



HAL
open science

Analyse et traitement de la matrice de covariance de données enregistrées sur des réseaux de stations sismiques

Léonard Seydoux

► **To cite this version:**

Léonard Seydoux. Analyse et traitement de la matrice de covariance de données enregistrées sur des réseaux de stations sismiques. Analyse de données, Statistiques et Probabilités [physics.data-an]. Institut de Physique du Globe de Paris, 2016. Français. NNT: . tel-03505145v2

HAL Id: tel-03505145

<https://theses.hal.science/tel-03505145v2>

Submitted on 7 Jan 2022

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 préparée à l'Institut de Physique du Globe de Paris
École Doctorale n° 560 des Sciences de la Terre et de
l'Environnement et Physique de l'Univers de Paris (STEP'UP)

ANALYSE ET TRAITEMENT DE LA MATRICE DE
COVARIANCE DE DONNÉES ENREGISTRÉES
SUR DES RÉSEAUX DE STATIONS SISMIQUES

par
Léonard Seydoux



présentée et soutenue publiquement le
3 octobre 2016
devant un jury composé de

Stefan Catheline, Directeur de recherche (Université Claude Bernard, Lyon) Rapporteur
Lapo Boschi, Maître de Conférence (UPMC, Paris) Rapporteur
Philippe Roux, Directeur de recherche (ISTerre, Grenoble) Examineur
Jean-Pierre Vilotte, Physicien (IPGP, Paris) Examineur
Julien de Rosny, Directeur de recherche (ESPCI, Paris) Directeur de thèse
Nikolaï M. Shapiro, Directeur de recherche (IPGP, Paris) Directeur de thèse



Thèse préparée à l'Institut de Physique du Globe de Paris
École Doctorale n° 560 des Sciences de la Terre et de
l'Environnement et Physique de l'Univers de Paris (STEP'UP)

ANALYSE ET TRAITEMENT DE LA MATRICE DE
COVARIANCE DE DONNÉES ENREGISTRÉES
SUR DES RÉSEAUX DE STATIONS SISMIQUES

par
Léonard Seydoux

✧

présentée et soutenue publiquement le
3 octobre 2016
devant un jury composé de

Stefan Catheline, Directeur de recherche (Université Claude Bernard, Lyon) Rapporteur
Lapo Boschi, Maître de Conférence (UPMC, Paris) Rapporteur
Philippe Roux, Directeur de recherche (ISTerre, Grenoble) Examineur
Jean-Pierre Vilotte, Physicien (IPGP, Paris) Examineur
Julien de Rosny, Directeur de recherche (ESPCI, Paris) Directeur de thèse
Nikolaï M. Shapiro, Directeur de recherche (IPGP, Paris) Directeur de thèse

Résumé

L'estimation des fonctions de Green par corrélation de bruit sismique ambiant repose entre autre sur l'hypothèse forte que les sources de bruit sont réparties de façon homogène dans le milieu. Or, les données sismiques réelles montrent que ces sources de bruit ne respectent pas ces conditions dans la plupart des cas. En particulier, les signaux très cohérents générés par des séismes tectoniques ou des sources de bruit directionnelles existent et sont néfastes à l'utilisation du bruit sismique en imagerie passive. Des techniques de traitement du signal temporelles et fréquentielles sont alors habituellement appliquées aux sismogrammes individuels afin d'améliorer la qualité de reconstruction des fonctions de Green. C'est dans la continuité des études visant à traiter les sismogrammes que ce travail s'inscrit. En s'inspirant de méthodes de traitement d'antenne, nous analysons la matrice de covariance de données enregistrées sur des réseaux de stations sismiques. À travers l'analyse des valeurs propres de cette matrice, nous montrons qu'il est possible de dégager des informations cruciales sur le champ d'onde mesuré, comme le degré de cohérence spatiale dans les dimensions temporelles et fréquentielles. Cette information est importante car elle permet d'identifier les domaines temps-fréquence du champ d'onde propice à l'estimation des fonctions de Green. Finalement, une méthode d'égalisation des valeurs propres de la matrice de covariance visant à atténuer l'anisotropie des sources est proposée. Cette dernière méthode est interprétée à la fois comme une extension de la normalisation spectrale à la dimension spatiale, et également dans l'analogie du retournement temporel, au filtre inverse passif qui a vu le jour récemment.

Summary

Green's function estimation from ambient seismic noise relies on the strong hypothesis that noise seismic sources are evenly distributed in the medium. Yet, observations of seismic data show that the noise sources do not provide such good conditions in real cases. Strongly coherent seismic sources or directional noise seismic sources may exist, and are harmful to the application of this in ambient seismic imaging. Several signal processing techniques are nowadays routinely applied to individual seismograms in frequency and temporal domain in order to improve the quality of reconstruction of the Green's function. The present work takes place in this context. Our approach is inspired by array-processing techniques, and is particularly focused on the covariance matrix of data recorded on seismic arrays. We show that the eigenstructure of this matrix provides crucial information about the seismic wavefield degree of coherence, as a function of time and frequency. This information is important because it allows to identify the time and frequency zones where the Green's function quality is ensured. An original array-processing technique is finally proposed, which consider the equalization of the covariance matrix eigenvalues, in order to attenuate the wavefield anisotropy such as earthquake-related signals or directive noise sources. We interpret this last method as an extension of the spectral whitening technique widely used in seismology to the spatial dimension encoded by the covariance matrix. We also invoke the analogy with the time-reversal technique that have recently led to a class of passive inverse filtering techniques.

Remerciements

Je tiens avant tout à remercier chaleureusement Nikolai et Julien pour m'avoir donné l'opportunité de travailler sur ce sujet passionnant. Cette thèse m'a beaucoup apporté, tant sur le plan scientifique que sur ma vision de la recherche en général. Merci à vous deux pour nos échanges fructueux qui ont permis à ce travail d'aboutir et d'avoir de l'avenir devant lui. Merci également pour m'avoir envoyé à de nombreuses conférences et écoles d'été qui ont été pour moi d'excellentes occasions d'apprendre de la communauté scientifique sur le thème du bruit ambiant, et pour votre encadrement remarquable. Je garderai un excellent souvenir de ces trois années.

Merci Nikolai pour avoir partagé tes grandes idées scientifiques avec moi de façon naturelle. Merci aussi pour ton coaching exemplaire et pour tes nombreux conseils avisés, en particulier en ce qui concerne la communication. J'ai beaucoup appris à travailler avec toi, et j'espère continuer à pouvoir collaborer dans le futur comme nous avons pu le faire au cours de cette thèse.

Julien, merci pour avoir partagé ta grande culture scientifique avec moi, et pour m'avoir donné envie d'apprendre de la science des domaines voisins sans appréhension. Merci pour tes idées lumineuses qui ont rendu ce travail passionnant, et qui m'ont permis de redécouvrir des concepts fondamentaux pour mieux me les approprier. J'espère que nous continuerons de travailler ensemble dans cette même ambiance.

Un grand merci à Jean-Pierre, Stefan, Lapo et Philippe pour avoir pris le temps d'évaluer ce travail et pour tous les commentaires et conseils qu'ils m'ont donnés.

Je remercie également toute l'équipe de sismologie qui accueille chaleureusement ses étudiants. C'est une grande équipe, tant de nombre que d'esprit. La sismologie était une science étrangère à mon arrivée, à l'heure actuelle j'ai plus qu'envie de continuer dans ce domaine passionnant, riche en choses à faire. Merci aux membres permanents de l'équipe pour leur ouverture à la discussion, en particulier, merci à Claudio, Pascal, Martin, Nobu, Éléonore, François, Jean-Pierre, Anne, Jean-Paul et Alexandre.

Je tiens également à remercier Matthieu qui m'a énormément inspiré, tant par sa vision scientifique que par sa puissante capacité à parler à l'oreille des clusters. Merci pour m'avoir propulsé au cœur de ma thèse en développant une grande partie des codes qui ont permis ce travail d'aboutir dans les temps, et pour m'avoir appris à parler dans des langues appropriées. Merci aussi pour nos échanges au-delà de la science!

Merci également à Michel pour m'avoir beaucoup aidé avec mes multiples crash d'ordinateurs (4 pour être précis!). Merci pour les coups de pouce qui m'ont facilité grandement la tâche.

Dans la même veine, merci Geneviève! Tu as su m'apprendre tellement de choses en informatique, en calcul intensif (au passage merci aussi à Alexandre Fournier pour les cours irréfutables de HPC, et pour nos échanges) et en breuvages divins de tous azimuts. Vivement un cours d'utilisation de ta toute dernière application!

Également, un grand merci à Sophie, Sylvie et Marie-Christine pour avoir été indulgentes avec ma relation peu avertie avec le côté administratif de la recherche.

Je tiens également à remercier Florent Brenguier, Philippe Roux et Claire Prada pour avoir participé à mes comités de thèse, et pour avoir partagé leur vision des choses avec moi.

Also I would like to thank Gregor Hillers for our very interesting discussions, and because you opened my mind to the inverse filter technique! It was a great step for me, and I hope we will share again in the future.

Je tiens aussi à remercier toute l'équipe étudiante qui façonne l'agréable atmosphère dans laquelle j'ai pris plaisir à bosser pendant ces 3 ans et demi.

En particulier, merci à la team du bureau 312, aka le pays du petitbacistan, équipe du tonnerre où matière grise et murs ne font qu'un, où le squash est religion, le petit bac de rigueur et où l'entraide est maître. Merci à Flo pour sa bonne humeur quotidienne et pour nos merveilleux échanges musicaux. Merci aussi pour tes phrases très spirituelles que tu as le don de sortir, qui te font rire très fort, et que tu appelles des "blagues". J'ai beau manger des clowns tous les matins, je n'arrive pas à ta cheville. Clara, merci pour tous nos bons moments de discussions et aussi pour nos débats enflammés qu'ils soient ronds et en caoutchouc ou bien de requêtes Google pour savoir si « tableau » est un objet noir ou non. Merci à Kevin pour son humour fin et sophistiqué qui ne me laissera jamais indifférent et pour nos échanges également fins et sophistiqués sur LA grande question que tout le monde se pose en géosciences : comment faire une superbe carte en un minimum de lignes de code? Au passage, merci Harsha pour nous avoir transmis l'Excalibur des sofas, on en a bien pris soin pendant Roland Garros! Georges, tu n'es pas dans le bureau, mais c'est tout comme! Merci pour nos échanges de tous genres, pour cette ambiance de travail et pour nos débats passionnants tous azimuts pleins de vie.

Quentin, tu n'es pas dans le labo, mais tu es bel et bien présent dans le bureau à travers ce superbe traité du signal que nous avons co-écrit jadis! Merci pour toutes nos discussions, et ce depuis des cinq dernières (superbes) années. C'est toujours un plaisir d'échanger avec toi, et j'espère pouvoir continuer à échanger quels que soient tes choix futurs (car je crois avoir compris que tu n'es point homme facile, et que tu as résisté avec bravoure à la séduction du monde de la recherche).

Merci également Jean pour notre collaboration de dernière minute qui m'a beaucoup apporté, tant pour ta bonne humeur que pour tes conseils avisés. Ce sera un grand plaisir de continuer dans cette quête de décryptage de ces matrices de covariance, dans les mois qui viennent comme dans des futurs plus lointains si tu es partant!

Je tiens beaucoup à remercier les personnes avec lesquelles j'ai pu discuter aussi bien sur le plan scientifique ou académique que sur le plan amical, pour tous les bons moments, merci Manon, Martha, Mathias, Lambert, Thomas, Marine, Vincent, Vincent, Lise, Aurélien, Thomas, William, Guigui, Laulau, Christian, Thjis, David, Pierre, Lambert, Amaury (Bill pour les intimes), Marthouille, Amandine, Amandine, Rémi, Chloé, Antonio, Alessandro, Diane, Dodo, Antoine, John, Kairly, Sergei, et à toutes les personnes que j'oublie sur le papier mais pas dans la tête, merci à tous!

Merci également à Lambert, Boris, Pierre, Nobu, Martin, Hugo, Agnès et Claudio pour nos grandes sessions musicales!

Je remercie également Boulet, Rendall Monroe, et Marc Antoine Matthieu, mes auteurs préférés pour m'avoir accompagné pendant ma thèse.

Merci aussi à la com', en particulier, Pierre-Olivier, Emmelyne et Joël. Aussi, merci à Christine pour ses cours d'anglais, et un grand merci à Laure Meynadier pour la bien-

veillance qu'elle porte à ses étudiants sans arrêt.

Un grand merci à Olivier Devauchelle pour m'avoir donné l'occasion de m'éclater en donnant des cours. Et merci à Pascal Godefroy pour nous y avoir aidé! Merci aussi à Zarie pour nos retrouvailles le midi.

Papa, merci pour tous tes bons conseils qui m'ont guidé tout au long de mon parcours, et ce bien avant la thèse. Je te dois beaucoup, et j'espère pouvoir continuer à profiter de cette énergie que tu as su me transmettre. Merci à Myriam pour sa patience et son précieux soutien connaissant ma petite nature! Aussi, merci à ma grande famille Sasa, Anto, Gigi, Roxane, Adri, Élo, Nathan, Laura et bien sûr Blandine et Olivier.

Et une spéciale dédicace à Clem, Hugo et Gagou — aka les loulous — merci du fond du cœur! Merci pour ces grandes soirées enflammées de débats à haute voix, agrémentées d'eau (et de quelques extraits de malt et de houblon) où nous avons toujours refait le monde en étant du même avis mais sans jamais être d'accord! Vous avez été là pour les coups durs comme pour les coups de mou, et c'était plus que bon pour moi de vous savoir à mes côtés. Même loin, je vous saurai dans le coin. (Cette dernière phrase est à lire avec l'accent Gaillacois bien sûr).

Aussi je voudrais remercier chaleureusement à Cril et Brigitte pour tous les super conseils en la matière et pour le soutien permanent!

Enfin, merci Maga pour ton soutien incessant et pour cette belle énergie avec laquelle tu sais m'aider à franchir les étapes. Surtout, tu as toujours su trouver les mots pour me rassurer avant ces moments où on monte sur des petites estrades pour parler science devant un immense public d'initiés (qui me fera croire qu'on peut un jour s'y faire?). Pour finir, merci pour m'avoir poussé à prendre des vacances quand il le fallait! Je serai là pour toi le moment venu.

Table des matières

Résumé	3
Summary	5
Remerciements	7
Table des matières	12
Liste des figures	16
Introduction	17
1 Du bruit sismique ambiant aux fonctions de Green	21
1.1 Ondes élastiques et fonctions de Green	22
1.1.1 Équation d'onde dans un milieu élastique homogène et isotrope	22
1.1.2 Les différents types d'ondes sismiques	22
1.1.3 La fonction de Green, solution particulière de l'équation d'onde	25
1.2 Estimation de la fonction de Green empirique depuis le bruit sismique	26
1.2.1 Anatomie du bruit sismique ambiant	26
1.2.2 Estimation de la fonction de Green	30
1.2.3 Limitations de la méthode	32
1.3 Motivations	34
2 Matrice de covariance	37
2.1 Formalisme de covariance et d'intercorrélation	38
2.1.1 Variance, covariance et matrice de covariance	38
2.1.2 Propriétés de la matrice de covariance dans le cas sismique	40
2.1.3 Valeurs propres de la matrice de covariance	41
2.1.4 Espace signal et espace bruit	43
2.2 Formation de voie	43
2.3 Matrice de covariance estimée depuis des données réelles	44
2.3.1 Estimateur de la matrice de covariance	44
2.3.2 Rang de l'estimateur	44
2.3.3 Revue des méthodes de séparation d'espaces signal et bruit	44
2.4 Largeur spectrale de la matrice de covariance	47
2.5 Calcul de la matrice de covariance depuis des données réelles	48
2.5.1 Données sismiques	48
2.5.2 Calcul de la matrice de covariance	50

3	Analyse des données du Piton de la Fournaise et de l'USArray	53
3.1	Analyse des données du Piton de la Fournaise	54
3.1.1	Piton de la Fournaise seismic network	54
3.1.2	Synthetic tests	55
3.1.3	Analysis of the Undervolc data	58
3.1.4	Discussion and conclusions	68
3.2	Application to the transportable component of the USArray in 2010	70
3.2.1	Selection of the parameters	70
3.2.2	Results	71
3.2.3	Observation of the teleseismic activity	71
3.2.4	Microseismic noise: surface waves generated by swells	74
3.2.5	Microseismic noise: teleseismic P-waves	74
3.2.6	26-second microseism	76
3.2.7	Conclusions	76
4	Ambient noise imaging from anisotropic ambient noise	79
4.1	Introduction	80
4.2	Theory	82
4.2.1	Analogy between time-reversal and cross-correlation	82
4.2.2	Optimal focusing with spatio-temporal inverse filter	84
4.2.3	Equalization of the covariance matrix spectrum	85
4.2.4	Derivation of the eigenvalue threshold from the Vekua theory	86
4.3	US Transportable Array geometry	88
4.4	Illustration of the method with a synthetic test	89
4.5	Numerical experiment	91
4.5.1	Numerical setup	91
4.5.2	Protocol	92
4.5.3	Inter-station cross-correlations in presence of a strong source	96
4.5.4	Equalization of the covariance matrix spectrum	96
4.5.5	Travel-time measurement error, and travel-time inversion	98
4.6	Application to the 2010, Maule M8.8 earthquake	101
4.6.1	Wavefield anisotropy in presence of earthquakes	101
4.6.2	Equalization of the M8.8 Maule earthquake	104
4.6.3	Eigenvector-based surface waves selection	104
4.7	Discussion and conclusions	109
	Conclusions et Perspectives	111
	Références	115

Table des figures

1	Séismes mondiaux de 1970 à aujourd'hui	17
2	Large Aperture Seismic Array	19
3	Global Seismographic Network	19
4	Réseau sismique dense de Long Beach, Californie	20
1.1	Types d'ondes élastiques	23
1.2	Exemple de sismogramme.....	24
1.3	Probabilité de densité spectrale de puissance du bruit microsismique	27
1.4	Génération de microséismes par le système Atmosphère-Océan-Terre	28
1.5	Répartition des sources de bruit et intercorrélation	33
2.1	Covariances spatiales théoriques	40
2.2	Estimation de la matrice de covariance	45
2.3	Données sismiques pour l'analyse de la covariance.....	49
2.4	Calcul de la matrice de covariance depuis des données réelles	52
3.1	Map of the Piton de la Fournaise monitoring system.....	55
3.2	Construction of the synthetic wavefield.....	56
3.3	Covariance matrix spectral width as a function of the spatial coherence	57
3.4	Numerical seismic noise covariance matrix spectral width	57
3.5	Spectral and temporal normalization techniques	59
3.6	Covariance matrix spectral width: the noise vs. earthquake test.....	60
3.7	Covariance matrix spectral width from one year of the UnverVolc data	62
3.8	Power spectra from Undervolc and RER stations.	63
3.9	Example of earthquake detection around June, 12 of 2010.	64
3.10	Observation of the teleseismic activity.	66
3.11	Earthquake detector from spectral width	66
3.12	Observation of the seismo-volcanic activity at frequencies above 1 Hz.....	67
3.13	Détectabilité of earthquakes from the covariance matrix spectral width	69
3.14	Location of the USArray 121 seismic stations	70
3.15	Covariance matrix spectral width from one year of the USArray data	72
3.16	Earthquake observed at USArray	73
3.17	Spectral features in the covariance matrix spectral width	75
4.1	Time-reversal operation.....	84
4.2	Vekua theory illustrated with Bessel functions of the first kind	88
4.3	Selected USArray stations used in the spatial equalization approach	89
4.4	Synthetic spatial equalization	90
4.5	Numerical setup of the 2-D acoustic finite-differences simulation.	92

4.6	Snapshots of the numerical estimation of the Green's functions.....	94
4.7	Numerical and real Green's comparison	95
4.8	Numerical cross-correlation and beamforming analysis	97
4.9	Numerical experiment: equalization technique.....	98
4.10	Numerical experiment: comparison of the measured travel-times	99
4.11	Numerical travel-times inversion.....	100
4.12	U.S.G.S. Earthquake catalog of 2010	102
4.13	Wavefield anisotropy in presence of earthquakes	103
4.14	Spatial equalization of the the 2010 M8.8 Maule earthquake.....	105
4.15	Selection of surface waves from the covariance matrix eigenvectors	106
4.16	Cross-correlation in presence of the Maule earthquake (aligned)	107
4.17	Cross-correlation in presence of the Maule earthquake (orthogonal)	108

Introduction

Les ondes sismiques sont des perturbations mécaniques qui se propagent dans le milieu terrestre. Elles se traduisent par des variations du déplacement à la surface du sol, mesurables à l'aide de sismomètres. Parmi les grandes problématiques de la sismologie figure l'analyse des propriétés élastiques de la Terre grâce à une étude détaillée des ondes sismiques, ainsi que la surveillance des objets géologiques actifs. De nombreuses études de la structure interne du globe par les ondes sismiques ont été réalisées depuis près d'un siècle et ont permis de faire d'importantes découvertes à des échelles globales comme la mise en évidence du noyau interne ([Lehmann 1936](#)). Des modèles de Terre ont vu le jour dès le début des années 1980, comme le *Preliminary Reference Earth Model* (PREM, [Dziewonski and Anderson 1981](#)), ou le modèle IASP91 ([Kennett and Engdahl 1991](#)) en analysant un grand nombre de signaux en fonction de différentes distances source-récepteur. Ces modèles décrivent les propriétés élastiques de la Terre dans la dimension radiale et sont à symétrie sphérique.

Des modèles tomographiques faisant état des hétérogénéités latérales ont également été obtenus depuis les années 1980, comme [Romanowicz \(1979\)](#), [Marone et al. \(2007\)](#) à l'échelle régionale du continent nord-américain et à échelle globale [Li and Romanowicz \(1996\)](#) et continuent d'être améliorés de nos jours. Il existe cependant des facteurs limitants les performances de la tomographie de séismes. D'abord, nous pouvons noter que les séismes prennent naissance aux frontières des plaques tectoniques, comme le montre la figure 1. Cette répartition irrégulière des séismes implique une résolution limitée dans certaines régions situées loin des jonctions des plaques. Ensuite, on remarque que le taux d'occurrence des séismes de forte magnitude est faible, ce qui limite considérablement les applications en matière de surveillance, c'est-à-dire d'études tomographiques réalisées à court terme pour observer une évolution des propriétés élastiques du milieu. À ces deux limitations principales s'ajoute également celle de la pollution du bruit sismique ambiant, qui perturbe fortement l'analyse de signaux de faible amplitude issus de séismes de faible magnitude. Des efforts d'amélioration ont lieu en traitement du signal pour permettre la détection de séismes de plus en plus faibles.

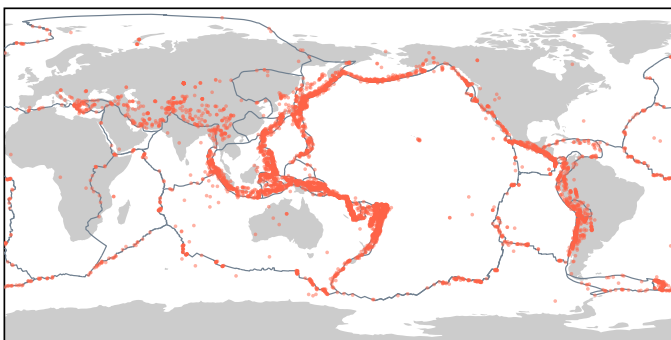


Figure 1 – Répartition globale des épicentres de séismes de magnitude supérieure ou égale à 6 de 1970 à 2016. Les épicentres sont indiqués par des points rouges. Les limites des plaques tectoniques sont représentées en lignes noires. La liste des épicentres est obtenue dans la base de données de l'United States Geological Survey ([USGS](#)).

Il existe également des études sismiques à échelle très locales menées à l'aide de sources actives, c'est-à-dire contrôlées par l'être humain. Il s'agit généralement de camions vibrants, ou bien de sources explosives. Ces dispositifs sont très coûteux à établir et servent souvent dans le domaine de l'exploration pétrolière pour obtenir des cartes du sous-sol à des échelles locales et à haute résolution.

Toutes ces approches se fondent sur l'estimation et l'analyse de la réponse élastique de la Terre, réponse qui contient toutes les informations relatives à la propagation des ondes sismiques. Dans un cadre formel, et dans l'hypothèse d'un milieu élastique linéaire, la réponse du milieu entre une source située au point de coordonnées \mathbf{r}_s et un point de mesure de coordonnées \mathbf{r} peut être caractérisée en fonction de la fréquence f par la fonction $G(\mathbf{r}, \mathbf{r}_s, f)$, également connue sous le nom de fonction de Green, du nom de son inventeur George Green qui les introduisit en 1828 dans le domaine de l'électromagnétisme ([Green 1828](#)). Ainsi, le déplacement $u(\mathbf{r}, f)$ enregistré sur une station sismique¹ située au point de mesure de coordonnées \mathbf{r} s'exprime alors comme combinaison linéaire de toutes les sources

$$u(\mathbf{r}, f) = \int G(\mathbf{r}, \mathbf{r}_s, f) s(\mathbf{r}_s, f) d^3 \mathbf{r}_s. \quad (1)$$

La fonction de Green est particulièrement recherchée en sismologie, car elle inclut tous les effets liés à la propagation et renseigne donc sur les propriétés internes de la Terre.

Une source est considérée comme ponctuelle lorsque son étendue est suffisamment petite en comparaison à la longueur d'onde. Si de plus, sa puissance est largement supérieure à celle des autres sources sismiques, alors le champ $u(\mathbf{r}, f)$ mesuré donne accès à la fonction de Green. En effet, l'équation 1 se transforme dans ce cas en

$$u(\mathbf{r}, f) = G(\mathbf{r}, \mathbf{r}_s, f) s(f). \quad (2)$$

Le problème fondamental posé en tomographie classique est l'inversion de cette dernière équation pour extraire $G(\mathbf{r}, \mathbf{r}_s, f)$. Récemment, des études ont montré que l'on pouvait tirer profit du bruit sismique ambiant naturellement présent sur Terre (par exemple, [Campillo 2003](#), [Shapiro and Campillo 2004](#), [Wapenaar 2004](#)). Effectivement, l'information déterministe que représente la fonction de Green est présente dans l'intercorrélation $R_{AB}(\tau)$ du champ d'onde diffus mesuré à deux points A et B , de telle sorte que

$$\frac{dR_{AB}(\tau)}{d\tau} \approx G(\mathbf{r}_A, \mathbf{r}_B, \tau) - G(\mathbf{r}_B, \mathbf{r}_A, -\tau) \quad (3)$$

Cette nouvelle classe de méthodes a révolutionné l'imagerie sismique passive, puisqu'elle permet des applications remarquables dans le domaine de la tomographie, et de la surveillance de zones actives. Bien que la pertinence de ces méthodes n'est plus à démontrer, nous verrons qu'elles nécessitent que le bruit sismique soit isotrope ce qui est rarement le cas en pratique. En effet, si la tomographie classique considère le bruit microsismique comme néfaste, la méthode passive supporte mal la présence de sources "cohérentes" telles que les séismes. Nous verrons que de nombreuses études ont été et sont encore menées pour définir des filtres qui visent à réduire l'influence de telles sources. Cette thèse a pris naissance dans cette même dynamique.

Le travail présenté dans ce manuscrit s'inspire largement de méthodes de traitement d'antenne et nécessite donc de travailler sur des données sismiques issues de réseaux

1. Par souci de simplicité, le caractère vectoriel des ondes sismiques ainsi que la réponse instrumentale sont omis sans perte de généralité

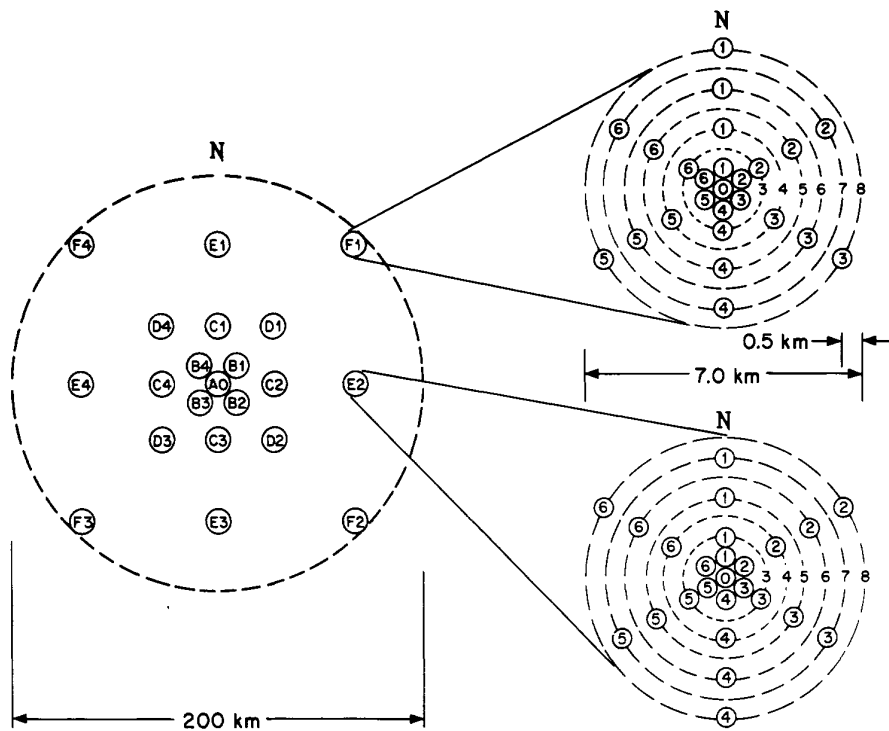


Figure 2 – Large Aperture Seismic Array. (figure extraite de [Frosch and Green 1966](#)). Déploiement du réseau expérimental dans l'est du Montana, aux États-Unis.

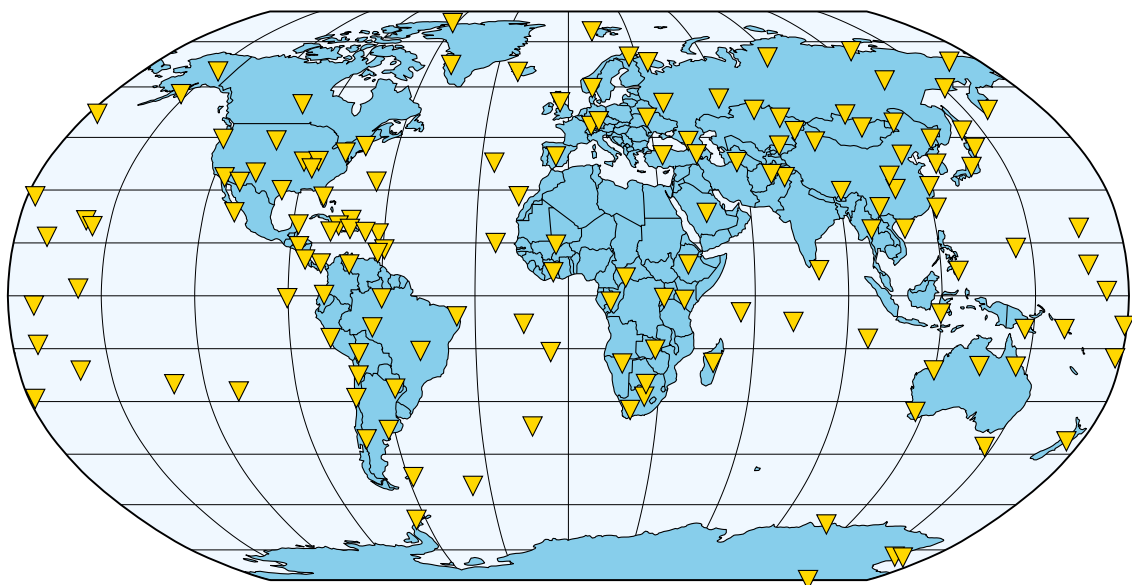


Figure 3 – The Global Seismographic Network. (figure disponible à l'adresse <https://www.iris.edu/hq/programs/gsn/maps>). Ce réseau collaboratif contient plus de 150 stations sismiques, indiquées par des étoiles, qui enregistrent et transmettent en temps réel les données sismiques au centre de gestion des données de l'IRIS. Ces données sont libres d'accès pour tous.

denses. Le déploiement de réseaux de stations sismiques a connu un essor au milieu des années 60 dans un contexte de guerre pour détecter et localiser des essais nucléaires.

Les deux premiers grands réseaux qui ont été déployés sont le *Large Aperture Seismic Array* (LASA, [Frosch and Green 1966](#)), présenté en figure 2, et le *NORwegian Seismic ARray* (ou NORSAR, voir par exemple [Bungum et al. 1971](#)). À l’heure actuelle, de plus en plus de réseaux sont installés dans le monde à des échelles locales (comme l’expérience menée à Long Beach en Californie [Nakata et al. 2015](#); présentée en figure 4) ou globales (comme le *Global Seismographic Network* présenté dans [Gee and Leith 2011](#); et en figure 3). L’utilisation de réseaux sismiques est donc plus que jamais d’actualité, ce qui justifie le développement des méthodes présentées dans ce manuscrit.

Les réseaux de stations sismiques sont des outils d’observation très puissants, car ils donnent accès à des échantillons spatio-temporels du champ d’onde. Une approche élémentaire en traitement de réseau, communément appelée formation de voie (*beamforming* en anglais), est l’estimation de la longueur et de la direction du champ incident. Dans l’optique d’améliorer les performances de la formation de voie, nous verrons qu’un grand nombre de méthodes haute-résolution ont vu le jour. Ces méthodes sont principalement fondées sur l’acquisition de la matrice de covariance du réseau. La définition mathématique de la matrice de covariance est

$$C_{ij}(f) = \mathbb{E}[u_i(f)u_j^*(f)]. \quad (4)$$

où $u_i(f)$ est le signal reçu sur la i -ème station et \mathbb{E} l’espérance mathématique.

Au cours de ma thèse, j’ai étudié de manière approfondie certaines propriétés de la matrice de covariance. En particulier, dans un premier temps, il est montré sur deux jeux de données sismiques que l’étude de la distribution des valeurs propres de $C_{ij}(f)$ permet une estimation fine de la “cohérence du champ”. Dans un second temps, nous proposons une méthode pour rendre isotrope le bruit par égalisation des valeurs propres dans la matrice de covariance.

Ainsi, le présent manuscrit est organisé comme il suit. D’abord, les théories principales nécessaires à la compréhension du manuscrit sont présentées dans le chapitre 1. Le chapitre 2 s’applique plus particulièrement à décrire la matrice de covariance, dans un cadre formel puis dans le contexte particulier de la sismologie de réseaux. Dans ce même chapitre est présentée une première méthode qui a été développée au cours de cette thèse, et qui permet d’analyser le degré de cohérence du champ d’onde sismique mesuré. Une application de la méthode sur deux jeux de données issus de deux réseaux différents est ensuite présentée en chapitre 3. Finalement, une seconde méthode qui s’intéresse plus particulièrement à l’amélioration de la reconstruction des fonctions de Green depuis un bruit ambiant mal conditionné est présentée en chapitre 4.

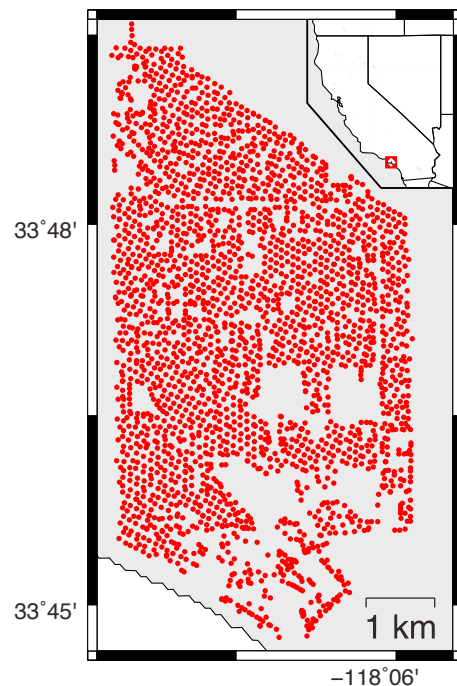


Figure 4 – Réseau sismique dense de Long Beach en Californie. (figure extraite de [Nakata et al. 2015](#)). La position des 2500 capteurs est indiquée en points rouges.

Chapitre 1

Du bruit sismique ambiant aux fonctions de Green : état de l'art

1.1	Ondes élastiques et fonctions de Green	22
1.1.1	Équation d'onde dans un milieu élastique homogène et isotrope.....	22
1.1.2	Les différents types d'ondes sismiques	22
1.1.3	La fonction de Green, solution particulière de l'équation d'onde	25
1.2	Estimation de la fonction de Green empirique depuis le bruit sismique	26
1.2.1	Anatomie du bruit sismique ambiant.....	26
1.2.2	Estimation de la fonction de Green	30
1.2.3	Limitations de la méthode	32
1.3	Motivations.....	34

Ce chapitre introduit les théories et les méthodes fondamentales au travail réalisé au cours de cette thèse. La section 1.1 présente brièvement l'équation qui régit la propagation des ondes dans un milieu élastique. La notion de fonction de Green est alors introduite comme une solution particulière de cette équation. Après avoir décrit le bruit sismique ambiant naturellement présent sur Terre, la section 1.2 présente une méthode récente d'estimation de la fonction de Green depuis l'inter-corrélation du bruit ambiant. L'application de cette méthode s'est avérée fructueuse dans divers domaines comme la sismologie, l'océanographie et l'acoustique. Néanmoins, nous montrons dans la section 1.2 que les hypothèses sur les propriétés du bruit qui assurent la convergence de l'inter-corrélation vers la fonction de Green ne sont pas toujours vérifiées dans des cas concrets. Nous présenterons un état de l'art des techniques de traitement actuellement utilisées pour tenter de corriger ces écarts. À partir de tous ces éléments, les motivations et les objectifs de ce travail de thèse sont précisés en section 1.3.

1.1 Ondes élastiques et fonctions de Green

1.1.1 Équation d'onde dans un milieu élastique homogène et isotrope

L'équation des ondes dans un milieu homogène et isotrope exprime le champ des déplacements en trois dimensions $\mathbf{u}(\mathbf{r}, t)$ en fonction de la position dans l'espace \mathbf{r} et du temps t de façon générale comme

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} = \frac{\lambda + \mu}{\rho} \nabla \cdot (\nabla \mathbf{u}) + \frac{\mu}{\rho} \nabla \times \nabla \mathbf{u}, \quad (1.1)$$

où $\nabla = (\partial/\partial x)\mathbf{e}_x + (\partial/\partial y)\mathbf{e}_y + (\partial/\partial z)\mathbf{e}_z$ représente le gradient, \times représente le produit vectoriel, et où λ et μ sont les coefficients de Lamé qui décrivent les propriétés mécaniques propres au milieu considéré.

L'équation 1.1 peut être résolue lorsque les conditions aux limites du système sont connues, c'est-à-dire l'état du système au temps $t = 0$ et aux frontières $\mathbf{r} \in \mathcal{S}$. Selon les conditions aux limites considérées, l'équation 1.1 n'admet pas les mêmes solutions et décrit différentes familles d'ondes. Ces ondes sont présentées dans les paragraphes suivants, dans le cadre simplifié d'un milieu homogène et isotrope. Nous verrons que même si dans la pratique, la Terre n'est ni homogène ni isotrope et que ses conditions aux frontières sont complexes, ces familles d'ondes apparaissent clairement dans les sismogrammes.

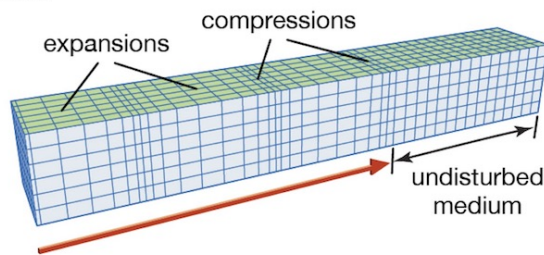
1.1.2 Les différents types d'ondes sismiques

Ondes de volume

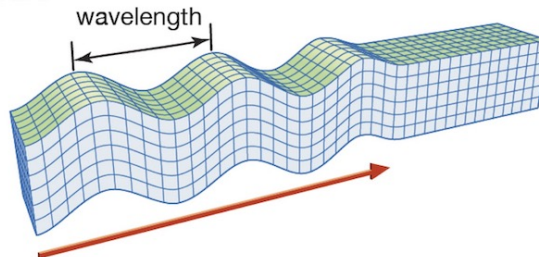
Dans un milieu élastique aux dimensions infinies, et en champ lointain, il est possible de montrer à partir de l'équation 1.1 que le déplacement se décompose comme une somme de deux ondes de volume : une onde de compression et une onde cisaillement (voir par exemple Catheline and Bencech 2015). En sismologie, ces deux ondes sont respectivement dénommées onde P (pour Primaire) et onde S (pour Secondaire). L'onde de compression, illustrée en haut de la figure 1.1, est une onde de célérité $\alpha = \sqrt{(\lambda + 2\mu)/\rho}$ polarisée suivant sa direction de propagation. L'onde de cisaillement (deuxième type d'onde sur la figure 1.1) dont la célérité s'exprime comme $\beta = \sqrt{\mu/\rho}$ est polarisée transversalement à la direction de propagation. Lorsque l'on considère un repère orienté (x, y, z) pour décrire le mouvement des particules, on distingue deux types d'ondes S : le type SV polarisé verticalement (dans la direction z) et le type SH polarisé horizontalement, c'est-à-dire

Main types of seismic waves

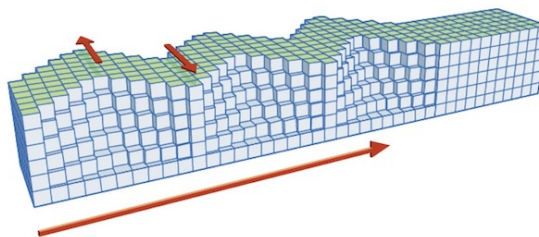
P wave



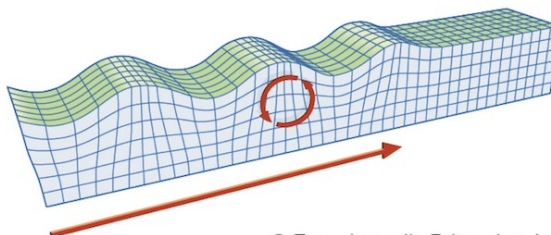
S wave



Love wave



Rayleigh wave



© Encyclopædia Britannica, Inc.

Figure 1.1 – Principaux types d’ondes sismiques. (Modifiée d’après <https://global.britannica.com/science/seismic-wave>, © Encyclopædia Britannica, Inc.) La direction de propagation est indiquée par une grande flèche rouge et le mouvement particulier des ondes de surface est précisé par des couples de petites flèches rouges. **Onde P** : onde volumique de compression polarisée dans la direction de propagation. **Onde S** : onde volumique de cisaillement, polarisée transversalement à la direction de propagation. On distingue les ondes S verticales (SV, représentée ainsi sur la figure) ou horizontales (SH). **Onde de Love** : onde de surface polarisée dans le plan de propagation. **Onde de Rayleigh** : onde de surface produisant un mouvement des particules vertical elliptique rétrograde. Les ondes de surface ont une amplitude qui décroît exponentiellement avec la profondeur.

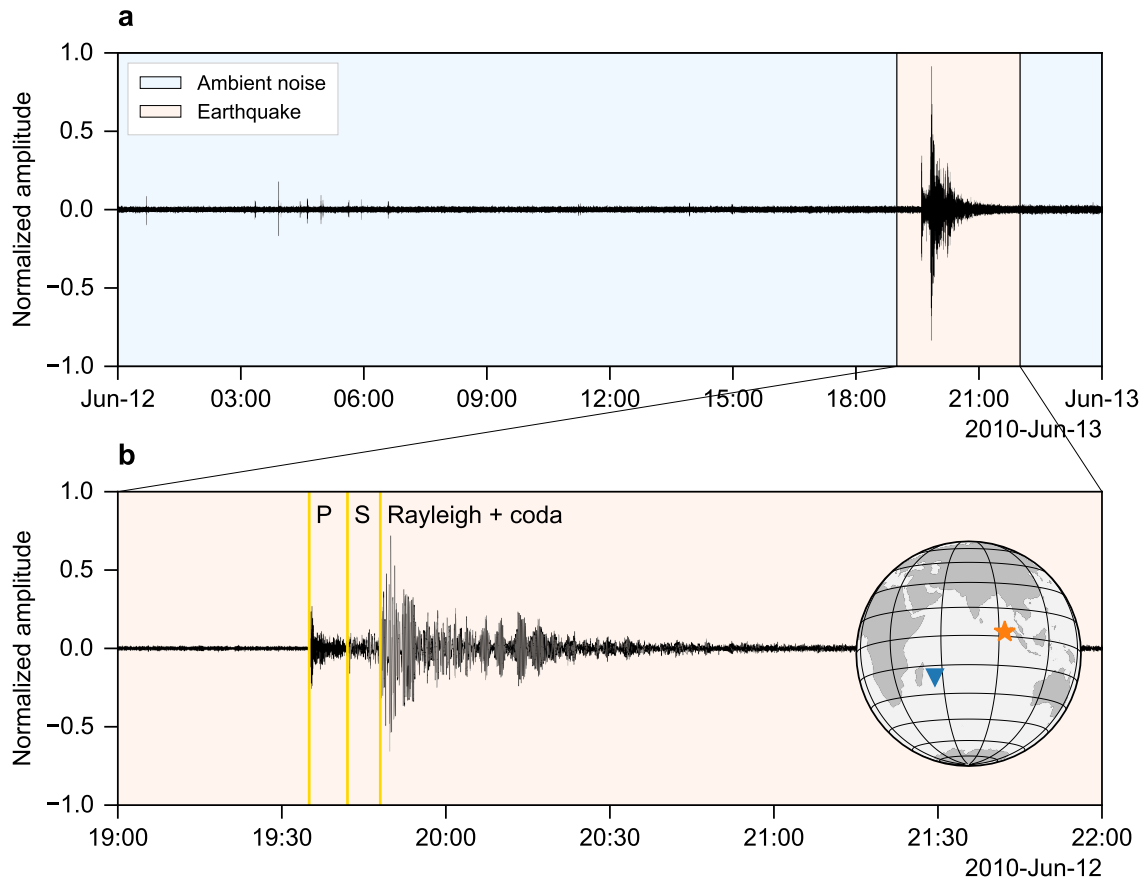


Figure 1.2 – Exemple de sismogramme enregistré par la composante verticale de la station sismique UV05 du réseau de surveillance du Piton de la Fournaise (île de la Réunion). (a) Enregistrement continu sur la journée du 12 juin 2010, indiquant la vitesse du sol en fonction du temps. En bleu, le bruit microsismique représente près de 90% de l’enregistrement. La zone hachurée indique un signal généré par le séisme de magnitude 7.5 des îles de Nicobar (à proximité de l’Inde, voir l’étoile rouge pour la position de l’épicentre sur le globe). (b) Zoom sur le signal généré par le séisme. Les ondes balistiques (en orange) correspondent à des ondes dont les trajets sont directs entre l’hypocentre et la station sismique. Les arrivées des ondes P, S et Rayleigh sont repérées grâce à l’algorithme TauP (Crotwell et al. 1999). La coda sismique (en vert) contient des arrivées plus diffuses.

dans le plan (x, y) . On observe directement que la célérité de l’onde P est plus grande que celle de l’onde S (c’est-à-dire $\alpha > \beta$), ce qui explique leurs dénominations respectives.

Ondes de surface

Dès lors que l’on considère un milieu semi-infini comme c’est le cas avec la surface terrestre, de nouveaux types d’ondes apparaissent. Il s’agit des ondes de surface, qui résultent des interactions entre les ondes P et S aux limites du système. Sur Terre, ce sont des ondes dispersives dont la vitesse dépend de la fréquence. Leur amplitude décroît exponentiellement en fonction de la profondeur, d’où leur nomenclature d’ondes de surface.

On distingue deux types d’ondes de surface : l’onde de Love, issue de l’interaction des ondes P et SH , et polarisée dans le plan de la surface ; et l’onde de Rayleigh, issue de l’interaction des ondes P et SV et polarisée dans le plan vertical (voir les deux schémas du bas en figure 1.1). Ces deux ondes sont souvent les plus destructrices, car elles présentent une forte amplitude. L’onde de Love est particulièrement à même de s’attaquer aux constructions urbaines avec sa polarisation particulière.

Observation des ondes sismiques sur un sismogramme

Un exemple de sismogramme enregistré pendant une journée entière par le canal vertical de la station sismique UV05 du réseau de surveillance du Piton de la Fournaise (volcan actif de l'île de la Réunion) est présenté en fonction du temps en figure 1.2a. La position de la station sismique est indiquée par un triangle noir sur la carte mondiale. Si cet enregistrement contient majoritairement du bruit sismique ambiant (à 90%), il inclut également un signal relatif à des ondes générées par le séisme du 12 juin 2010, atour de 19 : 20 UTC de magnitude 7.7 dont l'épicentre est proche des îles Nicobar (à proximité de l'Inde, voir l'étoile rouge sur le globe).

Un zoom autour de l'événement est présenté figure 1.2b. On y observe les premières arrivées d'ondes qualifiées d'ondes balistiques, car elles suivent un chemin direct entre l'épicentre et la station. Les arrivées respectives des ondes P , S et de Rayleigh sont repérées à l'aide de l'algorithme de calcul des temps de trajets TauP (Crotwell et al. 1999). Conformément à ce qui est énoncé précédemment, la première arrivée est celle de l'onde P , plus rapide, suivie de l'onde S . Les ondes de Rayleigh sont plus lentes, mais présentent une amplitude beaucoup plus forte. La suite du signal est plus diffuse et est nommée coda sismique. Elle est constituée des phases sismiques qui subissent des réflexions multiples au cours de leur trajet (voir par exemple Aki and Richards 2002; pour une description détaillée de la coda sismique).

1.1.3 La fonction de Green, solution particulière de l'équation d'onde

Lorsque l'on souhaite modéliser le processus de génération par une source, le formalisme des fonctions de Green est fréquemment utilisé. Sans perte de généralité, mais pour simplifier les dérivations suivantes, nous considérons l'équation des ondes pour une quantité scalaire, comme l'équation des ondes acoustiques qui régit le champ de pression. Si nous considérons également un terme source $s(\mathbf{r}, t)$, cette équation prend la forme

$$\left(\Delta - \frac{1}{c(\mathbf{r})^2} \frac{\partial^2}{\partial t^2} \right) u(\mathbf{r}, t) = s(\mathbf{r}, t), \quad (1.2)$$

où $c(\mathbf{r})$ est la vitesse des ondes que l'on suppose à présent dépendante de la position et où $\Delta = \nabla^2$ est l'opérateur Laplacien.

Les solutions de l'équation 1.2 peuvent s'exprimer dans le formalisme des fonctions de Green. Les fonctions de Green $G(\mathbf{r}, \mathbf{r}_s, t)$ sont définies comme les solutions de l'équation précédente, mais dont le terme source est une distribution de Dirac temporelle et spatiale (située au point de coordonnées \mathbf{r}_s)

$$\left(\Delta - \frac{1}{c(\mathbf{r})^2} \frac{\partial^2}{\partial t^2} \right) G(\mathbf{r}, \mathbf{r}_s, t) = \delta(\mathbf{r} - \mathbf{r}_s) \delta(t). \quad (1.3)$$

La fonction de Green G est donc la réponse du milieu à une impulsion spatiale et temporelle comme l'indique l'équation 1.3. La connaissance de ces fonctions particulières est intéressante, car elles permettent d'écrire le champ du déplacement $u(\mathbf{r}, t)$ comme une combinaison linéaire de ces dernières :

$$u(\mathbf{r}, t) = \int G(\mathbf{r}, \mathbf{r}_s, t - \tau) s(\mathbf{r}_s, \tau) d\tau d^3\mathbf{r}_s. \quad (1.4)$$

L'intégration est effectuée sur tout le volume et sur tous les temps. On remarquera ici que l'intégration sur le temps peut se réécrire comme un produit de convolution \otimes , c'est-à-

dire,

$$u(\mathbf{r}, t) = \int G(\mathbf{r}, \mathbf{r}_s, t) \otimes s(\mathbf{r}_s, t) d^3 \mathbf{r}_s. \quad (1.5)$$

Cette propriété provient de l'invariance par translation dans le temps de l'équation (1.2). Dans le domaine de Fourier, le produit de convolution devient un produit simple. Ainsi, le champ $u(\mathbf{r}, f)$ à la fréquence f s'écrit

$$u(\mathbf{r}, f) = \int G(\mathbf{r}, \mathbf{r}_s, f) s(\mathbf{r}_s, f) d^3 \mathbf{r}_s, \quad (1.6)$$

où $G(\mathbf{r}, \mathbf{r}_s, f)$ et $s(\mathbf{r}_s, f)$ sont les transformées de Fourier de $G(\mathbf{r}, \mathbf{r}_s, t)$ et $s(\mathbf{r}_s, t)$ respectivement. Avoir directement accès à une mesure de la fonction de Green en sismologie présente beaucoup d'intérêt, car elle permet de s'affranchir d'une description souvent complexe de la dépendance spatio-temporelle des sources.

Bien évidemment, dans la pratique, il est impossible d'avoir accès à la fonction de Green par l'expérience, car elle est issue d'une source impulsionnelle de durée nulle et d'amplitude infinie ! Néanmoins, par abus de langage, nous estimons avoir une évaluation de la fonction de Green, lorsque le champ mesuré est issu d'une source de faible dimension et de faible durée. Ainsi, au cours de ce manuscrit, et particulièrement dans le dernier chapitre, nous faisons volontairement l'amalgame entre la fonction de Green mathématique et la fonction de Green empirique, en gardant à l'esprit que ce sont deux objets distincts.

1.2 Estimation de la fonction de Green empirique depuis le bruit sismique

1.2.1 Anatomie du bruit sismique ambiant

Observations

Sur la figure 1.3a est représenté un sismogramme enregistré pendant deux jours sur la composante verticale de la station P31A de l'USArray (voir le triangle rouge sur le globe). Aucun événement ne domine de façon claire, cet enregistrement est donc majoritairement constitué d'un bruit sismique ambiant dont l'amplitude (ou plus exactement la variance) diminue faiblement au cours des deux journées.

Pour analyser le contenu spectral de ce sismogramme, un outil intéressant est l'histogramme des densités spectrales de puissance, présenté dans McNamara (2004). Cet outil permet d'étudier un signal à caractère aléatoire comme celui présenté en figure 1.3a. Comme nous le verrons plus tard dans ce manuscrit, l'estimation des densités spectrales de puissance d'un signal sismique est analogue à l'estimation de la matrice de covariance. Le sismogramme total est d'abord découpé en M fenêtres de signal $u_m(t)$ généralement recouvertes à 50%, dans lesquelles sont estimées les densités spectrales de puissances $\mathcal{P}_m(f)$. Pour cela, chaque fenêtre de signal $u_m(t)$ est à nouveau découpée en I sous-segments $u_{m,i}(t)$, eux-mêmes recouverts à 75%. L'énergie aux bords des sous-segments est atténuée par une fenêtre d'apodisation $\Omega(t)$. Ainsi, la densité spectrale de puissance de la fenêtre d'indice m s'exprime comme

$$\mathcal{P}_m(f) = \sum_{i=1}^I |\mathcal{F} [u_{m,i}(t)\Omega(t)](f)|^2, \quad (1.7)$$

où $\mathcal{F}[\cdot](f)$ désigne la transformation de Fourier.

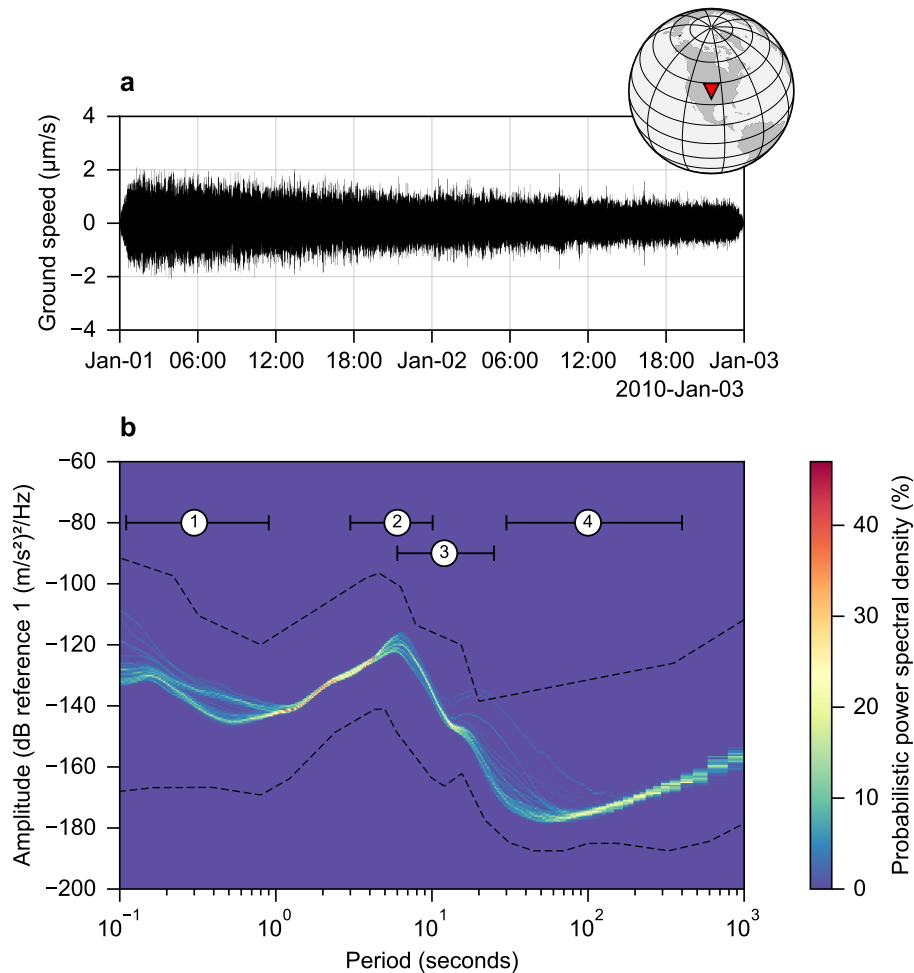


Figure 1.3 – Analyse de la densité spectrale de puissance du bruit sismique ambiant. (a) Sismogramme enregistré sur deux jours par le canal vertical de la station sismique USArray P31A, située dans l’État américain du Kansas (triangle rouge sur le globe). **(b)** Histogramme des densités spectrales de puissance de (a), méthode présentée par [McNamara \(2004\)](#) et détaillée dans le texte. Les niveaux de bruit sismique ambiant de références maximal et minimal observés dans le monde par [Peterson \(1993\)](#) sont indiqués en lignes grises. ① Sources de bruit locales (perturbation météorologique, activité anthropique...) ② Bruit microsismique secondaire ③ Bruit microsismique primaire ④ Hum sismique.

Dans notre cas, nous calculons les densités spectrales de puissance dans 43 fenêtres de 8 000 secondes, recouvertes à 50%. Le sismogramme original présente une mesure de vitesse, il est corrigé de sa réponse instrumentale dans le domaine fréquentiel pour obtenir une densité spectrale de puissance en accélération (c’est-à-dire, en $(\text{m/s}^2)^2/\text{Hz}$). Cette dernière est présentée en décibels ($\mathcal{P}_m^{dB} = 10 \log_{10} \mathcal{P}_m$), ce qui permet la comparaison de notre résultat avec les observations de [Peterson \(1993\)](#) sur les niveaux de bruit mondial minimal et maximal. L’histogramme est finalement construit à partir des 43 densités spectrales de puissances obtenues.

L’histogramme des densités spectrales de puissance est présenté en échelle de couleurs en figure 1.3b, en fonction de la période et de l’amplitude. Sur la même figure en lignes grises sont indiqués les niveaux de bruit minimal et maximal observés dans le monde par [Peterson \(1993\)](#). Les densités spectrales de puissances que nous obtenons sont bien encadrées par les deux courbes grises.

Nous observons clairement la présence d’un pic dominant autour de 7 secondes de pé-

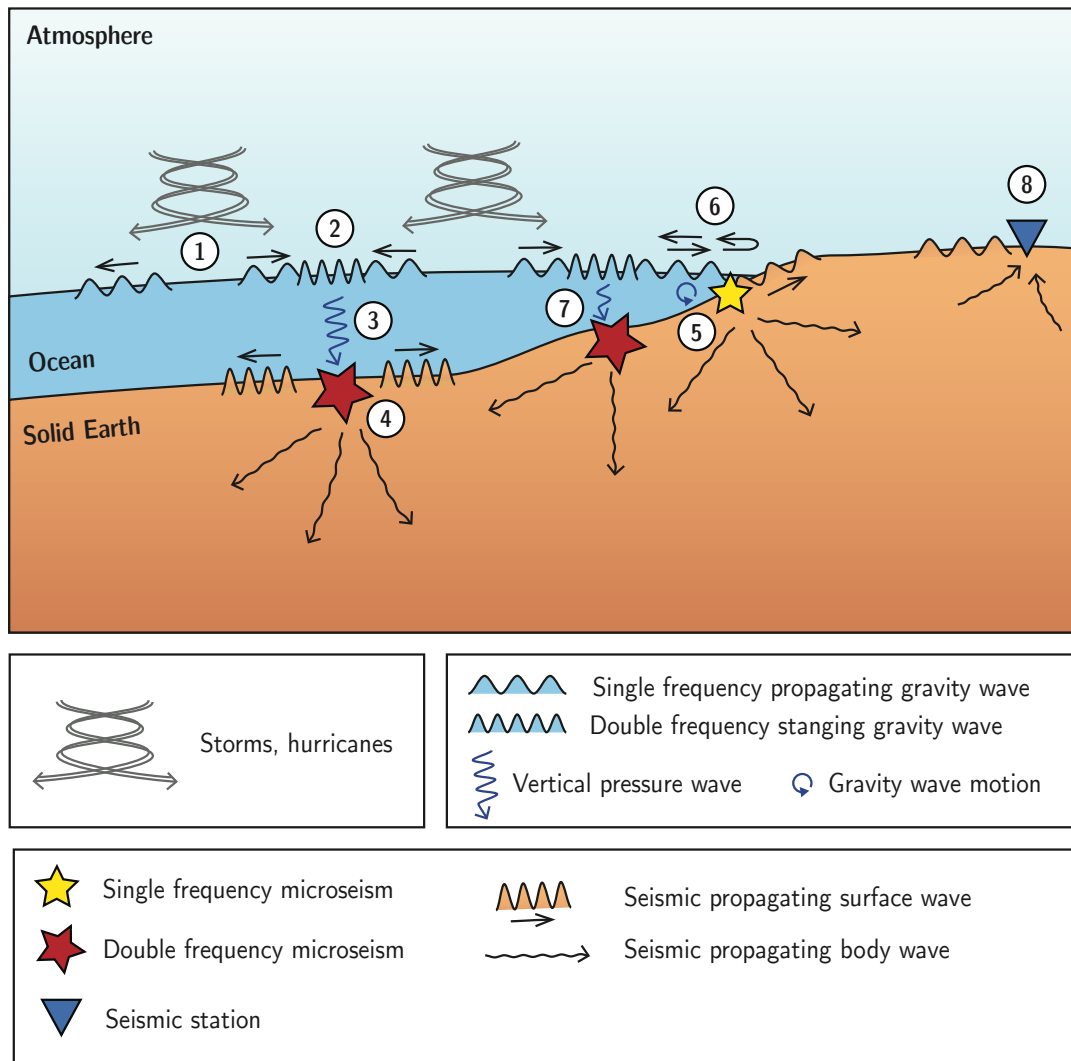


Figure 1.4 – Les microséismes issus du couplage Terre-Océan-Atmosphère.

- ① Les fluctuations de pression dans l'atmosphère produisent des ouragans, constitués de vents contraires à la source d'ondes de gravité se propageant à la surface des océans.
- ② Les ondes de gravité se propageant dans des directions opposées forment des ondes de gravité localement stationnaires.
- ③ Des ondes de pression causées par les ondes de gravité stationnaires se déplacent verticalement pour atteindre le sous-sol océanique.
- ④ Ces ondes verticales "frappent" le sol avec une intensité suffisante pour générer les microséismes secondaires, générant à leur tour des ondes sismiques de surface et de volume.
- ⑤ Les ondes de gravité primaires se propagent jusqu'au littoral, où elles interagissent directement avec le sous-sol océanique. Cette interaction est responsable de la génération des microséismes primaires, qui produisent également des ondes sismiques surfaciques et volumiques.
- ⑥ Une partie des ondes de gravité est réfléchiée par les côtes et interagit avec elle-même pour former des microséismes secondaires (étape 7).
- ⑦ Microséismes secondaires générés par l'étape 6.
- ⑧ Les ondes sismiques sont recueillies continuellement par des sismomètres.

riodes, suivi d'un second pic de plus faible amplitude autour de 12 secondes (secteurs (2) et (3) de la figure 1.3b). Ces deux pics sont d'origine océanique et sont enregistrables partout sur le globe, y compris en zone continentale comme c'est le cas dans notre exemple. Les paragraphes suivants s'efforcent de décrire ce bruit microsismique, en s'appuyant également du schéma récapitulatif proposé en figure 1.4. Deux autres secteurs (1) et (4) du domaine spectral sur la figure 1.3b sont également décrits dans les lignes qui suivent.

Le bruit microsismique secondaire ou microséismes à fréquence double

Les premières études de ce bruit microsismique observable partout sur le globe se sont directement accordées pour le situer d'origine océanique. L'énergie spectrale des microséismes secondaires se situe entre 3 et 12 secondes de période (ce qui correspond à la bande fréquentielle 0.08 à 0.35 Hz indiquée par le segment n°2 sur la figure 1.3b). Longuet-Higgins (1950) fut l'un des premiers à décrire ce bruit comme issu d'ondes de gravité océaniques de périodes égales se propageant dans des directions opposées et formant des ondes stationnaires de fréquence double (figure 1.4 ②), créant une pression oscillante sur le plancher océanique (figure 1.4 ③) et finalement responsable de la formation d'ondes sismiques dans le sous-sol océanique (figure 1.4 ④).

Ces ondes de gravité océaniques, communément appelées houle, sont dues à des variations de l'énergie potentielle de l'eau à la surface des océans occasionnées par des variations de pression atmosphérique (vents, ouragans). Elles sont dénommées ainsi, car la gravité est motrice dans la propagation de l'onde. Les différentes situations susceptibles de générer des ondes stationnaires sont décrites dans Arduin et al. (2011). Elles correspondent soit à un ouragan seul dont la vitesse de déplacement excède la vitesse des ondes de gravité ou plusieurs ouragans (figure 1.4 ①), ou encore à la réflexion sur les régions côtières des ondes de gravité (figure 1.4 ⑥ et ⑦).

Si ces microséismes sont majoritairement responsables de la génération d'ondes de Rayleigh (Kedar et al. 2008, Gualtieri et al. 2013), ils peuvent également générer des ondes de volume (Landès et al. 2010). Contrairement aux ondes de surface, les ondes de volume peuvent être utilisées pour retrouver la position des microséismes sur le globe (Gerstoft et al. 2008, Landès et al. 2010, Liu et al. 2016). Ces études montrent une relation forte entre l'origine des ondes de volume et la position des orages à la surface des océans, avec une forte variation saisonnière : en hiver, la majorité des sources est située dans l'hémisphère nord, et en été dans l'hémisphère sud. La profondeur de l'océan et la composition du sous-sol océanique sont des facteurs qui influent sur la puissance des microséismes secondaires comme le montrent entre autres Longuet-Higgins (1950), Kedar et al. (2008), Traer et al. (2012), Gualtieri et al. (2013).

Les microséismes primaires ou microséismes à fréquence simple

Les ondes de gravité responsables de la création des microséismes secondaires ont une amplitude qui décroît exponentiellement avec la profondeur dans l'océan. Lorsqu'elles se propagent jusqu'aux littoraux, elles peuvent également interagir directement avec le sous-sol océanique lorsque celui-ci est peu profond, comme le décrit par exemple Hasselmann (1963) (voir la figure 1.4 ⑦) dans la bande de périodes de 5 à 35 secondes (0.04 à 0.17 Hz, segment n°3 sur la figure 1.3b). La fréquence centrale de ce type de microséisme est directement liée à celle des ondes de gravité océaniques, d'où son appellation. Ces microséismes sont donc générés près des régions côtières, comme l'observent notamment Traer et al. (2012).

Le hum

À plus basse fréquence (entre 0.002 et 0.03 Hz, soit entre 30 et 500 s), un signal de plus faible amplitude est également observable (voir figure 1.3b 4^{ème} segment). Ce bruit d'amplitude environ 300 fois plus faible que celui des microséismes secondaires (≈ -40 dB) est l'expression des oscillations libres de la Terre (Nishida and Kobayashi 1999, Eks-tröm 2001, Rhie and Romanowicz 2004). Initialement, ces oscillations étaient considérées comme la répartition sur les modes propres de la Terre des excitations dues aux séismes de large amplitude. Tanimoto (2001) a cependant montré que l'énergie apportée par les séismes ne suffit pas à expliquer l'énergie visible dans le hum. Les avancées théoriques suggèrent alors que ce hum est généré par des ondes "infragravitaires" de très longues périodes. Ce signal est de suffisamment basse fréquence pour générer des ondes de Rayleigh multiorbitales susceptibles d'exciter les modes propres de la Terre (Webb 2008). Le hum est expliqué également par les modèles de Webb (2008), Traer and Gerstoft (2014) et de Arduin et al. (2015), où les théories de la génération des microséismes primaires et secondaires sont étendues aux très basses fréquences, en considérant des ondes infragravitaires. De la même façon que les microséismes primaires et secondaires, le hum peut être généré par l'interaction directe des ondes infragravitaires avec le sous-sol océanique, ou bien par interaction entre deux ondes infragravitaires se propageant dans des directions opposées.

1.2.2 Estimation de la fonction de Green

Ces sources de bruit naturellement présentes sur Terre sont responsables d'un signal à première vue stochastique dans les sismogrammes, comme le montre l'enregistrement de la figure 1.3a. Ce bruit est majoritairement présent dans les données sismologiques continues. De récentes études ont montré que le bruit ambiant présent permettait d'extraire une information déterministe comme la fonction de Green dans des milieux élastiques. La présente section s'intéresse au développement de cette théorie depuis le début des années 2000, et énonce les applications principales en tomographie sismique qui ont alors vu le jour.

De l'expérience à la théorie

Une méthode passive d'estimation de la fonction de Green a connu un essor grâce aux études menées notamment par Weaver and Lobkis (2001b) dans le domaine des ultrasons. Cette technique était déjà connue en héliosismologie pour mesurer les vitesses de propagation des ondes à la surface du Soleil (Duvall et al. 1993). Les acousticiens Weaver et Lobkis ont montré par l'expérience que les fluctuations de température qui excitent les modes acoustiques d'un bloc d'aluminium de façon aléatoire permettent de retrouver la fonction de Green entre deux points de la surface du bloc en corrélant les déplacements aléatoires mesurés en ces deux points. Ils utilisent alors le principe d'équipartition de l'énergie sur les différents modes acoustiques pour justifier théoriquement leurs observations.

Ils montrent également dans Weaver and Lobkis (2001a) que ce résultat est toujours valable pour un champ d'onde résultant de la diffusion d'une source impulsionnelle par un milieu réverbérant. Cette seconde approche s'explique par le fait qu'une source en milieu réverbérant finit par produire un champ d'onde dont l'énergie est équirépartie sur les différents modes de vibrations, c'est-à-dire que les modes ont une amplitude aléatoire avec une variance égale.

D'autres visions ont ensuite vu le jour pour expliquer et démontrer à nouveau ce théo-

rème. Toujours dans le milieu de l'acoustique, [Derode et al. \(2003\)](#) font l'analogie avec le retournement temporel — technique déjà utilisée en imagerie ultrasonore ([Fink 1992](#)) — pour démontrer le lien entre la fonction de Green et l'intercorrélation de champ diffus. Nous nous intéresserons particulièrement à cette approche dans le chapitre 4. Une autre approche est également proposée par [Snieder \(2004\)](#), [Roux et al. \(2005\)](#) qui généralise le théorème à des espaces ouverts en utilisant le principe de la phase stationnaire. Nous retiendrons également [Wapenaar \(2004\)](#) et [Wapenaar et al. \(2006\)](#) qui ont démontré comment cette théorie s'applique à un milieu élastique quelconque, dès lors qu'il est excité par un grand nombre de sources de bruit, c'est-à-dire un grand nombre de sources décorréelées et distribuées de façon homogène dans le milieu.

Nous pouvons ici traduire ces résultats dans un cadre plus formel. Si nous considérons un instant l'intercorrélation $R_{AB}(\tau)$ des sismogrammes $u_A(t)$ et $u_B(t)$ respectivement enregistrés à des stations sismiques A et B , telle que $R_{AB}(\tau) = \mathbb{E}[u_A(t)u_B(\tau - t)]$, avec \mathbb{E} représentant la moyenne statistique, nous avons (voir par exemple [Gouédard et al. 2008](#)) :

$$\frac{dR_{AB}(\tau)}{d\tau} \approx G(\mathbf{r}_A, \mathbf{r}_B, \tau) - G(\mathbf{r}_B, \mathbf{r}_A, -\tau) \quad (1.8)$$

où τ est le temps d'intercorrélation. Cette équation n'est valable que lorsque les conditions mentionnées plus haut sont vérifiées.

La partie de droite de l'équation 1.8 montre que nous ne retrouvons pas exactement la fonction de Green. Cependant, en réalité la différence entre la fonction de Green *causale* $G(\mathbf{r}_A, \mathbf{r}_B, \tau)$, qui concerne la propagation de la station B vers la station A et la fonction de Green *anti-causale* $G(\mathbf{r}_B, \mathbf{r}_A, -\tau)$ qui représente la propagation des ondes depuis la station A vers la station B . Dans un milieu réciproque dans un état proche de l'équilibre, l'équation 1.2 est symétrique par retournement temporel (c'est-à-dire en remplaçant t par $-t$), impliquant que $G(\mathbf{r}_B, \mathbf{r}_A, \tau) = G(\mathbf{r}_A, \mathbf{r}_B, -\tau)$. Ainsi dans le domaine spectral, il est intéressant de noter que l'équation précédente conduit à

$$C_{AB}(f) = \frac{2}{\omega} \text{Im}[G(\mathbf{r}_A, \mathbf{r}_B, f)] \quad (1.9)$$

où $C_{AB}(f) = \mathcal{F}R_{AB}(\tau)$ est la transformée de Fourier de la fonction d'intercorrélation, et où Im représente la partie imaginaire. Ici, $C_{AB}(f)$ représente la covariance spectrale des signaux $u_A(t)$ et $u_B(t)$. Cette remarque sera utile à l'introduction de la matrice de covariance dans le chapitre 2.

De la théorie à l'application en sismologie

Ce théorème fort implique que l'intercorrélation d'un champ diffus mesuré entre deux stations sismiques converge vers la fonction de Green du milieu entre ces deux stations, autrement dit, vers le signal qui serait mesuré sur une des stations si l'autre était remplacée par une source impulsionnelle. Cette approche permet par conséquent de convertir les différents sismomètres en sources virtuelles, puis d'utiliser, par exemple, les signaux issus de ces sources virtuelles dans une approche classique de tomographie.

En sismologie, la première application de cette approche été réalisée dans [Campillo \(2003\)](#), où la coda sismique (diffuse) de plusieurs séismes est utilisée pour reconstruire la fonction de Green entre deux stations sismiques. Cette première étude a constitué une preuve que l'application d'une telle méthode était possible en sismologie. Par la suite, [Shapiro and Campillo \(2004\)](#) reconstruisent les courbes de dispersion des ondes de Rayleigh sur de grandes distances à partir de la corrélation du bruit sismique ambiant. Ces

travaux se sont alors naturellement orientés vers une application en tomographie dans Shapiro et al. (2005) et Sabra et al. (2005b) où une carte haute résolution des vitesses des ondes de surface en Californie est obtenue par corrélation de bruit sismique ambiant.

La méthode est maintenant largement employée dans diverses régions du globe et à différentes échelles pour produire des analyses tomographiques du sous-sol terrestre. Des tomographies à échelle globale ont été obtenues à partir du hum dans Nishida et al. (2009) et plus récemment dans Haned et al. (2015). À échelle régionale, Yang et al. (2007) réalisent la tomographie d'ondes de surfaces du continent européen, suivi récemment par Behm et al. (2016) qui retrouvent également des ondes de Love dans l'est alpin. Les ondes de Love avaient été mises en évidence auparavant sur le continent nord américain notamment dans Ekström (2014). À des échelles locales, la méthode permet l'analyse de la structure interne de zones actives telles que les volcans, comme le montrent Brenguier et al. (2007) qui analysent la structure interne du Piton de la Fournaise grâce aux ondes de Rayleigh, mais également Jaxybulatov et al. (2014) pour l'analyse de la structure du supervolcan Toba, situé sur l'île de Sumatra ainsi que Lanza et al. (2016) dans l'analyse du volcan Pacaya au Guatemala. Toujours à échelle locale, la méthode est appliquée à l'analyse des failles comme dans Roux (2009) qui explore la structure de la faille de San Andreas à l'aide d'ondes de surface. Des applications voient également le jour dans le contexte de l'exploration pétrolière, comme il a été montré dans Mordret et al. (2013; 2014) sur l'analyse des données issues du champ pétrolifère de Valhall.

Les ondes de volumes peuvent également être reconstruites à partir des corrélations de bruit, comme le montrent Boué et al. (2013), Lin et al. (2013) qui parviennent à extraire des phases sismiques de noyau. À échelle locale, les ondes de volume sont également observées dans Nakata et al. (2015) sur le réseau dense de stations sismiques disposées à Long Beach en Californie.

Cette approche a également permis l'émergence d'une nouvelle technique de surveillance de zones actives, en analysant les corrélations de bruit dans des durées relativement courtes pour observer des variations de vitesse du sol de l'ordre de 1%. La première application est réalisée par Wegler and Sens-Schönfelder (2007) qui analyse les variations de la célérité dans le sol au Japon alors qu'un séisme a lieu, depuis l'autocorrélation du signal enregistré par une station seule, autocorrélation qui peut être interprétée comme une fonction de Green où la source et la station sont situées au même endroit. Sur le volcan du Piton de la Fournaise, Brenguier et al. (2008b) compare les temps de trajets obtenus sur des fonctions d'intercorrélation court terme avec ceux obtenus sur le long terme. Les auteurs mettent alors en évidence des précurseurs aux éruptions liées aux remontées de matériel mantellique dans l'édifice volcanique qui font varier la contrainte dans le sol. Une expérience similaire est proposée dans Brenguier et al. (2008a), où l'analyse est portée sur la relaxation de la contrainte après une rupture sismique le long de la faille de San Andreas en Californie. Plus récemment encore, Brenguier et al. (2014) a observé les augmentations brutales des contraintes au sein des édifices volcaniques japonais qui ont eu lieu en 2011 pendant la rupture sismique du séisme de magnitude 9.0 au niveau de la région du Tōhoku au Japon (Sato et al. 2011).

1.2.3 Limitations de la méthode

Même si ces applications remarquables de l'extraction des fonctions de Green à partir du bruit sismique ambiant ont pu voir le jour en sismologie, des limitations notables subsistent. En effet, deux remarques importantes de la description du bruit sismique ambiant réalisée en sous-section 1.2.1 sont à noter. D'abord, les sources de bruit ne sont pas sta-

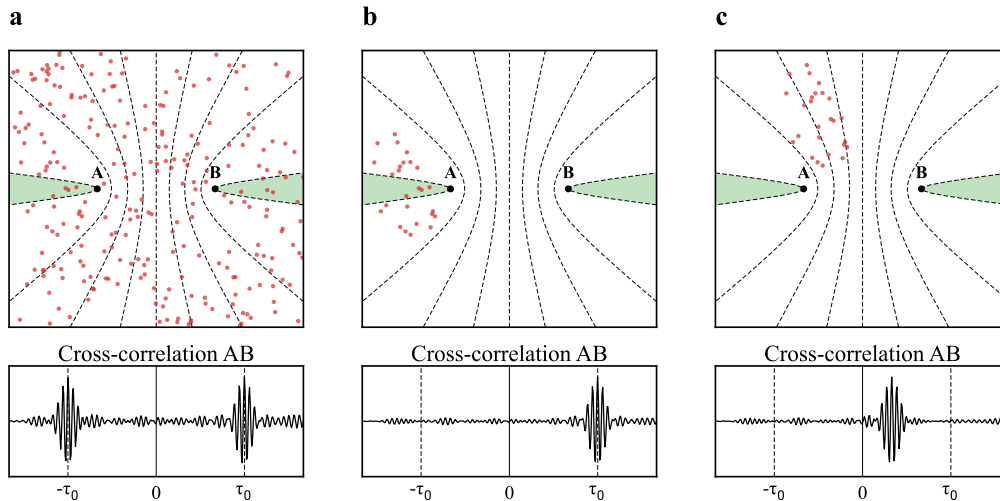


Figure 1.5 – Illustration de l'influence de la répartition des sources de bruit sur l'intercorrélation. (figure modifiée d'après [Stehly et al. 2006](#)). Les graphiques supérieurs illustrent la répartition des sources de bruit dans l'espace et les lignes pointillées indiquent les temps de trajets induits par les sources de bruit en fonction de leur position. **(a)** L'intercorrélation est symétrique lorsque les sources de bruit sont réparties aléatoirement et uniformément dans le milieu, impliquant l'émergence d'un front d'onde cohérent autour du temps de trajet τ_0 correspondant au trajet des ondes entre A et B. **(b)** L'intercorrélation est asymétrique et ne contient que des ondes qui se sont propagées de A vers B, car les sources sont situées proches de l'axe AB du côté de A. Le temps de trajet τ_0 qui correspond au trajet des ondes entre A et B est respecté. **(c)** L'intercorrélation est asymétrique, et cette fois le temps de trajet τ_0 est biaisé, car les sources de bruit sont localisées hors de l'axe AB.

tionnaires et présentent des variations d'amplitude saisonnières notamment liées aux positions des perturbations météorologiques. Ensuite, les sources de bruit sont générées à des positions préférentielles du globe, ne respectant ainsi pas les conditions nécessaires à la reconstruction des fonctions de Green.

Des études sur l'effet de la répartition des sources de bruit sur les fonctions d'intercorrélation montrent qu'une répartition anisotrope des sources de bruit implique l'émergence de fronts d'onde cohérents ne correspondant pas à des temps de trajet directs entre les stations ([Tsai 2009](#), [Weaver et al. 2009](#), [Froment et al. 2010](#)). En effet, les ondes qui arrivent avec une incidence proche de l'axe formé par le couple de stations sismiques interfèrent constructivement dans le processus de corrélation, alors que celles qui sont situées loin des axes ont tendance à s'annuler. Ce premier cas est illustré en figure 1.5a. Les lignes en pointillés indiquent les temps de trajet théoriques induits par les sources en fonction de leur position. Nous observons que seules les sources situées proches de l'axe ont contribué à l'émergence d'un front d'onde cohérent dans les parties causale et anti-causale, dont le temps de trajet correspond à celui des ondes qui se propagent de la station A à la station B par le chemin AB. Les corrélations sont donc symétriques quand la répartition des sources est symétrique elle-même.

Si la répartition des sources est anisotrope, la symétrie n'est pas respectée brisant ainsi l'annulation mutuelle des ondes à fort angle d'incidence. Certaines contributions non souhaitées peuvent apparaître, à des temps de trajets ne correspondant pas à celui du chemin réel qui relie les deux stations. Dans le cas où les sources sont toujours situées sur l'axe AB, mais d'un seul côté du système (figure 1.5b), seul le front d'onde qui s'est propagé de A vers B est retrouvé. L'amplitude de la fonction de corrélation est donc biaisée par

la mauvaise répartition des sources, comme le montre théoriquement entre autres Tsai (2009), Weaver et al. (2009), Froment et al. (2010). Également, lorsque les sources sont situées en dehors de l'axe AB uniquement, les temps de trajets sont biaisés et ne correspondent pas à des temps de trajet réels, c'est-à-dire au temps de trajet des ondes qui suivraient le chemin AB (figure 1.5c). Dans ce cas, les fonctions d'intercorrélation sont non seulement asymétriques, mais les temps de trajets mesurés sont faux. Les corrélations ne sont donc pas exploitables dans une telle configuration du champ d'onde comme le montrent également Kimman and Trampert (2010), Tian and Ritzwoller (2015).

Si les études tomographiques ont pu voir le jour, c'est en grande partie grâce au développement de traitements adaptés sur les traces sismiques, ayant pour but d'améliorer la répartition des sources et la stationnarité du bruit. D'abord, le signal issu des séismes qui sont par définition des sources de forte amplitude localisées à des régions préférentielles du globe (comme le montre la figure 1) ne sont pas pris en compte dans le processus de corrélation. Les fenêtres de signal contenant des événements sont souvent saturées en amplitude en considérant un seuil arbitraire, ou bien leur contribution est tout simplement réduite à l'aide d'une faible pondération (Sabra et al. 2005b, Hillers et al. 2014, Wu et al. 2016). Parfois, ces fenêtres sont tout simplement rejetées (Rhie and Romanowicz 2004, Shapiro et al. 2005, Brenguier et al. 2007, Seats et al. 2011), induisant une diminution de la quantité de données analysables. Si ce dernier point n'a que peu d'impact lorsque l'idée est de faire de la tomographie en moyennant les fonctions d'intercorrélation sur de longues durées, elle devient pénalisante dans le cadre de la surveillance qui nécessite un maximum de données pour effectuer des analyses dans le court terme.

L'anisotropie et la non-stationnarité des sources de bruit peut-être contrecarrée par des méthodes d'égalisation spectrale et temporelle des traces sismiques individuelles (traitement discuté dans Bensen et al. 2007). Ces techniques sont largement utilisées en sismologie et montrent une amélioration claire de l'extraction des fonctions de Green. Ces méthodes sont détaillées dans le chapitre suivant 2, et font l'objet d'une étude dans le chapitre 3.

Une technique consiste également à rejeter les corrélations dont l'asymétrie excède un certain niveau, soit en comparant les amplitudes des fronts d'onde cohérents causal et anti-causal, soit en regardant le rapport signal à bruit entre les fronts d'onde émergents et le bruit global observable sur le reste des fonctions d'intercorrélation. Là encore, cette technique se traduit par une diminution des données exploitables.

1.3 Motivations

Ce travail a pris naissance dans le but d'améliorer les techniques de corrélation de bruit afin de permettre l'extraction des fonctions de Green dans des contextes où le bruit sismique ambiant est inapproprié. Au cours de ce travail, nous nous sommes intéressés à un objet mathématique directement relié à la fonction d'intercorrélation qui est la covariance des signaux sismiques, définie comme étant le pendant fréquentiel de l'intercorrélation et présentée en première partie du chapitre 2. Plus précisément nous nous sommes intéressés à la matrice de covariance qui considère les signaux enregistrés par un réseau de stations sismiques.

Une première étape de ce travail a consisté à observer la "qualité" du champ d'onde sismique ambiant à travers les valeurs propres de la matrice de covariance. Par qualité, nous entendons l'isotropie et le niveau de cohérence du champ d'onde localement enregistré par le réseau de stations sismiques. Cette première étape nous a conduits au déve-

veloppement d'une méthode d'analyse du champ d'onde dans les dimensions temporelle et fréquentielle dans le but de définir correctement les zones temporelles et fréquentielles où les fonctions d'intercorrélation sont susceptibles de fournir une fonction de Green de qualité. Cette méthode est présentée en chapitre 2.

Si le degré de cohérence spatiale est analysé à travers la matrice de covariance, il est également possible de se servir des résultats obtenus pour analyser l'activité sismique, qu'il s'agisse de l'activité sismique locale ou télé-sismique. Nous avons également considéré cet aspect au cours du chapitre 3 en analysant l'activité sismique globale depuis le réseau de surveillance du Piton de la Fournaise. Une analyse de l'activité sismo-volcanique du Piton de la Fournaise est également proposée. Cette approche est d'un grand intérêt, car elle permet l'analyse de signaux émergents que les techniques de détection classiques peinent à détecter correctement. L'application de la méthode sur le jeu de données du réseau américain USArray nous a par ailleurs permis d'analyser des événements d'origine microsismique, et de retrouver des résultats principaux concernant les bandes de fréquence et les variations saisonnières des sources de bruit microsismique. Aussi, une analyse de l'efficacité des traitements d'égalisation des sismogrammes pour augmenter l'isotropie et la stationnarité des sources de bruit est proposée.

Ce travail s'est finalement penché en dernier lieu sur le développement d'une méthode de filtrage des sources à partir de la matrice de covariance des signaux, ce qui fait l'objet du chapitre 4. En s'inspirant des traitements du signal temporel et fréquentiel présentés dans [Bensen et al. \(2007\)](#), nous avons proposé d'étendre l'approche au spectre de la matrice de covariance, définissant ainsi une égalisation spatiale des signaux enregistrés. Cette méthode s'inscrit dans la continuité des études menées par [Gallot et al. \(2012\)](#) et [Carriere et al. \(2013\)](#), qui ont développé des méthodes similaires dans le domaine de la sismique et de l'océanographie.

Chapitre 2

Matrice de covariance de données enregistrées par un réseau de stations sismiques : théories et méthodes

2.1	Formalisme de covariance et d'intercorrélation	38
2.1.1	Variance, covariance et matrice de covariance.....	38
2.1.2	Propriétés de la matrice de covariance dans le cas sismique	40
2.1.3	Valeurs propres de la matrice de covariance	41
2.1.4	Espace signal et espace bruit	43
2.2	Formation de voie.....	43
2.3	Matrice de covariance estimée depuis des données réelles	44
2.3.1	Estimateur de la matrice de covariance	44
2.3.2	Rang de l'estimateur	44
2.3.3	Revue des méthodes de séparation d'espaces signal et bruit.....	44
2.4	Largeur spectrale de la matrice de covariance	47
2.5	Calcul de la matrice de covariance depuis des données réelles	48
2.5.1	Données sismiques	48
2.5.2	Calcul de la matrice de covariance	50

Ce chapitre se concentre plus particulièrement sur la matrice de covariance des signaux enregistrés sur un réseau de stations sismiques. On s'intéressera tout d'abord à une définition purement statistique de la matrice de covariance en section 2.1, où le comportement théorique attendu de la matrice de covariance est décrit à l'aide d'un modèle de signal simple. Cette première approche permet de décrire analytiquement le comportement attendu des valeurs propres de cette matrice de covariance. La méthode de formation de voie à partir de la matrice de covariance est également présentée en section 2.2. Une description de l'effet de l'estimation de la matrice de covariance depuis un nombre limité de données et de stations sismiques est ensuite proposée en section 2.3, en mettant l'accent sur le comportement des valeurs propres. Finalement, une méthode originale d'analyse du degré de cohérence spatiale du champ d'onde sismique qui a été développée au cours de cette thèse est présentée dans la section 2.4.

2.1 Formalisme de covariance et d'intercorrélation

Dans cette section, nous nous efforcerons de définir de manière rigoureuse les quantités que nous avons manipulées au cours de cette thèse, telle que la matrice de covariance et d'intercorrélation. Nous commencerons par une définition statistique de ces grandeurs puis nous verrons comment il est possible de les estimer à partir de données réelles.

2.1.1 Variance, covariance et matrice de covariance

Considérons une variable aléatoire X pouvant être définie dans \mathbb{C} , et $p(X)$ sa densité de probabilité théorique. Dans la plupart des cas réels, nous n'avons accès qu'à des réalisations de la variable aléatoire X et cherchons à déterminer la loi de probabilité qui la régit. Si dans la pratique, la visualisation directe des différentes réalisations statistiques de la variable X ne permet pas de déterminer sa loi de probabilité, nous pouvons déduire de ses différents *moments* une multitude d'informations. Les moments d'ordres un et deux sont connus respectivement sous les noms de moyenne, et de variance de X .

Variance Les moments statistiques sont définis à partir de la définition générale de l'espérance d'une fonction $f(X)$:

$$\mathbb{E}[f(X)] = \int_{-\infty}^{+\infty} f(X)p(X)dX, \quad (2.1)$$

où $p(X)$ est la densité de probabilité associée à la variable X . Nous pouvons alors considérer les deux premiers moments d'une variable aléatoire, la moyenne μ et la variance v qui s'écrivent respectivement

$$\mu = \mathbb{E}[X] \quad \text{et} \quad v = \mathbb{E}[(X - \mu)^2]. \quad (2.2)$$

Une interprétation de la variance d'une variable aléatoire peut être réalisée à partir de la reformulation suivante

$$v = \mathbb{E}[X^2 + \mu^2 - 2\mu X] \quad (2.3)$$

$$= \mathbb{E}[X^2] + \mu^2 - 2\mu\mathbb{E}[X] \quad (2.4)$$

$$= \mathbb{E}[X^2] - \mu^2. \quad (2.5)$$

Nous voyons, en effet, apparaître dans l'équation 2.3 la différence entre la moyenne des valeurs de X au carré et le carré de la moyenne de X . La variance est donc une mesure de la dispersion de la variable aléatoire X autour de sa valeur moyenne.

Covariance La covariance c_{XY} entre deux variables aléatoires X et Y peut-être interprétée comme une extension de la variance à une variable. Elle informe sur la dispersion conjointe des deux variables, telle que

$$c_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)^*], \quad (2.6)$$

où μ_X et μ_Y désignent respectivement les valeurs moyennes de X et Y , et où $*$ représente la conjugaison complexe. La covariance possède les propriétés suivantes :

- (1) X et Y sont deux variables aléatoires décorréélées si $c_{XY} = 0$
- (2) X et Y sont deux variables aléatoires corrélées partiellement si $0 < c_{XY} < \sqrt{v_X v_Y}$
- (3) X et Y sont deux variables aléatoires corrélées si $c_{XY} = \sqrt{v_X v_Y}$

Matrice de covariance Finalement, le formalisme de covariance entre deux signaux aléatoires s'étend naturellement à celui de matrice de covariance de N variables aléatoires x_n où $n = 1 \dots N$. Dans le cas où $\mathbb{E}[x_n] = 0$ pour tout n , la matrice de covariance est donnée par

$$\mathbf{\Gamma} = \mathbb{E}[\mathbf{xx}^\dagger] \quad (2.7)$$

où † représente la conjugaison transposée. Une description indicielle de la matrice de covariance donne

$$\Gamma_{ij} = \mathbb{E}[x_i x_j^*]. \quad (2.8)$$

Ce dernier formalisme indique que l'élément d'indices (i, j) de la matrice de covariance est la covariance des signaux aléatoires x_i et x_j . Certaines propriétés de la matrice de covariance statistique sont à noter :

- (1) La matrice de covariance est hermitienne par construction, puisque $\Gamma_{ij} = \Gamma_{ji}^*$
En effet, $\Gamma_{ji}^* = \mathbb{E}[x_j x_i^*]^* = \mathbb{E}[x_i x_j] = \Gamma_{ij}$
- (2) Sa diagonale est l'ensemble des variances des différents signaux aléatoires ($\Gamma_{ii} = v_i$)
Ici, $\Gamma_{ii} = \mathbb{E}[x_i x_i^*] = \mathbb{E}[x_i^2] = v_i$
- (3) Elle est semi-définie positive, et donc ses valeurs propres sont positives ou nulles

Covariance et intercorrélation La matrice de covariance est liée à la matrice d'intercorrélation par la transformation de Fourier. Si l'on considère la définition suivante pour l'intercorrélation $R(\tau)$ de deux signaux temporels $x_i(t)$ et $x_j(t)$:

$$R_{ij}(\tau) = \mathbb{E}[x_i(t)x_j(\tau - t)], \quad (2.9)$$

alors la matrice de covariance est l'expression de la matrice d'intercorrélation dans le domaine fréquentiel :

$$C_{ij}(f) = \mathcal{F} R_{ij}(\tau) = \mathcal{F} \mathbb{E}[x_i(t)x_j(\tau - t)] \quad (2.10)$$

$$= \mathbb{E}[\tilde{x}_i(f)\tilde{x}_j^*(f)], \quad (2.11)$$

où $\tilde{x}_i(f)$ représente la transformée de Fourier du signal $x_i(t)$. La dénomination de matrice de covariance est issue du fait que l'équation 2.11 est l'expression de la covariance interspectrale des signaux $\tilde{x}_i(f)$ et $\tilde{x}_j(f)$. Par abus de langage, les matrices de covariance et d'intercorrélation sont parfois toutes deux appelée matrice d'intercorrélation puisqu'elles représentent le même objet mathématique dans deux domaines différents (temporel et fréquentiel). La matrice de covariance est également appelée parfois matrice des densités de puissance interspectrale. Nous préférons le terme de matrice de covariance au sein de ce manuscrit dans un souci de rapprochement avec les terminologies présentes notamment dans le domaine du traitement du signal, et plus largement dans celui des statistiques.

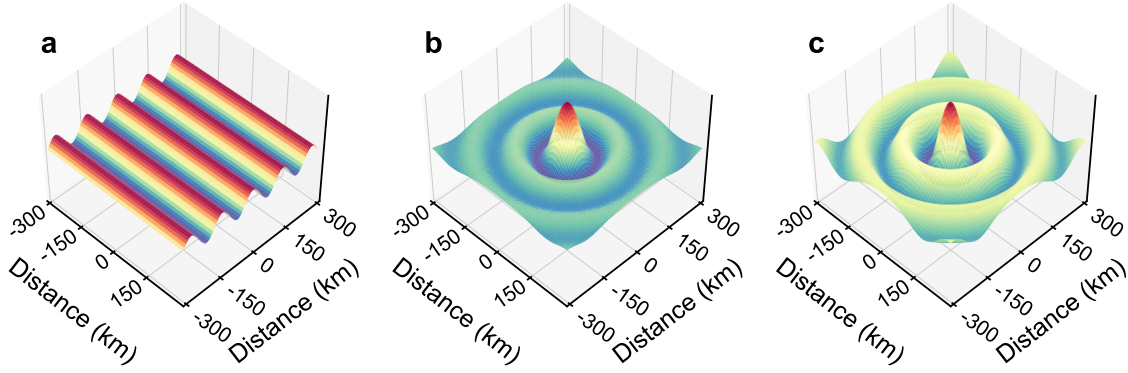


Figure 2.1 – Covariances théoriques obtenues dans un espace de dimensions $300 \times 300 \text{ km}^2$ avec l'origine située au centre de l'espace. La fréquence est fixée à 0.2 Hz et la célérité des ondes dans le milieu à 4 km/s. Les différents cas sont **(a)** une source unique située à une distance infinie de la zone, dans la direction parallèle à l'axe de droite, **(b)** un bruit isotrope à 3 dimensions et **(c)** un bruit isotrope à 2 dimensions (voir équation 2.15)

2.1.2 Propriétés de la matrice de covariance dans le cas sismique

Suivant le modèle général de signal proposé par Gerstoft et al. (2012), nous pouvons considérer les sismogrammes comme étant la superposition de trois contributions indépendantes :

$$\mathbf{x}(f) = \underbrace{\sum_{k=0}^K \alpha_k(f) \mathbf{s}_k(f)}_{K \text{ signaux déterministes}} + \underbrace{\mathbf{n}^p(f)}_{\text{bruit sismique}} + \underbrace{\mathbf{n}^u(f)}_{\text{bruit asismique}}, \quad (2.12)$$

écrites ici dans le domaine fréquentiel. Chacun des termes se décrit comme suit.

Les K signaux déterministes sont souvent considérés comme générés par des sources indépendantes statistiquement (le signal à la source est décorrélé entre chaque source) et localisées ponctuellement dans l'espace (c'est, par exemple, une hypothèse généralement faite sur les séismes). L'amplitude spectrale propre à chaque source d'indice $k = 1 \dots K$ est notée $\alpha_k(f)$ et la phase est contenue dans le vecteur $\mathbf{s}_k(f) = \exp(2i\pi f \gamma \mathbf{e}_k \cdot \mathbf{r})$, où \mathbf{e}_k est la direction de propagation, \mathbf{r} est le vecteur position des stations sismiques et où c est la vitesse de phase apparente des ondes.

Le bruit sismique correspond à du bruit généré par un grand nombre de sources indépendantes, comme le bruit microsismique. La notation p vient de l'anglais *propagative* pour décrire un bruit sismique qui se propage sous forme d'une multitude d'ondes.

Le bruit asismique n'est pas d'origine sismique et ne se propage donc pas d'un capteur à l'autre dans le milieu terrestre. À moins d'être généré par des phénomènes dont l'étendue spatiale avoisine celle du réseau de stations sismiques (pluie, vent, orage, réseau Wi-Fi, ...), il est souvent décorrélé entre les capteurs, d'où la notation u pour *uncorrelated*.

La matrice de covariance du modèle de signal proposé en équation 2.12 donne lieu à la matrice de covariance théorique $\Gamma(f)$ suivante, sous hypothèse d'indépendance statistique de chacun de ses termes entre eux :

$$\Gamma(f) = \Gamma^s(f) + \Gamma^p(f) + \Gamma^u(f), \quad (2.13)$$

Covariance des sources Son expression est facile à dériver analytiquement sous hypothèse de l'indépendance statistique ou du non-recouvrement spatio-temporel des différentes sources entre elles. Elle est donnée par

$$\Gamma^s(f) = \sum_{k=1}^K |\alpha_k(f)|^2 \mathbf{s}_k(f) \mathbf{s}_k^\dagger(f), \quad (2.14)$$

où $|\alpha_k(f)|^2$ est la densité spectrale de chaque source k . La covariance d'un signal unique est illustrée en figure 2.1a, en considérant une onde plane se propageant dans la direction parallèle à l'axe x .

Covariance du bruit sismique Même dans le cas où le bruit sismique est complètement aléatoire, autrement dit que le champ d'onde est constitué d'un grand nombre d'ondes planes décorréliées les unes des autres, il subsiste une certaine corrélation spatiale. Des expressions analytiques simples de la matrice de covariance ont été obtenus dans Eckart (1953), Jacobson (1962) dans des configurations de bruit sphérique (3D) et surfacique (2D). En conséquence, la matrice de covariance $\Gamma^p(f)$ est donnée par

$$\Gamma_{ij}^{3D}(f) = \text{sinc}\left(\frac{2\pi f \|\mathbf{r}_j - \mathbf{r}_i\|}{c}\right), \quad (2.15)$$

dans le cas 3D et où $\text{sinc}(x) = \sin(x)/x$. Dans le cas à deux dimensions,

$$\Gamma_{ij}^{2D}(f) = J_0\left(\frac{2\pi f \|\mathbf{r}_j - \mathbf{r}_i\|}{c}\right), \quad (2.16)$$

où J_0 est la fonction de Bessel d'ordre 0 et de première espèce. Les différentes formes de la matrice de covariance peuvent être visualisées sur la figure 2.1b pour du bruit 3D et sur la figure 2.1c pour du bruit 2D. Il est intéressant de noter que ces deux expressions de la matrice de covariance sont directement reliées à la fonction de Green par l'équation 1.9. En effet, elles sont l'expression de la partie imaginaire de la fonction de Green dans un milieu homogène 3D et 2D respectivement. Ce résultat est démontré, par exemple, dans Sanchez-Sesma (2006), Boschi et al. (2012).

Covariance du bruit asismique Si le bruit n'est pas corrélé d'un capteur à l'autre, alors la première propriété de la covariance indique que $\Gamma_{i \neq j} = 0$, la matrice est donc diagonale. De plus, si le niveau de bruit est le même sur tous les sismomètres, alors

$$\Gamma(f) = |v(f)|^2 \mathbf{I} \quad (2.17)$$

où $|v(f)|^2$ est la densité spectrale de puissance du bruit asismique enregistré sur chacun des capteurs, et où \mathbf{I} est la matrice identité.

2.1.3 Valeurs propres de la matrice de covariance

Nous allons voir que les valeurs propres de la matrice de covariance permettent d'extraire de fortes indications sur la nature du champ d'onde observé. La décomposition en valeurs propres de la matrice de covariance est un aspect fondamental en traitement du signal. Elle est également au cœur de l'analyse multivariée (Bartlett 1947, Lawley 1956) et de l'analyse en composantes principales (Anderson 1963) où le but est de déterminer l'intercorrélation d'un nombre souvent élevé de variables.

La matrice de covariance $\mathbf{\Gamma}$ étant de symétrie hermitienne par définition, elle est toujours diagonalisable. Les valeurs propres λ_i et les vecteurs propres ψ_i sont solutions de l'équation

$$\mathbf{\Gamma}\psi_i = \lambda_i\psi_i. \quad (2.18)$$

Les vecteurs propres forment une base orthonormée telle que

$$\psi_i^\dagger \psi_j = \delta_{ij}, \quad (2.19)$$

où δ_{ij} est le symbole de Kronecker. Comme la matrice de covariance est définie semi-positive, les valeurs propres sont toutes supérieures ou égales à 0. Ainsi, la matrice s'exprime dans sa base propre comme

$$\mathbf{\Gamma} = \sum_i \psi_i \lambda_i \psi_i^\dagger = \mathbf{\Psi}^\dagger \mathbf{\Lambda} \mathbf{\Psi}. \quad (2.20)$$

La matrice $\mathbf{\Lambda}$ est une matrice diagonale où les éléments de la diagonale sont les valeurs propres de la matrice de covariance ($\Lambda_{ij} = \delta_{ij}\lambda_i$).

Valeurs propres d'un bruit parfaitement décorrélé. Si nous considérons des bruits présents dans les enregistrements parfaitement décorrélés d'une station à l'autre, alors la première propriété de la covariance indique que tous les éléments non diagonaux Γ_{ij} sont nuls. La matrice est donc diagonale et les valeurs propres s'expriment naturellement en fonction de la variance des différents enregistrements

$$\lambda_i = v_i^2. \quad (2.21)$$

Ainsi, si nous considérons, par exemple, un bruit électronique présent sur chacun des capteurs d'un réseau de stations sismiques, sous l'hypothèse que ce bruit est d'égale variance v sur chacun des capteurs, alors la matrice de covariance statistique a une unique valeur propre v^2 de multiplicité N .

Valeurs propres dans le cas d'une source unique. Ce cas s'oppose particulièrement au cas précédent, puisqu'il considère une unique source de bruit (ou un signal transitoire, déterministe) qui "illumine" tous les capteurs (il s'agit du cas où $k = 1$ dans l'équation 2.14). Dans ce cas précis, et quelle que soit la forme temporelle de la source, la matrice de covariance s'exprime, à une fréquence f donnée comme

$$\mathbf{\Gamma}(f) = |\alpha(f)|^2 \mathbf{s}(f) \mathbf{s}^\dagger(f), \quad (2.22)$$

où $|\alpha(f)|^2$ représente la densité spectrale de puissance de la source, et où $\mathbf{s}(f)$ est un vecteur normé qui dépend de la direction d'incidence de la source d'arrivée et la phase du signal émis par la source. $\mathbf{\Gamma}(f)$ est générée par un unique vecteur, elle donc de rang 1 et ne présente qu'une seule valeur propre non nulle.

Valeurs propres dans le cas d'un nombre quelconque de sources décorrélées. Dans le cas où le nombre de sources décorrélées K est inférieur au nombre de stations sismiques N , il est facile de montrer que le rang de $\mathbf{\Gamma}(f)$, c'est-à-dire le nombre de valeurs propres non nulles est au maximum égal à K . Dans le cas où $K > N$ alors le rang de la matrice est maximum et est égal à N .

2.1.4 Espace signal et espace bruit

En pratique, une matrice de covariance idéale enregistrée expérimentalement est de rang plein (toutes les valeurs propres sont supérieures à 0) car les données sont toujours polluées par un minimum de bruit asismique. Lorsque le rapport signal sur bruit est suffisant, il est commun lors de la décomposition en valeurs propres de considérer les L valeurs propres les plus fortes comme correspondant à du signal et toutes les autres à du bruit. Ainsi, les L vecteurs propres associés aux L valeurs propres forment un sous-espace signal. Les autres vecteurs propres forment quant à eux le sous-espace bruit. Le sous-espace signal contient la partie des signaux sur lequel se projette l'information recherchée. L'espace bruit quant à lui va limiter l'exploitation des données.

Il est à noter que selon l'objectif à atteindre, le bruit sismique peut faire partie de l'espace bruit pour chercher, par exemple, une source sismique "déterministe" ou à l'espace signal dans le cadre, notamment, des corrélations de bruit. La difficulté de cette décomposition repose surtout dans le choix du rang L de l'espace signal. Néanmoins, nous verrons dans le dernier chapitre de la thèse une méthode d'estimation de ce rang fondée sur la décomposition en harmonique sphérique du champ d'onde.

2.2 Estimation des propriétés spectrales de la matrice de covariance par formation de voie

La matrice de covariance contient des informations relatives à la fréquence et à la direction d'arrivée des ondes qui composent le champ d'onde sismique étudié. Une méthode classique de l'estimation des directions d'arrivée des ondes peut être obtenue par décomposition de la matrice de covariance sur la base des ondes planes. Cette technique d'analyse est connue sous le nom de formation de voie, dérivé de l'anglais *beamforming*.

Si la technique est classiquement utilisée sur des antennes hertziennes ou dans des antennes acoustiques où la célérité des ondes est constante ou considérée comme tel ([Veen and Buckley 1988](#)), elle ne peut pas être appliquée directement dans le cas où les ondes ont des célérités variables, comme c'est le cas sur Terre. La recherche des ondes s'effectue donc dans l'espace des lenteurs $\boldsymbol{\gamma} = \gamma_x \mathbf{e}_x + \gamma_y \mathbf{e}_y$, où la lenteur se définit comme l'inverse de la vitesse de phase $\|\boldsymbol{\gamma}\| = 1/c$, (voir par exemple [Rost 2002](#); pour une revue sur les méthodes de traitement d'antenne en sismologie).

Si l'on considère alors l'onde plane $\mathbf{b}(f, \boldsymbol{\gamma}) = e^{i2\pi f \boldsymbol{\gamma} \mathbf{r}}$ théoriquement enregistrée sur le réseau de stations sismiques (où \mathbf{b} est de dimension $1 \times N$), le résultat de la formation de voie s'exprime comme

$$B(f, \boldsymbol{\gamma}) = \mathbf{b}(f, \boldsymbol{\gamma}) \mathbf{C}(f) \mathbf{b}^\dagger(f, \boldsymbol{\gamma}), \quad (2.23)$$

ce qui est l'expression de la décomposition en ondes planes de la matrice de covariance. Par exemple lorsqu'une onde plane de lenteur γ_0 se propage à travers le réseau de sismomètres, $B(f, \boldsymbol{\gamma})$ est maximum pour $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$.

Il est important de noter que la lenteur mesurée par formation de voie avec un réseau disposé à la surface du Globe correspond à la lenteur *apparente* des ondes. Celle-ci est en réalité la projection du vecteur lenteur total sur le plan du réseau, si nous faisons l'approximation locale d'un réseau plan et d'une géométrie cartésienne. Ainsi, une onde de volume qui arrive avec une incidence faible en comparaison à la verticale est susceptible de produire une lenteur apparente très proche de 0, car la projection dans le plan (x, y) du réseau est presque nulle.

2.3 Matrice de covariance estimée depuis des données réelles

Dans le cas où nous aurions accès à un nombre infini d'enregistrements dans le temps, nous pourrions parfaitement estimer la matrice de covariance. Cependant, en sismologie, la matrice de covariance ne peut être estimée que grâce à un nombre réduit M d'enregistrements. En effet, la principale limitation provient de la durée caractéristique des phénomènes que l'on cherche à mettre en lumière (par exemple, une éruption). Ainsi, la moyenne (ou stacking) ne peut-être effectuée que sur une durée inférieure à cette durée caractéristique. Nous verrons au sein de cette section les implications d'une telle limitation.

2.3.1 Estimateur de la matrice de covariance

Nous pouvons définir le vecteur des données d'un réseau sismique comme il suit

$$\mathbf{u}(f, t) = (u_1(f, t), u_2(f, t), \dots, u_N(f, t))^T, \quad (2.24)$$

où $u_i(f, t)$ est la transformée de Fourier du signal recueilli à la station i sur une fenêtre de signal comprise entre les temps t et $t + \delta t$. Une estimation de la matrice de covariance peut être alors formulée à partir des M fenêtres de signal comme

$$\mathbf{C}(f, t) = \sum_{m=0}^{M-1} \mathbf{u}(f, t + m\delta t/2) \mathbf{u}^\dagger(f, t + m\delta t/2), \quad (2.25)$$

où le facteur $1/2$ fait apparaître un recouvrement des fenêtres temporelles de 50%. Les étapes du calcul de la matrice de covariance depuis des données réelles sont détaillées sur la figure 2.2.

Même s'il est connu que la matrice de covariance estimée approche la matrice de covariance idéale sous certaines conditions (qui concernent souvent le nombre de fenêtres à prendre en compte et nombre de stations sismiques), la matrice de covariance estimée (ainsi que ses valeurs propres) reste toujours différente.

2.3.2 Rang de l'estimateur

Comme nous l'avons dit précédemment, le nombre de valeurs propres non nulles (ou autrement dit le rang) de la matrice de covariance n'est strictement parlant jamais directement lié au nombre de sources, car il existe toujours, aussi faible soit-il, un bruit électronique décorréolé d'une voie à l'autre. Cependant, contrairement la matrice de covariance idéale, le rang de la matrice de covariance donnée par l'équation 2.25 peut-être inférieur au nombre de voies. En effet, il suffit que le nombre d'acquisitions à partir duquel \mathbf{C} est évalué soit inférieur au nombre de voies du réseau. Ce cas de figure est de plus en plus fréquent avec l'accroissement du nombre de sismomètres dans les grands réseaux (comme l'USArray avec près de 400 sismomètres transportables, ou encore le réseau déployé à Long Beach en Californie présenté en figure 4 qui comprend 2500 stations sismiques). Ainsi le rang R est limité par le nombre d'acquisitions M , tel que :

$$R = \text{Min}(M, N). \quad (2.26)$$

2.3.3 Revue des méthodes de séparation d'espaces signal et bruit

Un bref retour historique est ici nécessaire pour comprendre les méthodes qui existent à l'heure actuelle pour décrire les valeurs propres de la matrice de covariance estimée. Dans

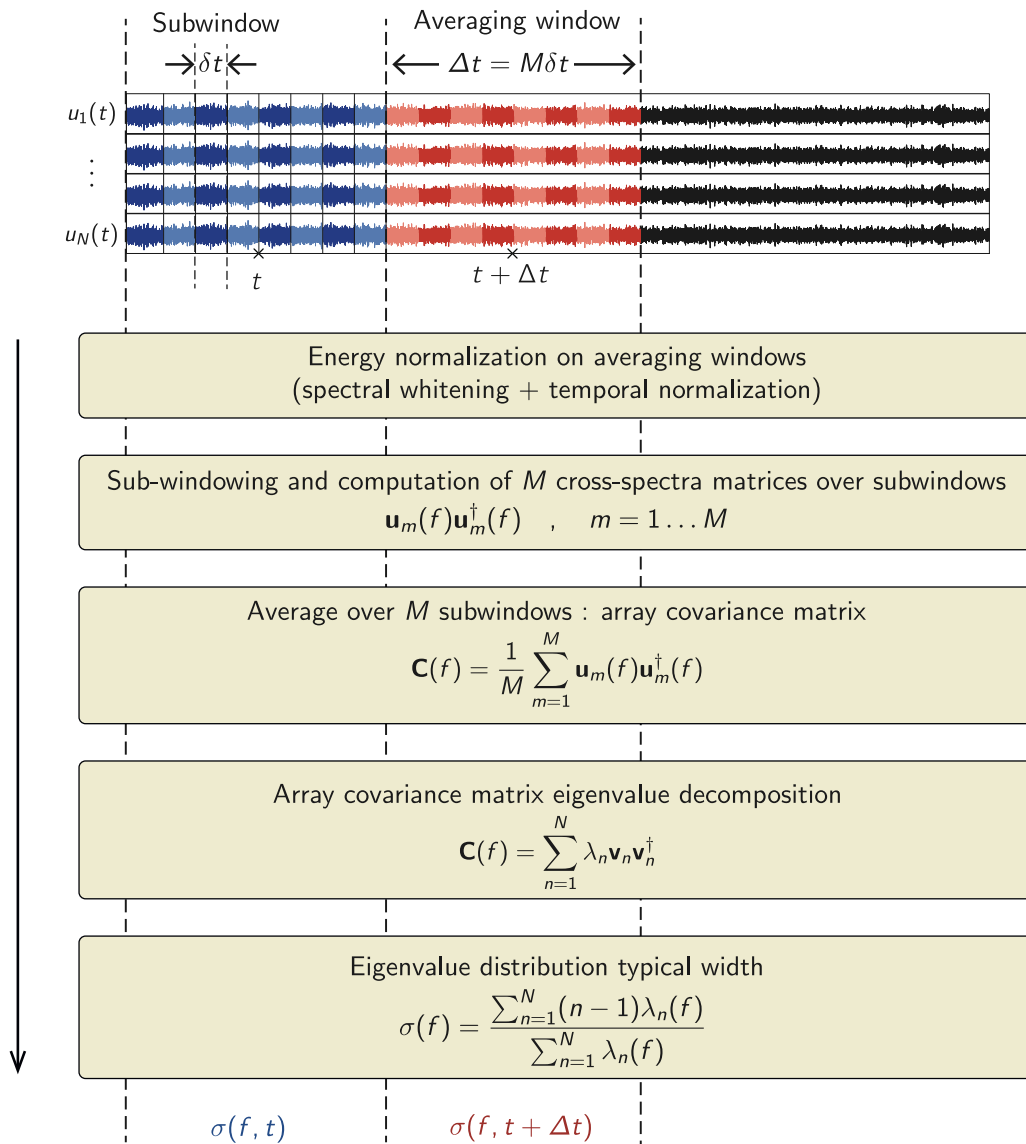


Figure 2.2 – Estimation de la matrice de covariance depuis des données réelles.

la plupart des cas, des méthodes sont proposées pour tester l'indépendance de grandeurs aléatoires mises sous forme de matrice de covariance, ce qui revient dans notre cas à tester, par exemple, la présence de K signaux dans le champ d'onde, où le nombre de signaux est généralement supposé plus faible que le nombre de stations sismiques N . L'hypothèse de la présence d'un bruit partiellement corrélé comme le bruit sismique n'arrive que tardivement, il est d'abord question de tester l'indépendance des grandeurs dans un contexte de bruit décorréolé, où les valeurs propres théoriques relatives à ce bruit sont censées être égales. En d'autres termes, il est question de chercher à séparer les sous-espaces signal et bruit.

Dans le domaine de la statistique, dans l'étude conjointe de plusieurs variables aléatoires (ou analyse multivariée), [Bartlett \(1954\)](#) et [Lawley \(1956\)](#) développent un test d'importance entre les valeurs propres de la matrice de covariance idéale et la matrice de covariance estimée. Ce test est utile pour évaluer l'indépendance globale des variables d'un vecteur aléatoire et est connu sous le nom de test de sphéricité. Ils utilisent un test χ^2 réalisé sur les valeurs propres dites « résiduelles » de la matrice de covariance estimée, qui sont les valeurs propres de rang élevé, supposées relatives au bruit et théoriquement égales. [Bartlett \(1954\)](#) et [Lawley \(1956\)](#) proposent alors d'effectuer des séries de tests qui permettent de déterminer la multiplicité des différentes valeurs propres. Si une seule valeur propre de multiplicité N est trouvée par le test, alors les vecteurs sont indépendants, sinon il existe une quantité corrélée dans les vecteurs, comme un signal cohérent. L'inconvénient de ce test réside dans le choix que l'utilisateur doit faire dans l'établissement de seuils. Ces travaux sont repris plus tard dans le cadre de l'analyse en composantes principales dans [Anderson \(1963\)](#).

Plus tard, dans le domaine du traitement du signal spécifique aux antennes, [Bienvenu and Kopp \(1980\)](#) proposent une méthode dite adaptative, car elle permet de chercher la base dans laquelle la matrice de covariance admet une valeur propre de grande multiplicité, c'est-à-dire une base dans laquelle le bruit seul donne à nouveau une matrice identité avec N valeurs propres égales. Cette méthode est adaptative au sens où elle s'adapte au bruit ambiant contextuel, en utilisant un modèle de bruit donné. Dans leur approche, les auteurs montrent également que des informations cruciales peuvent être tirées du sous-espace signal, en ce qui concerne la direction d'arrivée des ondes cohérentes par exemple. Dans la continuité des méthodes hautes résolutions d'estimation spectrales des signaux reçus sur un réseau de capteurs ont vu le jour, comme [Wax et al. \(1984\)](#) et [Schmidt \(1986\)](#).

Un point qui reste cependant difficile à résoudre est l'estimation du rang du sous-espace signal. [Wax and Kailath \(1985\)](#) proposent alors de formaliser la matrice de covariance observée avec un modèle de signal semblable à celui proposé dans la section précédente, et d'effectuer une sélection du meilleur modèle à l'aide du critère d'information théorique d'Akaike qui ne requiert aucun jugement subjectif, contrairement à [Bartlett \(1954\)](#) et [Lawley \(1956\)](#).

Les premières applications de ces méthodes haute résolution en sismologie sont présentées dans [Capon \(1974; 1969\)](#) pour l'analyse de la structure de la Terre et dans [Ringdal and Husebye \(1982\)](#) pour l'analyse d'explosions et de séismes à partir de réseau de stations. En sismologie, [Goldstein and Archuleta \(1987\)](#) utilisent le critère défini dans [Wax and Kailath \(1985\)](#), et obtiennent une estimation de l'espace signal en déterminant l'espace orthogonal à l'espace bruit. Cette approche donne de meilleurs résultats, car elle est mieux contrainte par le nombre généralement plus important de vecteurs propres liés à ce sous-espace. Cette technique est connue sous le nom de MUSIC (pour *MULTiple Signal Correlators*) et avait été préalablement présentée dans [Schmidt \(1986\)](#).

Les approches citées ci-dessus ne rendent cependant pas compte de l'effet de l'estimation de la matrice de covariance à partir d'un nombre fini de fenêtres de signal. Plus récemment, [Nadakuditi and Edelman \(2008\)](#) reformulent le problème de détecter des signaux dans du bruit avec un réseau dans la limite d'un faible nombre d'échantillons. Ils définissent un cadre mathématique dans lequel il est possible de faire de la détection de signaux à partir des valeurs propres estimées et le nombre maximal de signaux identifiables dans la matrice de covariance avec des méthodes comme celle présentée dans [Wax and Kailath \(1985\)](#).

Si l'objectif principal de ces méthodes est de déterminer un nombre de signaux noyés dans du bruit, elles ne permettent pas de rendre compte du comportement des valeurs propres de la matrice de covariance estimée, ce qui nous intéresse particulièrement ici. Ce fût l'objet de l'étude menée dans [Menon et al. \(2012a\)](#) qui proposent de déterminer des descriptions analytiques des valeurs propres de la matrice de covariance estimée et statistique dans le cas asymptotique où le nombre de capteurs N et le nombre d'acquisitions M vérifient tous deux $N, M \rightarrow \infty$ tout en conservant un rapport constant $N/M = c$. Les auteurs s'intéressent particulièrement à évaluer ces valeurs propres non pas pour détecter du signal, mais pour étudier du bruit uniquement (propagatif et asismique), afin d'interpréter des changements environnementaux dans le bruit. En utilisant les principes de la théorie des matrices aléatoires ([Marčenko and Pastur 1967](#), [Sengupta and Mitra 1999](#), [Mestre 2008](#)), des expressions de densité de probabilité des valeurs propres de la matrice de covariance pour un bruit propagatif et asismique sont déduits. La théorie des matrices aléatoires a également trouvé des applications intéressantes dans le domaine de l'acoustique ([Aubry and Derode 2009](#)), en océanographie ([Gerstoft et al. 2012](#), [Menon et al. 2012b](#)) et en télécommunications ([Muller 2002](#), [Tulino and Verdú 2004](#)).

2.4 Largeur spectrale de la matrice de covariance

De l'état de l'art ci-dessus, nous comprenons que déterminer un nombre exact de signaux présents dans le bruit est un véritable challenge. Au cours de cette thèse, nous n'avons pas souhaité, si tant est possible, de calculer la dimension exacte de l'espace signal. Nous avons plutôt souhaité déterminer sa dimension approximative en fonction du temps et de la fréquence. S'inscrivant dans cette démarche, [Wagner and Owens \(1996\)](#) ont déjà proposé d'analyser l'énergie portée par la première valeur propre de la matrice de covariance. Ce résultat peut être utilisé comme détecteur et a été appliqué par [Wagner and Owens \(1996\)](#) sur des données sismiques. C'est une approche intéressante qui permet de détecter un signal unique dans un bruit de variance relativement forte. Cependant, ce qui critère ne permet pas de distinguer du bruit de type propagatif et du bruit non propagatif.

Notre méthode se rapproche plus d'une méthode un peu plus ancienne, comme la mesure de la polarisation des ondes depuis les valeurs propres de la matrice de covariance des données recueillies par un réseau de stations présenté dans [Samson and Olson \(1980\)](#). Dans cet article, les auteurs proposent de comparer la dispersion des valeurs propres à une distribution de référence.

Nous proposons ici d'utiliser la largeur spectrale de la matrice de covariance pour tester le degré de cohérence de ses entrées. Cette largeur spectrale $\sigma(f, t)$ s'exprime en fonction des valeurs propres λ_i comme il suit

$$\sigma(f, t) = \frac{\sum_{i=1}^N (i-1) \lambda_i(f, t)}{\sum_{i=1}^N \lambda_i(f, t)}. \quad (2.27)$$

Le comportement de la largeur spectrale de la matrice de covariance est une informa-

tion qui dépend de plusieurs paramètres, à l’instar des valeurs propres de la matrice de covariance. Parmi ces paramètres, il y a la fréquence f , le nombre de capteurs N ainsi que le nombre de fenêtres M (d’acquisitions) prises en compte dans l’estimation de la matrice. Il est intéressant de noter que même si les valeurs propres associées au bruit asismique ne sont plus égales, mais sont déviées de leurs positions théoriques. La largeur spectrale de la matrice de covariance dans les deux cas reste à peu près égale à $R/2$ et relativement indépendante de la fréquence. Cette propriété intéressante permet de visualiser du bruit asismique facilement.

Le bruit propagatif quant à lui, a une expression théorique qui dépend de la fréquence, comme le montre l’équation 2.15. Les largeurs spectrales qui correspondent à ce bruit sont à déterminer numériquement, car elles dépendent de la géométrie du réseau (densité, ouverture) et de la fréquence. De façon qualitative, nous pouvons tout de même distinguer deux régimes pour la largeur spectrale σ dans ce bruit. Le premier régime correspond au cas où la longueur d’onde du bruit est très grande en comparaison avec l’ouverture du réseau. Dans ce cas, la matrice de covariance est quasiment de rang 1 et $\sigma \rightarrow 0$. Lorsque la fréquence augmente, la largeur spectrale augmente également jusqu’à atteindre une valeur maximale, $\sigma_{max}(f)$ qui peut être déterminée numériquement (voir chapitre 3) ou analytiquement (via une théorie présentée en chapitre 4). Si l’objet de l’étude est de différencier le bruit propagatif du bruit non propagatif, il est donc nécessaire d’utiliser des paramètres adéquats (nombre de fenêtres pour estimer la matrice de covariance et bandes de fréquences adaptées).

Enfin, les signaux “cohérents” contribuent à diminuer la valeur de σ , car ils se concentrent sur les premières valeurs propres. Il est donc possible de détecter les zones temporelles et fréquentielles où ces signaux sont présents à travers cette méthode.

2.5 Calcul de la matrice de covariance depuis des données réelles

Dans cette section, l’implémentation concrète de la méthode sur un jeu de données sismiques est présentée en détail. Deux grandes étapes sont à retenir : d’abord les données sont rapatriées depuis les plateformes d’échange des données sismiques et sont préparées pour l’analyse (phase d’homogénéisation des données...). Ensuite, les données sont analysées à travers un algorithme qui a été développé au cours de la thèse.

2.5.1 Données sismiques

Formats des données

En décembre 1987, après d’intensives discussions sur l’adoption d’un format standard d’échange de données, la communauté sismologique se mettait d’accord pour adopter le format d’échange SEED (pour *Standard for the Exchange of Earthquake Data*, IRIS 1993). Cette décision facilite grandement de nos jours le transfert des données dans un format adapté par un grand nombre de sismologues. Les jeux de données sont souvent en accès libre et disponibles sur requêtes dans des plateformes dédiées via le *File Transfer Protocol*, comme c’est le cas sur la plateforme de l’IRIS (voir la première étape de la figure 2.3). Le jeu de données de l’USArray est par exemple, disponible intégralement sur cette dernière.

Ce format est destiné à agréger les données sismiques recueillies par les sismomètres à l’aide de fichiers de données (fichiers *miniSEED*) et de métadonnées (fichiers *dataless*), qui fournissent toutes les informations nécessaires aux utilisateurs désireux d’exploiter

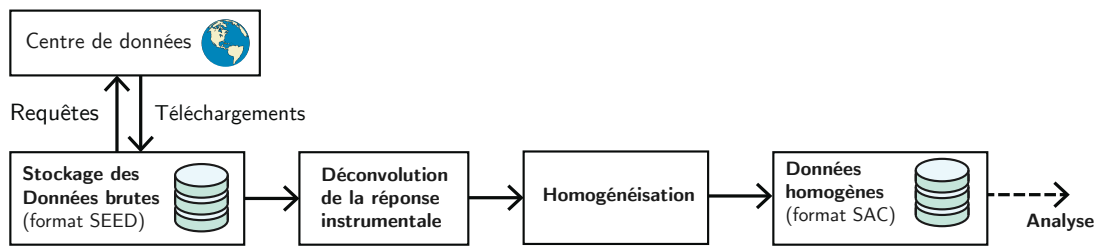


Figure 2.3 – Étapes de préparation des données, de leur récupération sur le centre des données à leur analyse.

ces données. Les métadonnées contiennent des informations relatives à l'instrument qui mesure les vibrations du sol, comme la fréquence d'échantillonnage, la position géographique, le temps, la réponse instrumentale, etc. Les données sont les enregistrements "bruts" directement issus du convertisseur analogique/numérique en sortie du sismomètre. La conservation des données brutes séparément des métadonnées garantit une forme de sécurité, car elle laisse le soin à l'opérateur d'effectuer lui-même les filtrages adéquats. La position spatiale est souvent donnée en coordonnées géographiques classiques (longitude, latitude et altitude). Elle est d'une importance capitale pour pouvoir analyser les données comme nous le verrons plus tard. La fréquence d'échantillonnage est la fréquence à laquelle sont échantillonnés les signaux sismiques.

La réponse instrumentale

Un sismomètre est un instrument qui sert à mesurer une grandeur physique comme le déplacement du sol, sa vitesse ou son accélération. Le processus qui permet de passer de cette grandeur physique à la forme d'onde numérique peut être succinctement décrit comme il suit :

1. La grandeur physique est mesurée à l'aide d'un système électromécanique. Dans le cas des vélocimètres, il s'agit généralement d'un pendule maintenu en situation d'équilibre par une force rétroactive. Lorsqu'une vibration a lieu, l'amplitude de la force varie pour maintenir en situation d'équilibre le pendule et donne ainsi une information sur la vibration. La mesure de la force est électronique, cette première étape constitue donc une conversion de la grandeur physique en grandeur électrique.
2. La donnée de la force rétroactive est digitalisée à l'aide d'un convertisseur analogique/numérique pour permettre sa mise en mémoire (typiquement une carte mémoire ou un disque dur). La donnée finale digitale est exprimée en entiers signés (des *counts*)

Il est donc nécessaire de connaître la réponse instrumentale (ou sensibilité) qui permet de retrouver l'unité de la grandeur mesurée. La réponse instrumentale totale s'exprime, par exemple, en Counts/(m/s) dans le cas d'un vélocimètre et varie en fonction de la fréquence. Elle est intégrée au fichier de métadonnées sous forme de coefficients de filtres fréquentiels. La correction de la réponse instrumentale s'effectue le plus souvent dans le domaine fréquentiel, en divisant le spectre du sismogramme par la réponse instrumentale, ce qui correspond à une opération de déconvolution dans le domaine temporel (voir la seconde étape de la figure 2.3).

Homogénéisation des données

Les données au format SEED sont regroupées sous forme de blocs. Chaque bloc contient des échantillons enregistrés entre un recalage de l'horloge interne du sismomètre avec le temps de référence accessible par transmission GPS. La durée des blocs varie en fonction de la fréquence de synchronisation de la station. Aussi est-il parfois question d'agréger les blocs, et de correctement vérifier que les données sont bien mesurées sur le temps de référence. Cette opération consiste également à vérifier que des données ne sont pas manquantes, ce qui peut être le cas parfois avec des réseaux locaux, qui utilisent des réseaux Wi-Fi pour transmettre les données. Si des trous dans les données sont présents, nous insérons à la place généralement des zéros.

Le but de cette opération est d'obtenir un jeu de données homogène (voir la troisième étape de la figure 2.3), c'est-à-dire des données sismiques recueillies sur plusieurs stations sismiques et échantillonnées à des instants identiques. Une fois les données homogénéisées, les fichiers sont stockés sur un disque dur pour permettre leur analyse via un ordinateur. Les fichiers sont habituellement découpés sur des durées de 24 heures. Un jeu de données peut occuper énormément de place selon la fréquence d'échantillonnage, le nombre de stations sismiques et la durée totale des enregistrements. Il est souvent utile de les stocker sur des disques durs à accès rapide, comme c'est le cas du système de fichier parallèle GPFS (pour *General Parallel File System*).

Les données sont également converties au format SAC (*Seismic Analysis Code*), un outil très pratique souvent utilisé pour visualiser les données rapidement, et effectuer des analyses simples comme l'analyse du spectre de puissance ou le filtrage des données. Les fichiers SAC réunissent les données et les métadonnées (voir la dernière étape de la figure 2.3).

2.5.2 Calcul de la matrice de covariance

Une fois les données préparées comme indiqué précédemment, nous pouvons entamer l'estimation de la matrice de covariance. Le code d'estimation de cette matrice a été initialement développé en Matlab, et est pleinement à même de calculer la matrice de covariance de réseaux dont la taille n'excède pas trente capteurs. Les résultats obtenus sur le jeu de données du Piton de la Fournaise présentés dans le chapitre suivant ont été calculés sur un ordinateur portable bicœur cadencé à 2.9 GHz sur une année entière en quelques heures. Dans un souci de partage du code, la méthode a été entièrement réécrite en Python 2 pour la gratuité du langage, mais également pour accélérer les calculs en passant une partie du code en C++, l'interface C++/Python supportant mieux différents systèmes d'exploitation que l'interface C++/Matlab.

Dès lors que les données sont issues d'un grand réseau comme c'est le cas de l'USArray (dans notre cas, nous avons sélectionné 121 capteurs), il est nécessaire d'envisager un algorithme plus performant et une exécution par un calculateur adapté. Les calculateurs sont des ordinateurs qui possèdent une architecture particulièrement adaptée à l'exploitation de données massives, et aux calculs intensifs grâce à un grand nombre de nœuds eux-mêmes comprenant un grand nombre de cœurs. Dans notre cas, les calculs ont été effectués dans le langage C++, sur le calculateur du Service de Calcul Parallèle et de Traitement de Données en sciences de la Terre (S-CAPAD, <http://webpublix.ipgp.fr/rech/scp/index.php>). La grille de calculs nœuds/cœurs permet la parallélisation du code sur deux niveaux, ce qui permet d'accélérer nettement le calcul de la matrice de covariance, comme indiqué sur l'organigramme en figure 2.4 qui présente les étapes prin-

cipales du code de calcul de la matrice de covariance.

Le code a été établi dans l'idée de fonctionner avec différents jeux de données. La première étape du code consiste à tester l'homogénéité des données, de façon élémentaire puisqu'il observe simplement si chaque trace contient le même nombre de points.

Les données sont ensuite mises en mémoire tampon, un niveau de mémoire d'accès encore plus rapide que celui du disque dur sur lequel sont stockées les données, car située au niveau même des nœuds. Le calcul est ensuite réparti sur les différents nœuds via la norme MPI, de façon à calculer la matrice de covariance à chaque fréquence de façon parallèle. Les données sont traitées à cette occasion avec les prétraitements classiques d'égalisation temporelle et spectrale.

Le calcul est ensuite distribué à nouveau au sein de chacun des nœuds sur les cœurs, en utilisant OpenMP. Cette dernière interface de programmation est automatique et permet d'accélérer les boucles imbriquées en utilisant les différents cœurs de façon simultanée. Ainsi, chaque matrice interspectrale est calculée indépendamment des autres sur chacun des cœurs simultanément. La lecture des données s'effectue dans la mémoire tampon depuis chaque cœur. Les matrices sont ensuite sommées pour obtenir la matrice de covariance à une fréquence donnée sur chaque nœud et l'opération est répétée pour compléter le calcul sur la bande de fréquence totale.

Un nœud désigné comme maître se charge finalement de regrouper les matrices de covariance calculées sur chaque fréquence dans la fenêtre globale considérée, et les trie en fonction de la fréquence, ce qui permet au résultat d'être intelligible par l'utilisateur. Le résultat est écrit dans un fichier binaire, qui minimise la taille du résultat. Deux fichiers sont écrits sur le disque à cette étape : un fichier qui contient la matrice de covariance et un second qui contient la largeur spectrale de la matrice. Les fichiers sont stockés sur chaque jour.

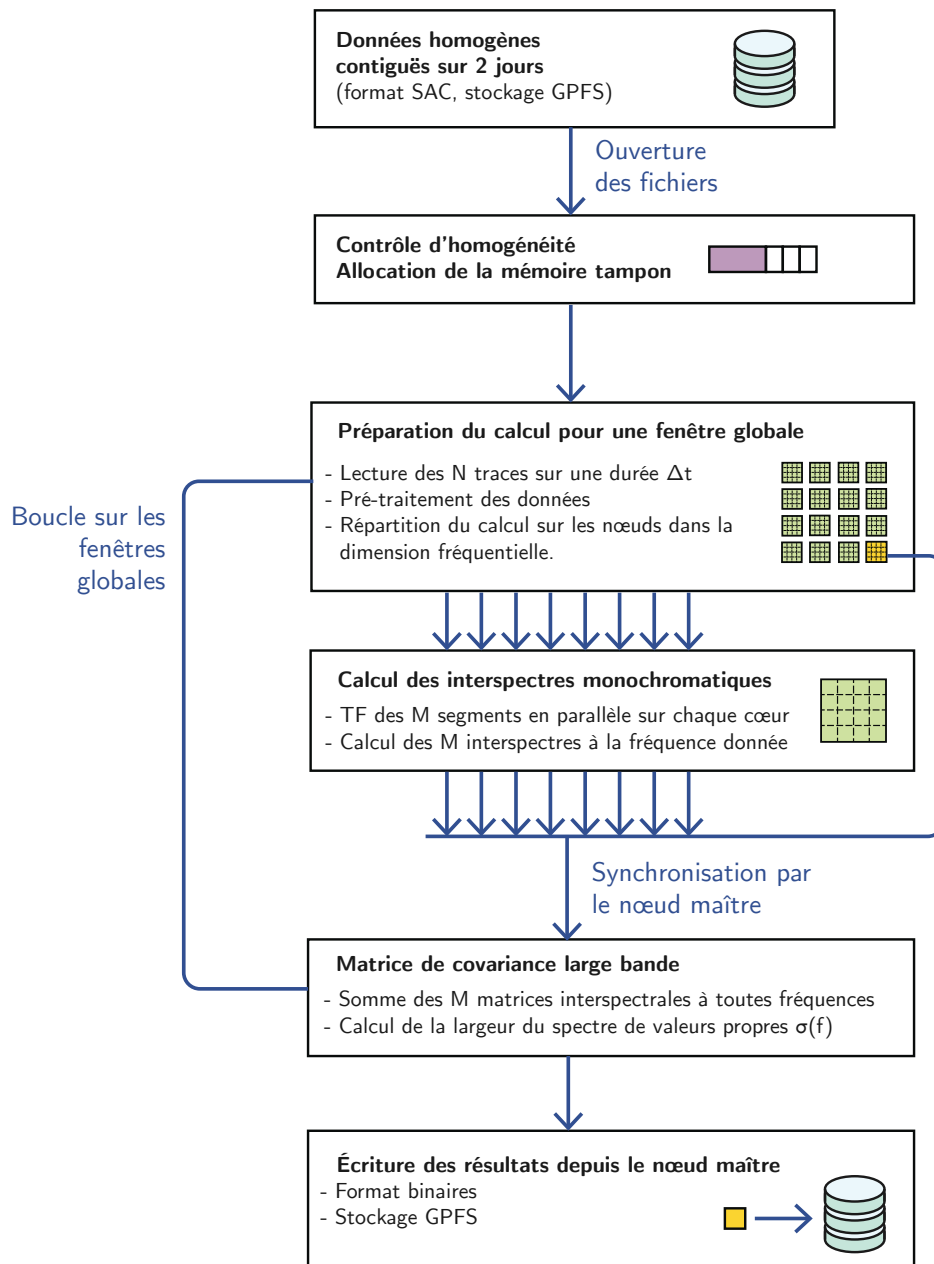


Figure 2.4 – Schéma fonctionnel du calcul de la matrice de covariance depuis des données réelles Les calculs sont effectués dans le langage C++, sur le calculateur du Service de Calcul parallèle et de Traitement de Données en sciences de la Terre (S-CAPAD, <http://webpublix.ipgp.fr/rech/scp/index.php>). Le calcul est distribué sur les nœuds du cluster via un premier niveau de parallélisation MPI (pour *Message Passing Interface*) dans la dimension fréquentielle. La distribution du calcul des interspectres sur les cœurs s’effectue grâce à un second niveau de parallélisation via OpenMP (pour *Open Multi-Processing*). Les données sont lues et écrites depuis un système de fichier parallèle GPFS (pour *General Parallel File System*) qui permet un accès très rapide aux données. Le calcul est orchestré par un nœud désigné comme “maître”, ce qui permet la synchronisation du calcul en parallèle et l’écriture des résultats de façon séquentielle.

Chapitre 3

Analyse de la matrice de covariance des données du Piton de la Fournaise et de l'USArray

3.1	Analyse des données du Piton de la Fournaise	54
3.1.1	Piton de la Fournaise seismic network	54
3.1.2	Synthetic tests	55
3.1.3	Analysis of the Undervolc data.....	58
3.1.4	Discussion and conclusions.....	68
3.2	Application to the transportable component of the USArray in 2010	70
3.2.1	Selection of the parameters	70
3.2.2	Results	71
3.2.3	Observation of the teleseismic activity.....	71
3.2.4	Microseismic noise: surface waves generated by swells.....	74
3.2.5	Microseismic noise: teleseismic P-waves.....	74
3.2.6	26-second microseism	76
3.2.7	Conclusions	76

Dans ce chapitre, nous explorons la méthode sur deux jeux de données issus de réseaux différents.

Le premier jeu de données provient du réseau de surveillance du Piton de la Fournaise situé sur l'île de la Réunion, mis en place par [Brenguier et al. \(2012\)](#) dans un rayon d'environ 10 km. Ces données ont été analysées pendant l'année 2010 entière dans des bandes de fréquence allant de ~ 0.01 à ~ 10 Hz. Cette première application a permis de valider la pertinence de la méthode dans deux gammes de fréquences. En basse fréquence (sous 0.1 Hz), nous montrons que la méthode détecte des événements d'origine téléseismique. Nous comparons nos résultats de détection avec la base de données de l'U.S. Geological Survey et montrons que 75% du catalogue mondial de séismes est détecté grâce à notre méthode. À plus haute fréquence, nous analysons l'activité sismo-tectonique du volcan du Piton de la Fournaise. Nous détectons clairement les trois trémors volcaniques qui ont eu lieu, ainsi que la crise sismique qui précédait la plus grande des trois. Nos résultats sont comparés avec les observations de [Roult et al. \(2012\)](#). En particulier, ils montrent qu'il est possible d'utiliser la largeur spectrale de la matrice de covariance pour détecter des signaux émergents, comme c'est le cas des trémors volcaniques, ce que des méthodes classiques de détection comme le STA/LTA ([Allen 1982](#)), ou le Kurtosis ([Baillard et al. 2013](#)) peinent à détecter.

Le deuxième jeu de données a été enregistré par le réseau américain USArray. Les données ont été enregistrées pendant l'année 2010 par 121 sismomètres répartis sur une zone d'environ 1500×500 km². Dans cette seconde application, dans un premier temps, nous nous intéressons à quantifier l'efficacité des méthodes d'égalisation spectrale et temporelle pour compenser l'inhomogénéité de la distribution des sources de bruit. Nous montrons alors que ces méthodes sont particulièrement efficaces pour soustraire l'influence de sources de forte amplitude situées à proximité de certaines stations, mais échouent à réduire l'influence des sources les plus "cohérentes" comme les événements téléseismiques ou certains événements microsismiques de forte amplitude.

Dans un second temps, cette observation nous a naturellement amenés à utiliser la largeur spectrale comme un outil d'analyse robuste. Comme pour le Piton de la Fournaise, l'activité téléseismique est observée à basse fréquence (< 0.1 Hz) et contribue toujours à réduire le rang de la matrice de covariance. À plus haute fréquence, les sources d'origine microsismique et des ondes de volumes probablement générées en océans profonds sont observées. Une source persistante monochromatique et quasi continue est également mise en avant, et est comparée à des observations similaires dans la littérature.

Les deux sections de ce chapitre sont extraites de deux articles. Le premier est accepté dans *Geophysical Journal International* ([Seydoux et al. 2016a](#)) et le second est actuellement en révision dans *Geophysical Research Letters*.

3.1 Analyse des données du Piton de la Fournaise

3.1.1 Piton de la Fournaise seismic network

We applied our method to the data from the UnderVolc experiment ([Brenguier et al. 2012](#)) when an array of 21 broad-band seismographs was installed around the Piton de la Fournaise volcano on La Réunion island as shown in Fig. 3.1. The network consists of 15 Gurlap 30-sec CMG-40T broad-band three-component seismometers and of 6 additional Kinemetrics Q330 associated with Gurlap 3ESPC three-components sensors. This array recorded many teleseismic earthquakes as well as seismo-volcanic events associated with the volcanic activity. This study focuses on the vertical-component records. An extension

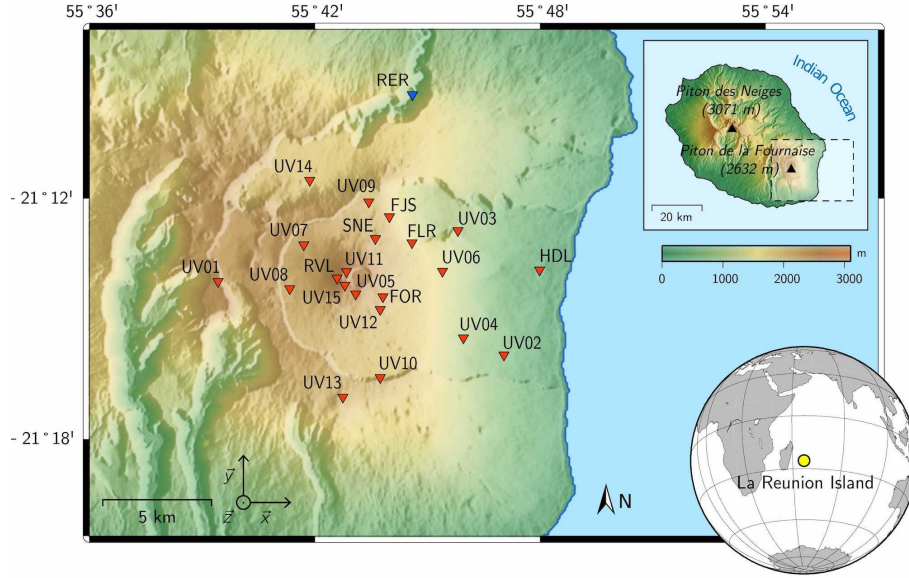


Figure 3.1 – Map of the Piton de la Fournaise volcano. The top right inset shows the location of the Piton de la Fournaise volcano on La Réunion Island, and the bottom right inset shows the location of La Réunion island in the Indian Ocean. The locations of UnderVolc stations are shown with red triangles, and the location of the RER seismic station of the Geoscope seismic network is shown in blue triangle.

of our method to a full analysis of three-component seismograms is under way. We analyze seismic data recorded during 2010 (from 1st of January to 31st of December) when all 21 stations were recording simultaneously.

3.1.2 Synthetic tests

We performed synthetic tests to characterize the covariance matrix spectral width as a function of the wavefield coherence, for propagating signals only. These tests were applied to the geometry of the UnderVolc network system (Fig. 3.1) and we ignored the topography.

Building synthetic data

We built the data as described in Fig. 3.2: we considered a number of K vertically polarized plane surface waves propagating isotropically in the network plane, with the same slowness and frequency (respectively γ and f). The expression of the k^{th} plane wave at coordinate $\mathbf{r} = (x, y)$ is given by

$$\Phi_k(\mathbf{r}, f) = \exp(-2i\pi f \gamma \mathbf{e}_k \cdot \mathbf{r} - i\varphi_k), \quad k = 1 \dots K, \quad (3.1)$$

where φ_k is the phase, $\mathbf{e}_k = \cos(\theta_k)\mathbf{e}_x + \sin(\theta_k)\mathbf{e}_y$ is the direction vector of the wave with θ_k homogeneously distributed between 0 and 2π . The synthetic data vector $\tilde{\mathbf{u}}$ is then obtained from the sum of all recorded plane waves at \mathbf{r}_i , the coordinates of the sensor i :

$$\tilde{u}_i(f) = \sum_{k=1}^K \Phi_k(\mathbf{r}_i, f). \quad (3.2)$$

The method requires averaging with time sub-windows over a given duration Δt . Therefore, we generated M different synthetic records $\tilde{\mathbf{u}}_m(f)$, $m = 1 \dots M$ corresponding to different time sub-windows in the real data. The seismic noise can be seen as an incoherent

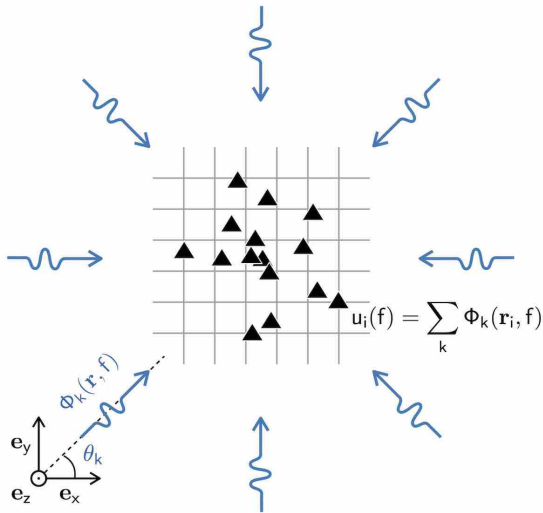


Figure 3.2 – Construction of the wavefield used in synthetic tests: a sum of plane waves recorded by a seismic array. The black triangles show the location of the seismic stations, and the blue arrows represent the incoming planewaves.

wavefield and we simulate it as a sum of plane waves with randomly varying phase from a given sub-window to another. A coherent wavefield is also simulated as a sum of plane waves. In this case, however, the phase of each wave does not change in the sub-windows (we can set φ_k to 0 for instance). We then built the synthetic covariance matrix using equation 2.25 from M different sets of synthetic signals in order to perform the ensemble average (combining equations 3.1 and 3.2)

$$\tilde{\mathbf{C}}(f) = \sum_{m=1}^M \tilde{\mathbf{u}}_m(f) \tilde{\mathbf{u}}_m^\dagger(f). \quad (3.3)$$

Example of the covariance matrix spectra

We set the frequency to 0.2 Hz and the slowness to $1/2000 \text{ s.m}^{-1}$. These parameters are approximately those of the Rayleigh wave propagation within La Réunion crust (Bren-guier et al. 2008b). We investigated the effect of the number of sources and the coherence of these sources within four cases. We used $M = 100$ generations of the wavefield to perform the ensemble average, and we compared the results obtained when $K = 3$ and $K = 100$ plane waves, for the coherent and the incoherent cases (constant and random phases, respectively). The eigenvalue distributions of the covariance matrices obtained in all four cases are presented in Fig. 3.3. Not surprisingly, when the plane waves are coherent, the number of nonzero eigenvalue remains equal to 1 regardless of the number of sources (Fig. 3.3a, b). However, there are 3 nonzero eigenvalues when the wavefield contains 3 incoherent plane waves (Fig. 3.3c), and the covariance matrix spectrum is wider for a wavefield composed of 100 incoherent plane waves (Fig. 3.3d).

These simulations illustrate the main idea used in our analysis, namely that the number of nonzero eigenvalues is related to the number of independent signals. Fig. 3.3(b) illustrates that waves generated by the same source and scattered by the media will produce a covariance matrix of rank 1, because they remain coherent. This is in agreement with the theoretical results provided in (e.g., Aubry and Derode 2009).

Selecting optimal parameters of the data analysis

Results of application of the proposed method to continuous seismic records depend on the choice of the parameters used in the covariance matrix computation, namely of the

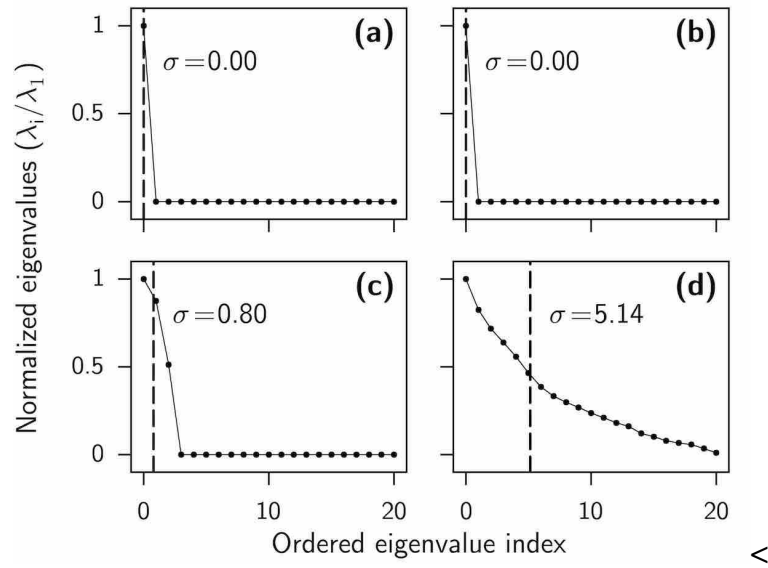


Figure 3.3 – Effect of the coherence or incoherence nature of the wavefield on the eigenvalue distribution of covariance matrices computed from synthetic wavefields, for the 4 cases considered in the section 3.1.2: **(a)** 3 coherent plane waves **(b)** 100 coherent plane waves **(c)** 3 incoherent plane waves **(d)** 100 incoherent plane waves. Values of the covariance matrix spectral width σ are indicated with vertical dashed lines.

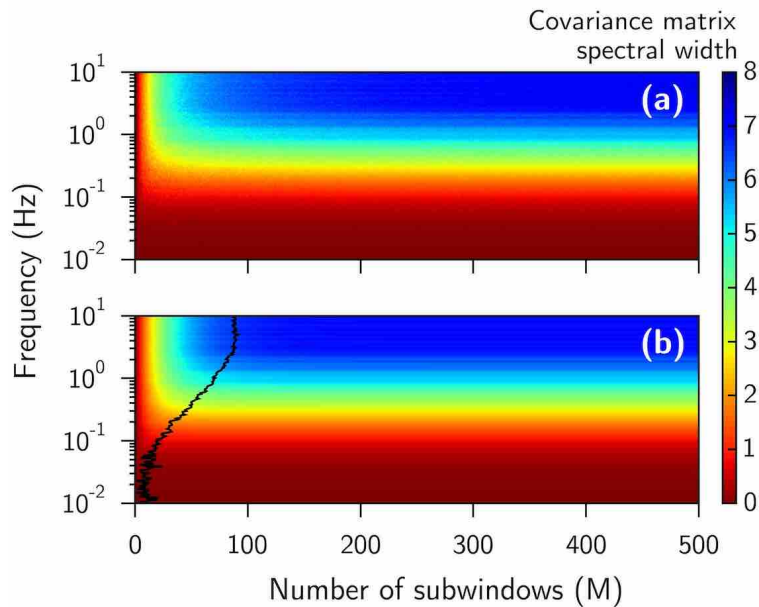


Figure 3.4 – Convergence of the covariance matrix spectral width $\sigma(f)$ as function of the number of sub-windows M (equation 2.25). The synthetic wavefield composed of 100 incoherent plane waves was used for these tests. **(a)** Direct results of synthetic tests. **(b)** Fit obtained with equation 3.4. The value $3M_0(f)$ is shown with the black solid line, and shows the limit where $\sigma(f)$ reaches 95% of its maximum value.

number of sub-windows M and of their length δt . This leads to a trade-off between a statistically robust estimation of the covariance matrix, and a good time resolution.

Length of the sub-windows δt The length of sub-windows δt is directly linked to the lowest frequency that can be resolved. In our case, the longest period that we aim to study is 20 seconds (0.05 Hz). Therefore, we can use 48s-long sub-windows in order to have at least two oscillations of the signal in every sub-window.

Number of sub-windows M When $M = 1$, the estimated covariance matrix is degenerated and its rank equals 1. With increasing M the estimation of the covariance matrix becomes more robust and its eigenvalue spectrum becomes wider. The covariance matrix spectral width $\sigma(f)$ converges to a maximum value $\sigma_{\max}(f)$. We used our synthetic signals to test the speed of this convergence at different frequencies between 0.01 and 10 Hz. We computed the covariance matrix spectral width as function of frequency and of sub-windows number $\sigma(f, M)$ as shown in Fig. 3.4a. This function can be reasonably fitted as proposed in Fig. 3.4b with the following expression:

$$\tilde{\sigma}(f, M) = \sigma_{\max}(f) \left(1 - e^{-M/M_0(f)} \right). \quad (3.4)$$

Equation 3.4 can be used to evaluate the maximal value reachable by the covariance matrix spectral width and, more importantly, to estimate the number of sub-windows required to approach this maximal value with a certain accuracy. Equation 3.4 indicates that 95% of the maximal value of the covariance matrix spectral width is reached after stacking $3M_0(f)$ sub-windows. We display $3M_0(f)$ with a black solid line on Fig. 3.4b. We can see that with considered network configuration a minimal number of 100 sub-windows is required to reach the 95% threshold at all considered frequencies.

3.1.3 Analysis of the Undervolc data

Data preparation

First, all seismograms are bandpassed between 0.01 and 10 Hz and downsampled from 100 to 20 Hz to accelerate the computations. Then, we apply the spectral and temporal normalizations to compensate the non-stationarity of seismic signals and to disregard any amplitude information. [Bensen et al. \(2007\)](#) presents several ways to normalize the data in this way. We apply the spectral whitening and the temporal normalization as described in Fig. 3.5, according to [Bensen et al. \(2007\)](#).

Spectral whitening consists in dividing the signal spectrum by a smooth version of its amplitude

$$u^W(f) = \frac{u(f)}{\langle\langle |u(f)| \rangle\rangle_{df}}, \quad (3.5)$$

where $|u(f)|$ represents the real absolute value of the spectrum, and where $\langle\langle \cdot \rangle\rangle_{df}$ stands for the df -long running average used to smooth the spectral amplitude. In our case, we used $df = 0.33$ Hz. The operator in equation 3.5 preserves the phase, and the amplitude of the whitened spectrum is close to 1 for all frequencies.

We then apply a temporal normalization in the time domain:

$$u^N(t) = \frac{u^W(t)}{\langle\langle |u^W(t)| \rangle\rangle_{dt}} \quad (3.6)$$

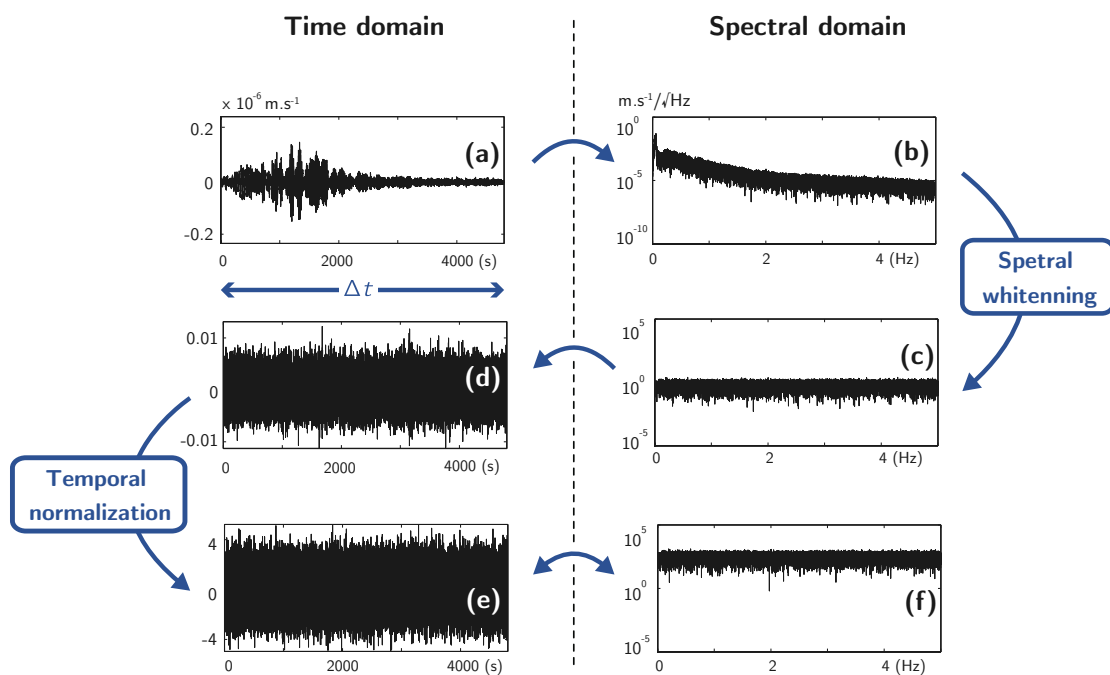


Figure 3.5 – Illustration of the pre-processing of seismic records. (a) Raw data. (b) Raw data amplitude Fourier spectra. (c) Whitened spectral amplitude. (d) Real part the inverse Fourier transform of the of whitened spectrum. (e) Temporally normalized time series. (f) Fourier amplitude spectra of (e).

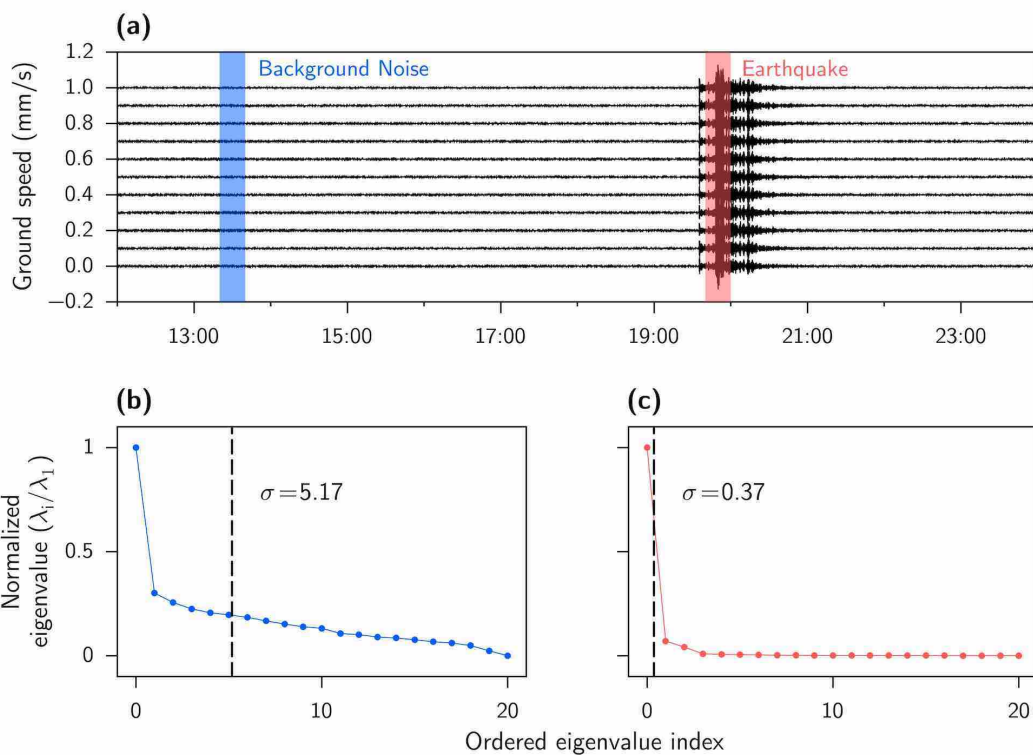


Figure 3.6 – Example of covariance matrix spectra computed from real records (shown in **(a)** after band passing filtered around 0.07 Hz). Covariance matrices were computed at $f = 0.07$ Hz. **(b)** Covariance matrix spectrum for noise. **(c)** Covariance matrix spectrum for signal. Values of the covariance matrix spectral width σ are indicated with vertical dashed lines.

where $u^W(t)$ is the real part of the inverse Fourier transform of $u^W(f)$ and $\langle\langle\cdot\rangle\rangle_{dt}$ is the dt -long running average used to smooth the temporal normalization. In our case, we use a 1.25 second-smoothing window.

Difference between the background noise and an earthquake

We applied our technique to the records of the $M \approx 7.5$ Nicobar island earthquake which occurred around 7:00 PM UTC on June, 12. We analyzed the signals over two time periods shown in Fig. 3.6a: a *noise* window where no significant event occurred and an *event* window where the Nicobar Island earthquake is recorded by the UnderVolc array. We use a hundred of 48s-long 50% overlapping sub-windows to compute the two covariance matrices. Therefore the full window durations are $\Delta t \approx 48 \times 100/2 = 2400$ seconds. We compute the covariance matrix eigenvalues at a frequency $f = 0.07$ Hz.

The covariance matrix spectrum computed from the noise window shown in Fig. 3.6b contains no null eigenvalues and its width is large, indicating the low level of the signal coherence. On the contrary, the signal generated by the Nicobar Island earthquake is strongly coherent across the network and the corresponding covariance matrix spectrum shown in Fig. 3.6c has a single eigenvalue that is significantly greater than zero. In this case, the spectral width σ is very small, as expected.

Analysis of continuous records in 2010

We apply the method over the whole year 2010. Following the results of the synthetic tests, we compute covariance matrix on 2400s-long windows. Each window contains 100 of half overlapping 48s-long sub-windows. The spectral width of the covariance matrix as a function of time and frequency is displayed In Fig. 3.7a.

As discussed in section 3.1.2, the covariance matrix spectral width does not converge to the same value at all frequencies. Therefore, we normalize the raw result shown in Fig. 3.7a by using the maximal value $\sigma_{\max}(f)$ obtained in section 3.1.2. The normalized covariance matrix spectral width shown in Fig. 3.7b exhibits strong variability in time and frequency. While most of this variability is related to variations of the level of coherency of the recorded wavefield, some of the anomalies can be caused by mis-functioning stations. For this reason, we show in Fig. 3.7c the percentage of working stations as a function of time and can see that some of the apparent increases of coherence, in particularly those in April and November, are caused by missing data.

At frequencies below 0.15 Hz, we observe that the normalized spectral width of the covariance matrix is higher than 1, highlighting that the coherence level is much lower than the synthetic results. This indicates that the noise recorded at these relatively long periods does not correspond to propagating long-period seismic waves that would be expected to be coherent across the network. Indeed, because of the non-perfect condition of temporary installation of seismic sensors on the Piton de la Fournaise volcano during the UnderVolc experiment, the long period noise is dominated by local perturbations likely of meteorological origin.

To estimate the amplitude of this non-propagating noise, we compare in Fig. 3.8b the spectral density of the records provided by the UnderVolc sensors over one day with the RER seismic station from the Geoscope seismic network located at la Réunion Island (see Fig. 3.1). We see that below 0.1 Hz, the spectral amplitude of the noise recorded by the UnderVolc stations is high compared to the one recorded by RER. The two microseismic peaks are clearly visible in the RER data (at 0.06 and 0.14 Hz) whereas only the secondary microseismic peak is recorded by the UnderVolc stations. We also compare the spectra of the Nicobar Is. earthquake from record of UnderVolc and RER stations in Fig. 3.8c,

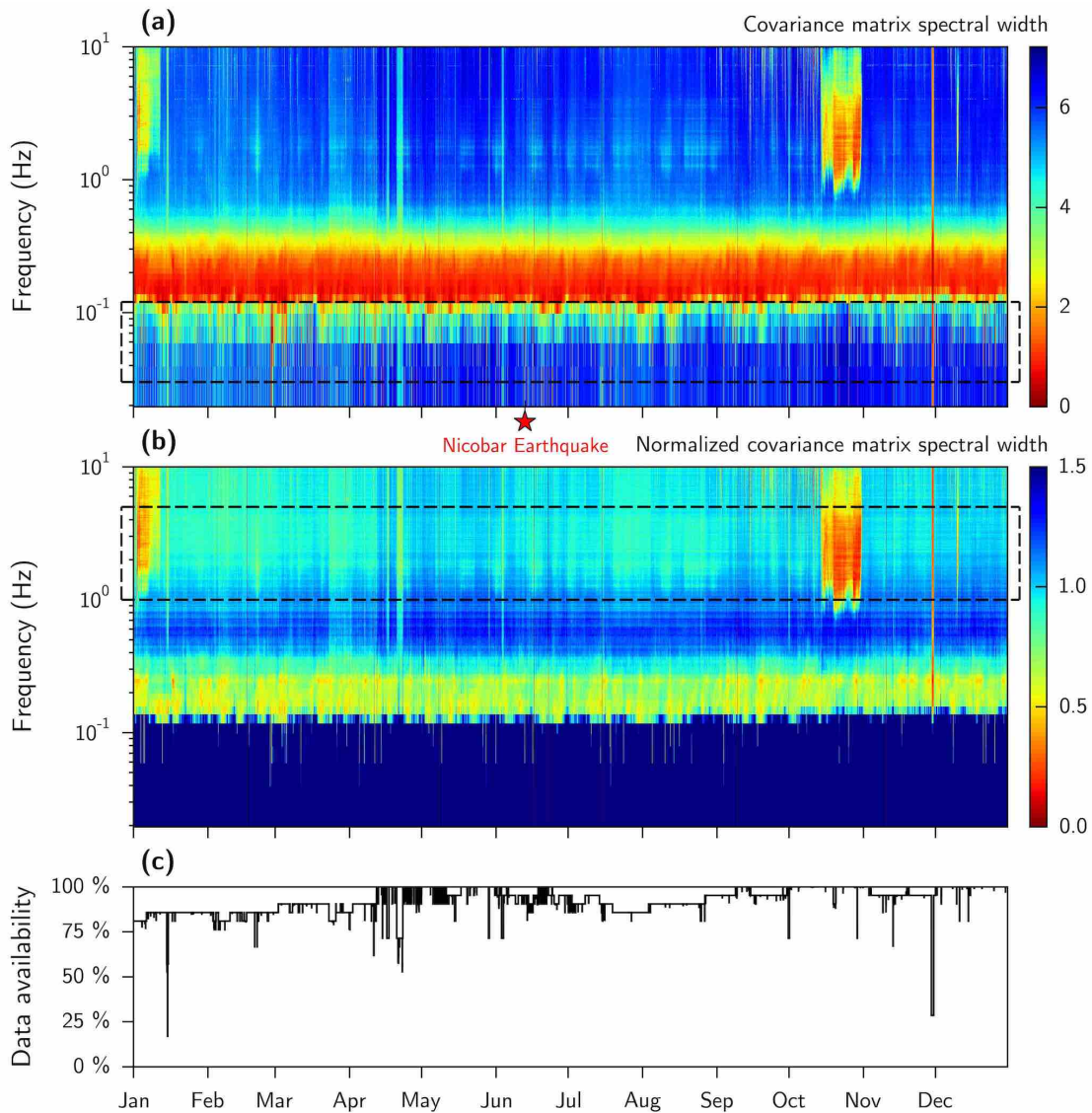


Figure 3.7 – Results of analysis of continuous records of the UnderVolc seismic network in 2010. (a) Spectral width of covariance matrix σ as function of time and frequency. **(b)** Normalized covariance matrix spectral width obtained after dividing the values shown in (a) by the maximal expected width estimated in section 3.1.2. **(c)** Amount of working station as a function of time.

and observe that it does not differ from UnderVolc stations to RER, at frequencies higher than 0.02 Hz. We thus conclude that earthquake detection is possible. This additional information is clearly in agreement with the hypothesis of non-propagating noise recorded by the UnderVolc stations, and with the results obtained in Fig. 3.7.

Around 0.2 Hz, many coherent arrivals are also visible all year long around 0.2 Hz corresponding to the secondary microseismic peak. This is in agreement with the standard model considering that this peak consists of waves emitted by strong sources originating from the interaction between the oceanic gravity waves and the seafloor (Longuet-Higgins 1950, Friedrich et al. 1998). Waves from every individual microseismic source are coherent across the network.

Finally, in the frequency range between 0.5 and 5 Hz, the normalized spectral width of the covariance matrix is most of the time close to 1, indicating that incoherent signals are recorded by the network. Strongly coherent signals are appearing a few times in the

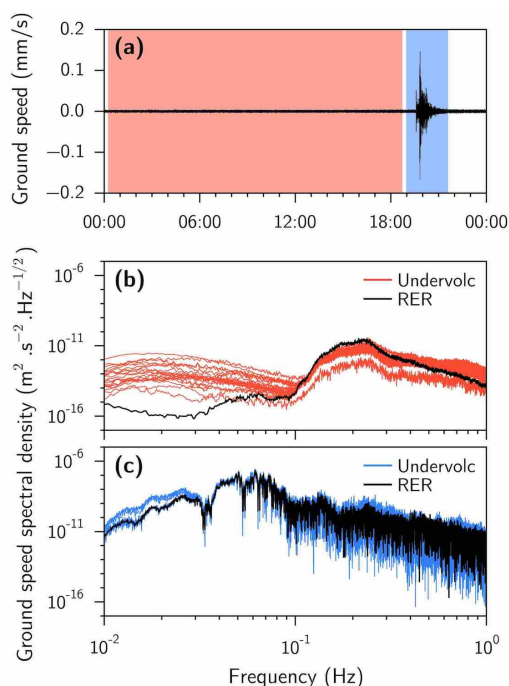


Figure 3.8 – Comparison of amplitude spectra from records of the Undervolc stations and from the broadband GEOSCOPE station RER. We analyse the records of July, 12, when the Nicobar Island earthquake occurred around 7:00 PM. The power spectral densities of the noise records (collected from 0:00 to 6:20 PM) were computed using Welch’s method (Welch 1967). The event (signal recorded from 6:20 PM to 8:20 PM) amplitude spectra were computed using the Fast Fourier Transform.

year in this frequency range and are linked to seismo-volcanic activity of the Piton de la Fournaise volcano as described later in this paper.

Example of coherent signals from teleseismic earthquakes

Strong peaks of coherence are clearly visible on Fig 3.7a, at frequencies below 0.15 Hz. Most of these peaks coincide in time with expected arrivals of surface waves generated by teleseismic earthquakes. This indicates that the analysis of the wavefield coherence could be used for earthquake detection. Therefore, we establish a simple detection criteria based on our method and compare its results with the Global Centroid Moment Tensor (GCMT) catalog (Ekström et al. 2012). We use the covariance matrix spectral width shown in Fig. 3.7a, because the normalized results obtained in Fig. 3.7b are saturated by the non-propagating noise, as discussed in the previous section.

Observation and detection of teleseismic earthquakes during 8 days in June 2010 is illustrated in Fig. 3.9. Signals from strongest events (such as Nicobar earthquakes occurred on June 12) are clearly seen on seismograms Fig. 3.9d. In addition, the analysis of the covariance matrix width (Figs. 3.9a and b) reveals peaks from weaker earthquakes not directly observable on seismograms. Fig. 3.9b shows the covariance matrix width $\hat{\sigma}(t)$ averaged between 0.03 and 0.12 Hz (delimited with a black frame in Fig. 3.9a) that we use in our simple detection algorithm. First, we identify all intervals when the averaged spectral width is below the median of $\hat{\sigma}(t)$ (shown with the horizontal dotted line in Fig. 3.9b). To reduce the influence of the small noisy peaks, we establish a detection threshold σ_{th} (shown with the horizontal dashed line in Fig. 3.9b) and keep only the intervals whose absolute minima are below this threshold. We choose $\sigma_{th} = 3.3$ in this example. The retained intervals are considered as *alarms* and are shown with colored boxes in Figs. 3.9b-d.

Finally, we compare the obtained detections with the Global Centroid Moment Tensor (GCMT) catalog. For every earthquake from this catalog, we compute the arrival time of Rayleigh waves at the position of the Undervolc array. An earthquake is considered as *detected* when its arrival time at La Réunion Island is within an *alarm* window (colored

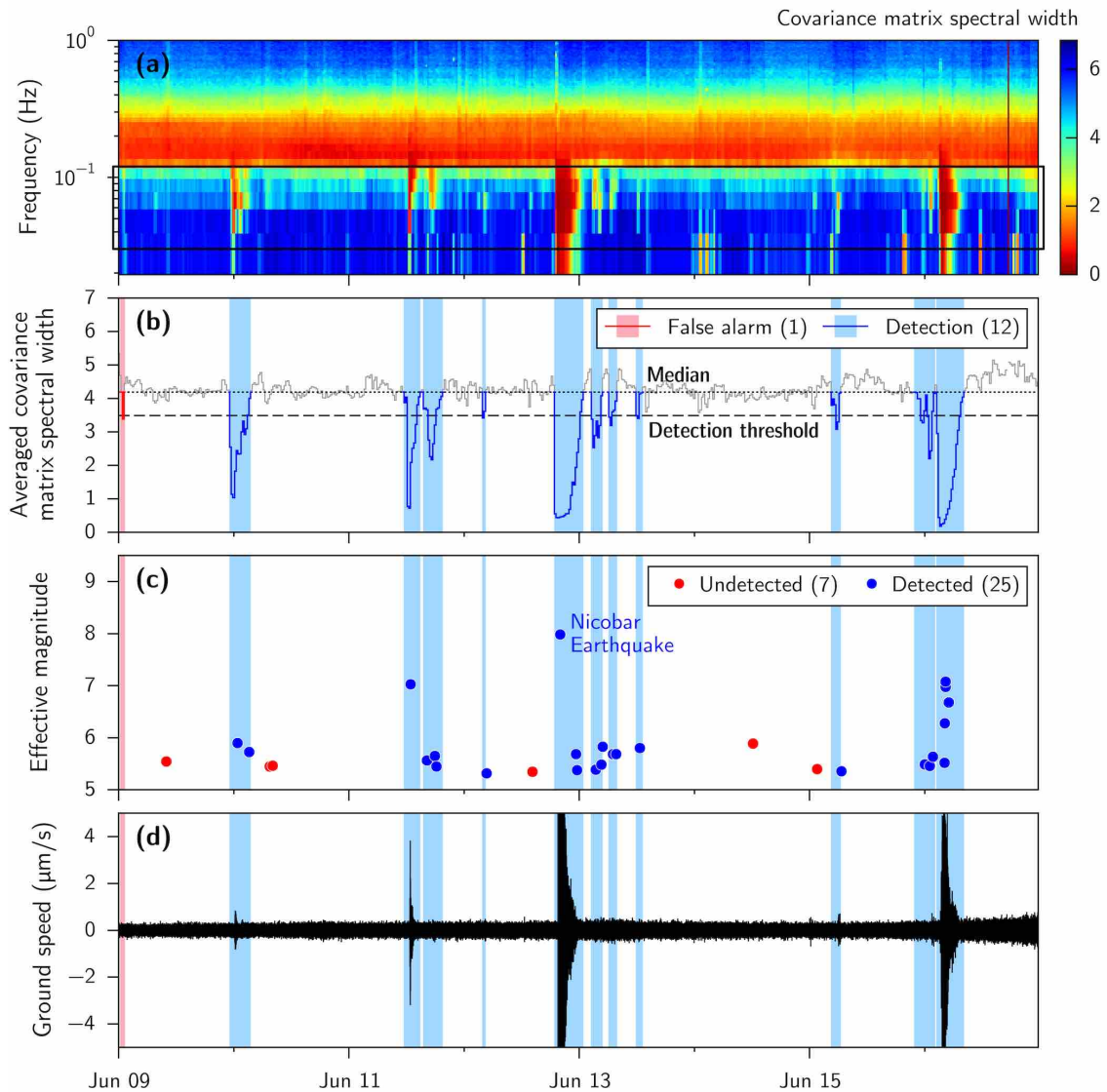


Figure 3.9 – Example of earthquake detection around June, 12 of 2010. Observation of teleseismic earthquakes at frequencies below 0.12 Hz. (a) Raw covariance matrix spectral width. The black frame indicates the zone where the covariance matrix spectral width is averaged. (b) Covariance matrix spectral width averaged between 0.03 and 0.12 Hz. Colored segments indicate time period where a detection is declared. Blue segment are *detections* that correspond to at least one event from the GCMT catalog, red segments are *false alarms*. (c) Arrival times of earthquakes from the GCMT catalog as a function of time and effective magnitude (equation 3.7). Blue circles show the *detected* events, and red circles show the *undetected* ones. (d) Seismic trace recorded by the Undervolc station UV14 as a function of time, bandpass-filtered between 0.03 and 0.12 Hz.

boxes in Fig. 3.9b-d).

Even after removing the amplitude information with the pre-processing technique present in section 3.1.3, the spatial coherence of the recorded signals related to the earthquake and the expected detection efficiency depend on the signal-to-noise ratio. This ratio depends in turn on the earthquake magnitude and on the epicentral distance between the earthquake and La Réunion Island. To simplify this relationship, we normalize the magnitudes to compensate the differences in the epicentral distances Δ . The corrected magnitudes M_{eff} correspond to distance of 90 degrees (derived from Aki and Richards 2002; app. 2) :

$$M_{eff} = M_s + 1.656 \log(\Delta) + 1.656 \log(90). \quad (3.7)$$

Corrected times and magnitudes of earthquakes from the GCMT catalog are shown with colored circles in Fig. 3.9c. During the considered period from June, 9 to June 17, the GCMT catalog lists 32 earthquakes with effective magnitude greater than 5. Earthquakes with arrival times that are not present in a detection window are considered as *undetected*, and are shown in red circles. Otherwise, they are considered as *detected*, and are shown in blue circles. In a similar way, time segments of $\hat{\sigma}(t)$ that contain no earthquakes are considered as *false alarms* and are represented in red color, while those that contain one or more earthquakes are considered as *detections* and are represented in blue.

Observation of the teleseismic activity during 2010

We applied the method described in the previous sub-section to the continuous records of 2010. The relative amounts of *detections* and *false alarms* and of *detected* and *undetected* events depends on the choice of the detection threshold σ_{th} and of the minimal effective magnitude of earthquakes taken into account. We define two ratios R^{real} and R^{succ} . The first one, R^{real} , is defined as the ratio between the number of *detections* and the total number of *alarms*. The second one, R^{succ} , is the ratio of the *detected* earthquakes over the total number of earthquakes listed in the GCMT catalog. We vary the detection threshold from 2 to 3.5 and the minimal effective magnitude from 4 to 7.5, and observe the variation of R^{real} and R^{succ} in Fig. 3.10a and b, respectively. We found that our simple detection algorithm is ‘optimized’ with the detection threshold of 3.3 and with the minimal effective magnitude of 5.3. With these parameters, 75% of *alarms* are successful *detections* and we detect 75% of earthquakes (951 *detections* out of 1246 total events). The final detection results during the year 2010 are shown in Fig. 3.11.

Observation of the seismo-volcanic activity

We use the covariance matrix spectral width at high frequencies to detect the seismo-volcanic activity because the volcanic sources are close to the network and the high-frequency radiation is not strongly attenuated as was the case for the teleseismic earthquakes. We average the normalized covariance matrix spectral width between 1 and 5 Hz (frequency region delimited with a black dashed frame in Fig. 3.7b), and present it as a function of time in Fig. 3.12a. We see that the signal is not coherent (the average covariance matrix spectral width is close to 1) most of the time. A set of coherent events are detected as minima of spectral width. We compare their occurrences and durations with the catalog of Piton de la Fournaise activity described in Roult et al. (2012). Fig. 3.12b, c and d are zoomed in periods of known volcanic eruptions.

In the three cases presented in Fig. 3.12b, c and d, the occurrence and duration of volcanic tremors are well observed. In these three cases, the yellow windows indicate the durations of the volcanic tremors, and show great agreement with the periods of increased

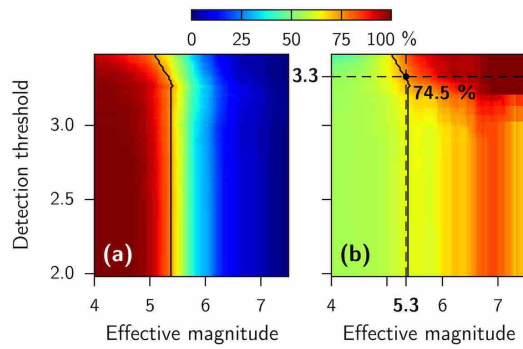


Figure 3.10 – Analysis of observation of teleseismic earthquakes. (a) R^{real} as function of the detection threshold and the minimal effective magnitude. The black line indicates points where $R^{real} = 75\%$. (b) R^{succ} as function of the detection threshold and the minimal effective magnitude. The solid black line corresponds to the coordinates where $R^{real} = 75\%$. The maximal value of R^{succ} on these coordinates is 74.5% and is reached when the detection threshold is 3.3 and the minimal effective magnitude is 5.3.

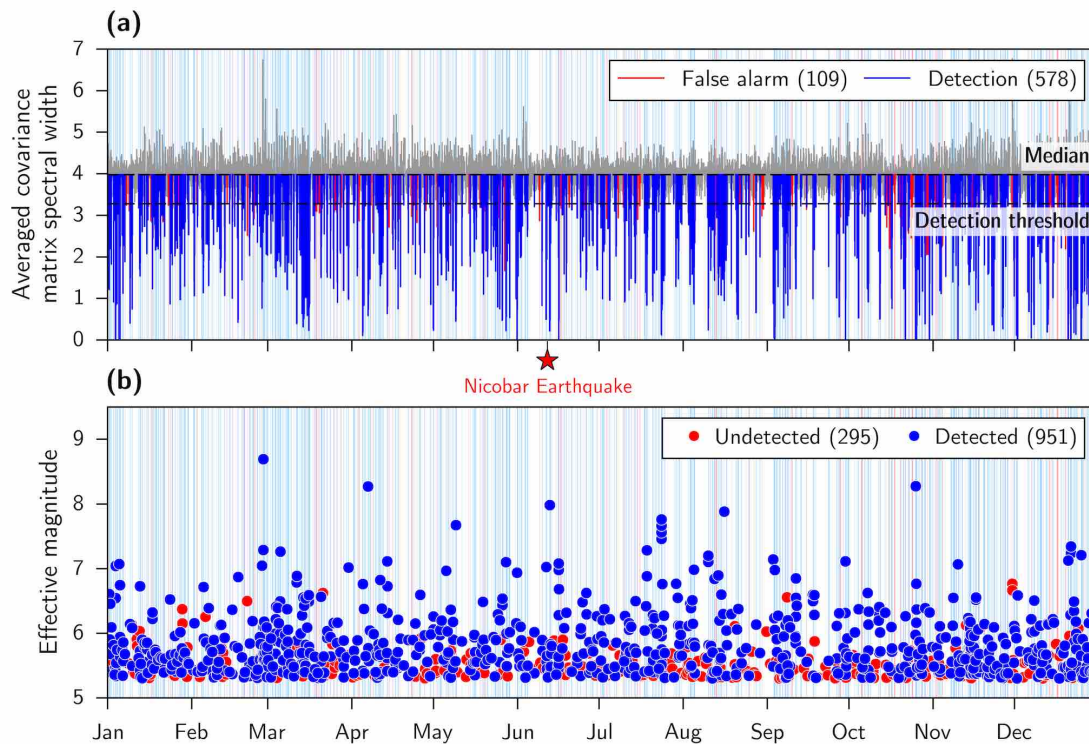


Figure 3.11 – Earthquake detector. Observation of teleseismic earthquakes at frequencies below 0.1 Hz. (a) Normalized covariance matrix spectral width σ averaged between 0.03 and 0.12 Hz. Blue segments indicate time periods when the averaged value of σ that correspond to at least one event from the GCMT catalog and red segments indicate time periods with no matching events from the GCMT catalog. (b) Earthquakes from the GCMT catalog as a function of time and effective magnitude (equation 3.7). Blue and red circles show 'detected' and 'undetected' events, respectively.

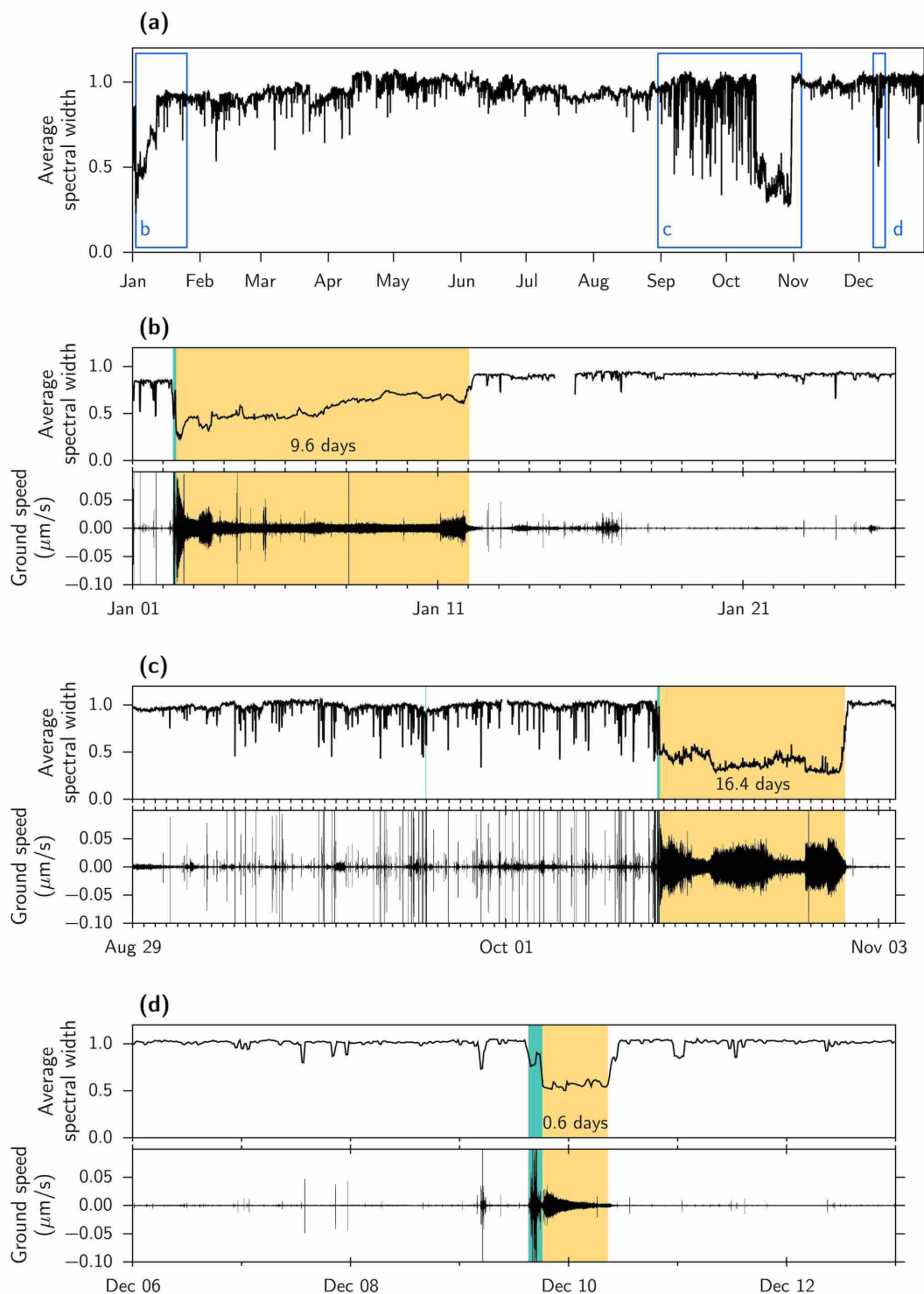


Figure 3.12 – Observation of the seismo-volcanic activity at frequencies above 1 Hz. (a) Covariance matrix spectral width σ averaged between 1 and 5 Hz and shown for the whole year 2010. Blue frames indicate periods corresponding to three eruptive crises occurred in 2010. **(b)** Zoom between January 1 and January 20. **(c)** Zoom between August 28 and November 16. **(d)** Zoom between December 3 and December 14. In each case, green boxes indicate pre-eruptive seismic crisis, and yellow boxes indicate starting and ending dates of eruptions from [Roult et al. \(2012\)](#). Records from the UV05 seismic station filtered between 1 and 5 Hz are presented in each case.

coherence. The green windows show the seismic crises recorded before the eruptions. The eruption of October is particularly visible because it was the most energetic in 2010. The smaller eruptions of January and December are also clearly distinguishable.

Numerous minima of the covariance matrix spectral with averaged between 1 and 5 Hz are also observed outside the cataloged periods of volcanic activity. A few of them correspond to days with missing data (Fig. 3.7c). However, the large majority of these minima correspond to small volcanic earthquakes. In particular, we observe a long period (~ 1.5 months) of increased seismic activity preceding the eruption in October 2010. A detailed comparison with seismograms shows that some small earthquakes are not reflected in the presented results of the covariance matrix analysis. This is related to the choice of the duration of the time windows δt and Δt in equation 2.25. In this study, we selected relatively long windows to be able to analyze the long-period part of the signal, while a more detailed study of the seismo-volcanic activity would require focusing at frequencies above 1 Hz and, therefore, running the covariance matrix analysis in shorter windows.

3.1.4 Discussion and conclusions

We developed an approach that allows to quantify the level of spatial coherence of a wavefield recorded by a network of sensors with a single parameter that is the spectral width of the covariance matrix. We first tested this method with a set of synthetic signals and illustrated it with an application to the real data, namely, to one year of continuous vertical-component records from 21 broad-band seismographs installed around the Piton de la Fournaise volcano at La Réunion island. The main result of our analysis is presented in Fig. 3.7 and shows that the level of the seismic wavefield coherence varies strongly with respect to the time and frequency, and that clearly distinguishable coherent events are seen in different frequency ranges.

We distinguish three frequency bands with different characteristic behavior of the wavefield coherence. Between 1 and 5 Hz, the wavefield is dominated most of the time by the propagating noise, or diffuse seismic field and the coherent signals are emitted by sources located very close to the network, mainly related to the seismo-volcanic activity of the Piton de la Fournaise volcano. For such local sources the high frequency radiation is not attenuated. In the secondary microseismic band around 0.2 Hz, the wavefield is on average more coherent than at lower and higher frequencies. A detailed analysis of Fig. 3.7b shows that the microseismic noise in this frequency range is composed of many separated arrivals characterized by the increased values of the coherency. The duration of these microseismic coherent arrivals varies between a few hours and a few tens of hours and they likely correspond to signals generated by separated storms or swells. At lowest frequencies, below 0.1 Hz, the records are dominated by the non-propagating noise because of the non-perfect conditions of temporary installation of seismic sensors on a volcano. The coherent arrivals from the teleseismic earthquakes are clearly distinguishable in this frequency in contrast with the incoherent noise leading to a possibility to detect them based on the analysis of the covariance matrix spectra.

The presented example clearly shows that the array covariance matrix is very sensitive to the coherent signals generated by strong earthquakes and by the volcano. At the same time, the simple approach discussed in our paper does not aim to outperform more classical earthquake detection algorithms. The UndeVolc array is not and well suited for the detection of teleseismic earthquakes. This temporary seismic network was installed on an active volcano where most of the sites were affected by strong environmental noise

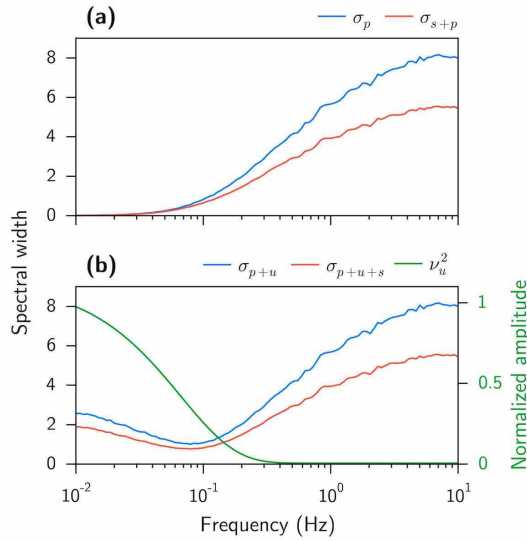


Figure 3.13 – (a) Covariance matrix spectral width as function of frequency estimated for a combination of two first terms from equation 2.12: with an propagating noise only (σ_p) and adding one coherent signal (σ_{s+p}). (b) Similar to (a) but after adding the third term: the non-propagating noise (whose spectral intensity $v_n^2(f)$ is schematically illustrated with the green line).

generated either by volcanic or by meteorological activity. The detection based on the covariance matrix spectral width can strongly improve the analysis of emergent signals such as, for example, volcanic tremors illustrated in Fig. 3.12. The detection of such signals could be significantly enhanced with a choice of parameters used in the covariance matrix computation (the number of sub-windows M and of their length δt) specifically optimized for high frequencies and not for a broadband analysis presented in this paper. Such a more detailed study is, however, out of the scope of this initial paper mainly aimed at presenting the theory and the method of the covariance matrix analysis.

Fig. 3.13 illustrates relative contributions of different types of signals described in equation 2.12 within these bands. We see in Fig. 3.13a that covariance matrix spectral width of the propagating noise (in blue line) decreases with the decreasing frequency. This indicates that the spatial coherence of the propagating noise increases when the corresponding wavelength becomes too wide in comparison with the inter-station spacing and the array aperture. The red line shows the same propagating noise, with an additional 10-times stronger coherent source. The detection of this coherent source is therefore hard to accomplish because the difference between the propagating noise and the source embedded in propagating noise spectral width is weak at frequencies below 0.15 Hz. The covariance matrix spectral width of the data shows that the wavefield contains incoherent signal at frequencies below 0.1 Hz which can be explained by the presence of non-propagating noise at the sensors, most probably due to the weather perturbations. The detectability of the seismic source in the presence of this additional non-propagative noise is presented in Fig. 3.13b. The green line shows the non-propagating noise spectral amplitude. The noise amplitude is chosen to be the same as the signal's one and rapidly decays with the increasing frequency. The source is now detected, and this model explains the detection capacity of our method for this dataset.

The results of the covariance matrix analysis are not unique and depend on the choice of the time windows δt and Δt in equation 2.25. In the presented example of the application of the method to the real data we decided, for simplicity, to use the constant set of these windows for all frequencies and selected the long windows lengths that allowed us to analyze the longest periods. As a consequence of this choice, the results of the analysis of the seismo-volcanic activity shown in Fig. 3.12 are not optimal, as discussed in the previous section. This shows that in the future studies the choice of the windows δt and

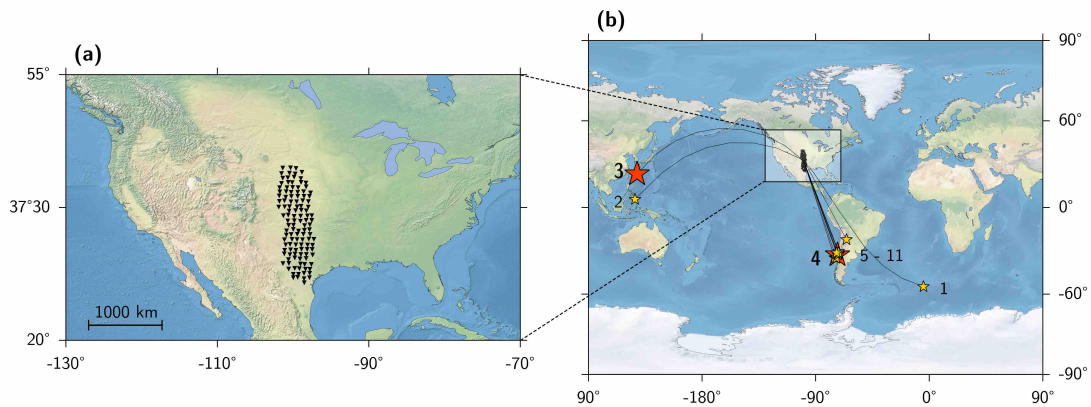


Figure 3.14 – (a) Locations of the 121 seismic stations (black triangles) of the US Transportable Array Experiment that were continuously recording the seismic wavefield from January 1 to December 31 of 2010. (b) Epicenters of earthquakes of magnitude higher than 5.3 that occurred between Feb. 26 and Mar. 02, obtained from the USGS database. Red stars stand for magnitude higher than 7. Numbering indicates the order of occurrence and corresponds to the one used in Fig. 3.16. The great circle path between each epicenter and the center of the USArray is shown in black line.

Δt should be optimized considering the specific properties of the studied signals and, in particular, their dominant frequency range.

The analyzed seismic records were pre-processed with applying a spectral whitening and a temporal normalization to compensate for inhomogeneity of amplitudes, which is a standard practice in many applications based on correlations of the ambient seismic noise (e.g., [Bensen et al. 2007](#)). Our results show that, while these procedures compensate for the effect of strongest amplitudes, they are not sufficient to fully ‘randomize’ the records and that a clear imprint of different seismic sources remains in the wavefield coherency.

3.2 Application to the transportable component of the USArray in 2010

We analyze the vertical-component continuous records from the US Transportable Array Experiment (USArray) during the year 2010. We selected the records of 121 stations shown as black triangles in Fig. 3.1a that were operating continuously from January 1 to December 31.

3.2.1 Selection of the parameters

As explained in [Seydoux et al. \(2016a\)](#), the choice of the sub-windows lengths used in equation 2.25 is constrained by the minimal analyzed frequency and by the maximal travel-time spent by the slowest seismic waves to propagate across the network. With the selected seismic array, the largest inter-station spacing is about 1200 kilometers. Considering a minimal wave speed of 1.25 km/s, the maximal travel-time is of 960 seconds. We therefore use 960 seconds-long sub-windows, giving access to frequencies above ~ 0.002 Hz.

The number of sub-windows M is related to the convergence of the covariance matrix estimation. In our case, we wanted to estimate the covariance matrix while keeping

a maximal temporal resolution. We therefore selected 20 overlapping subwindows of width 960 seconds, leading to a duration of the computational window of $\Delta t \approx M\delta t/2 = 20 \times 960/2 = 9600$ seconds (hence, 2.60 hours). With a synthetic test similar to the one described by [Seydoux et al. \(2016a\)](#) we observe that the covariance matrix spectral width of the seismic noise varies between 4 at low frequencies ($f = 0.01$ Hz) to 5.5 at frequencies higher than 0.2 Hz (which correspond to $M_{\text{eff}}/2$, where M_{eff} is the effective number of overlapping subwindows). We notice that this choice of parameter does not allow to distinguish between uncorrelated noise and seismic noise at frequencies higher than 0.2 Hz. A higher number of subwindows M should be taken into account in this specific goal, with a substantial degradation in the temporal resolution of the analysis. We selected this parameter on purpose, because we considered the data to be free from uncorrelated noise, and because our main goal is to investigate the presence of coherent seismic sources in the seismic data.

3.2.2 Results

The covariance matrix spectral width $\sigma(f, t)$ computed from one year of raw data is shown in Fig. 3.15a. During most of the time, the seismic wavefield is dominated by apparently coherent signals (with low values of σ) at high and low frequencies. This reflects the fact that the wavefield represented by the ensemble of the continuous raw seismic records by the USArray stations is strongly non-stationary and very far from being equipartitioned. A close inspection of the records shows that most of the low-sigma parts of the wavefield correspond to local noise recorded by one or a few closely located stations. The influence of this type of non-stationary signals can be efficiently suppressed with the energy normalization of raw records in time and frequency domain often used in the analysis of the seismic noise cross-correlations (e.g. [Bensen et al. 2007](#)). We therefore apply this normalisation to the ensemble of analyzed records. We start with a time-normalization and divide raw traces by the average amplitude computed in a 1.25 second-long sliding time window. We then whiten the data spectrum within every Δt window. This is done by dividing the spectra by its amplitude smoothed in 0.8 Hz sliding-frequency window.

The covariance matrix spectral width computed from the normalized seismograms is shown in Fig. 3.15b. Compared to the results obtained from the raw records (Fig. 3.15a), most of the low-sigma parts disappear showing that the pre-processing of individual records, indeed, helps to ‘homogenize’ the wavefield. Nevertheless, some strongly coherent arrivals are still present even after this energy normalization. Results of a more detailed analysis of most typical of this signals indicated with white frames in Fig. 3.15b are presented in Figs 3.16, 3.17a, 3.17c, and 3.17e and are described in the following subsections.

3.2.3 Observation of the teleseismic activity

At frequencies below 0.1 Hz, many short-time drops of the covariance matrix spectral width are observable in Fig. 3.15b indicating that short-duration coherent signals are detected. These peaks correspond to waves emitted by teleseismic earthquakes. The high frequency part of this radiation is strongly attenuated and remains visible only for strongest earthquakes.

A typical example of earthquake related signals is shown in Fig. 3.16. In order to illustrate the relationship of the covariance matrix spectral width with the wavefield coherence, two examples of the covariance matrix spectra are provided in Fig. 3.16b and

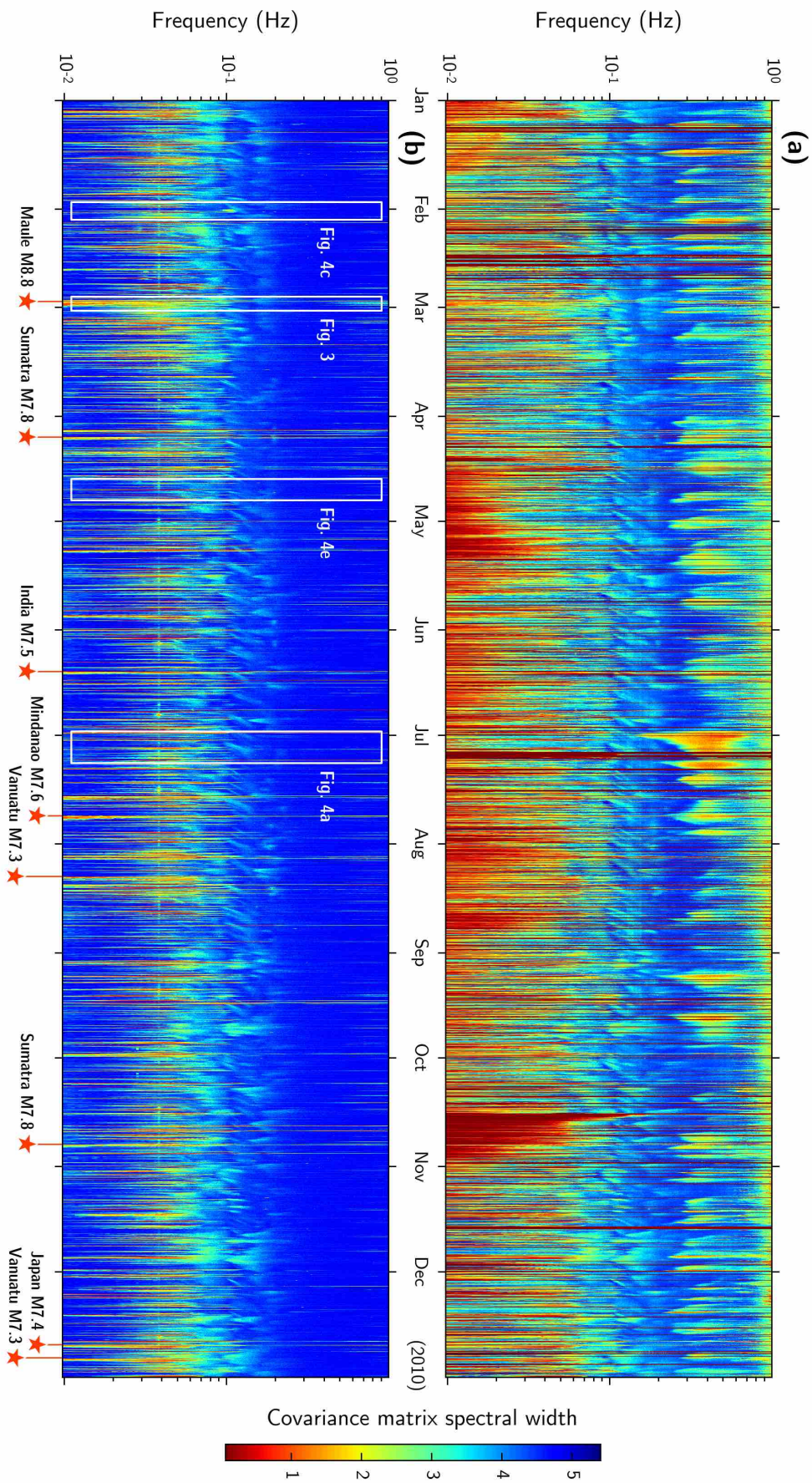


Figure 3.15 – Covariance matrix spectral width versus time and frequency. The colorscale indicates the covariance matrix spectral width. Windows where the covariance matrix spectral width is close to 0 contain coherent signals, and windows where the covariance matrix spectral width is close to 5 likely contain incoherent propagating noise. **(a)** No normalization is applied to the data before computation. **(b)** Spectral and temporal normalizations are applied to the data before computation. White frames show the time and frequency limits of the zooms in Fig. 3.16, 3.17a, 3.17c and 3.17e. Red stars indicate the arrival times of Rayleigh waves emitted by $M \geq 7.3$ earthquakes (obtained from the USGS earthquake database)

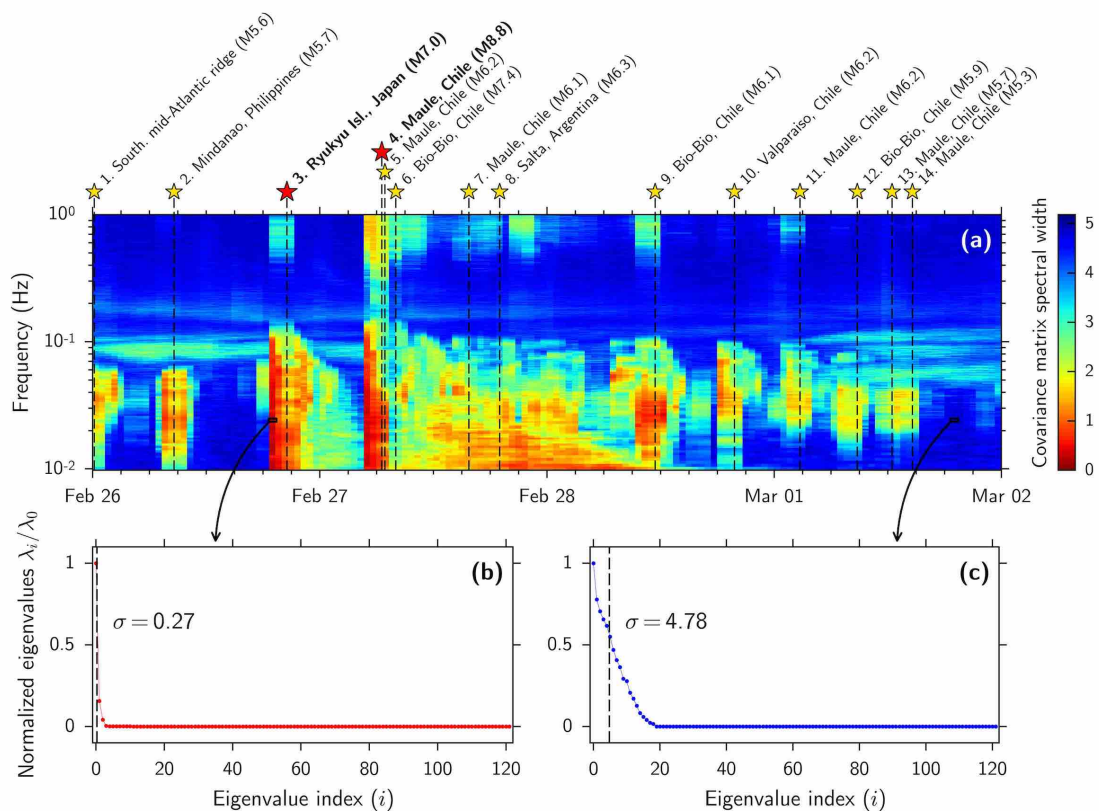


Figure 3.16 – (a) Example of earthquakes influence on covariance matrix spectral width obtained from data normalized with the process proposed in [Bensen et al. \(2007\)](#). Earthquakes names are given with their corresponding magnitude; the related epicenters and great circle paths are reported on the global map shown in [Fig. 3.1b](#). **(b)** Covariance matrix spectrum of the M7.0 Ryukyu IIs. earthquake, at 0.25 Hz. **(c)** Covariance matrix spectrum taken in a time window when coherent event occurs, at 0.25 Hz.

c. The covariance matrix spectrum obtained from a time window when the M7.0 Ryukyu Islands earthquake occurred and at 0.25 Hz (Fig. 3.16b) shows up to be dominated by the first eigenvalue, in contrast with the covariance matrix spectrum obtained in a window that looks free from any coherent signal (Fig. 3.16c). The values of the corresponding spectral width is explicated in each case.

The analyzed window shown in Fig. 3.16 contains several events including the M=8.8 Maule megathrust earthquake that occurred near the coast of central Chile (Delouis et al. 2010). The arrival times of the Rayleigh waves from the earthquakes of magnitude higher than 5.3 are indicated with black dashed lines on the figure (only the ones with magnitude higher than 6 are indicated within the 24 hours after the main shock of the Maule earthquake in order not to saturate the figure). The corresponding epicenter locations and the great circle path of Rayleigh waves are indicated in Fig. 3.1b. The extended duration of the earthquakes-related drops of σ is explained by the 9600 seconds long windows used in the computation. All seismic events are clearly distinguishable as low-sigma ($\lesssim 1$) vertical bars at frequencies below 0.1 Hz. For strongest events, the increase of coherence is also observed at high frequencies. So far, the M=8.8 Maule earthquake appears to be highly coherent at all frequencies between 0.01 and 1 Hz. Moreover, it is followed by more than 24 hours of coherent coda whose duration is clearly increasing with decreasing frequencies. The long-period coda remains observable over longer time periods because of the slower attenuation of long-period seismic waves within the Earth.

3.2.4 Microseismic noise: surface waves generated by swells

At frequencies between 0.09 and 0.2 Hz, that correspond to the secondary ocean-generated microseismic band, the covariance matrix spectral width presented in Fig. 3.15b shows many arrivals with increased coherence. Typical characteristics of these signals are their relatively long durations (a few days) and their dispersion with higher frequencies arriving after the low frequencies. A typical example of such dispersive arrival is shown in Fig. 3.17a.

We apply the plane wave beamforming analysis described in section 2.2.3 to the covariance matrix averaged in time and frequency: between 0.09 and 0.12 Hz and over the whole day of 2010/07/04 (Fig. 3.17b). The beamforming output shows aliasing artifacts because of the short wavelength of the incoming waves compared to the inter-station spacing. The location and shape of the principal lobe, however, indicates that the waves are incoming from the south south-west of the USArray (with an azimuth between 175 and 225 degrees). The estimated slowness is about 0.34 s.km^{-1} and corresponds to a Rayleigh wave.

Considering the provenance and the dispersive nature of the signal analyzed in Fig. 3.17b the most plausible explanation of its origin is that it was generated in vicinity of the coast of Mexico by the non-linear interaction of an incoming oceanic swell with a relatively shallow near-coastal sea floor (e.g. Bromirski et al. 2013) and then propagate as Rayleigh waves from the coast to the USArray. The long signal duration and visible dispersion is produced by the slowly propagating oceanic swell, whose source is somewhere far from the coast in the Pacific ocean.

3.2.5 Microseismic noise: teleseismic P-waves

While many signals detected in the frequency band between 0.1 and 0.2 Hz are dispersive and correspond to surface waves generated by distant swells, some of them have different

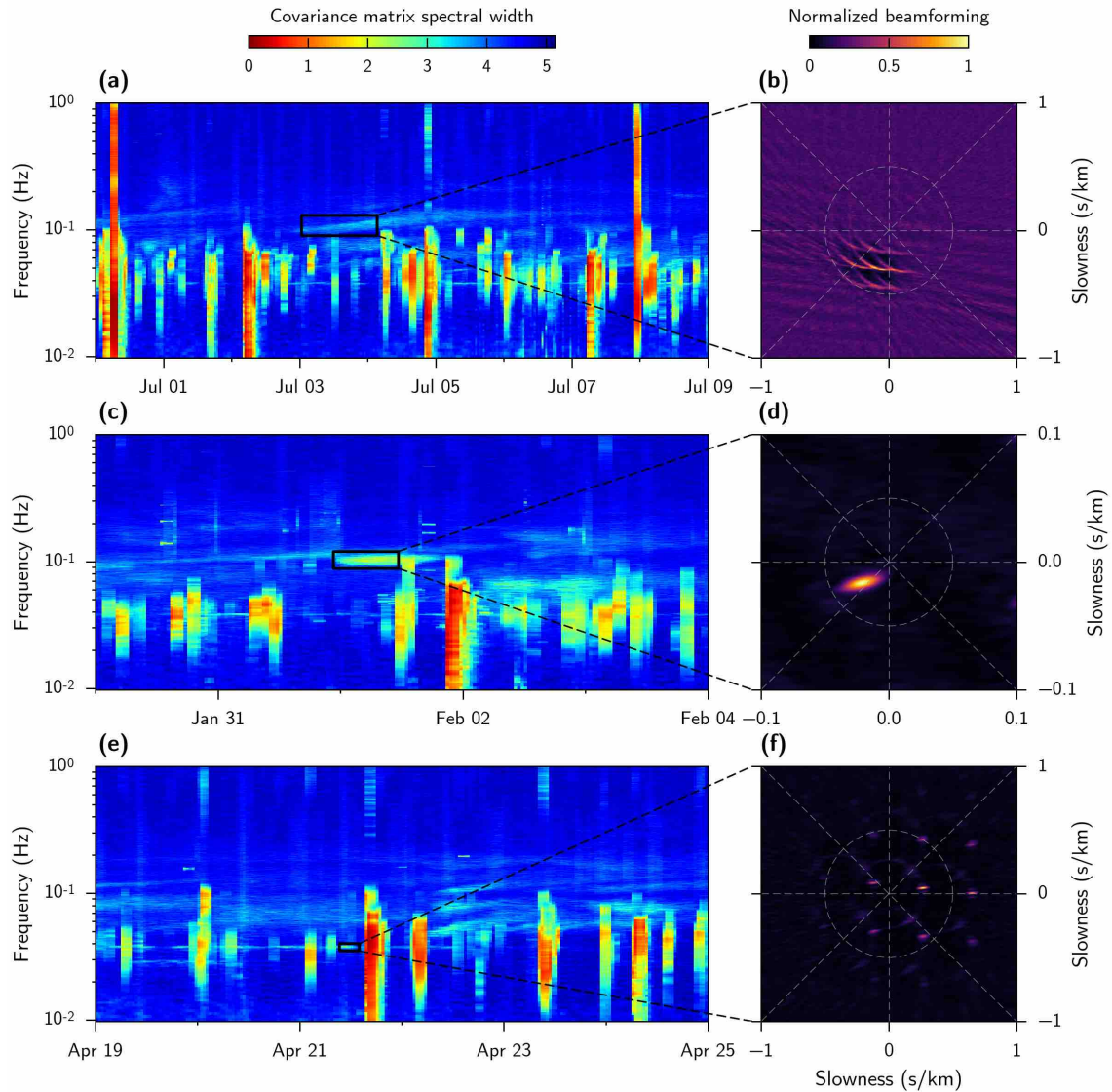


Figure 3.17 – **Left:** Zooms of the covariance matrix spectral width obtained from the pre-processed data (Fig. 3.15b). **Right:** Plane-wave beamforming analyses obtained from the average of the covariance matrix in the time and frequency limits delimited by the respective black frames on the left parts. (a) Dispersive signal analysed in (b) as Rayleigh waves incoming from between 175 and 205 degrees with a slowness of $\sim 0.34 \text{ s.km}^{-1}$ (c) Non-dispersive signal corresponding to (d) body waves incoming from about 230 degrees with a slowness of 0.0269 km.s^{-1} . (e) 5-hour long portion of the nearly-continuous and quasi-monochromatic signal, (f) composed of Rayleigh waves and incoming from 80 degrees of azimuth.

properties. As an example, we show in Fig. 3.17c the analysis of a signal recorded during February 1, 2010. This signal is very different from the microseisms generated by distant swells, i.e., it is not dispersive and its frequency range (0.09-0.12 Hz) remains constant. The plane-wave beamforming analysis of this signal (Fig. 3.17d) shows a clear maximum with a backazimuth of ~ 230 degrees and a slowness of $\sim 0.027 \text{ km.s}^{-1}$, corresponding to teleseismic P waves. A more detailed study of this event has been reported by [Retailleau et al. \(2015\)](#) who demonstrated that the P waves have been generated by the non-linear interaction of oceanic wave from a strong storm occurred in the Indian Ocean.

3.2.6 26-second microseism

Another remarkable feature seen in Fig. 3.15a is a nearly constantly present monochromatic signal (horizontal line) at frequencies about 0.034 Hz (nearly 26-second period). This signal corresponds to previously reported '26-second microseism' (e.g. [Holcomb 1980; 1998, Shapiro et al. 2006](#)). The beamforming analysis of a portion (5 hours) of this signal presented in Fig. 3.17f shows a typical aliased pattern corresponding to surface (Rayleigh waves) and its main lobe indicates a backazimuth of ~ 80 degrees and a slowness of $\sim 0.27 \text{ s.km}^{-1}$, which is in perfect agreement with the known source of this microseism in the Gulf of Guinea.

3.2.7 Conclusions

We studied the spatial coherence of the seismic wavefield recorded during one year (2010) by 121 stations of the Transportable component of the USArray. For this goal, we applied the recently developed method ([Seydoux et al. 2016a](#)) that analyzes the distribution of the eigenvalues of the array covariance matrix and use its 'spectral width' to characterize the level of coherence of the recorded wavefield. This method is attractive for its robustness to separate coherent information from the background seismic noise, and also because no prior knowledge on the underlying medium is required to be applied.

We first see that the raw record are strongly instationary in time and frequency (Fig. 3.15a). Since the contribution of the seismograms amplitudes is present in this first result, we cannot investigate it in term of spatial coherence. Yet, the pre-processing technique proposed in [Bensen et al. \(2007\)](#) mainly contributes to normalize the energy of individual records, leading to a reduction of its contribution in the second result (Fig. 3.15a). Hence, we can interpret this second result in terms of spatial coherence and therefore consider the strong heterogeneities visible in the raw records being due to some seismograms 'contaminated' by local noises, i.e., strong signals recorded by only a few closely located stations.

This normalization technique, widely used in the data pre-processing for computing noise cross-correlations (e.g. [Bensen et al. 2007](#)), was designed to reduce the contribution of strongly directional seismic sources that contribute to the anisotropy of the seismic wavefield. We show at the same time that it is not sufficient to fully homogenize the recorded wavefield and, even after applying the normalization, the covariance matrix (and by consequence the cross-correlations) contain clear imprints of many different sources. In particular, the influence of the strongest earthquakes on the cross-correlation of continuous records cannot be ignored. Also, multiple surface and body wave signals emitted by microseismic sources remain clearly visible.

An important consequence of the presented observations is that the cross-correlations of the continuous seismic records computed with the recently developed methods of the signal pre-processing cannot be considered as correlations of the equipartitioned wave-

field and are not exact proxies of the Earth's Green's functions (especially for amplitudes). Therefore, the inhomogeneous source distribution must be taken into account when interpreting these cross-correlations.

Chapitre 4

Enhancing cross-correlation travel-time measurements in presence of earthquakes: equalization of the covariance matrix spectrum

4.1	Introduction	80
4.2	Theory	82
4.2.1	Analogy between time-reversal and cross-correlation	82
4.2.2	Optimal focusing with spatio-temporal inverse filter.....	84
4.2.3	Equalization of the covariance matrix spectrum.....	85
4.2.4	Derivation of the eigenvalue threshold from the Vekua theory	86
4.3	US Transportable Array geometry	88
4.4	Illustration of the method with a synthetic test	89
4.5	Numerical experiment	91
4.5.1	Numerical setup	91
4.5.2	Protocol	92
4.5.3	Inter-station cross-correlations in presence of a strong source.....	96
4.5.4	Equalization of the covariance matrix spectrum.....	96
4.5.5	Travel-time measurement error, and travel-time inversion	98
4.6	Application to the 2010, Maule M8.8 earthquake	101
4.6.1	Wavefield anisotropy in presence of earthquakes.....	101
4.6.2	Equalization of the M8.8 Maule earthquake	104
4.6.3	Eigenvector-based surface waves selection	104
4.7	Discussion and conclusions	109

4.1 Introduction

Fifteen years ago, [Weaver and Lobkis \(2001a\)](#) demonstrated that the Green's function between two receivers recording a diffuse acoustic wavefield could be retrieved from the cross-correlation of the receivers' records. The diffuse acoustic wavefield is therein defined as the equipartitionned acoustic energy over modes, *i.e.*, where the mode amplitudes are uncorrelated (random) with equal variances. Since then, several papers presented other theoretical demonstrations. In [Derode et al. \(2003\)](#), the analogy between the cross-correlation of a diffuse wavefield and the time-reversal operation helps to demonstrate the link between the cross-correlation and the Green's function. Later, [Wapenaar \(2004\)](#) and [Wapenaar et al. \(2006\)](#) extended this theory to a general elastodynamic and inhomogeneous medium, when it is illuminated by many uncorrelated noise sources evenly distributed in the medium.

First applications of this theory in seismology ([Campillo 2003](#), [Shapiro and Campillo 2004](#), [Shapiro et al. 2005](#), [Sabra et al. 2005a](#)) and oceanography ([Roux et al. 2004](#), [Sabra et al. 2005c](#)) have led to a class of routinely used passive imaging methods in both domains. Ambient-noise based surface-wave tomography studies appeared at global scale using hum ([Nishida et al. 2009](#), [Haned et al. 2015](#)), at regional scale ([Shapiro et al. 2005](#), [Sabra et al. 2005b](#), [Behm et al. 2016](#)), at local scales (*e.g.* for the study of faults and volcanoes [Brenguier et al. 2007](#), [Roux 2009](#), [Jaxybulatov et al. 2014](#), [Lanza et al. 2016](#), [Hillers et al. 2016](#)) and in industrial contexts ([Mordret et al. 2014; 2013](#)). Body-waves can also be extracted from the cross-correlation of ambient noise at global scale ([Boué et al. 2013](#), [Lin et al. 2013](#)) and at local scale ([Nakata et al. 2015](#)). Moreover, several applications have been done in passive monitoring (known as 4-D tomography) of active zones such as faults ([Wegler and Sens-Schönfelder 2007](#), [Brenguier et al. 2008a](#), [Wu et al. 2016](#)) or volcanoes ([Sens-Schönfelder and Wegler 2006](#), [Duputel et al. 2009](#), [Brenguier et al. 2008b; 2014; 2015](#)).

Ambient-noise imaging requires the ambient seismic wavefield to be generated by uncorrelated sources evenly distributed in the medium ([Derode et al. 2003](#), [Wapenaar 2004](#)). Yet, this condition is often not verified: strong events such as earthquakes (which are localized in time and space) often produce highly coherent transient signals. Furthermore, the ambient microseismic noise ([Longuet-Higgins 1950](#)) is generated at specific regions of the Earth's surface, and shows a seasonal dependence due to its generating mechanisms (*e.g.* [Stehly et al. 2006](#)). While the primary microseisms are likely due to the direct interaction between oceanic swells and the seafloor in shallow water (*e.g.* [Traer et al. 2012](#), [Ermert et al. 2015](#)), the secondary microseisms are likely generated in deep ocean by the non-linear interaction of oceanic gravity waves, from which are generated pressure oscillations on the seafloor responsible for seismic waves generation (*e.g.* [Longuet-Higgins 1950](#), [Landès et al. 2010](#), [Bromirski et al. 2013](#)). Furthermore, localized and persistent sources may exist and contaminate the cross-correlations. This is the case of the 26-seconds microseismic source previously discussed in [Holcomb \(1980\)](#), [Shapiro et al. \(2006\)](#), [Xia et al. \(2013\)](#), [Seydoux et al. \(2016b\)](#). The effect of anisotropically distributed noise sources have been discussed in [Tsai \(2009\)](#), [Weaver et al. \(2009\)](#), [Froment et al. \(2010\)](#), [Kimman and Trampert \(2010\)](#), [Tian and Ritzwoller \(2015\)](#). They show that anisotropic seismic source distributions contribute to produce spurious arrivals or asymmetry of the cross-correlations functions, resulting in a bias of the travel-time measurements.

Pre-processing techniques therefore are usually applied to the seismic data to reduce

the effect of noise anisotropy and strong events influence. First, the data time-segments that contain strongly coherent signals are clipped above a given threshold or down-weighted (e.g. [Sabra et al. 2005b](#), [Hillers et al. 2014](#), [Wu et al. 2016](#)), or even rejected (e.g. [Rhie and Romanowicz 2004](#), [Shapiro et al. 2005](#), [Brenquier et al. 2007](#), [Gouédard et al. 2008](#), [Seats et al. 2011](#)), which results in a decrease in exploitable data. Moreover, the individual traces are often equalized in both spectral and temporal domains (see Fig.3.5 in chapter 3 of this thesis) as described in [Bensen et al. \(2007\)](#). This contributes to randomize the seismic wavefield, and attenuates the strong directional sources. Nevertheless, even if these pre-processing techniques strongly improve the seismic wavefield isotropy, some anisotropic features of the seismic wavefield may remain in the cross-correlations, as it has been recently discussed in [Seydoux et al. \(2016a\)](#), as well as in chapter 3. In particular, earthquakes and strong microseismic events generally resist to this amplitude equalization performed on individual traces.

More recently, array-based filtering techniques have been developed to enhance the recovery of Green's functions from the cross-correlations ([Gallot et al. 2012](#), [Leroy et al. 2012](#), [Carriere et al. 2013](#)). A passive inverse filter was proposed in [Gallot et al. \(2012\)](#) to enhance the isotropy of surface waves in the ambient seismic noise, and the results are compared with an active seismic experiment. In ocean acoustics, [Leroy et al. \(2012\)](#) proposed to improve the emergence rate of Green's function with a spatio-temporal filter built up from the first singular vectors of a reference matrix. Finally, [Carriere et al. \(2013\)](#) proposed a spatial-filter approach based on an angular filtering of volumic noise using least-square truncated pseudo-inversion of a steering matrix. This last study allowed to filter unwanted directive sources from the cross-correlation, such as shipping-related signals in underwater acoustics, but cannot be used automatically because the filter has to be designed regarding the source back-azimuth.

Our approach is inspired by [Gallot et al. \(2012\)](#) and [Carrière et al. \(2014\)](#), and is directly concerned with the covariance matrix computed from array seismic data. This is an advantage compared with [Gallot et al. \(2012\)](#) because it can be applied to a pre-existing cross-correlation database. We propose to extend the widely used spectral equalization process presented in [Bensen et al. \(2007\)](#), which is applied to the frequency spectrum of individual seismograms, to the equalization of the covariance matrix spectrum (*i.e.*, the ordered eigenvalues of the covariance matrix). Since the entries of the covariance matrix have been measured at different spatial locations, we interpret our approach as a *spatial* equalization process. This method allows correcting for the wavefield anisotropy in two fashions: (1) the influence of strong directive sources is substantially attenuated, and (2) the weakly excited modes are reinforced, allowing to partially recover the conditions that ensure a good-quality Green's function retrieval. It presents the advantage to be automatic, and does not require the knowledge of the source back-azimuth to be used, contrarily to [Carrière et al. \(2014\)](#).

As the covariance matrix is obtained from real data, uncorrelated noise may be present in the covariance matrix. It has been shown in many studies [Bienvenu and Kopp \(1980\)](#), [Wax and Kailath \(1985\)](#), [Gerstoft et al. \(2012\)](#) that the first eigenvalues of the covariance matrix are related to coherent information, whereas the last ones are more noise-related. Similarly to [Gallot et al. \(2012\)](#), [Carrière et al. \(2014\)](#), our technique requires to be performed on a low-rank approximation of the covariance matrix, in order not to consider uncorrelated noise in the equalization process, which could result in strong artifacts in the cross-correlation functions. In the present chapter, we show how the appropriate truncation can be derived from the number of degrees of freedom of the seismic wavefield

within the Vekua theory (Walsh and Vekua 1969).

Finally, we present an eigenvector-based spatial filter useful to distinguish between surface and body waves. This last filter is to be used together with the equalization of the eigenvalue spectrum. We apply it in the case of the equalization of earthquake-related signals when high-amplitude body-waves are present, and show that we can remove them from the cross-correlation functions.

The chapter is organized as follows: Section 4.2 presents the theoretical statements used to properly explain our approach. The equalization of the covariance matrix spectrum is defined therein, with the analogy of the passive inverse filter. Section 4.2 also presents the Vekua theory which is useful to theoretically derive the eigenvalue-threshold used in the equalization process. Section 4.3 presents the seismic array used in both numerical and real experiments. Because we partially investigated the wavefield isotropy with conventional beamforming, we selected a 34-elements square-shaped array from available stations of the USArray during 2010. The method is then illustrated with a synthetic example in section 4.4 where a source embedded in idealistic isotropic noise is equalized. In section 4.5, we investigate the effectiveness of our method with using two-dimensional numerical experiments with heterogeneous velocity model. We show that our technique allows to substantially reduce the impact of a strong seismic source embedded in isotropic seismic noise in the cross-correlations. In particular, we investigate the effectiveness of our approach. To that end, the travel-time measurement error remaining after having applied the method is quantified. Finally, we apply the technique to the data recorded USArray during 2010 in section 4.6. In particular, we select 30 days of data centered around the 2010 M8.8 Maule earthquake (Delouis et al. 2010), and show that the effect of the earthquake is strongly attenuated in the cross-correlations.

4.2 Theory

4.2.1 Green's function retrieval from ambient noise correlation: a time-reversal operation

Green's function retrieval from cross-correlation

It is a well-established theory (Weaver and Lobkis 2001a, Derode et al. 2003, Wapenaar 2004) that the cross-correlation $R_{ij}(\tau)$ of the ambient noise recorded at two sensors located at positions \mathbf{r}_i and \mathbf{r}_j leads to

$$\frac{dR_{ij}(\tau)}{d\tau} = G_{ij}(-\tau) - G_{ij}(\tau), \quad (4.1)$$

where τ is the correlation time and $G_{ij}(\tau)$ is the time-domain Green's function between the two sensors *i.e.* the signal that would be recorded at sensor j if an impulsive source were located at sensor i . The theorem presented in Eq. 4.1 holds when the noise-field is generated by uncorrelated sources evenly distributed in the medium (e.g. Wapenaar 2004). The cross-correlation between signals $u_i(t)$ and $u_j(t)$ recorded by two sensors i and j is obtained from

$$R_{ij}(\tau) \equiv \mathbb{E}[u_i(t)u_j(t - \tau)], \quad (4.2)$$

where \mathbb{E} denotes expected value. In practice, we estimate the correlation from a finite number of snapshots, as detailed later in the present chapter. Often, Eq. 4.2 is computed in the frequency domain, $R_{ij}(\tau) = \mathcal{F}^{-1}[C_{ij}(f)]$, from the array covariance matrix $C_{ij}(f)$ defined as

$$C_{ij}(f) \equiv \mathbb{E}[u_i(f)u_j^*(f)], \quad (4.3)$$

where $u_i(f) = \mathcal{F}[u_i(t)]$ is the Fourier-transformed seismogram recorded in station i and $*$ represents complex conjugation. In the following, the derivations are performed in the frequency domain to ease the reading of equations.

Analogy with the time-reversal operation

Many authors have described the spatial cross-correlation operation between two records of the noise-field generated by uncorrelated seismic sources as a time reversal operation (see for instance [Derode et al. 2003](#), [Larose 2004](#), [Campillo 2006](#)). We briefly explain this analogy in this section. Let us consider a discrete number of S sources generating arbitrary signals $e_s(t)$, where $s = 1 \dots S$. We know that the signal $u_i(t)$ recorded at the receiver i can be analytically expressed from the source-to-receiver Green's function collection $G_{is}(t)$ as

$$u_i(t) = \sum_{s=1}^S G_{is}(t) \otimes e_s(t), \quad (4.4)$$

where \otimes represents the convolution product. In the Fourier domain, and adopting a matrix representation, equation 4.4 simply reads

$$\mathbf{u}(f) = \mathbf{G}(f)\mathbf{e}(f), \quad (4.5)$$

where $\mathbf{u}(f)$ is the $N \times 1$ array data vector, $\mathbf{e}(f)$ is the source data vector and $\mathbf{G}(f)$ is the $S \times N$ matrix that includes all the source-to-receiver Green's functions. This matrix is also known as the *monochromatic propagation operator* ([Tanter et al. 2001](#)), and can be experimentally derived when the sources are under control. If we consider the sources to emit uncorrelated noise of equal variance $\varepsilon(f)$, such as, $\mathbb{E}[e_s(f)e_{s'}^*(f)] = \varepsilon^2(f)\delta_{ss'}$ then the covariance matrix of the received signals is given by

$$\mathbf{C}(f) = \mathbb{E}[\mathbf{u}(f)\mathbf{u}^\dagger(f)] = \mathbf{G}\mathbb{E}[\mathbf{e}\mathbf{e}^\dagger]\mathbf{G}^\dagger = \varepsilon^2(f)\mathbf{G}\mathbf{G}^\dagger, \quad (4.6)$$

where the operator $\mathbf{G}\mathbf{G}^\dagger = \sum_{s=1}^S G_{is}(f)G_{sj}^*(f)$ is also known to be the *time-reversal operator* ([Fink 1992](#)).

In the present case, the time-reversal operation virtually uses the seismic sources as a time-reversal mirror, as illustrated in Figure 4.1. First, the time-reversal experiment requires the knowledge of the Green's function $G_{is}(t)$ between the sensor i and the S sources. It is obtained by recording at the sources a pulse emitted by the sensor i . Then, the time-reversal operation is performed as follows: (a) the record at the sources are time-reversed and re-emitted from the sources, inducing (b) a focusing of the wavefield at \mathbf{r}_i , followed (c) by a diverging wavefront. The cross-correlation $R_{ij}(\tau)$ corresponds to the signal recorded by the sensor j during steps (a) to (c). We see that it is composed of a causal part (step b) part and an anti-causal part (step c). Because of the reciprocity of the wave equation, this experiment should lead to the same result while taking station j as focal spot.

It is well known that the time-reversal operator allows obtaining a good focusing at the original source located in \mathbf{r}_i if the sources are numerous enough, and if the time-reversal mirror (here, the set of sources) is surrounding the medium ([Tanter et al. 2001](#), [Derode et al. 2003](#)). Otherwise, the time-reversal operation will suffer from the lack of knowledge of the wavefield, and undesired side-lobes may appear at different locations because the focusing operation is not constrained in space in the time-reversal operation \mathbf{r}_i ([Tanter et al. 2000](#)). In passive seismology, such ill-conditioned time-reversal mirror corresponds to an uneven distribution of the noise seismic sources; the induced spurious

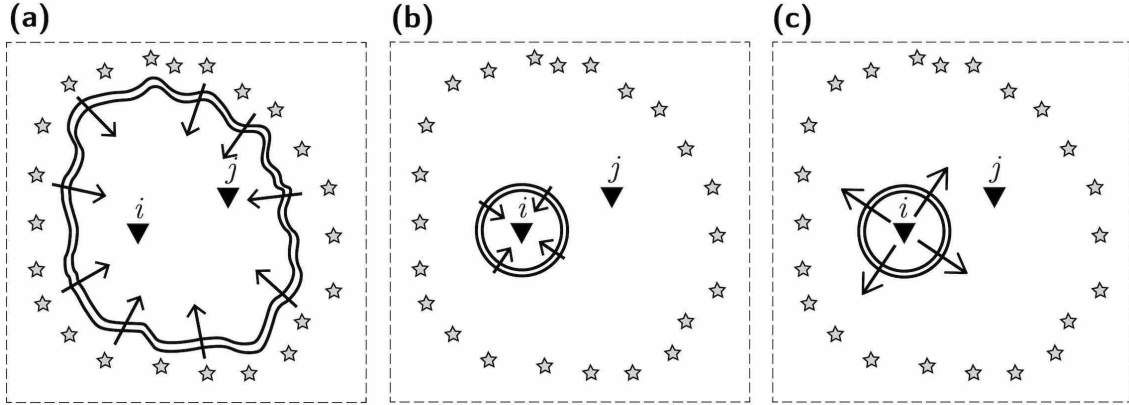


Figure 4.1 – Interpretation of cross-correlation process with time-reversal operation. The two sensors are indicated with inverted black triangles. The seismic sources are located on a surrounding curve at the surface. **(a)** The record of the pulse previously re-emitted by the station i is time-reversed and emitted by the sources. This is the basic principle of time-reversal operation. **(b)** waves are converging to the sensors i , and are about to focus at the position \mathbf{r}_i . **(c)** the wavefield then naturally diverges from the station i . The record obtained at station j during steps (a), (b) and (c) is the same as the one obtained with cross-correlation $R_{ij}(\tau)$ of ambient noise.

arrival or asymmetry of the cross-correlation functions are due to the degraded focus of the wavefield.

4.2.2 Optimal focusing with spatio-temporal inverse filter

In the context of finding an optimal focused beam in heterogeneous medium, [Tanter et al. \(2001\)](#) and [Aubry et al. \(2001\)](#) proposed a spatio-temporal inverse filter where a constraint on the focusing pattern is imposed. It consists in finding the source data vector $\mathbf{e}_0(f)$ to be emitted by the set of S sources to obtain a specific objective vector $\mathbf{u}_0(f)$ at the receivers. To obtain the optimally focused beam, they defined the objective vector as being 1 at the desired focusing receiver i_0 and 0 elsewhere:

$$\mathbf{u}_0(f) = \left(0 \cdots 0 \underset{\substack{\uparrow \\ i_0}}{1} 0 \cdots 0 \right)^T, \quad (4.7)$$

the signal $\mathbf{e}_0(f)$ to be emitted to obtain $\mathbf{u}_0(f)$ is derived from the inversion of Eq. 4.5. As the matrix \mathbf{G} is non-square, we compute the pseudo-inverse $\mathbf{G}^\# = (\mathbf{G}^\dagger \mathbf{G})^{-1} \mathbf{G}^\dagger$, leading to

$$\mathbf{e}_0(f) = \mathbf{G}^\#(f) \mathbf{u}_0(f). \quad (4.8)$$

In this last equation, $\mathbf{e}_0(f)$ is the signal to be emitted that is the most likely to produce $\mathbf{u}_0(f)$ in a least-square sense. The pseudo-inverse propagation matrix $\mathbf{G}^\#$ is computed from the singular-value decomposition of \mathbf{G} ([Golub and Kahan 1965](#))

$$\mathbf{G} = \mathbf{U} \mathbf{D} \mathbf{V}^\dagger \implies \mathbf{G}^\# = \mathbf{V} \mathbf{D}^\# \mathbf{U}^\dagger, \quad (4.9)$$

where \mathbf{U} is a $N \times N$ unitary matrix formed by the left-singular vectors, \mathbf{V} is a $S \times S$ unitary matrix that contains the right-singular vectors and \mathbf{D} is a $N \times S$ rectangular diagonal matrix of singular values. The inverse of the rectangular diagonal matrix \mathbf{D} is obtained by replacing the diagonal terms by their inverse. Because the received signals are experimentally acquired, some low-amplitude singular values related to the noise can be present,

likely to produce strong artifacts in the inversion. We therefore usually invert Eq. 4.8 for the low-rank approximation of \mathbf{G} (noted $\hat{\mathbf{G}}$), by keeping the L strongest singular values and with setting the following $N - L$ ones to zero. The corresponding low-rank pseudo-inverse propagation matrix is noted $\hat{\mathbf{G}}^\#$. Then, the most focused beam $\hat{\mathbf{u}}(f)$ is obtained with

$$\hat{\mathbf{u}}(f) = \mathbf{G}(f)\hat{\mathbf{G}}^\#(f)\mathbf{u}_0(f). \quad (4.10)$$

Here we observe that the time-reversal focusing operation $\mathbf{G}\mathbf{G}^\dagger$ has been replaced by the inverse filter operation, $\mathbf{G}\hat{\mathbf{G}}^\#$ which is expected to give an optimal focusing. This operator can also be interpreted as a projector into the signal subspace (Prada and Thomas 2003), where the signal subspace is formed by the signal-related singular vectors. This signal subspace is useful in many applications, such as high-resolution beamforming techniques (Wax et al. 1984, Goldstein and Archuleta 1987).

We can also express the optimally focused beam $\hat{\mathbf{u}}(f)$ with the singular values and vectors of the propagation operator with injecting equations 4.9 into equation 4.10:

$$\hat{\mathbf{u}}(f) = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger\mathbf{V}\hat{\mathbf{D}}^\#\mathbf{U}^\dagger\mathbf{u}_0(f) \quad (4.11)$$

$$= \mathbf{U}\hat{\Lambda}\mathbf{U}^\dagger\mathbf{u}_0(f), \quad (4.12)$$

because the unitary matrix \mathbf{V} leads to $\mathbf{V}^\dagger\mathbf{V} = \mathbf{I}$, and where the $N \times N$ matrix $\hat{\Lambda}$ is given by,

$$\hat{\Lambda} = \mathbf{D}\hat{\mathbf{D}}^\# = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & & & \vdots \\ \vdots & 0 & 1 & & & \vdots \\ \vdots & & & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix}. \quad (4.13)$$

We notice that only the left singular vectors appear in the inverse filter. This technique shows to substantially improve the focusing of elastic wave in a heterogeneous medium as described in Aubry et al. (2001), and is of particular interest when the time-reversal mirror does not surround the medium.

4.2.3 Equalization of the covariance matrix spectrum

We here expose our method of equalization of the covariance matrix spectrum and show that it is similar to the passive inverse filter proposed in Gallot et al. (2012) which is derived from the spatio-temporal inverse filter in the passive case where the sources are not under control.

The covariance matrix is hermitian by definition, we can therefore perform an eigenvalue decomposition on it, such as

$$\mathbf{C} = \Psi\Delta\Psi^\dagger, \quad (4.14)$$

where $\Delta = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ is the diagonal matrix formed by the eigenvalues λ_i and where Ψ is the matrix formed by the corresponding eigenvectors ψ_i . We propose to reconstruct the covariance matrix from its eigenvectors with the following equalized eigenvalues defined with

$$\hat{\Delta}_{ii} = \begin{cases} \lambda_i = 1 & \text{if } i \leq L, \\ \lambda_i = 0 & \text{otherwise.} \end{cases} \quad (4.15)$$

where L is a threshold above which we consider the eigenvectors to be noise-related only, similar to the inverse filter case. The threshold L thus corresponds to the low-rank approximation of the inverse filter technique previously mentioned. We derive theoretical expression for L in the next section. The equalized covariance matrix $\hat{\mathbf{C}}$ is then obtained from

$$\hat{\mathbf{C}} = \Psi \hat{\Delta} \Psi^\dagger \quad (4.16)$$

We can further interpret our technique as follows. Noticing that the covariance matrix can be expressed from the singular vectors of the time-reversal operator, (see Eqs. 4.6 and 4.9)

$$\mathbf{C} = \Psi \Delta \Psi^\dagger = \mathbf{U} \mathbf{D} \mathbf{V}^\dagger \mathbf{V} \mathbf{D}^\dagger \mathbf{U}^\dagger = \mathbf{U} \mathbf{D} \mathbf{D}^\dagger \mathbf{U}^\dagger, \quad (4.17)$$

one directly obtains $\Psi = \mathbf{U}$ and $\Delta = \Lambda$. Our method therefore is similar to the inverse filter directly applied to the array covariance matrix with

$$\hat{\mathbf{C}} = \mathbf{U} \hat{\Lambda} \mathbf{U}^\dagger \quad (4.18)$$

where $\hat{\Lambda}$ is defined from Eq. 4.13. This operation can also be seen as the extension of the spectral equalization process (Bensen et al. 2007) — which operates on the Fourier spectrum of the seismic traces — to the spatial case.

The generalization of the inverse filter to the passive case, where the sources are not under control, has been first done in Gallot et al. (2012) with seismic data. Instead of applying the singular value to the (unknown) monochromatic propagation operator, the authors exploit the definition of the covariance matrix $\hat{\mathbf{C}}$ estimated from real data

$$\hat{\mathbf{C}} = \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{u}_m \mathbf{u}_m^\dagger = \frac{1}{M} \mathbf{X} \mathbf{X}^\dagger \quad (4.19)$$

where \mathbf{X} is the $N \times M$ array data matrix built up from the array data vectors \mathbf{u}_m obtained from the m^{th} time-segment of duration δt (see Fig. 2.2 for more details). They then assume the array data matrix to be similar to the monochromatic propagation operator ($\mathbf{X} \sim \mathbf{G}$), with considering the noise-field to be generated by a significant number of overlapping Green's functions which are likely to be captured in each snapshot. The authors finally apply the pseudo-inversion described in Eq. 4.8 on the array data matrix. Despite showing similar results, it requires keeping the array data matrix, and perform the singular value decomposition on a $N \times M$ matrix where M is often large, inducing time-consuming computation.

4.2.4 Derivation of the eigenvalue threshold from the Vekua theory

We understand that the rank L (or eigenvalue-cutoff) of $\hat{\Lambda}$ has to be chosen carefully because it constitutes the single parameter of the filter. We here introduce the link between this rank and the degrees of freedom of the seismic wavefield. Our derivation shows similar results in the case of a 3-D noise geometry obtained within Carriere et al. (2013), and has the advantage to give also the eigenvalue-cutoff as a function of the frequency for the 2-D noise geometry.

Let consider the frequency-domain homogeneous wave equation (Helmholtz equation) given by $(\nabla^2 + k^2)u = 0$, where $k = 2\pi f\gamma$ is the wavenumber that depends on the frequency f and the slowness γ . In our case, the seismic wavefield is sampled on a typical spatial domain Ω covered by the seismic array. We know that in the two-dimensional case (which corresponds to Rayleigh waves, for instance), the Fourier-Bessel functions are a solution of the Helmholtz equation on the domain Ω . The vertical displacement $u_i(f)$ at point \mathbf{r}_i expressed in the polar basis as $\mathbf{r}_i = r_i\mathbf{e}_r + \theta_i\mathbf{e}_\theta$ of the domain can thus be expressed with

$$u_i(f) = \sum_{\ell=-\infty}^{+\infty} \alpha_\ell J_\ell(2\pi f\gamma r_i) e^{i\ell\theta_i} \quad (4.20)$$

where $J_\ell(x)$ are the ℓ -order Bessel functions of the first kind, and α_ℓ are the coefficients of decomposition of $u_i(f)$ onto the Fourier-Bessel basis. In the three-dimensional case, $\mathbf{r}_i = r_i\mathbf{e}_r + \theta_i\mathbf{e}_\theta + \phi_i\mathbf{e}_\phi$, the solution can be expressed with

$$u_i(f) = \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{\ell} \alpha_{\ell m} j_\ell(2\pi f\gamma r_i) Y_{\ell m}(\theta_i, \phi_i) \quad (4.21)$$

where $j_\ell(x)$ is the ℓ -order spherical Bessel functions of the first kind, and where $Y_{\ell m}$ are the spherical harmonics. In both 2-D and 3-D cases, the coefficients can be obtained analytically when the geometry of the domain Ω is simple (rectangular, elliptic) and when the values of the wavefield on the domain boundaries $\partial\Omega$ are known. When the geometry is not simple, it can only be obtained numerically.

In the Vekua theory (Walsh and Vekua 1969), we locally approximate the wavefield $u_i(f)$ at any point \mathbf{r}_i of the medium in the 2-D case with the following truncated approximation

$$u_i(f) \approx \sum_{\ell=-\mathcal{L}}^{\mathcal{L}} \alpha_\ell J_\ell(2\pi f\gamma r_i) e^{i\ell\theta_i}, \quad (4.22)$$

A similar approximation is obtained in the 3-D case with

$$u_i(f) \approx \sum_{\ell=0}^{\mathcal{L}} \sum_{m=-\ell}^{\ell} \alpha_{\ell m} j_\ell(2\pi f\gamma r_i) Y_{\ell m}(\theta_i, \phi_i). \quad (4.23)$$

The value of \mathcal{L} is obtained from Muiola et al. (2011)

$$\mathcal{L} = \lceil 2\pi f\gamma \bar{r} \rceil \quad (4.24)$$

where $\lceil \cdot \rceil$ stands for ceiling function. We see from Eqs. 4.22 and 4.23 that the number of degrees of freedom L of the wavefield differs in the 2-D and 3-D cases and are given by

$$\begin{cases} L^{2D} = 2\mathcal{L} + 1 \\ L^{3D} = (\mathcal{L} + 1)^2 \end{cases}, \quad (4.25)$$

and more specifically, one notices $L^{3D} > L^{2D}$. As \mathcal{L} is frequency-dependent, the truncation index can be explicitly written regarding the frequency

$$\begin{cases} L^{2D}(f) = 2\lceil 2\pi f\gamma \bar{r} \rceil + 1 \\ L^{3D}(f) = (\lceil 2\pi f\gamma \bar{r} \rceil + 1)^2 \end{cases} \quad (4.26)$$

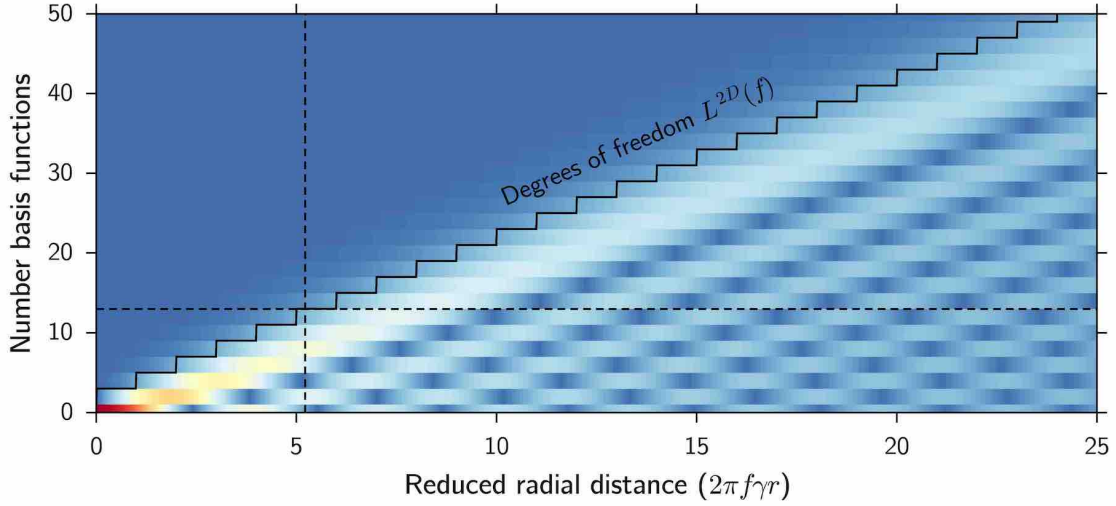


Figure 4.2 – Illustration of the Vekua theory for the 2-D case. The Bessel functions of the first kind $J_\ell(k\bar{r})$, where $\ell = -50 \dots 50$, are shown in colorscale. The number of degrees of freedom obtained from Eq. 4.26 is indicated in black solid line. The synthetic case presented in section 4.4 is highlighted with dashed lines.

The theory of Vekua can be understood more intuitively from the following in the 2-D case: if the measurement of the wavefield is performed on a limited region Ω with average inter-station spacing $\bar{r} = (N(N-1)/2)^{-1} \sum_i \sum_{j>i} r_{ij}$, then the Bessel functions of rank $\ell > \mathcal{L}$ do not contribute to the construction of the seismic wavefield because $J_{\ell>\mathcal{L}}(kr) \approx 0$ when $kr < k\bar{r}$. This is illustrated in Fig. 4.2, where the Bessel functions of order $-50 < \ell < 50$ are shown in colorscale (sorted with the absolute rank value $|\ell|$), regarding the reduced radial distance $2\pi f\gamma r$. The number of degrees of freedom L^{2D} is shown with black solid line. A similar argument can be obtained in the 3-D case.

The link with the covariance matrix rank is straightforward. If we note $E_{\ell i} = f_\ell$, where f_ℓ are the basis functions in the 2-D or 3-D case, then both Eqs. 4.22 and 4.23 lead to the following theoretical array data vector

$$\mathbf{u} = \mathbf{E}\mathbf{a}, \quad (4.27)$$

where \mathbf{a} contains the L coefficient of the decomposition. The matrix \mathbf{E} is a $L \times N$ rectangular matrix of rank $\min(L, N)$, indicating that the covariance matrix \mathbf{C} computed from \mathbf{u} is a square matrix rank $\min(L, N)$. We therefore use the frequency-dependent L as truncation index. We also notice that L is expressed with the slowness of the medium, which is unknown in the special case of seismic investigations. We nevertheless take an average slowness γ_0 to extend the theory to the heterogeneous case. A discussion of this choice is proposed at the end of the chapter.

4.3 US Transportable Array geometry

In the present chapter, we investigate the data continuously collected by the vertical channels of the USArray seismic stations during 2010. To observe the isotropy of the seismic wavefield, we use the conventional beamforming technique (e.g. Veen and Buckley 1988), which allows estimating the slowness and back-azimuth of the most coherent seismic waves in the seismic wavefield. The beamforming $B(f, \gamma)$ can be expressed as the projection of the covariance matrix $\mathbf{C}(f)$ onto plane-wave beamformers $\mathbf{b}(f, \gamma) = e^{-2i\pi f\gamma \cdot \mathbf{r}}$ as

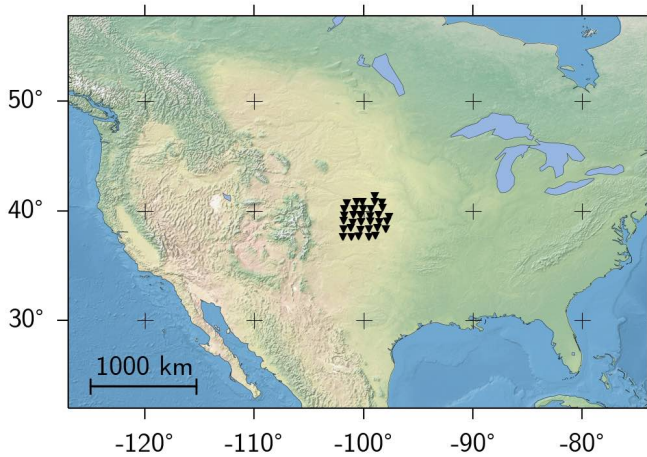


Figure 4.3 – Selected USArray stations. The location of the 34 seismic stations is shown as black triangles. The array aperture is about 500 km, and the typical inter-station spacing is 70 km. The stations were recording the seismic wavefield continuously during the year 2010.

following

$$\mathbf{B}(f) = \mathbf{b}^\dagger(f)\mathbf{C}(f)\mathbf{b}(f). \quad (4.28)$$

The plane-wave beamforming is obtained at each point of the spectrum and is represented regarding the 2-D apparent slowness vector $\boldsymbol{\gamma} = \gamma_x \mathbf{e}_x + \gamma_y \mathbf{e}_y$. The maxima of the beam power $|B(f, \boldsymbol{\gamma})|^2$ indicates the energy of the incoming waves as a function of the slowness and the azimuth encoded in $\boldsymbol{\gamma}$.

The beamforming resolution (size of the main-lobe) is inversely proportional to the array aperture (e.g. [Veen and Buckley 1988](#), [Rost 2002](#)). If the seismic array shape shows a larger extent in a specific direction (which is the case of the USArray full geometry used in Chapter 3, Fig. 3.14), then the slowness resolution of the conventional beamforming behaves differently regarding the azimuth. This is particularly inconvenient when our goal is to observe the isotropy of the seismic wavefield, that is, to observe the energy of the seismic waves in the slowness space regarding the azimuth.

We therefore select a square-shaped seismic array formed by 34 seismic stations of the USArray, shown in black triangles in figure 4.3. The typical inter-station spacing is 70 km and the spatial extent of the seismic array is about 500 km in both East-West and North-South directions. This induces a quasi-isotropic slowness resolution, visible in the beamforming pattern obtained from an idealistic 2-dimensional isotropic noise (Eq. 2.15) presented in Fig. 4.4d.

4.4 Illustration of the method with a synthetic test

This preliminary test aims at illustrating the method in the simple case where the medium is homogeneous with space-independent slowness $\gamma_0 = 0.25$ s/km, and where we fixed the frequency f_0 to 0.02 Hz in order not to induce aliasing artifacts in the beamforming output. The selected slowness is the typical Rayleigh wave slowness within the crust around this frequency.

First, we calculate the idealistic covariance matrix \mathbf{C}_{iso} directly from the theoretical expression of the 2-dimensional seismic noise

$$\mathbf{C}_{\text{iso}} = J_0(2\pi f_0 \boldsymbol{\gamma}). \quad (4.29)$$

The eigenvalues of \mathbf{C}_{iso} are presented in Fig. 4.4a. We observe that the eigenvalues are decaying steadily, and vanish with high eigenvalue indexes. This indicates that the number

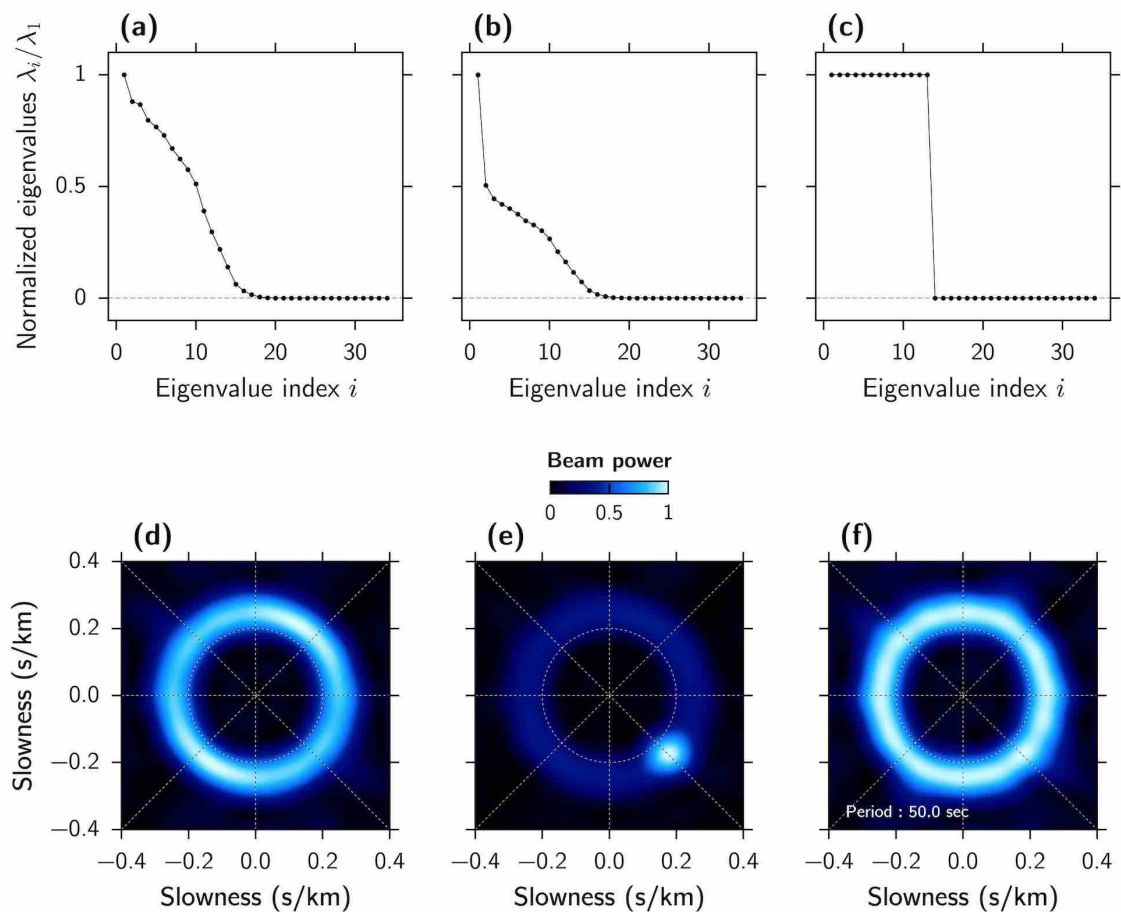


Figure 4.4 – Spatial equalization of a synthetic source embedded in idealistic 2-D isotropic seismic noise. The top panels show the covariance matrix eigenvalues, sorted in decreasing order. The corresponding beamforming is presented in each case at the bottom. They are obtained from the covariance matrix in each case with Eq. 4.28. We fixed the slowness $s_0 = 0.25$ s/km, and the period $T = 50$ seconds. **(a) and (d)** idealistic 2-D seismic noise obtained from Eq. 2.15. **(b) and (e)** idealistic 2-D noise plus a 10 times higher seismic source located at azimuth 135° . **(c) and (f)** case (b/e) where the covariance matrix is reconstructed from its first 11 eigenvectors with the corresponding eigenvalues set to 1.

of degrees of freedom of the seismic wavefield L^{2D} is smaller than the number of sensors N . Indeed, in this particular case, $L^{2D} = 2\lceil 2\pi f_0 \gamma_0 \bar{r} \rceil + 1 = 13$, which is in accordance with the obtained eigenvalue spectrum.

The corresponding beamforming obtained from \mathbf{C}_{iso} is presented in Fig. 4.4e. We observe that the energy is located near the slowness circle around $\gamma_0 = 0.25$ s/km, with an almost isotropic circular shape, indicating that the energy incomes from all azimuths with the same energy. The slight differences in amplitude around azimuths 45° and 225° are due to the non-uniform shape of the USArray selected stations.

We then calculate a second covariance matrix, where we embed a strong source in an idealistic 2-D isotropic noise, with an amplitude $a_s = 10$ times higher than the one from the seismic noise. Assuming that both the source and the seismic noise are independent, this second covariance matrix is obtained with

$$\mathbf{C}_{\text{src}} = \mathbf{C}_{\text{iso}} + a_s \mathbf{s} \mathbf{s}^\dagger, \quad (4.30)$$

where \mathbf{s} is the source-related array data vector such as $s_i = e^{2\pi f_0 \gamma_0 (\cos \theta x_i + \sin \theta y_i)}$, and θ is the source back-azimuth. We use $\theta = 135^\circ$ in this first test. The eigenvalue spectrum of \mathbf{C}_{src} is presented in Fig. 4.4b, where we clearly observe the strong dominance of the first eigenvalue, which is mainly related to the high-amplitude source. The corresponding beam power in Fig. 4.4f exhibits a high-amplitude spot located at azimuth 135° related to the source, and a background circular-shaped pattern related to the idealistic noise. This second covariance matrix aims at illustrating the case where the seismic data contain an earthquake-related signal, embedded in background seismic noise.

To reduce the influence of the source in the covariance matrix \mathbf{C}_{src} , we apply our spatial equalization method as follows. First, we get the eigenvalues λ_i and the corresponding eigenvectors ψ_i of \mathbf{C}_{src} . We then build the equalized covariance matrix from the truncated summation of the $L^{2D} = 13$ firsts eigenvectors:

$$\mathbf{C}_{\text{src, equal}} = \sum_{i=1}^{L^{2D}} \psi_i \psi_i^\dagger. \quad (4.31)$$

the L^{2D} first eigenvalues λ_i can be omitted in this last equation because they are all set to 1. The eigenvalues of $\mathbf{C}_{\text{src, equal}}$ are shown in Fig. 4.4c, where we simply observe that the 13 first eigenvalues are 1 and the other ones are null.

We see in Fig. 4.4f that the amplitude of the source and the background noise have been equalized, and the circular shape recovered. This simple test illustrates our equalization technique, with considering expected covariance matrices of idealistic noise and signal related to a strong far-field single source. One notices that the resulting beamforming is slightly degraded in Fig. 4.4f, with comparison with the original one in Fig. 4.4d. This indicates that small errors are induced with the equalization of the covariance matrix spectrum. These errors are investigated in terms of travel times measurements from the corresponding cross-correlation in the following section.

4.5 Numerical experiment

4.5.1 Numerical setup

We restricted our analysis to the vertical-component records, inducing that we observe only Rayleigh waves and body waves, the Love waves being polarized in the horizontal plane (see Figure 1.1). In this case study, we perform a 2-D acoustical simulation which

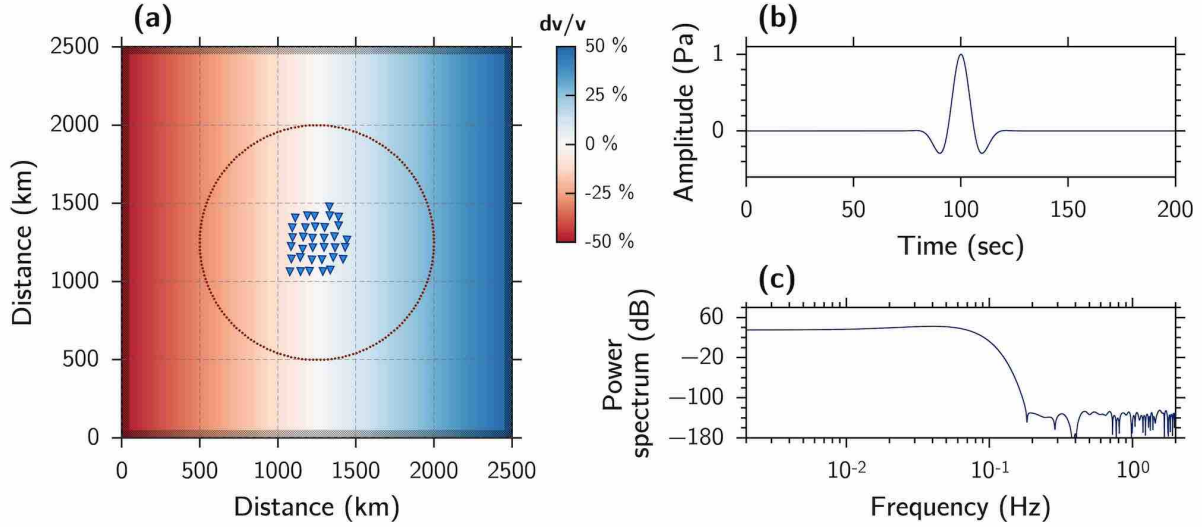


Figure 4.5 – Numerical setup of the 2D acoustic finite-differences simulation. (a) Location of the seismic stations (blue triangles) and of the seismic sources (red stars). The velocity model is represented in colorscale with the reference velocity $v_0 = 5$ km/s. (b) Ricker pulse $r(t)$ used to obtain the source-to-receiver numerical Green's function between the (see equation 4.32) (c) Ricker pulse Fourier transform $R(f)$ of the pulse signal $r(t)$ shown in (b). The cutoff frequency is around 0.1 Hz (see text for details).

considers pressure wave propagation. We therefore do not consider body waves in this simulation, and model only Rayleigh waves propagation. We considered two-dimensional finite-differences for solving the 2-D wave equation.

The numerical setup of the simulation is presented in Figure 4.5a. To use the geometry of the selected USArray seismic stations presented in Figure 4.3, we performed the simulation on a 2500×2500 km² grid. We chose a spatial grid step $d_x = 1.6$ km, and a time step $d_t = 0.065$ s to avoid numerical dispersion. The reflections of the pressure waves on the borders of the numerical domain are canceled out using 20 Perfectly Matched Layers (PML) derived from Liu (1997) and represented with hatching in the borders of the spatial domain in Figure 4.5a.

The velocity model is changing linearly in the x-direction from 2 km.s^{-1} to 6 km.s^{-1} . It is presented in colorscale in Figure 4.5 relatively to the average velocity $v_0 = 4$ km/s. The positions of the $N = 34$ seismic stations are shown with blue triangles. To simulate noise sources, we used a number of $S = 200$ seismic sources deployed around the seismic array (shown as red stars in Figure 4.5a). We consider this number of sources to ensure the sensor-to-sensor cross-correlations computed from numerical noise sources to have a correlation coefficient of more than 90% with the direct Green's functions when assuming sufficient averaging and equal source amplitude (see for instance Derode et al. 2003).

4.5.2 Protocol

Numerical Green's functions estimation

To estimate the numerical Green's function $G_{ab}(t)$ of the medium between two grid nodes a and b , we create a pulse-like signal $r(t)$ at a location a and consider the record of this pulse at any location b to consist in the Green's function. The pulse-like signal is chosen to be a Ricker wavelet (first proposed by Ricker 1953):

$$r(t) = (1 - (t - t_0)^2 \beta^2) e^{-(t - t_0)^2 \beta^2} \Leftrightarrow R(f) = R_0 f^2 e^{-f^2 / \beta^2} \quad (4.32)$$

where β is the frequency bandwidth, t_0 is the time shift of the wavelet, and R_0 is a constant. Eq. 4.32 indicates that both the temporal and frequency content of the pulse are second derivative of a Gaussian function, which is a symmetric function with zero-mean amplitude. Its half-height width is of $1/\beta$ in the temporal domain, and of β in the frequency domain. We performed the analysis between 100 and 10 seconds of period (0.01 and 0.1 Hz), so we chose $\beta = 0.1$ Hz. The temporal amplitude of the pulse is shown in Figure 4.5b where $t_0 = 100$ seconds, with its corresponding power spectrum in Figure 4.5c. We see that with our choice of parameters, the pulse has a flat spectrum from 0.01 to 0.1 Hz, and its energy vanishes beneath -150 dB at $f > 0.2$ Hz.

We estimate all numerical Green's functions $G_{is}(t)$ between each sensor $i = 1 \dots N$ and sources $s = 1 \dots S$. Because of the reciprocity of the wave equation, we notice that $G_{is}(t) = G_{si}(t)$. Therefore, we compute $G_{si}(t)$ with emitting the pulse from each seismic stations and record the wavefield at the sources to save computational time. (This allows to perform only 34 simulations against 200 if the pulse were sent from the seismic sources.) We also store the sensor-to-sensor ‘‘direct’’ Green's functions $G_{ij}(t)$ obtained from the records of the pulse emitted by a sensor i at a sensor j , to allow for comparison between the cross-correlation and the ‘‘direct’’ Green's functions.

Four selected snapshots of the pulse emission are shown in Fig. 4.6. The colorscale of the top panes represents the normalized pressure field. The bottom panels show the pressure recorded at the big red star location. The pulse is emitted from the most northerly seismic station. As we shifted the Ricker wavelet with $t_0 = 100$ seconds, half the pulse is emitted at time = 100.00 sec. The waves are then propagating towards all directions (see time = 312.00 sec), in a faster fashion towards the West, which is in good agreement with the numerical velocity model presented in Fig. 4.5a. The pressure recorded at the labeled seismic source indicates the passage of the pulse. At time = 312.00 sec, we have collected all the source-to-receiver impulse responses. The simulation ends at time = 532.67 sec. We see in the bottom plot how small are the residual reflections of the pressure waves on the border of the grid thanks to the PML.

Numerical inter-station cross-correlation functions

The numerical cross-correlation $R_{ij}(t)$ of the signals collected at sensors i and j can be analytically derived when the wavefield is generated by S seismic sources with known source-to-receiver Green's functions $G_{is}(t)$ and when the sources are emitting signals $n_s(t)$:

$$R_{ij}(t) = \mathbb{E} \left[\sum_{s,s'} (G_{is}(t) \otimes n_s(t)) \otimes (G_{js'}(-t) \otimes n_{s'}(-t)) \right], \quad (4.33)$$

where \otimes is the temporal convolution product. If in addition we consider the seismic sources to emit uncorrelated noise $n_s(t)$ of variances \mathbf{v} such as

$$\mathbb{E}[n_s(t) \otimes n_{s'}(-t)] = \mathbf{v}^2 \delta_{ss'}, \quad (4.34)$$

where $\delta_{ss'}$ is the Kronecker delta, and if the sources are evenly distributed on a surrounding curve around the seismic array, then the expected value of Eq. 4.33 reduces to (omitting the amplitude normalization factor)

$$R_{ij}(t) = \sum_{s=1}^S G_{is}(t) \otimes G_{js}(-t) = -G_{ij}(t) + G_{ij}(-t), \quad (4.35)$$

indicating that we retrieved Eq. 4.1. To save computational time, we compute the cross-correlation functions directly from Eq. 4.35, implicitly assuming uncorrelated noise sources

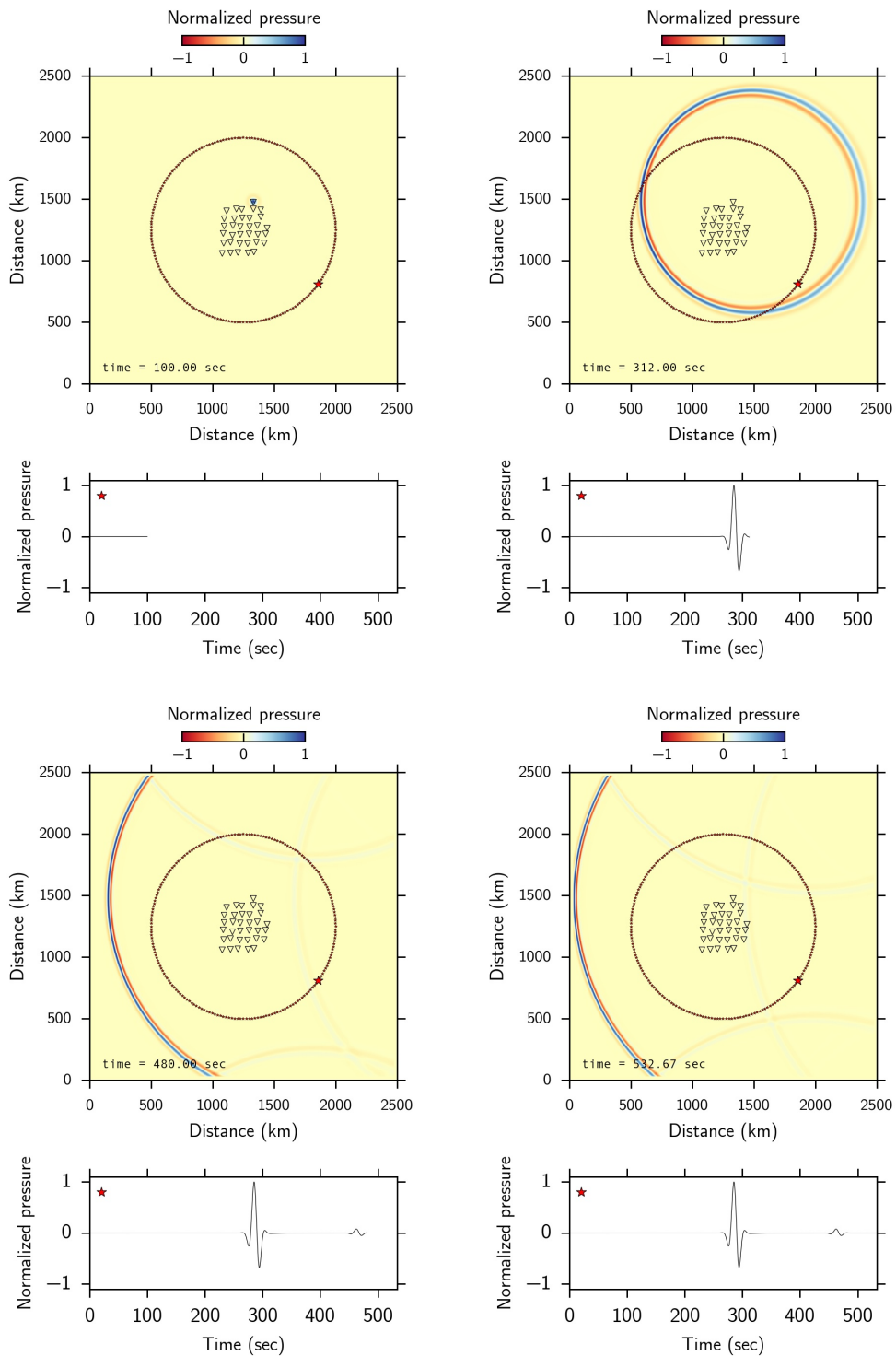


Figure 4.6 – Snapshots of the numerical estimation of the Green’s functions. The top panes show the normalized pressure field in colorscale, and the bottom panes represent the pressure recorded by the seismic source located at the big red star regarding the time. (time = 100.00 sec) the pulse is emitted from the most northerly seismic station. (time = 312.00 sec) the wavefronts are propagating towards all the directions, with a faster propagation speed towards the Eastern part of the domain (in accordance with the velocity model shown in Figure 4.5a). (time = 480.00 sec) The impulse response between the seismic sensor and all the sources have been recorded. (time = 532.67 sec) End of simulation. Notice the second wavefront arrival on the bottom panel due to the residual reflections on the grid borders.

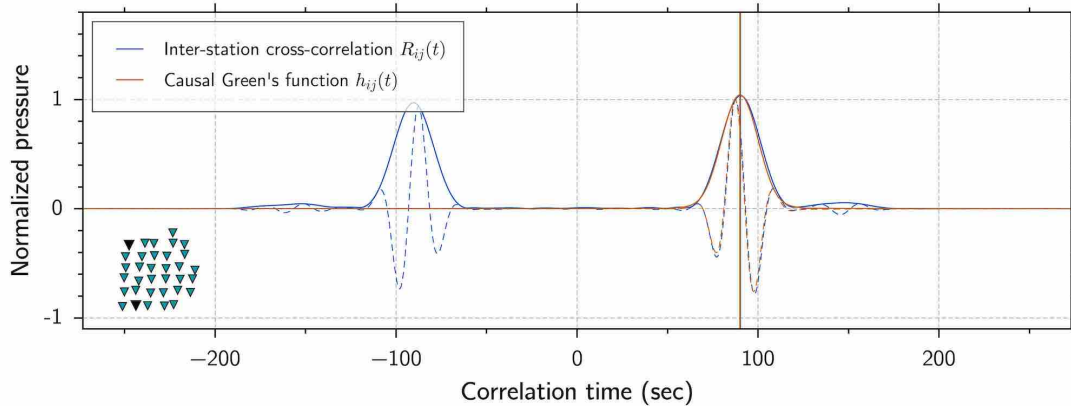


Figure 4.7 – Comparison between the direct causal Green’s function $G_{ij}(t)$ and inter-station cross-correlation $R_{ij}(t)$. The considered pair of stations i and j are shown with black triangles in the bottom-left inset. The direct Green’s function is shown with red dashed line, and the inter-station cross-correlation computed from Equation 4.35 is shown with blue dashed line. The seismic sources have the same amplitude, implying this simulated isotropic 2-dimensional seismic noise. The envelope computed from the Hilbert transform (Eq. 4.36) of the two signals are represented in solid line with corresponding colors. The travel-times corresponding to the maximum of the envelope are indicated with vertical solid lines.

and sufficient averaging.

Comparison between Green’s functions and inter-station cross-correlation functions

We compare the direct causal numerical Green’s function $G_{ij}(t)$ (red dashed line) and the inter-station cross-correlation $R_{ij}(t)$ (blue dashed line) in Fig. 4.7. The two functions are bandpass-filtered between 25 and 100 seconds of period. To compare the arrival times, we compute the envelope of the two functions from the Hilbert transformation. If we consider the signal $x(t)$, then the envelope $\tilde{x}(t)$ is defined as

$$\tilde{x}(t) = |x(t) + \iota \mathcal{H}x(t)|, \quad (4.36)$$

where $|\cdot|$ is the complex modulus, ι is the imaginary unit and \mathcal{H} denotes Hilbert transformation defined as $\mathcal{H}x(t) = (\pi)^{-1} \int_{-\infty}^{+\infty} x(\tau)(t - \tau)^{-1} d\tau$. The envelope’s maximum is used to extract the waves travel-times between the considered station pair. The selected sensor pair is shown with black triangle in the bottom-left inset of Fig. 4.7. We see that the causal part of the inter-station cross-correlation properly fits the causal Green’s function, with an error of 2.5% between the travel-times obtained in both cases. This comparison allows us to consider the inter-station cross-correlation equal to the Green’s function when the seismic sources are of equal amplitude and well-distributed regarding the azimuth.

The cross-correlations computed between all $34 \times (34 - 1)/2 = 561$ pairwise combinations of seismic stations are presented in colorscale in Fig. 4.8a sorted with the inter-station spacing and against time. We applied a narrow bandpass filter between 25 and 100 seconds (0.01 - 0.04 Hz) to the cross-correlation functions to allow for comparison of the results with beamforming analysis that is computed at a single point of the frequency spectrum. We observe the “V” shape formed by the cross-correlation functions, which is a typical expression of the good azimuthal coverage of the seismic sources with equal amplitudes because it induces symmetric inter-station cross-correlations functions. The cross-correlation function presented in blue line in Fig 4.7 is indicated with a black dashed line on Fig. 4.8a.

We perform the planewave-beamforming analysis of the covariance matrix obtained

from the inverse Fourier transform of the cross-correlation functions. The beam power obtained at $T = 54.6$ seconds (which approximately corresponds to the center¹ of the frequency band used to filter the cross-correlations in Fig. 4.8a) is presented in Fig. 4.8d. We observe the typical circular shape, also synonymous with evenly distributed equal-amplitude seismic sources (*i.e.*, it indicates the energy incomes from all directions with the same amplitude). Most of the beam power is nearby the average slowness circle (0.25 s/km). The slight amplitude asymmetry is due to the changing velocity in the medium.

4.5.3 Inter-station cross-correlations in presence of a strong source

We perform the same experiment as in the isotropic case, but we rewrite Eq. 4.34 as following

$$\mathbb{E}[n_s(t) \otimes n_{s'}(-t)] = \nu \delta_{ss'} + \nu_0 \delta_{ss_0}, \quad (4.37)$$

where s_0 is the source index of a strong seismic source of amplitude $\nu_0 \gg \nu$. This new relationship correspond to a strong seismic source of amplitude ν_0 embedded in isotropic seismic noise of variance ν . If the sources are evenly distributed, which is the case in the present simulation, the source index s_0 is simply related to a given azimuth θ_0 .

We compute the inter-station cross-correlation functions with a source located at azimuth $\theta_0 = 158^\circ$ (corresponding to the one from the 2010 M8.8 Maule earthquake observed from the USArray, see Fig 4.12b), and with an amplitude ten times higher than the background isotropic noise (*i.e.*, $\nu_0 = 10\nu$). The results are presented in Fig. 4.8b, in colorscale, sorted with the inter-station spacing and regarding time. We see that the “V” shape related to the background isotropic noise is contaminated with many spurious waves arrivals with smaller travel-times. These spurious peaks are due to the strong source, and their travel-times are shifted relatively to the azimuth between the inter-station axis and the direction-of-arrival of the wave θ_0 . The selected cross-correlation function indicated with a dashed line in Fig. 4.8b is presented with red color in Fig. 4.9. We observe that the maximum of the envelope does not correspond to the one obtained in the homogeneous case (also indicated in the Figure with a blue solid line).

The corresponding beamforming analysis is presented in Fig. 4.8e. We clearly observe the presence of the strong seismic source since most of the energy is incoming from azimuth 158° . The isotropic seismic noise is still visible in the background, with the typical circular shape with amplitude 10 times smaller than the main spot.

4.5.4 Equalization of the covariance matrix spectrum

We then apply our equalization technique to this last covariance matrix, built up from isotropic noise plus a strong source. We perform the eigen-decomposition at each frequency, and reconstruct the covariance matrix from the frequency-dependent number of equalized eigenvectors as described by Eqs. 4.18 and 4.26. The average radius \bar{r} of the selected stations is obtained from the average of all pairwise-combination inter-station spacing, and is about 166 km. We used γ_0 equal to the average velocity in the medium 0.25 s/km in this approach.

The results obtained from the filtered covariance matrix are presented in Fig. 4.8c, and the corresponding beamforming in Fig. 4.8f. We observe that the symmetry of the cross-correlation functions is strongly improved, suggesting that the effect of the source have

1. the real center actually is 33 seconds, but we chose 54.6 seconds in order not to have aliasing artifacts that contaminate the visibility of the results, without any loss of generality

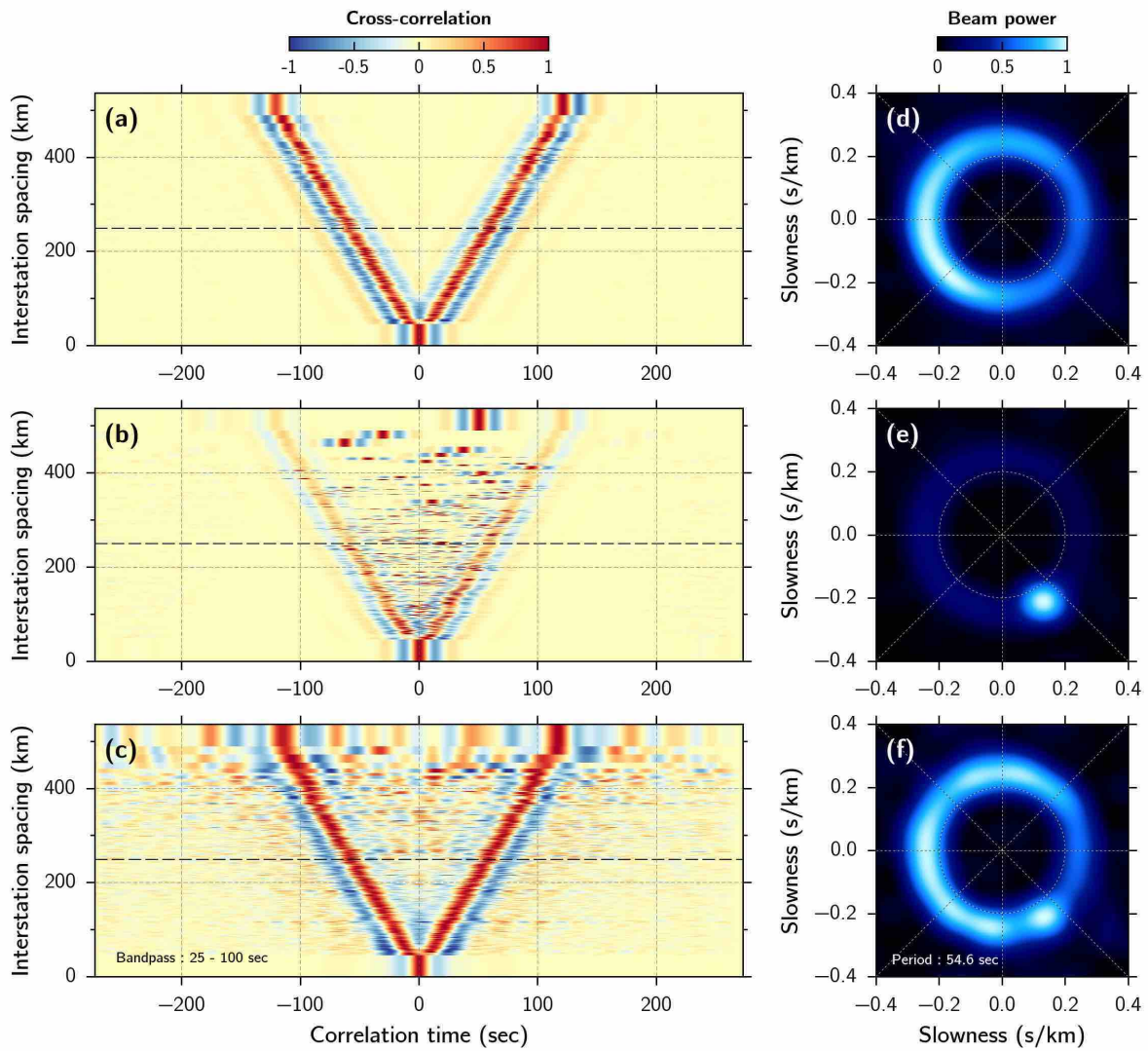


Figure 4.8 – Numerical inter-station cross-correlation functions (left) and corresponding plane wave beamforming (right). The cross-correlation functions are computed from Eq. 4.35, and are bandpass-filtered between 25 and 100 seconds. The planewave-beamforming analysis is performed at $T = 54.6$ seconds. **(a)** Isotropic case, where the seismic sources are evenly distributed regarding the azimuth and have equal amplitude. **(b)** Anisotropic case, where all sources have the same amplitude except one located at azimuth 158° (corresponding to the Maule earthquake azimuth, see Fig. 4.12). This case simulates a strong source embedded in ambient and isotropic seismic noise. **(c)** Equalized case, obtained from the covariance matrix (b) equalized with Eq. Cases **(d-f)** show the beamforming analyses respectively corresponding to cases (a-c).

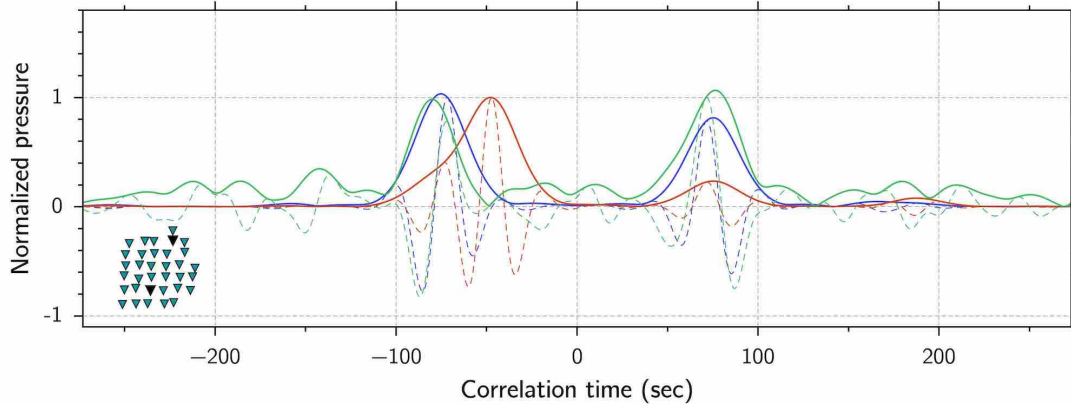


Figure 4.9 – Cross-correlation functions obtained in three different source configurations. The considered pair of stations i and j are shown with black triangles in the bottom-left inset. The corresponding (blue) the sources have the same amplitude (red) one of the sources has an amplitude 10 times higher than the other ones (green) similar than (b) after equalization of the covariance matrix spectrum (Eq. 4.18). The envelopes are indicated in thick solid lines, and the correlation function in thin dashed lines.

been substantially attenuated. The spatial equalization introduces artifacts, particularly for the cross-correlations with large inter-station spacing (above 400 km).

The corresponding beamforming shows a clear improvement in comparison with the raw result shown in Fig. 4.8e. The circular shape is greatly improved, indicating that we reinforce the weak isotropic background noise. The presence of the seismic source is still visible at the azimuth $\theta_0 = 158^\circ$, meaning that the source has not been removed, in contrast with the method presented in Carriere et al. (2013), but its energy has been equalized.

The correlation function that corresponds to the dashed line in Fig. 4.8c is presented in green color in Fig. 4.9. The envelope of the filtered correlation is closer to the reference case in blue line. Some amplitude differences are visible, and the travel-times are shifted with a small value compared with the reference case.

4.5.5 Travel-time measurement error, and travel-time inversion

We finally investigate the overall error in travel-time measurements in Fig. 4.10, within two cases (a) when the wavefield is contaminated with the strong-amplitude source and (b) with the equalization of the eigenvalues applied to (a). The travel-time measurements are obtained from the time-lag corresponding to the maximum of the envelope (see Eq. 4.36), and with taking the absolute value to get positive travel-times. The reference travel-times are obtained from the isotropic cross-correlation functions.

The overall travel-time measurement error ε is obtained in both cases from

$$\varepsilon = \sum_{n=1}^{N_c} \frac{|tt_n - tt_{ref,n}|}{tt_{ref,n}} \quad (4.38)$$

where tt_n is the travel-time of the n th cross-correlation function, $tt_{ref,n}$ is the reference travel-time for the n th reference cross-correlation, and $N_c = N(N-1)/2$ is the total number of cross-correlation functions.

The results are presented in Fig. 4.10. Fig. 4.10a shows that the strong source induce an overall error in the travel-time measurement of 37.7%, with travel-times spread out between 0 and $tt_{ref,n}$. This overall error is reduced to 6.5% after equalizing the covariance

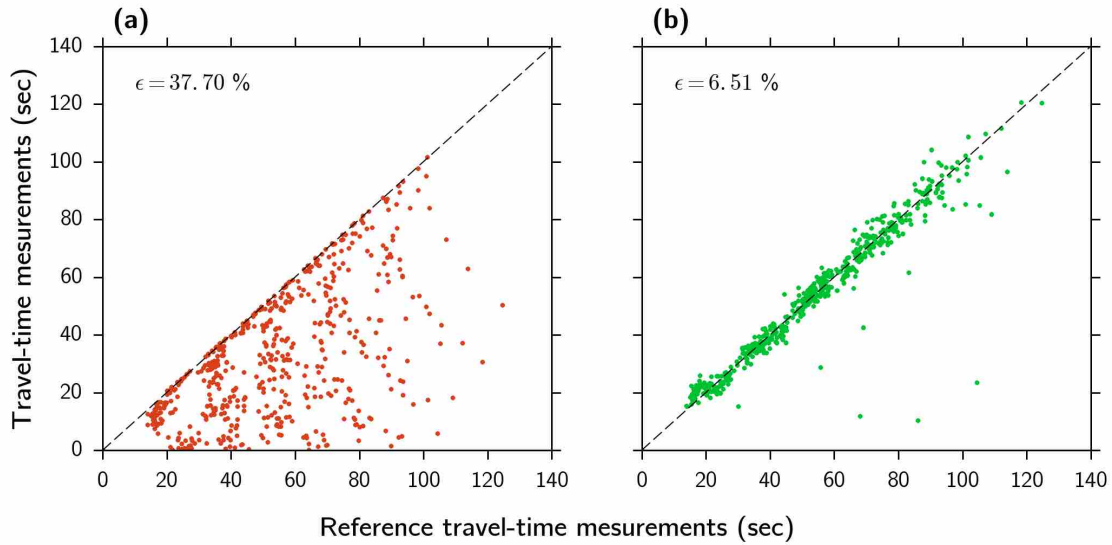


Figure 4.10 – Travel-times measurement from all the cross-correlation functions compared with the reference travel-times. The travel-times are measured from the maximum of the envelopes of the cross-correlation functions. The reference travel-times correspond to those obtained in the isotropic case presented in Fig. 4.8a. **(a)** one of the sources has an amplitude 10 times higher than the other ones **(b)** similar than (b) after equalizing the covariance matrix spectrum (Eq. 4.18). The overall error in travel-times measurement is indicated in each case.

matrix spectrum in Fig. 4.10b, where the travel-times are, indeed, closer to the expected ones.

We present the results of the inversion of the travel-times to recover the velocity model in Fig. 4.11 where (a) is obtained in the isotropic case, (b) corresponds to the isotropic case + a strong seismic source and (c) is the case (b) where we equalized the covariance matrix spectrum. The input velocity model is reminded in Fig. 4.11e. The inversion is performed using the travel-time measurements between all inter-station cross-correlation function which paths are shown in Fig. 4.11d, using the open-source Python package Fatiando (Uieda et al. 2013). We performed the inversion on a 136×136 grid.

We see that the velocity model obtained in the isotropic case (Fig. 4.11a) is in good agreement with the input model. There are no path coverage outside the box between $x, y \in [1100, 1400]$ which explains why the inversion leads to underestimated velocities on the border of the grid. As expected, the case where a strong source is located in $\theta_0 = 158^\circ$ lead to strong bias in the travel-time measurements, which induce strong anomalies in the velocity inversion. The colorscale is fixed, so the results in this case are saturated. Finally, the equalization of the covariance matrix spectrum allows to strongly improve the velocity model exactness, and we see that it is close to the one obtained in the isotropic case. Some anomalies nevertheless subsist, which are related to the errors in the travel-times measurements shown in Fig. 4.10b.

The method therefore shows good improvements of travel-time measurements in presence of strong sources, that are herein interpreted as strong anisotropic features in the seismic wavefield. The methods show up to work under assumption that the sources are evenly distributed around the studying zone, but not necessarily with equal amplitudes. This constitutes the assumption we need to make in the real case.

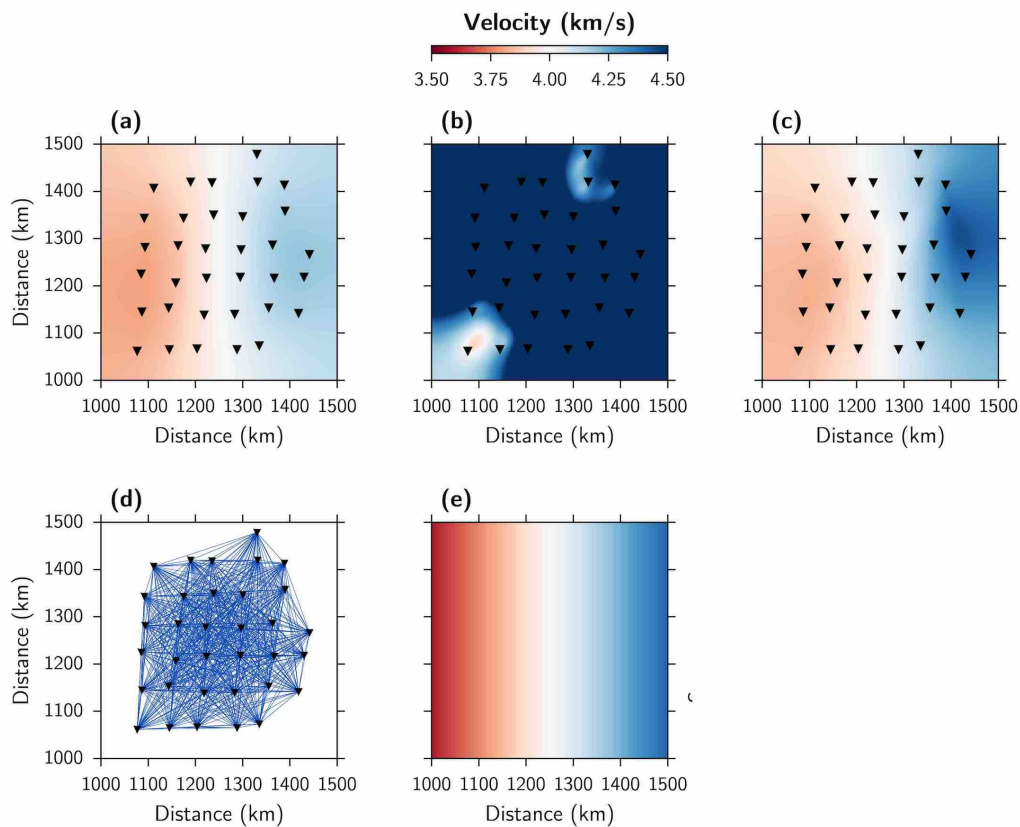


Figure 4.11 – Inversion of the travel-times measured from the numerical cross-correlation function presented in Fig. 4.8a, b and c. (a, b, c) tomographic results respectively obtained in (a) an isotropic source configuration, (b) with a strong source illuminating from $\theta_0 = 158^\circ$ and (c) from the spatial equalization applied on (b). **(d)** Inter-station paths. **(e)** Input velocity model.

4.6 Application to the data collected by the USArray: equalization of the 2010, M8.8 Maule earthquake

4.6.1 Wavefield anisotropy in presence of earthquakes

Earthquakes of 2010

We show the magnitudes of $M \geq 5$ earthquakes as black dots regarding the time in figure 4.12a. We narrowed the focus on three 30-days long time-segments of data indicated with colored boxes in Figure 4.12a. The yellow box contains earthquakes of magnitude smaller than 6.3, with epicenter locations shown as yellow dots in the azimuthal world map in Figure 4.12b. The red and blue windows are centered around two major earthquakes of the year 2010, respectively the M8.8 Maule, Chile megathrust earthquake (e.g. [Delouis et al. 2010](#)), and the Moro Gulf, Mindanao, Philippines series of three earthquakes (M7.3, 7.6 and 7.4 [USGS 2010](#)) which we will refer as the M7.6 earthquake in the present study.

Estimation of the covariance matrix from the data collected by the USArray

The covariance matrix $\mathbf{C}(t, f)$ is estimated from the USArray data using equation 4.19 (see also Fig. 2.2 for more details). The duration δt of the sub-windows is 960 seconds, and we used a number of $M = 20$ half-overlapping sub-windows (*i.e.* the covariance matrix is obtained within $\Delta t = 9600$ -seconds long time-segments of data). Before computation, we pre-process the data over each 9600-sec long windows with temporal and spectral smoothing techniques described in [Bensen et al. \(2007\)](#), and detailed in see Section 3.1.3.

We then stack it over 24-hours to obtain daily estimates $\mathbf{C}_d(t, f)$. The effective number of sub-windows is then $M_{\text{eff}} \approx 4 \times 24 \times 3600 / \Delta t = 720$, the covariance matrix is thus robustly estimated on each day because $M_{\text{eff}} \gg N$, where N is the number of seismic stations (e.g. [Menon et al. 2012a](#)).

Beamforming analysis

To investigate the wavefield isotropy of the three time-segments of data presented in Figure 4.12a, we stack the daily covariance matrices obtained in each period of time. The planewave beamforming analysis is then performed onto the stacked covariance matrices. We show in Figure 4.13 the beamforming obtained within the yellow window (Figure 4.13a), red window (Figure 4.13b) and blue window (Figure 4.13c).

The influence of the epicenter repartition is clearly visible as for Fig. 4.13(a). Most of the energy is incoming from the East (between -135° and -45°), which corresponds to the Pacific subduction region where most of the epicenters are located during the considered time. We also see that the wavefield is energetic between 110° and 180° . This is in agreement with the location of epicenters located in South America, and especially those located in the subduction zone. The wavefield is mostly composed of Rayleigh waves because most of the energy is located near the 0.25 s/km slowness circle.

Figure 4.13b is obtained from the time-segment when the Maule M8.8 earthquake occurred. Even if the covariance matrix is stacked over 30 days including 15 days before the earthquake nucleation, we observe a strong dominance of Rayleigh waves incoming from the earthquakes (the waves are incoming with an apparent slowness of 0.25 s/km and with about 160° of azimuth). We also observe the presence of body waves with apparent slowness close to 0.

Fig. 4.13c shows a similar result obtained around the Moro Gulf M7.6 earthquake. The dominance of Rayleigh waves incoming from the earthquake is also clearly visible

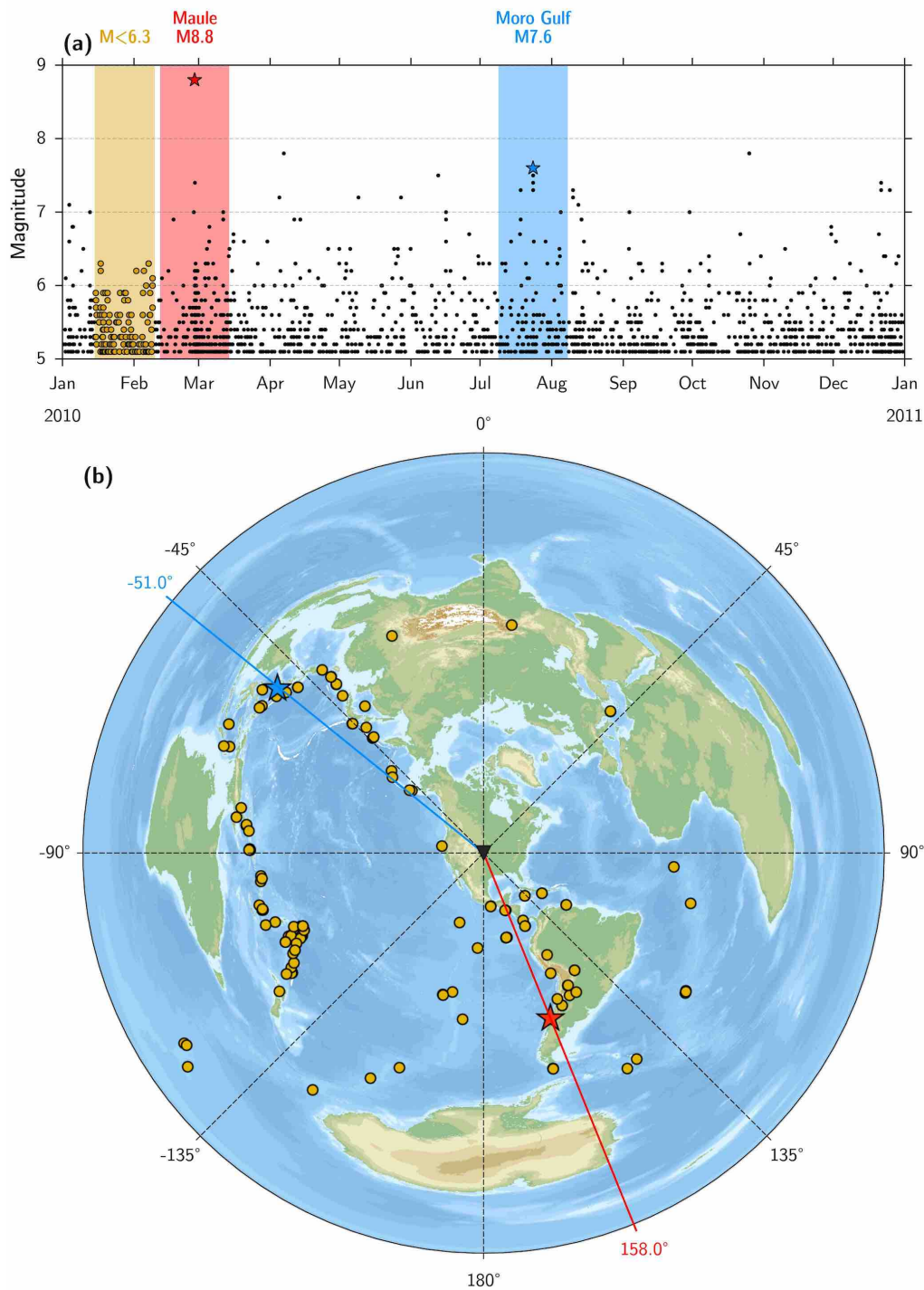


Figure 4.12 – U.S.G.S. Earthquake catalog of 2010 **(a)** Magnitude and date of $M \geq 5$ earthquakes that occurred in 2010, obtained from the U.S. Geological Survey database <http://earthquake.usgs.gov/earthquakes/search/>. The colored boxes indicate three time-segments (30-days long each) over which we analyzed the wavefield isotropy. The yellow box is free from earthquakes of magnitude higher than 6.3. The corresponding epicenter locations are shown with yellow dots in the azimuthal map **(b)**. The red and blue boxes contain two major earthquakes of 2010: the Maule M8.8 megathrust earthquake and the Moro Gulf M7.6 earthquake. Their corresponding epicenter locations are shown with identical symbol in **(b)**. **(b)** Azimuthal global map centered on the USArray barycenter (black triangle). We indicate the direction of arrival of Rayleigh waves incoming from the Maule earthquake (red line) and the Moro Gulf earthquake (blue line).

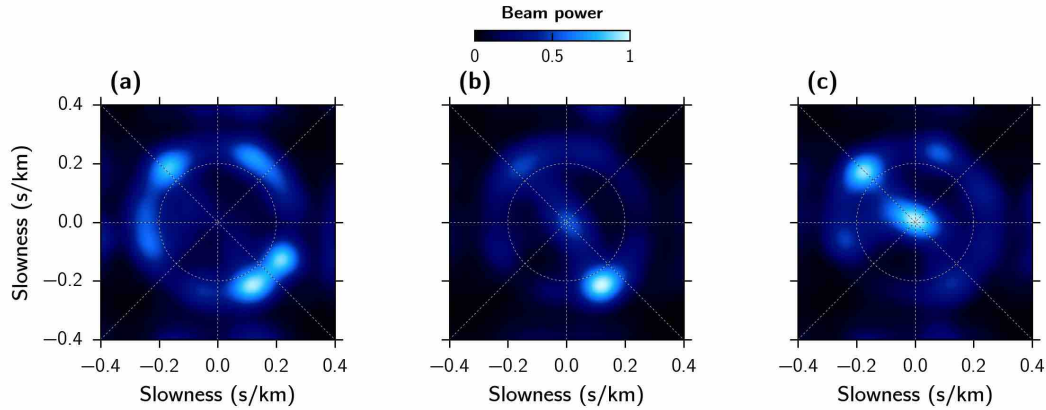


Figure 4.13 – Beamforming computed over the three windows of data shown in Figure 4.12. The data is pre-processed with temporal and spectral equalization before computation. The beamforming’s output is averaged over the 40 and 50 seconds of period. **(a)** Yellow window **(b)** Red window **(c)** Blue window

(the waves are incoming with an apparent slowness of 0.25 s/km and with about -45° of azimuth). In this case, we observe a stronger amplitude of body waves with apparent slowness close to 0.

These three results were obtained from the covariance matrix estimated from the data pre-processed with the classical temporal and spectral equalization. We see that, in agreement with the results obtained in Chapter 3, the equalization process does not allow to fully isotropify the seismic wavefield, and the uneven distribution of the seismic sources (Fig. 4.13a) or the presence of spatially localized coherent sources (Fig. 4.13b and c) remain clearly encoded in the covariance matrix.

Standard cross-correlation

The specific case of the Maule earthquake is herein further investigated. We compute the cross-correlations functions between all pairs of selected stations from the inverse Fourier transform of the covariance matrix, and present them in Figure 4.3, as a function of time and sorted regarding the inter-station spacing. The cross-correlation functions are bandpass-filtered between 25 and 100 seconds, and are shown in colorscale in Fig. 4.14a. The maximal inter-station spacing is about 525 km, and the correlation time is shown between -300 and 300 seconds. Furthermore, the corresponding beamforming is presented in Fig. 4.14c, averaged between 41 and 51.2 seconds of period.

We can distinguish between several features in this first result. First, we observe in the background the typical “V” shape formed by the Rayleigh waves due to the ambient seismic noise. The same observation can be done from the corresponding beamforming, with the low-amplitude circular shape at slowness 0.25 s/km in Fig. 4.14c. This is an important observation because it allows to consider the seismic wavefield to be generated by seismic sources with a quasi total azimuthal coverage, with different amplitudes, which a required condition to apply our method. A second important feature is the strong peak present in many cross-correlations near the zero time-lag. This peak can be both related to (1) the spurious arrivals due to the Rayleigh waves incoming from the Maule earthquake at the pairs of stations which axis is orthogonal to the earthquake azimuth, and (2) to the body waves that income from the earthquake, also responsible for the low-energy peak visible in the beamforming, at apparent slowness close to 0.

4.6.2 Equalization of the M8.8 Maule earthquake

We apply our method onto daily covariance matrix, and we then stack the filtered covariance matrices, allowing to increase the performances of the method. Even if the Rayleigh waves are known to be dispersive within the Earth, we use a constant slowness $\gamma_0 = 0.25$ s/km (observed on the beamforming at $T \in [40, 50]$ seconds of period) to define the eigenvalue-cutoff $L(f)$ from Eq. 4.25. A more precise result should be obtained with considering the typical Rayleigh waves slowness regarding the frequency (dispersive approach). As we want to retrieve the travel-times of Rayleigh waves, we choose $L(f) = L^{2D}(f)$.

The inter-station cross-correlations obtained from the equalized covariance matrix is presented in Fig. 4.14b, and the corresponding beamforming is shown in Fig. 4.14d. We clearly observe that the equalization process has increased the energy of the background noise. Some azimuthal regions are still not excited (around azimuth -135° and near 90°), because of the lack of seismic energy from these directions. Furthermore, we see that the strong spot of energy related to the Maule earthquake Rayleigh wave is still visible in the beamforming, with an amplitude quasi-equal to the background noise. This is in accordance with the numerical result presented in Section 4.5, see Fig. 4.8c.

Finally, we see that a continuum of energy had been reinforced in Fig. 4.14d, with a pattern oriented with the earthquake azimuth. Because the beamforming only gives access to an apparent slowness, the waves related to this pattern are necessarily body waves, likely generated by the Maule earthquake. These body waves have been equalized by our method, because of its inability to distinguish between 2-dimensional and 3-dimensional wavefields. Indeed, we choose the eigenvalue $L(f) = L^{2D}(f)$, but as mentioned in Sec. 4.2, $L^{3D} > L^{2D}(f)$. We therefore see that from its fundamental definition (see Eqs. 4.22 and 4.23) the equalization of the covariance matrix spectrum as well as the passive inverse filter approach can reinforce body waves if they are present in the wavefield, and if they are spanned by eigenvectors ψ_i with index $i < L^{2D}(f)$.

4.6.3 Eigenvector-based surface waves selection

To remove the influence of the unwanted body waves observed in Fig. 4.14d after having equalized the covariance matrix spectrum, we propose to select surface waves from the covariance matrix eigenvectors. The selection can be done with the beamforming analysis, applied to the covariance matrix eigenvectors ψ_n , with rejecting those that contain energy inside an arbitrary slowness threshold γ_{hr} . The beamforming \mathbf{B}_n applied to the n^{th} eigenvector where $n = 1 \dots N$ is the eigenvector index is given by

$$\mathbf{B}_n = \mathbf{b}^\dagger \psi_n \psi_n^\dagger \mathbf{b}, \quad (4.39)$$

where \mathbf{b} is the beamformer defined in Section 4.2. The beamforming analyses of the first 27 eigenvectors of the array covariance matrix computed from the Maule earthquake (Fig. 4.14c) are presented in Fig. 4.15. Each eigenvector show a specific beam pattern, and some of them clearly contain body waves (for instance, eigenvectors #4, 8, 14, 15, and 22), while the other ones does not. Therefore, we design a rejection criterion upon which the eigenvectors that contain energy inside an arbitrary circle are not considered in the equalization procedure. In the present case, we rejected all the eigenvectors that show a maximum of energy inside the slowness circle $\gamma_{hr} = 0.15$ s/km. The rejected eigenvectors are visible with a red label in Fig. 4.15. The selection is performed at each frequency.

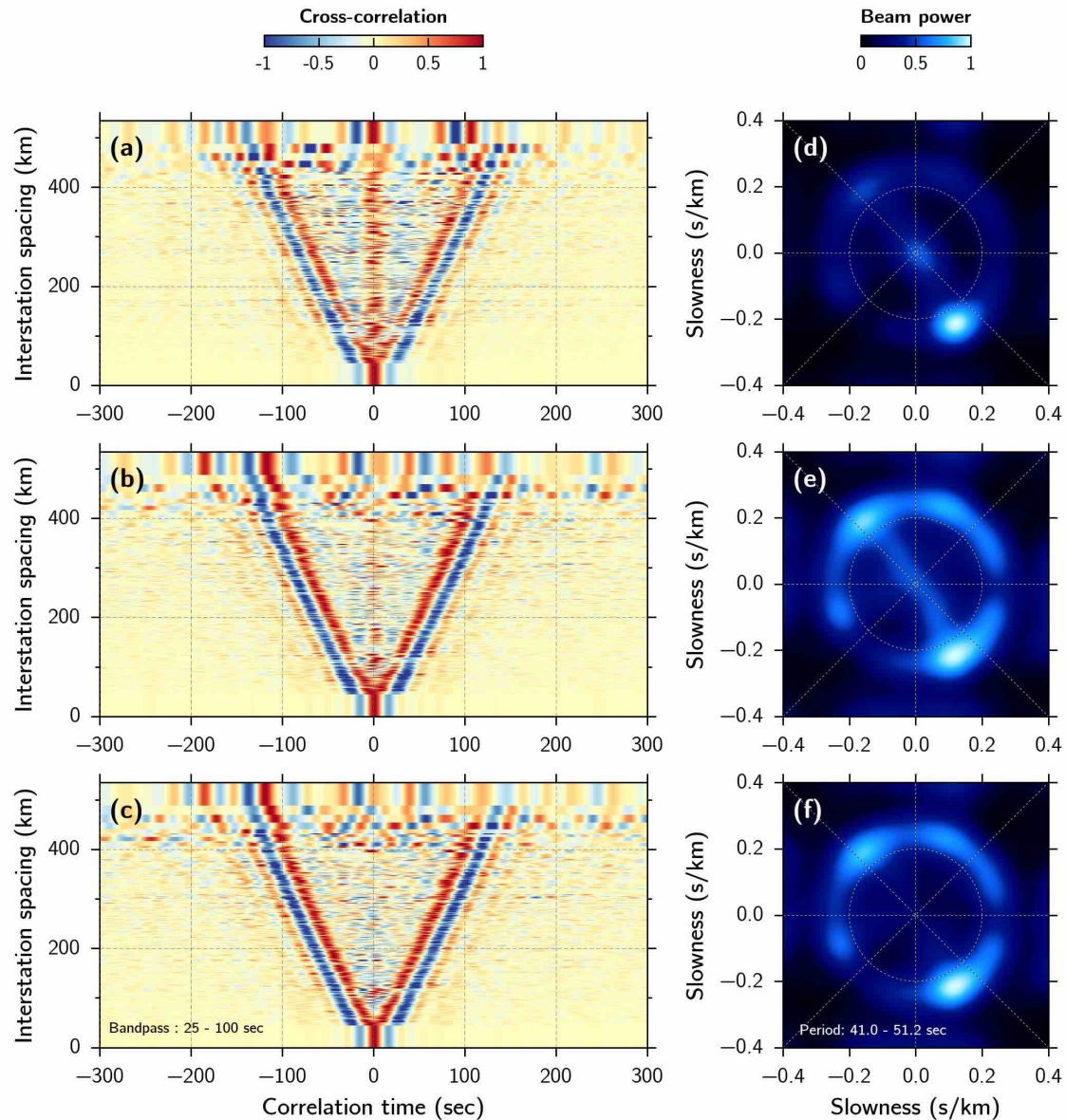


Figure 4.14 – Spatial equalization of the 2010 M8.8 Maule earthquake. Cross-correlation functions obtained between all the 34 seismic stations, regarding the inter-station spacing and the correlation time. The cross-correlations are bandpass-filtered between 25 and 100 seconds. **(a)** Cross-correlation obtained from the data smoothed in both the temporal and the spectral domain with the normalization technique proposed in [Bensen et al. \(2007\)](#). **(b)** Same as (a) obtained from the equalized covariance matrix. **(c)** Same as (b) with the additional eigenvector selection. The eigenvectors with energy inside the 0.15 km/s slowness circle are systematically rejected. **(d-f)** Beamforming analyses of cases (a-c) respectively, averaged between 41 and 51.2 seconds of period.

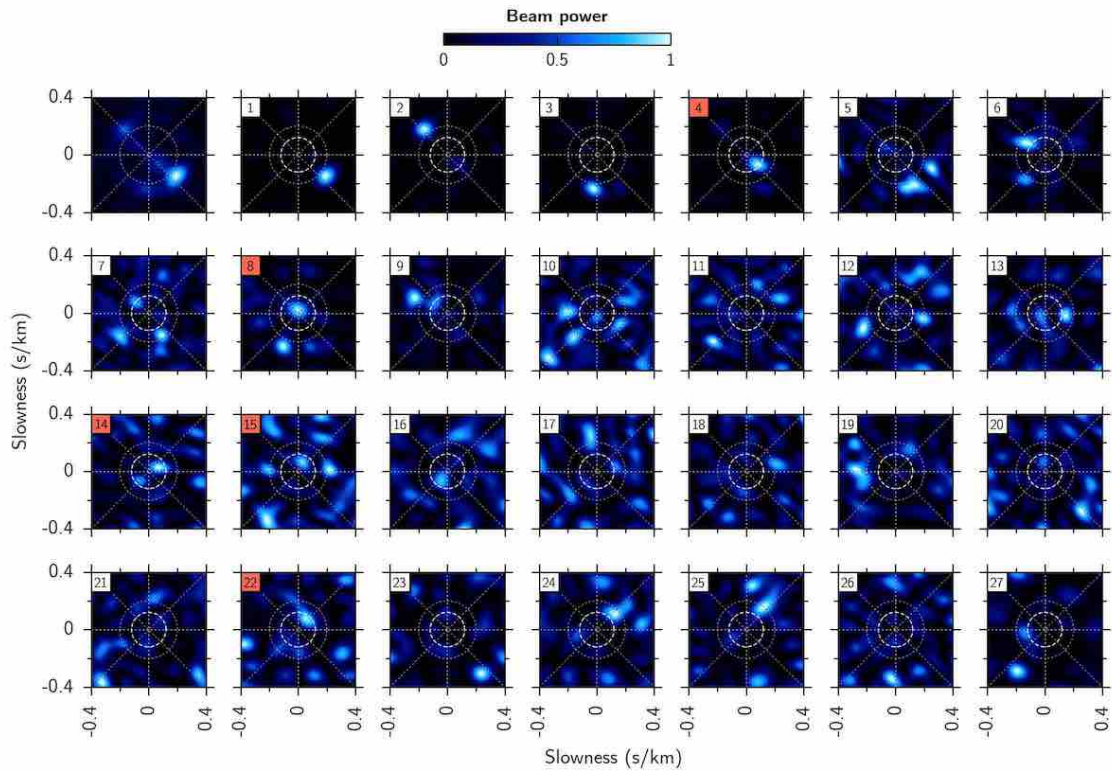


Figure 4.15 – Selection of the covariance matrix eigenvectors that are related to Rayleigh waves. Each plot show the beamforming performed on one eigenvector of the covariance matrix computed from the M8.8 Maule earthquake. The full-rank beamforming is presented in the top-left plot. In each case, the number correspond to the eigenvector index. The dashed circle corresponds to the slowness circle of rejection, γ_{hr} and the rejected eigenvectors are indicated with red labels.

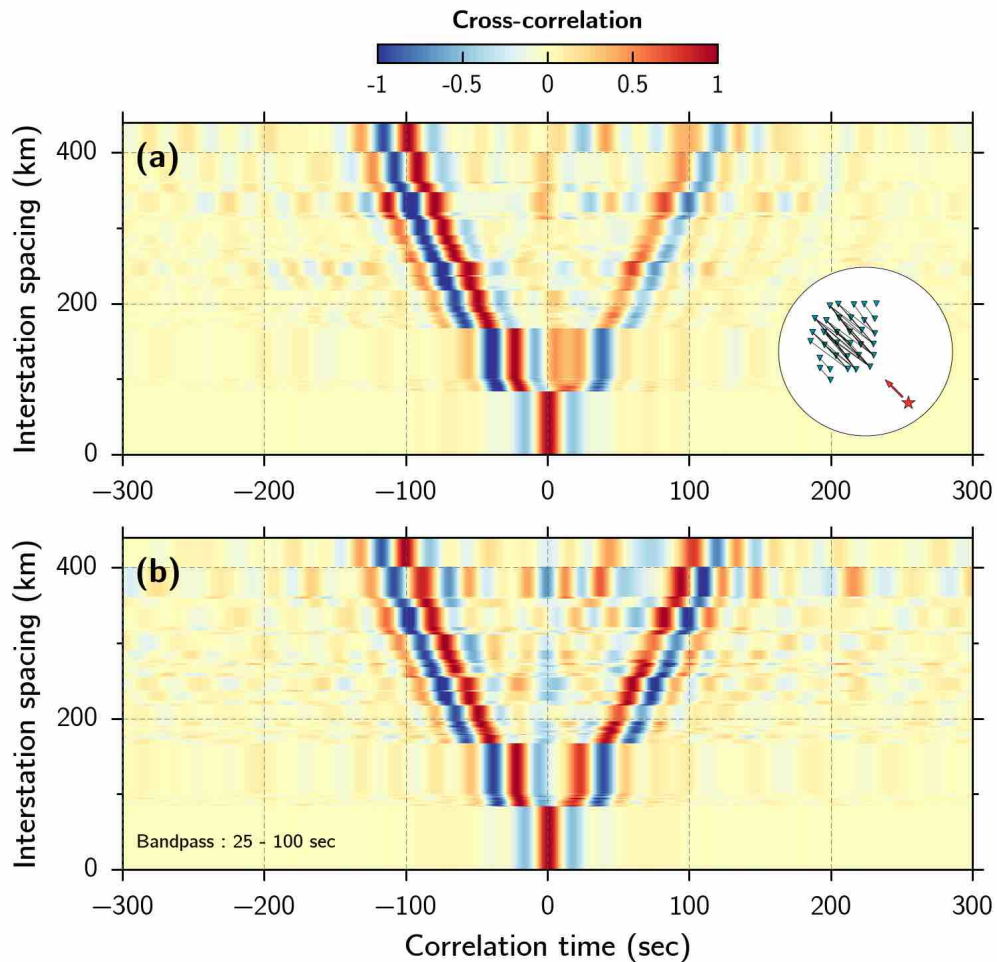


Figure 4.16 – Comparison between the standard cross-correlations and the cross-correlations obtained from the equalized covariance matrix and the eigenvector selection. The selected station pairs are aligned with the earthquake azimuth, with a tolerance angle of 10° , as indicated in the inset in (a). **(a)** Standard cross-correlations **(b)** cross-correlations obtained with the spatial equalization and the eigenvector selection.

We perform the equalization of the covariance matrix with additional eigenvector selection, and present the corresponding cross-correlation functions and beamforming in Fig. 4.14b and d respectively. We observe that the beamforming pattern related to the body waves that were incoming from the Maule earthquake have been removed. The corresponding peaks visible near the zero time-lag in the cross-correlations have been attenuated as well.

To better understand the effects of this equalization process, we present two particular cases in Figs. 4.16 and 4.17, where the stations pairs have been selected to be respectively oriented and orthogonal with the earthquake azimuth.

The standard cross-correlations that are aligned with the earthquake azimuth are presented in Fig. 4.16a. As expected by the theory, the Rayleigh waves incoming from the earthquake do not induce spurious arrivals in the cross-correlations because they contribute to the construction of coherent wavefronts that propagate along the path between the two stations. Nevertheless, the amplitude of the cross-correlation functions are asymmetrical, since most of the energy incomes from one side of the system. The same result after having applied the spatial equalization and eigenvector selection shows up to

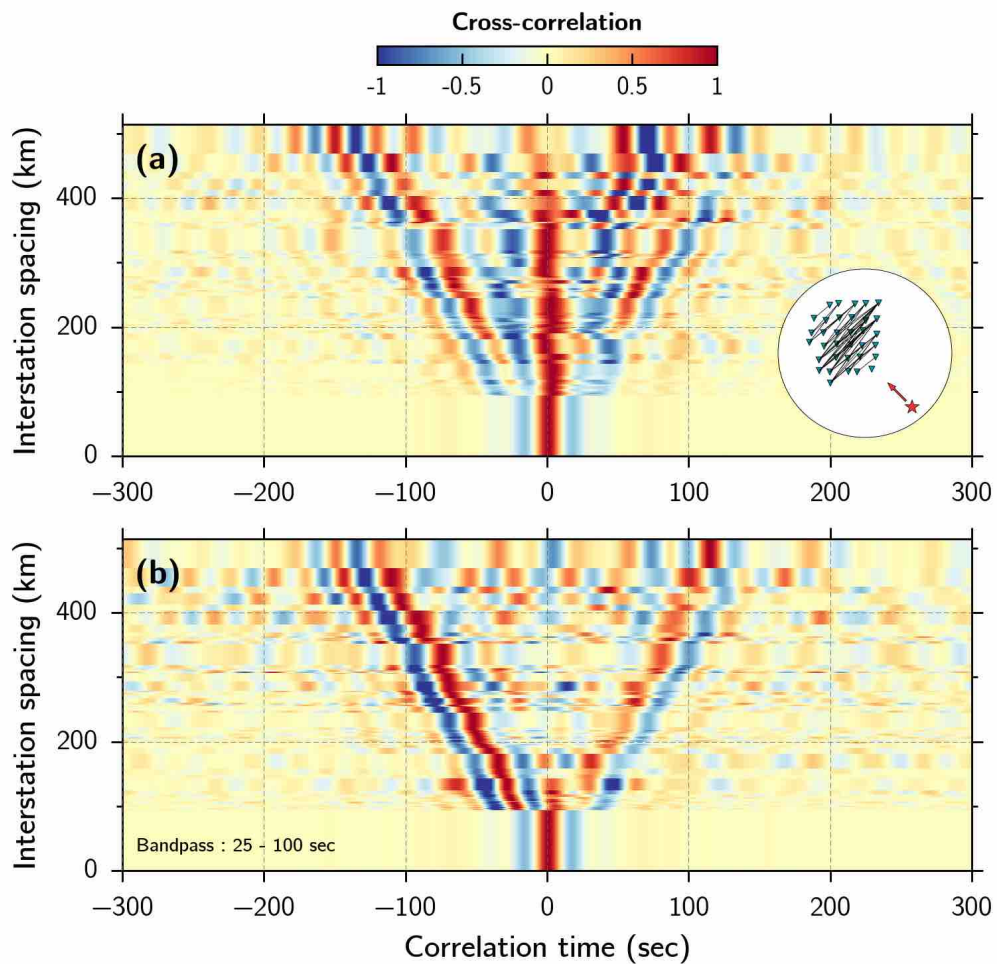


Figure 4.17 – Comparison between the standard cross-correlations and the cross-correlations obtained from the equalized covariance matrix and the eigenvector selection. The selected station pairs are orthogonal with the earthquake azimuth, with a tolerance angle of 10° , as indicated in the inset in (a). **(a)** Standard cross-correlations **(b)** cross-correlations obtained with the spatial equalization and the eigenvector selection.

strongly improve the symmetry of these cross-correlations.

The station pairs which axis is orthogonal with the earthquake azimuth show standard cross-correlation functions with strong peak near the zero time-lag in Fig. 4.17a. This can be simply explained geometrically: waves impinging on the stations from the earthquake arrive at the same moment to both the stations of this orientation. This strong peak is almost completely canceled from cross-correlation function obtained after the filtering of the covariance matrix (Fig. 4.17b). We see that the symmetry of the cross-correlation functions is not as good as in the oriented case, but it is suitable for travel time measurements.

4.7 Discussion and conclusions

We have developed an approach that is useful to enhance Green's function, recovering from ambient seismic noise in presence of strong seismic sources. Our approach is inspired by the passive inverse filter presented in Gallot et al. (2012) and by the least-square truncated spatial filter presented in Carrière et al. (2014). A theoretical expression for the eigenvalue-cutoff which is involved in our method as well as in Gallot et al. (2012) and Carrière et al. (2014) is derived, in both 2-D and 3-D noise configurations. We use this theoretical derivation to automatically define the eigenvalue threshold as a function of the frequency, without any prior knowledge of the seismic wavefield spatial distribution. The technique here developed has the advantage to be directly performed onto the array covariance matrix, and therefore, onto pre-existing cross-correlation databases. Finally, we propose an eigenvector-based spatial-filter to be combined with the inverse-filter, which consider the selection of the covariance matrix eigenvectors to remove unwanted seismic waves such as body waves.

We performed a numerical simulation with using the geometry of the USArray to investigate the performance of our method. We simulate the effect of a strong source embedded in isotropic ambient noise, where the source has an amplitude 10 times higher than the background noise. Strong bias in the travel-time measurements with an error of more than 37% are observed on the standard cross-correlation computed from this highly anisotropic wavefield configuration. The equalization of the covariance matrix spectrum applied to this last case shows to be a powerful tool to attenuate the effect of the source. The source amplitude is strongly attenuated while the background isotropic noise is reinforced, leading to significantly better travel-time measurements, with overall errors of about 6%.

Moreover, we see that our derivation of the spatial inverse filter requires the knowledge of the medium slowness. Even if the medium is inhomogeneous, we herein considered the average slowness. This is likely to produce the fluctuations in the filtered cross-correlations because the reconstruction from a too much number of Bessel function (both spherical and cylindrical) implies undesired high-frequency fluctuation. A good improvement of this approach will be to consider an initial slowness model, at least frequency-dependent, to improve the accuracy of the cutoff. Furthermore, the definition of the average radius \bar{r} is important. We here decided to take the average inter-station spacing $\bar{r} = (N(N-1)/2)^{-1} \sum_i \sum_{j>i} r_{ij} \approx 2/3 r_{max}$, where r_{max} is the radius of the circle that incorporate all the seismic stations. This is likely to provide a good focusing at the seismic stations located inside the seismic array, while the ones located at the border will have a degraded one.

We have shown that the isotropy of the seismic wavefield obtained from the widely

used spectral and temporal pre-processing (Bensen et al. 2007) is strongly affected by the presence of earthquakes. Using beamforming, we identify that the Rayleigh waves related to the background seismic noise are weakly present, when strong peaks of energy are related to the earthquakes. We apply the equalization technique to 30 days of the USArray data, centered around the date of the 2010 M8.8 Maule earthquake. The standard cross-correlation functions obtained from the data pre-processed with temporal and spectral equalization (Bensen et al. 2007) still suffers from the presence of the earthquake, with both spurious coherent arrivals and asymmetry of the causal and anti-causal parts. The equalization of the covariance matrix spectrum shows up to strongly reduce both artifacts.

However, we highlight the inability of this approach to be designed to separate the body and surface waves, as we observe that the earthquake-related body wave are reinforced as well. We therefore propose an eigenvector-based selection to cancel body waves. Using beamforming analysis on eigenvectors, we reject the eigenvectors that contain energy in an arbitrary slowness domain, related to body waves. This eigenvector selection, combined with the equalization of the covariance matrix spectrum, strongly improve the cross-correlation symmetry and reduces the spurious arrivals due to the earthquake.

We believe this approach is strongly attractive because it allows to perform the cross-correlation analysis even when the wavefield does not present the required conditions that ensure the Green's function retrieval. It is of particular interest if it is applied to the monitoring of active zones, where seismic activity is strong and directive, and when the window rejection is not appropriate because it may considerably lower the amount of exploitable data.

Conclusions et Perspectives

Conclusions

Au cours de la dernière décennie, une méthode d'imagerie passive à partir du bruit ambiant a vu le jour. Elle a en particulier trouvé des applications remarquables en sismologie. Dans le cas idéal, la réponse transitoire entre deux stations sismiques est estimée depuis l'enregistrement continu du bruit sismique ambiant sur lesdites stations. Appliquée à un réseau de stations sismiques, cette méthode permet l'acquisition de sismogrammes inter-stations comme si un séisme "virtuel" avait eu lieu sur chaque station. Il est alors possible d'appliquer des méthodes de tomographie classique avec une grande résolution.

L'extraction des fonctions de Green par corrélation de bruit repose sur le principe d'équipartition du champ d'onde sismique. Ce principe constitue ainsi une hypothèse fondamentale que doivent vérifier le champ d'onde sismique. En d'autres termes, le champ d'onde sismique doit être généré par des sources de bruit sismique réparties de façon homogène dans le milieu.

En pratique, cette hypothèse n'est que partiellement vérifiée. En effet, comme nous avons pu le voir au cours de ce manuscrit, le bruit sismique est généré par une grande variété de sources sismiques qui présentent des caractéristiques différentes. Aucune ne génère un champ parfaitement ergodique. Des méthodes de traitement des données sismiques ont donc été développées pour améliorer la qualité de reconstruction des fonctions de Green. Le travail ici présenté s'est inscrit dans cette dynamique.

Cette thèse s'est inspirée de méthodes de traitement d'antenne pour analyser et traiter des données sismiques de réseaux. Ces méthodes nous ont naturellement conduits à construire la matrice de covariance des données sismiques, objet mathématique situé au cœur des techniques de traitement d'antenne classiques. Deux axes principaux se sont alors dégagés.

Dans un premier temps, nous avons développé une méthode robuste d'estimation du degré de cohérence du champ d'onde sismique enregistré sur le réseau. En exploitant les propriétés statistiques de la matrice de covariance, nous avons proposé d'utiliser sa largeur spectrale comme un indicateur du degré de cohérence du champ d'onde sismique. À travers une série de tests synthétiques où différentes configurations de champs d'onde sont simulées, nous avons montré que ce critère permet de distinguer des signaux cohérents du bruit sismique et du bruit non corrélé entre les capteurs. Plusieurs paramètres essentiels à cette analyse doivent être considérés comme la largeur des fenêtres temporelles analysées et le nombre de fenêtres d'estimation de la matrice de covariance. Ces paramètres dépendent essentiellement de la géométrie du réseau et de la bande de fréquence analysée et sont estimés dans ce manuscrit via des tests synthétiques.

La méthode a été appliquée sur deux jeux de données aux caractéristiques différentes. Le premier a été enregistré par le réseau de surveillance du Piton de la Fournaise de façon continue pendant l'année 2010. L'ouverture de ce premier réseau est d'environ 10 km et

les distances inter stations de l'ordre de 3 km donnant accès à des fréquences allant de 0.02 à 10 Hz pour analyser la cohérence du champ d'onde. Le second jeu de données est issu d'une sélection de 121 stations sismiques de l'USArray, qui ont enregistré le champ d'onde pendant l'année 2010 en continu également. Ce sous-réseau de l'USArray s'étend sur environ 500 km impliquant un choix de paramètres différents. L'analyse du champ d'onde a donc été effectuée à plus basse fréquence, entre 0.01 Hz et 1 Hz.

Ainsi, nous avons montré que cette méthode qui permet d'estimer le degré de cohérence du champ d'onde en fonction du temps et de la fréquence présente une capacité à détecter l'activité sismique. Dans les basses fréquences (< 0.1 Hz), l'analyse des deux jeux de données montre que l'activité sismique globale est détectée. Une analyse automatique de ces signaux détectés sur le jeu de données du Piton de la Fournaise montre par ailleurs que 75% du catalogue mondial des séismes de magnitude supérieure à six est retrouvé. À plus haute fréquence, l'activité sismo-volcanique du Piton de la Fournaise est extraite des données. Les dates et les durées de ces événements concordent bien avec les catalogues issus d'observatoires. Appliquée sur le jeu de données de l'USArray, la méthode met en évidence également des événements d'origine microsismiques dans les fréquences entre 0.07 et 0.14 Hz, et des signaux persistants comme le pic micro-sismique à 26 secondes déjà observé auparavant dans la littérature scientifique. La détection de signaux émergents comme les tremors volcaniques ou les microséismes d'origine océanique est un aspect très intéressant, car ces signaux sont actuellement difficiles à observer avec des méthodes de détections classiques.

L'information de cohérence du champ d'onde contenue dans ces résultats est également cruciale dans le cadre de l'imagerie sismique à partir de bruit ambiant. Or, les méthodes d'imagerie passive fondées sur l'intercorrélacion de bruit supportent difficilement la présence de signaux "cohérents", comme nous avons pu le voir au cours de cette thèse. Cette méthode peut donc être également utilisée dans l'esprit de définir les zones temporelles et fréquentielles à partir desquelles il est possible de faire des reconstructions des fonctions de Green de qualité. Dans ce même esprit, nous avons également utilisé la méthode pour étudier les limites d'usage des égalisations temporelles et spectrales qui visent à réduire l'effet de sources cohérentes, et qui sont couramment utilisées en imagerie de bruit passive à l'heure actuelle. Nos résultats montrent que si ces méthodes permettent d'atténuer en grande partie des signaux enregistrés localement par quelques stations sismiques, elle ne parvient pas à atténuer les sources les plus cohérentes.

Finalement, ce travail s'est orienté vers le développement d'une seconde méthode visant cette fois à atténuer les ondes les plus cohérentes et à amplifier les régions faiblement excitées par le bruit sismique ambiant. Cette seconde méthode s'inspire de la méthode d'égalisation du spectre fréquentiel des sismogrammes et consiste à égaliser un nombre limité de valeurs propres de la matrice de covariance jusqu'à un certain ordre. Nous proposons d'interpréter cette méthode comme une égalisation *spatiale* du champ d'onde, à l'instar de l'égalisation spectrale des sismogrammes, mais également comme une application directe d'un filtre inverse passif sur la matrice de covariance. Le filtre inverse passif a été proposé récemment en sismologie et s'inspire naturellement des méthodes utilisées en imagerie acoustique. Ces méthodes sont présentées dans ce manuscrit et nous permettent d'argumenter notre approche.

Notre méthode de filtrage de la matrice de covariance est construite à l'aide d'un unique paramètre qui est le nombre de valeurs propres à prendre en compte dans l'égalisation de la matrice de covariance. Nous avons proposé au sein de ce manuscrit de déterminer ce nombre à partir du nombre de degrés de liberté du champ d'onde mesuré

par un réseau de stations sismiques. Ce nombre de degrés de liberté peut être déduit des propriétés fondamentales du champ d'onde comme la fréquence et la lenteur des ondes, ainsi que de l'étendue spatiale moyenne du réseau de stations sismiques. La lenteur des ondes est bien entendu inconnue, puisqu'elle constitue l'objet de recherche en imagerie passive. En revanche, nous proposons d'utiliser un *a priori* de cette lenteur pour égaliser le champ d'onde. Des extensions de la méthode peuvent être envisagées et sont discutées dans les perspectives, en fin de ce chapitre. Les performances de notre méthode en matière d'erreurs de mesure sur les temps de trajets à partir des fonctions d'intercorrélation sont également étudiées à l'aide de simulations aux différences finies d'un champ d'onde acoustique en milieu hétérogène. Nous montrons alors que les temps de trajets obtenus avec la matrice de covariance filtrée sont grandement améliorés par rapport au temps de trajets obtenus avec la matrice de covariance polluée par une source.

Enfin, cette seconde méthode est appliquée au jeu de données de l'USArray, sur des périodes où des séismes de forte magnitude sont enregistrés et "polluent" le bruit sismique ambiant. Les signaux très cohérents tels que ceux issus du séisme de magnitude 8.8 à Maule au Chili peuvent être égalisés par notre méthode.

Enfin, nous proposons en dernier lieu de séparer les informations relatives aux ondes de surface et aux ondes de volume. Nous effectuons cette séparation à l'aide d'un critère de sélection des vecteurs propres de la matrice de covariance dans l'espace des lenteurs, grâce à la formation de voie appliquée sur ces vecteurs propres. Les vecteurs sont alors rejetés lorsqu'ils contiennent de l'énergie en dessous d'une certaine valeur de lenteur fixée arbitrairement. Cette méthode montre qu'il est possible de soustraire les ondes de volume pour ne garder que les ondes de surface. C'est un point intéressant, car il permet de travailler sur un seul type d'onde à la fois.

Perspectives

Nous avons vu que notre méthode d'analyse de la cohérence du champ d'onde peut être employée comme une méthode de détection. Cette méthode présente l'avantage d'être robuste, c'est-à-dire qu'elle permet d'observer des signaux de faible amplitude par rapport au bruit. Sa résolution temporelle est toutefois faible devant celle des méthodes de pointé, en raison de la taille (nécessaire) des fenêtres sur lesquelles sont estimés les spectres des sismogrammes. En termes de détection, la méthode constitue donc un *a priori* qui peut être utile à l'utilisation des méthodes de pointé beaucoup plus sensibles au bruit sismique.

Des améliorations de la résolution temporelle sont également envisageables de plusieurs façons. D'une part, en remarquant que la taille des fenêtres d'analyse est contrainte uniquement par la plus basse fréquence analysée, nous pouvons imaginer une gestion adaptative de la taille des fenêtres garantissant une résolution optimale à toute fréquence. Également, nous pourrions considérer une autre transformation que la transformée de Fourier, comme la transformée en ondelettes complexes qui peut être estimée sur des fenêtres de plus courtes durées. Ce second point nécessite également une étude formelle du comportement de la matrice de covariance estimée à partir de la nouvelle transformation.

Une autre utilisation pratique d'une telle méthode est la classification automatique de signaux sismiques. Au cours de ce manuscrit, les différentes détections (activité télé-sismique, ou sismo-volcanique) ont été réalisées dans des bandes de fréquences sélectionnées par nos soins. Nous pouvons cependant imaginer une classification automatique des différentes structures présentes dans l'analyse temps-fréquence de la largeur spectrale, fondée sur des outils comme l'analyse en composantes principales ou l'analyse en composantes indépendantes. Cette approche permettrait de détecter automatiquement les évé-

nements microsismiques, par exemple, et ainsi leur analyse systématique par formation de voie, ou bien une étude un peu plus statistique de leurs caractéristiques temps-fréquence.

Pour finir, la largeur spectrale de la matrice de covariance peut être utilisée comme témoin de la faisabilité des corrélations de bruit, en fonction du temps et de la fréquence. Si plusieurs critères comme l'asymétrie des fonctions de corrélation, ou bien l'amplitude maximale des fenêtres de signal sont utilisés à l'heure actuelle pour sélectionner les "bons" segments de signal, la largeur de la matrice de covariance est un témoin indépendant, qui peut être employé facilement comme critère de rejet. Dans cette même idée, il est possible de l'employer pour déterminer l'efficacité du prétraitement utilisé pour réduire les événements de forte amplitude. Ces méthodes de prétraitement sont régies par des paramètres qui peuvent être sélectionnés à l'aide de notre analyse lorsque les événements les plus cohérents sont les plus atténués.

La méthode d'égalisation spatiale, quant à elle nécessite encore plusieurs études avant d'être définitivement validée. En effet, de nombreuses questions restent ouvertes et devront faire l'objet d'études futures, comme l'analyse d'autres formes d'égalisation des valeurs propres, moins brutales, par exemple, de façon analogue aux méthodes de filtrage fréquentiel qui évitent le plus souvent d'utiliser des fenêtrages abrupts responsables d'artefacts indésirables dans les signaux traités.

Aussi, il serait intéressant de déterminer quantitativement les capacités de cette approche, comme la vitesse de convergence des fonctions d'intercorrélation vers la fonction de Green, ou bien le nombre minimum de fenêtres à prendre en compte pour l'appliquer. Finalement, la méthode mériterait sans doute d'être testée sur un jeu de données réellement problématique, comme la tomographie 4D d'une zone à forte activité sismique.

Enfin, la méthode proposée en toute fin du dernier chapitre de sélection des vecteurs propres liés aux ondes de surface mérite également d'être étudiée plus en profondeur. D'autres méthodes de filtrage des ondes de volume existent, dans le domaine des nombres d'onde, par exemple, et ne peuvent être appliquées que lorsque le réseau a une répartition régulière de stations sismiques. Dans notre cas, nous avons souhaité envisager une solution plus directe, qui concerne directement les vecteurs propres de la matrice de covariance. Cette méthode fonctionne dans notre cas, mais une application plus fine de cette sélection de vecteurs propres est à considérer. Par exemple, il serait sans doute possible d'effectuer une sélection par filtrage avant d'égaliser les valeurs propres, afin de conserver un maximum d'information relative aux ondes de surface.

Références

- Aki, K. and Richards, P. (2002). *Quantitative Seismology*. Geology (University Science Books). : Seismology. University Science Books.
- Allen, R. (1982). Automatic phase pickers : Their present use and future prospects. *Bull. Seismol. Soc. Am.*, 72(6B) :S225–S242.
- Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, 34(1) :122–148.
- Ardhuin, F., Gualtieri, L., and Stutzmann, E. (2015). How ocean waves rock the earth : Two mechanisms explain microseisms with periods 3 to 300 s. *Geophys. Res. Lett.*, 42(3) :765–772.
- Ardhuin, F., Stutzmann, E., Schimmel, M., and Mangeney, A. (2011). Ocean wave sources of seismic noise. *J. Geophys. Res.*, 116(C9).
- Aubry, A. and Derode, A. (2009). Random matrix theory applied to acoustic backscattering and imaging in complex media. *Phys. Rev. Lett.*, 102(8) :084301.
- Aubry, J.-F., Tanter, M., Gerber, J., Thomas, J.-L., and Fink, M. (2001). Optimal focusing by spatio-temporal inverse filter. II. experiments. application to focusing through absorbing and reverberating media. *The Journal of the Acoustical Society of America*, 110(1) :48–11.
- Baillard, C., Crawford, W. C., Ballu, V., Hibert, C., and Mangeney, A. (2013). An automatic kurtosis-based p- and s-phase picker designed for local seismic networks. *Bull. Seismol. Soc. Am.*, 104(1) :394–409.
- Bartlett, M. S. (1947). Multivariate Analysis. *Supplement to the Journal of the Royal Statistical Society*, 9(2) :176.
- Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 296–298.
- Behm, M., Nakata, N., and Bokelmann, G. (2016). Regional ambient noise tomography in the eastern alps of Europe. *Pure Appl. Geophys.*, pages 1–28.
- Bensen, G. D., Ritzwoller, M. H., Barmin, M. P., Levshin, A. L., Lin, F., Moschetti, M. P., Shapiro, N. M., and Yang, Y. (2007). Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophys. J. Int.*, 169(3) :1239–1260.

- Bienvenu, G. and Kopp, L. (1980). Adaptivity to background noise spatial coherence for high resolution passive methods. In *ICASSP '80. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 307–310. Institute of Electrical & Electronics Engineers (IEEE).
- Boschi, L., Weemstra, C., Verbeke, J., Ekstrom, G., Zunino, A., and Giardini, D. (2012). On measuring surface wave phase velocity from station-station cross-correlation of ambient signal. *Geophys. J. Int.*, 192(1) :346–358.
- Boué, P., Poli, P., Campillo, M., Pedersen, H., Briand, X., and Roux, P. (2013). Tele-seismic correlations of ambient seismic noise for deep global imaging of the earth. *Geophys. J. Int.*, 194(2) :844–848.
- Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M., and Larose, E. (2008a). Postseismic relaxation along the san andreas fault at parkfield from continuous seismological observations. *Science*, 321(5895) :1478–1481.
- Brenguier, F., Campillo, M., Takeda, T., Aoki, Y., Shapiro, N. M., Briand, X., Emoto, K., and Miyake, H. (2014). Mapping pressurized volcanic fluids from induced crustal seismic velocity drops. *Science*, 345(6192) :80–82.
- Brenguier, F., Kowalski, P., Ackerley, N., Nakata, N., Boué, P., Campillo, M., Larose, E., Rambaud, S., Pequegnat, C., Lecocq, T., Roux, P., Ferrazzini, V., Villeneuve, N., Shapiro, N. M., and Chaput, J. (2015). Toward 4D noise-based seismic probing of volcanoes : Perspectives from a large- n experiment on piton de la fournaise volcano. *Seismol. Res. Lett.*, 87(1) :15–25.
- Brenguier, F., Kowalski, P., Staudacher, T., Ferrazzini, V., Lauret, F., Boissier, P., Catherine, P., Lemarchand, A., Pequegnat, C., Meric, O., Pardo, C., Peltier, A., Tait, S., Shapiro, N. M., Campillo, M., and Muro, A. D. (2012). First results from the UnderVolc high resolution seismic and GPS network deployed on piton de la fournaise volcano. *Seismol. Res. Lett.*, 83(1) :97–102.
- Brenguier, F., Shapiro, N. M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O., and Nercessian, A. (2008b). Towards forecasting volcanic eruptions using seismic noise. *Nat. Geosci.*, 1(2) :126–130.
- Brenguier, F., Shapiro, N. M., Campillo, M., Nercessian, A., and Ferrazzini, V. (2007). 3-D surface wave tomography of the piton de la fournaise volcano using seismic noise correlations. *Geophys. Res. Lett.*, 34(2) :L02305.
- Bromirski, P. D., Stephen, R. A., and Gerstoft, P. (2013). Are deep-ocean-generated surface-wave microseisms observed on land? *J. Geophys. Res. Solid Earth*, 118(7) :3610–3629.
- Bungum, H., Husebye, E. S., and Ringdal, F. (1971). The NORSAR array and preliminary results of data analysis. *Geophys. J. Int.*, 25(1-3) :115–126.
- Campillo, M. (2003). Long-range correlations in the diffuse seismic coda. *Science*, 299(5606) :547–549.

- Campillo, M. (2006). Phase and Correlation in ‘Random’ Seismic Fields and the Reconstruction of the Green Function. *Pure Appl. Geophys.*, 163(2-3) :475–502.
- Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8) :1408–1418.
- Capon, J. (1974). Characterization of crust and upper mantle structure under lasa as a random medium. *Bull. Seismol. Soc. Am.*, 64(1) :235–266.
- Carriere, O., Gerstoft, P., and Hodgkiss, W. (2013). Spatial filtering in ambient noise crosscorrelation. In *ICA 2013 Montreal*, pages 070043–070043. ASA.
- Carrière, O., Gerstoft, P., and Hodgkiss, W. S. (2014). Spatial filtering in ambient noise interferometry. *The Journal of the Acoustical Society of America*, 135(3) :1186–1196.
- Catheline, S. and Bencech, N. (2015). Longitudinal shear wave and transverse dilatational wave in solids. *The Journal of the Acoustical Society of America*, 137(2) :EL200–EL205.
- Crotwell, H. P., Owens, T. J., and Ritsema, J. (1999). The TauP toolkit : Flexible seismic travel-time and ray-path utilities. *Seismological Research Letters*, 70(2) :154–160.
- Delouis, B., Nocquet, J.-M., and Vallée, M. (2010). Slip distribution of the February 27, 2010 $m_w = 8.8$ maule earthquake, central Chile, from static and high-rate GPS, InSAR, and broadband teleseismic data. *Geophys. Res. Lett.*, 37(17) :L17305–7.
- Derode, A., Larose, E., Tanter, M., de Rosny, J., Tourin, A., Campillo, M., and Fink, M. (2003). Recovering the green’s function from field-field correlations in an open scattering medium (I). *The Journal of the Acoustical Society of America*, 113(6) :2973–2976.
- Duputel, Z., Ferrazzini, V., Brenguier, F., Shapiro, N., Campillo, M., and Nercessian, A. (2009). Real time monitoring of relative velocity changes using ambient seismic noise at the piton de la fournaise volcano (la réunion) from January 2006 to June 2007. *J. Volcanol. Geotherm. Res.*, 184(1-2) :164–173.
- Duvall, T. L., Jefferies, S. M., Harvey, J. W., and Pomerantz, M. A. (1993). Time-distance helioseismology. *Nature*, 362(6419) :430–432.
- Dziewonski, A. M. and Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the Earth and Planetary Interiors*, 25(4) :297–356.
- Eckart, C. (1953). The Theory of Noise in Continuous Media. *The Journal of the Acoustical Society of America*, 25(2) :195–199.
- Ekström, G., Nettles, M., and Dziewoński, A. (2012). The global CMT project 2004–2010 : Centroid-moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Inter.*, 200-201 :1–9.
- Ekström, G. (2001). Time domain analysis of Earth’s long-period background seismic radiation. *J. Geophys. Res.*, 106(B11) :26483–26493.

- Ekström, G. (2014). Love and Rayleigh phase-velocity maps, 5–40 s, of the western and central USA from USArray data. *Earth and Planetary Science Letters*, 402 :42–49.
- Ermert, L., Villaseñor, A., and Fichtner, A. (2015). Cross-correlation imaging of ambient noise sources. *Geophys. J. Int.*, 204(1) :347–364.
- Fink, M. (1992). Time reversal of ultrasonic fields. i. basic principles. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 39(5) :555–566.
- Friedrich, A., Krüger, F., and Klinge, K. (1998). Ocean-generated microseismic noise located with the gräfenberg array. *J. Seismolog.*, 2(1) :47–64.
- Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A., and Weaver, R. L. (2010). Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations. *GEOPHYSICS*, 75(5) :SA85–SA93.
- Frosch, R. A. and Green, P. E. (1966). The Concept of a Large Aperture Seismic Array. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 290(1422) :368–384.
- Gallot, T., Catheline, S., Roux, P., and Campillo, M. (2012). A passive inverse filter for green’s function retrieval. *The Journal of the Acoustical Society of America*, 131(1) :EL21–7.
- Gee, L. S. and Leith, W. S. (2011). The global seismographic network. Technical report, US Geological Survey.
- Gerstoft, P., Menon, R., Hodgkiss, W. S., and Mecklenbrauker, C. (2012). Eigenvalues of the sample covariance matrix for a towed array. *The Journal of the Acoustical Society of America*, 132(3) :2007.
- Gerstoft, P., Shearer, P. M., Harmon, N., and Zhang, J. (2008). Global p, PP, and PKP wave microseisms observed from distant storms. *Geophys. Res. Lett.*, 35(23) :L23306.
- Goldstein, P. and Archuleta, R. J. (1987). Array analysis of seismic signals. *Geophys. Res. Lett.*, 14(1) :13–16.
- Golub, G. and Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 2(2) :205–224.
- Gouédard, P., Stehly, L., Brenguier, F., Campillo, M., de Verdière, Y. C., Larose, E., Margerin, L., Roux, P., Sánchez-Sesma, F. J., Shapiro, N. M., and Weaver, R. L. (2008). Cross-correlation of random fields : mathematical approach and applications. *Geophys. Prospect.*, 56(3) :375–393.
- Green, G. (1828). *An essay on the application of mathematical analysis to the theories of electricity and magnetism*. Wezäta-Melins Aktiebolag.
- Gualtieri, L., Stutzmann, E., Capdeville, Y., Arduin, F., Schimmel, M., Mangeney, A., and Morelli, A. (2013). Modelling secondary microseismic noise by normal mode summation. *Geophysical Journal International*, 193(3) :1732–1745.

- Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S., Davaille, A., and Yelles-Chaouche, A. (2015). Global tomography using seismic hum. *Geophys. J. Int.*, 204(2) :1222–1236.
- Hasselmann, K. (1963). A statistical analysis of the generation of microseisms. *Rev. Geophys.*, 1(2) :177.
- Hillers, G., Campillo, M., Ben-Zion, Y., and Roux, P. (2014). Seismic fault zone trapped noise. *J. Geophys. Res. Solid Earth*, 119(7) :5786–5799.
- Hillers, G., Roux, P., Campillo, M., and Ben-Zion, Y. (2016). Focal spot imaging based on zero lag cross correlation amplitude fields : Application to dense array data at the san jacinto fault zone. *Journal of Geophysical Research : Solid Earth*.
- Holcomb, L. G. (1980). Microseisms : A twenty-six-second spectral line in long-period earth motion. *Bulletin of the Seismological Society of America*, 70(4) :1055–1070.
- Holcomb, L. G. (1998). Spectral structure in the earth's microseismic background between 20 and 40 seconds. *Bulletin of the Seismological Society of America*, 88(3) :744–757.
- IRIS (1993). Standard for the exchange of earthquake data—reference manual.
- Jacobson, M. J. (1962). Space-Time Correlation in Spherical and Circular Noise Fields. *The Journal of the Acoustical Society of America*, 34(7) :971–978.
- Jaxybulatov, K., Shapiro, N. M., Koulakov, I., Mordret, A., Landes, M., and Sens-Schonfelder, C. (2014). A large magmatic sill complex beneath the toba caldera. *Science*, 346(6209) :617–619.
- Kedar, S., Longuet-Higgins, M., Webb, F., Graham, N., Clayton, R., and Jones, C. (2008). The origin of deep ocean microseisms in the north atlantic ocean. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 464(2091) :777–793.
- Kennett, B. L. N. and Engdahl, E. R. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2) :429–465.
- Kimman, W. P. and Trampert, J. (2010). Approximations in seismic interferometry and their effects on surface waves. *Geophys. J. Int.*, 182(1) :no–no.
- Landès, M., Hubans, F., Shapiro, N. M., Paul, A., and Campillo, M. (2010). Origin of deep ocean microseisms by using teleseismic body waves. *J. Geophys. Res.*, 115(B5) :B05302.
- Lanza, F., Kenyon, L. M., and Waite, G. P. (2016). Near-surface velocity structure of pacaya volcano, Guatemala, derived from small-aperture array analysis of seismic tremor. *Bull. Seismol. Soc. Am.*, pages 1–8.
- Larose, E. (2004). Imaging from one-bit correlations of wideband diffuse wave fields. *J. Appl. Phys.*, 95(12) :8393.
- Lawley, D. N. (1956). Tests of Significance for the Latent Roots of Covariance and Correlation Matrices. *Biometrika*, 43(1/2) :128–136.

- Lehmann, I. (1936). P' : Publications du bureau central séismologique international. *A14*, pages 87–115.
- Leroy, C., Lani, S., Sabra, K. G., Hodgkiss, W. S., Kuperman, W. A., and Roux, P. (2012). Enhancing the emergence rate of coherent wavefronts from ocean ambient noise correlations using spatio-temporal filters. *The Journal of the Acoustical Society of America*, 132(2) :883–893.
- Li, X.-D. and Romanowicz, B. (1996). Global mantle shear velocity model developed using nonlinear asymptotic coupling theory. *J. Geophys. Res.*, 101(B10) :22245–22272.
- Lin, F.-C., Tsai, V. C., Schmandt, B., Duputel, Z., and Zhan, Z. (2013). Extracting seismic core phases with array interferometry. *Geophys. Res. Lett.*, 40(6) :1049–1053.
- Liu, Q., Koper, K. D., Burlacu, R., Ni, S., Wang, F., Zou, C., Wei, Y., Gal, M., and Reading, A. M. (2016). Source locations of teleseismic p, SV, and SH waves observed in microseisms recorded by a large aperture seismic array in China. *Earth Planet. Sci. Lett.*, 449(C) :39–47.
- Liu, Q.-H. (1997). The perfectly matched layer for acoustic waves in absorptive media. *The Journal of the Acoustical Society of America*, 102(4) :2072–2082.
- Longuet-Higgins, M. S. (1950). A Theory of the Origin of Microseisms. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 243(857) :1–35.
- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 1(4) :457–483.
- Marone, F., Gung, Y., and Romanowicz, B. (2007). Three-dimensional radial anisotropic structure of the north american upper mantle from inversion of surface waveform data. *Geophysical Journal International*, 171(1) :206–222.
- McNamara, D. E. (2004). Ambient noise levels in the continental United States. *Bulletin of the Seismological Society of America*, 94(4) :1517–1527.
- Menon, R., Gerstoft, P., and Hodgkiss, W. S. (2012a). Asymptotic eigenvalue density of noise covariance matrices. *IEEE Transactions on Signal Processing*, 60(7) :3415–3424.
- Menon, R., Gerstoft, P., and Hodgkiss, W. S. (2012b). Cross-correlations of diffuse ocean noise using eigenvalue based statistical inference. *The Journal of the Acoustical Society of America*, 132(3) :2007.
- Mestre, X. (2008). On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Transactions on Signal Processing*, 56(11) :5353–5368.
- Moiola, A., Hiptmair, R., and Perugia, I. (2011). Vekua theory for the helmholtz operator. *Z. Angew. Math. Phys.*, 62(5) :779–807.
- Mordret, A., Landes, M., Shapiro, N. M., Singh, S. C., and Roux, P. (2014). Ambient noise surface wave tomography to determine the shallow shear velocity structure at valhall : depth inversion with a neighbourhood algorithm. *Geophys. J. Int.*, 198(3) :1514–1525.

- Mordret, A., Landes, M., Shapiro, N. M., Singh, S. C., Roux, P., and Barkved, O. I. (2013). Near-surface study at the valhall oil field from ambient noise surface wave tomography. *Geophys. J. Int.*, 193(3) :1627–1643.
- Muller, R. (2002). A random matrix model of communication via antenna arrays. *IEEE Trans. Inform. Theory*, 48(9) :2495–2506.
- Nadakuditi, R. and Edelman, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions on Signal Processing*, 56(7) :2625–2638.
- Nakata, N., Chang, J. P., Lawrence, J. F., and Boué, P. (2015). Body wave extraction and tomography at long beach, california, with ambient-noise interferometry. *J. Geophys. Res. Solid Earth*, 120(2) :1159–1173.
- Nishida, K. and Kobayashi, N. (1999). Statistical features of earth's continuous free oscillations. *J. Geophys. Res.*, 104(B12) :28741–28750.
- Nishida, K., Montagner, J.-P., and Kawakatsu, H. (2009). Global surface wave tomography using seismic hum. *Science*, 326(5949) :112–112.
- Peterson, J. R. (1993). Observation and modeling of the seismic background noise. Technical Report 93-322, Albuquerque Seismological Laboratory.
- Prada, C. and Thomas, J.-L. (2003). Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix. *The Journal of the Acoustical Society of America*, 114(1) :235.
- Retailleau, L., Shapiro, N. M., Guilbert, J., Campillo, M., and Roux, P. (2015). Detecting and locating seismic events with using USArray as a large antenna. *Adv. Geosci.*, 40 :27–30.
- Rhie, J. and Romanowicz, B. (2004). Excitation of Earth's continuous free oscillations by atmosphere-ocean-seafloor coupling. *Nature*, 431(7008) :552–556.
- Ricker, N. (1953). THE FORM AND LAWS OF PROPAGATION OF SEISMIC WAVELETS. *GEOPHYSICS*, 18(1) :10–40.
- Ringdal, F. and Husebye, E. (1982). Application of arrays in the detection, location, and identification of seismic events. *Bull. Seismol. Soc. Am.*, 72(6B) :S201–S224.
- Romanowicz, B. A. (1979). Seismic structure of the upper mantle beneath the United States by three-dimensional inversion of body wave arrival times. *Geophysical Journal International*, 57(2) :479–506.
- Rost, S. (2002). Array seismology : Methods and applications. *Rev. Geophys.*, 40(3) :1008.
- Roult, G., Peltier, A., Taisne, B., Staudacher, T., Ferrazzini, V., and Muro, A. D. (2012). A new comprehensive classification of the piton de la fournaise activity spanning the 1985–2010 period. search and analysis of short-term precursors from a broad-band seismological station. *J. Volcanol. Geotherm. Res.*, 241-242 :78–104.

- Roux, P. (2009). Passive seismic imaging with directive ambient noise : application to surface waves and the san andreas fault in parkfield, CA. *Geophys. J. Int.*, 179(1) :367–373.
- Roux, P., Kuperman, W. A., and the NPAL Group (2004). Extracting coherent wave fronts from acoustic ambient noise in the ocean. *The Journal of the Acoustical Society of America*, 116(4) :1995–9.
- Roux, P., Sabra, K. G., Kuperman, W. A., and Roux, A. (2005). Ambient noise cross correlation in free space : Theoretical approach. *The Journal of the Acoustical Society of America*, 117(1) :79–84.
- Sabra, K., Gerstoft, P., Roux, P., Kuperman, W., and Fehler, M. (2005a). Surface wave tomography from seismic ambient noise in southern california. *The Journal of the Acoustical Society of America*, 117(4) :2431–2431.
- Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W. A., and Fehler, M. C. (2005b). Surface wave tomography from microseisms in southern california. *Geophys. Res. Lett.*, 32(14) :n/a–n/a.
- Sabra, K. G., Roux, P., and Kuperman, W. A. (2005c). Emergence rate of the time-domain green’s function from the ambient noise cross-correlation function. *The Journal of the Acoustical Society of America*, 118(6) :3524–3531.
- Samson, J. C. and Olson, J. V. (1980). Some comments on the descriptions of the polarization states of waves. *Geophys. J. Int.*, 61(1) :115–129.
- Sanchez-Sesma, F. J. (2006). Retrieval of the green’s function from cross correlation : The canonical elastic problem. *Bulletin of the Seismological Society of America*, 96(3) :1182–1191.
- Sato, M., Ishikawa, T., Ujihara, N., Yoshida, S., Fujita, M., Mochizuki, M., and Asada, A. (2011). Displacement above the hypocenter of the 2011 Tohoku-oki earthquake. *Science*, 332(6036) :1395–1395.
- Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propagat.*, 34(3) :276–280.
- Seats, K. J., Lawrence, J. F., and Prieto, G. A. (2011). Improved ambient noise correlation functions using Welch’s method. *Geophys. J. Int.*, 188(2) :513–523.
- Sengupta, A. M. and Mitra, P. P. (1999). Distributions of singular values for some random matrices. *Physical Review E*, 60(3) :3389–3392.
- Sens-Schönfelder, C. and Wegler, U. (2006). Passive image interferometry and seasonal variations of seismic velocities at merapi volcano, Indonesia. *Geophys. Res. Lett.*, 33(21).
- Seydoux, L., Shapiro, N., de Rosny, J., Brenguier, F., and Landès, M. (2016a). Detecting seismic activity with a covariance matrix analysis of data recorded on seismic arrays. *Geophys. J. Int.*, 204(3) :1430–1442.

- Seydoux, L., Shapiro, N., de Rosny, J., and Landes, M. (2016b). Spatial coherence of the seismic wavefield continuously recorded by the usarray. *Geophysical Research Letters* (Under review).
- Shapiro, N. M. and Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophys. Res. Lett.*, 31(7) :n/a–n/a.
- Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. H. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, 307(5715) :1615–1618.
- Shapiro, N. M., Ritzwoller, M. H., and Bensen, G. D. (2006). Source location of the 26 sec microseism from cross-correlations of ambient seismic noise. *Geophys. Res. Lett.*, 33(18) :L18310.
- Snieder, R. (2004). Extracting the green's function from the correlation of coda waves : A derivation based on stationary phase. *Physical Review E*, 69(4) :046610.
- Stehly, L., Campillo, M., and Shapiro, N. M. (2006). A study of the seismic noise from its long-range correlation properties. *J. Geophys. Res.*, 111(B10) :B10306.
- Tanimoto, T. (2001). Continuous free oscillations : Atmosphere-solid earth coupling. *Annu. Rev. Earth Planet. Sci.*, 29(1) :563–584.
- Tanter, M., Aubry, J.-F., Gerber, J., Thomas, J.-L., and Fink, M. (2001). Optimal focusing by spatio-temporal inverse filter. i. basic principles. *The Journal of the Acoustical Society of America*, 110(1) :37–11.
- Tanter, M., Thomas, J.-L., and Fink, M. (2000). Time reversal and the inverse filter. *The Journal of the Acoustical Society of America*, 108(1) :223.
- Tian, Y. and Ritzwoller, M. H. (2015). Directionality of ambient noise on the juan de fuca plate : implications for source locations of the primary and secondary microseisms. *Geophys. J. Int.*, 201(1) :429–443.
- Traer, J. and Gerstoft, P. (2014). A unified theory of microseisms and hum. *J. Geophys. Res. Solid Earth*, 119(4) :3317–3339.
- Traer, J., Gerstoft, P., Bromirski, P. D., and Shearer, P. M. (2012). Microseisms and hum from ocean surface gravity waves. *J. Geophys. Res.*, 117(B11) :n/a–n/a.
- Tsai, V. C. (2009). On establishing the accuracy of noise tomography travel-time measurements in a realistic medium. *Geophys. J. Int.*, 178(3) :1555–1564.
- Tulino, A. M. and Verdú, S. (2004). Random matrix theory and wireless communications. *Foundations and Trends™ in Communications and Information Theory*, 1(1) :1–182.
- Uieda, L., Jr, V. C. O., and Barbosa, V. C. F. (2013). Modeling the earth with fatiando a terra. In van der Walt, S., Millman, J., and Huff, K., editors, *Proceedings of the 12th Python in Science Conference*, pages 96 – 103.
- USGS (2010). M7.3, 7.6, and 7.4 moro gulf, mindanao, philippines, earthquakes of 23 july 2010. Technical report, U.S. Geological Survey.

- Veen, B. V. and Buckley, K. (1988). Beamforming : a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2) :4–24.
- Wagner, G. S. and Owens, T. J. (1996). Signal detection using multi-channel seismic data. *Bull. Seismol. Soc. Am.*, 86(1A) :221–231.
- Walsh, J. and Vekua, I. N. (1969). New methods for solving elliptic equations. *The Mathematical Gazette*, 53(384) :210.
- Wapenaar, K. (2004). Retrieving the elastodynamic green's function of an arbitrary inhomogeneous medium by cross correlation. *Phys. Rev. Lett.*, 93(25) :254301.
- Wapenaar, K., Slob, E., and Snieder, R. (2006). Unified green's function retrieval by cross correlation. *Phys. Rev. Lett.*, 97(23) :234301.
- Wax, M. and Kailath, T. (1985). Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2) :387–392.
- Wax, M., Shan, T.-J., and Kailath, T. (1984). Spatio-temporal spectral analysis by eigenstructure methods. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(4) :817–827.
- Weaver, R., Froment, B., and Campillo, M. (2009). On the correlation of non-isotropically distributed ballistic scalar diffuse waves. *The Journal of the Acoustical Society of America*, 126(4) :1817.
- Weaver, R. L. and Lobkis, O. I. (2001a). On the emergence of the green's function in the correlations of a diffuse field. *The Journal of the Acoustical Society of America*, 109(5) :2410–2410.
- Weaver, R. L. and Lobkis, O. I. (2001b). Ultrasonics without a source : Thermal fluctuation correlations at MHz frequencies. *Phys. Rev. Lett.*, 87(13) :134301.
- Webb, S. C. (2008). The earth's hum : the excitation of earth normal modes by ocean waves. *Geophysical Journal International*, 174(2) :542–566.
- Wegler, U. and Sens-Schönfelder, C. (2007). Fault zone monitoring with passive image interferometry. *Geophysical Journal International*, 168(3) :1029–1033.
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra : A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.*, 15(2) :70–73.
- Wu, C., Delorey, A., Brenguier, F., Hadziioannou, C., Daub, E. G., and Johnson, P. (2016). Constraining depth range of s-wave velocity decrease after large earthquakes near parkfield, california. *Geophys. Res. Lett.*, pages 1–20.
- Xia, Y., Ni, S., and Zeng, X. (2013). Twin enigmatic microseismic sources in the gulf of Guinea observed on intercontinental seismic stations. *Geophys. J. Int.*, 194(1) :362–366.
- Yang, Y., Ritzwoller, M. H., Levshin, A. L., and Shapiro, N. M. (2007). Ambient noise Rayleigh wave tomography across Europe. *Geophysical Journal International*, 168(1) :259–274.

Résumé

L'estimation des fonctions de Green par corrélation de bruit sismique ambiant repose entre autre sur l'hypothèse forte que les sources de bruit sont réparties de façon homogène dans le milieu. Or, les données sismiques réelles montrent que ces sources de bruit ne respectent pas ces conditions dans la plupart des cas. En particulier, les signaux très cohérents générés par des séismes tectoniques ou des sources de bruit directionnelles existent et sont néfastes à l'utilisation du bruit sismique en imagerie passive. Des techniques de traitement du signal temporelles et fréquentielles sont alors habituellement appliquées aux sismogrammes individuels afin d'améliorer la qualité de reconstruction des fonctions de Green. C'est dans la continuité des études visant à traiter les sismogrammes que ce travail s'inscrit. En s'inspirant de méthodes de traitement d'antenne, nous analysons la matrice de covariance de données enregistrées sur des réseaux de stations sismiques. À travers l'analyse des valeurs propres de cette matrice, nous montrons qu'il est possible de dégager des informations cruciales sur le champ d'onde mesuré, comme le degré de cohérence spatiale dans les dimensions temporelles et fréquentielles. Cette information est importante car elle permet d'identifier les domaines temps-fréquence du champ d'onde propice à l'estimation des fonctions de Green. Finalement, une méthode d'égalisation des valeurs propres de la matrice de covariance visant à atténuer l'anisotropie des sources est proposée. Cette dernière méthode est interprétée à la fois comme une extension de la normalisation spectrale à la dimension spatiale, et également dans l'analogie du retournement temporel, au filtre inverse passif qui a vu le jour récemment.

Summary

Green's function estimation from ambient seismic noise relies on the strong hypothesis that noise seismic sources are evenly distributed in the medium. Yet, observations of seismic data show that the noise sources do not provide such good conditions in real cases. Strongly coherent seismic sources or directional noise seismic sources may exist, and are harmful to the application of this in ambient seismic imaging. Several signal processing techniques are nowadays routinely applied to individual seismograms in frequency and temporal domain in order to improve the quality of reconstruction of the Green's function. The present work takes place in this context. Our approach is inspired by array-processing techniques, and is particularly focused on the covariance matrix of data recorded on seismic arrays. We show that the eigenstructure of this matrix provides crucial information about the seismic wavefield degree of coherence, as a function of time and frequency. This information is important because it allows to identify the time and frequency zones where the Green's function quality is ensured. An original array-processing technique is finally proposed, which consider the equalization of the covariance matrix eigenvalues, in order to attenuate the wavefield anisotropy such as earthquake-related signals or directive noise sources. We interpret this last method as an extension of the spectral whitening technique widely used in seismology to the spatial dimension encoded by the covariance matrix. We also invoke the analogy with the time-reversal technique that have recently led to a class of passive inverse filtering techniques.