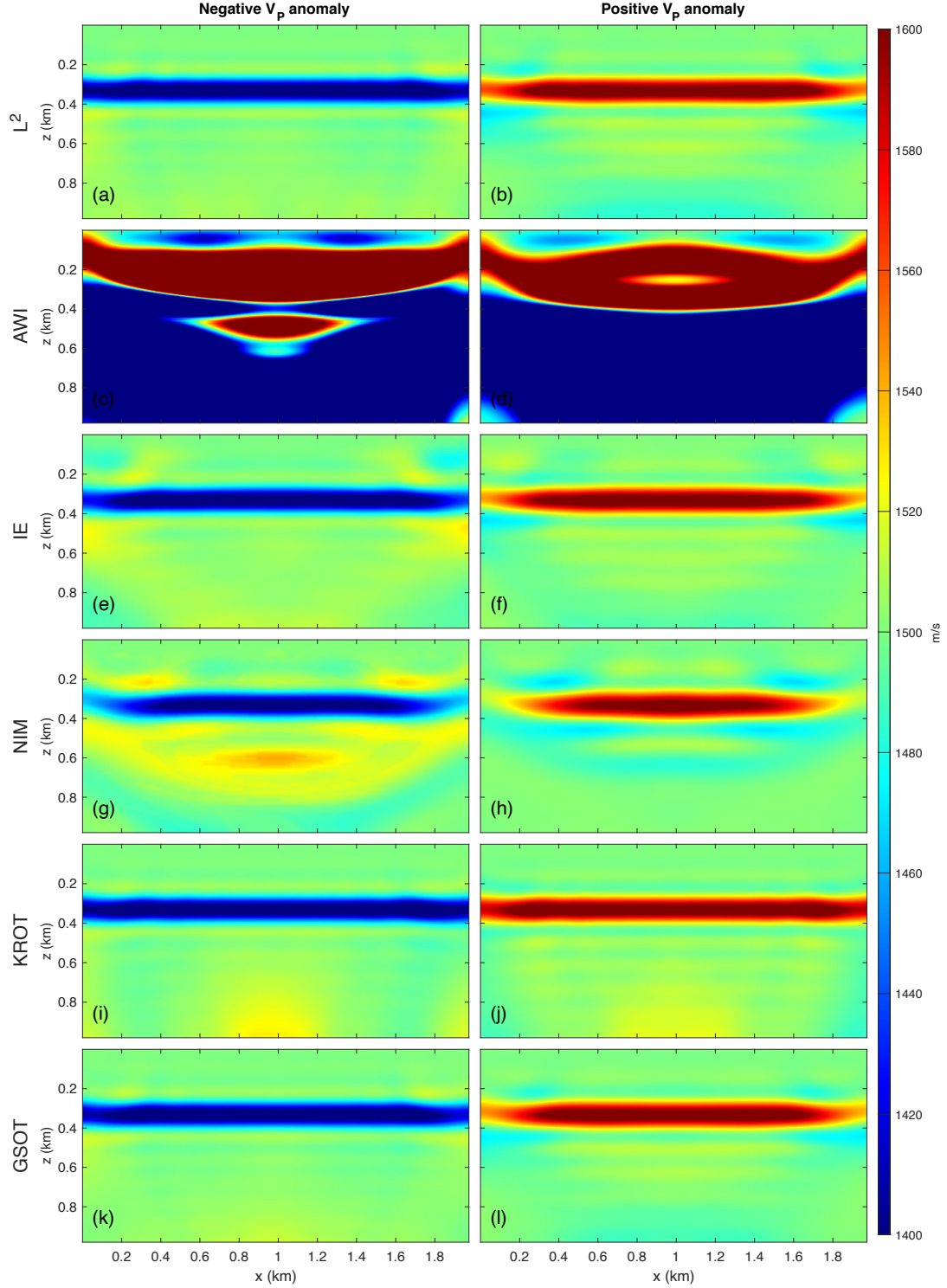


Figure 2.12: FWI Test 2: FWI layer benchmark results. Left column corresponds to a negative V_P anomaly while the right column to positive V_P . The subfigures under respectively correspond to final reconstructed V_P model obtained with FWI using L^2 (a,b), AWI (c,d), IE (e,f), NIM (g,h), KROT (i,j) and GSOT (k,l).

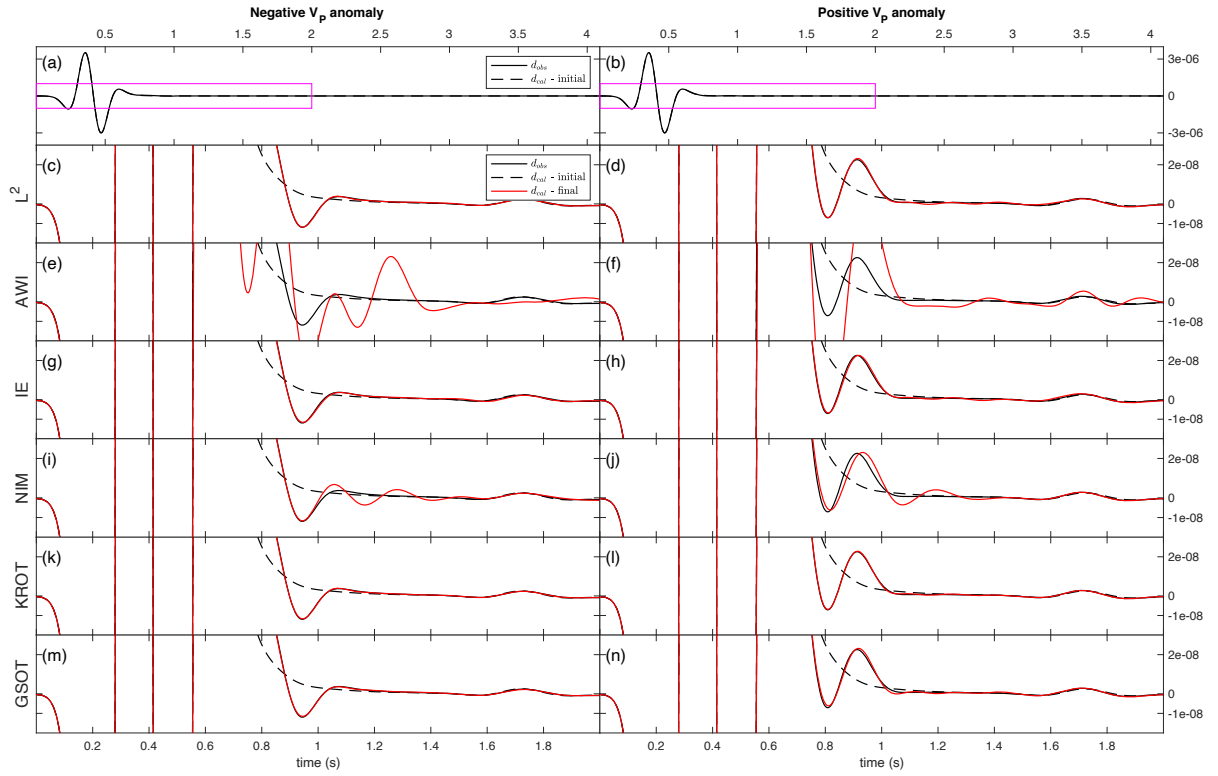


Figure 2.13: FWI Test 2: Extracted traces for a center shot at zero offset. Observed data are in solid black, synthetic data in the initial model in dashed black, and final reconstructed synthetic data in solid red. The left column corresponds to a negative V_P anomaly while the right column to positive V_P . (a,b) shows the complete traces. Each subfigures under are cropped on the magenta box to emphasize the polarity reversal introduced by the layer. They correspond to traces calculated in the reconstructed V_P model using respectively L^2 (c,d), AWI (e,f), IE (g,h), NIM (i,j), KROT (k,l) and GSOT (m,n) misfits functions.

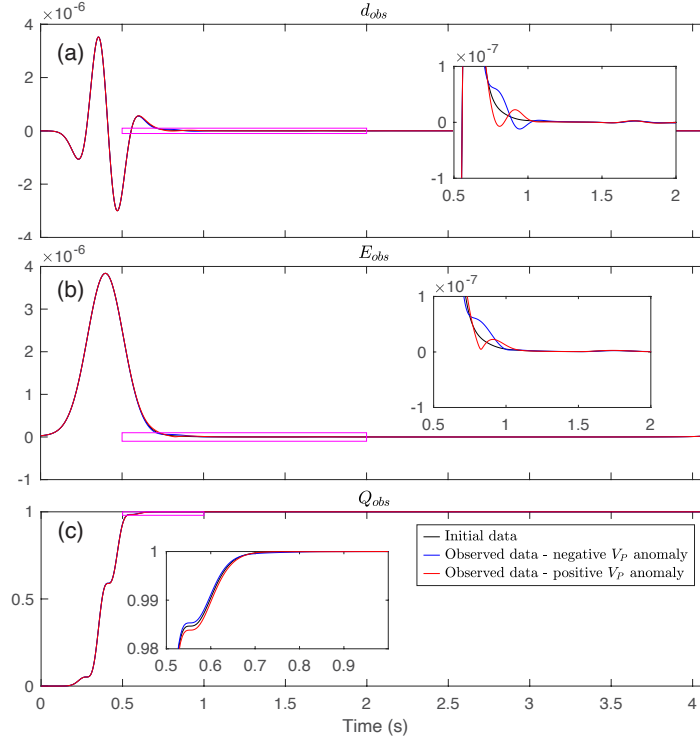


Figure 2.14: FWI Test 2: Extracted traces for (a) d_{obs} and d_{cal} , (b) E_{obs} and E_{cal} , and (c) Q_{obs} and Q_{cal} for center shot at zero offset. Data associated to negative V_P anomaly are in solid blue, and positive V_P anomaly in solid red. Calculated data is in solid black. Magenta box corresponds to the enlarged area.

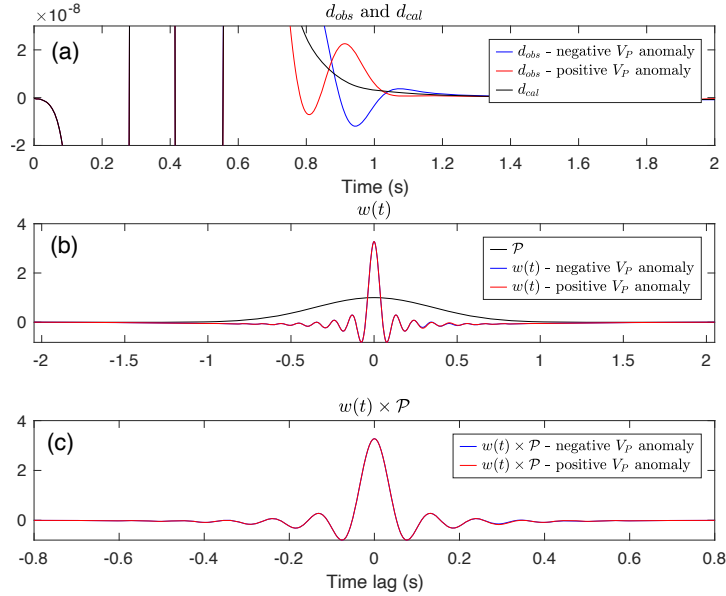


Figure 2.15: FWI Test 2: (a) Extracted traces d_{obs} , (b) Wiener filters $w(t)$ and \mathcal{P} , and (c) $w(t) \times \mathcal{P}$ for center shot at zero offset at first iteration. Data associated to negative V_P anomaly are in solid blue, and positive V_P anomaly in solid red.

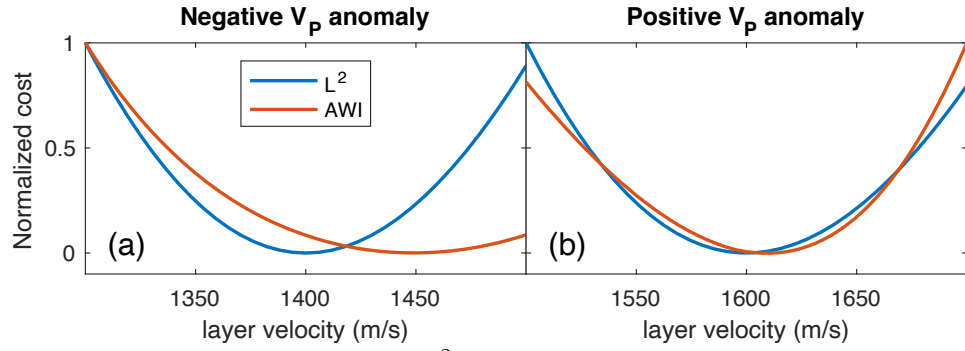


Figure 2.16: FWI Test 2: Misfit value for L^2 and AWI with respect to the layer velocity for the two inversion cases: (a) negative and (b) positive V_P anomaly. Misfit is calculated between d_{obs} (data in the true medium) and a $d_{cal}(V_P)$ generated in a medium with a correct background $V_P = 1500 \text{ m.s}^{-1}$ and a as only varying parameter the layer V_P , ranging from $\pm 100 \text{ m.s}^{-1}$ around the original velocity of layer in d_{obs} .

2.5 Marmousi case study: toward a more realistic case study

2.5.1 Common framework

We design a synthetic case study using the Marmousi II P-wave velocity model (Figure 2.17) (Martin et al., 2006) to continue our analysis on a more realistic FWI configuration. We use a fixed spread surface acquisition model with 128 sources and 169 receivers. The source spacing is 132 m, and the receiver spacing is 100 m. The data is generated with a 4 Hz centered Ricker wavelet high-pass filtered to remove energy below 2 Hz (wavelet is visible in Figure 2.24). The recording time is set to 7 s. PML absorbing layers are used on the bottom and lateral sides of the model to mimic a medium of infinite extension in these directions, while a free surface condition is applied on top.

In the first case, referred to as “inverse crime inversion”, we model the data in the constant density acoustic approximation and use the same grid for modeling and inversion to remain in the inverse crime settings. The mesh spacing is 25 m in this case.

In the second case, referred to as “more realistic inversion”, we use the variable density Marmousi II model and a refined 10 m grid to generate the data. White noise bandpassed between 2 Hz and 10 Hz is added to the data to reach a signal to noise ratio of 15%. The inversion is done on a 25 m grid, using a density model derived from the initial V_P model through Gardner’s law (Gardner et al., 1974) (Figure 2.18). A wavelet estimation is done before inversion. Performing the inversion in this more realistic framework, away from the usual inverse crime settings, makes it possible to assess the effects of incorrect amplitude prediction on the different misfit functions and better judge their usability toward field data applications.

The optimization is performed using the l -BFGS algorithm. The regularization of the gradient is defined as 0.3 of the local wavelength. Pseudo-hessian preconditioning is used (Choi and Shin, 2008; Yang et al., 2018a). The lower and upper V_P boundaries for the inversion are respectively set to 1000 m.s^{-1} and 5200 m.s^{-1} . Inversion is performed without any frequency continuation approaches or other multi-scale strategies for all the misfit functions considered (including L^2).

2.5.2 Inverse crime inversion

2.5.2.1 Case study description

We rely here on two starting models. The first one, called $S500$ (Figure 2.17 c), is derived from the true V_P model using a Gaussian smoothing with a correlation length of 500 m. This starting model preserves the long wavenumber content of the true model. The second one is a linearly increasing vertical 1D (Figure 2.17 d) model, ramping from 1500 m.s^{-1} at seabed to 4500 m.s^{-1} at depth. This initial model does not contain long-wavelength structures inherited from the true model. The data-fit obtained through this initial model (Figure 2.19 b) is affected by cycle-skipping. In comparison, the data-fit obtained with $S500$ is globally better (Figure 2.19 a), with more in-phase arrivals (especially on 0 to 3 km offset diving waves).

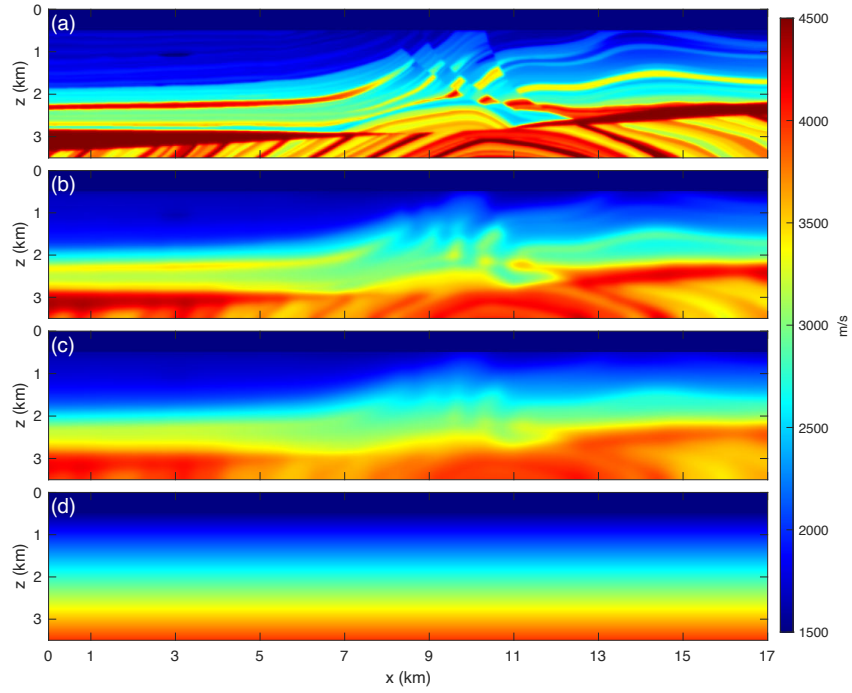


Figure 2.17: Test 3.a: (a) True Marmousi II model. (b) $\mathcal{S}250$ initial model, (c) $\mathcal{S}500$ initial model and (d) 1D initial model.

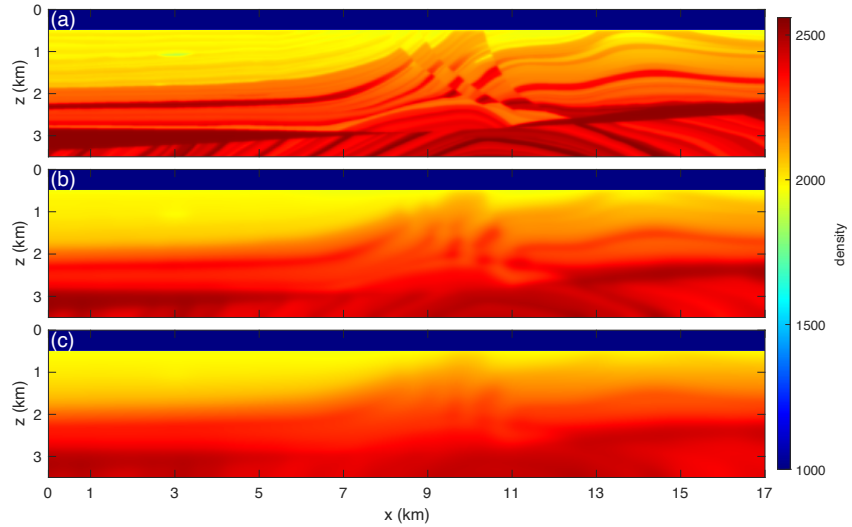


Figure 2.18: Test 3.a: Density model obtained from V_P using Gardner's law for: (a) True Marmousi II model, (b) $\mathcal{S}250$ initial model and (c) $\mathcal{S}500$ initial model.

2.5.2.2 Results starting from $\mathcal{S}500$ initial model

Reconstructed V_P results for all the selected misfit functions are presented in the left column of Figure 2.20. The associated data-fit obtained after FWI are presented in Figure 2.21 (for a common shot gather in the middle of the acquisition).

In this model, the L^2 results give, at first order, a good reconstruction of the Marmousi model. However, we can observe on the left part that the horizontal layers are not correctly

reconstructed and present an up-shift ($0 < x < 3$ km), associated with a low-velocity anomaly on the shallow left part of the model (around $x = 1$ and $z = 0.8$ km). This corresponds to the part where strong reflections are generated. Because the background velocity is incorrectly predicted in the early iterations, the arrivals corresponding to these reflections are cycle-skipped. As we illustrated earlier, L^2 misfit function being unable to tackle cycle-skipping effects, FWI cannot update the medium correctly to fit these arrivals.

AWI provides a clear improvement over classical L^2 : the horizontal layers are correctly positioned on the left part. The central part at depth ($8 < x < 13$ km, ($z > 2$ km) is improved compared to L^2 , with better contrast and more lateral coherency in the layers structure. Moreover, the low-velocity anomaly on the shallow left part of the model is removed. These results are obtained with $\sigma = 0.25$ s and $\zeta = 10^{-5}$.

IE also improves the reconstructed model. The relatively small improvement in cycle-skipping robustness introduced by the envelope is enough to mitigate the artifacts on the left part of the model ($0 < x < 3$ km) and flattens the layers compared to the L^2 result. One drawback is the slight degradation in the reconstruction of the central part at depth ($9 < x < 13$ km, $z > 2$ km). Still, such a simple formulation is enough to improve the FWI workflow over the classical L^2 in this case.

The case of NIM misfit is interesting. Here, we can see that it fails to converge, producing an erroneous reconstructed model. This illustrates the limitation of NIM when applied to more realistic cases where the data contains multiple arrivals, multiple phases, and potentially mixed phases. Integrating all these pieces of information into a single observable (the cumulative distribution) does not make it possible to reconstruct the subsurface velocity. As we can see in the data-fit, NIM can also not fit the vast majority of the signal.

The KROT misfit function, as AWI and IE, can prevent the appearance of the left side artifacts observed with the L^2 reconstruction ($0 < x < 3$ km). As for IE, the relatively small improvement in terms of attraction valley width provided by KROT is sufficient to improve the results significantly. Besides, KROT can account for the lateral coherency of the data, which might also help stabilize the inversion.

Finally, GSOT also produces a significant improvement over the L^2 result, with almost no artifacts on the reconstructed model. We use $\tau = 0.25$ s, similarly to AWI parameterization. This illustrates that GSOT, as AWI, while being able to significantly enlarge the valley of attraction of the misfit function on simple convexity cases, can also be used in a more realistic framework that mixes transmitted and reflected energy with relatively complex multi-arrival data.

Regarding the data-fit, excepted for NIM, all the misfit functions can provide a good data-fit in this case.

2.5.2.3 Results starting from 1D initial model

Reconstructed models from the 1D initial model are presented in the right column of Figure 2.20. The associated data-fit obtained after FWI are presented in Figure 2.22.

Starting from this initial model, strong artifacts appear on the L^2 results. We observe long-wavelength low-velocity anomalies on both left and right parts of the models typical of cycle-skipping induced artifacts. The data-fit analysis confirms this observation: only early

arrivals in the near offset are correctly fitted. Diving waves arriving at larger offsets on the left (5 s and -7.5 km offset) and right parts (4 s and 6 km offset) of the gather are cycle-skipped.

Without any surprise, NIM cannot provide a meaningful estimate of the V_P model, as it is already the case starting from the $\mathcal{S}500$ initial model.

As the 1D initial model generates large time-shifts, and since both IE and KROT are only marginally improving cycle-skipping robustness, it is not surprising to observe artifacts on the associated reconstructed V_P models. IE results present strong artifacts, mainly on the left part of the model ($0 < x < 7$ km), while the shallow right part ($11 < x < 16$ km, $x \leq 2$ km) presents an improved reconstruction compared to L^2 . This is confirmed by the data-fit, where all the arrivals on the right part (offset between 1 and 8 km) are correctly predicted, whereas data-fit on the left part (offset between -8 and 0 km) is degraded compared to L^2 data-fit. Conversely, KROT provides a more accurate reconstruction in the left part of the model. The strong low-velocity anomalies observed in the left part ($0 < x < 7$ km) of the L^2 and IE reconstructions are reduced and appear only in the deep part of the model ($z > 2$ km). This is consistent with the data-fit, where we see that using KROT, the long offset diving waves for negative offsets of the shot gather are correctly fitted.

AWI manages to provide a clear improvement over classical L^2 , mainly on the center part of the model ($5 < x < 14$ km), while some artifacts on both sides of the reconstructed model are still present. These parts are more difficult to reconstruct as they are illuminated by waves traveling along the longest paths of the medium, increasing cycle-skipping risk. We set $\sigma = 0.6$ s to try to capture as large as possible time-shifts. To get the best results possible, we used $\zeta = 10^{-2}$ to obtain the helping smoothing effect required to tackle large time-shifts introduced by this initial model. The data-fit obtained with AWI is good, with only some out-of-phase arrivals for late diving waves (around -8 to -6 km offset).

Using GSOT, the reconstructed V_P model also presents a clear improvement over classical L^2 . We use $\tau = 0.6$ s in this case. At the first order, most of the artifacts are removed. Still, some artifacts are present close to the edges ($0 < x < 1$ and $15 < x < 17$ km), which is expected from the lack of illumination in these parts. Some other artifacts are visible in the center part of the structure at depth ($10 < x < 12$ km and $2 < z < 3.5$ km). The data-fit appears to be good, with no out of phase arrivals for all offsets.

2.5.2.4 Error reduction analysis

Besides this qualitative analysis of the results, we can provide quantitative comparisons by analyzing the data error and model error evolution along with iterations. We use here the following relative L^1 model error definition

$$Err(V_P) = \frac{100}{M} \sum_{i=1}^M \frac{|V_{P,i} - V_{P,i}^{true}|}{V_{P,i}^{true}} \quad (2.38)$$

where V_P^{true} is the true model, M the number of points in the model and i denotes one pixel of the grid used to describe the models at the discrete level.

For the second experiment only (1D initial model), we present the evolution of

- the convergence rate (misfit error with respect to the iterations);

- the L^2 convergence rate (L^2 error with respect to the iterations);
- the model convergence rate (model error with respect to the iterations);
- the model vs. data convergence rate (model error with respect to the misfit error).

For model error, we truncate the model by 1 km on the left and right sides and 625 m at depth to remove the model areas that are not well illuminated.

For alternative misfit function definition, the L^2 -based convergence rate is interesting as moving away from L^2 local minima should be made visible by an increase of the L^2 error with respect to the iterations. Also, the fourth item is interesting, as, ideally, we look for a monotonic decrease of the model error with respect to the misfit error. Besides, to improve the readability, we have excluded from these figures the results corresponding to NIM. The method does not produce reliable results in both cases.

The error reduction analysis is shown in Figure 2.23. First, we observe that KROT and AWI present a relatively slow convergence rate on the cost evolution, while IE and GSOT have a faster convergence rate. L^2 convergence is in between. KROT follows more or less the same as the L^2 misfit function. This is somehow expected, as the valley of attraction of KROT is expected to be similar to the one of the L^2 misfit function. Note, however, that in the early iterations, KROT displays a small increase of the L^2 error, which clearly states that the two misfit functions follow a different minimization path. IE, AWI, and GSOT display another trend: the L^2 error is increased in the first iterations before being strongly decreased in a second stage. The substantial decrease of the L^2 error appears the latest for AWI (after 100 iterations) and the earliest for GSOT (after 30 iterations). GSOT achieves the smallest L^2 misfit, followed by AWI and KROT. The model convergence rate classifies the misfit functions into two groups: one that does not reduce model error compared to the starting point, with L^2 , IE, and KROT; and a second group that decreases the model error with AWI and GSOT. In the second group, only GSOT provides a constant decrease with respect to the iterations, while AWI start to increase the model error until 100 iterations, followed by a decrease. The final reduction of model error obtained with KROT and IE are smaller than the one attained by the L^2 , still, this does not explicitly compared to better interpretable results overall. AWI and GSOT obtain the best reduction of model error. Finally, looking at the model vs. data convergence, only GSOT provides a quasi-monotonic decrease. For all the others, the model error starts by increasing with the reduction of the misfit.

2.5.3 A more realistic inversion

2.5.3.1 Case study description

Similar to the previous inverse crime inversion, we perform FWI starting from two different initial models. The first one is derived from the true Marmousi model using a lighter Gaussian smoothing, referred to as the $S250$ model with a correlation length of 250 m (Figure 2.17 b). The second one is the $S500$ model already used in the inverse crime settings (Figure 2.17 c). We did not re-use the 1D initial model as it proves to be too difficult for any alternative misfit functions to provide convergence in this more realistic case. Not predicting the data to machine precision generates a more challenging benchmark.

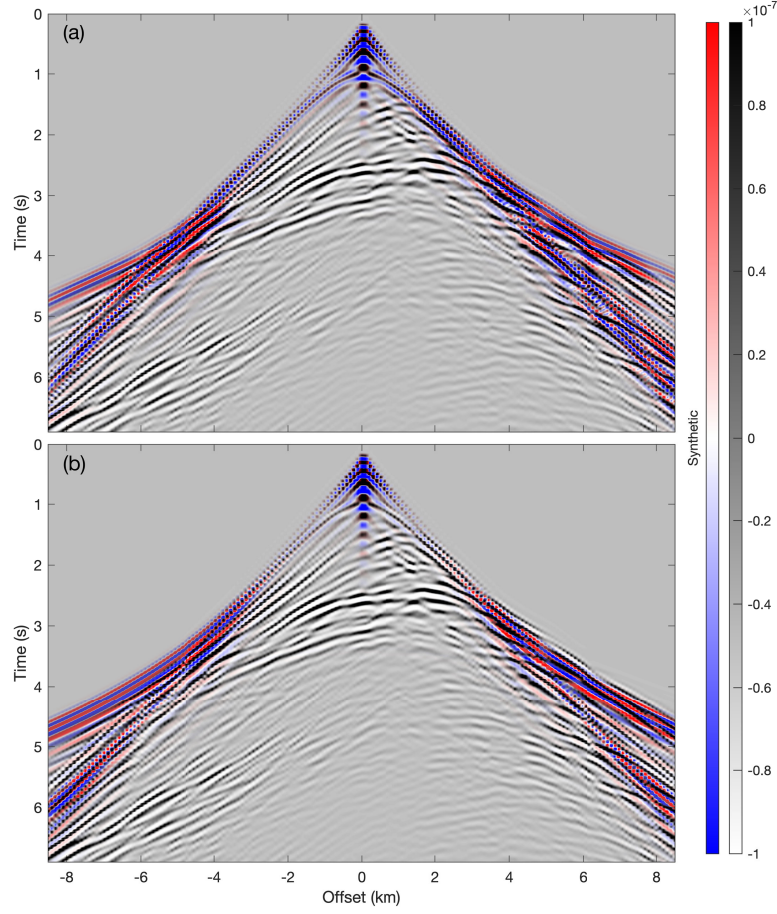


Figure 2.19: Test 3.a: Common shot gather (CSG) for field data overlapped by synthetic data in (a) $S500$ initial model, and (b) 1D initial model. Field data in black and white, overlapped by red to blue synthetic data with transparency. Red and white visible mean out of phase, black and blue mean in phase.

The wavelet used for FWI is obtained through a source estimation in the initial model based on short offset (100 m) only to decouple the influence of the initial V_P model as much as possible. The obtained inverted wavelet is presented in Figure 2.24. We can observe that the inverted wavelet is close to the true wavelet, but some noticeable amplitude and waveform differences are visible. These amplitude effects are induced by the use of a “true” density model for the data generation compared to Gardner’s one (Figure 2.18) used for wavelet estimation and to the white noise added to the data.

The data-fits for these two initial models are presented in (a) of Figures 2.26 and 2.27. As expected, the data-fit is better using $S250$ initial model, while the data-fit generated with $S500$ initial model displays more out of phase arrivals.

We compare the results obtained using L^2 , AWI, IE, KROT, and GSOT misfit functions. We do not include NIM results here, as we have already shown how the method fails to produce meaningful results in the previous inverse crime settings. A maximum of 500 FWI iterations is performed for both initial models.

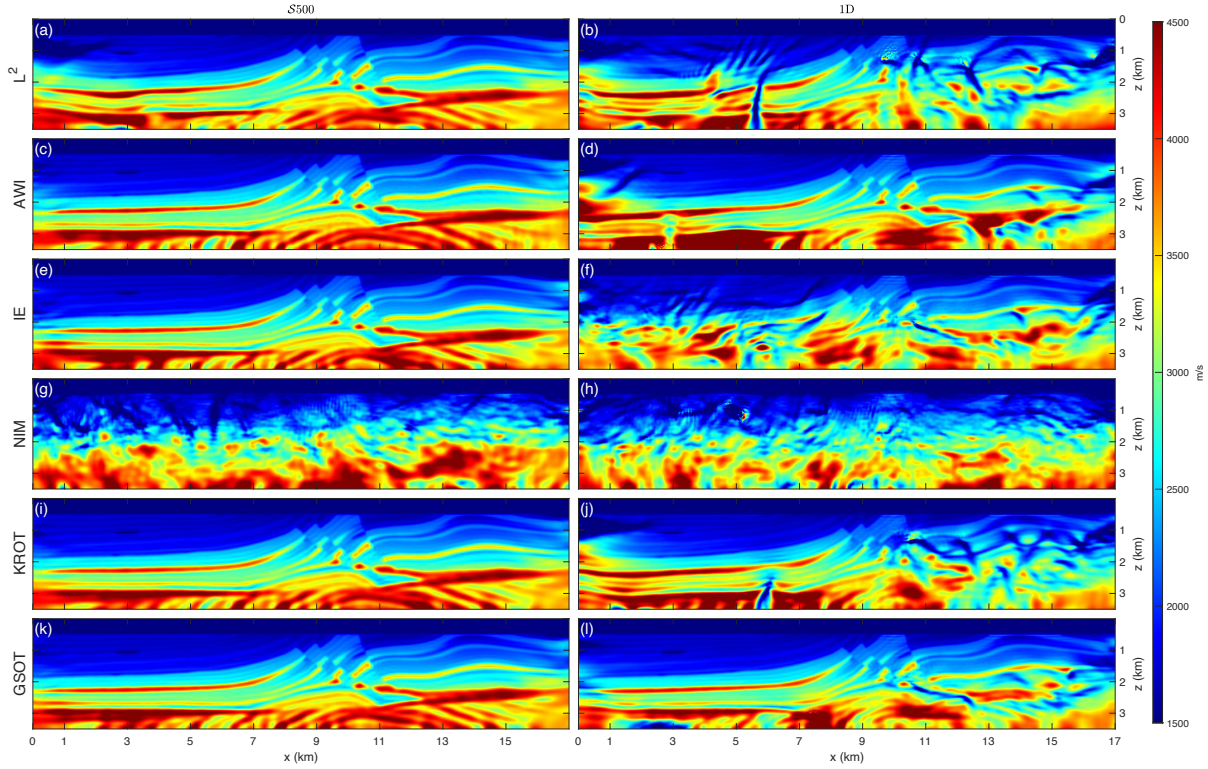


Figure 2.20: Test 3.a: Inverse crime FWI final reconstructed V_P model for Marmousi. Left column corresponds to $S500$ initial model, right column to 1D initial model. The lines respectively correspond to the final reconstructed V_P model using L^2 (a,b), AWI (c,d), IE (e,f), NIM (g,h), KROT (i,j) and GSOT (k,l).

2.5.3.2 Results starting from $S250$ initial model

Starting from the $S250$ initial model (Figure 2.25 left column), the main expected difference with the previous “inverse crime” setup is an inaccurate amplitude prediction (which would be the case if considering field data). Data-fits are presented in Figure 2.27. Classical L^2 can provide an acceptable result. Good reconstruction in the well-illuminated area is achieved, with no visible artifacts in the center part and only a small low-velocity artifact visible at $x = 2$ km $z = 0.8$ km and a high-velocity artifact at $x = 16$ km $z = 1$ km. The data-fit obtained with L^2 is quite satisfying with most of the arrivals in phase.

This time, IE results are clearly degraded compared to the classical L^2 one. The reconstructed V_P model is tainted with high wavenumber oscillation and strong artifacts. This is an indication that the IE approach is sensitive to a correct amplitude prediction. This validates the interest of a more realistic framework, making us able to detect this kind of limitation. The data-fit presents many out-of-phase arrivals, coherent with the small artifacts present everywhere in the reconstructed V_P .

Again, AWI improves over the L^2 results. We use a relatively small $\sigma = 0.2$ s here as the maximum time-shifts expected are relatively small with this good initial model. We used $\zeta = 10^{-2}$ as noise requires a relatively large amount of damping, moreover as illustrated before, a larger damping value helps when facing challenging FWI setups. The deep center part is improved with a more coherent deep-layer structure. The left ($x = 2$ km and $z = 0.8$ km) and

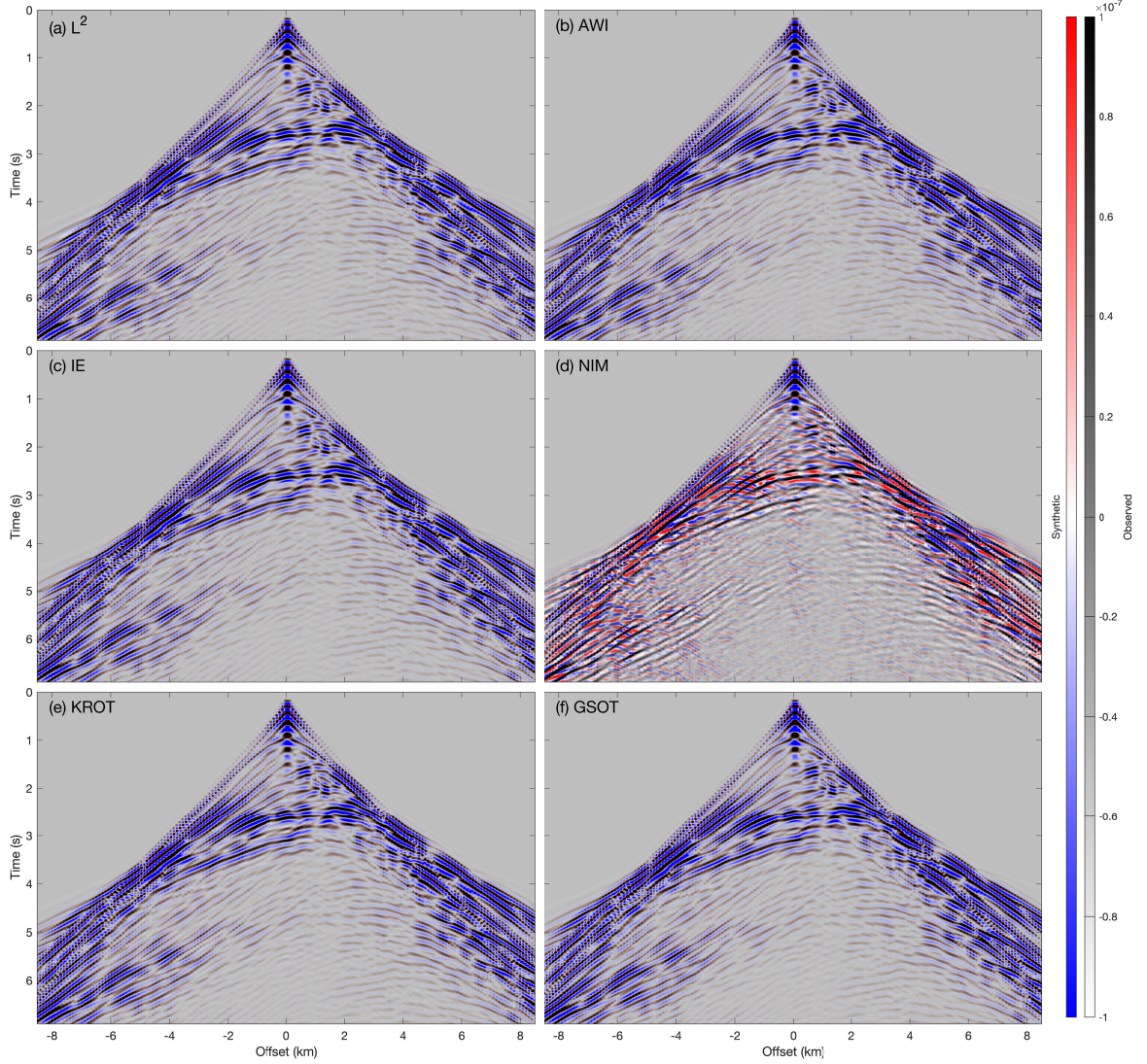


Figure 2.21: Test 3.a: Overlapped common shot gathers for synthetic data in the final reconstructed V_P model starting from $S500$ initial model vs field data. Each subfigure corresponds to misfit function, with L^2 (a), AWI (b), IE (c), NIM (d), KROT (e), and GSOT (f).

right ($x = 16$ km $z = 1$ km) side artifacts present in L^2 results are also partially mitigated. Surprisingly, the data-fit obtained with AWI is poor for large offset arrivals (from -8 km to -3 km and 3 to 8 km). This degradation of the data-fit is slightly counter-intuitive and does not correlate with the improvement of the reconstructed V_P model observed.

Finally, KROT and GSOT reconstructed models both present similar improvement compared to the L^2 one. We can observe an increase in terms of high wavenumber content. Interestingly, the deep center part ($9 < x < 13$ km, $z > 2$ km), which is the main target of interest of the Marmousi model (an anticlinal structure) is more resolved using KROT and GSOT compared to L^2 . For GSOT, we use $\tau = 0.2$ s in this case. The data-fit obtained with both methods is good, with almost all arrivals in phase. Only some first arrivals between -4 to -2 km offset are still not well explained. The GSOT data-fit appears to be slightly better

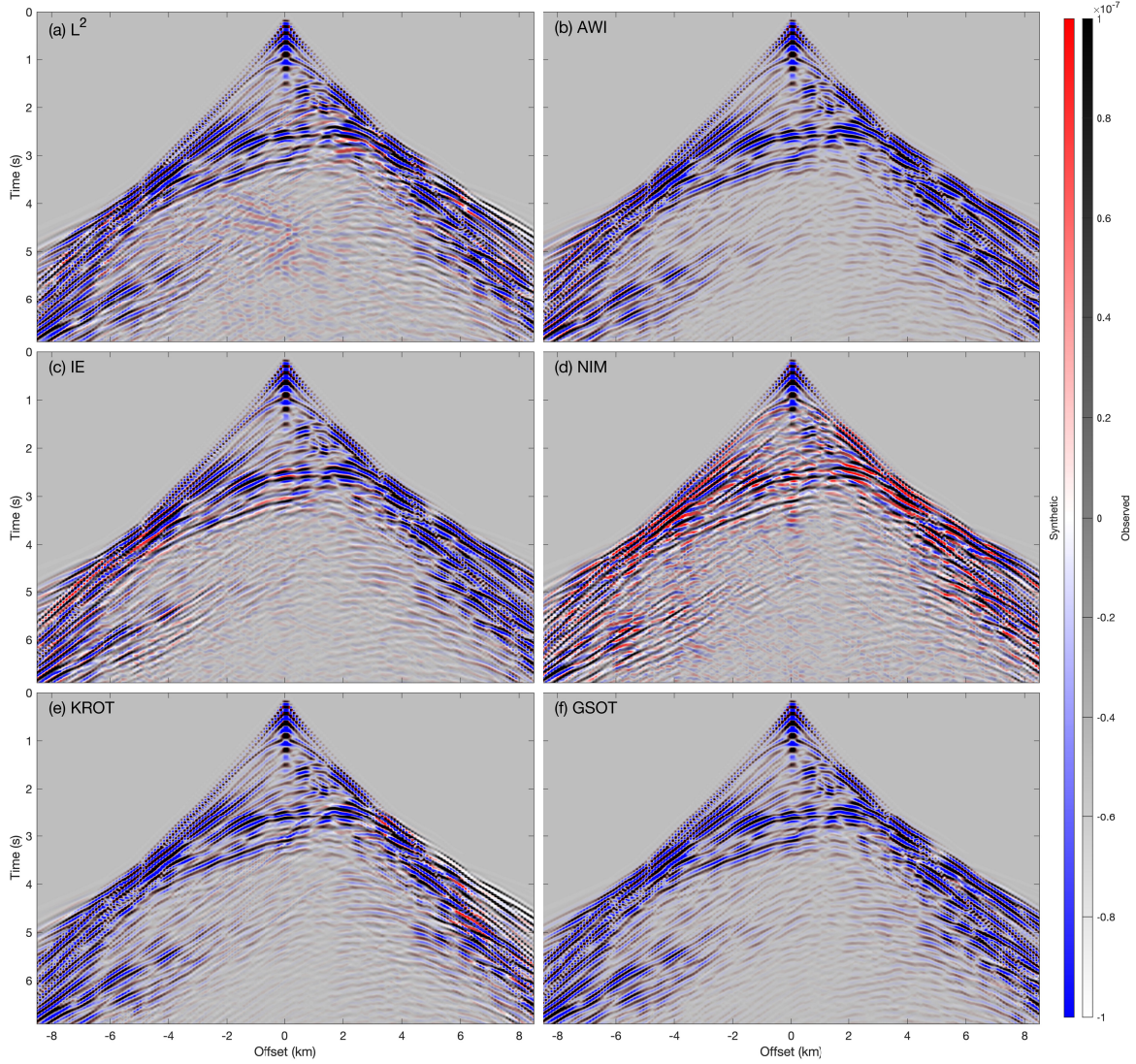


Figure 2.22: Test 3.a: Same as Figure 2.21 but starting from 1D initial model.

than the KROT one.

2.5.3.3 Results starting from $S500$ initial model

Starting from the $S500$ model, reconstructed V_P results are presented in Figure 2.25 right column, while data-fit are presented in Figure 2.27.

Here, the classical L^2 fails to reconstruct a meaningful V_P model. Many artifacts are present on the model that may come in part from cycle-skipping. This would prevent any interpretation of the reconstructed model. The data-fit present out-of-phase arrivals, even if the majority would appear to be in-phase. This again illustrates potential convergence toward a local minimum that makes possible to fit the data with non-meaningful V_P updates.

With no surprise, IE fails to reconstruct a meaningful V_P estimate. The reconstructed V_P model suffers from many artifacts. The data-fit is clearly degraded compared to L^2 , which is

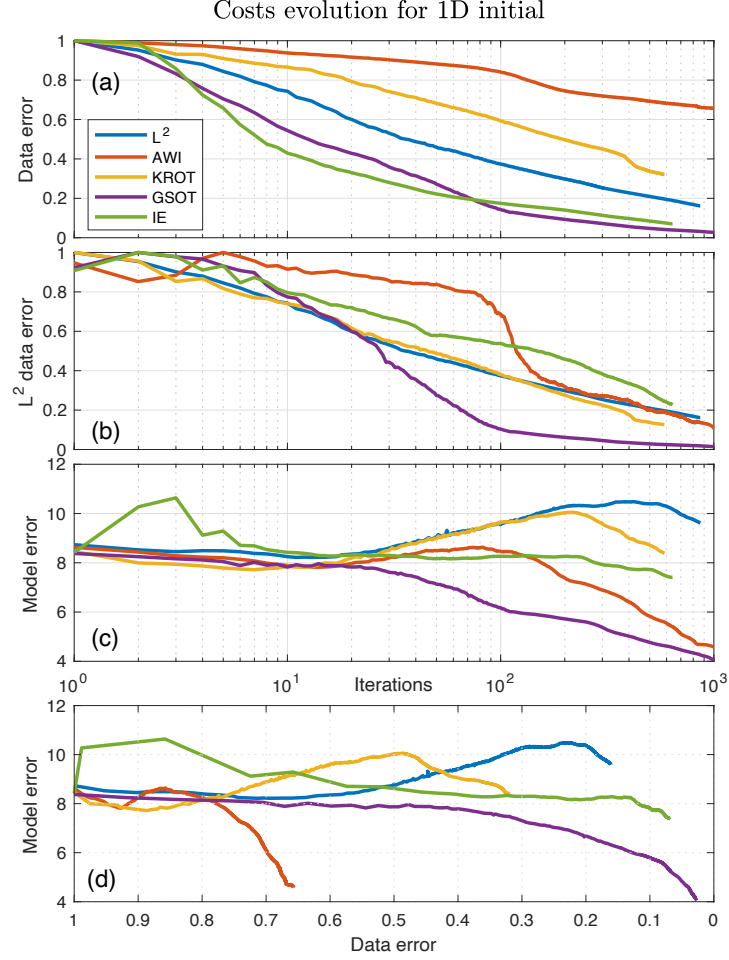


Figure 2.23: Test 3.a: Costs evolution in the inverse crime Marmousi for 1D initial model. (a) evolution of cost functions over iterations, (b) true L^2 cost evolution over iterations, (c) model error reduction over iterations and finally (d) model error vs. the data error reduction.

likely explained by the difficulty faced by IE in tackling wrong amplitude predictions compared to classical L^2 .

AWI reconstructed model produces here an improvement over L^2 or IE, with the central part and right part of the Marmousi model more or less retrieved. However, significant artifacts are present in the left part of the model ($1 < x < 6$ km) associated with an erroneous reconstruction of the central part at depth ($9 < x < 13$ km, $z > 2$ km). Here we increase σ to 0.4 s, and keep $\zeta = 10^{-2}$. The data-fit is degraded with out of phase arrivals for offsets between -8 to -3 km as well as between 2 to 8 km.

KROT produces satisfactory results here. This is interesting as KROT only marginally improves cycle-skipping robustness. Here, it manages to perform well in this complexified case. This is a good indication that the difficulties induced in this more realistic inversion are not only cycle-skipping but also amplitude mismatch (due to density) and noise. As KROT introduces lateral coherency and has a regularizing effect on noise, it is not surprising to observe a better behavior in this case. The data-fit obtained with KROT is good with almost all arrivals in phase, except for some transmitted waves from -3 to -1 km offset and some long offset arrivals

around -8 to -7 km.

Finally, GSOT provides a good reconstructed V_P model. The central part ($9 < x < 13$ km, $z > 2$ km) is well reconstructed. The layers show more lateral coherency compared with KROT. Furthermore, left side artifacts are reduced compared to KROT. Again and similarly to AWI, τ is increased to 0.4 s to account for the larger time-shifts introduced by the degraded initial model. The data-fit is also good, with improvement over the KROT for the long offset arrivals around -8 to -7 km.

2.5.3.4 Error reduction analysis

A similar analysis for the different misfit functions is presented for this inverse crime inversion of Marmousi. The model error is calculated in a similar zone as in the previous experiment.

Starting from the $S500$ initial model (Figure 2.28), we observe that KROT and AWI present again a relatively slow convergence rate (AWI being the slower), while L^2 , IE and GSOT have a faster convergence rate. The L^2 data-error is again interesting, with GSOT and KROT performing the most substantial reduction of L^2 data error (with an initial jump to pass a L^2 local minimum for GSOT at the first iteration). While IE increases the cost drastically for the first two iterations, it then fails to reduce the data error. We can note that KROT is not following L^2 misfit function behavior anymore compared to the inverse crime Marmousi case. Regarding AWI, we can observe that it starts to increase the L^2 data error until 20 iterations, then rapidly reduce for 10 iterations, to finish with a constant increase afterward. This time, the model error displays a strong increase for L^2 and IE misfit functions, which is coherent with the artifacts present in the reconstructed V_P models. AWI is also increasing the model error as it is also affected by artifacts, but less drastically than L^2 and IE, which is visible on the reconstructed V_P model. KROT and GSOT manage to decrease the model error continuously. Looking at the model vs. data convergence, only KROT and GSOT present monotonic behavior, while L^2 , IE, and AWI are increasing the model error.

2.5.3.5 Computational cost

The computational overhead induced by the alternative misfit function selected in this review varies from +2 to +30% compared to L^2 misfit. These values are coherent with the values documented in the literature. The key feature here is that even a +30% computational overhead is not a blocking feature and is affordable with modern computing facilities. For us, the key feature is the “physical” performance of the misfit function that translates into an improvement of FWI robustness.

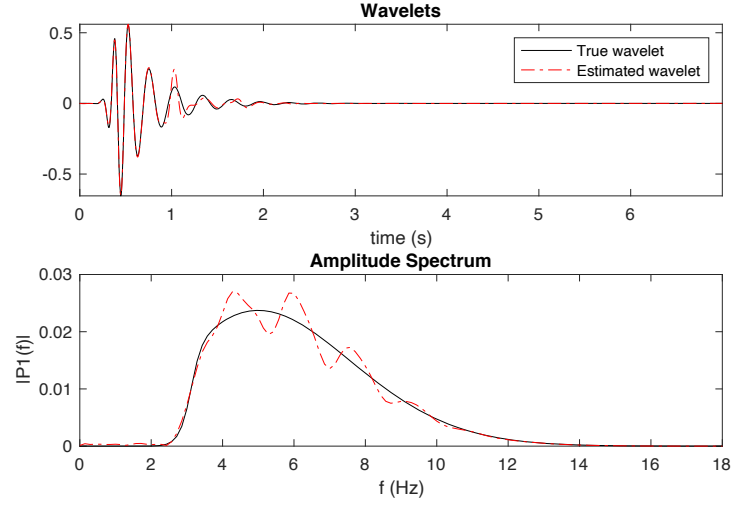


Figure 2.24: Test 3.b: Comparison of true wavelet used to generate the observed data (for both test 3.a and 3.b) versus the inverted wavelet used for inversion in more realistic test 3.b. Under the associated amplitude spectrum of both true wavelet and estimated wavelets.

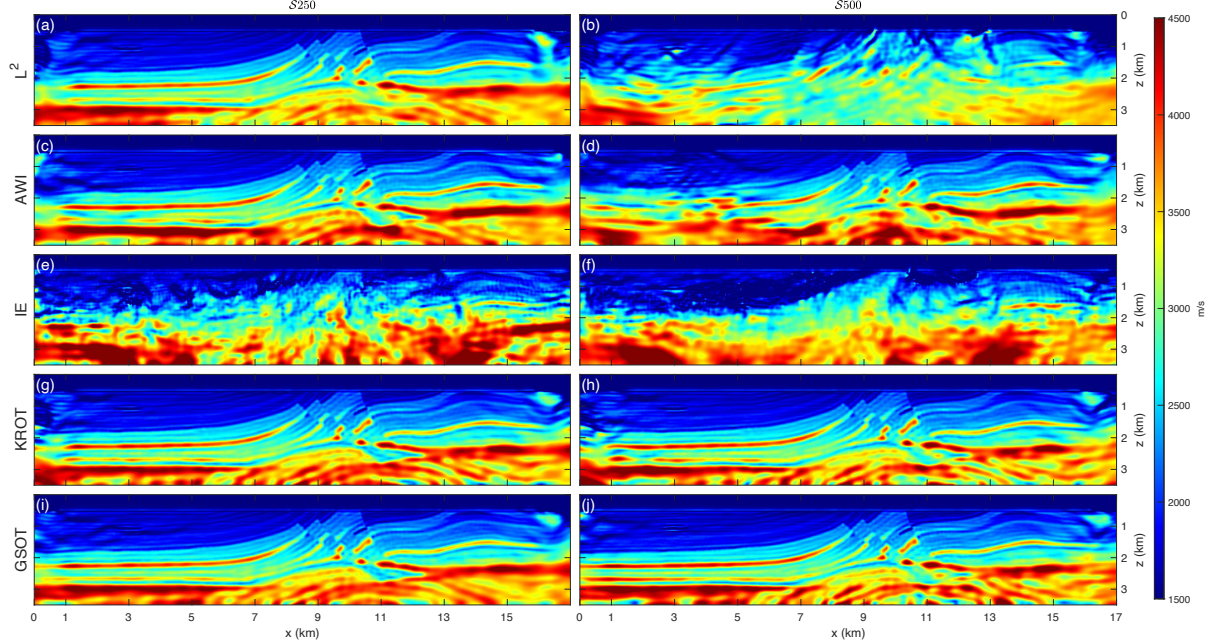


Figure 2.25: Test 3.b: More realistic FWI final reconstructed V_P model for Marmousi. Left column corresponds to $S250$ initial model, right column to $S500$ initial model. The lines respectively correspond to the final reconstructed V_P model using L^2 (a,b), AWI (c,d), IE (e,f), KROT (g,h) and GSOT (i,j).

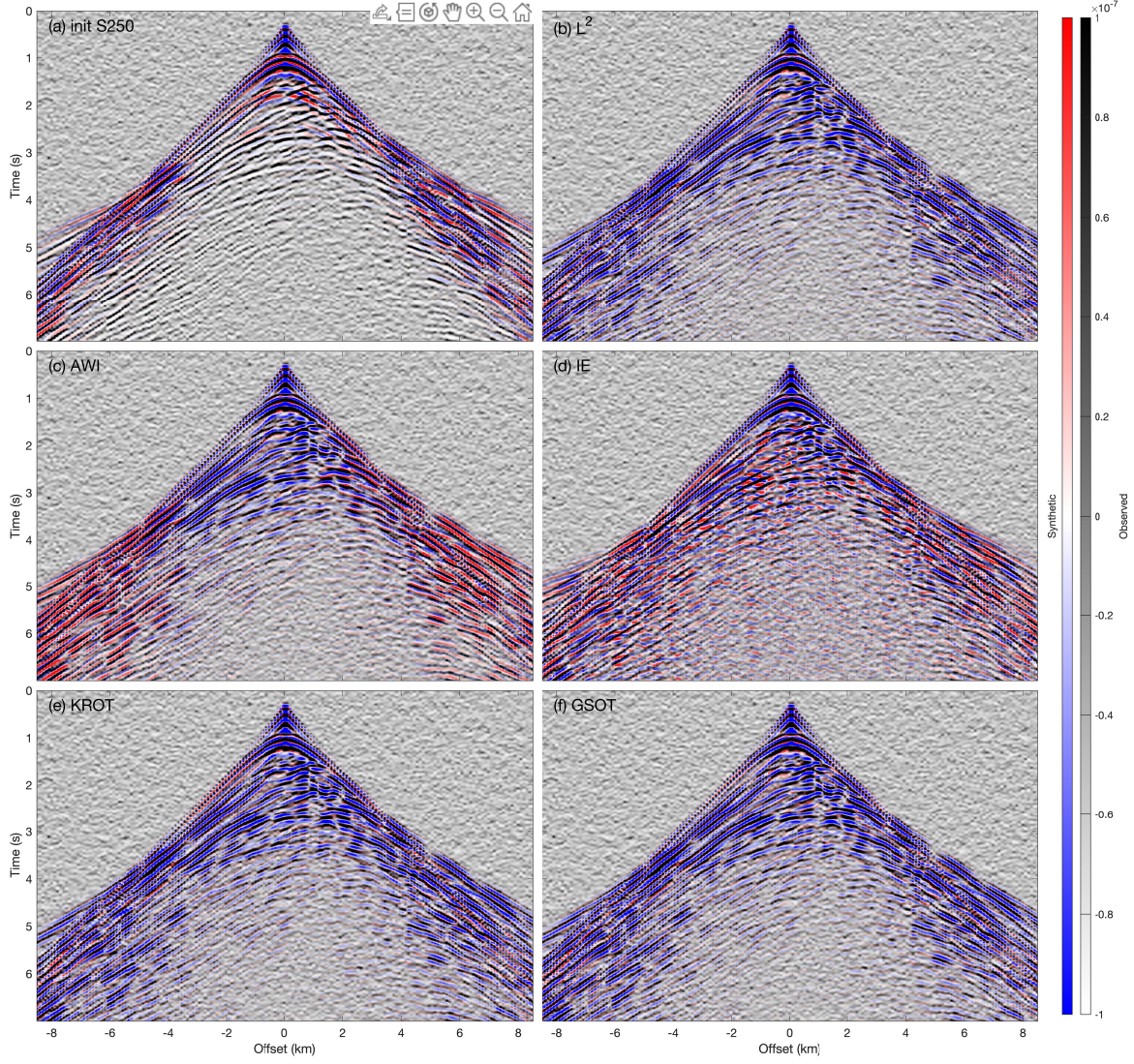


Figure 2.26: Test 3.b: Overlapped common shot gathers for synthetic data in the final reconstructed V_P model starting from $S250$ initial model vs field data. (a) corresponds to the data-fit in the $S250$ initial model. Then, each subfigure corresponds to misfit function, with L^2 (b), AWI (c), IE (d), KROT (e), and GSOT (f).

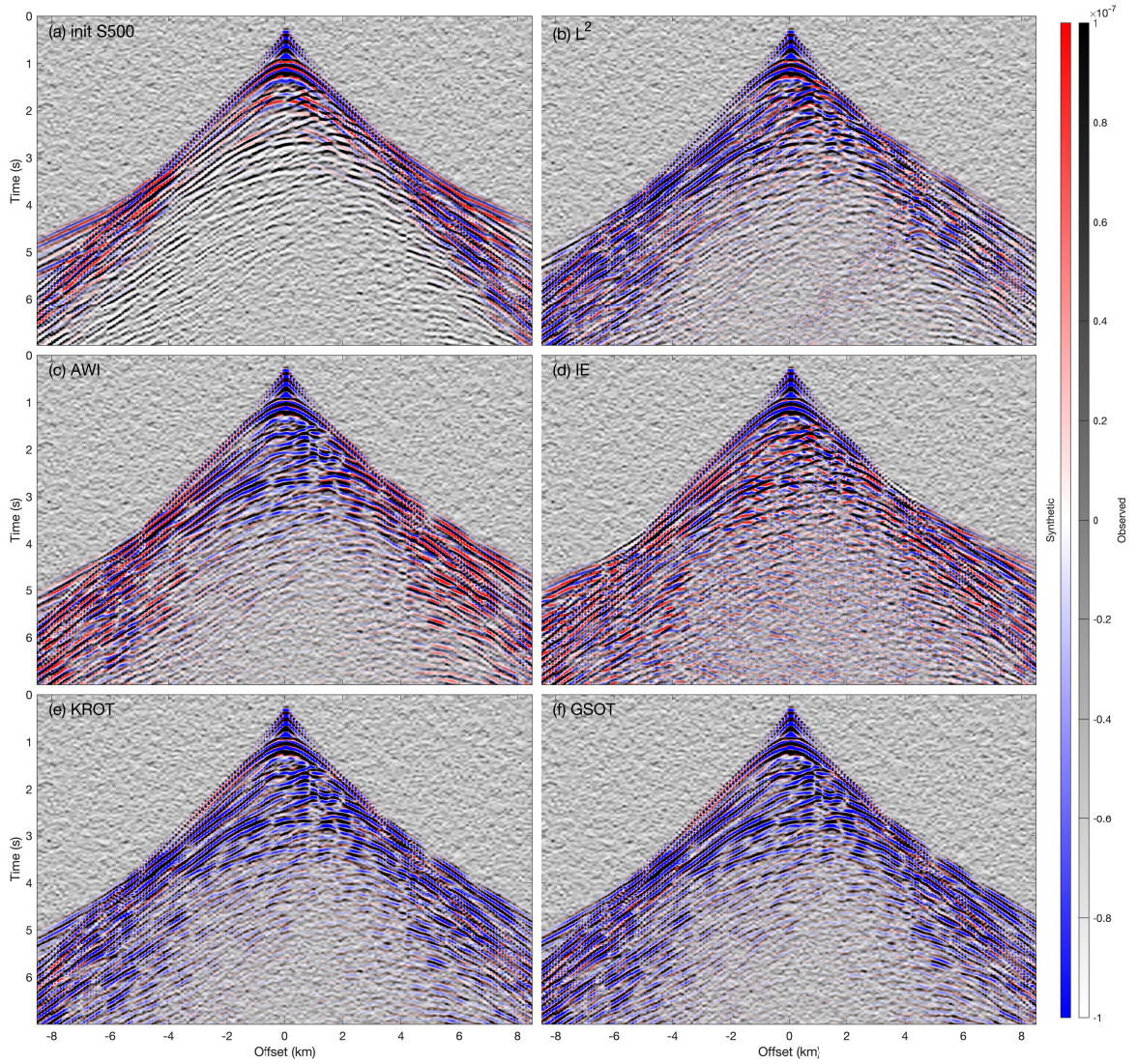


Figure 2.27: Test 3.b: Same as Figure 2.26 but starting from $S500$ initial model.

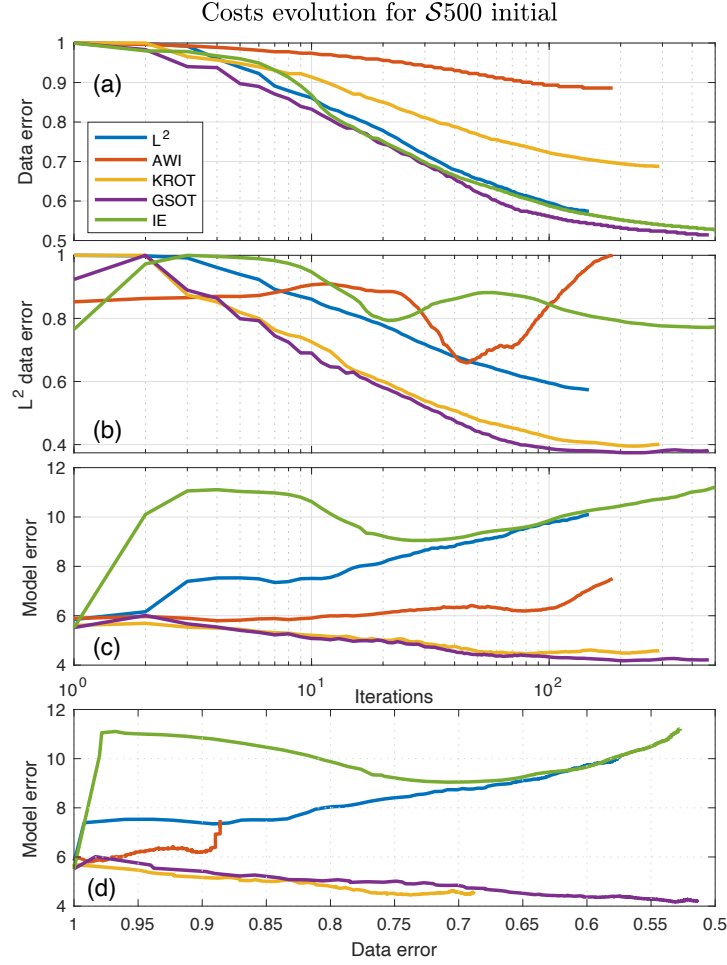


Figure 2.28: Test 3.b: Costs evolution in the more realistic Marmousi for $\mathcal{S}500$ initial model. (a) evolution of cost functions over iterations, (b) true L^2 cost evolution over iterations, (c) model error reduction over iterations and finally (d) model error vs. the data error reduction.

2.6 Discussion

Among the five misfit functions compared here, namely NIM, IE, AWI, KROT, GSOT, three of them show a significant improvement in convexity with respect to a time-shift: NIM, AWI, and GSOT. However, when applied to a realistic case (Marmousi), NIM fails to produce a meaningful V_P estimate. Conversely, for AWI, while difficulties are identified on schematic examples, including multiple arrivals (a situation known to be problematic for correlation and deconvolution approaches), satisfactory results are obtained when applied to the Marmousi case, both within and without the inverse crime settings. GSOT also appears as an interesting strategy, providing satisfactory results in all the tests performed here.

Interestingly, while IE and KROT show less robustness to strong cycle-skipping, the small increase in the valley of attraction they provide is sufficient to enhance the velocity reconstruction in the Marmousi test in inverse crime settings. However, IE fails when it comes to non-inverse crime settings, that is when noise corrupts the data, and amplitude prediction cannot be guaranteed anymore. On the contrary, KROT reveals relatively robust to these settings, probably benefiting from its ability to account for the lateral continuity of events in shot-gather representation and for the robustness of optimal transport based distances with respect to the presence of noise (Engquist et al., 2016).

From the experiment performed in this article, KROT, AWI, and GSOT appear as an interesting alternative to the least-squares distance from the perspective of field data application. In cases where no strong cycle-skipping is expected, KROT should perform well, and this is supported by several field data applications already performed on exploration data (Messud and Sedova, 2019; Sedova et al., 2019; Carotti et al., 2020). The computational cost of KROT is relatively higher than that of AWI and GSOT; however, its ability to account for the lateral coherency of the data in shot-gather panels makes it an appealing strategy. For 3D data cubes, cutting it into 2D slices and summing over the slices is a good compromise. To deal with larger kinematics inaccuracy, AWI and GSOT should be preferred options. AWI has already been successfully applied to field data (Warner and Guasch, 2015; Ravaut et al., 2017; Debens et al., 2017; Roth et al., 2018; Guasch et al., 2019; Warner et al., 2019). We, however, show here that it could suffer from some limitations in the case of complex data containing multiple arrivals. GSOT has been mostly applied to synthetic data by now (He et al., 2019a; Provenzano et al., 2020a). Nevertheless, field data applications are ongoing (Pladys et al., 2020; Górszczyk et al., 2020).

2.7 Conclusion

This article is dedicated to comparing misfit function reformulation for FWI, which aims at mitigating cycle-skipping. The first result drawn is that the link between cycle-skipping and the non-convexity with respect to time-shifts of the least-squares distance is evident from the different tests we provide. However, when no such cycle-skipping occurs (sufficiently accurate initial model), least-squares FWI performs well, even for complex data including multiple phases, mixed phases, noise, and when amplitude prediction cannot be performed accurately (as is the case for field data). Therefore, efficient reformulation of the FWI misfit function should not rely only on a better convexity to time-shifts to replace the least-squares norm advantageously but should also exhibit robustness with respect to these settings, which are always met on field data applications.

Chapter 3

Robust FWI with graph space optimal transport: application to 3D OBC Valhall data

Contents

3.1	Introduction	75
3.2	Methodology	78
3.2.1	Modeling	78
3.2.2	Inversion	79
3.3	Field data presentation and full-waveform inversion workflow	81
3.3.1	Geological situation, dataset and initial models	81
3.3.2	Full-waveform inversion workflow	85
3.4	Full-waveform inversion results	94
3.4.1	Starting from the reflection tomography model	94
3.4.2	Starting from the 1D initial model	96
3.4.3	Computational costs	100
3.5	Discussion	107
3.6	Conclusion	107

This chapter directly comes from the submitted paper Pladys et al. (2021a).

abstract

Improving full-waveform inversion to make it more robust to cycle-skipping has been the subject of a large number of studies. From the several families of approaches developed, one of the most documented consists in modifying the least-squares distance defining the discrepancy between observed and calculated data. From all the propositions made to improve and replace the least-squares distance, only a few of them have been applied to field data. One of the methods proposed recently, the graph space optimal transport distance, presents appealing properties for field data applications. This study compares it with the least-squares distance in an analysis performed on the three-dimensional ocean bottom cable data from the Valhall field. This data has already been at the heart of several full-waveform inversion studies, making it a good candidate to evaluate the properties of this new misfit function. We first perform this comparison starting the inversion from the reflection traveltime tomography model used in previous studies. We then perform a second comparison from a crude, linearly varying in-depth one-dimensional velocity model. Starting from this model, least-squares-based full-waveform inversion fails to provide a meaningful estimate of the pressure-wave velocity model due to cycle skipping. We illustrate how the graph-space optimal transport-based full-waveform inversion mitigates this issue. A meaningful estimate of the pressure-wave velocity model is obtained in the zone sampled by both diving and reflected waves, down to almost two kilometers depth. To our knowledge, this is the first application of a graph space optimal transport-based full-waveform inversion to three-dimensional field data.

3.1 Introduction

Full waveform inversion (FWI) is a seismic imaging method aiming at reconstructing high-resolution models (up to half the shortest wavelength) of the mechanical properties of the subsurface (Devaney, 1984; Pratt and Shipp, 1999; Plessix and Perkins, 2010; Raknes et al., 2015; Górszczyk et al., 2017). The method is an iterative process based on minimizing a misfit function between observed and calculated data over a space of model parameters describing the subsurface. The improvement of resolution that FWI provides over standard tomography methods makes it possible to significantly improve depth-migration images or directly produce interpretable images of the subsurface physical properties. This method is used at multiple scales; from global and regional scales (Fichtner et al., 2010; Tape et al., 2010; Bozdağ et al., 2016) to seismic exploration targets for oil & gas industry (Plessix and Perkins, 2010; Stopin et al., 2014; Operto et al., 2015) and even near-surface scale (Bretaudié et al., 2013; Groos et al., 2014; Schäfer et al., 2013; Irnaka et al., 2019). A thorough review of FWI and its applications is given in Virieux et al. (2017).

As powerful as this method is, it suffers from a significant shortcoming in its classical formulation: the non-convexity with respect to time-shifts of the least-squares (L^2) misfit function used to calculate the distance between observed and synthetic data. This non-convexity of the misfit function is an issue as the iterative process used in FWI is based on local-optimization algorithms. This leads to the so-called cycle-skipping issue. This limitation of FWI in its classical formulation has been documented since FWI has been introduced (Gauthier et al., 1986), and it has been of great interest to overcome it.

A large amount of studies has been published, proposing different approaches. One of

the historical approaches to overcome this limitation in practical cases is to rely on a data hierarchy workflow (Bunks et al., 1995; Pratt, 1999; Shipp and Singh, 2002; Wang and Rao, 2009; Brossier et al., 2009). This approach consists in interpreting first the lowest frequency available, generally 2 – 4 Hz for seismic exploration targets, then progressively introducing higher frequency data following a multi-scale approach (Sirgue and Pratt, 2004). A second level of data-hierarchy can be defined by modifying the temporal and/or offset selection of the data used during inversion. The idea is to reduce the number of propagated wavelengths that are interpreted simultaneously. Current industrial applications generally rely on these two levels of data hierarchy, combined with a robust starting model, obtained, for instance, through reflection traveltime tomography or stereotomography (Lambaré, 2008).

Nonetheless, the conditions to apply this workflow are not always satisfied. For instance, low-frequency data around 2 – 4 Hz are not always available or of sufficient quality. Moreover, obtaining low frequency increases the cost of the acquisition and can also compromise the quality of the high frequency needed to obtain high-resolution model reconstructions. Accurate initial model building can also be a time-consuming and challenging task requiring strong human expertise as it generally requires accurate traveltime and/or reflected event picking. Besides, prior information coming from geology and well logs are often needed. These constraints make FWI less robust and reduce its potential range in terms of applications.

Several methods have been introduced to improve robustness to cycle-skipping. The first group can be named as “extension strategies” and relies on introducing supplementary degrees of freedom to the FWI problem (Symes, 2008, 2015; Huang et al., 2018a; van Leeuwen and Herrmann, 2013; Wang et al., 2016; Aghamiry et al., 2020), which can be used to artificially match the data at early iterations of the FWI process, avoiding cycle-skipping.

The second group consists in reformulating the FWI problem using an alternative measure of the distance between the observed and calculated data, namely a different misfit function. Numerous approaches have been proposed, such as cross-correlation (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010) and deconvolution based misfit function (Luo and Sava, 2011; Warner and Guasch, 2016), or by modifying the signal itself, making the L^2 norm between this new observable more convex with, for instance, instantaneous envelope (Fichtner et al., 2008; Bozdağ et al., 2011). We want to keep in mind that replacing the L^2 norm is not an easy task, as, despite its simplicity, the L^2 misfit presents excellent and interesting properties. First, it is robust to Gaussian noise. Second, it presents an excellent resolution power, translating into high-resolution reconstruction that FWI is well known for. Third, it is straightforward to implement and the computational cost of the misfit function evaluation is negligible compared to most of the proposed alternative misfit functions. These advantages have made the L^2 misfit the “state of the art” for FWI at exploration scales and could explain why L^2 is still widely used even if many alternative misfit functions have been proposed to mitigate the cycle-skipping issue.

Indeed, there is a discrepancy between the many propositions for alternative misfit functions compared to the number of actual field data applications. We think that this discrepancy could be explained by the - often not deeply discussed - intrinsic limitations of these alternative formulations. For instance, cross-correlation-based misfit functions might have difficulties handling complex data, including missing arrivals or multiples. Deconvolution-based strategies require a penalization/weighting function, which can be difficult to set. Such settings are often case-dependent, making FWI less of an automated process. Instantaneous envelope intrinsically modifies the signal shape and discards information coming from the phase (which

is essential to interpret the polarity of reflected events in the data correctly).

A new class of misfit functions based on optimal transport (OT) has been introduced recently. The motivation is to benefit from the convexity of the optimal transport distance with respect to translation and dilation, which provides a misfit function convex with respect to time and amplitude shifts, a good proxy towards convexity with respect to velocities perturbations (Engquist and Froese, 2014; Métivier et al., 2018). Another important motivation to use OT as a misfit function is the ability to take into account the coherency of the seismic signal in an adequate space, be it a common shot or receivers gather. However, OT can only be applied to positive quantities and cannot be directly applied to seismic traces. To circumvent this difficulty, three main strategies have been developed.

The first one proposes to bring back the problem to the comparison of positive quantities by modifying the signal before solving the OT problem (Engquist and Froese, 2014; Qiu et al., 2017; Yang et al., 2018b; Yang and Engquist, 2018). A nonlinear transform is applied to the data in a trace-by-trace framework to transform each of them as probability measures. However, modifying the signal and altering the polarity information might be detrimental to a stable and satisfactory reconstruction of the subsurface mechanical properties.

The second one relies on the dual formulation of a specific instance of optimal transport distance, namely the 1-Wasserstein distance (Métivier et al., 2016a,b,c). This formulation can be naturally extended to the comparison of signed data. It benefits from its ability to be applied directly to 2D and 3D data, taking into account the coherency of the seismogram in the receiver and/or sources direction. However, even if the attraction valley to the global minimum is enlarged compared with the least-squares approach, the application of the 1-Wasserstein distance to signed data loses the convexity with respect to time-shift, which was the original motivation to use OT in the framework of FWI (see Métivier et al., 2018 for a review on different OT formulations). This strategy has been successfully applied to several field datasets (Poncet et al., 2018; Messud and Sedova, 2019; Sedova et al., 2019; Carotti et al., 2020; Hermant et al., 2020).

Finally, the third one considers each discrete seismic trace as point clouds and computes the optimal transport distance between point clouds associated with synthetic and observed traces. This method is called the graph space optimal transport (GSOT) and presents the main characteristic of preserving the convexity with respect to time shifts (Métivier et al., 2018, 2019). GSOT has already been successfully applied to 3D synthetic and field data (He et al., 2019a; Pladys et al., 2019; Li et al., 2019; Górszczyk et al., 2019) and is a promising candidate to tackle cycle-skipping on field datasets.

Therefore, this study focuses on applying the GSOT strategy against the classical L^2 misfit on a 3D OBC dataset from the North Sea, the Valhall field data. This dataset has been one of the first used to make proof of concept of the resolution power that FWI can bring on field data as shown in Sirgue et al. (2010). Since then, this dataset has been used several times for FWI application (Prioux et al., 2011; Gholami et al., 2013; Prioux et al., 2013; Operto and Miniussi, 2018; Kamath et al., 2020). This dataset can be seen as a “calibrated reference” for testing FWI formulations, such as frequency-domain FWI in Operto et al. (2015) or in time-domain with attenuation in Kamath et al. (2020). Here, we are using this dataset to compare the GSOT misfit function to the conventional L^2 norm through a time-domain 3D visco-acoustic VTI FWI.

To make this comparison, we first consider a canonical case where the initial model is the

same as the one used in the aforementioned studies, which is derived from reflection tomography. This initial model ensures FWI converges toward a plausible estimation of the subsurface using conventional L^2 misfit function.

Then, we introduce a very “crude” initial model in which calculated data are shifted by more than one cycle compared to observed data, which is the typical case scenario for cycle-skipping. We show how L^2 -based FWI fails from this “crude” starting model, whereas GSOT-based FWI manages to correctly interpret data to provide a plausible reconstruction of the subsurface (down to almost 2 km depth, which corresponds to the zone of the data sampled by both diving and reflected waves). This constitutes one of the first applications of the graph-space optimal transport misfit function to 3D field data at the exploration scale.

The paper is organized as follows. In the next section, we present the modeling and inversion algorithm used for our FWI application. Then we detail the global methodology of GSOT for FWI. This is followed by a presentation of the Valhall field application, from the geological situation to the initial model and dataset presentation. We detail our FWI workflow and analyze the results in two cases: from the reflection traveltime tomography initial model and then from a 1D initial model. Results are then discussed, followed by conclusion and perspectives, which are given in a final Section.

3.2 Methodology

3.2.1 Modeling

This study is performed in the frame of 3D time-domain FWI. We rely on the anisotropic visco-acoustic time-domain modeling and inversion algorithm developed by Yang et al. (2018a), based on the following partial differential equations:

$$\begin{cases} \rho \partial_t v_x &= \partial_x g \\ \rho \partial_t v_y &= \partial_y g \\ \rho \partial_t v_z &= \partial_z q \\ \partial_t g &= c_{11}(\partial_x v_x + \partial_y v_y) + c_{13} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{11} \xi_\ell^g + c_{13} \xi_\ell^q] \\ \partial_t q &= c_{13}(\partial_x v_x + \partial_y v_y) + c_{33} \partial_z v_z - \sum_{\ell=1}^L Y_\ell [c_{13} \xi_\ell^g + c_{33} \xi_\ell^q] \\ \partial_t \xi_\ell^g &= -\omega_\ell \xi_\ell^g + \omega_\ell (\partial_x v_x + \partial_y v_y), \quad \ell = 1, 2, \dots, L \\ \partial_t \xi_\ell^q &= -\omega_\ell \xi_\ell^q + \omega_\ell \partial_z v_z, \quad \ell = 1, 2, \dots, L. \end{cases} \quad (3.1)$$

In system 3.1, c_{11} , c_{13} , and c_{33} are the stiffness tensor coefficients, ρ is the density, v_x , v_y , v_z are the horizontal and vertical displacement velocities respectively, while g and q are related to the normal stress components σ_{xx} , σ_{yy} and σ_{zz} through

$$\begin{aligned} g &= \sigma_{xx} = \sigma_{yy} \\ q &= \sigma_{zz}. \end{aligned} \quad (3.2)$$

This simplification is due to the VTI approximation. Similarly, the memory variables ξ_ℓ^g and ξ_ℓ^q are related to the memory variables ξ^{xx} , ξ^{yy} and ξ^{zz} associated with the normal stress components through

$$\begin{aligned} \xi_\ell^g &= \xi^{xx} + \xi^{yy} \\ \xi_\ell^q &= \xi^{zz}. \end{aligned} \quad (3.3)$$

These memory variables are used to model the viscosity of the medium following the generalized Maxwell body theory. Each represents one relaxation mechanism. We use three relaxation mechanisms to approximate a constant attenuation within the considered frequency band ($L=3$). The variables Y_ℓ are therefore calibrated depending on the target quality factor representing the attenuation in the considered media. This calibration is done through the solution of a least-squares problem. The details of this calibration can be found in Yang et al. (2016a) for instance.

In the VTI approximation, the stiffness tensor coefficients are related to the vertical P-wave velocity, the density, and the Thomsen anisotropy parameters ϵ and δ through

$$\begin{aligned} c_{11} &= \rho V_P^2 (1 + 2\epsilon) \\ c_{33} &= \rho V_P^2 \\ c_{13} &= \rho V_P^2 (1 + 2\delta). \end{aligned} \quad (3.4)$$

The discretization of this system of partial differential equations is performed using a fourth-order in space and second-order in time staggered grid finite-difference method (Virieux, 1986; Levander, 1988).

A flat free surface condition is applied on top of the model to represent the water/air interface. Sponge layers (Cerjan et al., 1985) are applied on the other faces of the model to mimic a medium of infinite extensions in these directions. This numerical method is preferred over the use of perfectly matched layers (PML, Bérenger, 1994) both because of the anisotropy, which might cause PML instabilities, as well as the presence of attenuation and relaxation mechanisms, which complexifies the PML implementation.

Finally, windowed sinc interpolation is used to simulate source and receivers off-grid points accurately (Hicks, 2002).

3.2.2 Inversion

3.2.2.1 General formulation

FWI is an iterative process which relies on the minimization of a misfit function. Classically, it is the L^2 misfit function defined as follows

$$f_{L^2}[m] = \sum_s \sum_r h(d_{cal}[m](x_r, t; x_s), d_{obs}(x_r, t; x_s)), \quad (3.5)$$

where

$$h(d_1(t), d_2(t)) = \frac{1}{2} \int_{t=0}^T |d_1(t) - d_2(t)|^2 dt. \quad (3.6)$$

The observed and synthetic trace calculated in model m , associated with source x_s and receiver x_r , are denoted by $d_{obs}(x_r, t; x_s)$ and $d_{cal}[m](x_r, t; x_s)$ respectively.

To solve this minimization problem, we rely on a local optimization scheme. We use a pre-conditioned quasi-Newton l-BFGS algorithm (Nocedal, 1980), implemented in the SEISCOPE optimization toolbox (Métivier and Brossier, 2016).

We compute the gradient following the adjoint state method (Plessix, 2006). This method makes it possible to easily replace the L^2 norm as a change of the misfit function only translates

to a modification of its associated adjoint-source to obtain the gradient. See Yang et al. (2018a) for a review. Since the method relies on reverse time propagation of the wavefield, which is numerically unstable with attenuation, our code relies on the checkpoint-assisted reverse forward simulation (CARFS) strategy proposed by Yang et al. (2016b) to provide a stable and yet efficient implementation for large scale problems with attenuation.

The preconditioner we use is the wavefield preconditioner presented in Kamath et al. (2020). An anisotropic non-stationary Gaussian smoothing is also applied to the gradient. Inversions are performed for P-wave velocities (V_P) only.

3.2.2.2 Optimal transport for FWI

Optimal transport (OT) distances are derived from the OT theory introduced by the French mathematician Gaspard Monge more than two centuries ago (Monge, 1781). More precisely, they rely on the OT relaxation proposed by Kantorovich (1942). The distances, also called Wasserstein distances, have an intrinsic property of particular interest for the definition of inverse problems: they are convex with respect to translation and dilation of the compared quantities. This convexity with respect to translation has been, in particular, the motivation to introduce it in the framework of FWI to obtain a distance measurement convex with respect to time-shifts (Engquist and Froese, 2014). However, OT distances are defined for comparing probability distributions, which are by definition positive and normalized. So this new distance cannot be directly applied to seismic data, which is oscillatory (a generalization of OT to signed distribution is still an open question from a mathematical point of view, see Ambrosio et al. (2011) and Mainini (2012) for instance).

This study focuses on a recent proposition made to apply OT to seismic data: the graph space optimal transport (GSOT) strategy, proposed in Métivier et al. (2019).

This formulation of OT distance should preserve the signal unmodified while also preserving the convexity to shifted patterns. It relies on the idea of comparing the discrete graph of the data rather than the data itself: each 1D trace in time becomes a point cloud of Dirac delta functions (of amplitude 1) in a 2D space made of the time dimension, and a new amplitude dimension. This transformation does not affect the signal shape but makes it possible to deal with positive mass (the Dirac delta functions). Thus the OT distance can be applied directly.

The corresponding misfit function is formulated as

$$f_{GSOT}[m] = \sum_s \sum_r h(d_{cal}[m](x_r, t; x_s), d_{obs}(x_r, t; x_s)), \quad (3.7)$$

where this time

$$h(d_1, d_2) = \min_{\sigma \in S(N_t)} \sum_{i=1}^{N_t} c_{i\sigma(i)}(d_1, d_2). \quad (3.8)$$

$S(N_t)$ denotes the ensemble of permutations of $\{1, \dots, N_t\}$, and c_{ij} the L^2 distance between the discrete points of the graph $(t_i, d_1(t_i))$ and $(t_j, d_2(t_j))$:

$$c_{ij}(d_1, d_2) = |t_i - t_j|^2 + \eta^2 |d_1(t_i) - d_2(t_j)|^2. \quad (3.9)$$

The function h corresponds to the 2-Wasserstein distance between the discrete graph of the “calculated” trace $d_1(t)$ and the “observed” trace $d_2(t)$.

The scaling parameter η controls the convexity of the misfit function f_{GSOT} with respect to time shifts. In practice, we define it as

$$\eta = \frac{\tau}{A}, \quad (3.10)$$

where τ is a user-defined parameter corresponding to the maximum expected time shift between observed and calculated data in the initial model, and A is the maximum peak amplitude difference between observed and calculated data. This ensures the convexity of the GSOT distance for time up to approximately τ .

The final cost function we use for the purpose of FWI application with N_s shots containing N_r receivers is defined as:

$$\min_m f_{GSOT}[m] = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} w^{s,r} h(d_{cal}^{s,r}[m], d_{obs}^{s,r}), \quad (3.11)$$

where $w^{s,r}$ is a trace-by-trace weighting factor, typically used to restore the AVO trend in the data. This trend is removed from the trace-by-trace GSOT approach, as the amplitude of each trace is treated separately through the normalization factor η .

The adjoint source of the misfit function $f_{GSOT}[m]$ is computed from $\frac{\partial h}{\partial_{cal}}$ using the adjoint-state strategy (Plessix, 2006). The following result is proved in Métivier et al. (2019). Denoting σ^* the minimizer in eq. 3.7, we have

$$\frac{\partial h}{\partial_{cal}} = 2 \left(d_{cal} - d_{obs}^{\sigma^*} \right), \quad (3.12)$$

where

$$d_{obs}^{\sigma^*}(t_i) = d_{obs}(t_{\sigma^*(i)}). \quad (3.13)$$

The GSOT approach can thus be viewed as a generalization of the L^2 distance: The adjoint source is equal to the difference between calculated and observed data at time samples connected by the optimal assignment σ^* . The solution of the problem eq. 3.7 provides the information to compute both the misfit function and the adjoint source.

To solve eq. 3.7 efficiently, we use the auction algorithm (Bertsekas and Castanon, 1989), dedicated to the solution of linear sum assignment problems such as eq. 3.7. Despite a relatively high computational complexity in $O(N_t^3)$, it is quite efficient for small instances of such problems. Resampling the data close to the Nyquist frequency (at the exploration scale, under acoustic approximation, traces are often around few hundreds of time steps after resampling) yields such small-scale problems making the GSOT feasible for realistic scale FWI applications, as seen in the application presented here.

3.3 Field data presentation and full-waveform inversion workflow

3.3.1 Geological situation, dataset and initial models

3.3.1.1 Geological situation

The Valhall field is located in the southern part of the Norwegian sector in the North Sea, approximately 300 km southwest of Stavanger (Fig. 3.1). It is a shallow environment with a

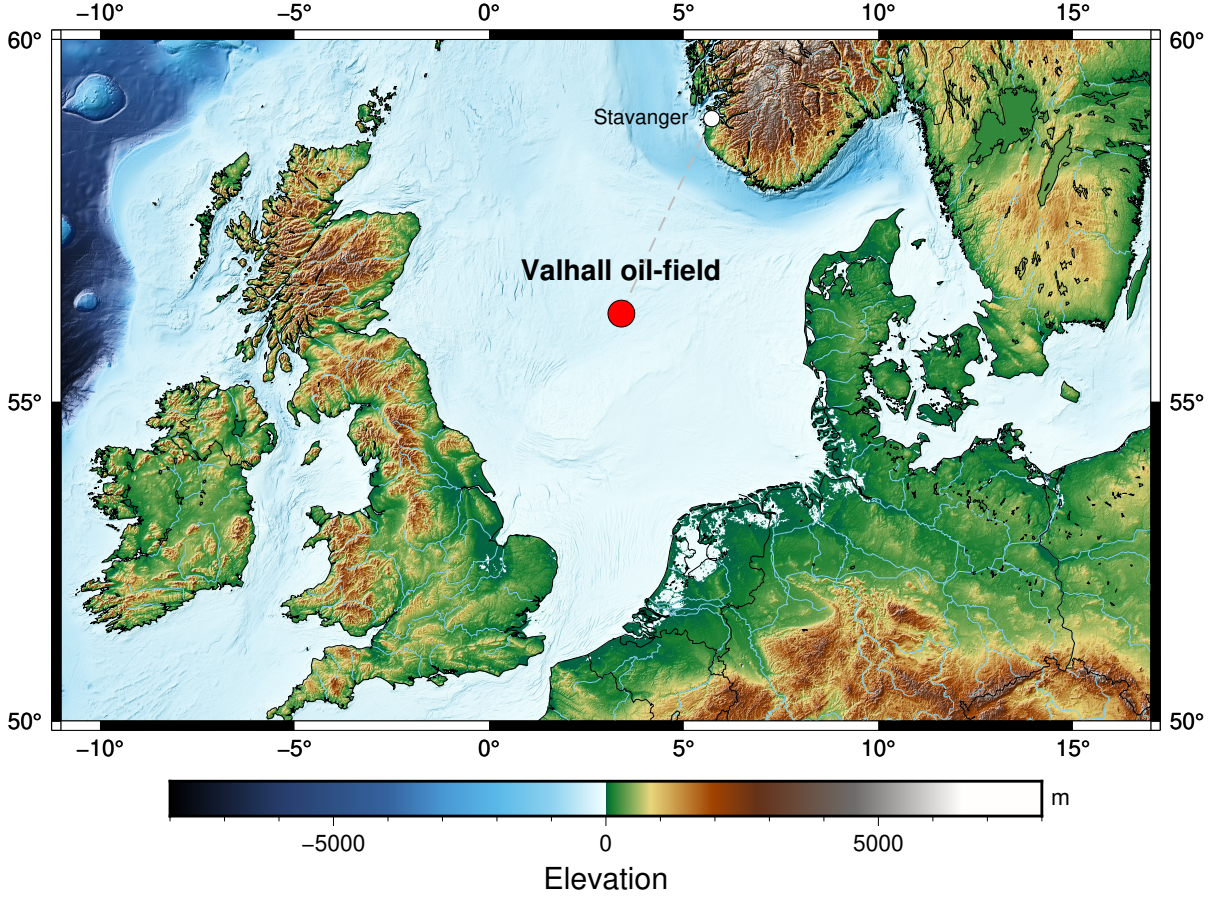


Figure 3.1: Location of the Valhall field on the North Sea (from Thurin, 2020)

nearly constant water depth of 70 m. Valhall reservoir lies along the Lindesnes Ridge, which trends NNW (Munns, 1985; Leonard and Munns, 1987). The field has been discovered in 1975 and is used since then for oil production. It is characterized as an anticlinal in chalk in the Upper Cretaceous Hod and Tor formations, which form the reservoir at a depth of approximately 2400 m. Trapped gas in Tertiary shale is present above the reservoir (Sirgue et al., 2010; Prioux et al., 2011, 2013; Operto et al., 2015). The Tertiary overburden is relatively simple and free of complex structure (Hall et al., 2002). Chalk compaction resulting from pressure depletion and water weakening has led to seabed subsidence (Field: VALHALL - Norwegianpetroleum.no <https://www.norskipetroleum.no/en/facts/field/valhall>).

3.3.1.2 Dataset presentation

The seismic data are 4-components acquired by ocean-bottom cables (OBC), with wide aperture/azimuth acquisition. The covered zone is a surface of 145 km². Twelve receiver cables are deployed on the seabed, containing 2048 receivers with an inline spacing of 50 m and a cable spacing of 300 m. A total of 50824 shots are available, located 5 m below the sea surface. The layout of this 3D acquisition is presented in Figure 3.2. In this study, we use only the pressure component of one acquisition performed in 2011 as part of the Valhall Life of Field Seismic (LoFS) project (Barkved et al., 2003).

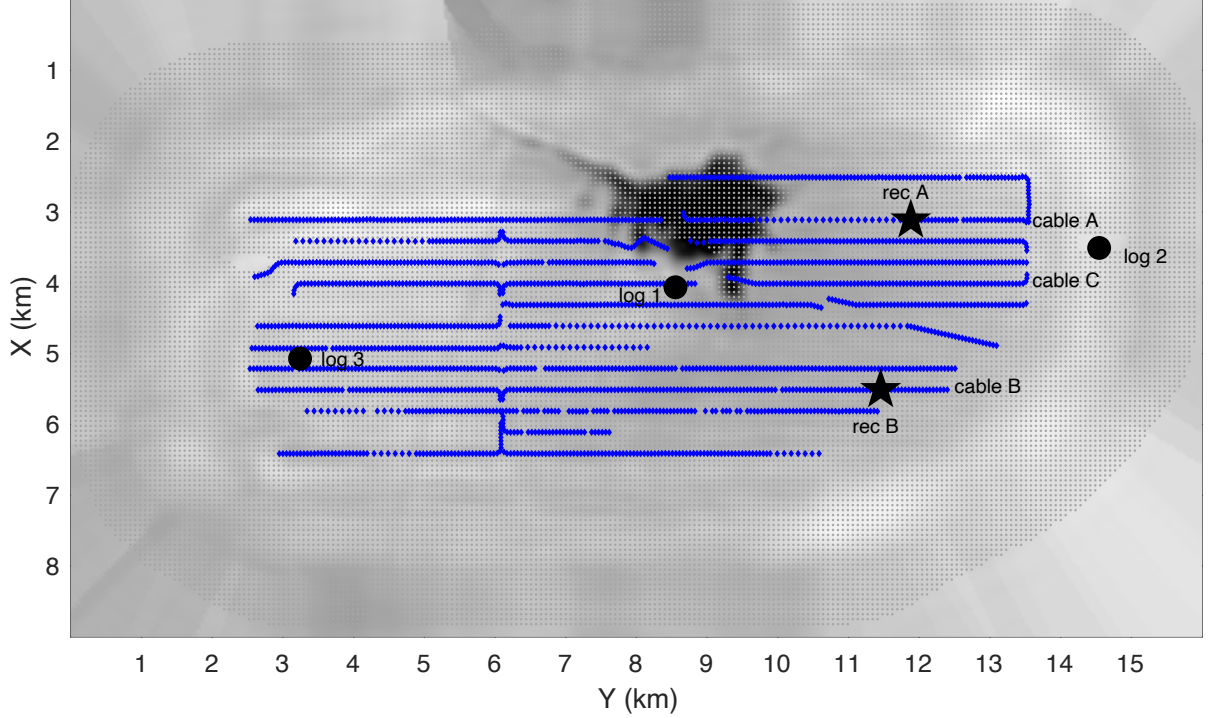


Figure 3.2: Layout of the Valhall acquisition overlapped on an horizontal V_P slice at 1 km obtained by GSOT-based FWI (from this study). Location of sources (gray dots) and receivers (blue diamonds). Two receivers positions (A and B) are located with black stars. Cables A ($x = 2950$ m), B ($x = 5530$ m) and C ($x = 3080$ m) are identified. Black dots denote the position of the three V_P well sonic-logs.

From the raw data provided in SEG Y format, only a simple pre-processing is applied. As our FWI code relies on source parallelization, source-receiver reciprocity is applied to process the hydrophone as explosive sources and the shots as hydrophones, hence sensibly reducing the impact on computer resources. The data is then de-spiked before a quality control over the complete dataset: the energy (RMS) of each gather is calculated to manually remove faulty gathers (the one with a large variation of RMS amplitude). The last step is to create frequency bands for the inversion using a minimum-phase band-pass filter in two distinct bands: 2.5–5 Hz and 2.5 – 7 Hz (referred to as band 1 and band 2 in the following). This goes in hand with time-decimation; from $\Delta t = 4$ ms to $\Delta t = 8$ ms on the first band, and from $\Delta t = 4$ ms to $\Delta t = 5$ ms on the second band (Fig. 3.3).

3.3.1.3 Initial models

An initial V_P model was made accessible to us thanks to AkerBP. It has been obtained through reflection traveltome tomography. It is referred to as TOMO initial and is presented in Figure 3.4. This initial model has been used in several publications using this Valhall dataset and has proven its capacity to give satisfactory results with L^2 -based FWI as it predicts the arrival within half a period of the considered starting frequencies (2.5 – 5 Hz), avoiding cycle-skipping issue (Prioux et al., 2011; Operto et al., 2015; Operto and Miniussi, 2018). The associated density model is derived from V_P TOMO using Gardner’s law (Gardner et al., 1974), defined as

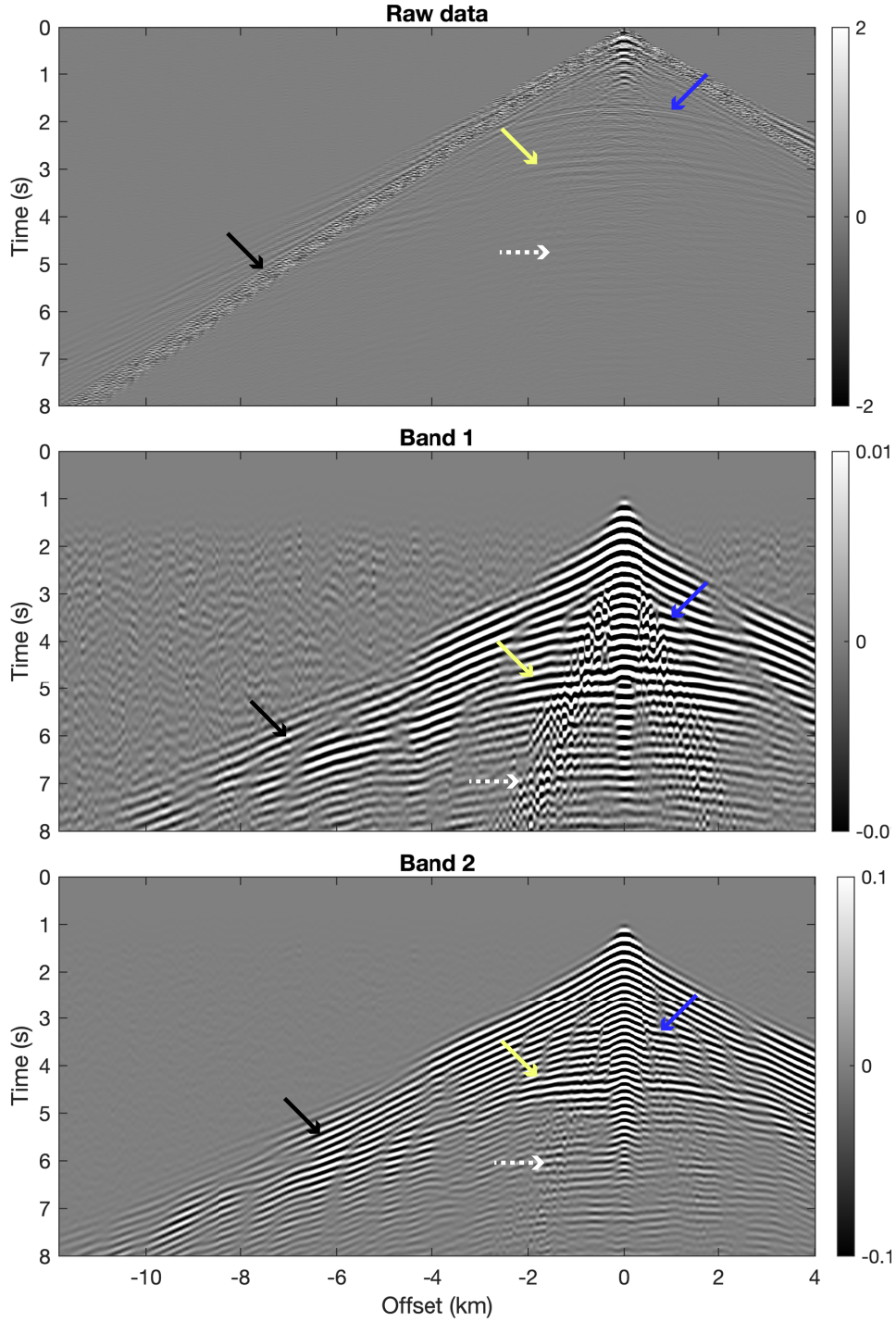


Figure 3.3: 2D common-receiver gathers extracted for receiver A along cable A for: raw data (top), band 1 data 2.5 – 5 Hz (middle) and band 2 data 2.5 – 7 Hz (bottom). White dashed arrows point on the Schölte waves which are muted for the inversion. Blue and yellow arrows respectively point on the reflexion from the low velocity zone and the reflexion on the top of the reservoir. Black arrows point on the diving waves.

$\rho = 309.6 * V_P^{0.25}$. This relation is a fair average for brine-saturated rock (excluding evaporites), which is coherent with the expected geology of the Valhall field. AkerBP also provided us the anisotropy model ϵ and δ , and while their structure is not complicated, they are of significant influence in the modeling. The anisotropic parameter η define by Alkhalifah (1998) as

$$\eta = \frac{\epsilon - \delta}{1 - 2\delta} \quad (3.14)$$

is shown in Figure 3.6. We can see that maximum anisotropy reach values around 15% near the reservoir. Finally, to introduce attenuation, a simple two-layer Q_P model, with 1000 in the water column and 200 in the sediments, is used (Operto et al., 2015). While it is a simple model, it has been proven to be of great importance to explain the data. Using a more complex Q_P model (for example derived from V_P) only results in marginal improvements. Moreover, it introduces complexity and uncertainties that we prefer not to deal with (Kamath et al., 2020).

As the TOMO initial model is good enough to match the data within half a period of the 2.5 – 5 Hz frequency-band used to start FWI (Fig. 3.7), it does not represent any challenge regarding the cycle-skipping issue. This is why we introduce a new “crude” V_P starting model called 1D initial presented in Figure 3.5. It is a purely 1D vertical starting model, based on a linearly increasing profile with one main interface around 2400 m depth. This 1D starting model generates strong cycle-skipping, as made visible in Figure 3.8. As traditional L^2 -based FWI cannot tackle this cycle-skipping in the data, it should be a good candidate to benchmark the capability of GSOT FWI. Associated with this V_P 1D model, the density model is derived using the preceding Gardner’s law. The anisotropy model ϵ and δ , as the attenuation model Q_P are kept similar to those used in the TOMO setup.

3.3.2 Full-waveform inversion workflow

To obtain the final FWI reconstructed V_P model, we rely on several elements that compose our complete workflow. Each part of it plays a critical role in obtaining the best results possible, from wavelet estimation, data selection, to post-processing the FWI model. The generic workflow that we use in this article is presented in Figure 3.9. We detail the different parts composing the workflow in the following sections.

3.3.2.1 Source wavelet estimation

The first step before running FWI is to obtain a proper source wavelet that is used during the inversion to generate synthetic data. This is a crucial step as an incorrect wavelet estimation could induce artifacts into the reconstructed model.

The wavelet is estimated by solving a linear deconvolution problem in the frequency domain, following the methodology described in Pratt (1999). The wavelet inversion is performed on a single random subset of 240 shot-gathers. The sample of 240 wavelets (one per shot-gather) is then averaged to produce a single wavelet, assuming that all hydrophones have the same coupling response.

To minimize the cross-talk between the V_P model and the wavelet, we rely on a carefully designed data weighting strategy, focusing on short-offset only as presented in Figure 3.10. The data weighting consists in using 400 m of offset at full amplitude, then ramping down to zero

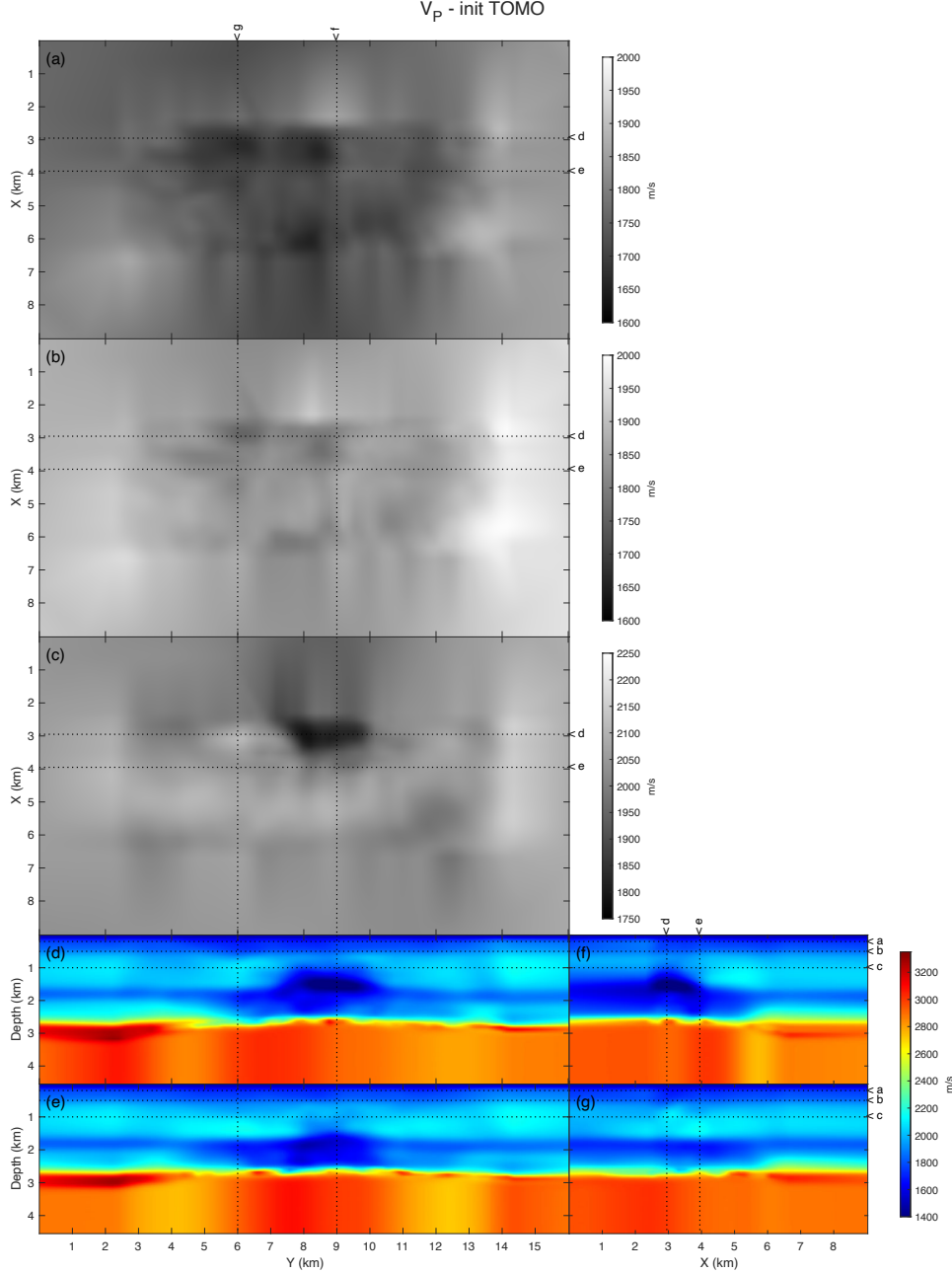


Figure 3.4: Slices of the initial model V_P TOMO . (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.

at 1200 m offset. A tail mute is applied to remove the Schölte waves. Before being used for inversion, the wavelet is then manually checked and tapered to ensure its causality. With this methodology, the wavelet is only estimated at the beginning of each frequency band and kept fixed during FWI steps. The wavelet for the first frequency band and its associated spectrum is presented in Figure 3.10. We can see that no oscillations are present after 2.7 s, and wavelets generated from TOMO or 1D initial models are similar, which validates that this careful data selection mitigates the potential leakage in the wavelet estimation, which could come from the

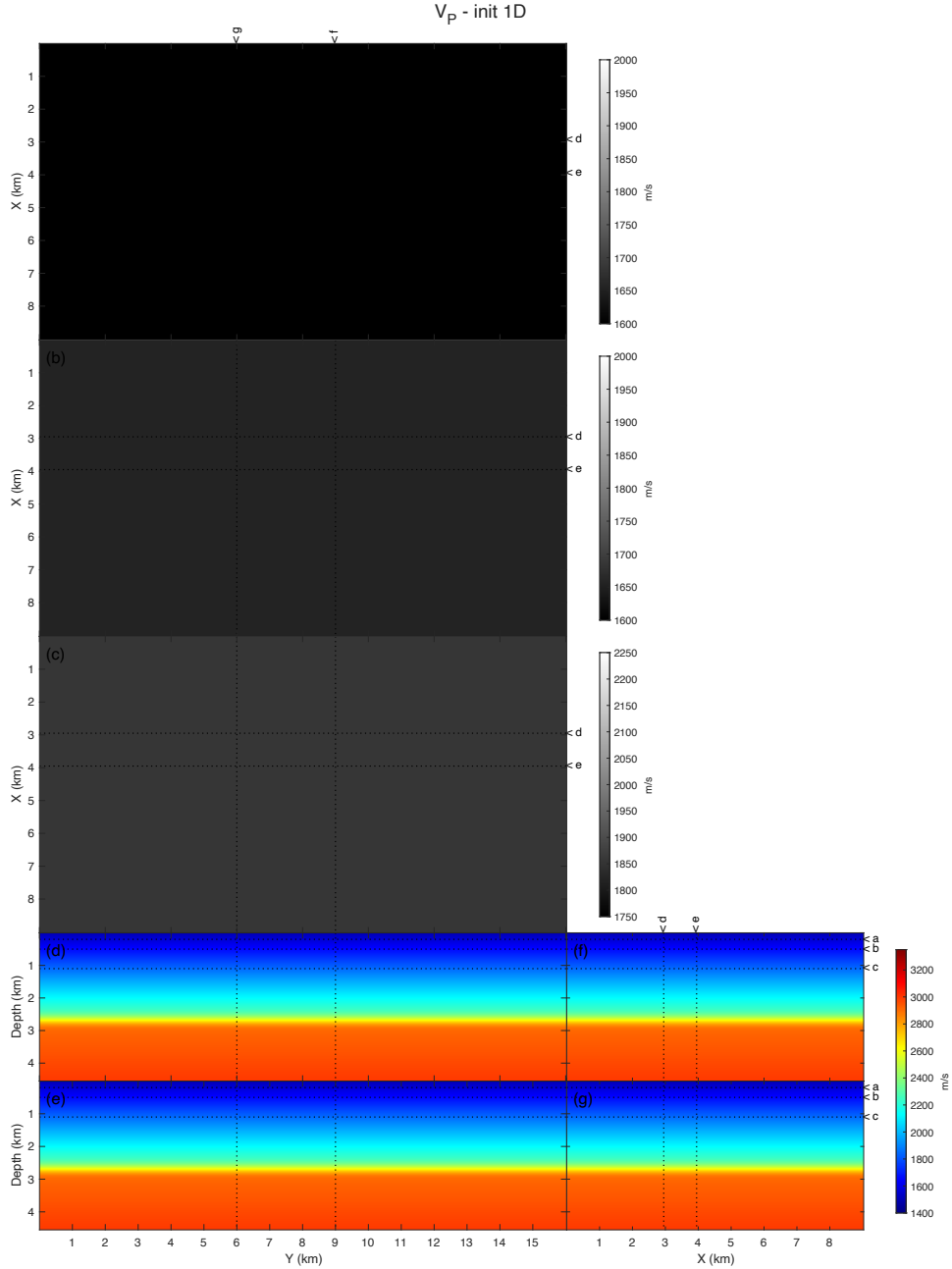


Figure 3.5: Same as Figure 3.4 for initial model V_P 1D .

P-wave velocity model. Finally, the wavelet spectrum is coherent with the data (2.5 to 5 Hz).

3.3.2.2 Random shot subsampling and frequency continuation

The FWI workflow relies on a frequency continuation approach. Inversion is performed first on the frequencies ranging between 2.5 – 5 Hz, then on a second band between 2.5 – 7 Hz. Respectively, the model grid spacing is set to 70 m and 50 m, ensuring at least five grid points for the smallest wavelength. These parameters are the same as in Operto et al. (2015) and

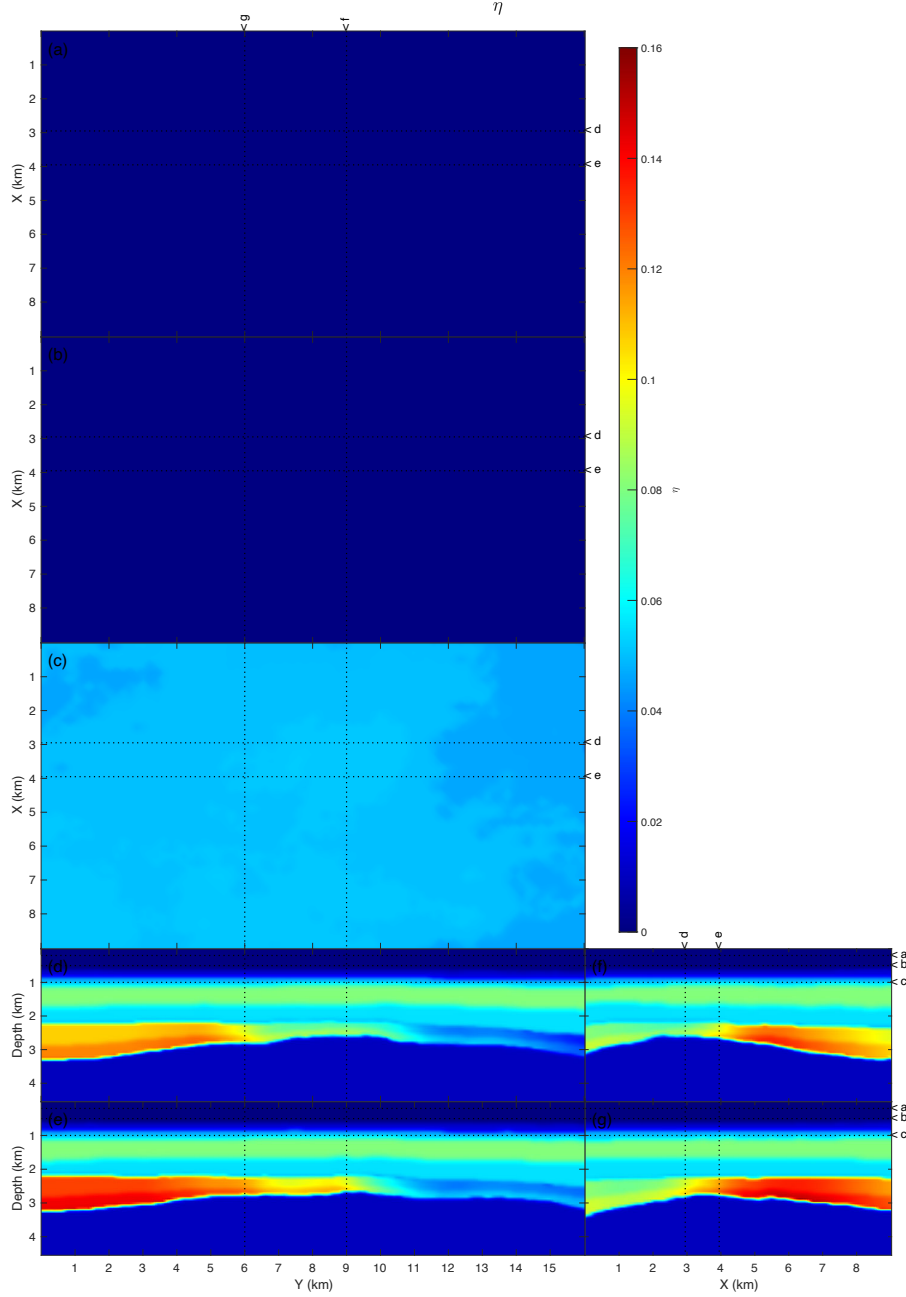


Figure 3.6: Same as Figure 3.4 for anisotropic parameter η .

Kamath et al. (2020).

As previously mentioned, the dataset contains 2048 shot-gathers. Since our FWI code relies on source-parallelization, the complete dataset could not be reasonably fit in a single inversion on the HPC facility we have access to. Therefore, we rely on a source-subsampling strategy, similar to the one described in Warner et al. (2013). This makes it possible to divide the dataset into batches of pseudo-randomly selected shot-gathers (120 in our case). The pseudo-random selection implies that the previously used shot-gather could not be selected in the next subsample until all of the available ones are used once. The subsample of shot-gathers

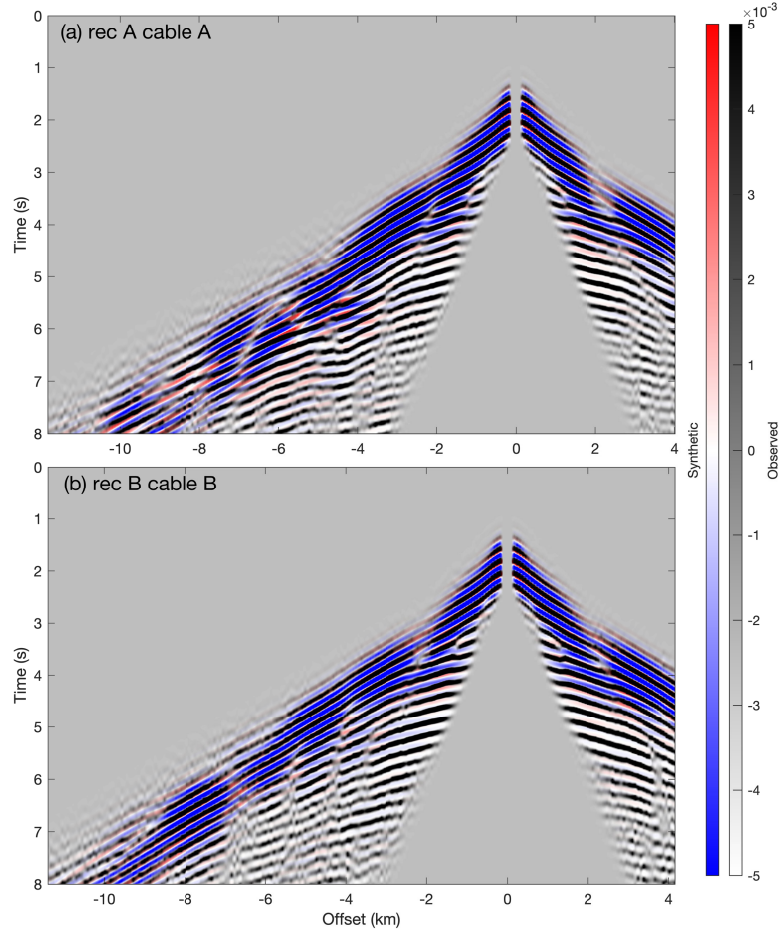


Figure 3.7: 2D common-receiver gathers at 5 Hz. Synthetic data generated into the initial model TOMO are displayed in a blue/white/red color scale, field data are overlapped in grayscale with transparency. The best result is achieved when black and blue are the only colors visible. Red and white are shown when data are not in phase. (a) receiver A along cable A (through the low velocity anomaly). (b) receiver B along cable B.

is changed every time the memory limit for l -BFGS is reached (3 iterations is the maximum memory of l -BFGS in this study). The source-subsampling strategy makes it possible to fit the FWI problem on relatively small HPC facilities while mitigating the acquisition foot-print in the reconstructed model.

3.3.2.3 Hierarchical data weighting strategy

The first data weighting strategy tried on the dataset for FWI is simple: only remove the Schölte waves and muted the trace near-zero offset in a radius of 350 m (see the approach in Kamath et al. 2020). While this direct and straightforward approach (using all the data directly) could be justified with TOMO initial models as it is not supposed to generate cycle-skipping, this approach could not be applied to tackle a crude initial model as the 1D one, even with the GSOT misfit function. Because of the strong cycle-skipping generated with the

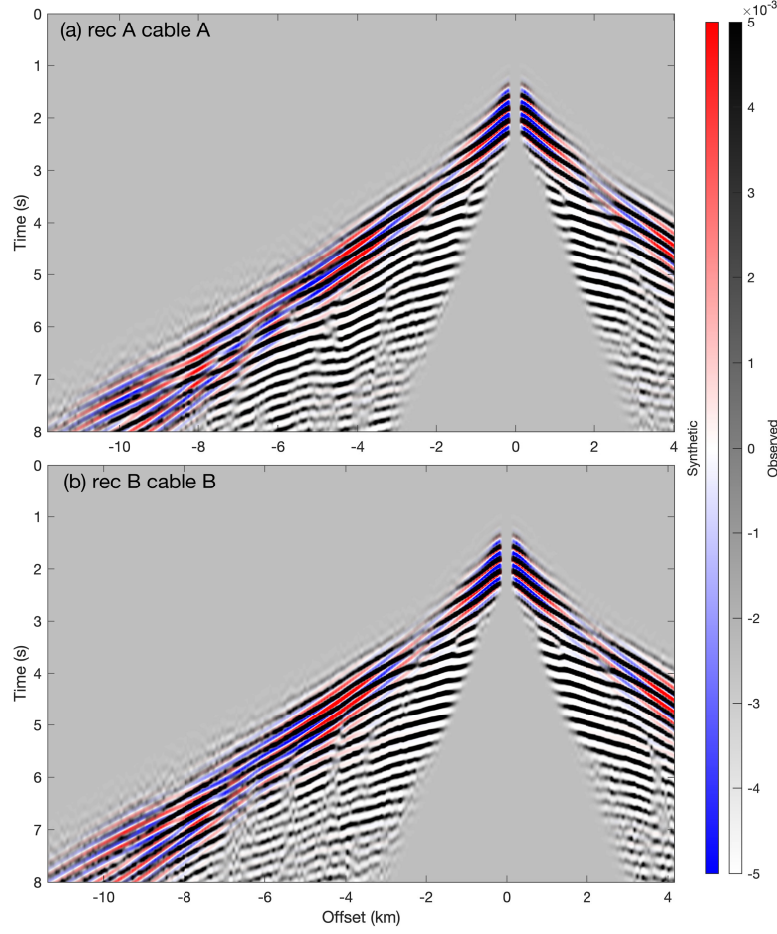


Figure 3.8: Same as Figure 3.7 into the 1D initial model.

1D initial model, we need to rely on a more careful data weighting strategy to maximize the capability of the GSOT. This leads to introduce a six-step data weighting strategy, presented in Figure 3.11. The three first steps only focus on diving waves with a strict time windowing while restricting to offset to the first 4 km (later referred to as **DW SO** for **Diving Wave - Short Offsets**), 8 km offset (later referred to as **DW MO** for **Diving Wave - Medium Offsets**), and full offset (later referred to as **DW FO** for **Diving Wave - Full Offsets**). Then, the three next steps release the time windowing progressively, starting with 8 km offset (later referred to as **RT1 MO** for **Release Time 1 - Medium Offsets**), then on full offset (later referred to as **RT1 FO** for **Release Time 1 - Full Offsets**); to finally finish with full offset and a complete release of time windowing (later referred to as **RT2 FO** for **Release Time 2 - Full Offsets**). Please also note that for each data weighting, the Schölte waves are additionally muted, as we cannot model them in the acoustic approximation.

3.3.2.4 Model post-processing

The last step of the FWI workflow is a post-processing applied to the updated V_P model at the end of each FWI step. It consists in cutting the model using a stencil (based on the shape of the acquisition) and extrapolating the value outside the stencil with a nearest-neighbors algorithm. This is performed for each depth slice. The stencil shaped is adapted for each depth slice to consider the maximum illumination the acquisition can provide at depth. This aims to remove the area on the edges of the model that are never updated during FWI due to the lack of illumination and remove artificial low-velocity zones created on the border of the well-illuminated zone.

To be consistent and perform a fair comparison between our two starting models TOMO and 1D, the complete workflow detailed before is applied to both starting models. By doing so, we validate that our workflow can tackle the dataset properly and provide satisfactory results. Then, changing only the misfit function from L^2 to GSOT, we recondut the complete inversion process to check how GSOT compares to L^2 in this controlled environment.

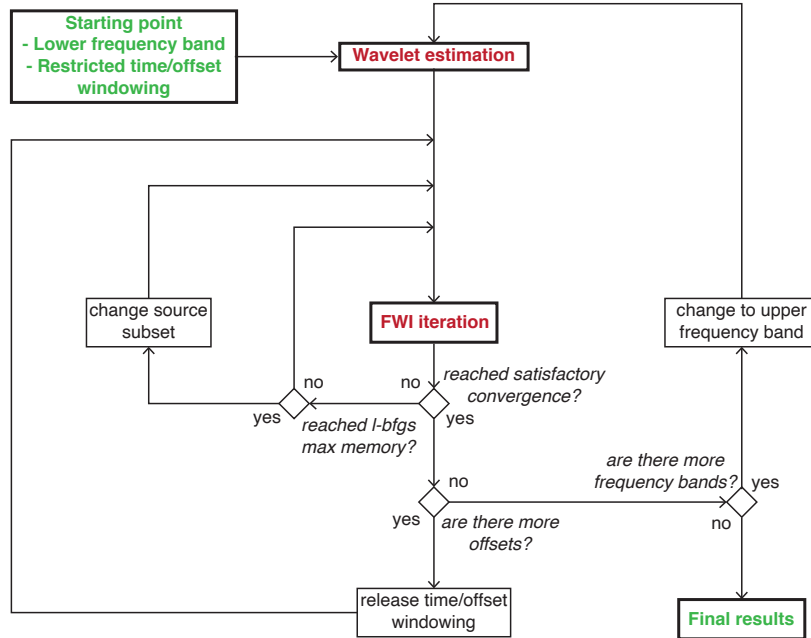


Figure 3.9: Complete FWI workflow used on the OBC Valhall dataset. At the core of the process lies the FWI iterations. Then several loops are nested one into another, from internal FWI iterations, source-subsampling, data selection, to finally the outer one of frequency continuation.

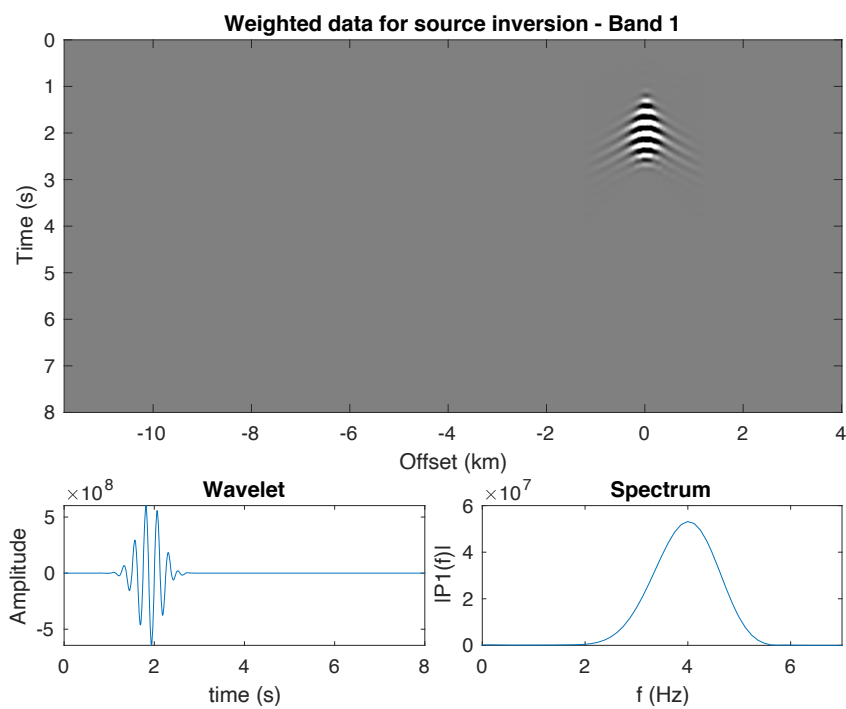


Figure 3.10: On top, weighted data for source inversion displayed on a 5 Hz 2D common-receiver gather (receiver A cable A). On the bottom, the estimated wavelet (left) and associated spectrum (right).

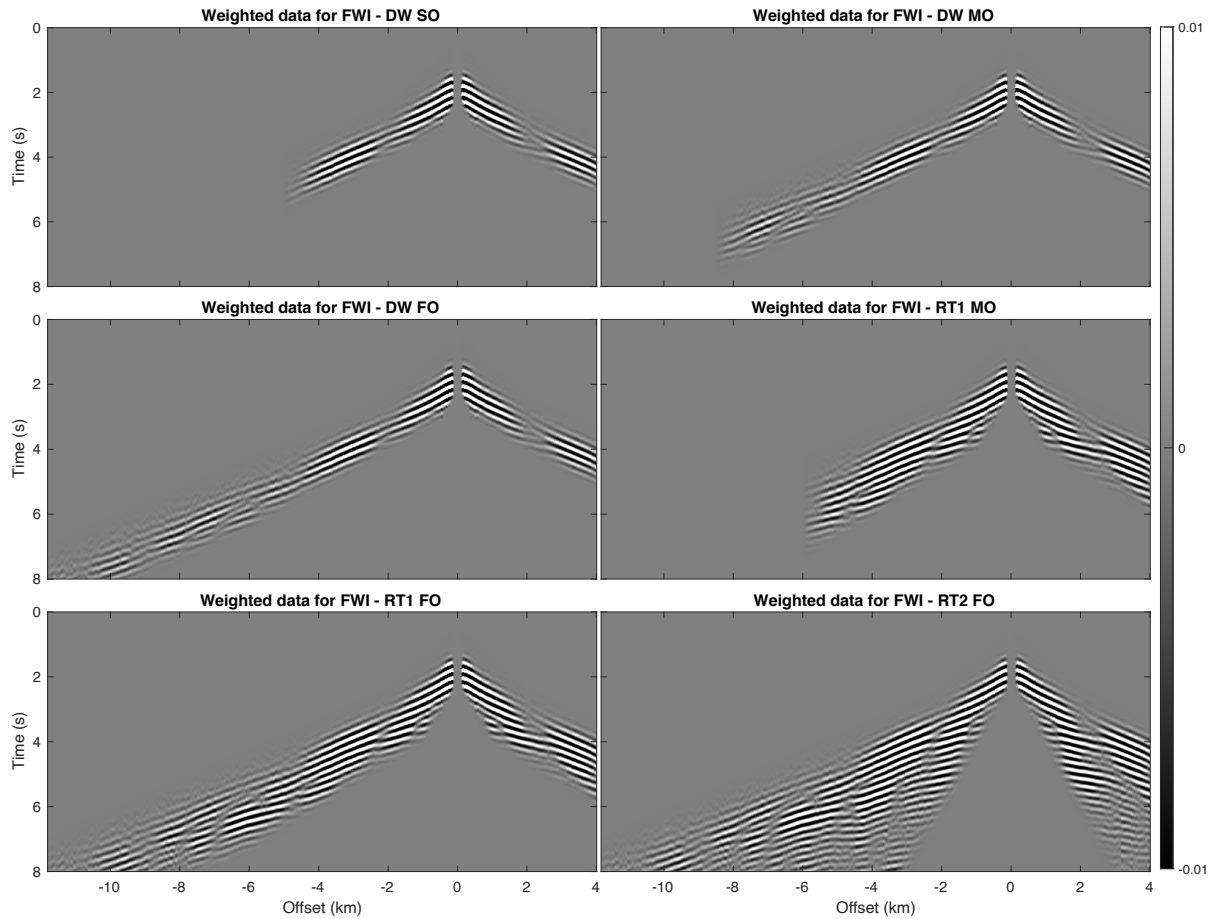


Figure 3.11: 2D common-receiver gathers extracted for receiver A along cable A with different data weighting applied on them. From top to bottom: first break & short offset (DW SO), first break & medium offset (DW MO), first break & full offset (DW FO), first time release & medium offset (RT1 MO), first time release & full offset (RT1 FO), second time release & full offset (RT2 FO).

3.4 Full-waveform inversion results

3.4.1 Starting from the reflection tomography model

The first FWI results presented in this article are based on the TOMO initial model. The final reconstructed V_P at 7 Hz is shown in Figure 3.12 using the L^2 misfit function, and in Figure 3.13 using the GSOT misfit function. Starting from the TOMO initial model, and as we do not expect cycle skipping, we keep the τ parameter of GSOT to 0.2 s. This value is low enough to always allow for fast convergence (similar to L^2) while preserving the potential improvement that GSOT can provide. Using a smaller value of τ would make results similar to L^2 while making τ larger would degrade the capacity of convergence of FWI due to a flatter attraction basin which is not needed in this case. The results are almost identical between both formulations on this setup. Some minor differences can, however, be observed. On the shallow depth slice (a) at 200 m, a small reduction of the acquisition imprint is observable using the GSOT misfit function. On the depth slice (c) at 1 km, a slight improvement of contrast between the low velocity anomaly and the sediment background is also visible. Finally, on the vertical slices (d-g), more lateral coherency in the geological structures is visible using the GSOT misfit function. The difference globally remains marginal, but this similarity between the L^2 and GSOT misfit functions results is satisfying in itself. Indeed, alternative misfit functions generally bring some drawbacks, such as loss of resolution power, which is not the case here. When performing QC on the V_P profile extracted from the reconstructed model versus the sonic log (Fig. 3.22), we can see that the GSOT result is almost perfectly following L^2 result, which is consistent with the observation made directly on the V_P model. A slight improvement of the well velocity fit at depth for GSOT misfit function can be observed for the well log 1 (which passes through the center of the target). Generally, FWI results improve the fit to the log compared to the initial model, showing that our FWI workflow provides robust and reliable results.

The data fit presented in Figure 3.15 shows that GSOT can improve over the L^2 FWI data fit (which is already very satisfying). On the *rec A cable A* 2D CRG (through the low velocity anomaly), we can observe some data fit improvement with GSOT, for example, at -6.5 km offset and 7 s. On the *rec B cable B* 2D CRG, in an area with relatively mild variations in velocities and well away from the “gas cloud”, we can observe that, globally, the data fit is almost the same for L^2 and GSOT, with all phases correctly explained. This is not surprising as this CRG focuses on the part of the model with sedimentary geology only, which the initial model better explains. However, we can still see the advantages of GSOT this time with more arrivals present for late time, as visible in the -7 to -4 km offset, from 5 to 7 s.

This “reference” result validates two essential aspects of this study. This first one is that our FWI workflow is adapted to the dataset and can provide satisfactory results when using L^2 -based FWI when a good enough initial model is used as the TOMO model. This point is of first importance as it validates that our FWI workflow is consistent with the literature on this dataset. The second conclusion is that the GSOT misfit function can tackle field data problems without significant issues. Modifying the misfit function and keeping all other parameters similar, GSOT can provide comparable results to those provided by the L^2 FWI. We can even see slight improvement both in terms of the P-wave velocity model estimation and data fit with the GSOT misfit function. These encouraging results make it possible to push the analysis forward: can GSOT help tackle the cycle-skipping issue and improve FWI

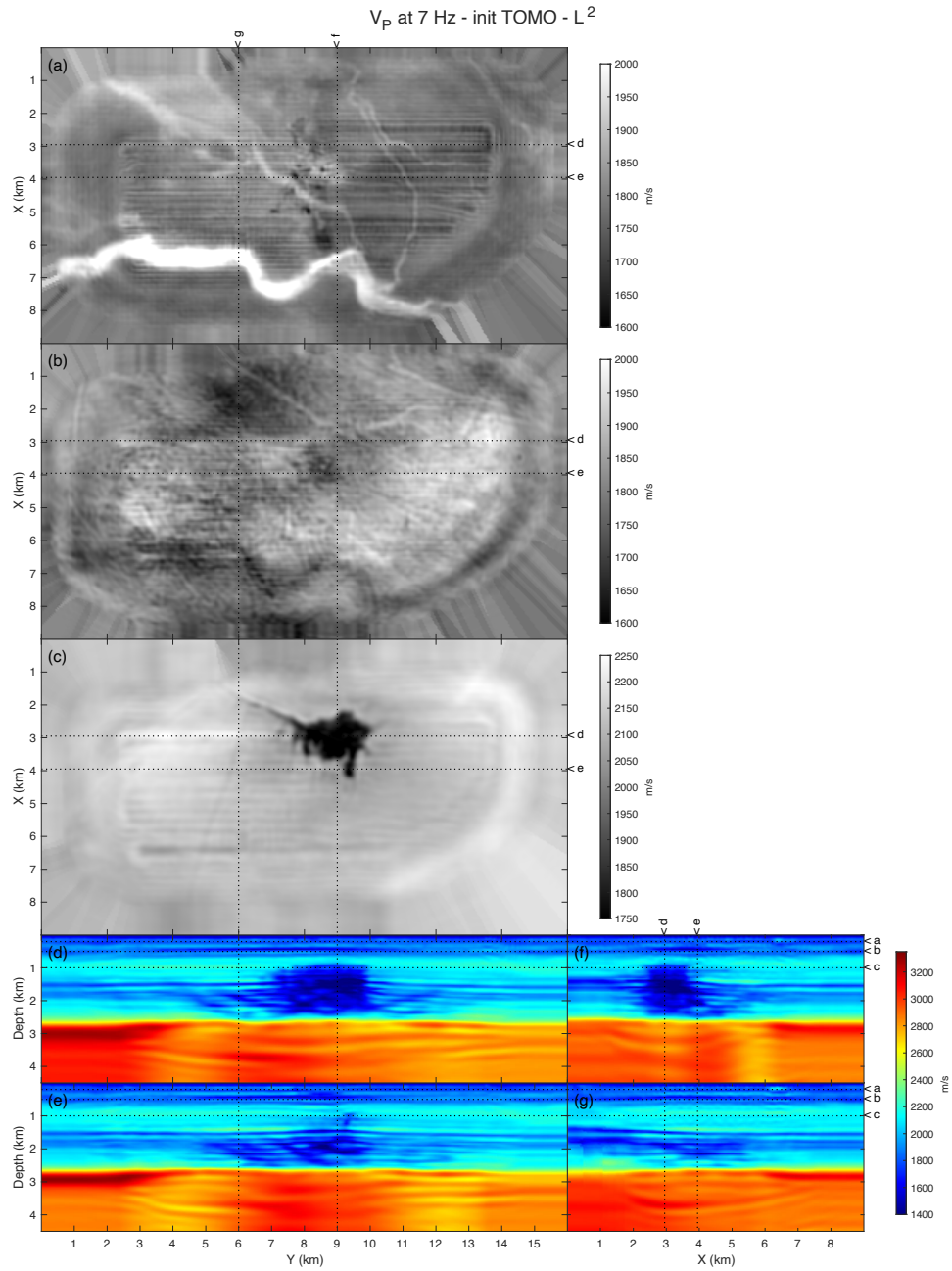


Figure 3.12: Slices of the 7 Hz FWI reconstructed V_P using L^2 misfit function starting from TOMO initial model. (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.

robustness compared to L^2 -based FWI?

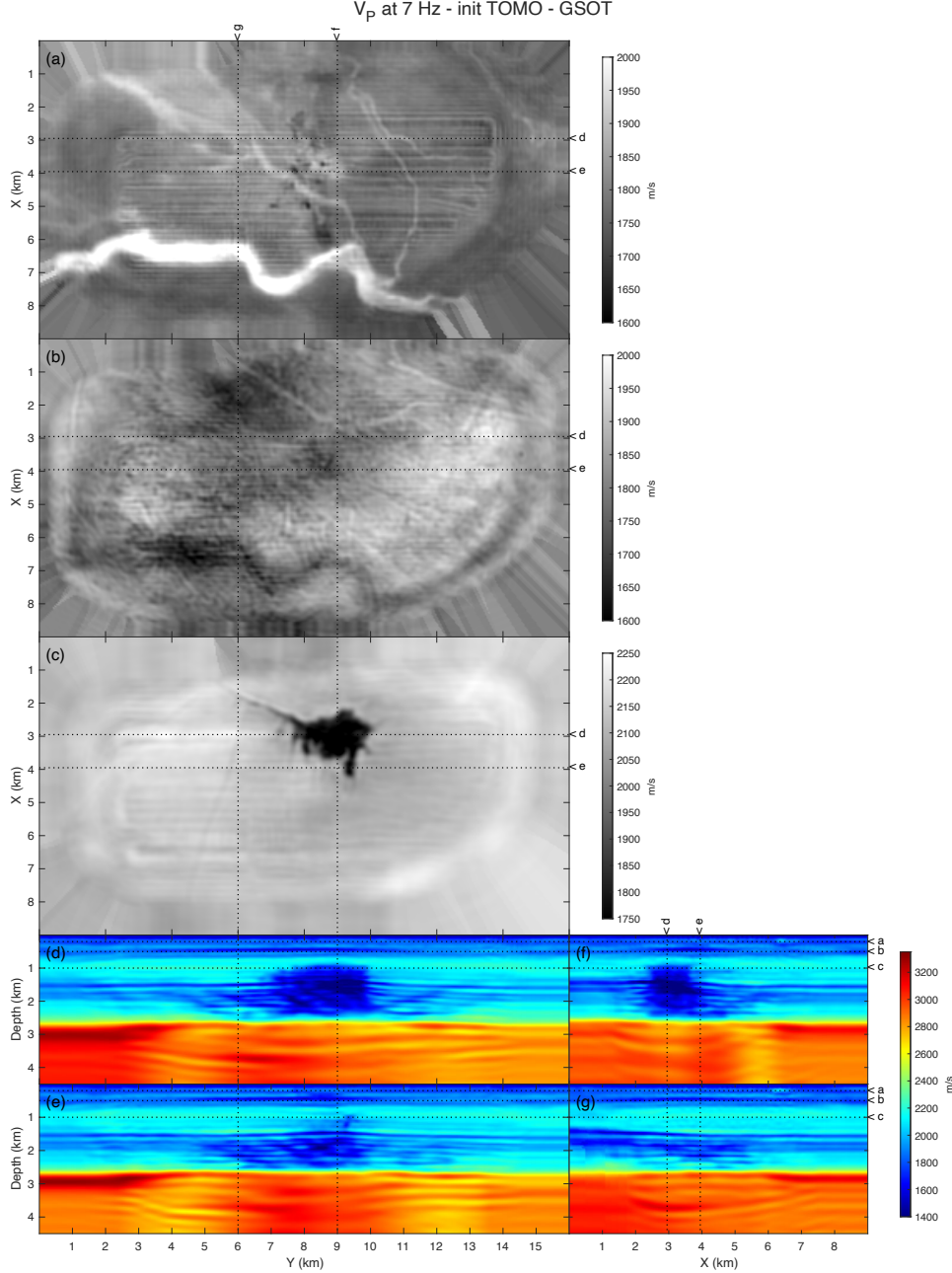


Figure 3.13: Same as Figure 3.12 using GSOT misfit function.

3.4.2 Starting from the 1D initial model

To validate the capability of GSOT to tackle large cycle-skipping, we use the 1D initial model that we introduced previously. This 1D initial model generates cycle-skipping, even on the mid-offset diving waves (mainly at -4 and -8 km offset), as clearly illustrated in Figure 3.8.

First, we compare the data fit obtained with L^2 and GSOT at the 2nd step of our FWI workflow as presented in Figure 3.16. For this first two steps of the FWI workflow, τ is set to 0.35 s, which is enough to handle the cycle-skipping generated in the data. First-order

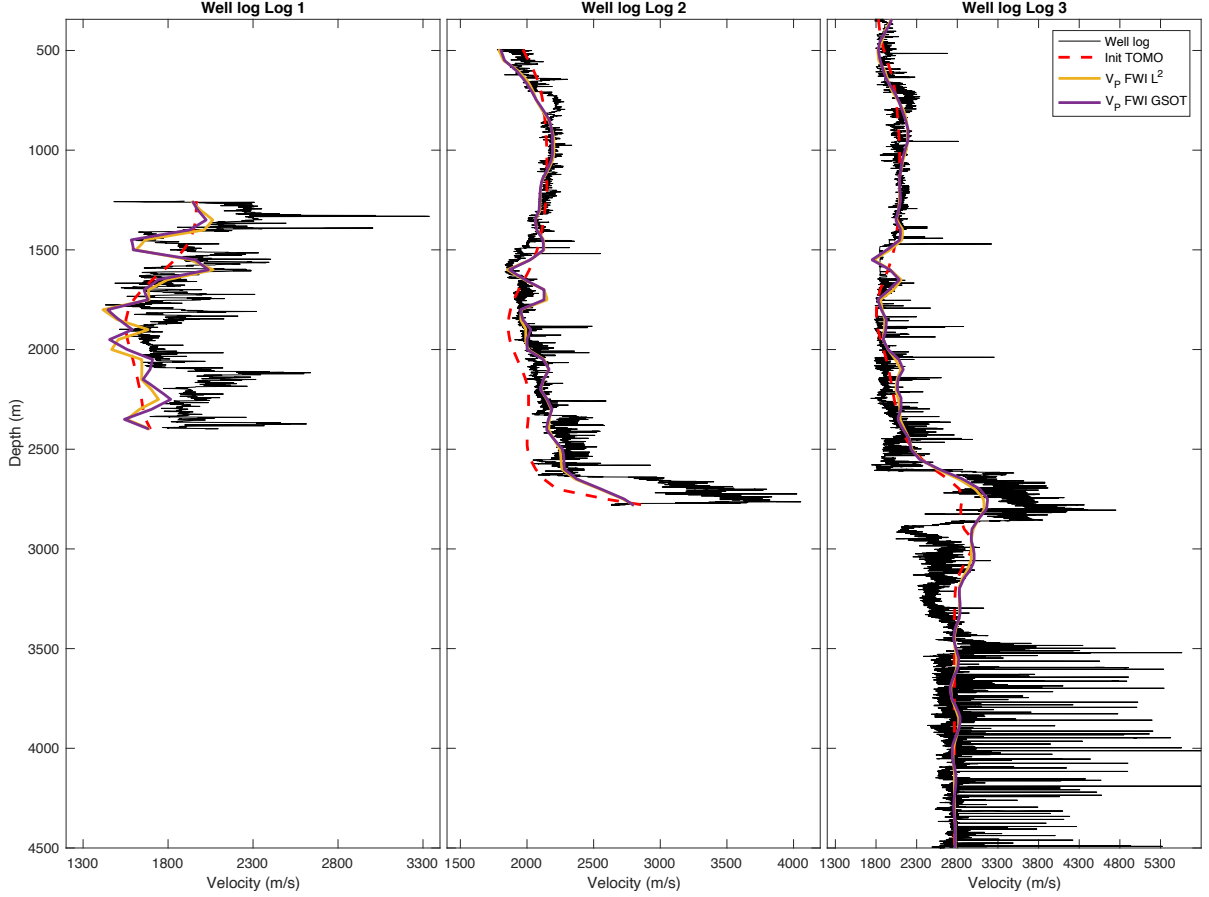


Figure 3.14: Comparison of V_P profiles extracted from the TOMO initial model (dashed red), 7 Hz FWI models using L^2 (solid yellow) and 7 Hz FWI model using GSOT (solid purple) with sonic log (solid black). Left subfigure corresponds to the Log 1 at the center of the target. Middle subfigure to the Log 2, and right subfigure to Log 3 (which is far away from the target).

observation may indicate a similar data fit with GSOT and L^2 , but a more in-depth analysis shows several differences. At offset higher than 5 km, we observe a degradation with out-of-phase arrivals for L^2 on the *rec A cable A* CRG, while GSOT results present a significant improvement on this part. Then, late arrivals events are better explained with GSOT. On the second *rec B cable B* CRG, we observe a more continuous reconstruction of the first events while also reducing out of phase one. Late arrivals are also better reconstructed with GSOT. On the second CRG, the data fit obtained with GSOT is already quite good for such an early stage of the inversion. In Figure 3.17 is presented the same CRG, but without applying the data weighting used at this early stage (DW MO data weighting), but instead the final relaxed data weighting (RT2 FO). Here, we can see that we are starting to predict data at larger offset when using GSOT compared to L^2 .

After the first two steps of FWI (DW SO and DW MO), we obtain the reconstructed model presented in Figures 3.18 and 3.19. It is clear here that L^2 -based FWI result displays heavy artifacts on the reconstructed V_P model for depth larger than 300 m. Even if the very shallow part of the model is correctly reconstructed, the deeper part of the model is not (see shallow slice (a) compared with deeper slices (b) and (c)). Starting from a crude 1D initial model,

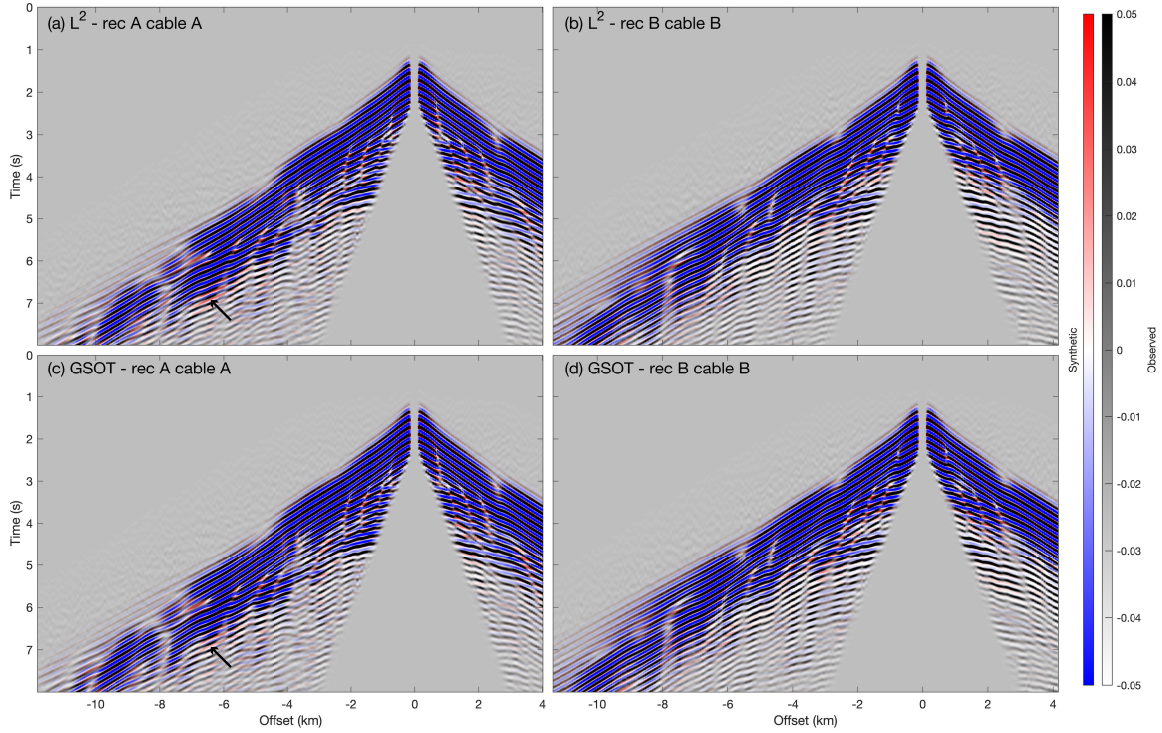


Figure 3.15: 2D common-receiver gathers at 7 Hz starting from the TOMO initial model. Synthetic data (blue/white/red color scale) generated into the final reconstructed V_P using (a,b) L^2 misfit function, (c,d) GSOT misfit function. (a,c) receiver A along cable A (through the low velocity anomaly). (b,d) receiver B along cable B. Field data are overlapped in grayscale with transparency. Black arrows point to improvement obtained with GSOT.

L^2 reaches its limit and is likely affected by strong cycle-skipping. This is why we stop the L^2 inversion at this stage of our workflow: pushing forward the inversion by introducing more data does not help. Conversely, GSOT-based FWI can provide promising and meaningful V_P updates, with the recovery of correct background velocities at depth and even key features such as the definition of strong low velocity anomalies (slices (c) (d) (f)).

As GSOT results are encouraging after only two passes on the first frequency band, we apply our complete workflow and perform the complete inversion similar to our reference inversion, with six passes on 5 Hz data and 7 Hz data. Regarding the τ parameter for GSOT, as said earlier, the first two steps on the first frequency band (2.5 to 5 Hz) were performed using $\tau = 0.35$ s. The next four steps on the first band used a reduced $\tau = 0.2$ s as it is enough to tackle the shift present in the data (which already illustrates the improvement that GSOT achieves in the first step of the inversion). For the second frequency band (2.5 to 7 Hz), only the first step of the workflow uses a slightly relaxed τ of 0.25 s, while the remaining five steps use $\tau = 0.2$ s. Again, τ is relaxed at the beginning to mitigate cycle-skipping, while a smaller τ is used afterward to preserve resolution power and speed up convergence.

We present the final data fit at 7 Hz for GSOT in Figure 3.20. We can see that a relatively good data fit is obtained, with most of the arrivals correctly explained. Still, some late arrivals are out of phase, mainly for large offsets (larger than 6 km)

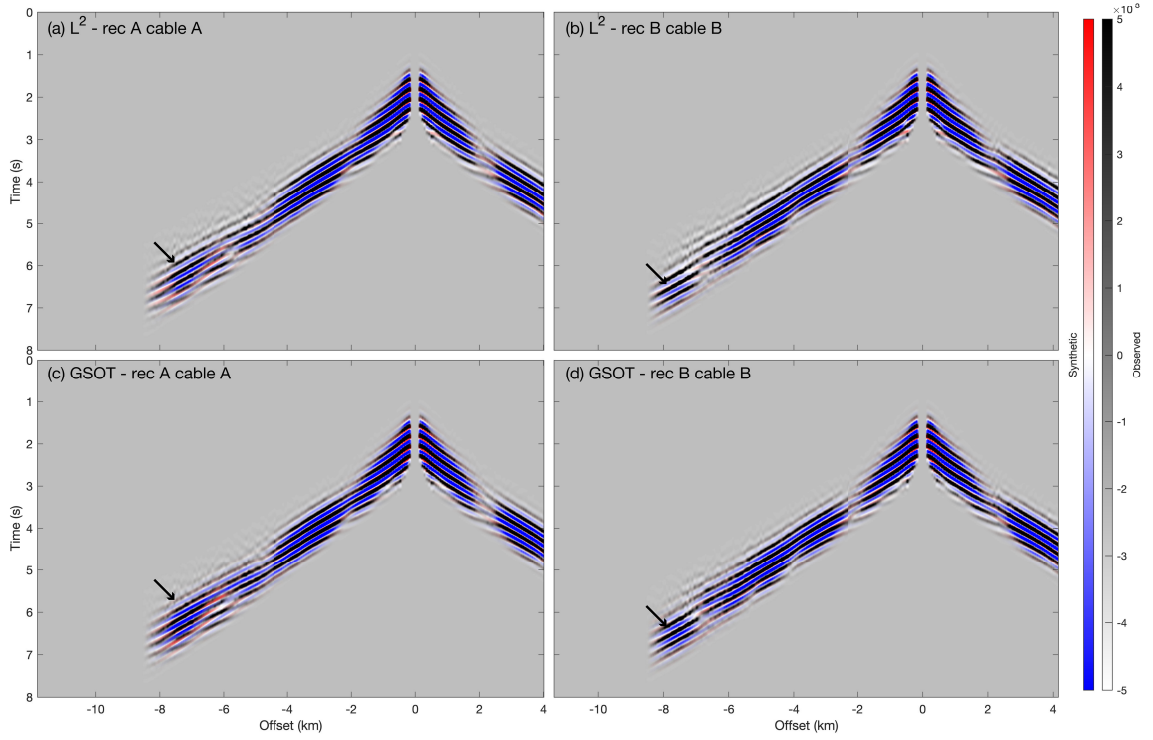


Figure 3.16: 2D common-receiver gathers at 5 Hz starting from the 1D initial model with data weighting apply (DW MO). Synthetic data (blue/white/red color scale) generated into the final reconstructed V_P at 2nd workflow step using: (a,b) L^2 misfit function, (c,d) GSOT misfit function. (a,c) receiver A along cable A (through the low velocity anomaly). (b,d) receiver B along cable B. Field data are overlapped in grayscale with transparency. Black arrows point to area where GSOT improves the datafit.

We obtain the final reconstructed V_P presented in Figure 3.21. The results are promising and show a clear improvement in resolution compared to the early results at 5 Hz, with the main target structures retrieved above 2 km depth. The shape of the low velocity anomaly is correctly retrieved (slice (c) (d) (f)). Lateral resolution is very significantly improved, allowing the definition of narrow low-velocity (150 m wide) anomalies not resolved in the L^2 inversion (slice (e)).

When performing a QC using the sonic log presented in Figure 3.22, comparing the final GSOT results with the early L^2 results (that could not be pushed further with meaningful updates) clearly illustrates that GSOT-based FWI performs meaningful updates of the model that follows the trends of the V_P sonic log. Conversely, the L^2 -based FWI updates are not correcting for the wrong initial model, and updates go in the wrong direction at several depths (for example, 1200 m and 1750 m). A comparison between the sonic log and GSOT-based FWI results for the two starting models is shown in Figure 3.23. For the two logs outside the center target area (Log 2 & 3), a good agreement of the FWI results is observed, with reconstructed V_P models following the same trend. For Log 1 in the center of the target area, we can expect that reconstructed V_P results will be incorrect below approximately ≈ 2 km due to the poor illumination. Above 1.4 km depth, a good agreement can be observed between the reconstructed models obtained using the two starting models.

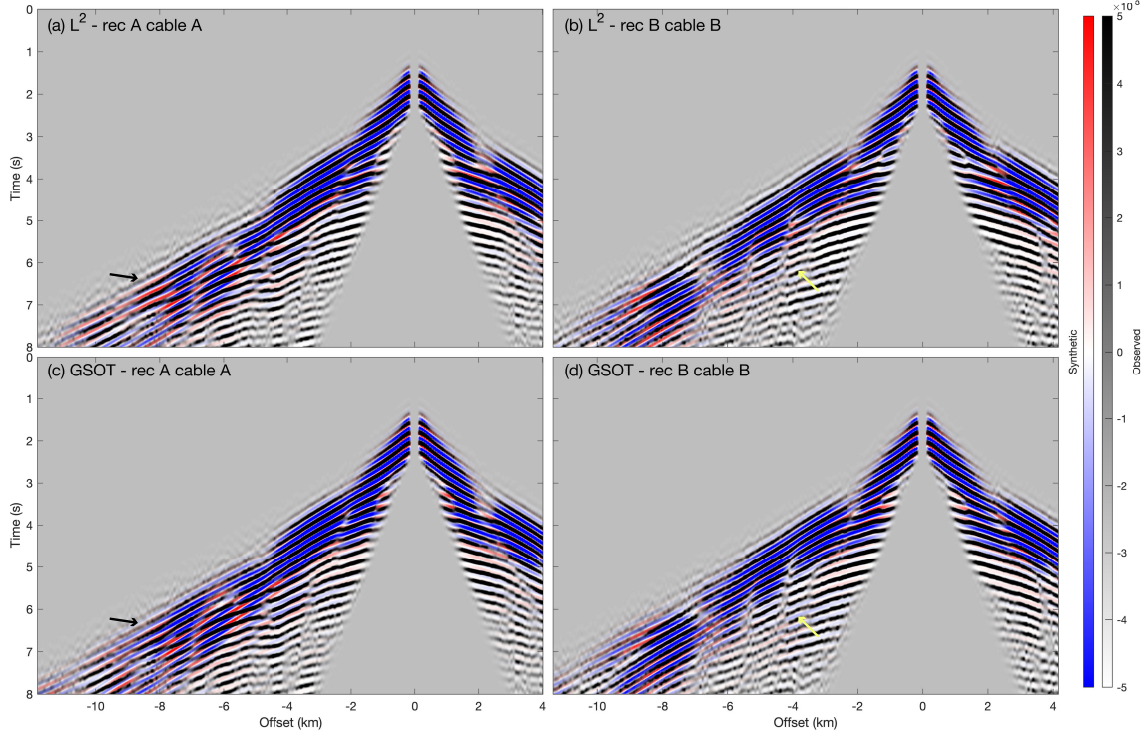


Figure 3.17: Same as Figure 3.16 but with final relaxed data weighting (RT2 FO) for display. This exhibits the improvement of datafit obtained in area which are not yet inverted. Here, black arrows point to improve fit and coherency of the diving wave with GSOT, whereas yellow arrows point to improved data-fit of reflected events with GSOT.

The results are still not perfect, and the main issue is the presence of a low-velocity update around 500 m depth, as made visible on slice (b). This low-velocity update (not present in the reference FWI results starting from TOMO initial model) introduces a down vertical shift of layer under this perturbation. This is why slice (c) is extracted 100 m under (1.1 km instead of 1 km depth). This vertical shift does not affect the shape of the low velocity anomaly but only its depth. One possible way to avoid this artifact would be to modify the early stage of the inversion, for example, with different data-selections or modifications of the initial model for a slightly better one (for example, based on sonic log information). While this would have probably improved the results, we decided to keep a crude 1D initial model to stay as generic as possible and assess the capacity of GSOT in a setup without prior information.

3.4.3 Computational costs

Computational cost analysis is performed for one gradient estimation for both L^2 and GSOT misfit functions. Computation is performed on Haswell E5-2690V3@2.6 GHz Intel nodes containing 24 cores and 64 Gb of memory per node. We use 4 cores per source as our finite difference scheme uses an OpenMP parallelization.

The computational costs on the first frequency band (2.5 – 5 Hz) and second frequency band (2.5 – 7 Hz) are summerized in the table.

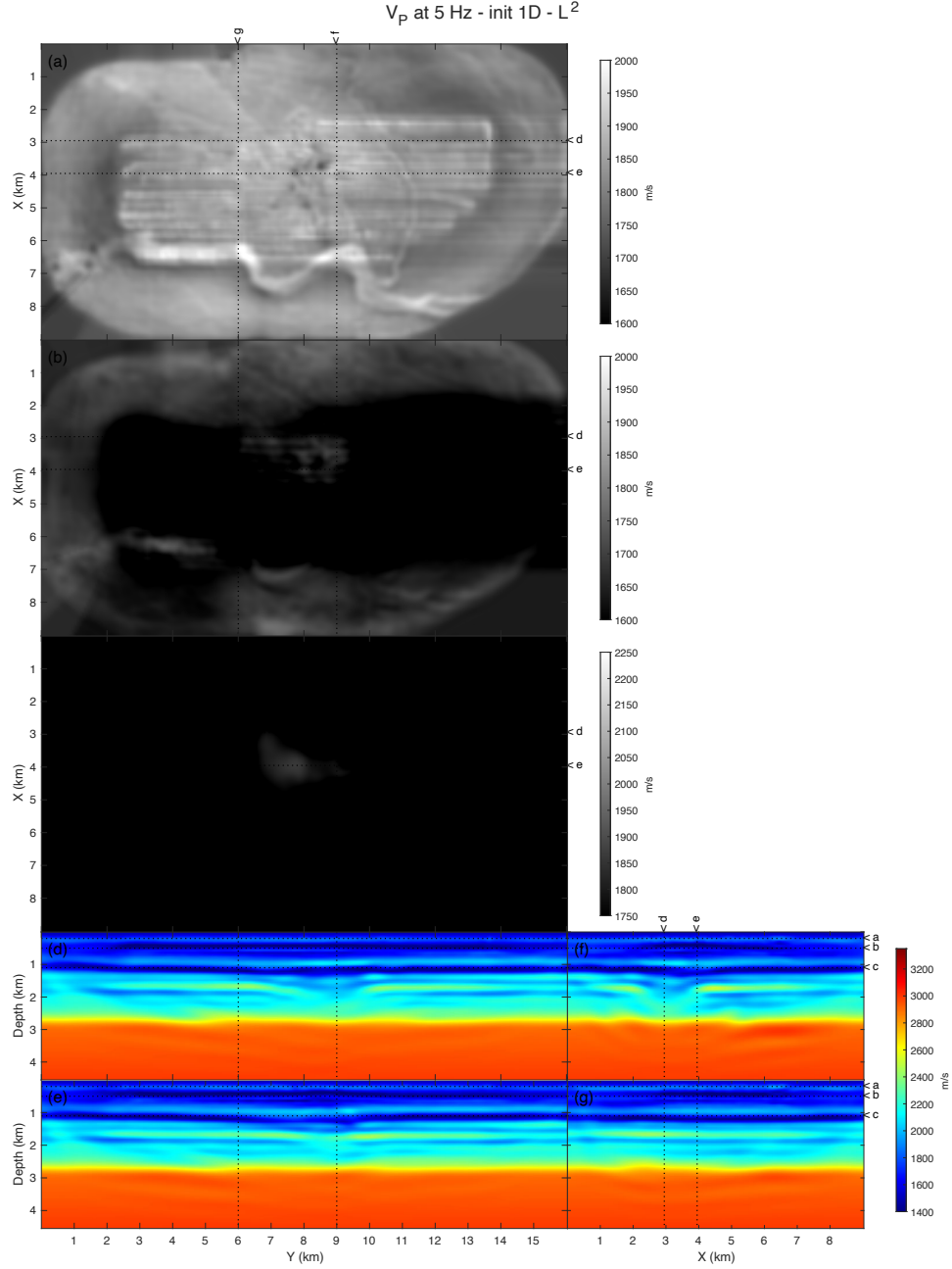


Figure 3.18: Slices of the 5 Hz FWI reconstructed V_P using L^2 misfit starting from 1D initial model. (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1.1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.

	Gradient	Misfit	Total time	Ratio
L^2 - 5 Hz	243 s	1 s	254 s	100 %
GSOT - 5 Hz	243 s	55 s	308 s	121 %
L^2 - 7 Hz	898 s	1 s	912 s	100 %
GSOT - 7 Hz	898 s	101 s	1012 s	111 %

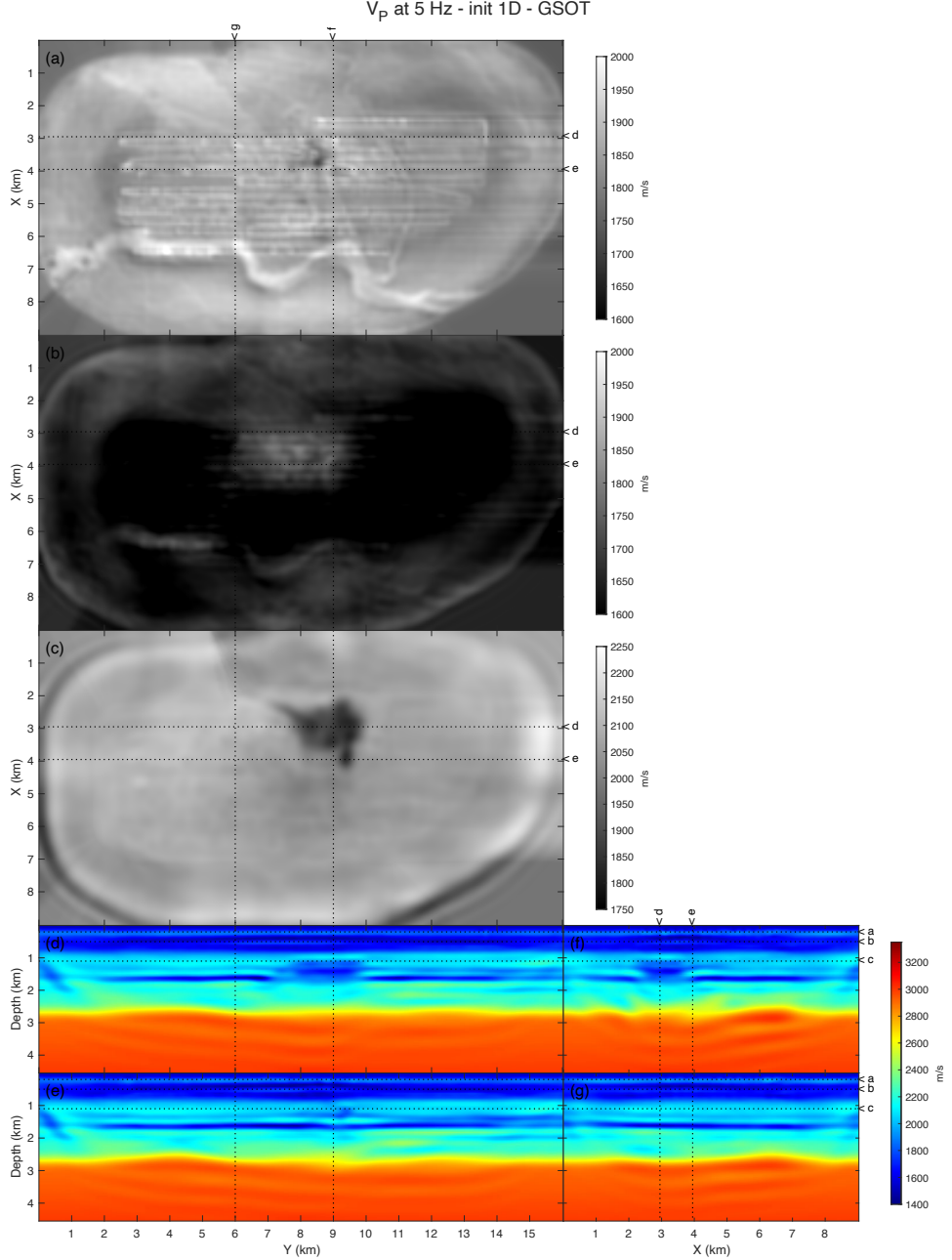


Figure 3.19: Same as Figure 3.18 using GSOT. Here, results are consistent until 1.5 to 2 km depth compared to L^2 -based FWI. Characteristic structures of the Valhall field are recovered.

The gradient column corresponds to the time spent to calculate the gradient (containing all required wave modelings), while the misfit column isolates the time spent computing the misfit function. The total time is the sum of gradient and misfit, and the ratio compares L^2 (put at 100%) with GSOT. This analysis shows that while the computational complexity of the solution of the gradient estimation scales to $O(\omega^4)$, the computation complexity of the GSOT computation is in $O(\omega^3)$, as noted in Métivier et al. (2019).

The overhead cost induced by GSOT is therefore reduced on the higher frequencies, which

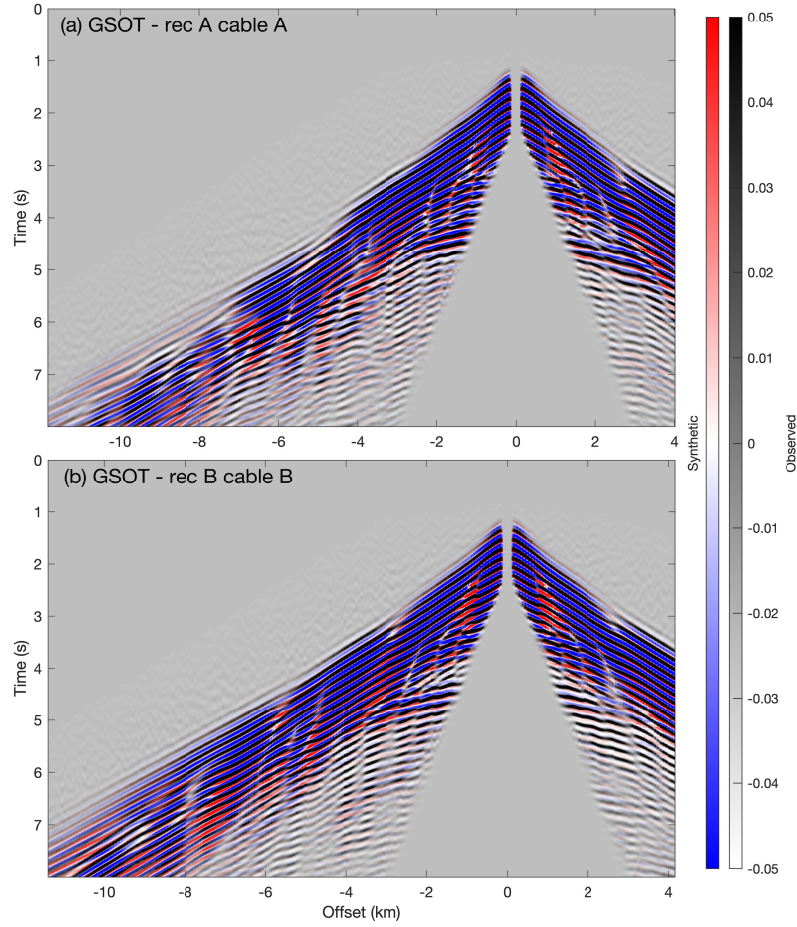


Figure 3.20: 2D common-receiver gathers at 7 Hz starting from the 1D initial model. Synthetic data (blue/white/red color scale) generated into the final reconstructed V_P using GSOT. (a) receiver A along cable A (through the low velocity anomaly). (b) receiver B along cable B. Field data are overlapped in grayscale with transparency.

are the most expensive ones computationally speaking (for example, with a maximum frequency of 5 Hz, the first frequency band is relatively fast to compute, making a 20% overhead acceptable). This is one key feature that makes the GSOT misfit function able to tackle field data applications as higher frequency drastically increases the computational cost, and GSOT overhead will become smaller.

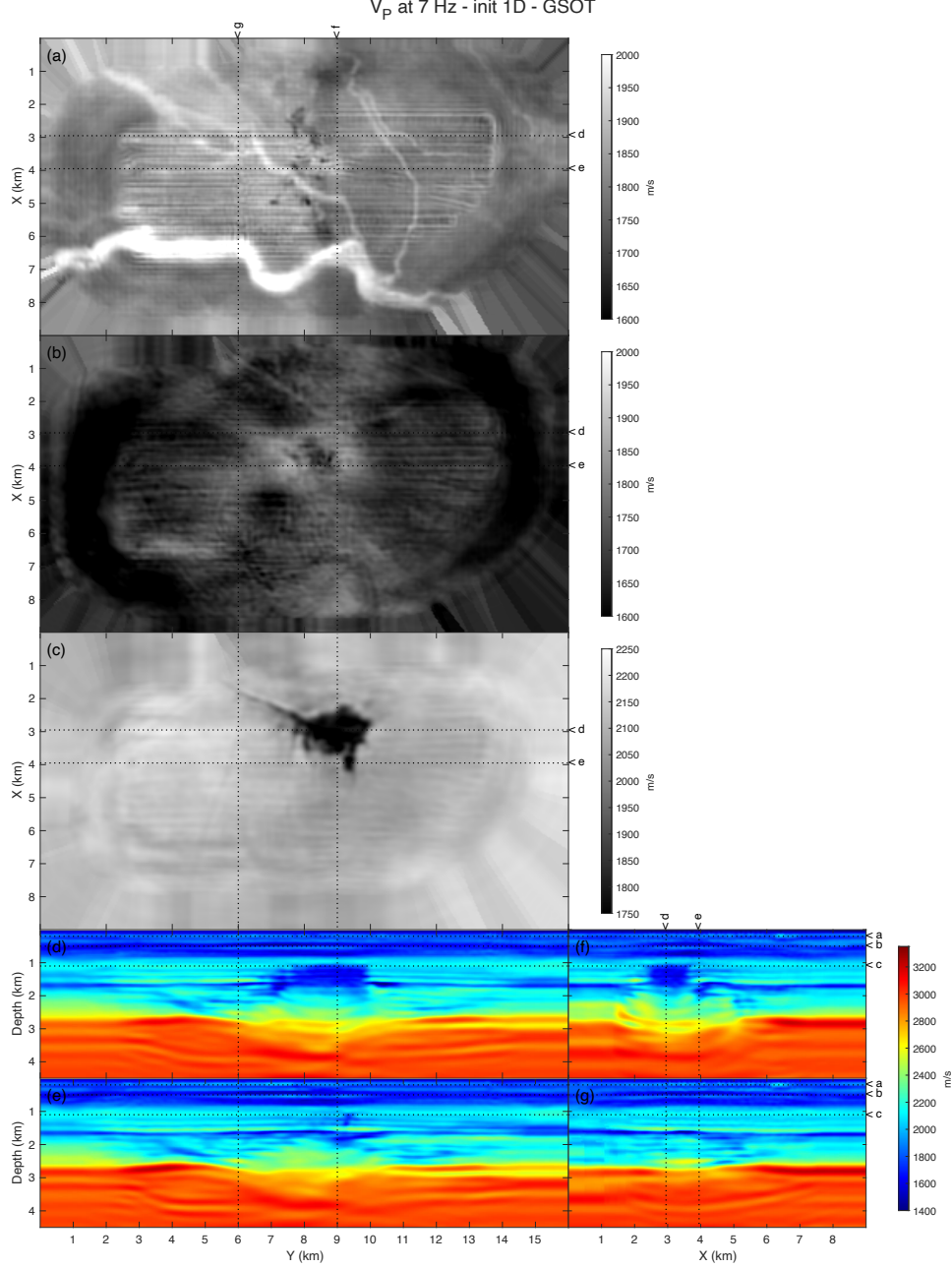


Figure 3.21: Slices of the 7 Hz FWI reconstructed V_P using GSOT misfit starting from 1D initial model. (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1.1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.

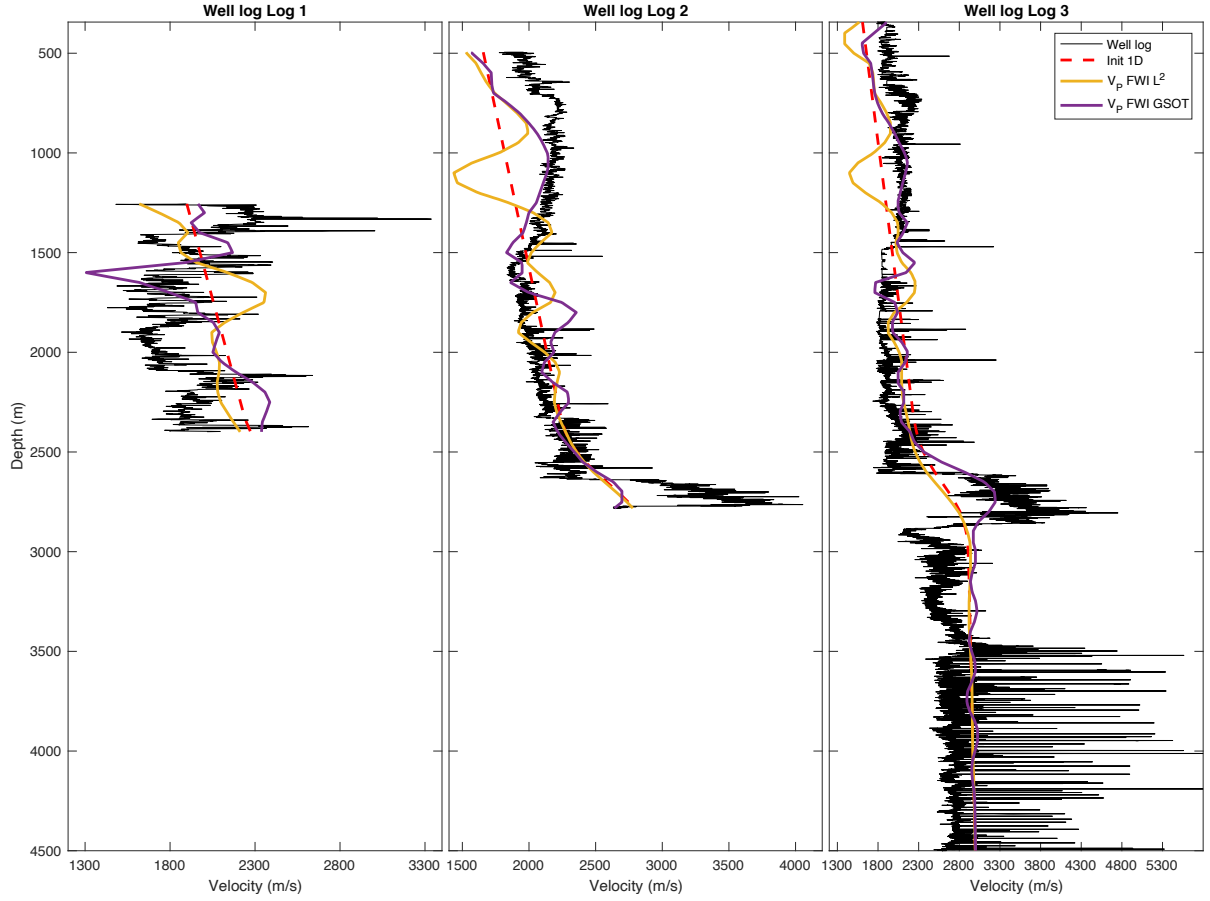


Figure 3.22: Comparison of V_P profiles extracted from the 1D initial model (dashed red), early FWI models using L^2 (solid yellow) and 7 Hz FWI model using GSOT (solid purple) with sonic log (solid black). Left subfigure corresponds to the Log 1 at the center of the target. Middle subfigure to the Log 2, and right subfigure to Log 3 (which is far away from the target). Updates of velocity model obtained with GSOT are following the logs trend until ≈ 2 km depth.

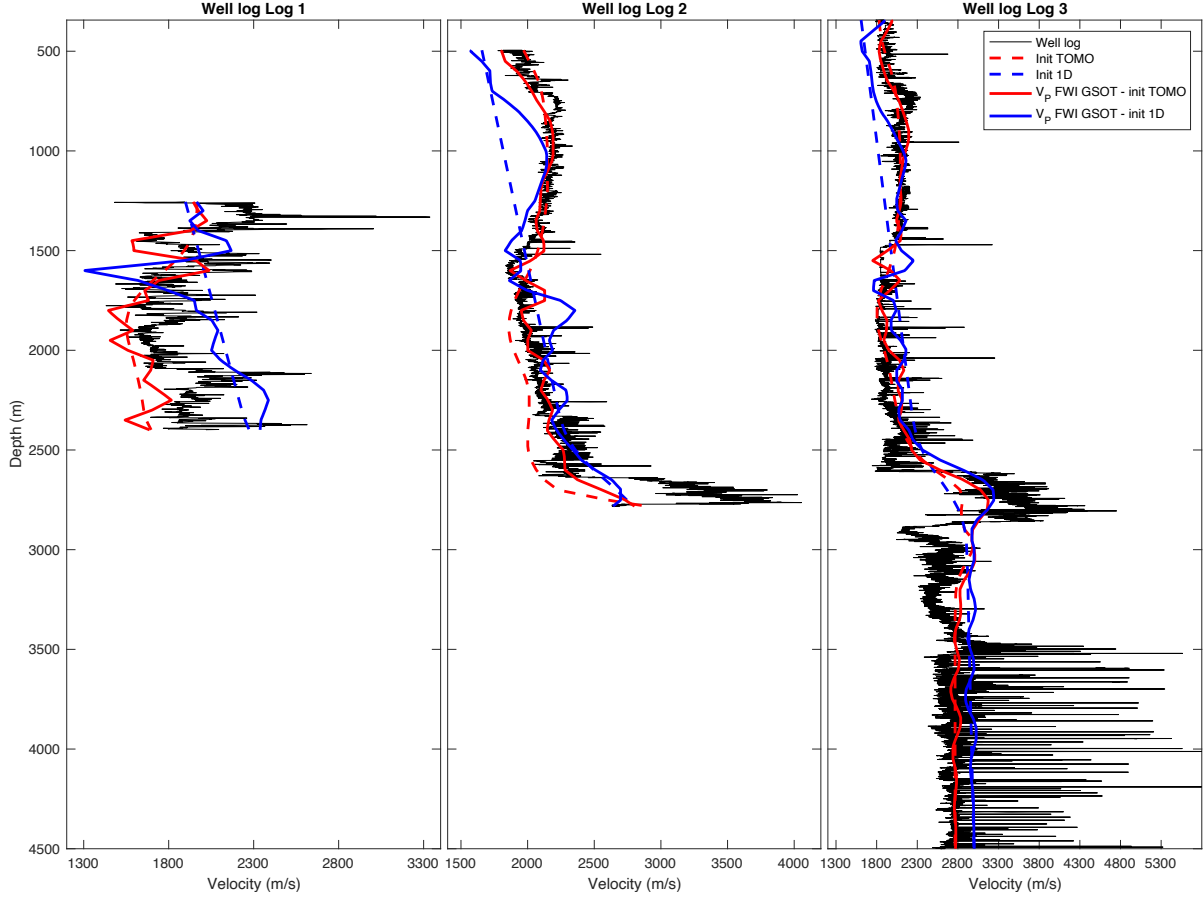


Figure 3.23: Comparison of V_P profiles extracted from the TOMO and 1D initial models (respectively dashed red and dashed blue), GSOT-based FWI reconstructed models at 7 Hz starting from TOMO and 1D initial models (respectively solid red and solid blue), with sonic log (solid black). Logs 2 and 3 show that results from the two different starting models are globally following the same trend. Results from Log 1 passing through the target are following the same trend until 1.4 km depth.

3.5 Discussion

Facing challenging field data applications with alternative misfit functions is not widely documented in the literature. One of the only other alternative misfit functions that has been applied successfully to field data is adaptive waveform inversion (AWI) (Warner and Guasch, 2015; Ravaut et al., 2017; Debens et al., 2017; Roth et al., 2018; Guasch et al., 2019; Warner et al., 2019) or Kantorovich-Rubinstein optimal transport (KROT) (Poncet et al., 2018; Messud and Sedova, 2019; Sedova et al., 2019).

Regarding the final results obtained with the GSOT misfit function, we first validate that it can improve over L^2 -based FWI in a controlled environment (starting from a good enough V_P TOMO) with improved datafit and improved lateral coherency and reduced acquisition imprint on the model side. The computational overhead induced by GSOT stays limited thanks to the computational complexity of GSOT being one order of magnitudes smaller than the complexity of the gradient estimation for a given maximum frequency. This behavior translates into a good scaling property when facing high-frequency data, which is now one trend for field data FWI applications. When tackling a difficult 1D initial model, which generates cycle-skipping, we show that GSOT-based FWI provides good model updates and good datafit where classical L^2 -based FWI fails. The control of the convexity provided by the GSOT misfit function is performed through the choice of the τ parameter. This parameter is directly linked to the observed time-shift between calculated and observed data, making it easy to tune, depending on the initial fit of the data and the expected maximum time shift. We adapt it from 0.2 s to 0.35 s in the initial stage of the workflow when we switch from the initial tomography model to the initial 1D model.

We also emphasize that results under 2 km suffer from a substantial lack of illumination and are therefore limited to the resolution of our initial model, explaining why no meaningful updates are present for depth superior to 2 km. This limitation in terms of depth reconstruction (under 2 km) is not surprising as the FWI alone is not expected to present enough illumination. In this case, it would require the use of reflected wave inversion (RWI) or joint full-waveform inversion (JFWI). Some preliminary but encouraging results were obtained combining GSOT and JFWI (Provenzano et al., 2020b): GSOT adds the convexity necessary to predict the reflected data, enabling robust velocity updates of the model at depth.

3.6 Conclusion

In this work, we focus on the application of a new misfit function: graph space optimal transport. This formulation shows a clear improvement over L^2 in a controlled environment, unleashing the potential of FWI to perform meaningful updates when starting from a crude, 1D initial model. This clear improvement in cycle-skipping robustness, combined with a simple setup (only one *physical* parameter to define) and a reasonable computational overhead, illustrates that GSOT is a good candidate to improve FWI robustness and therefore make FWI more accessible and easily applicable.

Chapter 4

Exploring new ideas and overcoming practical limitations

Contents

4.1	Overcoming practical issue for 3D KROT applications	109
4.2	Exploring new ideas of alternative misfit functions	112
4.2.1	Instantaneous phase and unwrapped phase misfit functions	112
4.2.2	Spline based envelope misfit function	117
4.2.3	Diffusion based misfit function	121
4.2.4	Beat-tone misfit function	125
4.3	Investigation published during my thesis	129

This chapter presents some of the ideas and workaround that I have developed during my years of research. This could be seen as the “graveyard” of multiples ideas, that could seem interesting but were never pushed further for multiples reasons. Some ideas were simply not suitable, some were interesting but present problematic flaws, and better ones simply replaced some.

I think it is important to remind that science is not a straight path. It often requires going through multiple paths to have a better understanding of the possibilities. Unfortunately, the scientific community often focuses only on working solutions but not so much on what is not working. This leads researchers worldwide to repeatedly try the same “doomed-from-the-beginning” idea that nobody ever published about.

In a small and humble attempt to improve this, I present some of what I could refer to as “interesting failures”, that still brings meaningful insights on the road ahead for better and more robust FWI by possibly removing some dead-ends for future research in this field.

4.1 Overcoming practical issue for 3D KROT applications

As illustrated in Figure 4.1, the Kantorovich-Rubinstein optimal transport (KROT) does not translate into drastic convexity improvement. The main benefit comes from the fact that it can be defined in 1D, 2D, or 3D, allowing us to consider the lateral coherency of shot gathers. By taking lateral coherence information of the data into account, KROT allows having a beneficial effect on the convergence of the FWI schemes. It also helps “regularize” the results, with a denoising effect that is welcome for field data application. This has been illustrated in recent publications (Messud and Sedova, 2019; Sedova et al., 2019; Carotti et al., 2020).

For 3D field data, keeping computational costs low is essential, and it is essential to minimize the potential overhead introduced by the KROT formulation. To mitigate the computational overhead of 3D KROT, we rely on a new idea: extract lines of receivers (Figure 4.2) inside the 3D seismic cube (Poncet et al., 2018). This approach creates 2D panels on which 2D KROT can be applied. The 3D residual will be reconstructed using all 2D residuals, and the cost function will be the summation of every panel cost.

The algorithm for this new optimal transport work as follows:

- Choose the direction of receivers line (in X or Y) and number of receivers lines to extract in the 3D cube (these parameters should make sense with the acquisition)
- A width of X or Y values for a line is calculated depending on the size of the acquisition
- A line is extracted from the 3D cube by selecting all receivers inside a given range of x_{rec} or y_{rec} values
- 2D KROT is applied on the panel. The cost function is increased by the value of the cost for this 2D panel, outputted 2D residual for the line is remapped to the 3D residual (each traces of 2D residual is put back where it was in the complete 3D residual)
- Code move to next line until all lines are done

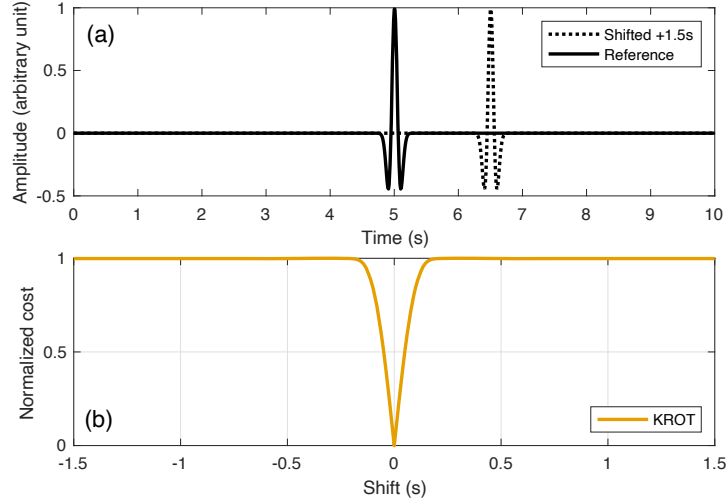


Figure 4.1: Simple time-shifted Ricker wavelet test for crosscorrelation misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the KROT misfit function with respect to time-shift.

This yields a more flexible KROT implementation for 3D data, which presents a reasonable computational overhead (increase around 25% max compared to L^2) while preserving the ability of KROT to take advantage of lateral coherency of the data compared to 1D KROT.

The only potential issue is that multi-2D implementation assumes that the extracted 2D line is straight. This can be circumvented by performing receivers re-interpolation on a regular cartesian grid directly on the data, which allows performing a regular line extraction afterward (Figure 4.3).

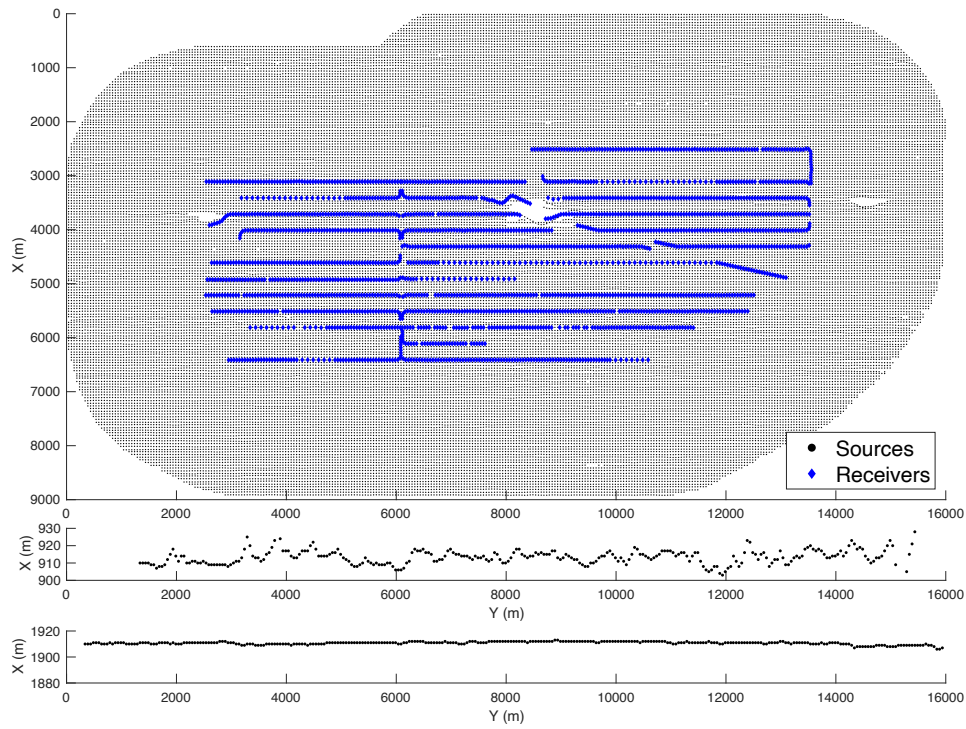


Figure 4.2: Complete (original) acquisition of a 3D OBC North Sea Valhall data. Sources are represented by black dots, receivers by blue diamond. Two subpanels shows a 2D extracted lines of receivers inside the acquisition.

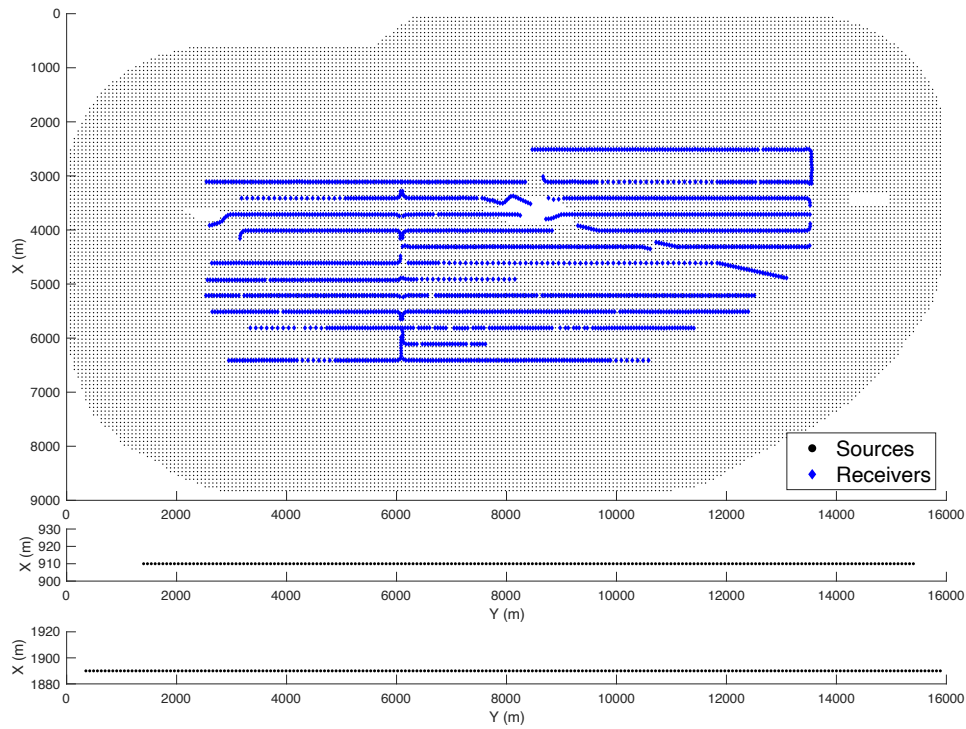


Figure 4.3: Same as Figure 4.2, but for the interpolated dataset. Here the extracted 2D lines are now perfectly straight, with regular receivers spacing.

4.2 Exploring new ideas of alternative misfit functions

In this section, I present some ideas that I have explored. These ideas have been explored mainly on an early “prototyping” stage, focusing on increasing convexity for time-shifts on simple 1D shifted patterns tests. These ideas have never passed this early stage and have never been implemented into FWI for different reasons that I will detail.

4.2.1 Instantaneous phase and unwrapped phase misfit functions

First, in this section, I re-introduce the concept behind instantaneous phase and envelope (as proposed by Bozdağ et al., 2011). One way to separate the signal phase and envelope information is to rely on the analytical signal based on the Hilbert transform.

4.2.1.1 Hilbert transform and analytical signal

The analytical signal relies on the utilization of the Hilbert function \mathcal{H} which can be defined in the time domain as

$$\mathcal{H}[d(t)] = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{d(\tau)}{t - \tau} d\tau, \quad (4.1)$$

where P stands for the Cauchy principal value.

In practical application, we do not rely on this time formulation to get the Hilbert transform but rather on a frequency formulation proposed by Marple (1999). This method relies on a simple three-step approach:

1. Compute the Fourier transform of $d(t)$ using an efficient FFT solver
2. Change the negative frequency to zero
3. Compute the inverse FFT

This simple algorithm gives direct access to the analytical signal in a computationally efficient manner. From this analytical signal, we can simply retrieve the Hilbert transform as the imaginary part of the analytical signal $\tilde{d}(t)$

$$\mathcal{H}[d(t)] = \mathcal{I}[\tilde{d}(t)]. \quad (4.2)$$

This analytical signal can be written regarding the instantaneous phase $\phi(t)$ and the instantaneous envelope $E(t)$ as

$$\tilde{d}(t) = E(t)e^{i\phi(t)}, \quad (4.3)$$

where the instantaneous envelope is defined as

$$E(t) = \sqrt{\mathcal{R}[\tilde{d}(t)]^2 + \mathcal{I}[\tilde{d}(t)]^2}, \quad (4.4)$$

and the instantaneous phase as

$$\phi(t) = \text{atan} \frac{\mathcal{I}[\tilde{d}(t)]}{\mathcal{R}[\tilde{d}(t)] + \varepsilon}, \quad (4.5)$$

or alternatively using the `atan2` that has been introduced in Fortran and was originally intended to return a correct and unambiguous value for the angle in cartesian coordinates

$$\phi(t) = \text{atan2} \left(\mathcal{I}[\tilde{d}(t)], \mathcal{R}[\tilde{d}(t)] \right). \quad (4.6)$$

The `atan2` formulation of instantaneous phase is preferred as it is unconditionally stable and does not require a water level ε .

This decomposition between the instantaneous phase and instantaneous envelope can be used to define alternative misfit functions as detailed in the following. In chapter 2, we have already seen that instantaneous envelope could provide a potential improvement over the L^2 norm. In the following, I want to give more details on the instantaneous phase and some other ways to obtain “envelope-like” observables that could be used for alternative misfit functions.

4.2.1.2 Instantaneous phase misfit

Using the instantaneous phase defined on equation 4.6, an instantaneous phase misfit has been proposed by Bozdağ et al. (2011). The motivation behind the utilization of the phase information is to reduce the sensitivity to amplitude variation, which mainly impacts global seismology. At the crustal scale, these amplitude problems are less present.

A representation of the instantaneous phase of a signal is shown on Figure 4.4. It exhibits that the instantaneous phase is oscillatory and could not increase cycle-skipping robustness as the width of the temporal support carrying information is now smaller than the original signal.

This instantaneous phase misfit function is defined as

$$F_{IP} = \frac{1}{2} \int_0^T [\phi_{cal}(t) - \phi_{obs}(t)]^2 dt. \quad (4.7)$$

This misfit function is not designed to improve the convexity for time delays (Figure 4.5), but rather improve robustness to amplitude miss-match. It is sensitive to polarity information, and its resolution is prone to be good. One of the issues of removing amplitude information is that the weighting between noise and the arrivals will be the same, leading to an increased sensitivity to noise.

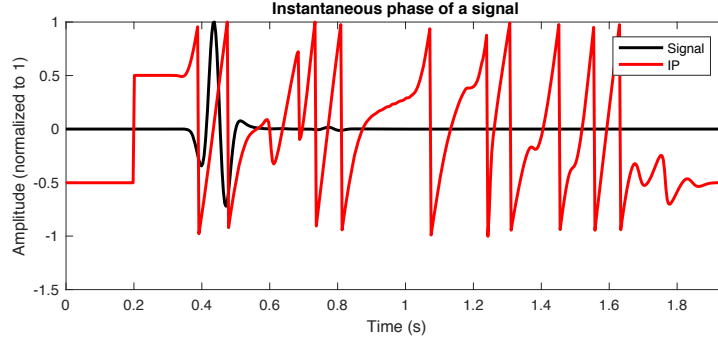


Figure 4.4: Representation of a simple signal $d(t)$ (black line) and the associated instantaneous phase $\phi(t)$ (red line).

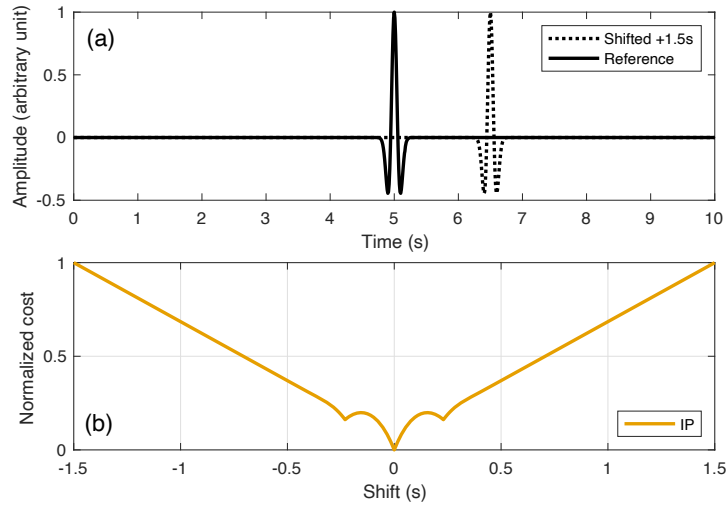


Figure 4.5: Simple time-shifted Ricker wavelet test for the instantaneous phase misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the instantaneous phase misfit function with respect to time-shift.

4.2.1.3 Unwrapped phase misfit

From the instantaneous phase can be derived the unwrapped phase, which can potentially lead to an interesting formulation. The unwrapping of the phase is done by a simple algorithm: whenever the jump between consecutive angles is greater than or equal to π , the algorithm shifts the angles by adding multiples of $\pm 2\pi$ until the jump is less than π .

Using an unwrapped phase, we can obtain a drastic improvement of convexity for time delays, which mimic NIM misfit function (Figure 4.6). Unfortunately, unwrapping of phase is an arduous task that is incredibly sensitive to noise (Figure 4.7), and is still challenging today in optics or electronic fields (Wang et al., 2019). It would be impossible to extract a clean and stable unwrapped phase of a field data signal. Therefore all the attempts to use the unwrapped phase were not further explored.

This idea remains an excellent example of tempting ideas that could work well on synthetic traces but are doomed to failure when analyzed a bit closer. It represents the large amount of “dead-ends” that I have faced when trying to find alternative misfit functions.

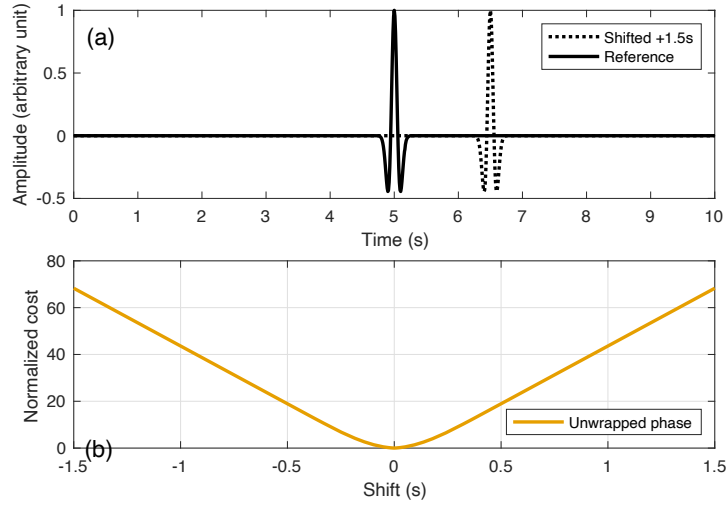


Figure 4.6: Simple time-shifted Ricker wavelet test for the unwrapped phase misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the unwrapped phase misfit function with respect to time-shift.

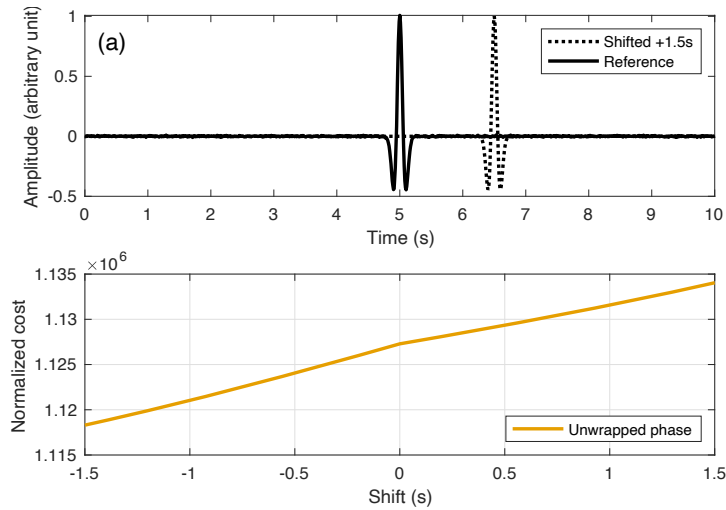


Figure 4.7: Same as Figure 4.6 but with a very small white noise on the reference signal.

4.2.2 Spline based envelope misfit function

The Hilbert transform is one way to extract the “true” envelope of the signal. By “true” it directly relates to the structure of a signal, which is composed of phase and amplitude, which translates into the fact that the instantaneous envelope can be recombined with the instantaneous phase to recreate the original signal. However, another possibility is to extract an “envelope” from the signal in a more physical sense, with no direct link to the phase.

To create this new envelope, we first need to extract all the maxima (and minima) of the signal; then, spline interpolation is performed between these maxima points (or minima point) to create a spline envelope approximation of the signal as proposed by Han et al. (2016). Thus, we can define two separate envelopes: the upper part (from the maxima) and the lower part (from the minima). This is the first drastic difference to the unique envelope created by the Hilbert transform. These lower and upper envelopes can be merely average (Hu et al., 2016) to create an “average” spline envelope (ASE). The original idea is performed using cubic spline interpolant. From the extensive test performed, a better version was obtained using a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP).

The processing of the signal is performed in two steps. The first one (Algorithm 1) computes the first and second derivative of the signal, then a second part (Algorithm 2) extracts the top and bottom peaks. From the peaks, the PHCIP algorithm can be applied to get the top and bottom envelope.

An example of a spline envelope and average spline envelope is presented on Figure 4.8. In this case, on synthetic data, the average envelope obtained from the top and bottom envelope could make an excellent observable as the width of the temporal support has been increased while preserving phase and amplitude information.

The convexity improvement obtained with the ASE misfit is slightly improved compared to the instantaneous envelope, as shown in Figure 4.9. The advantages of ASE over IE would not only be related to slightly increased convexity but also to the fact that it integrates some polarity (phase) information that IE discards. This could potentially translate into improved results for practical application where phase information is crucial.

The effect of noise on the spline envelope is interesting. Test of data without noise vs.

Algorithm 1 calculate the first and second derivative of the signal and extract binary version of the first derivative

```

1:  $sn = s + \text{small white noise}$ 
2: for  $i = 2$  to  $N_t - 1$  do
3:    $ds(i) = (sn(i + 1) - sn(i - 1)) / (2 * dt)$ 
4:    $dds(i) = (sn(i + 1) - 2 * sn(i) + sn(i - 1)) / (dt * dt)$ 
5: end for
6: for  $i = 1$  to  $N_t$  do
7:   if  $ds(i) \geq 0$  then
8:      $sign(i) = 1$ 
9:   else if  $ds(i) < 0$  then
10:     $sign(i) = -1$ 
11:   end if
12: end for

```

Algorithm 2 extract top and bottom peaks

```

1:  $p_{top} = 0$ 
2:  $p_{bot} = 0$ 
3: for  $i = 2$  to  $N_t - 1$  do
4:   if  $dds(i) < 0$  and  $sign(i) \neq sign(i + 1)$  then
5:      $p_{top} = p_{top} + 1$ 
6:      $t_{top}(p_{top}) = t(i)$ 
7:      $env_{top}(p_{top}) = s(i)$ 
8:   else if  $dds(i) > 0$  and  $sign(i) \neq sign(i + 1)$  then
9:      $p_{bot} = p_{bot} + 1$ 
10:     $t_{bot}(p_{bot}) = t(i)$ 
11:     $env_{bot}(p_{bot}) = s(i)$ 
12:   end if
13: end for

```

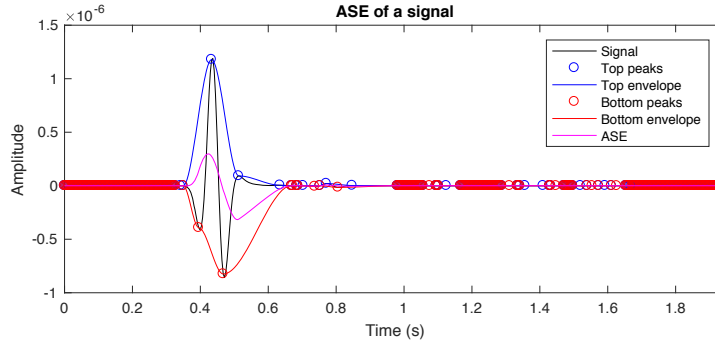


Figure 4.8: Spline envelopes and average spline envelopes for a given seismic trace (solid black). The top peaks (blue circle) and bottom peaks (red circle) are extracted using algorithms 1 and 2. From these minima and maxima, the top envelope (blue line) and bottom envelope (red line) are extracted using a piecewise cubic Hermite interpolating polynomial. The mean of these two envelopes, corresponding to the average spline envelope (purple line), presents a nice shape that could be seen as a low-frequency version of the original signal.

two levels of noise is presented on Figure 4.10. It can be observed that for noiseless data (Figure 4.10a) or data with a small amount of noise (Figure 4.10b), the ASE presents wider temporal support, which will translate into increased cycle-skipping robustness. We can even observe that the ASE common shot gather is slightly cleaner at zero offset for the signal with a small amount of noise. However, when strong noise is present (Figure 4.10c), the resulting signal is a denoised version, without the increased temporal support width. While this denoising effect is impressive on its own, it does not translate into improvement for cycle-skipping robustness.

The conclusion on ASE would be that the algorithms 1 and 2 present two main issues:

- It is not a differentiable operation, meaning that any attempt to derive an adjoint source from any misfit based on this operation will be unsuccessful.
- It relies on the addition of subtle white noise to stabilize the results. For synthetic data, there will be no issue. For real data with strong noise, the peaking will result in too many minima and maxima. When interpolated, this will only create a "denoise" version of the signal and not an envelope.

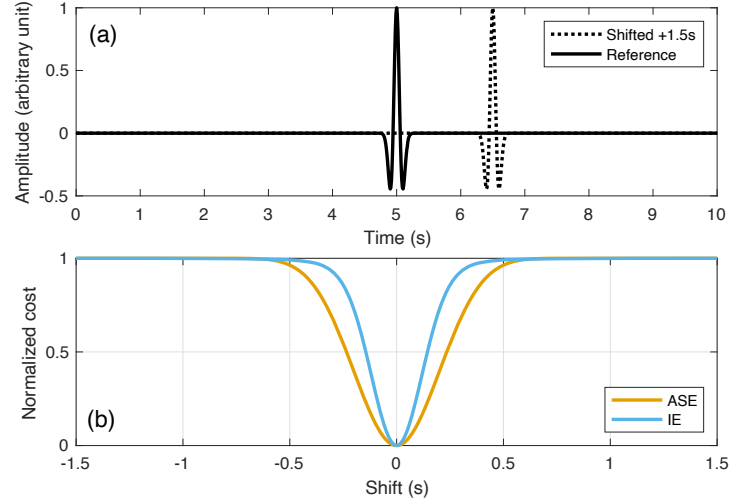
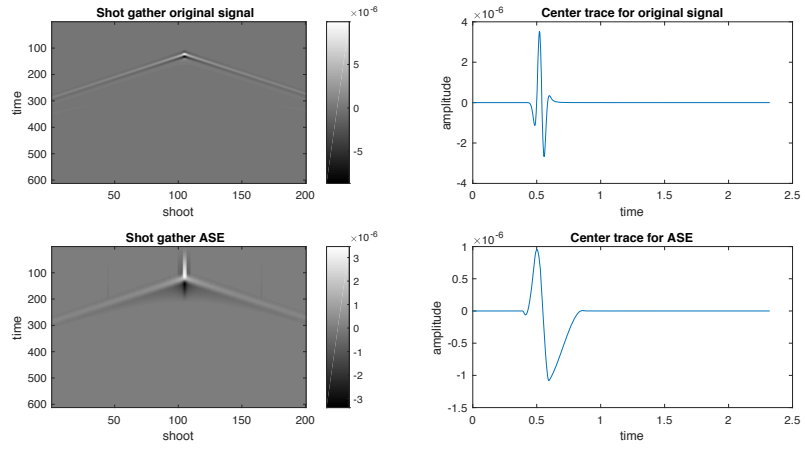
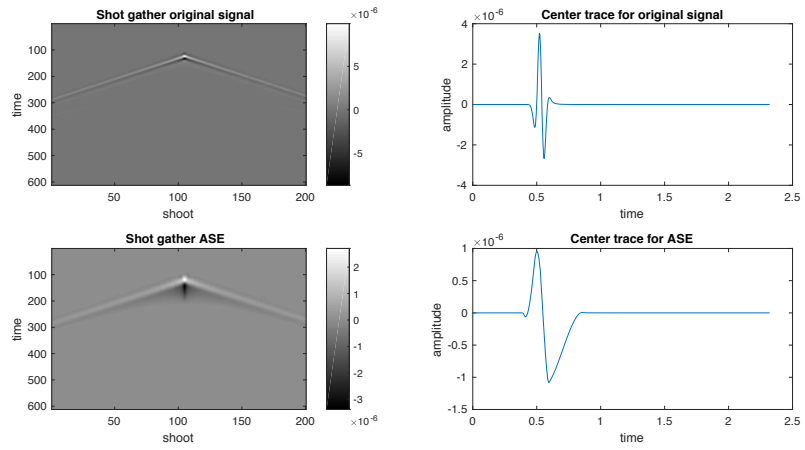


Figure 4.9: Simple time-shifted Ricker wavelet test for ASE misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the ASE and IE misfit functions with respect to time-shift.

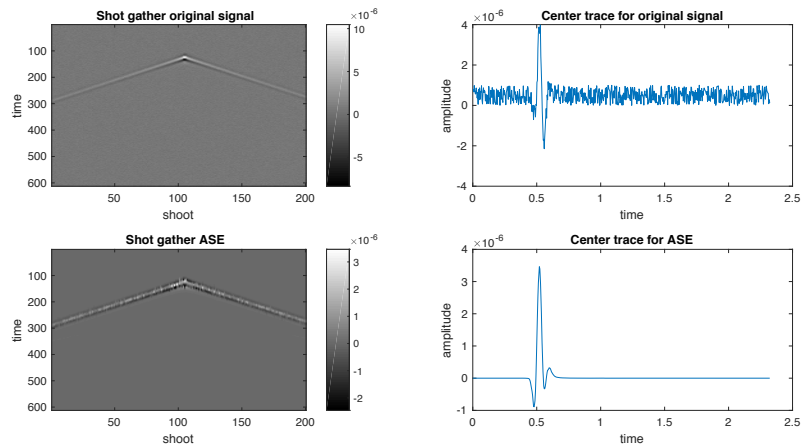
Peak extraction is not a differentiable operation: this is a dramatic drawback and prevents us from applying this misfit function to FWI when using the adjoint state method to obtain gradient. Still, ASE properties seem interesting, and from toy experiment on traces shows it could be promising. The strong denoising effect could also be of interest for FWI as a pre-processing step to clean out data strongly contaminated by noise.



(a) No noise added to the signal



(b) Very small noise added to the signal



(c) Strong noise added to the signal

Figure 4.10: Extraction of average spline envelope based using PCHIP for three levels of noise (a: without noise, b: small amount of noise, c: strong noise).

4.2.3 Diffusion based misfit function

Another idea explored for a new observable to improve the L^2 norm misfit function is a Diffusion-based observable. The simple idea is to take each trace and set it as the initial condition for a 1D diffusion problem solved by finite differences. By doing so, we hope to extract a “smooth” version of the observable with a wider temporal support. Since it is a diffusion problem, we can finely control and adjust the “diffusion time”, which allows for precise control on how much the signal is diffused and therefore smooth. At first, we can diffuse for a long time (until we obtained a smooth enough observable to mitigate the cycle-skipping issue), then during the non-linear iteration of the FWI process, slowly reduce the diffusion time to make the observable less diffused and closer to the original signal. Ideally, in the end, it will be a non-diffusion misfit that will be similar to the classical L^2 waveform misfit. This ability to smoothly control the widening of temporal support translate into control of the cycle-skipping robustness, which is very appealing.

Let us remind the formulation of the diffusion equation as

$$\frac{\partial u}{\partial t} + \kappa \frac{\partial^2 u}{\partial x^2} = 0, \quad (4.8)$$

where κ is the diffusion parameters.

In our case, since we do not care about the physical meaning of the diffusion we performed, we can set $\kappa = 1$. Since we set $d(t)$ as an initial space condition, we can rewrite the diffusion equation as

$$\frac{\partial d}{\partial \tau} + \frac{\partial^2 d}{\partial t^2} = 0. \quad (4.9)$$

The new quantity $D(t, \tau)$ is computed as the diffusion of the original signal $d(t)$ for a given time τ (the diffusion time). If $\tau = 0$, we do not diffuse and $D(t, \tau = 0) = d(t)$. The more τ increases, the more $D(t, \tau)$ will be a diffused version of $d(t)$.

Instead of considering the least-squares misfit (L^2 norm) between d_{obs} and d_{cal} , we consider the least-squares difference between D_{obs} and D_{cal} .

$$F_{diffusion}(\tau) = \frac{1}{2} \int_0^T [D_{cal}(d_{cal}, t, \tau) - D_{obs}(d_{obs}, t, \tau)]^2 dt. \quad (4.10)$$

where $D_{cal} = D(d_{cal}, t, \tau)$ and $D_{obs} = D(d_{obs}, t, \tau)$ are modified versions of the wavefield $d_{cal}(t)$ and $d_{obs}(t)$ based on the diffusion equation for a given diffusion time τ .

An illustration of the effect of diffusion on a signal is visible on Figure 4.11. This clearly illustrates the increased temporal support for the diffused version compared to the original signal. Also, it shows the ability to control “how much” increased temporal support is possible by modifying the diffusion “time”.

This potentially translates into increased convexity of the misfit function with respect to time-shift as shown in Figure 4.12.

This behavior is further illustrated on Figure 4.13, with two Ricker wavelets, one unfiltered and one high passed above 2 Hz. For the unfiltered Ricker wavelet, we can see a clear improvement in the temporal support width, which is the feature we want to observe. However, for

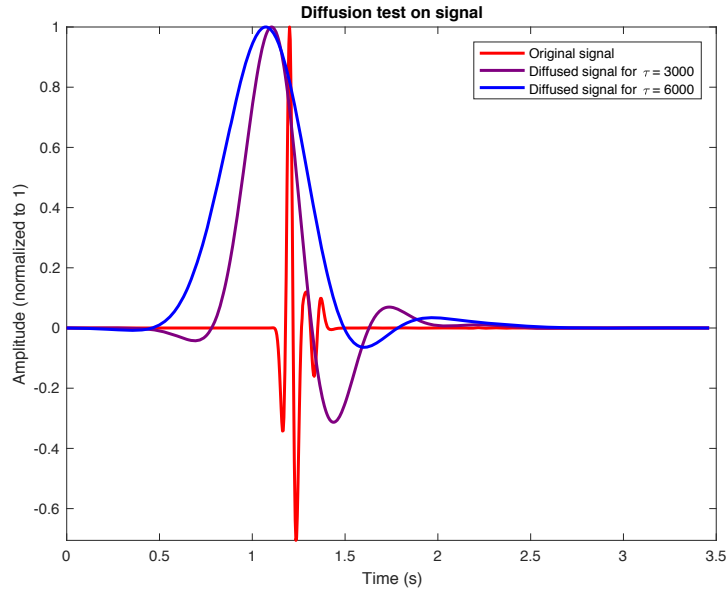


Figure 4.11: Test of diffusion on a simple signal (solid red) for two diffusion values. The less diffused version (solid purple) and the two times more diffused version (solid blue).

the highpass filtered version, it can be observed that the signal shape is not preserved, while the associated spectrum is still good and coherent. This raises concern with the suppression of zero-frequency information. To push the analysis deeper, noise can be added as presented on Figure 4.14. Here, it clearly illustrates that adding white noise to the data starts to interfere with the diffused version and altered the signal. To validate this hypothesis, we observed the effect of diffusion on a field-data trace (Figure 4.15). It can be observed that the spectrum of the diffused data is not clean and does not represent a pure shift down in frequency. This behavior can be linked to the fact that diffusion is expected to behave like a linear low-pass filter, meaning it amplifies information from the pre-existing low frequencies in the data. When looking at a high-passed Ricker wavelet or a field data (which contains slight noise and no zero frequency information), it can be observed that there is no added benefit of using diffusion to create a new observable.

This idea illustrates the potential false “good idea” which can be appealing at first while presenting no advantages to field data application. It is a good reminder that playing with synthetic data and especially Ricker wavelet (which contains close to zero frequency information) can introduce a skew conclusion. This kind of artificial boosting of a Ricker wavelet’s low-frequency content has been observed some times on some ideas presented during conferences and should raise a warning.

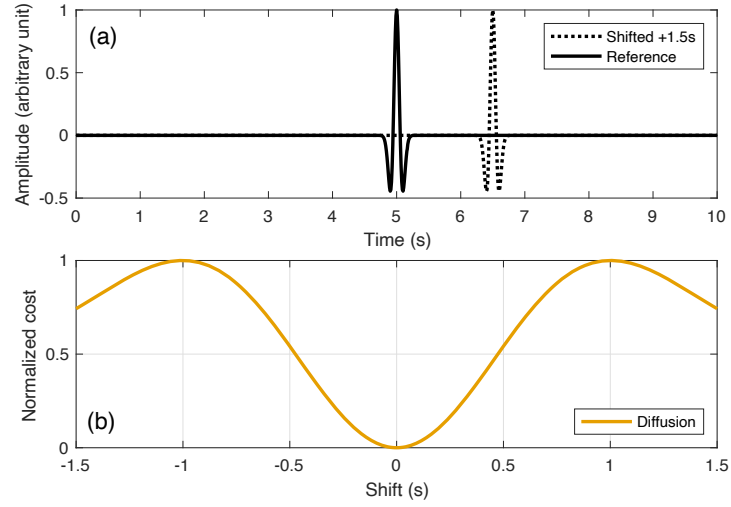


Figure 4.12: Simple time-shifted Ricker wavelet test for crosscorrelation misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the diffusion-based misfit function with respect to time-shift. Here the diffusion time τ is set to 3000.

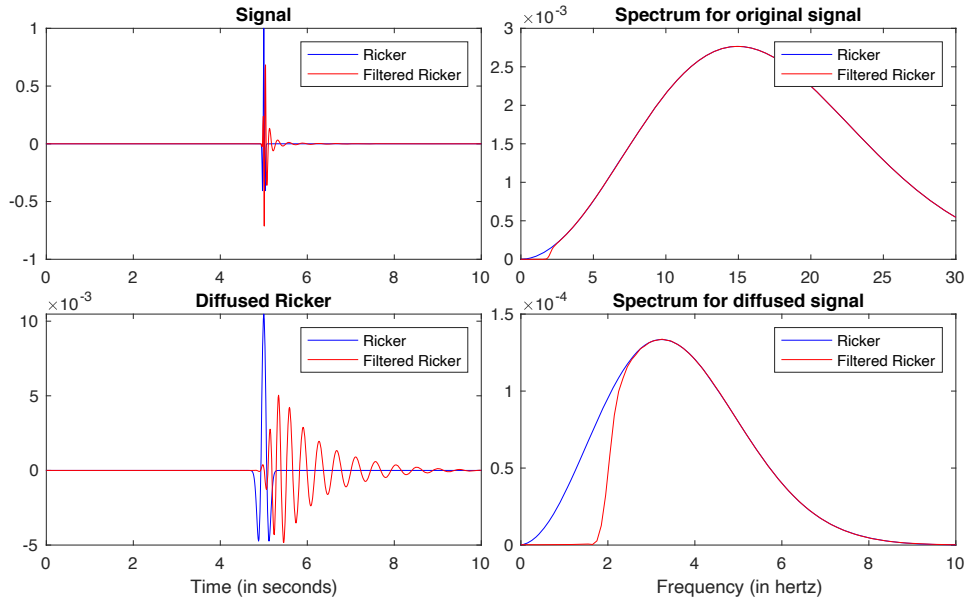


Figure 4.13: Test of diffusion on two Ricker wavelets, one unfiltered (solid blue) and one highpass filtered above 1.5 Hz (solid red). The right column corresponds to the associated spectrum. The top line represents the non-diffused signal while the bottom line represents signals after diffusion.

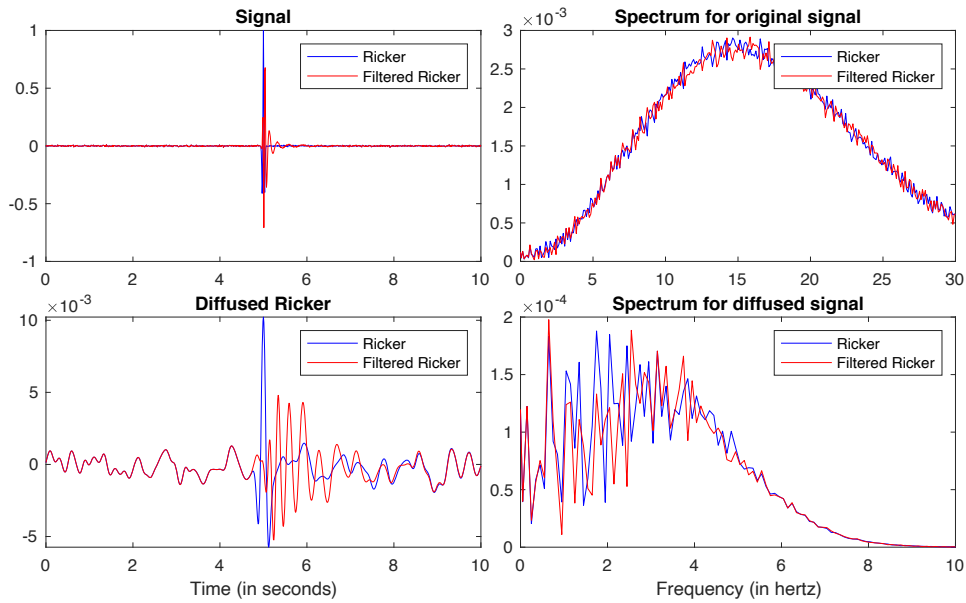


Figure 4.14: Same setup as Figure 4.13 but with noise added on the signal.

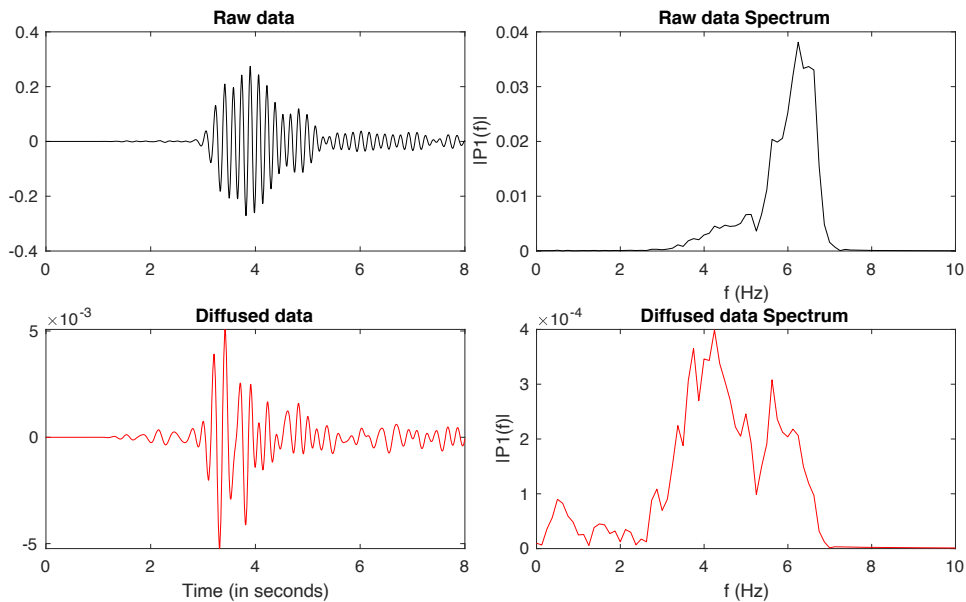


Figure 4.15: Test of diffusion on a field data trace coming from Valhall from the 2.5 to 5 Hz frequency band. Top line corresponds to the raw data and associated spectrum. Bottom line corresponds to the diffused data and its associated spectrum.

4.2.4 Beat-tone misfit function

Finally, another explored idea is to generate a beat tone using a cosine function on the data. To do so, we multiply our signal $d(t)$ in time with a cosine at a given frequency f_c :

$$B(t) = d(t) * \cos(2\pi f_c t) \quad (4.11)$$

If we have a signal of a given central frequency f , the beat generated should have a high and a low-frequency component such as $f - f_c$ and $f + f_c$. We then have to remove the high-frequency component by applying a low pass filter, for example keeping only the frequencies under the central frequency f of the original signal. This would theoretically allow shifting the frequencies of the signal toward the low frequency. This behavior is illustrated on Figure 4.16 using a simple Ricker wavelet at 10 Hz. In this case, introducing a carrier at $f_c = 2$ Hz results in a split of the frequency content, with a part around 5 Hz on the low-frequency side and one around 15 Hz for the high-frequency side. After applying a low-pass filter under 7 Hz we obtain a low-frequency version of our original signal, which looks appealing.

Again, this should theoretically translate into increased convexity of the misfit function as illustrated in Figure 4.17.

To further illustrate the effect, we apply beat tone with $f_c = 4$ Hz on a trace from field data from Valhall as presented on Figure 4.18. The signal is then low-passed with a maximum cut-off frequency of 5 Hz. The original signal central frequency is around 6 Hz, and after applying the beat tone, we again have a split of the spectrum with peaks around 2.5 and 10 Hz. After the low-pass filter, we obtain a low-frequency version of the original signal. However, one of the first issues would be the sensitivity to the carrier frequency f_c as illustrated on Figure 4.19. Here, with $f_c = 12$ Hz, no beat tone effect is present. There is no “automatic and easy” way of finding the proper carrier frequency to optimize the beat tone effect from the experiment. Moreover, a second issue is also present and corresponds to the very unpredictable behavior of beat tone as it depends on the cosine phase. This effect is illustrated in Figure 4.20. A slight shift in time could lead to drastic modification of the beat tone version. This unstable behavior will effectively lead to a basin of attraction as large as the classical L^2 misfit.

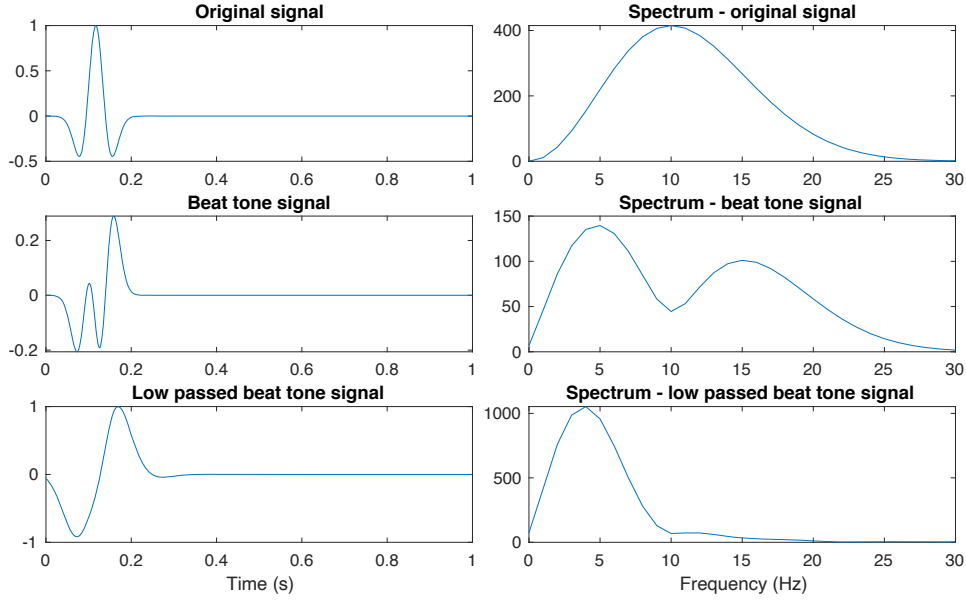


Figure 4.16: Beat tone illustration on a simple unfiltered Ricker wavelet and associated spectrum. First line corresponds to the raw Ricker wavelet, the second line to the beat tone version and the third line to a low-passed version of the beat tone version to remove high-frequency content.

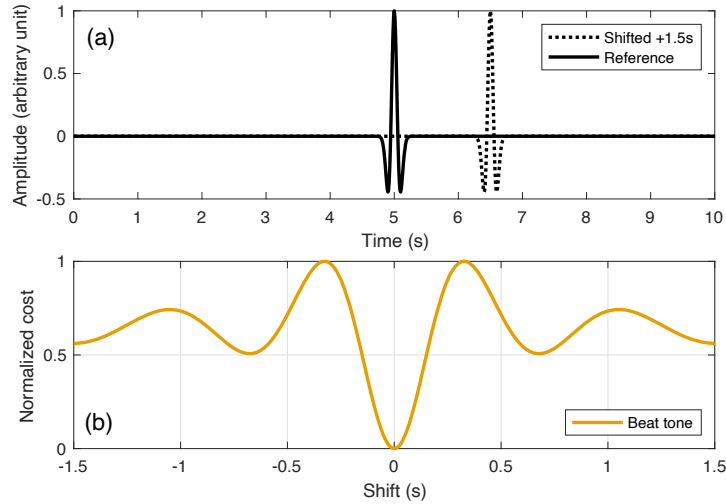


Figure 4.17: Simple time-shifted Ricker wavelet test for crosscorrelation misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the beat-tone misfit function with respect to time-shift. Here $f_c = 0.5$ Hz and the low-pass filter is set to 1 Hz.

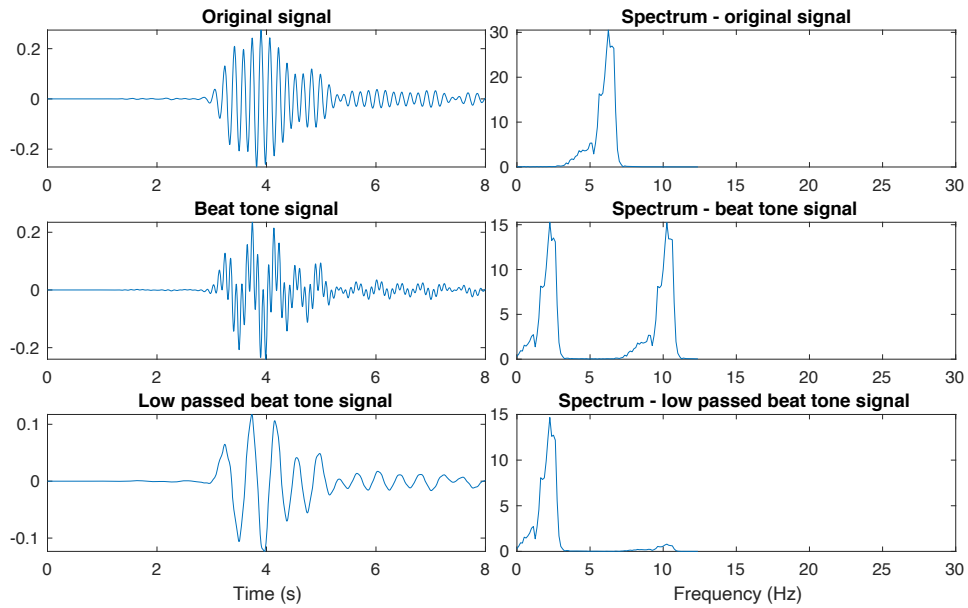


Figure 4.18: Same setup as on Figure 4.16 but using a field data trace coming from Valhall instead of the Ricker wavelet.

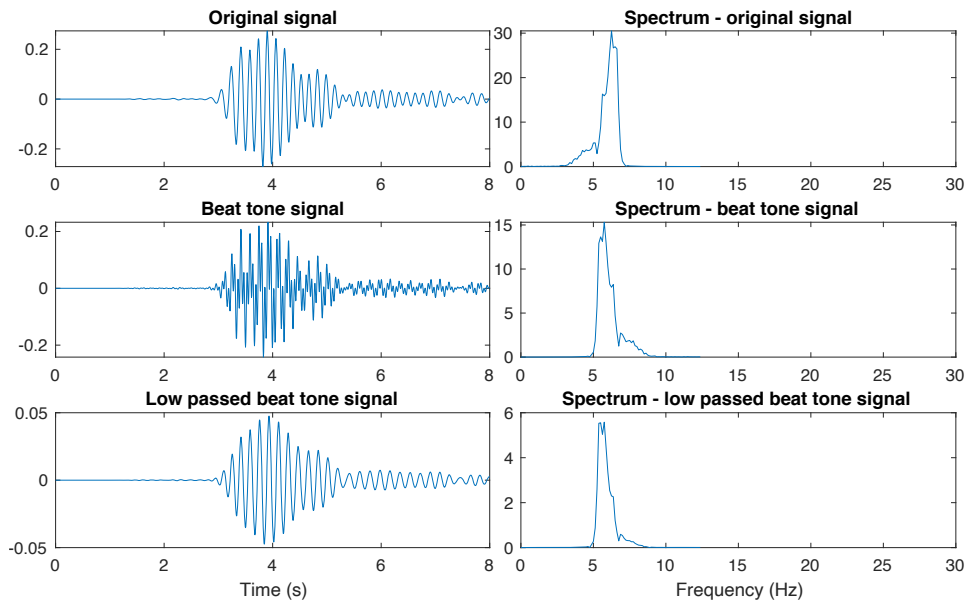


Figure 4.19: Same setup as on Figure 4.18 but with a less effective parameterization of the beat tone resulting in no shift toward the low frequencies.

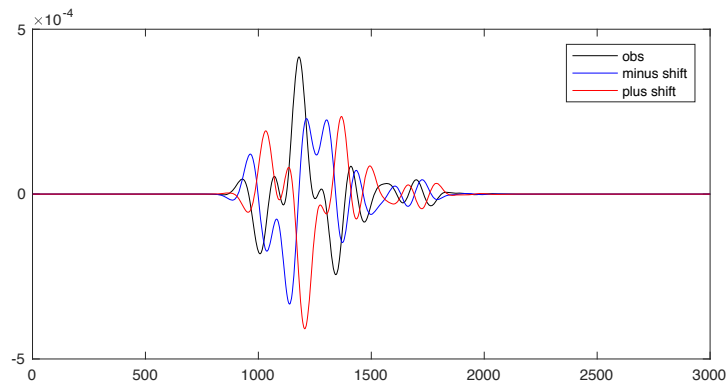


Figure 4.20: Exemple of the variability of beat tone on a Valhall field trace. The reference beat tone version is in solid black. In solid blue and red are two beats obtained from a time-shifted version of the observed data. We can observe here that as the carrier is always the same, a possible shift in time of the wavelet will dramatically change the signal shape.

4.3 Investigation published during my thesis

Here, I want to present a list of the investigations I have performed and published during my thesis that contains meaningful insights on robust FWI and alternative misfit functions. Some of the ideas are presented in-depth in this thesis, but some were put aside for the clarity of the manuscript. I would refer interested readers to my work for broader pictures on robust FWI and alternative misfit functions.

- First, my Master's thesis, focusing on improving FWI robustness with alternative misfit functions: Pladys, A. (2016). Sensibility of several misfit functions for Full Waveform Inversion: theoretical analysis, implementation and comparative study on multiple realistic cases. Master's thesis, Univ. Grenoble Alpes.
- Some work on key properties that alternative misfit functions should satisfy: Pladys, A., Brossier, R., and Métivier, L. (2017). FWI alternative misfit functions: what properties should they satisfy. In *79th EAGE Conference and Exhibition 2017, Paris*, page Tu P1 01.
- Case study for GSOT-based FWI on Valhall: Pladys, A., Brossier, R., Irnaka, M., Kamath, N., and Métivier, L. (2019). Assessment of optimal transport based FWI: 3d OBC valhall case study. In *SEG Technical Program Expanded Abstracts 2019*, pages 1295–1299.
- Improving cycle-skipping robustness with GSOT-based FWI on Valhall: Pladys, A., Brossier, R., and Métivier, L. (2020). Graph space optimal transport based FWI: 3D OBC valhall case study. In *SEG Technical Program Expanded Abstracts 2020*.
- Published paper on alternative misfit functions for FWI: Pladys, A., Brossier, R., Li, Y., and Métivier, L. (2021b). On cycle-skipping and misfit functions modification for full-wave inversion: comparison of five recent approaches. *Geophysics*, in press.
- Submitted paper on 3D OBC field data application of GSOT-based FWI: Pladys, A., Brossier, R., Kamath, N., and Métivier, L. (2021a). Robust fwi with graph space optimal transport: application to 3d obc valhall data. *Geophysics*, submitted.

I also wanted to express my regrets for not sharing my work performed on a 3D OBC North Sea dataset in collaboration with Equinor. Unfortunately, the ownership of the dataset was transferred during my thesis to another company, and all the authorizations were lost. This was my first time getting around field data, and the application focused on 1D KROT analysis, then on a multi-2D KROT analysis.

Conclusion and perspectives

Conclusion

This thesis manuscript aims at comparing alternative misfit functions for full-waveform inversion. In the beginning, a general introduction of the history of seismic imaging and, more generally, imaging in geophysics is performed. This introduces the different methods developed along the year and introduces the method at the core of this manuscript: full-waveform inversion (FWI). The pros and cons of this method are developed and make a transition toward the first Chapter. In Chapter 1, a generic introduction to the FWI formalism is made, from the forward modeling part of the problem to the inversion part with the multiple formalisms that can be used to solve inverse problems. The non-convexity issue is introduced, followed by a presentation of the cycle-skipping issue. An introduction to possible solutions to mitigate FWI non-convexity is then performed. In Chapter 2, alternative misfit functions are benchmarked on several synthetic tests, from simple 1D examples to more complex 2D FWI cases to assess the pros and cons of each proposition. In Chapter 3, an application of the most promising method (graph space optimal transport) is applied to a 3D field dataset. Finally, in Chapter 4, some ideas for other alternative misfit functions and some workaround performed for practical applications are presented.

Throughout this work, several important points have been discussed, and I want to re-emphasize them as a conclusion:

- **Is robustness to cycle-skipping the only important feature of an alternative misfit function?**

As introduced in Chapter 1, the primary and central issue of FWI comes from the non-convexity of the misfit function conventionally used: the least-squares norm (L^2 norm). This L^2 norm is at the origin of the cycle-skipping issue. Therefore, when the idea came of replacing the L^2 norm with alternative misfit functions, it seemed obvious - at first - to focus on reducing the non-convexity of the misfit function. This led to the proposition of many more convex (with respect to time-shifts) misfit functions. This should have *theoretically* solved the cycle-skipping issue and improve FWI robustness. Nonetheless, the number of field data applications using alternative misfit functions remains small compared to the number of proposed solutions.

This discrepancy is, for me, explained by the fact the L^2 norm is pretty good for all other aspects: it preserves signal shape (no alteration in amplitude/phase information), it exhibits strong noise robustness, it - when not affected by cycle skipping - delivers very high-resolution results that FWI is known for and finally, the computational cost

is negligible. These advantages have made the L^2 norm the “way to go” for FWI. By looking at cycle-skipping robustness alone while not preserving the other important *key features* that a misfit function should exhibit, propositions of alternative misfit functions can reveal themselves not adapted to realistic FWI usage (such as field data), even if they reduce cycle-skipping sensitivity “on paper”.

- **How to assess the *real world* capacity of an alternative misfit function?**

Assessing misfit functions in the most *generic* and *reproducible* way is important. The literature misses much cross-comparison between several propositions. As an attempt to address this issue, I proposed several synthetic cases (presented in Chapter 2). From 1D *toy* example that only focuses on shifted patterns, to simple 2D FWI tests in transmission and reflection, to finally more realistic 2D cases with Marmousi, with a shared parameterization between misfit functions, cross-comparison is made easier and understanding potential drawbacks of a given formulation is made simpler. This allowed us to restrict the potential good down to what is for us the best candidate and further test it on field data in Chapter 3.

If an alternative misfit function is directly tested on a more realistic case, it can be easily tweaked to provide artificially good results, or it can fail and be difficult to identify why. I want to promote a more systematic approach to validate alternative misfit functions capacity and hope this work is a step in this direction.

- **Are alternative misfit function some kind of *magic button* that could make FWI automatic?**

After extended tests, using what is for me the best misfit function at my disposal, from simple 2D FWI setup to 3D field data application, it never behaves as a “magic button”. Even if, on simple synthetic cases, the increase of robustness that graph space optimal transport (GSOT) provides can simplify the workflow or allows to start from really crude initial models, extended tests have shown that it still requires the use of appropriate workflows when trying to tackle more complex situations. This is not a drawback, as in such cases, L^2 -based FWI simply fails to provide results even when using all the possible approaches. For me, GSOT is a clear improvement to FWI robustness, which constantly (for the tests I performed) improves results over L^2 , but at a given computational overhead. FWI remains a complex imaging process that is far from *automatic*, and providing tools to increase its robustness is a step in the good direction but not a magic leap toward a completely robust tool.

Perspectives

This work can be separated into two parts. The first one is a presentation of several misfit functions with an attempt at cross-benchmarking them. The analysis can be pushed forward on multiple aspects:

- **Non-exhaustive selection of misfit functions.**

Indeed, the selection of misfit functions that are benchmarked together is non-exhaustive. Moreover, we compared some of the best challengers to some simpler propositions, which were never pushed further on. This was made as an attempt to exhibit pros and cons

and to create more systematic ways of comparing misfit functions for FWI. By focusing on time-domain FWI formulation and only on misfit functions, we restricted ourselves from possible good candidates who rely on other kinds of reformulation of the FWI problem. Other misfit functions could be considered and benchmarked against our selection using the same framework.

- **Exploring and benchmarking other types of solutions.**

In this thesis, my focus has never been on extension strategies. Nevertheless, it could be of interest to compare them with alternative misfit functions directly. Making a direct comparison of alternative misfit function to extension strategies could bring new insight into robust FWI.

The second part of this work consists of a field data application comparing GSOT to L^2 norm. Here, several possibilities should be considered to improve the results and push the analysis forward:

- **Considering other misfit functions for field data.**

From the analysis performed on several misfit functions, we decided to focus only on the most promising candidates at our disposal. Moreover, being the solution developed in the team, we had the greatest confidence with it, which is important when tackling complex and large datasets that require *substantial* computational resources. For the amount of time at my disposal, pushing the analysis on the best candidate make sense. Future works, comparing other good alternative misfit functions, such as AWI, would bring more insight into their behaviors when facing field data.

- **Improving results using complementary methods.**

As it was shown, results obtained on the Valhall dataset when starting from the 1D model were impressive. Still, some imperfections in the shallow layer reconstruction are visible compared to the results obtained from the reflection tomography initial model. These kinds of issues could be tackled with other methods such as RWI or JFWI (Zhou, 2016; Zhou et al., 2016; Brossier et al., 2017; Zhou et al., 2018; Li et al., 2019; Provenzano et al., 2020b; Yao et al., 2020). Also, considering a more complex physic for the modeling could help explain the data better. This would provide an exciting insight into the capacity of alternative misfit functions for more complex workflow.

- **Pushing inversion further.**

The last line of investigation would be to push the inversion further. The choice of stopping at 7 Hz was mainly motivated by computational resources at our disposal. This link with the previous point and pushing the FWI workflow further will bring more knowledge on alternative misfit function behavior on field data.

I see this work as a stepping stone to encourage more rigorous and systematic analysis and cross-comparison of alternative misfit functions for FWI. Pushing FWI forward by increasing its robustness can only make this method more appealing, while unlocking targets that are still out of reach for now because of lacking good initial model. We show that field data applications using one of the last and most promising methods (GSOT) is feasible and yield very encouraging results and hope that FWI will continue moving forward with more interesting applications in the years to come.

List of Figures

1	(a) Some of the possible ray paths for seismic waves passing through the Earth (from Fowler et al., 1990). (b) Jeffreys-Bullen (J-B) travel-times curves for an earthquake focus at the surface. Travel-times are represented as a function of epicentral distance. Each line represents a given travel path through the Earth's interior (after Bullen and Bolt, 1985).	2
2	Schematic of an onshore and offshore setup in seismic exploration surveys (Energy Information Australia).	4
3	Refraction/reflection FWI results at exploration scale. (a) correspond to the initial model used to start FWI workflow. (b) final 25 Hz FWI results. From Dickinson et al. (2017).	4
4	Small scale FWI results at 65 Hz. Here the target is the Ettlingen trench near Karlsruhe, Germany. The low-velocity anomaly in the center corresponds to fill-in materials used to level the ancient trench. From Irnaka, 2021.	5
5	Example of global FWI for vertically polarized shear-wave-speed perturbations in GLAD-M25 at 250 km depth (Lei et al., 2020). Remarkable hotspots are highlighted with white arrow.	5
1.1	Schematic illustration of the accuracy and resolution of tomography, full-waveform inversion and migration methods. The solid black line corresponds to the resolution gap between tomography and migration technics (as proposed by Claerbout, 1985). This gap is supposed to be filled by full-waveform inversion (figure from Yao et al., 2020).	10
1.2	Couple of source/receiver of a fixed diffraction point in a homogeneous medium.	10
1.3	Representation of misfit function $f[m] : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ presenting three local minima, two of which are global and denoted by a red dot and a third one denoted by a blue dot. From Thurin (2020).	19
1.4	Schematic example of the cycle-skipping/phase-ambiguity issue on sinusoidal signals. As soon as the initial shift is larger than half a period, the signal's fit using a least-squares distance is performed up to one or several phase shifts. One may try to fit the $n+1$ dashed wriggle of the top signal with the n continuous wriggle of the middle signal moving in the wrong direction. The bottom dashed signal predicts the n wriggle in less than a half-period leading to a correct updating direction (from Virieux and Operto, 2009).	24

LIST OF FIGURES

1.5	Simple time-shifted Ricker wavelet (4 Hz) test for L^2 misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the L^2 norm with respect to time-shift.	25
1.6	FWI schematic example for cycle skipping: (a) true model with spherical inclusion, (b) homogeneous initial model 1, (c) homogeneous initial model 2.	25
1.7	FWI schematic example for cycle skipping: traces for a source-receiver couple at $z = 500$ m for d_{obs} (solid black), d_{cal} in the initial model (dashed black) and d_{cal} in final reconstructed V_P model (solid red). (a) correspond to initial model 1, (b) to initial model 2.	26
1.8	FWI schematic example for cycle skipping: reconstructed V_P models using FWI starting from (a) homogeneous initial model 1, (b) homogeneous initial model 2.	26
1.9	Same 1D setup as Figure 1.5, but this time comparing the influence of frequency on the convexity. (a) 0.5 Hz Ricker wavelet, (b) 2 Hz Ricker wavelet, (c) 4 Hz Ricker wavelet. (d) L^2 misfit function values with respect to time-shifts for the three different frequency cases. The width of the basin of attraction of the misfit function is enlarged for lower frequencies.	27
1.10	Illustration of frequency-continuation approach used in FWI; here, on a shallow near-surface target of 30 m (from Irnaka, 2021). FWI is started on a 3 to 25 Hz frequency band, then higher frequency are progressively introduced with an upper limit of 45 Hz then finally 65 Hz. A clear improvement in resolution (smaller wavenumber content) is observed for higher frequency data.	27
2.1	Comparison of several misfit functions in a simple 1D case for one shifted arrival. The arrival is set to be a Ricker wavelet with a central frequency of 4 Hz. (a) represents the signal used for the test (with only one arrival at the center). The fixed reference signal is displayed in continuous black. The shifted signal is displayed in dotted black (here for +1.5 s). (b) represents the normalized misfit function values with respect to the time-shift (from -1.5 s to 1.5 s).	43
2.2	(a) quantities Q_{obs} and Q_{cal} for three time-shifts (-1.5 s in green, -0.1 s in blue and “in phase” in dashed red). (b) the area under the curve for $Q_{obs} - Q_{cal}$ for the three time-shifts.	43
2.3	E_{obs} and E_{cal} for three time-shifts (-0.5 s, -0.2 s and in phase).	43
2.4	Detail for $\bar{\varphi}(t)$ from KROT. (a) the setup with d_{obs} (in black) and d_{cal} for two time-shifts (-0.3 s in red and -0.15 s in blue). (b) shape of L^2 and KROT misfit function with respect to time-shifts, red and blue cross represent the positions of the two time-shifts selected. (c) and (d) respectively display $\bar{\varphi}(t)$ and $d_{obs} - d_{cal}$ for the two time-shifts of -0.3 s and -0.15 s. (e) the area under the curve for $\bar{\varphi}(t)(d_{obs} - d_{cal})$ quantity for the two time-shifts. This last quantity is used to get the misfit function value after time integration.	44
2.5	Same as Figure 2.1 but with two Ricker wavelets with one shifted (left) and one in phase (right).	44

2.6	AWI analysis with two setups: test A and test B. (a) test A both wavelets shift, (b) test B only the left wavelet is shifted, the right one being always in phase (similarly to Figure 2.5). (c) shape of AWI misfit function with respect to time-shift in both cases. The Wiener filters presented under are shown for a time-shift of -0.5 s (black cross on the misfit). (d) Wiener filters ($w(t)$) and the Gaussian penalty function $\mathcal{P}(t)$. (e) the Wiener filters multiplied by the penalty function.	45
2.7	FWI Test 1: (a) true model, (b) initial model 1 with $V_P = 1300$ m.s $^{-1}$, (c) initial model 2 with $V_P = 1700$ m.s $^{-1}$ and (d) initial model 3 with $V_P = 1900$ m.s $^{-1}$.	48
2.8	FWI Test 1: V_P results from FWI. First line corresponds to initial model 1, second line to initial model 2 and third line to initial model 3. Each column corresponds to reconstructed V_P model from FWI using respectively L^2 (a-c), AWI (d-f), IE (g-i), NIM (j-l), KROT (m-o) and GSOT (p-r) misfit functions.	48
2.9	FWI Test 1: Extracted traces along the shortest path (horizontal straight line at 500 m depth passing through the spherical inclusion). Observed data are in solid black, synthetic data in the initial model in dashed black and final reconstructed synthetic data in solid red. First column corresponds to initial model 1, second column to initial model 2 and third column to initial model 3. Each line corresponds to reconstructed V_P model from FWI using respectively L^2 (a-c), AWI (d-f), IE (g-i), NIM (j-l), KROT (m-o) and GSOT (p-r) misfit functions.	49
2.10	FWI Test 2: (a) Homogeneous initial model ($V_P = 1500$ m.s $^{-1}$). (b) True model with a negative V_P anomaly, (c) true model with a positive V_P anomaly.	52
2.11	FWI Test 2: 2D common shot gathers for (a) observed data for negative V_P anomaly and (b) positive V_P anomaly. Strong clipping is applied to enhanced the reflected waves which are approximately 100 times smaller in amplitude than the transmitted waves. Polarity reversal of the reflected waves is pointed with black arrows.	53
2.12	FWI Test 2: FWI layer benchmark results. Left column corresponds to a negative V_P anomaly while the right column to positive V_P . The subfigures under respectively correspond to final reconstructed V_P model obtained with FWI using L^2 (a,b), AWI (c,d), IE (e,f), NIM (g,h), KROT (i,j) and GSOT (k,l).	54
2.13	FWI Test 2: Extracted traces for a center shot at zero offset. Observed data are in solid black, synthetic data in the initial model in dashed black, and final reconstructed synthetic data in solid red. The left column corresponds to a negative V_P anomaly while the right column to positive V_P . (a,b) shows the complete traces. Each subfigures under are cropped on the magenta box to emphasize the polarity reversal introduced by the layer. They correspond to traces calculated in the reconstructed V_P model using respectively L^2 (c,d), AWI (e,f), IE (g,h), NIM (i,j), KROT (k,l) and GSOT (m,n) misfits functions.	55
2.14	FWI Test 2: Extracted traces for (a) d_{obs} and d_{cal} , (b) E_{obs} and E_{cal} , and (c) Q_{obs} and Q_{cal} for center shot at zero offset. Data associated to negative V_P anomaly are in solid blue, and positive V_P anomaly in solid red. Calculated data is in solid black. Magenta box corresponds to the enlarged area.	56

2.15	FWI Test 2: (a) Extracted traces d_{obs} , (b) Wiener filters $w(t)$ and \mathcal{P} , and (c) $w(t) \times \mathcal{P}$ for center shot at zero offset at first iteration. Data associated to negative V_P anomaly are in solid blue, and positive V_P anomaly in solid red.	56
2.16	FWI Test 2: Misfit value for L^2 and AWI with respect to the layer velocity for the two inversion cases: (a) negative and (b) positive V_P anomaly. Misfit is calculated between d_{obs} (data in the true medium) and a $d_{cal}(V_P)$ generated in a medium with a correct background $V_P = 1500 \text{ m.s}^{-1}$ and a as only varying parameter the layer V_P , ranging from $\pm 100 \text{ m.s}^{-1}$ around the original velocity of layer in d_{obs}	57
2.17	Test 3.a: (a) True Marmousi II model. (b) $\mathcal{S}250$ initial model, (c) $\mathcal{S}500$ initial model and (d) 1D initial model.	59
2.18	Test 3.a: Density model obtained from V_P using Gardner's law for: (a) True Marmousi II model, (b) $\mathcal{S}250$ initial model and (c) $\mathcal{S}500$ initial model.	59
2.19	Test 3.a: Common shot gather (CSG) for field data overlapped by synthetic data in (a) $\mathcal{S}500$ initial model, and (b) 1D initial model. Field data in black and white, overlapped by red to blue synthetic data with transparency. Red and white visible mean out of phase, black and blue mean in phase.	63
2.20	Test 3.a: Inverse crime FWI final reconstructed V_P model for Marmousi. Left column corresponds to $\mathcal{S}500$ initial model, right column to 1D initial model. The lines respectively correspond to the final reconstructed V_P model using L^2 (a,b), AWI (c,d), IE (e,f), NIM (g,h), KROT (i,j) and GSOT (k,l).	64
2.21	Test 3.a: Overlapped common shot gathers for synthetic data in the final reconstructed V_P model starting from $\mathcal{S}500$ initial model vs field data. Each subfigure corresponds to misfit function, with L^2 (a), AWI (b), IE (c), NIM (d), KROT (e), and GSOT (f).	65
2.22	Test 3.a: Same as Figure 2.21 but starting from 1D initial model.	66
2.23	Test 3.a: Costs evolution in the inverse crime Marmousi for 1D initial model. (a) evolution of cost functions over iterations, (b) true L^2 cost evolution over iterations, (c) model error reduction over iterations and finally (d) model error vs. the data error reduction.	67
2.24	Test 3.b: Comparison of true wavelet used to generate the observed data (for both test 3.a and 3.b) versus the inverted wavelet used for inversion in more realistic test 3.b. Under the associated amplitude spectrum of both true wavelet and estimated wavelets.	69
2.25	Test 3.b: More realistic FWI final reconstructed V_P model for Marmousi. Left column corresponds to $\mathcal{S}250$ initial model, right column to $\mathcal{S}500$ initial model. The lines respectively correspond to the final reconstructed V_P model using L^2 (a,b), AWI (c,d), IE (e,f), KROT (g,h) and GSOT (i,j).	69
2.26	Test 3.b: Overlapped common shot gathers for synthetic data in the final reconstructed V_P model starting from $\mathcal{S}250$ initial model vs field data. (a) corresponds to the data-fit in the $\mathcal{S}250$ initial model. Then, each subfigure corresponds to misfit function, with L^2 (b), AWI (c), IE (d), KROT (e), and GSOT (f).	70
2.27	Test 3.b: Same as Figure 2.26 but starting from $\mathcal{S}500$ initial model.	71

2.28	Test 3.b: Costs evolution in the more realistic Marmousi for $S500$ initial model. (a) evolution of cost functions over iterations, (b) true L^2 cost evolution over iterations, (c) model error reduction over iterations and finally (d) model error vs. the data error reduction.	72
3.1	Location of the Valhall field on the North Sea (from Thurin, 2020)	82
3.2	Layout of the Valhall acquisition overlapped on an horizontal V_P slice at 1 km obtained by GSOT-based FWI (from this study). Location of sources (gray dots) and receivers (blue diamonds). Two receivers positions (A and B) are located with black stars. Cables A ($x = 2950$ m), B ($x = 5530$ m) and C ($x = 3080$ m) are identified. Black dots denote the position of the three V_P well sonic-logs. . .	83
3.3	2D common-receiver gathers extracted for receiver A along cable A for: raw data (top), band 1 data 2.5 – 5 Hz (middle) and band 2 data 2.5 – 7 Hz (bottom). White dashed arrows point on the Schölte waves which are muted for the inversion. Blue and yellow arrows respectively point on the reflexion from the low velocity zone and the reflexion on the top of the reservoir. Black arrows point on the diving waves.	84
3.4	Slices of the initial model V_P TOMO . (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.	86
3.5	Same as Figure 3.4 for initial model V_P 1D	87
3.6	Same as Figure 3.4 for anisotropic parameter η	88
3.7	2D common-receiver gathers at 5 Hz. Synthetic data generated into the initial model TOMO are displayed in a blue/white/red color scale, field data are overlapped in grayscale with transparency. The best result is achieved when black and blue are the only colors visible. Red and white are shown when data are not in phase. (a) receiver A along cable A (through the low velocity anomaly). (b) receiver B along cable B.	89
3.8	Same as Figure 3.7 into the 1D initial model.	90
3.9	Complete FWI workflow used on the OBC Valhall dataset. At the core of the process lies the FWI iterations. Then several loops are nested one into another, from internal FWI iterations, source-subsampling, data selection, to finally the outer one of frequency continuation.	91
3.10	On top, weighted data for source inversion displayed on a 5 Hz 2D common-receiver gather (receiver A cable A). On the bottom, the estimated wavelet (left) and associated spectrum (right).	92
3.11	2D common-receiver gathers extracted for receiver A along cable A with different data weighting applied on them. From top to bottom: first break & short offset (DW SO), first break & medium offset (DW MO), first break & full offset (DW FO), first time release & medium offset (RT1 MO), first time release & full offset (RT1 FO), second time release & full offset (RT2 FO).	93

LIST OF FIGURES

3.12	Slices of the 7 Hz FWI reconstructed V_P using L^2 misfit function starting from TOMO initial model. (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.	95
3.13	Same as Figure 3.12 using GSOT misfit function.	96
3.14	Comparison of V_P profiles extracted from the TOMO initial model (dashed red), 7 Hz FWI models using L^2 (solid yellow) and 7 Hz FWI model using GSOT (solid purple) with sonic log (solid black). Left subfigure corresponds to the Log 1 at the center of the target. Middle subfigure to the Log 2, and right subfigure to Log 3 (which is far away from the target).	97
3.15	2D common-receiver gathers at 7 Hz starting from the TOMO initial model. Synthetic data (blue/white/red color scale) generated into the final reconstructed V_P using (a,b) L^2 misfit function, (c,d) GSOT misfit function. (a,c) receiver A along cable A (through the low velocity anomaly). (b,d) receiver B along cable B. Field data are overlapped in grayscale with transparency. Black arrows point to improvement obtained with GSOT.	98
3.16	2D common-receiver gathers at 5 Hz starting from the 1D initial model with data weighting apply (DW MO). Synthetic data (blue/white/red color scale) generated into the final reconstructed V_P at 2nd workflow step using: (a,b) L^2 misfit function, (c,d) GSOT misfit function. (a,c) receiver A along cable A (through the low velocity anomaly). (b,d) receiver B along cable B. Field data are overlapped in grayscale with transparency. Black arrows point to area where GSOT improves the datafit.	99
3.17	Same as Figure 3.16 but with final relaxed data weighting (RT2 FO) for display. This exhibits the improvement of datafit obtained in area which are not yet inverted. Here, black arrows point to improve fit and coherency of the diving wave with GSOT, whereas yellow arrows point to improved data-fit of reflected events with GSOT.	100
3.18	Slices of the 5 Hz FWI reconstructed V_P using L^2 misfit starting from 1D initial model. (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1.1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.	101
3.19	Same as Figure 3.18 using GSOT. Here, results are consistant until 1.5 to 2 km depth compared to L^2 -based FWI. Characteristic structures of the Valhall field are recovered.	102
3.20	2D common-receiver gathers at 7 Hz starting from the 1D initial model. Synthetic data (blue/white/red color scale) generated into the final reconstructed V_P using GSOT. (a) receiver A along cable A (through the low velocity anomaly). (b) receiver B along cable B. Field data are overlapped in grayscale with transparency.	103
3.21	Slices of the 7 Hz FWI reconstructed V_P using GSOT misfit starting from 1D initial model. (a-c) Horizontal slices at (a) 200 m depth, (b) 500 m depth and (c) 1.1 km depth. (d-e) Inline vertical slices for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f-g) Cross-line vertical slices at (f) $y = 9$ km and (g) $y = 6$ km.	104

3.22	Comparison of V_P profiles extracted from the 1D initial model (dashed red), early FWI models using L^2 (solid yellow) and 7 Hz FWI model using GSOT (solid purple) with sonic log (solid black). Left subfigure corresponds to the Log 1 at the center of the target. Middle subfigure to the Log 2, and right subfigure to Log 3 (which is far away from the target). Updates of velocity model obtained with GSOT are following the logs trend until ≈ 2 km depth.	105
3.23	Comparison of V_P profiles extracted from the TOMO and 1D initial models (respectively dashed red and dashed blue), GSOT-based FWI reconstructed models at 7 Hz starting from TOMO and 1D initial models (respectively solid red and solid blue), with sonic log (solid black). Logs 2 and 3 show that results from the two different starting models are globally following the same trend. Results from Log 1 passing through the target are following the same trend until 1.4 km depth.	106
4.1	Simple time-shifted Ricker wavelet test for crosscorrelation misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the KROT misfit function with respect to time-shift.	110
4.2	Complete (original) acquisition of a 3D OBC North Sea Valhall data. Sources are represented by black dots, receivers by blue diamond. Two subpanels shows a 2D extracted lines of receivers inside the acquisition.	111
4.3	Same as Figure 4.2, but for the interpolated dataset. Here the extracted 2D lines are now perfectly straight, with regular receivers spacing.	111
4.4	Representation of a simple signal $d(t)$ (black line) and the associated instantaneous phase $\phi(t)$ (red line).	114
4.5	Simple time-shifted Ricker wavelet test for the instantaneous phase misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the instantaneous phase misfit function with respect to time-shift.	114
4.6	Simple time-shifted Ricker wavelet test for the unwrapped phase misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the unwrapped phase misfit function with respect to time-shift.	116
4.7	Same as Figure 4.6 but with a very small white noise on the reference signal.	116
4.8	Spline envelopes and average spline envelopes for a given seismic trace (solid black). The top peaks (blue circle) and bottom peaks (red circle) are extracted using algorithms 1 and 2. From these minima and maxima, the top envelope (blue line) and bottom envelope (red line) are extracted using a piecewise cubic Hermite interpolating polynomial. The mean of these two envelopes, corresponding to the average spline envelope (purple line), presents a nice shape that could be seen as a low-frequency version of the original signal.	118
4.9	Simple time-shifted Ricker wavelet test for ASE misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the ASE and IE misfit functions with respect to time-shift.	119

LIST OF FIGURES

4.10	Extraction of average spline envelope based using PCHIP for three levels of noise (a: without noise, b: small amount of noise, c: strong noise).	120
4.11	Test of diffusion on a simple signal (solid red) for two diffusion values. The less diffused version (solid purple) and the two times more diffused version (solid blue).122	
4.12	Simple time-shifted Ricker wavelet test for crosscorrelation misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the diffusion-based misfit function with respect to time-shift. Here the diffusion time τ is set to 3000.123	
4.13	Test of diffusion on two Ricker wavelets, one unfiltered (solid blue) and one highpass filtered above 1.5 Hz (solid red). The right column corresponds to the associated spectrum. The top line represents the non-diffused signal while the bottom lined represents signals after diffusion.	123
4.14	Same setup as Figure 4.13 but with noise added on the signal.	124
4.15	Test of diffusion on a field data trace coming from Valhall from the 2.5 to 5 Hz frequency band. Top line corresponds to the raw data and associated spectrum. Bottom line corresponds to the diffused data and its associated spectrum. . . .	124
4.16	Beat tone illustration on a simple unfiltered Ricker wavelet and associated spectrum. First line corresponds to the raw Ricker wavelet, the second line to the beat tone version and the third line to a low-passed version of the beat tone version to remove high-frequency content.	126
4.17	Simple time-shifted Ricker wavelet test for crosscorrelation misfit function. (a) presentation of the setup, with reference signal (seen as d_{obs}) in solid black, and shifted signal in dashed black (seen as d_{cal}). (b) Value of the beat-tone misfit function with respect to time-shift. Here $f_c = 0.5$ Hz and the low-pass filter is set to 1 Hz.	126
4.18	Same setup as on Figure 4.16 but using a field data trace coming from Valhall instead of the Ricker wavelet.	127
4.19	Same setup as on Figure 4.18 but with a less effective parameterization of the beat tone resulting in no shift toward the low frequencies.	127
4.20	Exemple of the variability of beat tone on a Valhall field trace. The reference beat tone version is in solid black. In solid blue and red are two beats obtained from a time-shifted version of the observed data. We can observe here that as the carrier is always the same, a possible shift in time of the wavelet will dramatically change the signal shape.	128

Acknowledgments

This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, SHELL, SINOPEC, SISPROBE and TOTAL. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d’Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS/TGCC under the allocation 046091 made by GENCI.

I thank AKERBP ASA and their partner Pandion Energy for providing the dataset and permission to present this work. Special thanks to Einar Jarle Kjos, Ross Milne and Espen Birger Raknes for their help and support on the Valhall data.

I thank Denes Vigh and Xin Cheng from Schlumberger for their support and warm welcome, for their in-depth and interesting scientific discussion, and for providing me with the opportunities to come to their office to present my work.

I thank Equinor for the opportunities to work on 3D OBC real data applications in their office for two summer internships. I thank Lyubov Skopintseva and Joachim Mispel for their help and support. I sincerely thank Céline Ravaut for the opportunities you provided to me, for the help you provided when I started working for the first time on field data, and for the interesting scientific discussion.

Bibliography

- Aghamiry, H., Gholami, A., and Operto, S. (2019). Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers. *Geophysics*, 84(1):R139–R162.
- Aghamiry, H., Gholami, A., and Operto, S. (2020). Accurate and efficient wavefield reconstruction in the time domain. *Geophysics*, 85(2):A7–A12.
- Aki, K., Christoffersson, A., Husebye, E., and Powell, C. (1974). Three-dimensional seismic velocity anomalies in the crust and upper-mantle under the usgs, california seismic array. *Eos Trans. AGU*, 56:1145.
- Aki, K., Christoffersson, A., and Husebye, E. S. (1977). Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, 82(2):277–296.
- Aleardi, M. and Mazzotti, A. (2016). 1d elastic full-waveform inversion and uncertainty estimation by means of a hybrid genetic algorithm–gibbs sampler approach. *Geophysical Prospecting*.
- Alkhalifah, T. (1998). Acoustic approximations for processing in transversely isotropic media. *Geophysics*, 63:623–631.
- Ambrosio, L., Mainini, E., and Serfaty, S. (2011). Gradient flow of the Chapman Rubinstein Schatzman model for signed vortices. *Annales de l’Institut Henri Poincaré (C) Non Linear Analysis*, 28(2):217–246.
- Bardsley, J. M., Solonen, A., Haario, H., and Laine, M. (2014). Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Numerical Analysis*, 36(4):A1895–A1910.
- Barkved, O., Bærheim, A., Howe, D., Kommedal, J., and Nicol, G. (2003). Life of Field Seismic Implementation - Another “first at valhal”. In *65th EAGE Workshop, Stavanger*.
- Bedle, H. and Lee, S. V. D. (2009). S velocity variations beneath north america. *Journal of Geophysical Research: Solid Earth*, 114(B7).
- Bérenger, J.-P. (1994). A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.
- Bertsekas, D. P. and Castanon, D. (1989). The auction algorithm for the transportation problem. *Annals of Operations Research*, 20(1):67–96.

- Billette, F. and Lambaré, G. (1998). Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International*, 135(2):671–680.
- Bishop, T. N., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., and Spinder, D. A. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50:903–923.
- Biswas, R. and Sen, M. (2017). 2d full-waveform inversion and uncertainty estimation using the reversible jump hamiltonian monte carlo. In *SEG Technical Program Expanded Abstracts 2017*, pages 1280–1285.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D. (2016). Global adjoint tomography: first-generation model. *Geophysical Journal International*, 207(3):1739–1766.
- Bozdağ, E., Trampert, J., and Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2):845–870.
- Bretaudeau, F., Brossier, R., Leparoux, D., Abraham, O., and Virieux, J. (2013). 2D elastic full waveform imaging of the near surface: Application to synthetic and a physical modelling data sets. *Near Surface Geophysics*, 11:307–316.
- Bretaudeau, F., Brossier, R., Métivier, L., and Virieux, J. (2014). First-arrival delayed tomography using 1st and 2nd order adjoint-state method. In *Expanded Abstracts*, pages 4757–4762. Society of Exploration Geophysics.
- Brossier, R., Operto, S., and Virieux, J. (2009). Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118.
- Brossier, R., Virieux, J., and Operto, S. (2008). Parsimonious finite-volume frequency-domain method for 2-D P-SV-wave modelling. *Geophysical Journal International*, 175(2):541–559.
- Brossier, R., Zhou, W., Operto, S., Virieux, J., and Yang, P. (2017). From rwi to jfwi: including diving waves in reflection-based velocity model building. In *79th EAGE Conference and Exhibition 2017-Workshops*, pages cp–519. European Association of Geoscientists & Engineers.
- Bullen, K. E. and Bolt, B. A. (1985). *An introduction to the theory of seismology*. Cambridge University Press, Berlin, Heidelberg.
- Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.
- Carcione, J. M. (2001). *Wave fields in real media, Theory and numerical simulation of wave propagation in anisotropic, an elastic and porous media*. Pergamon Press.
- Carotti, D., Hermant, O., Masclet, S., Reinier, M., Messud, J., Sedova, A., and Lambaré, G. (2020). Optimal transport full waveform inversion - applications. In *82th Annual EAGE Meeting (Amsterdam)*. European Association of Geoscientists & Engineers.

- Cary, P. and Chapman, C. (1988). Automatic 1-D waveform inversion of marine seismic refraction data. *Geophysical Journal of the Royal Astronomical Society*, 93:527–546.
- Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):2117–2131.
- Červený, V. (2001). *Seismic Ray Theory*. Cambridge University Press, Cambridge.
- Chapman, C. (2004). *Fundamentals of seismic waves propagation*. Cambridge University Press, Cambridge, England.
- Choi, Y. and Shin, C. (2008). Frequency-Domain Elastic Full Waveform Inversion Using the New Pseudo-Hessian Matrix: Experience Of Elastic Marmousi 2 Synthetic Data. *Bulletin of the Seismological Society of America*, 98(5):2402–2415.
- Claerbout, J. (1971). Towards a unified theory of reflector mapping. *Geophysics*, 36:467–481.
- Claerbout, J. (1985). *Imaging the Earth’s interior*. Blackwell Scientific Publication.
- Cocher, E., Chauris, H., and Plessix, R. (2017). Seismic Iterative Migration Velocity Analysis: two strategies to update the velocity model. *Computational Geosciences*, 21(4):759–780.
- Combettes, P. L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, volume 49 of *Springer Optimization and Its Applications*, pages 185–212. Springer New York.
- Cruse, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic non-linear waveform inversion: application to real data. *Geophysics*, 55:527–538.
- Dablain, M. (1986). The application of high order differencing for the scalar wave equation. *Geophysics*, 51:54–66.
- Datta, D. and Sen, M. K. (2016). Estimating starting models for full waveform inversion using a global optimization method. *Geophysics*, 81(4):R211–R223.
- Debens, H. A., Mancini, F., Warner, M., and Guasch, L. (2017). Full-bandwidth adaptive waveform inversion at the reservoir. In *SEG Technical Program Expanded Abstracts 2017*, pages 1378–1382.
- Devaney, A. (1984). Geophysical diffraction tomography. *Geoscience and Remote Sensing, IEEE Transactions on*, GE-22(1):3–13.
- Dickinson, D., Mancini, F., Li, X., Zhao, K., and Ji, S. (2017). A refraction/reflection full-waveform inversion case study from north west shelf offshore australia. In *SEG Technical Program Expanded Abstracts 2017*, pages 1368–1372. Society of Exploration Geophysicists.
- Donno, D., Chauris, H., and Calandra, H. (2013). Estimating the background velocity model with the normalized integration method. In *EAGE Technical Program Expanded Abstracts 2013*, page Tu0704.
- Duveneck, E. and Bakker, P. M. (2011). Stable P-wave modeling for reverse-time migration in tilted TI media. *Geophysics*, 76(2):S65–S75.

- Duveneck, E., Milcik, P., Bakker, P. M., and Perkins, C. (2008). Acoustic VTI wave equations and their application for anisotropic reverse-time migration. *SEG Technical Program Expanded Abstracts*, 27(1):2186–2190.
- Emmerich, H. and Korn, M. (1987). Incorporation of attenuation into time-domain computation of seismic wavefield. *Geophysics*, 52:1252–1264.
- Engquist, B. and Froese, B. D. (2014). Application of the Wasserstein metric to seismic signals. *Communications in Mathematical Science*, 12(5):979–988.
- Engquist, B., Froese, B. D., and Yang, Y. (2016). Optimal transport for seismic full waveform inversion. *Communications in Mathematical Sciences*, 14(8):2309–2330.
- Etienne, V., Hu, G., Operto, S., Virieux, J., Barkved, O., and Kommedal, J. (2012). Three-dimensional acoustic full waveform inversion: algorithm and application to Valhall. In *Expanded Abstracts, 74th Annual EAGE Conference & Exhibition, Copenhagen*. EAGE.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophysical Journal International*, 175:665–685.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2010). Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle. *Earth and Planetary Science Letters*, 290(3-4):270–280.
- Fichtner, A. and Trampert, J. (2011). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2):775–798.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., and nor, A. V. (2013). Multiscale full waveform inversion. *Geophysical Journal International*, 194:534–556.
- Fichtner, A., Zunino, A., and Gebraad, L. (2018a). Hamiltonian monte carlo solution of tomographic inverse problems. *Geophysical Journal International*, 216(2):1344–1363.
- Fichtner, A., Zunino, A., and Gebraad, L. (2018b). A tutorial introduction to the hamiltonian monte carlo solution of weakly nonlinear inverse problems.
- Fowler, C. M. R., Fowler, C. M. R., and Fowler, M. (1990). *The solid earth: an introduction to global geophysics*. Cambridge University Press.
- French, S. W. and Romanowicz, B. A. (2015). Broad plumes rooted at the base of the earth’s mantle beneath major hotspots. *Nature*, 525(7567):95.
- Gardner, G. F., Gardner, L., and Gregory, A. (1974). Formation velocity and density—the diagnostic basics for stratigraphic traps. *Geophysics*, 39(6):770–780.
- Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics*, 51(7):1387–1403.

- Gebraad, L., Boehm, C., and Fichtner, A. (2020). Bayesian elastic full-waveform inversion using hamiltonian monte carlo. *Journal of Geophysical Research: Solid Earth*, 125(3):e2019JB018428. e2019JB018428 10.1029/2019JB018428.
- Gebraad, L. and Fichtner, D. A. (2018). Bayesian elastic full-waveform inversion using hamiltonian monte carlo. In *AGU Fall Meeting Abstracts*.
- Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., and Virieux, J. (2013). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 2: application to Valhall. *Geophysics*, 78(2):R107–R124.
- Giannakis, I., Giannopoulos, A., and Warren, C. (2019). A machine learning-based fast-forward solver for ground penetrating radar with application to full-waveform inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4417–4426.
- Górszczyk, A., Brossier, R., and Métivier, L. (2021). Graph-space optimal transport concept for time-domain full-waveform inversion of ocean-bottom seismometer data: Nankai trough velocity structure reconstructed from a 1d model. *Journal of Geophysical Research: Solid Earth*, 126(5):e2020JB021504. e2020JB021504 2020JB021504.
- Górszczyk, A., Métivier, L., and Brossier, R. (2019). Mitigating the nonlinearity of the crustal scale full waveform inversion through the graph space optimal transport misfit function. In *AGU Fall Meeting Abstracts*, volume 2019, pages S41A–03.
- Górszczyk, A., Métivier, L., and Brossier, R. (2020). Relaxing the initial model constraint for crustal-scale full-waveform inversion with graph space optimal transport misfit function. In *Expanded Abstracts, 82nd Annual EAGE Meeting (Amsterdam)*.
- Górszczyk, A., Operto, S., and Malinowski, M. (2017). Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited. *Journal of Geophysical Research: Solid Earth*, 122(6):4601–4630.
- Groos, L., Schäfer, M., Forbriger, T., and Bohlen, T. (2014). The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves. *Geophysics*, 79(6):R247–R261.
- Guasch, L., Agudo, O. C., Tang, M.-X., Nachev, P., and Warner, M. (2020). Full-waveform inversion imaging of the human brain. *NPJ digital medicine*, 3(1):1–12.
- Guasch, L., Warner, M., and Ravaut, C. (2019). Adaptive waveform inversion: Practice. *Geophysics*, 84(3):R447–R461.
- Gutenberg, B. (1914). Über erdbebenwellen viia. beobachtungen an registrierungen von fernbeben in göttingen und folgerungen über die konstitution des erdkörpers. *Nachrichten von der Königlich-Gesellschaft der Wissenschaften zu Göttinge, Mathematisch-Physikalische Klasse*, pages 125–176.
- Hall, S., Kendall, J., and Barkved, O. (2002). Fractured reservoir characterization using P-wave AVOA analysis of 3D OBC data. *The Leading Edge*, pages 777–781.

- Han, L., Hu, Y., Zhang, P., and Bai, L. (2016). Spline envelope full-waveform inversion. In *SEG Technical Program Expanded Abstracts 2016*, pages 1496–1500. Society of Exploration Geophysicists.
- He, W., Brossier, R., and Métivier, L. (2019a). 3D elastic FWI for land seismic data: A graph space OT approach. In *SEG Technical Program Expanded Abstracts 2019*, pages 1320–1324.
- He, W., Brossier, R., Métivier, L., and Plessix, R.-É. (2019b). Land seismic multi-parameter full waveform inversion in elastic VTI media by simultaneously interpreting body waves and surface waves with an optimal transport based objective function. *Geophysical Journal International*, 219(3):1970–1988.
- He, W., Brossier, R., Métivier, L., and Plessix, R.-É. (2019c). Land seismic multi-parameter full waveform inversion in elastic VTI media by simultaneously interpreting body waves and surface waves with optimal transport based objective function. In *Expanded Abstracts, 81th Annual EAGE Meeting (London)*.
- Hermant, O., Aziz, A., Warzocha, S., and Al Jahdhami, M. (2020). Imaging complex fault structures on-shore oman using optimal transport full waveform inversion. 2020(1):1–5.
- Hicks, G. J. (2002). Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions. *Geophysics*, 67:156–166.
- Hu, Y., Han, L., Zhang, P., Cai, Z., and Ge, Q. (2016). Multistep full-waveform inversion based on waveform-mode decomposition. In *SEG Technical Program Expanded Abstracts 2016*, pages 1501–1505. Society of Exploration Geophysicists.
- Huang, G., Nammour, R., and Symes, W. W. (2018a). Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation. *Geophysics*, 83(5):R449–R461.
- Huang, G., Nammour, R., and Symes, W. W. (2018b). Volume source-based extended waveform inversion. *Geophysics*, 83(5):R369–387.
- Irnaka, T., Brossier, R., Métivier, L., Bohlen, T., and Pan, Y. (2019). Towards 3d 9c elastic full waveform inversion of shallow seismic wavefields - case study ettlingen line. In *Expanded Abstracts, 81th Annual EAGE Conference & Exhibition, London*, page We P01 04. EAGE.
- Irnaka, T. M. (2021). *3D elastic full waveform inversion for subsurface characterization: Study of a shallow seismic multicomponent field data*. Theses, Université Grenoble Alpes.
- Irnaka, T. M., Brossier, R., Metivier, L., Bohlen, T., and Pan, Y. (2020). 3D 9C Full Waveform Inversion for Shallow Seismic Target: Ettlingen Line Case Study. *Geophysical Journal International*, submitted.
- Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., and Trezeguet, D. (1989). Wavelengths of Earth structures that can be resolved from seismic reflection data. *Geophysics*, 54(7):906–910.

BIBLIOGRAPHY

- Jin, S., Madariaga, R., Virieux, J., and Lambaré, G. (1991). Two dimensional asymptotic iterative elastic inversion. In *Expanded Abstracts*, volume SI5.6, pages 1013–1016. Society of Exploration Geophysics.
- Jin, S., Madariaga, R., Virieux, J., and Lambaré, G. (1992). Two-dimensional asymptotic iterative elastic inversion. *Geophysical Journal International*, 108:575–588.
- Kamath, N., Brossier, R., Metivier, L., Pladys, A., and Yang, P. (2020). Multiparameter full-waveform inversion of 3D OBC data from the valhall field. *Geophysics*, in press.
- Kamei, R., Pratt, R. G., and Tsuji, T. (2013). On acoustic waveform tomography of wide-angle obs data - strategies for preconditioning and inversion. *Geophysical Journal International*, 192:1250–1280.
- Kantorovich, L. (1942). On the transfer of masses. *Dokl. Acad. Nauk. USSR*, 37:7–8.
- Kelly, K., Ward, R., Treitel, S., and Alford, R. (1976). Synthetic seismograms - a finite-difference approach. *Geophysics*, 41:2–27.
- Klotzsche, A., van der Kruk, J., Linde, N., Doetsch, J., and Vereecken, H. (2013). 3-D characterization of high-permeability zones in a gravel aquifer using 2-D crosshole GPR full-waveform inversion and waveguide detection. *Geophysical Journal International*, 195(2):932–944.
- Klotzsche, A., van der Kruk, J., Meles, G. A., Doetsch, J. A., Maurer, H., and Linde, N. (2010). Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the Thur River, Switzerland. *Near Surface Geophysics*, 8.
- Klotzsche, A., van der Kruk, J., Meles, G. A., and Vereecken, H. (2012). Crosshole GPR full-waveform inversion of waveguides acting as preferential flow paths within aquifer systems. *Geophysics*, 77(4):H57–H62.
- Klotzsche, A., Vereecken, H., and van der Kruk, J. (2019). Review of crosshole ground-penetrating radar full-waveform inversion of experimental data: Recent developments, challenges, and pitfalls. *Geophysics*, 84(6):H13–H28.
- Köhn, D., Wilken, D., De Nil, D., Wunderlich, T., Rabbal, W., and Werther, L. (2018). 2d full waveform inversion applied to a strongly-dispersive love wave field dataset. In *80th EAGE Conference and Exhibition 2018*.
- Lailly, P. (1983). The seismic problem as a sequence of before-stack migrations. In Bednar, J., editor, *Conference on Inverse Scattering: Theory and Applications*. SIAM, Philadelphia.
- Lambaré, G. (2008). Stereotomography. *Geophysics*, 73(5):VE25–VE34.
- Lambaré, G., Virieux, J., Madariaga, R., and Jin, S. (1992). Iterative asymptotic inversion in the acoustic approximation. *Geophysics*, 57:1138–1154.
- Lavoué, F., Brossier, R., Métivier, L., Garambois, S., and Virieux, J. (2014). Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset GPR data: a frequency-domain quasi-Newton approach. *Geophysical Journal International*, 197(1):248–268.

- Lay, T. and Wallace, T. C. (1995). *Modern global seismology*. Academic Press.
- Le Bouteiller, P., Benjema, M., Métivier, L., and Virieux, J. (2019). A discontinuous Galerkin fast-sweeping Eikonal solver for fast and accurate traveltimes computation in 3D tilted anisotropic media. *Geophysics*, 84(2):C107–C118.
- Le Meur, H., Virieux, J., and Podvin, P. (1997). Seismic tomography of the gulf of Corinth: a comparison of methods. *Annali Di Geofisica*, XL(1):1–24.
- Lei, W., Ruan, Y., Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D. (2020). Global adjoint tomography—model glad-m25. *Geophysical Journal International*, 223(1):1–21.
- Lelièvre, P. G., Farquharson, C. G., and Hurich, C. A. (2011). Computing first-arrival seismic traveltimes on unstructured 3-D tetrahedral grids using the fast marching method. *Geophysical Journal International*, 184:885–896.
- Leonard, R. and Munns, J. (1987). *Valhall Field in Geology of Norwegian Oil and Gas Fields*. Graham and Trotman.
- Levander, A. R. (1988). Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53(11):1425–1436.
- Li, Y., Brossier, R., and Métivier, L. (2019). Joint FWI for imaging deep structures: A graph-space OT approach. In *SEG Technical Program Expanded Abstracts 2019*, pages 1290–1294.
- Li, Y., Brossier, R., and Métivier, L. (2020). On the comparison of MUMPS and STRUMPACK for 3D frequency-domain elastic wave modeling. In *Expanded Abstracts, 82nd Annual EAGE Meeting (Amsterdam)*.
- Liu, P. and Archuleta, R. J. (2006). Efficient modeling of q for 3d numerical simulation of wave propagation. *Bulletin of the Seismological Society of America*, 96(4A):1352–1358.
- Lu, Y., Stehly, L., Brossier, R., Paul, A., and ng Group, A. W. (2020). Imaging Alpine crust using ambient noise wave-equation tomography. *Geophysical Journal International*, 222(1):69–85.
- Lu, Y., Stehly, L., Paul, A., and Group, A. W. (2018). High-resolution surface wave tomography of the european crust and uppermost mantle from ambient seismic noise. *Geophysical Journal International*, 214(2):1136–1150.
- Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1):2788–2792.
- Luo, Y. and Schuster, G. T. (1991). Wave-equation traveltimes inversion. *Geophysics*, 56(5):645–653.
- Mainini, E. (2012). A description of transport cost for signed measures. *Journal of Mathematical Sciences*, 181(6):837–855.
- Malinowski, M. and Operto, S. (2006). Full-waveform modeling and inversion of wide-aperture data from the Polish basin. In *Expanded Abstracts*. EAGE.

- Marfurt, K. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49:533–549.
- Marple, L. (1999). Computing the discrete-time” analytic” signal via fft. *IEEE Transactions on signal processing*, 47(9):2600–2603.
- Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for Marmousi. *The Leading Edge*, 25(2):156–166.
- Martin, J., Wilcox, L., Burstedde, C., and Ghattas, O. (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal of Scientific Computing*, 34(3):A1460–A1487.
- Marty, P., Boehm, C., and Fichtner, A. (2021). Acoustoelastic full-waveform inversion for transcranial ultrasound computed tomography. In *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, volume 11602, page 1160211. International Society for Optics and Photonics.
- Mazzotti, A., Bienati, N., Stucchi, E., Tognarelli, A., Aleardi, M., and Sajeve, A. (2016). Two-grid genetic algorithm full-waveform inversion. *The Leading Edge*, 35(12):1068–1075.
- Messud, J. and Sedova, A. (2019). Multidimensional optimal transport for 3d FWI: Demonstration on field data. In *Expanded Abstracts, 81th Annual EAGE Meeting (London)*.
- Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2018). Optimal transport for mitigating cycle skipping in full waveform inversion: a graph space transform approach. *Geophysics*, 83(5):R515–R540.
- Métivier, L. and Brossier, R. (2016). The seiscopes optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. *Geophysics*, 81(2):F11–F25.
- Métivier, L. and Brossier, R. (2020). Extension strategy for time-domain full waveform inversion using receiver relocalization. *Geophysics*, submitted.
- Métivier, L., Brossier, R., Mérigot, Q., and Oudet, E. (2019). A graph space optimal transport distance as a generalization of L^p distances: application to a seismic imaging inverse problem. *Inverse Problems*, 35(8):085001.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016a). Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy. *The Leading Edge*, 35(12):1060–1067.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016b). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, 205:345–377.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016c). An optimal transport approach for seismic tomography: Application to 3D full waveform inversion. *Inverse Problems*, 32(11):115008.
- Middleton, G. V. (1973). Johannes walther’s law of correlation of facies. *Geological Society of America Bulletin*, 84:979–988.

- Minet, J., Lambot, S., Slob, E. C., and Vanclooster, M. (2010). Soil surface water content estimation by full-waveform GPR signal inversion in the presence of thin layers. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3):1138–1150.
- Moczo, P. and Kristek, J. (2005). On the rheological models used for time-domain methods of seismic wave propagation. *Geophysical Research Letters*, 32(1).
- Mohorovičić, A. (1909). Das beben vom 8. x. 1909. *Jb. Met. Obs. Zagreb (Agram)*, 9:1–63.
- Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*.
- Munns, J. W. (1985). The Valhall field: a geological overview. *Marine and Petroleum Geology*, 2:23–43.
- Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151):773–782.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.
- Nolet, G. (1987). *Seismic tomography with applications in global seismology and exploration geophysics*. D. Reidel publishing Company.
- Oldham, R. D. (1906). The constitution of the interior of the earth, as revealed by earthquakes. *Quarterly Journal of the Geological Society*, 62(1-4):456–475.
- Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prioux, V., Ribodetti, A., and Virieux, J. (2013). A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice. *The Leading Edge*, Special section Full Waveform Inversion(September):1040–1054.
- Operto, S. and Miniussi, A. (2018). On the role of density and attenuation in 3D multiparameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea. *Geophysical Journal International*, 213:2037–2059.
- Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., and Virieux, J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation. *Geophysical Journal International*, 202(2):1362–1391.
- Panning, M. P., Lekić, V., and Romanowicz, B. A. (2010). Importance of crustal corrections in the development of a new global model of radial anisotropy. *Journal of Geophysical Research: Solid Earth*, 115(B12).
- Pavlenkova, G. and Pavlenkova, N. (2006). Upper mantle structure of the northern eurasia from peaceful nuclear explosion data. *Tectonophysics*, 416(1-4):33–52.
- Pladys, A. (2016). Sensibility of several misfit functions for Full Waveform Inversion: theoretical analysis, implementation and comparative study on multiple realistic cases. Master’s thesis, Univ. Grenoble Alpes.

- Pladys, A., Brossier, R., Irnaka, M., Kamath, N., and Métivier, L. (2019). Assessment of optimal transport based FWI: 3d OBC valhall case study. In *SEG Technical Program Expanded Abstracts 2019*, pages 1295–1299.
- Pladys, A., Brossier, R., Kamath, N., and Métivier, L. (2021a). Robust fwi with graph space optimal transport: application to 3d obc valhall data. *Geophysics*, submitted.
- Pladys, A., Brossier, R., Li, Y., and Métivier, L. (2021b). On cycle-skipping and misfit functions modification for full-wave inversion: comparison of five recent approaches. *Geophysics*, in press.
- Pladys, A., Brossier, R., and Métivier, L. (2017). FWI alternative misfit functions: what properties should they satisfy. In *79th EAGE Conference and Exhibition 2017, Paris*, page Tu P1 01.
- Pladys, A., Brossier, R., and Métivier, L. (2020). Graph space optimal transport based FWI: 3D OBC valhall case study. In *SEG Technical Program Expanded Abstracts 2020*.
- Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.
- Plessix, R. E. (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver. *Geophysics*, 74(6):WCC53–WCC61.
- Plessix, R.-E., Baeten, G., de Maag, J. W., and ten Kroode, F. (2012). Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set. *Geophysical Prospecting*, 60:733 – 747.
- Plessix, R. E. and Perkins, C. (2009). 3D full-waveform inversion with a frequency-domain iterative solver. In *Expanded Abstracts*. European Association of Geoscientists & Engineers.
- Plessix, R. E. and Perkins, C. (2010). Full waveform inversion of a deep water ocean bottom seismometer dataset. *First Break*, 28:71–78.
- Poncet, R., Messud, J., Bader, M., Lambaré, G., Viguier, G., and Hidalgo, C. (2018). Fwi with optimal transport: a 3D implementation and an application on a field dataset. In *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*.
- Popovici, M. and Sethian, J. (1998). Three dimensional traveltimes using the fast marching method. In *Expanded Abstracts*, pages 1–22. Eur. Ass. Expl. Geophys.
- Pratt, R. G. (1990). Frequency-domain elastic modeling by finite differences: a tool for crosshole seismic imaging. *Geophysics*, 55(5):626–632.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model. *Geophysics*, 64:888–901.
- Pratt, R. G. and Goulty, N. R. (1991). Combining wave-equation imaging with traveltome tomography to form high-resolution images from crosshole data. *Geophysics*, 56(2):204–224.
- Pratt, R. G., Shin, C., and Hicks, G. J. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133:341–362.

- Pratt, R. G. and Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, part II: Fault delineation in sediments using crosshole data. *Geophysics*, 64:902–914.
- Pratt, R. G., Song, Z. M., Williamson, P. R., and Warner, M. (1996). Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International*, 124:323–340.
- Pratt, R. G. and Worthington, M. H. (1990). Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method. *Geophysical Prospecting*, 38:287–310.
- Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., and Kommedal, J. (2011). On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, 187:1495–1515.
- Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 1: imaging compressional wavespeed, density and attenuation. *Geophysical Journal International*, 194(3):1640–1664.
- Provenzano, G., Brossier, R., and Métivier, L. (2020a). Joint FWI of diving and reflected waves using a graph space optimal transport distance: synthetic tests on limited-offset surface seismic data. In *SEG Technical Program Expanded Abstracts 2020*.
- Provenzano, G., Brossier, R., Métivier, L., and Li, Y. (2020b). *Joint FWI of diving and reflected waves using a graph space optimal transport distance: Synthetic tests on limited-offset surface seismic data*, pages 780–784.
- Qiu, L., Ramos-Martínez, J., Valenciano, A., Yang, Y., and Engquist, B. (2017). Full-waveform inversion with an exponentially encoded optimal-transport norm. In *SEG Technical Program Expanded Abstracts 2017*, pages 1286–1290.
- Raknes, E. B., Arntsen, B., and Weibull, W. (2015). Three-dimensional elastic full waveform inversion using seismic data from the sleipner area. *Geophysical Journal International*, 202(3):1877–1894.
- Ravaut, C., Maaß, F., Mispel, J., Osen, A., Warner, M., Guasch, L., and Nangoo, T. (2017). Imaging beneath a gas cloud in the north sea without conventional tomography. In *EAGE, 79th Conference and Exhibition, Expanded abstracts*, page We A3 04.
- Ravaut, C., Operto, S., Improta, L., Herrero, A., Virieux, J., and Dell’Aversana, P. (2002). Seismic imaging of a thrust belt by traveltimes and frequency-domain waveform inversions. In *Extended Abstracts*. Eur. Ass. Expl. Geophys.
- Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and dell’Aversana, P. (2004). Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International*, 159:1032–1056.
- Romanowicz, B. (2003). Global mantle tomography: progress status in the past 10 years. *Annual Review of Earth and Planetary Sciences*, 31:303–328.

- Roth, T., Nangoo, T., Shah, N., Riede, M., Henke, C., and Warner, M. (2018). Improving seismic image with high resolution velocity model from awi starting with 1d initial model - case study barents sea.
- Sajeva, A., Aleardi, M., and Mazzotti, A. (2017). Genetic algorithm full-waveform inversion: uncertainty estimation and validation of the results. *Bollettino di Geofisica Teorica ed Applicata*, 58(4).
- Schäfer, M., Groos, L., Forbriger, T., and Bohlen, T. (2013). 2D full waveform inversion of recorded shallow seismic Rayleigh waves on a significantly 2D structure. In *Proceedings of 19th European Meeting of Environmental and Engineering Geophysics, Expanded Abstracts, Bochum, Germany*.
- Sears, T. J., Barton, P. J., and Singh, S. C. (2010). Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to alba field, u. k. north sea. *Geophysics*, 75(6):R109–R119.
- Sedova, A., Messud, J., Prigent, H., Royle, G., and Lambaré, G. (2019). Acoustic land full waveform inversion on a broadband land dataset: the impact of optimal transport. In *Expanded Abstracts, 81th Annual EAGE Meeting (London)*.
- Sen, M. and Stoffa, P. (1991). Non-linear one-dimensional seismic waveform inversion using simulated annealing. *Geophysics*, 56:1624–1638.
- Shen, X., Jiang, L., Dellinger, J., Brenders, A., Kumar, C., James, M., Etgen, J., Meaux, D., Walters, R., and Abdullayev, N. (2018). High-resolution full-waveform inversion for structural imaging in exploration. In *SEG Technical Program Expanded Abstracts 2018*, pages 1098–1102.
- Shipp, R. M. and Singh, S. C. (2002). Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. *Geophysical Journal International*, 151:325–344.
- Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.
- Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.
- Smith, J. A., Borisov, D., Cudney, H., Miller, R. D., Modrak, R., Moran, M., Peterie, S. L., Sloan, S. D., Tromp, J., and Wang, Y. (2019). Tunnel detection at yuma proving ground, arizona, usa—part 2: 3d full-waveform inversion experimentstunnel detection at ypg—part 2: 3d fwi. *Geophysics*, 84(1):B95–B108.
- Sourbier, F., Operto, S., Virieux, J., Amestoy, P., and L’Excellent, J.-Y. (2009). FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data—Part 2: Numerical examples and scalability analysis. *Computers & Geosciences*, 35(3):496 – 514.
- Stehly, L., Campillo, M., and Shapiro, N. (2006). A study of the seismic noise from its long-range correlation properties. *Journal of Geophysical Research: Solid Earth*, 111(B10):n/a–n/a. B10306.

- Stehly, L., Fry, B., Campillo, M., Shapiro, N., Guilbert, J., Boschi, L., and Giardini, D. (2009). Tomography of the Alpine region from observations of seismic ambient noise. *Geophysical Journal International*, 178:338–350.
- Stoffa, P. L. and Sen, M. K. (1991). Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane-wave seismograms. *Geophysics*, 56(11):1794–1810.
- Stopin, A., Plessix, R.-E., and Al Abri, S. (2014). Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman. *Geophysics*, 79(3):WA69–WA77.
- Sultanov, D., Murphy, J., and Rubinstein, K. D. (1999). A seismic source summary for soviet peaceful nuclear explosions. *Bulletin of the Seismological Society of America*, 89(3):640–647.
- Symes, W. (2015). Algorithmic aspects of extended waveform inversion. In *77th EAGE Conference and Exhibition 2017-Workshops*.
- Symes, W. W. (2008). Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56:765–790.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180:433–462.
- Tape, C., Liu, Q., and Tromp, J. (2007). Finite-frequency tomography using adjoint methods—methodology and examples using membrane surface waves. *Geophysical Journal International*, 168(3):1105–1129.
- Tarantola, A. (1984a). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.
- Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.
- Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128:365–399.
- Telford, W. M., Geldart, L. P., and Sheriff, R. E. (1990). *Applied Geophysics - Second Edition*. Cambridge University Press.
- Thurin, J. (2020). *Uncertainties estimation in Full Waveform Inversion using Ensemble methods*. Theses, Université Grenoble Alpes.
- Tran, K. and Hiltunen, D. (2011). Two-dimensional inversion of full waveforms using simulated annealing. *Journal of Geotechnical and Geoenvironmental Engineering*, 138(9):1075–1090.
- Trinh, P., Brossier, R., Lemaistre, L., Métivier, L., and Virieux, J. (2019a). 3d elastic fwi with a non-linear model constraint: Application to a real complex onshore dataset. In *Expanded Abstracts, 81th Annual EAGE Conference & Exhibition, London*, page Th P01 07. EAGE.
- Trinh, P. T., Brossier, R., Métivier, L., Tavard, L., and Virieux, J. (2019b). Efficient 3D time-domain elastic and viscoelastic Full Waveform Inversion using a spectral-element method on flexible Cartesian-based mesh. *Geophysics*, 84(1):R75–R97.

- Trinh, P. T., Brossier, R., Métivier, L., Tavard, L., and Virieux, J. (2019c). Efficient 3D time-domain elastic and viscoelastic Full Waveform Inversion using a spectral-element method on flexible Cartesian-based mesh. *Geophysics*, 84(1):R75–R97.
- Tromp, J. (2019). Seismic wavefield imaging of earth’s interior across scales. *Nature Reviews Earth & Environment*, pages 1–14.
- Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160:195–216.
- van Leeuwen, T. and Herrmann, F. (2016). A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):1–26.
- van Leeuwen, T. and Herrmann, F. J. (2013). Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195(1):661–667.
- van Leeuwen, T. and Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation traveltimes tomography. *Geophysical Journal International*, 182(3):1383–1394.
- Vidale, D. (1990). Finite-difference calculation of traveltimes in three dimensions. *Geophysics*, 55:521–526.
- Vigh, D., Cheng, X., Jiao, K., Sun, D., and Kapoor, J. (2014). Multiparameter tti full waveform inversion on long-offset broadband acquisition: a case study. In *Expanded Abstracts*, pages 1061–1065. Society of Exploration Geophysics.
- Virieux, J. (1984). SH wave propagation in heterogeneous media: Velocity-stress finite difference method. *Geophysics*, 49:1259–1266.
- Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite difference method. *Geophysics*, 51:889–901.
- Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., and Zhou, W. (2017). An introduction to Full Waveform Inversion. In Grechka, V. and Wapenaar, K., editors, *Encyclopedia of Exploration Geophysics*, pages R1–1–R1–40. Society of Exploration Geophysics.
- Virieux, J. and Lambaré, G. (2007). Theory and observations - body waves: ray methods and finite frequency effects. In Romanowicz, B. and Diewonski, A., editors, *Treatise of Geophysics, volume 1: Seismology and structure of the Earth*. Elsevier.
- Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.
- Wang, C., Yingst, D., Farmer, P., and Leveille, J. (2016). *Full-waveform inversion with the reconstructed wavefield method*, pages 1237–1241.
- Wang, K., Li, Y., Kemao, Q., Di, J., and Zhao, J. (2019). One-step robust deep learning phase unwrapping. *Opt. Express*, 27(10):15100–15115.
- Wang, Y. and Rao, Y. (2009). Reflection seismic waveform tomography. *Journal of Geophysical Research*, 114(B3):1978–2012.

- Warner, M. and Guasch, L. (2015). Robust adaptive waveform inversion. In *SEG Technical Program Expanded Abstracts 2015*, pages 1059–1063.
- Warner, M. and Guasch, L. (2016). Adaptive waveform inversion: Theory. *Geophysics*, 81(6):R429–R445.
- Warner, M., Nangoo, T., Pavlov, A., and Hidalgo, C. (2019). Extending the velocity resolution of waveform inversion below the diving waves using awi. 2019(1):1–5.
- Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Stekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A. (2013). Anisotropic 3D full-waveform inversion. *Geophysics*, 78(2):R59–R80.
- Wittkamp, F., Athanasopoulos, N., and Bohlen, T. (2018). Individual and joint 2-d elastic full-waveform inversion of rayleigh and love waves. *Geophysical Journal International*, 216(1):350–364.
- Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM Review*, 11.
- Wu, R.-S., Luo, J., and Wu, B. (2014). Seismic envelope inversion and modulation signal model. *Geophysics*, 79(3):WA13–WA24.
- Wu, R. S. and Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52:11–25.
- Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016a). A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium. *Geophysical Journal International*, 207(1):129–149.
- Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016b). Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method. *Geophysics*, 81(6):R349–R362.
- Yang, P., Brossier, R., Métivier, L., Virieux, J., and Zhou, W. (2018a). A Time-Domain Pre-conditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion. *SIAM Journal on Scientific Computing*, 40(4):B1101–B1130.
- Yang, Y. and Engquist, B. (2018). Analysis of optimal transport and related misfit functions in full-waveform inversion. *GEOPHYSICS*, 83(1):A7–A12.
- Yang, Y., Engquist, B., Sun, J., and Hamfeldt, B. F. (2018b). Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion. *Geophysics*, 83(1):R43–R62.
- Yao, G., Wu, D., and Wang, S.-X. (2020). A review on reflection-waveform inversion. *Petroleum Science*, 17(2):334–351.
- Zelt, C. and Smith, R. B. (1992). Seismic traveltime inversion for 2-D crustal velocity structure. *Geophysical Journal International*, 108:16–34.
- Zhang, Y., Zhang, H., and Zhang, G. (2011). A stable TTI reverse time migration and its implementation. *Geophysics*, 76(3):WA3–WA11.

- Zhou, W. (2016). *Full Waveform Inversion of Early Arrivals and Reflections for Velocity Model Building and Case Study with Gas Cloud Effect*. PhD thesis, Univ. Grenoble Alpes.
- Zhou, W., Brossier, R., Operto, S., and Virieux, J. (2016). Joint full waveform inversion of early arrivals and reflections: a real obc case study with gas cloud. In *SEG Technical Program Expanded Abstracts 2016*, pages 1247–1251. SEG.
- Zhou, W., Brossier, R., Operto, S., Virieux, J., and Yang, P. (2018). Velocity model building by waveform inversion of early arrivals and reflections: a 2d ocean-bottom-cable study with gas cloud effects. *Geophysics*, 83(2):R141–R157.
- Zhu, H., Bozdağ, E., and Tromp, J. (2015). Seismic structure of the European upper mantle based on adjoint tomography. *Geophysical Journal International*, 201(1):18–52.

