



HAL
open science

Estimation des effets de site particuliers à partir du bruit sismique ambiant mesuré sur un réseau dense de capteurs

Loïc Gisselbrecht

► **To cite this version:**

Loïc Gisselbrecht. Estimation des effets de site particuliers à partir du bruit sismique ambiant mesuré sur un réseau dense de capteurs. Sciences de la Terre. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALU036 . tel-04431529

HAL Id: tel-04431529

<https://theses.hal.science/tel-04431529>

Submitted on 1 Feb 2024

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : STEP - Sciences de la Terre de l'Environnement et des Planètes

Spécialité : Sciences de la Terre et de l'Environnement

Unité de recherche : Institut des Sciences de la Terre

**Estimation des effets de site particuliers à partir du bruit sismique
ambiant mesuré sur un réseau dense de capteurs**

**Site effects estimation from ambient seismic noise measured on a
dense array of sensors**

Présentée par :

Loïc GISSELBRECHT

Direction de thèse :

Pierre BOUÉ
Enseignant-chercheur, UGA
Bérénice FROMENT
IRSN

Directeur de thèse

Co-encadrant de thèse

Rapporteurs :

Sébastien CHEVROT
DIRECTEUR DE RECHERCHE, CNRS délégation Occitanie Ouest
Luis Fabian BONILLA HIDALGO
DIRECTEUR DE RECHERCHE, Université Gustave Eiffel

Thèse soutenue publiquement le **5 décembre 2023**, devant le jury composé de :

Pierre BOUÉ MAITRE DE CONFERENCES, Université grenoble Alpes	Directeur de thèse
Sébastien CHEVROT DIRECTEUR DE RECHERCHE, CNRS délégation Occitanie Ouest	Rapporteur
Luis Fabian BONILLA HIDALGO DIRECTEUR DE RECHERCHE, Université Gustave Eiffel	Rapporteur
Cécile CORNOU DIRECTRICE DE RECHERCHE, IRD délégation Sud Est	Examinatrice
Thomas LECOCQ CHARGE DE RECHERCHE, Observatoire Royal de Belgique	Examineur
Helle PEDERSEN PHYSICIENNE, Université Grenoble Alpes	Présidente

Invités :

Bérénice Froment
DOCTEURE EN SCIENCES, Institut de Radioprotection et de Sûreté Nucléaire - IRSN
Vincent Perron
DOCTEUR EN SCIENCES, CEA Cadarache





Charte anti-plagiat

Art 1 : Définition du plagiat

Le plagiat consiste à reproduire un texte, une partie d'un texte, une illustration ou des idées originales d'un auteur, sans lui en reconnaître la paternité par un référencement bibliographique ou iconographique adéquat (Cf. art. 3).

Art 2 : Objet des travaux universitaires

Sont considérés comme travaux universitaires tous les documents réalisés par les étudiants et les enseignants, les chercheurs et les enseignants-chercheurs dans le cadre des activités de formation et de recherche. Ces travaux universitaires doivent toujours avoir pour ambition de produire un savoir inédit et d'offrir une lecture critique, nouvelle et personnelle du sujet.

Art 3 : Méthodologie de référencement bibliographique

La méthodologie d'un travail universitaire implique que les emprunts (par exemple par copier/coller) soient clairement identifiés et que le nom de l'auteur et la source de l'extrait soient mentionnés.

Les citations textuelles y compris dans une traduction personnelle, doivent obligatoirement être placées entre guillemets et être accompagnées d'une référence bibliographique à la suite de la citation, ou en note de bas de page.

Les emprunts non textuels (tableaux, graphiques, photos, formules scientifiques, etc.) doivent également être accompagnés d'une référence bibliographique à leur suite ou en note de bas de page.

En complément, toutes les références des documents cités, empruntés ou adaptés, doivent figurer en bibliographie.

Art 4 : Détection du plagiat

L'Université Grenoble Alpes est dotée d'un outil permettant de contrôler systématiquement les travaux universitaires et de détecter les similitudes, dans le but de rechercher le plagiat.

Art 5 : Sanctions disciplinaires pour plagiat

Les auteurs présumés de plagiat seront traduits devant la section disciplinaire compétente qui pourra prendre des sanctions pouvant aller jusqu'à l'exclusion définitive de tout établissement d'enseignement supérieur. La procédure disciplinaire n'exclut pas d'éventuelles poursuites judiciaires.

Art. 6 : Engagement

Les étudiants et les personnels s'engagent à ne pas commettre de plagiat dans leurs travaux universitaires. À cette fin, ils reconnaissent avoir pris connaissance des obligations décrites dans les articles 2 et 3 de la présente charte et s'engagent à s'y conformer.

Je certifie, Gisselbrecht Loïc (indiquer nom prénom)
avoir pris connaissance de la charte anti-plagiat et à la respecter.

Date et signature : 20/09/2023

Remerciements

Les thésards connaissent souvent des hauts et des bas. C'est d'ailleurs une des premières choses que les thésards en fin de thèse enseignent à ceux qui débutent la leur. Quelle que soit la période traversée, l'entourage du thésard est plus qu'important.

Mes premiers remerciements sont naturellement destinés à mes encadrants de thèse, Bérénice Froment et Pierre Boué, qui m'ont fait confiance sur ce projet ambitieux. J'aimerais saluer tout particulièrement le temps que vous avez dédié à mon encadrement et votre soutien inflexible tout au long de mon parcours de thèse. Je retiendrai notamment votre rigueur ainsi que votre goût communicatif pour la recherche scientifique. Travailler à vos côtés pendant 3 ans a été une expérience enrichissante et j'espère un jour pouvoir à nouveau travailler avec vous !

Je remercie en second lieu l'équipe du BERSSIN qui m'a accueillie. J'aurais aimé travailler un peu plus en « présentiel » avec vous mais un virus en a décidé autrement... Quoiqu'il en soit je retiendrai la très bonne atmosphère de travail ainsi que la solidarité qui règne dans ce bureau !

Je souhaite remercier ensuite l'ensemble de mon jury de thèse pour l'intérêt porté à mes travaux ; à mes rapporteurs Luis-Fabian Bonilla Hidalgo et Sébastien Chevrot ; à mes examinateurs, Cécile Cornou, Helle Pedersen et Thomas Lecocq. Un merci particulier à Vincent Perron pour avoir accepté de participer à ma soutenance en tant qu'invité.

Un grand merci à Céline Gélis et François Lavoué avec qui j'ai beaucoup interagi ainsi qu'aux autres membres du projet DARE qui se sont toujours rendus disponibles pour répondre à mes questions !

Enfin, j'ai évidemment une pensée pour ma famille et mes amis qui m'ont toujours encouragé. Je m'excuse d'avance si je ne vous nomme pas tous ici. A Audrey qui a partagé la froideur du bureau Kilohertz avec moi pendant les longs mois d'hiver. A Na'Omi pour sa bonne humeur éternelle. Au couz' Ben. Et bien sûr une ligne particulière en direction de Jolyanne, pour m'avoir appris à grimper et surtout pour tous les moments partagés !

Résumé

De nombreux travaux ont montré l'impact des couches géologiques superficielles sur le mouvement du sol lors d'un séisme. La modification du mouvement incident par ces couches est nommée « effet de site ». Certaines configurations géologiques, comme les anciennes vallées remplies de sédiments, sont propices à des effets de site significatifs pouvant mener à une augmentation de l'amplitude du mouvement sismique en surface et à l'allongement de sa durée. Il est donc important de prendre en compte ce phénomène dans l'évaluation de l'aléa sismique.

Depuis plusieurs années, l'IRSN mène des études pour améliorer l'estimation des effets de site, notamment via des approches empiriques. Ces approches nécessitent de nombreux enregistrements de séismes. En zone de sismicité faible à modérée, comme en France métropolitaine, le temps de retour des séismes modérés à forts est long, et donc limitant. Emerge alors l'intérêt du bruit sismique. Le bruit sismique correspond aux faibles vibrations de la surface de la Terre. Généré par des activités naturelles ou anthropiques, il a pour caractéristique d'être permanent. Ainsi, il constitue une source de données exploitable, même en région de faible sismicité.

Cette thèse exploite le bruit sismique enregistré sur un réseau de 400 capteurs nodes déployé dans la vallée du Tricastin (paléovallée du Rhône) dans le cadre du projet DARE (Dense ARray for seismic site effect Estimation, 2020-2023). Cette vallée est propice à générer des effets de site importants et complexes. Les objectifs sont d'étudier le potentiel du bruit sismique dans la caractérisation des effets de site, d'évaluer l'amplification associée à la vallée en cas de séisme à partir du bruit sismique et de proposer de nouvelles méthodes d'estimation de l'amplification.

Après un contrôle qualité des données ayant abouti à une première sélection des enregistrements de bruit sismique, l'amplification associée à la vallée du Tricastin a été estimée à travers le SSRn et le SSRh (méthode hybride combinant données de bruit et de séismes). Les résultats ont été comparés à des mesures SSR. A basse fréquence ($f < 1$ Hz), l'amplification estimée est comparable aux mesures SSR et est cohérente avec la structure de la vallée. A plus hautes fréquences ($f > 1$ Hz),

l'estimation de l'amplification est plus critique en raison de l'influence des sources anthropiques locales de bruit sismique (zone fortement industrialisée). Le SSRh, qui utilise un capteur de référence au site, fournit tout de même un facteur d'amplification comparable au SSR jusqu'à 4 Hz.

Pour retrouver l'amplification liée aux effets de site à plus haute fréquence, il faut limiter l'influence des sources locales. Un algorithme de clustering a été utilisé afin de sélectionner pour chaque enregistrement et de manière automatique, les fenêtres temporelles pour lesquelles l'activité des sources locales est moins forte. L'amplification estimée par le SSRh a ainsi pu être améliorée dans les zones impactées par des sources transitoires, mais pas dans les zones impactées par des sources permanentes. Ces résultats démontrent la difficulté d'implémenter une méthode se basant uniquement sur l'amplitude du bruit sismique en zone fortement industrialisée. Nous avons dès lors travaillé sur l'estimation de l'amplification à partir des déconvolutions de bruit sismique (ou ANIRF) calculées entre les paires de capteurs. L'idée est que l'information de phase que contient la déconvolution peut aider à filtrer le champ de bruit sismique et ainsi limiter l'impact des sources locales, notamment des sources permanentes qui sont plus difficiles à traiter. Les déconvolutions ont été calculées en utilisant 7 sources virtuelles distribuées autour de la zone. En retirant les signaux monochromatiques de forte amplitude et en sélectionnant la partie des déconvolutions la moins impactée (sélection temporelle), nous montrons qu'il est possible d'atténuer l'impact des sources permanentes jusqu'à environ 4 Hz. Les observations n'ont cependant pas pu être étendues à plus haute fréquence en raison du manque d'enregistrements de séismes de qualité suffisante.

Les résultats obtenus permettent de discuter les avantages et limitations des différentes approches développées et testées dans ce travail de thèse et d'ouvrir des perspectives sur des travaux futurs pour améliorer l'étude des effets de site à partir du bruit sismique dans les zones industrialisées.

Mots-clés : effets de site ; amplification ; bassin sédimentaire ; mouvement du sol ; bruit sismique ; réseau dense.

Abstract

Numerous studies have shown the impact of superficial layers on ground motion during an earthquake. The modification of the incident wavefield by these layers is known as "site effect". Certain geological structures are prone to significant site effects, which can lead to an increase in the amplitude of seismic motion at the surface and a lengthening of its duration. It is therefore important to take this phenomenon into account when assessing the seismic hazard.

The IRSN is conducting studies to improve the estimation of site effects, in particular by using empirical approaches. These approaches require a large number of earthquake recordings. In areas of low-to-moderate seismicity, as in mainland France, the return period of moderate-to-strong earthquakes is long, and therefore limiting. It is therefore of great interest to exploit the ambient noise. Ambient noise corresponds to weak vibrations of the Earth's surface. Generated by natural or anthropogenic activities, its characteristic feature is that it is permanent. Hence, ambient noise is exploitable even in areas of low seismicity.

This thesis exploits the ambient noise recorded on a 400 nodes array deployed in the Tricastin Valley (French Rhône Valley) as part of the DARE project (Dense ARray for seismic site effect Estimation, 2020-2023). This valley is prone to generate significant and complex site effects. The objectives are to study the potential of ambient noise in characterizing site effects, to evaluate the amplification associated with the valley in the event of an earthquake based on ambient noise, and to propose new methods for estimating amplification.

After performing a data quality control that led to an initial selection of ambient noise recordings, the amplification associated with the Tricastin Valley was estimated using SSRn and SSRh (a hybrid method combining noise and earthquake data). The results were compared with SSR measurements. At low frequency ($f < 1$ Hz), the estimated amplification is comparable to SSR measurements and is consistent with the expected geology of the valley. At higher frequencies ($f > 1$ Hz), amplification estimation is more critical due to the influence of local anthropogenic sources of

ambient noise (highly industrialized area). SSRh, which uses a reference sensor at the site, nevertheless provides an amplification factor comparable to SSR up to 4 Hz.

To recover the amplification associated with site effects at higher frequencies, the influence of local sources must be attenuated. A clustering algorithm was used to automatically select, for each recording, the time windows for which local sources are less active. The amplification estimated by SSRh was thus improved in areas impacted by transient sources, but not in areas impacted by permanent sources. These results show the difficulty of implementing a method based solely on ambient noise amplitude in highly industrialized areas. We therefore worked on estimating amplification from ambient noise deconvolutions (or ANIRF) calculated between sensor pairs. The idea is that the phase information contained in the deconvolution can help to filter the ambient noise wavefield and thus limit the impact of local sources, in particular permanent sources, which are more difficult to process. Deconvolutions were calculated using 7 virtual sources distributed around the study area. By removing high-amplitude monochromatic signals and selecting the least impacted part of the deconvolutions (temporal selection), we show that it is possible to attenuate the impact of permanent sources up to around 4 Hz. However, the observations could not be extended to higher frequencies due to the lack of earthquake recordings of sufficient quality.

The results obtained allow us to discuss the advantages and limitations of the different approaches developed and tested in this thesis work, and to open up prospects for future work to improve the study of site effects from ambient noise in industrialized areas.

Keywords: site effects; amplification; sedimentary basin; seismic motion; ambient noise; dense array.

Table des matières

Remerciements	1
Résumé	2
Abstract	4
Table des matières	6
Introduction générale	10
Chapitre 1 : Concepts physiques et méthodes	15
1.1 Introduction des effets de site en structure sédimentaire	15
1.1.1 L'effet stratigraphique 1D	15
1.1.2 Les effets 2D/3D associés à la géométrie de la structure	18
1.1.3 Conclusion sur les différents effets de site	21
1.2 Prise en compte des effets de site dans les études d'aléa	22
1.2.1 Généralités sur l'évaluation de l'aléa sismique	22
1.2.2 Différentes approches de prise en compte des effets de site dans les études d'aléa	23
1.3 Introduction au bruit sismique	25
1.3.1 Origine du bruit sismique	25
1.3.2 Nature du bruit sismique	27
1.4 Méthodes site-spécifiques d'estimation des effets de site	28
1.4.1 Le SSR (Standard Spectral Ratio)	28
1.4.2 Le SSRn (noise-based Standard Spectral Ratio)	29
1.4.3 Le SSRh (hybrid Standard Spectral Ratio)	30

Chapitre 2 : Acquisition des données à Tricastin, traitements préliminaires et observations des sources de bruit sismique.....	32
2.1 Contexte de l'étude	32
2.1.1 Formation de la vallée du Tricastin.....	32
2.1.2 Motivation d'une étude site-spécifique à Tricastin.....	33
2.1.3 Données de sismique active	33
2.1.4 Des enjeux importants dans la zone d'étude	34
2.1.5 Le séisme du Teil.....	36
2.2 Campagnes d'acquisition du projet DARE	38
2.2.1 Présentation des deux réseaux déployés	38
2.2.2 Modèle 3D de la vallée du Tricastin.....	39
2.3 Traitement des données de bruit sismique.....	41
2.3.1 Jeu de données des nodes	41
2.3.2 Fiabilité des nodes à basse fréquence	42
2.3.3 Contrôle qualité des données.....	44
Chapitre 3 : Evaluation des conditions d'application de deux méthodes site-sur-référence calculées à partir du bruit sismique dans le cas de la vallée du Tricastin	58
3.1 Introduction	58
3.2 Insights into the conditions of application of noise-based spectral ratios in a highly industrialized area: a case study in the French Rhone Valley	59
3.2.1 Abstract	59
3.2.2 Introduction	60
3.2.3 Context & data acquisition	64
3.2.4 Characterization of the local sources	68
3.2.5 Proposed workflow to mitigate local sources impact on amplification assessment.....	71
3.2.6 Spectral ratios computation & results.....	75
3.2.7 Discussion & conclusion	82
3.2.8 Acknowledgments.....	84
3.2.9 Data availability.....	84
3.2.10 Appendix	85

Chapitre 4 : Analyse préliminaire des corrélations de bruit sismique entre une source virtuelle et l'ensemble du réseau	86
4.1 Introduction	86
4.1.1 Principe de l'interférométrie par corrélation de bruit	86
4.1.2 Configurations permettant la convergence de la fonction de corrélation vers la fonction de Green	88
4.1.3 La distribution des sources de bruit sismique comme principale limitation de l'interférométrie par corrélation de bruit	89
4.1.4 Sur l'utilisation de l'amplitude des fonctions de corrélation	91
4.2 Corrélations de bruit sismique entre ADHE et l'ensemble du réseau	92
4.2.1 Calcul et analyse des corrélations brutes et blanchies.....	92
4.2.2 Analyse des signaux monochromatiques dans les fonctions de corrélation brutes	96
4.2.3 Détection et suppression des signaux monochromatiques	98
4.2.4 Analyse des corrélations traitées.....	101
4.3 Jeu de données utilisé pour le calcul de l'ANIRF	105
Chapitre 5 : Estimation de l'amplification associée à la vallée du Tricastin en utilisant l'ANIRF	109
5.1 Introduction	109
5.2 Amplitude de l'ANIRF à l'échelle du réseau à basse fréquence (entre 0.35 et 0.9 Hz).....	113
5.2.1 Propagation des fronts d'onde reconstruits à basse fréquence	113
5.2.2 Amplitude de l'ANIRF à l'échelle du réseau à basse fréquence d'aléa	124
5.3 Amplitude de l'ANIRF à l'échelle du réseau à haute fréquence (entre 3 et 6 Hz).....	127
5.3.1 Amplitude de l'ANIRF à l'échelle du réseau pour des temps d'arrivée particuliers	127
5.3.2 Amplitude de l'ANIRF à l'échelle du réseau à haute fréquence.....	131
5.4 Evaluation de l'amplification site/rocher à partir de l'ANIRF	135
5.4.1 Présentation de la procédure de calcul	135
5.4.2 Présentation des résultats	138
5.5 Discussion/conclusion	149

Conclusion générale	152
Bibliographie.....	157
Annexe	172

Introduction générale

Une des branches principales parmi le vaste champ d'études de la sismologie est la compréhension des séismes et de leur impact. Les séismes sont des phénomènes au potentiel destructeur en raison de la grande quantité d'énergie qu'ils libèrent. Cette énergie est générée au moment de la rupture brutale des roches de la croûte terrestre soumises à une accumulation de contraintes sur des temps longs. Elle se propage à travers la Terre sous forme d'ondes élastiques. Au cours de leurs trajets, ces ondes interagissent avec le milieu qu'elles traversent. Ainsi leur comportement (vitesse, direction, amplitude) varie en fonction des propriétés mécaniques du milieu traversé. En surface, des structures géologiques comme des bassins ou des vallées sédimentaires sont susceptibles de piéger les ondes et générer des effets locaux qui modifient le mouvement sismique (Figure I1). Cette modification du mouvement sismique en lien avec la géologie de surface est couramment appelée « effet de site ». Les effets de site se caractérisent souvent par une augmentation de l'amplitude des secousses sismiques (amplification) et par l'allongement de leur durée.

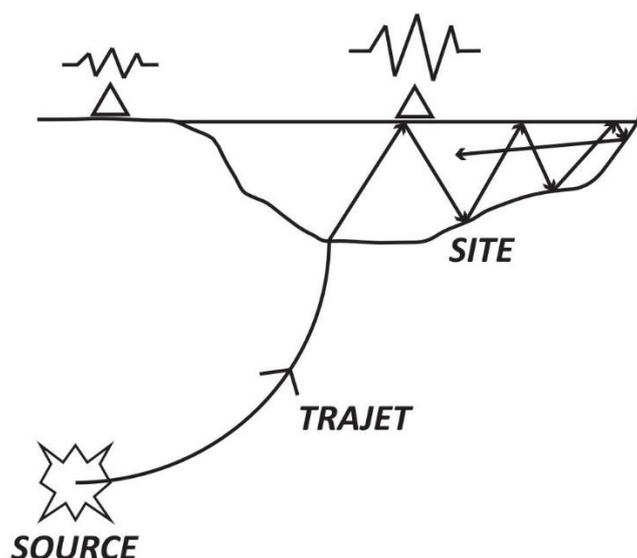


Figure I1 : Illustration des différents termes (SOURCE-TRAJET-SITE) qui contribuent au mouvement sismique en surface enregistré sur des stations sismologiques (triangles).

Depuis les années 80, un grand nombre d'études a montré l'impact des effets de site et leurs conséquences potentiellement dévastatrices (e.g. Bard et Bouchon 1985 ; Celebi *et al.* 1987 ; Campillo *et al.* 1989 ; Lermo et Chávez-García 1994 ; Kawase 1996 ; Huang et Teng 1999 ; Roten *et al.* 2006). Le séisme de Michoacán (Mexico, Mw = 8.2) en 1985 a été l'un des premiers révélateurs de l'influence des effets de site sur la gravité et la répartition des dégâts. La ville de Mexico, bien que localisée à environ 350 km de l'épicentre, a en effet été particulièrement impactée en raison du bassin sédimentaire sur laquelle elle est construite. Plus généralement, beaucoup de villes et de mégalopoles sont bâties sur ce genre de configurations sédimentaires, et sont donc potentiellement sujettes aux effets de site. De la conjonction d'une probabilité d'occurrence d'un séisme, des effets de propagation (incluant de potentiels effets de site) et des enjeux sociétaux (population, hôpitaux, sites nucléaires, usines ...) peut résulter un risque sismique local important. On comprend alors qu'il est capital, humainement et économiquement, de prendre en compte de tels effets dans l'évaluation de l'aléa sismique, comme le stipule notamment la Règle Fondamentale de Sécurité RFS 2001-01 pour les installations nucléaires.

Les effets de site sont généralement intégrés aux études d'aléa via des approches le plus souvent « ergodiques » considérant une classification du type de sol et/ou l'utilisation de proxys simples. La vitesse moyenne des ondes de cisaillement dans les 30 premiers mètres (V_{s30}) ou la fréquence de résonance fondamentale du site (f_0) sont les proxys les plus utilisés. Cependant, ces proxys se cantonnent généralement à décrire le comportement de milieux géologiques simples, typiquement des empilements sédimentaires plans (tabulaires). Dès que la géologie se complexifie, par exemple en considérant la géométrie dans les 3 dimensions d'une paléovallée, née de phénomènes érosifs, ces proxys ne suffisent plus pour décrire totalement les effets locaux du site. La transposition directe de proxys calibrés pour un site vers un autre site est également discutable. Des études dites « spécifiques » au site sont alors nécessaires. Parmi ces études site-spécifiques, deux approches sont généralement utilisées. L'une est basée sur la simulation du mouvement sismique à la surface à partir d'une modélisation numérique de la propagation des ondes dans un modèle du sous-sol (approche numérique). L'autre, est basée sur une mesure directe des propriétés que l'on souhaite évaluer ; à savoir typiquement, l'amplification et ses caractéristiques spectrales, ou encore la durée de l'ébranlement (approche empirique). Ces deux approches sont complémentaires et présentent des avantages et des limitations qu'il convient de prendre en compte dans leur mise en œuvre. L'approche numérique est par exemple dépendante du modèle au sein duquel les simulations sont réalisées et nécessite ainsi une très bonne connaissance de la structure géologique afin de

construire un modèle le plus réaliste possible. L'approche empirique nécessite en revanche de nombreux enregistrements de séismes afin d'obtenir une mesure statistiquement robuste. En zone de sismicité faible à modérée (comme en France métropolitaine), cela signifie des temps d'acquisition longs et représente donc une contrainte pour les besoins de l'ingénierie (constructions, dimensionnement des infrastructures...).

Dans les zones de faible sismicité l'intérêt de l'utilisation du bruit sismique pour l'estimation des effets de site est étudié depuis de nombreuses années (e.g. Kagami *et al.* 1982, 1986 ; Field *et al.* 1990 ; Rovelli *et al.* 1991 ; Lermo et Chávez-García 1994 ; Ferritto 1995 ; Zhao *et al.* 1998 ; Horike *et al.* 2001 ; Roten *et al.* 2006). Faibles vibrations de la surface de la Terre émises par des sources naturelles ou anthropiques, le bruit sismique présente le double avantage d'être permanent et observable partout sur Terre. Son enregistrement peut donc être réalisé partout, en tout temps. La quantité de données recueillies peut donc être suffisante, quel que soit l'endroit, en des temps d'acquisition relativement courts (y compris en zone de faible sismicité). En parallèle des méthodes introduites utilisant le bruit sismique, les progrès liés à l'instrumentation et aux capacités de calculs numériques ont permis de déployer des réseaux de capteurs de plus en plus denses, et d'accéder ainsi à une meilleure connaissance des propriétés du sous-sol, tout en limitant les coûts.

Au sein de l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN), le BERSSIN (Bureau d'Évaluation des Risques Sismiques pour la Sûreté des Installations Nucléaires) mène depuis plusieurs années des études de recherche et développement sur les effets de site et leur prise en compte dans l'évaluation de l'aléa sismique autour des sites nucléaires français. Compte-tenu de l'intérêt évident du bruit sismique en France métropolitaine, le BERSSIN a lancé des travaux ces dernières années autour de son utilisation pour l'étude des effets de site. C'est dans ce contexte que le projet franco-allemand DARE (Dense ARray for seismic site effect Estimation) a été lancé en 2019. Ce projet, financé par les agences nationales de recherche française et allemande (ANR, DFG) et par l'IRSN, a impliqué quatre partenaires (IRSN, Université de Potsdam, Université Grenoble Alpes et le GFZ Potsdam) et a duré 42 mois (avril 2020 - septembre 2023). A travers le projet DARE, la vallée sédimentaire du Tricastin (située dans la paléovallée du Rhône au sud-est de la France) a été instrumentée sur une zone d'environ 10 km par 10 km. Présentant un fort contraste lithologique entre le remplissage sédimentaire et le substratum rocheux ainsi qu'une géométrie incisée, la vallée du Tricastin est en proie à générer des effets de site importants et complexes. La zone d'étude comprend de surcroît plusieurs infrastructures à risque dont les

installations nucléaires de Tricastin (TNS : Tricastin Nuclear Site) ainsi que 5 villes pour une population totale de 45 000 habitants. Deux campagnes d'acquisition ont permis le déploiement : - d'un réseau d'environ 400 capteurs de type « nodes » (c.-à-d. des géophones autonomes) visant à enregistrer le bruit sismique pendant 1 mois - d'un réseau d'environ 50 capteurs large-bande visant à enregistrer l'activité sismique locale et régionale pendant 8 mois. L'objectif du projet est d'utiliser la complémentarité des deux réseaux pour caractériser les effets de site à l'échelle de la vallée via des approches numériques et empiriques et proposer de nouvelles méthodes visant à améliorer l'estimation des effets de site, en particulier en zone de sismicité faible à modérée.

Cette thèse fait partie du volet empirique¹ du projet DARE et utilise les données de bruit sismique issues du réseau de 400 nodes. Dans ce travail, on s'intéresse à évaluer le potentiel du bruit sismique dans l'estimation de l'amplification associée aux effets de site en structure sédimentaire. Les études traitant de l'utilisation du bruit sismique pour estimer l'amplification associée aux effets de site montrent en général des limitations au-delà de 1 Hz liées à la proximité des sources anthropiques de bruit sismique par rapport au site d'étude (e.g. Field *et al.* 1990 ; Zhao *et al.* 1998 ; Bard 1999 ; Horike *et al.* 2001). Il est cependant important d'avoir une mesure robuste de l'amplification sur une gamme de fréquences large, typiquement jusqu'à 10 Hz pour les besoins de l'ingénierie (évaluation de la réponse des bâtiments et infrastructures). Récemment, une méthode hybride combinant données de bruit sismique et de séismes a été introduite (SSRh, hybrid Standard Spectral Ratio ; Perron *et al.* 2018). Cette méthode a été appliquée à 3 sites sédimentaires différents (Provence, au sud de la France ; Argostoli en Grèce ; Sion, dans la vallée suisse du Rhône ; Perron *et al.* 2018, 2022) et a montrée des résultats prometteurs.

Les objectifs de ce travail de thèse sont de mieux cerner les conditions d'application des méthodes utilisant le bruit sismique pour l'estimation de l'amplification, d'évaluer l'amplification associée à la vallée du Tricastin à partir du bruit sismique et de proposer des approches se basant sur le bruit sismique qui seraient potentiellement généralisables à d'autres sites. Le cas d'étude de la vallée du Tricastin permettra notamment d'étendre les travaux de Perron *et al.* (2018, 2022) en appliquant le SSRh dans une zone fortement industrialisée.

¹ Concernant le volet numérique, le lecteur pourra notamment se référer aux premiers travaux en cours de publication à ce jour : Lavoué *et al.* (soumis).

Ce manuscrit présente le travail réalisé et s'articule comme suit :

- le chapitre 1 introduit tout d'abord les concepts physiques qui régissent les effets de site en structure sédimentaire avant de présenter la façon dont ces effets sont pris en compte dans les études d'aléa. L'origine et la nature du bruit sismique sont ensuite discutées. Enfin, 3 méthodes « site-sur-référence » d'estimation de l'amplification sont introduites : le SSR (Standard Spectral Ratio) qui se base sur des données de séismes ; le SSRn (noise-based Standard Spectral Ratio) qui se base sur des données de bruit sismique ; le SSRh qui combine données de séismes et de bruit sismique ;
- le chapitre 2 présente les données et leur acquisition. Dans un premier temps, le contexte géologique de la zone d'étude ainsi que les deux campagnes d'acquisition sont introduits. Un contrôle qualité des données de bruit sismique est ensuite conduit. Ce contrôle qualité a été présenté dans Froment *et al.* (2022a) et permet notamment de commencer à discuter de l'impact des sources locales de bruit sismique en vue des calculs d'amplification ;
- le chapitre 3 reprend un article publié dans Geophysical Journal International (GJI ; Gisselbrecht *et al.* 2023). Cet article introduit une procédure de sélection automatique des fenêtres de temps visant à améliorer la mesure d'amplification fournie par le SSRn et le SSRh et discute de leur applicabilité en zone fortement industrialisée ;
- enfin, les chapitres 4 et 5 discutent de la possibilité d'utiliser les méthodes dérivées de l'interférométrie (ou corrélations) de bruit, en particulier l'ANIRF (e.g. Prieto et Beroza 2008 ; Prieto *et al.* 2011 ; Denolle *et al.* 2013, 2014 ; Viens *et al.* 2014, 2015, 2016 ; Bowden 2015 ; Perron 2017 ; Denolle *et al.* 2018) pour estimer l'amplification associée aux effets de site. L'ANIRF (Ambient Noise Impulse Response Function) est introduite dans le chapitre 4 ainsi que les traitements précédant son utilisation. Le chapitre 5 présente une procédure d'estimation de l'amplification à partir de l'ANIRF et discute des résultats ainsi obtenus.

Chapitre 1 : Concepts physiques et méthodes

1.1 Introduction des effets de site en structure sédimentaire

Comme présenté dans l'introduction, les effets de site sont les conséquences d'une configuration géologique locale sur le champ d'ondes sismiques incident, se propageant depuis un séisme. Dans le cas des structures sédimentaires, on distingue habituellement deux types d'effets de site : l'effet stratigraphique 1D (vertical), associé au contraste de propriétés mécaniques entre les sédiments et le substratum rocheux ; les effets 2D/3D associés à la géométrie de la structure. Ces deux types d'effets de site sont ici introduits de manière distincte (respectivement en section 1.1.1 et 1.1.2).

1.1.1 L'effet stratigraphique 1D

Une structure sédimentaire tabulaire peut être schématisée comme une couche de sol plane reposant sur le substratum rocheux. L'amplification du mouvement sismique au sein de la couche de sol résulte à la fois de l'augmentation de l'amplitude de l'onde transmise dans la couche et du phénomène de résonance. Ces effets sont les conséquences physiques directes du contraste de propriétés mécaniques à l'interface des deux milieux.

Augmentation de l'amplitude de l'onde transmise dans la couche de sol

On fait souvent l'approximation que les ondes provenant d'un séisme en profondeur remontent vers la surface en traversant une succession de couches homogènes et isotropes. Leur vitesse de propagation diminue généralement entre les couches le long du trajet vers la surface² et, par conservation de l'énergie, leur amplitude augmente. On peut caractériser cette augmentation de l'amplitude à travers deux couches par les

² La vitesse de propagation des ondes est fonction de la résistance à la déformation du milieu. Pour une onde de cisaillement par exemple (onde S), la vitesse de propagation est $\sqrt{(\mu/\rho)}$, avec μ le module de cisaillement (mesure de la résistance au cisaillement) et ρ la densité du milieu traversé. S'il est vrai que ρ diminue généralement le long du trajet de l'onde jusqu'à la surface, il diminue moins rapidement que μ . Ainsi, la vitesse de propagation de l'onde diminue en moyenne jusqu'à la surface.

coefficients de transmission et réflexion, respectivement notés T et R . Considérons le cas d'une couche de sol (Figure 1.1) d'épaisseur h définie par les paramètres de densité et de vitesse des ondes S (ρ_1, V_{S1}) surplombant un demi-espace (le substratum rocheux) défini par (ρ_2, V_{S2}). Une onde plane arrive à incidence quasi-normale sur la couche de sol du fait de la diminution de la vitesse et de la loi de Snell. Pour satisfaire les conditions de continuité à l'interface, l'amplitude de l'onde transmise doit être égale à la somme de l'amplitude de l'onde incidente et réfléchie, d'où : $T = 1 + R$. Partant de cette relation très simple, il est possible de démontrer que $T = \frac{2C}{C+1}$, avec C le contraste d'impédance à l'interface défini comme $\frac{\rho_2 V_{S2}}{\rho_1 V_{S1}}$. Comme expliqué précédemment, généralement $\rho_1 \ll \rho_2$ et $V_{S1} \ll V_{S2}$, impliquant $C \gg 1$. Le contraste d'impédance prévoit ainsi que l'amplitude de l'onde transmise dans la couche de sol peut aller jusqu'à doubler.

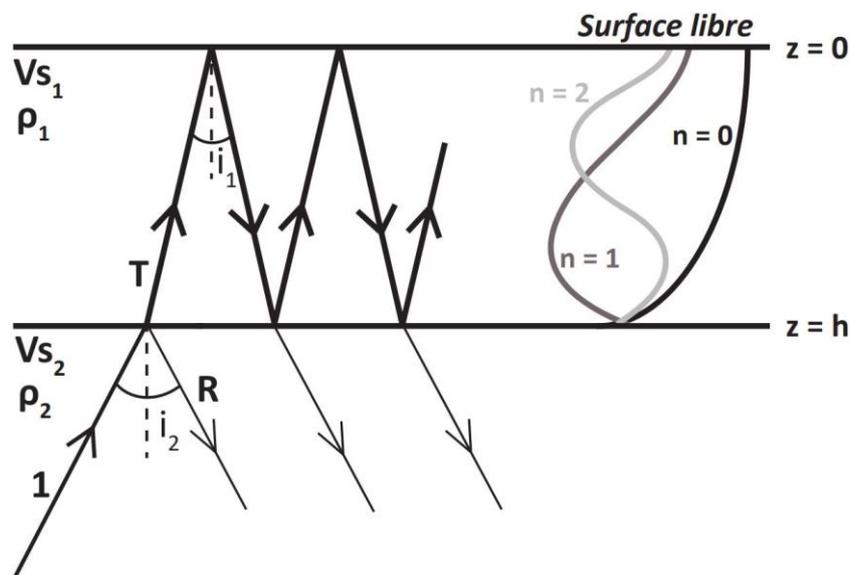


Figure 1.1 : Illustration de l'effet d'une couche de sol surplombant un demi-espace sur une onde plane incidente. ρ , V_s et i définissent respectivement la densité, la vitesse des ondes S ainsi que l'angle d'incidence pour chaque milieu. Les coefficients de transmission et réflexion sont respectivement notés T et R . L'onde montante et descendante donnent naissance à une onde stationnaire dont le mode fondamental et les deux premiers harmoniques sont représentés.

Phénomène de résonance verticale

Si le contraste d'impédance explique que l'amplitude de l'onde transmise dans la couche de sol est plus forte que l'amplitude de l'onde incidente, il explique aussi que l'onde réfléchie à la surface libre a du mal à se propager dans le demi-espace. Il en

résulte un piégeage partiel de l'énergie dans la couche de sol à l'origine de réflexions multiples. La superposition d'ondes montantes et descendantes donne naissance à une onde stationnaire : un phénomène de résonance verticale apparaît (figures 1.1 et 1.2). La résonance survient pour des longueurs d'onde caractéristiques et donc des fréquences caractéristiques qui sont fonction des propriétés de la couche de sol (Haskell 1960 ; Kramer 1996) :

$$f_n = (2n + 1) \frac{v_{s1}}{4h}, \quad (1)$$

avec n les modes de résonance (fondamental : $n = 0$; 1^{er} harmonique : $n = 1$; ...). L'amplification associée est donnée par la fonction d'amplification $A(f_n)$, définie comme le rapport dans le domaine fréquentiel entre le mouvement à la surface libre de la couche de sol et un mouvement de référence à la surface libre d'un modèle sans couche de sol. Dans le cas d'une couche de sol viscoélastique surplombant un demi-espace lui aussi viscoélastique, on peut montrer :

$$A(\omega_n) = \frac{2c}{2 + c\pi\xi(2n+1)}, \quad (2)$$

avec ω_n la pulsation des modes de résonance ($\omega_n = 2\pi f_n$). $A(\omega_n)$ possède des maxima aux fréquences de résonance. Plus le contraste d'impédance est important, plus les maxima sont grands. Ces maxima diminuent avec la fréquence en raison du phénomène d'amortissement, caractérisé par le paramètre ξ (eq. 2, Figure 1.2).

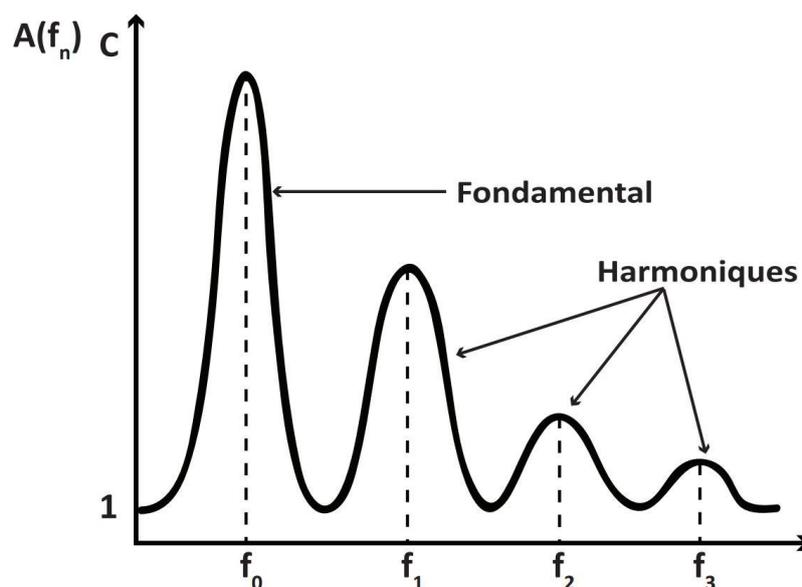


Figure 1.2 : Fonction d'amplification 1D entre le mouvement à la surface libre de la couche de sol et un mouvement de référence à la surface libre d'un modèle sans couche de sol.

1.1.2 Les effets 2D/3D associés à la géométrie de la structure

En réalité, les structures sédimentaires présentent bien souvent une géométrie non tabulaire, plus ou moins encaissée, caractérisée par les variations tridimensionnelles de l'interface sol/substratum. La géométrie de la structure peut engendrer la coexistence de différents types d'ondes ainsi que des propagations d'ondes complexes (Figure 1.3). A l'effet stratigraphique 1D vu précédemment s'ajoutent alors des effets 2D/3D. Ces effets 2D/3D, étudiés depuis de nombreuses années (e.g. Bard et Bouchon 1980 ; Bard et Bouchon 1985 ; Graves 1993 ; Kawase 1996 ; Graves *et al.* 1998 ; Cornou *et al.* 2003a et b ; Somerville *et al.* 2004 ; Roten *et al.* 2006 ; Maufroy *et al.* 2017), sont souvent à l'origine d'une amplification du mouvement sismique sur une large gamme de fréquences. Les principaux effets 2D/3D sont ici présentés de manière succincte.

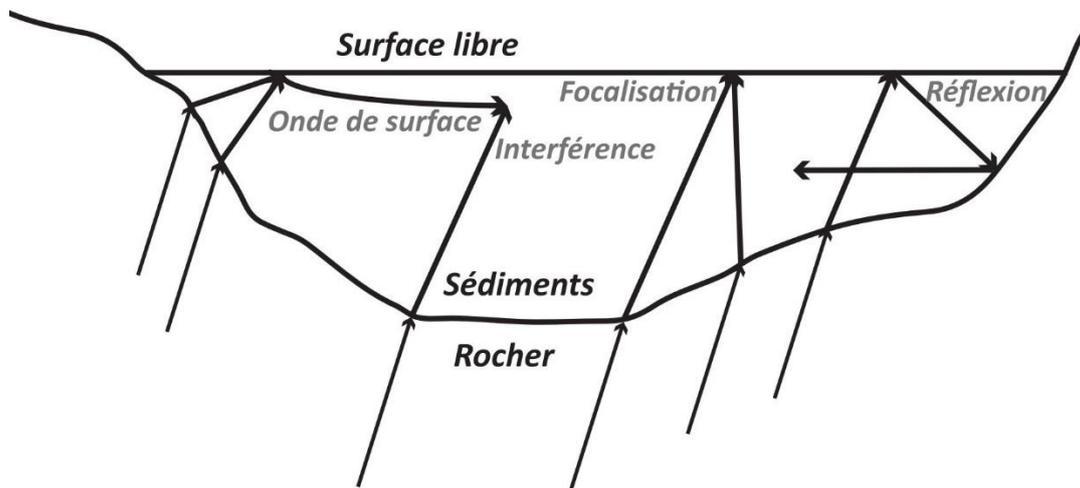


Figure 1.3 : Exemples d'effets 2D/3D liés à la géométrie de la structure sédimentaire.

Génération d'ondes de surface depuis les bords de la structure

Les bords inclinés de la structure sédimentaire ont tendance à rediriger les ondes de volume incidentes depuis un séisme en profondeur avec une trajectoire rentrante. Ceci favorise le jeu des interférences constructives au sein de la structure. Une partie des ondes de volume peut ainsi être convertie en ondes de surface (e.g. Aki et Richards 1980). Par exemple, la superposition d'ondes SH totalement réfléchies à l'interface sol/substratum (réflexion post-critique) engendre une onde de Love se propageant horizontalement dans la structure (e.g. Somerville *et al.* 2004). La superposition d'une onde P et d'une onde SV de même vitesse horizontale apparente engendre quant à elle une onde de Rayleigh. Les ondes de surface sont généralement de forte amplitude et,

de surcroît, s'atténuent moins rapidement avec la distance que les ondes de volume³. Il en résulte une amplification plus importante et un allongement significatif de la durée du mouvement sismique (e.g. Maufroy *et al.* 2017).

La génération d'ondes de surface depuis les bords de la structure peut fortement varier selon la position du séisme et l'angle d'incidence des ondes de volume. Gélis *et al.* (2008) ont par exemple illustré cet effet pour le bassin sédimentaire de Nice. Dans cette étude, un modèle de bassin 3D construit sur la base de données géotechniques a été soumis à une onde plane pour différents angles d'incidence (normale ; -35° ; +35°). Premièrement, les résultats ont montré une amplification plus importante pour un angle d'incidence incliné (-35° ou +35°) que pour une incidence normale. Deuxièmement, l'amplification est d'autant plus importante du côté du bassin où arrive l'onde : lorsque l'onde arrive du côté sud-est (incidence de -35°), l'amplification est plus importante du côté sud-est du bassin ; lorsque l'onde arrive du côté sud-ouest (incidence de +35°), l'amplification est plus importante du côté sud-ouest du bassin. Cette amplification concerne une large bande de fréquences, ce qui est caractéristique de la signature des ondes de surface et montre ainsi la variabilité de la génération d'ondes de surface selon l'angle d'incidence des ondes de volume.

Dans des cas de séismes proches du site, le mécanisme de rupture peut également influencer la génération d'ondes de surface et ainsi l'importance des amplifications. Olsen *et al.* (2006) ont par exemple simulé le mouvement sismique dans la région de Los Angeles pour deux scénarios de rupture de la faille de San Andreas (Figure 1.4). Dans le premier scénario (a), le sens de propagation de la rupture est SE-NW, dans le deuxième (b) NW-SE. Pour les deux scénarios, la distribution spatiale du PGV (Peak Ground Velocity) est représentée. La comparaison des deux scénarios révèle un mouvement sismique plus important au nord de la région de Los Angeles pour un sens de propagation de la rupture vers le nord (a), ce qui est cohérent avec des effets de directivité et donc globalement attendu. En revanche, l'aggravation de l'amplification dans le bassin de Los Angeles due à cet effet de directivité est remarquable. Olsen *et al.* (2006) expliquent ce résultat par la capacité de la directivité de la rupture à exciter des ondes de surface (ici Love) dans le bassin.

³ L'atténuation géométrique induit une décroissance d'amplitude en $1/r$ pour les ondes de volume, r étant la distance parcourue par l'onde. Contrairement aux ondes de volume, la divergence géométrique des ondes de surface n'est pas sphérique mais cylindrique (elles sont guidées à la surface de la Terre). Leur amplitude décroît donc en $1/\sqrt{r}$.

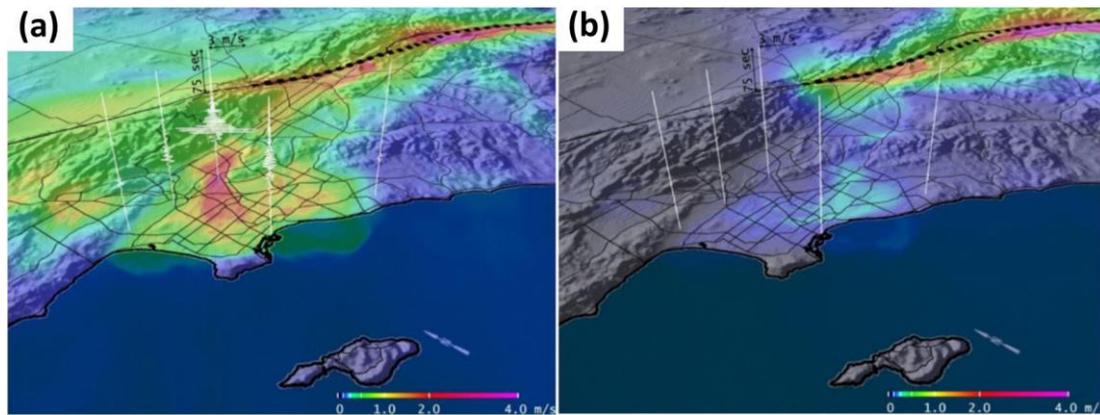


Figure 1.4 : Distribution spatiale du PGV dans la région de Los Angeles pour deux scénarios de rupture de la faille de San Andreas : (a) SE-NW ; (b) NW-SE. Olsen *et al.* (2006).

Interférences ondes de surface/ondes de volume

Plusieurs effets de site peuvent se combiner et apporter une amplification additionnelle. Des jeux d'interférences complexes entre ondes de surface et ondes de volume peuvent notamment se produire. Kawase (1996) montre par exemple que l'amplitude des secousses dans le bassin de Kobe durant le séisme de Hyogo-Ken Nanbu (1995) s'explique en partie par des interférences constructives entre des ondes S à incidence subverticale et des ondes de surface diffractées se propageant horizontalement depuis les bords du bassin. Ce type d'interférence, communément appelé « basin-edge effect », aggrave l'amplification et la durée du mouvement sismique aux bords de la structure sédimentaire (e.g. Narayan 2005 ; Hallier *et al.* 2008). King et Tucker (1984), puis Bard et Bouchon (1985) mettent quant à eux en évidence un phénomène de résonance 2D/3D. La résonance 2D/3D résulte de la combinaison entre les interférences latérales des ondes de surface et les interférences verticales des ondes de volume. La condition d'existence de ce phénomène est principalement contrôlée par le rapport d'aspect de la structure (rapport entre l'épaisseur sédimentaire et la largeur de la structure). Ainsi, ce type de résonance intervient en général pour des structures encaissées, dont la largeur est comparable à l'épaisseur sédimentaire⁴. Les motifs de résonance 2D/3D font intervenir la structure dans son ensemble, provoquant de larges amplifications, tant en termes d'amplitude que de durée du mouvement sismique.

⁴ Pour des structures allongées, les ondes de surface générées aux bords arrivent bien après les principales réflexions verticales des ondes de volume au centre de la structure. Les ondes de surface et la résonance des ondes de volume sont alors deux phénomènes séparés et la résonance 2D/3D n'a pas lieu.

Effet de focalisation des ondes

La variabilité et la multiplication des trajets au sein de la structure sédimentaire peuvent amener les ondes à converger simultanément vers certaines zones. Cet effet de focalisation est responsable d'une grande concentration de l'énergie et peut ainsi localement aggraver l'amplification du mouvement sismique. Lors du séisme de Northridge (1994) par exemple, la ville de Santa Monica (bassin de Los Angeles) a été particulièrement impactée par un effet de focalisation des ondes sismiques. Dans ce cas précis, la présence d'une structure convexe en profondeur formée par la base du bassin de Los Angeles et la faille de Santa Monica a redirigé, à la manière d'une lentille, les rais sismiques sous la ville (e.g. Gao *et al.* 1996 ; Davis *et al.* 2000). L'effet de focalisation s'est ici produit pour un faisceau de rais contenus dans un intervalle restreint d'angles d'incidence et explique la concentration anormale des dégâts à Santa Monica en comparaison des villes voisines.

1.1.3 Conclusion sur les différents effets de site

Dans les structures sédimentaires non tabulaires, les ondes sismiques provenant d'un séisme en profondeur subissent des effets 1D (verticaux) liés au contraste de propriétés mécaniques à l'interface sédiments/substratum et des effets 2D/3D liés à la géométrie de la structure. Ces effets de site peuvent varier selon les paramètres du séisme (position, mécanisme de rupture...) et l'angle d'incidence. Ils sont à l'origine d'une propagation d'ondes complexe au sein de la structure et sont souvent responsables d'une amplification importante des secousses sismiques. Maintenant que les effets de site ont été introduits, il est intéressant de comprendre comment ils sont pris en compte dans les études d'aléa.

1.2 Prise en compte des effets de site dans les études d'aléa

1.2.1 Généralités sur l'évaluation de l'aléa sismique

Au cours de la seconde moitié du 20^{ème} siècle, l'expansion des villes, la construction de grands ouvrages, le développement de l'industrie nucléaire ont augmenté le risque associé aux tremblements de terre. Les structures doivent être pensées pour résister aux futurs séismes susceptibles de se produire dans la zone sur la totalité de leur durée de vie (plusieurs dizaines d'années). Il faut donc, avant de construire, évaluer les mouvements sismiques auxquels elles peuvent être soumises : c'est le domaine de l'évaluation de l'aléa sismique.

L'évaluation de l'aléa sismique se base sur la connaissance des séismes passés et nécessite deux étapes : l'identification des sources sismiques et l'estimation du mouvement associé. La première étape repose en partie sur la construction de catalogues de sismicité les plus exhaustifs possibles. Ces catalogues comprennent la sismicité historique et instrumentale et renseignent notamment sur leur magnitude et localisation. La collecte de ces informations associée à des données géologiques provenant notamment de l'étude des failles actives sert de fondement pour définir ensuite un zonage sismotectonique au sein duquel chaque entité (zone) est considérée homogène d'un point de vue du potentiel sismique (on considère que tous les séismes qui se sont produits dans une zone peuvent s'y reproduire à l'avenir en n'importe quel endroit de la zone). Concernant la seconde étape (estimation du mouvement sismique en surface), elle repose la plupart du temps sur des équations de prédiction du mouvement (GMPEs). Ces GMPEs sont des relations empiriques définies à partir de grandes bases de données et permettent la prédiction d'indicateurs du mouvement à partir de paramètres simples tels que la magnitude et la distance d'un séisme.

Deux approches sont communément utilisées pour évaluer l'aléa sismique : l'approche probabiliste et déterministe. L'approche probabiliste s'intéresse à l'ensemble des séismes passés dans la zone en tenant compte de leur probabilité d'occurrence (estimation du taux d'occurrence annuel pour chaque magnitude). Cette approche permet d'estimer la probabilité de dépassement d'un mouvement cible sur une certaine période de temps. L'approche déterministe au contraire, prend en compte un séisme de référence : c'est le séisme de la zone qui produit les intensités et mouvements les plus pénalisants. Cette approche consiste ensuite à déterminer le mouvement sismique au site en considérant un scénario dans lequel les paramètres du séisme de référence sont souvent majorés (augmentation de la magnitude, diminution

de la distance au site). En France, l'approche déterministe constitue la règle en vigueur pour les installations nucléaires, comme la stipule la RFS 2001-01 (Règle Fondamentale de Sûreté). Quelle que soit l'approche considérée, l'objectif est d'évaluer un mouvement sismique de référence, susceptible de se produire à l'avenir dans la zone. Ce mouvement de référence⁵ est ensuite utilisé comme niveau d'aléa sismique à considérer dans l'ingénierie civil pour établir et/ou vérifier le dimensionnement des ouvrages.

1.2.2 Différentes approches de prise en compte des effets de site dans les études d'aléa

Au début des années 70, plusieurs auteurs reconnaissent la nécessité de prendre en compte les conditions de sol dans l'évaluation de l'aléa sismique, montrant notamment des accélérations plus importantes du mouvement sismique pour les sols meubles. Par la suite, des séismes particulièrement marquants où les conditions de sol ont joué un rôle prépondérant dans la gravité et la répartition des dégâts (comme celui de Michoacán en 1985 ; e.g. Campillo *et al.* 1989) se sont produits et au cours des années 80 et 90, les règles de constructions parasismiques ont progressivement intégré ces conditions de sol (e.g. Borchardt 1994).

Comme mentionné plus haut, les études d'aléa reposent en général sur des équations de prédiction du mouvement sismique (GMPEs). Lorsque les conditions de sol sont prises en compte dans les GMPEs, elles s'intègrent le plus souvent dans une approche dite « ergodique » via une classification du type de sol et/ou l'utilisation de proxys simples. La vitesse moyenne des ondes de cisaillement dans les 30 premiers mètres (V_{s30}) ou la fréquence de résonance fondamentale du site (f_0) sont les proxys les plus utilisés. Cependant, il est désormais établi que ces proxys ne peuvent pas rendre compte des effets du site dans leur intégralité (et complexité). Ceci est particulièrement vrai dans le cas d'une structure sédimentaire présentant une géométrie complexe pouvant entraîner des effets 2D/3D (voir section 1.1.2). Dans ces contextes, des études dites « spécifiques » au site sont donc recommandées afin d'appréhender correctement cette complexité (Figure 1.5).

Deux approches sont généralement adoptées dans les études site-spécifiques. La première approche repose sur l'estimation empirique (ou expérimentale) des effets de site par enregistrement des mouvements sismiques sur le site en question. La

⁵ Le mouvement de référence étant fonction de la fréquence d'oscillation, il est souvent défini sous la forme d'un spectre de réponse.

deuxième approche est basée sur la simulation du mouvement sismique à la surface à l'aide d'une modélisation numérique de la propagation des ondes dans un modèle du sous-sol. Ces deux approches sont complémentaires et présentent des avantages et des limitations qu'il convient de prendre en compte dans leur mise en œuvre. Comme mentionné dans l'introduction du manuscrit, cette thèse s'inscrit dans le cadre d'une approche empirique de l'estimation des effets de site.

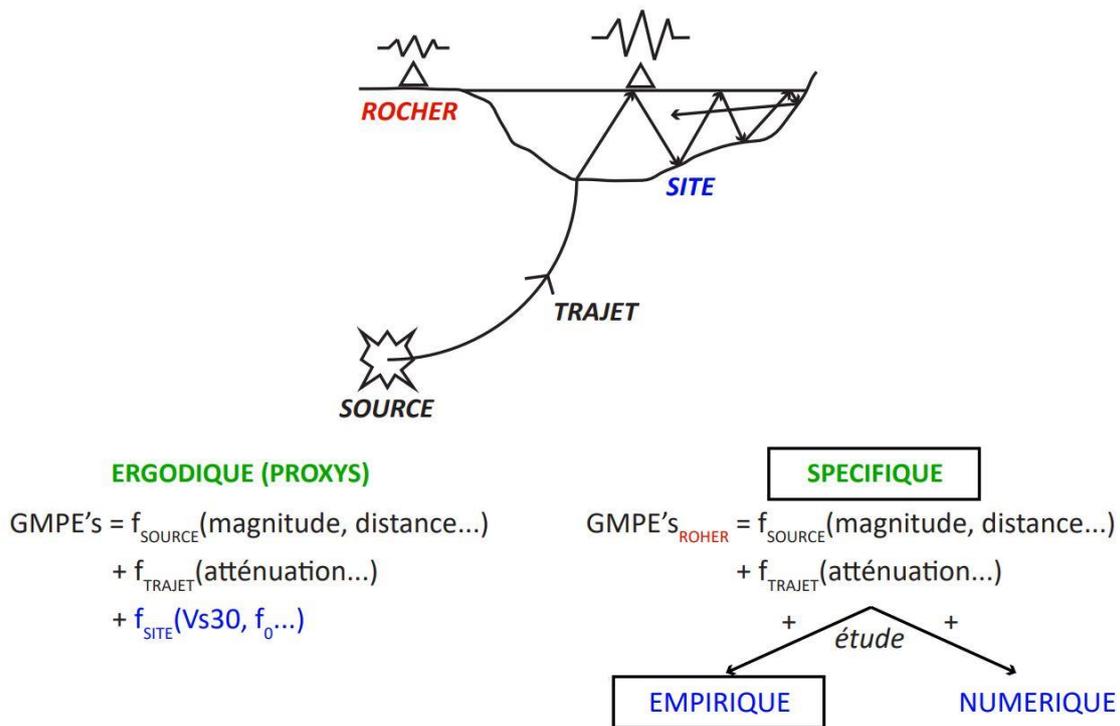


Figure 1.5 : Schéma résumant les différentes approches de prise en compte des effets de site dans les études d'aléa.

L'estimation empirique des effets de site nécessite néanmoins de nombreux enregistrements de séismes. En zone de sismicité faible à modérée, comme en France métropolitaine, le temps de retour des séismes modérés à forts est long, et donc limitant pour de telles études. Emerge alors l'intérêt du bruit sismique. Le bruit sismique correspond aux faibles vibrations de la surface de la Terre et a pour caractéristique d'être permanent et observable partout sur Terre. Ainsi, il constitue une source de données exploitable même en région de faible sismicité et, depuis les années 80, son potentiel pour l'estimation des effets de site est étudié. Dans cette thèse, le potentiel du bruit sismique pour l'estimation des effets de site est étudié à partir de plusieurs méthodes, présentées en section 1.4. Avant d'introduire ces méthodes, il est intéressant de se pencher un peu plus sur l'origine et la nature du bruit sismique.

1.3 Introduction au bruit sismique

1.3.1 Origine du bruit sismique

Le bruit sismique a été observé pour la première fois grâce au pendule de Bertelli (Bertelli 1872). Celui-ci a constaté que même en l'absence de séisme, le pendule était en perpétuel mouvement et que ces mouvements étaient corrélés aux conditions météorologiques et à la pression atmosphérique. Dès les premières études, on note une corrélation entre le bruit sismique et les phénomènes naturels. Banerji (1924, 1925) observe par exemple la concordance entre l'amplitude du bruit sismique et la mousson. Longuet-Higgins (1950) démontre que la houle océanique génère du bruit sismique par interférences entre des groupes de vagues de même fréquence se propageant en sens opposé. Il fait le lien avec les cyclones qui seraient propices à créer ces conditions d'interférences. Gutenberg (1958) établit lui aussi un lien entre les perturbations atmosphériques et le bruit sismique produit par l'océan. Il propose également que la météorologie locale (vent, tempête, orage...) ainsi que l'activité humaine génèrent du bruit sismique et propose une classification des origines du bruit sismique en fonction de la fréquence.

Par la suite, de nombreux auteurs ont proposé une telle classification (pour une revue étendue, le lecteur peut se référer à Bonnefoy-Claudet *et al.* 2006a). Il en ressort classiquement une limite autour de 1 Hz : en dessous de 1 Hz, le bruit sismique provient majoritairement de sources naturelles ; au-dessus de 1 Hz, le bruit sismique est de moindre intensité et donc rapidement dominé par des sources anthropiques. Un tel changement de sources se traduit généralement par un changement de direction dominante de propagation du champ de bruit (e.g. Satoh *et al.* 2001) et par le contraste entre une amplitude stable à l'échelle saisonnière en dessous de 1 Hz et des variations journalières et hebdomadaires au-dessus de 1 Hz caractéristiques de l'activité humaine (e.g. Bonnefoy-Claudet 2004). Néanmoins, cette frontière autour de 1 Hz n'est pas une limite absolue. Seo (1997) observe par exemple des variations journalières de l'amplitude du bruit sismique à des fréquences inférieures à 1 Hz dans le bassin de Mexico. Il suggère que la frontière entre le bruit sismique d'origine naturelle et anthropique dépende en partie de la géologie et qu'elle peut être décalée vers les basses fréquences pour des bassins sédimentaires profonds. Bonnefoy-Claudet (2004) fait des observations similaires pour le bassin grenoblois où elle caractérise cette limite autour de 0.75 Hz.

Si l'on résume, les sources de bruit sismique varient en fonction de la fréquence. Globalement, elles sont d'origine naturelle et plus ou moins distantes du site d'observation à basse fréquence ($f < 1$ Hz), d'origine anthropique et locales à plus haute fréquence ($f > 1$ Hz). Concernant les sources d'origine naturelle ($f < 1$ Hz), on distingue souvent deux catégories (e.g. Nakata *et al.* 2019) : la météorologie locale et régionale ($f \sim 0.5$ Hz) ; les océans et les conditions météorologiques à grande échelle ($f < 0.5$ Hz). Cette seconde catégorie est dominée par deux pics micro-sismiques autour de 14 s (1^{er} pic) et 7 s (2nd pic) de période. Le 1^{er} pic micro-sismique, bien que mal compris, serait dû à l'interaction de la houle océanique avec la côte (Haubrich *et al.* 1963). Le 2nd pic micro-sismique est généré par l'interférence entre deux houles de fréquence voisine se propageant en sens opposée (Longuet-Higgins 1950). Le schéma ci-dessous résume les différentes sources de bruit sismique en fonction de la fréquence/période.

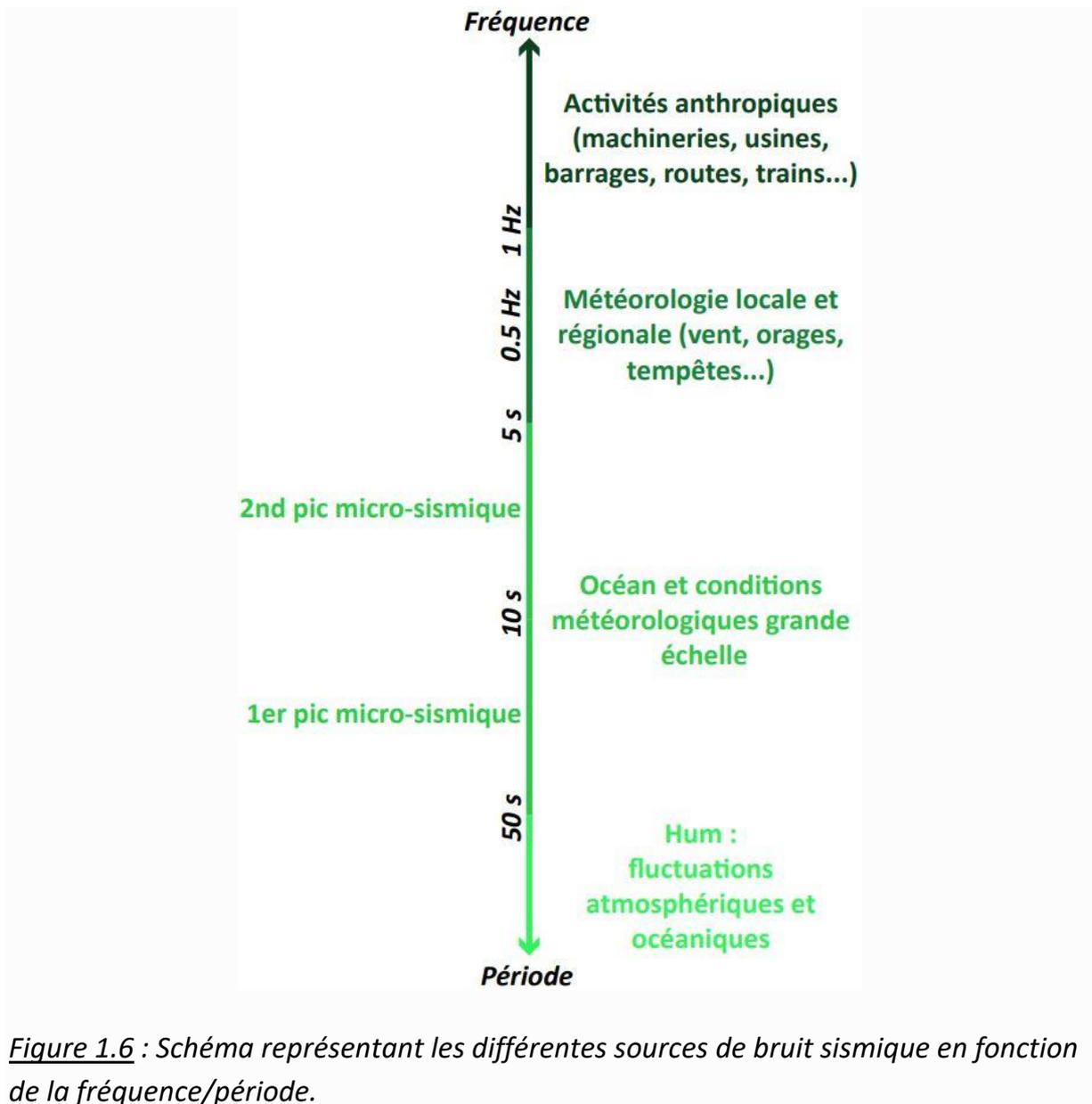


Figure 1.6 : Schéma représentant les différentes sources de bruit sismique en fonction de la fréquence/période.

1.3.2 Nature du bruit sismique

Plusieurs auteurs ont cherché à caractériser la nature du champ de bruit, c'est-à-dire à identifier les types d'ondes qui le compose et dans quelle proportion (e.g. Li *et al.* 1984 ; Horike 1985 ; Arai et Tokimatsu 1998 ; Yamamoto 2000 ; Cornou 2002 ; Cornou *et al.* 2003a, b ; Bonnefoy-Claudet *et al.* 2006a, b ; Di Giulio *et al.* 2006 ; Köhler *et al.* 2007 ; Endrun *et al.* 2010). Les études montrent en général la prépondérance des ondes de surface par rapport aux ondes de volume. Les proportions relatives semblent néanmoins dépendre de la géologie du site étudié ainsi que de la distance source-récepteur et du type d'excitation (e.g. Bonnefoy-Claudet *et al.* 2006b). Bonnefoy-Claudet (2004) montre par exemple, sur la base de synthétiques de bruit calculés à la surface de différents milieux à couches horizontales excités par des sources aléatoires, que les ondes de surface dominent uniquement pour des contrastes d'impédance élevés (> 3). Pour des contrastes d'impédance plus faibles en revanche (< 3), le champ de bruit semble être un mix entre ondes de surface et ondes de volume. Concernant les ondes de surface, il est difficile de discerner la proportion entre ondes de Rayleigh et ondes de Love, ainsi qu'entre les différents modes. Le mode fondamental de Rayleigh semble dominer en dessous de 1 Hz (e.g. Horike 1985 ; Yamanaka *et al.* 1994). Au-delà de 1 Hz, les conclusions divergent (car davantage dépendantes du site d'observation). Li *et al.* (1984) et Horike (1985) identifient principalement des harmoniques supérieurs de Rayleigh tandis que Arai et Tokimatsu (1998), Yamamoto (2000), Di Giulio *et al.* (2006) ou encore Endrun *et al.* (2010) trouvent une large proportion d'ondes de Love.

1.4 Méthodes site-spécifiques d'estimation des effets de site

Plusieurs méthodes site-spécifiques d'estimation des effets de site ont été introduites à partir des années 70. Ce manuscrit s'intéresse aux approches site-sur-référence. Dans la suite du manuscrit, trois méthodes site-sur-référence d'estimation des effets de site seront utilisées : le SSR (Standard Spectral Ratio) ; le SSR bruit, ou SSRn (noise-based Standard Spectral Ratio) ; le SSR hybride, ou SSRh (hybrid Standard Spectral Ratio). Ces méthodes sont présentées ci-dessous.

1.4.1 Le SSR (Standard Spectral Ratio)

Comme décrit en introduction de ce manuscrit, le mouvement sismique enregistré à la surface d'une structure superficielle peut se décrire sous la forme des contributions de la source (magnitude, position, directivité, mécanisme de rupture...), du trajet des ondes dans la croûte terrestre (atténuation géométrique, absorption, diffusion...) et de la réponse de la structure elle-même (effets de site). Dans le domaine fréquentiel (spectral), il s'agit d'une multiplication :

$$MOUVEMENT\ EN\ SURFACE = SOURCE \times TRAJET \times SITE . \quad (3)$$

L'idée du SSR, introduit par Borchardt (1970), est de considérer la séparation de ces contributions dans le domaine fréquentiel via un simple rapport spectral entre deux enregistrements d'un même séisme, l'un obtenu à une station au site, l'autre obtenu à une station en-dehors du site (référence au rocher). Si la distance inter-stations est largement inférieure à la distance épacentrale, alors les fonctions *SOURCE* et *TRAJET* sont identiques aux deux stations et la fonction *SITE* est isolée par le rapport. Mathématiquement, ce rapport se traduit par la normalisation du spectre d'amplitude (FAS, Fourier Amplitude Spectrum) de l'enregistrement à la station au site (*SX*) par le spectre d'amplitude de l'enregistrement à la station de référence (*Rref*) :

$$SSR_{SX/Rref} = \frac{FAS_{SX}}{FAS_{Rref}} . \quad (4)$$

Le SSR est considéré comme une méthode robuste d'estimation des effets de site (e.g. Rovelli *et al.* 1991 ; Aki 1993 ; Field et Jacob 1995 ; Field 1996 ; Huang et Teng 1999 ; Cultrera *et al.* 2014). Le choix de la position de la station de référence au rocher reste toutefois un point de discussion (e.g. Laurendeau 2013 ; Theodoulidis *et al.* 2018). La station de référence doit en effet être proche du site tout en subissant des effets locaux négligeables. En réalité, tout point en surface de la Terre, même à considérer

que celui-ci soit situé à l’affleurement de ce que l’on appelle physiquement un « bon rocher », subit un effet local. Il peut donc s’avérer non trivial de sélectionner un rocher dont les effets locaux sont négligeables. Ainsi, caractériser le mouvement sismique à la station de référence est tout aussi important que de caractériser le mouvement sismique au site, puisque la mesure est relative. Comme élément de réponse à cette problématique, certains auteurs utilisent une mesure moyenne de plusieurs stations au rocher pour réduire les incertitudes (e.g. Bonilla *et al.* 1997). D’autres, proposent d’utiliser une référence à la base de la couche de sol avec une mesure en forage (e.g. Steidl *et al.* 1996), mais cela amène à ne pas considérer les effets de la surface libre à la référence. Pour plus d’informations concernant cette problématique, le lecteur pourra se référer à Laurendeau (2013).

En zone de sismicité faible à modérée comme en France métropolitaine, la principale limitation du SSR reste le temps d’acquisition nécessaire à l’enregistrement d’un nombre de séismes suffisant (e.g. Perron *et al.* 2015 ; Perron *et al.* 2020). Ceci est d’autant plus vrai compte tenu de la variabilité des effets de site dépendamment de la source (voir section 1.1.2). Dès lors, il peut être intéressant de chercher à utiliser son équivalent appliqué au bruit sismique : le SSRn.

1.4.2 Le SSRn (noise-based Standard Spectral Ratio)

Le SSRn a été introduit par Irikura et Kawanaka (1980). Comme mentionné juste au-dessus, la méthode est équivalente au SSR (eq. 4), mais plutôt que de s’appliquer sur des enregistrements de séismes, elle s’applique sur des fenêtres synchrones de bruit sismique. Les premières études utilisant le SSRn dans le but d’évaluer les effets de site sont celles de Kagami *et al.* (1982, 1986). Ils ont appliqué le SSRn à basse fréquence (entre 1 et 0.2 Hz) dans la plaine de Niigata (Japon) ainsi que dans le bassin de Los Angeles. Dans les deux cas, l’amplification estimée par le SSRn s’est avérée qualitativement cohérente avec la structure sédimentaire. Par la suite, plusieurs auteurs ont comparé le facteur d’amplification obtenu par le SSRn, soit avec la géologie du site étudié (e.g. Ferritto 1995 ; Ibs-von Seht et Wohlenberg 1999 ; Roten *et al.* 2006), soit avec le facteur d’amplification obtenu par le SSR (e.g. Rovelli *et al.* 1991 ; Lermo et Chávez-García 1994). La plupart de ces études a montré que pour des fréquences inférieures à 1 Hz, le SSRn est cohérent avec la géologie du site et peut fournir des mesures comparables au SSR. Ceci est possible car à ces fréquences, le bruit sismique est principalement dominé par des sources naturelles lointaines. En conséquence, la condition nécessaire à l’application de la méthode site-sur-référence (la distance inter-stations doit être largement inférieure à la distance site-source) est satisfaite. Au-delà de 1 Hz en revanche, la proximité des sources liées à l’activité humaine est susceptible

de briser cette hypothèse de travail (e.g. Field *et al.* 1990 ; Zhao *et al.* 1998 ; Horike *et al.* 2001). Le facteur d'amplification du SSRn est alors biaisé : il surestime souvent le facteur d'amplification du SSR et la variation spatiale de l'amplification observée n'est, au premier ordre, plus le reflet de la structure géologique. Il reste cependant important d'avoir une mesure robuste de l'amplification sur une gamme de fréquences large, typiquement jusqu'à 10 Hz, pour les besoins de l'ingénierie (évaluation de la réponse des bâtiments et infrastructures).

Récemment, une nouvelle approche site-sur-référence combinant le SSRn et le SSR et pouvant potentiellement atténuer ces limitations a été introduite. Il s'agit du SSRh.

1.4.3 Le SSRh (hybrid Standard Spectral Ratio)

Le SSRh a été introduit par Perron *et al.* (2018). L'idée est d'abord d'estimer l'amplification site/rocher à une station de référence (S_{ref}) au site, puis de reporter les variations spatiales site-à-site de l'amplification à n'importe quel capteur (SX). L'amplification site/rocher est estimée via le SSR classique. Les variations spatiales site-à-site de l'amplification sont reportées via un SSRn site/site (Figure 1.7).

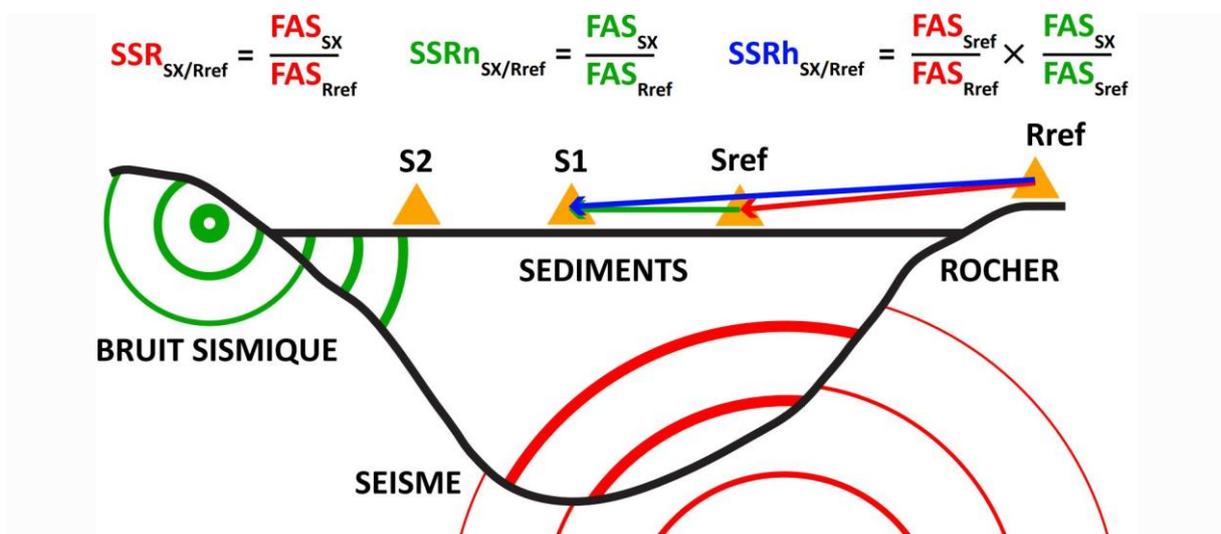


Figure 1.7 : Schéma présentant le principe du SSR hybride, ou SSRh (hybrid Standard Spectral Ratio).

Perron *et al.* (2018) ont appliqué le SSRh et le SSRn dans les bassins sédimentaires de Provence (sud de la France) et d'Argostoli (Grèce). Dans les deux cas, le SSRn a fourni un facteur d'amplification comparable au SSR sur une gamme de fréquences limitée par la fréquence fondamentale de résonance f_0 du bassin. La limite

d'applicabilité du SSRn n'a pas ici été associée à la proximité des sources anthropiques qui, comme mentionné précédemment, dominant généralement le champ de bruit à partir de 1 Hz. Perron *et al.* (2018) ont plutôt proposé que le bassin agisse comme une barrière naturelle qui piège le bruit sismique au-delà de f_0 . En conséquence, au-delà de f_0 , seule une partie limitée de l'énergie est transmise à la station de référence au rocher et le SSRn surestime le facteur d'amplification du SSR. Le SSRh étant basé sur l'introduction d'une station de référence sous les mêmes conditions géologiques que la station à laquelle l'amplification est estimée, il n'est pas limité par cette problématique. Ainsi, cette méthode a pu fournir un facteur d'amplification comparable au SSR sur une gamme de fréquences large (jusqu' à 12 Hz pour le bassin de Provence et 20 Hz pour le bassin d'Argostoli).

Plus tard, Perron *et al.* (2022) ont appliqué le SSRh dans la région de Sion (vallée sédimentaire du Rhône, Suisse). Ils ont globalement reporté une bonne corrélation avec le SSR sur une gamme de fréquences large, avec toutefois des biais pour les stations situées aux bords de la vallée. Dans le chapitre 3 (Gisselbrecht *et al.* 2023), le SSRh et le SSRn seront appliqués dans la vallée du Tricastin (vallée sédimentaire du Rhône, sud-est de la France). La vallée du Tricastin est caractérisée par une forte industrialisation, ce qui pourra notamment permettre d'évaluer l'applicabilité du SSRh dans une zone très anthropisée.

Chapitre 2 : Acquisition des données à Tricastin, traitements préliminaires et observations des sources de bruit sismique

2.1 Contexte de l'étude

Dans cette partie, plusieurs niveaux d'informations liées au contexte de nos travaux ainsi que de la zone d'étude seront introduits. Ces informations sont regroupées sur la figure 2.1. Précisons donc que la figure 2.1 constitue la figure principale de cette partie et que nous allons nous y référer régulièrement pour la découvrir pas-à-pas.

2.1.1 Formation de la vallée du Tricastin

La vallée du Tricastin est située dans la basse vallée du Rhône, au sud-est de la France (Figure 2.1). Sa formation remonte à la crise de salinité messinienne (e.g. Hippolyte *et al.* 2021). Cet événement débute par la fermeture du Corridor de Gibraltar il y a environ 6 millions d'années. Isolée de l'Atlantique, la Méditerranée présente un bilan alimentation en eau-évaporation négatif, provoquant son assèchement. La chute du niveau marin entraîne les fleuves méditerranéens (dont le Rhône) à creuser de profondes vallées pour rétablir leur profil d'équilibre. Au début du Pliocène, la réouverture du détroit de Gibraltar entraîne la remise en eau soudaine de la Méditerranée et l'ennoiement des vallées messiniennes, creusées bien en-dessous du niveau marin. Cet ennoiement entraîne le remplissage des vallées par des sédiments (sables et argiles) datant du Pliocène. C'est dans ce contexte que s'est formée la vallée du Tricastin, dans les grès et calcaires du Crétacé (en vert sur la figure 2.1). Elle est aujourd'hui remplie par des couches sédimentaires du Pliocène (en rose) recouvertes par des terrasses quaternaires (en gris).

2.1.2 Motivation d'une étude site-spécifique à Tricastin

En 2016, en amont du projet DARE, une campagne géophysique a permis d'instrumenter deux sites sédimentaires, BOLL et PAUL (respectivement localisés par un triangle vert et bleu sur la figure 2.1), ainsi que le site au rocher affleurant ADHE (triangle rouge en-dehors de la vallée sédimentaire). Notons dès à présent que ces trois stations seront utilisées comme stations de référence dans l'ensemble du manuscrit. Gélis *et al.* (2022) ont regroupé les informations issues de cette campagne, des logs géologiques des forages présents dans la région (deux forages profonds atteignent notamment le fond de la vallée fournissant des informations sur l'épaisseur des sédiments) et d'études spécifiques réalisées sur les trains à grande vitesse (TGV ; Mocochain *et al.* 2006). Leurs travaux ont notamment permis une première estimation des bords de la vallée (Figure 2.1, trait bleu épais) ainsi que de l'axe profond (Figure 2.1, trait bleu fin) et d'estimer les profils verticaux de vitesses des ondes S à BOLL et ADHE. Au-dessous de BOLL, la vitesse moyenne obtenue dans la couche sédimentaire est d'environ 850 m/s, contre 3100 m/s dans le substratum rocheux. Ceci traduit un contraste de vitesse supérieur à 3 à l'interface sédiments/rocher. A cet endroit, la base du remplissage repose directement sur le substratum rocheux. Une coupe géologique E-O de Gélis *et al.* (2022) montre également qu'au niveau de BOLL, la profondeur du remplissage est supérieure à 500 m, pour une largeur inférieure à 4 km, révélant une forte incision. La géométrie incisée de la vallée ainsi que le fort contraste lithologique avec le substratum rocheux sont propices à générer des effets de site importants et complexes. Les premières mesures SSR de Gélis *et al.* (2022) semblent d'ailleurs aller dans ce sens, révélant une amplification locale sédiments/rocher supérieure à 6 sur une large gamme de fréquences (marqueur d'effets 2D/3D, chapitre 1 section 1.1.2).

Ces premiers travaux suggérant une structure géologique complexe, ils appuient le besoin d'études spécifiques pour mieux comprendre et caractériser les effets liés à cette configuration.

2.1.3 Données de sismique active

En 2020, l'IRSN a eu accès aux données de 11 profils sismiques (6 lignes E-O et 5 lignes N-S) acquises par la compagnie DMT en 2008 pour le compte de Gaz de France (actuellement ENGIE) et sous-traitées par la société Electrabel France. Ces données de sismique active quadrillent la vallée du Tricastin et ont fait l'objet d'une analyse dans l'objectif de construire un modèle géologique de la zone (Cushing et collaborateurs) en parallèle (et complément) du projet DARE. Les mesures ont donc été interpolées pour

contraindre la géométrie des horizons géologiques et ainsi la profondeur de la vallée sur une zone de 10 km par 10 km (Figure 2.1, encadré bleu). Cette profondeur correspond à l'interface entre les sédiments du Pliocène et les roches du Crétacé, appelée surface d'érosion messinienne (MES). Elle est représentée dans le coin supérieur droit de la figure 2.1(a). La profondeur de la surface d'érosion messinienne indique que le remplissage sédimentaire de la vallée du Tricastin s'approfondit vers le sud en suivant le cours du Rhône. Comme mentionné précédemment, la vallée est fortement incisée à cet endroit. Dans la partie S-O, on distingue le canyon profond de l'Ardèche ainsi que la zone de confluence entre le Rhône et l'Ardèche. Cette zone de confluence creuse un îlot, dont la roche est quasi-affleurante, appelé île de Lapalud. Notons que cette image interpolée de la profondeur de la surface d'érosion messinienne sera utilisée comme fond de carte principale à travers la suite du manuscrit.

2.1.4 Des enjeux importants dans la zone d'étude

La partie de la zone d'étude délimitée par l'encadré bleu est fortement industrialisée. Les installations principales sont représentées dans le coin supérieur droit de la figure 2.1(a). Cette partie comprend notamment le site nucléaire de Tricastin (TNS, trait épais noir). Le TNS est le site nucléaire le plus vaste de France. Il s'étend sur plus de 600 hectares et regroupe la centrale EDF (4 réacteurs) ainsi que d'importantes installations liées à la fabrication, au stockage et au recyclage des déchets. La zone comprend également le barrage hydroélectrique de Donzère-Mondragon (cercle vert au sud-est de la zone), la nationale N7 (en vert), l'autoroute A7 (en violet) ainsi que deux lignes de fret et une ligne à grande vitesse (LGV, en marron) traversant la zone sur un axe N-S. Elle comprend aussi 5 villes pour une population totale de 45 000 habitants.

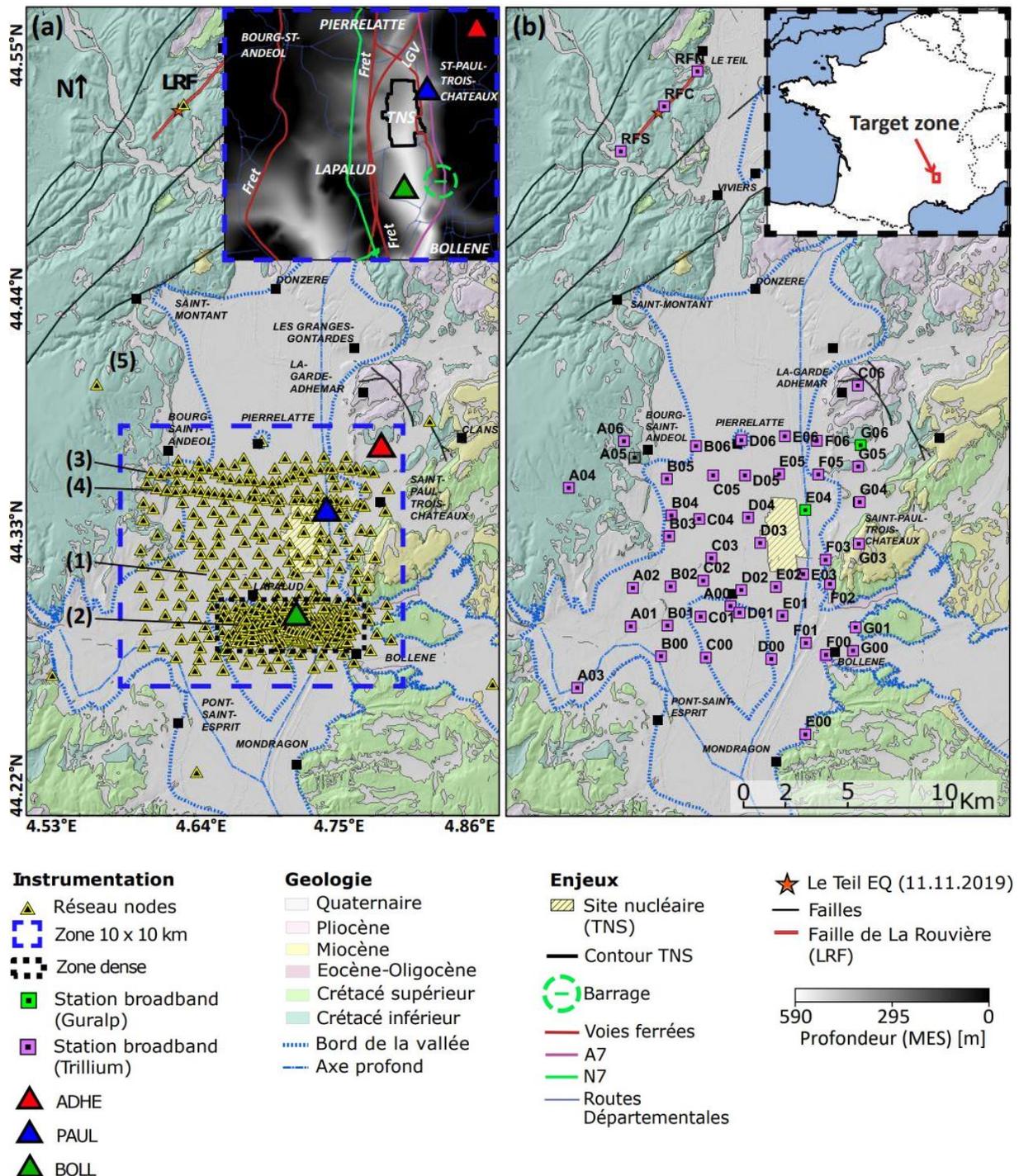


Figure 2.1 : Déploiement du réseau de nodes (a) et stations large-bande (b) sur fond de carte géologique. La station A05 a été grisée car les données sont inutilisables. ADHE est un site de référence au rocher. BOLL et PAUL sont des sites de référence aux sédiments. Les nodes à ADHE, BOLL et PAUL sont colocalisés respectivement avec les stations large-bande G06, E01 et E04. La faille de La Rouvière (LRF) ainsi que le site nucléaire (TNS) sont représentés sur les deux cartes. Les pointillés noirs (a) délimitent la zone dense du réseau de nodes (distance moyenne inter-stations de 250 mètres).

L'encadré bleu délimite une zone de 10 km par 10 km pour laquelle la profondeur du remplissage sédimentaire est représentée (coin supérieur, Bagayoko 2021 ; Do Couto et al., soumis). Cette profondeur correspond à l'interface entre les sédiments du Pliocène et les roches du Crétacé, appelée surface d'érosion messinienne (MES). Les principales infrastructures y sont aussi représentées. Le réseau de nodes (a) a été déployé avec l'aide d'EGIS et SISPROBE. Reproduite en intégrant les figures de Froment et al. (2022a) et Gisselbrecht et al. (2023).

2.1.5 Le séisme du Teil

Le 11 novembre 2019, un séisme de magnitude 4.9 s'est produit sur la faille de La Rouvière (LRF, en rouge sur la carte géologique de la figure 2.1), localisée à environ 20 km du TNS. Ce séisme est le plus important en France métropolitaine depuis 1967. Des dommages importants ont été observés dans les villages situés à proximité de la zone de rupture. Le village du Teil a été particulièrement impacté (Ritz *et al.* 2020 ; Causse *et al.* 2021 ; Cornou *et al.* 2021). Schlupp *et al.* (2021) ont estimé, sur la base d'observations de terrain et de témoignages, une intensité maximale (EMS98) d'environ VII-VIII au Teil et d'au moins VI dans 30 municipalités voisines (Figure 2.2).

A la suite de l'évènement, l'ASN (l'Autorité de Sûreté Nucléaire) a déclaré qu'aucun dégât n'avait été observé sur les deux centrales nucléaires à proximité (Tricastin à environ 20 km au sud-est et Cruas à environ 15 km au nord-est). En revanche, le mouvement sismique a atteint le seuil d'alerte à Cruas, provoquant l'arrêt de la centrale jusqu'en décembre 2019 pour des vérifications approfondies. La survenue de ce séisme a également entraîné des demandes de l'ASN auprès des exploitants des deux sites nucléaires de réaliser des études géologiques et géophysiques complémentaires, notamment concernant la caractérisation des failles dans la zone, et ce, afin de consolider la définition de l'aléa sismique sur ces deux sites.

Le séisme du Teil met en avant la réalité du risque sismique dans la zone de Tricastin. Outre cet évènement marquant, la région du Tricastin est impactée par des essais sismiques fréquents (documentés depuis la fin du XVIII^{ème} siècle) qui durent parfois plusieurs mois et se produisent à faible profondeur (quelques centaines de mètres ; e.g. Thouvenot *et al.* 2009 ; Larroque *et al.* 2021). Le séisme du Teil ainsi que les nombreux essais sismiques révèlent une activité sismique régulière dans la zone présentant une certaine diversité de mécanismes et comportements.

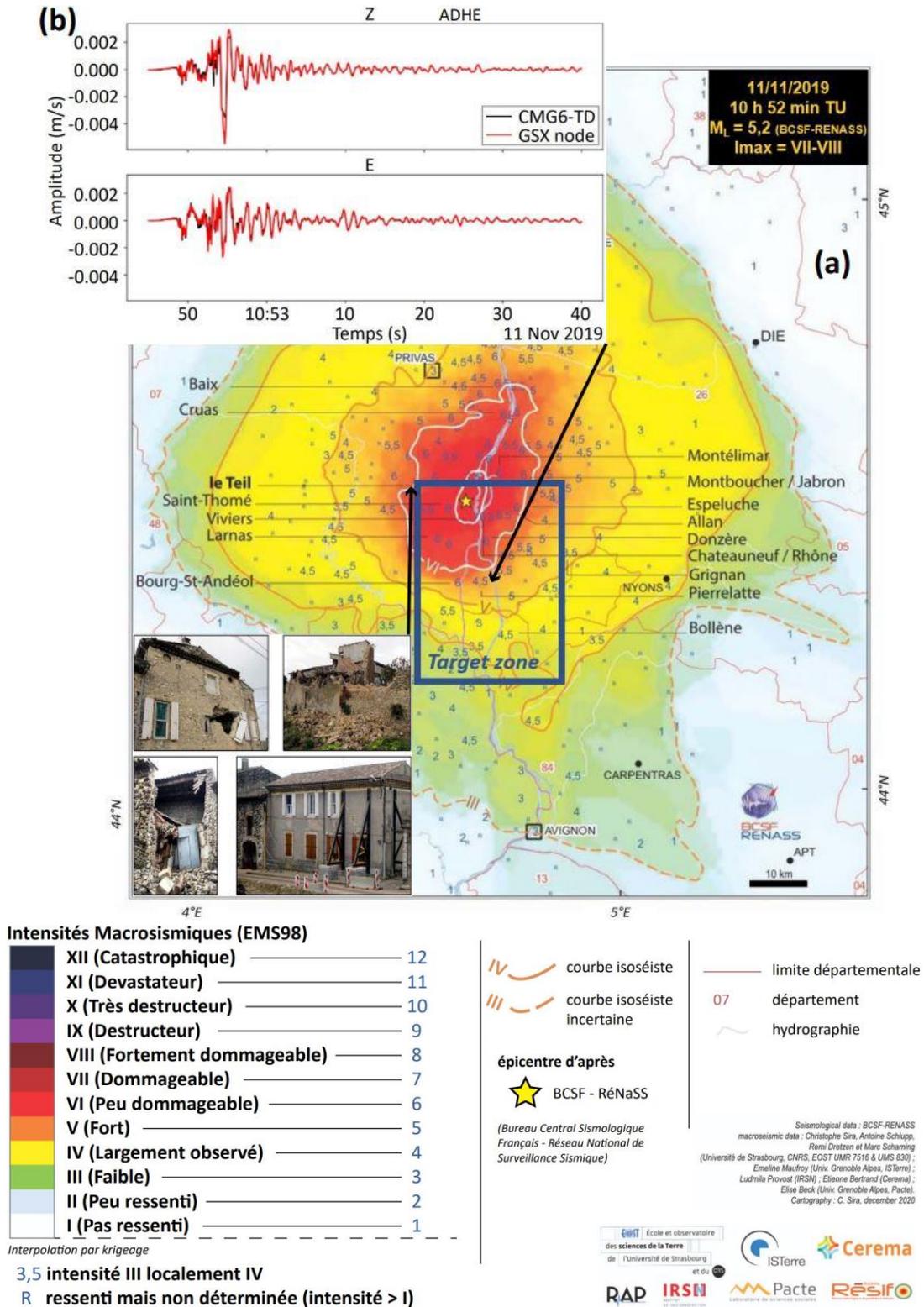


Figure 2.2 : (a) Carte d'intensités macrosismiques pour le séisme du Teil (11/11/2019) avec quelques illustrations des dégâts dans la municipalité du Teil. Modifiée de Schlupp et al. (2021). (b) Sismogrammes (entre 0.2 et 20 Hz) du séisme du Teil enregistré à deux capteurs (node en rouge et large-bande en noir) au site ADHE, située à environ 20 km de l'épicentre, pour la composante E et Z. Modifiée de Froment et al. (2022a).

2.2 Campagnes d'acquisition du projet DARE

2.2.1 Présentation des deux réseaux déployés

Comme mentionné dans l'introduction de ce manuscrit, deux réseaux de capteurs ont été installés dans le cadre du projet DARE dans le but de mener des études site-spécifiques à Tricastin. Le premier réseau implique 409 capteurs nodes (géophones autonomes 3 composantes) déployés pendant 1 mois, entre le 17 février 2020 et le 18 mars 2020 (triangles jaunes et noirs, Figure 2.1a). Ce réseau a enregistré le bruit sismique en continu. Il est constitué de 5 sous-réseaux :

- (1) une zone de 10 km par 10 km (encadré bleu, Figure 2.1a). Cette partie comprend 164 nodes espacés de 400 à 1300 m ;
- (2) une zone dense (pointillés noirs, Figure 2.1a). Cette partie comprend 179 nodes espacés d'environ 250 m et couvrant la zone attendue la plus profonde et incisée de la vallée ;
- (3) et (4) deux lignes E-O de 30 nodes espacés de 400 m qui longent deux routes départementales au nord ;
- (5) 6 nodes distants couvrant un azimuth de 360° autour de la zone dense. L'un de ces nodes a été placé sur LRF.

La géométrie formée par ces 5 sous-réseaux permet à la fois de couvrir l'ensemble de la zone et d'augmenter la densité spatiale dans certaines zones spécifiques (e.g. où la vallée est plus étroite).

Le second réseau implique 49 stations large-bande (47 Trillium compacts et 2 Guralp) déployés pendant plus de 8 mois, de septembre 2020 à mai 2021 (carrés violets et verts, Figure 2.1b). Ce réseau a eu pour objectif d'enregistrer la sismicité en continu. Il couvre l'ensemble de la zone définie par les nodes (encadré bleu, Figure 2.1a). La plupart des stations (hormis A0 et G4) ont été placées en dehors des bâtiments pour éviter d'inclure la réponse de la structure dans les enregistrements. Les stations A04, A06, G06 et D06 ont été placées sur des sites « rocher » selon des critères géologiques (à l'aplomb de roches dures datant du Crétacé inférieur). Enfin, notons que les stations RFS, RFC et RFN ont été placées le long de la faille de la Rouvière.

Comme précisé dans l'introduction du manuscrit, le projet DARE a pour objectif d'exploiter la complémentarité des deux réseaux via deux approches : numérique et empirique. Rappelons que cette thèse est basée sur l'exploitation du réseau de nodes

(Figure 2.1a) via l'approche empirique pour estimer l'amplification associée aux effets de site dans la vallée du Tricastin. Toutefois, la sismicité extraite du réseau de stations large-bande (Figure 2.1b) ainsi que des 3 stations Guralp installées lors de la campagne géophysique de 2016 (1 an d'acquisition ; Gélis *et al.* 2022) servira de point de comparaison pour évaluer les résultats issus du bruit sismique. A ce sujet, notons que les nodes à ADHE, BOLL et PAUL ont été colocalisés respectivement avec les stations large-bande G06, E01 et E04 ainsi que les 3 stations Guralp. Pour plus de précisions sur le déploiement des deux réseaux, le lecteur peut se référer à Froment *et al.* (2022a).

2.2.2 Modèle 3D de la vallée du Tricastin

Même si dans cette thèse nous nous concentrons sur l'approche empirique, il est intéressant de présenter le modèle de vitesse à la base de l'approche numérique. Cela permettra à ce stade d'avoir une vision 3D de la vallée du Tricastin en complément de la MES tirée de l'analyse des profils de sismique active. Ce modèle a été construit par tomographie de bruit sismique en utilisant l'ensemble des 400 nodes (Lavoué *et al.*, soumis). La tomographie de bruit sismique utilise les propriétés de dispersion des ondes de surface reconstruites par corrélation de bruit entre chaque paire de capteurs (Ambient Noise Surface Wave Tomography, ANSWT ; e.g. Shapiro et Campillo 2004). La dispersion fréquentielle de la vitesse de groupe des ondes de surface est mesurée par analyse temps-fréquence pour chaque paire (courbe de dispersion). Ces mesures sont régionalisées, via une inversion linéarisée supposant la trajectoire des rais, afin d'obtenir des cartes de vitesse à différentes fréquences. Enfin, une seconde inversion de la dispersion en chaque point de la grille est réalisée pour obtenir des profils verticaux (en profondeur) de vitesse des ondes S. L'ensemble des profils verticaux sur une surface 2D constitue le modèle 3D.

Le modèle de vitesse est représenté sur la figure 2.3. Notons que la surface représentée correspond à l'iso-vitesse 1200 m/s des ondes S. L'iso-vitesse 1200 m/s (vitesse à la base de la couche sédimentaire) correspond au premier ordre à la surface d'érosion messinienne (MES, en nuance de gris sur la figure 2.1a) et est par conséquent interprétée dans cette imagerie comme l'interface sédiments/rocher. Les différences entre l'iso-vitesse 1200 m/s et la MES sont toutefois discutées dans Froment *et al.* (2022b). Le modèle obtenu confirme le fort contraste de vitesse entre les sédiments et le substratum et met en lumière la géométrie complexe de la vallée. L'axe N-S de la paléovallée du Rhône est clairement visible. Comme décrit dans le paragraphe 2.1.2, la vallée s'approfondit vers le sud et devient plus étroite. La vue en 3D met en évidence des bords très incisés du côté est, presque verticaux, qui contrastent avec une pente

plus douce côté ouest. Dans la partie S-O, on identifie le canyon de l'Ardèche et la zone de confluence qui forme l'île de Lapalud.

Au premier ordre, les principales structures géologiques sont donc bien définies par le modèle. Notons toutefois que les travaux de Lavoué *et al.* (soumis) ont pointés certaines limitations intrinsèques liées au manque de sensibilité des ondes à la proche surface. En conséquence, le modèle manque de bords clairs remontant jusqu'à la surface capables de piéger efficacement les ondes dans la vallée. Les résultats présentés dans Lavoué *et al.* (soumis) montrent que les simulations numériques de propagation des ondes dans ce modèle ne permettent pas de reproduire fidèlement les effets 2D/3D observés par Gélis *et al.* (2022).

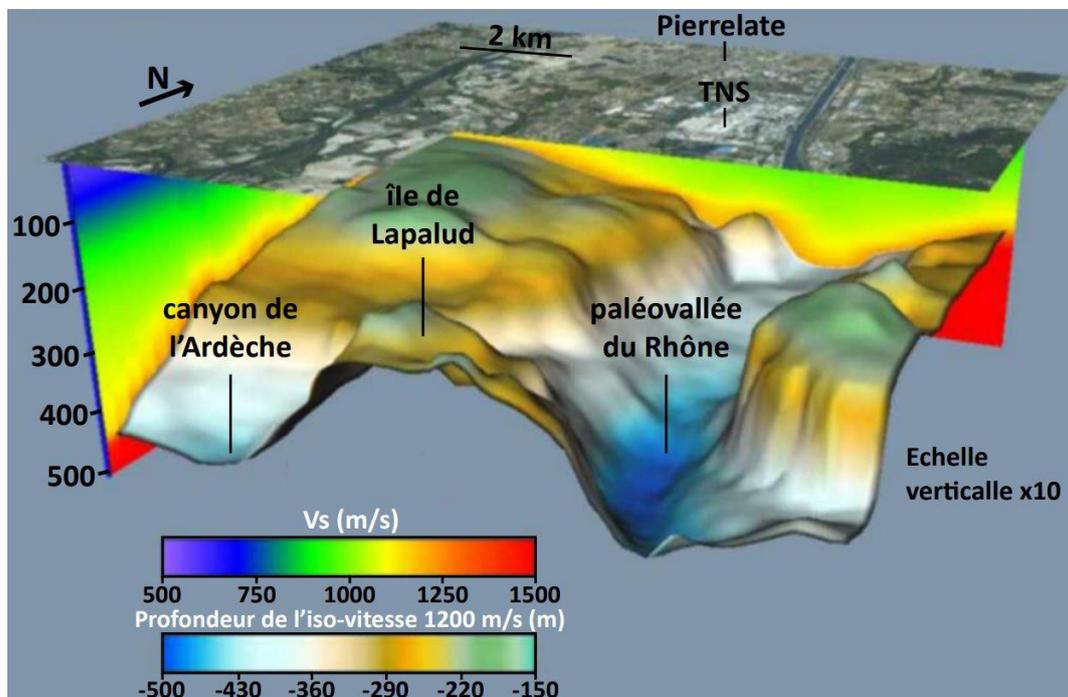


Figure 2.3 : Image 3D de la vallée du Tricastin obtenue par tomographie de bruit sismique (Ambient Noise Surface Wave Tomography, ANSWT) avec le réseau de nodes. Modifiée de Froment *et al.* (2022b). La surface représentée correspond à l'iso-vitesse 1200 m/s des ondes S. L'iso-vitesse 1200 m/s correspond au premier ordre à la surface d'érosion messinienne (MES, en nuance de gris sur la figure 2.1a) et est par conséquent interprétée dans cette imagerie comme l'interface sédiments/rocher. Les différences entre l'iso-vitesse 1200 m/s et la MES sont toutefois discutées dans Froment *et al.* (2022b).

2.3 Traitement des données de bruit sismique

2.3.1 Jeu de données des nodes

L'installation du réseau de nodes a eu lieu entre le 17 et le 20 février 2020. La désinstallation a eu lieu entre le 16 et le 18 mars 2020. Sur les 409 nodes déployés, 400 ont été retrouvés intacts. Ces 400 nodes constituent le jeu de données principal de cette thèse. Sur ces 400 nodes, 22 ont été déterrés avant la fin de l'acquisition. Les fichiers ont été retirés à partir de la date correspondant au déterrement. 24 autres nodes présentent un enregistrement écourté. Au total, 46 nodes (11.5%) présentent donc un enregistrement incomplet. Sur ces 46 nodes, 3 présentent un enregistrement particulièrement court (bar plot jaune, Figure 2.4), entre 13 et 32 heures d'acquisition. Enfin 2 nodes présentent des arrêts intermédiaires (gaps) de plusieurs heures. La figure 2.4 représente la période d'acquisition à chaque node.

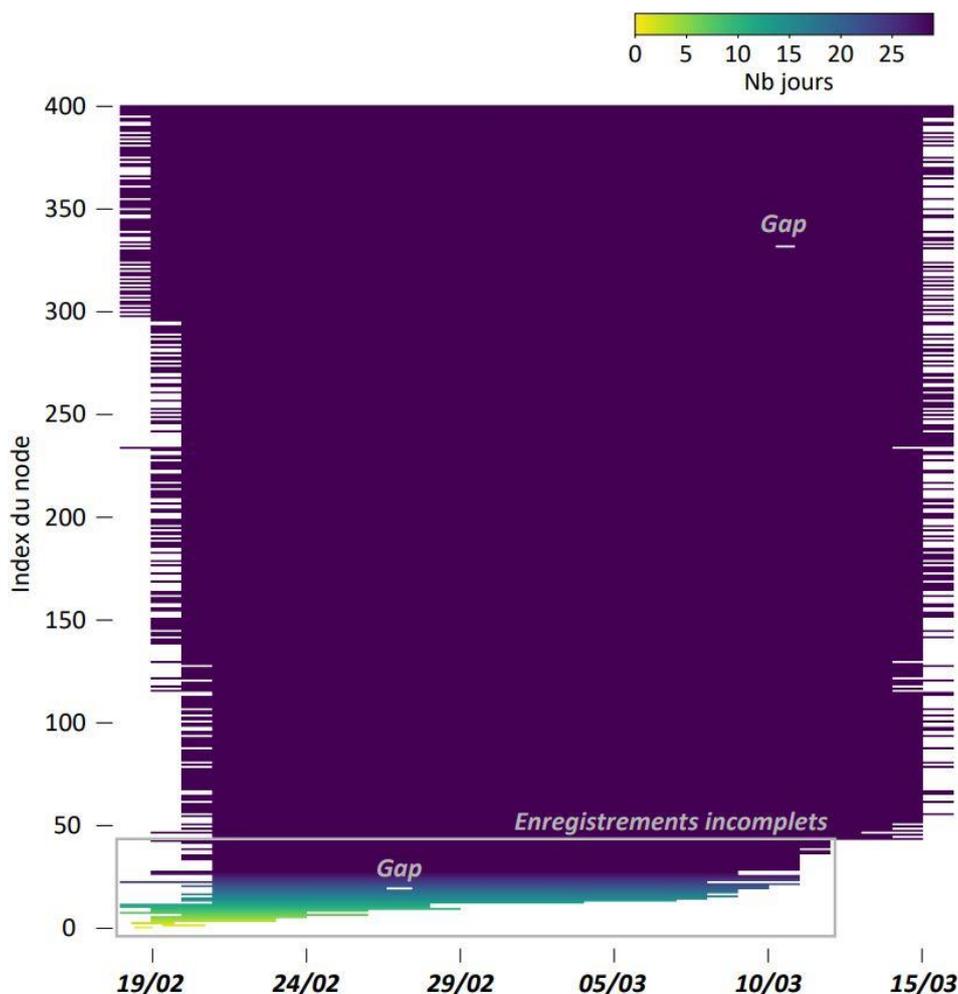


Figure 2.4 : Période d'acquisition à chaque node. Les nodes sont classés en fonction de la durée de l'acquisition en ordre croissant.

Notons que pour chaque node, l'enregistrement est découpé en traces sismiques journalières (fichiers de 24H). Chaque trace a été rééchantillonnée à 50 Hz (initialement 250 Hz). Nous avons ensuite corrigé la pente de chaque trace (detrend) et déconvolué la réponse instrumentale. Le volume total du jeu de données est d'environ 1 TB.

2.3.2 Fiabilité des nodes à basse fréquence

D'après Gélis *et al.* (2022), la fréquence fondamentale de résonance (f_0) de la vallée du Tricastin se situe autour de 0.5 Hz. Les fréquences d'intérêt pour cette étude englobent donc des fréquences bien inférieures à la fréquence de coupure des nodes (5 Hz). Il faut donc s'assurer de leur bon fonctionnement dans ces gammes de fréquences. Pour cela, nous avons calculé les PPSD (densité spectrale de puissance, e.g. Nakata *et al.* 2019) aux nodes situés à ADHE et BOLL pour les 3 composantes du mouvement (E, N, Z). Les PPSD permettent d'évaluer la distribution des spectres de puissance. Cette distribution est généralement comparée à deux modèles de référence représentatifs d'un bruit calme (new low noise model, NLNM) ou élevé (new high noise model NHNM). Ces modèles sont issus de Peterson (1993). La distribution des spectres et la comparaison aux modèles peut donner des informations sur la performance d'un capteur dans une bande de fréquences donnée. Les PPSD ont été calculés sur des fenêtres d'1 heure sans recouvrement sur la totalité de l'acquisition (plus de 600 segments). Les résultats sont présentés sur la figure 2.5. Au-delà de 0.1 Hz, la distribution est contenue entre les modèles de niveau de bruit (lignes grises) et semble stable. On note toutefois un niveau de bruit élevé à BOLL, qui peut s'expliquer par sa localisation (à l'aplomb de la partie sédimentaire profonde et à proximité de sites industriels, Figure 2.1). En dessous de 0.1 Hz, la distribution dépasse le NHNM et la densité spectrale sur la composante Z reflète clairement du bruit numérique, ce qui signifie que la sensibilité du node est atteinte pour cette composante. Ces résultats sont cohérents avec certains tests réalisés en laboratoire pour des géophones équivalents (e.g. Ringler *et al.* 2018 ; nodes Fairfield 5 Hz) et semblent indiquer que les nodes sont effectivement fiables dans une gamme de fréquences bien inférieure à la fréquence de coupure et qui contient f_0 (jusqu'à environ 0.1 Hz).

Froment *et al.* (2022a) ont comparé les PPSD de ces nodes aux PPSD obtenus pour des stations large-bande colocalisées. La comparaison a montré des résultats similaires sur la plage 0.1-20 Hz. Dans cette étude, l'enregistrement du séisme du Teil a également été comparé entre ces mêmes capteurs nodes et large-bande entre 0.2 et 20 Hz. Les sismogrammes ont révélé une bonne corrélation, tant en phase qu'en

amplitude. Toutes ces observations confirment la fiabilité des nodes entre 0.2 et 20 Hz. La donnée finale a donc été filtrée entre 0.2 et 20 Hz. Cette plage de fréquences permet de couvrir la bande d'intérêt pour nos calculs.

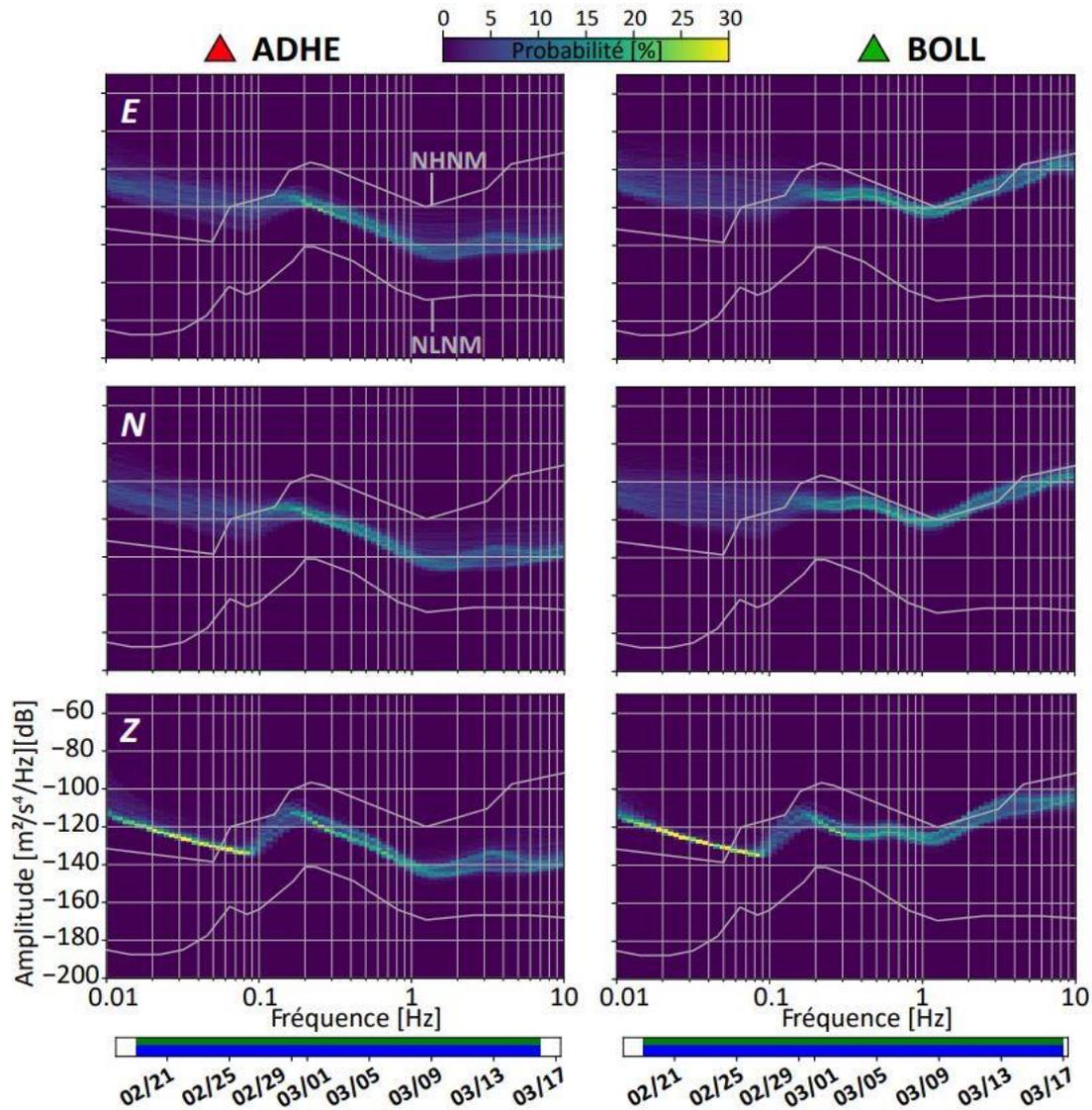


Figure 2.5 : PPSD calculés à ADHE (colonne de gauche) et BOLL (colonne de droite) pour les 3 composantes du mouvement (E, N, Z). Les PPSD ont été calculés sur des fenêtres d'1 heure sans recouvrement sur la totalité de l'acquisition (plus de 600 segments). Les lignes grises représentent le NLNM (new low noise model) et le NHNM (new high noise model) issus de Peterson (1993).

2.3.3 Contrôle qualité des données

Un jeu de données aussi dense (400 capteurs, 3 composantes, 1 mois d'acquisition) nécessite une évaluation approfondie afin de ne pas appliquer des calculs à l'aveugle sur des données dont on ne connaît pas la qualité. Après avoir procédé aux traitements préliminaires (section 2.3.1) et s'être assuré de la fiabilité des nodes dans la bande de fréquences d'intérêt (section 2.3.2), un contrôle qualité a été conduit. Ce travail doit permettre : (1) d'identifier les nodes défectueux ; (2) d'évaluer la qualité globale de la donnée à chaque node ; (3) de caractériser au premier ordre les sources de bruit sismique dans la zone. Le contrôle qualité s'est déroulé en deux étapes. Dans un premier temps, un catalogue permettant d'explorer rapidement les données (partie A) a été construit. Ensuite, nous avons cherché à caractériser l'amplitude du bruit sismique dans la zone (partie B).

A. Le catalogue

Le catalogue construit comprend 400 pages, chacune d'entre elles correspondant à 1 node. Ce catalogue est proposé pour les 3 composantes du mouvement (1200 pages en tout). Sur chaque page figurent 4 représentations : la distribution des FAS calculés à partir de segments de 10 min, la trace temporelle, le spectrogramme et une carte indiquant la position du node dans le réseau. L'idée est de combiner les informations des 4 représentations pour explorer rapidement les données et proposer une classification simple des nodes en fonction de la qualité des enregistrements. Ce catalogue a donc été pensé comme support à l'utilisation de ces données (en cours ou future). Il est disponible à travers les suppléments électroniques de Froment *et al.* (2022a). Les fichiers y ont été nommés File S1, S2, S3 et correspondent respectivement à la composante N, E, et Z. Par la suite, nous nous appuyerons sur quelques exemples extraits du catalogue pour illustrer nos observations. Notons que les pages du catalogue correspondant à chacun des nodes décrits dans l'ensemble du manuscrit figurent en annexe (figures A1 à A16).

Juger de la qualité des enregistrements pour proposer une classification des capteurs n'est pas trivial et dépend de l'application. Pour ce travail, nous avons donc classé les nodes après une étude qualitative du catalogue et en gardant à l'esprit que nous utiliserons principalement ces enregistrements pour des calculs d'amplification. La classification utilise un feu signalétique : rouge, l'enregistrement n'est pas considéré utilisable et le node ne sera pas utilisé dans les calculs ; orange, l'enregistrement est considéré utilisable mais présente des signatures particulières pouvant impacter les

calculs d'amplification ; vert, l'enregistrement est jugé de bonne qualité. La figure 2.6(a) montre le résultat de cette analyse avec le pourcentage de nodes classés en vert, orange ou rouge en fonction de la composante. Les résultats sont globalement similaires sur les 3 composantes : environ 2/3 des nodes sont classés en vert, 1/3 en orange. Deux nodes sur les composantes horizontales et un sur la composante verticale ont présenté des signes clairs de dysfonctionnement et ont donc été classés en rouge. Notons d'ailleurs à ce stade que l'ensemble des observations issues du contrôle qualité se sont avérées similaires sur les 3 composantes. Par simplicité, nous présentons donc dans la suite du chapitre uniquement les observations pour la composante Z.

Pour avoir une idée de la répartition spatiale des 3 classes, une carte est présentée sur la figure 2.6(b). Au premier ordre, on ne semble pas pouvoir distinguer de cohérence spatiale évidente. Cependant, lorsque l'on regarde de plus près, on remarque une certaine cohérence entre les nodes classés en orange et la position de grands sites industriels de la zone, comme dans le quart sud-est du TNS (comprenant les 4 réacteurs nucléaires EDF), dans un rayon d'1 km autour du barrage (cercle vert), le long des voies ferrées (en marron) ou encore de l'autoroute A7 (en violet).

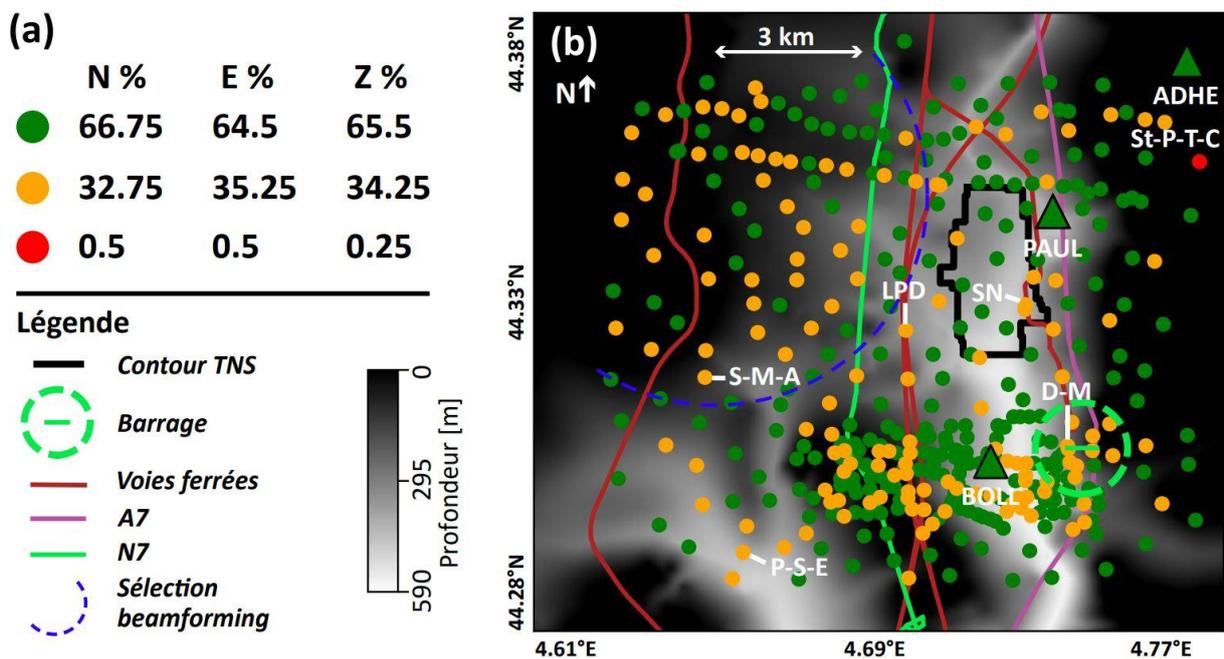


Figure 2.6 : (a) Pourcentage de nodes classés vert, orange ou rouge en fonction de la composante d'après le contrôle qualité proposé. (b) Cartographie des classes de nodes pour la composante Z et sélection des nodes pour le calcul de beamforming. Nous ne représentons que la composante Z, les résultats étant similaires sur les 3 composantes.

Pour mieux comprendre les résultats de cette classification, quelques pages du catalogue ont été extraites (Figure 2.7, Figure 2.8 et Figure 2.9). Dans un premier temps, nous comparerons un node classé en vert (BOLL) et un node classé en rouge (St-P-T-C, Figure 2.6b) sur la figure 2.7. Ensuite, nous essaierons d'identifier les signatures à l'origine de l'attribution des nodes à la classe orange. Pour cela, 5 nodes classés en orange seront étudiés. Sur ces 5 nodes, 2 montrent des signatures particulières à basse fréquence (Figure 2.8), 3 à haute fréquence (Figure 2.9).

Comparaison entre un node classé en vert et un node classé en rouge

Commençons par la comparaison d'un node classé en vert (BOLL) et d'un node classé en rouge (St-P-T-C, voir la localisation sur la figure 2.6b). A BOLL (Figure 2.7a), la trace temporelle et le spectrogramme révèlent des variations journalières et hebdomadaires de l'amplitude du bruit sismique au-delà de 1 Hz, reflétant l'activité anthropique dans la zone. En dessous de 1 Hz, l'amplitude est stable à l'échelle de temps observée. La distribution des FAS (panneau de droite) est centrée autour de la médiane sur toute la plage de fréquences, indiquant au premier ordre que l'enregistrement est stable. A St-P-T-C (Figure 2.7b) en revanche, la distribution des FAS est largement dispersée. Cette dispersion reflète des interruptions intermédiaires sur la totalité de l'enregistrement (de surcroît écourté) visibles sur la trace temporelle et le spectrogramme et indiquant un dysfonctionnement de l'acquisition. L'enregistrement a donc été considéré inutilisable.

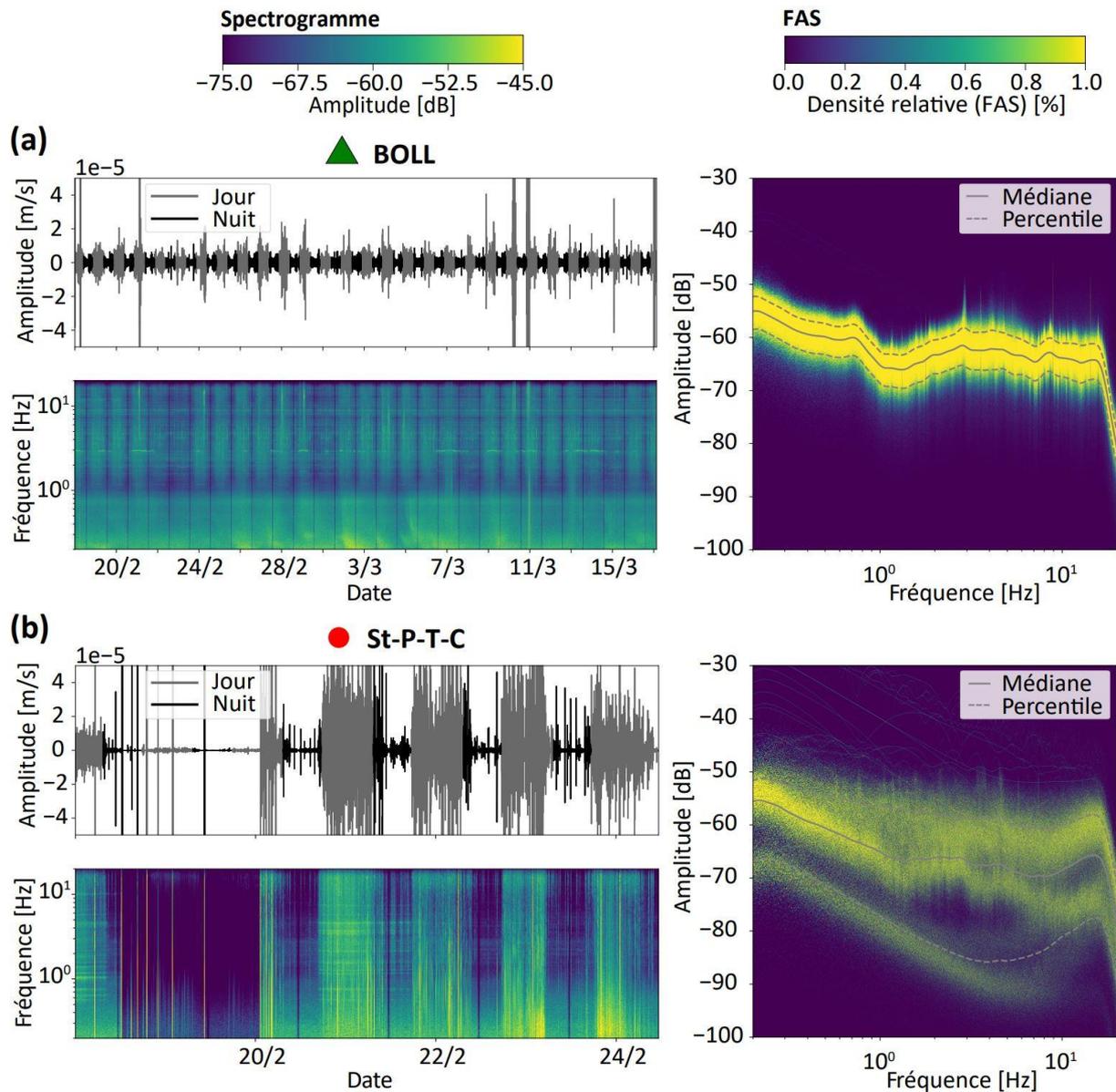


Figure 2.7 : Comparaison basée sur le catalogue entre un node classé en vert (a) et un node classé en rouge (b). La trace temporelle, le spectrogramme et la distribution des FAS sont représentés respectivement sur le panneau supérieur gauche, inférieur gauche et droit. Concernant la trace temporelle, les enregistrements diurnes et nocturnes sont respectivement en gris et noirs. La plage diurne est 6H-20H. Le spectrogramme et les FAS ont été calculés à partir de fenêtres de 10 min sans recouvrement. La médiane ainsi que les percentiles 84 et 16 des FAS sont représentés respectivement en traits pleins gris et pointillés gris. Pour des raisons pratiques, la disposition des panneaux a été changée par rapport au catalogue fourni dans la section supplément de Froment et al. (2022a). Les observations sont présentées pour la composante Z.

Nodes classés en orange en raison de signatures particulières à basse fréquence

On s'intéresse ici aux nodes S-M-A et P-S-E (Figure 2.8 a et b). On distingue nettement des variations quasi-journalières de l'amplitude du bruit sismique dans les spectrogrammes. Ces variations sont d'autant plus marquées à basse fréquence et se caractérisent par une distribution dispersée des FAS en-dessous de 1.5 Hz. Notons qu'à travers le catalogue, ce phénomène a été observé pour plusieurs nodes distribués de manière éparsée sur la zone d'étude sans cohérence spatiale apparente. Pour comprendre ces variations, nous nous sommes intéressés à la météorologie locale entre le 21 février et le 16 mars. Les données des rafales de vent pour des plages de 2H (vent le plus fort enregistré sur une fenêtre de 2H) ont été collectées (source : <https://www.infoclimat.fr>). On observe une forte cohérence entre la courbe des rafales de vent (en bleu ciel) et ces variations d'amplitude à basse fréquence : les plages temporelles de vents forts correspondent aux plages temporelles où l'amplitude du bruit sismique est maximale.

Pour aller plus loin, nous avons cherché à déterminer la direction dominante du bruit sismique en fonction du temps grâce au beamforming (e.g. Capon 1969 ; Lacoss *et al.* 1969). Le beamforming (Figure 2.8c) a été calculé sur deux plages temporelles de 2H chacune à partir de fenêtres de 2 min avec 50 % de recouvrement. Nous nous sommes concentrés sur une plage temporelle durant laquelle le vent est calme (le 29 février entre 22H00 et 0H00, à gauche) et une plage temporelle durant laquelle le vent est fort, comprenant notamment la plus forte rafale sur la totalité de l'acquisition (le 13 mars entre 15H00 et 17H00, à droite). Notons que les nodes utilisés pour le calcul du beamforming sont situés en-dehors de la partie sédimentaire profonde (sélection indiquée par un arc de cercle en pointillés bleus sur la figure 2.6b) afin d'éviter la diffraction des ondes par la vallée et de se rapprocher le plus possible de l'hypothèse d'une onde plane (condition à la base du beamforming). Notons également que les enregistrements ont été filtrés entre 0.45 et 0.9 Hz afin (1) de garantir une largeur de bande supérieure à la longueur d'onde maximale observée (limite basse de la bande de fréquences) pour résoudre correctement la direction (2) d'éviter le repliement spectral (limite haute de la bande de fréquences). Les résultats montrent que les sources principales de bruit sismique sont situées au N-O et S-O de la zone pour une plage temporelle durant laquelle le vent est calme. Ces directions pointent respectivement l'océan Atlantique et la Méditerranée. En revanche, pour une plage temporelle durant laquelle le vent est fort, aucune cohérence directionnelle n'est trouvée et les lenteurs ne sont pas contraintes. Ceci est certainement dû à l'action du vent qui agit, de manière directe (interaction directe avec le capteur) ou indirecte

(retransmission par une structure vibrante à proximité), comme une source locale brisant la condition d'approximation en ondes planes.

En conclusion, si le réseau est dominé dans son ensemble par des sources distantes (océan Atlantique et mer Méditerranée), certains nodes (comme S-M-A et P-S-E) sont cependant très exposés au vent. En dominant les enregistrements à certains nodes, le vent peut potentiellement masquer l'information liée à la réponse de la structure géologique (effets de site) localement et être ainsi parasite en vue des calculs d'amplification. Les nodes impactés par le vent ont donc été classés en orange.

Notons que plusieurs études ont fait état de l'effet du vent sur les enregistrements à des capteurs nodes ou large-bande (e.g. De Angelis et Bodin 2012 ; Naderyan *et al.* 2016 ; Lott *et al.* 2017). Si une partie de l'énergie du vent peut être retransmise dans le sol par des structures vibrantes au voisinage du capteur (arbres, clôtures...), l'effet le plus important reste l'interaction directe entre le vent et le capteur (Naderyan *et al.* 2016). Enterrer les nodes pourrait donc atténuer ces effets parasites et améliorer la qualité des enregistrements. Cela réduirait cependant leur facilité d'installation, principal atout de ces capteurs.

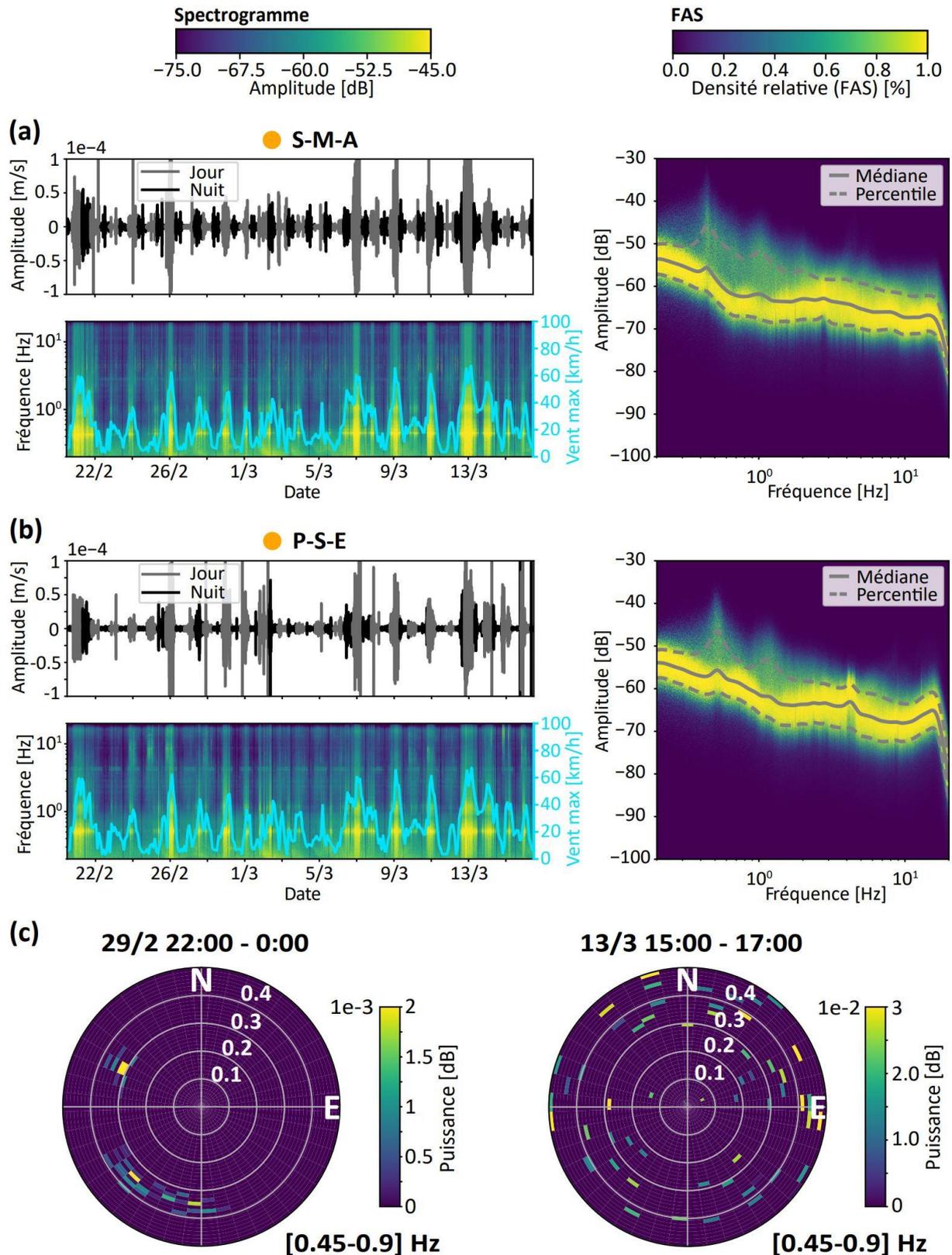


Figure 2.8 : (a) et (b) Idem que pour la figure 2.7 pour des nodes classés en orange. La courbe des rafales (vent le plus fort sur une plage de 2H) est représentée en bleu ciel. Les données de cette courbe ont été extraits du site <https://www.infoclimat.fr>. (c)

Beamforming pour le 29/2 entre 22H00 et 0H00 (à gauche) et le 13/3 entre 15H00 et 17H00 (à droite) calculé à partir de fenêtres de 2 min avec 50 % de recouvrement. Les enregistrements ont été filtrés entre 0.45 et 0.9 Hz. Les observations sont présentées pour la composante Z.

Nodes classés en orange en raison de signatures particulières à haute fréquence

Concentrons-nous maintenant sur 3 nodes présentant des signatures particulières à haute fréquence (LPD, D-M et SN, Figure 2.9). Au node LPD par exemple (a), le spectrogramme présente des variations journalières d'amplitude importantes à partir de 7 Hz. Ces variations sont bien visibles sur la trace temporelle et se traduisent par une distribution bimodale des FAS : l'amplitude des spectres calculés à partir des segments diurnes est nettement plus importante (jusqu'à une quinzaine de dB) pour des fréquences supérieures à 7 Hz. LPD étant situé à la jonction entre une ligne à grande vitesse et une ligne de fret (Figure 2.6b), nous expliquons ces variations journalières d'amplitude par le passage des trains.

Concernant les nodes D-M et SN (Figure 2.9 b et c), le spectrogramme révèle au contraire une amplitude étonnamment stationnaire à haute fréquence (à partir de 6 Hz pour D-M, 3 Hz pour SN). Ceci est cohérent avec la distribution très centrée des FAS autour de la médiane. Par ailleurs, l'amplitude dans ces gammes de fréquences est très élevée (jusqu'à environ -45 dB). Un rapide coup d'œil à la position de ces nodes permet de constater que D-M est situé à proximité du barrage de Donzère-Mondragon et que SN est situé dans le quart sud-est du TNS. Le quart sud-est du TNS comprend les 4 réacteurs nucléaires EDF et ainsi concentre l'activité principale du site. L'activité quasi-permanente du barrage et du TNS est cohérente avec la stabilité des enregistrements à haute fréquence.

En conclusion, ces signatures haute fréquence sont attribuables à l'activité humaine (TGV, barrage, activité autour du TNS). Les nodes présentant ces signatures ont été classés en orange. Les observations à ces 3 nodes permettent de comprendre la cohérence spatiale entre les nodes classés en orange et la position des principaux sites industriels (Figure 2.6b).

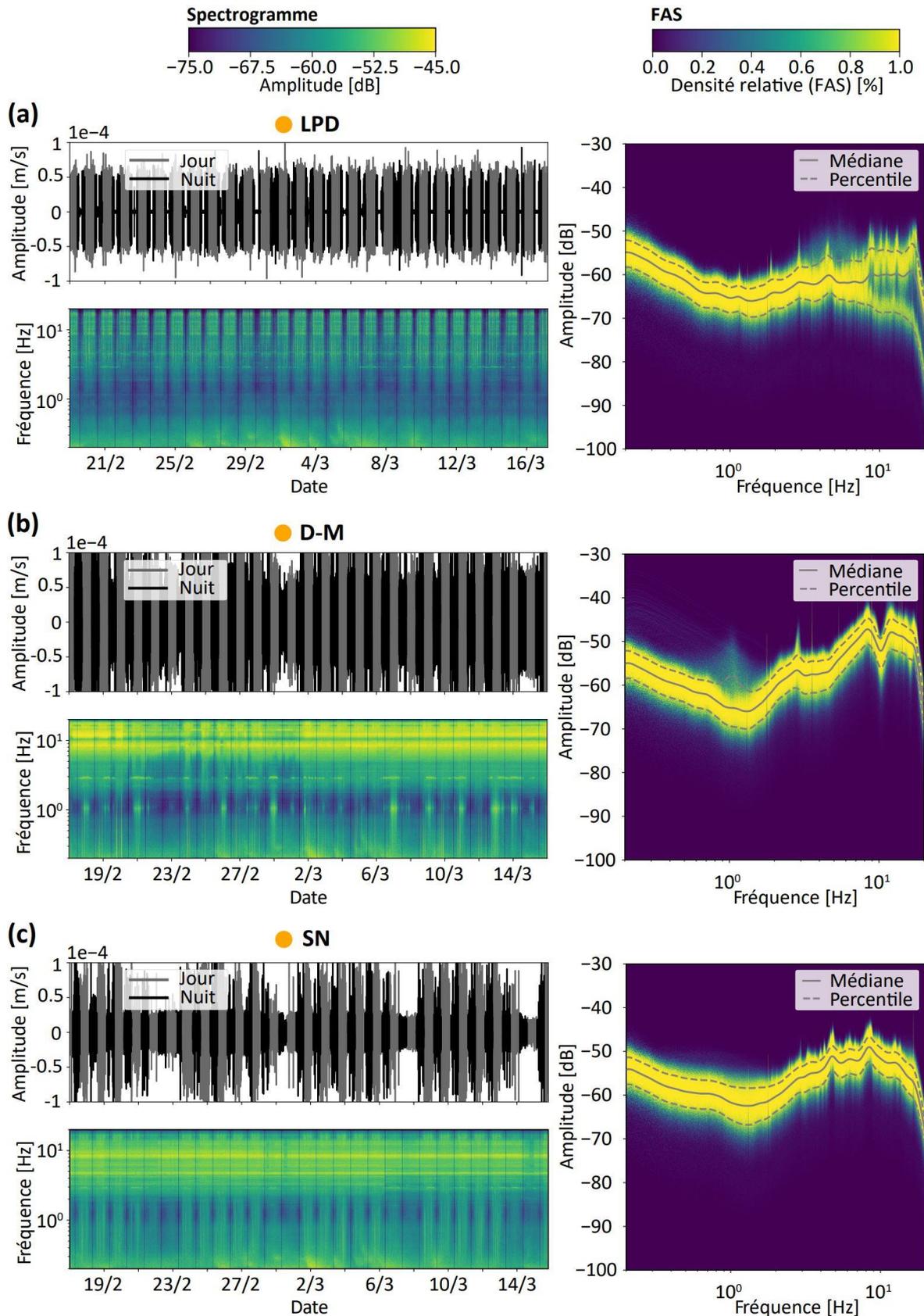


Figure 2.9 : Idem que pour la figure 2.7 pour des nodes classés en orange. Les observations sont présentées pour la composante Z.

Conclusion sur les nodes classés en orange

En résumé les signatures particulières dans les enregistrements des nodes classés en orange correspondent à des sources locales. A basse fréquence, il s'agit du vent, à haute fréquence de l'activité humaine. En dominant les enregistrements, ces sources peuvent potentiellement masquer l'information liée à la réponse de la structure géologique (effets de site) et être ainsi parasites en vue des calculs d'amplification (voir chapitre 1, section 1.4). Pour évaluer leur impact en amont des calculs d'amplification (chapitre 3), nous avons cherché à caractériser l'amplitude du bruit sismique, dans un premier temps à proximité des sources anthropiques majeures (le TNS, le barrage, le TGV et l'A7 ; Figure 2.10), dans un second temps à travers l'ensemble du réseau (Figure 2.11).

B. Caractérisation de l'amplitude du bruit sismique

A proximité des sources anthropiques majeures

Les nodes proches de chacune des sources anthropiques majeures ont été sélectionnés (Figure 2.10a). La sélection comprend donc 4 groupes de nodes : les nodes situés sur le périmètre du TNS (en noir), les nodes situés à moins de 250 m de l'A7 (en violet) et de la LGV (en bleu), les nodes situés à moins d'1 km du barrage (en vert). Pour chaque groupe (c.-à-d. chaque source), nous avons calculé la médiane des FAS à chaque node en séparant les enregistrements diurnes et nocturnes. Les résultats sont présentés dans les panneaux (b), (c), (d) et (e) associés respectivement au TNS, à l'A7, au barrage et au TGV. Premièrement, une vue générale de ces 4 panneaux révèle une différence d'amplitude entre les spectres médians diurnes et nocturnes au-delà de 1 Hz. Cette différence est très marquée pour les nodes situés le long de la LGV et s'explique par le passage des trains. Elle reste cependant assez faible concernant le TNS, l'A7 et le barrage. Concernant le TNS et le barrage, ceci est cohérent avec les observations précédentes et confirme l'activité quasi-permanente de ces deux sources. Concernant l'A7, l'observation faite est plus surprenante si l'on considère que le trafic s'atténue la nuit. En second lieu, on remarque une large dispersion des spectres médians pour le TNS et le barrage. La bande de fréquences pour laquelle cette dispersion est importante coïncide avec la bande de fréquences dans laquelle ces sources semblent dominantes (à partir de 3 Hz pour le TNS, 6 Hz pour le barrage ; Figure 2.9). Cette dispersion est simplement liée à la distance des nodes à la source : les spectres médians dont l'amplitude est la plus élevée correspondent aux nodes à proximité

directe de la source. Rappelons que concernant le TNS, les activités principales se concentrent dans le quart sud-est.

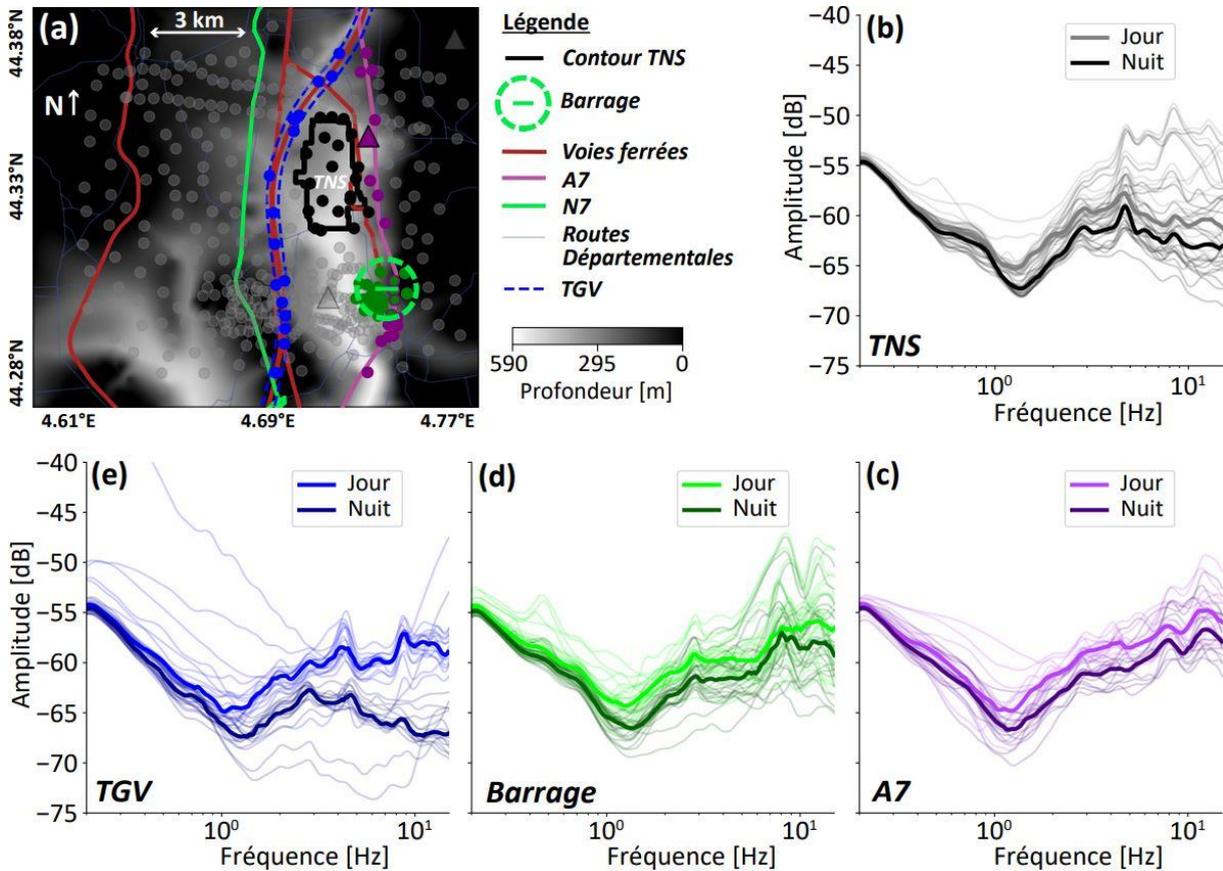


Figure 2.10 : Médiane des FAS (traits fins) calculés pour des segments diurnes et nocturnes aux nodes situés sur le périmètre du TNS (en noir, a et b), aux nodes situés à moins de 250 m de l'A7 (en violet, a et c) et de la LGV (en bleu, a et e), ainsi qu'aux nodes situés à moins d'1 km du barrage (en vert, a et d). La médiane des courbes diurnes et nocturnes est représentée en trait épais. Toutes les courbes ont été lissées en utilisant le filtre Konno-Ohmachi (Konno et Ohmachi 1998). Les FAS ont été calculés à partir de fenêtres de 10 min sans recouvrement. La plage diurne est 6H-20H. Les observations sont présentées pour la composante Z.

A l'échelle du réseau

L'amplitude du bruit sismique à chaque node a été représentée pour des bandes de fréquences particulières (Figure 2.11). Plus précisément, nous avons moyenné la médiane des FAS calculés à chaque node dans 4 bandes de fréquences : 0.45-0.9, 2.5-3, 5.5-6 et 8.5-9 Hz. Encore une fois, les enregistrements diurnes (colonne de gauche) et nocturnes (colonne de droite) ont été séparés. Concentrons-nous d'abord sur les résultats à basse fréquence (0.45-0.9 Hz, panneau a). Dans cette gamme de

fréquences, les amplitudes sont, au premier ordre, en accord avec la structure géologique : les amplitudes les plus importantes dessinent les parties profondes et sédimentaires de la vallée, quelle que soit la fenêtre temporelle utilisée (diurne ou nocturne). Au second ordre en revanche, on remarque plusieurs nodes présentant des amplitudes importantes dans des parties non sédimentaires. Ceci est particulièrement vrai pour les enregistrements diurnes. Après vérification, nous nous sommes aperçus que la plupart de ces nodes ont été identifiés dans le contrôle qualité comme étant fortement impactés par le vent (Figure 2.8). Ceci confirme que le vent joue localement un rôle important sur les amplitudes mesurées, notamment en ce qui concerne les enregistrements diurnes.

A haute fréquence (> 1 Hz, panneau b), pour toutes les bandes de fréquences, nous avons tendance à observer des amplitudes plus faibles la nuit, en particulier dans la partie centrale de la zone. Ceci est particulièrement nette entre 8.5 et 9 Hz, bande de fréquences pour laquelle des amplitudes beaucoup plus élevées sont observées le long de la ligne à grande vitesse pendant la journée. Entre 2.5 et 3 Hz, pour les enregistrements diurnes et nocturnes, des amplitudes plus élevées sont observées sur certains nodes situés au sud de l'empreinte TNS. Dans les bandes de fréquences 5.5-6 et 8.5-9 Hz, toujours pour des enregistrements diurnes et nocturnes, des niveaux d'amplitude très élevés sont observés le long de l'A7 et autour de l'empreinte du TNS. Enfin, de fortes amplitudes sont également observées à proximité du barrage hydroélectrique entre 8.5 et 9 Hz tant pour les enregistrements diurnes que nocturnes. La cohérence spatiale frappante entre la position des sources anthropiques principales et la distribution des plus fortes amplitudes illustre leur impact au-delà du Hertz. Au premier ordre, les niveaux d'amplitude semblent davantage refléter l'illumination est-ouest du réseau par ces sources que la structure géologique.

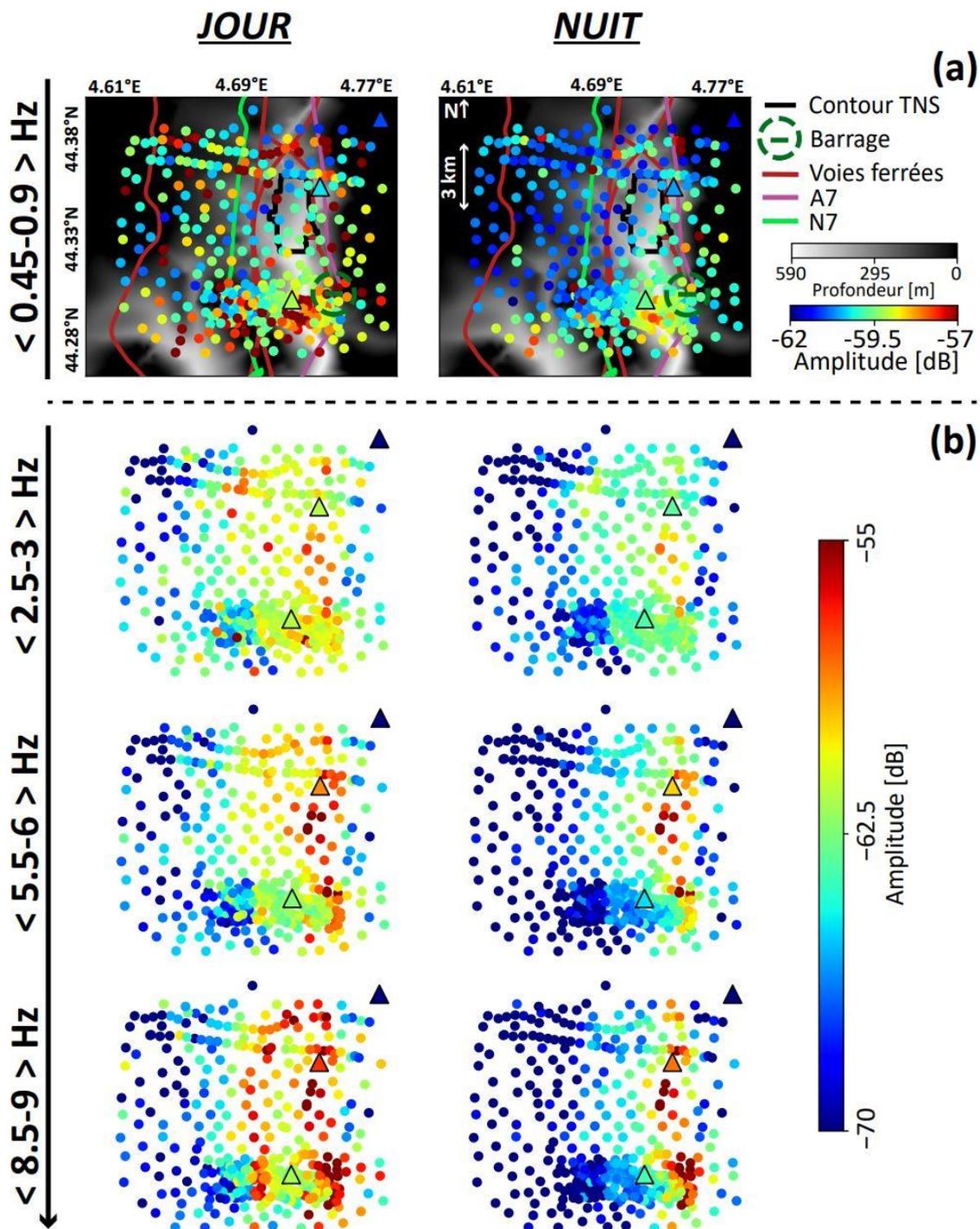


Figure 2.11 : Amplitude du bruit sismique à travers le réseau à basse (a) et haute fréquence (b) pour des enregistrements diurnes (colonne de gauche) et nocturnes (colonne de droite). A chaque node, la médiane des FAS a été moyennée dans 4 bandes de fréquences : 0.45-0.9, 2.5-3, 5.5-6 et 8.5-9 Hz. Les observations sont présentées pour la composante Z.

2.4 Conclusion

La vallée du Tricastin est propice à générer des effets de site importants et complexes en raison de sa géométrie incisée et du fort contraste lithologique entre l'épaisseur sédimentaire et le rocher sous-jacent. Les premières mesures SSR de Gélis *et al.* (2022) semblent aller dans ce sens, révélant une amplification locale sédiments/rocher supérieure à 6 sur une large gamme de fréquences. Dans la suite du manuscrit, nous chercherons à évaluer l'amplification associée aux effets de site dans la vallée via l'utilisation du bruit sismique mesuré sur le réseau de 400 nodes. En amont des calculs d'amplification, le contrôle qualité réalisé a permis de caractériser l'impact de sources locales qui semblent dominer les enregistrements au-delà de 1 Hz et masquer au premier ordre l'information liée à la réponse de la vallée. Certaines de ces sources, comme le vent et le TGV, s'atténuent la nuit. Ainsi, dans les calculs d'amplification, seules les fenêtres nocturnes seront considérées. En revanche, d'autres, comme le barrage de Donzère-Mondragon, le TNS ou encore l'A7, semblent très actives de jour comme de nuit. Dans le chapitre 3, une procédure sera notamment introduite pour sélectionner automatiquement les fenêtres de temps les moins impactées par des sources locales dans le but de fournir une estimation robuste de l'amplification.

Chapitre 3 : Evaluation des conditions d'application de deux méthodes site-sur-référence calculées à partir du bruit sismique dans le cas de la vallée du Tricastin

3.1 Introduction

Dans l'étude qui constitue le corps de ce chapitre, on cherche à évaluer l'amplification associée à la vallée du Tricastin à travers le SSRn et le SSRh (méthodes détaillées dans la section 1.4 du chapitre 1). La plupart des études a montré les limitations du SSRn au-delà de 1 Hz en raison de la proximité des sources liées à l'activité humaine (e.g. Field *et al.* 1990 ; Zhao *et al.* 1998 ; Bard 1999 ; Horike *et al.* 2001). Le SSRh en revanche a montré des résultats prometteurs pour 3 sites sédimentaires : Provence, au sud de la France, Argostoli en Grèce et Sion, dans la vallée suisse du Rhône ; Perron *et al.* (2018, 2022). Le chapitre 2 a révélé l'impact de plusieurs sources anthropiques locales qui dominent les enregistrements au-delà de 1 Hz et peuvent potentiellement biaiser l'estimation de l'amplification. L'objectif est ici d'évaluer le potentiel de ces méthodes, notamment le SSRh, dans l'estimation de l'amplification en zone fortement industrialisée. Une procédure de sélection automatique des fenêtres de temps les moins impactées par les sources locales y est également introduite dans le but de fournir une estimation plus robuste de l'amplification. Ce travail a été publié dans Geophysical Journal International (GJI ; Gisselbrecht *et al.* 2023).

3.2 Insights into the conditions of application of noise-based spectral ratios in a highly industrialized area: a case study in the French Rhone Valley

3.2.1 Abstract

The local ground motion amplification related to the geology at a specific site (i.e. the so-called site effects) may be classically quantified through the SSR (Standard Spectral Ratio) technique applied on earthquake recordings. However, such a quantification might be challenging to carry out in low-to-moderate seismicity regions. Methods based on background ambient noise, such as noise-based Standard Spectral Ratio (SSRn), might be of great interest in these areas. But noise-derived amplification is particularly sensitive to local anthropogenic sources, which may introduce biases in the evaluation of site effects by dominating the geological effects, especially for frequencies higher than 1 Hz. A hybrid approach (SSRh), developed to reduce biases in noise-based spectral ratios by combining classical earthquake-based spectral ratio (SSR) and SSRn, was recently introduced and relies on a site-reference. We here investigate the applicability of SSRn and SSRh in a heavily industrialized environment in the Tricastin Valley (South-East France), where critical facilities are located. We continuously recorded ambient noise from February to March 2020 on a 400-sensor seismic array covering an area of about 10 km by 10 km. We demonstrate that SSRn and SSRh computed below 1 Hz are able to reproduce amplification factors provided by SSR. By contrast, at frequencies higher than 1 Hz, SSRn strongly deviates from SSR. SSRh shows closer results to SSR but presents a dependence to the choice of the local site-reference, thereby questioning the possibility to use SSRh blindly to estimate the local amplification in our context. These discrepancies reflect the impact of local anthropogenic sources. We therefore introduced a two-step workflow to mitigate the influence of local sources. The first step is to define a characteristic time window to optimally isolate significant transient signals. The second step consists in selecting the time segments that do not contain these transients with a clustering-based approach. By applying this workflow, we were able to remove some strong anthropogenic transient signals likely to be generated by local sources at some sites and therefore to locally improve the amplification assessment through noise-based spectral ratios. However, stationary sources, whose impact cannot be removed through our procedure, remain a major issue. This study provides some insights into the application of SSRn and SSRh in noisy industrialized areas, especially regarding the impact of local noise sources. It illustrates the difficulty of having a procedure for mitigating the impact

of these sources that is efficient everywhere inside such a complex anthropized environment, where different types of sources (including stationary sources) cohabit.

Key words: Site effects; Seismic noise; Spectral ratio; Earthquake ground motions; Fourier analysis; Earthquake hazards.

3.2.2 Introduction

It is now well known that shallow complex geological structures can modify seismic motion. From a seismic hazard point of view, this modification of ground motion caused by local geological structures is usually referred to as site effects. In the case of sedimentary basins (or valleys), these site effects are mainly caused by the impedance contrast at the interface between soft soil layers filling the sedimentary basin and the bedrock. Several phenomena can thus take place within the basin, including body wave resonance, edge-generated surface waves and focusing effects. These phenomena can lead to a dramatic ground motion amplification and an increase in the duration of shaking when an earthquake occurs (e.g. Bard & Bouchon 1985; Celebi *et al.* 1987; Kawase 1996; Semblat *et al.* 2005). As cities and critical infrastructures may be located on top of such geological structures, it is of great importance, for the purpose of seismic hazard assessment at these specific sites, to have a reliable method in predicting the local amplification that would occur during an earthquake.

Earthquake and noise-based spectral ratios for local amplification assessment

Several approaches have been introduced since the 70s to assess the local amplification associated with site effects. The well-known SSR (Standard Spectral Ratio, Borcherdt 1970), has been widely used since its introduction as it provides reliable estimates of amplification factors (e.g. Field & Jacob 1995; Raptakis *et al.* 1998; Huang & Teng 1999). This method consists in computing the ratio of the Fourier Amplitude Spectra (FAS) recorded at two stations for a given earthquake. Local amplification is estimated relatively to a reference station. The numerator FAS corresponds to the ground motion observed at a station located at the site under study (i.e. on top of the sedimentary filling) and the denominator FAS corresponds to a rock-reference station. The choice of the reference station is sometimes not straightforward, as it must ensure that site effects are negligible at this station, and has been discussed in numerous papers (e.g. Steidl *et al.* 1996). The inter-station distance must be small with respect to the earthquake distance, so that both stations are supposed to experience the same source and path effects on the seismic motion. As a result, the source and path effects

cancel out in the SSR, thereby isolating the site effects. This can be summarized as follows:

$$SSR_{s/r,k}(f) = \frac{FAS_{s,k}}{FAS_{r,k}} = \frac{S_k(f) \times P_k(f) \times H_{s,k}(f)}{S_k(f) \times P_k(f) \times H_{r,k}(f)} = H_{s,k}(f), \quad (1)$$

where f is the frequency. $SSR_{s/r,k}(f)$ is the SSR between station s located at the site under study and the rock-reference station r , for an earthquake k . $S_k(f)$ and $P_k(f)$ are respectively the source and path terms. $H_{s,k}(f)$ is the amplitude of the site transfer function at s . By assuming that site effects are negligible at the rock-reference station, $H_{r,k}(f)$ is considered to be equal to 1. Therefore $H_{s,k}(f)$ estimates the local ground motion amplification factor between s and r .

SSR technique may be however difficult to apply in low-to-moderate seismicity areas as it requires the recordings of many earthquakes with a good signal to noise ratio over a wide frequency range to get a statistically robust estimate (e.g. Perron *et al.* 2020). Ambient noise on the other hand appears to be an appealing source of data in low-to-moderate seismicity zones. This weak seismic wavefield, generated by natural or anthropogenic sources, has the characteristics of being permanent and exploitable everywhere on the surface of the Earth. Therefore, ambient noise-based approaches may be of great interest where good quality earthquake recordings are challenging to get. In this context, Irikura & Kawanaka (1980) introduced the noise-based Standard Spectral Ratio (SSRn), which is equivalent to the SSR (eq. 1) but relies on ambient noise recordings. The first studies using this method for the purpose of site responses assessment were these of Kagami *et al.* (1982, 1986). They applied the SSRn using long-period microtremor recordings (1-5 sec) in the Niigita Plain (Japan) and in Los Angeles (California, USA), where the geological properties at the surface may have played an important role in the strong ground motion observed during the Mw 7.6 Niigita and Mw 6.6 San Fernando earthquakes, which respectively occurred in 1964 and 1971. The amplification estimated through SSRn was found to be qualitatively consistent with the local sedimentary deposit, and the authors concluded that this easy-to-use method should be applied to other surveys at low frequencies. Later, several studies qualitatively compared SSRn results, either with the geology of the site under study (e.g. Ferritto 1995), or with the SSR amplification factor (for an extended review, see for instance Bard 1999 or Perron 2017). Most of these studies have shown that SSRn can provide comparable measurements to SSR at low frequency but largely overestimates the amplitude of the site transfer function obtained through SSR approach at higher frequencies if the site to reference distance is greater than several hundreds of meters (e.g. Zhao *et al.* 1998). This discrepancy is generally explained by

local anthropogenic noise sources that are close to the sensors at the site, but far from the rock-reference sensor (e.g. Horike *et al.* 2001). Hence, the source signature at both locations is different and the method is failing in isolating the amplitude of the site transfer function. In other words, the background noise wavefield is not necessarily fulfilling the required hypothesis underlying the simplifications of eq. 1 which occur for a distant earthquake and are at the basis of the SSR approach.

Perron *et al.* (2018) introduced a new approach combining SSR and SSRn techniques for site effects estimation in sedimentary basins and that may mitigate such limitations. The bottom line of the so-called SSRh (hybrid Standard Spectral Ratio, eq. 2) is first to use earthquake-based SSR to measure the amplification at an intermediate station *sref* (located within the sedimentary basin) relative to a rock-reference station. Then the amplification variability between *sref* and a station *sx* in the basin is estimated through site-to-site noise-based SSRn computation. *sref* is therefore used as a local site-reference. This might be expressed as:

$$SSRh_{sx/r} = SSR_{sref/r} \times SSRn_{sx/sref} = \frac{FAS_{sref}}{FAS_r} \times \frac{FASn_{sx}}{FASn_{sref}}, \quad (2)$$

where $SSRh_{sx/r}$ is the SSRh at a sensor *sx* located at the site under study. *sref* and *r* are respectively the local intermediate reference station located within the basin and the rock-reference station outside the basin. *FAS* and *FASn* are the Fourier Amplitude Spectra computed respectively with earthquakes and noise recordings. Perron *et al.* (2018) have compared both SSRh and classical SSRn to SSR for two basins, in Provence (France) and in Argostoli (Greece). Their results showed that SSRh was able to reproduce SSR for a wide frequency range (up to 12 Hz in Provence and 20 Hz in Argostoli). They also observed unrealistic amplification factors with SSRn for frequencies higher than the minimal resonance frequency f_0 of the basins. However, as the rock-reference station was located on a noisier area than the stations within the basin in Argostoli, and as no clear relation with site-to-reference distance was found in the SSRn results for both basins, they did not assign this discrepancy to the influence of local anthropogenic noise sources. They rather proposed that the sedimentary basin acts as a natural barrier, so that part of the ambient noise wavefield is trapped within it at frequencies higher than f_0 . As a result, only a limited part of the wave energy generated within (or crossing) the basin would reach the rock-reference station above f_0 . Such behavior is not observed in SSR because body waves generated by a distant earthquake reach the Earth's surface with a quasi-vertical incidence. This interpretation explains not only the discrepancy between the rock-referring SSRn and the SSR above f_0 , but also the agreement observed in Perron *et al.* (2018) between

SSRh (that uses SSRn between 2 stations under the same geological conditions) and SSR over a wide frequency band. The authors concluded that SSRh should be tested to other sites in order to better characterize its conditions of applicability. More recently, Perron *et al.* (2022) proposed a site-specific hazard map derived from an empirical amplification model based on SSRh for the Sion area (Swiss Rhone Valley).

Applicability of noise-based spectral ratios

Perron *et al.* (2018) show that SSRh is a promising method to mitigate the impact of the difference in the geological conditions between sensors at the site and the reference location in noise-based spectral ratios. The authors also explained the good results of SSRh by showing that site-to-site SSRn is in good agreement with site-to-site SSR up to 12 Hz in Provence basin for both horizontal and vertical components, whatever the phase of the signal that is considered for SSR computation. This result tends to validate an important assumption for noise-based spectral ratios applicability to assess site response, which is that the seismic response of a site to an ambient noise wavefield, usually considered as dominated by surface waves (e.g. Bonnefoy-Claudet *et al.* 2006a), must be comparable to that of the incident seismic (body and surface) waves generated by earthquakes (Lermo *et al.* 1994; Mori *et al.* 2016; Kawase *et al.* 2019). However, a second assumption is necessary, in which the only resulting factor when computing the spectral ratio corresponds to the amplitude of the site transfer function, meaning that this ratio makes it possible to isolate the site influence in the ambient noise wavefield. This assumption is challenged by the presence of strong local sources of seismic energy. Ignoring the influence of such strong local sources would mean relying on some kind of “local ergodicity” assumption, i.e. any point of the array experiences a fully equipartitioned noise field, even if not excited by the same sources. In other words, all oscillating modes are locally excited at any point of the array meaning that a spectral ratio between the wavefield amplitude at two locations would mainly capture the local geological structure imprint, whatever the sources. Although not straightforward to verify, it is most likely not a valid hypothesis for most highly anthropized sites, especially at high frequency where localized and dominant sources like a railway or a persistently vibrating dam could itself break such assumption.

Purpose of our study

In this study, we investigate the potential of SSRn and SSRh in a highly industrialized region in the area of the Tricastin Nuclear Site (TNS, South-East France). This will allow us to complement the work by Perron *et al.* (2018, 2022) by investigating SSRn and

SSRh in a very anthropized area. Our objective is to give some further insights into the conditions of application of SSRn and SSRh for them to reproduce SSR, especially regarding the impact of local noise sources. In the following, we first briefly introduce the target area and the data acquisition before identifying the main anthropized features of our site of study. Such analysis is important to better understand amplification results that reflect here, the influence of strong sources in local ground motion. Hence, we propose a two-step workflow to attenuate source effects in local amplification assessment through noise-based approaches. We then show results of noise-based spectral ratios, considering our workflow or not. We discuss the impact of our workflow for the utilization of SSRn and SSRh in noisy industrialized environments. We here speculate that by selecting quiet periods within acquisition, and thus muting strong transient signals emanating from local sources, we are getting closer to the assumption we refer to as “local ergodicity”, at least in some parts of the array.

3.2.3 Context & data acquisition

Context of the study

The Tricastin region (French Rhone Valley, Fig. 1a) is located on top of an elongated valley that was dug during the Messinian Salinity Crisis in Cretaceous sandstones and limestones and that is now filled with Pliocene and Quaternary sediments (sands and clays) (Gélis *et al.* 2022). The site is interesting to study for several reasons. First, in terms of seismic hazard, since 1) the expected strong lithological contrast between the sedimentary filling and the bedrock, as well as the valley's incised geometry, is prone to generate strong and complicated site effects; 2) the area hosts critical infrastructures including the large Tricastin Nuclear Site (TNS) and the hydroelectric dam of Donzère-Mondragon, as well as 5 towns with a total population over 45 000. Second, in terms of experimenting SSRh method in a different context than in Perron *et al.* (2018), as the Tricastin Valley is larger than the studied basins in Provence and Argostoli and as the targeted area is heavily industrialized. Indeed, in addition of hosting the TNS and the hydroelectric dam, the site under study is crossed from north to south by 2 freight lines and 1 high-speed railway line as well as the busy A7 highway and N7 national road. These main expected anthropogenic sources are displayed in Fig. 1(b). In the context of investigating noise-based spectral ratios computations, it is worth noting the unequal repartition between noisy environments, mostly located on top of the sedimentary filling (i.e. mostly the center area of the valley), and quiet environments, rather located on top of the cretaceous outcrops (on both sides of the valley).

In November 2019, a Mw 4.9 earthquake occurred on La Rouvière Fault (LRF), located about 20 km away from the targeted area (Fig. 1a). This event, which is the strongest in metropolitan France since 1967, took place near the Le Teil village. Strong ground motion and severe damage were observed in the villages located in the vicinity of the rupture zone (Ritz *et al.* 2020; Causse *et al.* 2021; Cornou *et al.* 2021). Schlupp *et al.* (2021) estimated, on the basis of field observations and communal surveys, a maximum intensity (EMS98) of about VII-VIII in Le Teil and at least VI in 30 nearby localities. This recent earthquake further highlights the importance of conducting seismic hazard studies in the area.

Data acquisition

For the purpose of conducting seismic hazard studies in the area of Tricastin, we continuously recorded ambient noise from February to March 2020 on a 400-sensor seismic array (all-in-one 3-components Geospace GSX nodes) in the framework of the French-German DARE (Dense ARray for seismic site effect Estimation) project. The array was designed on the basis of 5 sub-arrays (Fig. 1a): a loose grid over the entire zone; a very dense grid located 2 to 3 km south of the TNS and covering the narrowest part of the valley with an inter-station distance of about 200-250 m; 2 dense East-West lines that follow 2 departmental roads in the north of the targeted area; 6 distant nodes covering different azimuths. One of these distant nodes was located on LRF. The total size of the combination of the first 4 sub-arrays, that is the core of our target zone, is about 10 km by 10 km (Fig. 1). For more information on the experimental design and the global dataset, we invite the reader to refer to Froment *et al.* (2022a). Three nodes were co-located with three seismic stations (called ADHE, BOLL and PAUL, Fig. 1) that have been deployed intermittently since 2016-2017 and have therefore recorded the seismicity for several months (Gélis *et al.* 2022). BOLL and PAUL sites are located on top of the sedimentary filling whereas ADHE is placed on top of the cretaceous outcrop. Gélis *et al.* (2022) showed almost flat H/V ratio with a value around 1 between 0.2 and 10 Hz at ADHE and therefore used this station as rock-reference for SSR computation.

In order not to blindly use continuous data in the spectral ratios, we evaluated at each of the 400 nodes the data quality for the East-West, North-South and vertical components. For this purpose, we constructed a catalog in which the continuous time series, the FAS based on a 10-min long segmentation as well as the spectrogram are computed for each sensor. Based on a qualitative examination of this catalog, 2 sensors for the East-West and North-South components, as well as 1 sensor for the vertical component were considered as not suitable for the purpose of this study (and

therefore not considered) as no time window and frequency range were considered exploitable for these nodes. For more information about this survey, see Froment *et al.* (2022a) since the complete catalog is attached as electronic supplement.

With the aim of investigating the spatial variability of the low frequency noise amplitude, we computed the median over all the 10-min long windows FAS (used in the data quality survey) at each sensor and averaged within the [0.25-0.75 Hz] frequency band (Fig. 1b). Note that in our computations, the FAS correspond to the Fourier transform of non-overlapping sliding windows and that no smoothing filter was applied here, neither to the FAS nor to the median of the FAS (called median FAS through the paper). Fig. 1(b) reveals a good agreement between the noise amplitude and the geology revealed by the topography of the bottom of the canyon deduced from the work of Bagayoko (2021, *M. Sc. thesis*). Indeed, higher amplitudes are observed on top of the sedimentary filling where the valley is the deepest (lightest area on the background map). By contrast, lower noise amplitudes are observed for nodes located on top of outcropping bedrock (darkest areas). It is worth noting that, according to Froment *et al.* (2022b), the minimal resonant frequency f_0 of the valley is about 0.5 Hz (hence situated in the middle of the frequency band considered). This suggests that the ambient noise captures, at least qualitatively, the effect of local amplification around the resonant frequency of the valley, and that we may relate to the effect of the overall geological structure. Note that the results are shown for the E-W component.

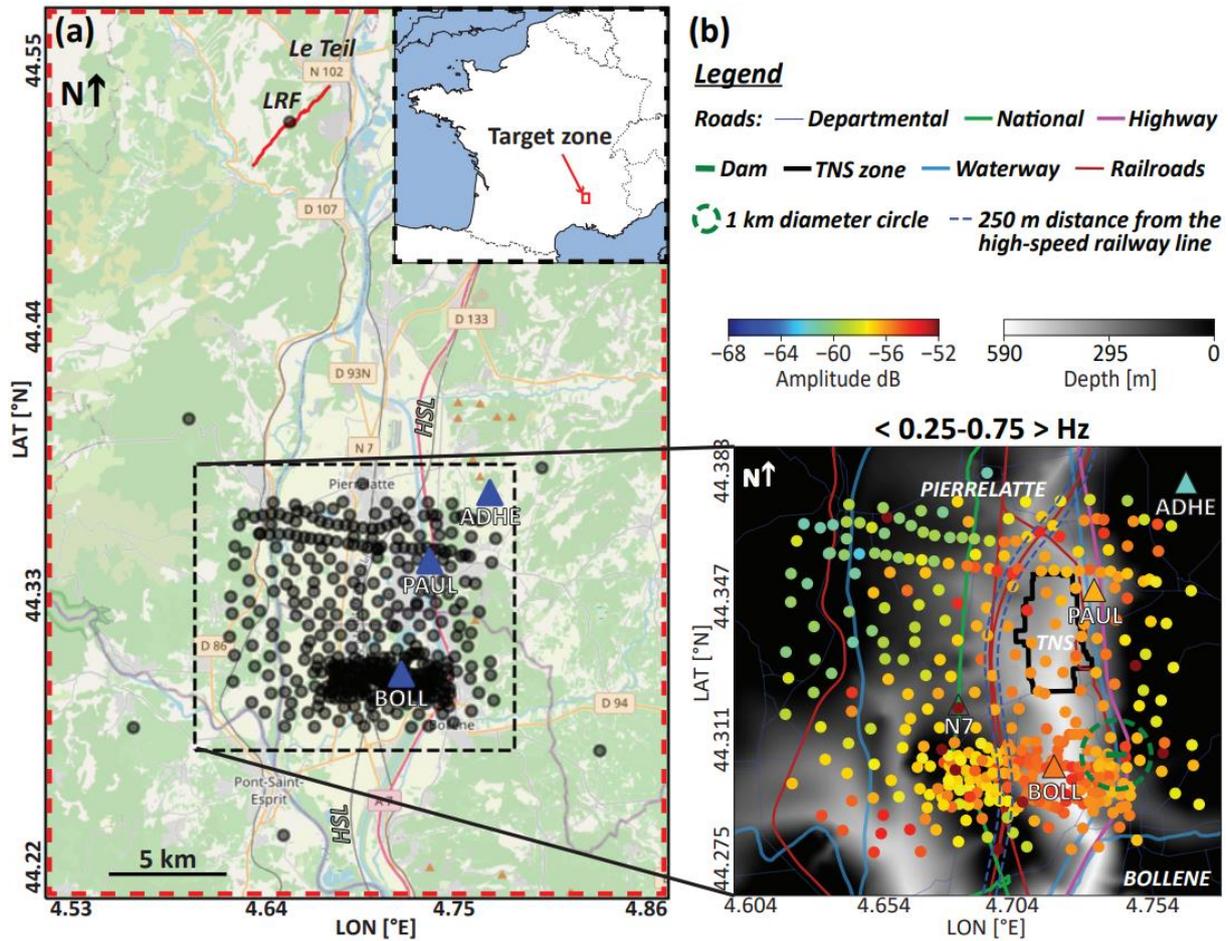


Figure 1. (a) Location of the study area. The thin red line represents the La Rouvière Fault (LRF). The nodes are represented in grey dots. BOLL, PAUL and ADHE stations are represented by blue triangles. The dashed black rectangle delimits the core of our target zone. (b) Median FAS at each sensor, averaged between 0.25 and 0.75 Hz (colored dots), superimposed with the depth of the bedrock (grey scale; Bagayoko 2021, M. Sc. thesis). The FAS are computed from 10-min long signal windows over the whole recordings. The main anthropized features are displayed and referenced in the legend. The results in (c) are shown for the E-W component.

3.2.4 Characterization of the local sources

In order to better understand amplification results obtained through noise-based methods at a specific site and to investigate the impact of local noise sources, we start our analysis by examining the different noise sources in the study zone. For this purpose, we averaged the median FAS within higher frequency bands than in section 2.2 (Fig. 2a). We separated daytime (from 6 a.m. to 8 p.m.) and nighttime (8 p.m. - 6 a.m.) recordings, to better characterize the impact of human activities on the seismic noise wavefield. Indeed, it is usually assumed that above 1 Hz, ambient noise is mostly dominated by anthropogenic sources near inhabited areas (Bonney-Claudet *et al.* 2006a). First, Fig. 2(a) shows that for all frequency bands, we tend to observe smaller noise amplitudes at night, especially in the central part of the area. This is particularly clear within 8.5-9 Hz and 11.5-12 Hz, for which much higher amplitudes are observed along the high-speed railway line during the day. Between 2.5 and 3 Hz, for both daytime and nighttime recordings, higher amplitudes are observed at some sensors located to the south of the TNS imprint (Fig. 1b). In the [5.5-6 Hz], [8.5-9 Hz] and [11.5-12 Hz] frequency bands, again for both daytime and nighttime recordings, very high noise amplitude level is found for sensors placed all along the Western side of the A7 highway and around the TNS imprint. Finally, strong amplitudes are also seen nearby the hydroelectric dam in the [8.5-9 Hz] and [11.5-12 Hz] frequency bands for both daytime and nighttime records. The striking spatial consistency between the distribution of the main identified noise sources (Fig. 1b) and high noise amplitude patterns illustrates the impact of local anthropogenic sources for frequencies of a few Hertz (Fig. 2a).

Let us now focus on the impact at specific sensors close to dominant sources. We plotted in Fig. 2(b), the median FAS (thin solid lines) for sensors located in the vicinity of the high-speed railway line (left panel) and of the hydroelectric dam (right panel). We consider to be in the vicinity, sensors located up to 250 m away from the high-speed railway line (i.e. dashed blue lines in Fig. 1b) and within a 1 km radius from the dam (dashed green circle in Fig. 1b). We smoothed each median FAS with a Konno-Ohmachi smoothing (Konno & Ohmachi 1998), taking b parameter equal 40 (eq. A1 in appendix section). First of all, the ambient noise amplitude in the vicinity of the high-speed railway line is significantly higher during the day than at night for frequencies higher than 1 Hz, which is consistent with the associated daily activity. By contrast, almost no difference is seen between daytime and nighttime recordings for sensors close to the hydroelectric dam (right panel), that is indeed not expected to stop during nighttime. Looking more into details on the left panel, we notice two peaks between 4

and 4.7 Hz and around 9 Hz for daytime. During the quality survey, these peaks have been found to be observed mainly for sensors located in the vicinity of the high-speed railway line at day. Hence, we consider these peaks as the spectral signature of high-speed trains traffic. Finally, regarding the right panel, the spectral signature of the dam has been found to be characterized by a broadband peak between 7 and 10 Hz. This analysis illustrates the variety of spectral signatures of the different anthropogenic sources we are dealing with in the area. Note that, as mentioned in section 2.2, we simply consider the Fourier transform of non-overlapping sliding windows for the FAS computation. We can notice that alternative processing, for instance the use of a multitaper approach (Prieto *et al.* 2009; Burjanek *et al.* 2018) to compute the FAS, could be explored to further discriminate sharp frequency peaks, as the ones discussed here and related to human activities.

Spectrograms of recordings at BOLL and PAUL stations are represented in Fig. 2(c) to have a closer look at the spectral content at these stations that will be used later as site-reference in the SSRh. On both spectrograms, we can see that below 0.9 Hz, despite some slight fluctuations, the noise amplitude level remains quite stable over time which is consistent with the commonly assumed natural origins of the background noise at these frequencies (oceanic activity and large-scale meteorological conditions). At higher frequencies, we observe daily and weekly variations as expected for human activities. When paying more attention, we notice two daily peaks around 1 Hz and 3 Hz. As no clear spatial consistency was found to properly discriminate these peaks during the quality survey, we do not interpret their origin as being related to processes at the scale of the basin. These features are likely to reveal very local effects. For BOLL station, a clear energy peak is also present around 9 Hz. It is however probably not related to the railway activity (as the 9 Hz feature discussed above) given its stability through time. For PAUL station, the daily variations are not that pronounced, which is consistent with the quasi non-stop activity of the nuclear facilities in the vicinity.

Fig. 2 highlights the complexity of the noise signal above 1 Hz inside the valley. This complexity is related to the specific distribution of local sources. In such an anthropized context and for noise-based spectral ratios applications, it is crucial to be able to decouple the contribution of the geology from the influence of these local sources on the ground motion.

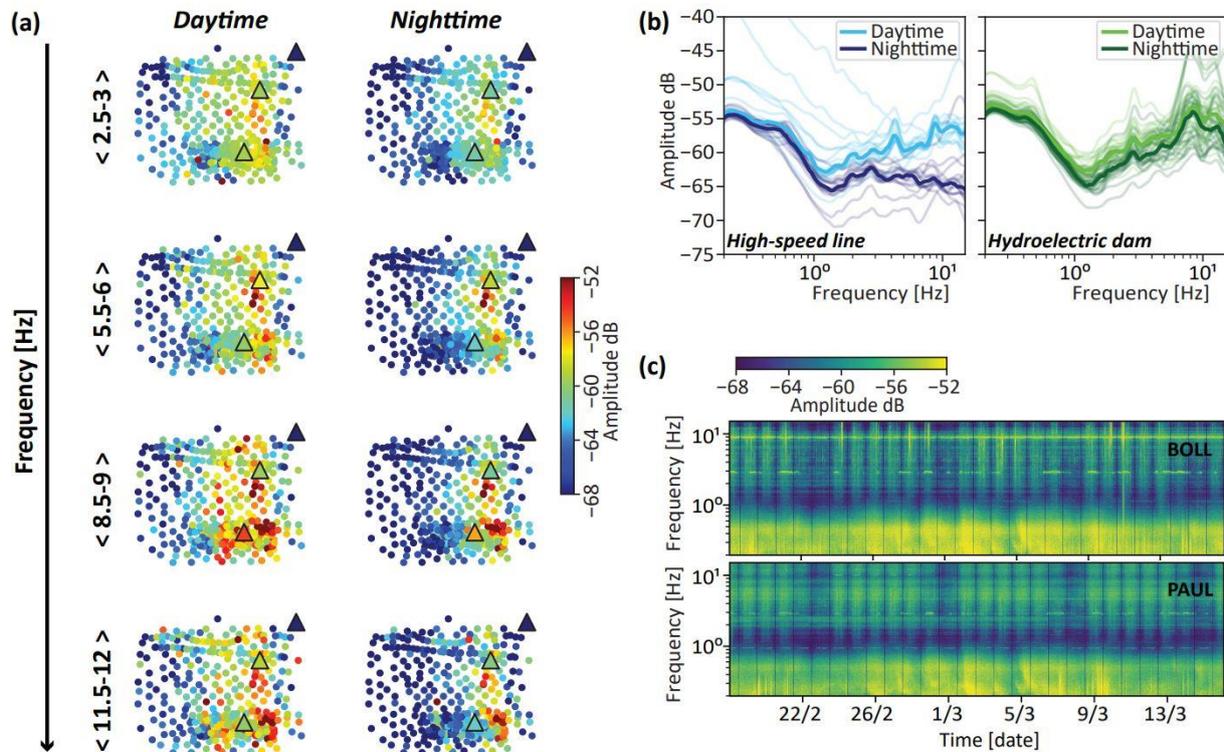


Figure 2. (a) Median FAS at each sensor for daytime and nighttime recordings, averaged over different frequency bands. In the published paper (Gisselbrecht et al. 2023), the color bar is missing. (b) Median FAS (thin solid lines) for sensors located in the vicinity of the high-speed railway line (blue lines in left panel) and of the hydroelectric dam (green lines in right panel) for daytime and nighttime recordings. We consider to be in the vicinity, sensors located up to 250 m away from the high-speed railway line (i.e. dashed blue lines in Fig. 1b) and within a 1 km radius from the dam (dashed green circle in Fig. 1b). The blue and green thick solid lines show the median for daytime (light) and nighttime (dark) curves. (c) Spectrogram at BOLL and PAUL stations for frequencies between 0.2 and 15 Hz. All results in this figure correspond to the E-W component.

3.2.5 Proposed workflow to mitigate local sources impact on amplification assessment

From the analysis of local sources realized in the previous section, comparing local ground motion without removing their effects might end up highlighting abnormally high and localized patterns of amplification which are not related to the underlying geological structure. For the computation of the so-called H/V ratios and noise correlation functions, some strategies have been proposed to mitigate biases due to local sources in noise recordings by removing time-impulsive signals in the data (e.g. SESAME team 2004; Bensen *et al.* 2007; Denolle *et al.* 2013; Viens & Iwata 2020). For SSRn and SSRh computation, Perron *et al.* (2018) followed the processing proposed by Denolle *et al.* (2013). These transient signals are indeed likely to be generated by local transient disturbances like footsteps, cars, trains, etc. One commonly used strategy to mitigate transients, is the so-called STA/LTA anti-triggering filter (e.g. SESAME team 2004). However, some tests performed on our dataset suggest the difficulty of choosing optimal time windows in the STA/LTA procedure to properly detect and discriminate the transients when their repeatability in time is high. In this study, we introduce an alternative procedure based on a two-step workflow. Note that in comparison to the STA/LTA procedure, the proposed workflow aims at isolating and removing transient signals based on a spectral representation of the wavefield. Our workflow is therefore consistent with the purpose of spectral ratios computation.

The first step of our workflow consists in defining an optimal duration of the time window before computing the FAS. The initial definition of 10-min long durations was ad hoc. Like classical tradeoff in time-frequency analysis, the longer the window the better the frequency resolution and the representativity of the current FAS as compared with the continuous data, but the lower our ability to separate transient signals. Since our goal is to remove the contribution of significant transient signals, the optimal window would then be the largest window able to properly discriminate the different transient sources whatever their time and frequency characteristics. In order to better characterize the impact of the time window duration (T_w) on the data, we compared the FAS relative density for $T_w = 10$ min (Fig. 3a) and $T_w = 1$ min (Fig. 3b) at a specific sensor located nearby the N7 national road (the so-called N7 site in Fig. 1b). The FAS relative density simply gives the relative amount of spectra at a given amplitude and a specific frequency. When looking at the highest amplitudes over the whole frequency band, lower density is found for $T_w = 1$ min compared to $T_w = 10$ min. We also notice two distinct behaviors whether looking below 4 Hz or above. Below 4 Hz, several spectra separate from the rest of the data reaching high amplitudes and

causing a large dispersion around the median FAS for $T_w = 10$ min while a very low amount of spectra reaches these amplitudes for $T_w = 1$ min. On the other hand, above 4 Hz, all the spectra are relatively concentrated around the median FAS for $T_w = 10$ min while several separate from the rest of the data and reach very high amplitudes for $T_w = 1$ min. From these observations, we suggest that, depending on the frequency, a certain amount of 10-min long segments contain some strong time-impulsive signals of duration shorter than 10 minutes and that weigh the amplitude when computing the FAS. Below 4 Hz, some 10-min long segments are impacted. Above 4 Hz, the repetition in time of the characteristic transients may be high enough to impact the large majority of the segments. By doing very short time windows (i.e. $T_w = 1$ min, Fig. 3b), we concentrate these transients into some segments while the rest of the data is now free from these spurious arrivals and thus characterized by a lower amplitude.

To further understand how the transients are isolated according to the duration of the time window, we evaluated at each frequency, the median of the absolute deviations (MAD) from the median FAS (thin solid red line in Figs 3a and b) depending on T_w (Fig. 3c). The MAD is a measure of dispersion. The benefit of using the MAD instead of the standard deviation is that it might be more robust to outliers (e.g. Peter J. Rousseeuw & Christophe Croux 1993; Leys *et al.* 2013). This analysis may be seen as a quantification of the distribution of transient signals over time. The two behaviors described above are clearly emphasized. Below 4 Hz (blueish curves), the MAD (i.e. the dispersion) is quite high for long segments between $T_w = 8$ min and $T_w = 60$ min; this is because a significant portion of temporal segments contains strong transients, causing a large dispersion (like in Fig. 3a). Reducing T_w ends up densifying the data around the median FAS by decreasing the portion of segments impacted. Hence, the MAD decreases with T_w from $T_w = 8$ min. On the other hand, above 4 Hz (greenish to reddish curves), the MAD is low for long segments (i.e. T_w superior or equal to 10 min); this is because the large majority of the segments may contain strong transient signals. As a result, while isolating transients by decreasing T_w , it increases the dispersion produced by high amplitude transients as seen in Figs 3(a) and (b). This dispersion reaches its maximum when T_w is approximately equal to what could be the actual characteristic length of a transient signal (see the maximal deviation just around 1-min time window for high-frequency range, Fig. 3c). Below this peak, the probability of splitting a transient into 2 different segments (and therefore reducing its relative impact on each segment) increases, decreasing the MAD. The fact that this peak is shifted to lower T_w when the frequency is increasing is consistent with the assumed length of a transient depending on its frequency.

For T_w higher than 60 min, each segment becomes more and more representative of the average amplitude level of the whole recording, so that all the spectra will again converge towards the median FAS (i.e. the noise is stationary on a long-term basis), no matter the frequency we are looking at. This explains the very low MAD from $T_w = 60$ min.

Based on this analysis we chose $T_w = 1$ min as the optimal time window in order to separate transients at high frequency (Fig. 3c) since we aim at mitigating source effects impacting the signal at high frequency. This time window also ensures a suitable frequency resolution of the FAS.

The second step of our proposed workflow is to further identify families of FAS reflecting different behaviors to isolate time segments cleaned from strong transients. We used the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise, e.g. Deng 2020), for this purpose. The underlying idea is that FAS associated with strong transients could form specific clusters of high amplitude that can then be easily removed. Based on this idea, we discarded the cluster tagged as -1 in the classification result of each sensor, which corresponds to non-consistent data according to DBSCAN, and retained the cluster for which the median FAS is the lowest. More information about DBSCAN algorithm can be found in appendix section. Fig. 3(d) shows all the FAS computed using 1-min long segments (in green) and the FAS forming the retained cluster (in grey) for site N7. Retained FAS represent $\sim 9.6\%$ of all the FAS. By comparing to Fig. 3(b), we can assert that the resulting selection is able to put aside most of the transient signals of high amplitude, most probably due to traffic along the national road. This translates into a lower resulting median FAS (dashed red line). Computing SSR_n and SSR_h on the basis of this selection should then help reducing the impact of local sources, thereby reflecting better the actual amplification associated with the underlying geology.

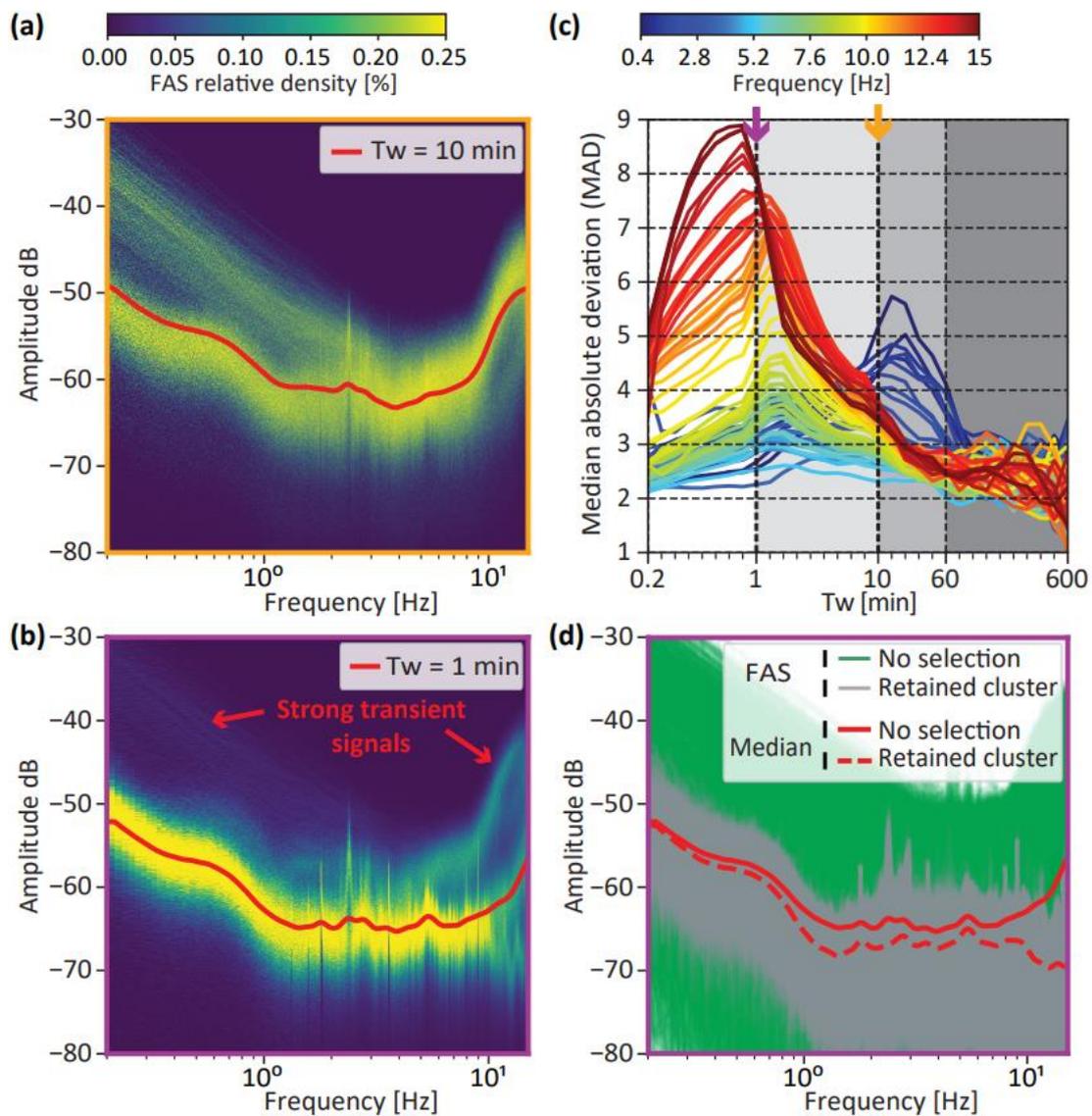


Figure 3. (a) and (b) FAS relative density for respectively 10 min and 1 min long windows (T_w). The median FAS is in solid red line. (c) Median absolute deviation (MAD) as a function of T_w and frequency. Purple and orange arrows indicate respectively $T_w = 1$ min and $T_w = 10$ min. Note that the horizontal axis is not linear and reflects different time scales highlighted by shaded areas. (d) Comparison between all the FAS (in green) and the FAS forming the cluster retained by DBSCAN (in grey) for $T_w = 1$ min. The median over all the FAS and over the retained FAS are respectively in solid and dashed red line. All results in Fig. 3 are derived for the sensor located at site N7 (Fig. 1b) for the E-W component. The color of the frame in Figs 3(a), (b) and (d) stands for the T_w used in the respective panels (i.e. orange: $T_w = 10$ min; purple: $T_w = 1$ min).

3.2.6 Spectral ratios computation & results

Spectral ratios computation

In the following sections, we present the results of SSRn and SSRh with or without considering our workflow in the computations.

First, the section 5.2 is dedicated to the results obtained without considering our workflow. We present in this section a comparison of SSRn and SSRh with classical earthquake-based SSR (Fig. 4), as well as an analysis of SSRh amplification factors across the entire array (Fig. 5). No pre-processing in the noise-based spectral ratios except a selection of nighttime data segments to remove major, and obvious anthropogenic sources (like trains passing, Figs. 2a and b) was applied. We recall that 10-min long segments are chosen to compute initial noise-based ratios. We thus computed the ratios between 10-min long windows FAS obtained at each sensor and the synchronous time windows FAS at rock-reference site ADHE between 0.2 and 15 Hz for the three components of motion. To attenuate the impact of outliers, we defined our SSRn at a sensor sx as the median over these ratios and smoothed the median using Konno-Ohmachi smoothing (Konno & Ohmachi 1998), taking b parameter equal 40 (eq. 3). Note that we do not smooth the FAS, as it may be commonly done, given the amount of data and the related difficulty due to the duration of such processing. However, some tests have been performed on a limited set of data revealing comparable results for the two processing. We used the Konno-Ohmachi smoothing because of its ability of achieving a uniform smoothing over the whole frequency band through an adaptative window size (eq. A1 in appendix section). Hence, one can avoid over-smoothing low frequencies. 40 is commonly chosen as the value of parameter b : this has been found to be a good compromise to efficiently smooth the spectra without over-smoothing. We benefit from SSR computed by Gélis *et al.* (2022) at sites BOLL and PAUL relative to ADHE (following eq. 4) for comparison. We invite the reader to refer to Gélis *et al.* (2022) for details regarding the SSR computation. Using co-located nodes at BOLL and PAUL (see section 2.2), we computed the SSRh at each sensor sx , either considering BOLL or PAUL as a local site-reference (eq. 5). As the node sensor at site PAUL is considered unusable for the N-S component (see quality control in Froment *et al.* 2022a), we did not estimate the SSRn and the SSRh using this station for this component. The following are the mathematical expressions of our SSRn, SSR and SSRh computation:

$$SSRn_{sx/ADHE} = \left\{ \left\langle \frac{FASn_{sx,Tw(i)}}{FASn_{ADHE,Tw(i)}} \right\rangle \right\}, \quad (3)$$

$$SSR_{s/ADHE} = \left\{ \left| \frac{FAS_{s,k}}{FAS_{ADHE,k}} \right| \right\}, \quad (4)$$

$$SSRh_{sx/ADHE} = \left\{ \left| \frac{FAS_{s,k}}{FAS_{ADHE,k}} \right| \right\} \times \left\{ \left\langle \frac{FASn_{sx,Tw(i)}}{FASn_{s,Tw(i)}} \right\rangle \right\}, \quad (5)$$

where sx is a sensor of the array, s is either BOLL or PAUL station, $Tw(i)$ is the time window at time i , k is the considered earthquake, $\langle . \rangle$, $|. |$ and $\{. \}$ are respectively the median over the noise-based FAS ratios, the geometric mean over the SSR and the Konno-Ohmachi smoothing.

Second, sections 5.3 and 5.4 are dedicated to the results obtained following our workflow. In these sections, we evaluate its impact by comparing SSRn to our so-called DBSCAN_SSRn (Fig. 6) and SSRh to our so-called DBSCAN_SSRh (Fig. 7). In these figures, SSRn and SSRh are estimated considering all synchronous 1-min long windows FAS between each sensor and the reference station. DBSCAN_SSRn and DBSCAN_SSRh are estimated considering only the synchronous FAS between the retained cluster at each sensor and the reference. Still only nighttime segments are used in the computations.

Results of SSRn and SSRh without considering our workflow

In Fig. 4(a), we compare $SSR_{PAUL/ADHE}$, $SSRn_{PAUL/ADHE}$ and $SSRh_{PAUL/ADHE}$ using BOLL as local site-reference station for the E-W component. Below 1 Hz, a good agreement is found between $SSRn_{PAUL/ADHE}$, $SSRh_{PAUL/ADHE}$ and $SSR_{PAUL/ADHE}$. Moreover, as expected from the spatial variability of low frequency noise amplitude discussed in section 2.2 (Fig. 1b), $SSRn$ value at 0.5 Hz across the area (Fig. 4b) is well correlated with the underlying geology, showing higher amplification on top of the sedimentary filling where the valley is expected the deepest (Fig. 1b). This confirms that, below 1 Hz, noise-based spectral ratios seem to be able to capture site effects. Above 1 Hz, $SSRn_{PAUL/ADHE}$ strongly deviates from $SSR_{PAUL/ADHE}$ (Fig. 4a), reaching values that are unlikely to be correct. As 1 Hz is usually described as a threshold in the origin of noise (i.e. between ambient noise generated by natural sources at low frequency and noise coming from human activities at high frequency), the observations above 1 Hz suggest the impact of local anthropogenic sources close to PAUL site on $SSRn_{PAUL/ADHE}$ (e.g. A7 highway and the TNS). $SSRh_{PAUL/ADHE}$ presents a better agreement to $SSR_{PAUL/ADHE}$, by giving close estimate of amplification until around 2 Hz and in between 3 and 4 Hz. For the rest of the frequency band, $SSRh_{PAUL/ADHE}$ still tends to give higher estimates than $SSR_{PAUL/ADHE}$, but it is able to retrieve its global shape with some peaks at specific frequencies. Note that we represent the amplification using a linear scale and not a logarithmic scale (that

may be more commonly used for SSR representation) in order to unambiguously compare the different curves that cover a limited range of amplification. The difference between $SSR_{hPAUL/ADHE}$ and $SSR_{PAUL/ADHE}$ seems to be related to the intermediate $SSR_{nPAUL/BOLL}$ used in the SSR_h computation (dashed green line). This site-to-site ratio is indeed significantly above 1 for the frequencies where $SSR_{hPAUL/ADHE}$ is overestimating $SSR_{PAUL/ADHE}$, reaching a factor of 2 at 6 Hz. Hence, we reported SSR_h value at this frequency across the array, either considering BOLL (Fig. 5a) or PAUL (Fig. 5b) as local site-reference station. SSR_h estimation is supposed to give similar results whatever the intermediate site-reference used. Yet, results show lower SSR_h values across the array when PAUL is used as site-reference. This is likely to be due to the difference in the source signature experienced at 6 Hz between PAUL, that is close to the A7 highway and the TNS, and BOLL that is further away from these strong anthropogenic sources. This result reveals that considering a local intermediate site-reference is not sufficient to fully cancel source effects at high frequency. This illustrates the limitations of the application of SSR_h in our highly industrialized area using simple criteria to select noise windows (10-min long windows, nighttime recordings).

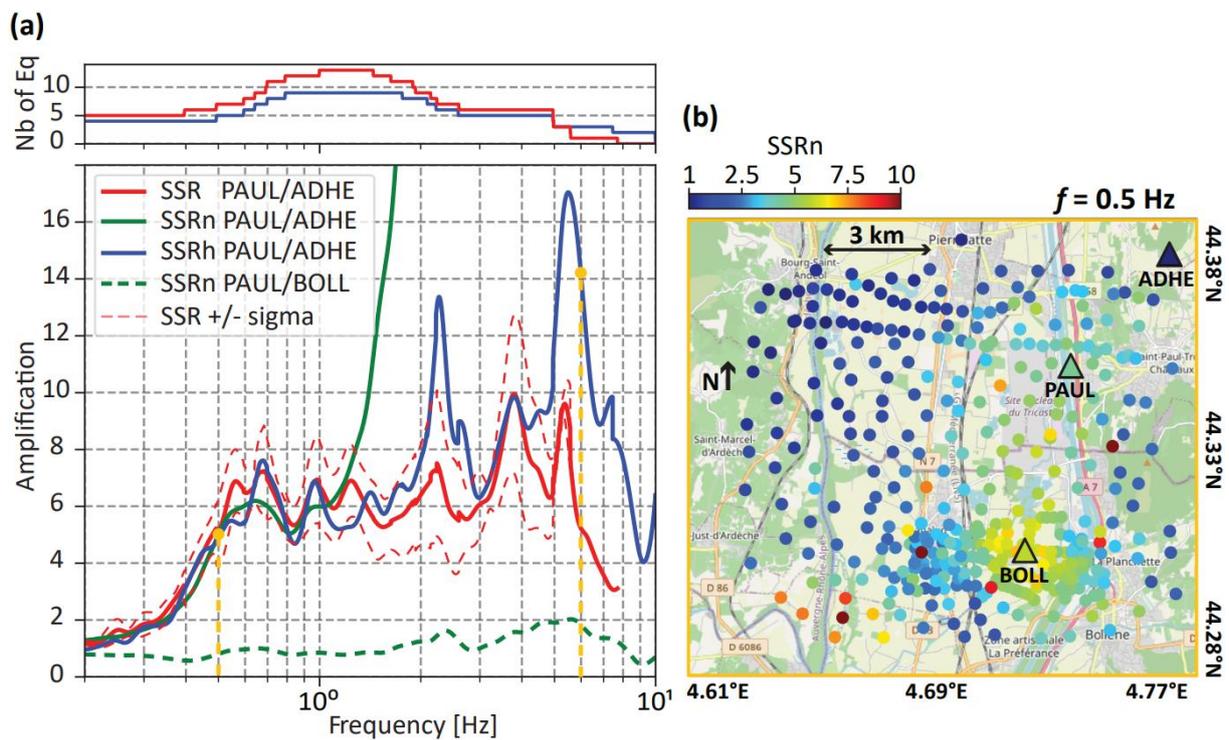


Figure 4. (a) Comparison between $SSR_{PAUL/ADHE}$, $SSR_{nPAUL/ADHE}$ and $SSR_{hPAUL/ADHE}$ with BOLL as local site-reference station, respectively represented by the solid red, green and blue curves. The site-to-site $SSR_{nPAUL/BOLL}$ is shown in dashed green line. Plus and minus one standard deviation to SSR are plotted in dashed red lines. The number of earthquakes used to compute $SSR_{PAUL/ADHE}$ and $SSR_{hPAUL/ADHE}$ is represented on the top

panel respectively in thin solid red and blue line. Orange dots indicate $SSRn_{PAUL/ADHE}$ value at 0.5 Hz and $SSRh_{PAUL/ADHE}$ value at 6 Hz. (b) Map of $SSRn$ value at 0.5 Hz computed at each sensor of the array. All results in this figure correspond to the E-W component.

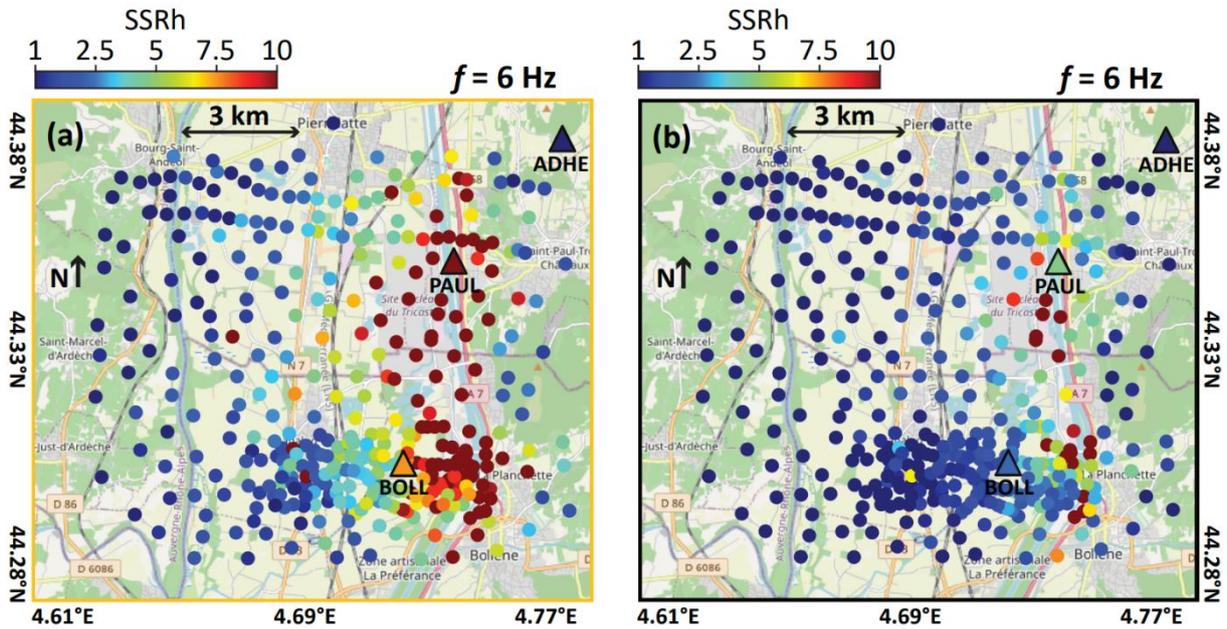


Figure 5. Maps of $SSRh$ value at 6 Hz computed at each sensor respectively using BOLL (a) and PAUL (b) as local site-reference station. Given eq. 2, note that the $SSRh$ value at BOLL in (a) and at PAUL in (b) corresponds to the SSR value. All results in this figure correspond to the E-W component.

Impact of our workflow on the resulting $SSRn$ at N7

A comparison between $DBSCAN_SSRn$ and $SSRn$ at site N7 is shown in Fig. 6 as grey and green lines respectively. As expected, $DBSCAN_SSRn$ exhibits lower amplification factors than $SSRn$ over the whole frequency range for all components. The discrepancy between $DBSCAN_SSRn$ and $SSRn$ is generally increasing with the frequency, which highlights the importance of a careful data selection for high frequencies. It is worth noting that this observation holds for all components.

Fig. 6 shows that computing $SSRn$ following our workflow ends up attenuating unrealistically high values. However, the resulting amplification factors remain very high, especially for the horizontal components. Amplification reaching values of 20 to 30 at frequencies of a few Hertz are much higher than values observed at PAUL using earthquake-based SSR (see red curve in Fig. 4a). Even if the very local geology may influence the amplification within the valley between PAUL and N7 sites, N7 is not

located on top of the axis of the canyon (by contrast to PAUL) and is not expected to be experiencing much larger amplification (i.e. as the ones estimated through DBSCAN_SSRn) at these frequencies. This suggests that DBSCAN_SSRn still presents some limitations to provide a reasonable estimate of the local amplification, at least at this site.

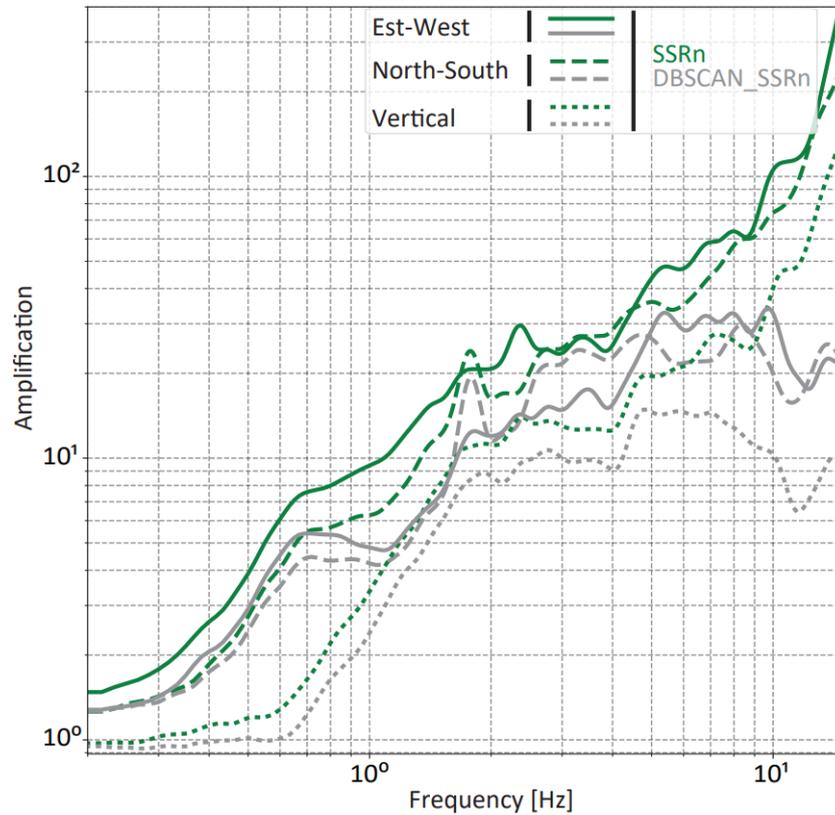


Figure 6. Comparison between SSRn (in green) and DBSCAN_SSRn (in grey) at site N7 for the E-W (solid lines), N-S (dashed lines) and vertical (dotted lines) components. The logarithmic scale has been used here to visualize properly the amplification that covers several orders of magnitude.

Impact of our workflow on the resulting SSRh over the entire array

Figs 7(a) and (b) show respectively DBSCAN_SSRh results and the relative difference $D_{f,sx}$ between DBSCAN_SSRh and SSRh at 6 Hz across the array for the E-W component. In addition to the broadband vision at a specific sensor given in Fig. 6, Figs 7(a) and (b) allow us to quantify the impact of our workflow on SSRh amplification factors across the array at a given frequency. BOLL and ADHE (Fig. 1) are the site-reference and rock-reference stations respectively. $D_{f,sx}$ at a sensor sx and frequency f is computed as follows:

$$D_{f,sx} = \frac{SSRh_{f,sx} - DBSCAN_SSRh_{f,sx}}{SSRh_{f,sx}} \times 100 . \quad (6)$$

Fig. 7(b) shows that a positive difference, which means that DBSCAN_SSRh is lower than SSRh, is seen nearby significant roads on the Western part of the array. It is particularly visible along the N7 national road (orange line crossing the center of the array from north to south) where most of the sensors exhibit a positive difference of more than 25%. These observations are consistent with the capacity of our workflow to attenuate unrealistically high values in the spectral ratio by filtering the data from strong transient noise produced by the road traffic on this busy local road. Nevertheless, on the Eastern part of the array, almost no difference is found between DBSCAN_SSRh and SSRh. This means that the first step of our workflow fails to isolate transient signals among the collection of 1-min segments, suggesting that the main sources in this area, like the TNS, the hydroelectric dam and the A7 highway rather act as stationary sources where no transient signals can be isolated on a 1-min basis.

To extend this analysis on the proposed workflow impact to other frequencies, we represent $D_{f,sx}$ in Fig. 7(c) as a function of distance from BOLL site for sensors located inside an azimuth of $\pm 20^\circ$ east and west of BOLL, for $f = 0.5$ Hz, 3 Hz, 6 Hz and 9 Hz. Almost no relative difference is seen for the sensors located east of BOLL, regardless of the distance and the frequency considered. On the other hand, as the frequency is increasing, more sensors exhibit high positive $D_{f,sx}$ when located west of BOLL, especially around the N7 national road location. This dependence to higher frequencies simply reflects the spectral characteristics of the signal produced by cars and trucks traffic (typically a few tens of hertz). These series of plots confirm that the proposed workflow performs well at all considered frequencies for sensors nearby the N7 national road, but does not help reducing the influence of the hydroelectric dam whatever the frequency. The contrast in the processing response between east and west of BOLL station emphasized in Fig. 7 reveals the difficulty of having a procedure that is efficient everywhere inside such a complex anthropized environment, where different types of sources cohabit.

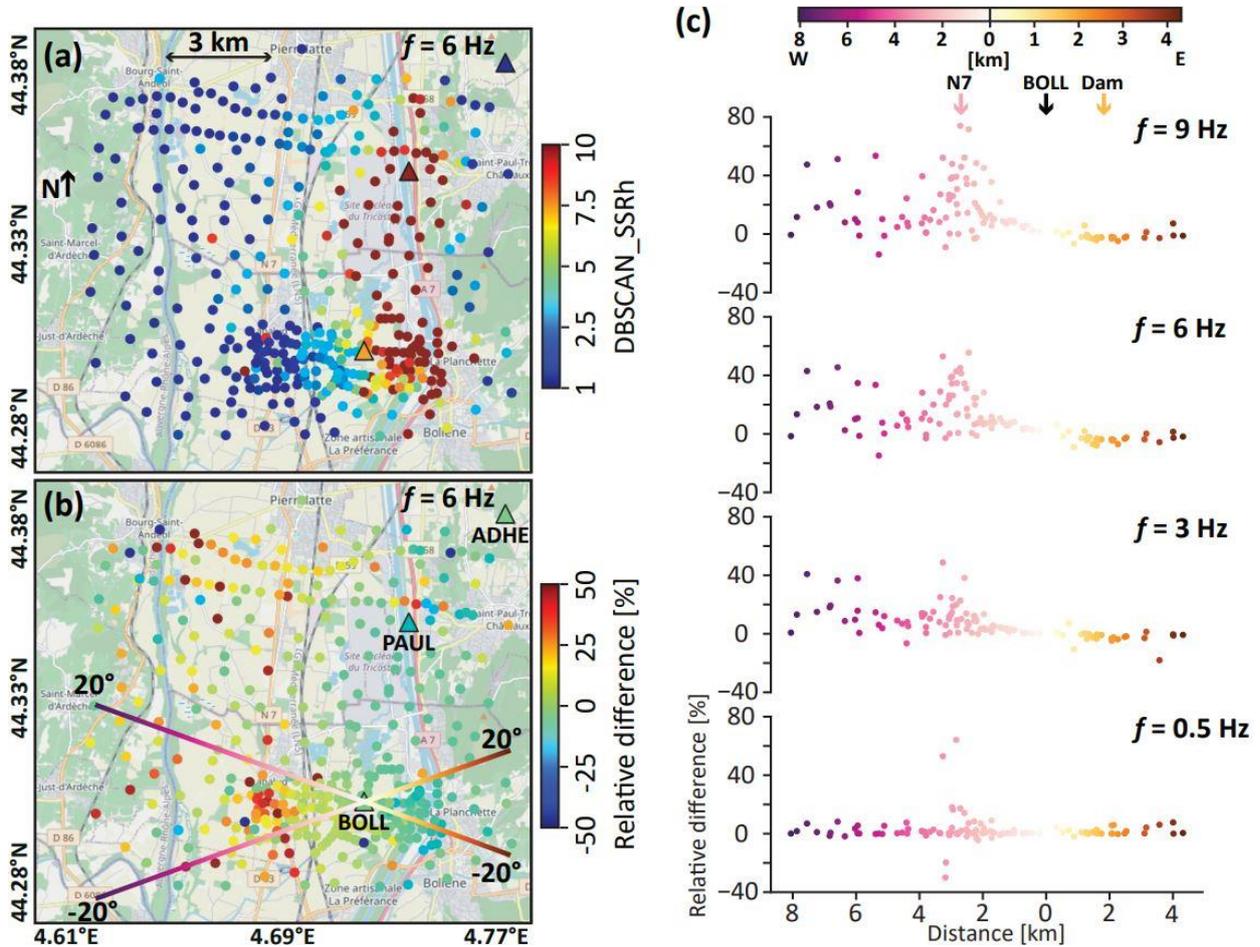


Figure 7. DBSCAN_SSRh results (a) and the relative difference (b) between DBSCAN_SSRh and SSRh at 6 Hz at each sensor. (c) Relative difference value as a function of distance from BOLL site, for sensors located inside an azimuth of $\pm 20^\circ$ east and west of BOLL, for $f = 0.5$ Hz, 3 Hz, 6 Hz and 9 Hz. BOLL and ADHE (Fig. 1) are respectively the site-reference and rock-reference stations considered for the SSRh computation. All results in Fig. 7 are derived for the E-W component.

3.2.7 Discussion & conclusion

In this paper, we investigated the applicability of SSRn and SSRh in a heavily industrialized environment in the Tricastin area (French Rhône Valley, South-East France). In such anthropized conditions, comparing local ground motion without removing source effects might end up highlighting abnormally high and localized patterns of amplification which are not related to the underlying geological structure (Fig. 2). In order to tackle this problem and mitigate these source effects, we introduced a two-step workflow to remove strong impulsive signals (Fig. 3). These so-called transient signals are indeed likely to be generated by local disturbances like traffic (cars, trucks, trains, etc.), or any natural or human related sources. After selecting only nighttime data segments, a first step is to define a characteristic time window to optimally isolate significant transients in the continuous time series. This time window must ensure a suitable frequency resolution of the FAS. We here chose a time window of 1 min. The second step consists in selecting the time segments that do not contain these significant transients based on a clustering of the FAS according to their spectral characteristics.

For the purpose of testing SSRn and SSRh applicability in our case, we computed a first estimation across the array without applying our workflow or any other pre-processing than a selection of nighttime data segments (Figs. 4 and 5). Below 1 Hz, SSRn amplification factor seems to be consistent with the expected geology of the site and able to capture the site effects. This is quantitatively confirmed by a comparison with earthquake-based SSR amplification factor. However, as expected, SSRn strongly deviates from SSR above 1 Hz. This is likely to be due to anthropogenic sources of noise in the vicinity of some sensors, mainly within the valley. SSRh, which involves a local intermediate reference, gives closer estimate of amplification to SSR above 1 Hz. However, its amplification factor still shows some differences with SSR curve that we attribute to the impact of local sources. SSRh limitations at high frequency are emphasized by the fact that taking two different site-reference stations leads to differences in amplification estimates. It is worth noting that the differences revealed must be discussed depending on the applications of these estimates and the targeted degree of precision. However, these features need to be examined since they contribute to understand what controls noise-based spectral ratios, thereby gaining insight into their applicability as tools for estimating local site amplification. The results show that the amplification provided by the SSRh approach may be affected by local noise sources, especially at high frequencies and in highly industrialized areas. This can

be interpreted as a need of having a fairly well equipartitioned wavefield within the site under consideration; this obviously does not hold in our case.

Figs. 4 and 5 illustrate the motivation for applying careful pre-processing steps and data selection before computing SSRn and SSRh to address the problem of local source effects in the resulting amplification factors. Hence, we estimated SSRn and SSRh following our two-step procedure and evaluated the impact of our proposed workflow on these source effects. Comparison between SSRn computed with or without our data selection at site N7 (Fig. 6) shows that computing SSRn following our workflow ends up attenuating unrealistically high values. However, the resulting amplification factors remain abnormally high as the frequency is increasing, especially for the horizontal components. We quantitatively evaluated the impact of our approach on SSRh across the array for different frequencies (Fig. 7). We concluded that the proposed workflow is indeed effective to attenuate the influence of some transient sources (like some busy roads), rather located on the Western part of the array, but has almost no effect on stationary sources (i.e. the nuclear site, hydroelectric dam, A7 highway), mainly located on the Eastern part. It is worth noting that the traffic may be almost continuous on the highway, by contrast to the N7 national road for which we can find some quiet time windows. Our results reveal the difficulty of having a simple and automatic procedure that is efficient everywhere inside such a complex anthropized environment, where different types of sources cohabit, and in particular stationary sources.

Our procedure focused on mitigating source effects. It is directly related to the spectral content of the noise wavefield and therefore consistent with the purpose of spectral ratios computation. The time window is chosen by analyzing the data in order to discard transient signals. Another key point is that it is a fully automatic procedure, which is decisive for the processing of dense array measurements. As an area for improvement though, we think that this approach should be tested to other sites with an adaptation of the characteristic time window and clustering parameters depending on the sensor location and a careful analysis of the main local sources, for the method to be fully data-adaptative.

Another idea would be to model the remaining stationary sources to estimate and correct their impact. Based on these models, one could then be able to mitigate the remaining differences between noise-based spectral ratios and SSR in order to bring to the forefront the amplification associated with the site's geology. However, this would rely on strong assumptions (e.g. on the source radiation, attenuation within

the propagation medium) whose related uncertainties could not be easily quantified in the resulting estimated amplification.

Finally, one could also include the phase information when computing the spectral ratio, which corresponds to a full noise-based deconvolution (e.g. Prieto *et al.* 2011). Similar to noise correlation functions (NCFs), deconvolutions computed between different station pairs allow to spatially filter the ambient wavefield and mitigate remaining local sources to retrieve the relative amplitude through the array. Although not straightforward as it is strongly dependent on the spatial distribution of noise sources (e.g. Tsai 2011; Stehly & Boué 2017), some studies demonstrated that amplitude of NCFs can be exploited, with specific processing, to map local amplification and 3D propagation effects associated with sedimentary basins (e.g. Prieto *et al.* 2011; Denolle *et al.* 2014; Viens *et al.* 2016; Boué *et al.* 2016).

3.2.8 Acknowledgments

We gratefully acknowledge the editor and assistant editor, as well as Matsushima Shinichi and an anonymous reviewer for their useful comments and critics. This work was funded by a public grant overseen by the French National Research Agency (Grant # ANR19-CE31-0029). We would like to thank the French and German colleagues of ISTerre and GFZ Potsdam as well as the University of Potsdam for their collaboration. Special thanks must go to François Lavoué and Vincent Perron for interesting discussions on this work as well as Nanaba Bagayoko, Edward Marc Cushing and Damien Do Couto for their important work on the geology of the Tricastin Valley. We also would like to cite colleagues who helped installing the nodes array (Flomin Tchawe Nziha, Christophe Clément, Charles Beard, Sophie Beauprêtre, Gilbert Bouzat, Florent Brenguier, Edward Marc Cushing, Aurore Laurendeau, Olivier Magnin, Laurent Stehly, Alexandre Tourette) and thank the private land owners and communities that have hosted this array.

3.2.9 Data availability

Data of the nodes array (Froment *et al.* 2023, doi:10.15778/RESIF.XG2020) will be available on the French RESIF datacenter (<https://www.resif.fr/en/>) at the end of the DARE project (end of 2023).

3.2.10 Appendix

A. Konno-Ohmachi smoothing

The following equation is the Konno-Ohmachi formula as introduced in Konno & Ohmachi (1998).

$$W(f, f_c) = \left[\frac{\sin\left(\log_{10}\left(\frac{f}{f_c}\right)^b\right)}{\log_{10}\left(\frac{f}{f_c}\right)^b} \right]^4, \quad (\text{A1})$$

where f and f_c are the frequency and a center frequency respectively and b a bandwidth coefficient.

B. DBSCAN clustering algorithm

DBSCAN is a density-based data clustering algorithm. A key point of DBSCAN for big data applications is that it is an unsupervised clustering method. Therefore, we don't need a prior information to set an initial number of clusters. Moreover, DBSCAN is able to classify data depending on their density, which suits perfectly our need since FAS relative density is changing when isolating transient signals. The only two parameters for this clustering procedure are chosen empirically after manual inspection for some examples. One parameter is the maximum Euclidian distance between the sample point and the neighborhood boundary. We set this parameter to 150. The other is the density threshold, which corresponds to the minimal number of sample points in the neighborhood to form a cluster. We set this parameter equal to the total number of spectra divided by 60.

Chapitre 4 : Analyse préliminaire des corrélations de bruit sismique entre une source virtuelle et l'ensemble du réseau

4.1 Introduction

Le chapitre 3 a illustré (dans notre cas) la difficulté d'anticiper l'amplification au site en cas de séisme au-delà de 1 Hz à travers un simple rapport d'amplitude. Les sources locales de bruit sismique brisent l'hypothèse selon laquelle un simple rapport d'amplitude isolerait totalement la réponse du site. Pour isoler la réponse du milieu, une possibilité serait de modéliser ces sources pour corriger les signaux de leurs signatures (e.g. Lavoué *et al.* 2020). Cela demande néanmoins une connaissance précise des paramètres des sources. Une autre approche consiste à se servir de l'information de phase contenue dans le champ de bruit sismique : c'est l'interférométrie par corrélation de bruit.

4.1.1 Principe de l'interférométrie par corrélation de bruit

Pour introduire l'interférométrie, rappelons qu'un signal sismique émis par une source et enregistré à un capteur porte (au premier ordre) l'information de la source (énergie, mécanisme, distance, directivité, durée...) et du trajet des ondes sur le chemin source-capteur, c'est-à-dire la fonction de Green. La fonction de Green est la réponse du milieu à une source impulsionnelle en temps et en espace (mathématiquement définie par un Dirac $\delta(x)\delta(t)$). C'est donc ce que l'on cherche à connaître. Dans le cas du bruit sismique, le signal enregistré à un capteur est issu d'une multitude de sources et de trajets source-capteur. L'idée de l'interférométrie par corrélation de bruit est d'utiliser cette complexité pour reconstruire la fonction de Green entre 2 capteurs. Dans ce schéma-là, l'un des 2 capteurs joue le rôle de source impulsionnelle, l'autre de récepteur. La fonction de Green ainsi reconstruite contient toute l'information, quelle

que soit la fréquence, sur la propagation des ondes entre la source virtuelle et le récepteur.

Prenons par exemple 2 capteurs distants A et B qui ont enregistré le bruit sismique sur une plage de temps synchrones (Figure 4.1). Le signal u_A enregistré au capteur A et u_B enregistré au capteur B sont la combinaison d'une multitude de sources S_i (étoiles jaunes) et de fonctions de Green décrivant la propagation entre ces sources et les 2 capteurs. Parmi ces sources, certaines contribuent à la fois au signal enregistré en A et au signal enregistré en B. Plus il y a de sources et de trajets, plus l'information commune entre A et B est importante. L'opération de corrélation étant une analyse de la similarité entre deux points de mesure, l'information commune entre A et B est extraite par corrélation des enregistrements ($C_{AB}(\tau)$). On reconstruit ainsi un signal déterministe correspondant à la propagation entre A et B, c'est-à-dire à la fonction de Green $G_{AB}(\tau)$, à partir d'un champ de bruit complexe. L'un des 2 capteurs joue alors le rôle de source virtuelle, l'autre de récepteur. La partie causale $G_{AB}^+(\tau)$ de la fonction de Green décrit la propagation de A vers B (A est la source virtuelle) et la partie anti-causale $G_{AB}^-(-\tau)$ décrit la propagation de B vers A (B est la source virtuelle). Si la distribution des sources est homogène, les deux parties sont symétriques car autant d'énergie se propage de A vers B que de B vers A.

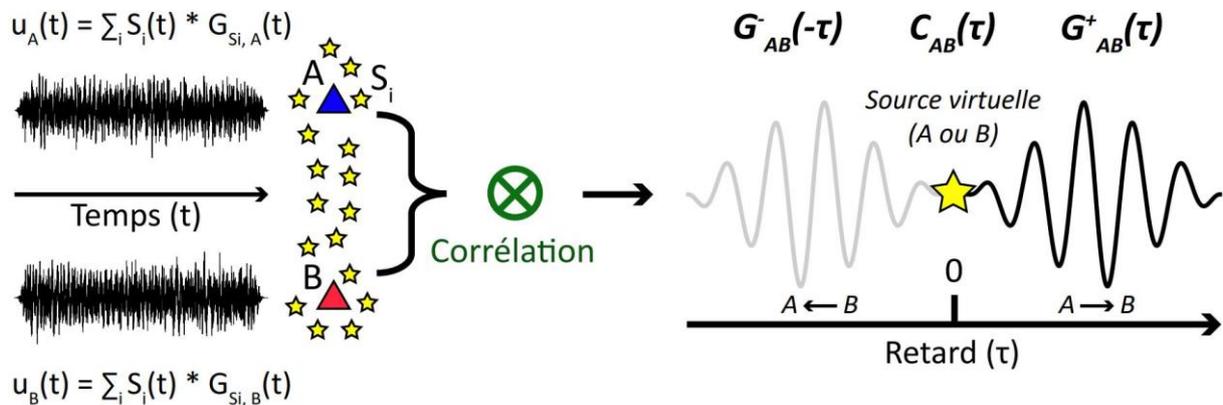


Figure 4.1 : Illustration du principe de corrélation de bruit sismique. Le signal u_A enregistré au capteur A et u_B enregistré au capteur B sont la combinaison d'une multitude de sources S_i (étoiles jaunes) et de fonctions de Green décrivant la propagation entre ces sources et les 2 capteurs. La corrélation des enregistrements permet de reconstruire la fonction de Green entre A et B. La partie causale $G_{AB}^+(\tau)$ de la fonction de Green reconstruite décrit la propagation de A vers B (A est la source virtuelle) et la partie anti-causale $G_{AB}^-(-\tau)$ décrit la propagation de B vers A (B est la source virtuelle).

4.1.2 Configurations permettant la convergence de la fonction de corrélation vers la fonction de Green

Le lien entre la fonction de corrélation et la fonction de Green a été étudié sous différents aspects. Dans le domaine de l'ultra-sonore, Lobkis et Weaver (2001) ont démontré la convergence de la fonction de corrélation vers la fonction de Green dans le cas d'un champ diffus obtenu par réverbérations multiples sur les parois d'un volume d'aluminium. Derode *et al.* (2003) font quant à eux le parallèle entre expérience de retournement temporel (e.g. Fink 1992) et corrélation et démontrent la reconstruction de la fonction de Green entre 2 récepteurs dans le cas où une distribution de sources spatialement décorréliées entoure de manière continue ces 2 récepteurs. Les hétérogénéités du milieu agissant comme sources secondaires (principe de Huygens-Fresnel) peuvent cependant pallier une distribution imparfaite. Snieder (2004) puis Roux *et al.* (2005) établissent une relation de proportionnalité en se servant du théorème de phase stationnaire. La dérivée temporelle de la fonction de corrélation C_{AB} entre 2 récepteurs A et B est proportionnelle à la fonction de Green du milieu entre A et B :

$$\frac{\partial C_{AB}(t)}{\partial t} \propto G_{AB}^+(t) - G_{AB}^-(-t). \quad (1)$$

Dans cette démonstration, il n'est pas nécessaire d'avoir un continuum de sources tout autour des 2 récepteurs. Seules les sources dans l'alignement de la paire (zones de Fresnel) contribuent à la reconstruction de la fonction de Green, les autres interfèrent de manière destructive. Notons au passage que l'équation 1 fait apparaître la partie causale $G_{AB}^+(t)$ et anti-causale $G_{AB}^-(-t)$ de la fonction de Green décrivant respectivement le trajet de A vers B (où A est la source virtuelle) et de B vers A (où B est la source virtuelle).

En définitive, deux grandes configurations permettent la convergence de la fonction de corrélation vers la fonction de Green. L'une d'entre elles repose sur la condition d'un champ diffus. En sismologie, on parle de champ diffus lorsque l'énergie totale du champ d'ondes est distribuée uniformément entre les différents modes élastiques qui le composent. Lorsque toutes les ondes et tous les trajets sont représentés, le champ atteint un régime d'équipartition (e.g. Hennino *et al.* 2001). Ce type de champ est obtenu en présence d'une distribution homogène de sources ou après que les ondes aient été multiples fois diffractées par les hétérogénéités du milieu. Un champ diffus a priori chaotique conserve malgré tout une certaine cohérence. Comme décrit précédemment, l'opération de corrélation permet d'extraire

cette cohérence entre deux points de mesure. On reconstruit ainsi un signal déterministe correspondant à la propagation entre ces deux points (fonction de Green) à partir d'un champ diffus. L'autre configuration implique une distribution adéquate des sources par rapport aux points de mesure (autour de la paire de récepteurs ou dans les zones de Fresnel). La fonction de Green émerge de la moyenne sur les sources des fonctions de corrélation.

4.1.3 La distribution des sources de bruit sismique comme principale limitation de l'interférométrie par corrélation de bruit

Concernant le champ de bruit sismique, les deux configurations présentées ne sont jamais parfaitement réalisées et la fonction de Green n'est que partiellement reconstruite dans la réalité. Il est néanmoins possible d'améliorer la qualité de la convergence des fonctions de corrélation en considérant des enregistrements suffisamment longs. La distribution des sources a tendance à s'homogénéiser et les ondes sont de plus en plus diffractées avec le temps.

De nombreuses études montrent cependant l'impact d'une distribution anisotrope de sources de bruit (e.g. Stehly *et al.* 2006 ; Van der Neut et Bakulin 2009 ; Tsai 2009 ; Froment *et al.* 2010). Cette anisotropie engendre des biais dans la fonction de corrélation. En d'autres termes, les corrélations ne convergent pas vers la fonction de Green du milieu et présentent des différences (biais) tant en termes d'amplitude que de temps d'arrivée (information liée à la phase). Reprenons le cas de 2 capteurs distants A et B illustré précédemment, mais pour une distribution de sources non-homogène, majoritairement situées à proximité de A. Premièrement, l'anisotropie dans la distribution des sources cause une dissymétrie dans l'amplitude des parties causale et anti-causale de la fonction de Green reconstruite car plus d'énergie se propage de A vers B que de B vers A. Deuxièmement, les sources n'étant pas situées tout autour de la paire de capteurs ou dans les zones de Fresnel (zones grises), les temps d'arrivée des ondes reconstruites sont biaisés et peuvent mener à une surestimation des vitesses apparentes (ici particulièrement visible concernant la partie causale de la fonction de Green).

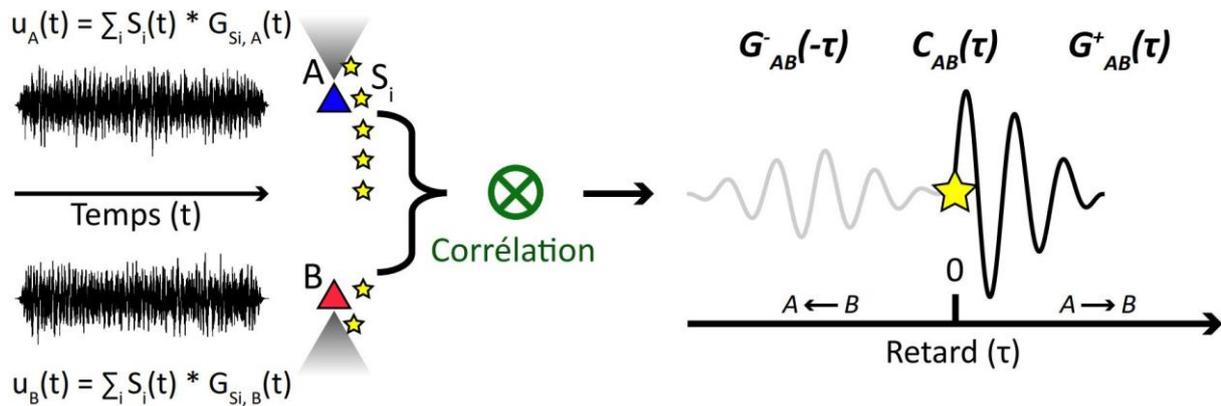


Figure 4.2 : Idem que pour la figure 4.1 mais dans le cas d'une distribution de sources anisotrope. L'anisotropie cause une dissymétrie dans l'amplitude des parties causale et anti-causale de la fonction de Green reconstruite et introduit un biais important dans l'estimation des temps d'arrivée.

Plusieurs méthodes ont été introduites pour améliorer la convergence des corrélations vers la fonction de Green. La plupart de ces méthodes priorisent une bonne convergence (et stabilité) de l'information de phase des signaux, au détriment de l'amplitude. Ainsi, des mesures robustes de temps d'arrivée pour la tomographie et/ou le suivi temporel peuvent être réalisées (e.g. Bensen *et al.* 2007). En pratique, la plupart des méthodes traitent les traces sismiques (c.à.d. le bruit) de manière individuelle avant corrélation, comme la normalisation temporelle 1-Bit (e.g. Larose *et al.* 2004), ou certaines méthodes d'écrêtage (e.g. Sabra *et al.* 2005). D'autres méthodes s'appliquent sur les fonctions de corrélation elles-mêmes, comme par exemple la corrélation de la coda des corrélations qui vise à exploiter au mieux la part diffuse du champ initial (e.g. Stehly *et al.* 2008 ; Froment *et al.* 2010). Récemment, plusieurs techniques de filtrage en réseau ont également vu le jour pour moduler directement l'amplitude du champ en fonction de la direction d'incidence, au prix d'un lissage spatial (e.g. Gallot *et al.* 2012). L'une des méthodes les plus utilisées, au moins dans le cadre des mesures de vitesse, reste le blanchiment des traces sismiques avant corrélation, couplé ou non avec une normalisation temporelle. Cette technique, dénommée pre-whitening en anglais, consiste à égaliser le spectre d'amplitude à toutes les fréquences d'intérêts. La majorité de ces méthodes contribuent à améliorer la robustesse des mesures de déphasage en jouant sur la distribution effective des sources, ou plus exactement sur les propriétés du champ incident. Elles ont donc été particulièrement utilisées dans le domaine de la tomographie sismique car elles n'affectent pas, ou peu, l'information de phase. Ainsi, des mesures de vitesse peuvent être réalisées avec un minimum de biais à partir des temps d'arrivée pour imager la

structure aux différentes échelles (e.g. Shapiro *et al.* 2005 ; Stehly *et al.* 2009 ; Ritzwoller *et al.* 2011).

4.1.4 Sur l'utilisation de l'amplitude des fonctions de corrélation

Si les méthodes auxquelles nous venons de faire référence n'affectent pas la phase des fonctions de corrélation, elles en altèrent en général l'amplitude. Par conséquent, elles ne peuvent être utilisées dans le but d'étudier l'amplitude des corrélations. Certaines études relativement récentes se sont malgré tout penchées sur le sujet de l'utilisation de l'amplitude des corrélations, soit pour étudier l'atténuation du milieu de propagation (e.g. Prieto *et al.* 2011 ; Stehly et Boué 2017), soit pour étudier la réponse des bassins sédimentaires (e.g. Prieto et Beroza 2008 ; Prieto *et al.* 2011 ; Denolle *et al.* 2013, 2014 ; Viens *et al.* 2014, 2015, 2016 ; Perron 2017). Dans ce cas, il convient de normaliser (dans le domaine fréquentiel) la fonction de corrélation par l'autocorrélation à une station au rocher utilisée comme source virtuelle. La variation d'amplitude (amplification) entre le réseau (dans le bassin sédimentaire) et la source virtuelle (au rocher) peut ainsi être conservée. Le résultat de cette normalisation correspond mathématiquement à une déconvolution, parfois appelée ANIRF (Ambient Noise Impulse Response Function) :

$$ANIRF(f) = \left\langle \frac{U_A(f)U_B^*(f)}{\{|U_B(f)|^2\}} \right\rangle , \quad (2)$$

avec $U_A(f)$ et $U_B(f)$ la transformée de Fourier du signal aux capteurs A et B, $|U_B(f)|$ le spectre d'amplitude du signal au capteur B, $\{.\}$ une fonction de lissage et $\langle . \rangle$ la moyenne arithmétique sur les fenêtres de temps utilisées. La notation suit ici Prieto *et al.* (2011). $U_A(f)U_B^*(f)$ dénote la fonction de corrélation entre A et B où B est la source virtuelle. Le complexe conjugué est représenté par le symbole $*$.

Dans le chapitre 5, nous nous intéresserons à estimer l'amplification associée à la vallée du Tricastin à travers l'ANIRF. En amont, les fonctions de corrélation, à la base des calculs d'amplification du chapitre 5, sont ici analysées et traitées.

4.2 Corrélations de bruit sismique entre ADHE et l'ensemble du réseau

4.2.1 Calcul et analyse des corrélations brutes et blanchies

Les corrélations de bruit sismique ont été calculées entre ADHE (source virtuelle) et chaque node du réseau selon 2 cas. Dans le premier cas les 2 traces sismiques mensuelles ont été blanchies avant corrélation (corrélations blanchies) tandis que dans le second aucun blanchiment n'a été appliqué (corrélations brutes). Les étapes de calcul sont données ci-dessous :

- (1) rotation des composantes horizontales des 2 traces sismiques mensuelles (nord-est \rightarrow radiale-transverse)⁶ ;
- (2) blanchiment des 2 traces sismiques entre 0.2 et 10 Hz ou non ;
- (3) calcul des corrélations pour des segments de 30 min sans recouvrement ;
- (4) moyenne arithmétique des fonctions de corrélation ;
- (5) correction de l'atténuation géométrique⁷.

Précisons que les corrélations n'ont pas été calculées pour les composantes croisées du tenseur. En d'autres termes, seules les composantes TT (transverse-transverse), RR (radiale-radiale) et ZZ (verticale-verticale) ont été considérées.

Les figures 4.3 et 4.4 représentent les corrélations brutes et blanchies (respectivement colonne de gauche et colonne de droite) classées en fonction de la distance séparant ADHE (source virtuelle) du node associé à la paire (récepteur). Les corrélations sont filtrées entre 0.35 et 0.9 Hz sur la figure 4.3, entre 3 et 6 Hz sur la figure 4.4. Les panneaux (a), (b) et (c) correspondent respectivement aux composantes

⁶ Il est courant d'effectuer une rotation des composantes horizontales (N) et (E) aux 2 capteurs avant corrélation. Après rotation, les capteurs sont virtuellement orientés dans la direction donnée par l'azimut de la paire. Cette opération fournit la composante radiale (R) et transverse (T) du mouvement par rapport à cette direction et isole ainsi théoriquement les ondes de Rayleigh et Love respectivement sur l'une et l'autre composante. La corrélation entre les 2 capteurs peut se calculer selon 9 combinaisons de composantes qui constituent le tenseur des corrélations : (TT, RR, ZZ, TR, RT, TZ, ZT, RZ, ZR).

⁷ Le bruit sismique est principalement constitué d'ondes de surface dont l'atténuation géométrique induit une décroissance d'amplitude en $1/\sqrt{d}$, où d est la distance parcourue par l'onde. Pour corriger l'atténuation géométrique, l'amplitude des corrélations a été multipliée par $\sqrt{d_{\text{interstation}}}$, où $d_{\text{interstation}}$ est la distance entre la source virtuelle et le récepteur.

TT, RR et ZZ. Les temps positifs correspondent à la partie causale de la fonction de Green partiellement reconstruite (réponse impulsionnelle du milieu entre ADHE et le réseau), les temps négatifs à la partie anti-causale (réponse impulsionnelle du milieu entre le réseau et ADHE). Entre 0.35 et 0.9 Hz (Figure 4.3), des trains d'ondes traversant le réseau sont clairement visibles sur les corrélations brutes et blanchies. Hormis pour la composante RR, l'amplitude des trains d'ondes est plus élevée pour les temps négatifs, ce qui signifie que durant la période d'acquisition, plus d'énergie s'est propagée du réseau vers ADHE que de ADHE vers le réseau. Ceci est cohérent avec les observations du chapitre 2 qui désignent l'océan Atlantique et la mer Méditerranée comme sources principales de bruit sismique à ces fréquences. Les corrélations brutes sont dominées par un train d'ondes dont la période d'oscillation est supérieure à 2 s, ce qui correspond à la limite basse de la plage de fréquences. La vitesse associée est légèrement supérieure à 2 km/s. Le blanchiment des traces sismiques avant corrélation fait apparaître des trains d'ondes de périodes plus courtes se propageant à des vitesses inférieures (entre 0.5 et 1 km/s) responsables d'une large dispersion.

Entre 3 et 6 Hz (Figure 4.4), on peut faire deux observations majeures. Premièrement, on remarque un changement des temps d'arrivées des ondes (des temps négatifs vers les temps positifs) lorsque la distance à ADHE augmente en suivant l'amplitude des corrélations blanchies. La plupart de l'énergie transite du réseau vers ADHE pour des distances inférieures à 8 km, et d'ADHE vers le réseau au-delà. Ceci implique que les sources majeures dans cette bande de fréquences (le TNS et l'autoroute A7 d'après le chapitre 2) sont globalement plus proches du réseau que d'ADHE en-dessous de 8 km et inversement au-delà. Deuxièmement, les corrélations brutes montrent des signaux sans cohérence apparente et de forte amplitude entre 7 et 11 km de distance. Ces signaux ne semblent pas s'atténuer dans la fenêtre de temps d'observation. Pour visualiser clairement l'un de ces signaux, la fonction de corrélation entre ADHE et D-M (node proche du barrage de Donzère-Mondragon et situé à environ 8 km d'ADHE, voir la figure 2.6 du chapitre 2) pour la composante ZZ est représentée au bas de la figure 4.4. On constate que l'amplitude est très peu variable et que la période d'oscillation est la même dans la fenêtre de temps observée. Cela reflète le caractère monochromatique de ces signaux.

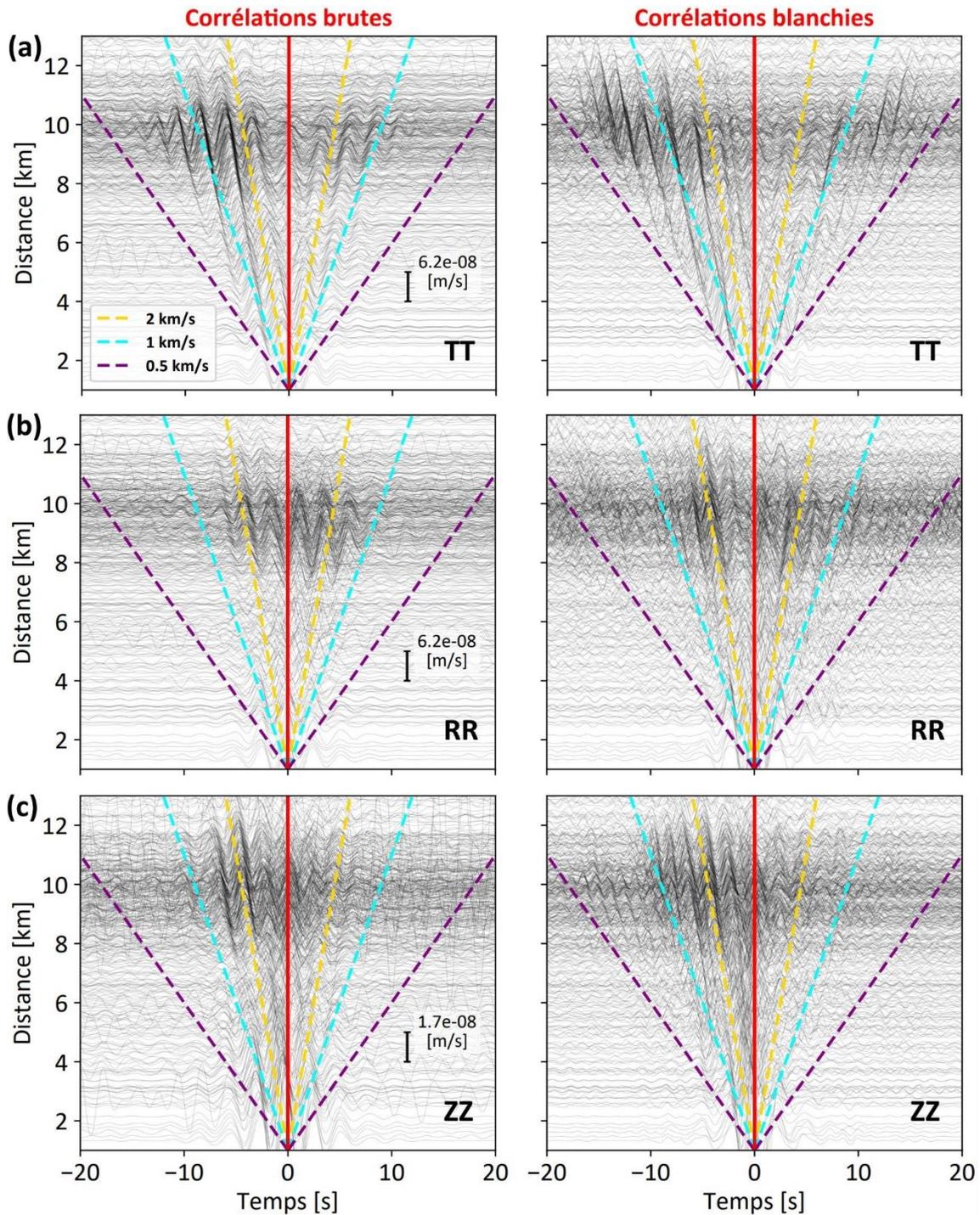


Figure 4.3 : Corrélations brutes et blanchies (respectivement colonne de gauche et colonne de droite) classées en fonction de la distance séparant ADHE (source virtuelle) du node associé à la paire (récepteur). Les corrélations sont filtrées entre 0.35 et 0.9 Hz. Les panneaux (a), (b) et (c) correspondent respectivement aux composantes TT, RR et ZZ. Les temps positifs correspondent à la partie causale de la fonction de Green partiellement reconstruite, les temps négatifs à la partie anti-causale. A titre indicatif, les lignes en pointillés correspondent à différentes vitesses de propagation.

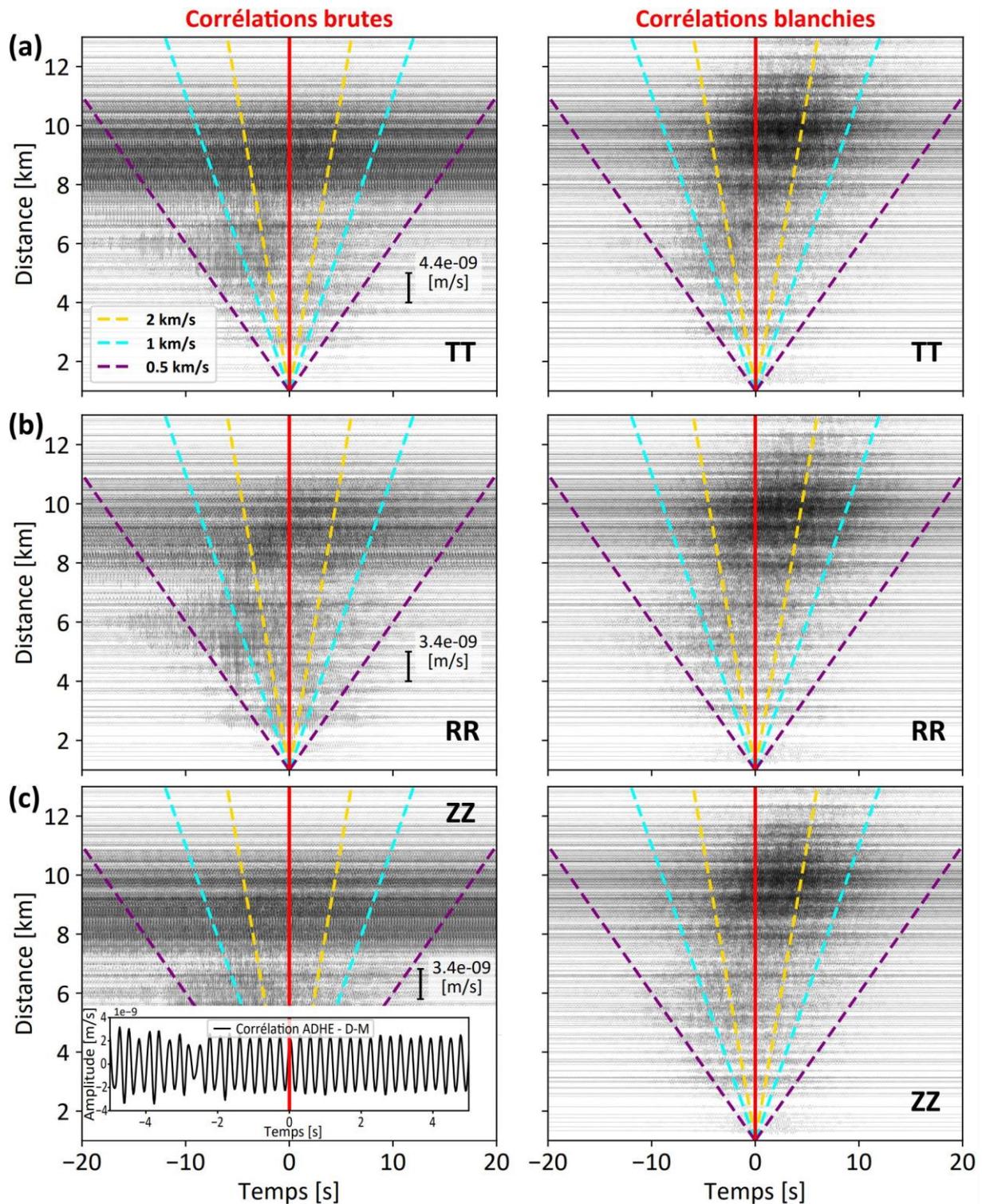


Figure 4.4 : Idem que pour la figure 4.3, entre 3 et 6 Hz. La fonction de corrélation brute entre ADHE et D-M (node proche du barrage de Donzère-Mondragon et situé à environ 8 km d'ADHE, voir la figure 2.6 du chapitre 2) pour la composante ZZ est représentée au bas de la figure.

4.2.2 Analyse des signaux monochromatiques dans les fonctions de corrélation brutes

Le spectre d'amplitude de chaque corrélation (cross-spectre) a été calculé pour la composante ZZ pour analyser les signaux monochromatiques observés sur la figure 4.4. La totalité de la fonction de corrélation a été considérée dans les calculs (2 x 15 min). La figure 4.5(a) donne la distribution des spectres d'amplitude entre 0.2 et 10 Hz. La surface noircie contient l'ensemble des valeurs entre les percentiles 16 et 84. La médiane est représentée en gris. La distribution des spectres met évidence des pics étroits de forte amplitude. Ces pics sont la signature fréquentielle des signaux monochromatiques observés dans le domaine temporel. Afin de comprendre leurs origines, on s'est concentré sur 2 bandes de fréquences étroites et contiguës (Figure 4.5b, panneau du haut). L'une est centrée autour d'un pic de forte amplitude (bande rouge, [3.56 ; 3.58] Hz), l'autre ne contient aucun pic de forte amplitude (bande bleue, [3.54 ; 3.56] Hz). L'idée est de comparer pour chaque spectre, l'amplitude médiane entre ces 2 bandes de fréquences. En d'autres termes, pour chaque spectre, la médiane des amplitudes dans la bande rouge a été normalisée par la médiane des amplitudes dans la bande bleue. Les résultats des rapports ont été associés au node correspondant à chaque spectre (node formant la paire avec ADHE) et cartographiés (Figure 4.5b, panneau du bas). Cette carte permet de visualiser les nodes les plus impactés par ce pic d'amplitude. Le facteur d'amplification résultant est très élevé dans un rayon légèrement supérieur à 1 km du barrage (jusqu'à 10) et diminue rapidement avec la distance au barrage. Le même travail a été réalisé pour un deuxième pic observé à plus haute fréquence (Figure 4.5c), entre [7.875 ; 7.895] Hz (bande rouge) et [7.855 ; 7.875] Hz (bande bleue). Le facteur d'amplification est cette fois très élevé autour de l'empreinte du TNS. Cette distribution spatiale indique que ces signaux monochromatiques sont directement liés aux sources anthropiques majeures dans la zone.

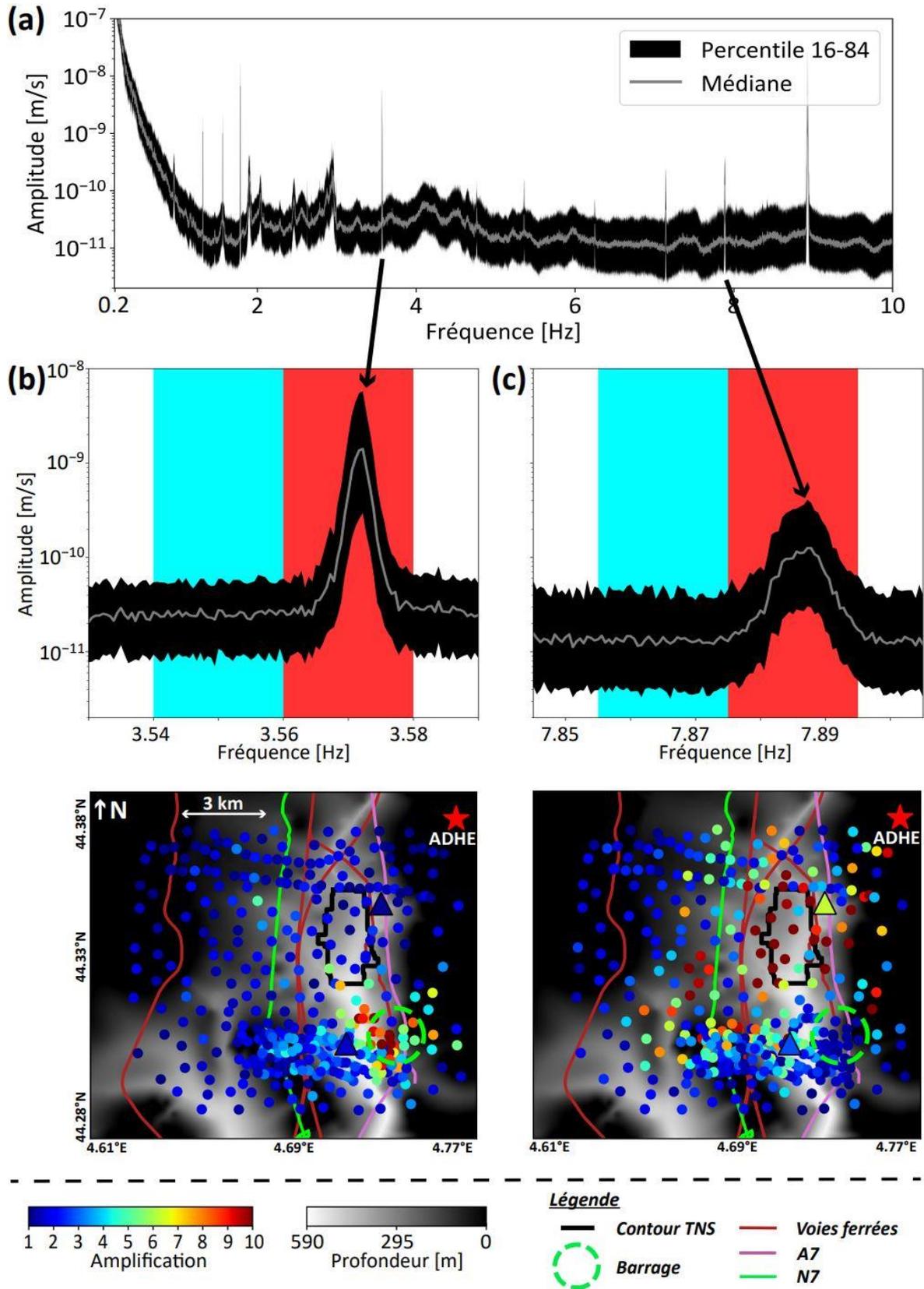


Figure 4.5 : (a) Distribution des spectres d'amplitude entre 0.2 et 10 Hz pour la composante ZZ. La surface noircie contient l'ensemble des valeurs entre les percentiles

16 et 84. La médiane est représentée en gris. (b) Panneau du haut : zoom sur 2 bandes de fréquences étroites et contiguës, l'une centrée autour d'un pic de forte amplitude (bande rouge, [3.56 ; 3.58] Hz), l'autre ne contenant aucun pic de forte amplitude (bande bleue, [3.54 ; 3.56] Hz) ; panneau du bas : amplification résultant du rapport à chaque node de la médiane des amplitudes dans la bande rouge par la médiane des amplitudes dans la bande bleue. (c) Idem que pour (b) entre [7.875 ; 7.895] Hz (bande rouge) et [7.855 ; 7.875] Hz (bande bleue).

4.2.3 Détection et suppression des signaux monochromatiques

Plusieurs approches de détection et suppression des signaux monochromatiques ont été introduites dans différents domaines géophysiques. La plupart se basent sur la modélisation de ces signaux pour les soustraire à la trace sismique brute (e.g. Xu *et al.* 2013). Ces approches sont souvent coûteuses en termes de temps de calcul. Concernant l'unique étape de suppression, les signaux monochromatiques sont généralement éliminés avec l'utilisation d'un filtre coupe-bande. Cependant, le spectre de réponse du filtre doit être adapté à la fréquence centrale et à la largeur de chaque pic. De plus, l'amplitude résultante est quasi-nulle dans la bande de fréquences d'application du filtre. Une procédure alternative de détection et suppression automatique des signaux monochromatiques est ici introduite. La procédure s'applique directement sur le spectre d'amplitude de la fonction de corrélation. Elle ne nécessite aucun a priori sur le spectre et est instantanée pour un nombre de pics détecté inférieur à 100.

La figure 4.6 illustre l'application de la procédure au spectre d'amplitude de la fonction de corrélation entre ADHE et D-M pour la composante ZZ (en noir). Tout d'abord, le spectre d'amplitude brut de la fonction de corrélation est lissé (courbe grise, Figure 4.6a). Un filtre de Butterworth (e.g. Butterworth, 1930) passe-bas zéro phase d'ordre 1 est ici utilisé. La fréquence de coupure f_c du filtre doit correspondre à la largeur (à sa base) du pic de forte amplitude le plus large pour lisser au maximum les signaux monochromatiques, sans toutefois trop lisser l'ensemble du spectre. Nous avons déterminé expérimentalement cette largeur comme étant environ égale à 0.01 Hz (donc f_c est ici égale à 0.01 Hz). Cependant, pour de futures applications, il serait judicieux de déterminer cette largeur en se basant sur la physique des ondes. Dans l'idée, les ondes interagissent avec les hétérogénéités du milieu dans des gammes de longueurs d'ondes d'une certaine largeur, et par conséquent, dans des gammes de fréquences d'une certaine largeur. La largeur de la bande de fréquences la plus étroite représentative de l'interaction ondes-structure pourrait être utilisée comme

fréquence de coupure. Ensuite, à chaque fréquence, l'amplitude du spectre brut (en noir) est comparée à celle du spectre lissé (en gris). Les amplitudes 3 x supérieures au spectre lissé sont détectées (croix jaunes). Les détections contenues dans une bande de fréquences de largeur inférieure à f_c sont considérées comme un ensemble distinct marquant un pic de forte amplitude. L'amplitude maximale de l'ensemble correspond à l'amplitude maximale du pic (A_{max} , croix verte). La fréquence correspondant à A_{max} est notée f_p . A ce stade, tous les pics de forte amplitude sont théoriquement détectés.

Pour illustrer la façon dont ils sont supprimés à travers la procédure, nous avons zoomé sur l'un d'entre eux dans la bande de fréquences [3.553 ; 3.585] Hz (Figure 4.6b ; pic déjà identifié sur la figure 4.5b, panneau du haut). La largeur du pic à droite (dfd) et à gauche (dfg) de f_p est évaluée à sa demi-hauteur ($A_{max}/2$, croix grise). D'après nos observations, la largeur totale d'un pic à sa base est inférieure à $6 \times dfg + 6 \times dfd$. Le spectre brut est donc ensuite supprimé dans la bande de fréquences [$f_p - 6 \times dfg$; $f_p + 6 \times dfd$] Hz. Finalement, la donnée manquante est remplacée par une interpolation linéaire dont les extrémités gauche et droite sont calibrées respectivement par la médiane dans la bande de fréquences [$f_p - 6 \times dfg - f_c$; $f_p - 6 \times dfg$] Hz et [$f_p + 6 \times dfd$; $f_p + 6 \times dfd + f_c$] Hz. La tendance locale du spectre est ainsi conservée. Notons qu'à travers la procédure, la phase n'a pas été altérée.

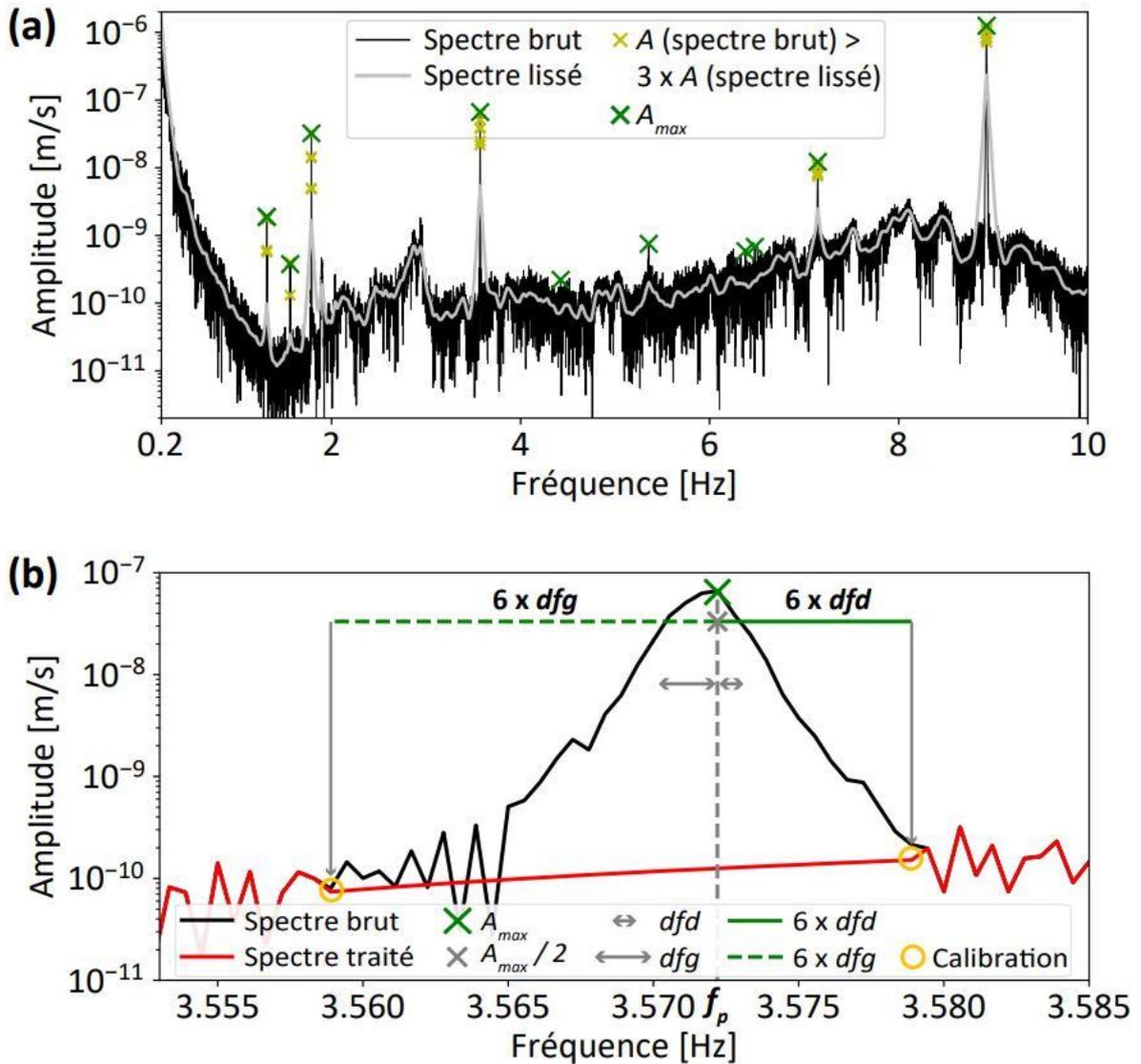


Figure 4.6 : Illustration de la procédure de détection et suppression des signaux monochromatiques. La procédure est ici appliquée au spectre d'amplitude de la fonction de corrélation entre ADHE et D-M pour la composante ZZ (en noir). (a) La courbe grise représente le spectre lissé. Les amplitudes 3 x supérieures au spectre lissé sont notées par des croix jaunes. L'amplitude maximale de chaque pic détecté est notée par une croix verte. (b) Zoom sur un pic détecté dans la bande de fréquences [3.553 ; 3.585] Hz. La largeur du pic à droite (dfd , petite double flèche grise) et à gauche (dfg , grande double flèche grise) de la fréquence centrale (f_p) est évaluée à sa demi-hauteur ($A_{max}/2$, croix grise). Une interpolation linéaire, dont les extrémités gauche et droite sont fixées par des points de calibration (cercles jaunes), remplace le spectre brut dans la bande de fréquences [$f_p - 6 \times dfg$; $f_p + 6 \times dfd$] Hz. Le spectre traité résultant est en rouge.

4.2.4 Analyse des corrélations traitées

La procédure a été appliquée à l'ensemble des spectres d'amplitude issus des fonctions de corrélation entre ADHE et le réseau pour les 3 composantes (TT, RR, ZZ). Sur la figure 4.7, la médiane des spectres d'amplitude traités (en rouge) est comparée à la médiane des spectres d'amplitude bruts (en gris). Les panneaux (a), (b) et (c) correspondent respectivement aux composantes TT, RR et ZZ. Tous les pics de forte amplitude présents sur la médiane des spectres bruts sont fortement atténués, voire supprimés sur la médiane des spectres traités quelle que soit la composante. Les fonctions de corrélation traitées et brutes ont également été comparées dans le domaine temporel entre 0.35 et 0.9 Hz (Figure 4.8) et entre 3 et 6 Hz (Figure 4.9). Les panneaux (a), (b) et (c) correspondent respectivement aux composantes TT, RR et ZZ. La colonne de gauche et la colonne de droite correspondent respectivement aux corrélations brutes et traitées. Entre 0.35 et 0.9 Hz (Figure 4.8), les corrélations sont similaires car aucun signal monochromatique n'a été détecté sur les corrélations brutes. Entre 3 et 6 Hz (Figure 4.9), la comparaison confirme l'atténuation des signaux monochromatiques après traitement pour les 3 composantes.

Les observations faites dans le domaine fréquentiel et temporel valident l'efficacité de la procédure. L'ANIRF (chapitre 5) sera calculée à partir des corrélations traitées.

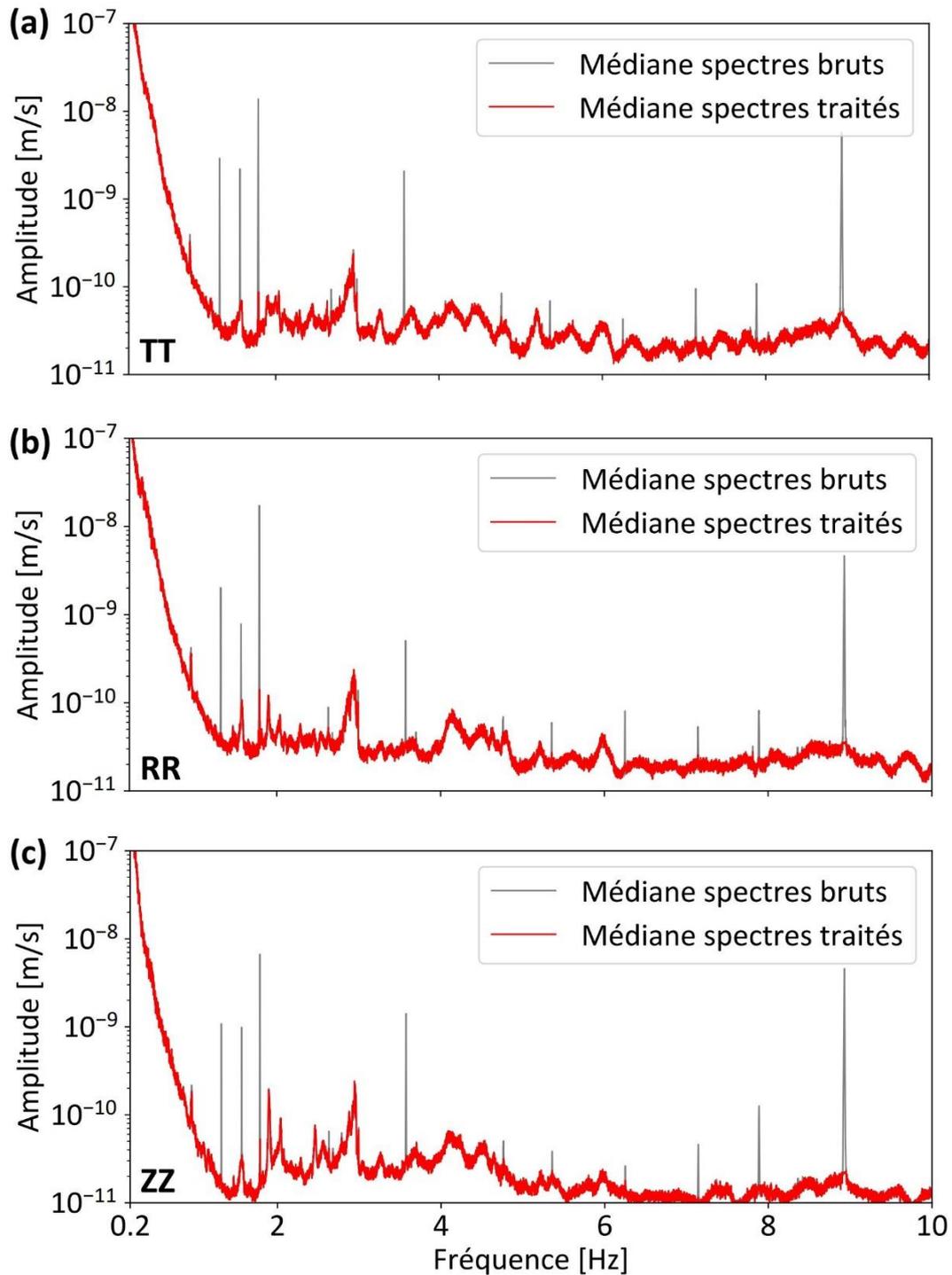


Figure 4.7 : Médiane des spectres d'amplitude traités (en rouge) comparée à la médiane des spectres d'amplitude bruts (en gris) entre 0.2 et 10 Hz. Les panneaux (a), (b) et (c) correspondent respectivement aux composantes TT, RR et ZZ.

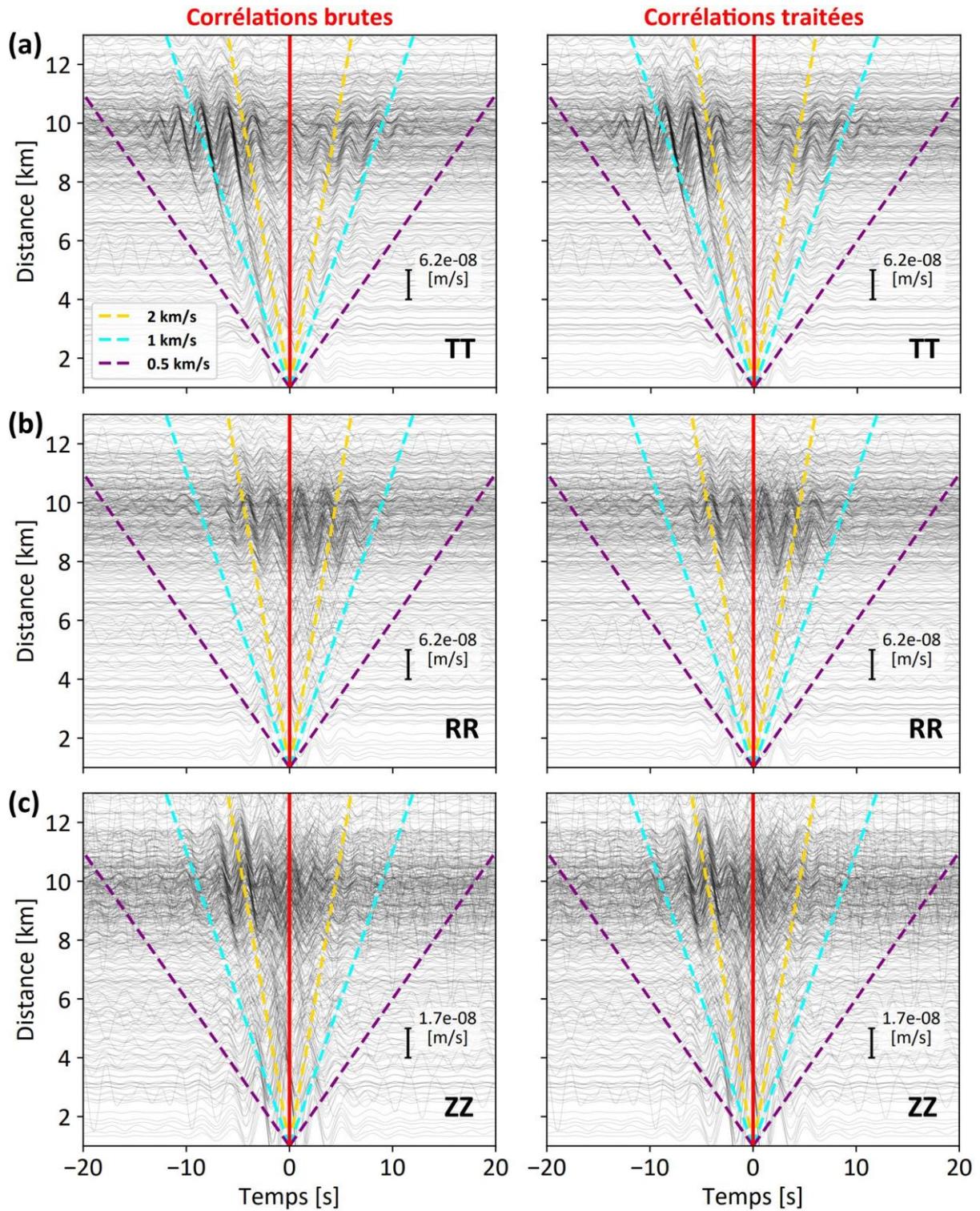


Figure 4.8 : Idem que pour la figure 4.3, pour des corrélations brutes (colonne de gauche) et traitées (colonne de droite).

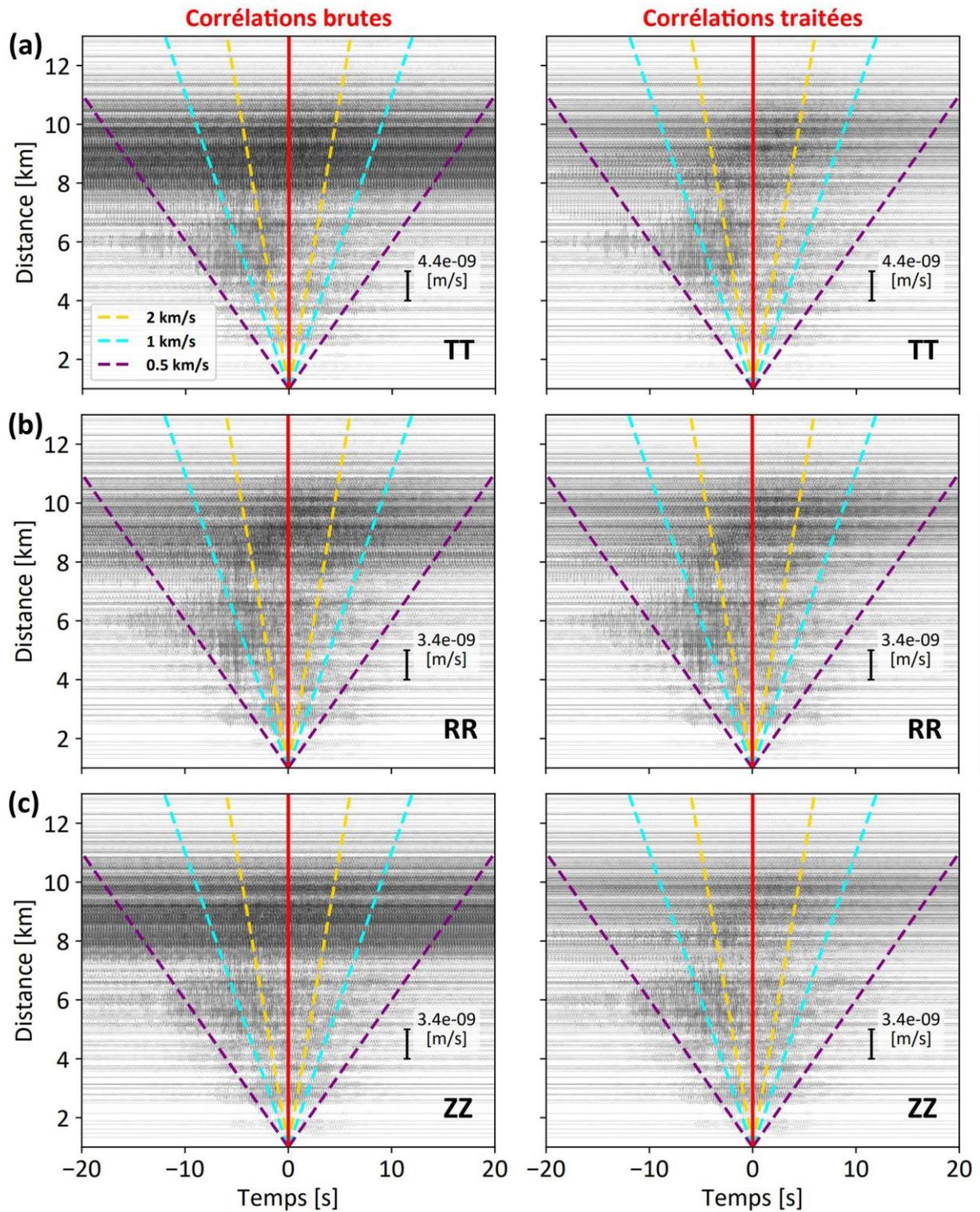


Figure 4.9 : Idem que pour la figure 4.8, entre 3 et 6 Hz.

4.3 Jeu de données utilisé pour le calcul de l'ANIRF

Comme mentionné dans l'introduction de ce chapitre, nous chercherons dans le chapitre 5 à estimer l'amplification associée à la vallée du Tricastin à travers l'ANIRF. Nous évaluons ici l'applicabilité de la procédure introduite dans le chapitre 3 (Gisselbrecht *et al.* 2023) dans le cadre des calculs de l'ANIRF. Rappelons que celle-ci consiste à limiter l'influence des sources locales de bruit sismique en sélectionnant automatiquement les fenêtres de temps les moins impactées à chaque node. Cette procédure s'est avérée efficace pour traiter les sources transitoires. Le jeu de données résultant s'en trouve néanmoins significativement réduit à certains nodes. La question est de savoir si la donnée sélectionnée à ces nodes est suffisante dans le cadre des calculs de l'ANIRF. Pour cela, il faudrait que les fenêtres de temps sélectionnées à ces nodes, synchrones avec celles sélectionnées à la source virtuelle (ici ADHE), couvrent une durée totale suffisante pour permettre la convergence des fonctions de corrélation.

Dans la procédure présentée dans le chapitre 3 (Gisselbrecht *et al.* 2023), les fenêtres de temps utilisées sont des fenêtres nocturnes d'1 min (entre 20H00 et 6H00). La figure 4.10(a) scanne la durée consécutive synchrone entre chaque node et ADHE à travers les plages nocturnes sur la totalité de l'acquisition en considérant les fenêtres d'1 min sélectionnées par la procédure. Les nodes sont classés en fonction de la durée totale synchrone avec ADHE. Pour les nodes indexés de 200 à 400, la durée synchrone avec ADHE dépasse régulièrement 30 min consécutives. Pour les nodes indexés de 100 à 200, celle-ci atteint souvent 10 min consécutives. En revanche, pour une centaine de nodes (index 0 à 100), presque aucune plage de temps synchrone ne dépasse 5 min consécutives et donc la donnée sélectionnée par la procédure serait potentiellement faible en vue des calculs de l'ANIRF. Ces nodes sont ceux pour lesquels la sélection opérée par la procédure est la plus drastique. Parmi ces nodes, 69 sont concernés par une différence significative (supérieure à 20 %) entre le SSRh classique et celui calculé sur la base de la procédure (appelé DBSCAN_SSRh ; Gisselbrecht *et al.* 2023). Pour chacun de ces 69 nodes, il est nécessaire de voir si la donnée sélectionnée serait suffisante pour les calculs de l'ANIRF. La durée totale synchrone avec ADHE a donc été calculée en fonction de fenêtres de temps allant de 1 à 30 min (Figure 4.10b). Cette durée diminue avec la largeur de la fenêtre considérée. Pour des fenêtres courtes de 2 min, 46 % de ces nodes dépassent 10 h (soit 1 nuit) de données synchrones avec ADHE, contre seulement 4 et 2 % pour des fenêtres de 10 et 20 min.

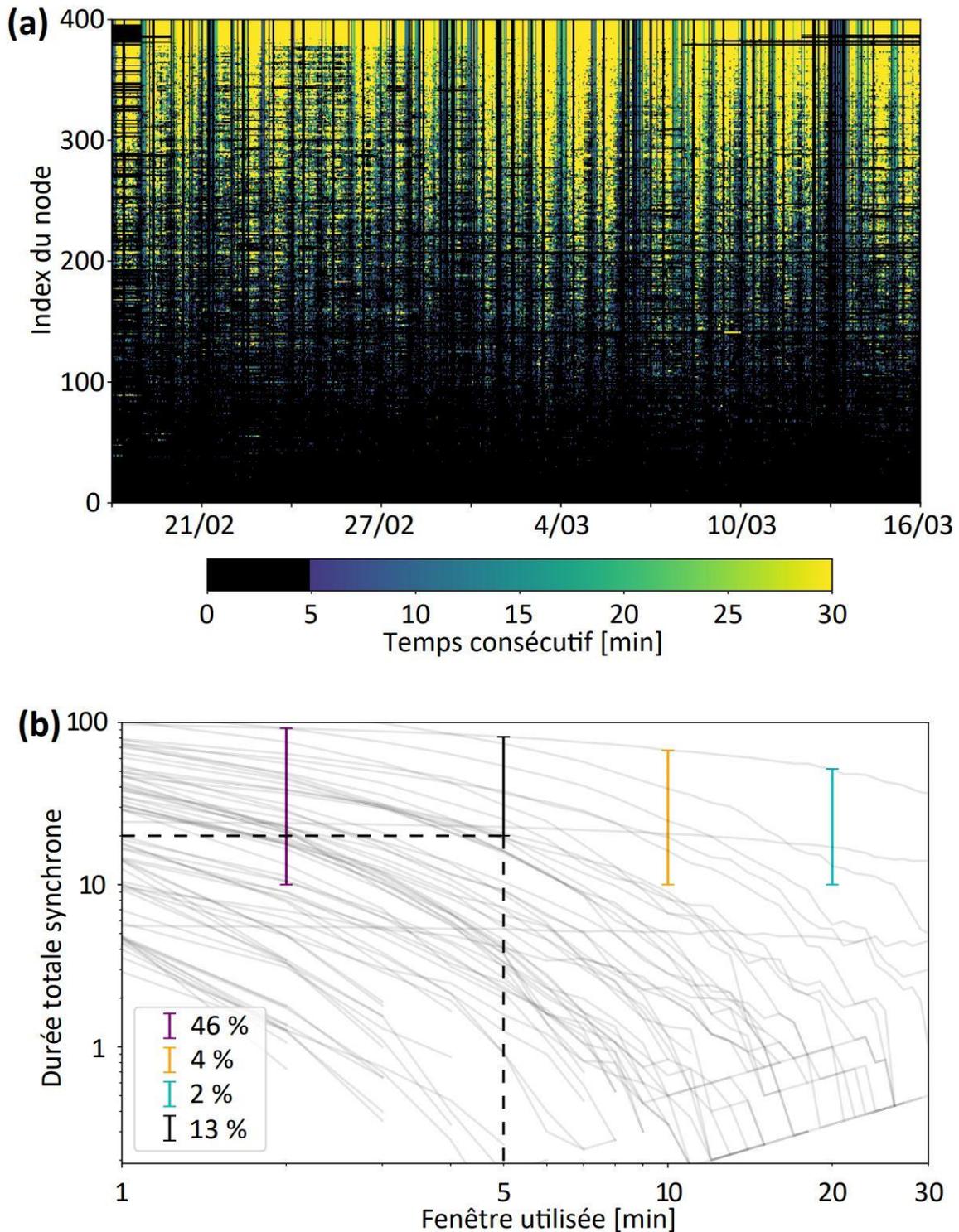


Figure 4.10 : (a) Durée consécutive synchrone entre chaque node et ADHE à travers les plages nocturnes (20H00 - 6H00) sur la totalité de l'acquisition en considérant les fenêtres d'1 min sélectionnées par la procédure. (b) Durée totale synchrone (courbes grises) avec ADHE en fonction de différentes fenêtres de temps (de 1 à 30 min) pour 69 nodes pour lesquels l'impact de notre procédure est significatif (Gisselbrecht et al. 2023).

Pour maximiser la donnée, il faudrait donc calculer l'ANIRF sur la base de fenêtres courtes. Ces fenêtres doivent cependant être suffisamment larges pour capter la totalité de la propagation entre la source virtuelle et le réseau. Une largeur de fenêtre de 5 min semble être à priori un bon compromis. Pour avoir une idée de la donnée nécessaire à la convergence des corrélations, l'évolution du coefficient de corrélation⁸ (ou Ressemblance, R) en fonction du temps (sur la base de fenêtres de 5 min) pour la paire ADHE-BOLL a été représentée (Figure 4.11). Aucun recouvrement temporel n'a été appliqué. Les corrélations ont été filtrées entre 0.35 et 0.9 Hz (a) et entre 3 et 6 Hz (b). Les courbes en dégradé de bleu, vert et rouge représentent respectivement les composantes TT, RR et ZZ. La convergence est plus rapide pour les composantes horizontales, en particulier à basse fréquence. Entre 0.35 et 0.9 Hz (a), 20 h sont nécessaires pour atteindre un coefficient de corrélation de 0.9 pour les composantes horizontales, 60 h pour la composante verticale. Entre 3 et 6 Hz (b), la convergence est globalement plus lente : 60 h sont nécessaires pour atteindre un tel coefficient pour les composantes horizontales, 100 h pour la composante verticale. Or, seul 13 % des nodes pour lesquels la procédure à un impact significatif dépasse 20 h de données synchrones avec ADHE (Figure 4.10b). Cette analyse montre donc que la donnée disponible après application de la procédure de sélection proposée dans le chapitre 3 est globalement insuffisante à la convergence des fonctions de corrélation (même en considérant des fenêtres courtes). Cette procédure semble ainsi difficilement applicable dans le cadre des calculs de l'ANIRF (chapitre 5).

Le taux de convergence varie selon plusieurs paramètres (milieu de propagation, distance avec la source virtuelle, fréquence, largeur de la fenêtre temporelle, pourcentage de recouvrement des fenêtres... ; e.g. Seats *et al.* 2012). Pour réellement statuer sur l'applicabilité de la procédure dans le cadre de l'ANIRF, une étude approfondie de la convergence pour chacune des 69 paires serait nécessaire. Une approche adaptative pourrait être envisagée, impliquant des largeurs de fenêtre variables dépendamment de la paire considérée pour maximiser la donnée. L'effet de l'application d'un recouvrement temporel pourrait également être étudié. Une telle étude n'a pas pu être réalisée ici. Néanmoins, les observations faites tendent à montrer qu'un temps d'acquisition plus long est nécessaire pour permettre son application en amont d'un calcul de déconvolution simple et homogène à toutes les paires. L'ANIRF

⁸ Le coefficient de corrélation est une mesure de la ressemblance entre deux signaux. L'évolution du coefficient de corrélation décrit la convergence des fonctions de corrélation progressivement moyennées sur les fenêtres de temps vers la fonction de corrélation finale (moyennée sur toutes les fenêtres de temps).

(chapitre 5) ne sera donc pas calculée sur la base de la procédure. Seule une sélection des fenêtres nocturnes (20H00 - 6H00) sera appliquée.

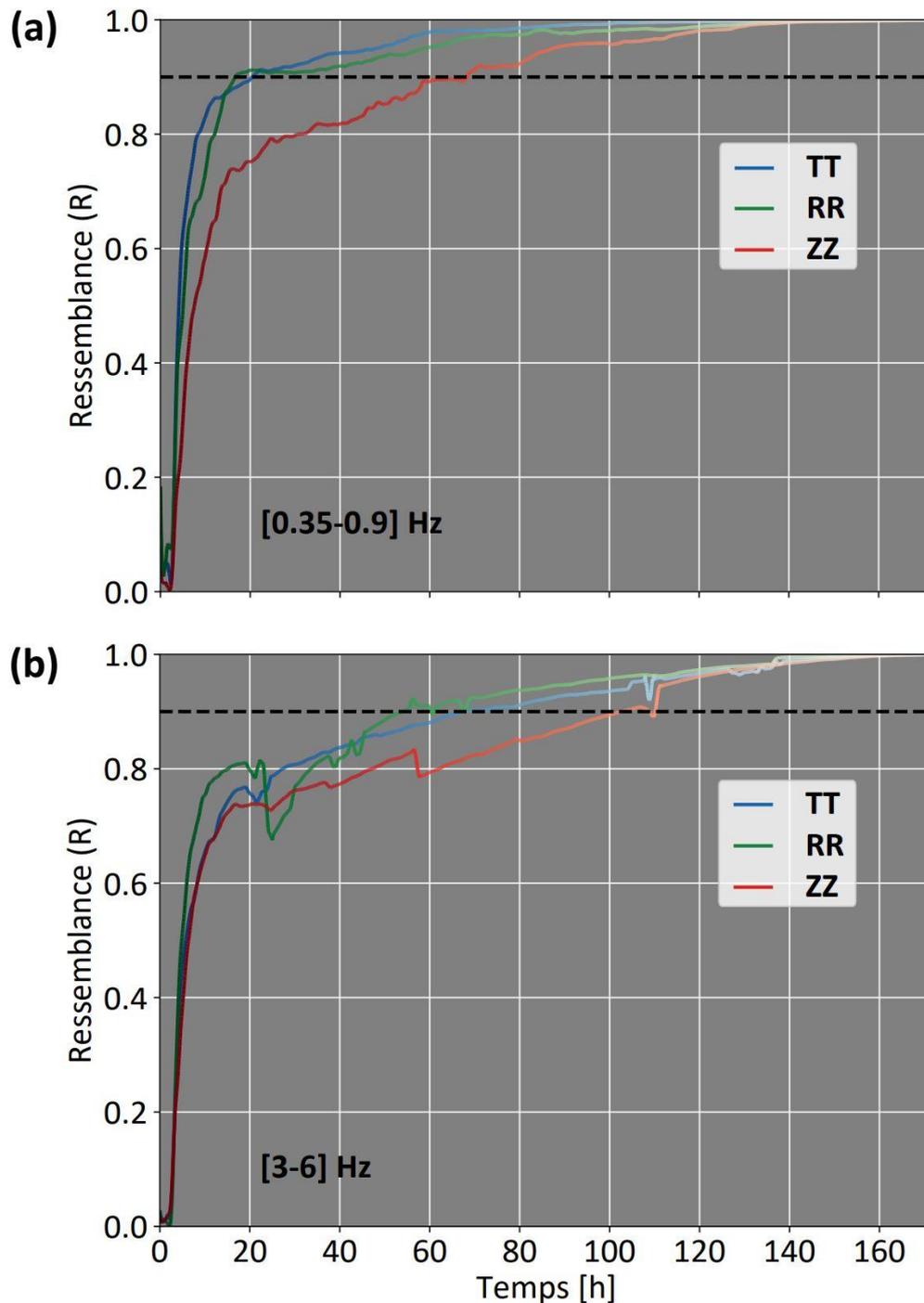


Figure 4.11 : Evolution du coefficient de corrélation (ou Ressemblance, R) en fonction du temps (sur la base de fenêtres de 5 min) pour la paire ADHE-BOLL. Aucun recouvrement temporel n'a été appliqué. Les corrélations ont été filtrées entre 0.35 et 0.9 Hz (a) et entre 3 et 6 Hz (b). Les courbes en dégradé de bleu, vert et rouge dénotent respectivement les composantes TT, RR et ZZ.

Chapitre 5 : Estimation de l'amplification associée à la vallée du Tricastin en utilisant l'ANIRF

5.1 Introduction

Ce chapitre discute de la possibilité d'utiliser l'ANIRF (introduit dans le chapitre 4) pour estimer l'amplification associée aux effets de site. A ce titre, l'ANIRF a été utilisée dans des études relativement récentes (Prieto et Beroza 2008 ; Prieto *et al.* 2011 ; Denolle *et al.* 2013, 2014 ; Viens *et al.* 2014, 2015, 2016 ; Bowden 2015 ; Perron 2017 ; Denolle *et al.* 2018).

Prieto et Beroza (2008) ont comparé les sismogrammes du séisme de Big Bear (en 2001, Mw 4.63) aux sismogrammes virtuels obtenus par l'ANIRF à plusieurs stations situées dans et autour du bassin de Los Angeles. Un facteur de calibration d'amplitude au séisme (commun à toutes les paires) a été appliqué. Une forte similitude, en termes d'amplitude, de phase et de durée des signaux a été observée entre les sismogrammes du séisme et les sismogrammes virtuels. Les discordances subsistantes s'expliquent par : (1) la difficulté d'extraire la cohérence du champ de bruit sur une large gamme de fréquences (ici la fonction de Green n'a pu être reconstruite qu'à basse fréquence, entre 0.1 et 0.25 Hz) ; (2) la différence de trajet entre les ondes reconstruites par l'ANIRF et les ondes générées par le séisme (due à la différence de position entre les sources virtuelles en surface et le séisme en profondeur) ; (3) la différence du mécanisme source (source impulsionnelle pour l'ANIRF, double couple de forces pour un séisme).

Denolle *et al.* (2013) ont cherché à traiter les difficultés (2) et (3). Ils ont corrigé l'ANIRF en convertissant la source impulsionnelle en surface en un point source en profondeur dont le mécanisme correspond à un double couple de forces. L'ANIRF corrigée a été appelée VEA (Virtual Earthquake Approach). L'ANIRF et la VEA ont ensuite été comparées aux sismogrammes de 4 séismes d'intensité modérée (Chino

Hills, Mw 5.4 ; Hector Road, Mw 5.1 ; San Fernando, Mw 4.2 et San Bernardino, Mw 4.5) pour différentes stations dans et autour du bassin de Los Angeles. La comparaison a été réalisée entre 0.1 et 0.25 Hz pour les 3 composantes diagonales du tenseur de Green (TT, RR, ZZ). Les résultats ont montré qu'au premier ordre, l'ANIRF est comparable aux sismogrammes des 4 séismes pour la majorité des stations, ce qui démontre sa capacité à retranscrire les informations liées à la propagation des ondes et à la réponse du bassin (amplitude relative, atténuation géométrique, amplification et allongement de la durée associés aux effets de site). Au second ordre en revanche, la VEA a montré un meilleur accord que l'ANIRF avec les sismogrammes, en particulier concernant les 2 séismes les plus forts pour lesquels l'impact du mécanisme de la source est plus important. En sommant les contributions de plusieurs points sources en profondeur, Denolle *et al.* (2014) ont modifié la VEA initiale (valable pour les séismes d'intensité modérée) en une VEA qui intègre un mécanisme de rupture finie (valable pour les séismes forts). Ils ont notamment pu observer l'influence de la directivité de la source sur la capacité des bassins de Los Angeles, Ventura et Salton Trough à piéger les ondes. L'amplitude du mouvement sismique dans ces différents bassins sédimentaires augmente en effet lorsque la rupture se propage dans leur direction. Un facteur 3 à 5 environ sépare les amplitudes mesurées entre une rupture qui se propage vers le bassin ou dans le sens opposé.

Tout comme Prieto et Beroza (2008) et Denolle *et al.* (2013), Viens *et al.* (2014, 2015) ont observé de fortes similitudes à basse fréquence (< 0.25 Hz) entre l'ANIRF et plusieurs sismogrammes de séismes d'intensité modérée ($4 < Mw < 6$) dans la région de Tokyo. Plus tard, Viens *et al.* (2016) ont été capables de comparer l'ANIRF avec les sismogrammes de 2 séismes (Mw 4.9 et Mw 5.8) aux stations du réseau MeSO-net dans le bassin de Kanto (baie de Tokyo) jusqu'à 1 Hz. Dans cette même étude, l'amplitude a été cartographiée à l'échelle du bassin à 6 s de période (proche de la période de résonance fondamentale du bassin). Les cartes obtenues avec l'ANIRF ont montré des variations d'amplitude similaires aux cartes obtenues pour les 2 séismes.

Toutes ces études confirment que l'ANIRF retranscrit bien au premier ordre l'information sur la propagation des ondes et la réponse des bassins sédimentaires à basse fréquence (en général jusqu'à 0.25 Hz et dans certains cas jusqu'à 1 Hz)⁹. Une des clefs de voûte de l'ANIRF à ces fréquences réside dans la capacité de la fonction de corrélation à extraire la cohérence du champ de bruit et à capturer les effets 2D/3D du bassin (réflexion et conversion d'ondes en bord de bassin, piégeage, interférences,

⁹ Notons que les bassins sédimentaires étudiés (notamment Los Angeles et Kanto) ont une fréquence de résonance fondamentale proche du 2nd pic micro-sismique (voir section 1.3, chapitre 1).

focalisation, génération de modes supérieurs ; e.g. Boué *et al.* 2016). Cependant la bande de fréquences d'intérêt pour l'ingénierie est bien plus large (jusqu'à environ 10 Hz) et au-delà de 1 Hz, il peut s'avérer difficile de conserver l'information de propagation (liée à la phase) à l'échelle de la structure géologique. En revanche l'information sur l'amplitude portée par l'ANIRF est potentiellement utilisable au-delà de 1 Hz. Perron (2017) a par exemple montré, dans le cas du bassin d'Argostoli (Grèce), que le facteur d'amplitude entre 2 fonctions ANIRF partageant la même source virtuelle, l'une impliquant un récepteur au site, l'autre impliquant un récepteur au rocher, est comparable au SSR site/rocher entre ces 2 récepteurs jusqu'à 4 Hz.

Dans ce chapitre, on se propose d'examiner la possibilité d'utiliser l'ANIRF pour estimer l'amplification associée à la vallée du Tricastin sur une large gamme de fréquences. L'ANIRF telle que définie par Prieto *et al.* (2011) a ici été modifiée pour accélérer les temps de calcul : plutôt que de moyenner sur les fenêtres de temps les rapports entre la fonction de corrélation et l'autocorrélation à la source virtuelle (eq. 1), nous avons considéré le rapport de ces 2 termes déjà moyennés (eq. 2). Cela limite la répétition de l'opération de lissage¹⁰. Il pourrait néanmoins être intéressant de vérifier que ces deux opérations donnent bien des résultats équivalents dans différents environnements. Dans le cas où les propriétés statistiques du champ de bruit sont relativement stationnaires, on s'attendrait à ce que l'ANIRF (eq. 1) et l'ANIRF modifiée (eq. 2) soient identiques. Toutefois, cette hypothèse de stationnarité du champ de bruit pourrait être plus largement discutée, notamment à haute fréquence.

$$ANIRF(f) = \left\langle \frac{U_A(f)U_B^*(f)}{|U_B(f)|^2} \right\rangle \quad (1) \rightarrow ANIRF(f) = \frac{\langle U_A(f)U_B^*(f) \rangle}{\langle |U_B(f)|^2 \rangle} \quad (2) ,$$

avec $U_A(f)$ et $U_B(f)$ la transformée de Fourier du signal aux capteurs A et B, $|U_B(f)|$ le spectre d'amplitude du signal au capteur B, $\{ . \}$ une fonction de lissage et $\langle . \rangle$ la moyenne arithmétique sur les fenêtres de temps utilisées. $U_A(f)U_B^*(f)$ dénote la fonction de corrélation entre A et B où B est la source virtuelle. Le complexe conjugué est représenté par le symbole *. L'ANIRF modifiée (eq. 2) a été calculée entre 7 sources virtuelles (Figure 5.1) et chacun des 400 nodes du réseau selon les composantes diagonales du tenseur de Green (TT, RR, ZZ). Des fenêtres nocturnes de 30 min sans recouvrement ont été utilisées. Pour corriger l'atténuation géométrique, l'amplitude de l'ANIRF entre chaque paire a été multipliée par un facteur \sqrt{d} (où d est la distance

¹⁰ La fonction de lissage utilisée ici est un filtre Konno-Ohmachi (Konno et Ohmachi, 1998). Cette fonction permet de lisser le spectre de manière uniforme sur toute la bande de fréquences. En revanche, elle est chronophage en termes de temps de calcul car repose sur l'utilisation d'une fenêtre de lissage de taille adaptative (voir chapitre 3).

entre le récepteur et la source virtuelle utilisée). Les 7 sources virtuelles utilisées sont situées tout autour de la zone d'étude et pourront ainsi permettre d'étudier la réponse de la vallée dépendamment de la direction du champ incident (Figure 5.1).

Le chapitre s'organise comme suit. Dans les parties 5.2 et 5.3, l'amplitude de l'ANIRF est étudiée respectivement à basse et haute fréquence en considérant 3 sources virtuelles (V2, ADHE et V4). Dans la partie 5.4, l'amplification site/rocher associée à la vallée du Tricastin est estimée à partir de l'ANIRF en suivant une procédure en 4 étapes.

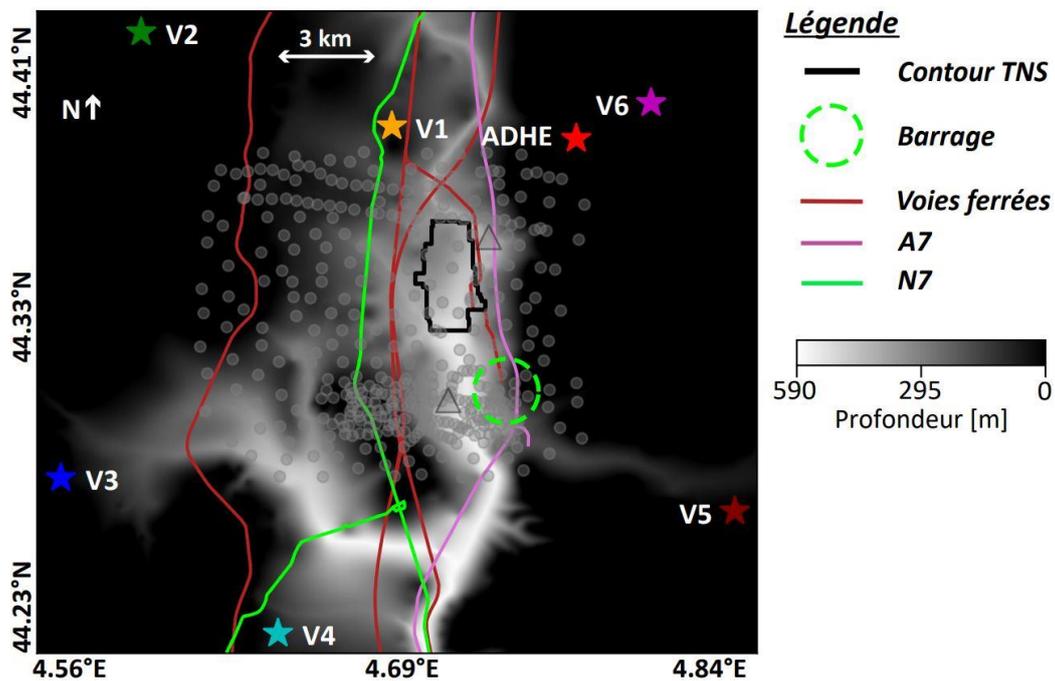


Figure 5.1 : Illustration des 7 sources virtuelles utilisées pour le calcul de l'ANIRF : ADHE (en rouge), V1 (en orange), V2 (en vert), V3 (en bleu), V4 (en bleu ciel), V5 (en marron), V6 (en violet).

5.2 Amplitude de l'ANIRF à l'échelle du réseau à basse fréquence (entre 0.35 et 0.9 Hz)

Cette partie porte sur l'analyse de l'amplitude de l'ANIRF (eq. 2) à basse fréquence. Les résultats sont présentés au cas par cas pour 3 sources virtuelles : V2, ADHE, V4 (Figure 5.1). En section 5.2.1, l'amplitude de l'ANIRF est cartographiée à différents temps d'arrivée pour suivre la propagation des fronts d'onde reconstruits à travers le réseau. L'idée est d'étudier l'interaction avec la vallée dépendamment de la direction d'incidence du champ reconstruit. En section 5.2.2, l'amplitude de l'ANIRF est cartographiée en intégrant l'ensemble des temps d'arrivée correspondant à la partie cohérente du signal reconstruit à chaque paire.

5.2.1 Propagation des fronts d'onde reconstruits à basse fréquence

Avec V2 comme source virtuelle

Les figures 5.2 et 5.3 montrent la propagation du champ d'ondes reconstruit par l'ANIRF entre la source virtuelle V2 et le réseau pour la composante TT et ZZ respectivement. L'amplitude de l'ANIRF a été cartographiée à 4 temps d'arrivée différents (panneau b) pour la partie causale et anti-causale (respectivement colonne de droite et colonne de gauche). Ces temps d'arrivée sont écrits à côté de chaque carte (panneau b) et indiqués par des barres verticales (barres vertes) sur les sections temporelles de l'ANIRF (fonctions ANIRF classées selon la distance séparant la source virtuelle du récepteur) représentées sur le panneau (a). Des flèches de couleur indiquent le suivi d'un même front d'onde. L'ANIRF a été filtrée entre 0.35 et 0.9 Hz. Les fonctions ANIRF dont le rapport signal à bruit (SNR) est inférieur à 3 n'ont pas été considérées.

Concentrons-nous d'abord sur la composante TT (Figure 5.2). L'amplitude de l'ANIRF est plus élevée pour la partie causale (plus d'énergie s'est propagée de V2 vers le réseau que du réseau vers V2 durant la période d'acquisition, voir chapitre 4). Ceci est cohérent avec les observations du chapitre 2 qui désignent l'océan Atlantique et la mer Méditerranée comme sources principales de bruit sismique à ces fréquences. Le suivi des fronts d'onde reconstruits est néanmoins possible sur les 2 parties (causale et anti-causale). Une vue générale permet de distinguer nettement plusieurs fronts d'onde traversant la vallée sur un axe NO-SE. Ces différents fronts d'onde sont marqués par une distorsion importante qui résulte d'une différence de vitesse de propagation : la partie nord-est des fronts d'onde, qui interagit avec la partie nord de la vallée du

Rhône (partie sédimentaire profonde), se propage plus lentement que la partie sud-ouest. Sur la partie causale, l'amplitude des fronts d'onde est plus élevée au centre et au nord de la vallée du Rhône. Elle est maximale lorsque ceux-ci atteignent le bord est de la vallée ($t = 5.61$ s ; $t = 6.61$ s ; $t = 7.61$ s), ce qui révèle l'interaction entre le champ d'ondes incident et la géométrie de la structure géologique.

Pour la composante ZZ (Figure 5.3), bien que la propagation d'un même front d'onde soit plus difficile à suivre, des observations similaires sont faites en termes d'amplitude, notamment en ce qui concerne le bord est de la vallée (sur la partie causale, l'amplitude est maximale lorsque le front d'onde atteint le bord est de la vallée du Rhône).

cartographiée à 4 temps d'arrivée différents (panneau b) pour la partie causale et anti-causale (respectivement colonne de droite et colonne de gauche). Ces temps d'arrivée sont écrits à côté de chaque carte (panneau b) et indiqués par des barres verticales (barres vertes) sur les sections temporelles de l'ANIRF (panneau a). Des flèches de couleur indiquent le suivi d'un même front d'onde. L'ANIRF a été filtrée entre 0.35 et 0.9 Hz.

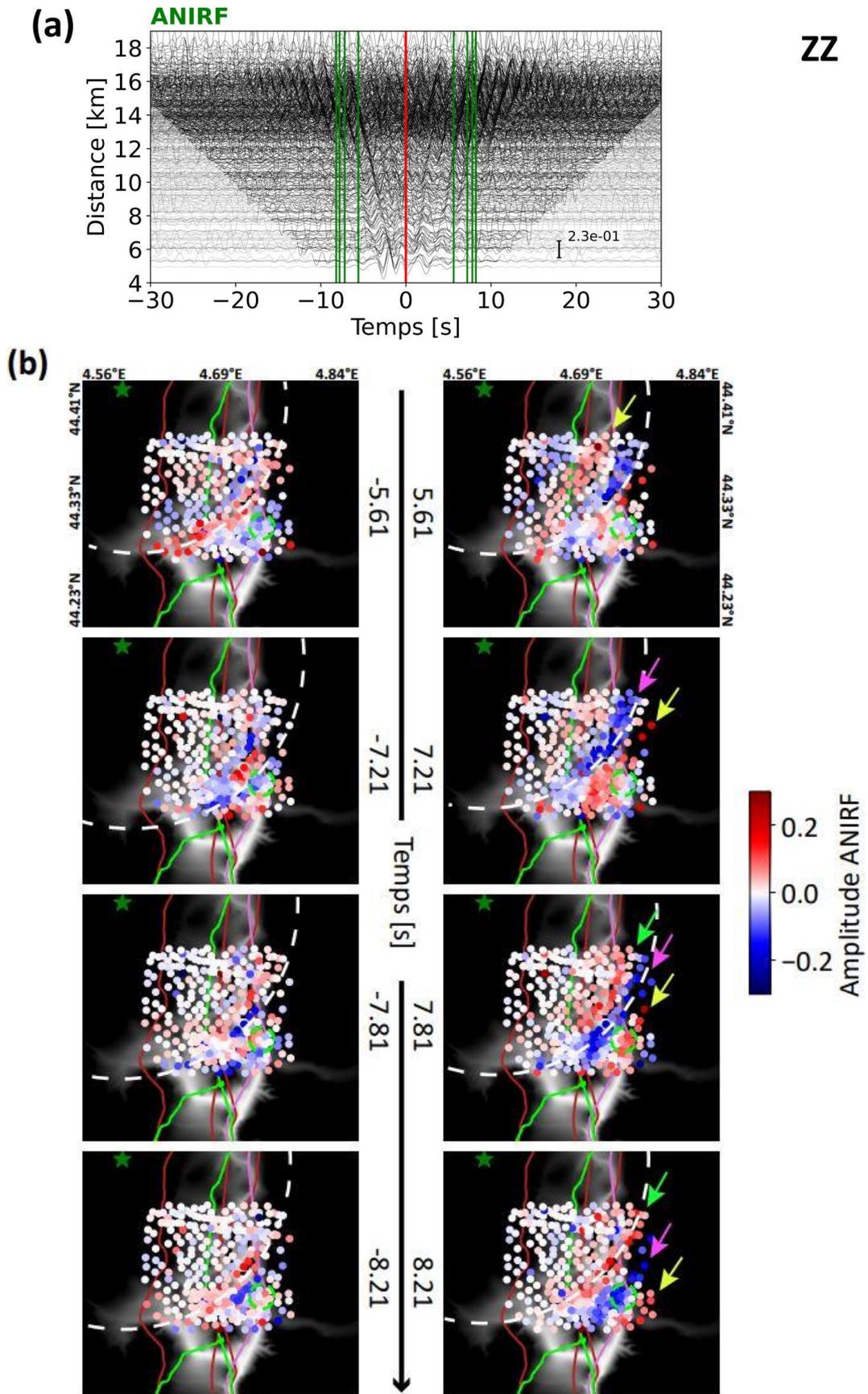


Figure 5.3 : Idem que pour la figure 5.2 pour la composante ZZ.

Avec ADHE et V4 comme sources virtuelles

Le même travail a été réalisé en utilisant ADHE comme source virtuelle (Figures 5.4 et 5.5, respectivement composantes TT et ZZ). L'amplitude de l'ANIRF est cette fois plus élevée pour la partie anti-causale (plus d'énergie s'est propagée du réseau vers ADHE que de ADHE vers le réseau durant la période d'acquisition, voir chapitre 4). Pour la composante TT (Figure 5.4), la propagation des différents fronts d'onde est bien visible (partie anti-causale, colonne de gauche). Le phénomène de distorsion des fronts d'onde causé par la diminution de la vitesse de propagation dans la vallée profonde du Rhône apparaît clairement à $t = -4.81$ s et $t = -5.61$ s. A $t = -11.01$ s et $t = -11.81$ s, l'amplitude de l'ANIRF est maximale au niveau du bord est de l'île de Lapalud (flèche noire). Durant notre analyse, nous avons observé que l'amplitude de l'ANIRF était maximale pour cette zone entre approximativement 9 et 13 s de temps de propagation. Une amplitude aussi élevée dans cette zone sur un temps aussi long (environ 4 s) révèle possiblement une forte interaction des ondes avec l'île de Lapalud. Concernant la composante ZZ (Figure 5.5), bien que la propagation d'un même front d'onde soit plus difficile à suivre, une observation similaire peut être faite entre -4.41 s et -6.01 s. Une rupture assez nette du front d'onde (marquée par une flèche noire) est d'ailleurs visible.

Enfin, les figures 5.6 et 5.7 (respectivement composantes TT et ZZ) montrent la propagation des fronts d'onde reconstruits en utilisant V4 comme source virtuelle. La partie causale de l'ANIRF concentre la plupart de l'énergie. Pour la composante TT (Figure 5.6), l'amplitude de l'ANIRF est maximale dans le canyon de l'Ardèche, la partie sud de la vallée du Rhône et au niveau du bord est de l'île de Lapalud. Cette zone, qui marque dans ce cas le point d'entrée dans la vallée profonde du Rhône, est caractérisée par une large distorsion des fronts d'onde. Cette distorsion est notamment très marquée pour la composante ZZ (Figure 5.7) à $t = 3.41$ s et $t = 7.21$ s. Notons pour cette composante l'amplitude relativement importante de l'ANIRF sur la partie N-O de la zone d'étude, pourtant située en-dehors de la vallée sédimentaire profonde.

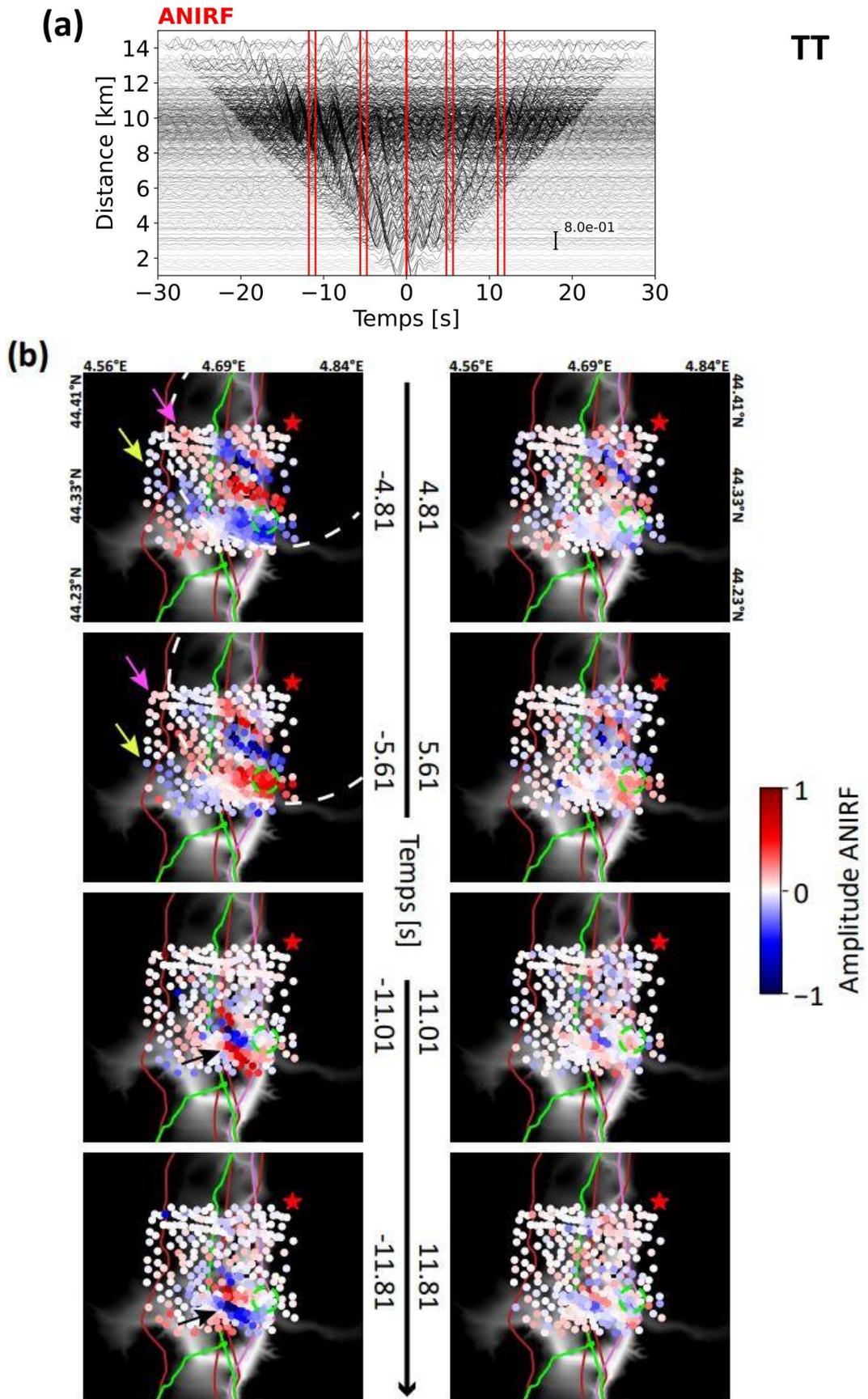


Figure 5.4 : Idem que pour la figure 5.2 avec ADHE comme source virtuelle.

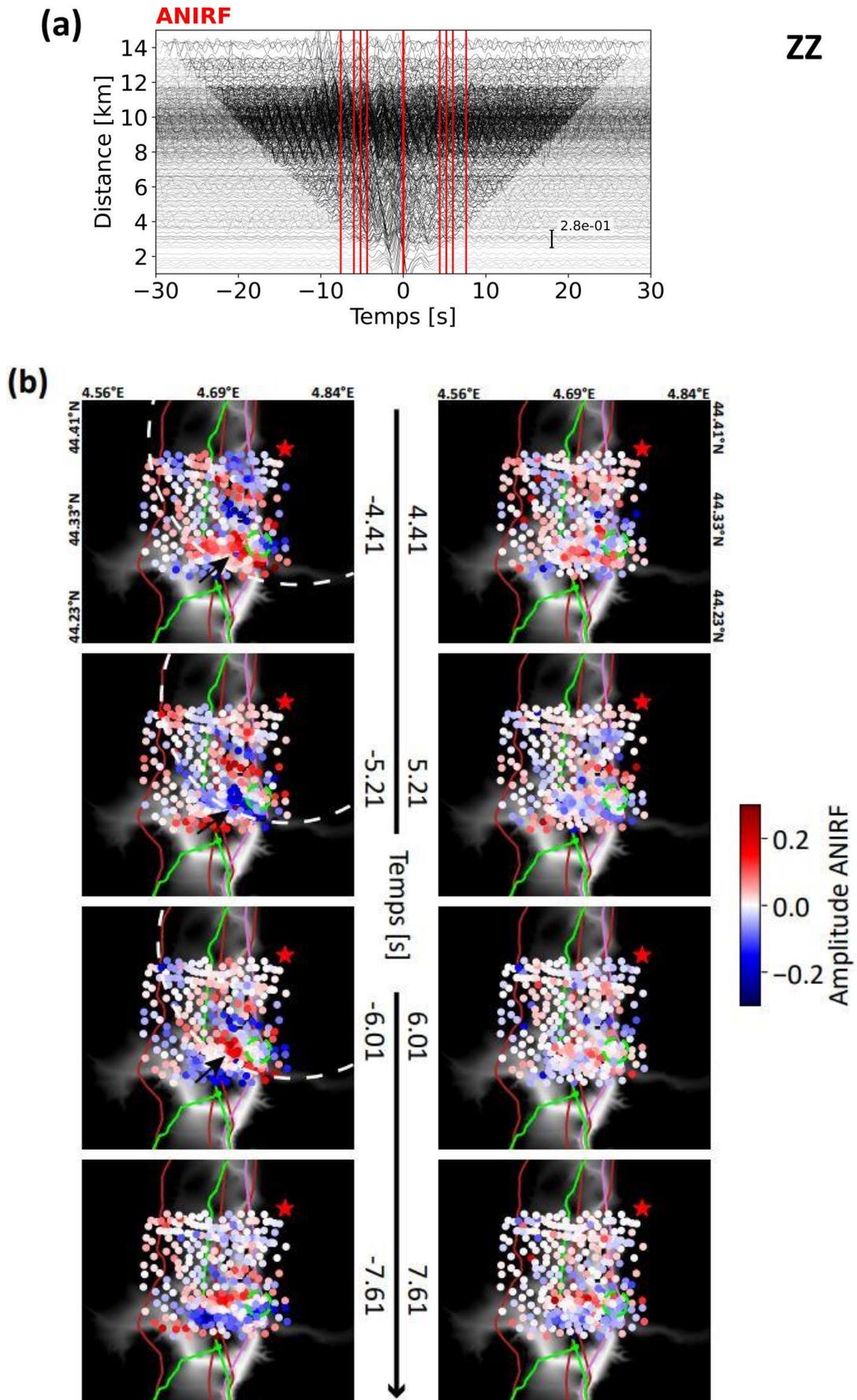


Figure 5.5 : Idem que pour la figure 5.4 pour la composante ZZ.

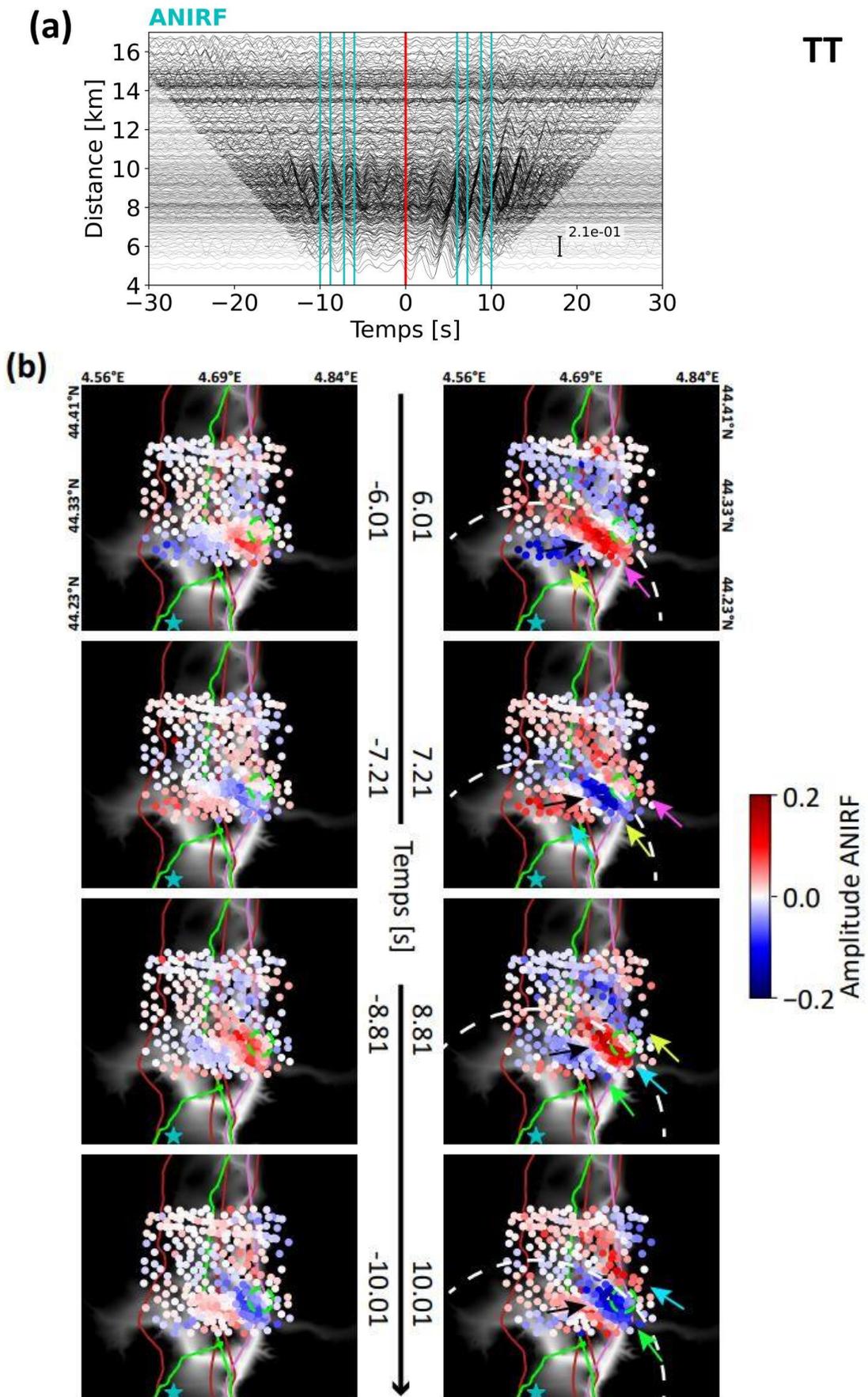


Figure 5.6 : Idem que pour la figure 5.2 avec V4 comme source virtuelle.

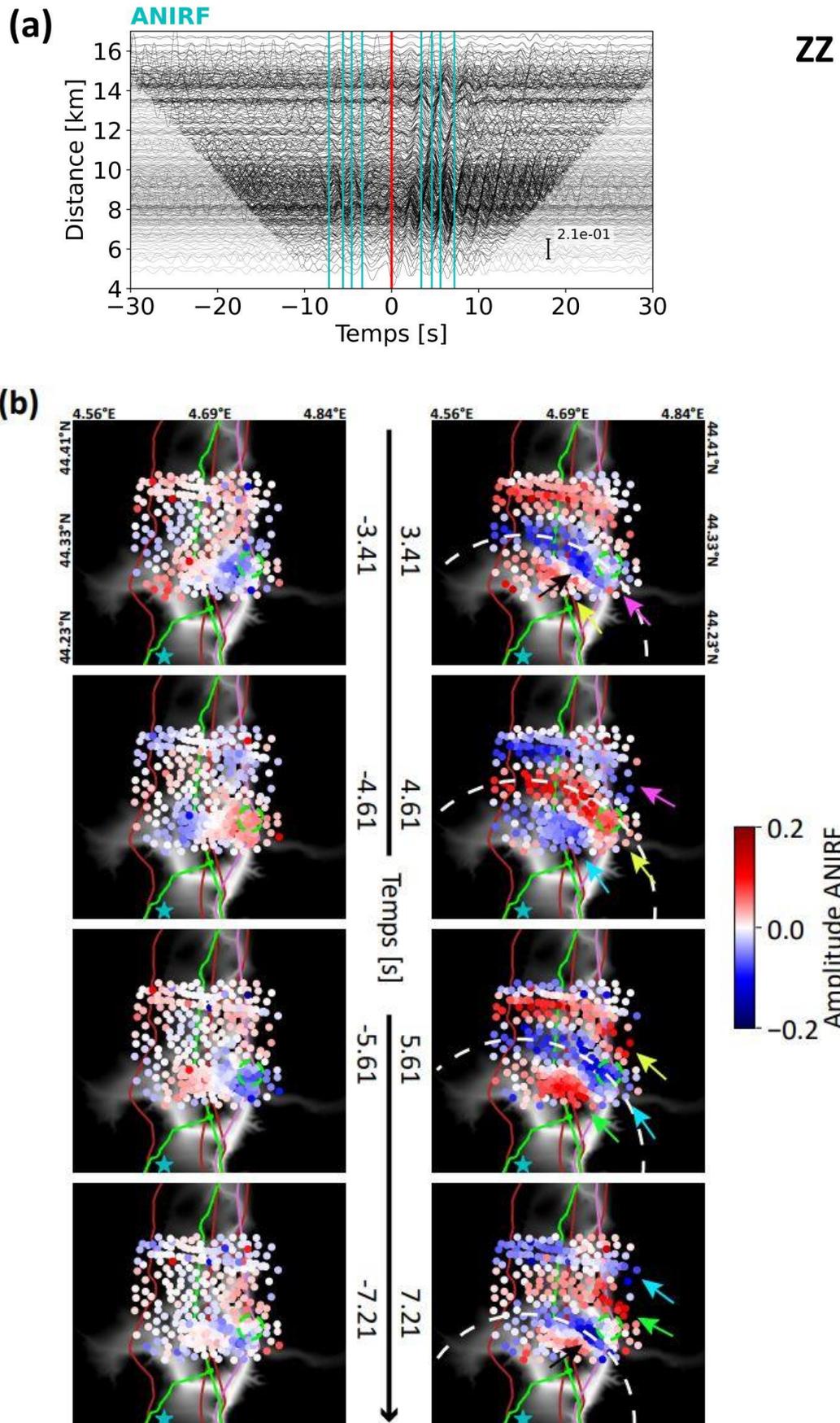


Figure 5.7 : Idem que pour la figure 5.6 pour la composante ZZ.

Conclusion sur la propagation des fronts d'onde reconstruits à basse fréquence

L'étude de l'amplitude de l'ANIRF pour différents temps d'arrivée a montré la reconstruction de fronts d'onde propagatifs à travers le réseau, ce qui confirme que l'ANIRF est capable d'extraire la cohérence du champ de bruit à basse fréquence (< 1 Hz). Le suivi des fronts d'onde révèle l'interaction avec la structure géologique. Lors de l'entrée des ondes dans la vallée, le front d'onde se distord du fait de la diminution de la vitesse de propagation dans les zones sédimentaires. L'amplitude de l'ANIRF est également plus élevée dans ces zones (canyon de l'Ardèche et vallée du Rhône). Elle est maximale lorsque le front d'onde atteint les bords de la structure, ce qui peut être le signe que l'ANIRF capture bien les effets 2D/3D associés à sa géométrie.

L'utilisation de 3 sources virtuelles situées tout autour de la vallée a permis d'observer des différences spatiales d'amplitude dépendamment de la direction d'incidence du champ entrant. L'amplitude de l'ANIRF est généralement maximale lorsque le front d'onde atteint les bords de la structure qui lui sont à peu près parallèles (et donc à peu près perpendiculaires à la direction de propagation). Ainsi lorsque la propagation se fait sur un axe NO-SE (comme c'est le cas en utilisant V2 comme source virtuelle), l'amplitude de l'ANIRF est maximale au niveau du bord est de la vallée du Rhône. En revanche, lorsque la propagation se fait sur un axe NE-SO (comme c'est le cas en utilisant ADHE ou V4 comme source virtuelle), l'amplitude de l'ANIRF est maximale au niveau du bord est de l'île de Lapalud. Simuler numériquement la propagation des ondes pourrait aider à mieux comprendre leur interaction avec les bords de la structure dépendamment de la direction d'incidence du champ entrant.

5.2.2 Amplitude de l'ANIRF à l'échelle du réseau à basse fréquence

Les observations présentées en section 5.2.1 ont été faites à partir de l'étude de l'amplitude de l'ANIRF pour des temps d'arrivée particuliers. Néanmoins, le signal reconstruit n'est pas impulsionnel, mais concerne une certaine plage de temps, qui s'étale naturellement en raison de la dispersion des ondes de surface. Ainsi, en vue des calculs d'amplification (partie 5.4), il est peut-être préférable de mesurer l'amplitude de l'ANIRF en considérant l'ensemble de la partie cohérente du signal reconstruit afin d'intégrer l'ensemble de l'information.

Dans cette section, l'amplitude de l'ANIRF est mesurée en intégrant l'ensemble de la partie considérée comme cohérente du signal reconstruit à chaque paire. Le but est d'observer l'amplitude résultante à travers le réseau en amont de l'évaluation de l'amplification (partie 5.4) qui sera basée sur les mêmes calculs. Pour contraindre la plage de temps associée à la partie cohérente du signal à chaque paire (au moins en première approximation)¹¹, nous avons simplement considéré la vitesse de groupe¹² minimale observée (entre 0.35 et 0.9 Hz) à travers le réseau des ondes de surface reconstruites et la distance source virtuelle-récepteur. Le spectre d'amplitude de la partie considérée comme cohérente¹³ a ensuite été calculé dans la bande [0.35 - 0.9] Hz. Pour observer l'amplitude de l'ANIRF à travers le réseau, la médiane des spectres d'amplitude a été cartographiée (Figure 5.8). Les sources virtuelles utilisées sont les mêmes qu'en section 5.2.1 (V2, panneau a ; ADHE, panneau b ; V4, panneau c). Les colonnes gauche et droite représentent respectivement les composantes TT et ZZ.

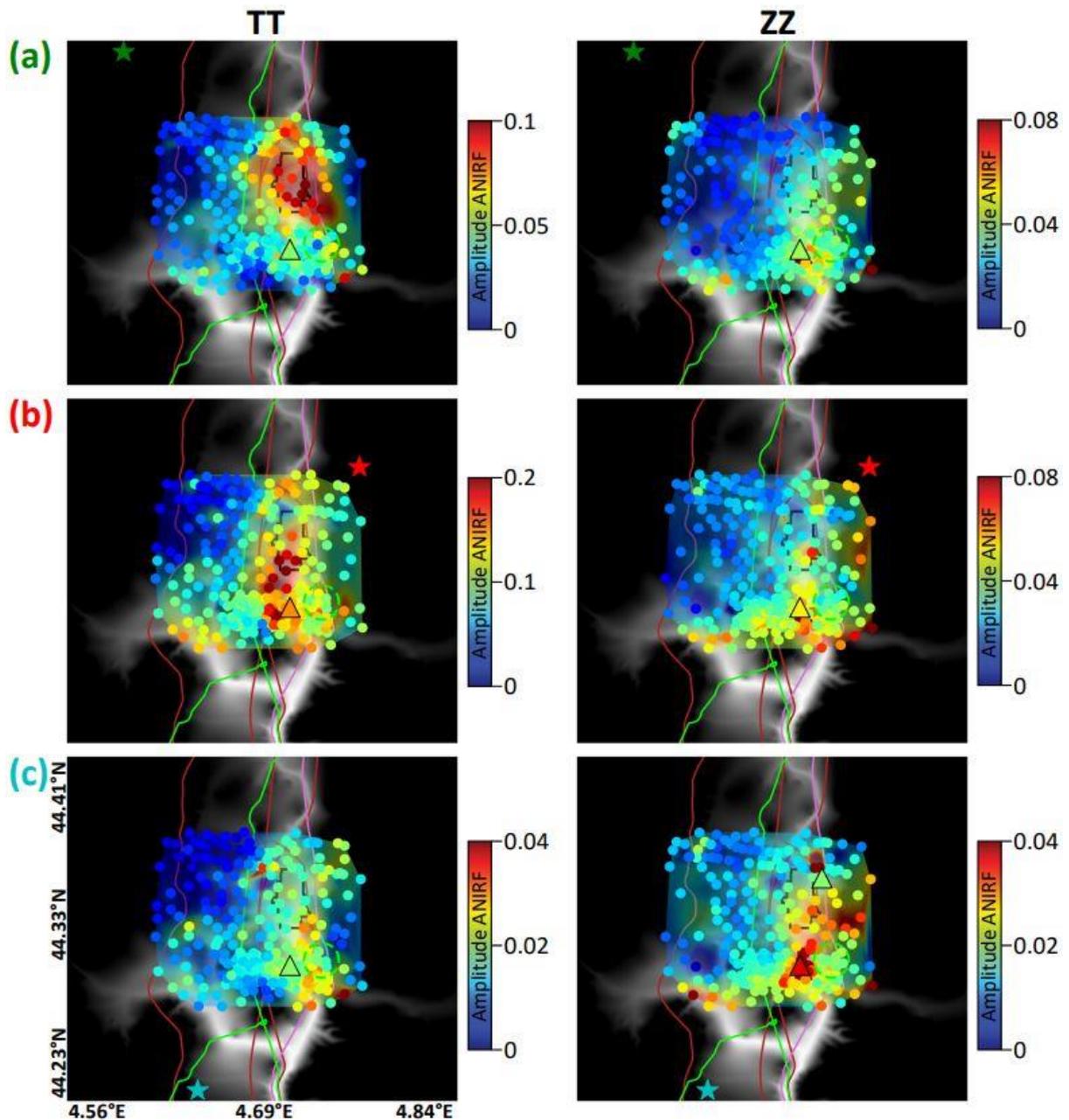
Une observation générale permet globalement de confirmer l'analyse faite sur la propagation des fronts d'onde : lorsque la propagation se fait sur un axe NO-SE (e.g. V2 comme source virtuelle, panneau a), l'amplitude de l'ANIRF est élevée (composante ZZ), voir maximale (composante TT) au niveau du bord est de la vallée du Rhône. En revanche, lorsque la propagation se fait sur un axe NE-SO (ADHE ou V4 comme source virtuelle, respectivement panneau b et c), l'amplitude de l'ANIRF est maximale au niveau du bord est de l'île de Lapalud, à l'exception toutefois de la composante TT

¹¹ Pour chaque paire, la plage de temps associée à la partie cohérente du signal reconstruit est fonction de la distance source virtuelle-récepteur, de la vitesse de propagation des ondes le long du trajet et de l'angle formé par l'azimut de la paire avec la position des différentes sources de bruit sismique.

¹² Les vitesses de groupe des ondes de surface ont été obtenues de Lavoué *et al.* (soumis).

¹³ La partie considérée comme cohérente de l'ANIRF à chaque paire a été noircie sur les sections temporelles (panneau a des figures 5.2 à 5.7).

lorsque V4 est la source virtuelle (panneau c, colonne de gauche). Il est difficile de proposer une explication sur la base de ces observations. On peut toutefois émettre l'hypothèse d'un piégeage d'énergie plus important lorsque les ondes atteignent un bord à peu près perpendiculaire à la direction de propagation. Encore une fois, simuler numériquement la propagation des ondes depuis chaque source virtuelle pourrait aider à mieux comprendre ces résultats. Notons l'amplitude importante de l'ANIRF sur le bord est du réseau pour la composante ZZ (colonne de droite), quelle que soit la source virtuelle. Ceci est surprenant puisque la zone est en-dehors de la vallée profonde du Rhône. En réalité, des unités sédimentaires peu profondes, non visibles sur la carte des profondeurs, surplombent cette région (Lavoué *et al.*, soumis), ce qui peut expliquer l'observation. Enfin, on remarque que les amplitudes sont faibles, et de surcroît très différentes dépendamment de la source virtuelle. Or, dans les conditions adéquates favorisant la convergence de l'ANIRF vers la réponse du milieu, l'amplitude absolue de l'ANIRF devrait simplement correspondre à la variation d'amplitude sur le trajet source virtuelle-récepteur, c-à-d ici, à l'amplification liée aux effets de site dans la vallée. On perçoit ici que ces conditions n'étant pas réunies, l'amplitude absolue de l'ANIRF ne correspond pas directement à l'amplification site/rocher recherchée et que seule l'information d'amplitude relative à travers le réseau est utilisable. Afin de poursuivre notre objectif d'utiliser l'ANIRF pour l'évaluation de l'amplification site/rocher (partie 5.4), nous établirons par la suite une étape de calibration au rocher de l'amplitude relative à travers le réseau.



Légende

- Contour TNS
 - Barrage
 - Voies ferrées
 - A7
 - N7
- 590 295 0
 Profondeur [m]

Figure 5.8 : Amplitude de l'ANIRF mesurée entre 0.35 et 0.9 Hz en intégrant l'ensemble de la partie considérée comme cohérente du signal reconstruit à chaque paire. Les sources virtuelles utilisées sont les mêmes qu'en section 5.2.1 (V2, panneau a ; ADHE, panneau b ; V4, panneau c). Les colonnes gauche et droite représentent respectivement les composantes TT et ZZ.

5.3 Amplitude de l'ANIRF à l'échelle du réseau à haute fréquence (entre 3 et 6 Hz)

La partie 5.2 a démontré la capacité de l'ANIRF à extraire la cohérence du champ de bruit sismique pour reconstruire un signal déterministe qui se propage à travers la structure géologique et qui en porte ces effets. Cette observation est tout du moins valable à basse fréquence (< 1 Hz). Cette partie traite des résultats de l'ANIRF à plus haute fréquence (entre 3 et 6 Hz). Un travail similaire à celui présenté en partie 5.2 a été conduit.

5.3.1 Amplitude de l'ANIRF à l'échelle du réseau pour des temps d'arrivée particuliers

Sur les figures 5.9 et 5.10 (respectivement TT et ZZ), l'amplitude de l'ANIRF a été cartographiée à différents temps d'arrivée pour la source virtuelle ADHE. Contrairement aux observations faites à basse fréquence (section 5.2.1), les cartes ne montrent aucun front d'onde se propageant à travers le réseau. Cela reflète la difficulté d'extraire la cohérence de phase du champ de bruit à ces fréquences. En d'autres termes, l'ANIRF ne permet pas de reconstruire un signal propagatif clair, mais est dominée par l'information d'amplitude contrôlée par l'énergie à la source virtuelle et au récepteur. Cette information d'amplitude est de surcroît biaisée par les sources locales : les zones où l'amplitude est maximale sont centrées autour du TNS et du barrage.

En comparant la partie causale (colonne de droite) et anti-causale (colonne de gauche), on s'aperçoit néanmoins que l'impact de ces sources semble atténué sur la partie causale. Dans les faits, les sources anthropiques majeures dans la zone sont plus proches du récepteur que de la source virtuelle pour la majorité des paires. Par conséquent, l'énergie émise par ces sources est principalement retranscrite sur la partie anti-causale de l'ANIRF (voir l'introduction du chapitre 4). On entrevoit ici ce qui peut être un intérêt majeur de l'utilisation de l'ANIRF par rapport à un simple rapport d'amplitude (SSRn) dans le but d'évaluer l'amplification au site : les différentes contributions des sources sont potentiellement séparables en fonction des temps d'arrivée. Ainsi, sélectionner les parties de l'ANIRF les moins impactées par les sources locales majeures pourrait aider à limiter le biais sur l'estimation de l'amplification.

Précisons qu'à travers notre étude, des résultats similaires ont été trouvés quelle que soit la source virtuelle et la composante (TT, RR ou ZZ) utilisée. A titre

d'exemple, les résultats pour les sources virtuelles V2 et V4 sont présentés en annexe (Figures B1 à B4).

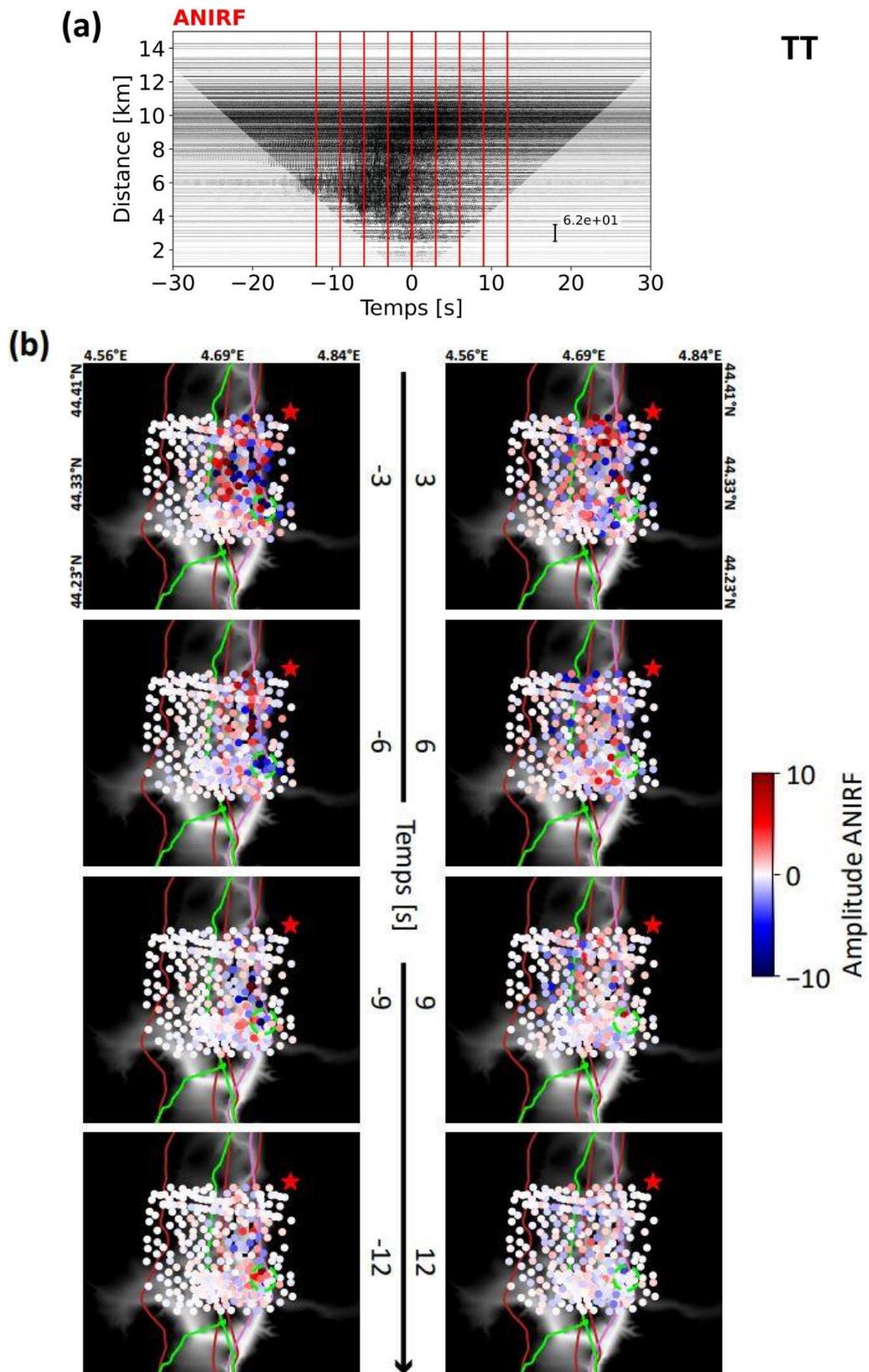


Figure 5.9 : Amplitude de l'ANIRF cartographiée à différents temps d'arrivée pour la partie causale et anti-causale (respectivement colonne de droite et colonne de gauche) en utilisant ADHE comme source virtuelle. L'ANIRF est ici filtrée entre 3 et 6 Hz. Les observations sont présentées pour la composante TT.

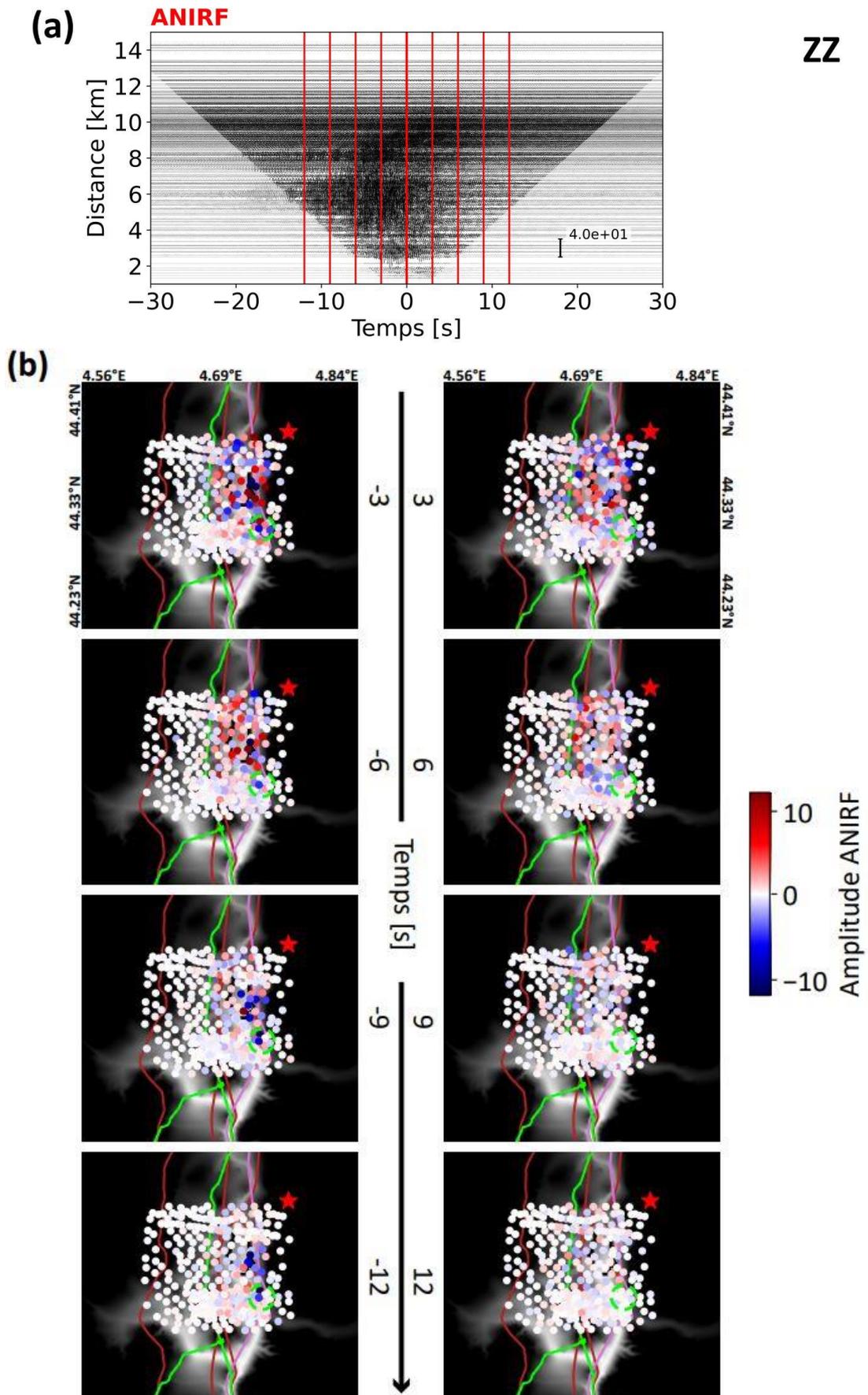


Figure 5.10 : Idem que pour la figure 5.9 pour la composante ZZ.

5.3.2 Amplitude de l'ANIRF à l'échelle du réseau à haute fréquence

Pour chacune des 3 sources virtuelles, le spectre d'amplitude de l'ANIRF a été calculé pour chaque paire dans la bande [3 - 6] Hz selon 2 cas. Dans les 2 cas, la partie de l'ANIRF considérée à chaque paire a été délimitée par la vitesse de groupe minimale observée à travers le réseau à 3 Hz des ondes de surface reconstruites (la vitesse de groupe n'a pas pu être contrainte au-delà de 3 Hz) et la distance source virtuelle-récepteur. Dans le premier cas, les parties causale et anti-causale ont été considérées. Dans le second cas, seule la partie causale a été considérée. La médiane des spectres d'amplitude a ensuite été cartographiée selon ces 2 cas pour chacune des 3 sources virtuelles (V2, panneau a ; ADHE, panneau b ; V4, panneau c) pour la composante TT (Figure 5.11) et ZZ (Figure 5.12). La colonne de gauche dénote le premier cas (parties causale et anti-causale considérées). La colonne de droite dénote le second cas (partie causale considérée uniquement).

Une vue générale permet de constater, comme attendu, la domination des sources anthropiques locales sur l'amplitude estimée. Ceci est vrai quel que soit le cas, la source virtuelle, ou la composante considérée. Sur les 2 composantes, le TNS (en noir), le barrage (cercle vert) et l'autoroute A7 (en violet) semblent être les sources les plus impactantes. Sur la composante ZZ (Figure 5.12), on peut également percevoir la signature N-S de la LGV (en marron) et de la nationale N7 (en vert). Certaines zones d'amplitude élevée illuminent tout de même des régions relativement distantes de ces sources (flèches blanches). On relève notamment le sud de la vallée du Rhône lorsque ADHE est la source virtuelle (panneau b, Figures 5.11 et 5.12), le canyon de l'Ardèche lorsque V4 est la source virtuelle (panneau c, Figures 5.11 et 5.12) et la zone de confluence au nord du canyon de l'Ardèche lorsque V2 est la source virtuelle (composante ZZ, panneau a, Figure 5.12). Ces régions correspondent à des zones sédimentaires bénéficiant d'un éclairage favorable par la source virtuelle. Ceci illustre la possibilité, dans ce cas, d'extraire la réponse de la structure à des fréquences relativement hautes en utilisant l'ANIRF. Néanmoins, la réponse de la structure dans son ensemble reste biaisée par la signature des sources anthropiques locales.

On constate toutefois une certaine atténuation de l'influence des sources anthropiques majeures lorsque seule la partie causale est considérée (colonne de droite), ce qui concorde avec les observations de la partie 5.3.1. Pour quantifier cette atténuation, la différence relative d'amplitude a été calculée à chaque node (eq. 3), puis moyennée selon 5 zones différentes. Quatre zones incluent des sources anthropiques majeures, comprenant respectivement les nodes situés sur le périmètre du TNS, les nodes situés à moins de 250 m de l'A7 et de la LGV, les nodes situés à moins

d'1 km du barrage. Une zone de référence comprend tous les nodes du réseau qui ne sont pas dans les 4 zones précédemment citées. Les résultats sont présentés pour chaque source virtuelle au bas des figures 5.11 et 5.12, les panneaux (d), (e) et (f) correspondant respectivement à V2, ADHE et V4. Par rapport à la zone référence, la différence relative moyenne est plus élevée pour les 4 zones correspondant aux sources anthropiques majeures, quelle que soit la source virtuelle et la composante (à l'exception de l'A7 pour V4 et la composante ZZ ; panneau f, Figure 5.12). Globalement, la différence relative moyenne est importante concernant le TNS, l'A7 et le barrage. Lorsque ADHE est la source virtuelle par exemple, la différence relative moyenne observée pour ces 3 zones est plus de 2 fois supérieure à celle observée pour la zone référence. Ces résultats confirment de manière quantitative l'atténuation de l'influence des sources anthropiques majeures (e.g. TNS, A7, barrage de Donzère-Mondragon) lorsque seule la partie causale de l'ANIRF est considérée dans les calculs.

$$Différence\ relative = \frac{ANIRF_{Causale\ et\ anti-causale} - ANIRF_{Causale}}{ANIRF_{Causale\ et\ anti-causale}} \times 100 . \quad (3)$$

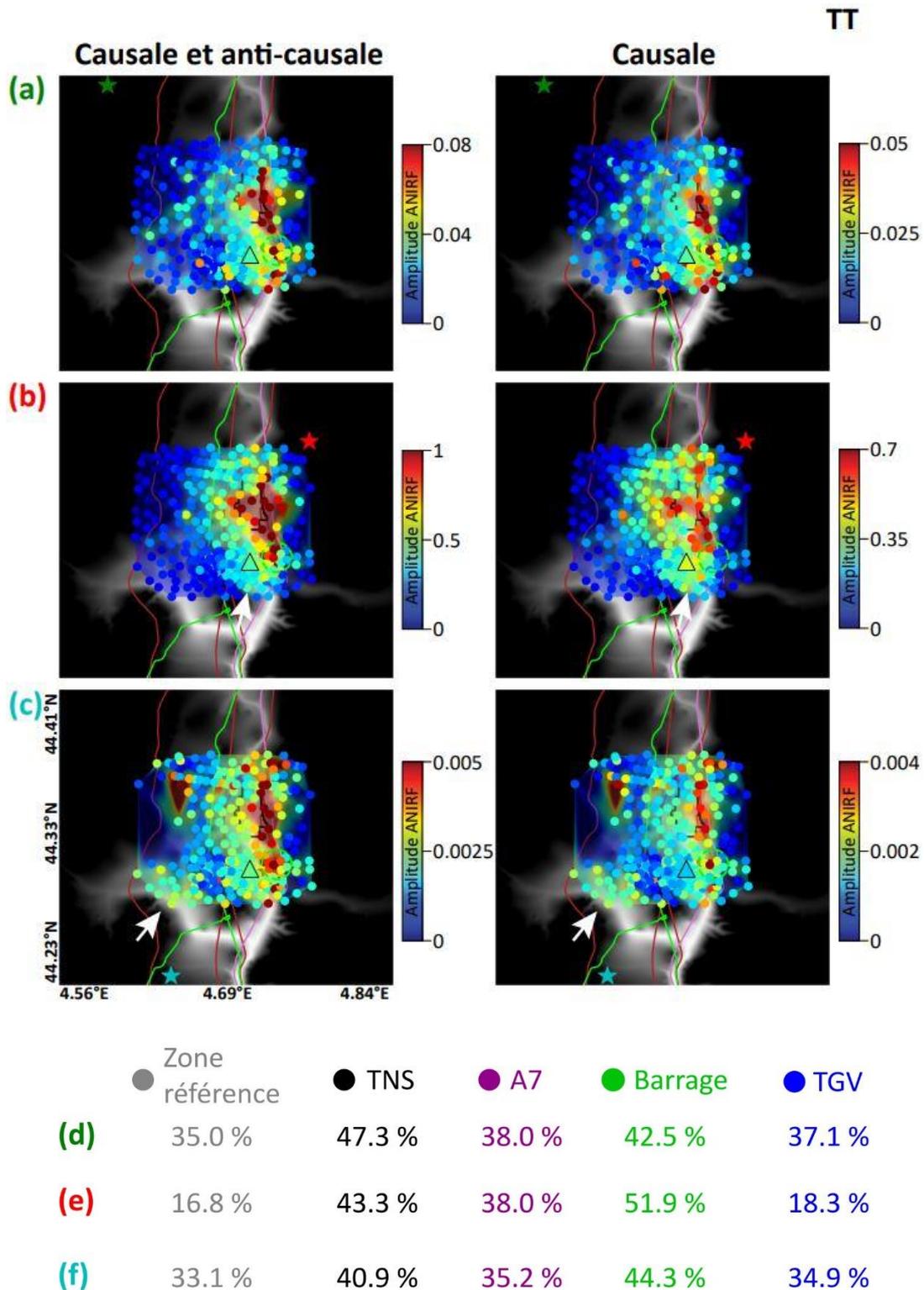


Figure 5.11 : Les panneaux (a)-(d), (b)-(e) et (c)-(f) correspondent respectivement aux sources virtuelles V2, ADHE et V4. (a), (b) et (c) : amplitude de l'ANIRF mesurée entre 3 et 6 Hz en considérant la partie causale et anti-causale (colonne de gauche) ou uniquement la partie causale (colonne de droite). (d), (e) et (f) : différence relative d'amplitude entre les deux cas, moyennée sur un ensemble de nodes compris dans 5 zones différentes. Les observations sont faites pour la composante TT.

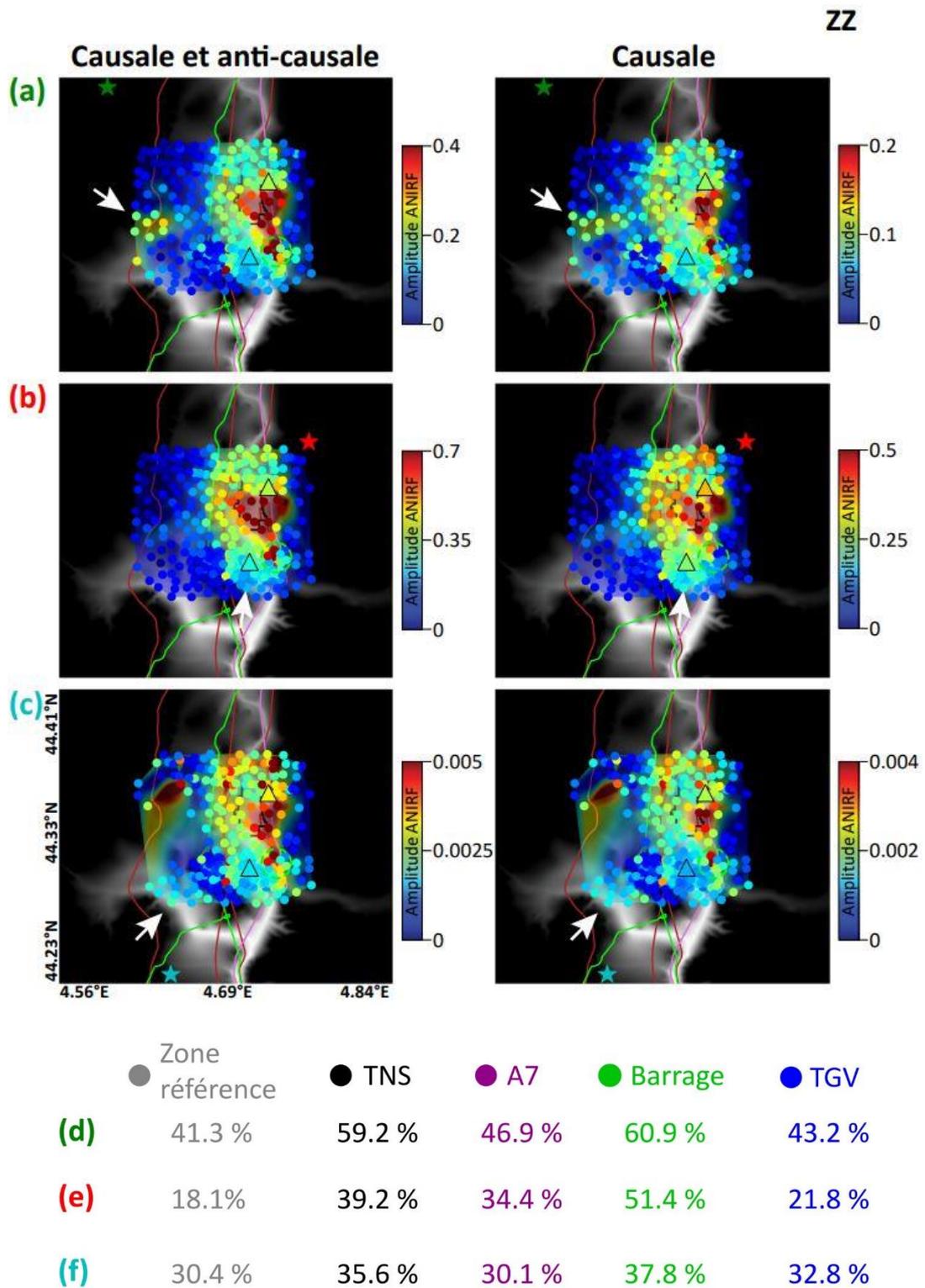


Figure 5.12 : Idem que pour la figure 5.11 pour la composante ZZ.

5.4 Evaluation de l'amplification site/rocher à partir de l'ANIRF

L'utilisation de 3 sources virtuelles situées autour de la zone d'étude a mis en évidence la variabilité de la réponse de la vallée, du moins à basse fréquence, dépendamment de la direction du champ incident. Ceci montre l'importance du choix de la position de la source virtuelle si le but est d'anticiper l'amplification site/rocher pour un séisme spécifique (scénario de rupture). Dans cette étude, le but est au contraire d'anticiper la réponse globale de la vallée à un champ d'ondes pouvant provenir de toutes les directions. Ainsi, la considération de plusieurs sources virtuelles distribuées tout autour de la vallée prévaut. L'amplification site/rocher associée à la vallée est ici évaluée en considérant 7 sources virtuelles situées autour de la zone d'étude (Figure 5.1). Une procédure en 4 étapes a été appliquée. L'idée de la procédure est d'abord d'extraire une mesure d'amplitude relative à travers le réseau à partir de l'ANIRF, puis de « calibrer » cette mesure site-à-site par un SSR site/rocher pour évaluer l'amplification site/rocher en tout point du réseau.

5.4.1 Présentation de la procédure de calcul

Chacune des étapes de la procédure introduite est décrite sur la figure 5.13 :

- (1) l'ANIRF (eq. 2) est calculée entre 7 sources virtuelles (SV) et chacun des 400 nodes (Si) du réseau selon les composantes diagonales du tenseur de Green (TT, RR, ZZ). Rappelons que des fenêtres nocturnes de 30 min sans recouvrement sont ici utilisées et que l'amplitude de l'ANIRF pour chaque paire $SV-Si$ est multipliée par un facteur \sqrt{d} (où d est la distance entre le récepteur et la source virtuelle utilisée) pour corriger l'atténuation géométrique. L'ANIRF à chaque paire est ensuite filtrée dans les bandes [0.2 - 1] Hz et [1 - 10] Hz. Les fonctions ANIRF dont le rapport signal à bruit (SNR) est inférieur à 3 ne sont pas considérées ;
- (2) à basse fréquence (< 1 Hz), la partie de l'ANIRF sélectionnée à chaque paire est fixée par la vitesse de groupe minimale observée à travers le réseau entre 0.2 et 1 Hz des ondes de surface reconstruites et par la distance source virtuelle-récepteur. Les parties causale et anti-causale sont considérées. A haute fréquence (> 1 Hz), la partie de l'ANIRF sélectionnée à chaque paire est fixée par la vitesse de groupe minimale observée à travers le réseau entre 1 et 3 Hz (la vitesse de groupe n'a pas pu être contrainte au-delà de 3 Hz) et par la distance source virtuelle-récepteur. Dans un premier cas, les parties causale et anti-

causale sont considérées. Dans un second cas, seule la partie causale est considérée ;

- (3) pour chaque paire $SV-S_i$, le spectre d'amplitude de la partie sélectionnée de l'ANIRF est calculé. Puisque seule l'amplitude relative à travers le réseau est utilisable (voir section 5.2.2), le spectre d'amplitude de chaque paire $SV-S_i$ est normalisé par celui de la paire $SV-BOLL$ (SV étant la source virtuelle correspondante et $BOLL$ le capteur de référence au site). A ce stade, nous avons estimé une amplification site/site à l'échelle du réseau pour chacune des 7 sources virtuelles et des composantes TT, RR et ZZ ;
- (4) afin d'estimer une amplification site/site qui serait la signature d'une réponse globale de la vallée à un champ d'ondes pouvant provenir de toutes les directions, la médiane sur les 7 sources virtuelles de l'amplification site/site est calculée à chaque node S_i . Concernant les composantes horizontales (TT et RR), la moyenne quadratique est considérée. L'amplification site/site médiane à chaque node est ensuite multipliée par une mesure SSR site/rocher (ici le $SSR_{BOLL/A04}$, obtenu de Lavoué *et al.*, soumis) pour évaluer l'amplification site/rocher en tout point du réseau. Cette démarche hybride combinant données de bruit sismique et de séismes s'inspire directement du SSRh (Perron *et al.* 2018).

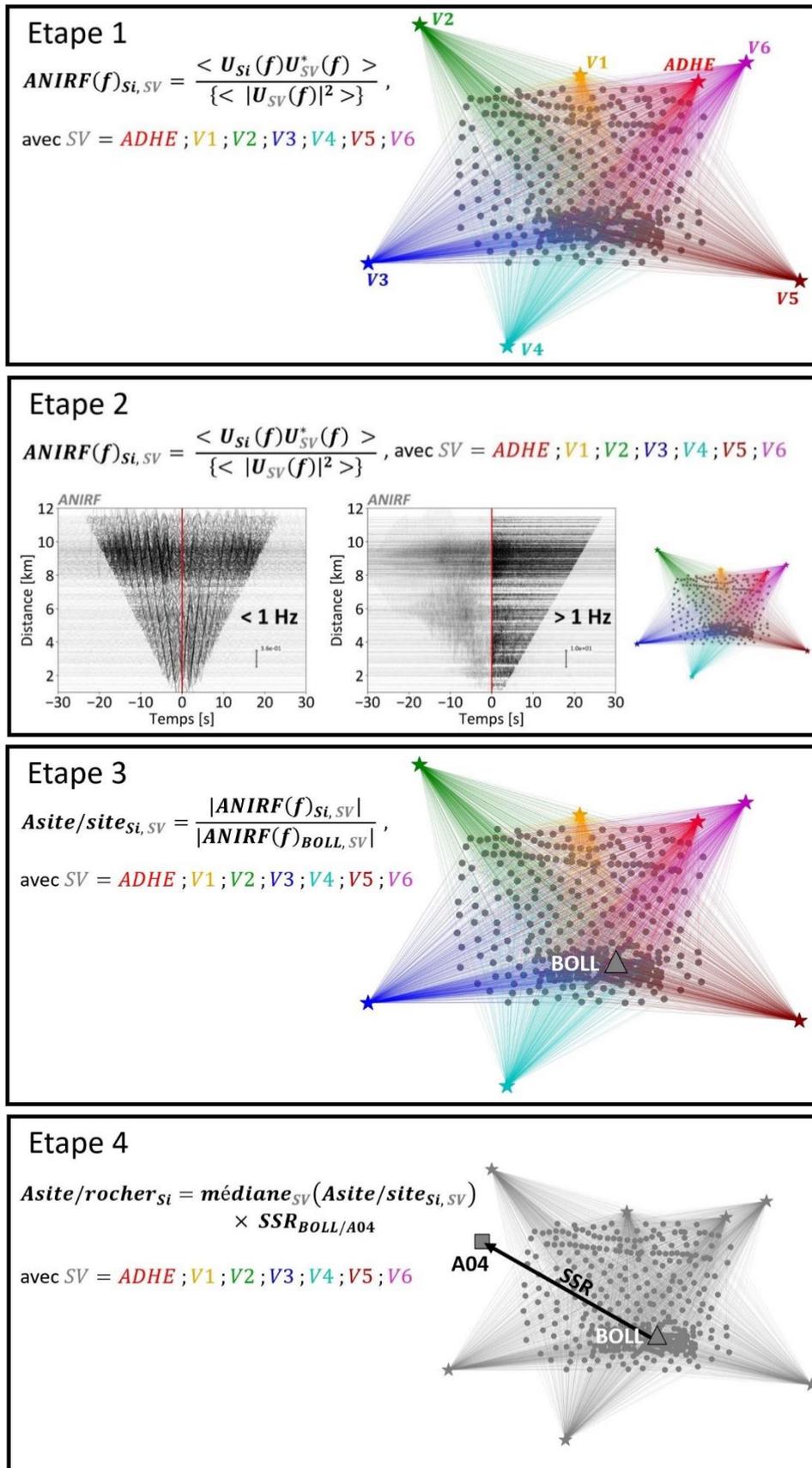


Figure 5.13 : Illustration des 4 étapes de la procédure de calcul de l'amplification site/rocher associée à la vallée.

5.4.2 Présentation des résultats

Les résultats sont présentés sur les figures 5.14 à 5.17 pour la composante horizontale (H) et les figures 5.18 à 5.21 pour la composante verticale (Z). Ils sont comparés aux mesures SSR obtenues par Lavoué *et al.* (soumis) sur le réseau de stations large-bande (présenté dans le chapitre 2). La structure des figures est identique :

- le panneau (a) présente la comparaison entre le SSR mesuré pour 7 stations large-bande spécifiques et le facteur d'amplification obtenu par l'ANIRF au node le plus proche. La distance inter-station est indiquée dans le coin supérieur droit. Le nom de la station large-bande servant à la comparaison est noté juste au-dessus. La courbe continue rouge représente le SSR (moyenné sur plusieurs enregistrements de séismes), les courbes en pointillés rouges représentent le SSR moyen + ou - 1 écart type, les courbes continues bleues et violettes représentent l'amplification obtenue par l'ANIRF (respectivement en considérant les parties causale et anti-causale ou la partie causale uniquement) ;
- le panneau (b) présente la carte interpolée des valeurs médianes dans la bande [0.35 – 0.9] Hz de l'amplification obtenue par l'ANIRF (réseau de nodes). La valeur médiane du SSR dans la même bande de fréquences est superposée pour chaque station large-bande (représentées par des carrés).

Au total, 28 comparaisons « fréquence à fréquence » ont été effectuées (panneau a) pour chaque composante. Entre 0.2 et 1 Hz, le facteur d'amplification obtenu par l'ANIRF (courbe bleue) est comparable¹⁴ au SSR (courbe rouge) pour environ 57 % des cas sur la composante horizontale (figures 5.14 à 5.17) et 61 % des cas sur la composante verticale (figures 5.18 à 5.21). Il s'agit de la comparaison avec A00, A01, A02, B00, B01, C00, C02, C03, C04, D00, D02, D03, D04, D05, E05 et F01 (composante horizontale) et avec A00, A01, B00, B02, C00, C02, D00, D01, D02, D03, D04, D05, E03, E05, F01, F02 et G01 (composante verticale). Dans quelques cas, l'amplification obtenue par l'ANIRF surestime assez largement le SSR (B03, C01, C05, F05, G00). Ces cas correspondent à des mesures proches des bords de la structure, où le champ de bruit est très complexe. Au-delà de 1 Hz, le facteur d'amplification obtenu par l'ANIRF est globalement plus proche du SSR lorsque seule la partie causale est considérée (courbes violettes). Ainsi, la mesure fournie par l'ANIRF est comparable au

¹⁴ L'amplification obtenue par l'ANIRF est contenue dans (ou est à minima proche de) la distribution des mesures SSR.

SSR jusqu'à environ 4 Hz dans 54 % des cas sur la composante horizontale (A00, A02, B05, C00, C02, C03, C04, C05, D01, D02, D04, D05, E03, E05, G01) et 39 % sur la composante verticale (A00, A01, C00, C02, C03, C04, D02, D04, E03, E05, G01) malgré la faible dispersion des mesures SSR à certaines stations due au manque d'enregistrements de séismes de qualité suffisante. Notons que la majorité des nodes est relativement éloigné de la station large-bande de comparaison. Dans 61 % des cas, la distance inter-station est supérieure à 200 m, ce qui peut expliquer certains biais à haute fréquence (au-delà de 3 Hz, la longueur d'onde dans les zones sédimentaires des ondes de surface reconstruites est de l'ordre de la centaine de mètres).

Entre 0.35 et 0.9 Hz, la carte interpolée de l'amplification obtenue par l'ANIRF (panneau b) est cohérente avec les mesures SSR ainsi que la structure de la vallée quelle que soit la composante, montrant une amplification plus importante à l'aplomb des zones sédimentaires profondes. Notons que l'amplification est plus élevée pour la composante horizontale (jusqu'à 8 dans le canyon de l'Ardèche et au niveau du bord est de la vallée du Rhône) et que les zones où l'amplification est maximale sont différentes entre les composantes. Pour la composante horizontale, l'amplification estimée est importante (entre 5 et 8) dans l'ensemble de la vallée du Rhône et du canyon de l'Ardèche. Pour la composante verticale, l'amplification se concentre au sud et à l'est (unités sédimentaires peu profondes) de la vallée du Rhône ainsi que dans le canyon de l'Ardèche. Des simulations numériques de la propagation du champ d'ondes depuis les différentes sources virtuelles pourraient aider à mieux comprendre ces différences d'amplification entre les composantes.

Les résultats obtenus ici montrent la capacité de l'ANIRF à extraire la réponse du site à basse fréquence et laissent entrevoir encore davantage les potentiels bénéfiques de l'utilisation de ce type d'approche pour des évaluations d'amplification même au-delà du hertz.

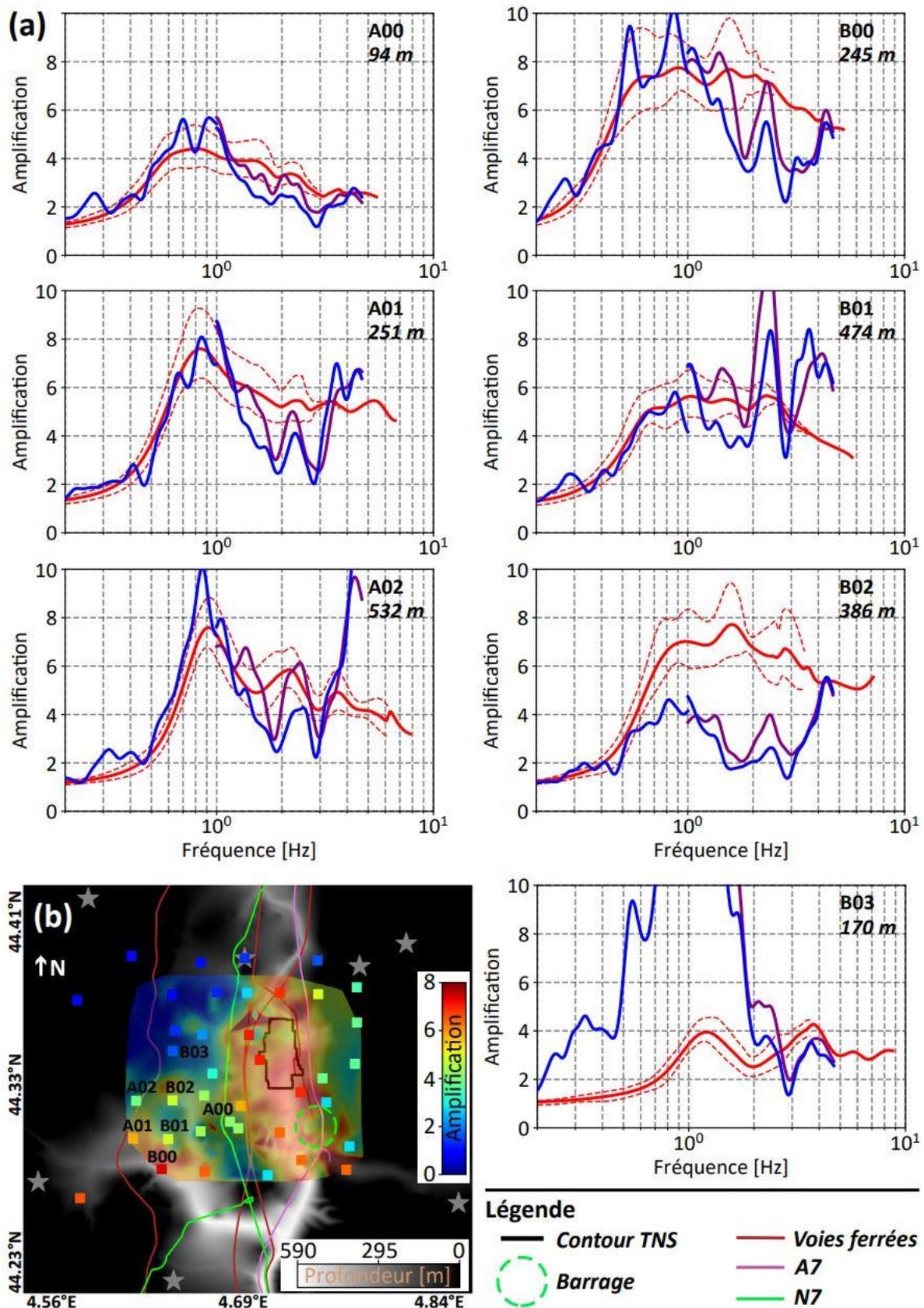


Figure 5.14 : (a) Comparaison entre le SSR mesuré pour 7 stations large-bande spécifiques (A00, A01, A02, B00, B01, B02, B03) et le facteur d'amplification obtenu par

l'ANIRF au node le plus proche. La distance inter-station est indiquée dans le coin supérieur droit. Le nom de la station large-bande servant à la comparaison est noté juste au-dessus. La courbe continue rouge représente le SSR (moyenné sur plusieurs enregistrements de séismes), les courbes en pointillés rouges représentent le SSR moyen + ou - 1 écart type, les courbes continues bleues et violettes représentent l'amplification obtenue par l'ANIRF (respectivement en considérant les parties causale et anti-causale ou la partie causale uniquement). (b) Carte interpolée des valeurs médianes dans la bande [0.35 – 0.9] Hz de l'amplification obtenue par l'ANIRF (réseau de nodes). La valeur médiane du SSR dans la même bande de fréquences est superposée pour chaque station large-bande (représentées par des carrés). Les résultats sont présentés pour la composante H.

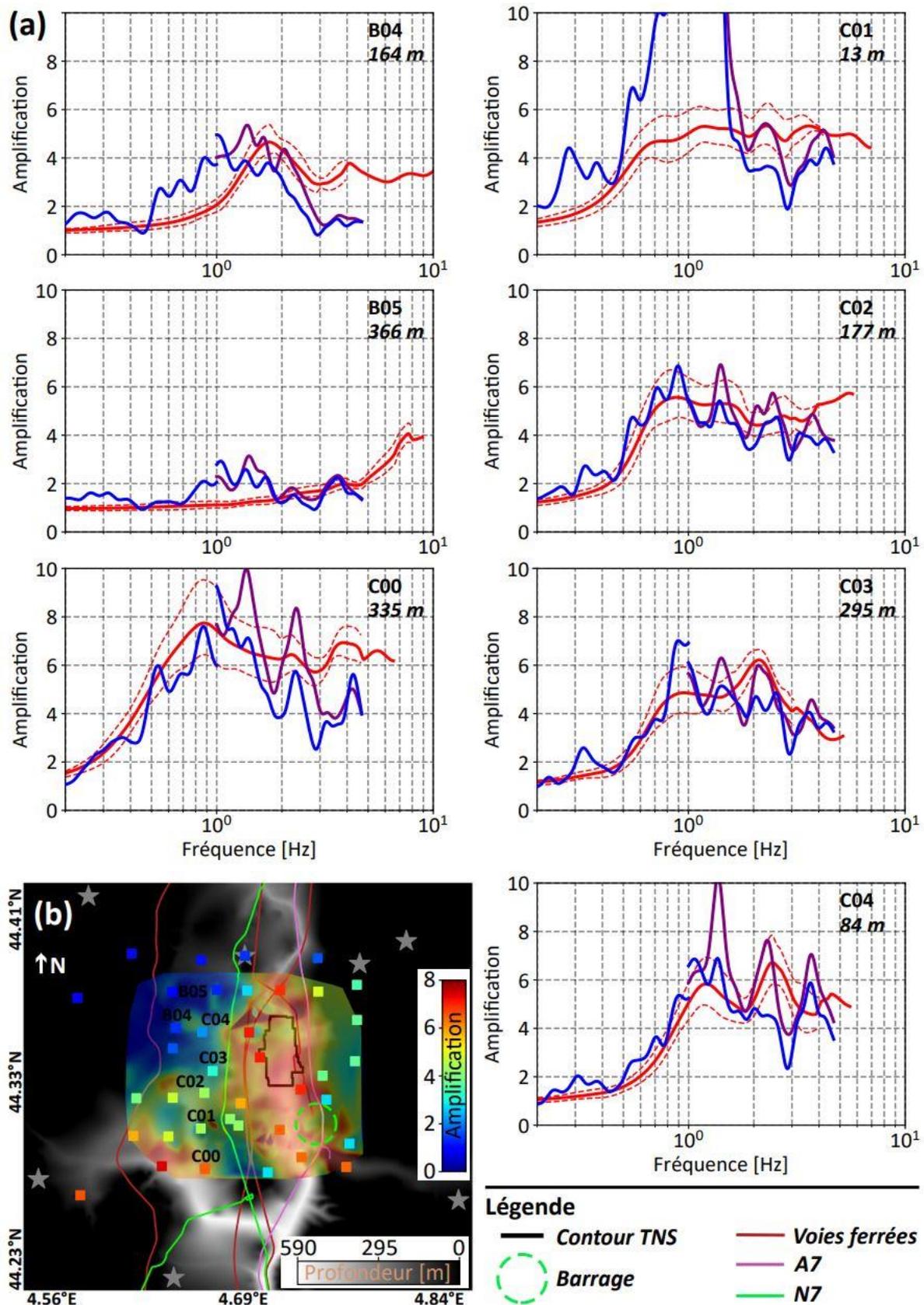


Figure 5.15 : Idem que pour la figure 5.14 pour les stations large-bande B04, B05, C00, C01, C02, C03 et C04.

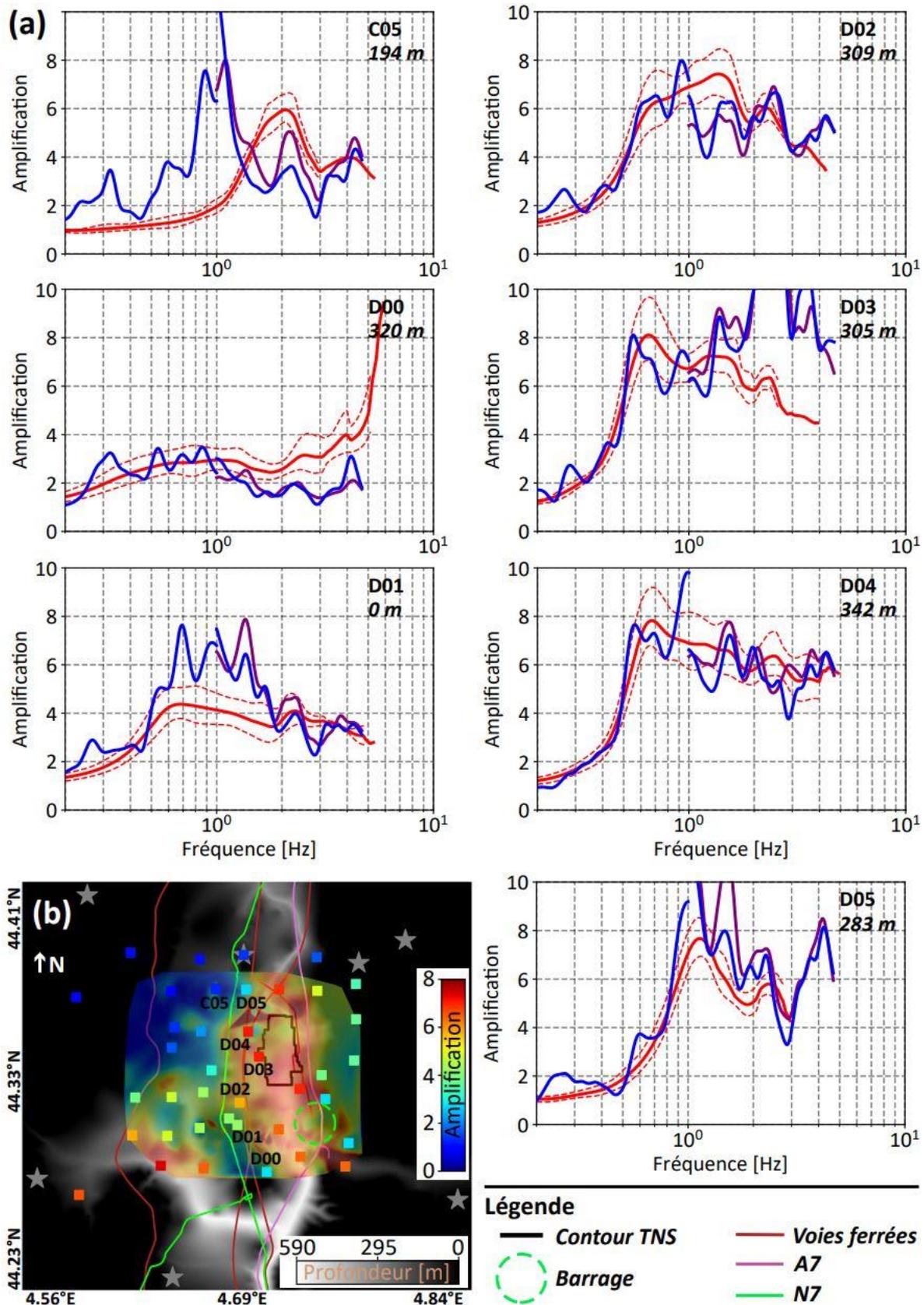


Figure 5.16 : Idem que pour la figure 5.14 pour les stations large-bande C05, D00, D01, D02, D03, D04 et D05.

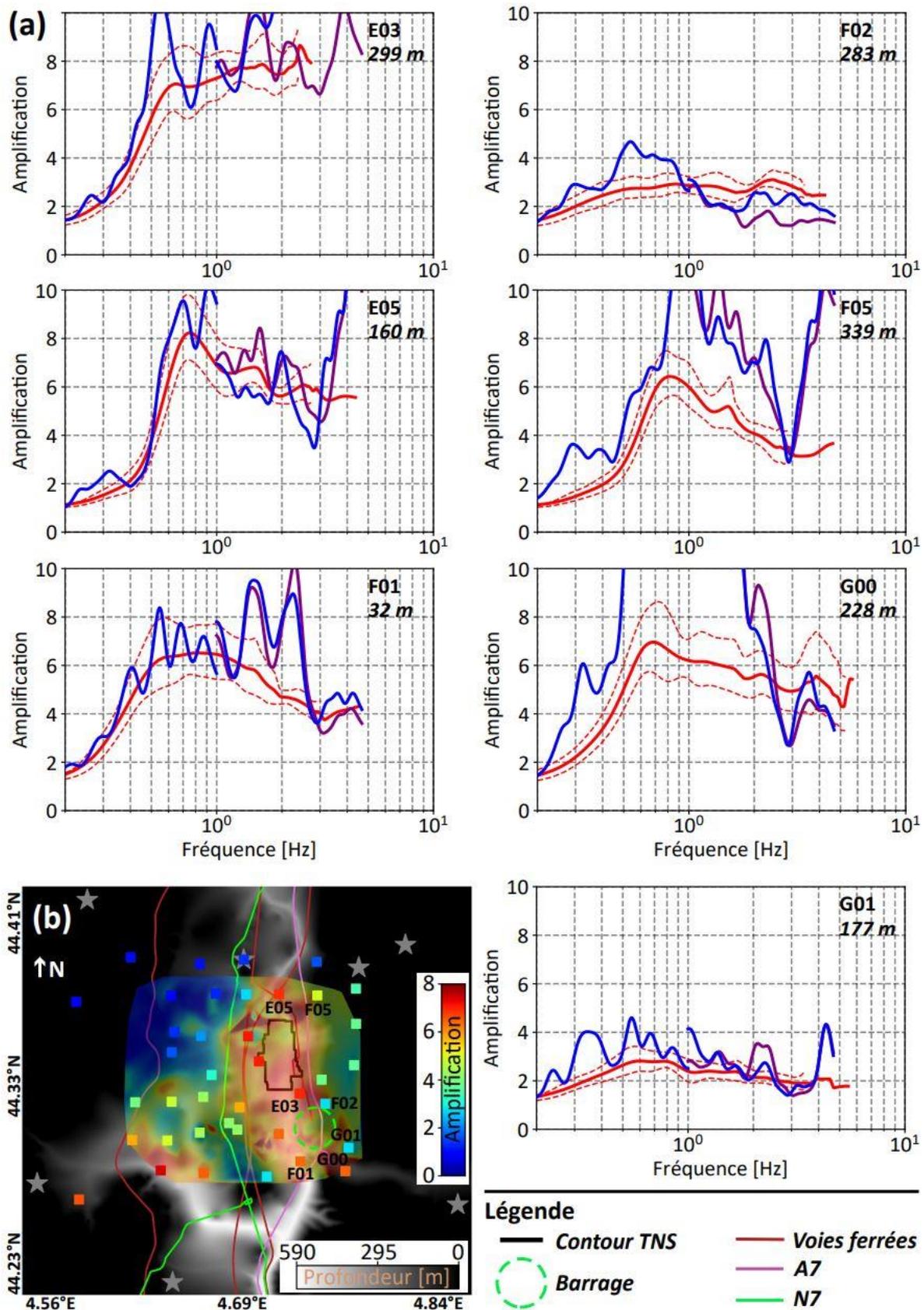


Figure 5.17 : Idem que pour la figure 5.14 pour les stations large-bande E03, E05, F01, F02, F05, G00 et G01.

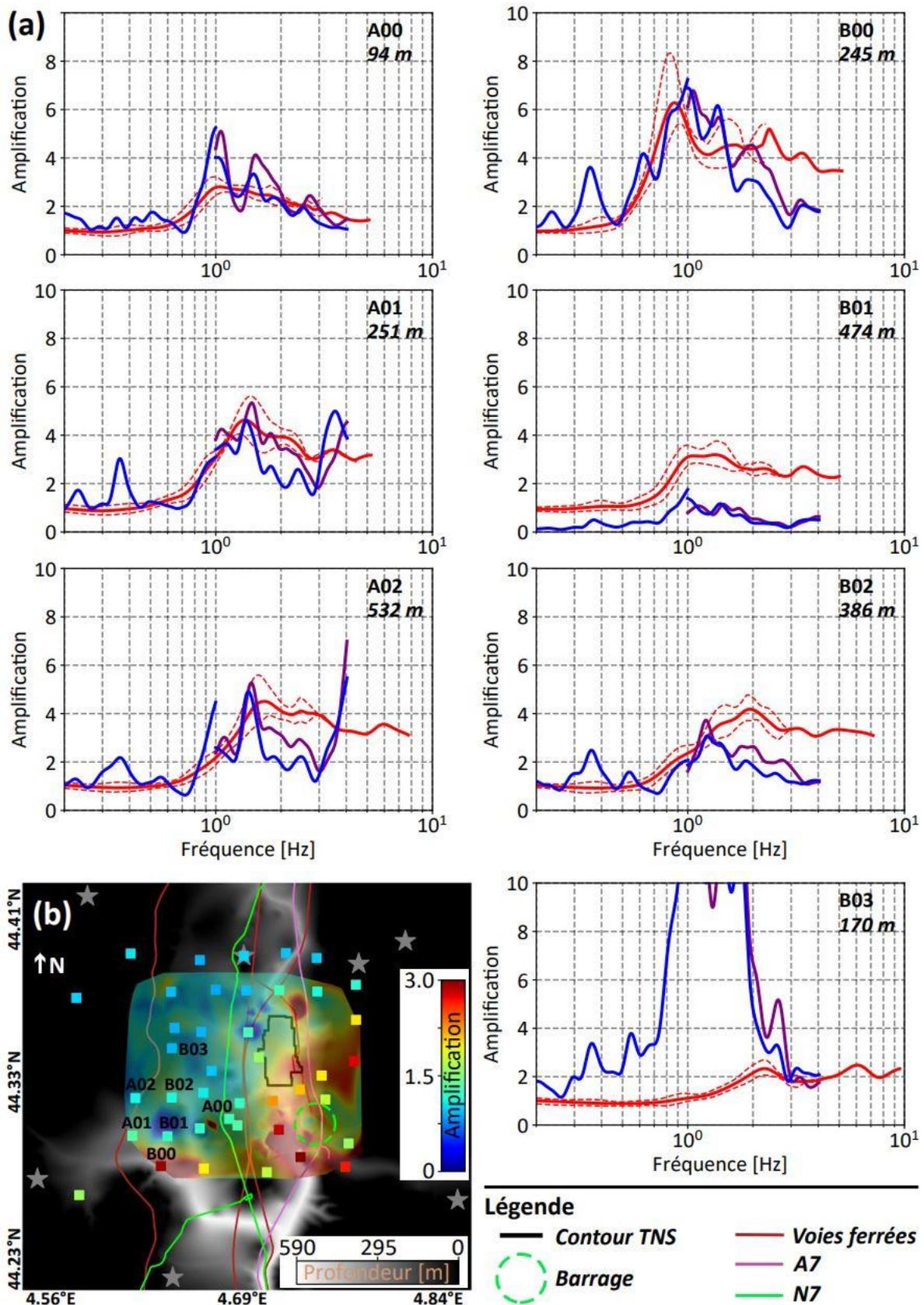


Figure 5.18 : Idem que pour la figure 5.14 pour la composante Z.

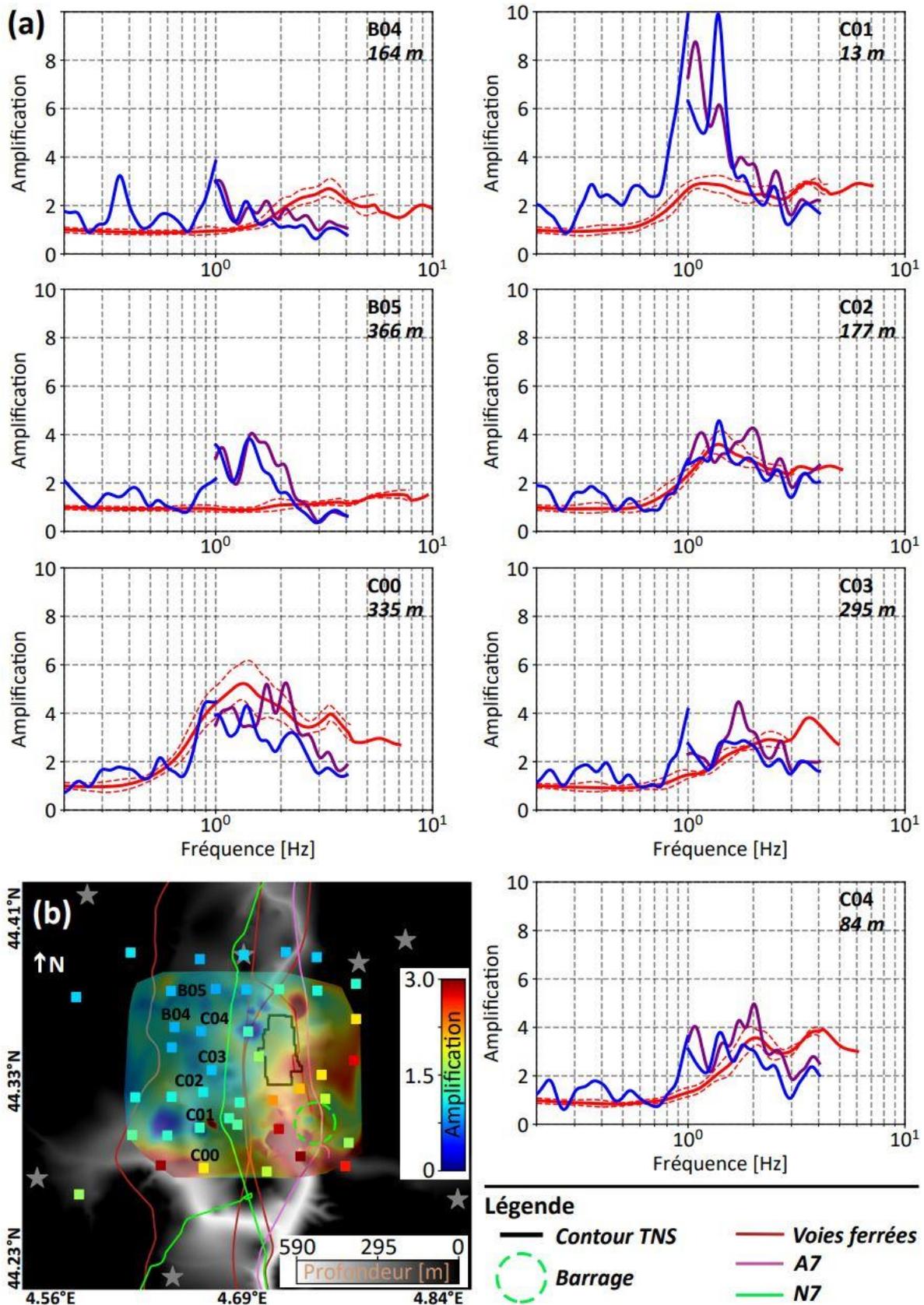


Figure 5.19 : Idem que pour la figure 5.15 pour la composante Z.

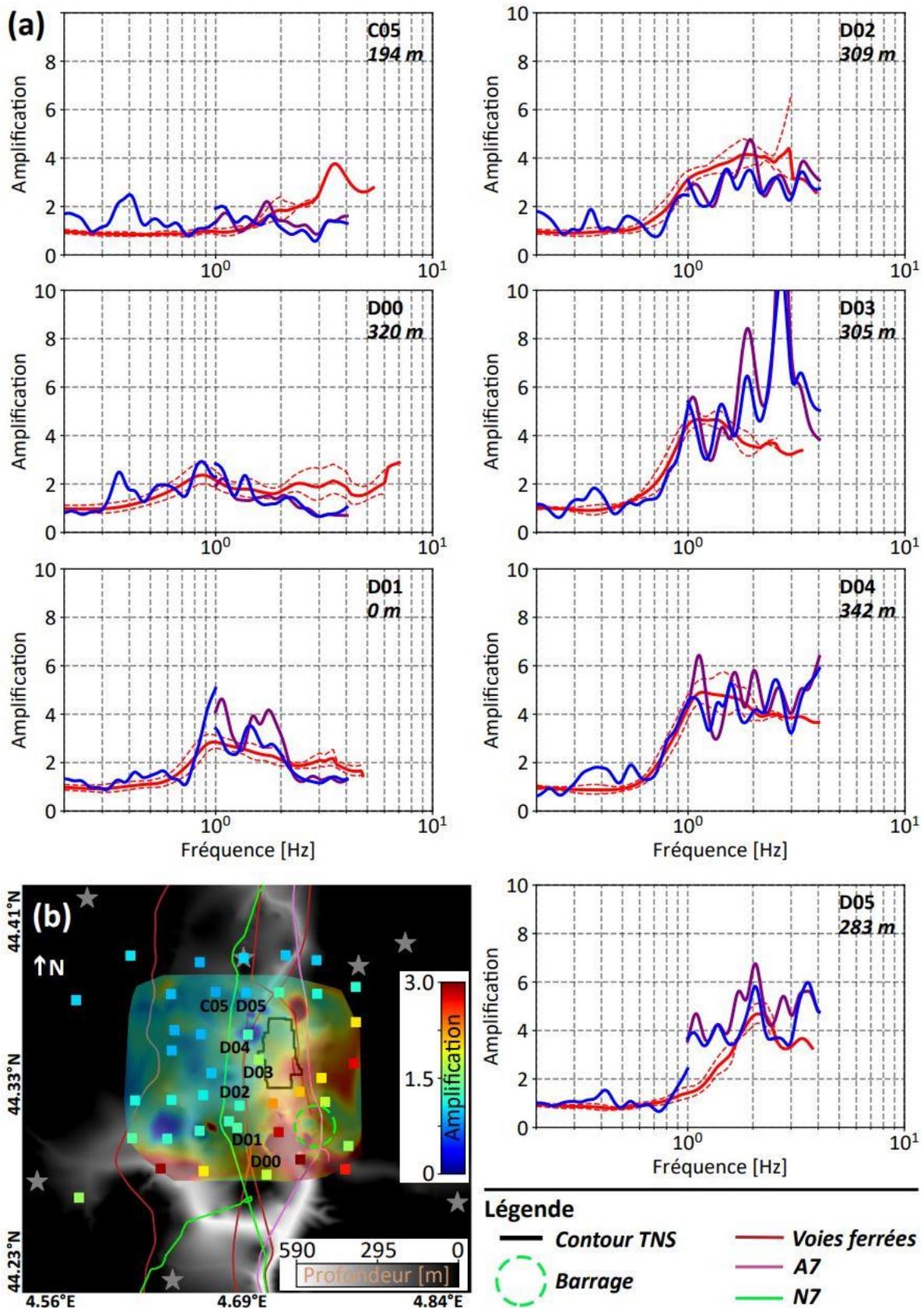


Figure 5.20 : Idem que pour la figure 5.16 pour la composante Z.

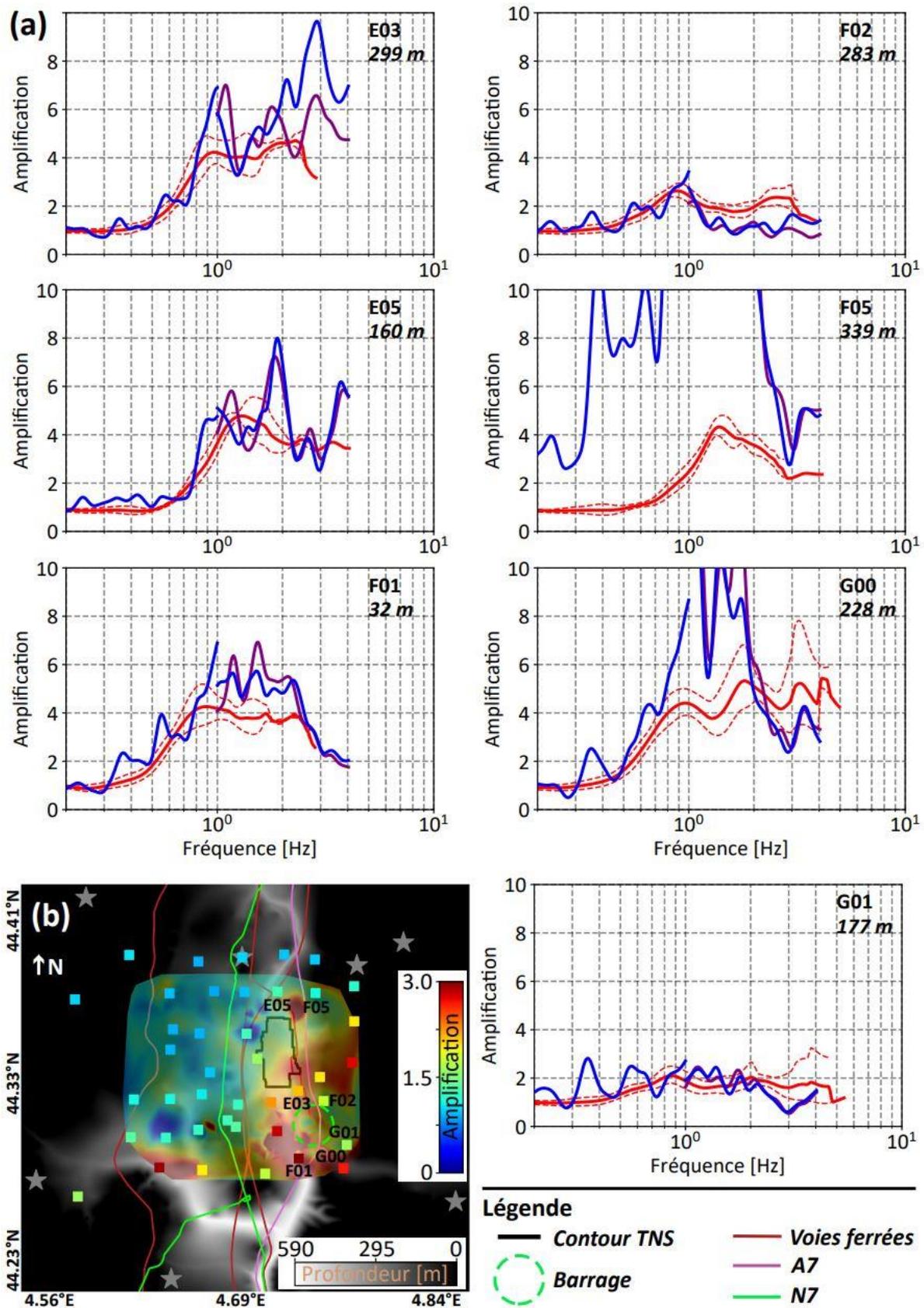


Figure 5.21 : Idem que pour la figure 5.17 pour la composante Z.

5.5 Discussion/conclusion

A basse fréquence (< 1 Hz), le suivi des fronts d'onde a montré la capacité de l'ANIRF à extraire la cohérence du champ de bruit sismique pour reconstruire un signal déterministe qui se propage à travers la structure géologique et qui en porte ces effets. L'utilisation de plusieurs sources virtuelles situées tout autour de la zone d'étude a mis en évidence la variabilité de la réponse de la vallée dépendamment de la direction d'incidence du champ entrant. L'amplitude de l'ANIRF est généralement maximale lorsque le front d'onde atteint les bords de la structure qui lui sont à peu près parallèles (et donc à peu près perpendiculaires à la direction de propagation). Des simulations numériques de la propagation du champ d'ondes depuis les différentes sources virtuelles pourraient aider à mieux comprendre leur interaction avec la géométrie de la structure dépendamment de la direction d'incidence.

A haute fréquence (entre 3 et 6 Hz), l'ANIRF n'a pas permis de reconstruire un signal propagatif clair, ce qui montre la difficulté d'extraire la cohérence de phase du champ de bruit à ces fréquences. En d'autres termes, l'ANIRF est simplement dominée par l'information d'amplitude contrôlée par l'énergie à la source virtuelle et au récepteur. Cette information d'amplitude est de surcroît biaisée par les sources anthropiques locales. Cependant, ce biais peut être atténué en sélectionnant les segments temporels de l'ANIRF les moins impactés. Les sources anthropiques majeures dans la zone d'étude étant globalement plus proches du récepteur que de la source virtuelle, leur signature est principalement retranscrite sur la partie anti-causale de l'ANIRF. Il convient alors de sélectionner uniquement la partie causale. Cette sélection très simple est néanmoins valable pour ce cas précis. Dans un cas général, une sélection plus fine des temps d'arrivée serait nécessaire. Dans l'idée, une telle sélection devrait tenir compte de l'angle formé par l'azimut de chaque paire avec la position des sources que l'on souhaite atténuer, avec la difficulté toutefois que les sources de bruit sismique sont rarement assimilables à une position unique dans l'espace (e.g. une autoroute).

L'amplification site/rocher associée à la vallée du Tricastin a été estimée sur la base de la partie sélectionnée de l'ANIRF à chaque paire en suivant une procédure en 4 étapes. Pour évaluer une réponse globale de la vallée à champ d'ondes pouvant provenir de toutes les directions, 7 sources virtuelles situées tout autour de la zone d'étude ont été utilisées. L'idée de la procédure est d'abord d'extraire une mesure d'amplitude relative à travers le réseau à partir de l'ANIRF, puis de « calibrer » cette mesure site-à-site par un SSR site/rocher. A basse fréquence (< 1 Hz), l'amplification

site/rocher ainsi obtenue est cohérente avec les mesures SSR ainsi que la structure de la vallée, montrant une amplification plus importante à l'aplomb des zones sédimentaires profondes. Ceci confirme la capacité de l'ANIRF à extraire la réponse du site à basse fréquence. La comparaison « fréquence à fréquence » entre les stations large-bande et le node le plus proche révèle néanmoins des biais entre les 2 mesures d'amplification (ANIRF et SSR) à proximité des bords de la vallée où le champ de bruit est très complexe. A haute fréquence (> 1 Hz), le facteur d'amplification obtenu par l'ANIRF est globalement plus proche du SSR lorsque seule la partie causale est considérée. L'ANIRF reproduit le facteur d'amplification du SSR jusqu'à environ 4 Hz pour à peu près la moitié des cas. Ceci laisse entrevoir le potentiel de cette méthode pour évaluer l'amplification site/rocher au-delà du hertz dans les zones fortement industrialisées. Le nombre d'enregistrements de séismes de qualité suffisante empêche toutefois d'étendre nos observations au-delà de 4 Hz.

Enfin, précisons que la sélection temporelle telle qu'appliquée ici prend en compte les premiers temps d'arrivée (à partir de $t = 0$ s). Or le champ d'ondes émis depuis une source virtuelle et correspondant à la fonction de Green partiellement reconstruite met un certain temps avant d'arriver sur le réseau. Par exemple dans le cas où ADHE est la source virtuelle (Figure 5.22), le front d'onde correspondant à ce champ atteint les premiers nodes (partie nord-est du réseau) à environ 1 s de temps de propagation (cercle blanc, dernier cadrant de la partie causale). A des temps de propagation plus courts, le champ qui domine l'ANIRF à travers le réseau est en réalité la signature de la/les source(s) dominante(s) dans la gamme de fréquences étudiée (ici 0.35 - 0.9 Hz). Le front d'onde correspondant à ce champ est plan et orienté sur un axe NE-SO, ce qui concorde avec l'océan Atlantique. Si ce champ d'ondes semble être porteur des effets de site (l'amplitude est plus importante dans la partie sédimentaire profonde du nord de la vallée du Rhône), il ne correspond néanmoins pas au scénario dans lequel un champ incident arriverait depuis cette source virtuelle et, par conséquent, pourrait induire un biais sur l'estimation de l'amplification qui se produirait pour un tel scénario. Pour de futures études, il serait peut-être préférable de filtrer ce champ avant de sélectionner la partie de l'ANIRF à utiliser dans l'évaluation de l'amplification.

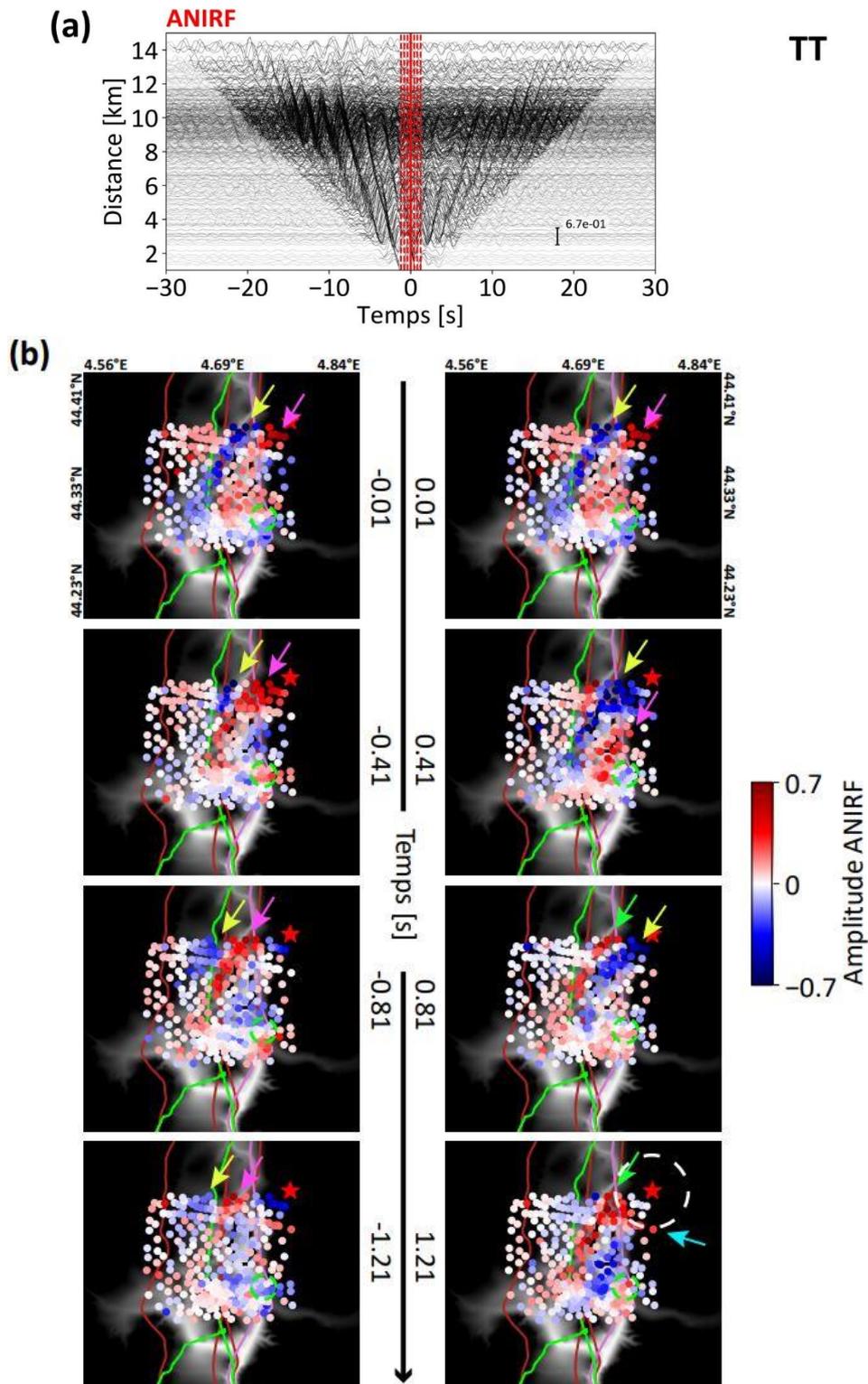


Figure 5.22 : Observation d'un front d'onde plan à basse fréquence (0.35-0.9 Hz) traversant le réseau sur un axe NE-SO pour des temps d'arrivée courts. L'orientation du front d'onde concorde avec la position de l'océan Atlantique par rapport au réseau (source de bruit dominante à ces fréquences). A 1.21 s, le front d'onde correspondant au champ émis par la source virtuelle ADHE arrive sur le réseau. Ce front d'onde est tangent au premier.

Conclusion générale

Les effets de site associés aux structures sédimentaires (bassins/vallées) sont à la fois fonction du contraste de propriétés mécaniques entre les sédiments et le substratum rocheux et de la géométrie de la structure. La géométrie de la structure peut engendrer des effets complexes 2D/3D que les proxys simples classiquement utilisés dans l'évaluation de l'aléa (e.g. V_s30) ne peuvent pas caractériser. Dans ce cas, des études spécifiques au site sont nécessaires afin d'appréhender cette complexité et d'intégrer correctement ces effets dans l'évaluation de l'aléa.

Les études « site-spécifiques » sont généralement menées via des approches numériques et/ou empiriques. L'approche numérique repose sur un modèle du sous-sol pour simuler le mouvement sismique en surface et nécessite ainsi une très bonne connaissance du milieu afin de contraindre au mieux ce modèle. L'approche empirique repose quant à elle sur une mesure directe des propriétés que l'on souhaite évaluer sur le site en question et nécessite de nombreux enregistrements de séismes pour obtenir une mesure robuste. En zone de sismicité faible à modérée, le temps de retour des séismes modérés à forts est long (plusieurs années) et la qualité des enregistrements de séismes de faible magnitude ou se produisant à grande distance est largement dépendante de l'amplitude du bruit sismique local. Ainsi, obtenir un nombre suffisant d'enregistrements exploitables demande des temps d'acquisition longs ; ce qui augmente d'autant le risque, tout en nécessitant une logistique potentiellement lourde pour la maintenance des instruments. Il peut alors être intéressant de se tourner vers l'utilisation du bruit sismique qui constitue une source de données permanente et exploitable partout à la surface de la Terre. En plus de pouvoir être enregistré partout, le bruit sismique peut être enregistré sur des capteurs faciles à déployer. Ainsi, les progrès technologiques récents en instrumentation permettent de densifier les réseaux de capteurs, améliorant la couverture spatiale de la zone d'étude et permettant de mieux prendre en compte la variabilité du mouvement.

Dans cette thèse, le potentiel du bruit sismique pour l'estimation de l'amplification associée aux effets de site a été étudié sur un réseau de capteurs dense

(400 nodes). Des méthodes de rapports spectraux (dites site-sur-référence) ont été appliquées dans la vallée du Tricastin (paléovallée du Rhône comprenant le site nucléaire de Tricastin). Tout d'abord, un contrôle qualité des données a été mené et a abouti à une première sélection des enregistrements de bruit sismique (chapitre 2, Froment *et al.* 2022a). Ensuite, deux méthodes basées sur des rapports spectraux calculées à partir de données de bruit sismique, le SSRn (Irikura et Kawanaka 1980) et le SSRh (Perron *et al.* 2018), ont été appliquées. Pour évaluer leur applicabilité, elles ont été comparées au SSR (Borcherdt 1970) estimé dans la zone (calculé à partir de données de séismes).

En dessous de 1 Hz, le SSRn fournit un facteur d'amplification comparable au SSR évalués sur séismes, et est cohérent avec nos connaissances de la structure géologique de la vallée du Tricastin ; en particulier, l'amplification estimée est plus importante (jusqu'à un facteur 6 autour de la fréquence de résonance fondamentale de la vallée ; e.g. Gélis *et al.* 2022 ; Gisselbrecht *et al.* 2023) à l'aplomb des zones sédimentaires profondes. Ceci confirme que le champ de bruit est porteur de la réponse de la structure géologique. Au-delà de 1 Hz en revanche, le facteur d'amplification du SSRn surestime largement celui du SSR. La principale problématique du SSRn réside dans la répartition des sources qui dominent le champ de bruit. En dessous de 1 Hz, celles-ci sont suffisamment distantes pour que leur contribution au champ de bruit enregistré sur la zone d'étude soit égale en tout point du réseau. Ainsi, un simple rapport d'amplitude entre 2 capteurs suffit à isoler la réponse du site. Au-delà de 1 Hz en revanche, le bruit sismique est dominé par des sources locales (liées à l'activité humaine) qui excitent ce champ de manière inégale sur la zone d'étude. Le rapport d'amplitude site-sur-référence entre deux capteurs n'est donc plus capable d'isoler les effets de site des contributions des différentes sources. Dans notre cas, les sources majeures de bruit sismique étant situées dans la vallée (et donc à proximité du « site »), le SSRn surestime le facteur d'amplification du SSR. Le SSRh, qui utilise un capteur de référence au site (relativement proche des sources majeures), fournit tout de même un facteur d'amplification comparable au SSR jusqu'à environ 4 Hz. Les résultats obtenus (présentés dans le chapitre 3) ont cependant montré une dépendance au choix du capteur de référence au site, questionnant ainsi la possibilité d'utiliser cette méthode de manière automatique dans un contexte fortement industrialisé comme celui étudié ici. Notons que pour les besoins de l'ingénierie (évaluation de la réponse des bâtiments et infrastructures ; e.g. Borcherdt 1994), il peut être intéressant de chercher à augmenter la gamme de fréquences sur laquelle il est possible de mesurer l'amplification, typiquement jusqu'à 10 Hz.

Pour isoler l'amplification liée aux effets de site de manière robuste à des fréquences de quelques hertz, il paraît important de limiter l'influence des sources locales. En premier lieu, nous avons cherché à traiter cette problématique en considérant un mode de sélection des enregistrements continus. Un algorithme de clustering a été utilisé afin de sélectionner pour chaque enregistrement et de manière automatique, les fenêtres temporelles pour lesquelles l'activité des sources locales est moins forte. En suivant cette procédure, l'amplification estimée par le SSRh a pu être améliorée dans les zones du réseau impactées par des sources transitoires. Cependant, l'activité de certaines sources varie peu dans le temps. La procédure de sélection introduite a donc un impact limité les concernant. Les enregistrements à certains capteurs sont perpétuellement dominés par des sources locales. Ainsi, dans les zones où ces sources dominent le champ de bruit, le biais sur l'estimation de l'amplification reste important. Si cette procédure de sélection peut s'avérer efficace pour les sites où des sources transitoires dominent le champ de bruit, ces résultats montrent aussi la difficulté d'implémenter une méthode se basant uniquement sur l'amplitude du bruit sismique dans les zones fortement industrialisées, où des sources permanentes (barrage, site nucléaire, usine) peuvent être dominantes.

Outre l'information d'amplitude, il est possible d'inclure l'information portée par la phase des enregistrements continus, et en particulier sa cohérence spatiale à travers le bassin, pour essayer de mieux contraindre l'amplification. La déconvolution entre des enregistrements distants est une méthode interférométrique permettant d'aller dans ce sens, notamment dans le cas où l'effet des sources locales semble non négligeable. L'idée d'utiliser la déconvolution de bruit sismique (ou ANIRF) pour estimer l'amplification est que l'information de phase qu'elle contient peut aider à filtrer spatialement le champ de bruit et ainsi limiter l'impact des sources locales, notamment des sources permanentes qui sont plus difficiles à traiter. L'ANIRF a ici été calculée en utilisant 7 sources virtuelles distribuées autour de la zone d'étude. Dans notre cas, l'amplitude absolue de l'ANIRF ne correspond pas directement à l'amplification site/rocher recherchée (les conditions de convergence vers la fonction de Green, c-à-d la réponse du milieu, ne sont pas satisfaites). L'amplitude relative à travers la zone en revanche semble refléter les effets liés à la structure. Pour évaluer l'amplification site/rocher, nous avons calibré cette mesure d'amplitude site-à-site avec un SSR site/rocher. Le facteur d'amplification ainsi obtenu reproduit le facteur d'amplification du SSR jusqu'à environ 4 Hz pour à peu près la moitié des comparaisons effectuée. Ceci laisse entrevoir le potentiel de cette méthode pour évaluer l'amplification site/rocher au-delà du hertz dans les zones fortement industrialisées.

Les résultats obtenus permettent de discuter les avantages et limitations des différentes approches développées et testées dans ce travail de thèse et d'ouvrir des perspectives sur des travaux futurs pour améliorer l'étude des effets de site à partir du bruit sismique. En particulier, l'ANIRF semble être une méthode prometteuse en vue d'estimer l'amplification au site sur une large gamme de fréquences dans les zones industrialisées. Quelques perspectives sont présentées ci-dessous afin de compléter les pistes de travail mentionnées au sein des différents chapitres et prolonger la discussion autour des nouvelles approches discutées dans ce manuscrit.

Le principe du SSRh et de notre procédure sur l'ANIRF est d'estimer une amplification site-à-site à travers le réseau, puis de « calibrer » cette amplification par une mesure SSR pour évaluer l'amplification site/rocher. A l'avenir, il serait intéressant de faire cette calibration au rocher autrement que par une mesure SSR, dans le but de s'affranchir totalement de l'utilisation d'enregistrements de séismes. Il serait également intéressant d'évaluer l'intérêt de considérer d'autres références au site choisies en fonction de leur localisation au sein de la structure sédimentaire. L'idée pourrait être d'utiliser plusieurs stations de référence au site, chacune se rapportant à une zone au sein de laquelle les conditions en termes de sources de bruit pourraient être considérées comme à peu près homogènes. L'estimation site-à-site de l'amplification serait alors calculée zone par zone, puis calibrée au rocher pour chacune d'entre elles. La définition de ces zones pourrait être faite de manière automatique par le biais d'un algorithme de clustering capable de classer les signaux en fonction de leurs caractéristiques (forme d'onde, contenu fréquentiel, amplitude...) et ainsi d'identifier les différents types de sources. Dans cette thèse, le DBSCAN s'est par exemple révélé être un outil prometteur pour une classification dans le domaine spectral. Steinmann *et al.* (2022) ont, par exemple, récemment proposé une méthode automatisée de classification des signaux et d'identification des types de sources par regroupement hiérarchique des formes d'ondes. Ils ont montré la performance de cette approche pour isoler la signature des effets de gel/dégèle en surface dans les signaux continus dans un environnement urbain. Cette approche serait intéressante à tester sur ce jeu de données puisqu'elle se rapproche de notre objectif d'isoler la réponse physique du milieu. Le domaine de l'apprentissage machine (actuellement en plein essor) est en tout cas un vaste champ d'études dont les applications pour la sismologie sont potentiellement multiples. Avec la prolifération de réseaux de capteurs de plus en plus denses grâce au progrès de l'instrumentation, les jeux de données deviennent de plus en plus volumineux et il devient essentiel de pouvoir les explorer/traiter de manière automatique.

Une autre voie serait d'explorer la possibilité de modéliser l'ensemble des sources locales majeures de bruit sismique dans la zone pour pouvoir les intégrer dans une modélisation avancée du mouvement sismique dans la zone (basée sur un modèle de sol – comme ce qui est fait actuellement dans le cadre du projet DARE – et sur un modèle de sources) et/ou de retirer leurs signatures dans les enregistrements. Ceci demanderait bien-sûr de connaître les paramètres associés à chacune d'entre elles (amplitude, contenu fréquentiel, durée, position...) mais nous avons vu que par l'intermédiaire d'une exploration approfondie des données (chapitre 2), ces paramètres ne sont pas inaccessibles. Certains auteurs ont par exemple été capables de modéliser des signaux associés au passage des trains (e.g. Lavoué *et al.* 2020) ou à la machinerie de certains navires (Traverso *et al.* 2012). Ces travaux proposent des pistes pour aborder ces problématiques liées à la modélisation des sources. Là encore, les méthodes d'apprentissage machine peuvent être intéressantes pour adresser ces questions dans une configuration complexe, comme celle étudiée ici.

Bibliographie

Aki K., Richards P.G., 1980. Quantitative Seismology.

Aki K., 1993. Local site effects on weak and strong ground motion. *Tectonophysics*, 218, 93–111.

Arai, H., Tokimatsu, K., 1998. Evaluation of local site effects based on microtremor H/V spectra, in *Proceeding of the Second International Symposium on the Effects of Surface Geology on Seismic Motion*, Yokohama, Japan, pp. 673–680.

Bagayoko, N., 2021. *Intégration de données de sismiques réflexion et de données géologiques dans un modèle structural 3D du canyon messinien du Rhône : Etude du site du Tricastin*, M.Sc. thesis, Sorbonne Université, Paris.

Banerji, S.K., 1924. Microseisms associated with the incidence of the south-west monsoon. *Nature*, 114(2868), 576.

Banerji, S.K., 1925. Microseisms and the Indian monsoon. *Nature*, 116(2928), 866.

Bard, P.-Y. & Bouchon, M., 1980. The seismic response of sediment-filled valleys. Part 1. The case of incident *SH* waves. *Bulletin of the Seismological Society of America*, 70(4), 1263–1286, doi: <https://doi.org/10.1785/BSSA0700041263>.

Bard, P.-Y. & Bouchon, M., 1985. The two-dimensional resonance of sediment-filled valleys, *Bull. seism. Soc. Am.*, 75(2), 519-541, doi:10.1785/BSSA0750020519.

Bard, P.-Y., 1999. Microtremor measurements: a tool for site effect estimation? in *Proceedings of the 2th IASPEI/IAEE (ESG2)*, Yokohama, Japan, 1998 December 1-3, pp. 1251-1279.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, 169, 1239-1260, doi:10.1111/j.1365-246X.2007.03374.x.

Bertelli, T., 1872. Osservazioni sui piccolo movimenti dei pendoli in relazione ad alcuni fenomeni meteorologiche. *Bullettino Meteorologico dell'Osservatorio dell'Collegio Romano*, 101, (In Italian).

Bonilla LF, Steidl JH, Lindley GT, Tumarkin G, Archuleta R, 1997. Site amplification in the San Fernando Valley, California: variability of site-effect estimation using the S-wave, coda, and H/V methods, *Bull Seismol Soc Am*, 87, 710–730, doi:10.1785/BSSA0870030710.

Bonnefoy-Claudet, S., 2004. Nature du bruit de fond sismique : implications pour les études des effets de site, thèse, Université Joseph Fourier, Grenoble, France.

Bonnefoy-Claudet, S., Cotton, F. & Bard, P.-Y., 2006a. The nature of noise wavefield and its applications for site effects studies, *Earth Sci. Rev.*, 79, 205–227, doi:10.1016/j.earscirev.2006.07.004.

Bonnefoy-Claudet, S., Cornou, C., Bard, P.-Y., Cotton, F., Moczo, P., Kristek, J., Fäh, D., 2006b. H/V ratio: a tool for site effects evaluation. Results from 1D noise simulations. *Geophysical Journal International*, 167, 827–837, doi:10.1111/j.1365-246X.2006.03154.x.

Borcherdt, R.D., 1970. Effects of local geology on ground motion near San Francisco Bay, *Bull. seism. Soc. Am.*, 60, 29-61, doi:10.1785/BSSA0600010029.

Borcherdt, R.D., 1994. Estimates of Site-Dependent Response Spectra for Design (Methodology and Justification), *Earthquake Spectra*, 10(4).

Boué, P., Denolle, M.A., Hirata, N., Nakagawa, S. & Beroza G.C., 2016. Beyond basin resonance: characterizing wave propagation using a dense array and the ambient seismic field, *Geophys. J. Int.*, 206, 1261-1272, doi:10.1093/gji/ggw205.

Bowden DC, Tsai VC, Lin FC (2015) Site amplification, attenuation, and scattering from noise correlation amplitudes across a dense array in Long Beach, CA. *Geophys Res Lett* 42:1360– 1367. doi: 10.1002/2014GL062662.

Burjánek, J., Gischig, V., Moore, J. R. & Fäh, D., 2018. Ambient vibration characterization and monitoring of a rock slope close to collapse, *Geophys. J. Int.*, 212(1), 297-310, doi:10.1093/gji/ggx424.

Butterworth, S., 1930. *On the Theory of Filter Amplifiers*, *Wireless Engineer*, 7, p. 536-541.

Campillo, M., Gariel, J.C., Aki, K., Sánchez-Sesma, F. J., 1989. Destructive strong ground motion in Mexico city: Source, path, and site effects during great 1985 Michoacán earthquake, *Bulletin of the Seismological Society of America*, 79(6), 1718–1735, doi:<https://doi.org/10.1785/BSSA0790061718>.

Capon, J., 1969. High-resolution frequency–wavenumber spectrum analysis. *IEEE* 1408–1419.

Causse, M., Cornou, C., Maufroy, E., Grasso, J.R., Baillet, L. & El Haber, E., 2021. Exceptional ground motion during the shallow M_w 4.9 2019 Le Teil earthquake, France, *Commun Earth Environ* 2, 14, doi:[10.1038/s43247-020-00089-0](https://doi.org/10.1038/s43247-020-00089-0).

Celebi, M., Dietel, C., Prince, J., Onate, M. & Chavez, G., 1987. Site amplification in Mexico City (determined from 19 September 1985 strong motion records and from recordings of weak motions), *Dev. Geotech. Eng.*, 44, 141-151, doi:[10.1016/B978-0-444-98956-7.50013-1](https://doi.org/10.1016/B978-0-444-98956-7.50013-1).

Cornou, C., 2002. Traitement d'antenne et imagerie sismique dans l'agglomération grenobloise (Alpes françaises) : implications pour les effets de site, thèse, Université Joseph Fourier, Grenoble, France.

Cornou C., Bard, P.-Y., Dietrich M. 2003a. Contribution of Dense Array Analysis to the Identification and Quantification of Basin-Edge-Induced Waves, Part I: Methodology, *Bull. seism. Soc. Am.*, 93(6), 2604–2623, doi:[10.1785/0120020139](https://doi.org/10.1785/0120020139).

Cornou C., Bard, P.-Y., Dietrich M. 2003b. Contribution of Dense Array Analysis to the Identification and Quantification of Basin-Edge-Induced Waves, Part II: Application to Grenoble Basin (French Alps), *Bull. seism. Soc. Am.*, 93(6), 2624–2648, doi:[10.1785/0120020140](https://doi.org/10.1785/0120020140).

Cornou, C., Ampuero, J.-P., Aubert, C., Audin, L. et al., 2021. Rapid response to the M_w 4.9 earthquake of November 11, 2019 in Le Teil, Lower Rhône Valley, France, *Compt. Rendus. Geosci.*, 353(S1), 441-463, doi:[10.5802/crgeos.30](https://doi.org/10.5802/crgeos.30).

Cultrera G, De Rubeis V, Theodoulidis N, et al., 2014. Statistical correlation of earthquake and ambient noise spectral ratios, *Bull Earthq Eng.* doi:[10.1007/s10518-013-9576-7](https://doi.org/10.1007/s10518-013-9576-7).

Davis P., Rubinstein J., Liu K., Gao S., Knopoff L., 2000. Northridge Earthquake Damage Caused by Geologic Focusing of Seismic Waves, *Science*, 289, 1746-50, doi:[10.1126/science.289.5485.1746](https://doi.org/10.1126/science.289.5485.1746).

De Angelis S., Bodin P., 2012. Watching the Wind: Seismic Data Contamination at Long Periods due to Atmospheric Pressure-Field-Induced Tilting, *Bulletin of the Seismological Society of America*, 102(3), 1255–1265, doi:<https://doi.org/10.1785/0120110186>.

Deng, D., 2020. DBSCAN Clustering Algorithm Based on Density, *7th International Forum on Electrical Engineering and Automation (IFEEA)*, 2020, pp. 949-953, doi:10.1109/IFEEA51475.2020.00199.

Denolle, M.A., Dunham, E.M., Prieto, G.A. & Beroza, G.C., 2013. Ground motion prediction of realistic earthquake sources using the ambient seismic field, *J. Geophys. Res. Solid Earth*, 118, 2102-2118, doi:10.1029/2012JB009603.

Denolle, M.A., Dunham, E.M., Prieto, G.A. & Beroza, G.C., 2014. Strong Ground Motion Prediction Using Virtual Earthquakes, *Science*, 343(6169), 399-403, doi:10.1126/science.1245678.

Denolle, M. A., Boué, P., Hirata, N., & Beroza, G. C., 2018. Strong shaking predicted in Tokyo from an expected M7+ Itoigawa-Shizuoka earthquake, *Journal of Geophysical Research: Solid Earth*, 123, 3968–3992. <https://doi.org/10.1029/2017JB015184>.

Derode, A., Larose, E., Campillo, M., Fink, M., 2003. How to estimate the green's function of a heterogeneous medium between two passive sensors? application to acoustic waves. *Applied Physics Letters*, 83(15), 3054–3056, doi:10.1063/1.1617373.

Di Giulio, G., Cornou, C., Ohrnberger, M., Wathelet, M., Rovelli A., 2006. Deriving Wavefield Characteristics and Shear-Velocity Profiles from Two- Dimensional Small-Aperture Arrays Analysis of Ambient Vibrations in a Small-Size Alluvial Basin, Colfiorito, Italy, *Bulletin of the Seismological Society of America*, 96(5), 1915–1933, doi:<https://doi.org/10.1785/0120060119>.

Do Couto, D., Cushing, E.M., Mocochain, L., Rubino, J.L., Miquelis F., Hanot, F., Froment, B., Gélis, C., Camus M., Bagayoko, N., Bellier, O., soumis. Reappraisal of the Messinian canyons morphology of the Rhône and Ardèche (Rhône valley): new insights from seismic profiles.

Endrun, B., Ohrnberger, M. & Savvaidis, A. 2010. On the repeatability and consistency of three-component ambient vibration array measurements. *Bull Earthquake Eng*, 8, 535–570, <https://doi.org/10.1007/s10518-009-9159-9>.

Ferritto, J., 1995. Ground Motion Amplification Using Microseisms, in *Int. Conf. Recent. Adv. Geotech. Earthq. Eng. Soil. Dyn.*, Paper 4, 561-566, St. Louis, Missouri, 1995 April 2-7.

Field, E. H., Hough, S. E., Jacob K. H., 1990. Using microtremors to assess potential earthquake site response: A case study in Flushing Meadows, New York City. *Bulletin of the Seismological Society of America*, 80 (6A), 1456–1480. doi:<https://doi.org/10.1785/BSSA08006A1456>.

Field, E.H. & Jacob, K.H., 1995. A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent, *Bull. seism. Soc. Am.*, 85(4), 1127-1143, doi:[10.1785/BSSA0850041127](https://doi.org/10.1785/BSSA0850041127).

Field, E.H., 1996. Spectral amplification in a sediment-filled valley exhibiting clear basin-edge-induced waves, *Bull Seismol Soc Am*, 86, 991–1005.

Fink, M., 1992. Time reversal of ultrasonic fields. i. basic principles. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 39(5), 555–566, ISSN 0885-3010, doi : [10.1109/58.156174](https://doi.org/10.1109/58.156174).

Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A., Weaver R.L., 2010. Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations, *Geophysics*, 75(5), SA85–SA93, doi:[10.1190/1.3483102](https://doi.org/10.1190/1.3483102).

Froment, B., Olivar-Castano, A., Ohrnberger, M., Gisselbrecht, L. et al., 2022a. Complementary dense seismic campaigns in a low-to-moderate seismicity area for characterizing site effects: application to the French Rhône Messinian canyon, *Seismol. Res. Lett.*, in press.

Froment, B., Cushing, E.M., Gélis, C., Bagayoko, N., Beauprêtre, S., Boué, P., Do Couto, D., Magnin, O., Mocochain, L., Mordret, A., Rubino, J.-L. & Tourette, A., 2022b. First 3D characterization of the Rhône Messinian Canyon in the Tricastin area from complementary geophysical approaches, in *Proceedings of the 17th European Conference on Earthquake Engineering (17ECEE)*, Bucharest, Romania, 4-9 September 2022.

Froment, B., Cushing, E.M., Gélis, C., Gisselbrecht, L., Beauprêtre, S., Tourette, A. & RESIF, 2023. France 2020, Dense nodal seismic array in the Rhône Valley, DARE project [Data set]. RESIF-Réseau Sismologique et géodésique Français, doi:[10.15778/RESIF.XG2020](https://doi.org/10.15778/RESIF.XG2020).

Gallot, T., Catheline, S., Roux, P., Campillo M. 2012. A passive inverse filter for Green's function retrieval. *J. Acoust. Soc. Am.*, 131(1), EL21–EL27, doi:<https://doi.org/10.1121/1.3665397>.

Gao S., Liu H., Davis P.M., Knopoff L., 1996. Localized Amplification of Seismic Waves and Correlation with Damage due to the Northridge Earthquake: Evidence for Focusing in Santa Monica, *Bulletin of the Seismological Society of America*, 86(1B), 209–230, doi:<https://doi.org/10.1785/BSSA08601BS209>.

Gélis C., Bonilla L.F., Regnier J., Bertrand E., Duval A.M., 2008. On the use of Saenger's finite difference stencil to model 2D P-SV nonlinear basin response: application to Nice, France.

Gélis, C., Cauchie, L., Cushing, E.M., Froment, B., Franco, S., Jomard, H., Moiriat, D., Provost, L., Sariguzel, B. & Tebib, H., 2022. Estimation of the local seismic amplification on an industrialized site in the French Rhône Valley, *Pure appl. geophys.*, doi:10.1007/s00024-022-03069-x, in press.

Gisselbrecht L., Froment B., Boué P., Gélis C., 2023. Insights into the conditions of application of noise-based spectral ratios in a highly industrialized area: a case study in the French Rhone Valley, *Geophysical Journal International*, 234(2), 985–997, <https://doi.org/10.1093/gji/ggad108>.

Gutenberg, B., 1958. Microseisms. *Advances in Geophysics*, 5, 53–92.

Graves W. 1993. Modeling three-dimensional site response effects in the Marina District Basin, San Francisco, California, *Bulletin of the Seismological Society of America*, 83(4), 1042–1063, doi:<https://doi.org/10.1785/BSSA0830041042>.

Graves W., Pitarka A., Somerville P., 1998. Ground-motion amplification in the Santa Monica area: Effects of shallow basin-edge structure, *Bulletin of the Seismological Society of America*, 88(5), 1224–1242. doi:<https://doi.org/10.1785/BSSA0880051224>.

Haubrich, R.A., Munk, W.H., Snodgrass, F.E., 1963. Comparative spectra of microseisms and swell. *Bulletin of the Seismological Society of America*, 53(1), 27–37.

Hallier, S., Chaljub, E., Bouchon, M. *et al.* Revisiting the Basin-edge Effect at Kobe During the 1995 Hyogo-Ken Nanbu Earthquake. *Pure appl. geophys.* 165, 1751–1760 (2008). <https://doi.org/10.1007/s00024-008-0404-y>.

Haskell, N. A. (1960), Crustal reflection of plane *SH* waves, *J. Geophys. Res.*, 65(12), 4147–4150, doi:10.1029/JZ065i012p04147.

Hennino, R. et al. 2001. Observation of equipartition of seismic waves, *Physical Review Letters*, 86, 3447–3450.

Horike, M., 1985. Inversion of phase velocity of long-period microtremors to the S-wave-velocity structure down to the basement in urbanized areas. *Journal of Physics of the Earth*, 33, 59–96.

Horike, M., Zhao, B. & Kawase, H., 2001. Comparison of site response characteristics inferred from microtremors and earthquake shear waves, *Bull. seism. Soc. Am.*, 91(6), 1526-1536, doi:10.1785/0120000065.

Huang, H.C. & Teng, T.L., 1999. An Evaluation on H/V Ratio vs. Spectral Ratio for Site-response Estimation Using the 1994 Northridge Earthquake Sequences, *Pure appl. geophys.*, 156, 631-649, doi:10.1007/s000240050316.

Hippolyte, J.-C., Suc, J.-P., Gorini, C., Rubino, J.-L., Do Couto, D., 2021. La Méditerranée s'assèche à la fin du Messinien. Bouridey J. La géologie des Bouches du Rhône, Roches et paysages remarquables, *BRGM Edition*, 145-160.

Ibs-von Seht M., Wohlenberg J., 1999. Microtremor measurements used to map thickness of soft sediments, *Bull Seismol Soc Am*, 89, 250–259, doi:https://doi.org/10.1785/BSSA0890010250.

Irikura, K. & Kawanaka, T., 1980. Characteristics of microtremors on ground with discontinuous underground structure, *Bull. disaster prev. res. inst.*, 30, 81-96.

Kagami, H., Duke, C.M., Liang, G.C. & Ohta, Y., 1982. Observation of 1-to 5-second microtremors and their application to earthquake engineering. Part II. Evaluation of site effect upon seismic wave amplification due to extremely deep soil deposits, *Bull. seism. Soc. Am.*, 72(3), 987-998, doi:10.1785/BSSA0720030987.

Kagami, H., Okada, S., Shiono, K., Oner, M., Dravinski, M. & Mal, A.K., 1986. Observation of 1-to 5-second microtremors and their application to earthquake engineering. Part III. A two-dimensional study of site effects in the San Fernando Valley, *Bull. seism. Soc. Am.*, 76(6), 1801-1812, doi:10.1785/BSSA0760061801.

Kawase, H. (1996). The cause of the damage belt in Kobe: 'The Basin-Edge Effect', constructive interference of the direct S-wave with the basin-induced

diffracted/rayleigh waves. *Seismological. Research Letters*, 67(5), 25–34. <https://doi.org/10.1785/gssrl.67.5.25>.

Kawase, H., Matsushima, S., Satoh, T. & Sanchez-Sesma, F.J., 2015. Applicability of theoretical horizontal-to-vertical ratio of microtremors based on the diffuse field concept to previously observed data, *Bull. seism. Soc. Am.*, 105(6), 3092-3103, doi:10.1785/0120150134.

Kawase, H., Nagashima, F., Nakano, K. & Mori, Y., 2019. Direct evaluation of S-wave amplification factors from microtremor H/V ratios: Double empirical corrections to “Nakamura” method, *Soil Dyn. Earthqu. Eng.*, 126, doi:10.1016/j.soildyn.2018.01.049.

King JL, Tucker BE (1984) Observed variations of earthquake motion across a sediment-filled valley. *Bull Seismol Soc Am*, 74:137–151.

Konno, K. & Ohmachi, T., 1998. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. seism. Soc. Am.*, 88, 228-241, doi:10.1785/BSSA0880010228.

Köhler, A., Ohrnberger, M., Scherbaum, F., Wathelet, M., Cornou, C., 2007. Assessing the reliability of the modified three-component spatial autocorrelation technique, *Geophysical Journal International*, 168(2), 779–796, doi:<https://doi.org/10.1111/j.1365-246X.2006.03253.x>.

Kramer SL (1996) Geotechnical earthquake engineering. Pearson Education India.

Lacoss, R.T., Kelly, E.J., Nafi, T.M., 1969. Estimation of seismic noise structure using arrays. *Geophysics*, 34(1), 21–38.

Larose, E., Derode, A., Campillo, M. & Fink, M., 2004. Imaging from one-bit correlations of wideband diffuse wavefields, *J. Appl. Phys.*, 95, 8393– 8399.

Larroque, C. et al., 2021. Seismotectonic of the South-East France: from the Jura mountains to Corsica. *C. R. Géosci.*, 353(S1), 105-151, doi:10.5802/crgeos.69.

Laurendeau A., 2013. Définition du mouvement sismique « au rocher », thèse, Université Grenoble Alpes, Grenoble, France.

Lavoué et al., soumis. Potential and limitations of noise-based surface-wave tomography for 2 numerical site effect estimation: a case study in the French Rhône valley, *Comptes Rendus. Géoscience*.

Lavoué, F., O. Coutant, P. Boué, L. Pinzon-Rincon, F. Brenguier, R. Brossier, P. Dales, M. Rezaeifar, C. J. Bean, 2020. Understanding Seismic Waves Generated by Train Traffic via Modeling: Implications for Seismic Imaging and Monitoring, *Seismol. Res. Lett.*, 1–14, doi:10.1785/0220200133.

Lermo, J. & Chávez-García, F.J., 1994. Are Microtremors useful in site response Evaluation?, *Bull. seism. Soc. Am.*, 84(5), 1350-64.

Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, *J. Exp. Soc. Psychol.*, 49(4), 764-766, doi:10.1016/j.jesp.2013.03.013.

Li, T.M.C., Ferguson, J.F., Herrin, E., Durham, H.B., 1984. High-frequency seismic noise at Lajitas, Texas. *Bulletin of the Seismological Society of America*, 74 (5), 2015–2033.

Lobkis O. I., Weaver. R. L., 2001. On the emergence of the green's function in the correlations of a diffuse field. *Journal of Acoustical Society of America*, 110(6), 3011–3017, doi : 10.1121/1.1417528.

Longuet-Higgins, M.S., 1950. A theory of the origin of microseisms. *Philosophical Transactions of the Royal Society of London*, A 243 (a), 1–35.

Lott, F.F., Ritter, J.R.R., Al-Qaryouti, M. et al. 2017. On the Analysis of Wind-Induced Noise in Seismological Recordings, *Pure Appl. Geophys.*, 174, 1453–1470, <https://doi.org/10.1007/s00024-017-1477-2>.

Maufroy, E., et al., 2017. Source-Related Variability of Site Response in the Mygdonian Basin (Greece) from Accelerometric Recordings and 3D Numerical Simulations, *Bull. Seismol. Soc. Am.*, 107(2), doi:10.1785/0120160107.

Mocochain, L., Clauzon, G., Bigot, J.-Y., 2006. The Ardèche endokarstic responses to the eustatic variations resulting from the Messinian salinity crisis [Réponses de l'endokarst ardéchois aux variations eustatiques générées par la crise de salinité messinienne], *Bulletin de La Societe Géologique de France*, 177(1), 27-36, doi:<https://doi.org/10.2113/177.L27>.

Mori, Y., Matsushima, S., Kawase, H. & Nagashima, F., 2016. Comparison of observed earthquake and microtremor horizontal-to-vertical spectral ratios and inversion of velocity structures based on their empirical ratios, *J. Jpn. Assoc. Earthq. Eng.*, 16(9), 13-32, doi:10.5610/jaee.16.9_13.

Naderyan, V., Hickey, C. J., Raspet, R., 2016. Wind-induced ground motion, *J. Geophys. Res. Solid Earth*, 121, 917–930, doi:10.1002/2015JB012478.

N. Nakata, L. Gualtiéri et A. Fichtner, 2019. Seismic ambient noise, *Cambridge University Press*.

Narayan J. P., 2005. Study of Basin-edge Effects on the Ground Motion Characteristics Using 2.5-D Modelling, *Pure appl. geophys.*, 162, 1–17, doi:0.1007/s00024-004-2600-8.

Olsen K. B., Day S. M., Minster J. B., Cui Y., Chourasia A., Faerman M., Moore R., Maechling P., Jordan T., 2006. Strong shaking in Los Angeles expected from southern San Andreas earthquake, *geophysical research letters*, 33, doi:10.1029/2005gl025472.

Perron V, Hollender F, Bard P-Y, Gélis C., Guyonnet-Benaize C., 2015. Site instrumentation usefulness: implementation recommendations and application examples of empirical site effect estimation in low to moderate seismicity context, In pre: 9ème Colloque National AFPS. Marne-la-Vallée.

Perron, V., 2017. Apport des enregistrements de séismes et de bruit de fond pour l'évaluation site-spécifique de l'aléa sismique en zone de sismicité faible à modérée, thèse, Université Grenoble Alpes, Grenoble, France.

Perron, V., Gélis, C., Froment, B., Hollender, F., Bard, P.-Y., Cultrera G. & Cushing E.M., 2018. Can broad-band earthquake site responses be predicted by the ambient noise spectral ratio? Insight from observations at two sedimentary basins, *Geophys. J. Int.*, 215, 1442-1454, doi:10.1093/gji/ggy355.

Perron, V., Bergamo, P., Danciu, L. & Fäh, D., 2020. Statistic of the earthquake's site response: Determination of the minimum number of earthquakes, Final Rept., ETH Zurich, doi:10.3929/ethz-b000479043.

Perron, V., Bergamo, P. & Fäh, D., 2022. Site Amplification at High Spatial Resolution from Combined Ambient Noise and Earthquake Recordings in Sion, Switzerland, *Seismol. Res. Lett.*, 93(4), 2281-2298, doi:10.1785/0220210289.

Prieto GA, Beroza GC, 2008. Earthquake ground motion prediction using the ambient seismic field, *Geophys Res Lett*, 35, doi:10.1029/2008GL034428.

Prieto, G. A., Parker, R. L. & Vernon, F. L., 2009. A Fortran 90 library for multitaper spectrum analysis, *Comput. Geosci.*, 35, 1701-1710, doi:10.1016/j.cageo.2008.06.007.

Prieto, G.A., Denolle, M.A., Lawrence, J.F. & Beroza, G.C., 2011. On amplitude information carried by the ambient seismic field, *Compt. Rendus. Geosci.*, 343, 600-614, doi:10.1016/j.crte.2011.03.006.

Peterson J. 1993. Observations and modeling of seismic background noise, *U.S. Geological Survey*, 93-322, doi: 10.3133/ofr93322.

Raptakis, D., Theodulidis, N. & Pitilakis, K., 1998. Data Analysis of the Euroseistest Strong Motion Array in Volvi (Greece): Standard and Horizontal-to-Vertical Spectral Ratio Techniques, *Earthquake Spectra*, 14(1), 203-224, doi:10.1193/1.1585996.

Ringler, A., Anthony, R., Karplus, M., Holland, A., Wilson, D., 2018. Laboratory Tests of Three Z-Land Fairfield Nodal 5-Hz, Three-Component Sensors, *Seismological Research Letters*, 89. 10.1785/0220170236.

Ritz, JF., Baize, S., Ferry, M., Larroque, C., Audin, L., Delouis, B. & Mathot, E., 2020. Surface rupture and shallow fault reactivation during the 2019 Mw 4.9 Le Teil earthquake, France, *Commun Earth Environ* 1, 10, doi:10.1038/s43247-020-0012-z.

Ritzwoller, M. H., Lin, F.-C., Shen, W., 2011. Ambient noise tomography with a large seismic array, *Comptes Rendus Geoscience*, 343, 558–570, doi:10.1016/j.crte.2011.03.007.

Roten, D., Fah, D., Cornou, C., & Giardini, D. (2006). Two-dimensional resonances in Alpine valleys identified from ambient vibration wavefields. *Geophysical Journal International*, 165(3), 889–905. <https://doi.org/10.1111/j.1365-246X.2006.02935.x>.

Rousseeuw, P. J. & Croux, C., 1993. Alternatives to the Median Absolute Deviation, *J. Am. Stat. Assoc.*, 88(424), 1273-1283, doi:10.1080/01621459.1993.10476408.

Roux, P., Sabra, K. G., Kuperman, W. A., Roux, A., 2005. Ambient noise cross correlation in free space: Theoretical approach, *The Journal of the Acoustical Society of America*, 117(1), 79–84, doi:10.1121/1.1830673.

Rovelli A, Singh SK, Malagnini L, Amato A, Cocco M. 1991. Feasibility of the use of Microtremors in Estimating Site Response during Earthquakes: Some Test Cases in Italy, *Earthquake Spectra*, 7(4), 551-561, doi:10.1193/1.1585643.

Satoh, T., Kawase, H., Shin'ichi, M., 2001. Estimation of S-wave velocity structures in and around the Sendai Basin, Japan, using arrays records of microtremors. *Bulletin of the Seismological Society of America*, 91 (2), 206–218.

Schlupp, A., Sira, C., Maufroy, E., Provost, L., Dretzen, R., Bertrand, E., Beck, E. & Schaming, M., 2021. EMS98 intensities distribution of the “Le Teil” earthquake, France, 11 November 2019 (Mw 4.9) based on macroseismic surveys and field investigations, *Compt. Rendus. Geosci.*, 353(S1), 465-492, doi:10.5802/crgeos.88.

Semblat, J.F., Kham, M., Parara E., Bard, P.-Y., Pitilakis, K., Makra, K. & Raptakis, D., 2005. Seismic wave amplification: Basin geometry vs soil layering, *Soil Dyn. Earthqu. Eng.*, 25(7-10), 529-538, doi:10.1016/j.soildyn.2004.11.003.

Seo, K., 1997. Comparison of measured microtremors with damage distribution. *JICA*, Research and Development Project on Earthquake Disaster Prevention.

SESAME team, 2004. Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations: measurements, processing and interpretation, SESAME European research project.

Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M.C., 2005. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.*, 32, L03310, doi:10.1029/2004GL021862.

Seats, K. J., Lawrence, J. F., & Prieto, G. A., 2012. Improved ambient noise correlation functions using Welch's method, *Geophysical Journal International*, 188(2), 513-523.

Shapiro, N. M., Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, 31, L07614, doi:10.1029/2004GL019491.

Shapiro, N. M., Campillo, M., Stehly, L., Ritzwoller, M., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, 307(5715), 1615–1618, doi:10.1126/science.1108339.

Snieder, R., 2004. Extracting the green's function from the correlation of coda waves: A derivation based on stationary phase, *Phys. Rev. E*, 69, doi:10.1103/PhysRevE.69.046610.

Somerville P., Collins N., Graves R., Pitarka A., 2004. An engineering ground motion model for basin generated surface waves, in *Proceedings of the 13th World Conference on Earthquake Engineering*, Vancouver, B.C., Canada, Paper Number: 515.

Stehly, L., Campillo, M., Shapiro, N., 2006. A study of the seismic noise from its long-range correlation properties. *Journal of geophysical research*, 111, B10306, doi:10.1029/2005JB004237.

Stehly, L., Campillo, M., Froment, B., Weaver, R.L., 2008. Reconstructing green's function by correlation of the coda of the correlation (c3) of ambient seismic noise. *Journal of Geophysical Research: Solid Earth*, 113(B11). ISSN 2156-2202, doi:10.1029/2008JB005693.

Stehly, L. et al., 2009. Tomography of the alpine region from observations of seismic ambient noise. *Geophysical Journal International*, 178(1), 338–350, ISSN 1365-246X, doi:10.1111/j.1365-246X.2009.04132.x.

Stehly, L. & Boué, P., 2017. On the interpretation of the amplitude decay of noise correlations computed along a line of receivers, *Geophys. J. Int.*, 209, 358-372, doi:10.1093/gji/ggx021.

Steidl, J.H., Tumarkin, A.G. & Archuleta, R.J., 1996. What is a reference site?, *Bull. seism. Soc. Am.*, 86(6), 1733-1748, doi:10.1785/BSSA0860061733.

Steinmann, R., Seydoux, L., Beaucé, É., & Campillo, M., 2022. Hierarchical exploration of continuous seismograms with unsupervised learning, *Journal of Geophysical Research: Solid Earth*, 127, doi:https://doi.org/10.1029/2021JB022455.

Theodoulidis et al. 2018. Basin effects on ground motion: the case of a high resolution experiment in Cephalonia (Greece), *Bull Earthquake Eng*, 16, 529–560, doi:https://doi.org/10.1007/s10518-017-0225-4.

Thouvenot, F., Jenatton, L. & Gratier, J., 2009. 200-m-deep earthquake swarm in Tricastin (lower Rhône Valley, France) accounts for noisy seismicity over past centuries. *Terra Nova*, 21(3), 203-210, doi:10.1111/j.1365-3121.2009.00875.x.

Traverso F., Vernazza G. & Trucco A., 2012. Simulation of non-White and non-Gaussian underwater ambient noise, *Oceans - Yeosu*, Yeosu, Korea (South), 1-10, doi:10.1109/OCEANS-Yeosu.2012.6263385.

Triantafyllidis, P., Theodoulidis, N., Savvaidis, A., Papaioannou, C. & Dimitriu, P., 2006. Site effects estimation using earthquake and ambient noise data: the case of lefkas town (w. Greece), in *Proceedings of the 13th European Conference on Earthquake Engineering (13ECEE)*, Geneva, Switzerland, 2006 September 3-8, Paper Number: 1249.

Tsai, V. C., 2009. On establishing the accuracy of noise tomography travel time measurements in a realistic medium, *Geophysical Journal International*, 178(3), 1555–1564, doi:10.1111/j.1365-246X.2009.04239.x

Tsai, V.C., 2011. Understanding the amplitudes of noise correlation measurements, *J. Geophys. Res. Solid Earth*, 116, B09311, doi:10.1029/2011JB008483.

Van der Neut, J., Bakulin A., 2009. Estimating and correcting the amplitude radiation pattern of a virtual source, *Geophysics*, 74(2), S127–S136, doi:10.1190/1.3073003.

Viens L, Laurendeau A, Bonilla LF, Shapiro NM, 2014. Broad-band acceleration time histories synthesis by coupling low-frequency ambient seismic field and high-frequency stochastic modelling, *Geophys J Int*, 199, 1784–1797, doi:10.1093/gji/ggu362.

Viens L, Miyake H, Koketsu K, 2015. Long-period ground motion simulation of a subduction earthquake using the offshore-onshore ambient seismic field, *Geophys Res Lett*, 42, 5282– 5289, doi: 10.1002/2015GL064265.

Viens, L., Koketsu, K., Miyake, H., Sakai, S. & Nakagawa, S., 2016. Basin-scale Green's functions from the ambient seismic field recorded by MeSO-net stations, *J. Geophys. Res. Solid Earth*, 121, 2507-2520, doi:10.1002/2016JB012796.

Viens, L. & Iwata, T., 2020. Improving the Retrieval of Offshore-Onshore Correlation Functions With Machine Learning, *J. Geophys. Res. Solid Earth*, 125(8), doi:10.1029/2020JB019730.

Xu, Jin & Wang, Wei & Gao, Jinghuai & Chen, Wenchao., 2013. Monochromatic Noise Removal via Sparsity-Enabled Signal Decomposition Method, *Geoscience and Remote Sensing Letters*, IEEE. 10, 533-537, doi:10.1109/LGRS.2012.2212271.

Yamamoto, H., 2000. Estimation of shallow S-wave velocity structures from phase velocities of Love- and Rayleigh-waves in microtremors, in *Proceedings of the 12th World Conference on Earthquake Engineering*, Auckland, New Zealand.

Yamanaka, H., Takemura, M., Ishida, H., Niwa, M., 1994. Characteristics of long-period microtremors and their applicability in exploration of deep sedimentary layers. *Bulletin of the Seismological Society of America*, 84 (6), 1831–1841.

Zhao, B.M., Horike, M. & Takeuchi, Y., 1998. Reliability of estimation of seismic ground characteristic by microtremor observation, in *Proceedings of the 11th European Conference on Earthquake Engineering (11ECEE)*, Paris, France, 1998 September 6-11, pp. 6-11.

Annexe

A Contrôle qualité des données

Les figures A1 à A16 montrent 16 pages issues du catalogue de contrôle qualité des données (Froment *et al.* 2022). Les observations sont présentées pour la composante Z. Les nodes représentés sont respectivement ADHE, PAUL, BOLL, St-P-T-C, S-M-A, P-S-E, LPD, D-M, SN, V1, V2, V3, V4, V5, V6, N7.

B Amplitude de l'ANIRF entre 3 et 6 Hz

Les figures B1 à B4 représentent l'amplitude de l'ANIRF (entre 3 et 6 Hz) sur le réseau de nodes pour différents temps d'arrivée pour les composantes TT et ZZ. Les sources virtuelles utilisées sont V2 et V4. Notons l'absence d'information sur la partie ouest du réseau en raison du faible rapport signal-à-bruit (SNR < 3) des fonctions ANIRF calculées entre V4 et les récepteurs situés dans cette zone.

C Rapport entre l'autocorrélation à chaque source virtuelle VX et l'autocorrélation à ADHE

La figure C1 donne un aperçu de la variabilité de l'amplitude enregistrée entre les 7 sources virtuelles. Pour les 3 composantes (TT, RR, ZZ), l'autocorrélation à chaque source virtuelle VX a été normalisée par l'autocorrélation à ADHE :

$$Amplification = \frac{\{ \langle |U_{VX}(f)|^2 \rangle \}}{\{ \langle |U_{ADHE}(f)|^2 \rangle \}} ,$$

avec $|U_{VX}(f)|$ et $|U_{ADHE}(f)|$ respectivement le spectre d'amplitude du signal à VX (X allant de 1 à 6) et ADHE, $\langle . \rangle$ la moyenne arithmétique sur les fenêtres de temps synchrones entre VX et ADHE et $\{ . \}$ une fonction de lissage.

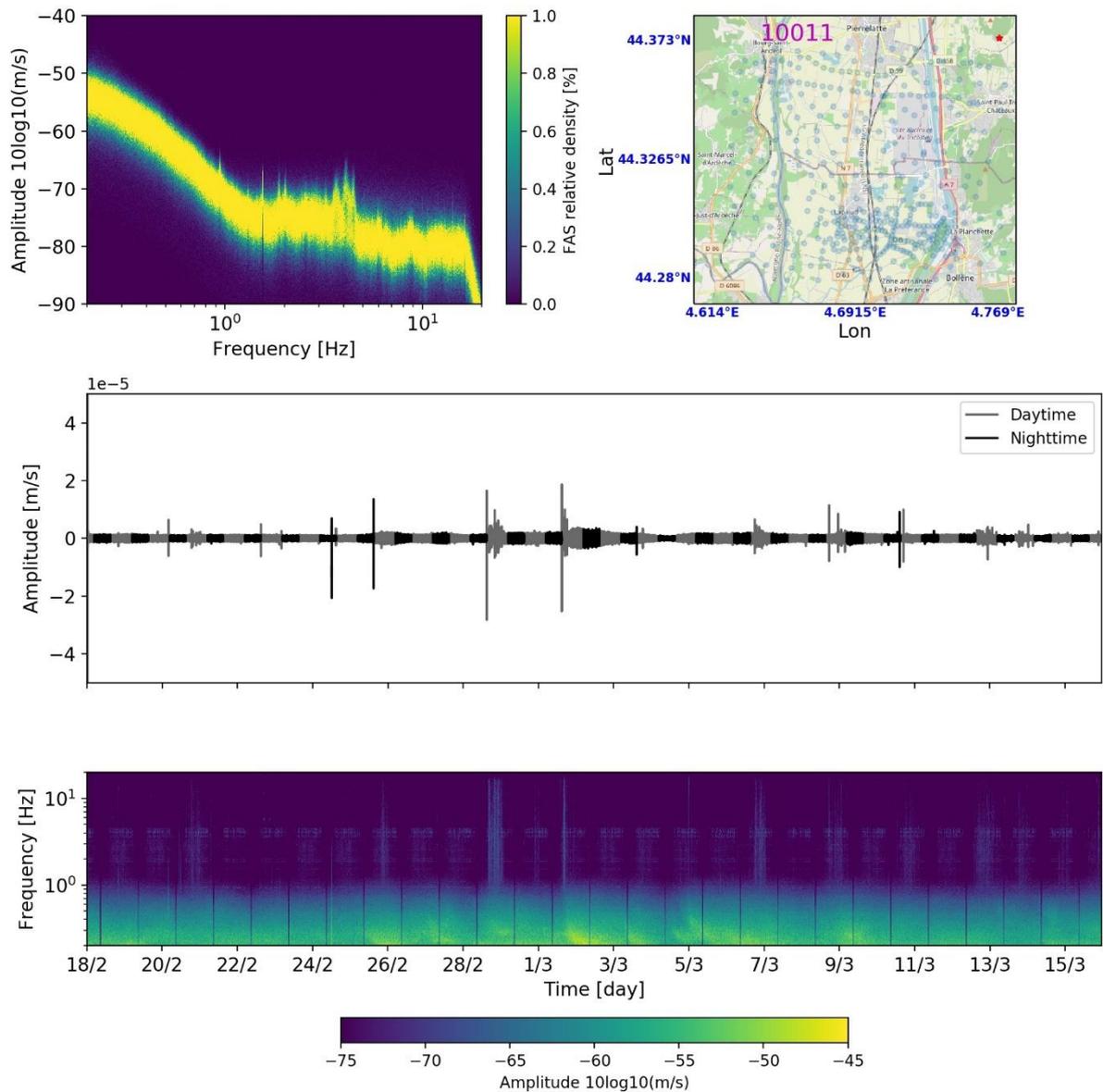


Figure A1 : Contrôle qualité des données à ADHE. Panneau en haut à gauche : distribution des spectres d'amplitude sur la base de segments temporelles de 10 min. Panneau en haut à droite : position du node en question. Panneau central supérieur : trace temporelle filtrée entre 0.2 et 20 Hz. Panneau central inférieur : spectrogramme. Les observations sont présentées pour la composante Z sur l'ensemble de l'acquisition.

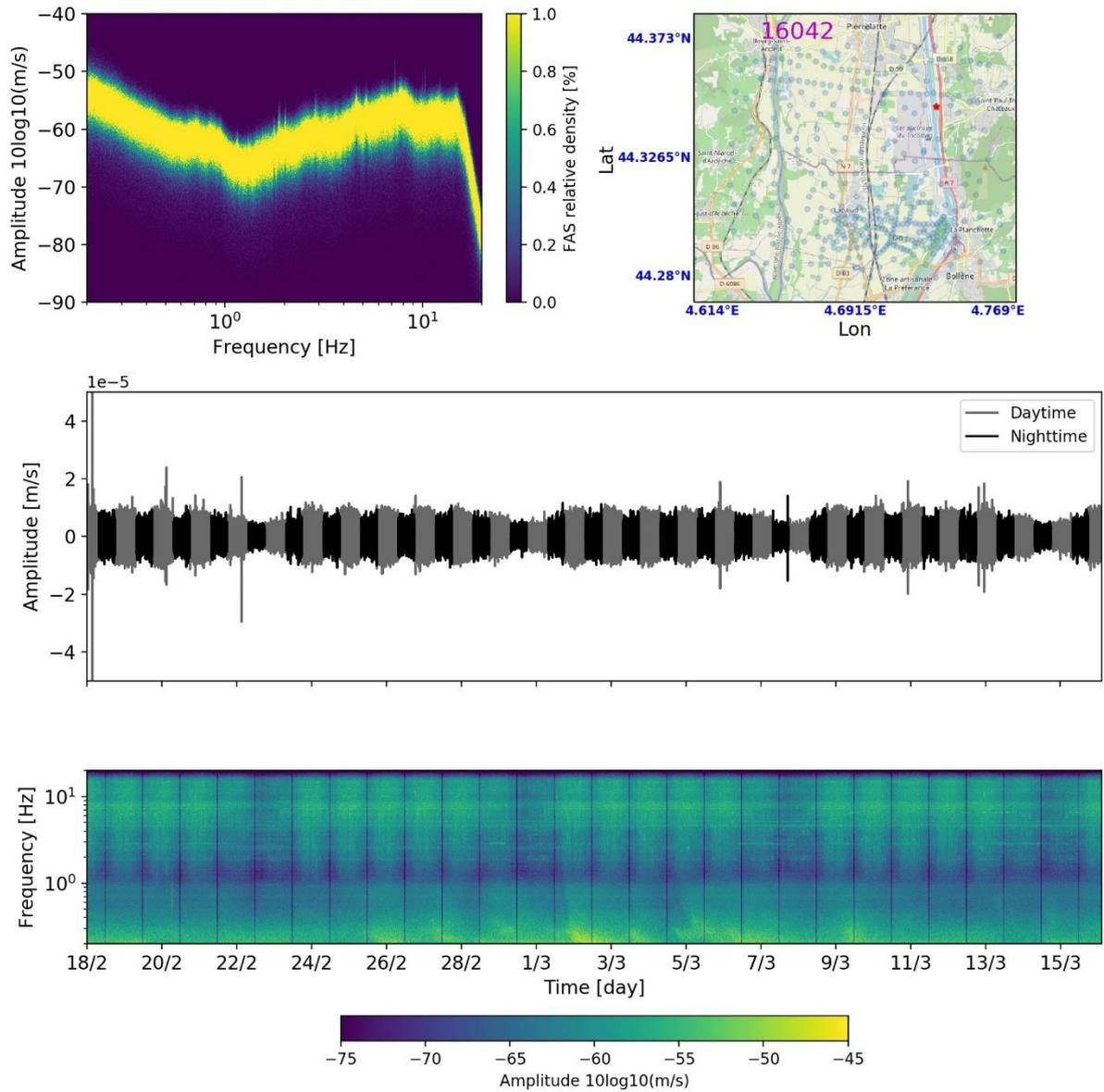


Figure A2 : Idem pour le node PAUL.

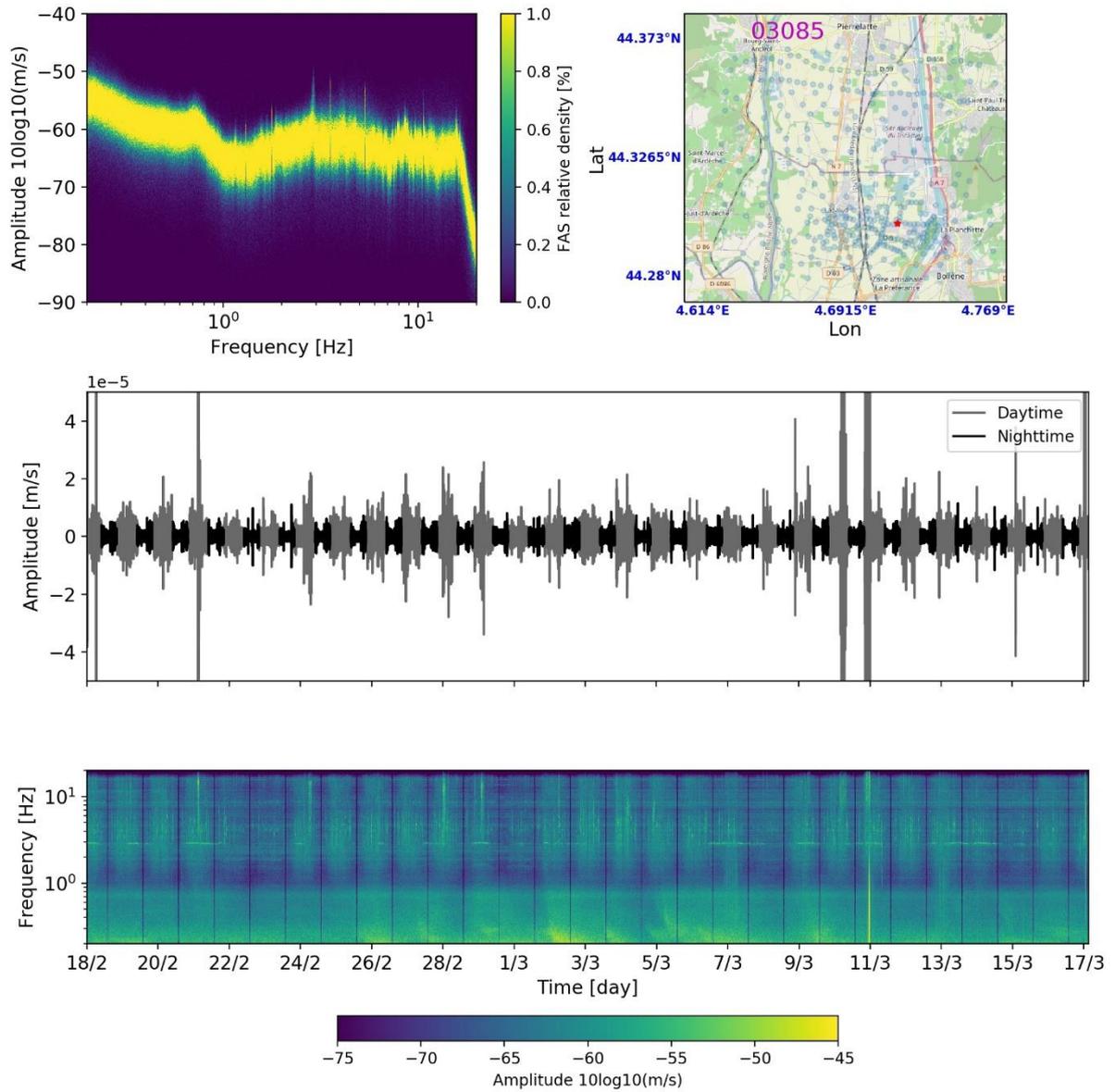


Figure A3 : Idem pour le node BOLL.

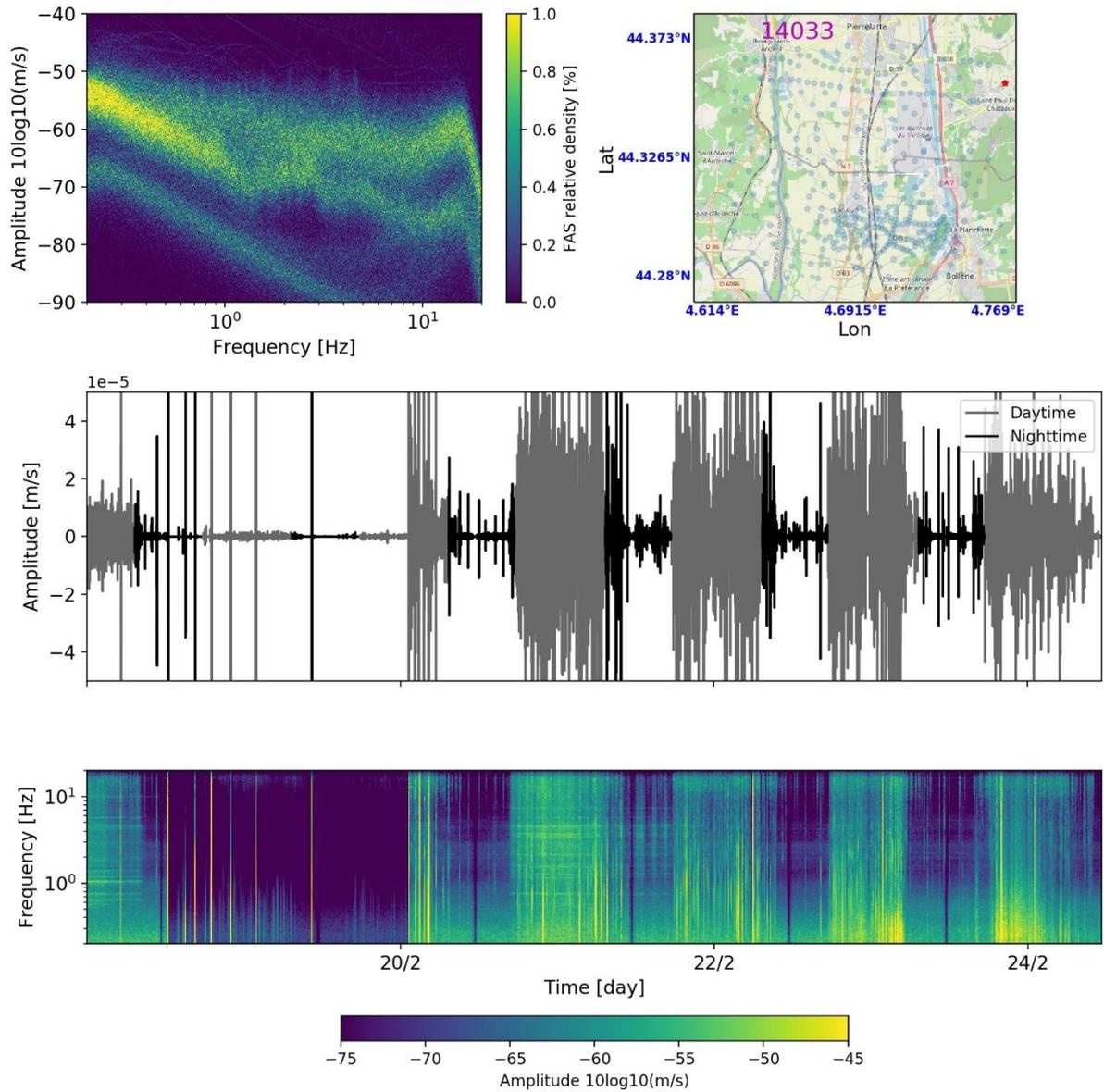


Figure A4 : Idem pour le node St-P-T-C.

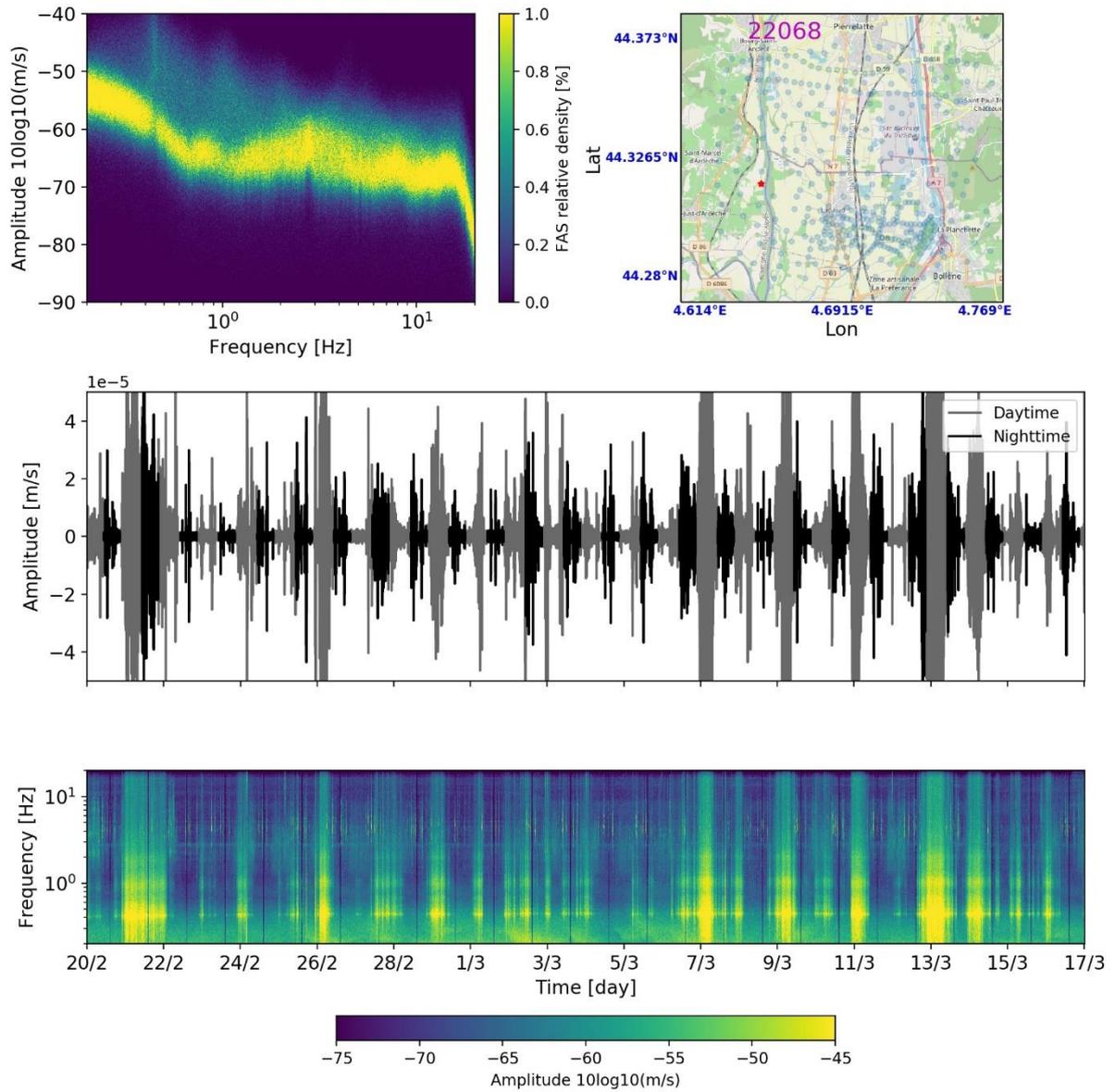


Figure A5 : Idem pour le node S-M-A.

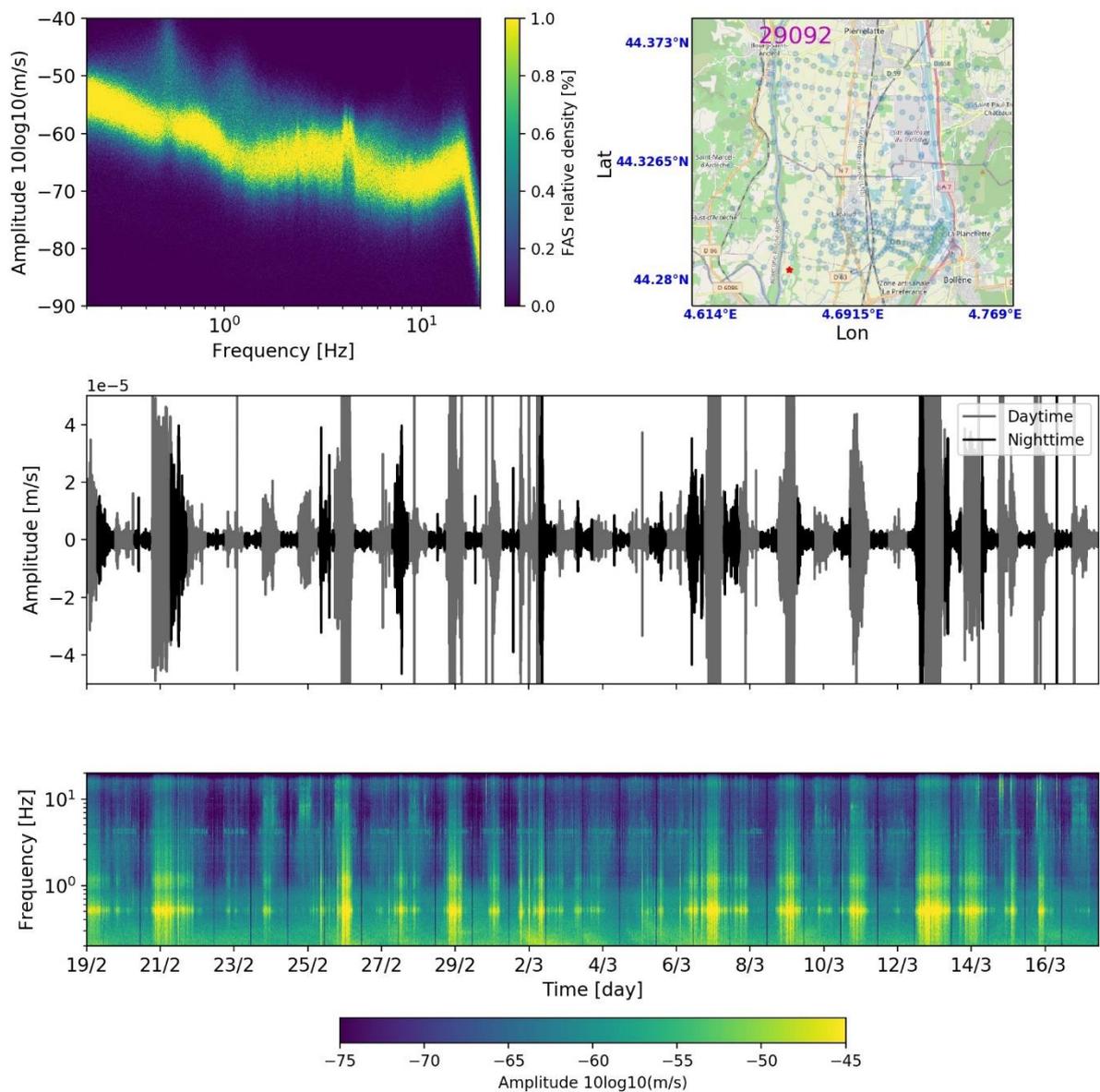


Figure A6 : Idem pour le node P-S-E.

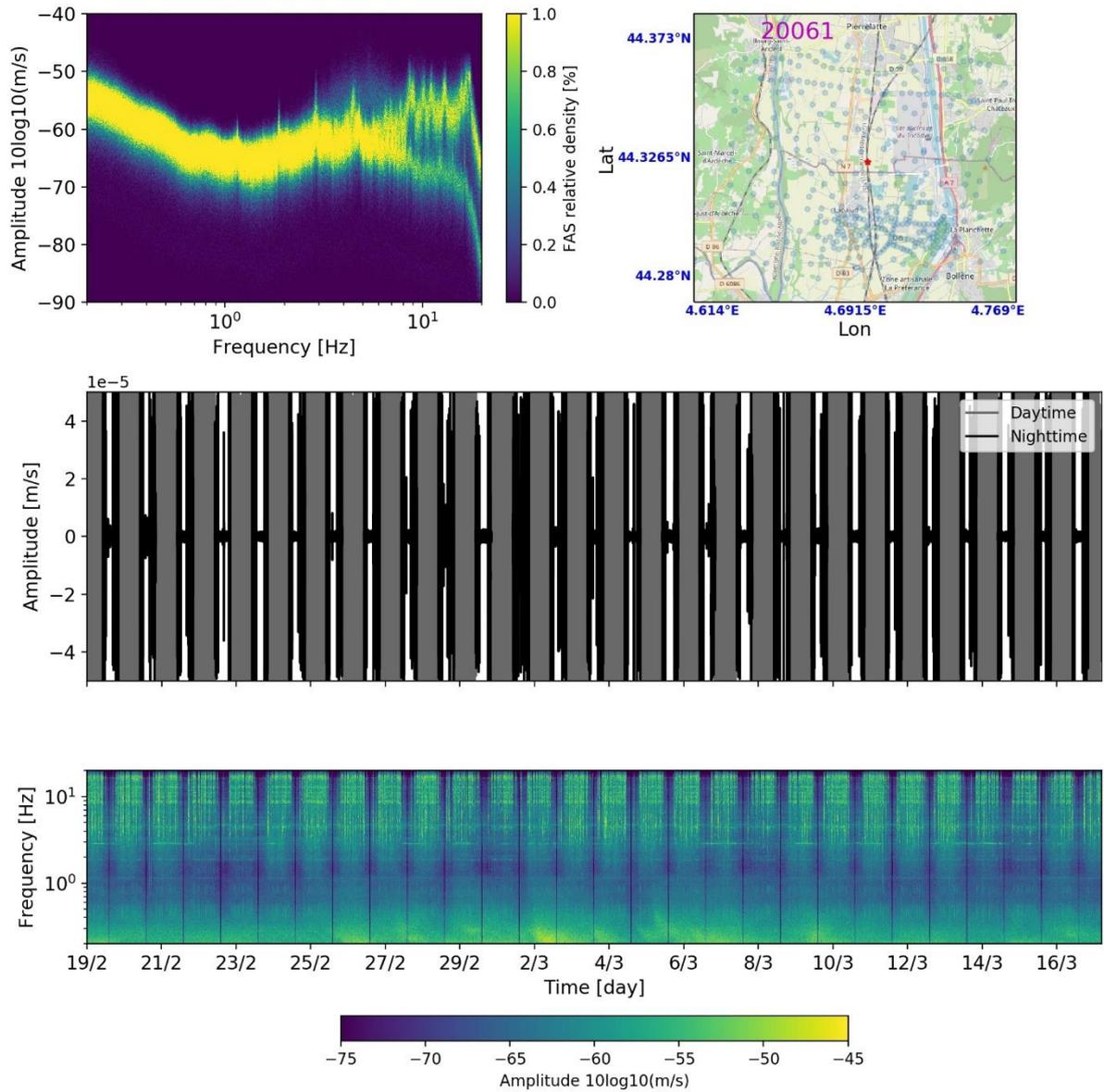


Figure A7 : Idem pour le node LPD.

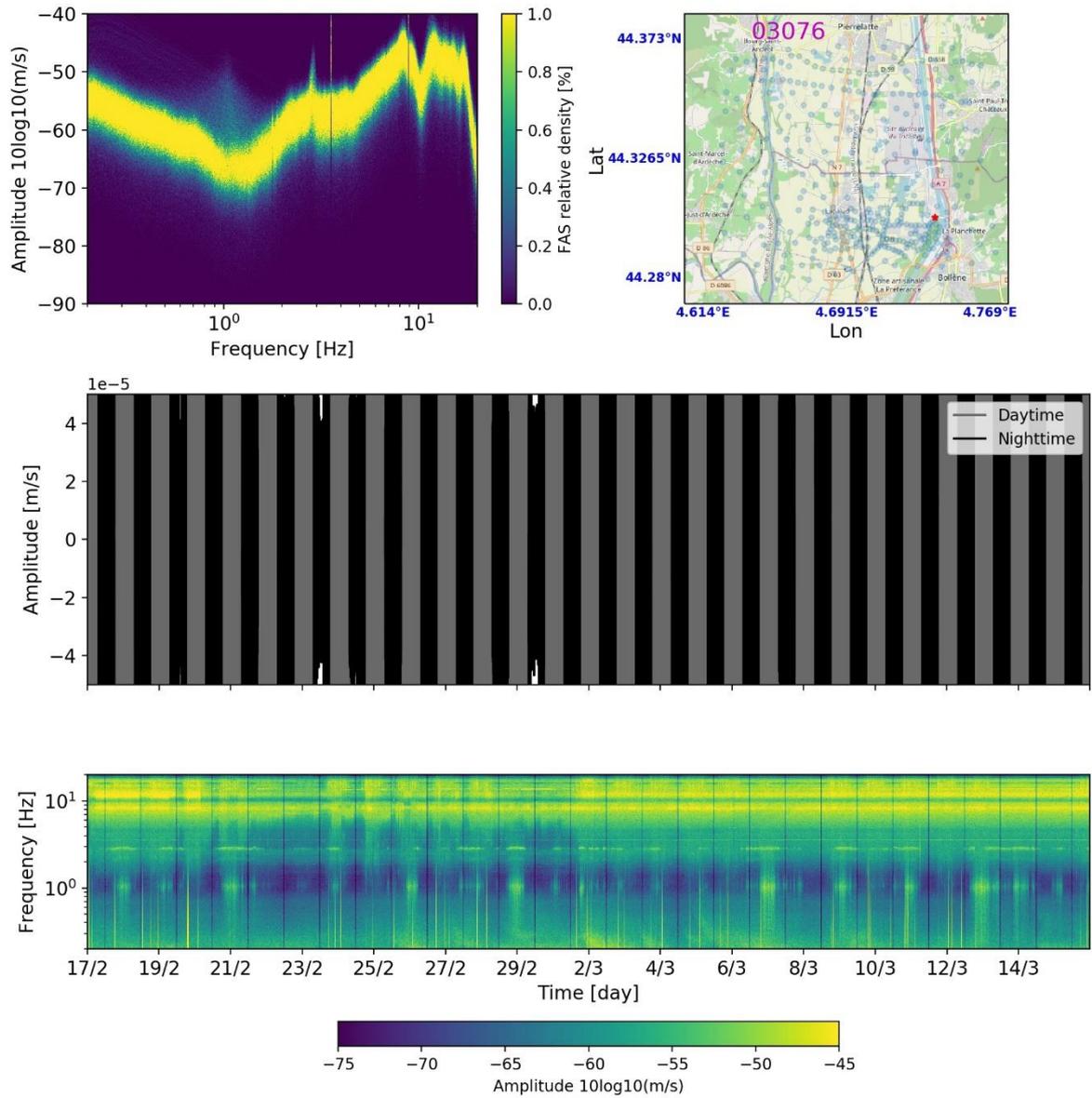


Figure A8 : Idem pour le node D-M.

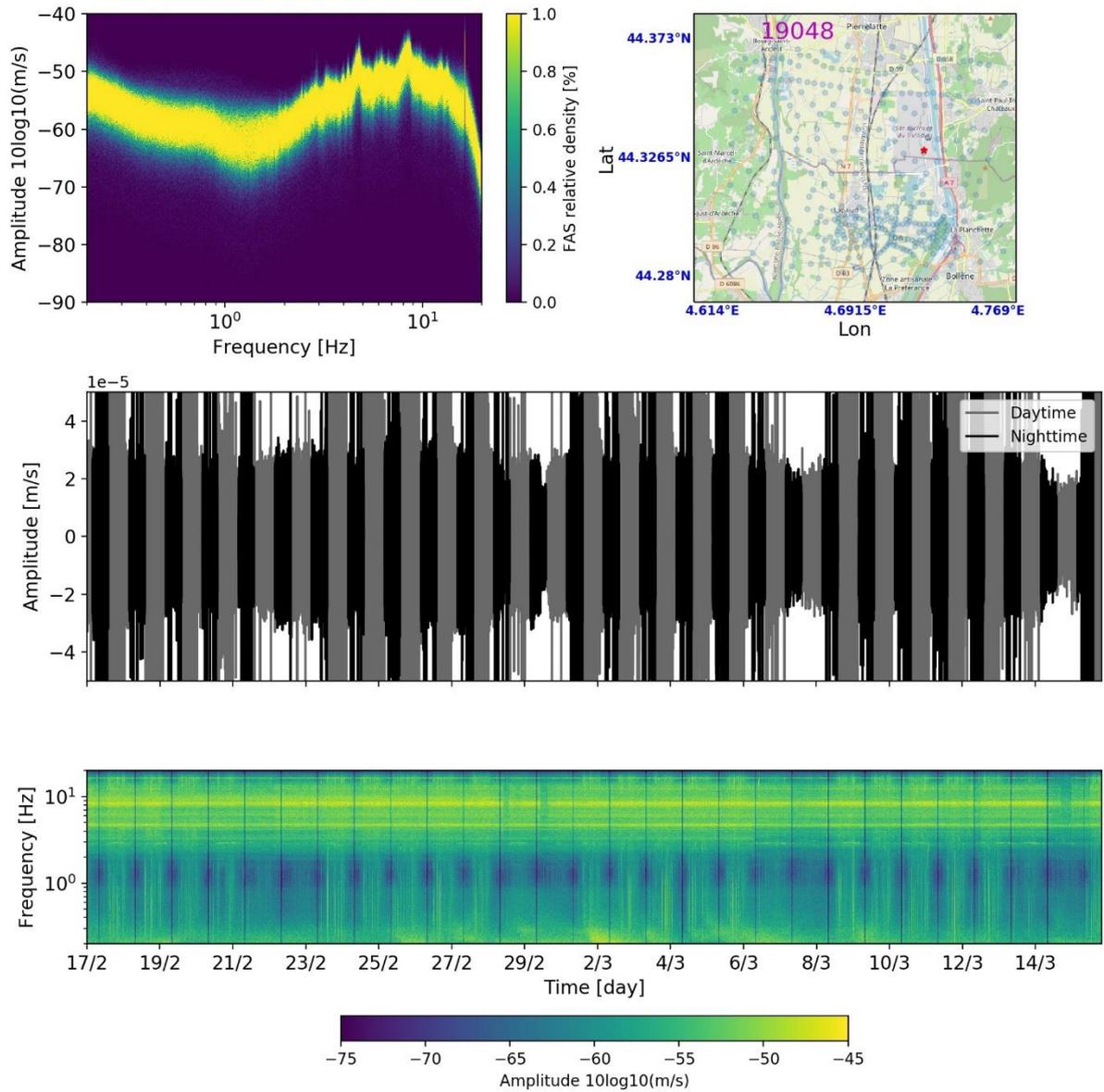


Figure A9 : Idem pour le node SN.

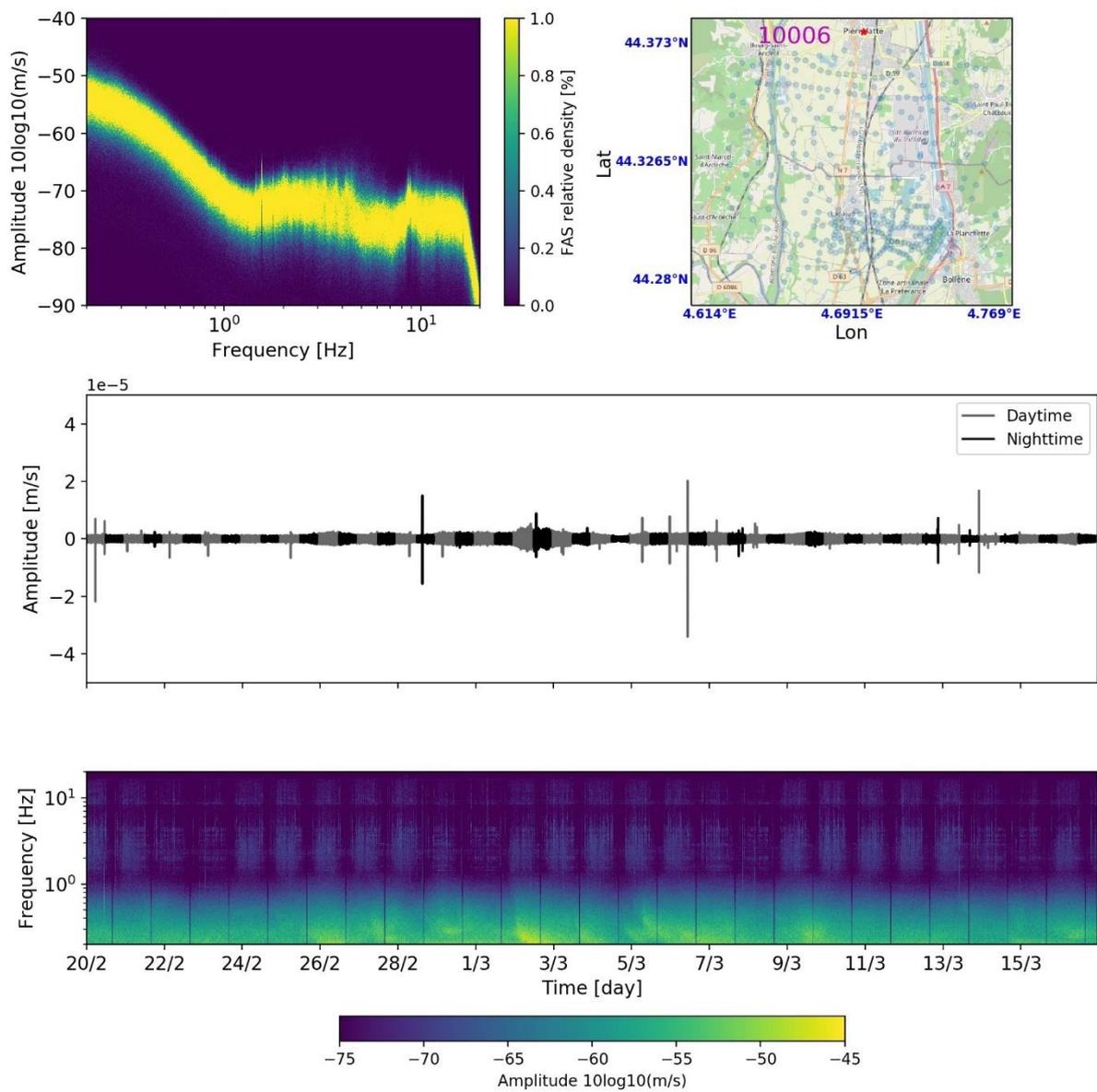


Figure A10 : Idem pour le node V1.

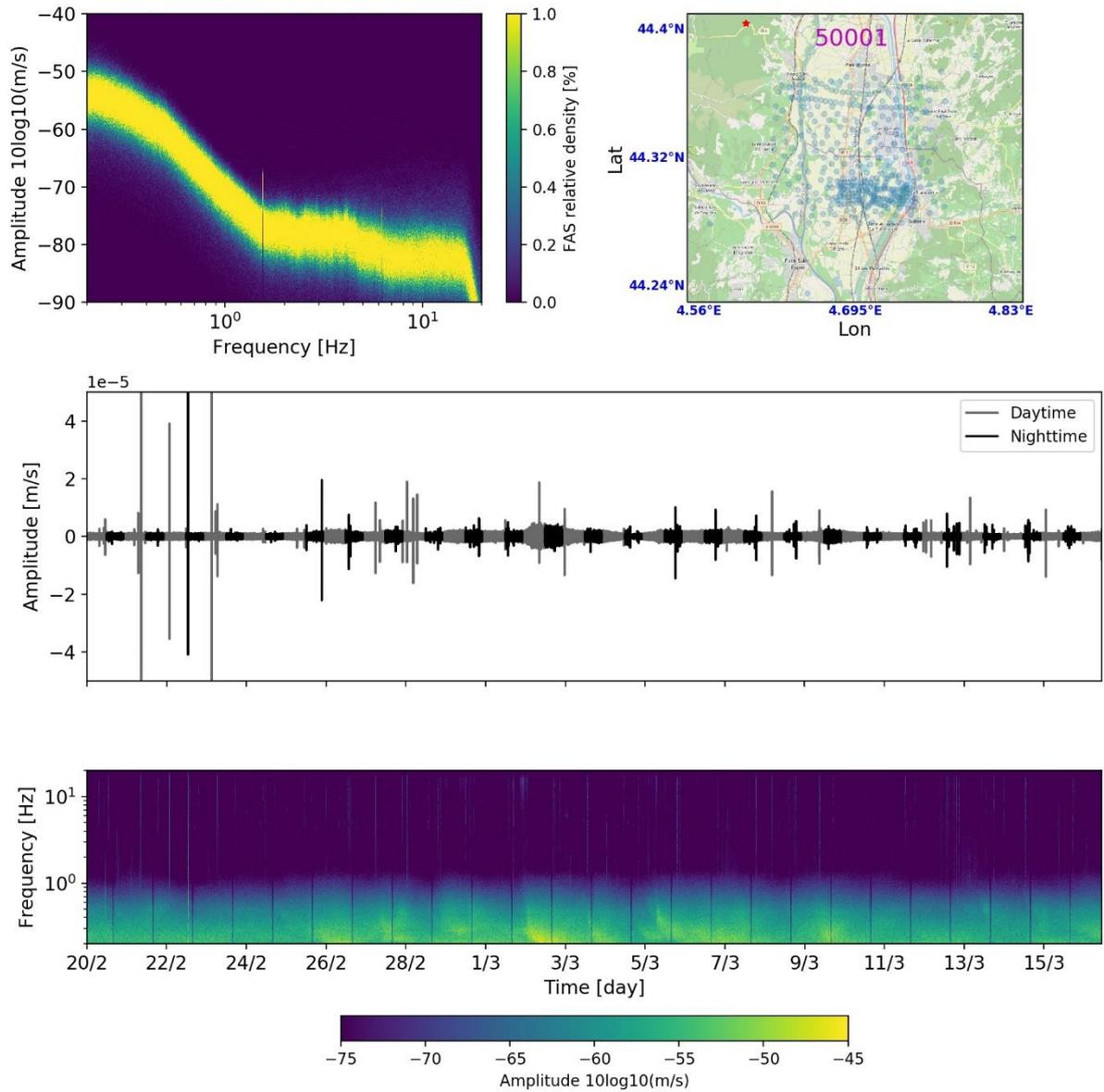


Figure A11 : Idem pour le node V2.

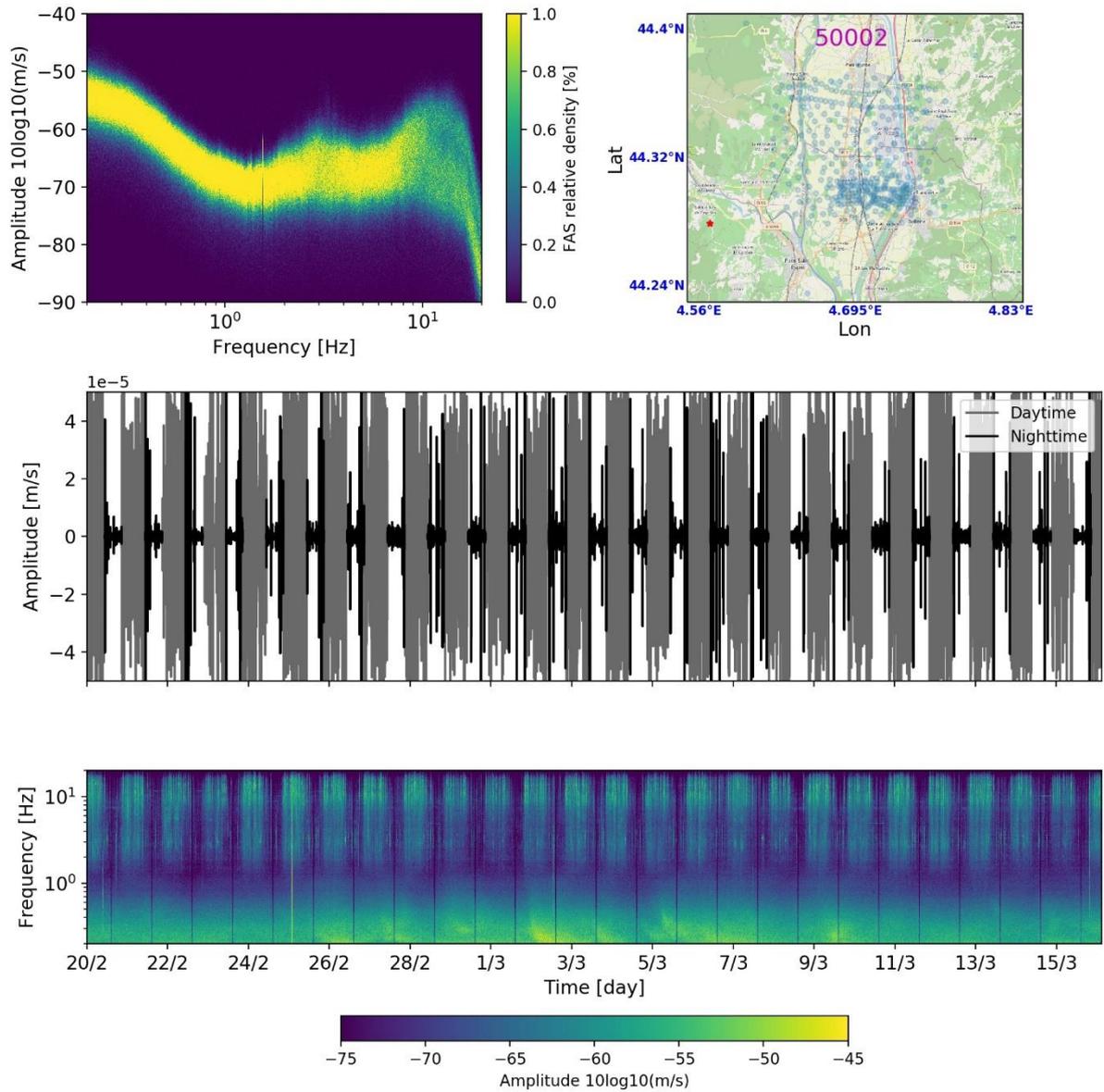


Figure A12 : Idem pour le node V3.

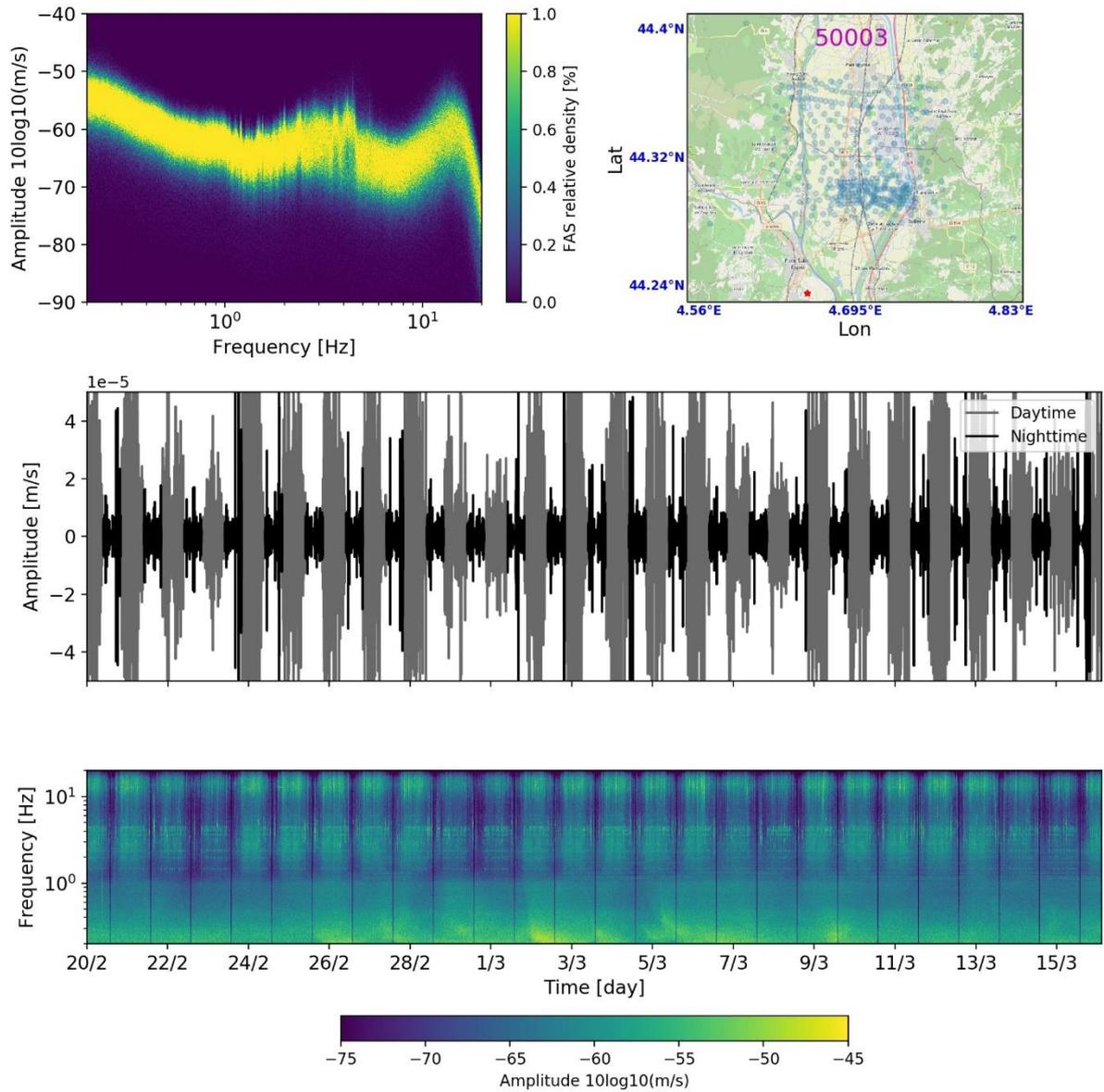


Figure A13 : Idem pour le node V4.

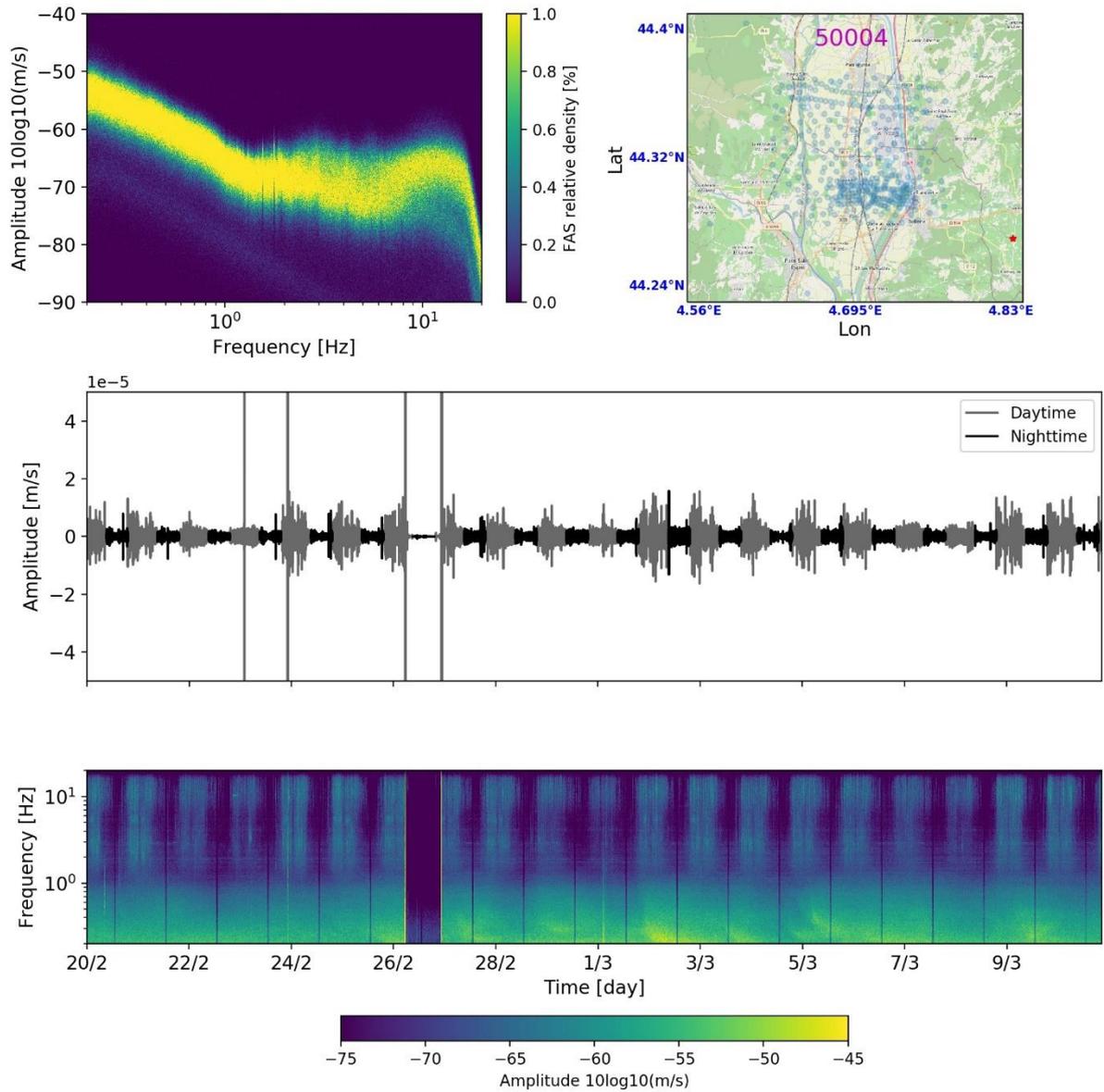


Figure A14 : Idem pour le node V5.

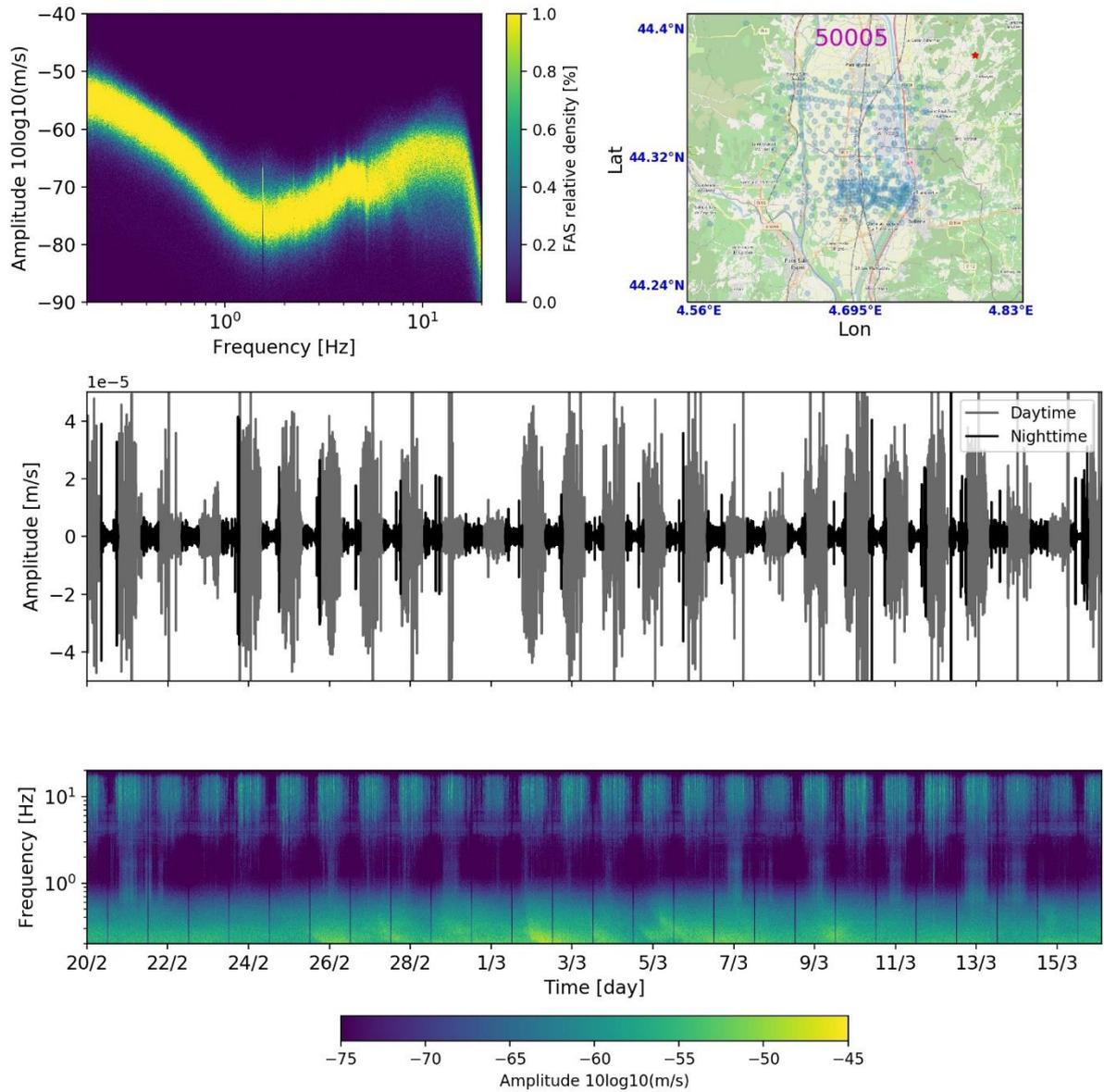


Figure A15 : Idem pour le node V6.

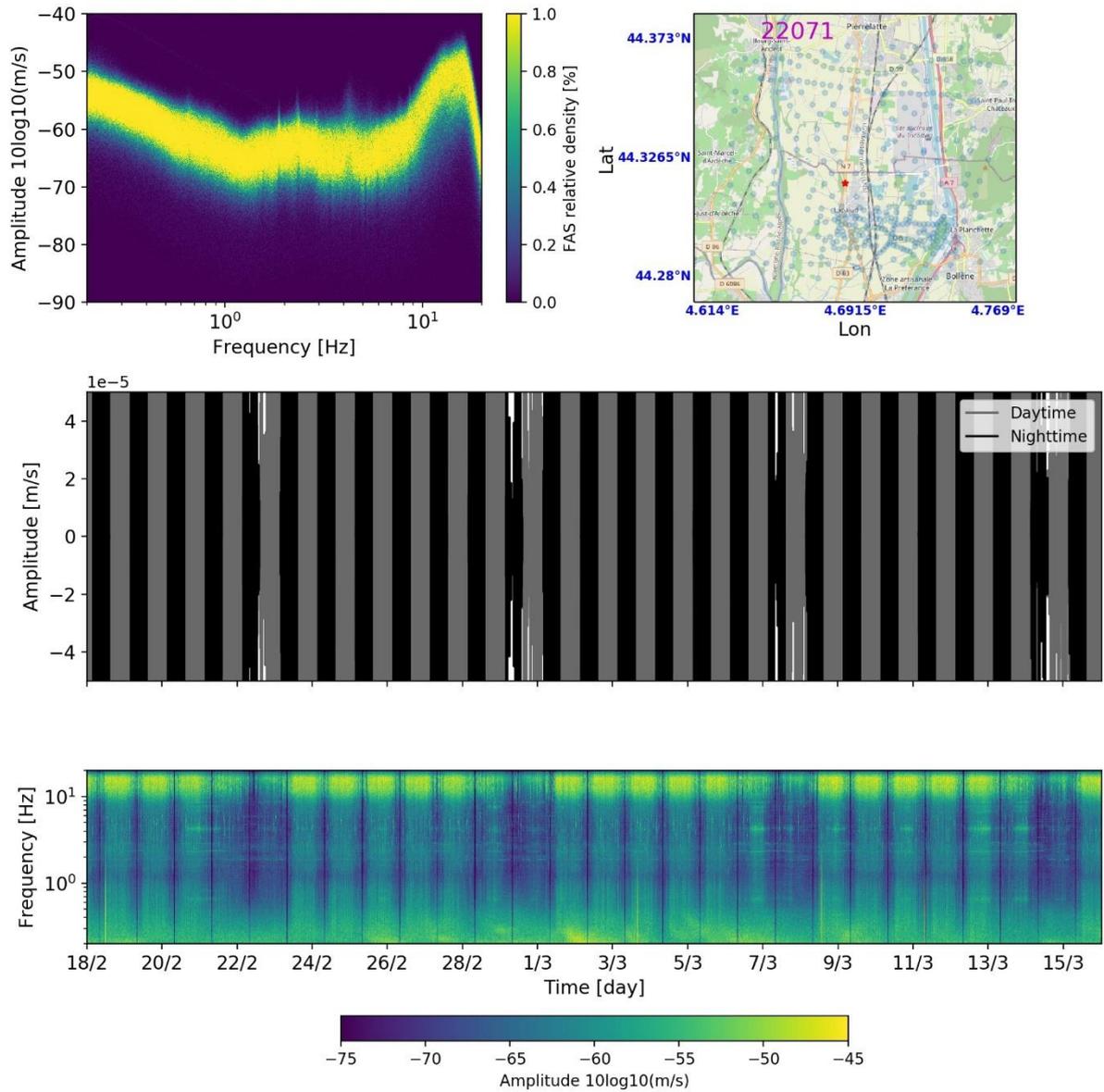


Figure A16 : Idem pour le node N7.

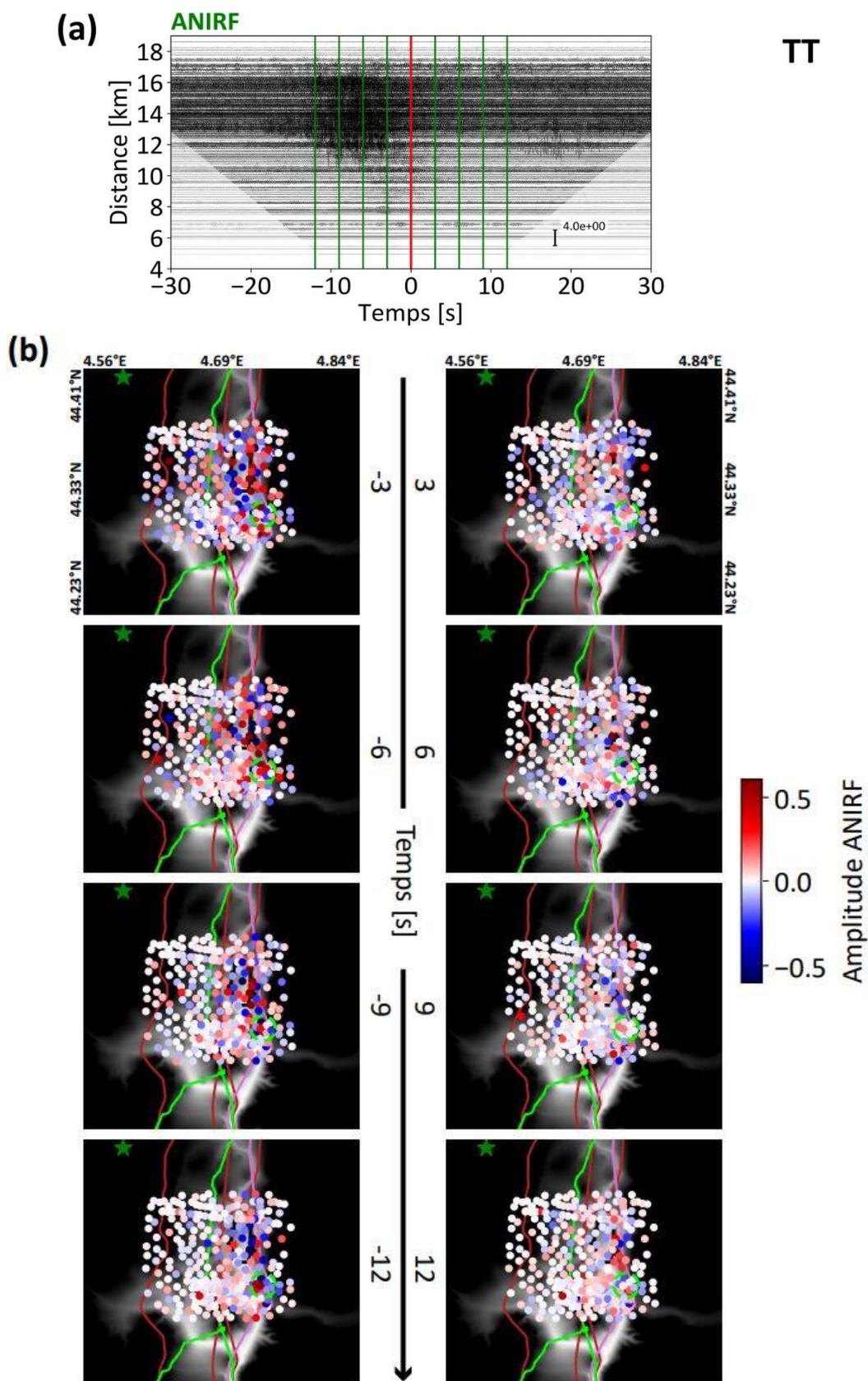


Figure B1 : (a) Section temporelle de l'ANIRF. (b) Amplitude de l'ANIRF (entre 3 et 6 Hz) sur le réseau de nodes pour différents temps d'arrivées pour la composante TT. La source virtuelle utilisée est V2.

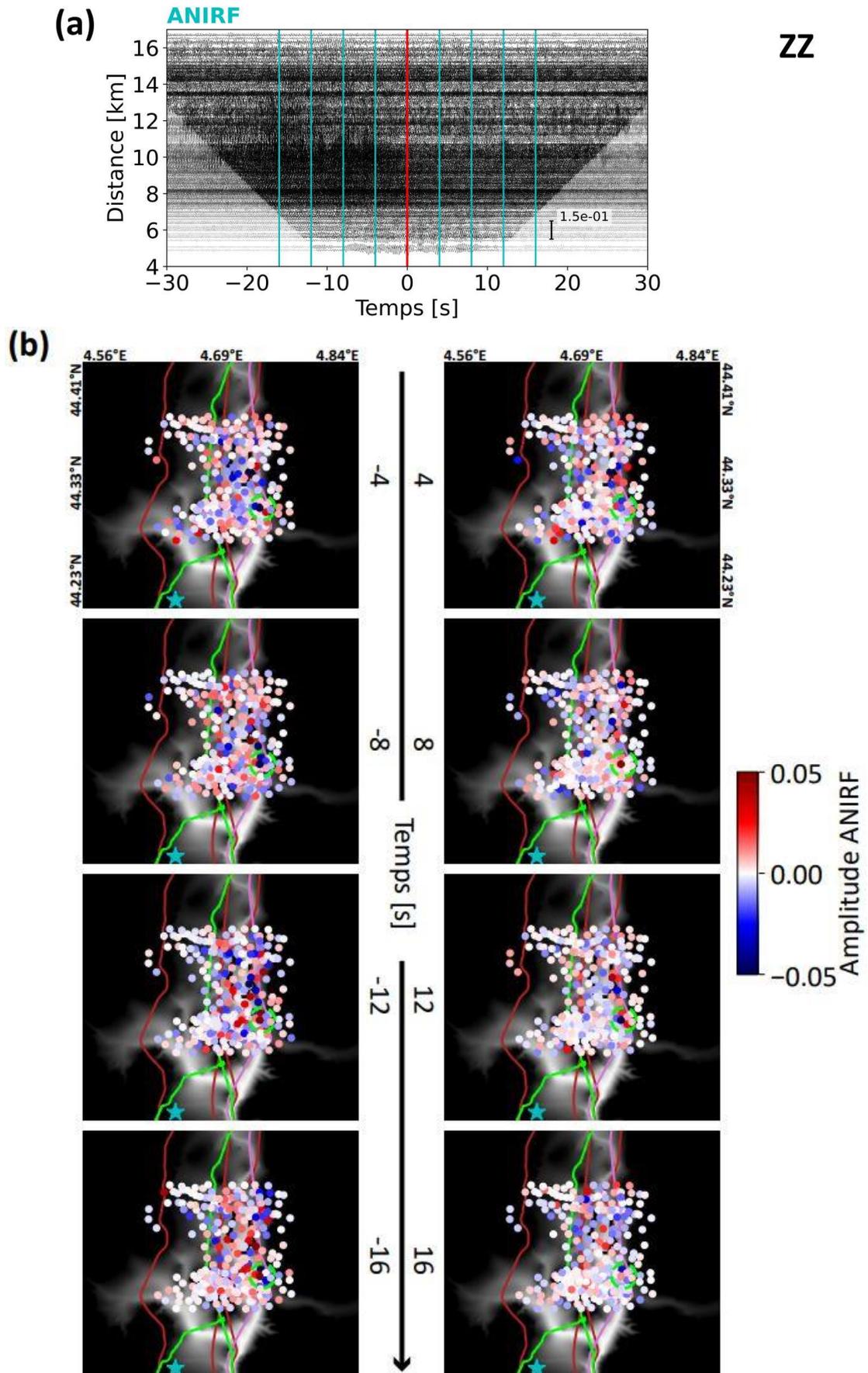


Figure B4 : Idem que la Figure B3 pour la composante ZZ.

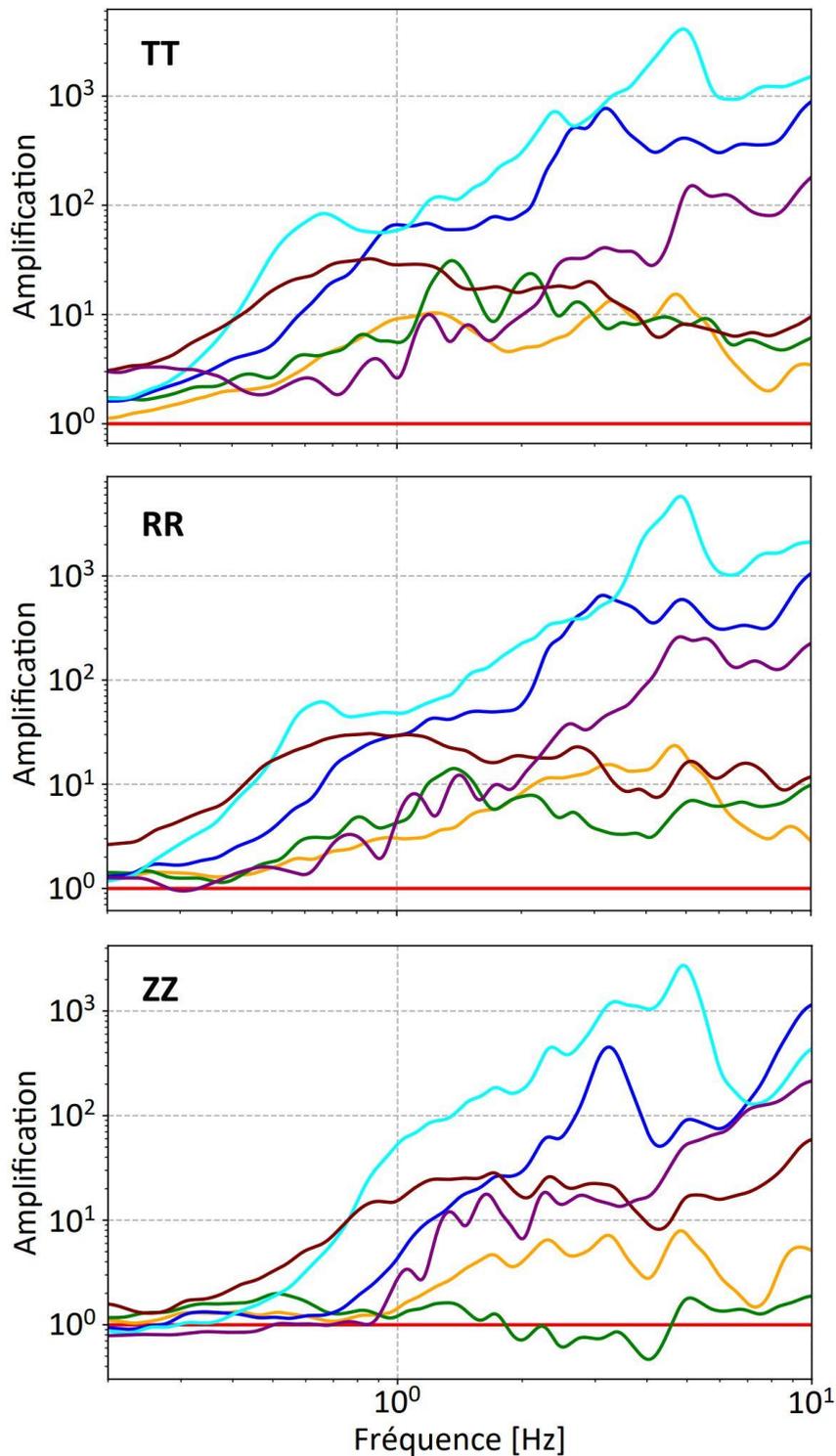


Figure C1: Rapport entre l'autocorrélation à chaque source virtuelle VX et l'autocorrélation à ADHE pour les 3 composantes (TT, RR, ZZ). Le code couleur correspond à celui utilisé pour représenter chaque source virtuelle VX sur la figure 5.1.