Ambient noise spectral amplitude distortions above heterogeneities : feasability study for multi-fluid reservoir exploration and monitoring

Alexandre Kazantsev

▶ **To cite this version:**

Alexandre Kazantsev. Ambient noise spectral amplitude distortions above heterogeneities : feasability study for multi-fluid reservoir exploration and monitoring. Geophysics [physics.geo-ph]. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLEM075 . tel-02143918

HAL Id: tel-02143918
https://pastel.hal.science/tel-02143918

Submitted on 29 May 2019

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

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

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Ambient noise spectral amplitude distortions above heterogeneities: feasability study for multi-fluid reservoir exploration and monitoring

Perturbations d'amplitude du bruit ambiant au droit des hétérogéneités: étude de faisabilité pour l'exploration et la surveillance de réservoirs multi-fluide

École doctorale n°398

GÉOSCIENCES-RESSOURCES NATURELLES-ENVIRONNEMENT

Spécialité GÉOSCIENCES ET GÉOINGÉNIERIE

COMPOSITION DU JURY :

Pr. Nikolaï SHAPIRO
IPGP, Président du jury

Pr. Valérie MAUPIN
Université d'Oslo, Rapporteur

Pr. Erik H. SAENGER
Hochschule Bochum, Rapporteur

Dr. René-Edouard PLESSIX
Shell, Examinateur

Dr. Gilles GRANDJEAN
BRGM, Examinateur

Dr. Pierre DUBLANCHET
MINES ParisTech, Examinateur

Pr. Hervé CHAURIS
MINES ParisTech, Examinateur

M. Frédéric HUGUET
Storengy, Examinateur

Dr. Jean-Philippe METAXIAN
IRD/IPGP, Invité

Soutenue par **Alexandre KAZANTSEV**
le 03/12/2018

Dirigée par **Pr. Hervé CHAURIS**

MINES ParisTech | PSL RESEARCH UNIVERSITY PARIS

Late in the evening, a policeman notices
someone looking for his keys near a lamppost
- Did you lose the keys here?
- No, but the light is much better.

A popular joke

Remerciements

Je me souviens bien d'une certaine journée de novembre 2014, où j'ai décidé de postuler pour un stage à Storengy. Le sujet concernait une vague *anomalie sismique basse fréquence*. Il pleuvait très fort et j'étais en retard pour l'entretien, à peu près comme je le suis maintenant pour le rendu du manuscrit. J'ai donc dû prendre le vélo pour aller plus vite que le bus. Arrivé à Storengy totalement dégoulinant (mais tout de même en chemise), j'ai vu défiler devant moi des spectres, et ensuite d'autres spectres. J'ai compris alors qu'il s'agissait d'une anomalie spectrale. Il fallait comprendre ce que l'on pouvait en faire en termes de monitoring de stockage et d'exploration géothermique. Damien et Frédéric n'avaient pas l'air de faire grand cas de mon piteux état. Ils étaient intéressants et sympathiques, alors je me suis lancé.

Très rapidement, j'ai compris qu'il n'y avait pas forcément besoin de venir en chemise à Storengy, et c'est là le premier point de mes remerciements. Merci Tiphaine, Florian, Hélène, Charlotte, Yannick, Floriane, Aden et tous les autres pour avoir participé à ma sociabilisation. Merci David, Marc, Stéphane, Julien, Simon pour les matchs de basket à midi en temps estival, pour énerver la canicule. Merci Greg pour les parties d'échecs. Merci Vincent, Jean-Fred et Benoît, les gourous de Linux, Fortran et autre MPI. Merci Typhaine, Floriane, Myriam, Frédéric, Marie-Paule, Yvan, Gladys, Pauline et tant d'autres pour les mails commençant par *gâteau à la machine à café*, cet excellent prétexte pour remettre à plus tard le debug d'un code qui ne marche pas plus à 9h30 du matin qu'il ne marchait la veille à minuit. Merci Delphine, Thomas et Arnaud pour la distribution quasi-quotidienne de chocolats sur le coup des 16h, alors que le code ne marche toujours pas. Et merci Frédéric pour l'ail des ours, le mélilot, et plus largement pour l'initiaiton à la flore sauvage et son bon usage. En restant dans la veine du terroir, merci à Yann de m'avoir initié aux nombreuses subtilités entre pommes à couteau et pommes à cidre, avec travaux pratiques. Même si j'oublie beaucoup de gens, le lecteur aura compris, Storengy c'est bien. D'abord, parce qu'on y mange bien. Mais quid de l'anomalie spectrale ?

Premier constat: elle a l'air de vraiment exister! On est en janvier 2015. Le terme *capteur basse fréquence* domine alors l'histogramme de mon vocabulaire en résidence universitaire, ex aequo avec *la vague océanique qui fait vibrer la terre*. C'est le moment de remercier tous les copains qui ont su me supporter à ce moment-là, et surtout Adrien, Yolande et Pauline (qui aura droit à son paragraphe, et qui peut donc sauter la suite). Mes encadrants à Storengy sont presque aussi enthousiastes que moi, et suivent avec attention l'évolution des cartes d'anomalie au fil des semaines: d'attribut en attribut, la *tache rouge* se concentre de plus en plus au-dessus du réservoir. C'est le moment de dire officiellement merci à Frédéric Huguet, Damien Lavergne et Patrick Egermann pour leur temps et l'intérêt qu'ils ont porté à mon travail, dès mes premiers jours à Storengy, et qui m'ont ouvert de très nombreuses opportunités. Et merci à Raouf Nabil, le chef du département, de m'avoir proposé de poursuivre l'aventure en thèse, pour espérer enfin cerner la nature de l'anomalie basse fréquence. Du côté de Storengy, un merci particulier va aussi à Catherine Formento, experte en Petrel et en géostatistique, qui m'a consacré beaucoup de temps pour produire de nombreuses images parmi celles présentées dans le Chapitre 2 de ce manuscrit.

Hervé Chauris, qui était alors mon professeur de géophysique à l'Ecole de Mines, a accepté d'encadrer ce travail avec enthousiasme, là encore. Tout au long de cette thèse, il a été d'une disponibilité, d'une patience et d'une attention rares. Qui plus est, nous avons fait ensemble un certain nombre de trails en Forêt de Fontainebleau, profitant de l'emplacement plutôt favorisé du Centre de Géosciences. Merci aussi à Pierre Dublanchet pour son expertise en physique, qui a contribué à guider ce travail. Du côté des Mines, merci à toute la sympathique équipe de Géophysique: Alexandrine, Mark, Pierre-François, Nidhal, Michelle, Tyaniou, Hao, Yubing, Yves-Marie, Charles-Antoine et d'autres. Un merci particulier à Keurfon, qui s'est carrément investi pour m'aider dans ma thèse, avec son expertise en code et en inversion d'ondes de surface. Merci à Véronique Lachasse, qui a su faire face à ma désorganisation administrative en restant d'une gentillesse inconditionnelle.

Grâce à l'enthousiasme de Patrick, Frédéric et Raouf, cette thèse m'a emmené jusque dans la jungle indonésienne où, à défaut d'avoir rencontré un tigre, j'ai rencontré Jean-Philippe Métaxian. Ce dernier m'a appris beaucoup de choses, parmi lesquelles la technique d'installation des stations large-bande permanentes, que j'ai essayé, parfois avec succès, de mettre en place sur le site de stockage de Saint-Illiers. Merci aussi à Rizal et Teja de l'IRD, ainsi qu'à Irvan, Redwan, et tout le personnel de Supreme Energy, avec qui nous avons arpenté les sentiers. Du côté de Saint-Illiers, quel soulagement que de ne pas avoir travaillé seul, grâce à l'aide de Marc Peruzzetto, Simon Lejart et plus récemment Jérôme Vergne et Damian Kula, avec qui nous avons partagé les joies du terrain. Heuresement, le personnel du site de Saint-Illiers était toujours chaleureux et avenant pour faire en sorte que nos interventions pour la maintenance du réseau se déroulent dans les meilleurs conditions. Un merci particulier à Marc Peruzzetto dont le travail de stage s'est totalement inscrit dans l'avancement de ma thèse.

Pauline, qui a rendu son propre manuscrit il y a quelques heures, a eu le lot le plus difficile: partager avec moi le moment de la rédaction, et la bonne ambiance qui va avec. Merci pour ton soutien permanent, et d'avoir cru en ce que je faisais, souvent plus fort que je n'y croyais moi-même. On remercie tous les deux Charik.

Merci à l'association des Gogols d'Islande, à Zézé et l'Homme en Slip, au Baba Yetu et à la barque Alleluia, n'en déPlesse à Sherlock. Merci au grand camarade Breton en peignoir et ses histoires du Kirghizistan Éthiopien (dit Bukistan). Merci à Océane pour la francophilie, l'héritage communiste et la soupe d'orties. D'une manière plus générale, merci à tout ce qui court et qui se jette.

Last but not least, merci à ma famille qui par son humour bienveillant a toujours su dédramatiser les enjeux. Bon courage à Gena. Mes plus fortes pensées vont à *Babouchka*, partie il y a deux ans, sans qui je ne serais ni en France, ni tout court.

Enfin, merci aux membres du jury d'avoir accepté ce rôle, et désolé de les avoir fait attendre pour l'envoi du manuscrit.

General summary

A considerable amount of experimental evidence in the literature reports strong vertical component amplification of the ambient noise above hydrocarbon reservoirs at frequencies of several Hertz. In this PhD work, three datasets recorded in different contexts (two Undeground Gas Storages and one high-enthalpy geothermal field) were analysed, and spectral amplitude distortions were indeed found above the different reservoirs. The spectral signature, however, is different in gas storage and geothermal contexts. A non-supervised algorithm for amplitude spectrum classification was developed and patented, allowing to extract relevant attributes for constructing anomaly maps at the surface. The general purpose of this PhD work was to understand the physical mechanism behind the amplitude perturbations, yet unclear in the literature. More precisely, I focused on what could be explained with elastic modelling.

As a first step, a wavefield characterisation methodology relying on a combination of MUSIC and ambient-noise cross-correlation was designed and applied to the largest of the available datasets in order to determine the composition of the ambient noise. Besides the common fundamental mode Rayleigh-waves, it revealed the presence of strong overtones and L_g phase, which itself is a mixture of overtones. Guided by this observation, I attempted to explain the observed spectral amplitude anomalies based on multi-modal Rayleigh wave scattering by the reservoir. Numerical 2D elastic modelling was used to simulate the propagation of Rayleigh waves across a reservoir-like inclusion embedded within a realistic geological structure. Using reasonable elastic contrasts for a gas-saturated reservoir, only *small* amplitude perturbations were obtained at the surface. The strong amplitude distortions observed in the real data could not be explained by the presence of an elastic reservoir alone. However, a realistic structural anticline was able to generate spectral amplitude perturbations of the correct order of magnitude and frequency content in the numerical simulations.

Using the Born approximation, I was able to invert the *small* Rayleigh wave amplitude perturbations arising in the numerical simulations for the position of reservoir-like embedded scatterers, in simple background models. This method could in theory be used for processing small time-lapse amplitude variations caused by the reservoir alone. However, several obstacles remain to be overcome before a real-data application, which is discussed in this thesis. Imaging structural heterogeneities, such as reservoir-bearing anticlines, would require a non-linear approach, as I demonstrate that the Born approximation based on a flat background model fails to predict the effects of a realistic structural deformation at the surface above it. Visco-elasticity, which might be relevant for explaining fluid-specific effects on the wave propagation, was not addressed in this work. The 3D case was also not studied. Thus *this work does not exclude the possibility of strong reservoir-specific spectral anomalies* and should be followed by further research.

Acronyms

BEM Boundary Element Method

CLG Conjugate Linear Gradient algorithm

DFA Diffuse Wavefield Assumption

FK A family of grid-search techniques in the frequency-wavenumber domain (e.g. Capon et al., 1967) to determine the parameters of the strongest plane waves propagating across an array of seismometers

FWI Full-Waveform Inversion

HM Homogeneous model

HMT Hydrocarbon Micro-Tremors

HRFK High-Resolution FK, a modification of FK

LFPS Low-Frequency Passive Seismics, as a generic name for methods relying on hydrocarbon micro-tremors and other similar effects

LOH Layer-over-halfspace model

MSM Microseismic Sounding Method

MUSIC MUltiple SIgnal Classification

MUSIC-CC Workflow which consists of applying MUSIC to a shot gather obtained by ambient noise cross-correlation

PSD$_i$ Power Spectral Density of the signal on the component i; z-component is the default when the component is not specified. Defined in equation 2.2 for individual time windows.

SEM Spectral-Element Method

SPECFEM2D open-source SEM solver (`https://github.com/geodynamics/specfem2d` Komatitsch et al., 1999)

UGS Underground Gas Storage

V/H Vertical-to-horizontal spectral ratio of the ambient noise, defined as a ratio of the power spectral densities (equation 2.4)

WIFF Wave-Induced Fluid Flow

WIGED Wave-Induced Gas Exsolution-Dissolution

Contents

Acronyms **5**

1 Introduction **11**
1.1 Industrial context of the PhD thesis: Storengy and natural gas storage 12
1.2 Low-Frequency Passive Seismics: a review 15
1.2.1 Quick overview 15
1.2.2 Case studies and controversial issues 18
1.2.3 A lacking mechanism 33
1.3 Propagation modes of elastic waves: basic properties 40
1.3.1 Wave equation and Green's function 40
1.3.2 Body waves 41
1.3.3 Surface waves 43
1.3.4 Coda waves 52
1.4 Which methods for ambient noise ? 52
1.4.1 Composition of the ambient noise 52
1.4.2 Surface wave analysis methods 54
1.5 Rayleigh wave scattering as a possible mechanism for hydrocarbon microtremors 58
1.5.1 Reasons for analysing Rayleigh wave scattering 59
1.5.2 Specific features of near-receiver scattering 60
1.5.3 Numerical modelling 61
1.5.4 Modelling strategy 62
1.6 Effective properties of a gas reservoir 64
1.6.1 Concept of effective medium 65
1.6.2 Elastic modelling: the Gassmann relation 66
1.6.3 Attenuation due to patchy saturation 71
1.6.4 Attenuation due to Wave-Induced Fluid Flow in heterogeneous media 75
1.6.5 Conclusions 78
1.7 Contributions and outline 79
1.7.1 PhD work organisation 79
1.7.2 Manuscript outline 80
1.7.3 Main contributions 82

2 Experimental observations of amplitude distortions **84**
2.1 Dataset 1: Chémery underground gas storage 86
2.1.1 Site presentation 86

2.1.2 Gas accumulation evolution . . . 87
2.1.3 Seismic network . . . 89
2.1.4 Signal processing . . . 93
2.1.5 Spectral anomalies and attribute profiles . . . 95
2.1.6 Interpolated attribute maps . . . 100
2.1.7 Automated spectrum classification . . . 103
2.1.8 Main conclusions . . . 106
2.2 Dataset 2: Saint-Illiers underground gas storage . . . 108
2.2.1 Site presentation . . . 108
2.2.2 Seismic network and gas accumulation evolution . . . 108
2.2.3 Spectral anomalies . . . 112
2.2.4 Active or passive anomaly source ? . . . 115
2.2.5 Attribute evolution vs. gas stock . . . 117
2.2.6 Main conclusions . . . 119
2.3 Small-scale field experiment . . . 120
2.3.1 Experimental setup . . . 120
2.3.2 Processing and results . . . 120
2.4 Dataset 3: Muara Laboh high-enthalpy geothermal field . . . 122
2.4.1 Context and introduction . . . 124
2.4.2 Acquisition over a geothermal site . . . 125
2.4.3 Results . . . 125
2.4.4 Wavefield composition . . . 130
2.4.5 Discussion . . . 132
2.4.6 Main conclusions . . . 132
2.5 General conclusion on the field data . . . 134

3 Ambient noise characterisation 135
3.1 Introduction . . . 136
3.2 Data . . . 138
3.3 Methodology . . . 141
3.3.1 Aliasing and resolution limits . . . 142
3.3.2 Cross-correlation . . . 142
3.3.3 MUSIC . . . 144
3.4 Results . . . 146
3.4.1 Dispersion plots . . . 146
3.4.2 Back-azimuths . . . 148
3.4.3 Day/night variation . . . 151
3.4.4 Inversion for V_S . . . 153
3.5 Discussion . . . 156
3.5.1 Ambient noise sources . . . 156
3.5.2 Methodology and inversion . . . 159
3.5.3 Benefits of wavefield characterisation . . . 160
3.6 Conclusion . . . 161
Appendices . . . 162
3.A Group and phase velocity . . . 162
3.B MUSIC and automated signal subspace determination . . . 162
3.C Synthetic tests of the MUSIC algorithm . . . 165

3.C.1 Dispersion curves retrieval . . . 167
3.C.2 Back-azimuth retrieval . . . 168
3.D Spectral-element simulations of the wavefield composition . . . 168
3.E L_g phase and overtones . . . 170

4 Elastic forward modelling of amplitude distortions 177
4.1 Introduction . . . 179
4.2 Definition of the observables . . . 180
4.3 Numerical forward-modelling . . . 182
4.3.1 Background models . . . 182
4.3.2 Inclusion parameters . . . 183
4.3.3 Source distributions and average PSD . . . 184
4.4 Results: modelled spectral anomalies . . . 185
4.4.1 Background model influence . . . 186
4.4.2 Incident wavefield composition influence . . . 189
4.4.3 Inclusion geometry influence . . . 191
4.4.4 Structural effects vs. reservoir effects . . . 193
4.5 Interpretation of the spectral anomalies with Born modelling . . . 195
4.5.1 Main elements of the Born approximation . . . 196
4.5.2 Born modelling of the reservoir effects . . . 196
4.5.3 Structural effects: limits of the Born approximation . . . 201
4.6 Discussion . . . 203
4.6.1 Objectives and scope . . . 203
4.6.2 Amplitude anomaly strength and behaviour . . . 204
4.6.3 Single-scattering interpretation . . . 204
4.6.4 Role of numerical modelling . . . 205
4.6.5 Potential for practical applications . . . 206
4.6.6 Limitations . . . 207
4.7 Conclusion . . . 208
Appendices . . . 209
4.A Born sensitivity kernels derivation . . . 209
4.B Born sensitivity kernels implementation and validation . . . 211
4.C Dominant frequency trends under modal summation . . . 214

5 Characteristics of Rayleigh wave diffraction 218
5.1 Introduction . . . 221
5.2 Born approximation, sensitivity kernels and radiation patterns . . . 222
5.3 Equivalent source terms . . . 225
5.3.1 Numerical illustration . . . 225
5.3.2 Rayleigh waves . . . 228
5.3.3 SV waves . . . 229
5.3.4 P waves . . . 232
5.4 Analytic radiation patterns in unbounded space . . . 232
5.5 Numerical radiation patterns in 2D half-space . . . 234
5.5.1 Validation of the numerical implementation . . . 234
5.5.2 Free-surface effects . . . 238
5.6 Discussion . . . 243

5.7 Conclusion . . . 246
Appendices . . . 248
5.A Analytic expressions for equivalent source terms . . . 248
5.A.1 Rayleigh waves . . . 248
5.A.2 SV waves . . . 249
5.A.3 P waves . . . 250
5.B Analytic sensitivity kernels derivation in unbounded space . . . 250
5.C Pattern asymmetry for a ρ-contrast . . . 252

6 Spectral amplitude imaging with Rayleigh waves 255
6.1 Introduction . . . 256
6.2 Forward and inverse problems . . . 258
6.2.1 Investigated models . . . 259
6.2.2 Observables . . . 261
6.2.3 Inversion . . . 266
6.3 Multi-parameter inversion: choosing the best parametrisation . . . 267
6.4 Inversion results . . . 273
6.4.1 Regularisation . . . 273
6.4.2 Fundamental mode vs. overtones . . . 274
6.4.3 Standard vs. PSD observable . . . 277
6.4.4 Comparison with active-source P waves . . . 279
6.5 Discussion . . . 281
6.5.1 Why do overtones perform better ? . . . 282
6.5.2 Possible real-data applications . . . 283
6.5.3 Ambient wavefield characterisation . . . 284
6.5.4 Forward modelling and inversion: improvement potential . . . 285
6.6 Conclusion . . . 286
Appendices . . . 287
6.A Sensitivity kernels under different parametrisations . . . 287
6.B Quantitative criteria for assessing parametrisation quality . . . 290
6.B.1 Spatial crosstalk . . . 290
6.B.2 Inter-parameter crosstalk . . . 291
6.B.3 Conditioning . . . 292
6.B.4 Sensitivity . . . 293
6.C Quantitative criteria for assessing inversion quality . . . 293
6.D Objective function sensitivity to the background model . . . 294

7 Conclusions and perspectives 297
7.1 Experimental feasibility . . . 299
7.2 What to expect from the elastic modelling methodology ? . . . 299
7.3 Effects of stronger contrasts in the reservoir . . . 301
7.4 Improving the Rayleigh wave scattering modelling . . . 302
7.5 Diffuse or coherent incident wavefield ? . . . 305

Bibliography 308

Chapter 1

Introduction

Résumé (français)

Dans ce chapitre d'introduction, on présente dans un premier temps les motivations de ce travail de thèse, ensuite une revue bibliographique sur l'état de l'art des techniques dites de *sismique passive basse fréquence*, basées sur l'amplitude du bruit ambiant, et enfin l'approche de modélisation adoptée dans la suite du manuscrit.

La motivation initiale est le monitoring des stockages de gaz naturel en aquifère de Storengy, leader européen dans ce domaine. Le concept du stockage en aquifère est brièvement expliqué. On présente également les premiers résultats montrant les anomalies du bruit ambiant enregistrées au-dessus de l'un des sites de stockages, qui constituent l'élément de départ de la thèse.

L'approche consistant à interpréter ces anomalies comme indicatrices de la présence d'hydrocarbures est très controversée dans la littérature scientifique. Une revue bibliographique de cette controverse est effectuée, avec d'un côté des cas d'étude encourageants, et de l'autre des contre-exemples (corrélation des anomalies avec des paramètres indépendants de la présence d'hydrocarbures). En somme, il est difficile de faire la part des choses sans comprendre le mécanisme physique à l'origine des anomalies observées. L'investigation de ces mécanismes à l'aide de la modélisation numérique est l'axe principal de ce travail.

Les différents mécanismes proposés dans la littérature sont analysés afin de positionner notre approche par rapport à ceux-ci. On justifie le choix d'un modèle basé sur la diffraction élastique des ondes de surface en tant que point de départ. Pour faciliter la compréhension, quelques notions de base concernant la propagation des ondes, et en particulier celle des ondes de surface, sont rappelées. Enfin, on donne des pistes pour complexifier ce modèle de départ, notamment en rajoutant un comportement visco-élastique au sein du réservoir.

Le chapitre se conclut par un plan explicitant le rôle des différentes parties du manuscrit. Une liste détaillée des contributions (logiciels, brevets, publications) issues de la thèse est dressée.

1.1 Industrial context of the PhD thesis: Storengy and natural gas storage

The present work is funded by Storengy, a company of Engie, the French leader of underground storage of natural gas. Gas storage is important for insuring a vital stock for the country, in case of political instabilities or extreme climate events (such as anomalously cold winters). It also allows to avoid purchasing gas at too expensive rates in winter, when the demand is high. This PhD work is part of a research project initially aiming to improve gas storage monitoring, referred to as *Low Frequency Passive Seismics* (LFPS). Extensions of the method to geothermal applications in steam-bearing zones are also considered, as Storengy progressively invests this emerging domain. One dataset recorded in the context of geothermal exploration is presented in this PhD. However, Underground Gas Storage (UGS) remains the main focus of this manuscript.

An example of a typical aquiferous UGS on which the passive seismic method should be applied is the one in Chémery, where gas is injected into a sandstone layer at a depth of 1100 m. Fig. 1.1 offers a simplified view of how such a storage is operated. Gas is injected by producing wells (also injecting wells) and trapped in an anticlinal structure. Monitoring wells at the periphery of the structure (spill point monitoring wells) make sure there is no lateral gas leakage. Other wells are located at the top of the upper aquifer (upper aquiferous monitoring wells) to make sure there is no vertical gas leakage, for example through faults or damaged well casings.

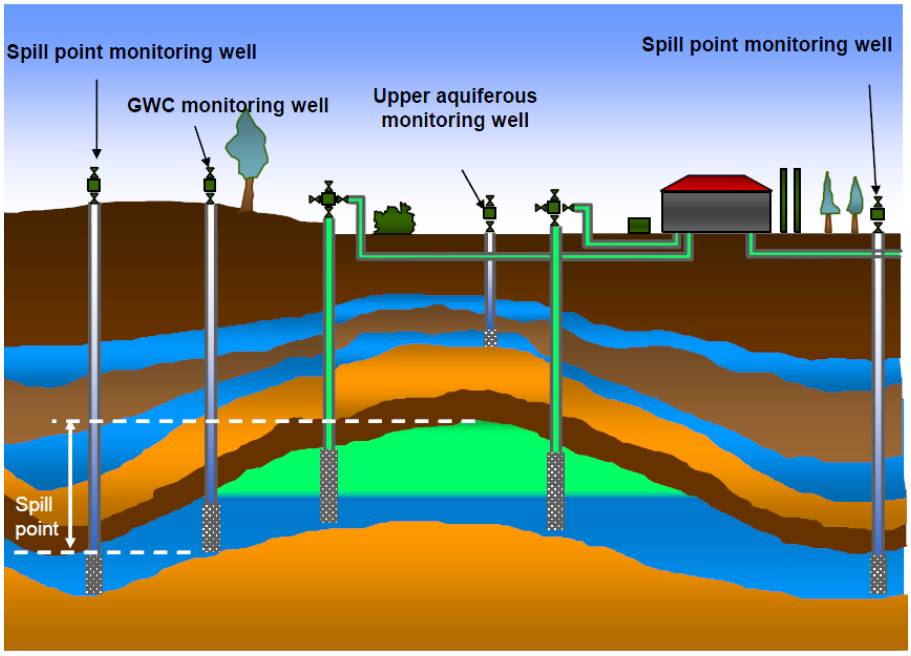


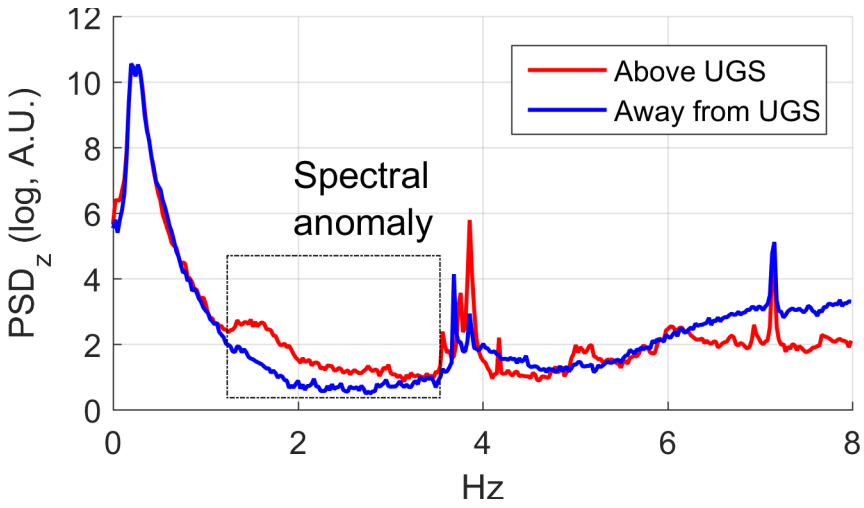
Figure 1.1: Schematic view of an underground gas storage (source: Storengy).

Drilling a controlling well is a costly operation, which explains why measures from controlling wells are sparse. Gas arrival in such a well must absolutely be avoided, because of evident environmental reasons. Gas bubble control is achieved by an appropriate injection strategy. At present, the latter is guided by the available sparse

well data, namely the gas pressure and the gas/water contact depth. Continuously mapping the gas front would be an invaluable help for piloting producing wells, such as to control the maximum lateral extent of the gas bubble.

The purpose of this PhD thesis is to investigate the physical basis and the resolution of a method reported to detect underground hydrocarbon presence from passive seismic data. Broadband seismometers are deployed above and further away from a gas reservoir, and record the ambient noise. The processing is based on the power spectral density (PSD) of ambient noise recordings. It appears that the latter often exhibits a low-frequency anomaly at the surface above hydrocarbon-bearing zones. Prior to the PhD work, the subject of my master-degree internship in Storengy was to extract such an anomaly from a dataset recorded above the Chémery UGS. Several attributes characteristic of the low-frequency anomaly were defined. One of them, based on the slope of the vertical component power spectral density, yielded the anomaly map shown in Fig. 1.2 (see Chapter 2 for processing details). Visually, it seems to correlate well with the effective thickness of the gas column ($H_g = \sum_i s_i \Phi_i H_i$, with s_i the gas saturation, Φ_i the porosity and H_i the thickness of the different reservoir layers), estimated from the reservoir numerical model. Namely, the fault system at the south-west of the structure seems well spotted. It was decided to deeper investigate the nature of this anomaly, in order to find a method for its quantitative interpretation.

(a) Spectral anomaly pattern



(b) Attribute maps

(c) Known gas accumulation

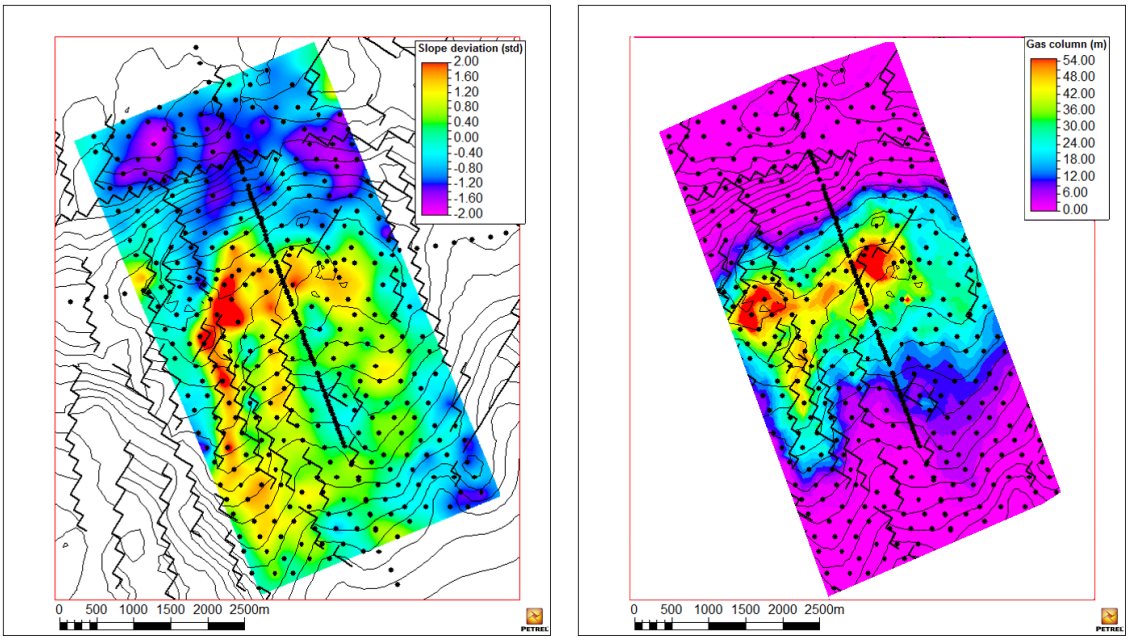


Figure 1.2: (a) Two PSD_z spectra recorded simultaneously above the gas bubble (red) and outside it (blue). Data from the Chémery UGS, median spectra for 4 hours of recording (00:00-04:00, 23/04/2010). (b) Map of an anomaly-based attribute above the Chémery UGS (average spectrum slope between 1.3 and 1.6 Hz). The isolines of the reservoir top boundary are shown with this lines. Faults are shown with thick lines. From Kazantsev (2015). (c) Map of the effective gas thickness in the Chémery UGS, based on a flow simulation constrained by well injection/production data.

This Chapter is organised as following. In the next section (1.2), I present a review of previous applications of this technique, interpretations proposed for the physical origin of the observed anomaly, as well as the associated controversial debate in the scientific community. Based on this review, I choose passive scattering as the hypothesis to investigate in my work. Then, in Section 1.3, follows an overview of different

types of elastic waves, with a particular focus on surface waves, as they are the central point of most of the developments presented in this manuscript. In Section 1.4 the composition of the ambient noise, such as decribed in the literature, is presented. I also make a short review of surface wave methods which have found applications in ambient noise analysis. In Section 1.5, I explain why I choose to study *Rayleigh-wave* scattering as a potential mechanism for explaining the observed amplitude anomalies, and analyse the consequences of this choice in terms of modelling strategy to be followed. As modelling a scattering process requires to define a heterogeneity, the plausible elastic and visco-elastic parameters for describing a gas reservoir are tackled in Section 1.6. Finally, in Section 1.7, I provide the outline of the manuscript and the list of the main contributions made during my PhD work.

1.2 Low-Frequency Passive Seismics: a review

One of the objectives of this section is to familiarise the reader with the LFPS thematic. For this purpose, a quick summary of this method's emergence and evolution is provided in Section 1.2.1. In Section 1.2.2, a more detailed analysis of the reported observations is proposed in order to understand the controversy around this subject and to list the main arguments. Finally, in Section 1.2.3, I focus on the mechanisms which were proposed as potential explanations for the observed effects, and choose one particular category among them as a guideline for the subsequent investigations.

1.2.1 Quick overview

Low-frequency effects associated to hydrocarbon reservoirs were first described in Soviet/Russian literature where seismic stimulation was used to enhance oil recovery (Beresnev & Johnson, 1994; Nikolaevskiy et al., 1996). Signals of higher amplitude were recorded at the surface above hydrocarbon-bearing zones, both during the active source emission period and, more surprisingly, also after it stopped (Kurlenya & Serdyukov, 1999, see Fig. 1.3a). Such effects typically spanned over the frequency range around 10-30 Hz. Some more references from the Russian/Soviet literature dedicated to this subject, often published in Russian language, can be found in Zhukov et al. (2007). Low-Frequency Passive Seismic (which I refer to as LFPS in this manuscript), in the sense of hydrocarbon exploration using ambient noise amplitude spectra, appeared as a subject with the discovery by Dangel et al. (2003) of amplification anomalies in the ambient noise spectra measured above 15 hydrocarbon reservoirs throughout the world. The affected frequency range was considerably lower than in the active case: 1.5-4 Hz (Fig. 1.3b).

(a) Post-stimulation effect (b) Ambient noise anomaly

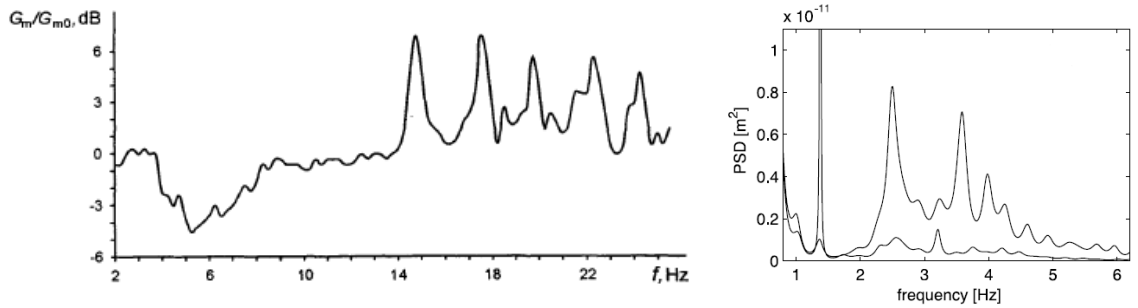


Figure 1.3: Comparison between a post-stimulation amplitude anomaly (a) and a passive ambient noise anomaly (b). (a) Ratio between the spectra computed 2 minutes after the end and 30 minutes before the start of the vibrational stimulation of the Mortymya-Teterevsky oil field (near Khanty-Mansiysk, Russia). From Kurlenya & Serdyukov (1999). (b) Ambient noise vertical component spectra in a gas field in central Italy. Upper line: above the gas accumulationn; lower line: 600 m from the first recording, outside the gas accumulation. From Dangel et al. (2003).

Dangel et al. (2003) used the term «tremor» to describe the supposed hydrocarbon signature in the ambient noise spectrum because of an «astonishing similarity» with volcanic tremors. The latter are usually associated with bubble dynamics in migrating magmatic fluids, though several theories exist concerning their precise nature (Ferrick et al., 1982; Ripepe & Gordeev, 1999; Urquizú & Correig, 1999, and other references in Dangel et al. (2003)). This analogy lead to a preliminary interpretation of the «hydrocarbon micro-tremor» (HMT) as a result of oil/gas bubble oscillation within water-wetted pores of the reservoir rock (Holzner et al., 2009; Frehner et al., 2009). In the context of hydrocarbon exploration, this method looked promising, since it could allow to get a hydrocarbon potential map at low cost, without using any active source of seismic emission. In the late 2000s - early 2010s, a technology based on HMT was widely promoted by the Spectraseis company under the name *HyMas* (hydrocarbon micro-tremor analysis, Graf et al., 2007). Several aspects of this technology were patented (Saenger, 2008; Saenger et al., 2009; Saenger, 2009; Podladchikov et al., 2010; Kelly et al., 2013), mainly concerning particular ways of extracting relevant attributes from the ambient noise, statistically treating them, and converting them into hydrocarbon potential maps. Another patent concerned reservoir imaging based on the time-reverse imaging (TRI) exploiting the HMT signals in the time domain (Saenger et al., 2010). Other companies developed technologies based on HMT. Namely, PGES (Petroleum & Geothermal Exploration Services) use(s/d) a similar approach, which they call IPSS (Infrasonic Passive Seismic Spectroscopy) (http://www.pgexploration.com/geophysical-services/passive-seismic/psspectroscopy). Also, the Geospectra IPDS ® technology (Infrasonic Passive Differential Spectrosopy) is reported as being patented by GeoDynamics Research Group (gww.uk.com), though I was unable to find the patent reference.

In 2009, Spectraseis launched a wide consortium (Low-Frequency Seismic Partnershift) in close collaboration with the Swiss Federal Institute of Technology (ETH Zürich). The aim of the consortium was to assess the physical nature of HMT,

through numerical modelling, sample tests in laboratory and repeated field surveys. This gave rise to a significant number of interesting publications (e.g. Steiner et al., 2008b; Lambert et al., 2009; Saenger et al., 2009; Quintal, 2012; Lambert et al., 2013). The subject rapidly became controversial, as some authors reported failures of the method for individual hydrocarbon fields. An intense debate arised on the pages of *Geophysical Prospecting* after the publication by Lambert et al. (2009) of the results of a passive survey above the Voitsdorf oil and gas field, Austria, and concluding to some degree of correlation between the ambient noise amplitude attributes and the actual location of hydrocarbon-bearing zones. The methods and conclusions of Lambert et al. (2009) were severely criticized by Green & Greenhalgh (2010), according to whom the claimed HMT was in fact a combination of artificial noise and shallow layer effects. Lambert et al. (2010) then replied to this criticism, insisting that neither the effects mentionned by Green & Greenhalgh (2010), nor the hydrocarbon-related nature of the «microtremors» could be excluded. Within the frame of the consortium program, Spectraseis recorded two time-lapse dense datasets over the Chémery underground gas storage (UGS), owned and operated by Storengy (formerly the Infrastructure branch of GDF Suez), the funding company of the present PhD work. The data were recorded when the gas content in the storage was at its minimum (April 2010) and maximum (November 2010) values. The advantage of the Chémery site was twofold:

- The cyclic injection/withdrawal of gas into the aquifer reservoir trapped in an anticline (depth$\sim$ 1100 m) should allow to check if the amplitude of HMT changes accordingly.

- A fully water-saturated geological trap, similar to the gas-bearing structure, is located about 7 km away, which should allow to check if HMT could be a structural effect.

The results of the first stage of this survey (April 2010, baseline) were presented in a conference paper (Duclos et al., 2011) and looked very promising: a HMT between 1.7-3.2 Hz was visible only above the gas reservoir, and the latter could be located at the correct depth by time-reverse imaging. Surprisingly, none of these results appeared in journal articles. To my knowledge, the time-lapse results in terms of HMT were not published. Instead, Riahi et al. (2013a) treated this dataset in terms of time-lapse variations of surface wave velocity anisotropy, which is yet a distinct subject. Based on the internal presentations provided by Spectraseis to Storengy, and my current results shown in Section 2.1, the change of HMT between the two surveys was not conclusive enough.

Progressively, the methods based on the supposed HMT were categorized as unreliable and somewhat mysterious by the geophysical community. The interest for HMT decreased (to my knowledge, no major case studies/modelling works since 2014), while Spectraseis turned towards micro-earthquake location methods as main activity (`www.spectraseis.com`). However, Storengy continued working internally on the datasets recorded in Chémery, as well as on new data recorded as part of this PhD work.

After this quick overview of the rapid emergence and decline of the HMT-related works, the next section offers a more detailed literature review in order to summarize previous field applications of the method and possible physical origins of HMT, as suggested by different authors. It also covers a very similar methodology, the so-called *Microseismic Sounding Method* (MSM), developed by A. Gorbatikov, V. Tsukanov and collaborators from the Russian Academy of Sciences, though these works are not directly linked to the hydrocarbon or the geothermal thematic.

1.2.2 Case studies and controversial issues

The pioneering publication by Dangel et al. (2003) observes spectral anomalies of similar shape and frequency content (1-4.5 Hz) on the vertical component ambient noise recordings above 15 oil and gas reservoirs at different locations throughout the world. Since then, several cases of successful HMT observations above hydrocarbon fields were reported. In their critical paper, Green & Greenhalgh (2010) stress that most of these works were published in conference papers (Holzner et al., 2005; Bloch & Akrawi, 2006; Birialtsev et al., 2006; Rached, 2006; Schmalholz et al., 2006; Lambert et al., 2007; Saenger et al., 2007; Steiner et al., 2007, 2008a; Mastrigt & Al-Dulaijan, 2008; Duclos et al., 2011) and «lightly reviewed journals» (Holzner et al., 2005; Graf et al., 2007; Walker, 2008). I selected six articles published in «fully peer-reviewed journals» for detailed analysis. They are summarised in Table 1.1. In Table 1.2, I also summarise four articles which reported failures of LFPS surveys. In this section, I attempt to identify the main arguments from both sides. The possible mechanisms of the HMT generation are addressed in Section 1.2.3.

Typical attributes correlated with hydrocarbons

Lambert et al. (2009) and Saenger et al. (2009) introduce several attributes to quantify the strength of HMT, which can be mapped over the survey area and presented as «hydrocarbon potential maps». I summarise these attributes below.

1. Vertical component power spectral density (PSD_z)

   The first attribute is directly based on the vertical component power spectral density (PSD_z), as suggested by Dangel et al. (2003). The PSD spectrum is estimated individually for each receiver by the Welch's method, i.e. averaging the spectra over many time segments, typically lasting about 40 s, in order to get enough resolution in the frequency domain down to the mircroseismic peaks (0.1-0.2 Hz), and also enough time segments to obtain a smooth average spectrum from only several hours of recording. Simple strategies of time segment selection are presented in Goertz et al. (2012) and Riahi et al. (2013b), while a thourough study of advanced pre-processing strategies is performed by Lambert et al. (2011). Once a smooth spectrum has been obtained, it is integrated within the frequency band where the «anomaly» is observed (presumably due to HMT). The bounds of the integration, f_1 and f_2, are fixed manually based on the visual inspection of the anomaly pattern. In this manuscript, I introduce

the following notation for this attribute:

$$\mathrm{PSD}_z|_{f_1}^{f_2} = \int_{f_1}^{f_2} \mathrm{PSD}_z(f)df \ . \tag{1.1}$$

As a variant, Saenger et al. (2009) use a variable lower bound of integration, corresponding to the first spectral minimum after 1 Hz, which I write as $\mathrm{PSD}_z|_{f_1+}^{f_2}$.

2. V/H spectral ratio

This attribute is the inverse of the well-known H/V ratio introduced by Nakamura (1989) to measure the thickness of the soft near-surface layers via the value of the peak frequency. There are several possible implementations to estimate the spectral ratio, summarised by Saenger et al. (2009) in their fig. A-1. The latter are expected to influence the absolute values, but not the relative trends between different receivers, neither the values of the peak frequencies. In the HMT context, Lambert et al. (2006, 2007) observe that the positive PSD_z anomalies described by Dangel et al. (2003) can be accompanied by positive anomalies of the V/H ratio, indicating a preferentially vertical polarisation of HMT. To quantify this effect, Lambert et al. (2009) propose to use the maximum V/H value within the HMT frequency band. Other authors integrate the V/H ratio over this frequency band (e.g. Goertz et al., 2012), or just visually examine the shape of the spectral curves (e.g. Martini et al., 2013). No matter which precise attribute is extracted, I introduce the notation $\mathrm{V/H}|_{f_1}^{f_2}$ to indicate that the V/H spectral ratio is studied between f_1 and f_2. Saenger et al. (2009) suggests that the V/H ratio can be considered as anomalous when it is greater than 1, so they introduce a modified integral attribute:

$$\mathrm{V/H}|_{f_1}^{f_2}(>1) = \int_{f\in[f_1,f_2]\left|\mathrm{V/H}(f)>1\right.} \mathrm{V/H}(f)df \ , \tag{1.2}$$

which is zero if the V/H remains below 1 between f_1 and f_2. Note that f_1 and f_2 are not necessarily exactly the same for PSD and V/H attributes for a given dataset (e.g. Saenger et al., 2009).

3. Peak frequencies of PSD_z and PSD_x

Introduced by Lambert et al. (2009), they are the frequencies corresponding to the maximum PSD_z and PSD_x values between f_1 and f_2, which I write respectively as $f_{\max}^{\mathrm{PSD}_z}|_{f_1}^{f_2}$ and $f_{\max}^{\mathrm{PSD}_x}|_{f_1}^{f_2}$. They are important attributes in the works by M.-A. Lambert and his co-authors, who suggest that the peak frequency of HMT should be maximum above the reservoir and shift towards lower values outside it. They support this conjecture by modelling the spectra at the surface for an embedded vertically polarised source (Lambert, 2010, his Chapter 4 and Appendix C), and extend this model to a random source distribution in Lambert et al. (2010).

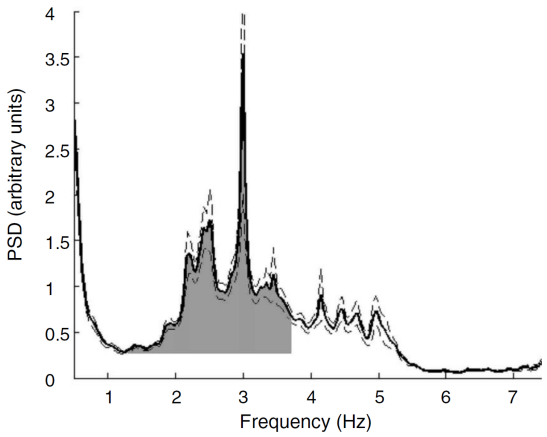
In Fig. 1.4 I provide illustrations of the different attributes, taken from the corresponding papers. In addition to these attributes, the signal polarisation within the

HMT frequency band is often invoked (e.g. Saenger et al., 2009), typically measured following Jurkevics (1988): rectilinearity, dip and back-azimuth are obtained from the particle motion eigen-decomposition at each time sample. However, these attributes are rarely plotted in a map view, and are mostly analysed for individual stations in order to confirm the preferentially vertical polarisation of HMT above the reservoir, already deduced from the V/H ratio. Based on my own experience, the absence of polarisation maps could be due to the instability of the polarisation attributes in both space and time at frequencies above 1 Hz. In Fig. 1.5, I show examples of attribute maps/profiles which are visually correlated with hydrocarbon-bearing zones, for the oil and gas field in Voitsdorf, Austria (Lambert et al., 2009), and the gas field in Burgos, Mexico (Saenger et al., 2009). In the case of Saenger et al. (2009), six additional wells were drilled after the LFPS survey, four of which encountered hydrocarbons. The two dry wells were close to the non-anomalous zones of the HMT attribute (Fig. 1.5b). This is representative of the typical output expected from a LFPS survey: a map of the hydrocarbon potential in order to optimise the location of the new wells to be drilled. A more quantitative approach is introduced by Riahi et al. (2013b), who demonstrate that the statistical correlation between the attribute ($\mathrm{PSD}_z|^3_{1.5}$) and the known gas accumulation (Jonas field, Wyoming, USA) is stronger than the correlation with other possible factors which could explain the supposed HMT, such as the ground elevation and the well density (as a proxy for human activity and production noise). They also show that the relationship between $\mathrm{PSD}_z|^3_{1.5}$ and the gas accumulation in their field is monotonic, but not linear. Thus they use a correlation rank which measures the monotony of the relation between the statistical variables (Spearman rank), instead of the traditional Pierson correlation coefficient (see references in Riahi et al. (2013b)). Further following the statistical approach, a Bayesian method of attribute processing is patented by Spectraseis (Kelly et al., 2013).

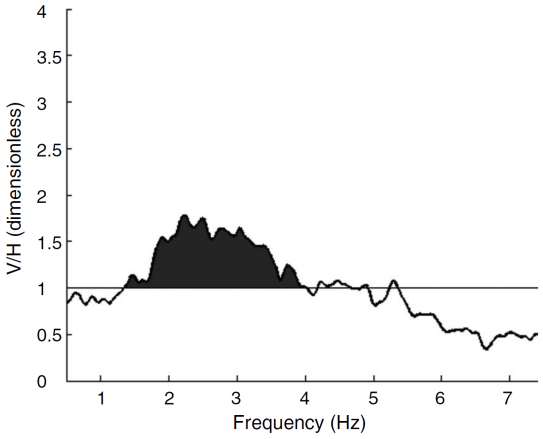
Cases of attribute correlation with other parameters

On the other hand, Hanssen & Bussat (2008) report several possible misinterpretations of non-hydrocarbon signals as HMT, based on a dataset recorded in Libya. First, they show that human-generated noise, especially the production noise in the vicinity of the wells, can take place within the HMT frequency band (several Hz), with an associated spectral signature similar to the HMT peak. Second, processing only the quietest noise windows (low artificial noise, typically night-time), they conclude to a correlation between the attribute $\mathrm{PSD}_z|^3_2$ and the ground elevation. Ali et al. (2010) report a false-positive HMT signal outside the known hydrocarbon-bearing zone in the UAE. They observe that the evolution of the energy is uncorrelated between the supposed HMT (2.5 Hz) and the oceanic microseism (0.25 Hz), contradicting the assumption of the microseism being the trigger of a non-linear interaction generating HMT (Holzner et al., 2009, see next section). They observe a strong variation of the microtremor energy between day- and night-periods, similarly to Hanssen & Bussat (2008), which indicates that human activity contaminates the corresponding frequency band. By performing array-processing, they measure the typical velocities at the microtremor frequency (2.5 Hz), and find them consistent with Rayleigh wave propagation (about 1 km/s, i.e. not body waves propagating vertically from the

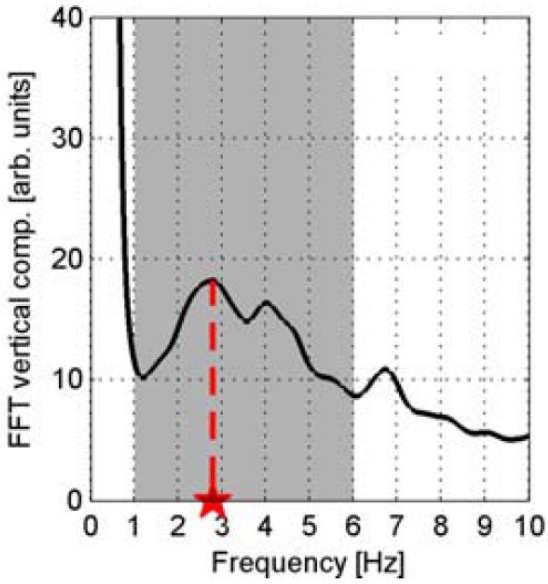
(a) Attribute $\mathrm{PSD}_z|_{1+}^{3.7}$



(b) Attribute $\mathrm{V/H}|_1^6(>1)$



(c) Attribute $f_{\max}^{\mathrm{PSD}_z}|_1^6$



(d) Attribute $f_{\max}^{\mathrm{PSD}_x}|_1^6$

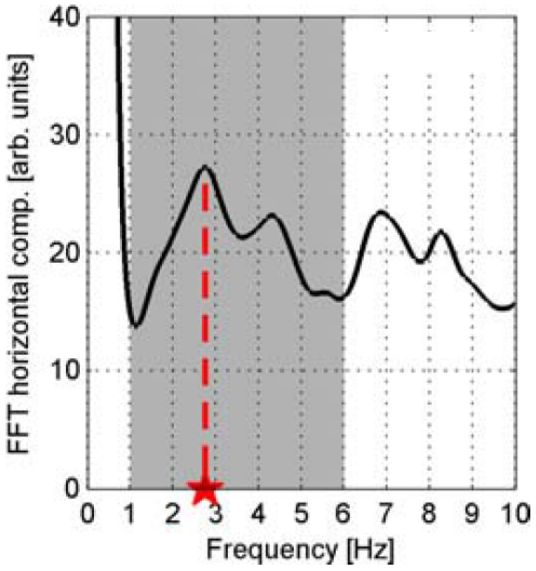
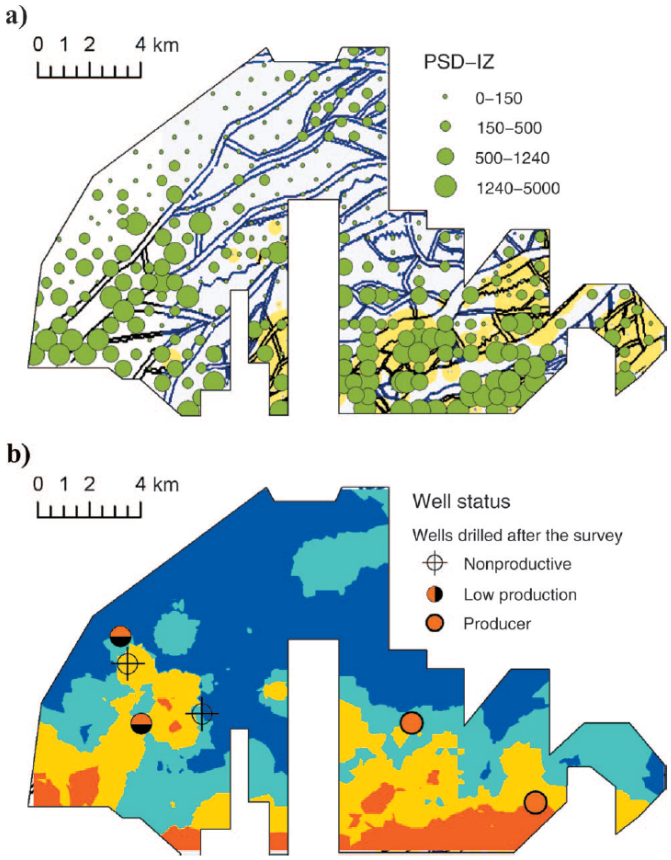


Figure 1.4: Illustration of the different attributes. (a)-(b) from Saenger et al. (2009); (c)-(d) from Lambert et al. (2009).

Hydrocarbon potential maps after Saenger et al. (2009)



Hydrocarbon potential profiles after Lambert et al. (2009)

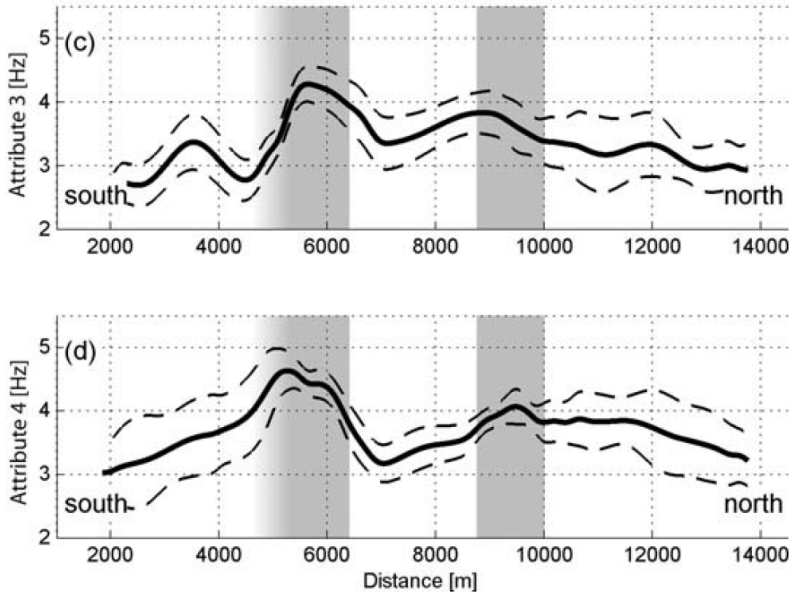


Figure 1.5: Hydrocarbon potential maps from a dense array (top, Saenger et al., 2009) and from a dense linear profile (bottom, Lambert et al., 2009) of seismometers. Plotted attributes: (a) $\mathrm{PSD}_z|_{1+}^{3.7}$ (green circles); (b) $\mathrm{PSD}_z|_{1+}^{3.7}$ (interpolated coulours); (c) (resp. (d)) $f_{\max}^{\mathrm{PSD}_z}|_1^6$ (resp. $f_{\max}^{\mathrm{PSD}_x}|_1^6$), dashed lines indicate the corresponding plus/minus 1 standard deviation. Field information: (a) drainage radii of producing wells (yellow areas) and reservoir fault system (blue lines); (b) status of wells drilled after the survey; (c-d) reservoir extent (grey areas).

reservoir). They suggest that the spectral amplitude pattern, and namely the H/V ratio, can be interpreted in terms of shallow layer resonance, or standard site effets (Nakamura, 1989; Bard, 1999; Bonnefoy-Claudet et al., 2006a). Green & Greenhalgh (2010) propose a sceptical review of the HMT-related publications, where they criticise the HMT attributes introduced by Lambert et al. (2009) for their instability (namely the peak frequency of the individual component spectra), and also reject the HMT method as a whole, claiming that the overwhelming part of the ambient noise energy between 1 and 10 Hz can be explained by surface wave propagation and well documented site effects (i.e. the previously mentionned Nakamura, 1989; Bard, 1999, and other references in Green & Greenhalgh (2010)). Martini et al. (2013) report the capital influence of the geological structure in the HMT frequency band, based on a case study in Uganda. While they find some correlation between oil-bearing zones and HMT attributes, they also encounter a false-negative attribute value in an oil-bearing zone. They invert the H/V ratio at individual stations assuming it can be explained by the Rayleigh wave ellipticity. The inverted ground models appear mutually consistent, and also consistent with the available well data. The authors conclude that the spectral H/V (or equivalently V/H) ratio, including its peaks and troughs, can be explained in terms of ambient wavefield propagation in a layered (in their case, dipping) structure, rather than by the presence of hydrocarbons. They also insist that the presence of day/night variations at frequences above 1.2 Hz indicates a human origin of the noise in this frequency band, making it unusable as a hydrocarbon indicator. This is in contrast to some other authors, who think that the contamination by human activities can be sufficiently reduced by selecting the quietest time windows during night-time (Goertz et al., 2012; Riahi et al., 2013b). Finally, a geothermal application is attempted by Woith et al. (2014) in a gas-reach field in Turkey. Their initial assumption is that the dynamics of the water steam bubbles in a hot aquifer should not be fundamentally different from the methane bubbles in a conventional gas field. The authors detect a strong signal around 6 Hz above the center of the geothermal reservoir, with a high associated V/H ratio value. However, this signal stops abruptly when some machinery is turned off during several hours, eliminating the option of a fluid-induced microtremor. The authors do not provide any spectral analysis results for the quiet hours of the recording (i.e. when the machinery is turned off).

Time-reverse imaging

An important breakthrough is achieved by Steiner et al. (2008b), who apply time-reverse imaging (TRI), a technique based on the acoustic time reversal theory (e.g. Fink et al., 2000), to locate the source of the HMT recorded in Voitsdorf. The source is found at the known position of the hydrocarbon-bearing zones. The idea of the time reversal theory is as following. If a dense enough array of transducers surrounds a source of acoustic emission, the time-reversed recordings re-emitted at the position of each transducer collapse at the source location (retro-propagation). In imaging applications, the retro-propagation is performed by numerical resolution of the time-reversed wave equation, so a velocity model must be available as an input. In seismology, when the receivers are at the surface, it is impossible to surround a source in the sub-surface. Still, Steiner et al. (2008b) perform synthetic tests which

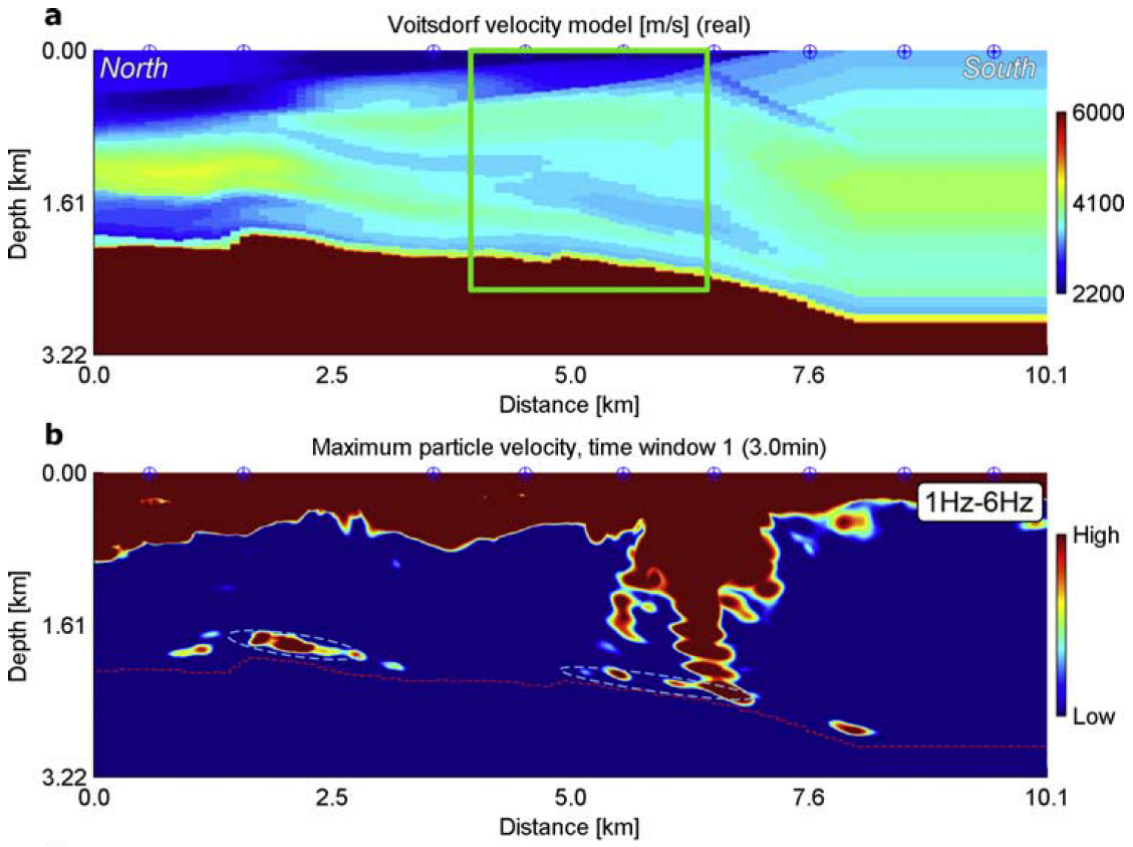


Figure 1.6: Result of 2D TRI after Steiner et al. (2008b) for the Voitsdorf field (Austria). (a) Velocity model (V_p). (b) Time-reversed image. The dashed ellipses show the two known reservoirs. They correspond to the gray shadowed areas in Figs 1.5 (c) and (d).

demonstrate the possibility of using a 2D finite-difference retro-propagation from a dense receiver line to re-focus the energy in an embedded seismic emission zone beneath the receiver line. Strong artifacts are also present in the upper part of the model because of the violation of the time-reversal theoretical hypotheses is such a configuration. In the real-data application at Voitsdorf, Steiner et al. (2008a) obtain severe near-surface artifacts, but the reservoir remains visible (Fig. 1.6). A key element in TRI is the choice of the *imaging condition*, which measures the «focus» in each cell of the model. Steiner et al. (2008b) use the maximum absolute particle velocity as imaging condition. According to the synthetic study by Steiner & Saenger (2012), the maximum energy provides better results. They also demonstrate that 2D retro-propagation of signals obtained with 3D forward modelling yields acceptable results in terms of focusing, though some characteristic artifacts appear. Artman et al. (2010) perform a synthetic study using the auto-correlation of the wavefield at zero time-lag (which is equivalent to the total energy integrated over all time) as imaging condition. They perform wavefield separation in the image domain, which consists of extracting P- and S- wave energies by taking respectively the divergence and the curl of the numerically retro-propagated signals. By modifying the imaging condition, which becomes a correlation of P-P, S-S or P-S wavefields, they retrieve the radiation pattern at the source, and thus the source mechanism. A further improvement is

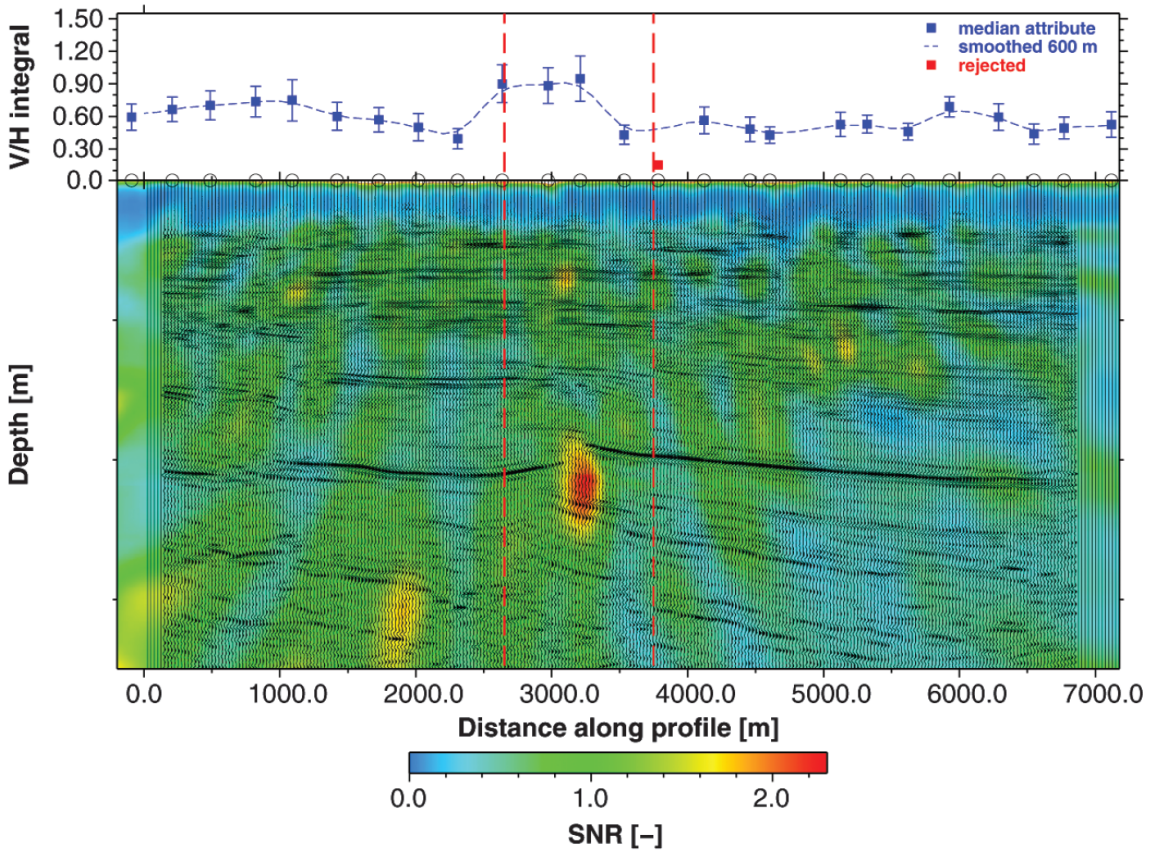


Figure 1.7: Result of 2D TRI after Goertz et al. (2012) for an oil field in Germany. The upper panel shows the $V/H|_1^{3.5}$ attribute profile. The lower panel shows the signal-to-noise estimate of the time-reversed image, overlayed onto the depth-migrated seismic section along the receiver line.

achieved by Witten & Artman (2011), who propose a method to eliminate the near-surface artifacts in the time-reserved images, which generated strong contamination in both synthetic and real-data examples presented by Steiner et al. (2008b). The idea is to separately retro-propagate random signals obtained with a (e.g. gaussian) noise model, and to deconvolve the time-reversed image of the true data by a smooth version of the time-reversed image of the random signals. In the resulting image, the true focus is much better visible than in the raw time-reversed image. This is referred to as *signal-to-noise estimate of the time-reversed image*. Adding this additional step to the workflow described by Artman et al. (2010), the authors apply TRI to two passive datasets in Germany and Egypt. 2D retro-propagation with a modified P-S imaging condition is applied to the German dataset, using three minutes of recording within the quietest night-time period, band-passed between 1 and 4 Hz, i.e. within the frequency band of HMT. A strong focus is obtained in the signal-to-noise image, perfectly collocated with the geological fault bounding the reservoir (Fig. 1.7, lower panel). According to the authors, this spot is also collocated with the known reservoir. However, the same figure is shown in Goertz et al. (2012), who mention that «the image focus is somewhat shallower than the reservoir location», which they first interpret as a resolution limit due to the too small array aperture, and finally also consider the option of a diffraction event on the fault itself, independant from

the reservoir. Goertz et al. (2012) thoroughly analyse the «German dataset» presented by Witten & Artman (2011). They show that a V/H ratio anomaly is present above the focus of the retro-propagated noise, consistently with the reservoir location (Fig. 1.7, upper panel). They present a detailed analysis of the TRI procedure, and show that the well-focused image obtained for the 3 minute time interval is due to constructive stacking of subsequent 10 second intervals, in which the correct focus is less prominent. Here, one could wonder why only a 3 minute time interval was extracted from a whole night of recordings. If HMT is stationary, taking longer (quiet) time windows should further enhance the SNR of the final image. Finally, Witten & Artman (2011) apply a TRI algorithm with 3D retro-propagation to their Egypt dataset, and manage to locate a 3D emission zone at a depth of more than 2 km, consistently with the lateral position of a discovery well and the vertical position of geological traps. The authors do not mention precisely the depth of the hydrocarbon discovery, nor do they provide any detail about the length of the retro-propagated time series, apart they consist of «night-time selections». To my knowledge, this is the only reported case of 3D TRI applied to HMT data.

As a first summary, the two main approaches to exploit the HMT are:

- Constructing attribute maps at the surface and supposing them representative of the fluid distribution in depth.
- Retro-propagating the signal in the time domain to get idea of the depth of the microtremor source.

Anticipating on the few next chapters, my work consisted of developing a third type of approach: *imaging an attribute defined in the frequency domain* via an inverse problem (see Chapter 6).

Before discussing the possible mechanisms of HMT generation (Section 1.2.3), I further summarise and discuss the arguments on both sides of the debate on whether HMT do or do not exist.

Table 1.1: Summary of positive HMT observations published in fully peer-reviewed journals since Dangel et al. (2003).

Publication	Location	Type	Depth	Positive attributes	Suggested Mechanism
Steiner et al. (2008b)	Voitsdorf, Austria	oil/gas	2 km	2D TRI$\|_1^6$	bubble oscillation
Lambert et al. (2009)		idem.		$f_{\max}^{\mathrm{PSD}_z}\|_1^6$ $f_{\max}^{\mathrm{V/H}}\|_1^6$	bubble oscillation (vertically polarized)
Saenger et al. (2009)	Burgos, Mexico	gas	2 km	$\mathrm{PSD}_z\|_{1+}^{3.7}$ $\mathrm{V/H}\|_1^6(>1)$ $f_{\max}^{\mathrm{PSD}_z}\|_{1.5}^{3.7}$ polarization (not mapped)	bubble oscillation (vertically polarized, triggered by the microseims)
Witten & Artman (2011)	Unknown field, Egypt	oil	2.5 km	3D TRI$\|_2^6$	(not discussed)
Goertz et al. (2012)	Speyer, Germany	oil	2 km	$\mathrm{V/H}\|_1^{3.5}$ 2D TRI$\|_1^4$	passive scattering (viscoelastic or elastic) of vertically incident body waves
Riahi et al. (2013b)	Wyoming, USA	gas	2.4-2.8 km	$\mathrm{PSD}_z\|_{1.5}^3$	viscoelastic passive scattering or bubble oscillation

Table 1.2: Summary of reported HMT failures

Publication	Location	Type	Depth	Tested attributes	Evidence
Hanssen & Bussat (2008)	Murzuq Basin, Libya	oil	?	$\mathrm{PSD}_z\vert_2^4$	correlation with human activity in time; correlation with surface topography in space (quiet periods)
Ali et al. (2010)	Abu Dhabi, UAE	oil	?	$\mathrm{PSD}_z\vert_2^3$	false-positive attribute values; correlation with human activity in time; microseisms not correlated in time with HMT (i.e. not the trigger); no body waves detected (array processing);
Martini et al. (2013)	Albertine Graben, Uganda	oil	< 1 km	$\mathrm{PSD}_z\vert_{0.5}^{1.2}$ $\mathrm{V/H}\vert_{0.25}^{5}$	false-negative attribute values; attribute correlation with basin thickness; no body waves detected above 0.5 Hz; V/H ratio well explained by a layered model
Woith et al. (2014)	Heybeli, Turkey	geoth. steam	180 m	$\mathrm{PSD}_z\vert_{6-}^{6+}$ $\mathrm{V/H}\vert_{6-}^{6+}$	attribute due to machinery (spectral line)

Argument summary

Both successful and unsuccessful reported LFPS surveys are summarised respectively in Tables 1.1 and 1.2. According to this review, the vision on the «optimistic» side, i.e. the authors who suggest that HMT could indeed serve as hydrocarbon indicators despite some uncertainties, can be summarised in the following few points.

1. The primary manifestation of HMT is the PSD_z positive anomaly somewhere between 1 and 6 Hz above reservoirs (Dangel et al., 2003; Saenger et al., 2009; Riahi et al., 2013b).

2. It was demonstrated that the attribute spatial distribution was, to some extent, correlated with known hydrocarbon resources (mainly Riahi et al. (2013b), but also Lambert et al. (2009), Saenger et al. (2009) and Goertz et al. (2012)). From these studies, I would also conclude that a spectral peak at an individual receiver is not necessarily an indicator in itself, but its spatial concentration in a limited zone is.

3. V/H ratio positive anomalies often recorded in the frequency band of the PSD_z anomaly indicate that HMT mainly consists of vertically polarised P-waves originating from the reservoir (Lambert et al., 2009; Saenger et al., 2009; Goertz et al., 2012).

4. Time-reversed images of the noise records band-passed in the HMT frequency band focus at the reservoir position for some datasets, confirming that the reservoir emits seismic waves (Steiner et al., 2008a; Witten & Artman, 2011; Goertz et al., 2012).

5. As human activities influence the noise spectrum above 1 Hz, the quietest time-periods should be selected prior to further processing in order to retrieve HMT (Lambert et al., 2009; Witten & Artman, 2011; Goertz et al., 2012; Riahi et al., 2013b). Saenger et al. (2009) agree with this assertion in general, but observe the same anomalies during night- and day-time in their data.

6. As structural features, such as the layering, influence the H/V ratio, their effect should ideally be calculated and removed from the observed spectral ratio prior to the V/H ratio interpretation in terms of HMT attributes (Lambert et al., 2006; Goertz et al., 2012).

On the sceptical side, the following arguments are opposed to each of the previous points:

1. All of the PSD_z *spectral peaks* are either due to artificial noise or to site effects (Hanssen & Bussat, 2008; Ali et al., 2010; Martini et al., 2013; Woith et al., 2014).

2. Same as point 1: the *lateral variations* of the structure or the proximity of human activity zones explain the *spatial patterns* of the attributes (Hanssen & Bussat, 2008; Martini et al., 2013).

3. V/H ratio peaks are due to the well-known surface layer resonance. There are no body waves detected by array processing (Ali et al., 2010; Martini et al., 2013). This is also the sense of fig. 5 in Hanssen & Bussat (2008).

4. (To my knowledge, no reported arguments).

5. Same as point 1; artificial noise is likely to dominate even at night-time (Ali et al., 2010).

6. *The whole* V/H ratio curve can be explained as in point 3 (Ali et al., 2010; Martini et al., 2013).

The absence of a clear mechanism explaining HMT, discussed in the next Section (1.2.3), yields another argument in favour of the sceptical side. Regarding the point 3, there is one point on which both sides agree: HMT, if they exist, should consist of body waves originating from the reservoir. Array-processing of the wavefield measured above the reservoir is usually performed by the «sceptical» authors to demonstrate it mainly consists of surface waves, and not body waves, while the «optimistic» authors usually

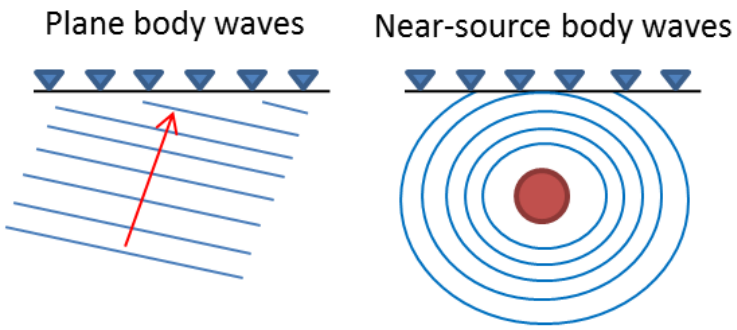


Figure 1.8: Difference between plane and circular wavefronts.

use polarisation or V/H ratio as indicators of body wave presence. This point is crucial, as the reported HMT are quite strong (up to several times the unperturbed PSD) and should in principle be measurable as coherent waves. However, based on my own experience, as well as on the absence of positive examples in the literature, I must admit that the supposed HMT are indeed not detected by array methods. Instead, waves with smaller apparent velocities dominate in the HMT frequency band. Two reasons, other than «HMT do not exist», can be invoked to explain this observation:

- Standard array techniques, such as FK (Capon et al., 1967), are well-suited for plane incident waves. However, the reservoir is typically located beneath the array and close enough to the surface. If it emitted body waves, it would generate a circular wavefront at the array scale (Fig. 1.8). For such cases, specific array techniques exist, such as the Matched Field Processing (MFP, e.g. Cros et al., 2011), or correlation-based techniques (e.g. Almendros et al., 1999).

- Otherwise, the HMT could also not be standard independent body waves. If they are single-scattered waves of some incident wavefield, and the reservoir lies within the first Fresnel zone of this wavefield with respect to the receiver array, then the HMT do not appear as a separate arrival, but simply modify the incident wavefield amplitude and phase. This could explain why they are not detected by array processing. My work is in fact based on this vision.

Regarding the point 6, it should be stressed that modelling the H/V ratio, even in the absence of tremor sources directly beneath the array, is not an easy task and relies on a set of debatable hypotheses the analyst must make on the incident wavefield composition. As shown by Bonnefoy-Claudet et al. (2006a) and Lambert (2010) (his Chapter 2), the H/V ratio in a layered structure can be interpreted both in terms of Rayleigh wave ellipticity and standing SH-wave resonance. The latter approach was introduced by Nakamura (1989) and is referred to as *Nakamura's method*. Interestingly, both approaches yield the same peak frequencies:

$$f_0 = \frac{V_s}{4H} \ , \tag{1.3}$$

with V_s the shear wave velocity and H the thickness of the shallowest layer. In theory, each strong contrast due to a soft layer overlaying a much stiffer one should

yield a specific peak frequency which can be estimated using the quarter-wavelength rule (equation 1.3) between the contrast depth and the free surface (Lambert, 2010, his Chapter 2). However, the amplitude of the peaks and the aspect of the H/V spectral curve is not the same for Rayleigh- and SH-wave interpretations. Yet another body-wave interpretation can be used if P-waves are taken into account, as done by Martini et al. (2013) below 0.5 Hz. Finally, the contribution of Love waves to the H/V ratio is quite significant (Endrun, 2011), but not taken into account in any of the previous models. Fortunately, all the interpretations predict the same value of the dominant peak frequency (equation 1.3), which can thus be regarded as a robust feature of the H/V spectral ratio. However, modelling and inverting the whole spectral H/V curve appears delicate, as it requires a heavy set of hypotheses which are difficult to verify. A way of dealing with this issue is proposed by Endrun (2011), via the estimation of the frequency-dependant proportions of the different wave modes present in the wavefield, and their contributions to the V/H ratio. Being aware of these limitations, the use of the modelled H/V ratio can appear ambiguous on both sides of the HMT controversy. Let us first take the example of Martini et al. (2013), which is a «sceptical» paper. The authors perform array analysis using two relatively small arrays with a maximum aperture of 400 m to retrieve the Rayleigh wave fundamental mode dispersion curve between 0.5 and 2 Hz. Though they mention that frequencies below 0.5 Hz are below the resolution limit, they still interpret the high apparent velocities observed in this range as body waves. The latter is unusual, as the frequencies close to the secondary microseism peak are usually dominated by surface waves (Bonnefoy-Claudet et al., 2006b; Nishida, 2017). They separately invert:

- the Rayleigh wave dispersion curve between 0.5 and 2 Hz, using Geopsy software (Wathelet, 2008);
- the H/V ratio between 0.5 and 10 Hz, interpreted as the Rayleigh wave fundamental mode ellipticity, again using Geopsy;
- the H/V ratio below 0.5 Hz, interpreted as the ratio between the amplification coefficients of vertically incident P- and SH-waves, using the routines of Herak (2008).

The inverted ground models appear mutually consistent, and also consistent with the available well data, which gives authors confidence about the fact the V/H ratio does not depend on the oil reservoir. However, the hypothesis of vertically incident body waves below 0.5 Hz is rather debatable. Moreover the contribution of Love waves to the H/V ratio above 0.5 Hz is totally neglected. On the other side, Goertz et al. (2012), in their «optimistic» paper, follow the vision proposed by Lambert et al. (2006). Acknowledging the fact that layered structures lead to H/V ratio peaks and troughs, the idea is to model the theoretical H/V due to the layered structure and to subtract the resulting V/H ratio to the total measured V/H. The difference should be due to HMT, i.e. body waves propagating from the reservoir towards the surface. They assume that the day/night variation of the ambient noise affects the wavefield composition, with body waves being detectable during the night for some frequency bands where human sources (i.e. surface waves) dominate during the day.

They consider that day-time noise is fully dominated by Rayleigh waves, so they invert the day-time H/V as a measure of Rayleigh wave ellipticity for a 1D V_s profile, individually for each station, similarly to Martini et al. (2013). Subsequently, they show that the night-time measured H/V ratio deviates from the curve predicted by the V_s profile obtained from the day-time data. The authors conclude that this is a manifestation of body waves appearing in the wavefield during the quietest periods. This conclusion is somewhat debatable. The argument could be returned the other way: if the night-time data was used to constrain the inversion, then the day-time V/H ratio would present a deviation from the predicted curve, and this could as well be interpreted as a manifestation of body waves. Moreover, the match between the observed and the fitted H/V curves during the day-time is not so good (fig. 9 in Goertz et al., 2012). Theses cases illustrate the ambiguity of H/V modelling with too simple hypotheses.

Regarding the points 1, 2 and 5, i.e. the role of the artificial noise, the sceptical position is valid if there is some human activity which never vanishes *within the receiver array*. If the human sources dominate the wavefield, but are located far from the array, they cannot explain the spatial variations of the noise amplitude within the array.

Finally, the point 4 appears as a strong argument in favour of the existence of the HMT. This is especially true as the noise band-passed in other frequency bands than the HMT-band did not focus at the reservoir location for the Voitsdorf case (Steiner et al., 2008b). However, some doubt can persist concerning the stability of the time-reversed image focus in the HMT band, as it remains unclear why the authors use so small time intervals for retro-propagation, while HMT are supposed to be a stationary process.

Only some of the discussed aspects are treated in this manuscript. Points 1 and 2 are of course adressed as they are the first object of investigation when treating LFPS data. PSD_z amplification is confirmed in the UGS context (Sections 2.1 and 2.2). The V/H ratio is also investigated (Point 3), but no clear correlation between the standard V/H attribute and the gas presence is observed (UGS context, Section 2.1). On the other hand, in the geothermal context, the V/H attribute is correlated with steam presence, but there is no PSD_z amplification. As UGS was the primary application domain of my work, I focused on PSD_z amplification and did not attempt to model the V/H ratio. TRI (Point 4) is totally beyond the scope of my work. Point 5 is addressed in Section 2.2, where we can observe that the PSD_z amplification, contrary to the absolute noise level, does not present any systematic difference between day- and night-periods. However, we can also observe an important day/night variation in the ambient wavefield composition in Chapter 3 (another dataset). Subsequently, my modelling work explicitely assumes the night-time wavefield composition (Chapter 4). Point 6 is not addressed in relation with the V/H ratio, but with the PSD_z amplification, as structural effects are investigated in Section 2.1 and in Chapter 4 both experimentally and numerically.

Now that we have discussed the phenomenology of the HMT, which is about finding evidence of seismic emission at the reservoir location, remains the question of the

mechanism behind such an emission. The latter is addressed in the next section (1.2.3).

1.2.3 A lacking mechanism

Dangel et al. (2003) constantly compare the newly detected HMT to volcanic tremors, because of a very similar spectral content. They investigate whether the two types of tremors could have a similar origin. Though the volcanic tremors are commonly considered as a manifestation of migrating magmatic fluids, there precise origin is under debate. While some authors suggest their spectral signature is due to a source effect (i.e. a specific mechanism of seismic emission), others interpret them in terms of a path effect (layered structure, vent geometry, etc.). For a discussion on this point, see the references in the introduction of Urquizú & Correig (1999). Regarding HMT generation, Dangel et al. (2003) eliminate the hypothesis of a path effect because of the «universality» of the detected hydrocarbon signatures, despite the different reservoir depths and the different geological environments at the surveyed sites. Among the possible source effects, Dangel et al. (2003) distinguish fluid flow (e.g. Julian, 1994) and bubble interface effects (e.g. Ripepe & Gordeev, 1999). They observe that volcanic tremors are transient, while HMT appear as stationary. Also, after applying a panel of signal processing tools, they conclude that HMT in the time domain are similar to band-passed signals. These two observations conduct them to prefer a model involving a stationary excitation of the fluid mixture in the reservoir by the «seismo-acoustic background», and a relaxation process involving phenomena at the scale of the fluid bubble. This relaxation would act as band-pass filter. Finally, they cite the model developed by Ripepe & Gordeev (1999) as the closest equivalent in volcanology, where the seismic part of the tremor is interpreted in terms of gas bubble coalescence. The advantage of this vision is that it allows to interpret long-lasting seismic emissions above hydrocarbon reservoirs observed after some active stimulation experiments (see Fig. 1.3a) and after earthquakes (see fig. 18 in Dangel et al. (2003) and Nguyen et al. (2008)), thus unifying both active and passive observations.

Bubble oscillation

The first modelling works conducted afterwards followed this hypothesis. Holzner et al. (2009) and Frehner et al. (2009) proposed simplified 1D oscillators as the equivalent models describing hydrocarbon bubble oscillation at the pore-scale, driven by the surface tension at the hydrocarbon/water contact. When the bubble moves inside the pore, the contact surface at both ends of the bubble becomes asymmetric, which generates a restoring force bringing the bubble back to its equilibrium position. The most simple oscillator is linear in a bi-conical pore (Fig. 1.9), while non-linearity appears when a spherical pore shape is used. The resonant frequency is of the same order of magnitude for both geometries, if the characteristic size of the pore remains the same:

$$\nu_0 = \frac{1}{2\pi h}\sqrt{\frac{6\gamma}{r\rho_L}}\ , \tag{1.4}$$

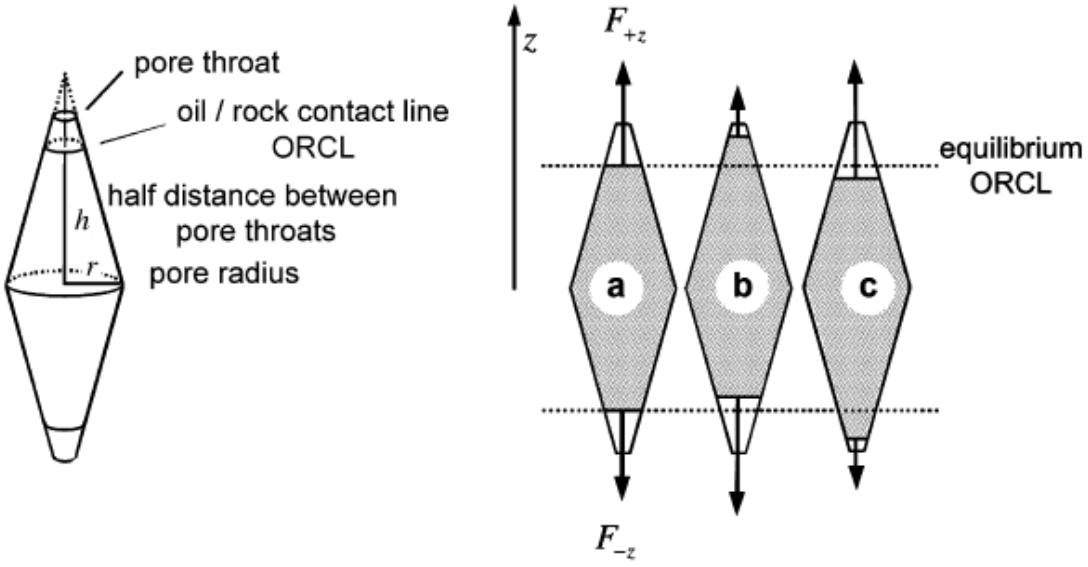


Figure 1.9: Schematic representation of a simple bi-conical pore geometry which enables low frequency oscillations of the bubble along z. The oil/rock contact line (ORCL) where capillary forces occur is formed between the fluid phases and the rock material. Different oscillation stages are: (a) liquid in equilibrium: the opposite capillary forces F_{+z} and F_{-z} balance each other; (b) Situation after a small displacement of the liquid toward positive z: F_{+z} decreases and F_{-z} increases. The resulting restoring force drives the bubble back towards its equilibrium position; (c) same as (b) with an initial bubble displacement towards negative z. Adapted after Holzner et al. (2009).

where ρ_L is the hydrocarbon density, γ the hydrocarbon/water surface tension coefficient, and r and h describe the bi-conical pore geometry shown in Fig. 1.9. The authors made numerical evaluations of ν_0 for an oil bubble with different viscosity values and pore radii of 1 mm and 3 mm. The resulting frequency values ν_0 ranged from 0.53 Hz to 5.5 Hz, so they concluded that the model was able to yield an interpretation of the amplification phenomenon over hydrocarbon reservoirs. However, equation (1.4) shows a strong dependance on the characteristic size l of the pore : $\nu_0 \propto l^{-3/2}$, which contradicts the «universality» requirement. Moreover, a pore radius of 1 mm for a sandstone reservoir is unusually large. A typical value of pore radius measured on the core samples from the sandstone reservoir in Chémery is about 10 μm. Using the latter value, one finds a resonance frequency of the order of a kHz, very far from the field observations. Further examining the theory's universality for hydrocarbon reservoirs, considering gas instead of oil decreases ρ_L and increases γ (water/gas interface), which results in a further increase of ν_0. Thus the model appears too dependent on the pore geometry and the fluid properties to explain the relative similarity of observations on reservoirs of different types pointed out by Dangel et al. (2003). These are essentially the conclusions of an extensive quantitative study by Broadhead (2010) testing the models of Holzner et al. (2009) and Frehner et al. (2009). Moreover, Broadhead (2010) was only able to reproduce an amplification effect when the damping term in the oscillatory model is unrealistically low. Goertz et al. (2012) argue that one should consider the typical *patch* size, assuming *patchy saturation* (e.g. Mavko & Mukerji, 1998) in the reservoir rock, instead of the pore

size, which decreases the resonant frequency. One could also argue that a continuous pore size distribution could include large pores generating strong effects. However, these arguments remain qualitative, and they do not solve the problem of the model sensitivity with respect to the fluid nature. If one still agrees to assume an amplification mechanism such as bubble oscillation, it is possible, via numerical simulation, to reproduce the spectral perturbations at the surface by placing random sources at the reservoir depth. When such sources act in addition to some random background noise, they guarantee an overall energy increase above the reservoir (contrary to the scattering mechanisms, which can result in both constructive and destructive interference, while the total energy is conserved). The increase of both PSD_z and V/H ratio is well reproduced by Lambert (2010) in his section 3.4, where he simulates an active emission zone inside the reservoir. However, this result does not justify the hypothesis of a seismic emission inside the reservoir.

While the bubble resonance theory does not, in its current state, appear satisfactory as the generation mechanism of the HMT, the next paragraph treats another theory, which does not rely on any kind of in-situ amplification.

Scattering due to attenuation contrast

As demonstrated experimentally by Korneev et al. (2004), the fluid saturation of a sandstone can significantly influence its reflection coefficient, and this effect is frequency-dependent. While Korneev et al. (2004) explain this feature in terms of a generic viscous-diffusive equation with ad-hoc coefficients (equation 6 in their paper), a considerable effort was made by other researchers to find a quantitative link between the rock-matrix and fluid saturation properties and the resulting frequency-dependent behaviour of the rock's quality factor. I provide more detail on this topic in Section 1.6. The general idea is as following. In standard visco-elasticity (Aki & Richards, 2002, section 5.5), each relaxation mechanism has a characteristic frequency f_0, where the quality factor is minimum (maximum attenuation). If the relaxation mechanism is due to partial saturation in a porous rock (e.g. mixture of water and hydrocarbons), the reflection coefficient of the reservoir is maximum at f_0. Then if f_0 is consistent with the typical frequencies of HMT, the latter could be interpreted in terms of visco-elastic scattering of the ambient waves on the reservoir, due to an *abnormally high* reflection coefficient. The particular mechanisms governing the wave propagation in a saturated porous medium are described by the Biot's theory of poro-elasticity (Biot, 1956, 1962). The latter involves the so-called *Wave-Induced Fluid Flow*, or WIFF, which dissipates energy at seismic frequencies (see Müller et al. (2010) for a review on WIFF). The characteristic frequency of this mechanism can be as low as several Hz, but depends on the characteristics and the heterogeneity of the solid matrix, the parameters of the fluids involved, as well as on the size of the typical fluid patches (e.g. Quintal et al., 2009). Quality factors for P- and S-waves behave differently with respect to these parameters (Quintal et al., 2012). After a series of numerical and experimental works investigating these effects, it was found that the characteristic relaxation frequency f_0 of WIFF could indeed be of several Hz for typical sandstone reservoirs (Quintal et al., 2011a,b, 2012). Lambert et al. (2013) integrate the predicted effective properties of a typical partially-saturated reservoir

in a wave-propagation model at macro-scale. Using random sources uniformly distributed in depth (both on the sides and beneath the reservoir), they analyse the spectra recorded at the surface, obtained with 10 different scenarii of reservoir properties. In a laterally invariant model with a soft surface layer, they predict an increase of the V/H ratio from 1 far from the reservoir up to 1.4 above the reservoir, for a 77% oil saturation scenario. Adding random sources at the surface (surface noise) progressively overwhelms the reservoir effect, which becomes barely detectable as the SNR ratio between deep and surface sources drops beneath 1. The authors conclude that body waves should dominate the wavefield in the HMT frequency band, in order be able to explain the V/H ratio anomalies above reservoirs by the WIFF-driven visco-elastic scattering. They refer to Zhang et al. (2009) as an example of body waves dominating the ambient wavefield between 0.6 and 2 Hz. They also mention that their model predicts a slight decrease of the PSD_z above the reservoir, which is in contradiction with the main reported HMT effect, i.e. PSD_z amplification. Since that time, Tisato et al. (2015) discovered an additional mechanism of P-wave attenuation acting at small values gas saturation: the *Wave-Induced Gas Exsolution-Dissolution* (WIGED), with characteristic frequencies of the order of several Hz. To my knowledge, no new upscaled simulations (similar to Lambert et al. (2013)) of ambient noise amplitude distortions accounting for this mechanism have yet been performed.

Elastic scattering

Elastic scattering can be seen as a particular case of visco-elastic scattering when there is no intrinsic attenuation contrast in the inclusion. This case corresponds to the simulation no. 10 in Lambert et al. (2013). They use the low-frequency limit of the P- and S- velocities in the inclusion, and the high-frequency limit in the background model, which generates an elastic contrast. This is not exactly similar to using the Gassmann equation (see Section 1.6), since the latter assumes the shear modulus independent from the fluid saturation, while the shear modulus in the model of Lambert et al. (2013) exhibits fluid-induced dispersion. Nevertheless, even with a V_s contrast inside the inclusion, pure elastic scattering is unable to generate a significant V/H ratio perturbation at the surface in their simulation. However, understanding the behaviour of the simple elastic scattering of the ambient noise by an inclusion is a first step before adding more complex mechanisms, such as visco-elasticity or multiple/resonant scattering. I have chosen elastic scattering as the starting point of the present work, and it revealed itself sufficiently complicated to become its central subject, though it does not allow, alone, to explain the observed HMT. Elastic scattering appears as a central phenomenon in the *Miscroseismic Sounding Method* (MSM, Gorbatikov et al., 2008), which is very similar to LFPS surveys in terms of input data. Ambient noise spectra are recorded with broadband seismometers during several hours, and the frequency-dependent amplitude lateral perturbations on the vertical component are assumed to be representative of the subsurface lateral heterogeneities. The method was originally introduced by Gorbatikov et al. (2004), who observed a correlation between the ambient noise amplitude perturbations and the position of a magmatic intrusion (imaged by other geophysical methods) on the Lanzarote Island. The difference with HMT is that the amplitudes are analysed below 1 Hz. Rayleigh wave fundamental mode domination in the ambient wavefield is

assumed, and the amplitude perturbations are interpreted in terms of Rayleigh wave scattering. The MSM does not consider scattered waves as independent body-wave arrivals, but rather as perturbations of the incident Rayleigh waves. The depth information is supposed to be contained in the frequency of the perturbations. An empirical coefficient $K \approx 0.4 - 0.5$ is introduced by Gorbatikov et al. (2008) to tie each frequency to the corresponding depths:

$$H(f) = K\lambda(f) \ , \tag{1.5}$$

where $H(f)$ is the depth associated to the frequency f and $\lambda(f)$ the corresponding wavelength of the Rayleigh fundamental mode. This allows to construct «intensity sections» in depth, simply using equation (1.5) to locally transform the measured PSD_z from the $(x, y) - f$ domain to the $(x, y) - z$ representation. Applications to deep structures such as salt domes, fault structures and subsident basements are presented. On empirical basis, in the case of the Smolensk fault where the method is applied, Gorbatikov et al. (2008) claim a lateral resolution of 4% of the wavelength, because, according to them, the resolution of their method is not limited by the size of the Fresnel zone. Numerical modelling, however, leads to a lateral resolution of 20% of the wavelength. Gorbatikov & Tsukanov (2011) perform a thorough numerical study of Rayleigh wave fundamental mode scattering by a subsurface inclusion, as the mechanism at the basis of the MSM. They conclude from the simulations that slow inclusions generate positive amplitude anomalies as main effect, which is in contradiction with the results from the Lanzarote Island, where the intrusion (positive velocity contrast) was associated with positive amplitude anomalies. Gorbatikov et al. (2013) apply the MSM to image the deep structure of the El Hierro Island down to 25 km, separating an old intrusive body and a modern magmatic reservoir. Again, according to the authors, the separation of the two bodies implies a better lateral resolution than the standard $\lambda/4$ criterion (super-resolution). From equation (1.5), it can be seen that MSM «super-resolution» implies low-frequency amplitude variation over small distances in the field. The authors interpret this effect in terms of very low Poisson ratio in the inclusion, plausible for highly-fractured materials. Tsukanov & Gorbatikov (2015) numerically study and confirm the claimed «super-resolution» effect in case of an anomalously low Poisson ratio. Tsukanov & Gorbatnikov (2018) extend their numerical modelling to the H/V ratio and to random Rayleigh wavefields. They find that a low-velocity inclusion increases the V/H ratio, while a high-velocity inclusion decreases it. Yanovskaya (2017) suggests to interpret the amplitude anomalies used by the MSM in terms of layered medium 1D eigenfunctions, instead of scattering. However, while this could work to explain an amplitude anomaly above the middle of a sufficiently wide inclusion, the edge effects, crucial for assessing the lateral resolution, are scattering effects. This point is further worked out in this manuscript (Chapters 5 and 6).

Resonant scattering

Other interpretations were proposed for HMT, based on different possible kinds of trapped waves. The latter include:

- Standing waves between the reservoir and the free surface (Birialtsev et al., 2006).
- Slow Stoneley waves trapped into fluid-saturated fractures. Lambert et al. (2009) and Lambert et al. (2011) indeed mention the paper by Korneev (2008) dedicated to this subject.
- Scattering of guided waves, if the reservoir lies in a stratigraphic waveguide (layer of lower velocity where the ambient noise amplitudes are higher than in the neighbouring layers, Mardanov & Kipot, 2008).

While potentially interesting, I did not study these approaches in much detail. Quantitative studies of their application to HMT are lacking in the literature. In my modelling work, I considered single elastic scattering as the first effect to be understood.

Which source term ?

Interestingly, relatively few attention is paid to the precise wavefield characterisation (i.e. which wave types dominate) in most of the LFPS surveys (e.g. Saenger et al., 2009; Lambert et al., 2009). However, most of the proposed HMT models imply the ambient noise as a forcing term. In case of the bubble oscillation theory, as energy transfer between different frequencies is allowed via the model's non-linearity, one can speculate about the precise trigger, e.g. the oceanic microseism (Rayleigh waves, Saenger et al., 2009), or even tidal stress (Gerivani et al., 2012). To verify this, one should compare the time evolution of the energy within the HMT frequency band and within the band of the conjectured trigger. Such a comparison failed for Ali et al. (2010), which is another argument against the bubble oscillation theory. For elastic scattering models, no non-linear energy transfer is possible between the different frequencies, so the ambient wavefield at the frequency of the detected anomalies is the only possible source term. It must definitely be characterised, at least to know whether it mainly consists of body waves or of surface waves; of the fundamental or of higher modes, in order to choose a proper modelling framework. The modelling works by Lambert (2010) and Lambert et al. (2013), for example, imply the necessary presence of body wave illumination. On the other hand, in Gorbatikov et al. (2008) and other publications in relation with the MSM, Rayleigh wave fundamental mode domination is assumed without any array processing to check this hypothesis. My opinion is that wavefield characterisation should be performed along with HMT attribute extraction, in order to be able to better understand which propagation modes one can invoke in order to explain the anomalies. This is the starting point of my modelling work in Chapter 3. It is important to stress that spectral modifications *are* possible due to purely elastic scattering, via the interference patterns it generates, and also via the frequency-dependent penetration depth of the surface waves.

Conclusion

As a conclusion, we can distinguish two wide families of models for the interpretation of HMT. One of them implies complicated and poorly understood phenomena, such as amplification through bubble oscillation, which for now appear contradictory.

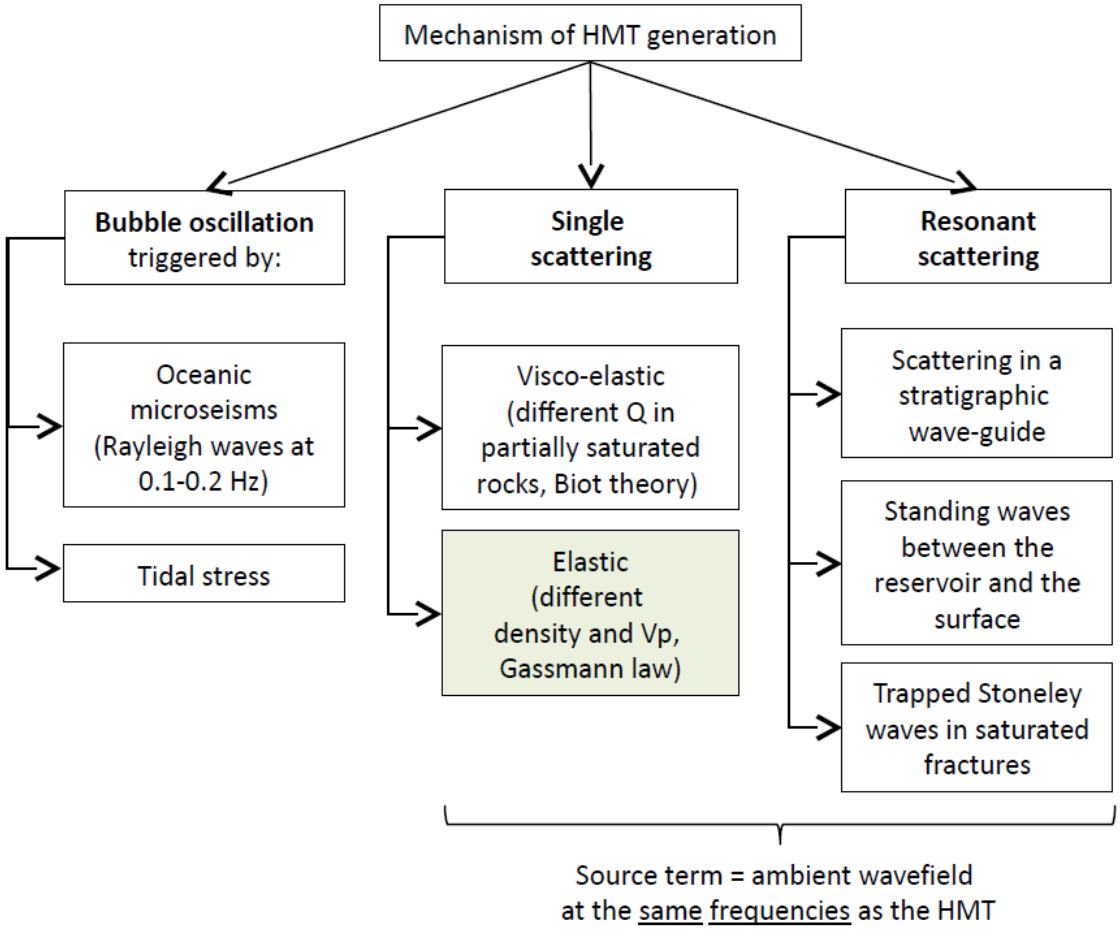


Figure 1.10: Summary of the suggested generation mechanisms for hydrocarbon micro-tremors. Elastic scattering is highlighted in green as it is the mechanism I chose to investigate in this manuscript.

The second one relies on more standard and well assessed scattering phenomena, either elastic or viscoelastic, and is of more general use. It also applies to MSM, proposed for structural imaging with ambient noise amplitude (Gorbatikov et al., 2008). A third family, which can be categorised as «resonant scattering» (Korneev, 2008, 2009; Mardanov & Kipot, 2008), is not studied in detail in this review, as I found no published quantitative considerations on its application to the HMT. The summary of the mentionned mechanisms is given in Fig. 1.10. It is difficult to reproduce the full HMT effect with scattering models (either elastic or visco-elastic), at least in terms of PSD_z increase (Lambert et al., 2013). It is easier to achieve an increase of PSD_z by explicitly embedding a seismic emission zone at the reservoir position (Lambert, 2010, chapter 3), as suggested by the bubble oscillation theory, but the latter remains unproved (Broadhead, 2010). On the other hand, a proper scattering modelling requires ambient wavefield characterisation in terms of wave types.

In the present work, I start with wavefield characterisation performed on our largest dataset (Chémery), after which I investigate the simplest possible scattering model, i.e. elastic single scattering (highlighted in green in Fig. 1.10). I consider this ap-

proach as a starting point before incorporating more complex phenomena within the reservoir. While purely elastic contrasts inside a hydrocarbon reservoir itself are unlikely to generate strong amplitude perturbations at the surface, the purely elastic approach can be valuable to assess the influence of lateral structural heterogeneity. Hydrocarbon reservoirs, in fact, are often characterised by the presence of a structural trap, such as a bounding fault or an anticline (e.g. Dangel et al., 2003; Goertz et al., 2012). It is yet unclear to which extent such structural features could not be the origin of the detected HMT, via an elastic scattering process. This option is considered by Goertz et al. (2012) in their discussion.

1.3 Propagation modes of elastic waves: basic properties

In this section, I provide a very brief and simplified overview of the different wave types existing in elastic media. This is useful to introduce some of the notions that will be used in the remainder of the manuscript. I try to comment on the role of these particular notions in the modelling workflow developed during this PhD work.

1.3.1 Wave equation and Green's function

Elastic waves consist of coupled propagation of stress and displacement within an elastic support. Seismic waves can be well described in terms of elastic wave propagation. Mathematically, they arise as the solutions of the elastic wave equation, which can be written as a coupled first-order system for the stress tensor σ_{ij} and the velocity vector v_i:

$$\begin{cases} \rho \partial_t v_i = \partial_j \sigma_{ij} + f_i^v, \text{ sum over } j, \\ \partial_t \sigma_{ik} = c_{ikjl} \partial_l v_j + f_{ik}^\sigma, \text{ sum over } (j,l), \end{cases} \tag{1.6}$$

where c_{ikjl} is the stiffness tensor, while f_i^v and f_{ik}^σ represent the source terms (respectively a force and a moment tensor rate (Virieux et al., 2011)). Differentiating the first equation with respect to time and substituting the second equation allows to eliminate the stress and leads to a second-order equation for the velocity. The latter can be written in the frequency domain by substituting the operator ∂_t by the complex multiplier $i\omega$. Finally, replacing v_i by $i\omega u_i$, leads to the displacement formulation that will be used in the remainder of this manuscript:

$$\left(-\rho\omega^2 \delta_{ij} - \partial_k c_{ikjl} \partial_l\right) u_j = f_i \text{ (sum over } j,k,l), \tag{1.7}$$

with $f_i(\omega) = f_i^v(\omega) + \partial_j f_{ij}^\sigma(\omega)/(i\omega)$. In the isotropic case, the stiffness tensor takes the simple form $c_{ikjl} = \lambda \delta_{ik}\delta_{jl} + \mu\left(\delta_{ij}\delta_{kl} + \delta_{il}\delta_{kj}\right)$, where λ and μ are the Lamé parameters of the elastic medium. The second Lamé parameter is equal to the shear modulus of the medium. The first Lamé parameter λ depends on both bulk (K) and shear moduli:

$$\begin{cases} K = \lambda + 2/3\mu \text{ in 3D}, \\ K = \lambda + \mu \text{ in 2D}. \end{cases} \tag{1.8}$$

In the following, I work only with the Lamé parameters, except in Section 1.6 dealing with the effective properties of a partially saturated medium. Provided a set of boundary conditions, the solution of the wave equation is often formulated for an impulsive point force source placed at the position $\mathbf{x}_s$, written as $f_j \delta(\mathbf{x} - \mathbf{x}_s) \mathbb{1}_f$ on the component j, with $\mathbb{1}_f$ the constant spectrum of Dirac distribution. For a receiver at the position $\mathbf{x}_r$, this solution is referred to as the *Green's function* $\mathbf{G}(\mathbf{x}_r, \mathbf{x}_s, \omega)$ for the source-receiver couple (r,s). In the elastic case, this function is a tensor, as it links the displacement vector components u_i to the force (source term) components f_j: $u_i = G_{ij} f_j$. The Green's function is unique and contains all the existing wave types. It also contains the response at all the frequencies. In practice, the source term has a finite support in the frequency domain, described by the source spectrum $s(\omega)$ instead of the Dirac distribution $\mathbb{1}_f$. Then the medium response is simply the convolution of the Green's function with the source wavelet, or a product in the frequency domain:

$$u_i(\omega) = s(\omega) G_{ij}(\omega) f_j \ . \tag{1.9}$$

In case of uncontrolled sources, different natural or anthropogenic processes have different frequency ranges, which I further discuss in Section 1.4.1 and in Chapter 3, in relation with the ambient noise composition. A visualisation of the Green's function (convolved with a source wavelet) in the space-time domain is provided by a seismic gather, which is a two-dimensional representation of the particle displacement, velocity, or acceleration. An example is provided in Fig. 1.11 for the vertical-component signal $u_z(x,t)$, obtained from the ambient noise analysis by Nakata et al. (2015). Surface and body waves can be distinguished due to different propagation velocities. Moreover, body waves appear as continuous lines, while surface waves form wide packets. The latter property is due the dispersive character of the surface waves, which is discussed a little further.

1.3.2 Body waves

Body waves arise as the only far-field solutions of the wave equation in an unbounded medium (no free surface). They are of two types: P-waves and S-waves, respectively corresponding to compressional and shear motions. Their velocities can be expressed in terms of the Lamé parameters as

$$\begin{cases} V_p = \sqrt{\dfrac{\lambda + 2\mu}{\rho}} \ , \\ V_s = \sqrt{\dfrac{\mu}{\rho}} \ . \end{cases} \tag{1.10}$$

V_p and V_s are also often written α and β. These velocities do not depend on frequency, except in the vicinity of a characteristic frequency of some dissipation mechanism, which is discussed in Section 1.6. For the scope of this manuscript, body waves can be considered as non-dispersive. In the time domain, body waves in a homogeneous isotropic unbounded three-dimensional space are a time-delayed version of the exci-

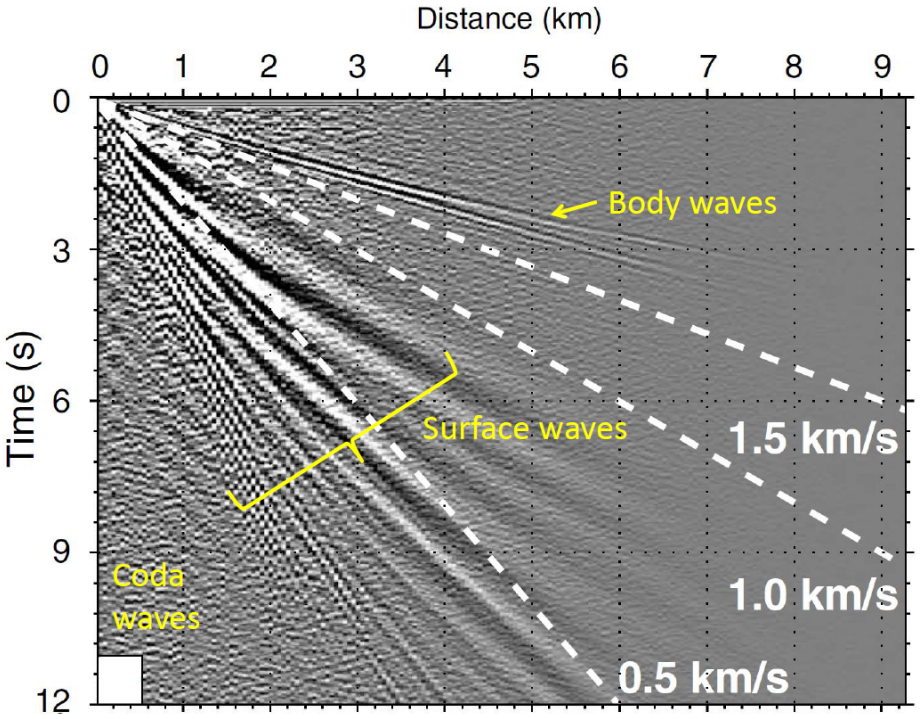


Figure 1.11: Example of vertical particle velocity (grey scale) represented as a function of time and distance from the source. Surface and body waves differ by their velocity and dispersion characteristics. White dashed lines correspond to constant velocity propagation in the $x-t$ space. From Nakata et al. (2015).

tation force (Aki & Richards, 2002):

$$\begin{cases} G_{np}^{(P,3D)} = \dfrac{1}{4\pi\rho\alpha^2}\gamma_n\gamma_p\dfrac{1}{r}\delta\left(t-\dfrac{r}{\alpha}\right) \\ G_{np}^{(S,3D)} = -\dfrac{1}{4\pi\rho\beta^2}\left(\gamma_n\gamma_p-\delta_{np}\right)\dfrac{1}{r}\delta\left(t-\dfrac{r}{\beta}\right) \end{cases}, \tag{1.11}$$

with r the distance from the source to the receiver and $\gamma_i = \cos(x_i/r)$ the direction cosines imposing the correct radiation patterns along with the longitudinal and shear motions respectively for P- and S-waves. Importantly, the amplitude decay of body waves is proportional to r^{-1} in 3D. The 2D case, *which I study in this work*, corresponds to assuming the medium invariance in one the the three directions (say y). The waves split in two uncoupled systems: the P-SV system, which corresponds to the motion within the $x-z$ plane (in-plane motion), and the SH system, which corresponds to the motion in the invariant direction y (antiplane motion). The antiplane motion only supports S-waves, while the in-plane motion supports both P- and S-waves. Because of the invariance in the y diraction, point forces become line forces. In 2D, the propagation of body waves is not sharp anymore, as the response at the receiver is a superposition of signals arriving from increasingly distant points of a *line source* (Strauss, 2008). The corresponding Green's functions can be found in Kausel (2006). While the propagation velocities remain the same, the amplitude decay is now proportional to $r^{-1/2}$, as the wave's energy (amplitude squared) spreads over a circle instead of a sphere. In Section 5, I show that the different near-field

behaviour of 2D (in-plane) and 3D unbounded medium Green's functions can have a considerable effect on the Rayleigh wave scattering radiation patterns, even when the analysis is restricted to the radial-vertical plane.

In bounded/heterogeneous media, body waves represented in terms of geometrical optics undergo both reflection and refraction. These phenomena govern the trajectory of the body wave rays.

Body waves from earthquakes

Body waves excited by earthquakes are categorised in different *phases*, according to their type (P or S), and to their path inside the Earth. They split in crustal phases, mantle phases and core phases (Fig. 1.12a-c), depending on how deep they penetrate in the Earth interior. Refraction plays an important role in the body wave propagation, as the rays can be trapped along the interfaces between layers with different seismic velocities. P_n and S_n, for example, are phases which propagate along the Moho (boundary between the crust and the upper-mantle).

Body waves in active experiments

Diving waves are the counterpart of refracted waves in media with a velocity gradient, where the rays continuously bend and re-emerge at the surface at some point. Fig. 1.12d illustrates a continuous mixture of diving and refracted waves in the shallow crust, which is typically encountered in active seismic experiments. Refracted waves are usually used for tomography, i.e. retrieving the velocity model based on the first arrival times. This typically yields a very smooth velocity model. On the contrary, reflected waves (Fig. 1.12e) allow to determine the two-way travel time between the surface and the reflector with a precision of a fraction of the wave's period. This technique is typically used in exploration seismic for obtaining high-resolution images of the geological layers.

It is fascinating to realise that the ambient noise can contain all of the described types of body waves, as discussed in Section 1.4.1. In this work, however, I focus on surface waves, which generally constitute the largest part of the ambient noise energy.

1.3.3 Surface waves

Surface waves arise when a free surface condition is added on some the medium boundaries. This is a zero-traction condition:

$$\sigma_{ik} n_k = 0 \ , \tag{1.12}$$

with n_k the normal to the surface. Surface waves can be described in terms of Rayleigh and Love waves, existing respectively in the P-SV and the SH systems. This work always remains in 2D, so Rayleigh and Love waves are uncoupled. As I only investigate vertical component motion in this work, Love waves have no influence (transverse motion). However, in 3D, Rayleigh and Love waves are coupled through scattering (e.g. Snieder, 1986). In the following, I enumerate some of the main characteristics of Rayleigh waves which are useful for reading the manuscript.

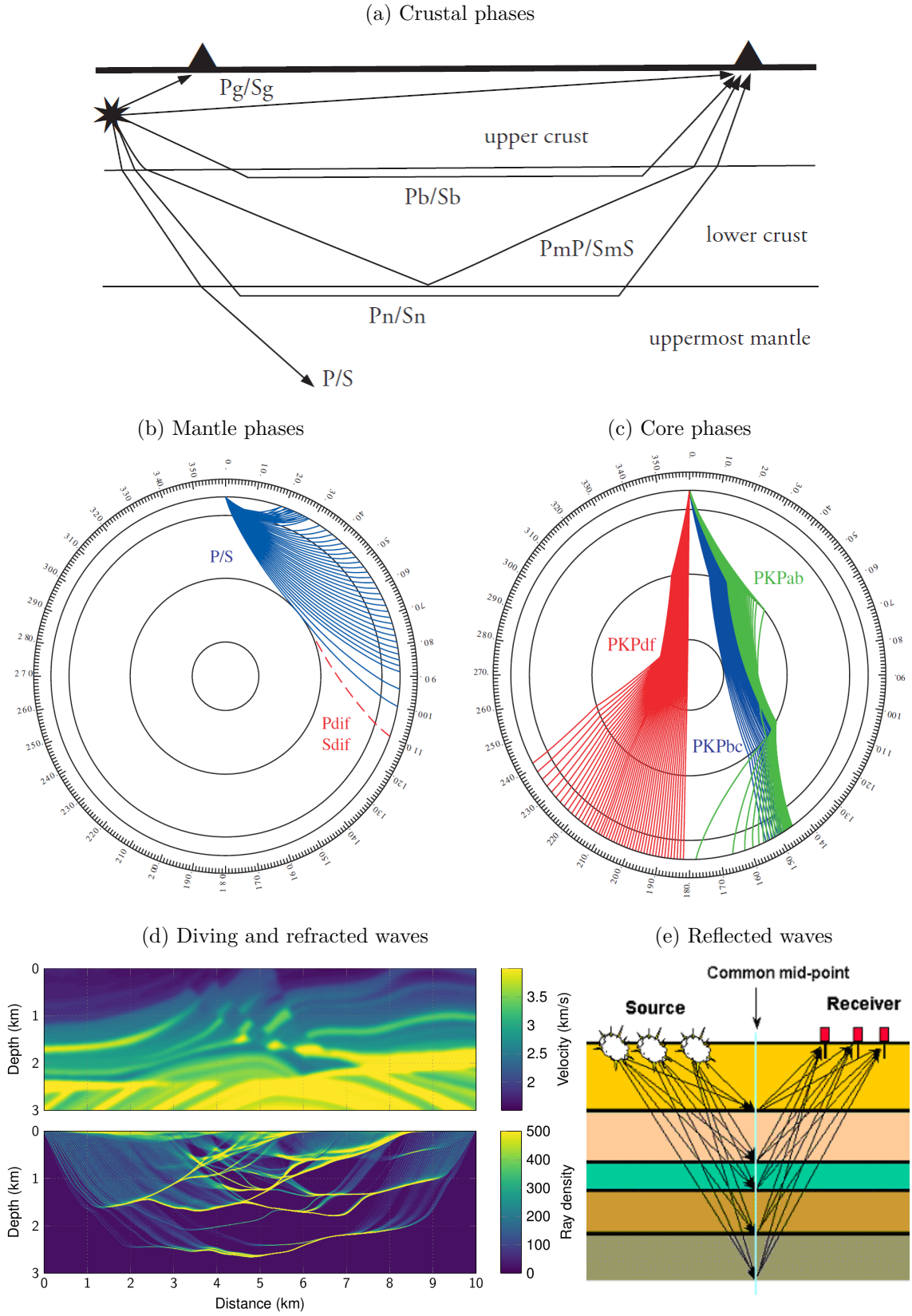


Figure 1.12: (a)-(c) Different types of phases according to the IASPEI classification (Storchak et al., 2003). (d) Upper panel: Marmousi velocity model. Lower panel: Associated diving and refracted waves (visualisation through ray density, courtesy of Dr. Keurfon Luu). (e) Schematic representation of reflected waves (source: Schlumberger website).

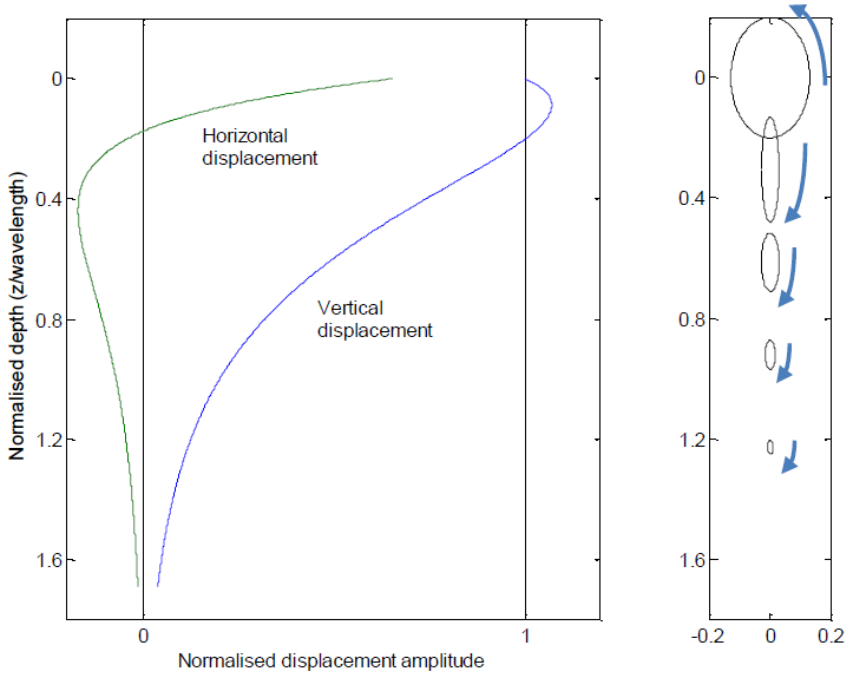


Figure 1.13: Left: Rayleigh wave amplitude on vertical and horizontal components as function of depth. Right: Associated particle motion. From Gedge & Hill (2012).

Particle motion and ellipticity

The particle motion of Rayleigh waves is elliptic, since vertical and horizontal components are shifted with respect to each other by $\pi/2$. At the surface, the motion of Rayleigh waves is retrograde, but the sense of the rotation switches to prograde at some depth (Fig. 1.13). The ratio between the amplitude of the horizontal and the vertical motions is called *ellipticity*. An ellipticity of 1 corresponds to a circular particle motion. In Section 5, it is shown that this change of propagation regime affects the radiation pattern of Rayleigh wave scattering. As the motion switches from prograde to retrograde, its horizontal component vanishes at some depth. As mentioned in Section 1.2, Saenger et al. (2009) suggested that this region of purely vertical motion could correspond to the reservoir depth for the low-frequency (0.1-0.2 Hz) microseisms. Then it could trigger vertically polarised microtremors. However, such a mechanism requires a non-linear energy transfer from low to high frequencies, which is a hypothesis that I discarded for the scope of the present work.

Fundamental and higher modes

In a homogeneous half-space, only the fundamental Rayleigh wave mode is present. The relative evolution of its amplitude with depth is shown in Fig. 1.13 for both vertical an horizontal components. The two curves in Fig. 1.13 are called Rayleigh *eigenfunctions*, and often written as U_z and U_x. Their derivation can be found in Aki & Richards (2002). The motion is retrograde/prograde according to the relative sign of the eigenfunctions. The fundamental mode velocity c_R in a homogeneous half-space

is defined as the root of the so-called Rayleigh characteristic function (Rayleigh, 1885)

$$\left(2-\frac{{c_R}^2}{\beta^2}\right)^2-4\left(1-\frac{{c_R}^2}{\alpha^2}\right)^{1/2}\left(1-\frac{{c_R}^2}{\beta^2}\right)=0\ . \tag{1.13}$$

Though the precise value of c_R also depends on α, it typically ranges from 0.87β to 0.95β. Rayleigh waves in homogeneous media are non-dispersive, as the above velocity does not depend on frequency. In Fig. 1.13, we see that the depth penetration of the fundamental mode is about one wavelength. This means that lower frequencies (larger wavelengths) penetrate deeper than higher frequencies.

In a layered half-space, Rayleigh waves can propagate according to several modes. An example of the associated eigenfunctions is shown in Fig. 1.14a. First, we observe that the higher modes (overtones) penetrate deeper than the fundamental mode at a given frequency, and thus are sensitive to deeper structures. This property is important in Chapter 3, where overtones are used to derive a V_s velocity model and to estimate the depth of the contact between the sedimentary cover and the bedrock. It is also important in Chapters 4 and 6, where it is shown that overtones are more efficiently scattered by a deep reservoir, compared to the fundamental mode.

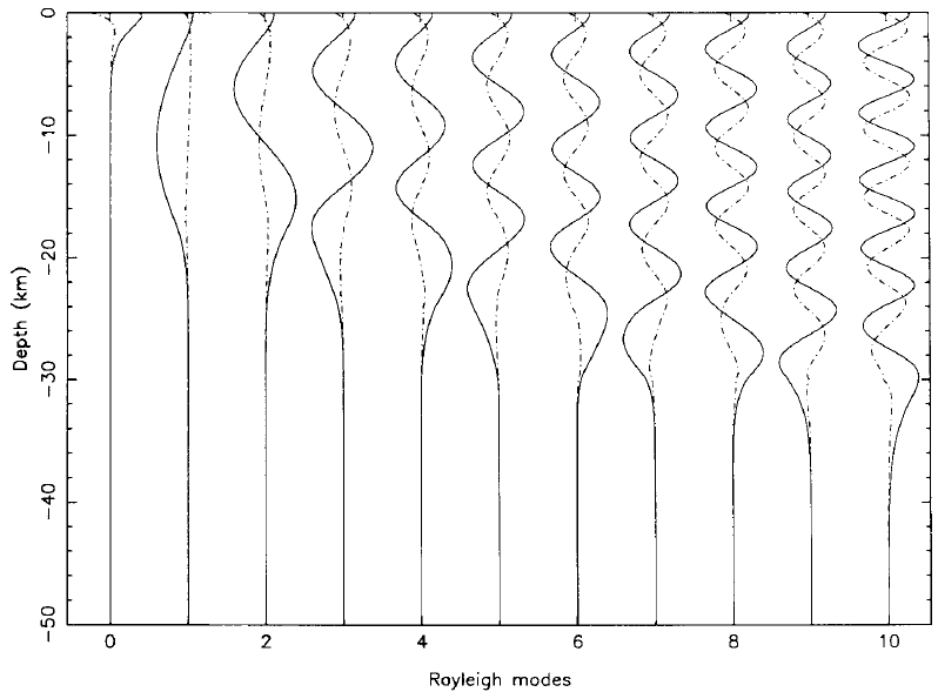
Dispersion

Another important property of Rayleigh waves which appears in layered media is dispersion. Intuitively, as the low frequencies penetrate deeper, they are sensitive to faster rocks, and propagate faster than the high frequencies, which are concentrated in the shallow layers. The evolution of the velocity with frequency is described by the dispersion curve. Each mode has its own dispersion curve, and higher modes propagate faster. The dispersion curves associated with the eigenfunctions in Fig. 1.14a are shown in Fig. 1.14b. The dispersion curves measured from field data can be inverted for the layered half-space properties, as illustrated in Fig. 1.15. This inversion is rather ill-posed, as many models can explain the observed data equally well. Nevertheless, this type of inversion is at the basis of most of the ambient noise applications, as will be shown in the next section. It is also used in this work in Chapter 3. Another issue with this technique is mode skipping, illustrated in Fig. 1.16, which generates an apparent dispersion curve containing several overtones. The latter are sometimes hard to correctly label and can severely bias the inversion. The issue of the correct mode labeling is encountered and discussed in Chapter 3. Finally, the dispersive nature of Rayleigh waves gives rise to two distinct velocities, which are the group and the phase velocity. The first measures the propagation velocity of the wave packet (envelope). An earthquake arrival times, for example, are predicted by the group velocity along the wave path from the source to the receiver. The phase velocity measures how fast the signal itself streams inside the envelope. An example of the dispersion curves for phase and group velocities is shown in Fig. 1.17. The group velocity can be derived from the phase velocity using the relation

$$v_g=v_\phi\left(1-\frac{\omega}{v_\phi}\frac{\partial v_\phi}{\partial\omega}\right)^{-1}\ . \tag{1.14}$$

The other way is more delicate, as an arbitrary constant appears due to the integration. Thus knowing the phase velocity presents some advantage.

(a) Rayleigh wave displacement eigenfunctions at $f = 1\ Hz$



(b) Rayleigh wave dispersion curves

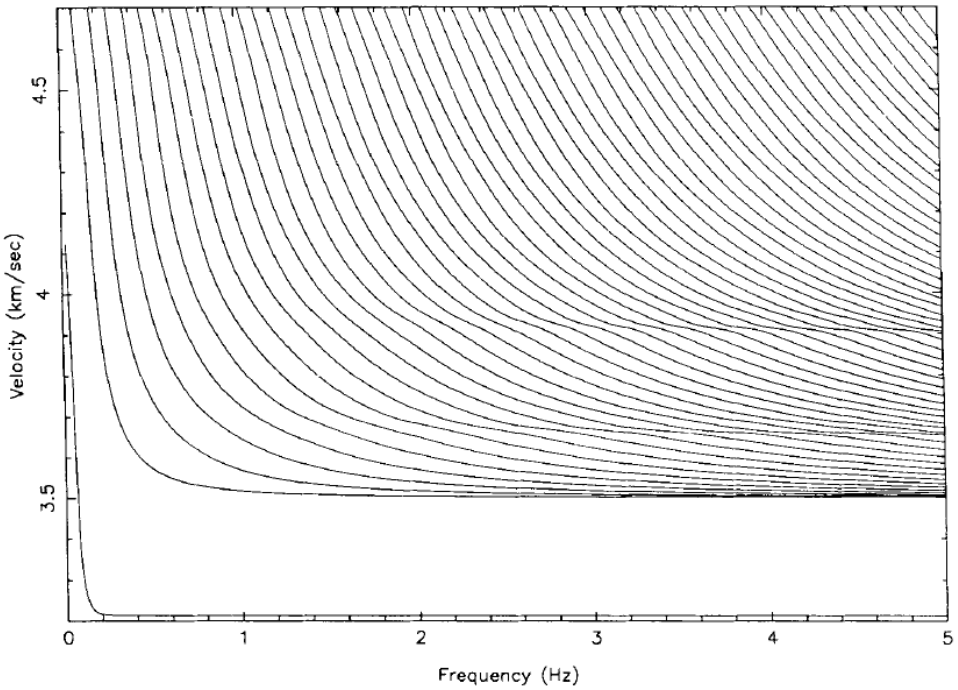


Figure 1.14: Example of (a) Rayleigh wave displacement modal eigenfunctions and (b) associated dispersion curves for a given ground model. From Chen (1993).

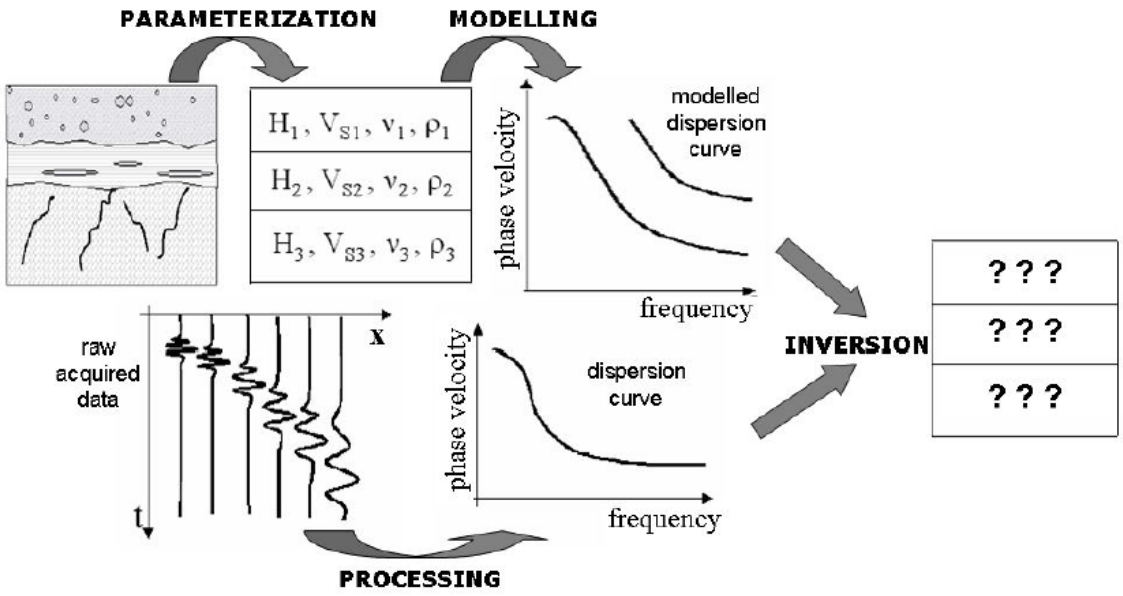


Figure 1.15: Dispersion curve inversion workflow. From Strobbia (2002), adapted by Solano et al. (2014).

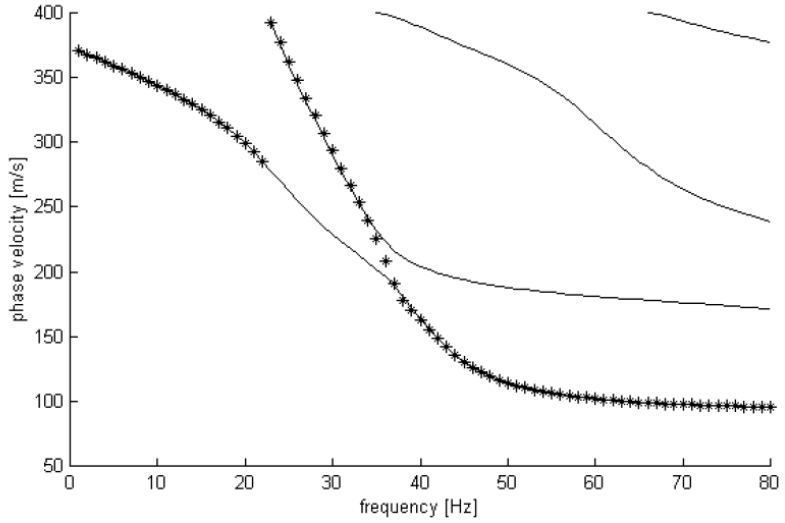


Figure 1.16: Apparent dispersion curve due to modal skipping (asterisks). Solid lines - theoretical dispersion curves. From Strobbia (2002).

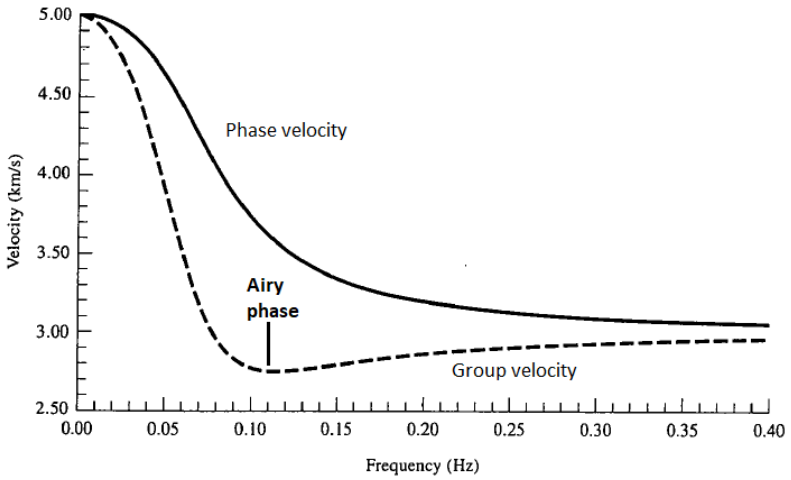


Figure 1.17: Example of phase and group velocity dispersion curves. From Aki & Richards (2002).

Airy phase

The Airy phase is a particular wave packet which develops around the frequency corresponding to a group velocity minimum, such as the one visible in Fig. 1.17. Details on why the group velocity minima concentrate the propagating energy can be found around equation (7.18) in Aki & Richards (2002). In this work the Airy phase phenomenon is important in relation with the L_g phase - a particular crustal phase which is detected as part of the ambient wavefield in Chapter 3. The L_g phase is a superposition of several Rayleigh wave overtones which have their group velocity minima at the same velocity, but at increasing frequencies, as illustrated in Fig. 1.18. The resulting superposition appears as a non-dispersive wave. Such a particular configuration of the modal dispersion curves is likely to arise due to a strong contrast between two layers, provided the upper layer is thick enough with respect to the shear wavelength. In the case of the L_g phase, this contrast is the Moho. In Chapters 4 and 6, I do not take the Moho into account, but a superposition of overtones arises in numerical simulations due to the contrast between the sediments and the underground bedrock. Details on the L_g phase propagation characteristics can be found, for example, in Kennett (1986) and Shapiro et al. (1996).

Propagation equations

The surface wave displacement on any component i can be written by separating vertical and horizontal variables (Aki & Richards, 2002)

$$\begin{cases} u_i(x,y,z,\omega) = f_i(z,\omega)g_i(x,y,\omega) \text{ in 3D}, \\ u_i(x,z,\omega) = f_i(z,\omega)g_i(x,\omega) \text{ in 2D}. \end{cases} \tag{1.15}$$

In 2D, requiring surface waves to be harmonic leads to $g(x) = \exp(ikx)$, with k to be determined. In the frequency domain, injecting these expressions into the equations of motion 1.6 without any source term (homogeneous equation) leads to a simple

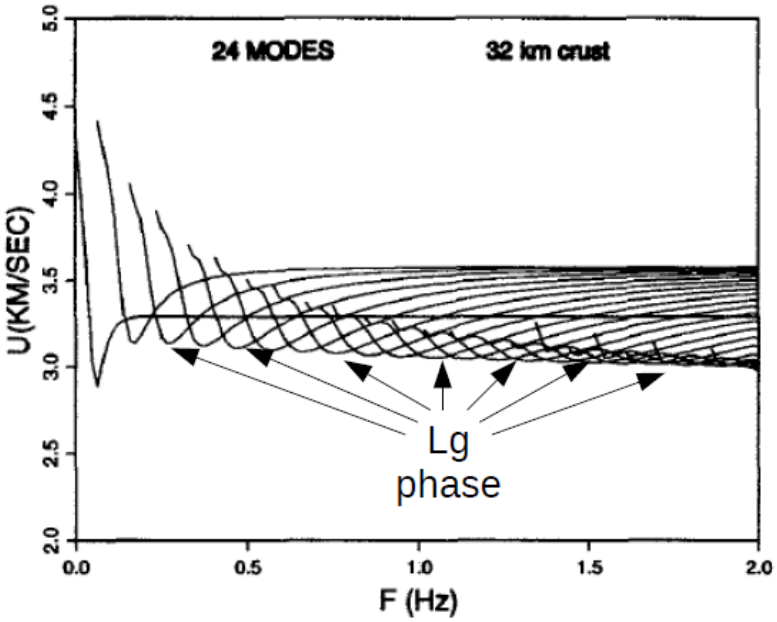


Figure 1.18: The L_g phase as a result of dense Airy phases due to several Rayleigh wave overtones. From Zhang & Lay (1995).

matrix equation of the form

$$\frac{d\mathbf{f}}{dz}(z) = \mathbf{A}(k, \omega, \rho, \lambda, \mu)\mathbf{f}(z) \ , \tag{1.16}$$

with $\mathbf{f}$ is a vector containing the individual vertical decay functions $f_i(z)$ for stresses and displacements (this vector has 4 elements for Rayleigh waves and 2 elements for Love wave). $\mathbf{A}$ is a square matrix of the size of $\mathbf{f}$. For a given angular frequency ω, this eigenvalue problem only admits solutions for some discrete values of $k_n(\omega)$, which depend on the medium properties ρ, λ, μ. Each $k_n(\omega)$ defines a mode of propagation, and yields its dispersion curve, e.g. those represented in Fig. 1.14b. The associated eigenvector $\mathbf{f}_n$ contains the vertical decay of the amplitude, i.e. the eigenfunctions $U_{x,n}$ and $U_{z,n}$ (e.g. those shown in Fig. 1.14a), specific to each mode. In a layered medium, the integration of the equation (1.16) can be performed numerically using the propagator matrix method, introduced by Thomson (1950) and Haskell (1953). The boundary conditions of the integration are provided by the radiation condition in depth (vanishing amplitudes) and the zero-traction condition at the free surface. This yields the direct problem for predicting dispersion curves and ellipticity $\chi = U_{x,n}/U_{z,n}$ for each mode.

Knowing the vertical dependence of the amplitude, the homogeneous solution for surface waves in 2D is

$$u_i^{(\text{hom, 2D})}(x, z, \omega) = \sum_n U_{i,n}(z, \omega) \exp\left[i\left(k_n(\omega)x + \phi_{i,n}\right)\right] \ , \tag{1.17}$$

with n the index of the mode and $U_{i,n}$ the displacement eigenfunction on the component i for the mode n. $\phi_{i,n}$ translates the phase shift between the two components (elliptic motion). Such a representation is referred to as modal summation, and is

usually used in 3D.

In 3D, the vertical dependence of the amplitude remains the same as in 2D, but $g(x, y)$ is determined as a solution of a 2D acoustic equation (Helmholtz equation, Wielandt, 1993). The latter is specific to each propagation mode defined by $k_n(\omega)$:

$$\left([k_n(\omega)]^2 + \nabla^2\right) g_n(x, y) = 0 \ , \tag{1.18}$$

where ∇^2 is the Laplace operator. The solutions of this equation belong to the family of Bessel functions. The final solution for the displacement u_i in a laterally homogeneous medium can thus be written as

$$u_i^{(\text{hom, 3D})}(r, z, \omega) = \sum_n U_{i,n}(z, \omega)(\omega) J\left[k_n(\omega) r + \phi_{i,n}\right] \ . \tag{1.19}$$

with J some combination of Bessel functions, with phase-shifts $\phi_{i,n}$ specific to the component. The precise expressions for the different type of waves (Rayleigh and Love) are given in the chapter 7 of Aki & Richards (2002). In the far field, the Bessel functions decay as $r^{-1/2}$. When the medium is not laterally homogeneous, but still smooth enough to separate the vertical and horizontal dependence (equation 1.15), the Helmholtz equation (1.19) can be adapted by making $k_n(\omega)$ depend on (x, y). This means that $k_n(\omega)(x, y)$ is determined by the resolution of the equation (1.16) with different 1D ground models (ρ, λ, μ) at different locations (x, y). In this formulation, the Helmholtz equation (1.19) can be solved numerically. One possible method is the eikonal ray tracing (high-frequency approximation, Gjevik, 1974). A discussion on the validity of this representation in presence of heterogeneity can be found in Wielandt (1993) and Yang & Forsyth (2006).

In this work, numerical modelling is performed to solve the full wave equation (1.7) in 2D by the spectral element method (SPECFEM2D, Komatitsch et al., 1999). In terms of numerical modelling, a 3D code leads to 2D surface waves, while a 2D code leads to 1D surface waves. Contrary to the 3D case where the Rayleigh wave amplitude decays with distance as $r^{-1/2}$, it does not decay at all in our case (2D), if there is no intrinsic attenuation. In Chapter 4, this generates standing waves as a superposition of the 1D incident and scattered wavefields. This effect does not exist in 3D wavefields, where surface waves propagate in 2D. This is an important drawback of my modelling workflow, and one of the motivations for using a 3D code, discussed in Chapter 7.

In 2D, addressing a non-homogeneous equation with a vertical point force source term at some depth z_0 leads to (equation 7.111 in Aki & Richards, 2002)

$$u_z(x, z, t) \propto \sum_n U_{i,n}(z_0, \omega) U_{i,n}(z, \omega) \exp\left[i(kx - \omega t)\right] \ . \tag{1.20}$$

As we have seen that the eigenfunctions $U_{z,n}$ penetrate deeper for higher modes, deeper sources are more efficient for exciting higher modes than shallow sources. This property is used in the spectral element simulations performed in Chapters 4 and 6 to obtain Rayleigh wave overtones.

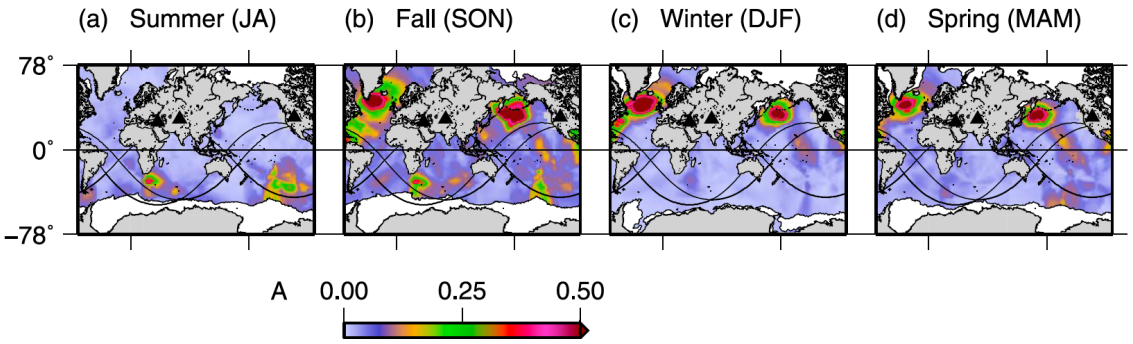


Figure 1.19: Seasonal ecolution of the microseism source zones. From Hillers et al. (2012)

.

1.3.4 Coda waves

Contrary to *ballistic* surface and body waves, i.e. those which propagate with a velocity predicted by the wave equation, coda waves do not propagate coherently, but rather diffuse apart in the wake of a ballistic wave (Margerin & Nolet, 2003). They are due to multiple scattering in a randomly heterogeneous medium and can be represented as a random walk of the rays from one scatterer to another. Well visible in Fig. 1.11 for an ambient noise example, they are also present in earthquake recordings and appear as long-lasting noise after the main arrivals. Despite the chaotic behaviour of individual rays, the coda wave part of a seismogram, as a whole, is a stable feature characteristic of the medium properties. Another important aspect of the coda waves is that they contribute to the emergence of a *diffuse* wavefield, which is close relationship with the ambient noise cross-correlation methods presented in the next Section (see 1.4.2).

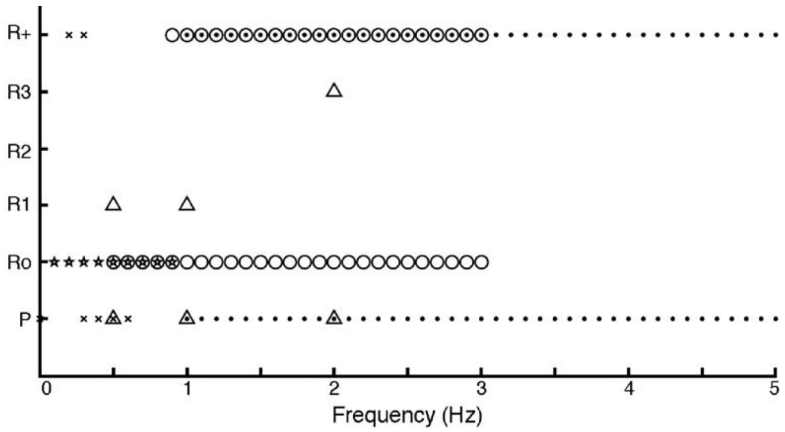
1.4 Which methods for ambient noise ?

The purpose of this section is to see which wave types are likely to be present in the ambient noise at the frequencies relevant for this study, and how they can be used.

1.4.1 Composition of the ambient noise

Ambient noise is generated by different types of sources, according to the geographic location and the frequency band. The most energetic frequency band is located around 0.1-0.2 Hz and consists of two peaks in the power spectrum, refered to as primary and secondary microseisms. The theory for the generation of these waves was developped by Longuet-Higgins (1950) and Hasselmann (1963), and is related to the ocean wave dynamics. A review of the different mechanisms generating these waves is proposed by Nishida (2017). The particular ocean zones responsible for the miscroseism generation are relatively well known. As illustrated in Fig. 1.19, they are the most active in winter in each hemisphere. The different waves excited by the oceanic microseims have been actively studied in the recent years. Nishida (2017) provides a nearly exhaustive list of references. In the on-shore noise recordings, the frequency band of the microseisms is usually dominated by surface waves (mainly

(a) Bonnefoy-Claudet et al. (2006b)



(b) Koper et al. (2010)

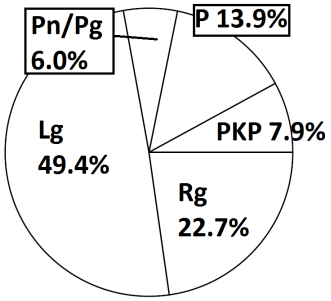


Figure 1.20: (a) «Synthesis of the type of waves (body waves or Rayleigh waves), according to frequency, contained in the seismic noise wavefield. Letter P refers to body waves and letter R to Rayleigh waves (subscript indicate the order of Rayleigh mode: 0 for fundamental mode, 1 for first mode, 2 for second mode, 3 for third mode, and + when there is no order precision)». From Bonnefoy-Claudet et al. (2006b). (b) «Phase distribution of noise from the 18 arrays over the time period of 2007-2008. Apparent velocities used to separate the phases are: >25 km/s for PKP, 25-8 km/s for teleseismic P, 8-5 km/s for regional P, 5-3.5 km/s for Lg, and 3.5-2.5 km/s for Rg». From Koper et al. (2010).

the fundamental mode), while body waves are also present and can be separated by velocity analysis (e.g. beamforming, Landès et al., 2010)). Our frequency band of interest, however, is above the frequency of the microseism, as we work above 1 Hz. For this frequency band, Fig. 1.20a illustrates the variety of the possible wave types as a function of frequency (from Bonnefoy-Claudet et al. (2006b), based on 5 papers analysing the ambient noise composition). For comparison, Fig. 1.20b shows the statistical results of Koper et al. (2010), based on the analysis of 18 arrays around the world in the frequency range [0.4-4] Hz. If we consider the L_g phase recorded on the vertical component as a superposition of many Rayleigh wave overtones, as explained in Section 1.3.3, both studies indicate that the largest amount of energy above 1 Hz is carried by the Rayleigh wave overtones. Of course, this depends on the location of the array. In the vicinity of active human activity zones, the fundamental mode tends to prevail (Bonnefoy-Claudet et al., 2006b). In Chapter 3, I show how the natural L_g and P_g phases emerge during the quietest periods (night-time), while the fundamental mode dominates during the day. Interestingly, according to the study by Koper et al. (2010) (Fig. 1.20b), a large variety of body waves can also dominate the ambient wavefield in some time windows: refracted phases P_g and P_n (see Fig. 1.12a), mantle phases P (see Fig. 1.12b) and even core phases PKP (see Fig. 1.12c). This is confirmed by number of other studies. Zhang et al. (2009) report wind-driven P_g around 1 Hz. Moho-refracted PmP and SmS (see Fig. 1.12a) are reported by Poli et al. (2012). Teleseismic phases (Figs 1.12(b) and (c)) are observed by Pratt et al. (2017). Diving waves in the ambient noise (Fig. 1.12d) are detected by Roux et al. (2005) and Nakata et al. (2015). Finally, reflections (Fig. 1.12e) are

reported by Draganov et al. (2009) and Ruigrok et al. (2011). This illustrates the variety of possible searching directions when investigating the origins of hydrocarbon micro-tremors. Rayleigh waves are the most straightforward choice, as they usually represent the largest amount of energy among the ballistic phases. However, other possibilites should definitely be explored, such as for example coda waves. Though not detected by array methods, the latter could in fact carry more energy than the most energetic ballistic phases, and thereby be more relevant for interpreting the raw PSD spectra, in which the HMT signatures are reported. For now, ambient noise coda have essentially been studied for monitoring purposes by tracking small phase distortions («stretching»). For this topic, I refer to Snieder (2006); Sens-Schönfelder & Wegler (2006); Obermann et al. (2013, 2016).

As I focus on Rayleigh waves in my work, I now present the main surface wave methods.

1.4.2 Surface wave analysis methods

I attempt to summarise the main surface wave methods in Fig. 1.21. A surface wave method is to be understood as an imaging tool, i.e. obtaining a velocity model based on surface wave recordings. The objective is to classify the methodology developed in this manuscript (i.e. *inversion of lateral perturbations of Rayleigh wave amplitudes*) with respect to standard methods. For this purpose, I identify three overlapping families of methods. The first separation is made between the methods which mainly rely on the phase and on the amplitude information. The second separation is between the methods which do or do not require a local 1D assumption.

Methods based on the phase

Most of phase methods rely on the dispersion curve inversion. Since Thomson (1950) and Haskell (1953) formulated a forward modelling method for predicting the surface wave dispersion curves from a 1D multi-layered model (ρ,V_s,V_p), the inverse problem has been extensively used as an imaging method based on the measured dispersion curves. As V_s has a dominant influence on the predicted dispersion curves, the inverted models are usually also obtained for V_s. The Thomson-Haskell formalism relies on the local 1D assumption for the ground model (no lateral variations). Dispersion curves can be measured for both phase and group velocities, depending on the source and on the receiver configuration. When a dense array of receivers spaced by less then half the wavelength is available, it is relatively easy to retrieve the phase velocity. Various array methods exist for this purpose: SPAC (Aki, 1957), FK (e.g. Capon et al., 1967), HRFK (Capon, 1969), MUSIC (Schmidt, 1986). These methods can be applied to both ambient noise and earthquake recordings. The active source counterpart of these methods is the Multichannel Analysis of Surface Waves (MASW, Park et al., 1999). Obtaining a dispersion curve for a given receiver array yields a local 1D V_s profile. I use this type of approach in Chapter 3 for retrieving a 1D V_s model to be used as background model in the subsequent spectral-element simulations performed in Chapter 4.

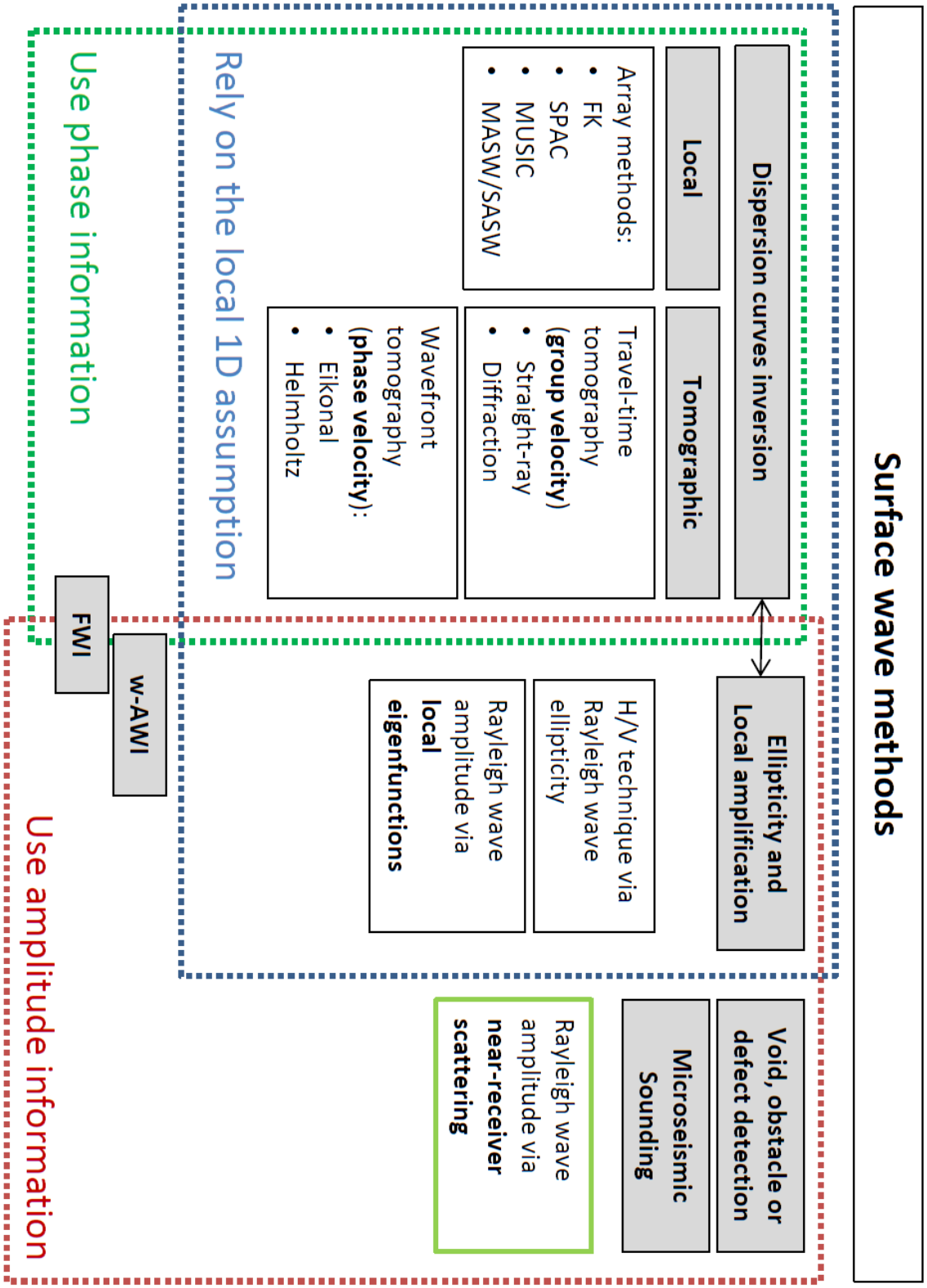


Figure 1.21: Synthetic view of the existing surface wave methods

In the early 1980s, surface wave inversion gained interest in global seismology, as global fundamental mode group velocity models for Rayleigh and Love waves could be constructed via arrival time tomography of teleseisms (e.g. Nakanishi & Anderson, 1982). Obtaining such maps at different frequencies yields a dispersion curve per model cell. Each of them can be inverted for a 1D V_s vertical profile with the Thomson-Haskell formalism, yielding a 3D model. As the periods of the teleseims were of the order of 100 s, the depth of the retrieved models was of several hundreds of kilometres, though with poor resolution. Shortly after, phase velocity measurements were also developed for single stations, and ever shorter periods were integrated in the analysis, resulting in ever better resolved global crustal models (e.g. Ekström et al., 1997). A review of these techniques is provided by Romanowicz (2002). Ritzwoller et al. (2002) extended the tomography based on the ray theory to a diffraction tomography, accounting for the arrival time perturbations due to scattering occurring within the first Fresnel zone (i.e. scattered wave in phase with the main wave to within $\pi/2$). This technique uses the Born approximation and the associated Fréchet derivatives (or sensitivity kernels), which are important in Chapters 4 and 6 of this manuscript for predicting and inverting Rayleigh wave amplitude perturbations. However, I do not use the same Green's functions to describe the *scattered* waves, as in our case they are not necessarily surface waves.

Surface wave tomography was replicated to the ambient noise analysis after the emergence of the ambient noise cross-correlation theory in the early 2000s (Shapiro & Campillo, 2004). A correlation between two ambient noise signals over a long enough time period was found to yield the surface wave part of the Green's function (see Section 1.3.1). Other parts of the Green's functions were recovered in other datasets. Many of the examples of body wave detection in ambient noise, listed in the previous section, are also due to the ambient noise cross-correlation. Using this technique, a virtual source can be reconstructed at each passive receiver, allowing for a tomography with the number of rays equal to $N(N-1)/2$, with N the number of passive receivers. One of the first applications was performed on the Réunion Island, France, by Brenguier et al. (2007). They used the frequency-time analysis (FTAN, Levshin & Pisarenko, 1972) to measure the group velocities for different station pairs. This allowed to image a high-velocity intrusive body in a volcanic context.

For dense arrays, another type of tomography was proposed by Lin et al. (2009) and Lin & Ritzwoller (2011), using the continental scale USArray. It is based on fundamental mode wavefront tracking. $g(x,y)$ in the Helmholtz equation (1.19) can be written in a general form

$$g(\mathbf{x},\omega) = A(\mathbf{x},\omega)\exp(i\omega\tau(\mathbf{x}))\ , \tag{1.21}$$

τ(x) being homogeneous to a phase arrival time at $\mathbf{x}=(x,y)$. The Helmholtz equation then splits in two coupled equations for A and τ

$$\begin{cases} ||\nabla\tau||^2 - \dfrac{\nabla^2 A}{\omega^2 A} = \dfrac{1}{\left[c_0(\omega,x,y)\right]^2}\ \text{(«Helmholtz» equation)}, \\ 2\nabla A\cdot\nabla\tau + A\nabla^2\tau = 0\ \text{(Transport equation)}, \end{cases} \tag{1.22}$$

where $c_0(\omega, x, y) = \omega / k_0(\omega, x, y)$ is the phase velocity of the fundamental mode. The transport equation describes the amplitude variation through wavefront focusing/defocusing (Lin et al., 2012). Assuming the frequency high enough or the amplitude field smooth enough in the upper equation (1.22) leads to the Eikonal approximation, via dropping the second term

$$||\nabla \tau||^2 = \frac{1}{[c_0(\omega, x, y)]^2} \text{ (Eikonal equation).} \tag{1.23}$$

Thus a phase velocity map can be obtained directly from the phase front. The phase arrival times on a dense array can be measured by a technique similar to phase unwrapping (Lin et al., 2008). One of the advantages of this method is that Rayleigh-Rayleigh scattering within the first Fresnel zone is naturally accounted for in this approach, contrary to the straight ray tomography. The Eikonal approximation (1.23) was applied to the ambient noise data by Lin et al. (2009) (Eikonal tomography), while the full upper equation (1.22) was used by Lin & Ritzwoller (2011) for teleseisms and Mordret et al. (2013b) for ambient noise. In the latter case, contrary to all the previously described tomography methods, the amplitude starts to be taken into account, though the major constraint still comes from the phase. The Helmlholtz tomography, in fact, does not use all of the amplitude information, but only the curvature of the amplitude field. Interpreting the full amplitude variations accross the array, however, is the original motivation of the present work. The next paragraph presents the methods which address this issue.

Methods based on the amplitude

Lin et al. (2012) write the full measured amplitude as a product between the amplitude governed by the wave propagation effects $|g(x, y)|$ (see equation 1.15) and a local amplification term $\beta(x, y)$, related to the spatialised $f_z(z)$ in equation (1.15). They also add an anelastic damping term in the Helmholtz equation (1.19) in order to account for amplitude decay with distance, as previously suggested by Prieto et al. (2009). While the Eikonal approximation remains unchanged, the transport equation now contains focusing/defocusing, anelastic attenuation and local amplification terms. As far as the local 1D approxiation at the basis of the Helmholtz equation holds, this is sufficient to fully describe the amplitude. Jaxybulatov (2017) proposed an ambient noise application for this method. The method looks promissing for being applied to dense datasets such the one recorded in Chémery (Section 2.1) at frequencies below 1 Hz, where coherent surface waves are dominating. It seems complicated, however, to use this approach to address the HMT-related amplitude perturbations, for the following reasons:

1. The wavefield at the HMT frequencies is multi-modal in Chémery (see Chapter 3), so the fundamental mode phase front would be hard to track.

2. The coherence of the cross-correlations above 1 Hz drops brutally because of the increasing scattering (Lehujeur et al., 2015), so, again tracking a coherent wavefield from individual virtual sources across the array would be hard.

3. The amplitude anomalies in Chémery are detected in the raw ambient noise recordings. It is yet unclear whether they are present in the cross-correlations (necessary to reconstruct wave fronts).

Other techniques based on the amplitude of surface waves rely exclusively on the the 1D eigenfunctions and do not account for the propagation phenomena. This is the case of the H/V ratio inversion interpreted as the Rayleigh wave elliplicity already discussed in Section (1.2). Finally, the last family of techniques address the amplitude perturbations due to Rayleigh wave scattering, i.e. to secondary (not necessarily Rayleigh) waves generated when an incident Rayleigh wave encounters a sharp obstacle. These are used in non-destructive concrete testing (Jagnoux & Vincent, 1988; Hévin, 1998), burried void/rigid body detection (Gucunski et al., 1996; Grandjean & Leparoux, 2004; Gelis et al., 2005; Nasseri-Moghaddam et al., 2005, 2007; Tallavó et al., 2009; Yu & Dravinski, 2009; Chai et al., 2012) , near-receiver scattering suppression in active seismics (Campman et al., 2004, 2005), as well as imaging of structural heterogeneities (Microseismic Sounding Method, Gorbatikov et al., 2008, see discussion and other references in Section 1.2.3). As already stated in Section 1.2.3, I choose this category of methods as the closest to the empirical observations of the vertical component motion amplification above reservoirs. The reasons of choosing this approach and the consequences of this choice are discussed in Section 1.5.

Before, two relatively recent methods of surface wave inversion must be mentioned, mostly targeting active source acquisitions. The first is the Full Waveform Inversion (FWI), used by Brossier et al. (2009) to jointly invert body and surface waves. Surface waves appeared to increase the non-linearity of the inversion because of their dispersive behaviour and strong sensitivity to the near-surface layers. Solano et al. (2014) proposed an alternative method, the windowed amplitude waveform inversion (w-AWI), based on the inversion of f-k amplitude spectra constructed from successive overlapping receiver subgroups. While much of the phase information is discarded by this method, the dispersion curves are still indirectly used, as they are contained in the f-k amplitude spectra. w-AWI proved more robust than FWI for surface wave inversion. Both methods rely on the full numerical modelling of the wavefield, without explicit extraction of surface waves. To some extent, the inversion method I develop in Chapter 6 can be regarded as a linearised version of w-AWI for small perturbations, and with the receiver subgroup reduced to one sensor. An effort is currently being made by a reasearch group to extend FWI to ambient noise applications (Sager et al., 2018).

1.5 Rayleigh wave scattering as a possible mechanism for hydrocarbon microtremors

In this section, I expose my modelling strategy for attempting to explain the nature of the hydrocarbon microtremors. In Section 2.1, I show that the vertical component amplification is the attribute which exhibits most clear correlation with the hydrocarbon presence for our largest dataset, consistently with the originally reported HMT (see Section 1.2). The frequency band where this effect is the most visible is [1.2-2.4] Hz. The V/H ratio increase (another frequently used attribute, see Section 1.2) in the same frequency band is not well correlated with the known reservoir extent. *Thus I choose the vertical component amplification as the first phenomenon to be understood. This is only feature of HMT I address in the modelling part of this PhD*

work.

1.5.1 Reasons for analysing Rayleigh wave scattering

A passive mechanism

At the end of the *Low Frequency Passive Seismic* review in Section 1.2, I concluded that a passive effect (i.e. not an active emission) was the least contradictory explanation for this phenomenon. On the other hand, based on the published studies on the short-period ambient noise content (see Section 1.4.1), one can expect surface waves to dominate the frequency band of interest (one to several Hz), in most of cases. Thus I decide to investigate whether surface wave scattering could be the cause of the reported HMT. As the surface wave penetration depth is frequency-dependent, a reservoir placed at a certain depth is likely to generate effects in some particular frequency band, which might be consistent with the HMT observations reported at several Hz. Rayleigh wave amplitude perturbations due to scattering beneath the array are indeed reported both at smaller (Nasseri-Moghaddam et al., 2007; Tallavó et al., 2009) and at larger (Gorbatikov et al., 2008, 2013) scales, so it appears interesting to test this mechanism at reservoir scale.

Testing geology as an alternative explanation

Moreover, neither the results from the literature nor my own experimental investigations (Chapter 2) allow to discard the geological structure as the origin of the observed anomalies. As hydrocarbon reservoirs are often associated with structural tops, the latter could also be the origin of HMT. From this point of view, it is useful to model the propagation of Rayleigh waves across a realistic anticline structure and to analyse the associated amplitude distortions (Chapter 4).

Potential difficulties of the theory

The first problem which arises with this mechanism is that reservoirs as deep as 1 km may be beyond the penetration depth of the fundamental mode Rayleigh waves. Assuming that they propagate at $c = 800$ m/s at $f = 2$ Hz, which is what is observed in the Chémery dataset (see Chapter 3), the associated wavelength is $\lambda = 400$ m, which is then also the fundamental mode penetration depth. Thus the presence of Rayleigh wave overtones is required to expect an interaction with a reservoir at kilometric depth at the typical HMT frequencies. In the Chémery dataset, overtones are indeed observed during night-time (L_g phase, see Chapter 3 and also Koper et al. (2010)). Because of that, one cannot discard to possibility of Rayleigh waves being scattered by a deep heterogeneity such as a reservoir.

The second problem is that it is hard to expect exclusively an amplification effect from a scattering mechanism, since it generally generates an alternating amplification/attenuation pattern over frequencies (constructive/destructive interference with the incident field). However, in case of near-receiver Rayleigh wave scattering, the

receiver lies in the near-field with respect to the scatterer, and I was a priori not sure of the type of effects I could expect from such a situation. The radiation patterns for this type of scattering are investigated in Chapter 5, while the effects on the amplitude are modelled in Chapter 4.

Why not local 1D amplification?

As the primary objective is to model a reservoir effect, the scale of the reservoir must be compared to the wavelength. As explained above, we expect the reservoir to interact with waves of wavelength comparable with the reservoir depth, i.e. kilometric. On the other hand, in the case of the Chémery réservoir I study in this work, the lateral extent is of about 2 km, comparable to the wavelength. Thus I think a scattering approach is more appropriate is this context, rather than locally considering a laterally infinite reservoir. As already stated, the local 1D hypothesis is behind the local amplification coefficient used by Lin et al. (2012). Moreover, the numerical modelling part of my work (in particular the Chapter 4) does not depend on how we call the mechanism, i.e. «scattering» or «1D amplification». If the major reason of the amplitude distortions turns out to be a smooth geological structure, 1D amplification might be a more appropriate analytic tool than scattering modelling, but both effects are contained in the numerical simulation. Focusing/defocusion effects, however, are a 3D phenomenon which remains beyond the scope of this work.

1.5.2 Specific features of near-receiver scattering

The most straightforward approach to model the propagation of Rayleigh waves above a reservoir and/or a geological structure is by direct numerical modelling. This is what is done in Section 4. However, in order to better understand the sensitivity of the Rayleigh amplitude with respect to the sub-surface perturbation, I also implement a semi-analytical approach based on the Born approximation (Hudson & Heritage, 1981) in Chapters 5 and 4).

The scatterer is below the receiver

In Sections 1.3.3 and 1.4.2, we saw that surface waves are usually represented as propagating in the horizontal XY plane. In this representation, when surface-wave scattering is addressed analytically (e.g. Snieder, 1986; Friederich et al., 1993; Maupin, 2017), both the incident and the scattered wavefields are described by the surface wave part of the Green's function. The latter is explicitly available in a layered half-space via the modal summation theory (Aki & Richards, 2002, their Chapter 7). However, waves scattered from below the receiver array are not well described by such an approach, as they are essentially body waves (in fact, near-field waves at low enough frequencies). An explicit analytic Green's function containing both surface and body waves in a layered half-space is not available, to my knowledge. Instead, it is formulated as an integral in the frequency-slowness or frequency-wavenumber domain. An integration in this domain must be performed, for example using the Cagniard-de Hoop method (see Kennett (1983); Aki & Richards (2002); Maupin (1996)). At some point, this integration must be either performed numerically, or some approximation

must be made to obtain an explicit expression in space-frequency or space-time domain, for example by choosing some particular wave-type contributions. This leads to heavy mathematical derivations I do not investigate in my work. A semi-analytical tool for near-surface scattering modelling, based on the numerical integration over wavenumbers, was proposed by Riyanti & Herman (2005), and applied by Campman & Riyanti (2007) to model and invert the scattered Rayleigh waves for the scatterer properties. The problem they consider is very close the ours, except they use a single source instead of a random wavefield, and they model the full waveform, instead of the spectral amplitude.

In this work, I use numerically estimated Green's functions, which contain all the available propagation modes. More precisely, I use numerical Green's functions for both incident and scattered wavefields, which is useful for addressing a background medium containing arbitrary structural perturbations (see Chapter 4). However, using analytical Green's functions obtained by modal summation for the incident Rayleigh waves is an option that should be considered to speed up calculations and to insure there are no body wave contributions.

1.5.3 Numerical modelling

Numerous numerical algorithms exist for solving the elastic wave equation (1.6). A review is proposed by Virieux et al. (2011). The authors distinguish between three main families of methods, differing by their spatial discretisation strategy: spectral methods, strong formulation methods and weak formulation methods. My choice was to use the spectral-element method (SEM) in 2D (Komatitsch & Vilotte, 1998; Komatitsch et al., 1999, SPECFEM2D), which belongs to the category of the weak formulation methods.

This choice was motivated by rather practical reasons, while other numerical techniques could be more relevant in the different parts of this manuscript. First, SEM allows to address variable geometries, which is useful for modelling the effect of an anticline structure (Chapter 4). This could probably also be achieved with a finite-difference scheme (Virieux, 1986; Saenger et al., 2000), as the geometry of the anticline is smooth enough. However, implementing a workflow based on SEM allows for studying more complex layers and free-surface topographies in the future. Second, the numerical dispersion with SEM is low, which allows to take a limited number of points per wavelength when studying surface waves (5 instead of 15 with finite-differences, Komatitsch & Vilotte, 1998). Third, the SPECFEM2D code has been extensively tested and is open-source, which represents a considerable gain of time. Meshes to be used as input to the spectral-element simulations were generated using Gmsh software (Geuzaine & Remacle, 2009).

According to Virieux et al. (2011), spectral techniques are very efficient in layered laterally invariant media. While they could be insufficient to model the scattering itself, or the propagation across the anticline structure, they could efficiently estimate the reference medium Green's function in the laterally invariant models studied in Section 4. The discrete wavenumber method (Bouchon, 1979) is an example of spec-

tral method.

The strong formulation methods, such as the finite-difference method, are generally less time-consuming than SEM and should be considered as an alternative. However, they generally use uniform grids. This means that the element size everywhere in the model is imposed by the 15 points per wavelength condition with respect to the slowest Rayleigh wave velocity in the model, while the mesh size can be adapted with SEM (see Chapter 4). Also, the few tests I realised with finite-differences in the beginning of my PhD work exhibited a divergent behaviour for long simulation times, which are required to estimate the spectra with enough precision at low frequency. This complication also motivated the choice of SPECFEM2D.

Finally, semi-analytical methods are also to be mentionned. Yu & Dravinski (2009) used the boundary-element method (BEM) to model the Rayleigh wave scattering by an embedded cavity, based on the piece-wise knowledge of the Green's function in different parts on the model. Riyanti & Herman (2005) also proposed a method for surface wave scattering modelling based on the analytical Green's functions in a layered medium, without performing the Born approximation. While these methods are suitable and efficient for simple configurations, they would not allow to address the propagation across the anticline structure.

1.5.4 Modelling strategy

To summarise this part, my modelling strategy is as following. First, I perform full spectral element simulations of distant random sources generating preferentially Rayleigh waves at the array. A reservoir and/or a geological structure are embedded into the model, and the respective effects are compared. I use the Born approximation to reproduce the numerical results for small contrasts and to prepare a framework for a linear inversion of the amplitude perturbations observed at the surface (Chapter 4). Second, I investigate the characteristics of the near-receiver Rayleigh wave scattering in terms of equivalent secondary sources and radiation patterns, in order to see if the effects of different elastic parameters can be distinguished (Chapter 5). For this part, I use a homogeneous halfspace as background model. Finally, in Chapter 6, I investigate the possibility of inverting the Rayleigh wave amplitude perturbations at surface for the scatterer shape, position and properties, in very simple media only (homogeneous half-space/layer-over-halfspace). This workflow is summarised in Fig. 1.22. The block resulting in the definition of the reservoir elastic properties is described in the next section (see 1.6.2).

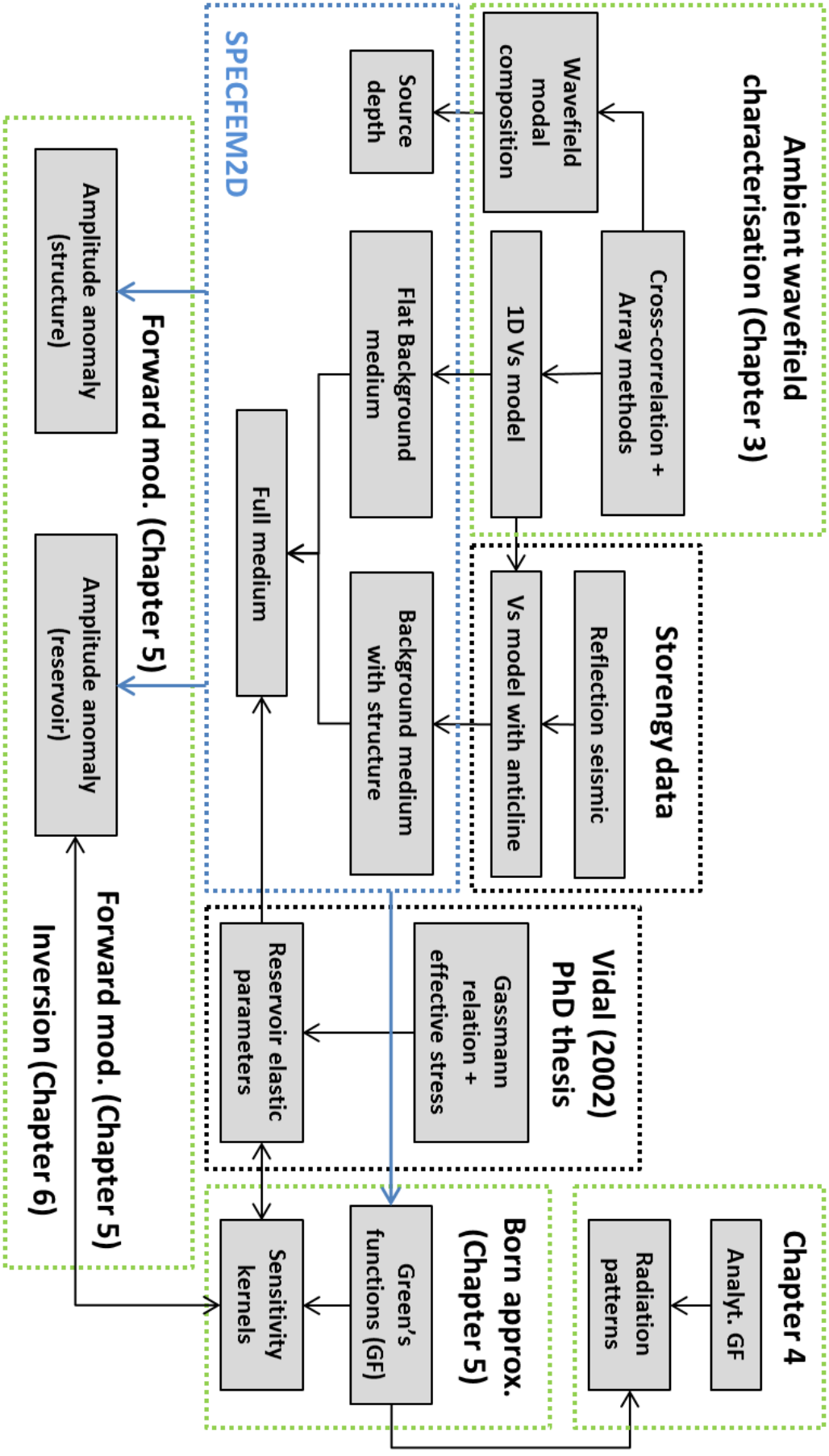


Figure 1.22: Modelling workflow. Green boxes correspond to the parts of the workflow implemented during this PhD work.

1.6 Effective properties of a gas reservoir

As scaterring was chosen as a potential HMT mechanism (Section 1.2.3), the knowledge of the contrast value in the reservoir is required. The strength of this contrast conditions the strength of the scattered signal. In this section, I provide a very brief overview of how to account for the fluid saturation in the reservoir when assigning its elastic and visco-elastic properties, i.e. the density ρ, the seismic velocities V_p, V_s, and the quality factors Q_p, Q_s. As attenuation plays an important role in this section, I add a preamble to define the main related notions.

Attenuation and quality factor

The quality factor is a generic way to describe attenuation due to any energy loss mechanism affecting the wave propagation. Considering the Fourier component ω of a plane wave propagating in the x direction, the amplitude decay can be included by introducing a complex wavenumer $k(\omega)$, so that (Ursin & Toverud, 2002)

$$u(x,t) = A(\omega) \exp i \left[k(\omega)x - \omega t\right] , \tag{1.24}$$

with

$$k(\omega) = \frac{\omega}{c_p(\omega)} + i\alpha(\omega) , \tag{1.25}$$

where $c_p(\omega)$ is a *real* phase velocity and $\alpha(\omega)$ is the attenuation. On the other hand, the full *complex* phase velocity is formally defined as

$$c(\omega) = \frac{\omega}{k(\omega)} = \sqrt{\frac{M(\omega)}{\rho}} , \tag{1.26}$$

where $M(\omega)$ is some complex rigidity. The quality factor is conventionnaly defined as

$$Q(\omega) = \left|\frac{M_R}{M_I}(\omega)\right| , \tag{1.27}$$

with M_R and M_I the real and complex parts of M. The relation between Q and α is

$$Q = \frac{\omega}{2\alpha c_p} - \frac{\alpha c_p}{2\omega} , \tag{1.28}$$

where the second term on the right-hand-side can be neglected in most of cases. While Q is the intrinsic attenuation factor, O'Connell & Budiansky (1978) gave an expression for the *seismic* quality factor $Q^{(s)}$, which measures an energy loss per oscillation:

$$Q^{(s)} = 2\pi \left[1 - \exp\left[\left(Q^2+1\right)^{1/2} - Q\right]\right]^{-1} . \tag{1.29}$$

Distinguishing both difinitions is important for comparing the published results, since both are used in the literature. However, the values are of similar order, and asymptotically equivalent for high Q. The quality factors for plane P- and S- waves are Q_p and Q_s, and can be regarded as medium characteristics. Typical values of $Q_{P,S}$ are of several hundreds for well consolidated rocks on Earth. Q values as low as 10 can be regarded as extremely low. I refer to Ursin & Toverud (2002) and references herein for any further detail on the quality factor definition and properties.

1.6.1 Concept of effective medium

In a homogeneous isotropic 3D medium, the velocities of P- and S-waves are given by

$$V_p = \sqrt{\frac{K + 4/3\mu}{\rho}} \ , \tag{1.30}$$

$$V_s = \sqrt{\frac{\mu}{\rho}} \ , \tag{1.31}$$

where K and μ are respectively the bulk and shear moduli of the medium. In this work, I always stay in an isotropic context. Thus knowing the two elastic moduli and the density is sufficient to obtain the seismic velocities. If we consider the true medium is a composite material, with $K^{(i)}$ and $\mu^{(i)}$ the moduli of each phase (solid of fluid), the *effective medium* is a fictitious homogeneous medium which correctly describes the seismic wave propagation in the whole medium. The density of the effective medium is readily obtained as

$$\rho_{\text{eff}} = \sum_i s_i \rho_i \ , \tag{1.32}$$

where s_i are the volumetric fractions of each phase. Mavko & Mukerji (1998) show that the elastic moduli (generically written M) of the effective medium are bounded by the Reuss (minimum) and the Voigt (maximum) averages of the moduli of the individual phases (M_i). The Voigt average is

$$M_V = \sum_i s_i M_i \ , \tag{1.33}$$

and corresponds to the case of an iso-displacement propagation. This means that the pressure gradients do not have time to relax during the wave's period. The Reuss average is

$$\frac{1}{M_R} = \sum_i \frac{s_i}{M_i} \ , \tag{1.34}$$

and corresponds to the case of an iso-stress wave propagation. This means that the pressure gradients arising at the boundaries of the different phases (due to the different stress-strain relations) relax on a time scale smaller than the wave's period. As the Voigt average is the upper limit of the effective medium's velocity, this means that high-frequency waves travel faster than low-frequency waves. High-frequency and low-frequency limits are separated by a characteristic frequency f_0, which depends on the scale of the material heterogeneity. Wavelengths far below this scale are in the high-frequency regime, while wavelengths far above this scale are in the low-frequency regime. At the transition (i.e. close to f_0), the wave propagation exhibits dispersion and attenuation (Aki & Richards, 2002). This is illustrated in Fig. 1.23. Johnston et al. (1979) suggest that at seismic frequencies the attenuation can be considered as frequency-independent and dominated by friction in the rock matrix, for both dry and saturated rocks. However, recent laboratory measurements and numerical simulations in partially saturated sandstones (Chapman et al., 2016, 2017; Subramaniyan et al., 2017) demonstrate a frequency-dependant fluid-specific attenuation with low characteristic frequencies (several Hz in some cases). In this short overview, I only

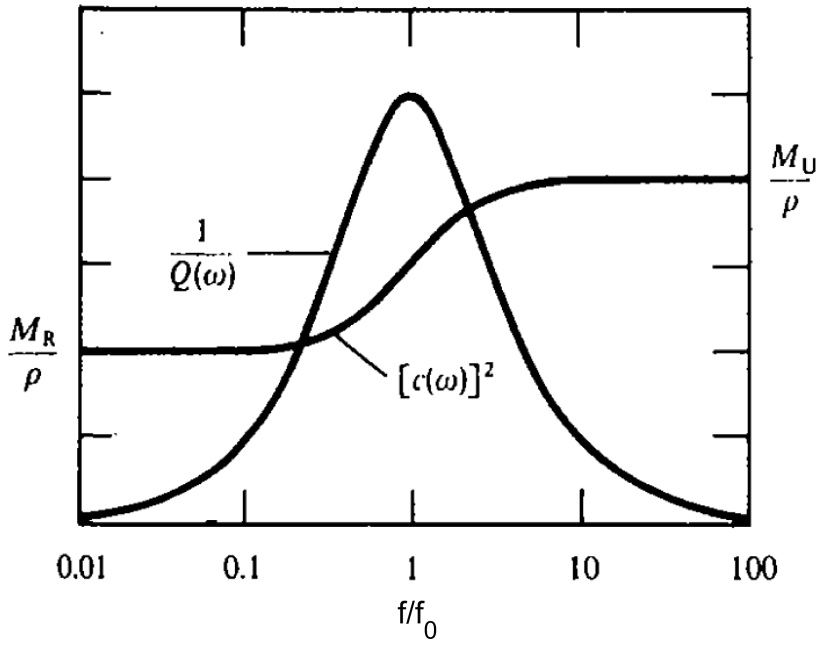


Figure 1.23: Dispersion and attenuation associated with the transition from the low-frequency (relaxed) to the high-frequency (unrelaxed) propagation regime. M_U and M_R stand for unrelaxed and relaxed elastic moduli. M/ρ is the squared phase velocity $[c(\omega)]^2$. From Aki & Richards (2002).

focus on this frequency-dependent attenuation, well described by the generic model in Fig. 1.23.

In our case, we are interested in predicting porous rock effective properties due to fluid (e.g. gas) saturation. The rocks surrounding the reservoir are fully water-saturated, while those inside the reservoir have a non-zero gas or oil saturation. Thus there are two questions to be answered for predicting the contrast between the reservoir and the surrounding rocks:

1. What are the effective properties of a rock fully saturated with one fluid ?
2. What are the effective properties of a rock saturated with a mixture of fluids ?

In the next two Sections (1.6.2 and 1.6.3), we consider that the solid matrix of the porous rock is homogeneous, and the only source of heterogeneity is the presence of fluid phases. In Section 1.6.4, we also consider the effect of matrix heterogeneity on the effective properties. The latter case is especially relevant in the context of highly fractured (e.g. geothermal) or highly heterogeneous reservoirs.

1.6.2 Elastic modelling: the Gassmann relation

In the first case (full saturation with one fluid), the scale of the heterogeneity is the pore size, which rarely exceeds 1 mm. Thus, at seismic frequencies (1-100 Hz), we are definitely in the low-frequency limit. The associated elastic moduli of the saturated

medium are then given by the Gassmann's relation (Gassmann, 1951)

$$\begin{aligned} \frac{K_{sat}}{K_{gr} - K_{sat}} &= \frac{K_{dry}}{K_{gr} - K_{dry}} + \frac{K_{fl}}{\Phi \left(K_{gr} - K_{fl}\right)} \ , \\ \mu_{sat} &= \mu_{dry} \ , \end{aligned} \tag{1.35}$$

where Φ is the porosity; K_{gr}, K_{dry}, K_{sat}, K_{fl}, are respectively the bulk moduli of the solid grains, the dry matrix, the saturated rock and the pore fluid; μ_{sat} and μ_{dry} are the shear moduli of the saturated and dry rock. An important consequence of the Gassmann's theory is that the shear modulus is not sensitive to the saturation in the low-frequency limit. As we are far from f_0 at seismic frequencies, no additional attenuation is expected due to the fluid saturation.

When two fluid phases are present, this introduces an additional scale of heterogeneity, depending of how the phases are mixed. If the mixture occurs at the pore scale (uniform saturation), as illustrated in Fig. 1.24a, we are again in the low-frequency limit. Then an effective fluid can be defined with a Reuss-averaged bulk modulus over the different fluid phases (equation 1.34). For example, in the case of mixture of gas with water, the effective fluid properties are

$$\begin{aligned} \rho_{\text{fl, eff}} &= s_w \rho_w + s_g \rho_g \\ \frac{1}{K_{\text{fl, eff}}^{(\text{unif})}} &= \frac{s_w}{K_w} + \frac{s_g}{K_g} \ , \end{aligned} \tag{1.36}$$

where $s_{w,g}$, $\rho_{w,g}$ and $K_{w,g}$ are respectively the water and gas saturations, densities, and bulk moduli. This effective fluid can now be injected into the Gassmann's relation (equation 1.35) to obtain the full effective medium properties.

However, if the fluid mixture takes the shape of patches larger than the pore scale (*patchy saturation*, Fig. 1.24b), but still smaller than the wavelength, the order of averaging changes. The Reuss average should now be applied to patches saturated with different phases. The effective properties of each type of patches (e.g. water- or gas-saturated) are obtained via the Gassmann's relation (equation 1.35) with a single fluid phase. Mavko & Mukerji (1998) show that for well consolidated rocks, where the dry matrix bulk modulus is much larger than the fluid bulk modulus, the effective properties describing patchy saturation (i.e. Reuss average over patches) can be obtained by defining an effective fluid as a Voigt average of the different fluid phases (and not the Reuss average as in the case of uniform saturation).

$$K_{\text{fl, eff}}^{(\text{patchy approx})} = s_w K_w + s_g K_g \ . \tag{1.37}$$

Such an effective fluid can be directly injected into the Gassmann's relation, like in the case of a uniform saturation, without averaging over the patches. The different approaches for computing effective properties with the Gassmann's relation for both uniform and patchy saturation are summarised in Fig. 1.25. The model predictions for both patchy and uniform saturation are illustrated in Fig. 1.26 for a typical well consolidated sandstone. The uniform saturation prediction is in excellent agreement with laboratory data during the sample imbibition phase. Velocities are measured

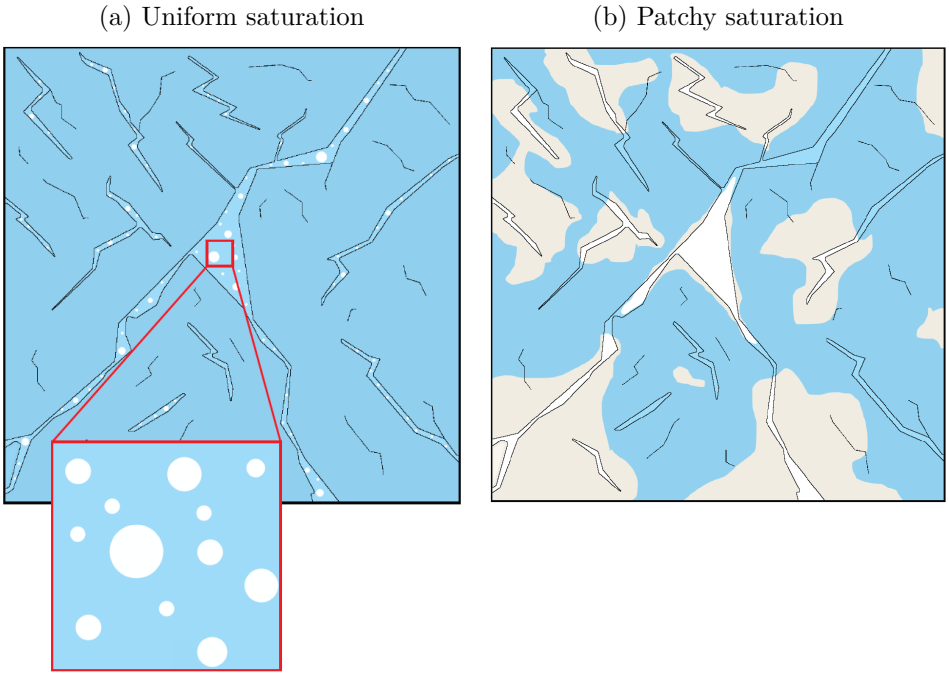


Figure 1.24: Illustration of uniform (a) and patchy (b) saturation of a porous fractured medium. Adapted after Grab et al. (2017a).

by the resonant bar technique (frequencies about 1 kHz, still in the low-frequency limit (e.g. Lucet et al., 1991)). An important outcome of the Gassmann's relation applied to a uniform saturation is that most of the effect (i.e. decrease of K_{eff} and V_p) occurs that when the first bubbles of gas start to form in an initially water-saturated rock. As the gas saturation further increases, it only has a limited influence of K_{eff} (gradual increase). On the other hand, V_s is only sensitive to the gas saturation via the density, and undergoes quite limited variation. Thus, if the Gassmann's relation is used to predict the reservoir effective properties, the simulated attributes (such as HMT strength) are expected to be sensitive to the presence or absence of gas via V_p, and to the total level of saturation via the density.

In my work, I assume uniform saturation in a sandstone reservoir representative of the Chémery UGS. This situation was modelled by Vidal (2002) for the Céré-la-Ronde UGS. The latter site is located 30 km away from Chémery (more detail in Chapter 2.1). It can be regarded as equivalent to Chémery in terms of petrophysical properties (same reservoir layers) and geomechanical behaviour (similar depth). In addition to the uniform saturation approach (Fig. 1.25a), Vidal (2002) also accounted for the effect of the fluid pore pressure, due to pumping/injection, on the dry matrix elastic moduli. The fluid pore pressure was obtained via a multi-phase flow simulation within a 3D reservoir model constrained by the available pressure and flow rate imposed at the wells. An increase in the pore pressure opposes the confining pressure of the overburden, thus decreasing the effective stress of the matrix, and results in a decrease of both shear and elastic moduli. Vidal (2002) modelled the relation between the elastic moduli of the matrix and the effective stress via the Hertz-Mindlin model (Mindlin, 1948). Their workflow for reservoir effective properties prediction is sum-

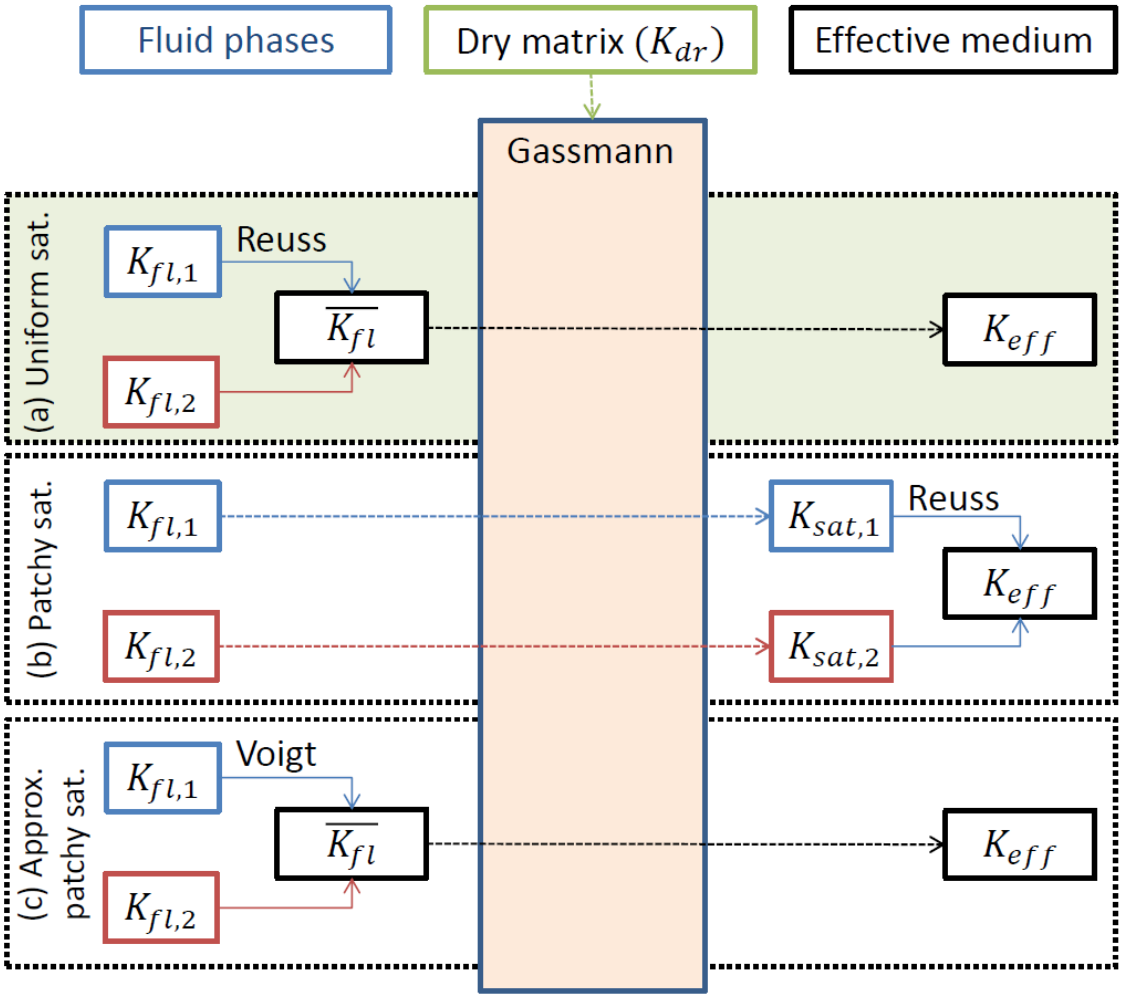


Figure 1.25: Effective elastic properties computed with the Gassmann's relation. (a) Uniform saturation. (b) Patchy Saturation. (c) Patchy saturation approximation for well consolidated rocks ($K_{dry} \gg K_{fl}$). The elastic reservoir properties used in this work are based on the workflow highlighted in green.

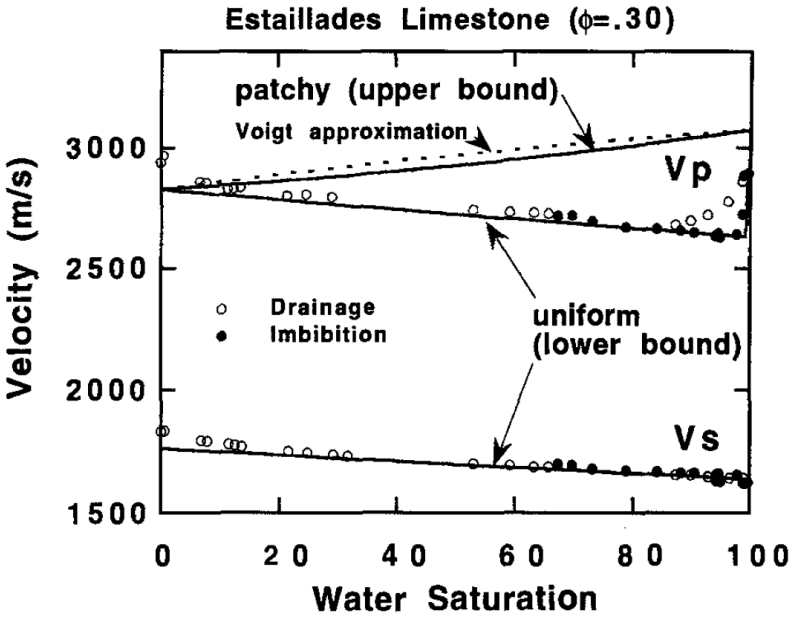


Figure 1.26: Gassmann's equation predictions for uniform and patchy saturation, compared to real data from (Cadoret & Poirier, 1993). Solid lines show the predictions for uniform and patchy saturation. The dotted line shows the patchy saturation approxiation for well consolidated rocks (Voigt-averaged effective fluid). Open (resp. filled) circles show real data during drainage (resp. imbibition) of the sample. Note the abrupt drop of V_p for uniform saturation when the first bubbles of gas appear (water saturation close to 100%). From Mavko & Mukerji (1998).

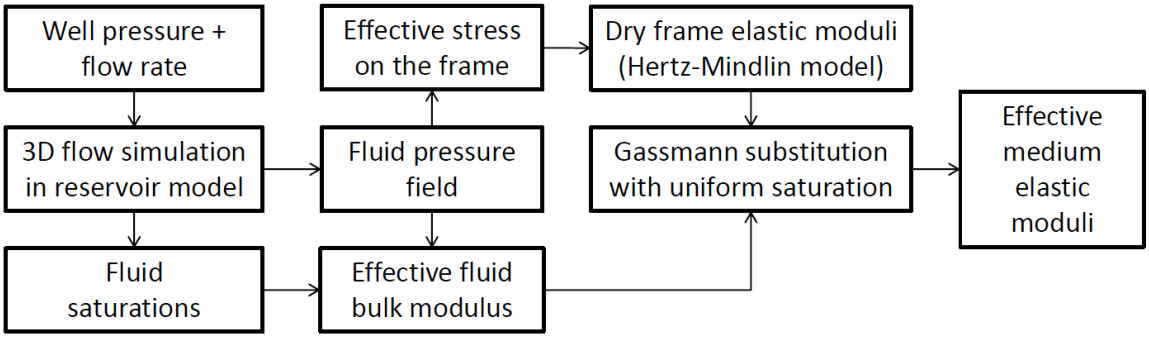


Figure 1.27: Workflow used by Vidal (2002). Dynamic flow simulation in a 3D reservoir model is used to compute fluid pressure in each cell and correct the matrix elastic moduli for effective stress variation. The Gassmann's relation for uniform saturation is then used for computing the effective moduli in the partially-saturated reservoir.

marised in Fig. 1.27. The model was applied to the first injection cycle in the UGS, i.e. when the reservoir transited from a full water saturation to a full gas saturation in some of its parts. The model predictions, validated by the sonic log data, are shown in Fig. 1.28 for V_p, V_s and ρ perturbations due to gas injection. *For the numerical simulations of elastic wave propagation (Chapter 4), I choose the maximum predicted effects in the reservoir, that is:*

$$\begin{cases} \Delta V_p/V_p^{(0)} = -18\% \\ \Delta V_s/V_s^{(0)} \approx 0\% \\ \Delta \rho/\rho^{(0)} = -10\% \ . \end{cases} \tag{1.38}$$

I assume these perturbations are equal to the contrast between the reservoir and the surrounding medium, since the initial state of the reservoir (blue curves in Fig. 1.28) corresponds to a full water saturation.

1.6.3 Attenuation due to patchy saturation

Until now, we have considered we were in the low-frequency limit with respect to the characteristic frequency f_0, even in the patchy saturation model. Following Mavko & Mukerji (1998), when the dry matrix is homogeneous, this frequency can be estimated as

$$f_0 \approx \frac{\kappa K_{fl}}{\eta L_c^2}, \tag{1.39}$$

with κ the permeability, L_c the characteristic size of a fluid patch, K_{fl} and η respectuvely the bulk modulus and the viscosity of the most viscous fluid phase. It is clear that f_0 in equation (1.39) decreases with the increase of the characteristic patch size L_c. Attenuation can be expected when f_0 corresponds to the typical frequencies of the seismic survey, i.e. several Hz in our case. Using the same reservoir sandstone properties as Kuteynikova et al. (2014), listed in Table 1.3, this would require patches of size $L_c \sim 1$ m for $f_0 = 3$ Hz. Physically, the attenuation corresponds to a relaxation mechanism called *Wave-Induced Fluid Flow*, or WIFF (see Müller

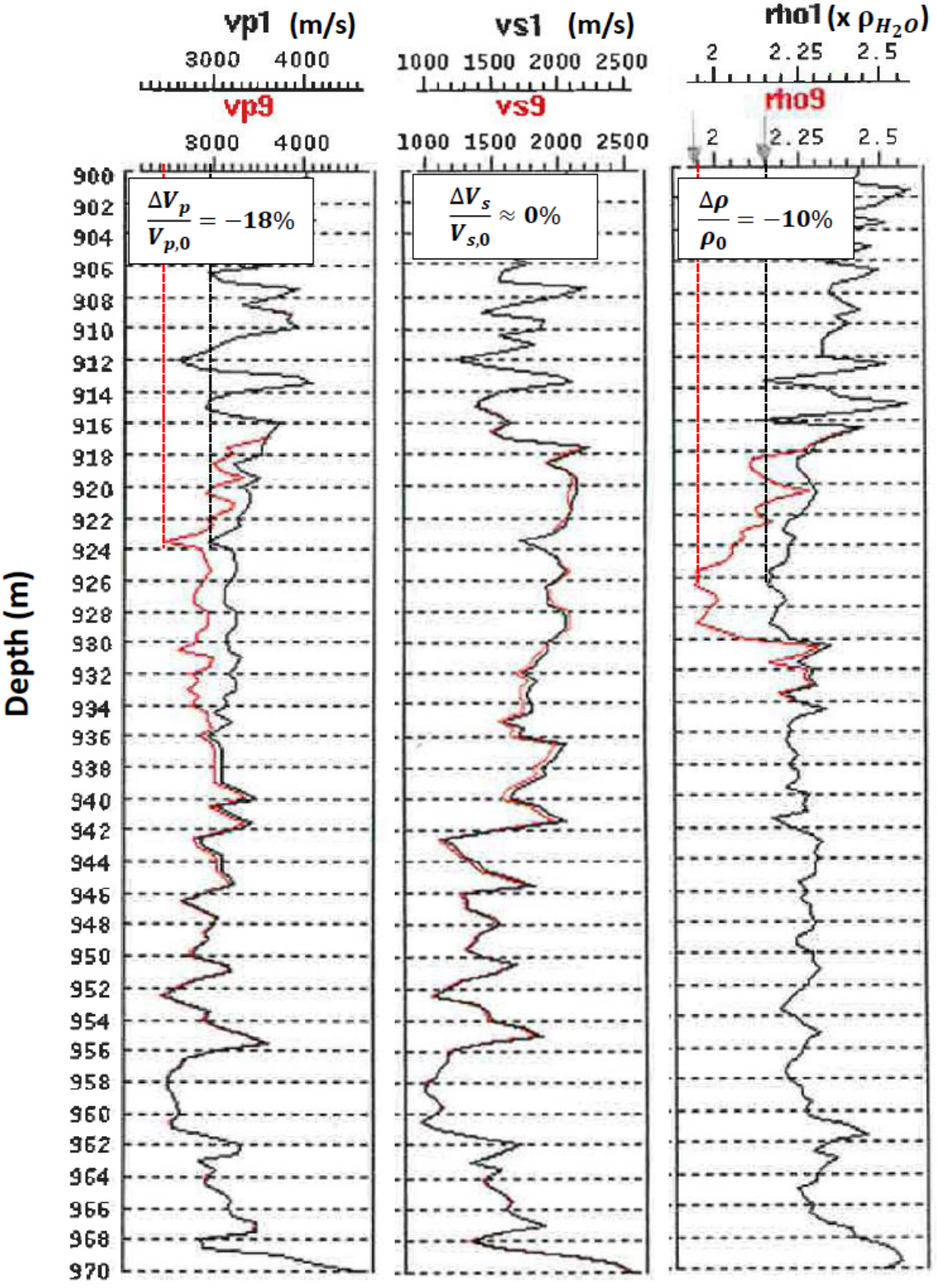


Figure 1.28: Predictions of the model by Vidal (2002) for the full substitution of water by gas in Céré-la-Ronde UGS. Black (resp. red) curves show the water- (resp. gas-) saturated state. Left: V_p, middle: V_s, right: ρ. The maximum predicted perturbations are chosen for numerical wave propagation simulations in my work.

Table 1.3: Sandstone elastic and petrophysical properties used by Kuteynikova et al. (2014). «?» stands for a meaningful parameter not given in the paper.

Region:	Solid grains	Dry matrix	Water	Air
ρ (kg/m^3)	2650	2120	1000	1
K (GPa)	36	7	2.2	$1 \cdot 10^{-4}$
μ (GPa)	?	4.2	-	-
Φ (%)	-	20	-	-
κ (mD)	-	600	-	-
η (Pa.s)	-	-	$1 \cdot 10^{-3}$	$2 \cdot 10^{-5}$

et al. (2010) for an extensive review). While equation (1.39) results from dimensional analysis, WIFF is governed at fine scale by the Biot's theory of poroelasticity (Biot, 1956, 1962). Though analytical solutions exist for the corresponding equations in different simple configurations (e.g. White, 1975), the most popular approach in the last years seems to be the numerical resolution via finite-elements, as described in Quintal et al. (2011b). A *Representative Elementary Volume* (REV) containing one fluid patch and some portion of the surrounding medium is used as a sample. An example is shown in Fig. 1.29. A step load is applied, for example, on the top of the sample. Then a simplified static version of the Biot's equations (consolidation equations, Biot, 1941) is resolved, as the dynamic terms are assumed to be negligible at seismic frequencies. Time-dependent stress σ_{ij} and strain ϵ_{ij} are calculated during the simulation. Their time-derivatives are averaged over the simulated sample and Fourier-transformed. Knowing these two frequency-dependent values, the effective frequency-dependent complex P-wave modulus $H(\omega)$ and shear modulus $\mu(\omega)$ are obtained via the Biot's stress-strain relation (Biot, 1962) (see appendix D in Quintal et al. (2011b) for details). It turns out that the real parts of these moduli exhibit a characteristic dispersive behaviour similar to the one shown in Fig 1.23, around the characteristic frequency of WIFF mechanism. Then the frequency-dependent quality factors are obtained according to equation (1.27):

$$\begin{aligned} Q_P &= \frac{\mathfrak{Re}(H)}{\mathfrak{Im}(H)} \ , \\ Q_S &= \frac{\mathfrak{Re}(\mu)}{\mathfrak{Im}(\mu)} \ . \end{aligned} \tag{1.40}$$

This is the intrinsic quality factor, i.e. not the one directly measuring the energy loss per cycle. Kuteynikova et al. (2014) use this numerical setup in 3D in a homogeneous background model with a characteristic water patch size of $L_c = 18$ cm. For the parameter values listed in Table 1.3, they obtain $f_0 \sim 30$ Hz, with $Q_P \sim 25$ in the partially saturated medium, instead of $Q_P > 100$ in the dry medium. In comparison, equation (1.39) yields $f_0 = 40$ Hz, which is consistent. This prediction, however corresponds to the case of an abrupt boundary between saturated and dry zones. An important finding of Kuteynikova et al. (2014) is that when the saturation gradually decreases across the patch boundary, the characteristic frequency shifts towards higher

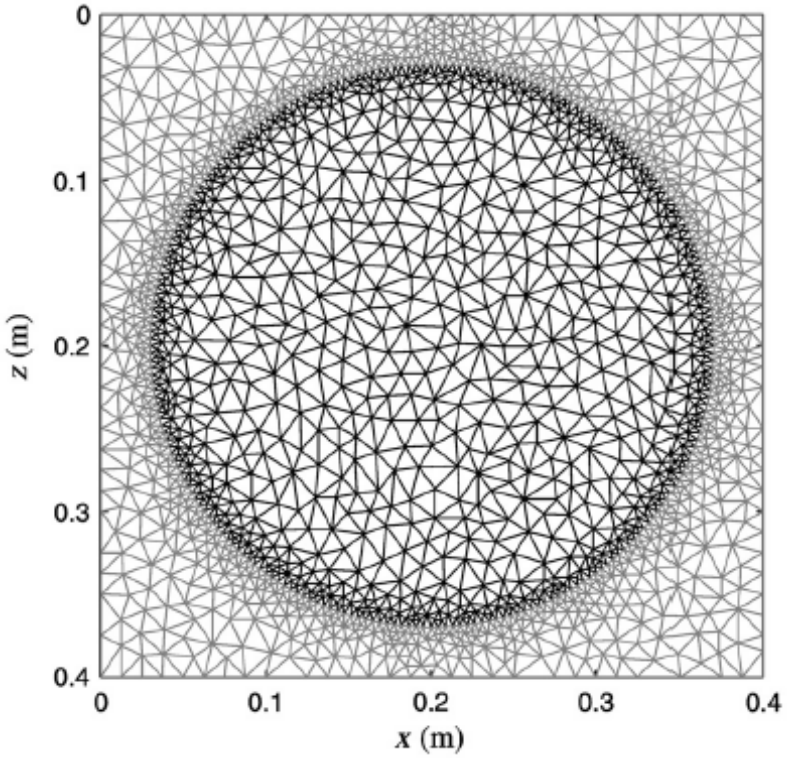


Figure 1.29: Mesh of the Representative Elementary Volume (REV), similar to the mesh used by Quintal et al. (2012) and Lambert et al. (2013) for the numerical resolution of the Biot's equations.

frequencies by one order of magnitude. That is, attenuation due to patchy saturation can happen at arbitrary low frequencies depending on the size of the patches, but it also strongly depends on the saturation distribution across the patch boundaries.

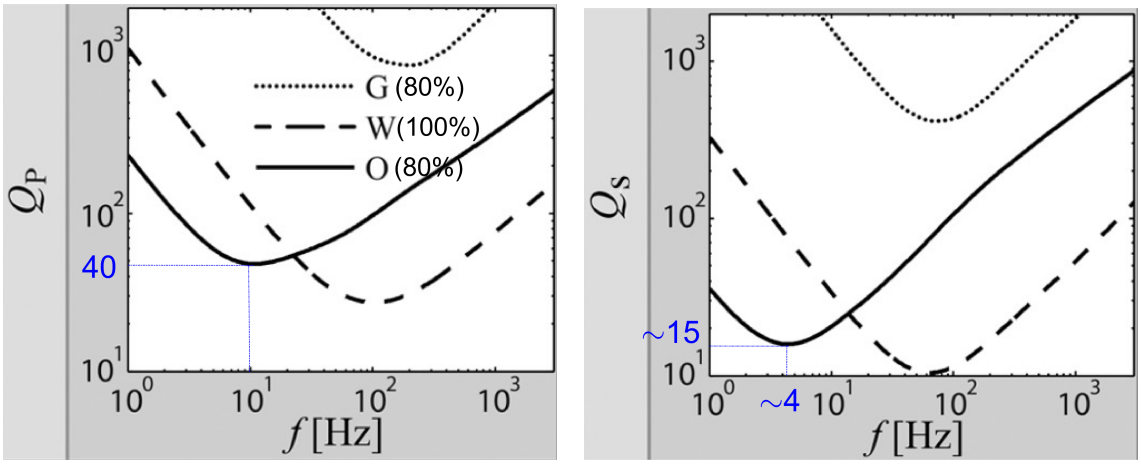
1.6.4 Attenuation due to Wave-Induced Fluid Flow in heterogeneous media

WIFF behaves differently in case of a heterogeneous matrix containing regions of different stiffness and permeability. Such a hetrogeneous medium enforces patchy saturation, as the low-permeability regions are preferentially saturated by the wetting phase (usually water). As explained in Müller et al. (2010), the presence of heterogeneities facilitates WIFF, because fluids are easily ejected from the compliant (soft) regions during compression cycles of the wave. Quintal (2012) numerically model the effect of such heterogeneities on the wave attenuation using the same approach as explained in the previous section. The medium is assumed to contain small spherical inclusions of 16 cm radius with lower permeabilities and porosities, and a different stiffness. The low-permeability inclusions are fully water-saturated, while the background medium contains a mixture of water and hydrocarbons (80% saturation with either gas or oil). The authors obtain f_0 as low as 10 Hz for P-waves (Fig. 1.30a, solid curve). S-waves behave differently with respect to WIFF compared to P-waves. In order to exhibit a strong wave-induced attenuation, S-waves need a permeability barrier around the matrix heterogeneities (undrained conditions on the boundaries of the REV in Fig. 1.29).

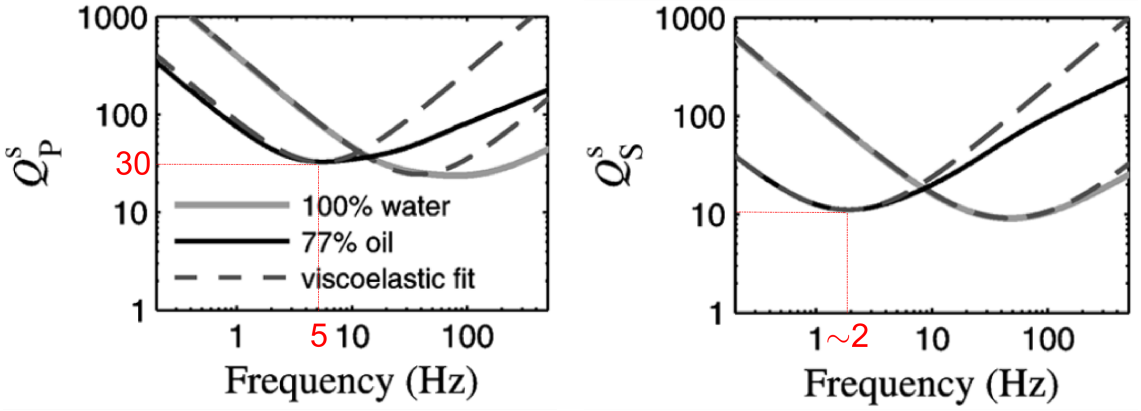
The WIFF theory was used by Lambert et al. (2013) as the viscoeslatic mechanism in the reservoir allowing to reproduce one of the charcteristic HMT attributes (V/H ratio increase above the reservoir). In Figs. 1.30a and b, I compare the results, in terms of frequency-dependent Q_P and Q_S quality factors, obtained by Quintal et al. (2012) and Lambert et al. (2013) with their respective matrix/heterogeneity parameters listed in Table 1.4. Undrained boundary conditions were applied around the REV in both papers. The contrasts of seismic velocities between the dry background medium ($V_p^{\mathrm{bg}} = 5500$ m/s, $V_s^{\mathrm{bg}} = 3500$ m/s) and the dry heterogeneity ($V_p^{\mathrm{het}} = 1600$ m/s, $V_s^{\mathrm{het}} = 940$ m/s) are already very high in Quintal et al. (2012).

It is clear from Table 1.4 that Lambert et al. (2013) systematically enhanced the contrast between the background and the heterogeneity (both elastic moduli and permeability), thus increasing WIFF effects. The seismic velocities in the heterogeneity are $V_p^{\mathrm{het}} = 1290$ m/s, $V_s^{\mathrm{het}} = 665$ m/s for Lambert et al. (2013), which can be considered as unusually low. The oil viscosity is increased by a factor of 2.5 compared to Quintal et al. (2012), which tends to decrease the characteristic frequency f_0. This explains why, in Fig. 1.30, the characteristic frequencies obtained by Lambert et al. (2013) are divided by two compared to Quintal et al. (2012), while the minimum Q_P and Q_S are divided by 1.5 and 2.5, respectively (compare blue values in Figs 1.30(a) and (c)). That is, one should be concious that the high values of the HMT attribute V/H obtained by Lambert et al. (2013) and discussed in Section 1.2.3 rely not only on the presence of incident body waves, but also on a reservoir model which is very favourable to the development of the WIFF phenomenon. While such a model is probably not unrealistic for some particular cases, I do not think it can be considered

(a) Quintal et al. (2012)



(b) Lambert et al. (2013)



(c) Correspondence between definitions of Q

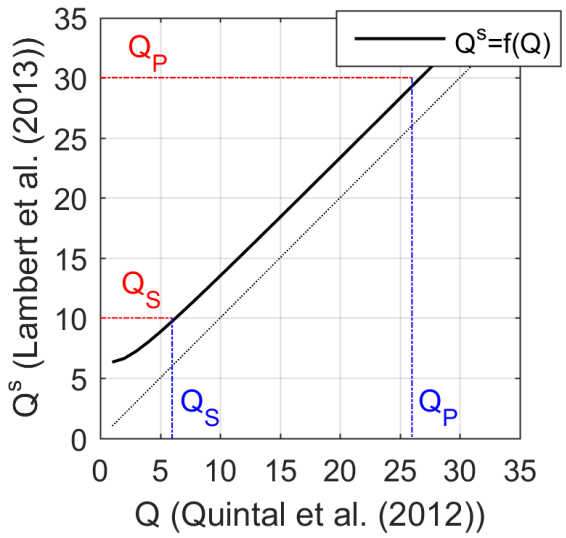


Figure 1.30: Frequency-dependent quality factors obtained by (a) Quintal et al. (2012) and (b) Lambert et al. (2013) with the parameters listed in Table 1.4. Since both papers do not use the same definition of Q, the values from Lambert et al. (2013) (77% oil) are brought to the scale of Quintal et al. (2012) in (c), using the relation $Q^{(s)} = f(Q)$ (equation 1.29).

Table 1.4: Matrix and heterogeneity parameters used by Quintal et al. (2012) (case A in their table 3) and Lambert et al. (2013) (their table 1). Different parameters are highlighted in blue for Quintal et al. (2012) and in red for Lambert et al. (2013). Subscripts gr, dry and fl refer respectively to the properties of the solid grains, dry matrix and fluids.

Region:	Background	Heterogeneity	Water	Oil	Gas
ρ_{gr} (kg/m^3)	2700	2700	-	-	-
K_{gr} (GPa)	40	48	-	-	-
Φ (%)	6	26	-	-	-
Quintal et al. (2012), case A					
κ (mD)	1000	40	-	-	-
K_{dry} (GPa)	36	4	-	-	-
μ_{dry} (GPa)	32	2	-	-	-
K_{fl} (GPa)	-	-	2.4	1.4	0.04
ρ_{fl} (kg/m^3)	-	-	1010	880	160
η (Pa.s)	-	-	0.001	0.02	$2 \cdot 10^{-5}$
Lambert et al. (2013)					
κ (mD)	1000	20	-	-	-
K_{dry} (GPa)	36	3	-	-	-
μ_{dry} (GPa)	32	1	-	-	-
K_{fl} (GPa)	-	-	2.4	1.5	-
ρ_{fl} (kg/m^3)	-	-	1010	880	-
η (Pa.s)	-	-	0.001	0.05	-

as representative of most of the hydrocarbon reservoirs where LFPS surveys were carried on. In Fig. 1.30a, we also observe that a water-gas system exhibits almost no attenuation, which is again problematic with respect to HMT interpretation, because similar seismic signatures are observed for oil and gas reservoirs in the real data (see Section 1.2.2).

As mentionned in Section 1.2.3, Tisato et al. (2015) discovered an additional low-frequency relaxation mechanism acting on microscopic gas bubbles, i.e. at small values of gas saturation: the *Wave-Induced Gas Exsolution-Dissolution* (WIGED). With characteristic frequencies of the order of several Hz, this mechanism does not require matrix heterogeneities and is solely due to gas exsolution/dissolution during decompression/compression wave cycles. According to the experimental results on Berea sandstone published by Tisato et al. (2015), $Q_P \approx 25$ can be obtained from this mechanism, which is comparable the value of $Q_P \approx 30$ used by Lambert et al. (2013) based on WIFF. However, no effect on S-waves is expected from WIGED. For S-wave attenuation, Quintal et al. (2014) numerically studied the effect of the fault connectivity within the reservoir, using the same method as Quintal et al. (2012). Fractures were considered as very porous, permeable and soft elliptic heterogeneities ($V_p \approx 200$ m/s and $V_s \approx 100$ m/s within the fracture). Only water-saturation was considered. In case of unconnected faults with a mutual 45° orientation, they obtained $Q_S < 10$ peaking around 1 Hz. However, this remains a very particular situation, as the peak frequency was found around 100 Hz when the faults were connected, and Q_S reached 20 when unconnected faults had a perpendicular orientation.

1.6.5 Conclusions

This section can be summarised as following. I have first presented the approach used to estimate the elastic properties (ρ, V_p, V_s) of the partially-saturated gas reservoir modelled in this work, based on the uniform saturation assumption and the Gassmann's equation (Vidal, 2002). I have also discussed how the anelasticity could be handled (predicting Q_P, Q_S and velocity dispersion). Pure elasticity, indeed, only partially describes the wave propagation. At seismic frequencies, attenuation is also expected to play a role, as field experiments report Q_P values as low as 5 sometimes observed in gas-saturated rocks, while Q_S is not reported very sensitive with respect to gas saturation (Klimentos, 1995; Walls et al., 2005). In a perspective of modelling these effects, I have briefly described a specific mechanism (WIFF) previously invoked as possibly responsible for fluid-specific attenuation at the HMT frequencies (several Hz, Lambert et al., 2013). This mechanism is able to generate fluid-specific frequency-dependant attenuation and dispersion. However, based on the analysis of the simulation parameters used in the literature, my conclusion is that WIFF should not be invoked as «directly» interpreting HMT, for the following two reasons:

1. WIFF exhibits strong dependence on the reservoir rock properties (heterogeneity, fracture connectivity, permeability, porosity, stiffness, ...).

2. WIFF usually does not yield the same effect (characteristic frequency, attenuation value) for partial saturation with oil and gas.

It appears unlikely that the anomalous attributes peaking at low frequencies, such as those presented in Fig. 1.4, could be a direct manifestation of a frequency-dependent attenuation peaking at the same frequencies. Nevertheless, they could justify the existence of a larger contrast between the reservoir and the surrounding medium, compared to the purely elastic case. In that sense, upscaling strategies for integrating attenuation mechanisms in the wave propagation modelling, such as the one proposed by Lambert et al. (2013), are extremely useful, and could be adapted to particular contexts, yielding more realistic results than the pure elastic modelling performed in my work.

In this work, *I focus on the elastic scattering of Rayleigh waves, as a crucial mechanism to be understood for interpreting HMT*, possibly responsible for the particular frequency-dependant pattern of the amplitude anomalies. In this context, adding visco-elasticity (which I do not do) would increase the contrast of the reservoir. Note that body wave scattering was the dominant phenomenon in Lambert et al. (2013). However, surface waves are mainly sensitive to V_s. If WIFF is invoked as the attenuation mechanism, S-wave attenuation and dispersion are obtained only in very specific conditions (strong matrix heterogeneity, permeability barriers, unconnected fractures), and cannot be considered as universally expected features. That is, my preliminary conclusion is that WIFF could have only a limited effect on Rayleigh wave scattering by the reservoir, in the general case.

For typical Underground Gas Storages (UGS) operated by Storengy, the reservoir rock is usually a fairly homogeneous sandstone, and the presence of strong heterogeneities proposed by Lambert et al. (2013) is unlikely, so WIGED (rather than WIFF) appears as the most relevant attenuation mechanism, since it does not require matrix heterogeneity. In the geothermal context (boiling water within a fractured matrix), I refer to Grab et al. (2017a), who modelled effective viscoelastic properties considering two mechanisms: WIFF on fractures, and an additional dissipation mechanism due to the phase change (latent heat dissipation in the liquid water/steam two-phase system). I suggest these two approaches (i.e. Tisato et al. (2015) and Grab et al. (2017a)) as guidelines for future modelling works accounting for visco-elasticity, respectively in UGS and geothermal contexts.

1.7 Contributions and outline

In this section, I briefly describe the main activities I was involved in as part of my PhD work, as well as I present the outline of the manuscript along with the main contributions of my work.

1.7.1 PhD work organisation

Based on the results obtained during my master work and in the beginning of the present PhD work, a new methodology of processing the recorded spectra in order to extract and map the low-frequency anomalies was patented by Storengy (Huguet et al., 2017). It is based on automated spectra classification (see Section 2.1.7). The major aim of the present work was to understand the physical link between the

recorded anomalies and the presence of gas in the subsurface, *and more specifically to investigate the possibility of an elastic mechanism.* Additionally to modelling, field acquisitions were an important part of my work. For this purpose, Storengy purchased a pool of 6 broadband seismometers (Nanometrics Trillium Compact 20s), which were used to instrument the Saint-Illiers UGS, France, with a permanent array (Section 2.2). I was in charge of installing and maintaining the array. I was helped by master interns Marc Peruzzetto (2017) and Simon Lejart (2018), as well as by Jérôme Vergne and Damian Kula (Ecole et Observatoire des Sciences de la Terre) starting from April 2018, to whom I am grateful. Another acquisition (see Section 2.4) took place in June 2016 on a geothermal field operated by Supreme Energy in Muara Laboh (Western Sumatra, Indonesia), organised in collaboration with IRD (Institut pour la Recherche et le Développement), and in particular with Jean-Philippe Métaxian. The field works lasted for about 3 weeks, preceded and followed by permanent station maintenance carried out by IRD. I was involved in the experiment preparation, the acquisition design, the field acquisition itself, as well as I was in charge of the final reporting for the Indonesian client. In this manuscript, for sake of brevity, the complete report of this experiment is replaced by the conference paper based on its main results. All together, the experimental part of my PhD work represented about one day per week over the three years.

On the other hand, three master internships related to my topic, which I supervised, were hosted by Storengy. The second of these internships (Marc Peruzzetto in 2017), dedicated to the ambient wavefield characterisation, resulted in the article published in GJI in July 2018, and which forms the Chapter 3 of this manuscript. The work on this topic represented about a year, as I started implementing the ambient noise cross-correlation before the beginning of Marc's internship, and continued working on several aspects after its end (namely the automated signal subspace determination and the numerical modelling to reproduce the L_g phase excitation at continental margins). The internship of Simon Lejart in 2018 dealed with the cross-correlations on the radial and transverse components, and the quantification of the coherent part of the total wavefield energy (see Section 7.5 for the motivation of this analysis). While interesting results were obtained, they could not be included in this manuscript because the internship ended shortly before the submission date.

1.7.2 Manuscript outline

The manuscript starts with the presentation of the experimental results, which exhibit spectral signatures similar to the hydrocarbon microtremors described in the literature. Three aspects are relatively new:

- A clustering procedure applied to classify the entire spectral shaped, and not the attributes (Section 2.1).
- The long-time evolution of the spectral anomalies is analysed in Section 2.2.
- An V/H anomaly is described above a steam-bearing geothermal reservoir (Section 2.4), while most of the previous applications focused on hydrocarbon reservoirs.

In Chapter 3, the ambient noise composition in the Chémery dataset is analysed. The most interesting aspects of this part are

- The detection of the L_g phase above 2 Hz and the numerical simulation of its excitation at a continental margin. The presence of this wave justifies the presence of many Rayleigh wave overtones in the subsequent modelling part.

- The application of array techniques to a cross-correlation common offset gather, which allows a better dispersion curve reconstruction compared to the direct application of the array techniques.

In Chapter 4, I numerically simulate the interaction of a multi-modal random Rayleigh wavefield with an elastic reservoir and with a structural anticline, and compare both effects. This is new in the context of hydrocarbon microtremor studies, as Lambert et al. (2013) modelled a reservoir embedded in a laterally invariant medium. This also applies to the microseismic sounding method (MSM), where simulations are usually performed in a homogeneous half-space (Gorbatikov et al., 2008; Tsukanov & Gorbatnikov, 2018). Moreover, the source term consisting of a multimodal Rayleigh wavefield is a new hypothesis. The Born approximation is introduced in this chapter. The sensitivity kernels (Fréchet derivatives) are modelled based on SPECFEM2D displacement field outputs. The wavefield gradients necessary for the Fréchet derivative computation are implemented in Matlab using a simple finite-difference differentiation.

In Chapter 5, I study the Rayleigh wave scattering in a rather simple context (elastic half-plane). To my knowledge, this is the first time when the radiation patterns for Rayleigh wave scattering are derived in the vertical $x - z$ plane, as Rayleigh wave scattering is most often studied in the horizontal 2D view. The results show that different elastic parameters have different radiation patterns, which opens the perspective of a multi-parameter inversion of the Rayleigh wave amplitude perturbations in the context of near-receiver scattering.

Chapter 6 presents examples of Rayleigh wave amplitude perturbations at the surface inverted for the scatterer position, shape, and elastic properties. Completely discarding the phase information in the Rayleigh wave analysis is essentially the idea of the *Microseismic Sounding Method* (MSM). The novelty of my work is to introduce a physical inversion method, contrary to the semi-empirical depth-frequency coefficient K used in MSM to link the depth of the scatterer to the frequency of the anomaly. The inversion is performed by implementing the conjugate gradient algorithm in Matlab and using the full Fréchet derivatives computed in Chapter 4 to form the Hessian matrix (Gaussian-Newton formulation). The optimal parametrisation choice is studied using the available full Hessian. This is an attempt to develop an imaging method based on HMT attributes (here, PSD perturbations), in order to go beyond an attribute map constructed at the surface.

However, at the end of Chapter 4, it becomes clear that the reservoir-related perturbations obtained with elastic 2D modelling are too weak compared to the spectral anomalies recorded in the field, so that the inversion procedure in Chapter 6 cannot

be expected to apply to real data with the current state of the forward problem. The final Chapter 7 discusses how the forward modelling part could be improved, and which could be the alternatives to scattering modelling in order to explain the HMT spectral signatures, which I was not able to achieve in this work.

1.7.3 Main contributions

Software

For internal real-data applications in Storengy, all of the developments presented in Chapters 2 and 3 have been integrated into a Matlab toolbox which is flexible to work with different datasets. Following functionalities are supported:

- Map-view interactive data and time period selection
- Visual interactive quality control of PSD and V/H spectra
- Computation and mapping of integral attributes (see Chapter 2)
- Automated spectra classification (see Chapter 2)
- Ambient-noise cross-correlation including pre-processing (see Chapter 3)
- Cross-correlation beam-forming (see Chapter 3)
- Array processing via FK, HRFK and MUSIC methods (see Chapter 3), mostly developed by Marc Peruzzetto.

For the modelling part, another Matlab toolbox was developed, with the following functionalities:

- Creating robust mesh files for SPECFEM2D simulations for layered models with specified structural deformations
- Creating SPECFEM2D parameter, source, receiver files, for forward and adjoint simulations
- Reading SPECFEM2D outputs with automatic determination of receiver positions, time-step and velocity model used for simulation
- Born sensitivity kernels computation based on SPECFEM2D outputs
- Linear-Conjugate gradient inversion of amplitude/waveform perturbations based on the sensitivity kernels
- Radiation pattern visualisation based on the sensitivity kernels

As several aspects of this works were published or are currently in the publishing process (list provided below), the manuscript is essentially a compiliation of self-standing papers. This explains some redundancy in several aspects of Chapters 4, 5 and 6 (namely the composition of the ambient noise, the justification of working with Rayleigh wave scattering in the vertical plane, the basic equations for Born approximation, etc).

Patent

Section 2.1.7: Huguet, F., Kazantsev, A., Lavergne, D., & Egermann, P., 2017. Procédé de caractérisation du sous-sol d'une region utilisant des signaux sismiques passifs, et système correspondant, Patent WO2017220918; Assignee: Storengy

Published works

Chapter 4: Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2019. Rayleigh wave amplitude distortions above a reservoir: new insights from elastic modelling, *Geophysical Journal International*, **accepted 04/02/2019**

Chapter 3: Peruzzetto, M., Kazantsev, A., Luu, K., Métaxian, J.-P., Huguet, F., & Chauris, H., 2018. Broad-band ambient noise characterization by joint use of cross-correlation and MUSIC algorithm, *Geophysical Journal International*, **215**(2), 760–779

Appendix 3.E: Kazantsev, A., Peruzzetto, M., Chauris, H., Dublanchet, P., & Huguet, F., 2018b. Origins Of Rayleigh Wave Overtones In Ambient Noise, in *Seventh EAGE Workshop on Passive Seismic 2018*

Section 2.4: Kazantsev, A., Egermann, P., Ramadhan, I., Huguet, F., Formento, C., Peruzzetto, M., Chauris, H., & Métaxian, J.-P., 2017b. An innovative methodology based on Low Frequency Passive Seismic data analysis to map geothermal reservoir steam saturated areas, in *5th International Indonesia Geothermal Conference and Exhibition (IIGCE)*, Jakarta, Indonesia

Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2017a. Near-field Elastic Scattering of Rayleigh Waves - A Model for Interpreting Hydrocarbon Microtremors, in *79th EAGE Conference and Exhibition 2017*

Submitted/in preparation

Chapter 5: Kazantsev, A. & Chauris, H., 2018. Radiation patterns for 2d rayleigh-wave scattering in the vertical plane: impact of the free-surface, *Geophysical Prospecting*, **submitted 29/07/2018**

Chapter 6: Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2018a. Spectral amplitude imaging with Rayleigh waves, *GJI*, **in preparation**.

Chapter 2

Experimental observations of amplitude distortions

Résumé (français)

Dans ce chapitre, on présente les résultats de l'analyse des attributs spectraux calculés sur trois jeux de données enregistrés à l'échelle du réservoir, et un jeu de données de petite échelle. On met en évidence l'apparente corrélation entre certains attributs spectraux et la présence de gaz dans le sous-sol.

Les données présentées dans les parties 2.2 à 2.4 ont été recueillies dans le cadre de ce travail de thèse. Les données de la partie 2.1 ont été enregistrées par Spectraseis dans le cadre du consortium *Low Frequency Seismic Partnership* (*LFSP*, voir partie 1.1). Dans les parties 2.1 et 2.2, on analyse les données enregistrées au-dessus de deux sites de stockage de gaz souterrain de Storengy en France. Dans la partie 2.3, on conduit une expérience à échelle réduite qui met en évidence les effets de la substitution de fluide dans un réservoir sur la diffraction des ondes de Rayleigh. Enfin, dans la partie 2.4, on analyse les attributs spectraux dans un jeu de données enregistré en Indonésie dans un contexte d'exploration géothermique. Pour chacun des réservoirs analysés, on présente le contexte géologique et les principales caractéristiques, ainsi que les anomalies spectrales observées. On observe à la fois des anomalies du spectre sur la composante verticale (PSD_z) et sur le ratio V/H. Cependant, elles sont différentes entre les réservoirs de gaz naturel (amplification) et le réservoir de vapeur (atténuation). On se restreint par la suite au cadre des réservoirs de gaz naturel, où on identifie l'amplification de PSD_z comme l'effet le plus clairement corrélé à la présence de gaz. Le travail de modélisation dans la suite du manuscrit porte uniquement sur cette amplification.

Summary (English)

In this chapter, we present results of «standard» spectral analysis performed on three reservoir-scale sites, and one small-scale field experiment. In Sections 2.1 and 2.2, we analyse data recorded above two underground gas storage facilities (UGS) operated by Storengy in France. Their respective locations are shown in Fig. 2.1. In Section 2.3, we present a simple small-scale experiment which demonstrates fluid substitution effects on Rayleigh wave scattering. Finally, in Section 2.4, we present a dataset recorded in Indonesia in the context of geothermal exploration. The data presented in Sections 2.2 to 2.4 were recorded as part of the present PhD work. The data presented in Section 2.1 was recorded in 2010 by Spectraseis, as part of the *Low Frequency Seismic Partnership* (see Section 1.1). For each reservoir, we present the geological context and the main characteristics, as well as the evidences for spectral anomalies. Both PSD_z-based and V/H ratio-based attributes exhibit some degree of correlation with the gas presence (either natural gas or steam, though the signatures are not the same). Nevertheless, restricting ourselves to the UGS context, we identify the PSD_z amplification as the most conclusive effect, and focus on its modelling in the remainder of the manuscript.

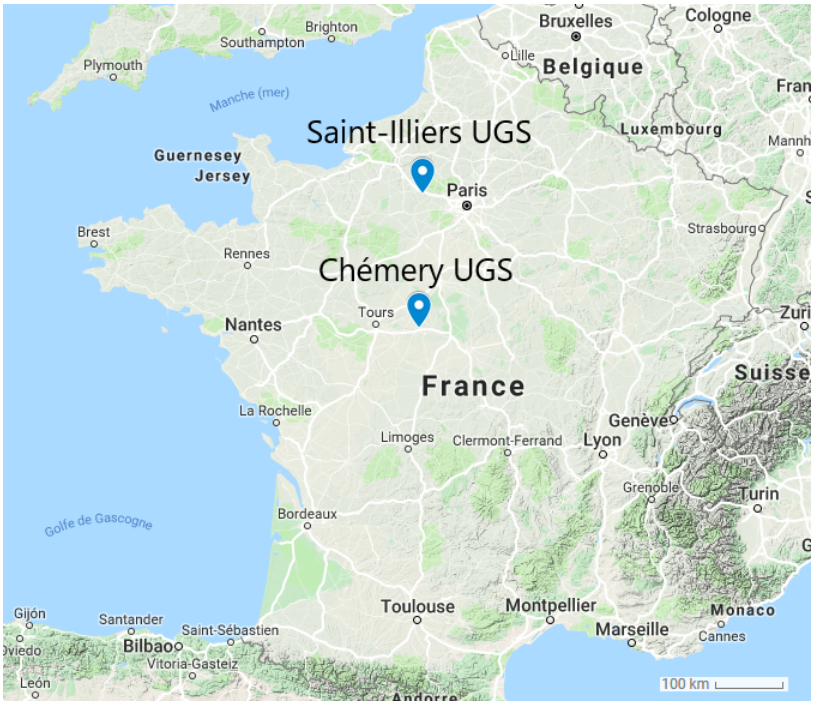


Figure 2.1: Location of the Chémery and Saint-Illiers underground gas storage facilities.

2.1 Dataset 1: Chémery underground gas storage

2.1.1 Site presentation

The Chémery UGS has been operated by Storengy since 1968. It is located at the periphery of the Paris sedimentary Basin. The oldest main-sequence sedimentary deposits at the bottom of the basin date back to the Triassic. The maximum sedimentary thickness in the center of the basin (roughly beneath Paris) is about 3000 m, (Fig. 2.2). However, older Permian and Carboniferous basins are present beneath the Triassic deposits in some areas of the Paris Basin (Perrodon & Zabek, 1990). In these areas, the thickness of the sedimentary cover is larger than the conventional Paris Basin thickness. Chémery is located in such area. While the Paris Basin thickness in this peripheral location should not exceed 1500 m (see Fig. 2.2), the deepest well in Chémery (CS01) was drilled down to about 2600 m and did not reach the crystalline basement. The presence of a Permian basin beneath Chémery is reported in Mégnien (1980) down to 3000 m. However, it is unclear whether this Permian basin lies on a crystalline bedrock or on a Carboniferous deposit. The question of the bedrock depth is important for Rayleigh wave propagation analysis and will be further discussed in the context of dispersion curves inversion in Chapter 3.

The reservoir itself consists of Permo-Triassic sandstones located at a depth of 1100-1180 m with respect to the surface (or 960-1040 m with respect to the sea level). The total thickness of the reservoir layers is between 70 and 85 m, with a porosity of 15-20% and a permeability of several hundreds mD. The cap rock consists of Levallois clays (variable thickness around 10 m), overlaid with another 175 m of alternating

layers of marls and clays. A typical geological log for the Chémery site is shown in Fig. 2.3.

The geological trap, used for gas storage, is formed by an anticline structure. The latter is due to successive tectonic deformations during the Alpine and Pyrenean orogenies. The topography of the upper reservoir boundary is shown in Fig. 2.4a. The elevated zone due to the anticline is well visible in the center of the map (yellow to red colours). The anticline structure is bounded by a structural fault, of EW orientation in the northern part, and NE-SW orientation in the western part. The anticline is also affected by a system of NS faults in its south-western part.

2.1.2 Gas accumulation evolution

Natural gas (methane) is injected and withdrawn from the reservoir in annual cycles. Gas is usually injected during the warmest part of the year (from Spring to Autumn), when the gas market demand is low, and withdrawn during the coldest period, when the demand is high. Pressurised gas replaces water in the porous reservoir rock during the injection phase, and the aquifer helps pushing gas back to the reservoir top during the withdrawal phase. After a transient period covering the few first cycles of the UGS, when gas is injected into a virgin aquifer, the system reaches a stable regime of pressure quasi-equilibrium between the gas bubble and the surrounding aquifer. In this permanent regime, about half the total reservoir capacity is cycled each year (e.g. blue curve in Fig. 2.5). At the end of the UGS life cycle, about 20-30% of the injected volume will not be recovered, since it would require unreasonable pumping power.

Importantly, *the volume occupied by the gas inside the reservoir is not equal to the injected volume* because of a strong gas compression in the reservoir to compensate the geostatic pressure gradient. For deeper reservoirs the compression factor is higher. As the gas volume decreases with compression, injecting new gas does not trigger an immediate expansion of the gas bubble in the reservoir. The red curve in Fig. 2.5 shows the simulated evolution of the bottomhole bubble volume for the Chémery site during three cycles (injected volumes shown in blue). The simulation is performed by the «MULTI» model, a finite-difference 3D solver of gas-water flow equations in a porous medium, developed internally in Storengy. A description of this tool can be found in Sonier et al. (1993). The maximum bubble volume is typically reached 1-2 months after the peak stock level. This time shift occurs because the highly compressible gas is immediately pressurised when a new injection starts, while the bubble expansion happens on a longer time scale through pressure diffusion in the surrounding aquifer.

Seismic monitoring of such a system is challenging, since the in-situ gas volume difference between high and low stock levels is not so large as most of the gas injection is accommodated through pressure variations. From Fig. 2.5, the estimated relative volume change between the two time-lapse acquisitions in Chémery (black vertical lines) is of only 5%. In Fig. 2.6, we show the simulated maps of gas accumulation,

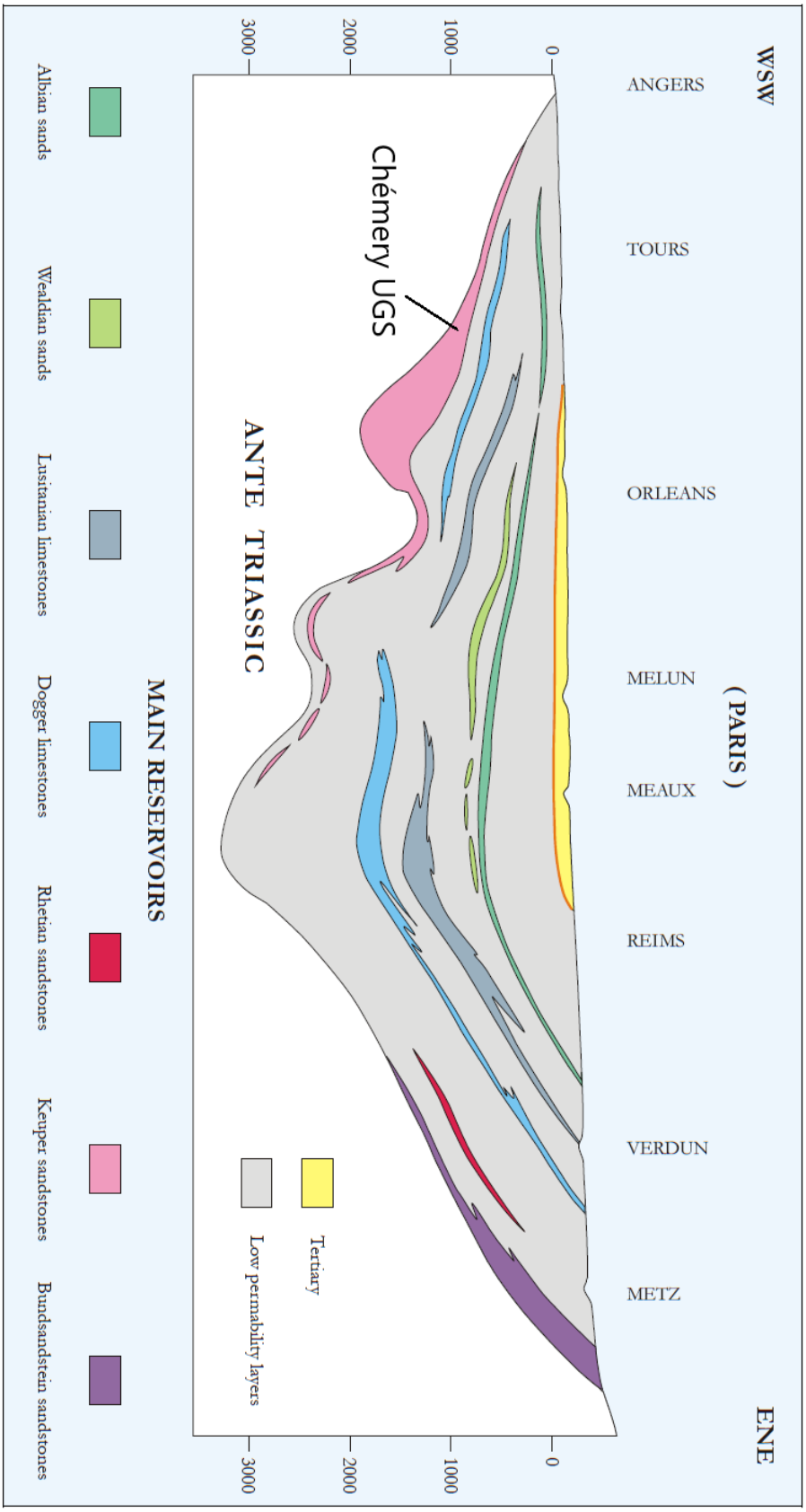


Figure 2.2: Schematic representation of the main reservoir deposits in the Paris Basin along a WSW-ENE profile. The approximate location of the Chémery UGS is indicated. From Bonijoly et al. (2003).

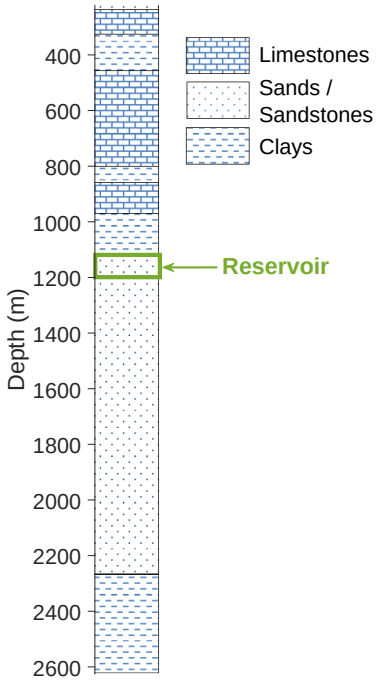


Figure 2.3: Simplified geological log of the Chémery site.

measured as the effective gas column thickness

$$H_g = \sum_i s_i \Phi_i h_i \ , \tag{2.1}$$

with s_i, Φ_i and h_i respectively the saturation, the porosity and the thickness of the different reservoir layers. Figs 2.6a and b show gas accumulations respectively at low and high stock levels, corresponding to the two passive seismic acquisitions in 2010. No major difference is visible between the two maps, consistently with the very small relative volume change observed in Fig. 2.5. The difference between the two maps is shown in Fig. 2.6c, where a small positive evolution (1-2 m) is observed at the periphery of the gas bubble, while small negative changes are also visible at somewhat randomly distributed locations.

2.1.3 Seismic network

About 100 broadband Trillium 40 s seismometers (Nanometrics) were deployed in April 2010 and November 2010 with various geometries and a 100 Hz sampling frequency. All of the surveyed locations studied in this work are shown with black dots in Fig. 2.4. All the signals were not recorded at the same time, as the sensors were moved every day. While small aperture arrays were also deployed during the survey, we only analyse dense rectangular configurations and linear profiles. Rectangular configurations were acquired from 20 to 23 April (4 days), and from 4 to 11 November (8 days). Most of the sensors were operational between 3 PM and 6 AM local time (UTC+2 in April and UTC+1 in November), and were shifted by about 250 m

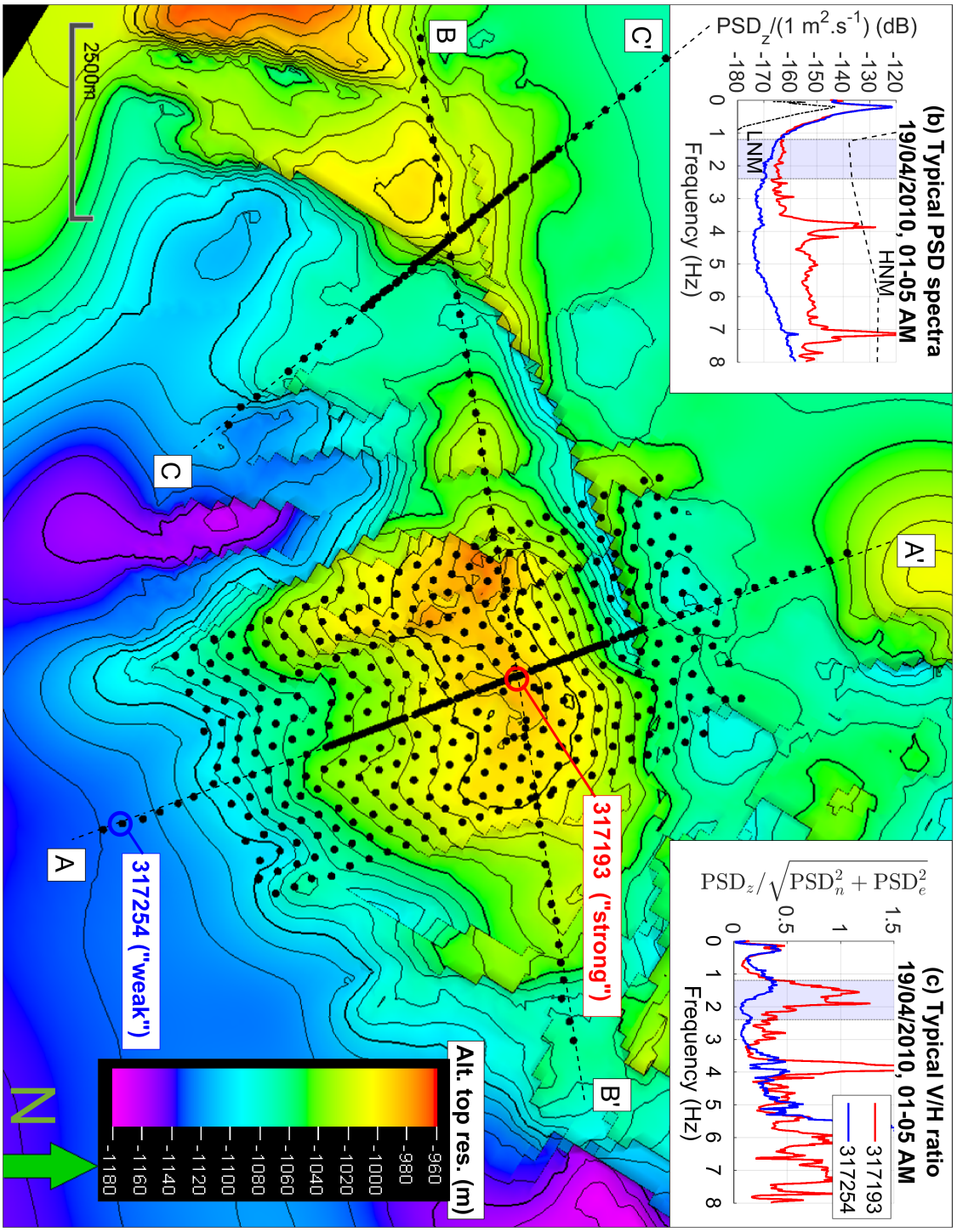


Figure 2.4: (a) Seismic network superposed on the structural map of the reservoir top layer. Each black dot represents a temporary recording location. All the locations were not recorded simultaneously. 2D profiles are spotted as AA', BB' and CC'. (b) Typical «low HMT»(blue) and «high HMT»(red) PSD_z spectra, recorded simultaneously for the two stations spotted on the map. Dotted lines show the low and the high noise models from Peterson (1993). The frequency band used for the attribute $PSD_z|_{1.2}^{2.4}$ (equation 1.1) is highlighted in blue. (c) V/H spectral ratio for the same stations.

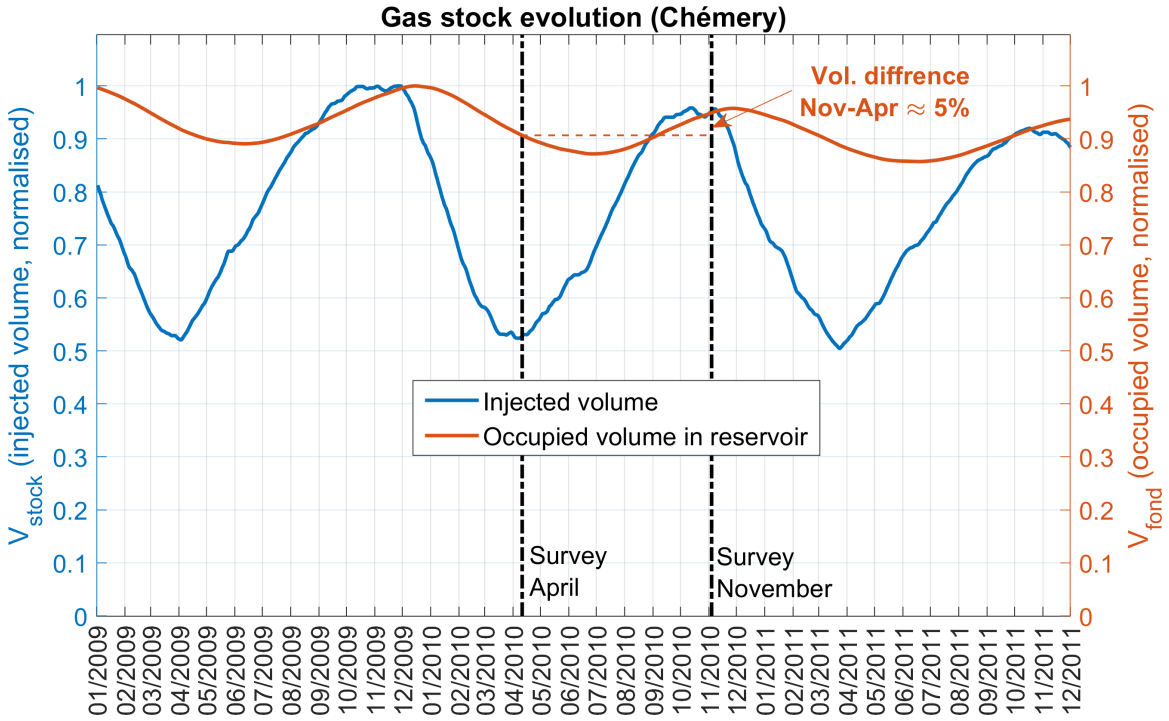


Figure 2.5: Evolution over 3 years of the injected gas volume in surface conditions (V_{stock}, blue lines) and of the simulated volume occupied by the gas in reservoir conditions (V_{fond}, red line) between 2009 and 2012. The time lapse passive seismic surveys are spotted by the black vertical lines. The corresponding V_{fond} difference in shown in red (5%).

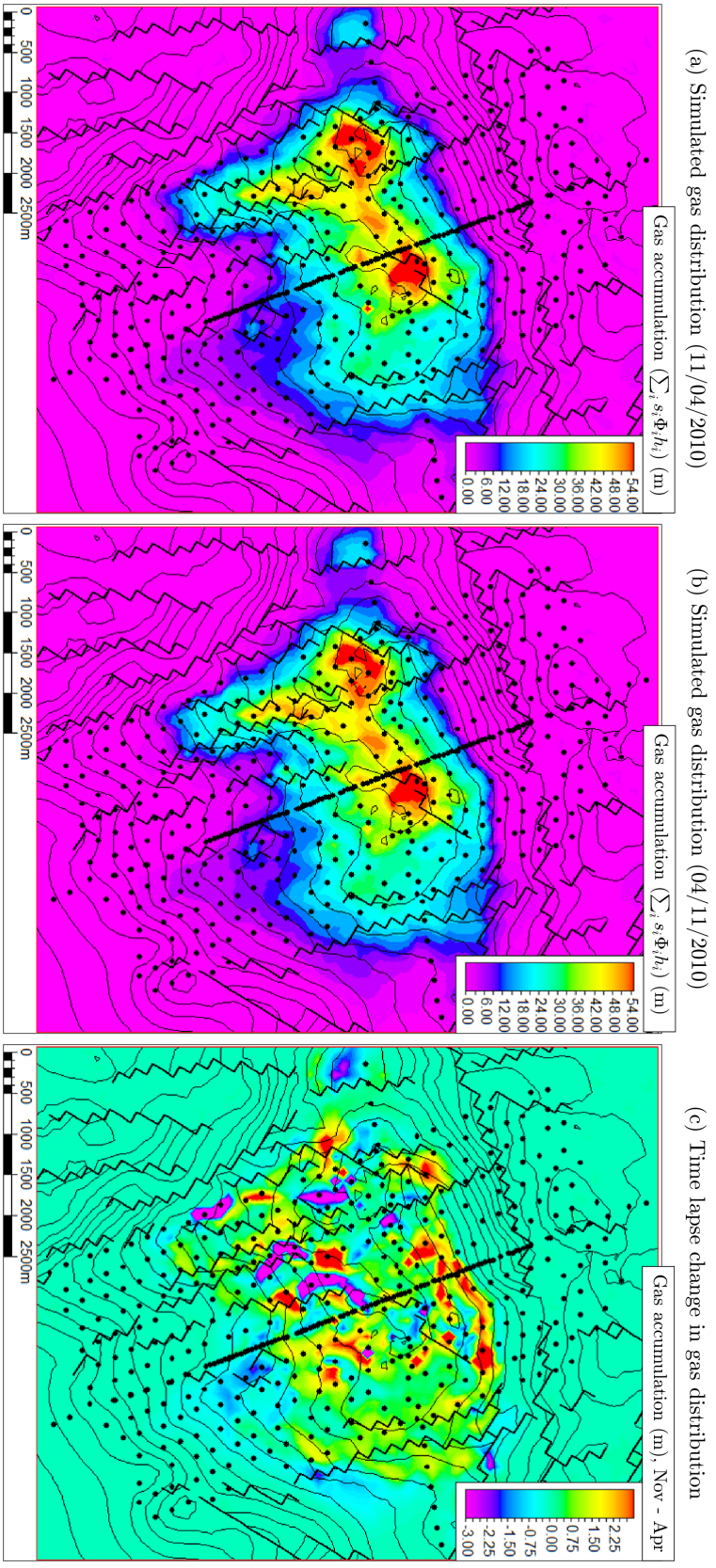


Figure 2.6: Maps of the simulated effective gas column height in (a) April 2010 (first survey) and (b) November 2010 (second survey). The difference (b)-(a) is shown in (c). The simulation is constrained by fifty years of history-matching with the well data. The background map shows the topography of the reservoir layer (upper boundary). The known faults are shown with black zigzag solid lines.

every day. There was an almost uniform spacing of 500 m between the sensors installed over a 3×6 km area located above the UGS. For the low-stock survey, linear profiles were recorded on 16 April (profile BB' in Fig. 2.4), 18-19 April (profile AA') and 24 April (profile CC'). For the high-stock survey, the profile AA' was recorded on 12-13 November, but the data were of very bad quality, presumably because of a storm which occurred at the same time. On 15 November, a crossed configuration was recorded (both AA' and BB' at the same time), which yields us at least one AA' profile for the high-stock survey, though with rather sparse receivers. The other profiles were recorded on 16 November (BB') and 17-18 November (CC'). The spacing in the linear profiles was of 250 m. The spacing of the profile AA' was locally refined to 50 m above the structure of interest. The profile CC' aimed to check the response of a virgin anticline (without gas), and was also refined to 50 m spacing above the target zone. However, it appears from Fig. 2.4 that this profile was slightly misplaced compared to the top of the structure.

In the remainder of this section, we extract attributes from the passive seismic data and compare their spatial and temporal variations to the maps shown in Fig. 2.6, as well as to the available structural maps of the site. In Chapter 3, we analyse the wavefield composition of the ambient noise recorded during this experiment.

2.1.4 Signal processing

As all the seismometers used in the experiment were of the same model, performing a full instrumental correction would not change the relative amplitudes of the signals. To obtain an estimate of the absolute signal strength, we correct the raw signals by the constant gain provided by the manufacturer. The particle velocity recordings on each component are divided into 40 second segments $u_i(t)$ overlapping by 50%. After tapering each time window in the time-domain, Fourier transformation is performed to obtain $u_i(\omega)$. The Power Spectral Density (PSD) of each time window is estimated following Bormann & Wielandt (2013):

$$p_i(f) = \frac{\Delta t}{N_T} \left[u_i(f)\right]^2 \ , \tag{2.2}$$

with N_T the number of samples in the $\Delta t = 40$ s time window (2^{12} samples in our case). We use decibel with respect to 1 $m^2.s^{-1}$ to represent the absolute PSD level. In order to obtain a PSD representative of one night of recording, we take the median $p_i(f)$ values independently at each frequency, after selecting a subset of «quiet» time windows. Though it is more common practice to use the arithmetic mean or the maximum probability density function (PDF) estimate of $p_i(f)$ (e.g. Lambert et al., 2011; Riahi et al., 2013b), we found that the median estimate was quite robust with respect to the time window selection, which we illustrate below.

The idea of selecting quiet time windows is based on the paradigm that the hydrocarbon-related signal is carried by body waves which can be hidden by strong surface waves propagating in the near-surface during noisy time periods (e.g. Riahi et al., 2013b). In Section 2.2, where we present the data recorded by a permanent array, we specifically analyse the influence of the processed time period on the presumably hydrocarbon-related spectral anomalies. For now, we restrict the processed data to night-time

periods (01-05 AM local time) as a first step to reduce the influence of the artificial noise. In a second step, in order to mute transient events, we test four different strategies based on the amplitude evolution of individual recordings. To illustrate the behaviour of these strategies, we choose two demo stations in Fig. 2.4a, shown with blue and red circles. These two stations can be considered as typical in the sense that one them is located above the reservoir at the crossing of profiles AA' and BB' (station 317193, red), while the other one is located outside the reservoir, at the southern end of the profile AA' (station 317254, blue, in the bottom of Fig. 2.4a). The 4-hour vertical component velocity time series recorded for both stations and band-passed between 1 and 3 Hz are shown in blue in the two columns of Fig. 2.7. The choice of the frequency band is guided by a preliminary data analysis which pointed towards a possible reservoir signature in this frequency band. We note that both the transients and the stationary part of the signal appear amplified at the station 317193, which we label as *strong HMT*, with respect to the station 317254, which we label as *weak HMT*. We test two selection criteria based on the removal of individual time samples as outliers based on their amplitude, and two another criteria based on the removal of some time windows based on their energy. For each tested selection criterion (presented below), we indicate the rejected time windows by the red shadowed zones in Fig. 2.7.

For *sample-based* criteria, if the amplitude of one sample in the band-passed signal is above a certain threshold, the two overlapping time windows containing this samples are rejected. The instantaneous amplitude is measured via the envelope $e(t) = |s_a(t)|$ of the time-domain signal, with $s_a(t)$ the analytic signal associated to the particle velocity recording $u(t)$. The first tested criterion is to reject all the samples of the envelope beyond $\text{med}(e) + 3\sigma(e)$, with $\text{med}(e)$ and $\sigma(e)$ the median and the standard deviation of the $e(t)$ samples over the four hours of recording. This strategy appears as rather aggressive applied to the weak HMT signal (Fig. 2.7c), but seems to efficiently remove the transients from the strong HMT signal (Fig. 2.7d). A far less aggressive strategy is to define the envelope threshold as a high quantile (99.9% in our case) on the whole sample series. As shown in Figs 2.7(g) and (h), only the few time windows containing the strongest spikes are removed in this case.

For *window-based* criteria, if the energy of the window is above a threshold value, the window is rejected. That is, $p_i(f)$ is computed for each of the 720 windows of 40 s according to equation (2.2). Each $p_i(f)$ is integrated between 1 and 3 Hz. This yields a series of 720 scalar energies $p_i|_1^3$. For this population, we tested similar selection criteria as for the envelope sample population described above. The very agressive 10% quantile criterion (Figs 2.7a and b) was proposed Riahi et al. (2013b), except in their case it was applied individually to each frequency, instead of integrating single-window $p_i(f)$ over a fixed frequency band before statistical analysis. In their case, both day and night data were processed together, so the effect of this very selective criterion was partly dedicated to keep the night-time data, which we did from the very beginning. Also, they took the mean value of the PSD in the remaining windows, instead of the median in our case. On the other hand, as the time-window energies are much less dispersed that the envelope samples, a threshold formulated as $\text{med}(p|_1^3) + \sigma(p|_1^3)$ keeps the vast majority of time windows (Figs. 2.7(e) and (f)).

Computing the representative PSD spectra following three of the tested strategies (except the 10% quantile on windows) leads to very similar results. The three corresponding spectra are almost perfectly superimposed in Fig. 2.8, despite the different amount of the selected time windows. This is due to the robustness of the median estimate compared to the mean estimate, since it is not sensitive to anomalously energetic time windows which can be missed by the window-selection step. In Fig. 2.8, we also show the relative anomaly

$$\eta = \log\left(\frac{\mathrm{PSD}_s}{\mathrm{PSD}_w}\right) \tag{2.3}$$

of the strong HMT station with respect to the weak one, measured in decibel, as a function of frequency. The relative anomalies are almost identical for three aforementioned window selection procedures (all except the 10% quantile on windows). Working with relative anomalies is a convenient procedure to mitigate the temporal variations of the ambient noise amplitude. Also, the interpretation of the HMT attributes is usually performed based on their spatial variations, i.e. relative differences between stations. Thus a robust estimation of the relative anomalies is more important than an accurate estimation of the absolute noise level. The problem of the 10% quantile on windows is that it keeps a too small amount of windows from a 4-hour recording to obtain a smooth representative spectrum. This results into a noisy and thus less reliable anomaly pattern in Fig. 2.8c. This issue could be fixed by smoothing the spectra, but this would introduce an additional processing step, with additional user-fixed parameters. Based on the results shown in this paragraph, we prefer choosing the less aggressive window selection procedure, namely the 99.9% quantile on the envelope samples. As the representative PSD is estimated as a median over time windows, we know that the final result is not too much sensitive to the choice of the window selection criterion.

Another interesting observation which can be made from Fig. 2.8b is that the estimated PSD for the 10% quantile criterion is significantly below the other spectra *only* in the 1-3 Hz band (shaded blue), used for band-pass filtering prior to the window selection. This indirectly shows that the energy of this frequency band is not correlated in time with other frequency bands, neither the microseismic peak at 0.2 Hz, nor the human noise around 4 Hz and above.

2.1.5 Spectral anomalies and attribute profiles

Attribute definition

As in most of the previous case studies, we analyse the PSD_z spectra, estimated as explained in the previous paragraph, and the V/H ratio. We compute V/H using the *median* PSD spectra (i.e. *not* averaging V/H over individual windows):

$$\mathrm{V/H} = \frac{\mathrm{PSD}_z}{\sqrt{\frac{1}{2}\left(\mathrm{PSD}_n^2 + \mathrm{PSD}_e^2\right)}}\ , \tag{2.4}$$

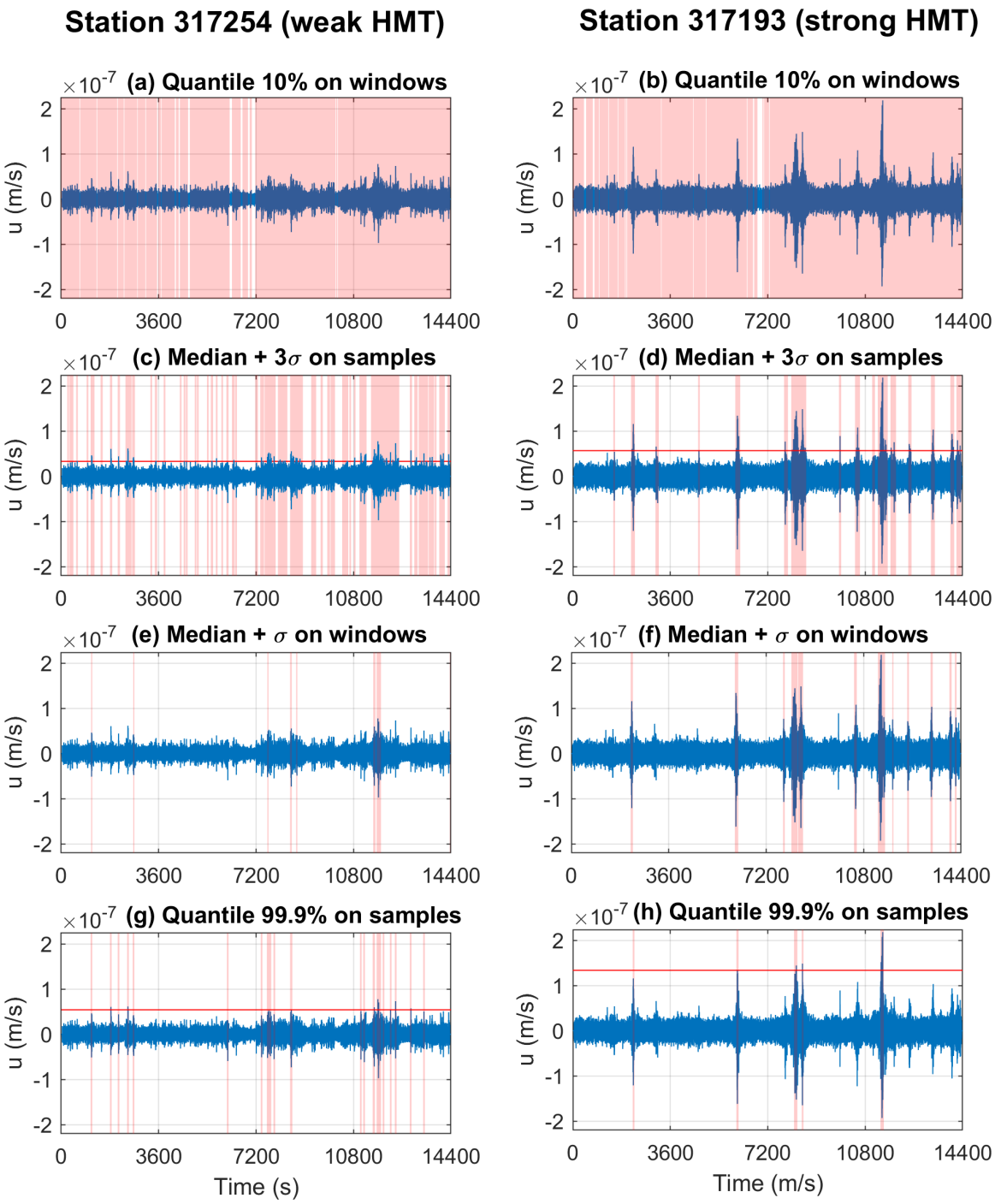


Figure 2.7: 4 hours of signal recorded at stations 317254 (left column) and 317193 (right column), respectively blue and red in Fig. 2.4. The displayed signals are band-passed between 1 and 3 Hz. From top to bottom, the red areas indicate the discarded portion of the signal for four different time selection criteria.

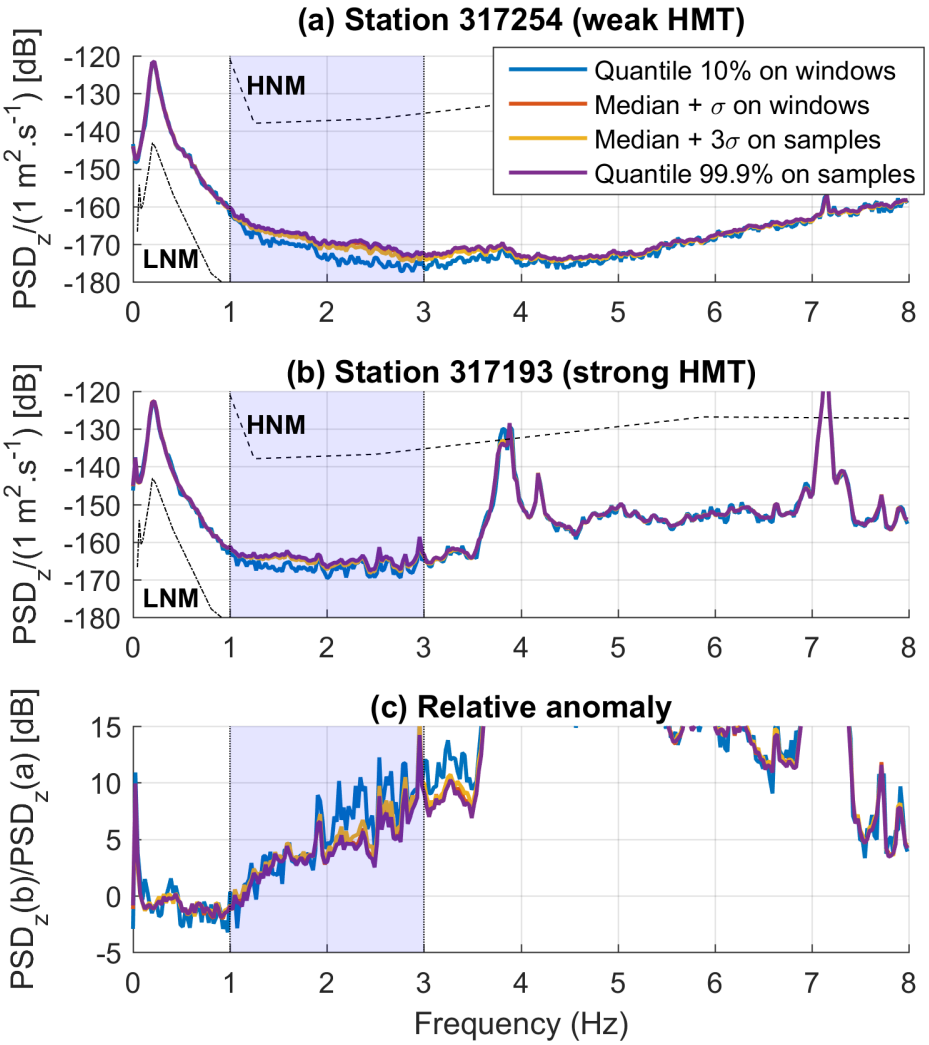


Figure 2.8: Median PSD_z spectra on the vertical component over the remaining (white) time windows. Results are compared for the four time selection criteria shown in Fig. 2.7. The frequency band used for applying the selection criteria is highlighted in blue (1-3 Hz). (a) Station 317254. (b) Station 317193. Dashed lines show the high and low noise models from Peterson (1993). (c) Relative PSD_z anomaly between the spectra shown in a and b. Note the similarity of the anomalies for the three last time selection criteria.

where $\mathrm{PSD}_{n,e}$ are the PSD spectra of the two horizontal components estimated using the time selection procedure based on the vertical component. The PSD_z and V/H spectra are compared for the two demo stations in Figs 2.4(b) and (c), respectively. While the PSD_z for the strong HMT station (317193) seems anomalous over a wide frequency band (Fig. 2.4b), we observed that this pattern was most stable below 2.5 Hz. For the definition of the integral attributes $\mathrm{PSD}_z|_{f_1}^{f_2}$ and $V/H|_{f_1}^{f_2}$, we refer to equation (1.1). However, we choose to work with *mean* attributes, which implies dividing the integral attributes by $(f_2 - f_1)$. Still, we keep the same notation as for integral attributes for sake of simplicity. First, we define the mean attribute $\mathrm{PSD}_z|_{1.2}^{2.4}$ to construct anomaly profiles and maps. In Fig. 2.4c, the V/H pattern also appears to exhibit strong values in the 1.2-2.4 Hz band, which motivates the definition of a second attribute $\mathrm{V/H}|_{1.2}^{2.4}$. Prior to mapping these attributes, a QC step is performed to reject stations with abnormal spectra. This step is performed by visual inspection of all the simultaneously recorded PSD_z spectra and manual picking of evident outliers. The procedure is then repeated for the V/H ratio spectra. Up to 30% of stations are rejected for some nights during the November survey, while this proportion does not exceed 5% in April. For the remaining stations, the evolution of the two attributes along profiles AA', BB' and CC' is shown in Fig. 2.9.

$\mathbf{PSD}_z|_{1.2}^{2.4}$ profiles

The $\mathrm{PSD}_z|_{1.2}^{2.4}$ attribute values are normalised with respect to a reference station located on the profile at x_0, far from the reservoir. The plotted value is thus

$$\eta|_{1.2}^{2.4}(x) = \frac{\mathrm{PSD}_z(x)|_{1.2}^{2.4} - \mathrm{PSD}_z(x_0)|_{1.2}^{2.4}}{\mathrm{PSD}_z(x_0)|_{1.2}^{2.4}} \ . \tag{2.5}$$

The available topography of the geological layers above the reservoir up to the surface is shown along these profiles in Figs 2.9a, b and c, respectively. For the AA' profile, $\mathrm{PSD}_z|_{1.2}^{2.4}$ exhibits some correlation with the structural topography (gradual increase (Fig. 2.9a) towards the top of the anticline). However, a strong increase occurs in the gas bearing zone, shaded in green in Fig. 2.9d. The profiles for 18 and 19 April (Fig. 2.9d) are quite stable and exhibit a peaking PSD anomaly of 150% above the reservoir. Overall anomaly values above the reservoir range from 50 to 100%. The profile for 15 November exhibits somewhat lower anomalies in the reservoir zone (peaking slightly above 50%). However, the latter profile is not very reliable because of the sparsity of the receivers and the high number of rejected stations with evidently abnormal PSD_z spectra. For the BB' profile, the attribute anomaly values are very similar on 16 April and on 15-16 November outside the gas bearing zone (Fig. 2.9d). Shortly before $x \approx 2000$ m, the increase of the attribute anomaly coincides with a structural top of the shallow layers (above 400 m). Another increase of the attribute anomaly occurs as the profile crosses the structural fault at $x \approx 6000$ m. The magnitude of these two local anomaly maxima is of 50-60%. The strongest anomalies, however, seem to be correlated with the gas-bearing zone (shaded in green). Here, some discrepancy arises between the different profiles. On 16 April, the anomaly peaks before $x = 8000$ m with a magnitude about 200%, and then definitely drops below 100%. On 16 November, on the contrary, the magnitude of the first peak is lower (150%), but a second broad peak of similar magnitude is visible around $x \approx 9000$ m.

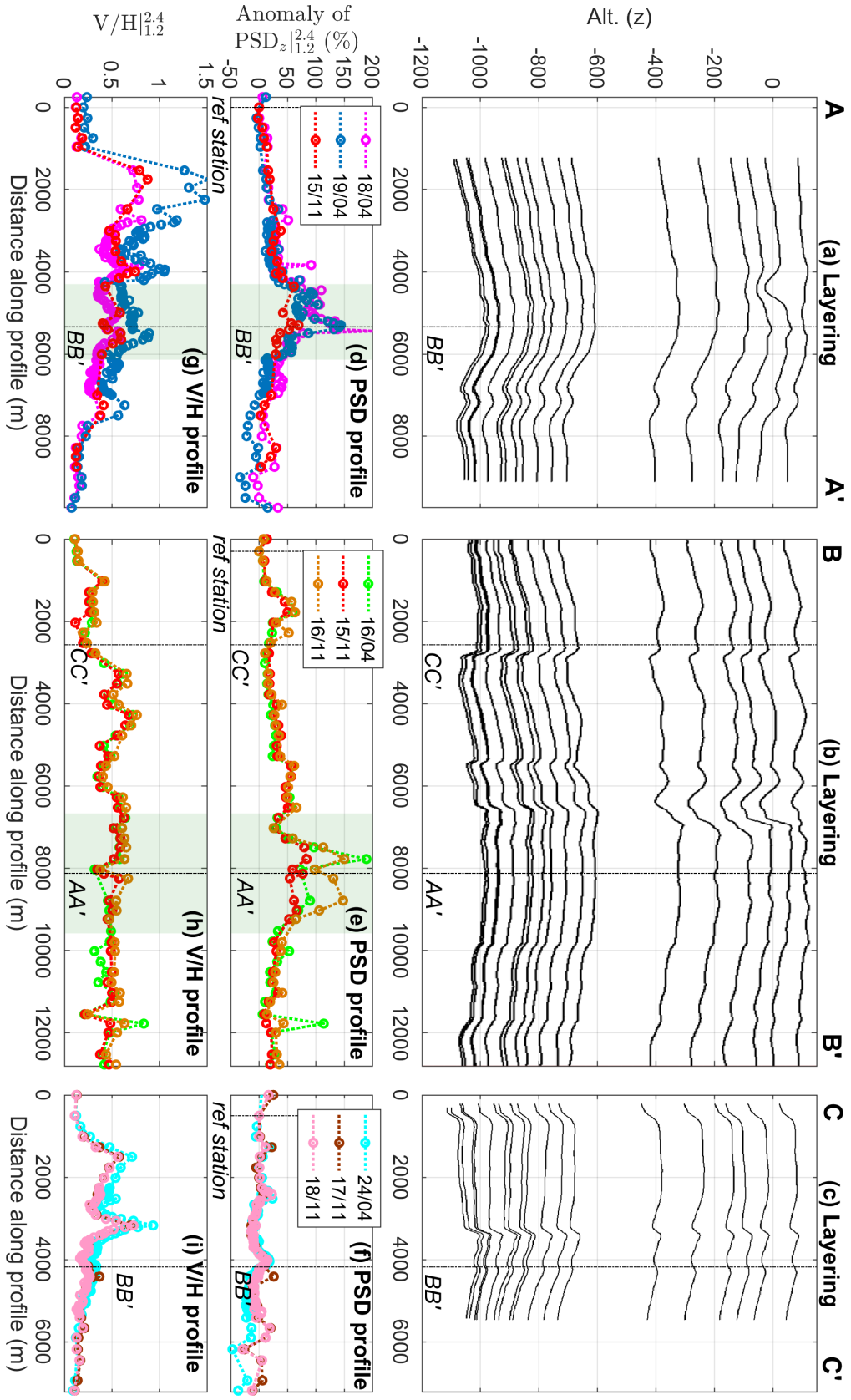


Figure 2.9: Attribute profiles along 2D seismometer lines. (a)-(c) Known geological layers along the profiles AA', BB' and CC' shown in Fig. 2.4a. (d)-(f) Corresponding $\eta|_{1.2}^{2.4}$ profiles (relative PSD$_z$ anomaly, see equation 2.5). The positions of the reference stations are shown with black dotted lines, as well as the intersections with other profiles. (g)-(i) Same as (d)-(f) for the V/H$|_{1.2}^{2.4}$ attribute, with raw values (no normalisation by a reference station).

The data of 15 November suffer from sparsity, but seem to exhibit lower anomalies in the gas-bearing zone, compared to 16 November. This gives an idea of the uncertainty of the computed attributes. Ideally, long time-series should be analysed, as done in Section 2.2. Finally, the CC' profile exhibits a rather flat $\mathrm{PSD}_z|_{1.2}^{2.4}$ profile, consistently with the limited structural perturbations (see Figs 2.9c and 2.4a), as well as with the absence of gas.

$\mathbf{V/H}|_{1.2}^{2.4}$ profiles

The $\mathrm{V/H}|_{1.2}^{2.4}$ profiles along AA', shown in Fig. 2.9f, are consistent with the observation made from Fig. 2.4c: V/H ratio between 1.2 and 2.4 Hz was higher in the center of the structure compared to the southern end of the AA' profile. However, the $\mathrm{V/H}|_{1.2}^{2.4}$ profiles seem correlated neither to the layer topography in Fig. 2.4a, nor to the highlighted gas-bearing zone. A strong variation of the ratio is observed between 18 and 19 April, which is surprising. As the vertical component PSD remained stable for these two dates (Fig. 2.4a), this means that the horizontal component PSD_z dropped on 19 April. We have not further investigated the reasons of such an evolution. The $\mathrm{V/H}|_{1.2}^{2.4}$ profiles along BB' (Fig. 2.9g) also exhibit no visible correlation with either topography or gas presence. Finally, the $\mathrm{V/H}|_{1.2}^{2.4}$ profiles along CC' display a significant increase between $x = 1000\ m$ and $x = 4000\ m$, while this interval contains neither gas nor significant geological structures. Thus we decide to discard $\mathrm{V/H}|_{1.2}^{2.4}$ as an attribute for further analysis. However, we incorporate the information carried by the V/H ratio in Section 2.1.7 through automated spectra classification.

2.1.6 Interpolated attribute maps

$\mathbf{PSD}_z|_{1.2}^{2.4}$ maps

The 2D profiles of the attribute $\mathrm{PSD}_z|_{1.2}^{2.4}$, seemed correlated to both the geological structure and the gas presence in the previous section. Consequently, it is of interest to represent the variations of this attribute as a surface map to be compared with the simulated gas saturation (Fig. 2.6) and the geological layers topography. The data recorded by the dense rectangular arrays on 20-23 April and 4-11 November are used for this purpose. As the rectangular arrays are shifted every day, we do not have any fix reference station for the relative anomaly calculation. To circumvent this problem, after a preliminary analysis of the attribute distributions for individual days, we define a zone were the attribute values are systematically the lowest. This zone is shaded in grey in Fig. 2.10a (north-western part). For each processed night, we use the median $\mathrm{PSD}_z|_{1.2}^{2.4}$ in this zone as a reference for normalisation, and then compute the relative anomaly as in the previous section (equation 2.5). This allows to efficiently remove the temporal variations of the ambient noise amplitude, yielding smooth anomaly maps. They are displayed in Figs 2.10a and b, respectively for April (low stock) and November (high stock) surveys. The difference $\eta_{\mathrm{Nov}} - \eta_{\mathrm{Apr}}$ is shown in Fig. 2.10c. As the normalisation reference has changed with respect to the previous section, the absolute values of the anomalies cannot be compared. Neverthess, they are of the same order of magnitude, with maximum values around 150%. The maps are overlaid over the topography of the reservoir layer upper boundary (isolines),

as a proxy for the structural topography, which is in fact rather similar for all the layers above the reservoir (see Figs 2.9a, b and c). The map obtained in November is more consistent with the saturation maps (Figs 2.6a and b) than the map obtained in April, as the former presents two distinct anomalous zones, on both sides of the profile AA'. However, the western anomaly is significantly shifted towards the east compared with the simulated saturation maximum which is adjacent to the western boundary of the rectangular seismometer network. In April, the anomalies on the eastern side of the AA' profile are small, consistently with the BB' profile analysed in the previous section (Fig. 2.9e). While at first glance such a difference could appear consistent with the higher stock level in November, the saturation simulations show that only a small increase of the gas bubble extension is expected in November, and it should occur along the bubble exterior boundary. This is not what we observe in the time-lapse difference map in Fig. 2.10c, where the increase of the anomaly is well visible on the eastern side of the AA' profile, while a decrease is observed along the fault zone in the south-western part of the array. Thus we cannot make a straightforward link between the gas stock evolution and the anomaly strength evolution. Other factors could influence the change of the anomaly pattern, such as changes in the ambient noise source distribution, addressed in Chapter 3, or the water table evolution between April and November, invoked by Riahi et al. (2013a) as a possible way to explain Rayleigh wave anisotropy change observed in the Chémery dataset.

PSD local slope

In practice, the anomaly maps depend on the precise choice of the attribute, as well as on the precise bounds of the frequency band. This is an important factor of the method subjectivity. In this paragraph, we show how introducing a modified attribute based on PSD_z allows to obtain anomaly maps more consistent with the gas saturation maps. The new attribute is based on the logarithmic slope of PSD_z within the frequency band of the anomaly (see the different slopes in the grey band in Fig. 2.4a). We choose to use the mean slope in the most stable region within the anomaly band, i.e. between 1.3 and 1.6 Hz. At each frequency, the local slope is estimated by performing a linear regression of $\log\left(\mathrm{PSD}_z(f)/1\ \mathrm{m^2.s^{-1}}\right)$ over a 0.5 Hz window centered on the frequency of interest. We call this attribute $\mathrm{PSD}_z^{\mathrm{slope}}|_{1.3}^{1.6}$. Since it does not measure any physically meaningful quantity, we treat it statistically instead of using anomalies with respect to a reference station/zone. For each processed night, we compute the mean $\overline{\mathrm{PSD}_z^{\mathrm{slope}}|_{1.3}^{1.6}}$ and the standard deviation $\sigma(\mathrm{PSD}_z^{\mathrm{slope}}|_{1.3}^{1.6})$ over the whole array, and then scale the distribution as

$$\widetilde{\mathrm{PSD}}_z^{\mathrm{slope}}|_{1.3}^{1.6} = \frac{\mathrm{PSD}_z^{\mathrm{slope}}|_{1.3}^{1.6} - \overline{\mathrm{PSD}_z^{\mathrm{slope}}|_{1.3}^{1.6}}}{\sigma(\mathrm{PSD}_z^{\mathrm{slope}}|_{1.3}^{1.6})} \ . \tag{2.6}$$

The resulting scaled quantity measures the deviation of the slope attribute from the mean value, with one standard deviation as a unit. Such a scaling also corrects for the temporal variation of the ambient noise amplitude, just as the normalisation by a reference station. The resulting maps are shown in Fig. 2.11a and b, respectively for April and November. While they lead to the same conclusions as the normalised $\mathrm{PSD}_z|_{1.2}^{2.4}$ maps presented in the previous paragraph, they are smoother and present

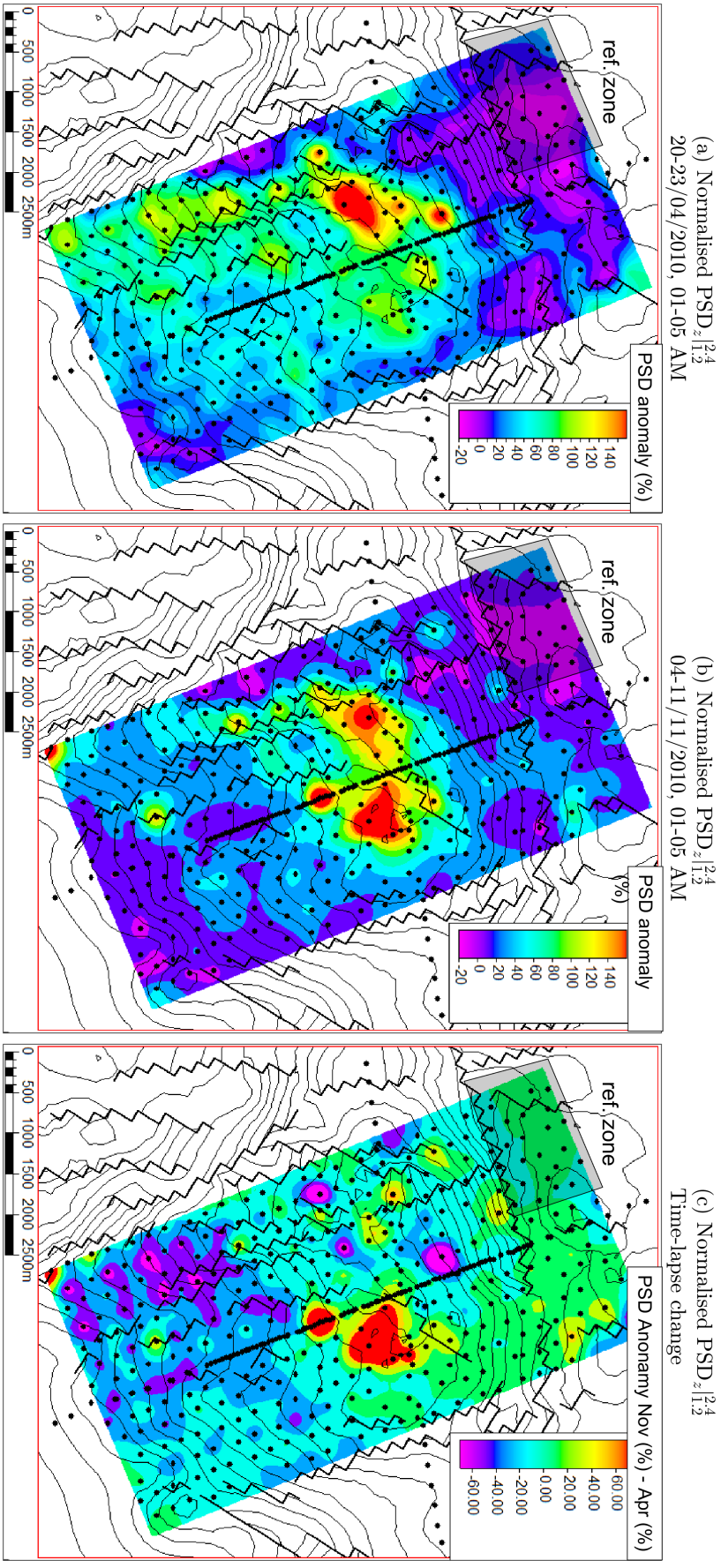


Figure 2.10: Maps of the relative PSD$_z$ anomalies $\eta|_{1.2}^{2.4}$ (see equation 2.5). The reference zone is shown with the grey rectangle. (a) April 2010. (b) Novermber 2010. (c) Difference (b)-(a). The background map is the same as in Fig. 2.6.

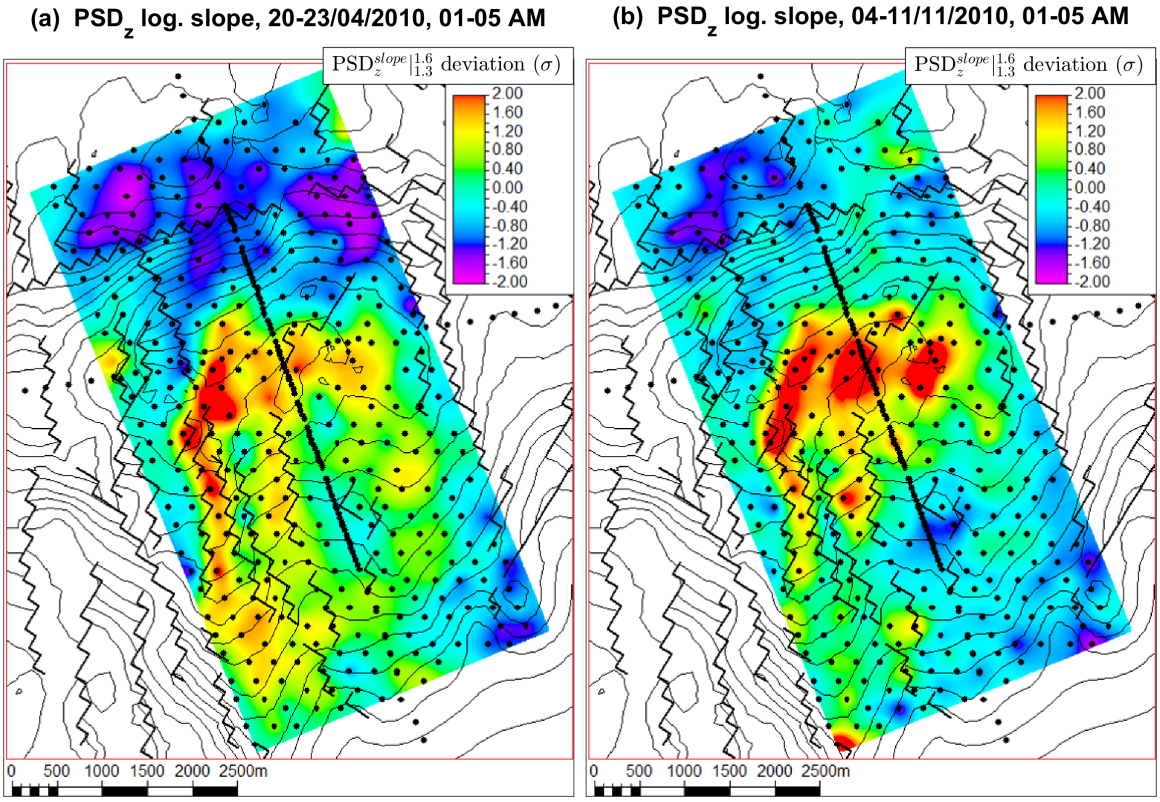


Figure 2.11: Maps of the scaled slope-based attribute $\widetilde{PSD}_z^{slope}|_{1.3}^{1.6}$ (see equation 2.6). (a) April 2010. (b) Novermber 2010. The background map is the same as in Fig. 2.6.

a clearer correlation with the fault corridors in the south-western part of the array, which could indeed contain some quantity of gas, as can be seen in the saturation simulation results in Figs 2.6a and b. These maps were presented in the introduction of the manuscript as motivating results for a deeper investigation of the HMT existence and nature. However, they are based on a subjective choice of the attribute and a fine tuning of the frequency band. In the next section, we introduce a more objective approach to produce spectral anomaly maps, based on automated spectrum classification

2.1.7 Automated spectrum classification

In this section, we present an automated methodology to integrate multi-attribute data in a single map. This methodology has been patented by Storengy (Huguet et al., 2017). We propose to consider each station as a data point evolving in a multi-dimensional data space. The size of the data space corresponds to the sum of the discrete frequencies considered for each non-integrated spectral attributes (such as $PSD_z(f)$ or $V/H(f)$):

$$N_d = \sum_i A_i(f) N_{i,f} \ , \tag{2.7}$$

where i refers to different attributes A_i. Using integral attributes is equivalent to projecting the data points from the full multi-dimensional data space on a linear one-dimensional sub-space. We improve this projection by introducing a one-dimensional *manifold*, which can be intuitively represented as a *deformed* linear sub-space. This

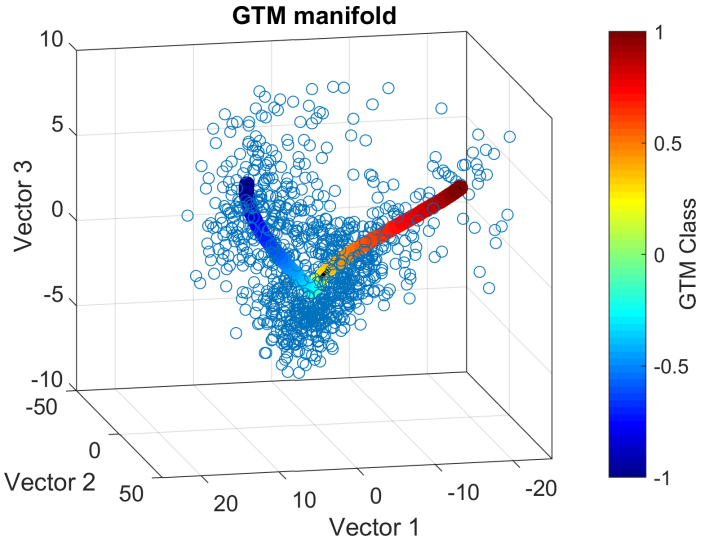


Figure 2.12: Visualisation of the GTM manifold in the three-dimensionnal space generated by the three major eigenvectors of the data cloud. The data points are shown in blue. The class centers lie on the manifold and form its uniform 1D grid with values ranging from from -1 to 1.

is achieved via the *Generative Topographic Mapping* (GTM) algorithm, invented by Bishop et al. (1998).

GTM algorithm

We use the available GTM algorithm implementation from the NETLAB toolbox, coded in Matlab, developed by Nabney (2002). The idea of the GTM algorithm is to use a combination of radial basis functions (RBF, Broomhead & Lowe, 1988) to map the data space on a manifold of reduced dimension. In our application, we use a one-dimensionnal manifold because it allows to directly display the projection in geographical coordinates, as for the integral attributes. A curvilinear coordinate is introduced along the manifold. The cluster centers are constrained to form a uniform grid of this coordinate, arbitrary set from -1 to 1. Then the cluster positions are optimised with respect to the data cloud in the data space, using an expectation-maximisation (EM) algorithm with a Gaussian Mixture Model (see Nabney (2002) for implementation details). At each iteration, the parameters of the RBF network are updated to adapt the shape of the one-dimensional manifold to the shape of the data cloud. At the end of the algorithm, the manifold optimally fits data cloud, as illustrated in Fig. 2.12. This yields a continuous representation of the data with a scalar class number ranging from -1 to 1, as the class centers form a grid from -1 to 1. That is, for each data point, we can represent the assigned class number in a map view.

Application to Chémery data

The procedure is applied to jointly cluster the PSD_z and the V/H spectra, respectively between [0.8,2.5] Hz and [0.3,1.5] Hz. The definition of the frequency band is still required, since the supposed HMT anomaly must be *the* structuring element of the

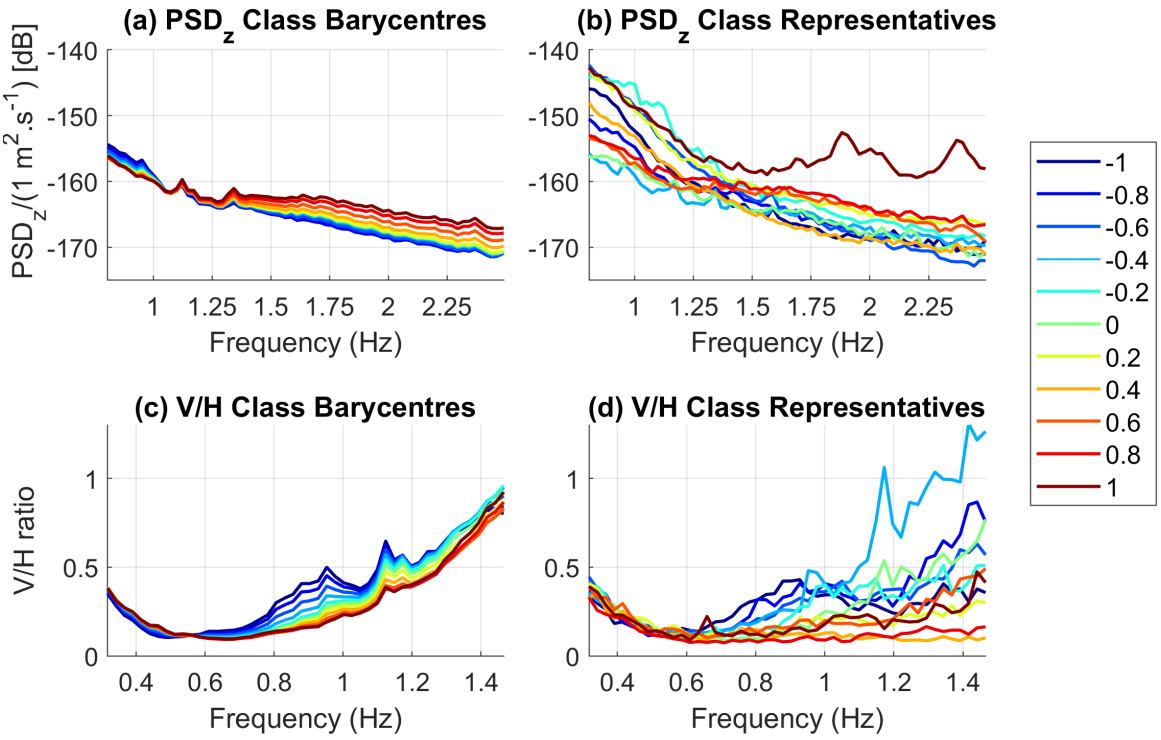


Figure 2.13: Spectral signatures associated to each GTM class. The class centers are represented as (a) PSD spectra and (c) V/H spectra. The true spectra corresponding to the data points nearest to each class center are shown in (b) (PSD spectra) and (d) (V/H spectra).

data cloud for the algorithm to produce relevant maps. However, the sensitivity of the final map to the precise bounds is reduced with respect to the integral attributes. The choice of the [0.3,1.5] Hz for the V/H ratio is motivated by Fig. 2.4c, where minimum V/H ratio of the red curve (around 0.8 Hz) is shifted towards a higher frequency compared to the blue curve. The next steps are:

1. Each spectral quantity is scaled *daily* over all the available receivers, similarly to equation (2.6), but individually at each frequency.

2. Principal Component Analysis (PCA, Pearson, 1901) is performed on the resulting population to reduce the data space to three dimensions by a linear projection on the subsbace generated by the three major eigenvectors.

3. The GTM algorithm is then applied in the three-dimensional data space to obtain a continuous clustering of the data points between -1 and 1 (Fig. 2.12).

4. By applying the Moore-Penrose pseudo-inverse (Moore, 1920) of the linear projector used in step 2, each class center can be mapped back to the full data space

5. Unscaling the obtained data points (inverse of equation 2.6) allows to represent each class center as a PSD_z spectrum or a V/H ratio (Figs 2.13a and c). For unscaling, the mean and the standard deviation are chosen arbitrary from one of the processed nights, as the absolute noise levels are variable over time.

6. For each class center, the nearest-lying real data point in the three-dimensional subspace (see step 3) can also be represented to get an idea of the representativity of the class centers (Figs 2.13b and d).

7. The cluster numbers between -1 and 1 attributed to each station (i.e. each data point) are interpolated and represented on a map (Fig. 2.14). A given colour on the map corresponds to an associated spectral signature in terms of both PSD_z and V/H spectra in Fig. 2.13.

The spectral signature of the high class numbers (close to 1) in comparison to the low class numbers (close to -1) is threefold. The first effect is the increase of the PSD_z spectra above 1.2 Hz (see the red spectra in Fig. 2.13a), which we used for the integral attribute maps and profiles in the previous sections. The second effect is a slight decrease of the PSD_z spectra below 1 Hz. The third effect is the shift of the V/H ratio minimum towards higher frequencies above the reservoir (red spectra in Fig. 2.13c). The best class representatives among the real PSD_z spectra appear dispersed in Fig. 2.13b because of the temporal variation of the ambient noise amplitude. However, their shape is consistent with the associated class centers shown in Fig. 2.13a. The same observation holds below 1.1 Hz for the best class representatives among the real V/H spectra shown in Fig. 2.13d. The highest class numbers (red curves) have lower values above 0.6 Hz, which is characteristic of the V/H minimum being shifted towards higher frequencies. Above 1.1 Hz, there is less similarity between the class centres and the class best representatives, which means that these frequencies did not contribute significantly to the three major eigenvectors used for projecting the data during the step 2.

The clustering maps (Fig 2.14) conserve the main features already observed in Figs 2.10 and 2.11, namely the anomaly concentration along the fault corridor in the south-western part of the array (both Figs 2.14a and b), and the stronger anomaly in November on the eastern side of the NS profile (BB'). The time-lapse difference of the clustering maps (Fig. 2.14) exhibits only small variations in the main reservoir zone in the center of the array, consistently with the small simulated saturation variations in Fig. 2.6. However, a strong negative evolution is observed in the south-western fault corridor (southern end), as well as a positive evolution in north-eastern part of the array. The latter variations are not well understood in the current state of our work.

2.1.8 Main conclusions

From this study, we draw the following conclusions:

- The shape of the spectral anomalies observed in Chémery is consistent with the HMT phenomenology described in the literature (see spectra in Fig. 2.4a).

- The PSD_z-based attribute exhibits more correlation with the gas presence than the associated increase of the V/H ratio (see attribute profiles in Fig. 2.9).

- The PSD_z-based attribute appears to be a combination of a structural and a gas effect (see attribute profiles in Fig. 2.9).

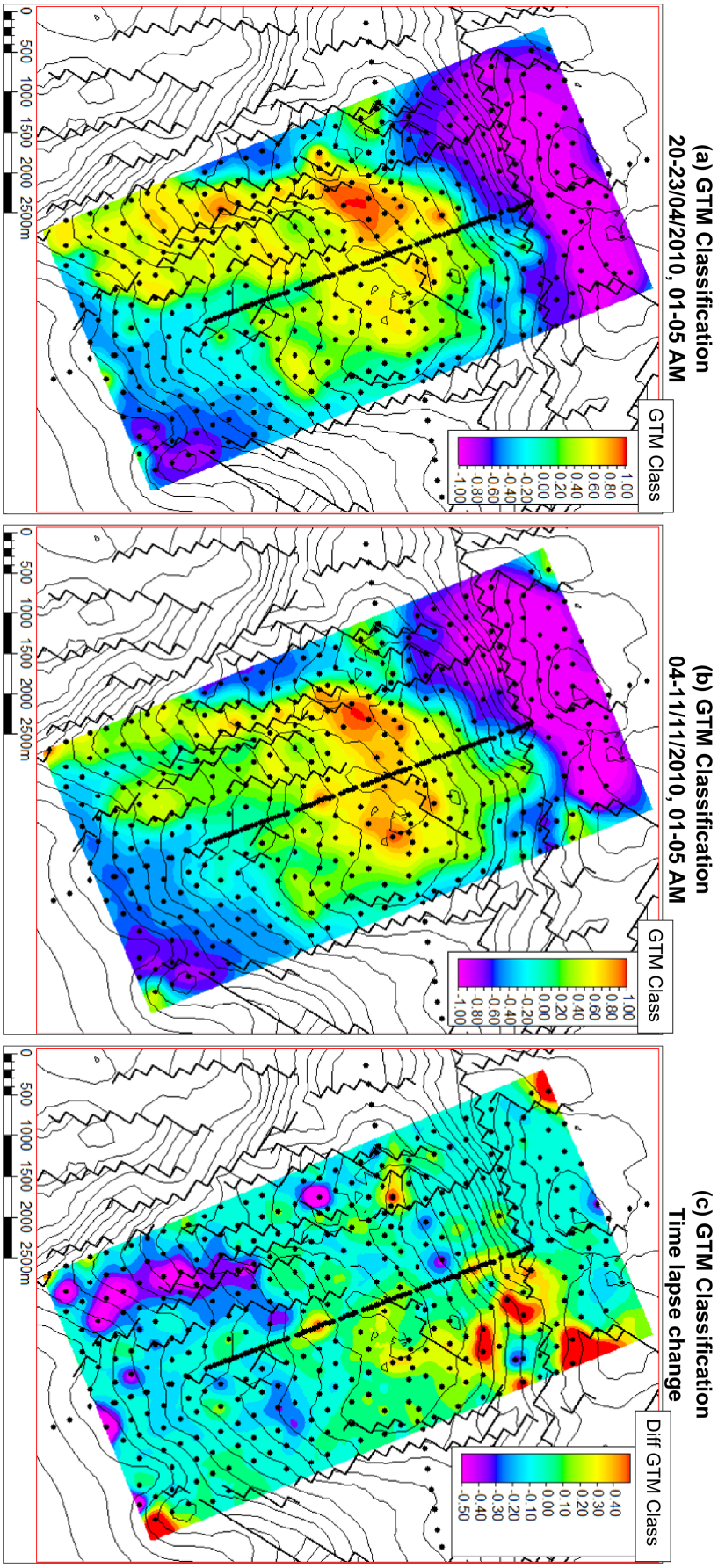


Figure 2.14: Same as Fig. 2.10, but with the GTM class represented instead of the PSD anomaly.

- The time-lapse surveys are consistent with each other, but their small difference is not consistent with the simulated gas saturation. The latter is probably due to the variability of the ambient noise rather than to the medium evolution (compare Fig. 2.6 to Figs 2.10c and 2.14c).
- The automated classification of the spectra appears as a promising method to detect and map multi-attribute HMT signatures.

While in this section we analysed two snapshot datasets with a large number of seismometers, the next section presents a long-time survey with a limited number of seismometers. This complements the first study as it yields an insight in the continuous temporal evolution of the presumed HMT signals.

2.2 Dataset 2: Saint-Illiers underground gas storage

2.2.1 Site presentation

The Saint-Illiers UGS has been operated by Storengy since 1965. Compared to the Chémery UGS, it is located closer to the center of the Paris Basin. The volume of gas stored in Saint-Illiers is about 5 times lower than in Chémery. The reservoir consists of Late Jurassic sands located at a depth of about 500-600 m with respect to the surface (or 350-450 m with respect to the sea level). The average thickness of the reservoir layers is about 30 m. The main reservoir layers have a porosity of 25-30% and a permeability ranging from 0.1 mD to 3 D. The cap rock consists of a 16 m thick deposit of alternating compact marls and limestones. This deposit is overlaid with another 180 m of alternating layers of impermeable marls and limestones. The topographic map of the site is shown in Fig. 2.15, with the production wells concentrated in the lower-middle part (small black dots). The geological profiles along the lines AA' and BB' (Fig. 2.15) are shown in Fig. 2.16. The geological series described above are indicated by the red annotations. The maximum allowed volume of the gas bubble is shown in green. The structural anticline (geological trap) is very abrupt in the south and more smooth in the north. No major faults are reported to affect the site.

2.2.2 Seismic network and gas accumulation evolution

We started the continuous monitoring experiment in May 2016, during the first year of the present PhD work. The experiment contains three distinct phases. The corresponding stations are shown with different colours in Fig. 2.15.

Phase 1 (May-August 2016)

During the first phase (black dots), only the central zone of the gas bubble was surveyed. As the production well pads are equipped with small buildings with electric plugs, we installed the sensors inside these buildings, directly on the floor, plugged to the current via a 12 V AC/DC converter. This phase lasted three months, during an injection cycle (Fig. 2.17). As shown in the following, the noise level recorded during

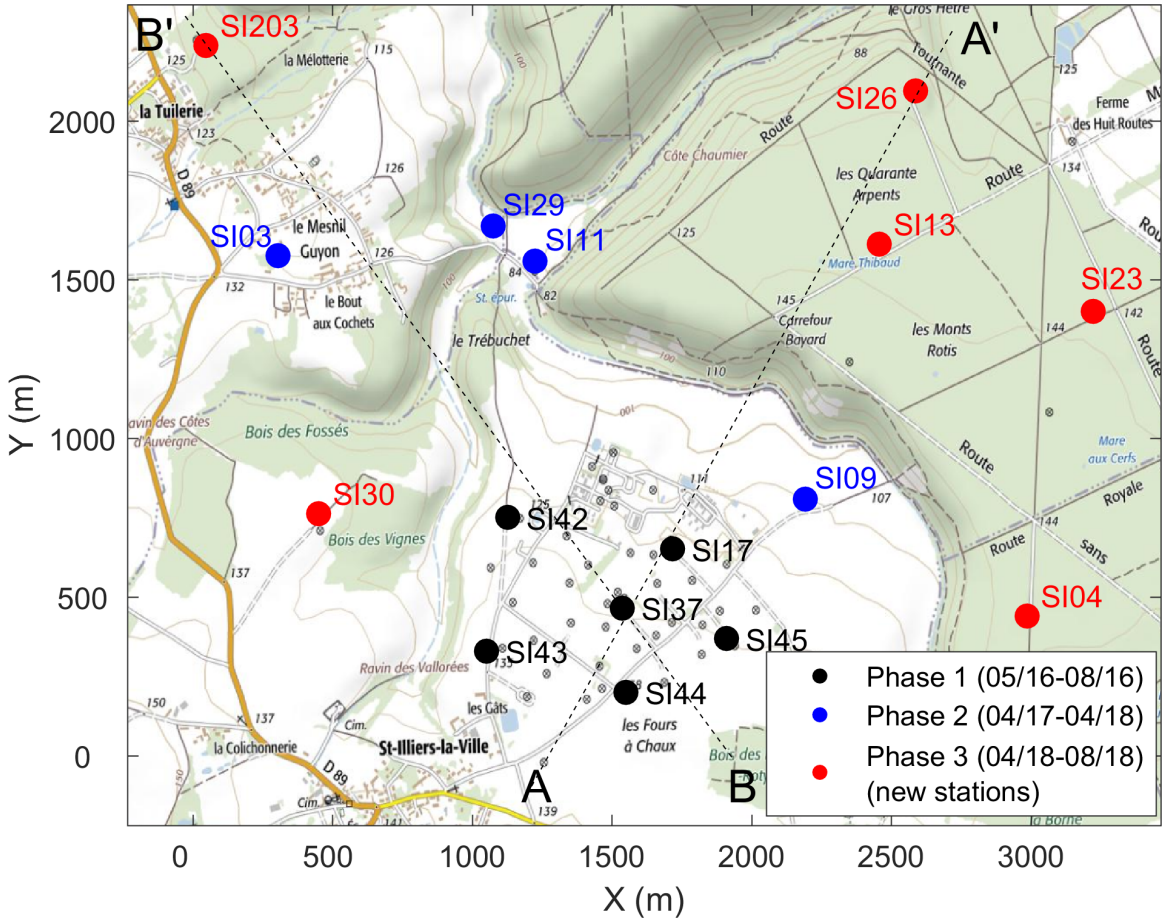
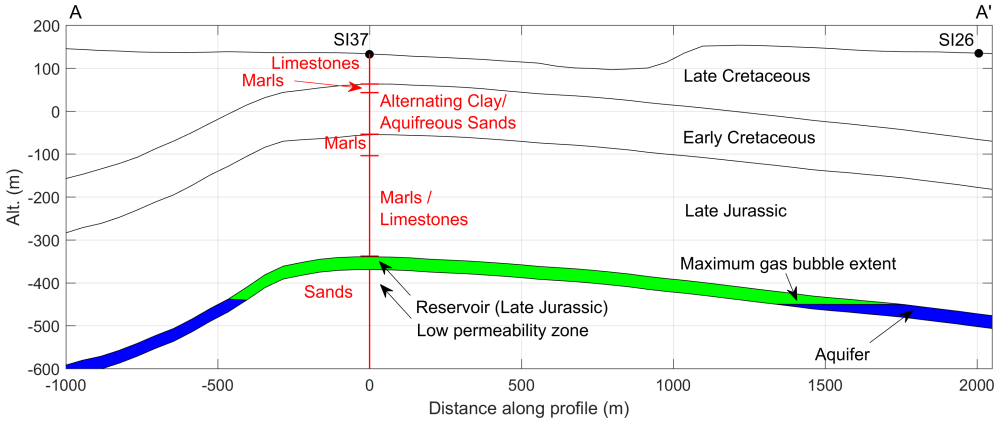


Figure 2.15: Topographic map of the Saint-Illiers site. The small black dots show the well pads. The stations installed during the different phases of the experiment are shown with different colours. The phase 3 contains both blue and red stations. SI37 is common to all the phases.

(a) Profile AA'



(b) Profile BB'

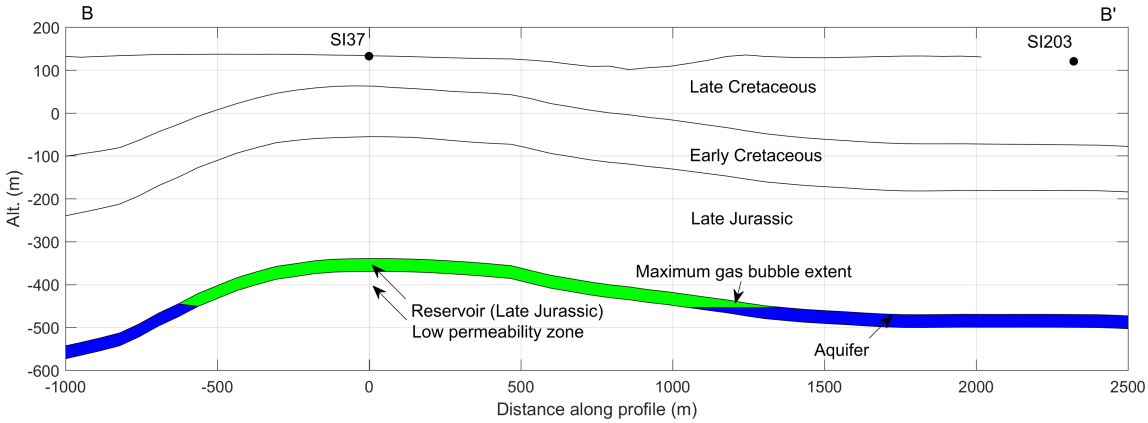


Figure 2.16: Geological profiles along the lines AA' and BB' shown in Fig. 2.15. The maximum extent of the gas bubble allowed by the spill point ($z = -460$ m) is shown in green. The aquifer is shown in blue. The ages of the main stratigraphic series are indicated (black), as well the main encountered geologic facies (red). Note the differences between the surface topography and the underlying structure.

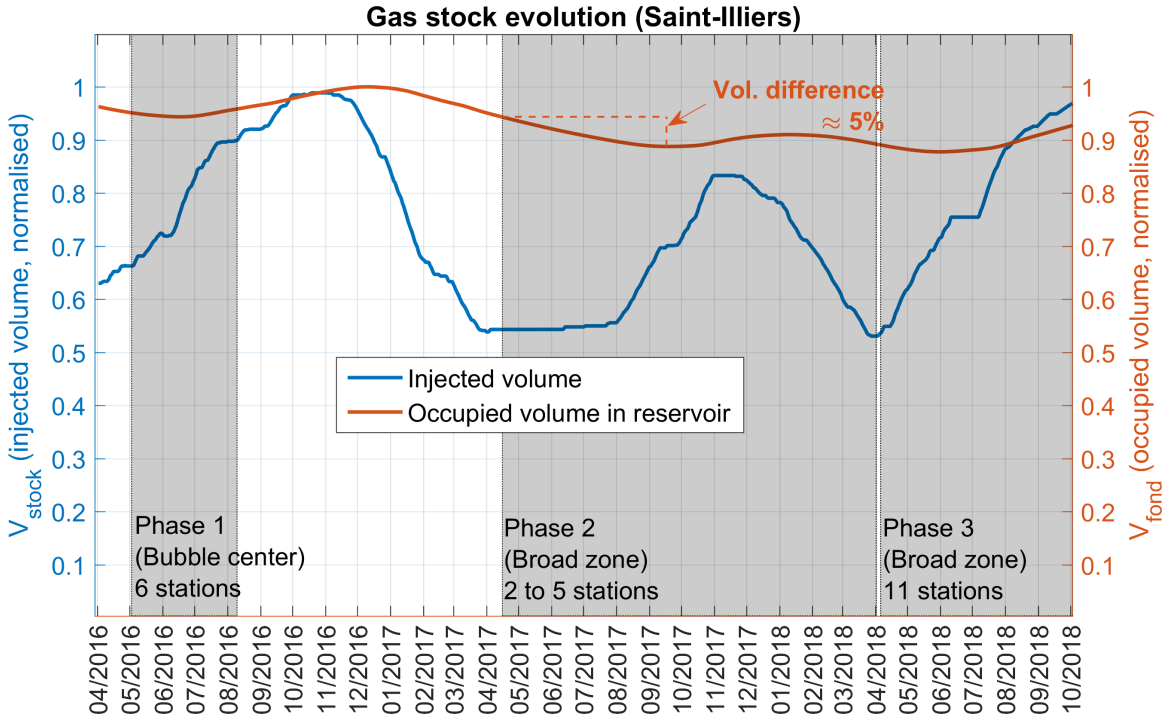


Figure 2.17: Evolution of the injected (blue) and occupied (red) gas volume in Saint-Illiers between 2016 and 2018. The grey zones show the different phases of the experiment. The maximum recorded V_{fond} difference during the experiment in shown in red (5%).

this phase did not present any anomalies due the installation inside a building, though it is not a recommended practice in seismology.

Phase 2 (April 2017 - April 2018)

The phase 2 started in April 2017 with 4 stations (blue points in Fig. 2.15), complemented by SI37 again installed in a building. Unfortunately, only two of the four installations were waterproof as expected (SI03 and SI11). SI29 was severely flooded in July 2018, after which the station was dismounted. SI09 also presented limited water inflows after heavy rains, so we also dismounted this station in October 2017. The phase 2 ended in April 2018 with three good quality stations, instead of the 5 initially planned. However, these three stations (SI03, SI11, SI37) were of particular interest, as SI03 was installed near a well never reached by the gas bubble, while SI11 was installed near a well periodically switching from gas to water saturation.

Phase 3 (April 2018 - last update January 2019)

For this phase, we were able to design a robust waterproof installation (exterior view in Fig. 2.18). 5 new Trillium 20s seismometers were provided by the *Ecole et Observatoire des Sciences de la Terre* (EOST), who now actively takes part into the network maintenance. This new acquisition with 11 stations was launched in early April 2018. It is planned to last at least two years (phase 3). Projected on the reservoir topography, this array covers the «smooth» northern flank of the structure

Figure 2.18: Seismometer deployment during phase 3

(Fig. 2.19).

Gas stock evolution

The gas stock evolution along with the occupied volume are plotted in Fig. 2.17 between 2016 and 2018. As in the case of Chémery, approximately half of the total volume is cycled every year, while the variation of the occupied volume does not exceed 15%. Only one full cycle was covered by the experiment (phase 2). In 2017, the injected volume was lower than usually, and the maximum occupied volume variation during the phase 2 was of 5%. We expect the phase 3 to cover a normal injection cycle with an increased number of seismometers and yield more conclusive results. Nevertheless, some interesting insights are provided by the already available data.

2.2.3 Spectral anomalies

Vertical component PSD

The same signal processing workflow as in Section 2.1.4 is applied to compute the PSD_z spectra, except we are using two-hour periods, instead of the four-hour periods in Section 2.1.4. This is done to obtain spectrograms with a better time resolution. In Fig. 2.20, we superpose the spectra recorded simultaneously by 9 stations during the phase 3. The line colours range from blue to red, inversely proportional to the distance of the corresponding station to SI37, which is located above the reservoir top (see Fig. 2.19). Thus the stations represented with cold colours are far from the gas bubble center. This convention is followed up in the remainder of the section dedicated to Saint-Illiers. In Fig. 2.20, we see that the central stations SI37 and SI09 exhibit a positive PSD_z anomaly roughly between 1.5 and 4.5 Hz. This anomaly is well observable both on Sunday (Figs 2.20a and c) and during the working week (Figs 2.20b and d), as well as during night time (Figs 2.20a and b) and day time (Figs 2.20c and d). The sign and the frequency band of the anomaly are consistent with the HMT spectral signature reported in literature, as well as with the results obtained in Chémery (Section 2.1). Regarding the geological profiles in Fig. 2.16, an important conclusion is that the surface topography is not, in this case, correlated with the PSD_z anomaly, contrary to what was reported by Hanssen & Bussat (2008).

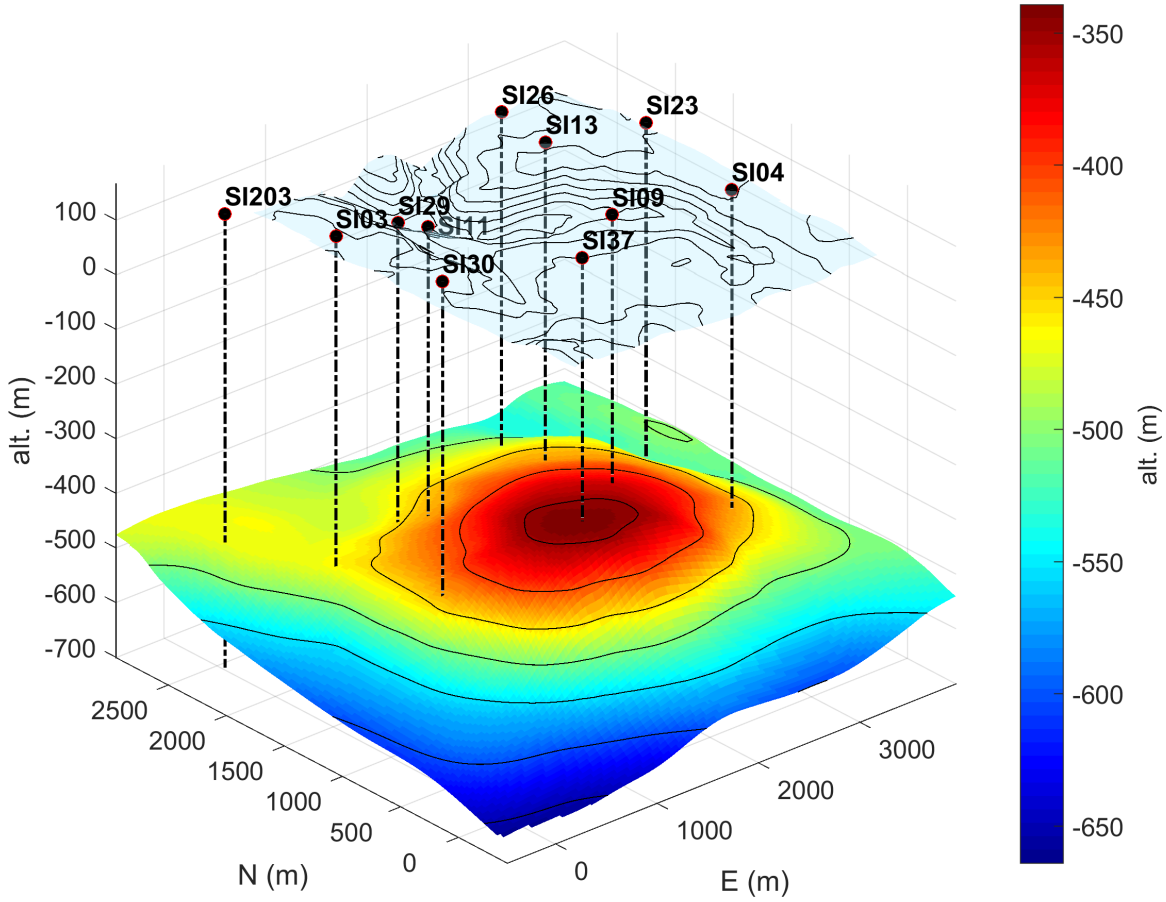


Figure 2.19: Seismometer positions (phase 3) projected on the reservoir top boundary in 3D view. The grey sheet is the surface topography.

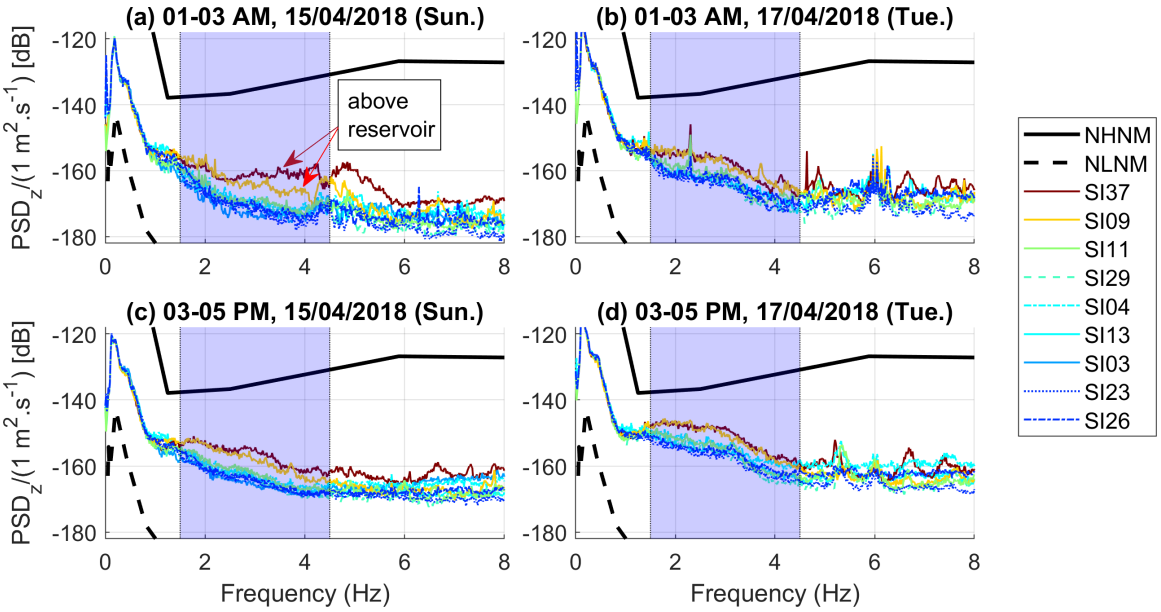


Figure 2.20: PSD_z spectra recorded simultaneously during the phase 3 at different time periods. (a) Week-end night-time; (b) Working day night-time; (c) Week-end day-time; (d) Working day day-time. The low and high noise models (Peterson, 1993) are shown in black. The colour scale is proportional to the distance from the reservoir structural top (beneath SI37). Red stations are closer to the top than blue stations

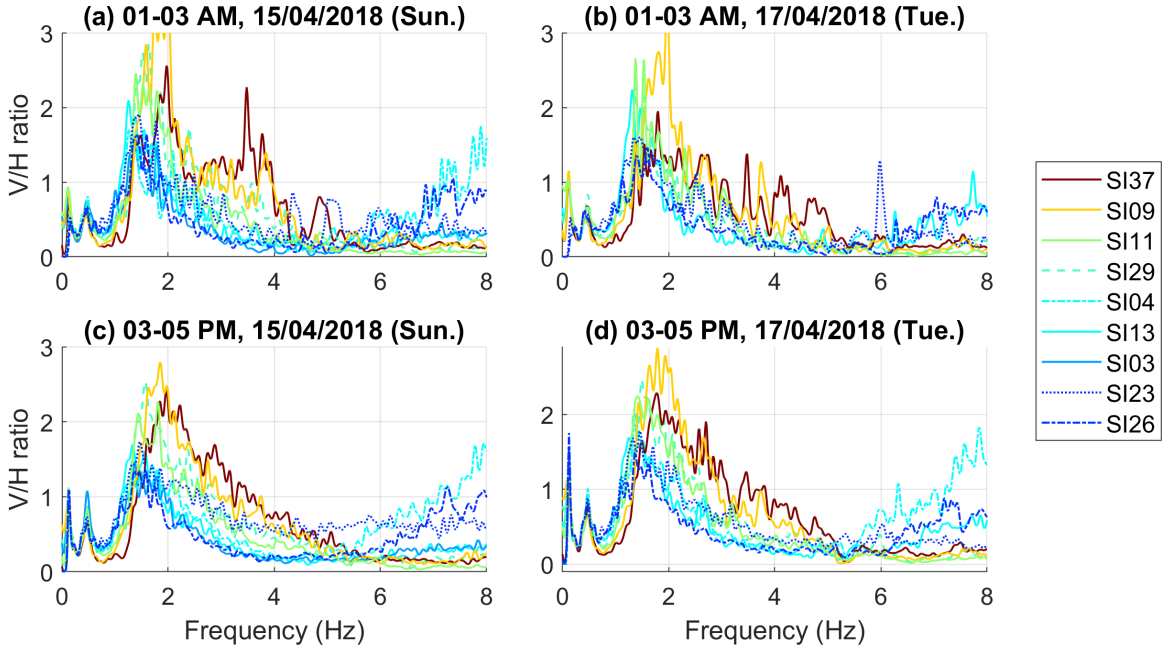


Figure 2.21: Same as Fig. 2.20 for the V/H spectral ratio.

V/H ratio

In Fig. 2.21, a plot similar to Fig. 2.20 is displayed for the V/H spectra. Again, a gradual transition from low to high V/H ratio values is observed as the stations get closer to the bubble center. However, the two central stations (SI37 and SI09) are less separated from the rest of the cloud, compared to the PSD_z spectra in Fig. 2.20. Some similarity can be observed between the V/H spectral ratios in Fig. 2.21 for Saint-Illiers and Fig. 2.4c for Chémery. For the Chémery dataset, we saw that the V/H increase within the frequency band of the PSD_z amplification was correlated neither with the geological structure, nor with the presence of gas. For this reason, the analysis of the Saint-Illiers dataset is restricted to PSD_z in our work.

2.2.4 Active or passive anomaly source ?

Spectrogram analysis

Extracts from PSD_z spectrograms are compared in Figs 2.22a and b, respectively for phase 1 and 3, between 1 and 5 Hz (i.e. an interval containing the attribute definition band). All of them present a characteristic daily and weekly variation associated with human activity. The site is indeed located 50 km away from Paris, 10 km away from the town of Mantes-la-Jolie (40.000 inhabitants), and 4 km away from the busy A13 highway. There is no stationary noise within the attribute definition band which could be attributed to machinery noise, contrary to the case of Woith et al. (2014). The industrial operator confirmed that no activity (such as traffic, working people, and other contaminating factors discussed by Hanssen & Bussat (2008)) took place at night, except in some specific periods. Such periods can be unambiguously distinguished on the spectrograms (see the night-time energy bursts highlighted by the black arrows in Fig. 2.22a). Thus we suggest that most of the energy is due to human source located *outside* the array. On the other hand, the stations located above the reservoir (all stations in Fig. 2.22a and only SI37-SI09 in Fig. 2.22b) exhibit a higher amplitude of day/night variation compared to the peripheral stations. Because the energy sources appear to be exterior to the array, we suggest to interpret these amplitude differences between the different stations as *site effects* (Bard, 1999).

Relative anomaly daily and weekly evolution

To confirm this hypothesis, we define the integral attribute $PSD_z|_{1.5}^{4.5}$ (equation 1.1), and plot the time series of its relative anomalies for all the stations available during the phase 3 (Summer 2018), taking the remote station SI29 as a reference (Fig. 2.23). The times series shown in each of the four plots were extracted during different time periods, which are the same as those for which we showed individual PSD spectra in Fig. 2.20: Sunday night-time (a), Sunday day-time (b), Tuesday night-time (c) and Tuesday day-time (d). We see that the strong positive anomaly ranging between 200% and 400% (3-6 dB), visible for SI37 and SI09, is observed during all the aforementioned time periods. On the other hand, the absolute PSD variations between the different time periods, given by the spectrograms in Fig. 2.22, are more than 10 dB. Thus if the anomaly on SI37 and SI09 was due to an independent source of given intensity, it would be periodically overwhelmed by the ambient noise, which is not observed

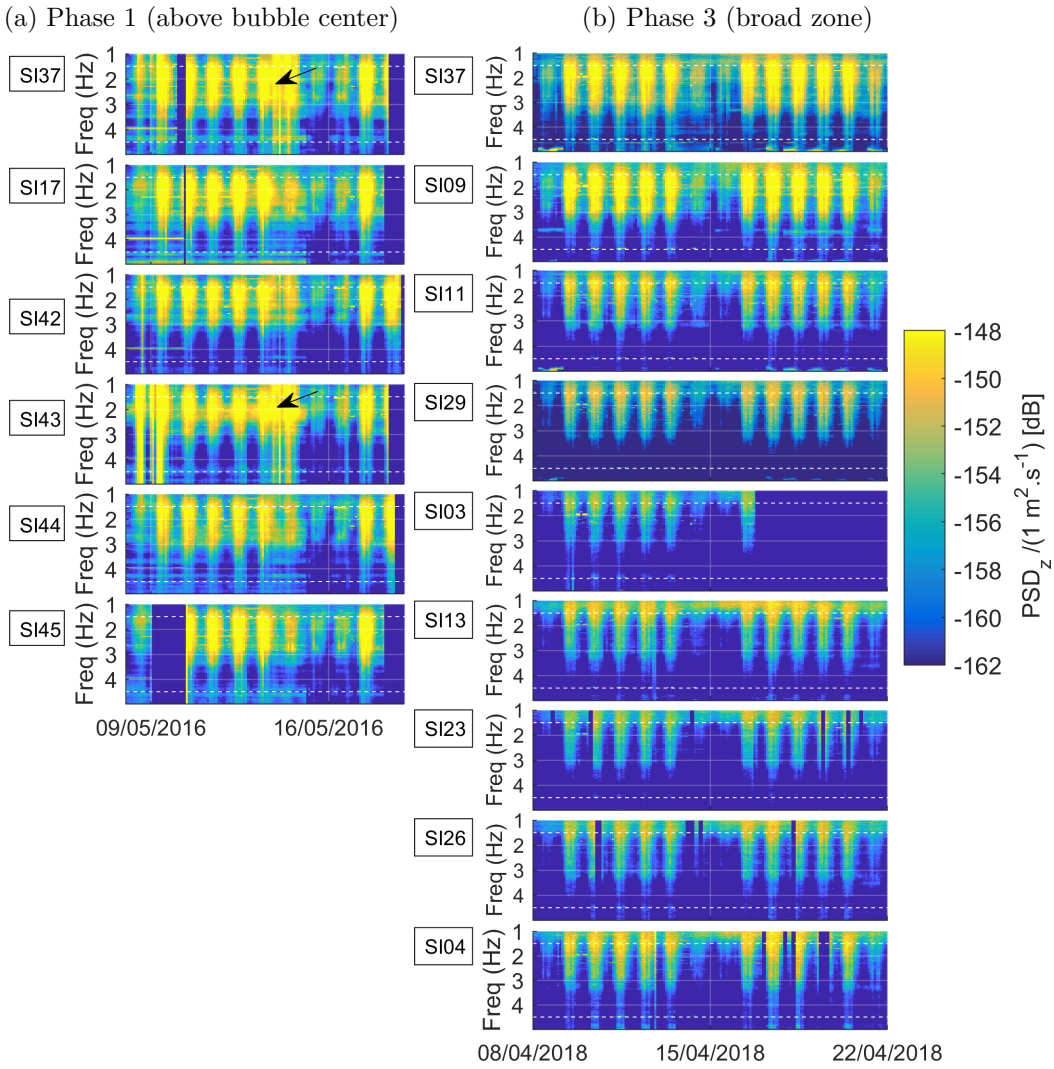


Figure 2.22: Spectrograms for recording extracts of phase 1 (a) and phase 3 (b) between 1 and 5 Hz. The 1.5-4.5 Hz band, used for the attribute definition, is shown by the white dotted lines. Note the larger day/night variation at the stations close to the reservoir top.

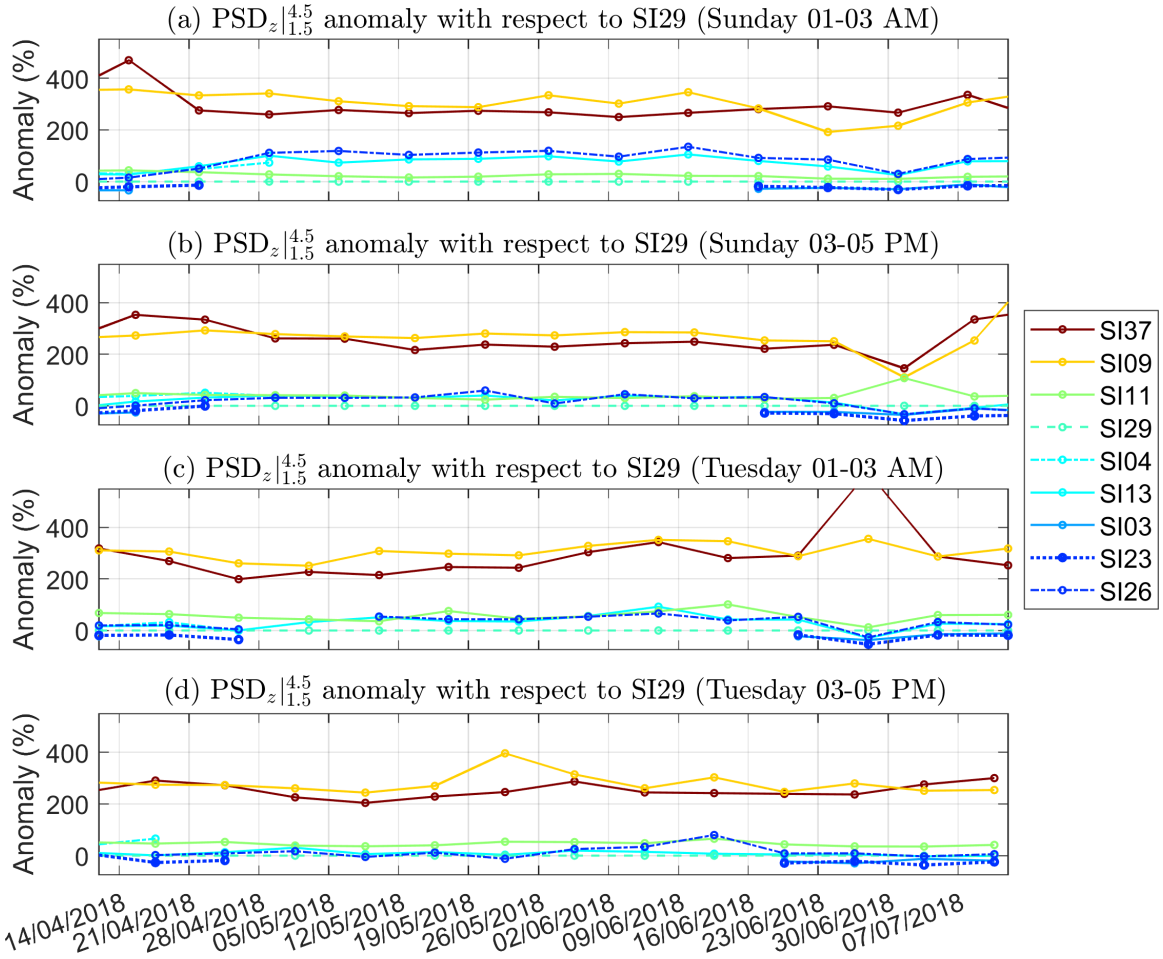


Figure 2.23: Evolution of the relative PSD$_z|_{1.5}^{4.5}$ anomaly with respect to SI29 (reference station) over three months of data in Summer 2018. (a) Week-end night-time; (b) Week-end day-time; (c) Working day night-time; (d) Working day day-time.

in Fig. 2.23. This confirms that the anomaly is not due to any inner source of active seismic emission within the array (neither human nor hydrocarbon-related), but rather to a different level amplification of incident waves. However, this site effect could be related to the presence of hydrocarbons, to the structural topography, or to a combination of both factors. Regarding the discussion in Section. 1.2.3, this confirms that passive scattering is the most plausible mechanism, *if* the site effect is indeed related to the presence of gas.

2.2.5 Attribute evolution vs. gas stock

Finally, we investigate if there exist any correlation between the gas stock and the attribute temporal evolution. The gas stock and the occupied volume are plotted in Fig. 2.24a (same as Fig. 2.17, but normalised between 0 and 1 in order to better see the evolution). In this part of the study, we use only Sunday night-time data (i.e. the quietest available period), though the previous paragraph suggests that the PSD$_z|_{1.5}^{4.5}$ anomaly is not specific to this time period.

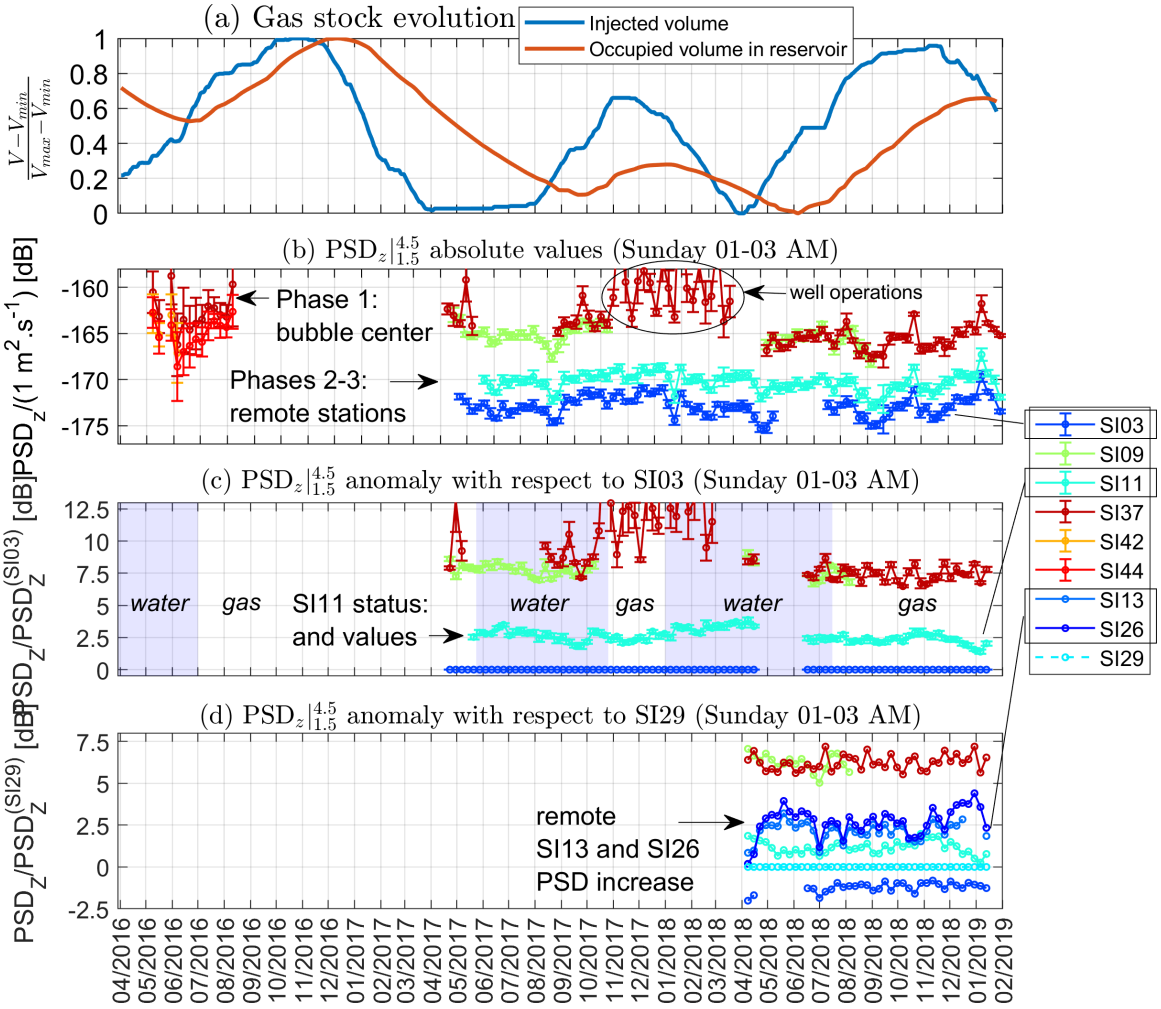


Figure 2.24: Evolution of the gas stock (a) compared to the attribute evolution: (b) absolute $PSD_z|_{1.5}^{4.5}$ attribute, (c) relative $PSD_z|_{1.5}^{4.5}$ anomaly with respect to SI03; (d) relative $PSD_z|_{1.5}^{4.5}$ anomaly with respect to SI29.

Absolute values of $\mathbf{PSD}_z|_{1.5}^{4.5}$

In Fig. 2.24b, we plot the absolute values of the attribute in decibel. During the phase 1, all of the stations exhibit similarly high attribute values, consistently with the fact they were all installed close to each other above the reservoir top. During the phases 2 and 3, the remote stations are 2-8 dB below the central SI09 and SI37.

Relative attribute anomalies

In Fig. 2.24c, we plot the relative attribute anomalies with respect to the remote SI03 station, constantly available during the phase 2. Based on well observations, we know that gas never reached this station during the observation period. On the other hand, SI11 was episodically reached by the gas bubble, which is indicated by the white stripes in Fig. 2.24c. Any clear signature is observed neither for gas arrival in the well SI11 in late October 2017, nor for the gas departure in January 2018, nor the new gas arrival in July 2018. This means that HMT are probably unusable as a real-time direct indicator of gas presence beneath the sensor in a monitoring perspective. However, the relative anomaly of SI11 (cyan curve in Fig. 2.24c) varies significantly during the observation period. For instance, it seems to exhibit some anticorrelation with the gas volume in the reservoir (red curve in Fig. 2.24a). This behaviour could be related to the gas bubble dynamics, but several full cycles of gas injection/withdrawal have to be observed to better understand the possible correlation. The evolution of the ground water table height should also be considered as a possible alternative factor influencing the ambient noise characteristics (Sens-Schönfelder & Wegler, 2006; Riahi et al., 2013a).

Because of the SI03 outage at the beginning of the phase 3, due to a memory card issue, we used the remote SI29 station as a reference to analyse this time period (Fig. 2.24d). It is interesting to observe the increase of the anomaly for the remote SI13 and SI26 stations, which starts simultaneously with the gas injection in April 2018. The anomaly seen by the central stations, on the contrary, remains stationary. SI13 and SI26 are located in the heart of the Rosny forest, so human influence on Sunday during night-time is unlikely. Once again, we cannot yet conclude on the origin of this increase, because of a too short observation period.

2.2.6 Main conclusions

The analysis of Saint-Illiers data brings some important complements to the Chémery data presented in Section 2.1. First, a similar effect is observed above the gas reservoir in terms of vertical component motion amplification. Second, the possibility of an active emission mechanism in the reservoir, uncorrelated with the background noise *at the same frequencies*, is eliminated through relative anomaly time series analysis. The hypothesis of human sources on the site is eliminated by the same arguments. Thus the mechanism of the anomaly generation is due to a passive scattering effect by either the reservoir or the surrounding geological structure (site effect). Third, the surface topography as a possible reason of the amplitude anomaly has also been eliminated. Fourth, even if the amplitude of the anomaly is *related* to the gas bubble in the reservoir, it cannot be used as a direct indicator of gas presence beneath the

receiver, as we have seen by comparing the data from SI11 and SI03.

As the ambient wavefield on the vertical component is likely to be dominated by Rayleigh waves at the frequency of the anomaly (Bonnefoy-Claudet et al., 2006b), the first mechanism to investigate is the Rayleigh-wave scattering by embedded obstacles, which can be either structural heterogeneities or fluid-saturated reservoirs. *This is the main subject of the modelling part of this PhD work.* Theoretical considerations on this topic start in Chapter 5. Before that, in the next section, we present the results of a very simple small scale experiment which demonstrates how the fluid saturation in a buried reservoir can affect the Rayleigh wave amplitude.

2.3 Small-scale field experiment

In our experiment, we focus on the amplitude variation of the main surface wave arrival above a scatterer with continuously changing properties (water saturation in unconsolidated sand). A buried plastic box initially filled with dry sand is continuously saturated with water. Sledgehammer shots are repeated at different levels of saturation. After a deconvolution step aiming to correct for the variability of the source amplitude, the effect of saturation on the surface waves amplitude is investigated.

2.3.1 Experimental setup

A line of 48 geophones (vertical component only) designed for active seismic experiments was deployed in a sand dune (forest of Fontainebleau) with a regular spacing of 0.5 m (Fig. 2.25a), . Each acquisition consisted of 5 sledgehammer strikes at the source position $x_S = 34$ m. The source position was chosen so that well separated surface waves could be observed at the receivers. Two plastic boxes filled with dry sand, of dimensions L80-W60-H15 cm (Fig. 2.25b), were buried next to each other beneath the central receiver (nr. 24). Their depth was of $h = 0.65$ m, measured between the boxes and the surface. A reference acquisition was performed. Then water was progressively pored through the tubes inside the boxes (Fig. 2.25c) until the maximum level of saturation was reached (100%), with a new acquisition performed every time the saturation increased by 20%. The maximum water level was roughly estimated based on the hypothesis of a 30% porosity for the sand, yielding a total water volume of 40 L. The final excavation at the end of the experiment showed that the sand inside the boxes became fully saturated, confirming that this hypothesis was acceptable.

2.3.2 Processing and results

Sledgehammer strikes performed by a human, even after stacking, are not identical from one shot to another. The empirically estimated variable source wavelet must be deconvolved from the signal prior to the amplitude analysis. This operation transforms the raw signal $s(x,t)$ into a normalised signal $\bar{s}(x,t)$. We use the following deconvolution procedure:

(a) Geophone line (b) Burried box (c) Water injection

(d) Acquisition configuration

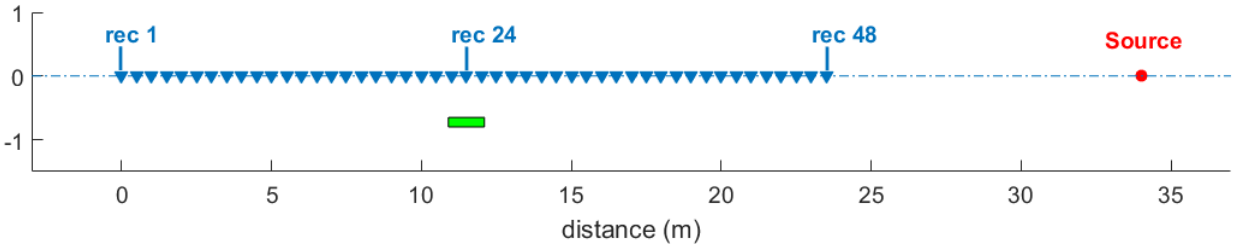


Figure 2.25: Experimental setup: (a) View of the geophone line. (b) Box filled with dry sand before excavation burial. (c) Continuous saturation of the burried box with water. (d) Scheme of the acquisition configuration. Green zone: burried boxes.

1. The amplitude correction coefficient $\sqrt{x_i - x_s}$ is applied to each receiver i to compensate for the geometrical spreading.

2. The closest receiver to the source ($x = 23.5$ m) is chosen as a reference station. We consider the Fourier-transformed signal $\psi(\omega)$ recorded during the reference shot (0% saturation) on the reference station as a proxy for the source wavelet («reference trace» in the frequency domain).

3. Each studied shot $u(t, x)$ is Fourier-transformed and deconvolved by the signal recorded at the reference station for this shot, and then convolved by the reference trace, at each receiver i:

$$\bar{u}(\omega, x_i) = \frac{u(\omega, x_i)u(\omega, x = 23.5\text{m})^*}{|u(\omega, x = 23.5\text{m})|^2 + \epsilon}\psi(\omega), \tag{2.8}$$

   where * stands for the complex conjugate, and $\epsilon = 0.001 \cdot \max |u(\omega, x = 23.5\text{m})|^2$ is a water-level parameter. The normalized waveform $\bar{u}(t, x_i)$ is then obtained via inverse Fourier transform.

4. Differential shot gathers are computed the for each studied shot in order to isolate the effect of the saturation change:

$$\Delta\bar{u}(x, t) = \bar{u}(x, t) - \bar{u}_{\text{ref}}(x, t) \ , \tag{2.9}$$

   where the subscript ref stands for the reference shot at 0% saturation. These shot gathers are shown in figure 2.26, left column.

5. The normalized PSD difference is computed with respect to the maximum PSD value at the reference station during the reference shot (0% saturation):

$$d_{obs}^{\text{EXP.}} = \frac{\Delta\text{PSD}}{\max \text{PSD}} = \frac{\text{PSD}(x_i, \omega) - \text{PSD}_{\text{ref}}(x_i, \omega)}{\max |\psi(\omega)|^2} \tag{2.10}$$

   This is different from the relative anomaly defined in equation 2.5 and used to process the real ambient noise. The problem with single-shot active data is that the associated spectra exhibit a strong comb effect (peaks and notches), which greatly complicates the normalisation by a reference station. The normalized PSD anomalies in the $x - f$ space are shown in Fig. 2.26, right column.

From Fig. 2.26 it is clear that the major part of time-lapse change in both space-time and space-frequency domains is localised above the buried boxes, except at short offsets. The spectral anomaly above the box is positive and located roughly between 30 and 70 Hz. Its maximum value is about 15%, which is not directly comparable to the values obtained in the previous sections because of a different normalisation procedure. Other differences in the spectral content, visible at the right end of the array, are due to the imperfect deconvolution by the reference trace. The latter do not consistently change with water saturation. As a conclusion, growing water-saturation of a buried plastic box filled with sand generates growing amplification of the Rayleigh-wave dominated incident wavefield in a limited frequency band. The precise nature of this interaction is to be investigated in the following, starting with an elastic scattering model.

In the next section, we present the results of a third reservoir-scale test performed during this PhD work, this time in the geothermal context.

2.4 Dataset 3: Muara Laboh high-enthalpy geothermal field

This section appears almost identical to the conference paper presented at the 5th International Indonesian Geothermal Congress and Exhibition (IIGCE) in 2017:
Kazantsev, A., Egermann, P., Ramadhan, I., Huguet, F., Formento, C., Peruzzetto, M., Chauris, H., & Métaxian, J.-P., 2017b. An innovative methodology based on Low Frequency Passive Seismic data analysis to map geothermal reservoir steam saturated areas, in *5th International Indonesia Geothermal Conference and Exhibition (IIGCE)*, Jakarta, Indonesia.

While the results are interesting from an operational point of view, the signature of a geothermal reservoir appears quite different from what was observed in the previous sections on the UGS sites. Instead of the amplification of the vertical component, we observe an attenuation of all the components. Since the attenuation for the horizontal components is stronger than for the vertical one, the resulting signature is an increase of the V/H ratio. In fact, this is similar to the simulation results of Lambert et al. (2013). Compared to the gas storage context, higher fluid-related attenuation may be expected in the geothermal context because of a higher matrix heterogeneity (fracture

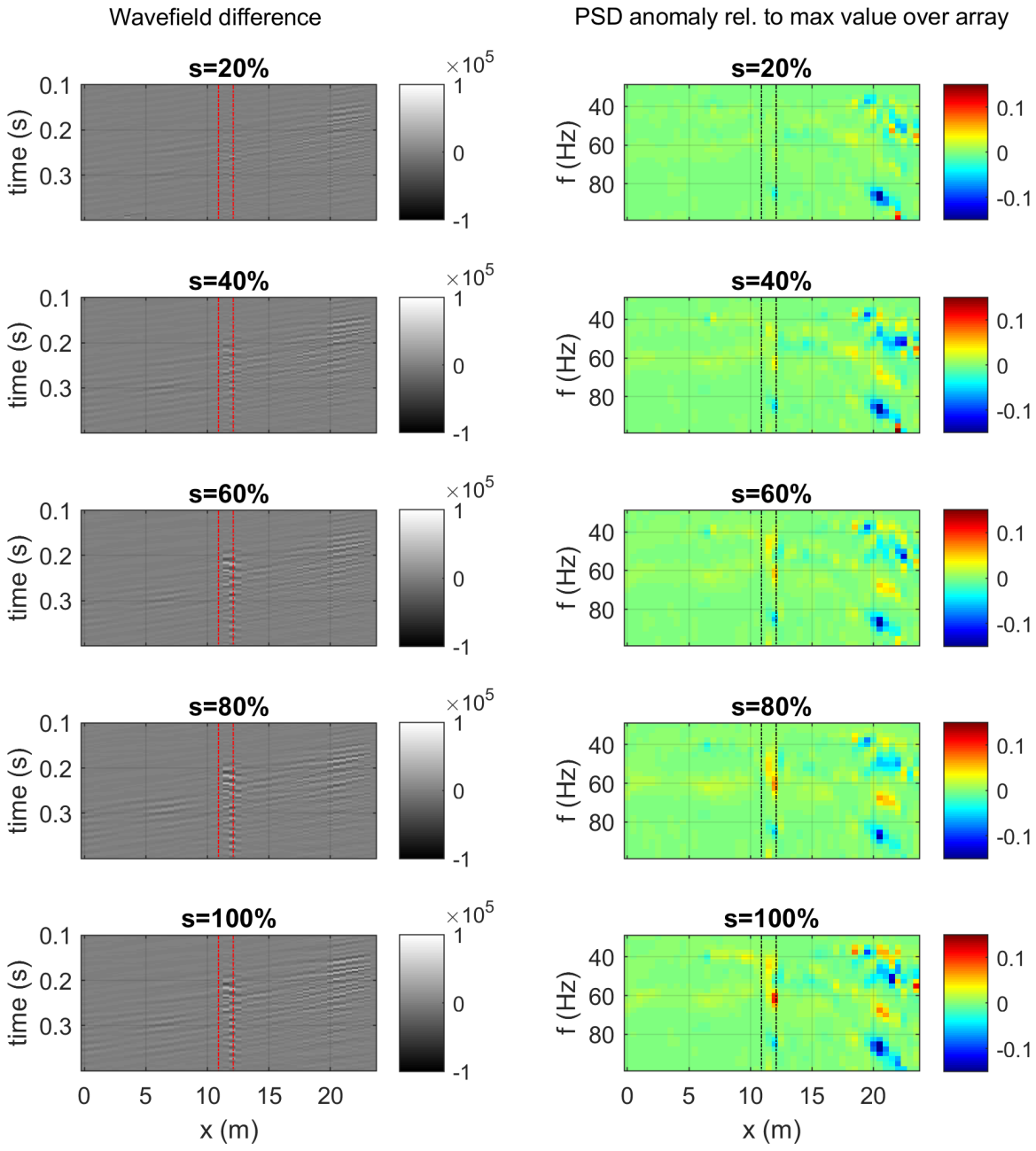


Figure 2.26: Experiment data after deconvolution procedure. Left column: difference between the shotgathers for increasing levels of saturation and the reference shotgather. Right column : difference between the PSD computed for the shots at increasing levels of saturation and the reference shot, normalized to the maximum PSD recorded by the sensor nearest to source. The dashed vertical lines represent the limits of the excavation.

porosity), and a thicker fluid system (Grab et al., 2017b), which gives more validity to the simulation parameters used by Lambert et al. (2013). Nevertheless, since I do not address anelastic effects in this PhD work, this dataset did not guide the modelling works and can be regarded as a complement, to be analysed in the future.

Summary

In the context of exploration, development and monitoring, it is highly valuable to have access to the steam saturated regions of a geothermal resource. During exploration phase, it consists of a strong evidence that a sufficiently hot system is still present before the drilling decision. Then it allows to identify the best locations for the future exploration and development wells. During the production phase, it is important to monitor the steam saturated area extension due to the field depletion (production induced reservoir pressure decrease) in order to optimize the positioning of the make-up wells. This work presents an innovative methodology to detect relevant reservoir steam saturated areas which can be applied at different stages of high enthalpy geothermal projects. A field survey was recently conducted in the Muara Laboh geothermal field. The obtained results were in complete agreement with observations made on all the exploration wells. They enable to map the extension of the potential steam saturated region at the top of the structure and also to identify another potential steam area associated with a secondary top. The identified steam cap extension remains to be confirmed by the development wells which will be drilled and tested during the course of this year. These preliminary results have demonstrated the applicability and the potential of this promising new steam detection method for high enthalpy geothermal resources.

2.4.1 Context and introduction

A new reservoir characterization technique based on the interpretation of ambient, natural, low frequency passive seismic signal meets currently a growing success to detect subsurface reservoirs saturated with multiphase fluids. It has been used in various domains such as hydrocarbon detection and volcanic eruption prediction (Ferrick et al., 1982; Dangel et al., 2003). This technique was also successfully evaluated by Storengy on a natural gas storage asset using an advanced and innovative workflow developed internally and patented (Huguet et al., 2017). As the spectral anomaly is related to the presence of several coexisting phases in a reservoir and the basic properties of a gas/water and steam/brine system are of the same order in terms of surface tension, density and viscosity, there is no fundamental reasons why this technique could not be applicable also in the geothermal context to detect steam saturated areas.

An accurate mapping of geothermal reservoir two-phase regions is of primary importance to guide the targeting of development and even exploration wells. On a longer term, it is also very important to map the extension of the steam area related to the reservoir depletion in order to guide the drilling of future make-up wells. Unfortunately, mapping two-phase regions is usually difficult in geothermal fields since the top reservoir is picked from the MT data whose vertical resolution is low, even

though the steam/water interface is identified in some exploration wells. Therefore, it is highly valuable to develop new techniques to access this information directly in an accurate and cost effective manner. This is the main objective of this paper.

2.4.2 Acquisition over a geothermal site

The LFPS methodology has been tested on the Muara Laboh geothermal field (West Sumatra) which was extensively described by Situmorang et al. (2016). PT Supreme Energy Muara Laboh (SEML) drilled six full diameter exploration wells (Fig. 2.27a) which yielded an estimation of temperature distributions, resource size, permeable structures, and hydrology of the system. The conceptual model relies on a hot geothermal fluid upwelling from the deep part of the resource in the southern part (Fig. 2.27b). The fluid then flows globally to the northern direction. The reservoir fluid temperature ranges between 200°C to 310°C. In most areas, the reservoir contains a single liquid phase, but a small extension steam cap was detected at the shallowest part of the reservoir from wellpad A, as shown in Fig. 2.27b.

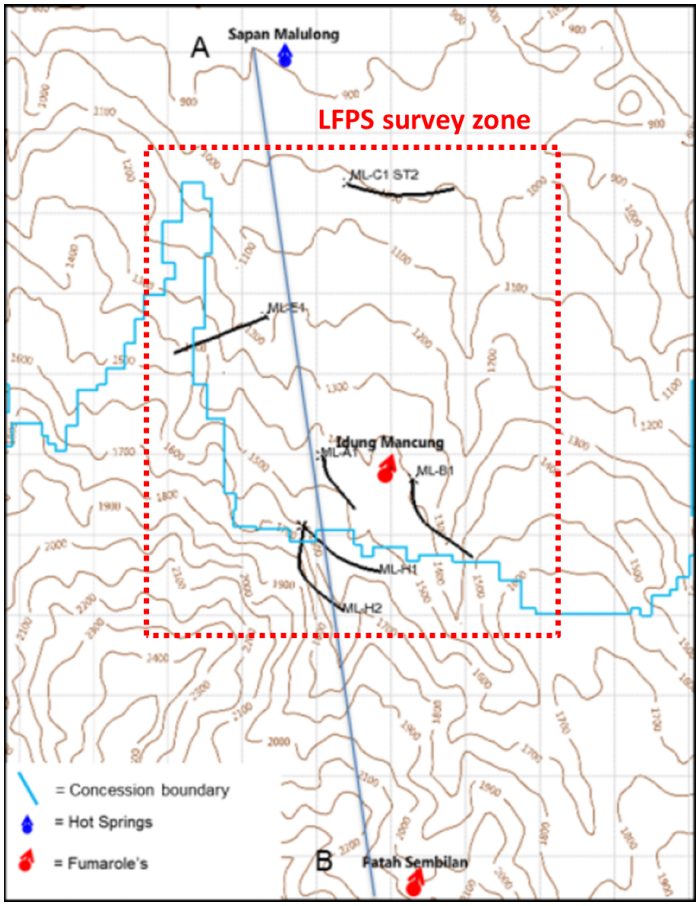
11 low frequency 3-component broadband seismometers (Güralp 6TD 30s) were deployed from 31/05/2016 until 14/06/2016. Three of them (REF01, REF02, REF03) remained at permanent locations. Another 7 stations were deployed in a circular configuration whose location changed every two days. This configuration enables the array processing of the ambient noise in order to characterize the incident wavefield. The network is shown in Fig. 2.28. Another permanent station (REF04) was added on 15/06/2016. The permanent stations remained installed until 14/08/2016 in order to check the stability of the measured attributes.

2.4.3 Results

After the same pre-processing as described in Section 2.1.4, the power spectral density (PSD) was estimated on all the 3 components. An example is shown in Fig. 2.29. The relative amplitude anomaly maps with respect to REF02 are shown in Fig. 2.30. It appears that in the southern part of the survey zone, where steam was encountered, the horizontal components exhibit strong attenuation between 1 and 2 Hz.

The V/H spectral ratio was calculated as in equation (2.4). Following the GTM method described in Section 2.1.7, an automated classification was applied to the recorded V/H spectral curves according to their shape. The classification output is quasi-continuous between -1 and 1. The spectra with values close to 1 exhibit a stronger V/H ratio between 1 and 2.5 Hz (Fig. 2.31). The evolution of the mean V/H ratio between 1 and 2.5 Hz recorded by the reference stations during two months following the survey is shown in Fig. 2.32. It can be seen that the high V/H ratio at REF01 is a stable feature over time. The fact the value of the mean V/H ratio presents important fluctuations only at the anomalous location REF01 is intriguing, as it could be related to an active circulation of geothermal fluids (boiling in the vicinity of the steam cap).
The classification results were then mapped onto the surveyed area by kriging, taking

(a) Topographical map, hot springs, fumaroles, and wells



(b) Conceptual 2D profile along the N-S line in (a)

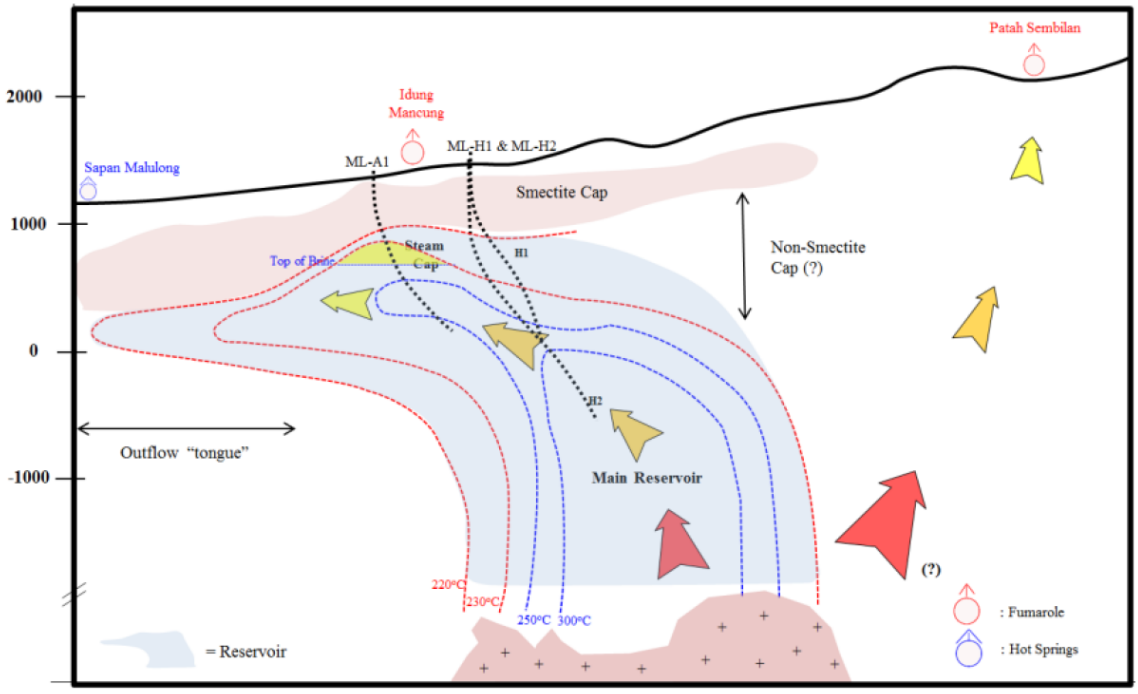


Figure 2.27: Muara Laboh geothermal site after (Situmorang et al., 2016). (a) Map view; (b) Conceptual model along a N-S section.

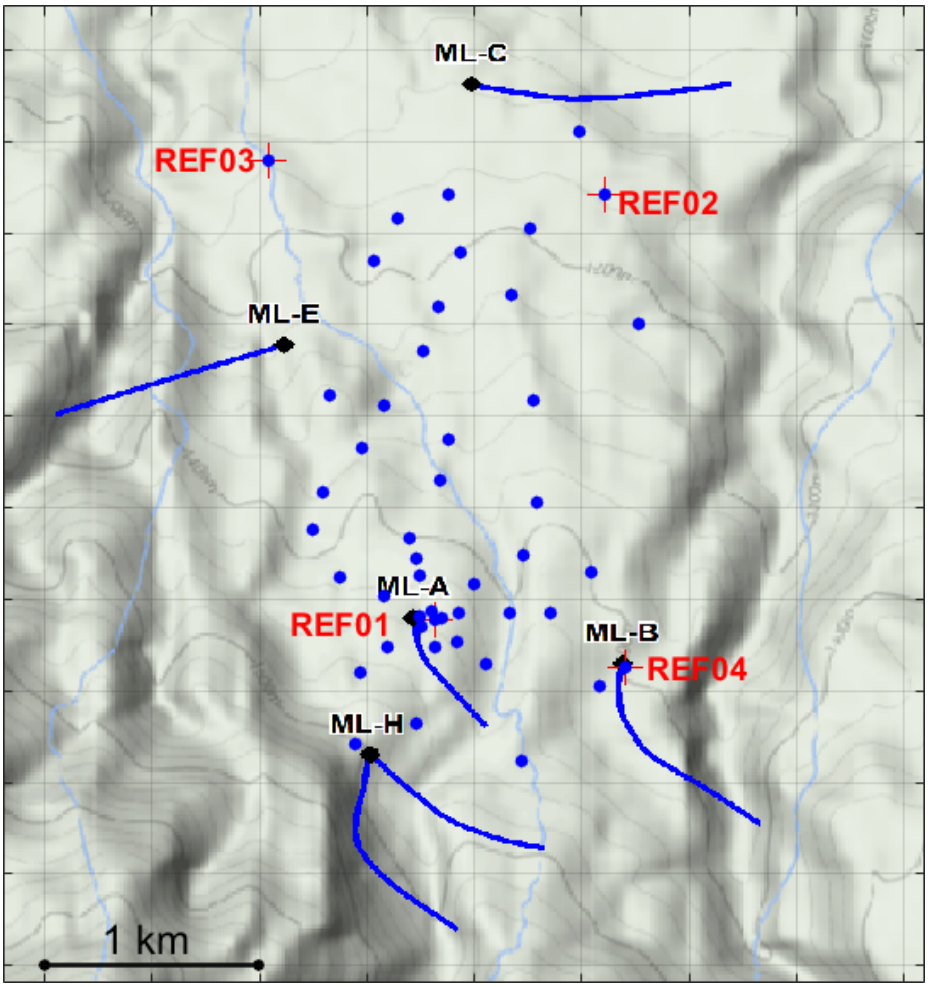


Figure 2.28: Broadband seismic network deployed in Muara Laboh. Each blue location corresponds to 48 hours of available recording. The red stations remained at fix locations during the survey.

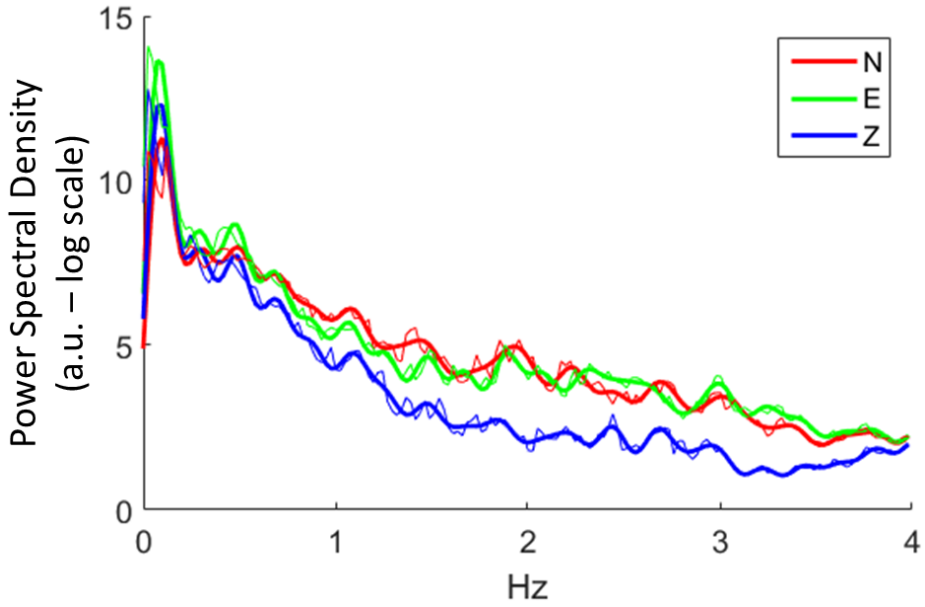


Figure 2.29: Example of spectra recorded at REF03, 03/06/2016. Bold curve : smoothed spectra.

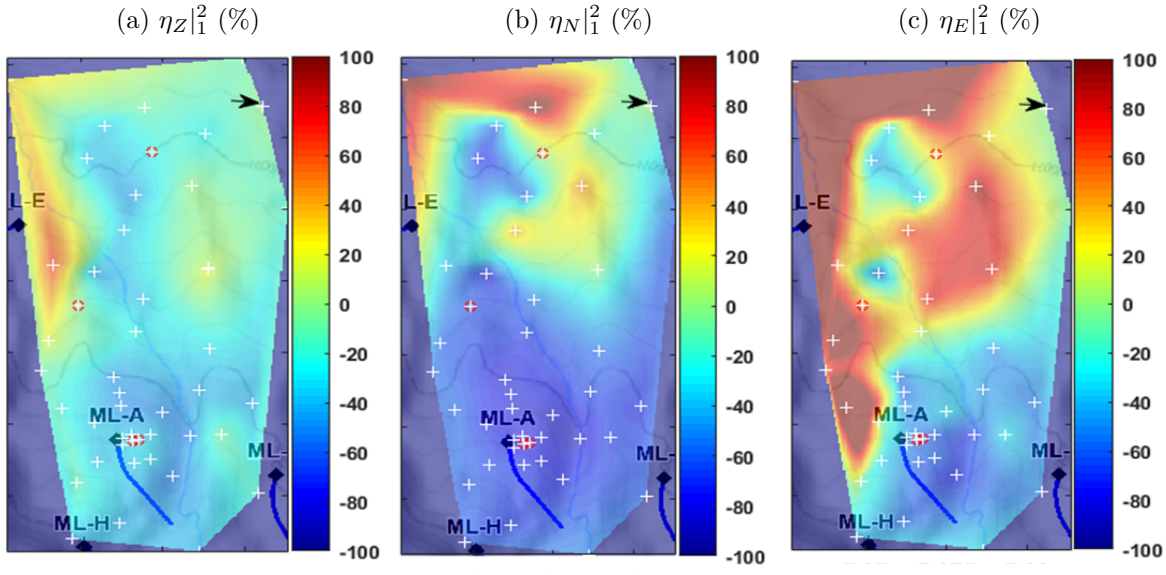


Figure 2.30: Relative PSD anomaly maps (see equation (2.5) for the definition) on the three components, between 1 and 2 Hz, with respect to REF02 (reference station shown with the black arrow).

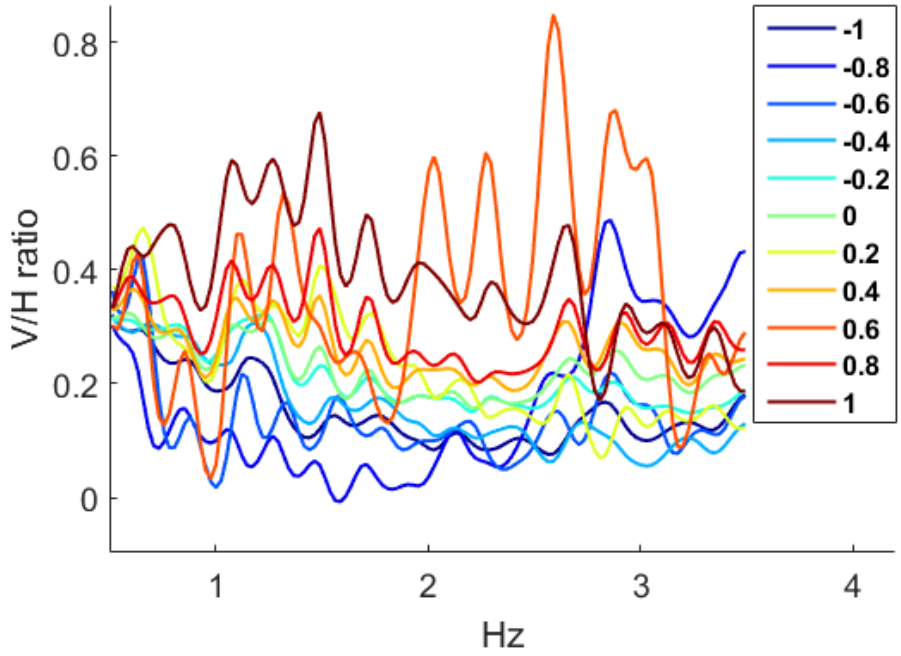


Figure 2.31: Result of the V/H spectral curves classification.

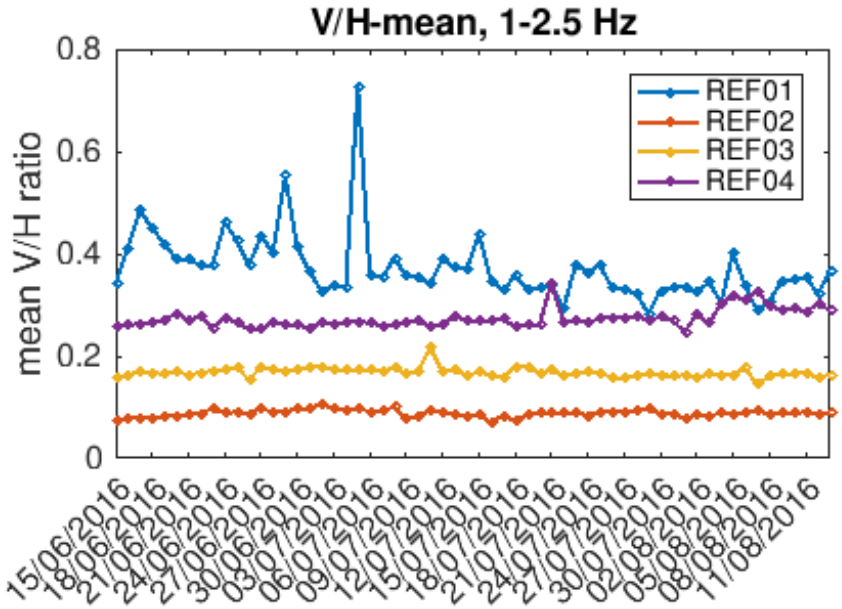


Figure 2.32: Evolution of the mean V/H ratio between 1 and 2.5 Hz over a two-month recording period.

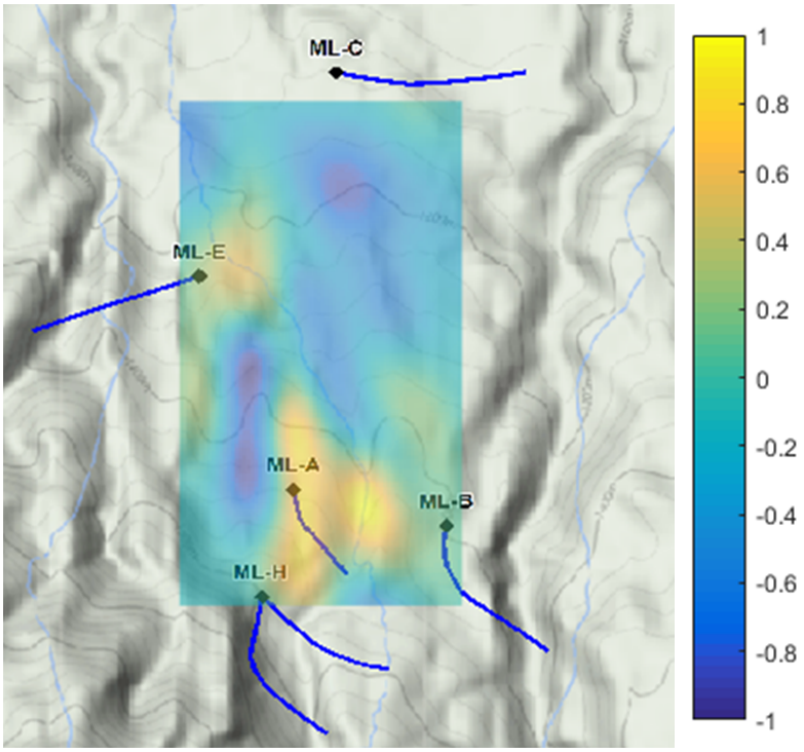


Figure 2.33: Map of GTM classes after classification and kriging.

into account the uncertainties related to the time variability of the spectra (Figure 2.33). This map shows two «anomalous» zones characterized by a high V/H ratio (classification output close to 1): a stronger one in the south of the field close to the well A, and a weaker one in the eastern vicinity of the well E.

2.4.4 Wavefield composition

Our current hypothesis is that the anomaly zones are generated due to the scattering of the ambient wavefield by the steam bearing zones. The ambient wavefield was analysed with two arrays of different apertures in 2 frequency bands : [1-2] Hz and [5-7.5] Hz (see Figure 2.34 for configuration and response functions). MUSIC algorithm (Schmidt, 1986) was applied to retrieve histograms of apparent slowness and back-azimuth over overlapping 5 minutes time windows, during the whole available recording time. The results displayed in Fig. 2.35 suggest that there is no consistent velocity decrease with frequency in the [1-2] Hz band (array 2) contrary to what is expected for surface waves. Moreover, the values of the apparent velocities range from 1500 to 3000 m/s, which is more typical for horizontally travelling body waves. The results of the array 1 exhibit velocities between 1000 and 2000 m/s into the [5-7.5] Hz band, which again seems too high for surface waves, though a decrease over frequencies is observed. Finally, the back–azimuth is stable over frequencies and is close to the Kerinci volcano direction, as shown in the back-azimuth histograms in Fig. 2.35. This leads us to conclude that the wavefield in the frequency band of interest is dominated by body waves originating from volcanic tremors. The next step of the work should consist of modelling the scattering of such waves by a contrast with properties relevant for a steam-bearing zone.

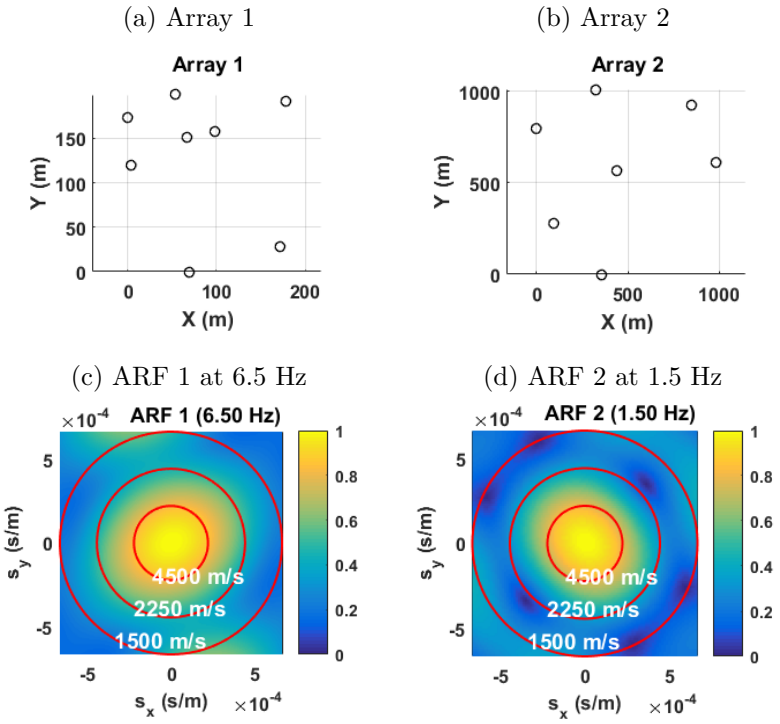


Figure 2.34: Geometry of the two arrays (top) and their response functions at the relevant frequencies (bottom)

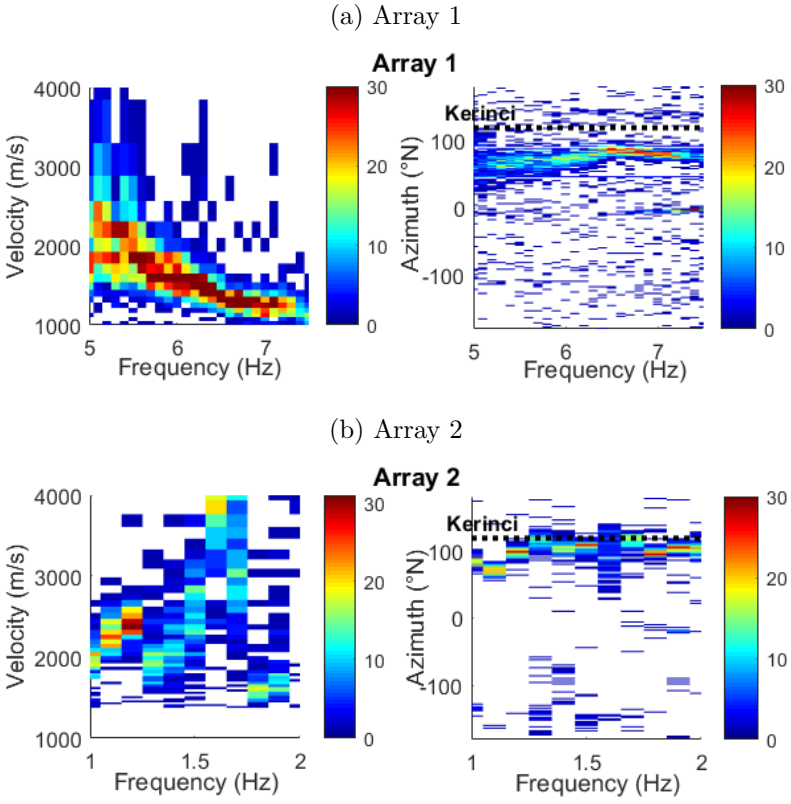


Figure 2.35: MUSIC results: velocity-frequency and back-azimuth histograms over the analysed time windows

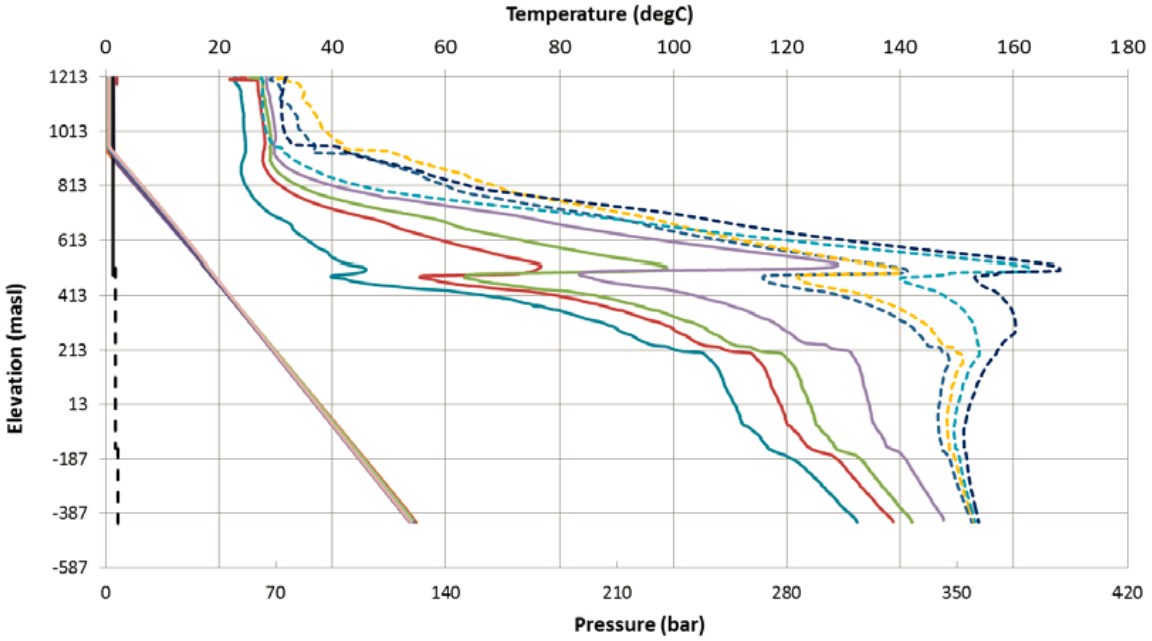


Figure 2.36: Anomaly in temperature recovery behind casing observed on E1 well that could be attributed to shallow steam.

2.4.5 Discussion

The results obtained by the empirical LFPS approach presented above are in complete agreement with the observations made on all the exploration wells (Fig. 2.36). They enable to map the extension of two steam saturated regions :

- The main one at the top of the structure (A wellpad) were the most prolific exploration well have been drilled with a steam production under bottom hole conditions.
- In the northern part near the area of wellpad E. This latest steam anomaly is less intense and could be associated with a shallow steam zone. This is suggested by the heating-up temperature profiles recorded after drilling which are showing a significant temperature anomaly behind the casing at shallow depth (Fig. 2.36).

Although the identified steam cap extension remains to be confirmed by development wells in the course to be drilled and tested this year, these first results are very encouraging and consistent with the current knowledge of this geothermal reservoir.

2.4.6 Main conclusions

A new methodology based on the acquisition and the analysis of Low Frequency Passive Seismic (LFPS) data has been successfully tested on the Muara Laboh geothermal reservoir to map the steam saturated areas. The protocol for data acquisition is simple and the survey can be conducted over a short period of time which makes it a cost effective solution. The data analysis and interpretation consist of the key part to obtain accurate energy spectra and relevant attributes representative of the spectral anomaly modification due to steam saturated reservoir areas. The first results are

very promising since already consistent with the current knowledge of the reservoir. The ongoing field development drilling campaign should enable to strengthen the accuracy of these preliminary results. The application of this new steam saturated areas detection method for high enthalpy geothermal resources is very promising for both the exploration and field development phases. It may improve the success ratio of the wells through a better targeting of the most prolific areas.

2.5 General conclusion on the field data

Three datasets were analysed in terms of hydrocarbon microtremor (HMT) attributes. Two of them were recorded above underground gas storage (UGS) facilities operated by Storengy in France. The third dataset was recorded above a steam-reach geothermal field in Indonesia. The first dataset consists of a dense array of sensors recording during a short time period. These data were used to visually investigate the spatial correlation between the HMT attributes, the gas saturation distribution, as well as the known geological structure of the site. The amplification of the vertical component power spectral density (PSD_z) appeared to be the most prominent attribute correlated with the presence of gas. However, the total PSD_z amplification (up to 2-3 dB) is likely a combined effect of structural topography and gas presence. A methodology of automated multi-attribute spectrum classification was introduced for constructing less user-dependent anomaly maps. The second dataset is a long-time recording by a few seismometers. Once again, a strong PSD_z amplification was observed above the reservoir (up to 8 dB). The analysis of the temporal evolution of the amplification demonstrated that this effect is not due to human sources on the site. This data also showed that the PSD_z anomaly is not related to the surface topography, contrary to the observations made by Hanssen & Bussat (2008). However, the relative contributions of the geological structure and of the fluid presence are still unclear, as the variation of the gas volume *in the reservoir* (not to be confused with the injected volume) during the observation period was only of 5%. The arrival of the gas front to an individual well did not trigger any immediately detectable change of the PSD_z spectrum. However, overall changes of the amplitudes recorded by the network *could* be sensitive to the gas volume evolution, though this has to be confirmed by a longer observation period. As the analysis of the first two datasets points towards a passive scattering mechanism behind the MHT, a small-scale field experiment was performed to investigate the influence of fluid saturation in a buried reservoir on an incident Rayleigh wave amplitude. A detectable modification of the spectral pattern above the reservoir was observed, motivating us for a further investigation of the Rayleigh wave scattering as a possible mechanism. Finally, the geothermal dataset in Indonesia revealed a different spectral signature above a steam-saturated reservoir: a V/H ratio increase through a preferential attenuation of the horizontal components. We applied our automated classification procedure to map the anomaly zone, which was in good correlation with the known steam zone. However, *I do not further investigate this dataset in this PhD work, as I focus on the PSD_z amplification modelling.*

The modelling strategy for PSD_z amplification was discribed in Fig. (1.22). I choose the Chémery site as the first to be modelled, because of the large amount of available data, and a very good knowledge of the geological structure due to 2D/3D reflection seismic data. The two other sites presented in this Chapter are not addressed in the manuscript from a modelling perspective. The first question I address is the ambient noise composition in Chémery (Chapter 3). Subsequently, this composition is used to define the source term in the numerical modelling of amplitude distortions (Chapter 4). Finally, Chapters 5 and 6 investigate the potential of inverting the modelled amplitude perturbations.

Chapter 3

Ambient noise characterisation

Mise en contexte (français)

Ce chapitre a un double objectif dans le schéma de modélisation adopté dans ce travail et représenté en Fig. 1.22. Le premier objectif est de déterminer la composition du bruit ambiant afin de choisir la position de la source dans les simulations numériques. Le deuxième objectif est de déterminer un modèle 1D approximatif en V_s à utiliser dans les simulations. Ceci est d'autant plus utile que les logs aux puits sont disponibles uniquement pour une plage de profondeur limitée. Ce chapitre est le résultat d'un travail commun mené avec Marc Peruzzetto, dont j'ai eu la chance d'encadrer le stage de fin d'études en 2017. Il a récemment été publié sous forme d'article:
Peruzzetto, M., Kazantsev, A., Luu, K., Métaxian, J.-P., Huguet, F., & Chauris, H., 2018. Broad-band ambient noise characterization by joint use of cross-correlation and MUSIC algorithm, *Geophysical Journal International*, **215**(2), 760–779

Context (English)

Within the modelling workflow in Fig. 1.22, this chapter has a double objective. The first is to determine the composition of the ambient noise in order to choose the source position in the numerical simulations. The second is to determine an approximate 1D V_s velocity model to be used in the simulations, as the well log data are only available over a limited depth interval. This chapter is a result of close collaboration with Marc Peruzzetto, whose master internship I supervised in 2017. It was recently published as
Peruzzetto, M., Kazantsev, A., Luu, K., Métaxian, J.-P., Huguet, F., & Chauris, H., 2018. Broad-band ambient noise characterization by joint use of cross-correlation and MUSIC algorithm, *Geophysical Journal International*, **215**(2), 760–779

Résumé (français)

Dans ce chapitre, plusieurs jours d'enregistrement du bruit ambiant (composante verticale) sont analysés. Le réseau d'écoute est dense, rectangulaire (3x6 km), constitué d'environ 100 sismomètres large-bande. L'objectif est la caractérisation du champ d'onde entre 0.1 et 3 Hz. L'azimut est déterminé à l'aide de l'algorithme MUSIC pour des fréquences inférieures à 1 Hz, et à l'aide du beamforming non-cohérent sur corrélations pour des fréquences supérieures à 1 Hz, car cette seconde approche est moins sensible aux effets d'aliasing. Une nouvelle approche pour la détermination de la vitesse est introduite, consistant à appliquer MUSIC à un stack des corrélations groupées par inter-distance commune (antenne linéaire virtuelle). Le mode fondamental et trois modes supérieurs des ondes de Rayleigh (R0, R1, R2, R3) sont détectés en dessous de 1 Hz. Au-dessus de 1.5 Hz, la phase L_g est observée, alors que R0 et R1 sont également présents. Environ entre 1 et 1.5 Hz, une onde plus rapide, probablement P_g, est présente. Les phases P_g et L_g sont dominantes pendant la période nocturne, suggérant qu'elles sont d'origine naturelle, ce qui est corroboré par leur azimut pointant en direction de l'Atlantique. Des simulations numériques 2D à grande échelle comportant des sources côtières et des sources en eaux profondes confirment la possibilité de l'excitation de la phase L_g par des sources océaniques. Ainsi, même au-dessus de 1 Hz, une part importante de l'énergie du bruit ambiant peut être due à des sources naturelles, pendant les heures calmes de la journée.

Summary (English)

Several days of passive seismic broadband recordings (vertical component) from a dense 3x6 km array installed near Chémery (France), with about 100 seismometers, are analysed for wavefield characterisation between 0.1 and 3 Hz. Back-azimuth is determined by using the MUSIC algorithm at frequencies below 1 Hz, and non-coherent cross-correlation beamforming above 1 Hz, since the latter is less sensitive to aliasing issues. A novel method of determining the wavefield velocity is introduced, consisting of processing a cross-correlation common-offset gather by the MUSIC algorithm. The fundamental and three higher modes of Rayleigh waves (R0, R1, R2, R3) are identified under 1 Hz. Above 1.5 Hz, the L_g phase is detected, while R0 and R1 are also present. Roughly between 1 and 1.5 Hz, a quicker phase, probably P_g, is detected. Both P_g and L_g are dominant during night-time, suggesting they have a natural origin, which is also consistent with their back-azimuth pointing towards the Atlantic. Large scale 2D spectral-element simulations using deep- and shallow-water ocean sources confirm the possibility of the L_g phase excitation. Thus, even above 1 Hz, natural sources can explain the major part of the ambient noise energy during quiet time periods.

3.1 Introduction

Ambient seismic noise applications are of growing interest in various contexts (Larose et al., 2015), boosting the development of numerous analysis methods. After the pioneering works by Douze (1964, 1967), Bonnefoy-Claudet et al. (2006b) reviewed

the techniques for the ambient wavefield composition, while Koper et al. (2010) compared data from 18 seismic arrays around the world to reveal some general trends. The frequency band of interest for most industrial applications (ambient noise tomography, H/V ratio studies for seismic hazard, direct hydrocarbon indicators, etc.) is roughly between 0.1 and 5 Hz. The wavefield at the peak frequencies of the primary and secondary microseisms (approximately 0.07 and 0.14 Hz, respectively) is usually dominated by the fundamental mode associated to Rayleigh and Love waves (Bonnefoy-Claudet et al., 2006b). For higher frequencies between 0.2 and 1 Hz, higher modes of Rayleigh and Love waves (Bonnefoy-Claudet et al., 2006b; Riahi et al., 2013a; Lehujeur et al., 2017a), as well as regional body waves (Poli et al., 2012) or teleseismic body waves (Pratt et al., 2017) may dominate the wavefield. Frequencies above 1 Hz are generally assumed to be dominated by artificial noise, with sources at the surface generating mostly Rayleigh waves of fundamental mode (Bonnefoy-Claudet et al., 2006a). However, Koper et al. (2010) highlighted L_g phase could also be predominant at these frequencies. The crust/mantle contact might thus play a major role in the short period ambient noise composition (Kennett, 1986).

Applications usually focus on one particular type of wave as representative of the ambient noise. The most widespread approach is to extract the surface wave (Rayleigh and/or Love) dispersion curves and to invert for a vertical shear-wave velocity profile. Array methods such as SPAC (Aki, 1957), FK (Capon et al., 1967) or High-Resolution FK (HRFK, Capon, 1969) were traditionally used for this purpose. The relatively recent development of the ambient-noise cross-correlation allows to extract the medium's Green's function between two passive seismic receivers (e.g. Shapiro & Campillo, 2004). Dispersion curves can thus be estimated for each station pair of 2D receiver arrays, allowing to invert for a smooth 3D shear-wave velocity cube (e.g. Brenguier et al., 2007; Mordret et al., 2013a). For some data sets, cross-correlation based techniques also recover the body-wave part of the Green's function (e.g. Roux et al., 2005). Nakata et al. (2015) used the direct diving body waves for a 3D travel-time tomography, while Draganov et al. (2007) and Ruigrok et al. (2011) managed to image continuous structural interfaces by extracting reflected P-waves from the cross-correlations. Teleseismic P-waves were used by Landès et al. (2010); Obrebski et al. (2013); Pratt et al. (2017) to track hurricanes and other meteorological perturbations affecting the oceans. Furthermore, heterogeneities within the Earth continuously scatter the ballistic surface and body waves, generating coda waves at later arrival times. They can be used to determine the mean properties of the medium. Sens-Schönfelder & Wegler (2006) presented an application for water-table monitoring.

Three-component data offer the possibility to compute the spectral ratio of the horizontal to vertical displacements. The H/V technique was introduced by Nakamura (1989) to derive SH wave resonance and an estimate of the contact depth between the bedrock and the overburden from the H/V ratio peak frequency, despite debatable hypotheses (Bonnefoy-Claudet et al., 2006a).

Dangel et al. (2003) observed an amplification of the vertical component as well as a low H/V ratio for frequencies ranging from 1 to 5 Hz above hydrocarbon reservoirs. This resulted into a debate, some authors claiming that the amplification could be used as a direct hydrocarbon indicator («DHI», see Lambert et al., 2009), while others objected that it was due to either higher artificial noise level near the hydrocarbon

extraction facilities or shallow geological structural effects (e.g. Green & Greenhalgh, 2010, and references herein). A quantitative modeling of the amplification of the vertical component would require a detailed knowledge of the ambient noise content over the frequency band of interest. This is the main objective of this work.

We propose here a methodology based on a joint use of a high-resolution array method (MUSIC, Schmidt, 1986; Goldstein & Archuleta, 1987) and ambient cross-correlation on the vertical data recorded by a dense array of broadband seismometers. In this way, the dispersion curves and the back-azimuths of the wavefield can be determined over a wide frequency band. Applying MUSIC (Multiple Signal Characterisation) to the windowed recorded signals provides a better resolution at low frequencies compared to FK or HRFK (Cornou, 2002), provided the number of incoming waves is inferred correctly. For higher frequencies, cross-correlation based beamforming techniques (e.g. Mordret et al., 2013a; Nakata et al., 2015; Lehujeur et al., 2017b) allow to obtain reliable slowness for inter-station distances beyond the wavelength (aliasing limit). This is achieved by aligning inter-station correlations instead of aligning raw waveforms. We suggest the combination of these methods provides a better insight into the wavefield composition over a wider frequency range that the one associated to a specific approach.

We first present the data set recorded in Central France (section 3.2), and then the methodology (section 3.3). In section 3.4 we focus on the real data set application. We discuss the evidence for the presence of Rayleigh wave modes and two crustal phases P_g and L_g into the [0.1-3] Hz frequency band, and the possibility to invert for a 1D shear wave velocity profile. In section 3.5, we use the spectral-element method to simulate the wavefield and to distinguish between artificial local surface sources and natural distant deep- or shallow-water sources, based on the wavefield content determined by our methodology.

3.2 Data

The study site is located inside the Paris sedimentary basin. Successive sedimentary layers are mainly composed of limestones, sandstones and clays. The reservoir itself is a layer of Triassic sandstones at a depth of approximately 1130 m (blue contour lines in Fig. 3.1a and green box in Fig. 3.1b). The thickness of the layer varies from 30 to 60 m. The depth of the basement is unknown at the reservoir location. It was neither reached by any of the wells nor seen in active seismic data, because the targeted reservoir layers were much shallower. The deepest exploration well in the area (CS01, see location in Fig. 3.1a), drilled in 1959 by Mobil Repga, stopped at 2680 m, where Permo-Triassic sediments were still dominating (Fig. 3.1b). Such a deep basement in this area is consistent with the reported presence of a Permian basin (Gély & Hanot, 2014, see the map enclosed to their book).

About 100 broadband Trillium 40 s seismometers (Nanometrics) were deployed in April 2010 and November 2010 with various geometries and a 100 Hz sampling frequency. Spectrograms reconstructed over the whole time period from the stations successively located near the center of the study area (within the dashed green rectangle in Fig. 3.1a) are presented in Fig. 3.2. As expected, above 1 Hz, they exhibit lower PSD during night time and on Sundays, as well as during lunch pauses at noon. These features, typical of anthropogenic noise, motivate a separate analysis of day-

(a) Array deployment 20/04/2010 (b) CS01 simplified log

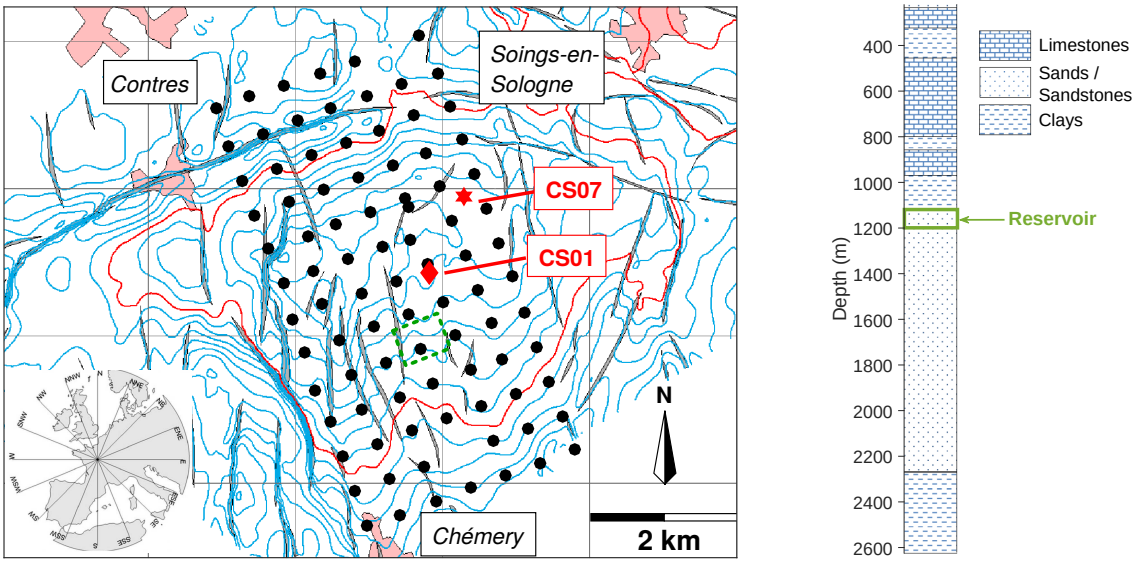
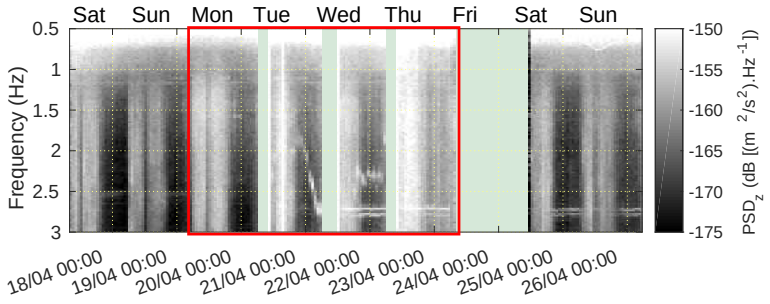


Figure 3.1: (a) Seismic network. Black dots mark the positions of the sensors deployed on 20 April. Reservoir's contour lines are shown in blue, and its maximum possible extent is given by the red curve. The black lines patch known faults. Urban ares are shaded in pink. The positions of wells CS01 and CS07 are shown with red markers. The green dashed line delimits the area where sensors were taken to compute the spectrograms in Fig. 3.2. The site location map in the left bottom corner was readily taken from Riahi et al. (2013a). (b) Simplified fundamental log at well CS01. The reservoir is shown by the green box. Log data were provided by Storengy.

(a) Power Spectral Density in April



(b) Power Spectral Density in November

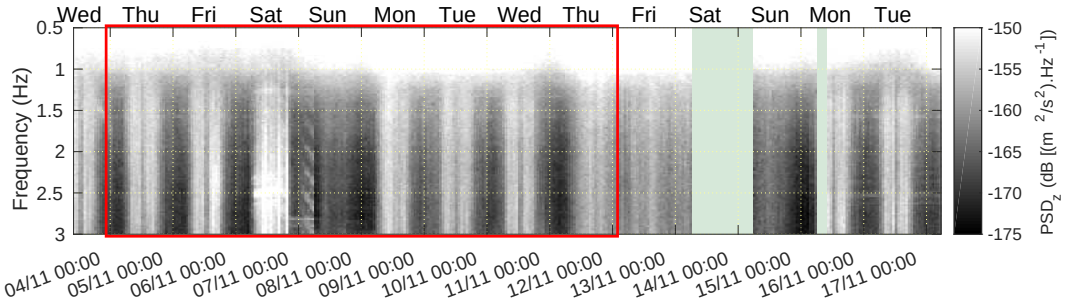


Figure 3.2: Power Spectral Density of the vertical component velocity for the data recorded in April 2010 (a) and November 2010 (b). They are computed using one station within the green rectangle in Fig. 3.1a. The exact station location is not necessarily the same from one day to another. Times when no stations were recording in the chosen area are highlighted in green. The red rectangles delimit the time period when rectangular arrays were used, which corresponds to the time period analysed in this work.

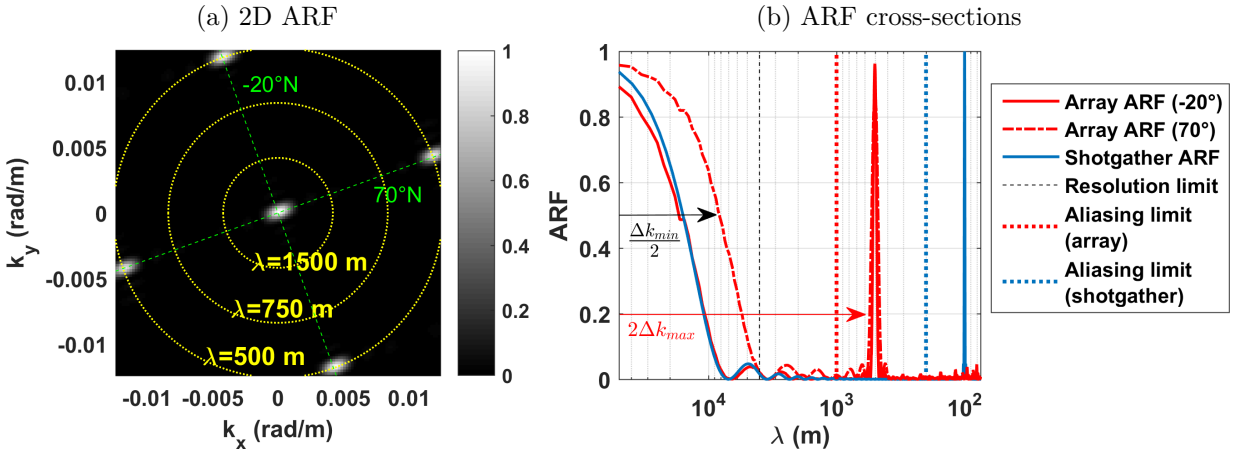


Figure 3.3: (a) Array Response Function (ARF) of the rectangular array deployed on 20 April. Yellow circles indicate wavelengths in the wave-vector plane. (b) ARF along the two main directions of the array (red plain and dash-dotted curves), compared to the 1D ARF of the virtual shot gather obtained from the cross-correlations (blue curve). The black dash-dotted line represents the theoretical maximum detectable wavelength (resolution limit). The red and blue dotted lines represent the minimum detectable wavelength (aliasing limit) for the 2D array and the shot gather, respectively.

and night-time periods, because the dominating wavefield sources might alternate between natural and artificial. In our study we specifically focus on the time periods when the array had a rectangular configuration. From 20 to 24 April, and from 4 to 11 November, most of the sensors were operational between 3 PM and 6 AM local time (UTC+1), and were shifted by a about 250 m every day. There was an almost uniform spacing of 500 m between the sensors installed over a 3×6 km area located above the UGS. The deployment geometry for 20 April is shown in Fig. 3.1a. We used 3 hours of continuous recording, either 1-4 AM or 3-6 PM local time, respectively referred to as «night-time» and «day-time», and processed separately. Such 3 hour intervals were concatenated for all the available days (4 in April and 8 in November). Riahi et al. (2013a) already analysed our data set using a three-component beamforming algorithm into a narrower frequency band [0.4-1.1] Hz. Both results will be compared in section 3.5.

3.3 Methodology

For both cross-correlation and MUSIC approaches, recordings were first split into elementary time windows of about 328 seconds (2^{15} samples), overlapping by 50%. Each window was tapered in time-domain and signals were then filtered between 0.1 and 3 Hz. Tapered square cosine windows with a transition width of 10% of the total length of the taper were used in both time and frequency domains.

3.3.1 Aliasing and resolution limits

We compute the array response function (ARF) following, among others, Krim & Viberg (1996); Cornou (2002); Foti et al. (2011). An example of ARF for the array configuration we used in our real-data application (Fig. 3.1a) is shown in Fig. 3.3a. The estimated aliasing limit (minimal unambiguously detectable wavelength) is twice the shortest inter-station distance (here, $\lambda_{min} = 2d_{min} = 1000$ m). The resolution limit (maximum detectable wavelength) can be estimated as the full width of the ARF's central lobe at half maximum (in our case, $\lambda_{max} = 4000$ m). Beyond this wavelength, the array resolution is not good enough for distinguishing a real incident wave from a wave with an infinite apparent velocity. It should be stressed that these limits are suitable for standard FK analysis. In the following, we will see that high-resolution methods sometimes allow to obtain reliable results beyond these theoretical limitations, as already mentioned by Foti et al. (2011).

3.3.2 Cross-correlation

Each tapered time-window was pre-processed following Bensen et al. (2007), including spectral whitening and one-bit normalisation to filter out earthquakes. Correlations between pre-processed signals s_i and s_j recorded at stations i and j were averaged over the available L time windows, yielding

$$\Gamma_{ij}(t) = \frac{1}{L}\sum_{l=1}^{L}\int s_i^{(l)}(\tau)s_j^{(l)}(\tau+t)d\tau. \tag{3.1}$$

In case of evenly distributed sources, $\Gamma_{ij}(t)$ approaches the Green's function between the two stations to within an amplitude factor (Boschi et al., 2013). For an uneven distribution, the causal part is correctly reconstructed for the station pairs aligned with the major source direction, while it is biased for station pairs with other orientations (Lehujeur et al., 2017a). In the latter case, the cross-correlations keep an imprint of the source azimuthal distribution, and can be used for measuring this distribution inside a routine beamforming procedure (section 3.3.2). On the contrary, if the purpose is to measure the Green's function of the medium to derive dispersion curves, the uneven source distribution must be corrected for. This is done by averaging cross-correlations over similar station pairs (section 3.3.2).

Cross-correlation beamforming (CC-beam)

Cross-correlation beamforming can be carried out either in the frequency domain (Ruigrok et al., 2017), or in the time domain. In our study we choose the latter solution since it allows an easy implementation of the non-coherent beamforming (see below). Assuming a single plane wave with slowness vector $\mathbf{s}$, the time delay between two stations is

$$\tau_{ij}(\mathbf{s}) = \mathbf{s}\cdot\mathbf{r_{ij}}\ , \tag{3.2}$$

where $\mathbf{r_{ij}}$ is the position vector going from station i to station j. $\Gamma_{ij}(t)$ is maximum at $t = \tau_{ij}(\mathbf{s})$ for any couple (i,j). Thus, the sum $\sum_{i>j}\Gamma_{ij}(\tau_{ij}(\mathbf{s}))$ is maximum if $\mathbf{s}$ is actually the slowness of the wave propagating across the array (Rost & Thomas, 2002). In practice, the signal-to-noise ratio in the cross-correlation can be quite low

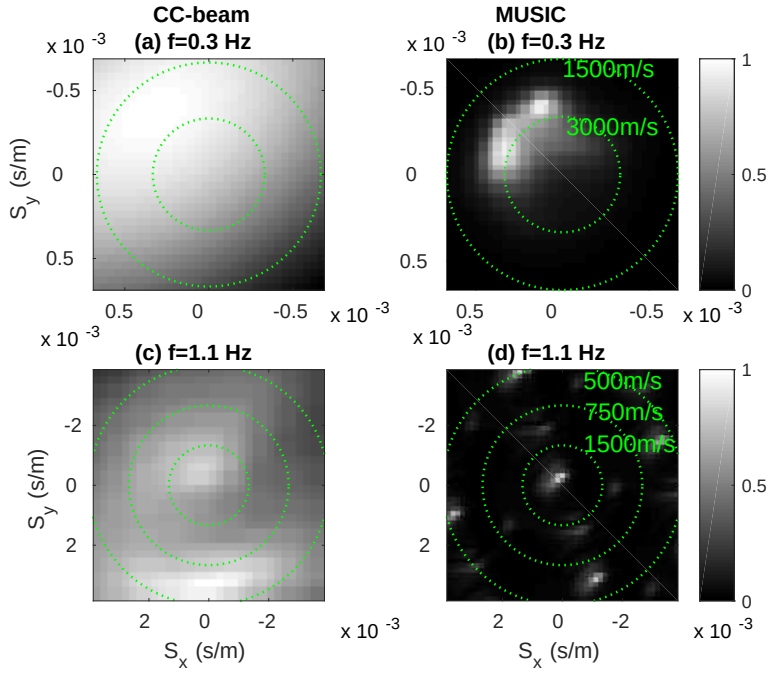


Figure 3.4: Compared beam-formers for CC-beam and MUSIC (20-23 April). (a) CC-beam (0.3 Hz), (b) MUSIC (0.3 Hz), (c) CC-beam (1.1 Hz), (d) MUSIC (1.1Hz). Slowness axes represent group slowness for CC-beam and phase slowness for MUSIC. The green circles show several velocity values, which are labeled in the right column.

at frequencies above 1 Hz because of strong scattering. Under such conditions, it is more robust to sum envelopes of the correlations:

$$D_{CC-beam}(\mathbf{s}) = \sum_{i>j} \left| \mathcal{H}\left\{\Gamma_{ij}\right\}\left(\tau_{ij}(\mathbf{s})\right)\right|, \tag{3.3}$$

where $\mathcal{H}$ is the Hilbert transform. This will be called *non-coherent* beamforming in the following. If several uncorrelated waves propagate across the array, $D_{CC-beam}$ will exhibit several peaks at the corresponding slowness vectors. In order to obtain slowness vectors into different frequency bands, correlations were filtered with a 0.2 Hz wide tapered square cosine windows centered on the frequency of interest. Examples are shown in Figs 3.4a and c. Such beam-formers yield an estimate of the back-azimuth distribution.

This method can also be used for group velocity dispersion curve estimation (see Appendix 3.A, Fig. 3.A.1). However the grid search over the slowness space at each requested frequency is rather CPU consuming and more importantly the resolution at low frequency is very low. This motivates another way of exploiting the cross-correlations for the dispersion curve retrieval, which will be addressed in the next paragraph.

Common offset stacking

Most of the time, sources are not uniformly distributed around on-shore arrays, preventing a proper reconstruction of the medium Green's function from cross-correlations. This problem can be efficiently fixed if we accept to drop the azimuthal information.

In the theoretical derivation (e.g. Boschi et al., 2013), the azimuthal integration over the source distribution is indeed equivalent to an azimuthal integration over the station pairs orientation. If enough station pairs with close offsets(i.e. distance between stations) and different orientations are available, the uniform source distribution condition is fulfilled for their average correlation. As already done by Poli et al. (2012); Mordret et al. (2013a); Nakata et al. (2015) and others, we divided all the available station pairs into 100 m offset bins, and averaged the cross-correlations for station pairs inside each bin. To correct for the longer NW-SE array extent, we also divided each offset bin into 10° azimuthal bins, and computed the average cross-correlation as

$$\Gamma(r,t) = \frac{\sqrt{r}}{\sum w_{ij}} \sum_{|\mathbf{r_{ij}}| \in [r, r+\Delta r]} w_{ij} \Gamma_{ij}(t), \tag{3.4}$$

with $w_{ij} = 1/N_\theta(i,j)$ being a weight inversely proportional to $N_\theta(i,j)$ the number of pairs in the azimuthal bin containing the couple (i,j), and $\sqrt{r}$ a correction term for surface wave geometrical spreading. This yields a 2D-shot gather with one trace per offset, and a spacing equal to the offset bin width Δr. Examples of such shot gathers are shown in Figs 3.5a and b. Δr is often smaller than the typical inter-station distance, which shifts the aliasing limit towards smaller wavelengths (Fig.3.3b):$\lambda_{min} = 2\Delta r$. We assume the resolution limit does not change with respect to the 2D array's ARF defined in section 3.3.1. In order to obtain the dispersion curves, the shot gather can now be processed as a linear array, using standard FK processing (FK-CC, Figs 3.5c and d), or, more interestingly, using the MUSIC algorithm.

3.3.3 MUSIC

The MUSIC algorithm separates recorded data into signal and noise subspaces (Schmidt, 1986; Goldstein & Archuleta, 1987). To achieve this separation, the data are first gathered into the cross-spectral matrix (CSM) $R(f)$ with elements given by

$$R_{ij}(f) = \langle S_i(f) \overline{S_j(f)} \rangle, \tag{3.5}$$

where $S_i(f)$ and $S_j(f)$ are the Fourier transforms of a one-component signal at stations i and j, $\overline{\ }$ is the complex conjugate and $\langle\ \rangle$ stands for an averaging operation, performed over adjacent frequencies (spectral smoothing), different time windows (temporal smoothing), similar sub-arrays or similar station pairs (spatial smoothing).

Given a priori knowledge of the number Q of plane waves to be detected, the MUSIC functional minimises the projection of the data onto the noise subspace (see Appendix 3.B). It is difficult in practice to determine Q when analysing seismic noise. Besides, in real noisy data, one plane wave can be «spread» over several eigenvalues. Several techniques were proposed for an automated detection of Q (see the review in Cornou (2002)). Here we developed a slightly different method, more robust for the purpose of pure ambient noise analysis. It is described in detail in Appendix 3.B. The idea is to use the slope break in the logarithmic eigenvalue decay in case of a strong signal-to-noise ratio (SNR), and a simple eigenvalue magnitude criterion in case of a low SNR.

From the definition (3.5) of the CSM, it will only be of rank 1 (which means one non-zero eigenvalue) if the averaging operation is not performed. The smoothing operation in equation (3.5) is thus necessary for the derivation presented in Appendix 3.B to hold (see Bokelmann & Baisch, 1999; Cornou, 2002). The traditional technique for *spatial smoothing* is to average CSMs obtained from identical sub-arrays, which results into a smaller CSM with increased rank (Shan et al., 1985). However this technique can only be applied to almost perfectly regular arrays. An alternative method consists in averaging individual terms in the CSM corresponding to nearly identical station couples (Bokelmann & Baisch, 1999). We used either option for spatial smoothing, depending on the type of the data to which MUSIC was applied.

2D array

Though the 2D arrays in our data set were quite regular, we preferred the second alternative for *spatial smoothing* (i.e. Bokelmann & Baisch, 1999) in order to make the method easily repeatable on less regular arrays. The station pairs were divided into 100 m offset bins ranging from 0 to 7000 m, and 5° azimuth bins ranging from 0° to 180°. CSM values were then averaged inside each bin. *Spectral smoothing* was also performed over a 0.1 Hz interval centered on the frequency of interest. After smoothing and diagonalising the CSM, D_M can be estimated from (3.B.5) and mapped over the two-dimensional wave vector space, or equivalently the slowness vector space, provided $\mathbf{k} = 2\pi f \mathbf{s}$. For a given frequency, such slowness maps are computed for each time window and then averaged, yielding the final beam-former (see examples shown in Figs 3.4b and d). In the following, we refer to the method described in this subsection as «direct» MUSIC, since the algorithm is applied directly to the recorded waveforms. While the resolution at low frequency (0.3 Hz, Fig. 3.4b) is much better compared to CC-beam described into the previous section (Fig. 3.4a), the MUSIC beam-former is severely aliased at higher frequencies (for instance at 1.1 Hz, the minimum velocity of the non-aliased zone is 1100 m/s, Fig. 3.4d). On the contrary, the CC-beam beam-former (here 1.1 Hz, Fig. 3.4c), though poorly resolved, does not encounter aliasing. This is because it operates on the time delay of wave packet propagation, and not the phase delay. The comparison suggests a strong complementarity between both approaches for back-azimuth determination. The MUSIC beam-former can finally be integrated over azimuth for each frequency. Concatenating the resulting 1D curves yields a dispersion plot in (f, s) or (f, v_ϕ) domain (in Figs 3.5g and h). The validity of such a dispersion plot between the aliasing and the resolution limits is also confirmed in a simple synthetic test performed in Appendix 3.C. The consistency between phase and group velocity dispersion curves derived respectively by MUSIC and CC-beam is verified in Appendix 3.A. However, neither of these two methods applied individually is as efficient for dispersion curve determination as the MUSIC algorithm applied to a cross-correlation common-offset gather. The latter approach presented in the next section.

Linear array with regular spacing

MUSIC was applied to virtual shot gathers obtained by cross-correlation (MUSIC-CC). Such a shot gather is equivalent to a linear array with a constant spacing equal to the width of the offset bin used during the common offset stacking of the cross-

correlations (section 3.3.2). Since it is very easy to define identical sub-arrays in this case, we used the traditional *spatial smoothing* described in Shan et al. (1985). CSMs were thus averaged over 20 sub-arrays. *Spectral smoothing* was again implemented with a 0.1 Hz smoothing interval around the frequency of interest. Here D_M can be directly plotted on a 1D slowness or phase velocity axis as the virtual array is linear. Concatenating such curves obtained for successive frequencies again yields a dispersion plot in (f, s) or (f, v_ϕ) domain. Such a dispersion plot (see examples in Figs 3.5e and f) exhibits a better resolution at low frequencies and less aliasing at high frequencies compared to «direct» MUSIC (Figs 3.5g and h). MUSIC-CC's resolution is also better than for the standard FK processing of the virtual shot gather (FK-CC, Figs 3.5c and d).

3.4 Results

In this section, we present the results obtained with our methodology on the data acquired above an Underground Gas Storage (UGS) near Chémery (Central France, Fig. 3.1a) in April 2010 and November 2010. The UGS was at its maximum filling level in November. These experiments originally aimed at a time-lapse observation of a low-frequency amplitude anomaly above the UGS as a direct hydrocarbon indicator (DHI). We compare the dispersion plots in the frequency band [0.3-3] Hz obtained by the different methods from section 3.3 (FK-CC, MUSIC-CC and «direct» MUSIC). Several propagating modes are identified, and their back-azimuths are studied using MUSIC and CC-beam approaches, outside and inside the aliasing zone, respectively. Differences in the wavefield composition between day- and night-period are investigated. Finally, the dispersion curves identified for different Rayleigh wave modes are jointly inverted for a 1D shear wave profile in order to check the compatibility of their identification with the local geology.

3.4.1 Dispersion plots

Cross-correlations are computed as described in section 3.3.2 between 0.1 and 3 Hz. Dispersion plots are derived into the frequency range [0.15-2.85] Hz, which is slightly smaller than the initial range because the MUSIC-based approaches use spectral smoothing over a 0.1 Hz interval. Results are presented in Fig. 3.5, where the left column is for April data, and the right column for November data.

The shot gathers computed for April and November (Figs 3.5a and b) look very similar, confirming the relative stability of the wavefield. The standard FK processing of the shot gathers (FK-CC, Figs 3.5c and d) exhibits three distinct dispersive patches below 1 Hz which it would be natural to interpret as three Rayleigh wave modes. However, several modes might be mixed into one apparent patch in case of an insufficient resolution. The velocity of the patches remains below 3500 m/s most of the time, but exhibits sharp peaks up to 8000 m/s, which are unrealistic for a surface wave dispersion curve. Compared to FK-CC, MUSIC-CC (Figs 3.5e and f) exhibits a qualitatively similar pattern with a better resolution. Two distinct modes can now be distinguished below 0.4 Hz, where FK-CC just identified one wide patch. Separating these two low-frequency modes is crucial for the V_S profile inversion in depth.

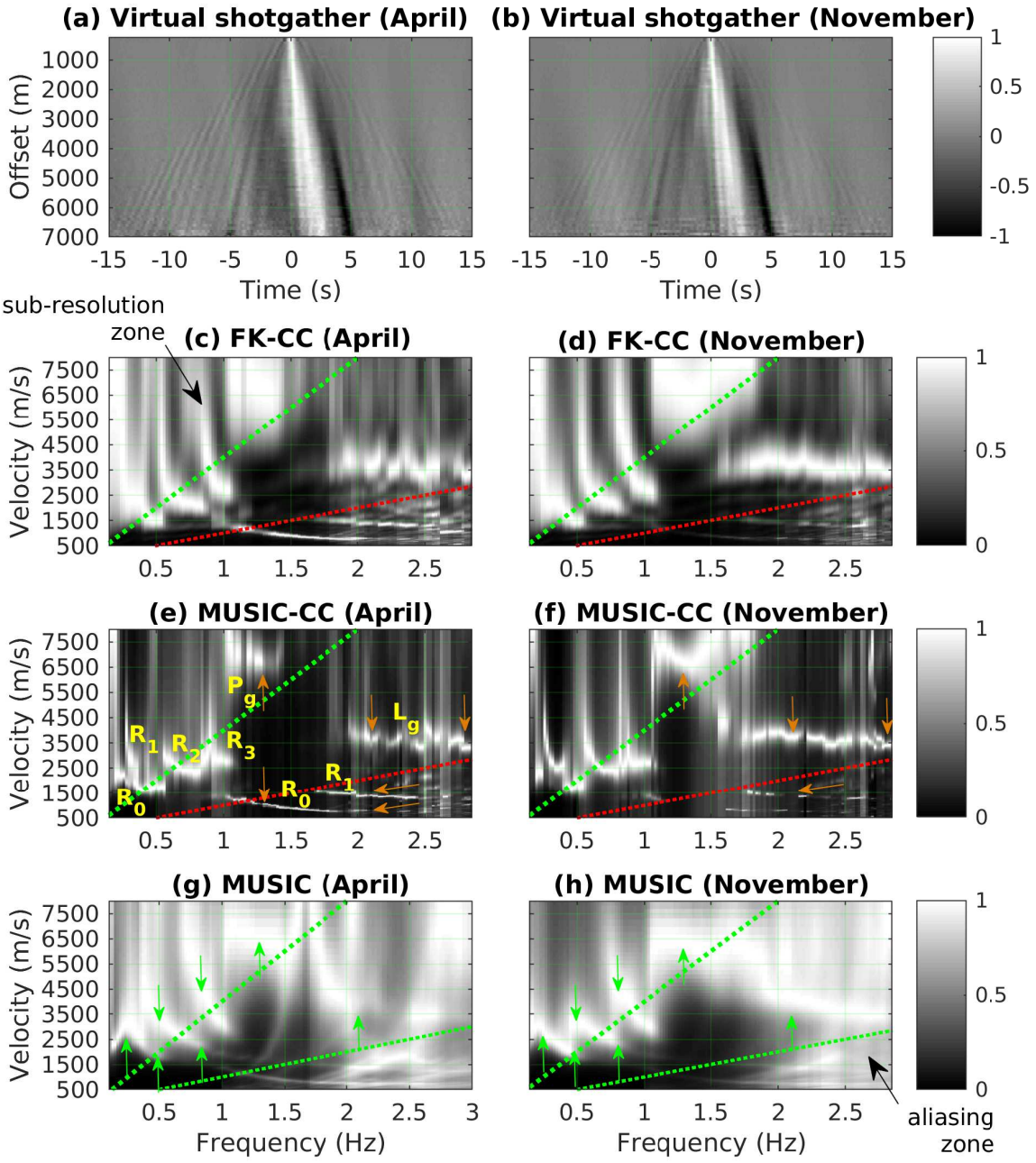


Figure 3.5: Night-time dispersion study in April (left column) and November (right column). (a,b) Virtual shot gathers from cross-correlation; (c,d) FK-CC (standard FK processing of the virtual shot gathers); (e,f) MUSIC-CC (MUSIC processing of the virtual shot gathers). (g,h) «direct» MUSIC processing of the signal recorded by the 2D array (without cross-correlation). MUSIC-CC yields the best results: for instance, compare panel (e) to (c) and panel (e) to (g). Green dotted lines mark the theoretical resolution and aliasing limits of the processed arrays. The red dotted line reported in (c), (d), (e) and (f) is the aliasing limit of the 2D array. The yellow labels identify different detected modes. The green (resp. orange) arrows are plotted at frequencies for which beam-formers are shown in Fig. 3.6 (resp. Fig. 3.7); they point towards the modes expected in the beam-formers at these frequencies.

Above 1 Hz, in April, two clear dispersive modes are detected below 2000 m/s, identified as the Rayleigh fundamental mode and the first overtone. A quick non-dispersive phase ($v_\phi \approx 7000$ m/s) is observed between 1 and 1.5 Hz in April, and between 1 and 1.8 Hz in November, which we interpret as P_g. It is indeed too quick for a guided S-wave, and too slow for a teleseismic arrival (e.g. Obrebski et al., 2013). Finally, above 2 Hz, an apparently non-dispersive phase propagating slightly quicker than 3500 m/s is detected. As Koper et al. (2010), we interpret it as L_g, which we will further discuss in section 3.5. All the identified modes are labeled in yellow in Fig. 3.5e.

The dispersion plot obtained from the direct application of the MUSIC algorithm to the noise recorded by the 2D array (Figs 3.5g and h), exhibits a lower resolution compared to MUSIC-CC (Figs 3.5e and f), and suffers from artifacts inside the aliasing zone. Branches due to the waves from the aliasing zone can also contaminate the aliasing-free zone, as for example the steeply increasing branch between 1.3 and 1.7 Hz in Fig. 3.5g. Below 0.4 Hz, only one modal branch is visible, with roughly the average phase velocity of the two modes $R0$ and $R1$ identified in Fig. 3.5e. This behavior is expected for MUSIC applied to noisy multimodal waveforms with random emission times and azimuths, as shown for synthetic data in Appendix 3.C. Thus, MUSIC-CC appears to be the best suited method for dispersion curves retrieval at all frequencies (for instance, for April night-time data, compare Fig. 3.5e to Fig. 3.5c, and Fig. 3.5e to Fig. 3.5g). The automated signal subspace determination plays a major role in the efficiency of the method (see Appendix 3.B for details).

3.4.2 Back-azimuths

As explained in section 3.3, «direct» MUSIC is used to extract beam patterns outside the aliasing zone (Fig. 3.6), while non-coherent cross-correlation beam-forming (CC-beam) is used inside it (Fig. 3.7). All the plotted beam-formers are normalised between 0 and 1.

The knowledge of the dispersion plots (Fig. 3.5) allows to better assess the mode labels in the beam-formers, benefiting from a continuous representation over a wide frequency range. Green arrows in Figs 3.5g and 3.5h indicate the modes which are expected to show up in the MUSIC beam-formers, at the frequencies for which they are plotted in Fig. 3.6, namely 0.25, 0.5, 0.8, 1.3 and 2.1 Hz. The arrows below 1 Hz correspond to several dispersive Rayleigh wave modes, while the arrows at 1.3 and 2.1 Hz correspond to the non-dispersive P_g and L_g phases, respectively. In the same way, brown arrows in Figs 3.5e and 3.5f indicate the expected modes for the CC-beam images in Fig. 3.7 at 1.3, 2.1 and 2.8 Hz. Phase velocities from Fig. 3.5 were converted into group velocities for the dispersive modes R0 and R1, in order to place the arrows at the right distance from the center in Fig. 3.7. The conversion procedure is described in Appendix 3.A. If a mode expected from the dispersion plot is not found in the beam-former, its label is put inside brackets. If an unexpected mode is identified, it is labeled with brackets and without any arrow.

Outside aliasing zone (MUSIC)

In both April and November, a continuous change of the dominant back-azimuth can be observed over frequencies. At low frequencies (0.25 Hz), the dominant direction

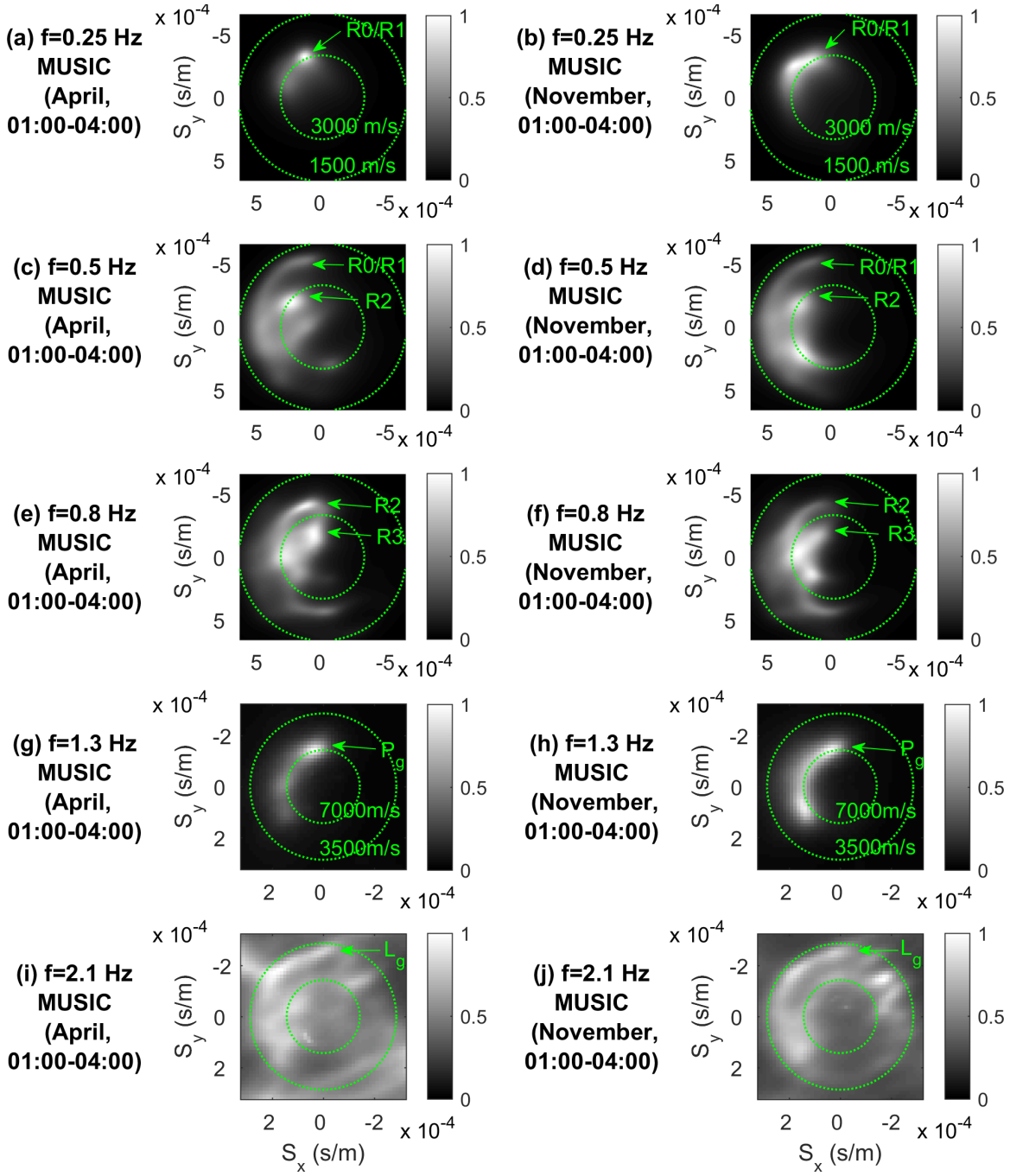


Figure 3.6: Night-time back-azimuths in April (left column) and November (right column), provided by MUSIC outside the aliasing zone. Analysed frequencies are 0.25 Hz (a,b), 0.5 Hz (c,d), 0.8 Hz (d,f), 1.3 Hz (g,h) and 2.1 Hz (i,j). The green circles indicate some velocity values; if they are not labeled, the velocity values are those of the subplot above. The mode labels are those from Fig. 3.5(e). The green arrows in the left (resp. right) column are those from Fig. 3.5g (resp. 3.5h).

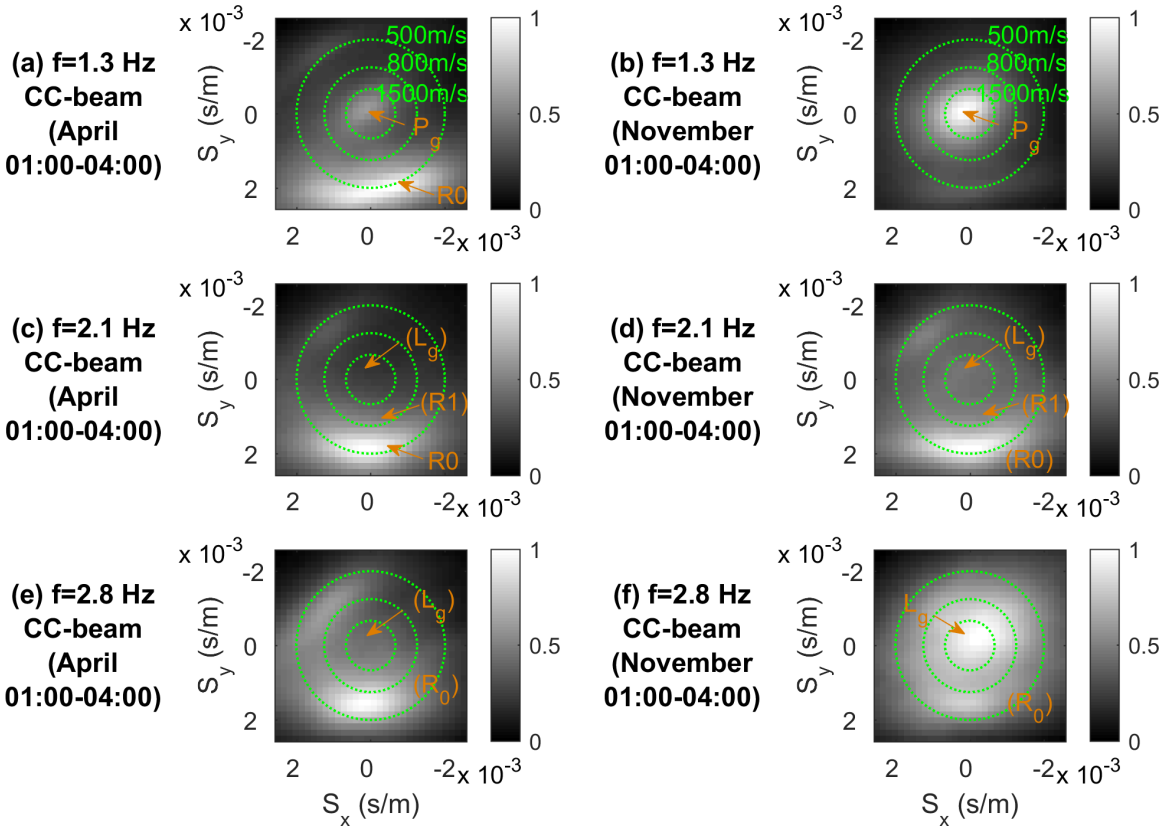


Figure 3.7: Night-time back-azimuths in April (left column) and November (right column), provided by CC-beam inside the aliasing zone. Analysed frequencies are 1.3 Hz (a,b), 2.1 Hz (c,d) and 2.8 Hz (e,f).

is NNW (Figs 3.6a and b), consistent with the well-known secondary microseim generation zone in the Northern Atlantic (Ardhuin et al., 2011; Stutzmann et al., 2012). At 0.5 and 0.8 Hz (Figs 3.6c to f), back-azimuths cover directions from N to SSW, suggesting a generation area along the Atlantic coast. The P_g phase (1-1.5 Hz, Figs 3.6g and h) exhibits a remarkably regular beam-former in November, covering back-azimuths from N to SW. The same distribution is observed in April with a strong peak for the N direction. The L_g phase (above 2 Hz, Figs 3.6i and j) appears with a noisier beam-former (especially in April), probably because of higher scattering and contamination by branches coming from the aliasing zone (see Fig. 3.5g). The back-azimuth for both April and November ranges from NE to SW, suggesting a contribution from the Scandinavian Northern coast, which is also known for being a secondary microseism generation zone (Essen et al., 2003).

Inside aliasing zone (CC-beam)

Compared to the previous paragraph, a radical change of the back-azimuth distribution is observed in the aliasing zone above 1 Hz (Fig. 3.7). In April, the SE-S directions are dominant at all the plotted frequencies (1.3, 2.1 and 2.8 Hz, see Figs 3.7a, c and e, respectively).

Only the R0 mode is detected by CC-beam, also when it is unexpected from the dispersion plot (Figs 3.7d, e and f). R1 is never detected, although expected in Figs 3.7c and d. This is probably due to the fact that the R1 mode does not form an independent wave packet in the cross-correlations, making it undetectable by a non-coherent beamforming approach.

Though the quick non-dispersive phases do not lie in the aliasing zone, they should still be seen in the beam-formers. They are better visible in November (P_g in Fig. 3.7b and L_g in Fig. 3.7f) compared to April (only a weak P_g peak in Fig. 3.7a). This confirms those phases are more energetic in November, as expected from Figs 3.5g and 3.5 h. However, it is still unclear to us why L_g is completely missed by the CC-beam method in Figs 3.7c, d and e: while R1 is mixed with R0 and thus not detected, the L_g phase does not seem to be mixed with another phase.

3.4.3 Day/night variation

Day-time virtual shot gathers and the corresponding MUSIC-CC dispersion plots are shown in Fig. 3.8. It is striking to observe that the quick P_g and L_g phases, dominant at night (Figs 3.5e and f), almost completely disappear the day, over-shaded by the R0 and R1 modes (Figs 3.8c and d). This suggests a different origin for P_g/L_g and R0/R1 above 1 Hz, which will be discussed in section 3.5. This analysis is completed by the day-time back-azimuth visualisation below and above 1 Hz (Figs 3.9 and 3.10 respectively). While the diurnal variation is minimal below 1 Hz (compare Figs 3.6 and 3.9), it is quite strong above 1 Hz (compare Figs 3.7 and 3.10). During the day, the back-azimuth distribution above 1 Hz clusters around NNW and S directions, while only the Southern directions dominate during the night. Also, as expected from Figs 3.8c and d, the quick non-dispersive phases P_g and L_g are almost totally over-shaded by the slow dispersive phases (R0 in Fig. 3.10).

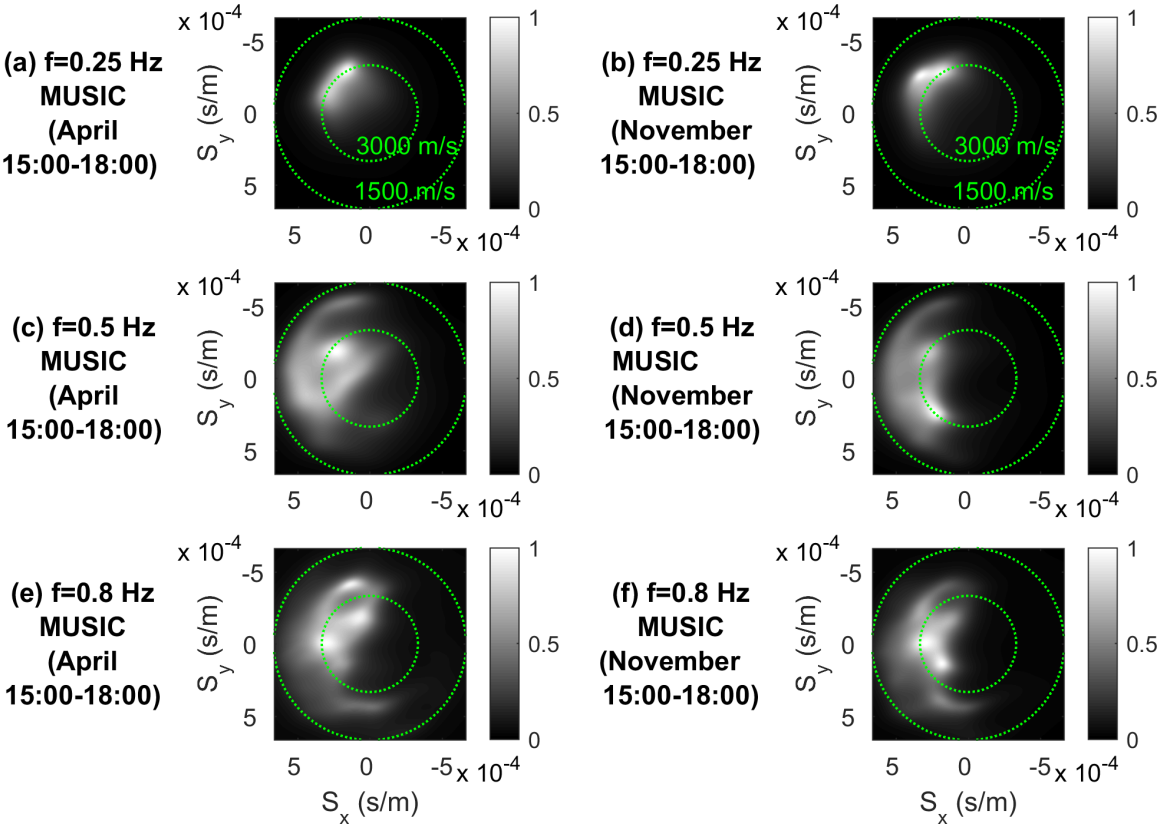


Figure 3.8: Day-time dispersion study in April (left column) and November (right column). (a,b) Virtual shot gathers from cross-correlation; (c,d) MUSIC-CC (MUSIC processing of the shot gather). The green lines mark the theoretical resolution limit of the virtual shot gather.

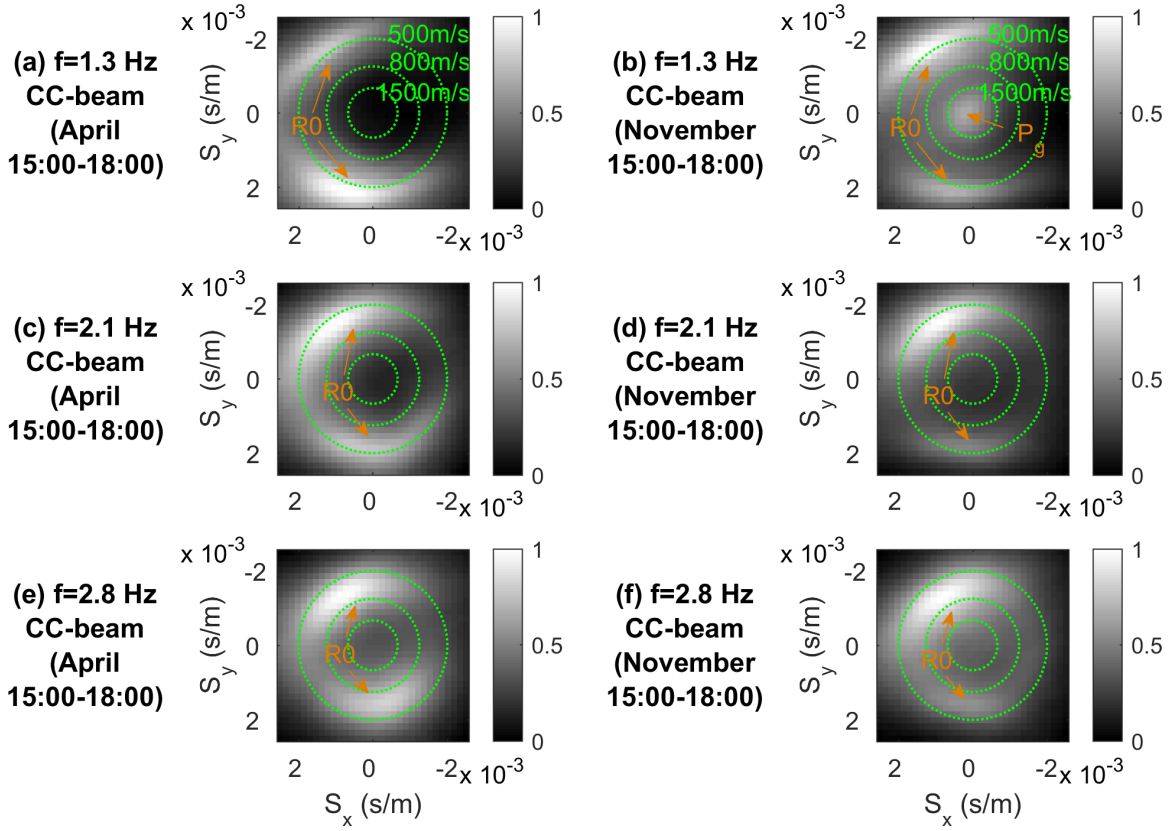


Figure 3.9: Same as Fig. 3.6, but for day-time and below 1 Hz.

3.4.4 Inversion for V_S

In order to check the compatibility of the mode identification with the available geological knowledge, and to investigate the potential for exploration purposes, we inverted Rayleigh wave dispersion for a vertical S-wave velocity profile. Dispersion curves were manually picked from the MUSIC-CC dispersion plots for April (Fig. 3.5(e), all modes except R0 between 2.1 and 2.35 Hz) and for November (Fig. 3.5(f), R0 between 2.1 and 2.35 Hz). The picked curves are shown with circles in Fig. 3.11(a). Only R0, R1 at high frequency (1.7-2.5 Hz) and R2 were used for inversion, since we were less confident about picking R1 at low frequency and R3 (large uncertainty and narrow spectral extent). Since the L_g phase velocity (approximately 3500 m/s, red dotted line in Fig. 3.11a) is close to V_S in the crust, the latter value was constrained between 3400 and 3600 m/s. Theoretical dispersion curves were estimated via the Thomson-Haskell method (Thomson, 1950; Haskell, 1953). The picked dispersion curves were inverted using a Competitive Particle Swarm Optimiser (Luu et al., 2018), with 20 independent inversions. For each inversion, we used a swarm size of 50 and a maximum number of iterations of 500. We parameterised the velocity model with 9 layers, and inverted for the S-wave velocity V_S, the V_P/V_S ratio, and the thickness of each layer. The mean V_P/V_S profile obtained is close to 2 in all the layers with high uncertainties as the forward problem is rather insensitive to V_P. This value is however consistent with ratios obtained from V_P and V_S logs available for a nearby Storengy gas storage facility (Céré la Ronde) with a similar sedimentary layering structure. In the following, a constant value of $V_P/V_S = 2$ is thus assumed.

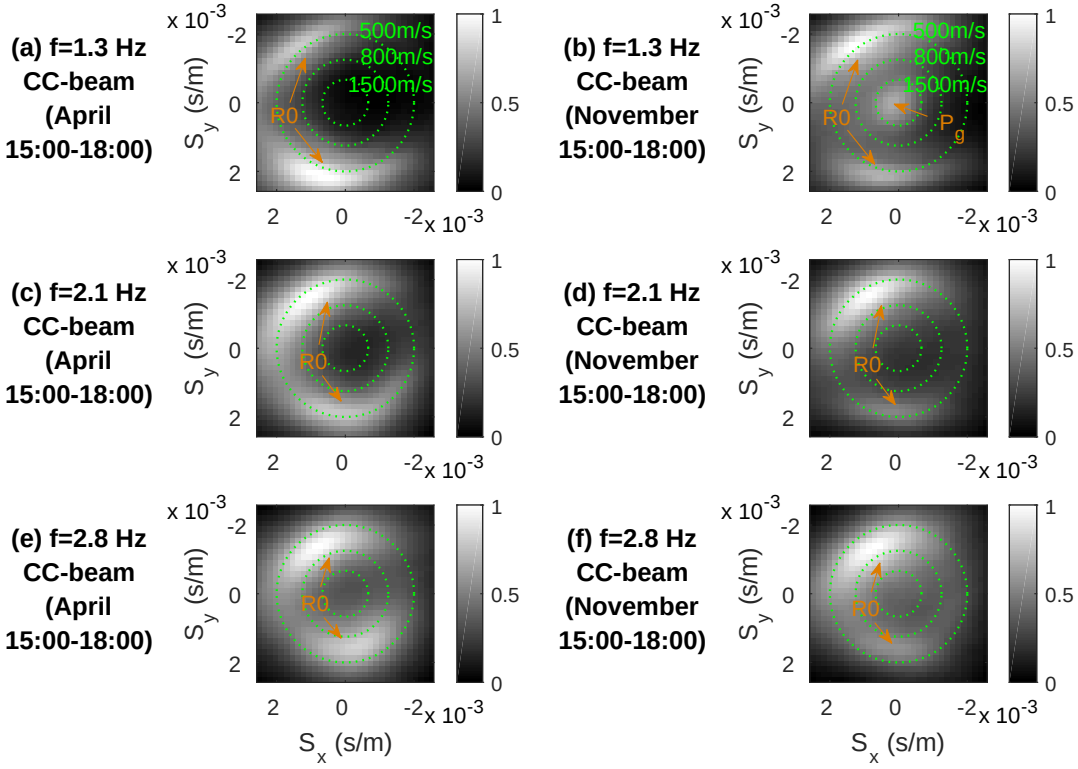
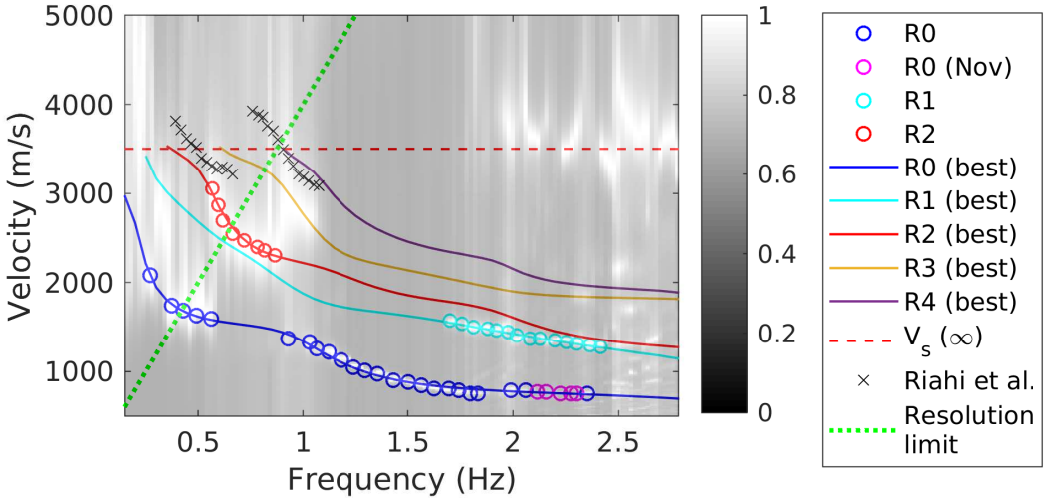


Figure 3.10: Same as Fig. 3.7, but for day-time. Only the modes detected in the beam-former are labeled.

(a) Dispersion curves



(b) Velocity model

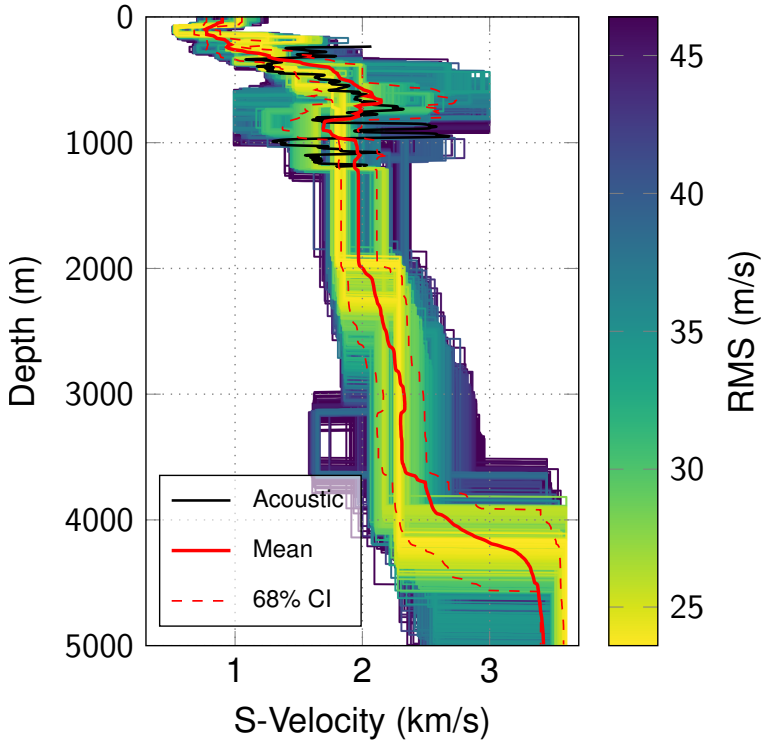


Figure 3.11: Dispersion curves picking and inversion. (a) Dispersion curves overlaid onto the MUSIC-CC dispersion plot for April (1-4 AM); circles : picked curves; plain pines : forward modeled dispersion curves for the best model; dashed line : high-frequency velocity limit for the L_g phase, equal to V_S in the crust, considered as an infinite half space during the inversion; crosses : dispersion curves from Riahi et al. (2013a). (b) Inverted mean velocity profile (red) along with the acoustic log provided by Storengy (black). The velocity models sampled by the different runs are represented in the background with the colorscale indicating their RMS value. The red dashed lines delimit the 68% confidence interval.

A subset of the best models, the mean model and the 68% confidence interval are shown in Fig. 3.11b. The mean V_S model is consistent with the sonic log available from the CS07 well (drilled in 1967 by Gaz de France, see well location in Fig. 3.1a) between 230 and 1190 m (V_P converted to V_S assuming $V_P/V_S = 2$). The results of the inversion indicate the presence of a fast layer between 550 and 800 m. The basement is found at approximately 4 km depth. Since our result is an averaged measure over the array area, it can be only qualitatively compared to individual log data. The fast layer probably corresponds to the thick limestone level visible on Fig. 3.1b between 450 and 800 m. On the other hand, the basement depth below 4000 m is consistent with the fact it was not seen in the CS01 log, as the well drilling stopped at 2623 m. Besides, the theoretical dispersion curves for the first five modes estimated for the best model are shown with plain lines in Fig. 3.11a. The modes R0, R1 at high frequency and R2 exhibit a good fit with the picked dispersion curves. R1 at low frequency and R3 are also compatible with the dispersion plot, although they were not used for the inversion.

3.5 Discussion

3.5.1 Ambient noise sources

Anthropogenic or natural?

Below 1 Hz both modal composition and back-azimuth distribution remained stable between day and night periods (Figs 3.6 and 3.9). This suggests an exclusively natural origin for these waves, with source zones located in the Atlantic. Above 1 Hz, though natural sources can show diurnal variations, ambient noise amplitude often follows the human activity cycles, usually weaker at night, on week-ends, and on holidays (e.g. Lehujeur et al., 2015). Such cycles can be spotted in Fig. 3.2: above 1 Hz the power spectral density is clearly lower during the night, at midday and on week-ends (e.g. Fig. 3.2b, see Sunday, 7th. November 2010). At these frequencies, results from section 3.4.3 showed a strong variation of the noise content between day and night in terms of both modal content and back-arriving azimuths. The Rayleigh wave fundamental mode originates from the South during night-time (Fig. 3.7). During day-time, its back-azimuth distribution clusters around NNW and S directions (Fig. 3.10). The day-time back-azimuth distribution might be due to the noise from the cities of Blois and Tours for the North-Western directions, and the A85 motorway close to the Southern end of the seismic network. Interpreting the night-time distribution is more ambiguous. Rayleigh waves originating from the South at frequencies near 1 Hz were also observed by Lehujeur et al. (2017a) near Strasbourg (Eastern France), which they interpreted as microseisms arriving from the Mediterranean sea. However, we cannot exclude the existence of industrial facilities with persistent night-time activity, or a night-time traffic on the A85.

As P_g and L_g phases are not visible during the day, we can infer they are not linked to anthropogenic activity. Taking into account the back-azimuths of the two quick phases (Figs 3.6g to j), the strongest source zones (from N to SW) are located within the Atlantic Ocean direction. The beam-former for the L_g phase is more wide and diffuse than for P_g, suggesting other generation zones (such as the Scandinavian

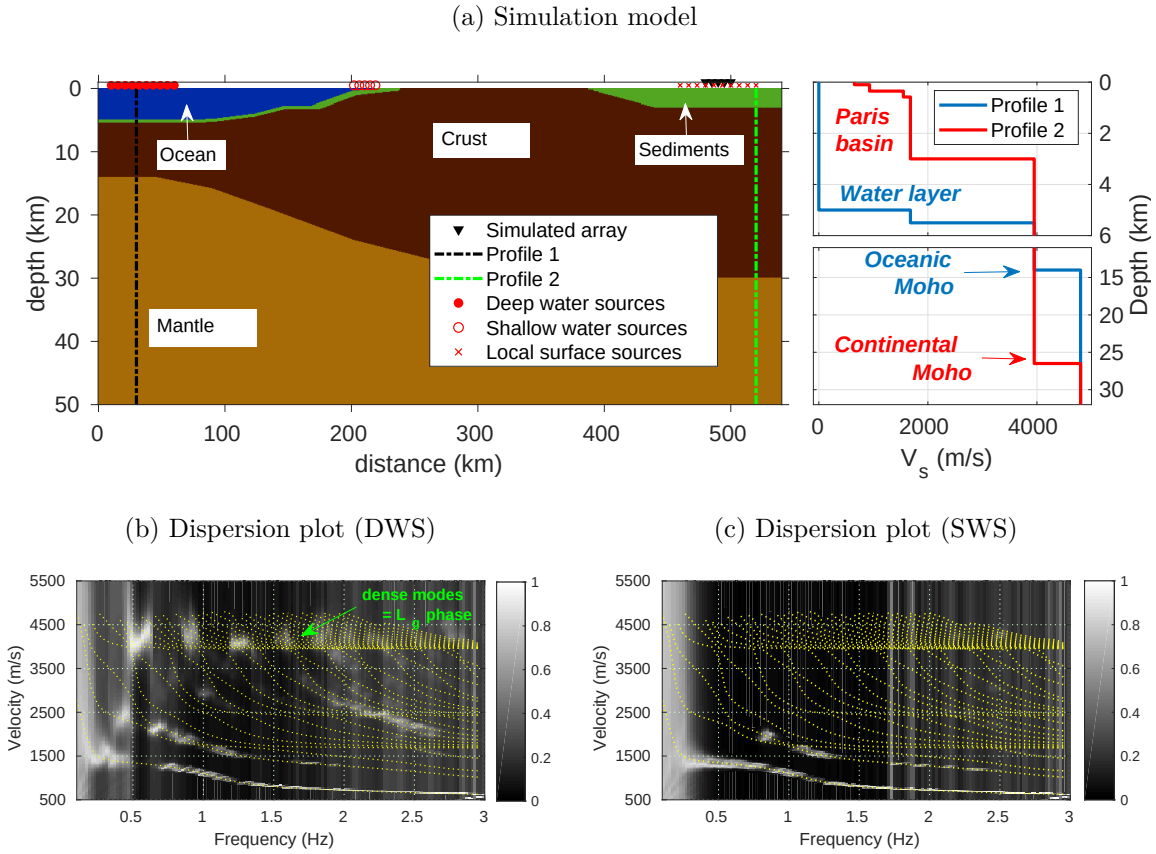


Figure 3.12: Spectral element simulation results. (a) 2D Model used for simulations. Red filled and unfilled dots respectively represent deep- and shallow-water oceanic sources. Red crosses represent local anthropogenic sources. Black and green dash-dotted lines indicate the oceanic and continental profiles for which shear velocity is displayed on the right. Black triangles show the location of the simulated seismic array. (b) One realization of the log-normal distribution used for the dominant frequency of the oceanic sources. (b) Dispersion plot obtained by applying MUSIC on 30 independent realisations of the wavefield with local and deep-water sources. (c) Same as (b), but with local and shallow-water sources.

Northern coastline) and/or strong scattering.

It is interesting to note that the observations made by Riahi et al. (2013b), who analysed differences between day- and night-time cross-correlations on the Jonas field in the USA, could also be interpreted in terms of alternating natural and anthropogenic noise domination.

Excitation of L_g and higher modes

So far, we are in line with the conclusions of Koper et al. (2010), according to which L_g is of natural (oceanic) origin, based on back-azimuth analysis from several arrays around the world. The seismic wave generation mechanism via non-linear ocean waves interactions (Longuet-Higgins, 1950) is indeed valid beyond the traditional secondary microseism band. For instance, the ambient noise amplitude above 1 Hz was well explained in terms of oceanic short-wave interaction for some mid-ocean

islands (Gimbert & Tsai, 2015). However, a major issue is to know if the L_g phase is likely to be excited by such sources, knowing that this wave cannot propagate inside the oceanic crust (Zhang & Lay, 1995). This is because of the nature of the L_g phase, which is a superposition of high-order overtones of both Rayleigh and Love waves (Kennett, 1986). These overtones are due to the crust/mantle interface (Moho), and their number increases with the thickness of the crust. The too thin oceanic crust does not offer enough available overtones for L_g to develop. We used numerical simulation in order to check whether it is realistic to observe an L_g phase arrival on the continent from oceanic sources. Simple elastic 2D spectral-element simulations of the wave propagation from the ocean towards the continent across a passive margin representative of the French Atlantic coast were performed. We used exactly the same profile of the margin as Gualtieri et al. (2015). Compared to their work, we added the mantle below the crust, in order to enable the existence of the L_g phase. The approximate mantle depth was taken from Artemieva & Thybo (2013). The model is sketched in Fig. 3.12a. In terms of seismic velocities, no difference was made between continental and oceanic crust. Sedimentary layers were added below the array in order to mimic the Paris basin. The maximum basin depth was of 3 km, which is shallower than the results from the dispersion curves inversion, but more representative of the Paris Basin. The wavefield generated at the array by deep- or shallow-water natural oceanic sources (respectively called DWS and SWS), modeled separately, was mixed with the one generated by local surface sources (LSS), reproducing human activity. For each source type, 500 vertical point-force sources were used, with dominant frequency, emission time and position randomly chosen within defined ranges. Details about seismic source implementation and wavefield mixture are given in Appendix 3.D.

The resulting dispersion plots for DWS and SWS are shown in Figs 3.12 b and c, respectively. The theoretical dispersion curves for the first 30 modes are overlaid. They were estimated using *Computer Programs for Seismology* software (Herrmann, 2013) for the profile 2 in Fig. 3.12a. The dispersion plot for DWS exhibits patches of energy into the L_g phase band, where the higher modes are concentrated. Such patches are present up to 3 Hz. At some frequencies, individual higher modes are excited (R1/R2 below 1 Hz, R10/R11 between 2 and 2.6 Hz). There is a gap into the R0 branch between 0.5 and 0.7 Hz, where higher modes dominate. A similar pattern is observed for the real data (see Fig. 3.5). On the other hand, the dispersion plot for SWS exhibits a continuous R0 excitation over the entire frequency range. R1 is excited between 0.8 and 1.4 Hz, and between 1.7 and 2.4 Hz. No higher modes above R1 are excited. From these observations, we can confirm the possible existence of the modes detected into the real data set, except for P_g. While the first two modes R0 and R1 can be excited by any type of sources (local and oceanic), the L_g phase and higher modes are specific to distant deep-water sources. From the wavefield snapshots not shown here, we saw that the L_g phase excitation took place as the wavefield reached the continental margin. Before reaching the margin, the wavefield propagates as a superposition of several individual modes specific to an oceanic environment with a water layer on its top, as described by Gualtieri et al. (2015). On the contrary, coastal shallow-water merely excites the fundamental mode. Such different excitation could probably be explained through modal summation, as done by Gualtieri et al. (2015), but taking the mantle into account. This however is out of the scope of our work.

P_g was not observed in the numerical simulations. This can be due to the fact we neglected viscous attenuation, which is higher in shallow layers, and thus tends to enhance phases propagating in depth, as does P_g. Unfortunately, the CPU-cost of a large scale visco-elastic simulation with high-frequency content was prohibitive. Hence, we are unable to discriminate whether P_g is as a signature of DWS or SWS. Another explanation is the simple homogeneous model used to represent the crust: introducing lower velocities layers could help guide P_g waves near the surface. Besides, the L_g phase can be spotted down to 0.5 Hz in the simulations, while it disappears below 2 Hz in the data (Figs 3.5e and f). This might be due to a weaker modal density at low frequencies (Fig. 3.12b), down to some threshold frequency beneath which the weakening L_g phase gets over-shaded by stronger phases.

An interesting development would be to find a robust way to determine the true amplitudes of the P_g and L_g phases, and to look for correlations with particular source zones, provided an ocean wave model for short waves analogous to WAVEWATCH III (Tolman, 1991), as done for example by Essen et al. (2003). However, reliably measuring of the non-linear interaction term for short waves remains a challenge (Peureux & Ardhuin, 2016).

3.5.2 Methodology and inversion

Back-azimuth retrieval

Compared to Riahi et al. (2013a), we were able to extend the frequency band of the wavefield analysis towards higher frequencies by using incoherent beamforming with the envelope of cross-correlations. For frequencies 0.2, 0.54 and 0.81 Hz, where the beam-formers were visualised in their work, our back-azimuth distribution is similar to theirs (NW below 0.3 Hz and wide range NNW-S at 0.5 and 0.8 Hz, see our Fig. 3.6 and their fig. 6). Nevertheless, the beam-formers we derive at 0.5 and 0.8 Hz (Figs 3.6e to h) distinguish between two modes at each frequency, while their beam-former at 0.54 Hz displays a single patch for Rayleigh waves. At 0.81 Hz, their dominant identification matches what we identify as R3 with similar back-azimuths. Another small patch of prograde Rayleigh waves, which is visible in their fig. 6 at $f = 0.81$ Hz, is compatible with our mode R2, but with limited back-azimuth match, as they only find a Southerly direction, while we have a wider back-azimuth distribution.

An intrinsic limitation arises in the statistical analysis carried out by Riahi et al. (2013a) or Koper et al. (2010), where only one maximum is picked from the beam-former per time window. If the wavefield is steady enough over time, the dominant phases will almost always be picked and the histogram of detections (e.g. fig. 6 in Riahi et al. (2013a)) will systematically miss weaker phases. This might be the reason why we observe more modes compared to Riahi et al. (2013a). On the other hand, we were not able to separate different polarisation states with our one-component method, contrary to their three-component beamforming approach.Other non-traditional array methods, such as CLEAN-PSF (e.g. Gal et al., 2016), remove iteratively the contribution of main seismic phases from the cross spectral matrix, which enables an efficient recovery of weaker phases.

Dispersion curves retrieval

We implemented dispersion curves retrieval both directly from the recorded waveforms (MUSIC, Figs 3.5g and h) and from the cross-correlations (FK-CC: Figs 3.5c and d / MUSIC-CC: Figs 3.5e and f). We found that MUSIC-CC was the most appropriate method, since it works in both sub-resolution and aliasing zones. This is in keeping with the conclusions drawn in the synthetic study by Gouédard et al. (2008): using cross-correlations enables to widen the aliasing-free zone, while high resolution methods (HRFK was used in Gouédard et al. (2008)) yield better results beyond the resolution limit. The benefit was crucial for the inversion. Indeed, only MUSIC-CC was able to resolve R0 and R1 below 0.5 Hz. Other methods exhibited a wide patch including both modes. A natural identification of this patch as R0 would yield a model with an erroneous (too shallow) basement position. The implementation of a stable automated detection of the signal subspace dimension for the MUSIC algorithm greatly improved the results, as shown in Appendix 3.B.

Dispersion curves from Riahi et al. (2013a) were also reported in Fig. 3.11(a) (black crosses). Their identification by the authors as higher modes is confirmed by our results. Still, the phase velocities picked by Riahi et al. (2013a) exceed the L_g phase high-frequency velocity limit (3500 m/s). One possible explanation is that the higher modes are influenced by the mantle, as suggested by the theoretical dispersion curves in Fig. 3.12b. Otherwise, velocities can be biased as they lie beyond the theoretical resolution limit for standard beamforming, or close to it. Velocity over-estimation into the sub-resolution zone was also highlighted by Gouédard et al. (2008).

Finally, it should be pointed out that the dispersion curves from Riahi et al. (2013a) are representative of the isotropic part of the wavefield, which can explain some discrepancy between their higher-mode dispersion curves and ours. As they found a non-negligible level of anisotropy (up to 10%), it might be useful to introduce a correction for anisotropy in our methodology. This however is beyond the scope of this paper.

3.5.3 Benefits of wavefield characterisation

For surface-wave based methods, the wavefield characterisation yields a major indication on the mode labels to be used according to the location of the pointed dispersion curve into the $f-v_\phi$ plane. For regionalised inversion methods based on ambient noise cross-correlation, it would allow to choose reliable station-pairs aligned with the major source direction. The knowledge of the L_g phase velocity at relatively high frequencies (above 2 Hz) yields a useful constraint on V_S into the infinite half-space during the inversion. For body-wave methods, wavefield characterisation would give an indication about the frequency bands where they are most likely to be found (1-1.5 Hz for the present data set). Whatever the targeted wave-type, the knowledge of the wavefield composition evolution over time is useful for selecting the appropriate time windows to be processed. For example, using day-time records would be preferable for Rayleigh wave dispersion curves extraction above 1 Hz, both in terms of strength and source distribution, while night-time records would allow to extract body waves between 1 and 1.5 Hz. Finally, for methods with unclear wave-type assumptions, such a characterisation of the incident wavefield should guide the numerical simulation. For example, Lambert et al. (2013) modeled the ambient wavefield as a superposition

of randomly excited deep and surface sources, in order to investigate the nature of the hydrocarbon-related amplitude anomalies. Such a setup completely misses the L_g phase, which however seems to carry a significant amount of energy at the frequencies of interest (several Hz). A model similar to the one shown in Fig. 3.12a would better approach the reality, though at a higher computational cost.

Acknowledgments

The authors are grateful to Jérôme Vergne, Maximilien Lehujeur, Nikolaï Shapiro, Eleonore Stutzmann and Fabrice Ardhuin for helpful discussions, as well as to Catherine Formento and Simon Lejart (Storengy) for their help with the well logs. They are also grateful to Storengy, an Engie company, for funding this research, and to the reviewers for their comments and advice.

3.6 Conclusion

We suggest to combine methods as following for a wide-band one-component ambient wavefield characterisation with a dense array:

- Back-azimuth retrieval: «direct» MUSIC in the sub-resolution and working zones of the array, and non-coherent cross-correlation beamforming in the aliasing zone (CC-beam).
- Dispersion curves retrieval and mode labeling: MUSIC applied to a common-offset gather of the inter-station cross-correlations (MUSIC-CC).

This methodology was applied to the vertical component recordings acquired above an Underground Gas Storage in Central France by a 3x6 km array with about 100 sensors. Between 0.1 and 3 Hz, 4 Rayleigh waves modes (R0, R1, R2, R3) were identified, as well as L_g and P_g phases. The comparison of day-time and night-time recordings made it possible to distinguish between anthropogenic seismic sources (mainly R0 and R1 above 1 Hz) and natural seismic sources (higher Rayleigh modes, P_g and L_g) in deep ocean and along the Atlantic coast. Numerical simulations were carried out to confirm this interpretation. Eventually, surface wave dispersion curves were inverted to yield a V_s profile consistent with the available sonic log, and compatible with the presence of a Permian basin in this area of the Paris Basin. Given the growing interest of the scientific and industrial communities in seismic noise, our work opens interesting perspectives. We developed a robust technique to analyse data recorded by dense arrays which are ever more used in the industry. A direct application are 1D V_s inversions, as our method helps to correctly identify and label different surface wave modes. Furthermore, having a precise knowledge of the incoming wavefield is of prior importance when studying possible links between hydrocarbon reservoirs and amplitude features at surface.

Appendix

3.A Group and phase velocity

The consistency between group and phase velocities measured respectively by non-coherent cross-correlation beam-forming and MUSIC-CC is investigated. Knowing the phase velocity v_ϕ, the group velocity v_g can be computed as

$$v_g = v_\phi \left(1 - \frac{\omega}{v_\phi}\frac{\partial v_\phi}{\partial \omega}\right)^{-1} . \tag{3.A.1}$$

Relation (3.A.1) is applied to the smoothed interpolated dispersion curve of the R0 mode (blue line in Fig. 3.A.1) picked from the MUSIC-CC dispersion plot (see Fig. 3.11(a)). The result is again smoothed and plotted in Fig. 3.A.1 (red line). This is the expected group velocity from the phase velocity measurements. On the other hand, CC-beam was applied to April's night-time cross-correlations between 0.7 and 2.7 Hz, with a step of 0.2 Hz. For each obtained beam-former, a Gaussian was fitted to the azimuthal section crossing the beam-former's maximum. This yields an estimate of the group velocity (Gaussian's mean) and the associated uncertainty (Gaussian's standard deviation) at each considered frequency. Results were reported on Fig. 3.A.1 (red circles with uncertainty bars). Group and phase velocity measurements are clearly consistent with each other.

The same velocity conversion procedure was applied for the R1 mode in order to place the arrows in Fig. 3.7 at the expected group velocity for this mode. However, Fig. 3.A.1 confirms that only the fundamental mode (R0) is detected by the CC-beam approach.

3.B MUSIC and automated signal subspace determination

In this section we present the derivation of the MUSIC algorithm and a method to determine the dimension of the signal subspace.

We drop the frequency dependence from the notations, simply writing R instead of $R(f)$ for the cross spectral matrix introduced in equation (3.5). At a given frequency, if a wavefield containing Q uncorrelated plane waves of amplitudes A_q is recorded by N stations affected by white noise of intensity η (e.g. instrumental noise), the CSM can be written as

$$R_{ij} = \sum_{q=1}^{Q} |A_q|^2 e^{i\mathbf{k_q}\cdot(\mathbf{x_i}-\mathbf{x_j})} + \eta^2 \delta_{ij}, \tag{3.B.1}$$

where δ is the Kronecker symbol. Assuming $N > Q$, R can be diagonalised as following:

$$R = E_s \Lambda_s E_s^\dagger + E_n \Lambda_n E_n^\dagger , \tag{3.B.2}$$

with Λ_s being a $Q \times Q$ diagonal matrix containing the Q biggest eigenvalues, and Λ_n a $(N-Q)\times(N-Q)$ diagonal matrix containing the $(N-Q)$ remaining eigenvalues.

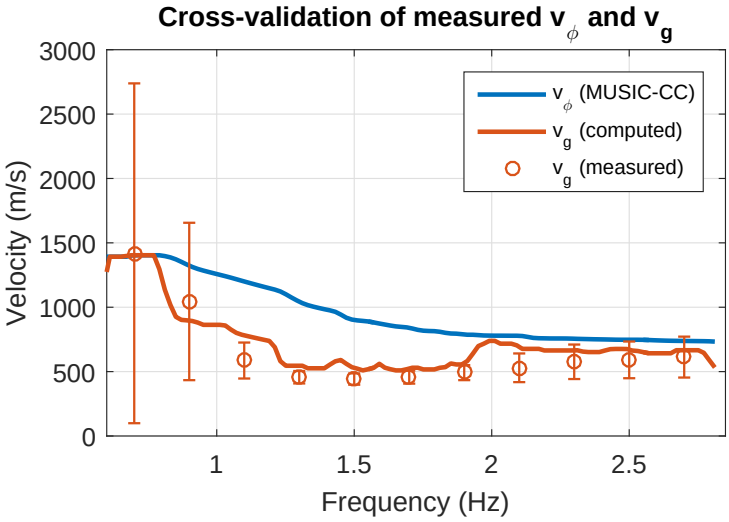


Figure 3.A.1: Phase and group velocity consistency check. Blue curve - interpolated and smoothed phase velocity dispersion curve for the fundamental mode, picked from the MUSIC-CC dispersion plot for April (night-time). Red curve - smoothed theoretical group velocity dispersion curve computed from the blue curve using (3.A.1). Red circles with error bars - group velocities measured by CC-beam (April, night-time), with associated uncertainties.

Using simple linear algebra, Schmidt (1986) showed that $\Lambda_n = \eta^2 I$. The orthonormal eigenvectors e_l which form the noise space basis (columns of E_n) then verify

$$a(\mathbf{k_1})^\dagger e_l = \ldots = a(\mathbf{k_Q})^\dagger e_l = 0 \ , \tag{3.B.3}$$

with

$$a(\mathbf{k}) = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{\mathbf{ik\cdot x_i}} \\ \vdots \\ e^{\mathbf{ik\cdot x_N}} \end{pmatrix} . \tag{3.B.4}$$

That is, the vector $a(\mathbf{k_q})$ is orthogonal to the noise subspace generated by the columns of E_n if $\mathbf{k_q}$ corresponds to a plane wave which actually propagates across the array. The principle of the MUSIC algorithm is to seek such optimal $\mathbf{k}$ vectors by minimising the projection of $a(\mathbf{k})$ on the noise subspace, and thus maximising the so-called MUSIC *functional*

$$D_M(\mathbf{k}) = \frac{1}{\left|\sum_{l=1}^{N-Q} a(\mathbf{k})^\dagger e_l\right|^2} = \frac{1}{a(\mathbf{k})^\dagger E_n E_n^\dagger a(\mathbf{k})} . \tag{3.B.5}$$

In comparison, the classical FK method (or standard frequency domain beamforming) differs insofar as it seeks the vector $\mathbf{k}$ maximising

$$D_{FK}(\mathbf{k}) = a(\mathbf{k})^\dagger R a(\mathbf{k}) \ , \tag{3.B.6}$$

without performing any diagonalisation.

Automated ways to choose the right dimension of the signal subspace for MUSIC, ie Q, use the eigenvalues profile (Cornou, 2002). One possible approach is to look

for the slope break into the logarithmic decay of the eigenvalues, while another one relies on comparing the slope of the logarithmic decay to the one obtained for random noise. Both approaches require threshold parameters that depend on the signal-to-noise ratio. Since the latter is unknown for the real ambient noise data, we developed a slightly different method which allows to address frequency ranges with high and low signal-to-noise ratios at once. A criterion on the eigenvalue ratio $R_i = \log \frac{\lambda_i}{\lambda_1}$, with λ_1 being the major eigenvalue, is introduced:

$$\begin{cases} C_R(i) = 1 \text{ if } |R_i| \leq n_R \\ C_R(i) = 0 \text{ if } |R_i| > n_R \ , \end{cases} \tag{3.B.7}$$

where n_R is a threshold parameter to be optimised. The threshold eigenvalue number is the defined as

$$i_{\text{mag}} = \max \{i | C_R(i) = 1\} \ . \tag{3.B.8}$$

This simply means that the first i_{mag} eigenvalues have a non-negligible magnitude compared to the major eigenvalue. On the other hand, the local logarithmic slope of the eigenvalue decay is defined as

$$S_\lambda(i) = \tan^{-1} \left\{ \log \left(\frac{\lambda_{i+1}}{\lambda_i} \right) \right\} \ . \tag{3.B.9}$$

The slope break corresponds to the maximum value of the slope derivative with respect to the eigenvalue number:

$$i_{\text{slope}} = \text{argmax} \left(\left\| \frac{\partial S_\lambda}{\partial i} \right\| \right) \ . \tag{3.B.10}$$

We choose to define the signal subspace dimension as

$$n_s = \max(i_{\text{slope}}, i_{\text{mag}}) \ . \tag{3.B.11}$$

The maximum signal subspace dimension is limited by the CSM smoothing, as explained in section 3.3.3. In order to take this into account, the CSM smoothing and diagonalisation is first applied to pure white noise of same duration and sampling as the analyse signal. The resulting eigenvalue profile typically exhibits a strong jump, which is detected using the slope break criterion defined above. The number of the eigenvalue corresponding to the this jump ($i_{\text{slope}}^{(\text{noise})}$) is considered to be the maximum available signal subspace dimension for the given CSM smoothing procedure.

Simple synthetic tests were performed with a linear array analogous to the virtual shot gathers shown in Figs 3.5a and b. Dispersive phases with known phase velocities and known SNR were propagated in frequency domain following the procedure described in Appendix 3.C. A misfit function was defined for the dispersion plots and the SNR estimated by the MUSIC algorithm. A grid search was performed to determine the number K of sub-arrays and the frequency bandwidth Δf used to smooth the CSM (see sections 3.3.3 and 3.3.3), and the eigenvalue magnitude threshold parameter n_R. After repeated tests with different SNR levels, the values $K = 20$, $\Delta f = 0.1$ Hz and $n_R = 2$ were chosen.

For the real-data virtual shot gather (April - night time), the dispersion plot obtained with the automated signal-subspace determination is compared to the one

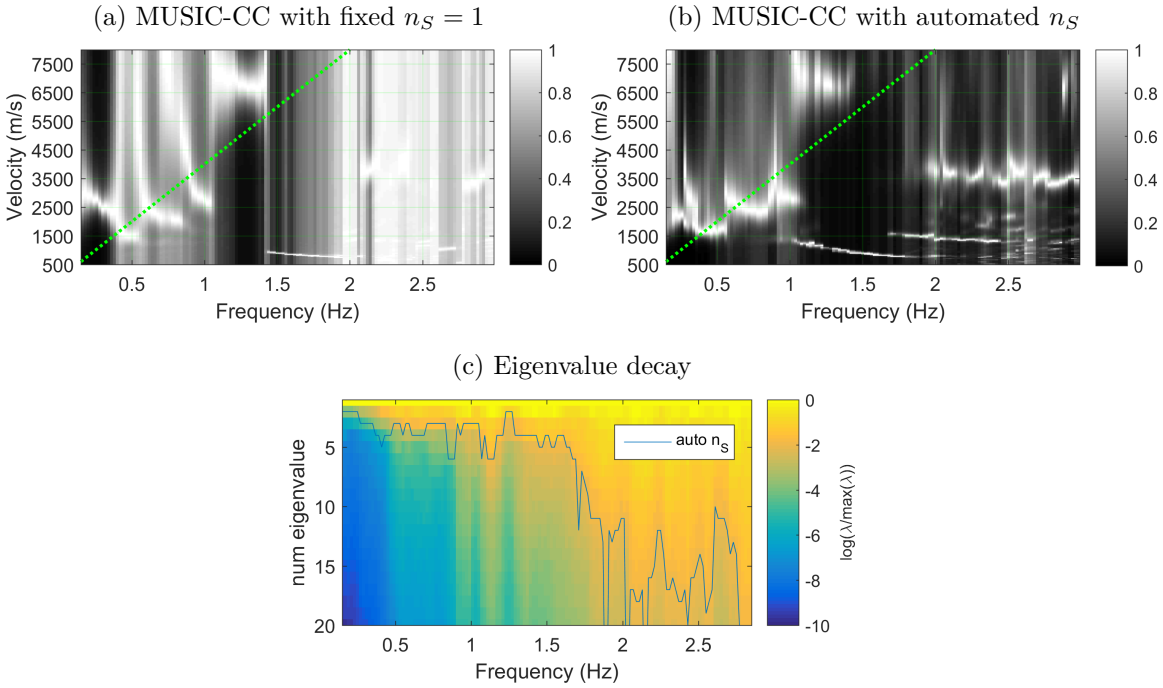


Figure 3.B.1: MUSIC-CC dispersion plot obtained for April (night-time data) with (a) fixed $n_S = 1$; (b) automated n_S determination (same as Fig. 3.5e). The green dotted line shows the resolution limit. (c) Eigenvalue magnitudes (logarithmic ratio to the major eigenvalue) at different frequencies (colored image). The blue curve shows the determined n_S as function of frequency.

obtained assuming $n_S = 1$ in Fig. 3.B.1. The latter appears much noisier and doesn't distinguish between R0 and R1 below 0.5 Hz. R0 is also not identified between 1 and 1.5 Hz. Eigenvalue magnitudes are displayed in Fig. 3.B.1c. The low frequencies (below 1 Hz), characterised by a high SNR, exhibit only several strong eigenvalues, and are governed by the slope break criterion. Higher frequencies (above 1 Hz), characterised by a smaller SNR, exhibit a more gradual eigenvalue decay and are governed by the eigenvalue magnitude criterion $n_R = 2$. More eigenvalues are thus kept: plane waves are indeed «spread» over several eigenvalues, which must all be included into the signal subspace to properly retrieve the correct phase velocity.

3.C Synthetic tests of the MUSIC algorithm

In this appendix, the MUSIC algorithm with automated signal subspace dimension determination is tested on synthetic signals with known back-azimuth and dispersion relation. Both MUSIC-CC (linear array - single time window) and «direct» MUSIC (2D array - several time windows) are implemented and compared.

Synthetic plane waves recorded by an array of sensors with positions $\mathbf{x_i}$ were generated in Fourier domain with respect to a fictitious source[1] located at $\mathbf{x_s}$, taken

[1]We use the term «fictitious» because a strictly plane wave cannot have any localised source. Here the «source» is only used as a reference point for fixing a back-azimuth and a phase-delay.

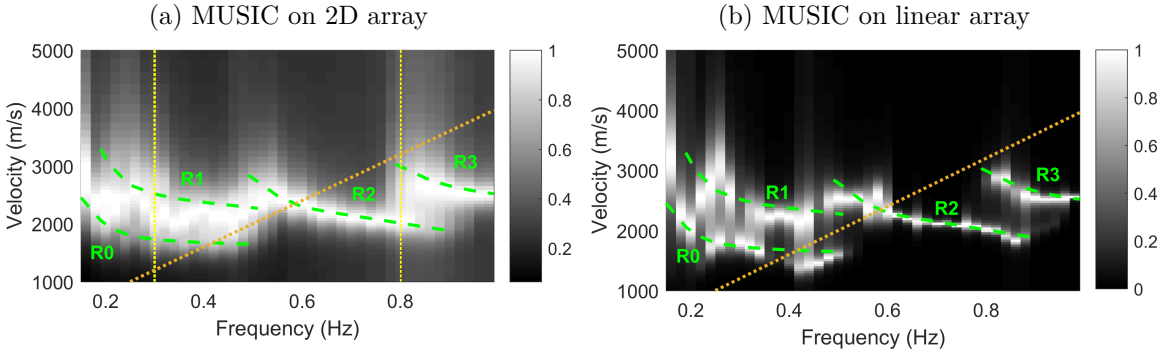


Figure 3.C.1: Dispersion plots of synthetic signals. (a) 2D array with random source parameter distributions from Table 3.C.1 - yellow dotted lines indicate the frequencies at which beam-formers are plotted in Fig. 3.C.3; (b) linear array with a single shot. Brown dotted lines show the resolution limit for both arrays.

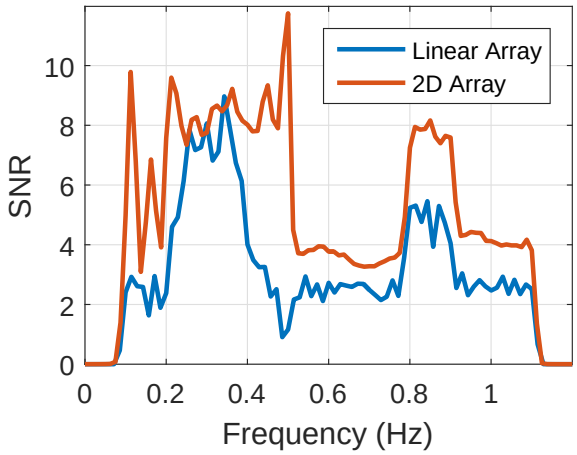


Figure 3.C.2: SNR in the simulated seismograms. Blue line - linear array; red line - 2D array

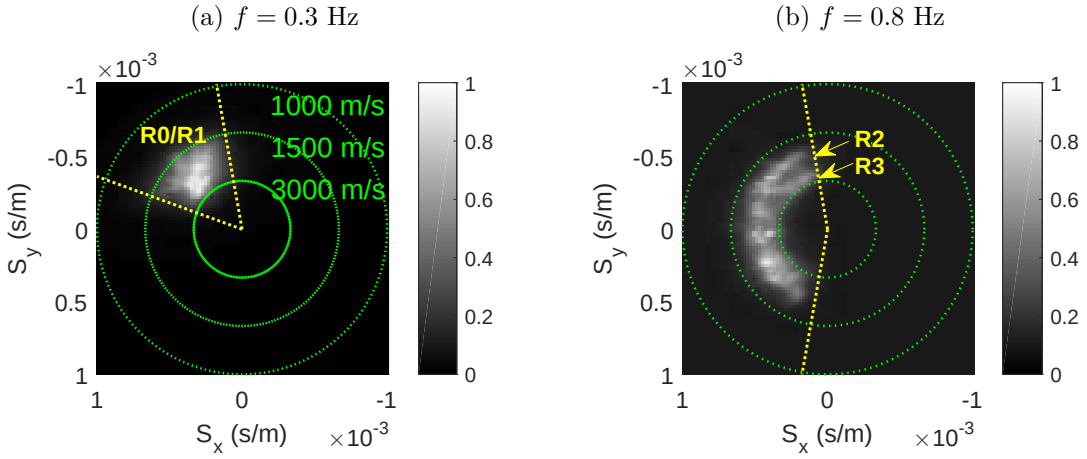


Figure 3.C.3: MUSIC beam-formers for synthetic signals. (a) f=0.3 Hz; (b) f=0.8 Hz. Yellow dotted lines indicate the bounds of the back-azimuth uniform distribution (a) for R0-R1, and (b) for R2-R3.

Table 3.C.1: Frequency and back-azimuth bounds of the simulated dispersive phases for MUSIC synthetic tests

Parameter	R0	R1	R2	R3
f (Hz)	$[0.05, 0.5]$	$[0.2, 0.5]$	$[0.5, 0.9]$	$[0.8, 1.1]$
BAZ (°)	$[-70; -10]$	$[-70; -10]$	$[-170, -10]$	$[-170, -10]$

as the time reference:

$$u(\omega, \mathbf{x_i}) = s(\omega) \exp\left[\mathrm{i}\omega\left(t_0 - \frac{(\mathbf{x_i} - \mathbf{x_s}).\mathbf{e}_\theta}{v_\phi(\omega)}\right)\right], \tag{3.C.1}$$

where ω is the angular frequency, $s(\omega)$ is the source spectrum, t_0 is the emission time at $\mathbf{x_s}$, $\mathbf{e}_\theta$ is the unit direction vector along the wave's back-azimuth θ, and $v_\phi(\omega)$ is the phase velocity dispersion relation. The fictitious source location was always chosen at a distance of 10 km from the array's center, in the direction of the wave's back-azimuth. Time domain seismograms were obtained by inverse Fourier-transformation. They were then re-sampled at 100 Hz as for the real data set. The array configuration shown in Fig. 3.1a was used for synthetic tests. Four distinct dispersive phases (green dashed curves in Fig. 3.C.1) were simulated below 1 Hz in order to roughly reproduce those identified as R0, R1, R2 and R3 in Fig. 3.5(e). 50 time windows of 100 seconds were generated. Each time window contained signals generated by 10 fictitious sources with back-azimuths following a uniform distribution, with bounds given in Table 3.C.1. Each fictitious source emitted 10 wave trains reaching the array within 100 s, with random emission times. The source spectrum associated to each wave train was a cosine tapered window with bounds given in Table 3.C.1. Random white noise was added to the synthetic signals, with an average resulting SNR ratio between 4 and 8 (Fig. 3.C.2, red curve). The MUSIC algorithm (section 3.3.3) was then applied to the resulting 50 time windows as described in Section 3.3.

3.C.1 Dispersion curves retrieval

The mean dispersion plot (Fig. 3.C.1a) does not allow to separate $R0$ and $R1$, while R2 and R3 are retrieved with the correct phase velocities. For comparison, we propagated the same dispersive phases on a linear array of 7000 m aperture with 100 m spacing to reproduce the virtual shot gathers cross-correlations would yield. Only one wave train was excited for each dispersive phase, all of them emitted simultaneously by the same fictitious source. The latter is aligned with the array, and located 200 m apart from the closest receiver. Random white noise was added to the simulated wavefield, so that the resulting SNR was close to the one used for the 2D array (Fig. 3.C.2, blue line). The MUSIC algorithm was applied to this single noisy shot gather, as explained in section 3.3.3. The obtained dispersion plot, shown in Fig. 3.C.1(b), was able to separate all the four dispersive phases. MUSIC applied to a single shot gather has thus a better resolution power than MUSIC applied to several time windows recorded

Table 3.D.1: Truncation bounds used for the source parameter uniform distributions in spectral-element simulations

Parameter	DWS	SWS	LSS
x_s (km)	$[10, 60]$	$[202, 219]$	$[460, 520]$
f_0 (Hz)	(log-norm)	(log-norm)	$[1, 3]$
t_0 (s)	$[20, 520]$	$[20, 440]$	$[5, 190]$

by a 2D array. Hence MUSIC-CC should be preferred to the «direct» MUSIC for dispersion curve retrieval from dense array data.

3.C.2 Back-azimuth retrieval

The MUSIC beam-formers correctly retrieve the back-azimuth distribution, both below and above 0.5 Hz (Figs 3.C.3a and b, respectively). While at 0.3 Hz R0 and R1 are not separated (Fig. 3.C.3(a)), as expected from the dispersion plot analysis above, R2 and R3 at 0.8 Hz are only separated for back-azimuths aligned with the direction of the longest array extent (NNW). Similar observations can be made on the real-data results in Figs 3.6c to f, where the circular patterns typically become more refined in the NNW direction. As stressed by Cornou (2002), the theoretical resolution power of the MUSIC algorithm is asymptotically infinite as the SNR tends to infinity. The true resolution power depends on the SNR. The SNR is however hard to quantify, since both random and coherent noise are present in real data. Thus, the only conclusion we draw is that applying MUSIC allows to obtain reliable results beyond the theoretical resolution limit. However, we are unable to quantify the new resolution power precisely, and suggest a specially designed synthetic study similar to the present one in order to investigate the performance of a given array on given targeted seismic modes.

3.D Spectral-element simulations of the wavefield composition

Synthetics were generated using the spectral-element based SPECFEM2D code (Komatitsch et al., 1999) coupled to Gmsh software used for generating a quadrangular mesh (Geuzaine & Remacle, 2009). The simulations were performed into a vertical 2D plane in order to avoid too costly computations associated with 3D simulations. Absorbing Stacey conditions (provided option in SPECFEM2D) were applied at the left, right and bottom edges of the model. The typical quadrangle size used in different regions, was chosen so that there were at least 4 points per minimal S wavelength at $f = 3$ Hz, which is the maximum frequency analysed. We checked that taking a more refined grid did not affect our results significantly. The time step was set to 3.10^{-4} s, so that the stability condition was fulfilled.

Each simulated source was a vertical point-force source emitting a Ricker wavelet with dominating frequency f_0 and emission date t_0. These parameters along with the source's position followed a truncated uniform probability distribution. An exception was made for the oceanic sources dominating frequency, where we used a log-normal distribution centered on 0.25 Hz with a standard deviation of 0.7 Hz, truncated between 0.1 and 3 Hz. Truncation bounds of the uniform distributions used for each type of sources are listed in Table 3.D.1, with x_s being the source's horizontal location. 30 sets of 500 random realisations of these distributions were used to simulate 30 independent realidations of seismic noise generated by 500 sources of each type. The total simulation times were of 600 s for Deep Water Sources (DWS), 480 s for Shallow Water Sources (SWS) and 220 s for Land Shallow Sources (LSS), allowing the wavefield from the latest excited source to reach the array. Time intervals extracted for processing were [400-580] s for DWS, [300-480] s for SWS, and [20-200] s for LSS, allowing to work with a fully developed wavefield where all possible phases are mixed.

The simulated array of receivers spans over 20 km with a 100 m spacing. It has a bigger offset compared to the real data, which allows to unambiguously resolve the low-frequency content of the simulated wavefield. Particle velocity wavefields recorded by the array, resulting from natural and anthropogenic sources, respectively written through $u_1(x,t)$ and $u_2(x,t)$, were mixed with a rate α. White random noise $\epsilon(x,t)$ was also added to the simulated data. The final signal can thus be written as

$$u(x,t) = u_1(x,t) + \alpha u_2(x,t) + \epsilon(x,t). \tag{3.D.1}$$

The frequency-dependent amplitude ratio between u_1 and u_2 measured at the array being

$$\xi(f) = \sqrt{\frac{\langle u_1^2(x_i,f)\rangle_i}{\langle u_2^2(x_i,f)\rangle_i}}\ , \tag{3.D.2}$$

where the index i runs over the 200 receivers, the relative contributions of the natural sources with respect to the anthropogenic sources and the white noise in the final signal are

$$\eta^{(12)}(f) = \sqrt{\frac{\langle u_1^2(x_i,f)\rangle_i}{\langle (\alpha u_2)^2(x_i,f)\rangle_i}} = \xi(f)/\alpha \text{ , and} \tag{3.D.3}$$

$$\eta^{(1\epsilon)}(f) = \sqrt{\frac{\langle u_1^2(x_i,f)\rangle_i}{\langle \epsilon^2(x_i,f)\rangle_i}}\ , \tag{3.D.4}$$

respectively. While $\xi(f)$ is fixed by the simulation, we choose α so that $\eta^{(12)}(1\text{ Hz}) = 10$, and the white noise amplitude so that $\eta^{(1\epsilon)}(1\text{ Hz}) = 100$. This allows to reproduce the situation where both natural and anthropogenic sources can influence the dispersion plot above 1 Hz, as in the real data. In order to get the dispersion plots in Figs 3.12c and d, we directly applied MUSIC to the simulated linear antenna, and averaged the dispersion plots over the 30 independent realisations of the seismic noise, as described in section 3.3.3.

3.E L_g phase and overtones

In order to better understand the L_g excitation and propagation across the continent, an additional set of numerical simulations involving single sources placed at different locations was performed in this Appendix. It highlights the importance of the L_g phase for exciting overtones in sedimentary basins. This Appendix was published as a conference paper:
Kazantsev, A., Peruzzetto, M., Chauris, H., Dublanchet, P., & Huguet, F., 2018b. Origins Of Rayleigh Wave Overtones In Ambient Noise, in *Seventh EAGE Workshop on Passive Seismic 2018*

Summary

2D spectral element simulations in the frequency range 0.1-3 Hz were performed on a largescale model (1200x50 km). Ocean water, sediments, granitic crust and mantle were included. Two seismic arrays were modeled on the continent in order to analyze the wavefield composition on the granitic crust and above a sedimentary basin. Four source locations were tested: ocean surface in deep and shallow water, ocean bottom and continental surface near the array. Finally, the effect of removing the mantle from the simulation was investigated. The deep-water surface source was able to excite a clear Lg phase, which disappeared when the mantle was removed. The L_g phase, in turn, excited the overtones above the sedimentary basin. The Rayleigh fundamental mode was dominant for the other investigated source locations. This might explain the presence of the L_g phase and Rayleigh overtones in the short-period seismic noise, which is probably a signature of short-period sources near the surface in deep ocean.

Introduction

Based on the available literature, the ambient seismic noise sources are of oceanic / meteorological origin below 1 Hz, and of human origin above 1 Hz (Bonnefoy-Claudet et al., 2006b). Fundamental mode Rayleigh waves are often assumed to be dominant on the vertical component, which can be questionable in some cases (see Bonnefoy-Claudet et al., 2006b, and references herein). In particular, Rayleigh overtones were reported by several authors above sedimentary basins (e.g. Ma et al., 2016). Overtones can yield useful constraints on deep structures for surface wave phase velocity or amplitude inversion since they have a deeper sensitivity. However, their origin is somewhat unclear. While they can sometimes be detected above 1 Hz, they are unlikely to be excited by local artificial surface sources, as shown below. In our work, we use 2D numerical simulations to investigate the possibility of overtone excitation by sources at different locations. Our modelling is guided by an intriguing observation made by Koper et al. (2010), who reported an omni-present L_g phase in the frequency range [0.4-2] Hz, based on noise measurements from 18 arrays around the world. The L_g phase is due to the crust/mantle contact (Moho). It is a result of a high concentration of available surface wave modes (both Love and Rayleigh) that exist in the thick continental crust overlaying a more rigid mantle (Kennett, 1986). Its group velocity usually ranges from 3500 to 4500 m/s, while the phase velocity of the individual modes composing this phase is slightly higher. We investigate if such a

wave can be observed on the continent due to a source located near the water surface in deep ocean, and how the L_g phase can be converted into Rayleigh overtones in a sedimentary basin.

Method

A two-step procedure was used. In a first step, synthetic data were generated using spectral-element method (Komatitsch et al., 1999). In a second step, a high-resolution array-processing method (Schmidt, 1986) was applied to the synthetics in order to retrieve the phase velocity dispersion curves at the array location and to deduce the wavefield structure. Different source positions (ocean surface in deep/shallow water, ocean bottom and earth surface near the array) were tested in order to see which of them were able to excite the L_g phase and/or the Rayleigh overtones. An additional simulation was performed without the mantle in order to check the reliability of the L_g phase identification.

Numerical simulation

In order to allow the existence of the Lg phase, the mantle was included into a large-scale model (1200x50 km, Fig. 3.E.1). The model is a schematic representation of the Atlantic margin at the French Western coast. The shape of both the margin and the oceanic sedimentary basin were taken from (Gualtieri et al., 2015). A continental 1400 m deep sedimentary basin was added to mimic the Parisian basin. The approximate depth of the Moho discontinuity in the region was taken from (Artemieva & Thybo, 2013). A vertical point-force source was placed at four different locations, referred to as s1 to s4, shown by the stars in Fig. 3.E.1. Sources s1 and s2 were located at the water layer surface, while sources s3 and s4 were embedded into the earth. The source time function was always the sum of two simultaneous Ricker wavelets of central frequencies $f = 0.5$ Hz and $f = 1.5$ Hz, so that a wide enough frequency range could be covered. Absorbing boundary conditions were used at the left, right and bottom boundaries. The spectral element simulation was performed on a mesh containing about 863000 quandrangles generated with Gmsh software (Geuzaine & Remacle, 2009). A zoom of the mesh comprising the continental sedimentary basin and some granitic crust is shown in Fig. 3.E.2. The typical quadrangle size used in different regions, listed in Table 3.E.1, was chosen so that there were at least 4 points per minimal S wavelength in all the medium. This was our compromise between accurate wavefield sampling and affordable computation time. Two linear arrays of 200 stations with 100 m spacing and 20 km aperture were simulated. One array was placed on the apparent part of the granitic crust ("crust array"), while another one was located above the continental sedimentary basin ("basin array").

Array processing

The multiple signal classification, or MUSIC, is a high-resolution array technique first introduced by Schmidt (1986). It relies on the eigendecomposition of the cross-spectral matrix, containing the cross-spectra of all the station pairs in the array. A grid search is performed over the wave vector space, whereby the theoretical vibration

Table 3.E.1: Element size and wavefield sampling

Medium	**Element size (m)**	**Points per λ_S at 3 Hz**
Ocean	140	-
Sediments	140	4
Crust	330	4
Mantle	400	4

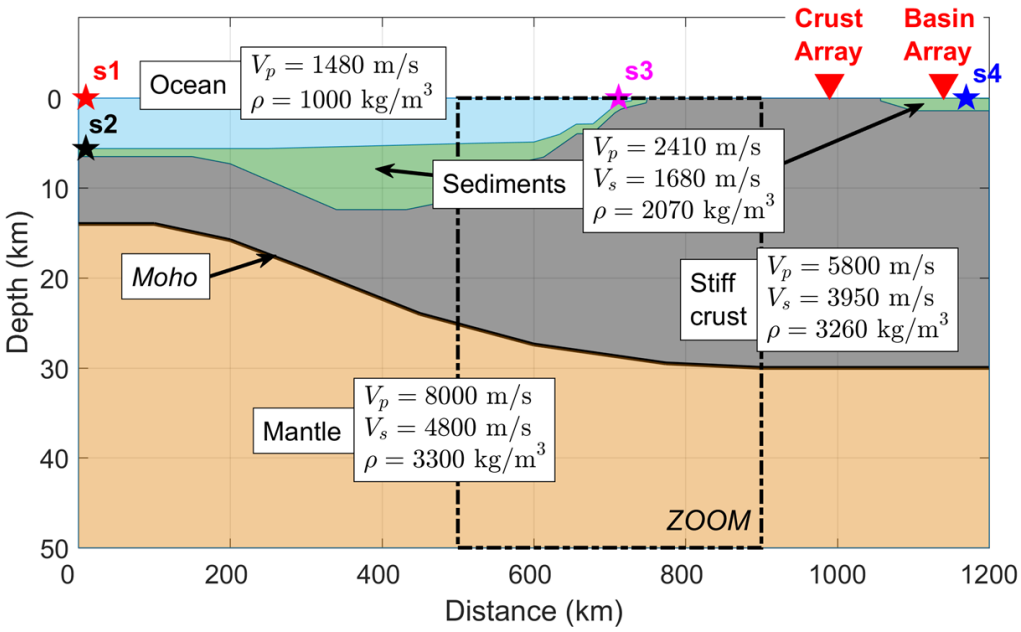


Figure 3.E.1: Model used for spectral element simulations. Stars stand for sources. Blue lines separate regions with different elastic or mesh properties.

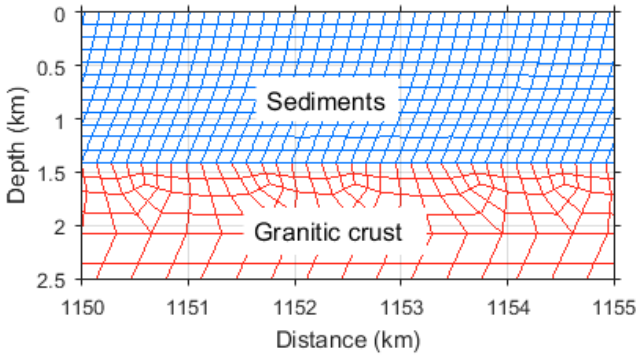


Figure 3.E.2: Zoom on the mesh (continental sedimentary basin).

state associated to each tested vector is projected onto the "noise" subspace. The latter is generated by the eigenvectors associated to the N-M smallest eigenvalues, where N is the number of stations, and M the a priori number of plane waves to be detected. The best estimates of the wave vector minimize this projection, or maximize its inverse (MUSIC functional).

Prior to the eigendecomposition, the cross-spectral matrix was smoothed by averaging it over a sliding 0.1 Hz boxcar window. The MUSIC functional was normalized to its maximum at each considered frequency. Since we worked in 2D, a surface wave vector was uniquely characterized by the phase velocity, so that the output was a dispersion curve. M was empirically fixed to 10, allowing to retrieve several Rayleigh wave overtones. The typical seismogram duration for each simulated shot, from the first P arrival until the end of the surface wave train, was about 10 minutes, sampled at 50 Hz. It was processed at once, without any time-domain windowing.

Results

The MUSIC outputs at both arrays are shown in Fig. 3.E.3. Cases s1 to s4 correspond to the different source positions shown in figure 1. Several observations can be made from these plots:

1. If the source is in deep ocean (s1), Lg phase is excited into the continental crust, visible as a wide band of phase velocities between 3500 and 4500 m/s in Fig. 3.E.3a. Such type of sources at low frequencies about 0.1-0.3 Hz were described in detail by Ardhuin et al. (2011). Pressure fluctuations at the ocean bottom at frequencies up to more than 10 Hz were also reported (see Farrell & Munk, 2008, , and references herein). From the wavefield snapshots, displayed in Fig. 3.E.4, we see that the L_g phase excitation takes place as the wavefield excited at s1 reaches the continental margin (around $t = 550$ s). On the other hand, the modes excited by s2 do not exhibit the characteristic blockage as they enter the thin oceanic crust. The L_g phase excitation from s1 occurs despite the fact the oceanic crust itself does not support the L_g phase propagation because of a weak modal concentration (Zhang & Lay, 1995). A similar study recently conducted by Gualtieri et al. (2015) predicted exclusively fundamental mode propagation on the continent because the crust was considered as an infinite half-space, without taking the mantle into account.

2. When placing the source at any other position than s1 (Figs 3.E.3(c) to (h)), L_g phase is not excited, highlighting the importance of the ocean site effect (multiples) on the L_g phase excitation. For any other source position, almost all the energy is transmitted directly to the solid crust, where it dominantly propagates as a fundamental mode Rayleigh wave. This is in contradiction with the hypothesis proposed by Koper et al. (2010) about the L_g phase being mostly generated by shallow water sources along coastlines.

3. The presence of the L_g phase is capital for overtone excitation over a sedimentary basin (Figs 3.E.3(b) vs. (d), (f), (h), (j)). This is because the eigenfunctions associated to the overtones are maximum at larger depths compared to

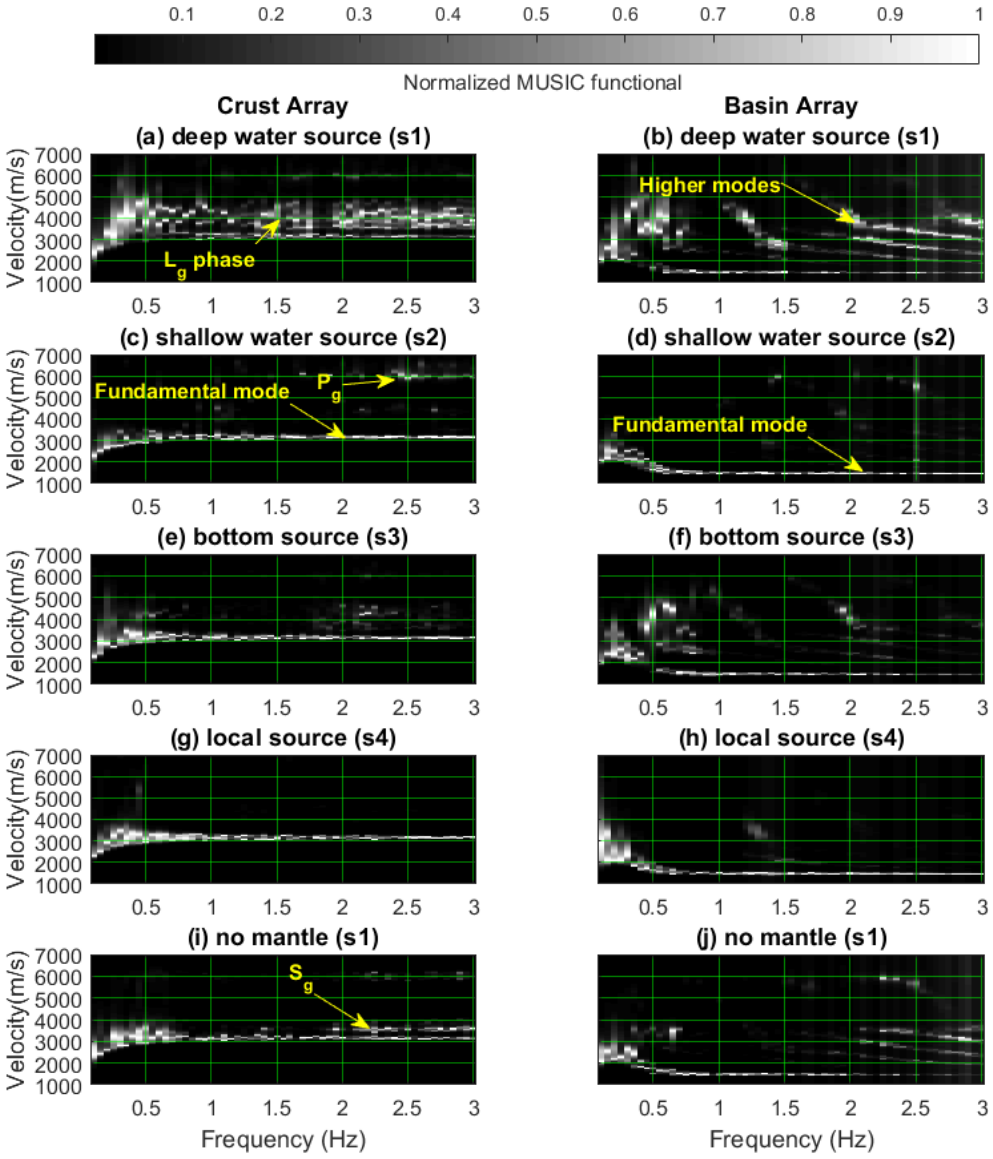


Figure 3.E.3: Results of the MUSIC algorithm. Left column: crust array; right column: basin array. (a) to (h): different source positions; (i) to (j): mantle properties replaced by crust properties (with the source s1).

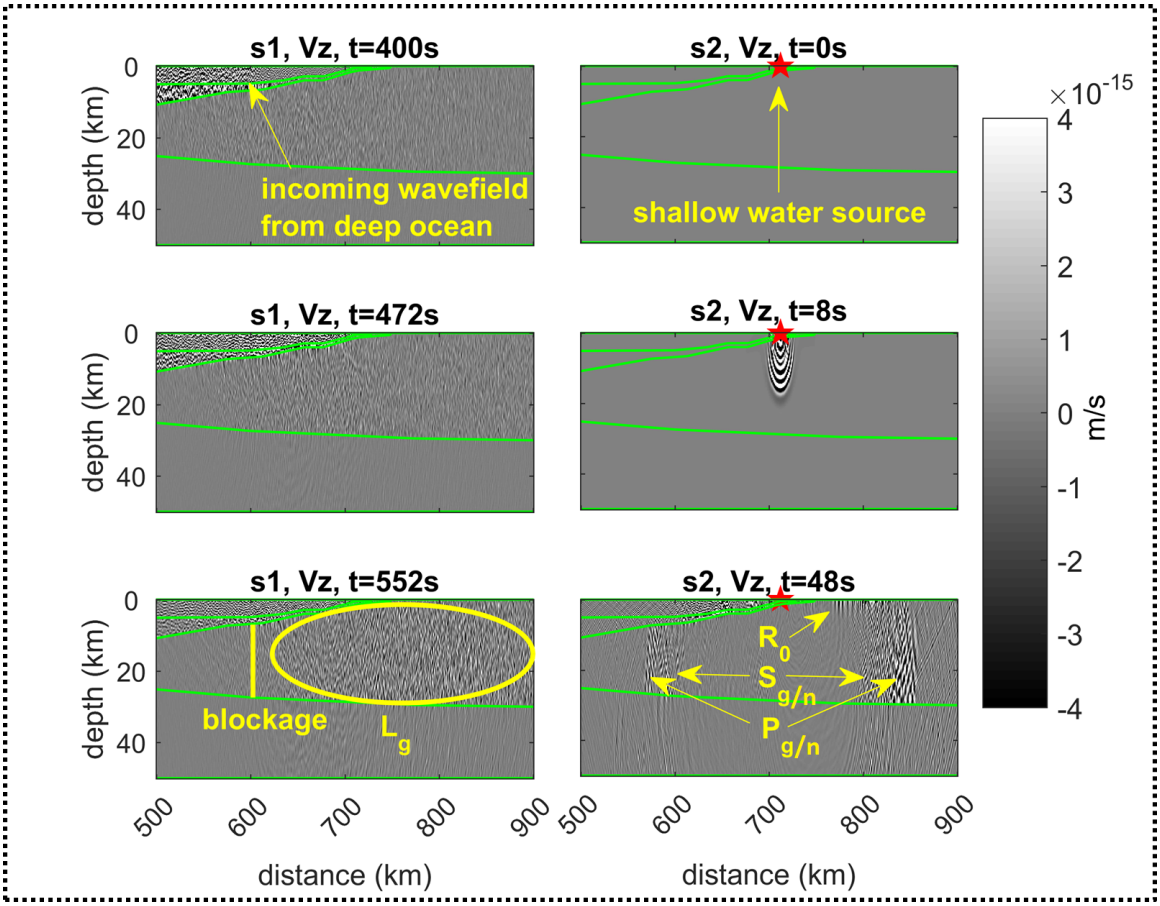


Figure 3.E.4: Vertical velocity wavefield snapshots for sources s1 and s2 for three different times. The snapshots correspond to the zoom zone in Fig. 3.E.1

.

the fundamental mode (e.g. Chen, 1993), and thus need deeper excitation to be activated. L_g phase brings this kind of excitation since it propagates with particle motion spread over the whole thickness of the crust.

4. What we identify as L_g phase is not an artefact since it disappears when we assign the elastic properties of the crust to the mantle region (Figs 3.E.3(i) to (j)). In this case, the crustal S wave (S_g) is visible at the crust array (Fig. 3.E.3i). This wave also excites some overtones in the sedimentary basin (Fig. 3.E.3j), but weaker than the L_g phase.

Conclusion

Based on the analysis of dispersion curves obtained by array processing of synthetic seismograms, we conclude that the presence of several overtones into the short-period seismic noise [0.1-3] Hz can be considered as evidence of a distant deep-water source, probably of the same nature as the microseism sources of class I or III described by Ardhuin et al. (2011), but acting in a wider frequency band. In our model, above a sedimentary basin, overtones could only be excited by the Lg phase propagating into the crust, itself generated when ocean multiples hit the continental margin. This gives a better understanding of the overtones and Lg phase observations in ambient noise. Though the attenuation was not taken into account in our study, it should mostly affect the fundamental mode Rayleigh wave because of its near-surface propagation, resulting into an even more prominent L_g phase.

Chapter 4

Elastic forward modelling of amplitude distortions

Mise en contexte (français)

Dans ce chapitre, les résultats du chapitre précédent (composition du champ d'onde) sont intégrés dans des simulations numériques par éléments spectraux. Puisque les simulations à grande échelle, comprenant les sources dans l'océan et le Moho (pour la génération de la phase L_g) seraient trop coûteuses pour un travail efficace, on génère des modes supérieurs en utilisant des sources profondes au sein d'un modèle relativement réduit (100x6 km). On utilise le modèle 1D en V_s obtenu dans le chapitre précédent. Afin de prendre en compte la structure géologique, on applique aux interfaces de ce modèle 1D la déformation donnée par la coupe selon le profil NS de Chémery (Fig. 2.9). Le réservoir est modélisé comme une perturbation élastique avec les valeurs données dans la Fig. 1.28. Ce chapitre a été publié sous forme d'article:
Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2019. Rayleigh wave amplitude distortions above a reservoir: new insights from elastic modelling, *Geophysical Journal International*, **accepted 04/02/2019**.

Context (English)

The results of the previous chapter in terms of wavefield composition are integrated in spectral-element simulations in this chapter. As large-scale simulations including the ocean and the Moho would be too time-intensive for an efficient work, I generate overtones by using deep sources in a relatively small model (100x6 km). I use the 1D V_s model obtained in the previous chapter. In order to take the geological structure into account, I apply the deformations along the NS profile in Chémery (Fig. 2.9) to the 1D model. The reservoir is as an elastic perturbation with the parameters from Fig. 1.28. This part was published as
Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2019. Rayleigh wave amplitude distortions above a reservoir: new insights from elastic modelling, *Geophysical Journal International*, **accepted 04/02/2019**.

Résumé (français)

Les ondes de surface dominent généralement le bruit sismique ambiant au-dessus de la fréquence des sources océaniques (~ 0.1 Hz). Leurs courbes de dispersion sont souvent inversées pour des profils de vitesse V_s, alors que leur amplitude reste largement inexploitée. Cependant, elle peut être porteuse d'informations utiles. On s'intéresse ici au problème de la prédiction des anomalies d'amplitude des ondes de Rayleigh lorsque celles-ci rencontrent une hétérogénéité du sous-sol, telle qu'un réservoir de gaz, à des fréquences allant de 0.5 à 5 Hz. Comme la méthodologie proposée est vouée à l'étude bu bruit ambiant, on utilise des fenêtres temporelles multiples contenant des signaux générés par des sources aléatoires. En utilisant la méthode des éléments spectraux pour la modélisation numérique 2D dans le plan radial-vertical, on montre que des anomalies d'amplitude spectrale (dépendantes de la fréquence) apparaissent à la surface à proximité de l'hétérogénéité. Les anomalies dues à une hétérogénéité profonde sont plus fortes lorsque le champ incident est constitué d'ondes de Rayleigh de modes supérieurs, du fait de leur plus grande profondeur de pénétration. En fonction du contenu modal du champ incident et de la complexité du modèle, les anomalies à la surface peuvent atteindre 20% de la densité de puissance spectrale des ondes incidentes. Pour comparaison, les anomalies générées par une structure géologique anticlinale au sein d'un bassin sédimentaire réaliste sont un ordre de grandeur audessus dans nos simulations. Les noyaux de sensibilité basés sur l'approximation de Born sont utiles pour l'interprétation de ces résultats. Ainsi les anomalies sont les plus sensibles à des perturbations en V_s au sein de l'hétérogénéité, suivies par ρ et V_p, comme attendu pour les ondes de surface. La sensibilité est concentrée dans les couches superficielles lentes lorsque celles-ci sont présentes dans le modèle.

Summary (English)

Surface waves usually dominate the ambient noise above the microseism frequency (~ 0.1 Hz). Their dispersion curves are routinely inverted for shear velocity profiles, while their amplitude is often neglected. Amplitude, however, can also carry useful information. We address the problem of predicting Rayleigh wave amplitude anomalies due to the interaction of incident Rayleigh waves with an embedded elastic inclusion (gas reservoir) between 0.5 and 5 Hz. The proposed method is designed for ambient noise applications, so we use multiple time-windows containing signals from randomly excited sources. Using the spectral-element method in the radial-vertical 2D plane, we show that frequency-dependent spectral anomalies arise at the surface in the vicinity of the inclusion. The anomalies generated by a deep inclusion are enhanced for the overtones, because of their deeper penetration. Depending on the wavefield modal content and the complexity of the background model, the anomalies can reach about 20% of the initial power spectral density. For comparison, the anomalies generated by a structural anticline in a realistic sedimentary basin are one order of magnitude higher in our simulations. The Born-based finite-frequency amplitude sensitivity kernels are useful for the interpretation of the results. The anomalies are most sensitive to V_s inside the inclusion, followed by ρ, as expected for surface waves. The sensitivity is concentrated in slow shallow layers when they are present in the model.

4.1 Introduction

Rayleigh waves are known to interact with shallow sub-surface sharp heterogeneities, such as cavities or rigid obstacles. Gucunski et al. (1996) numerically investigated the effects of rectangular inclusions on the surface wave dispersion curves estimated using Spectral Analysis of Surface Waves (SASW). Nasseri-Moghaddam et al. (2007) studied amplitude effects associated to an embedded void both numerically and experimentally (using ambient vibrations between 12 and 22 Hz). For a wide enough void, the wavefield above it was identified as trapped Lamb waves, usually observed within finite-thickness plates. Moreover, because of the alternating destructive and constructive interferences between the incident and the reflected waves, ripples before the void were observed in the amplitude-offset domain, similarly to the results of Gucunski et al. (1996) for dispersion curves. Tallavó et al. (2009) noticed an amplification localised above a zone containing buried rigid trestles. Amplitude effects due to less sharp heterogeneities, such as geological structures (e.g. Gorbatikov et al., 2008) or hydrocarbon reservoirs (e.g. Dangel et al., 2003) were also reported from low frequency ambient noise observations (typically around 1 Hz).

Rayleigh wave scattering by an embedded heterogeneity could be the mechanism common to all these observations. Several modelling approaches were proposed to address it. Gucunski et al. (1996) used a simple horizontal 1D model where the total wavefield was the sum of the incident wavefield and of two reflections originating from both edges of a rectangular inclusion. They were able to model their observations with limited accuracy, partly because the reflection coefficients were chosen without accounting for the depth-dependent ground motion associated to Rayleigh waves. Riyanti & Herman (2005) proposed a method for calculating the scattered wavefield in 3D based on the integral representation. Knowing the reference medium Green's function, this method yields an implicit equation governing the scattered wavefield, which must be solved iteratively. Chai et al. (2012) used this method to model the amplitude distribution at the surface and found that the vertical component was attenuated above the scatterer. Yu & Dravinski (2009) proposed a method for calculating the wavefield scattered by a spherical cavity, based on the boundary element method (BEM). They predicted an amplification of the horizontal component, while the vertical component was attenuated (fig. 7 in their paper). Gorbatikov & Tsukanov (2011) used a finite-difference scheme to model the amplitude anomalies of the fundamental mode Rayleigh waves due to scattering by an inclusion embedded in a homogeneous medium. In order to retrieve the inclusion position from the observed anomalies, they used a simple relationship to tie each frequency to the corresponding depth:

$$H(f) = K\lambda(f) \ , \tag{4.1}$$

where $H(f)$ is the depth associated to the frequency f, $\lambda(f)$ the corresponding wavelength of the Rayleigh fundamental mode, and K an ad-hoc numerical coefficient which must be optimised, with typical values of 0.4-0.5. Based on their modelling and available field observations, they conjectured that a positive amplitude anomaly at a given frequency is always due to a negative perturbation of the velocities at the corresponding depth.

In the present work, we use numerical simulations (Spectral Element Method (SEM), Komatitsch et al., 1999) and Born approximation to show that depending on the frequency band, both amplification and attenuation can be driven by purely elastic scattering. Our approach could be useful for ambient noise amplitude modelling above smooth structures, such as anticlines, or above fluid-saturated reservoirs (hydrocarbons, carbon dioxide, geothermal vapour). The Born approximation yields a linear link between the heterogeneity properties and the amplitude anomaly, useful for interpreting the numerical results. We show that this link reveals itself somewhat more complicated than described by Gorbatikov & Tsukanov (2011), and namely that the sensitivity of the amplitude to velocity perturbations can be both positive and negative. Moreover, using SEM simulations allows us to address heterogeneous background models.

After introducing the observables to measure amplitude anomalies in Section 4.2, we present the numerical modelling workflow in Section 4.3. The spectral amplitude anomalies obtained by numerical simulation in various configurations are presented in Section 4.4. The influence of the background medium is studied by computing amplitude anomalies in four models of increasing complexity, ranging from the homogeneous half-space to a realistic layered model with structural topography. The influence of the incident wavefield composition and of the heterogeneity parameters (depth and size) are also investigated. Finally, the anomalies generated by a reservoir-like inclusion are compared to those generated by a realistic structural feature such as an anticline. The Born approximation is introduced in Section 4.5. Numerical Green's functions are obtained via SEM in the background model. By using Born approximation, we derive amplitude sensitivity kernels (Appendix 4.A). For a given heterogeneity, they predict the scattered wavefield and the associated amplitude anomalies. The prediction is validated by comparison with numerical simulations in the full model (Appendix 4.B), and the limits of the Born approximation are discussed. The main objective is to better understand the modifications of the spectral content in the context of elastic wave propagation.

4.2 Definition of the observables

In this part, we define two observables which are sensitive to the presence of lateral heterogeneities in a 2D medium. Working in 2D is motivated by two main factors. First, it allows to reduce the computational time of the costly spectral element simulations, especially when we model several realisations of random source distributions (see Section 4.3.3). Second, it allows to store the complete Fréchet derivatives of the wavefield in the computer memory, which is useful in Section 4.5. Our study is performed in the P-SV system (radial-vertical plane), containing Rayleigh but not Love waves. While Rayleigh waves transport energy on the radial and the vertical components, Love waves transport energy on the transverse component (SH system). Both P-SV and SH systems are coupled through scattering, i.e. Rayleigh waves can be scattered as Love waves and vice-versa (e.g. Maupin, 2017). Such coupling is not taken into account in our 2D P-SV modelling.

In this work, we only consider the spectral amplitude of the vertical component signal. The latter can be measured by the power spectral density (PSD), defined as

$$\mathrm{PSD}_z(\omega) = |u_z(\omega)|^2 \ , \tag{4.2}$$

with $u_i(\omega)$ the Fourier transform of the displacement signal $u_i(t)$, and $i \in x, z$. The main objective of our work is to study how the PSD is affected by the medium heterogeneities. The general idea is that the PSD should be perturbed in the vicinity of the sub-surface heterogeneities. To measure the spatial variations of the PSD, we define a reference station in $\mathbf{x}_0$ chosen far enough from the expected medium heterogeneities, and we introduce the normalised observable

$$\eta(\mathbf{x}_r, \mathbf{x}_0, \omega) = \log\left[\frac{\mathrm{PSD}(\mathbf{x}_r, \omega)}{\mathrm{PSD}(\mathbf{x}_0, \omega)}\right] \tag{4.3}$$

If the anomalies are small enough, η simplifies as

$$\eta(\mathbf{x}_r, \mathbf{x}_0, \omega) \approx \frac{\mathrm{PSD}(\mathbf{x}_r, \omega) - \mathrm{PSD}(\mathbf{x}_0, \omega)}{\mathrm{PSD}(\mathbf{x}_0, \omega)} \ , \tag{4.4}$$

so that it represents a relative anomaly which can be measured in percent. If we consider that a heterogeneity (which we also refer to as "scatterer") is embedded in a laterally invariant medium, then the power spectral density is expected constant in the background medium, as far as the geometrical spreading from distant sources can be neglected at the array scale. Then the variations of η are solely due to the effect of the heterogeneity. If, however, the reference medium is not laterally invariant, η is not only sensitive to the scatterer, but also to the lateral variations of the background model. For the latter case, we define a differential observable

$$\eta'(\mathbf{x}_r, \mathbf{x}_0, \omega) = \eta(\mathbf{x}_r, \mathbf{x}_0, \omega) - \eta_0(\mathbf{x}_r, \mathbf{x}_0, \omega) \ , \tag{4.5}$$

where η_0 is measured in the reference medium, possibly including structural heterogeneities. From a practical point of view, η is of interest in the exploration context. It is well suited to instantaneous measures of the noise amplitude on sites which have a reasonably homogeneous lateral structure, except in some specific locations containing a fluid reservoir, a fault zone, a void, or any other geological heterogeneity of interest. In this case, η is just a particular value of η' with $\eta_0 \approx 0$. η' with non-zero η_0 corresponds to a situation some localised medium evolution on the PSD is superimposed on the effect of a static heterogeneity. This situation can be met in the monitoring context, relevant when time-lapse variations of the ambient noise amplitude above a changing medium are available. This can concern subtle changes of the medium, for example due to pumping or injecting fluids into a thin reservoir layer. Since η' has a more general definition, we focus on this observable in the remainder of the paper. Note that the formulation (4.5) for time-lapse variations still involves a reference station. This is useful to correct for the time-lapse evolution of the source amplitude, which can be particularly strong for the ambient noise sources (Gorbatikov et al., 2008). In the next section, we present a numerical approach to model η'.

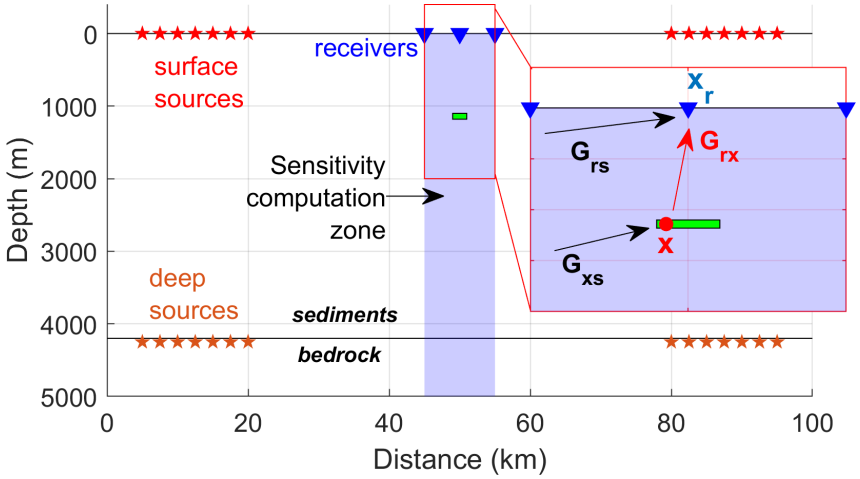


Figure 4.1: Example of the Layer-Over-Half-space (LOH) model used in the study. Distant sources at the surface (red stars) and in depth (brown stars) are used to create incident wavefields with different modal compositions. The blue shaded zone is where sensitivity kernels are investigated. It can contain a reservoir-type inclusion (green rectangle). Zoomed area: synthetic view of the terms implied in the Born approximation. The signal at the receiver $\mathbf{x_r}$ is the sum of an incident (black) and a scattered (red) wavefield. $\mathbf{x}$ is one of the scattering points over which an integration is performed to recover the total scattered wavefield at $\mathbf{x_r}$.

4.3 Numerical forward-modelling

The spectral element method (SEM, Komatitsch et al., 1999) is used to numerically model the amplitude spectral anomalies above an inclusion embedded into different background models. The studied frequency interval is [0.5-5] Hz. The typical mesh element (quadrangle) size is of 100 m. The mesh is locally refined in the vicinity of the scattering inclusion (blue zone containing the green rectangle in Fig. 4.1), and in the shallow layers, where the typical element size can be as low as 20 m. This insures that the number of elements per S-wavelength at 5 Hz is larger than 4 everywhere in the model. The refined zones typically impose time steps of several tenths of a millisecond. Modelling η' requires two simulations, one performed in the complete medium containing a heterogeneity, an another one in the reference background medium. Since our purpose is to model the ambient noise, we introduce random source distributions and measure the average PSD. In this section, we provide details on the implementation of the different steps.

4.3.1 Background models

Four different background models are tested. The first three are the homogeneous half-space model (HM, Table 4.1a), the layer-over-half-space model (LOH, Table 4.1b), and the flat realistic layered model (Table 4.1c). The fourth model is obtained by superimposing a realistic anticline structure onto the flat realistic model. The corresponding mesh is shown in Fig. 4.2. The size of all the models is 100x5 km (Fig. 4.1). Such a large model is used to attenuate body waves by geometrical spreading, and to concentrate the study on surface waves. The LOH model is the simplest possible

Table 4.1: Medium parameters in the simulations used for model validation. L∞ stands for the layer with an absorbing boundary condition at the bottom, which represents an infinite half-space.

(a) Homogeneous half-space

Parameter	L∞
h (m)	-
ρ (kg.m^{-3})	2300
V_s (m.s^{-1})	2000
V_p (m.s^{-1})	4000

(b) Layer-over-half-space (LOH)

Parameter	L1	L∞
h (m)	4200	-
ρ (kg.m^{-3})	2300	3850
V_s (m.s^{-1})	2000	3500
V_p (m.s^{-1})	4000	7000

(c) Realistic layered model

Parameter	L1	L2	L3	L4	L5	L6	L7	L∞
h (m)	350	313	280	135	232	960	1930	-
ρ (kg.m^{-3})	1700	2100	2300	2400	2200	2300	2400	3850
V_s (m.s^{-1})	810	1300	2040	2400	1740	2030	2330	3500
V_p (m.s^{-1})	1620	2600	4080	4800	3480	4060	4660	7000

respresentation of a sedimentary basin, with a stiff half-space overlaid by a looser layer. The last two models correspond to the environment of the Chémery underground gas storage (UGS) operated by Storengy. The 1D structure of the layered model (Table 4.1c) is close to the one obtained by Peruzzetto et al. (2018) by inverting the dispersion curves extracted from ambient noise above the Chémery UGS. The topography of the layers used in the last model is extracted from the NNW-SSE cross-Section of the 3D geological model provided by Storengy. Absorbing boundary conditions are applied on the left, right and bottom edges of the models.

4.3.2 Inclusion parameters

In the first three models (HM, LOH and flat layered), the inclusion is modelled as a rectangle with a width of 2 km, a thickness of 80 m, located at a depth of 1100 m. It is located in the middle of the model, between 49 and 51 km (green rectangle in the zoomed zone on the right in Fig. 4.1). For the last model (containing the anticline deformation), the inclusion mimics a gas accumulation trapped beneath a cap rock. The top of the gas bubble is at a depth of 1100 m, while the bottom is at a constant depth of 1180 m, as for the rectangular inclusion. $\delta V_p/V_p = -18\%$ and $\delta\rho'/\rho' = -10\%$ are used as the elastic perturbations representative of a full substitution of water by gas in a sandstone reservoir. These values were obtained using the Gassmann law

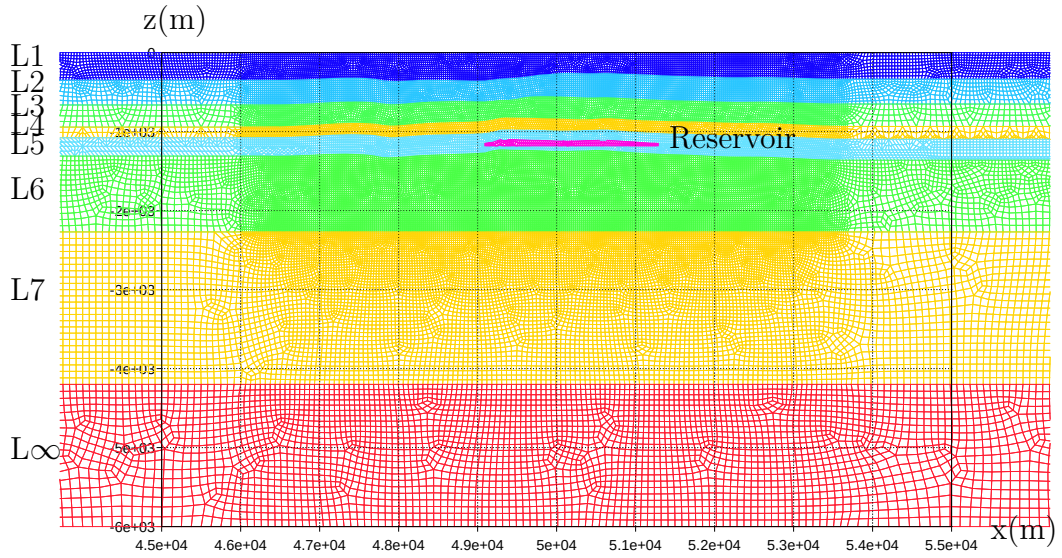


Figure 4.2: Mesh of the realistic model with anticline structure, zoomed in the central zone. The layer labels refer to those in Table 4.1c. Generated with Gmsh software (Geuzaine & Remacle, 2009).

for fluid substitution, and account for the effects of the effective pressure on the rock matrix bulk and shear moduli (Vidal, 2002). With such a modelling scheme applied to the Underground Gas Storage (UGS) reservoir operated by Storengy in Céré-la-Ronde (France), Vidal (2002) showed that $\delta V_s = \delta\sqrt{\mu/\rho} \approx 0$, because the drop in density is compensated by the drop in the shear modulus induced by the pore pressure increase. More generally, the Gassmann law relies on the hypothesis that the shear wave modulus itself is insensitive to the fluid saturation.

4.3.3 Source distributions and average PSD

The ambient noise can be approximated as a superposition of wavefields due by numerous randomly excited sources located within specific source regions. The PSD of the ambient noise is generally estimated using the Welch's method, which consists of splitting the record into several overlapping time windows and averaging the PSD over all the windows:

$$\langle \mathrm{PSD}(\omega) \rangle = \frac{1}{N_{\mathrm{win}}} \sum_{l=1}^{N_{\mathrm{win}}} \mathrm{PSD}^{(l)}(\omega) \ , \tag{4.6}$$

where N_{win} is the total number of windows (see Saenger et al., 2009, appendix A). Each time window can be simulated with one realisation of a random source distribution. Measuring the averaged η' anomalies then requires 2 simulations per random source distribution realisation (one simulation with and one without the inclusion).

In this work, the sources are modelled as vertical point forces with a Ricker wavelet as time function. The precise shape of the time function does not influence η' since the latter measures relative PSD *anomalies*, and not the shape of the noise spectrum. The

Table 4.2: Bounds used for uniform distributions of the source parameters. (L.) and (R.) refer to the independent distributions modelled respectively on the left and the right sides of the background model.

Parameter	Bounds (L.)	Bounds (R.)
x_s	$[5, 20]$ km	$[80, 95]$ km
f_0	$[0.5, 3]$ Hz	
t_0	$[5, 40]$ s	

time functions we use are not supposed to reproduce any natural process. The central frequencies of Ricker wavelets, however, determine the frequency range of the modelled noise. For generating realisations of the noise wavefield, we use distant sources characterised by random values of horizontal position, central frequency and emission time. These values are obtained from truncated uniform probability distributions with bounds listed in Table 4.2. Two independent distributions are used at both ends of the physical model (stars in Fig. 4.1) to form a 2D equivalent of a wavefield with a homogeneous back-azimuth distribution. One wavefield realisation uses 400 random sources generated by these distributions. In total, we used 20 independent wavefield realisations. Each realisation corresponds to a seismic noise recording lasting about a minute.

Both natural and human sources can be the origin of the ambient wavefield at frequencies from 0.5 up to several Hz (see mechanisms in Farrell & Munk (2008); Peureux & Ardhuin (2016) and observations in Koper et al. (2010); Gimbert & Tsai (2015); Peruzzetto et al. (2018)). While human sources are expected to excite mainly the fundamental mode, natural sources can also excite overtones on the continent via a diffraction phenomenon on the continental margin (Koper et al., 2010; Kazantsev et al., 2018b). As we analyse the influence of the wavefield modal content on the PSD anomalies, each source distribution in the numerical simulations was designed to excite one particular set of modes. The sources within one distribution are all kept either on the surface, exciting mostly the fundamental mode, or at a constant depth of $z = 4250$ m, exciting a mixture of overtones. As the sources are in all cases located far from the simulated receiver array, we assume the body wave contribution is weak in the incident wavefield. This hypothesis is further tested in Section 4.4.2. To summarise, our modelling assumes ambient noise of known modal composition (via fixing the source depth) and back-azimuth distribution (in 2D, via balancing the number of sources on the left and right sides of the model).

4.4 Results: modelled spectral anomalies

In this section, results of η' modelling are presented for the four background models listed in Section 4.3.1. Surface sources are used for the homogeneous model, and deep sources for the flat and deformed realistic models. For the LOH model, we test both surface and deep sources, which yields a total of 5 configurations. The reference receiver position is chosen at $x_0 = 47$ km. The inclusion parameters are those given in Section 4.3.2. In Section 4.4.1, we present the resulting η' anomalies, and analyse

how they are influenced by the background model. Then in Section 4.4.2 we focus on the effect of the wavefield composition on the η' anomalies by switching from deep to surface sources in the LOH model. The sensitivity of η' anomalies with respect to the inclusion geometry is addressed is Section 4.4.3. Finally, the modelled η' due to structural effects within the background model is shown in Section 4.4.4, and compared to the previously computed η' due to the reservoir-like inclusion.

Spectral anomalies (η') obtained by SEM for different models and source distributions are shown in Fig. 4.3. Fig. 4.3a displays the anomalies extracted above the middle of the inclusion. In the homogeneous and LOH models with sources at the surface, the anomaly is almost zero above 0.8 Hz, while between 0.5 and 0.8 Hz it is slightly positive, reaching about 1% at maximum (green and pink curves). On the contrary, when the sources are located in depth, the anomaly reaches -6% for the LOH model (blue curve). This effect is due to a different modal content of the incident wavefield, depending on the source depth, which is discussed in Section 4.4.2. The anomaly further increases in a flat model with realistic layering (Table 4.1c), where peak values are about +10% and -10% (black curve). Finally, for the model containing an anticline structure, the maximum anomaly reaches 17% (red curve). Figs 4.3(b) to (f) display the spatial distribution of the anomalies. When surface sources are used (i.e. the fundamental mode dominates, see Section 4.4.2), the anomaly is rather concentrated vertically above the inclusion (between the two green lines in Figs 4.3(b) and (c)). On the contrary, the anomaly is more diffuse in space for the other configurations (Figs 4.3(d) to (f)). The symmetry of the anomaly pattern decreases for the model with the anticline (Fig. 4.3f) because the structure itself is asymmetric.

4.4.1 Background model influence

After this general presentation of the anomaly patterns in each configuration, we investigate the stability of the observed patterns with respect to changes in bedrock depth (LOH model, Section 4.4.1) and in the layering (realistic model with anticline structure, Section 4.4.1). The study in this section is peformed for deep sources, which appear to be the only able to generate measurable anomalies due to a reservoir at kilometric depth.

Bedrock depth influence in the LOH model

In order to determine whether a precise knowledge of the bedrock depth in the LOH model is necessary for a correct anomaly prediction, we test a range of values between 1250 and 4000 m, and model the anomaly for each value. Results are shown in Fig. 4.4. Generally, the amplitude of the anomaly appears to decrease with the increase of the bedrock depth value. Below $\approx$ 2.2 Hz, the anomaly patterns are similar for all the values larger than 2000 m. At higher frequencies, some discrepancies persist for depths between 2000 and 3000 m. For depths larger than 3000 m, the patterns look similar at all the considered frequencies. Thus, as a practical guidance, we can retain that a good knowledge of the bedrock depth (to within $\sim$ 100 m) is required if the bedrock is closer to the inclusion than $\sim$ 1000 m (we remind that the anomaly patterns were calculated for an inclusion embedded at z = 1100 m). For deeper bedrocks, a

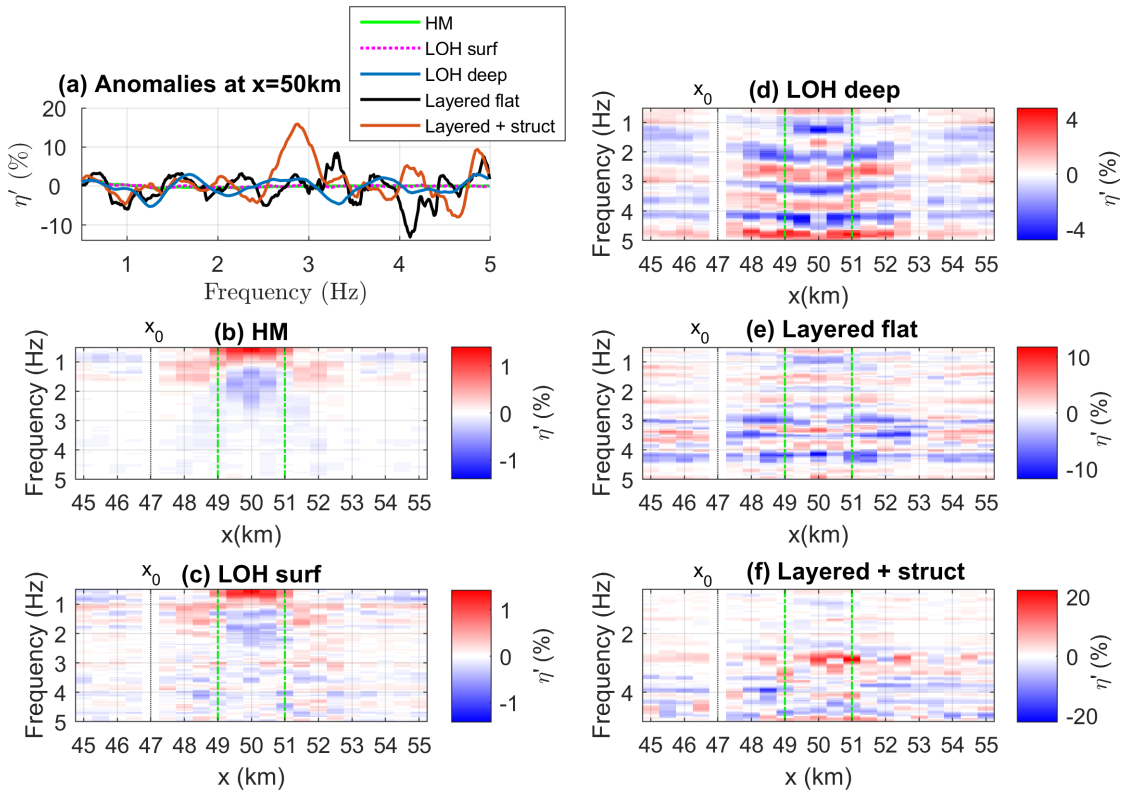


Figure 4.3: $\langle \mathrm{PSD}_z \rangle$ spectral anomalies (observable η') for the receiver at $x_r = 50$ km (above the middle of the inclusion) and the reference station at $x_0 = 47$ km, computed by SEM simulations in different background models with multiple random sources at different depth. Sources at the surface: green (homogeneous medium) and pink (LOH). Sources at $z = 4250$ m: blue (LOH), black (realistic layered medium) and red (realistic layered medium with anticline). (a): Spectral anomalies extracted above the middle of the reservoir. (b) to (f): Spectral anomalies plotted in the frequency-distance plane for each of the five configurations. The horizontal limits of the inclusion are shown with green vertical lines. The reference station location is shown with the black vertical line.

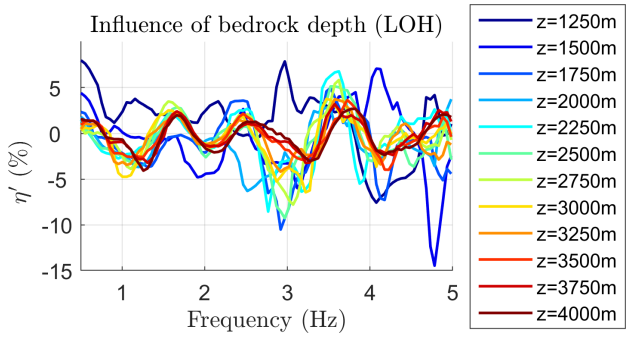


Figure 4.4: $\langle \mathrm{PSD}_z \rangle$ spectral anomalies (observable η') for the receiver at $x_r = 50$ km (above the middle of the inclusion) and the reference station at $x_0 = 47$ km, computed by SEM simulations in LOH-type models for a range of bedrock depths, and distant multiple random sources at $z = 4250$ m.

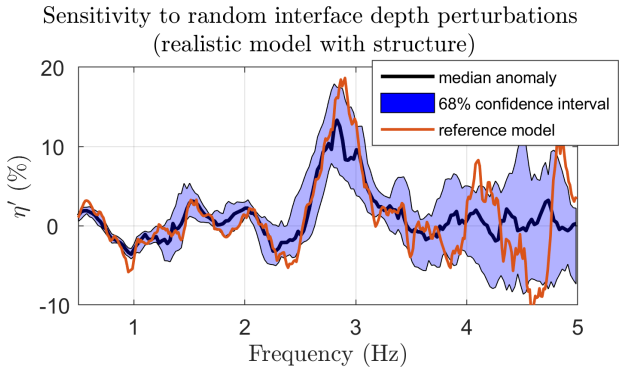


Figure 4.5: Same as Fig. 4.4, but for a set of 20 realisations of realistic layered models with structure (randomised interface depth), and distant multiple random sources at $z = 4250$ m. Reference layering (Table 4.1c) - red line, median anomaly over realisations - black line, 68% confidence interval - shaded area.

significant uncertainty on the bedrock depth does not generate a major distortion of the anomaly pattern.

Layering influence in the realistic model

For the realistic layered model, another key parameter is the thickness of different layers. We investigate how the uncertainty on this thickness affects the anomaly pattern. To model the uncertainty, we consider that the depth z_i of each of the first 6 interfaces i follows a truncated normal distribution with following standard deviation and bounds:

$$\begin{cases} \sigma &= 0.2 \cdot (z_{i+1}^{(0)} - z_{i-1}^{(0)}) \\ z_i^{\max} &= z_i^{(0)} + 0.3 \cdot (z_{i+1} - z_i) \\ z_i^{\min} &= z_i^{(0)} - 0.3 \cdot (z_i - z_{i-1}) \ , \end{cases} \tag{4.7}$$

where the exponent $^{(0)}$ refers to the values in the reference non-perturbed model given in Table 4.1c. We generate 20 sets of independent realisations of z_i following this normal distribution. The interfaces are vertically translated according to these values, in both flat and deformed zones. The inclusion depth and size remain unchanged (see reservoir in Fig. 4.2). We then perform SEM simulations with random distant deep sources to compute the PSD anomalies in each of the 20 perturbed models. At each frequency, we compute the median anomaly and the 68% confidence interval. Results are shown in Fig. 4.5. Below ≈ 3.5 Hz, the anomaly pattern remains stable. The 68% confidence interval contains the anomaly computed in the reference non-perturbed model, indicating that the anomaly presented in Fig. 4.3 is robust with respect to reasonable uncertainties on the layering. However, above 3.5 Hz, the median anomaly is close to 0, and the 68% confidence interval is equally spread over positive and negative values, which means that the anomalies are highly sensitive to layering perturbations in this frequency range, preventing a reliable amplitude prediction. Therefore, an approximate knowledge of the background model can be sufficient for anomaly prediction only at lower frequencies. We interpret this threshold frequency in terms of phase shift between the incident and the scattered wavefields in Section 4.5.2.

Another source of uncertainty for amplitude prediction is the incident wavefield. Since the phase information is ignored, the source-time functions of the ambient noise sources are not required. Instead, the important parameter is the wavefield modal composition, which is addressed in the next section.

4.4.2 Incident wavefield composition influence

As previously mentioned in Section 4.3.3, natural distant ambient-noise sources are mainly located in the oceans (e.g. Ardhuin et al., 2011). However, including the ocean within a large-scale model would be too costly in terms of repeated SEM simulations. This was done in Kazantsev et al. (2018b), where it was shown how oceanic sources were able to generate higher mode Rayleigh waves above sedimentary basins on the continent. On the other hand, local human sources located at the surface mainly excite the fundamental mode. In this work, we used distinct ad-hoc source distributions aiming to excite either the fundamental mode (surface sources) or the overtones (deep sources), as described by Bonnefoy-Claudet et al. (2006a). The only physical meaning of taking deep sources was to generate higher modes in a simple way. In this section, we check which overtones are generated in the numerical simulations, and how their presence affects the amplitude sensitivity.

In order to measure the modal content, we apply the FK transform to the synthetic incident wavefield simulated by SEM in different models without the reservoir. The wavefield for this analysis is generated by sources located only on the left side of the model, and recorded by a large array of 1201 receivers spanning from 30 to 90 km, with a 50 m spacing. Such a large array is needed for the FK transform to have enough resolution at low frequency. We display the results for both vertical (Figs 4.6(a) to (c)) and horizontal (Figs 4.6(d) to (f)) components. The theoretical dispersion curves for the phase velocity of the different modes, computed using *Computer Programs for Seismology* software (CPS, Herrmann, 2013), are overlaid on the result. The maximum P- and S-wave velocities of the model are indicated with yellow lines. No energy is present at velocities beyond the maximum P wave velocity, meaning that there are no steeply-incident body waves. Moreover, body waves in a non-attenuating medium should be non-dispersive and appear as straight lines crossing the origin. We observe only dispersive events in the dispersion plot. All the events follow (or are connected to) the theoretical dispersion curves for Rayleigh wave modes. There are some dispersive events beyond the maximum S-wave velocity in Figs 4.6(b), (c), (e) and (f), while Rayleigh wave modes cannot propagate at such velocities. All of these events connect to a Rayleigh wave mode as they cross the maximum S-wave velocity line. This is consistent with the *leaky modes*, «which behave as Rayleigh-like waves spreading faster than the S-wave velocity», according to García-Jerez & Sánchez-Sesma (2015). Now, comparing Figs 4.6(a) and (b) (LOH model, vertical component), or (d) and (e) (horizontal component), indicates that the incident wavefield excited by surface sources is dominated by the fundamental mode at all frequencies, while for deep sources a superposition of many overtones is observed. Overtones are also dominating the more sophisticated layered medium (Figs 4.6(c) and (f)). We checked that the results with a small array of 10 km above the inclusion are of similar nature, but it is hard to separate the modal dispersion

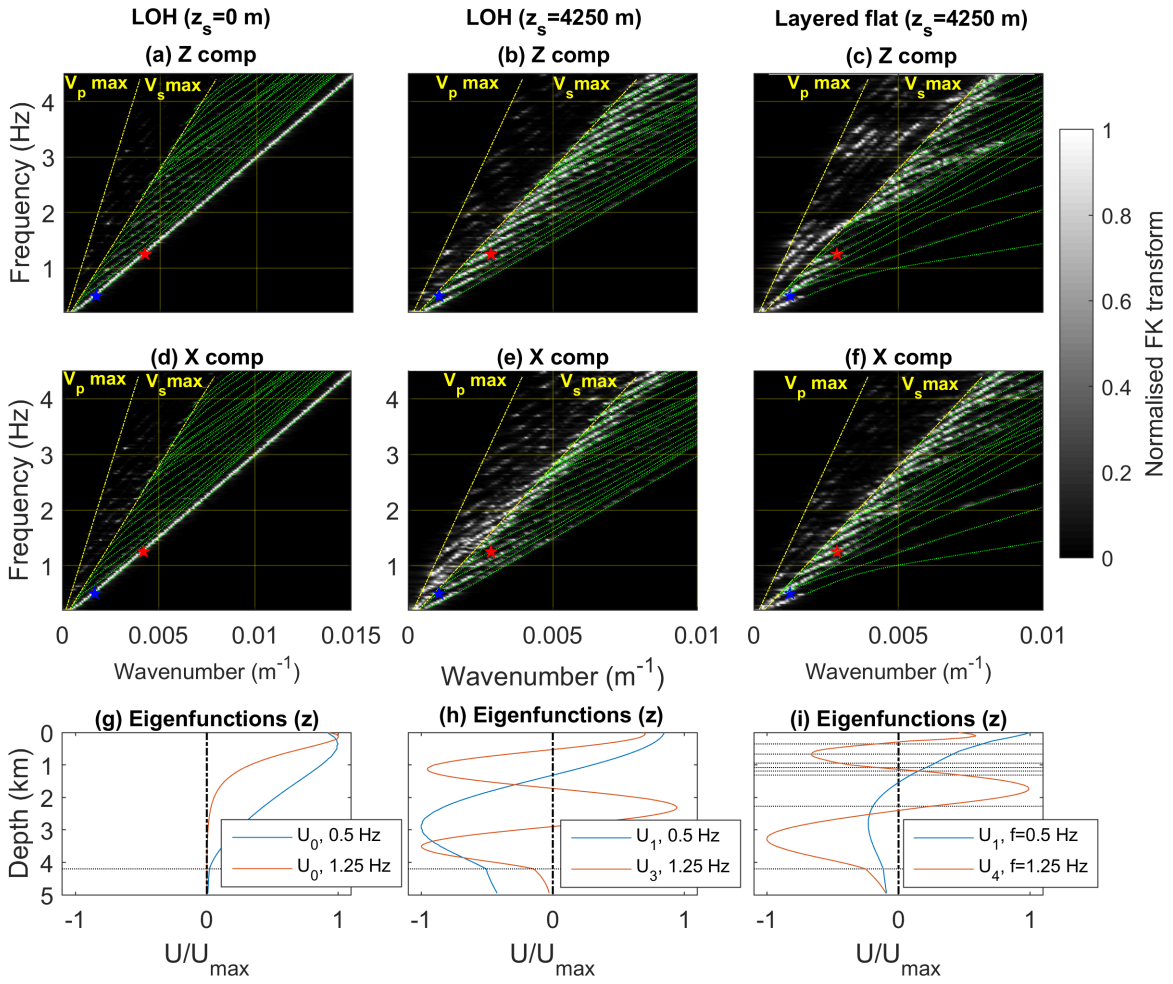


Figure 4.6: Upper two rows: FK transforms of the synthetic incident wavefields at the surface. (a)-(c): Z component. (d)-(f): X component. (a) and (d): LOH model, sources at the surface; (b) and (e): LOH model, sources at $z = 4250$ m; (c) and (f): Realistic flat layered model, sources at $z = 4250$ m. Green curves: theoretical dispersion curves for different modes. The blue (resp. red) star indicates the mode dominating the incident wavefield at $f = 0.5$ Hz (resp. $f = 1.25$ Hz), i.e. at the frequencies for which the sensitivity is shown in Figs 4.10 and 4.11. Bottom row: vertical component Rayleigh wave eigenfunctions computed for the modes spotted by the stars on the dispersion plots above, at $f = 0.5$ Hz (blue lines) and $f = 1.25$ Hz (red lines). (g), (h) and (i) correspond to the dispersion plots in (a), (b) and (c), respectively. The black dotted lines show the interfaces between the layers.

curves.

The difference between shallow and deep sources can be explained in terms of modal eigenfunctions. In a laterally homogeneous model, the depth dependent amplitude of the mode n respects the relationship

$$A_n(z,\omega) \propto U_n(z_s,\omega)U_n(z,\omega) \ , \tag{4.8}$$

where U_n is the vertical component eigenfunction for the mode n, while z_s and z are respectively the source and receiver depths (Aki & Richards, 2002). Knowing the 1D layered model, the eigenfunctions can be computed using CPS. The higher modes usually have eigenfunctions with stronger values in depth, while the fundamental mode eigenfunction is nearly zero beyond one wavelength. Thus, deeper source distributions are expected to excite strong higher modes via the term $U_n(z_s,\omega)$, which is in line with the wavefield modal content observed in the numerical simulations. Now, we compute the eigenfunctions for the relevant modes at the receiver (term $U_n(z_r = 0,\omega)$). The blue (resp. red) stars in Fig. 4.6 show which particular mode is dominant at $f = 0.5$ Hz (resp. $f = 1.25$ Hz). The fundamental mode eigenfunctions are relevant for the LOH medium with surface sources (Fig. 4.6g). When the sources move in depth, the mode 1 becomes dominant at $f = 0.5$ Hz, and the mode 3 at $f = 1.25$ Hz. The eigenfunction of the mode 3 (Fig. 4.6h) is much stronger in depth compared to the fundamental mode. The same observation can be made for the layered model, where the mode 4 dominates at $f = 1.25$ Hz (eigenfunction in Fig. 4.6i). Because of the deeper penetration of the higher-mode eigenfunctions, we expect them to present a stronger sensitivity to deep inclusions, compared to the fundamental mode.

4.4.3 Inclusion geometry influence

For each considered combination of background model and source depth, we now analyse the anomaly sensitivity to the reservoir depth, width and thickness. The results presented in this section are obtained by using a linearised version of the forward-modelling (Born approximation). We show in Appendix 4.B that this approximation is in good agreement with the full SEM simulations for predicting the η' anomalies due to a *small* contrast such as a reservoir. Thus we use it here to avoid performing too many SEM simulations. A rectangular reservoir is considered for all the models. The reservoir depth is defined at the upper bound. We use the RMS of η' over the frequency range [0.5-5] Hz as a measure of the average anomaly strength. The RMS anomalies for a range of reservoir depth, width and thickness are shown in Figs 4.7(a) to (c). The frequencies corresponding to the maximum anomaly are displayed for the same range of parameters in Figs 4.7(d) to (f) as a scatter plot. These dominant frequencies are shown only if the maximum anomaly is significant (above 1%) for the considered model. Several conclusions can be made:

1. The strength of the anomaly always decreases for increasing reservoir depth. When the fundamental mode dominates the wavefield, the anomaly drops to nearly zero below 1 km (Fig. 4.7a, HM and LOH surf.), because of a low sensitivity in depth. When higher modes are dominating (other curves in Fig. 4.7a),

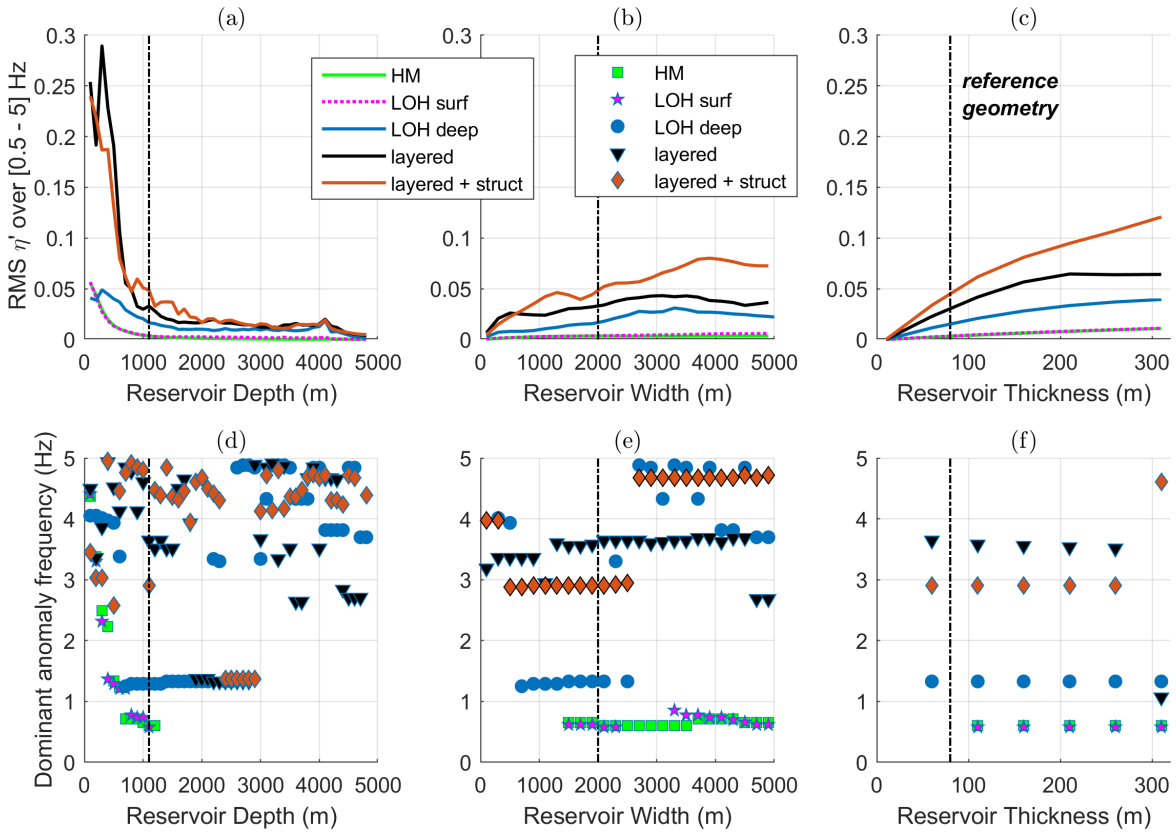


Figure 4.7: Upper row: RMS η' anomaly over the frequency range [0.5-5] Hz for different background models, as function of the reservoir depth (a), width (b) and thickness (c). The black dash-dotted line indicates the parameters of the reference reservoir used in the previous sections. Lower row: (d) to (f) - frequency of maximum $= \eta'$ as function of the same parameters as in (a) to (c), displayed only if $\max(\eta') > 1\%$. In (d), note the chaotic behaviour of the dominant frequency with respect to depth for models dominated by higher modes. In (e)-(f), note the relatively stable behaviour with respect to width and thickness.

a small anomaly (RMS $\approx$ 2%) persists over depth. This is due to a deeper sensitivity of the higher modes (see Section 4.4.2).

2. The anomaly drops when the reservoir is located in the stiff regions of the model (e.g. transition to deeper sediments at $z \approx 500$ m, and then to the bedrock at $z = 4200$ m for the layered models, see black and red curves Fig. 4.7a).

3. The strength of the anomaly generally increases with the reservoir size (width and thickness, see Figs 4.7b and c), but reaches a plateau beyond some threshold value.

4. The dominant frequency of the anomaly is mainly sensitive to the reservoir depth (Fig. 4.7d). However, a clear trend is observable only for the fundamental mode dominated homogeneous models (HM and LOH surf.), where the dominant frequency decreases with the reservoir depth. This also corresponds to the prediction of the simplified relation (4.1). For the other models, the dominant frequency exhibits strong jumps. This illustrates the fact that the widespread paradigm of higher frequencies being associated with shallower depths breaks down when higher modes dominate the wavefield at higher frequencies. We address this point in more detail in Appendix 4.C, where we show that this kind of irregular patterns are indeed predicted by the modal summation theory.

5. The dominant frequency of the anomaly is rather insensitive to the reservoir size (Figs 4.7e and f). Only a few jumps occur, due to the competition between several peaks present in the anomaly pattern at different frequencies (e.g. the positive peaks near 2.9 Hz and 4.8 Hz for the realistic model with structure in Fig. 4.3 (red curve), between which the red diamonds jump in Fig. 4.7e for a reservoir width value of 2800 m).

In the next section, we compare the effect of a reservoir to that of a geological structure such as an anticline.

4.4.4 Structural effects vs. reservoir effects

Local deviations from a laterally invariant velocity model generate an anomaly η' (eqn. (4.5) with $\eta_0 = 0$). The deviation from the layered flat 1D model (given in Table 4.1c) towards the model including an anticline is shown in Fig. 4.8a in terms of elastic perturbations. The expected effect of such a structural heterogeneity on the wavefield amplitude is higher than for a typical reservoir with fluid substitution for the following reasons:

1. The value of the relative elastic perturbations can be high (up to 100% an more) when replacing a portion of a geological layer by a much stiffer/looser layer.

2. The structural elastic perturbations also affect V_s, contrary to the perturbations due to fluid substitution. We show in Section 4.5 that the sensitivity with respect to V_s is often the highest among the three elastic parameters.

3. The spatial extent of the structural heterogeneity is usually larger than that of a reservoir.

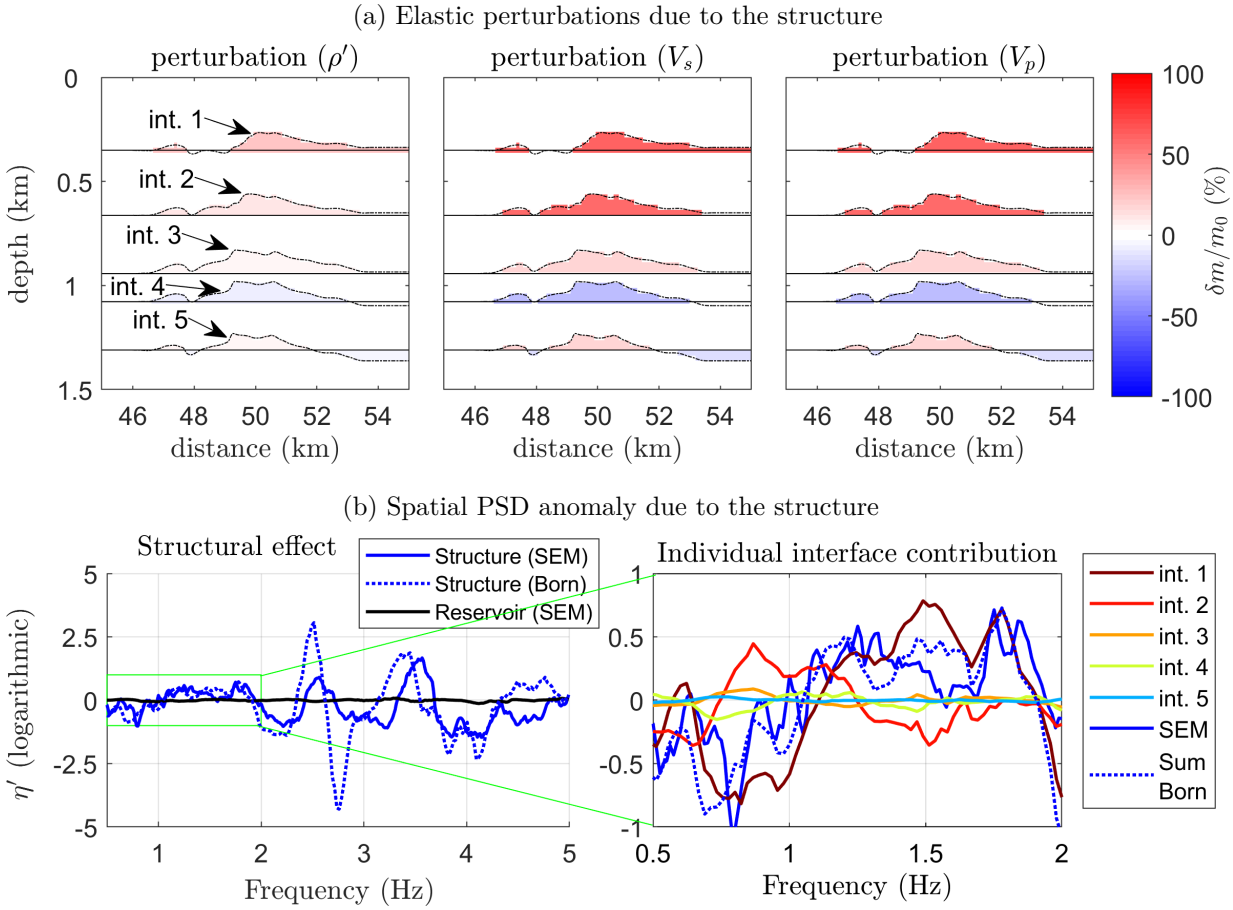


Figure 4.8: (a) Equivalent relative elastic perturbations ($\delta m/m$, with m standing for ρ, λ, μ) in the layered model with structure (Fig. 4.2) with respect to the 1D layered model (Table 4.1c). (b) η' anomalies generated by the structure at the receiver at $x = 50$ km with a reference station at $x = 47$ km, with respect to the reference layered 1D model with deep sources. Note the high magnitude of the structural anomaly compared to the reservoir-induced anomaly (black curve). The result of the SEM simulation (blue solid line) is compared to the prediction of the Born approximation (blue dotted line). Zoom on the right: contributions from different interfaces (int. 1 to 5) to the total anomaly predicted by the Born approximation.

The η-anomaly induced by the structural perturbations, computed by SEM simulations, is shown in Fig. 4.8b (blue solid line). The peak value of the structural anomaly reaches 1.67, which corresponds to a relative anomaly of 430% (or 7.25 dB), about an order of magnitude above the anomalies computed for a typical gas reservoir (black curve), in line with the qualitative predictions. The structural anomaly contains three main positive peaks: a broad peak between 1 and 2 Hz, and two narrower peaks around 2.5 and 3.5 Hz. This is consistent with the frequency range of the strong anomalies reported as direct hydrocarbon indicators by Dangel et al. (2003). In this section, we did not comment on the dotted line, referring to the Born approximation failure for *strong* contrast values, which are discussed in Section 4.5.3.

4.5 Interpretation of the spectral anomalies with Born modelling

In this section, we investigate whether the results of the previous section, obtained with SEM simulations, can be interpreted in terms of interaction between the incident Rayleigh waves and the associated scattered wavefield. Saenger et al. (2009) proposed this mechanism as a qualitative model for explaining spectral anomalies above a hydrocarbon reservoir, but involving a non-linear amplification process due oil-bubble resonance (their fig. 14). Our modelling, in contrast, is restricted to pure elasticity. As a validation procedure, we check that this simplified single-scattering model (Born approximation) is consistent with the SEM predictions. Such a representation yields a linear relationship between a model heterogeneity and the associated PSD anomalies at the surface. It presents the advantage of sparing the forward-modelling computation time, and also opens the perspective of designing a linear inversion scheme.

The application of the Born approximation to surface waves is formalised in Snieder (1986) with explicit analytic equations. The Born approximation is equivalent to consider only single-scattered waves, and to neglect multiple scattering. Friederich et al. (1993) investigate the effect of including the contribution of multiple scattering and conclude that it is significant when the horizontal extent of the heterogeneity exceeds one wavelength. Both Snieder (1986) and Friederich et al. (1993) represent surface waves as wave fronts propagating in the XY horizontal plane, with amplitude and velocity depending on the 3D properties of the medium. The scattering is treated in terms of surface wave modes only. As stressed by Maupin (1996), such a representation does not account for scattered radiation modes, including body waves which are not trapped within the structure. Since we are mainly interested in the amplitude anomalies *above or near* the heterogeneity (i.e. at steep propagation angles), these modes may have an important contribution (Maupin, 1996). In our work, all the surface wave modes and body waves excited in 2D by a given source distribution are taken into account, as we use SEM simulations to calculate the incident and the scattered fields, instead of analytic Green's functions based on modal summations.

4.5.1 Main elements of the Born approximation

Under the Born approximation, the total wavefield is written as a superposition of an incident wavefield $\mathbf{u}^{(0)}$, which would develop in the reference background model m_0, and a single-scattered wavefield $\mathbf{u}^{(1)}$ due to a small localised perturbation δm of that medium:

$$\mathbf{u}^{(\mathrm{tot})} = \mathbf{u}^{(0)} + \mathbf{u}^{(1)} + o(\mathbf{u}^{(1)}) \ . \tag{4.9}$$

Higher-order terms, corresponding to multiple-scattering, are neglected. Such a decomposition allows one to linearise the wave equation with respect to a perturbation of the medium. The linearisation leads to a sensitivity kernel $K_m^d(\mathbf{x})$ for an observable d, which tells how sensitive the observable is with respect to a perturbation of the parameter m of the model. For a given distribution of medium perturbations $\delta m(\mathbf{x})$, one can write the perturbation of the observable as a space integral:

$$\delta d = \int_\Omega K_m^d(\mathbf{x}) \delta m(\mathbf{x}) d\mathbf{x} \ , \tag{4.10}$$

where Ω is the investigated portion of the 2D plane. Eqn. (4.10) yields the direct problem for the observable prediction. While standard numerical simulations could be sufficient for solving the direct problem, deriving sensitivity kernels allows to better understand the behaviour of the spectral anomalies. Moreover, this is the first step towards the inversion of spectral amplitude anomalies for the position and the parameters of the medium perturbations. In Appendix 4.A, we briefly recall the derivation of the sensitivity kernels for the spectral amplitude. A more detailed derivation can be found in Dahlen et al. (2000); Dahlen & Baig (2002). The step by step procedure to compute the linearised scattered wavefields and PSD anomalies, as well as the validation of the approximation, are given in Appendix 4.B. We recall that our kernels reflect the sensitivity of all the wave types present in the incident wavefield, and not exclusively Rayleigh waves. However, we assume that most of the sensitivity is due to Rayleigh waves because they appear to be the dominant wave type in the simulated incident wavefield (see Section 4.4.2). In the next section, the Born-predicted η'-anomalies are compared to those modelled by SEM, shown in Fig. 4.3. The anomalies are then interpreted in terms of sensitivity kernels.

4.5.2 Born modelling of the reservoir effects

In the definition (4.5), η' can be seen as a deviation from a reference value η_0. From the Born approximation point of view, this means that $\mathbf{u}^{(0)}$ in eqn. (4.9) refers to the configuration (background medium and sources) where η_0 is calculated. Thus the sensitivity kernels must be calculated in this reference configuration. As a starting point, we consider deep sources in the LOH background medium (multi-modal wavefield). η' above the middle of the reservoir is shown with blue lines in Fig. 4.9a. The solid line, corresponding to the full SEM simulation, is the same as in Fig. 4.3a. The dotted line shows the Born approximation prediction, in good agreement with the SEM simulation. The strongest observed anomaly is about -6%, reached at $f \approx 1.25$ Hz. This anomaly is mainly driven by the density contrast, since the sensitivity to V_p is very low. The PSD sensitivity kernel for ρ' at the corresponding frequency (integrated between 1.22 and 1.32 Hz), computed via eqn. (4.A.10), is shown in Fig. 4.9b. The

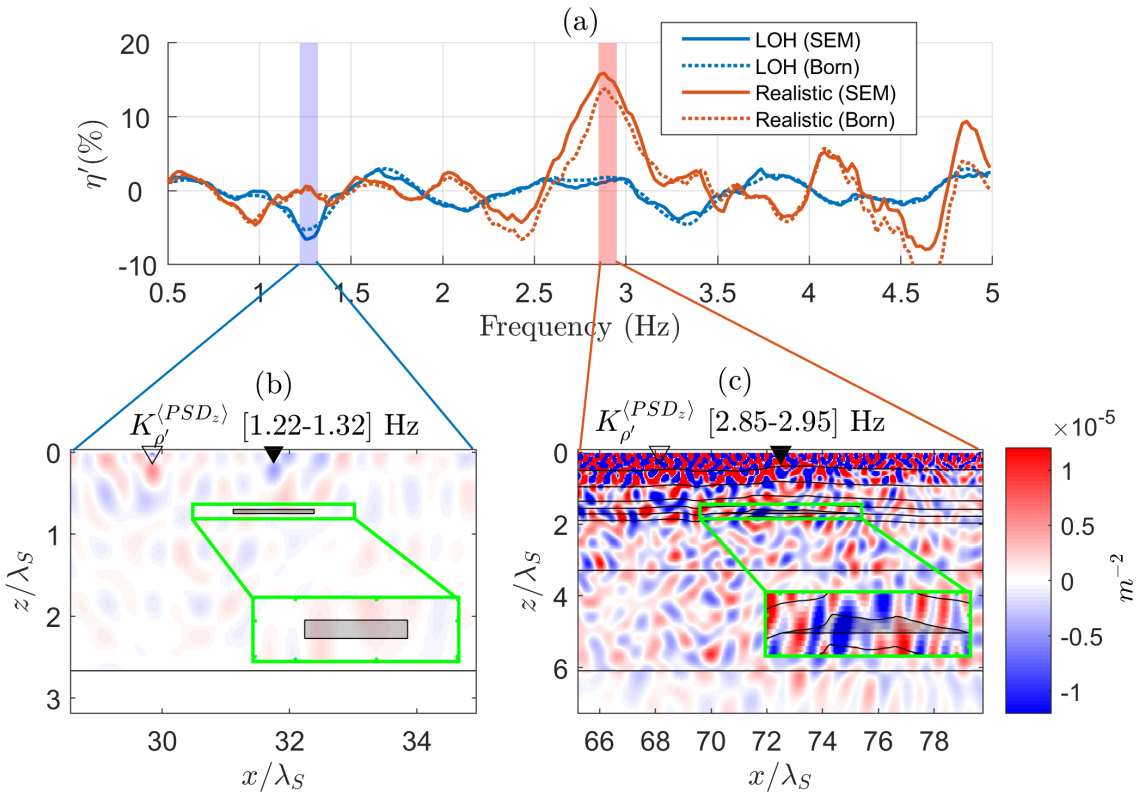


Figure 4.9: (a) $\langle \mathrm{PSD}_z \rangle$ spectral anomalies (observable η') for the receiver at $x = 50$ km (above the middle of the inclusion) and the reference station at $x = 47$ km, computed from numerical sensitivity kernels (dash-dotted lines) and SEM simulations (solid lines) in the LOH background model (blue lines) and a realistic layered background model including the anticline structure (red lines). For both cases, the incident wavefield is generated by multiple random deep sources located at both right and left ends of the model. The elastic perturbation values inside the inclusion are typical for full water saturation being replaced by gas saturation. (b) η' sensitivity kernel for ρ' integrated between 1.22 and 1.32 Hz in the LOH background medium. The black filled (resp. unfilled) triangle shows the receiver (resp. reference station) position. Horizontal and vertical coordinates are normalised by the S-wavelength assuming $V_S = 2000$ m/s, which corresponds the upper layer of the LOH model. Green box: zoom on the sensitivity inside the inclusion (shaded zone). The dominantly positive values explain the negative η' anomaly observed in this frequency band for the negative density contrast due to the substitution of water by gas. (c) Same as (b), but for the realistic background model, and integrated between 2.85 and 2.95 Hz. The zones of strong negative sensitivity inside the reservoir explain the positive η' anomaly.

dominant light-red colour (i.e. positive kernel value) inside the inclusion explains the negative η', which is observed for a negative density contrast. SEM and Born predictions are also compared for the more realistic layered model with an anticline structure (see Fig. 4.2). The corresponding η'-anomalies are shown with red lines in Fig. 4.9a. The strongest anomaly is about +17%, reached at $f \approx 2.9$ Hz. In this region we notice an increase of the mismatch between the full SEM and the Born approximation (dotted line). This is because the Born approximation is the most accurate for small wavefield perturbations (see Appendix 4.B). The PSD sensitivity kernel for ρ' at the corresponding frequency (integrated between 2.85 and 2.95 Hz), computed in the realistic background model, is shown in Fig. 4.9c. The observed positive anomaly can be explained by the strong blue patch (i.e. negative kernel value) crossing the reservoir at $x/\lambda_S \approx 74.5$. The stronger anomaly for the last model is due to the stronger heterogeneity, which generates additional spatial gradients in the incident and adjoint wavefields at the layer interfaces. The scattered wavefield, as given by eqn. (4.A.5), is sensitive to the spatial derivatives of these wavefields. This example illustrates how the PSD anomalies can be predicted from the sensitivity kernels, and shows that the background model has a significant influence on the sensitivity.

Now, in Figs 4.10 and 4.11, we show individual receiver kernels $K^{\langle \mathrm{PSD} \rangle}$ for the five configurations considered in Section 4.4. The two figures correspond to the kernels integrated respectively in the intervals $[0.45, 0.55]$ Hz and $[1.2, 1.3]$ Hz. At low frequency, the fundamental mode or the first overtone dominate for any source depth (see Fig. 4.6). This is why the sensitivity preserves a similar pattern for all the models in Fig. 4.10. This pattern is characterised by a patch of negative sensitivity below the receiver. This patch is also the prediction of the simplified model by Gorbatikov & Tsukanov (2011), where positive amplitude anomalies at the surface were expected from negative perturbations of the velocities in depth. However, the sensitivity can also have the opposite sign with a non-negligible magnitude in other parts of the model. On the contrary, at higher frequencies (Fig. 4.11), the sensitivity is higher for the realistic layered models, and rather concentrated in the shallow low-velocity layers. Such concentration can be explained by the properties of the adjoint wavefield $\mathbf{G^{rx}}$ in eqn. (4.A.3), describing the propagation between the scatterer and the receiver. This wavefield excited at the receiver gets trapped in the shallow low-velocity layers. This process explains why the sensitivity observed in the basement is lower than in the sedimentary cover, whatever the model and the source depth. The two consecutive drops of the anomaly strength pointed out in Fig. 4.7a as the reservoir enters stiffer layers can thus be interpreted in terms of amplitude sensitivity kernels. It was also observed in Figs 4.7(b) and (c) that the anomaly strength generally increases with the reservoir size (width and thickness), but reaches a constant value at some point. This can be interpreted in terms of high-frequency sensitivity kernels (Fig. 4.11). As the reservoir size increases beyond the dominant wavelength, it spans over zones with opposite-sign kernel values, which cancel out during the integration.

The single-scattering can also be invoked to explain why the spectral anomalies became unstable with respect to interface perturbations in the realistic medium in Section 4.4.1 above 3.5 Hz. From eqn. 4.A.7, we see that the spectral anomalies are governed by the phase shift $\Delta\phi$ between $u^{(1)}$ and $u^{(0)}$ at the receiver. This equation

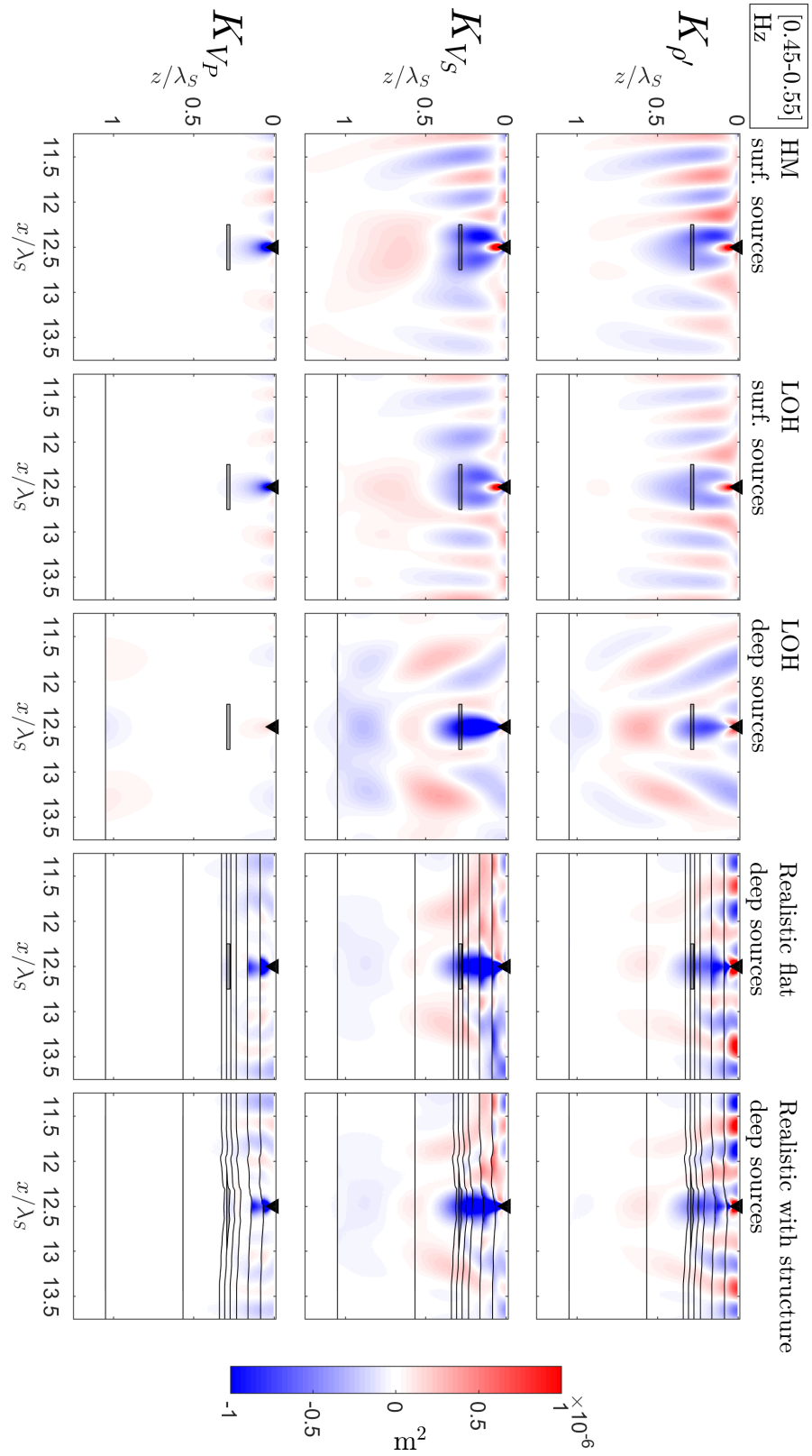


Figure 4.10: Sensitivity kernels $K_{\rho,\lambda,\mu}^{\langle \mathrm{PSD}_z \rangle}$ (see definition (4.A.9)) for average relative PSD anomaly, integrated between 0.45 and 0.55 Hz. From left to right: different combinations of models and source distributions. From top to bottom: sensitivity for ρ' (density perturbations at constant velocities), V_s and V_p. Black contours show the interfaces in the heterogeneous models.

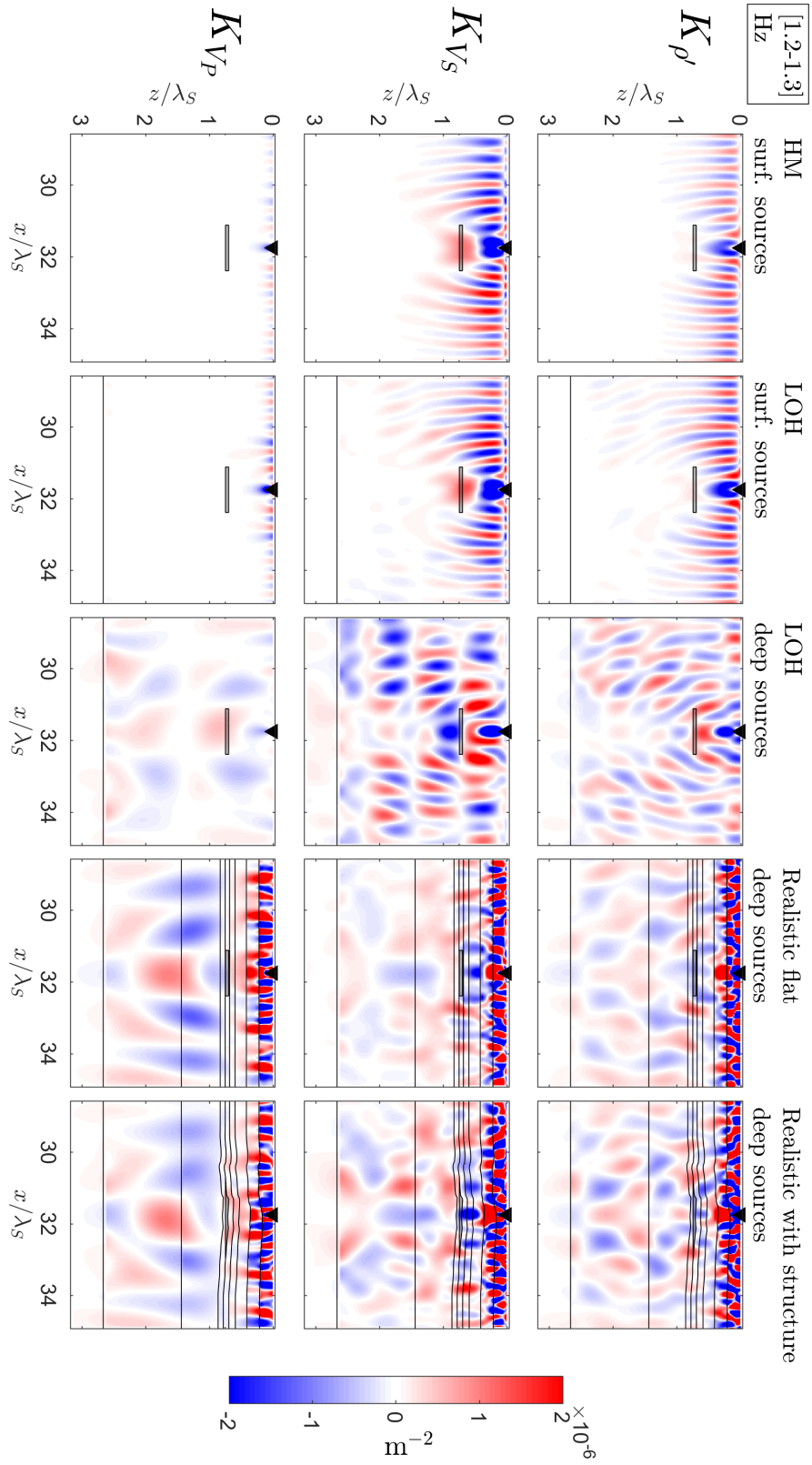


Figure 4.11: Same as Fig. 4.10, but integrated between 1.2 and 1.3 Hz.

can indeed be written as

$$\delta d^{\mathrm{PSD}}(\mathbf{x}_r, \omega) = 2\frac{|u^{(1)}|}{|u^{(0)}|}\cos(\Delta\phi) \ . \tag{4.11}$$

The spectral anomaly becomes unstable when the perturbation of the layering induces a too large additional phase shift $\delta\phi$. For example, if the additional phase shift reaches $\pi/2$, the sign of the computed spectral anomaly changes with respect to the true value, which can be regarded as a criterium of complete unstability. In order to determine at which frequencies the unstability is expected to occur, we build a conceptual model for the scattered wave propagation. We note that a significant portion of path from the scattered to the receiver corresponds to near-field propagation, where P and S waves contribution to compressional and shear motions are mixed (Wu & Ben-Menahem, 1985). If we first assume that the scattered waves propagate at the velocity V_p, we have

$$\delta\phi = \omega \sum_{i=1}^{n} \Delta z_i \left(\frac{1}{V_p^{(i)}} - \frac{1}{V_p^{(i+1)}} \right) , \tag{4.12}$$

where n is the total number of layers, Δz_i is the depth perturbation of the bottom interface of the layer i, and $V_p^{(i)}$ is the P-wave velocity of the layer. The dominant contribution to $\delta\phi$ comes from interfaces with a strong slowness contrast. The typical depth perturbation Δz_i corresponds to $\sigma = 0.2 \cdot (z_{i+1}^{(0)} - z_{i-1}^{(0)})$ in eqn. (4.7). Thus we can estimate the threshold frequency f_{crit}:

$$\delta\phi_{\mathrm{crit}} = \frac{\pi}{2} \sim 0.2 \cdot 2\pi f_{\mathrm{crit}} \sum_{i=1}^{5} (h_i + h_{i+1}) \left(\frac{1}{V_p^{(i)}} - \frac{1}{V_p^{(i+1)}} \right) , \tag{4.13}$$

with h_i the thickness of the layer i in Table 4.1c, yielding $f_{\mathrm{crit}} \sim 5.6$ Hz. If we had assumed the scattered wavefield propagates at the shear-wave velocity (requiring a slightly oblique incidence to have a non-zero vertical component), the result would be $f_{\mathrm{crit}} \sim 2.3$ Hz. Thus both estimates yield an interval containing the observed value $f_{\mathrm{crit}} = 3.5$ Hz in Section 4.4.1. Of course this interpretation is qualitative, and only intends to illustrate that the threshold frequency of the sensitivity to layering perturbations is controlled by the phase shift between the incident and the scattered wavefields.

The conclusion here is that the Born approximation allows to qualitatively interpret the SEM-predicted reservoir effects on the spectral anomalies through the analysis of the sensitivity kernels. In the next section, we discuss the limits of the Born approximation.

4.5.3 Structural effects: limits of the Born approximation

In this section we illustrate the limits of the linearisation by comparing the prediction of the Born theory to the SEM-simulated spectral anomaly generated by structural effects (see Section 4.4.4). As stressed by Mercerat & Nolet (2013), a Born-predicted attribute of a seismogram (arrival time in their case) contains two linearisation steps. The first step neglects higher-order terms (related to multiple-scattering) in the full scattered waveform (eqn. 4.9). The second step assumes that the extracted attribute

depends *linearly* on the single-scattered wavefield $u^{(1)}$. In our case, the second step takes place in eqn. (4.A.7), where we linearise the PSD anomaly. This step is likely to generate an under-estimation of positive PSD anomalies, since the neglected term $|u^{(1)}/u^{(0)}|^2$ is positive. The first linearisation is limited by the importance of the multiple-scattering with respect to single-scattering, while the second is limited by the importance of the single-scattered wavefield with respect to the incident wavefield. In Appendix 4.B, Fig. 4.B.3, we display the approximation errors on the PSD, which contain the cumulated effect of both linearisation steps. This allows us to verify that the Born approximation can be used for predicting the PSD anomalies generated by small enough localised inclusions with reasonable values of elastic contrasts, such as the simplified gas reservoir in Section 4.5.2. In order to see if the Born approximation can be used for predicting Rayleigh wave amplitude anomalies due to structural heterogeneities, we apply the sensitivity kernel computed in the flat layers medium (Table 4.1c) to the perturbations displayed in Fig. 4.8. The Born approximation (Fig. 4.8b, black dotted line) fails at frequencies above 2 Hz. We suggest that this is due to the breakdown of the first linearisation step, as multi-scattered trapped waves appear within the anticline. This interpretation is supported by the overestimated positive anomalies around 2.5 and 3.5 Hz, which would be underestimated if the error was due to the second linearisation step. $f = 2$ Hz then appears as the lowest frequency at which the waves start being efficiently trapped by the anticline. In Fig. 4.12, we display η (logarithmic PSD anomaly with respect to the reference surface station at $x = 47$ km, eqn. (4.3)) in the central part of the model, at depths between 0 and 1.5 km, for increasing frequencies from 0.5 to 3.6 Hz. This allows to see how the wavefield energy is distributed spatially within the model at different frequencies, and to check if some of its amount is indeed trapped by the anticline as suggested above. At frequencies below 2 Hz (Figs 4.12a and b), the lateral variations of the PSD are limited. The amplitude anomaly variations with depth correspond well to the fundamental mode (Fig. 4.12a) and higher-mode (Fig. 4.12b) Rayleigh-wave eigenfunctions propagating in the flat layered model (see Fig. 4.6i). Above 2 Hz, however, high-energy zones arise in both flanks of the anticline structure. This supports the idea of trapped multi-scattered waves above 2 Hz.

We suggest to interpret this threshold frequency in terms of S-wave resonance in the different layers. The fundamental resonant frequency for an individual layer of thickness h and velocity V_s, not in contact with the stiff bedrock (no π phase shift on reflection), is

$$f_0 = \frac{V_s}{2H} \ . \tag{4.14}$$

Assuming $V_S \sim 1300$ m/s, $f_0 = 2$ Hz corresponds to a typical layer thickness of $h \sim 325$ m, compatible with the thickness of the the layer L2 in Table 4.1c, where the energy is trapped in Fig. 4.12c.

Nevertheless, the Born approximation below 2 Hz is acceptable despite the strong model perturbations, since the latter affect a small enough zone compared to the wavelength. This allows one to analyse the contributions from the individual interface perturbations (int. 1 to 5 in Fig. 4.8a) to the total anomaly at these low frequencies (green zoom in Fig. 4.8b). It turns out that the topography of the first two interfaces

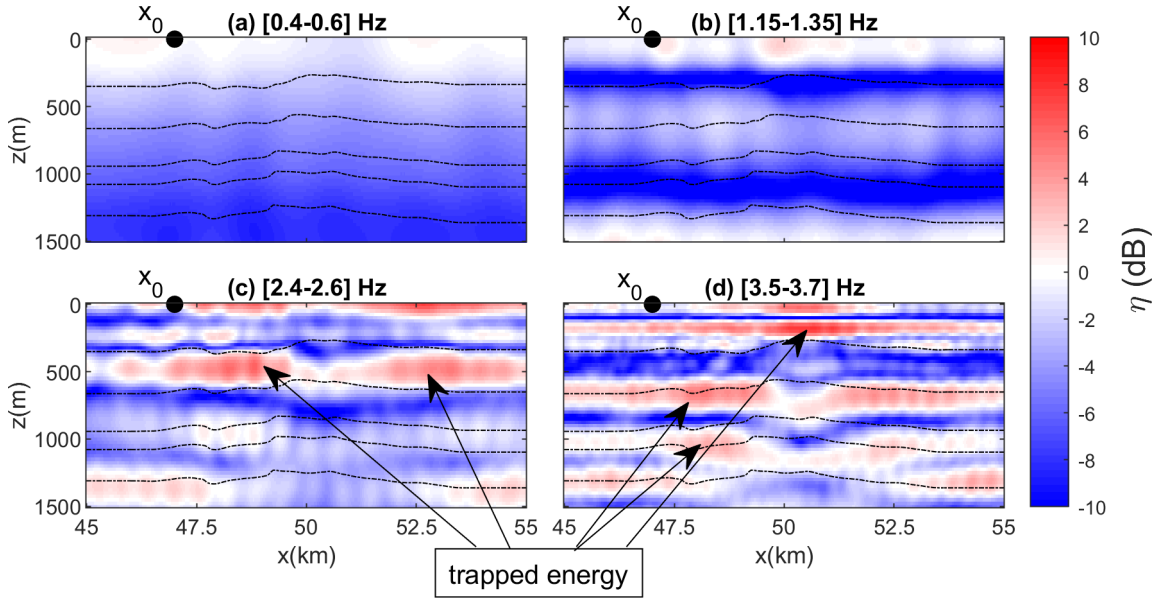


Figure 4.12: SEM-simulated η (measured in decibel) within the realistic model with anticline structure. Interfaces between layers are shown with dotted lines. The reference receiver is the black dot. From (a) to (d) - approximate η averaged in four different frequency bands.

(brown and red lines) are alone able to explain the pattern of the anomaly between 0.5 and 2 Hz.

4.6 Discussion

4.6.1 Objectives and scope

In this study, we propose an elastic modelling procedure for the ambient wavefield amplitude above an embedded heterogeneity at frequencies around 1 Hz. We tried to ensure that Rayleigh waves dominate the simulated signals. In the field data, Rayleigh waves at these frequencies can be present as fundamental mode due sources at the surface, but also as a mixture of overtones, likely due to natural sources in the ocean (Koper et al., 2010; Peruzzetto et al., 2018). In our work, we addressed both situations separately by adjusting the depth of the random source distributions in the simulations. The final objective was to study which kind of amplitude perturbations these two types of wavefields could induce at the surface due to interaction with a gas reservoir. We did not study the behaviour of individual overtones, which could be achieved by applying particular initial conditions in SEM simulations (plane wave with depth-dependant amplitude following a particular modal eigenfunction). Though individual overtones are also likely to be present in the ambient wavefield (e.g. Ma et al., 2016), the two situations we studied here can be considered as the two limiting cases of shallow- and deep- penetrating incident wavefields. They should thus yield bounds on the orders of magnitude which can be expected for the amplitude perturbations.

4.6.2 Amplitude anomaly strength and behaviour

Our study confirmed the possibility of frequency-dependent amplitude anomalies above an inclusion representative of a gas reservoir embedded at $z = 1100$ m, driven by purely elastic effects. Direct SEM simulations (Fig. 4.3) showed that the PSD spectral anomalies of the fundamental mode, measured by the normalised observable η' (eqn. (4.5)), were small (less then 1%) compared to the anomalies of an overtone mixture (about 17%). Consequently, we focused on the latter case (simulations with deep sources) in the remainder of the study, as it offers the most realistic perspective of measurable amplitude pertrubations. For the overtone mixture, we investigated the stability of the observed anomalies, the influence of the reservoir parameters, as well as the influence of a 2D geological structure (an anticline).

The stability was studied via the sensitivity of the spectral anomalies with respect to the uncertainty on the background model (the depth of the bedrock in the LOH model (Table 4.1b), and the depth of the interfaces between the layers in the realistic model (Fig. 4.2)). The pattern was stable with respect to the bedrock depth as far as the bedrock was not too close to the reservoir (Section 4.4.1). The pattern was also stable with respect to the interface depth (Section 4.4.1), as far as the cumulated uncertainty on the scattered wave phase shift did not exceed $\pi/2$ at the considered frequency.

The anomaly strength decreases with the reservoir depth, and increases with the reservoir size, though the latter increase stops as the reservoir becomes $\sim$ 4 km wide and $\sim$ 200 m thick (see Section 4.3.2). We also found that the dominant frequency of the anomaly obtained with deep sources (overtone-dominated field) did not exhibit any clear decreasing trend with increasing reservoir depth, contrary to a fundamental-mode-dominated wavefield, which we further verified with the modal summation theory in Appendix 4.C.

In the same time, because of the small strength of the maximum simulated anomaly, elastic scattering of Rayleigh waves has been eliminated as a potential mechanism of strong amplitude distorsions of several dB above reservoirs, such as reported by Dangel et al. (2003). On the other hand, such strong site effects could well be explained by elastic scattering structural features (e.g, anticlines) often associated with hydrocarbon reservoirs (see Fig. 4.8).

4.6.3 Single-scattering interpretation

In a second step, we investigated if the single-scattering theory (Born approximation) was enough to explain the modelled anomalies in a linear way. Here it is interesting to stress that using diffraction modelling for surface wave amplitude is unusual, since the latter can also be interpreted through the layered medium eigenfunctions (modal summation Aki & Richards, 2002). The latter approach relies on the assumption that the medium is laterally invariant (locally), and was successfully applied at continental scale (Lin et al., 2012). However, the lateral invariability is no longer valid for inclusions with sizes comparable to the wavelength, and diffraction modelling becomes

necessary. At $f = 1$ Hz, the typical Rayleigh wavelength is kilometric in a sedimentary basin. Thus, diffraction modelling is required for reservoir-scale heterogeneities.

Implementing the Born approximation conducted us to compute the sensitivity kernels for spectral anomalies with different background models and wavefield compositions. Their characteristics can be summarised as following:

- For the fundamental mode, there is a strong decay with depth, whatever the background model, which means weak sensitivity to deep reservoirs. This is in line with experimental results obtained in previous works (Nasseri-Moghaddam et al., 2007; Tallavó et al., 2009).
- For the overtones, there is a stronger sensitivity in depth than for the fundamental mode, consistently with the modal eigenfunctions.
- The kernels are amplified in the shallow slow layers, if they are present, and also in the vicinity of the receivers, where there is a singularity.

Some asymmetry can be observed in the sensitivity kernels, while one would expect them to be perfectly symmetric for symmetric sources at both ends of the model. This is due to an imperfect convergence of the average sensitivity kernels. Each realisation of the source distribution is not perfectly symmetric, and more realisations would be needed to achieve a perfect symmetry of the average kernel. For ambient noise applications, however, such a symmetry is not expected. In case-specific numerical 3D simulations, the number of random sources for each azimuthal bin should be proportional to the corresponding beamforming energy. In general, this would result in asymmetric kernels. Another interesting observation is that the sensitivity to V_p at higher frequencies (Fig. 4.11) is so much enhanced by the background model heterogeneities that is exceeds the sensitivity to V_s in the deep parts of the realistic models (the two right-most columns in Fig. 4.11). The reasons of this phenomenon are probably related to the properties of the tensor gradients and divergences used in eqn. (4.A.3) and deserve a closer analysis, beyond the scope of our work.

We found the Born approximation was enough for predicting the small reservoir effects. This opens the possibility for a linear inversion for the reservoir position and properties. However, the high sensitivity in the vicinity of the receiver, as well in the shallow low-velocity layers, might complicate the use of the kernels for spectral anomaly inversion. Preconditioning could be needed to balance the sensitivity over depth and distance from the receiver. We also found that the Born approximation fails to predict strong structural effects, which we explain in terms of an increasing importance of multi-scattering due to trapped waves within the structure beyond a frequency of about 2 Hz.

4.6.4 Role of numerical modelling

The first advantage of using numerical modelling as compared to modal summation is to take into account all the wave types in the scattered wavefield. In particular, the Born approximation involves two terms: the regular and the adjoint wavefields

(see eqn. (4.A.3)), which describe respectively the source-inclusion and the inclusion-receiver propagation (Fig. 4.1). Since the inclusion can be located at the same time at a sub-wavelength distance and at a sub-vertical angle with respect to the receiver, it is important to ensure that both near-field effects and body waves are included in the Green's tensor used for the adjoint wavefield calculation. This was achieved by using the Spectral Element Method (SPECFEM2D, Komatitsch et al., 1999) to numerically estimate all the Green's tensors. Moreover, the advantage of the numerical simulation is that arbitrary reference media can be used, while the use of analytic Green's functions (e.g. Maupin, 2017) is restricted to layered horizontally invariant media.

Since we use numerical simulations and not analytic Green's functions, the simulated incident fields do not contain exclusively surface waves. Therefore, it could be debated to which extent the sensitivity of other wave types contributes to our results. In Section 4.4.2, we demonstrate the wavefront contains only dispersive arrivals, which indicates that the body wave contribution is weak. However, the so called leaky modes are found in the dispersion plots (Fig. 4.6). These modes could bias the results, as they have not been reported as part of the ambient wavefield, to our knowledge. Thus a comparison of our sensitivity kernels with analytic kernels for surface waves in laterally invariant medium would be a valuable contribution.

4.6.5 Potential for practical applications

The practical applications of our modelling work could concern ambient-noise based imaging and monitoring. Monitoring applications would require a good knowledge of the velocity model (including structural deformations) and of the frequency-dependent modal composition of the ambient noise. Using these two pieces of information, sensitivity kernels can be computed, allowing one to invert small time-lapse variations of the ambient noise amplitude for localised elastic parameter variations due to fluid substitution within a reservoir assuming the noise modal content is not changing between the two snapshots in time. In 3D, assuming constant back-azimuth between the snapshots would also be necessary. While the stability of the modal composition is likely to be verified (Peruzzetto et al., 2018), the back-azimuth is expected to present strong variations (Hillers et al., 2012; Lehujeur et al., 2015). Selecting time periods with similar back-azimuth distribution (and modal content) could thus be a necessary step for real-data applications. Note that constant amplitude assumption is not needed, since the observable η' is normalised by a reference station. The knowledge of the reservoir location would be a very useful constraint for this application, as it would reduce the model space and avoid artifacts in high-sensitivity zones (shallow layers). On the other hand, imaging applications would exploit the spatial variations of the noise amplitude above a site with unknown structure. As a first step, the 1D velocity model and the ambient noise modal composition could be obtained from cross-correlation or array methods (e.g. Peruzzetto et al., 2018). As a second step, the sensitivity kernels could be computed in order to invert the amplitude variations for the local perturbations of the 1D velocity model. However, the frequencies between 1 and 5 Hz appear to be mainly sensitive to the perturbations of the shallow

low-velocity layers, even in presence of higher-mode Rayleigh waves (see zoom in Fig. 4.8b). Also, structural effects are likely to be too strong to be handled by the Born approximation. As a consequence, a non-linear inversion would be necessary, with an appropriate preconditioning to compensate for the sensitivity concentration in shallow layers.

4.6.6 Limitations

One limitation of our study is that it was performed in 2D. Gorbatikov & Tsukanov (2011) showed that standing waves developed in 2D due to interferences between incident and scattered waves, generating side-lobes in the amplitude anomalies on both sides of the heterogeneity. In 3D, this effect disappeared if the source wavefield was evenly distributed over azimuths. The same effect was observed by Maupin (2017), where the side-lobes in the 3D sensitivity kernels disappeared after averaging over all the source directions. Thus, the side-lobes observed in our sensitivity kernels on both sides of the receiver (Figs 4.10 and 4.11) might be a 2D effect. Nevertheless, we believe that these 2D results, "cheap" in terms of computation time compared to 3D, are useful to introduce the proposed methodology, and can serve as a guide for a more realistic 3D-modelling. It would definitely be interesting to compare the results in 2D and 3D.

It must also be mentioned that our study is purely elastic, and thereby neglects fluid-specific visco-elastic effects which might be important. Therefore, it does not rule out the possibility of stronger effects due to fluid-saturated reservoirs on the ambient noise amplitude. Quintal (2012) and Quintal et al. (2014) numerically studied the effect of hydrocarbon saturation on the attenuation of both P and S waves by using a low-frequency (static) version of the poro-elastic equations. Provided some particular reservoir rock properties (elastic matrix heterogeneities, double porosity, fault orientations), they found that a two-phase mixture of gas and liquid (patchy saturation) could induce a strong decrease of the quality factor Q_p and sometimes Q_s. Tisato et al. (2015) showed that for very low saturation values, microscopic gas bubbles at the pore scale introduced an additional dissipation mechanism with a characteristic frequency of about 1 Hz, which they called "Wave Induced Gas Exsolution-Dissolution" (WIGED). This also results in a drop of Q_p. Our method must thus be extended to include the sensitivity of the surface wave amplitude to the quality factor perturbations for better modelling fluid-specific effects. This is planned as future work. As an example, fluid-specific viscoelastic effects on the ambient noise amplitude were addressed by Lambert et al. (2013), but their random source distributions covered areas directly beneath the reservoir. In our opinion, this generates an unrealistically high proportion of body waves in the wavefield.

Another limitation concerns the assumption that the multi-modal Rayleigh waves completely dominate the wavefield energy on the vertical component. P-waves have also been reported within the [1-5] Hz band in some ambient noise datasets (e.g. Poli et al., 2012; Nakata et al., 2015; Peruzzetto et al., 2018). Stronger sensitivity to V_p can be expected for these waves, while V_p is also the most sensitive to fluid substitution

among the elastic parameters. A similar study with an appropriate source distribution for generating P-waves would definitely bring some additional insights. In general, the knowledge of the different wave type contributions (Rayleigh, P and S) to the total energy on the vertical component is required at each frequency to achieve a proper quantitative interpretation of the amplitude anomalies. This still remains a challenge for the ambient noise analysis.

4.7 Conclusion

Rayleigh wave amplitude anomalies (vertical component) at the surface above an embedded reservoir-scale elastic inclusion were computed in 2D between 0.5 and 5 Hz. The waves represented a potential ambient noise field excited by a distribution of random sources. Both deep and shallow sources were tested to excite modal compositions typical for human noise (dominant Rayleigh fundamental mode) and distant natural sources (dominant higher modes). Several background models were tested. The study was performed by means of both fully numerical simulation (Spectral Element Method) and semi-analytic Born approximation based on SEM-estimated Green's functions in the background medium. The good agreement between both approaches confirmed that diffraction was the relevant mechanism for this problem at these frequencies. Frequency-dependent anomalies were predicted by our purely elastic model. The presence of higher mode Rayleigh waves (deep sources) was necessary for obtaining a non-zero sensitivity at depths beyond the S-wavelength. The amplitude sensitivity within the reservoir region was primarily associated with V_s, then with ρ, and was quite weak for V_p. On the other hand, the gas saturation in a reservoir mainly affects V_p and ρ. Thus, for realistic elastic perturbations associated to a gas reservoir in sandstone ($\delta V_p/V_p = -18\%$ and $\delta\rho/\rho = -10\%$), embedded in a realistic background model, the amplitude anomalies were limited (under 20%, or 0.8 dB). For comparison, the effect of a realistic structural deformation (anticline) generated anomalies up to 430%, or 7.25 dB, again with deep sources. As expected, the Born approximation fails for such strong perturbations. Because the structural effect appears to be predominant, time-lapse observations of the amplitude would be necessary to retrieve small fluid-related anomalies ("monitoring"). On the other hand, a snapshot of the spatial variations of the amplitude could yield information about the geological structure beneath the receivers ("imaging"). For small perturbations, the Born-approximation based sensitivity kernels open the possibility of imaging the model heterogeneities from the observed amplitude anomalies, though the maximum depth of investigation can be limited by the shallow layers without a proper preconditioning. Extending the method to 3D, as well as implementing an inversion, are beyond the scope of this study, dedicated to forward 2D modelling.

Acknowledgments

The authors are grateful to Storengy for granting this work and providing access to the clusters for numerical simulations. They also thank the two anonymous reviewers for their constructive critics and suggestions.

Appendix

4.A Born sensitivity kernels derivation

In this Appendix, we briefly describe the derivation of the sensitivity kernels used in Section 4.5.2 to interpret the results of SEM-modelling. In the reference medium, the wavefield excited by a source located at $\mathbf{x_s}$ and recorded by a receiver located at $\mathbf{x_r}$ is fully described by the Green's tensor $\mathbf{G}(\mathbf{x_r}, \mathbf{x_s})$, or $\mathbf{G^{rs}}$. By virtue of the superposition principle, knowing the first-order perturbation of the Green's tensor $\delta\mathbf{G^{rs}}$ due to an elastic perturbation δm inside Ω, the scattered wavefield at the receiver in the frequency domain can be written as

$$\begin{pmatrix} u_x^{(1)}(\mathbf{x_r}, \omega) \\ u_z^{(1)}(\mathbf{x_r}, \omega) \end{pmatrix} = \begin{pmatrix} \delta G_{xx}^{\mathrm{rs}}(\omega) & \delta G_{xz}^{\mathrm{rs}}(\omega) \\ \delta G_{zx}^{\mathrm{rs}}(\omega) & \delta G_{zz}^{\mathrm{rs}}(\omega) \end{pmatrix} \begin{pmatrix} s_x(\omega) \\ s_z(\omega) \end{pmatrix} , \tag{4.A.1}$$

where $s_x(\omega)$ and $s_z(\omega)$ are the Fourier transforms of the source-time function in both components of a directional point force. As shown in Dahlen et al. (2000, their eqn. 36), the perturbed Green's tensor can be approximated to first order in the Lamé coefficients and density perturbations $(\delta\lambda(\mathbf{x}), \delta\mu(\mathbf{x}), \delta\rho(\mathbf{x}))$ as

$$\delta\mathbf{G^{rs}} = \int_\Omega \frac{\delta\rho}{\rho_0}(\mathbf{x})\mathbf{K}_\rho(\mathbf{x_r}, \omega) + \int_\Omega \frac{\delta\lambda}{\lambda_0}(\mathbf{x})\mathbf{K}_\lambda(\mathbf{x_r}, \omega) + \int_\Omega \frac{\delta\mu}{\mu_0}(\mathbf{x})\mathbf{K}_\mu(\mathbf{x_r}, \omega) \ , \tag{4.A.2}$$

with the following expressions of the sensitivity kernels

$$\begin{cases} \mathbf{K}_\rho(\mathbf{x_r}, \omega) = \rho_0(\mathbf{x})\omega^2 \mathbf{G^{rx}} \cdot \mathbf{G^{xs}} \\ \mathbf{K}_\lambda(\mathbf{x_r}, \omega) = \lambda_0(\mathbf{x})(\nabla^{\mathrm{x}} \cdot (\mathbf{G^{rx}})^{\mathrm{T}} (\nabla^{\mathrm{x}} \cdot \mathbf{G^{xs}}) \\ \mathbf{K}_\mu(\mathbf{x_r}, \omega) = \mu_0(\mathbf{x}) (\nabla^{\mathrm{x}}\mathbf{G^{rx}})^{\mathrm{T}} : \left[\nabla^{\mathrm{x}}\mathbf{G^{xs}} + (\nabla^{\mathrm{x}}\mathbf{G^{xs}})^{\mathrm{T}}\right] \ , \end{cases} \tag{4.A.3}$$

where the transposes and the double contraction are on the first two indices. The upper-script in $\nabla^{\mathbf{x}}$ indicates that the differential operator must be computed at the scatterer location. The terms involved in this expression are illustrated in Fig. 4.1 in terms of wave propagation (see the zoomed area). Multiplying eqns (4.A.2) and (4.A.3) by the source term $(s_x, s_z)^{\mathrm{T}}$, and using the Einstein index notation, we get the linearised scattered wavefield on the component i as

$$\delta u_i^{(1)} = \int_\Omega \frac{\delta\rho}{\rho_0}(\mathbf{x})K_{\rho,i}(\mathbf{x_r}, \omega) + \int_\Omega \frac{\delta\lambda}{\lambda_0}(\mathbf{x})K_{\lambda,i}(\mathbf{x_r}, \omega) + \int_\Omega \frac{\delta\mu}{\mu_0}(\mathbf{x})K_{\mu,i}(\mathbf{x_r}, \omega) \ , \tag{4.A.4}$$

with the associated *waveform sensitivity kernels*:

$$\begin{cases} K_{\rho,i} = \omega^2 \widetilde{G}_{ik}^{\mathrm{rx}} u_k^{(0)} \\ K_{\lambda,i} = -(\partial_k^{\mathrm{x}} \widetilde{G}_{ik}^{\mathrm{rx}})(\partial_l^{\mathrm{x}} u_k^{(0)}) \\ K_{\mu,i} = -\partial_l^{\mathrm{x}} \widetilde{G}_{ki}^{\mathrm{rx}} \left[\partial_k^{\mathrm{x}} u_l^{(0)} + \partial_l^{\mathrm{x}} u_k^{(0)}\right] \ . \end{cases} \tag{4.A.5}$$

Now, we show how to obtain suitable amplitude kernels based on the waveform kernels, and how to change the parametrisation. In order to work with relative anomalies of the PSD, we design an auxiliary logarithmic observable:

$$d^{\mathrm{PSD}}(\mathbf{x}_r, \omega) = \log[\mathrm{PSD}_z(\mathbf{x}_r, \omega)/\mathrm{PSD}_0], \tag{4.A.6}$$

with $\mathbf{x}_r$ the position of the receiver and PSD_0 some arbitrary constant. Differentiating d^{PSD}, we get

$$\begin{aligned}\delta d^{\mathrm{PSD}}(\mathbf{x}_r,\omega) &= \frac{\delta \mathrm{PSD}_z(\mathbf{x}_r,\omega)}{\mathrm{PSD}_z^{(0)}(\mathbf{x}_r,\omega)} \\ &= 2\cdot\mathfrak{Re}\left(\frac{u_z^{(1)}(\mathbf{x}_r,\omega)u_z^{*(0)}(\mathbf{x}_r,\omega)}{u_z^{(0)}(\mathbf{x}_r,\omega)u_z^{*(0)}(\mathbf{x}_r,\omega)}\right)+o\left(\frac{u_z^{(1)}}{u_z^{(0)}}\right)\ .\end{aligned} \tag{4.A.7}$$

Comparing this to the eqn. (39) in Liu & Zhou (2016), one can see that the kernels for d^{PSD} are twice the classical amplitude kernels. In eqn. (4.A.7), the only unknown is the single-scattered wavefield $u_z^{(1)}(\mathbf{x},\omega)$, which was expressed as a function of the model perturbations in eqn. (4.A.4). The PSD sensitivity kernels are readily obtained by expressing $u_z^{(1)}(\mathbf{x},\omega)$ in eqn. (4.A.7) through the waveform kernel (eqn. (4.A.5)), yielding

$$K^{\mathrm{PSD}}_{\rho,\lambda,\mu}(\mathbf{x_r},\omega)=2\cdot\mathfrak{Re}\left(\frac{K_{(\rho,\lambda,\mu),z}(\mathbf{x_r},\omega)u_z^{*(0)}(\mathbf{x_r},\omega)}{\left|u_z^{(0)}(\mathbf{x_r},\omega)\right|^2}\right)\ . \tag{4.A.8}$$

We consider the case of both individual and multiple sources. For the latter case, we extend the definition of the PSD sensitivity kernel for predicting the average PSD introduced in eqn. (4.6). Once the waveform kernel $K_{\rho,\lambda,\mu}$ has been computed for each realisation (time window), the kernel for the average PSD can be obtained as following:

$$K^{\langle\mathrm{PSD}\rangle}_{\rho,\lambda,\mu}=2\cdot\mathfrak{Re}\left(\frac{\langle K_{(\rho,\lambda,\mu),z}.u_z^{(0)*}\rangle}{\langle|u_z^{(0)}|^2\rangle}\right)\ . \tag{4.A.9}$$

After several realisations of wavefields with similar modal content, $K^{\langle\mathrm{PSD}\rangle}_{\rho,\lambda,\mu}$ converges to a stable value. The number of the necessary realisations depends on the background model and on the range of source parameters covered by the random distribution. In our case, taking 20 realisations (see Section 4.3.3) yielded a satisfactory result (i.e. the shape of the kernel became relatively stable). From the definition (4.5), the sensitivity kernel for η' is

$$K^{\eta'}_{\rho,\lambda,\mu}(\mathbf{x}_r,\mathbf{x}_0,\omega)=K^{\mathrm{PSD}}_{\rho,\lambda,\mu}(\mathbf{x}_r,\omega)-K^{\mathrm{PSD}}_{\rho,\lambda,\mu}(\mathbf{x}_0,\omega)\ , \tag{4.A.10}$$

where the subscript "PSD" refers indifferently to a single-window or an average estimation. η' is well suited for Born-modelling because it is defined with respect to a reference medium, i.e. the one where the kernels should be calculated. In order to express the kernels in terms of (V_p,V_s,ρ) instead of the Lamé coefficients, we use eqns (19) from Maupin (2017):

$$\begin{aligned}K_{\rho'} &= K_\rho+K_\lambda+K_\mu \\ K_{V_s} &= 2\left(K_\mu-2\frac{\mu_0}{\lambda_0}K_\lambda\right) \\ K_{V_p} &= 2\left(1+2\frac{\mu_0}{\lambda_0}\right)K_\lambda\ .\end{aligned}$$

This parametrisation is of practical convenience for working with numerical simulations since it allows one to change ρ while maintaining constant V_p and V_s. The kernel $K_{\rho'}$ takes into account the changes of λ and μ which are necessary to keep V_p and V_s

constant. In the following, we refer to density perturbations at constant velocities as $\delta\rho'$.

We use the reciprocity of the Green's tensor

$$\mathbf{G^{xr}} = (\mathbf{G^{rx}})^{\mathrm{T}} \tag{4.A.11}$$

in order to implement eqns (4.A.5) with only two numerical simulations: one with a source in $\mathbf{x_s}$ (regular wavefield) and another with a source in $\mathbf{x_r}$ (adjoint wavefield). A sum of two Ricker wavelets of central frequencies $f_0 = 0.5$ Hz and $f_0 = 2.5$ Hz is used as the source time function for both the regular and the adjoint sources in order to cover a large frequency band. Eventually, the resulting sensitivity kernels must be deconvolved by this source-time function.

4.B Born sensitivity kernels implementation and validation

In this Appendix, we validate the Born approximation prediction of the single-scattered wavefields and the amplitude anomalies associated with inclusions similar to the one described in Section 4.3.2, containing different elastic perturbations. The procedure we use to compute the SEM-based sensitivity kernels is as following:

1. Define Ω, the domain where the incident wavefield and the adjoint Green's tensors must be stored (tabulated). We use a 10 x 5 km region centred on the central receiver, disrectised in 20 m x 20 m cells.

2. Compute and store over Ω the regular time-domain wavefield $\mathbf{u^{(0)}}(\mathbf{x}, t)$ using SPECFEM2D (Komatitsch et al., 1999). Only vertical point force sources are used. The source-time function is a sum of two Ricker wavelets of central frequencies $f_0 = 0.5$ and $f_0 = 2.5$ Hz.

3. Compute, transpose and store the adjoint time-domain Green's tensor $\widetilde{\mathbf{G}}_{\mathbf{xr}}(\mathbf{x}, t)$ obtained from two SPECFEM2D simulations with a point force source placed at the receiver. The source-time function is the same as in step (ii). The $\widetilde{\ }$ indicates that the variable contains the source-time function signature which is to be deconvolved at the end. One simulation with a horizontal force and one with a vertical force are required, so that the four terms of $\widetilde{\mathbf{G}}_{\mathbf{rx}} = \widetilde{\mathbf{G}}_{\mathbf{xr}}^{\mathrm{T}}$ can be retrieved.

4. Perform a temporal Fourier transform on all the previously defined time-domain variables.

5. Using the finite-difference numerical gradient, implement eqn. (4.A.5) to obtain the waveform sensitivity kernels $\widetilde{K}$.

6. Deconvolve the wavefield kernel from the source-time function:

$$K_{(\rho,\lambda,\mu),z} = \frac{\widetilde{K}_{(\rho,\lambda,\mu),z}.s^*}{|s|^2+\epsilon_{\mathrm{deconv}}}, \tag{4.B.1}$$

where we include ϵ_{deconv} as a water-level parameter for sake of generality, but which we take equal to zero since the source spectrum never vanishes in our frequency band (between 0.5 and 5 Hz).

7. Obtain the PSD kernel by implementing eqn. (4.A.8):

$$K^{PSD}_{\rho,\lambda,\mu} = 2 \cdot \mathfrak{Re} \left(\frac{K_{(\rho,\lambda,\mu),z} . u_z^{(0)}}{\left| u_z^{(0)} \right|^2 + \epsilon_0} \right), \tag{4.B.2}$$

where $\epsilon_0 = 10^{-3} \cdot \max(\left| u_z^{(0)} \right|^2)$ is a water-level parameter for the reference wavefield spectrum, which can vanish at certain frequencies, contrary to the source time function. At this point the kernels have been computed and can be stored. We store all the frequencies between 0.5 and 5 Hz, with a step of 0.04 Hz. Storing such a kernel for one component and one parameter, computed within the blue zone in Fig. 4.1, represents around 280 MB with double precision.

8. For a given perturbation distribution $\delta m(\mathbf{x})$, numerically integrate the kernels over the domain Ω to obtain the PSD anomaly and the scattered wavefield in the frequency domain.

9. Perform inverse Fourier transform to convert the scattered wavefield into the time domain.

The validation consists of comparing the kernel-based scattered wavefields and PSD anomalies to those directly computed by numerical simulations, without relying on the Born approximation. For this validation step, no reference station is used, since we aim to verify the kernels at each individual receiver. Hence we show the values of d^{PSD} anomalies predicted by the K^{PSD} kernels, rather than using η' and $K^{\eta'}$. First, we take a single vertical point force located at $x_s = 10$ km and $z_s = 4250$ m, generating a multi-modal incident wavefield the LOH model (see Section 4.4.2). The scattered signals obtained via SEM simulations and Born approximation at the receiver located on the surface at $x_r = 50$ km (above the middle of the inclusion) are compared in Fig. 4.B.1 for negative 10% contrasts $\delta\rho'/\rho'$, $\delta V_s/V_s$ and $\delta V_p/V_p$. The agreement for all the 3 types of elastic perturbations is quite satisfactory (see correlation coefficients in Fig. 4.B.1). The few small discrepancies are due to the intrinsic limitations of the Born approximation (neglected higher orders in eqn. (4.9)).

Spectral anomalies recorded above the middle of the realistic inclusion (i.e. with gas-saturated reservoir parameters, see Section 4.3.2) are shown in Fig. 4.B.2 for both the single source used in the previous paragraph, and the multiple sources described in Section 4.3.3, acting in the LOH model. The match between SEM and Born approximation is good in both cases (see correlation coefficients shown in blue and red in Fig. 4.B.2). The unstable pattern of the anomaly obtained for a single source (blue curves) is due the presence of peaks and notches in the incident wavefield spectrum. For the average PSD anomaly (red curves), the smoothing is due to the random variations of the notch positions in the spectrum for different realisations of the source distribution.

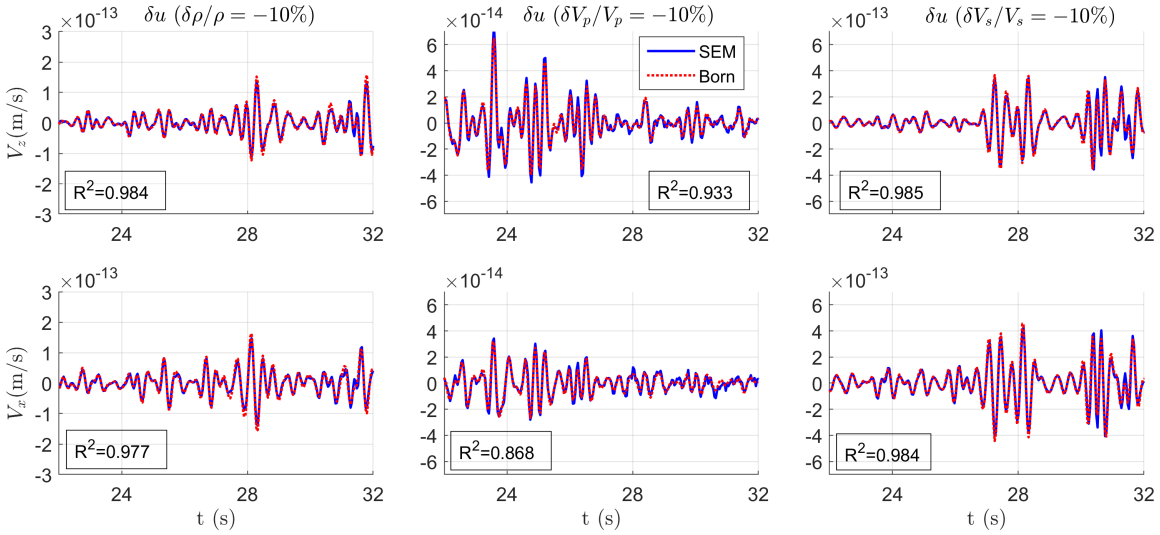


Figure 4.B.1: Time-domain comparison between scattered wavefields obtained in the LOH background medium from the waveform sensitivity kernels (dashed red lines) and from SEM (solid blue lines), at the receiver located vertically above the middle of a rectangular inclusion ($x = 50$ km). From left to right, different types of elastic contrasts are tested inside the inclusion ($\delta\rho'/\rho' = -10\%$, $\delta V_p/V_p = -10\%$ and $\delta V_s/V_s = -10\%$). Upper row: vertical component, lower row: horizontal component. Correlation coefficients are shown in the black boxes.

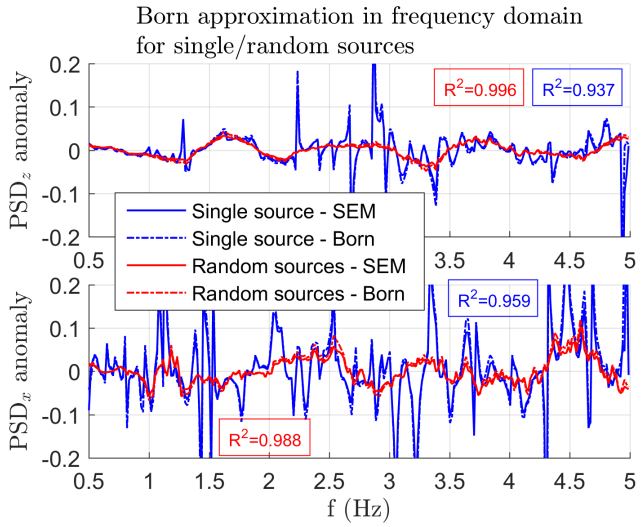


Figure 4.B.2: Relative PSD anomalies (δd^{PSD}) for the receiver at $x = 50$ km above the middle of an inclusion within the LOH background model. Anomalies are shown for an incident wavefield generated by a single source (blue lines) and by multiple random sources (red lines). They are computed both from numerical sensitivity kernels (dash-dotted lines) and SEM simulations (solid lines). Correlation coefficients between kernel predictions and SEM simulations are shown in the boxes of respective colors. The elastic perturbation values inside the inclusion are typical for full water saturation being replaced by gas saturation ($\delta V_p/V_p = -18\%$ and $\delta\rho'/\rho' = -10\%$). Top: vertical component. Bottom: horizontal component.

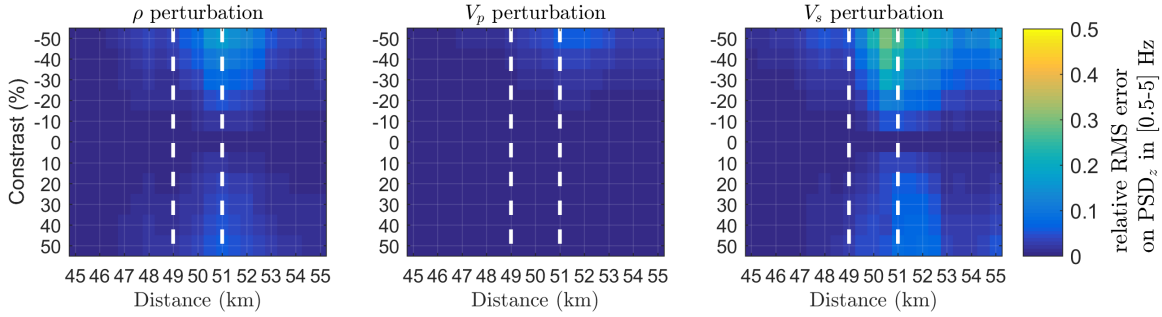


Figure 4.B.3: Approximation errors on $\langle \mathrm{PSD}_z \rangle$ in the LOH model with multiple random sources at $z = 4250$ m (at the left end of the model), as function of receiver location and perturbation strength inside the inclusion. The white dashed lines show the horizontal location of the inclusion.

The accuracy of the Born approximation in the LOH model with random multiple sources on the left-side of the model was tested for relative elastic perturbations inside the inclusion ranging from -50% to 50% with a step of 10%, for each of the three elastic parameters taken individually. For each perturbation value, the resulting average PSD was compared to the corresponding prediction of the Born approximation. The relative RMS prediction errors over the frequency range [0.5-5] Hz are shown in Fig. 4.B.3 as a function of the receiver position and the contrast value in the inclusion. As expected, the estimation error increases with the contrast. The error is stronger for negative contrasts compared to positive contrasts. Low-impedance zones generally tend to trap waves, increasing the multiple-scattering effects neglected in the Born approximation. We suggest this could explain the better accuracy for positive contrasts (stronger impedance). The error downstream the inclusion is stronger than upstream (with respect to the source distribution at the left end of the model). The errors for the elastic parameters range as follows: $\mathrm{RMS}_{V_s} > \mathrm{RMS}_{\rho} > \mathrm{RMS}_{V_p}$. This order also corresponds to the decreasing sensitivity (see Figs 4.10 and 4.11), meaning that the errors are stronger for stronger scattered wavefields. The error never exceeds 40% for the whole range of contrasts. Thus, the Born approximation appears to be robust enough to address problems involving relatively strong localised inhomogeneities at the reservoir scale.

4.C Dominant frequency trends under modal summation

The objective of this Appendix is to give a physical explanation of the dominant frequency patterns observed in Fig. 4.7d with respect to the reservoir depth. The latter contain both jumps and constant trends. Here we use the analytical mode summation theory and the Born approximation in order to show that such trends are indeed expected for a multi-modal incident wavefield. We work in our *layered flat* model (Table 4.1c). We compute the PSD perturbation $d^{\mathrm{PSD}}(\mathbf{x}_r, \omega)$ (see eqn. 4.A.6) recorded vertically above a reservoir and analyse its frequency content. As in the rest of the paper, we only consider the vertical component. For simplicity, we con-

sider only a density perturbation $\delta\rho_0 = -300$ kg/m^{-3} inside the reservoir. In order to model scattering under modal summation, we closely follow the derivations by Maupin (2001). As the latter are performed in 3D, we consider a cylindrical reservoir of radius $A = 1000$ m and thickness $h = 100$ m, and of variable depth z_0. Modal eigenfunctions, phase and group velocities are obtained using the Computer Programs for Seismology software (Herrmann, 2013).

Under the Born appoximation, $d^{\mathrm{PSD}}(\mathbf{x}_r, \omega)$ is a function of the incident and single-scattered fields $u_z^{(0)}$ and $u_z^{(1)}$ (see eqn. 4.A.7). We assume that the incident field at the receiver can be decomposed as a sum of Rayleigh wave modes propagating in the x-direction:

$$u_z^{(0)}(x_r, z_r = 0, \omega) = u_0 \sum_n \alpha_n(\omega) \exp\left(-ik_n x_r\right) , \tag{4.C.1}$$

with u_0 a constant amplitude and $\alpha_n(\omega)$ the dimensionless weight of each mode in the incident wavefield, normalised so that $\sum_n \alpha_n(\omega) = 1$ at any frequency. $\alpha_n(\omega)$ are determined by fitting a mixture of 1D gaussian models to each line (corresponding to ω_i) in Fig. 4.6c:

$$D_{\mathrm{FK}}(\omega_i, k) = \sum_n \alpha_n(\omega_i) \exp\left[-\left(\frac{k - k_n(\omega_i)}{\sigma_n}\right)^2\right] , \tag{4.C.2}$$

with D_{FK} the pixel values along the line i, and $k_n(\omega_i)$ the known theoretical wavenumber for the mode n. $\alpha_n(\omega_i)$ and σ_n are optimised using the Matlab Optimisation Toolbox. k_n is also allowed to vary slightly around its theoretical value during the optimisation. The obtained $\alpha_n(\omega_i)$ are smoothed at each frequency by taking a running average over 3 adjacent modes. For further calculations, the incident wavefield at any position (x, z) can be written in terms of modal potentials $\Phi_n^{(0)}(x, \omega) = \beta_n(\omega)\Phi_0 \exp\left(-ik_n x\right)$:

$$u_z^{(0)}(x, z, \omega) = \sum_n U_n(z)\Phi_n^{(0)}(x, \omega) , \tag{4.C.3}$$

with $U_n(z)$ the vertical Rayleigh-wave eigenfunction of the mode n and $\beta_n(\omega)$ the normalised dimensionless weights of the different modal potentials. β_n are obtained by identification of eqns (4.C.1) and (4.C.3) at $z = 0$:

$$\beta_n = \left(\sum_j \frac{\alpha_j}{U_j(z=0)}\right)^{-1} \frac{\alpha_n}{U_n(z=0)} . \tag{4.C.4}$$

α_n and β_n are different because the modal composition varies with depth through $U_n(z)$. Both $\alpha_n(\omega)$ and $\beta_n(\omega)$ are displayed in Fig. 4.C.1. Consistently with Fig. 4.6c, higher modes dominate at higher frequencies.

The single-scattered field can also be decomposed in Rayleigh-wave modes:

$$u_z^{(1)}(x_r, z = 0, \omega) = \sum_m U_m(z = 0)\Phi_m^{(1)}(x_r, \omega) . \tag{4.C.5}$$

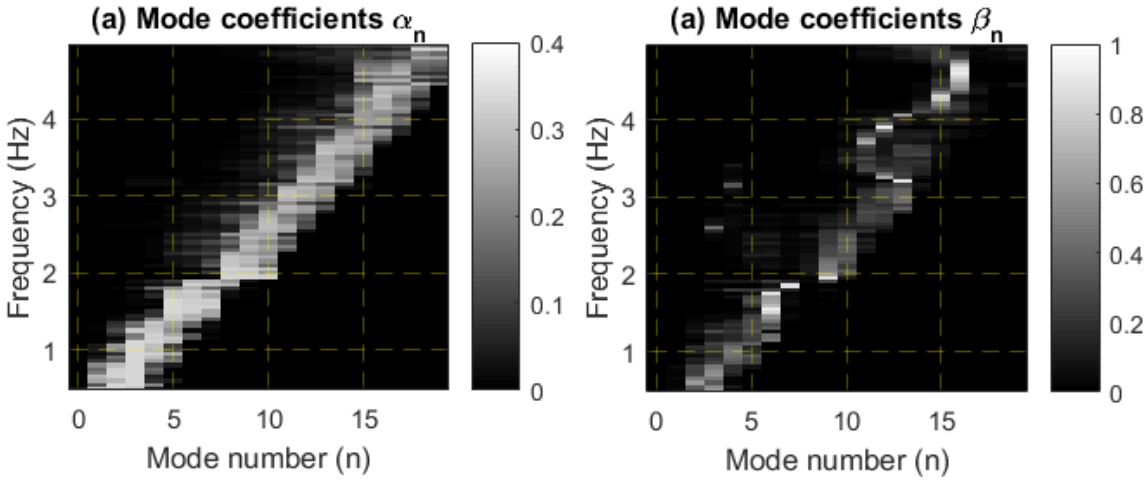


Figure 4.C.1: (a) Mode weights in the incident wavefield at the receiver. (b) Mode weights in the Rayleigh wave potential.

Using eqn. (19) from Maupin (2001), the scattered potential $\Phi_m^{(1)}$ can be expressed in a general form

$$\Phi_m^{(1)}(x_r,\omega) = -\frac{i}{8c_m u_m J_1^m}\sum_n \beta_n \Phi_0 \exp(-ik_n x_r)\frac{k_m}{k_n}\sum_{a,b}\left(\int_0^\infty C_{mn}^{(ab)}(z')dz'\right) B_a I_b(k_m,k_n,A)\ , \tag{4.C.6}$$

where c_m and u_m are respectively the phase and group velocity of the mode m, and $J_1^m = \int_0^\infty \rho(z)z^2\left[(U_m(z))^2 + (V_m(z))^2\right]dz$ is the energy integral (with V_m the radial eigenfunction). a and b are respectively row and column indices in table 1 of Maupin (2001), where the corresponding terms $C_{mn}^{(ab)}$ are reported. Analytical expressions for the integrals I_b can be found in appendix B of Maupin (2001). In our notation, I_b corresponds to the integral which arises when $C_{mn}^{(ab)}$ is non-zero for a given b. As we assume only density contrasts and incident field invariance with respect to the y-direction, the only non-zero contributions come from $C_{mn}^{(43)}(z') = \delta\rho_0\omega^2\frac{1}{k_m}V_m(z')V_n(z')$ and $C_{mn}^{(65)}(z') = \delta\rho_0\omega^2\frac{k_n}{k_m}U_m(z')U_n(z')$. The relevant integrals for $b \in \{3,5\}$ are then I_3 and I_5. In the notations of Maupin (2001), they correspond respectively to I_2 (her eqns B16 and B19) and I_0 (her eqns B15 and B18). Finally, we have $B_4 = -ik_n$ (resulting from ∂_x applied to $\exp(-ik_n x)$ in the incident potential) and $B_6 = 1$ (resulting from no differentiation applied to the incident potential). Before evaluating eqn. (4.C.6), we normalise all the eigenfunctions so that $-i/8c_m u_m J_1^m = 1$. We also assume that $C_{mn}^{(ab)}$ varies smoothly with z' at the reservoir scale, so that $\int_0^\infty C_{mn}^{(ab)}(z')dz' \approx hC_{mn}^{(ab)}(z_0)$. Once we have $\Phi_m^{(1)}(x_r,\omega)$, the PSD anomaly is obtained from eqn. (4.A.7):

$$d^{\mathrm{PSD}}(\mathbf{x}_r,\omega) = 2\cdot\mathfrak{Re}\left(\frac{u_z^{(1)}}{u_z^{(0)}}\right) = 2\cdot\mathfrak{Re}\left(\frac{\sum_m U_m(z=0)\Phi_m^{(1)}(x_r,\omega)}{\sum_n U_n(z=0)\Phi_n^{(0)}(x_r,\omega)}\right)\ . \tag{4.C.7}$$

For each tested reservoir depth, we normalise $|d^{\mathrm{PSD}}|$ by its maximum value, and display the resulting image in Fig. 4.C.2a. The observed dominant frequency pattern

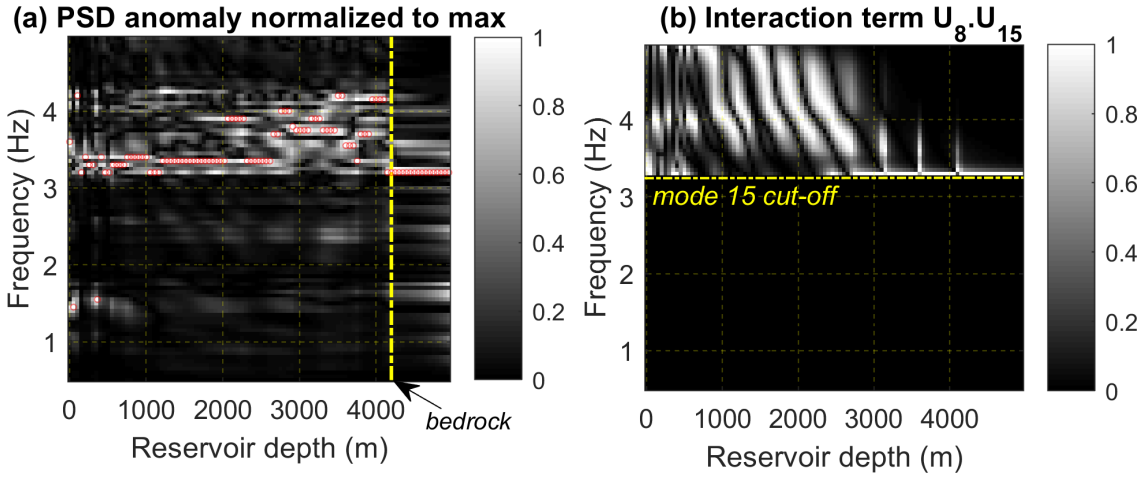


Figure 4.C.2: (a) PSD perturbation ($\left|d^{\mathrm{PSD}}\right|$), modelled vertically above the reservoir, and normalised by the maximum for each reservoir depth. Dominant frequencies are spotted with red circles. (b) Depth-frequency pattern of an individual intercation term between two modes (8 and 15), normalised by the maximum for each reservoir depth.

with respect to the reservoir depth presents jumps and constant zones, as previously observed in Fig. 4.7d for multi-modal incident wavefields, based on SEM simulations. In this Appendix, we recover this effect with modal summation. This indicates that alternating dominant modes according to frequency and depth is a relevant explanation for the observed dominant frequency patterns. The latter are not monotonously decreasing for increasing reservoir depth, as it would be the case for a pure fundamental mode. As an example, the coupling term (product of vertical eigenfunctions at reservoir depth) between modes 8 and 15 is displayed in Fig. 4.C.2b. It exhibits several discontinuous branches. The complete PSD anomaly in Fig. 4.C.2a is in fact a weighted superposition of a large amount of such branches due to different interacting pairs of modes. From eqn. (4.C.7), it can be seen the anomaly is the strongest at frequencies for which $U_n(z=0)$ is weak for the dominant modes. Physically, this means that at these frequencies an important amount of incident seismic energy propagates in depth and is not seen at the surface. A deep heterogeneity radiates a part of this «hidden» energy towards the surface, where strong anomalies are observed.

Chapter 5

Characteristics of Rayleigh wave diffraction

Mise en contexte (français)

Dans le chapitre précédent, l'approximation de Born a été introduite en tant qu'approximation linéarisée pour les faibles perturbations au sein de l'hétérogénéité. Cependant, l'étude a été menée à l'échelle du réservoir, et les mécanismes de la diffraction des ondes de Rayleigh n'étaient pas étudiés en détail. Cela constitue l'objet du présent chapitre, où l'on montre comment la polarisation elliptique caractéristique des ondes de Rayleigh (voir partie 1.3.3) influence le diagramme de radiation associé à la diffraction, en comparaison avec les ondes de volume, qui possèdent une polarisation rectiligne. En principe, la connaissance du digramme de radiation est utile pour choisir une bonne paramétrisation dans les applications d'imagerie (e.g. Virieux & Operto, 2009). Les diagrammes de radiation obtenus dans ce chapitre ne seront cependant pas directement utilisés dans le chapitre suivant, dédié à l'imagerie, car la matrice Hessienne complète contient plus d'informations pour le choix de la paramétrisation. Ce chapitre a été soumis pour publication dans *Geophysical Prospecting*:
Kazantsev, A. & Chauris, H., 2018. Radiation patterns for 2d rayleigh-wave scattering in the vertical plane: impact of the free-surface, *Geophysical Prospecting*, **submitted 29/07/2018**

Context (English)

In the previous chapter, the Born approximation was introduced as a linear approximation for small amplitude perturbations. However, the study was performed on a macro-scale, and the mechanism of Rayleigh wave scattering was not investigated in detail. This is done in this chapter, where I show how the elliptical particle motion of Rayleigh waves (see Section 1.3.3) influences the diffraction radiation patterns with respect to the case of body waves. The knowledge of the radiation patterns can be useful for choosing the right parametrisation in imaging applications (e.g. Virieux & Operto, 2009). The derived radiation patterns, however, will not be explicitely used in the imaging application presented in the next chapter, as the full Hessian matrix

contains more information for the parametrisation choice. This part was submitted to *Geophysical Prospecting* as

Kazantsev, A. & Chauris, H., 2018. Radiation patterns for 2d rayleigh-wave scattering in the vertical plane: impact of the free-surface, *Geophysical Prospecting*, **submitted 29/07/2018**

Résumé (français)

La modélisation de la diffraction nécessite de décrire les ondes incidentes et diffractées. Dans ce chapitre, on utilise une expression simple valable en champ lointain pour décrire les ondes de Rayleigh incidentes. Les approches classiques considèrent que les ondes diffractées sont également des ondes de surface, négligeant ainsi les conversions entre les ondes de surface et les ondes de volume. Dans notre cas, pour obtenir les diagrammes de radiation, on utilise l'approximation de Born sans faire cette hypothèse. Dans l'approximation de Born, une perturbation élastique du milieu agit comme une source sismique équivalente. Afin d'obtenir le diagramme de radiation associé à cette source, on utilise dans un premier temps les tenseurs de Green analytiques 2D et 3D en milieu homogène infini, qui sont relativement simples. Ceci revient à négliger l'effet de la surface libre sur les ondes diffractées. Dans un deuxième temps, on utilise des tenseurs de Green numériques pour calculer les diagrammes de radiation complets tenant compte de la surface libre. Ceci permet d'identifier les propriétés des diagrammes de radiation liées à la polarisation elliptique des ondes de Rayleigh et celles liées aux effets de la surface libre. En particulier, on s'aperçoit que la polarisation elliptique est responsable d'une asymétrie entre la diffraction vers l'arrière et la diffraction vers l'avant. D'autre part, sur la composante verticale, la surface libre a pour effet une simple amplification des ondes diffractées pour des angles d'incidence proches de la verticale à $\pm 60°$. L'effet de la surface libre sur la composante horizontale est plus complexe.

Summary (English)

This work investigates the radiation patterns for Rayleigh wave scattering in the vertical xz plane (P-SV system). Those patterns are of practical interest for diffraction-based imaging using Rayleigh waves. We use a simple far-field expression is to describe the incident Rayleigh waves. Classical approaches treat the diffracted waves as surface waves, neglecting the mode conversions from surface to body waves. Instead, we use the Born approximation without this assumption. Under the Born approximation, an elastic perturbation of the medium acts as an equivalent seismic source. We first use the relatively simple 2D and 3D Green's tensors in unbounded homogeneous medium to construct radiation patterns for such an equivalent source, neglecting the effect of the free-surface on the scattered waves. Then, we numerically compute the full radiation patterns accounting for the free-surface in 2D, using the spectral element method. This allows to distinguish between the pattern features which are related to the Rayleigh wave incident motion from those due to the free-surface effects on the scattered wavefield. Namely, we find that the rotating Rayleigh wave motion generates asymmetry between forward- and back- scattering. On the other hand, the free-surface magnifies the unbounded medium radiation pattern on the vertical component for steep scattering angles (within $\pm 60°$). The free-surface contribution has a more complicated shape on the horizontal component and for sub-horizontal angles.

5.1 Introduction

Surface wave scattering can be used by small- or intermediate-scale receiver arrays located directly above the investigated region. These experiments usually aim to detect either strong contrasts such as voids and buried objects (Gucunski et al., 1996; Nasseri-Moghaddam et al., 2007; Tallavó et al., 2009), or weaker contrasts such as fluid reservoirs (Saenger et al., 2009). Different types of modelling can be relevant according to the target. Since voids tend to trap the scattered waves, the single-scattering approximation (e.g. Born approximation, Snieder, 1986) is likely to fail. Instead, Yu & Dravinski (2009) used the boundary element method (BEM) for solving the scattering of P, SV and Rayleigh waves by an embedded 2D cylindrical cavity. In our work, we address the case of scattering by a weak contrast, for which the Born approximation is relevant. The latter requires to model incident and scattered wavefields with an appropriate Green's function.

We make the assumption that the incident wavefield is composed of Rayleigh waves. Surface waves arise due to the presence of a free-surface at the top of an elastic half-space. In flat 3D layered geometries, they are usually studied in the horizontal plane, where they approximately obey the 2D Helmholtz equation for each individual mode (see Wielandt (1993) for discussion). The vertical dependence of the motion can be obtained as a solution of an eigenvalue problem, for example by using the propagator matrix method (Thomson, 1950; Haskell, 1953). As a result, surface waves excited by a point force or a moment tensor can be represented as cylindrical wave-fronts with uncoupled vertical and horizontal dependence:

$$u_i(r,z,t) = \sum_n A_i^{(n)}(z,z_0) J(k_n r) \exp(-i\omega t) \ , \tag{5.1}$$

with u_i the displacement on the i-th component, t the time, ω the angular frequency, r the radial distance to the source, (z, z_0) respectively the receiver and the source depth, $A_i^{(n)}(z,z_0)$ the depth-dependent amplitude of the n-th mode, k_n the horizontal wavenumber of the n-th mode, and J some combination of Bessel functions. This type of representation is referred to as mode summation. Explicit expressions for different terms are summarised in Aki & Richards (2002, chapter 7), for both Rayleigh and Love waves.

When addressing the scattered wavefield, one usually stays into the horizontal plane, and considers each mode is scattered either on itself or on other surface wave modes (mode conversion) (Snieder, 1986; Friederich et al., 1993; Maupin, 2017). That is, expressions of type (5.1) are used for describing both incident and scattered wavefields. In theory, however, the scattered wavefield should be described by the full Green's function of the medium containing all the wave types, and not only the part related to the surface waves. As discussed by Maupin (1996), this is not the case of the expression (5.1), which neglects the radiating modes (non-trapped body waves). The surface waves indeed dominate in the far-field because of their lower geometrical spreading, but this is not necessarily the case right above the scatterer. Our work is motivated by this lack of the body wave contributions in standard methods.

We focus on the near-field scattering (i.e. immediately around and above the scatterer), and we use the full elastodynamic Green's tensor containing all the wave types. Since the latter is easier to obtain in 2D, we move to the vertical (xz) plane including the scatterer, instead of observing the process from above (xy plane). In this configuration, we derive the radiation patterns as a tool to provide an intuitive understanding of the Rayleigh wave scattering. In our analysis, we combine both analytic and numerical methods. First, we recall the type of equivalent source terms (point force or moment tensor) which arise in the Born approximation for density contrasts $\delta\rho$ and Lamé coefficient contrasts $\delta\lambda$, $\delta\mu$. Then we evaluate these terms based on simple far-field expressions for *incident* plane P, SV and Rayleigh waves, and verify these predictions by comparison with 2D numerical simulations (spectral element method, or SEM, Komatitsch et al., 1999). In a second step, we couple these terms to an unbounded medium Green's tensor, which allows to describe the Rayleigh wave scattering at early times, before the interaction with the free-surface. Typically, this kind of analysis would be impossible with a purely numerical approach, since incident Rayleigh waves would not exist in an unbounded medium. We compare the obtained patterns to their body wave equivalents. We also discuss differences between 2D and 3D Green's tensors, as well between far- and near-field behaviour. Finally, we use SEM to get a numerical 2D Green's tensor which contains all the wave types that can arise in presence of a free-surface. The contribution of the free-surface is analysed both in terms of scattered energy and pattern shape.

5.2 Born approximation, sensitivity kernels and radiation patterns

Our purpose is to describe the behaviour of the Rayleigh waves scattered by the elastic perturbations of the medium. The most simple approach is to linearise the wave equation and to consider that the scattered wavefield is a first order effect with respect to the reference wavefield. This is commonly referred to as Born approximation:

$$\mathbf{u} = \mathbf{u}^{(0)} + \mathbf{u}^{(1)} + o(\mathbf{u}^{(1)}), \tag{5.2}$$

where $\mathbf{u}$ is the full displacement wavefield, $\mathbf{u}^{(0)}$ the unperturbed (incident) wavefield, and $\mathbf{u}^{(1)}$ the leading (first order) term of the scattering series, corresponding to single scattering. In order to get a simple expression for $\mathbf{u}^{(1)}$, the elastic wave equation is first written in its general form:

$$\mathcal{L}\mathbf{u} = \mathbf{f}, \tag{5.3}$$

with the $\mathcal{L}$ the elastic operator and $\mathbf{f}$ the source term (point force). A small medium perturbation δm, related to either the density ρ or the two Lamé coefficients λ and μ, causes a perturbation $\delta\mathcal{L}$ of the operator. Then the following two equations must be satisfied by the reference and the single-scattered wavefields:

$$\mathcal{L}^{(0)}\mathbf{u}^{(0)} = \mathbf{f}\ , \tag{5.4}$$

$$\mathcal{L}^{(0)}\mathbf{u}^{(1)} = -\delta\mathcal{L}\mathbf{u}^{(0)}\ . \tag{5.5}$$

The single-scattered wavefield is thus a solution of the unperturbed wave equation with a modified source term ($-\delta\mathcal{L}\mathbf{u}^{(0)}$). Physically, this source is located at the

position of a scatterer, where $-\delta\mathcal{L}$ is non-zero because of the medium perturbation. It can be shown (e.g. Dahlen et al., 2000) that the component i of the scattered wavefield recorded at the receiver location $\mathbf{x}_r$ is given by an integral over the contributions of all the scatterers distributed within the investigated space Ω:

$$u_i^{(1)}(\mathbf{x}_r,\omega) = \sum_{m\in\rho,\lambda,\mu} \int_\Omega K_i^{(m)}(\mathbf{x}_r,\mathbf{x},\omega)\frac{\delta m}{m_0}(\mathbf{x})d\mathbf{x} \ . \tag{5.6}$$

The integrands K, of dimension [$\mathrm{m}^{-1}.\mathrm{Hz}^{-1}$] in 2D and [$\mathrm{m}^{-2}.\mathrm{Hz}^{-1}$] in 3D, are called the waveform sensitivity kernels for *relative* medium perturbations $\delta m/m_0$. Their derivation requires the knowledge of the reference medium's Green's tensor. This tensor yields the solution of the wave equation (5.4) for an impulsive source term. That is, if the source term on the component i can be written as $F_i(\omega) = f_i\delta(\mathbf{x} - \mathbf{x}_s)$, with δ the Dirac distribution and f_i a constant force spectrum, the Green's tensor $\mathbf{G}(\mathbf{x}_r,\mathbf{x}_s,\omega)$ for the source-receiver couple (r,s), which we write as $\mathbf{G}^{\mathrm{rs}}$ in the following, satisfies $\mathcal{L}_{ij}G_{jk}f_k = f_i$, and $u_j = G_{jk}f_k$ is the solution wavefield. The elastic operator is given by

$$\mathcal{L}_{ij}(\rho,\lambda,\mu) = -\rho\omega^2\delta_{ij} - \partial_k(c_{ikjl}\partial_l) \text{ [summation over } k,l] \ , \tag{5.7}$$

with subscripts referring to the displacement components, and c_{ikjl} being the stiffness tensor, which takes the simple form $c_{ikjl} = \lambda\delta_{ik}\delta_{jl} + \mu\left(\delta_{ij}\delta_{kl} + \delta_{il}\delta_{kj}\right)$ in isotropic media. In this case, the sensitivity kernels are given by

$$K_i^{(\rho)}(\mathbf{x}_r,\mathbf{x},\omega) = \rho_0(\mathbf{x})\omega^2 G_{ik}^{\mathrm{rx}}u_k^{(0)} \tag{5.8a}$$

$$K_i^{(\lambda)}(\mathbf{x}_r,\mathbf{x},\omega) = -\lambda_0(\mathbf{x})(\partial_k G_{ik}^{\mathrm{rx}})(\partial_l u_l^{(0)}) \tag{5.8b}$$

$$K_i^{(\mu)}(\mathbf{x}_r,\mathbf{x},\omega) = -\mu_0(\mathbf{x})\partial_l G_{ki}^{\mathrm{rx}}\left[\partial_k u_l^{(0)} + \partial_l u_k^{(0)}\right] \ , \tag{5.8c}$$

with $\mathbf{u^{(0)}}(\mathbf{x})$ the incident wavefield, and all the wavefield derivatives evaluated at the scatterer location $\mathbf{x}$. The detailed derivation of these expressions, which implies integration by parts with a careful treatment of surface integrals, can be found in Dahlen et al. (2000). One important property of the Green's tensor is that it also yields the solution for an arbitrary source distribution $\mathbf{f}(\mathbf{x}_s,\omega)$ as $u_i(\mathbf{x}_r,\omega) = \int_\Omega G_{ij}^{\mathrm{rs}}(\omega)f_j(\mathbf{x}_s,\omega)d\mathbf{x}_s$. Another important feature is the Green's tensor reciprocity, $G_{ij}^{\mathrm{rx}} = G_{ji}^{\mathrm{xr}}$, which is crucial for numerical estimations of the sensitivity kernels via the adjoint-state method (Tromp et al., 2005). In this paper, we compare both numerical and analytic implementations.

In the remainder of this section, we use 2D integrals and surface elements in our notations, though the extension to 3D is straightforward. For example, let us consider a medium perturbation localised in $\mathbf{x_0}$ (i.e. $\delta m(\mathbf{x}) = \epsilon m_0\delta(\mathbf{x} - \mathbf{x_0})$). Then the equation (5.6) yields the scattered wavefield as

$$u_i^{(1)}(\mathbf{x}_r,\omega) = \epsilon\delta S K_i^{(m)}(\mathbf{x}_r,\mathbf{x_0},\omega) \ , \tag{5.9}$$

with δS an infinitesimal surface containing $\mathbf{x_0}$. We are interested in the strength and spatial distribution of $u^{(1)}$ when the incident wavefield consists of Rayleigh waves.

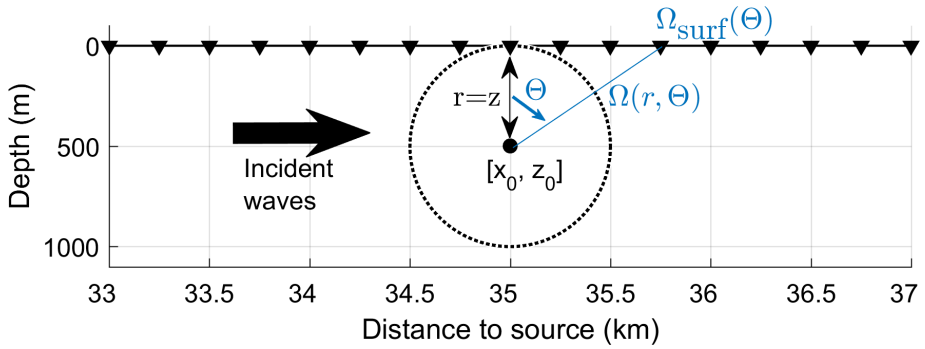


Figure 5.1: Geometry of the scattering problem. The incident wavefield arrives from the left. The scattered waves emitted at $[x_0, z_0]$ can be either studied along the dotted circle of radius r (radiation pattern $\Omega(r, \Theta)$, suitable for unbounded media), or at the surface ($\Omega_{\text{surf}}(\Theta)$, suitable for a half-space). In the latter case, for a given scattering angle Θ, the distance to the scatterer is constrained by the depth z_0. Black triangles mimic a receiver array at surface.

To quantify those, we choose to work with the energy spectral density (ESD) of the displacement radiated by the scatterer and received at $\mathbf{x}_r$, which is proportional to

$$\left|u_i^{(1)}(\mathbf{x}_r, \omega)\right|^2 = (\epsilon\delta S)^2 \left|K_i^{(m)}(\mathbf{x}_r, \mathbf{x_0}, \omega)\right|^2 \ . \tag{5.10}$$

Though the term «energy» is somewhat misplaced here since we work with displacement instead of velocity, our ESD is related to the true energy via a factor ω^2. We define the radiation pattern at the scattering location $\mathbf{x_0}$ as

$$\Omega^{(m,i)}(r, \Theta, \omega) = \left|K_i^{(m)}(r, \Theta, \omega)\right|^2 = \left|K_i^{(m)}(\mathbf{x}_r, \mathbf{x_0}, \omega)\right|^2 \ , \tag{5.11}$$

where the origin has been moved to the scattering point $\mathbf{x_0}$, and (r, Θ) are the polar coordinates of the receiver, Θ being measured clockwise from the vertical. Θ is linked to the canonical polar angle θ as $\Theta = \pi/2 - \theta$. In a realistic situation of a receiver line located at the surface and a scatterer buried at a depth z_0, the radiation pattern is modified to

$$\Omega_{\text{surf}}^{(m,i)}(\Theta, \omega) = \Omega^{(m)}(|z_0/\cos(\Theta)|, \Theta, \omega) \ . \tag{5.12}$$

The geometry of the problem is illustrated in Figure 5.1.

Here we can draw some parallel between our work and the paper by He & Plessix (2017), who studied the effect of anisotropic medium perturbations on body waves in a perspective of multi-parameter full waveform inversion (FWI). They choose a radiation pattern definition which differs from ours in three main points. First, our radiation pattern contains both finite-frequency effects (phase term) and smooth amplitude variations, while He & Plessix (2017) used an asymptotic high-frequency approximation (see their equation (1)). Second, in our notation their definition would be equivalent to $\Omega^{(m)} = K^{(m)}$, provided that amplitude and phase terms were previously discarded from $K^{(m)}$. This quantity represents the raw amplitude of the displacement perturbation radiated by the scatterer. It can be negative, or even complex if the incident displacement components are out of phase, as it is the case for Rayleigh waves. Our modified definition leads to a real positive value, which removes ambiguity from

the polar plots. Finally, we represent the pattern with respect to the angle between the vertical and the receiver-scatterer line. For surface waves, which always propagate horizontally, $\Theta = 0°$ corresponds to a direction perpendicular to the incident wave propagation. This is different from the radiation patterns for body waves, which are usually represented for a source located exactly above the scatterer position, and thus $\Theta = 0°$ is the incident wave propagation direction.

We also define a measure of the total ESD radiated by the scatterer on the component i as an integral over all the scattering angles perceivable from the surface, per square relative elastic perturbation and per square surface of the medium perturbation:

$$E_i^{(1)}(\omega) = \int_{-\pi/2}^{\pi/2} \Omega_{\text{surf}}^{(m,i)}(\Theta, \omega) d\Theta \ . \tag{5.13}$$

If all the scattered waves were propagating as circular wave fronts from $\mathbf{x_0}$ (i.e. in the absence of the free-surface), $E^{(1)}$ would be proportional to an energy flux across an upper half-circle surrounding the scatterer. In order to get an idea of the importance of $E_i^{(1)}$ compared to the incident wavefield, we normalise it by the mean unperturbed ESD recorded at the surface:

$$E_i^{(0)}(\omega) = \frac{1}{x_{\max} - x_{\min}} \int_{x_{\min}}^{x_{\max}} \left| u_i^{(0)}(x_r, \omega) \right|^2 dx_r \ . \tag{5.14}$$

5.3 Equivalent source terms

It is formally apparent from equations (5.8) that the single-scattered wavefield can be seen as a response to a point force applied in $\mathbf{x_0}$ for a ρ perturbation, an explosive source for a λ perturbation, and a moment tensor $M_{kl} = \epsilon\delta S\mu^{(0)} \left[\partial_k u_l^{(0)} + \partial_l u_k^{(0)}\right]$ for a μ perturbation. This is important, since the different elastic parameters must have different signatures for a multi-parameter inversion to be feasible. In this section, we first perform a simple numerical simulation to achieve a visual illustration of the different source types, and then derive analytic expressions for the source terms assuming horizontally incident Rayleigh waves. We compare them to the corresponding source terms for incident SV and P waves, which allows an intuitive understanding of the Rayleigh-wave scattering characteristics. Note the virtual source term depends only on the incident motion, while it is independent from the free-surface influence on the scattered wavefield.

5.3.1 Numerical illustration

We performed numerical simulations (SPECFEM2D, Komatitsch et al., 1999) into an homogeneous elastic 2D half-space with a free-surface. The source (vertical point force) is located at the origin, and the model contains a circular inclusion of centre $[x_0, z_0] = [35 \cdot 10^3, 500]$ m and radius $R = 100$ m. The half-space parameters are $\alpha^{(0)} = 4000$ m/s (compression wave velocity), $\beta^{(0)} = 2000$ m/s (shear wave velocity) and $\rho^{(0)} = 2300$ kg/m^3, yielding the Lamé coefficients $\lambda^{(0)} = \rho(\alpha^2 - 2\beta^2) = 1.84 \cdot 10^{10}$ Pa and $\mu^{(0)} = \rho\beta^2 = 9.2 \cdot 10^9$ Pa. The perturbations $\Delta\lambda = -0.2\lambda^{(0)}$, $\Delta\mu = -0.2\mu^{(0)}$ and $\Delta\rho = -0.2\rho^{(0)}$ are successively applied inside the inclusion. We obtained the scattered

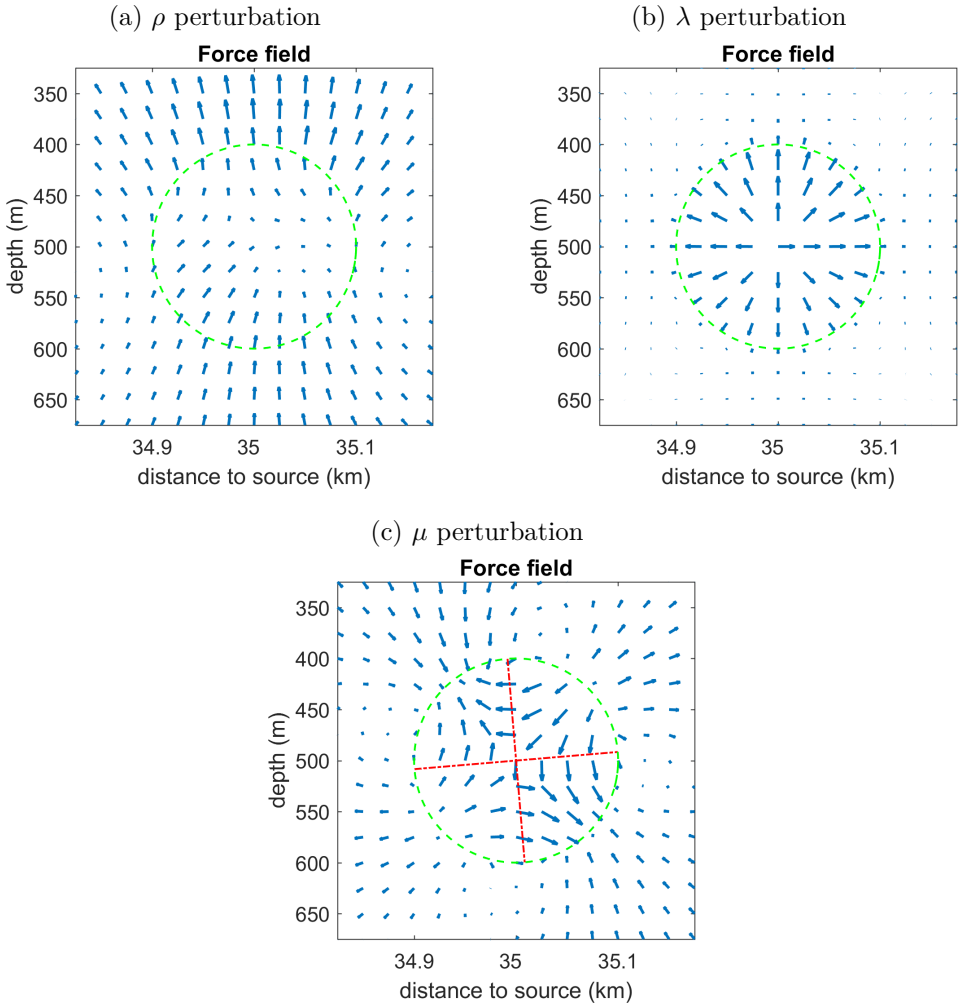


Figure 5.2: (a)-(c) Equivalent forces for the scattered wavefield. Each arrow represents a force acting on circular shell centred on the inclusion centre containing contrasts of different types (ρ, λ, μ). The inclusion is shown with the green dashed line. In (c), the principal axes of the equivalent moment tensor are shown with red dash-dotted lines.

wavefields by subtracting the reference wavefield simulated in the homogeneous half-space without inclusion from the full simulation containing the perturbation:

$$\mathbf{u}^{(\mathrm{sc})}(\mathbf{x},t) = \mathbf{u}^{(\mathrm{full})}(\mathbf{x},t) - \mathbf{u}^{(0)}(\mathbf{x},t) \tag{5.15}$$

We then calculated the equivalent differential stress field at each time step using the isotropic stress-strain relation

$$\tau_{ij} = \lambda\delta_{ij}\epsilon_{kk} + 2\mu\epsilon_{ij} \ , \tag{5.16}$$

with $\epsilon_{ij} = \dfrac{1}{2}\left(\dfrac{\partial u_i^{(\mathrm{sc})}}{\partial x_j} + \dfrac{\partial u_j^{(\mathrm{sc})}}{\partial x_i}\right)$ the strain tensor. At each grid node, we computed the force acting on the circle of centre $[x_0, z_0]$ and containing the node as

$$f_i(\mathbf{x}) = \tau_{ij}(\mathbf{x})n_j(\mathbf{x}) \ , \tag{5.17}$$

where $n_j(\mathbf{x}) = [\cos(\theta), \sin(\theta)]$ is the normal to the circle at the node location $\mathbf{x}$, with θ the polar angle with respect to $[x_0, z_0]$. When the circle corresponds to the inclusion, the associated force represents the action of the scatterer on the surrounding medium. Otherwise, the forces are interior (either to the scatterer or to the surrounding medium). Those forces are represented in Figure 5.2 for a time step corresponding to the Rayleigh wave propagation across the inclusion. In the case of a density contrast (Figure 5.2a), the net force integrated over the inclusion surface is non-zero, so the inclusion tends to act as a point force applied at its centre (pointing upwards for the displayed snapshot). On the other hand, the net force tends to vanish for both λ and μ (Figs 5.2b-c) contrasts, which means that the equivalent source is a moment tensor, consistently with the Born approximation formula (5.8). The moment tensor is clearly isotropic for the λ contrast (diverging arrows from the centre). For the μ contrast, we have to perform an eigendecomposition to identify its isotropic and double-couple components:

$$\begin{aligned}
\mathbf{M}^{(\mu)} &= \epsilon\delta S\mu^{(0)} \begin{pmatrix} 2\partial_x u_x^{(0)} & \partial_x u_z^{(0)} + \partial_z u_x^{(0)} \\ \partial_x u_z^{(0)} + \partial_z u_x^{(0)} & 2\partial_z u_z^{(0)} \end{pmatrix} \\
&= \epsilon\delta S\mu^{(0)} \mathbf{P}^\dagger \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \mathbf{P} \\
&= \epsilon\delta S\mu^{(0)} \mathbf{P}^\dagger \left[M_{\mathrm{iso}}^{(\mu)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + M_{\mathrm{dc}}^{(\mu)} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right] \mathbf{P} \ ,
\end{aligned} \tag{5.18}$$

with $\mathbf{P}$ the transformation matrix containing the principal axes of the moment tensor on its columns, and the isotropic and double-couple components being given by

$$\begin{aligned}
M_{\mathrm{iso}}^{(\mu)} &= \frac{1}{2}(\lambda_1 + \lambda_2) \\
M_{\mathrm{dc}}^{(\mu)} &= \frac{1}{2}(\lambda_1 - \lambda_2) \ .
\end{aligned} \tag{5.19}$$

The principal axes obtained by numerical diagonalisation of $\mathbf{M}^{(\mu)}$ are represented with red dashed lines in Figure 5.2c. As expected, they separate compression zones from dilatation zones. Thus, from this simple numerical example, we can see that

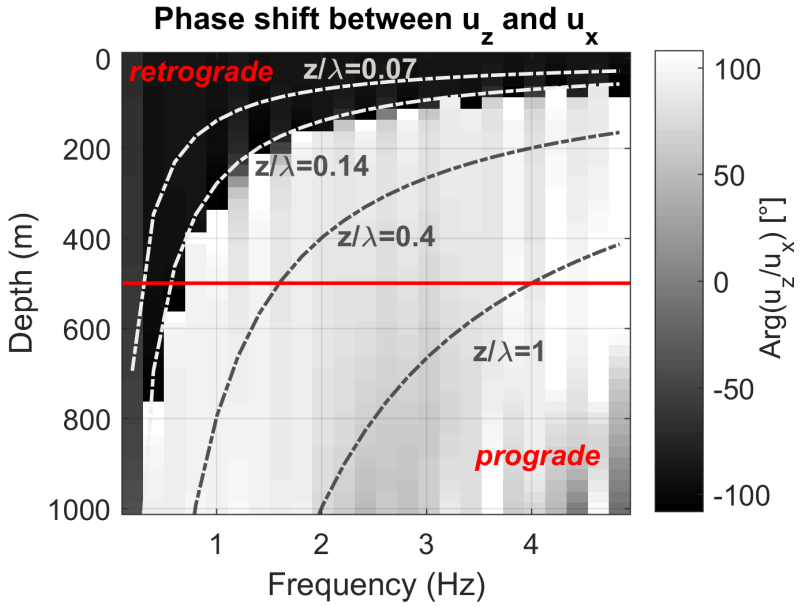


Figure 5.3: Phase shift between the simulated vertical and horizontal displacement as function of frequency and depth. Two regimes of elliptic motion (prograde and retrograde) are separated by the iso-line $z/\lambda = 0.14$. Other iso-lines are shown with dash-dotted lines. The red solid line at $z = 500$ m shows the physical depth at which the analysis is performed in the remainder of the paper.

the scattered wavefield is indeed generated by force distributions associated with some specific types of seismic sources, according to the contrast type. In the next paragraphs we analytically investigate the characteristics of the equivalent source terms for Rayleigh waves, and compare them to the case of horizontally incident P- and SV-waves.

5.3.2 Rayleigh waves

In the case of an incident wavefield entirely consisting of plane Rayleigh waves, we can write

$$\begin{cases} u_x^{(0)}(x,z,\omega) = iA_x(z,\omega)\exp(ikx) \\ u_z^{(0)}(x,z,\omega) = A_z(z,\omega)\exp(ikx)\ , \end{cases} \tag{5.20}$$

with $k = \omega/c_R$ the wavenumber of the horizontally propagating Rayleigh waves and $A_x(z,\omega)$, $A_z(z,\omega)$ the depth-dependant amplitudes on each component, proportional to the fundamental mode eigenfunctions U_x, U_z. The particle motion described by (5.20) is of rotating type because of the $\pi/2$ phase shift between components. In the shallowest part of the half-space, both amplitude terms are positive and the motion is retrograde. At some depth, A_x becomes negative and the motion becomes prograde. This can be well observed in Figure 5.3, where the phase shift between vertical and horizontal components is plotted for different depths and frequencies, at the horizontal position corresponding to the centre of the inclusion in Figure 5.2. The switch from -90° to +90° corresponds to the sign change of A_x, and always occurs for $z/\lambda \approx 0.14$.

In Appendix 5.A.1, we derive the expressions of the equivalent source terms for different types of elastic contrasts placed into the Rayleigh wavefield. We show that a ρ-contrast acts as a rotating point force (equation (5.A.1)), a λ-contrast as an isotropic moment tensor (equation (5.A.4)), and a μ-contrast as a moment tensor with both isotropic and rotating double-couple components (equations (5.A.9) and (5.A.10)). In Figure 5.4 we check these predictions by numerically performing the diagonalisation in equation (5.18), at each time step within the Rayleigh wave arrival (Figure 5.4a), at a depth of 500 m, and after filtering the signals around 0.3 Hz and 1 Hz. This corresponds to the dimensionless depth $z/\lambda = 0.07$ and $z/\lambda = 0.25$, respectively. The moment tensor component relative proportions $w_{\mathrm{dc,iso}} = \dfrac{|M_{\mathrm{dc,iso}}|}{|M_{\mathrm{dc}}| + |M_{\mathrm{iso}}|}$ are displayed in Figure 5.4b and c, for both frequencies. They periodically oscillate between a pure double-couple and a mix of a double couple with an isotropic component, with two cycles per Rayleigh wave cycle. The maximum reached isotropic component appears higher at low frequency. We display the displacement fields filtered around $f = 1$ Hz at two time steps corresponding to the minimum (Figure 5.5a) and maximum (Figure 5.5b) isotropic component (spotted by the green dotted lines in Figure 5.4c). The maximum isotropic component corresponds to a displacement field with non-zero divergence at the red star location(Figure 5.5b), while the pure double couple corresponds to a purely rotational incident wave motion (Figure 5.5a).

On the other hand, the principal axes of the moment tensor rotate in the sense of the incident Rayleigh wave motion, i.e. retrograde (counter-clockwise, Figure 5.4d) or prograde (clockwise, Figure 5.4e). The rotation speed is not constant, as it tends to accelerate when the axes get close to vertical or horizontal. This moment also corresponds to the maximum isotropic component (compare Figure 5.4b to d, and c to e).

The match between numerical and analytic computations is acceptable during the main arrival (between 17 and 21 s), while some deviations can be observed on both edges of the plot. They are due to the numerical wavefield being no more necessarily dominated by Rayleigh waves. Some contamination by S-waves can also take place within the main arrival, since they propagate at a velocity close to Rayleigh waves ($V_{\mathrm{Rayleigh}} \approx 0.95\beta$ in a homogeneous half-space), but with stronger geometrical spreading. Also, while the analytic result is estimated for one precise frequency, the numerical decomposition is based on imperfect band-pass filtering, which may also introduce some artefacts.

In the next two paragraphs we repeat the derivations for horizontally incident SV and P waves, in order to highlight the differences of the equivalent source terms.

5.3.3 SV waves

The derivations of Appendix 5.A.1 are repeated in Appendix 5.A.2 for horizontally propagating SV waves. A λ-contrast does not scatter SV waves because their velocity is insensitive to λ. On the other hand, because the particle motion associated with SV wave is not rotative, a ρ-contrast now acts as a vertical point force (equation (5.A.12)). Finally, we show (equations (5.A.14) and (5.A.15)) that a μ-contrast now behaves as

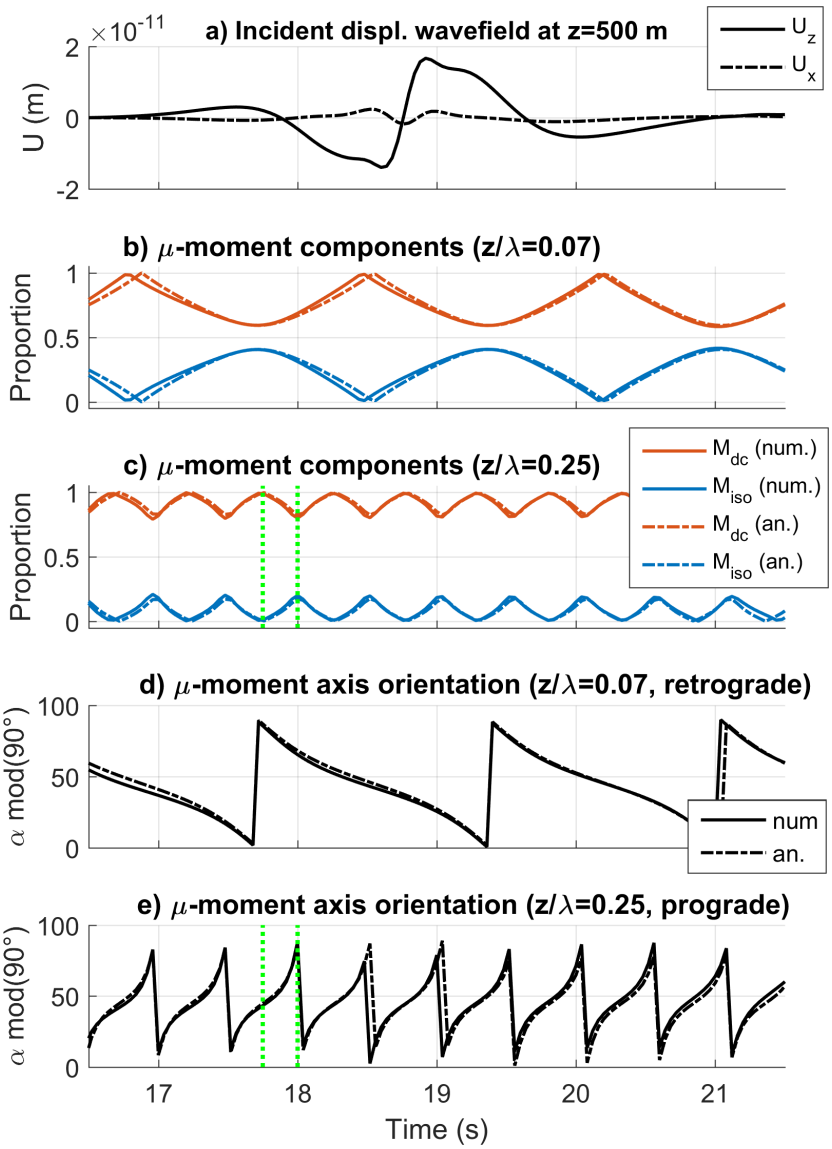


Figure 5.4: Equivalent moment tensor characteristics for a μ-contrast with incident Rayleigh waves. (a) Numerically simulated incident displacement wavefield at the scatterer location, on vertical (solid line) and horizontal (dashed line) components. (b) Isotropic (blue) and double-couple (red) moment tensor components obtained at f=0.3 Hz, corresponding to $z/\lambda = 0.07$. Results are compared between the numerical eigendecomposition (solid lines) and the analytic equation (5.A.9) (dashed lines). (c) Same as (b) for f=1 Hz ($z/\lambda = 0.25$). (d) Orientation of one of the moment tensor axes within the first quadrant (between 0° and 90°), compared between the numerical eigendecomposition (solid lines) and analytic equation (5.A.10) (dashed lines), obtained for $z/\lambda = 0.07$. (e) Same as (d) for $z/\lambda = 0.25$. The green dotted lines show the time step at which the displacement field is displayed in Figure 5.5. Note the change in the sense of the axis rotation between the retrograde (d) and prograde (e) regimes.

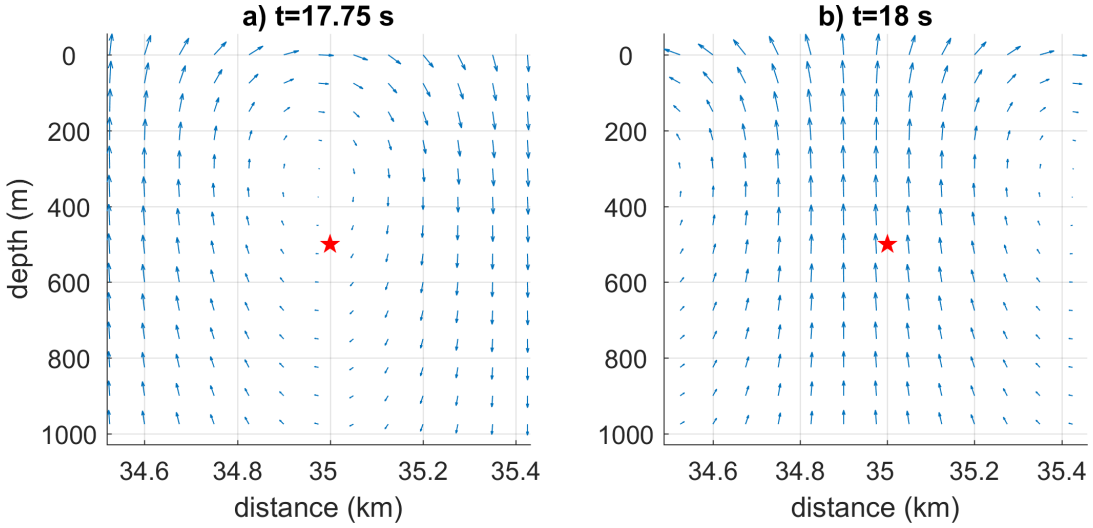


Figure 5.5: Rayleigh wave displacement fields (numerical) filtered around f=1 Hz, plotted at times 17.75 s and 18 s, corresponding respectively to minimum and maximum isotropic component in Figure 5.4c (green dotted lines). The red star represents the location where the moment tensor calculations were performed for Figure 5.4.

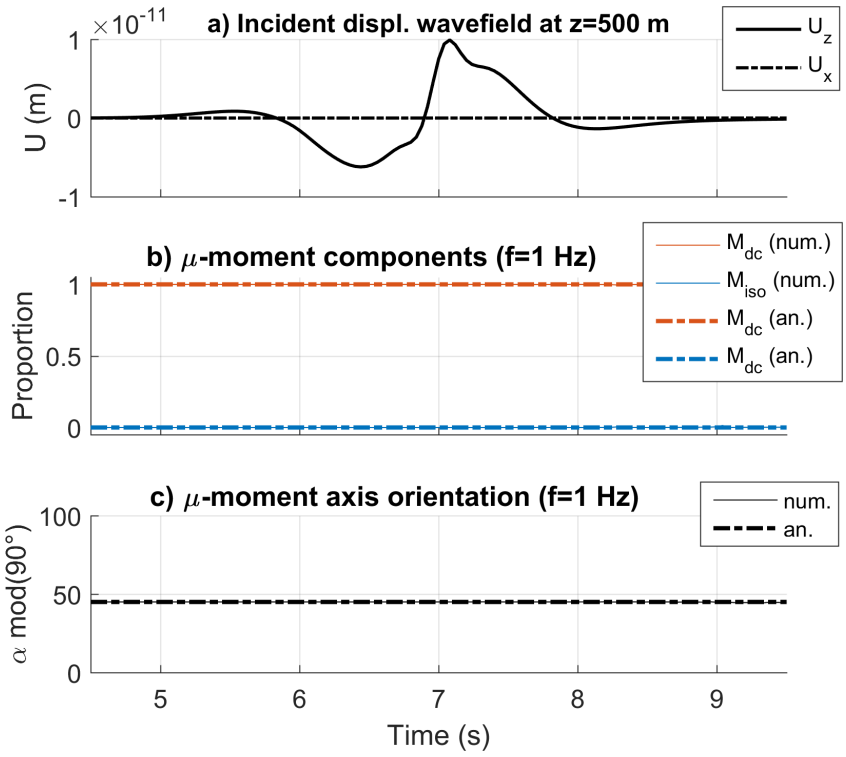


Figure 5.6: Same as Figure 5.4 with horizontally incident SV waves, at f=1 Hz only.

a pure double-couple moment tensor whose axes have a constant orientation

$$\alpha_{\pm} = \pm 45^{\circ} \ . \tag{5.21}$$

Again, we numerically check these predictions in Figure 5.6. In order to generate an approximately plane SV wave in the simulation, we removed the free-surface from the top of the model, and placed a vertical point force source at the same depth as the studied location ($z_s = 500$ m). In the previous example with the Rayleigh wave, we used a very distant source to make sure the body waves were almost completely attenuated at the receiver by geometrical spreading. Here, this issue disappeared and we moved the source to $x_s = 32.5 \cdot 10^3$ m in order to reduce the simulation time. The notion of depth-frequency scaling does not exist for SV waves, so we just display the results for $f = 1$ Hz. The numerical calculation follows the constant values predicted analytically.

5.3.4 P waves

The derivations of Appendix 5.A.1 are also repeated for horizontally propagating P waves in Appendix 5.A.3. A ρ-contrast now acts as a horizontal point force (equation (5.A.18)), because the incident wavefield is now polarized horizontally. As for Rayleigh waves, a λ-contrast acts as an isotropic moment tensor (equation (5.A.20)). This behaviour, predicted by equation (5.8b), is general for λ-contrasts placed into arbitrary wavefields with non-zero displacement divergence. A μ-contrast is equivalent a linear dipole along the x-axis (equations (5.A.22) and (5.A.23)). Because these conclusions are relatively evident directly from the considered P-wavefield, we skip the numerical validation for brevity.

In conclusion, we see that the major characteristic of Rayleigh waves compared to body waves is that they generate *rotating* virtual forces and double couple moments for ρ- and μ-contrasts, respectively. λ-contrasts behave in the same way (isotropic moment tensor) for any type of incident waves. The magnitude of the moment, however, depends on the type of the wave and determines how sensitive to λ-contrasts it is. P waves are quite sensitive to λ, because the motion is polarised in the same direction as the wave propagation, which enhances the wavefield divergence. SV waves are totally insensitive to λ because polarised perpendicular to the propagation (zero displacement divergence). Spatial snapshots of the Rayleigh wave motion exhibit alternating rotation and compression/dilatation (see Figure 5.5), so they *are* sensitive to λ. However, as we show in the last section, this sensitivity is very low compared to ρ and μ. In the next section, we analyse the radiation patterns produced by the equivalent source terms. We start with deriving analytic expressions assuming these sources are placed in an unbounded plane or space. This can be thought of as early-time scattering, when the wavefield has not yet reached the free-surface.

5.4 Analytic radiation patterns in unbounded space

The radiation patterns defined in equation (5.11) equal the squared sensitivity kernels, which in turn are a product of the equivalent source terms computed into the

previous section with the medium's Green's tensor (GT) or its spatial derivatives (equation 5.8). The analytic P-SV GT for a vertical half-plane with a free-surface has a complicated expression (see for example Kausel (2006, section 5.4)), and spatial derivatives computation would be rather intractable. However, the GT has a simple expression if the free-surface is neglected (see Sánchez-Sesma et al. (2006) for 2D and Wu & Ben-Menahem (1985) for 3D). Both 2D and 3D GT have a similar form in frequency domain. They only differ in their radial decay and in the value of some constants. If we call the dimension $D \in \{2,3\}$ the general expression is:

$$G_{ij}(\mathbf{x},\mathbf{0},\omega) = C_D\left[\delta_{ij}A_D(r) - (D\gamma_i\gamma_j - \delta_{ij})B_D(r)\right] \ , \tag{5.22}$$

with:
$C_{2D} = \dfrac{-i}{8\rho}$, $C_{3D} = \dfrac{-i}{12\pi\rho}$ - constants;
$i,j \in \{1,3\}$;
$\mathbf{x} = [x_1, x_3] = [x, z]$ - receiver position with respect to a source at the origin;
r - source-receiver distance;
$\gamma_x = x/r = \cos(\theta) = \sin(\Theta)$, $\gamma_z = z/r = \sin(\theta) = \cos(\Theta)$ - direction cosines;
δ_{ij} - Kronecker delta.
The radial terms A_D and B_D are given in Appendix 5.B, equations (5.B.1) and (5.B.2). They involve Hankel functions of the second kind with different weights between P- and S-wave terms according to the space dimension (2D or 3D). In Appendix 5.B, we use this GT and its analytic spatial derivatives to derive explicit expressions for the sensitivity kernels (equations (5.B.6)-(5.B.8)). These expressions are valid for any incident wave type, which controls the terms involving $u^{(0)}$. The derivatives of the incident wavefield can be computed analytically using $\partial_i^{\mathbf{x}} u_j = \partial_i G_{jk}(\mathbf{x}, \mathbf{x_s}, \omega) s_k$, with s_k being the components of the source (point force). Otherwise, a simplified far-field expression accounting for only specific wave types can be used (e.g. Rayleigh waves in equation (5.20)). In this section we use the latter approach. The vertically dependent amplitudes $A_z(z)$ and $A_x(z)$ are extracted from the numerical simulation. We analyse the shape of the resulting radiation patterns, as they describe the *early-time* scattering (before interaction with the free-surface). In section 5.5, we use these analytic radiation patterns to validate the numerical implementation of the Born approximation.

Let us first consider the patterns of P- and SV-waves. The corresponding patterns in both 2D and 3D media are shown at dimensionless scatterer-receiver distances $r/\lambda = [0.07, 0.4, 1, 10]$ in Figure 5.7 for P-waves and Figure 5.8 for SV-waves. Whatever the distance from the scatterer and whatever the contrast type, they appear symmetric with respect to both z- and x- axes. Now, considering the case of Rayleigh-type incident motion (see section 5.3.2), we select the dimensionless distances corresponding to the black iso-lines in Figure 5.3, i.e. $r/\lambda = [0.07, 0.14, 0.4, 1]$. Values beyond 1 are unrealistic, since fundamental mode Rayleigh waves do not penetrate beyond one wavelength. The corresponding patterns are shown in Figure 5.9. Their characteristic feature is the asymmetry between both back/forward and down-going/up-going scattering for ρ- and μ- contrasts. The sense of the asymmetry reverts around the transition frequency between the retrograde and the prograde propagation modes ($z = r \approx 0.14\lambda$, see Figure 5.3). This suggests that the asymmetry is directly related to the sense of rotation of the equivalent point force (ρ-contrast) and moment tensor axes (μ-contrast). In Appendix 5.C, we investigate which terms in the analytic

expression (5.B.6) are responsible for the asymmetry of the ρ-pattern. We find that these terms are large for $r/\lambda < 1$. Since Rayleigh waves do not penetrate beyond one wavelength anyway, this suggests an asymmetric ρ-pattern for all the real-data applications. This observation holds for both 2D and 3D radiation patterns. Nevertheless, the patterns themselves exhibit strong discrepancies between 2D and 3D for $r/\lambda < 1$, due to a different near-field behaviour.

In the next section, we use 2D numerical simulations to investigate the effect of the free-surface on the scattered wavefield.

5.5 Numerical radiation patterns in 2D half-space

In the previous sections, the only numerical input for Rayleigh waves were the scalar amplitudes of motion A_x and A_z, along with their vertical derivatives (see section 5.3.2). We used them to compute the excitation terms in the fully analytic expressions of the sensitivity kernels in equations (5.B.6)-(5.B.8), which allowed to plot radiation patterns in Figure 5.9. In this section, on the contrary, we directly estimate all the terms in equation 5.8 from 2D numerical simulations. This allows to address background models for which the analytic Green's tensor is unknown or too complicated to manipulate. Namely, we address the scattering problem in a half-space containing a free-surface (i.e. where the free-surface also influences the *scattered* wavefield). The numerical estimation of the sensitivity kernels requires two simulations: one to estimate the incident wavefield (forward simulation), and another one to estimate the tensor $\mathbf{G}^{\mathrm{rx}}$ (adjoint simulation). The theory for this method can be found in Tromp et al. (2005), while our frequency-domain implementation in 2D is described in Chapter 4. First, we validate the numerical approach using the analytic results from the previous section, and then we apply it to investigate the free-surface effect.

5.5.1 Validation of the numerical implementation

In order to validate the numerical approach, we perform adjoint simulations into an unbounded medium without free-surface (model extended to positive z with absorbing boundary conditions on the top), while the forward simulation is performed into a half-space with a free-surface. In this way the free-surface influences exclusively the incident wavefield, which corresponds to the assumptions made in the previous section for analytic kernel derivations. Also, from now and for the remainder of the paper, we switch to the surface representation of the radiation patterns (see definition (5.12)). The adjoint sources (i.e. the points for which the value of the radiation pattern is computed) are set at the receiver locations in Figure 5.1, with a 25 m spacing. In practice, only one adjoint simulation is needed because the medium in invariant under horizontal translation. Analytic and numerical results are shown respectively with solid and dotted lines in Figure 5.10. Because of the surface representation, the analytic patterns appear deformed compared to those in Figure 5.9, as r/λ now varies with the scattering angle for a fixed z/λ. A very good match is observed between analytic and numerical predictions. Namely, the predicted asymmetry is verified with the numerical simulations. The largest discrepancy is observed for the ρ-pattern at

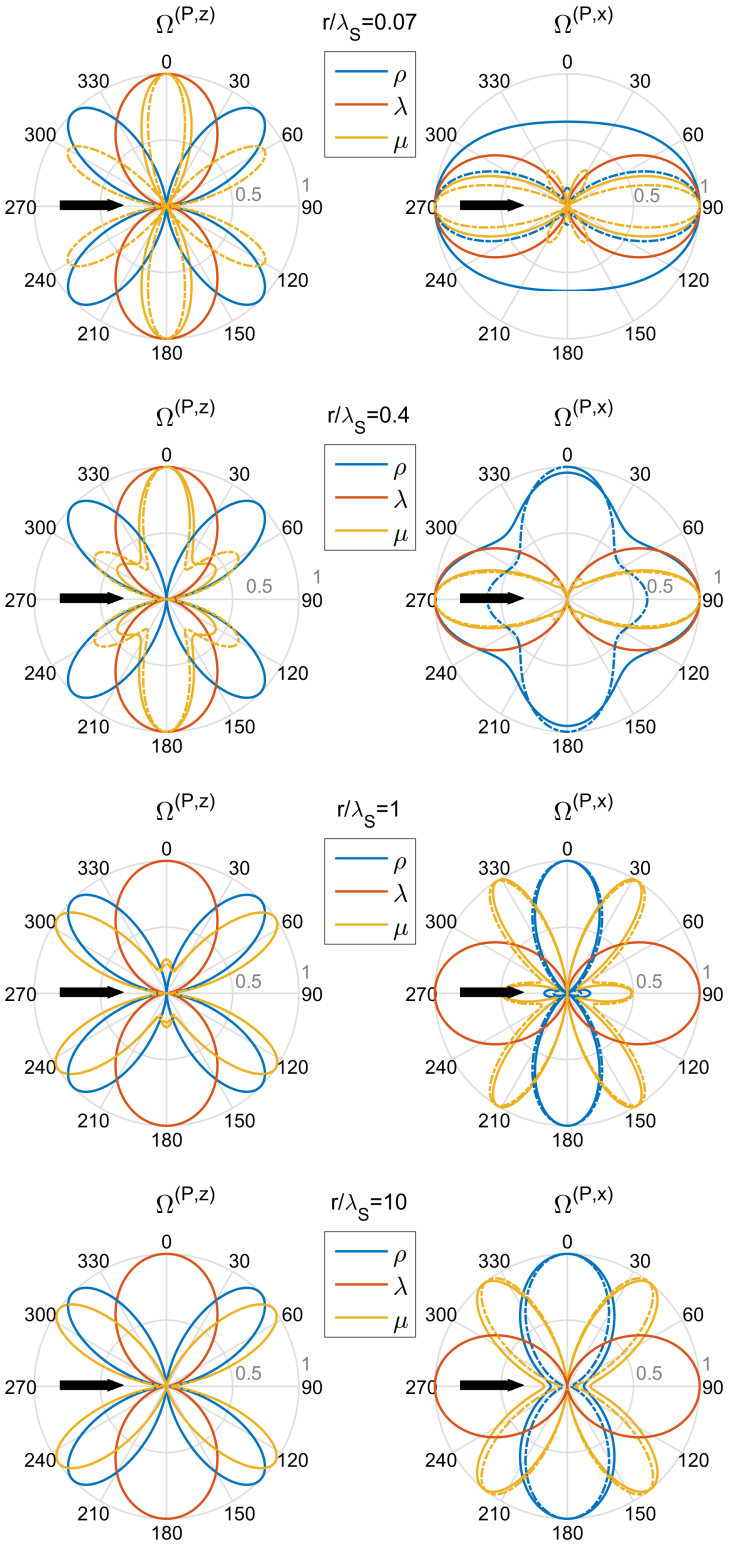


Figure 5.7: Radiation patterns $\Omega(r, \Theta, \omega)$ for horizontally incident P waves. Each pattern is normalized to its maximum value. Vertical (resp. horizontal) component is shown on the left (resp. right). 2D - solid lines; 3D - dashed lines. The black arrow indicates the incident wave propagation direction.

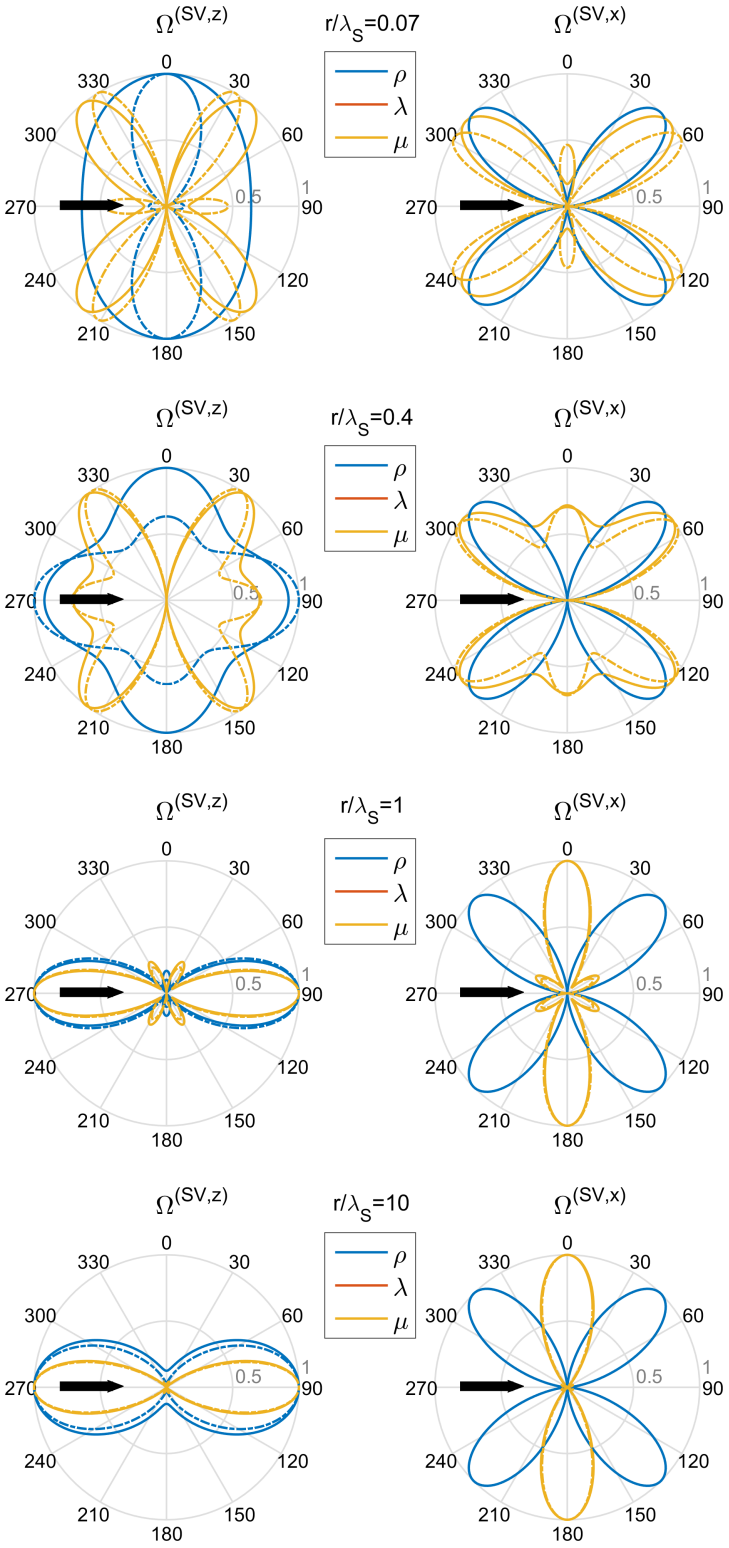


Figure 5.8: Same as Figure 5.7 for horizontally incident SV waves. Each pattern is normalized to its maximum value.

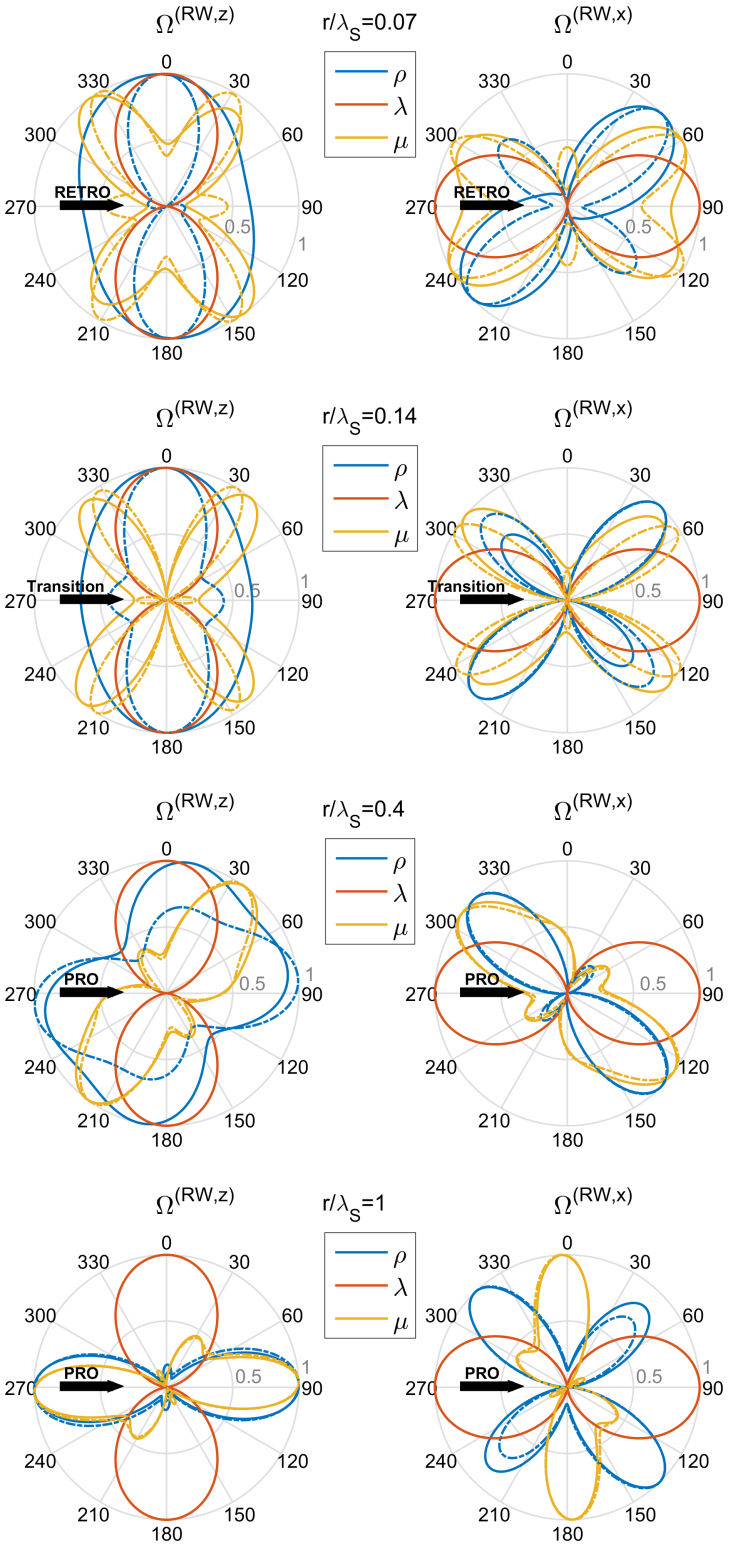


Figure 5.9: Same as Figure 5.7 for Rayleigh waves. Each pattern is normalized to its maximum value.

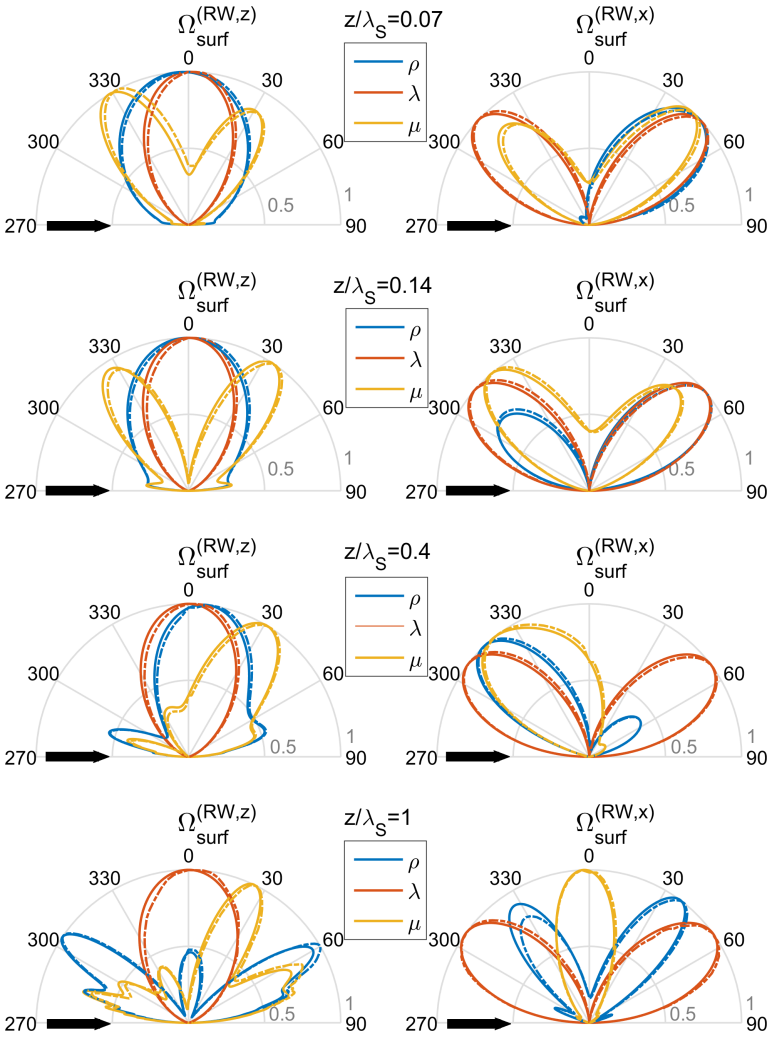


Figure 5.10: Numerically (dotted lines) and analytically (solid lines) estimated radiation patterns Ω_{surf} (see definition 5.12) for Rayleigh waves propagating along x-axis, assuming scattering in 2D unbounded plane (no free-surface). Results are displayed for the same values of dimensionless depth z/λ as in Figure 5.9. Each pattern is normalized to its maximum value.

$z/\lambda = 1$, on the horizontal component. As in Figure 5.4, this is likely due to the fact the numerical simulation does not contain only pure Rayleigh waves. Small body wave contributions are still present, which can be observed in Figure 5.3, where the phase shift between horizontal and vertical motion is not exactly $\pm 90°$ everywhere. These contributions would be somewhat complicated to filter out, since SV waves propagate at a velocity very close to Rayleigh waves. Nevertheless, we consider that the current numerical implementation correctly renders the radiation patterns without the free-surface. In the following, we add the free-surface to the background model for the adjoint simulations to study its impact.

5.5.2 Free-surface effects

Radiation patterns obtained after adding the free-surface to the adjoint simulations (solid lines in Figure 5.11) are compared to those neglecting it (dotted lines in Fig-

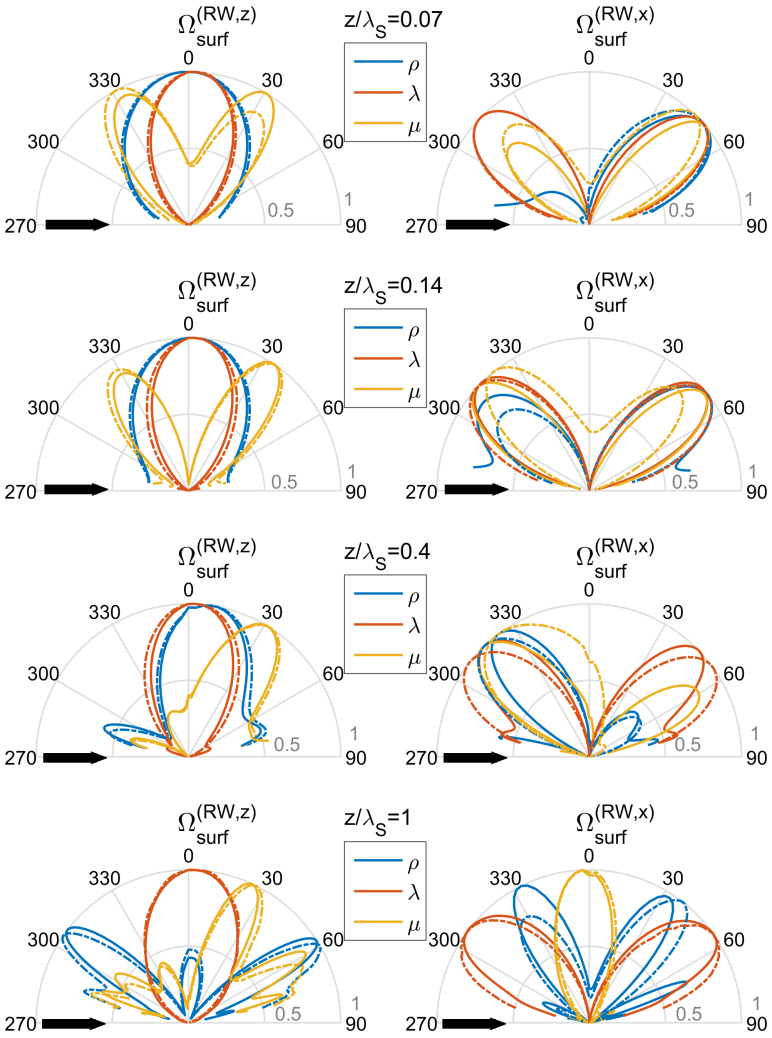


Figure 5.11: Dotted lines: same as Figure 5.10 (no free-surface). Solid lines: numerical patterns accounting for the free-surface (scattering in a 2D half-plane). Each pattern is normalized to its maximum value.

ure 5.11, which are the same as in Figure 5.10). From this comparison, we conclude that the presence of the free-surface significantly modifies the radiation patterns for the horizontal component, while the effects are limited on the vertical component. For an easier visual interpretation, the differences between the radiation patterns with and without the free-surface are shown in Figure 5.12. For the vertical component, the discrepancies grow at higher frequencies ($z/\lambda = 1$) for angles close to $\Theta = \pi/2$. In presence of a free-surface, the scattered energy at these angles propagates as Rayleigh waves which suffer no geometrical spreading in 2D, while in the absence of the free-surface, the scattered energy at $z = 0$ m tend to vanish for wide angles because of the geometrical spreading. This explains the observed difference. On the horizontal component, strong discrepancies can arise at any scattering angle, especially for the μ-contrast, meaning that the free-surface has a capital influence on the scattering process.

Let us now consider the total scattered energy, for which we have defined the quantity

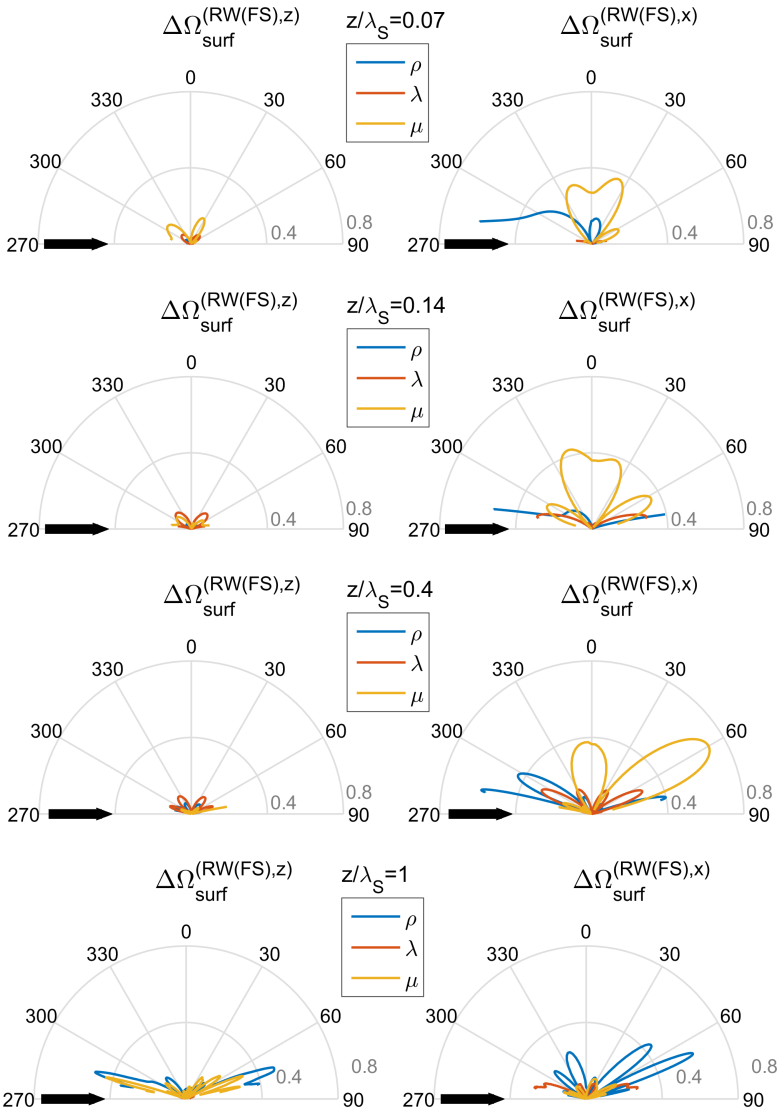


Figure 5.12: Difference between solid and dashed lines in Figure 5.11.

$E^{(1)}$ in section 5.2, equation (5.13). Normalised by the mean energy of the incident wavefield $E^{(0)}$, it is represented in Figure 5.13a for ρ-, λ- and μ-contrasts. The unit is m^{-4} because, as it can be seen in equation (5.10), this normalised ESD is given per relative medium perturbation and per squared perturbation area. That is, a ρ perturbation of -10% spanning over 10^5 m^2 within the interval $0.4 < z/\lambda_S < 0.9$ generates single-scattered energy which amounts to about $10^{-2}E^{(0)}$. We notice the scattering is the most efficient for a scatterer located between 0.4 and $0.9z/\lambda$. This can be interpreted as follows :

- Rayleigh wave fundamental mode does not penetrate beyond one wavelength (upper limit).
- Low frequencies are not sensitive to very shallow formations (lower limit).

μ- and ρ-contrasts clearly dominate the scattering, since they are up to 4 orders of magnitude above λ in terms of radiated energy. This is consistent with Rayleigh wave being weakly sensitive to λ. An important remark concerning the ESD plots in Figure 5.13a is that they are not self-similar if we change the scale of the problem, unlike the radiation patterns. Making them self-similar would require the medium perturbation area to be normalised by the squared wavelength in equation (5.10):

$$\left|u_i^{(1)}(\mathbf{x}_r,\omega)\right|^2 = \left(\epsilon\frac{\delta S}{\lambda^2}\right)^2 \left|\lambda^2 K_i^{(m)}(\mathbf{x}_r,\mathbf{x_0},\omega)\right|^2 \ . \tag{5.23}$$

$\lambda^2 K$ should then be used instead of K in all the subsequent equations. This would not affect the radiation patterns, which are plotted for a fixed λ. However, this would introduce a decreasing trend $\sim (z/\lambda)^{-4}$ in Figure 5.13a. We have chosen to avoid such a representation because the reader could have an artificial impression that lower frequencies generate stronger scattering.

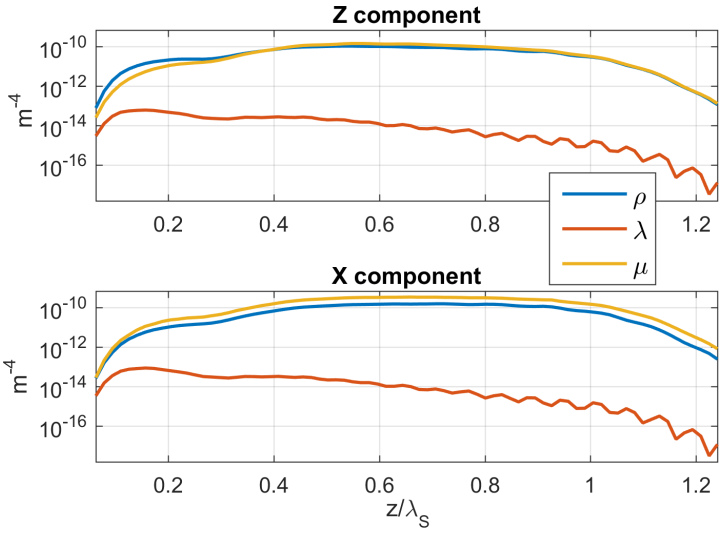
The contribution of the free-surface to the scattered energy is defined with respect to the scattering without the free-surface:

$$\frac{\Delta E}{E_0} = \frac{E^{(1)}_{\mathrm{FS}} - E^{(1)}_{\text{no FS}}}{E^{(1)}_{\text{no FS}}} \ . \tag{5.24}$$

We display it in Figure 5.13b on both components for all types of elastic contrasts. The scattered field energy related to the free-surface is in fact several times higher than the one obtained in the absence of the free-surface. The dimensionless depth corresponding to the strongest free-surface response depends on the elastic parameter. While for λ the shallowest depth of the scatterer generates the strongest free-surface response, the peak is shifted towards a bigger depth for ρ ($z/\lambda \approx 0.15$) and for μ ($z/\lambda \approx 0.3$). On the vertical component, considering the relatively small pattern deviations observed in Figure 5.12, together with the strong energy difference due to the free-surface, we suggest that the free-surface acts, to some extent, as a magnification factor applied to the unbounded medium radiation pattern.

To conclude this study, it would be interesting to visualise the free-surface contribution to the scattered wavefield in the full model. The comfortable solution would be to

(a) Energy Spectral Density (ESD) of the scattered wavefield ($E^{(1)}/E^{(0)}$)



(b) Free-surface contribution to $E^{(1)}$

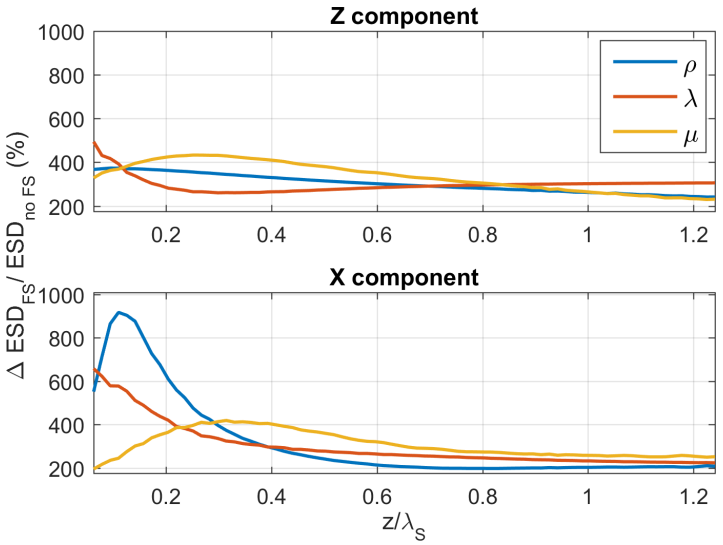


Figure 5.13: Normalised scattered energy vs. dimensionless depth.

extract it directly from numerical simulations, without using the Born approximation in each point of the model. The full scattered wavefield can indeed be obtained as a difference between two simulations, one being performed into the reference model, and another in the perturbed one, as was already done in Figure 5.2. However, the free-surface contribution is not so easy to extract because incident Rayleigh waves will not form if the free-surface is suppressed in one of the simulations. To circumvent this difficulty, we take advantage of the equivalent source terms derived in section 5.3.2. For instance, a rotating force is representative of ρ-scattering of Rayleigh waves. Such a source can be implemented as a superposition of two orthogonal directional point forces shifted in time by a quarter of the dominant period. If we use Ricker wavelets of dominant frequencies $f_0 = 1$ Hz, the horizontal source being delayed by $\Delta T = 0.25$ s, we get a prograde rotating force at $f = 1$ Hz. This source, placed at a depth of 500 m ($z/\lambda = 0.25$), is used in two numerical simulations, one performed in an unbounded plane, and another one containing a free-surface. The difference between both simulations corresponds to the free-surface induced part of the ρ-scattered wavefield (except it assumes a circular particle motion, i.e. $A_x = A_z$, instead of the elliptic one). The squared amplitude of this difference is plotted in Figure 5.14b. In comparison, the case of a vertical source is shown in Figure 5.14a. The expected unbounded-medium analytic radiation patterns are plotted around the source for both cases (white curves). The amplitudes at surface are plotted in the upper panel of the figure. As expected from symmetry considerations, a symmetric unbounded space radiation pattern triggers a symmetric free-surface response (Figure 5.14a). Asymmetry in the pattern generates asymmetry in the free-surface response (Figure 5.14b). In the vicinity of the source, the free-surface response presents stronger amplitudes for stronger pattern directions, while surface waves form in the far-field. This is an illustration of how the free-surface magnifies the unbounded medium radiation patterns on the vertical component at steep angles (within $\Theta \approx \pm 60°$, see Figure 5.12). From a close inspection of the surface amplitudes plotted in Figure 5.14, the reader might notice the energy of the free-surface-induced contribution peaks at about 30% of the total scattered wavefield, while one could expect a more significant contribution based on Figure 5.13b, where the free-surface effect seems overwhelming. To explain this apparent contradiction, we write the total scattered energy as

$$\left(u^{(1)}_{\text{tot}}\right)^2 = \left(u^{(1)}_{\text{FS}}\right)^2 + \left(u^{(1)}_{\text{no FS}}\right)^2 + 2\left(u^{(1)}_{\text{FS}}\right)\left(u^{(1)}_{\text{no FS}}\right)^* \ . \tag{5.25}$$

The plots in Figure 5.13b correspond to $\left[\left(u^{(1)}_{\text{tot}}\right)^2 - \left(u^{(1)}_{\text{no FS}}\right)^2\right] / \left(u^{(1)}_{\text{no FS}}\right)^2$, which equals ≈ 5 for $z/\lambda = 0.25$. On the other hand, the surface amplitude ratio in Figure 5.14 corresponds to $\left(u^{(1)}_{\text{FS}}\right)^2 / \left(u^{(1)}_{\text{tot}}\right)^2$, and reaches 0.3 at maximum. This simply means that the cross-term $2\left(u^{(1)}_{\text{FS}}\right)\left(u^{(1)}_{\text{no FS}}\right)^*$ carries about half the total scattered energy at the surface. Such a strong correlation is not surprising since $u^{(1)}_{\text{FS}}$ is excited via reflection or mode conversion of $u^{(1)}_{\text{no FS}}$ arriving at the surface.

5.6 Discussion

Our study, combining analytic and numerical tools, allows to discuss perspectives of the near-field sub-vertical Rayleigh wave scattering as an imaging tool. First, from

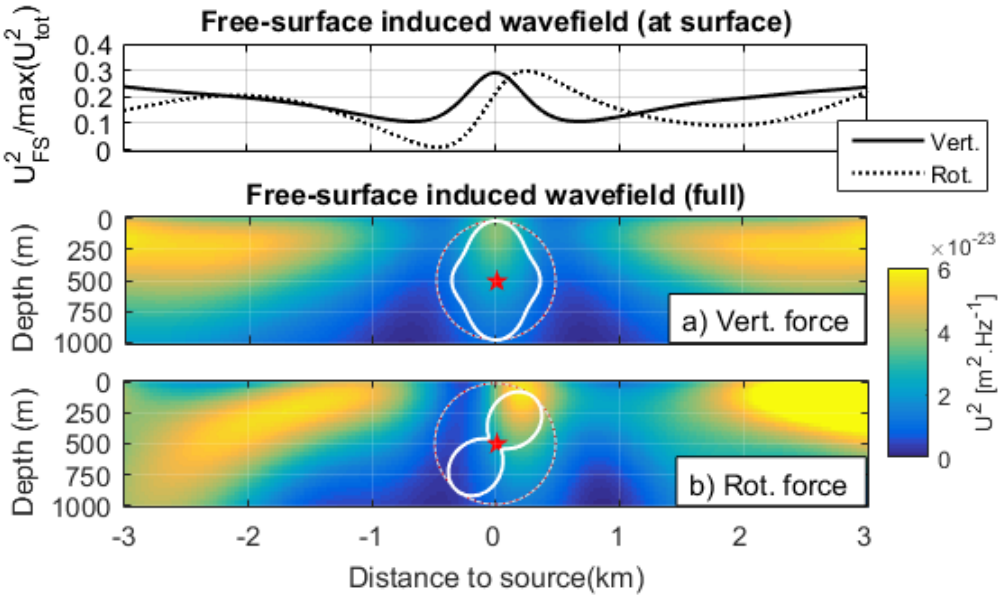


Figure 5.14: ESD of displacements wavefield induced by the free-surface in response to an excitation by a vertical point force (a) or a prograde rotating point force (b), at a frequency of 1 Hz ($z/\lambda_S = 0.25$). The corresponding radiation patterns in unbounded medium are plotted with white lines around the source (red star). The upper plot corresponds shows the same quantity extracted at surface for both vertical (solid) and rotating (dotted) source sources. Note the asymmetry of the dotted line.

the full radiation patterns in Figure 5.11, we see that the signatures of the three elastic parameters are different, and vary with frequency in a different way, especially on the vertical component (left column). While ρ presents a pattern dominated by a central lobe around 0° at low frequency ($z/\lambda_S = 0.07$), side-lobes around ±60° progressively grow and start dominating at $z/\lambda_S = 1$. The evolution is different for μ: symmetric side-lobes around ±30° dominate at low frequency, but the one at −30° progressively disappears as the frequency grows ($z/\lambda_S = 0.4$). At $z/\lambda_S = 1$, multiple small side-lobes also develop over the whole range of angles. Finally, the shape of the λ-pattern is invariant with frequency because of its explosive nature (pure P-wave emission). On the horizontal component (right column), the ρ-pattern switches from forward- to back-scattering at $z/\lambda = 0.14$, with the associated lobes located at ±45°. Note that the asymmetry of the scattering has opposite directions on the two components. As the frequency grows, the pattern tends to recover a symmetric shape ($z/\lambda = 1$), because of the weakening of the asymmetric term illustrated in Figure 5.C.1. Such a different behaviour of the radiation patterns suggests that measuring scattered signals over a wide enough range of frequencies could possibly allow for a multi-parameter inversion (see Virieux & Operto (2009) for an overview). However, Figure 5.13a tells us that the effect of λ-scattering is expected to be totally over-shaded by μ and ρ. We recover the well-known fact that surface waves are rather insensitive to λ. So, when posing an inverse problem based on sub-vertical Rayleigh wave scattering, an acceptable conditioning can be expected only when inverting for 2 parameters, both of which should strongly depend on ρ and μ. For example, these could be density and shear velocity or impedance.

It is interesting to note that Yu & Dravinski (2009) found an asymmetric radiation from their scattering cavity exposed to P and SV waves (their figs 8 and 9). In their case, the emission is probably governed by the trapped waves rotating along the internal cavity surface. The Born approximation would fail in this case, which totally deviates from single-scattering. This is a limitation of our approach (Born) compared to theirs (BEM): the ours is limited to small perturbations, not cavities. However, the advantage of the Born approximation is that it yields a physical understanding of the single-scattering in terms of secondary source excitation. It allowed us to study what the Rayleigh wave radiation pattern *would be* if there was no free-surface. Another advantage is that the Born-based sensitivity kernels can be directly used for inversions, as they yield the gradient of the least-squares misfit function between the modelled and the observed wavefields (Tromp et al., 2005).

Probably, one of the most interesting application domain would be the study of the ambient noise. Currently, most tomographic studies are based on Rayleigh wave dispersion inversion through Green's function extraction by cross-correlation (e.g. Brenguier et al., 2007; Mordret et al., 2014). This technique only uses direct waves without accounting for scattering by macro-inclusions. However, Lobkis et al. (2006) and Sánchez-Sesma et al. (2006) demonstrated that seismic interferometry can be a promising tool for extracting the scattered part of the wavefield. A linearised forward model for Rayleigh wave scattering would allow to invert this new field observable. Because of the radiation pattern shape (see Figure 5.11), important amplitude variations are expected in the scattered wavefield above the scatterer. Seen from a surface receiver array, this means that the stations which are located more or less above the scatterer are expected to measure stronger amplitude perturbations. Using the ambient noise amplitude variations, if correctly measured, opens the perspective of high-resolution imaging compared to standard tomographic inversions, where only the phase is used. Similar to background/reflectivity separation in active seismic, standard surface wave tomography could be used to retrieve a layered background model, and amplitude information added afterwards to retrieve sharp scattering zones. The inverse problem could be handled by numerical sensitivity kernels similar to those used in section 5.5, but computed within a layered background model.

Let us now discuss some of the limitations. In this study, we mainly focused on the vertical plane 2D scattering of Rayleigh waves in a homogeneous half-space. This allowed us to realise that the radiation patterns were asymmetric with respect to both horizontal and vertical axes for scattering by ρ- and μ-contrasts, contrary to the body wave analogues (P and SV). In 3D, we analytically verified that this asymmetry was also expected. However, we noticed significant discrepancies in the near-field between 2D and 3D unbounded medium patterns (Figure 5.9). These discrepancies would definitely persist in numerical simulations accounting for the free-surface. Full numerical 3D modelling (same as in section 5.5, but using 3D forward and adjoint wavefields) would thus be necessary for a quantitative Born-modelling of Rayleigh wave scattering in real datasets. The resulting sensitivity kernels would also account for 3D effects we did not address in this paper, such as:

– scattering in other planes than the plane of the incident particle motion (xz).

– coupling to Love waves.

– effects of model heterogeneity in the transverse (y) direction.

Accounting for model heterogeneity in the x-direction would also require a numerical approach, even in 2D. Another argument in the favour of the numerical modelling is that the analytic part of our work cannot be directly extended to Rayleigh wave overtones, which can form in heterogeneous layered media. First, since their penetration depth is larger at a given frequency, their scattering might be observed from greater distances than one wavelength. Second, it would be difficult to plot general radiation patterns for overtones, since the self-similarity breaks down in layered media, and, to our knowledge, there is no such dimensionless parameter as z/λ_S which would entirely determine the shape of the pattern. Instead, one could compute the individual eigenfunctions for each mode for a given model (for example using software such as Computer Programs for Seismology (Herrmann, 2013)), in order to extract A_z, A_x, $\partial_z A_x$ and $\partial_z A_x$, and inject them into equations (5.B.6)-(5.B.8) to obtain the patterns. However, since several modes are often mixed at the same frequencies, the modal decomposition of a real wavefield is not an easy task, especially for ambient noise (see for example Chapter 3). On the other hand, numerically estimated sensitivity kernels would automatically account for all the wave types. An appropriate source distribution must be used in the simulations, so that the overall modal distribution is similar to the real one (see Section 3.E for a discussion on the source position influence on the modal content).

Acknowledgements

The authors thank Pierre Dublanchet for a helpful discussion concerning the equivalent moment tensors. They are also grateful to Storengy for funding this research.

5.7 Conclusion

In this work we addressed the problem of elastic Rayleigh wave scattering in the vertical plane, unlike the more common horizontal plane view. This approach allows to model what happens at steep angles above the scatterer, where the modal summation does not correctly account for the scattered non-trapped body waves. We used a homogeneous background model. The Born approximation coupled to the reference medium Green's tensor allowed to construct the corresponding radiation patterns. Each elastic parameter acts as a specific equivalent source, which we derived analytically and verified numerically for plane P, SV and Rayleigh waves. We found that the Rayleigh wave elliptic motion generates rotating source terms. Due to that, the radiation patterns for density and the second Lamé coefficient μ are asymmetric in both 2D and 3D for Rayleigh wave scattering, contrary to P and SV wave scattering. The direction of the asymmetry is closely related to the prograde/retrograde regime of the Rayleigh wave motion. The asymmetry is significant when the scatterer is buried at less than one S-wavelength, meaning that the radiation patterns for real applications are indeed expected to be asymmetric. On the vertical component, within $\pm 60°$ from

the vertical, the pattern in 2D with a free-surface preserves the shape expected before the scattered wavefield interaction with the free-surface. The amplitude of the free-surface contribution to the scattered wavefield, however, is several times above the scattered wavefield amplitude in the unbounded medium. In terms of scattered energy, μ- and ρ-scattering dominate over λ-scattering by several orders of magnitude. The different shapes of μ- and ρ-patterns open a perspective for a multi-parameter imaging using near-field Rayleigh wave scattering. Nevertheless, 2D and 3D patterns exhibit significant differences in the near-field, suggesting that 2D modelling, though useful to get some intuitive notions, might be insufficient for real experiments.

Appendix

5.A Analytic expressions for equivalent source terms

5.A.1 Rayleigh waves

In this Appendix we derive the equivalent source terms for elastic contrasts placed into the wavefield given in equation (5.20) (Rayleigh waves). Using (5.8a), we find that a ρ-contrast localised in $[x_0, z_0]$ acts as a rotating point force:

$$\begin{cases} F_x^{(\rho)}(x,z,\omega) = i\epsilon\delta S\rho^{(0)}\omega^2 A_x(z,\omega)\Phi_{RW}(x,z,\omega) \\ F_z^{(\rho)}(x,z,\omega) = \epsilon\delta S\rho^{(0)}\omega^2 A_z(z,\omega)\Phi_{RW}(x,z,\omega) \ , \end{cases} \tag{5.A.1}$$

where we have introduced the term

$$\Phi_{RW}(x,z,\omega) = \exp(ikx)\delta(x-x_0)\delta(z-z_0) \tag{5.A.2}$$

containing the incident wavefield phase and the Dirac distribution at the scatterer position. The divergence of the incident wavefield is

$$\partial_l u_l^{(0)} = \left[\partial_z A_z(z,\omega) - kA_x(z,\omega)\right]\exp(ikx) \ . \tag{5.A.3}$$

Replacing this into (5.8b) yields that a λ-contrast is equivalent to an isotropic moment tensor of magnitude

$$M^{(\lambda)}(x,z,\omega) = -\epsilon\delta S\lambda^{(0)}\left[\partial_z A_z(z,\omega) - kA_x(z,\omega)\right]\Phi_{RW}(x,z,\omega) \ . \tag{5.A.4}$$

Similarly, replacing the derivatives of (5.20) into (5.8c), we can represent a μ-contrast by a rotating moment tensor

$$\mathbf{M}^{(\mu)}(x,z,\omega) = -\epsilon\delta S\mu^{(0)}\begin{pmatrix} -2kA_x & i(\partial_z A_x + kA_z) \\ i(\partial_z A_x + kA_z) & 2\partial_z A_z \end{pmatrix}\Phi_{RW}(x,z,\omega) \ . \tag{5.A.5}$$

It is of interest to assess the principal axes of this moment tensor and to quantify its isotropic and double-couple components. This requires to diagonalise the matrix in the expression of $\mathbf{M}^{(\mu)}$. In order to avoid complications with interpreting complex eigenvalues and eigenvectors, we switch to time domain considering monochromatic waves of frequency ω_0. Thus, the incident wave motion becomes

$$\begin{cases} u_x^{(0)}(x,z,t) = -A_x(z,\omega_0)\sin(\omega_0 t - kx) \\ u_z^{(0)}(x,z,t) = A_z(z,\omega_0)\cos(\omega_0 t - kx) \ , \end{cases} \tag{5.A.6}$$

and the matrix we wish to diagonalise is

$$\mathcal{M} = \begin{pmatrix} 2kA_x\cos(\phi) & (kA_z - \partial_z A_x)\sin(\phi) \\ (kA_z - \partial_z A_x)\sin(\phi) & 2\partial_z A_z\cos(\phi) \end{pmatrix} \equiv \begin{pmatrix} a & b \\ b & c \end{pmatrix} \ , \tag{5.A.7}$$

where $\phi = \omega_0 t - kx$. $\mathcal{M}$ has two eigenvalues of the form

$$\begin{aligned} \lambda_\pm &= \frac{1}{2}(a + c \pm \sqrt{\Delta}) \ , \text{ with} \\ \Delta &= (a-c)^2 + 4b^2 \ . \end{aligned} \tag{5.A.8}$$

Replacing a, b and c by their explicit expressions given in (5.A.7), and using (5.19), the two components of the tensor are

$$\begin{aligned} M_{\text{iso}}^{(\mu)}(x,z,t) &= (\partial_z A_z + kA_x)\cos(\phi) \\ M_{\text{dc}}^{(\mu)}(x,z,t) &= \sqrt{(\partial_z A_z - kA_x)^2\cos^2(\phi) + (kA_z - \partial_z A_x)^2\sin^2(\phi)}\ . \end{aligned} \tag{5.A.9}$$

From this equation we can expect the isotropic part to be enhanced when $\partial_z A_z$ and kA_x are of opposite signs. For the fundamental mode, since $\partial_z A_z$ is always negative, this corresponds to a scatterer located in the shallow zone where the propagation is retrograde ($A_x > 0$). After some algebra, we find the orientation of one of the principal axes, α_+, via

$$\tan(\alpha_+(t)) = 2\frac{\partial_z A_z\cos(\phi) + \sqrt{(\partial_z A_z - kA_x)^2\cos^2(\phi) + (kA_z - \partial_z A_x)^2\sin^2(\phi)}}{(kA_z - \partial_z A_x)\sin(\phi)}\ . \tag{5.A.10}$$

Note that the principal axes are expected to evolve with time because of $\cos(\phi)$ and $\sin(\phi)$ terms.

5.A.2 SV waves

In this Appendix we repeat the derivations of the Appendix 5.A.1 replacing Rayleigh waves by SV waves. The wave motion associated with horizontally incident SV waves can be written as

$$\begin{cases} u_x^{(0)}(x,z,\omega) = 0 \\ u_z^{(0)}(x,z,\omega) = A_S\exp(ikx)\ , \end{cases} \tag{5.A.11}$$

with $k = \omega/\beta$ the S-wavenumber. The equivalent force for a ρ contrast is now oscillating along the vertical direction instead of rotating:

$$\begin{cases} F_x^{(\rho)}(x,z,\omega) = 0 \\ F_z^{(\rho)}(x,z,\omega) = \epsilon\delta S\rho^{(0)}\omega^2 A_S\Phi_{SV}(x,z,\omega)\ , \end{cases} \tag{5.A.12}$$

where Φ_{SV} has the same definition as Φ_{RW} (see equation 5.A.2), but using the wavenumber appropriate for SV waves. Because the incident amplitude does not depend on z any more, while the horizontal component is zero, the divergence of the incident wavefield vanishes. Thus we recover that shear waves are insensitive to λ-contrasts. Finally, the equivalent moment tensor for a μ-contrast is

$$\mathbf{M}^{(\mu)}(x,z,\omega) = -\epsilon\delta S\mu^{(0)}\begin{pmatrix} 0 & kA_S \\ kA_S & 0 \end{pmatrix}\Phi_{SV}(x,z,\omega)\ . \tag{5.A.13}$$

The diagonalisation is evident directly in frequency domain: the eigenvalues are

$$\lambda_\pm = \pm kA_S\ , \tag{5.A.14}$$

with associated eigenvectors

$$V_\pm = \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}\ . \tag{5.A.15}$$

This is a pure double-couple moment tensor whose axes have a constant orientation

$$\alpha_\pm = \pm 45^\circ\ . \tag{5.A.16}$$

5.A.3 P waves

In this Appendix we repeat the derivations of the Appendix 5.A.1 replacing Rayleigh waves by P waves. The wave motion associated with horizontally incident P waves is

$$\begin{cases} u_x^{(0)}(x,z,\omega) = A_P \exp(ikx) \\ u_z^{(0)}(x,z,\omega) = 0 \ , \end{cases} \tag{5.A.17}$$

with $k = \omega/\alpha$ the P-wavenumber. The equivalent force for a ρ contrast is oscillating along the horizontal direction:

$$\begin{cases} F_x^{(\rho)}(x,z,\omega) = \epsilon\delta S \rho^{(0)} \omega^2 A_P \Phi_P(x,z,\omega) \\ F_z^{(\rho)}(x,z,\omega) = 0 \ , \end{cases} \tag{5.A.18}$$

where Φ_P has the same definition as Φ_{RW} (see equation 5.A.2), but using the wavenumber appropriate for P waves. The divergence of the incident wavefield is

$$\partial_l u_l^{(0)} = -kA_P \exp(ikx) \ . \tag{5.A.19}$$

The equivalent isotropic moment for a λ-contrast is thus

$$M^{(\lambda)}(x,z,\omega) = -\epsilon\delta S \lambda^{(0)} k A_P(z,\omega) \Phi_P(x,z,\omega) \ . \tag{5.A.20}$$

The equivalent moment tensor for a μ-contrast is

$$\mathbf{M}^{(\mu)}(x,z,\omega) = -\epsilon\delta S \mu^{(0)} \begin{pmatrix} -kA_P & 0 \\ 0 & 0 \end{pmatrix} \Phi_P(x,z,\omega) \ . \tag{5.A.21}$$

The only eigenvalue of the matrix in the above equation is

$$\lambda = -kA_P \ , \tag{5.A.22}$$

with the associated eigenvector

$$V = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ . \tag{5.A.23}$$

This is a linear dipole along the x-axis.

5.B Analytic sensitivity kernels derivation in unbounded space

In this Appendix we derive explicit analytic expressions for the sensitivity kernels associated with different elastic contrasts in unbounded 2D and 3D media. The first step is to write the appropriate Green's tensor (GT) and to find its spatial derivatives. The latter are then replaced into equation (5.8) to yield the desired kernels. The general expression of the GT for both 2D and 3D media (see Sánchez-Sesma et al. (2006) and Wu & Ben-Menahem (1985), respectively), is given in equation (5.22). It involves radial terms which have the following expressions in 2D:

$$\begin{aligned} A_{2D}(r) &= \frac{H_0^{(2)}(qr)}{\alpha^2} + \frac{H_0^{(2)}(kr)}{\beta^2} \\ B_{2D}(r) &= \frac{H_2^{(2)}(qr)}{\alpha^2} - \frac{H_2^{(2)}(kr)}{\beta^2} \ , \end{aligned} \tag{5.B.1}$$

and in 3D:

$$\begin{aligned} A_{3D}(r) &= \frac{k_\alpha H_0^{(2)}(qr)}{\alpha^2} + 2\frac{k_\beta H_0^{(2)}(kr)}{\beta^2} \\ B_{3D}(r) &= \frac{k_\alpha H_2^{(2)}(qr)}{\alpha^2} - \frac{k_\beta H_2^{(2)}(kr)}{\beta^2} \ , \end{aligned} \tag{5.B.2}$$

with:
α, β - P and S velocities;
$q = \omega/\alpha$, $k = \omega/\beta$ - P and S wave-numbers;
$H_m^{(2)}$ - Hankel function of the second kind and order m.

The derivatives required in equations (5.8b) and (5.8c) are respectively $\partial_j G_{jk} = \nabla_0 \cdot \mathbf{G}$ (tensor divergence = vector) and $\partial_i G_{jk} = \nabla_0 \mathbf{G}$ (tensor gradient = 3D tensor). The spatial derivatives for the kernel computation must be computed at the scatterer. If we compute them at the receiver, thanks to the reciprocity principle, the expression remains unchanged, with $\mathbf{x}$ being replaced by $-\mathbf{x}$ (source-receiver exchange). This is equivalent to evaluating $-\partial_j G_{jk}$ and $-\partial_i G_{jk}$ at $\mathbf{x}$ from equation (5.22). Thus, we obtain:

$$\begin{aligned} \partial_i G_{jk} = -C_D \Big[&(\delta_{jk}\gamma_i A'_D(r) - (D\gamma_j\gamma_k - \delta_{jk})B'_D(r))\,\gamma_i \\ &- D\frac{B_D(r)}{r}\left(\gamma_j(\delta_{ik} - \gamma_i\gamma_k) + \gamma_k(\delta_{ij} - \gamma_i\gamma_j)\right)\Big] \end{aligned} \tag{5.B.3}$$

and

$$\partial_j G_{jk} = -C_D \left[A'_D(r) + (1-D)B'_D(r) + D(1-D)\frac{B_D(r)}{r} \right] \gamma_k \ . \tag{5.B.4}$$

The derivatives $A'_D(r)$ and $B'_D(r)$ are obtained from the the definitions (5.B.1) and (5.B.2) by using the Hankel functions property

$$2\left(H_m^{(2)}\right)' = H_{m-1}^{(2)} - H_{m+1}^{(2)} \ . \tag{5.B.5}$$

Substituting (5.B.3) and (5.B.4) (5.8), we get after some simplification:

$$\begin{aligned} K_x^{(\rho)} &= -\omega^2 C_D \Big\{ \left[A_D(r) + \left(1 - \frac{D}{2}\right) B_D(r) + \frac{B_D(r)}{2}\cos(2\Theta) \right] u_x^{(0)} \\ &\qquad - D\sin(\Theta)\cos(\Theta) B_D(r) u_z^{(0)} \Big\} \\ K_z^{(\rho)} &= -\omega^2 C_D \Big\{ \left[A_D(r) + \left(1 - \frac{D}{2}\right) B_D(r) - \frac{B_D(r)}{2}\cos(2\Theta) \right] u_z^{(0)} \\ &\qquad - D\sin(\Theta)\cos(\Theta) B_D(r) u_x^{(0)} \Big\} \ ; \end{aligned} \tag{5.B.6}$$

$$K_x^{(\lambda)} = C_D\Big\{A'_D(r) + (1-D)B'_D(r) + D(1-D)\frac{B_D(r)}{r}\Big\}\left(\partial_x u_x^{(0)} + \partial_z u_z^{(0)}\right)\sin(\Theta)$$
$$K_z^{(\lambda)} = C_D\Big\{A'_D(r) + (1-D)B'_D(r) + D(1-D)\frac{B_D(r)}{r}\Big\}\left(\partial_x u_x^{(0)} + \partial_z u_z^{(0)}\right)\cos(\Theta)\ ; \tag{5.B.7}$$

$$K_x^{(\mu)} = C_D\Big\{Z_{1+}^{(\mu)}\cos(\Theta)\left(\partial_x u_z^{(0)} + \partial_z u_x^{(0)}\right) + Z_{2+}^{(\mu)}\sin(\Theta)\left(2\partial_x u_x^{(0)}\right) - Z_{3+}^{(\mu)}\sin(\Theta)\left(2\partial_z u_z^{(0)}\right)\Big\}$$
$$K_z^{(\mu)} = C_D\Big\{Z_{1-}^{(\mu)}\sin(\Theta)\left(\partial_x u_z^{(0)} + \partial_z u_x^{(0)}\right) + Z_{2-}^{(\mu)}\cos(\Theta)\left(2\partial_z u_z^{(0)}\right) - Z_{3-}^{(\mu)}\cos(\Theta)\left(2\partial_x u_x^{(0)}\right)\Big\}\ , \tag{5.B.8}$$

where the following auxiliary functions have been introduced for shortening the notations:

$$\begin{aligned}
F(r) &= \frac{B'(r)}{2} - \frac{B(r)}{r}\\
Z_{1\pm}^{(\mu)} &= A'(r) + D\frac{B(r)}{r} + (1-D)B'(r) \pm 2DF(r)\cos(2\Theta)\\
Z_{2\pm}^{(\mu)} &= A'(r) + (1-\frac{D}{2})B'(r) - D\frac{B(r)}{r} \pm DF(r)\cos(2\Theta)\\
Z_{3\pm}^{(\mu)} &= D\frac{B(r)}{2} \pm F(r)\cos(2\Theta)\ .
\end{aligned} \tag{5.B.9}$$

5.C Pattern asymmetry for a ρ-contrast

In this Appendix we investigate which terms in the analytic expression of the ρ-radiation pattern for Rayleigh waves are responsible for the emission asymmetry. Replacing the Rayleigh wave motion given in equation (5.20) into (5.B.6), and then using (5.11), we get the following explicit expression for the ρ-pattern:

$$\Omega_\rho^{(\mathrm{RW,z})}(r,\Theta,\omega) = \omega^4 C_D^2 Y(\Theta), \text{ with}$$
$$Y(\Theta) = \left|\left[A_D(r) + \left(1-\frac{D}{2}\right)B_D(r) - \frac{B(r)}{2}\cos(2\Theta)\right]A_z - iD\sin(\Theta)\cos(\Theta)B_D(r)A_x\right|^2 \tag{5.C.1}$$

If we decompose the radial functions $A_D(r)$, $B_D(r)$ into their real and imaginary components,

$$\begin{aligned}
A_D(r) &= \mathfrak{Re}(A_D(r)) + i\mathfrak{Im}(A_D(r))\\
B_D(r) &= \mathfrak{Re}(B_D(r)) + i\mathfrak{Im}(B_D(r))\ ,
\end{aligned}$$

we have

$$Y(\Theta) = \Bigg[\left\{ \mathfrak{Re}(A_D(r)) + \left(1 - \frac{D}{2}\right) \mathfrak{Re}(B_D(r)) - \frac{\cos(2\Theta)}{2} \mathfrak{Re}(B_D(r)) \right\} A_z + D \sin(\Theta) \cos(\Theta) \mathfrak{Im}(B_D(r)) A_x \Bigg]^2 + \Bigg[\left\{ \mathfrak{Im}(A_D(r)) + \left(1 - \frac{D}{2}\right) \mathfrak{Im}(B_D(r)) - \frac{\cos(2\Theta)}{2} \mathfrak{Im}(B_D(r)) \right\} A_z - D \sin(\Theta) \cos(\Theta) \mathfrak{Re}(B_D(r)) A_x \Bigg]^2 . \tag{5.C.2}$$

Expanding the square sums leads to

$$\begin{aligned} &Y(\Theta) = Y_{\text{sym}}(\Theta) + Y_{\text{asym}}(\Theta), \text{ with} \\ &Y_{\text{asym}} = 2D \left[\mathfrak{Im}(B_D(r)) \mathfrak{Re}(A_D(r)) - \mathfrak{Re}(B_D(r)) \mathfrak{Im}(A_D(r)) \right] A_x A_z \sin(\Theta) \cos(\Theta) \ , \end{aligned} \tag{5.C.3}$$

while $Y_{\text{sym}}(\Theta)$ contains the squared terms of the expansion. $Y_{\text{sym}}(\Theta)$ is symmetric with respect to both vertical and horizontal axes, while $Y_{\text{asym}}(\Theta)$ is anti-symmetric with respect to both axes because it is proportional to $\sin(\Theta)\cos(\Theta)$. The sum of both terms results into the observed pattern asymmetry. The asymmetry should disappear if $\mathfrak{Im}(B_D(r))\mathfrak{Re}(A_D(r)) \approx \mathfrak{Re}(B_D(r))\mathfrak{Im}(A_D(r))$. Both terms are plotted for the 2D case in Figure 5.C.1, from which we can conclude that important asymmetry is expected for the range $r/\lambda < 1$. This corresponds to the penetration depth of Rayleigh wave fundamental mode. In practice, this means that asymmetric scattering is expected for most real-data applications. A similar analysis can be performed for μ-scattering, though we did not attempt it.

When considering body waves in isotropic media, the motion is always polarised either parallel or perpendicular to the wave propagation direction, which we can choose along x without loss of generality. Thus, in the radiation pattern expression, there can be no cross-product between the coefficients in front of u_x and u_z, so all the sine/cosine terms appear squared. This means that the radiation patterns are symmetric with respect to both the propagation direction and its orthogonal.

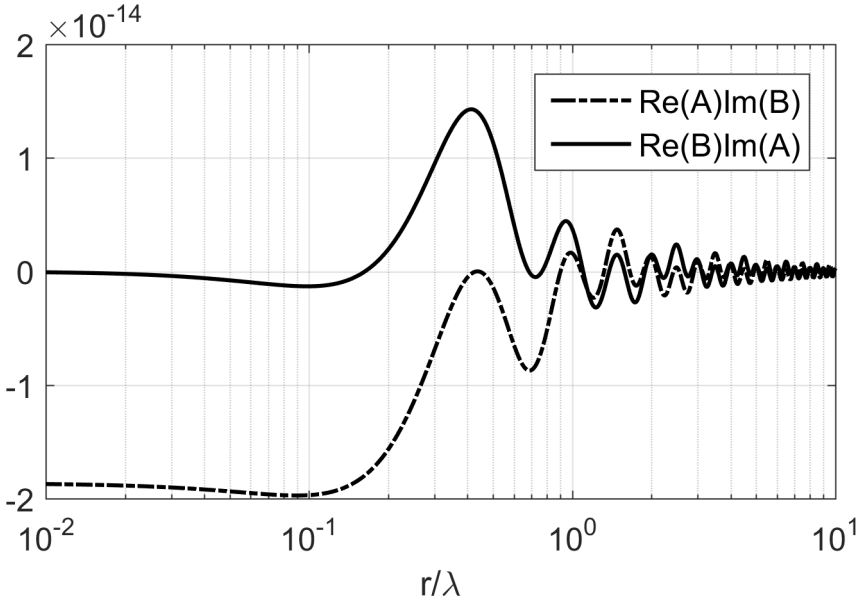


Figure 5.C.1: Terms $\mathfrak{Im}(B_{2D}(r))\mathfrak{Re}(A_{2D}(r))$ (solid) and $\mathfrak{Re}(B_{2D}(r))\mathfrak{Im}(A_{2D}(r))$ (dash-dotted) vs. dimensionless scatterer-receiver distance. Close values imply low asymmetry of the ρ-pattern.

Chapter 6

Spectral amplitude imaging with Rayleigh waves

Mise en contexte (français)

Dans ce chapitre, l'approximation de Born est utilisée pour traiter le problème inverse. Il s'agit de retrouver la forme, la position et les propriétés élastiques d'une hétérogénéité à partir de sa signature spectrale à la surface. Cela peut être vu comme une suite logique du chapitre 4, puisque les mêmes noyaux de Fréchet (noyaux de sensibilité) sont utilisés. L'étude des diagrammes de radiation effectuée dans la partie précédente est ici complétée par l'étude de la matrice Hessienne avec l'amplitude spectrale en tant qu'observable, ce qui permet de discuter le choix de la paramétrisation optimale. Il est important de garder à l'esprit que le schéma d'inversion linéarisée proposé peut s'appliquer uniquement à des petites perturbations d'amplitude spectrale (telles qu'obtenues numériquement dans le chapitre 4), et non pas à celles observées sur le terrain (chapitre 2), qui atteignent plusieurs centaines de pourcents. Cette partie est en cours de préparation pour une soumission, et se présente donc sous forme d'article.

Context (English)

In this chapter, the Born approximation is used to address the inverse problem: finding the shape, the position and the properties of an elastic inclusion based on its spectral signature at the surface. It can be regarded as a continuation of the Chapter 4, as the same Fréchet derivatives (sensitivity kernels) are used. The radiation pattern study performed in the previous chapter is here complemented by the study of the Hessian matrix for the spectral amplitude observable, which allows to discuss the optimal parametrisation choice. It is important to keep in mind that the linearised inversion can only apply to small spectral amplitude perturbations, unlike the field data where the perturbations are of several hundred percent (see Chapter 2). This part is in preparation for a submission, so it also has an article structure.

Résumé (français)

On met en évidence la possiblité d'inverser les ondes de Rayleigh basse fréquence (0.2-4 Hz) pour les propriétés élastiques du milieu à des profondeurs kilométriques, en 2D. Le schéma d'inversion s'apparente à une inversion des formes d'ondes par moindres carrés, avec une observable modifiée (amplitude spectrale à la place de la forme d'onde complète). L'information de phase est délibérément écartée, puisqu'elle n'est pas facilement accessible dans le cas du bruit ambiant. La présence de modes supérieurs dans les données améliore nettement la qualité des modèles inversés, à la fois en termes de résolution et de profondeur d'investigation, par rapport au mode fondamental pur. Le choix de la paramétrisation optimale est abordé via l'analyse de la matrice Hessienne, visant à réduire la diaphonie entre les différents paramètres et les différentes régions du modèle. Les meilleurs résultats sont obtenus pour une paramétrisation à trois paramètres (ρ, V_s, V_p), au moins pour ce qui est de la forme géométrique de l'hétérogénéité, malgré un mauvais conditionnement vis-à-vis de V_p (ou, de manière équivalente, du coefficient de Lamé λ).

Summary (English)

We demonstrate the possibility of inverting the 2D lateral elastic properties at a kilometer depth from the analysis of low-frequency Rayleigh waves (0.2-4 Hz). The inversion process is a linearised least-squares waveform inversion, only considering the diffracted wavefield, with a modified observable (spectral amplitudes instead of classical time-domain waveforms). The phase information is deliberately discarded as it is not easily available in the case of ambient noise. The presence of higher modes in the simulated data greatly improves the quality of the inversion both in terms of accuracy and investigation depth compared to a pure fundamental mode. We discuss the optimal model parameterisation through the analysis of the Hessian in terms of spatial and inter-parameter crosstalk. 3-parameter inversion (ρ, V_s, V_p) yields the best reconstructed models, at least in terms of inverted inclusion shapes, despite the poor conditioning of the inversion with respect to V_p (or equivalently the Lamé coefficient λ).

6.1 Introduction

Seismic ambient noise consists of a mixture of a very large number of arrivals generated by random sources (Gutenberg, 1958). Its amplitude can be measured directly from the raw signal. On the contrary, retrieving a coherent phase requires transforming receivers into virtual sources by the using cross-correlation techniques (Shapiro & Campillo, 2004). Ambient noise amplitude on the vertical component at frequencies around 1 Hz can contain sharp variations in space in the vicinity of geological structures (Gorbatikov et al., 2004, 2008) and/or fluid-bearing reservoirs (Dangel et al., 2003; Saenger et al., 2011; Riahi et al., 2013b). Our work is motivated by the need for a quantitative imaging method exploiting these effects. The on-shore ambient vibrations on the vertical component are often assumed dominated by fundamental mode Rayleigh waves (Bonnefoy-Claudet et al., 2006b). However, Bonnefoy-Claudet

et al. (2006b) insist that overtones can also carry an important fraction of energy. As a particular case of overtone mixture, the L_g phase was reported by Koper et al. (2010) as the propagation mode carrying about the half of the total ambient noise energy at frequencies around 1 Hz, on average in the world. This orientates our work towards multi-modal Rayleigh wavefields.

Standard imaging methods using surface waves are based on the phase. On addition, they often require assuming a laterally invariant medium, at least locally. This is the case of the dispersion curves inversion for a 1D layered model (for a review, see Strobbia (2002)). Prieto et al. (2009) proposed to use the decay of the cross-correlation amplitude with inter-station distance to retrieve the anelastic attenuation coefficient, assuming a radial propagation from the virtual source. Lin et al. (2012) worked out a generalised surface wave amplitude imaging method accounting for anelastic attenuation, wavefront (de-)focusing and local site amplification. This method treats surface waves as a solution of a 2D damped wave equation in which both velocity and local amplification are imposed by the local 1D structure. Thus, locally, this method also assumes propagation in a laterally invariant medium. On the other hand, Rayleigh wave scattering was extensively studied in the literature (Snieder, 1986; Friederich et al., 1993; Liu & Zhou, 2016; Maupin, 2017). Most of the studies are designed for applications in global or regional earthquake tomography (e.g. Ritzwoller et al., 2002). The idea is to correct the ray-theory prediction of the observed surface wave arrival times and amplitudes by accounting for the scattering which happens along the travel path from the source to the receiver. Thus only the scattering of the surface waves on themselves is considered, neglecting the conversion to body waves (Maupin, 1996).

In this work, we propose to use the Rayleigh wave amplitude to constrain the elastic parameters directly beneath the receivers, based on the scattering theory instead of the 1D eigenfunctions (Lin et al., 2012). The full scattering must then be modelled, and not only the Rayleigh-to-Rayleigh coupling. The Rayleigh-wave part of the scattered wavefield does not have time to form vertically above the scatterer. At low frequencies, the scattered wavefield in this area is a near-field mixture of body wave modes interacting with the free-surface. The associated radiation patterns for the fundamental mode scattering in the radial-vertical 2D plane were investigated in Chapter 5. In the present work, we move to a more complex 2D layer-over-halfspace model (LOH) which can also support Rayleigh wave overtone propagation. The Born approximation based on numerical Green's functions in the LOH model allows us to construct the sensitivity kernels (or Fréchet derivatives) for the amplitude at the surface, closely related to the gradient of a least-square misfit function. They can be used to invert the Rayleigh wave amplitude perturbations due to single-scattering beneath the receiver array. In this paper, we mostly focus on retrieving the scatterer size and position, rather than the correct elastic parameters inside the scatterer. Fundamentally, our problem is analogue to the one addressed by Campman & Riyanti (2007), except that we invert the *amplitude* perturbations of a random wavefield (instead of a single-scattered waveform from an individual source), and that our data contains overtones. Also, Campman & Riyanti (2007) performed a non-linear inversion in 3D with semi-analytic forward-modelling, while our forward modelling is based on

the fully numerical spectral-element method (SEM, Komatitsch et al., 1999) in 2D. Another parallel can be drawn with the non-linear windowed-amplitude waveform inversion (w-AWI), developped by Solano et al. (2014). Their observable is the amplitude of the FK spectrum of the wavefield recorded by successive segments of the receiver line. They give up some of the phase information by taking the amplitude of the FK spectrum, but the latter still contains the dispersion curves. In our case, all of the phase information is discarded, since we only use the amplitude spectra estimated at individual receivers. Since our method is based on the Born approximation (single-scattering), we need a reliable background model (not updated during the inversion) for computing the misfit function gradient. This is in contrast to the non-linear methods, which update the complete model and re-calculate the gradient at each iteration. The sensitivity of the single-scattering amplitude perturbations with respect to the uncertainties on the background model were addressed in Chapter 4. In Section 6.2, we present the inversion framework, introduce the amplitude sensitivity kernels and define the observables. In Section 6.3, we jusitfy the parametrisation choice suitable for the scatterer detection objective. Finally, we present the inversion results in Section 6.4. The discussion (Section 6.5) is dedicated to the field application potential of the method, as well as to its assumptions and limitations.

6.2 Forward and inverse problems

In this part, we define the observables used for imaging and present the inversion framework. The forward modelling part used to synthesise the observables was described in Chapter 4 and relies on the numerical spectral-element method (Komatitsch et al., 1999). An example of simulation setup is shown in Fig. 6.1. It consists of different source distributions at both ends of the model, a receiver array in the central part, beneath which the investigated zone is located (shaded blue).

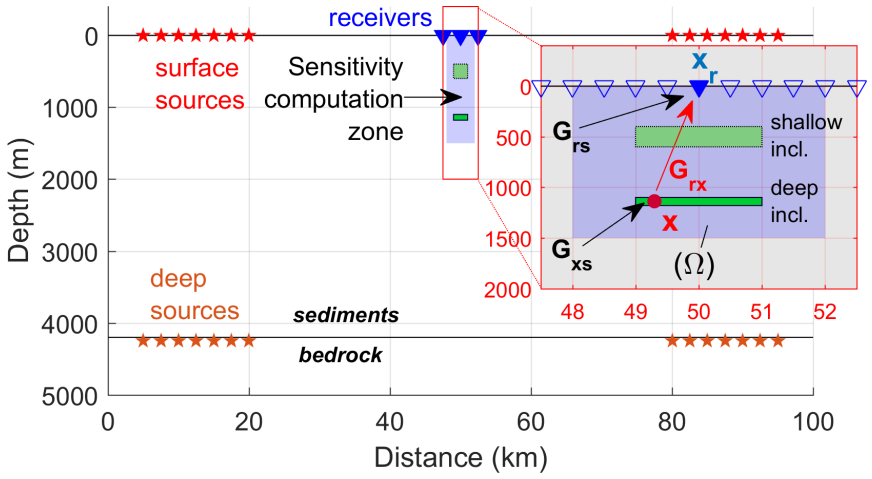


Figure 6.1: Layer-Over-Half-space (LOH) model used in the study (adapted from Fig. 4.1). Distant sources at the surface (red stars) and in depth (brown stars) are used to create incident wavefields with different modal compositions. The blue shaded zone is where sensitivity kernels are computed. It can contain two inclusions (green rectangles). Zoomed area on the right: synthetic view of the terms implied in the Born approximation formulas. The signal at the receiver $\mathbf{x_r}$ is the sum of an incident (black) and a scattered (red) wavefield. $\mathbf{x}$ is one of the scattering points over which an integration is performed to recover the total scattered wavefield at $\mathbf{x_r}$.

6.2.1 Investigated models

In this chapter, we use the two simplest models presented in chapter 4, i.e. the homogeneous and the layer-over-halfspace (LOH) models. Their parameters are those given in Table 4.1a and b. Two elastic inclusions (imaging targets) are tested (green ractangles in Fig. 6.1). Both inclusions span horizontally over 2 km, between 49 to 51 km from the left edge of the model. One of them is shallow and thick ($z \in [400, 600]$ m), and contains perturbations of all the three parameters ρ, V_p and V_s. The other one is deeper and thinner ($z \in [1100, 1180]$ m) and only contains perturbations of ρ and V_p (see zoomed zone in Fig. 6.1). The thin inclusion serves as a plausible representation of a gas reservoir with the parameters from Fig. 1.28, while the thick one is designed to test the fundamental mode inversion, which has a limited sensitivity and depth penetration. In the presented inversion examples, we compare several combinations of background model, source type, observable and inclusion to be imaged. They are summarized in Table 6.1.

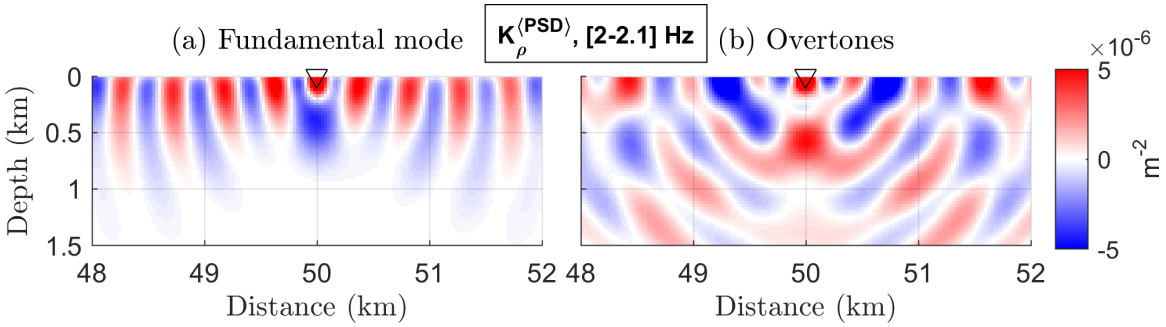


Figure 6.2: Example of sensitivity kernels $K_{(\rho)}^{\langle \mathrm{PSD} \rangle}$ for the receiver located on the surface at $x = 50$ km (black triangle). (a) Fundamental-mode dominated wavefield. (b) overtone-dominated wavefield.

Table 6.1: Summary of inversion tests presented in the paper. W = wave type, FS = presence of the free-surface, S = type of source (random distribution or single), f_0 = source central frequency, x_s = horizontal source position, z_s = source depth, Obs. = observable, Incl. = inclusion parameters; WF = waveform. The target inclusion is shallow and thick for Tests #1 and #2, while it is deep and thin for tests #3 to #5. The notation $[x_{\mathrm{start}} : \Delta x : x_{\mathrm{end}}]$ for test #5 indicates a series of equally spaced active shots recorded separately.

#	Model	W	FS	S	f_0(Hz)	x_s(km)	z_s(m)	Obs.	Incl.
1	HM	R0	yes	rand.	cf. Tab. 4.2		0	PSD	$\delta\rho/\rho_0 = -10\%$,
2	LOH	R+	yes	rand.	cf. Tab. 4.2		4250	PSD	$\delta V_p/V_p = -18\%$, $\delta V_s/V_s = -10\%$ $z \in [400, 600]$ m
3	LOH	R+	yes	rand.	cf. Tab. 4.2		4250	PSD	$\delta\rho/\rho_0 = -10\%$,
4	LOH	R+	yes	sing.	[0.5,2]	5	4250	WF	$\delta V_p/V_p = -18\%$,
5	LOH	P	no	sing.	[0.5,2]	[45:2.5:55]	0	WF	$\delta V_s/V_s = 0\%$ $z \in [1100, 1180]$ m

We saw in Chapter 4 that the Born sensitivity kernel depends on the average depth of the source distribution because the latter influences the modal composition of the wavefield. Shallow sources excite the fundamental mode, while deeper sources excite a mixture of overtones. As a reminder, the amplitude sensitivity kernels are shown in Figs 6.2a and b respectively for the fundamental mode- and overtone-dominated wavefields. Note that the sensitivity of the fundamental mode rapidly decays with depth, while that of the overtones remains strong in the whole displayed region.

For the first inclusion, we compare inversion results for the fundamental mode (R0, test #1) and the higher modes mixture (R+, test #2), using the averaged PSD as observable. For the second inclusion, we evaluate the influence of the observable on the inverted models by comparing PSD inversion (test #3) to the standard waveform inversion (test #4). Finally, we compare surface wave inversion to P-waves inversion (test #5). In the next section, we describe the observables in more detail.

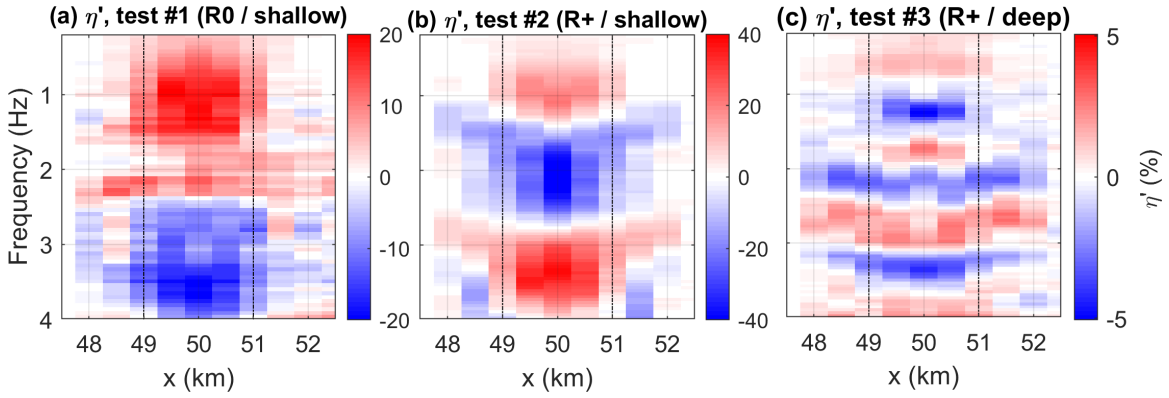


Figure 6.3: SEM-modelled spectral perturbations (η' in percent) for: (a) test #1 (fundamental mode in HM and shallow inclusion); (b) test #2 (overtones in LOH and shallow inclusion); (c) test #3 (overtones in LOH and deep inclusion). The observable is plotted in the distance-frequency-domain. The vertical black dash-dotted lines indicate the lateral position of the inclusion.

6.2.2 Observables

Rayleigh wave spectral amplitude

The Rayleigh wave spectral amplitude is the most important observable and the main novelty of this work. It is estimated by averaging the spectra over multiple random-source realisations, as explained in Section 4.3.3. Examples of wavefield realisations are displayed in Figs 6.4a and b, obtained respectively in the homogeneous and the LOH models. They consist of many arrivals coming from both ends of the model, each corresponding to one particular source among the random distribution. The earliest parts of the seismograms might be dominated by well-separated body waves coming from the early excited remote sources. The green lines in Figs 6.4a and b indicate the latest S-wave arrival from the remotest source, if it was excited at the earliest time allowed by the distribution bounds ($t_0 = 5$ s, see Table 4.2). In the center of the array this arrival time corresponds $t_S - t_0 = d_{max}/V_{s,\min} = 22.5$ s. Here we window the signal after $t_{S,\max} - t_0 > 25$ s to be sure to remove the possibly body-wave dominated part. The kept interval is shaded yellow in Figs 6.4a and b. The Rayleigh wave domination within the kept interval is confirmed by the dispersion curves of the windowed signals, shown in Figs 6.4(c) and (d), respectively for the homogeneous and the LOH models. The non-dispersive wave propagating at $v \approx 2000$ m/s in Fig. 6.4c corresponds to the fundamental mode in the homogeneous model. On the other hand, in Fig. 6.4d, the dispersion curves piece-wisely follow the theoretical curves for the first 20 Rayleigh-wave modes computed for the LOH model using the *Computer Programs for Seismology* package (Herrmann, 2013). Thus we recover the expected difference in modal content due to different background models and source depth, also observed by Bonnefoy-Claudet et al. (2006a), and already explained in Chapter 4.

For imaging, we use the normalised observable $\eta'(\mathbf{x}_r, \mathbf{x}_0, \omega)$ defined for each receiver $\mathbf{x}_r$ with respect to a common reference $\mathbf{x}_0$. The definition is given in equation (4.5). This observable is obtained as the defference between a simulation performed in a medium containing the eslatic inclusion, and a simulation performed in the background model.

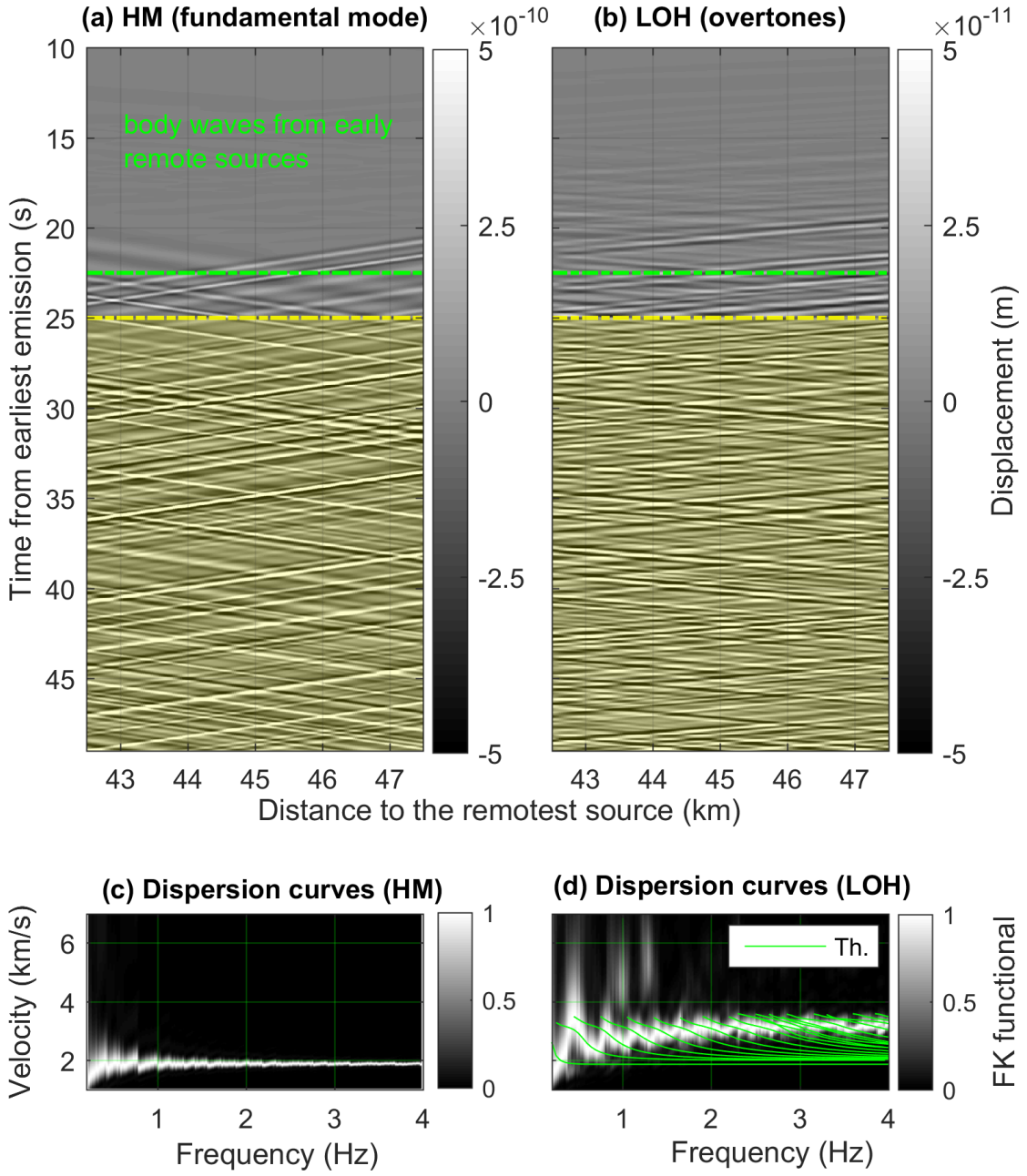


Figure 6.4: (a)-(b) Example of simulated random wavefield realisations in time domain. Averaging the spectra of several such realisations yields the average ⟨PSD⟩. The green line indicates the latest theoretical arrival time of body waves from the earliest and remotest sources. The yellow shadowed zone indicates the time interval used for processing. Simulations are performed in (a) the homogeneous model (dominated by the fundamental mode) and (b) the LOH model (dominated by the overtones). (c)-(d) The corresponding dispersion plots obtained by the FK method (Capon et al., 1967) applied to the yellow time interval. The green curves in (d) indicate the theoretical dispersion curves of the first 20 Rayleigh wave modes for the LOH model.

The modelled η' for tests #1 to #3 is displayed in Fig. 6.3. The reference receiver is the left-most one, located at $x = 47.5$ km, so the observable is zero by definition at this location. For all the three configurations, the strongest perturbations are located vertically above the inclusion, shown by the vertical dash-dotted lines. However, the spectra are also perturbed on both sides of the inclusion. The spectral perturbations consist of alternating positive and negative anomalies. Their shape and amplitude change according to the wave type and the inclusion parameters. Namely, the perturbations generated by the deep thin inclusion, illuminated by overtones (test #3, Fig. 6.3c), are about 10 times weaker than those generated by the shallow thick inclusion (test #2, Fig. 6.3b). The pattern of the observed spectral anomaly is symmetric with respect to the inclusion, mainly because of the two-sided source distribution.

The real-data equivalents of this observable were discussed in Chapter 4. To summarise, if the true reference medium (i.e. the real-life medium without the reservoir) is laterally invariant, we can assume $\eta_0 = 0$ in the real data, and $\eta' = \eta$ corresponds to a spatial PSD anomaly recorded during one single acquisition (exploration context). If the reference medium contains heterogeneities other that the imaging target, $\eta_0 \neq 0$ in the real data, and η' can only be used in case of time-lapse acquisitions, so that the baseline acquisition yields the otherwise unknown η_0. This can serve to retrieve the medium variations, due for example to pumping/injecting fluids into a reservoir (monitoring context).

The other observables we introduce in this paper are also obtained via the difference between two numerical simulations. They mainly serve as benchmark examples to be compared with the new amplitude-based observable.

Rayleigh waveform

In the test #4, the incident wavefield is generated by a single source (superposition of two simultaneous Ricker wavelets of $f_0 = 0.5$ and 2 Hz), and we invert the scattered waveform. Both the incident and the scattered wavefields are shown in Fig. 6.5a and b, respectively, within the inverted time interval ([20-38.5] s). In Fig. 6.5a, the point is to make sure the incident wavefield consists of surface wave arrivals, as we did it for random-source seismogram windowing in section 6.2.2. This time, the number of arrivals is limited because we work with a single source, and the windowing aims to choose a wide enough time-interval to incorporate as much available information as possible. The chosen time interval does not contain any P-arrivals, the latest of which are expected at $t_P - t_0 = (x_r - x_s)/v_{P,\min} = 11.25$ s for $(x_r - x_s) = 45$ km. Still, it could contain the latest S-arrivals, which are expected at $t_S - t_0 = (x_r - x_s)/v_{S,\min} = 22.5$ s. However, nearly at this time, Fig. 6.5a exhibits an arrival with a slope indicating a phase velocity of about 3.5 km/s. This cannot be a direct S-wave propagating into the slowest layer, whose phase velocity would be of 2 km/s. Thus this arrival is a Rayleigh wave overtone, corresponding to the curve portion above 2 Hz in the dispersion plot in Fig. 6.4d, where several overtones contribute to an apparently non-dispersive wave. Subsequent events are also Rayleigh waves overtones, as their late arrival times indicates a group velocity lower than the phase velocity, which is characteristic of dispersive waves. In Fig. 6.5b, the scattered wavefield is dominated

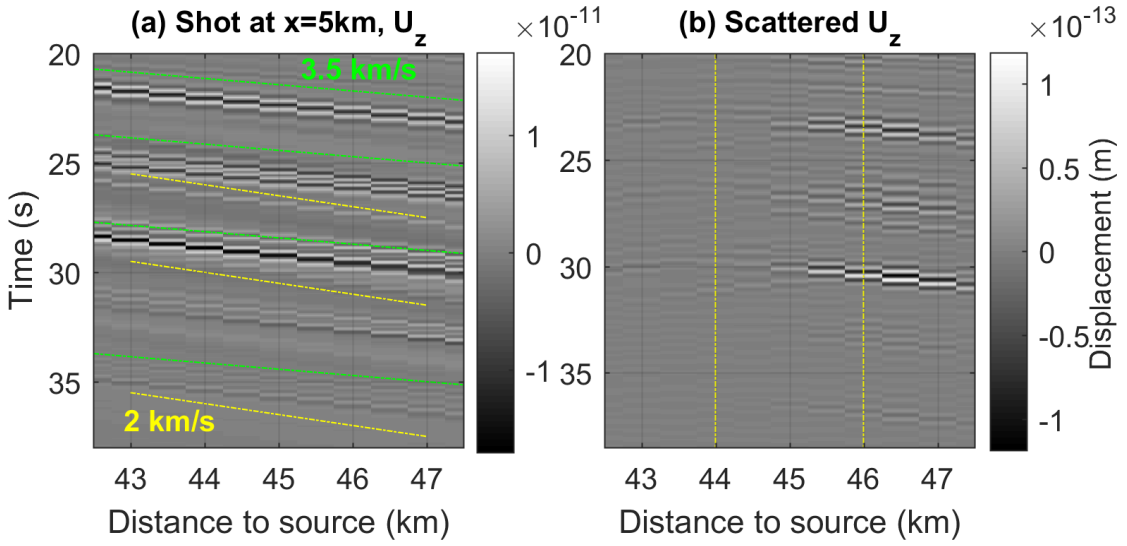


Figure 6.5: Observed time domain signals for test #4. (a) Incident wavefield consisting of Rayleigh overtones. The horizontal axis is given with respect to the source location, and the time axis with respect to the emission time. The slope corresponding to the minimum (resp. maximum) shear velocity in the model, $V_{s,\min} = 2$ km/s (resp. $V_{s,\max} = 3.5$ km/s), is shown with yellow (res. green) lines. Compatibly with Rayleigh overtones, the slopes of all the events lie in between. (b) Scattered wavefield (observable for the inversion). The vertical yellow lines indicate the lateral position of the inclusion.

by forward-scattered waves. Still, looking closely at times around 30 s, we can notice back-scattered waves originating from both ends of the inclusion.

P-waveform

In the test #5, the purpose is to compare surface wave inversion (tests #1 to #4) to the more standard body wave inversion, mostly to check our implementation behaves in the expected way. For generating pure P-waves, the model is extended upwards up to $z = -10$ km, and the free-surface is removed in order to prevent the formation of surface waves. Explosive sources are used instead of vertical point forces. The source functions are the same as in the test #4. Five shots are simulated separately, at $x_s = [45, 47.5, 50, 52.5, 55]$ km, labeled from n°1 to n°5 in this order. The direct P-wave has a zero amplitude on the vertical component for all the receivers. All the motion is horizontal: the corresponding shot gathers for shots n°3 and n°1 are shown in Figs 6.6a and b, respectively. The shot performed above the inclusion (n°3) generates a characteristic reflection move-out (Fig. 6.6c). On the contrary, the shot performed on the left-side of the inclusion (n°1) generates a diffraction event at each edge of the inclusion (Fig. 6.6d). The multiple generated by the interface of the LOH model, though present, was found to produce negligible scattered wavefield after interaction with the inclusion. We discarded it by windowing the signal in the time domain between 3.5 and 8 s.

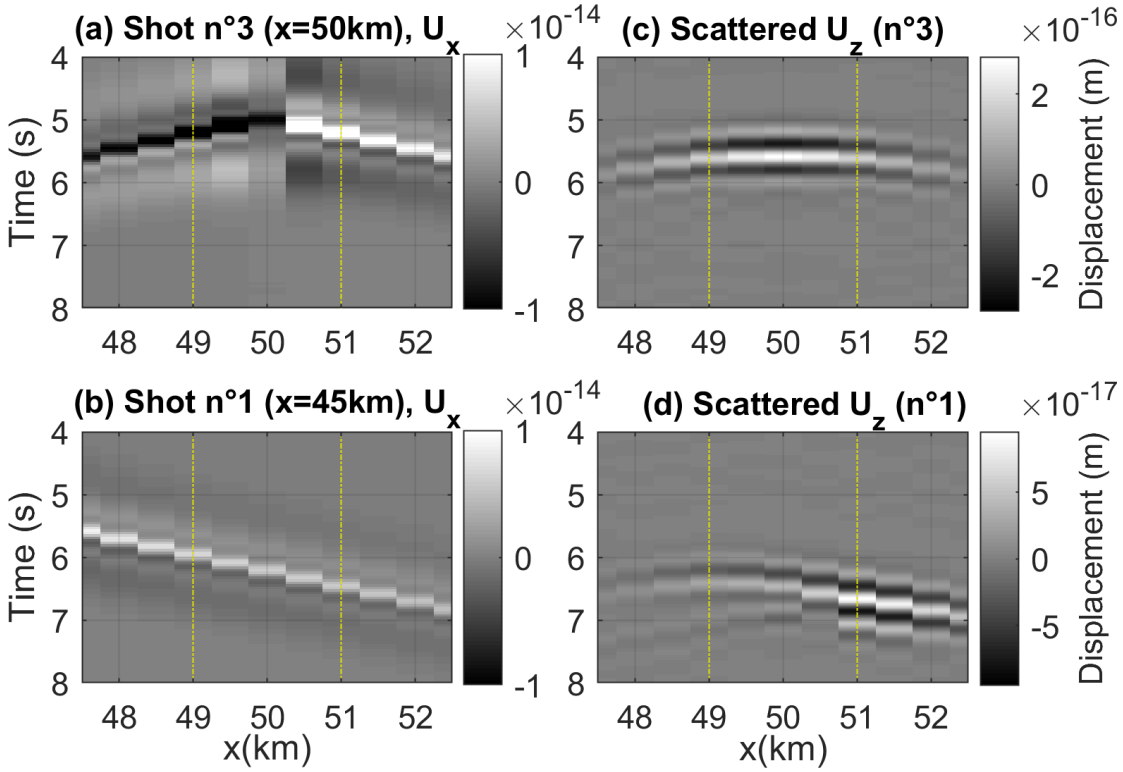


Figure 6.6: Observed time domain signals for test #5. Emission starts at $t_0 = 5$ s. Because of the explosive sources used in the simulation, the incident wavefield is nearly zero on the vertical component. (a) Incident wavefield on the horizontal component for shot n°3 ($x_s = 50$ km) (exactly above the inclusion). (b) Same as (a) for shot n°1 ($x = 45$ km) (on the left from the inclusion). (c) Scattered wavefield on the vertical component corresponding to shot °3. (d) Same as (c) for shot °1. The inclusion lateral position is indicated by the yellow lines.

6.2.3 Inversion

In this part, we define the general framework of our imaging scheme, a least-square iterative migration, which is a linearised version of the full-waveform inversion (FWI) (Jin et al., 1992). The solution model m (a set of elastic parameters as function of space) is found by minimising a least-square objective function of the form

$$J(m) = \frac{1}{2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_f} \left| d_{\text{calc}}(\mathbf{x}_i, \omega_j) - d_{\text{obs}}(\mathbf{x}_i, \omega_j) \right|^2 , \tag{6.1}$$

with d_{obs} a possibly complex-valued observable (for waveform observables), d_{calc} its prediction by a forward-modelling scheme within the model m_0, $\mathbf{x}_i$ the positions of the receivers, N_r the total number of receivers, ω_j the angular frequencies, and N_f the total number of inverted frequencies. For simplifying notations, we omit the additional summation over sources which arises when data from several independent shots are inverted simultaneously. In our case, the purpose is to retrieve a small perturbation δm of the reference medium m_0, acting as a scatterer. As we are using only synthetic data in this paper, d_{obs} is obtained by *numerical modelling*. In all our examples, d_{obs} is related to the scattered wavefield, i.e. the difference between two numerical simulations, one performed in the background model m_0, and another one performed in the complete model $m = m_0 + \delta m$ containing the inclusion which we wish to image. Based on Section 6.2.2, we consider two different types of observables, the standard observed data (waveform) and the amplitude spectrum. We refer to the corresponding objective functions as J_{LSM} and J_{PSD}, respectively. The gradient of J with respect to the model is related to the gradient of the observable as

$$\frac{\partial J}{\partial m} = \mathfrak{Re} \left\langle \frac{\partial \mathbf{d}_{\text{calc}}}{\partial m}, \mathbf{d}_{\text{calc}} - \mathbf{d}_{\text{obs}} \right\rangle , \tag{6.2}$$

with $\langle \cdot, \cdot \rangle$ the Hermitian scalar product in $\mathbb{C}^{N_f \cdot N_r}$, $\mathfrak{Re}$ the real part, and $\mathbf{d}_{\text{calc}}$, $\mathbf{d}_{\text{obs}}$ the $[1, N_f \cdot N_r]$ vectors corresponding to the discretised d_{calc}, d_{obs}. The gradient of the observable is the Fréchet derivative (or sensitivity kernel) derived in Chapter 4 for both the waveform and the amplitude spectrum (see equations 4.A.5, 4.A.8 and 4.A.9). A necessery condition for a minimum of J to be reached is that the gradient must be zero. Discretizing the integration domain, this leads to the normal equation

$$\mathfrak{Re} \left(\mathbf{K}^\dagger \mathbf{K} \right) \mathbf{\Delta m} = \mathfrak{Re} \left(\mathbf{K}^\dagger \mathbf{d}_{\text{obs}} \right) , \tag{6.3}$$

with $\mathbf{\Delta m}$ the discretised model perturbation vector of dimension $[1, N_p \cdot N_m]$ (N_m = number of model cells, N_p = number of medium parameters), $\mathbf{K}$ the discretised sensitivity kernel of size $[N_f \cdot N_r, N_p \cdot N_m]$ and $\mathbf{K}^\dagger$ the transposed conjugate of $\mathbf{K}$. The right-hand-side in equation (6.3) is the gradient of J in the background model m_0. To first order in the wavefield perturbation, the matrix on the left-hand-side of equation (6.3) is the Hessian of J:

$$\mathbf{H} = \mathfrak{Re} \left(\mathbf{K}^\dagger \mathbf{K} \right) . \tag{6.4}$$

This formulation of the problem is referred to as the Gauss-Newton method (see Virieux & Operto (2009) for an overview of different possible formulations). Computing $\mathbf{H}^{-1}$ represents too much computational cost, so we prefer using an iterative

method to solve the linear equation (6.3), namely the Conjugate Linear Gradient algorithm (CLG, Hestenes & Stiefel, 1952). In the literature, the Hessian matrix $\mathbf{H}$ is usually too heavy even for being stored as a whole, which motivates the use of the matrix-free algorithms, such as the adjoint-state method (Plessix, 2006). In our case, we stay in 2D and work at low frequencies ([0.2-4] Hz), which allows us to discretise the model into cells as large as 25x25 m. We manage to store the sensitivity kernel $\mathbf{K}$ and directly estimate $\mathbf{H}$ as the matrix product in equation (6.4). This yields a square real matrix of size $[N_p \cdot N_m, N_p \cdot N_m]$. With an investigated zone spanning 4000x1500 m, this means $N_m = 9.8 \cdot 10^3$ cells for the whole model. Inverting for three elastic parameters ($N_p = 3$) yields a Hessian matrix of size $[28.8 \cdot 10^3,\ 28.8 \cdot 10^3]$. With double precision, such a matrix occupies 6.6 GB of memory, which remains affordable with a powerful machine. Since we are able to use the standard CLG with a matrix formulation, we do not need to perform any new simulations in the reference ground model m_0, and the gradient of J at the iteration n is simply given by $\nabla_n J = \nabla_0 J + \mathbf{H}\boldsymbol{\Delta}\mathbf{m}^{(n)}$, with $\boldsymbol{\Delta}\mathbf{m}^{(n)}$ the model update at iteration n. The difference with standard FWI is that in our case the gradient is always estimated in the background model.

The scattered wavefield can also be expressed through other parameter combinations (or parametrisations), such as (ρ, V_p, V_s). The expressions of the sensitivity kernels associated with all the parametrisations we study in this paper are summarised in Appendix 6.A. The next section focuses on the choice of the best parametrisation. The parametrisation choice is guided by the detection objective, i.e.e correctly retrieving the inclusion position and shape, rather than the precise values of the parameters (more relevant for the LFPS application). Whatever the parametrisation, the results are given in terms of the reconstructed $\delta\rho/\rho_0$, $\delta\mu/\mu_0$ and $\delta\lambda/\lambda_0$ models.

6.3 Multi-parameter inversion: choosing the best parametrisation

A key element in the inversion is the Hessian matrix $\mathbf{H}$ (see equation (6.3)). Analysing its properties provides some useful insights into the feasibility of the multi-parameter inversion and the optimal parametrisation choice (Virieux & Operto, 2009; Operto et al., 2013; Plessix & Cao, 2011). Taking the example of tests #2 and #3, both involving Rayleigh wave overtones and the PSD observable, the Hessian matrix under the (ρ, λ, μ)-parametrisation is shown in Fig. 6.7a, and in Fig. 6.7b under the (ρ', V_p, V_s)-parametrisation. The notation ρ' is used in this section referring to the density under all parametrisations other than (ρ, λ, μ). This is a 3x3 block matrix, because three parameters are involved. Each block is a square matrix of size $[N_m, N_m]$. Each of its elements represents the interaction between two model cells. As $N_m = N_x \cdot N_z$, each row or column of a block corresponds to a concatenation of the N_x columns of the physical model. If one moves either horizontally or vertically inside the Hessian, there is a column skip every N_z cells. This is why each of the 9 big blocks is further subdivided into elementary square sub-blocks of size $[N_z, N_z]$. The big blocks $\mathbf{H}_{m,m}$ (those of size $[N_m, N_m]$) which lie on the diagonal represent the interaction between the cells associated with the same parameter m_0. The ideal situation would

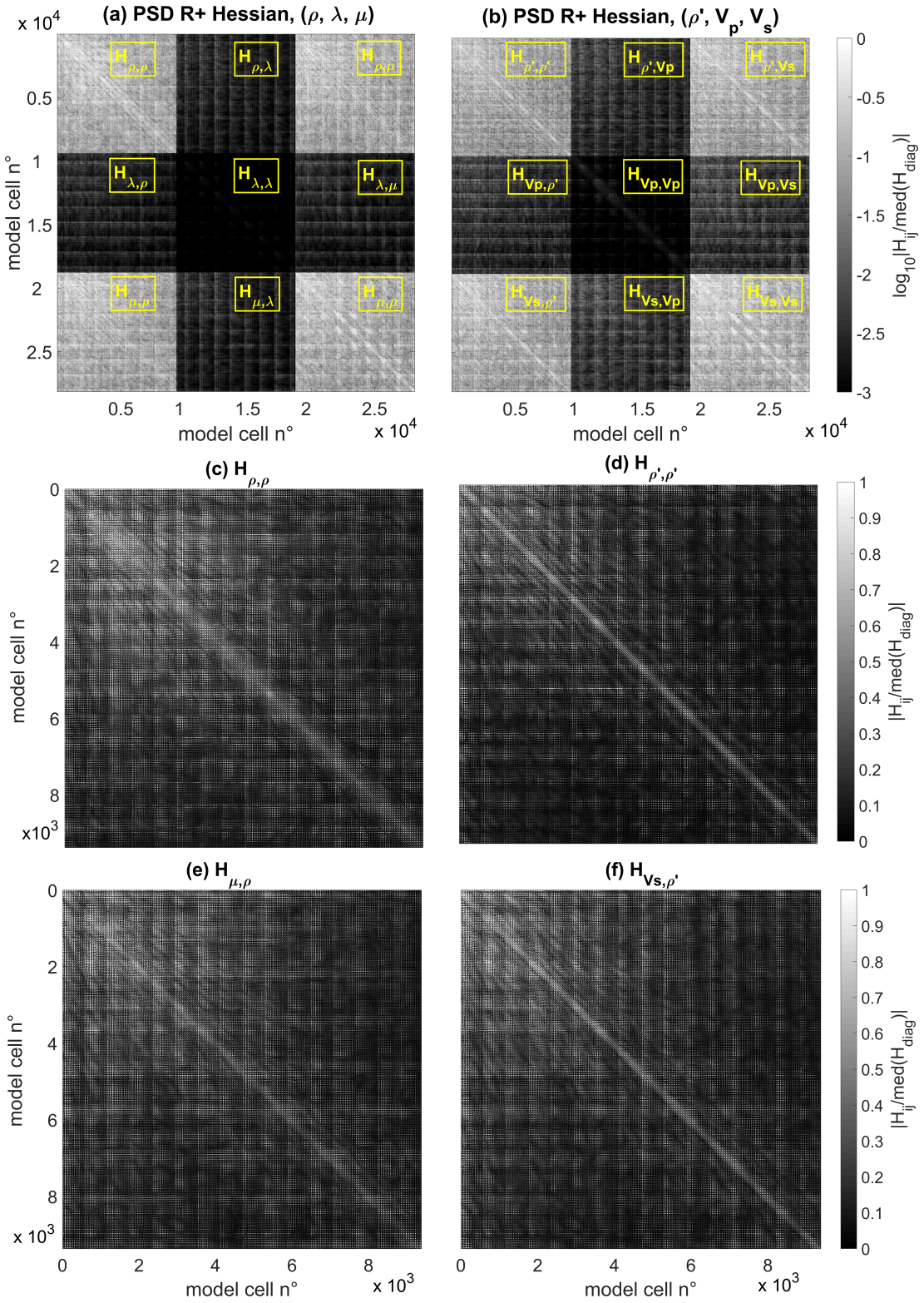


Figure 6.7: Hessian matrix for the overtone-dominated wavefield with the averaged PSD observable. In a and b the matrix elements are normalised to the median value of the diagonal elements and shown with a logarithmic scale. (a) (ρ, λ, μ)-parametrisation. (b) (ρ', V_p, V_s)-parametrisation. (c)-(f) Zoom on particular blocks shown with a linear scale.

be to have purely diagonal $\mathbf{H}_{m,m}$: in this case there would be no ambiguity between different cells. The ambiguity is carried by the off-diagonal elements, which represent the *spatial crosstalk*. The off-diagonal blocks $\mathbf{H}_{m_i,m_j}(i \neq j)$, in turn, carry both *spatial* and *inter-parameter* crosstalks. Their diagonal elements (k, k) represent the pure inter-parameter crosstalk. Choosing the best parametrisation implies a trade-off between spatial and inter-parameter crosstalk. We can observe that the diagonal block $\mathbf{H}_{\rho',\rho'}$ (Fig. 6.7d) for the (ρ', V_p, V_s)-parametrisation is more diagonally-dominant than the block $\mathbf{H}_{\rho,\rho}$ (Fig. 6.7c) for the (ρ, λ, μ)-parametrisation, implying a weaker spatial crosstalk for (ρ', V_p, V_s). However, the diagonal of the block $\mathbf{H}_{\mu,\rho}$ (Fig. 6.7e) is less pronounced for the (ρ, λ, μ)-parametrisation, compared to $\mathbf{H}_{V_s,\rho'}$ (Fig. 6.7f) for the (ρ', V_p, V_s)-parametrisation, implying a weaker inter-parameter crosstalk for (ρ, λ, μ). If the main objective is to correctly locate the inclusion, the best parametrisation is the one which minimizes the spatial crosstalk (here, (ρ', V_p, V_s)). This is the point of view we adopt in our work. However, if the location of the inclusion is known, and one wishes to accurately estimate the elastic parameters inside it, one should prefer the parametrisation which minimizes the inter-parameter crosstalk (here, (ρ, λ, μ)). Another key-point for a successful multi-parameter inversion is that the direct problem should present enough sensitivity for the observable to be recorded. In the same time, the sensitivities for each parameter should have comparable magnitudes to avoid some parameters overshadowing others during the inversion. It is obvious from Fig. 6.7 that the sensitivities with respect to both λ and V_p are extremely low compared respectively to $(\rho,\ \mu)$ and (ρ', V_s), meaning that the inverse problem is poorly conditioned in case of a 3-parameter inversion. To quantify all the described features of the Hessian, we propose simple scalar criteria in Appendix 6.B. These criteria are C_I for inter-parameter crosstalk, C_S for spatial crosstalk, C_κ for conditioning and S for the absolute sensitivity. Both C_κ and C_I are based on the diagonals of the Hessian blocks. Initially defined for each model cell, they are averaged over the area covered by the inclusion. C_κ measures the ratio between the weakest and the strongest radiation produced by the different elastic parameters per unit relative perturbation, in a given cell. C_I measures the mutual orthogonality between the radiation patterns associated to different elastic parameters involved in the parametrisation. Thus C_I is defined by only one value for 2-parameter inversions, and by three different values for 3-parameter inversions. C_S measures the amplitude remaining in the inclusion after the action of the Hessian on a model vector which takes the value 1 inside the inclusion and 0 elsewhere. It is based on the concept of the *point spread function* (PSF, see Pan, 2017, for a review). C_S is defined for each parameter individually, so it takes 2 or 3 values respectively for 2- and 3-parameter inversions. The sensitivity S corresponds to an average ratio $u^{(1)}/u^{(0)}$ between the scattered and the incident wavefields, per unit surface, normalised by 1 m^{-2}. Large values of the criteria always indicate a better behaviour of the Hessian. This allows to compare different parametrisations for each of the five test cases listed in Table 6.1. For tests #1 and #2, we consider only 2-parameter inversions: (ρ, μ), (ρ', V_s), (ρ', I_s) and (V_s, I_s), with $I_s = \rho V_s$ the S-wave impedance. For tests #3 to #5, we also study the effect of adding a third parameter: (ρ, μ, λ), (ρ', V_s, V_p), (ρ', I_s, I_p) and (V_s, I_s, λ).
In Figs 6.8(a) to (e), all the criteria defined in Appendix 6.B are shown for tests #1 and #2. For example, comparing first and second columns in Figs 6.8(c) and (d), we conclude that the (ρ', V_s)-parametrisation presents a lower spatial crosstalk compared

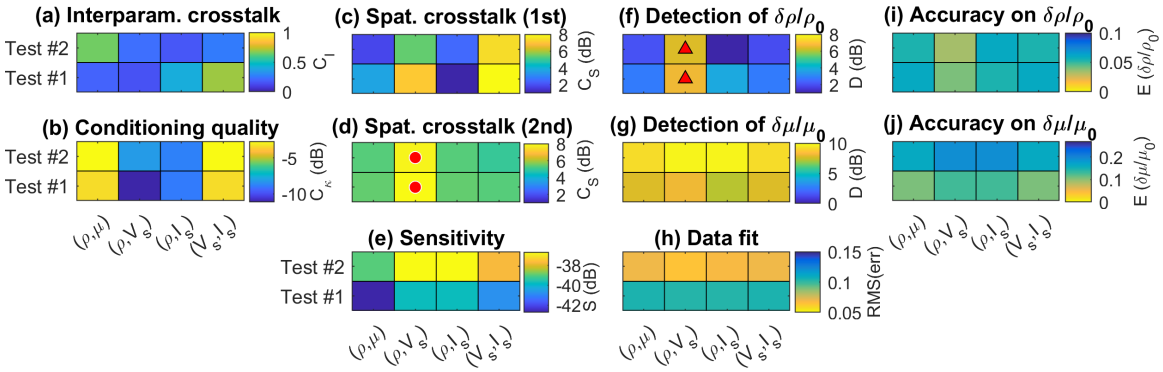


Figure 6.8: (a)-(e): a priori quality criteria for several 2-parameter inversions in the case of the thick shallow inclusion (tests #1 and #2). Yellow colors always correspond to preferable values of the criteria. C_S - spatial crosstalk criterion for the different parameters (1st, 2nd), in the same order as listed in the parametrisation labels; C_I - inter-parameter crosstalk criterion (just one value for 2-parameter inversions, see Appendix 6.B); C_κ - conditioning quality; S - sensitivity. (f)-(i): a posteriori inversion quality criteria for the same cases: D - inclusion detection; E - absolute error on the parameter relative perturbation estimation inside the inclusion. The maximum value of the color scale corresponds to the true value of the perturbation. The red dots indicate the preferable parametrisations in the panels which were deciding for the parametrisation choice. The best parametrisations (resp. inverted models) are marked by red dots (resp. triangles) in the discriminating cells.

to (ρ, μ) ($[C_s^{(\rho)}, C_s^{(Vs)}] = [7.1, 9.1]$ against $[C_s^{(\rho)}, C_s^{(\mu)}] = [3.7, 5.5]$ for test #1, and $[C_s^{(\rho)}, C_s^{(Vs)}] = [5.4, 7.9]$ against $[C_s^{(\rho)}, C_s^{(\mu)}] = [1.8, 5.3]$ for test #2), in line with the visual analysis of Fig. 6.7. The discriminating cells allowing to choose the (ρ', V_s)-parametrisation among all the others are spotted with red dots in Fig. 6.8d. On the other hand, the (ρ', V_s)-parametrisation is the worst in terms of the inter-parameter crosstalk (Fig. 6.8a), also confirming the visual analysis of Fig. 6.7 (strong diagonal in the off-diagonal block $\mathbf{H}_{\rho', V_s}$). In order to determine the relevance of the different criteria for the parametrisation choice, we invert the data (see Section 6.4 for details) under all the considered parametrisations, and use another two criteria to quantify the quality of the obtained models: D for the accurate location of the inclusion detection, and E for the absolute error on the parameter relative perturbation inside the inclusion. Those are defined in Appendix 6.C. For tests #1 and #2, they are shown in Figs 6.8(f) and (g). A lower E corresponds to smaller errors, so E is represented with a reversed color scale to ease the visual analysis. It appears that (ρ', V_s) yields the best results in terms of correct detection (Fig. 6.8f, red triangles), while all the parametrisations fit the data equally well (Fig. 6.8h). The accuracy of ρ and μ estimation is roughly the same for all the parametrisations (Figs 6.8(i) and (j)). Thus, in overall, the (ρ', V_s)-parametrisation performs the best for both tests #1 and #2. However, both the conditioning and the inter-parameter corsstalk analysis would disadvantage this parametrisation, while the sensitivity analysis would not be able to discriminate between (ρ', V_s) and (ρ', I_s). Hence we suggest that the spatial crosstalk criterion C_S is the most relevant for the parametrisation choice in our case. For this reason, when presenting results of the Hessian analysis for tests #3 to #5, we only show this criterion. The corresponding results are displayed along with the

inversion quality criteria in Fig. 6.9 for 2-parameter inversions, and in Fig. 6.10 for 3-parameter inversions.

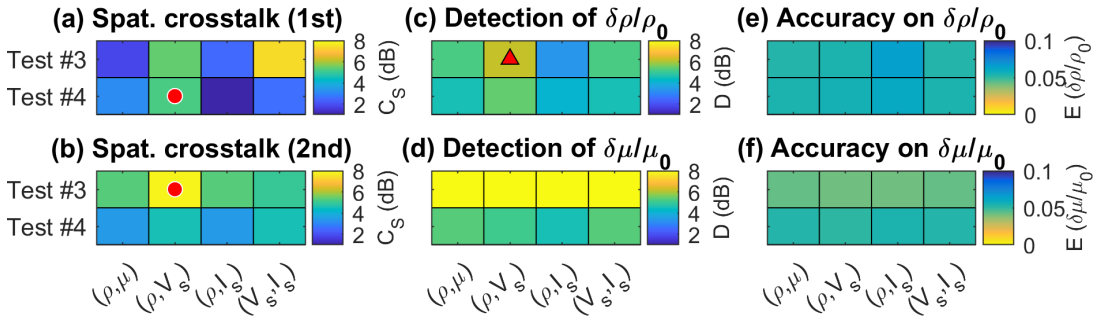


Figure 6.9: Same as Fig. 6.8 (only the spatial crosstalk criterion C_S is shown), but for the deep thin inclusion (tests #3 to #5) and 2-parameter inversions.

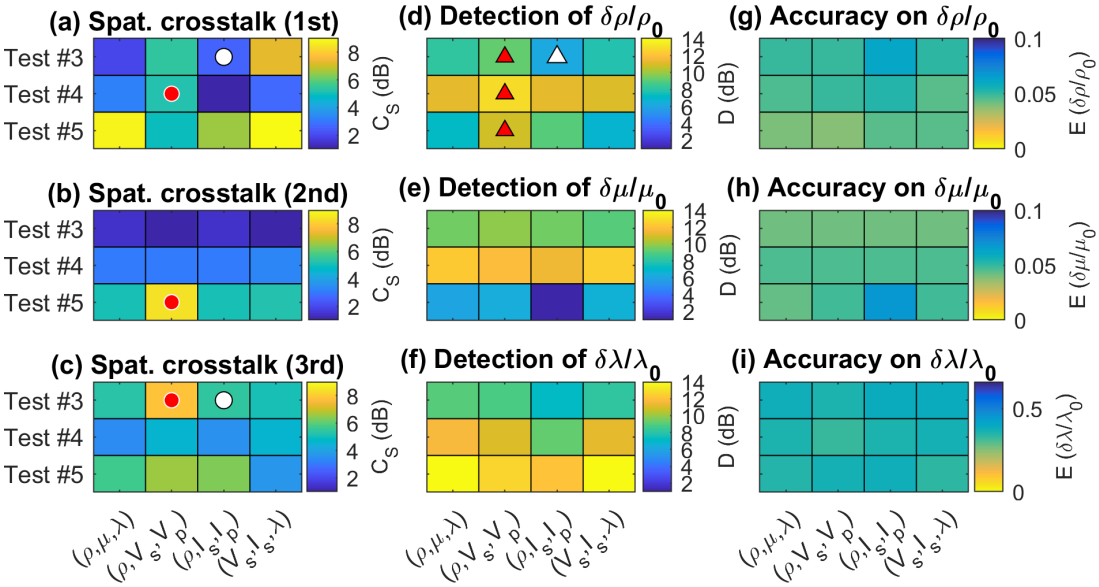


Figure 6.10: Same as Fig. 6.8 (only the spatial crosstalk criterion C_S is shown), but for the deep thin inclusion (tests #3 to #5) and 3-parameter inversions. The white dot (resp. triangle) indicates the cell where the «bad» parametrisation (resp. inversion) behaves clearly worse than the «good» one. The corresponding inverted models are compared in section 6.4.

Despite a stronger ambiguity between different parametrisations (compared to tests #1 and #2), we choose (ρ', V_s) and (ρ', V_s, V_p) as the most appropriate parametrisations, respectively for 2-parameter and 3-parameter inversions, for all the tests. Discriminating, for example, between (ρ', V_s, V_p) and (V_s, I_s, λ) for test #3, based on the spatial crosstalk, is very subjective (Figs 6.10(a) and (c)). However, it is clear from the Hessian analysis that the (ρ, I_s, I_p)-parametrisation should be avoided because it presents a significantly larger spatial crosstalk on both 1st and 3rd parameters (white dots in Figs 6.10(a) and (c)). For this parametrisation, the inversion results are indeed worse than for the (ρ', V_s, V_p)-parametrisation (Fig 6.10d, compare cells containing white and red triangles in the row labeled «test #3»). The corresponding

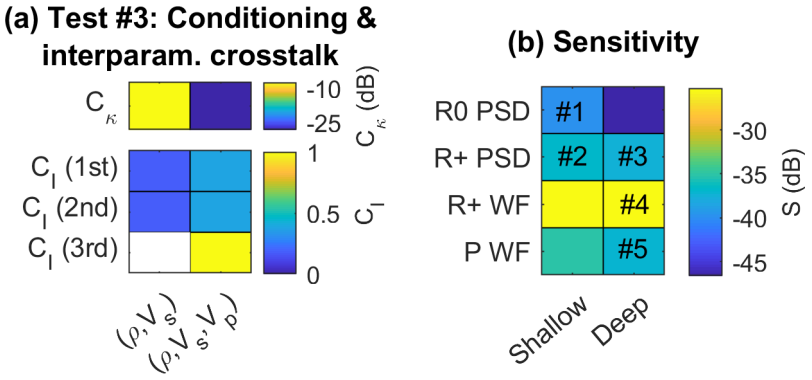


Figure 6.11: (a) Conditioning quality C_κ and inter-parameter crosstalk criterion C_I compared between two- and 3-parameter inversions for test #3 (overtones with PSD observable). (b) Absolute sensitivity S compared for several combinations of incident wavefield, observable, and inclusion type («shallow» or «deep», see Table 6.1). The configurations corresponding to the tests described in Table 6.1 are labeled accordingly. Results are displayed for the (ρ', V_s, V_p)-parametrisation.

inverted ρ models are compared in section 6.4.

Now that the parametrisation has been chosen, another important question is to know whether a two- or 3-parameter inversion should be preferred. Taking the example of the test #3, we compare the corresponding conditioning and inter-parameter crosstalk criteria in Fig. 6.11a. The conditioning criterion worsens dramatically (by more than 15 dB) for the 3-parameter inversion. However, the crosstalk between the third parameter V_p and the two original ones (ρ', V_s) is almost zero ($C_I(V_p) \approx 1$). Testing both two- and 3-parameter inversions reveals that including V_p improves the results (compare color scales of the detection criterion in Figs 6.9 and 6.10). This means that the presence of a V_p-contrast generates effects on the spectra which cannot be explained by a combination of ρ-and V_s-contrasts, despite the low average sensitivity with respect to V_p. That is, V_p could act as an additional parameter needed to explain the data, but for which we do not expect to obtain a meaningful model. This is for example the case for density is some acoustic inversions (Forgoes & Lambaré, 1997).

We also compare the absolute sensitivity of the different tests in Fig 6.11b. Results are displayed for the (ρ', V_s, V_p)-parametrisation for all the cases, but are identical to within 0.1 dB to those obtained with the (ρ', V_s)-parametrisation. First, they confirm the drop of the fundamental mode sensitivity for the deep inclusion, which is the reason why we do not study this case in our work. Second, they reveal that the sensitivity of the overtones with the PSD observable is comparable to that of the active-source P-waves. Finally, the sensitivity of the overtones with the waveform observable is more that 10 dB above its PSD counterpart, meaning that a considerable amount of information is lost when the phase is not available.

Table 6.2: Number of inverted frequencies for different tests.

Test	N_f
#1	72
#2	95
#3	95
#4	288
#5	45

6.4 Inversion results

In this section, we present the inversion procedure and the resulting models. By comparing different tests, we analyse the influence of both the wavefield modal compostion and the definition of the observable. Finally, we apply our linearised inversion to P-waves as a validation step.

We use the Conjugate Linear Gradient algorithm (CLG, Hestenes & Stiefel, 1952) to find a least-square solution of a regularised version of the normal equation (6.3). We do not use any preconditioning. The maximum number of iterations is set to 1000. The iterations stop when $\widetilde{J}$ stops decreasing. As a criterion, we compute the logarithmic slope of $\widetilde{J}/\widetilde{J}_0$ with a linear regression over the last ten iterations, and stop iterating when its value exceeds -0.04 ($\widetilde{J}_0$ refers to the initial model). All the frequencies, linearly sampled between 0.2 and 4 Hz, are inverted simultaneously. The number of inverted frequencies varies from one test to another, depending on the length of the time window used to estimate the spectral amplitude (tests #1, #2, #3) or the scattered waveform in the frequency-domain (tests #4, #5). These numbers are summarised in Table 6.2. Since we use large time windows containing all the relevant arrivals for each test, the more or less important number of inverted frequencies represents redundancy rather than additional constraint, and should not considerably influence the inversion.

6.4.1 Regularisation

To avoid overfitting, the misfit function is modified to penalise models which contain too strong parameter variations (Tikhonov regularisation):

$$\widetilde{J}(m) = J(m) + \frac{1}{2}\epsilon \left|\left| R(m) \right|\right|^2 , \tag{6.5}$$

where R is the regularisation operator and ϵ the regularisation coefficient chosen via the L-curve inspection (Hansen & O'Leary, 1993). The most simple regularisation operator is the identity: $R(m) = m$. However, in order to better remove the oscillatory solutions, we prefer using the Laplace regularisation: $R = L$. The action of L on m can be seen as a 2D convolution by a 3x3 kernel:

$$L(m) = m \star \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \cdot \frac{1}{(\Delta l)^2} , \tag{6.6}$$

with Δl the size of the square cells (25 m in our case). The matrix $\mathbf{L}$ representing the action of L is of size $[N_m \cdot N_p, N_m \cdot N_p]$ (same as the Hessian), each of its blocks representing the Laplace operator action on the model associated to one parameter. Each block can be compactly written as

$$L_{ij} = (D_i \delta_{ij} - A_{ij}) \cdot \frac{1}{(\Delta l)^2} \ , \tag{6.7}$$

with D_i the number of neighbours of the cell i (2 in the corners, 3 on the edges, and 4 for inner cells), and $\mathbf{A}$ the adjacency matrix (taking the value 1 for cells sharing and edge, and 0 otherwise).

Repeating the same steps as in section 6.2.3, but under the constraint $\partial \widetilde{J}/\partial m = 0$, leads to the following modification of the normal equation (6.3):

$$\left(\mathbf{H} + \epsilon \mathbf{L}^\dagger \mathbf{L}\right) \mathbf{\Delta m} = -\nabla_0 J \ . \tag{6.8}$$

Since the magnitude of $\mathbf{H}$ can vary according to the incident wave type and the chosen observable, we scale the above equation by the median value h_{med} of the Hessian diagonal. The median of the diagonal of $\mathbf{L}^\dagger\mathbf{L}$, on the other hand, is $l_{med} = 20/(\Delta l)^2$. Thus we get the following scaled equation:

$$\left(\widetilde{\mathbf{H}} + \widetilde{\epsilon}\widetilde{\mathbf{L}^\dagger\mathbf{L}}\right) \mathbf{\Delta m} = -\frac{1}{h_{\text{med}}} \nabla_0 J \ , \tag{6.9}$$

with $\widetilde{\mathbf{H}} = \mathbf{H}/h_{\text{med}}$, $\widetilde{\mathbf{L}^\dagger\mathbf{L}} = \mathbf{L}^\dagger\mathbf{L}/l_{\text{med}}$ and $\widetilde{\epsilon} = \epsilon \cdot l_{\text{med}}/h_{\text{med}}$. The scaled $\widetilde{\epsilon}$ measures the relative weight of the regularisation term with respect to the original Hessian. The values of $\widetilde{\epsilon}$ are tested on a log-scale, ranging from 1 to 10^7, to construct the L-curves for the 5 tests presented in Table 6.1. Such high values highlight the ill-posedness of the problem. In order to construct the L-curve, a zero-mean gaussian noise of standard deviation $\sigma = 0.1\sqrt{\langle d^2_{\text{obs}} \rangle}$ is added to $\mathbf{d}_{\text{obs}}$ in equation (6.3) in order to get a threshold beneath which it is useless to reduce the RMS misfit. All the inversions are performed using the same convergence criterion, described in the beginning of Section 6.4. The resulting L-curves, which display the normalised RMS misfit $\sqrt{||\mathbf{K}\mathbf{\Delta m} - \mathbf{d}_{\text{obs}}||^2/||\mathbf{d}_{\text{obs}}||^2}$ with respect to the regularisation term $||\mathbf{L}\mathbf{\Delta m}||^2$, are shown in Fig. 6.12 for the 5 proposed tests. Two- and 3-parameter inversions are studied separately in Figs 6.12a and b, respectively. The chosen regularisation coefficients are summarised in Table 6.3. Since several points can be reasonably considered as corners, we choose the one for which the normalised RMS misfit lies the closest to the noise threshold (0.1). For each test, the L-curve was constructed under the parametrisation recommended in Section 6.3 based on the Hessian analysis. On the other hand, the invesions quality criteria presented in Figs 6.8, 6.9 and 6.10 were all obtained with the same regularisation parameter.

6.4.2 Fundamental mode vs. overtones

Let us first consider tests #1 and #2, i.e. those involving a thick, shallow and strongly contrasting inclusion (containing a non-zero $\delta V_s/V_s$), illuminated respectively with the fundamental mode and the overtones. This inclusion is still within the penetration

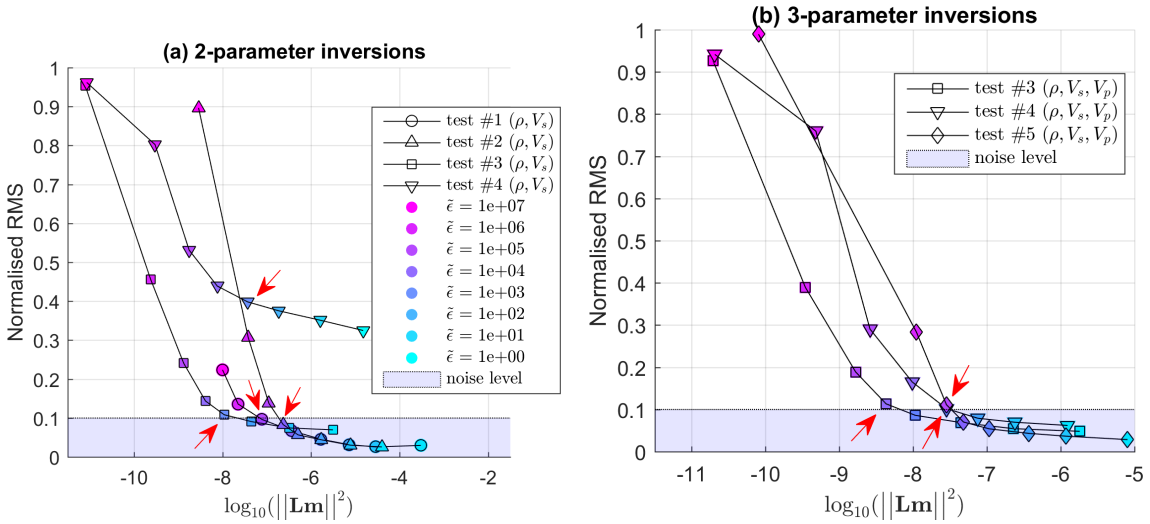


Figure 6.12: L-curves constructed for the different tests in Table 6.1 and $\widetilde{\epsilon}$ ranging from 1 to 10^7, indicated by the marker colors. (a) 2-parameter inversions; (b) 3-parameter inversions. The different tests differ by the marker shapes. The shaded area correponds to the level of the added gaussian noise, beneath which the RMS reduction is due to ovefitting. All the inversions have been performed with the same convergence criterion used to stop the iterations (see beginning of Section 6.4). The chosen regularisation coefficients are spotted with the red arrows.

Table 6.3: Summary of the chosen regularisation coefficients $\widetilde{\epsilon}$, for different tests, with two- and 3-parameter parametrisations.

Test	(ρ', V_s)	(ρ', V_s, V_p)
#1	10^5	-
#2	10^4	-
#3	10^3	10^3
#4	10^3	10^3
#5	-	10^5

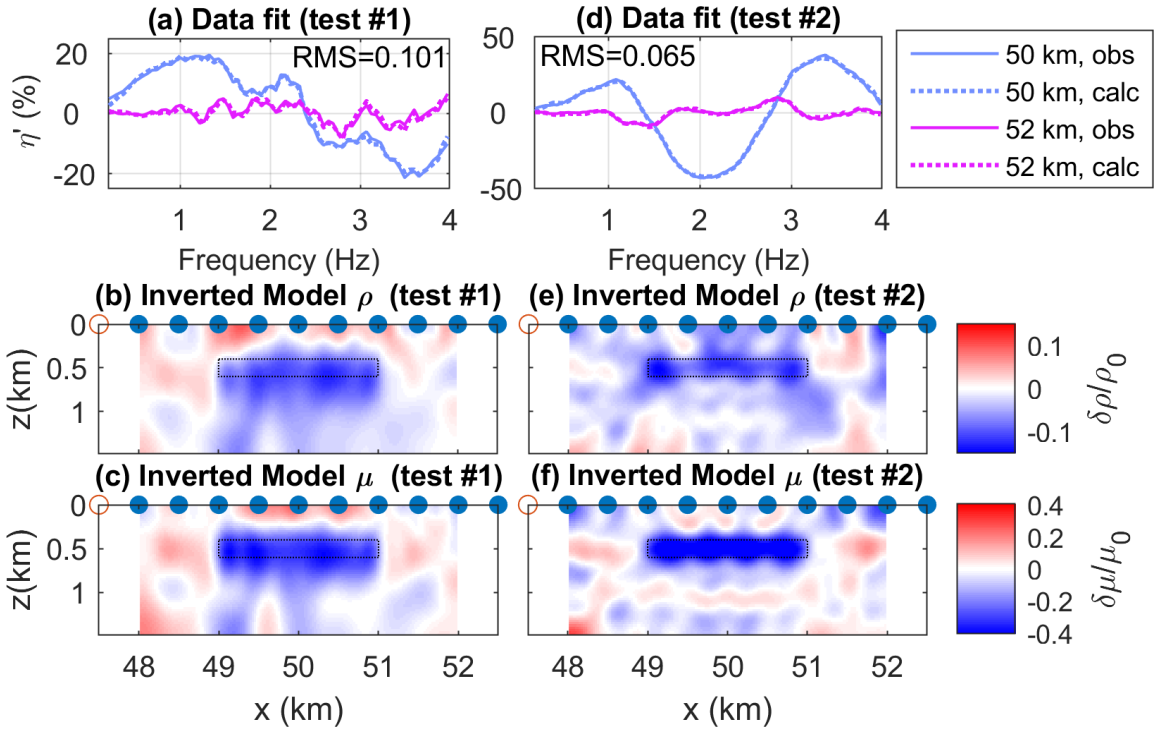


Figure 6.13: Left side: inversion results for test #1 (fundamental mode in HM) under the parametrisation (ρ', V_s). (a) Spectral perturbation η' fit at two receivers for the final model; (b) Final inverted model for density perturbations; (c) Final inverted model for μ perturbations. The receivers are shown with circles. The reference receiver is transparent. Right side, (d)-(f): same as (a)-(c) for test #2 (overtones in LOH). The correct model perturbations are: $\delta\rho/\rho_0 = -0.1$, $\delta\mu/\mu_0 = -0.27$.

range of the fundamental mode (see Fig. 6.11b), allowing to compare results obtained with these two types of wavefields. The observables generated by such an inclusion are shown in Figs 6.3a and b, respectively for the fundamental mode and the overtones. Inversion results are shown in Fig. 6.13. In both cases, the observed data are quite well explained (Figs 6.13(a) and (d), compare solid vs. dash-dotted lines). The ρ-model (Figs 6.13(b) and (e)) is less successfully reconstructed compared to the μ-model (Figs 6.13(c) and (f)), as a result of a poor conditioning (see Fig. 6.8b). Both wavefields have the same accuracy for $\delta\rho/\rho_0$ estimation (Fig. 6.8i, where the maximum of the colormap correponds to 100% error with respect to the true model perturbation). The accuracy of the estimated $\delta\mu/\mu_0$ perturbation is slightly better for the fundamental mode wavefield (test #1, Fig. 6.8j). In particular, the strength of the $\delta\mu/\mu_0$ perturbation resulting from the overtones inversion is overestimated (saturated blue colour within the rectangle in Fig. 6.13f). However, the μ-model is sharper retrieved with the overtone illumination. Also, the observed perturbation amplitude is higher for the overtones because of their stronger sensitivity (see Fig. 6.11b). Hence, at least in the exploration context (inclusion detection objective), Rayleigh wave overtones appear as a more advantageous incident wavefied based on the comparison of tests #1 and #2, even for a shallow inclusion.

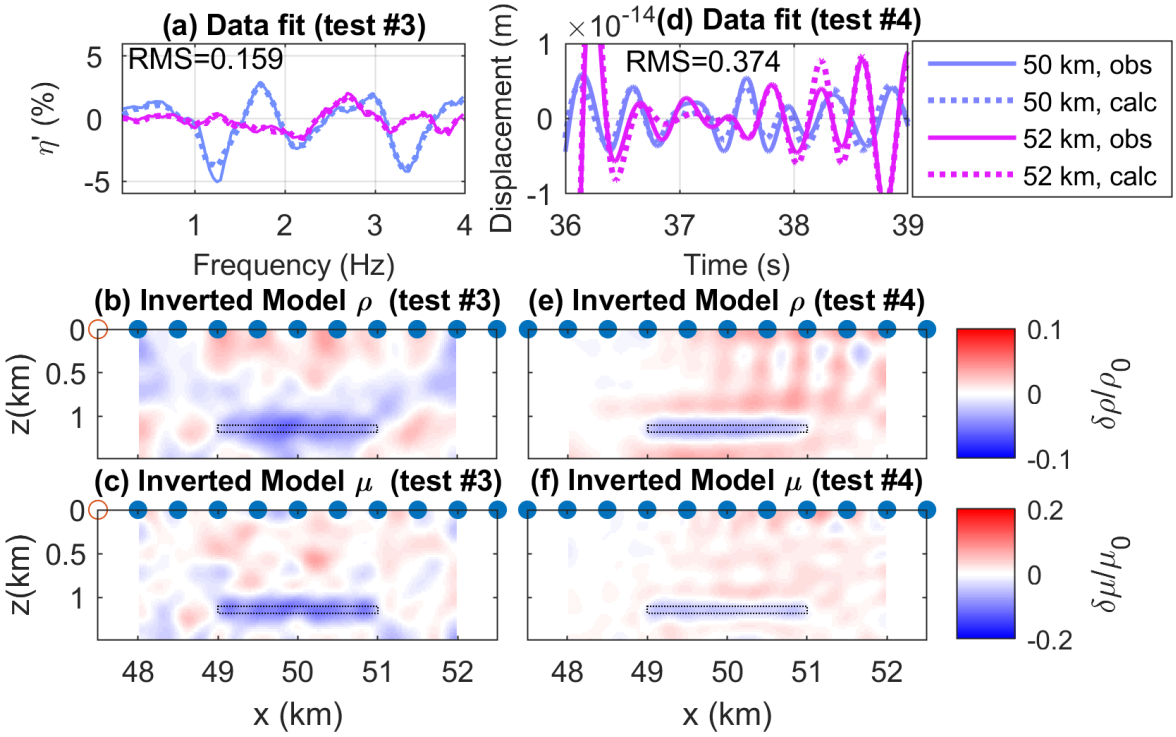


Figure 6.14: Same as Fig. 6.13, but comparing tests #3 (overtones in LOH with PSD observable) and #4 (overtones in LOH with standard observable), both with (ρ', V_s)-parametrisation. (d) displays the scattered waveform fit at two receivers, zoomed in a 3-second interval within the inverted wave train, for the final model. The correct model perturbations are: $\delta\rho/\rho_0 = -0.1$, $\delta\mu/\mu_0 = -0.1$.

6.4.3 Standard vs. PSD observable

Secondly, comparing tests #3 and #4 allows to study the influence of the PSD observable compared to the standard full waveform observable, for an overtone-dominated wavefield. In the same time, we move to the case of the more realistic deep and thin inclusion containing no V_s perturbation. The corresponding observables are shown in Figs 6.3(c) and 6.5(b), respectively for the PSD and the full waveform observables. The 2-parameter inversion results are displayed in Fig. 6.14, arranged in the same way as Fig. 6.13, with test #3 (PSD) on the left and test #4 (waveform) on the right. For the PSD observable, μ (Fig. 6.14c) is better reconstructed than ρ (Fig. 6.14b). The SNR of the reconstructed μ model (Fig. 6.14c) is lower than in the previous case of the strong, thick and shallow inclusion (Fig. 6.13f). Still, the inclusion is clearly visible at the correct position. On the other hand, the position of the inclusion is even better spotted when using the time-domain waveform observable, for both ρ and μ models (Figs 6.14(e) and (f)). However, this is achieved at a price of an almost constant positive artifact spreading over the right part of the model. The data misfit for the PSD observable (0.159, Fig. 6.14a) is higher than in the previous case (0.065, Fig. 6.13d), namely because the negative peak at $f \approx 1.2$ Hz requires the third elastic parameter to be taken into account. For the waveform observable the data misfit is quite large (0.374, Fig. 6.14d), which also highlights the need for the third parameter to be considered. Thus, we add the third parameter, switching to the (ρ', V_s, V_p)-parametrisation. The obtained results, shown in Fig. 6.15, considerably

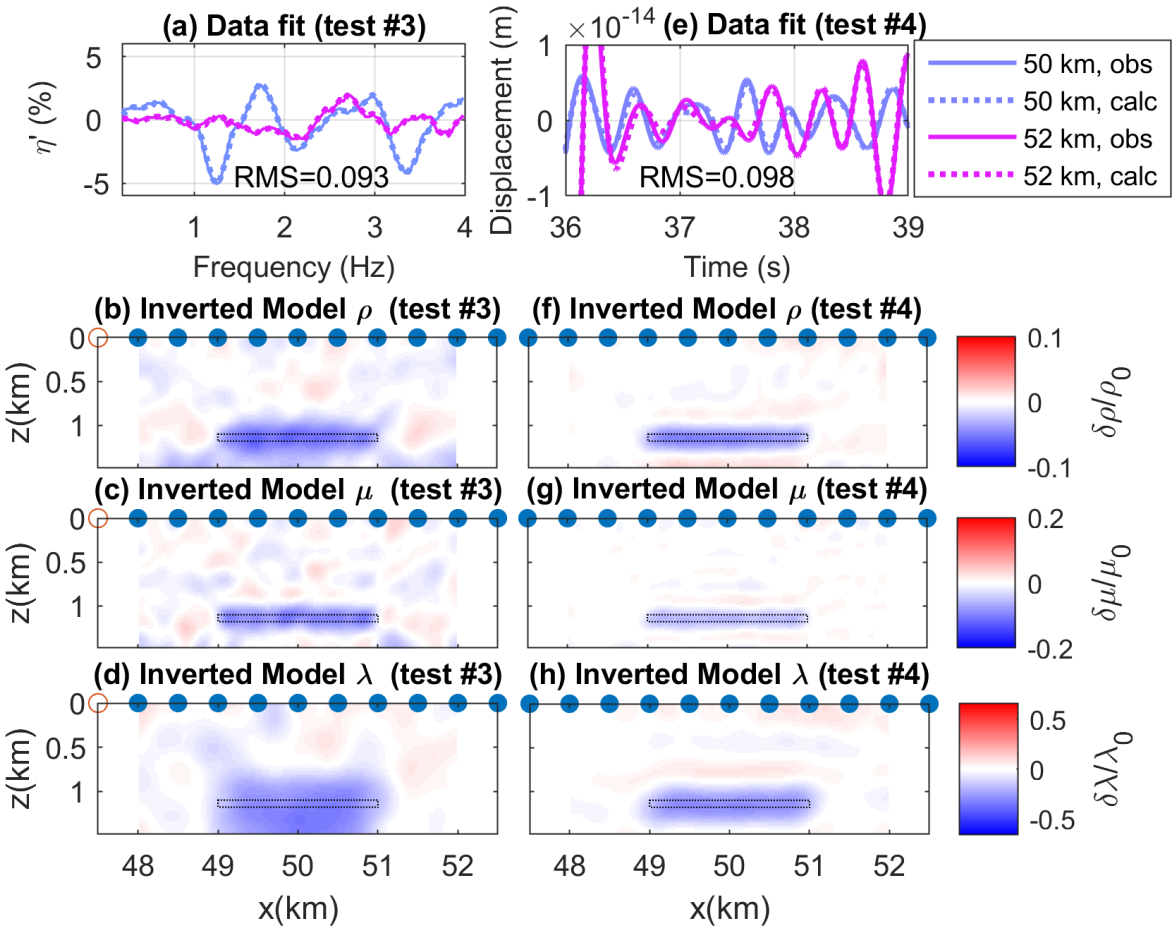


Figure 6.15: Same as Fig. 6.14, but with the (ρ', V_s, V_p)-parametrisation. Due to the added third parameter, λ can be reconstructed. The corresponding inverted models are displayed in (d) and (h). The correct model perturbations are: $\delta\rho/\rho_0 = -0.1$, $\delta\mu/\mu_0 = -0.1$, $\delta\lambda/\lambda_0 = -0.66$.

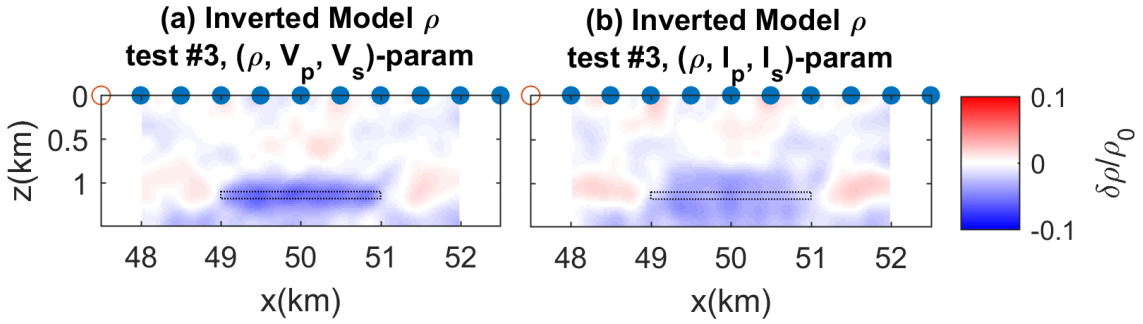


Figure 6.16: Example of bad parametrisation influence for test #3 (3-parameter): inverted density model for test #3, under (a) the (ρ', V_s, V_p)-parametrisation (red dots in Fig. 6.10) and (b) the (ρ, I_s, I_p)-parametrisation (white dots in Fig. 6.10)

improve those obtained with the 2-parameter inversion (Fig. 6.14), both in terms of data misfit and model quality for ρ and μ. For the PSD observable, the RMS misfit drops from 0.159 to 0.093, and the spectral peak at $f = 1.2$ Hz is now well explained (Fig. 6.15a). The noise level outside the inclusion in ρ and μ models strongly decreases (Figs 6.15(b) and (c)). This is also visible comparing Figs 6.9(c)-(d) to Figs 6.10(d)-(e), and observing the increase of the detection ratio when switching to the 3-parameter inversion. The same trend is observed for the waveform observable, where the positive artifact disappears (Figs 6.15(f) and (g)). The RMS misfit drops from 0.374 to 0.098, which corresponds to an almost perfect waveform match (Fig. 6.15e). Finally, the $\delta\lambda/\lambda_0$ perturbation reconstructed by the 3-parameter inversion is less well-resolved than for the first two parameters, but correctly located with the correct correct sign (negative anomaly, Figs 6.15(d) and (h)).

We also use the test #3 to highlight the influence of choosing the right parametrisation. Figs 6.16a and b display the inverted ρ models respectively under (ρ', V_s, V_p)- and (ρ', I_s, I_p)-parametrisations. The detection in Fig. 6.16a is clearly sharper. This difference was quantified by the «detection coefficient» D in Fig. 6.10d, in the cells marked by the red and the white dots. The choice of the correct parametrisation was predicted based on the spatial crosstalk in the Hessian (Figs 6.10(a) and (c), compare cells spotted with white and red dots). This example illustrates the importance of the Hessian analysis to avoid an inappropriate parametrisation choice.

6.4.4 Comparison with active-source P waves

Now that we have demonstrated the bevahiour of the surface wave inversion, we compare it to active P-wave inversion (test #5). The purpose is mostly to check the implementation by applying the method to a different type of wavefield which is known to be well suited for imaging. The corresponding observables (scattered waveforms) can be found in Figs 6.6(c) and (d) for two different explosive shots. The inverted models obtained from all the 57 frequencies linearly sampled between 0.2 and 4 Hz are shown in Figs 6.17(c), (e) and (f). The shape and position of the inclusion are correctly retrieved, as expected with P-waves, which gives us confidence about the implementation. As expected, the λ-model (Fig. 6.17f) is better reconstructed

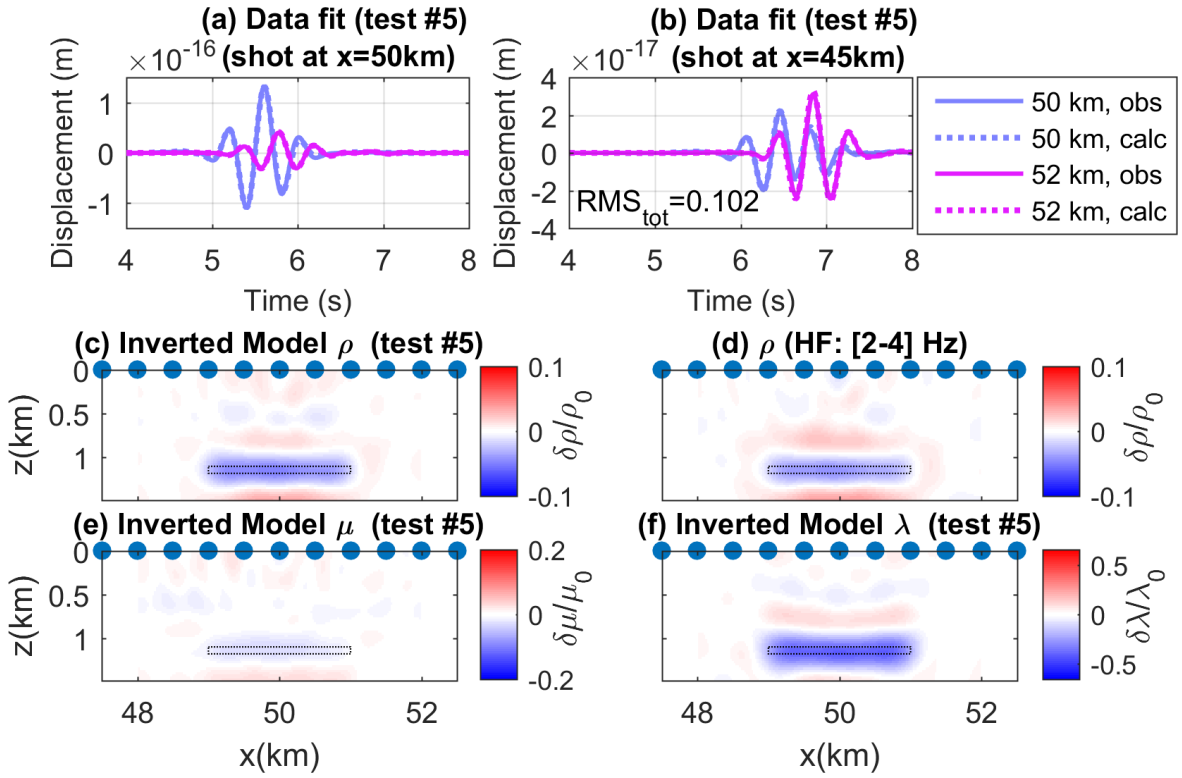


Figure 6.17: Inversion results for test #5 (active shots generating P-waves), under the (ρ', V_s, V_p)-parametrisation. (a) Scattered waveform fit at two receivers for the final model, obtained for the shot n°3 ($x_S = 50$ km, above the center of the inclusion); (b) same as (a) but for the shot n°1 ($x_S = 45$ km, on the left from the inclusion); (c),(e),(f) Final inverted models for $\rho-, \mu-, \lambda-$perturbations, respectively, based on the inversion of the full frequency interval ([0.2-4] Hz); (d) inverted model of ρ-perturbations based on frequencies between 2 and 4 Hz only.

than for any of the tests involving surface waves (see also Fig. 6.10f), highlighting the stronger constraint of the P-waves on this parameter. However, the μ-model (Fig. 6.17e) is better retrieved from surface wave experiments (see also Fig. 6.10f). This is despite the fact that the accuracy on μ is similar for both surface- and P-waves (Fig. 6.10h), since the former overstimate the model perturbation (Fig. 6.15c), while the latter underestimate it (Fig. 6.17e). The ρ-model is better retrieved from surface wave experiments with the waveform observable (Fig. 6.15f), but not with the PSD observable (Fig. 6.15b), compared to Fig. 6.17c. The typical spatial extent of the artifacts in Figs 6.17(c), (e) and (f) is larger than for the models derived from surface waves. This is probably related to the shortest wavelength involved in the problem, which is lower for Rayleigh waves compared to P-waves, because they propagate at smaller velocities.

In practice, frequencies as low as 0.2 Hz are not available in active experiments. In Fig. 6.17d, we display the ρ model resulting from the inversion of the test #5 data filtered between 2 and 4 Hz. Compared to the full-spectrum model in Fig. 6.17c, we observe more oscillations due to the lack of the low-frequency data. Thus low-frequency Rayleigh waves contained in the ambient noise may complement active seismic surveys for detecting inclusions in the sub-surface.

6.5 Discussion

The main conclusion of these investigations is that the vertical component power spectral density of Rayleigh waves alone can in some cases contain enough information to retrieve the position and the shape of an embedded scatterer. The inverted models are contaminated by more or less pronounced artifacts depending on the modal content of the incident wavefield, the depth of the inclusion and the number of inverted parameters. The values of the parameter perturbations inside the inclusion are predicted with an error ranging from 25% to 50% most of the time (see the right-most columns in Figs 6.8, 6.9 and 6.10, where the blue end of the colorbar corresponds to a 100% relative error on the parameter perturbation, while the yellow end corresponds to a 0% error). The perturbation magnitudes can be either over-estimated (Fig. 6.13f) or under-estimated (Figs 6.14 and 6.15), depending on the particular configuration, so we cannot conclude to a general behaviour of the Rayleigh wave inversion from this point of view. A widely used technique for choosing the optimal parametrisation is the radiation pattern analysis (e.g. Plessix & Cao, 2011), which aims to minimise the inter-parameter cross-talk. In our case, however, the parametrisation choice was guided by the detection objective (minimising spatial crosstalk), rather than by the parameter estimation objective (minimising inter-parameter crosstalk). In Section 6.3, we show that there is no universal optimal parametrisation allowing to minimize both types of crosstalk. We realised that the strong spatial cross-talk could prevent the sharp inclusion detection (Fig 6.16). In this case, an accurate parameter estimation is not so relevant, since the inclusion is not enough focused, and the elastic parameter perturbation values are diluted within a larger area. If we considered the location of the inclusion as known, as done for example by Campman & Riyanti (2007), then we would probably choose other parametrisations (based on inter-parameter crosstalk minimisation), and the accuracy of the parameter estima-

tion would be better. An important conclusion of Chapter 5 was that the three elastic parameters had distinct patterns, because the virtual sources associated to ρ, μ and λ were respectively a rotating point force, a rotating double-couple moment with a pulsating isotropic component, and a pure isotropic moment. Though we do not explicitly use the radiation patterns in the present work (we introduce a scalar criterion C_I which measures their orthogonality), the fact that we recover the inclusion in the inverted models associated to different parameters is a consequence of the radiation pattern difference. It is interesting to observe that introducing the third parameter V_p (or equivalently λ) improves the data fit and the inverted models of the two other parameters, despite the bad conditioning. In the synthetic examples presented in this work, the model retrieved for λ was still meaningful (Figs 6.10(d) and (h)), though poorly resolved compared to the two other parameters. We do not think it is realistic to expect a meaningful λ or V_p model from the noisier real datasets. However, we suggest that the third parameter should be kept in the parametrisation in order to improve the data fit and to de-noise the models of the other parameters. As far as we perform a linear inversion, hierarchical approaches (e.g. Campman & Riyanti, 2007; Brossier et al., 2009; Solano et al., 2014, in the non-linear context) do not seem so relevant, since the final model is not expected to be so much influenced by the order in which the different parameters or frequencies are updated. Preconditioning , however, could be a useful improvement, as it would allow a quicker convergence. 3-parameter inversions presented in Section 6.4 typically required several hundreds of iterations of the CLG algorithm: this number could be reduced. Moreover, preconditioning is easy to implement with a Gauss-Newton formulation, since the whole Hessian matrix is reconstructed. The inverse diagonal of the Hessian could then be an example of preconditioner (Plessix & Mulder, 2004). Preconditioning could also play an important role when addressing more complicated background models containing slow shallow layers, where the sensitivity is likely to be concentrated (see Chapter 4). Such applications are planned as future research. In the remainder of the discussion, we focus on some specific aspects which we think deserve particular attention.

6.5.1 Why do overtones perform better ?

In this work, we modelled a mixture of many overtones (Figs 6.4(b) and (d)), which mimics the L_g phase, detected as one of the dominating phases in the short-period ambient noise by Koper et al. (2010) and also in Chapter 3. As already mentioned, the deeper sensitivity of the overtones is a consequence of the deeper penetration of their eigenfunctions (Aki & Richards, 2002). However, we saw in Fig. (6.13) that they also provide a better resolution. This is a consequence of the sign quickly alternating with depth in the sensitivity kernel shown in Fig. 6.2b. Such an oscillation means that the sign of the observable (PSD anomaly) at a given frequency yields an important constraint on the vertical position of the inclusion. For the fundamental mode sensitivity kernel (Fig. 6.2a), the sign is more or less constant along vertical sections. Thus, it is more difficult to deduce the inclusion depth from the sign of the PSD anomaly at a given frequency. The only tie between the frequency of the anomaly and the depth of the inclusion in this case is the frequency-dependant depth penetration of the fundamental mode. As a result, the inverted model appears smooth (Fig. 6.13(b) and (c)).

We suggest that the reason of the sign oscillations for the overtones might be due to frequent alternations of prograde and retrograde motions with depth, though we did not check this conjecture. We conclude that the presence of the overtones in the ambient wavefield is a favourable factor for amplitude inversion, and time windows containing overtones should be selected.

6.5.2 Possible real-data applications

We saw in Fig. 6.15 that the standard observable (waveform) performed better than the average PSD of random fields. Still, we beleive the PSD inversion is the most valuable result of this work, because the PSD, contrary to the coherent waveforms, can be easily measured from ambient noise records. So, from a practical point of view, what can the proposed method tell us about a reservoir, and can it be used as an imaging tool for ambient noise spatial variations described in the introduction? The ideal, but rather irrealistic case corresponds to a reservoir embedded in an otherwise laterally invariant medium, as the one presented in this paper. Then the ambient noise amplitude recorded in a single acquisition can be inverted for the reservoir size and position, as $\eta_0 = 0$ in equation (4.5). The first question for real acquistions is thus to know whether the investigated zone contains strong structural effects other than the reservoir. If yes, then the method cannot be applied in the exploration context (i.e. finding the position and the size of a static reservoir with a single acquisition). As shown in Chapter 4, the structural effects (such as an anticline deformation of several layers) on the ambient noise amplitude are likely to be about an order of magnitude stronger than the reservoir effects, so they dominate the spatial variations of the ambient noise, overshadowing the reservoir. Mathematically, this means that η_0 in equation (4.5) is non-zero and unknown. In this case, we could try to image the structural heterogeneity as a perturbation with respect to a laterally invariant model. Unfortunately, we concluded in Chapter 4 that the Born approximation is likely to fail in case of such strong contrasts. Thus a non-linear inversion must be designed. The situation is different in the monitoring context, where the reservoir is not static because of fluid pumping/injection, so that its elastic properties evolve over time. In this case, performing time-lapse acquisitions yields both η and η_0 in equation (4.5), so that the method can in theory be applied to retrieve the position and the size of the zone affected by the medium changes. The static structural effect is filtered out by the time-lapse acquisitions. However, the sensitivity kernels must computed into the reference medium containing the structure (ground model m_0 corresponding to η_0), which requires to have a good a priori knowledge of the layering topography. This is usually the case for underground gas storage facilities (UGS), where gas is injected and withdrawn in a cyclic manner for decades. UGS monitoring could thus be a promising application of the method. We are currently recording a continuous dataset over one such UGS operated by Storengy to see if significant cyclic ambient noise amplitude changes correlated with gas injection can be measured. To summarise, the linearised spectral amplitude imaging with Rayleigh waves can possibly be applied in two cases:

- Reservoir exploration: when the reservoir is the most prominent geological heterogeneity in the survey area;

- Reservoir monitoring: when time-lapse acquisitions are available as the reservoir properties (pressure, fluid content,...) evolve over time, and the background model (including the structure) is well known.

The simple examples presented in this paper rely on several assumptions on the background model and the ambient noise composition. In the next section, we discuss these assumptions, and also provide some ideas to overcome the associated limitations in order to be able to apply the method to real datasets.

6.5.3 Ambient wavefield characterisation

Several difficulties related to the assumptions we have to make on the ambient noise wavefield might arise when considering real-data applications of the proposed method. There are three main assumptions:

- The frequency-dependent modal composition of the ambient noise (wave type, velocity, back-azimuth) is known (and does not change between snapshots in case of time-lapse acquisitions).

- We are able to numerically simulate wavefields with this precise modal content.

- The wavefield entirely consists of *coherent* waves propagating across the array, generated by random uncorrelated sources.

An important advantage of the normalised PSD-based observables is that the knowledge of precise source-time functions of the ambient noise sources, and even the absolute noise level, are not required, because of the normalisation by the reference station in equation (4.3). The first assumption in the list comes from the influence of the modal content on the sensitivity kernels, which we demonstrate in this work. A thourough frequency-dependent wavefield characterisation is necessary for applying the relevant sensitivity kernels in the relevant frequency bands. Such kind of characterisation was proposed in Chapter 3, where we used a combination of the ambient noise cross-correlation with array methods. However, the drawback of the array methods, such as FK (Capon et al., 1967), HRFK (Capon, 1969) or MUSIC (Schmidt, 1986) is that their objective functions do not directly yield the true energy associated with each propagation mode. Additional steps might then be needed for partitioning the PSD into different propagation mode contributions. Regarding the back-azimuth, which is a 3D feature, its equivalent in 2D is the balance between the energy coming from both ends of the model. In the present work, we assumed symmetric illumination, which corresponds to an isotropic source distribution around the array in 3D. However, the number of sources at both ends (or at different azimuths in 3D simulations) could be adapted based on the real-data beamforming results.

Assuming the modal contributions to the wavefield energy are correctly estimated, it is also necessary to reproduce the same modal composition in the forward-modelling part of the workflow (item 2 in the list). This is relatively easy to achieve for the fundamental mode, which is the main mode excited by surface sources. In this work, we also managed to model a mixture of many overtones. However, modeling a precise balance of the overtones in numerical simulation depends on a fine tuning of

the source depth and can be a complicated task. Analytical modelling of the incident wavefield should be considered, as the modal summation theory implemented in software like *Computer Programs for Seismology* (Herrmann, 2013) allows to model seismograms (and thus also sensitivity kernels) specific to each particular mode. In this case, however, handling structural heterogeneities would become a challenge.

Finally, the modelling such as performed in our paper relies on random, but coherent wavefields. However, the diffuse field assumption (DFA), recently proposed by Sánchez-Sesma (2017) to model the H/V ratio, considers the wavefield is totally diffuse, with energy equipartionned between all the available modes of propagation (Weaver, 1982). In this case, the auto-correlation (same as the PSD) computed at a given receiver in x_r is proportional to the imaginary part of the Green's function obtained when the source and the receiver are both in x_r. This is clearly different from our modelling, which depends on particular dominating modes in the incident wavefield. In reality, at frequencies above ~ 1 Hz, the coherence of the ambient noise drops (Lehujeur et al., 2015). In the same time, the validity of the DFA increases. This means that it might also be necessary to separate the energies of coherent and diffuse components of the wavefield, prior to the PSD inversion.

Integrating all the necessary information in a quantitative way increases the number of pre-processing steps prior to the inversion, and generates an accumulation of uncertainty on the measured PSD and on the relevance of the sensitivity kernels. Another source of uncertainty is the numerical modelling framework, via the background model and the 2D assumption. This is discussed in the next paragraph.

6.5.4 Forward modelling and inversion: improvement potential

The most important limitation of our modelling is that it is performed in 2D. The main difference with the 3D case is that surface waves experience no geometrical spreading. This creates a standing-wave phenomenon throughout the model due to the interaction of incident and scattered wavefields (Gorbatikov & Tsukanov, 2011). An extension to 3D would certainly be necessary for most field applications. Because of the absence of standing waves in 3D, the anomalies are expected to be more focused above the inclusion compared to Fig. 6.3. As 3D SEM is more time-consuming, alternative approaches could be considered, such as finite-difference modelling (Saenger et al., 2000), or semi-analytical methods in flat layered media (Riyanti & Herman, 2005). Also, storing the full Hessian would not be realistic in 3D. The Gauss-Newton method could no more be used, and the matrix-free adjoint-state method could be adopted instead (Plessix, 2006). This would require to derive the proper adjoint source function, in order to retropropagate the the PSD residuals, as done by Solano et al. (2014) in their appendix B for the w-AWI observable. Implementing the adjoint-state method would offer a natural possibility to perform a non-linear inversion, simply by retro-propagating the residual in the *updated* model at each iteration. This would possibly allow to address stronger contrasts, such as the structural effects, where the Born approximation fails (see Chapter 4).

For now, as we present a linear inversion method, a good background model is a pre-requisite. In the exploration context (unknown field), it would require, for example, a preliminary ambient-noise based dispersion curve inversion for a 1D model (see Chapter 3). However, the sensitivity study in Chapter 4 showed that our PSD-based observable exhibits some tolerance with respect to erroneous background models, as far as the error on the background model does not generate a too large phase shift between the incident and the scattered wavefields. This issue is addressed in Appendix 6.D, which shows that the objective function J_{PSD} is expected to be more robust with respect to errors on the background velocity model compared to the standard J_{LSM}, in case of distant sources.

Despite the discussed limitations, the obtained results are quite valuable since they demonstrate the concept of PSD inversion as the only observable, without any phase information. Previously, Tsukanov & Gorbatnikov (2018) developed a workflow for linearised V/H ratio migration for the fundamental mode in a homogeneous medium, using an ad-hoc sensitivity kernel based on semi-qualitative considerations. Our results extend these pioneering works by linearising the true elastic wave equation. This uncovers the potential of the Rayleigh amplitude, particularly the overtones, for imaging the perturbations of the sub-surface. The comparison with P-waves, namely a comparable sensitivity (Fig. 6.11b), a better resolution on the μ parameter (section 6.4.4), and a natural availability of low frequencies can also open new perspectives for coupling passive and active surveys.

6.6 Conclusion

Two inclusions with different depths and elastic properties were added to a simple 2D reference medium. The Rayleigh wave spectral amplitude perturbations between 0.2 and 4 Hz induced by both inclusions were computed by numerical simulation (SPECFEM2D). The modelling was based on random wavefield generation by distant sources in order to approach the case of a coherent ambient noise. The inverse problem was posed based on the Born approximation (single-scattering). The full Fréchet derivative and the Hessian were stored in memory thanks to the small scale of the problem and its low-frequency content. The spectral amplitude was then inverted for the elastic perturbations of the medium, using the Gauss-Newton method with Tikhonov regularisation, and an iterative resolution by the Conjugate Linear Gradient algorithm. For the shallow and strongly contrasting inclusion, both Rayleigh fundamental mode and overtones yielded the correct inclusion position and shape, with a better vertical resolution for the overtones. For the deep and weaker contrasting inclusion, only the overtones had enough sensitivity, again yielding the correct inclusion position and shape. 3-parameter inversions under the (ρ', V_s, V_p)-parametrisation yielded the best results. The parameter values inside the inclusion were always predicted with the correct sign, but with a relative estimation error ranging from 25 to 50%. Adding phase information to the Rayleigh wave overtones inversion improved the resolution, and suppressed nearly all the artifacts. Compared to active shot P-wave inversion within the same framework (vertical component waveform, elastic modelling), Rayleigh wave overtones allow to reconstruct μ with a better resolution, while P-waves yield a better resolution in λ. Preconditioned inversion should be con-

sidered in background models containing slow shallow layers, not addressed in this study. Extension to 3D modelling would probably be necessary for most real-data applications. Finally, non-linear inversion might be used for addressing strong structural perturbations, which are beyond the validity of the Born approximation.

Appendix

6.A Sensitivity kernels under different parametrisations

In this appendix we derive the expressions of the sensitivity kernels under different parametrisations. We emphasize that the sensitivity kernel for a given parameter depends on the other parameters involved in the parametrisation, as well as on the number of the parameters involved. For elastic isotropic media, the perturbation of an observable d is a function of 3 independent parameters m_1, m_2, m_3. Under the Born approximation, this function is linearised and can be written as

$$\delta d = \int_\Omega K_{(m_1)} \delta m_1(\mathbf{x}) + \int_\Omega K_{(m_2)} \delta m_2(\mathbf{x}) + \int_\Omega K_{(m_3)} \delta m_3(\mathbf{x}) \ , \tag{6.A.1}$$

where $K_{(m_i)}$ are the sensitivity kernels. Their explicit expressions are given in equation (4.A.5) for *relative* parameter perturbations under the (ρ, λ, μ) parametrisation. Taking this parametrisation as a reference, any other 3-parameter inversion can be described by a function f mapping $\mathbb{R}^3$ on $\mathbb{R}^3$:

$$\begin{cases} m_1 = f_1(\rho, \lambda, \mu) \\ m_2 = f_2(\rho, \lambda, \mu) \\ m_3 = f_3(\rho, \lambda, \mu) \ . \end{cases}$$

Replacing this into (6.A.1), we obtain

$$\begin{aligned} \delta d = & \int_\Omega K_{(m_1)} \left(\partial_\rho f_1 \delta\rho(\mathbf{x}) + \partial_\lambda f_1 \delta\lambda(\mathbf{x}) + \partial_\mu f_1 \delta\mu(\mathbf{x}) \right) \\ & + \int_\Omega K_{(m_2)} \left(\partial_\rho f_2 \delta\rho(\mathbf{x}) + \partial_\lambda f_2 \delta\lambda(\mathbf{x}) + \partial_\mu f_2 \delta\mu(\mathbf{x}) \right) \\ & + \int_\Omega K_{(m_3)} \left(\partial_\rho f_3 \delta\rho(\mathbf{x}) + \partial_\lambda f_3 \delta\lambda(\mathbf{x}) + \partial_\mu f_3 \delta\mu(\mathbf{x}) \right) \ . \end{aligned} \tag{6.A.2}$$

By identification of the above expression with (6.A.1) evaluated under the (ρ, λ, μ) parametrisation, we get

$$[K_{(\rho)}, K_{(\lambda)}, K_{(\mu)}]^{\mathrm{T}} = \mathbf{Jac}_f^{\mathrm{T}} \cdot [K_{(m_1)}, K_{(m_2)}, K_{(m_3)}]^{\mathrm{T}} \ , \tag{6.A.3}$$

with $\mathbf{Jac}_f$ the Jacobian matrix of f. Finally, the desired new kernels are

$$[K_{(m_1)}, K_{(m_2)}, K_{(m_3)}]^{\mathrm{T}} = \left(\mathbf{Jac}_f^{\mathrm{T}} \right)^{-1} \cdot [K_{(\rho)}, K_{(\lambda)}, K_{(\mu)}]^{\mathrm{T}} \ . \tag{6.A.4}$$

Working with relative perturbations is easier because f becomes linear in most cases. As an example, the steps for obtaining the kernels under the (ρ',V_s,V_p) parametrisation

are given below (we use the notation ρ' for the density under parametrisations other than (ρ, λ, μ)). Since $V_s = (\mu/\rho)^{1/2}$, we have

$$\delta \ln V_s = \frac{1}{2} \left(\delta \ln \mu - \delta \ln \rho \right) \ .$$

Similarly, $V_p = ((\lambda + 2\mu)/\rho)^{1/2}$ yields

$$\begin{aligned} \delta \ln V_p &= \frac{\lambda_0}{2(\lambda_0 + 2\mu_0)} \delta \ln \lambda + \frac{\mu_0}{\lambda_0 + 2\mu_0} \delta \ln \mu - \frac{1}{2} \delta \ln \rho \\ &= c_\lambda \delta \ln \lambda + c_\mu \delta \ln \mu - \frac{1}{2} \delta \ln \rho \ . \end{aligned}$$

For convenience, we change the order of the parameters in the reference parametrisation to (ρ,μ,λ). Then the Jacobian matrix is

$$\mathbf{Jac}_{(\rho', Vs, Vp)} = \begin{pmatrix} 1 & 0 & 0 \\ -1/2 & 1/2 & 0 \\ -1/2 & c_\mu & c_\lambda \end{pmatrix} \ .$$

Noticing that $c_\lambda + c_\mu = 1/2$, we find the inverse of the transposed Jacobian as

$$\left(\mathbf{Jac}^{\mathrm{T}}_{(\rho', Vs, Vp)} \right)^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & -2c_\mu/c_\lambda \\ 0 & 0 & 1/c_\lambda \end{pmatrix} \ .$$

Thus the new kernels for *relative* perturbations are

$$\begin{aligned} K_{(\rho')} &= K_{(\rho)} + K_{(\lambda)} + K_{(\mu)} \\ K_{(Vs)} &= 2K_{(\mu)} - 4\frac{\mu_0}{\lambda_0} K_{(\lambda)} \\ K_{(Vp)} &= 2\frac{\lambda_0 + 2\mu_0}{\lambda_0} K_{(\lambda)} \ . \end{aligned}$$

If one wishes to design a 2-parameter inversion, a partial derivative of d with respect to one of the parameters of the reference parametrisation must be neglected. This sets one column of $\mathbf{Jac}_f$ to zero, making it a 2x2 invertible matrix. Based on the comparison of the Rayleigh wave sensitivity with respect to different parameters (see Chapter 4) we choose to set $\partial_\lambda d \approx 0$ for deriving the 2-parameter inversions presented in this paper. For example, the Jacobian matrix for the (ρ',V_s) parametrisation is

$$\mathbf{Jac}_{(\rho', Vs)} = \begin{pmatrix} 1 & 0 \\ -1/2 & 1/2 \end{pmatrix} \ ,$$

yielding

$$\begin{aligned} K_{(\rho')} &= K_{(\rho)} + K_{(\mu)} \\ K_{(Vs)} &= 2K_{(\mu)} \ . \end{aligned}$$

This approach does not impose any constraint on the third parameter (for example V_p under the (ρ',V_s) parametrisation). This is different from the case where only two parameters are updated under a 3-parameter inversion, which freezes the third parameter. In Table 6.A.1, we give the expressions of the kernels under all the parametrisations considered in this paper, as linear combinations of $K_{(\rho)}$, $K_{(\lambda)}$ and $K_{(\mu)}$.

Table 6.A.1: Expressions of the sensitivity kernels under different parametrisations as linear combinations of $K_{(\rho)}$, $K_{(\lambda)}$, $K_{(\mu)}$ (columns). Each entry corresponds to the dimensionless coefficient in front of the corresponding kernel.

Param.	Kernel	$K_{(\rho)}$	$K_{(\lambda)}$	$K_{(\mu)}$
(ρ',V_p,V_s)	$K_{(\rho')}$	1	1	1
	$K_{(Vp)}$	0	$2\left(1+2\frac{\mu_0}{\lambda_0}\right)$	0
	$K_{(Vs)}$	0	$-4\frac{\mu_0}{\lambda_0}$	2
(ρ',I_p,I_s)	$K_{(\rho')}$	1	-1	-1
	$K_{(Ip)}$	0	$2\left(1+2\frac{\mu_0}{\lambda_0}\right)$	0
	$K_{(Is)}$	0	$-4\frac{\mu_0}{\lambda_0}$	2
(V_s,I_s,λ)	$K_{(Vs)}$	1	0	-1
	$K_{(Is)}$	1	0	1
	$K_{(\lambda)}$	0	1	0
(ρ',V_s)	$K_{(\rho')}$	1	0	1
	$K_{(Vs)}$	0	0	2
(ρ',I_s)	$K_{(\rho')}$	1	0	-1
	$K_{(Is)}$	0	0	2
(V_s,I_s)	$K_{(Vs)}$	1	0	-1
	$K_{(Is)}$	1	0	1

6.B Quantitative criteria for assessing parametrisation quality

In this Appendix, we enumerate simple quantitative criteria facilitating the optimal parametrisation choice. They concern spatial crosstalk, inter-parameter crosstalk, conditioning and absolute sensitivity.

6.B.1 Spatial crosstalk

As shown in Section 6.3, the spatial crosstalk is carried by the off-diagonal elements of the blocks of the Hessian matrix. Pan (2017) propose to use Point Spread Functions (PSF) to quantify the leakage through spatial crosstalk. A PSF is the product of the Hessian with a test model vector $\mathbf{\Delta m}_0$ which is zero everywhere but in one point $\mathbf{x}_0$ where it equals 1 (or any other value):

$$\begin{cases} \mathbf{\Delta m}_0 = 1 \text{ if } \mathbf{x} = \mathbf{x}_0 \,, \\ \mathbf{\Delta m}_0 = 0 \text{ otherwise} \,; \end{cases} \tag{6.B.1}$$

and

$$\mathrm{PSF}(\mathbf{x}_0) = \mathbf{H}\mathbf{\Delta m}_0. \tag{6.B.2}$$

The PSF has the dimension of the model ($N_m = N_x \cdot N_z$). In case of low spatial crosstalk (i.e. $\mathbf{H}$ approaches the identity matrix), the PSF appears concentrated around $\mathbf{x_0}$. In our synthetic study, we wish to minimize the leakage from an inclusion occupying a known region Σ_0. A simple criterion is obtained by extending the PSF notion to a new test model vector:

$$\begin{cases} \mathbf{\Delta m}_0 = 1 \text{ if } \mathbf{x} \in \Sigma_0 \,, \\ \mathbf{\Delta m}_0 = 0 \text{ otherwise} \,. \end{cases} \tag{6.B.3}$$

It is natural to call the product of the Hessian with such a test model vector an *Object Spread Function*, or OSF. In case of a multi-parameter model with N_p parameters, the size of $\mathbf{\Delta m}_0$ is $N_p \cdot N_m$, and $\mathbf{\Delta m}_0 = 1$ in all the cells corresponding to the inclusion, whatever the parameter. The resulting OSF splits into N_p parts. Each part i represents the information leakage from the cells of Σ_0, summed over all the parameters, towards the full model of the parameter m_i. Taking the example of the (ρ, λ, μ)-parametrisation, and using the blocks of $\mathbf{H}$, this can be written as

$$\begin{cases} \mathrm{OSF}_\rho(\Sigma_0) = \mathbf{H}_{\rho,\rho}\mathbf{\Delta m}_0^{(\rho)} + \mathbf{H}_{\rho,\lambda}\mathbf{\Delta m}_0^{(\lambda)} + \mathbf{H}_{\rho,\mu}\mathbf{\Delta m}_0^{(\mu)} \\ \mathrm{OSF}_\lambda(\Sigma_0) = \mathbf{H}_{\lambda,\rho}\mathbf{\Delta m}_0^{(\rho)} + \mathbf{H}_{\lambda,\lambda}\mathbf{\Delta m}_0^{(\lambda)} + \mathbf{H}_{\lambda,\mu}\mathbf{\Delta m}_0^{(\mu)} \\ \mathrm{OSF}_\mu(\Sigma_0) = \mathbf{H}_{\mu,\rho}\mathbf{\Delta m}_0^{(\rho)} + \mathbf{H}_{\mu,\lambda}\mathbf{\Delta m}_0^{(\lambda)} + \mathbf{H}_{\mu,\mu}\mathbf{\Delta m}_0^{(\mu)} \,. \end{cases} \tag{6.B.4}$$

A scalar criterion of the spatial crosstalk affecting the reconstruction of the parameter m_i is obtained by taking the ratio between the root median square OSF inside and outside Σ_0, measured in decibel:

$$C_S^{(m_i)} = 10 \log_{10} \left(\frac{\sqrt{\mathrm{med}\left(\left\| \mathrm{OSF}_{m_i} \right\|^2 \right)_{\mathbf{x} \in \Sigma_0}}}{\sqrt{\mathrm{med}\left(\left\| \mathrm{OSF}_{m_i} \right\|^2 \right)_{\mathbf{x} \notin \Sigma_0}}} \right) \,. \tag{6.B.5}$$

With such a definition, the leakage from $m_{j\neq i}$ towards m_i at the correct spatial location does not result into a decrease of $C_S^{(m_i)}$. That is, while the OSF accounts for both spatial and inter-parameter crosstalks, $C_S^{(m_i)}$ focuses on the spatial crosstalk.

6.B.2 Inter-parameter crosstalk

In section 6.3, we have also seen that the pure inter-parameter crosstalk is carried by the diagonals of the off-diagonal blocks of the Hessian: $\mathbf{H}_{m_i,m_j}(i \neq j)$. Taking a the model cell corresponding to the position $\mathbf{x}_0$ cell under the (ρ, λ, μ) parametrisation, the corrseponding diagonal elements of the Hessian blocks can be arranged into a 3x3 matrix:

$$\mathcal{M}(\mathbf{x}_0) = \begin{pmatrix} H_{\rho,\rho}(\mathbf{x}_0,\mathbf{x}_0) & H_{\rho,\lambda}(\mathbf{x}_0,\mathbf{x}_0) & H_{\rho,\mu}(\mathbf{x}_0,\mathbf{x}_0) \\ H_{\lambda,\rho}(\mathbf{x}_0,\mathbf{x}_0) & H_{\lambda,\lambda}(\mathbf{x}_0,\mathbf{x}_0) & H_{\lambda,\mu}(\mathbf{x}_0,\mathbf{x}_0) \\ H_{\mu,\rho}(\mathbf{x}_0,\mathbf{x}_0) & H_{\mu,\lambda}(\mathbf{x}_0,\mathbf{x}_0) & H_{\mu,\mu}(\mathbf{x}_0,\mathbf{x}_0) \end{pmatrix} . \tag{6.B.6}$$

The interprameter crosstalk within the cell $\mathbf{x}_0$ is carried by the off-diagonal elements of $\mathcal{M}(\mathbf{x}_0)$. Following Plessix & Cao (2011), performing an eigendecomposition of $\mathcal{M}(\mathbf{x}_0)$ allows to quantify the ambiguity between different parameters which arises from the non-diagonal terms. Let $\lambda_{1,2,3}$ be the eigenvalues of $\mathcal{M}(\mathbf{x}_0)$ sorted in decreasing order. The associated eigenvectors are

$$\begin{cases} v_1 = \alpha_{1,\rho}\mathbf{e}_\rho + \alpha_{1,\lambda}\mathbf{e}_\lambda + \alpha_{1,\mu}\mathbf{e}_\mu \\ v_2 = \alpha_{2,\rho}\mathbf{e}_\rho + \alpha_{2,\lambda}\mathbf{e}_\lambda + \alpha_{2,\mu}\mathbf{e}_\mu \\ v_3 = \alpha_{3,\rho}\mathbf{e}_\rho + \alpha_{3,\lambda}\mathbf{e}_\lambda + \alpha_{3,\mu}\mathbf{e}_\mu \ , \end{cases} \tag{6.B.7}$$

with $\mathbf{e}_\rho, \mathbf{e}_\lambda, \mathbf{e}_\mu$ the canonical basis of $\mathbb{R}^3$. Ideally, in case of a diagonal $\mathcal{M}(\mathbf{x}_0)$ each eigenvector would be related to a single parameter. The mix of the parameters on different eigenvectors measures the inter-parameter crosstalk. As a simple criterion, we propose to use the average ambiguity of the eigenvectors over the target Σ_0:

$$C_I^{(\mathrm{raw},i)} = \left\langle \frac{\max\limits_{m\in\{\rho,\lambda,\mu\}} \left|\alpha_{i,m}(\mathbf{x})\right|}{\sum\limits_{m\in\{\rho,\lambda,\mu\}} \left|\alpha_{i,m}(\mathbf{x})\right|} \right\rangle_{\mathbf{x}\in\Sigma_0} , \tag{6.B.8}$$

where i is the number of the eigenvector. In the case of a 3-parameter inversion, the worst possible value of $C_I^{(\mathrm{raw})}$ is 0.33, obtained when all the $\alpha_{i,m}$ are equal. In the case of 2-parameter inversion, this value is 0.5. For a better comparability, we choose to recast the value of C_I between 0 and 1:

$$\begin{cases} C_I^{(i)} = \dfrac{C_I^{(\mathrm{raw},i)} - 0.5}{1 - 0.5} \text{ for 2-parameter inversions,} \\ C_I^{(i)} = \dfrac{C_I^{(\mathrm{raw},i)} - 0.33}{1 - 0.33} \text{ for 3-parameter inversions.} \end{cases} \tag{6.B.9}$$

A simple calculation shows that for 2-parameter inversions one always has $C_I^{(1)} = C_I^{(2)}$, while for 3-parameter inversions $C_I^{(i)}$ can take three distinct values. C_I can be seen as a measure of orthogonality between the radiation patterns associated to each parameter. For 2-parameter inversions, there are only two patterns to compare, while

there are three possible combinations in the 3-parameter case. Below we develop this parallel with the radiation patterns.

A traditional tool for measuring the inter-parameter crosstalk is the visual radiation pattern analysis, usually performed within the infinite-frequency approximation (references). Our inversion method, however, relies on the finite-frequency effects in the Rayleigh wave scattering. Another complication is that the scattered waves are observed in the near-field with respect to the scattering zones, so that the different scattered wave modes do not have time to separate from each other. Thus, it would be complicated to construct separate patterns for Rayleigh-Rayleigh, Rayleigh-P or Rayleigh-S diffraction. Instead, in Chapter 5, derived frequency-dependent radiation patterns for the mixture of all the modes resulting from Rayleigh wave scattering in a homogeneous medium. They were defined as a function of the receiver position with respect to the scattering point, and assuming that the incident Rayleigh wave propagates in the sense of the x-axis:

$$\Omega_{(m)}(r,\Theta,\omega) = K^*_{(m)}(r,\Theta,\omega)K_{(m)}(r,\Theta,\omega) = K^*_{(m)}(\mathbf{x}_r,\mathbf{x}_0,\omega)K_{(m)}(\mathbf{x}_r,\mathbf{x}_0,\omega) \ , \tag{6.B.10}$$

with (r,Θ) the polar coordinates of $\mathbf{x}_r$ with respect to $\mathbf{x}_0$. This quantity represents a scattered energy per surface surface element and per relative elastic parameter perturbation. Low inter-parameter crosstalk means that the patterns associated to different parameters should have as different shapes as possible, at as many frequencies as possible. Mathematically, this is translated by the following orthogonality relation:

$$\forall i \neq j, \sum_{k=1}^{N_r}\sum_{l=1}^{N_f} K_{m_i}(\mathbf{x}_k,\mathbf{x}_0,\omega_l)K^*_{m_j}(\mathbf{x}_k,\mathbf{x}_0,\omega_l) = 0 \ . \tag{6.B.11}$$

From the definition (6.4), the above equation is equivalent to zeroing the diagonal element corresponding to $\mathbf{x}_0$ in each the off-diagonal block of the Hessian:

$$\forall i \neq j, \mathbf{H}_{m_i,m_j}(\mathbf{x}_0,\mathbf{x}_0) = 0 \ . \tag{6.B.12}$$

This is the link between the radiation pattern analysis and the Hessian analysis for inter-parameter crosstalk evaluation. $C_I^{m_j} = 1$ is indeed equivalent to equation (6.B.12), i.e. the radiation pattern of m_j is orthogonal to the other patterns. The advantage of the Hessian analysis is that is naturally accounts for all the frequencies, while the radiation pattern analysis, in our case, would require to examine radiation patterns at many different frequencies. Another advantage of the Hessian analysis is to be quantitative rather than qualitative.

6.B.3 Conditioning

In this paragraph the conditioning is to be understood in the sense of the inter-parameter crosstalk and measures the overshadowing of the less sensitive parameters by the most sensitive ones, at a given spatial location. This value is defined as the ratio of the largest by the smallest eigenvalue of the matrix in equation (6.B.6):

$$\kappa(\mathbf{x}_0) = \frac{\lambda_{\max}}{\lambda_{\min}} \ . \tag{6.B.13}$$

One would wish to have a κ as close to 1 as possible. For an easier visual inspection, we define a criterion for which a larger value indicates a preferable configuration, and we average it over the target Σ_0:

$$C_\kappa = -10\log_{10}\left\langle \kappa(\mathbf{x})\right\rangle_{\mathbf{x}\in\Sigma_0} \ . \tag{6.B.14}$$

6.B.4 Sensitivity

We define the sensitivity as the ratio of the observable perturbation to the observable value in the unperturbed medium, per unit surface and per medium relative perturbation. If the observable is the scattered waveform $\mathbf{u}^{(1)}$ (see equation (4.9)), the sensitivity would be of the order of $\left|K_{(\rho,\lambda,\mu)}/u^{(0)}\right|$, with K the sensitivity kernel from equation (4.9). In order obtain a global measure of the sensitivity summed over all the frequencies and receivers, we use the diagonal terms of the Hessian (since $\mathbf{H} = \mathfrak{Re}\left(\mathbf{K}^\dagger\mathbf{K}\right)$), divided by the RMS amplitude of the incident wavefield over the array:

$$k_{\mathrm{WF}}(\mathbf{x}_0) = \frac{\sqrt{\mathbf{H}_{m,m}(\mathbf{x}_0,\mathbf{x}_0)}}{\sqrt{\left\langle \left(u^{(0)}(\mathbf{x}_r,f)\right)^2\right\rangle_{\mathbf{x}_r,f}}} \ , \tag{6.B.15}$$

with m the parameter associated with the dominant eigenvalue of the matrix $\mathcal{M}(\mathbf{x}_0)$ defined in equation (6.B.6). If the observable is the relative perturbation of the spectral amplitude, then the sensitivity kernel defined in equation (4.A.8) is already scaled by the incident wavefield, and the sensitivity can be measured as

$$k_{\mathrm{PSD}}(\mathbf{x}_0) = \sqrt{\mathbf{H}_{m,m}(\mathbf{x}_0,\mathbf{x}_0)} \ . \tag{6.B.16}$$

Finally, the sensitivity criterium S displayed in the main text is averaged over the target Σ_0 defined in decibel with respect to $k_0 = 1\ \mathrm{m}^{-2}$:

$$S = 10\log_{10}\left(\frac{\left\langle k(\mathbf{x})\right\rangle_{\mathbf{x}\in\Sigma_0}}{k_0}\right) \ . \tag{6.B.17}$$

6.C Quantitative criteria for assessing inversion quality

In order to be able to compare the a priori prediction of the best parametrisation based on the Hessian matrix analysis (see Appendix 6.B) to the actual inversion results, we also define two quantitative criteria for assessing the quality of an inverted model. The first criterion is based on the correct spatial location of the inclusion, and is analogous to C_S used for quantifying the spatial crosstalk (see equation 6.B.5). We call it the *detection criterion* for the parameter m_i (referring to a relative perturbation), measured in decibel:

$$D^{(m_i)} = 10\log_{10}\left(\frac{\sqrt{\mathrm{med}\left(\left\|\widehat{m_i}(\mathbf{x})\right\|^2\right)_{\mathbf{x}\in\Sigma_0}}}{\sqrt{\mathrm{med}\left(\left\|\widehat{m_i}(\mathbf{x})\right\|^2\right)_{\mathbf{x}\notin\Sigma_0}}}\right) \ , \tag{6.C.1}$$

with $\widehat{m_i}(\mathbf{x})$ the inverted model and Σ_0 the correct location of the inclusion. The second criterion measures the accuracy of the parameter estimation inside the inclusion. It is simply defined defined as the root mean square error inside the inclusion:

$$E^{(m_i)} = \sqrt{\left\langle \left(\widehat{m_i}(\mathbf{x}) - m_i^{\text{true}}(\mathbf{x}) \right)^2 \right\rangle_{\mathbf{x}\in\Sigma_0}} \,. \tag{6.C.2}$$

6.D Objective function sensitivity to the background model

In this work we present examples of scattered-wavefield inversion. This kind of inversion remains linear until the Born approximation holds. The latter can be considered as valid as far as the scattered wavefield is small enough compared to the incident wavefield (equation (4.9)). This is the case for our examples, so there are no non-linear effects such as cycle skipping, whatever the objective function. We introduce a new objective function J_{PSD}, which only takes the spectral amplitude into account, and compare it to the standard least-square migration objective function J_{LSM}, based on the full waveform (amplitude and phase). It is not surprising to find that the latter provides a better resolution (tests #3 vs. #4, section 6.4.3), since the phase information adds an extra constraint for the inversion. However, we did not address the question of the inversion sensitivity with respect to the background medium parameters. This is where the choice of the objective function starts to play a bigger role. Here we propose a simplified model to understand the behaviour of J_{PSD} when the background model velocity changes, and to compare it to J_{LSM}. We consider a source at the origin ($x_s = 0$), one single receiver located at the surface in x_r vertically above a very small scatterer reduced to one point (x_r, z). The reference receiver used in definitions (4.3) and (4.5) is located in x_0, far enough to consider it perceives only a negligible part of the scattered energy. We decompose the wavefield as in equation (4.9). Assuming $u^{(0)}$ is a plane wave arriving horizontally at the velocity V_0, and $u^{(1)}$ propagates from the scatterer towards the receiver as a single mode at the same velocity V_0, we have the following expressions of $u^{(0)}$ and $u^{(1)}$:

$$\begin{cases} u^{(0)}(x_r, f_0) = U_0 \exp\left(j\left(k_0 x_r + \phi_0\right)\right) \\ u^{(1)}(x_r, f_0) = \underline{\epsilon}(x_r, z, f_0) u^{(0)}(x_r, f_0) \exp\left(j k_0 z\right) \end{cases} , \tag{6.D.1}$$

with $\underline{\epsilon}$ is a scattering coefficient (possibly complex) derived from the Born approximation, and $k_0 = 2\pi f_0 / V_0$. Now, considering only the receiver at x_r at a single frequency f_0, the two objective functions to be compared are

$$\begin{cases} J_{\text{LSM}} = \left| u^{(1)}_{\text{obs}}(x_r, f_0) - u^{(1)}_{\text{calc}}(x_r, f_0) \right|^2 \\ J_{\text{PSD}} = \left| \eta'_{\text{obs}}(x_r, x_0, f_0) - \eta'_{\text{calc}}(x_r, x_0, f_0) \right|^2 \end{cases} . \tag{6.D.2}$$

The observed single-scattered wavefield is estimated as the difference between the total wavefield and the wavefield in the reference medium: $u^{(1)}_{\text{obs}} = u^{(\text{tot})}_{\text{obs}} - u^{(0)}_{\text{obs}}$. Let $u^{(0)}_{\text{obs}}$ be the true reference wavefield, and $u^{(0)}_{\text{err}}$ its forward-modelling estimation based on an erroneous background medium velocity V'_0. Then the calculated scattered wavefield

is $u^{(1)}_{\text{err}}$, obtained by inserting $u^{(0)}_{\text{err}}$ into the second equation in (6.D.1). Starting from the correct model, where $J = 0$, the deviation of the LSM objective function due to the error on the reference wavefield is

$$J_{\text{LSM}} = \left|u^{(0)}_{\text{obs}} - u^{(0)}_{\text{err}}\right|^2 + o\left(\left(u^{(0)}\right)^2\right). \tag{6.D.3}$$

After some basic algebra, we get

$$J_{\text{LSM}} \sim 2U_0^2 \left[1 - \cos\left(\left(k_0' - k_0\right) x_r\right)\right] \ , \tag{6.D.4}$$

with $k_0' = 2\pi f_0 / V_0'$. Thus J_{LSM} in this case is of the same order as $\left(u^{(0)}\right)^2$. Regarding J_{PSD}, we can first simplify its expression (6.D.2) to

$$J_{\text{PSD}} = \left| \frac{\Delta P_{\text{obs}}(x_r, f_0)}{P^{(0)}_{\text{obs}}(x_r, f_0)} - \frac{\Delta P_{\text{err}}(x_r, f_0)}{P^{(0)}_{\text{err}}(x_r, f_0)} \right|^2 \ , \tag{6.D.5}$$

where $P^{(0)}(x_r, f_0) = \left|u^{(0)}(x_r, f_0)\right|^2$ is the PSD in the reference medium, $\Delta P = \left|u^{(\text{tot})}(x_r, f_0)\right|^2 - \left|u^{(0)}(x_r, f_0)\right|^2$ is the PSD perturbation due to scattering, and P_{err} refers to the forward modelling in the erroneous background model. Equation (6.D.5) contains the aforementioned assumption $\Delta P(x_0, f_0) = 0$ for the remote reference station. To the first-order in $u^{(1)}$, the PSD perturbation is

$$\Delta P / P = 1 + \frac{2\,\mathfrak{Re}\left[u^{(1)} \cdot \left(u^{(0)}\right)^*\right]}{\left(u^{(0)}\right)^2} + o\left(\frac{u^{(1)}}{u^{(0)}}\right) \ . \tag{6.D.6}$$

Replacing this expression into (6.D.5) and using the simplified scattering model (6.D.1) yields the objective function deviation from zero:

$$J_{\text{PSD}} \sim 8\epsilon^2 \sin^2\left(\frac{k_0 + k_0'}{2} z\right) \sin^2\left(\frac{k_0 - k_0'}{2} z\right) \ , \tag{6.D.7}$$

with $\epsilon = |\underline{\epsilon}|$. Thus the modified objective function is of the second order in $u^{(1)}$. This fact alone implies that J_{PSD} is less sensitive to errors on the background model. Now, both objective functions present several minima, but they are not due to the same phenomenon. For J_{LSM}, they correspond to the cycle-skipping on the main arrival $u^{(0)}$, and are modulated by the source-receiver distance x_r. For J_{PSD}, they are due to alternating constructive/destructive interference between the incident and the scattered wavefields, and are modulated by the scatterer depth z. This means that for distant sources ($x_r \gg z$), typically relevant for surface wave recordings, the local minima of J_{PSD} due to errors on the background model velocity lie further from the true minimum than for J_{PSD}. Some examples are given in Fig. 6.D.1, where we take $f_0 = 1$ Hz, and scale both objective functions to their maximum values. J_{LSM} and J_{PSD} are shown for source-receiver and scatterer-receiver distances of 1 and 5 km, considering V_0 errors within ± 1000 m/s around the correct value of 2000 m/s. When the distance values are the same ($x_r - x_s = z$), the minima of both objective functions are at the same locations, and they are wider for J_{PSD}. For a realistic configuration, such as $(x_r - x_s) = 5$ km and $z = 1$ km, J_{PSD} presents a considerably wider attraction basin around the correct minimum. The counterpart of this stability is that J_{PSD} would be less efficient for updating the background model in a non-linear inversion workflow.

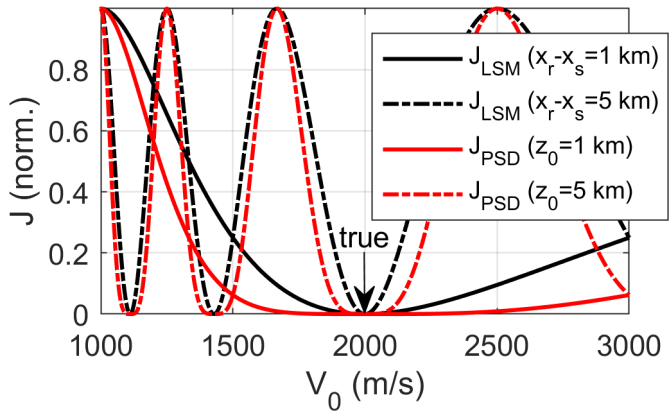


Figure 6.D.1: Leading terms of the objective functions evaluated by the simplified scattering model when the background model velocity V_0 deviates from the true value 2000 m/s.

Chapter 7

Conclusions and perspectives

Résumé (français)

Dans ce chapitre, on dresse le bilan de ce travail de thèse, on discute du potentiel pratique des développements réalisés, et on identifie les pistes pour poursuivre l'investigation des anomalies spectrales au-dessus des réservoirs multi-fluide. Une première conclusion est que l'on ne parvient pas à expliquer les observations de terrain avec la simple présence d'un réservoir traité en tant qu'hétérogénéité élastique. En revanche, une structure géologique anticlinale réaliste est en mesure de générer des anomalies spectrales ayant des ordres de grandeurs comparables à celles observées dans les données réelles, toujours dans le cadre de la modélisation élastique (chapitre 4). Cependant, les données réelles (en particulier la partie 2.1) suggèrent que la structure géologique n'est pas l'unique facteur expliquant la présence des anomalies, malgré un certain degré de corrélation. A partir de ces constats on identifie quatre scénarios quant aux potentiel et à la suite de l'étude menée:

1. La modélisation utilisée est bien adaptée: c'est la structure géologique qui est le facteur dominant à l'origine des anomalies observées.

   Alors il faudra étendre l'inversion linéarisée développée au chapitre 6 à un schéma d'inversion non-linéaire apte à traiter les fortes perturbations. On pourra alors espérer exploiter les anomalies spectrales pour explorer les structures géologiques du sous-sol, telles que les anticlinaux.

2. La modélisation utilisée est bien adaptée, mais les valeurs des hétérogénéités au sein du réservoir ont été sous-estimées (partie 1.6.2). En réalité, l'effet fluide dans le réservoir contribue de manière non-négligeable aux anomalies spectrales observées.

   Des simulations montrent en effet qu'en prenant des perturbations élastiques de -50% au sein du réservoir on parvient à retrouver les bons ordres de grandeur pour les anomalies spectrales. L'ajout d'un contraste d'atténuation pourrait également contribuer à renforcer la réponse du réservoir. Si l'on suit cette piste, la problématique se déplace vers la physique des roches: quelles sont les propriétés mécaniques d'une roche poreuse saturée par un mélange multiphasique à des fréquences de quelques Hertz ?

3. La modélisation utilisée est bien adaptée, mais la diffraction doit être décrite en 3D.

   Etudier cette piste suppose de répéter les simulations numériques en 3D, ce qui est assez couteux avec la méthode des éléments spectraux. Par ailleurs, des résultats dans la littérature montrent qu'il n'y a pas de raison d'attendre des anomalies plus fortes en 3D qu'en 2D, même si leur forme précise est différente (Gorbatikov & Tsukanov, 2011).

4. L'ensemble de la modélisation est à revoir: les anomalies observées ne sont pas liées à la diffraction des ondes de surface incidentes cohérentes à l'échelle du réseau de sismomètres, détectées dans le chapitre 3.

   Dans ce cas il faudrait traiter le champ incident en termes de champ diffus, généré par un grand grand nombre de diffractions qui ont lieu à l'échelle du réseau de sismomètres. Une telle description est intéressante car la densité spectrale de puissance peut être liée de manière simple à la fonction de Green en pseudo-réflexion du milieu (Sánchez-Sesma et al., 2011a), ce qui permet une modélisation facile permettant de tester cette hypothèse. Cette piste est définitivement à explorer.

The motivation for my PhD work was to investigate the feasibility of exploiting the ambient noise amplitude anomalies at frequencies of several Hertz for reservoir exploration and monitoring. This technique is referred to as *Low Frequency Passive Seismic* (LFPS). As part of this PhD work, I conducted several field experiments and performed some modelling work.

7.1 Experimental feasibility

Exploration

In terms of exploration, the three datasets analysed in Chapter 2 exhibited anomalies spatially coinciding with the reservoir location. For the two UGS facilities in the Paris Basin (Chémery and Saint-Illiers), the spectral signatures presented similarities. They mainly consisted of the amplification of the vertical component motion between 1 and 2-4 Hz, consistently with the so-called *hydrocarbon microtremors* (HMT) reported in the literature (see Section 1.2). For the case of the steam-rich geothermal field in Muara Laboh (Indonesia), the signature was different, as it consisted of a strong *attenuation* of the *horizontal* motion. Experimental results in the exploration context can be regarded as encouraging, as they are likely to help the correct well implantation, given the apparent correlation of the anomalies with gas-bearing zones.

Monitoring

In terms of monitoring, my conclusion is that the LFPS technique cannot be used as an indicator of the gas front reaching the seismometer location. This conclusion was formulated in Section 2.2 by analysing the signals recorded at a peripheral well intermittently reached by the gas front. The HMT signature appears as a rather stationary feature. One should bear in mind that the gas volume evolution is very low in the reservoir (below 15% over a full cycle). Nevertheless, a significant evolution of the amplitude anomaly with time is observed at some locations at the periphery of the gas reservoir, and the correlation of this evolution with the gas injection/withdrawal must be studied over several cycles to be able to conclude.

In terms of modelling, however, *my work does not allow to understand the mechanism behind the HMT*, and namely the amplification of the vertical component, as summarised in the next section.

7.2 What to expect from the elastic modelling methodology ?

I investigated the elastic scattering of Rayleigh waves contained in the ambient noise as a possible mechanism of the vertical motion amplification. Rayleigh waves were indeed found to carry a significant part of the wavefield energy (Chapter 3). It turns out that this mechanism does not allow to explain hydrocarbon microtremors by the presence of a reservoir, the magnitude of the modelled response being too small (Chapter 4). Nevertheless, this work demonstrates the potential of using *small* Rayleigh

amplitude perturbations in a quantitative way (Chapter 6). Perhaps does it bring a new contribution to the microseismic sounding method (Gorbatikov et al., 2008). This section describes the potential application scope of the methodology developed during my PhD work.

An important conclusion of the elastic modelling is that structural effects due to an anticline, such as in Chémery, are about one order of magnitude stronger than the reservoir effects, if the latter is modelled with a realistic elastic contrast. Four different hypotheses can be formulated based on this result:

1. The modelling approach is correct, but the presumed HMT are in fact due to structural features often correlated with the presence of reservoirs.

2. The modelling approach is correct, but the constrast in the reservoir is underestimated, for example by neglecting viscoelasticity.

3. The modelled mechanism is correct, but the modelling approach must be more realistic.

4. The modelled mechanism is irrelevant, and the measured amplitudes are not due to the scattering of coherent Rayleigh waves propagating across the receiver array.

The consequences are analysed in the following sections.

Geological structure exploration

Assuming the first hypothesis, the developed methodology could possibly be used for mapping the structural deformations below the array, for example in order to detect anticline tops. As these are strong perturbations, the Born approximation fails to treat the associated amplitude distortions as a result of single scattering (see Chapter 4). A non-linear inversion algorithm must be implemented as an extension of the Chapter 6 to achieve this goal. This would naturally account for multiple scattering, as the background model would be updated at each iteration. The risk, however, would be to converge in a local minimum. We refer to the discussion in Chapter 6 for further details on this subject.

Time-lapse reservoir monitoring

Further on, assuming the first hypothesis, amplitude differences between time-lapse acquisitions remove the strong amplitude perturbations due to the structure. The remaining effects should be due to the medium evolution, and possibly the gas volume evolution in the reservoir. For these small perturbations, the developed methodology could in theory be applied, provided some assumptions concerning the 2D wavefield approximation and the ambient noise stability However, the accuracy of the ambient noise amplitude measurements is a major limitation to such applications. In Chapter 4, we found maximum power spectral density perturbations of the order of 20% due to a reservoir placed within a realistic structure. The inversion procedure in Chapter 6 relies on matching amplitude perturbations as low as 1%. That is, the ambient noise relative amplitude anomalies should be measured with an accuracy of

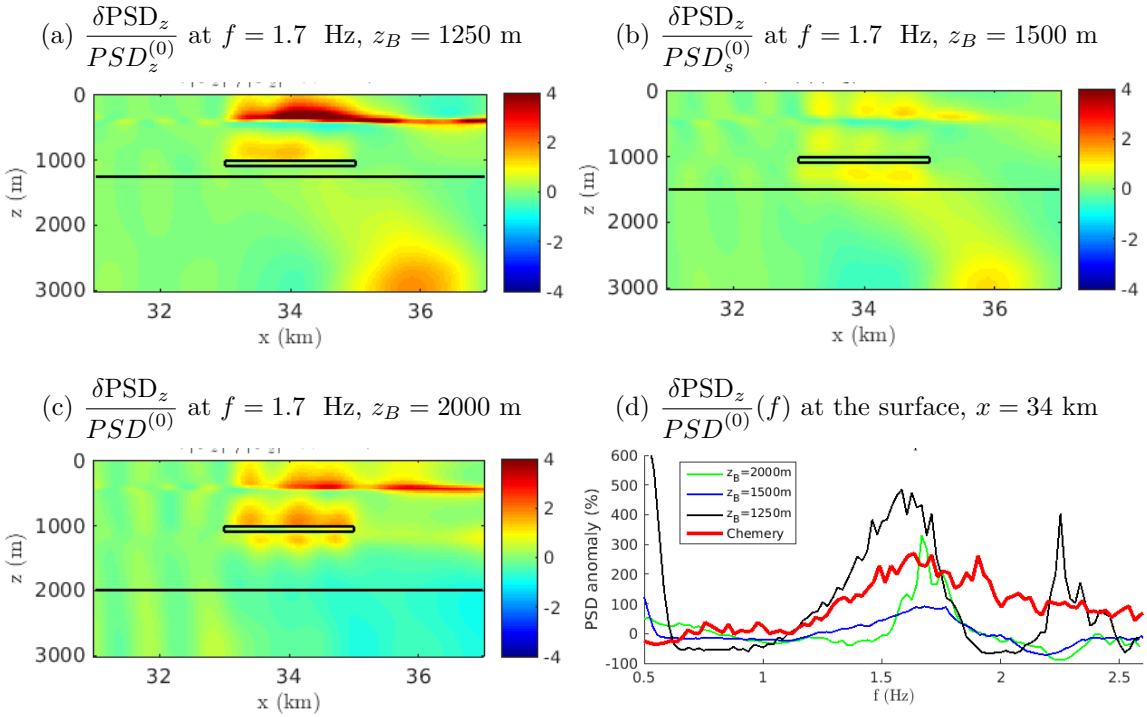


Figure 7.1: Modelled $\delta \mathrm{PSD}_z / PSD_z^{(0)}(f)$ (anomaly between two simulations, without reference station) with a strong constrast in the reservoir. (a-c) Spatial distribution of the PSD perturbations for different values of bedrock depth. (d) Spectral anomalies recorded above the middle of the reservoir for the (a), (b) and (c) (respectively black, blue and green), compared to the anomaly pattern recorded in Chémery (red line: anomaly $\eta(\mathbf{x_r}, \mathbf{x_0})$ with a reference station far from the reservoir, extracted from Fig. 1.2a).

the order of 1%. The results of continuous monitoring in Saint-Illiers (Chapter 2.2) exhibit uncertainties of the order of 100% (see temporal fluctuations in Fig. 2.23). Thus a huge progress in signal processing algorithms would be needed for a reservoir monitoring objective, assuming the wavefield modifications as small as predicted by elastic modelling.

7.3 Effects of stronger contrasts in the reservoir

Now assuming the second hypothesis (underestimated contrasts in the reservoir), I have carried out some additional numerical tests to investigate the effect of a very strong elastic contrast in the reservoir ($\delta V_p/V_{p,0} = \delta\rho/\rho_0 = \delta V_s/V_{s,0} = -50\%$) placed in a layer-over-halfspace model. The latter had nearly the same parameters as the one used in Chapter 4. The random distribution of deep sources was on the left side of the model. This test was performed at an early stage of the PhD work, before the dispersion curve inversion in Section 3.4.4. This is why the bedrock was taken shallower, between 1250 and 2000 m, based on the general knowledge of the Paris basin. Results are displayed in Fig. 7.1.

Figs 7.1a-c show the spatial distribution of the PSD anomaly at 1.7 Hz. It is clear that a strong positive anomaly is generated near the surface. Interestingly, the anomaly near the surface is stronger than near the reservoir, which acts as a secondary source. Testing different bedrock depths does not fundamentally affect this behaviour. The modelled spectral anomalies at the surface above the reservoir are shown in Fig 7.1(d) as function of frequency. They are compared to the anomaly recorded in Chémery (red line), corresponding to the relative difference of the two spectra in Fig. 1.2a. The simulation allows to retrieve the correct order of magnitude and the correct frequency content. This result is interesting, but the elastic contrasts in the reservoir are obviously unrealistic. This motivates the inclusion of viscoelasticity in the reservoir modelling, which could increase the reflectivity of the reservoir without taking unrealistically strong elastic contrasts. Such modelling was already performed by Lambert et al. (2013), but with a different source distribution. Guidelines for the visco-elastic extension were given in Section 1.6.

7.4 Improving the Rayleigh wave scattering modelling

Considering the third hypothesis, the main limitation of the scattering modelling performed in this work is its 2D character. As mentionned in the previous chapters, 3D scattering of Rayleigh waves is expected to present some difference with respect to the 2D case.

Difference between 2D and 3D scattering

This question was addressed by Gorbatikov & Tsukanov (2011) in a homogeneous elastic half-space. Their results are displayed in Fig. 7.2. in terms of amplitude perturbations above an embedded cylindrical scatterer, which is a rectangle projected in 2D. It is clear that the amplitude perturbations are better focused above the scatterer in 3D. However, the amplitude perturbation in 2D is reported 1.4 dB (about 50%) stronger than in 3D by Gorbatikov & Tsukanov (2011). This means that strong amplitude anomalies (comparable to the recorded ones) will be even harder to obtain in 3D.

Extension of the numerical modelling to 3D

Undertaking an extension of the developed methodology to 3D contains two main difficulties. The first is the computational cost of 3D spectral-element simulations. Komatitsch & Vilotte (1998) compared the cost of a 2D simulation on a grid of 50x30=1500 elements to a 3D simulation on a grid of 26x26x14=9464 elements, with the same time step. The cost in 3D was multiplied by 45, accounting for the different number of parallel cores involved in their 2D and 3D examples. In our case, the 2D grid of the realistic model, containing shallow layers and and anticline structure (Chapter 4), is composed of $\sim$ 167.000 mesh elements (quadrangles). Simulating one time window (about 1 minute) of ambient noise takes about 45 minutes with a parallelisation over 100 cores of the Storengy computation cluster. Such a simulation in 3D would definitively take several days just for one ambient noise realisatation, which

is a serious limitation.

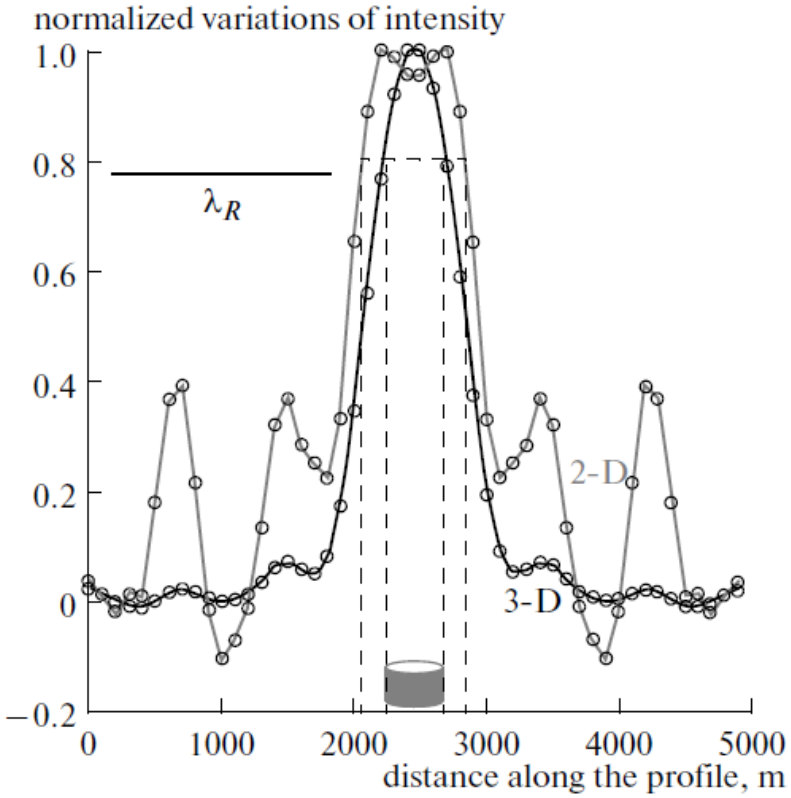


Figure 7.2: Difference between the spectral amplitude perturbations in 2D and 3D modelling. From Gorbatikov & Tsukanov (2011).

To some extent, this simulation time can be reduced by changing the ambient noise modelling strategy presented in Chapter 4. In the present implementation, all the random sources are explicitely simulated, as well as the 20 ambient noise windows. However, as the sources are distant and kept at constant depth, the signals recorded by the receivers can possibly be modelled as randomly time-delayed versions of a single Green's function, convolved with Ricker wavelets of random central frequencies. If this approximation is valid, a single-source simulation would be sufficient to reproduce the results of Chapter 4. However, a potential limitation is that the shape of the simulated dispersive wave trains above the reservoir depends on the distance to the source. If this dependence is significant over the lateral extent of the simulated source zone, time-shifting the Green's function might not be a satisfactory solution. I did not investigate this point during this PhD work.

An additional burden in 3D comes from the source distribution, which must now be randomised over an azimuth interval. Simulations for different source azimuths can be either performed simultaneously or sequentially. The first option is as time consuming as a single source simulation. However, the amplitude above the reservoir could suffer from a strong comb effect, which motivated the averaging over different time windows in Chapter 4. The second option is the one adopted by Gorbatikov & Tsukanov (2011), as they averaged the *intensities* for the different source azimuths

(and not the time domain wavefields). This option requires repeated sequential 3D simulations. Contrary to the previous paragraph, rotating a single Green's function to account for the source azimuth variation is not compatible with studying a 3D geological structure/reservoir.

The large size of the models which were used in this work is essentially due to the need of simulating well-separated Rayleigh waves and avoiding bias due to body waves. It would be useful to study down to which limit the model size can be reduced while preserving the same spectral anomalies as described in Chapter 4. This could greatly contribute to the reduction of the computation time in 3D. Finally, spectral-element simulations are more time-consuming than other techniques, such as the finite-differences (see Virieux et al., 2011).

Born sensitivity kernels in 3D

An important advantage of the 2D modelling was the easiness of the Born approximation implementation, due to an explicit storage of the Fréchet derivatives in the full model. This allowed a Gauss-Newton formulation of the inversion (see Chapter 6) and a parametrisation analysis based on the full Hessian. The size of a 2D Hessian was of 6.6 GB in Chapter 6. In 3D, adding additional 4000 m in the tranverse direction results in 160 aditionnal $x - z$ slices, thus about 1 TB weight for the Hessian matrix, and even more for the Fréchet derivatives, as they are not summed over receivers and frequencies. Such a memory demand is of course unreasonable. Matrix-free techniques, such as the adjoint-state method (Plessix, 2006) should be implemented (see Chapter 6 for discussion).

Accounting for body waves and ambient noise azimuthal distribution

It was shown in Chapter 3 that a quick phase identified as P_g also affected the ambient noise composition between 1 and 1.5-1.8 Hz, which lies within the presumed HMT frequency band. It could also be a P_n phase, since the resolution of the array used in Chapter 3 was insufficient to unambiguously distinguish both phases. In any case, the scattering of this type of waves should also be taken into account. An incident wave dominated by P-waves with apparent velocities close to the what was detected in Chapter 3 (i.e. 6-8 km/s) can be obtained by using distant deep explosive sources in the numerical simulation, instead of vertical point forces. Assuming all the modes propagating across the array are uncorrelated, the total incident power spectral density at a given receiver can be written as

$$\mathrm{PSD}^{(0)}(\mathbf{x}, \omega) = \sum_i \mathrm{PSD}_i^{(0)}(\mathbf{x}, \omega). \tag{7.1}$$

Similarly, the scattered wavefield can be decomposed in the contributions of the different modes. The amplitude sensitivity kernel K_i can be numerically computed for each mode following the procedure described in Chapter 4, by adapting the source distribution characteristics. The azimuthal distribution can differ from one propagation mode to another (e.g. Figs 3.6 and 3.7). In case of 3D extension, each mode should be taken into account with its back-azimuth distribution, e.g. the beam power

displayed in e.g. Fig. 3.7, which we write as $B_i(\theta)$. Then new formulation of the Born approximation (equation 4.10) becomes

$$\frac{\delta \mathrm{PSD}}{\mathrm{PSD}^{(0)}}(\mathbf{x},\omega) = \sum_i \alpha_i(\omega) \int_\Omega K_i(B_i(\theta),\mathbf{x},\omega) \frac{\delta m}{m_0}(\mathbf{x}) d\mathbf{x} \ , \tag{7.2}$$

with $\alpha_i = \mathrm{PSD}_i^{(0)}/\mathrm{PSD}^{(0)}$ the power proportion of each mode in the incident wavefield. The dependence of K_i on $B_i(\theta)$ means that the density of the random sources used in the numerical simulation is proportional to the measured beam power in the ambient wavefield. This results in the definition of a «total» sensitivity kernel

$$K_{tot}(\mathbf{x},\omega) = \sum_i \alpha_i(\omega) K_i(B_i(\theta),\mathbf{x},\omega) \tag{7.3}$$

accounting for all the wave types and azimuthal distributions. Measuring α_i, however, is not an easy task. As starting point, using the relative proportions between the «power» values attributed to the different modes in the dispersion plots such as in Fig. 3.5 could be a good compromise. However, the beam-formers of the array methods such as FK or MUSIC are not necessarily proportional to the physical power of the signal, but rather measure a likelihood for different candidate wavenumbers (Capon et al., 1967; Schmidt, 1986; Goldstein & Archuleta, 1987). Measuring a «true» power with MUSIC requires to identify one eigenvector per mode (Goldstein & Archuleta, 1987), which is complicated in the case of the ambient noise. In Appendix 3.B, we saw that several eigenvectors must generally be considered as «signal» in the MUSIC implementation in order to obtain a satisfactory result. On the other hand, standard beam-forming techniques (Rost & Thomas, 2002), which yield a true measure of the signal power, do not have enough resolution for a satisfactory mode separation with an array of the Chémery type (see Chapter 3).

Including viscoelasticity in the background medium

As a last improvement direction, adding a finite quality factor into the background model could also improve the modelling workflow. The effect would be twofold. First, as shallow layers have a lower quality factor compared to deep layers, the incident wavefield magnitude in depth would be enhanced compared to the amplitude at the surface. As the scattering by the reservoir happens in a relatively deep part of the model, the strength of the secondary source compared to the incident wavefield intensity would increase. This could result in an increase of the relative amplitude perturbation measured at the surface. Second, the low quality factors in the shallow layers are likely to affect the scattered wavefield by filtering the high frequencies, as these undergo more oscillation cycles on their way from the reservoir to the surface.

7.5 Diffuse or coherent incident wavefield ?

Finally, the fourth hypothesis in Section 7.2 was that coherent wave scattering is not the relevant mechanism for describing the observed vertical motion amplification. Invoking a dominantly diffuse field could be an alternative. In the Chémery data, we observed that ballistic arrivals are hard to identify in the individual cross correlations

above 1 Hz when the inter-station distance increases beyond 1 km (not shown in this manuscript). This indicates that an increasing part of the wavefield is dominated by the coda, i.e. waves multiply scattered by relatively small-scale model heterogeneities. As the layered models used in our numerical simulations were piece-wisely homogeneous, this effect was ignored.

Principles of the Diffuse Wavefield Assumption

Several authors have recently proposed to use the *diffuse wavefield assumption* (DFA) to model the H/V spectral ratio at frequencies of several Hz (Sánchez-Sesma et al., 2008; García-Jerez et al., 2011; Sánchez-Sesma et al., 2011b,c,a; García-Jerez et al., 2016; Perton et al., 2018). This includes our problem, as both horizontal and vertical spectral amplitudes must be predicted for modelling the H/V ratio. The main idea of the DFA theory is similar to the ambient noise cross-correlation in the limit when the virtual source and the receiver are located at the same location. In a diffuse field, an autocorrelation, or equivalently the PSD at an individual station, can then be interpreted in terms of the imaginary part of the Green's function between the receiver location $\mathbf{x_r}$ and itself (pseudo-reflection response)

$$\begin{aligned} \langle u_i(\mathbf{x_r},\omega)u_i^*(\mathbf{x_r},\omega)\rangle &= -4\pi E_S k^{-2}\mathfrak{Im}\left[G_{ii}(\mathbf{x_r},\mathbf{x_r},\omega)\right] \text{ in 2D,} \\ \langle u_i(\mathbf{x_r},\omega)u_i^*(\mathbf{x_r},\omega)\rangle &= -2\pi E_S k^{-1}\mathfrak{Im}\left[G_{ii}(\mathbf{x_r},\mathbf{x_r},\omega)\right] \text{ in 3D,} \end{aligned} \tag{7.4}$$

with E_S the energy density of shear waves in the diffuse field and k the shear wavenumber. A diffuse field can be defined as a wavefield where all the available propagation modes (body and surface waves) are represented with equal energies (Weaver, 1982). DFA assumes the energy is equipartitionned among both surface and body wave modes. The «standard» ambient noise theory often assumes equipartition among surface waves only, and the retrieved Green's functions are then those of the Helmholtz equation (1.19) (e.g. Boschi et al., 2013). On the contrary, the detection of reflected body waves in ambient noise cross-correlations (Draganov et al., 2009; Ruigrok et al., 2011) can be regarded as a non-zero offset version of DFA. The validity of surface/body wave vision of the equipartion depends on frequency and the scattering characteristics of the medium. We can expect the validity of the body wave representation to increase at higher frequencies, as the wavelength becomes closer to the characteristic size of the heterogeneities. In the same time, the surface wave terms of the Green's function become harder to reconstruct.

Potential for DFA application to the hydrocarbon microtremors

Compared to the coherent wavefield propagation modelling developped during this PhD work, DFA offers the advantage of being simple to implement and extremely cheap computationally, both in 2D and 3D. Only the Green's function $G_{ii}(\mathbf{x_r},\mathbf{x_r},\omega)$ must indeed be modelled in the forward problem under DFA. This allows to take a very small mesh compared the one required for simulations with distant sources. On the other hand, the prefactor in front of the Green's function in equation (7.4) would disappear when considering normalised amplitude anomalies such as η introduced in equation 4.3. This also offers a simple modelling procedure for the V/H ratio, another attribute which can be correlated to a multi-fluid reservoir. However, from

a quantitative perspective, it is unclear how the energy balance evolves between the ballistic and the diffuse parts of the wavefield as the frequency increases, i.e. starting from which frequency DFA becomes relevant. As a rough estimate, the ratio of the signal subspace eigenvalues determined by MUSIC in Chapter 3 over the sum of all the eigenvalues can be used to quantify the proportion of coherent (non-diffuse) energy in the wavefield. This frequency-dependent proportion could be used to weight the contributions of wavefield energies modelled under the distant-sources assumption (low frequencies) and DFA (high frequencies). Interestingly, if HMT were to be explained by DFA (i.e. the pseudo-recflection response of the medium containing a reservoir, comparable to a bright-spot in reflection seismic), the day/night variations of the ambient noise power due to the human activity would not affect the spectral anomalies, and would only concern the pre-factor E_S in equation (7.4). This would be compatible with the observations in Saint-Illiers (Section 2.2).

Bibliography

Aki, K., 1957. Space and Time Spectra of Stationary Stochastic Waves, with Special Reference to Microtremors, *Bulletin of Earthquake Research Institute*, **35**, 415–456.

Aki, K. & Richards, P. G., 2002. *Quantitative Seismology*, University Science Books.

Ali, M. Y., Berteussen, K. A., Small, J., & Barkat, B., 2010. Low-frequency passive seismic experiments in Abu Dhabi, United Arab Emirates: implications for hydrocarbon detection: Low-frequency passive seismic experiments, *Geophysical Prospecting*, **58**(5), 875–899.

Almendros, J., Ibanez, J. M., Alguacil, G., & Del Pezzo, E., 1999. Array analysis using circular-wave-front geometry:an application to locate the nearby seismo-volcanic source, *Geophysical Journal International*, **136**(1), 159–170.

Ardhuin, F., Stutzmann, E., Schimmel, M., & Mangeney, A., 2011. Ocean wave sources of seismic noise, *Journal of Geophysical Research*, **116**(C9).

Artemieva, I. M. & Thybo, H., 2013. A seismic model for Moho and crustal structure in Europe, Greenland, and the North Atlantic region, *Tectonophysics*, **609**, 97–153.

Artman, B., Podladtchikov, I., & Witten, B., 2010. Source location using time-reverse imaging: Source location using time-reverse imaging, *Geophysical Prospecting*, **58**(5), 861–873.

Bard, P.-Y., 1999. Microtremor measurements: a tool for site effect estimation?, in *Second International Symposium on the Effects of Surface Geology on seismic motion*, vol. 3, pp. 1251–1279, Balkema.

Bensen, G. D., Ritzwoller, M. H., Barmin, M. P., Levshin, A. L., Lin, F., Moschetti, M. P., Shapiro, N. M., & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, **169**(3), 1239–1260.

Beresnev, I. & Johnson, P., 1994. Elastic-wave stimulation of oil production: A review of methods and results, *Geophysics*, **59**(6), 1000–1017.

Biot, M., 1962. Mechanics of Deformation and Acoustic Propagation in Porous Media, *Journal of Applied Physics*, **33**(4).

Biot, M. A., 1941. General Theory of Three-Dimensional Consolidation, *Journal of Applied Physics*, **12**(2), 155–164.

Biot, M. A., 1956. Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range, *The Journal of the acoustical Society of america*, **28**(2), 168–178.

Birialtsev, E. V., Plotnikova, I. N., Khabibulin, I. R., & Shabalin, N. Y., 2006. The Analysis of Microseisms Spectrum at Prospecting of Oil Reservoir on Republic Tatarstan, in *2nd EAGE St Petersburg International Conference and Exhibition on Geosciences*.

Bishop, C. M., Svensén, M., & Williams, C. K. I., 1998. GTM: The Generative Topographic Mapping, *Neural Computation*, **10**(1), 215–234.

Bloch, G. & Akrawi, K., 2006. Application of Passive Seismic (IPDS) Surveys in Arabian Peninsula, in *First EAGE Passive Seismic Workshop - Exploration and Monitoring Applications*.

Bokelmann, G. H. & Baisch, S., 1999. Nature of narrow-band signals at 2.083 Hz, *Bulletin of the Seismological Society of America*, **89**(1), 156–164.

Bonijoly, D., Barbier, J., Robelin, C., Kervevan, C., Thiery, D., Menjoz, A., Matray, J., Cotiche, C., & Herbrich, B., 2003. Feasibility of CO2 storage in geothermal reservoirs: example of the Paris Basin - France, Tech. Rep. BRGM/RP-52349-FR, BRGM-CFG-ANTEA.

Bonnefoy-Claudet, S., Cornou, C., Bard, P.-Y., Cotton, F., Moczo, P., Kristek, J., & Fäh, D., 2006a. H/V ratio: a tool for site effects evaluation. Results from 1-D noise simulations, *Geophysical Journal International*, **167**(2), 827–837.

Bonnefoy-Claudet, S., Cotton, F., & Bard, P.-Y., 2006b. The nature of noise wavefield and its applications for site effects studies, *Earth-Science Reviews*, **79**(3-4), 205–227.

Bormann, P. & Wielandt, E., 2013. *Seismic Signals and Noise*, GFZ Potsdam/IASPEI.

Boschi, L., Weemstra, C., Verbeke, J., Ekstrom, G., Zunino, A., & Giardini, D., 2013. On measuring surface wave phase velocity from station-station cross-correlation of ambient signal, *Geophysical Journal International*, **192**(1), 346–358.

Bouchon, M., 1979. Discrete wave number representation of elastic wave fields in three-space dimensions, *Journal of Geophysical Research: Solid Earth*, **84**(B7), 3609–3614.

Brenguier, F., Shapiro, N. M., Campillo, M., Nercessian, A., & Ferrazzini, V., 2007. 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations, *Geophysical Research Letters*, **34**(2).

Broadhead, M. K., 2010. Oscillating oil drops, resonant frequencies, and low-frequency passive seismology, *Geophysics*, **75**(1), O1–O8.

Broomhead, D. S. & Lowe, D., 1988. Multivariable Functional Interpolation and Adaptive Networks, *Complex Systems*, **2**.

Brossier, R., Operto, S., & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.

Cadoret, T. & Poirier, J.-P., 1993. *Effet de la saturation eau-gaz sur les proprietes acoustiques des roches : Etude aux frequences sonores et ultrasonores*, Ph.d. thesis, Paris 7.

Campman, X. & Riyanti, C., 2007. Non-linear inversion of scattered seismic surface waves, *Geophysical Journal International*, **171**, 1118–1125.

Campman, X., Wijk, K., Riyanti, C., Scales, J., & Herman, G., 2004. Imaging scattered seismic surface waves, *Near Surface Geophysics*, **2**, 223–230.

Campman, X. H., van Wijk, K., Scales, J. A., & Herman, G. C., 2005. Imaging and suppressing near-receiver scattered surface waves, *Geophysics*, **70**(2), V21–V29.

Capon, J., 1969. High-resolution frequency-wavenumber spectrum analysis, *Proceedings of the IEEE*, **57**(8), 1408–1418.

Capon, J., Greenfield, R., & Kolker, R., 1967. Multidimensional maximum-likelihood processing of a large aperture seismic array, *Proceedings of the IEEE*, **55**(2), 192–211.

Chai, H.-Y., Phoon, K.-K., Goh, S.-H., & Wei, C.-F., 2012. Some theoretical and numerical observations on scattering of Rayleigh waves in media containing shallow rectangular cavities, *Journal of Applied Geophysics*, **83**, 107–119.

Chapman, S., Tisato, N., Quintal, B., & Holliger, K., 2016. Seismic attenuation in partially saturated Berea sandstone submitted to a range of confining pressures, *Journal of Geophysical Research: Solid Earth*, **121**(3), 1664–1676.

Chapman, S., Quintal, B., Tisato, N., & Holliger, K., 2017. Frequency scaling of seismic attenuation in rocks saturated with two fluid phases, *Geophysical Journal International*, **208**(1), 221–225.

Chen, X., 1993. A systematic and efficient method of computing normal modes for multilayered half-space, *Geophysical Journal International*, **115**(2), 391–409.

Cornou, C., 2002. *Traitement d'antenne et imagerie sismique dans l'agglomération grenobloise (Alpes françaises) : implications pour les effets de site*, Ph.D. thesis, Laboratoire de Géophysique Interne et Tectonophysique.

Cros, E., Roux, P., Vandemeulebrouck, J., & Kedar, S., 2011. Locating hydrothermal acoustic sources at Old Faithful Geyser using Matched Field Processing: Old Faithful viewed through array processing, *Geophysical Journal International*, **187**(1), 385–393.

Dahlen, F. A. & Baig, A. M., 2002. Fréchet kernels for body-wave amplitudes, *Geophysical Journal International*, **150**(2), 440–466.

Dahlen, F. A., Hung, S.-H., & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophysical Journal International*, **141**(1), 157–174.

Dangel, S., Schaepman, M., Stoll, E., Carniel, R., Barzandji, O., Rode, E.-D., & Singer, J., 2003. Phenomenology of tremor-like signals observed over hydrocarbon reservoirs, *Journal of Volcanology and Geothermal Research*, **128**(1-3), 135–158.

Douze, E. J., 1964. Rayleigh waves in short-period seismic noise, *Bulletin of the Seismological Society of America*, **54**(4), 1197–1212.

Douze, E. J., 1967. Short-period seismic noise, *Bulletin of the Seismological Society of America*, **57**(1), 55–81.

Draganov, D., Wapenaar, K., Mulder, W., Singer, J., & Verdel, A., 2007. Retrieval of reflections from seismic background-noise measurements, *Geophysical Research Letters*, **34**(4).

Draganov, D., Campman, X., Thorbecke, J., Verdel, A., & Wapenaar, K., 2009. Reflection images from ambient seismic noise, *Geophysics*, **74**(5), A63–A67.

Duclos, M., Artman, B., Birkelo, B., Huguet, F., Dutzer, J. F., & Habiger, R., 2011. Low Frequency Seismic Survey at a Gas Storage Reservoir, in *73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011*.

Ekström, G., Tromp, J., & Larson, E. W. F., 1997. Measurements and global models of surface wave propagation, *Journal of Geophysical Research: Solid Earth*, **102**(B4), 8137–8157.

Endrun, B., 2011. Love wave contribution to the ambient vibration H/V amplitude peak observed with array measurements, *Journal of Seismology*, **15**(3), 443–472.

Essen, H.-H., Krüger, F., Dahm, T., & Grevemeyer, I., 2003. On the generation of secondary microseisms observed in northern and central Europe., *Journal of Geophysical Research: Solid Earth*, **108**(B10).

Farrell, W. E. & Munk, W., 2008. What do deep sea pressure fluctuations tell about short surface waves?, *Geophysical Research Letters*, **35**(19).

Ferrick, M. G., Qamar, A., & St Lawrence, W. F., 1982. Fluid dynamic analysis of volcanic tremor, Tech. rep., COLD REGIONS RESEARCH AND ENGINEERING LAB HANOVER NH.

Fink, M., Cassereau, D., Derode, A., Prada, C., Roux, P., Tanter, M., Jean-Louis Thomas, & Wu, F., 2000. Time-reversed acoustics, *Reports on Progress in Physics*, **63**(12), 1933.

Forgoes, E. & Lambaré, G., 1997. Parameterization study for acoustic and elastic ray+born inversion, *J Seism Explor*, **6**, 253–277.

Foti, S., Parolai, S., Albarello, D., & Picozzi, M., 2011. Application of Surface-Wave Methods for Seismic Site Characterization, *Surveys in Geophysics*, **32**(6), 777–825.

Frehner, M., Schmalholz, S. M., & Podladchikov, Y., 2009. Spectral modification of seismic waves propagating through solids exhibiting a resonance frequency: a 1-D coupled wave propagation-oscillation model, *Geophysical Journal International*, **176**(2), 589–600.

Friederich, W., Wielandt, E., & Stange, S., 1993. Multiple forward scattering of surface waves: comparison with an exact solution and Born single-scattering methods, *Geophysical Journal International*, **112**(2), 264–275.

Gal, M., Reading, A., Ellingsen, S., Koper, K., Burlacu, R., & Gibbons, S., 2016. Deconvolution enhanced direction of arrival estimation using one- and three-component seismic arrays applied to ocean induced microseisms, *Geophysical Journal International*, **206**(1), 345–359.

García-Jerez, A. & Sánchez-Sesma, F. J., 2015. Slowly-attenuating P-SV leaky waves in a layered elastic halfspace. Effects on the coherences of diffuse wavefields, *Wave Motion*, **54**, 43–57.

García-Jerez, A., Luzón, F., Sánchez-Sesma, F. J., Albarello, D., Lunedei, E., Campillo, M., Santoyo, M. A., & Iturrarán-Viveros, r., 2011. Comparison between two methods for forward calculation of ambient noise H/V spectral ratios, in *AGU Fall Meeting*.

García-Jerez, A., Piña-Flores, J., Sánchez-Sesma, F. J., Luzón, F., & Perton, M., 2016. A computer code for forward calculation and inversion of the H/V spectral ratio under the diffuse field assumption, *Computers & Geosciences*, **97**, 67–78.

Gassmann, F., 1951. *Über die Elastizität poröser Medien*, Inst. für Geophysik an der ETH.

Gedge, M. & Hill, M., 2012. Acoustofluidics 17: Theory and applications of surface acoustic wave devices for particle manipulation, *Lab on a Chip*, **12**(17), 2998.

Gelis, C., Leparoux, D., Virieux, J., Bitri, A., Operto, S., & Grandjean, G., 2005. Numerical Modeling of Surface Waves Over Shallow Cavities, *Journal of Environmental & Engineering Geophysics*, **10**(2), 111–121.

Gerivani, H., Haghshenas, E., Moghaddas, N. H., Ghafoori, M., & lashkaripour, G. R., 2012. Frequency–amplitude range of hydrocarbon microtremors and a discussion on their source, *Journal of Geophysics and Engineering*, **9**(6), 632.

Geuzaine, C. & Remacle, J.-F., 2009. Gmsh: A 3-D finite element mesh generator with built-in pre-and post-processing facilities, *International journal for numerical methods in engineering*, **79**(11), 1309–1331.

Gimbert, F. & Tsai, V. C., 2015. Predicting short-period, wind-wave-generated seismic noise in coastal regions, *Earth and Planetary Science Letters*, **426**, 280–292.

Gjevik, B., 1974. Ray Tracing for Seismic Surface Waves, *Geophysical Journal International*, **39**(1), 29–39.

Gély, J.-P. & Hanot, F., 2014. *Le Bassin parisien: Un nouveau regard sur la géologie*, AGBP.

Goertz, A., Schechinger, B., Witten, B., Koerbe, M., & Krajewski, P., 2012. Extracting subsurface information from ambient seismic noise — A case study from Germany, *Geophysics*, **77**(4), KS13–KS31.

Goldstein, P. & Archuleta, R. J., 1987. Array analysis of seismic signals, *Geophysical Research Letters*, **14**(1), 13–16.

Gorbatikov, A. V. & Tsukanov, A. A., 2011. Simulation of the Rayleigh waves in the proximity of the scattering velocity heterogeneities. Exploring the capabilities of the microseismic sounding method, *Izvestiya, Physics of the Solid Earth*, **47**(4), 354–369.

Gorbatikov, A. V., Kalinina, A. V., Volkov, V. A., Arnoso, J., Vieira, R., & Velez, E., 2004. Results of Analysis of the Data of Microseismic Survey at Lanzarote Island, Canary, Spain, *pure and applied geophysics*, **161**(7), 1561–1578.

Gorbatikov, A. V., Stepanova, M. Y., & Korablev, G. E., 2008. Microseismic field affected by local geological heterogeneities and microseismic sounding of the medium, *Izvestiya, Physics of the Solid Earth*, **44**(7), 577–592.

Gorbatikov, A. V., Montesinos, F. G., Arnoso, J., Stepanova, M. Y., Benavent, M., & Tsukanov, A. A., 2013. New Features in the Subsurface Structure Model of El Hierro Island (Canaries) from Low-Frequency Microseismic Sounding: An Insight into the 2011 Seismo-Volcanic Crisis, *Surveys in Geophysics*, **34**(4), 463–489.

Gouédard, P., Cornou, C., & Roux, P., 2008. Phase-velocity dispersion curves and small-scale geophysics using noise correlation slantstack technique, *Geophysical Journal International*, **172**(3), 971–981.

Grab, M., Quintal, B., Caspari, E., Deuber, C., Maurer, H., & Greenhalgh, S., 2017a. The Effect of Boiling on Seismic Properties of Water-Saturated Fractured Rock, *Journal of Geophysical Research: Solid Earth*, **122**(11), 9228–9252.

Grab, M., Quintal, B., Caspari, E., Maurer, H., & Greenhalgh, S., 2017b. Numerical modeling of fluid effects on seismic properties of fractured magmatic geothermal reservoirs, *Solid Earth*, **8**(1), 255–279.

Graf, R., Schmalholz, S., Podlachikov, Y., & Saenger, E. H., 2007. Passive low frequency spectral analysis: Exploring a new field in geophysics, *World Oil*, pp. 47–52.

Grandjean, G. & Leparoux, D., 2004. The potential of seismic methods for detecting cavities and buried objects: experimentation at a test site, *Journal of Applied Geophysics*, **56**(2), 93–106.

Green, A. G. & Greenhalgh, S., 2010. Comment on 'Low-frequency microtremor anomalies at an oil and gas field in Voitsdorf, Austria' by Marc-André Lambert, Stefan Schmalholz, Erik H. Saenger and Brian Steiner, *Geophysical Prospecting* 57, 393-411, *Geophysical Prospecting*, **58**(2), 335–339.

Gualtieri, L., Stutzmann, E., Capdeville, Y., Farra, V., Mangeney, A., & Morelli, A., 2015. On the shaping factors of the secondary microseismic wavefield, *Journal of Geophysical Research: Solid Earth*, **120**(9), 6241–6262.

Gucunski, N., Ganji, V., & Maher, M. H., 1996. Effects of obstacles on Rayleigh wave dispersion obtained from the SASW test, *Soil Dynamics and Earthquake Engineering*, **15**(4), 223–231.

Gutenberg, B., 1958. Microseisms, in *Advances in Geophysics*, vol. 5, pp. 53–92, Elsevier.

Hansen, P. C. & O'Leary, D. P., 1993. The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, *SIAM Journal on Scientific Computing*, **14**(6), 1487–1503.

Hanssen, P. & Bussat, S., 2008. Pitfalls in the analysis of low frequency passive seismic data, *First Break*, **26**(6).

Haskell, N. A., 1953. The dispersion of surface waves on multilayered media, *Bulletin of the Seismological Society of America*, **43**(1), 17–34.

Hasselmann, K., 1963. A statistical analysis of the generation of microseisms, *Reviews of Geophysics*, **1**(2), 177–210.

He, W. & Plessix, R.-d., 2017. Analysis of different parameterisations of waveform inversion of compressional body waves in an elastic transverse isotropic Earth with a vertical axis of symmetry: Parameterisations of waveform inversion, *Geophysical Prospecting*, **65**(4), 1004–1024.

Herak, M., 2008. ModelHVSR—A Matlab® tool to model horizontal-to-vertical spectral ratio of ambient noise, *Computers & Geosciences*, **34**(11), 1514–1526.

Herrmann, R. B., 2013. Computer Programs in Seismology: An Evolving Tool for Instruction and Research, *Seismological Research Letters*, **84**(6), 1081–1088.

Hestenes, M. R. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *Journal of research of the National Bureau of Standards*, **49**, 409–436.

Hillers, G., Graham, N., Campillo, M., Kedar, S., Landès, M., & Shapiro, N., 2012. Global oceanic microseism sources as seen by seismic arrays and predicted by wave action models, *Geochemistry, Geophysics, Geosystems*, **13**(1), n/a–n/a.

Holzner, R., Eschle, P., Zürcher, H., Lambert, M., Graf, R., Dangel, S., & Meier, P. F., 2005. Applying microtremor analysis to identify hydrocarbon reservoirs, *First Break*, **23**(5).

Holzner, R., Eschle, P., Dangel, S., Frehner, M., Narayanan, C., & Lakehal, D., 2009. Hydrocarbon microtremors interpreted as nonlinear oscillations driven by oceanic background waves, *Communications in Nonlinear Science and Numerical Simulation*, **14**(1), 160–173.

Hudson, J. A. & Heritage, J. R., 1981. The use of the Born approximation in seismic scattering problems, *Geophysical Journal International*, **66**(1), 221–240.

Huguet, F., Kazantsev, A., Lavergne, D., & Egermann, P., 2017. Procédé de caractérisation du sous-sol d'une region utilisant des signaux sismiques passifs, et système correspondant, Patent WO2017220918; Assignee: Storengy.

Hévin, G., 1998. *Utilisation des ondes de surface pour l'auscultation des structures en génie civil : application à la caractérisation des fissures de surface*, Ph.d. thesis, Université Joseph-Fourier - Grenoble I.

Jagnoux, P. & Vincent, A., 1988. Etude des ondes de surface permettant la réalisation de cartographies de défauts débouchants, *Revue de Physique Appliquée*, **23**(8), 1383–1396.

Jaxybulatov, K., 2017. *Combining different types of geophysical and geological data to study volcano-magmatic feeding systems*, Ph.D. thesis, IPGP Cuvier.

Jin, S., Madariaga, R., Virieux, J., & Lambaré, G., 1992. Two-dimensional asymptotic iterative elastic inversion, *Geophysical Journal International*, **108**(2), 575–588.

Johnston, D. H., Toksöz, M. N., & Timur, A., 1979. Attenuation of seismic waves in dry and saturated rocks: II. Mechanisms, *Geophysics*, **44**(4), 691–711.

Julian, B. R., 1994. Volcanic tremor: Nonlinear excitation by fluid flow, *Journal of Geophysical Research: Solid Earth*, **99**(B6), 11859–11877.

Jurkevics, A., 1988. Polarization analysis of three-component array data, *Bull. Seismol. Soc. Am*, pp. 1725–1743.

Kausel, E., 2006. *Fundamental solutions in elastodynamics: a compendium*, Cambridge University Press, New York, OCLC: ocm61130818.

Kazantsev, A., 2015. Extraction et analyse de l'anomalie sismique basse fréquence sur le site de stockage de Chémery, Storengy internal report OI-CE/DESS-AKA-2015-00237.

Kazantsev, A. & Chauris, H., 2018. Radiation patterns for 2d rayleigh-wave scattering in the vertical plane: impact of the free-surface, *Geophysical Prospecting*, **submitted 29/07/2018**.

Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2017a. Near-field Elastic Scattering of Rayleigh Waves - A Model for Interpreting Hydrocarbon Microtremors, in *79th EAGE Conference and Exhibition 2017*.

Kazantsev, A., Egermann, P., Ramadhan, I., Huguet, F., Formento, C., Peruzzetto, M., Chauris, H., & Métaxian, J.-P., 2017b. An innovative methodology based on Low Frequency Passive Seismic data analysis to map geothermal reservoir steam saturated areas, in *5th International Indonesia Geothermal Conference and Exhibition (IIGCE)*, Jakarta, Indonesia.

Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2018a. Spectral amplitude imaging with Rayleigh waves, *GJI*, **in preparation**.

Kazantsev, A., Peruzzetto, M., Chauris, H., Dublanchet, P., & Huguet, F., 2018b. Origins Of Rayleigh Wave Overtones In Ambient Noise, in *Seventh EAGE Workshop on Passive Seismic 2018*.

Kazantsev, A., Chauris, H., Dublanchet, P., & Huguet, F., 2019. Rayleigh wave amplitude distortions above a reservoir: new insights from elastic modelling, *Geophysical Journal International*, **accepted 04/02/2019**.

Kelly, M. C., Riahi, N., Ruiz, M., & Weiwei, 2013. Bayesian DHI for seismic data, US Patent US8358561B2; Assignee: Spectraseis AG.

Kennett, B. L. N., 1983. *Seismic Wave Propagation in Stratified Media*, Cambridge University Press.

Kennett, B. L. N., 1986. Lg waves and structural boundaries, *Bulletin of the Seismological Society of America*, **76**(4), 1133–1141.

Klimentos, T., 1995. Attenuation of P- and S-waves as a method of distinguishing gas and condensate from oil and water, *Geophysics*, **60**(2), 447–458.

Komatitsch, D. & Vilotte, J.-P., 1998. The spectral element method: An efficient tool to simulate the seismic response of 2d and 3d geological structures, *Bulletin of the Seismological Society of America*, **88**(2), 368–392.

Komatitsch, D., Vilotte, J.-P., Vai, R., Castillo-Covarrubias, J. M., & Sánchez-Sesma, F. J., 1999. The spectral element method for elastic wave equations—application to 2-D and 3-D seismic problems, *International Journal for Numerical Methods in Engineering*, **45**(9), 1139–1164.

Koper, K. D., Seats, K., & Benz, H., 2010. On the composition of Earth's short-period seismic noise field, *Bulletin of the Seismological Society of America*, **100**(2), 606–617.

Korneev, V., 2008. Slow waves in fractures filled with viscous fluid, *Geophysics*, **73**(1), N1–N7.

Korneev, V., 2009. Resonant seismic emission of subsurface objects, *Geophysics*, **74**(2), T47–T53.

Korneev, V. A., Goloshubin, G. M., Daley, T. M., & Silin, D. B., 2004. Seismic low-frequency effects in monitoring fluid-saturated reservoirs, *Geophysics*, **69**(2), 522–532.

Krim, H. & Viberg, M., 1996. Two decades of array signal processing research: the parametric approach, *IEEE Signal Processing Magazine*, **13**(4), 67–94.

Kurlenya, M. V. & Serdyukov, S. V., 1999. Low-frequency resonances of seismic luminescence of rocks in a low-energy vibration-seismic field, *Journal of Mining Science*, **35**(1), 1–5.

Kuteynikova, M., Tisato, N., Jänicke, R., & Quintal, B., 2014. Numerical modeling and laboratory measurements of seismic attenuation in partially saturated rock, *Geophysics*, **79**(2), L13–L20.

Lambert, M., Steiner, B., Schmalholz, S. M., Holzner, R., & Saenger, E. H., 2006. Soft Soil Amplification of Ambient Seismic Noise – Field Measurements and Numerical Modeling of H/V Ratios, in *First EAGE Passive Seismic Workshop - Exploration and Monitoring Applications*.

Lambert, M., Schmalholz, S., Saenger, E., & Podladchikov, Y., 2007. Low-frequency anomalies in spectral ratios of single-station microtremor measurements: Observations across an oil and gas field in Austria, in *SEG Technical Program Expanded Abstracts 2007*, pp. 1352–1356.

Lambert, M.-A., 2010. *Characterization of ambient ground-motion using spectral analysis techniques*, Ph.D. thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 19047, 2010.

Lambert, M.-A., Schmalholz, S. M., Saenger, E. H., & Steiner, B., 2009. Low-frequency microtremor anomalies at an oil and gas field in Voitsdorf, Austria, *Geophysical Prospecting*, **57**(3), 393–411.

Lambert, M.-A., Schmalholz, S. M., Saenger, E. H., & Steiner, B., 2010. Reply to comment on 'Low-frequency microtremor anomalies at an oil and gas field in Voitsdorf, Austria' by Marc-André Lambert, Stefan M. Schmalholz, Erik H. Saenger and Brian Steiner, *Geophysical Prospecting* 57, 393-411, *Geophysical Prospecting*, **58**(2), 341–346.

Lambert, M.-A., Nguyen, T., Saenger, E. H., & Schmalholz, S. M., 2011. Spectral analysis of ambient ground-motion—Noise reduction techniques and a methodology for mapping horizontal inhomogeneity, *Journal of Applied Geophysics*, **74**(2-3), 100–113.

Lambert, M.-A., Saenger, E., Quintal, B., & Schmalholz, S., 2013. Numerical simulation of ambient seismic wavefield modification caused by pore-fluid effects in an oil reservoir, *Geophysics*, **78**(1), T41–T52.

Landès, M., Hubans, F., Shapiro, N. M., Paul, A., & Campillo, M., 2010. Origin of deep ocean microseisms by using teleseismic body waves, *Journal of Geophysical Research*, **115**(B5).

Larose, E., Carrière, S., Voisin, C., Bottelin, P., Baillet, L., Guéguen, P., Walter, F., Jongmans, D., Guillier, B., Garambois, S., Gimbert, F., & Massey, C., 2015. Environmental seismology: What can we learn on earth surface processes with ambient noise?, *Journal of Applied Geophysics*, **116**, 62–74.

Lehujeur, M., Vergne, J., Schmittbuhl, J., & Maggi, A., 2015. Characterization of ambient seismic noise near a deep geothermal reservoir and implications for interferometric methods: a case study in northern Alsace, France, *Geothermal Energy*, **3**(1).

Lehujeur, M., Vergne, J., Maggi, A., & Schmittbuhl, J., 2017a. Ambient noise tomography with non-uniform noise sources and low aperture networks: case study of deep geothermal reservoirs in northern Alsace, France, *Geophysical Journal International*, **208**(1), 193–210.

Lehujeur, M., Vergne, J., Maggi, A., & Schmittbuhl, J., 2017b. Vertical seismic profiling using double beamforming processing of non-uniform anthropogenic seismic noise: the case study of Rittershoffen, Upper Rhine Graben, France, *Geophysics*, pp. 1–38.

Levshin, A. L. & Pisarenko, V., 1972. On a frequency-time analysis of oscillations, *Annales Geophysicae*, **28**(2), 211–218.

Lin, F.-C. & Ritzwoller, M. H., 2011. Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure: Helmholtz surface wave tomography, *Geophysical Journal International*, **186**(3), 1104–1120.

Lin, F.-C., Moschetti, M. P., & Ritzwoller, M. H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophysical Journal International*, **173**(1), 281–298.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R., 2009. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array, *Geophysical Journal International*, **177**(3), 1091–1110.

Lin, F.-C., Tsai, V. C., & Ritzwoller, M. H., 2012. The local amplification of surface waves: A new observable to constrain elastic velocities, density, and anelastic attenuation, *Journal of Geophysical Research: Solid Earth*, **117**(B6), n/a–n/a.

Liu, K. & Zhou, Y., 2016. Travelling-wave Green tensor and near-field Rayleigh-wave sensitivity, *Geophysical Journal International*, **205**(1), 134–145.

Lobkis, O. I., Larose, E., & Weaver, R. L., 2006. Passive correlation imaging of a buried scatterer, *The Journal of the Acoustical Society of America*, **119**(5), 3307–3307.

Longuet-Higgins, M. S., 1950. A theory of the origin of microseisms, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **243**(857), 1–35.

Lucet, N., Rasolofosaon, P. N. J., & Zinszner, B., 1991. Sonic properties of rocks under confining pressure using the resonant bar technique, *The Journal of the Acoustical Society of America*, **89**(3), 980–990.

Luu, K., Noble, M., Gesret, A., Belayouni, N., & Roux, P.-F., 2018. A parallel competitive Particle Swarm Optimization for non-linear first arrival traveltime tomography and uncertainty quantification, *Computers & Geosciences*, **113**, 81–93.

Ma, Y., Clayton, R. W., & Li, D., 2016. Higher-mode ambient-noise Rayleigh waves in sedimentary basins, *Geophysical Journal International*, **206**(3), 1634–1644.

Mardanov, M. & Kipot, V., 2008. Nature of spectral anomalies of seismic waves above hydrocarbon-bearing layers (in russian), *Georesursy (in Russian)*, **24**(1).

Margerin, L. & Nolet, G., 2003. Multiple scattering of high-frequency seismic waves in the deep Earth: Modeling and numerical examples, *Journal of Geophysical Research: Solid Earth*, **108**(B5).

Martini, F., Lokmer, I., Jonsdottir, K., De Barros, L., Müllhoff, M., Bean, C. J., Hauser, F., Doherty, J., Ryan, C., & Mongan, J., 2013. A passive low-frequency seismic experiment in the Albertine Graben, Uganda: *A passive low-frequency seismic experiment*, *Geophysical Prospecting*, **61**, 39–61.

Mastrigt, P. v. & Al-Dulaijan, A., 2008. Seismic Spectroscopy Using Amplified 3c Geophones, in *70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008*.

Maupin, V., 1996. The radiation modes of a vertically varying half-space: a new representation of the complete Green's function in terms of modes, *Geophysical Journal International*, **126**(3), 762–780.

Maupin, V., 2001. A multiple-scattering scheme for modelling surface wave propagation in isotropic and anisotropic three-dimensional structures, *Geophysical Journal International*, **146**(2), 332–348.

Maupin, V., 2017. 3-D sensitivity kernels of the Rayleigh wave ellipticity, *Geophysical Journal International*, **211**(1), 107–119.

Mavko, G. & Mukerji, T., 1998. Bounds on low-frequency seismic velocities in partially saturated rocks, *Geophysics*, **63**(3), 918–924.

ed. Mégnien, C., 1980. *Stratigraphie et paléogéographie*, no. publiée sous la direction de Claude Mégnien ; Vol. 1 in Synthèse géologique du Bassin de Paris, Éd. du BRGM, Orléans, OCLC: 256350552.

Mindlin, R., 1948. Compliance of elastic bodies in contact, *Journal of Applied Mechanics*, **16**, 259–268.

Müller, T. M., Gurevich, B., & Lebedev, M., 2010. Seismic wave attenuation and dispersion resulting from wave-induced flow in porous rocks — A review, *Geophysics*, **75**(5), 75A147–75A164.

Moore, E.-H., 1920. On the reciprocal of the general algebraic matrix, *Bulletin of the American Mathematical Society*, **26**, 394–395.

Mordret, A., Landes, M., Shapiro, N. M., Singh, S. C., Roux, P., & Barkved, O. I., 2013a. Near-surface study at the Valhall oil field from ambient noise surface wave tomography, *Geophysical Journal International*, **193**(3), 1627–1643.

Mordret, A., Shapiro, N. M., Singh, S. S., Roux, P., & Barkved, O. I., 2013b. Helmholtz tomography of ambient noise surface wave data to estimate Scholte wave phase velocity at Valhall Life of the Field, *Geophysics*, **78**(2), WA99–WA109.

Mordret, A., Landes, M., Shapiro, N. M., Singh, S. C., & Roux, P., 2014. Ambient noise surface wave tomography to determine the shallow shear velocity structure at Valhall: depth inversion with a Neighbourhood Algorithm, *Geophysical Journal International*, **198**(3), 1514–1525.

Nabney, I. T., 2002. *NETLAB: Algorithms for Pattern Recognition*, Advances in Computer Vision and Pattern Recognition, Springer-Verlag, London.

Nakamura, Y., 1989. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surfac, *Quaterly Report of RTRI*, **30**.

Nakanishi, I. & Anderson, D. L., 1982. Worldwide distribution of group velocity of mantle Rayleigh waves as determined by spherical harmonic inversion, *Bulletin of the Seismological Society of America*, **72**(4), 1185–1194.

Nakata, N., Chang, J. P., Lawrence, J. F., & Boué, P., 2015. Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *Journal of Geophysical Research: Solid Earth*, **120**(2), 1159–1173.

Nasseri-Moghaddam, A., Cascante, G., & Hutchinson, J., 2005. A new quantitative procedure to determine the location and embedment depth of a void using surface waves, *Journal of Environmental & Engineering Geophysics*, **10**(1), 51–64.

Nasseri-Moghaddam, A., Cascante, G., Phillips, C., & Hutchinson, D., 2007. Effects of underground cavities on Rayleigh waves—Field and numerical experiments, *Soil Dynamics and Earthquake Engineering*, **27**(4), 300–313.

Nguyen, T. T., Saenger, E. H., Schmalholz, S. M., & Artman, B., 2008. Earthquake Triggered Modifications of Microtremor Signals above and nearby a Hydrocarbon Reservoir in Voitsdorf, Austria, in *70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008*.

Nikolaevskiy, V. N., Lopukhov, G. P., Yizhu, L., & Economides, M. J., 1996. Residual Oil Reservoir Recovery With Seismic Vibrations, *SPE Production & Facilities*, **11**(02), 89–94.

Nishida, K., 2017. Ambient seismic wave field, *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, **93**(7), 423–448.

Obermann, A., Planes, T., Larose, E., Sens-Schonfelder, C., & Campillo, M., 2013. Depth sensitivity of seismic coda waves to velocity perturbations in an elastic heterogeneous medium, *Geophysical Journal International*, **194**(1), 372–382.

Obermann, A., Planès, T., Hadziioannou, C., & Campillo, M., 2016. Lapse-time-dependent coda-wave depth sensitivity to local velocity perturbations in 3-D heterogeneous elastic media, *Geophysical Journal International*, **207**(1), 59–66.

Obrebski, M., Ardhuin, F., Stutzmann, E., & Schimmel, M., 2013. Detection of microseismic compressional (P) body waves aided by numerical modeling of oceanic noise sources, *Journal of Geophysical Research: Solid Earth*, **118**(8), 4312–4324.

O'Connell, R. J. & Budiansky, B., 1978. Measures of dissipation in viscoelastic media, *Geophysical Research Letters*, **5**(1), 5–8.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L., & Virieux, J., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice, *The Leading Edge*, **32**, 1040–1054.

Pan, W., 2017. *Waveform inversion for estimating subsurface properties: phase-encoding strategies, optimization methods, interparameter tradeoffs quantification and reduction*, Ph.D. Thesis, University of Calgary.

Park, C. B., Miller, R. D., & Xia, J., 1999. Multichannel analysis of surface waves, *Geophysics*, **64**(3), 800–808.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **2**(11), 559–572.

Perrodon, A. & Zabek, J., 1990. Paris Basin, in *Interior Cratonic Basin*, vol. 51 of **American Association of Petroleum Geologists Memoirs**, pp. 653–679.

Perton, M., Spica, Z., & Caudron, C., 2018. Inversion of the horizontal-to-vertical spectral ratio in presence of strong lateral heterogeneity, *Geophysical Journal International*, **212**(2), 930–941.

Peruzzetto, M., Kazantsev, A., Luu, K., Métaxian, J.-P., Huguet, F., & Chauris, H., 2018. Broad-band ambient noise characterization by joint use of cross-correlation and MUSIC algorithm, *Geophysical Journal International*, **215**(2), 760–779.

Peterson, J. R., 1993. Observations and modeling of seismic background noise, USGS Numbered Series 93-322, U.S. Geological Survey.

Peureux, C. & Ardhuin, F., 2016. Ocean bottom pressure records from the Cascadia array and short surface gravity waves., *Journal of Geophysical Research: Oceans*, **121**(5), 2862–2873.

Plessix, R.-. & Cao, Q., 2011. A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium: FWI in an acoustic VTI medium, *Geophysical Journal International*, **185**(1), 539–556.

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

Plessix, R.-E. & Mulder, W. A., 2004. Frequency-domain finite-difference amplitude-preserving migration, *Geophysical Journal International*, **157**(3), 975–987.

Podladchikov, Y., Lambert, M.-A., Dewarrat, R., & Schmalholz, S., 2010. VH Reservoir Mapping, Patent US7676326B2; Assignee: Spectraseis AG.

Poli, P., Pedersen, H. A., & Campillo, M., 2012. Emergence of body waves from cross-correlation of short period seismic noise, *Geophysical Journal International*, **188**(2), 549–558.

Pratt, M. J., Wiens, D. A., Winberry, J. P., Anandakrishnan, S., & Euler, G. G., 2017. Implications of sea ice on Southern Ocean microseisms detected by a seismic array in West Antarctica, *Geophysical Journal International*, **209**(1), 492–507.

Prieto, G. A., Lawrence, J. F., & Beroza, G. C., 2009. Anelastic Earth structure from the coherency of the ambient seismic field, *Journal of Geophysical Research*, **114**(B7).

Quintal, B., 2012. Frequency-dependent attenuation as a potential indicator of oil saturation, *Journal of Applied Geophysics*, **82**, 119–128.

Quintal, B., Schmalholz, S. M., & Podladchikov, Y. Y., 2009. Low-frequency reflections from a thin layer with high attenuation caused by interlayer flow, *Geophysics*, **74**(1), N15–N23.

Quintal, B., Frehner, M., Madonna, C., Tisato, N., Kuteynikova, M., & Saenger, E., 2011a. Integrated numerical and laboratory rock physics applied to seismic characterization of reservoir rocks, *The Leading Edge*, **30**(12), 1360–1367.

Quintal, B., Steeb, H., Frehner, M., & Schmalholz, S. M., 2011b. Quasi-static finite element modeling of seismic attenuation and dispersion due to wave-induced fluid flow in poroelastic media, *Journal of Geophysical Research*, **116**(B1).

Quintal, B., Steeb, H., Frehner, M., Schmalholz, S. M., & Saenger, E. H., 2012. Pore fluid effects on S-wave attenuation caused by wave-induced fluid flow, *Geophysics*, **77**(3), L13–L23.

Quintal, B., Jänicke, R., Rubino, J. G., Steeb, H., & Holliger, K., 2014. Sensitivity of S-wave attenuation to the connectivity of fractures in fluid-saturated rocks, *Geophysics*, **79**(5), WB15–WB24.

Rached, G. R., 2006. Surface Passive Seismic in Kuwait, in *First EAGE Passive Seismic Workshop - Exploration and Monitoring Applications*.

Rayleigh, L., 1885. On Waves Propagated along the Plane Surface of an Elastic Solid, *Proceedings of the London Mathematical Society*, **s1-17**(1), 4–11.

Riahi, N., Bokelmann, G., Sala, P., & Saenger, E. H., 2013a. Time-lapse analysis of ambient surface wave anisotropy: A three-component array study above an underground gas storage, *Journal of Geophysical Research: Solid Earth*, **118**(10), 5339–5351.

Riahi, N., Goertz, A., Birkelo, B., & Saenger, E. H., 2013b. A statistical strategy for ambient seismic wavefield analysis: investigating correlations to a hydrocarbon reservoir, *Geophysical Journal International*, **192**(1), 148–162.

Ripepe, M. & Gordeev, E., 1999. Gas bubble dynamics model for shallow volcanic tremor at Stromboli, *Journal of Geophysical Research: Solid Earth*, **104**(B5), 10639–10654.

Ritzwoller, M. H., Shapiro, N. M., Barmin, M. P., & Levshin, A. L., 2002. Global surface wave diffraction tomography, *Journal of Geophysical Research: Solid Earth*, **107**(B12), 2335.

Riyanti, C. D. & Herman, G. C., 2005. Three-dimensional elastic scattering by near-surface heterogeneities, *Geophysical Journal International*, **160**(2), 609–620.

Romanowicz, B., 2002. 11 Inversion of surface waves: A review, in *International Geophysics*, vol. 81, pp. 149–173, Elsevier.

Rost, S. & Thomas, C., 2002. Array seismology: Methods and applications, *Reviews of Geophysics*, **40**(3).

Roux, P., Sabra, K. G., Gerstoft, P., Kuperman, W. A., & Fehler, M. C., 2005. P-waves from cross-correlation of seismic noise, *Geophysical Research Letters*, **32**(19), L19303.

Ruigrok, E., Campman, X., & Wapenaar, K., 2011. Extraction of P-wave reflections from microseisms, *Comptes Rendus Geoscience*, **343**(8), 512–525.

Ruigrok, E., Gibbons, S., & Wapenaar, K., 2017. Cross-correlation beamforming, *Journal of Seismology*, **21**(3), 495–508.

Saenger, E., Torres, A., Rentsch, S., Lambert, M., Schmalholz, S., & Mendez-Hernandez, E., 2007. A hydrocarbon microtremor survey over a gas field: Identification of seismic attributes, in *SEG Technical Program Expanded Abstracts 2007*, pp. 1277–1281, Society of Exploration Geophysicists.

Saenger, E., Steiner, B., & Schmalholz, S., 2010. Time reverse reservoir localization, Patent US7675815B2; Assignee: Spectraseis AG.

Saenger, E. H., 2008. Seismic attributes for reservoir localization, Patent US20080288173A1; Assignee: Spectraseis AG.

Saenger, E. H., 2009. Signal integration measure for seismic data, Patent US7590491B2, Assignee: Spectraseis AG.

Saenger, E. H., Gold, N., & Shapiro, S. A., 2000. Modeling the propagation of elastic waves using a modified finite-difference grid, *Wave motion*, **31**(1), 77–92.

Saenger, E. H., Schmalholz, S. M., Lambert, M.-A., Nguyen, T. T., Torres, A., Metzger, S., Habiger, R. M., Müller, T., Rentsch, S., & Méndez-Hernández, E., 2009. A passive seismic survey over a gas field: Analysis of low-frequency anomalies, *Geophysics*, **74**(2), O29–O40.

Saenger, E. H., Kocur, G. K., Jud, R., & Torrilhon, M., 2011. Application of time reverse modeling on ultrasonic non-destructive testing of concrete, *Applied Mathematical Modelling*, **35**(2), 807–816.

Sager, K., Ermert, L., Boehm, C., & Fichtner, A., 2018. Towards full waveform ambient noise inversion, *Geophysical Journal International*, **212**(1), 566–590.

Schmalholz, S. M., Podladchikov, Y. Y., Holzner, R., & Saenger, E. H., 2006. Scientific Strategy to Explain Observed Spectral Anomalies over Hydrocarbon Reservoirs Generated by Microtremors, in *First EAGE Passive Seismic Workshop - Exploration and Monitoring Applications*.

Schmidt, R. O., 1986. Multiple Emitter Location and Signal Parameter Estimation, *IEEE Transactions on Antennas and Propagation*, **AP-34**(3).

Sens-Schönfelder, C. & Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia, *Geophysical Research Letters*, **33**(21), L21302.

Shan, T.-J., Wax, M., & Kailath, T., 1985. On spatial smoothing for direction-of-arrival estimation of coherent signals, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **33**(4), 806–811.

Shapiro, N., Béthoux, N., Campillo, M., & Paul, A., 1996. Regional seismic phases across the Ligurian Sea: Lg blockage and oceanic propagation, *Physics of the earth and planetary interiors*, **93**(3-4), 257–268.

Shapiro, N. M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise: correlations of the seismic noise, *Geophysical Research Letters*, **31**(7), n/a–n/a.

Situmorang, J., Martikno, R., Perdana Putra, A., & Ganefianto, N., 2016. A Reservoir Simulation of the Muara Laboh Field, Indonesia, Stanford University, Stanford, California.

Sánchez-Sesma, F. J., 2017. Modeling and inversion of the microtremor H/V spectral ratio: physical basis behind the diffuse field approach, *Earth, Planets and Space*, **69**(1).

Sánchez-Sesma, F. J., Pérez-Ruiz, J. A., Campillo, M., & Luzón, F., 2006. Elastodynamic 2d Green function retrieval from cross-correlation: Canonical inclusion problem, *Geophysical Research Letters*, **33**(13).

Sánchez-Sesma, F. J., Pérez-Ruiz, J. A., Luzón, F., Campillo, M., & Rodríguez-Castellanos, A., 2008. Diffuse fields in dynamic elasticity, *Wave Motion*, **45**(5), 641–654.

Sánchez-Sesma, F. J., Rodríguez, M., Iturrarán-Viveros, U., Luzón, F., Campillo, M., Margerin, L., García-Jerez, A., Suarez, M., Santoyo, M. A., & Rodríguez-Castellanos, A., 2011a. A theory for microtremor H/V spectral ratio: application for a layered medium, *Geophysical Journal International*, **186**(1), 221–225.

Sánchez-Sesma, F. J., Rodríguez-Castellanos, A., Perton, M., Luzón, F., & Ortiz-Alemán, C., 2011b. Diffuse seismic waves and site effects, *Journal of Geophysics and Engineering*, **8**(1), 109–114.

Sánchez-Sesma, F. J., Weaver, R. L., Kawase, H., Matsushima, S., Luzon, F., & Campillo, M., 2011c. Energy Partitions among Elastic Waves for Dynamic Surface Loads in a Semi-Infinite Solid, *Bulletin of the Seismological Society of America*, **101**(4), 1704–1709.

Snieder, R., 1986. 3-D linearized scattering of surface waves and a formalism for surface wave holography, *Geophysical Journal International*, **84**(3), 581–605.

Snieder, R., 2006. The Theory of Coda Wave Interferometry, *Pure and Applied Geophysics*, **163**(2-3), 455–473.

Solano, P., A, C., Donno, D., & Chauris, H., 2014. Alternative waveform inversion for surface wave analysis in 2-D media, *Geophysical Journal International*, **198**(3), 1359–1372.

Sonier, F., Lehuen, P., & Nabil, R., 1993. Full-Field Gas Storage Simulation Using a Control-Volume Finite-Element Model, Society of Petroleum Engineers.

Steiner, B. & Saenger, E. H., 2012. Comparison of 2d and 3d time-reverse imaging—A numerical case study, *Computers & Geosciences*, **46**, 174–182.

Steiner, B., Saenger, E. H., & Schmalholz, S. M., 2007. Time Reverse Modeling of Microtremors: A Potential Method For Hydrocarbon Reservoir Localization, in *2007 SEG Annual Meeting*, Society of Exploration Geophysicists.

Steiner, B., Saenger, E. H., & Schmalholz, S. M., 2008a. Case Studies on 2d- and 3d-Time Reverse Modeling of Low-frequency Microtremors - Application to Reservoir Localization, in *70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008*.

Steiner, B., Saenger, E. H., & Schmalholz, S. M., 2008b. Time reverse modeling of low-frequency microtremors: Application to hydrocarbon reservoir localization, *Geophysical Research Letters*, **35**(3).

Storchak, D. A., Schweitzer, J., & Bormann, P., 2003. The IASPEI Standard Seismic Phase List, *Seismological Research Letters*, **74**(6), 761–772.

Strauss, W. A., 2008. *Partial Differential Equations: An Introduction*, John Wiley & Sons, New York, 2nd edn.

Strobbia, C., 2002. *Surface wave methods, Acquisition, processing and inversion*, Ph.D. thesis, Politecnico di Torino.

Stutzmann, E., Ardhuin, F., Schimmel, M., Mangeney, A., & Patau, G., 2012. Modelling long-term seismic noise in various environments, *Geophysical Journal International*, **191**(2), 707–722.

Subramaniyan, S., Quintal, B., & Saenger, E. H., 2017. Forced oscillation measurements of seismic attenuation in fluid saturated sandstone, *Acta Geophysica*, **65**(1), 165–172.

Tallavó, F., Cascante, G., & Pandey, M., 2009. Experimental and numerical analysis of MASW tests for detection of buried timber trestles, *Soil Dynamics and Earthquake Engineering*, **29**(1), 91–102.

Thomson, W. T., 1950. Transmission of Elastic Waves through a Stratified Solid Medium, *Journal of Applied Physics*, **21**(2), 89–93.

Tisato, N., Quintal, B., Chapman, S., Podladchikov, Y., & Burg, J.-P., 2015. Bubbles attenuate elastic waves at seismic frequencies: First experimental evidence, *Geophysical Research Letters*, **42**(10), 3880–3887.

Tolman, H. L., 1991. A Third-Generation Model for Wind Waves on Slowly Varying, Unsteady, and Inhomogeneous Depths and Currents, *Journal of Physical Oceanography*, **21**(6), 782–797.

Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, **160**(1), 195–216.

Tsukanov, A. A. & Gorbatikov, A. V., 2015. Microseismic sounding method: Implications of anomalous Poisson ratio and evaluation of nonlinear distortions, *Izvestiya, Physics of the Solid Earth*, **51**(4), 548–558.

Tsukanov, A. A. & Gorbatnikov, A. V., 2018. Influence of Embedded Inhomogeneities on the Spectral Ratio of the Horizontal Components of a Random Field of Rayleigh Waves, *Acoustical Physics*, **64**(1), 70–76.

Urquizú, M. & Correig, A. M., 1999. On the spectral peaks of volcanic tremor at Stromboli, *Physics of the Earth and Planetary Interiors*, **110**(3), 247–261.

Ursin, B. & Toverud, T., 2002. Comparison of Seismic Dispersion and Attenuation Models, *Studia Geophysica et Geodaetica*, **46**(2), 293–320.

Vidal, S., 2002. *Méthodologie d'interprétation des paramètres réservoir à partir des mesures du monitoring sismique, intégrant l'aspect géomécanique*, Ph.D. thesis, Paris 6.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method, *Geophysics*, **51**(4), 889–901.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.

Virieux, J., Calandra, H., & Plessix, R.-d., 2011. A review of the spectral, pseudo-spectral, finite-difference and finite-element modelling techniques for geophysical imaging, *Geophysical Prospecting*, **59**(5), 794–813.

Walker, D., 2008. Recent developments in low frequency spectral analysis of passive seismic data, *First Break*, **26**(2).

Walls, J., Uden, R., Singleton, S., Shu, R., & Mavko, G., 2005. P- and S-wave seismic attenuation for deep natural gas exploration and development, Tech. Rep. DOE/NT/86227–4, 934923, Rock Solid Images.

Wathelet, M., 2008. An improved neighborhood algorithm: Parameter conditions and dynamic scaling, *Geophysical Research Letters*, **35**(9).

Weaver, R. L., 1982. On diffuse waves in solid media, *The Journal of the Acoustical Society of America*, **71**(6), 1608–1609.

White, J., 1975. Computed seismic speeds and attenuation in rocks with partial gas saturation, *Geophysics*, **40**(2), 224–232.

Wielandt, E., 1993. Propagation and structural interpretation of non-plane waves, *Geophysical Journal International*, **113**(1), 45–53.

Witten, B. & Artman, B., 2011. Signal-to-noise estimates of time-reverse images, *Geophysics*, **76**(2), MA1–MA10.

Woith, H., Parolai, S., Boxberger, T., Picozzi, M., Özmen, z. T., Milkereit, C., Lühr, B. G., & Zschau, J., 2014. Spatio-temporal variability of seismic noise above a geothermal reservoir, *Journal of Applied Geophysics*, **106**, 128–138.

Wu, R.-S. & Ben-Menahem, A., 1985. The elastodynamic near field, *Geophysical Journal International*, **81**(3), 609–621.

Yang, Y. & Forsyth, D. W., 2006. Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern California, *Journal of Geophysical Research*, **111**(B7).

Yanovskaya, T. B., 2017. On the theory of the microseismic sounding method, *Izvestiya, Physics of the Solid Earth*, **53**(6), 819–824.

Yu, C.-W. & Dravinski, M., 2009. Scattering of plane harmonic P, SV or Rayleigh waves by a completely embedded corrugated cavity, *Geophysical Journal International*, **178**(1), 479–487.

Zhang, J., Gerstoft, P., & Shearer, P. M., 2009. High-frequency P-wave seismic noise driven by ocean winds, *Geophysical Research Letters*, **36**(9).

Zhang, T.-R. & Lay, T., 1995. Why the Lg phase does not traverse oceanic crust, *Bulletin of the Seismological Society of America*, **85**(6), 1665–1678.

Zhukov, A. P., Loginov, K. I., Shneerson, M. B., Shulakova, V. E., Kharisov, R. G., & Ekimenko, V. A., 2007. Nonlinear properties of vibrator-generated wavefields and their application to hydrocarbon detection, *The Leading Edge*, **26**(11), 1395–1402.

Résumé

L'objet de cette thèse est l'étude des possibles mécanismes élastiques expliquant l'amplification du bruit sismique ambiant au droit de certains réservoirs multi-phasiques. Trois jeux de données sont traités. La signature spectrale observée d'un réservoir de vapeur géothermique est différente de celle d'un stockage de gaz naturel. Dans une approche empirique, un algorithme de classification permet d'extraire et de cartographier les anomalies que l'on présume liées au réservoir. Un travail de modélisation est effectué pour tenter d'expliquer les anomalies mesurées. Dans les données réelles, une forte présence de modes supérieurs d'ondes de Rayleigh est détectée. On modélise numériquement en 2D la propagation de ces modes à travers un réservoir placé au sein d'une structure géologique réaliste. La réponse simulée du réservoir se révèle trop faible par rapport aux observations de terrain. Néanmoins, on parvient à inverser les *faibles* perturbations d'amplitude synthétiques pour la position du réservoir, dans des modèles de référence simples. Cette méthode pourrait être utilisable pour l'imagerie à partir de faibles variations d'amplitude dans le cadre du monitoring. Pour ce qui est des *fortes* anomalies observées sur le terrain, il est à noter que les effets visco-élastiques, les effets 3D, et les effets liés à un éventuel champ incident diffus n'ont pas été pris en compte dans la modélisation. Ainsi *ce travail n'exclut pas la possibilité de telles anomalies liées à la présence d'un réservoir*.

Mots Clés

Bruit ambiant, Sismique passive, Imagerie sismique, Monitoring de réservoir, Stockage de gaz souterrain.

Abstract

This PhD work investigates the possible elastic mechanisms behind the ambient noise amplification above multi-phase fluid reservoirs. Three datasets are analysed above different reservoirs. The observed spectral signature is different in the natural gas storage and geothermal contexts. As an empirical approach, an automated classification algorithm allows to extract and map the spectral attributes presumably indicating a multi-phase fluid presence. In order to understand the nature of the observed anomalies, a modelling work is performed. In real data, the presence of strong Rayleigh overtones is detected. The propagation of these modes across a reservoir embedded in a realistic geological structure is modelled numerically in 2D. The modelled reservoir response is too weak compared to the real data. However, the *small* amplitude perturbations arising in the *synthetics* are successfully inverted for the position of the reservoir, in simple background models. The developed method could in theory be used for imaging small time-lapse amplitude variations (monitoring). In what concerns the *strong* anomalies observed in the field, it must be emphasized that neither visco-elastic, nor 3D, nor diffuse wavefield effects are taken into account in the modelling. Thus *this work does not exclude the possibility of such reservoir-specific anomalies*.

Keywords

Ambient noise, Passive seismic, Seismic imaging, Reservoir monitoring, Underground gas storage.