Polarization of surface waves : characterization, inversion and application to seismic hazard assessment

Manuel Hobiger

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences de la Terre, de l'Univers et de l'Environnement**

Arrêté ministériel : 7 août 2006

Présentée par

Manuel HOBIGER

Thèse dirigée par **Pierre-Yves BARD** et
codirigée par **Cécile CORNOU** et **Nicolas LE BIHAN**

préparée au sein de l'**Institut des Sciences de la Terre (ISTerre)**
dans **l'École Doctorale Terre, Univers, Environnement**

Polarisation des ondes de surface : Caractérisation, inversion et application à l'étude de l'aléa sismique

Thèse soutenue publiquement le **13 janvier 2011**,
devant le jury composé de :

Michel CAMPILLO
ISTerre, Grenoble, Président
Donat FÄH
ETH Zürich, Rapporteur
Pascal LARZABAL
ENS Cachan, Rapporteur
Matthias OHRNBERGER
Universität Potsdam, Examinateur
Ulrich WEGLER
BGR, Hannover, Examinateur
Pierre-Yves BARD
ISTerre, Grenoble, Directeur de thèse
Cécile CORNOU
ISTerre, Grenoble, Directrice de thèse
Nicolas LE BIHAN
GIPSA-Lab, Grenoble, Directeur de thèse

Polarization of surface waves: Characterization, inversion and application to seismic hazard assessment

Polarisation des ondes de surface : Caractérisation, inversion et application à l'étude de l'aléa sismique

Summary

The estimation of site effects is an important task in seismic hazard assessment. In order to understand and estimate these site effects, the local soil structure as well as the composition (body/surface waves, Love/Rayleigh waves) and properties of the wave field have to be investigated. Instead of using expensive methods such as borehole measurements or reflection or refraction seismics, these properties can be investigated using surface waves. The two most important types of seismic surface waves are Love and Rayleigh waves. These waves exhibit dispersion curves which are directly linked to the soil structure. Furthermore, the ellipticity, a parameter describing the elliptical motion of Rayleigh waves as a function of frequency, is also linked to the soil structure. In order to identify and extract different wave types, their respective polarization parameters have to be characterized.

In a first part of this work, new methods allowing the determination of the polarization parameters of seismic surface waves are developed. Two methods, DELFI and RayDec, estimate the ellipticity of Rayleigh waves using the three-dimensional record of a single seismic station. The first method tries to directly fit ellipses to parts of the signal, whereas the second one uses statistical properties to effectively suppress other wave types than Rayleigh waves. A third newly developed method, MUSIQUE, is the combination of the established MUSIC algorithm with a version using quaternions, an example of hypercomplex numbers. MUSIQUE uses seismic array recordings to discriminate between Love and Rayleigh waves, estimate the related dispersion curves and the Rayleigh wave ellipticity.

In the second part of the work, the new methods are applied to real data measurements. A theoretical investigation of the inversion of ellipticity curves shows that the right flank of the ellipticity peak and the peak frequency carry the important information on the soil structure. However, as the ellipticity curve is not unique, additional information, e.g. small-scale SPAC or MASW measurements, have to be added to the inversion. In a second step, ellipticity curves estimated by RayDec on real data measurements are inverted jointly with SPAC measurements. For the 14 investigated sites, the inversion results are in good agreement with direct dispersion curve measurements. Furthermore, the ellipticity inversion could help in identifying the correct Rayleigh wave modes. In the last chapter, the seismological wave field generated by 22 earthquakes and recorded with a dense seismic array in the Santa Clara Valley, California, is investigated using MUSIQUE. Large amounts of scattered Love waves arriving from southern directions are found. Furthermore, the estimated energy repartition between Love and Rayleigh waves for the different events is rather heterogeneous. Finally, combining the information of all earthquakes, the dispersion curves for both the fundamental and the first harmonic Love wave modes are retrieved. For the fundamental Rayleigh wave mode, neither the dispersion nor the ellipticity curve can be found, but for the first harmonic Rayleigh wave mode both properties are identified.

Résumé

L'analyse des effets de site joue un rôle décisif pour l'estimation de l'aléa sismique. Afin de comprendre et de quantifier ces effets de site, il faut étudier la structure locale du sol ainsi que la composition (ondes de volume/surface, ondes de Love/Rayleigh) et les propriétés du champ d'ondes. Au lieu de techniques coûteuses comme des mesures par forage ou la sismique par réflexion ou réfraction, ces propriétés peuvent être analysées en utilisant les ondes de surface. Les deux types d'ondes de surface de loin les plus importants en sismologie sont les ondes de Love et de Rayleigh. Ces ondes sont caractérisées par des courbes de dispersion directement liées à la structure du sol. En outre, l'ellipticité, un paramètre décrivant le mouvement elliptique des ondes de Rayleigh en fonction de la fréquence, est aussi liée à la structure du sol. Afin d'identifier et d'extraire les différents types d'ondes, leurs paramètres de polarisation doivent être caractérisés.

Dans la première partie de ce manuscrit, de nouvelles méthodes permettant la détermination des paramètres de polarisation des différents types d'ondes de surface sont développées. Deux de ces méthodes, DELFI et RayDec, évaluent l'ellipticité des ondes de Rayleigh à partir de l'enregistrement à trois composantes d'un seul capteur sismique. La première de ces méthodes essaie d'ajuster directement des ellipses à des parties du signal, tandis que la deuxième réprime tout type d'onde sauf les ondes de Rayleigh en exploitant des propriétés statistiques. Une troisième nouvelle méthode, MUSIQUE, représente la combinaison de l'algorithme MUSIC avec une version utilisant les quaternions, des nombres hypercomplexes de dimension 4. Cette méthode utilise les enregistrements de réseaux sismiques pour distinguer ondes de Love et ondes de Rayleigh, estimer les courbes de dispersion associées et l'ellipticité des ondes de Rayleigh.

La deuxième partie du manuscrit est dédiée à l'application des nouvelles méthodes sur des données réelles. Une 'etude théorique des inversions de courbes d'ellipticité montre que le flanc droit du pic d'ellipticité et la fréquence du pic comportent toutes les informations nécessaires pour caractériser la structure du sol. Cependant, la courbe d'ellipticité n'étant pas unique, des informations supplémentaires sur la vitesse des ondes de cisaillement de la proche surface doivent être inclues dans le processus d'inversion. Une seconde étude s'intéresse à l'inversion conjointe de courbes d'ellipticité mesurés par RayDec avec des mesures d'autocorrélation spatiale sur des données réelles acquises dans différents sites en Europe. Sur les 14 sites ainsi analysés, les résultats de l'inversion sont en bon accord avec les mesures directes de courbes de dispersion. Par ailleurs, la méthode montre son potentiel à discriminer les différents modes de Rayleigh observés. Dans le dernier chapitre, le champ d'ondes issu de l'enregistrement de séismes par un réseau de capteurs sismiques dans la vallée de Santa Clara en Californie est analysé par MUSIQUE. De grandes quantités d'ondes de Love diffusées arrivant de directions méridionales sont identifiées. En outre, la répartition entre ondes de Love et ondes de Rayleigh varie fortement pour les différents évènements. Finalement, en combinant les informations de tous les séismes, les courbes de dispersion pour le mode fondamental et le premier mode harmonique des ondes de Love sont identifiées. Pour le mode fondamental des ondes de Rayleigh, ni la courbe de dispersion ni la courbe d'ellipticité ne sont trouvées, tandis que pour le premier mode harmonique, les deux propriétés sont identifiées.

Acknowledgments

In the following, I would like to express my gratitude to those without whom this work could not have been done.

Trois directeurs de thèse, Cécile, Pierre-Yves et Nicolas, m'ont encadré pendant ces trois ans. Profitant ainsi de la combinaison de leurs savoir-faire et de trois points de vue différents, c'était une grande chance pour moi. Merci de m'avoir proposé ce sujet intéressant et d'avoir toujours été disponible à aider et donner de bons conseils.

I would also like to thank Donat Fäh and Pascal Larzabal for having willingly accepted to review this work and Michel Campillo, Matthias Ohrnberger and Ulrich Wegler for having consented to be part of the jury.

It was a big opportunity to use the abundance of data measured during the Neries project. Thanks to all who participated in these measurements. I would also like to thank Florence, Brigitte and Giuseppe for their help and good discussions and Marc for his availability and support with geopsy and dinver. It was also a big pleasure to work with Alekos and Nikos, especially during the field campaign in Greece. I will also keep good memories of the measurements with Seiji in Cadarache and La Tronche.

Another thank goes to Arthur Frankel and David Carter who willingly provided the data of the San Jose seismic array and Arthur Rodgers who helped with good discussions and the Bay Area model to analyze the data.

Merci aussi à Cécile (encore une fois) et Manu pour m'avoir accueilli, en 2005, pour le stage de master et m'avoir donné l'envie pour la sismologie. Ce stage était finalement la raison pour laquelle je suis revenu au LGIT pour passer la thèse.

C'était toujours un vrai plaisir de travailler au LGIT qui m'est devenu comme une deuxième maison. Un grand merci aux autres étudiants pour avoir rendu le travail le plus agreáble possible, mais aussi pour les différentes escapades qu'on a faites, les balades en montagne, le ski ou les nuits à regarder les étoiles.

Merci à l'équipe de foot du LGIT, c'était un plaisir et un honneur de représenter les couleurs du LGIT sur les terrains de foot, même si le succès n'était pas toujours de notre côté.

Enfin, merci aussi à ceux dont le travail reste assez souvent inaperçu, mais qui sont indispensables pour le bon déroulement d'une thèse : les équipes informatique et administrative et l'Ecole Doctorale.

Ein besonderer Dank geht schließlich an meine Eltern, auf deren Unterstützung ich mich auf diesem langen Weg immer verlassen konnte und die mir stets mit Rat und Tat zur Seite standen.

Contents

Introduction

During all times of history, earthquakes have influenced the development of mankind by destroying towns, rising the sea level and reshaping the landscape. Longtime considered as the revenge of gods, modern science proved that earthquakes are linked to processes in the Earth's interior. Due to the motion of tectonic plates, stress builds up in the crust and discharges in earthquakes.

Nevertheless, it is not only the brute force of an earthquake, measured as its magnitude, which determines the possible damages. Indeed, the seismic waves released during an earthquake interact with the structure of the medium through which they are traveling. For example, in sedimentary basins representing a strong velocity contrast to the surrounding rock formations, seismic waves can reflect many times and, to a certain degree, be trapped inside the basin structure. In this way, the sedimentary basin built by the deposits of a river can act as a wave guide. Inside the basin, the seismic motion is thus not only amplified, but also extended to longer durations. Many towns are located in sedimentary basins and the seimic risk for these towns can be several orders higher compared to locations in the vicinity. Therefore, even earthquakes of moderate magnitude can present a serious risk.

The effects modifying the seismic motion at a given site are called site effects. They have to be taken into account in order to assess the seismic hazard at a given site. In sediment-filled basins, the site effects range from one-dimensional to three-dimensional effects, depending mainly on the valley geometry and the properties of the involved rocks. The most important geometry parameter is the ratio between the horizontal dimensions of the valley and the sediment thickness. If this ratio is large, one-dimensional site effects predominate. These effects consist in standing waves in the vertical direction. For more complex structures, two-dimensional and three-dimensional site effects occur. The evaluation of site effects can be performed using earthquake recordings or by methods allowing the imaging of the soil structure. Especially in areas of moderate seismicity, such methods turn out to be rather useful.

Methods allowing the imaging of the soil structure include borehole measurements or classical geophysical prospection methods, such as reflection or refraction seismics or gravimetry. All these methods are quite expensive and are furthermore troublesome to carry out in densely populated areas. Methods using surface waves are therefore widely used in urban areas.

The seismic wave field is composed of four principal wave types. Shear and pressure waves are the two types of body waves, whereas Love and Rayleigh waves are surface waves. The motion of surface waves is confined to the Earth's surface, i.e. the interface between the soil and the air. Surface waves are rather adapted to investigate the soil structure. They are dispersive, i.e. their velocity is a function of frequency which is directly linked to the soil structure. Furthermore, the motion of Rayleigh waves is elliptical and the ellipticity, i.e. the ratio between the horizontal and the vertical axes, is also a function of frequency linked to the soil structure. Therefore, by analyzing surface waves the structure of the sedimentary layers can be investigated without penetrating the soil. Body waves, in contrast, are not dispersive. Besides, their apparent velocity at the surface depends on the angle of incidence. Hence, they are not adapted to an investigation of the soil structure. If the dispersion (or the Rayleigh wave ellipticity) of seismic surface waves can be measured with precision, the properties of the soil structure, above all the shear wave velocity profile, can be found by inverting these data.

Surely, surface waves could be generated actively, but these methods are not always feasible in urban areas and furthermore commonly restricted to the investigation of the very surficial layers only. Therefore, the use of ambient seismic vibrations (the so called "seismic noise") is tempting. These seismic motions, composed both of body and surface waves, are present at all times, but their amplitude can change in time. In different frequency ranges, they are generated by different sources. At low frequencies (below about 1 Hz), natural sources dominate. For example, ocean waves generate ambient seismic vibrations with frequencies around $\frac{1}{7}$ Hz, but the action of wind on trees or buildings can also induce waves in the soil. At higher frequencies, anthropogenic sources like traffic and industrial sources dominate.

Seismic noise recordings by means of arrays of seismic sensors and adapted processing techniques (SPAC [Aki, 1957], frequency-wavenumber analysis [Lacoss et al., 1969, Capon, 1969], noise correlation [Shapiro et al., 2005]) are nowadays largely used in urban areas in order to retrieve the shear-wave velocity profile. Another appealing approach allowing the use of recordings of a single seismic station only would be the inversion of the ellipticity curve in order to find the soil structure. Nevertheless, this approach requires the thorough investigation and extraction of Rayleigh waves in the seismic noise wave field.

Classical site effect estimation methods compare the ground motions recorded on the sediments with those recorded on a rocksite reference. This yields both site amplification and resonance frequency. Another widely used method, the H/V technique [Nogoshi and Igarashi, 1971, Nakamura, 1989], compares the ground motion recordings of the horizontal components with those of the vertical component at a given site in order to retrieve the resonance frequency. The site response can also be estimated using more sophisticated methods like generalized inversion. Nevertheless, none of these methods provides insights on the actual wave field composition. In the case of two- or three-dimensional site effects, however, there is a special interest in characterizing the wave field.

Several studies have shown that surface waves diffracted at the basin edges significantly modify the seismic wave field by contributing to amplification and duration lengthening of seismic motions [Kawase and Aki, 1989, Kawase, 1996, Field, 1996, Gaffet et al., 1998, Rovelli et al., 2001, Cornou et al., 2003a,b].

Studies investigating the properties of wave fields usually apply classical array processing methods (high-resolution frequency-wavenumber analysis [Capon, 1969], MUSIC [Schmidt, 1986]) to earthquake recordings of dense seismic arrays. These methods provide the azimuth and velocity of wave trains crossing the array. Waves coming directly from the source region can then be distinguished from surface waves diffracted at the valley edges. However, these techniques usually analyze the three components of seismic motion independently. Therefore, using these methods, it is difficult to clearly identify the surface wave types and their respective contributions to the wave field. The key parameter to identify the different wave types is their polarization. Considering all three components of a seismic record together would allow the estimation of the wave polarization and thus the correct identification of the wave types.

The first part of this PhD thesis will focus on the problem of determining the polarization properties of surface waves. Notably, three new methods will be presented which have been developed during this work. DELFI and RayDec have both been designed to estimate the ellipticity of Rayleigh waves using the three-component record of a single seismic station. The third method, MUSIQUE, is an advanced version of the MUSIC algorithm and uses seismic array recordings. It allows the identification of the wave type and the determination of the dispersion curves of surface waves and the Rayleigh wave ellipticity curve. The second part of the thesis shows the application of these methods to real data measurements. The inversion of ellipticity curves estimated using RayDec is shown. Furthermore, the seismological wave field generated by earthquakes is investigated using MUSIQUE.

Depuis le début de l'histoire humaine sur Terre, les séismes influent sur le développement de l'humanité en détruisant des villes, changeant le niveau de la mer ou transformant le paysage. Longtemps considérés comme la vengeance des dieux, la science moderne a découvert que les séismes sont liés à des processus à l'intérieur de la Terre. Suite aux mouvements de plaques tectoniques, le stress dans la croûte terrestre augmente et se décharge pendant des séismes.

Quoi qu'il en soit, ce n'est pas seulement la force brute d'un séisme, indiquée par sa magnitude, qui détermine les dégâts potentiels. En fait, les ondes sismiques dégagées lors d'un séisme interagissent avec la structure du médium à travers lequel ils voyagent. Par exemple, dans des bassins sédimentaires à fort contraste de vitesse aux rochers avoisinants, les ondes sismiques peuvent se réfléchir plusieurs fois et être, à un certain degré, piégées à l'intérieur de la structure du bassin. De cette façon, le bassin sédimentaire formé par les dépôts d'une rivière peut agir comme un guide d'ondes. Dans le bassin, le mouvement sismique n'est donc pas seulement amplifié, mais aussi prolongé. Un grand nombre de villes étant localisées dans des bassins sédimentaires, l'aléa sismique de celles-ci peut être plusieurs fois plus important que pour des endroits proches hors du bassin. C'est pourquoi même des séismes de moyenne magnitude peuvent présenter un aléa important.

Les effets modifiant le mouvement sismique à un site donné s'appellent effets de site. Ils doivent être pris en compte pour évaluer l'aléa sismique. Dans des bassins remplis de sédiments, les effets de site vont d'effets unidimensionnels jusqu'aux effets tridimensionnels, dépendant principalement de la géométrie de la vallée et des propriétés des roches impliquées, le paramètre géométrique le plus important étant le rapport entre les dimensions horizontales de la vallée et l'épaisseur des sédiments. Pour de fortes valeurs de ce rapport, les effets de site unidimensionnels prédominent. Ces effets consistent en ondes stationnaires dans la direction verticale. Pour des structures plus complexes, des effets de site bi- et tridimensionnels apparaissent. Les effets de site peuvent être évalués à l'aide de mesures de séismes ou par méthodes permettant l'imagerie de la structure du sol. Notamment dans des régions à sismicité modérée, ces dernières s'avèrent assez utiles.

Les méthodes permettant l'imagerie de la structure du sol incluent les mesures par forage ou les méthodes classiques de prospection géophysique comme les sismiques de réflexion ou refraction ou la gravimétrie. Toutes ces méthodes sont plutôt coûteuses et posent problème dans des régions très peuplées. C'est pourquoi les méthodes utilisant des ondes de surface sont très répandues dans les régions urbaines.

Le champ d'ondes sismiques est composé de quatre types d'ondes principaux. Les ondes de compression et les ondes de cisaillement forment les deux types d'ondes de volume, tandis que les ondes de Love et les ondes de Rayleigh sont des ondes de surface. Le mouvement de ces dernières se borne à la surface de la Terre, soit l'interface entre le sol et l'air. Etant dispersives, c'est-à-dire leurs vitesses étant des fonctions de la fréquence directement liées à la structure du sol, les ondes de surface sont plutôt appropriées pour étudier la structure du sol. En outre, le mouvement des ondes de Rayleigh suivant une ellipse, le rapport entre les axes horizontal et vertical de ce mouvement, l'ellipticité, est aussi une fonction de la fréquence liée à la structure du sol. En conséquence, la structure des couches sédimentaires peut être explorée sans pénétration du sol en analysant les ondes de surface. Les ondes de volume, quant à elles, ne sont pas dispersives, d'autant plus que leur vitesse apparente à la surface dépend de l'angle d'incidence. C'est pourquoi ces ondes ne sont pas adaptées à l'investigation de structures sédimentaires. Si les courbes de dispersion des ondes de surface (ou la courbe d'ellipticité des ondes de Rayleigh) peuvent être mesurées précisément, une inversion de ces données permet de retrouver les propriétés de la structure du sol, en particulier le profil de vitesse des ondes de cisaillement.

Evidemment, des ondes de surface pourraient être générées activement, mais ceci n'est pas toujours facilement faisable à l'intérieur d'agglomérations et en plus limité à l'investigation des couches les plus superficielles. C'est pourquoi l'utilisation du bruit sismique ambiant est alléchante. Ce bruit sismique, composé d'ondes de volume et d'ondes de surface, est toujours présent, son amplitude variant au cours du temps. Dans différentes gammes de fréquence, il est généré par des sources différentes. A basse fréquence (en-dessous de 1 Hz), les sources naturelles dominent. Par exemple, les ondes de la mer génèrent des vibrations sismiques ambientes avec des fréquences autour de $\frac{1}{7}$ Hz, mais l'action du vent sur les arbres ou les bâtiments peut aussi induire des ondes dans le sol. A plus haute fréquence, les sources anthropogéniques comme le trafic automobile ou les sources industrielles dominent.

Aujourd'hui, l'analyse des enregistrements de bruit sismique à l'aide de réseaux de capteurs sismiques par des techniques de traitement du signal adaptées (SPAC [Aki, 1957], analyse fréquence-nombre d'onde [Lacoss et al., 1969, Capon, 1969], corrélation de bruit [Shapiro et al., 2005]) est largement utilisé pour retrouver le profil de vitesse des ondes de cisaillement dans les agglomérations urbaines. Une autre approche alléchante tirant profit des enregistrements d'une seule station sismique serait donnée par l'inversion de courbes d'ellipticité en vue d'une détermination de la structure du sol, cette approche nécessitant cependant l'investigation et l'extraction soigneuses des ondes de Rayleigh du champ d'ondes ambiant.

Les méthodes classiques d'estimation des effets de site comparaissent les mouvements du sol enregistrés sur les sédiments avec ceux enregistrés sur un site de référence au rocher. Ceci donne l'amplification au site ainsi que la fréquence de résonance. Une autre technique largement utilisée, la méthode du rapport spectral H/V [Nogoshi and Igarashi, 1971, Nakamura, 1989], compare les enregistrements des mouvements du sol sur les composantes horizontales avec ceux de la composante verticale pour déterminer la fréquence de résonance d'un site donné. La réponse du site peut aussi être estimée en utilisant des méthodes plus sophistiquées comme l'inversion généralisée. Quoi qu'il en soit, aucune de ces méthodes ne permet d'évaluer la composition réelle du champ d'ondes.

Dans le cas d'effets de site bi- ou tridimensionnelles, il existe toutefois un intérêt spécial dans la caractérisation du champ d'ondes. Plusieurs études ont montrées que des ondes de surface diffractées aux bords d'un bassin modifient le champ d'ondes de manière significative en contribuant à l'amplification et à la prolongation temporelle des mouvements sismiques [Kawase and Aki, 1989, Kawase, 1996, Field, 1996, Gaffet et al., 1998, Rovelli et al., 2001, Cornou et al., 2003a,b].

Les études des propriétés de champs d'ondes appliquent normalement des méthodes de traitement de réseaux classiques (analyse fréquence-nombre d'onde à haute résolution [Capon, 1969], MUSIC [Schmidt, 1986]) à des enregistrements de séismes obtenus à l'aide de réseaux denses de capteurs sismiques. Ces méthodes donnent l'azimut et la vitesse des trains d'ondes traversant le réseau. Des ondes venant directement de la région de source peuvent ainsi être discernées des ondes de surface diffractées aux bords de la vallée. Cependant, ces techniques étudient les trois composantes du mouvement sismique indépendamment. Par conséquent, l'identification claire des types d'ondes de surface et leurs contributions au champ d'ondes sont difficiles à effectuer avec ces méthodes. Le paramètre clé de l'identification des différents types d'ondes est leur polarisation. La considération simultanée des trois composantes d'un enregistrement sismique permettrait d'estimer la polarisation des ondes et par conséquent l'identification correcte du type d'onde.

La première partie de ce manuscrit de thèse sera centrée sur la détermination des propriétés de polarisation des ondes de surface. Notamment, trois nouvelles méthodes développées dans le cadre de ce travail seront présentées. DELFI et RayDec ont tous les deux été désignés pour estimer l'ellipticité des ondes de Rayleigh à partir d'enregistrements d'un seul capteur sismique. La troisième méthode, MUSIQUE, est une version avancée de l'algorithme MUSIC et utilise des données de réseaux sismiques. Outre l'identification d'azimut et de vitesse d'ondes incidentes, cette méthode permet l'identification du type d'onde, la détermination des courbes de dispersion associées aux ondes de surface ainsi que de la courbe d'ellipticité des ondes de Rayleigh.
La deuxième partie du manuscrit est consacrée à l'application de ces méthodes à des données réelles. L'inversion de courbes d'ellipticité obtenues avec RayDec est montrée et le champ d'ondes sismiques générées par des séismes est analysé par MUSIQUE.

Thesis outline

The present document is organized in two parts. The first part is dedicated to the theoretical aspects of the estimation of seismic wave properties:

- Chapter 1 briefly recalls the properties of seismic waves which are necessary for the understanding of the work presented in the subsequent chapters.
- Chapter 2 deals with the use of single-sensor methods to estimate the polarization of seismic waves. Some existing methods are presented, before introducing DELFI and RayDec, two methods developed during this thesis work.
- Chapter 3 presents already existing methods using arrays of seismic sensors to retrieve the dispersion curves of seismic waves. These methods include beam-forming, frequency-wavenumber analysis, MUSIC and SPAC. The SPAC method is presented in detail and 2s-SPAC, a method using only two seismic sensors, is tested on seismic measurements.
- Chapter 4 introduces quaternions and biquaternions, two types of hypercomplex numbers. Using these numbers, MUSIC can not only estimate the dispersion curve but also the polarization of seismic waves. The possibilities and limits of this approach are shown before presenting MUSIQUE, a method developed during this thesis work. This method is a combination of the established classical and quaternion-MUSIC algorithms.
- In chapter 5, the previously introduced methods are applied to simulations of seismic signals to investigate their respective performances and limitations.

The second part of the work is focused on the application of the new methods to real seismic data:

- Chapter 6 investigates the theoretical aspects of the inversion of ellipticity data for the retrieval of the soil structure. It is determined which parts of the ellipticity curve actually carry the important information and which additional data can be used to constrain the structure in an unambiguous way. Furthermore, the effects of misestimated ellipticity curves are shown.
- In chapter 7, the lessons learned in chapter 6 are applied to real data measurements of 14 sites. These sites have been extensively investigated during the European NERIES project. For each site, a broad-band dispersion curve is available which allows the comparison of our inversion results with other measurements.
- In chapter 8, MUSIQUE is used to investigate the seismological wave field with an array of seismic sensors located on a sedimentary basin in California. For 22 different earthquakes, the azimuthal energy distribution and the energy repartition between Love and Rayleigh waves are measured and dispersion and ellipticity curves of the basin are retrieved combining the data of all earthquakes.

Part I

Theory

Chapter 1

Properties of seismic waves

This work deals with seismic waves, their identification and the determination of their respective properties. Therefore, this chapter will briefly present the properties of seismic waves. Starting with the wave equation, the difference between body and surface waves will be explained. Furthermore, the properties of surface waves, namely Love and Rayleigh waves, will be highlighted and the terms dispersion curve and ellipticity defined. For further details on the theory of seismic waves, we refer to standard literature as Aki and Richards [2002] or Lay and Wallace [1995].

Cet œuvre s'occupant d'ondes sismiques, de leur identification et de la détermination de leurs propriétés respectives, ce chapitre présentera brièvement les différents types d'ondes sismiques et leurs propriétés. Commençant par l'équation d'onde, les différences entre les ondes de volume et les ondes de surface seront expliquées avant d'élucider les propriétés des ondes de surface, notamment les ondes de Love et de Rayleigh et d'expliquer les termes courbe de dispersion et ellipticité. Pour des informations plus détaillées sur la théorie des ondes sismiques, nous nous référons à la littérature standard, par exemple Aki and Richards [2002] ou Lay and Wallace [1995].

1.1 Wave equation

The propagation of a three-dimensional plane wave $\vec{u}(\vec{r},t)$ in a homogeneous elastic medium is a solution of the wave equation

$$\frac{1}{v^2}\cdot\frac{\partial^2\vec{u}(\vec{r},t)}{\partial t^2}-\Delta\vec{u}(\vec{r},t)=0, \tag{1.1}$$

where $\vec{r}=\begin{pmatrix}x\\y\\z\end{pmatrix}$ indicates the coordinates of a point in three-dimensional space, t is time, v the wave velocity and $\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$ the Laplace operator.

1.2 Body waves

In elastic media, two types of seismic body waves exist which differ in their respective wave polarization [Aki and Richards, 2002]:

- Pressure waves, or P-waves, exhibit an associated wave motion which is longitudinal to the direction of propagation.
- Shear waves, or S-waves, display a motion which is transverse to the direction of propagation.

The wave velocities of both wave types are directly linked to the elastical properties of the medium, notably

$$v_P=\sqrt{\frac{\lambda+2\mu}{\rho}}, \tag{1.2}$$

$$v_S=\sqrt{\frac{\lambda}{\rho}}, \tag{1.3}$$

where λ and μ are the Lamé parameters and ρ the mass density of the medium. Both Lamé parameters parameterize the elastic moduli of homogeneous, perfectly elastic isotropic media and are linked by Poisson's ratio $\nu=\frac{\lambda}{2(\lambda+\mu)}$ [Lay and Wallace, 1995]. As both Lamé parameters are positive, the pressure wave velocitiy v_P is always larger than the shear wave velocity v_S. In a homogeneous medium, both P- and S-waves travel at their respective velocities at all frequencies and are therefore non-dispersive. As the polarization of S-waves is always perpendicular to the direction of propagation, two different types of S-waves can be defined related to their plane of polarization: The polarization of SV-waves lies in a vertical plane, whereas the polarization of SH-waves is horizontal.

1.3 Surface waves

The Earth's surface represents a fundamental border for seismic waves. At this interface, surface waves form. In contrast to body waves, the amplitude of surface waves decays exponentially with increasing distance from the interface. Consequently, the energy of surface waves decreases as $\frac{1}{r}$ with distance r from the wave's origin, but the energy of body waves decreases as $\frac{1}{r^2}$ [I]. Thus, in case of an earthquake, surface waves can cause damage at greater distances from the epicenter than body waves. Although other surface wave types exist (e.g. Stoneley waves), we will focus on Love and Rayleigh waves in the following, which are by far the most important surface waves in seismology.

1.3.1 Love waves

The existence of Love waves requires that the surface layer has a lower S-wave velocity than the underlying structure. Therefore, the simplest soil structure for which Love waves can occur consists of a single layer overlying a homogeneous half-space with higher S-wave velocity. In this case, SH-waves are reflected both at the surface and at the interface between layers. The waves which are trapped in this way, are called Love waves. As Love waves consist of SH-waves, the motion of a surface particle under the influence of a Love wave is parallel to the surface. Consequently, Love waves can only be recorded on the horizontal components of a seismic sensor. In contrast to body waves, Love waves travel with different velocities at different frequencies. Furthermore, harmonic modes of motion are possible. A very simple soil structure model is given in table 1.1. The P- and S-wave velocity profiles for this model are shown in figure 1.1 (a). Figure 1.1 (b) shows the Love wave dispersion curves for this model for the fundamental and the first four higher modes. This figure illustrates that the fundamental mode is present over the whole frequency range, while the harmonic modes occur at higher frequencies one after another. The Love waves' phase velocity is always comprised between the minimum and maximum S-wave velocities of the soil structure.

Table 1.1: Parameters of a very simple soil structure model: Thickness range d, pressure wave velocity V_P, shear wave velocity V_S and density ρ of the layers.

$d[m]$	$V_P[m/s]$	$V_S[m/s]$	$\rho[kg/m^3]$
$0-20$	1 500	400	2 000
$20-\infty$	5 600	3 200	2 000

[I] Assuming an undamped wave propagation, the total energy of a seismic wave will remain constant as it travels away from the source. For surface waves, at a distance r from the origin, this energy is dispersed on a ring of circumference $2\pi r$. For body waves, the energy is dispersed on the surface of a half-sphere $2\pi r^2$. As the energy of a wave is proportional to its amplitude squared, the amplitudes of surface and body waves decrease as $\frac{1}{\sqrt{r}}$ and $\frac{1}{r}$, respectively.

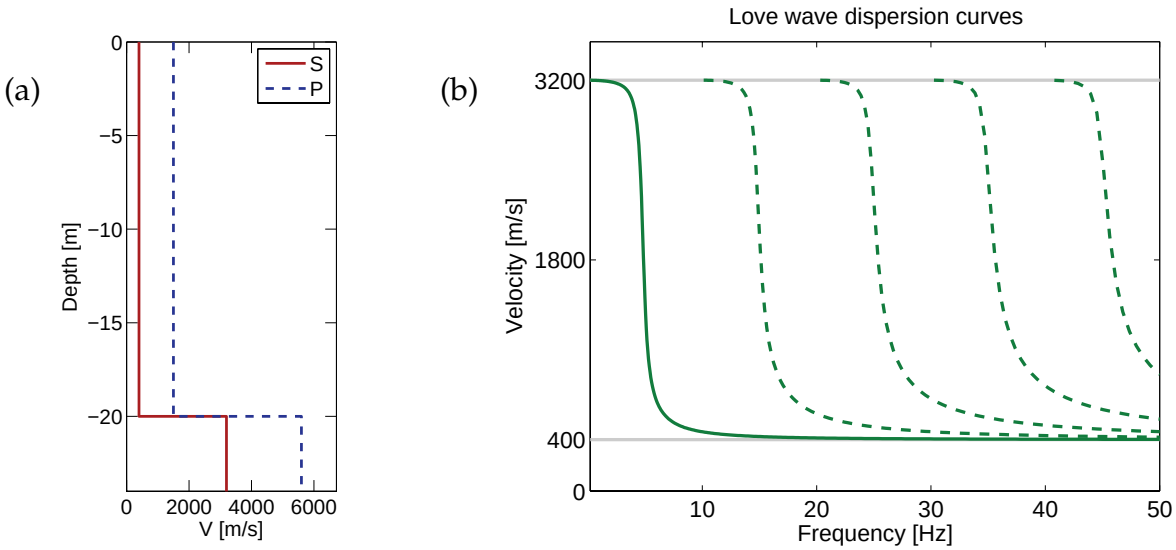


Figure 1.1: (a) S- and P-wave velocity profile for the model given in table 1.1. (b) Love wave dispersion curves for this model. The fundamental mode is plotted as a solid line, the dashed lines indicate the first four harmonic modes of this model.

1.3.2 Rayleigh waves

At the surface of the Earth, a coupled propagation of P- and SV-waves is possible. These waves are called Rayleigh waves. In contrast to Love waves, Rayleigh waves can form not only at layered soil structures, but also at the surface of a homogeneous half-space. The coupling of the SV-wave on the vertical component and the P-wave on the horizontal component is only possible if, disregarding the damping of waves, both components are phase-shifted by $\pm 90°$ [Aki and Richards, 2002]. However, due to anelasticity, the phase shift can differ from $\pm 90°$, but even for very anelastic material, this shift is in the order of $2°$ only [Boore and Toksöz, 1969].
The motion of a particle under the influence of a Rayleigh wave follows an ellipse. According to the actual sign of the phase shift, this motion is either retrograde or prograde (see figure 1.2), which describes the sense of rotation in relation to the propagation direction. Another property of Rayleigh waves is the ellipticity, which is defined as the ratio between the horizontal and the vertical amplitudes of the signal.
For the fundamental mode of a Rayleigh wave at the surface of a homogeneous half-space, the motion is always retrograde. In this case, the Rayleigh waves travel with the same velocity at all frequencies and are not dispersive. In this case, the Rayleigh wave velocity is a function of Poisson's ratio ν, but is always smaller than the S-wave velocity v_S. For example, for $\nu = 0.25$, the Rayleigh wave velocity of a homogeneous half-space is $v_{Rayleigh} = 0.92 v_S$. Furthermore, in this case the ellipticity is constant with frequency and also depends on ν [Malischewsky and Scherbaum, 2004].

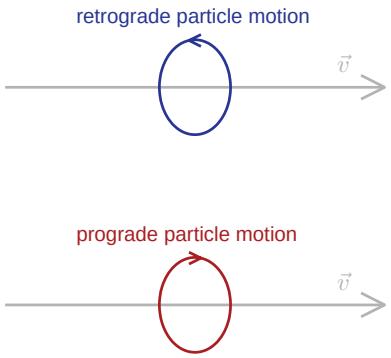


Figure 1.2: Diagram of retro- and prograde Rayleigh wave motion.

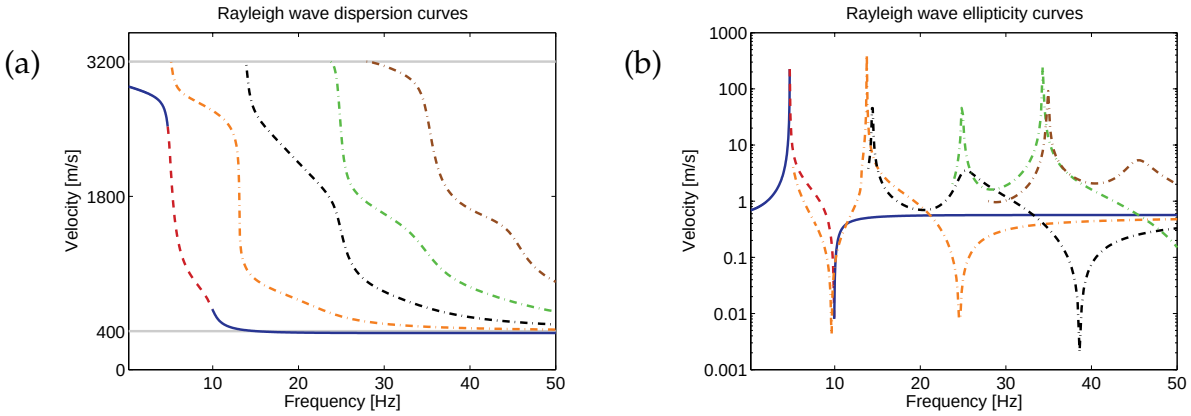


Figure 1.3: (a) Rayleigh wave dispersion curve for the fundamental and the first four harmonic modes for the model given in table 1.1. The fundamental mode is plotted as combination of a solid and a dotted line, where the solid line indicates retrograde particle motion and the dashed line prograde motion. The dispersion curves for the higher modes are given by dash-dotted lines. (b) Ellipticity curves for the same modes. At the right flank of the fundamental mode, the particle motion is prograde, whereas it is retrograde at the left flank and above the trough frequency. The harmonic modes are indicated by colored dash-dotted lines.

For a layered structure, however, Rayleigh waves are dispersive, the ellipticity varies with frequency and the wave motion can be prograde or retrograde, depending on frequency. The ellipticity as a function of frequency can show different behaviors, depending on the soil model. In case of a layer overlying a half-space, the ellipticity curve exhibits both a singularity peak and a trough if the impedance contrast between the uppermost layer and the half-space is strong (i.e. $v_{S2} > v_{S2}$). In this case, the particle motion for the fundamental Rayleigh wave mode is prograde in the frequency range between the peak and the trough, i.e. for the right flank of the peak. For the other parts of the ellipticity curve, the motion is retrograde. If the velocity profile does not exhibit a strong contrast, the ellipticity curve can still exhibit a peak, but neither a singular peak nor a trough. Then, the motion of the fundamental Rayleigh wave mode is retrograde for all frequencies.

In figure 1.3 (a), the dispersion curves for the fundamental and the first four harmonic modes of Rayleigh waves are shown for the model given in table 1.1. The fundamental mode is present at all frequencies, whereas each harmonic mode is present above a specific frequency only. At this frequency, the velocity of each mode equals the S-wave velocity at the second layer. At high frequencies, the fundamental mode does not reach deep enough to be influenced by the half-space. Therefore, its behavior is the same as on a half-space which has the properties of the surficial layer. Notably, for the model curves shown in figure 1.3 (a), the Rayleigh wave velocity is smaller than the respective S-wave velocity at frequencies above 15 Hz. In the opposite case, at very low frequencies, the Rayleigh wavelength is so large that the influence of the surficial layer becomes negligible and the wave behaves as on the top of a half-space. The ellipticity curves of the model are shown in figure 1.3 (b). For the fundamental mode, the curve exhibits a singularity peak and a trough and becomes flat at higher frequencies. Between the peak and trough frequencies, the Rayleigh wave motion is prograde and retrograde at the other frequencies. The ellipticity curves for the other modes show a more complicated behavior with multiple peaks and troughs.

1.4 Conclusion

The seismic wave field is dominated by four types of waves: P-, S-, Love and Rayleigh waves. Among them, Love and Rayleigh waves appear to be very interesting for retrieving information on the soil structure due to their dispersion properties. The dispersion of Love and Rayleigh waves can be depicted in a simple way. With increasing frequency, the wavelength decreases[II]. Surface waves are confined to the surface and their amplitudes decrease exponentially with depth. This decay depends on the wavelength. Consequently, a wave with a larger wavelength samples the soil structure to larger depths than a wave of smaller wavelength. Therefore, the low frequency part of the dispersion curve depends on a thicker part of the soil structure than the high frequency part. In general, the velocities of seismic waves increase with depth and low frequency surface waves travel faster than at higher frequencies.
Another important property carrying information on the soil structure is the ellipticity of Rayleigh waves. In contrast to the measurement of Love and Rayleigh wave dispersion curves, which can only be measured by using an array of seismic sensors, the ellipticity can be measured using a single seismic station. Nevertheless, the retrograde or prograde sense of rotation cannot be retrieved using a single seismic station because these notations only make sense in combination with the direction of propagation which cannot be measured by a single sensor.

[II]Frequency f and wavelength λ are linked via the wave velocity v (or the slowness $s = \frac{1}{v}$) in a simple way:

$$v = \lambda \cdot f.$$

Chapter 2

Single-sensor methods to estimate the polarization of seismic waves

This chapter presents methods which can be used to characterize the polarization of seismic waves using single sensors. First, some examples of established methods will be shown, including rectilinearity and planarity filters as well as a method using the analytic signal. Then, the widespread H/V technique is explained, before introducing DELFI and RayDec, two new methods which have been developed in the framework of this thesis work.

Ce chapitre présentera des méthodes pour caractériser la polarisation d'ondes sismiques en n'utilisant qu'un seul capteur. D'abord, quelques exemples de méthodes établies seront montrés, dont des filtres de réctilinéarité et de planarité et une méthode utilisant le signal analytique. Après, la technique très répandu du rapport H/V sera expliqué avant d'introduire DELFI et RayDec, deux méthodes nouvelles développées dans le cadre de ce travail de thèse.

2.1 Introduction

This chapter is devoted to the estimation of polarization parameters of seismic waves using single seismic sensors. The estimation of polarization is an important problem as it allows the discrimination between different wave types, e.g. body and surface waves, Love and Rayleigh waves. The knowledge of the wave properties and the composition of the wave field are important issues for various problems. These range from characterizing the wave source (i.e. estimating the azimuth and the incident angle) or the propagation medium (wave velocities of the different wave types, damping parameters) to the estimation of site effects (Rayleigh wave ellipticity). Some of these characteristics, for example the wave velocity, can only be estimated using simultaneous recordings of the wave field by multiple seismic sensors, but some others, like the Rayleigh wave ellipticity, can be determined using single seismic sensors. Seismic array methods will be treated in chapters 3 and 4, the focus of the present chapter lies on single-sensor methods.

The motion of a Rayleigh wave is confined in the plane formed by the vertical and the radial component. The motion of a Love wave, however, is linear, i.e. confined to a single horizontal direction. The basic idea of the methods which are presented in sections 2.2.1[Flinn, 1965, Montalbetti and Kanasewich, 1970, Jurkevics, 1988] is to develop filters which estimate the rectilinearity (i.e. the confinement of the signal in its dominant polarization direction) and planarity (i.e. the confinement in a plane). Using these filters, time windows with predominant rectilinear movement, i.e. P-, S- or Love waves, or planar movement, i.e. Rayleigh waves, can be extracted from the signal recordings. All these methods are based on eigenvalue decompositions of real signal covariance matrices. The method presented in section 2.2.2 [Vidale, 1986], uses the analytic signal instead of covariance matrices, but the objective of the method is the same. The use of the analytic signal allows in principle a more precise temporal signal characterization, as it is not necessary to average over multiple time samples as in the other methods.

The other presented methods aim to estimate the Rayleigh wave ellipticity. The first method, H/V [Nogoshi and Igarashi, 1971, Nakamura, 1989], presented in section 2.2.3, simply calculates the spectral ratio between the horizontal and the vertical components. This ratio is indeed linked to the resonance of the structure [Bonnefoy-Claudet et al., 2006] and corresponds to the Rayleigh wave ellipticity if the wave field consists exclusively of Rayleigh waves. However, in the presence of Love waves, the H/V curve misestimates ellipticity.

During this thesis work, two new methods have been developed which are specially designed to estimate ellipticity. The first method, DELFI, is based on the direct fitting of ellipses to time windows of the signal. The second method, RayDec, suppresses other wave types than Rayleigh waves by statistical means in order to estimate the ellipticity using ambient seismic vibrations. Both methods are presented in detail in sections 2.3 and 2.4, respectively.

2.2 Overview of existing methods

2.2.1 Rectilinear and planar polarization filters

2.2.1.1 Rectilinear polarization filter

The filter presented in the following was originally developed by Flinn [1965]. The filter detects the instantaneous linearity of a seismic signal and efficiently suppresses noise contributions. It is primarily designed to enhance the recordings of transient signals generated by earthquakes.
In a first step, the signals are rotated in the azimuth direction, which separates radial and transverse contributions[I]. For a time window of N data samples centered at t, the covariance matrix is calculated. Let

$$\mathbf{X}(t) = \begin{pmatrix} R(t_1) & R(t_2) & \cdots & R(t) & \cdots & R(t_N) \\ T(t_1) & T(t_2) & \cdots & T(t) & \cdots & T(t_N) \\ Z(t_1) & Z(t_2) & \cdots & Z(t) & \cdots & Z(t_N) \end{pmatrix} \tag{2.1}$$

be a $3 \times N$ data matrix describing the seismic records around the time t, where R, T and Z indicate the recordings of the radial, transverse and vertical components, respectively. Then the associated covariance matrix is given by

$$\mathbf{C}(t) = \frac{1}{N}\mathbf{X}^T(t)\mathbf{X}(t). \tag{2.2}$$

The eigenvalues of $\mathbf{C}(t)$ are λ_1, λ_2 and λ_3 (with $\lambda_1 > \lambda_2 > \lambda_3$). Now, a function describing the rectilinearity of the signal is defined by

$$G(\lambda_1, \lambda_2) = 1 - \frac{\lambda_2}{\lambda_1}. \tag{2.3}$$

If the signal has a principal polarization direction, λ_1 will be large compared to λ_2 and $G(\lambda_1, \lambda_2)$ close to unity. In the opposite case, λ_1 and λ_2 will be comparable and $G(\lambda_1, \lambda_2)$ small.
This filter technique was improved by Montalbetti and Kanasewich [1970] by adding an arbitrary parameter n in equation (2.3), which transforms to

$$F(\lambda_1, \lambda_2) = 1 - \left(\frac{\lambda_2}{\lambda_1}\right)^n. \tag{2.4}$$

The rectilinearity of the signal at time t is then given by

$$RL(t) = \left(F(\lambda_1, \lambda_2)\right)^J, \tag{2.5}$$

where J is an additional arbitrary parameter. $\vec{e}_1 = \begin{pmatrix} e_R \\ e_T \\ e_Z \end{pmatrix}$, the eigenvector associated to λ_1, indicates the direction of maximum polarization, i.e. the repartition of the wave energy

[I] The azimuth is either given if the earthquake epicenter is known or has to be detected by an array method.

on the different components. In this way, time-dependent direction functions can be defined by raising the different components of the eigenvector to an arbitrary power K, which yields

$$D_R(t) = (e_R)^K, \tag{2.6}$$
$$D_T(t) = (e_T)^K, \tag{2.7}$$
$$D_Z(t) = (e_Z)^K. \tag{2.8}$$

By changing the arbitrary constants n, J and K in equations (2.4) - (2.8) of the method, the results can be tuned[II]. By averaging over some neighbouring values, the rectilinearity $RL(t)$ and direction functions can be smoothed. The final signals are then obtained by

$$R_f(t) = R(t) \cdot RL(t) \cdot D_R(t), \tag{2.9}$$
$$T_f(t) = T(t) \cdot RL(t) \cdot D_T(t), \tag{2.10}$$
$$Z_f(t) = Z(t) \cdot RL(t) \cdot D_Z(t). \tag{2.11}$$

2.2.1.2 Planar polarization filter

Jurkevics [1988] expanded the use of the rectilinear filter described above to planar waves. The calculation of the covariance matrix is the same as given in equation (2.2). A slightly modified version of equation (2.4) is used to describe the rectilinearity:

$$R(\lambda_1, \lambda_2, \lambda_3) = 1 - \left(\frac{\lambda_2 + \lambda_3}{2\lambda_1} \right). \tag{2.12}$$

Still, this function is close to one if the signal is linearly polarized, i.e. if λ_1 is large compared to the other eigenvalues. Additionally, the degree of planarity of the signal is measured by the property

$$P(\lambda_1, \lambda_2, \lambda_3) = 1 - \left(\frac{2\lambda_3}{\lambda_1 + \lambda_2} \right). \tag{2.13}$$

As the motion of Rayleigh waves is confined to a plane, a combination of both functions can be used to discriminate against other wave types. Furthermore, to improve the signal-to-noise ratio, Jurkevics [1988] proposed to use an array of seismic sensors and to average the covariance matrices of the different sensors, which yields

$$\overline{\mathbf{C}}(t) = \frac{1}{N_S} \sum_{k=1}^{N_S} \mathbf{C}_k, \tag{2.14}$$

where N_S is the number of sensors and $\mathbf{C}_k$ the covariance matrix of the kth sensor.

[II] Montalbetti and Kanasewich [1970] indicate that good results are obtained by setting $n = 0.5$ or 1, $J = 1$ and $K = 2$.

2.2.2 Polarization analysis using the analytic signal

The following method has been proposed by Vidale [1986]. If the real-valued three-component measurements for a single seismic sensor are given by $x_{k,r}(t)$ (where $k \in \{1,2,3\}$ indicates the direction; 1 stands for eastern, 2 for northern and 3 for vertical component), the analytic signal is built by

$$x_k(t) = x_{k,r}(t) + i\mathcal{H}(x_{k,r}(t)), \tag{2.15}$$

where $\mathcal{H}$ represents the Hilbert transform, defined by $\mathcal{H}(f(x)) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(x)}{y-x}dx$. In this way, the real-valued signals become complex (without increasing the information content). In fact, the analytic signal corresponds to a real signal where all negative frequency contributions have disappeared without changing the signal's energy, i.e. in the Fourier transform the positive frequency contributions have doubled. By using the analytic signal, it is possible to calculate an instantaneous covariance function

$$\mathbf{C}(t) = \begin{pmatrix} x_1(t)x_1^*(t) & x_1(t)x_2^*(t) & x_1(t)x_3^*(t) \\ x_2(t)x_1^*(t) & x_2(t)x_2^*(t) & x_2(t)x_3^*(t) \\ x_3(t)x_1^*(t) & x_3(t)x_2^*(t) & x_3(t)x_3^*(t) \end{pmatrix}. \tag{2.16}$$

As the covariance matrix is by construction Hermitian, its eigenvalues λ_n ($n \in \{1,2,3\}$, $\lambda_1 > \lambda_2 > \lambda_3$) are real, whereas the normed eigenvectors $\vec{u}_n = \begin{pmatrix} u_{n,1} \\ u_{n,2} \\ u_{n,3} \end{pmatrix}$ are, in general, complex. The eigenvector associated to the largest eigenvalue λ_1 indicates the direction of maximum polarization. In fact, $\vec{u}_1$ represents a set of eigenvectors which are associated to λ_1. The product of $\vec{u}_1$ and $e^{i\alpha}$ (with an arbitrary α) is still an eigenvector, because $\mathbf{C}\vec{u}_1 e^{i\alpha} = \lambda_1 e^{i\alpha}$. From this set of eigenvectors, the vector whose real component is maximum is chosen. This is done by searching the α maximizing the length X of the real component of $\vec{u}_1 e^{i\alpha}$ given by

$$X = \left|\Re(\vec{u}_1 e^{i\alpha})\right|. \tag{2.17}$$

The elliptical polarization component is estimated by

$$P_E = \frac{\sqrt{1-X^2}}{X}. \tag{2.18}$$

Due to the normalization of $\vec{u}_1$, the length of the complex part of $\vec{u}_1$ is given by $\sqrt{1-X^2}$ and P_E represents the ratio between imaginary and real part of the eigenvector. For linear polarization, $P_E = 0$, and for circular polarization, $P_E = 1$. By using the components of $\vec{u}_1$, the azimuth ϑ and angle of incidence δ of the wave can be estimated by

$$\vartheta = \arctan\left(\frac{\Re(u_{1,2})}{\Re(u_{1,1})}\right), \tag{2.19}$$

$$\delta = \arctan\left(\frac{\Re(u_{1,3})}{\sqrt{(\Re(u_{1,1}))^2 + (\Re(u_{1,2}))^2}}\right), \tag{2.20}$$

where $u_{1,k}$ indicates the component of $\vec{u}_1$ in direction k. Further polarization parameters can be defined in a way similar to the rectilinearity and planarity filters of Jurkevics [1988]. The strength of polarization is defined in a way slightly different from equation (2.12) by

$$P_S = 1 - \left(\frac{\lambda_2 + \lambda_3}{\lambda_1} \right) \tag{2.21}$$

and the degree of planarity is estimated by

$$P_P = 1 - \left(\frac{\lambda_3}{\lambda_2} \right). \tag{2.22}$$

If the signal is polarized with a single dominant component, P_S is close to 1. If the signal in both other components is comparable to the dominant one, P_S is low. In the same way, P_P is close to 1 if the intermediate component is much larger than the smallest one and close to 0 if both components are comparable.
The advantage of using the analytic signal is that the analysis can be performed for each time sample independently, i.e. without having to average over a time window in the seismogram as would be necessary when using real-valued signals, but it is necessary to use the time-series of the signal to calculate the Hilbert transform. Nevertheless, the method can only work if a single seismic signal is present. If two waves arrive at the same time, calculating the analytical signal gives wrong results.

2.2.3 H/V

The horizontal-to-vertical (H/V) method was introduced by the publication (in Japanese language) of Nogoshi and Igarashi [1971] and popularized among the English-speaking science community by Nakamura [1989]. Since then, it is a widespread technique which helps to characterize the soil by simple means. The method consists in the calculation of the ratio between the horizontal and vertical Fourier spectra of the seismic noise recordings of a single three-component seismic sensor. If the Fourier transforms of the recordings of the three components are $X_{east}(f)$, $X_{north}(f)$ and $X_{vert}(f)$, respectively, then the H/V ratio is calculated by[III]

$$H/V = \frac{\sqrt{|X_{east}(f)|^2 + |X_{north}(f)|^2}}{|X_{vert}(f)|}. \tag{2.23}$$

Although the theory behind the H/V ratio is not yet completely understood, empirical evidence shows that a peak in the H/V curve is correlated with the resonance frequency of the structure [Bonnefoy-Claudet et al., 2006, Haghshenas et al., 2008]. Furthermore, for a single Rayleigh wave crossing the structure, the H/V ratio corresponds to the definition of the ellipticity curve. However, in the general case Love waves are also present [Köhler et al., 2006, Endrun, 2010] and the H/V ratio will not provide the ellipticity.

[III] Sometimes, this ratio is defined with an additional $\sqrt{2}$ in the denominator.

2.3 DELFI

The following method has been developed during this work. DELFI (Direct ELlipse FItting for Rayleigh wave ellipticity estimation) is a method which estimates the ellipticity of Rayleigh waves by directly fitting an ellipse to the filtered data [Hobiger et al., 2009b]. There are two processing steps, the first one determines the vertical plane where the signal is confined and projects the signal in this plane, the second fits an ellipse to the obtained two-dimensional signal. Figure 2.1 gives an overview of the principles of the method. The algorithm is explained in the following.

2.3.1 Plane detection

Consider the three-dimensional signal recorded by a single seismic sensor. This signal can be represented as a time series $\mathbf{S}(t) = [E(t)\ N(t)\ Z(t)]$, where $E(t)$, $N(t)$ and $Z(t)$ are the signals of the eastern, northern and vertical component, respectively. In this way the signal $\mathbf{S}(t)$ takes values in $\mathbb{R}^3$. In a first step, the signal $\mathbf{S}(t)$ is filtered around a central frequency f with a filter bandwidth df which yields the filtered signal $\mathbf{S_f}(t) = \left[E_f(t)\ N_f(t)\ Z_f(t)\right]$.
Then, for each frequency of interest, the following processing is performed independently. First, the filtered signal is cut into blocks of time length $T = n_P/f$, which corresponds to signals of n_P periods at the given frequency. If we denote the sample rate of the original signal by dt, each cut block of signal includes $N = T/dt$ data points. The three-component signal of one of these blocks $\mathbf{S}_b = [E_b\ N_b\ Z_b]$ forms a matrix of dimension $N \times 3$. The starting and ending times of a block are given by $t_{b,s}$ and $t_{b,e}$, respectively.
As the polarization of Rayleigh waves is confined to a plane formed by a horizontal and a vertical axis, the first step is to identify the vertical plane where the signal is confined the most. This is done by using the signal of both horizontal components. The plane can be defined by its normal vector $\vec{a} = (a_E, a_N)^T$. For each data point $\vec{p}_i = (E_i, N_i)$, the distance to the plane is given by $d_i = \vec{p}_i \cdot \vec{a}$. We fit the plane in a least-squares sense by minimizing the functional

$$D = \sum_{i=1}^{N} d_i^2, \tag{2.24}$$

which can be written in matrix form as

$$\begin{aligned} D &= (\mathbf{p} \cdot \vec{a})^T (\mathbf{p} \cdot \vec{a}) \\ &= \vec{a}^T \cdot \mathbf{p}^T \cdot \mathbf{p} \cdot \vec{a} \\ &= \vec{a}^T \cdot \mathbf{S} \cdot \vec{a}, \end{aligned} \tag{2.25}$$

where $\mathbf{p} = [E_b\ N_b]$ is the matrix of all data points and $\mathbf{S} = \mathbf{p}^T \cdot \mathbf{p}$ the position covariance matrix. The normalized eigenvectors $\vec{u}_i$ (with associated eigenvalues λ_i, for $i \in \{1, 2\}$, with $\lambda_1 < \lambda_2$) of $\mathbf{S}$ form a basis of the two-dimensional space.

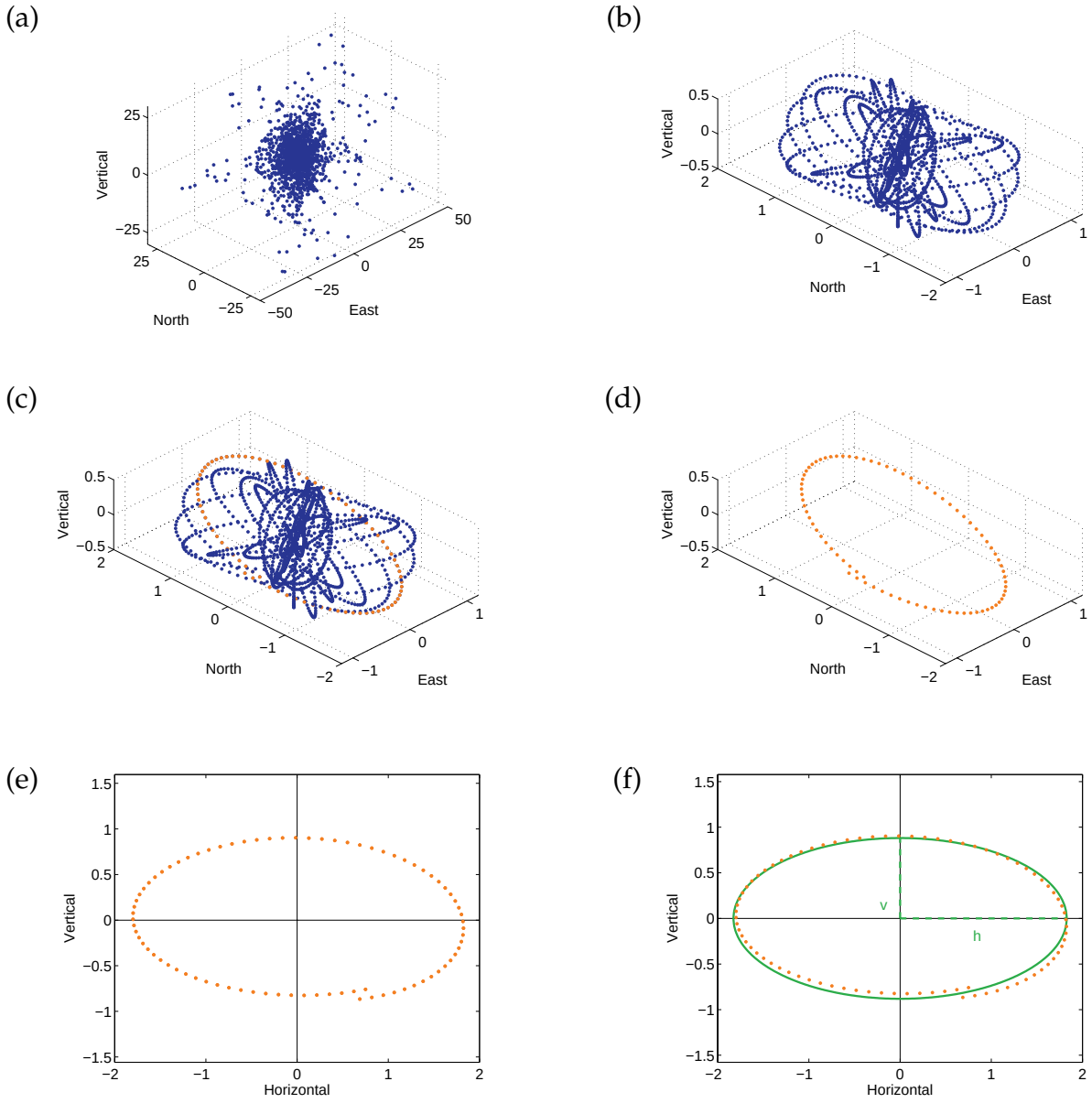


Figure 2.1: Graphic outline of the DELFI method: (a) A single three-component seismic sensor records a time series of data points in 3D space. (b) The signal is filtered around f, the frequency of interest. (c) The filtered signal is cut into blocks of equal length (typically one period). (d) Each of these blocks is extracted and the plane, where the signal is confined, is searched for. (e) The signal block is projected into this plane. (f) An ellipse is fitted. Repeated processing of steps (d)-(f) for every signal block cut in (c) yields the ellipticity estimation at frequency f.

Therefore, $\vec{a}$ can be expressed as a linear combination of $\vec{u}_1$ and $\vec{u}_2$ by

$$\vec{a} = \sum_{i=1}^{2} k_i \vec{u}_i. \tag{2.26}$$

Thus, exploiting the properties $\mathbf{S}\vec{u}_i = \lambda_i \vec{u}_i$ and $\vec{u}_j^T \vec{u}_i = \delta_{ij}$, the functional can be written as

$$\begin{aligned} D &= \sum_{i=1}^{2}\sum_{j=1}^{2} k_i k_j \vec{u}_i^T \mathbf{S} \vec{u}_j \\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} k_i k_j \vec{u}_i^T \vec{u}_j \lambda_j \\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} k_i k_j \delta_{ij} \lambda_j \\ &= k_1^2 \lambda_1 + k_2^2 \lambda_2 \end{aligned} \tag{2.27}$$

As the vector $\vec{a}$ is normalized, $k_1^2 + k_2^2 = 1$. As $\lambda_1 < \lambda_2$, D is minimized by $k_1 = 1$ and $k_2 = 0$. It follows that $\vec{a} = \vec{u}_1$. As $\vec{a}$ is the normal vector of the plane, the other eigenvector $\vec{u}_2$ and the vertical unit vector span the plane. The horizontal components E_b and N_b can now be projected onto the vector $\vec{u}_2$ to form the horizontal signal

$$H_b = [E_b \; N_b] \cdot \vec{u}_2. \tag{2.28}$$

In this way, the original three-dimensional signal is transformed into the two-dimensional signal

$$\mathbf{S}'_b = [H_b \; Z_b]. \tag{2.29}$$

2.3.2 Ellipse fitting

The next step consists in fitting an ellipse to the data matrix $\mathbf{S}'_b$. Fitzgibbon et al. [1999] proposed an algorithm which fits an ellipse to any two-dimensional cloud of data points. In general, six parameters are necessary to define a two-dimensional ellipse: two coordinates defining the center point and two two-dimensional vectors defining the semi-major and semi-minor axes, respectively[IV]. Of course, the ellipse can be parameterized in a different way without changing the number of necessary parameters. For the case where we search for Rayleigh waves, the number of parameters can be reduced considerably. Firstly, we know that the center point of the ellipse must lie on the origin of the sensor's coordinate system[V].

[IV] In the three-dimensional case, nine parameters are necessary: each of the three vectors is defined by three parameters. In the three-dimensional case, the plane of the ellipse can be easily found by using a three-dimensional covariance matrix. By projecting the signal on this plane, the problem reduces to the two-dimensional case.

[V] The presence of different signals, especially of frequency content different from the analyzed Rayleigh wave, could shift the ellipse including the center point. This effect is inhibited by the narrow-band filtering in the first step of the processing.

In addition, both axes of the ellipse of the Rayleigh wave motion lie in the horizontal and vertical direction if the soil is perfectly elastic. Boore and Toksöz [1969] showed that anelasticity leads to a shift of the phase angle between the vertical and the horizontal motion. However, they showed that even for very anelastic material ($Q_\alpha = 10$, $Q_\beta = 4$, $\beta = 0.55\alpha$) this shift does not exceed 2°. As such a shift is hardly ever measurable, we impose that both axes lie in horizontal and vertical direction, respectively. In this way, the Rayleigh wave motion can be parameterized by only two parameters, namely the horizontal and the vertical axes. Both are necessary to calculate ellipticity.
The following function describes an ellipse with a horizontal and a vertical axis:

$$ax^2 + by^2 - 1 = 0, \tag{2.30}$$

where a and b are parameters related to the horizontal and vertical axes, respectively. If we denote the lengths of the horizontal and the vertical axes by h and v, respectively, the following relations hold:

$$a = \frac{1}{h^2}, b = \frac{1}{v^2}. \tag{2.31}$$

The distance of an arbitrary point (x, y) from the ellipse is given by

$$F(x, y) = ax^2 + by^2 - 1. \tag{2.32}$$

To a set of data points (x_i, y_i) with $i \in \{1, 2, \ldots, N\}$, an ellipse can be fitted by minimizing the functional

$$D = \sum_{i=1}^{N} F(x_i,y_i)^2. \tag{2.33}$$

This leads to the following set of equations:

$$\frac{\partial D}{\partial a} = a \sum_{i=1}^{N} x_i^4 + b \sum_{i=1}^{N} x_i^2 y_i^2 - \sum_{i=1}^{N} x_i^2 = 0, \tag{2.34}$$

$$\frac{\partial D}{\partial b} = a \sum_{i=1}^{N} x_i^2 y_i^2 + b \sum_{i=1}^{N} y_i^4 - \sum_{i=1}^{N} y_i^2 = 0. \tag{2.35}$$

This can be written in matrix form as

$$\mathbf{S} \cdot \vec{A} = \vec{C}, \tag{2.36}$$

with

$$\mathbf{S} = \begin{pmatrix} \sum x_i^4 & \sum x_i^2 y_i^2 \\ \sum x_i^2 y_i^2 & \sum y_i^4 \end{pmatrix}, \tag{2.37}$$

$$\vec{A} = \begin{pmatrix} a \\ b \end{pmatrix}, \tag{2.38}$$

$$\vec{C} = \begin{pmatrix} \sum x_i^2 \\ \sum y_i^2 \end{pmatrix}, \tag{2.39}$$

where the summations stretch over all indices $1 \leq i \leq N$. This equation can be easily solved by[VI]

$$\vec{A} = \mathbf{S}^{-1} \cdot \vec{C}. \tag{2.40}$$

The properties a and b can be transformed into the horizontal and vertical axes by using equation (2.31).
The presented ellipse-fitting algorithm is applied to every signal block $\mathbf{S}_b$, which yields the horizontal and vertical axes h_b and v_b, as well as the value of the functional D which serves as weighting factor D_b. This processing is performed for all signal blocks and the ellipticity at frequency f is finally calculated by

$$\epsilon_{DELFI}(f) = \frac{\sum_b h_b / D_b}{\sum_b v_b / D_b}. \tag{2.41}$$

2.3.3 Free parameters

Table 2.1: Values for the horizontal and vertical axes as well as the resulting ellipticity for the ellipse fits of figure 2.2.

Used data points	Horizontal axis	Vertical axis	Ellipticity
original ellipse	2.000	1.000	2.00
all points	2.072	0.973	2.13
left half	2.058	0.951	2.16
right half	2.085	0.993	2.10
quarter 1	2.008	1.051	1.91
quarter 2	2.012	0.933	2.16
quarter 3	2.086	0.975	2.14
quarter 4	2.220	0.923	2.41

In the algorithm, there are only two free parameters: df, the bandwidth of the filter and n_P, the number of periods in each signal block. The filter bandwidth should be adjusted in such a way to exclude contributions of other frequencies to a maximum. The number of periods in the signal block should not exceed 1, as the amplitude of seismic waves can evolve with time. Therefore, it is improbable that two subsequent signal blocks have the same amplitude, even if their respective ellipticities are identical. The ellipse-fitting algorithm does not set a lower boundary of how much signal has to be processed. However, the sample rate dt of the signal sets a lower limit for n_P, as an ellipse estimation based on two or three data points does not make much sense. In the same time, if only a part of the ellipse is covered by data points, the ellipse estimation

[VI] In the general case, the position covariance matrix $\mathbf{S}$ is invertible. However, if the data points lie on a degenerated ellipse, i.e. on a line, $\mathbf{S}$ will be singular and cannot be inverted. Then, $\mathbf{S}$ can be transformed into an invertible form by adding small numbers on the diagonal entries.

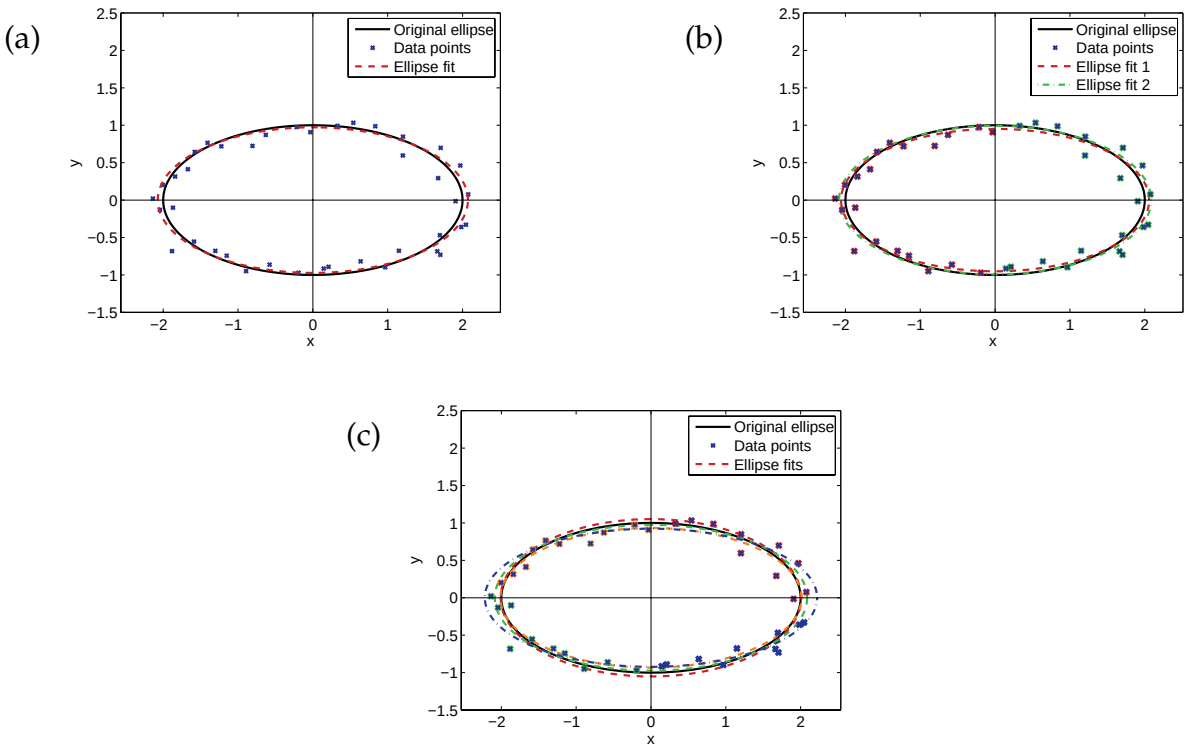


Figure 2.2: Examples of fitting different parts of an ellipticity with added noise. 40 data points lying on the original ellipse have been generated and noise of the same amplitude in horizontal and vertical direction has been added. In (a), an ellipse has been fitted to all data points, in (b) the left and the right half of the data points have been fitted separately and in (c) four quarters of the data (corresponding to the four quadrants of the coordinate system) have been fitted separately.

gets instable. Figure 2.2 shows an example for fitting different parts of an ellipse. The horizontal axis of the original ellipse in the figure has a length of 2, the vertical one is 1. By fitting data which are exactly on an ellipse, the ellipse-fitting algorithm would always yield the same result. Therefore, a small amount of noise has been added to the 40 original data points. Table 2.1 indicates the horizontal and vertical axes resulting from an ellipse fit of different parts of the data. In Figure 2.2 (a), all data points are used to fit the ellipse. The horizontal axis is slightly overestimated. Fitting the left or the right half of the data points (see figure 2.2 (b)) yields a qualitatively comparable result to the fit of all data points. When fitting only a quarter of the data points, however, the results are quite variable (see figure 2.2 (c)). As less data points are used, the ellipse estimation is less robust. The ellipses found for the different sets of data points are quite different. However, the additional noise amplitude is equal in horizontal and in vertical direction, but this leads to a systematic overestimation of the horizontal and underestimation of the vertical axis. The ellipticity value is biased to higher values. For the practical use of DELFI, it is advisable to use more than half a period of the signals. As in general, the ellipse estimation is more robust when using longer signals, we used $n_P = 1$ in the applications of DELFI.

2.4 RayDec

The name RayDec stands for "Rayleigh wave ellipticity estimation using the random decrement technique". Figure 2.3 gives an overview of the principles of the method. The technique was published in *Geophysical Review Letters*, the following article ("Single Station Determination of Rayleigh Wave Ellipticity by Using the Random Decrement Technique (RayDec)" by Hobiger, Bard, Cornou, and Le Bihan, 2009a) describes the method in detail.

2.4.1 Abstract

Rayleigh wave ellipticity as a function of frequency is closely linked to underground structure, i.e. shear wave velocity profile and sediment thickness. The possibility to calculate these underground properties by inverting ellipticity curves has recently been shown. We propose a new technique enabling the Rayleigh wave ellipticity to be recovered over a wide frequency range by using ambient noise recordings. Based on the random decrement technique commonly used to characterize dynamic parameters of buildings, this method eliminates all wave types except Rayleigh waves. We apply the method to noise synthetics simulated for different underground structures and show its applicability to real seismic noise data.

2.4.2 Introduction

In order to properly evaluate local site amplification it is important to have a good knowledge of the local soil properties, *i.e.* shear-wave velocities and sediment thickness. These soil properties can be obtained by inverting dispersion curves. One way is to use the dispersion curves of surface waves, as derived from active or passive seismic array measurements. However, it usually provides reliable information at frequencies higher than the resonance frequency and the inverted velocity profiles do not carry information on the deeper structures. Additional information around the resonance frequency can be added by using H/V spectra. Recent studies have shown the possibility to evaluate the S-wave velocity profile by inversion of H/V spectra [Fäh et al., 2003, Arai and Tokimatsu, 2004] or joint inversion of H/V spectra and dispersion curves [Arai and Tokimatsu, 2005, Picozzi et al., 2005]. The classical H/V technique [Nakamura, 1989] calculates the ratio between the horizontal and the vertical spectrum of ambient vibrations recorded on a single three-component seismic sensor by $H/V(f) = \frac{\sqrt{|E(f)|^2+|N(f)|^2}}{\sqrt{2}\,|Z(f)|}$ where $E(f)$, $N(f)$ and $Z(f)$ are the spectra of the east-west, north-south and vertical displacements of the sensor. Both Rayleigh and Love waves contribute to the H/V spectrum [Bonnefoy-Claudet et al., 2008] and the presence of both surface wave types has to be accounted for in inversion algorithms. Therefore, the proportion between Rayleigh and Love waves in the noise wave field has to be inferred or assumed prior to the inversion [Scherbaum et al., 2003, Arai and Tokimatsu, 2004, Köhler et al., 2006]. Extraction of Rayleigh wave ellipticity and its subsequent inversion [Boore and Toksöz, 1969, Fäh et al., 2001, Malischewsky and Scherbaum, 2004] would avoid making such assumptions.

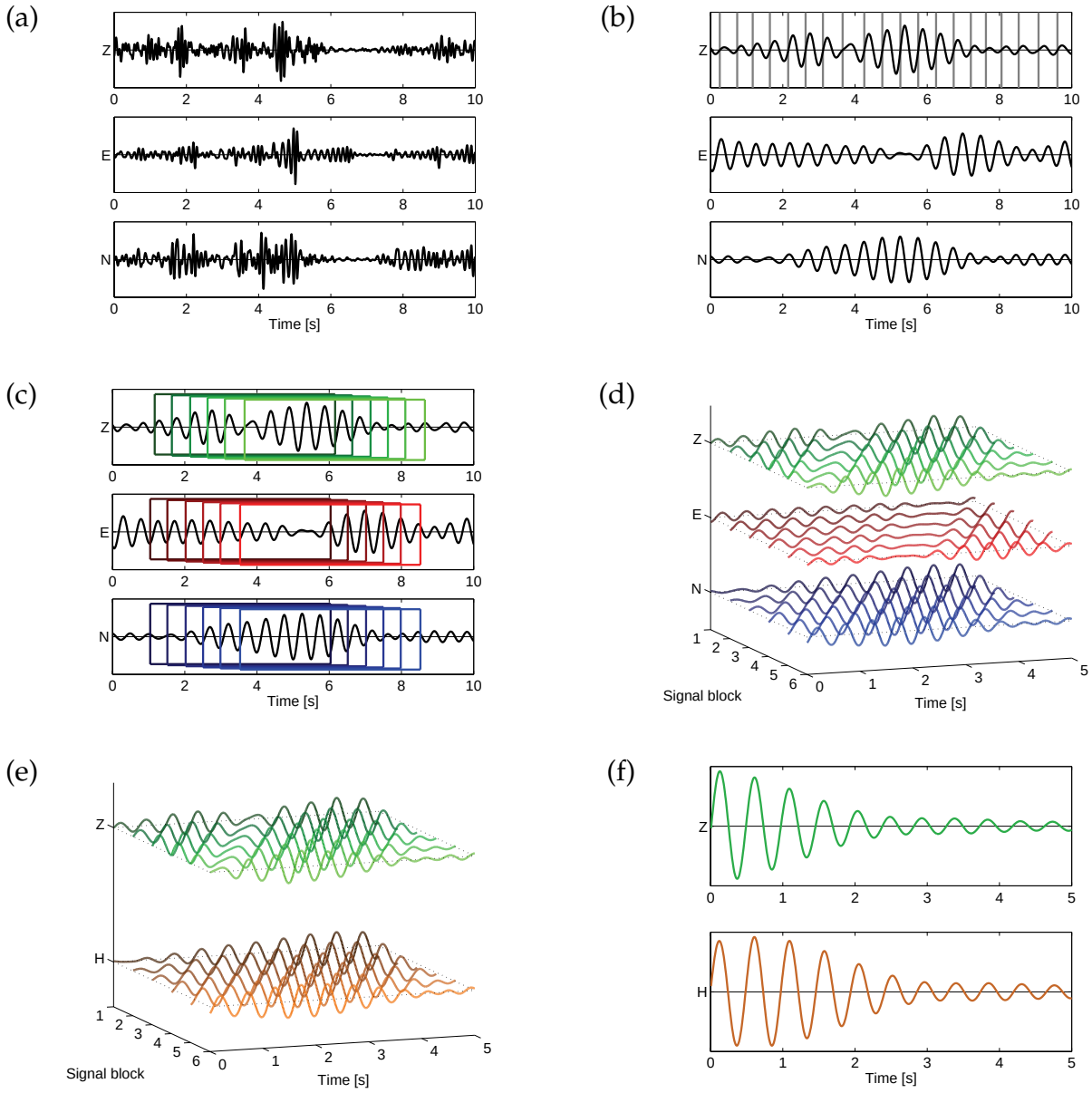


Figure 2.3: Graphic outline of the RayDec method: (a) A single three-component seismic sensor records a time series of data. (b) On the vertical component of the filtered signal, all zero crossings from negative to positive are searched for. (c) For every identified zero crossing, signals of length Δ are extracted on all three components, delaying the horizontal by $\frac{1}{4f}$, which corresponds to the vertical 90° phase shift to the vertical component. (d) These signal blocks are stacked. (e) Both horizontal signals are projected in the direction which maximizes the correlation to the vertical signal. (f) All these signals are summed, which yields a single vertical and horizontal signal. The ellipticity at frequency f is then estimated by comparing the respective energies of these summed signals.

We propose a new method (RayDec) to estimate the Rayleigh wave ellipticity spectrum by using the random decrement technique [Asmussen, 1997]. A usual application of the random decrement technique is the measurement of resonance frequencies and damping parameters of buildings [Dunand, 2005, Michel et al., 2008]. We propose to apply the random decrement technique to three-component seismometer records of seismic noise. By using the vertical component as a master trigger and stacking a large number of horizontal and vertical signals, Rayleigh waves will be emphasized with respect to Love and body waves. Therefore, the obtained ellipticity curve will be closer to the true curve than the H/V curve.
In this paper, we will present the method, apply it to synthetic seismic noise and finally show its applicability to real seismic noise records with one example.

2.4.3 Methodology

Assume that a three-component sensor record consists of three time-series of seismic noise ($e(t)$ in east-west direction, $n(t)$ in north-south direction and $v(t)$ in vertical direction). Each of these time-series has a time length T and consists of N data points. In order to extract the ellipticity of Rayleigh waves we suggest the following method, based on the random decrement technique [Asmussen, 1997]. The principles of the method are to emphasize Rayleigh waves with respect to other wave types by summing a large number of specially tuned signal windows, calculate the energy on the vertical and horizontal summed signals and henceforward estimate Rayleigh wave ellipticity. The first step consists of filtering the signals by using a narrow-band Chebyshev filter of order 4 and bandwidth df centered on frequency f. We use the Chebyshev filter because of its sharpness. Filtering was performed on a linear frequency scale. The resulting signals are called $e_f(t)$, $n_f(t)$ and $v_f(t)$.

Then we search for all times τ_i where the signal on the vertical component changes its sign from negative to positive ($v_f(\tau_i) \leq 0$, $v_f(\tau_i + dt) > 0$). Every time this condition is satisfied, we store signals of length Δ on all three components, leading to buffered signals $v_{f,b,i}(t)$, $e_{f,b,i}(t)$ and $n_{f,b,i}(t)$ defined in terms of the original signals as (for $0 \leq t \leq \Delta$):

$$\begin{aligned} v_{f,b,i}(t) &= v_f(\tau_i + t), \\ e_{f,b,i}(t) &= e_f(\tau_i - \frac{1}{4f} + t), \\ n_{f,b,i}(t) &= n_f(\tau_i - \frac{1}{4f} + t). \end{aligned} \tag{2.42}$$

The shift of $\frac{1}{4f}$ on the eastern and northern components accounts for the 90° phase shift between the vertical and horizontal components of Rayleigh waves. Then we build a horizontal signal $h_{f,b,i}(t)$ by projecting the east-west and north-south components onto an axis forming an azimuth angle ϑ_i with the northern direction by

$$h_{f,b,i}(t) = \sin(\vartheta_i) \cdot e_{f,b,i}(t) + \cos(\vartheta_i) \cdot n_{f,b,i}(t). \tag{2.43}$$

The angle ϑ_i is chosen in such a way that it maximizes the correlation between the vertical signal $v_{f,b,i}(t)$ and the horizontal signal $h_{f,b,i}(t)$ given by:

$$C_{f,b,i}(\vartheta_i) = \int_0^\Delta v_{f,b,i}(t) \cdot h_{f,b,i}(t)dt$$
$$C_{f,b,i}(\vartheta_i) = \sin(\vartheta_i) \int_0^\Delta v_{f,b,i}(t) \cdot e_{f,b,i}(t)dt + \cos(\vartheta_i) \int_0^\Delta v_{f,b,i}(t) \cdot n_{f,b,i}(t)dt. \quad (2.44)$$

The correlation is maximized by

$$\vartheta_i = \tan^{-1}\left(\frac{\int_0^\Delta v_{f,b,i}(t) \cdot e_{f,b,i}(t)dt}{\int_0^\Delta v_{f,b,i}(t) \cdot n_{f,b,i}(t)dt}\right). \quad (2.45)$$

There are two possible $\vartheta_i \in [0°, 360°[$ satisfying equation (2.45), with $\vartheta_{i,2} = \vartheta_{i,1} + 180°$, one of them leading to a negative, the other to a positive correlation. The ϑ_i leading to a positive correlation is chosen. The method does not distinguish between prograde and retrograde particle motion as prograde Rayleigh waves are treated as retrograde waves arriving from the opposite azimuth.

Although the chosen angle ϑ_i maximizes the correlation between $v_{f,b,i}(t)$ and $h_{f,b,i}(t)$, the correlation can be low. Thus, we make use of the correlation factor defined as

$$c_{f,b,i} = \frac{\int_0^\Delta v_{f,b,i}(t) \cdot h_{f,b,i}(t)dt}{\sqrt{\int_0^\Delta v_{f,b,i}^2(t)dt \cdot \int_0^\Delta h_{f,b,i}^2(t)dt}}. \quad (2.46)$$

In the case of perfect correlation between $v_{f,b,i}(t)$ and $h_{f,b,i}(t)$ this factor equals 1. On the other side, it will vanish when the vertical and horizontal signals are not correlated at all. The signals $v_{f,b,i}(t)$ and $h_{f,b,i}(t)$ are weigthed by $c_{f,b,i}^2$ and summed for all times τ_i to obtain two buffered signal windows of length Δ:

$$v_{f,S}(t) = \sum_i c_{f,b,i}^2 \cdot v_{f,b,i}(t) \quad (2.47)$$
$$h_{f,S}(t) = \sum_i c_{f,b,i}^2 \cdot h_{f,b,i}(t). \quad (2.48)$$

Finally, the ellipticity ϵ is calculated as the square root of the ratio of the energies in the buffered signal windows $h_{f,S}(t)$ and $v_{f,S}(t)$:

$$\epsilon = \sqrt{\frac{\int_0^\Delta (h_{f,S}(t))^2 dt}{\int_0^\Delta (v_{f,S}(t))^2 dt}}. \quad (2.49)$$

By repeated filtering of the signal over the entire frequency range and applying the presented method the ellipticity of Rayleigh waves as a function of frequency is determined.

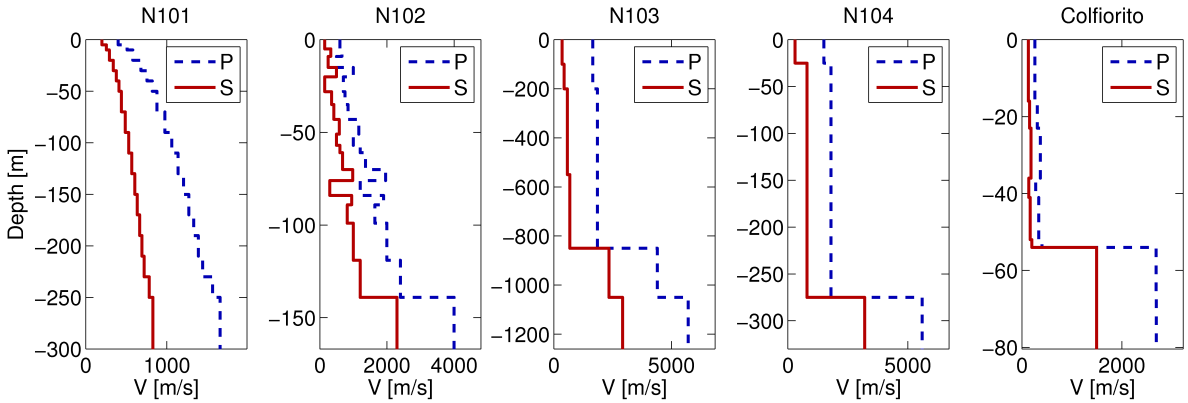


Figure 2.4: S- and P-wave velocity profiles for the models N101, N102, N103 and N104 [Cornou et al., 2009]. The original figure has been modified to improve the readability.

The free parameters of the method are the width of the frequency filter df and the length of the buffered signal Δ.

It could be supposed that Love and body waves influence the results. In an assumed isotropic medium, Love waves are not present on the vertical component and their horizontal movements are thus uncorrelated with the vertical signal. Although incidental correlations of Love and Rayleigh waves arriving at the same time cannot be excluded, such effects will rapidly lose importance because of stacking and can thus be neglected. Arriving P- and S-waves can be present on the vertical and horizontal components, but as their movements are not phase delayed, they will also disappear due to averaging over a sufficiently large number of windows.

2.4.4 Application to synthetic noise

In order to test the method, we used a set of synthetic noise simulated for the models N101, N102, N103 and N104 of the ESG 2006 benchmark [Cornou et al., 2009]. The compressional (P) and shear wave (S) velocity profiles of these models are displayed in figure 2.4.
The noise synthetics were obtained in the following way: Noise sources were approximated by forces of random orientation and amplitude located 0.5 m below the surface [Moczo and Kristek, 2002]. The forces are randomly distributed in time and space. The time function of each source is a delta-like signal with a flat Fourier frequency-amplitude spectrum between 0.1 and 20 Hz. In this frequency range, the associated wave field has been computed by using the wavenumber-based technique for horizontally layered structures proposed by Hisada [1994, 1995]. All types of surface and body waves are included in the simulations.

2.4.4.1 Influence of the parameters of the method

As mentioned above, there are two free parameters that can be adjusted to improve the performance of the method: df, the width of the frequency filter, and Δ, the length of the buffered signal. To investigate the optimal parameters, we used 30 minutes of noise simulated for the model N102. The noise sources are located in the far field as well as in the local field.
Figure 2.5 (a) shows the calculated ellipticity for different bandwidths df of the frequency filter compared to the theoretical ellipticity of the model. The used bandwidths are proportional to the central frequencies. Too weak filtering ($df = 0.5f$) as well as too strong filtering ($df = 0.1f$) do not give satisfying results. An intermediate setting of $0.2f$ reproduces the real curve the best. Therefore, we fixed the bandwidth to $0.2f$ in the following.
Figure 2.5 (b) shows the resulting ellipticity for different lengths Δ of the buffered signal windows in comparison to the theoretical curve. We fix Δ as a function of frequency, so that the signals consist of the same number of periods at all frequencies. As can be seen in the figure, at least ten cycles are necessary for a good coverage of the signal whereas further increasing of the window length does not improve the results. Therefore, $\Delta = 10/f$ is used in the following.

2.4.4.2 Minimum required signal length and temporal stability of the results

The required signal length is illustrated in figure 2.5 (c). The same signals as in subsubsection 2.4.4.1 were used and the first 1800, 600, 300 and 120 seconds of signal were analyzed. The results of the analysis for signals of 1800, 600 and 300 seconds only slightly differ. However, the curve obtained for 120 seconds of signal is less smooth than the other curves. At low frequencies, the signal statistics deteriorate due to the smaller number of buffered windows. It can be deduced that - for sites with resonance frequencies around 1 Hz - a signal length of five minutes can be sufficient to measure an accurate ellipticity curve, as long as enough Rayleigh waves are present in the signal. For smaller underground resonance frequencies, the measurement has to be longer. Indeed, longer measurements improve the statistical features and the reliability of the results at all frequencies.
The temporal stability of the obtained ellipticity curves is shown in figure 2.5 (d). Five different synthetic signals of ten minutes each were simulated and the ellipticity curves were calculated for each window. The overall behavior of the curves is identical. Only the fourth ten-minute-window shows some instabilities between 0.1 and 0.2 Hz emphasizing the need for longer signal lengths at low frequencies. For the other frequencies, the five ellipticity curves are very close. This illustrates the temporal stability of the results.

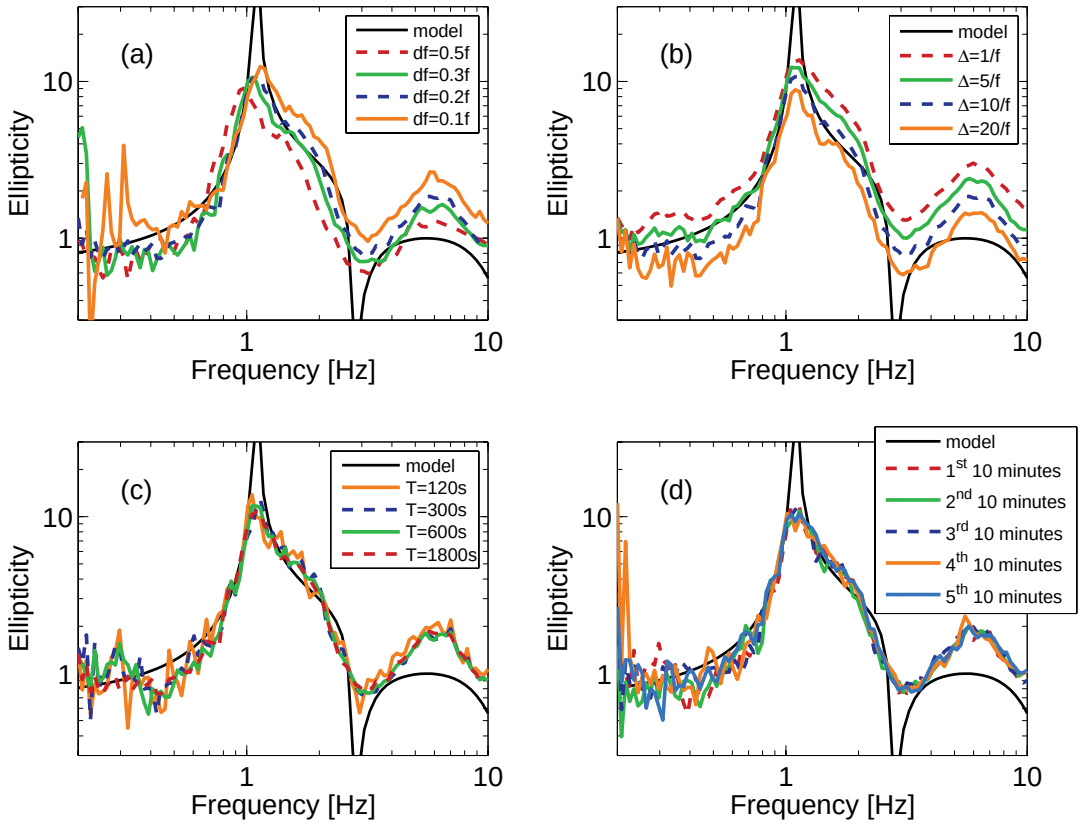


Figure 2.5: Sensitivity of the results with respect to the parameters of the method demonstrated on simulated seimic noise for the N102 model. The real ellipticity curve of the model is included for comparison. Curves (a) and (b) were calculated for 30 minutes of simulated seismic noise. (a) Effect of the filter bandwidth df on the resulting ellipticity curve ($\Delta = 10/f$) for $df = 0.5f$, $df = 0.3f$, $df = 0.2f$ and $df = 0.1f$. (b) Effect of the length Δ of the buffered signal ($df = 0.2f$) for $\Delta = 1/f$, $\Delta = 5/f$, $\Delta = 10/f$ and $\Delta = 20/f$. (c) Effect of the length of the signals (time length T, $df = 0.2f$, $\Delta = 10/f$) for $T = 1800$s, $T = 600$s, $T = 300$s and $T = 120$s. (d) Analysis of five different signals of ten minutes each ($df = 0.2f$, $\Delta = 10/f$). The original figure has been modified to improve the readability.

2.4.4.3 Results for the models N101, N103 and N104

RayDec has been applied to synthetic noise simulated for different models. The results of these simulations are shown in figure 2.6. For model N101 (figure 2.6 (a)), the curve is close to the real curve over the whole frequency range, although the peak ellipticity is overestimated. For this model layout, reflected body waves can deteriorate the results. Therefore, more signal would be needed to obtain a more robust ellipticity curve. For model N103 (figure 2.6 (b)), there is a good agreement between the simulated and the theoretical curves around the resonance frequency and at higher frequencies, with some minor deviations between 0.3 and 2 Hz.

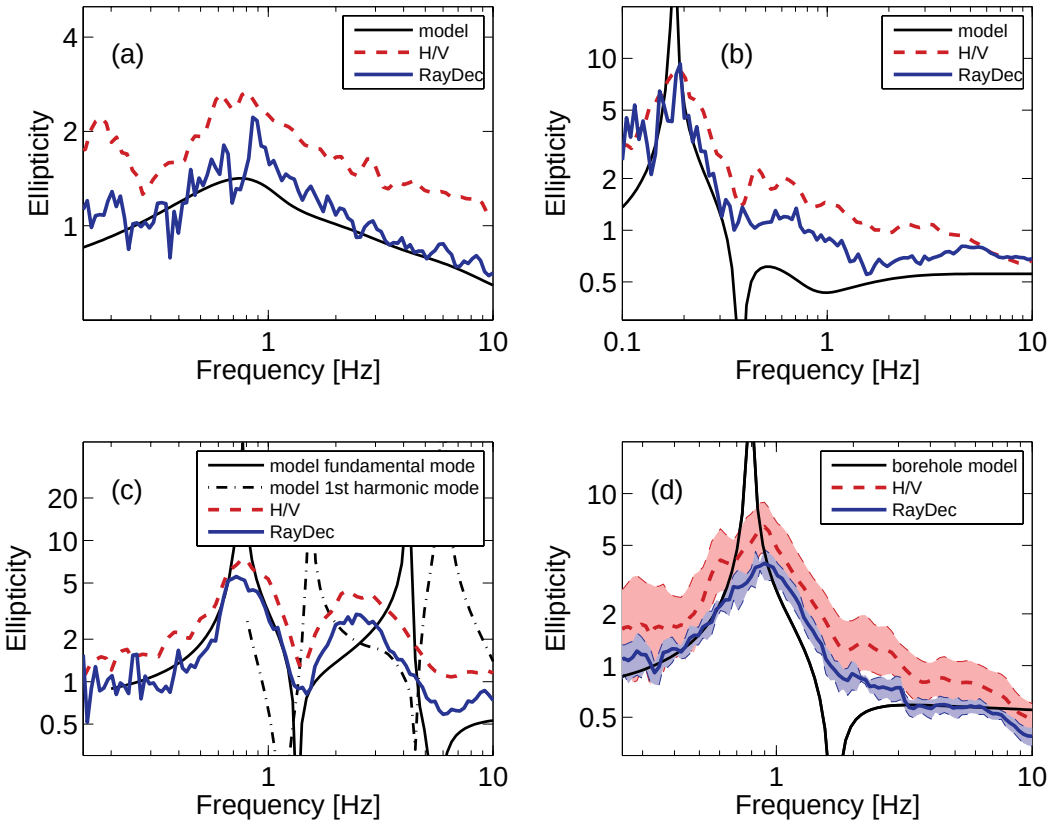


Figure 2.6: Ellipticity curves obtained for ten minutes of synthetic noise simulated for three different underground models (N101(a), N103 (b) and N104 (c)) and for one hour of real noise from Colfiorito basin (d). The used parameters are $df = 0.2f$ *and* $\Delta = 10/f$*. In the four graphs, the results are compared to the real ellipticity curves of the fundamental mode and the first harmonic mode (only in (c)) and to the results obtained by the H/V technique. Curve (d) also contains the error ranges of the H/V and RayDec curves.* The original figure has been modified to improve the readability.

Below 0.15 Hz, the curve differs considerably from the theoretical curve. Model N104 (figure 2.6 (c)) exhibits some unfavorable behavior, as the fundamental mode and the first harmonic mode of Rayleigh wave motion overlap above 1.4 Hz. RayDec cannot distinguish between two modes. Therefore, its results are a mixture of the two modes in this frequency range. For lower frequencies, RayDec reproduces well the theoretical curve around the first peak of the fundamental mode. The H/V curve follows the same trend, but the absolute values of the RayDec curve better reproduce the right flank of the fundamental peak.
It can be stated that, in general, the results of RayDec are much closer to the theoretical curves than the results obtained with the H/V method. However, RayDec regularly fails in reproducing very small ellipticity values. At these points, the horizontal components of the Rayleigh wave vanish.

Therefore, it is problematic to find horizontal signals correlated to the vertical signal. The same reasoning is valid for the peak frequency ellipticity in the case of a theoretically vanishing vertical component. This explains why the peak ellipticity cannot be satisfactorily estimated. On the other hand, small ellipticities can also be hidden by the presence of higher Rayleigh wave modes. In cases where higher Rayleigh wave modes are close to the fundamental modes, RayDec only indicates an apparent ellipticity which cannot be simply related with the true ellipticity of each mode.

2.4.5 Application to real noise data

To test the RayDec method on real data, we applied it to one hour of seismic noise recorded in the center of Colfiorito basin in central Italy. Colfiorito basin is a well-known site which has been extensively investigated by different geophysical methods [Di Giulio et al., 2006]. The theoretical ellipticity curve is calculated on the basis of S- and P-wave velocity profiles obtained by a borehole measurement (see figure 2.4). The parameters used were $df = 0.2f$ and $\Delta = 10/f$. The hour of signal was cut into six ten-minute signals which were analyzed separately. The resulting curves were then averaged. When comparing with the theoretical ellipticity curve, one should keep in mind that a borehole measurement is made at a selected point, whereas a seismic noise record is influenced by the whole surrounding underground structure. The RayDec curve is in better agreement with the theoretical curve than the H/V curve, especially on the slopes of the fundamental resonance frequency peak (figure 2.6 (d)). However, both the H/V and RayDec curve's peak frequencies are shifted with respect to the borehole model's peak frequency.

2.4.6 Conclusion

RayDec is a fast and simple method to determine the ellipticity of Rayleigh waves. As it suppresses Love and body waves efficiently, its results are closer to theory than H/V results. The necessary signal length depends on the resonance frequency of the underground structure, but some minutes of signal can be sufficient to get satisfying results, as long as enough Rayleigh waves are present in the wave field, although longer signal lengths are preferable. However, as it is a one-sensor technique, RayDec gives two possible azimuths and cannot distinguish between prograde and retrograde particle motion. By coupling RayDec with array measurements the correct azimuth and the sense of particle motion could be retrieved. Besides that, it could be envisaged to estimate the proportion of Rayleigh waves in the wave field by comparing RayDec results including only Rayleigh waves with H/V measurements including all wave types. The future work consists of trying to improve the inversion of S-wave velocity profiles by using RayDec's ellipticity estimations. This is a non-trivial endeavor which will be addressed in a future paper.

2.5 Conclusion

One of the main fields of interest of this thesis is the measurement and use of the ellipticity of Rayleigh waves. The rectilinearity and planarity filters indicate which wave type is predominant at a given time and can be used to discriminate against undesired wave types and enhance the desired signal part. However, these methods do not allow the direct retrieval of ellipticity and will not be considered in the following chapters. Nevertheless, they could be used as a preprocessing step for real data applications or for studies on specific seismic phases (P, S, surface waves).

The new methods DELFI and RayDec, as well as H/V, will be tested on synthetic seismic signals in chapter 5. In chapter 6, the potentials and limitations of the inversion of ellipticity curves will be investigated. These inversions will subsequently be applied to real data measurements collected during the NERIES project for 14 European sites in chapter 7. The ellipticity curves used for these inversions will be measured by RayDec. A concise overview of DELFI and RayDec was presented at the European Signal Processing Conference (EUSIPCO) 2009 in Glasgow (UK). The associated paper ("Rayleigh wave ellipticity estimation from ambient seismic noise using single and multiple vector-sensor techniques" by Hobiger, Le Bihan, Cornou, and Bard, 2009b) which includes a comparison with the MUSIQUE technique (see section 4.6) can be found in appendix A.

Chapter 3

Characterization of seismic waves using arrays of sensors

This chapter is dedicated to methods characterizing seismic waves using arrays of seismic sensors. The conventional beamforming method, frequency-wavenumber analysis and MUSIC techniques will be presented before focusing on the SPAC method. For this method, we will investigate the influence of the number of seismic sensors used on the resolution of the method, proofing mathematically that an increasing (odd) number of ring stations is directly linked to a resolution increase. Finally, 2s-SPAC, a method using only two seismic stations, is presented and tested on real data measurements.

Ce chapitre est dédié aux méthodes caractérisant les ondes sismiques à l'aide de réseaux de capteurs sismiques. La formation de voies conventionnelle, l'analyse fréquence-nombre d'onde et la méthode MUSIC seront présentées avant de présenter la méthode SPAC plus en détail. L'influence du nombre de capteurs utilisés sur la résolution de la méthode sera examinée, montrant mathématiquement qu'un nombre croissant (impair) de capteurs sur l'anneau induit directement une augmentation de résolution. Finalement, 2s-SPAC, une méthode n'utilisant que deux capteurs sismiques, sera présentée et testée sur des données sismiques réelles.

3.1 Introduction

Arrays of seismic sensors are an important tool to investigate the soil properties. In fact, a single seismic sensor cannot determine the complete wave vector, i.e. velocity and azimuth, of incident seismic waves. Nevertheless, in chapter 2, it has been shown that the polarization parameters of the wave can be estimated using a single sensor. The complete wave vector can be determined using arrays of seismic sensors. Therefore, two-dimensional arrays of seismic sensors are used in seismology. An example for the layout of such an array is given in figure 3.1 (a). The basic principle of every array method is that an incident wavefront arrives at the different stations at different times[I]. From the arrival time delays of the signals recorded at the respective stations, the wave velocity v (or its inverse, the slowness s) and the azimuth can be retrieved. It is clear that a one-dimensional seismic array (i.e. a line of sensors) is not able to retrieve azimuth or velocity, because it can only determine the component of the wave vector which lies in the array direction[II]. Azimuth and velocity can only be found when both horizontal components of the wave vector are resolved.

Furthermore, a two-dimensional seismic array can only find the "apparent" velocity of the wave. If a plane wave is arriving under an angle of incidence γ (see figure 3.1 (b)) from the vertical axis, the apparent velocity is given by $v_a = v/\sin\gamma$, where v is the actual wave velocity (the apparent slowness is given by $s_a = s \cdot \sin\gamma$, where s is the actual slowness). Notably, for an angle of incidence of $0°$, the apparent velocity is infinite and the wave arrives at the same time at all stations. In order to retrieve the actual velocity, azimuth and angle of incidence of a wave, a three-dimensional array would need to be used. This requires sensors in boreholes with depths comparable to the deployment size at the surface. Thus, apart from being quite expensive, such measurements are time-consuming.

However, only body waves as S- and P-waves can arrive under an angle of incidence different from $90°$. Surface waves, like Love and Rayleigh waves, are travelling horizontally. Therefore, surface waves can be analyzed using two-dimensional seismic arrays. In the following, some well-known investigation methods using seismic arrays will be presented in detail. Conventional beam-forming (section 3.2) is the simplest and most intuitive of these methods. The azimuth and velocity of an incident wave is determined by delaying the signals of the different sensors according to different azimuth and velocity values and identifying those values maximizing the total energy. The frequency-wavenumber method [Lacoss et al., 1969, Capon, 1969] presented in section 3.3 can be seen as the transformation of the conventional beam-forming method from the time domain in the frequency domain. The MUSIC method [Schmidt, 1986] shown in section 3.4 is another algorithm which allows the estimation of the number of signals and separates

[I] In general, a seismic wave radiating from a point source has a curved wave front. If the distance between a seismic array and the source is large compared to the array size, however, the wave front is almost plane.

[II] If either the azimuth or the wave velocity are known, the other parameter can be determined by a one-dimensional array. However, the azimuth estimation for a known velocity is still ambiguous.

signal from noise space. The advancement of this code by including quaternions and biquaternions will be the topic of chapter 4. All these methods allow the determination of wave velocity and azimuth of incoming waves.

A different approach to measure the wave velocity using ambient seismic vibrations is the spatial autocorrelation (SPAC) technique [Aki, 1957], which will be presented in more detail in section 3.5. This technique measures the wave velocity by averaging the correlations between stations located on a ring and the central station. After introducing the methodology, we will investigate the theoretical resolution of the technique for arrays of different sizes and show a mathematical proof for the resolution increase caused by an increase of the number of stations on the ring. The M-SPAC approach [Bettig et al., 2001] facilitates SPAC measurements as it suppresses the need of equidistant station spacing of the original method. Finally, the 2s-SPAC method [Morikawa et al., 2004] will be presented, which substitutes the spatial averaging over many station pairs by a temporal averaging of a single moving station pair. We will apply this method to measurements performed in Patras, Greece, and show its usefulness for real data measurements in section 3.5.7.

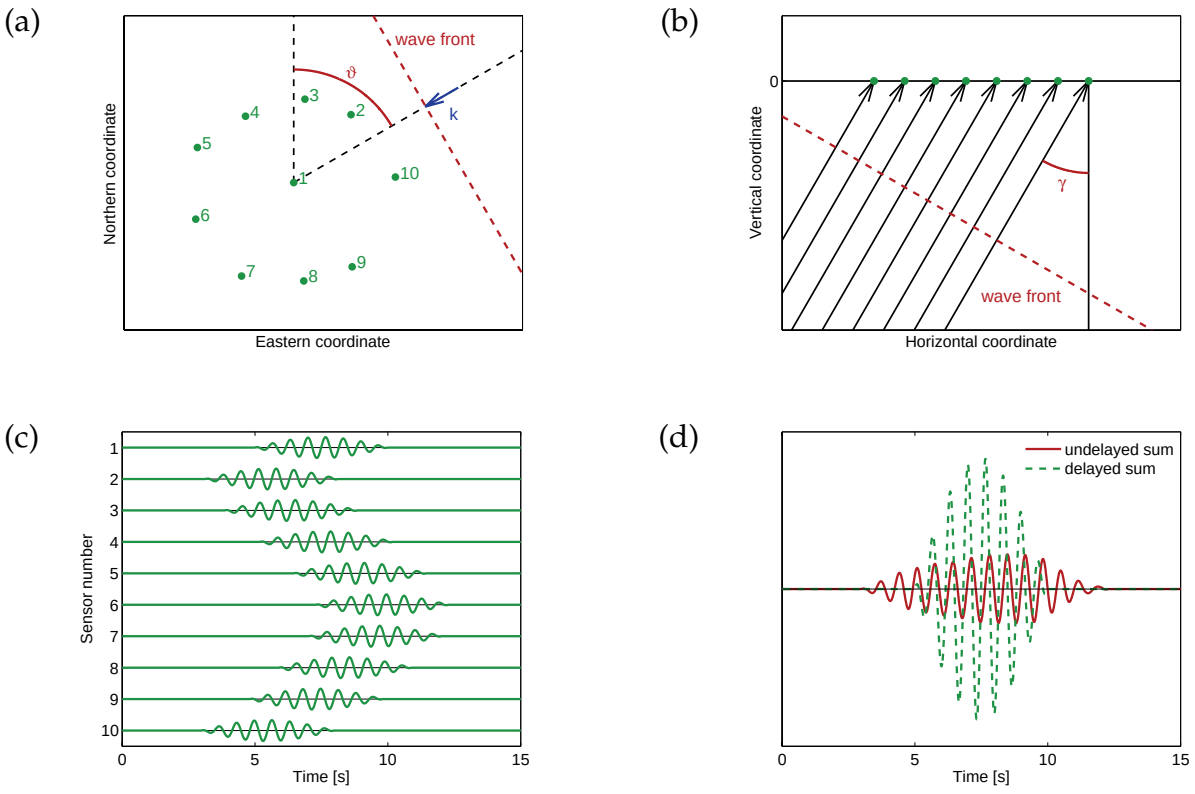


Figure 3.1: (a) Diagram of a seismic array with incident plane wave arriving under an azimuth angle ϑ. (b) If the wave is arriving with an angle of incidence different from 90°, the array will detect an apparent velocity $v_a = v/\sin\gamma$. (c) Theoretical signals of the different sensors for the incident wave in (a). (d) Undelayed sum und delayed sum of the signals from (c).

3.2 Delay-and-sum beamforming

The simplest and probably the most intuitive way of treating recordings of seismic arrays is delay-and-sum beamforming [Christensen and Hald, 2004]. The basic idea is to bring the recordings of the different seismic stations into phase by compensating the time delays induced by the respective travel time differences.
An incident wave will arrive at the different stations of the array at different times, as can be seen in figure 3.1 (c), where the signals of the different stations due to the incident wave in figure 3.1 (a) are shown. Let the slowness vector of the incident wave be given by

$$\vec{s}(f) = s(f) \cdot \begin{pmatrix} -\sin\vartheta \sin\gamma \\ -\cos\vartheta \sin\gamma \\ \cos\gamma \end{pmatrix}, \tag{3.1}$$

where $s(f) = |\vec{s}(f)|$ is the absolute value of the slowness vector, ϑ the azimuth and γ the angle of incidence of the wave. Defining the matrix $\mathbf{R} \in \mathbb{R}^{N\times 3}$ as the position matrix of the N seismic stations of the array, the delay vector $\vec{\Delta}(\vec{s}(f)) = \mathbf{R}\,\vec{s}(f)$ indicates the delays corresponding to the different seismic stations with respect to the origin (not necessarily a sensor position).
The recorded signals $b_i(t)$ ($i \in \{1, 2, \ldots, N\}$) of the seismic sensors are filtered around the frequency f, which results in band-pass filtered signals $b_{f,i}(t)$. They include a signal part $s_{f,i}(t)$ and a noise part $n_{f,i}(t)$:

$$b_{f,i}(t) = s_{f,i}(t) + n_{f,i}(t). \tag{3.2}$$

In order to identify the azimuth and slowness of an incident wave, a grid search can be performed. For each pair of slowness s and azimuth ϑ, the corresponding delay vector is calculated, the signals of the different sensors are delayed accordingly and summed to produce the final signal $b_{\vec{s}}(t)$[III]:

$$\begin{aligned} b_{f,\vec{s}}(t) &= \frac{1}{N}\sum_{i=1}^{N} w_{f,i}\, b_{f,i}(t + \Delta_i(\vec{s}(f))), \\ &= \frac{1}{N}\sum_{i=1}^{N} w_i\, s_{f,i}(t + \Delta_i(\vec{s}(f))) + \frac{1}{N}\sum_{i=1}^{N} w_i\, n_{f,i}(t + \Delta_i(\vec{s}(f))). \end{aligned} \tag{3.3}$$

Here, the w_i are optional weighting factors. The correct slowness vector can now be identified by calculating the energy content for the signals $b_{f,\vec{s}}(t)$ by

$$E_f(\vec{s}) = \sum_{l=1}^{L} b^2_{f,\vec{s}}(t_l)dt, \tag{3.4}$$

where $l \in 1, \ldots, L$ parameterizes the time samples. The $\vec{s}(f)$ corresponding to the maximum of $E_f(\vec{s})$ characterizes the incident wave. The associated $b_{f,\vec{s}}(t)$ can be seen as an enhanced version of the incident signal[IV].

[III] If the delays of the different sensors can be identified precisely, $\vec{s}$ can be calculated directly by $\mathbf{R}^{-1}\vec{\Delta}$.
[IV] As the signal parts in equation (3.3) superpose constructively whereas the noise parts do not, the signal can be improved by using a large number of seismic sensors. The signal-to-noise ratio increases as $\sqrt{N}$.

In figure 3.1 (d), the result of a direct summation of the signals at the different sensors is compared to the result of the delayed sum of signals. The direct summation is not representing the actual signal.
For simple cases, where a single wave is incident, the delay-and-sum beamforming can yield satisfactory results. Nonetheless, if multiple waves interfere, the resolution of this simple method may not be sufficient. Therefore, more sophisticated methods exist.

3.3 Frequency-wavenumber analysis

The frequency-wavenumber analysis [Lacoss et al., 1969] can be seen as beamforming in the frequency domain. If we consider a single incident plane wave whose slowness vector $\vec{s}_0(f)$ is given by equation (3.1) arriving at an array of N seismic sensors (with respective sensor positions $\vec{r}_i$, $i \in \{1, 2, \dots N\}$). The first sensor's position is the origin of the coordinate system. Then, the respective (single-component) recordings of the different stations can be written as

$$x_i(t) = \int_{-\infty}^{\infty} C(f) e^{i2\pi f(t - \vec{s}_0(f)\ \vec{r}_i)} df, \tag{3.5}$$

where $C(f)$ is the frequency-dependent complex amplitude of the wave. It is possible to define the array signal vector $\vec{x}(t)$ as

$$\vec{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{pmatrix} = \int_{-\infty}^{\infty} C(f) e^{i2\pi f(t - \mathbf{R}\ \vec{s}_0(f))} df, \tag{3.6}$$

where $\mathbf{R}$ is the matrix of sensor locations (of size $N \times 3$). The Fourier transform of the signal vector yields

$$\begin{aligned}
\vec{X}(f) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(f') e^{i2\pi f'(t - \mathbf{R}\ \vec{s}_0(f'))} e^{-i2\pi f t} df' dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} C(f') e^{-i2\pi f' \mathbf{R}\ \vec{s}_0(f')} \int_{-\infty}^{\infty} e^{i2\pi(f' - f)t} dt df' \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} C(f') e^{-i2\pi f' \mathbf{R}\ \vec{s}_0(f')} \delta(f' - f) df' \\
&= \frac{1}{\sqrt{2\pi}} C(f) e^{-i2\pi f \mathbf{R}\ \vec{s}_0(f)} \\
&= S_0(f) e^{-i\mathbf{R}\ \vec{k}_0(f)},
\end{aligned} \tag{3.7}$$

where $S_0(f) = \frac{C(f)}{\sqrt{2\pi}}$ is the Fourier transform of the signal recorded at the station located at the origin and $\vec{k}_0(f) = 2\pi f \vec{s}_0(f)$ the wave vector.

A steering vector can be defined by $\vec{a}(\vec{k}) = \frac{1}{\sqrt{N}} e^{-i\mathbf{R}\vec{k}}$. Multiplying $\vec{X}(f)$ by $\vec{a}^{\dagger}(\vec{k}_0)$ brings all the components of $\vec{X}(f)$ in phase and sums them. The energy of $\vec{X}(f)\vec{a}^{\dagger}(\vec{k})$ is calculated by

$$\begin{aligned} E_f(\vec{k}) &= \left| \vec{X}^{\dagger}(f)\vec{a}(\vec{k}) \right|^2 \\ &= \left| S_0(f) e^{i(\mathbf{R}\ \vec{k}_0(f))^T} \frac{1}{\sqrt{N}} e^{-i\mathbf{R}\vec{k}} \right|^2 \\ &= \left| S_0(f) \frac{1}{N} \sum_{i=1}^{N} e^{-i\vec{r}_i\ (\vec{k}-\vec{k}_0(f))} \right|^2 \\ &= |S_0(f)|^2 \left| \frac{1}{N} \sum_{i=1}^{N} e^{-i\vec{r}_i\ (\vec{k}-\vec{k}_0(f))} \right|^2 \leq |S_0(f)|^2 = E_f(\vec{k}_0). \end{aligned} \tag{3.8}$$

The correct wave vector $\vec{k}_0$ can therefore be found by performing a grid search over all possible values of $\vec{k}$ and identifying the maximum of $E_f(\vec{k})$. The theoretical array response is given by

$$A(\vec{k}, f) = \left| \frac{1}{N} \sum_{i=1}^{N} e^{-i\vec{r}_i\ \vec{k}} \right|^2, \tag{3.9}$$

which corresponds to $E_f(\vec{k}) / |S_0(f)|^2$ for an incident wave whose wave vector is the zero vector, corresponding to infinite velocity (or a wave arriving under an angle of incidence of 0°).

3.3.1 Resolution limits of seismic arrays

A seismic array can only resolve seismic waves in a region of the frequency-slowness plane which is defined by the array layout. For example, an equidistant linear sensor antenna cannot resolve waves whose wavelengths are inferior to twice the minimum sensor distance d_{min} because of aliasing, which would correspond to a maximum wave number of $k_{max} = \frac{2\pi}{2d_{min}}$. A rough estimation of the lower resolution limit can be given by $k_{min} = \frac{2\pi}{d_{max}}$, where d_{max} is the maximum inter-sensor distance in the array. However, for two-dimensional array layouts, the resolution limits are different for different azimuths. Using the array response function defined in equation (3.9), the resolution limits of the array can be investigated for each azimuth. The question of array responses of two-dimensional arrays has been largely discussed in the literature, for example by Woods and Lintz [1973], Asten and Henstridge [1984], Tokimatsu [1997], Gaffet et al. [1998] and Wathelet et al. [2008].

3.4 MUSIC

The MUSIC method was originally developed by Schmidt [1986]. The name MUSIC stands for "MUltiple SIgnal Classification". This algorithm allows the estimation of the number of signals and their associated directions of arrival for data recorded by an array of sensors. The basic idea behind the MUSIC algorithm is to separate the noise and signal subspaces and subsequently estimate the signal parameters. The method is explained in the following.
An array of N monocomponent seismic sensors records a time series of seismic data. K different uncorrelated seismic waves impinge. Each wave is characterized by its wave vector $\vec{k}_i = -2\pi f s_i(f) \begin{pmatrix} \sin\vartheta_i \\ \cos\vartheta_i \\ 0 \end{pmatrix}$, where $s_i(f)$ is the slowness and ϑ_i the backazimuth of the wave. Therefore, the signal $\vec{X}(f)$ recorded at the different seismic stations can be written as a linear combination of the signals of the different sources plus noise. In particular, this noise is not seismic noise, which would still carry information, but instrumental or other noise. Thus, it is assumed that the noise contributions at the different sensors are uncorrelated. An additional assumption is that the noise is uncorrelated to the true signal. This yields the following expression for the signal:

$$\begin{pmatrix} X_1(f) \\ X_2(f) \\ \vdots \\ X_N(f) \end{pmatrix} = \left(\vec{a}(\vec{k}_1),\ \ \vec{a}(\vec{k}_2),\ \ \ldots,\ \ \vec{a}(\vec{k}_K)\right) \cdot \begin{pmatrix} F_1(f) \\ F_2(f) \\ \vdots \\ F_K(f) \end{pmatrix} + \begin{pmatrix} W_1(f) \\ W_2(f) \\ \vdots \\ W_N(f) \end{pmatrix}, \tag{3.10}$$

or in matrix form

$$\vec{X}(f) = \mathbf{A}(f) \cdot \vec{F}(f) + \vec{W}(f). \tag{3.11}$$

The signals are represented in the frequency-phase domain, corresponding to a Fourier transform of the signals in time-space domain. The complex vector $\vec{F} \in \mathbb{C}^{K\times 1}$ indicates the amplitude and phase (in reference to a given point) of the source signals. The vectors $\vec{a}(\vec{k})$ in matrix $\mathbf{A} \in \mathbb{C}^{N\times K}$ describe the phase differences induced by the propagation between each source (defined by its wave vector $\vec{k} \in \mathbb{R}^{3\times 1}$) and each sensor. If $\mathbf{R} \in \mathbb{R}^{N\times 3}$ is the matrix of sensor positions,

$$\vec{a}(\vec{k}) = \frac{1}{\sqrt{N}} \exp(-i\mathbf{R} \cdot \vec{k}). \tag{3.12}$$

The vector $\vec{W} \in \mathbb{C}^{N\times 1}$ characterizes the noise at each sensor. In this way, the signal $\vec{X} \in \mathbb{C}^{N\times 1}$ is a linear combination of source signals and noise terms. In the following, the frequency arguments will be omitted, but it should not be forgotten that we work at a given frequency.
It is now possible to build a $N \times N$ covariance matrix $\mathbf{S}$ by

$$\mathbf{S} = \mathsf{E}\left(\vec{X} \cdot \vec{X}^{\dagger}\right) = \mathbf{A} \cdot \mathsf{E}(\vec{F}\vec{F}^{\dagger})\mathbf{A}^{\dagger} + \mathsf{E}(\vec{W}\vec{W}^{\dagger}), \tag{3.13}$$

where $\mathsf{E}(x)$ indicates the expectation value and the non-correlation of $\vec{F}$ and $\vec{W}$ is used. The correlation matrix simplifies to

$$\mathbf{S} = \mathbf{APA}^\dagger + \lambda \mathbf{W}_0, \tag{3.14}$$

where $\mathbf{P} = \mathsf{E}(\vec{F}\vec{F}^\dagger)$ and $\mathbf{W}_0$ is the noise covariance matrix. We assume that the elements of the noise vector W are distributed according to a normal distribution with zero mean and variance σ^2, which yields $\lambda \mathbf{W}_0 = \sigma^2 \mathbf{I}$. Therefore,

$$\mathbf{S} = \mathbf{APA}^\dagger + \sigma^2 \mathbf{I}. \tag{3.15}$$

In general, the number of seismic sensors N is larger than the number of incident waves K. Therefore, $\mathbf{APA}^\dagger$ is singular with a rank of K. The eigenvalue σ^2 of $\mathbf{S}$ will occur $(N-K)$ times. $\mathbf{S}$ has N eigenvectors $\vec{u}_i$ $(1 \leq i \leq N)$ with associated eigenvalues λ_i. We can write

$$\mathbf{S}\vec{u}_i = \mathbf{APA}^\dagger \vec{u}_i + \sigma^2 \mathbf{I}\vec{u}_i \tag{3.16}$$

$$\lambda_i \vec{u}_i = \mathbf{APA}^\dagger \vec{u}_i + \sigma^2 \vec{u}_i \tag{3.17}$$

$$\mathbf{APA}^\dagger \vec{u}_i = (\lambda_i - \sigma^2)\vec{u}_i. \tag{3.18}$$

As the eigenvalue σ^2 of $\mathbf{S}$ occurs $(N-K)$ times, the corresponding eigenvectors $\vec{u}_i$ (with $1 \leq i \leq N-K$) are orthogonal to $\mathbf{APA}^\dagger$ and to the space spanned by the column vectors of $\mathbf{A}$.
In this way, these $(N-K)$ eigenvectors span a noise subspace. The other K eigenvectors span the signal subspace. In a real data application, the number of sources is, in general, not known. Then, the number of signals can be estimated by considering the eigenvalues: the largest eigenvalues belong to the signal subspace and the remaining eigenvalues to the noise subspace.
We build the noise subspace $\mathbf{G} \in \mathbb{C}^{\mathbf{N}\times\mathbf{N}-\mathbf{K}}$ by

$$\mathbf{G} = \left(\vec{u}_1,\ \vec{u}_2,\ \ldots,\ \vec{u}_{N-K}\right). \tag{3.19}$$

For a given $\vec{a}(\vec{k})$, the squared distance d^2 to the signal subspace is calculated by

$$\begin{aligned} d^2 &= \|\mathbf{G}^\dagger\ \vec{a}(\vec{k})\|^2 \\ &= \left(\mathbf{G}^\dagger\ \vec{a}(\vec{k})\right)^\dagger \mathbf{G}^\dagger\ \vec{a}(\vec{k}) \\ &= \vec{a}^\dagger(\vec{k})\ \mathbf{G}\mathbf{G}^\dagger\ \vec{a}(\vec{k}), \end{aligned} \tag{3.20}$$

as the vectors spanning the noise subspace are orthogonal to the signal subspace. The $\vec{a}(\vec{k})$ which maximizes the MUSIC functional

$$P(\vec{k}) = \frac{1}{d^2} = \frac{1}{\vec{a}^\dagger(\vec{k})\ \mathbf{G}\mathbf{G}^\dagger\ \vec{a}(\vec{k})} \tag{3.21}$$

is in a certain way the most orthogonal to the noise subspace and therefore the most confined in the signal subspace. In this way the properties of the incident waves, i.e. azimuth and slowness, can be retrieved by a two-dimensional grid search.

3.5 Spatial autocorrelation technique (SPAC)

The spatial autocorrelation technique (SPAC) has been developed by Aki [1957][V]. As we use this method extensively for the analysis of real seismic noise data, its methodology will be presented in more detail. In contrast to the original derivation of Aki [1957], it is not necessary to assume that different waves are uncorrelated, as will be shown in the following.

3.5.1 Vertical component

First, we consider the vertical motion only. The motion of a seismic wave at a given frequency f can be written as a superposition of a number of seismic waves with different wave vectors $\vec{k}_n$ by

$$u(\vec{r}, f, t) = \sum_n A_n(f) \cos(2\pi f t - \vec{k}_n(f)\vec{r} + \varphi_n), \tag{3.22}$$

where φ_n is the phase of the wave and $\vec{r}$ the position of an arbitrary sensor. $k(f) = \left|\vec{k}_n(f)\right|$ is the absolute value of the wave vector. Wave vector and velocity at frequency f are linked by $k(f) = 2\pi f / v(f) = 2\pi f \, s(f)$. The correlation between two sensors located at $\vec{r}_1$ and $\vec{r}_2$ can then be written as

$$\begin{aligned}
\phi(\vec{r}_1, \vec{r}_2, f) =& \frac{1}{T} \int_0^T u(\vec{r}_1, f, t) \; u(\vec{r}_2, f, t) dt \\
\phi(\vec{r}_1, \vec{r}_2, f) =& \frac{1}{T} \sum_n \sum_m A_n \; (f) A_m(f) \\
& \cdot \int_0^T \cos(2\pi f t - \vec{k}_n(f) \; \vec{r}_1 + \varphi_n) \cos(2\pi f t - \vec{k}_m(f) \; \vec{r}_2 + \varphi_m) dt,
\end{aligned} \tag{3.23}$$

where T has to be a multiple of $1/f$[VI]. The original derivation of Aki [1957] assumed that different wave components are uncorrelated, i.e. $A_n \bar{A}_m = 0$ for $n \neq m$, where the bar represents the averaging operation. In fact, this property is not necessary to derive the SPAC equations as will be shown in the following.

[V] The name is misleading, because indeed cross correlations between different sensors are calculated.

[VI] Alternatively, T has to be so large that $\frac{1}{T} \int_0^T \cos(2\pi f t) dt \to 0$.

By substituting $a = 2\pi f t$, $b = \vec{k}_n(f)\,\vec{r}_1 - \varphi_n$ and $c = \vec{k}_m(f)\,\vec{r}_2 - \varphi_m$, the integral can be calculated as

$$\begin{aligned}
\int_0^{2\pi} \cos(a-b)\cos(a-c)da &= \int_0^{2\pi} (\cos a\cos b + \sin a \sin b)(\cos a \cos c + \sin a \sin c)da \\
&= \int_0^{2\pi} \cos^2 a \cos b \cos c \; da + \int_0^{2\pi} \sin^2 a \sin b \sin c \; da \\
&\quad + \int_0^{2\pi} \sin a \cos a (\sin b \cos c + \cos b \sin c) \; da \\
&= \pi(\cos b \cos c + \sin b \sin c) \\
&= \pi \cos(b-c). \qquad (3.24)
\end{aligned}$$

By resubstituting a, b and c, equation (3.23) simplifies to

$$\phi(\vec{r}_1, \vec{r}_2, f) = \frac{1}{2} \sum_n \sum_m A_n(f)\; A_m(f) \cos(\vec{k}_n(f)\vec{r}_1 - \vec{k}_m(f)\vec{r}_2 - \varphi_n + \varphi_m). \qquad (3.25)$$

Without restricting the universal validity of this formula, we can set the first sensor on the origin of the coordinate system and put a ring of seismic sensors at a radius r around the first sensor. Then, the inner product of $\vec{k}_m(f)$ and $\vec{r}_2$ can be written as $\vec{k}_m(f)\,\vec{r}_2 = -k(f) r \cos(\vartheta_m - \gamma)$, where ϑ_m is the azimuth of the mth incoming wave and γ the angle under which the sensor at $\vec{r}_2$ is seen from the origin. The mean correlation between the central sensor and all sensors on the ring can then be written as

$$\begin{aligned}
\overline{\phi(r,f)} = &\quad \frac{1}{2\pi} \int_0^{2\pi} \phi(0, \vec{r}_2 = r \cdot (\sin\gamma, \cos\gamma)^T, f) d\gamma \\
= &\quad \frac{1}{4\pi} \sum_n \sum_m A_n(f) A_m(f) \int_0^{2\pi} \cos(k(f) r \cos(\vartheta_m - \gamma) - \varphi_n + \varphi_m) d\gamma \\
= &\quad \frac{1}{4\pi} \sum_n \sum_m A_n(f) A_m(f) \int_0^{2\pi} \cos(k(f) r \cos\gamma) \cos(\varphi_n - \varphi_m) d\gamma \\
&+ \frac{1}{4\pi} \sum_n \sum_m A_n(f) A_m(f) \int_0^{2\pi} \sin(k(f) r \cos\gamma) \sin(\varphi_n - \varphi_m) d\gamma. \\
= &\quad \frac{1}{4\pi} \sum_n \sum_m A_n(f) A_m(f) \int_0^{2\pi} \cos(k(f) r \cos\gamma) \cos(\varphi_n - \varphi_m) d\gamma. \qquad (3.26)
\end{aligned}$$

By using $\int_0^{2\pi} \cos(x\cos\gamma)d\gamma = 2\pi J_0(x)$, where $J_0(x)$ indicates the Bessel function of the first kind of order 0[VII], this equation can be simplified to

$$\overline{\phi(r,f)} = \frac{1}{2}\sum_n \sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m)\cdot J_0(k(f)r). \tag{3.27}$$

The graph of $J_0(x)$ is plotted in figure 3.2. The autocorrelation of the central station is given by setting $\vec{r}_2 = \vec{r}_1$ (i.e. $r = 0$):

$$\phi(0,f) = \frac{1}{2}\sum_n \sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m). \tag{3.28}$$

Finally, dividing equation (3.27) by equation (3.28) yields

$$\rho_{vertical}(r,f) = \frac{\overline{\phi(r,f)}}{\phi(0,f)} = J_0\left(2\pi f r\, s(f)\right). \tag{3.29}$$

The meaning of this equation can be described as follows: If one disposes of a single seismic sensor surrounded by a large number of sensors, then the correlation functions between the central station and each ring sensor have to be calculated. Averaging all these correlation functions and normalizing the result by the autocorrelation of the central station yields a curve which can be identified as $J_0\left(2\pi f r\, s(f)\right)$. In this way, a direct link between the frequency and the slowness of the wave, i.e. its dispersion curve, is given.
In the derivation of this function, it was integrated over γ, i.e. in such way that the pairs of correlated seismic sensors cover all possible azimuths. Replacing the integration over γ by an integration over ϑ in the formula would not change the result. This would correspond to a single pair of seismic sensors and incoming waves covering all azimuths. This means that if the wave field was perfectly isotropic, i.e. if waves arrived from all directions with identical energies, a single pair of seismic sensors would suffice to measure the correct Bessel function.

[VII] The Bessel functions of the first kind $J_n(x)$ are solutions of the equation $x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - n^2)y = 0$:

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k(\frac{x}{2})^{2k}}{k!^2}, \qquad J_1(x) = \sum_{k=0}^{\infty} \frac{(-1)^k(\frac{x}{2})^{2k+1}}{(k+1)!\,k!}, \qquad J_2(x) = \sum_{k=0}^{\infty} \frac{(-1)^k(\frac{x}{2})^{2k+2}}{(k+2)!\,k!}.$$

3.5.2 Horizontal components

The derivation described above is only valid for the vertical component. For both horizontal components of the seismic sensors, the reasoning is slightly different because horizontal waves can be polarized either in the transverse or in the radial direction. In general, on both horizontal components of a seismic sensor both polarizations are recorded simultaneously.
Considering two seismic sensors, the radial component is oriented in the direction connecting both sensors. The transverse component is oriented in the perpendicular direction.
Let us assume in a first step that the horizontal wave field is composed exclusively of waves with radial polarization which arrive under respective azimuth angles ϑ_n. If the second seismic sensor is seen under an angle γ from position of the first sensor, i.e. the angle γ is defining the radial component of their respective recordings. Then the radial component recording of any of both seismic sensors is given by

$$u_r(\vec{r},\gamma,f,t) = \sum_n A_n(f)\cos(2\pi f t - \vec{k}_n(f)\vec{r} + \varphi_n)\cos(\vartheta_n - \gamma), \tag{3.30}$$

whereas its transverse recording is given by

$$u_t(\vec{r},\gamma,f,t) = \sum_n A_n(f)\cos(2\pi f t - \vec{k}_n(f)\vec{r} + \varphi_n)\sin(\vartheta_n - \gamma). \tag{3.31}$$

The calculation of the correlation between the two sensors is analog to equations (3.23) - (3.25) with additional sinus or cosinus terms and will therefore not be shown in detail. The equation analog to equation (3.26) for the radial component is then given by

$$\begin{aligned}\overline{\phi_r(r,f)} &= \frac{1}{4\pi}\sum_n\sum_m A_n(f)A_m(f) \\ &\quad \cdot \int_0^{2\pi} \cos(k(f)r\cos(\vartheta_m - \gamma) - \varphi_n + \varphi_m)\cos(\vartheta_n - \gamma)\cos(\vartheta_m - \gamma)\, d\gamma \\ &= \frac{1}{4\pi}\sum_n\sum_m A_n(f)A_m(f) \\ &\quad \cdot \int_0^{2\pi} \cos(k(f)r\cos\gamma - \varphi_n + \varphi_m)\cos(\gamma + \vartheta_m - \vartheta_n)\cos\gamma\, d\gamma \end{aligned} \tag{3.32}$$

and for the transverse component by

$$\begin{aligned}\overline{\phi_t(r,f)} &= \frac{1}{4\pi}\sum_n\sum_m A_n(f)\ A_m(f) \\ &\quad \cdot \int_0^{2\pi} \cos(k(f)r\cos\gamma - \varphi_n + \varphi_m)\sin(\gamma + \vartheta_m - \vartheta_n)\sin\gamma\, d\gamma. \end{aligned} \tag{3.33}$$

The integral in equation (3.32) can be calculated by

$$\begin{aligned}
&\int_0^{2\pi} \cos(k(f)r\cos\gamma - \varphi_n + \varphi_m)\cos(\gamma + \vartheta_m - \vartheta_n)\cos\gamma\, d\gamma \\
&= \quad \cos(\varphi_n - \varphi_m)\cos(\vartheta_m - \vartheta_n)\int_0^{2\pi} \cos(k(f)r\cos\gamma)\cos^2\gamma\, d\gamma \\
&\quad + \sin(\varphi_n - \varphi_m)\cos(\vartheta_m - \vartheta_n)\int_0^{2\pi} \sin(k(f)r\cos\gamma)\cos^2\gamma\, d\gamma \\
&\quad - \cos(\varphi_n - \varphi_m)\sin(\vartheta_m - \vartheta_n)\int_0^{2\pi} \cos(k(f)r\cos\gamma)\sin\gamma\cos\gamma\, d\gamma \\
&\quad - \sin(\varphi_n - \varphi_m)\sin(\vartheta_m - \vartheta_n)\int_0^{2\pi} \sin(k(f)r\cos\gamma)\sin\gamma\cos\gamma\, d\gamma.
\end{aligned}$$

In this equation, $\int_0^{2\pi} \cos(k(f)r\cos\gamma)\cos^2\gamma\, d\gamma = \pi(J_0(k(f)r) - J_2(k(f)r))$ and all other integrals vanish. The graph of $J_2(x)$ is shown in figure 3.2. Now, the correlation for the radial component can be expressed by

$$\overline{\phi_r(r,f)} = \frac{1}{4}\sum_n\sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m)\cos(\vartheta_m - \vartheta_n)(J_0(k(f)r) - J_2(k(f)r)). \tag{3.34}$$

For the transverse component, the calculation is completely analog and yields[VIII]

$$\overline{\phi_t(r,f)} = \frac{1}{4}\sum_n\sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m)\cos(\vartheta_m - \vartheta_n)(J_0(k(f)r) + J_2(k(f)r)). \tag{3.35}$$

The radial and transverse autocorrelations for the central station are calculated by

$$\phi_r(0,f) = \frac{1}{4}\sum_n\sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m)\cos(\vartheta_m - \vartheta_n), \tag{3.36}$$

$$\phi_t(0,f) = \frac{1}{4}\sum_n\sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m)\cos(\vartheta_m - \vartheta_n). \tag{3.37}$$

[VIII] Using $\int_0^{2\pi} \cos(k(f)r\cos\gamma)\sin^2\gamma\, d\gamma = \pi(J_0(k(f)r) + J_2(k(f)r))$.

This finally generates the normalized correlation functions $\rho_r(r,f)$ and $\rho_t(r,f)$ for radially polarized waves:

$$\rho_{radial,r}(r,f) = \frac{\overline{\phi_r(r,f)}}{\phi_r(0,f)} = J_0\left(2\pi f r\, s(f)\right) - J_2\left(2\pi f r\, s(f)\right), \tag{3.38}$$

$$\rho_{radial,t}(r,f) = \frac{\overline{\phi_t(r,f)}}{\phi_t(0,f)} = J_0\left(2\pi f r\, s(f)\right) + J_2\left(2\pi f r\, s(f)\right). \tag{3.39}$$

The reasoning for transversely polarized waves is completely analogue. In this case, however, the radial component's recording is given by

$$u_r(\vec{r},\gamma,f,t) = \sum_n A_n(f)\cos(2\pi f t - \vec{k}_n(f)\vec{r} + \varphi_n)\sin(\vartheta_n - \gamma), \tag{3.40}$$

whereas its transverse recording is given by

$$u_t(\vec{r},\gamma,f,t) = \sum_n A_n(f)\cos(2\pi f t - \vec{k}_n(f)\vec{r} + \varphi_n)\cos(\vartheta_n - \gamma). \tag{3.41}$$

This means that the equation for the radial component in this case is the same as for the transverse component in the case of radially polarized waves (equation (3.31)) and that the transverse component's record is described as the radial component in equation (3.30). Therefore, we can directly give the resulting autocorrelation functions $\rho_r(r,f)$ and $\rho_t(r,f)$ for transverse wave polarization:

$$\rho_{transverse,r}(r,f) = J_0\left(2\pi f r\, s(f)\right) + J_2\left(2\pi f r\, s(f)\right), \tag{3.42}$$

$$\rho_{transverse,t}(r,f) = J_0\left(2\pi f r\, s(f)\right) - J_2\left(2\pi f r\, s(f)\right). \tag{3.43}$$

3.5.3 Rayleigh and Love waves

The SPAC method is particularly adapted to analyze seismic noise recordings. In a real case, the noise wave field will always be composed of both Rayleigh and Love waves (neglecting body waves). The Rayleigh wave dispersion curve can then be retrieved by using the vertical autocorrelation function $\rho_{vertical}(r,f)$ defined in equation (3.29). The horizontal components are not easy to analyze, however, as Love and Rayleigh waves both contribute with an unknown percentage. Rayleigh waves are radially polarized and Love waves are transverse waves. Therefore, we can identify

$$\rho_{Rayleigh,v}(r,f) = \rho_{vertical}(r,f) \tag{3.44}$$

$$\rho_{Rayleigh,r}(r,f) = \rho_{radial,r}(r,f) \tag{3.45}$$

$$\rho_{Rayleigh,t}(r,f) = \rho_{radial,t}(r,f) \tag{3.46}$$

$$\rho_{Love,r}(r,f) = \rho_{transverse,r}(r,f) \tag{3.47}$$

$$\rho_{Love,t}(r,f) = \rho_{transverse,t}(r,f). \tag{3.48}$$

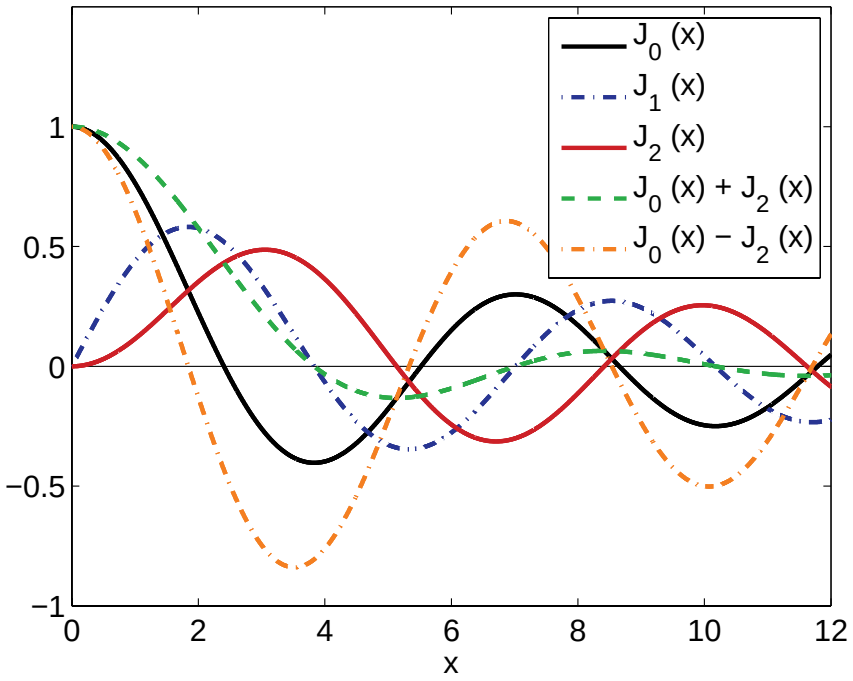


Figure 3.2: Graphs of the Bessel functions of the first kind $J_0(x)$, $J_1(x)$ and $J_2(x)$ as well as $J_0(x) - J_2(x)$ and $J_0(x) + J_2(x)$.

Nevertheless, if the ratio of Rayleigh waves in the wave field is called $\alpha(f)$, then the measurements yield the following properties:

$$\begin{aligned}
\rho_{vertical}(r,f) &= \rho_{Rayleigh,v}(r,f) \\
&= J_0\left(2\pi f r\, s_{Rayleigh}(f)\right), && (3.49)\\
\rho_{radial}(r,f) &= \alpha(f)\,\rho_{Rayleigh,r}(r,f) + (1-\alpha(f))\,\rho_{Love,r}(r,f) \\
&= \alpha(f)\left[J_0\left(2\pi f r\, s_{Rayleigh}(f)\right) - J_2\left(2\pi f r\, s_{Rayleigh}(f)\right)\right] \\
&\quad + (1-\alpha(f))\left[J_0\left(2\pi f r\, s_{Love}(f)\right) + J_2\left(2\pi f r\, s_{Love}(f)\right)\right], && (3.50)\\
\rho_{transverse}(r,f) &= \alpha(f)\,\rho_{Rayleigh,t}(r,f) + (1-\alpha(f))\,\rho_{Love,t}(r,f) \\
&= \alpha(f)\left[J_0\left(2\pi f r\, s_{Rayleigh}(f)\right) + J_2\left(2\pi f r\, s_{Rayleigh}(f)\right)\right] \\
&\quad + (1-\alpha(f))\left[J_0\left(2\pi f r\, s_{Love}(f)\right) - J_2\left(2\pi f r\, s_{Love}(f)\right)\right]. && (3.51)
\end{aligned}$$

$\rho_{vertical}$, ρ_{radial} and $\rho_{transverse}$ are determined by measurements and are in principle sufficient to fix the dispersion curves $s_{Rayleigh}(f)$ and $s_{Love}(f)$ as well as the ratio of Rayleigh waves α.

3.5.4 Applying the SPAC method

3.5.4.1 The Bessel functions

In figure 3.2, the Bessel functions of the first kind are plotted for the orders 0, 1 and 2. The calculation of the correlations between the central and the ring stations yields an oscillating function with decreasing amplitude. The roots of $J_0(x)$ are approximately $x = 2.405, 5.520, 8.654, 11.792, \ldots$ and the function has maxima at $x = 0, 7.016, \ldots$ and minima at $x = 3.832, 10.174, \ldots$. Therefore, the equation $J_0(x) = y$ has multiple solutions if $J_0(3.832) < y < J_0(7.016)$, which corresponds to $x > 1.869$. For the equations $J_0(x) + J_2(x)$ and $J_0(x) - J_2(x)$, the locations of the roots and extrema are different, but the shape of the functions is comparable to $J_0(x)$. Regarding the problem of finding the dispersion curve, this means that for higher frequencies, multiple slowness values exist which can explain the autocorrelation curve. Therefore, the use of autocorrelation curves should be limited to the first descending flank up to the first minimum of the curve. Additional measurements with different array radii can then be used to extend the investigation to different frequency ranges. However, it is possible to make a direct inversion of the soil structure without trying to retrieve the dispersion curve first [Wathelet et al., 2005]. Then, in principle, the whole SPAC curve could be used, but in general the curve beyond the second root is neither well constrained by the measurements (due to the array layout) nor improving the inversion.

3.5.4.2 Practical problems

An application of the SPAC method as it was derived is in principle impossible as an infinite number of seismic stations would be necessary so that the integration in equation (3.26) is mathematically possible. However, using less stations will still yield a curve which fits the Bessel function in an acceptable way. Formally, by using only N ring stations (regularly distributed on the ring of radius r), equation (3.26) transforms to

$$\begin{aligned}\overline{\phi(r,f)} = \; & \frac{1}{2N}\sum_n\sum_m A_n(f)A_m(f)\sum_{l=1}^{N}\cos\left(k(f)r\cos\left(\vartheta_m - \frac{2\pi\, l}{N}\right)\right)\cos(\varphi_n - \varphi_m) \\ & + \frac{1}{2N}\sum_n\sum_m A_n(f)A_m(f)\sum_{l=1}^{N}\sin\left(k(f)r\cos\left(\vartheta_m - \frac{2\pi\, l}{N}\right)\right)\sin(\varphi_n - \varphi_m),\end{aligned} \tag{3.52}$$

and

$$\phi(0,f) = \frac{1}{2}\sum_n\sum_m A_n(f)A_m(f)\cos(\varphi_n - \varphi_m). \tag{3.53}$$

If we suppose different waves to be uncorrelated, equations (3.52) and (3.53) simplify to

$$\overline{\phi(r,f)} = \frac{1}{2N}\sum_n A_n^2(f)\sum_{l=1}^{N}\cos\left(k(f)r\cos\left(\vartheta_n - \frac{2\pi\, l}{N}\right)\right), \tag{3.54}$$

$$\phi(0,f) = \frac{1}{2}\sum_n A_n^2(f). \tag{3.55}$$

In order to investigate how close the theoretical response will be to the Bessel function depending on the number of stations, we will suppose that a single plane wave is incident (i.e. $n = 1$ and $\vartheta_n = \vartheta$) and that the N ring sensors are located regularly on a ring around the central station. This yields

$$\rho(r,f,\vartheta) = \frac{1}{N}\sum_{l=1}^{N}\cos\left(k(f)r\cos\left(\vartheta - \frac{2\pi\, l}{N}\right)\right). \tag{3.56}$$

In figure 3.3, the autocorrelation function given by equation (3.56) is computed for arrays of different sensor numbers, varying the azimuth. For comparison reasons, the Bessel function $J_0(x)$ is also plotted. The first ring sensor is always placed exactly north of the central station. As the array is symmetric, the array responses of all azimuths can be reduced to the interval $[0^\circ, \vartheta_c]$, where $\vartheta_c = \frac{360^\circ}{4N}$ if N is odd and $\vartheta_c = \frac{360^\circ}{2N}$ if N is even. Therefore, the curves presented in figure 3.3 are displaying curves for ϑ varying from 0 to ϑ_c, which are representative for all possible azimuths. In the figure, it can be seen that for all curves the Bessel function is fitted up to a certain value $x = x_{limit}$ which depends on the number of stations. For two ring stations, the SPAC method cannot work if only waves from one azimuth arrive, which can also be seen in figure 3.3 (a). For odd numbers of ring stations, x_{limit} is larger than for even numbers of ring stations of comparable size. Notably, the resolution for three ring stations (figure 3.3 (b)) is superior than for four ring stations (figure 3.3 (c)), etc. The curves for 3 and 6 ring stations and for 5 and 10 ring stations are identical, respectively, because an array of 6 (10) sensors is built from two sub-arrays of 3 (5) stations. For these examples, in the array with less sensors, every vector connecting a ring and the central station occurs only once. In the respective array with more sensors, every connection vector occurs twice, without improving the resolution. In this way, every array whose number of ring sensors is the double of an odd number, has the same resolution than the respective array with half the number of stations. Actually, for every array of an even number of ring stations, an array with a smaller odd number and yet better resolution can be found.
Examining odd and even ring station numbers separately, the resolution of the array increases with the number of ring stations in both groups. Especially for the odd numbers, all arrays fit the Bessel function very well up to the first minimum at least. The three-station case resolves well up to the first minimum only, five stations up to the next maximum, seven stations up to the second minimum and so on. In this way, if the measurements shall be accurate up to a certain degree, the odd number of stations which are necessary to achieve this resolution can be given theoretically. Okada [2006] and Cho et al. [2008] found similar results. In section 3.5.4.3, a mathematical proof will be given illustrating that the Bessel function is fitted in a better way for an increasing (odd) number of ring stations.

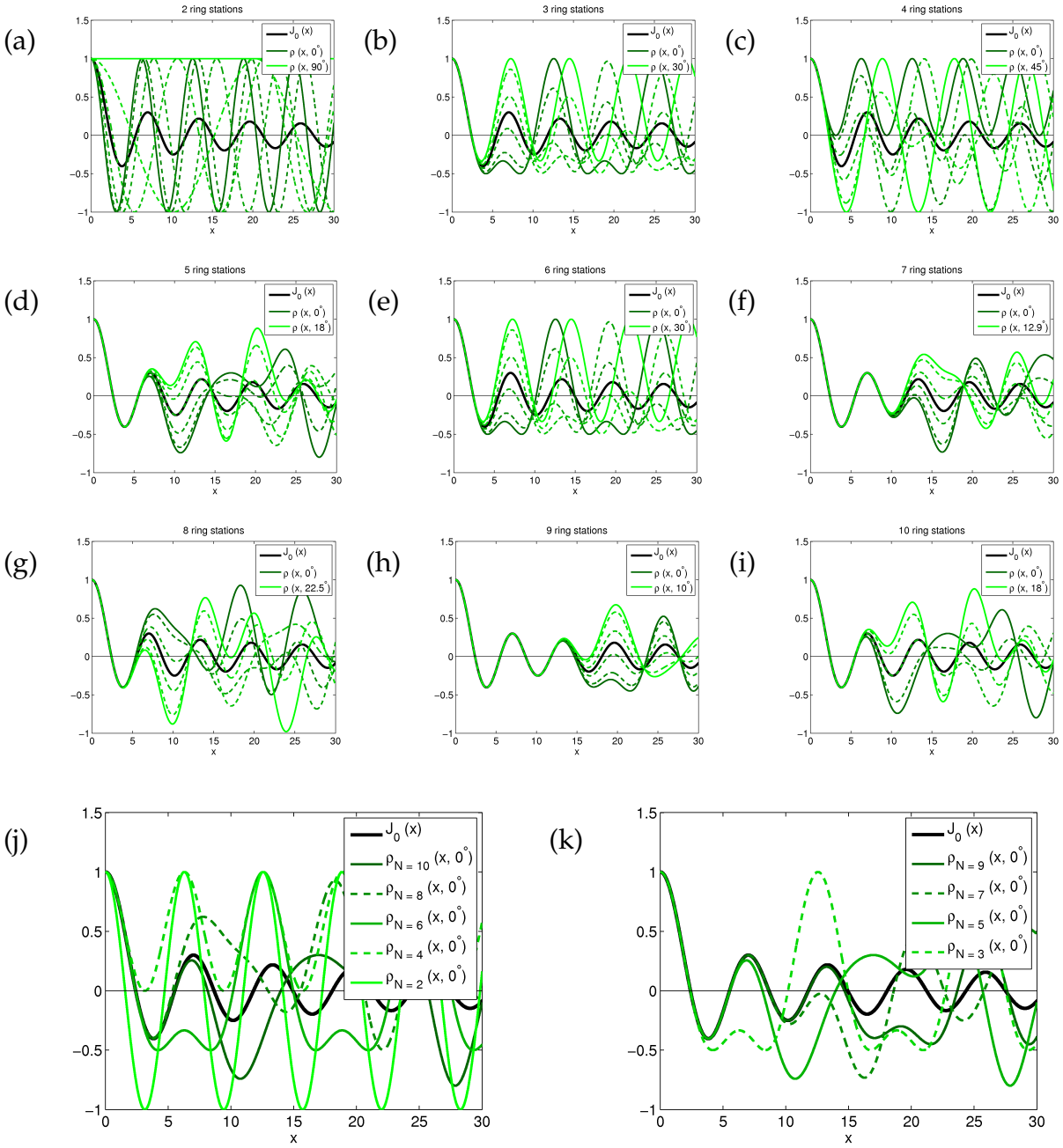


Figure 3.3: Top: Theoretical autocorrelation functions for seismic arrays composed of a central station and between two and ten ring stations for different azimuths. The azimuth values range from 0° to ϑ_c, with $\vartheta_c = \frac{360^\circ}{4N}$ if N, the number of ring sensors, is odd and $\vartheta_c = \frac{360^\circ}{2N}$ if N is even. Due to the symmetry of the respective array configurations, the value intervals are representative for all possible azimuths. In the legends, only the extreme azimuth values are indicated, the dashed lines correspond to the intermediate azimuth values. Bottom: Theoretical autocorrelation functions for even (left) and odd (right) numbers of ring stations for a single incident wave with an azimuth of 0°.

Even though with a larger number of ring sensors the theoretical resolution of the array increases, the non-uniqueness for values of x exceeding 1.869 subsists. This is why actually, the first descending flank of the Bessel function is often the only measurement used for the retrieval of the dispersion curve. According to figure 3.3, with three seismic sensors the Bessel function can already be retrieved up to the first minimum. Furthermore, the shown figures are obtained for a single incident wave, i.e. all wave energy is incident from a single azimuth. This is the worst possible scenario for using SPAC. In real case measurements, waves will be incident from all possible azimuths (principal azimuths with larger energy still being possible), and the results will correspond to a mixture of the curves for different azimuths shown in figure 3.3. If the waves arrive from all possible azimuths with the same energy (and different waves are uncorrelated), the measurements will correspond to the average of the curves for all azimuths. Actually, the average curves for all array sizes in figure 3.3 coincide with the Bessel function.
It should not be forgotten that the circular array layout with a central station is not mandatory. When using only the correlations between the central and the ring stations, the case of an array with three ring sensors surrounding a central station is indeed equivalent to the deployment of only three sensors forming an equilateral triangle. Nevertheless, the layout with a central station can be used to perform two distinguished SPAC measurements with different distance and thus frequency ranges: The first one between the central station and the ring stations and the second between the different ring stations. Different, but equivalent array layouts can always be formed because the importance lies in the distance vectors and not in the absolute positions.

3.5.4.3 Mathematical considerations relating to the number of seismic stations

In figure 3.3, it was shown that, for an odd number of ring stations, the Bessel function is fitted in a better way when increasing the number of seismic stations. Notably, when using 3 ring stations, the Bessel function was retrieved up to approximately the first minimum, for 5 ring stations up to the next maximum, for 7 stations up to the next minimum, and so on, independent of the azimuth of the incident wave.
We found a mathematical proof for this which will be given in the following. For the sake of simplicity, we suppose that a single seismic wave is arriving under an azimuth of 0°. This is one of both extreme cases which deviate the most from the true Bessel curve in figure 3.3.
The Bessel function $J_0(x)$ can be developed as a Taylor series, which yields

$$J_0(x) = \sum_{k=0}^{\infty} a_k x^{2k}, \tag{3.57}$$

with

$$a_k = \frac{(-1)^k}{2^{2k}(k!)^2}. \tag{3.58}$$

In the same way, we define the function $B_N(x) = \rho(x, \vartheta = 0)$, with $\rho(x = k(f) \cdot r, \vartheta)$ given by equation (3.56), where N is any odd number of seismic sensors on the ring:

$$B_N(x) = \frac{1}{N} \sum_{l=1}^{N} \cos\left(x \cos\left(\frac{2\pi l}{N}\right)\right). \tag{3.59}$$

Using the Taylor series of the cosine function[IX], $B_N(x)$ can be developed as a Taylor series as well:

$$\begin{aligned} B_N(x) &= \frac{1}{N} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!}\, x^{2k} \sum_{l=1}^{N} \left[\cos\left(\frac{2\pi l}{N}\right)\right]^{2k} \\ &= \sum_{k=0}^{\infty} b_k\, x^{2k}, \end{aligned} \tag{3.60}$$

with

$$b_k = \frac{1}{N} \cdot \frac{(-1)^k}{(2k)!} \sum_{l=1}^{N} \left[\cos\left(\frac{2\pi l}{N}\right)\right]^{2k}. \tag{3.61}$$

Applying the binomial theorem[X] to $\cos^n x$ yields

$$\cos^n x = \left(\frac{e^{ix} + e^{-ix}}{2}\right)^n = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} e^{ix(n-2k)}. \tag{3.62}$$

Now, b_k can be written as

$$b_k = \frac{1}{N} \cdot \frac{(-1)^k}{(2k)!} \sum_{l=1}^{N} \frac{1}{2^{2k}} \sum_{j=0}^{2k} \binom{2k}{j} \exp\left(i \frac{2\pi l}{N}(2k - 2j)\right). \tag{3.63}$$

[IX] The Taylor series of the cosine is given by

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}.$$

[X] The binomial theorem indicates the power of a binomial [Bronstein et al., 2001]:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient and $x!$ is the factorial of $x \in \mathbb{N}$ defined by $x! = x \cdot (x-1) \cdot (x-2) \cdots 2 \cdot 1$.

When calculating

$$\sum_{l=1}^{N} \exp\left(i\frac{2\pi l}{N}(2k-2j)\right), \tag{3.64}$$

two cases have to be considered separately.
In the first case, $k < N$ and therefore $k - j < N$. In this case, equation (3.64) equals zero except if $j = k$. Then, equation (3.64) equals $N\delta_{jk}$ and equation (3.63) simplifies to

$$b_k = \frac{(-1)^k}{(2k)!}\frac{1}{2^{2k}}\frac{(2k)!}{(k!)^2} = \frac{(-1)^k}{2^{2k}(k!)^2}, \tag{3.65}$$

which is equal to a_k defined in equation (3.58).
In the second case, $k \geq N$. Consequently, as in equation (3.63) a summation over j occurs, at least two values of j will be found for which $(k - j)$ is a multiple of N (e.g. for $k = N$, this is the case if $j = 0$ and $j = 2N$). Notably, if $k = N$, equation (3.64) gives the result $N(\delta_{jN} + \delta_{j0} + \delta_{j,2N})$. Then, b_N can be calculated as

$$\begin{aligned} b_N &= \frac{1}{N}\cdot\frac{(-1)^N}{(2N)!}\frac{1}{2^{2N}}\sum_{j=0}^{2N}\binom{2N}{j}N(\delta_{jN}+\delta_{j0}+\delta_{j,2N}) \\ &= \frac{(-1)^N}{(2N)!}\frac{1}{2^{2N}}\left[\frac{(2N)!}{(N!)^2}+2\right] = a_N + \frac{(-1)^N}{(2N)!}\frac{1}{2^{2N-1}}, \end{aligned} \tag{3.66}$$

which differs from a_k defined by equation (3.58). This result is in agreement with the results found by Okada [2006], using a different derivation.
This proves that for a given odd number N of regularly distributed seismic sensors on the ring, the SPAC measurement retrieves the Taylor series of the Bessel function up to the $k = 2(N-1)$th element. The terms for $k \geq N$ diverge from the terms of the Bessel function and have larger absolute values. In other words, $J_0(x) - B_N(x) = \mathcal{O}(x^{2N})$. This explains the results of figure 3.3. For $N = 3$, both curves' coefficients are identical up to x^4. Therefore, as the Bessel function is symmetrical ($J_0(-x) = J_0(x)$), the first two roots (a negative and a positive one) are well fitted by $B_3(x)$ and the first minima are approximately fitted as well. With each increment of N by 2, the agreement between $J_0(x)$ and $B_N(x)$ increases of two additional terms and consequently the curves are in agreement for the next root and approximately up to the next extremum.

3.5.5 Modified SPAC method (M-SPAC)

The original SPAC method forces the seismic sensors to have equal distances one from another. When performing field measurements, however, this requirement is almost impossible to fulfill. In urban areas in particular, buildings, infrastructure and plants have to be accounted for when deploying the sensors. The classical method cannot be applied there. Bettig et al. [2001] proposed the modified auto-correlation method (M-SPAC). This method allows the seismic sensors to be located in approximate rings or even completely different layouts. Sensor pairs of comparable distance are then assembled into rings.

3.5.5.1 Vertical component

For the vertical component, grouping all sensor pairs with distances between r_1 and r_2 should yield an autocorrelation function which is the average function for all radii between r_1 and r_2. This average is calculated by integrating equation (3.29) for radii between r_1 and r_2 and renorming the result:

$$\rho_{MSPAC,vertical}(r_1, r_2, f) = \frac{\int_{r_1}^{r_2} J_0(2\pi f r\, s(f)) r\, dr}{\int_{r_1}^{r_2} r\, dr}. \tag{3.67}$$

By using $\int_0^r J_0(x) x dx = r\, J_1(r)$, the integrals can be calculated:

$$\begin{aligned}
\int_{r_1}^{r_2} J_0(2\pi f r\, s(f)) r\, dr &= \int_0^{r_2} J_0(2\pi f r\, s(f)) r\, dr - \int_0^{r_1} J_0(2\pi f r\, s(f)) r\, dr \\
&= \frac{1}{4\pi^2 f^2\, s^2(f)} \left[\int_0^{r_2'} J_0(r') r'\, dr' - \int_0^{r_1'} J_0(r') r'\, dr' \right] \\
&= \frac{1}{4\pi^2 f^2\, s^2(f)} \left[r_2'\, J_1(r_2') - r_1'\, J_1(r_1') \right] \\
&= \frac{1}{2\pi f\, s(f)} \left[r_2\, J_1(2\pi f r_2\, s(f)) - r_1\, J_1(2\pi f r_1\, s(f)) \right], \qquad (3.68) \\
\int_{r_1}^{r_2} r\, dr &= \left[\frac{r^2}{2} \right]_{r_1}^{r_2} = \frac{r_2^2 - r_1^2}{2}. \qquad (3.69)
\end{aligned}$$

Inserting (3.68) and (3.69) in (3.67) yields

$$\rho_{MSPAC,vertical}(r_1, r_2, f) = \frac{1}{(r_2^2 - r_1^2)\pi f\, s(f)} \left[r_2\, J_1(2\pi f r_2\, s(f)) - r_1\, J_1(2\pi f r_1\, s(f)) \right]. \tag{3.70}$$

This is the theoretical formula. The measured correlations for all station pairs (i, j) with distance r_{ij} comprised between r_1 and r_2 are weighted and averaged:

$$\rho_{vertical}(r_1, r_2, f) = \frac{1}{\pi} \sum_{r_1 < r_{ij} < r_2} \rho(r_{ij}, \varphi_{ij}, f) \Delta\varphi_{ij}, \tag{3.71}$$

where φ_{ij} indicates the angle under which station j is seen from station i or vice versa ($0 < \varphi_{ij} < \pi$) and $\Delta\varphi_{ij}$ (with $\sum_{r_1 < r_{ij} < r_2} \Delta\varphi_{ij} = \pi$) is a weighting factor which takes the density of the distribution of sensor pairs at angle φ_{ij} into account (if many station pairs with comparable φ fall in the same ring, their respective weights are low). The autocorrelations $\rho(r_{ij}, \varphi_{ij}, f)$ for each station pair are calculated by

$$\rho(r_{ij}, \varphi_{ij}, f) = \frac{\int u_{i,vertical}(f,t) u_{j,vertical}(f,t) dt}{\sqrt{\int u^2_{i,vertical}(f,t) dt \int u^2_{j,vertical}(f,t) dt}}. \tag{3.72}$$

3.5.5.2 Horizontal components

The M-SPAC technique was extended to the horizontal components by Köhler et al. [2007]. In analogy to the vertical component, the autocorrelation functions for the horizontal components given by equations (3.50) and (3.51) are averaged over the radii between r_1 and r_2:

$$\rho_{MSPAC,radial}(r, f) = \frac{\int_{r_1}^{r_2} \rho_{radial}(r,f) r \, dr}{\int_{r_1}^{r_2} r \, dr}, \tag{3.73}$$

$$\rho_{MSPAC,transverse}(r, f) = \frac{\int_{r_1}^{r_2} \rho_{transverse}(r,f) r \, dr}{\int_{r_1}^{r_2} r \, dr}. \tag{3.74}$$

By using $\int_{r_1}^{r_2} [J_0(r) - J_2(r)]\, r\, dr = 2\,[J_0(r_2) - J_0(r_1) + r_2 J_1(r_2) - r_1 J_1(r_1)]$
and $\int_{r_1}^{r_2} [J_0(r) + J_2(r)]\, r\, dr = 2\,[J_0(r_1) - J_0(r_2)]$, this finally leads to

$$\begin{aligned}\rho_{MSPAC,radial}(r,f) = \frac{\alpha(f)}{r_2^2 - r_1^2}\Bigg(&\frac{1}{\pi^2 f^2\, s_R^2(f)}\,[J_0(2\pi f r_2\, s_R(f)) - J_0(2\pi f r_1\, s_R(f))] \\ &+ \frac{2}{\pi f\, s_R(f)}\,[r_2 J_1(2\pi f r_2\, s_R(f)) - r_1 J_1(2\pi f r_1\, s_R(f))]\Bigg) \\ &- \frac{1-\alpha(f)}{(r_2^2 - r_1^2)\pi^2 f^2 s_L^2(f)}\,[J_0(2\pi r_2 f s_L(f)) - J_0(2\pi r_1 f s_L(f))]\,,\end{aligned} \tag{3.75}$$

$$\begin{aligned}\rho_{MSPAC,transverse}(r,f) = &\frac{-\alpha(f)}{(r_2^2 - r_1^2)\pi^2 f^2 s_R^2(f)}\,[J_0(2\pi r_2 f s_R(f)) - J_0(2\pi r_1 f s_R(f))] \\ &+ \frac{1-\alpha(f)}{r_2^2 - r_1^2}\Bigg(\frac{1}{\pi^2 f^2 s_L^2(f)}\,[J_0(2\pi f r_2 s_L(f)) - J_0(2\pi f r_1 s_L(f))] \\ &+ \frac{2}{\pi f s_L(f)}\,[r_2 J_1(2\pi f r_2 s_L(f)) - r_1 J_1(2\pi f r_1 s_L(f))]\Bigg),\end{aligned} \tag{3.76}$$

where $s_R(f) = s_{Rayleigh}(f)$ and $s_L(f) = s_{Love}(f)$. In analogy to the vertical component, the measured correlations for all station pairs (i, j) with $r_1 < r_{ij} < r_2$ are averaged and weighted to form the radial and transverse correlation functions

$$\rho_{radial}(r_1, r_2, f) = \frac{1}{\pi} \sum_{r_1 < r_{ij} < r_2} \rho_{radial}(r_{ij}, \varphi_{ij}, f)\Delta\varphi_{ij}, \tag{3.77}$$

$$\rho_{transverse}(r_1, r_2, f) = \frac{1}{\pi} \sum_{r_1 < r_{ij} < r_2} \rho_{transverse}(r_{ij}, \varphi_{ij}, f)\Delta\varphi_{ij}, \tag{3.78}$$

where ρ_{radial} and $\rho_{transverse}$ are calculated as in equation (3.72).

3.5.6 Two-sites SPAC method (2s-SPAC)

Morikawa et al. [2004] proposed to perform SPAC measurements using only two seismic sensors (2s-SPAC). In this way, the logistical requests for the measurements can be minimized as less seismic stations and less manpower are necessary. In the classical SPAC method, it is averaged over a large number of seismic sensors placed at specified locations. The 2s-SPAC method replaces the spatial averaging by a temporal averaging. This is only possible if the composition of the wave field is constant in time. The temporal averaging is done by using two seismic sensors, one fixed at a central station and the other one moving to different locations during the measurement. The necessary measuring time at each location depends on the frequency of interest. In principle, the method can be used to measure an arbitrary number of station pairs with equal distance, but the constraint that the wave field composition must not change during the measurements implies that as few locations as necessary should be used. Therefore, the layout using only three station pairs for each distance is preferable for these measurements. Isotropy of the wave field is not required for 2s-SPAC measurements as the only additional requirement compared to classical SPAC measurements is the time constance of the wave field composition.

3.5.7 Application of 2s-SPAC to real data measurements

A good possibility to test the two-station SPAC method is given by measurements which we have performed in Patras, Greece, in July 2009, within the framework of an active and passive surface wave technique benchmark. This test was organized by USGS, the University of Patras and ITSAK (Thessaloniki, Greece). At two different locations (near a school and near a hospital), seismic noise measurements have been accomplished using arrays of ten seismic sensors. The smallest arrays at both locations are appropriate to test 2s-SPAC because the sensors were located regularly on circles of 10 m radius (school site) and 8 m radius (hospital site). The respective sensor locations are shown in figures 3.4 (a) and 3.6 (a). For the school site, the spacing of the sensors is perfectly regular, with angles of 40° between neighboring ring stations. For the hospital site, the location of station 10 was deviating from the regular position because of local constraints. Due to the regular array layouts of both sites, subarrays of four sensors can be built using the central station and the station sets (2-5-8), (3-6-9) and (4-7-10). For the hospital deployment, the comparisons include subarrays of the central station and the stations (2-5-8) and (3-6-9) only. The theoretical autocorrelation measurements for a single seismic wave arriving on arrays of 9 and 3 ring stations have been shown in figures 3.3 (b) and (h), respectively. Due to the constant station pair distances, the classical SPAC method (see 3.5) can be applied here.

3.5.7.1 Location near the school

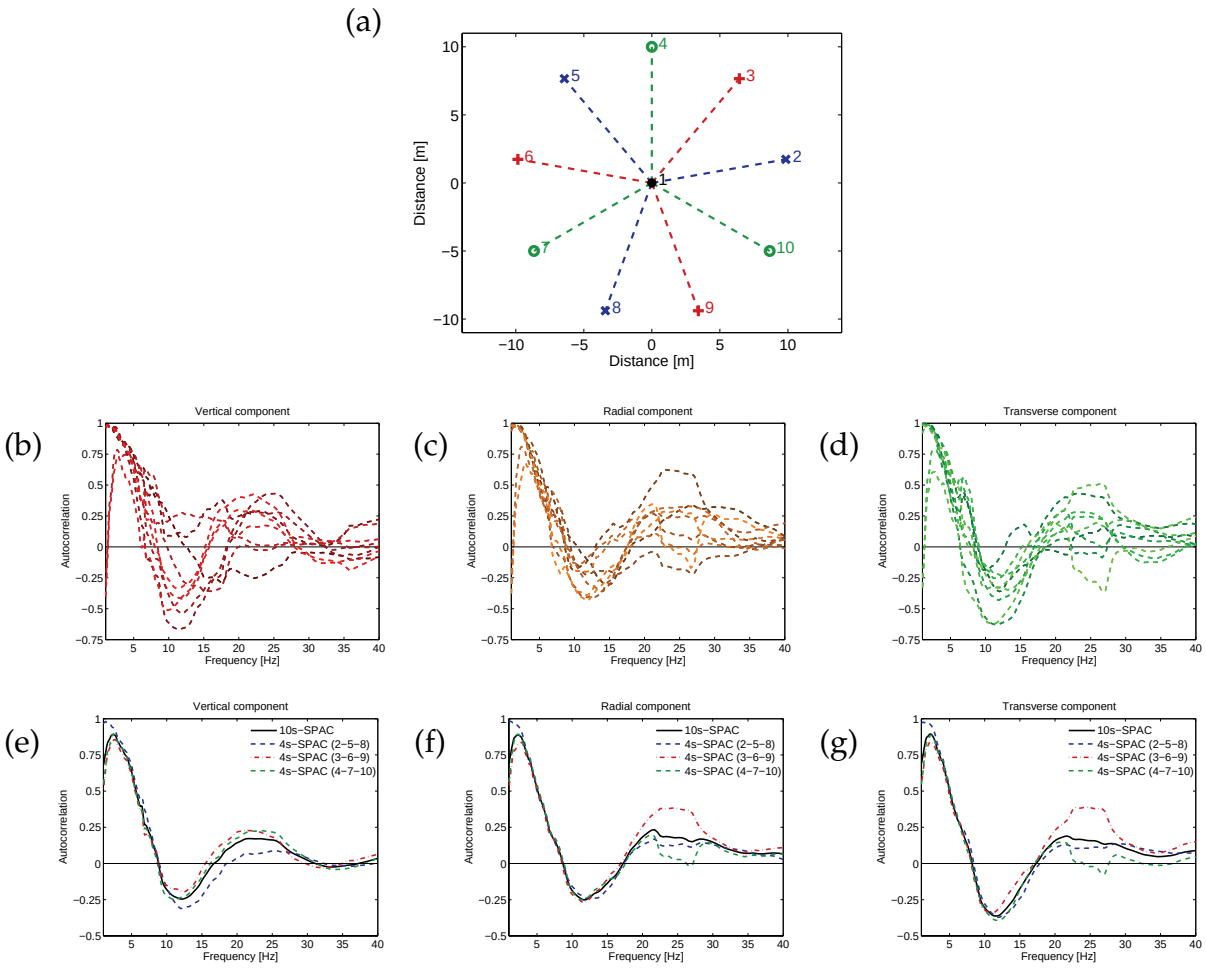


Figure 3.4: (a) Layout of the seismic array and definition of the different subarrays (all subarrays include the central station). Autocorrelation function between the central and all ring stations for (b) the vertical, (c) the radial and (d) the transverse component. Comparison between the average autocorrelation curve using all ten stations simultaneously (10s-SPAC, using 9 station pairs between the central and the ring stations) and the three subarrays (4s-SPAC, using 3 station pairs between the central and ring stations) for (e) the vertical, (f) the radial and (g) the transverse component.

The array layout of the measurements performed near the school is shown in figure 3.4 (a). The total duration of simultaneous seismic signal recordings is 45 minutes. The measured autocorrelation functions for every station pair consisting of the central and a ring station are shown in figures 3.4 (b) - (d) for the vertical, radial and transverse components, respectively. If the seismic noise wave field had been completely isotropic, the autocorrelation function for every seismic sensor pair should be the same.

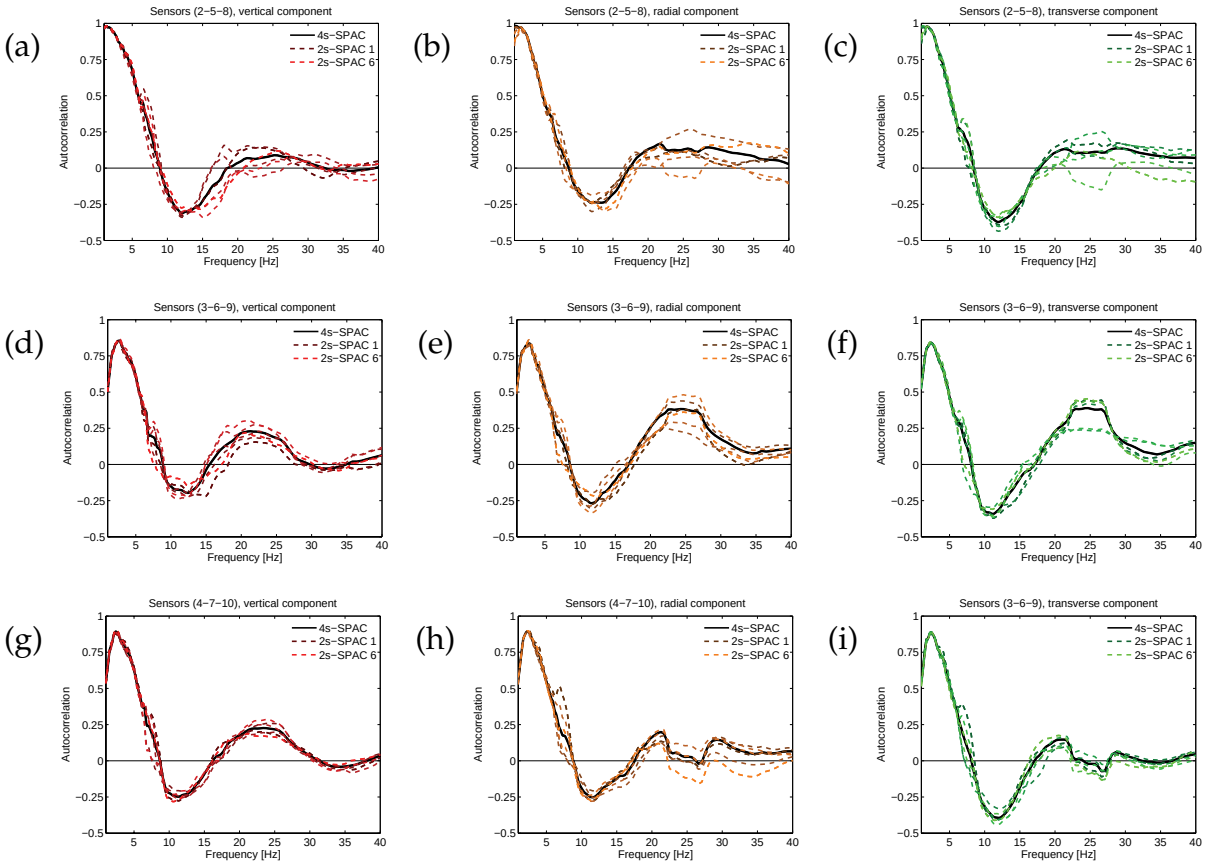


Figure 3.5: Comparison between 4s-SPAC and the six possible realizations of 2s-SPAC for the three subarrays and the three different signal components.

The fact that the curves for different station pairs are very different indicates that the noise generation is not homogeneously distributed in space and that principal noise azimuths exist. The different curves resemble the possible theoretical autocorrelation curves for single station pairs shown in figure 3.3 (a) (the case for two ring sensors being equivalent to the case of a single station pair).
In figures 3.4 (e) - (g), the averaged SPAC curves for all 9 station pairs and for the three subarrays (with 3 station pairs) are shown for the vertical, radial and transverse components, respectively. According to figures 3.3 (b) and (h), the average using 9 station pairs should fit the Bessel function up to the second non-zero maximum, whereas the average using 3 station pairs should only fit the first descending part of the Bessel function. Keeping this in mind, the shown curves are in very good agreement.
In the next step, we want to test the 2s-SPAC method. This method consists in measuring the autocorrelation between two stations for a given period of time, then moving one station to a different location and measuring for the same period of time and finally moving it to a third position and measuring a last period of time. Using the recordings of any subarray, this can be simulated in a very simple way.

As the total recording length is 45 minutes, we will calculate the autocorrelation using the first 15 minutes of signal for the central and the first ring station, the second 15 minutes of signal for the central and the second ring station and finally the third 15 minutes for the pair involving the third ring station.
In this way, using a single subarray, six different 2s-SPAC measurements can be simulated, depending on which ring station is the first, second and third one. The results of this procedure are given in figure 3.5. Figures 3.5 (a) - (c) give the results for using the subarray (2-5-8) for the vertical, radial and transverse components, respectively, (d) - (f) for the subarray (3-6-9) and (g) - (i) for the subarray (4-7-10). According to the theoretical considerations of figure 3.3, the measurements should fit the correct Bessel function up to the first minimum only. In this frequency range and apart from some minor discrepancies, the 2s-SPAC curves are in very good agreement with the classical SPAC curves using all four stations at the same time (4s-SPAC) for every combination of subarray and component. The non-isotropy of the noise wave field does not constitute a problem for the 2s-SPAC measurements. Besides, the fact that the 2s-SPAC measurements agree with the 4s-SPAC measurements indicates that the constitution of the seismic noise wave field did not change during the 45 minutes of measurement.

3.5.7.2 Location near the hospital

For the measurements at the location near the hospital, the array layout is given in figure 3.6 (a). At this location, 60 minutes of seismic signals have been recorded simultaneously on all stations. Figures 3.6 (b) - (d) show the autocorrelation functions for every station pair consisting of the central and a ring station. Here again, the noise wave field is not perfectly isotropic. Figures 3.6 (e) - (g) indicate the average autocorrelation functions using 9 seismic sensors and two subarrays of 4 sensors. As sensor 10 could not be placed on the exact location necessary for the regular array, it was placed about a meter aside. Therefore, this station is disregarded in the following. For all components, the measurements using only 4 stations are comparable with the measurements including all stations, but the respective curves are far away from being Bessel functions. The results for the six possible realizations of a simulated 2s-SPAC are shown for one of both subarrays only in figures 3.6 (h) - (j). For the other subarray, the results are comparable. For 2s-SPAC, the curves do not look like a Bessel function neither, but are still close to the 4s-SPAC using the same data. The strange autocorrelation curves could be linked to the presence of higher modes in the wave field or a too narrow seismic sensor spacing. In fact, if the wave velocity is high and the sensors are too closely located, it can happen that the SPAC measurements at higher frequencies deviate from the Bessel function.

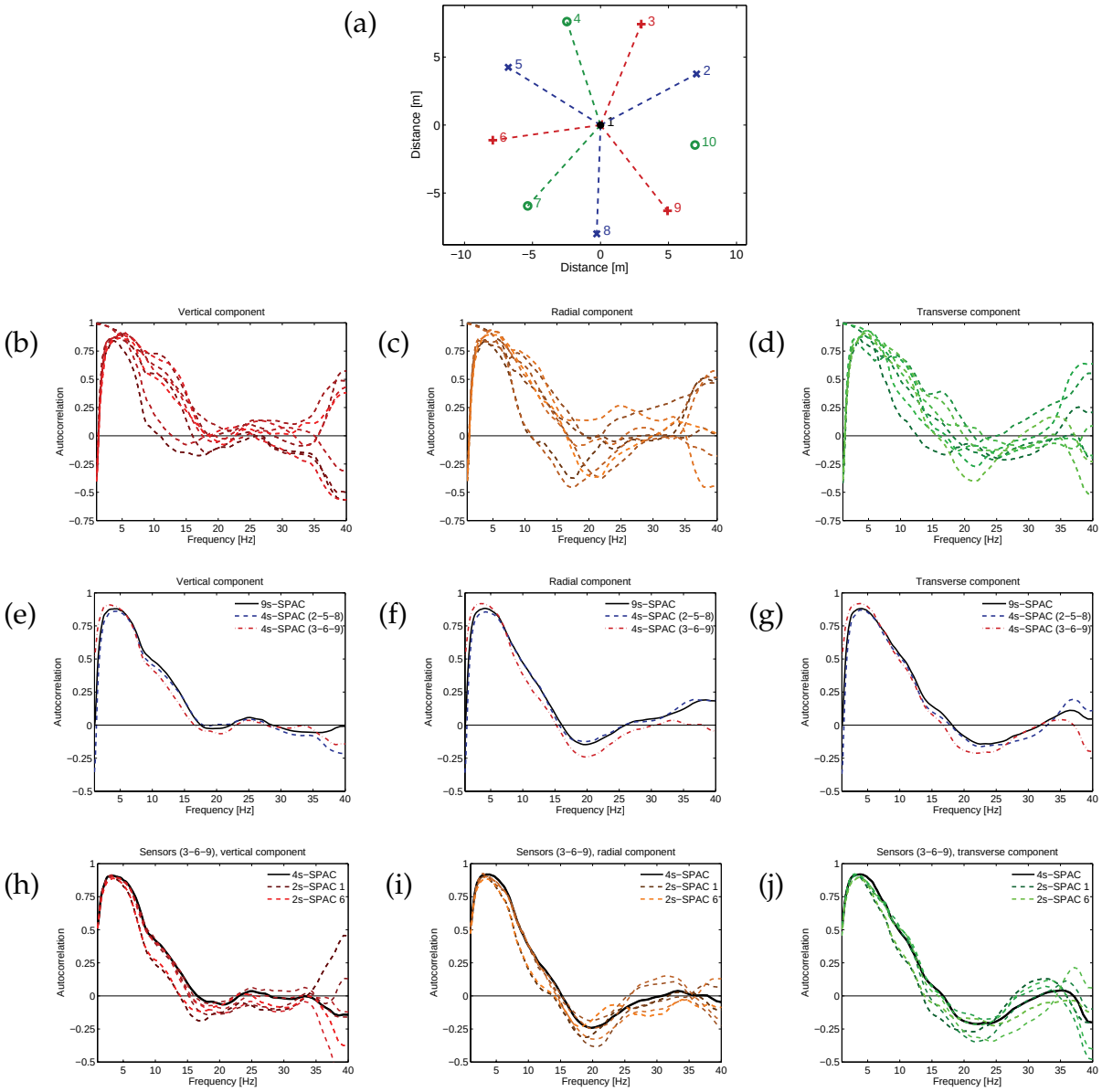


Figure 3.6: (a) Layout of the seismic array and definition of the different subarrays (all subarrays include the central station). Autocorrelation function between the central and all ring stations (except station 10) for (b) the vertical, (c) the radial and (d) the transverse component. Comparison between the average autocorrelation curves using the 8 station pairs between the central and the ring stations simultaneously (9s-SPAC) and the two subarrays (4s-SPAC, using 3 station pairs between the central and ring stations) for (e) the vertical, (f) the radial and (g) the transverse component. (h-j) Comparison between 4s-SPAC and the six possible realizations of 2s-SPAC for the subarray (3-6-9) and the three different signal components.

To investigate this, a measurement with larger sensor spacings is necessary. Due to local constraints, the sensors of the larger arrays could not be located very regularly. Nevertheless, it is possible to choose different station pairs with larger distances for the small array. For the following considerations, the central station of the array and the misplaced station 10 have been omitted. In figure 3.7 (a), the remaining 8 stations are grouped in 7 station pairs (the distances for these pairs being 15.75 m compared to 8 m in the previous case). The autocorrelation functions for each of these pairs are shown in figures 3.7 (b) - (d) for the three components. As already stated above for the smaller distance, the differences of the curves illustrate that the noise wave field is not isotropic. Using the station pairs, it is possible to define a "subarray" consisting of the station pairs (2, 6), (3, 8) and (5, 9). This subarray is not really a subarray as each station pair is formed by different stations, but the linking vectors form a regular equilateral triangle. In figure 3.7 (e) - (g), the average of the autocorrelation curves for the different station pairs are shown for using all station pairs (i.e. 7 pairs including 8 stations) and using only the station pairs forming the subarray (which will be called 3p-SPAC in the following, as three station pairs are used). Here, the concordance with a Bessel function is much more pronounced. On the vertical component, above the first root, however, the curve is deviating from the expected form in a way similar to the previous case of smaller distances. Nevertheless, using larger array distances shifts the theoretical Bessel function to lower frequencies (as the argument of the Bessel function is $2\pi frs(f)$) and shifts the first root out of the frequency range where the autocorrelation cannot be calculated correctly. For the radial and transverse components, however, the curves resemble a Bessel curve at higher frequencies as well. On all three components, the curves for using all 7 station pairs and only 3 station pairs are almost identical.
By using the subarray, six different realizations of 2s-SPAC can be simulated as well. In contrast to the classical 2s-SPAC method with a single moving sensor, this would correspond to a measurement with two moving sensors. The results for all six realizations are given in figures 3.7 (h) - (j). On all three components, the 2s-SPAC curves are in perfect agreement with the 3p-SPAC measurement up to the first root. At higher frequencies, however, the curves deviate considerably.

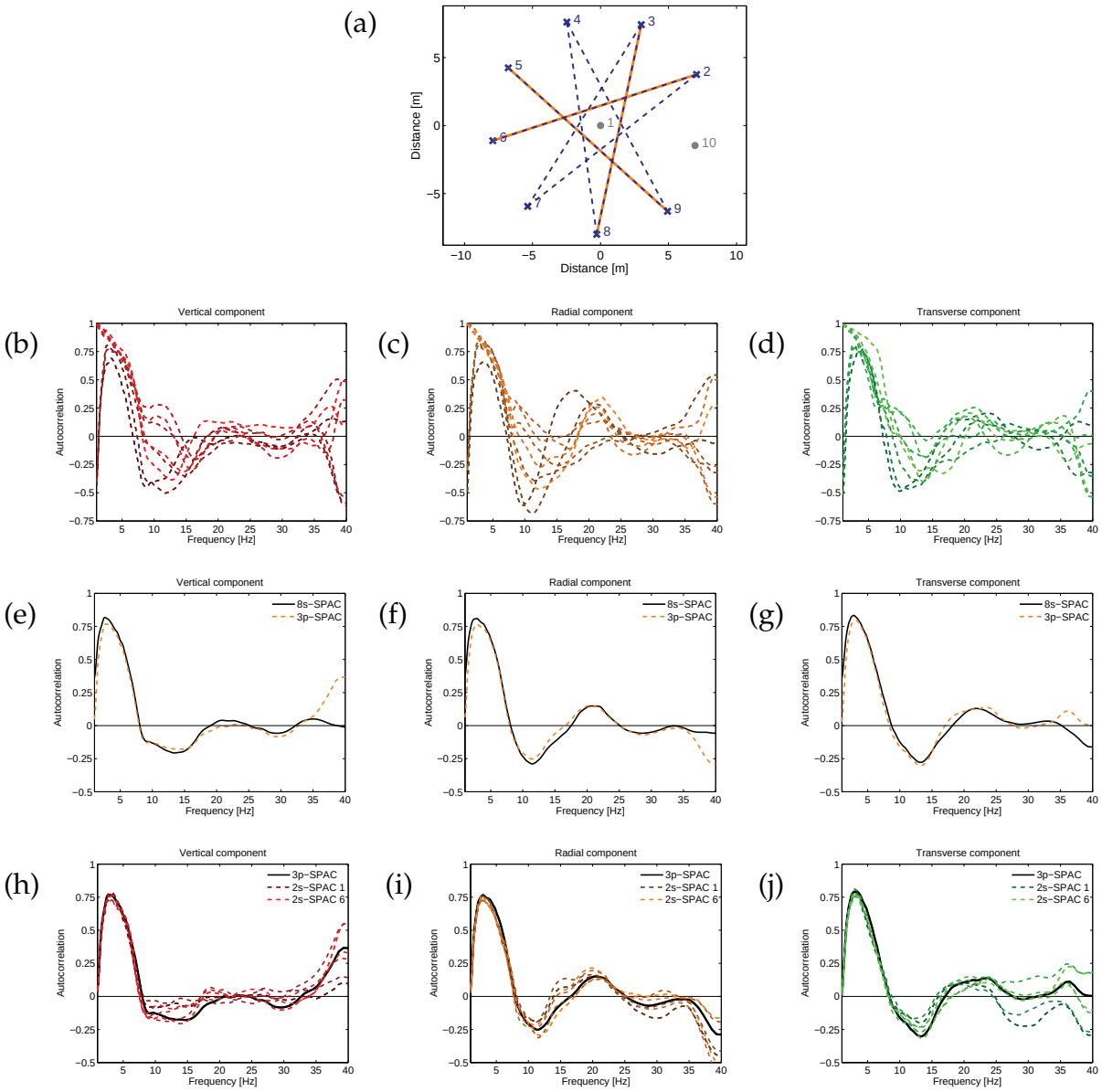


Figure 3.7: (a) Layout of the seismic array and definition of subarray (orange). Autocorrelation function for the pairs of stations indicated in (a) for (b) the vertical, (c) the radial and (d) the transverse component. Comparison between the average autocorrelation curves using all 7 station pairs (8s-SPAC) and the subarray (3p-SPAC, using the 3 orange station pairs) for (e) the vertical, (f) the radial and (g) the transverse component. (h-i) Comparison between 3p-SPAC and the six possible realizations of 2s-SPAC for the orange subarray and the three different signal components.

3.6 Conclusion

In this chapter, important methods to characterize seismic waves using arrays of sensors have been presented. Of course, many more methods exist [Marcos, 1998], which have not been presented here. The first presented methods (i.e. beamforming, frequency-wavenumber analysis and MUSIQUE) allow the estimation of an incident wave vector (i.e. the estimation of azimuth and velocity), whereas SPAC only allows the estimation of the velocity.

Theoretical considerations of the SPAC method demonstrated that an odd number of ring stations is always preferable to an even number of ring stations. Indeed, the classical layout with a ring of sensors surrounding a central station itself is not essential. By omitting the central station, the ring stations still form the same number of station pairs at each distance value. For example, an array of 3 stations surrounding a central station can be replaced by a smaller array of 3 stations only, positioned in a way to form an equilateral triangle. Nevertheless, for the layout including the central station, SPAC measurements for two different distance values are obtained.

The tests of the 2s-SPAC method did not show significant differences between SPAC measurements including 9 station pairs, 3 station pairs or 2s-SPAC. The 2s-SPAC method is therefore an interesting means of characterizing the wave field with the least possible measurement effort. The only additional prerequisite of the method is the temporal stationarity of the wave field. However, when performing a real 2s-SPAC measurement, this property can not be estimated easily. If the wave field composition changes during the measurement, the measured energy at the reference station is likely to change, too. However, it is still possible that the energy of the wave field stays constant, but the principal source azimuths change. A fourth measurement at the same location as the first one could affirm the temporal stationarity of the signals. To assure the temporal stationarity of the wave field, the measurement times of the 2s-SPAC method should be kept as short as possible.

SPAC measurements will be used to fix the absolute values of the velocities in the soil model in joint inversions with ellipticity curves: In chapter 6, the theoretical aspects of joint inversions of Rayleigh wave ellipticity and SPAC curves are investigated, before the results are applied to real data measurements in chapter 7.
In chapter 4, advanced versions of the MUSIC algorithm using quaternions or biquaternions are presented. This will then lead to the development of the MUSIQUE algorithm.

Chapter 4

Advanced seismic array methods using hypercomplex numbers

This chapter is dedicated to array methods which use hypercomplex numbers. First, the properties of quaternions and biquaternions, two examples of hypercomplex numbers, are indicated. Then, versions of the MUSIC algorithm exploiting these properties to retrieve the polarization parameters between different signal components are presented. Finally, the MUSIQUE algorithm is developed, which combines the classical MUSIC algorithm with a quaternion-MUSIC algorithm and allows the estimation of azimuth and velocity, the identification of Love and Rayleigh waves and the polarization estimation for Rayleigh waves.

Ce chapitre est dédié aux méthodes de réseau utilisant des nombres hypercomplexes. Dans un premier temps, les propriétés de quaternions et de biquaternions, deux exemples de nombres hypercomplexes, seront indiquées. Après, des versions de l'algorithme MUSIC tirant profit des propriétés de ces nombres pour retrouver la polarisation entre différentes composantes du signal seront présentées. Ceci mènera finalement au développement de l'algorithme MUSIQUE, une combinaison de l'algorithme classique avec l'algorithme quaternionique de MUSIC. Cet algorithme permet l'estimation d'azimut et de vitesse ainsi que l'identification d'ondes de Love et de Rayleigh et l'estimation de la polarisation de ces dernières.

4.1 Introduction

In section 3.4, the original MUSIC algorithm [Schmidt, 1986] was presented. This code deals with monocomponent signals only and is not particularly designed for multicomponent signals. Therefore, three-component signals can be either analyzed for each signal component separately or by putting the signals of the different components in a single data vector. However, the polarization information between the different components is lost in both cases. To keep the polarization relations between the different components, their respective data have to be stored separately. The MUSIC data matrix for a single component is complex valued. In order to store multiple components in a single data matrix, hypercomplex numbers can be used.

Examples of hypercomplex numbers are the quaternions (of dimension 4) and the biquaternions (of dimension 8). In this way, two complex-valued data matrices corresponding to different components can be stored in a single quaternion-valued data matrix or all three components of the seismic signals can be stored in a single biquaternion matrix. In fact, MUSIC versions using quaternions have been proposed by Miron et al. [2005b, 2006] and using biquaternions by Miron [2005].

This chapter starts with presenting the properties of quaternions and biquaternions which are necessary to understand the respective MUSIC versions (sections 4.2 and 4.3). In section 4.4, the original methodology of quaternion-MUSIC [Miron et al., 2006] is indicated. We show that in contrast to the original method, it is not necessary to perform a grid search to determine the two polarization parameters of a two-component wave (i.e. amplitude ratio and phase difference) if the wave vector (azimuth and velocity) is already known. In this case, the polarization parameters can be obtained analytically. Then, in section 4.5, the biquaternion version of MUSIC as proposed by Miron [2005] is presented. Here again, we show that the grid search for the four polarization parameters describing three-component waves (two amplitude ratios and two phase differences) can be largely simplified if the wave vector has been previously determined. However, we also show that the biquaternion approach fails in basic cases as identical signals on two or three components or two-dimensional signals.

These considerations finally lead to the development of MUSIQUE, which combines the classical MUSIC algorithm and the quaternion-MUSIC algorithm. MUSIQUE is designed to determine azimuth and velocity of incident waves, identify Love and possible Rayleigh waves and estimate the polarization parameters (among them ellipticity) of possible Rayleigh waves. The algorithm is presented in section 4.6.

4.2 Quaternions

Quaternions are an extension of the complex numbers and have been discovered by the Irish mathematician Sir William Rowan Hamilton in 1843[I](Hamilton [1844-50]). In fact, the French mathematician Olinde Rodrigues had already used quaternions in 1840 to describe rotations in 3-dimensional space (Rodrigues [1840]), but indeed without calling them quaternions. Later, his work fell into oblivion. Kantor and Solodovnikov [1989] give a good overview of quaternion algebra and Sweetser [2005] demonstrates how the use of quaternions can simplify basic equations in physics. Using quaternions, rotations can be calculated faster than using rotation matrices. Therefore, the use of quaternions includes video cards in modern computers. Here, only the basic rules of quaternion algebra which are necessary to understand the following section will be presented.

4.2.1 Definition

A quaternion is defined as

$$q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}, \tag{4.1}$$

where $q_0, \ldots, q_3 \in \mathbb{R}$ and $\mathtt{i}$, $\mathtt{j}$ and $\mathtt{k}$ are imaginary units defined by

$$\mathtt{i}^2 = \mathtt{j}^2 = \mathtt{k}^2 = \mathtt{ijk} = -1 \tag{4.2}$$

q_0 is called scalar part of the quaternion and $q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}$ forms the vector part. The set of all quaternions is denoted by $\mathbb{H}$. Based on the equations (4.2), it follows that

$$\begin{aligned} \mathtt{ij} &= \mathtt{k}, & \mathtt{ji} &= -\mathtt{k} \\ \mathtt{jk} &= \mathtt{i}, & \mathtt{kj} &= -\mathtt{i} \\ \mathtt{ki} &= \mathtt{j}, & \mathtt{ik} &= -\mathtt{j} \end{aligned} \tag{4.3}$$

Evidently, the multiplication of quaternions is not commutative.

[I]After spending many years on the fruitless quest for a three-dimensional extension of the complex numbers, in 1843, while on a walk with his wife, Hamilton got suddenly struck by the idea of a four-dimensional extension, the quaternions. As he was crossing the Broom Bridge in Dublin, Ireland, he immediately carved the fundamental formula defining quaternions into the bridge. Today, a plaque on the bridge commemorates this event. Its inscription says:

"Here as he walked by
on the 16th of October 1843
Sir William Rowan Hamilton
in a flash of genius discovered
the fundamental formula for
quaternion multiplication
$\mathtt{i}^2 = \mathtt{j}^2 = \mathtt{k}^2 = \mathtt{ijk} = -1$
& cut it on a stone of this bridge"

4.2.2 Properties

Be p and $q \in \mathbb{H}$ given by

$$p = p_0 + p_1\,\mathtt{i} + p_2\,\mathtt{j} + p_3\,\mathtt{k}, \tag{4.4}$$

$$q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}. \tag{4.5}$$

The addition of quaternions is then defined by

$$p + q = (p_0 + q_0) + (p_1 + q_1)\,\mathtt{i} + (p_2 + q_2)\,\mathtt{j} + (p_3 + q_3)\,\mathtt{k}. \tag{4.6}$$

From equations (4.3), it follows that the multiplication of p and q yields

$$\begin{aligned} pq =& (p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3) + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_3)\,\mathtt{i} \\ &+ (p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1)\,\mathtt{j} + (p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0)\,\mathtt{k}. \end{aligned} \tag{4.7}$$

It can be seen from equations (4.6) and (4.7) that the addition of quaternions is commutative, whereas the multiplication, in general, is not: $p\,q \neq q\,p$. However, quaternion multiplication is still associative, i.e.

$$(p\,q)\,r = p\,(q\,r) \tag{4.8}$$

for any $p, q, r \in \mathbb{H}$, which can be easily shown by direct calculation.
In analogy to the complex numbers, the conjugate of a quaternion $q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}$ is defined by

$$q^* = q_0 - q_1\,\mathtt{i} - q_2\,\mathtt{j} - q_3\,\mathtt{k}. \tag{4.9}$$

By using equation (4.7), it can be shown that the product of q with its conjugate q^* is given by

$$qq^* = q^*q = q_0^2 + q_1^2 + q_2^2 + q_3^2. \tag{4.10}$$

In this way, the norm or absolute value of a quaternion can be defined by

$$|q| = \sqrt{q^*q} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}. \tag{4.11}$$

The inverse value of a quaternion $q \neq 0$ is given by

$$q^{-1} = \frac{q^*}{|q|}. \tag{4.12}$$

The quaternion division is defined in this way, too.
We define the functions $\mathtt{S}(q)$, $\mathtt{I}(q)$, $\mathtt{J}(q)$ and $\mathtt{K}(q)$, which extract the different parts of the quaternion q defined in equation (4.1), by:

$$\mathtt{S}(q) = \frac{q + q^*}{2} = q_0, \tag{4.13}$$

$$\mathtt{I}(q) = \frac{\mathtt{i}\,(q - q^*) + (q - q^*)\,\mathtt{i}}{-4} = q_1, \tag{4.14}$$

$$\mathtt{J}(q) = \frac{\mathtt{j}\,(q - q^*) + (q - q^*)\,\mathtt{j}}{-4} = q_2, \tag{4.15}$$

$$\mathtt{K}(q) = \frac{\mathtt{k}\,(q - q^*) + (q - q^*)\,\mathtt{k}}{-4} = q_3. \tag{4.16}$$

4.2.3 Cayley-Dickson representation

An arbitrary quaternion $q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}$ can be written in the form

$$q = (q_0 + q_1\,\mathtt{i}) + (q_2 + q_3\,\mathtt{i})\,\mathtt{j} = z_0 + z_1\,\mathtt{j}, \tag{4.17}$$

where z_1 and $z_2 \in \mathbb{C}$. In this way, a quaternion can be represented as a complex number with complexified coefficients. This is called the Cayley-Dickson representation.

4.2.4 Quaternion vectors

A quaternion vector $\vec{q} \in \mathbb{H}^N$ is an N-dimensional vector whose elements are quaternions. For two vectors $\vec{q}, \vec{p} \in \mathbb{H}^N$, the inner product is given by (e.g. Miron [2005])

$$\langle \vec{p} \mid \vec{q} \rangle = \vec{q}^{\dagger}\vec{p}, \tag{4.18}$$

where $\vec{q}^{\dagger} = \left(\vec{q}^{T}\right)^{*}$ denotes the conjugate transpose of a quaternion vector.
The norm of a quaternion vector can then be defined by

$$\|\vec{q}\| = \sqrt{\langle \vec{q} \mid \vec{q} \rangle} = \sqrt{\vec{q}^{\dagger}\vec{q}}. \tag{4.19}$$

The set of quaternion-valued vectors forms a Hilbert space[II]. Two quaternion vectors $\vec{q}, \vec{p} \in \mathbb{H}^N$ are orthogonal if their inner product equals zero:

$$\langle \vec{p} \mid \vec{q} \rangle = 0. \tag{4.20}$$

[II] A quaternion Hilbert space H is a complete metric space with respect to the distance function induced by the inner product. This means that for every $\vec{p}, \vec{q}, \vec{r} \in H$ and $\alpha, \beta \in \mathbb{H}$, the following properties are true:

- $(\vec{p} + \vec{q}) \in H$, $(\alpha\vec{p}) \in H$, $\alpha(\beta\vec{p}) = (\alpha\beta)\vec{p}$.
- An inner product $\langle \vec{p} \mid \vec{q} \rangle$ is defined with:
  $\langle \vec{q} \mid \vec{q} \rangle \in \mathbb{R}_0^+$, $\langle \vec{q} \mid \vec{q} \rangle = 0$ iff $\vec{q} = \vec{0}$,
  $\langle \vec{p}\alpha + \vec{q}\beta \mid \vec{r} \rangle = \langle \vec{p} \mid \vec{r} \rangle\alpha + \langle \vec{q} \mid \vec{r} \rangle\beta$
  $\langle \vec{p} \mid \vec{q} \rangle = \langle \vec{q} \mid \vec{p} \rangle^{*}$.
- H is complete, i.e. every Cauchy sequence in H converges in H.

4.2.5 Quaternion matrices

As quaternions are not commutative, the properties of quaternion matrices differ from real or complex matrices. Considering a quaternion matrix $\mathbf{A} \in \mathbb{H}^{M\times N}$, we will denote by $\mathbf{A}^T$ the transpose, by $\mathbf{A}^*$ the conjugate and by $\mathbf{A}^\dagger$ the conjugatee transpose of $\mathbf{A}$. A good overview of quaternion matrices is given by Zhang [1997]. In the following, some results of this article will be presented. In particular, for two quaternion matrices $\mathbf{A} \in \mathbb{H}^{M\times N}$ and $\mathbf{B} \in \mathbb{H}^{N\times O}$, the following properties hold:

1. $(\mathbf{A}^*)^T = (\mathbf{A}^T)^*$
2. $(\mathbf{A}\,\mathbf{B})^\dagger = \mathbf{B}^\dagger\,\mathbf{A}^\dagger$
3. $(\mathbf{A}\,\mathbf{B})^* \neq \mathbf{A}^*\,\mathbf{B}^*$, in general
4. $(\mathbf{A}\,\mathbf{B})^T \neq \mathbf{B}^T\,\mathbf{A}^T$, in general
5. $(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}$, if $\mathbf{A}$ and $\mathbf{B}$ are invertible
6. $(\mathbf{A}^\dagger)^{-1} = (\mathbf{A}^{-1})^\dagger$, if $\mathbf{A}$ is invertible
7. $(\mathbf{A}^*)^{-1} \neq (\mathbf{A}^{-1})^*$, in general
8. $(\mathbf{A}^T)^{-1} \neq (\mathbf{A}^{-1})^T$, in general

Due to the non-commutativity of quaternions, a quaternion matrix $\mathbf{A} \in \mathbb{H}^{N\times N}$ has right and left eigenvalues, satisfying the equations

$$\mathbf{A}\vec{u}_r = \vec{u}_r\lambda_r, \tag{4.21}$$
$$\mathbf{A}\vec{u}_l = \lambda_l\vec{u}_l, \tag{4.22}$$

respectively. In general, both eigenvalues are different. The theory of left eigenvalues is not yet well understood in literature, but the properties of right eigenvalues are well established Zhang [1997]. Therefore, in the following, we will use exclusively right eigenvalues and eigenvectors.
In analogy to the Cayley-Dickson representation of quaternions (equation 4.17), any quaternion matrix $\mathbf{A} \in \mathbb{H}^{N\times N}$ can be written as a combination of two complex matrices $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{H}^{N\times N}$ by $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2\mathtt{j}$, as well as every vector $\vec{u} \in \mathbb{H}^N$ can be written as $\vec{u} = \vec{u}_1 + \vec{u}_2\mathtt{j}$. Now, the eigenvalue equation (4.21) can be written as

$$(\mathbf{A}_1 + \mathbf{A}_2\mathtt{j})(\vec{u}_1 + \vec{u}_2\mathtt{j}) = (\vec{u}_1 + \vec{u}_2\mathtt{j})\lambda \tag{4.23}$$
$$\mathbf{A}_1\vec{u}_1 - \mathbf{A}_2\vec{u}_2^* + (\mathbf{A}_1\vec{u}_2 + \mathbf{A}_2\vec{u}_1^*)\mathtt{j} = \lambda\vec{u}_1 + \lambda^*\vec{u}_2\mathtt{j}, \tag{4.24}$$

which is equivalent to the equations

$$\mathbf{A}_1\vec{u}_1 - \mathbf{A}_2\vec{u}_2^* = \lambda\vec{u}_1, \tag{4.25}$$
$$\mathbf{A}_1\vec{u}_2 + \mathbf{A}_2\vec{u}_1^* = \lambda^*\vec{u}_2. \tag{4.26}$$

Conjugation of equations (4.25) and (4.26) yields

$$\mathbf{A}_1^*\vec{u}_1^* - \mathbf{A}_2^*\vec{u}_2 = \lambda^*\vec{u}_1^*, \tag{4.27}$$
$$\mathbf{A}_1^*\vec{u}_2^* + \mathbf{A}_2^*\vec{u}_1 = \lambda\vec{u}_2^*. \tag{4.28}$$

Finally, equations (4.25) and (4.28) can be combined in matrix form as

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ -\mathbf{A}_2^* & \mathbf{A}_1^* \end{pmatrix} \begin{pmatrix} \vec{u}_1 \\ -\vec{u}_2^* \end{pmatrix} = \lambda \begin{pmatrix} \vec{u}_1 \\ -\vec{u}_2^* \end{pmatrix}, \tag{4.29}$$

and equations (4.26) and (4.27) combine to

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ -\mathbf{A}_2^* & \mathbf{A}_1^* \end{pmatrix} \begin{pmatrix} \vec{u}_2 \\ \vec{u}_1^* \end{pmatrix} = \lambda^* \begin{pmatrix} \vec{u}_2 \\ \vec{u}_1^* \end{pmatrix}. \tag{4.30}$$

The matrix $\chi_A = \left(\begin{smallmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ -\mathbf{A}_2^* & \mathbf{A}_1^* \end{smallmatrix}\right) \in \mathbb{C}^{2N\times 2N}$ is called adjoint matrix of $\mathbf{A}$.
χ_A has $2N$ eigenvectors with $2N$ associated (complex) eigenvalues. According to equations (4.29) and (4.30), if λ is an eigenvalue of χ_A, λ^* is also an eigenvalue. This means that the eigenvalues of χ_A occur in conjugated pairs. On the other side, according to equations (4.29) and (4.30), every eigenvalue of χ_A is also an eigenvalue of $\mathbf{A}$.
If λ is an eigenvalue of $\mathbf{A}$, then $\rho^{-1}\lambda\rho$ is an eigenvalue as well [Brenner, 1951]: As $\mathbf{A}\vec{u} = \vec{u}\lambda$, it follows that $\mathbf{A}(\vec{u}\rho) = (\vec{u}\rho)\rho^{-1}\lambda\rho$. Therefore, $(\vec{u}\rho)$ is an eigenvector corresponding to the eigenvalue $\rho^{-1}\lambda\rho$.
Notably, setting $\rho = \mathtt{j}$ or $\rho = \mathtt{k}$ yields $\rho^{-1}\lambda\rho = \lambda^*$. Therefore, if $\vec{u}_\lambda$ and $\vec{u}_{\lambda^*}$ are the eigenvectors corresponding to λ and λ^*, respectively, it follows that $\vec{u}_\lambda\mathtt{j}$ and $\vec{u}_\lambda\mathtt{k}$ are also eigenvectors of $\mathbf{A}$. In this way, an infinite number of eigenvalue and eigenvector pairs of $\mathbf{A}$ can be generated based on a single (complex) eigenvalue λ_0 by using any quaternion-valued ρ, which yields $\lambda' = \rho^{-1}\lambda\rho$ and $\vec{u}' = \vec{u}\rho$. However, all these eigenvectors are not independent among each other and form a class of eigenvalues. Notably, the eigenvectors $\vec{u}_\lambda$ and $\vec{u}_{\lambda^*}$ of equations (4.29) and (4.30) are not linearly independent and λ and λ^* correspond to the same eigenvalue class of $\mathbf{A}$. If χ_A has real eigenvalues, however, the application of $\rho^{-1}\lambda\rho$ does not change λ and consequently, this eigenvalue of χ_A occurs twice.
The matrix $\mathbf{A}$ has only N linearly independent eigenvectors. By taking one eigenvalue of χ_A per eigenvalue class, the corresponding eigenvectors $\vec{u} = \vec{u}_1 + \vec{u}_2\mathtt{j}$ of $\mathbf{A}$ can be obtained by identifying the eigenvector of χ_A as indicated in equation (4.29).
In the following application to quaternion-MUSIC, only Hermitian matrices are needed, i.e. matrices with $\mathbf{A}^\dagger = \mathbf{A}$. Be $\lambda \in \mathbb{C}$ an eigenvalue of a Hermitian quaternion matrix $\mathbf{A} \in \mathbb{H}^{\mathbf{N}\times\mathbf{N}}$ and $\vec{u} \in \mathbb{H}^N$ the associated normed eigenvector ($\vec{u}^\dagger\vec{u} = 1$). Then the eigenvalue λ is real-valued, as can be shown by the short calculation:

$$(\mathbf{A}\vec{u})^\dagger\vec{u} = (\vec{u}\lambda)^\dagger\vec{u} = \lambda^*\vec{u}^\dagger\vec{u} = \lambda^* \tag{4.31}$$
$$(\mathbf{A}\vec{u})^\dagger\vec{u} = \vec{u}^\dagger\mathbf{A}\vec{u} = \vec{u}^\dagger\vec{u}\lambda = \lambda. \tag{4.32}$$

Therefore, $\lambda^* = \lambda$ and $\lambda \in \mathbb{R}$.

If we consider two different eigenvalues $\lambda_1 \neq \lambda_2$ of a Hermitian matrix $\mathbf{A}$ with associated eigenvectors $\vec{u}_1$ and $\vec{u}_2$. These eigenvectors are orthogonal because:

$$(\mathbf{A}\vec{u}_1)^\dagger \vec{u}_2 = (\vec{u}_1\lambda_1)^\dagger \vec{u}_2 = \lambda_1(\vec{u}_1^\dagger \vec{u}_2) \tag{4.33}$$

$$(\mathbf{A}\vec{u}_1)^\dagger \vec{u}_2 = \vec{u}_1^\dagger \mathbf{A}\vec{u}_2 = (\vec{u}_1^\dagger \vec{u}_2)\lambda_2, \tag{4.34}$$

which can only be true, if $\vec{u}_1^\dagger \vec{u}_2 = 0$. Therefore, different eigenvectors of $\mathbf{A}$ are orthogonal and the set of eigenvectors of $\mathbf{A}$ can be used to form an orthonormal basis of the N-dimensional quaternion Hilbert space.

4.3 Biquaternions

4.3.1 Definition

In a way similar to the Cayley-Dickson representation of quaternions, where a quaternion is written as a complex number with complex coefficients, it is possible to define quaternions with complex coefficients (or complex numbers with quaternion coefficients). This numbers are called biquaternions or complex quaternions and have been introduced by Hamilton [1853]. Other hypercomplex numbers exists, notably the octonions [Kantor and Solodovnikov, 1989, Ward, 1997], but those will not be discussed here.
We denote by $\mathbb{H}_\mathbb{C}$ the set of biquaternions. Any biquaternion b can be written in the form

$$b = b_0 + b_1\,\mathtt{i} + b_2\,\mathtt{j} + b_3\,\mathtt{k}, \tag{4.35}$$

where $b_{0,\ldots,3} \in \mathbb{C}$ and $\mathtt{i}$, $\mathtt{j}$ and $\mathtt{k}$ are the quaternion imaginary units (see section 4.2). Different representations of biquaternions exist, which are equivalent to the given one, for example by writing a biquaternion as a complex number with quaternion-valued parameters:

$$\begin{aligned} b = &[\Re(b_0) + \Re(b_1)\,\mathtt{i} + \Re(b_2)\,\mathtt{j} + \Re(b_3)\,\mathtt{k}] \\ &+ i\,[\Im(b_0) + \Im(b_1)\,\mathtt{i} + \Im(b_2)\,\mathtt{j} + \Im(b_3)\,\mathtt{k}]\,, \end{aligned} \tag{4.36}$$

where i is the complex imaginary unit. The complex imaginary unit i commutes with the quaternion imaginary units $\mathtt{i}$, $\mathtt{j}$ and $\mathtt{k}$, i.e. $i\,\mathtt{i} = \mathtt{i}\,i$, $i\,\mathtt{j} = \mathtt{j}\,i$ and $i\,\mathtt{k} = \mathtt{k}\,i$.

4.3.2 Properties

As biquaternions are made of complex numbers and quaternions, different ways exist to define conjugates of biquaternions. In this way, it is possible [Tian, 2000] to define the dual conjugate by

$$b^d = b_0 - b_1\,\mathtt{i} - b_2\,\mathtt{j} - b_3\,\mathtt{k}, \tag{4.37}$$

and the complex conjugate by

$$b^c = b_0^* + b_1^*\,\mathtt{i} + b_2^*\,\mathtt{j} + b_3^*\,\mathtt{k}, \tag{4.38}$$

where $b^*_{0,\ldots,3}$ are the complex conjugates of the coefficients of b. Finally, the simultaneous application of the dual and the complex conjugate yields the Hermitian conjugate

$$b^* = (b^d)^c = b_0^* - b_1^* \, \mathtt{i} - b_2^* \, \mathtt{j} - b_3^* \, \mathtt{k}. \tag{4.39}$$

Applying any of these conjugates twice to a biquaternion b yields the original biquaternion.
As for quaternions, the multiplication of biquaternions is not commutative, but is still associative, i.e. for $b, p, q \in \mathbb{H}_{\mathbb{C}}$, $(b\ p)\ q = b\ (p\ q)$[III].
The inner product of biquaternions has to be defined in a different way than for quaternions (equation (4.18)). It can be shown that the inner product defined by

$$\langle b \mid p \rangle = \mathtt{S}(p^* b) \tag{4.40}$$

satisfies all requirements for an Hermitian inner product [Ward, 1997]. The function $\mathtt{S}(b)$ which extracts the (complex-valued) scalar part of a biquaternion b can be defined by

$$\mathtt{S}(b) = \frac{b + b^d}{2} = b_0. \tag{4.41}$$

Then, the inner product of two biquaternions

$$\begin{aligned} b &= b_0 + b_1 \, \mathtt{i} + b_2 \, \mathtt{j} + b_3 \, \mathtt{k} \\ p &= p_0 + p_1 \, \mathtt{i} + p_2 \, \mathtt{j} + p_3 \, \mathtt{k} \end{aligned}$$

is given by

$$\langle b \mid p \rangle = p_0^* b_0 + p_1^* b_1 + p_2^* b_2 + p_3^* b_3. \tag{4.42}$$

Using the inner product, the absolute value of a biquaternion b can be defined by

$$|b| = \sqrt{\langle b \mid b \rangle} = \sqrt{b_0^* b_0 + b_1^* b_1 + b_2^* b_2 + b_3^* b_3}. \tag{4.43}$$

It can be shown [Ward, 1997] that, in general, for two biquaternions p and q,

$$|pq| \neq |p||q|. \tag{4.44}$$

Therefore, the biquaternion algebra is not normed[IV].

4.3.3 Biquaternion vectors

In analogy to quaternion vectors, a biquaternion vector $\vec{b} \in \mathbb{H}_{\mathbb{C}}^N$ can be defined as an N-dimensional vector with biquaternion values. Different definitions of the inner product of two biquaternion vectors $\vec{b}, \vec{p} \in \mathbb{H}_{\mathbb{C}}^N$ exist. In the following, we will define it as

$$\langle \vec{b} \mid \vec{p} \rangle = \mathtt{S}\left(\vec{p}^{\dagger} \vec{b}\right), \tag{4.45}$$

[III] For octonions, the multiplication is not associative. For biquaternions, the associativity is due to the fact that the complex imaginary unit commutes with any of the quaternion imaginary units.
[IV] The octonions, in contrast, constitute a normed eight-dimensional algebra.

where $\vec{p}^{\dagger} = \left(\vec{p}^{T}\right)^{*}$ denotes the Hermitian conjugate transpose of a biquaternion vector. The biquaternion vectors $\vec{b}$ and $\vec{p}$ are orthogonal, if

$$\langle \vec{b} \mid \vec{p} \rangle = 0. \tag{4.46}$$

Consequently, the norm of a biquaternion vector can be defined by

$$\|\vec{b}\| = \sqrt{\langle \vec{b} \mid \vec{b} \rangle} = \sqrt{\mathbb{S}(\vec{b}^{\dagger}\vec{b})}. \tag{4.47}$$

Defining the inner product by equation (4.45), the set of biquaternion vectors forms a Hilbert space (see footnote [II] on page 97), with the restriction that the equation

$$\langle \vec{p}\alpha + \vec{q}\beta \mid \vec{r} \rangle = \langle \vec{p} \mid \vec{r} \rangle \alpha + \langle \vec{q} \mid \vec{r} \rangle \beta \tag{4.48}$$

is fulfilled for $\alpha, \beta \in \mathbb{C}$ only.

4.3.4 Biquaternion matrices

Be $\mathbf{A} \in \mathbb{H}_{\mathbb{C}}^{\mathbf{M}\times\mathbf{N}}$ a biquaternion-valued matrix. Then, we denote by A^T the transpose, by A^* the Hermitian conjugate and by $A^{\dagger}$ the Hermitian conjugate transpose of $\mathbf{A}$. Then, the following equations hold [Tian, 2000]:

1. $(\mathbf{A}^{\dagger})^{\dagger} = \mathbf{A}$.
2. $(\mathbf{AB})^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger}$.
3. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, if $\mathbf{A}$ and $\mathbf{B}$ are invertible.
4. $(\mathbf{A}^{\dagger})^{-1} = (\mathbf{A}^{-1})^{\dagger}$, if $\mathbf{A}$ is invertible.

As for quaternion matrices, biquaternion matrices have left and right eigenvectors. However, we will focus on the right ones only for reasons similar to those mentioned in the case of quaternion matrices. A biquaternion matrix $\mathbf{A} \in \mathbb{H}_{\mathbb{C}}^{\mathbf{N}\times\mathbf{N}}$ can be written as a sum of two quaternion matrices $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{H}^{N\times N}$ by $\mathbf{A} = \mathbf{A}_1 + i\,\mathbf{A}_2$. In the same way, any vector $\vec{u} \in \mathbb{H}_{\mathbb{C}}^{N}$ can be written as $\vec{u} = \vec{u}_1 + i\,\vec{u}_2$ with $\vec{u}_1, \vec{u}_2 \in \mathbb{H}^{N}$. Exploiting the fact that the complex imaginary unit i commutes with any of the quaternion imaginary units `i`, `j` and `k`, the (right) eigenvalue equation of $\mathbf{A}$ can be written as

$$\begin{aligned} \mathbf{A}\vec{u} &= \vec{u}\lambda, \\ (\mathbf{A}_1 + i\,\mathbf{A}_2)\,(\vec{u}_1 + i\,\vec{u}_2) &= (\vec{u}_1 + i\,\vec{u}_2)\lambda \\ \mathbf{A}_1\vec{u}_1 - \mathbf{A}_2\vec{u}_2 + i\,\mathbf{A}_1\vec{u}_2 + i\,\mathbf{A}_2\vec{u}_1 &= \vec{u}_1\lambda + i\,\vec{u}_2\lambda, \end{aligned} \tag{4.49}$$

which is equivalent to the equations

$$\mathbf{A}_1\vec{u}_1 - \mathbf{A}_2\vec{u}_2 = \vec{u}_1\lambda, \tag{4.50}$$

$$\mathbf{A}_1\vec{u}_2 + \mathbf{A}_2\vec{u}_1 = \vec{u}_2\lambda. \tag{4.51}$$

Equations (4.50) and (4.51) can be written in matrix form:

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ -\mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \begin{pmatrix} \vec{u}_1 \\ -\vec{u}_2 \end{pmatrix} = \begin{pmatrix} \vec{u}_1 \\ -\vec{u}_2 \end{pmatrix} \lambda. \tag{4.52}$$

Therefore, the eigenvalue problem of an $N \times N$ biquaternion matrix is equivalent to the eigenvalue problem of a $2N \times 2N$ quaternion matrix.
The matrix $\chi_A = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ -\mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \in \mathbb{H}^{2N \times 2N}$ is called biquaternion adjoint matrix of $\mathbf{A}$.
χ_A has $2N$ eigenvectors with $2N$ associated (quaternion-valued) eigenvalues. According to equation (4.52), every eigenvalue of χ_A is also an eigenvalue of $\mathbf{A}$. The $2N$ different eigenvalues of a $2N \times 2N$ quaternion matrix are independent among each other (see 4.2.5). Therefore, any $N \times N$ biquaternion matrix has $2N$ independent eigenvectors. For a given biquaternion matrix, the eigenvalues and eigenvectors are therefore retrieved by building the biquaternion adjoint matrix, calculating its eigenvalues and eigenvectors and transforming the (quaternion) eigenvectors of χ_A to (biquaternion) eigenvectors of $\mathbf{A}$.
An important case of biquaternion matrices are Hermitian biquaternion matrices. Actually, the biquaternion adjoint matrix of a Hermitian biquaternion matrix $\mathbf{A}$ is itself Hermitian. $\mathbf{A}^\dagger$ can be written as

$$\begin{aligned} \mathbf{A}^\dagger &= (\mathbf{A}_1 + i\mathbf{A}_2)^\dagger, \\ &= \mathbf{A}_1^\dagger - i\mathbf{A}_2^\dagger. \end{aligned} \tag{4.53}$$

Therefore, $\mathbf{A}^\dagger = \mathbf{A}$ iff

$$\mathbf{A}_1^\dagger = \mathbf{A}_1, \tag{4.54}$$

$$\mathbf{A}_2^\dagger = -\mathbf{A}_2. \tag{4.55}$$

Consequently,

$$\chi_A^\dagger = \begin{pmatrix} \mathbf{A}_1^\dagger & -\mathbf{A}_2^\dagger \\ \mathbf{A}_2^\dagger & \mathbf{A}_1^\dagger \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ -\mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} = \chi_A. \tag{4.56}$$

A Hermitian quaternion matrix has real eigenvalues and consequently, the eigenvalues of a Hermitian biquaternion matrix are also real-valued.
Furthermore, it can be shown that different eigenvectors of a Hermitian biquaternion matrix $\mathbf{A}$ are orthogonal. If λ_1 and λ_2 denote two different eigenvalues ($\lambda_1 \neq \lambda_2$) of $\mathbf{A}$ with associated eigenvectors $\vec{u}_1$ and $\vec{u}_2$, then

$$(\mathbf{A}\vec{u}_1)^\dagger \vec{u}_2 = (\vec{u}_1\lambda_1)^\dagger \vec{u}_2 = \lambda_1(\vec{u}_1^\dagger \vec{u}_2), \tag{4.57}$$

$$(\mathbf{A}\vec{u}_1)^\dagger \vec{u}_2 = \vec{u}_1^\dagger \mathbf{A}\vec{u}_2 = (\vec{u}_1^\dagger \vec{u}_2)\lambda_1, \tag{4.58}$$

which can only be true if $\vec{u}_1^\dagger \vec{u}_2 = 0$. Therefore, the $2N$ eigenvectors of $\mathbf{A}$ can be used to form an orthonormal basis of an N-dimensional biquaternion Hilbert space.

4.4 Quaternion-MUSIC

4.4.1 Methodology

For "classical" MUSIC, only monocomponent signals can be used. The data matrices in equation (3.11) are complex. When working with multidimensional data, there are two possibilities. Either some components of the signal are disregarded or the data of the different components are stocked in a single long vector (the long-vector approach) and processed together. In this case, the calculation of the covariance matrix in equation (3.13) is equivalent to the summation of the covariance matrices of the different components. However, the polarization parameters of the source signal are not retrieved.
In order to overcome this problem, Miron et al. [2006] proposed the MUSIC version using quaternions which is presented in the following. In this way, the polarization properties of incident waves can be estimated directly. By replacing a complex data matrix by a quaternion-valued one, two complex-valued signals can be stocked and processed at the same time. The reasoning for the algorithm presented in the following is analog to the classical MUSIC case.
We consider an array of N two-component seismic sensors with K incident waves. The respective signals of the sensors are $\vec{X}_i(f) \in \mathbb{C}^{N\times 1}$ with $i \in \{1,2\}$. The parameters i indicate the two different components of the signal. The quaternion-valued signal vector is defined as

$$\vec{X}_q(f) = \Re(\vec{X}_1(f)) + \mathtt{i}\cdot\Re(\vec{X}_2(f)) + \mathtt{j}\Im(\vec{X}_1(f)) + \mathtt{k}\cdot\Im(\vec{X}_2(f)), \tag{4.59}$$

where $\Re(x)$ yields the real part and $\Im(x)$ the imaginary part of a complex number x.
The vectors $\vec{a}(f,\vec{k})$ in equation (3.12) are transformed into quaternion-valued vectors in the following way:

$$\vec{a}_q(\vec{k},\rho,\varphi) = \frac{1}{\sqrt{N}}\left[\cos\rho + \mathtt{i}\sin\rho\cdot\exp(\mathtt{j}\varphi)\right]\cdot\exp(-\mathtt{j}\mathbf{R}\cdot\vec{k}). \tag{4.60}$$

ρ (with $0 \le \rho \le 90°$) parameterizes the amplitude ratio between both components and φ (with $0 \le \varphi < 360°$) their phase difference. For a Rayleigh wave, for example, if $\vec{X}_1(f)$ is the vertical component of the signal and $\vec{X}_2(f)$ the horizontal one, the ellipticity is given by $\epsilon = \sin\rho/\cos\rho = \tan\rho$ and the phase shift is equal to 90° or 270°, depending on whether the motion is retrograde or prograde. In this way, the equation corresponding to equation (3.11) is given by

$$\vec{X}_q(f) = \mathbf{A}_q(f)\cdot\vec{F}_q(f) + \vec{W}_q(f), \tag{4.61}$$

where the matrix $\mathbf{A_q}(f)$ is composed of the vectors defined in equation (4.60) and the vectors $\vec{F}_q(f)$ and $\vec{W}_q(f)$ are defined in an analog way to $\vec{X}_q(f)$ in equation (4.59).

The quaternion covariance matrix (of size $N \times N$) is now built by

$$\mathbf{S}_q = \mathsf{E}\left(\vec{X}_q \cdot \vec{X}_q^{\dagger}\right) = \mathbf{A}_q \cdot \mathsf{E}(\vec{F}_q \vec{F}_q^{\dagger}) \mathbf{A}_q^{\dagger} + \mathsf{E}(\vec{W}_q \vec{W}_q^{\dagger}), \tag{4.62}$$

where $\mathsf{E}(x)$ indicates the expectation value and the non-correlation of $\vec{F}_q$ and $\vec{W}_q$ is used. By construction, $\mathbf{S}_q$ is Hermitian (because $\left(\vec{X}_q \cdot \vec{X}_q^{\dagger}\right)^{\dagger} = \vec{X}_q \cdot \vec{X}_q^{\dagger}$) and its eigenvalues are real-valued. In a completely similar way to the classical MUSIC case, the eigenvectors and eigenvalues of $\mathbf{S}_q$ are calculated. The largest K eigenvalues correspond to the signal subspace and the smaller ones to the noise subspace. The noise subspace $\mathbf{G}_q \in \mathbb{H}^{N \times (N-K)}$ is then built by:

$$\mathbf{G}_q = (\vec{u}_1, \vec{u}_2, \ldots \vec{u}_{N-K}). \tag{4.63}$$

The properties of the sources are retrieved by searching the $\vec{a}_q(\vec{k}, \rho, \varphi)$ defined by equation (4.60) which minimizes the distance to the signal subspace. The squared distance to the signal subspace can be expressed in an analog way to the classical case by

$$\begin{aligned} d_q^2 &= \|\mathbf{G}_q^{\dagger}\, \vec{a}_q(\vec{k}, \rho, \varphi)\|^2 \\ &= \left(\mathbf{G}_q^{\dagger}\, \vec{a}_q(\vec{k}, \rho, \varphi)\right)^{\dagger} \mathbf{G}_q^{\dagger}\, \vec{a}_q(\vec{k}, \rho, \varphi) \\ &= \vec{a}_q^{\dagger}(\vec{k}, \rho, \varphi)\; \mathbf{G}_q \mathbf{G}_q^{\dagger}\, \vec{a}_q(\vec{k}, \rho, \varphi). \end{aligned} \tag{4.64}$$

The parameters $\vec{k}$, ρ and φ maximizing the quaternion-MUSIC functional

$$P_q(\vec{k}, \rho, \varphi) = \frac{1}{\vec{a}_q^{\dagger}(\vec{k}, \rho, \varphi)\; \mathbf{G}_q \mathbf{G}_q^{\dagger}\, \vec{a}_q(\vec{k}, \rho, \varphi)} \tag{4.65}$$

indicate the properties of the incident waves. In general, these properties have to be found by a grid search along four dimensions corresponding to azimuth, slowness and the polarization parameters ρ and φ. A four-dimensional grid search is a lot more time-consuming than a two-dimensional grid search[V]. However, the wave vector of an incident wave can be estimated without knowing the polarization. This would separate the problem in two smaller problems: First, the estimation of the wave vector and second, the estimation of polarization. Both problems can be solved by a two-dimensional grid-search. Actually, we found out that even the grid search for the polarization parameters is not necessary as they can be determined in an analytical way as will be shown in the following.

[V] For example, if the grid search covers 100 different values for each parameter, a two-dimensional search comprises $100^2 = 10^4$ data points, whereas a four-dimensional grid search comprises $100^4 = 10^8$ data points. In this example, a four-dimensional grid search is therefore about 10 000 times more time-consuming than a two-dimensional grid search.

4.4.2 Polarization parameter identification for known wave vector

If the azimuth and slowness values are already known, the polarization parameters ρ and φ can be identified as follows.
The steering vector is given by equation (4.60):

$$\vec{a}_q(\vec{k},\rho,\varphi) = \frac{1}{\sqrt{N}}\left[\cos\rho + \mathtt{i}\sin\rho\cdot\exp(\mathtt{j}\varphi)\right]\cdot\exp(-\mathtt{j}\mathbf{R}\cdot\vec{k}). \tag{4.66}$$

As slowness and azimuth are known, $\exp(-\mathtt{j}\mathbf{R}\cdot\vec{k})$ is a constant when searching for ρ and φ. Maximization of equation (4.65) corresponds to a minimization of its denominator

$$D = \vec{a}_q^{\dagger}(\vec{k},\rho,\varphi)\ \mathbf{G}_q\mathbf{G}_q^{\dagger}\ \vec{a}_q(\vec{k},\rho,\varphi). \tag{4.67}$$

Inserting equation (4.66) into equation (4.67) and setting $\vec{\delta} = \mathbf{R}\cdot\vec{k}$, D can be calculated by:

$$\begin{aligned}
D &= \frac{1}{N}\cdot\left[(\cos\rho + \mathtt{i}\sin\rho\cdot e^{\mathtt{j}\varphi})\cdot e^{-\mathtt{j}\vec{\delta}}\right]^{\dagger}\cdot\mathbf{G}_q\mathbf{G}_q^{\dagger}\cdot\left[(\cos\rho + \mathtt{i}\sin\rho\cdot e^{\mathtt{j}\varphi})\cdot e^{-\mathtt{j}\vec{\delta}}\right]\\
&= \frac{1}{N}\cdot e^{\mathtt{j}\vec{\delta}^{\dagger}}\left[\cos\rho - e^{-\mathtt{j}\varphi}\mathtt{i}\sin\rho\right]\cdot\mathbf{G}_q\mathbf{G}_q^{\dagger}\cdot\left[(\cos\rho + \mathtt{i}\sin\rho\cdot e^{\mathtt{j}\varphi})\cdot e^{-\mathtt{j}\vec{\delta}}\right]\\
&= \frac{1}{N}\cdot\left[\cos\rho\cdot e^{\mathtt{j}\vec{\delta}^{\dagger}} - \mathtt{i}e^{\mathtt{j}\varphi}\sin\rho\cdot e^{-\mathtt{j}\vec{\delta}^{\dagger}}\right]\cdot\mathbf{G}_q\mathbf{G}_q^{\dagger}\cdot\left[e^{-\mathtt{j}\vec{\delta}}\cos\rho + e^{\mathtt{j}\vec{\delta}}e^{-\mathtt{j}\varphi}\mathtt{i}\sin\rho\right]\\
&= \frac{1}{N}\cdot\left[\cos^2\rho\cdot(e^{\mathtt{j}\vec{\delta}^{\dagger}}\ \mathbf{G}_q\mathbf{G}_q^{\dagger}\ e^{-\mathtt{j}\vec{\delta}}) - \mathtt{i}e^{\mathtt{j}\varphi}\sin\rho\cdot(e^{-\mathtt{j}\vec{\delta}^{\dagger}}\ \mathbf{G}_q\mathbf{G}_q^{\dagger}\ e^{\mathtt{j}\vec{\delta}})e^{-\mathtt{j}\varphi}\mathtt{i}\sin\rho\right.\\
&\qquad \left. + \cos\rho\cdot(e^{\mathtt{j}\vec{\delta}^{\dagger}}\ \mathbf{G}_q\mathbf{G}_q^{\dagger}\ e^{\mathtt{j}\vec{\delta}})e^{-\mathtt{j}\varphi}\mathtt{i}\sin\rho - \mathtt{i}e^{\mathtt{j}\varphi}\sin\rho\cdot(e^{-\mathtt{j}\vec{\delta}^{\dagger}}\ \mathbf{G}_q\mathbf{G}_q^{\dagger}\ e^{-\mathtt{j}\vec{\delta}})\cos\rho\right].
\end{aligned} \tag{4.68}$$

Here, we used the properties

$$\begin{aligned}
\exp(\mathtt{j}x)\cdot\mathtt{i} &= (\cos x + \mathtt{j}\sin x)\cdot\mathtt{i}\\
&= \mathtt{i}\cdot(\cos x - \mathtt{j}\sin x)\\
&= \mathtt{i}\cdot\exp(-\mathtt{j}x)
\end{aligned} \tag{4.69}$$

and

$$\begin{aligned}
\exp(\mathtt{j}x)\cdot\exp(\mathtt{j}y) &= (\cos x + \mathtt{j}\sin x)\cdot(\cos y + \mathtt{j}\sin y)\\
&= \cos x\cos y + \mathtt{j}\sin y\cos x + \mathtt{j}\sin x\cos y - \sin x\sin y\\
&= (\cos y + \mathtt{j}\sin y)\cdot(\cos x + \mathtt{j}\sin x)\\
&= \exp(\mathtt{j}y)\cdot\exp(\mathtt{j}x).
\end{aligned} \tag{4.70}$$

The terms $e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{-\mathtt{j}\vec{\delta}}$ and $e^{-\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}}$ of the two first summands of equation (4.68) are real-valued:

$$\begin{aligned} e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{-\mathtt{j}\vec{\delta}} &= \left(\mathbf{G}_q^\dagger e^{-\mathtt{j}\vec{\delta}}\right)^\dagger \cdot \left(\mathbf{G}_q^\dagger e^{-\mathtt{j}\vec{\delta}}\right) \\ &= \left|\mathbf{G}_q^\dagger e^{-\mathtt{j}\vec{\delta}}\right|^2, \end{aligned} \tag{4.71}$$

which is the sum of the squared absolute values of the components of a vector and therefore real-valued. The reasoning for $e^{-\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}}$ is analog. For the last two summands of equation (4.68), it can be shown that

$$\left(e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}} e^{-\mathtt{j}\varphi\mathtt{i}}\right)^\dagger = -\mathtt{i}e^{\mathtt{j}\varphi} e^{-\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{-\mathtt{j}\vec{\delta}}. \tag{4.72}$$

Using the properties of equations (4.71) and (4.72), equation (4.68) can be simplified:

$$\begin{aligned} D = \frac{1}{N} \cdot [&\cos^2\rho \cdot (e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{-\mathtt{j}\vec{\delta}}) + \sin^2\rho \cdot (e^{-\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}}) \\ &+ 2\sin\rho\cos\rho \cdot \Re(e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}} e^{-\mathtt{j}\varphi\mathtt{i}})] \end{aligned} \tag{4.73}$$

By substituting

$$A = e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{-\mathtt{j}\vec{\delta}}, \tag{4.74}$$

$$B = e^{-\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}}, \tag{4.75}$$

$$C = e^{\mathtt{j}\vec{\delta}^\dagger}\ \mathbf{G}_q\mathbf{G}_q^\dagger\ e^{\mathtt{j}\vec{\delta}}, \tag{4.76}$$

equation (4.73) can be written as

$$D = \frac{1}{N} \cdot [\cos^2\rho \cdot A + \sin^2\rho \cdot B + 2\sin\rho\cos\rho \cdot \Re(Ce^{-\mathtt{j}\varphi\mathtt{i}})]. \tag{4.77}$$

The parameters ρ and φ minimizing D can be found by searching the roots of the derivatives $\frac{\partial D}{\partial \rho}$ and $\frac{\partial D}{\partial \varphi}$. This leads to

$$\begin{aligned} \frac{\partial D}{\partial \rho} &= \frac{1}{N}[-2\sin\rho\cos\rho \cdot A + 2\sin\rho\cos\rho \cdot B + 2(\cos^2\rho - \sin^2\rho) \cdot \Re(Ce^{-\mathtt{j}\varphi\mathtt{i}})] \\ 0 &= \sin\rho\cos\rho \cdot (B - A) + (\cos^2\rho - \sin^2\rho) \cdot \Re(Ce^{-\mathtt{j}\varphi\mathtt{i}}), \end{aligned} \tag{4.78}$$

$$\begin{aligned} \frac{\partial D}{\partial \varphi} &= \frac{1}{N}[2\sin\rho\cos\rho \cdot \Re(Ce^{-\mathtt{j}\varphi}(-\mathtt{j})\mathtt{i})] \\ 0 &= \sin\rho\cos\rho \cdot \Re(Ce^{-\mathtt{j}\varphi}\mathtt{k}). \end{aligned} \tag{4.79}$$

Equation (4.79) could be trivially solved by $\sin\rho = 0$ or $\cos\rho = 0$.

Then, one of both components of the motion would disappear and the phase difference φ would not make any sense. The general solution is therefore given by

$$\begin{aligned} \Re(Ce^{-\mathtt{j}\varphi}\mathtt{k}) &= 0 \\ \Re(C(\mathtt{k}\cos\varphi - \mathtt{i}\sin\varphi)) &= 0 \\ -\mathtt{K}(C)\cdot\cos\varphi + \mathtt{I}(C)\cdot\sin\varphi &= 0 \\ \tan\varphi = \frac{\mathtt{K}(C)}{\mathtt{I}(C)} &= \frac{\mathtt{K}(e^{\mathtt{j}\vec{\delta}^{\dagger}}\,\mathbf{G}_q\mathbf{G}_q^{\dagger}\,e^{\mathtt{j}\vec{\delta}})}{\mathtt{I}(e^{\mathtt{j}\vec{\delta}^{\dagger}}\,\mathbf{G}_q\mathbf{G}_q^{\dagger}\,e^{\mathtt{j}\vec{\delta}})}, \end{aligned} \tag{4.80}$$

which has two solutions φ_1 and φ_2 (with $0 \leq \varphi_{1,2} < 360^\circ$ and $\varphi_2 = \varphi_1 + 180^\circ$) corresponding to the minimum and maximum of D. The φ corresponding to the actual solution can be determined by considering the second derivative of D with respect to φ:

$$\begin{aligned} \frac{\partial^2 D}{\partial\varphi^2} &= \frac{2}{N}\sin\rho\cos\rho\cdot\Re(Ce^{-\mathtt{j}\varphi}(-\mathtt{j})\mathtt{k}) \\ &= -\frac{2}{N}\sin\rho\cos\rho\cdot\Re(Ce^{-\mathtt{j}\varphi}\mathtt{i}). \end{aligned} \tag{4.81}$$

As $e^{-\mathtt{j}\varphi_2} = e^{-\mathtt{j}(\varphi_1+\pi)} = -e^{-\mathtt{j}\varphi_1}$, one φ belongs to a maximum and the other to a minimum of D. The φ leading to a positive $\frac{\partial^2 D}{\partial\varphi^2}$ indicates the minimum of D.

Equation (4.78) can be solved by using the relationships $\sin\rho\cos\rho = \frac{1}{2}\sin(2\rho)$ and $\cos^2\rho - \sin^2\rho = \cos(2\rho)$:

$$\begin{aligned} \sin(2\rho)\cdot(B-A) + 2\cos(2\rho)\cdot\Re(Ce^{-\mathtt{j}\varphi}\mathtt{i}) &= 0 \\ \tan(2\rho) &= \frac{-2\Re(Ce^{-\mathtt{j}\varphi}\mathtt{i})}{B-A} \\ \tan(2\rho) &= \frac{-2\Re(e^{\mathtt{j}\vec{\delta}^{\dagger}}\,\mathbf{G}_q\mathbf{G}_q^{\dagger}\,e^{\mathtt{j}\vec{\delta}}e^{-\mathtt{j}\varphi}\mathtt{i})}{e^{-\mathtt{j}\vec{\delta}^{\dagger}}\,\mathbf{G}_q\mathbf{G}_q^{\dagger}\,e^{\mathtt{j}\vec{\delta}} - e^{\mathtt{j}\vec{\delta}^{\dagger}}\,\mathbf{G}_q\mathbf{G}_q^{\dagger}\,e^{-\mathtt{j}\vec{\delta}}}. \end{aligned} \tag{4.82}$$

By inserting the φ determined by equations (4.80) and (4.81) into equation (4.82), the parameter ρ (with $0 \leq \rho \leq 90^\circ$) is unambiguously determined.

4.5 Biquaternion-MUSIC

4.5.1 Methodology

As MUSIC treats one-component signals and quaternion-MUSIC two-dimensional signals, the next logical step would be a method which can treat more-dimensional signals. In fact, Miron [2005] proposed the method capable of working with four-dimensional data which is presented in the following. In this case, the data matrix would be built of biquaternion values. As our seismic signals have four components, the scalar part of the data vector will be zero and the data vector $\vec{X}_{biq}(f) \in \mathbb{H}_{\mathbb{C}}^{N\times 1}$ is built in the following way:

$$\vec{X}_{biq}(f) = \vec{X}_1(f)\mathtt{i} + \vec{X}_2(f)\mathtt{j} + \vec{X}_3(f)\mathtt{k}, \tag{4.83}$$

where $\vec{X}_i(f) \in \mathbb{C}^{N\times 1}$ are the signals of the different components (indicated by $i \in 1,2,3$) for all seismic sensors.
In analogy to equation (4.60), the biquaternion steering vector is given as

$$\begin{aligned}\vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2) = \frac{1}{\sqrt{N}}[&\cos\rho_1\cos\rho_2\ \mathtt{i} + \sin\rho_1\cos\rho_2\exp(i\varphi_1)\ \mathtt{j} \\ &+ \sin\rho_2\exp i\varphi_2\ \mathtt{k}]\cdot\exp(-i\mathbf{R}\cdot\vec{k}).\end{aligned} \tag{4.84}$$

Here, ρ_1 and ρ_2 (with $0 \le \rho_{1,2} \le 90°$) parameterize the amplitude ratios between the different components and φ_1 and φ_2 (with $0 \le \varphi_{1,2} \le 360°$) the phase differences between the first and the second or third component, respectively.
The biquaternion covariance matrix (of size $N \times N$) is built by

$$\mathbf{S}_{biq} = \mathsf{E}\left(\vec{X}_{biq}^{\dagger}\vec{X}\right). \tag{4.85}$$

In complete analogy to the "classical" or quaternion case, the eigenvalues $\lambda_{1,\ldots,2N}$ and eigenvectors $\vec{u}_{1,\ldots,2N}$ of $\mathbf{S}_{biq}$ are calculated. As $\mathbf{S}_{biq}$ is a biquaternion matrix, it has $2N$ eigenvectors. The $2K$ largest eigenvalues correspond to the signal subspace. The noise subspace $\mathbf{G}_{biq} \in \mathbb{H}_{\mathbb{C}}^{N\times(2N-2K)}$ is then composed of the smaller eigenvectors:

$$\mathbf{G}_{biq} = (\vec{u}_1, \vec{u}_2, \ldots \vec{u}_{2N-2K}). \tag{4.86}$$

The properties of the sources are retrieved by searching for the $\vec{a}_q(\vec{k},\rho,\varphi)$ (defined in equation (4.84)) which minimizes the distance to the signal subspace. The squared distance to the signal subspace can be expressed as

$$\begin{aligned} d_{biq}^2 &= \|\mathbf{G}_{biq}^{\dagger}\,\vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\|^2 \\ &= \mathtt{S}\left(\left(\mathbf{G}_{biq}^{\dagger}\,\vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\right)^{\dagger}\ \mathbf{G}_{biq}^{\dagger}\,\vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\right) \\ &= \mathtt{S}\left(\vec{a}_{biq}^{\dagger}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\ \mathbf{G}_{biq}\mathbf{G}_{biq}^{\dagger}\,\vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\right). \end{aligned} \tag{4.87}$$

The parameters $\vec{k}$, ρ_1, ρ_2, φ_1, φ_2 which maximize the biquaternion music functional

$$P_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2) = \frac{1}{\mathrm{S}\left(\vec{a}^{\dagger}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\ \mathbf{G}_{biq}\mathbf{G}^{\dagger}_{biq}\ \vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\right)} \tag{4.88}$$

indicate the properties of the incident waves. In general, these properties have to be found by a grid search along six dimensions corresponding to azimuth, slowness and the polarization parameters ρ_1, ρ_2, φ_1 and φ_2. Similar to the quaternion-MUSIC case, the problem can be separated in two problems: First, the estimation of the wave vector and second, the estimation of the four polarization parameters. The first problem can be solved by a two-dimensional grid search. The second problem would correspond to a four-dimensional, and thus time-consuming, grid search. Nevertheless, in a similar way to the quaternion-MUSIC case, we found out that the second grid search is not necessary.

4.5.2 Polarization parameter identification for known wave vector

If the azimuth and slowness values are already known, ρ_1, ρ_2, φ_1 and φ_2 can be found as follows.
As slowness and azimuth are known, $\exp(-i\mathbf{R}\cdot\vec{k}) = \exp(-i\vec{\delta})$ is given. Maximizing equation (4.88) is equivalent to the minimization of

$$D = \mathrm{S}\left(\vec{a}^{\dagger}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\ \mathbf{G}_{biq}\mathbf{G}^{\dagger}_{biq}\ \vec{a}_{biq}(\vec{k},\rho_1,\rho_2,\varphi_1,\varphi_2)\right). \tag{4.89}$$

Inserting equation (4.84) in equation (4.89) yields

$$\begin{aligned}
D =& \frac{1}{N}\,\mathrm{S}\Big(\Big[\left[\cos\rho_1\cos\rho_2\ \mathtt{i} + \sin\rho_1\cos\rho_2 e^{i\varphi_1}\ \mathtt{j} + \sin\rho_2 e^{i\varphi_2}\ \mathtt{k}\right]\cdot e^{-i\vec{\delta}}\Big]^{\dagger} \\
&\cdot \mathbf{G}_{biq}\mathbf{G}^{\dagger}_{biq}\left[\left[\cos\rho_1\cos\rho_2\ \mathtt{i} + \sin\rho_1\cos\rho_2 e^{i\varphi_1}\ \mathtt{j} + \sin\rho_2 e^{i\varphi_2}\ \mathtt{k}\right]\cdot e^{-i\vec{\delta}}\right]\Big) \\
D =& \frac{1}{N}\,\mathrm{S}\Big(\Big[e^{i\vec{\delta}^{\dagger}}\cdot\left[-\cos\rho_1\cos\rho_2\ \mathtt{i} - \sin\rho_1\cos\rho_2 e^{-i\varphi_1}\ \mathtt{j} - \sin\rho_2 e^{-i\varphi_2}\ \mathtt{k}\right]\Big] \\
&\cdot \mathbf{G}_{biq}\mathbf{G}^{\dagger}_{biq}\left[\left[\cos\rho_1\cos\rho_2\ \mathtt{i} + \sin\rho_1\cos\rho_2 e^{i\varphi_1}\ \mathtt{j} + \sin\rho_2 e^{i\varphi_2}\ \mathtt{k}\right]\cdot e^{-i\vec{\delta}}\right]\Big) \\
D =& \frac{1}{N}\,\mathrm{S}\Big(\left[-\cos\rho_1\cos\rho_2\ \mathtt{i} - \sin\rho_1\cos\rho_2 e^{-i\varphi_1}\ \mathtt{j} - \sin\rho_2 e^{-i\varphi_2}\ \mathtt{k}\right] \\
&\cdot e^{i\vec{\delta}^{\dagger}}\mathbf{G}_{biq}\mathbf{G}^{\dagger}_{biq}e^{-i\vec{\delta}}\cdot\left[\cos\rho_1\cos\rho_2\ \mathtt{i} + \sin\rho_1\cos\rho_2 e^{i\varphi_1}\ \mathtt{j} + \sin\rho_2 e^{i\varphi_2}\ \mathtt{k}\right]\Big).
\end{aligned} \tag{4.90}$$

By substituting

$$C = e^{i\vec{\delta}^{\dagger}}\ \mathbf{G}_{biq}\mathbf{G}^{\dagger}_{biq}\ e^{-i\vec{\delta}} \in \mathbb{H}_{\mathbb{C}}, \tag{4.91}$$

with $C = C_0 + C_1\mathtt{i} + C_2\mathtt{j} + C_3\mathtt{k}$ and $C_{0,\ldots,3} \in \mathbb{C}$, this simplifies to

$$\begin{aligned}
D = \frac{1}{N}\,\mathrm{S}\Big(& -\cos^2\rho_1\cos^2\rho_2\ \mathtt{i}C\mathtt{i} - \sin^2\rho_1\cos^2\rho_2\ \mathtt{j}C\mathtt{j} - \sin^2\rho_2\ \mathtt{k}C\mathtt{k} \\
&- \sin\rho_1\cos\rho_1\cos^2\rho_2 e^{i\varphi_1}\ \mathtt{i}C\mathtt{j} - \sin\rho_1\cos\rho_1\cos^2\rho_2 e^{-i\varphi_1}\ \mathtt{j}C\mathtt{i} \\
&- \cos\rho_1\sin\rho_2\cos\rho_2 e^{i\varphi_2}\ \mathtt{i}C\mathtt{k} - \cos\rho_1\sin\rho_2\cos\rho_2 e^{-i\varphi_2}\ \mathtt{k}C\mathtt{i} \\
&- \sin\rho_1\sin\rho_2\cos\rho_2 e^{i(\varphi_2-\varphi_1)}\ \mathtt{j}C\mathtt{k} - \sin\rho_1\sin\rho_2\cos\rho_2 e^{i(\varphi_1-\varphi_2)}\ \mathtt{k}C\mathtt{j}\Big).
\end{aligned} \tag{4.92}$$

Here, it is necessary to calculate the effect of multiplying C with one quaternion unit from left and one from right, which gives the following results:

$$\begin{array}{lll} \mathtt{S(iCi)} = -C_0 & \mathtt{S(jCi)} = -C_3 & \mathtt{S(kCi)} = +C_2 \\ \mathtt{S(iCj)} = +C_3 & \mathtt{S(jCj)} = -C_0 & \mathtt{S(kCj)} = -C_1 \\ \mathtt{S(iCk)} = -C_2 & \mathtt{S(jCk)} = +C_1 & \mathtt{S(kCk)} = -C_0 \end{array} \tag{4.93}$$

Inserting the properties (4.93) in equation (4.92) yields

$$\begin{aligned} D = \frac{1}{N}\Big(& \cos^2\rho_1\cos^2\rho_2 C_0 + \sin^2\rho_1\cos^2\rho_2 C_0 + \sin^2\rho_2 C_0 \\ & + \sin\rho_1\cos\rho_1\cos^2\rho_2\left(e^{-i\varphi_1} - e^{i\varphi_1}\right) C_3 \\ & + \cos\rho_1\sin\rho_2\cos\rho_2\left(e^{i\varphi_2} - e^{-i\varphi_2}\right) C_2 \\ & + \sin\rho_1\sin\rho_2\cos\rho_2\left(e^{i(\varphi_1-\varphi_2)} - e^{i(\varphi_2-\varphi_1)}\right) C_1\Big) \\ = \frac{1}{N}\Big(& C_0 - 2i\sin\rho_1\cos\rho_1\cos^2\rho_2\sin\varphi_1 C_3 \\ & + 2i\cos\rho_1\sin\rho_2\cos\rho_2\sin\varphi_2 C_2 \\ & - 2i\sin\rho_1\sin\rho_2\cos\rho_2\sin(\varphi_2-\varphi_1) C_1\Big). \end{aligned} \tag{4.94}$$

As D is the norm of a biquaternion vector, it is real-valued. Therefore, its imaginary part vanishes for every set of parameters (ρ_1, ρ_2, φ_1, φ_2). Consequently, $\Im(C_0) = 0$ and $\Re(C_k) = 0$ for $k \in \{1,2,3\}$. Equation (4.94) can be simplified by omitting its imaginary part. Therefore, we define that $C_{1I} = \Im(C_1)$, $C_{2I} = \Im(C_2)$ and $C_{3I} = \Im(C_3)$ are the respective imaginary parts. Then, equation (4.94) can be written as

$$\begin{aligned} D = \frac{1}{N}\Big(& C_0 + 2\sin\rho_1\cos\rho_1\cos^2\rho_2\sin\varphi_1 C_{3I} \\ & - 2\cos\rho_1\sin\rho_2\cos\rho_2\sin\varphi_2 C_{2I} \\ & + 2\sin\rho_1\sin\rho_2\cos\rho_2\sin(\varphi_2-\varphi_1) C_{1I}\Big). \end{aligned} \tag{4.95}$$

The parameters ρ_1, ρ_2, φ_1 and φ_2 minimizing D can now be found by searching the roots of the partial derivatives:

$$\frac{\partial D}{\partial \rho_1} = \frac{1}{N}\Big(2(\cos^2\rho_1 - \sin^2\rho_1)\cos^2\rho_2 \sin\varphi_1 C_{3I} + 2\sin\rho_1 \sin\rho_2 \cos\rho_2 \sin\varphi_2 C_{2I} + 2\cos\rho_1 \sin\rho_2 \cos\rho_2 \sin(\varphi_2 - \varphi_1) C_{1I}\Big), \tag{4.96}$$

$$\frac{\partial D}{\partial \rho_2} = \frac{1}{N}\Big(-4\sin\rho_1 \cos\rho_1 \sin\rho_2 \cos\rho_2 \sin\varphi_1 C_{3I} - 2\cos\rho_1(\cos^2\rho_2 - \sin^2\rho_2)\sin\varphi_2 C_{2I} + 2\sin\rho_1(\cos^2\rho_2 - \sin^2\rho_2)\sin(\varphi_2 - \varphi_1) C_{1I}\Big), \tag{4.97}$$

$$\frac{\partial D}{\partial \varphi_1} = \frac{1}{N}\Big(2\sin\rho_1 \cos\rho_1 \cos^2\rho_2 \cos\varphi_1 C_{3I} - 2\sin\rho_1 \sin\rho_2 \cos\rho_2 \cos(\varphi_2 - \varphi_1) C_{1I}\Big), \tag{4.98}$$

$$\frac{\partial D}{\partial \varphi_2} = \frac{1}{N}\Big(-2\cos\rho_1 \sin\rho_2 \cos\rho_2 \cos\varphi_2 C_{2I} + 2\sin\rho_1 \sin\rho_2 \cos\rho_2 \cos(\varphi_2 - \varphi_1) C_{1I}\Big), \tag{4.99}$$

which leads to the following set of equations (using $\cos^2 x - \sin^2 x = \cos(2x)$ and $\sin x \cos x = \frac{1}{2}\sin(2x)$ and supposing that for the general solution $\rho_{1,2} \neq 0°, 90°$):

$$\cos(2\rho_1)\cos\rho_2 \sin\varphi_1 C_{3I} + \sin\rho_1 \sin\rho_2 \sin\varphi_2 C_{2I} + \cos\rho_1 \sin\rho_2 \sin(\varphi_2 - \varphi_1) C_{1I} = 0 \tag{4.100}$$

$$-\sin(2\rho_1)\sin(2\rho_2)\sin\varphi_1 C_{3I} - 2\cos\rho_1 \cos(2\rho_2)\sin\varphi_2 C_{2I} + 2\sin\rho_1 \cos(2\rho_2)\sin(\varphi_2 - \varphi_1) C_{1I} = 0 \tag{4.101}$$

$$\cos\rho_1 \cos\rho_2 \cos\varphi_1 C_{3I} - \sin\rho_2 \cos(\varphi_2 - \varphi_1) C_{1I} = 0 \tag{4.102}$$

$$-\cos\rho_1 \cos\varphi_2 C_{2I} + \sin\rho_1 \cos(\varphi_2 - \varphi_1) C_{1I} = 0. \tag{4.103}$$

This set of equations is not easily solvable by analytic means. However, a trivial solution can be easily found. For $\varphi_{1,2} = 0, 180°$, equations (4.100) and (4.101) are true and ρ_1 and ρ_2 can be found by solving equations (4.102) and (4.103). This yields the four trivial solutions given in table 4.1. However, only one of these solutions is admissible ($0 \leq \rho_{1,2} \leq 90°$). Therefore, the signs of C_{1I}, C_{2I} and C_{3I} determine this solution.

Table 4.1: Trivial solutions for the equation set (4.100)-(4.103).

φ_1	φ_2	$\tan\rho_1$	$\tan\rho_2$
0°	0°	$\frac{C_{2I}}{C_{1I}}$	$\frac{C_{3I}}{C_{1I}}\cos\rho_1$
0°	180°	$\frac{C_{2I}}{C_{1I}}$	$-\frac{C_{3I}}{C_{1I}}\cos\rho_1$
180°	0°	$-\frac{C_{2I}}{C_{1I}}$	$\frac{C_{3I}}{C_{1I}}\cos\rho_1$
180°	180°	$-\frac{C_{2I}}{C_{1I}}$	$-\frac{C_{3I}}{C_{1I}}\cos\rho_1$

Nevertheless, biquaternion-MUSIC is intended to retrieve any polarization parameters of seismic motion. Therefore, the solutions given in 4.1 are not acceptable because they impose phase differences of 0° or 180° between the different components. Other solutions of the equations (4.100)-(4.103) cannot be retrieved by analytic means. A grid search would be necessary for that purpose. However, a two-dimensional grid search for ρ_1 and φ_2 is sufficient. Equation (4.103) then yields

$$\varphi_1 = \varphi_2 - \arccos\left(\frac{C_{2I}\cos\varphi_2}{C_{1I}\tan\rho_1}\right), \tag{4.104}$$

which is only valid if $\left|\frac{C_{2I}\cos\varphi_2}{C_{1I}\tan\rho_1}\right| < 1$, i.e.

$$\rho_1 > \arctan\left|\frac{C_{2I}}{C_{1I}}\cos\varphi_2\right|. \tag{4.105}$$

Therefore, the grid search in ρ_1 can be limited to the interval $[\arctan\left|\frac{C_{2I}}{C_{1I}}\cos\varphi_2\right|, 90^\circ]$, whereas the search for φ_2 stretches the interval $[0, 360^\circ[$. Equation (4.104) has two solutions in the intervall $[0^\circ, 360^\circ]$. According to equation (4.102), ρ_2 can then be determined by

$$\rho_2 = \arctan\left(\frac{C_{3I}\sin\rho_1\cos\varphi_1}{C_{2I}\cos\varphi_2}\right). \tag{4.106}$$

As $0 \leq \rho_2 \leq 90^\circ$, $\frac{C_{3I}\sin\rho_1\cos\varphi_1}{C_{2I}\cos\varphi_2}$ has to be positive. The grid search of all admissible solutions yields the set (ρ_1, ρ_2, φ_1, φ_2) which minimizes equation (4.95).

4.5.2.1 Case of identical signals for all three components

A special case is given if the signals for the three components are identical $\left(\vec{X}_1(f) = \vec{X}_2(f) = \vec{X}_2(f)\right)$. In this case, the three components of the correct steering vector (equation (4.84)) should be equal, i.e.

$$\cos\rho_1\cos\rho_2 = \sin\rho_1\cos\rho_2\exp(i\varphi_1) = \sin\rho_2\exp(i\varphi_2), \tag{4.107}$$

which is equivalent to

$$\rho_1 = 45^\circ, \tag{4.108}$$

$$\rho_2 = \arcsin\left(\frac{1}{\sqrt{3}}\right), \tag{4.109}$$

$$\varphi_1 = 0^\circ, \tag{4.110}$$

$$\varphi_2 = 0^\circ. \tag{4.111}$$

Calculating the covariance matrix of the data vector $\vec{X}_{biq}(f) = \vec{X}_1(f)\cdot(\mathtt{i}+\mathtt{j}+\mathtt{k})$ gives

$$\begin{aligned}\mathbf{S}_{biq} &= \vec{X}_{biq}^{\dagger}\vec{X}_{biq} \\ &= \vec{X}_1^{\dagger}\vec{X}_1\cdot(-\mathtt{i}-\mathtt{j}-\mathtt{k})\cdot(\mathtt{i}+\mathtt{j}+\mathtt{k}) \\ &= 3\vec{X}_1^{\dagger}\vec{X}_1. \end{aligned} \tag{4.112}$$

Therefore, only the scalar part of the covariance matrix is non-zero and $C_1 = C_2 = C_3 = 0$. Consequently, equation (4.95) simplifies to

$$D = \frac{C_0}{N} \tag{4.113}$$

and is independent of ρ_1, ρ_2, φ_1 and φ_2. Although this very simple special case is unrealistic for real data measurements, the method should be able to resolve the polarization parameters.

4.5.2.2 Case of identical signals for two components

If the signals for two components are identical, two polarization values are already known in theory. In the following, we will focus on the case $\vec{X}_1(f) = \vec{X}_2(f)$, which would be described by $\rho_1 = 45^\circ$ and $\varphi_1 = 0^\circ$. In this case, the data vector is given by $\vec{X}_{biq}(f) = \vec{X}_1(f)\,(\mathtt{i}+\mathtt{j}) + \vec{X}_2(f)\,\mathtt{k}$, which gives the covariance matrix

$$\begin{aligned}\mathbf{S}_{biq} &= \vec{X}_{biq}^{\dagger}\vec{X}_{biq} \\ &= \left[\vec{X}_1^{\dagger}(f)\,(-\mathtt{i}-\mathtt{j}) - \vec{X}_3^{\dagger}(f)\,\mathtt{k}\right]\left[\vec{X}_1(f)\,(\mathtt{i}+\mathtt{j}) + \vec{X}_3(f)\,\mathtt{k}\right] \\ &= 2\vec{X}_1^{\dagger}(f)\vec{X}_1(f) + \vec{X}_3^{\dagger}(f)\vec{X}_3(f) + \vec{X}_1^{\dagger}(f)\vec{X}_3(f)(-\mathtt{i}+\mathtt{j}) + \vec{X}_3^{\dagger}\vec{X}_1(f)(\mathtt{i}-\mathtt{j}).\end{aligned} \tag{4.114}$$

Consequently, $\mathtt{K}(\mathbf{S}_{biq}) = 0$ and $C_3 = 0$.
Furthermore, as $\left(\vec{X}_1^{\dagger}(f)\vec{X}_3(f)(-\mathtt{i}+\mathtt{j})\right)^{\dagger} = \vec{X}_3^{\dagger}(f)\vec{X}_1(f)(\mathtt{i}-\mathtt{j})$, $C_1 = -C_2$.
Therefore, equation (4.95) can be written as

$$D = \frac{1}{N}\left[C_0 + 2\sin\rho_2\cos\rho_2(\cos\rho_1\sin\varphi_2 + \sin\rho_1\sin(\varphi_2-\varphi_1))C_{1I}\right]. \tag{4.115}$$

Setting the derivatives of D with respect to ρ_1, ρ_2, φ_1 and φ_2 to 0 then yields

$$\frac{\partial D}{\partial \rho_1} = \frac{2C_{1I}}{N}\sin\rho_2\cos\rho_2(-\sin\rho_1\sin\varphi_2 + \cos\rho_1\sin(\varphi_2-\varphi_1)) = 0, \tag{4.116}$$

$$\frac{\partial D}{\partial \rho_2} = \frac{2C_{1I}}{N}(\cos^2\rho_2 - \sin^2\rho_2)(\cos\rho_1\sin\varphi_2 + \sin\rho_1\sin(\varphi_2-\varphi_1)) = 0, \tag{4.117}$$

$$\frac{\partial D}{\partial \varphi_1} = -\frac{2C_{1I}}{N}\sin\rho_2\cos\rho_2\sin\rho_1\cos(\varphi_2-\varphi_1) = 0, \tag{4.118}$$

$$\frac{\partial D}{\partial \varphi_2} = \frac{2C_{1I}}{N}\sin\rho_2\cos\rho_2(\cos\rho_1\cos\varphi_2 + \sin\rho_1\cos(\varphi_2-\varphi_1)) = 0. \tag{4.119}$$

For the general solution where no components vanish, ρ_1 and ρ_2 are not allowed to become 0° or 90°. Consequently, equations (4.116) - (4.119) are simplified to

$$-\sin\rho_1\sin\varphi_2 + \cos\rho_1\sin(\varphi_2-\varphi_1) = 0, \tag{4.120}$$

$$(\cos^2\rho_2 - \sin^2\rho_2)(\cos\rho_1\sin\varphi_2 + \sin\rho_1\sin(\varphi_2-\varphi_1)) = 0, \tag{4.121}$$

$$\cos(\varphi_2-\varphi_1) = 0, \tag{4.122}$$

$$\cos\rho_1\cos\varphi_2 + \sin\rho_1\cos(\varphi_2-\varphi_1) = 0. \tag{4.123}$$

Inserting equation (4.122) in equation (4.123) gives

$$\cos\varphi_2 = 0, \tag{4.124}$$

and therefore $\varphi_2 = 90^\circ$ or 270°. Then, equation (4.122) yields $\varphi_1 = 0^\circ$ or 180°. However, for $\varphi_1 = 180^\circ$, equation (4.120) has no solution for $0^\circ < \rho_1 < 90^\circ$. Therefore, $\varphi_1 = 0^\circ$. Inserting either value of φ_2 in equation (4.120) gives $\tan\rho_1 = 1$ and therefore $\rho_1 = 45^\circ$. Using the obtained values, equation (4.121) can only be solved if $\tan\rho_2 = 1$ and therefore $\rho_2 = 45^\circ$. Inserting the obtained values in the formula for D given in equation (4.115) yields $D = \frac{1}{N}\left(C_0 + \sqrt{2}\,C_{1I}\right)$ for $\varphi_2 = 90^\circ$ and $D = \frac{1}{N}\left(C_0 - \sqrt{2}\,C_{1I}\right)$ for $\varphi_2 = 270^\circ$. The sign of C_{1I} decides whether $\varphi_2 = 90^\circ$ or 270° corresponds to the minimum of D.
In the case where the signal of two components is identical, the polarization between these two components is well retrieved (i.e. $\rho_1 = 45^\circ$ and $\varphi_1 = 0^\circ$). However, the polarization of the third component of the signal is always found to be $\rho_2 = 45^\circ$ and $\varphi_2 = 90^\circ$ or 270° and therefore, the correct polarization is not retrievable in general.

4.5.2.3 Case of two-dimensional signals

The motion of Rayleigh waves is confined to a plane built by the vertical and radial components, Love waves are confined to the transverse component. Therefore, for a pure wave of one of both types, at least one component vanishes. Only in the presence of additional noise or both wave types in the same time, all three components are non-zero. Therefore, it is interesting to see how biquaternion-MUSIC is performing in these cases. In the following, it is supposed that the wave vector $\vec{k}$ has already been determined before searching for the polarization parameters.
If one of the three components vanishes, the biquaternion data vector $\vec{X}_{biq}(f)$ given in equation (4.83) will only have two non-zero quaternion parts and consequently, the biquaternion covariance matrix $\mathbf{S}_{biq}$ given by equation (4.85) has only two non-vanishing parts: The scalar part and the part for which the data vector vanishes (i.e. if $\mathtt{I}(\vec{X}_{biq}(f)) = \vec{0}$, then $\mathtt{J}(\mathbf{S}_{biq}) = \mathbf{0}$ and $\mathtt{K}(\mathbf{S}_{biq}) = \mathbf{0}$). Consequently, the same property is valid for the eigenvectors of $\mathbf{S}_{biq}$ and subsequently for C defined by equation (4.91), because the other components $\mathtt{i}$ and $\mathtt{j}$ cannot be generated by any product of a scalar part and $\mathtt{k}$. Therefore, two of the three components C_1, C_2 and C_3 vanish.
We will only show the case of a vanishing third component of $\vec{X}_{biq}(f)$, i.e. $\vec{X}_3(f) = \vec{0}$. Then, $C_1 = C_2 = 0$ and equation (4.95) simplifies to

$$D = \frac{1}{N}\left(C_0 + 2\sin\rho_1\cos\rho_1\cos^2\rho_2\sin\varphi_1 C_{3I}\right). \tag{4.125}$$

The four derivatives of D with respect to ρ_1, ρ_2, φ_1 and φ_2 can then be set to 0, which yields

$$\frac{\partial D}{\partial \rho_1} = \frac{2}{N}(\cos^2\rho_1 - \sin^2\rho_1)\cos^2\rho_2 \sin\varphi_1 C_{3I} = 0, \tag{4.126}$$

$$\frac{\partial D}{\partial \rho_2} = -\frac{2}{N}\sin\rho_1\cos\rho_1\sin\rho_2\cos\rho_2\sin\varphi_1 C_{3I} = 0, \tag{4.127}$$

$$\frac{\partial D}{\partial \varphi_1} = \frac{2}{N}\sin\rho_1\cos\rho_1\cos^2\rho_2\cos\varphi_1 C_{3I} = 0, \tag{4.128}$$

$$\frac{\partial D}{\partial \varphi_2} = 0. \tag{4.129}$$

The only parameter in equations (4.126) - (4.129) which depends on the measurements is C_{3I}. Furthermore, this term is cancelled out. Therefore, the solutions of these equations are independent of the measurements. In fact, equations (4.126) - (4.129) are solved by $\rho_2 = \pi/2$ and arbitrary values for the other parameters. Then, $D = \frac{C_0}{N}$ is constant and it is impossible to determine the polarization of the signal. In fact, the steering vector given by equation (4.84) tries to bring the different components of the signal to equal amplitude and phase. In the case of a vanishing component, this is only possible if the steering vector vanishes the other two components, which is achieved here by $\cos\rho_2 = 0$. If the first or second component of the signal vanishes, the reasoning is analog and will not be shown here. It results that biquaternion-MUSIC is unable to retrieve the polarization for signals with a vanishing component. Nevertheless, if a component vanishes, the polarization relationship of the remaining two components can be very well retrieved using quaternion-MUSIC. As Rayleigh waves are limited to two components, it is thus not necessary to use biquaternion-MUSIC for their analysis. However, biquaternion-MUSIC might help in analyzing any complicated three-dimensional signal.

4.6 MUSIQUE

As shown above, the biquaternion-MUSIC approach does not work in certain special cases (e.g. two-dimensional signal, identical signals for two or three components). Furthermore, the polarization of a pure seismic wave is always confined to two components maximum. Therefore, the quaternion-MUSIC approach is sufficient to characterize the polarization parameters of an incident wave. Consequently, the MUSIQUE algorithm consists in a combination of the "classical" MUSIC and the quaternion-MUSIC algorithms. Figure 4.1 gives an overview of the method's principle. The recordings of a seismic array consisting of N sensors are cut into signal blocks of length $T = n_P / f_0$, where n_P indicates the number of periods in the signal and f_0 is the frequency of interest. The data of a signal block are then given (in the time domain) by three data matrices $\mathbf{x}_i \in \mathbb{R}^{N \times N_T}$, where $i \in \{1,2,3\}$ stands for the vertical, eastern and northern components, respectively. N_T is the number of data points in the signal block ($N_T = T/dt$, dt indicating the time sampling of the signals). In order to better suppress wave contributions of different frequencies, the signals $\mathbf{x}_i$ are prefiltered around f_0 with a bandwidth of df, but the actual MUSIC processing is narrow-band.

4.6.1 Estimation of the wave vector by MUSIC

A Fourier transform of $\mathbf{x}_i$ yields the respective data matrices $\mathbf{X}_i \in \mathbb{C}^{N \times N_S}$ in the frequency domain. As we are interested in the signal at a given frequency f_0, we are interested in the signal vectors $\vec{X}(f_0) \in \mathbb{C}^N$. According to equation (3.13), the covariance matrix at frequency f_0 is built for each component by

$$\mathbf{S}_i = \mathsf{E}\left(\vec{X}_i(f_0)\vec{X}_i^{\dagger}(f_0)\right) = \sum_{f=f_0-\Delta}^{f_0+\Delta} \vec{X}_i(f)\vec{X}_i^{\dagger}(f). \tag{4.130}$$

Here, the expectation value is realized by summation over a small number of frequencies around f_0. The width of Δ depends on the frequency resolution of the Fourier transform. To ensure good results, it is sufficient to sum over five frequency values. Seismic signals can occur on a single component or on all components together, depending on the motion. The motion of Love waves, for example, is confined to the horizontal components. If we are only interested in Rayleigh waves, we could use the vertical component only. However, at the peak frequency of the ellipticity, the vertical Rayleigh wave motion vanishes. In order to correctly analyze the wave field, no component has to be omitted. Therefore, the MUSIQUE code is based on a covariance matrix which is simply built by adding the matrices of the single components:

$$\mathbf{S} = \sum_{i=1}^{3} \mathbf{S}_i. \tag{4.131}$$

The "classical" MUSIC code described in 3.4 is applied to the covariance matrix defined in equation (4.131). The eigenvalues and eigenvectors of $\mathbf{S}$ are calculated and the noise subspace is constructed. Then, the wave vector $\vec{k}$ which maximizing the MUSIC functional given in equation (3.21) is searched for.

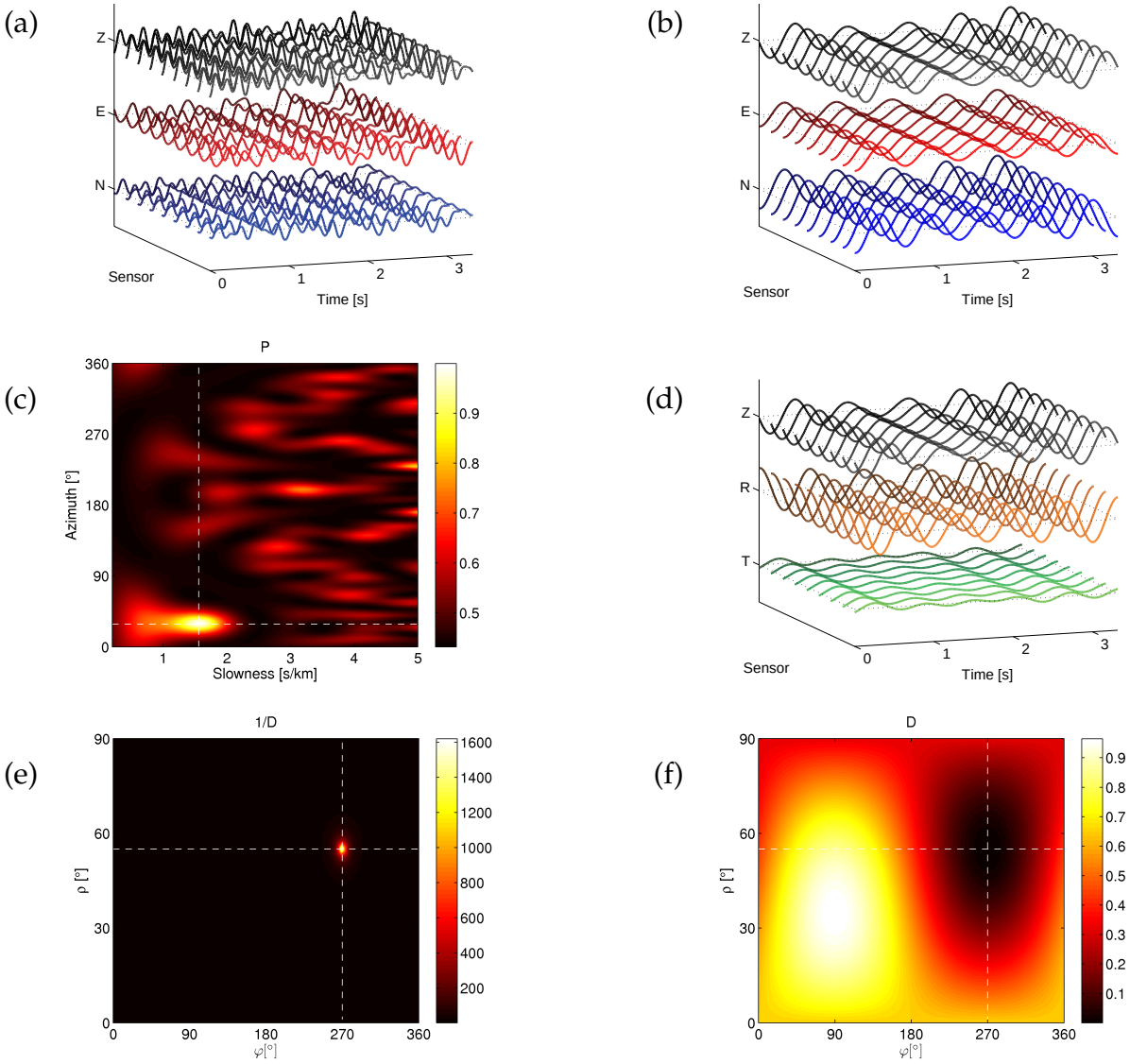


Figure 4.1: Graphic outline of the MUSIQUE method: The original three-component signal (a) of all stations of an array of seismic sensors is (b) prefiltered and cut into signal blocks of some periods of lenght. (c) A grid-search yields the wave vector (given by slowness and azimuth) which maximizes the "classical" MUSIC functional P given by equation (3.21). (d) By projecting the eastern and northern components of the signal in the azimuth and its orthogonal direction, the radial and transverse component signals are generated. Based on an energy criterium, the wave is identified either as a Love wave or as a possible Rayleigh wave. In the second case, (e) the polarization parameters ρ and φ are found, using the vertical and radial signals, by maximizing the quaternion-MUSIC functional $P_q = 1/D$, which is equivalent to (f) the minimization of D. Actually, ρ and φ can be calculated analytically and no grid search is necessary. The identification of the wave type (retrograde or prograde Rayleigh wave) is based on the value of φ.

The wave vector $\vec{k}$, i.e. azimuth and slowness corresponding to the predominant wave source are identified[VI]. The signals $\mathbf{x}_i$ are then projected in the direction of the identified wave vector $\vec{k}$ (with azimuth ϑ). Therefor, the vertical component $\mathbf{x}_1$ is left unchanged and the eastern component $\mathbf{x}_2$ and northern component $\mathbf{x}_3$ are transformed in the radial and transverse components by:

$$\begin{aligned}\mathbf{x}_{radial} &= -\sin\vartheta \cdot \mathbf{x}_2 - \cos\vartheta \cdot \mathbf{x}_3 \\ \mathbf{x}_{transverse} &= \cos\vartheta \cdot \mathbf{x}_2 - \sin\vartheta \cdot \mathbf{x}_3.\end{aligned}$$

The energies of the different components are calculated by

$$\begin{aligned}E_{vertical} &= \sum_{f=f_0-\Delta}^{f_0+\Delta} \vec{X}_1^{\dagger}(f)\ \vec{X}_1(f) \\ E_{east} &= \sum_{f=f_0-\Delta}^{f_0+\Delta} \vec{X}_2^{\dagger}(f)\ \vec{X}_2(f) \\ E_{north} &= \sum_{f=f_0-\Delta}^{f_0+\Delta} \vec{X}_3^{\dagger}(f)\ \vec{X}_3(f) \\ E_{radial} &= \sin^2\vartheta \cdot E_{east} + \cos^2\vartheta \cdot E_{north} \\ E_{transverse} &= \cos^2\vartheta \cdot E_{east} + \sin^2\vartheta \cdot E_{north}.\end{aligned}$$

If the predominant wave is a Love wave, the transverse energy will dominate over the vertical and radial energies. On the other hand, if a Rayleigh wave is predominant, the transverse energy will be smaller than both other energies. It should be noted that, depending on the ellipticity at the frequency f_0, either the vertical or the radial component can vanish although a Rayleigh wave is predominant. Therefore, the transverse energy is always compared to the sum of radial and vertical energies.
In the MUSIQUE code, if $E_{transverse} > E_{vertical} + E_{radial}$, the wave is identified as Love wave and the processing stops here. In the contrary case, the predominant wave is presumed to be a Rayleigh wave and its polarization parameters are estimated by applying quaternion-MUSIC.

[VI] In principle, MUSIC can identify multiple sources, i.e. correctly identify azimuths and slownesses if more than one source exist. However, the MUSIQUE code is principally built to retrieve the polarization parameters of Rayleigh waves, which are estimated in the further processing. This processing does, indeed, not work for more than one incident wave, because the signal is projected in the azimuth direction identified in the first step. This means that only the dominant arrival of each time window is analyzed.

4.6.2 Estimation of the polarization parameters by quaternion-MUSIC

The Fourier transform of the radial and vertical data matrices $\mathbf{x}_{radial}$ and $\mathbf{x}_1$ yields the matrices $\mathbf{X}_{radial} \in \mathbb{C}^{N \times N_S}$ and $\mathbf{X}_{vertical} = \mathbf{X}_1$. Using these matrices, the quaternion-valued signal vector (equation (4.59)) is constructed as

$$\vec{X}_{MUSIQUE}(f) = \Re(\vec{X}_{vertical}(f)) + \mathtt{i} \cdot \Re(\vec{X}_{radial}(f)) + \mathtt{j}\Im(\vec{X}_{vertical}(f)) + \mathtt{k} \cdot \Im(\vec{X}_{radial}(f)). \tag{4.132}$$

In analogy to equation (4.131), the covariance matrix $\mathbf{S}_{MUSIQUE}$ is built by

$$\mathbf{S}_{MUSIQUE} = \sum_{f=f_0-\Delta}^{f_0+\Delta} \vec{X}_{MUSIQUE}(f)\ \vec{X}^{\dagger}_{MUSIQUE}(f). \tag{4.133}$$

As described in 4.4, the eigenvalues and eigenvectors of $\mathbf{S}_{MUSIQUE}$ are calculated and the noise subspace is constructed. Because the wave vector $\vec{k}$ has already been determined by the classical version of MUSIC, the polarization parameters ρ and φ which describe the relation between both components of the motion, are obtained by equations (4.80) and (4.82) which emerge from the analytic calculation presented in 4.4.2.

The particle motion of Rayleigh waves can be retrograde or prograde. In theory, retrograde motion corresponds to a phase difference $\varphi = 90°$ and prograde motion to $\varphi = 270°$. We allow the phase differences to deviate 45° from these "perfect" phase differences, and identify waves with $45° < \varphi < 135°$ as retrograde Rayleigh waves and waves with $225° < \varphi < 315°$ as prograde ones. For other values of φ, the wave is left unclassified.

4.6.3 Assembling the results

The MUSIQUE code retrieves for every signal block:

- azimuth ϑ
- slowness s
- amplitude parameter ρ (for Rayleigh waves)
- phase difference φ (for Rayleigh waves)
- type of wave (Love wave, retrograde/prograde Rayleigh wave)

Thus, by analyzing a large number of signal blocks at different frequencies, the wave field can be characterized. We denote by $s(f,m)$ the slowness of the mth block at frequency f, by $\rho(f,m)$ the amplitude parameter, $E_{vertical}(f,m)$, $E_{radial}(f,m)$ and $E_{transverse}(f,m)$ the energies of the three components and by $l(f,m)$, $r_{retro}(f,m)$ and $r_{pro}(f,m)$ parameters which are 1 if the signal in the respective block was identified as Love wave, retrograde or prograde Rayleigh wave, respectively, and 0 everywhere else. $N_B(f)$ indicates the number of signal blocks analyzed at frequency f. Then, the dispersion curves are obtained by averaging over the different analyzed signal blocks in the following way:

$$s_{Love}(f) = \frac{\sum_{m=1}^{N_B(f)} l(f,m)\ E_{transverse}(f,m)\ s(f,m)}{\sum_{m=1}^{N_B(f)} l(f,m)\ E_{transverse}(f,m)}, \tag{4.134}$$

$$s_{Rayleigh,retro}(f) = \frac{\sum_{m=1}^{N_B(f)} r_{retro}(f,m)\ \left[E_{vertical}(f,m) + E_{radial}(f,m)\right]\ s(f,m)}{\sum_{m=1}^{N_B(f)} r_{retro}(f,m)\ \left[E_{vertical}(f,m) + E_{radial}(f,m)\right]}, \tag{4.135}$$

$$s_{Rayleigh,pro}(f) = \frac{\sum_{m=1}^{N_B(f)} r_{pro}(f,m)\ \left[E_{vertical}(f,m) + E_{radial}(f,m)\right]\ s(f,m)}{\sum_{m=1}^{N_B(f)} r_{pro}(f,m)\ \left[E_{vertical}(f,m) + E_{radial}(f,m)\right]}. \tag{4.136}$$

In a similar way, the Rayleigh wave ellipticity curves are obtained by:

$$\epsilon_{retro}(f) = \frac{\sum_{m=1}^{N_B(f)} r_{retro}(f,m)\sqrt{E_{vertical}(f,m) + E_{radial}(f,m)}\ \sin(\rho(f,m))}{\sum_{m=1}^{N_B(f)} r_{retro}(f,m)\sqrt{E_{vertical}(f,m) + E_{radial}(f,m)}\ \cos(\rho(f,m))}, \tag{4.137}$$

$$\epsilon_{pro}(f) = \frac{\sum_{m=1}^{N_B(f)} r_{pro}(f,m)\sqrt{E_{vertical}(f,m) + E_{radial}(f,m)}\ \sin(\rho(f,m))}{\sum_{m=1}^{N_B(f)} r_{pro}(f,m)\sqrt{E_{vertical}(f,m) + E_{radial}(f,m)}\ \cos(\rho(f,m))}. \tag{4.138}$$

However, misidentified signal blocks can introduce a large bias in the respective average dispersion or ellipticity estimations. Therefore, in real data applications, it is preferable to plot a histogram of all identified signal blocks before finally determining the respective curves.

4.7 Conclusion

Two advanced versions of the MUSIC algorithm have been presented, which can be used to retrieve the polarization parameters of seismic waves. The quaternion-MUSIC approach is capable of retrieving the polarization relationship for two-component seismic signals. The biquaternion-MUSIC approach can retrieve the polarization of three-component signals. At first, this method appears rather appealing. However, for a lot of special cases, this algorithm fails. Certainly, none of these special cases will ever occur for real data measurements. Nevertheless, it is difficult to assume that the algorithm yields correct results in the general case if it fails for simple special cases. Furthermore, as seismic waves are always confined to a maximum of two directions of polarization, the quaternion-MUSIC approach is sufficient to investigate these waves. This is what the MUSIQUE technique does. By identifying the azimuth and velocity (or slowness, respectively) of the wave with "classical" MUSIC first, it is possible to separate the radial and transverse components of the signal. Then, the polarization parameters between the vertical and the radial component can be estimated by using the quaternion-MUSIC approach. In contrast to other established array methods, MUSIQUE allows the distinction between retrograde and prograde Rayleigh wave motion.

MUSIQUE will be tested on synthetic seismic signals in chapter 5 and applied to real seismological data in chapter 8. Furthermore, a comparison between DELFI, RayDec and MUSIQUE was presented at the European Signal Processing Conference (EUSIPCO) 2009 in Glasgow, UK. The associated paper [Hobiger et al., 2009b] can be found in appendix A.

Chapter 5

Tests on synthetic data

In this chapter, the three methods DELFI, RayDec and MUSIQUE which were presented in previous chapters (DELFI in 2.3, RayDec in 2.4 and MUSIQUE in 4.6) are tested on synthetic seismic data. A first test investigates the behavior of the methods under the influence of additional white noise. In further tests, the methods are tested on signals of increasing complexity, starting from single source signals and ending with ambient seismic vibration simulations with a random source distribution.

Dans ce chapitre, les trois nouvelles méthodes DELFI, RayDec et MUSIQUE présentées dans les chapitres précédents (DELFI en 2.3, RayDec en 2.4 et MUSIQUE en 4.6), seront testées sur des données sismiques synthétiques. Un premier test examinera le comportement des différentes méthodes sous l'influence d'une addition de bruit blanc aux signaux. Dans les tests suivants, les méthodes seront étudiées en les appliquant à des signaux de complexité croissante, commençant par des signaux émanant d'une seule source jusqu'à des simulations de bruit sismique ambiant avec une distribution aléatoire de sources.

5.1 Introduction

This thesis work includes three new method developments, which have been presented above (DELFI in section 2.3, RayDec in section 2.4 and MUSIQUE in section 4.6). DELFI and RayDec only estimate ellipticity, MUSIQUE estimates the azimuth, slowness of seismic waves as well as the ellipticity parameter and phase difference between both the vertical and radial component for Rayleigh waves. In this chapter, the respective estimations of these methods will be tested on synthetic seismic data of increasing complexity. The first test (section 5.2) investigates how the methods behave if white noise is added to a perfect cosine signal. The white noise has a theoretical constant ellipticity of $\sqrt{2}$ because the energies of the three different components are equal. This test includes two different signal sets, one with an ellipticity of 5 and the other with an ellipticity of 0.3.

The next test (section 5.3) is based on a purely theoretical signal which is generated on the basis of a dispersion curve and an ellipticity curve. The summed energy of the vertical and horizontal components of the signal is kept constant, whereas the relative energy distribution is fixed by the ellipticity curve. The propagation of waves originating in a distance of 20 km is modeled corresponding to the dispersion curve. In this way, the wave parameters are completely controlled. The ellipticity estimations of DELFI and RayDec are compared to the H/V results and the MUSIQUE results for three arrays of different sizes are compared to the results for analyzing all stations together.

In section 5.4, ambient seismic vibrations are simulated. For two different distributions of the seismic sources (in the first one, all sources are located in a specific direction, in the second one, the source distribution is random), DELFI, RayDec and MUSIQUE are tested and the results compared.

5.2 Influence of the signal-to-noise ratio

5.2.1 Signal generation

This section investigates the influence of the signal-to-noise ratio (SNR) on the results for the three newly developed methods. The test is based on a synthetic, monochromatic retrograde Rayleigh wave signal. This signal consists of a pure cosine of a frequency of 0.77 Hz, propagating across an array of seismic sensors with a slowness of 0.59 s/km (i.e. a velocity of 1.69 km/s) under an azimuth of 30°. The seismic array consists of eight stations, regularly spaced on a ring of 1.5 km radius surrounding a central station. Its layout is displayed in figure 5.1.

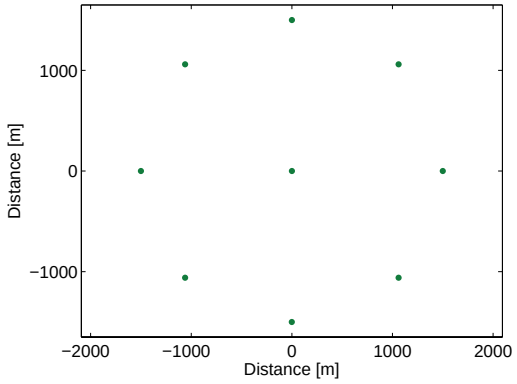


Figure 5.1: Layout of the seismic array used for the tests.

Two different signals have been generated, the first with an ellipticity of 5, the second with an ellipticity of 0.3.
White noise of variable energy content was added to the signal. The signal-to-noise ratio is defined as

$$SNR[dB] = 10 \log_{10} \left(\frac{E_{signal}}{E_{noise}} \right), \tag{5.1}$$

where the energies E_{signal} and E_{noise} are calculated on signals which are prefiltered around the analyzed frequency. Notably, these energies are calculated for a given three-component signal consisting of N data points as

$$E = \sum_{i=1}^{N} \left(v^2(t_i) + e^2(t_i) + n^2(t_i) \right), \tag{5.2}$$

where $v(t_i)$, $e(t_i)$ and $n(t_i)$ indicate the signals on the vertical, eastern and northern component, respectively.

For tests involving the MUSIQUE technique, a single window of 5 cycles length (i.e. 6.5 s long) was analyzed, adding random noise with an energy corresponding to the SNR value and repeating the test 1 000 times. For tests of DELFI and RayDec, a signal of 5 minutes length was generated, random noise with an energy corresponding to the SNR value was added and a total of 1 000 different realizations of this performed per SNR. Signal-to-noise ratios between 0 dB and 20 dB have been investigated. According to equation (5.1), a SNR of 0 dB corresponds to a signal where the energies of the signal part and the noise part are equal. For a SNR of 10 dB, the signal energy is ten times the noise energy. In figure 5.2, example signals for SNR values of 0 dB, 10 dB and 20 dB are shown for both used signals, i.e. with theoretical ellipticities of 5 and 0.3.

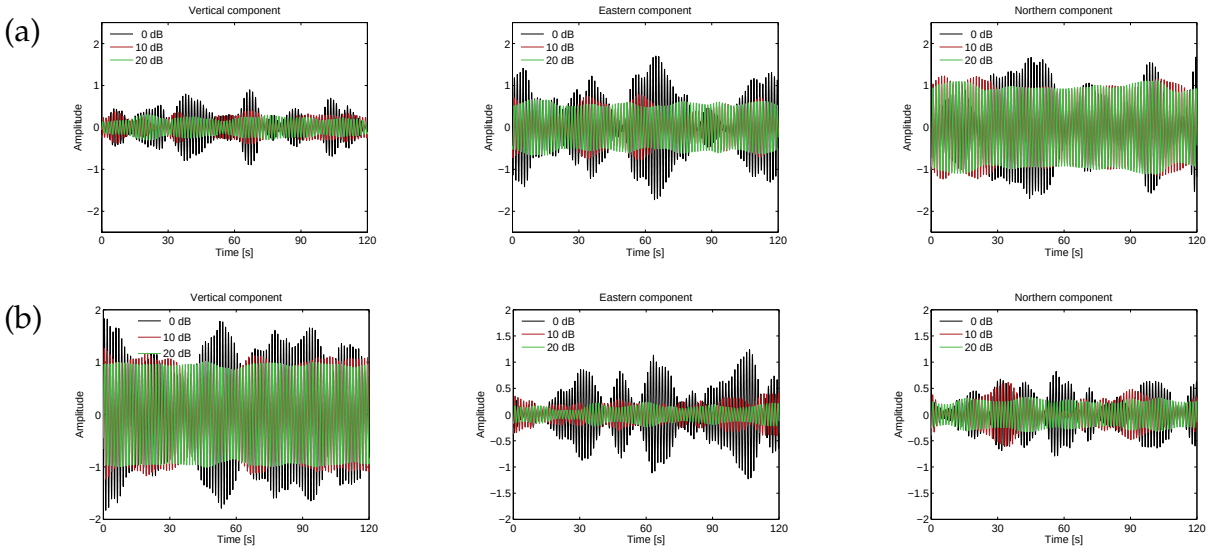


Figure 5.2: Example signals of the vertical (left), eastern (center) and northern (right) components for different signal-to-noise ratios for signals with a theoretical ellipticity of (a) 5 and (b) 0.3.

5.2.2 Analysis results

In figure 5.3, the results for the first test (with a theoretical ellipticity of 5) are shown. The mean azimuth, slowness, phase difference and ellipticity values obtained by averaging the results of analyzing 1 000 signal realizations at each SNR value using MUSIQUE are shown and compared to the signal parameters. The wave vector, i.e. azimuth and slowness, is obtained using the "classical" MUSIC method (see 3.4). It can be seen that for signal-to-noise ratios above 10, the correct azimuth and slowness values are retrieved for every signal realization (figures 5.3 (a) and (b)). The average value of the azimuth estimation is rather stable and deviates significantly from the true azimuth for SNR's below 5. For SNR = 0, the mean azimuth estimation equals 43°. In addition, the error bars of the mean azimuth estimation are negligible for SNR's above 10 and increase strongly for SNR's below 7. The mean slowness estimation is correct for SNR's above 7. Below this value, the slowness is systematically overestimated. In the same range, the error bars of this estimation increase strongly.
The polarization parameters, i.e. the phase difference φ between vertical and horizontal signal (figure 5.3 (c)) and the ellipticity (figure 5.3 (d)), are determined using the calculation which is based on the previous azimuth and slowness estimation and was presented in 4.4.2. Therefore, it cannot be expected that the polarization parameters are well estimated if the wave vector is misestimated. Nevertheless, for the phase difference φ, the mean estimation is in very good agreement with the theoretical value for all SNR ratios. However, for low SNR values the standard deviation increases. The average ellipticity estimation is in good agreement with the actual value for SNR's above 8. Below this value, the ellipticity is underestimated a lot. Nevertheless, over the whole SNR range, the error bars are rather large. Figures 5.3 (e) and (f) have been calculated in the same way as (c) and (d), but fixing the correct azimuth and slowness values before estimating the wave polarization. This eliminates the effects of the wrong slowness estimation at low SNR's. As a result, the average phase difference still fits the actual one, but the error bars are reduced. Now, the average estimation and the true curve of the ellipticity are in agreement for SNR's above 5 and the misestimation at low SNR values is reduced, with reduced error bars.

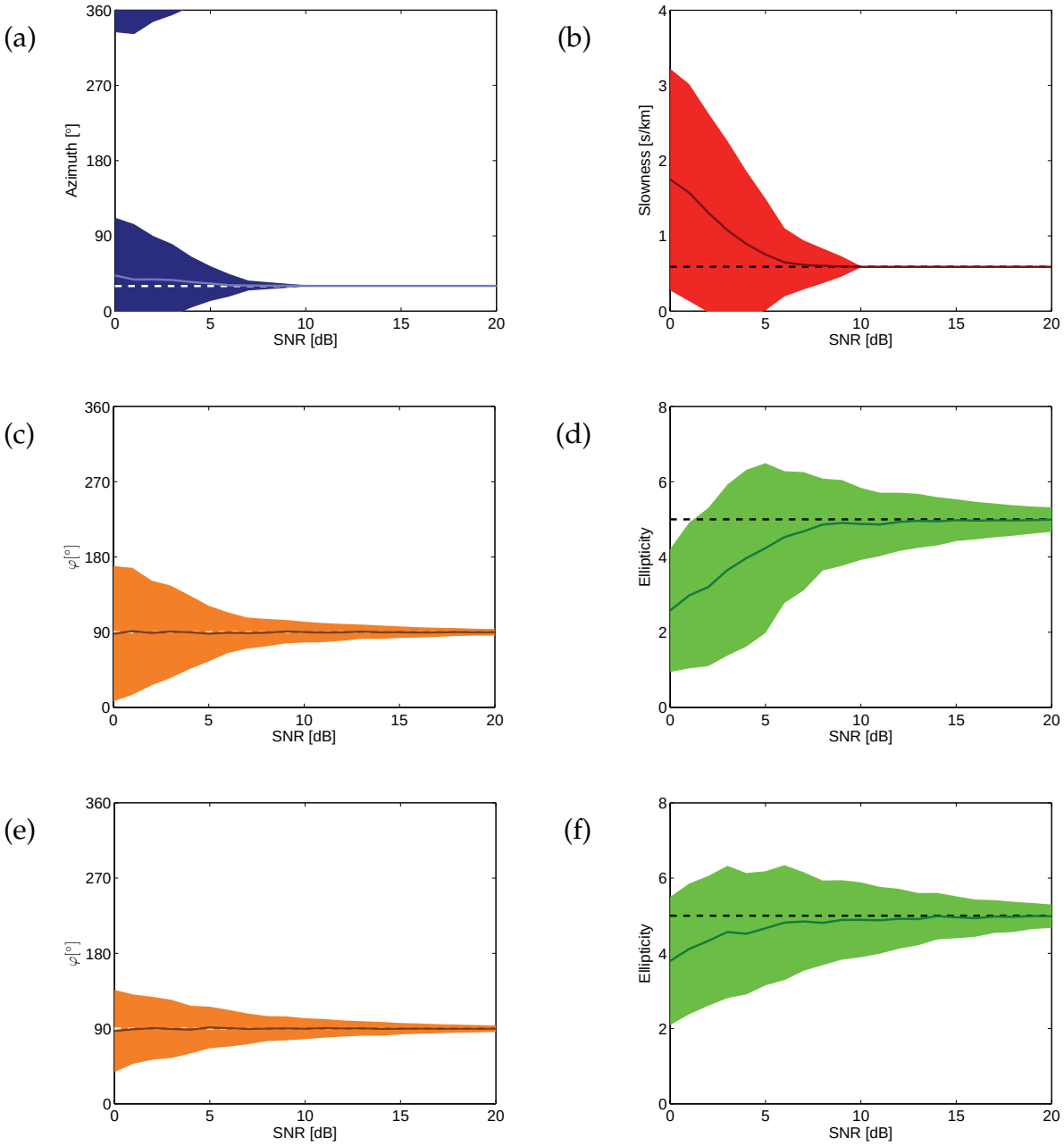


Figure 5.3: MUSIQUE results for adding white noise to a pure Rayleigh wave signal with an ellipticity of 5. 1 000 realizations per value of the signal-to-noise ratio (SNR) have been analyzed: average (a) azimuth, (b) slowness, (c) phase difference between the vertical and horizontal components and (d) ellipticity estimations. (e) Phase difference and (f) ellipticity for correctly estimated azimuth and slowness values. The solid line marks the average value, the shaded areas indicate the standard deviation. The dotted lines represent the actual values of the respective parameters for the signal.

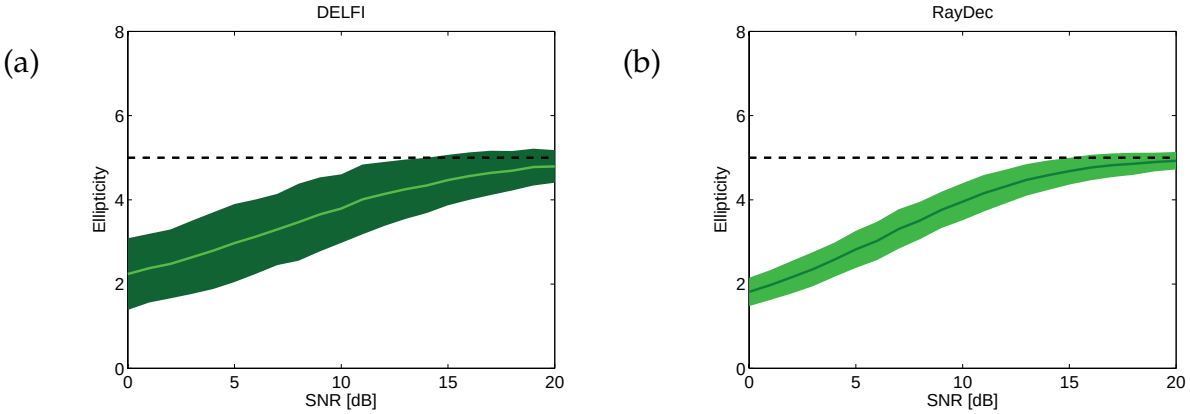


Figure 5.4: Results for adding white noise to a pure Rayleigh wave signal with an ellipticity of 5. Analyzing 1 000 realizations per SNR value using (a) DELFI, (b) RayDec. The solid line marks the average value, the shaded areas indicate the standard deviation. The dotted lines represent the actual values of the respective parameters for the signal.

The performances of DELFI and RayDec analyzing the same 1 000 realizations of theoretical signals per SNR are displayed in figures 5.4 (a) and (b), respectively. For both methods, the average ellipticity estimation decreases monotonically with decreasing SNR, without significant increase of the standard deviation. It is rather surprising that a quite small amount of noise largely influences the ellipticity estimation of these two methods. A possible explanation is that both methods exploit the time correlation between the vertical and horizontal components of a Rayleigh wave. With a theoretical ellipticity of 5, the combined energy of the horizontal components is 25 times larger than the energy of the vertical component of the signal part. As the added noise has the same energy on all three components, the noise added on the vertical component has a much larger influence than the SNR ratio would suggest[I]. Therefore, the added noise puts a relatively large amount of energy on the vertical component. Consequently, the ellipticity is underestimated. As RayDec uses the vertical component as a trigger, the method is rather vulnerable to added noise on the vertical component[II].

[I] For example, a SNR of 10 corresponds to $E_{signal} = 10\ E_{noise}$. The noise energy is equally distributed on the three components ($E_{noise,vert} = E_{noise,north} = E_{noise,east} = \frac{1}{3}\ E_{noise}$), whereas with an ellipticity of 5 the signal energy is mostly concentrated on the horizontal components ($E_{signal,north} + E_{signal,east} = 25\ E_{signal,vert} = \frac{25}{26}\ E_{signal}$. Consequently, the signal and noise contributions to the vertical component are almost equal ($E_{signal,vert} = \frac{30}{26} E_{noise,vert}$).

[II] The signal is stacked for every zero-crossing of the vertical component. All stacked signals are finally summed. If the noise on the vertical component is strong, it will also shift a large part of the zero-crossings and induce instabilities in the final ellipticity estimation.

The results for the second test with a theoretical ellipticity of 0.3 are shown in figure 5.5. As in the previous test, the results are obtained by averaging 1 000 signal realizations. For the azimuth and slowness values (figures 5.5 (a) and (b)), the results are almost identical to the previous case, with a correct azimuth and slowness retrieval for every run for SNR's above 10. For an SNR of 0, the mean azimuth is estimated as 42°. Nevertheless, the slowness overestimation at low SNR values is slightly larger than in the previous case.
The phase difference φ estimation is equivalent to the previous case, i.e. the average estimated value fits the actual value and the error bars increase monotonically with decreasing SNR. However, the ellipticity estimation is much more robust against additional noise than in the previous case. The average value fits the actual value even for low SNR's and the standard deviation rises less with a decrease of the SNR.
For DELFI (figure 5.5 (e)), at high SNR's the actual ellipticity value is well retrieved and increases monotonically with a decreasing SNR. RayDec (figure 5.5 (f)) exhibits the same behavior, but the absolute misestimation and the error bars are smaller. In this case with an ellipticity value of 0.3, the signal energy on the vertical component is 11 times the combined signal energy on the horizontal components. Therefore, the added noise acts very little on the vertical component and the RayDec estimation is more robust than in the previous case for a high ellipticity.

5.2.3 Discussion of the results

The results suggest that the azimuth estimation is much more stable against noise contributions than the slowness estimation. This can be seen in both cases, where even for a SNR of 0, the azimuth estimation error does not exceed 13°, whereas the slowness is overestimated to about three times its actual value.
A part of the ellipticity misestimation for MUSIQUE in the first test can be explained by the slowness misestimation, but even for a correct slowness estimation, the ellipticity is systematically underestimated for high noise levels. The added noise has the same energy level on all three components, which corresponds to a theoretical H/V ratio of $\sqrt{2}$ of this part[III]. Therefore, adding noise on data with an ellipticity exceeding $\sqrt{2}$ tends to underestimate the actual ellipticity value, while the effect for adding noise on data with lower ellipticity is the opposite. This can be seen in the results for all three methods. Furthermore, it seams that the ellipticity estimation using MUSIQUE is more robust against noise than using DELFI or RayDec. This might be linked to the fact that MUSIQUE uses an array of sensors. As the noise recorded on the different sensors is uncorrelated, an array can more efficiently suppress noise contributions.

[III] If the energies of all three components are equal, the H/V ratio is calculated as

$$\frac{H}{V} = \sqrt{\frac{E_{north} + E_{east}}{E_{vert}}} = \sqrt{2}.$$

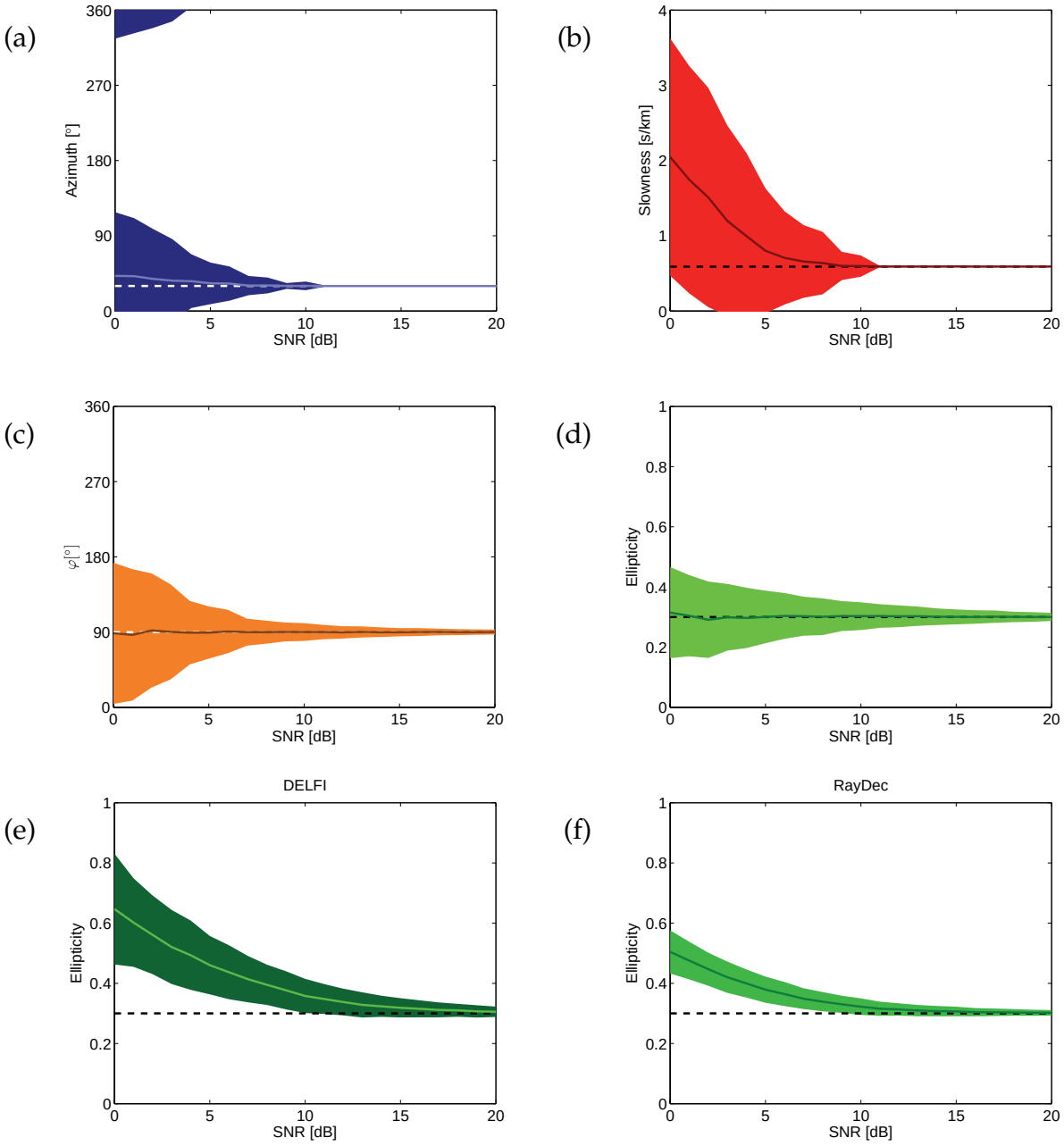


Figure 5.5: Results for adding white noise to a pure Rayleigh wave signal with an ellipticity of 0.3. (a)-(d) Analyzing 1 000 realizations per value of the signal-to-noise ratio (SNR) with MUSIQUE: average (a) azimuth, (b) slowness, (c) phase difference between the vertical and horizontal components and (d) ellipticity estimations. (e) Analyzing 1 000 realizations per SNR value using DELFI. (f) Analyzing 1 000 realizations per SNR value using RayDec. The solid line marks the average value, the shaded areas indicate the standard deviation. The dotted lines represent the actual values of the respective parameters for the signal.

5.3 Tests on a theoretical dispersive signal

5.3.1 Signal generation

This test is based on a completely theoretical signal. The reason for this is to test the methods on a signal where we control all properties. The dispersion curves $s_{Rayleigh}(f)$ and $s_{Love}(f)$ and ellipticity curve $\epsilon(f)$ shown in figures 5.6 (a) and (b) are used for the generation of the signal. For the stations of the seismic array (see figure 5.6 (c) for the array layout), plane Rayleigh and Love waves originating in a distance of $d_S = 20\ km$ from the array center with azimuth $\vartheta = 30^\circ$ are calculated as described in the following. The array is composed of three rings of eight sensors each at radii of 60, 300 and 1 500 meters, respectively, and a central station. The total energy at each frequency is kept constant by setting the sum of vertical energy $E_v(f)$ and radial energy $E_r(f)$ to 1. On the other hand, vertical and radial motions are linked through the ellipticity $\epsilon(f) = \sqrt{\frac{E_r(f)}{E_v(f)}}$. This fixes the values of $E_v(f)$ and $E_r(f)$ to:

$$E_v(f) = \frac{1}{\epsilon^2(f) + 1}, \tag{5.3}$$

$$E_r(f) = \frac{\epsilon^2(f)}{\epsilon^2(f) + 1}. \tag{5.4}$$

In a first simulation, only Rayleigh waves are accounted for. In a second test, Love waves are added. Thus, in the first simulation $E_t(f) = 0$ and in the second simulation, $E_t(f) = 1$, where $E_t(f)$ is the transverse energy.
The signals at the source are stationary cosine signals. An additional parameter $m(f)$ parameterizes the retrograde or prograde sense of motion of the Rayleigh wave: at frequencies below the ellipticity peak and beyond the trough, $m = 1$; between the peak and trough frequencies, $m = -1$. The vertical, radial and transverse signals at the source are defined by:

$$v_S(t) = \int_{f_1}^{f_2} \sqrt{E_v(f)} \cdot \sin(2\pi f t) df, \tag{5.5}$$

$$r_S(t) = \int_{f_1}^{f_2} \sqrt{E_r(f)} \cdot \sin(2\pi f t + m(f)\pi/2) df, \tag{5.6}$$

$$t_S(t) = \int_{f_1}^{f_2} \sqrt{E_t(f)} \cdot \sin(2\pi f t) df. \tag{5.7}$$

If the position of the ith sensor is given by $\vec{p}_i$ and the position of the source by $\vec{p}_S$, then their distance along the ray path is given by $d_i = |\vec{p}_S| - \frac{\vec{p}_i^T \vec{p}_S}{|\vec{p}_S|} = d_S - \frac{\vec{p}_i^T \vec{p}_S}{d_S}$.

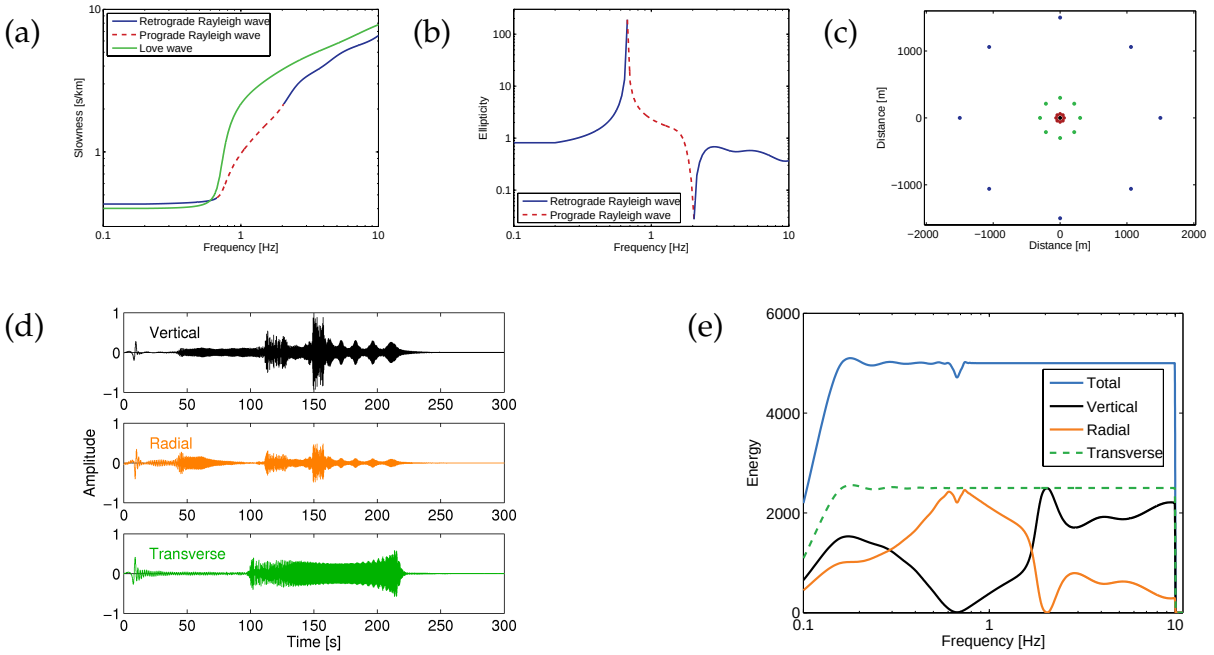


Figure 5.6: The Rayleigh and Love wave dispersion curves (a) and the Rayleigh wave ellipticity curve (b) used for simulating Rayleigh and Love waves arriving under an azimuth of 30° (i.e. from the northeast) at the array of seismic sensors (c). This yields the seismic signal (d) at the central station of the array. (e) The energy spectra of the individual components for the signal including Love waves.

At the ith sensor, the following signals are therefore recorded:

$$v_i(t) = \int_{f_1}^{f_2} \sqrt{E_v(f)} \cdot \sin(2\pi f(t - s_{Rayleigh}(f)\, d_i))\, df, \tag{5.8}$$

$$r_i(t) = \int_{f_1}^{f_2} \sqrt{E_r(f)} \cdot \sin(2\pi f(t - s_{Rayleigh}(f)\, d_i) + m(f)\pi/2)\, df, \tag{5.9}$$

$$t_i(t) = \int_{f_1}^{f_2} \sqrt{E_t(f)} \cdot \sin(2\pi f(t - s_{Love}(f)\, d_i))\, df. \tag{5.10}$$

For the simulations, the frequency limits are chosen as $f_1 = 0.1\,Hz$ and $f_2 = 10\,Hz$ and the signals integrated numerically for 10^6 different frequency values for a total signal length of 5 minutes. Finally, the eastern signals $e_i(t)$ and the northern signals $n_i(t)$ of the different sensors are obtained by projecting the radial and transverse signals in the right directions:

$$e_i(t) = \sin\vartheta \cdot r_i(t) + \cos\vartheta \cdot t_i(t), \tag{5.11}$$

$$n_i(t) = \cos\vartheta \cdot r_i(t) - \sin\vartheta \cdot t_i(t). \tag{5.12}$$

Although the signal is composed of cosine waves, the signal content is not the same at each time sample. In fact, a real Dirac signal is composed of an infinite number of cosine waves which interfere constructively at one point in time. This is the case of the presented signal at the source, although the energy content is not the same at all frequencies. For the stations of the array, the dispersion of the waves leads to a divergence of the different frequency contents. In this way, the produced signal is not too far away from a real seismic signal produced by a single hammer blow, for example. This can be seen in figure 5.6 (d) where the signal simulated for the central station of the array including Love waves is shown. In the case without Love waves, the vertical and radial signals are the same, but the transverse signal vanishes. The associated spectra of the individual components of this signal are shown in figure 5.6 (e).

5.3.2 Analysis results

5.3.2.1 Single-sensor ellipticity estimation

The results for applying the three single-sensor methods H/V, DELFI and RayDec to the theoretical signals are shown in figure 5.7.

H/V
In the case without Love waves, the H/V curve (figure 5.7 (a)) corresponds to the model's ellipticity curve. This is a direct consequence of the signal generation. In the case including Love waves, however, the H/V curve overestimates ellipticity because additional energy is present on the transverse component (figure 5.7 (b)).

DELFI
The DELFI results for the central station of the array are shown in figures 5.7 (c) and (d) for different filtering bandwidths. In the processing, signal blocks of one period length were analysed. As can be seen in figure 5.7 (c) for the signal without Love waves, the retrieved ellipticity curve is very close to the model curve over the whole frequency range, and the curve is best fitted for the smallest filtering bandwidth, i.e. $df = 0.1f$. In figure 5.7 (d), the results for the signal including Love waves are shown. At low frequencies, the ellipticity is clearly overestimated, but at frequencies above the peak frequency, the agreement is very satisfactory, again with the best results for the smallest filter bandwidth, $df = 0.1f$. This can be explained because at low frequencies, Love and Rayleigh waves have approximately the same slownesses. Therefore, the contributions of both wave types arrive at almost the same time at the array and overlap. Consequently, DELFI's estimation of the ellipticity is biased at these frequencies. At higher frequencies, the dispersion curves of Rayleigh and Love waves are different and therefore, the wave packets arrive at different times. DELFI succeeds in discriminating the Love waves in this case. Consequently, the ellipticity estimation is very close to the real curve here. However, it should be noted that DELFI cannot retrieve the correct ellipticity neither at the ellipticity peak nor at the trough. At these points, either the vertical or the radial component of the Rayleigh wave vanishes and therefore the ellipse degenerates into a linear motion. Thus, the fitting of an ellipse at these points must fail.

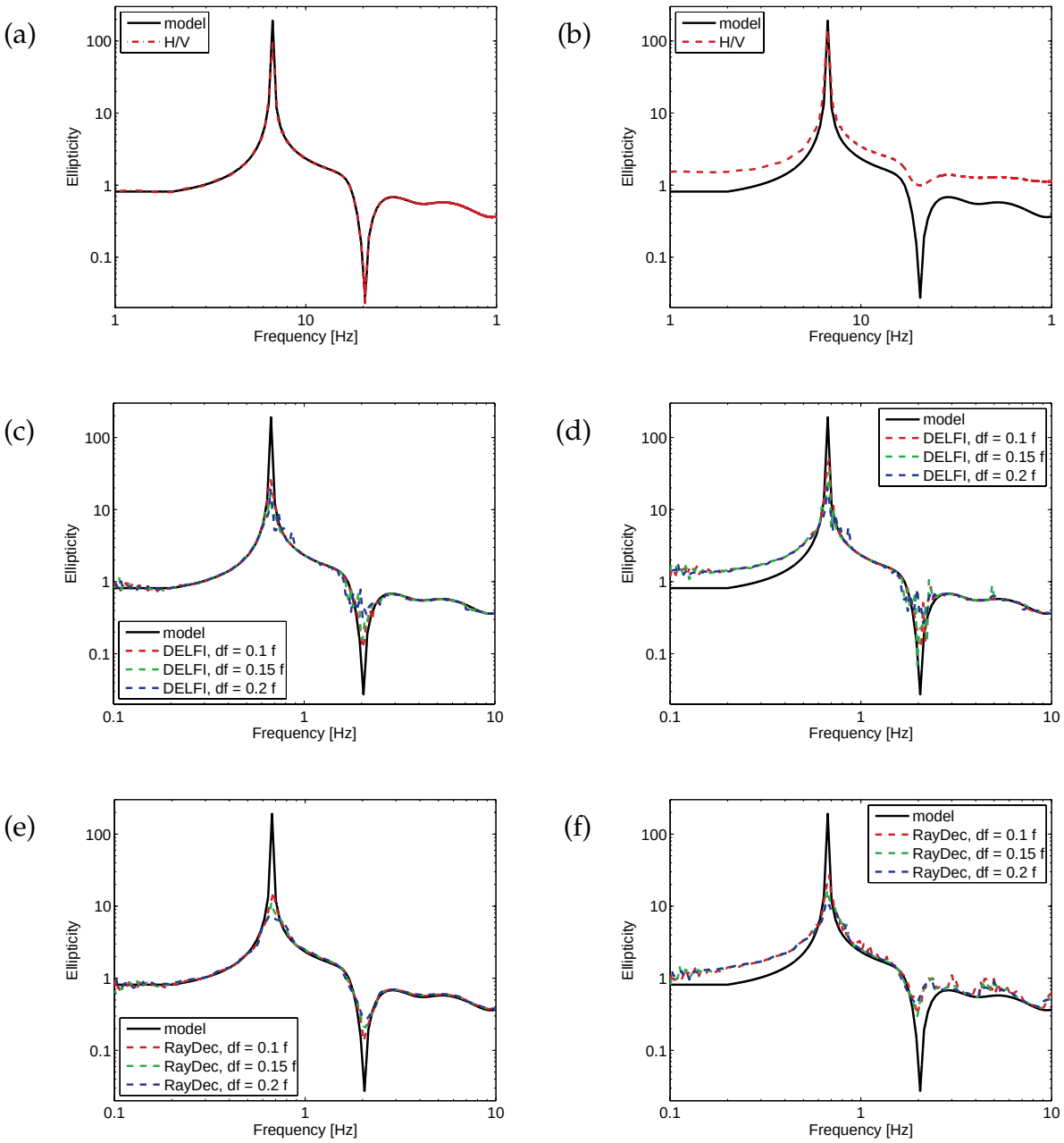


Figure 5.7: (a) and (b) H/V curves obtained for the synthetic signal without Love waves and including Love waves, respectively. (c) and (d) Curves obtained by DELFI for the respective cases with and without Love waves, both for different values of the frequency filter bandwidth df. (e) and (f) RayDec curves for the respective cases, for different values of the frequency filter bandwidth df.

RayDec
Figures 5.7 (e) and (f) show the results for applying RayDec to the signals at the central station of the array, again for different filter bandwidths. The number of cycles used in the code was set to 10. In figure 5.7 (e), no Love waves are present and the ellipticity estimation is in very good agreement for all filter bandwidths. At the peak and trough frequencies, the results are better for sharper filters. When adding Love waves (see figure 5.7 (f)), the results of the smallest filter bandwidth ($df = 0.1f$) are less stable than for the larger bandwidths. At frequencies below the peak, the ellipticity is overestimated. At the right flank of the peak, the measurement is in very good agreement with the model. Above the trough frequency, some instabilities in the ellipticity estimation can be seen. Again, these results can be explained by the close velocities of the Rayleigh and Love waves at low frequencies. In addition to the simultaneous arrival of both surface wave types at the array stations, the Love and Rayleigh waves of the simulation have a constant phase difference throughout the simulated signal. Therefore, the horizontal part of the Rayleigh wave is overestimated.
RayDec exploits the well-defined phase relationship between vertical and horizontal signal to discriminate against other wave types. Therefore, at points where one of both Rayleigh components vanishes, the ellipticity estimation will always be biased if other wave types are present. This explains why the estimated ellipticity curves cannot fit the real model data at the ellipticity's peak and trough.

Comparison
In the case of a theoretical signal without Love waves, all three methods succeed in retrieving the correct ellipticity curve. In the presence of Love waves, however, H/V overestimates the ellipticity curve. The left flank is overestimated by DELFI and RayDec, probably because Rayleigh and Love waves are present at the same time for these frequencies. DELFI retrieves the right flank of the peak and the part beyond the trough, but fails at the peak and trough frequencies. RayDec retrieves the correct right flank and deviates above the trough frequency. All methods succeed in finding the frequency of the ellipticity peak.

5.3.2.2 MUSIQUE

The signals have been analyzed by MUSIQUE both processing all stations at the same time and considering the different rings separately. For every analyzed frequency, the processing described in 4.6 was performed, i.e. the signal was cut into blocks of 5 periods length and for each block the azimuth and slowness values were estimated. Love and Rayleigh waves have been detected and the polarization parameters of the Rayleigh waves estimated. The dispersion curves shown in the following correspond to the mean value for the respective wave type at the analyzed frequency.
Figure 5.8 displays the results for analyzing all stations simultaneously for the cases without Love waves (figures 5.8 (a) and (b)) and including Love waves (figures 5.8 (c) and (d)). In both cases, the dispersion curves are well retrieved. Some prograde Rayleigh waves are found at higher frequencies, although the motion should be retrograde. Around 4 Hz, the array is losing its resolution and aliasing effects occur.

At frequencies where the Rayleigh wave motion changes from prograde to retrograde and vice versa, both curves continue across that frequency and deviate there. The corresponding ellipticity values are very well retrieved with some data points of misestimated polarization. In the case including Love waves, (c), both Love and Rayleigh wave dispersion curves are well estimated up to the resolution limit of the array (around 4 Hz). The ellipticity curve in (d) is well retrieved up to this frequency.
In theory, the prefiltering in the MUSIQUE algorithm should be redundant, as the code itself takes the signal at some frequencies only and the filtering should not bias the respective information. However, if strong sources at certain frequencies are present in the signal, they can "leak" into neighboring frequency ranges and alter the results. Figures 5.8 (e) and (f) are equivalent to (c) and (d), but the prefiltering process in the MUSIQUE code has been omitted. Evidently, the results are less clear with misestimated Love and prograde Rayleigh wave slowness values, especially at low frequencies. At higher frequencies, all dispersion curves are not as clearly retrieved as with the prefiltering. On the ellipticity curve, the effect is even more obvious as the retrograde and prograde wave motions are not well distinguished any more and the ellipticity values are misestimated. As these effects are already visible for a completely theoretical signal, they will be even more pronounced for real seismic signals, which justifies prefiltering the signals.
The data processing time increases with the number of seismic sensors[IV]. In fact, it cannot be expected that the array stations on the 1500-meter ring contribute to the analysis of high-frequency signals. Indeed, including too many sensors which do not contribute to the analysis can even worsen the results. Due to aliasing effects, both outer rings could not correctly estimate the properties of the waves for frequencies above 3 Hz in this example. Therefore, the estimation at higher frequencies is based on the smallest ring only and could be accelerated by omitting the larger rings. Surely, the presented array with an outer ring radius 25 times larger than the inner ring radius is a rather extreme example, but if the measurements are supposed to cover a large frequency range, such an array size is necessary.
In figure 5.9, the dispersion and ellipticity curves for analyzing the stations of one ring (plus the central station) only are shown, both without Love waves ((a) - (f)) and including Love waves ((g) - (l)). In the case without Love waves, dispersion and ellipticity curves are correctly estimated in every case. In the case including Love waves, the dispersion curves are well retrieved in the respective resolution limits of the single arrays. The ellipticity curves are better estimated than in the case of analyzing all stations at the same time. Notably the smallest array (figures 5.9 (i) and (l)) resolves better for frequencies around 4 Hz compared to the case including all stations in figures 5.8 (c) and (d). Therefore, by combining the results of the different subarrays, a better result can be achieved than for using all stations simultaneously. Furthermore, the computation time decreases.

[IV] Not only more data have to be processed, but in the algorithm, the eigenvalues and eigenvectors of an $N \times N$ data matrix have to be determined (where N indicates the number of sensors).

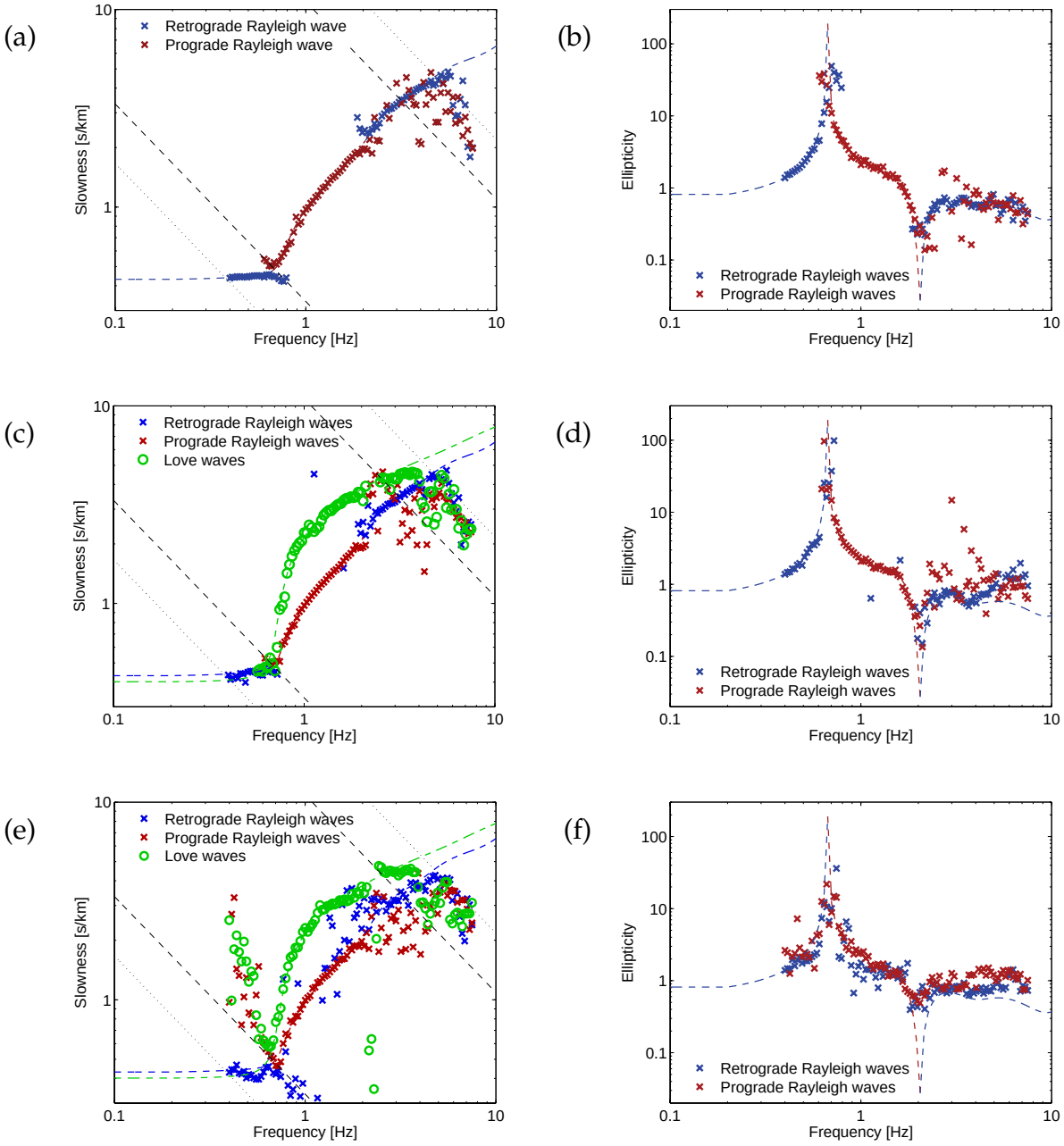


Figure 5.8: Results of applying MUSIQUE to the theoretical signals: (a) dispersion curve, (b) ellipticity curve for the case without Love waves; (c) dispersion curve, (d) ellipticity curve for the case with Love waves. (e) dispersion and (f) ellipticity curve including Love waves, but without prefiltering the signal. The dashed and dotted lines indicate roughly the resolution limits of the array, and correspond to $s = \frac{1}{2d_{max}f}$, $s = \frac{1}{d_{max}f}$, $s = \frac{1}{2d_{min}f}$ *and* $s = \frac{1}{d_{min}f}$.

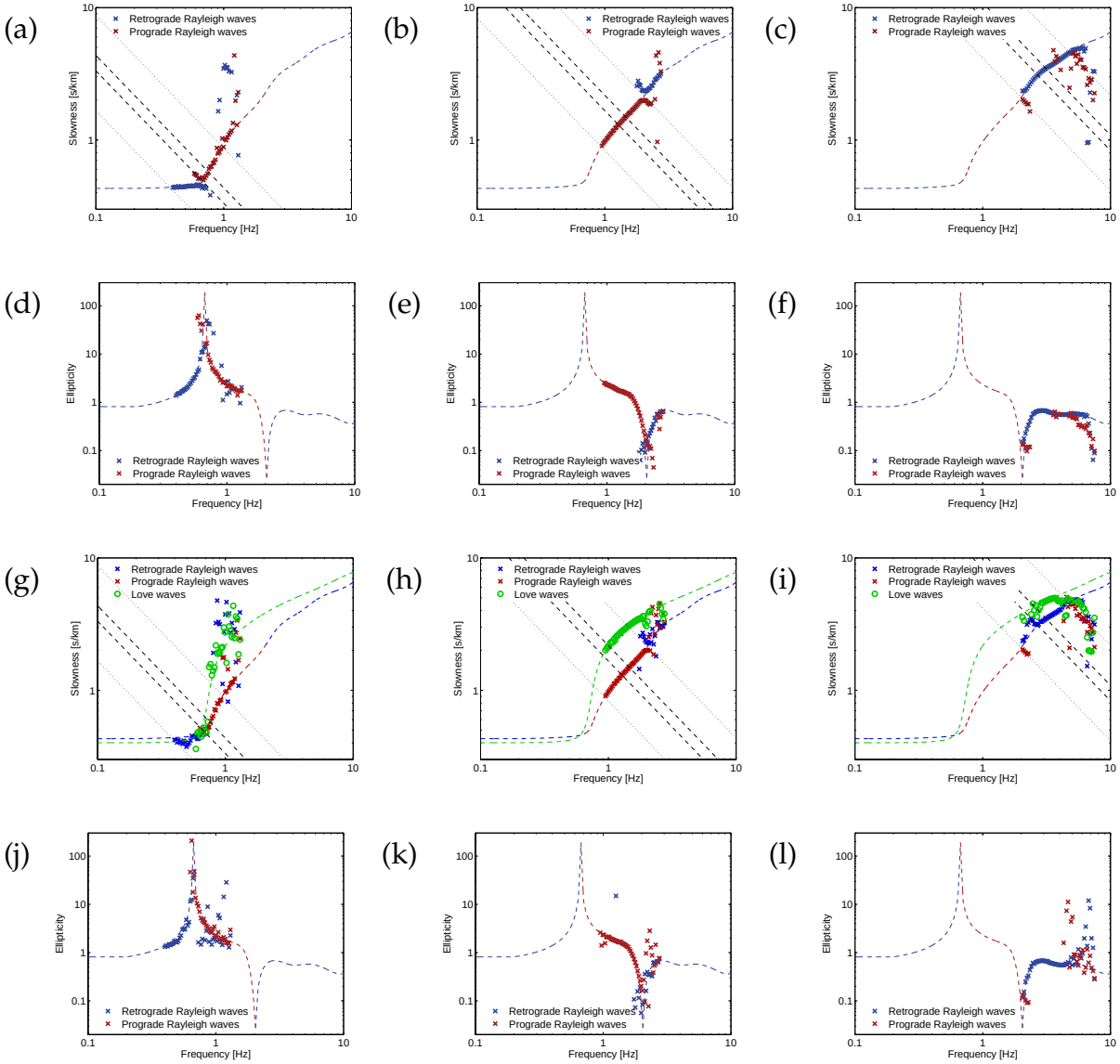


Figure 5.9: Results of applying MUSIQUE to the theoretical signals, using only one ring of sensors plus the central station: dispersion and ellipticity curves for the largest ring without Love waves (a, d) and with Love waves (g, j); The corresponding curves for the medium ring without (b, e) and with Love waves (h, k) and for the smallest ring without (c, f) and with Love waves (i, l). The dashed and dotted lines indicate roughly the resolution limits of the array, and correspond to $s = \frac{1}{2d_{max}f}$, $s = \frac{1}{d_{max}f}$, $s = \frac{1}{2d_{min}f}$ *and* $s = \frac{1}{d_{min}f}$.

5.4 Tests on simulated seismic signals

5.4.1 Signal generation

The following tests are based on a more realistic simulation of real seismic noise signals (body and surface waves), i.e. signals which are combinations of the signals of many sources [Bonnefoy-Claudet et al., 2006]. The signals were simulated using the wavenumber-based technique of Hisada [1994, 1995]. For given source signals and positions, the propagation of seismic waves and their recordings for a given array of seismic sensors are simulated, using the soil structure given in table 5.1. The dispersion and ellipticity curves corresponding to this model are shown in figures 5.10 (a) and (b). The array is composed of three concentric rings of 8 stations each with radii of 50, 200 and 400 meters, respectively, and a central station.
A first test simulates seismic waves arriving from a principal source region. The noise source positions are located inside a circle of 100 meter radius. The center of this circle lies in a distance of 1 250 meters from the array center under an azimuth of 30° (see figure 5.10 (c) for the source locations with respect to the array of seismic sensors).
In a second test, the seismic sources are distributed randomly around the array (see figure 5.10 (d)). The minimum distance of a noise source to the array center is 900 meters, in order to keep a minimum distance of 500 meters between any source and any sensor. In both tests, the source time functions are Dirac peaks. At each source point, a seismic source was simulated once during the total simulation duration of 17.8 minutes. In both tests, a total of 320 sources have been used and two simulations per test have been performed. The first one uses only vertically oriented source forces. Thus, no Love waves are generated. The second one also includes horizontally oriented forces and therefore Love waves.

Table 5.1: Parameters of the model used for the simulation using Hisada's code: Thickness range d, pressure wave velocity V_P, shear wave velocity V_S and density ρ of the layers, and the quality factors Q_P and Q_S with respect to P- and S-waves.

$d[m]$	$V_P[m/s]$	$V_S[m/s]$	$\rho[kg/m^3]$	Q_P	Q_S
0−20	1 500	400	2 000	200	100
20−∞	5 600	3 200	2 000	200	100

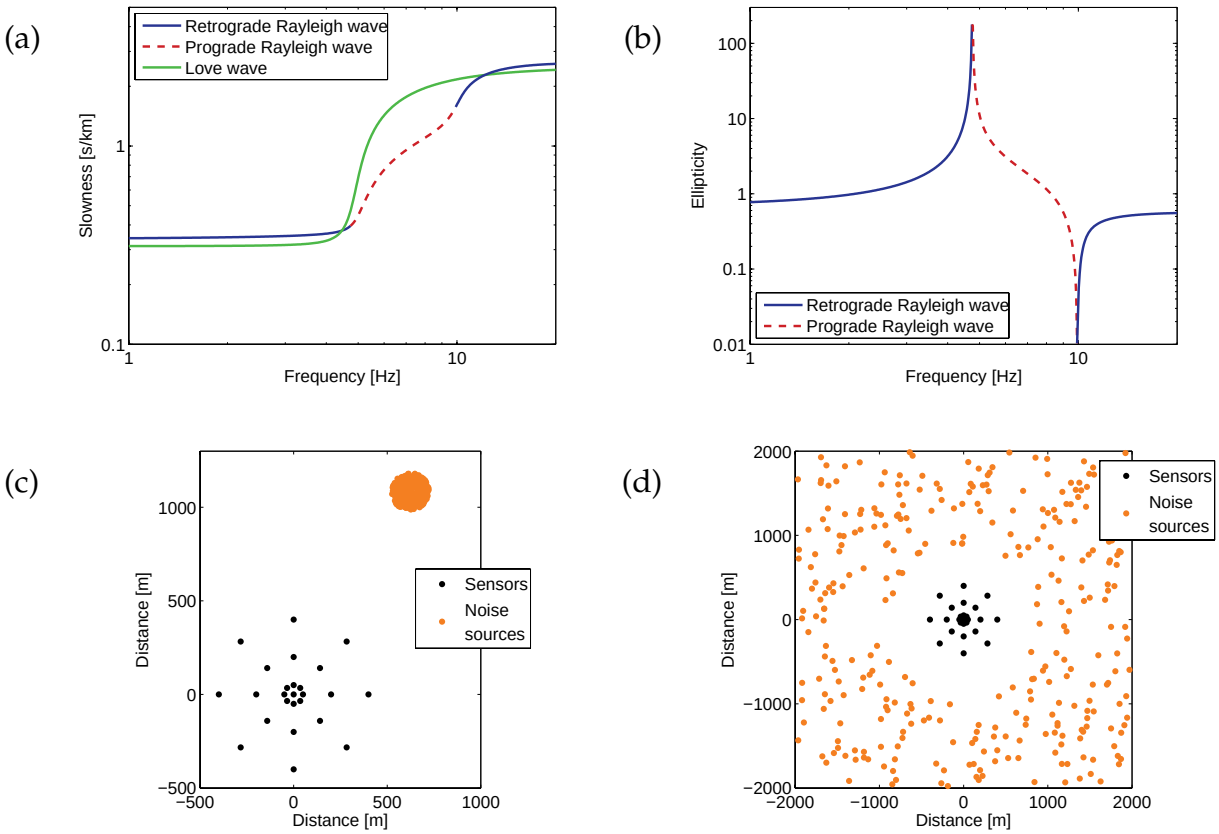


Figure 5.10: (a) The Rayleigh and Love wave dispersion curves and (b) the Rayleigh wave ellipticity curve corresponding to the model given in table 5.1 which was used for the simulation of a transient seismic signal. (c) and (d) Positions of the sources and the array sensors for seismic noise simulations with multiple sources. (a) The sources are located on a circle of 100 radius, which is located 1 250 m from the array center under an azimuth of 30°. (b) The source distribution is random, but no source is closer than 500 m to a seismic sensor.

5.4.2 Analysis results - Noise with a principal azimuth

5.4.2.1 Signals

In this simulation, the signal noise sources are located around a principal azimuth of 30° in a distance of about 1 250 m from the array center. This simulation yields the signals shown in figures 5.11 (a) and (b) at the central station of the seismic array, for the cases without and with Love waves, respectively. The associated spectra of the three components are shown in figures 5.11 (c) and (d), respectively. In the frequency range below 5 Hz, the energy of the signals is much smaller than at higher frequencies. In the case without Love waves, the ellipticity trough at 10 Hz can be seen on both horizontal components. In the case including Love waves, the energy of the horizontal components is larger.

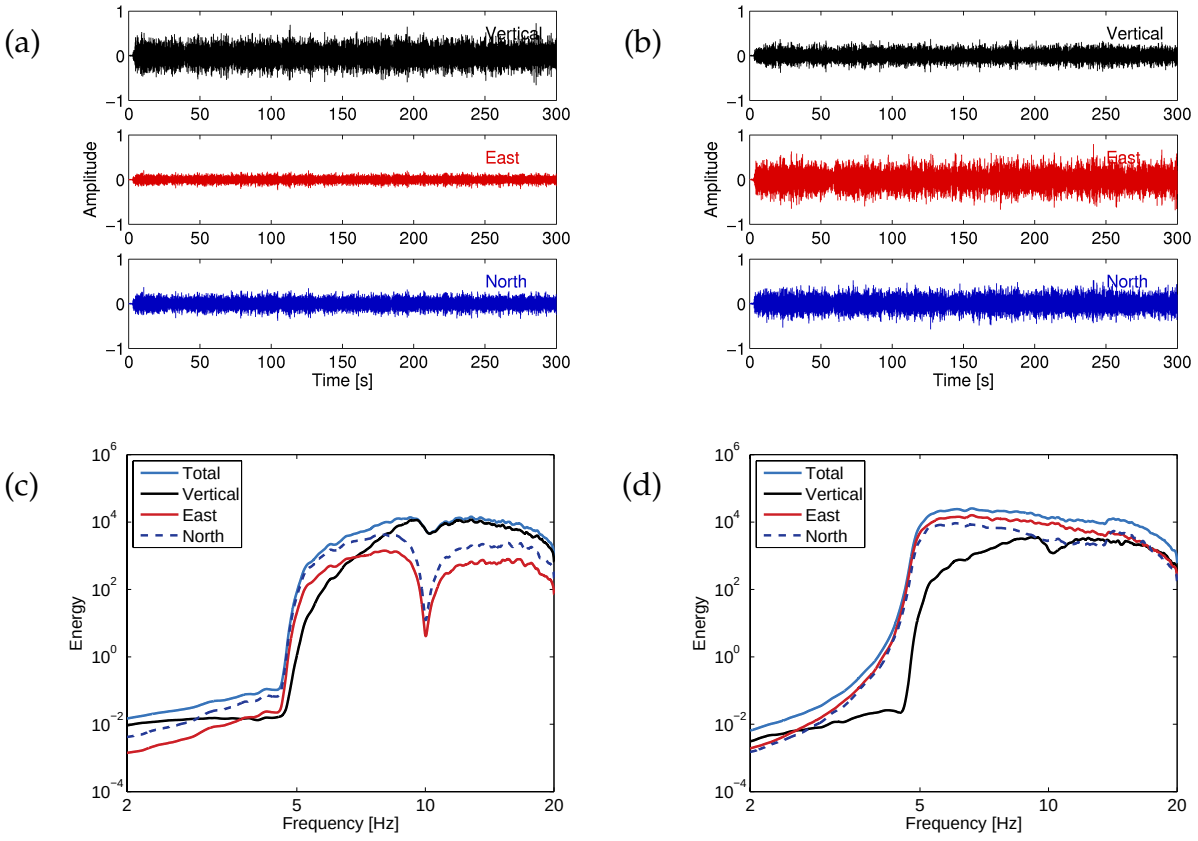


Figure 5.11: The signal simulated at the central station of the seismic array for the test of noise arriving under a principal azimuth (a) for the case without and (b) including Love waves. (c) and (d) The respective spectra of the signals shown in (a) and (b).

5.4.2.2 Single-sensor ellipticity estimation

In figure 5.12, the results of applying the three single-sensor methods H/V, DELFI and RayDec to the signals simulated for the central station are shown. The left column shows the results for the signal withouth Love waves and the right column including Love waves.

H/V
The H/V curve for the case without Love waves (figure 5.12 (a)) fits the theoretical ellipticity curve perfectly above 5 Hz and is in good agreement at the left ellipticity flank. For the case including Love waves (figure 5.12 (b)), the H/V curve is in disagreement with the ellipticity curve in any frequency range and largely overestimates particularly the right flank.

DELFI
The DELFI results using a block size of one period and different values for the filter bandwidth df are shown in figures 5.12 (c) and (d). The curves for both cases fit the right flank of the ellipticity peak and for every value of the frequency bandwidth parameter. However, below the peak frequency and around the trough, the theoretical curve is poorly fitted, especially in the case including Love waves.

RayDec
The results for applying RayDec to the same data are shown in figures 5.12 (e) and (f) for different filter bandwidths. The number of cycles used by RayDec was set to 10. For the case without Love waves (figure 5.12 (e)), the curves are in good agreement with the right flank of the ellipticity peak. For the case including Love waves , the ellipticity values of the right flank are slightly overestimated (figure 5.12 (f)), the results for the largest filter bandwidth still being the closest to the real curve. Below the peak frequency, the ellipticity curve is not fitted in both cases.

Comparison
The results below the peak frequency can be explained by the fact that the energies of the noise signals at low frequencies are very low on all components (see figures 5.11 (c) and (d)). Comparing the right flank of the ellipticity peak, in the case without Love waves, all methods yield accurate results. In the case including Love waves, H/V does not fit the ellipticity curve in a satisfactory way, while the results of DELFI and RayDec are much closer to the right flank of the true curve.

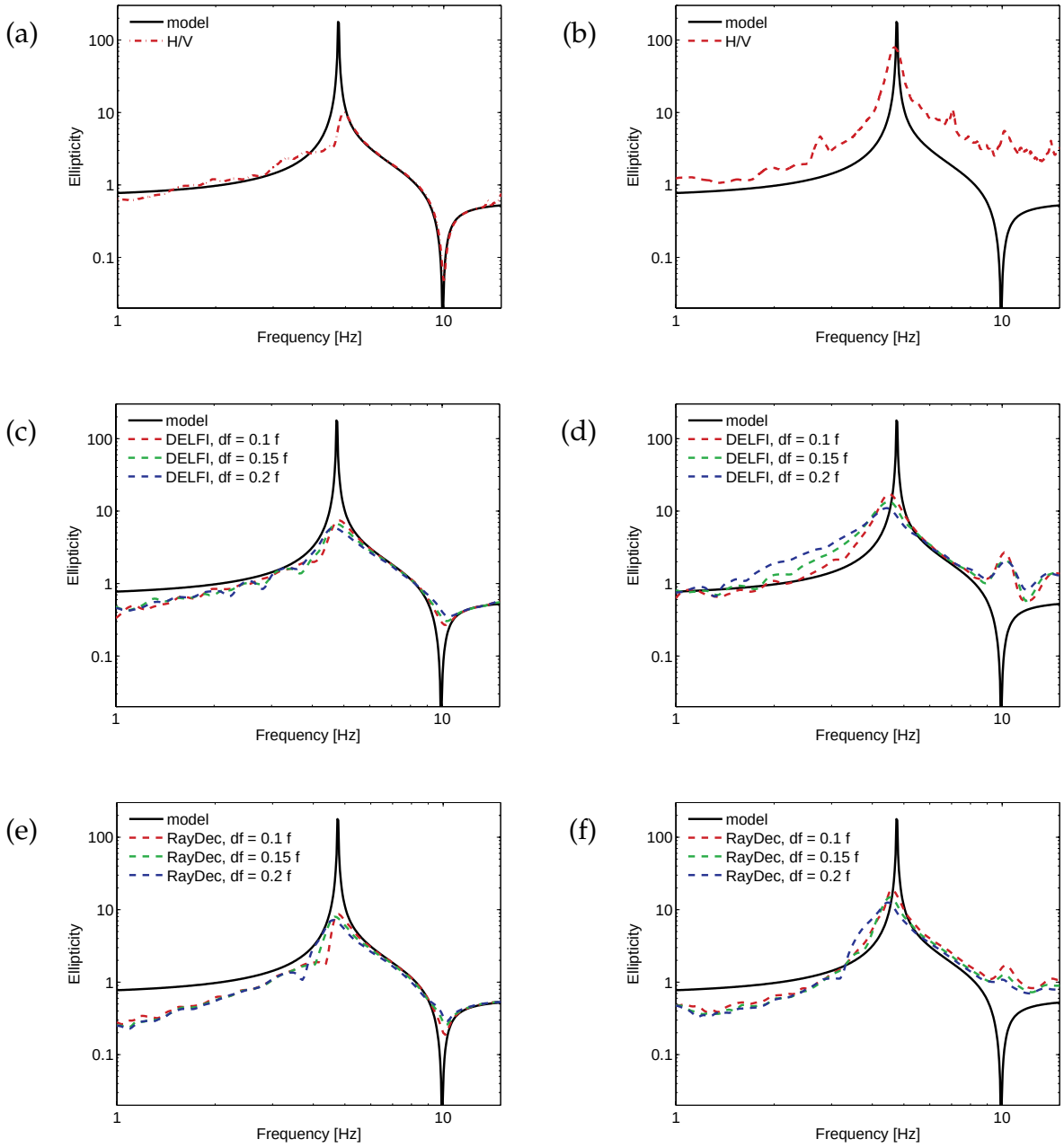


Figure 5.12: H/V curves for the signal simulated at the central array station with sources of a principal azimuth for (a) the case without Love waves and (b) with Love waves. The respective DELFI curves for different filter bandwidths df (c) without and (d) with Love waves. (e) and (f) give the respective curves for RayDec, for different filter bandwidths.

5.4.2.3 MUSIQUE

MUSIQUE has been applied to the simulations using 5 minutes of the signal for all stations of the array simulataneously. At each frequency, the analysis was performed on signal blocks with lengths of three cycles.

Dispersion curves
The resulting dispersion curves are shown in figure 5.13. The different figures show the slowness values for all time windows which have been identified as Love, retrograde or prograde Rayleigh waves, respectively, and are coded as distribution densities. The darker colors indicate that more time windows have been identified around the respective value.
It is evident that in the case with exclusively vertical source polarizations, i.e. without Love waves (left column of figure 5.13), the Love wave dispersion curve cannot be retrieved (figure 5.13 (a)). Only a few signals have been misidentified as Love waves. However, the energy of the respective time windows represents only 0.1 % of the total analyzed energy. This energy corresponds to the total energy of the time windows identified as Love waves and is calculated over the complete analyzed frequency range. The time windows attributed to retrograde and prograde Rayleigh waves account for 15.7 % and 54.9 % of the total analyzed energy, respectively, while the unclassified time windows represent 29.3 %. In the structure model, retrograde Rayleigh waves occur below 5 Hz and above 10 Hz. In figure 5.13 (c), both branches are correctly identified. In figure 5.13 (e), the prograde Rayleigh wave dispersion curve is well found in its respective frequency range.
In the case including Love waves (right column of figure 5.13), the Love wave dispersion curve is overall retrieved (figure 5.13 (b)), but the Rayleigh wave dispersion curve is also seen due to some misidentified Rayleigh waves. The retrograde Rayleigh wave dispersion curve is retrieved below 4 Hz, but at frequencies around 10 Hz, Love waves are identified as Rayleigh waves (figure 5.13 (d)). The prograde dispersion curve is well retrieved, but also a part of the Love waves is misidentified (figure 5.13 (d)). The analyzed time windows attributed to Love waves carry 59.3 % of the total analyzed energy in this case, retrograde Rayleigh waves 7.6 %, prograde Rayleigh waves 16.3 % and 16.8 % are unclassified.
In all figures, misestimated waves are present at high slowness values. Their linear shape in the log-log plot indicates that these are probably aliasing effects. Furthermore, some Love waves are misidentified as Rayleigh waves and vice versa. Therefore, an automatic identification and averaging of the results does not yield reliable results.

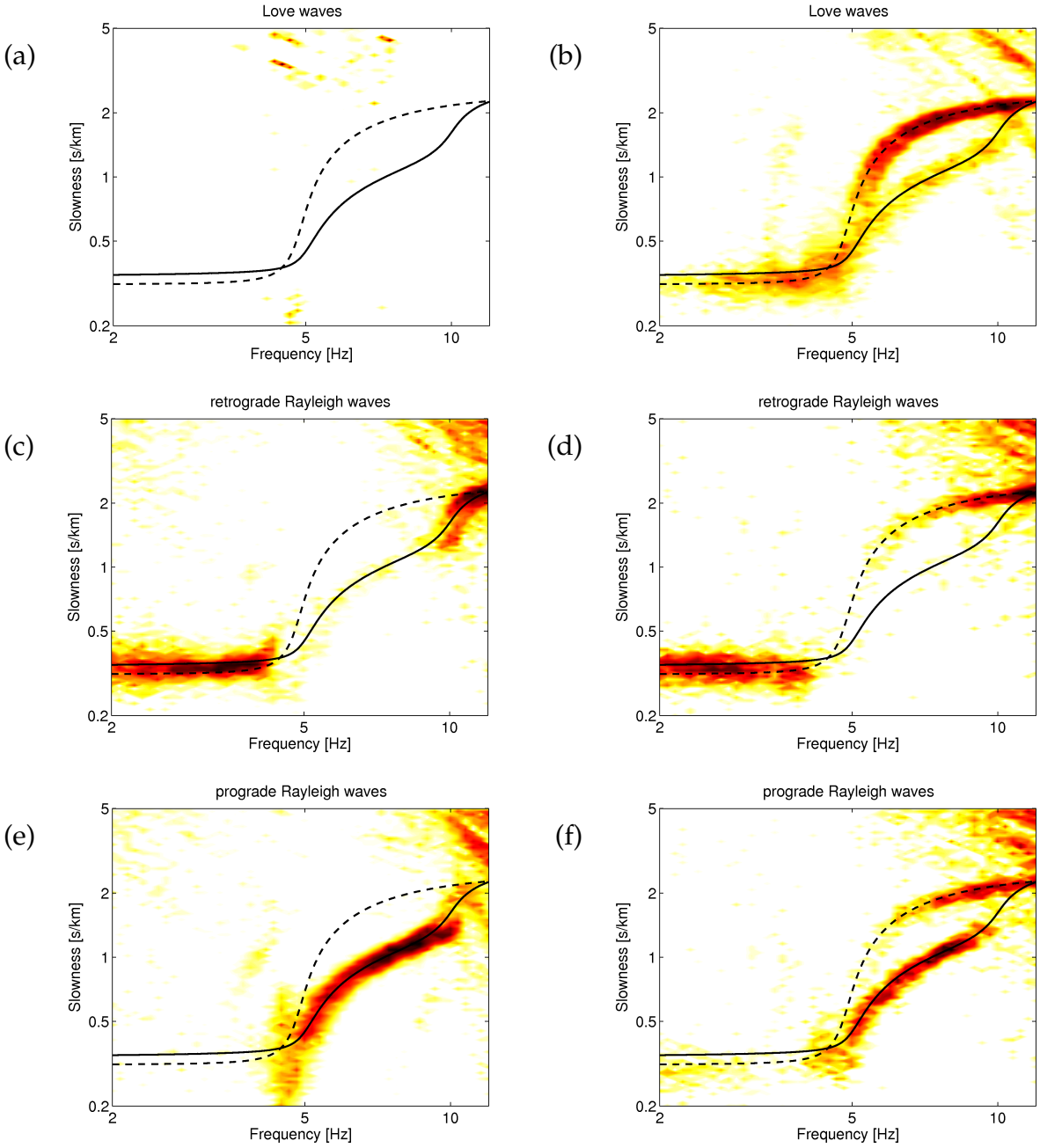


Figure 5.13: Dispersion curves obtained by MUSIQUE for the signals with principal azimuth. (a) and (b) Love wave dispersion curves for the cases without and with Love waves. (c) and (d) Retrograde Rayleigh wave dispersion curves, (e) and (f) Prograde Rayleigh wave dispersion curves for the respective cases. The dark colors indicate regions with many identified data points. The theoretical Love and Rayleigh wave dispersion curves of the model are plotted as dashed and solid black lines, respectively.

Ellipticity curves

The ellipticity values for the time windows identified as retrograde or prograde Rayleigh waves are plotted in figure 5.14. The retrograde Rayleigh wave ellipticities are well retrieved on the left flank of the peak in both cases (figures 5.14 (a) and (c)), but in the case including Love waves, the correct curve is slightly overestimated. Above the trough frequency, the results are very scattered.

At the right flank of the ellipticity peak, the Rayleigh waves are prograde. In the case without Love waves (figure 5.14 (c)), this part of the ellipticity curve is well retrieved. In the case including Love waves (figure 5.13 (d)), the agreement is slightly worse, but the correct curve is still found.

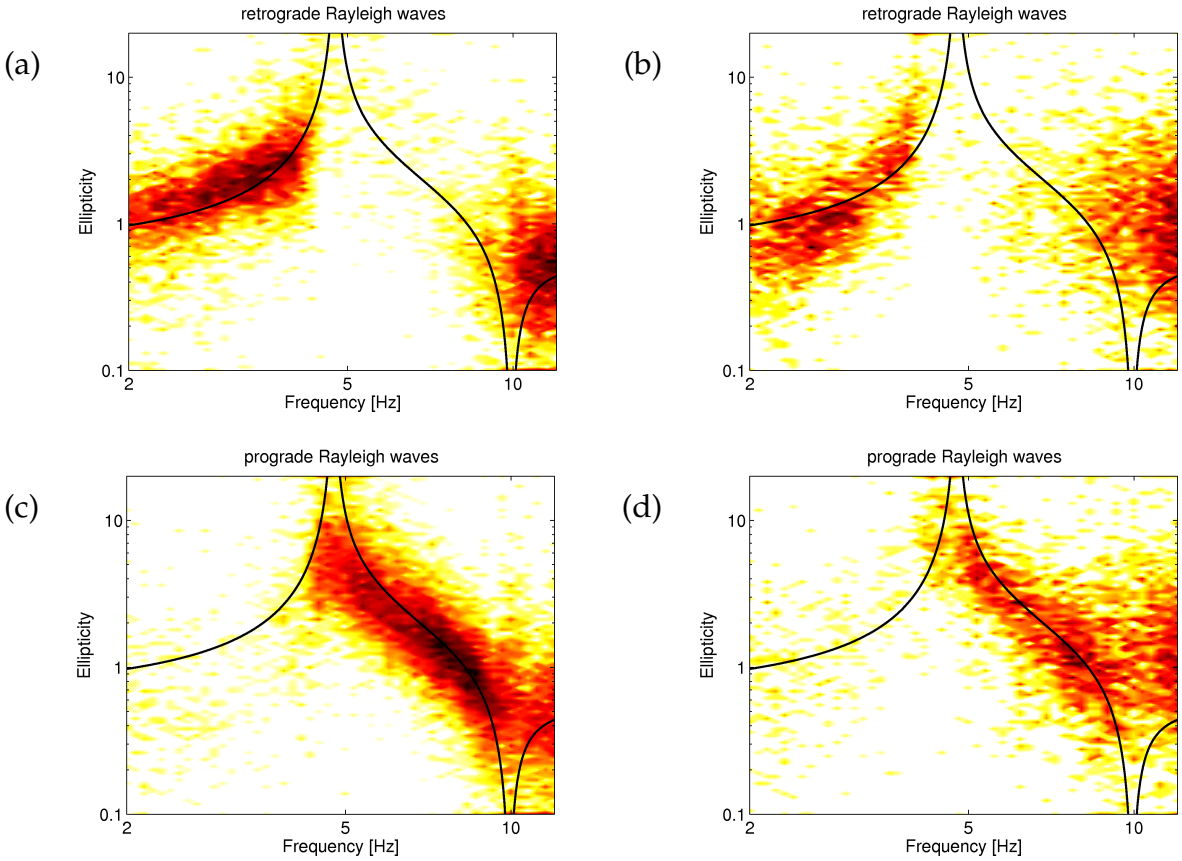


Figure 5.14: Ellipticity curves obtained by MUSIQUE for the signals with principal azimuth. (a) and (b) Retrograde Rayleigh wave ellipticity curves for the cases without and with Love waves. (c) and (d) Prograde Rayleigh wave ellipticity curves for the respective cases. The dark colors indicate regions with many identified data points. The solid black line indicates the theoretical ellipticity curve of the soil model.

5.4.3 Analysis results - Random noise source distribution

5.4.3.1 Signals

In this noise simulation, the locations of the noise sources are random (see figure 5.10 (d)). In figure 5.15, the signals and spectra for the simulation are shown for the central array station. For the case without Love waves, the signal and spectra are given in figures 5.15 (a) and (c), respectively. Figures 5.15 (b) and (d) indicate the signal and spectra for the case including Love waves. In the case without Love waves, the ellipticity trough is visible in the spectra of both horizontal components.

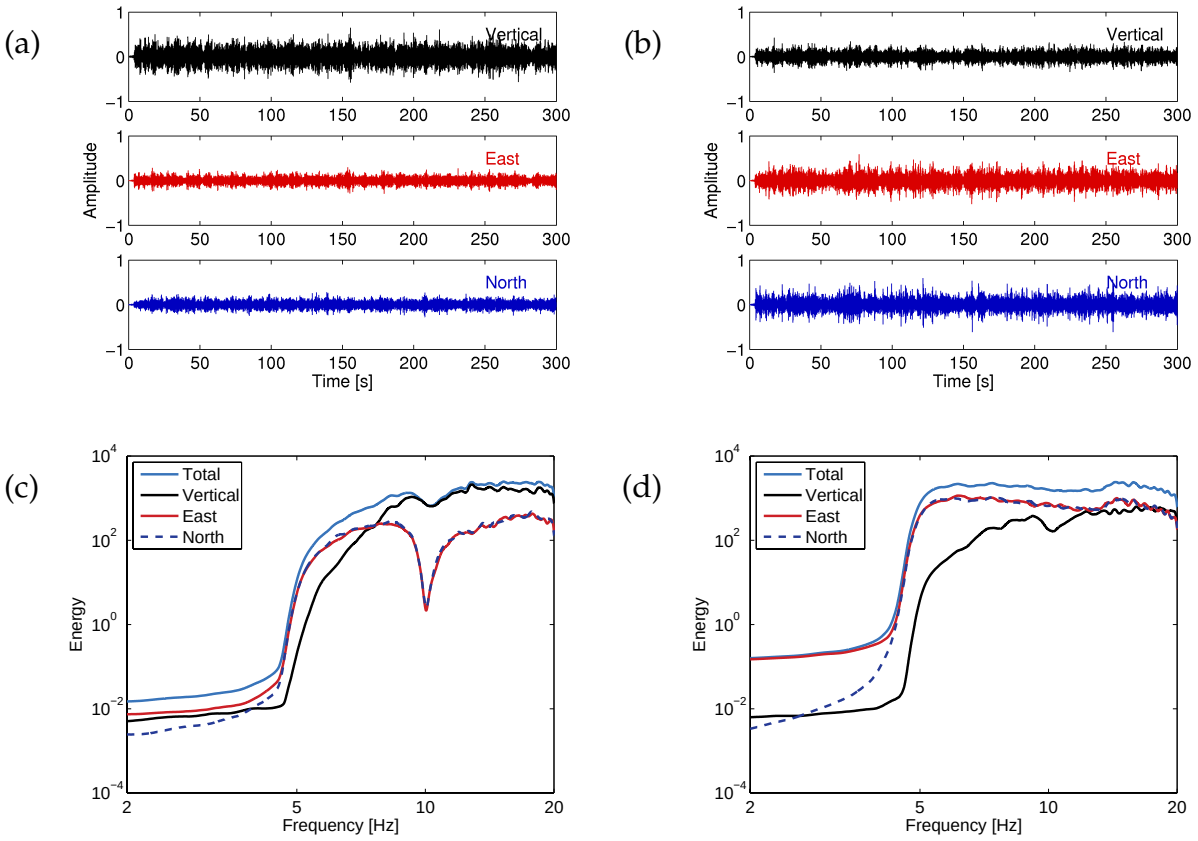


Figure 5.15: The signal simulated at the central station of the seismic array for the test of noise arriving from random positions (a) for the case without and (b) including Love waves. (c) and (d) The respective spectra of the signals shown in (a) and (b).

5.4.3.2 Single-sensor ellipticity estimation

Figure 5.16 shows the results of applying the three single-sensor methods H/V, DELFI and RayDec to the signals simulated for the central station with a random source distribution.

H/V
In the case without Love waves (5.16 (a)), the H/V curve overestimates the actual ellipticity curve over the whole frequency range (except for the peak), but is in acceptable agreement. For the case including Love waves (5.16 (b)), however, the overestimation is very strong over the whole frequency range.

DELFI
The DELFI results using signal blocks of one period and different values for the filter bandwidth df are shown in figures 5.16 (c) and (d). In the case without Love waves, the theoretical ellipticity curve is underestimated. This underestimation is more pronounced at the left flank of the peak, the right flank is actually well retrieved. For the case including Love waves (5.16), DELFI overestimates both the left and the right flank of the peak. No significant differences between the curves corresponding to different filter bandwidths df can be stated.

RayDec
The RayDec curves shown in figures 5.16 (e) and (f) have been obtained by setting the number of cycles used in the method to 10 and varying the filter bandwidth. For the case without Love waves (figure 5.16 (e)), RayDec fundamentally underestimates the actual ellipticity curve over the whole frequency range and for all filter bandwidths df. In the case including Love waves (5.16 (f)), the RayDec curves are in very good agreement with the theoretical ellipticity curve, especially at the left flank. For the right flank, the best agreement with the measurements is obtained for setting $df = 0.15\,f$, but the other values also yield very good results. Around the trough frequency, however, the RayDec curves do not follow the theoretical curve.

Comparison
As a conclusion, it can be stated that, considering the right flank of the ellipticity peak, the best result in the case without Love waves is obtained by DELFI, whereas the best result including Love waves is obtained by RayDec. If Love waves are present, no method succeeds in resolving the trough of the curve. The bad results for frequencies below the peak can be explained by the low energy content of the signal in this range. The underestimation of RayDec (and DELFI) of the right flank in the case without Love waves might be linked to some interaction between different wave types increasing the vertical signal or weakening the horizontal signals.
The correct peak frequency (4.77 Hz) is hardly retrieved by any method. In the case without Love waves, H/V (4.98 Hz) and RayDec (4.86 Hz) overestimate the frequency, while DELFI (4.78 Hz) finds the correct value.

In the case including Love waves, H/V (4.86 Hz) slightly overestimates, while DELFI and RayDec (both 4.64 Hz) underestimate the correct frequency.

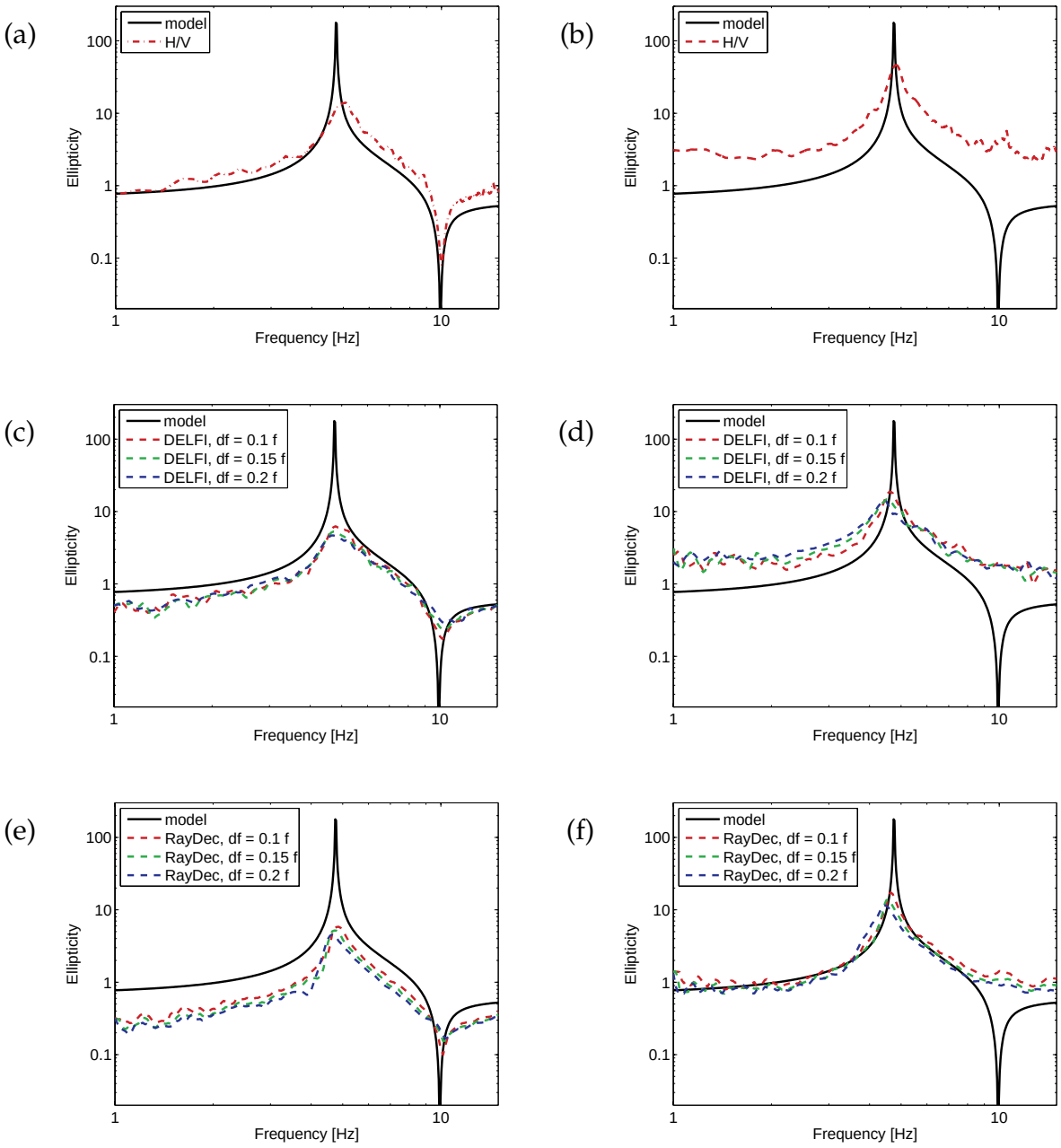


Figure 5.16: H/V curves for the signal simulated at the central array station for random noise source distribution for (a) the case without Love waves and (b) with Love waves. The respective DELFI curves for different filter bandwidths df (c) without and (d) with Love waves. (e) and (f) give the respective curves for RayDec, for different filter bandwidths.

5.4.3.3 MUSIQUE

The data set generated by a random distribution of noise sources has been analyzed by MUSIQUE as well, using 5 minutes of the signal for all stations of the array simultaneously. The length of the signal blocks was set to three cycles for the analysis.

Dispersion curves
The resulting dispersion curves are shown in figure 5.17. The different figures show the slowness values for all time windows which have been identified as Love, retrograde or prograde Rayleigh waves, respectively, and are coded as distribution densities. The darker colors indicate that more time windows have been identified at the respective value.
In the case without Love wave generation (left column of figure 5.17), some prograde Rayleigh waves are misidentified as Love waves (figure 5.17 (a)), but their energy represents only 0.5 % of the total analyzed energy. This energy corresponds to the total energy of the time windows identified as Love waves, covering the complete analyzed frequency range. The time windows attributed to retrograde and prograde Rayleigh waves account for 20.7 % and 32.7 % of the total analyzed energy, respectively, while the unclassified time windows represent 46.1 %. According to the soil structure model, the Rayleigh waves should be prograde in the frequency range between 5 and 10 Hz and retrograde below and above that frequency region. In figure 5.17 (c), the Rayleigh wave dispersion curve is well fitted in the regions corresponding to retrograde motion. The prograde Rayleigh wave dispersion curve is well retrieved in its respective frequency range as well (5.17 (e)).
In the case including Love waves (right column of figure 5.17), the Love wave dispersion curve is very well retrieved in the whole frequency range (figure 5.17 (b)). The retrograde Rayleigh wave dispersion curve is retrieved in the frequency range below 5 Hz (figure 5.17 (d)), but the high-frequency branch corresponds to misidentified Love waves. The prograde Rayleigh wave dispersion curve, however, is well retrieved in its frequency range between 5 and 10 Hz (figure 5.17 (f)), but a second branch of misidentified Love waves is present in this case as well. In this case, time windows attributed to Love waves carry 27.0 % of the complete analyzed energy, retrograde Rayleigh waves 17.4 %, prograde Rayleigh waves 19.7 % and 35.9 % of the energy is unclassified.
All figures exhibit aliasing artifacts, especially for high frequency and slowness values, but to a smaller extent at lower frequencies, too. These artifacts prevent automatic identification and averaging algorithms of the results from being reliable.

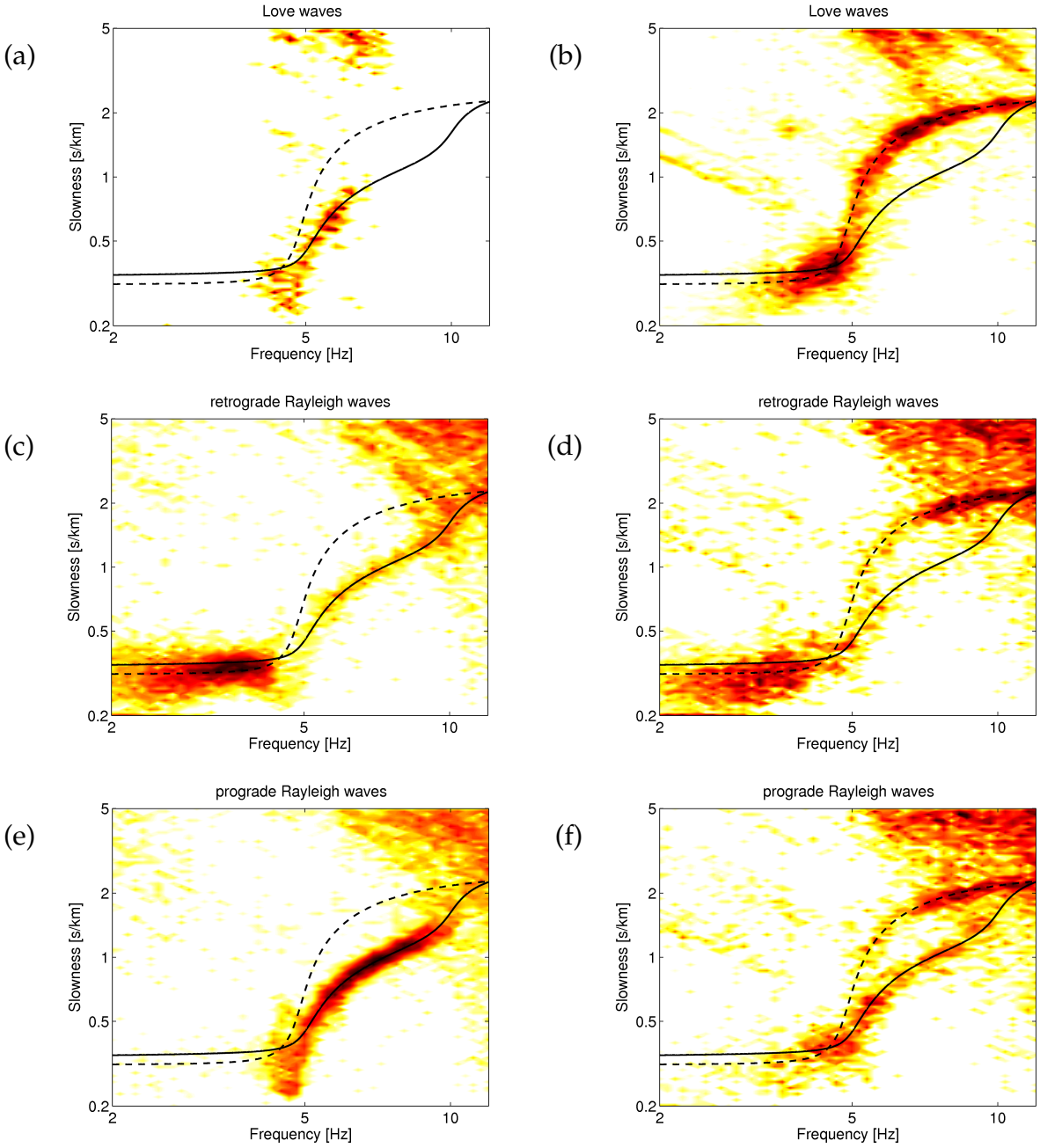


Figure 5.17: Dispersion curves obtained by MUSIQUE for the signals with random noise source distribution. (a) and (b) Love wave dispersion curves for the cases without and with Love waves. (c) and (d) Retrograde Rayleigh wave dispersion curves, (e) and (f) Prograde Rayleigh wave dispersion curves for the respective cases. The dark colors indicate regions with many identified data points. The theoretical Love and Rayleigh wave dispersion curves of the model are plotted as dashed and solid black lines, respectively.

Ellipticity curves

In figure 5.18, the ellipticity values for the time windows identified as retrograde or prograde Rayleigh waves are plotted. In the case without Love waves, the retrograde Rayleigh wave ellipticity is quite well retrieved at the left ellipticity peak (figure 5.18 (a)), but even in the case including Love waves, this part of the ellipticity curve is well fitted (figure 5.18 (b)).

The prograde part, i.e. the right flank of the Rayleigh wave ellipticity peak, is well retrieved in the case without Love wave generation (figure 5.18 (c)). In the case including Love waves (figure 5.18 (d)), however, it is not retrieved at all.

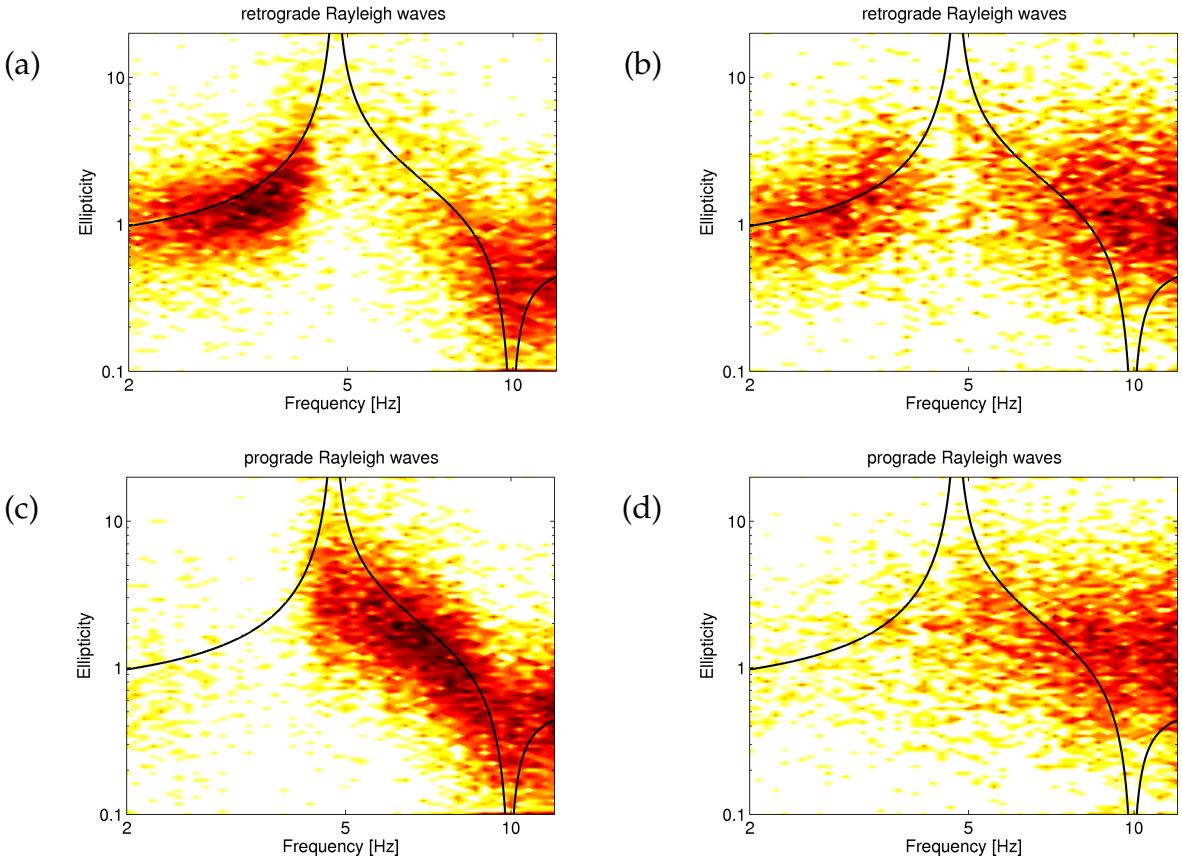


Figure 5.18: Ellipticity curves obtained by MUSIQUE for the signals with random noise source distribution. (a) and (b) Retrograde Rayleigh wave ellipticity curves for the cases without and with Love waves. (c) and (d) Prograde Rayleigh wave ellipticity curves for the respective cases. The dark colors indicate regions with many identified data points. The solid black line indicates the theoretical ellipticity curve of the soil model.

5.5 Conclusion

In this chapter, the performances of the newly developed methods DELFI, RayDec and MUSIQUE have been tested on synthetic data sets and compared to the results of the H/V method.

Test results
The first test investigated the robustness of the methods' results against additional white noise. For MUSIQUE, the azimuth estimation is more robust than the slowness estimation. The misestimation of both influences the subsequent polarization parameter estimation. For this estimation, the phase difference parameter between the horizontal and vertical components is more robust than the ellipticity estimation. For DELFI and RayDec, adding white noise tends to underestimate ellipticity values above $\sqrt{2}$ and to overestimate lower ellipticity values. Compared to DELFI, RayDec is more robust against the addition of white noise.

The next tests compared the estimations of the different methods for signals of increasing complexity.
For a simple theoretical signal, DELFI, RayDec and MUSIQUE retrieved the correct ellipticity curve (or at least the right flank of the peak). H/V only retrieved the correct curve in the case without Love waves. Furthermore, MUSIQUE correctly identifies the different wave types and discriminates between retrograde and prograde Rayleigh waves.
For multiple signals with a principal direction of arrival, H/V only fits the ellipticity curve if no Love waves are present, while RayDec and DELFI correctly estimate the right flank of the ellipticity peak, but not the rest of the curve. MUSIQUE identifies the different wave types and estimates correct dispersion curves, as well as the ellipticity curve.

The final test simulated real seismic noise with waves arriving from different azimuths. Of the single-sensor methods, only DELFI gives a correct ellipticity estimation of the right flank if no Love waves are present. In the general case with Love waves, the only method to retrieve the correct right ellipticity flank is RayDec. In this case, MUSIQUE correctly identifies the different wave types and estimates the respective dispersion curves. Although the right ellipticity peak is only retrieved if no Love waves are present, the distinction between prograde and retrograde Rayleigh waves helps in identifying their respective parts of the dispersion curve.

Field of application of the different methods
According to their respective qualities, the different methods should be used in different situations. The mono-sensor methods RayDec and DELFI can be used to characterize ambient seismic vibrations. As RayDec is based on the statistics of summing a multitude of signal blocks, it is especially suited for these investigations. The minimum required signal length depends on the frequency range of interest. Nevertheless, as less as five minutes of signal can be sufficient for reliable results[V]. However, the method is not suitable for signals of short duration, e.g. earthquake data. DELFI is less robust against additional Love wave contributions than RayDec and gives results which are similar to H/V for ambient seismic vibrations which are composed of both Rayleigh and Love waves with a random azimuth distribution. Therefore, the method should only be applied to pure Rayleigh wave data if it is supposed to measure ellipticity. Nonetheless, by applying both RayDec and DELFI to the same measurements of ambient seismic vibrations, the composition of the seismic wave field, i.e. the relation between Love and Rayleigh wave energy, can be estimated using a single seismic sensor.

As MUSIQUE uses array data, it can discriminate between retrograde and prograde Rayleigh waves, which is not possible with single-sensor methods. Furthermore, it allows the estimation of azimuth and slowness, i.e. the wave vector, of incoming waves. The correct identification between retrograde and prograde Rayleigh waves separates the Rayleigh wave dispersion curve in two distinct branches. This distinction is not possible with other established methods. However, for real ambient seismic vibration data, single-sensor methods should be preferred to estimate ellipticity. Although MUSIQUE can be applied to long-duration recordings, it requires a lot of computation time, notably because of the repeated use of Fourier transforms and eigenvalue decompositions. Therefore, it is more convenient for short data sets, e.g. earthquake data which consist of transient seismic signals.

The following part of this work will deal with the application of the developed methods to real data measurements. Chapter 6 deals with the theoretical aspects of the inversion of ellipticity curves in order to find the local soil structure. The obtained results will then be applied to real seismic data measurements in chapter 7, where RayDec is used for the ellipticity estimation. Finally, in chapter 8, real earthquake data measured in California will be analyzed using MUSIQUE.

[V] The necessary signal length depends on the frequency range of interest, five minutes are sufficient for frequencies above about 0.3 Hz.

Part II

Application

Chapter 6

Inversion of ellipticity curves - Theoretical aspects

In previous chapters, methods to retrieve the ellipticity of Rayleigh waves have been introduced (DELFI in section 2.3, RayDec in section 2.4 and MUSIQUE in section 4.6) and tested (chapter 5). The ellipticity will now be used to retrieve the soil structure by inversion, using SPAC measurements (section 3.5) as additional information to constrain the superficial structure. In this chapter, the fundamentals of the inversion of ellipticity curves will be developed. The following text represents the article "Inversion of Rayleigh wave ellipticity measurements, Part I: Theoretical aspects" by Hobiger, Cornou, Bard, Le Bihan, and Wathelet, 2010a, which was submitted to *Geophysical Journal International*.

Dans les chapitres précédents, des méthodes pour déterminer l'ellipticité des ondes de Rayleigh ont été introduites (DELFI en 2.3, RayDec en 2.4 et MUSIQUE en 4.6) et testées (chapitre 5). Maintenant, l'ellipticité sera utilisée pour retrouver la structure du sol par inversion, contraignant la structure superficielle par des mesures SPAC (cf. 3.5) additionnelles. Dans le chapitre suivant, les aspects théoriques de l'inversion de courbes d'ellipticité seront développés. Le texte du chapitre correspond à l'article "Inversion of Rayleigh wave ellipticity measurements, Part I: Theoretical aspects" par Hobiger, Cornou, Bard, Le Bihan, and Wathelet, 2010a, qui a été soumis à *Geophysical Journal International*.

6.1 Abstract

The knowledge of the local soil structure is important for the assessment of seismic hazards. A widespread, but expensive and time-consuming technique to retrieve the parameters of the local underground is the drilling of boreholes. Furthermore, the shear wave velocity profile at a given location can be obtained by inversion of surface wave dispersion curves. In order to ensure a good resolution for both superficial and deeper layers, the used dispersion curves need to cover a wide frequency range. This wide frequency range can be obtained using several arrays of seismic sensors or a single array composed of a large number of sensors. Consequently, these measurements are time-consuming. A simpler alternative is provided by the use of the ellipticity of Rayleigh waves. The frequency dependence of the ellipticity is tightly linked to the shear wave velocity profile. Furthermore, it can be measured using a single seismic sensor. As soil structures which are obtained by scaling of a given model exhibit the same ellipticity curve, an inversion of the ellipticity curve alone will always yield ambiguous results. Therefore, additional measurements which fix the shear wave velocity profile at some points have to be included in the inversion process. Small-scale spatial autocorrelation measurements or MASW measurements can provide the needed data. We will show which parts of the ellipticity curve have to be included in the inversion process to get a reliable result and which parts can be omitted. Furthermore, the use of autocorrelation or high-frequency dispersion curves will be highlighted and the bias on the inversion results which is produced by incorrect ellipticity curve measurements is shown. The rules for ellipticity inversion resulting from these studies are applied to real data measurements in an accompanying article.

6.2 Introduction

In order to assess the seismic hazard of a given site, the knowledge of the site conditions is important. The local soil can be investigated by different means. For example, the drilling of boreholes gives a good view of the ground structure, but is a time-consuming and expensive task. Refraction or reflection measurements using explosive sources are difficult to be carried out in urban areas. On the contrary, active sources like sledgehammering can be used in urban areas, but generate waves only in a high frequency range which does not penetrate the soil very deeply. Therefore, within agglomerations, the recording of seismic ambient vibrations is appropriate. Arrays of seismic sensors can be easily deployed and dispersion curves of surface waves can be measured. Then, the local soil structure is obtained by inverting the dispersion curves. However, in order to cover a broad frequency band, several seismic arrays have to be deployed [Wathelet et al., 2008]. As a result, such measurements are time-consuming. In urban areas, the existing infrastructure limits the possible deployment locations of the seismic sensors. Therefore, using a lower number of sensors is tempting.
The ellipticity of Rayleigh waves, i.e. the ratio between the horizontal and the vertical movement, strongly depends on the local soil structure and carries information on the complete sedimentary layers. The possibility of using ellipticity measurements to con-

strain the soil structure was already shown by Boore and Toksöz [1969]. A widespread technique for investigating the soil structure is the classical H/V technique [Nogoshi and Igarashi, 1971, Nakamura, 1989]. The H/V ratio simply indicates the spectral ratio between the horizontal and the vertical components. If the wave field was composed exclusively of single mode Rayleigh waves, the H/V ratio would indicate their ellipticity. In general, however, the seismic noise field is composed of various modes of Rayleigh waves, Love waves and body waves. The composition of the wave field depends on the noise sources and the site structure [Bonnefoy-Claudet et al., 2008]. However, the measurement of the ellipticity is not trivial in seismic noise recordings, since Love and body waves contribute to the wave field as well [Köhler et al., 2006, Endrun, 2010]. Therefore, all these waves contribute to the H/V ratio. As Love waves are only present on the horizontal components, the H/V ratio will, in general, give an overestimation of ellipticity.

By using assumptions on the wave field composition, the H/V curve can be corrected for Love wave contributions and the resulting curve can be inverted as an ellipticity curve [Fäh et al., 2001, 2003]. Scherbaum et al. [2003] showed that the inversion of Rayleigh wave dispersion curves can be improved by adding H/V peak and trough frequencies. Arai and Tokimatsu [2004] used H/V curves to fix the exact depths of shallow layers for sites where the shear wave velocities were known by borehole measurements. Parolai et al. [2005] inverted Rayleigh wave dispersion curves jointly with H/V curves. A similar study was performed by Picozzi et al. [2005], who concluded that an inversion of the dispersion curve constrains the sedimentary velocity profile, whereas an inclusion of the H/V curve sets additional constraints on the bedrock depth and velocity. This was confirmed by Arai and Tokimatsu [2005], who also showed the benefits of including the H/V curve in the inversion of Rayleigh wave dispersion curves. Furthermore, they proposed to use H/V data and small-scale array measurements instead of large-scale array measurements which are necessary to obtain broad-band dispersion curves.

However, inversions of H/V curves always include prior assumptions on the composition of the wave field, especially the ratio of Rayleigh to Love waves. Recent studies have shown that this ratio is, even for a given site, not a constant and varies with frequency and time [Köhler et al., 2006, Endrun, 2010].

Newly developed techniques allow the ellipticity of Rayleigh waves to be directly retrieved from seismic noise measurements. The RayDec method [Hobiger et al., 2009a] uses the random decrement technique [Asmussen, 1997] which is usually applied to measure the resonance frequencies and damping parameters of buildings [Dunand, 2005, Michel et al., 2008]. By stacking large numbers of vertical and horizontal signals and correlating them in the way typical for Rayleigh waves, Love and body wave contributions are efficiently suppressed. Poggi and Fäh [2010] proposed a method that uses high resolution f-k analysis in order to extract ellipticity from seismic arrays of about 10 sensors. With this method, even higher modes can be identified.

However, ellipticity measurements alone are never sufficient to retrieve the local soil structure. By scaling both the wave velocity and the depth values of a given soil structure model by the same factor, the ellipticity function remains the same [Scherbaum et al., 2003]. The ellipticity curve fixes only the shape of the shear wave velocity structure, but not the scale.

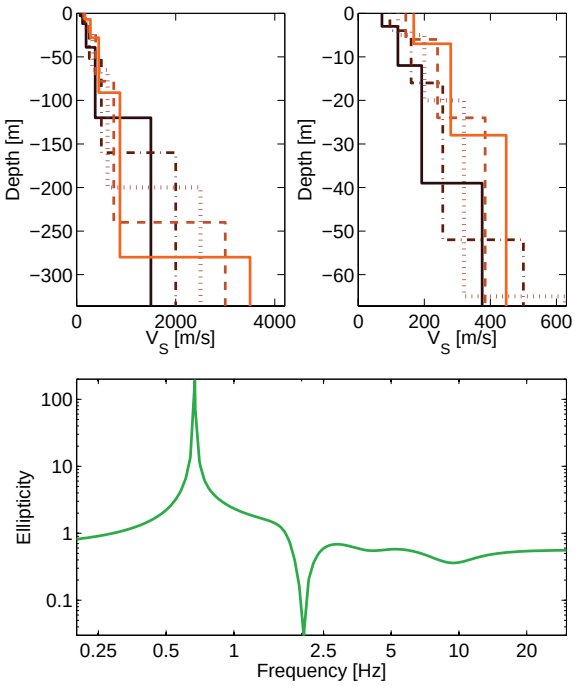


Figure 6.1: (Top left) Five different soil models which can be transformed one into another by multiplying the depth, shear and pressure waves by the same factor. (Top right) Zoom on the upper layers. (Bottom) The five models have the same corresponding ellipticity curve.

In figure 6.1, different shear wave velocities of soil models are shown that have the same ellipticity curve. These models were obtained by scaling the depth, shear and pressure wave velocities by the same factor. It is evident that an ellipticity inversion alone would yield ambiguous results by fixing the shape of the velocity profile only. If the superficial structure, however, is constrained precisely by additional measurements, the ellipticity can be used to constrain the deeper part of the structure.
Therefore, additional measurements have to be performed to fix the absolute velocity values. Small-scale spatial autocorrelation measurements [Aki, 1957] using a small number of sensors are sufficient to fix the shear wave profile at the surface. Another way to determine the superficial structure consists in MASW measurements [Park et al., 1999, Socco and Strobbia, 2004].
Although many studies have shown the utility of including H/V or ellipticity measurements in dispersion curve inversions, no systematic investigation has been published on which parts of the ellipticity curve are carrying the important information on the ground structure and which types of additional measurements fixing the shallow structure have to be included. This paper will focus on the theoretical aspects of inversions including the Rayleigh wave ellipticity and show how the soil structure can be retrieved. Therefore, the use of spatial autocorrelation as well as dispersion curve measurements for fixing the shallow structure will be shown.
Then, the question concerning which parts of the ellipticity curve have to be included in the inversion process and which parts can be omitted will be answered for models

with and without singularities in the ellipticity curve, i.e. with or without large velocity contrast at some depth. Finally, it will be shown how misestimations of the ellipticity curve influence the inversion results. The identified rules for ellipticity inversions will be applied to real data measurements in a companion paper.
As the ellipticity depends very strongly on the shear wave profile, but only slightly on the pressure wave velocities [Arai and Tokimatsu, 2004], the goal of the ellipticity inversion is to retrieve the shear wave velocity profile only. Therefore, in the following the pressure wave velocity profiles will be disregarded.

6.3 Inversion algorithm

The inversions are performed using the neighborhood algorithm which was proposed by Sambridge [1999a,b] and modified by Wathelet [2008]. For a random initial set of models, the associated misfit values are calculated. The next generation of models is located in the neighborhood of the models with the lowest misfit values. In this way, the search is guided by the best models without neglecting the possible existence of better-fitting models further away. The associated misfit value for a model with associated data M_i is calculated in the following way:

$$misfit = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{D_i - M_i}{\sigma_i} \right)^2}, \tag{6.1}$$

if the measured data to be inverted is given by a set of N data points D_i ($1 \leq i \leq N$) with associated measuring errors σ_i. For the inversions of theoretical data presented in the following, we assume equal measuring errors for each data point. As this results in a scaling of the misfit value, we can arbitrarily chose the value of σ_i and fix it, for the sake of simplicity, to 1. If two different data sets, for example an autocorrelation and an ellipticity curve, are to be inverted jointly, a separate misfit value will be calculated for each of them and the final misfit value will be a weighted average of both values. In the following, we will impose equal weights on both information. Hence, the final misfit value is the mean of both misfit values. It should also be noted that the two different data sets can be given in different frequency ranges.

6.4 Tests on theoretical data

This section addresses the following theoretical aspects of the inversion of ellipticity curves:

- By which additional measurements can the velocity profile be fixed at the surface?
- Which parts of the ellipticity curve have to be included in the inversion process and which parts can be omitted without deteriorating the results?
- How do erroneous ellipticity measurements influence the inversion results?

6.4.1 Model parameters

Table 6.1: Table of the properties of the used soil profiles: model A (top) and model B(bottom).

model A	Thickness [m]	Depth range [m]	$V_P[m/s]$	$V_S[m/s]$	$\rho[kg/m^3]$
layer 1	5	0−5	540	120	1800
layer 2	15	5−20	900	200	1800
layer 3	45	20−65	1440	320	1800
layer 4	135	65−200	2810	625	1800
layer 5	∞	>200	6250	2500	2000

model B	Thickness [m]	Depth range [m]	$V_P[m/s]$	$V_S[m/s]$	$\rho[kg/m^3]$
layer 1	5	0−5	540	120	1800
layer 2	15	5−20	900	200	1800
layer 3	45	20−65	1440	320	1800
layer 4	135	65−200	2430	540	1800
layer 5	∞	>200	2520	840	2000

In order to investigate the aforementioned questions, we performed inversion tests on theoretical data obtained on a sample soil profile. For a model with singularities (peaks and troughs) in the ellipticity (model A, see figures 6.2 (a) and (b)), we used a simplified version of the soil structure model of the Euroseistest site in Volvi, Greece [Raptakis et al., 1998]. We fixed the number of layers to four overlying a homogeneous half-space. By changing the shear and pressure wave velocities of the fourth and fifth layers of model A, and consequently reducing the impedance contrast, we obtained model B (figures 6.2 (c) and (d)) which does not exhibit ellipticity singularities. Table 6.1 indicates the properties of both soil profiles.

6.4.2 Goodness of the inversion results

For the known sample soil models, the exact theoretical ellipticity, dispersion and spatial autocorrelation curves can be calculated directly. As a result, these curves do not have error bars. In equation (6.1), a constant error value of 1 was attributed to each data point. As the inversions in this section are performed for theoretical curves, we know the exact soil structure and can therefore investigate directly how good the model is fitted. Therefore, it is desirable to define a parameter describing the goodness of fit. This parameter should be applicable for real data inversions as well.
The goodness of fit of the resulting soil structure models could be determined by summing the squared distances between the obtained model's shear wave velocity and the true shear wave velocity profile for all depths. However, this way of doing has two disadvantages. Firstly, in general, the reference model and the inverted model do not have the same bedrock depth. The choice of the depth down to which the models would

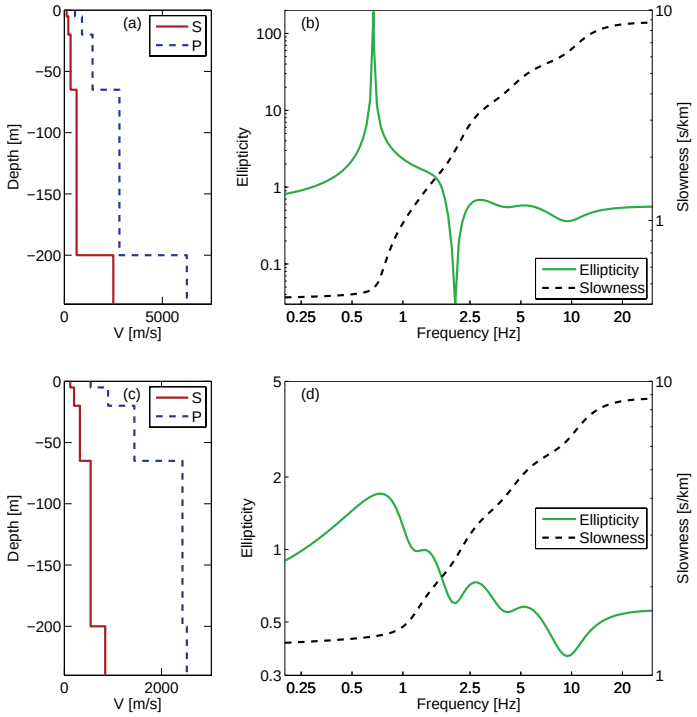


Figure 6.2: (a) Pressure and shear wave velocities of the sample ground structure profile of model A. (b) Corresponding dispersion and ellipticity curves of model A. (c) Pressure and shear wave velocity profiles of model B. (d) Dispersion and ellipticity curves for model B.

be compared would therefore over- or underemphasize the bedrock shear wave velocity. Secondly, for an inversion of real data measurements, the soil structure model is not exactly known. Therefore, a better way is to compare the Rayleigh wave dispersion curve calculated for the inverted models with a reference dispersion curve. This reference dispersion curve can be calculated for the given theoretical model or, in the case of real data measurements, directly measured at the surface.
Similarly to the definition of the misfit value, we define a proximity value comparing the inverted and the theoretical dispersion curves at N frequency samples in the following way:

$$T = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{s_{mod}(f_i) - s_{inv}(f_i)}{s_{mod}(f_i)}\right)^2}. \tag{6.2}$$

Here, $s_{inv}(f)$ indicates the dispersion curve for the inverted models and $s_{mod}(f)$ the theoretical dispersion curve for the given structure where the dispersion curves are expressed in slowness. As the theoretical models do not have measuring errors, σ_i from equation (6.1) is is replaced by $s_{mod}(f_i)$ in equation (6.2). In this way, T gives a sort of mean deviation of the best inversion model from the reference model and describes the goodness of fit. An inversion can have a low misfit value, but it is T which indicates if the inversion really fits the theoretical model. The calculation of the T value does not necessarily span the same frequency range as used for the inversion process.

For the theoretical models, we define that values of T below 0.1 are acceptable and that values below 0.05 indicate a good fit.

6.4.3 Fixing the shear wave velocity at the surface by spatial autocorrelation curves

Table 6.2: Misfit and T values for the best-fitting model of the inversions of autocorrelation curves corresponding to arrays of 5 m, 10 m and both 5 and 10 m radius.

Reference figure	Autocorrelation curve Array radius	Frequency range	Number of models	Minimum misfit	T value (0.7 - 30 Hz)	T value (2 - 30 Hz)
Fig. 6.3 (a)	5 m	2 - 30 Hz	50 100	$0.51 \cdot 10^{-3}$	0.132	0.010
Fig. 6.3 (b)	10 m	2 - 30 Hz	50 100	$0.68 \cdot 10^{-3}$	0.044	0.001
Fig. 6.3 (c)	5 m + 10 m	2 - 30 Hz	50 100	$0.64 \cdot 10^{-3}$	0.020	0.001

For the sample profile A, theoretical spatial autocorrelation curves for the vertical component have been calculated for sensors on circles of 5 and 10 m radius between 2 and 30 Hz. As the models A and B differ only below depths of 60 meters, the autocorrelation curves for both models are identical within this frequency range. Figure 6.3 (a) shows the results of the inversion of the autocorrelation curve only (without considering ellipticity information) for a radius of 5 m, (b) for 10 m and (c) for a joint inversion of the 5- and 10-meter curves. The three inversions (50 100 models per run) reproduce the shear wave velocity profile in a comparable way. The misfit and T values for the best-fitting model of the three inversions are shown in table 6.2. The minimum misfit values are low, indicating that the autocorrelation curves have been well fitted. The superficial layers are well constrained and fit the true model in all three cases equally well. The shear wave velocities are, however, very badly constrained at depths below 40 meters. When regarding the corresponding Rayleigh wave dispersion curves, the higher frequency part above 2 Hz is well fitted for all three inversions. This can also be seen in the very low T values for the frequency range between 2 and 30 Hz. At low frequencies, on the contrary, the dispersion curves for comparable misfit values are very scattered. The T values which are calculated between 0.7 and 30 Hz indicate that the 5-meter autocorrelation constrains the dispersion curve at lower frequencies worse than the 10-meter curve, but a joint inversion of both curves yields the best result. However, by looking at the dispersion curve and velocity profile of this inversion, it is evident that the deeper structure cannot be retrieved by using small-scale autocorrelation curves only.

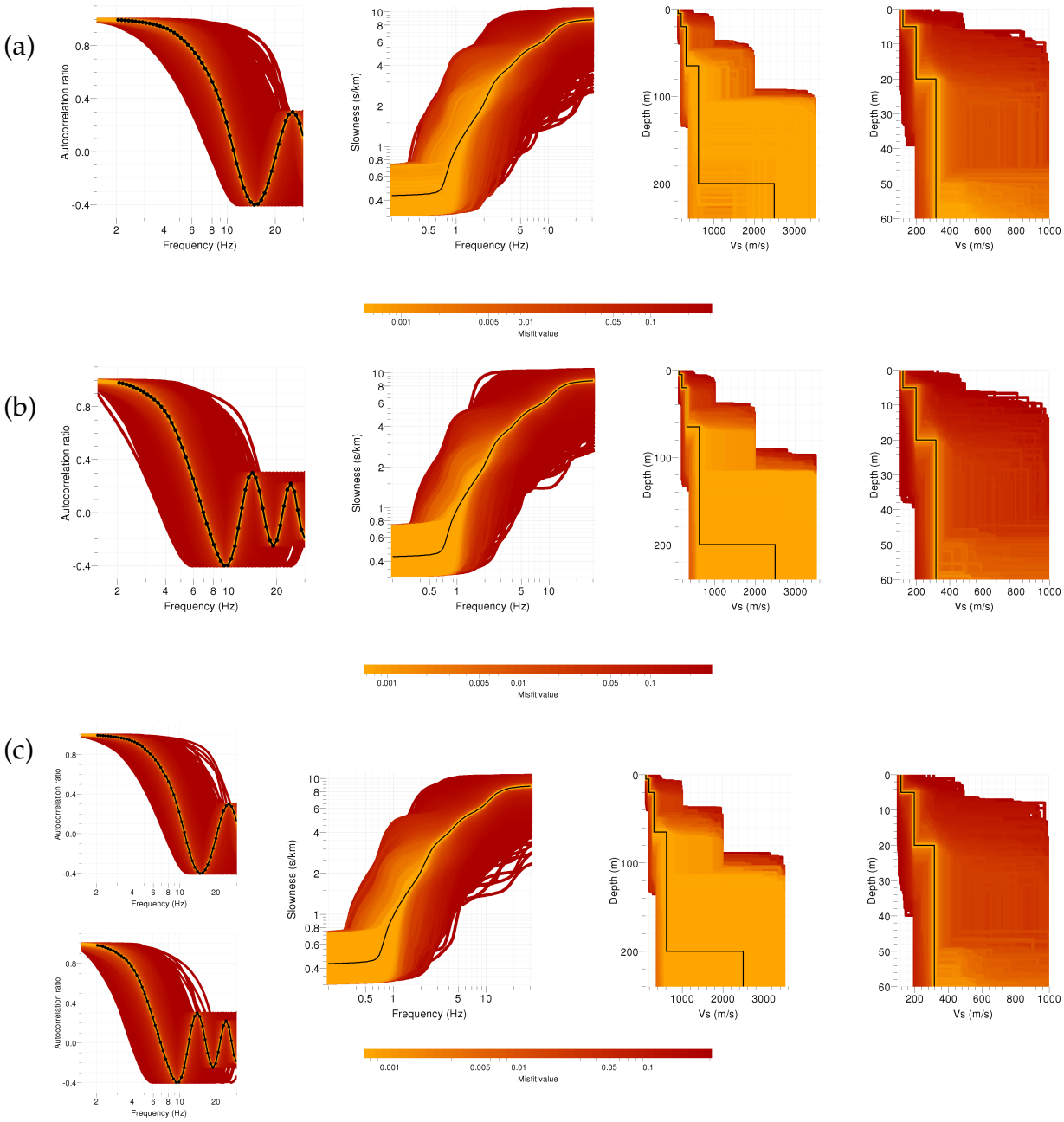


Figure 6.3: (a) Results for the inversion of the theoretical autocorrelation curve (left, with data points in black) for a ring with a 5 meter radius: Dispersion curves (center left), and shear wave (center right) velocity profiles. A zoom to the shallow structure of the shear wave velocity profiles is displayed at right. All models generated during the inversion process with misfit values lower than 1 are shown in the color of the respective misfit value. Darker curves are associated with higher misfit values, whereas lighter colors indicate the best-fitting models. The black line corresponds to the actual soil model. (b) The same as (a), but for a ring with a radius of 10 meters. (c) The same for a joint inversion of the 5-meter-ring (autocorrelation curve top left) and the 10-meter ring (autocorrelation curve bottom left).

6.4.4 Which parts of the ellipticity curve are necessary for the inversion process?

The deeper part of the structure can be fixed by using Rayleigh wave ellipticity measurements. In the following inversions, the theoretical autocorrelation curve for a radius of 5 m (between 2 and 30 Hz) and the theoretical ellipticity curves of profiles A and B will be inverted jointly. In order to determine which parts of the ellipticity curve carry the important information about the soil structure, different parts of the ellipticity curve will be used for the inversion process and the results will be compared for both a model with and without singularities (peak and trough) in the ellipticity curve.

6.4.4.1 Model with a singularity in the ellipticity curve

Table 6.3: Parts of the autocorrelation and ellipticity curves used for the inversions and the associated misfit and T values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for model A.

Reference figure	Autocorrelation curve Array radius	Autocorrelation curve Frequency range	Used frequency range of the ellipticity curve	Number of models	Minimum misfit	T value (0.7 - 30 Hz)
Fig. 6.4 (a)	5 m	2 - 30 Hz	0.2 - 0.65 Hz	100 100	$8.5 \cdot 10^{-3}$	0.241
Fig. 6.4 (b)	5 m	2 - 30 Hz	0.7 - 2.0 Hz	100 100	$15.4 \cdot 10^{-3}$	0.043
Fig. 6.4 (c)	5 m	2 - 30 Hz	0.2 - 2.0 Hz	150 200	$30.2 \cdot 10^{-3}$	0.140
Fig. 6.4 (d)	5 m	2 - 30 Hz	0.7 - 4.0 Hz	150 200	$21.2 \cdot 10^{-3}$	0.087
Fig. 6.4 (e)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	100 100	$4.1 \cdot 10^{-3}$	0.013
Fig. 6.4 (f)	5 m	2 - 30 Hz	0.9 - 1.7, 2.5 - 4.0 Hz	100 100	$3.9 \cdot 10^{-3}$	0.016
Fig. 6.4 (g)	5 m	2 - 30 Hz	0.9 - 1.7 Hz	100 100	$7.0 \cdot 10^{-3}$	0.154
Fig. 6.4 (h)	10 m	2 - 30 Hz	0.9 - 1.7 Hz	100 100	$1.1 \cdot 10^{-3}$	0.030
Fig. 6.4 (i)	5 m	2 - 30 Hz	0.7 - 4.0 Hz	150 200	$35.7 \cdot 10^{-3}$	0.093

Figure 6.4 (a) shows the results for an inversion of the left flank of the resonance peak only, i.e. low-frequency ellipticity values between 0.2 and 0.65 Hz. The generated ellipticity curves are badly constrained above the peak frequency and the dispersion curves exhibit a large variability over the whole frequency range. The true shear wave velocity profile is not matched at all and the T value is large. However, the low minimum misfit value indicates that the used part of the ellipticity curve and the autocorrelation curve are well fitted.
The inversion of the right flank of the ellipticity peak between 0.7 and 2 Hz (figure 6.4 (b)) yields rather good results. Although only the right flank of the ellipticity peak is inverted, the left flank is well retrieved as well. Beyond the trough frequency, the ellipticity curve is not well retrieved. The real dispersion curve is well retrieved and the shear wave velocity profiles are well fitted as well. Unsurprisingly, the bedrock velocity is not retrieved.
Combining both previous inversions by inverting the whole ellipticity peak (i.e. a frequency range from 0.2 to 2 Hz) gives an astonishing result (see figure 6.4 (c)). Although more models have been generated in this case (150 200), the results are worse than for the inversion of the right flank only. The dispersion curves show a significant deviation from

the true curve around 4.0 Hz. The shear wave velocity profile is not well constrained and the bedrock depth underestimated. The comparatively large misfit value indicates that the inversion process is slowing down in the presence of very high ellipticity values and the best model has not yet been found.
An inversion of the part of the ellipticity curve from 0.7 to 4.0 Hz, including the right flank of the peak and the trough (figure 6.4 (d)), matches the ellipticity curve very well, also at higher frequencies. Here, 150 200 models have been generated as well, but the minimum misfit value is still relatively high, indicating that a prolongated search for better models might have been successful. The T value of the best model proves the acceptable fit of the real dispersion curve, but the shear wave velocity profile's discontinuity depths are underestimated.
In figure 6.4 (e), the same frequency range as in (d) was inverted, omitting ellipticities around the trough frequency. The inverted frequency range lies between 0.7 and 1.7 and between 2.5 and 4.0 Hz. The minimum misfit and the associated T value are smaller than in (d). The dispersion curve and the shear wave velocity profile are very well retrieved. The inversion is less time-consuming than in (d) and yields better results.
In figure 6.4 (f), a smaller part of (e) was inverted, omitting ellipticity values exceeding 3. The frequency range of the inverted ellipticity data lies between 0.9 and 1.7 and between 2.5 and 4.0 Hz. The inversion has misfit and T values comparable to (e).
The same part of the right flank (0.9 - 1.7 Hz) without the data points at higher frequencies was inverted in figure 6.4 (g). Despite the small misfit value, the T value is not acceptable. The ellipticity curve at frequencies above the trough is badly fitted. The dispersion curve and the velocity profile are badly retrieved as well. The deviation in the dispersion curve around 4.0 Hz suggests that the overlap of the autocorrelation and ellipticity information are not sufficient in this case.
Therefore, in figure 6.4 (h), the autocorrelation for the 5-m ring has been replaced by the 10-m ring autocorrelation curve, but the ellipticity data is the same as in (g). Here, the misfit and T values are very small, as the dispersion curve and the shear wave velocity profile are very well retrieved.
In real data measurements, the ellipticity trough is sometimes not retrieved due to contributions of Love waves or higher Rayleigh wave modes. Such effects are investigated in figure 6.4 (i), where an ellipticity curve with wrong data is inverted. The ellipticity parts used for inversion (e), i.e. the true ellipticity curve between 0.7 and 1.7 and between 2.5 and 4.0 Hz, have been used. Instead of excluding the trough area, the data points between 1.7 and 2.5 Hz have been interpolated. The inversion succeeds in finding a soil model which exhibits a clear ellipticity peak, but no trough. The inversion result is still acceptable and the T value comparable with the one of inversion (d), where the correct ellipticity curve in the same frequency range was inverted. Nevertheless, inversion (e), where the data points between 2.5 and 4.0 Hz have been omitted, gives the best result.
The theoretical ellipticity curve of model A exhibits a peak at 0.67 Hz and a trough at 2.05 Hz. The parts of the ellipticity curve for the following inversions will be chosen in relation with these two frequencies. Unless otherwise stated, the following results were obtained after generating 100 100 soil models. The Rayleigh wave dispersion curve for frequencies below the resonance frequency is dominated by the shear wave velocity in the bedrock and the ellipticity curve is not sensitive to this velocity.

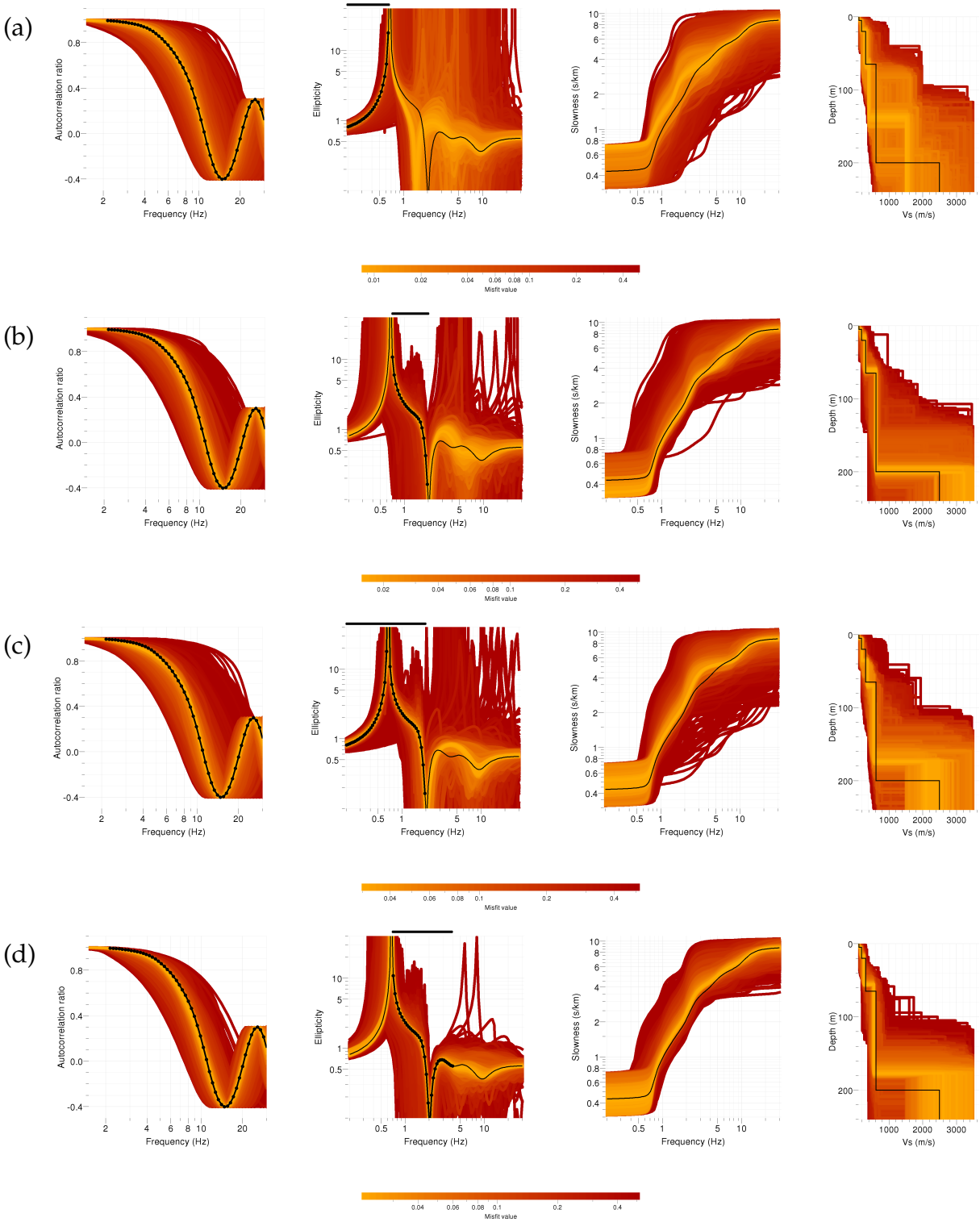


Figure 6.4: (a) Inversion results for joint inversions of the autocorrelation curve (left) and the ellipticity curve (center left): dispersion curves (center right) and shear wave velocity profiles (right). For the inversion, the used data points are shown as black dots, the curves corresponding to the real model are shown as black lines: Inversion of the ellipticity curve (a) between 0.2 and 0.65 Hz, (b) between 0.7 and 2 Hz, (c) between 0.2 and 2 Hz. (d) between 0.7 and 4 Hz.

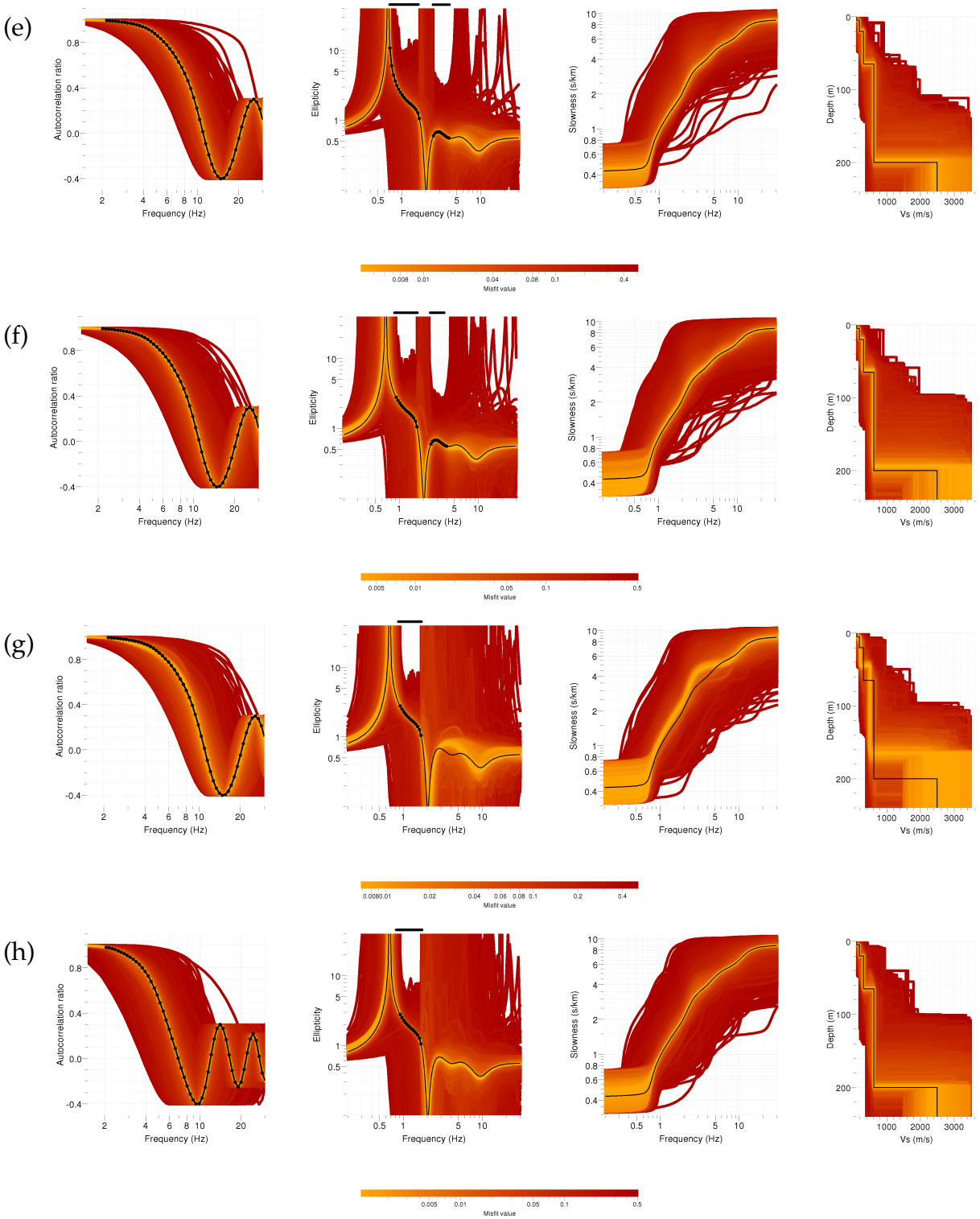


Figure 6.4: – continued. Inversion of the ellipticity curve (e) between 0.7 and 1.7 and 2.5 and 4.0 Hz, (f) between 0.9 and 1.7 and 2.5 and 4.0 Hz, (g) between 0.9 and 1.7 Hz, (h) between 0.9 and 1.7 Hz using the autocorrelation curve for the 10-m ring.

(i)

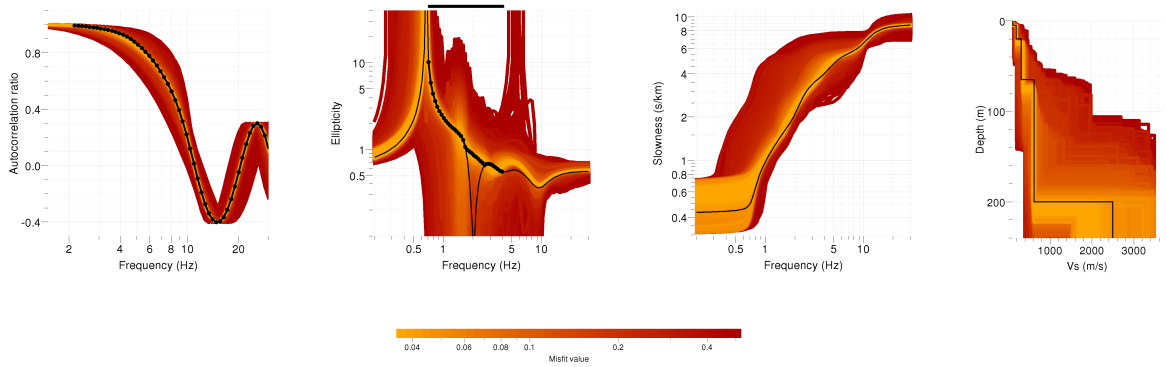


Figure 6.4: – continued. Inversion of an ellipticity curve (i) between 0.7 and 4.0 Hz, which corresponds to the real curve between 0.7 and 1.7 and between 2.5 and 4.0 Hz and is interpolated in the range between 1.7 and 2.5 Hz.

That is why the following T values are calculated for the frequency range between 0.7 and 30 Hz although the curves are shown for lower frequencies as well. Figure 6.4 shows the results obtained by inverting different parts of the ellipticity curve. The misfit and T values of the best-fitting models of these inversions are shown in table 6.3.
It should be noted that the shear wave velocity of the bedrock half-space is quite badly constrained for all inversions, even if the upper structure is very well retrieved. This is related to the penetration depth of the fundamental mode of Rayleigh waves. As the shear wave velocity in the bedrock is much higher than in the overlying sedimentary layers, the Rayleigh waves sample the bedrock only marginally and the velocity contrast between the bedrock and the overlying layer fixes the ellipticity curve at the peak frequency only. However, the effect of a changing bedrock velocity on the ellipticity curve is quite small and therefore not easy to measure.
Resuming the results of the inversions above, the lowest T values, indicating the best inversion results, were obtained for the inversions (e) and (f) with T values of 0.013 and 0.016, respectively. These inversions include a part of the right flank of the ellipticity peak and a part beyond the trough. The third best result was obtained for inversion (h) with a T value of 0.030, where only a part of the right flank was inverted with the autocorrelation curve for the larger ring. A still very good result (T = 0.043) was obtained for inversion (b), where the complete right flank was inverted. The other inversions yield worse results, even if they include more data points like inversion (d) or (c). The left flank only (a) or a small part of the right flank (g) without a sufficient frequency overlap with the autocorrelation data are not sufficient to retrieve the soil structure.
As a conclusion it can be said that the most important part of the ellipticity curve to measure is the right flank of the peak. The left flank of the peak is not important for the inversion and can be omitted. Including the ellipticity curve at the peak or trough itself, however, does not improve the results as it slows down the convergence of the inversion process and inhibits the retrieval of the best model. Adding information above the trough frequency increases the overlap between the frequency ranges sampled by the ellipticity and the autocorrelation curves. Consequently, the inversion results in this case are better.

The best results for real data inversions can be expected if the whole right flank is measured accurately. If the complete flank cannot be measured, a part beyond the trough frequency has to be included into the inversion process. However, if the trough frequency cannot be clearly measured, the ellipticity values in the questionable frequency range of the trough should be omitted.

6.4.4.2 Model without a singularity in the ellipticity curve

Table 6.4: Parts of the autocorrelation and ellipticity curves used for the inversions and the associated misfit and T values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for model B.

Figure reference	Autocorrelation curve Array radius	Autocorrelation curve Frequency range	Ellipticity curve	Number of models	Minimum misfit	T value (0.7 - 30 Hz)
Fig. 6.5 (a)	5 m	2 - 30 Hz	0.75 - 9.5 Hz	100 100	$7.0 \cdot 10^{-3}$	0.017
Fig. 6.5 (b)	5 m	2 - 30 Hz	0.75 - 5.2 Hz	100 100	$13.1 \cdot 10^{-3}$	0.047
Fig. 6.5 (c)	5 m	2 - 30 Hz	0.60 - 2.7 Hz	100 100	$17.8 \cdot 10^{-3}$	0.233

Model B does not exhibit an ellipticity singularity. The ellipticity curve shows a maximum of 1.71 at 0.73 Hz and a minimum of 0.36 at 9.44 Hz. Its ellipticity peak is broad and exhibits a sort of fine structure which is probably related to the different layer boundaries, but certainly not in a simple way. In order to determine which parts of the ellipticity curve have to be included in the inversion process in this case, we performed joint inversions of autocorrelation and ellipticity measurements. The theoretical autocorrelation curve for a distance of 5 m was used jointly with different parts of the ellipticity curve of model B. 100 100 models have been computed per inversion. The misfit and T values for the best-fitting models are shown in table 6.4.
Figure 6.5 (a) shows the results for an inversion of the right side of the broad ellipticity peak between 0.75 and 9.5 Hz. The ellipticity curve is well fitted, the dispersion curve and the shear wave velocity profile are retrieved. This is also indicated by the low T value.
Inverting the right side of the ellipticity peak between 0.75 and 5.2 Hz yields good results (see figure 6.5 (b)). The dispersion curve is well retrieved, but slight uncertainties in the estimation of the shear wave velocity profile can be seen.
Using the ellipticity values between 0.6 and 2.7 Hz only is not sufficient to find the correct soil structure (figure 6.5 (c)). The ellipticity at higher frequencies is badly estimated and the dispersion curve and velocity profile are badly fitted as well, although the minimum misfit value is in the same order as for the other inversions.
These inversions prove that even in the case of an ellipticity which does not show a clear peak or singularities, the ellipticity curve carries important information on the ground structure and can be used to find the soil structure. The best results are obtained by inverting the complete right side of the ellipticity peak.

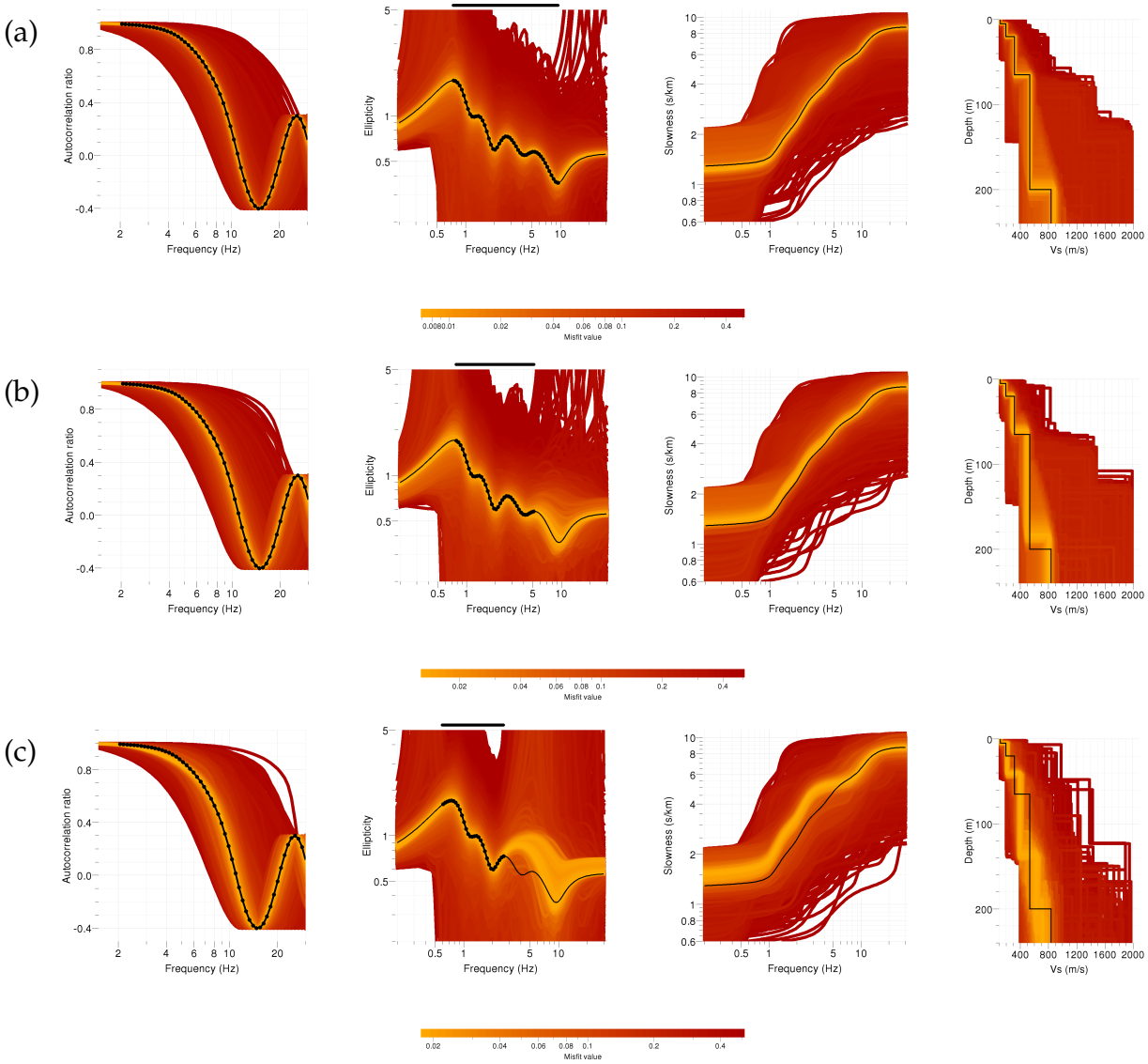


Figure 6.5: (a) Inversion results for joint inversions of the autocorrelation curve (left) and the ellipticity curve (center left): dispersion curve (center right) and shear wave velocity profile (right). For the inversion used data points as black dots, the curves corresponding to the real model are shown as black lines: Inversion of the ellipticity curve (a) between 0.75 and 9.5 Hz, (b) between 0.75 and 5.2 Hz, (c) between 0.6 and 2.7 Hz.

6.4.4.3 Joint inversion of dispersion and ellipticity curves

Table 6.5: Table of the misfit and T values for the best-fitting models of joint inversions of a part of the dispersion curve and a part of the ellipticity curve of model A.

Figure reference	Dispersion curve	Ellipticity curve	Number of models	Minimum misfit	T value (0.7 - 30 Hz)
Fig. 6.6 (a)	2 - 30 Hz	0.7 - 2.0 Hz	100 100	$20.4 \cdot 10^{-3}$	0.015
Fig. 6.6 (b)	4 - 30 Hz	0.7 - 2.0 Hz	100 100	$18.4 \cdot 10^{-3}$	0.079
Fig. 6.6 (c)	8 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	100 100	$4.1 \cdot 10^{-3}$	0.019
Fig. 6.6 (d)	10 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	100 100	$8.2 \cdot 10^{-3}$	0.092

The previous subsections dealt with joint inversions of autocorrelation and ellipticity curves. However, for an autocorrelation curve the frequency range where it really carries information is connected to the descending flank of the plateau of the Bessel function and therefore linked to its first root. Though, the exact frequency range where the autocorrelation curve carries information is not easy to determine. In the previous examples, the autocorrelation curve was used between 2 and 30 Hz. For a radius of 5 meters, the first zero-crossing of the autocorrelation occurs at 10 Hz, for the 10-meter ring it occurs around 7 Hz. As can be seen in figures 6.4 (g) and (h), it is evident that the 10-meter autocorrelation curve carries more information at low frequencies and constrains the velocity profile to greater depths.
SPAC is only one of the possible methods to measure Rayleigh wave dispersion curves. Another widespread method of measuring is MASW [Park et al., 1999, Socco and Strobbia, 2004]. These active measurements yield high-frequency dispersion curves. The combination of existing MASW mesurements with ellipticity measurements obtained using a single station would be a simple way of extending the knowledge of the soil structure to larger depths. It is clear that the joint inversion cannot succeed if the frequency ranges of the ellipticity and dispersion curves are too separated. Therefore, the objective of this subsection is to determine how close they really have to be.
In figure 6.6, the results of joint inversions of ellipticity and dispersion curves are shown. Each figure was obtained by generating 100 100 models. The minimum misfit and T values of the inversions are shown in table 6.5. In figure 6.6 (a) and (b), the ellipticity data correspond to those used in figure 6.4 (b), in 6.6 (c) and (d), the ellipticity data are those of figure 6.4 (e). One could argue that the calculation of the T value for an inversion of dispersion curves is the same as the calculation of the misfit. If the ellipticity data was not included in the inversion, this would be true. However, only a small part of the dispersion curve is used here and the remaining parts of the dispersion curve are constrained by the ellipticity curve. Therefore, the T value is still appropriate to assess the goodness of the inversion result.

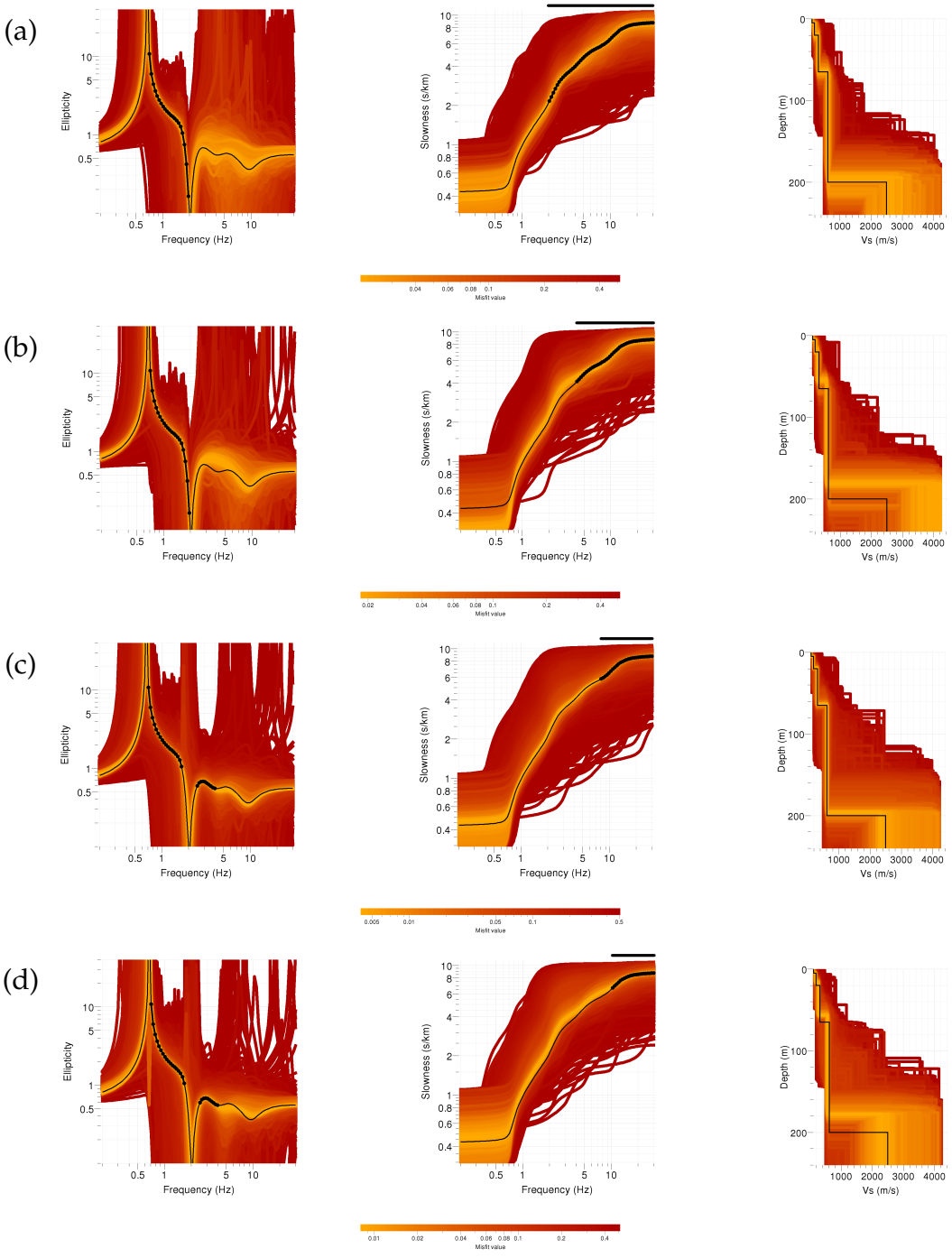


Figure 6.6: (a) Inversion results for joint inversions of the ellipticity curve (left) and the dispersion curve (center): resulting shear wave velocity profile (right). For the inversion used data points as black dots, the curves corresponding to the real model are shown as black lines: Inversion of the ellipticity curve between 0.7 and 2 Hz and the dispersion curve (a) between 2 and 30 Hz, (b) between 4 and 30 Hz, (c) between 6 and 30 Hz. Inversion of the ellipticity curve between 0.7 and 1.7 Hz and between 2.5 and 4 Hz and the dispersion curve (d) between 8 and 30 Hz, (e) between 10 and 30 Hz.

The first two inversions (figures 6.6 (a) and (b)) show the results for inverting the ellipticity curve between 0.7 and 2 Hz with different parts of the dispersion curve. Combining the ellipticity information with the dispersion curve between 2 and 30 Hz yields the correct soil structure. The T value of this inversion is better than the one for the inversion of figure 6.4 (b), where the autocorrelation curve was inverted instead of the dispersion curve. The correct dispersion curve is matched.
By restricting the dispersion curve to the range between 6 and 30 Hz (figure 6.6 (b)) the result degrades. The velocity profile is not well retrieved any more. The dispersion curve is still acceptably fitted, but the T value is five times the value of (a), although the minimum misfit is comparable. Omitting more dispersion curve data subsequently worsens the results.
Using the same ellipticity data as in figure 6.4 (e), i.e. between 0.7 and 1.7 and between 2.5 and 4 Hz, covers a wider frequency range. Consequently, the dispersion curve can be limited to higher frequencies. In this case, inverting the ellipticity curve jointly with the dispersion curve between 8 and 30 Hz is sufficient to retrieve the soil structure (figure 6.6 (a)). The fit of the dispersion curve is very good with a low T value.
However, using the dispersion curve between 10 and 30 Hz only (figure 6.6 (d)) gives a worse result. Although the inverted data are well fitted as well, the soil structure matches badly. The T value of the dispersion curve fit is still acceptable, but five times worse than in the previous case.
The examples above show that, in order to perform a successful joint inversion, the distinct datasets do not necessarily need to have an overlap in the frequency domain. Combining ellipticity data below 4 Hz with a dispersion curve above 8 Hz retrieves the correct soil structure. However, a further increase of this gap deteriorates the results until the two distinctive data sets are not linked any more. Beyond that point, a successful inversion becomes impossible. Furthermore, this result shows that the ellipticity and dispersion curve carry information about the same superficial layers at different frequencies. Otherwise, it would be impossible to invert a high-frequency dispersion curve measurement and a low-frequency ellipticity curve jointly, with both curves being defined in distinct, non-overlapping frequency ranges.

6.4.5 Inversion of erroneous ellipticity curves

The measurement of ellipticity curves can always be biased by systematic errors, introduced for example by a wrong estimation of the amount of Love waves or incomplete elimination of waves other than Rayleigh waves. In this subsection, the same autocorrelation curve as in the previous subsections will be inverted jointly with a wrongly estimated ellipticity curve of model A. As in subsection 6.4.4.1 the inversion of the ellipticity curve parts between 0.7 and 1.7 and 2.5 and 4 Hz gave the best results, only these parts will be used for the inversion here. The inversion shown in figure 6.4 (e) will therefore be taken as reference inversion.

6.4.5.1 Ellipticity misestimation by a constant factor

Table 6.6: Table of the misfit and T values for the best-fitting models of joint inversions of the autocorrelation curve and misestimated parts of the ellipticity curve of model A.

Figure reference	Autocorrelation curve Array radius	Autocorrelation curve Frequency range	Ellipticity curve	Ellipticity misestimation	Number of models	Minimum misfit	T value (0.7 - 30 Hz)
Fig. 6.7 (a)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	+ 10 % constant	100 100	$6.2 \cdot 10^{-3}$	0.028
Fig. 6.7 (b)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	+ 25 % constant	100 100	$12.5 \cdot 10^{-3}$	0.055
Fig. 6.7 (c)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	+ 40 % constant	100 100	$15.9 \cdot 10^{-3}$	0.051
Fig. 6.7 (d)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	- 25 % constant	100 100	$60.9 \cdot 10^{-3}$	0.150
Fig. 6.9 (a)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	+ 25 % maximum	100 100	$7.4 \cdot 10^{-3}$	0.026
Fig. 6.9 (b)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	+ 50 % maximum	100 100	$8.0 \cdot 10^{-3}$	0.038
Fig. 6.9 (c)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	- 25 % maximum	100 100	$26.4 \cdot 10^{-3}$	0.097
Fig. 6.9 (d)	5 m	2 - 30 Hz	0.7 - 1.7, 2.5 - 4.0 Hz	- 50 % maximum	100 100	$85.5 \cdot 10^{-3}$	0.194

For the following inversions, we overestimated the ellipticity curve by a constant factor over the whole frequency range. Then, we performed a joint inversion of the autocorrelation curve for a radius of 5 m between 2 and 30 Hz and the biased ellipticity curves in the ranges 0.7 - 1.7 and 2.5 - 4 Hz. The resulting curves are shown in figure 6.7 and the depth and shear wave velocity values for the soil structures of the best-fitting models of each inversion are given in table 6.7. The misfit and T values of these inversions are given in table 6.6.
The resonance frequency could be supposed to change significantly when biasing the ellipticity curve, but the maximum frequency shift is - 1.8 % for the - 25 % bias. For positive bias values, the resonance frequency shifts very slightly to higher frequencies. For the reference inversion, a minimum misfit of 0 is theoretically possible, as the true model exists and can be found by the inversion. For biased ellipticity curves, however, it is rather improbable that any model with exactly such an ellipticity curve exists. Therefore, the minimum misfit values are supposed to increase with an increasing introduced bias. This can be observed in table 6.6, but it should also be noted that for underestimated ellipticity curves the misfit values are higher than for overestimated ones. Consequently, if the ellipticity curve is overestimated, it is easier to retrieve the correct model than for underestimated ellipticity curves.
For the overestimated ellipticity curves (Fig. 6.7 (a) - (c)), the real dispersion curve is well fitted. For the underestimated ellipticity curve (Fig. 6.7 (d)), the T value is larger than for any of the overestimated inversions and the dispersion curve is not well fitted.
Introducing a positive bias on the ellipticity values increases the bedrock depths significantly (parameter D_4 in table 6.7). However, these values are inferior to the introduced bias. The first two layers are mainly controlled by the autocorrelation curve. Therefore, their properties do not significantly change, as well as the shear wave velocities of the first three layers, except for the significant underestimation of the interface between the second and the third layer. The depths of the third layer are not systematically shifted, the velocities of the fourth layer are overestimated in a way comparable to the overestimation of the bedrock depth. As the ellipticity inversions are not very sensitive to the bedrock velocity, the bad estimation of this parameter is not surprising.

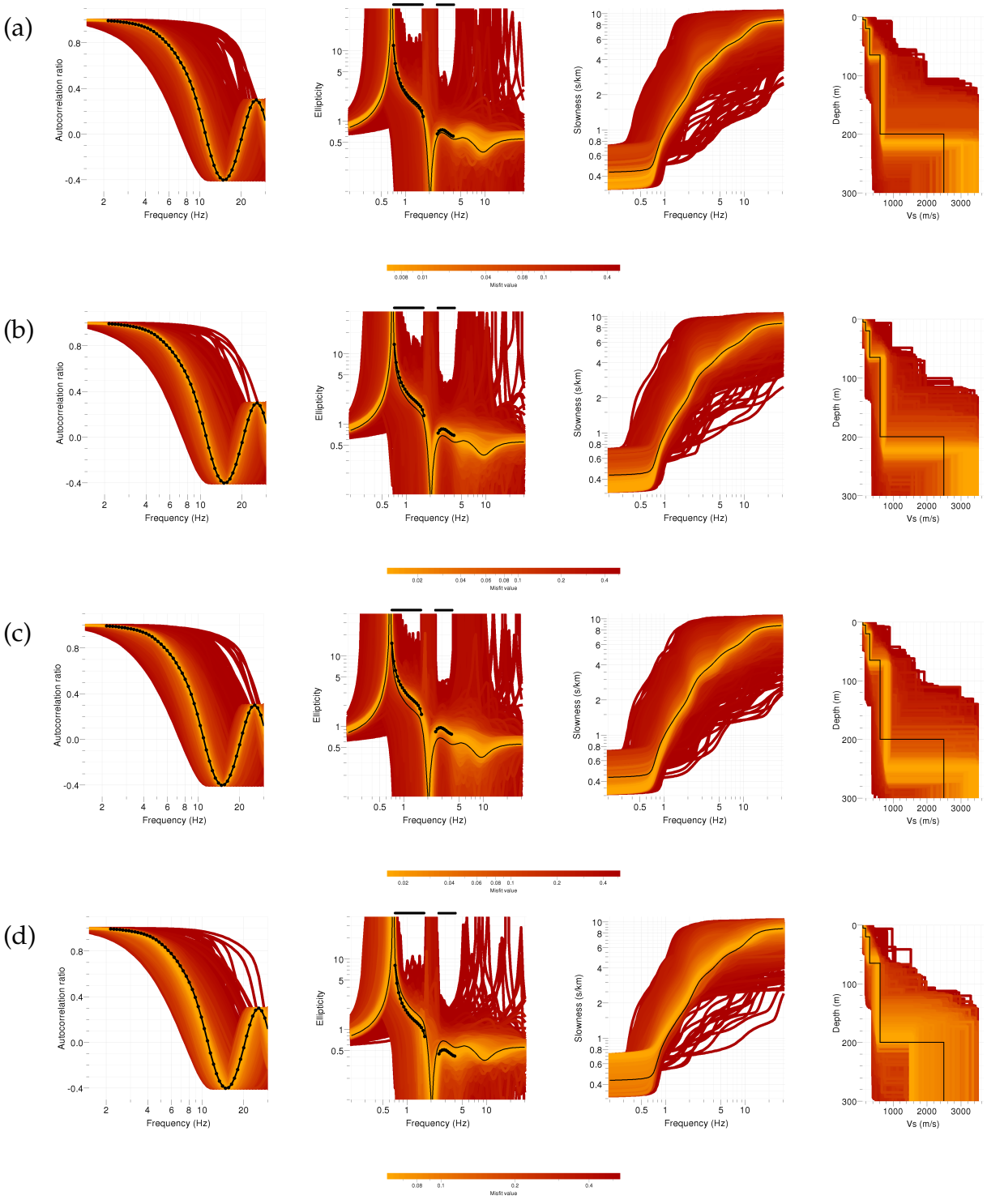


Figure 6.7: Inversion results for joint inversions of the autocorrelation curve (left) and biased ellipticity curves (center left): dispersion curves (center right) and shear wave velocity profiles (right). The black dots indicate the data points used for the inversion and the black lines the corresponding curves for the unbiased model. The complete ellipticity curves are biased by (a) + 10 %, (b) + 25 %, (c) + 40 %, (d) - 25 %.

Overall, it can be stated that the superficial structure, mainly controlled by the autocorrelation curve, is quite robust against overestimation of the ellipticity. The overestimation of the deeper structure, particularly the parameters of the layer directly above the bedrock, is still smaller than the bias on the real ellipticity curve.
Introducing a negative bias of - 25 % on the ellipticity values, however, gives a different result. The velocity of the superficial layer stays constant, but every parameter from the first layer's depth to the shear wave velocity of the fourth layer is shifted by approximately - 10 % (ranging from - 7.0 to - 12.0 %). The bedrock depth is underestimated by only - 3.9 %. The ellipticity underestimation also influences the superficial structure, but the bias induced in the whole structure is far below the introduced ellipticity bias.
The percental deviation of the bedrock depth D_4, the mean velocity v_{SB} and the shear wave travel time τ_{SB} between the surface and the bedrock depth are plotted as a function of the percental ellipticity misestimation in figure 6.8. The travel time calculation is based on the shear wave velocities and layer depths. The mean velocity is calculated as the ratio between the bedrock depth and the travel time. The travel time values vary only slightly with varying ellipticity bias. The increasing travel-time for the underestimated ellipticity curve is against the trend of the overestimated ellipticity values. In any case, the variations of the travel time (maximum 5.7 %) are much lower than the introduced ellipticity bias. The bedrock depth and mean velocity are more sensitive to the ellipticity misestimation. The scaling of soil profiles itself does not change travel times, but velocity and depth parameters of the single layers, and do provide very close site reponses.

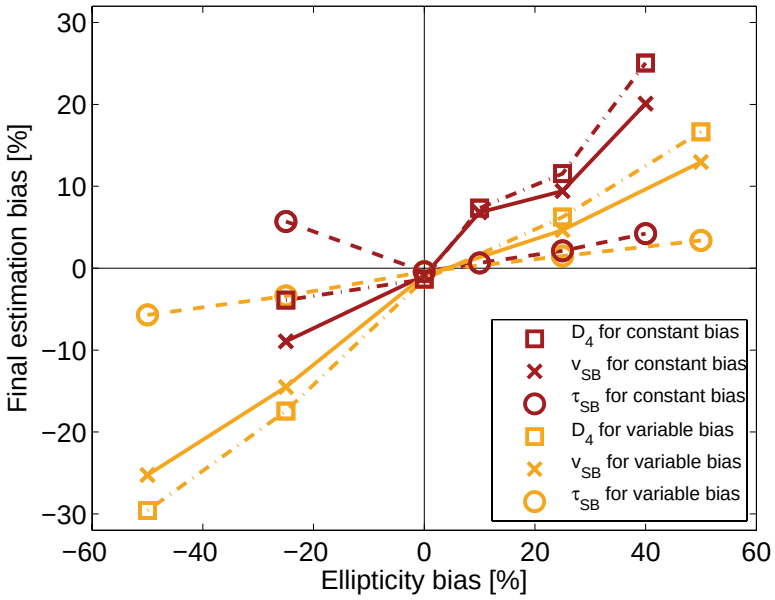


Figure 6.8: Percental deviation of bedrock depth D_4, mean velocity v_{SB} and shear wave travel time τ_{SB} from the surface to the bedrock in function of the introduced bias on the ellipticity curve. The results for a constant ellipticity misestimation are shown in dark color, for a variable misestimation in light color. For the variable misestimation, the x-axis indicates the maximum ellipticity deviation.

6.4.5.2 Frequency-dependent ellipticity misestimation

For the following inversions, the autocorrelation and ellipticity curves are inverted over the same frequency ranges as in the previous case, but the systematic error imposed on the ellipticity curve is frequency-dependent in the following way: At 0.7 Hz, the ellipticity value is not biased. At 1.7 Hz and between 2.5 and 4 Hz, the bias equals + 25 %, + 50 %, - 25 % and - 50 %. Between 0.7 and 1.7 Hz, the bias changes continuously. The overestimated cases are more representative for real ellipticity measurements than in the previous inversions as low ellipticities are more susceptible to overestimations than higher ellipticity values because an overestimation of the horizontal displacement due to Love waves has less impact if the horizontal Rayleigh wave displacement, and thus the ellipticity, is high. The results are shown in figure 6.9 and the depth and shear wave velocity values for the soil structures of the best-fitting models of each inversion are given in table 6.7. The minimum misfit and T values are given in table 6.6.
Ellipticity over- and underestimations both yield positive resonance frequency shifts, although the relative error is very small for overestimated ellipticities and acceptable for underestimated ellipticities.
Adding a positive bias of maximum + 25 % leads to misestimations of single properties lower than 10 %. In the case where the ellipticity curve is biased by at most + 50 %, the misestimation of the layer properties is lower than 17 %. In both cases, the dispersion curve is still well fitted.
Underestimating the ellipticity curves reveals a different picture. In both cases, the dispersion curve is not well retrieved. Except for the surface velocity, even the properties of the shallow layers are misestimated by over 10 %. The deeper layers are misestimated by up to 17.5 % (-25 % bias) and 30.0 % (-50 % bias). In these two cases, yet the autocorrelation curves do not seem perfectly fitted.
The percental deviations of the bedrock depth, the mean velocity and the shear wave travel time for frequency-dependent ellipticity misestimations are plotted in figure 6.8. Here, the travel times are even less biased than in the case of a constant ellipticity bias. The bedrock depths and mean velocities vary in a systematic way, but the slope of the corresponding curves is larger for negative ellipticity misestimations. This is further evidence indicating that the ellipticity inversion is more robust against overestimated ellipticity values than against underestimated ones.

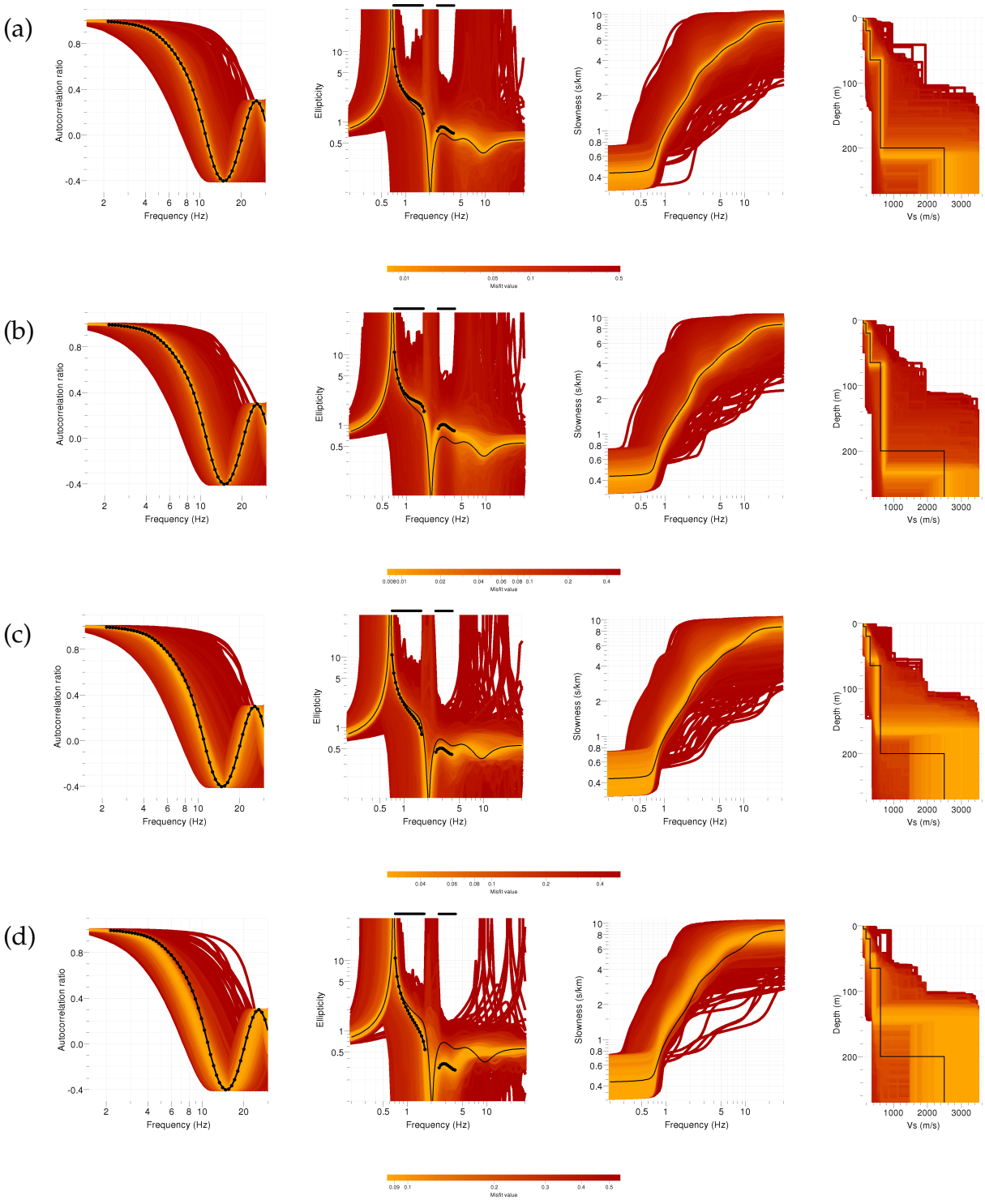


Figure 6.9: Inversion results for joint inversions of the autocorrelation curve (left) and biased ellipticity curves (center left): dispersion curves (center right) and shear wave velocity profiles (right). The black dots indicate the data points used for the inversion and the black lines the corresponding curves for the unbiased model. The ellipticity curves are biased by maximum (a) + 25 %, (b) + 50 %, (c) - 25 %, (d) - 50 %.

Table 6.7: Table of the layers' depths and shear wave velocities for the models with lowest misfit values resulting from biased inversions. Also shown are the travel-time from the surface to the bedrock τ_{SB} and the average velocity $v_{SB} = D_4/\tau_{SB}$ of the sedimentary layers. The reference model is the inversion shown in figure 6.4 (e). The biased models for the constant biases of +10 %, +25 %, +40 % and -25 % are shown in figures 6.7 (a), (b), (c) and (d), respectively. The biased models for the frequency-dependent biases of +25 %, +50 %, -25 % and -50 % maximum are shown in figures 6.9 (a), (b), (c) and (d), respectively. The given percentage values indicate the deviation of the best-fitting models' values from the real model values.

	Figure reference	V_{S1} $[m/s]$	D_1 $[m]$	V_{S2} $[m/s]$	D_2 $[m]$	V_{S3} $[m/s]$	D_3 $[m]$	V_{S4} $[m/s]$	D_4 $[m]$	V_{S5} $[m/s]$	f_0 $[Hz]$	τ_{SB} $[s]$	v_{SB} $[m/s]$
model		120.0	5.0	200.0	20.0	320.0	65.0	625.0	200.0	2500.0	0.668	0.473	422.6
reference	Fig. 6.4 (e)	119.6	4.96	198.7	19.6	313.0	64.7	625.9	197.3	2442.5	0.671	0.471	418.8
model		-0.3 %	-0.8 %	-0.7 %	-2.0 %	-2.2 %	-0.5 %	+0.1 %	-1.4 %	-2.3 %	+0.4 %	-0.4 %	-0.9 %
+10 %	Fig. 6.7 (a)	120.8	4.91	200.7	19.4	319.3	68.3	698.3	214.7	3040.2	0.667	0.476	451.4
constant		+0.7 %	-1.8 %	+0.4 %	-3.0 %	-0.2 %	+5.1 %	+11.7 %	+7.4 %	+21.6 %	-0.1 %	+0.6 %	+6.8 %
+25 %	Fig. 6.7 (b)	119.6	4.87	194.8	17.1	294.8	64.5	726.7	223.1	3494.7	0.672	0.483	462.4
constant		-0.3 %	-2.6 %	-2.6 %	-14.5 %	-7.9 %	-0.8 %	+16.3 %	+11.6 %	+39.8 %	+0.6 %	+2.1 %	+9.4 %
+40 %	Fig. 6.7 (c)	120.8	4.87	200.7	18.3	319.3	69.6	802.7	250.1	3391.9	0.675	0.493	507.5
constant		+0.7 %	-2.6 %	+0.4 %	-8.5 %	-0.2 %	+7.1 %	+28.4 %	+25.1 %	+35.7 %	+1.0 %	+4.2 %	+20.1 %
-25 %	Fig. 6.7 (d)	119.6	4.54	179.9	18.6	283.3	58.2	550.0	192.2	1515.0	0.656	0.500	384.8
constant		-0.3 %	-9.2 %	-10.1 %	-7.0 %	-11.5 %	-10.5 %	-12.0 %	-3.9 %	-39.4 %	-1.8 %	+5.7 %	-8.9 %
+25 %	Fig. 6.9 (a)	119.6	4.82	194.8	18.1	306.8	64.6	671.1	212.5	2567.1	0.670	0.480	442.3
maximum		-0.3 %	-3.6 %	-2.6 %	-9.5 %	-4.1 %	-0.6 %	+7.4 %	+6.3 %	+2.7 %	+0.3 %	+1.5 %	+4.7 %
+50 %	Fig. 6.9 (b)	120.8	4.96	204.7	18.5	322.4	68.3	726.7	233.3	2618.7	0.669	0.489	477.4
maximum		+0.7 %	-0.8 %	+2.4 %	-7.5 %	+0.8 %	-5.1 %	+16.3 %	+16.7 %	+4.7 %	+0.1 %	+3.4 %	+13.0 %
-25 %	Fig. 6.9 (c)	118.4	4.19	174.6	17.6	283.3	56.4	523.3	165.1	3010.1	0.670	0.457	361.3
maximum		-1.3 %	-16.2 %	-12.7 %	-12.0 %	-11.5 %	-13.2 %	-16.3 %	-17.5 %	+20.4 %	+0.3 %	-3.4 %	-14.5 %
-50 %	Fig. 6.9 (d)	124.5	4.45	153.4	16.0	261.6	48.1	437.5	140.8	2892.7	0.674	0.446	315.9
maximum		+3.8 %	-11.0 %	-23.3 %	-20.0 %	-18.3 %	-26.0 %	-30.0 %	-29.6 %	+15.7 %	+0.9 %	-5.7 %	-25.2 %

6.5 Conclusion

This paper showed the theoretical possibilities, constraints and limits of inversions including the Rayleigh wave ellipticity in order to retrieve the local soil structure. First of all, it was shown that an inversion of an ellipticity curve must be performed jointly, e.g. with an autocorrelation or dispersion curve. The frequency ranges of both measurements do not necessarily have to overlap, but should be as close as possible. In the case of an ellipticity curve with singularities, the ellipticity curve should be truncated in order to exclude too high or too low values, which might strongly reduce the exploration of the complete parameter space. As the left flank of the ellipticity peak does not carry important information, the inversion should comprise only the frequency range above the peak frequency. If the ellipticity curve does not exhibit singularities, the complete right side of the peak which is, in general, broader than in the previous case has to be included in the inversion. For misestimated ellipticity curves, the induced errors in the soil structure are, in general, smaller than the introduced bias (about half the original bias).

In the present paper, all ellipticity and autocorrelation curves were calculated theoretically. For ellipticity curves obtained from real data measurements, a perfect measurement of the real ellipticity curve cannot be expected. Even if all contributions of other wave types are perfectly excluded, higher modes of Rayleigh waves can still bias the ellipticity estimation, especially at high frequencies. In the present study, these effects were disregarded. The identification of such effects in the ellipticity curve, however, is impossible in most cases. Therefore, it is advisable to limit the use of the ellipticity curve to the right flank of the peak. Consequently, the station distance used for the autocorrelation measurements should be chosen as small as possible for a good resolution of the superficial layers and in the same time as large as necessary for a good link between the frequency ranges of the ellipticity and autocorrelation curves. To estimate the correct array size for the autocorrelation measurements, preliminary information on the underground structure have to be known or assumed (e.g. H/V resonance frequency, geology).

Another problem with real data ellipticity curves is that the estimation of the properties of Rayleigh waves usually exploits the correlation between the vertical and the horizontal movement. Yet, at the singularities of the ellipticity curve, one of both components vanishes. Therefore, at those points, both components are not correlated any more. Consequently, at the trough frequency, the movement of Rayleigh and Love waves cannot be distinguished any more (at least when using a single sensor) and the ellipticity will be overestimated here. At the peak frequency, the vertical signal will be overestimated in an analog way. Therefore, the ellipticity value will be underestimated here. Thus, it is advisable to exclude the ellipticity estimation around the singular values from the ellipticity inversion.

In the companion paper, the lessons learned in the present paper for the inversion of ellipticity curves will be used to invert real data measurements.

Chapter 7

Inversion of ellipticity curves - Application to real data measurements

In chapter 6, the theoretical aspects of the inversion of ellipticity curves have been presented. It has been investigated which parts of the ellipticity curve are actually carrying the information on the soil structure and which additional measurements are necessary to make the inversion unambiguous. Furthermore, how misestimated ellipticity data influence the inversion result was shown. In this chapter, the lessons learned in chapter 6 will be applied to real data measurements which have been performed during the European NERIES project (2006-2010). The following text represents the article "Inversion of Rayleigh wave ellipticity measurements, Part II: Application to real data measurements" by Hobiger, Cornou, Di Giulio, Endrun, Renalier, Savvaidis, Bard, Hailemikael, Le Bihan, Ohrnberger, Theodulidis, and Wathelet, 2010b, which was submitted to *Geophysical Journal International*.

Le chapitre 6 a présenté les aspects théoriques de l'inversion de courbes d'ellipticité. Il a été déterminé quelles parts de la courbe d'ellipticité comportent les informations sur la structure du sol et quelles mesures additionnelles sont nécessaires afin de rendre l'inversion unique. En outre, l'influence de misestimations d'ellipticité sur le résultat d'inversion a été démontré. Dans le chapitre suivant, les leçons du chapitre 6 seront appliquées à des données réelles mesurées dans le cadre du projet européen NERIES (2006-2010). Le texte suivant représente l'article "Inversion of Rayleigh wave ellipticity measurements, Part II: Application to real data measurements" par Hobiger, Cornou, Di Giulio, Endrun, Renalier, Savvaidis, Bard, Hailemikael, Le Bihan, Ohrnberger, Theodulidis, and Wathelet, 2010b, qui a été soumis à *Geophysical Journal International*.

7.1 Abstract

Site effects can have a large impact on the damages occurring during earthquakes. Therefore, it is important to know the local soil structure in order to assess the seismic hazard. Established methods include borehole drilling and active reflection or refraction measurements. These measurements are, however, either expensive or not very appropriate for the use in urban areas. Another way is to obtain the dispersion curves of surface waves by measuring ambient seismic vibrations using arrays of seismic sensors. However, to ensure a good depth coverage, the measurements have to be performed at several array sizes or using a single array composed of a large number of stations. Therefore, such measurements are time-consuming. A simple alternative is the use of Rayleigh wave ellipticity, which can be measured by a single seismic sensor. Ellipticity as a function of frequency is closely linked to the shear wave velocity profile. Though, the ellipticity cannot be inverted unambiguously because a scaling of the velocity profile does not change the ellipticity curve. Therefore, additional measurements fixing the absolute values of the velocities at certain depths are necessary. Small scale SPAC or MASW measurements can provide this data. In a companion article, the theoretical aspects of joint inversions of ellipticity and SPAC or MASW measurements have been shown. This article applies the method to real data measurements collected at 14 different sites during the European NERIES project. It is found that the ellipticity inversion results are in good agreement with dispersion curve measurements and that the determined Vs30 values agree with borehole measurements for most sites. Furthermore, the method can help in identifying the mode of Rayleigh waves in dispersion curve measurements.

7.2 Introduction

The knowledge of the local soil structure is an important component of seismic hazard assessment as site effects can amplify ground motion and increase damages. Different methods permit the retrieval of the local soil properties, among them the drilling of boreholes and seismic refraction or reflection measurements. These methods are quite expensive or hardly suitable for urban areas. Alternatively, the soil can be investigated by measurements which generate surface waves actively, e.g. by sledgehammering. As the hammering generates waves at high frequencies only, this method investigates the very superficial layers only [Socco and Strobbia, 2004]. Another way of measuring in urban areas is to use ambient seismic vibrations which are generated by natural sources like the ocean waves or by human activities like traffic or industry, depending on the frequency range. By deploying arrays of seismic sensors, the dispersion curves of Rayleigh and Love surface waves can be measured. In order to cover a large frequency band, however, arrays of different sizes or a single array including many sensors have to be used [Wathelet, 2008]. Consequently, such measurements are time-consuming. By exploiting the ellipticity of Rayleigh waves, the number of sensors needed for the measurements can be reduced. The ellipticity indicates the ratio between the horizontal and the vertical movement of a Rayleigh wave and is a function of frequency. The possibility to constrain the soil structure by ellipticity data was first shown by Boore and Toksöz [1969].

A widespread technique used to characterize site effects is the H/V method which was proposed by Nogoshi and Igarashi [1971] and promoted by Nakamura [1989]. This method indicates the spectral ratio between the horizontal and the vertical components. If other wave types than Rayleigh waves did not contribute to the ambient noise spectra, the H/V curve would match the ellipticity curve. The exact composition of the noise wave field depends on the source characteristics of seismic noise and on the site structure [Bonnefoy-Claudet et al., 2008]. The main wave types are Rayleigh and Love waves, but their ratio is frequency-dependent or even time-dependent [Köhler et al., 2006, Endrun, 2010]. As Love waves are only present on the horizontal components, the H/V ratio will, in general, misestimate ellipticity. Assuming an a priori proportion between Rayleigh and Love waves, one can remove the Love wave contributions on the H/V ratio and use it as an ellipticity curve [Fäh et al., 2001, 2003]. Although the interpretation of H/V is still controversial, previous studies have already shown the benefits of including H/V data to inversions of Rayleigh wave dispersion curves [for instance Scherbaum et al., 2003, Parolai et al., 2005, Picozzi et al., 2005, Arai and Tokimatsu, 2005]. Arai and Tokimatsu [2004] used H/V data to fix the exact depths of shallow layers where the shear wave velocities were known by borehole measurements.

Either the use of H/V curves is limited to the use of the resonance frequency only or additional assumptions have to be made to correct other wave types' contributions if the whole curve is to be used. Newly developed methods exist which allow the direct measurement of ellipticity curves. The RayDec method [Hobiger et al., 2009a] discriminates other waves than Rayleigh waves by statistical means for single seismic sensors. Poggi and Fäh [2010] propose a method using f-k analysis which can even identify higher modes using arrays of several seismic sensors.

As scaling the depth and velocity properties of the soil structure by the same factor does not change the ellipticity curve [Scherbaum et al., 2003, companion paper], an inversion of ellipticity data alone will never yield the correct velocity profile. In order to fix the absolute shear wave velocity values of the profile, additional measurements are needed. For this purpose, small scale spatial autocorrelation measurements using few seismic sensors [Aki, 1957] or MASW measurements [Park et al., 1999, Socco and Strobbia, 2004] can be used.

A companion paper investigated the theoretical aspects of joint inversions of ellipticity and autocorrelation or dispersion curves in detail. It shows that the part of the ellipticity curve which carries the important information is the right flank of the peak whether the ellipticity curve does exhibit a singularity or not. The exact ellipticity values at the peak and trough frequencies of the curve are difficult to measure, but too high or too low ellipticity values can be omitted for the inversion without corrupting the results. The frequency ranges of the inverted ellipticity and autocorrelation curves do not necessarily have to overlap as long as the gap between them is not too large (see companion paper). If the ellipticity measurement is biased, the resulting errors on the final soil structure inversion results are, in general, smaller than the introduced bias.

This article will focus on the application of the theory developed in the previous article. We invert the ellipticity curves for 14 sites investigated during the European NERIES project.

For a representative selection of five of them, the inversions will be shown in detail. SPAC measurements using all stations of the array and for a subset of stations only will be jointly inverted with ellipticity curves for these sites. For the remaining sites, only the main results will be shown.

7.3 Measurement methodology

7.3.1 Ellipticity measurements

The ellipticity measurements have been performed using the RayDec method [Hobiger et al., 2009a]. This method uses the data of a single 3-component seismic sensor to retrieve the ellipticity of Rayleigh waves. The algorithm is based on the random decrement technique which is commonly used to characterize the dynamic parameters of buildings or other oscillatory systems. The 3-component signal is cut in small time windows starting at the positive zero-crossings of the vertical component. For each time window, the signal of the horizontal components is projected into the direction which maximizes the correlation to the vertical component with a 90° phase shift which is typical of Rayleigh waves. Subsequently, the signals obtained in this way are stacked together and the ratio between the horizontal and vertical movements is estimated by analyzing the energy content of the vertical and horizontal stack. We will not describe the method in more detail here, but refer the interested reader to the publication of Hobiger et al. [2009a].

7.3.2 Spatial autocorrelation measurements

As the ellipticity inversion cannot be successful without additional information, we need to include further data in the inversion process. Aki [1957] introduced the spatial autocorrelation (SPAC) technique.

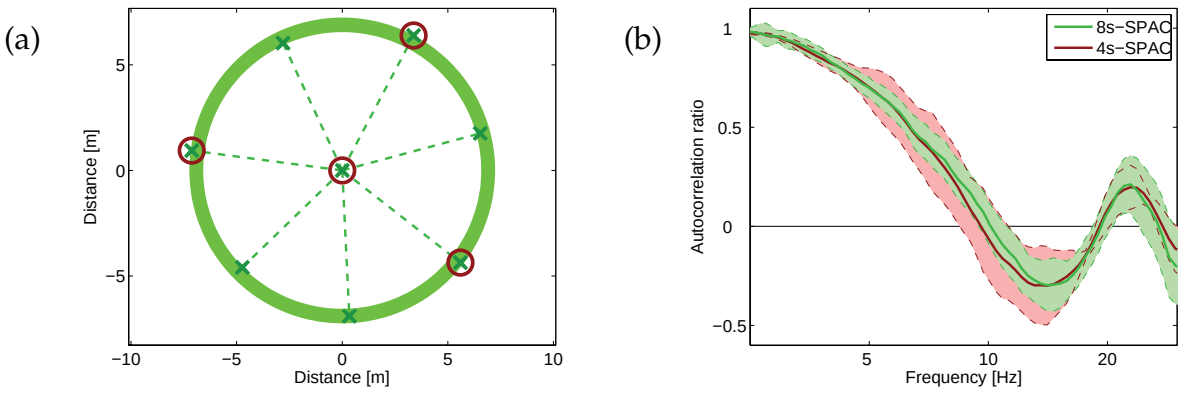


Figure 7.1: (a) Layout of the smallest seismic array deployed in Nestos. SPAC measurements were performed using the correlations between the central and: all ring stations (8s-SPAC), the three marked ring stations (4s-SPAC). (b) Results for 8s-SPAC and 4s-SPAC.

This technique uses the correlation between different seismic sensors to retrieve the dispersion curve. However, the original layout imposed that the distances between each station pair are equal. Therefore, the seismic stations were to be deployed on a ring around a central station. In addition, the number of seismic stations should be large to ensure a wide coverage of azimuths for incoming waves. As such a layout is, due to spatial limitations, hardly feasible in urban areas, Bettig et al. [2001] further developped the spatial autocorrelation method by suppressing the need of equal distances. The measurement is now possible for rings with a certain thickness. In this way, SPAC measurements are possible even for unsymmetric array layouts. The SPAC measurements presented in this paper are obtained using this method. For the measurements performed during the NERIES project, a total of 8 seismological stations [Earthdata acquisition systems equipped with 5s-Lennartz 3-component sensors, Endrun, 2010] were used for each array of sensors. One station served as central station whereas the other 7 stations were deployed on a ring around it, more or less symmetric depending on local constraints.
Okada [2006] showed that an array layout with three ring stations surrounding a central station is sufficient for good results, independent of the azimuth distribution of the seismic noise field. This is illustrated in figure 7.1. An example array layout is shown in figure 7.1 (a), corresponding to one of the investigated sites (Nestos). The 7 ring stations have distances to the central station between 6.59 and 7.24 m. We used one hour of measurements recorded on all 8 stations simultaneously. A 7-station pair SPAC measurement has been performed using only pairs consisting of the central and any ring station. The three marked stations on the ring build a good triangle around the central station and are between 7.09 and 7.24 m distant from it. The SPAC measurements resulting of the correlation of these three stations with the central station are compared to the 7-station SPAC in figure 7.1 (b). Both measurements yield qualitatively identical results, but the error bars are smaller when using more stations.
Morikawa et al. [2004] suggested 2s-SPAC, a method using only two seismic stations to perform SPAC measurements. In this way, the reference station is fixed whereas the second station changes its position during the measurement. This method presumes that the seismic noise wave field is stationary in time. Nevertheless, the focus of this paper is not to investigate the temporal stationarity of the wave field or the applicability of the 2s-SPAC method. However, the use of 2s-SPAC could further decrease the costs of seismic surveys. In the following, we will show inversions of ellipticity curves measured at the central station combined with SPAC measurements using all available stations as well as for three station pairs only (i.e. the extreme cases).

7.4 Inversion algorithm

We used the dinver code to perform the inversions [Wathelet et al., 2004, Wathelet, 2008]. This code generates, for a given parametrization, models for the shear wave velocity and assigns a misfit value to each of them. The misfit values are calculated in the following way:

$$misfit = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{D_i - M_i}{\sigma_i} \right)^2}. \tag{7.1}$$

Here, the inverted data set is called D_i and consists of N data points ($i \in \{1, 2, \ldots, N\}$) with associated measuring errors σ_i. M_i describes the same type of data calculated on the basis of the corresponding soil structure model. For joint inversions of two distinct data sets, a single misfit value is calculated for each of them. Their average value then gives the final misfit value.
The used algorithm is based on the neighborhood algorithm proposed by Sambridge [1999a,b]. This algorithm produces a random initial set of models and calculates the associated misfit values. Then, a new generation of models is searched for in the neighborhood of the models corresponding to the lowest misfit values. In this way, the best models yet found guide the inversion process, ensuring a good coverage of the whole parameter space as well.

7.4.1 Goodness of the inversion results

All sites for which soil structure inversions were performed in this paper had been previously studied by different methods [Picozzi et al., 2007]. During the European NERIES project, seismic noise has been recorded with arrays of different sizes for each site [Endrun and Renalier, 2008]. Additional MASW measurements have also been performed. Each method yields a dispersion curve in a certain frequency range. By combining all of them, a dispersion curve covering a wide frequency band is obtained. The goal of the present paper is to retrieve the soil structure using as little information and minimum field effort as possible. Therefore, the ellipticity curve which is measurable with a single sensor is inverted jointly with an autocorrelation curve which can be measured using only three (or even two) sensors. Alternatively, a high-frequency dispersion curve measured by the MASW method can replace the autocorrelation curve. In this way, the presented inversions are based on a small number of sensors. The inversion results can then be compared to the broad-band dispersion curve obtained during the NERIES project.
The minimum misfit value indicates if the data used for the inversion is fitted correctly. However, it does not indicate the fit between the inversion result and reality. The inversion data might be insufficient to constrain the soil structure in an adequate way, although the inversion yields a very low misfit value. Therefore, the misfit value can only tell us if the inversion data has been fitted in a proper way, but not if the inversion gives a satisfactory result for the considered site.

In order to compare the results of our inversions to reference data, we can compare them either to the borehole data or to the broad-band dispersion curves measured during the NERIES project. The borehole data gives the velocity profile of the site, but we have no information about the goodness of these measurements or the associated error bars. Furthermore, it is difficult to compare two velocity profiles as it is not clear up to which depth the comparison should reach or how many layers should be considered. Moreover, some borehole measurements did not even reach the bedrock. Therefore, we chose to compare our inversions to the direct dispersion curve measurements.
In the first article, we defined a proximity value comparing the inverted and the theoretical dispersion curves (which did not exhibit error bars) at N frequency samples in the following way:

$$T = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{s_{mod}(f_i) - s_{inv}(f_i)}{s_{mod}(f_i)}\right)^2}, \tag{7.2}$$

where s_{mod} indicates the dispersion curve of the theoretical model and s_{inv} the dispersion curve of any model generated during the inversion process. Similarly to this definition, we define a proximity value which takes into account the measurement errors of the direct dispersion curve measurements by:

$$P = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{s_{meas}(f_i) - s_{inv}(f_i)}{\sigma(f_i)}\right)^2}. \tag{7.3}$$

Here, $s_{inv}(f)$ indicates the dispersion curve for any inverted model and $s_{meas}(f)$ the dispersion curve of the direct measurements with error bars $\sigma(f)$, both expressed in terms of slowness. The comparison should range the same frequencies as the measured dispersion curve. In this way, P corresponds to the misfit value the profile would have if we had inverted the dispersion curve. For a model whose dispersion curve lies within the error range at all frequencies, the P value will be less than 1. Therefore, we will accept inversions with P values below 1. As different models which fit the ellipticity data in a comparable way can exhibit quite different dispersion curves and therefore different P values, it would be arbitrary to calculate the P value for the model with the smallest misfit value only. Therefore, we will calculate the P values for every model generated during the inversion process and deduce a mean P value for the models whose misfit values are not more than 10 percent larger than the minimum misfit value. Nevertheless, there is no general agreement on how to determine mean values for a model population. Therefore, we will consider the model with the minimum misfit value as being representative for the population.

7.4.2 Inversion strategy

The inversion strategy has to be adapted to the ellipticity curves. The spatial autocorrelation curves can be inverted without ellipticity values to get a first idea of the shallow structure, but this is not crucial for a successful inversion. As shown in the companion article, different parts of the ellipticity curve have to be inverted depending on whether clear peaks can be seen in the ellipticity curve or not. Therefore, the first step consists in looking at the ellipticity curve. If a clear peak can be seen, we will use its right flank for the inversion. As the measurements of ellipticity are problematic at the singularity location because either the horizontal or the vertical displacement of Rayleigh waves vanishes at these points, we will also omit very high or very low ellipticity values. Of course, the threshold has to be chosen with respect to the given ellipticity curve, but as a rough guideline we can say that ellipticity values above 3 should not be taken into account. Ellipticity values below 0.5 should be omitted as well. If the frequency ranges of the autocorrelation and ellipticity curves differ too much, additional information beyond the trough frequency should be included for the ellipticity inversion. However, the trough frequency is, in general, not well determined, and beyond this frequency, the measured ellipticity curve can be affected by higher modes. Therefore, it is preferable to limit the ellipticity inversion to the right flank of the peak and to use a larger layout for the spatial autocorrelation measurements. By inverting the right flank alone, the frequency of the peak is poorly constrained. Therefore, after each inversion, the inversion's peak frequency has to be controlled. If it does not coincide with the measured frequency, including the left flank of the peak can help. The part of the left flank should range the same ellipticity values as the right flank.
If no clear peak can be seen in the ellipticity curve, there is no reason to omit very high or very low ellipticity values. In that case, a broader but smaller ellipticity peak can be usually seen and inverted completely.

7.5 Application to real data measurements

The inversion strategy described above has been applied to measured data obtained during the NERIES European project. We selected 14 (6 in Italy, 5 in Greece, 3 in Turkey) of the 20 sites originally investigated during this project. The locations of these sites are shown in figure 7.2. The selection includes sites exhibiting clear ellipticity peaks and troughs as well as sites without singular ellipticities. For these sites, dispersion curves from array and MASW measurements are in agreement and could be combined easily. This yields directly measured dispersion curves with which the ellipticity curve inversions can be compared.
During the NERIES measurements, for each site ambient seismic vibrations have been recorded for three to four arrays of different sizes. Each array consists of a central station surrounded by a more or less perfect ring of seven seismometers. Additional MASW measurements have been performed as well. Using this data, it is possible to obtain a broad-band Rayleigh wave dispersion curve. In the following analysis, we will limit us to the use of the smallest array for each site. The ellipticity curve of the central station will be used for the inversions and the surrounding stations will be used for additional spatial autocorrelation measurements. The array radii at the different sites lie between 5 and 15 meters. The exact array layouts are shown in the separate analysis part for each site. Except for the three Turkish sites (Bolu, Düzce and Sakarya), for all sites independent borehole measurements are available which serve as a reference. For the Turkish sites, the reference measurements are former independent inversions of MASW measurements. Borehole measurements give the soil profile of a single point only, whereas broad-band dispersion curve measurements average over a larger part of the structure.
Table 7.1 lists the selected sites and indicates their respective soil classification according to the European EC8 code, the bedrock depth determined by borehole measurements and the size of the smallest seismic array used for the measurements.
Figure 7.3 shows the ellipticity curves which were obtained for each station of the smallest array for each analyzed site and compares them with the H/V curves for the same sites and the theoretical ellipticity curves derived from the reference measurements. Almost all sites show qualitatively identical ellipticity curves for each seismic sensor indicating that the soil structure does not change significantly over small distances. However, some sites show identical ellipticity curves for all sensors in a certain frequency range and different ellipticities at different frequencies (e.g. Norcia at low and high frequencies). The site of Knidi exhibits a completely atypical behavior with different ellipticity curves for each sensor. The peaks of Aigio and Buia (and Sturno) at 1.45 Hz can be identified as artificial peaks and are not related to the soil structure. Qualitatively, some sites exhibit very clear ellipticity peaks (Nestos and Volvi), or even two clearly distinguished peaks like Colfiorito, where the second peak could correspond to the fundamental mode or to a mixture of higher modes. For other sites, the ellipticity curves show a dominant peak which is however not singular (Aigio, Benevento, Bolu, Düzce, Korinthos, Norcia and Sakarya). Another set of sites exhibits a rather flat ellipticity curve (Bolu, Forlì and Sturno), but even these curves carry important information on the soil structure. For Knidi, no qualitative conclusions can be drawn as every station shows a different curve although the stations are located not farther away than 10 meters one from each other.

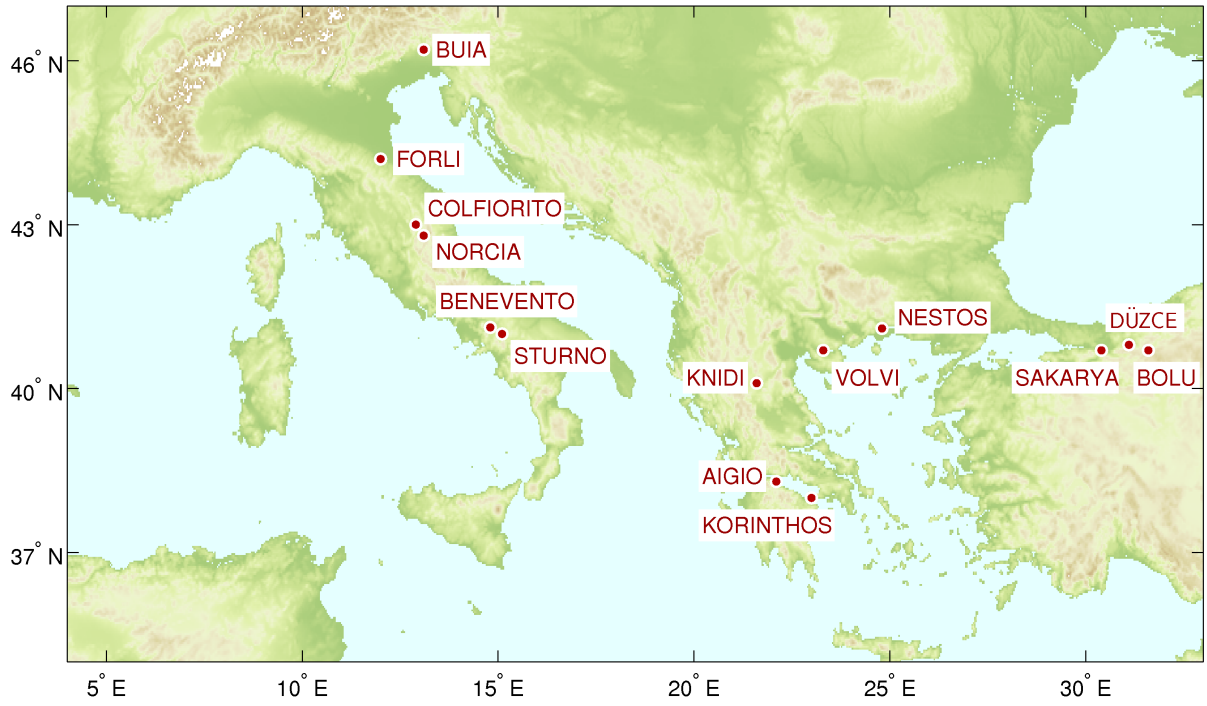


Figure 7.2: Location of the NERIES sites used in this study.

Table 7.1: Table of investigated NERIES sites, including their respective classification after the European EC8 code, the bedrock depth determined by borehole measurements (N/A indicates that the borehole did not reach the bedrock) and the radius of the smallest array of seismic sensors used for the NERIES measurements.

Site	EC8 category	Bedrock depth [m]	Smallest array radius [m]
Aigio	B	20	not used
Benevento	B	130	7.14−9.48
Bolu	C	N/A	7.56−9.29
Buia	C	48	9.32−11.50
Colfiorito	D	54	12.79−16.00
Düzce	C	N/A	8.01−8.83
Forlì	C	N/A	7.41−9.35
Knidi	E	16	7.13−8.59
Korinthos	C	N/A	9.41−10.44
Nestos	C	N/A	6.59−7.24
Norcia	B	N/A	8.67−9.97
Sakarya	B	N/A	11.83−13.67
Sturno	B	N/A	8.15−8.77
Volvi	C	196	10.42−13.23

We cannot show the inversions performed for each site in detail here. Therefore, we will focus on the results for some typical sites before giving an overview of the inversion results for the other sites.

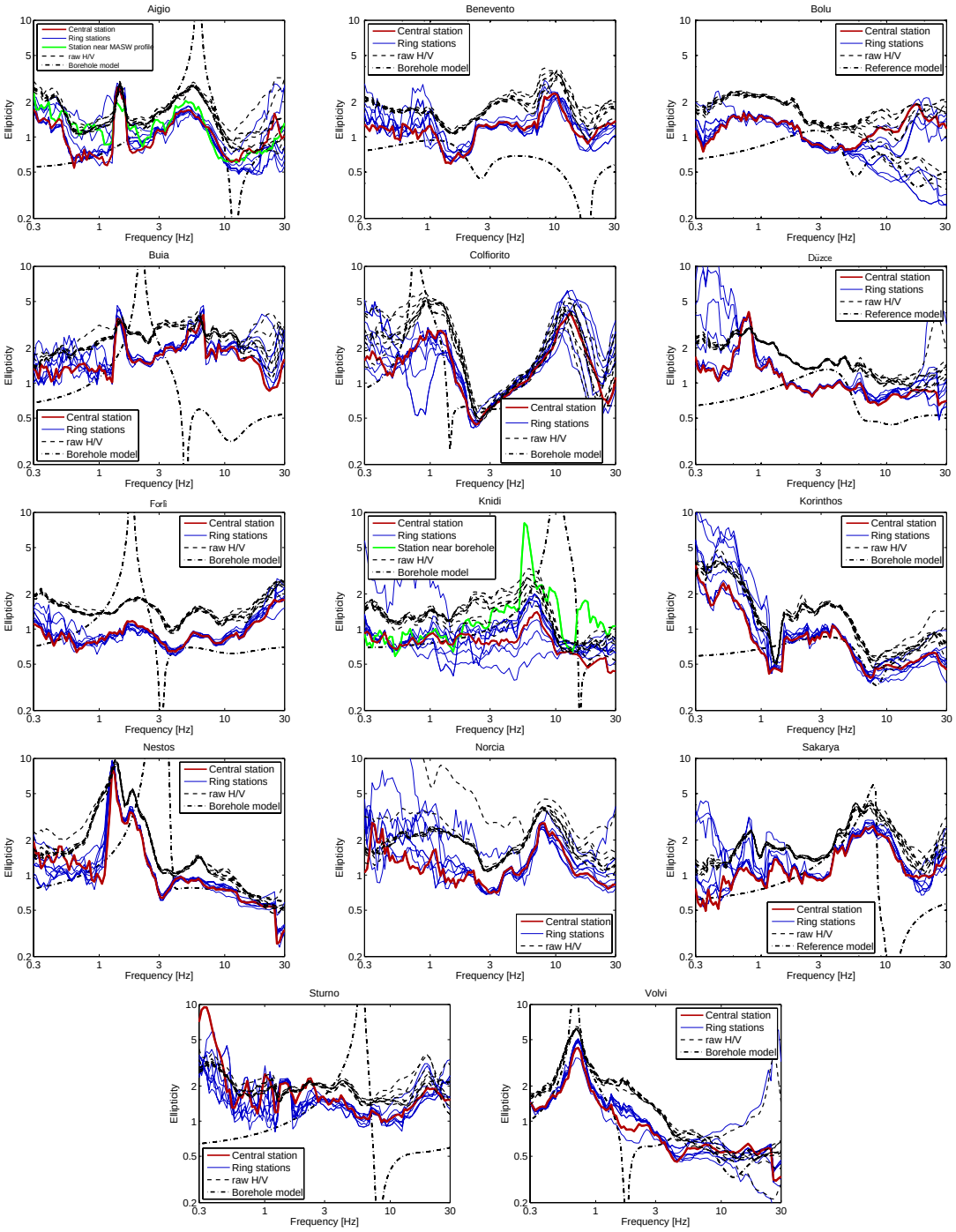


Figure 7.3: Overview of the ellipticity curves measured with the RayDec technique for the stations of the smallest array for each site. For comparison reasons, the corresponding raw H/V curves are shown as well as the ellipticity curve for the reference measurements. For Norcia, the low velocity zones of the borehole model did not allow the calculation of a theoretical ellipticity curve.

We chose Volvi and Nestos as sites with a clear ellipticity peak and Colfiorito as its two peaks are worth a more detailed investigation. Futhermore, Aigio will serve as an example for a joint inversion of ellipticity and MASW dispersion curve measurements, as local constraints on the array layout do not allow to perform small-scale SPAC measurements. As a second site without a clear peak, the inversion of Korinthos will be presented.

7.5.1 Models with clear ellipticity peak

7.5.1.1 Volvi

The Euroseistest site located in Volvi, Greece, is a well-studied test site for seismic measurements [Raptakis et al., 1998]. The valley is filled by sediments up to a depth of 200 meters. For the analysis presented here, we used the smallest array of sensors of the NERIES measurements. The relative positions of the individual sensors are shown in figure 7.4 (a). 8-station and 4-station spatial autocorrelation measurements were performed. The 8-station measurements were performed for seven pairs of stations with distances between 10.42 and 13.23 m. The three station pairs used for the 4-station measurement are located at distances between 12.86 and 15.41 m. A total of 45 minutes of seismic noise recordings were analyzed. The corresponding spatial autocorrelation curves are shown in figure 7.4 (b) with their respective measurement errors. Although the mean distances of the corresponding autocorrelation measurements differ slightly, both curves have the first root at the same frequency. At higher frequencies, both curves differ, but are still in agreement considering the error bars. Due to the averaging over less station pairs, the autocorrelation curve's error bars are larger for the 4-station case than for the 8-station case by a factor of about 1.4. Nevertheless, both curves are very similar and similar inversion results can be expected for both inversions.
The ellipticity curves for all seismic stations of the array are shown in figure 7.3. All curves exhibit a clear ellipticity peak at 0.73 Hz. The trough frequency seems to be at 4.35 Hz, but the ellipticity value of 0.5 would be quite large for the trough. The ellipticity for the borehole model which is also shown in the figure, is quite similar to the measurements around the peak frequency, but has a trough at 1.76 Hz. At higher frequencies, the measured curves are comparable to the borehole model. Figure 7.4 (c) shows the ellipticity curve for the central station which was obtained by cutting the signal into five pieces of 9 minutes each, applying RayDec to each of them and averaging the resulting curves. The averaging generates error bars as well. As the ellipticity curves are susceptible to misestimation both around the peak and around the trough frequencies, we disregard ellipticity values above 3.

Table 7.2: Table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for the Volvi data set.

Figure reference	Autocorrelation curve			Ellipticity curve	Number of layers (incl. bedrock)	Number of models	Minimum misfit	P value (0.65 - 28.0 Hz)
	Stations	Distance range	Frequency range					
Fig. 7.4 (d)	8	10.42 - 13.23 m	2.7 - 20 Hz	0.40 - 0.62 Hz, 0.82 - 1.25 Hz	5	100 100	0.270	0.794 ± 0.038
Fig. 7.4 (e)	4	12.86 - 15.41 m	2.7 - 20 Hz	0.40 - 0.62 Hz, 0.82 - 1.25 Hz	5	100 100	0.302	0.718 ± 0.041

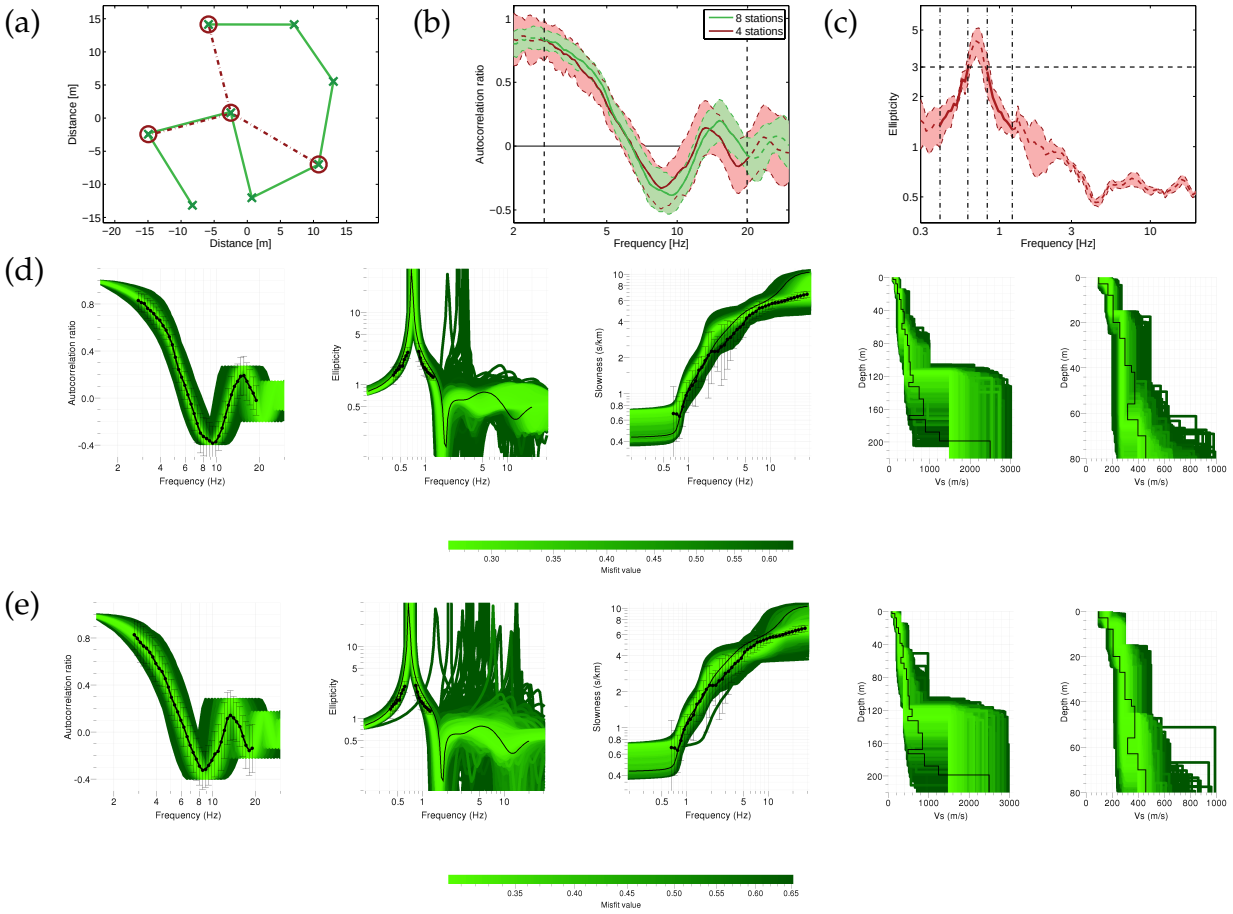


Figure 7.4: (a) Relative positions of the seismic sensors used for the study of Volvi. The station pairs contributing to the spatial autocorrelation curves for the 8-station and the 4-station case are also shown. (b) Spatial autocorrelation curves using 8 stations (7 station pairs, distances between 10.42 and 13.23 m) and 4 stations (3 station pairs, distances between 12.86 and 15.41 m). The shaded areas indicate the measurement errors. The bold parts of the curves, delimited by the dashed lines, were used for the inversions. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into five parts of 9 minutes each, applying RayDec and averaging the results. The bold parts of the curve, delimited by the dashed lines, were used for the inversions. (d) and (e): Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Volvi data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right): (d) using the 8-station SPAC measurement (e) using the 4-station SPAC measurement.

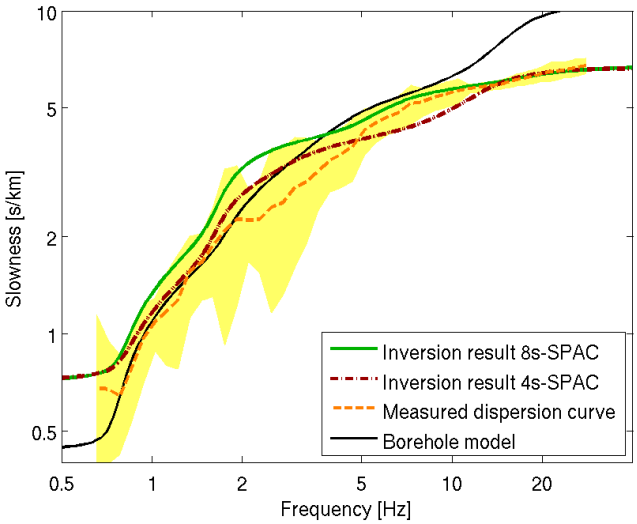


Figure 7.5: Dispersion curves for Volvi: Best models of the inversion using the 8-station and 4-station autocorrelation curves, the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area.

Beyond 1.25 Hz, the error bars are relatively large suggesting that the ellipticity estimation is unstable at that frequency. The same instability can be seen in figure 7.3 for the seismic stations on the ring as well. Therefore, we chose to invert the right flank of the peak between 0.82 and 1.25 Hz only. However, an inversion of the right flank of the peak only yields soil models with too low peak frequencies (below 0.6 Hz). Therefore, the peak frequency has to be constrained as well. We could constrain it directly, but then it is not obvious how this information has to be weighted compared to the other ellipticity data. Therefore, it is simpler to constrain the peak frequency by including the left flank of the peak in the inversion process. Consequently, the parts of the ellipticity curve we used for the inversion process consist of the frequency ranges between 0.40 and 0.62 Hz and between 0.82 and 1.25 Hz.
We inverted the ellipticity curve (Fig. 7.4 (c)) jointly with the spatial autocorrelation curves (Fig. 7.4 (b)), parameterizing the soil structure as 4 homogeneous layers overlying a homogeneous half-space. Figure 7.4 (d) shows the results obtained for using the 8-station autocorrelation curve, (e) the results for using the 4-station autocorrelation curve. The autocorrelation and ellipticity curves are well fitted in both inversions (see table 7.2). For both the 4-station and 8-station inversion, the borehole model's dispersion curve is overall quite well fitted, with major deviations at high and low frequencies. The inverted shear wave velocity profiles are in agreement with the borehole model down to 120 meters of depth. The bedrock depth and subsequently the bedrock velocity are underestimated.
The P values given in table 7.2 are calculated using the direct dispersion curve measurement as $s_{meas}(f)$ in equation (7.3). The dispersion curves are compared in figure 7.5, where for clarity reasons only the best-fitting model for each inversion is indicated. The dispersion curves for all models generated in the inversion are shown in figure 7.4.

The P value is nevertheless averaged over the set of best-fitting models. The inversion results using the 8-station and the 4-station autocorrelation curve are both in good agreement with the directly measured dispersion curves. Furthermore, between 1 and 10 Hz, the dispersion curves resulting from the inversions and the direct measurement are more or less in agreement with the borehole model dispersion curve, but at lower and higher frequencies, the curves dissent significantly.

7.5.1.2 Nestos

The relative positions of the sensors used for the Nestos data are shown in figure 7.6 (a). The autocorrelation curves shown in figure 7.6 (b) have been calculated using the measurements of 8 sensors and 4 sensors, respectively. Both autocorrelation curves are consistent, the error bars using 8 station pairs are about half the error bars using 3 station pairs. The ellipticity curves estimated for each of the 8 stations are shown in figure 7.3. All stations show a clear peak at 1.27 Hz and a minimum at 3.1 Hz. The ellipticity curves of all stations are identical above the peak frequency. Therefore, the small bump in the right flank could correspond to a real feature of the site. The corresponding curve obtained by applying RayDec to six 10-minute time windows of signal on the central station and averaging the results is shown in figure 7.6 (c). We chose to invert the right flank of the ellipticity curve between the bump and the trough and to include the left flank only to constrain the peak frequency. As the peak is closer to the left than to the right flank, special attention has to be paid on the resulting models' peak frequency. The autocorrelation curves of figure 7.6 (b) have been inverted jointly with the parts of the ellipticity curve shown in figure 7.6 (c). The parameters and resulting misfit and P values of the inversions are given in table 7.3. Figure 7.6 (d) and (e) show the results using the 8-station and 4-station autocorrelation curves, respectively. In both cases, the inversion targets are well fitted. Both inversions yield a bedrock depth of about 70 meters. The borehole does not reach the seimic bedrock, i.e. a strong velocity contrast. The borehole model's dispersion curve is not well fitted, which is not surprising when considering the discrepancy between the measured ellipticity peak frequency and the one corresponding to the borehole model. However, the inversion results are in good agreement with the direct dispersion curve measurements. The different dispersion curves are shown in figure 7.7. The dispersion curves for both inversions are compatible with the measured dispersion curve. Furthermore, both curves deviate considerably from each other for frequencies above 16 Hz.

Table 7.3: table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for the Nestos data set.

Figure reference	Autocorrelation curve			Ellipticity curve	Number of layers (incl. bedrock)	Number of models	Minimum misfit	P value (1.6 - 36.3 Hz)	P value (1.6 - 16.4 Hz)
	Stations	Distance range	Frequency range						
Fig. 7.6 (d)	8	6.59 - 7.24 m	3 - 20 Hz	0.80 - 1.12, 1.75 - 3.0 Hz	4	100 100	0.423	0.799 ± 0.063	0.592 ± 0.018
Fig. 7.6 (e)	4	7.09 - 7.24 m	3 - 20 Hz	0.80 - 1.12, 1.75 - 3.0 Hz	4	100 100	0.262	1.158 ± 0.248	0.549 ± 0.028

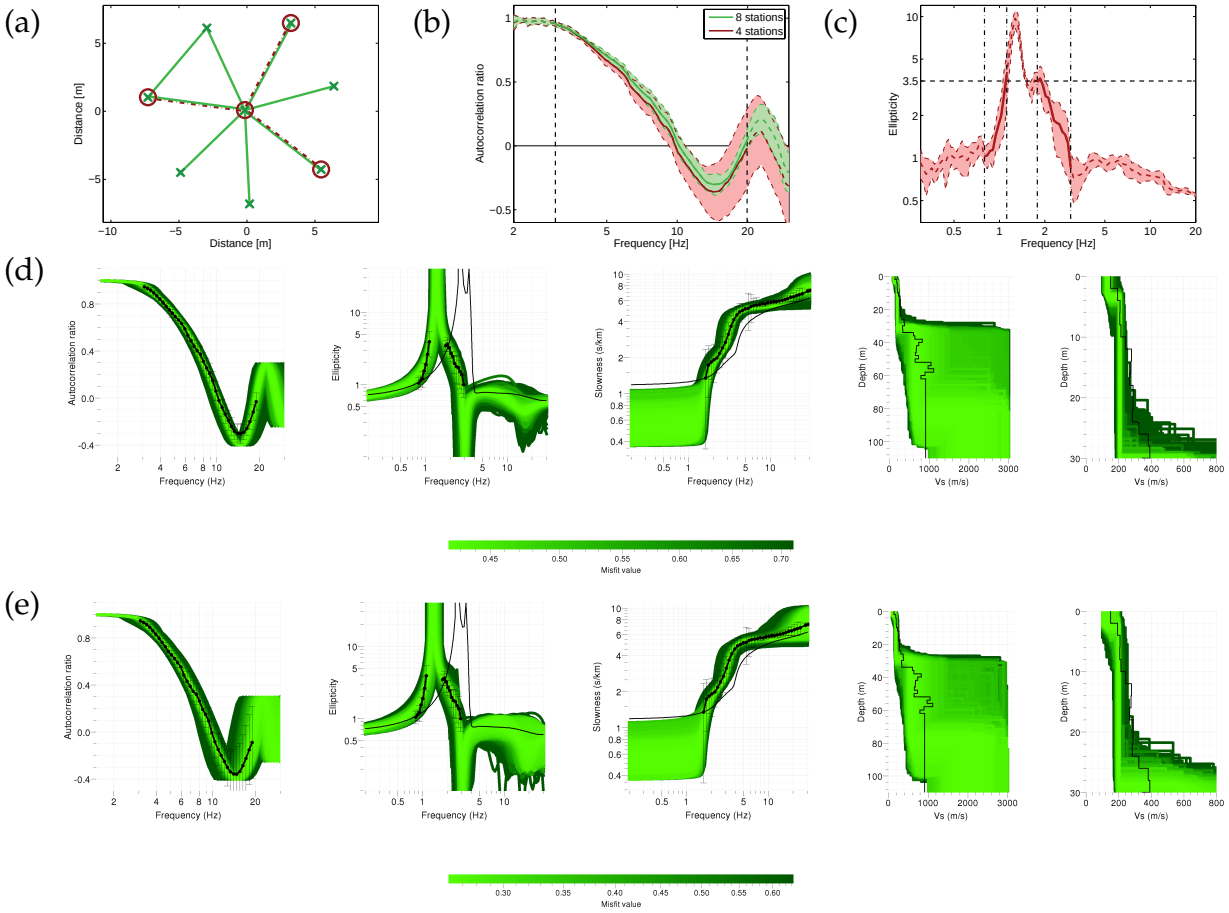


Figure 7.6: (a) Relative positions of the seismic sensors used for the study of Nestos. The station pairs contributing to the spatial autocorrelation curves for the 8-station and the 4-station case are also shown. (b) Spatial autocorrelation curves using 8 stations (8 station pairs, distances between 6.59 and 7.24 m) and 4 stations (3 station pairs, distances between 7.09 and 7.24 m). The shaded areas indicate the measurement errors. The bold parts of the curves, delimited by the dashed lines, were used for the inversions. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into six parts of 10 minutes each, applying RayDec and averaging the results. The bold parts of the curve, delimited by the dashed lines, were used for the inversions. (d) and (e): Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Nestos data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right): (d) using the 8-station SPAC measurement (e) using the 4-station SPAC measurement.

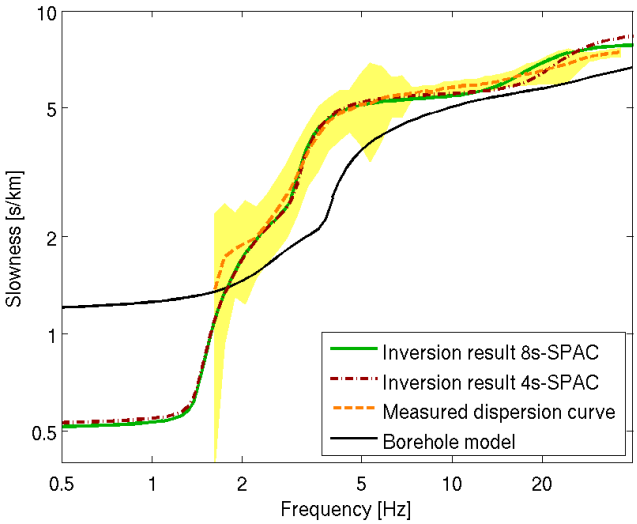


Figure 7.7: Dispersion curves for Nestos: Best models of the inversion using the 8-station and 4-station autocorrelation curves, the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area.

7.5.1.3 Colfiorito

Colfiorito basin is a well-known site extensively investigated by different geophysical methods [Di Giulio et al., 2006], where borehole measurements reached the bedrock at a depth of about 50 meters. The locations of the seismic sensors are shown in figure 7.8 (a). We used all eight sensors to perform spatial autocorrelation measurements at distances between 12.79 and 16.00 m. A second autocorrelation measurement using only three stations with distances between 14.03 and 15.62 m was performed using the same data. The resulting autocorrelation curves are shown in figure 7.8 (b). Below the first root, both curves are almost identical, although the measurement errors are much smaller for the 8-station autocorrelation. At higher frequencies, the differences are larger, but each curve is still within the error bars of the other curve.
The ellipticity measurements for each station are shown in figure 7.3. All curves exhibit two different peaks, one around 0.96 Hz and a second at 12.4 Hz. We identify the first peak as the fundamental peak. The interpretation of the second peak is not straightforward, but its frequency is too different from the fundamental frequency to be an associated harmonic mode.

Table 7.4: Table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for the Colfiorito data set.

Figure reference	Autocorrelation curve			Ellipticity curve	Number of layers (incl. bedrock)	Number of models	Minimum misfit	P value (2.3 - 11.2 Hz)
	Stations	Distance range	Frequency range					
Fig. 7.8 (d)	8	12.79 - 16.00 m	2 - 12 Hz	0.50 - 0.83, 1.3 - 2.0 Hz	3	100 100	0.432	0.959 ± 0.015
Fig. 7.8 (e)	3	14.03 - 15.62 m	2 - 12 Hz	0.50 - 0.83, 1.3 - 2.0 Hz	3	100 100	0.469	0.609 ± 0.039
Fig. 7.8 (f)	3	14.03 - 15.62 m	2 - 12 Hz	0.50 - 0.83, 1.3 - 2.0, 3.5 - 10.0, 15.5 - 23.3 Hz	3	100 100	0.463	0.585 ± 0.017

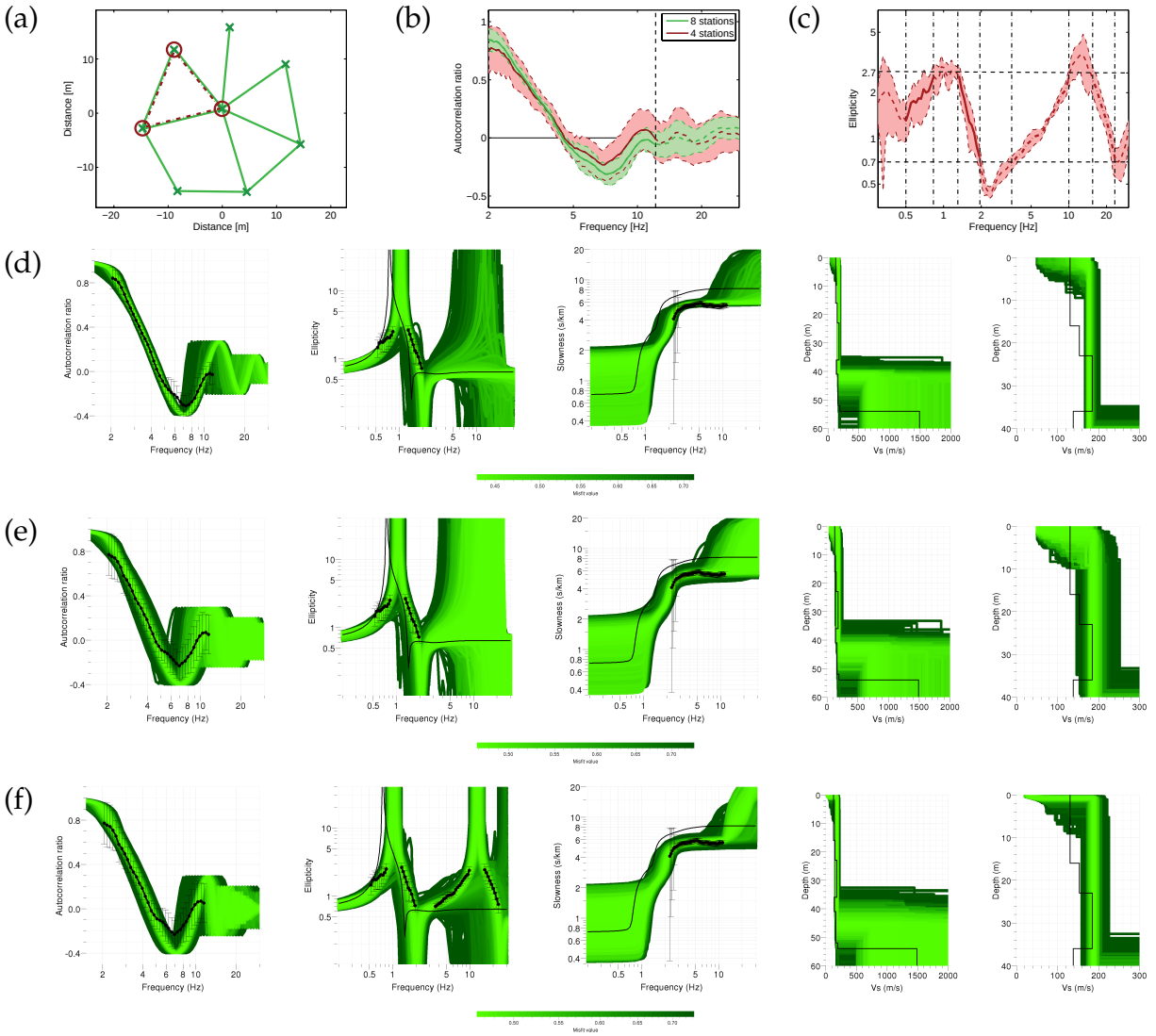


Figure 7.8: (a) Relative positions of the seismic sensors used for the study of Colfiorito. The station pairs contributing to the spatial autocorrelation curves for the 8-station and the 3-station case are also shown. (b) Spatial autocorrelation curves using 8 stations (11 station pairs, distances between 12.79 and 16.00 m) and 3 stations (3 station pairs, distances between 14.03 and 15.62 m). The shaded areas indicate the measurement errors. The bold parts of the curves, delimited by the dashed lines, were used for the inversions. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into five parts of 9 minutes each, applying RayDec and averaging the results. The bold parts of the curve, delimited by the dashed lines, were used for the inversions. For inversion (f), the dash-dotted part of the curve was also included in the inversion. (d) - (f): Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Colfiorito data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right): (d) using the 8-station SPAC measurement, (e) using the 3-station SPAC measurement, (f) using the 3-station SPAC measurement and including the second ellipticity peak.

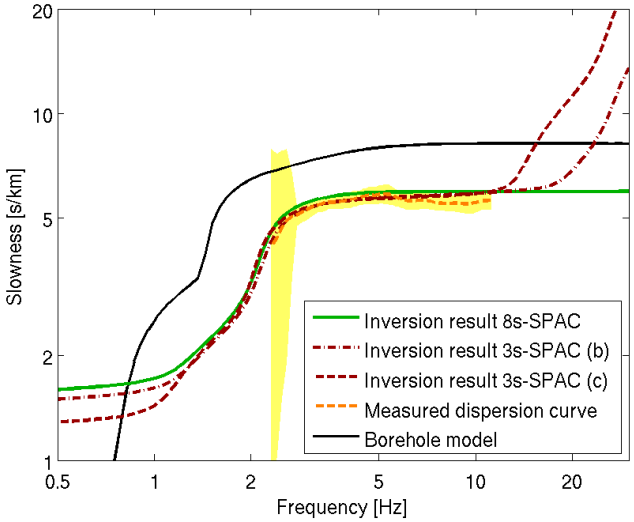


Figure 7.9: Dispersion curves for Colfiorito: Best models of the inversion using the 8-station and 3-station autocorrelation curves, the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area.

Therefore, we suggest that it is related to a small-scale resonance of superficial layers [Guéguen et al., 1998]. Figure 7.8 (c) shows the mean of the ellipticities obtained by cutting the signal of the central station into five pieces of 9 minutes each, applying RayDec to them and averaging the results. The bold parts were used for the inversions. As the first peak is smaller than 3, we omitted ellipticity values above 2.7 Hz, in order to allow for soil models with a singular peak. Here again, the left flank of the first peak serves to constrain the peak frequency only. Ellipticity values below 0.7 were omitted as well, in order to leave the ellipticities around the trough unconstrained. For the inversion including the right peak, we used the same threshold values for both peaks.
The given curves were inverted for models consisting of two homogeneous layers over a homogeneous half-space. As this parameterization succeeds in retrieving a soil structure explaining the measurements, we did not search for more complex parameterizations. The parameters and resulting misfit and P values of the inversions are given in table 7.4. In figures 7.8 (d) and (e), only the first peak has been inverted using the 8-station and the 3-station autocorrelation curve, respectively. Both inversions give almost identical results. The respective autocorrelation and ellipticity curves are well fitted, the inversions yield a bedrock depth of about 42 meters. The resulting dispersion curves are in agreement with the direct dispersion curve measurement. The 3-station inversion yields a better P value than the 8-station inversion. Both the ellipticity measurement and the inversion results are in contradiction with the borehole model. In figure 7.9, the dispersion curves resulting from the inversion are compared to the borehole model's and the directly measured curves.

The indetermination of the resulting ellipticity curves at higher frequencies suggested that including the second ellipticity peak might improve the inversion. The resulting inversion obtained by using the second peak and attributing it to the fundamental mode are shown in figure 7.8 (f) for the 3-station autocorrelation curve only. The autocorrelation and ellipticity curves are still well fitted. The resulting dispersion curve exhibits a practically identical P value than without including the second peak. In figure 7.9, the inclusion of the second peak leaves the dispersion curve unchanged below 12 Hz, but changes it drastically above. The inversion algorithm models the second peak by introducing a large velocity contrast between the first and the second layer. For the best-fitting model, this leads to a first layer which is 75 cm thick with a velocity of 40 m/s. The deeper layers are not affected by the second ellipticity peak. Finally, all inversions yield models which agree with the direct dispersion curve measurements. Including the second ellipticity peak does not significantly improve the inversion result compared with the direct dispersion curve measurements. This result is compatible with the presence of a very thin and very slow superficial layer, but as the inversion without the second peak yields comparable results without assuming such a layer, this hypothesis can neither be proven nor refuted. However, the attribution of the second peak to the fundamental Rayleigh wave mode would require that no higher modes which would alter the measurement were present in the wave field at all. Additional MASW measurements at high frequencies could help in identifying the nature of the second ellipticity peak.

7.5.2 Sites without clear ellipticity peak

7.5.2.1 Aigio

At Aigio, the array layout of the smallest ring was not adapted for small-scale spatial autocorrelation measurements and the next ring had a diameter of 30 meters. As we want to show the possibility to retrieve the soil structure by small scale measurements, we decided to use the dispersion curve obtained by MASW measurements instead. This dispersion curve is shown in figure 7.10 (a). As the part of the curve between 12 and 15 Hz is quite unstable, we use the curve between 15 and 28 Hz only. The ellipticity curve shown in figure 7.10 (b) was measured by the seismic station which was closest to the MASW profile. It is the mean curve for six 10-minute time windows of seismic noise recordings. Calculating the apparent damping of the signal at 1.45 Hz identified the artificial origin of this peak. Therefore, the fundamental peak is the broad second peak with an amplitude of 2 around 5 Hz. As the peak is broad, the complete peak between 3.4 and 10.5 Hz was inverted.

Table 7.5: Table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for the Aigio data set.

Figure reference	Dispersion curve	Ellipticity curve	Number of layers (incl. bedrock)	Number of models	Minimum misfit	P value (7.3 - 28.5 Hz)
Fig. 7.10 (d)	15 - 28 Hz	3.4 - 10.5 Hz	4	100 100	0.144	0.506 ± 0.014

The high frequency of the ellipticity peak justifies the use of the high-frequency dispersion curve above 15 Hz.
The misfit and P values of the joint inversion of the dispersion and ellipticity data are given in table 7.5, the inversion results are shown in figure 7.10. Models consisting of three homogeneous layers overlying a half-space were necessary to correctly fit the measurements. The inversion targets are well fitted, the dispersion curve disagrees with the borehole model's curve below 15 Hz. The directly measured dispersion curve was obtained between 7.3 and 28.5 Hz. In this range, the curve is in good agreement with the inversion result. Figure 7.10 (c) shows a comparison of the directly measured, borehole model and inverted dispersion curves. The shear wave velocity profiles match the borehole model quite well, except for the layer boundary which is given at 20 m by the borehole and found at about 42 m by our inversions.

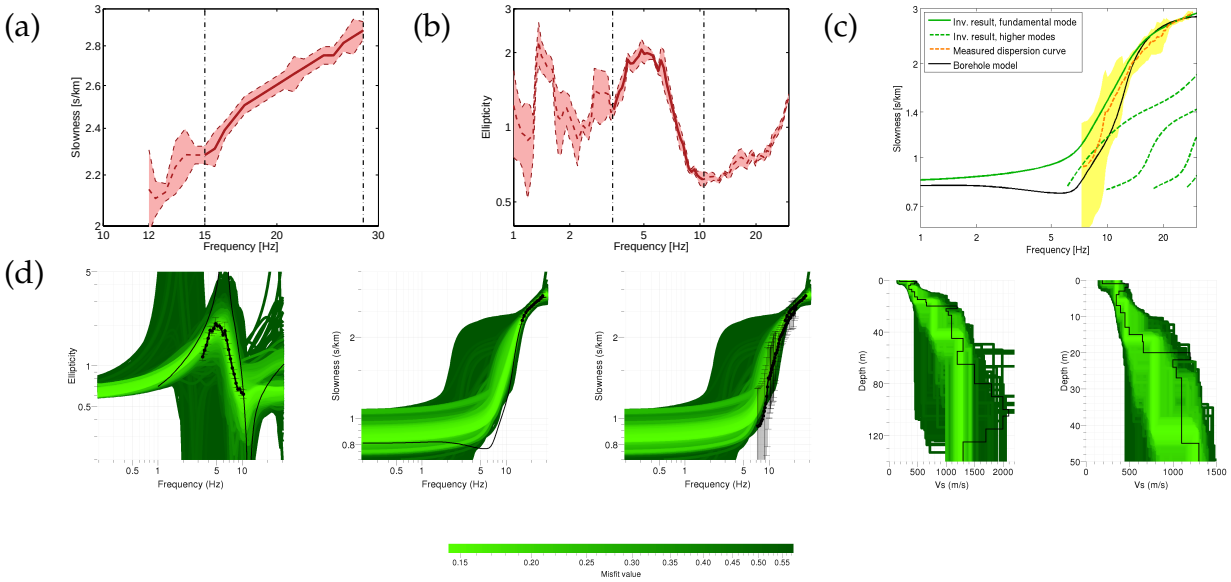


Figure 7.10: (a) Dispersion curve obtained by MASW measurements. The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (b) Mean ellipticity and standard deviation obtained by cutting the signal of station next to the MASW measurements into six parts of 10 minutes each, applying RayDec and averaging the results. The shaded areas indicate the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Dispersion curves for Aigio: Best models of the inversion, the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (d) Inversion results for joint inversions of the ellipticity curve (left, with data points) and dispersion curve (center left: data points with borehole model, center: with dispersion curve data from array measurements) for the Aigio data set. The resulting shear wave velocity profiles are shown on the right (center right: velocity profiles, right: zoom on the superficial 40 meters). The black lines the curves for the borehole model.

7.5.2.2 Korinthos

The locations of the seismic sensors used for the measurements at Korinthos are shown in figure 7.11 (a). The distances used for the autocorrelation measurements are slightly different in the 7-sensor and the 3-sensor case. This explains the differences in the spatial autocorrelation curves shown in figure 7.11 (b). The ellipticity curves for the different sensors of the array (see figure 7.3) are in good agreement above 1.5 Hz. At lower frequencies, every station exhibits a clear peak, but only the right flank is resolved. The trough at 1.3 Hz is visible for all sensors and can be attributed to an artificial source on the vertical component. Figure 7.11 (c) shows the ellipticity curve obtained by averaging the results of an application of RayDec to four different time windows of 7.5 minutes each. Only the right flank of the peak between 0.5 and 1.1 Hz is used for the inversions. The parameters of the inversions using both autocorrelation curves are given in table 7.6. The results of both inversions are shown in figure 7.11. Both inversions succeed in fitting the inversion targets and yield comparable results. The 7-station and 3-station inversions find the bedrock at depths of 213 and 204 meters, respectively. The borehole measurements did not exceed 40 meters. In figure 7.11, the dispersion curves generated by our inversion are in good agreement with the directly measured curve for frequencies above 7 Hz only. A comparison of the different dispersion curves is given in figure 7.12. For the best-fitting models of both inversions, the first four higher modes are indicated as well. Both ellipticity inversions give similar results. The ellipticity inversions suggest that the directly measured dispersion curve corresponds to the fundamental mode at frequencies above 7 Hz only. At lower frequencies, a mixture of higher modes dominates the dispersion curve measurements. Nevertheless, it is assumed for the inversion that the inverted ellipticity curve corresponds to the fundamental mode. The deviation of the inversion results from the borehole measurements is acceptable. The small depression in both the borehole and the directly measured dispersion curves between 20 and 50 Hz indicates a low velocity zone just beneath the surface (or a small surface layer of high velocity). This feature cannot be found in the ellipticity inversions because the autocorrelation curves for distances of about 10 meters are insensitive to such a surface effect.
As the ellipticity inversion indicates that the directly measured dispersion curve belongs to different modes, the P values which are shown in table 7.6 are calculated on the basis of the lowest three modes, taking the mode which is closest to the measured curve at each frequency value.

Table 7.6: Table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for the Korinthos data set.

Figure reference	Autocorrelation curve			Ellipticity curve	Number of layers (incl. bedrock)	Number of models	Minimum misfit	P value (2.3 - 60.5 Hz)
	Stations	Distance range	Frequency range					
Fig. 7.11 (d)	7	10.43 - 11.90 m	5 - 20 Hz	0.50 - 1.10 Hz	3	100 100	0.249	0.652 ± 0.021
Fig. 7.11 (e)	3	9.41 - 10.44 m	5 - 20 Hz	0.50 - 1.10 Hz	3	100 100	0.200	0.587 ± 0.028

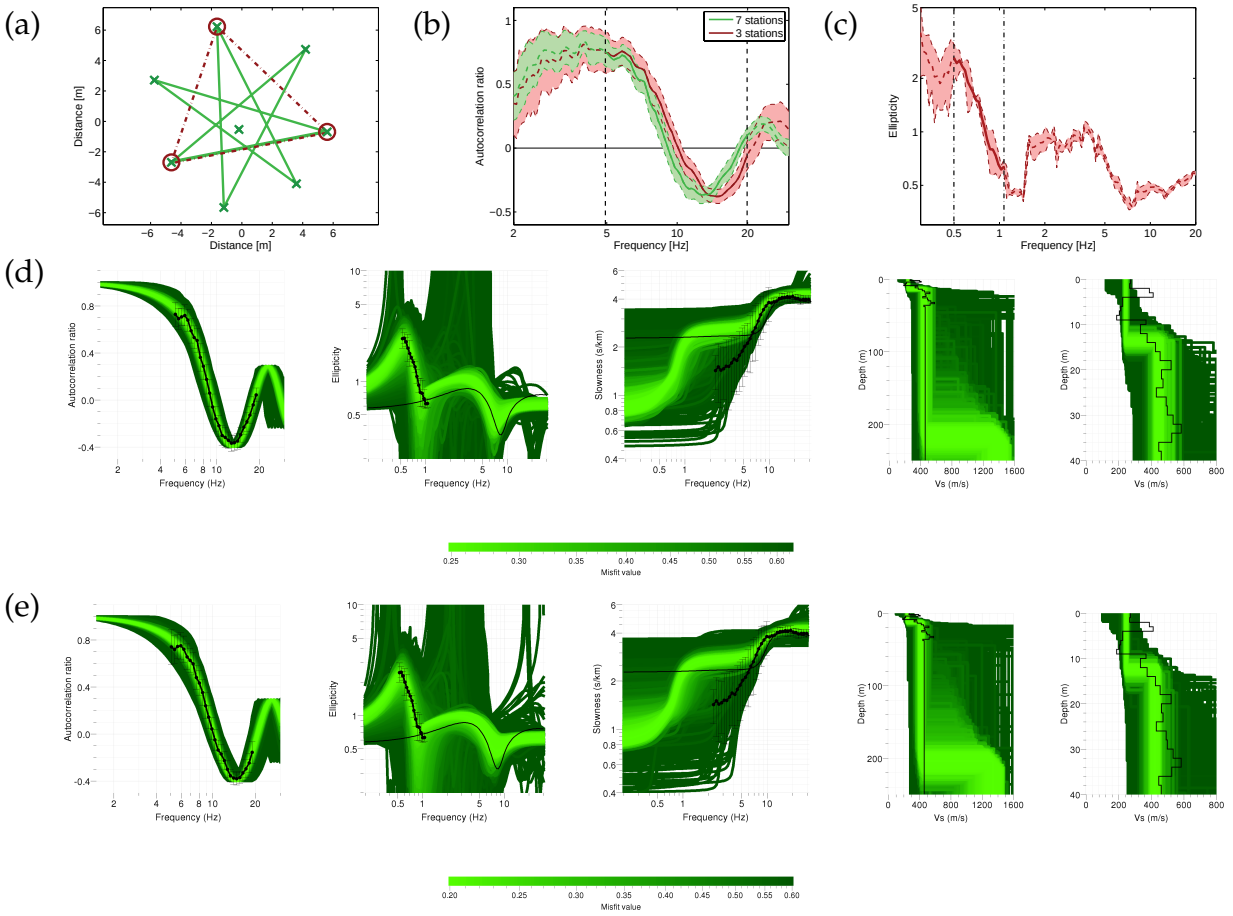


Figure 7.11: (a) Relative positions of the seismic sensors used for the study of Korinthos. The station pairs contributing to the spatial autocorrelation curves are also shown. (b) Spatial autocorrelation curves using 7 stations (7 station pairs, distances between 10.43 and 11.90 m) and 3 stations (3 station pairs, distances between 9.41 and 10.44 m). The shaded areas indicate the measurement errors. The bold parts of the curves, delimited by the dashed lines, were used for the inversions. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into four parts of 7.5 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) and (e): Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Korinthos data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right): (d) using the 7-station SPAC measurements (e) using the 3-station SPAC measurement.

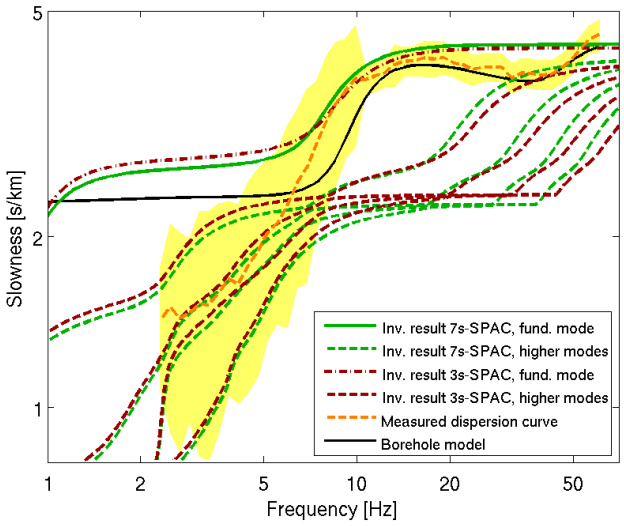


Figure 7.12: Dispersion curves for Korinthos: Best models of the inversion using the 8-station and 3-station autocorrelation curves compared to the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area.

7.5.3 Inversion results for the other sites

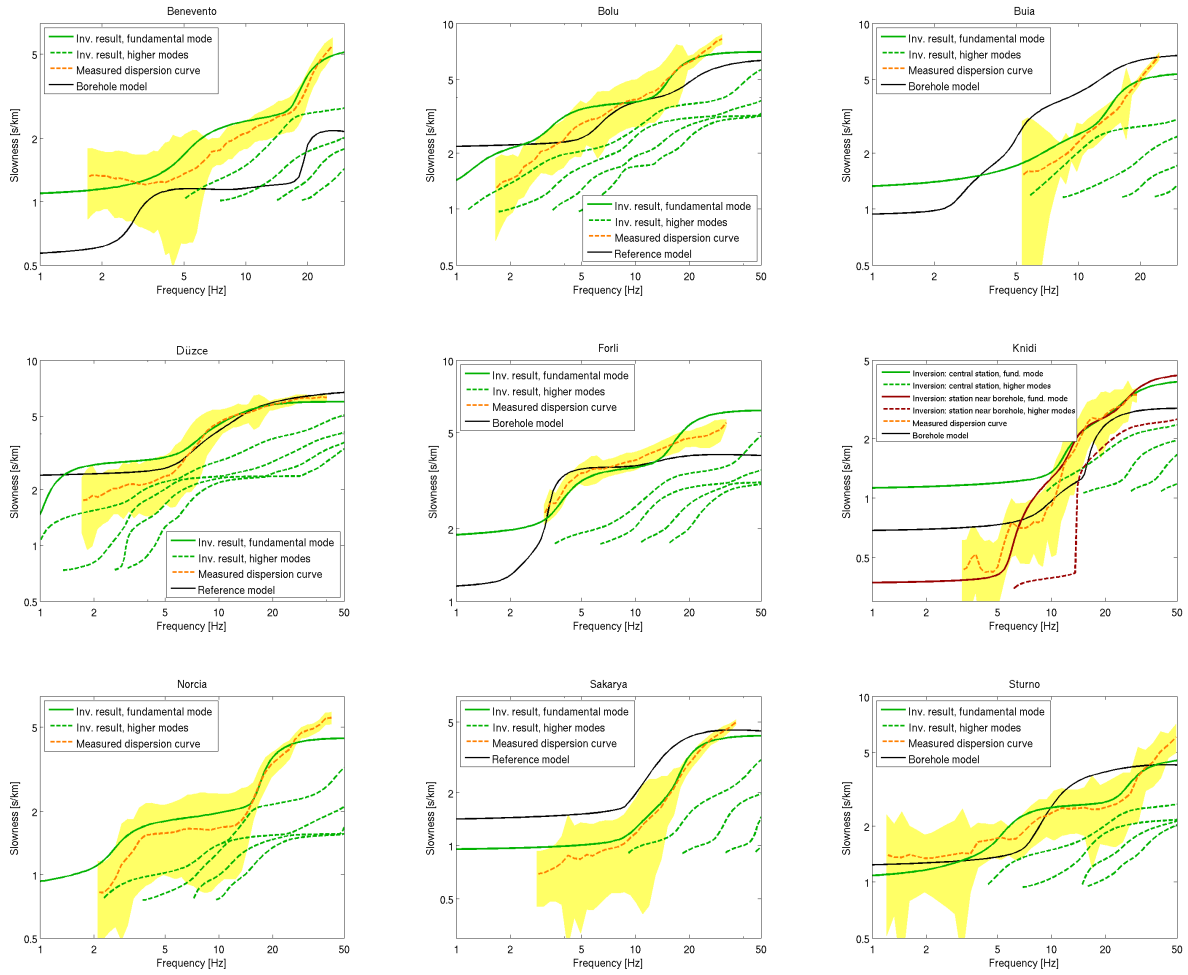


Figure 7.13: Overview of the dispersion curves for the sites which are not treated in detail: The results for the best models resulting from joint inversions of ellipticity and autocorrelation curves are compared with the dispersion curves obtained by direct measurements and the reference curves.

In the appendix, the inversions for the nine remaining sites are shown in detail. Figure 7.13 compares the dispersion curves of the best-fitting models with the borehole model dispersion curve as well as the direct dispersion curve measurements. The parameters of the respective inversions, among them the P values, are given in table 7.7. Except for Forlì, P values below 1 could be achieved for all sites. For the sites of Buia, Knidi, Norcia and Sakarya, this can only be achieved by excluding areas of high frequency (above 20 Hz, see table 7.7 for details) from the comparison. These frequencies are constrained by the very superficial layers only, which are not well constrained by the data used for the inversion. An autocorrelation curve which has its first root around 10 Hz does not include sufficient information on the shallowest layers. By including higher frequency measurements, e.g. MASW, in the inversion, these superficial layers could be constrained in a better way.

The ellipticity and autocorrelation measurements used for the inversions presented here are not defined in that frequency range and therefore do not constrain the very superficial layers. Consequently, it cannot be expected that our inversions fit the dispersion curve measurements at high frequencies and this part should not be compared. When neglecting high frequencies, for Benevento, Buia, Norcia, Sakarya and Sturno, a good agreement between the inversions and the direct measurements is given over the whole frequency range. For Bolu and Düzce, the fundamental mode resulting from the inversions and the direct measurements diverge at frequencies below 5 Hz. However, looking at the higher modes of the best model suggests that the direct measurement at these frequencies corresponds to the first harmonic mode rather than to the fundamental one. In this way, the inversion of ellipticity curves can help to determine the modal nature of a measured dispersion curve.
For Knidi, a comparison of the ellipticities of the individual sensors (see figure 7.3) already indicated that the site is quite inhomogeneous and that a special attention has to be paid for the inversion of ellipticity data. In fact, none of the stations of the smallest array shows a clear ellipticity peak, whereas the station near the borehole (which is located about 40 meters west of the array) exhibits a clear peak. Consequently, the inversion results are different depending on which ellipticity curve is used. The inversion of the ellipticity of the station near the borehole is in better agreement with the dispersion curve than the inversion using the ellipticity of the central station of the array (the autocorrelation information used in both inversions are identical). This could indicate that the deep structure of the soil is different below the smallest array and that the soil structure at the borehole is closer to the large-scale soil structure which is measured by the seismic arrays (the radius of the largest array measuring up to 140 meters).

Table 7.7: Table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves for the sites which are not shown in detail.

Site	Autocorrelation curve			Ellipticity curve	Number of layers	Number of models	Minimum misfit	P value	
	Stations	Distance range	Frequency range					Frequency range	value
Benevento	6	7.14 - 9.48 m	6 - 25 Hz	2.0 - 19.1 Hz	3	100 100	0.386	1.7 - 26.4 Hz	0.696 ± 0.093
Bolu	8	7.56 - 9.29 m	3 - 20 Hz	0.4 - 4.0 Hz	4	100 100	0.247	1.7 - 30.3 Hz	0.691 ± 0.173
Buia	6	9.32 - 11.50 m	6 - 20 Hz	0.6 - 10.0 Hz	4	100 100	0.318	5.3 - 24.5 Hz 5.3 - 20.4 Hz	1.300 ± 0.008 0.631 ± 0.013
Düzce	7	8.01 - 8.83 m	6 - 20 Hz	0.5 - 2.7 Hz	4	100 100	0.406	1.7 - 40.0 Hz	0.537 ± 0.035
Forlì	7	7.41 - 9.35 m	3 - 20 Hz	0.7 - 10.0 Hz	3	100 100	0.415	3.1 - 31.9 Hz	1.173 ± 0.032
Knidi	7	8.56 - 10.37 m	10 - 30 Hz	5.0 - 11.6 Hz (central station)	3	100 100	0.284	3.1 - 44.6 Hz 3.1 - 29.9 Hz	2.130 ± 0.044 2.232 ± 0.044
				6.4 - 11.6 Hz (station near borehole)	3	100 100	0.420	3.1 - 44.6 Hz 3.1 - 29.9 Hz	1.198 ± 0.060 0.508 ± 0.014
Norcia	7	8.67 - 9.97 m	8 - 35 Hz	1.0 - 20.1 Hz	3	100 100	0.554	2.1 - 42.5 Hz 2.1 - 28.9 Hz	1.048 ± 0.098 0.632 ± 0.164
Sakarya	7	11.89 - 13.67 m	8 - 25 Hz	3.0 - 5.4 Hz	3	100 100	0.492	2.8 - 36.3 Hz 2.8 - 26.4 Hz	1.064 ± 0.032 0.517 ± 0.021
Sturno	7	8.15 - 8.77 m	6 - 20 Hz	1.75 - 10.0 Hz	3	100 100	0.420	1.2 - 49.4 Hz	0.501 ± 0.021

7.6 Overview of the results for all 14 sites

7.6.1 Dispersion curves

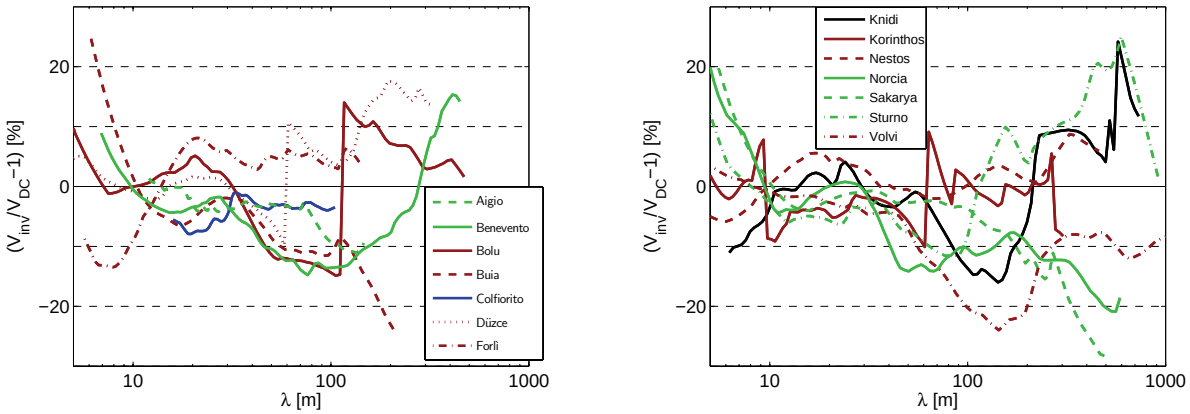


Figure 7.14: Comparison of the percental differences between the dispersion curves of the best-fitting models resulting from the joint inversions of ellipticity and autocorrelation curves with the directly measured dispersion curves for all analyzed sites.

Figure 7.14 shows the percental differences between the directly measured dispersion curves and the ones resulting from the ellipticity inversions. Only the dispersion curve for the respective models with lowest misfit values are shown. In the inversions shown above, spatial autocorrelation curves including all available stations as well as for three station pairs only have been used. As the differences between both cases are not significant, for the other sites only an inversion including the autocorrelation for the whole seismic array (6-8 stations) is shown. Both dispersion curves have been transformed in the wavelength domain before comparison. For Bolu, Düzce and Korinthos, where the inversion results indicate that the measured dispersion curve belongs to different modes, the comparison is shown for the mode of the inversion result which is closest to the direct measurement. This explains the jumps in the respective curves.
For wavelengths between 10 and 100 meters, the differences between both dispersion curves do not exceed 10 % for most of the sites. The deviation for Volvi for wavelengths between 50 and 300 meters can be attributed to an insufficient link between the ellipticity and autocorrelation curves used for inversion, the maximum frequency of the ellipticity target being 1.25 Hz and the minimum frequency of the autocorrelation target 2.7 Hz. This is in agreement with the results of the companion paper, where it is shown that both information used for the inversion must be defined in proximate frequency ranges.
At wavelengths below 10 and above 100 m, the differences between direct measurements and inversions are larger. Short wavelengths correspond to the very superficial structure, which is not well constrained by the used autocorrelation measurements and might be constrained in a better way by active MASW measurements as are included in the directly measured dispersion curves.

For large wavelengths, the bedrock velocity influences the dispersion curves. As has been shown in the first paper on the theoretical aspects of ellipticity inversions, this parameter is badly constrained by the ellipticity curve. A loss of agreement to the direct measurements at these wavelengths is therefore not surprising.

7.6.2 Vs30 values

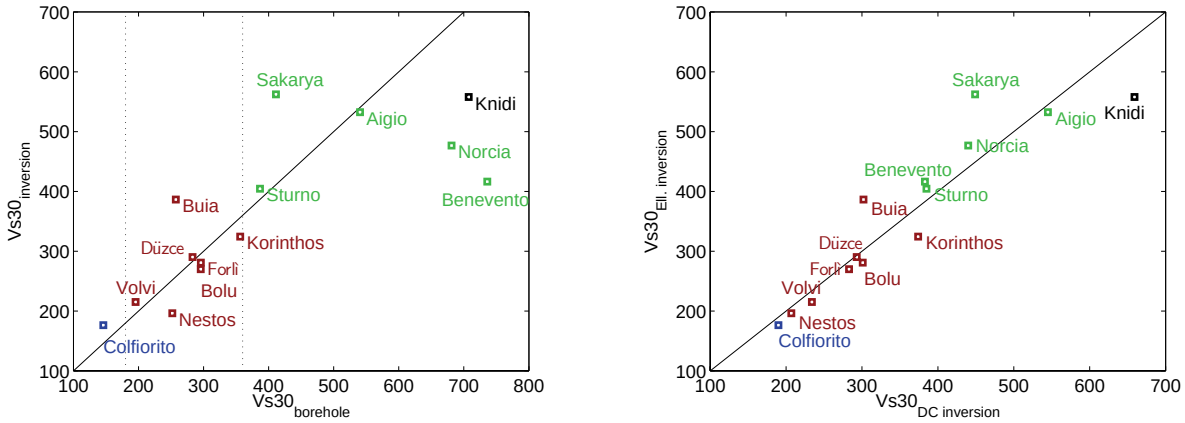


Figure 7.15: Left: Comparison of the Vs30 values from borehole measurements (or reference measurements for Bolu, Düzce and Sakarya) with the values resulting from the joint inversions of ellipticity and autocorrelation curves. The dotted lines indicate the limits of the EC8 classes. Right: Comparison of the Vs30 values obtained by inversion of a broad frequency-range dispersion curve with the joint inversions of ellipticity and autocorrelation curves.

The Vs30 value indicates the mean shear wave velocity over the uppermost 30 meters. This value is used in the actual seismic codes for buildings. In the article on the theoretical aspects of ellipticity inversion, it has been shown that the superficial layers are principally constrained by the autocorrelation curves. However, when measuring autocorrelation curves, the ellipticity can be easily calculated using the same data without further measurement efforts and be included in the inversion process. In this way, joint ellipticity and autocorrelation inversions represent an easier and cheaper way of measurement than borehole measurements. In figure 7.15, the Vs30 values resulting from our ellipticity inversions are compared to the values obtained by borehole measurements (or reference MASW measurements for Bolu, Düzce and Sakarya). For 9 of the 14 analyzed sites, both values are in very good agreement. All the other sites are stiff or rock sites. At least for soft sites, joint inversions involving ellipticity curves present a simple and cheap method to characterize the soil structure.
In figure 7.15, the Vs30 values resulting from our ellipticity inversions are also compared to the Vs30 values which result from the inversion of broad-band dispersion curves [Di Giulio et al., 2010]. Except for some sites exhibiting a complicated soil structure, the Vs30 values resulting from both inversions are in very good agreement.

7.7 Conclusion

In this article, it has been shown that joint inversions of ellipticity and spatial autocorrelation or dispersion curves yield soil structure models which are in good agreement with direct dispersion curve measurements. Furthermore, using only a small number of stations for the SPAC measurements yields results which are in complete agreement with measurements including more sensors. In this way, our method presents an easy and cheap way to investigate the soil structure. For 13 of the 14 investigated sites, P values below 1 can be found if high frequencies are disregarded. An inclusion of information at high frequencies as obtained by MASW measurements could enhance the constraint of the shallowest structure. However, it can be disputed if the expected improvements justify the additional measurement effort. Nonetheless, the presented method is not capable to determine the exact bedrock velocity.
Furthermore, at some sites, the inversions can even help in identifying the fundamental and harmonic Rayleigh wave modes in cases where the directly measured dispersion curve, if regarded separately, could reasonably represent the fundamental mode over the whole frequency range as well. For Buia, Düzce and Korinthos, the ellipticity curve inversions clearly suggest that, at low frequencies, the directly measured dispersion curves are dominated by the first or second harmonic mode.
Although the present study's aim was not to determine Vs30 values, but to find the whole soil structure, it yields very good agreements with borehole measurements for most of the sites. All sites with a worse consistency are stiff or rock sites. At least for soft sites, the presented method can be very helpful for seismic hazard assessment. As the measurements can be performed quickly, a dense survey of a given area can be performed in order to find small-scale irregularities.

Acknowledgments

We would like to thank all those who helped to perform the measurements:
Mehmet Akif Alkan, Marios Anthimidis, Stéphane Drouet, Markus Gurk, Yıldız İravul, Ioannis Kalogeras, Andreas Köhler, Armand Mariscal, Jean-Marc Nicole, Özgür Tuna Özmen, Carsten Riggelsen, Abdullah Sandıkkaya, Paula Teves-Costa, Bekir Tüzel, Daniel Vollmer and Ali Zeynel.

7.8 Appendix: Inversion results for the other sites in detail

7.8.1 Benevento

The Benevento measurements have been performed on a stiff site. The locations of the seismic sensors are shown in figure 7.16 (a). Six sensors forming eight station pairs have been used to calculate the autocorrelation curve (see figure 7.16 (b)). The ellipticity curve (figure 7.16 (c)) was obtained by using 55 minutes of seismic noise recordings of the central station (averaging five windows of eleven minutes). The peak at 10 Hz could be a singularity, but none of our inversions succeeded in fitting a singular peak to this point. Therefore, we take the peak as a flat peak and include it in the inversion process as a whole. In the frequency range between 2.0 and 19.1 Hz, which was used for the inversion, the ellipticity curves of the central station and the ring stations are in good agreement.
The parameters of the inversion are given in table 7.8, the results are shown in figure 7.16 (e). The autocorrelation and ellipticity data are well fitted, the dispersion curve of the inversion is in qualitative agreement with the directly measured one (see figure 7.16 (d)). Both dispersion curves do not fit the borehole measurement.

Table 7.8: Table of the misfit and P values for the best-fitting models of joint inversions of autocorrelation and ellipticity curves.

Site	Figure reference	Autocorrelation curve			Ellipticity curve	Number of layers	Number of models	Minimum misfit	P value	
		Stations	Distance range	Frequency range					Frequency range	value
Benevento	Fig. 7.16 (e)	6	7.14 - 9.48 m	6 - 25 Hz	2.0 - 19.1 Hz	3	100 100	0.386	1.7 - 26.4 Hz	0.696 ± 0.093
Bolu	Fig. 7.17 (e)	8	7.56 - 9.29 m	3 - 20 Hz	0.4 - 4.0 Hz	4	100 100	0.247	1.7 - 30.3 Hz	0.691 ± 0.017
									1.7 - 22.6 Hz	0.373 ± 0.005
Buia	Fig. 7.18 (e)	6	9.32 - 11.50 m	6 - 20 Hz	1.85 - 10.0 Hz	4	100 100	0.293	5.3 - 24.5 Hz	1.300 ± 0.008
									5.3 - 20.4 Hz	0.631 ± 0.013
Düzce	Fig. 7.19 (e)	7	8.01 - 8.83 m	6 - 20 Hz	0.5 - 2.7 Hz	4	100 100	0.406	1.7 - 40.0 Hz	0.537 ± 0.035
Forlì	Fig. 7.20 (e)	7	7.41 - 9.35 m	3 - 20 Hz	0.7 - 10.0 Hz	3	100 100	0.415	3.1 - 31.9 Hz	1.173 ± 0.032
	Fig. 7.20 (f)	7	7.41 - 9.35 m	3 - 20 Hz	-	3	100 100	0.718	3.1 - 31.9 Hz	1.637 ± 0.065
Knidi	Fig. 7.21 (e)	7	8.56 - 10.37 m	10 - 30 Hz	5.0 - 11.6 Hz	3	100 100	0.284	3.1 - 44.6 Hz	2.130 ± 0.044
									3.1 - 29.9 Hz	2.232 ± 0.044
	Fig. 7.21 (f)	7	8.56 - 10.37 m	10 - 30 Hz	6.4 - 11.6 Hz	3	100 100	0.420	3.1 - 44.6 Hz	1.198 ± 0.060
									3.1 - 29.9 Hz	0.508 ± 0.014
	Fig. 7.21 (g)	3	7.13 - 8.59 m	10 - 30 Hz	6.4 - 11.6 Hz	3	100 100	0.308	3.1 - 44.6 Hz	3.584 ± 0.206
									3.1 - 29.9 Hz	0.547 ± 0.013
Norcia	Fig. 7.22 (e)	7	8.67 - 9.97 m	8 - 35 Hz	1.0 - 20.1 Hz	3	100 100	0.554	2.1 - 42.5 Hz	1.048 ± 0.098
									2.1 - 28.9 Hz	0.632 ± 0.164
Sakarya	Fig. 7.23 (e)	7	11.83 - 13.67 m	8 - 35 Hz	3.0 - 5.4, 9.8 - 19.1 Hz	3	100 100	0.492	2.8 - 36.3 Hz	1.064 ± 0.032
									2.8 - 26.4 Hz	0.517 ± 0.021
Sturno	Fig. 7.24 (e)	7	8.15 - 8.77 m	6 - 30 Hz	1.75 - 10.0 Hz	3	100 100	0.420	1.2 - 49.4 Hz	0.501 ± 0.021

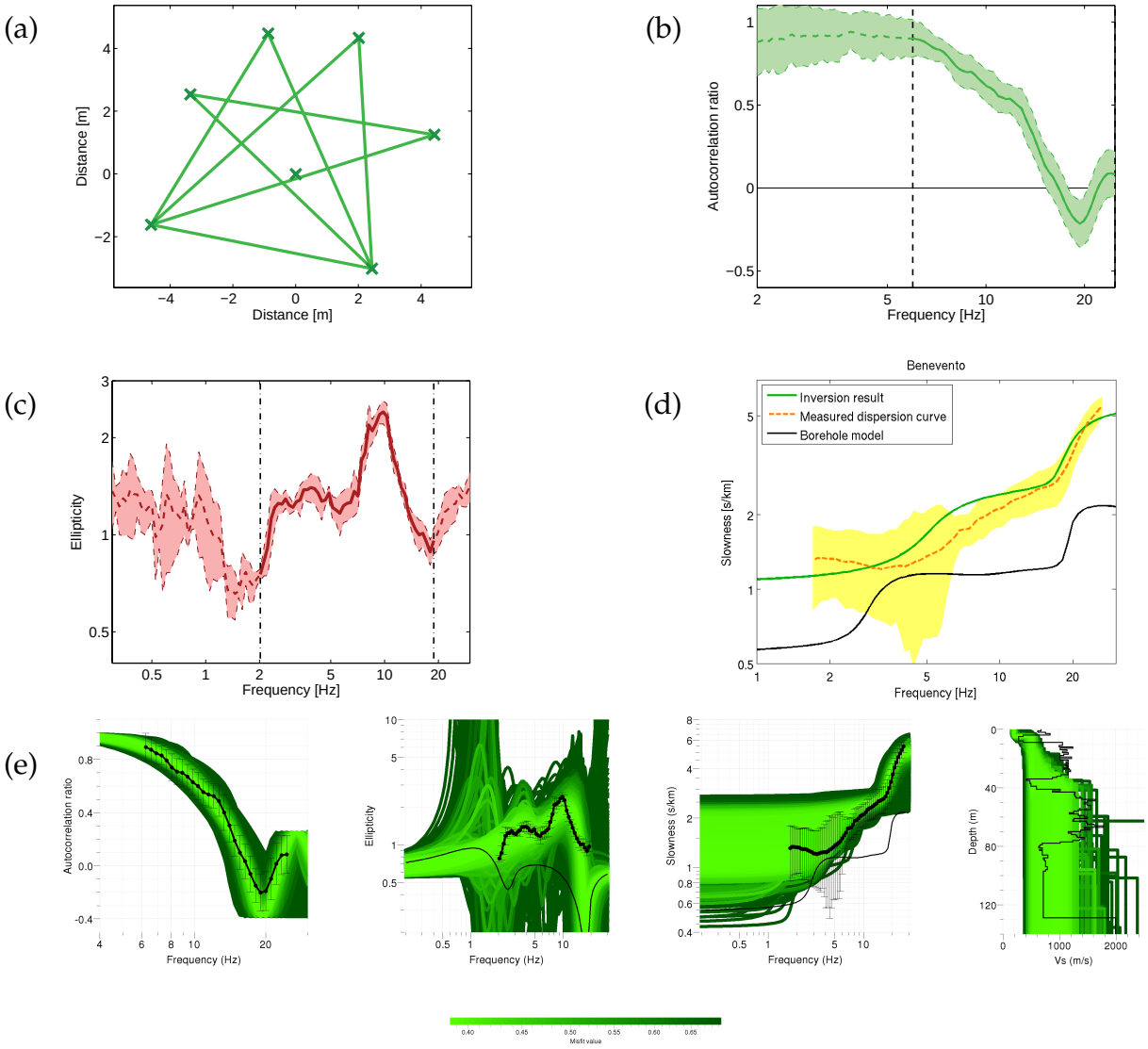


Figure 7.16: (a) Relative positions of the seismic sensors used for the study of Benevento. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 6 stations (8 station pairs, distances between 7.14 and 9.48 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into five parts of 11 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the borehole model's dispersion curve. The incertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Benevento data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center right, with dispersion curve data from array measurements) and shear wave velocity profiles (right).

7.8.2 Bolu

The layout of the array of seismic sensors used for the Bolu measurements is shown in figure 7.17 (a). Six station pairs involving eight seismic sensors were used to calculate the autocorrelation curve of figure 7.17 (b). The ellipticity curves of the different sensors shown in figure 7.3 are in good mutual agreement below 5 Hz. The ellipticity curve shown in figure 7.17 (c) results from an averaging of four ten-minute windows of seismic noise signals of the central station. As the ellipticity curve is rather flat, the bump between 0.4 and 4.0 Hz is used for the inversion.
Both the autocorrelation and the ellipticity curve are well fitted by the inversion (see table 7.8 for the parameters and figure 7.17 (e) for the inversion results). The dispersion curves obtained by the inversion are compatible with the directly measured ones at least between 5 and 20 Hz, as can be seen in figure 7.17 (d). The higher modes for the best inversion model are shown as well. At frequencies below 3 Hz, the inversion suggests that the directly measured dispersion curve corresponds to the first harmonic mode rather than the fundamental one. According to this interpretation, the measured dispersion curve between 3 and 8 Hz would correspond to a mixture of the two lowest modes. The calculation of the P value in table 7.8 is based on both modes, whichever is closest to the data.

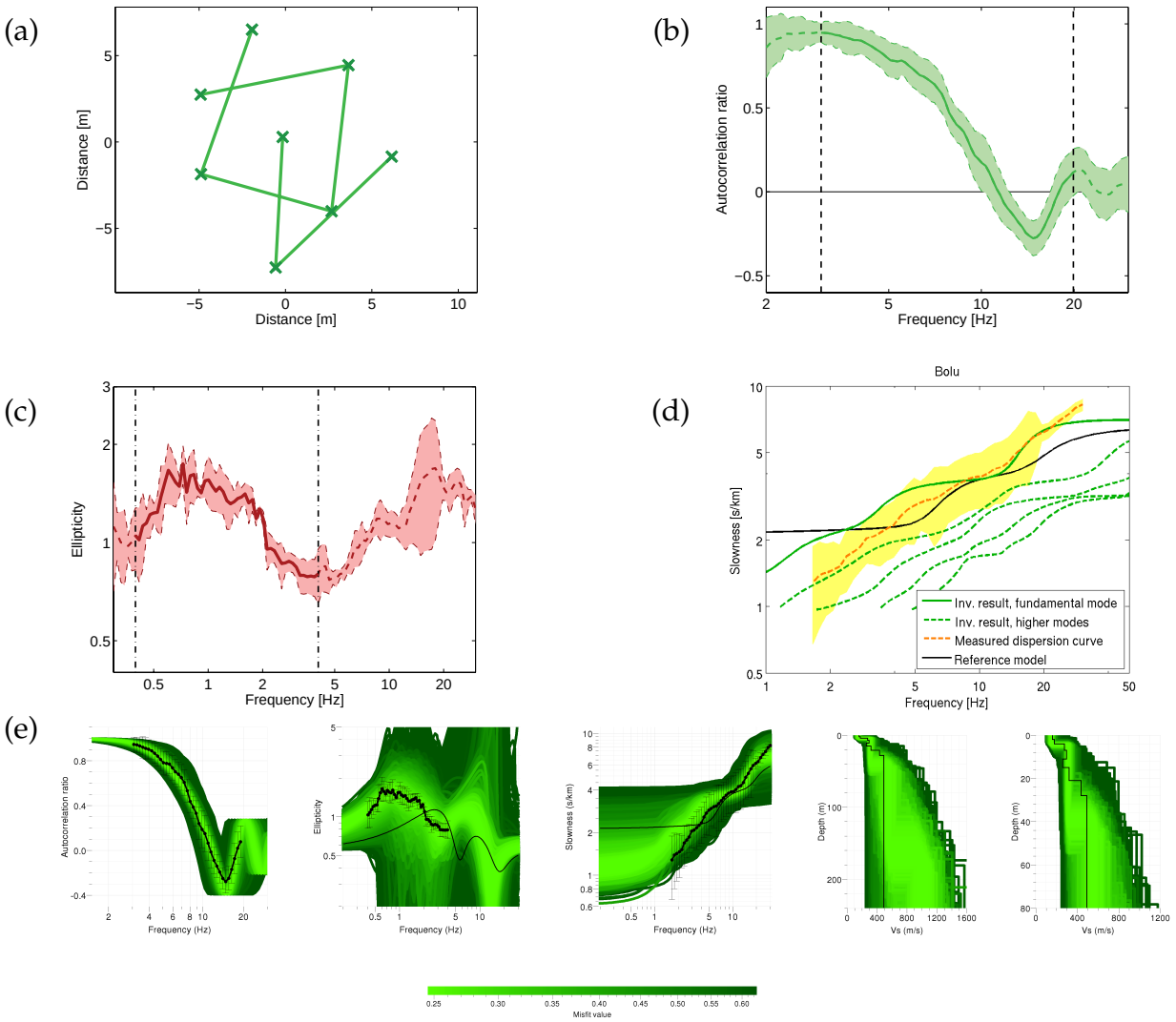


Figure 7.17: (a) Relative positions of the seismic sensors used for the study of Bolu. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 8 stations (6 station pairs, distances between 7.56 and 9.29 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into four parts of 10 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the reference model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Bolu data set. The black dots indicate the data points used for the inversion and the black lines the curves for the reference model. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right).

7.8.3 Buia

For the Buia measurements, the autocorrelation curve (figure 7.18 (b)) was obtained by using the seven station pairs indicated in figure 7.18 (a). The ellipticity curves of all of these stations show an artificial peak at 1.45 Hz (see figure 7.3). The second peak at 6.8 Hz is, despite its shape, not artificial, therefore we identify it as a soil feature and invert the ellipticity curve for the central station between 1.85 and 10.0 Hz (see figure 7.18 (c)).
For the inversion, the parameters are given in table 7.8 and the results in figure 7.18 (e). The autocorrelation and ellipticity data are well fitted. The dispersion curve resulting of the inversion is in good agreement with the direct measurement for frequencies below 20 Hz (see figure 7.18 (d)). Above that frequency, the inversion results are not well constrained as the autocorrelation curve carries information between 6 and 20 Hz only.

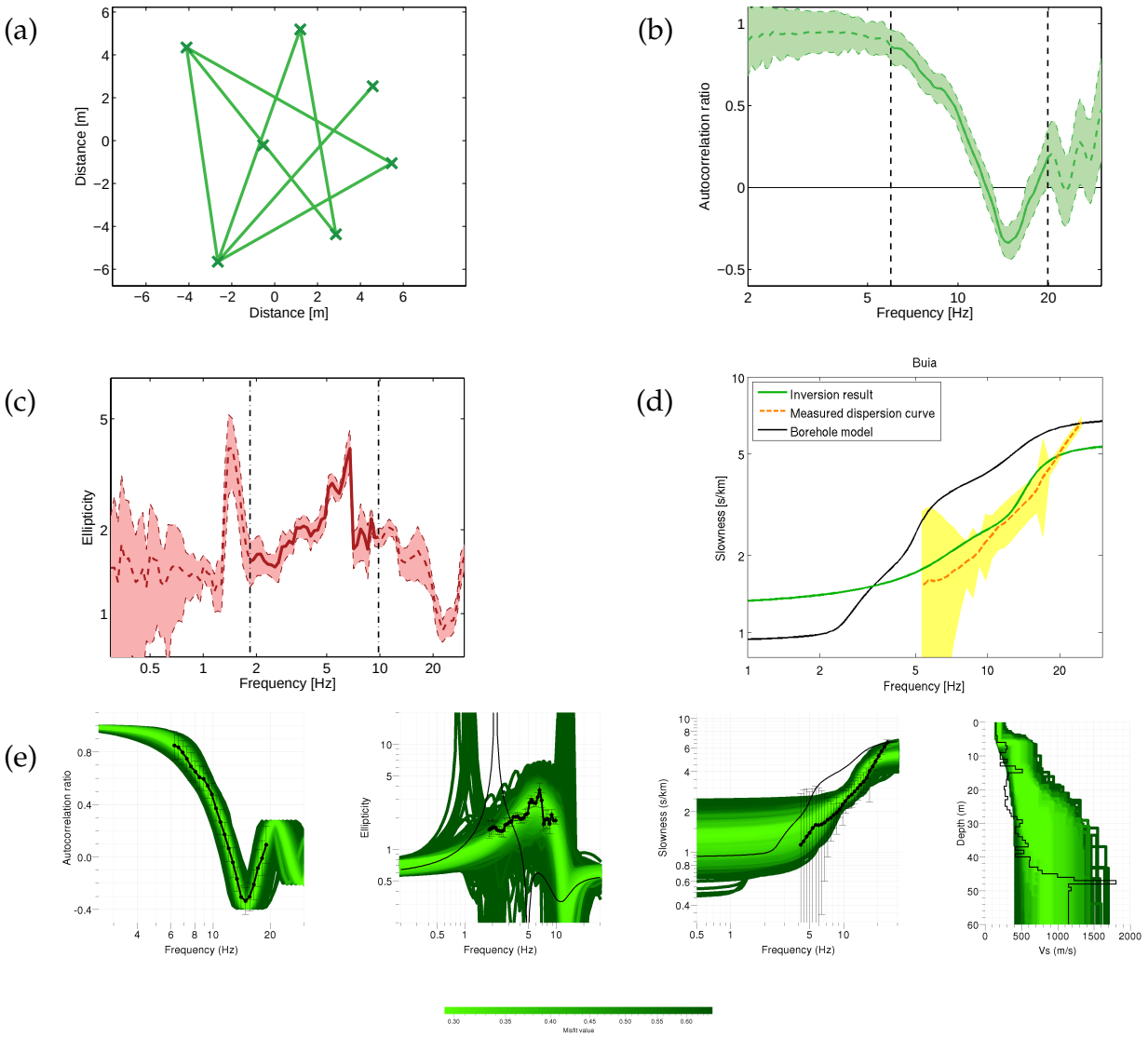


Figure 7.18: (a) Relative positions of the seismic sensors used for the study of Buia. The station pairs contributing to the spatial autocorrelation curves are also shown. (b) Spatial autocorrelation curve using 6 stations (7 station pairs, distances between 9.32 and 11.50 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into six parts of 10 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Buia data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center right, with dispersion curve data from array measurements) and shear wave velocity profiles (right).

7.8.4 Düzce

The array of seismic sensors used for the Düzce measurements is displayed in figure 7.19 (a). The seven ring sensors were used to measure the autocorrelation curve of figure 7.19 (b). The ellipticity curve (figure 7.19 (c)) obtained by analyzing 50 minutes of seismic noise recordings of the central station corresponds to the curves of the ring stations (figure 7.3). The peak at 0.7 Hz could represent a real singularity, but our trials of forcing the inversion to find a singularity at this position did not succeed. Therefore, we took it as a non-singular peak and inverted the ellipticity curve between 0.5 and 2.7 Hz.
The results of the joint inversion of the autocorrelation and the ellipticity curve are shown in figure 7.19 (e). The best model's dispersion curve is in very good agreement with the direct measurement above 8 Hz (see figure 7.19 (d)). At lower frequencies, the first harmonic mode of the inversion's best model is closer to the direct measurement than the fundamental mode. This result suggests that the dispersion curve measurement misinterprets the Rayleigh wave modes. The reference measurement at this site was not obtained by borehole measurements, but by independent MASW measurements. Therefore, the reference profile does not penetrate deeper than about 30 meters. Nevertheless, the inversion results are close to the reference model.

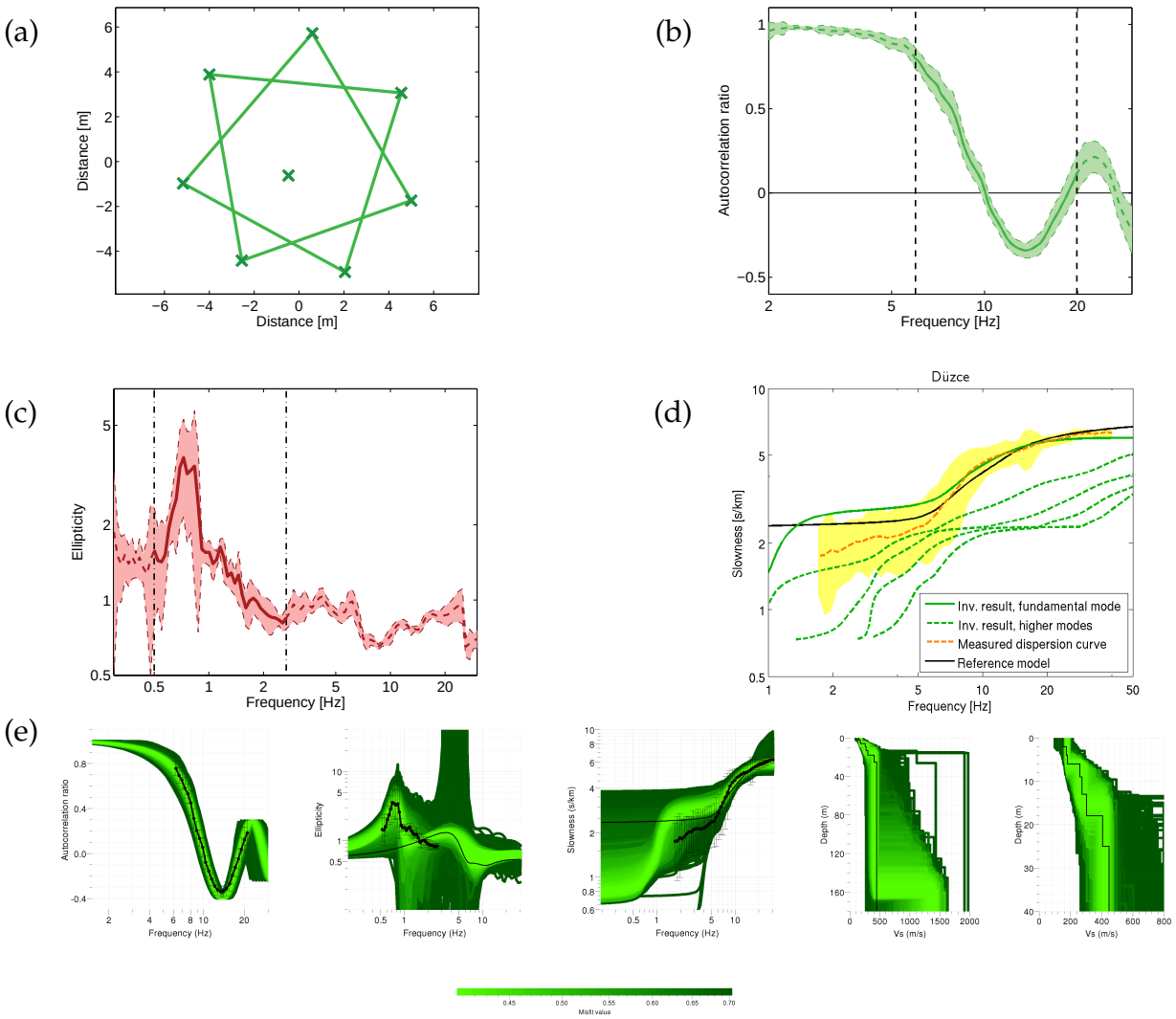


Figure 7.19: (a) Relative positions of the seismic sensors used for the study of Düzce. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 7 stations (7 station pairs, distances between 8.01 and 8.83 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into five parts of 10 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the reference model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (7-station measurement, left) and ellipticity curve (center left) for the Düzce data set. The black dots indicate the data points used for the inversion and the black lines the curves for the reference model. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right).

7.8.5 Forlì

At Forlì, all stations (see figure 7.20 (a) for the locations) exhibit a flat ellipticity curve with values hardly exceeding 1. Nevertheless, the ellipticity curve has a bump around 2 Hz. We invert the ellipticity curve calculated for 45 minutes of seismic noise for the central station of the array between 0.7 and 10.0 Hz (see figure 7.20 (c)) jointly with the autocorrelation curve calculated by using nine station pairs of the seven ring stations (figure 7.20 (b)).
As the autocorrelation curve used for the inversion is defined in the same frequency region (3 - 20 Hz) as the ellipticity curve, one could argue that the both informations are redundant. Therefore, we performed a second inversion of the autocorrelation curve only. The parameters of both inversions are given in table 7.8 and the respective results in figure 7.20 (e) and (f).
It can be seen that the inclusion of the ellipticity information improves the inversion result by further constraining the possible solutions and acting particularly on the deeper layers. In figure 7.20 (d), the dispersion curves of the best models of both inversions are compared to the measured dispersion curve. Above 5 Hz, both inversions' results are identical. At lower frequencies, the inversion including the ellipticity data is very close to the measured dispersion curve, whereas the autocorrelation inversion deviates considerably. This effect is reflected in the P values as well. At frequencies above 15 Hz, both inversions are in contradiction with the dispersion curve measurement, certainly because the autocorrelation curve does not sufficiently constrain the very superficial layers. Therefore, the P values of both inversions do not go below 1.

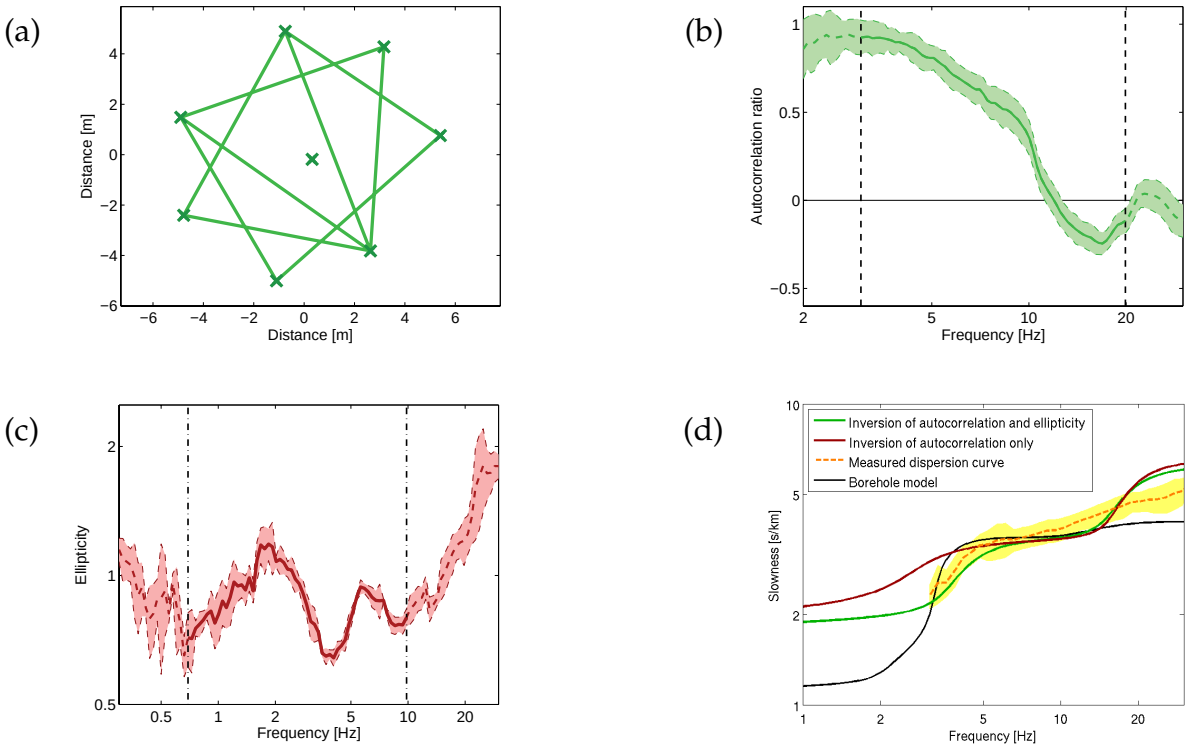


Figure 7.20: (a) Relative positions of the seismic sensors used for the study of Forlì. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 7 stations (9 station pairs, distances between 7.41 and 9.35 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into four parts of 11.25 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area.

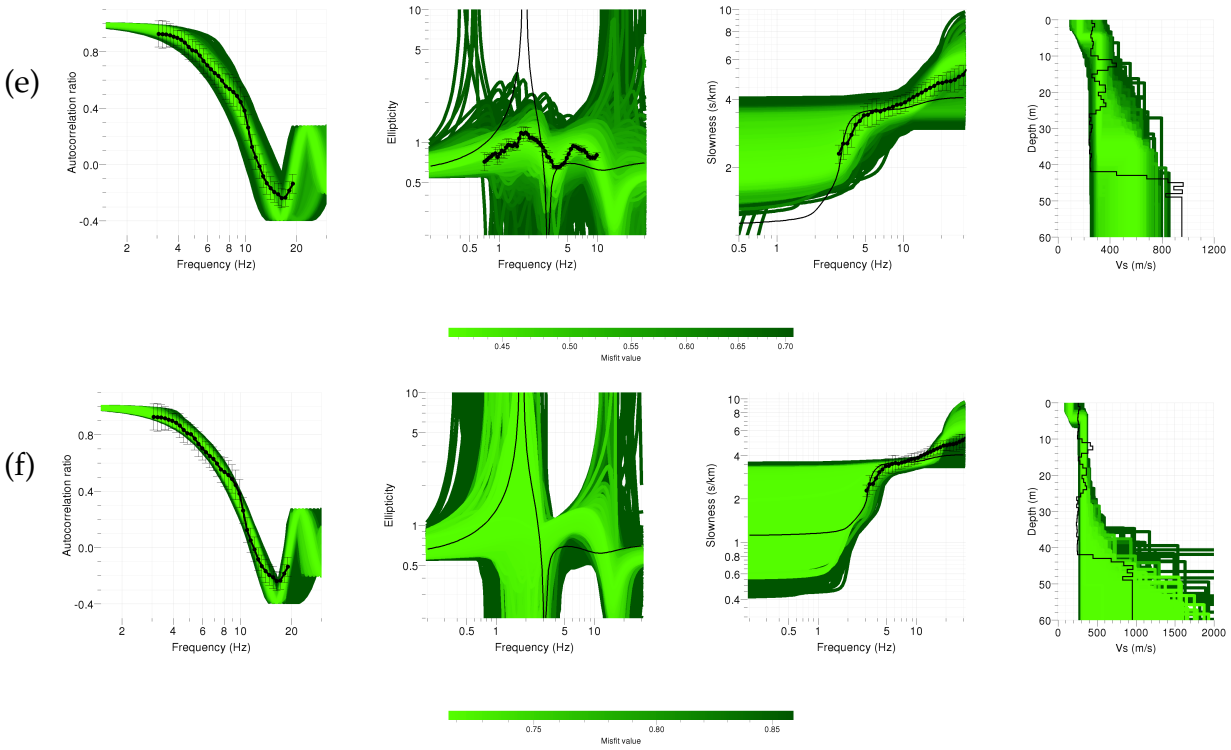


Figure 7.20: – continued. Inversion results for (e) a joint inversion of the autocorrelation curve (7-station measurement, left) and the ellipticity curve (center left) and (f) an inversion of the autocorrelation curve only for the Forlì data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curves (center right, with dispersion curve data from array measurements) and shear wave velocity profiles (right).

7.8.6 Knidi

At Knidi, the smallest ring of seismic sensors was located about 40 meters east of a downhole measurement serving as reference. A station of a larger ring was located closer to the borehole, about 15 meters to the west. Figure 7.21 (a) shows the relative positions of the small ring of seismic sensors. We used seven station pairs and three station pairs for the respective SPAC measurements shown in figure 7.21 (b). The ellipticity curves of the ring stations in figure 7.3 are quite inhomogeneous, although the stations are located about 10 meters away from each other. The ring stations' ellipticity does not have a clear peak and does not exceed 2 in amplitude. The station located near the borehole, however, exhibits a clear peak at 5.7 Hz.

In a first step, we inverted the autocorrelation curve calculated by using seven station pairs of the small array and the ellipticity curve of the central station of this array between 5.0 and 11.6 Hz (figure 7.21 (c)). The results are shown in figure 7.21 (e). As the ellipticity curve does not have a clear peak, the inversion does not yield a strong velocity contrast. Although the autocorrelation and ellipticity curves are well fitted, the resulting dispersion curves are neither in agreement with the borehole model nor with the dispersion curve measured by using four arrays of different sizes. Consequently, the velocity model of the borehole is only retrieved at the surface.

These results led to a second inversion, using the same autocorrelation data, but the ellipticity curve for the station near the borehole. In contrast to the other curves shown in figure 7.3, this ellipticity curve exhibits a clear peak around 5.7 Hz, and a stronger velocity contrast can be expected. As the inversion of the right flank of the peak fixed the peak frequency well enough, the left flank of the peak was not included in the inversion process which yields the results shown in figure 7.21 (f). The autocorrelation and ellipticity curves are well fitted, but the shear wave velocity profile is different from the borehole model. The dispersion curves of the inversion and the borehole model differ over the whole frequency range, but the inversion result is in good agreement with the directly measured dispersion curve for frequencies below 30 Hz, as can also be seen in figure 7.21 (d).

An inversion of the same ellipticity information, but using the 3-station autocorrelation curve of figure 7.21 (b) leads to the results shown in figure 7.21 (g). The results are qualitatively the same as in the previous case, but the corresponding dispersion curve deviates considerably at frequencies above 20 Hz. However, the concordance with the directly measured dispersion curve is still given for frequencies up to 30 Hz. When disregarding frequencies above 30 Hz, both the 7-station and the 3-station inversion fit the directly measured dispersion curve equally well.

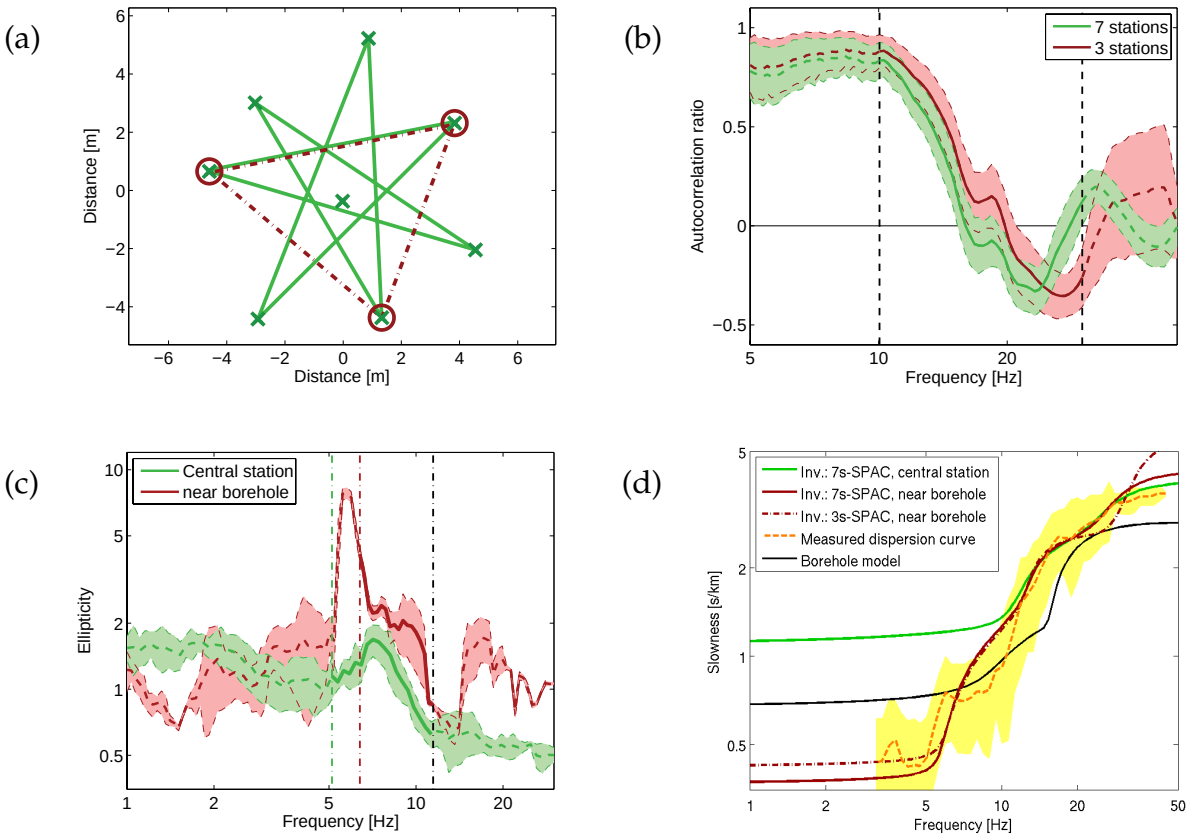


Figure 7.21: (a) Relative positions of the seismic sensors used for the study of Knidi. The station pairs contributing to the spatial autocorrelation curves for the 7-station and the 3-station case are also shown. The location of the borehole site 50 meters west of the array and of the station near the borehole (15 meters from the borehole) are not shown. (b) Spatial autocorrelation curves using 7 stations (7 station pairs, distances between 8.56 and 10.37 m) and 3 stations (3 station pairs, distances between 7.13 and 8.59 m). The shaded areas indicate the measurement errors. The bold parts of the curves, delimited by the dashed lines, were used for the inversion. The different distances of both measurements explain the shift between both curves. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the station near the borehole into four parts of 7.5 minutes each, applying RayDec and averaging the results. The bold parts of the curves, delimited by the dashed lines, were used for the inversions. (d) Dispersion curves for Knidi: Best models of the inversion using the 7-station and 3-station autocorrelation curves compared to the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area.

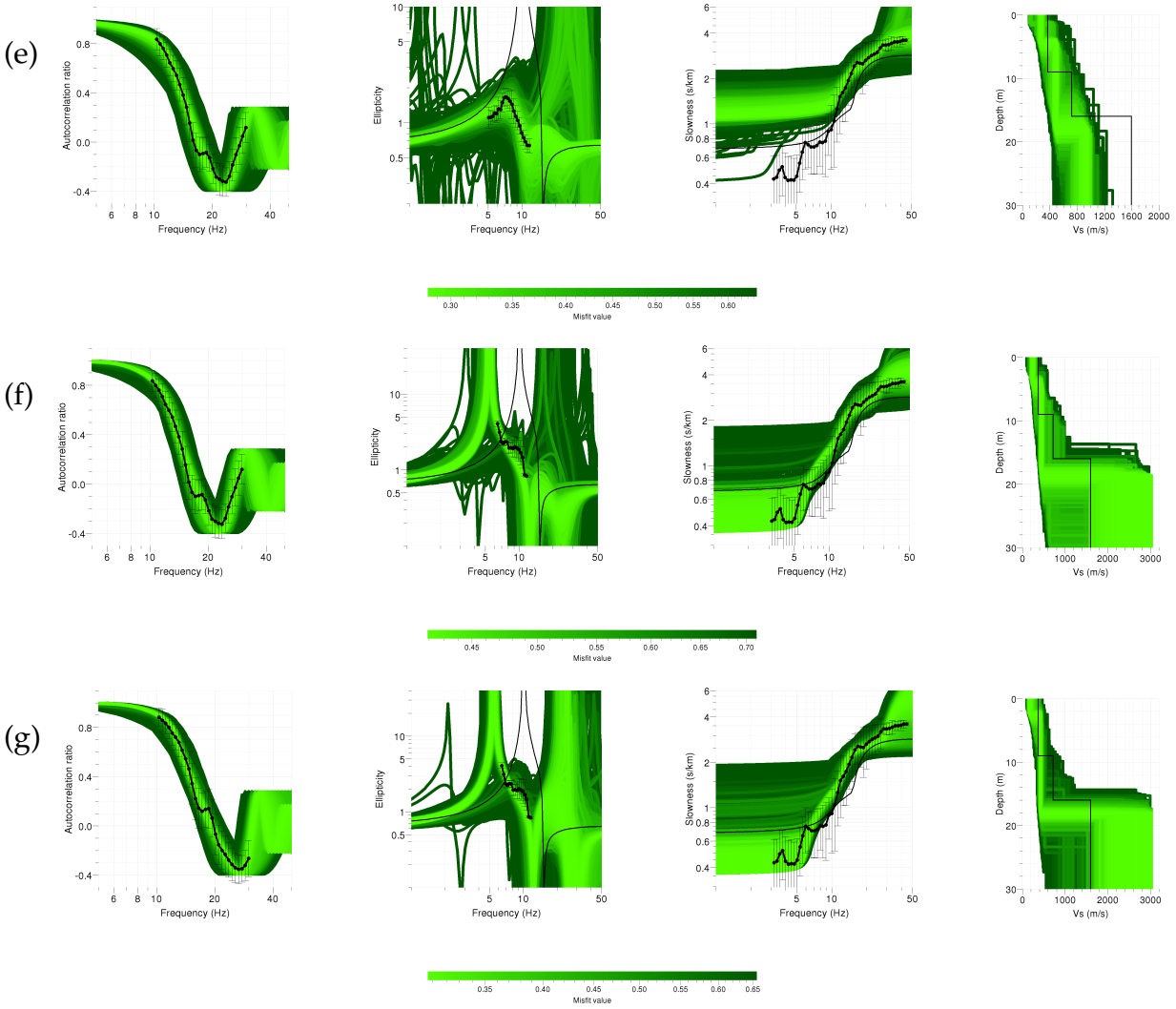


Figure 7.21: – continued. Inversion results for joint inversions of the autocorrelation curve (left) and ellipticity curve (center left) for the Knidi data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center right, with dispersion curve data from array measurements) and shear wave velocity profiles (right): (e) using the 7-station SPAC measurement and the ellipticity curve for the central station of the array (f) using the 7-station SPAC measurement and the ellipticity curve for the station near the borehole (g) using the 3-station SPAC measurement and the station near the borehole.

7.8.7 Norcia

The positions of the sensors of the seismic array used for the study of Norcia are given in figure 7.22 (a). The autocorrelation curve (figure 7.22 (b)) results from analyzing eight station pairs using the ring stations. The different stations exhibit major differences in the ellipticity curves below 1 Hz (figure 7.3). Around the peak frequency of 8 Hz, all stations exhibit virtually identical curves. As no inversion with an ellipticity singularity at that frequency could be found, we took the peak as a non-singular one. The ellipticity curve in figure 7.22 (c) was obtained by analyzing 45 minutes of seismic noise recorded by the central station. A joint inversion of the autocorrelation and the ellipticity curves yields the results shown in figure 7.22 (e). Although the ellipticity curve is not too well fitted, the resulting dispersion curve (figure 7.22 (d)) is very close to the measured one, at least below 30 Hz.

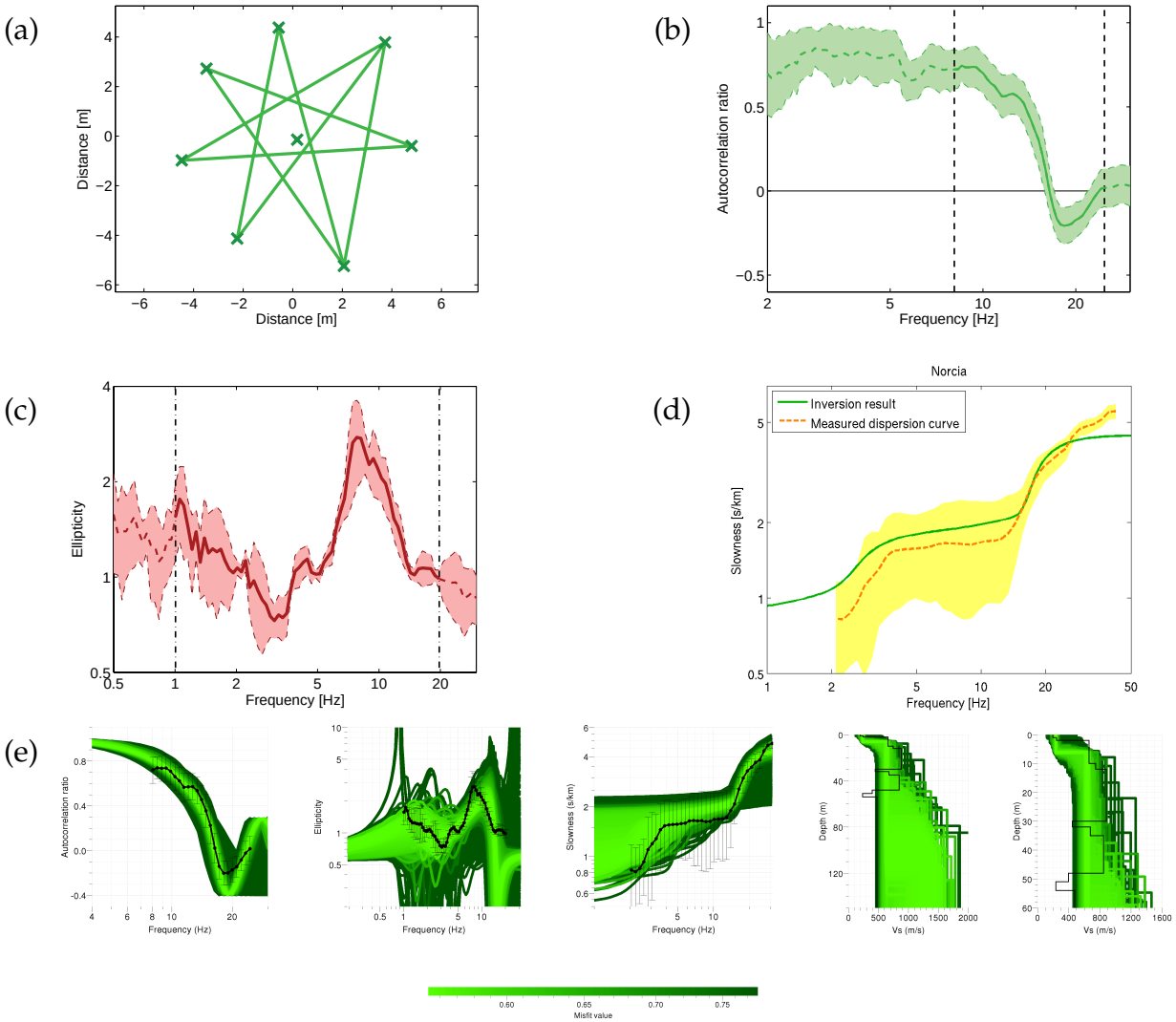


Figure 7.22: (a) Relative positions of the seismic sensors used for the study of Norcia. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 7 stations (8 station pairs, distances between 8.67 and 9.97 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into four parts of 8.75 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (7-station measurement, left) and ellipticity curve (center left) for the Norcia data set. The black dots indicate the data points used for the inversion. Dispersion curve (center, with dispersion curve data from array measurements) and shear wave velocity profiles (center right, zoom on the shallow layers on the right). The low velocity zones of the borehole model (black line in the figures on the right) did not allow the calculation of dispersion or ellipticity curves.

7.8.8 Sakarya

The location of the seismic sensors, the resulting autocorrelation curve using seven pairs of these sensors and the ellipticity curve of the central station for 50 minutes of measurements are shown in figure 7.23 (a) - (c). As it is unclear if the ellipticity peak is singular or not, we only constrain the left and right flanks of it. The inversion (see table 7.8 for parameters) yields the results of figure 7.23 (e). A model with ellipticity singularity explains the data in a better way than other models. When disregarding frequencies above 25 Hz, the best model's dispersion curve is in good agreement with the directly measured one (figure 7.23 (d)).

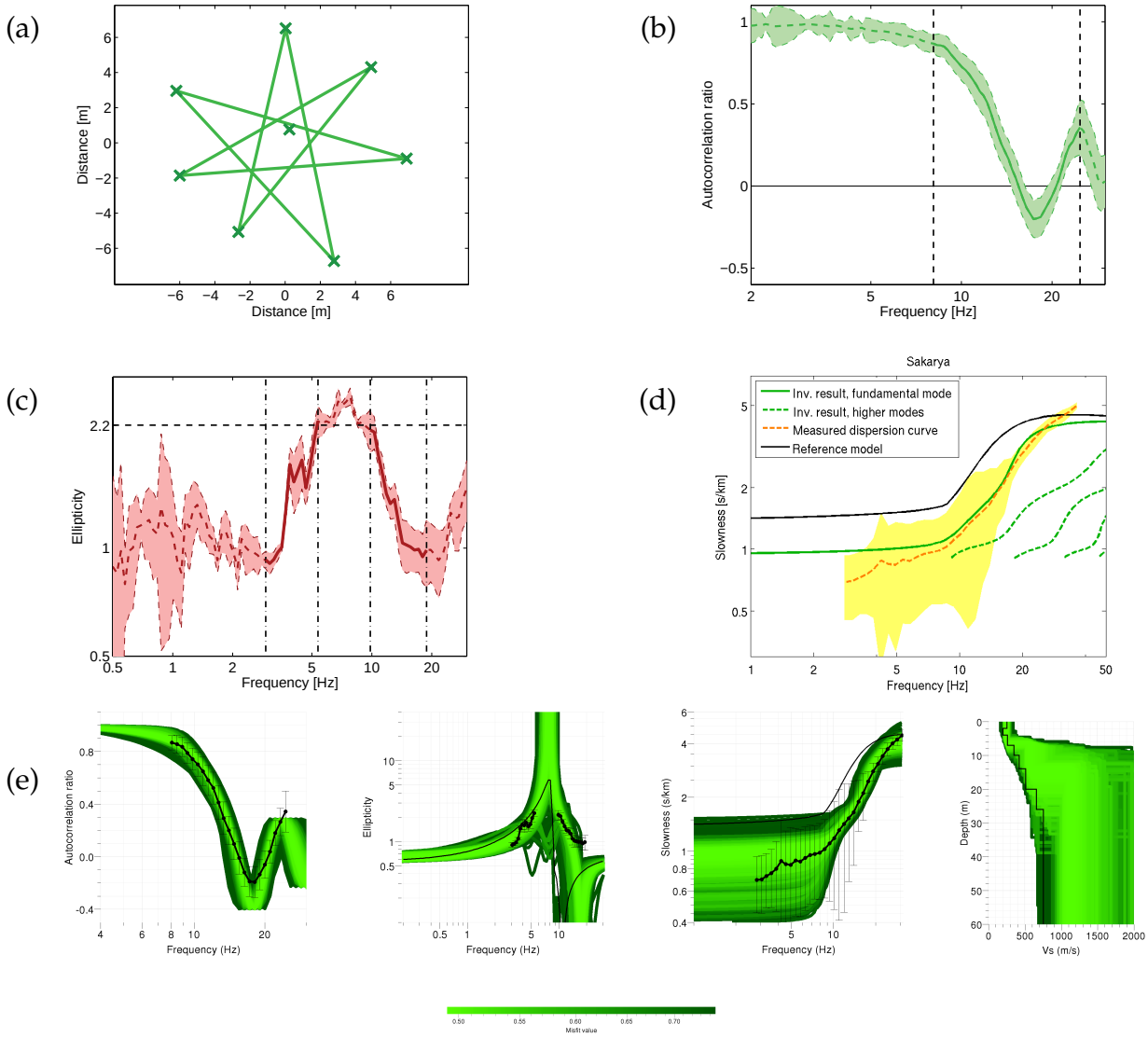


Figure 7.23: (a) Relative positions of the seismic sensors used for the study of Sakarya. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 7 stations (7 station pairs, distances between 11.89 and 13.67 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into five parts of 10 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the reference model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (7-station measurement, left) and ellipticity curve (center left) for the Sakarya data set. The black dots indicate the data points used for the inversion and the black lines the curves for the reference model. Dispersion curve (center right, with dispersion curve data from array measurements) and shear wave velocity profiles (right).

7.8.9 Sturno

The ellipticity curves of Sturno (see figure 7.3) are quite heterogeneous at low frequencies and exhibit a not very distinctive industrial peak at 1.45 Hz. Above that frequency, the ellipticity curves are rather flat. The array of seismic sensors shown in figure 7.24 (a) yields the autocorrelation curve of figure 7.24 (b). Taking the ellipticity curves for all stations into account, we decided to invert the bump of the ellipticity curve of the central station above the industrial peak frequency (figure 7.24 (c)) which is calculated for 40 minutes of data.
Using the parameters of table 7.8, the joint inversion yields the results shown in figure 7.24 (e). Both inversion targets are well fitted, the dispersion curve of the best model is in very good agreement with the direct measurement (see figure 7.24 (d)), except for frequencies above 40 Hz.

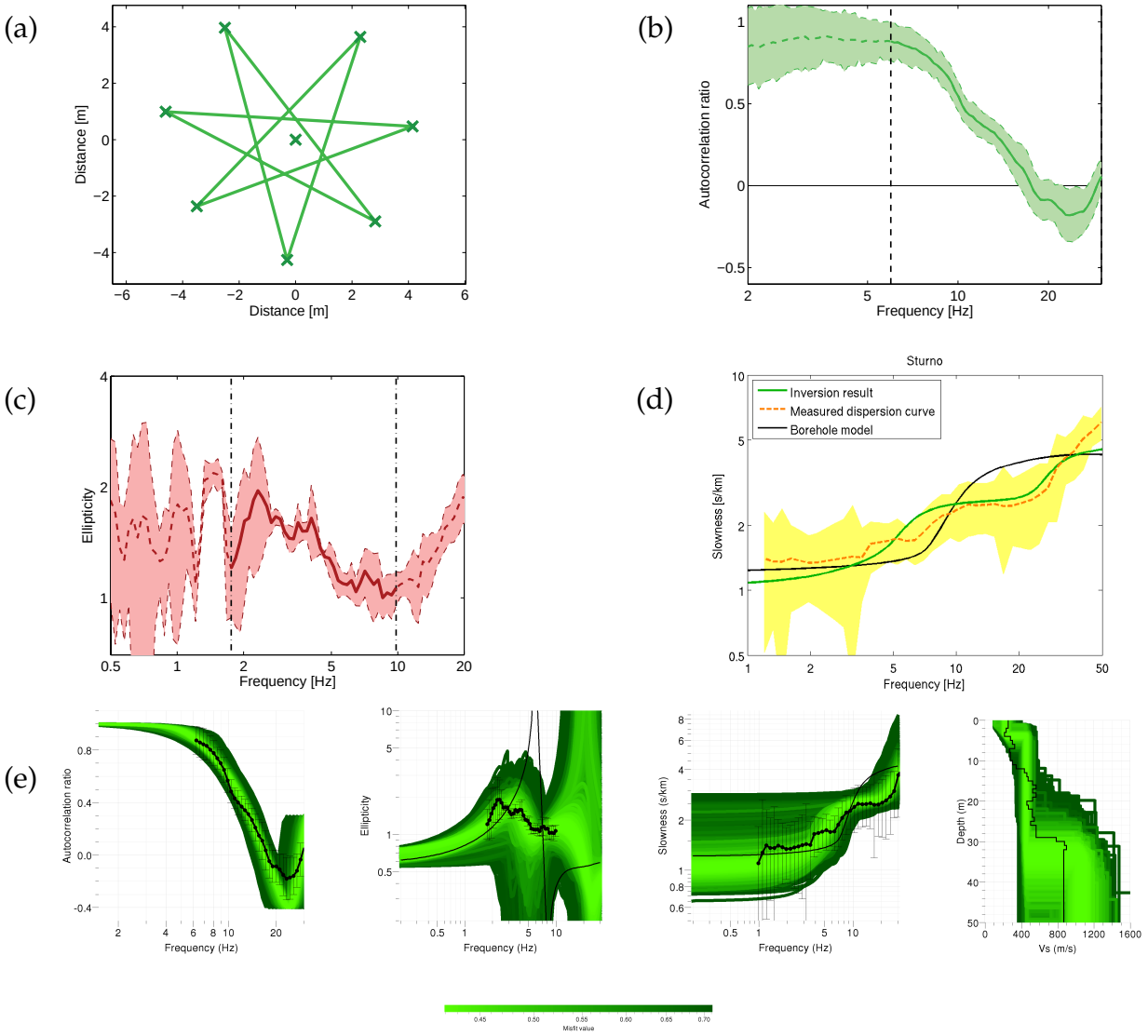


Figure 7.24: (a) Relative positions of the seismic sensors used for the study of Sturno. The station pairs contributing to the spatial autocorrelation curve are also shown. (b) Spatial autocorrelation curve using 7 stations (7 station pairs, distances between 8.15 and 8.77 m). The shaded area indicates the measurement errors. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (c) Mean ellipticity and standard deviation obtained by cutting the signal of the central station into four parts of 10 minutes each, applying RayDec and averaging the results. The bold part of the curve, delimited by the dashed lines, was used for the inversion. (d) Dispersion curves: Best models of the inversion compared to the directly measured dispersion curve and the borehole model's dispersion curve. The uncertainty region of the direct dispersion curve measurement is indicated by the shaded area. (e) Inversion results for joint inversions of the autocorrelation curve (7-station measurement, left) and ellipticity curve (center left) for the Sturno data set. The black dots indicate the data points used for the inversion and the black lines the curves for the borehole model. Dispersion curve (center right, with dispersion curve data from array measurements) and shear wave velocity profiles (right).

Chapter 8

Investigation of the seismological wave field: Earthquake array measurements in California

This chapter investigates the seismological wave field generated by earthquakes. A data set of 22 earthquakes recorded by the San Jose Dense Seismic Array in California is analyzed using MUSIQUE. The array is located on the Evergreen Basin, near the eastern edge of the Santa Clara Valley. The azimuthal energy distribution of the different events is compared, including for some events a comparison between the early arrivals and the later signal. Furthermore, the energy repartition between Love and Rayleigh waves is analyzed in different frequency ranges. Finally, the analysis results for the different events are combined to global dispersion and ellipticity curves of the basin.

Ce chapitre examinera le champ d'ondes sismique produit par des séismes. Un jeu de données de 22 séismes enregistrés par un réseau de capteurs sismiques localisé à San Jose en Californie est analysé avec MUSIQUE. Le réseau a été deployé sur le bassin d'Evergreen, proche de la frontière est de la vallée de Santa Clara. La distribution d'énergie suivant l'azimut sera comparé pour les différents évènements, y inclus une comparaison entre les premières arrivées et les arrivées tardives pour quelques séismes. En outre, la répartition d'énergie entre les ondes de Love et les ondes de Rayleigh sera analysée pour différentes gammes de fréquence. Finalement, les analyses des différents évènements seront combinées pour déterminer des courbes de dispersion et d'ellipticité globales pour le bassin.

8.1 Introduction

The topic of this chapter is the investigation of the seismological wave field generated by earthquakes. This analysis is performed using the MUSIQUE technique introduced in 4.6. In the following, this technique will be applied to real data measurements obtained by The San Jose Dense Seismic Array in California, a temporary array recording between 1999 and 2004. The array is located near the eastern edge of the Santa Clara Valley, where deep sedimentary layers form the Evergreen Basin. The south-western edge of the Santa Clara Valley is formed by the Cupertino basin, which is separated from the Evergreen Basin by a subsurface ridge. The geology of the Evergreen Basin is shown in figure 8.1.

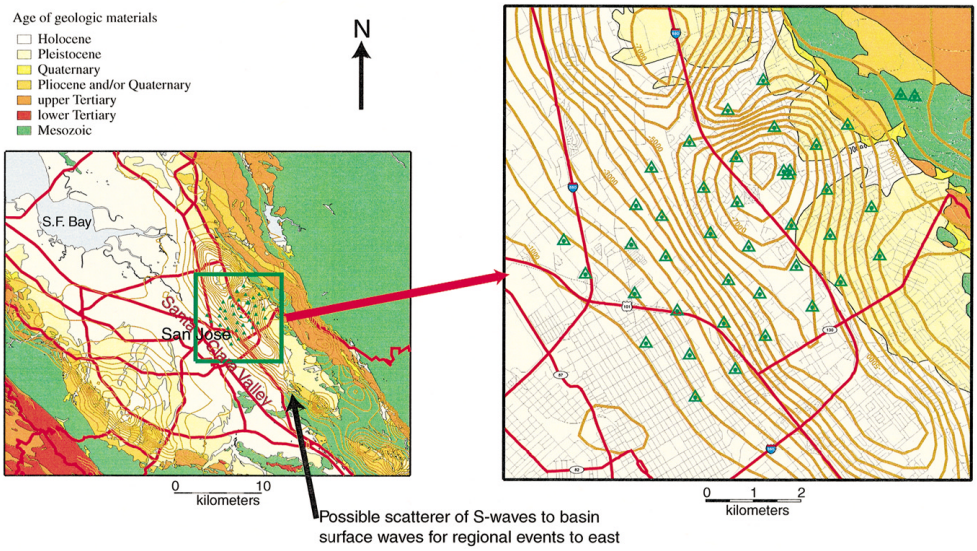


Figure 8.1: (Left) Map showing the surficial geology [by Wentworth, 1997] of the Santa Clara Valley and the location of the San Jose Dense Seismic Array in the Evergreen Basin inside the valley. In the zoom on the array (right), the array stations and the depth contour layers [by Brocher et al., 1997] of the Basin are shown in 500-m-intervals. The basin reaches a maximum depth of about 8 km. Figure extracted from Frankel et al. [2001].

A previous study using data of the San Jose Dense Seismic Array was published by Frankel et al. [2001]. They found that the amplification of seismic waves is stronger towards the south-western edge of the Evergreen Basin. Furthermore, they observed patterns of waves traveling across the array. For earthquakes located at the east, surface waves from the south follow the direct S-wave arrivals. Frankel et al. [2001] identified these surface waves as Love waves which are probably generated by S-wave scattering at the southern border of the Santa Clara Valley.
Another study using the San Jose array was performed by Hartzell et al. [2003] who investigated the site response for local and regional earthquakes and identified possible wave scatterers in the valley. Hartzell et al. [2006] developed a detailed 3D velocity model of the Santa Clara Valley and showed the importance of surface waves for the ground motion of sedimentary basins.

In a further study by Hartzell et al. [2010], it was shown that the maximum of amplification for the horizontal motion occurs at the southern edge of the basin, whereas it occurs at the deepest parts for the vertical component. The authors suggest that this can either be attributed to a different interaction of Love and Rayleigh waves with the basin shape or that Love waves are more influenced by the lower shear wave velocity region at the south-western edge of the basin.
Other studies investigated the same area without using the San Jose Dense Seismic Array. Fletcher et al. [2003] analyzed the data of a different array covering the whole Santa Clara Valley. They investigated the P- and S-wave arrival times from both local deep events and teleseismic events and found that the arrivals are delayed for stations located over the Evergreen or Cupertino Basins, but arrive earlier for stations located on the ridge between both basins. Furthermore, they determined that the attenuation is larger at the basins than on the ridge. Dolenc et al. [2005] used the same array recordings and showed that the P-wave arrival time delays and amplification parameters for teleseismic events are correlated. Another study was done by Dolenc and Dreger [2005], which was based on the same array but for microtremor records. This study indicates that the energy of the seismic noise recorded on the array is related to the wave height on the ocean and that the H/V ratio for the different sensors is stable with time.
The objective of our study is to perform a systematic analysis of the propagation of surface waves in the basin, not only for a single earthquake, but for a set of earthquakes in order to study the correlation between surface wave types (Love, Rayleigh) and earthquake characteristics (location, magnitude). Therefore, 22 regional earthquakes in distances between 4 and 600 km are selected and the repartition and azimuth angles of Love and Rayleigh waves are investigated in different frequency ranges. Furthermore, we combine the data for the different events to obtain Love and Rayleigh wave dispersion curves as well as Rayleigh wave ellipticity curves.

8.2 Array setup and basin model

In early 1999, the United States Geological Survey (USGS) deployed an array of Kinemetrics K2 accelerometers, which recorded until May 2004. The different instruments were operated in triggered mode, i.e. did not record continuously, but only if the acceleration exceeds a predefined threshold. The array layout is shown in figure 8.2 (a) and the array response corresponding to the selected stations is shown in figure 8.2 (b). The P- and S-wave velocities along the profile marked in figure 8.2 (a) are indicated in figures 8.3 (a) and (b), respectively. Along this profile, the maximum depth of the sedimentary filling of the basin reaches up to 5 km. The velocity profiles were obtained using the USGS seismic velocity model of the San Francisco Bay Area[I], which is based on the works of Brocher [2005] and Phelps et al. [2008]. Based on this model, the theoretical Rayleigh and Love wave dispersion curves for the structure beneath the different array stations have been calculated.

[I] Version USGSBayAreaVM-08.3.0 of the model was used.
This model is available at http://earthquake.usgs.gov/regional/nca/3Dgeologic/.

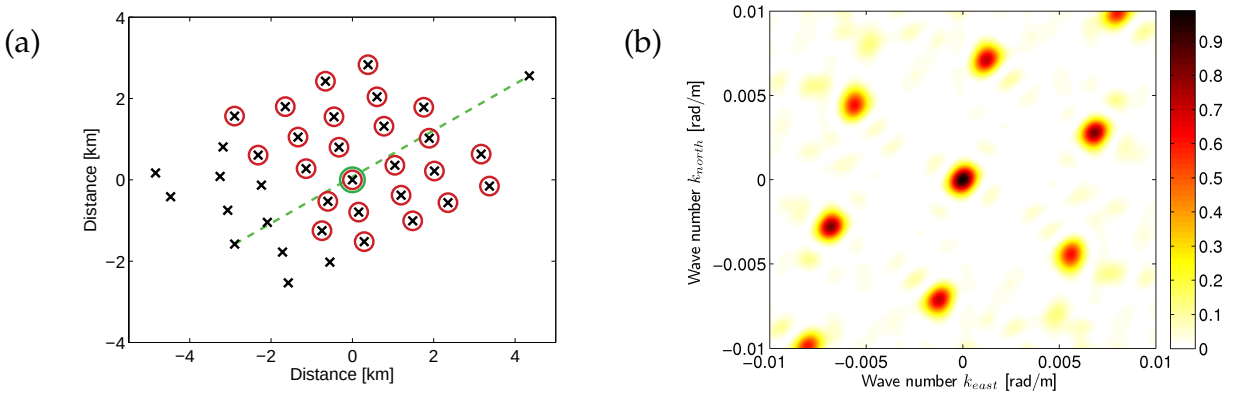


Figure 8.2: (a) Layout of the San Jose array. Only the marked stations have been used for this study. The green line indicates the profile for which the shear and pressure wave velocities are indicated in figure 8.3. The dispersion and ellipticity curves for the station which is marked by the large green circle are shown in figure 8.3. (b) Theoretical array response for the array of selected stations.

The south-western stations of the array are located on the border of the Evergreen Basin. Therefore, their dispersion curves are different from those of the central stations. Consequently, we used only the 25 seismic stations on the basin center for this study, which are indicated by the red circles in figure 8.2 (a). As the structure of the basin is rather homogeneous under these stations, their respective dispersion and ellipticity curves are very similar as shown in figures 8.3 (c) and (d). The outliers are located on the edges of the used array configuration. To give an example, the dispersion curves for the station marked by a green circle in figure 8.2 (a) are shown in figure 8.3 (e) and the ellipticity curve is shown in figure 8.3 (f). As the velocity contrast of the model is not very pronounced, the group slownesses of both surface wave types do not show a distinct Airy phase. Furthermore, the theoretical ellipticity curve for the fundamental mode does not exhibit a singularity. Consequently, the motion of this mode should be retrograde at all frequencies. As the motion of the first harmonic mode is prograde, both modes should be distinguishable by their particle motion.

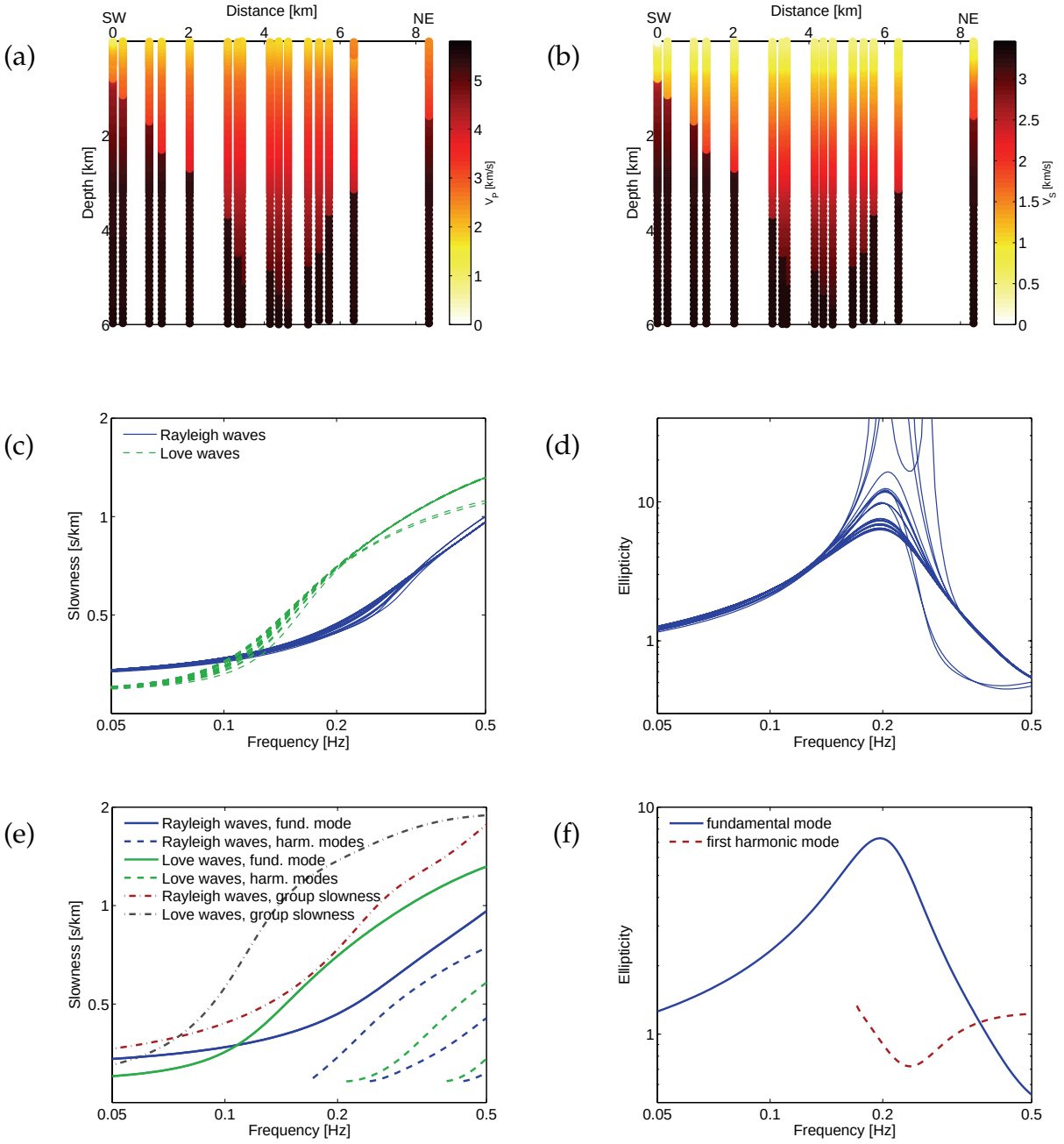


Figure 8.3: (a) Pressure wave and (b) shear wave velocity profiles according to the basin model along the cross section indicated by the dashed line in figure 8.2 (a). (c) Dispersion curves for the fundamental Rayleigh and Love wave modes calculated for the basin model at all used stations. (d) Fundamental Rayleigh wave mode ellipticity curve calculated at all used stations. (e) Dispersion curve for phase and group slownesses of Rayleigh and Love waves calculated for the basin model at the central station marked in figure 8.2 (a). (f) Rayleigh wave ellipticity curve calculated for the basin model at the same station.

8.3 Database

The San Jose array recorded a total of 199 earthquakes between May 1999 and April 2006. However, some stations of the array were subsequently moved to locations at other places in Santa Clara Valley, which reduces the number of recording stations for earthquakes occuring in the late phase of the deployment. As the different seismic sensors were triggered, the number of stations actually recording is dependent on the magnitude and distance of the earthquake. In order to have a good signal-to-noise ratio and as much seismic sensors as possible for the array analysis, we selected the 22 earthquakes whose properties are given in table 8.1 for our analysis. An example accelerogram is shown in figure 8.5 (a) for event 5, the Hector Mine earthquake, which is the strongest event in the database and for which the recording length is the longest. The corresponding energy as a function of time is shown in figure 8.5 (b). This energy is calculated by squaring the relative acceleration values. After the initial signal, the energy decreases monotonically without reaching a constant value. However, the energy in the late signal after 900 s is at least a hundred times smaller than at the strongest signal peak.

Table 8.1: List of analyzed earthquakes. For each earthquake, the origin date and time, location, depth and magnitude are indicated. Furthermore, the distances between the epicenter and the center of the array of seismic sensors are given with the theoretical azimuth under which the array should detect the seismic waves. Finally, the number of seismic stations contributing to the analysis of the given earthquake is indicated as well as the recorded signal length for which the indicated number of stations recorded simultaneously and the peak ground acceleration on the horizontal components.

Event ID	Date	Origin time (UTC)	Latitude $[°N]$	Longitude $[°W]$	Depth $[km]$	Magnitude		Distance to array $[km]$	Incident azimuth	Number of stations	Signal length [s]	Peak ground acceleration $[cm/s^2]$
1	05/15/1999	13:22	37.53	118.82	5.6	5.6	Mw	269	85°	23	436	0.79
2	05/15/1999	17:54	37.51	118.83	7.4	4.7	Mw	268	86°	23	136	0.28
3	08/01/1999	16:06	37.40	117.12	1.3	4.6	Md	420	88°	24	540	1.00
4	08/18/1999	01:06	37.91	122.69	6.8	5.0	ML	94	310°	19	203	2.07
5	10/16/1999	09:46	34.59	116.27	0.0	7.1	Mw	590	120°	23	1317	1.41
6	09/03/2000	08:36	38.38	122.41	10.1	5.2	ML	122	337°	21	305	2.66
7	09/26/2000	07:20	38.66	119.53	9.2	4.2	ML	250	54°	22	104	1.04
8	12/02/2000	15:34	39.38	120.45	14.3	4.9	ML	255	28°	12	107	0.08
9	08/10/2001	20:19	39.81	120.62	4.0	5.5	ML	292	21°	18	216	0.41
10	12/28/2001	21:14	36.64	121.25	6.8	4.7	ML	98	146°	19	191	0.42
11	05/14/2002	05:00	36.97	121.60	7.2	4.9	ML	51	153°	19	106	12.86
12	09/25/2002	07:08	36.59	121.20	8.0	3.9	ML	105	146°	13	76	0.16
13	11/24/2002	14:54	37.76	121.95	11.0	3.9	ML	44	350°	15	98	0.35
14	11/25/2002	03:57	37.76	121.94	10.0	3.8	ML	44	350°	15	74	0.26
15	11/25/2002	18:22	37.76	121.94	11.0	3.5	ML	44	351°	12	67	0.15
16	11/26/2002	12:38	37.76	121.95	10.6	3.6	ML	44	350°	15	72	0.16
17	01/07/2003	22:29	36.81	121.39	8.9	4.3	Mw	76	146°	15	134	1.00
18	02/02/2003	16:22	37.75	121.95	16.8	3.6	ML	42	350°	15	70	0.61
19	02/02/2003	18:22	37.74	121.94	16.7	4.2	ML	42	350°	15	103	0.52
20	02/02/2003	18:47	37.75	121.95	16.9	4.0	ML	42	350°	15	103	0.25
21	02/02/2003	19:02	37.75	121.95	17.1	3.5	ML	43	350°	16	56	0.57
22	05/25/2003	07:09	38.46	122.70	5.4	4.3	ML	141	329°	12	145	0.24

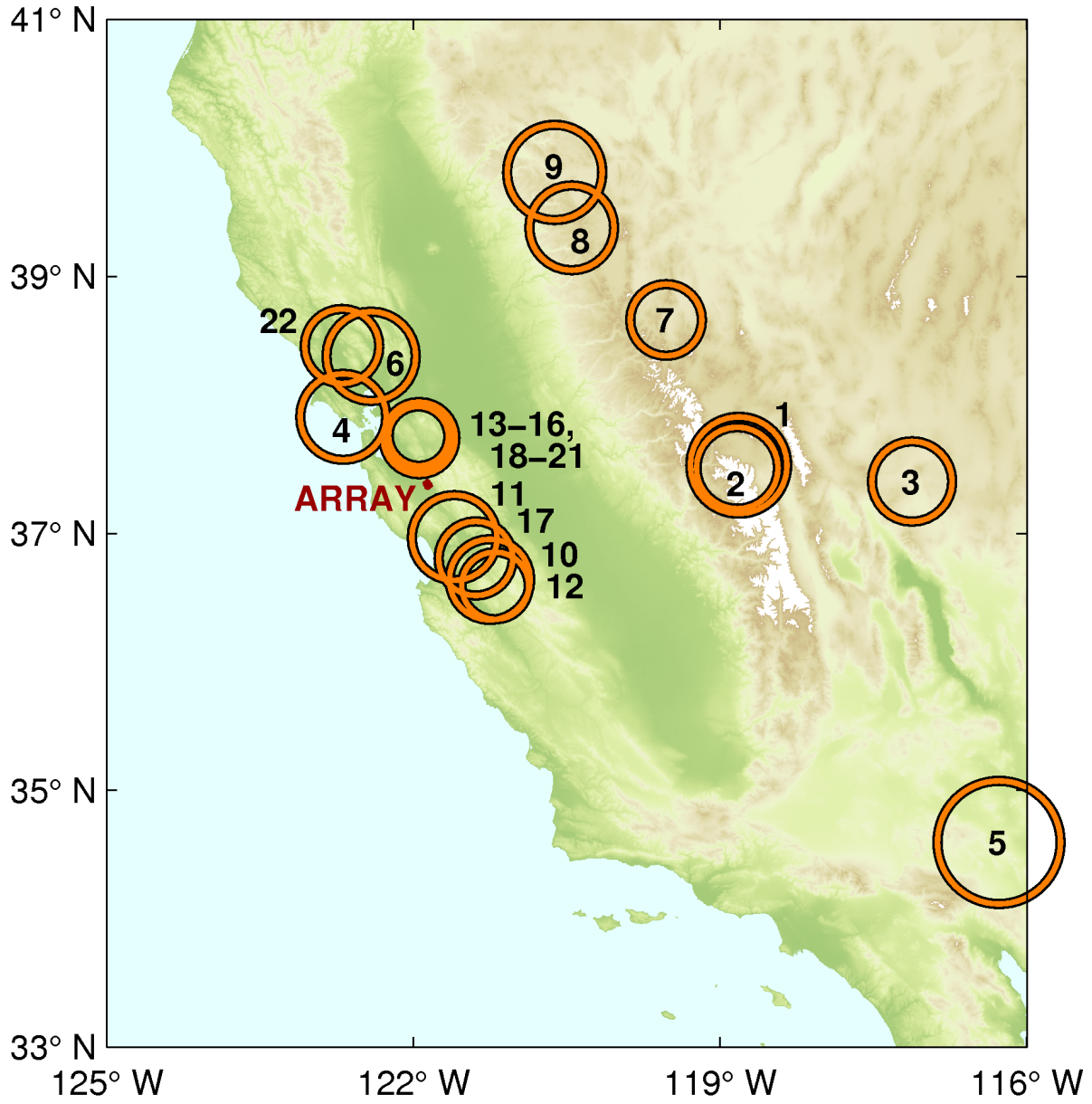


Figure 8.4: Map showing the epicenters of the analyzed earthquakes with respect to the location of the San Jose array. The diameter of the circles indicates the magnitude of the respective earthquake.

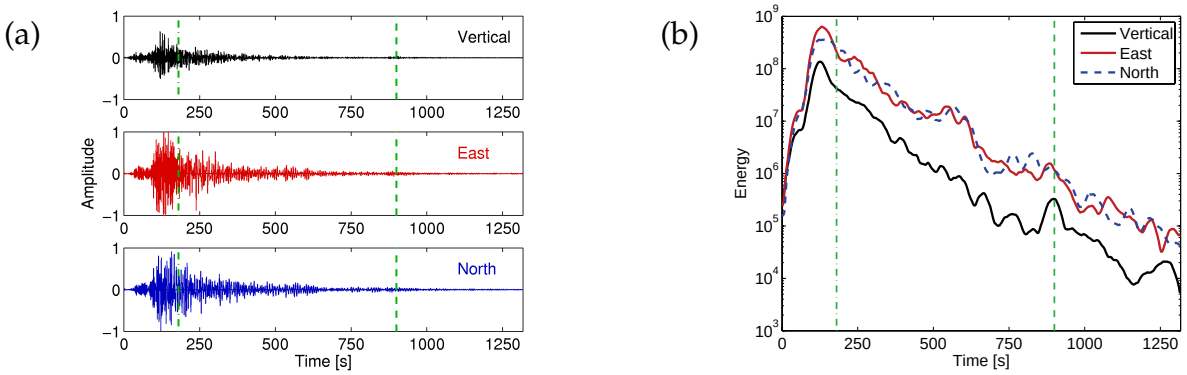


Figure 8.5: (a) Relative accelerogram for the three components of the station marked in figure 8.2 (a) recorded during event 5. (b) The corresponding energy of the signal as a function of time. The vertical lines delimitate the early signal including the first arrivals and the late signal which are both used in the following.

8.4 Data processing

The signal recordings for each of the events shown in table 8.1 have been analyzed in the following way. The signals of those of the 25 stations marked by a red circle in figure 8.2 (a) which recorded the event (i.e. for which the signal was strong enough to trigger the recording) form the array recording. The number of array stations which recorded the respective event is also given in table 8.1. The signals have been analyzed using MUSIQUE (see section 4.6) for 50 different frequency values between 0.05 and 0.5 Hz, using a logarithmic frequency scale. This frequency range was chosen using the basin model's dispersion curves (figure 8.3 (e)) in combination with the array response (figure 8.2 (b)). According to both information, the dispersion curve can be best retrieved between 0.1 and 0.5 Hz. Below 0.1 Hz, however, the recordings for most events are of very low energy.
The actual signal processing was performed as described in the following. The signals are prefiltered using a Chebyshev filter with a bandwidth of 0.1 times the respective frequency. For example, analyzing the signal at 0.2 Hz, the signal is filtered between 0.19 and 0.21 Hz. Then, the signals are cut in time windows of three periods length. For each time window, the wave vector, i.e. azimuth and slowness, of the dominant signal is determined using the classical MUSIC algorithm. Then, the signal is projected in the identified azimuth direction to separate the transverse and radial components. The energy contents of the three components are calculated using the squared absolute values of the Fourier transform for 5 frequency values around the analyzed frequency. Based on the energy of the respective components, the algorithm identifies Love waves (if the transverse energy exceeds the sum of the radial and vertical energies) and possible Rayleigh waves (in the opposite case). If the wave is identified as a possible Rayleigh wave, its polarization parameters are estimated by using quaternion-MUSIC.

The final Rayleigh wave identification is based on the estimation of the phase difference φ: For $45^\circ < \varphi < 135^\circ$, the wave is identified as retrograde Rayleigh wave, for $225^\circ < \varphi < 315^\circ$ as prograde wave and for other values, it is left unclassified.

8.5 Azimuthal energy distribution

8.5.1 Comparison between the the early and late signals

According to table 8.1, there are four earthquakes (events 1, 3, 5 and 6) for which the recorded signal duration is significantly longer than three minutes. The beginning of the record for these events is clearly dominated by the arrival of direct waves from the earthquake. Once these waves have left the array, the signal will exhibit features linked to the basin structure. Therefore, analyzing the beginning and end of these signals independently can teach us on the basin structure from its effects on the seismological wave field. Phillips and Aki [1986] and Riepl et al. [1998] suggested that the beginning of the coda of earthquake data starts at about two times the shear wave travel time from the earthquake to the array. To be in the safe range, in our investigation of the early arrivals, the early signal ends at about 1.5 times the shear wave travel time. Due to the use of accelerometers, for most events, the signals are of very low energy at low frequencies (around 0.1 Hz).

8.5.1.1 Event 1

In figure 8.6, the energy distribution with respect to the incident angle for event 1 is plotted for Rayleigh and Love waves in different frequency ranges. The energy distribution for the first 75 seconds of signal is indicated in figure 8.6 (a), for the signal after 200 seconds in figure 8.6 (b). The relative energy distributions of Rayleigh and Love waves which are indicated by the respective diagrams are normalized in such way that the sum of the total energies of Love and Rayleigh waves over all frequency ranges equals 1. In the figure below, it is indicated which part of the total energy in the respective frequency range is carried by the identified Love and Rayleigh waves.
In all frequency ranges, the main first Love wave arrivals are principally oriented in the theoretical azimuth direction, i.e. arriving under an azimuth of 85°. The small deviations to the north can be explained by wave refraction on the eastern basin edge. Actually, the north-western to south-eastern orientation of the basin refracts waves arriving from the east by some degrees to the south, so that the array detects them from a direction shifted to the north.

At frequencies below 0.2 Hz, Rayleigh waves arrive principally from the theoretical azimuth, but their amplitude is very low. At higher frequencies, they are still arriving from eastern directions, but are more scattered than Love waves.
In the late arrivals, the Love wave energy is slightly larger than the Rayleigh wave energy. Nevertheless, the energy content in the first signal (the first 75 s) is ten times higher than in this late signal (after 200 s). At all frequencies, a dominant Love wave contribution arriving from the south can be seen, with minor contributions from other directions.

This signal from the south has already been observed by Frankel et al. [2001] who attributed it to a scattered Love wave. The scatterer could either be a large rock formation located south of the array or the southern basin edge. For Rayleigh waves, the energy distribution is rather inhomogeneous, but at lower frequencies (below 0.3 Hz), waves arrive principally from the north-west, whereas the principal direction is south-east for higher frequencies.

8.5.1.2 Event 3

Event 3 is located at the same azimuth (88°), but in a larger distance than event 1. The azimuthal energy distribution for this event is shown in figure 8.7 both for the first 110 seconds of signal (a) and for the signal after 250 seconds (b). Both Love and Rayleigh waves arrive principally from the theoretical azimuth. Between 0.2 and 0.3 Hz, however, the Love wave arrival is dominated by the southern direction. For the late signal arrivals, a large part of energy is still arriving from the main azimuth. Between 0.15 and 0.2 Hz, however, both wave types are arriving under a large variety of azimuths. In the next frequency band between 0.2 and 0.3 Hz, Love waves exhibit three principal azimuths: north, the theoretical azimuth and south.

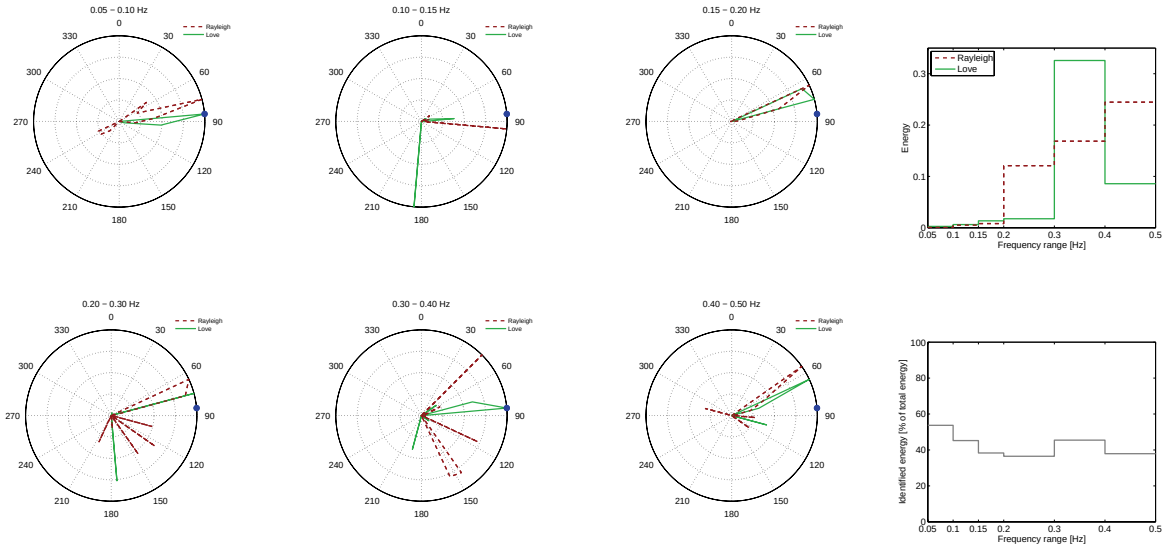
8.5.1.3 Event 5

Event 5 represents the Hector Mine earthquake, with a magnitude of 7.1 the largest recorded earthquake in the data set. With a recording length of 22 minutes, it represents the longest record, too. The earthquake was located at a theoretical azimuth of 120°. In figure 8.8, the azimuthal energy distributions of the first 180 seconds of signal as well as for the signal after 900 seconds are shown for different frequency ranges. In the beginning of the signal, waves are mainly arriving from south-eastern directions, but slightly shifted from the theoretical azimuth. At low frequencies (below 0.15 Hz), Love waves arrive under azimuths slightly shifted to the north, whereas for higher frequencies, a large amount of Love wave energy is arriving from the south. This azimuth was already dominant for the late arrivals of event 1. For Rayleigh waves, no systematic behavior can be observed.
In the late arrivals, smaller amounts of Love waves can be observed from different azimuths, but the main energy contribution is associated to the frequency range between 0.15 and 0.2 Hz, where the waves are mainly arriving from the east, and to the frequency range between 0.2 and 0.3 Hz, where the waves are principally arriving from the south. Here again, Rayleigh waves are more scattered with a multitude of azimuths, but the main energy arrives from southern to south-eastern directions.

8.5.1.4 Event 6

The last event of longer duration is event 6, located under an azimuth of 337°, i.e. in the north of the Santa Clara Valley. The azimuthal energy distributions associated to the first 50 seconds of this signal as well as for the signal after 150 seconds are shown in figure 8.9. In the early signal, the energy is concentrated around the theoretical azimuth. Nevertheless, most of the Love waves are arriving under azimuths which are systematically shifted, e.g. around 0° for the frequency range between 0.2 and 0.4 Hz. These waves could be direct waves which have been reflected once or multiple times at the basin borders. Rayleigh waves are of very low energy below 0.4 Hz. Above that frequency, most of the energy is arriving from the direct azimuth.
For the late signal, the azimuthal energy distribution is wider with more azimuths involved, but no energy arrives from south-eastern directions. However, the most energetic Love wave arrival in this signal part is associated to azimuths centered around the theoretical azimuth. For Rayleigh waves, the energy distribution is wider and more waves arrive from azimuths different from the main azimuth.

(a) Event 1 - first 75 s of signal



(b) Event 1 - signal after 200 s

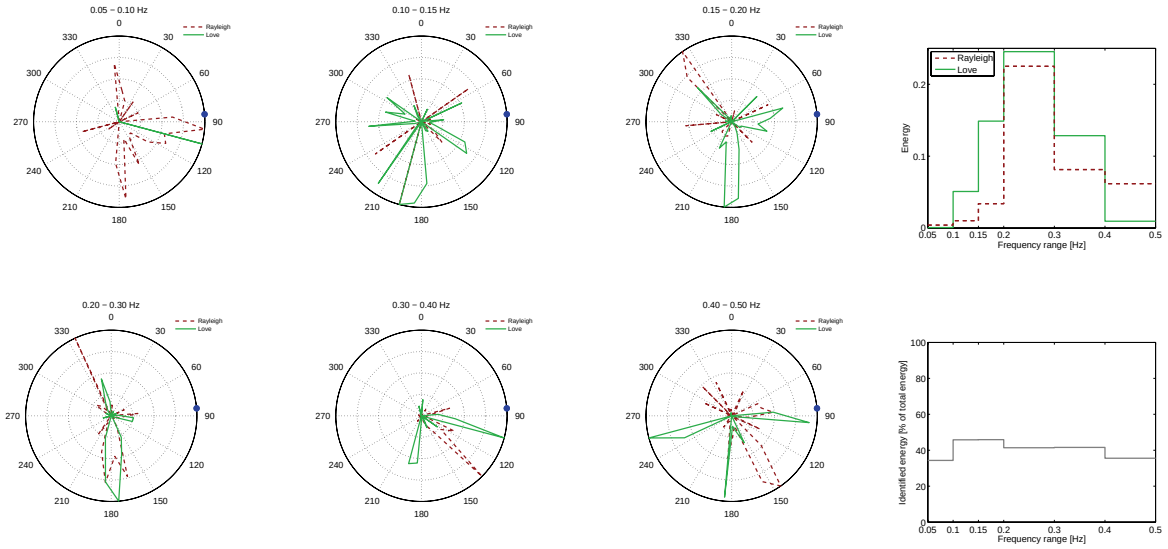
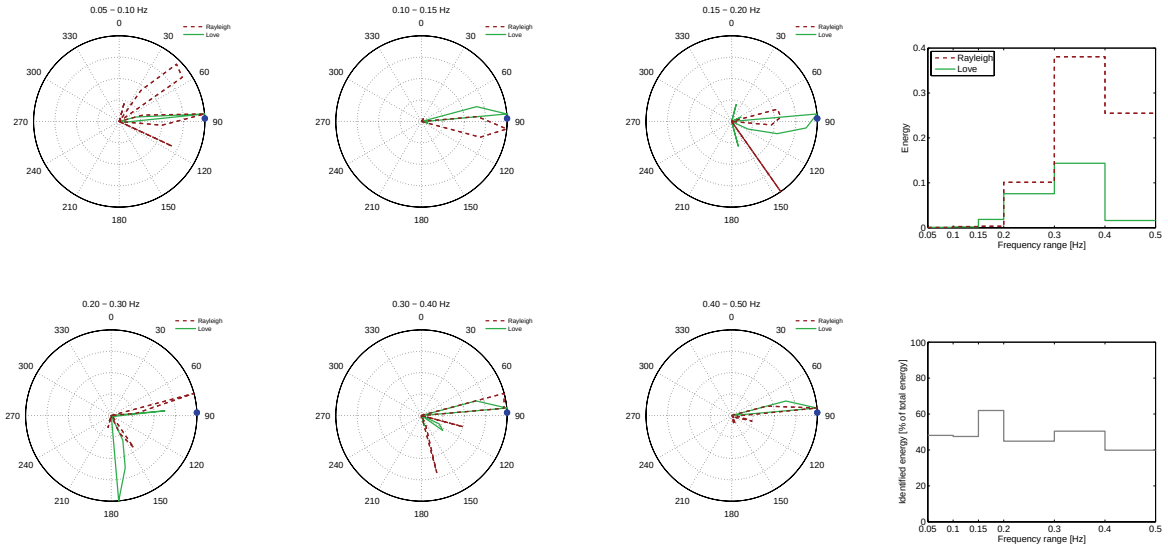


Figure 8.6: Azimuthal energy distribution for different frequency ranges for (a) the first 75 s of signal of event 1 and (b) the signal after 200 s, i.e. after all direct waves have passed the array (bottom). The blue dot indicates the theoretical azimuth of the event. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 3 - first 110 s of signal



(b) Event 3 - signal after 250 s

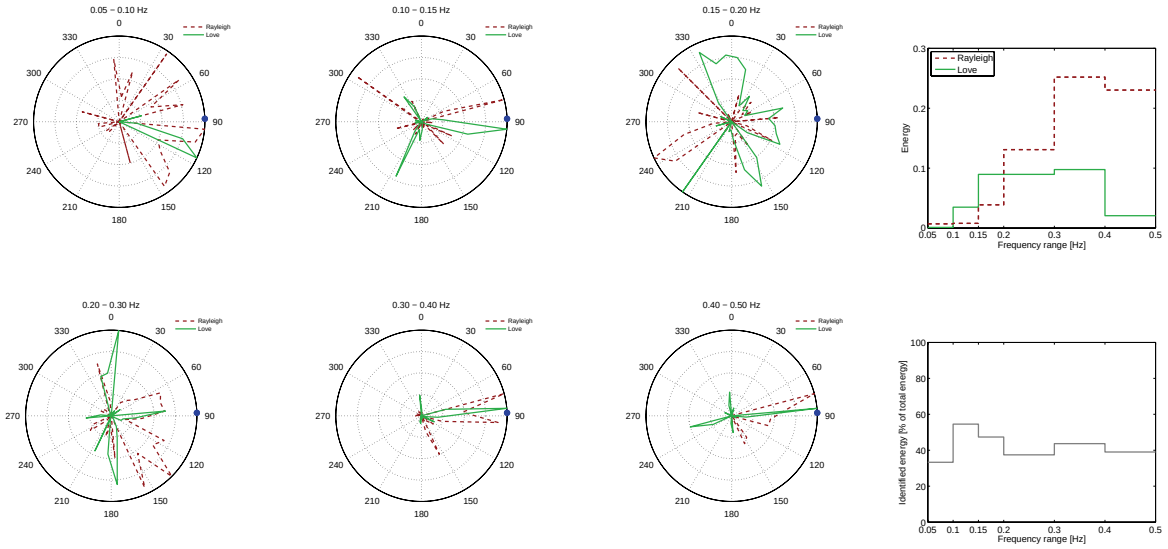
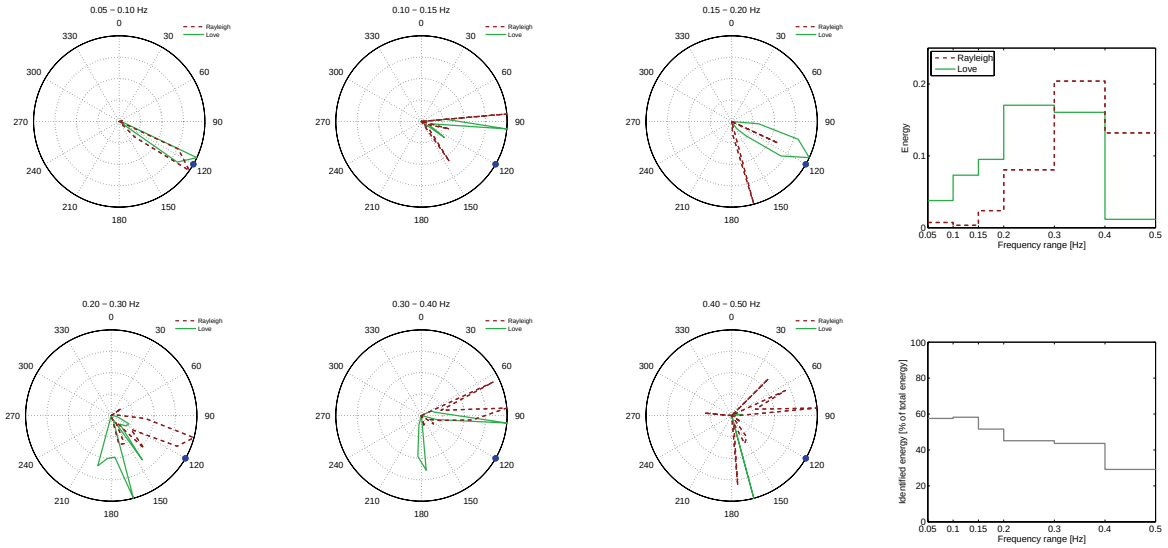


Figure 8.7: Azimuthal energy distribution for different frequency ranges for (a) the first 110 s of signal of event 3 and (b) the signal after 250 s, i.e. after all direct waves have passed the array. The blue dot indicates the theoretical azimuth of the event. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 5 - first 180 s of signal



(b) Event 5 - signal after 900 s

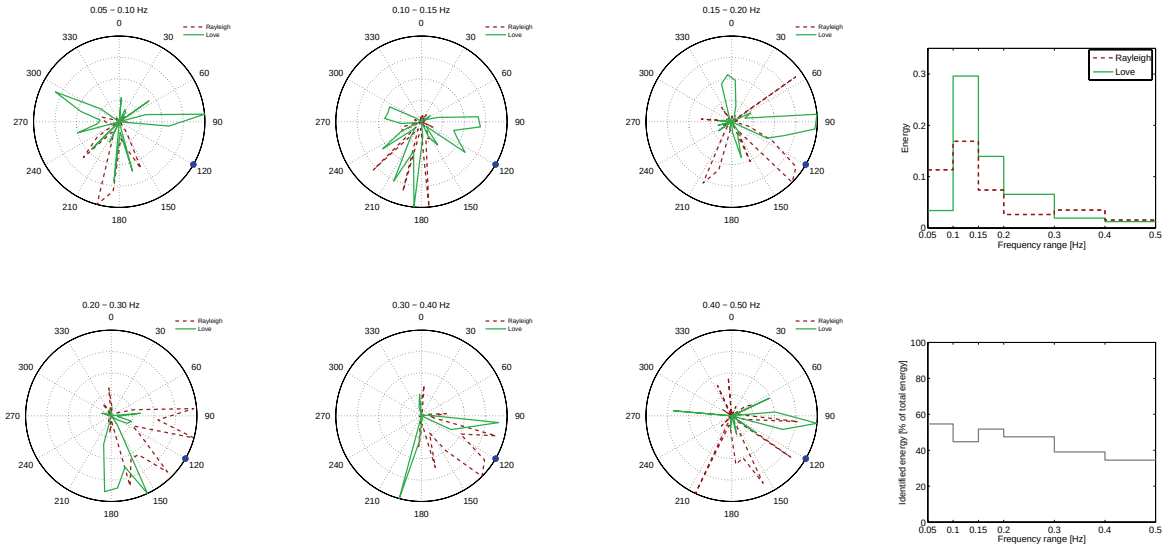
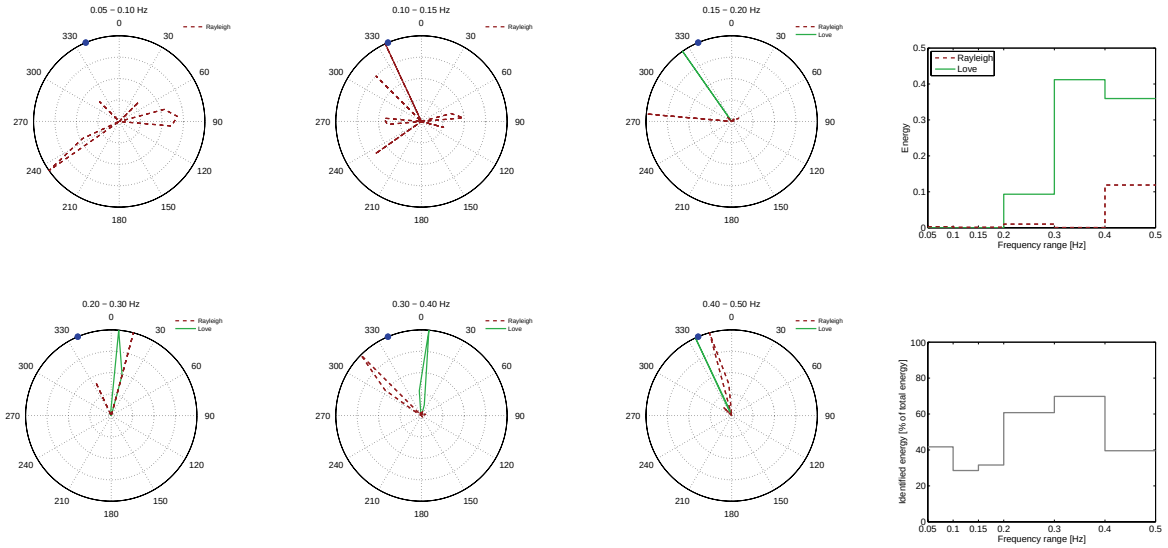


Figure 8.8: Azimuthal energy distribution for different frequency ranges for (a) the first 180 s of signal of event 5 and (b) the signal after 900 s, i.e. after all direct waves have passed the array. The blue dot indicates the theoretical azimuth of the event. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 6 - first 50 s of signal



(b) Event 6 - signal after 150 s

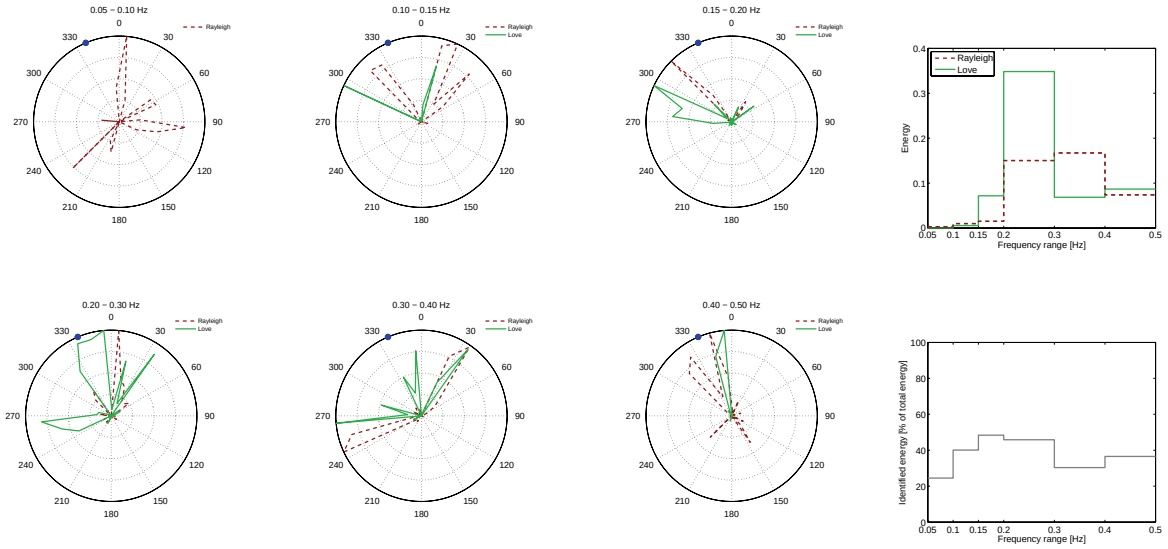


Figure 8.9: Azimuthal energy distribution for different frequency ranges for (a) the first 50 s of signal of event 6 and (b) the signal after 150 s, i.e. after all direct waves have passed the array. The blue dot indicates the theoretical azimuth of the event. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

8.5.2 Temporal Love wave azimuth evolution

In figures 8.6, 8.7 and 8.8, a large amount of Love waves arriving from the south was identified. The main frequency range of these arrivals lies between 0.15 and 0.3 Hz. Therefore, the temporal evolution of the azimuth of Love waves in this frequency range is shown in figure 8.10 for the events 1, 5 and 6. Figure 8.6 indicated that the early Love wave field is dominated by arrivals from the theoretical azimuth direction, whereas the late wave field is dominated by arrivals from the south. This effect can also be seen in figure 8.10. In all analyzed frequency bands, the first arrivals come approximately from the direct azimuth, slightly deviated to the north. Later arrivals are dominated by waves arriving at azimuths of 180°. For frequencies between 0.25 and 0.30 Hz, both main azimuths are not well separated, i.e. arrivals from 180° are directly following arrivals from 90° and vice versa.
For event 5, due to the length of the signal, a lot more time windows have been analyzed. Between 0.15 and 0.20 Hz, waves arrive under all azimuths between 60 and 180° during the whole signal length. In the next frequency range (0.20 - 0.25 Hz), two principal azimuths at 90° and 160° can be identified, but waves arrive as well from azimuths between these two values. In the frequency range between 0.25 and 0.30 Hz, finally, waves arrive almost exclusively from two distinct directions, 90° and 180°. The first arrivals in these frequency range are not coming from the direct azimuth, but from southern directions. However, in all frequency ranges, the azimuthal energy distribution does not change much with time. Towards the end of the signal, some low-energy arrivals from azimuths different from the two main directions can be seen.
For events 1 and 5, the deviation of the direct arrivals (75° instead of 85° for event 1, 90° instead of 120° for event 5) can be explained by refraction of the direct waves at the basin borders. The arrivals from the south in both events can be explained by a large scatterer located at the south. This result has already been found by Frankel et al. [2001].
For event 6, the theoretical azimuth is 337° and the source is located in north of the Santa Clara Valley. For the frequency range between 0.15 and 0.20 Hz, the main Love wave azimuth is about 270°. However, the very first arrivals are in the theoretical azimuth. The same can be observed for larger frequencies (0.20 - 0.25 Hz), but another part of the signal is arriving with azimuths around 0°. In the frequency range from 0.25 to 0.30 Hz, finally, the main part of the energy is arriving under azimuths around 0°, including the first arrival. Some later energy is arriving from 270°, too. Here again, the principal energy distribution is rather constant in time. These results can be explained by the location of the source in the approximate extension of the basin. Waves originating at the source are reflected at the lateral basin borders. The array detects these waves as arrivals from larger or lower azimuth values.

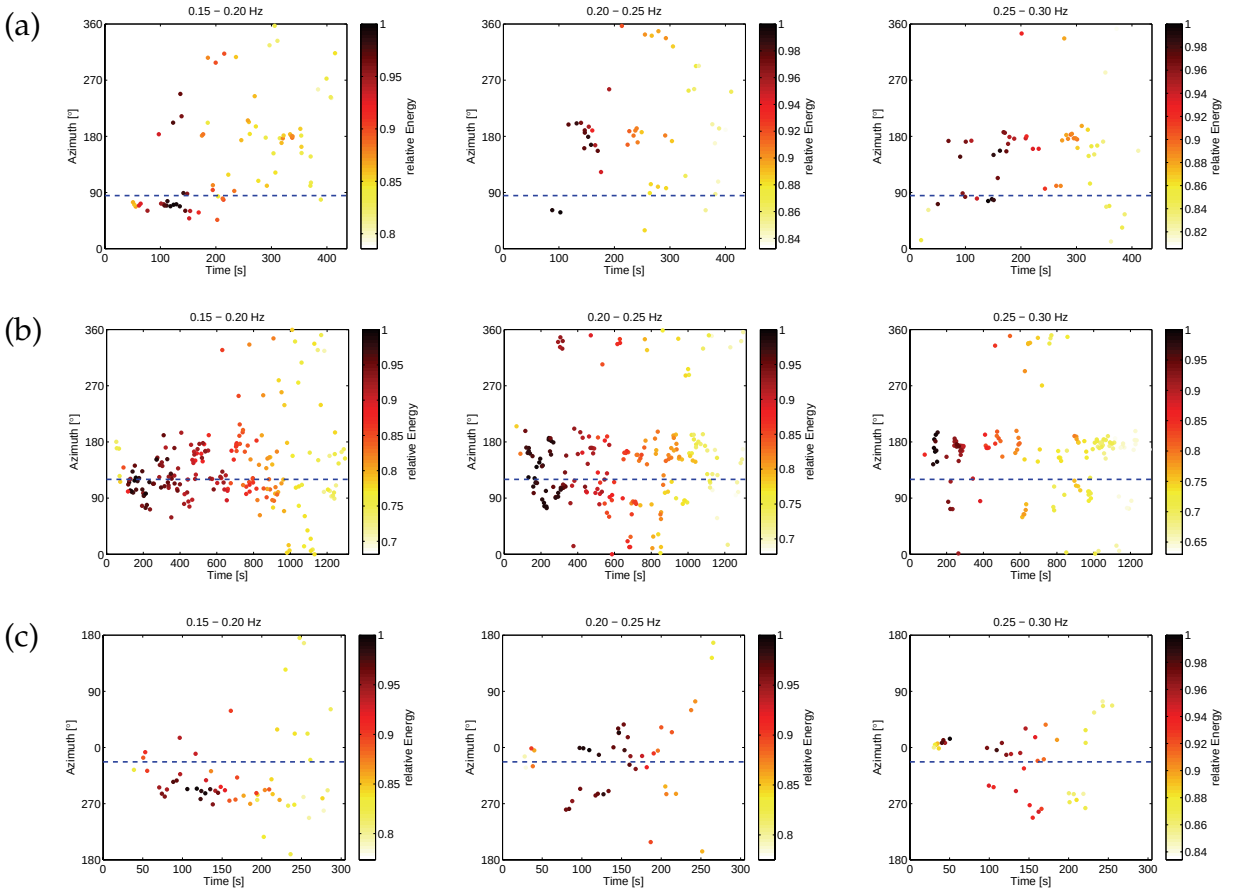


Figure 8.10: Evolution of the azimuth of Love waves with time for different frequency ranges for (a) event 1, (b) event 5 and (c) event 6. For all analyzed frequencies which lie inside the indicated frequency ranges (left: 0.15 - 0.20 Hz, center: 0.20 - 0.25 Hz, right: 0.25 - 0.30 Hz), the azimuths for all time windows identified as Love waves are shown. The color indicates the Love wave energy of the respective time window in a logarithmic scale.

8.5.3 Results for the other events

For the events with longer signal durations, the azimuthal energy distribution has been shown both for the beginning and for the end of the signal in figures 8.6 - 8.9. For the other events, the signal duration is too short to exhibit a significant end signal. Therefore, for these events, the azimuthal energy distributions for the complete signal are shown in figures 8.11 - 8.16. The events are shown according to their respective azimuths in clockwise direction, starting in the north-west.

8.5.3.1 Events 4 and 22

Events 4 and 22 (figures 8.11 and 8.12 (a)) are located in the extension of the Santa Clara Valley to the north-west, close to event 6 (see figure 8.9). For both events, at frequencies below 0.3 Hz, most ot the energy is coming from azimuths close to the theoretical azimuth. In the range between 0.2 and 0.3 Hz, the Rayleigh wave energy is concentrated around the theoretical azimuth, but the Love waves arrive under different azimuths which could correspond to waves reflected at the lateral basin edges. At frequencies above 0.3 Hz, both events show Rayleigh and Love waves arriving under a multitude of different azimuths concentrated in the north-western hemicycle. For event 22, at frequencies above 0.4 Hz a signal from the south can be seen, even though it is of low energy.

8.5.3.2 Events 13-16 and 18-21

For the two groups of events 13-16 and 18-21 the results are shown in figures 8.12 (b) and 8.13 (a), respectively. All of these events are located under theoretical azimuths of about 350°, at a distance of about 40 km from the array. For all these events, the most energetic arrival is composed of Love waves in the frequency range between 0.3 and 0.4 Hz. These waves arrive from the azimuth direction, slightly deviated to the east, which is probably due to refraction at the basin border. The low-energy arrivals at lower frequencies are mainly arriving from directions close to the theoretical azimuth. In the frequency range above 0.4 Hz, the Love wave arrival is slightly deviated to the east, whereas the Rayleigh waves arrive from the theoretical azimuth. For these events, no energy arriving from the south-eastern hemicycle can be observed.

8.5.3.3 Events 7, 8 and 9

For event 9 (figure 8.13 (b)), the arrivals below 0.3 Hz are out of azimuth for both wave types. However, the Rayleigh wave arrivals above 0.3 Hz which contribute most of the energy to the wave field are arriving approximately under the theoretical azimuth, but slightly deviated to the south. Love waves are principally arriving from the theoretical azimuth between 0.3 and 0.4 Hz, but are coming from the south above 0.4 Hz.
For event 8 (figure 8.14 (b)), most of the energy is carried by Rayleigh waves arriving slightly south of the theoretical azimuth. Love waves of significant energy are arriving from the north between 0.2 and 0.3 Hz and from the south above 0.3 Hz.

For event 7 (figure 8.14 (b)), the dominant Rayleigh wave arrivals are exactly in the theoretical azimuth direction for all frequencies above 0.15 Hz. The Love wave arrivals are more scattered, but the most energetic Love waves in the range from 0.2 to 0.4 Hz are coming from the south. The arrival of Rayleigh waves from the exact theoretical azimuth might be explained by the fact that the theoretical azimuth is perpendicular to the basin edge. Therefore, the waves pass the border without being deviated.

8.5.3.4 Events 2

Most Love wave arrivals for event 2 (figure 8.15 (a)) arrive from the theoretical azimuth direction with slight differences, but a large part of waves in the frequency range between 0.3 and 0.4 Hz is arriving from the south as well. Rayleigh waves, which are dominating the wave field, are mainly incident from angles north of the theoretical azimuth.

8.5.3.5 Events 10/12, 11 and 17

For the combination of events 10 and 12, which are located to the south-east, the main Love wave arrivals are coming from south-eastern directions (figure 8.15 (b)), but deviated to the north. The main Rayleigh wave arrivals are associated to the theoretical azimuth.
For event 17 (figure 8.16 (a)), the only significant part of Love waves between 0.2 and 0.3 Hz comes from the east, whereas Rayleigh waves are mainly associated to the theoretical azimuth.
The main Love wave arrival of event 11 (figure 8.16 (b)) is associated to southern and south-eastern directions, but in the frequency range between 0.3 and 0.4 Hz, Love waves come from the east. For the higher-energy arrivals of Rayleigh waves above 0.2 Hz, the deviation from the theoretical azimuth increases with frequency. Between 0.2 and 0.3 Hz, the waves arrive from about 130°, between 0.3 and 0.4 Hz from 120° and above 0.4 Hz from about 90°.

8.5.4 Summary

The events can be divided into two principal groups, events arriving from the north and events arriving from the east. For events arriving from the north (i.e. events 4, 6, 8 and 9, 13 through 16, 18 through 22), the wave energy is confined to the north-western hemicycle and no energy is arriving from south-eastern directions, except for Love wave arrivals from the south, which can be seen for almost all of these events in the frequency range between 0.4 and 0.5 Hz. Furthermore, evidence for Love and Rayleigh wave reflections at the lateral basin edges has been found that leads to significant deviations of the wave arrivals.

For events arriving from the south (i.e. events 1, 2, 3, 5, 10 through 12 and 17), the main Love wave arrivals are only slightly deviated from the theoretical azimuth. For all events, Love waves arriving from the south can be seen, both in the early and late signals. In contrast to the northern events, where arrivals from the south occured above 0.4 Hz, these waves are present over the whole frequency, with a maximum contribution between 0.2 and 0.3 Hz. As waves already arrive from this azimuth the very beginning of the signals, these waves could be associated to primary surface or shear waves. The Rayleigh wave arrivals are more scattered for these events.
Because its theoretical azimuth is perpendicular to the basin border, event 7 represents a special case. In this case, the Rayleigh waves arrive mainly from the direct azimuth, but Love waves are more scattered and arrive predominantly from the south between 0.2 and 0.4 Hz.

Event 4

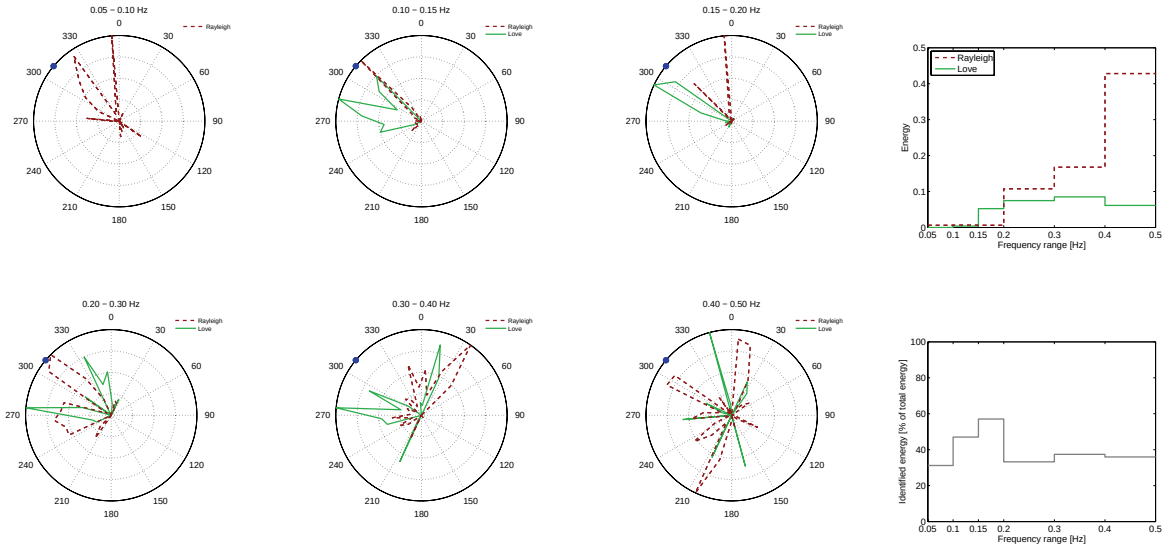
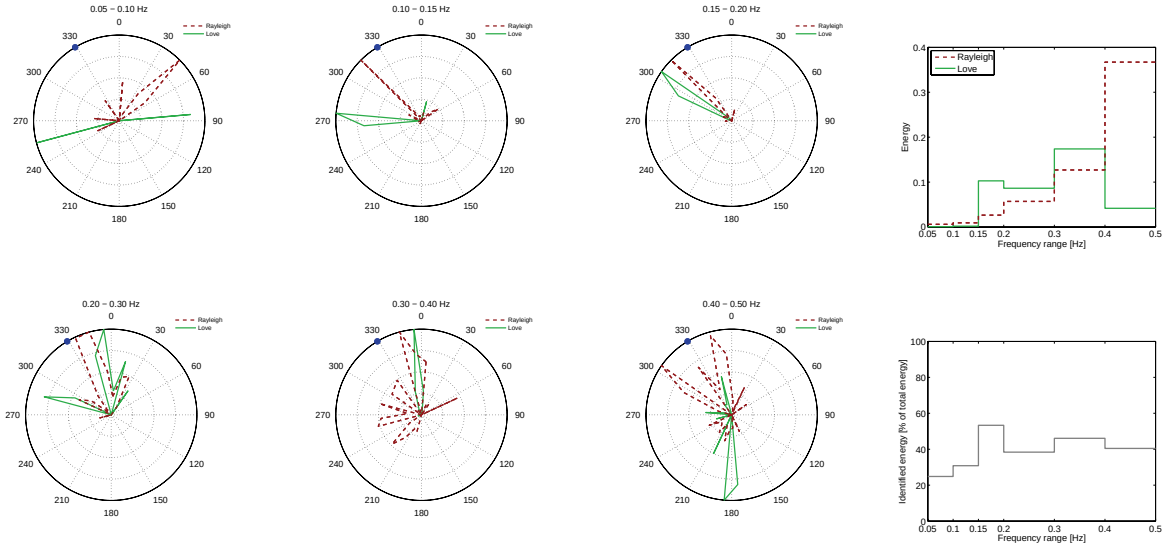


Figure 8.11: Azimuthal energy distribution for different frequency ranges for event 4. The blue dot indicates the theoretical azimuth of the events. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 22



(b) Events 13-16

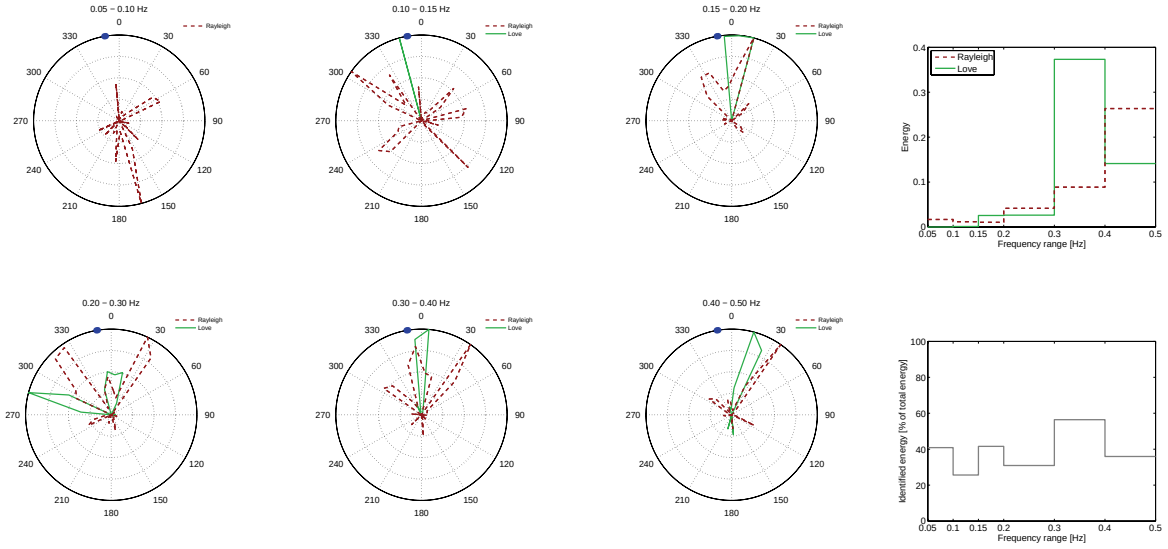
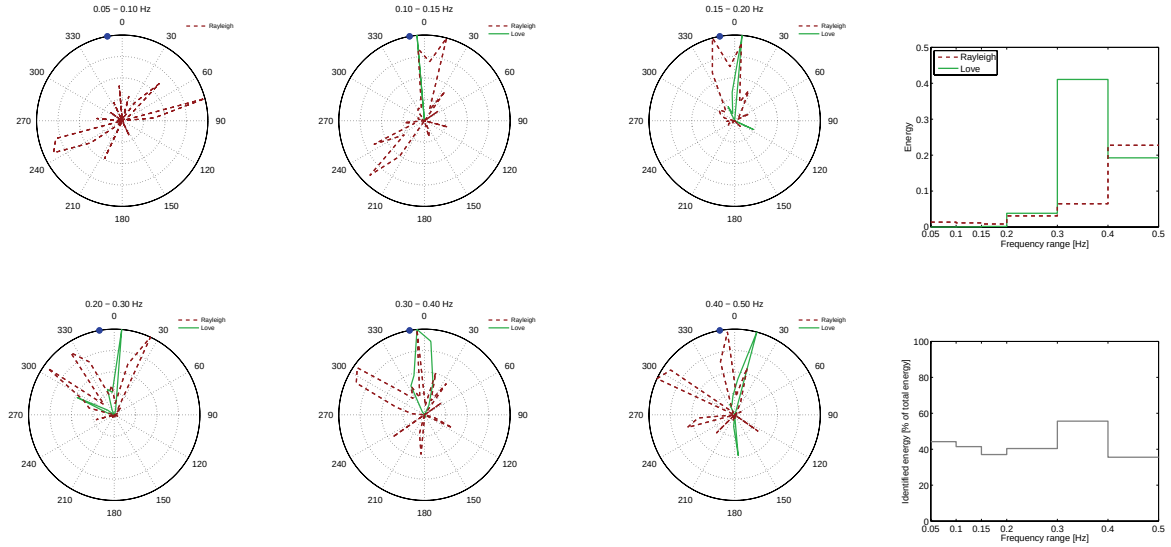


Figure 8.12: Azimuthal energy distribution for different frequency ranges for (a) event 22 and (b) events 13-16. The blue dot indicates the theoretical azimuth of the events. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Events 18-21



(b) Event 9

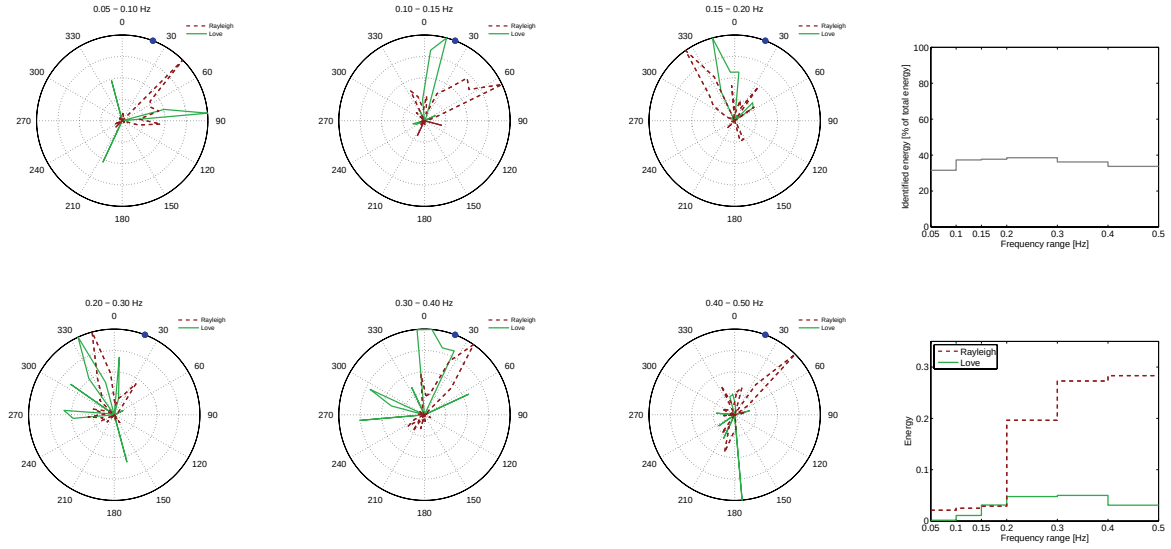
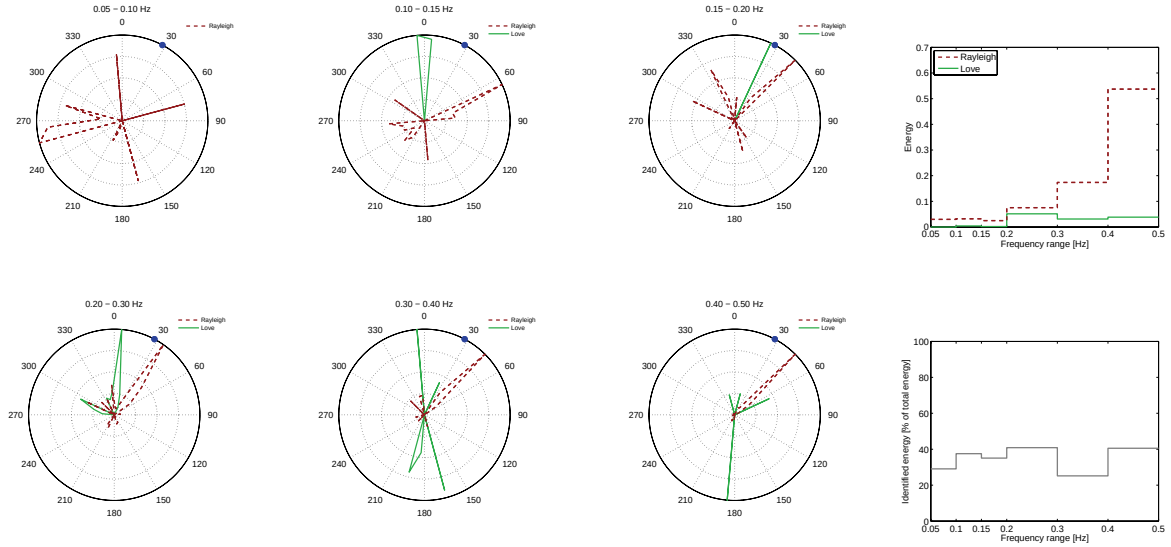


Figure 8.13: Azimuthal energy distribution for different frequency ranges for (a) events 18-21 and (b) event 9. The blue dot indicates the theoretical azimuth of the events. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 8



(b) Event 7

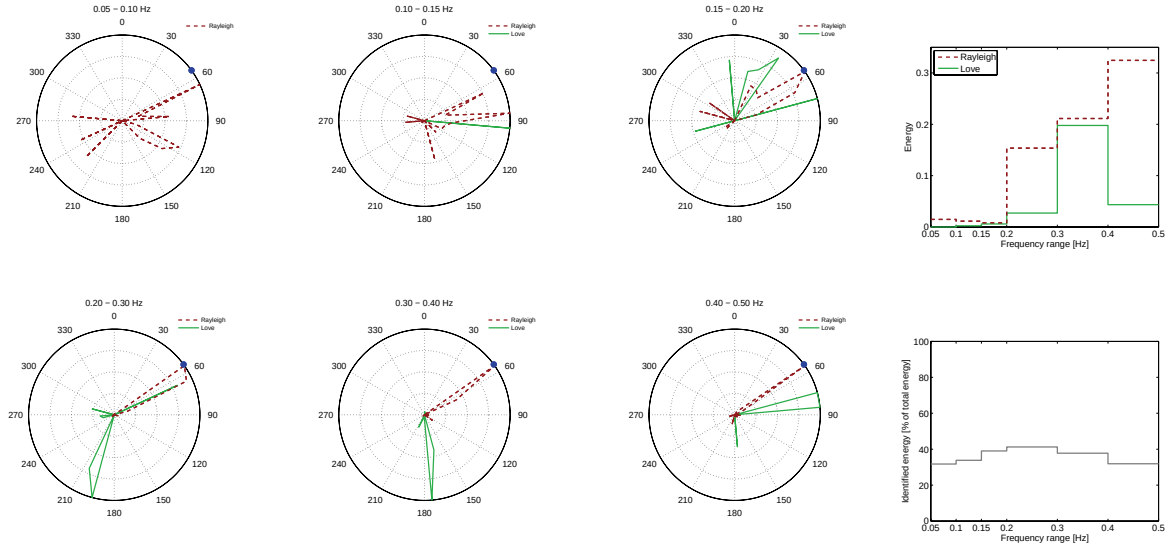
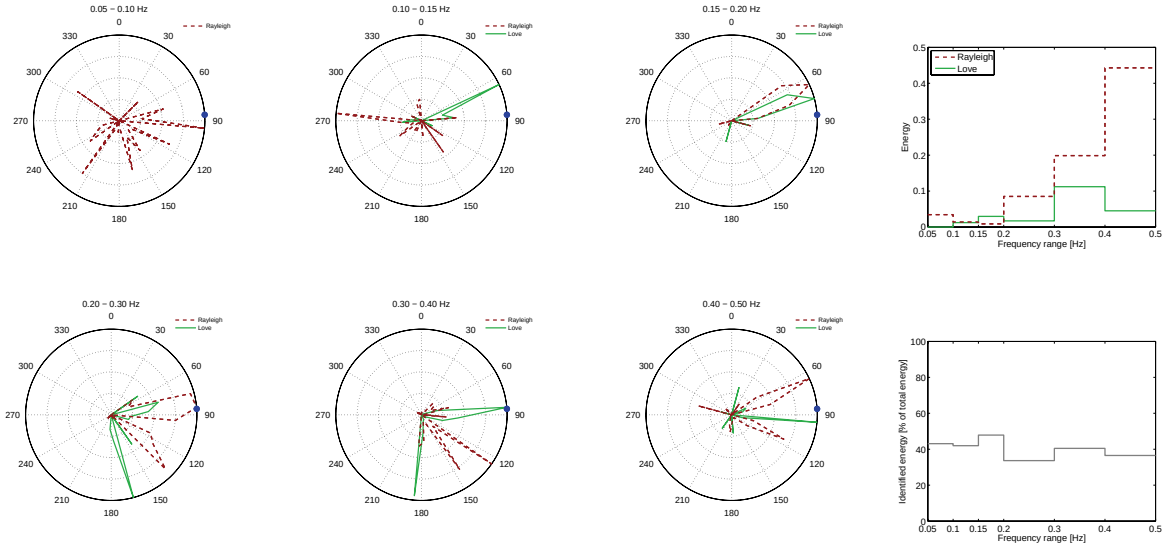


Figure 8.14: Azimuthal energy distribution for different frequency ranges for (a) event 8 and (b) event 7. The blue dot indicates the theoretical azimuth of the events. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 2



(b) Events 10 and 12

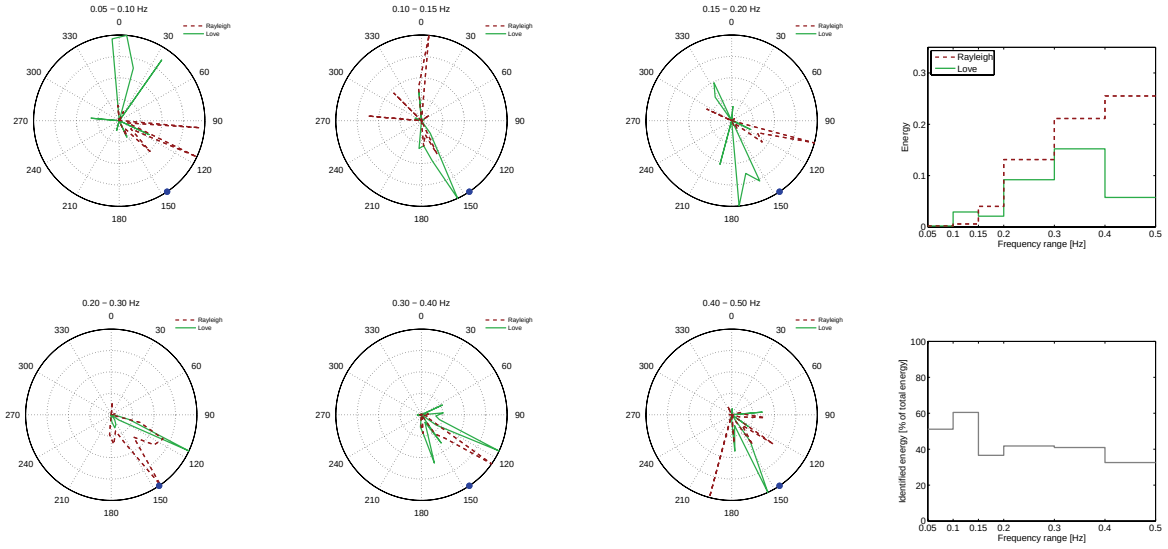
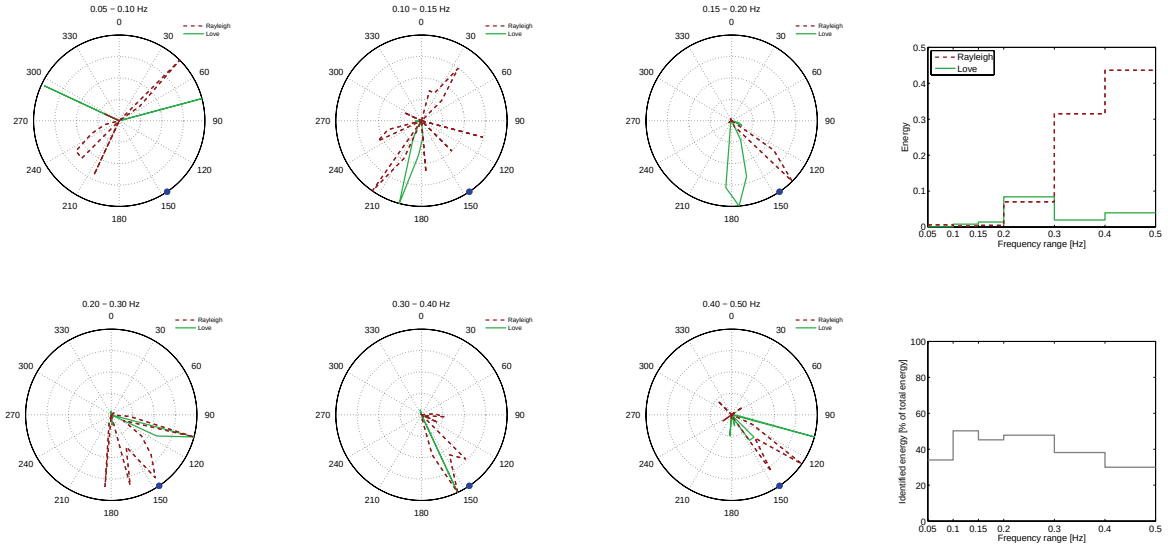


Figure 8.15: Azimuthal energy distribution for different frequency ranges for (a) event 2 and (b) events 10 and 12. The blue dot indicates the theoretical azimuth of the events. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

(a) Event 17



(b) Event 11

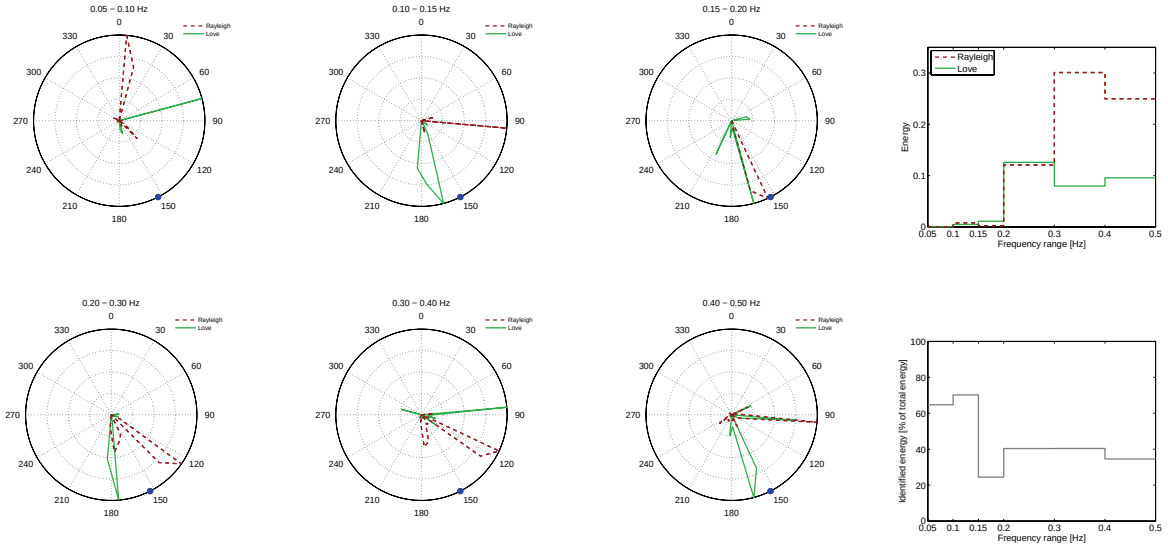


Figure 8.16: Azimuthal energy distribution for different frequency ranges for (a) events 17 and (b) event 11. The blue dot indicates the theoretical azimuth of the events. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of both wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.

8.6 Energy repartition between Love and Rayleigh waves

The energy repartition between the different surface wave types is given for each event in table 8.2. The indicated energies are the total energies for all 50 analyzed frequencies in the range between 0.05 and 0.50 Hz. We decided in the MUSIQUE code to identify the signal of a time window as Love wave if the transverse energy in the respective time window is dominant. The Love wave energy indicated in the table corresponds to the summed transverse energies of all time windows which have been attributed to Love waves. The radial and vertical energies of these windows are also included in the table. These latter energies correspond to the less energetic waves in the signal, which can include Love or Rayleigh waves arriving under different azimuths. The energy of these "wrong" components is, by construction of the method, always smaller than the energy of the components corresponding to the identified wave. For the values given in the table, in average, the energies not corresponding to the identified dominant wave type carry about two-thirds of the energy corresponding to the identified wave type (i.e. in average, the radial and vertical components of the windows identified as Love waves carry about 67 % of the energy of the respective transverse components). For the Rayleigh waves, the signal energy is carried by the radial and vertical components, whereas the transverse component is composed by inferior wave contributions.
The energy repartition between the different wave types varies largely for the different events. The proportion of Love waves ranges from 4.5 % (event 8) to 30.9 % (event 19), with an average of 17.4 %. The retrograde Rayleigh wave proportion is quite stable for the different events, taking values from 7.0 to 12.8 %, whereas the prograde Rayleigh wave proportion is more variable, accounting for 5.4 to 25.2 % of the energy. The unclassified waves contribute between 25.0 and 40.5 % to the total energy, but, in fact, also the contributions of the components not corresponding to the identified polarization of each time window are not attributed to a wave type. The energy repartition between classified Love and Rayleigh waves is indicated in the table as well.
The last column of the table indicates the proportion of Love wave energy which is arriving from the south (with azimuths between 150° and 210°). This proportion is also represented in function of the theoretical earthquake azimuth in figure 8.17. For events located north of the array, this proportion ranges from 0 to 12.8 %. For events located in the north-east, the proportion is rather high, with values of 12.4 % (event 9), 30.1 % (event 8) and the highest proportion of all events is observed for event 7 with 76.8 % of the Love wave energy arriving from the south. For events located at the east, about 30 % of the Love wave energy comes from the south. Finally, for events located south-east of the array, 31.0 to 58.7 % of the Love wave energy is incident from southern directions.

Table 8.2: List of the energy partition for all analyzed earthquakes, sorted by increasing azimuths starting in the north-west. The columns indicate the transverse and combined radial and vertical energies of all time windows identified as Love waves and retrograde or prograde Rayleigh waves, plus the energy attributed to time windows which have not been classified. The ninth column indicates the sum of the transverse energy of windows identified as Love waves and the radial and vertical components of the windows identified as Rayleigh waves, i.e. the sum of the second, fourth and sixth columns. The values are indicated as percentages of the total energy content in the frequency range between 0.05 and 0.50 Hz. The next columns indicates the energy repartition between Love and Rayleigh wave energy in the classified time windows. In the last column, the energy proportion of Love waves arriving from southern directions (with azimuths between 150° and 210°) compared to the total Love wave energy is indicated.

Event ID	Energies (in % of the total energy) of the windows identified as							sum of all classified components	Repartition Love/Rayleigh in classified components	Percentage of Love waves arriving from the south
	Love waves		retrograde Rayleigh waves		prograde Rayleigh waves		unclassified			
	E_t	$E_r + E_v$	$E_r + E_v$	E_t	$E_r + E_v$	E_t	$E_t + E_r + E_v$			
4	10.2	7.7	9.0	6.2	17.5	12.0	37.4	36.7	27.8/72.2	10.6
22	17.4	12.2	11.1	7.7	14.2	10.5	26.9	42.7	40.7/59.3	6.5
6	20.9	14.2	11.2	8.2	6.6	4.4	34.5	38.7	54.0/46.0	1.0
13	26.7	17.4	7.9	5.3	9.2	6.3	27.2	43.8	61.0/39.0	6.4
14	21.2	11.7	12.4	7.3	7.4	4.4	35.6	41.0	51.7/48.3	0.0
15	23.8	12.1	11.1	6.3	11.8	7.4	27.5	46.7	51.0/49.0	1.4
16	21.5	12.4	8.8	5.6	8.6	6.0	37.1	38.9	55.3/44.7	0.0
18	22.3	13.1	12.8	8.4	6.5	3.7	33.2	41.6	53.6/46.4	0.0
19	30.9	18.3	7.5	4.9	5.6	3.3	29.5	44.0	70.2/29.8	12.8
20	25.2	15.2	12.7	7.5	5.4	3.5	30.5	43.3	58.2/41.8	1.4
21	26.0	15.8	7.9	3.6	9.1	5.8	31.8	43.0	60.5/39.5	5.2
9	6.2	5.1	12.2	7.6	17.5	11.6	39.8	35.9	17.3/82.7	12.4
8	4.5	3.7	9.3	5.6	21.7	14.8	40.5	35.5	12.7/87.3	30.1
7	9.9	7.4	7.0	4.7	18.9	11.7	40.4	35.8	27.7/72.3	76.8
1	22.5	16.0	9.5	6.7	10.9	7.4	27.0	42.9	52.4/47.6	30.0
2	8.2	6.4	11.1	7.6	18.7	11.7	36.3	38.0	21.6/78.4	27.1
3	13.0	9.9	7.4	5.5	25.2	14.0	25.0	45.6	28.5/71.5	25.0
5	27.8	17.2	9.6	6.9	8.1	5.4	25.0	45.5	61.1/38.9	32.7
10	13.5	10.1	10.6	6.5	14.1	9.8	35.4	38.2	35.3/64.7	38.3
12	12.6	9.7	9.2	5.6	18.3	13.1	31.5	40.1	31.4/68.6	32.1
17	5.8	4.4	7.2	5.4	21.9	15.7	39.6	34.9	16.6/83.4	31.0
11	12.1	7.3	10.5	7.6	15.5	10.2	36.8	38.1	31.8/68.2	58.7

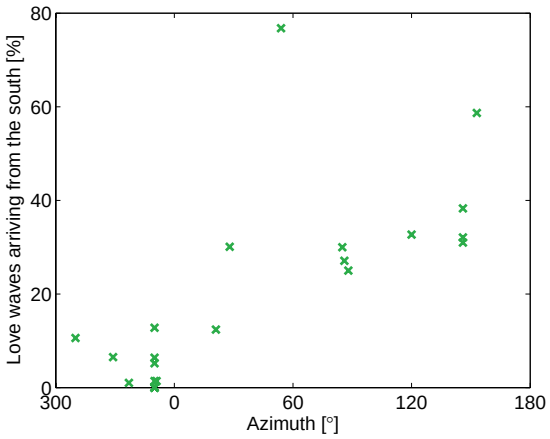


Figure 8.17: Proportion of Love wave energy arriving from southern directions (azimuths between 150° *and* 210°*) in function of the theoretical azimuth for all analyzed earthquakes.*

In figure 8.18, the energy repartition between Love, retrograde and prograde Rayleigh waves is shown in different frequency ranges for all analyzed events. Some events which are closely located and exhibit similar behavior have been grouped together for clarity reasons. A strong peak of the Love wave energy can be seen for frequencies between 0.15 and 0.20 Hz for all events except event 5 and the group of events 18-21. For event 5, the strongest magnitude event in the dataset, however, the Love wave energy is very strong in the whole low-frequency range. The two groups of events 13-16 and 18-21 both show very energetic Love waves between 0.3 and 0.4 Hz. With distances of 44 km and 42 km, respectively, these events are the closest to the array. For the other events, it might be that Love waves have been generated in this frequency range, but are already attenuated when arriving at the array. Nevertheless, a smaller Love wave energy peak at these frequencies can be seen for events 22, 6, 7, 1 and 2, which are all located under azimuths between 329° and 86°.
For retrograde Rayleigh waves, no specific behavior can be observed, for some events the energy content is almost constant with frequency (events 13-16, 3 and 5), while for other events the amount of energy changes strongly with frequency (e.g. events 4, 18-21, 7), without following a systematic scheme.
Prograde Rayleigh waves are strongest at higher frequencies for most events. According to the basin model, the fundamental Rayleigh wave mode is retrograde at all frequencies. Therefore, prograde Rayleigh waves correspond to the first harmonic mode, which does not exist below 0.17 Hz. Consequently, prograde Rayleigh waves found at lower frequencies are probably misidentified. The peak of prograde Rayleigh waves observed for events 18-21 in the range between 0.15 and 0.20 Hz is not visible for other events.

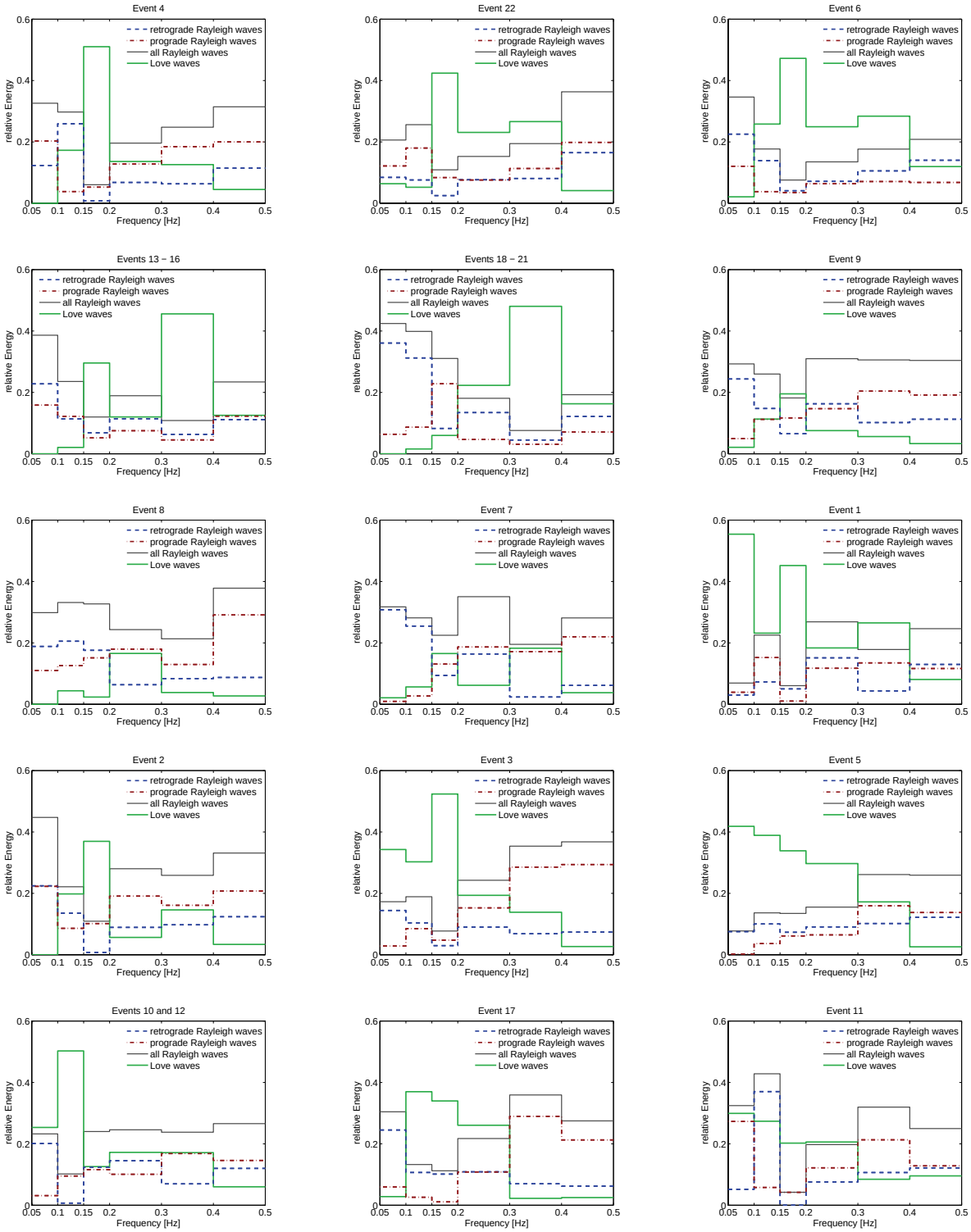
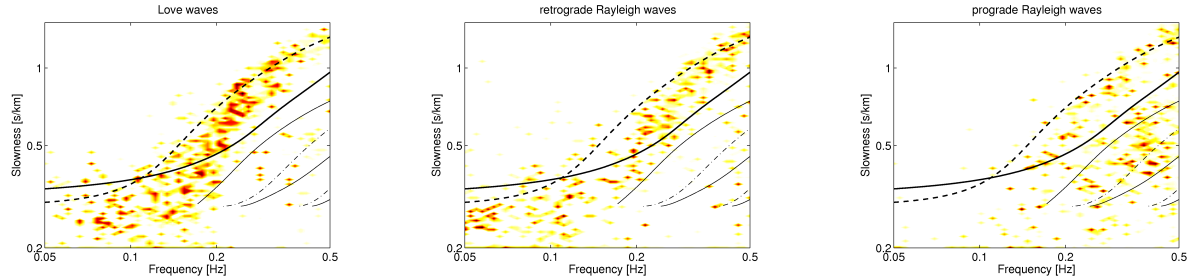


Figure 8.18: Repartition of energy between Love waves, retrograde and prograde Rayleigh waves in different frequency ranges for all events (grouping some events which are close and have similar curves). The order of the events is defined by their azimuth, beginning in the north-west, in clockwise direction. The total energy in every frequency range is normalized to 1.

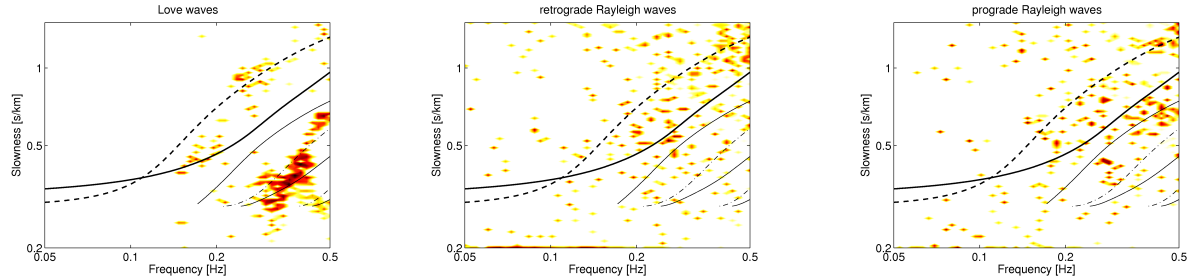
8.7 Dispersion curves

In subsection 8.6, it was shown that the energy repartition between Love and Rayleigh waves is varying from one earthquake to another. Consequently, the different events will yield different parts of the dispersion curves. We will not show the dispersion curves for all events, but limit us to showing some extreme cases. The most energetic event, event 5, is also the one with the second highest proportion of Love waves (see table 8.2). The dispersion curves for this event are shown in figure 8.19 (a) and compared to the theoretical dispersion curves for the basin model. The Love wave dispersion curve of the fundamental mode is relatively well retrieved. For the retrograde and prograde Rayleigh waves, however, the data are not easy to interpret. The retrograde Rayleigh wave curve seems to follow the same pattern as the Love wave data and might as well be composed of misidentified Love waves. Higher modes are not found.
The two series of events 13-16 and 18-21 exhibit a large Love wave proportion for the frequency range between 0.3 and 0.4 Hz, which is not very pronounced for the other events. In figure 8.19 (b), these data points form a dispersion curve which is actually very close to the theoretical dispersion curve for the first harmonic mode of Love waves. In the data for other events, the first higher Love wave mode cannot be seen. In the data of events 13-16 and 18-21, neither a retrograde nor a prograde dispersion curve can be identified.
The dispersion curves for the event with the largest proportion of prograde Rayleigh waves, event 3, are shown in figure 8.19 (c). The dispersion curve for the fundamental mode of Love waves is retrieved in this case, higher modes cannot be seen. The data points for the retrograde dispersion curve cannot be attributed to a dispersion curve. For prograde Rayleigh waves, some clear data points can be identified between 0.2 and 0.5 Hz which are very close to the first harmonic mode's dispersion curve for the basin model.
The slowness values which were calculated for all events separately are combined to a single dispersion curve for each wave type in figure 8.19 (d). This curve has been obtained by summing the data for the single events which have been normalized by their respective total energies. In this way, the information of less energetic events actually contributes to the results. The dispersion curve of the fundamental Love wave mode can be retrieved between 0.15 and 0.5 Hz, the first harmonic mode is clearly seen between 0.3 and 0.5 Hz. Both modes are shifted to lower slowness values at their respective low frequency parts compared to the basin model. This indicates that the velocities of the deeper layers are underestimated in the basin model. For retrograde Rayleigh waves, some data points are seen between 0.2 and 0.5 Hz. As their values are very close to the measured Love wave dispersion curve, it is not clear if these are well-identified Rayleigh waves or misidentified Love waves. For the prograde Rayleigh waves, clear data can only be seen above 0.2 Hz. The large spot identified at 0.4 Hz could well represent the first harmonic Rayleigh wave mode and is clearly separated from the first harmonic Love wave mode. For the retrograde Rayleigh waves, no clear curve can be retrieved. This is certainly due to the fact that the wave field is dominated by Love waves.

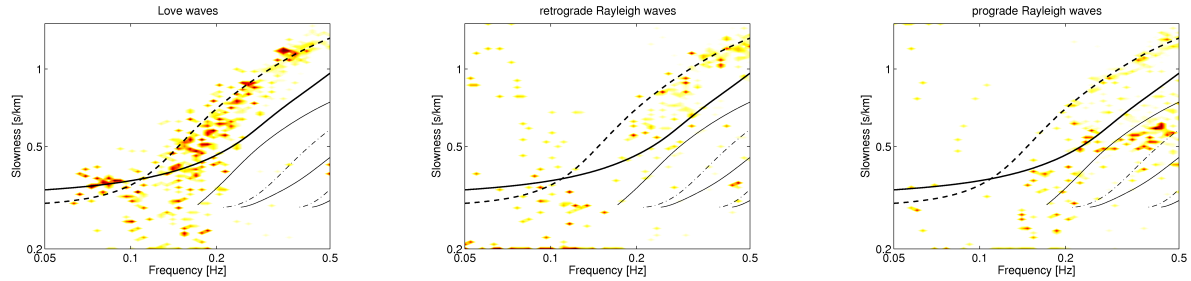
(a) Event 5



(b) Events 13-16 and 18-21



(c) Event 3



(d) Combining all events

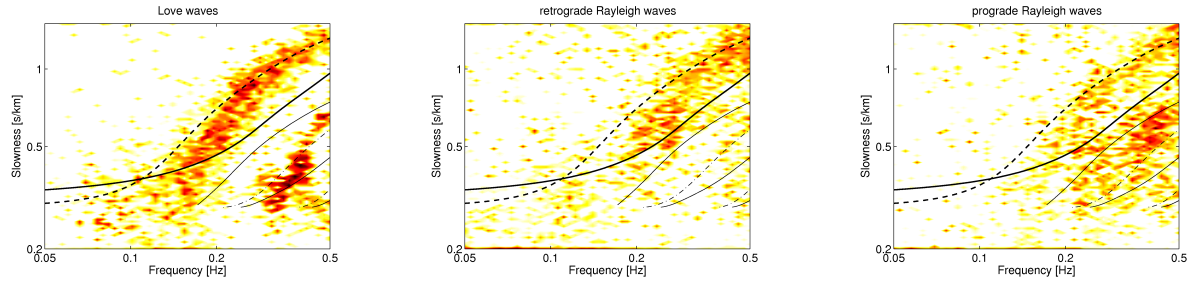


Figure 8.19: Measured slowness values for Love waves (left), retrograde Rayleigh waves (center) and prograde Rayleigh waves (right). The slowness values for all measured time windows attributed to the respective wave type are included in the figure, weighted by their respective energy on the components corresponding to the wave type. The darkest colors indicate high data concentrations. The solid black lines represent the fundamental and harmonic Rayleigh wave dispersion curves corresponding to the basin model for the station indicated in figure 8.2 (a), the dashed and dash-dotted lines to the fundamental and harmonic Love wave dispersion curves. The figures show the results for: (a) event 5, (b) events 13-16 and 18-21, (c) event 3 and (d) the combination of all events.

An alternative way of analyzing the data without taking Love waves into account is to use the vertical component of the signals only. However, the polarization characteristics of the waves, i.e. the retrograde or prograde sense of motion as well as the ellipticity, cannot be retrieved in this case. Figure 8.20 (a) shows the combination of the retrograde and prograde Rayleigh wave dispersion curves of figure 8.19 (d). The resulting dispersion curve which is obtained by averaging the results of analyzing the vertical component of the signal only for all events is shown in figure 8.20 (b). The data form a large cloud of points which spans the theoretical dispersion curves for the basin model's fundamental and first harmonic Rayleigh wave modes. It cannot be determined to which mode the data points correspond or if they correspond to a mixture of both modes. However, the cloud of points is clearly distinguished from the Love wave dispersion curve measured in figure 8.19 (d). This suggests that the retrograde Rayleigh wave dispersion curve of figure 8.19 (d) is actually biased by Love waves. Nevertheless, the prograde Rayleigh wave of figure 8.19 (d) is not retrieved in figure 8.20 (b) either. It might be masked by the large cloud of points or be biased to lower slowness values by the harmonic Love wave mode.

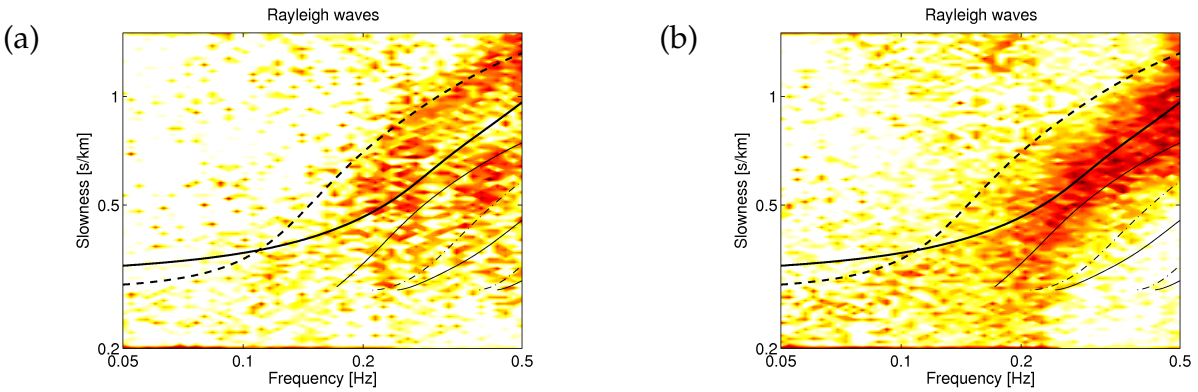


Figure 8.20: Measured slowness values for both Rayleigh wave types, averaged for all events. The slowness values for all measured time windows attributed to Rayleigh waves are included in the figure, weighted by their respective energy on the radial and vertical components. The darkest colors indicate high data concentrations. The solid black lines represent the fundamental and harmonic Rayleigh wave dispersion curves corresponding to the basin model for the station indicated in figure 8.2 (a), the dashed and dash-dotted lines to the fundamental and harmonic Love wave dispersion curves. (a) shows the combination of the results for retrograde and prograde Rayleigh waves shown in figure 8.19 (d). (b) was obtained analyzing the vertical components only and thus eliminating any Love wave contributions.

8.8 Ellipticity curves

The ellipticity curves estimated by MUSIQUE using the results of all events together are shown in figure 8.21 for retrograde and prograde Rayleigh waves. The retrograde ellipticity curve is badly estimated and no ellipticity curve can be identified. For the prograde Rayleigh waves, a scattered ellipticity curve is retrieved, which can be attributed to the first harmonic mode. Below about 0.18 Hz, very few prograde waves are identified compared to the higher frequency part. According to the basin model, the first harmonic mode does not exist below 0.17 Hz. Actually, the retrieved ellipticity values are in good agreement with the first harmonic mode. This is a further indication that the fundamental mode does not exhibit a singularity. If such a singularity existed, the motion of the fundamental mode Rayleigh waves would be prograde for the frequencies of the right flank of the peak, which would certainly bias the ellipticity estimation of the first harmonic mode.

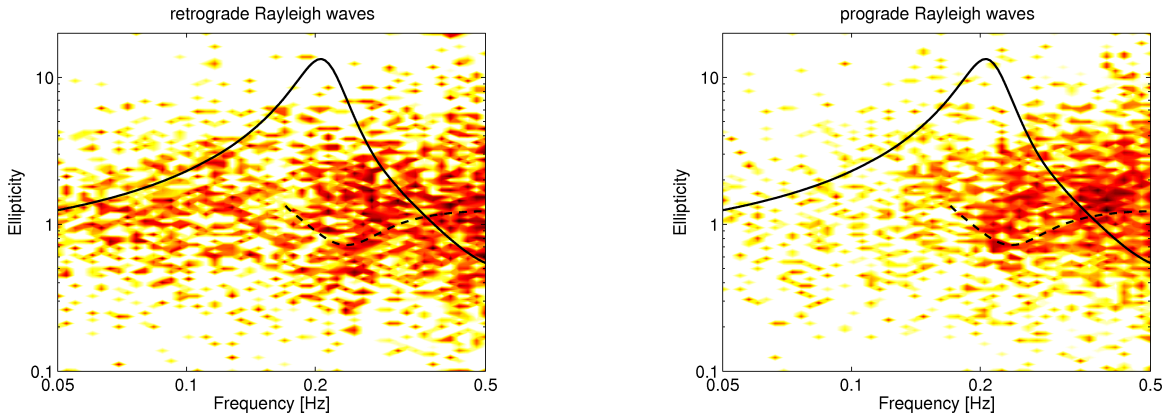


Figure 8.21: Measured ellipticity values for retrograde Rayleigh waves (left) and prograde Rayleigh waves (right) for considering all events together. The ellipticity values for all measured time windows attributed to the respective wave type are included in the figure, weighted by their respective amplitude. The darkest colors indicate high data concentrations. The solid black line represents the ellipticity curve of the fundamental mode and the dashed line of the first harmonic mode of Rayleigh waves corresponding to the basin model for the station indicated in figure 8.2 (a).

As it is based on statistics, a direct application of RayDec on earthquake signals would not make sense because large energy arrivals which do not consist of Rayleigh waves could never be compensated. However, for event 5, the recording length of 22 minutes allows the application of the method to the late signal (see figure 8.5). We used the last seven minutes of signal (starting 900 seconds after the signal beginning). The results of applying RayDec, DELFI and H/V to this signal are shown in figure 8.22. The ellipticity curve estimations are rather instable, showing many little peaks for all methods. Nevertheless, the results suggest lower peak frequencies than predicted by the basin model. This model predicts a value of 0.20 Hz, whereas RayDec and DELFI find a value of 0.18 Hz and H/V a value of 0.16 Hz. However, the signal is still quite short for an analysis at such low frequencies and the signal energy is still decreasing over the whole signal duration. At least RayDec supposes the signal amplitude to be quite stationary in time, which is not the case here. Furthermore, the presence of the first harmonic Rayleigh wave mode above 0.17 Hz is supposed to bias the estimation of the fundamental mode's ellipticity. The three methods used here, however, are not able to distinguish between retrograde and prograde waves or different modes and will yield a mixed estimation if more than a single mode are present.

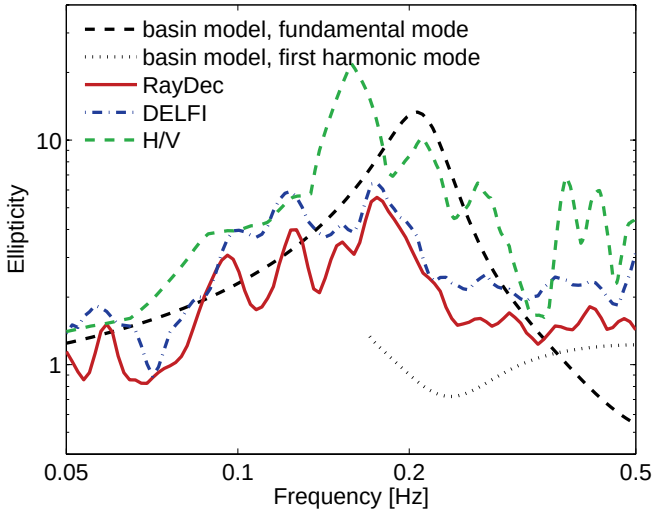


Figure 8.22: Ellipticity curves measured on the late signal of event 5 (seconds 900 to 1317) using RayDec, DELFI and H/V.

8.9 Conclusion

This study investigated the seismological wave field generated by regional earthquakes in the Evergreen Basin, California, using MUSIQUE. In addition to the determination of wave azimuth and velocity, the technique allows the identification of Love, retrograde and prograde Rayleigh waves. The basin response to an earthquake depends on the location of the event. The azimuthal energy distribution is clearly different for waves incident from north-western or south-eastern directions. In both cases, the energy at the array is almost exclusively arriving from the respective hemicycle, with an important exception. A large part of the Love wave energy arrives from the south, even for northern earthquakes. The extreme case is event 7, for which over 75 % of Love waves arrive from the south, i.e. an azimuth which is about 120° different from the direct connecting line. Frankel et al. [2001] already found evidence for a large Love wave scatterer located south of the array and attributed it either to a rock formation or to the southern basin edge. Apart from these southern Love waves, Love waves arrive under azimuths which are slightly deviated from the theoretical earthquake azimuths. The deviations are likely to be explained by refraction of incident waves at the lateral borders of the basin. In general, we find that Rayleigh waves are more scattered and arrive under a multitude of different azimuths. As a detailed model of the San Francisco Bay Area exists, numerical simulations could help to better understand the propagation patterns and to identify the most important wave scatterers.

By analyzing the energy partition between Love and Rayleigh waves, we found large differences between the different earthquakes. The attributed Love wave energy ranges from about 5 to 30 %, whereas the retrograde Rayleigh wave energy is quite stable, taking values between 7 and 13 %. Between 5 and 25 % of the energy have been attributed to prograde Rayleigh waves, the total attributed energy constituting about 40 % of the total signal energy for all events. We identify prograde Rayleigh waves as the first harmonic mode. For the events located the closest to the array (about 40 km), the first harmonic Love wave mode can be found. For the other events, no such waves of significant amplitude can be identified. They might be produced as well during these earthquakes, but attenuated on their way to the array or this might be a feature attributed to the special location of these earthquakes.

By combining the information of all earthquakes, we succeeded in retrieving the Love wave dispersion curves for the fundamental and the first higher mode. The dispersion curve found for retrograde Rayleigh waves is very close to the measured Love wave dispersion curve and could be biased by Love waves as well, particularly because the curve is completely different from the curve obtained by analyzing the vertical components of the signal only. Furthermore, no ellipticity curve is found for these waves. For prograde Rayleigh waves, however, it is possible to identify the dispersion curve and assign an ellipticity curve.

The inter-sensor spacing of the array limits the upper frequency limit of our analysis. Furthermore, the structure of the basin is not homogeneous under the total array size. Therefore, a smaller, but denser array located in the central part of the array would allow the study of the higher-frequency part of incident waves. This might help to increase the understanding of the mechanisms of the Santa Clara Valley.

Acknowledgments
We would like to thank Arthur Frankel and David Carver for kindly providing the data and Arthur Rodgers for supplying the data of the San Francisco Bay Area model and for fruitful discussions.

Conclusion

Ellipticity measurement and inversion using single-station methods

A part of this work focused on the development of new methods allowing the characterization of surface waves. Two methods, DELFI and RayDec, have been designed to estimate the ellipticity of Rayleigh waves from single-sensor measurements.
The first method, DELFI, tries to retrieve this property by fitting ellipses to parts of the signal. If Rayleigh and Love waves are well separated in the signal, the method gives reliable results. However, if both wave types are present in the signal at the same time, the Love wave contributions lead to a misestimation of the ellipticity.
The second method, RayDec, uses statistical properties to suppress other wave types than Rayleigh waves. The technique is especially adapted for the application to ambient seismic vibration measurements.

Ellipticity inversion
The ellipticity curve can be inverted in order to retrieve the local soil structure (i.e. the S-wave velocity profile). However, as the ellipticity function is not unambigious, additional information has to be included in the inversion process, e.g. small-scale SPAC measurements or MASW measurements. The requirements and limits for such inversions have been investigated for theoretical data. The main results are that the right flank of the ellipticity peak and the peak frequency itself carry all important information on the structure. The left flank can be included in the inversion to better fix the peak frequency. The actual velocity values can be fixed both by SPAC or MASW measurements, but the frequency range of this additional information should be close to the frequency range of the ellipticity curve used for inversion. Furthermore, it was shown that the errors on the soil structure model induced by a misestimation of the ellipticity curve are about half the original misestimation. Nevertheless, the soil structure can only be retrieved down to the bedrock depth, as the absolute value of the shear wave velocity in the bedrock is badly constrained.

Application to real data sets
The inversion scheme developed using the theoretical data has then been applied to measurements of ambient seismic vibrations performed during the European NERIES project (2006-2010), where RayDec was used to measure the ellipticity. The results for 14 different well-known sites, for which borehole measurements exist, show that the joint inversion of ellipticity and small-scale SPAC data are in good agreement with broad-band dispersion curve measurements.

No significant differences between the use of SPAC including larger (up to 8) or smaller (down to 3) numbers of sensors could be found. In this way, our method presents an easier and cheaper way for the investigation of the soil structure than using array noise measurements. For some sites, the directly measured dispersion curves are in disagreement with the fundamental Rayleigh wave mode of the models obtained by ellipticity inversion, but in good agreement with the first or second harmonic modes for the same models. Considering the measured dispersion curve alone, however, it is difficult to decide which mode is actually measured. In these cases, an ellipticity inversion can help in identifying the correct mode. Another result of the analysis is that, at least for soft sites, the Vs30 parameter obtained by joint ellipticity inversions is in agreement with borehole measurements and can be helpful for seismic hazard assessment.

Investigation of wave field polarization parameters

MUSIQUE

The third method developed during this thesis work, MUSIQUE, is the combination of two different MUSIC algorithms. The first algorithm, "classical" MUSIC, is a well-established method which has already been successfully applied to seismic measurements. The second algorithm, quaternion-MUSIC, is a rather new development in signal processing and has never been applied to real seismic data before. This code is limited to two-component signals. Certainly, the use of quaternions increases the mathematical and computational complexity of the method, but it allows the simultaneous use of two components and therefore the estimation of the polarization between these two components. This finally extends the field of possible applications of array techniques by discriminating between Love and Rayleigh waves. The additional benefit of using quaternion-MUSIC is the identification of the retrograde or prograde sense of motion and the estimation of ellipticity. In this way, MUSIQUE estimates the contributions of Love and Rayleigh waves to the wave field and characterizes the respective properties.

Biquaternion-MUSIC

A further aspect of the work was linked to biquaternion-MUSIC, an even more complex method which could, in theory, be applied to four-dimensional signals. Therefore, using this method, the three components of the seismic signal could be processed together in order to retrieve their respective polarization properties. Nevertheless, we found that the method cannot work in simple special cases, e.g. a purely two-dimensional signal. Modifications to the code might improve its performances and make the method applicable to complex real data measurements. However, each type of seismic waves is always confined to a maximum of two different directions and can be characterized using quaternion-MUSIC.

Application of MUSIQUE
In tests on synthetic seismic data, MUSIQUE proved its ability to separate not only Love from Rayleigh waves, but also retrograde from prograde Rayleigh waves. Furthermore, the respective ellipticities have been well retrieved.
The application of MUSIQUE to seismological data measured by a dense accelerometer array located in the Santa Clara Valley, California, identified wave types which contributed about 40 % to the total signal energy. The diffracted surface wave field is especially dominated by Love waves generated at the southern valley edge. Combining all earthquakes, the dispersion curves of the fundamental and first harmonic Love wave modes could be retrieved. Although it was possible to identify the dispersion and ellipticity curves of the first harmonic Rayleigh wave mode, the phase velocity estimations of the fundamental Rayleigh wave mode seem to be largely contaminated by Love wave contributions. Nevertheless, this data set appeared not to be the best suited for testing all the potentials of MUSIQUE because the fundamental Rayleigh wave mode does not exhibit a singularity and therefore does not change its sense of motion. In this sense, this site cannot be considered as a school test site for the method. Furthermore, the analysis is restricted to a limited frequency range due to the array geometry. An application of the method to other dense seismic array measurements might fully exploit all of its capabilities. The potential of the method might be fully tested by applying it to simulations of the propagation of seismic waves in sediment-filled basins.

Perspectives

The joint inversion of ellipticity curves as presented in this work is a very promising technique. However, further tests are needed to proof its applicability to seismic hazard assessment at sites for which noise data are already available (e.g. Beirut, alpine valleys).

A possible perspective of analyzing array measurements using MUSIQUE could be the application to other sites, but the number of available array measurements of high quality with a small enough array size is limited. Nevertheless, the method could be further tested on synthetic signals generated for precise basin simulations as the San Francisco Bay Area model, the Euroseistest site in Volvi, Greece, or the Grenoble basin in France. A main objective of such investigations would consist in identifying if a systematic correlation between structure properties as the basin geometry or velocity profiles and the wave type, i.e. Love and Rayleigh wave, exists.

Furthermore, it should be possible to still improve the method so that more than the dominant wave can be identified. Such an algorithm could for example be composed by a first identification of the dominant wave, followed by a subtraction of the respective signal from the recordings. Then, the properties of the second wave could be estimated, again followed by a subtraction of the respective signal and so on. Of course, misestimations in the different algorithm steps would accumulate for the weaker signal parts, but this way of doing should allow the identification of the main wave contributions.

Mesures d'ellipticité à l'aide de méthodes monocapteurs et inversion

Une partie de ce manuscrit était centrée sur le développement de nouvelles méthodes permettant la caractérisation d'ondes de surface. Deux méthodes, DELFI et RayDec, ont été conçues pour estimer l'ellipticité d'ondes de Rayleigh à partir d'enregistrements d'un seul capteur.
La première méthode, DELFI, cherche à retrouver cette propriété en ajustant des ellipses à des parties du signal. Si les ondes de Love et les ondes de Rayleigh sont bien séparées dans le signal, cette méthode donne des résultats fiables. Cependant, si les deux types d'ondes coexistent en même temps, les contributions des ondes de Love engendrent une mauvaise estimation de l'ellipticité.
La deuxième méthode, RayDec, profite de propriétés statistiques pour supprimer les ondes qui ne sont pas des ondes de Rayleigh. La technique est particulièrement adaptée aux enregistrements de vibrations sismiques ambiantes.

Inversion d'ellipticité

Une courbe d'ellipticité peut être inversée afin de retrouver la structure locale du sol (c'est-à-dire le profil de vitesse des ondes de cisaillement). Comme l'ellipticité n'est pas une fonction unique, des informations complémentaires doivent être inclues dans l'inversion, par exemple des mesures SPAC ou MASW pour contraindre les vitesses en surface. Les conditions et limites de ces inversions ont été évaluées pour des données théoriques. Les résultats principaux sont que le flanc droit du pic d'ellipticité et la fréquence du pic elle-même comportent toutes les informations nécessaires sur la structure. Le flanc gauche du pic, quant à lui, peut être utilisé dans l'inversion pour mieux fixer la fréquence du pic. Les valeurs de vitesse réelles peuvent être fixées soit par des données de mesures SPAC, soit par des mesures MASW, mais les bandes de fréquence de ces informations complémentaires doivent être proches des gammes de fréquence de la courbe d'ellipticité inversée. En plus, il a été montré que les erreurs sur la structure du sol induits par des erreurs d'estimation de la courbe d'ellipticité ne représentent qu'à peu près la moitié de l'erreur d'estimation originale. Quoi qu'il en soit, la structure du sol ne peut être retrouvé que jusqu'à la profondeur du bedrock, les valeurs absolues de la vitesse des ondes de cisaillement dans le bedrock étant mal contraintes.

Application aux données réelles

Le schéma d'inversion développé à l'aide des données théoriques a ensuite été appliqué à des mesures de vibrations sismiques ambiantes effectuées dans le cadre du projet européen NERIES (2006-2010), utilisant RayDec pour mesurer les courbes d'ellipticité. Les résultats pour 14 sites bien connus, où des mesures de référence par méthode en forage existent, montrent que les inversions conjointes de courbes d'ellipticité et de courbes SPAC à petite échelle sont en bon accord avec des mesures de dispersion large-bande. En plus, aucune différence significative entre l'utilisation d'un plus grand nombre (jusqu'à 8) ou d'un plus petit nombre (jusqu'à 3) de capteurs pour les mesures SPAC n'a pu être identifiée. Notre méthode d'inversion représente donc une alternative plus simple et moins coûteuse pour estimer la structure en vitesse que des mesures par réseaux. Pour quelques sites, à basse fréquence, la courbe de dispersion mesurée directement est en désaccord avec la courbe pour le mode fondamental des ondes de Rayleigh obtenue par inversion d'ellipticité, mais en bon accord avec le premier ou le second mode harmonique. Dans ces cas, l'inversion de la courbe d'ellipticité peut aider à identifier les différents modes d'ondes de Rayleigh présents dans les courbes de dispersion mesurées. Un autre résultat de l'analyse est que les paramètres Vs30 déterminés par inversion de l'ellipticité sont en bon accord avec les mesures en forage.

Investigation des paramètres de polarisation d'un champ d'ondes

MUSIQUE

La troisième méthode développée dans le cadre du travail présenté, MUSIQUE, représente la combinaison de deux algorithmes distincts de MUSIC. Le premier de ceux-ci, la version «classique » de MUSIC, est une méthode bien établie qui a déjà été appliquée à des mesures sismiques avec succès. Le deuxième algorithme, MUSIC quaternionique, représente un développement plutôt récent en traitement du signal et n'avait jamais été appliqué à des mesures sismiques auparavant. Certainement, l'utilisation de quaternions augmente la complexité mathématique, mais permet d'utiliser deux composantes différentes en même temps et par conséquent l'estimation de la polarisation entre ces deux composantes. Ainsi, les applications possibles de techniques en réseau sont étendues par la discrimination entre ondes de Love et de Rayleigh. Le bénéfice complémentaire de l'utilisation de MUSIC quaternionique est l'identification du sens de rotation retrograde ou prograde du mouvement et l'estimation de l'ellipticité. De cette manière, MUSIQUE évalue les contributions des ondes de Love et de Rayleigh au champ d'ondes et caractérise leurs propriétés.

MUSIC biquaternionique

Un autre aspect de ce travail était lié à MUSIC biquaternionique, une méthode encore plus complexe qui pourrait, en théorie, être appliquée à des signaux à quatre dimensions. Par conséquent, avec cette méthode les trois composantes du signal sismique pourraient être traitées simultanément afin de retrouver les propriétés de polarisation entre elles. Quoi qu'il en soit, nous avons trouvé que cette méthode échouait dans des cas spéciaux simples, par exemple pour des signaux purement bidimensionnels. Le code pourrait éventuellement être amélioré pour l'appliquer à des données réelles complexes. Néanmoins, chaque type d'ondes de surface étant toujours confiné au maximum à deux directions différentes, il peut être caractérisé par MUSIC quaternionique.

Application de MUSIQUE

Dans des tests avec des données sismiques synthétiques, MUSIQUE a confirmé sa capacité de distinguer non seulement entre ondes de Loves et ondes de Rayleigh, mais aussi entre ondes de Rayleigh rétrogrades et prograde. En plus, les courbes d'ellipticité associées ont été bien déterminées. En l'appliquant à des données sismiques mesurées par un réseau d'accéléromètres localisé dans la vallée de Santa Clara en Californie, MUSIQUE a identifié des ondes représentant environ 40 % de l'énergie totale des signaux. Le champ d'ondes de surface diffractées est dominé par des ondes de Love générées aux bords méridionaux du bassin. En combinant tous les séismes analysés, les courbes de dispersion du mode fondamental ainsi que du premier mode harmonique ont été trouvées pour les ondes de Love. Bien que les courbes de dispersion et d'ellipticité du premier mode harmonique des ondes de Rayleigh aient pu être identifiées, les estimations de vitesses de phase pour le mode fondamental semblent être largement contaminées par des contributions d'ondes de Love. Par ailleurs, ce jeu de données n'est probablement pas le meilleur pour tester toutes les capacités de MUSIQUE car le mode fondamental des ondes de Rayleigh n'a pas de singularité et ne change pas de sens de rotation. Dans ce sens, ce site ne représente pas un site d'école pour la méthode. En outre, la géométrie du réseau limite la bande de fréquence qui peut être analysée. L'ensemble des capacités de la méthode pourrait éventuellement être mieux testé en l'applicant à d'autres mesures de réseaux sismiques et à des simulations de la propagation d'ondes sismiques à l'intérieur de bassins sédimentaires.

Perspectives

L'inversion conjointe de courbes d'ellipticité telle que présentée dans ce manuscrit est une technique prometteuse. Néanmoins, des tests plus approfondis sont nécessaires pour démontrer son applicabilité à l'estimation de l'aléa sismique. Ces tests pourraient être faits pour des sites où des données de bruit sismique sont déjà disponibles (p. ex. Beyrouth, vallées alpines).

Une perspective possible d'analyse de données de réseaux par MUSIQUE pourrait être l'application à d'autres sites, avec un nombre d'enregistrements sismologiques de qualité et suffisant et les géométries des réseaux adaptées. Néanmoins, la méthode pourrait être testée sur des signaux synthétiques générés par des simulations précises de bassins sédimentaires telles qu'elles existent pour la région de la baie de San Francisco, le site Euroseistest à Volvi en Grèce ou le bassin de Grenoble en France. Un objectif majeur de ces investigations serait la recherche de corrélations systématiques entre propriétés du bassin telles que la géométrie, les profils de vitesse et la composition du champ d'ondes, c'est-à-dire la recherche d'effets caractéristiques pour les ondes de Love ou de Rayleigh.

Finalement, il devrait être possible d'améliorer la méthode afin que plusieurs ondes puissent être identifiées en même temps. Une réalisation possible d'un tel algorithme comporterait l'identification de l'onde dominante, suivi par une soustraction du signal identifié des enregistrements. Après, les propriétés de la seconde onde pourraient être évaluées, à nouveau suivi par une soustraction du signal identifié et ainsi de suite. Bien sûr, les erreurs d'estimation des différents étapes de l'algorithme s'accumuleraient pour les parties les plus faibles du signal, mais cette démarche devrait au moins permettre l'identification des contributions majeures du champ d'ondes.

Appendix A

EUSIPCO paper

The following article ("Rayleigh wave ellipticity estimation from ambient seismic noise using single and multiple vector-sensor techniques" by Hobiger, Le Bihan, Cornou, and Bard 2009b) was presented at the European Signal Processing Conference (EUSIPCO) 2009 in Glasgow (UK). It gives a concise overview of the methods DELFI (see section 2.3), RayDec (see section 2.4 and MUSIQUE (see section 4.6). It should be noted that the DELFI and MUSIQUE algorithms in the text correspond to early versions of the respective codes.

Le texte suivant a été présenté à l'EUSIPCO (European Signal Processing Conference) 2009 à Glasgow (Royaume-Uni) sous le titre "Rayleigh wave ellipticity estimation from ambient seismic noise using single and multiple vector-sensor techniques", [Hobiger, Le Bihan, Cornou, and Bard, 2009b]. Il donne une vue d'ensemble concise des méthodes DELFI (cf. section 2.3), RayDec (cf. section 2.4) et MUSIQUE (cf section 4.6). Les algorithmes DELFI et MUSIQUE présentés dans le texte suivant représentent des versions antérieures des codes présentés dans les sections respectives.

A.1 Abstract

In this paper, we present three new methods to evaluate the ellipticity of polarized signals generated by seismic ambient vibrations. Classically, the estimation of Rayleigh wave ellipticity is performed using a spectrum ratio approach [Nakamura, 1989]. Here, we propose to use two single-sensor techniques (based on ellipse fitting [Fitzgibbon et al., 1999] and random decrement [Asmussen, 1997]) and one multi-sensor technique (based on quaternion-MUSIC [Miron et al., 2005b]) to substitute the existing algorithm. Rayleigh wave ellipticity is an important geophysical parameter because its estimation as a function of frequency leads to velocity vs. depth estimation, *i.e.* identification of the underground structure. We illustrate the three algorithms on simulation data and compare them to the existing approach.

A.2 Introduction

Geophysical signals carry information about the medium they have propagated through. In seismology it is common use to take advantage of the constant seismic activity to perform passive tomography [Aki, 1957]. The "ambient seismic noise" is indeed built of several types of waves. By analyzing them, conclusions about the medium characteristics can be drawn. In this paper, we are interested in surface waves called Rayleigh waves. In contrast to "classical" noise, these waves are polarized. The particle movement of Rayleigh waves describes an ellipse. The ratio of the horizontal axis to the vertical axis of the ellipse is called ellipticity. It is a function of frequency. In this paper, we focus on the estimation of the frequency dependency of Rayleigh wave ellipticity from passive recordings of ambient ground vibrations. In order to record polarization parameters, it is necessary to use vector-sensors whose outputs consist of three signals measuring displacements in three orthogonal directions of space.
In seismology, a standard way to access ellipticity is to use the spectral ratio between horizontal and vertical displacement measurements [Bard, 1998]. This way of doing has advantages in terms of simplicity but may not be very accurate as other wave types than Rayleigh waves also contribute to the wave field [Bonnefoy-Claudet et al., 2008]. The P- and S-wave velocity profile (velocity as a function of depth) of the underground structure can be obtained by inversion of the ellipticity curve (ellipticity ε as a function of frequency ν, [Boore and Toksöz, 1969, Arai and Tokimatsu, 2004]). Therefore, errors in the ellipticity curve lead to inaccurate estimation of the velocity profile of the underground structure.
In order to improve the estimation of ellipticity, we propose three different new methods: DELFI (Direct ellipse fitting for Rayleigh wave ellipticity estimation), RayDec (Rayleigh wave ellipticity estimation by using the random decrement technique) and MUSIQUE (MUSIC using quaternions for Rayleigh wave ellipticity estimation). DELFI and RayDec use single vector-sensors, MUSIQUE takes an array of multiple vector-sensors into account. The proposed techniques are compared and their limitations are given.

A.3 Rayleigh wave ellipticity in seismology

First, we introduce the term of Rayleigh wave ellipticity as understood in seismology. After briefly recalling the underlying physics, we present the signals we will use to determine the ellipticity.

A.3.1 Surface waves and underground structure

Four main wave types are important in seismology. P- and S-waves represent pressure and shear body waves, respectively. Rayleigh and Love waves are surface waves, *i.e.* they propagate along the earth's surface. Their energy decreases as $1/r$ compared to $1/r^2$ for body waves where r is the distance to an earthquake hypocenter. Therefore surface waves are important for seismic hazard assessment.
Love waves are transverse waves which are confined to the surface layer. Rayleigh waves, which are also called "ground roll", are waves with a vertical and a horizontal component. The movement of a particle under the influence of a Rayleigh wave describes an ellipse in the plane formed by the vertical axis and the wave propagation direction. At the surface of homogeneous halfspaces, the movement of the fundamental mode is retrograde, *i.e.* the horizontal movement advances the vertical movement by $\pi/2$ in phase. Prograde movement corresponds to a phase difference of $-\pi/2$.
The underground structure of the earth's surface is in general not homogeneous but consists of different sedimentary layers superposing the bedrock. Every layer has different mechanical properties, *e.g.* P- and S-wave velocities. The properties of Rayleigh waves are closely related to the underground structure. An important polarization parameter of Rayleigh waves is the ellipticity ε. It is a function of frequency ν. Figure A.1 shows a realistic model of the P- and S-wave velocity profile of an underground structure and the corresponding $\varepsilon(\nu)$-curve. Inversion techniques exist allowing the underground P- and S-wave velocity profiles to be derived from the $\varepsilon(\nu)$-curve of the Rayleigh waves [Arai and Tokimatsu, 2004].
The knowledge of the soil structure of a given site is important in order to estimate the ground motion and the associated risk for buildings in the case of an earthquake.

A.3.2 Sensors and arrays

In seismology, the earth's movements are recorded using vector-sensors called geophones. Such sensors are made of three co-located sensors recording the ground motion in three orthogonal directions. As a result, the output signal is a 3D vector-valued time-series. By convention, the three sensors are aligned in north-south, east-west and vertical direction. A number of vector-sensors can be arranged to form an array. Thus, this configuration enables the measurement of the azimuth and velocity of incoming waves. In the following, the east-west signal of the vector-sensor will be denoted by $x(t)$, the north-south signal by $y(t)$ and the vertical signal by $z(t)$.

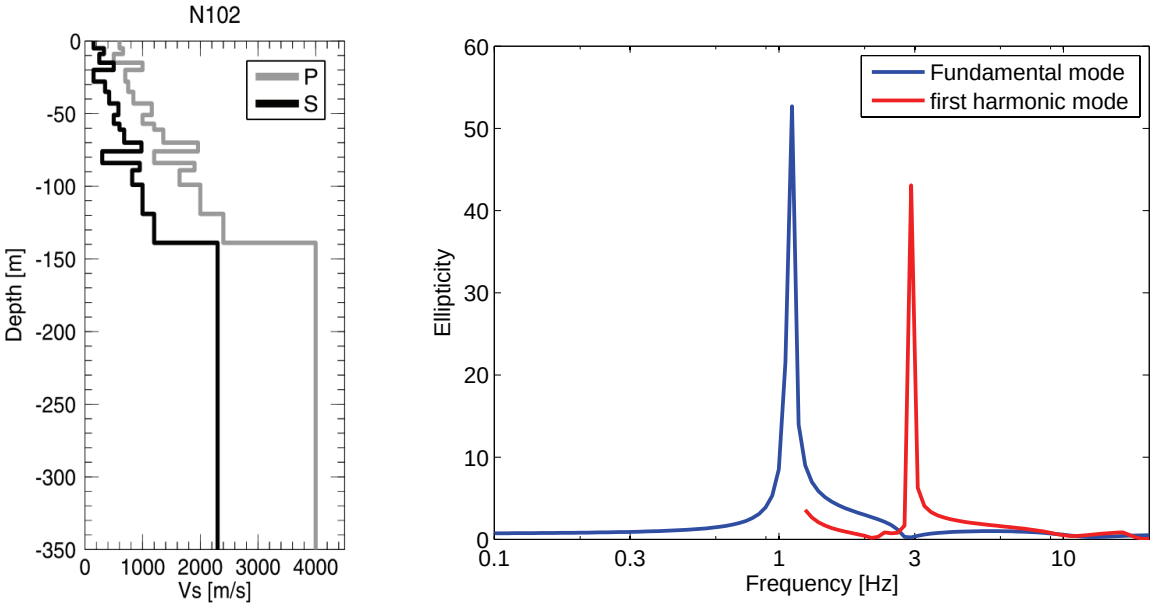


Figure A.1: On the left, the used P- and S-wave velocity profile model. On the right, the theoretical frequency-ellipticity curve corresponding to this model for the fundamental and the first harmonic Rayleigh wave mode.

A.4 Existing technique: H/V spectrum ratios

A widespread approach to estimate ellipticity ε as a function of frequency ν for Rayleigh waves is the so-called H/V technique [Nakamura, 1989]. It is a single-sensor technique which consists of calculating the direct ratio between the horizontal and the vertical spectrum:

$$\varepsilon_{H/V}(\nu) = \frac{\sqrt{|X(\nu)|^2 + |Y(\nu)|^2}}{\sqrt{2}\,|Z(\nu)|}. \tag{A.1}$$

Note that for some frequencies, $|Z(\nu)|$ may vanish, which corresponds to singular points in the $\varepsilon(\nu)$ curve. In addition to Rayleigh waves, Love and body waves may also be recorded. As the movement of Love waves is restricted to the horizontal plane, the H/V approach systematically over-estimates ellipticity. Using assumptions on the proportion of Rayleigh and Love waves present in the wave field, it is possible to correct the over-estimation errors. However, there is no guarantee that the assumptions are in accordance with reality. This is a major source of error in the H/V approach. This motivates the search for other techniques for the estimation of $\varepsilon(\nu)$.

A.5 Single-sensor approach

In this section, we present two approaches to perform ellipticity estimation using the signal recorded on a single three-component vector-sensor. Their performances will be illustrated in Section A.7.

A.5.1 Direct ellipse fitting (DELFI)

Consider the vector-valued signal $\vec{\mathbf{s}}(t)$ describing the movement of a soil particle in 3D space with time, given as $\vec{\mathbf{s}}(t) = [x(t)\, y(t)\, z(t)]^T$ taking values in $\mathbb{R}^3$. First, $\vec{\mathbf{s}}(t)$ is band-pass filtered around a frequency ν with bandwidth $d\nu$. Then, the following processing is performed independently for each frequency.
The filtered signal is cut in blocks of time length $T = 1/\nu$. The total length of the signal is assumed to be much longer than T. Each time-window consists thus of $N = T/dt$ samples where dt is the sampling period. The three-component signal of one time-window is noted $\mathbf{S}_w = [X_w\ Y_w\ Z_w]^T$ and is of dimension $3 \times N$. The polarization of Rayleigh waves is confined to a vertical plane, in opposition to the movement of noise signals which lie in 3D space. In order to project the signal $\mathbf{S}_w$ onto a vertical plane, we use the covariance matrix $\mathbf{C}_w$ built by the two horizontal components by $\mathbf{C}_w = [X_w\ Y_w]^T[X_w\ Y_w]$.
The Eigenvalue Decomposition (EVD) of $\mathbf{C}_w$ gives access to its eigenvalues λ_1 and λ_2 (ranked in increasing order, *i.e.* $\lambda_1 < \lambda_2$) and associated eigenvectors $\vec{\mathbf{u}}_1$ and $\vec{\mathbf{u}}_2$. The eigenvectors are normal vectors to two orthogonal vertical planes. The associated eigenvalues are the sums of the squared distances between the data points and the respective planes.
The 2D plane where the signal $\vec{\mathbf{s}}_w(t)$ is mostly confined is the one orthogonal to $\vec{\mathbf{u}}_1$ as it is associated to the smallest eigenvalue λ_1, *i.e.* the corresponding plane is spanned by the eigenvector $\vec{\mathbf{u}}_2$ and the vertical unit vector. The horizontal components of the time-window signal $\mathbf{S}_w$ are then projected onto an axis to build a horizontal signal H_w by $H_w = [X_w\ Y_w] \cdot \vec{\mathbf{u}}_2$. The 2D signal is then built by $\mathbf{S}'_w = [H_w\ Z_w]^T$.
Finally, the ellipticity measurement is performed using a modified version of the ellipse fitting algorithm proposed by *Fitzgibbon et al.* [Fitzgibbon et al., 1999]. The modifications of the algorithm ensure that the axes of the ellipse are in horizontal and vertical directions. The algorithm yields the horizontal axis h_w, the vertical axis v_w and a parameter d_w^2 indicating the actual sum of squared distances between the data points and the fitted ellipse.
The ellipticity $\varepsilon(\nu)$ is calculated by averaging the found horizontal and vertical axes of all signal blocks for the corresponding frequency using the distance parameter d_w^2 as weighting factor:

$$\varepsilon_{DELFI}(\nu) = \frac{\sum h_w/d_w^2}{\sum v_w/d_w^2}. \tag{A.2}$$

The only free parameter of the method is $d\nu$.

A.5.2 Random decrement (RayDec)

The random decrement technique is commonly used to characterize dynamic parameters of buildings [Dunand, 2005, Michel et al., 2008]. It is classically applied to one-component (scalar valued) signals. We propose a three-component version of this technique in order to estimate $\varepsilon(\nu)$.
The following processing is performed independently for each frequency. First, the three components of $\vec{\mathbf{s}}(t)$ are filtered in a frequency range of bandwidth $d\nu$ around a central frequency ν. The filtered components are called $x_\nu(t)$, $y_\nu(t)$ and $z_\nu(t)$.
The next step is to search for the times τ_i where the signal sign changes from negative to positive on the vertical component ($z_\nu(\tau_i) \leq 0$, $z_\nu(\tau_i + dt) > 0$). For each of these τ_i a signal block of length Δ is stored on all three components. The number of buffered blocks thus depends on the analyzed signal itself. In order to compensate the natural $\pi/2$ phase shift, *i.e.* $1/(4\nu)$ in time, between the vertical and horizontal component signals of a Rayleigh wave, a phase shift is applied to x_ν and y_ν. The buffered signals then are:

$$\left\{ \begin{array}{rcl} z_{\nu,b,i}(t) & = & z_\nu(\tau_i + t), \\ x_{\nu,b,i}(t) & = & x_\nu(\tau_i - \frac{1}{4\nu} + t), \\ y_{\nu,b,i}(t) & = & y_\nu(\tau_i - \frac{1}{4\nu} + t). \end{array} \right. \tag{A.3}$$

for $0 \leq t \leq \Delta$. Index i specifies the change of sign position while index b stands for *buffered*. The east-west and north-south components, namely $x_{\nu,b,i}(t)$ and $y_{\nu,b,i}(t)$, of every buffered signal $\vec{\mathbf{s}}_{\nu,b,i}(t)$ are then projected onto a single axis to build a *horizontal component* in the following way:

$$h_{\nu,b,i}(t) = \sin\vartheta_i \cdot x_{\nu,b,i}(t) + \cos\vartheta_i \cdot y_{\nu,b,i}(t). \tag{A.4}$$

The azimuth angle ϑ_i is chosen to maximize the correlation $C(\vartheta_i) = \int_0^\Delta z_{\nu,b,i}(t) \cdot h_{\nu,b,i}(t)dt$ between the vertical signal $z_{\nu,b,i}(t)$ and the horizontal signal $h_{\nu,b,i}(t)$. The azimuth angle ϑ_i is thus obtained by

$$\vartheta_i = \tan^{-1}\left(\frac{\int_0^\Delta z_{\nu,b,i}(t) \cdot x_{\nu,b,i}(t)dt}{\int_0^\Delta z_{\nu,b,i}(t) \cdot y_{\nu,b,i}(t)dt} \right). \tag{A.5}$$

From the two possible $\vartheta_i \in [0°, 360°[$ satisfying equation (A.5) we choose the one leading to a positive correlation. At this point, the method does not distinguish between prograde and retrograde particle motion as prograde Rayleigh waves are treated as retrograde waves arriving from the opposite azimuth.
Although the angle ϑ_i maximizes the correlation, $z_{\nu,b,i}(t)$ and $h_{\nu,b,i}(t)$ can still be weakly correlated. Therefore, a correlation factor defined by

$$c_{\nu,b,i} = \frac{\int_0^\Delta z_{\nu,b,i}(t) \cdot h_{\nu,b,i}(t)dt}{\sqrt{\int_0^\Delta z^2_{\nu,b,i}(t)dt \cdot \int_0^\Delta h^2_{\nu,b,i}(t)dt}} \tag{A.6}$$

is used as weighting factor in the summation of the signals $z_{\nu,b,i}(t)$ and $h_{\nu,b,i}(t)$:

$$z_{\nu,s}(t) = \sum_i c^2_{\nu,b,i} \cdot z_{\nu,b,i}(t) \tag{A.7}$$

$$h_{\nu,s}(t) = \sum_i c^2_{\nu,b,i} \cdot h_{\nu,b,i}(t). \tag{A.8}$$

Finally, the ellipticity is obtained by

$$\varepsilon_{RayDec}(\nu) = \sqrt{\frac{\int_0^\Delta h^2_{\nu,s}(t)dt}{\int_0^\Delta z^2_{\nu,s}(t)dt}}. \tag{A.9}$$

Repeated execution of this algorithm over the whole frequency range yields the frequency-ellipticity curve. The length of the buffered signal Δ and the width of the frequency filter $d\nu$ can be chosen arbitrarily.

A.6 Multiple-sensor approach (MUSIQUE)

In this section, we propose a way to take advantage of array measurements. The proposed approach is based on MUSIC algorithm and uses two of its versions: "classical" MUSIC to perform wave identification and "quaternion" MUSIC for ellipticity estimation. The array consists of N_S three-component sensors. The N_S time sample records are cut into blocks of $N_T = n_P/(\nu\ dt)$ data points each, where n_P is the number of periods corresponding to the frequency ν in the block and dt the sampling period. The cut time samples are stored in three $N_S \times N_T$ data matrices $\mathbf{S}_i(t)$ with $i = 1$ for *vertical*, $i = 2$ for $east - west$ and $i = 3$ for $north - south$ component. Fourier transform of these three data matrices yields the three matrices $\mathbf{S}_i(\nu)$ of dimensions $N_S \times N_T$.

A.6.1 First step: Azimuth estimation

The first step of our approach is to perform a "classical" MUSIC to identify the azimuth and velocity of an incoming wave. Note that this is a narrow-band MUSIC of bandwidth $d\nu$. Therefore, the *spectral matrix* is computed. Each one-component *spectral matrix* $\mathbf{M}_i(\nu) \in \mathbb{C}^{N_S \times N_S}$ is given by $\mathbf{M}_i(\nu) = \mathbf{S}_i(\nu)\mathbf{S}_i^\dagger(\nu)$. Then, the *total* sample covariance matrix $\mathbf{M}(\nu) \in \mathbb{C}^{N_S \times N_S}$ is obtained by $\mathbf{M}(\nu) = \sum_{i=1}^3 \mathbf{M}_i(\nu)$.
The N_{big} biggest eigenvalues of the matrix $\mathbf{M}(\nu)$ are identified to belong to the signal subspace and the other $(N_S - N_{big})$ to the noise subspace. It should be noted that due to the later performed projection into a plane the technique can only yield reliable results if the wave field is dominated by one wave and therefore $N_{big} = 1$. The noise subspace $\mathbf{G}(\nu) \in \mathbb{C}^{N_S \times (N_S - N_{big})}$ is hence built using the associated eigenvectors.
An incident Rayleigh wave with azimuth ϑ and phase velocity c is assumed. The associated wave vector is thus $\vec{\kappa}(\vartheta) = [-\sin\vartheta\ -\cos\vartheta\ 0]^T$ and the *time-delay vector* $\vec{\delta}(\nu, \vartheta, c) \in \mathbb{R}^{N_S \times 1}$ is given by

$$\vec{\delta}(\nu, \vartheta, c) = \frac{2\pi\nu}{c}\mathbf{R}\vec{\kappa}(\vartheta), \tag{A.10}$$

where $\mathbf{R} \in \mathbb{R}^{N_S \times 3}$ is the matrix of the positions of the array sensors and c the wave velocity. This leads to the steering vector $\vec{\mathbf{fl}}(\nu, \vartheta, c) = \exp(-i\vec{\delta}(\nu, \vartheta, c))/\sqrt{N_S}$. The MUSIC functional $F(\nu, \vartheta, c)$ is finally given by:

$$F(\nu,\vartheta,c) = \frac{1}{\vec{\gamma}^{\dagger}(\nu,\vartheta,c)\mathbf{G}(\nu)\mathbf{G}^{\dagger}(\nu)\vec{\gamma}(\nu,\vartheta,c)}. \tag{A.11}$$

A grid search is performed in order to find the set of parameters (ϑ_0, c_0) maximizing this functional.
Now, as the movement of Rayleigh waves is confined to a plane spanned by the wave vector and the vertical axis, the original signals are projected into the vertical plane spanned by the wave vector just found. This means that the original vertical signal is left unchanged and the two horizontal signals are rotated using ϑ_0. The new signal $\mathbf{S}_p(t)$ consists of two matrices, $\mathbf{S}_{p1}(t)$ and $\mathbf{S}_{p2}(t)$, as it is confined to a 2D plane (instead of 3 components). It is obtained from the original signal by:

$$\mathbf{S}_{p1}(t) = \mathbf{S}_1(t) \tag{A.12}$$

$$\mathbf{S}_{p2}(t) = -\sin\vartheta_0 \cdot \mathbf{S}_2(t) - \cos\vartheta_0 \cdot \mathbf{S}_3(t). \tag{A.13}$$

Fourier transforms of $\mathbf{S}_{p1}(t)$ and $\mathbf{S}_{p2}(t)$ give the matrices $\mathbf{S}_{p1}(\nu)$ and $\mathbf{S}_{p2}(\nu)$. Now that the signals are confined in the propagation plane of the Rayleigh wave, we can proceed to the estimation of ellipticity.

A.6.2 Second step: Ellipticity estimation

The second step of our approach consists in using a quaternion MUSIC algorithm to estimate the polarization parameters of the identified wave. Miron et al. [2006] demonstrated its ability to perform polarization estimation. Therefore, we first build a quaternion valued data matrix $\mathbf{S}_q(\nu) \in \mathbb{H}^{N_S \times N_T}$ in the frequency domain [Miron et al., 2006]:

$$\mathbf{S}_q(\nu) = \Re(\mathbf{S}_{p1}(\nu)) + \mathtt{i} \cdot \Re(\mathbf{S}_{p2}(\nu)) + \mathtt{j} \cdot \Im(\mathbf{S}_{p1}(\nu)) + \mathtt{k} \cdot \Im(\mathbf{S}_{p2}(\nu)), \tag{A.14}$$

where `i`, `j` and `k` are the three quaternion imaginary units [Ward, 1997]. The quaternion valued spectral matrix $\mathbf{M}_q(\nu)$ is given by $\mathbf{M}_q(\nu) = \mathbf{S}_q(\nu)\ \mathbf{S}_q^{\dagger}(\nu)$. Here again, the noise subspace $\mathbf{G}_q(\nu)$ is built by the $(N_S - N_{big})$ eigenvectors corresponding to the smallest eigenvalues, where N_{big} is the same number of sources as defined in the first step. In agreement with Miron et al. [2006], the quaternion steering vector is given by:

$$\vec{\gamma}_q(\nu,\rho,\varphi) = [\cos(\rho) + \mathtt{i}\sin(\rho)e^{\mathtt{j}\varphi}]e^{-\mathtt{j}\vec{\delta}(\nu,\vartheta_0,c_0)}/\sqrt{N_S} \tag{A.15}$$

where ρ and φ are the polarization parameters of the wave and $\vec{\delta}(\nu, \vartheta_0, c_0)$ is the array delay vector given by equation (A.10) with $\vartheta = \vartheta_0$ and $c = c_0$. The ellipticity ε is the ratio of the amplitudes of horizontal and vertical motion, *i.e.*

$$\varepsilon_{MUSIQUE} = \sin\rho/\cos\rho = \tan\rho. \tag{A.16}$$

φ is the phase difference between the horizontal and the vertical signal. Note that we omitted the ν dependency of ρ and φ in eq. (A.15), but it should not be forgotten that their values depend on ν. A grid search yields the set of parameters (ρ_0, φ_0) that maximizes the quaternion MUSIC functional

$$\mathcal{F}_q(\nu,\rho,\varphi) = \frac{1}{\vec{\gamma}_q^{\dagger}(\nu,\rho,\varphi)\mathbf{G}_q(\nu)\mathbf{G}_q^{\dagger}(\nu)\vec{\gamma}_q(\nu,\rho,\varphi))} \tag{A.17}$$

Instead of using the parameter ρ to describe the ellipticity ϵ of the wave, it is possible to use ellipticity itself. However, as ellipticity values are in principle not limited, one would have to perform the parameter search for $0 \leq \varepsilon \leq \infty$. Therefore, it is advantageous to limit the search to $0 \leq \rho \leq \pi/2$ and to define ellipticity by equation (A.16). The phase difference φ is searched in the intervall $[0, 2\pi[$. For a retrograde Rayleigh wave, the phase difference will be found around $\pi/2$, for a prograde Rayleigh wave it will be about $3\pi/2$. The preceding processing is performed for every signal block of length N_T. For each frequency ν, the ellipticities of the signal blocks are averaged using the reciprocal value of the difference of φ from the the theoretical phase difference ($\pi/2$ for retrograde, $3\pi/2$ for prograde) and the signal energy of the signal block as weighting factors. In this way, the ellipticity curve $\varepsilon(\nu)$ is estimated.
The performances of the proposed multiple-sensor method are presented in the next section.

A.7 Simulation

In order to show the performances of the three proposed algorithms, we applied them to a set of synthetic seismic ambient vibrations. Model N102 of an international blind test benchmark [Cornou et al., 2009] was used as underground structure. The P- and S-wave velocity profiles of the model, as well as the associated theoretical frequency-ellipticity curve, are shown in figure A.1. 620 seconds of synthetic ambient vibrations were simulated as follows. Signal sources were modeled by forces with random orientation and amplitude located 0.5 m below the surface [Moczo and Kristek, 2002]. The forces were distributed randomly in time and space. The time function of each source was a delta-like signal with a flat Fourier frequency-amplitude spectrum between 0.1 and 20 Hz. The associated wave field was computed by using the wavenumber-based technique of Hisada [Hisada, 1994, 1995] and contains all body and surface wave types. As the goal of this investigation is the Rayleigh wave ellipticity as a function of frequency, Love and body waves can be considered as "noise". However, the signal-to-noise ratio is unknown and the robustness of the methods under different signal-to-noise ratios cannot be investigated. As the signal generation mimics the generation of real seismic ambient vibrations this case can though be considered realistic.
The recordings were simulated for sixty seismic sensors. Their spatial distribution is shown in figure A.2. The noise sources distribution covered all possible azimuths and distances up to around 1200 m from the origin. Therefore, every seismic sensor represents a different set of data. As the underground structure is the same at every point of the model, the theoretical ellipticity is the same for each sensor.

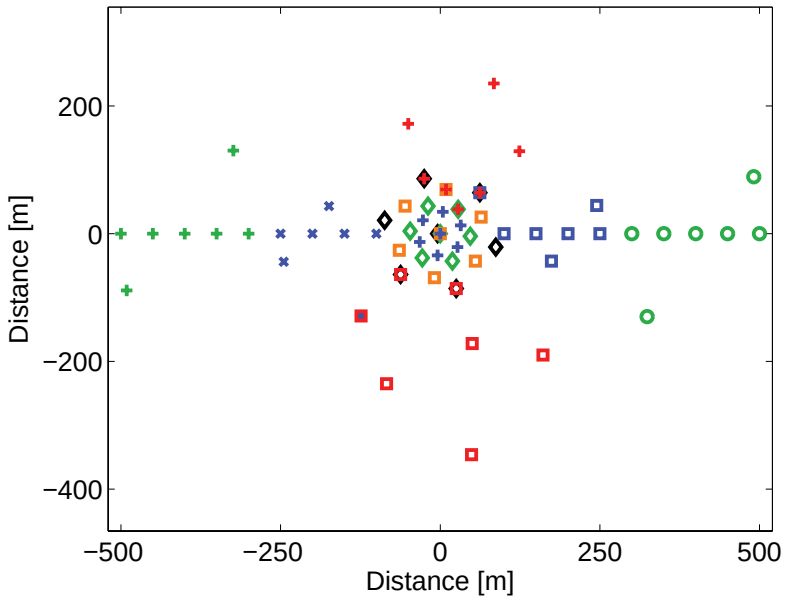


Figure A.2: Layout of the sixty sensors forming ten seven-sensor arrays. Some stations belong to several arrays.

For H/V, DELFI and RayDec, the sixty sensors were used independently. To test MUSIQUE, ten arrays of seven sensors each were formed, some sensors belonging to several arrays. The Rayleigh wave ellipticity estimations of the three presented new methods and the classical H/V technique are shown in figure A.3 in comparison to the theoretical ellipticity curve of the model.
The used free parameters are: for DELFI, $d\nu = 0.2\nu$; for RayDec, $d\nu = 0.2\nu$ and $\Delta = 10/\nu$; for MUSIQUE, $d\nu = 0.2\nu$ and $n_P = 10$.
The H/V technique overestimates ellipticity considerably, especially at frequencies above and below the peak frequency. Compared to H/V, DELFI only improves the peak frequency estimation. As it works on the raw signals, it does not eliminate Love and body waves and therefore its results are close to H/V results. The RayDec curve is very close to the theoretical curve over the whole analyzed frequency range. RayDec is a statistical method which averages over a large number of time-windows to efficiently eliminate other wave types than Rayleigh waves. That is why its results are very close to the theoretical curve. For MUSIQUE, the results are in good agreement with theory over the whole frequency range, but compared to RayDec, the error bars are larger. However, MUSIQUE does not suppress Love and body waves but also shows good results in identifying ellipticity. As the curve is the average for ten arrays, its results could propably be better with more arrays or more signal analysed. Optimisation of the array layouts (array aperture and geometry) could also improve the results.

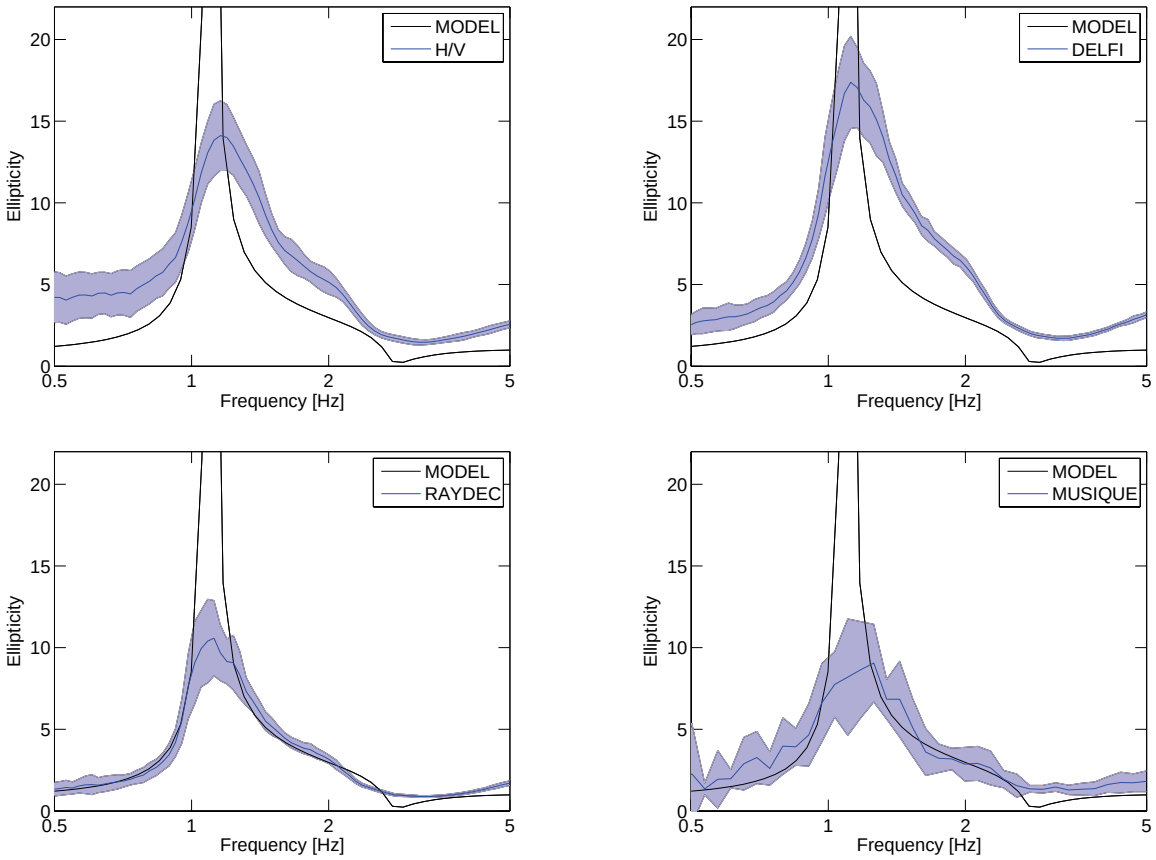


Figure A.3: Rayleigh wave ellipticity curves obtained by analyzing 620 seconds of synthetic seismic ambient vibrations by H/V, DELFI, RayDec and MUSIQUE techniques in comparison to the theoretical ellipticity curve of the model. For H/V, DELFI and RayDec, the average curves of 60 sensors are shown with the corresponding error range. For MUSIQUE, the average curve with error margin for ten arrays is shown.

In testing the methods, we have stated that RayDec requires a minimum signal length of some minutes to give stable results. DELFI and MUSIQUE, in contrast, do not require a minimum signal length and are thus applicable to non-stationary wave trains as produced in earthquakes. As MUSIQUE uses array measurements, it also gives an estimate of the azimuth and wave velocity. Therefore, it is also possible to distinguish between prograde and retrograde Rayleigh wave movement. DELFI and RayDec do not give the wave velocity and give two possible azimuths. Therefore, they cannot distinguish between prograde and retrograde Rayleigh waves.

A.8 Conclusion

The three new methods have been succesfully applied to synthetic noise data and have proven their ability to estimate the frequency-ellipticity curves of Rayleigh waves. The further work is to quantify the differences between the results of the method and the theory and to test the methods on other simulation data. Completing those tasks would validate the applicability of the proposed methods to real seismic data.

Acknowledgments

Noise synthetics were performed at the Service Commun de Calcul Intensif (SCCI) of Grenoble observatory, France (OSUG). This work is supported by the NERIES European project and the ANR QSHA project.

Bibliography

K. Aki. Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bull. Earthquake Res. Inst. Tokyo Univ.*, 35:415–456, 1957.

K. Aki and P. G. Richards. *Quantitative Seismology*. University Science Books, 2 edition, 2002.

H. Arai and K. Tokimatsu. S-wave velocity profiling by inversion of microtremor H/V spectrum. *Bull. Seismol. Soc. Am.*, 94:53–63, 2004.

H. Arai and K. Tokimatsu. S-wave velocity profiling by joint inversion of microtremor dispersion curve and horizontal-to-vertical (H/V) spectrum. *Bull. Seismol. Soc. Am.*, 95: 1766–1778, 2005. doi: 10.1785/0120040243.

H. Arai and K. Tokimatsu. Evaluation of local site effects based on microtremor H/V spectra. In *The Effects of Surface Geology on Seismic Motion*. Balkema, Rotterdam, 1998.

J. C. Asmussen. *Modal analysis based on the random decrement technique*. PhD thesis, University of Aalborg, 1997.

M. W. Asten and J. D. Henstridge. Array estimators and the use of microseisms for reconnaissance of sedimentary basins. *Geophysics*, 49:1828–1837, 1984. doi: 10.1190/1.1441596.

P.-Y. Bard. Microtremor measurements: a tool for site effect estimation? In *Second International Symposium on the Effects of Surface Geology on seismic motion, Yokohama, December 1-3, 1998*, volume 3, pages 1251–1279. Balkema, 1998.

T. Bardainne, P. Gaillot, N. Dubos-Sallée, J. Blanco, and G. Sénéchal. Characterization of seismic waveforms and classification of seismic events using chirplet atomic decomposition. Example from the Lacq gas field (Western Pyrenees, France). *Geophys. J. Int.*, 166:699–718, 2006. doi: 10.1111/j.1365-246X.2006.03023.x.

C. Beauval, P.-Y. Bard, P. Moczo, and J. Kristek. Quantification of frequency-dependent lenghtening of seismic ground-motion duration due to local geology: Applications to the Volvi area (Greece). *Bull. Seismol. Soc. Am.*, 93:371–385, 2003.

B. Bettig, P.-Y. Bard, F. Scherbaum, J. Riepl, F. Cotton, C. Cornou, and D. Hatzfeld. Analysis of dense array noise measurements using the modified spatial auto-correlation method (SPAC): application to the Grenoble area. *Boll. Geof. Teor. Appl.*, 42:281–304, 2001.

S. Bonnefoy-Claudet. *Nature du bruit de fond sismique : implications pour les études des effets de site*. PhD thesis, Université Joseph Fourier Grenoble, 2004.

S. Bonnefoy-Claudet, C. Cornou, P.-Y. Bard, F. Cotton, P. Moczo, J. Kristek, and D. Fäh. H/V ratio: a tool for site effects evaluation. Results from 1-D noise simulations. *Geophys. J. Int.*, 167:827–837, 2006. doi: 10.1111/j.1365-246X.2006.03154.x.

S. Bonnefoy-Claudet, A. Köhler, C. Cornou, M. Wathelet, and P.-Y. Bard. Effects of Love waves on microtremor H/V ratio. *Bull. Seismol. Soc. Am.*, 98:288–300, 2008. doi: 10.1785/0120070063.

D. M. Boore and M. N. Toksöz. Rayleigh wave particle motion and crustal structure. *Bull. Seismol. Soc. Am.*, 59:331–346, 1969.

M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, 7th edition, 1999.

J. L. Brenner. Matrices of quaternions. *Pacific J. Math.*, 1:329–335, 1951.

T. M. Brocher. Compressional and shear wave velocity versus depth in the San Francisco Bay Area, California: Rules for USGS Bay Area velocity model 05.0.0. U.S. Geological Survey Open-File Report 2005-1317 Version 1.0, 2005.

T. M. Brocher. Compressional and shear-wave velocity versus depth relations for common rock types in Northern California. *Bull. Seismol. Soc. Am.*, 98:950–968, 2008. doi: 10.1785/0120060403.

T. M. Brocher, E. E. Brabb, R. D. Catchings, G. S. Fuis, T. E. Fumal, R. C. Jachens, A. S. Jayko, R. E. Kayen, R. J. McLaughlin, T. Parsons, M. J. Rymer, R. G. Stanley, and C. M. Wentworth. A crustal-scale 3-D seismic velocity model for the San Francisco Bay area, California. *Eos Trans. AGU*, 78:F435, 1997.

I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 2001.

C. Brosseau. *Fundamentals of Polarized Light: A Statistical Optics Approach*. Wiley-Interscience, 1998.

J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57: 1408–1418, 1969.

I. Cho, T. Tada, and Y. Shinozaki. Assessing the applicability of the spatial autocorrelation method: A theoretical approach. *J. Geophys. Res.*, 113:B06307, 2008. doi: 10.1029/2007JB005245.

J. J. Christensen and J. Hald. Beamforming. Technical report, Brüel & Kjær, 2004.

C. Cornou. *Traitement d'antenne et imagerie sismique dans l'agglomération grenobloise (Alpes françaises) : implications pour les effets de site*. PhD thesis, Université Joseph Fourier Grenoble, 2002.

C. Cornou, P.-Y. Bard, and M. Dietrich. Contribution of dense array analysis to the identification and quantification of basin-edge induced waves, part I: Methodology. *Bull. Seismol. Soc. Am.*, 93:2604–2623, 2003a.

C. Cornou, P.-Y. Bard, and M. Dietrich. Contribution of dense array analysis to the identification and quantification of basin-edge induced waves, part II: Application to Grenoble basin (French Alps). *Bull. Seismol. Soc. Am.*, 93:2624–2648, 2003b.

C. Cornou, M. Ohrnberger, D. M. Boore, K. Kudo, and P.-Y. Bard. Derivation of structural models from ambient vibration array recordings: Results from an international blind test. In *Proc. 3rd Int. Symp. on the Effects of Surface Geology on Seismic Motion, Grenoble, 30 August - 01 September, 2006*, volume 2, pages 1127–1217. LCPC Editions, 2009.

K. De Meersman, M. van der Baan, and J.-M. Kendall. Signal extraction and automated polarization analysis of multicomponent array data. *Bull. Seismol. Soc. Am.*, 96:2415–2430, 2006.

G. Di Giulio, C. Cornou, M. Ohrnberger, M. Wathelet, and A. Rovelli. Deriving wavefield characteristics and shear-velocity profiles from two-dimensional small-aperture arrays analysis of ambient vibrations in a small-size alluvial basin, Colfiorito, Italy. *Bull. Seismol. Soc. Am.*, 96:1915–1933, 2006. doi: 10.1785/0120060119.

G. Di Giulio, A. Savvaidis, N. Theodulidis, M. Ohrnberger, B. Endrun, M. Wathelet, C. Cornou, F. Renalier, and P.Y. Bard. Inversion of surface wave dispersion at European strong motion sites using a multi-model parameterization and an information-theoretic approach. In *Proc. 14th European Conference on Earthquake Engineering, 30 August - 03 September 2010, Ohrid, Republic of Macedonia*, 2010.

D. Dolenc and D. Dreger. Microseisms observations in the Santa Clara Valley, California. *Bull. Seismol. Soc. Am.*, 95:1137–1149, 2005. doi: 10.1785/0120040060.

D. Dolenc, D. Dreger, and S. Larsen. Basin structure influences on the propagation of teleseismic waves in the Santa Clara Valley, California. *Bull. Seismol. Soc. Am.*, 95: 1120–1136, 2005. doi: 10.1785/0120040059.

F. Dunand. *Pertinence du bruit de fond sismique pour la caractérisation dynamique et l'aide au diagnostic sismique des structures de génie civil*. PhD thesis, Université Joseph Fourier Grenoble, 2005.

B. Endrun. Love wave contribution to the ambient vibration H/V amplitude peak observed with array measurements. *J. Seismol.*, 2010. doi: 10.1007/s10950-010-9191-x.

B. Endrun and F. Renalier. Report on in-situ measurements at the 20 selected sites. NERIES deliverable JRA4 D2, available at http://www.neries-eu.org, 2008.

B. Endrun, M. Ohrnberger, and A. Savvaidis. On the repeatability and consistency of three-component ambient vibration array measurements. *Bull. Earthquake Eng.*, 8: 535–570, 2010. doi: 10.1007/s10518-009-9159-9.

D. Fäh, F. Kind, and D. Giardini. A theoretical investigation of average H/V ratios. *Geophys. J. Int.*, 145:535–549, 2001.

D. Fäh, F. Kind, and D. Giardini. Inversion of local S-wave velocity structures from average H/V ratios, and their use for the estimation of site-effects. *Journal of Seismology*, 7:449–467, 2003.

D. Fäh, G. Stamm, and H.-B. Havenith. Analysis of three-component ambient vibration array measurements. *Geophys. J. Int.*, 172:199–213, 2008. doi: 10.1111/j.1365-246X.2007.03625.x.

D. Fäh, M. Wathelet, M. Kristekova, H. Havenith, B. Endrun, G. Stamm, V. Poggi, J. Burjanek, and C. Cornou. Using ellipticity information for site characterisation. NERIES deliverable JRA4 D4, available at http://www.neries-eu.org, 2009.

E. H. Field. Spectral amplification in a sediment-filled valley exhibiting clear basin-edge-induced waves. *Bull. Seismol. Soc. Am.*, 86:991–1005, 1996.

A. W. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least-squares fitting of ellipses. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21:476–480, 1999.

J. B. Fletcher, J. Boatwright, and A. G. Lindh. Wave propagation and site response in the Santa Clara Valley. *Bull. Seismol. Soc. Am.*, 93:480–500, 2003.

E. A. Flinn. Signal analysis using rectilinearity and direction of particle motion. *Proc. IEEE*, 53:1874–1876, 1965.

A. Frankel, D. Carver, E. Cranswick, T. Bice, R. Sell, and S. Hanson. Observations of basin ground motions from a dense seismic array in San Jose, California. *Bull. Seismol. Soc. Am.*, 91:1–12, 2001.

S. Gaffet, C. Larroque, A. Deschamps, and F. Tressols. A dense array experiment for the observation of waveform perturbations. *Soil Dynamics and Earthquake Engineering*, 17:475–484, 1998.

A. V. Gorbatikov, M. Yu. Stepanova, and G. E. Korablev. Microseismic field affected by local geological heterogeneities and microseismic sounding of the medium. *Izvestiya, Physics of the Solid Earth*, 44:577–592, 2007. doi: 10.1134/S1069351308070082.

P. Gouédard, C. Cornou, and P. Roux. Phase-velocity dispersion curves and small-scale geophysics using noise correlation slantstack technique. *Geophys. J. Int.*, 172:971–981, 2008. doi: 10.1111/j.1365-246X.2007.03654.x.

A. Grandi, A. Mazzotti, and E. Stucchi. Multicomponent velocity analysis with quaternions. *Geophysical Prospecting*, 55:761–777, 2007. doi: 10.1111/j.1365-2478.2007.00657.x.

P. Guéguen, J.-L. Chatelain, B. Guillier, H. Yepes, and J. Egred. Site effect and damage distribution in Pujili (Ecuador) after the 28 March 1996 earthquake. *Soil Dynamics and Earthquake Engineering*, 17:329–334, 1998.

E. Haghshenas, P.-Y. Bard, N. Theodulidis, and SESAME WP04 team. Empirical evaluation of microtremor H/V spectral ratio. *Bulletin of Earthquake Engineering*, 6:75–108, 2008.

W. R. Hamilton. On quaternions, or on a new system of imagineries in algebra. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 1844-50.

W. R. Hamilton. On the geometrical interpretation of some results obtained by calculation with biquaternions. *Proc. of the Royal Irish Academy*, 5:388–390, 1853.

D. G. Harkrider. Surface waves in multilayered elastic media, I. Rayleigh and Love waves from buried sources in a multilayered elastic half-space. *Bull. Seismol. Soc. Am.*, 54:627–679, 1964.

D. G. Harkrider and D. L. Anderson. Computation of surface wave dispersion for multilayered anisotropic media. *Bull. Seismol. Soc. Am.*, 52:321–332, 1962.

S. Hartzell, D. Carver, R. A. Williams, S. Harmsen, and A. Zerva. Site response, shallow shear-wave velocity, and wave propagation at the san jose, california, dense seismic array. *Bull. Seismol. Soc. Am.*, 93:443–464, 2003. doi: 10.1785/0120020080.

S. Hartzell, S. Harmsen, R. A. Williams, D. Carver, A. Frankel, G. Choy, P.-C. Liu, R. C. Jachens, T. M. Brocher, and C. M. Wentworth. Modeling and validation of a 3D velocity structure for the Santa Clara Valley, California, for seismic-wave simulations. *Bull. Seismol. Soc. Am.*, 96:1851–1881, 2006. doi: 10.1785/0120050243.

S. Hartzell, L. Ramirez-Guzman, D. Carver, and P. Liu. Short baseline variations in site response and wave-propagation effects and their structural causes: Four examples in and around the Santa Clara Valley, California. *Bull. Seismol. Soc. Am.*, 100:2264–2286, 2010. doi: 10.1785/0120090278.

Y. Hisada. An efficient method for computing Green's functions for a layered halfspace with sources and receivers at close depths. *Bull. Seismol. Soc. Am.*, 84:1456–1472, 1994.

Y. Hisada. An efficient method for computing Green's functions for a layered halfspace with sources and receivers at close depths (part 2). *Bull. Seismol. Soc. Am.*, 85:1080–1093, 1995.

M. Hobiger, P.-Y. Bard, C. Cornou, and N. Le Bihan. Single station determination of Rayleigh wave ellipticity by using the random decrement technique (RayDec). *Geophys. Res. Lett.*, 36:L14303, 2009a. doi: 10.1029/2009GL038863.

M. Hobiger, N. Le Bihan, C. Cornou, and P.-Y. Bard. Rayleigh wave ellipticity estimation from ambient seismic noise using single and multiple vector-sensor techniques. In *17th European Signal Processing Conference, Glasgow, UK, 24-28 August, 2009*, 2009b.

M. Hobiger, C. Cornou, P.-Y. Bard, N. Le Bihan, and M. Wathelet. Inversion of Rayleigh wave ellipticity measurements, Part I: Theoretical aspects. submitted to Geophys. J. Int., 2010a.

M. Hobiger, C. Cornou, G. Di Giulio, B. Endrun, F. Renalier, A. Savvaidis, P.-Y. Bard, S. Hailemikael, N. Le Bihan, M. Ohrnberger, N. Theodulidis, and M. Wathelet. Inversion of Rayleigh wave ellipticity measurements, Part II: Application to real data measurements. submitted to Geophys. J. Int., 2010b.

C. I. Huerta, J. M. Roesset, and K. H. Stokoe. Evaluation of the random decrement method for in-situ soil properties. In *The Effects of Surface Geology on Seismic Motion*, pages 749–756. Balkema, Rotterdam, 1998.

A. Jurkevics. Polarization analysis of three-component array data. *Bull. Seismol. Soc. Am.*, 78:1725–1743, 1988.

I. L. Kantor and A. S. Solodovnikov. *Hypercomplex Numbers*. Springer-Verlag, 1989.

H. Kawase. The cause of the damage belt in Kobe: "the basin-edge effect", constructive interference of the direct S-wave with the basin-induced diffracted/Rayleigh waves. *Seism. Res. Lett.*, 67:25–34, 1996.

H. Kawase and K. Aki. A study on the response of a soft basin for incident S, P, and Rayleigh waves with special reference to the long duration observed in Mexico City. *Bull. Seismol. Soc. Am.*, 79:1361–1382, 1989.

A. Köhler, M. Ohrnberger, and F. Scherbaum. The relative fraction of Rayleigh and Love waves in ambient vibration wavefields at different European sites. In *Proc. 3rd Int. Symp. on the Effects of Surface Geology on Seismic Motion, Grenoble, 30 August - 01 September, 2006*, volume 1, pages 351–360, 2006.

A. Köhler, M. Ohrnberger, F. Scherbaum, M. Wathelet, and C. Cornou. Assessing the reliability of the modified three-component spatial autocorrelation technique. *Geophys. J. Int.*, 168:779–796, 2007. doi: 10.1111/j.1365-246X.2006.03253.x.

K. Konno and T. Ohmachi. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bull. Seismol. Soc. Am.*, 88:228–241, 1998.

R. T. Lacoss, E. J. Kelly, and M. N. Toksöz. Estimation of seismic noise structure using arrays. *Geophysics*, 34:21–38, 1969.

J.-L. Lacoume, P.-O. Amblard, and P. Comon. *Statistiques d'ordre supérieur pour le traitement du signal*. Masson, 1997.

T. Lay and T. C. Wallace. *Modern Global Seismology*. Academic Press, 1995.

N. Le Bihan and J. Mars. Singular value decomposition of quaternion matrices: A new tool for vector-sensor signal processing. *Signal Processing*, 84:1177–1199, 2004.

B. LeBrun, D. Hatzfeld, and P.-Y. Bard. Site effect study in urban area: Experimental results in Grenoble (France). *Pure appl. geophys.*, 158:2543–2557, 2001.

H. C. Lee. Eigenvalue and canonical forms of matrices with quaternion coefficients. *Proc. of the Royal Irish Academy*, 52:253–261, 1949.

J. Lermo and F. J. Chávez-García. Are microtremors useful in site response evaluation? *Bull. Seismol. Soc. Am.*, 84:1350–1364, 1994.

P. G. Malischewsky and F. Scherbaum. Love's formula and H/V-ratio (ellipticity) of rayleigh waves. *Wave motion*, 40:57–67, 2004.

S. Marcos, editor. *Les méthodes à haute résolution - traitement d'antenne et analyse spectrale*. Editions HERMES, 1998.

J. Max and J.-L. Lacoume. *Méthodes et techniques de traitement du signal et applications aux mesures physiques - 1. Principes généraux et méthodes classiques*. Masson, 5 edition, 1996.

C. Michel, P. Guéguen, and P.-Y. Bard. Dynamic parameters of structures extracted from ambient vibration measurements: An aid for the seismic vulnerability assessment of existing buildings in moderate seismic hazard regions. *Soil Dynamics and Earthquake Engineering*, 28:593–604, 2008.

S. Miron. *Méthodes multilinéaires et hypercomplexes en traitement d'antenne multicomposante haute résolution*. PhD thesis, INP Grenoble, 2005.

S. Miron, N. Le Bihan, and J. Mars. Vector-sensor MUSIC for polarized seismic sources localization. *EURASIP J. Appl. Signal Process.*, pages 74–84, 2005a.

S. Miron, N. Le Bihan, and J. I. Mars. High resolution vector-sensor array processing using quaternions. In *IEEE International Conference on Statistical Signal Processing, Bordeaux, France, 2005*, 2005b.

S. Miron, N. Le Bihan, and J. Mars. Quaternion-MUSIC for vector-sensor array processing. *IEEE Trans. Signal Process.*, 54:1218–1229, 2006.

P. Moczo and J. Kristek. FD code to generate noise synthetics. Sesame report D02.09, available at http://SESAME-FP5.obs.ujf-grenoble.fr, 2002.

J. F. Montalbetti and E. R. Kanasewich. Enhancement of teleseismic body phases with a polarization filter. *Geophys. J. R. astr. Soc.*, 21:119–129, 1970.

M. L. Moran and D. G. Albert. Source location and tracking capability of a small seismic array. Technical report, CRREL, 1996.

H. Morikawa, S. Sawada, and J. Akamatsu. A method to estimate phase velocities of Rayleigh waves using microseisms simultaneously observed at two sites. *Bull. Seismol. Soc. Am.*, 94:961–976, 2004.

Y. Nakamura. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly reports of the Railway Technical Research Institute Tokyo*, 30:25–33, 1989.

M. Nogoshi and T. Igarashi. On the amplitude characteristics of microtremor (part 2). *J. Seism. Soc. Japan*, 24:26–40, 1971. in Japanese with English abstract.

H. Okada. Theory of efficient array observations of microtremors with special reference to the SPAC method. *Exploration Geophysics*, 37:73–85, 2006.

C. B. Park, R. D. Miller, and J. Xia. Multichannel analysis of surface waves. *Geophysics*, 64:800–808, 1999.

S. Parolai, M. Picozzi, S. M. Richwalski, and C. Milkereit. Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings using a genetic algorithm, considering higher modes. *Geophys. Res. Lett.*, 32:L01303, 2005. doi: 10.1029/2004GL021115.

G. A. Phelps, R. W. Graymer, R. C. Jachens, D. A. Ponce, R. W. Simpson, and C. M. Wentworth. Three-dimensional geologic map of the Hayward Fault zone, San Francisco Bay region, California. U.S. Geological Survey Scientific Investigations Map 3045, http://pubs.usgs.gov/sim/3045/, 2008.

W. S. Phillips and K. Aki. Site amplification of coda waves from local earthquakes in central California. *Bull. Seismol. Soc. Am.*, 76:627–648, 1986.

M. Picozzi, S. Parolai, and S. M. Richwalski. Joint inversion of H/V ratios and dispersion curves from seismic noise: Estimating the S-wave velocity of bedrock. *Geophys. Res. Lett.*, 32:L11308, 2005. doi: 10.1029/2005FL022878.

M. Picozzi, F. Sabetta, N. Theodulidis, S. Zacharopoulos, A. Savvaidis, P.-Y. Bard, C. Cornou, P. Guéguen, D. Fäh, I. Kalogeras, S. Akkar, D. Rinaldis, and G. Tanircan. Selected sites and available information. NERIES deliverable JRA4 D1, available at http://www.neries-eu.org, 2007.

A. Plešinger, M. Hellweg, and D. Seidl. Interactive high-resolution polarization analysis of broad-band seismograms. *Geophysics*, 59:129–139, 1986.

V. Poggi and D. Fäh. Estimating Rayleigh wave particle motion from three-component array analysis of ambient vibrations. *Geophys. J. Int.*, 180:251–267, 2010. doi: 10.1111/j.1365-246X.2009.04402.x.

D. Raptakis, N. Theodulidis, and K. Pitilakis. Data analysis of the Euroseistest strong motion array in Volvi (Greece): Standard and Horizontal-to-Vertical spectral ratio techniques. *Earthquake Spectra*, 14:203–224, 1998.

F. Renalier. *Caractérisation sismique de sites hétérogènes à partir de méthodes actives et passives : variations latérales et temporelles*. PhD thesis, Université Joseph Fourier Grenoble, 2010.

F. Renalier and B. Endrun. Comparative analysis of classical measurements and newly developed methods. NERIES deliverable JRA4 D6, available at http://www.neries-eu.org, 2009.

J. Riepl, P.-Y. Bard, D. Hatzfeld, C. Papaioannou, and S. Nechtschein. Detailed evaluation of site-response estimation methods across and along the sedimentary valley of Volvi (EURO-SEISTEST). *Bull. Seismol. Soc. Am.*, 88:488–502, 1998.

O. Rodrigues. Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *Journal de mathématiques pures et appliquées*, pages 380–440, 1840.

S. Rost and E. J. Garnero. Array seismology advandes research into Earth's interior. *EOS*, 85:301,305–306, 2004.

S. Rost and C. Thomas. Array seismology: Methods and applications. *Rev. Geophys.*, 40: 1008, 2002. doi: 10.1029/2000RG000100.

A. Rovelli, L. Scognamiglio, F. Marra, and A. Caserta. Edge-diffracted 1-sec surface waves observed in a small-size intramountain basin (Colfiorito, central Italy). *Bull. Seismol. Soc. Am.*, 91:1851–1866, 2001.

M. Sambridge. Geophysical inversion with a neighbourhood algorithm - I. Searching a parameter space. *Geophys. J. Int.*, 138:479–494, 1999a.

M. Sambridge. Geophysical inversion with a neighbourhood algorithm - II. Appraising the ensemble. *Geophys. J. Int.*, 138:727–746, 1999b.

F. Scherbaum, F. K. Hinzen, and M. Ohrnberger. Determination of shallow shear wave velocity profiles in the Cologne, Germany area using ambient vibrations. *Geophys. J. Int.*, 152:597–612, 2003.

M. Schimmel and J. Gallart. The use of instantaneous polarization attributes for seismic signal detection and image enhancement. *Geophys. J. Int.*, 155:653–668, 2003.

R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas and Propagation*, 34:276–280, 1986.

F. Schwab and L. Knopoff. Surface wave dispersion computations. *Bull. Seismol. Soc. Am.*, 60:321–344, 1970.

O. Sèbe. *Déconvolution aveugle et sismologie : estimation de la source sismique et des effets de site*. PhD thesis, Université Joseph Fourier Grenoble, 2004.

J. L. Sexton, A. J. Rudman, and J. Mead. Ellipticity of Rayleigh waves recorded in the Midwest. *Bull. Seismol. Soc. Am.*, 67:369–382, 1977.

N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller. High-resolution surface-wave tomography from ambient seismic noise. *Science*, 307:1615–1618, 2005. doi: 10.1126/science.1108339.

L. V. Socco and C. Strobbia. Surface-wave method for near-surface characterization: a tutorial. *Near Surface Geophysics*, 2:165–185, 2004.

D. B. Sweetser. Doing physics with quaternions. http://quaternions.com, 2005.

T. Tada, I. Cho, and Y. Shinozaki. New circular-array microtremor techniques to infer Love-wave phase velocities. *Bull. Seismol. Soc. Am.*, 99:2912–2926, 2009. doi: 10.1785/0120090014.

T. Tanimoto and L. Rivera. Prograde Rayleigh wave particle motion. *Geophys. J. Int.*, 162: 399–405, 2005. doi: 10.1111/j.1365-246X.2005.02481.x.

M. E. Taylor. *Noncommutative Harmonic Analysis*. American Mathematical Society, 1986.

Y. Tian. Matrix theory over the complex quaternion algebra. *ArXiv Mathematics e-prints*, 2000.

K. Tokimatsu. Geotechnical site characterization using surface waves. In *Earthquake Geotechnical Engineering*, pages 1333–1368. Balkema, Rotterdam, 1997.

K. Tokimatsu, S. Wakai, and H. Arai. Three-dimensional soil stratification using surface waves in microtremors. In *Geotechnical Site Characterization*, pages 537–542. Balkema, Rotterdam, 1998.

J. E. Vidale. Complex polarization analysis of particle motion. *Bull. Seismol. Soc. Am.*, 76: 1393–1405, 1986.

V. D. Vrabie, N. Le Bihan, and J. I. Mars. Multicomponent wave separation using HOSVD/unimodal-ICA subspace method. *Geophysics*, 71:V133–V143, 2006. doi: 10.1190/1.2335.387.

G. S. Wagner. Resolving diversely polarized, superimposed signals in three-component seismic array data. *Geophys. Res. Lett.*, 23:1837–1840, 1996.

G. S. Wagner. Regional wave propagation in southern California and Nevada: Observations from a three-component seismic array. *J. Geophys. Res.*, 102:8285–8311, 1997.

J. P. Ward. *Quaternions and Cayley Numbers: Algebra and Applications*. Kluwer Academic Publishers, 1997.

M. Wathelet. An improved neighborhood algorithm: Parameter conditions and dynamic scaling. *Geophys. Res. Lett.*, 35:L09301, 2008. doi: 10.1029/2008GL033256.

M. Wathelet. *Array recordings of ambient vibrations: surface-wave inversion*. PhD thesis, Université de Liège, 2005.

M. Wathelet, D. Jongmans, and M. Ohrnberger. Surface wave inversion using a direct search algorithm and its application to ambient vibration measurements. *Near Surface Geophysics*, 2:211–221, 2004.

M. Wathelet, D. Jongmans, and M. Ohrnberger. Direct inversion of spatial autocorrelation curves with the neighborhood algorithm. *Bull. Seismol. Soc. Am.*, 95:1787–1800, 2005.

M. Wathelet, D. Jongmans, M. Ohrnberger, and S. Bonnefoy-Claudet. Array performances for ambient vibrations on a shallow structure and consequences over V_S inversion. *J. Seismol.*, 12:1–19, 2008.

G. S. Watson. *Statistics on spheres*. Wiley-Interscience, 1983.

C. M. Wentworth. Geologic materials of the San Francisco Bay region. U.S. Geological Survey Open-File Report 97-744, part 5, 1997.

J. W. Woods and P. R. Lintz. Plane waves at small arrays. *Geophysics*, 38:1023–41, 1973. doi: 10.1190/1.1440393.

A. Zerva and O. Zhang. Estimation of signal characteristics in seismic ground motions. *Probabilistic Eng. Mech.*, 11:229–242, 1996.

F. Zhang. Quaternions and matrices of quaternions. *Linear Algebra and its Applications*, 251:21–57, 1997.

Polarisation des ondes de surface : Caractérisation, inversion et application à l'étude de l'aléa sismique

L'aléa sismique d'un site donné peut être largement influencé par des effets de site. Afin d'évaluer ces effets, la structure locale du sous-sol ainsi que les propriétés du champ d'onde doivent être étudiées. Les ondes de surface (ondes de Love et de Rayleigh) s'avèrent utiles, leurs propriétés (courbes de dispersion, ellipticité des ondes de Rayleigh) étant directement liées à la structure du sous-sol. Le paramètre clé pour l'identification du type d'onde est la polarisation. Dans la première partie de la thèse, de nouvelles méthodes pour l'estimation de paramètres de polarisation d'ondes de surface sont développées. Deux méthodes, DELFI et RayDec, estiment l'ellipticité des ondes de Rayleigh à partir d'enregistrements d'un seul capteur sismique. La troisième méthode, MUSIQUE, est basée sur la méthode MUSIC et utilise les enregistrements multi-composantes de réseaux sismiques afin de distinguer ondes de Love et ondes de Rayleigh et d'estimer leurs propriétés. Dans la deuxième partie de la thèse, une étude théorique de l'inversion de courbes d'ellipticité montre quelles parties de ces courbes véhiculent les informations importantes sur la structure du sol et comment l'inversion peut être améliorée. Le schéma d'inversion résultant est alors testé en l'appliquant à des données réelles mesurées pour 14 sites européens. Finalement, 22 séismes enregistrés par un réseau de capteurs dans la vallée de Santa Clara en Californie sont analysés par MUSIQUE. La répartition azimutale, les courbes de dispersion, la courbe d'ellipticité et les énergies des différents types d'ondes sont analysées et soulignent l'importance des ondes de surface diffractées dans le champ d'ondes enregistré.

Mots-clés :
sismologie - ondes de surface - analyse de polarisation - ellipticité des ondes de Rayleigh - effets de site

Polarization of surface waves: Characterisation, inversion and application to seismic hazard assessment

The seismic hazard of a given site can be largely influenced by site effects. In order to estimate these effects, the local soil structure and the wave field properties have to be investigated. This can be done using surface waves (Love and Rayleigh waves) since their properties (dispersion curves, Rayleigh wave ellipticity) are closely linked to the soil structure. The key parameter for the correct wave type identification is the polarization. In a first part of this thesis, three new methods to estimate the polarization parameters of surface waves have been developed. Two methods, DELFI and RayDec, estimate the Rayleigh wave ellipticity using the recordings of a single seismic station. The third method, MUSIQUE, is a further developed version of the MUSIC algorithm, and uses seismic array recordings to discriminate between Love and Rayleigh waves and estimate their respective properties. In the second part of the thesis, a theoretical investigation of the inversion of ellipticity curves shows which parts of an ellipticity curve carry the important information on the soil structure and by which means the inversion can be disambiguated. The developed inversion scheme is then tested on real data measurements obtained at 14 different European sites. Finally, the seismological wave field in the Californian Santa Clara valley is investigated applying MUSIQUE to seismic array recordings for 22 earthquakes. The azimuthal energy repartition between the different wave types is investigated and the respective dispersion curves as well as the Rayleigh wave ellipticity curve are estimated. This study outlines the importance of diffracted surface waves in the recorded wave field.

Keywords: seismology - surface waves - polarization analysis - Rayleigh wave ellipticity - site effects