UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FíSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

IMPLEMENTATION OF SYNTHETIC SCENARIOS FOR THE STUDY OF THE AMBIENT NOISE TOMOGRAPHY TECHNIQUE

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL ELÉCTRICA

ROMINA CAROLINA ARRIAZA BARRIGA

PROFESOR GUÍA:
MARCOS DÍAZ QUEZADA
PROFESOR CO-GUÍA:
DIANA COMTE SELMAN

MIEMBROS DE LA COMISIÓN:
MARCOS ORCHARD CONCHA
JORGE SILVA SANCHEZ
STEVEN W. ROECKER

Este trabajo ha sido parcialmente financiado por PROYECTO CORFO 14IDL4-300327
EXPLORACIÓN MINERA POR TOMOGRAFÍA SÍSMICA

SANTIAGO DE CHILE
2017

RESUMEN DE LA TESIS/MEMORIA PARA OPTAR AL
TÍTULO DE Ingeniera Civil Eléctrica y al grado de
Magíster en Ciencias de la Ingeniería, Mención Eléctrica
FECHA: 2017
PROF. GUÍA: Marcos Díaz Quezada

IMPLEMENTATION OF SYNTHETIC SCENARIOS FOR THE STUDY OF THE AMBIENT NOISE TOMOGRAPHY TECHNIQUE

La Tomografía de Ruido Ambiente (ANT) consiste en una técnica de busca modelar áreas y estructuras cuya novedad es que utiliza el ruido ambiente como herramienta. Este tipo de señal suele ser eliminada en técnicas como la Tomografía Sísmica (TS), la cual es la más ampliamente utilizada, especialmente en el área de la minería para la exploración de yacimientos. A diferencia de ANT, ésta hace uso señales de alta energía como terremotos y explosiones artificiales, por lo tanto, dentro de su procesamiento elimina este ruido. En zonas con baja actividad sísmica no es posible utilizar TS, por lo que es necesario desarrollar técnicas alternativas. ANT es una de las opciones que cumple, además de ser amigable con el medio ambiente.

Una de las problemáticas tanto de ANT es que no cuenta con un ground-truth. Dado esto, no se puede saber con certeza qué tan cercano es el modelo estimado a la realidad. Para intentar resolver esto, actualmente sismólogos contrastran los resultados provenientes de ST o usan información de previo conocimiento del área, pero no es suficiente.

Era necesario el desarrollo de una plataforma sintética, que fuera lo más cercano a la realidad posible. Sobre ésta se implementa la técnica de tomografía de ruido, con el objetivo de identificar las variables que influyen en la estimación, tales como la posición de estaciones, la frecuencia de la fuente, velocidad del medio, entre otras. También esta plataforma debe cuenta con un ground-truth para determinar los errores de estimación.

En el presente trabajo, ésta plataforma fue implementada en lenguaje Python. Cada etapa de la metodología y fue probada sobre tres escenarios sintéticos con diferentes caracteríasticas de estructura. Se estudiaron las variables de interés y se realizó un análisis sobre cada etapa. También se logró entregar una visualización clara y apropiada de los resultados.

Este trabajo sirvió para crear una base de conocimiento sobre esta nueva técnica para que pueda continuar siendo desarrollada y mejorada. Diversas mejoras potenciales fueron identificadas y se entregaron diversas soluciones para su posterior implementación en futuras campañas para el mejor provecho de los recursos existentes.

RESUMEN DE LA TESIS/MEMORIA PARA OPTAR AL
TÍTULO DE Ingeniera Civil Eléctrica y al grado de
Magíster en Ciencias de la Ingeniería, Mención Eléctrica
FECHA: 2017
PROF. GUÍA: Marcos Díaz Quezada

IMPLEMENTATION OF SYNTHETIC SCENARIOS FOR THE STUDY OF THE AMBIENT NOISE TOMOGRAPHY TECHNIQUE

Ambient Noise Tomography (ANT) consists in a technique that seeks to model areas and structures and whose novelty is the use of ambient noise as a tool. This kind of signal is usually disregarded in techniques such as in Seismic Tomography (ST), which is more widely used specially in the mining area of ore deposits exploration. Unlike ANT, ST uses high energy signals such as earthquakes and artificial explosions, thus its processing internally ignores this noise. In areas of low seismic activity, it is not possible to use ST, therefore it is necessary to develop alternative methods. ANT is one of the options that can replace it, and in an environmentally friendly way.

One of the problems of ANT is that it has no ground-truth. Consequently, it is not possible to know exactly how similar the estimation is from reality. To solve this, seismologists try to contrast the results coming from ST or they use prior information from the area, but this is not always enough.

For studying ANT, it was necessary to develop a synthetic simulator that is as close to reality as possible, and implement the noise tomography technique on it. The objective was to identify the variables that affect the estimation such as the position of stations, the source?s frequency and velocity of the medium. This platform must have included a ground-truth to determine the estimation errors.

In this work, the platform was implemented in Python. Each step of the methodology was tested on three synthetic scenarios with different structure characteristics. The interest variables were studied and an analysis was performed on each step of the process. In addition, a clear and proper visualization was achieved for showing the results.

This work helped creating a basal knowledge of this new technique for allowing future development and enhancement. Many potential improvements were identified, and several solutions were given. These should help future campaigns in order to make better use of the existing resources.

"We're all stories at the end. Just make it a good one, eh?"
The Doctor

Acknowledgements

I would like to thank my family for their support. To my mother Patricia, my brother Sebastián, my sister Loreto and my niece Agustina. I am grateful for the attention, concern and advices. And of course the constant remindings for me to rest.

I would also like to thank my teacher Diana Comte for not only teaching me in the academic field, but also teaching me how to be a professional. Thank you very much for the support, the concern and the many tools that you have given to me. I will always work hard to keep you proud.

I want to thank Professor Steven Roecker. I would like to thank you for your good disposition and all the help you have given to me since the beginning of my thesis. Your contribution was vital for my training.

I am grateful to the members of my committee who supported me in the whole process of the thesis: To my mentor teacher Marcos Diaz. To my professors of commission Jorge Silva and Marcos Orchard.

Many thanks to Sergio Liberman for his invaluable help and concern. You have been an excellent support. Thank you for been there.

Of course I have to thank Constanza Z. Csori for being an unconditional friend and always having a coffee for me on the hard times and a pitcher for the celebrations.

Many thanks to my friend Catalina Elzo. Without the countless beers and conversations this process would have been too monotonous. Thank you for always being there and giving me such good times.

I am grateful to the Opamps for their good companionship in my passage through the career. Joaquín alias Pato, Iván alias Capitaaaaaaaan, César Beodo D 'iaz, Gerardo, Joseph, Caro, Canitas, Peladito, and Lander.

Finally and the most important: to my grandfather. Thank you very much for helping me with the multiplication tables.

Many thanks to CORFO, who, through the CORFO PROJECT 14IDL4-300327 MINING EXPLORATION BY SOMMERIC TOMOGRAPHY, financed my magister.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

The development of techniques to identify the structure and properties of the Earth's subsurface is significant for the study of complex geological structures. In Chile, there is an interest in working on these techniques, with the objective of having a new low cost, that is also environmentally friendly, alternative for the deposit exploration in mining companies. Despite that the models achieved with this methods have improved because of the computational advances and seismic network expansion, there's still a lot of work to do in the area.

In order to develop a model of the Earth's subsurface, seismologists use high energy sources such as earthquakes or artificial explosions. Seismologist watch how waves move in order to make assumptions of the structure and properties of the crust. This method is called Seismic Tomography. Within the measurements, another type of signals were present in the records of the network: background noise – a signal recorded when there is no identified source – and seismic coda waves – the result from several scattered parts of seismic waveforms – [2]. Background noise is commonly called *ambient seismic noise*, which is a low energy signal that could be caused by several sources such as the action of the ocean and wind on the continent, rock fracturing and human activity [4].

Regarding seismic tomography, high energy signals are the only ones we want to study. Thus, ambient noise used to be attenuated or filtered out from the recordings in order to enhance these coherent signals. Nevertheless, seismically quiet regions do not have enough information which makes it difficult to use local seismic tomography methods [4]. Given that noise travels through the Earth, it should contain the missing information. After this process, the objective is to find a way to unravel surface's characteristics from those complex-looking waveforms. The use of interferometry techniques on ambient noise allows to create an artificial surface wave from it, which represents the system's response. The need to solve this problem is how a new technique was created, known as Ambient Noise Tomography. Given that this method works with noise, its resolution concentrates on the surface, and it is unable to achieve models of greater depth, unlike classic tomography does. However, the depth range that the ambient noise tomography can reach is enough for the study of mining

deposits.

The emergence of interferometry theory applied to noise allowed seismologists to discover new information of the Earth, specially on relatively aseismic areas. Moreover, this technique does not need to estimate where the sources are located since it is known where the stations are. This simplifies one of the problems of the classical earthquake tomography method [4].

The common practice of the ambient noise tomography allowed the improvement of the technique. A pre-processing method was established in order to increase the signal to noise ratio of the final cross-correlation signal [1], and the characteristics that must meet the scenario to study [12, 11].

Noise tomography is a prominent tool for exploration and it is a technique highly developed in recent years allowed by computational advances. Nevertheless, it has a drawback: it does not have an appropriate method to identify how close the estimated models to reality are. Nowadays, the methods used are a comparison with results from seismic tomography or the contraposition with prior information of the area. If the estimated model corresponds with any of those two, the result will be consider acceptable.

It is proposed that one way to formally approve these estimations is developing several synthetic scenarios in order to understand the variables which influence the inversion problem. Those variables are resolution, the size of the area of study, the location of the stations, among others. For instance, the analysis of two sensor's performances: Broad Band and Short Period seismometers. The first kind is significantly more expensive than a 4.5 Hz seismometer because of the large range of periods it can identify. Nevertheless, it is interesting to analyse the benefits and disadvantages of each sensor, in order to understand if it is possible to estimate certain structures using low-cost sensors and under which conditions.

There are previous works in the state of the art from scientists who made their first synthetic simulations. [9] who corroborated the interference of sources within the Fresnel zone, [8] and [7] who has inferred an interferometric modelling of wave propagation through elastic media using acoustic waves. This work seeks to go a step further. The implementation of the entire controlled scenario where all hypothesis can be proved.

1.2 Hypothesis, goals and methodology

1.2.1 Hypothesis

Although ambient noise tomography is widely implemented, there is no rigorous way to know how well this technique works. Nowadays, the methods used are a comparison of performance with seismic tomography or the contraposition with prior information of the area. However, there is no certainty of how close the result is to the real scenario or how it could be improved for future campaigns.

Given this, the hypothesis for this work is:

The implementation of a controlled synthetic platform will prove the parameter sensitivity of the methodology. That is, the consequences of the variation of parameters such as station distance and inversion problem settings. Therefore, the selection of parameters through the L-curve method and the guide of the established restrictions will lead to results with an error less than 10%. And this platform can be processed and computed on any of the current operating systems with the minimum requirements for them.

1.2.2 Main goal

The main objective of this work is the development of different synthetic controlled scenarios in order to know which variables influence the final model, in which way they do and the conditions for this to happen.

1.2.3 Specific goals

1. To adopt and implement widely used techniques for the implementation of each stage of the ambient noise tomography.
2. To be able to simulate different cases, close to reality. That is, to be adaptable so that any variable of the problem can be modified such as station locations and surface structure among others.
3. To estimate the recordings of each station given different setups.
4. To give an appropriate and clear visualization of the results. A comparison between estimated model and ground-truth, and cross-correlation signals.
5. To determine how different variables influence in the inverse problem.

1.2.4 Methodology

The steps that will be followed for the implementation and analysis of the synthetic scenarios will be:

1. To study of the state of the art of the ambient noise tomography technique.
2. To implement different surface scenarios, station locations and noise sources to simulate the stations recordings.
3. To implement different stages of the inverse problem. Each stage is tested under different conditions to ensure its correct operation.
4. To implement the correction stage of the technique in order to improve the final model.
5. To analyse the influence of each variable envolved in the methodology. This influence is observed with the comparison of the final estimated model and the ground-truth.

1.3 Thesis structure

1. **Chapter 2: State of the art and background.** This chapter reviews the concepts, stages and methodologies used for the implementation of the ambient noise tomography in the world.
2. **Chapter 3: Synthetic scenario.** This chapter focuses in how the synthetic scenarios were designed, programmed and tested.
3. **Chapter 4: Results and discussion.** This section reveals the results of each stage the tomography implemented on the synthetic scenarios. This outcomes are discussed and new experiments are performed.
4. **Chapter 5: Conclusion.** This chapter makes a list of the principal conclusions obtained through the previous work. It establishes necessary points to study in future works.

Chapter 2

State of the art and background

2.1 Types of tomography

Tomography involves the modelling of a system using the signals that travel through it. Having an identified signal emitter and receiver, and analysing the recordings of the last one, it is possible to identify how much time has taken to this signal to travel between those two points. With several emitters and receivers located all over the system, numerous paths can be obtained. With that, zones where the signal travelled faster than another or, in other words, zones with different composition can be identified .

In the geological field, we can mainly identify two types of tomography: Seismic Tomography (ST) and Ambient Noise Tomography (ANT). The main difference between them is the type of source they use. On one hand, ST uses waves from earthquakes and explosions. On the other hand ANT - as its name suggest - uses low energy signals, that is ambient noise. Along this section both techniques will be explored.

2.1.1 Seismic Tomography

This is a technique which uses signals originated from earthquakes and explosions in order to create a 3D-model of the Earth's interior. Observing the behaviour of this signals while they pass through the system, it is possible to know the internal properties of the Earth's crust. It is a similar technique to a CAT-scan, which uses X-rays transmitted through a person's body at different directions. A mathematical procedure is applied with the objective of explaining the absorption that the rays suffer throughout the body. However, in the case of the CAT-scan, X-rays travel in straight paths which is not the case of seismic tomography. The Earth's crust is broadly layered (different densities are present), and because of that seismic rays are reflected and refracted across boundaries which do not follow a straight line. Nevertheless, for simplification purposes, in this work it will be assumed that they do.

Given a station network displayed in the figure 2.1, each station is responsible for the recording of the arrival time of seismic waves. This allows to have a notion of the wave speed along the traveled path. Seismic tomography uses this information and mathematical modelling of P and S waves to create a map with the perturbations in the crust. The difference in speed propagations between P and S waves is due to various material properties which control the velocity and absorption of seismic signals. The study needs to consider the magnitudes and travel times in order to infer the existence of features within the Earth.

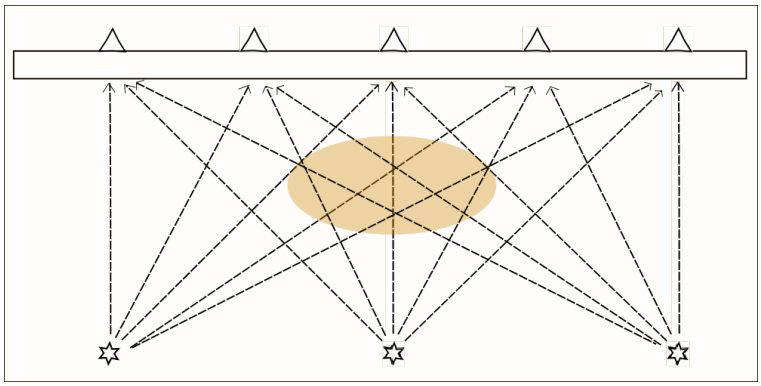


Figure 2.1: Schematic depicts 5 seismic stations (triangles) recording information from 3 earthquakes (stars). The anomaly (brown) can be detected using the recorded data that indicates slower or faster wave speeds.

The variables that influence in a best resolution of the model are:

1. High-density network of seismic stations covering the area of study
2. Large number of earthquakes and explosions recorded by every station
3. Earthquakes signals coming to stations from different parts of the world

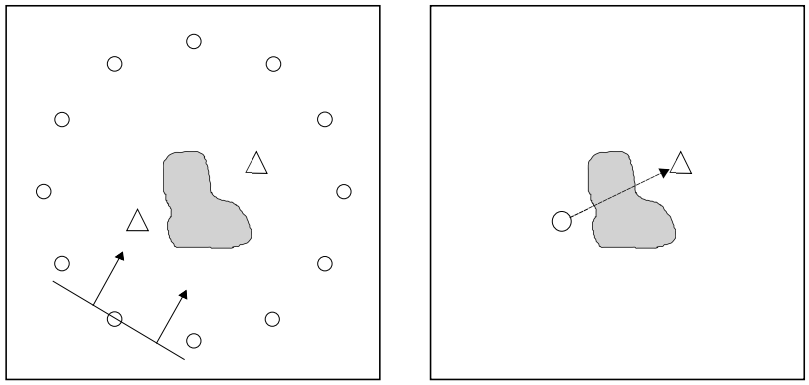
Besides, as the seismic waves bring complications to the method because of their refraction and bending trajectory, they also can bounce off of sharp boundaries such as core-mantle.

2.1.2 Ambient Noise Tomography

As it was explained previously, this technique uses ambient noise waves. That is, low energy signals such as the action of the ocean waves, wind, rock fracturing and anthropogenic activity. The objective is to construct an inference of the subsurface of the Earth identifying faster and slower speed zones.

Consider figure 2.2a. Two receivers (seismometers represented by two triangles) are surrounded by several sources. As the plane-waves travel and pass through the pair of stations, the time it takes to that signal to travel from one station to another can be estimated from the cross-correlation between recordings. In that way, it is possible to have an estimation of the main speed associated to the ray path that join those stations.

Besides, theoretical studies have shown that this cross-correlation signal provides an estimation of the Green's Function (i.e Impulse Response of the system), which represents a seismogram that one station could have recorded if the source was located on the other station, also called artificial surface wave (figure 2.2a) [4].



(a) (b)

Figure 2.2: (a) Two stations (triangles) and several sources (circles) travelling from different directions. Each source represents an incoming plane-wave which travels through both stations. Using cross-correlation, it is possible to estimate the Green's Function associated to the path between them. (b) From the Green's Function it can measure the travel time as one of the receivers had been a source. This time is an average because of the heterogeneity of the medium (darker estructure).

Then, the cross-correlation signal is particularly a representation of a surface wave traveling from one station to another. Within a package of surface wave longer periods travel faster than the shorted ones. This can be represented by a dispersion curve which is a velocity versus period plot. The Ambient Noise Tomography creates a speed map for each coherence period presented within the surface wave, that is why it is so important to analyse each dispersion curve.

The benefits of the Ambient Noise Tomography:

1. It does not need to know where the sources are. Because of the surface wave estimation, all that is needed is the location of the stations, and that is a variable that is always known.
2. On relatively aseismic regions, Ambient Noise Tomography could get a better resolution model than seismic tomography.

2.2 Ambient Noise Tomography Methodology

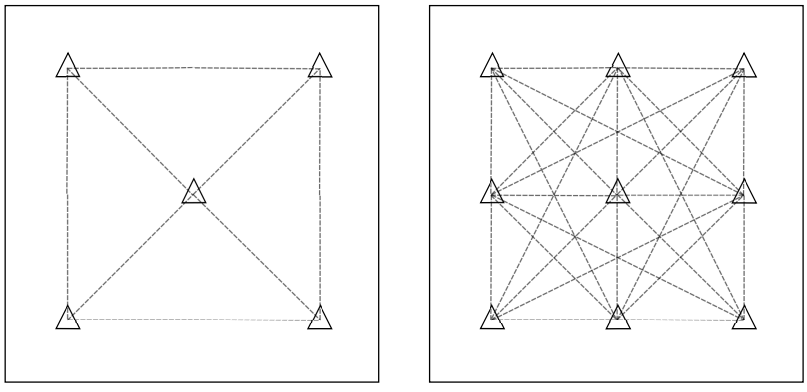
The process to obtain velocity maps from noise signals is divided in four steps: Data acquisition, Pre-processing, Inverse problem and Correction. The first one concerns about the station network setup which depends on the type of the sensor, area of study and the way stations are displayed. The pre-processing step consists on reveal travel times (observations) from de noise data. The Inverse Problem focuses on taking the observations for each station pair and create a velocity-map model from it. And the correction process makes an estimation of the ambient noise respect the azimuth. This adds extra information to the problem and adjust the result. Each of these steps is explained on detail below.

2.2.1 Data Acquisition

Station localization

Once the area of study is selected, stations must be located, and their position is ruled by mainly two parameters. Firstly, the distance between stations is related to the analysed noise wavelength, as it is indicated in equation 2.1 [12]. For example, if two stations are positioned at 12 kms from each other, it is possible to study waves with a maximum wavelength magnitude of 4 kms. Secondly, the number of station displayed is related to the resolution level of the final model. Figures 2.3a and 2.3b show the difference between two station dispositions. The more stations are available, the more ray paths are available, hence more travel time observations.

$$\Delta \geq 3\lambda \tag{2.1}$$



(a) (b)

Figure 2.3: (a) Stations (triangles) displayed on an area. Dashed lines indicate ray paths which have time travel observation associated. (b) Same escenario as (a), with the difference that more stations are displayed, that means that more observations are provided.

For the estimation of the artificial surface wave it is necessary to understand the properties of the Fresnel Zone. This is an area where sources experience constructive interference. This kind of behaviour is required in order to form a Green's Function with high SNR.

The Fresnel zone can be estimated through equation 2.2 [11]. Where Δ_{AB} represents the distance between stations, θ is the angle between the north and the direction of the incoming wave, φ is the angle between the north and the line that connects the stations and λ is the wavelength of the source. Figure 2.4 shows an schematic of the variables to consider for the study of the Fresnel zones.

$$|\Delta_{AB} cos(\theta - \phi) - \Delta_{AB}| < \lambda/2 \tag{2.2}$$

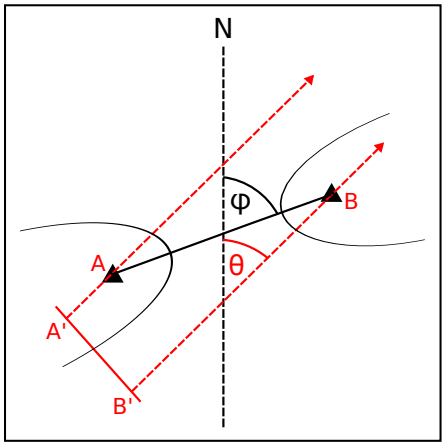


Figure 2.4: Schematic of the variables to consider for the study of the Fresnel zones. The stations are represented by black triangles. The wavefront direction (in red), θ, and φ are variables to consider to identify which sources contribute to a constructive interference.

Wave Properties

Each source implemented in the synthetic escenario will emit a a plane wave as described in equation2.3.

$$y(t,x) = \begin{cases} A\mathrm{e}^{-\alpha x} cos(kx - \omega t + \phi) & t \in [t_A, t_A + T_n] \\ 0 & elsewhere \end{cases} \tag{2.3}$$

Where:

- A is the initial wave's amplitude.
- α is the medium attenuation factor. To simulate a lossless media, this parameter must be set at cero.
- x is the traveled distance.
- t is the time.
- k is the wave number.
- ϕ is the phase.

- t_A is the arrival time.
- T_n is the width of the signal in seconds. The sent signal can be a pulse or several of them as the parameter T_n is set.

Each of this sources emits a perpendicular polarized plane wave. In the event that a wave (Iω) faces a medium change, one part of it travels through the second medium (Tω) and the other reflects (Rω) as showed in figure 2.5. With the objective of computing the properties of each of those resulted waves, the Fresnel's Equations are used (2.4 - 2.8).

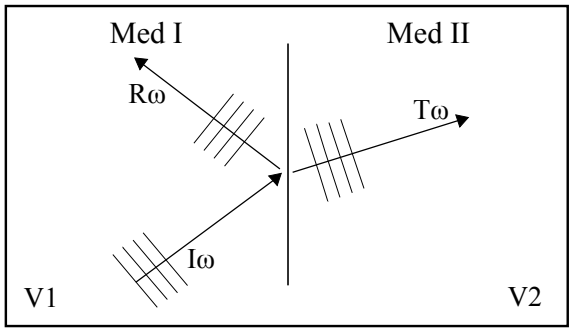


Figure 2.5: The incoming wave (Iω) splits into two new waves. One part transmits through the second medium (Tω) and the other one reflects (Rω). Their properties are governed by the Fresnel's equations.

$$I\omega = A_{\mathrm{i}} cos(\omega t - kx - \phi) \tag{2.4}$$

$$R\omega = A_r cos(\omega t - kx - \phi) \tag{2.5}$$

$$T\omega = A_t cos(\omega t - kx - \phi) \tag{2.6}$$

$$A_r = \frac{n_{\mathrm{i}} cos(\theta_{\mathrm{i}}) - n_t cos(\theta_t)}{n_{\mathrm{i}} cos(\theta_{\mathrm{i}}) + n_t cos(\theta_t)} \tag{2.7}$$

$$A_t = \frac{2n_{\mathrm{i}} cos(\theta_{\mathrm{i}})}{n_{\mathrm{i}} cos(\theta_{\mathrm{i}}) + n_t cos(\theta_t)} \tag{2.8}$$

$$\theta_c = asin\left(\frac{v_1}{v_2}\right) \tag{2.9}$$

Where:

- n_{i} and n_t represent the refraction indices for the incidence and transmission medium respectively.
- θ_{i} and θ_t correspond to the incidence and transmission angles respectively.

Another consideration is the critical angle. Expression in 2.9 shows how it is computed. If an incident angle is higher, there is total reflection. However, this phenomenon presents only when $v_1 < v_2$.

2.2.2 Pre-processing: Superficial waves estimation

The Pre-processing stage is divided in several steps or phases. In this work it is used the sequence suggested by [1] which is described in figure 2.6.

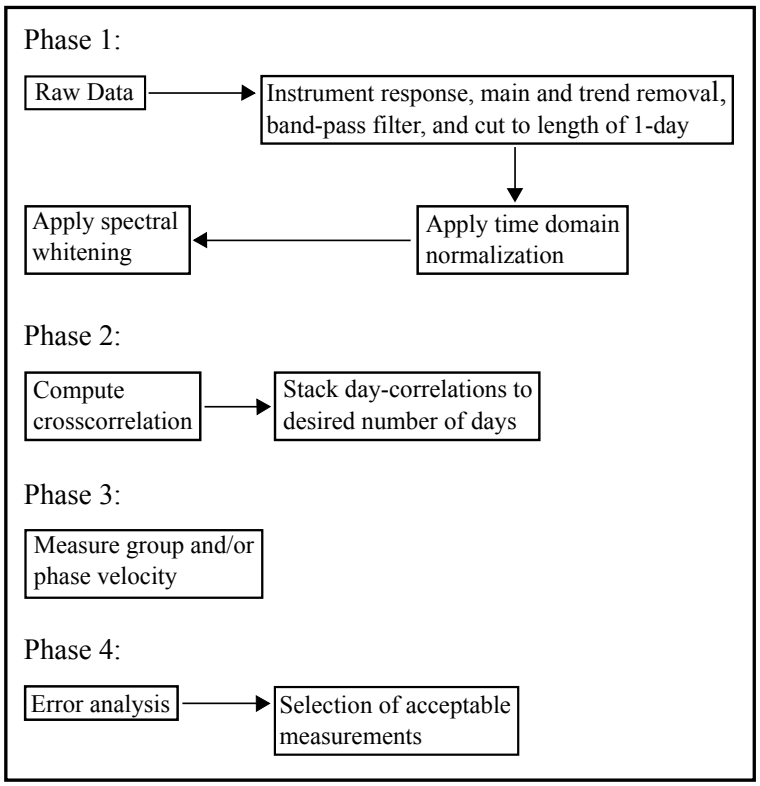


Figure 2.6: The pre-processing stage is divided in four phases which reveal the travel time from noise raw data.

Phase 1

It is focused on attenuate or remove all unnecessary data from the recordings. The sensors which form the network do not always have the same instrumental properties. It is better to work with the raw data by eliminating each sensor response from the recordings. Then, main and trend are removed from each record in order to fix all stations' data. As no all frequencies are of interest, a band pass filter is applied. Finally, the data is divided in traces of one day length. This is a standard measure with the objective of increase the SNR on the cross-correlation signals by the average of several traces (stacking).

The time domain normalization attenuates high energy signals such as earthquakes and increase the amplitudes affected by absorption. There are several techniques to achieve this goal explained in detail on [1]. In this work it is going to be used the one-bit normalization. This procedure is easy to implement and, according to the literature, has brought the best results in comparison to others such as water level and clipped waveform method.

The spectral normalization, known as whitening, forces the raw-data spectrum to be as

white as possible in order to simulate the impulse response. To accomplish this, the amplitud of the spectrum is modified by force without changing the phase. This technique is used in order to approximate the data to an impulse response of the system.

Phase 2

It is responsible for computing the cross-correlations and increase its signal to noise ratio (SNR) through stacking. This technique consist on taking one-day correlation signals and compute an average in order to highlight the high energy component of the signal. Therefore, higher amount of recordings, makes possible to the SNR increase. The size of the recording used on this procedure can be from months to years. This depends on the presence of noise on the area.

Phase 3

The travel time from the cross-correlation signals is measured. There are two most known ways to obtain the arrival time from those signals. The first one is by locating the peak of the Empirical Green's Function, which is computed from the cross-correlation signal as is shown in equation 2.10 [11]. This maximum value indicates the time travel of the signal which have been demonstrated mathematically by [9]. The second one involves the analysis of the spectrum of the cross-correlation signal. The phase (ϕ) of the coherent frequency of interest is computed and the travel time is obtain through equation 2.11 where ω represents the angular frequency. Once the time is obtained the velocity is estimated since the station distance is known.

$$\frac{\mathrm{d}C_{AB}(\omega,t)}{\mathrm{d}t} = -\tilde{G}_{AB}(\omega,t) + \tilde{G}_{AB}(\omega,-t) \tag{2.10}$$

$$\delta t = \delta\phi/\omega \tag{2.11}$$

Phase 4

Finally, the error analysis in phase 4 seeks to avoid erroneous velocity measurements. One way to do that is ensure that all pairs of stations respect the remoteness rule for the period being analyzed. Another way is by comparing velocity measurements with earthquake analysis of another accepted readings [10].

2.2.3 Inverse Problem

Once all travel times have been estimated for each station pair, it is possible proceed with the inverse problem. The first step consist on establish a grid on the area of study in order to divide it in smaller sections called cells. Each trajectory is divided in sections by the grid as it is observed in figure 2.7. Each cell has an associated velocity, therefore the sum of each section divided by the respective cell's velocity is equal to the original travel time associated to the station pair. In order to establish a linear relation, this methodology uses slowness which corresponds to the reciprocal of the speed as shows equation 2.12. Here $\mathrm{d_{i}}_j$, $V_{\mathrm{i}j}$ and $S_{\mathrm{i}j}$ correspond to the distance, velocity and slowness associated to cell j of the station pair i, and T_{i} corresponds to the total travel time.

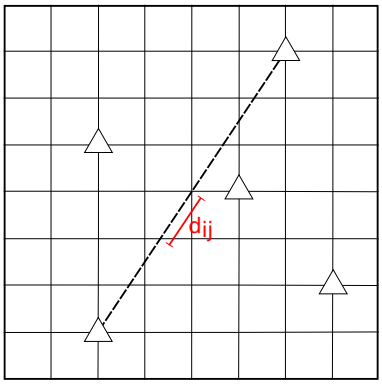


Figure 2.7: On ambient noise tomography the area is divided in a grid. The inverse problem seeks to estimate the wave velocity at each cell.

$$T_{\mathrm{i}} = \sum_{j=1}^{n} \frac{\mathrm{d_{i}}_j}{V_{\mathrm{i}j}} = \sum_{j=1}^{n} \mathrm{d_{i}}_j S_{\mathrm{i}j} \tag{2.12}$$

Equation 2.12 is raised for each pair of stations. That way the inverse problem is established as indicated by equations 2.13 to 2.16, where G corresponds to a mxn matrix which contains the trajectories per cells associated to each pair of stations, m is the parameter array of n unknown variables (in this case the value of slowness of each cell of the grid), and d is a mx1 array which contains the observations (in this case travel times associated to each cell involved in the trajectory).

$$Gm = \mathrm{d} \tag{2.13}$$

$$G = \begin{bmatrix} \mathrm{d}_{11} & \mathrm{d}_{12} & x_{13} & \dots & \mathrm{d}_{1n} \\ \mathrm{d}_{21} & \mathrm{d}_{22} & x_{23} & \dots & \mathrm{d}_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \mathrm{d}_{m1} & \mathrm{d}_{m2} & \mathrm{d}_{m3} & \dots & \mathrm{d}_{mn} \end{bmatrix} \tag{2.14}$$

$$m = \begin{bmatrix} t_1 & t_2 & t_3 & \dots & t_m \end{bmatrix}^T \tag{2.15}$$

$$\mathrm{d} = \begin{bmatrix} s_1 & s_2 & s_3 & \dots & s_n \end{bmatrix}^T \tag{2.16}$$

Since trajectories do not go through all the cells of the grid, this is an ill-conditioned inverse problem. To solve this issue it is necessary to add information to the problem. Firstly, the speed value of each cell is not highly different from the values of its neighbors, this property is called *smoothness*. Secondly, with the objective of forcing the parameters to do not adopt unreal values a likeness norm is applied, this is called *damping*. Therefore, a new penalty function is achieved which represents an over-conditioned inverse problem in equation 2.17.

$$S(m) = (Gm - \mathrm{d})^T C_{\mathrm{d}}^{-1}(Gm - \mathrm{d}) + \varepsilon(m - m_o)^T C_m^{-1}(m - m_o) + \eta m^T L^T L m \tag{2.17}$$

Where:

- C_{d} is the covariance matrix which includes prediction and measure error.
- m_o is the reference model. It indicates the values which the result should be approached.
- C_m is the covariance matrix between estimation and reference model.
- ε and η are the weight of damping and smoothness components.
- L is the matrix which forces to each cell to have similar speed values with its neighbors.

Least squares problem

In order to solve the inverse problem, some decompositions must be done. Using the Cholesky decomposition, C_{d} and C_m are rewritten as follows in equation 2.18 which establishes the optimization problem [6], [3], [5].

$$\min_m (Gm - \mathrm{d})^T W_x^T W_x (Gm - \mathrm{d}) + \varepsilon(m - m_o)^T W_m^T W_m (m - m_o) + \eta m^T L^T L m \tag{2.18}$$

This is solved defining new variables

$$F = \begin{bmatrix} W_x G \\ \varepsilon W_m \\ \eta L \end{bmatrix} \qquad D = \begin{bmatrix} W_x \mathrm{d} \\ \varepsilon W_m m_o \\ 0 \\ \dots \\ 0 \end{bmatrix} \tag{2.19}$$

Then, using least squares, the optimal matrix parameter is given by equation 2.20.

$$m_{est} = (F^T F)^{-1} F^T D \tag{2.20}$$

Smoothness setup

The Laplacian operator is chosen to implement the smoothness component of the problem. Therefore, the L matrix of 2.18 corresponds to a representation of this operator. This is defined by the finite difference stencil represented in figure 2.8. A weight is associated to each cell and to its neighbors in order to enforce similarity between them. There are three stencil cases: Corner, centred and border cells described in the Figure. With this configuration L is set as a sparse matrix [5].

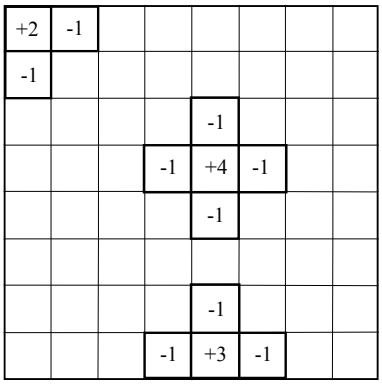


Figure 2.8: The stencil technique associates a weight to each cell and to its neighbors in order to enforce similarity between them. The image illustrates the weight values used in the three possible cases: corner, centred and border cell.

Weight selection

With the objective of selecting appropriate values for ε and η, on this work the L-curve method is explored. The objective is to trace a curve comparing the residual norm ($||Gm - \mathrm{d}||_2^2$) and the smoothness/damping component (L_{error}) [5]. In this way it is easy to see the compromise between the values of these two quantities and select an appropriate regularization parameter. This method follows the procedure explained below:

1. A range of values for ε is selected. This values are selected as a guess. Posterior results will indicate if it is necessary a modification.
2. The optimization problem is modified as equation 2.21 shows and the optimization problem is solved for each ε value.

$$\min_m \underbrace{(Gm - \mathrm{d})^T W_x^T W_x (Gm - \mathrm{d})}_{RMSE} + \varepsilon \underbrace{(m - m_o)^T W_m^T W_m (m - m_o)}_{L_{error}} \tag{2.21}$$

3. The root mean square error (RMSE) and L_error (error associated to the a prior information) are calculated and graphed as figure 2.9 shows.
4. A range of values is selected which are located in the convex section of the curve in order to optimize the error values. If a curve of these characteristics is not generated from this procedure, it means that the range of values chosen must be changed.
5. Analogous procedure is used to select the values for η.

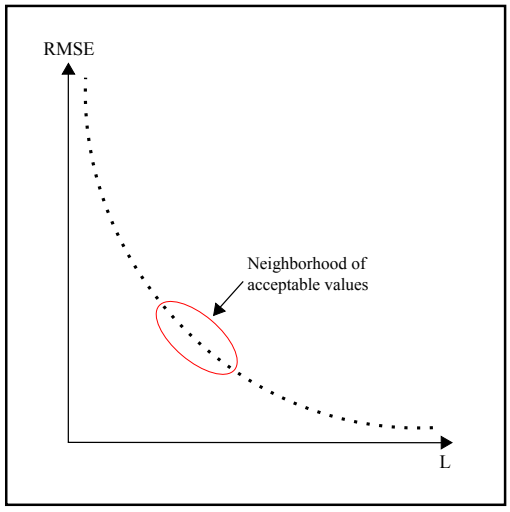


Figure 2.9: The L-curve method estimates the root mean square error and a priori component error for each parameter value (ε or η) in order to identify a neighborhood of aceptable values for the inverse problem.

2.2.4 Correction Process

Once the model is estimated using the least squares method, it is needed a correction stage. [11] suggests that uneven distribution of ambient noise energy(E_p) could produce a bias in the estimation of speed values. Moreover, even if the distribution of noise is uniform, it is recommended to make the correction process anyway.

This distribution is unknown and has a non linear relationship with the cross-correlation signals as it is shown in equations 2.22 and 2.23. Nevertheless, [11] affirms that E_p can be estimated from the synthetic and empirical cross-correlations signals which involves an

inversion. Therefore, it is possible to compute a bias and correct all estimated phase velocities.

$$C_{AB}(\omega, t) = \int_o^{2\pi} E_p(\omega, \theta) cos[\omega(t - \delta t)] H(t, \delta t) \mathrm{d}\theta \tag{2.22}$$

$$H(t, \delta t) = \begin{cases} 0.5\{1 + cos[2\pi(t - \delta t)/T^*]\} & t \in [\delta t - T^*/2, \delta t + T^*/2] \\ 0 & elsewhere \end{cases} \tag{2.23}$$

Where:

- C_{AB} is the Green's Function between stations A and B.
- E_p is the noise distribution of energy.
- ω corresponds to the frequency.
- θ is the angle between the wave direction and the north.
- δt is the time travel between stations A and B.
- H is the time domain taper function for the cross-correlation signal.
- T^* is the time width of H.

The process involves 4 steps:

1. Obtaining the 2D phase velocity map. For the first iteration, the speed between stations is obtained through the empirical Green's Functions. For the following iterations, they are achieved through the update of the phase velocities in Step 3.
2. Using the inversion problem proposed in [11] it is possible to estimate E_p. The starting point is the 2D velocity map from Step 1.
3. Estimate the bias ($\mu^{(k)}$) by comparison of the empirical and theoretical Green's Function (using equations 2.22 - 2.25). This parameter is used to correct the velocities as equation 2.26 indicates.
4. If velocities or E_p converges, the iterations stop.

$$\delta\varphi_{AB}(\omega) = \Theta[\tilde{G}_{AB} W(t, \Delta_{AB})] - \Theta[G_{AB} W(t, \Delta_{AB})] \tag{2.24}$$

$$\mu = \frac{\delta\varphi_{AB}(\omega)}{\omega t_{AB}} \tag{2.25}$$

$$c^{(k+1)} = \frac{\tilde{c}_{AB}}{(1 + \mu^{(k)})} \tag{2.26}$$

2.3 Background

Ambient noise tomography has been used worldwide on several mediums of different scales. For instance, [4] applied this technique on the British Isles achieving the first wave group velocity maps of the Scottish Highlands. [11] demonstrated how ambient noise anisotropy influences in the computation of phase velocities on a study at the Tibet.

Within the pre-processing stage, there are several techniques whose objetive is to eliminate the influence of seismic signals [1]. Nevertheless, the one-bit normalization is a widely accepted technique because of its good results and easy implementation.

Once the cross-correlation signals have been computed, the stacking stage takes place. It is normal to use months or even years of data in order to achieve a SNR value as high as possible.

A key point in the inversion problem stage is how to choose the parameters which influence in how much it is possible to believe on each a prior information: damping and smoothness. A common practice is to work with the damping component and leave the smoothness out of the equation. However, this restriction is used by implementing filters to the final model. With this procedure, seismologist save time and they do not have to concern about how to pick two parameters for one problem. They just have to take care of one, the damping weight. The most known technique that allow to choose an adequate weight is the L-curve. Nevertheless, in practice, the way this value is chosen is by guessing. Then, the observation of the resulted model will tell us how to modify this parameter in order to achieve a model which is close to the observations. This procedure is adopted to obtain low computational cost process.

This entire procedure was used to estimate a model of a mine in northern Chile. This work is looking for finding out if this can be improved. The selection of parameters through the heuristic method can be seen as a blind technique. It is necessary to find out if techniques like the L-curve can deliver better results.

Chapter 3

Synthetic scenario

In this chapter, different techniques used widely in seismic tomography articles are adopted in order to create different synthetic scenarios. The criteria that is followed in order to select these techniques are:

- *They have to be able to perform simulations under any conditions*: This platform can display diverse cases when varying parameters of the source (e.g. amplitude, frequency, duration), station's location, subsurface structure (whether homogeneous or heterogeneous), among others.
- *They count only contrasted to a ground-truth*: This in order to compare a variety of scenarios' performances with a given reference.

Ambient noise tomography counts with four stages which are implemented in this platform. These range from simulating the stations' recordings, to estimating the structure of the subsurface with its respective correction process.

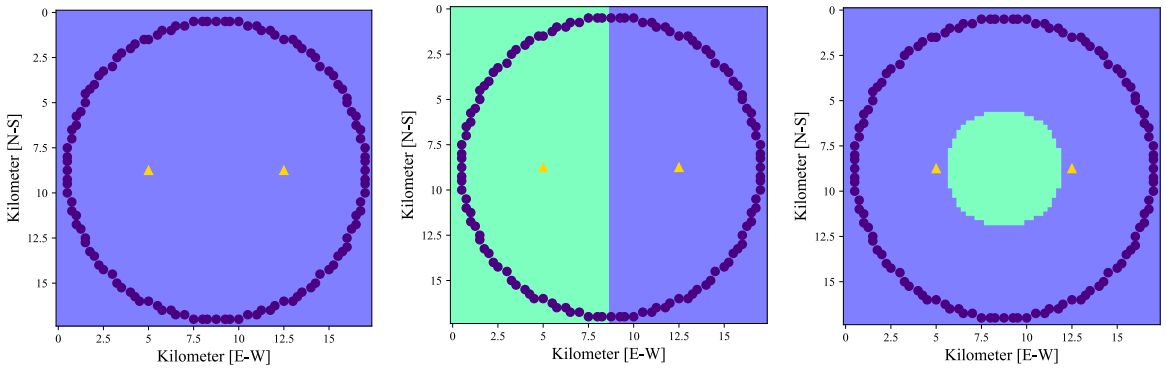
3.1 Data acquisition

To simulate the data acquisition, the procedure proposed by [9] will be followed. It consists in two steps. The first one considers the estimation of the stations' recordings under different conditions: homogeneous or heterogeneous media, lossy/lossless media, among others. The second one corroborates that the simulated measurements are consistent with the imposed scenario.

Figure 3.1 shows the three implemented scenarios. Stations are represented by yellow triangles and surrounded by uniformly distributed sources (purple dots). Each source emits a plane wave pulse, this approximation is acceptable under the assumption of real sources which are distant from the stations [11].

Each scenario involves several simulations or tests. Several variables change from test to test such as: number of stations, number of sources and size of the cells.

One vital point to keep in mind, is the boundary shown in equation 2.1. If this restriction is not respected while performing the simulations, the results contain an error that might be considerable, reducing the reliability of the computation.



(a) (b) (c)

Figure 3.1: Scenarios implemented in the synthetic study. Stations are represented by yellow triangles, while sources are represented by purple dots. (a) Homogeneous media (b) Heterogeneous media displayed as a medium change (c) Heterogeneous media. Structure inserted in a homogeneous environment.

3.1.1 Homogeneous scenario

For this type of synthetic scenario (and the next ones) each source represents an incoming plane wave. This means that the wave front points which are going to reach each station are calculated as figure 3.2 suggests. For illustration, the image shows a fewer amount of sources than the used in the simulations. This is a more realistic way to calculate the arrival times, different from the one used in [9]. As the medium is supposed homogeneous, there is no diffraction in the trajectory.

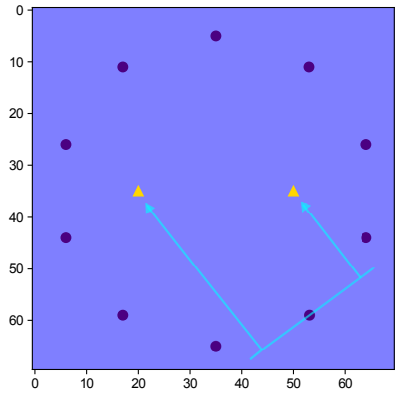
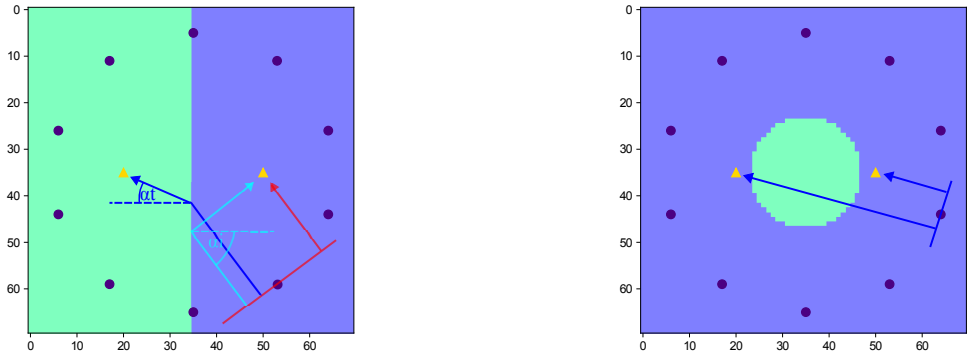


Figure 3.2: Each wave front point that reaches a station is calculated.

From each source, a single pulse is emitted. Since the velocity in the medium is a known factor, it is possible to compute the travel times priorly. The platform is able to simulate media with and without loss. Equation 2.3 expreses the plane wave. To simulate a lossless media, the parameter α must be set at zero.

3.1.2 Heterogeneous scenarios

For the implementation of the synthetic scenario 3.1b, there are several considerations taken into account. Each wave front is infinite and every point of it reaches the medium change. From each of these a direct (red), reflected (cyan) and transmitted (blue) wave will be created as it is shown in figure 3.3a. Equations 2.4 to 2.9 explain the behavior of the signals through the medium.



(a) (b)

Figure 3.3: (a) Plane waves divide into transmitted and reflected wave. In the case the incident angle is bigger than the critical, there is no transmission to the second medium (total reflection). (b) For simplification, waves follow a straight path. Each path goes through several cells. Each trajectory is divided in sections in order to estimate the distance associated to each cell.

For the second case of heterogeneous medium, no reflections or transmissions are considered for simplification as figure 3.3b suggests. The arrival time of each plane wave is computed dividing the trajectory into little sections. Each of them are associated to a cell of the grid and divided by its velocity obtaining the time the wave took to travel through it.

3.2 Pre-processing: Artificial surface wave estimation

Normally, it is necessary to eliminate all non useful data such as the trend and the response of the instrument. However, on this synthetic scenarios zero mean data is developed, with no instrument response. Then, there is no need to implement such correction.

Before computing the crosscorrelation signals, the implementation of a temporal and spectral normalization is necessary. This kind of process is not used on all synthetic data sets; just on those whose SNR require to be increased.

3.2.1 Temporal normalization: One-bit method

The implementation of the one-bit normalization is simple. All amplitudes are replaced by 1 or -1 depending of the sign of the original amplitude [1]. Figure 3.4 shows an example of a trace going through this process. As a consequence of using this technique, the attenuation that waves suffer during the journey have less impact.

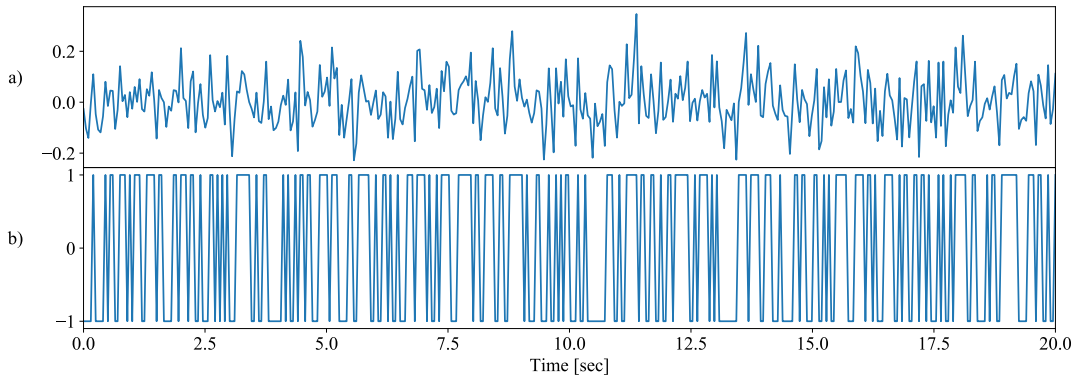


Figure 3.4: Example of One-Bit normalization method where all amplitudes of the original trace (a) are replaced by an unitary value with its respective sign as shown in (b).

3.2.2 Spectral normalization: Whitening

The objective of applying a whitening transform to the waveform is to increase the signal to noise ratio (SNR).

The implementation of the spectrum normalization consists of six stages:

1. The trace spectrum is computed by a Fast Fourier Transform.
2. For each frequency, the polar form of the spectrum is obtained. In this representation the amplitude is decoupled from the phase so that any perturbation applied to one doesn't affect the other.
3. We obtain the amplitude's envelope through interpolation. Then, the inverse of the envelope is computed.
4. The inverse of the envelope is multiplied by the amplitude. From the product the whitened spectrum is obtained.
5. The final spectrum is filtered by a low pass filter. This step is necessary in order to eliminate higher frequencies which are not wanted for the analysis.
6. The signal is returned to the temporal domain through Inverse Fast Fourier Transform.

The figure 3.5 shows an example of the use of the spectral whitening. Firstly, the original signal is presented in (a). Then, in (b) the spectrum and its corresponding envelope is shown in blue and red respectively. Finally, in (c) and (d) the resulting signal and its spectrum is displayed. In this case, the SNR increases from 0.0414 to 0.0984.

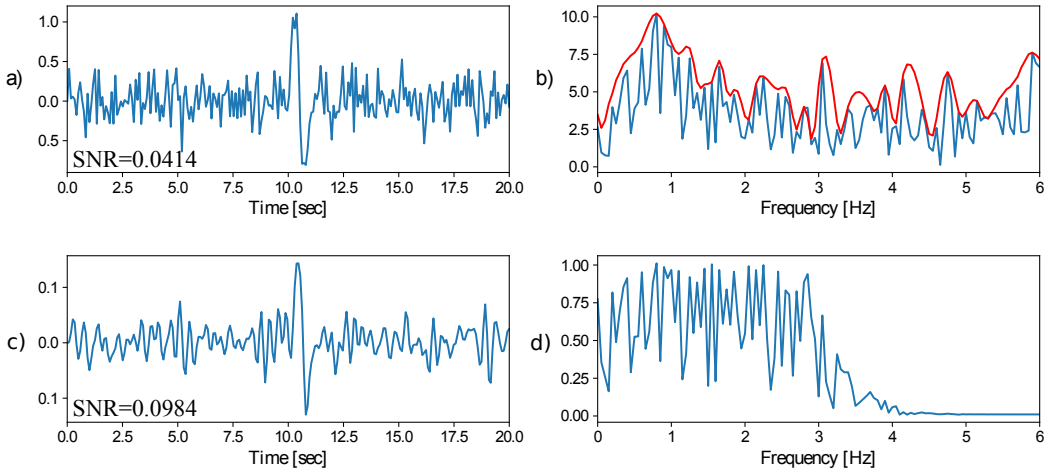


Figure 3.5: Example of Spectral Whitening method. (a) Original signal (b) original spectrum in blue and its envelope in red (c) Resulting signal after use of whitening (d) Spectrum after normalization and corresponding filtering.

3.2.3 Crosscorrelation and stacking

For the artificial surface wave estimation, the procedure proposed by [9] is followed. Firstly, the crosscorrelation signal is computed between each source of each station. Then, all cross-correlations signals are added obtaining the final cross-correlation signal.

As the first simulation represents the results from pulse sources, there is no need to use stacking because of the absence of noise.

3.3 Inverse problem

As discussed in chapter 2, the inverse problem is ill-conditioned. The prior information added to solve this issue is damping and smoothness norm. The first one is implemented setting a reasonable guess of the subsurface structure. For the second norm, a stencil matrix is elaborated as explained in chapter 2.

Then, the G matrix is constructed. An approximation method is used for this. The entire trajectory is divided into several sections (ds) as shown in figure 3.6. After counting how many sections are within each cell, it is possible to obtain an estimate of the size of the trajectory section that corresponds to the cell. The parameter vector (m) corresponds to the

matrix map written in column wise form. For a start, the Wm matrix is set as the identity matrix, but its values are changed in each iteration as it adopts the estimation errors.

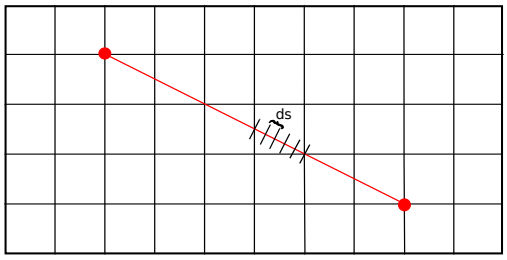


Figure 3.6: Each trajectory is divided in several sections (ds). After counting how many sections are within each cell it is possible to obtain an estimation of the size of the trajectory section that corresponds to each cell.

Before computing the model estimation, it is needed to select the weight parameters for the different norms. For this implementation, the L-curve method is used with the expression shown in figures 3.1 and 3.2. Therefore, several tests are needed in order to identify where the optimal parameter value is located.

$$\min_m \underbrace{(Gm - \mathrm{d})^T W_x^T W_x (Gm - \mathrm{d})}_{RMSE} + \varepsilon \underbrace{(m - m_o)^T W_m^T W_m (m - m_o)}_{L_{error}} \tag{3.1}$$

$$\min_m \underbrace{(Gm - \mathrm{d})^T W_x^T W_x (Gm - \mathrm{d})}_{RMSE} + \eta \underbrace{m^T L^T L m}_{L_{error}} \tag{3.2}$$

The inverse problem is solved iteratively. Each cycle consist on estimating the model and applying the correction explained in section 3.4.

3.4 Correction and Adjusting results for visualization

The estimation of the ambient noise distribution is made following the procedure described in section 2.2.4. As a reference model (m_o) an homogeneous area was chosen. This process is finished when the energy distribution of noise or the velocities converge.

After the final estimation is obtained, a gaussian blur is used to give more uniformity to all the sections of the map. The dimensions of the filter vary according to the size of the grid. Finally, the posible velocities rage is divided in clusters. In this way it is easier to visualize zones with similar velocity and possibly of the same material.

Chapter 4

Results and discussion

In this section, the results of the implementation of the ambient noise tomography are presented.

Firstly, the recordings of each station are simulated for each scenario and with them, the travel times. Secondly, the inversion problem is implemented.

4.1 Data Acquisition

4.1.1 Homogeneous Medium

Figure 4.1 shows the first scenario of study. It consists of two stations surrounded by 4.5 Hz plane wave sources. Each cell of the grid has a size of 250 by 250 meters. Therefore, the stations are 7.5 kilometres away from each other. Given those parameters, the restriction imposed by equation 2.1 is respected.

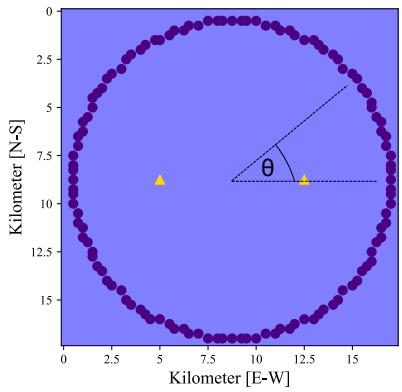


Figure 4.1: First synthetic scenario. Two stations (yellow triangles) located on an homogeneous medium surrounded by plane wave sources (purple dots). θ represents the angle of each source location with respect to the horizontal line.

Figure 4.2 show the recordings in function of the source location (each source is characterized by its angle, θ, as figure 4.1 shows). Despite this experiment was performed with 500 sources, just a few of them are displayed for simplicity.

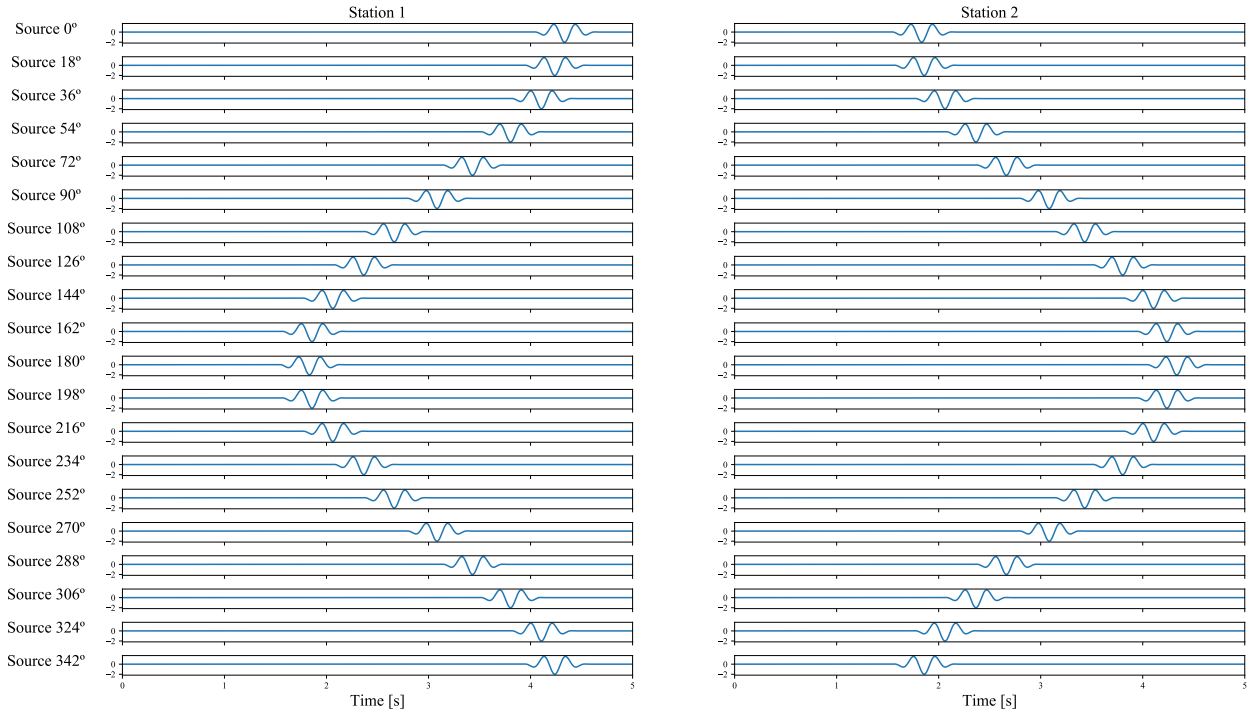


Figure 4.2: Recordings from two stations in function of the source position for an homogeneous case. Each signal represents the system response to 5 Hz pulse sources sent through a lossless media with a velocity of 3 km/s.

In the next step, the records of each station are cross-correlated by source. Figure 4.3 shows the resulting signals. Each raw of the image represents a cross-correlation function seen from above. Using the parameters of this scenario, it was possible to compute the opening angle of the Fresnel zone. The signals associated to sources inside this area are inside the red boxes.

It is easy to notice that those signals interact in a constructive interference which constructs the peaks of the final cross-correlation signal. The rest of the signals do not contribute any useful information; in fact they interfere destructively with each other. The result of the stacking process which represents the system response is shown below the cross-correlation image from the figure.

For this case the estimated velocity between the stations resulted in 2.97 km/s (0.86% of error).

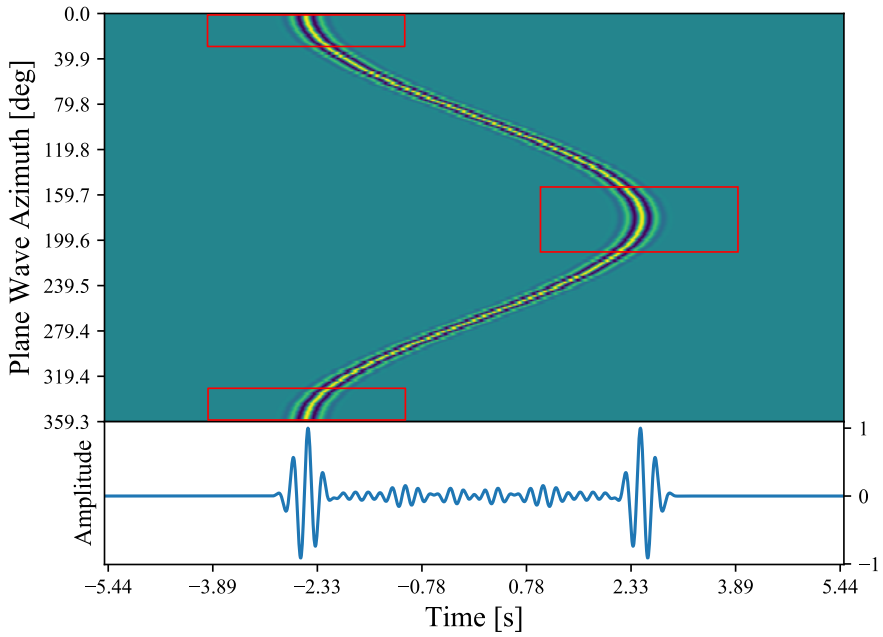


Figure 4.3: Crosscorrelation signals computed by source. Signals which interfere with each other constructively are inside the red boxes. The rest of the signals do not contribute with useful information. After stacking all the crosscorrelation signals, it is obtained the response of the system between the stations (trace below the image).

In figure 4.4 there is a comparison between EGF (blue) and theoretical GF (red) proposed in [11]. For this, it was computed an average between the causal and acausal part of the signal. There is a perfect match inside the coherent part of the signal which shows the high performance of the technique.

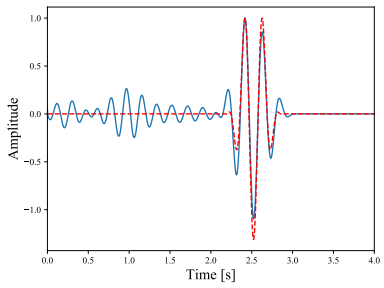


Figure 4.4: Comparison between Empirical Green's Function (blue) and theoretical Green's Function (red) for the case of homogeneous medium.

4.1.2 Heterogeneous Medium 1

The results of the implementation of the first heterogeneous medium are explored. Figure 4.5 displays an environment with a change of phase. Two materials with different velocities are simulated. As in the previous experiment, the two stations are surrounded by 500 plane wave sources.

Figure 4.6 displays the stations response to a 4.5 Hz incoming wave. As discussed in the previous section, because of the change of medium, other kind of waves are received by the stations. Because of the reflection that waves suffer in the interface. between one medium and the other, there are cases where stations receive more than one pulse. When the incoming wave has a direction greater than the critical angle, it suffers complete reflection. That is why, there are cases where stations do not record some of the sources.

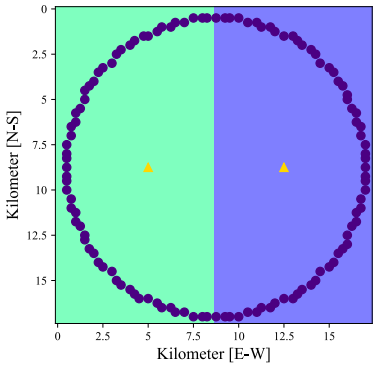


Figure 4.5: Second synthetic scenario. Two stations (yellow triangles) located on an heterogeneous medium surrounded by plane wave sources (purple dots). The media is composed by two materials with different velocities.

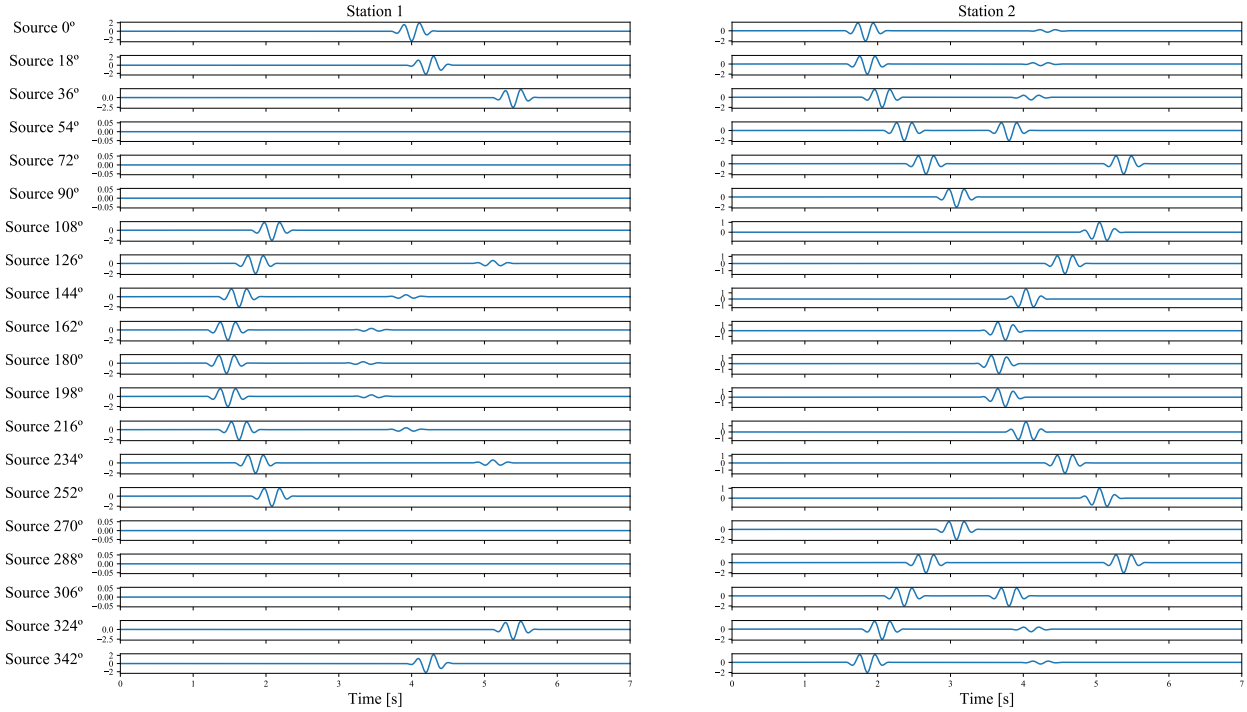


Figure 4.6: Recordings from two stations in function of the source position. Each signal represents the system response to 4.5 Hz pulse sources sent through a lossless media with a change of medium.

In figure 4.7, the crosscorrelation signals by source can be observed. Signals inside the red

boxes correspond to sources inside the Fresnel zone. The form of the stacked signal is more chaotic than in the previous case. The signals resulted from reflection presents constructive interference, but with low energy so it does not interfere with the process. The estimated velocity was of 3.5 km/s (0.25% of error).

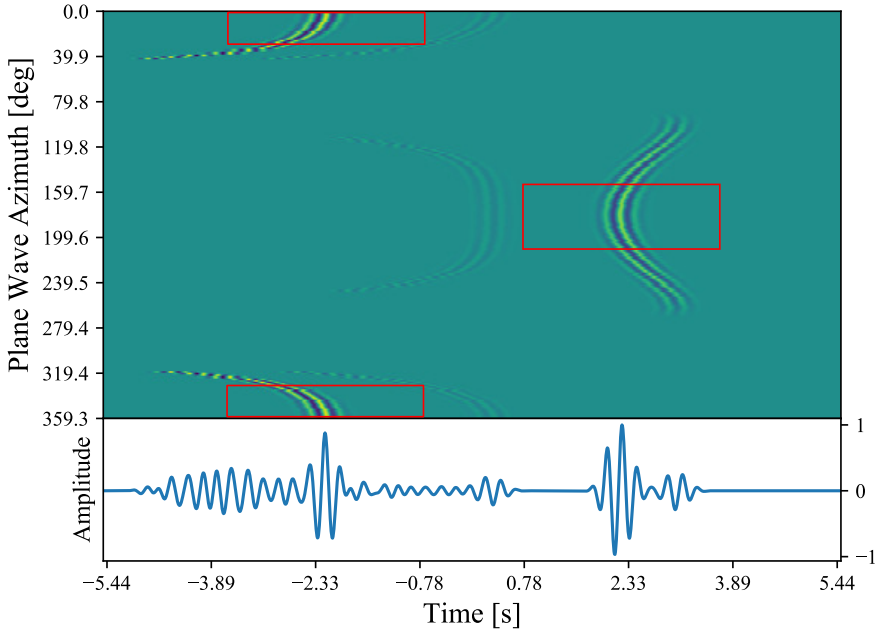


Figure 4.7: Crosscorrelation signals computed by source. Signals which interfere with each other constructively are inside the red boxes. The rest of the signals do not contribute with useful information. After stacking all the crosscorrelation signals, the response of the system between the stations (trace below the image) is obtained.

In figure 4.8 there is a new comparison between EGF (blue) and theoretical GF (red) for this scenario. There is a perfect match inside the coherent part of the signal.

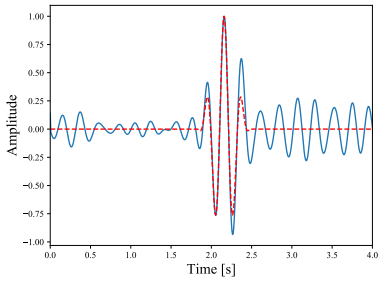


Figure 4.8: Comparison between Empirical Green's Function (blue) and theoretical Green's Function (red) for the case of heterogeneous medium 1.

4.1.3 Heterogeneous Medium 2

The second heterogeneous medium consists of a structure immersed in an homogeneous medium as figure 4.9 displays. Firstly, the response of two stations located on the area is

analyzed. In figure 4.10 the recordings from each station by source can be observed. In station 2, for sources located between 126^o and 234^o , there is an odd behavior in the arrival times such as station 1 between 0^o and 43^o. In this range, the plane waves travel through the structure, therefore the recordings adopt this form.

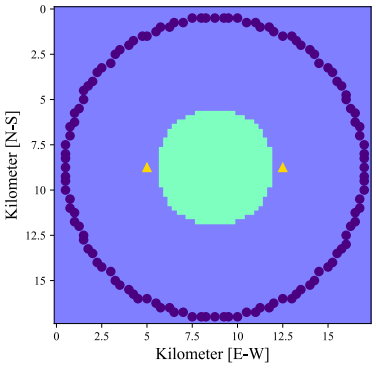


Figure 4.9: Third synthetic scenario. Two stations (yellow triangles) located on an heterogeneous medium surrounded by plane wave sources (purple dots).

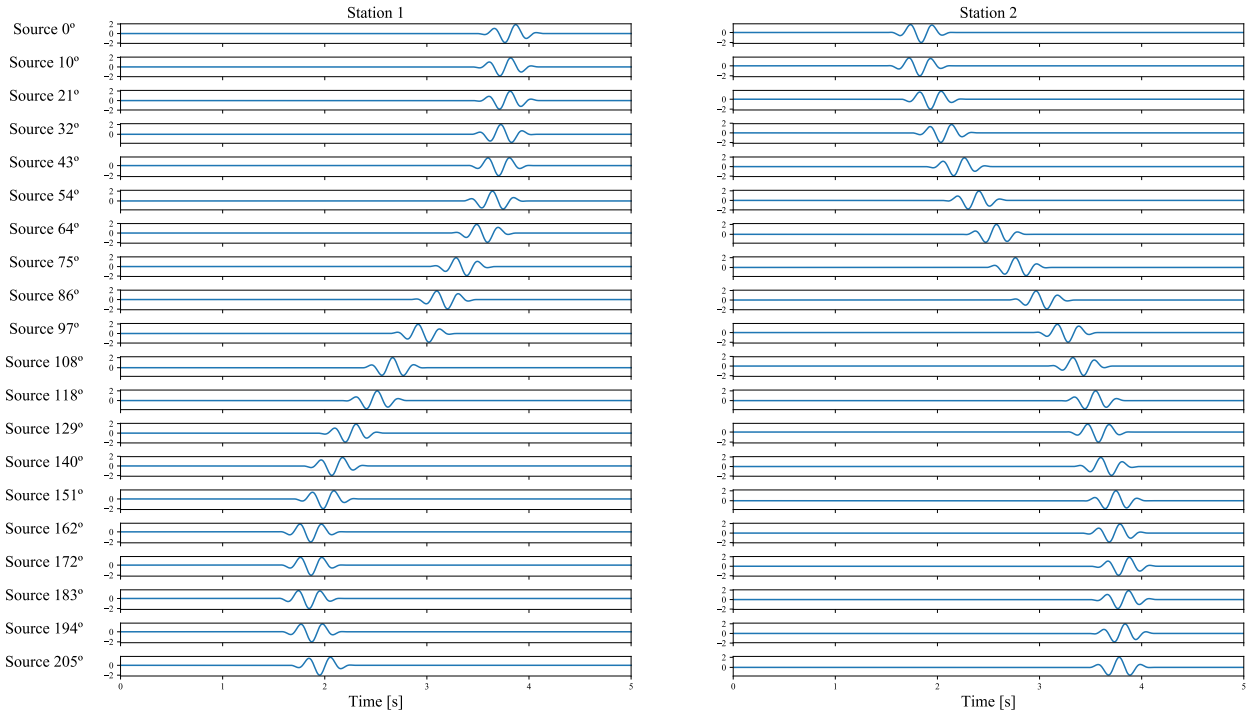


Figure 4.10: Recordings from two stations in function of the source. Each signal represents the system response. Sources send a 4.5 Hz pulse through a lossless heterogeneous media.

In figure 4.11, the cross-correlation signals by source that corresponds to this model are presented. Signals from sources inside the Fresnel zone are inside the red boxes, while the final cross-correlation signal resulted from the stacking process is below the cross-correlation image. The velocity computed was of 3.9 km/s (0.34% of error).

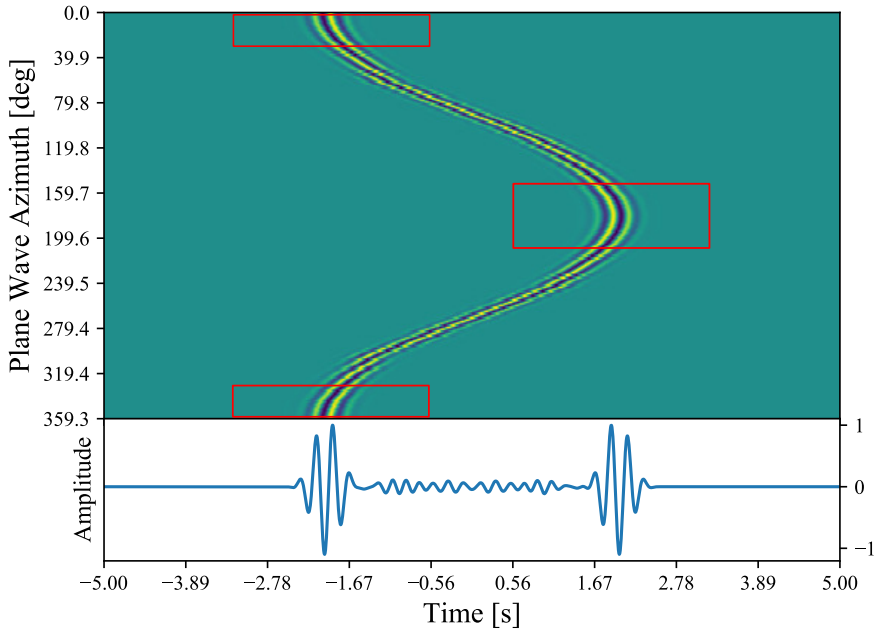


Figure 4.11: Cross-correlation signals computed by source for the second case of heterogeneous medium. Signals which interfere with each other constructively are shown inside the red boxes. The final crosscorrelation signal resulted from the stacking process is displayed below the image.

In figure 4.12 there is a comparison between EGF (blue) and theoretical GF (red) for the second case of heterogeneous medium. Again, there is a perfect match inside the coherent part of the signal.

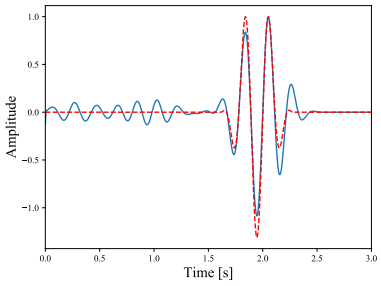


Figure 4.12: Comparison between Empirical Green's Function (blue) and theoretical Green's Function (red) for the case of heterogeneous medium 2.

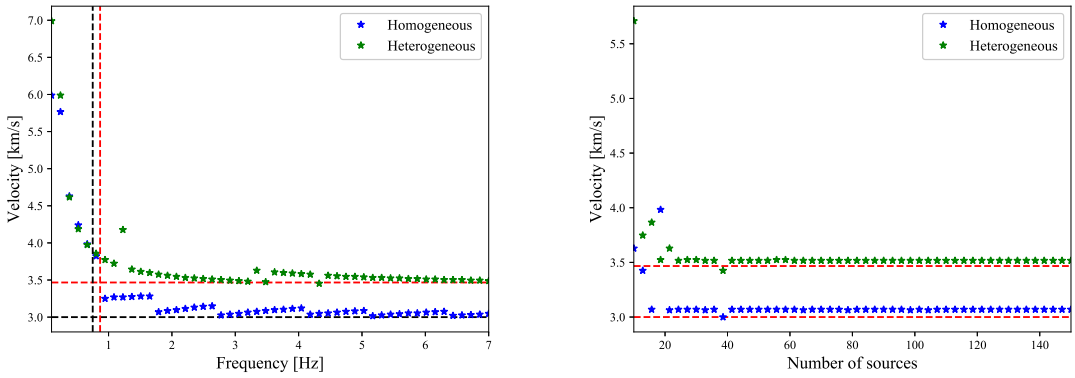
4.1.4 Parameter sensitivity analysis

The speed estimation between two stations highly depends on the parameter setup. Here two characteristics are analyzed: station distance and source density. For this analysis, we use two kind of mediums: homogeneous and heterogeneous 1.

Firstly, the scenario presented in 4.1 is used. In it, stations are 7.5 kilometres apart in a medium where waves move at 3 km/s. Given this conditions, the minimum frequency

possible to analyze is 1.2 Hz. However, lower frequencies are studied in order to display the consequences of not respecting the boundaries. The estimation of speed is repeated, and in each iteration the frequency of the source is changed. This process is used in the heterogeneous scenario 4.5. The results are displayed in figure 4.13a. The estimation results related to the homogeneous and heterogeneous scenarios are displayed in green and blue, respectively. The horizontal lines represent the target velocity value and the vertical ones indicate the critical frequency, explained by equation 2.1. The results show how the error increases while the frequency approaches to its critical value in both cases.

Secondly, another experiment is developed. Figure 4.13b show how the estimation error increases due the low density of the source. This density is expressed as the number of point sources located around the stations. They represent a plane wave generator. In addition, this experiment shows how the excessive increment of the emitters does not give any additional information in both cases.



(a) (b)

Figure 4.13: (a) Estimation of velocity between two stations in function of the frequency of the source. When the frequency analyzed is closer to the critical one, the estimation error increases. (b) Velocity estimation in function of the sources density. The more sources are inside the Fresnel Zone, the more accurate the procedure is.

4.2 Inversion Problem, correction and visualization

The estimation of the velocities between two stations is easy to achieve. However, when an interest for revealing the complete model of an area appears more stations are needed and an inversion problem has to be established. The following simulation outcomes were generated for the previous synthetic scenarios, with the difference that the number of stations was increased.

Those results are later going through the correction and visualization process explained in section 3.4.

4.2.1 Homogeneous medium

On an uniform area, 25 stations are displayed as figure 4.14 exhibits. As in the previous section, they are surrounded by 500 plane wave sources. The grid is composed by cells of 250 by 250 meters. For each possible pair of stations, a travel time is estimated (called an observation). Then, the inversion problem is established and the first model is achieved.

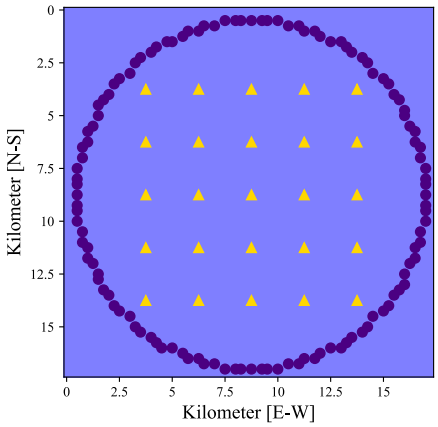
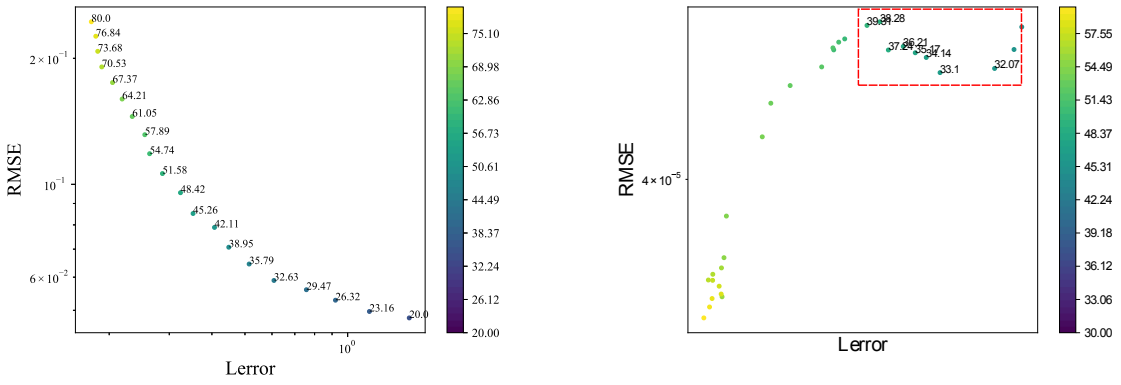


Figure 4.14: Homogeneous area with 25 stations displayed (yellow triangles). They are surrounded by plane wave sources (purple dots).

In order to implement the inverse problem, it is necessary to select the parameters values – ε and η. The L-curve method is used. Figures 4.15a and 4.15b display the resulted curves. In the first figure it is easy to identify an optimal value. However, in the case of η, the curve has an irregular shape. In this cases it is necessary to identify a section of the curve which is similar to an L-curve. Figure 4.15b shows this section inside the dashed red box. The optimal values found in this stage are 45 and 35 for the damping and smoothness weights respectively.

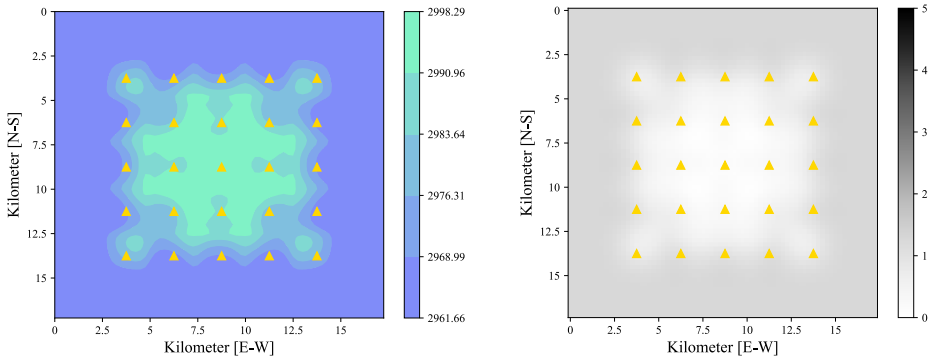


(a) (b)

Figure 4.15: (a) L-curve associated to the damping parameter ε. Given this, the optimal value selected is 45. (b) L-curve associated to the smoothness parameter η. This irregular curve requires to find a section with the corresponding L-curve shape. Then, the value selected for this weight is 35.

Using those values, the inverse problem can be solved. After obtaining the first model, the ambient noise energy distribution is computed in order to recompute the estimations. The convergence of the ambient noise as achieve after the third iteration. Figure 4.16a shows the final estimation of the map. If we observe the values in the map, they oscillate inside a narrow range of 0.037 km/s long. This is an accurate homogeneous medium estimation.

Figure 4.16b shows the percentage of error associated to the estimations. Ignoring the area outside the region of interest, the errors range between 0 and 1.5%. Despite the irregularities, this outcome is close to the real model.



(a) (b)

Figure 4.16: (a) Final estimated map for homogeneous case. It presents an homogeneous estimation. (b) Error map which displays the percentage error associated to each cell's velocity estimation.

A test is made on the same homogeneous medium, the difference being the size and the quantity of cells. This time each cell measures 600 by 600 meters. This way, the resolution

decreases. This new scenario is shown in figure 4.17. The new optimal value for ε and η is now 59 and 50, respectively.

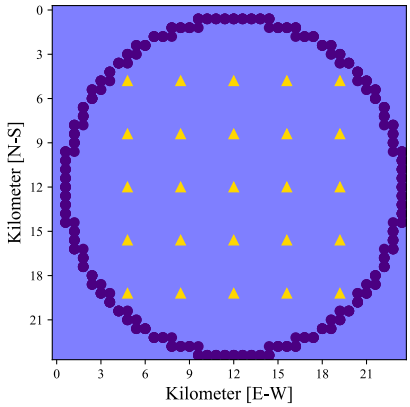
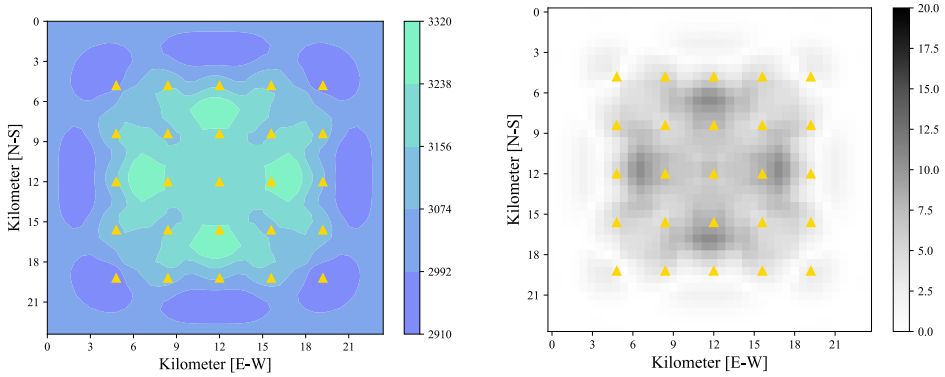


Figure 4.17: Homogeneous area with 25 stations displayed (yellow triangles). They are surrounded by plane wave sources (purple dots).

Figure 4.18a presents a similar result as figure 4.16a. The velocity values of the cells simulate an homogeneous medium. Figure 4.18b shows an decrease of performance of this estimation compared to the previous one. The values fluctuate inside a 0.41 km/s range hence the estimation resembles an homogeneous medium. The estimation errors decreased to values between 0 and 10%.



(a) (b)

Figure 4.18: (a) Final estimated map for homogeneous case with lower resolution. (b) Error map which displays the percentage of error associated to each cell's velocity estimation.

4.2.2 Heterogeneous medium 1

In this case we will observe the results of estimating a model with non uniform medium. Figure 4.19 shows the heterogeneous medium of study. There are 25 stations making up the network with 500 plane wave sources surrounding them.

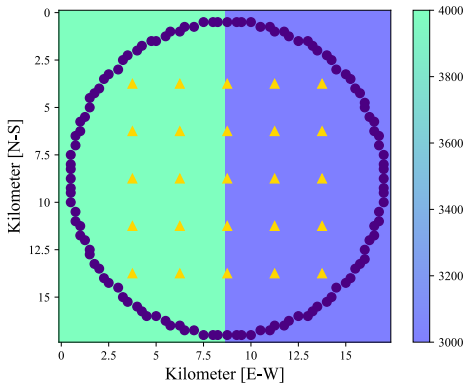
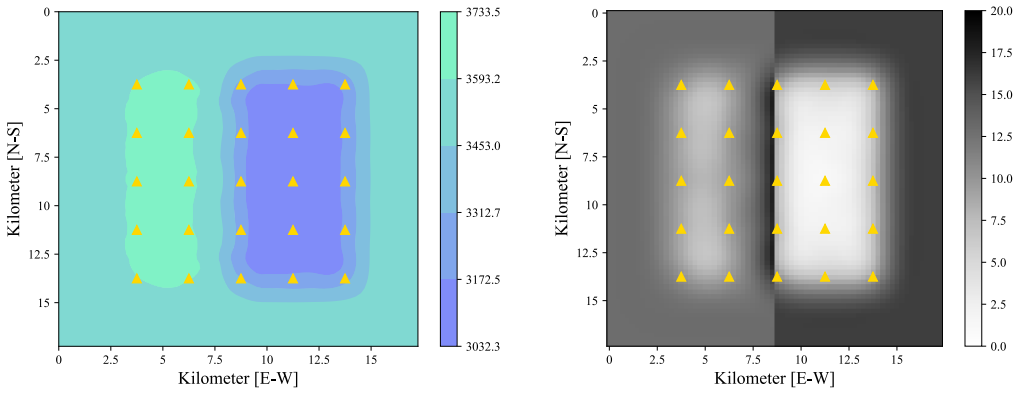


Figure 4.19: Heterogeneous area characterized by a change of medium. There are 25 stations displayed (yellow triangles) and they are surrounded by plane wave sources (purple dots).

As the previous case, an estimation of the parameters is executed. For this case the used values for ε and η are 45 and 50, respectively. Although the intersection between the two mediums in figure 4.20a looks blurred, the estimation could recognize different velocity zones.

Figure 4.20b shows that higher velocity zones have a higher associated error. As the velocity of the medium increases, further remoteness is recommended. Thus, It makes sense that the estimations of zones of greater speed have also a greater error compared to slower areas.



(a) (b)

Figure 4.20: (a) Final estimated map for the first heterogeneous case. (b) Error map which displays the percentage of error associated to each cell's velocity estimation.

Figure 4.21 displays the same scenario with a lower resolution. The size of the grid decreased and each cell measures 600 by 600 meters. The rest of the parameters are preserved.

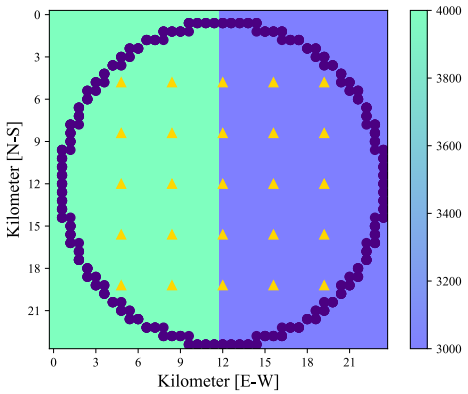
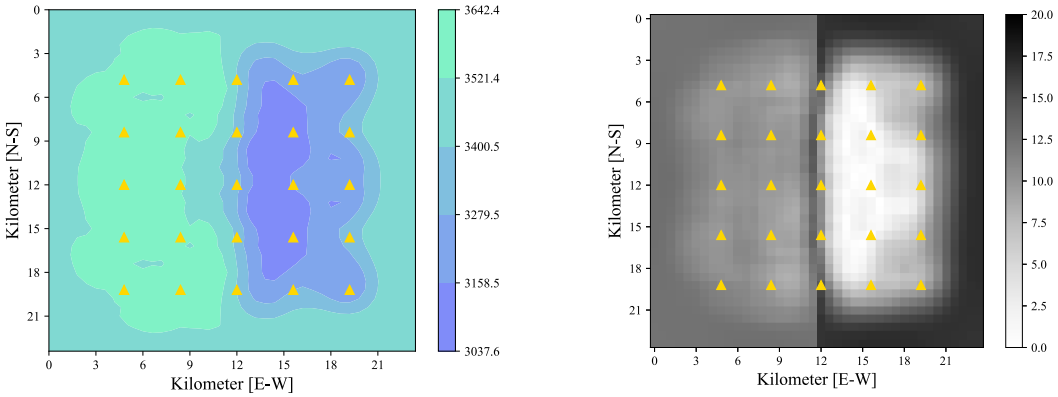


Figure 4.21: First heterogeneous medium with a decrease in resolution. There are 25 stations displayed (yellow triangles) and they are surrounded by plane wave sources (purple dots).

Figures 4.22a and 4.22b show the results of the same experiment with a change of parameters. The technique manages to differentiate different speed zones with lower efficiency than the previous one. The smoothness component in the inverse problem and the added gaussian blur produce an imprecise estimation of the mediums contact.



(a) (b)

Figure 4.22: (a) Final estimated map for the first heterogeneous case with a lower resolution. (b) Error map which displays the percentage of error associated to each cell's velocity estimation.

4.2.3 Heterogeneous medium 2

Finally, the last scenario to simulate is the case of an structure immersed in an homogeneous medium. Figure 4.23 shows the distribution of the network on this area surrounded by 500 sources. The structure has a higher speed associated as it is indicated in the diagram.

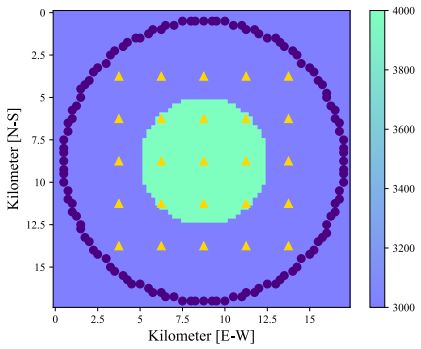
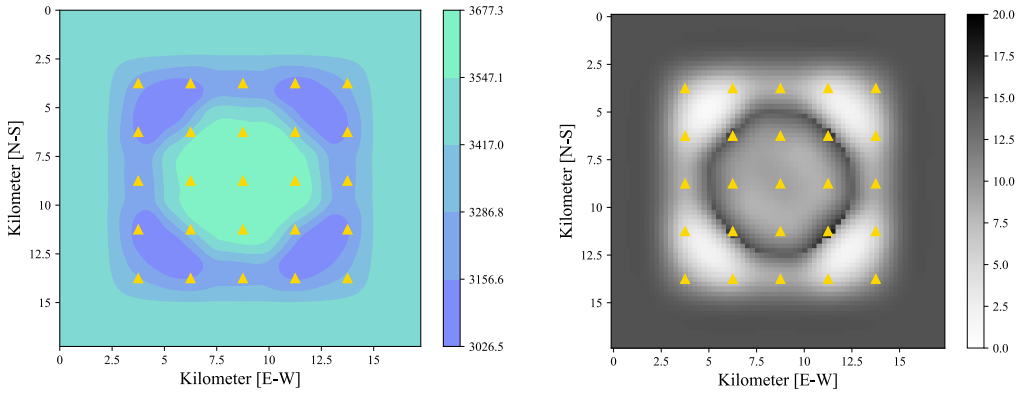


Figure 4.23: Heterogeneous area characterized by a structure inserted inside an homogeneous medium. There are 25 stations displayed (yellow triangles) and they are surrounded by plane wave sources (purple dots).



(a) (b)

Figure 4.24: (a) Final estimated map for the first heterogeneous case. (b) Error map which displays the percentage of error associated to each cell's velocity estimation.

Figure 4.24a shows how the estimation methodology could identify the anomaly. As the previous case, the is an important error associated to the model of the intersection between media.

As in previous cases, a change of the grid's parameters is made. The size of each cell is of 600 by 600 meters and the grid's size decreases as it is shown by figure 4.25.

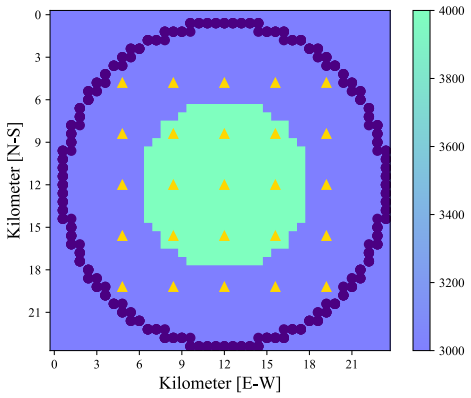
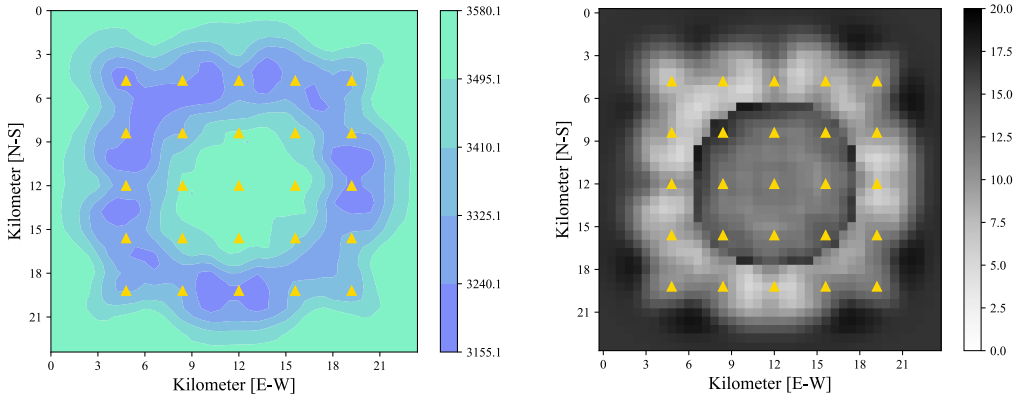


Figure 4.25: Heterogeneous area characterized by a change of medium. There are 32 stations displayed (yellow triangles) and they are surrounded by plane wave sources (purple dots).

This case shows a clear identification of the central structure (figure 4.26a). Nevertheless, the higher speed zones have associated a higher estimation error as it is indicated by figure 4.26b.



(a) (b)

Figure 4.26: (a) Final estimated map for the first heterogeneous case. (b) Error map which displays the percentage of error associated to each cell's velocity estimation.

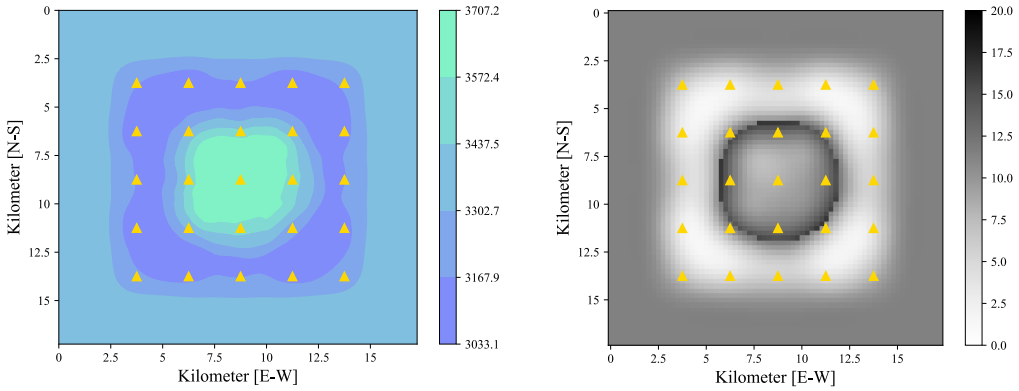
Variation in structure

Another experiment is developed in order to know what is the minimum size for a structure to be detected. The model in figure 4.23 was adopted, and at each iteration the radio of the structure was changed. In figure 4.27a it can be appreciated that the tomography was capable to identify a structure of 3.25 kilometres of ratio. The error map in figure 4.27b has zones where the borders are not accurately estimated (black borders). However, the entire model has in average an error of 15 % if the zone outside the network is ignored.

The ratio structure is later shrinked to 2.5 kilometres. Despite the model in figure 4.28a detects the presence of a structure, it is not capable of estimating its velocity accurately.

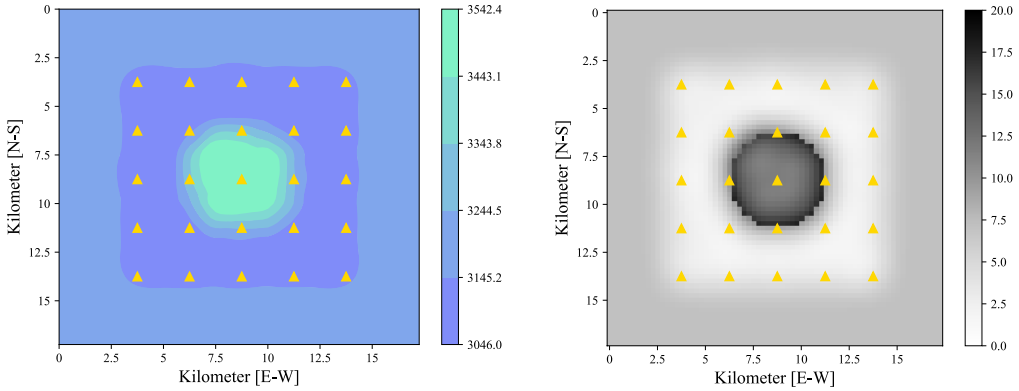
Figure 4.28b shows how it is possible to estimate the surrounding area and the anomaly with errors that move between 0 and 15 % approximately.

Finally, figure 4.29a displays the result from reducing the ratio to 1.25 kilometres. The technique was capable of identifying the high speed structure with an error in its ratio. The highest error is associated to the anomaly velocity estimation. With this we can see that given the parameters, there are some anomalies that the methodology is able to identify even with a small size.



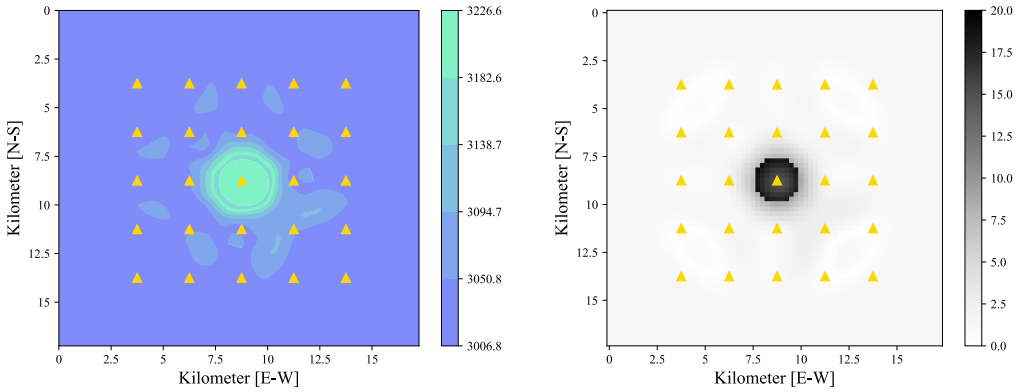
(a) (b)

Figure 4.27: (a) Estimation of a structure of 3.25 kilometres of ratio. Model detects the anomaly size and the speed value with a mean error of 10%. (b) Map of the error associated to the estimation.



(a) (b)

Figure 4.28: (a) Estimation of a structure of 2.5 kilometres of ratio. Model detects the anomaly size but not the speed value. (b) Map of the error associated to the estimation.



(a) (b)

Figure 4.29: (a) Estimation of a structure of 1.25 kilometres of ratio. Model detects neither the anomaly size nor the speed value. (b) Map of the error associated to the estimation.

4.2.4 Simulation using noise sources

As an additional experiment, uncorrelated noise sources are implemented. Firstly, an homogeneous scenario is set with a velocity of 3 km/s, 0.9 by 0.9 kilometers cells and two stations as seen in figure 4.1. The signal emitted from the sources corresponds to noise centered at 1.0 Hz (high period). This is a similar experiment to the one implemented by [9].

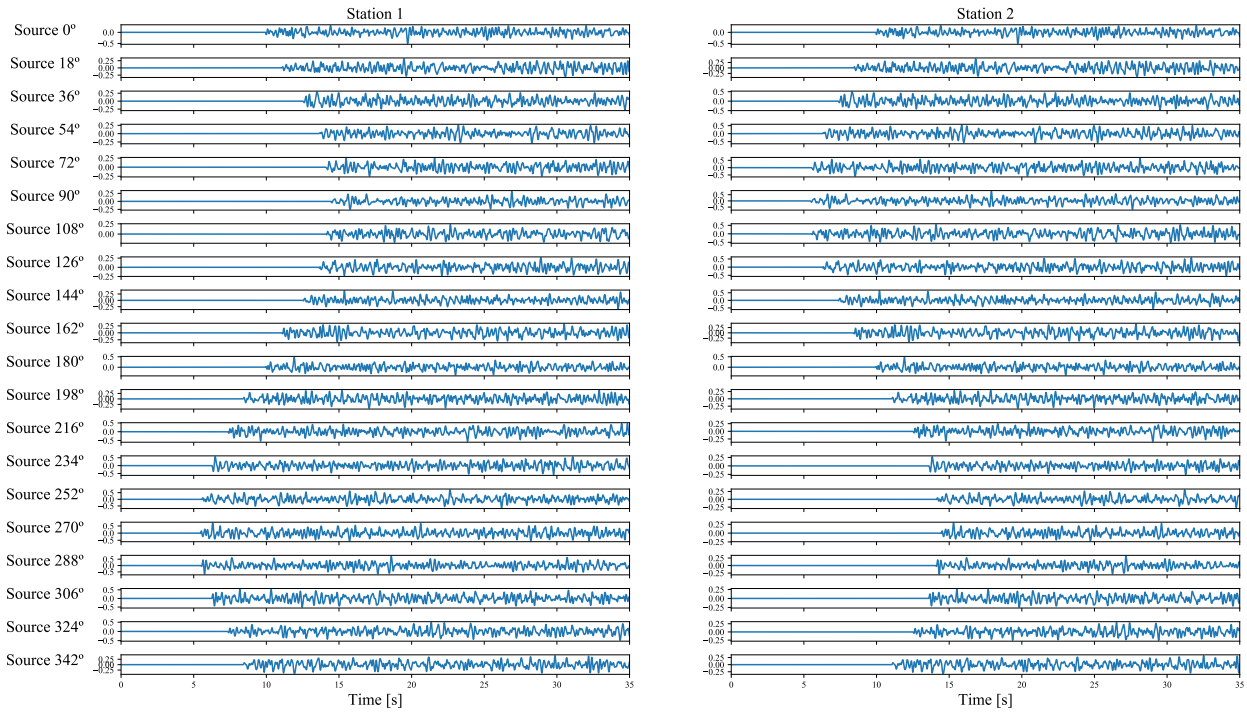


Figure 4.30: Recordings of two stations by source position for an homogeneous case. Uncorrelated noise sources were sent through a lossless media with a velocity of 3 km/s.

The station recordings in function of the source are presented in figure 4.30. A time normalization is needed for these signals. For this, the One-Bit method is used on the recordings, and its results are displayed in figure 4.31. In addition, a spectral normalization is applied in order to continue with the cross-correlations computing and obtain the impulse response. Figure 4.32 the resulted cross correlation signal which has the expected shape can be seen.

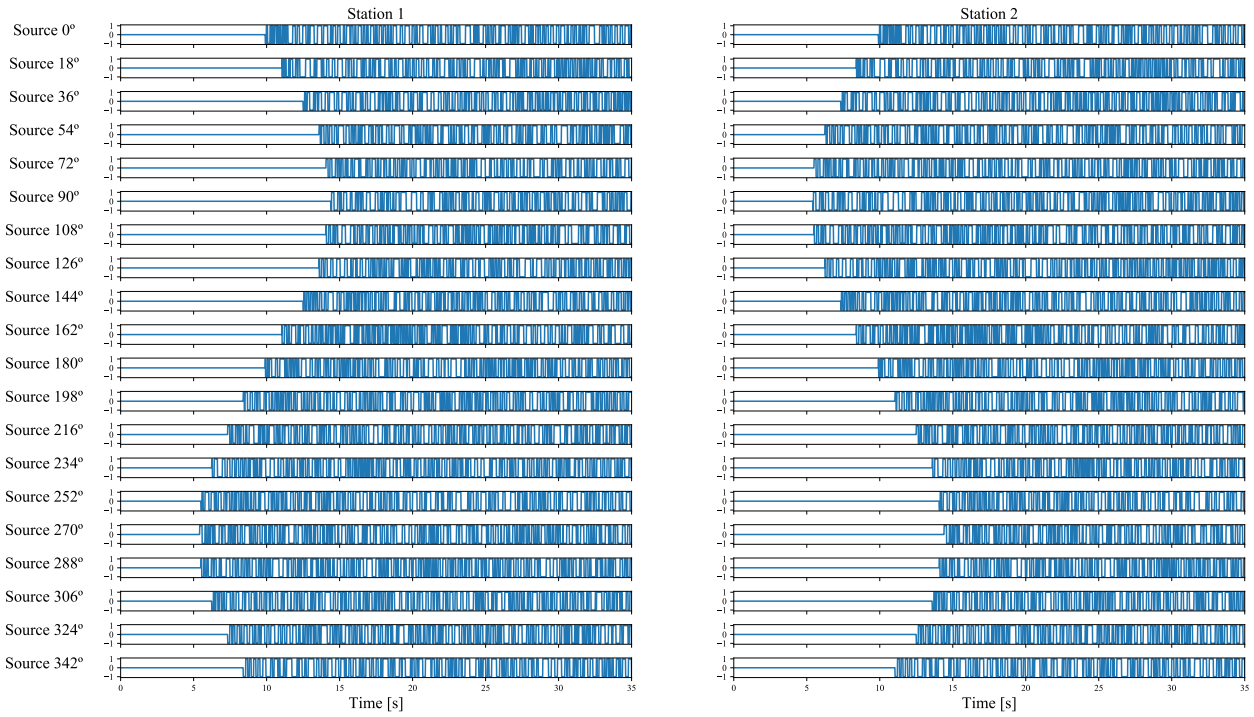


Figure 4.31: Result of the temporal normalization: One-Bit, on the recordings.

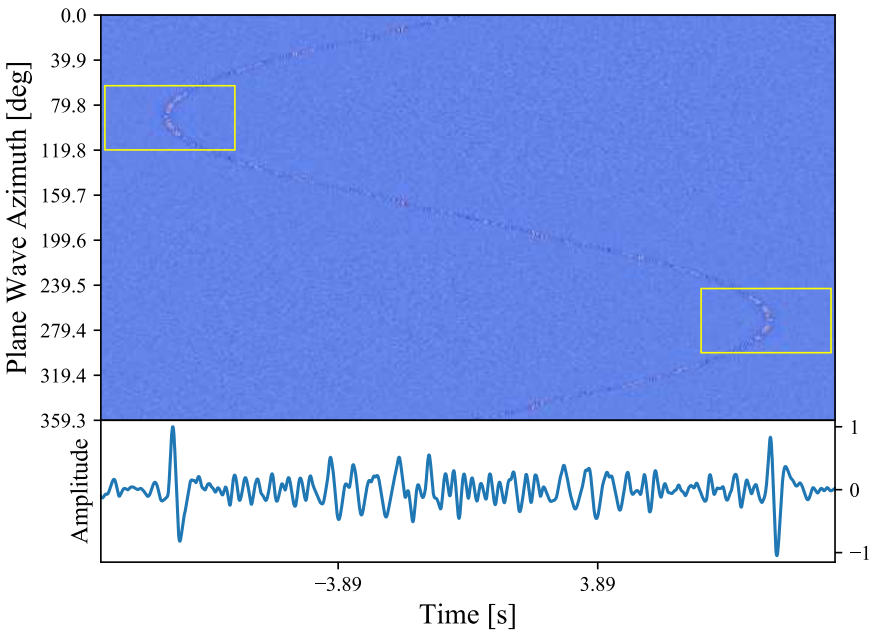
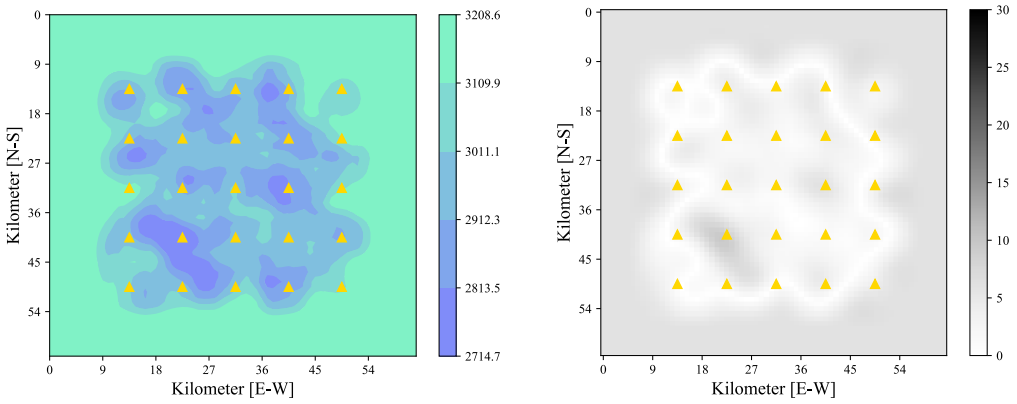


Figure 4.32: Cross-correlation signals in function of the source. After the stacking of all them, the final cross-correlation signal is obtained (blue graphs below the image.

Finally, after stacking the signals and applying a final low-pass filter, the final cross-correlation signal can be seen. The estimated velocity is 2.928 km/s (2.38 % of error). This is an adequate result, hence the analysis of a network follows.

The three synthetic cases are analyzed. Figure 4.33a displays the result of the estimation of an homogeneous medium. Although there irregularities can be observed inside the network area, the minimum and maximum values indicate that this is an uniform area. Figure 4.33b is a proof of this assumption. The estimation errors locate between 0 and 5 %.

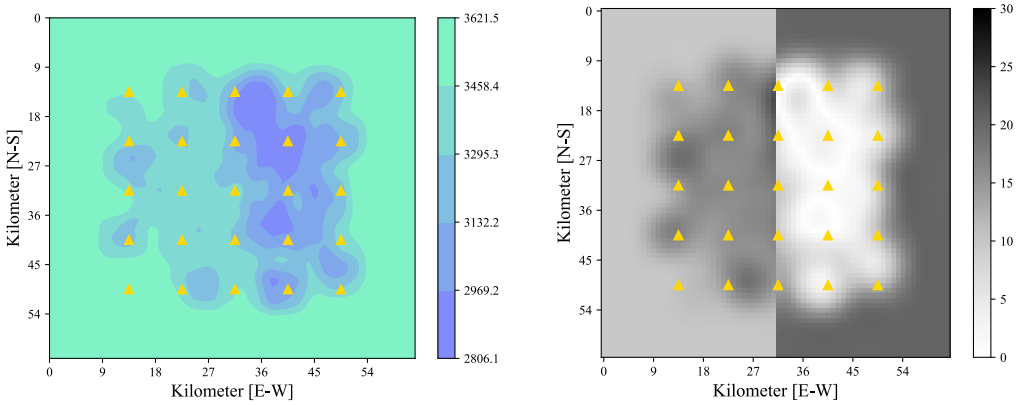


(a) (b)

Figure 4.33: (a) Model estimation using noise sources for an homogeneous medium. (b) Map of the error associated to the estimation.

Figures 4.34a and 4.34b display the estimation of the first heterogeneous case. The lower

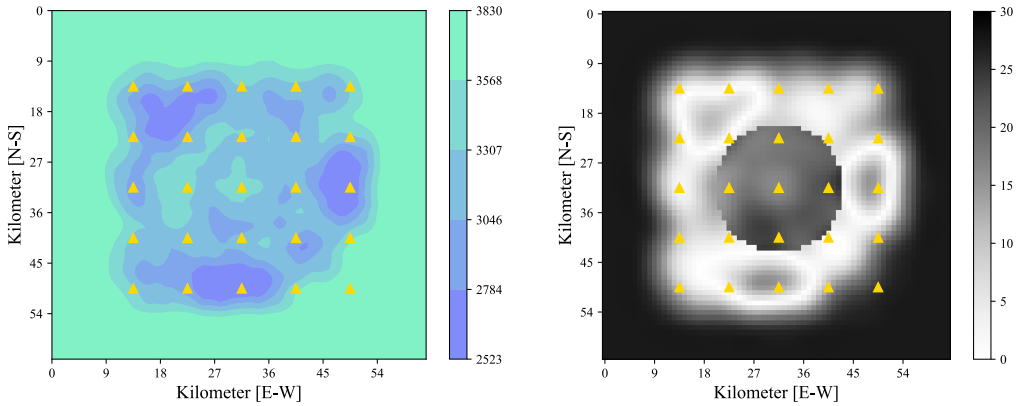
speed section was identified with a small related error. However, the change in fase is not quite identified.



(a) (b)

Figure 4.34: (a) Model estimation using noise sources for an heterogeneous medium. (b) Map of the error related to the estimation.

Figures 4.35a and 4.35b exhibit the results from the second heterogeneous medium estimation. Despite the model discovers there is a zone of higher speed in the center, it is not able to identify its shape.



(a) (b)

Figure 4.35: (a) Model estimation using noise sources for an heterogeneous medium. (b) Map of the error associated to the estimation.

The results associated to this experiment presented higher errors than the previous ones. Because of the impulse shape of the cross-correlation signals, it is more difficult to achieve the desired interference in the stacking stage. The results may be improved with a denser array of sources.

Chapter 5

Conclusion

This work achieved its goal of helping to understand the ambient noise tomography technique. Several variables influence notoriously on this method. Given this, the implementation of synthetic scenarios - which are representative of reality - is a complex task. The simulations presented in this thesis work as a starting point for a complete methodology study. Furthermore, the scenarios show how different variables influence the final estimated models. It also corroborates the established rules widely used in the state of the art.

Seismologists simplify different stages of the methodology. This, with the objective of reducing computational costs. However, this is not the best procedure to follow. For instance, the L-curve technique is used for the selection of parameter values for the inverse problem. Seismologists used to make this selection guessing whilst considering prior information. Then, those values are changed according to the results. However, this procedure could signify more computational costs. The process must be repeated each time a riddle is performed. In addition, this problem is highly sensitive. A wrong estimate could produce significant differences between the real model and its estimation.

The presence of a ground-truth did more than corroborate the technique worked. It added information of the associated estimation errors and how parameters affected them. It was possible to know that if the restrictions of the parameters are not respected, there is an associated error. For example, the case of stations distance. Also, it was possible to know that there are characteristics of the medium that are not going to be identified depending on the setup. And despite different speed zones being identified, higher velocity areas have a bigger associated error.

Although an automatized procedure is preferred, several stages need to count with prior information. It was possible to identify the influence of station positions on the Fresnel Zones. Prior information can help to picture an idea of the noise distribution. Then, the network location can be designed accordingly in order to exploit this resource.

Given this, the hypothesis established is correct. It is possible to implement a synthetic scenario in order to study different aspects of noise tomography. Later, they can be applied in future campaigns.

5.1 Future work

1. To explore new methodologies for the selection of parameters. It is needed to replace the parameters guessing by a more stable parameter selection technique.
2. To develop a sensitivity study of the technique. With inadequate restrictions and parameter values, the estimation could diverge. It is necessary the development of better restrictions and its consequences in the results.
3. To study the ambient noise tomography technique for anisotropic distribution of noise. This work used isotropic noise distribution only. It is necessary to understand the influence of this variable on the velocity model.
4. To study more complex escenarios. In order to achieve more realistic experiments, it is needed to consider subsurfaces with more than two type mediums and to try different transitions.
5. To implement new techniques for the execution of different stages of the tomography. Despite the methodologies implemented worked, fewer error could be achieved with new techniques.
6. To compare different sensors' performances. For example, to establish the benefits of using short period sensors instead of broad-band sensors.

Chapter 6

Bibliography

[1] GD Bensen, MH Ritzwoller, MP Barmin, AL Levshin, F Lin, MP Moschetti, NM Shapiro, and Yanyan Yang. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, 169(3):1239–1260, 2007.

[2] Andrew Curtis, Peter Gerstoft, Haruo Sato, Roel Snieder, and Kees Wapenaar. Seismic interferometry?turning noise into signal. *The Leading Edge*, 25(9):1082–1092, 2006.

[3] William Menke. *Geophysical data analysis: discrete inverse theory: MATLAB edition*, volume 45. Academic press, 2012.

[4] Heather Nicolson, Andrew Curtis, Brian Baptie, and Erica Galetti. Seismic interferometry and ambient noise tomography in the british isles. *Proceedings of the Geologists' Association*, 123(1):74–86, 2012.

[5] Brian Borchers Richard C. Aster and Clifford H. Thurber. *Parameter Estimation and Inverse Problems: Second Edition*. ELSEVIER, 2013.

[6] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

[7] Dirk-Jan van Manen, Andrew Curtis, and Johan O Robertsson. Interferometric modeling of wave propagation in inhomogeneous elastic media using time reversal and reciprocity. *Geophysics*, 71(4):SI47–SI60, 2006.

[8] Dirk-Jan van Manen, Johan OA Robertsson, and Andrew Curtis. Modeling of wave propagation in inhomogeneous media. *Physical Review Letters*, 94(16):164301, 2005.

[9] Kees Wapenaar, Deyan Draganov, Roel Snieder, Xander Campman, and Arie Verdel. Tutorial on seismic interferometry: Part 1?basic principles and applications. *Geophysics*, 75(5):75A195–75A209, 2010.

[10] Yingjie Yang, Michael H Ritzwoller, Anatoli L Levshin, and Nikolai M Shapiro. Ambi-

ent noise rayleigh wave tomography across europe. *Geophysical Journal International*, 168(1):259–274, 2007.

[11] Huajian Yao and Robert D Van Der Hilst. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to se tibet. *Geophysical Journal International*, 179(2):1113–1132, 2009.

[12] Huajian Yao, Robert D van Der Hilst, and Maarten V De Hoop. Surface-wave array tomography in se tibet from ambient seismic noise and two-station analysis?i. phase velocity maps. *Geophysical Journal International*, 166(2):732–744, 2006.