



HAL
open science

Caractérisation de la sismicité naturelle d'une région continentale stable : application au Massif armoricain

Céline Hourcade

► **To cite this version:**

Céline Hourcade. Caractérisation de la sismicité naturelle d'une région continentale stable : application au Massif armoricain. Sciences de la Terre. Nantes Université, 2023. Français. NNT : 2023NANU4066 . tel-04653909

HAL Id: tel-04653909

<https://theses.hal.science/tel-04653909>

Submitted on 19 Jul 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 DE DOCTORAT

NANTES UNIVERSITE

ECOLE DOCTORALE N° 596
Matière, Molécules, Matériaux et Géosciences
Spécialité : *Sciences de la Terre et des Planètes*

Par

Céline HOURCADE

Caractérisation de la sismicité naturelle d'une région continentale stable : application au Massif armoricain

Thèse présentée et soutenue à Nantes, le 18 octobre 2023

Unité de recherche : Laboratoire de Planétologie et Géosciences, UMR CNRS 6112

Rapporteurs avant soutenance :

Alessia Maggi Professeur des universités, Université de Strasbourg
Stéphane Mazzotti Professeur des universités, Université de Montpellier

Composition du Jury :

Président : Stéphane Mazzotti Professeur des universités, Université de Montpellier

Examineurs : Marc Grunberg Ingénieur de recherche CNRS, Université de Strasbourg
 Marie Calvet Physicienne CNAP, Université Toulouse III Paul Sabatier

Dir. de thèse : Éric Beucler Professeur des universités, Nantes Université

Co-enc. de thèse : Mickaël Bonnin Physicien-adjoint CNAP, Nantes Université

Invité(s)

Jean Letort Physicien-adjoint CNAP, Université de Toulouse III Paul Sabatier
Christel Tiberi Directrice de recherche CNRS, Université de Montpellier
Yoann Cano Agent CEA, Commissariat d'Énergie Atomique et d'Énergie Alternatives

Remerciements

Les travaux de recherche présentés dans ce manuscrit de thèse ont été rendus possibles grâce à la contribution et au soutien de nombreuses personnes, auquel je souhaite exprimer ma sincère reconnaissance. Je tiens tout d'abord à adresser mes remerciements aux membres du jury **Marie Calvet** et **Marc Grunberg** d'avoir accepté d'évaluer mon travail et aux rapporteurs **Alessia Maggi** et **Stéphane Mazzotti** d'avoir pris le temps d'étudier ce manuscrit.

J'adresse mes plus vifs remerciements à mes encadrants de thèse, **Éric Beucler** et **Mickaël Bonnin**, pour leur soutien inconditionnel et leur disponibilité constante tout au long de cette aventure de recherche. Merci **Mickaël** de m'avoir fait confiance et accompagné, depuis mon stage de Master 1 pour enfin co-encadrer ma thèse. Merci **Éric** pour ton soutien constant et les nombreuses heures que nous avons passées à explorer les pistes de réflexion dès que des problèmes se sont présentés. Je suis tellement reconnaissante d'avoir pu bénéficier de votre encadrement et de vos conseils éclairés, qui ont véritablement façonné ma compréhension du domaine et m'ont permis d'en arriver là où j'en suis aujourd'hui. Vos influences positives resteront gravées dans mon parcours et je vous suis infiniment reconnaissante pour votre soutien. Je souhaite que chacun puisse bénéficier d'un encadrement aussi bienveillant, rassurant et instructif.

Je souhaite également remercier **Clément Perrin**, **Antoine Mocquet** et **Yann Capdeville**, pour leurs conseils et leurs idées durant nos réunions « sismo » du vendredi. Je tiens également à remercier toutes les personnes qui participent à l'acquisition des données sismologiques au sein du LPG et de l'OSUNA, en particulier **Damien Fli-giel**.

Je remercie **Christel Tiberi** et **Jean Letort**, membres de mon comité de suivi, pour leur écoute et leur bienveillance tout au long de cette thèse. Merci **Jean** de m'avoir invitée pour mon tout premier séminaire à l'OMP à Toulouse!

Ces trois années n'auront pas été les mêmes sans toutes les rencontres que j'ai pu

faire aux cours des différentes conférences et workshops. Merci à **Benoit Derode** pour nos discussions stimulantes et enrichissantes en me raccompagnant à la gare. Merci à tous les doctorantes et doctorants que j'ai pu croiser, **Laeticia, Alisson, Marion**.

Pour tous les membres du Laboratoire, travailler avec vous a été une expérience vraiment gratifiante. Je tiens à exprimer ma sincère reconnaissance envers chacun d'entre vous pour avoir contribué à créer une ambiance chaleureuse et conviviale, perceptible à travers les nombreux événements (petits dej', barbecues). Cette période a été source d'un épanouissement personnel, car j'ai eu la chance de faire la connaissance de personnes exceptionnelles.

Parmi les doctorants (mais pas que) de véritables amitiés se sont naturellement tissées, enrichissant ainsi encore davantage cette aventure humaine et professionnelle. Merci à **Pauline, Sami, Simon, Justine V., Meven, Guillaume, Maëva, Justine G., Anthony, Anna, Axel, Marion, Valentine** et **Kevin** ! Sans vous l'aventure aurait été clairement différente. Nos (trop) nombreuses pauses, nos sorties et nos voyages, particulièrement au Parc Asterix, resteront gravés dans ma mémoire. Je tiens à remercier tout particulièrement mes co-bureaux **Arthur** et **Benoit** ! J'ai eu de la chance de passer mes trois années de thèse avec vous dans le même bureau. Nos heures passées ensemble à discuter, échanger des idées et partager nos avancées ont été parmi les moments les plus précieux de ces années de thèse. Votre soutien et vos conseils ont été inestimables tout au long de cette aventure. Comment oublier notre mascotte bien-aimée, Charlotte la marmotte, qui a su égayer nos journées malgré les nombreux vols du bureau 126 (Meven et Pauline, on vous a vus!).

Je souhaite adresser mes plus chaleureux remerciements à mes amis extraordinaires, **Nathan, Camille, Léa L., Pierre, Valentin, Valériane** et **Bastien**, qui étaient toujours là pour moi. Avec nos soirées jeux de sociétés du jeudi et durant mes répétitions de présentations (où vous compreniez sûrement un mot sur deux), vous avez toujours été présents pour m'encourager. Merci à **Léa C.** d'avoir toujours été là quand je revenais chez mes parents à Limoges.

Bien sûr, cette incroyable aventure n'aurait jamais pu se réaliser sans le soutien indéfectible de ma famille, et je tiens à leur exprimer toute ma gratitude. Merci du fond du cœur **Maman, Papa, Nico** pour tout ce que vous avez fait pour moi et pour vos encouragements constants, même lorsque je doutais et remettait tout en question. Et pour finir, merci à toi **Lucas** pour tout ce que tu m'apportes au quotidien et pour ton soutien constant malgré ces derniers mois difficiles.

Un peu d'histoire...

AU sein des entrailles de notre planète, un phénomène intrigant se déroule depuis des millénaires : les séismes. Ces secousses telluriques, produites par les mouvements des plaques tectoniques, ont toujours captivé l'humanité. Elles suscitent une gamme complexe d'émotions allant de la fascination à la peur, rappelant à l'homme sa position fragile sur cette Terre en constante évolution.

Depuis les temps les plus reculés, l'homme a ressenti ces vibrations troublantes qui ébranlent le sol sous ses pieds. Au départ, ces secousses étaient perçues comme des manifestations divines ou des signes mystérieux envoyés par les forces cosmiques. Au fil des siècles, les légendes et les récits se sont tissés autour de ces phénomènes naturels, transmis de génération en génération.

C'est en Chine ancienne que l'on retrouve les premières traces d'un instrument conçu pour enregistrer les tremblements de terre, appelé Houfeng Didong Yi, traduit littéralement par « instrument pour mesurer les vents saisonniers et les mouvements de la Terre ». Il a été inventé, en 132, par le chinois Zhang Heng, qui a vécu sous la dynastie Han.

Ce sismographe chinois, un dispositif ingénieux, était doté d'un design remarquable : il prenait la forme d'un dragon, symbole de puissance et de sagesse, perché sur un support. Ce dragon abritait des boules de bronze placées dans sa gueule (Fig 1). Lorsqu'un séisme se produisait, les secousses faisaient vibrer le pendule du sismographe, provoquant la chute des boules de bronze. Chaque boule tombée indiquait l'intensité et la direction du séisme, offrant ainsi une lecture tangible des mouvements du sol. Les observateurs, fascinés par cet instrument, cherchaient à décoder les informations transmises par les boules pour mieux comprendre les caprices de la Terre.

Ainsi, au fil du temps, les séismes sont devenus des phénomènes étudiés et scrutés avec attention. Les avancées scientifiques ont permis de développer des sismomètres plus



FIGURE 1 – Réplique du Houfeng Didong Yi. (Source : *Science and Society Picture Library*)

sophistiqués et précis, ouvrant la voie à une meilleure compréhension de ces événements naturels. Les secousses telluriques continuent de nous rappeler l'imprévisibilité et la puissance de notre planète, nous invitant à percer les mystères qui se cachent sous nos pieds.

Sommaire

1	Introduction	1
1.1	Motivations générales	1
1.2	Le Massif armoricain : un contexte géologique polyphasé	5
1.2.1	Orogenèse panafricaine (cadomienne), 650 Ma - 540 Ma	6
1.2.2	Orogenèse varisque (hercynienne), 420 Ma - 290 Ma	8
1.2.3	Événements post-hercynien	9
1.3	Structure générale du Massif armoricain	10
1.3.1	Zone Nord armoricaine	12
1.3.2	Zone Centrale armoricaine	13
1.3.3	Zone Sud armoricaine	13
1.3.4	Domaine varisque du Léon	13
1.4	Apports de la sismologie	14
1.4.1	Profondeur du Moho	14
1.4.2	Tomographie	15
1.4.3	Atténuation des ondes sismiques	17
2	Sismicité d'une région continentale stable	21
2.1	Sismicité historique	23
2.2	Sismicité instrumentale	26
2.2.1	Réseaux de stations permanentes	26
2.2.2	Sismicité diffuse	27
2.3	Magnitude de complétude élevée	31

2.4	Sismicité non tectonique	35
2.5	Champs de contrainte	38
2.6	Vers une meilleure compréhension de cette sismicité?	42
3	Méthodes existantes et apport de l'intelligence artificielle pour la dé-	
	tection des séismes	49
3.1	État de l'art des méthodes classiques	50
3.1.1	Détection par analyse de l'énergie	50
3.1.2	Détection par corrélation de segments temporels	53
3.1.3	Discrimination des événements sismiques	56
3.2	Apport du <i>machine learning</i> dans l'observation des événements sismiques	58
3.2.1	Introduction - Définitions	58
3.2.2	Détection et identification des phases sismiques	66
3.2.3	Discrimination des événements sismiques	70
3.3	Bilan des approches	71
4	Nouvelle approche pour la détection et la discrimination d'événe-	
	ments sismiques de faible magnitude	73
4.1	Détection des signaux de faible amplitude	74
4.1.1	Quelle méthode de détection utiliser?	74
4.1.2	Quels paramètres utiliser pour optimiser la détection?	78
4.2	Discrimination des signaux sismiques des fausses détections	82
4.3	Discrimination entre les événements naturels et anthropiques	90
4.3.1	New CNN-based tool to discriminate anthropogenic from natural low magnitude seismic events	91
4.3.2	Complément : Discrimination à l'échelle de la station et du réseau	119
4.4	Lecture des temps d'arrivée des ondes	121
5	Conséquences sur la compréhension de la sismicité du Massif armo-	
	ricain	131
5.1	Calcul de la magnitude de durée (M_d)	133

5.2	Application à la sismicité armoricaine de 2018 à 2021	138
5.2.1	Détails des différentes étapes	138
5.2.2	Nouvelle carte de sismicité de 2018 à 2021	146
5.3	Études préliminaires sur l'origine de la sismicité	152
5.3.1	Distribution spatiale de la sismicité	152
5.3.2	Facteurs environnementaux	159
5.3.3	Bilan des éléments permettant de caractériser la sismicité	163
5.4	Conséquences en termes d'observation sismologique	165
5.4.1	Analyse d'un réseau <i>a posteriori</i>	165
5.4.2	Comparaison avec d'autres méthodes de détection	168
6	Application à la campagne post-sismique du séisme de La Laigne	171
6.1	Le séisme de La Laigne (16 juin 2023, $M_w=4,9$)	171
6.2	Application au signal des stations permanentes	175
6.3	Application au signal des stations temporaires	176
6.3.1	Réseau de stations temporaires post-sismique	177
6.3.2	Détection des répliques	177
6.4	Implications	183
	Conclusion et perspectives	185

Annexes

A	Bases de sismologie	A
A.1	Intensité d'un séisme	A
A.2	Magnitude d'un séisme	B
A.3	Mécanisme au foyer	B
A.4	Types d'ondes sismiques	D
	Table des figures	I

Liste des tableaux

VI

Bibliographie

VII

Liste des abréviations

BCSF Bureau Central Sismologique Français

CNA Cisaillement Nord-Armoricain

CNN Convolutional Neural Network

CSA Cisaillement Sud-Armoricain

DAM-CEA Direction des applications Militaires au sein du Commissariat à l’Energie Atomique et aux énergies alternatives

FCNN Fully Connected Neural Network

GAN Generative Adversarial Network

LDG Laboratoire de Détection et de Géophysique du DAM-CEA

LPG Laboratoire de Planétologie et Géosciences

OSUNA Observatoire des Sciences de l’Univers de Nantes l’Atlantique

OSU Observatoire des Sciences de l’Univers

RAP Réseau Accélérométrique Permanent

RF Random Forest

RLBP Réseau Large-Bande Permanent

RNNA Recurrent Neural Network with Attention

RNN Recurrent Neural Network

RÉNASS Réseau National de Surveillance Sismique

RÉSIF-EPOS Réseau Sismologique et géodésique Français -EuropeanPlateObservingSystem

SI-HEX Sismicité Instrumentale de l’Hexagone

Chapitre 1

Introduction

Sommaire

1.1 Motivations générales	1
1.2 Le Massif armoricain : un contexte géologique polyphasé	5
1.2.1 Orogenèse panafricaine (cadomienne), 650 Ma - 540 Ma	6
1.2.2 Orogenèse varisque (hercynienne), 420 Ma - 290 Ma	8
1.2.3 Événements post-hercynien	9
1.3 Structure générale du Massif armoricain	10
1.3.1 Zone Nord armoricaine	12
1.3.2 Zone Centrale armoricaine	13
1.3.3 Zone Sud armoricaine	13
1.3.4 Domaine varisque du Léon	13
1.4 Apports de la sismologie	14
1.4.1 Profondeur du Moho	14
1.4.2 Tomographie	15
1.4.3 Atténuation des ondes sismiques	17

1.1 Motivations générales

L'activité sismique mondiale est principalement localisée au niveau des limites de plaques tectoniques. Néanmoins, en dehors de ces zones, une sismicité dite « intra-plaque » peut être observée. Les processus qui contrôlent la déformation, la sismicité, l'aléa et le risque sismique, dans ces régions, sont, à l'heure actuelle, fortement débattus

(Stein et Mazzotti, 2007; Clark *et coll.*, 2012; Calais *et coll.*, 2016; Soto-Cordero *et coll.*, 2018). La persistance de l'activité sismique dans ces zones est un phénomène observé dans diverses régions du monde, telles que l'Australie (par exemple Hillis *et coll.*, 2008; Clark *et coll.*, 2012), l'Amérique du Nord (par exemple Zoback, 1992; Calais *et coll.*, 2005) ou dans le Massif armoricain en France.

Étant donné qu'il n'existe pas encore de modèle spécifique adapté à ces zones, ces événements sismiques sont généralement considérés, de manière traditionnelle, comme similaires aux séismes qui se produisent le long des limites de plaques. Ces séismes se produisent lorsque l'accumulation de contraintes tectoniques sur des failles actives atteint un seuil critique, ce qui provoque leurs ruptures (par exemple Kanamori et Brodsky, 2004). Une fois que la faille a été déchargée par le tremblement de terre, elle a besoin d'un certain laps de temps pour se recharger jusqu'au point de rupture. Ce processus dépend de la résistance de la faille, de l'ampleur de la réduction des contraintes et de la vitesse de rechargement. La contrainte, associée à la tension le long de la faille, continue de s'accumuler, pour, finalement, que la faille se décharge à nouveau, entraînant la répétition régulière ou irrégulière du cycle sismique (Fig 1.1).

En comparaison avec d'autres régions françaises, connues pour leur plus forte activité sismique, tels que l'arc alpin, le Fossé rhénan ou les Pyrénées, la région nord-ouest de la France se distingue par une sismicité fréquente et de magnitude faible à modérée. Malgré cette sismicité bien réelle, les observations géodésiques ne permettent pas de déterminer avec précision le taux de déformation en raison du déplacement en surface insignifiant. Actuellement, le Massif armoricain est soumis à des vitesses de déplacement mesurées de l'ordre de 1 mm/an (Nocquet, 2012). Dans une autre région intra-plaque, à New Madrid aux États-Unis, l'étude de Calais *et coll.* (2010) montre que les mesures du système mondial de localisation (GNSS), effectuées sur une période de 20 ans, ne révèlent aucun déplacement visible. Or, dans cette région, trois séismes d'une magnitude 7 ou plus sont survenus en 1811-1812. Des tremblements de terre de moindre ampleur se produisent encore aujourd'hui, délimitant les segments de faille qui auraient pu se rompre lors des événements de 1811-1812.

Par conséquent, une question se pose : les séismes dans les zones continentales stables obéissent-ils aux mêmes mécanismes que les séismes le long des limites de plaques, mais avec des cycles de chargement/relaxation si longs que les mouvements sismiques ne sont pas encore détectables par les mesures GNSS, ou bien sont-ils causés par d'autres facteurs (Calais *et coll.*, 2016)? Il est possible que ces séismes puissent

s'expliquer davantage par des perturbations transitoires des contraintes locales ou de la résistance des failles, libérant ainsi l'énergie élastique accumulée (Fig 1.1).

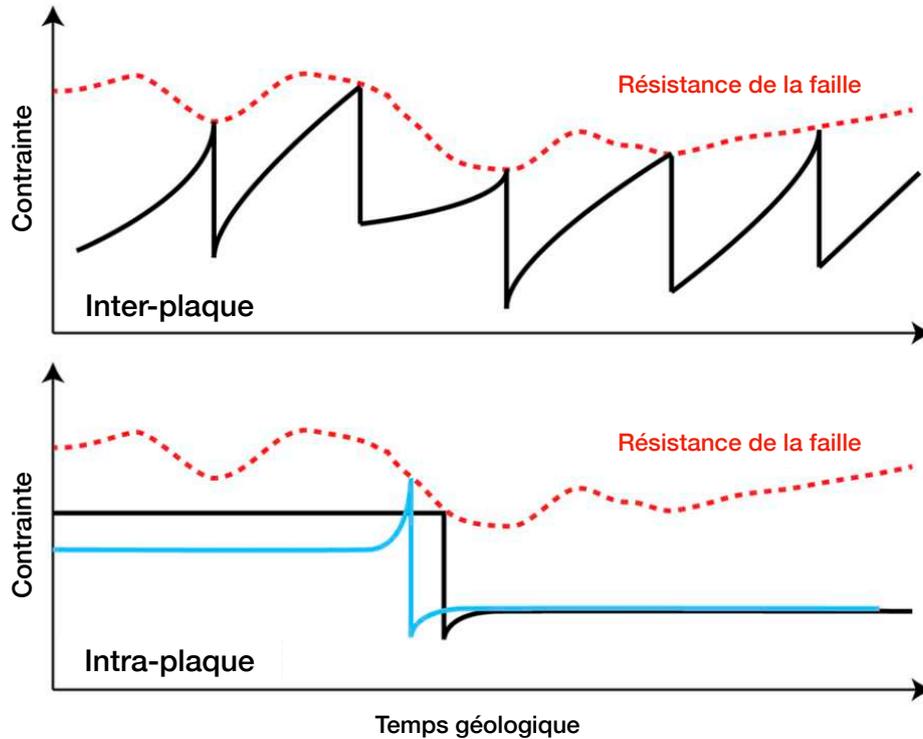


FIGURE 1.1 – Diagrammes montrant les changements de contraintes et séquence de tremblements de terre. (en haut) Cycle sismique illustré par une séquence de séismes à la limite des plaques se produisant sous l'effet de la charge tectonique, à un rythme qui peut légèrement varier avec le temps et les variations temporelles de la résistance de la faille (Kanamori et Brodsky, 2004). (en bas) Dans les environnements intra-plaques, les contraintes s'accumulent à des rythmes très lents et les séismes peuvent se produire à la suite d'une modification de la résistance de la faille (ligne noire, par exemple, augmentation de la pression interstitielle des fluides à la profondeur sismogénique) ou de perturbations transitoires des contraintes (ligne bleue, par exemple, modification de la charge hydrologique ou sédimentaire) (modifié d'après Calais et coll., 2016).

Ces perturbations transitoires ou variations locales des contraintes, associées à des facteurs agissant à différentes échelles de temps, peuvent jouer un rôle significatif dans la sismicité des zones continentales stables (Levandowski et coll., 2018). Si ces variations de contrainte ou de résistance de la faille sont suffisamment importantes, elles peuvent déclencher une rupture (par exemple Long, 1988; Calais et coll., 2010), libérant l'énergie élastique d'une lithosphère précontrainte (par exemple Feldl et Bilham, 2006) (Fig 1.1).

Dans l'ensemble, la sismicité régionale du nord-ouest de la France implique probablement des interactions complexes entre différents processus. Cela comprend des chargements tectoniques extrêmement lents et des contraintes résiduelles, des réactivations de failles avec des intervalles de temps longs entre les événements, des séismes isolés dus à une faible résistance des failles, ainsi que des perturbations transitoires des contraintes. Ainsi, des déclencheurs internes ou externes (par exemple, rebond post-glaciaire, érosion, etc.) conjugués à l'héritage structural (par exemple, système de failles majeures du cisaillement armoricain) peuvent expliquer la sismicité dans l'ouest de la France.

Ces différents déclencheurs sont généralement éclipsés, en limites de plaques, par les effets des processus tectoniques classiques. En nous concentrant sur la sismicité du Massif armoricain, en région continentale stable, nous pouvons mettre en lumière ces processus qui affectent la sismicité de la région.

Afin de mieux comprendre la sismicité dans cette région, nous mettons l'accent sur la détection des événements sismiques. Notre objectif est d'accroître la fiabilité statistique de l'étude des comportements sismiques à court terme en optimisant la détection des séismes de faible amplitude. Une détection plus précise de ces petits séismes ouvre la voie à une description plus détaillée de l'évolution spatio-temporelle des séquences sismiques. Pour atteindre cet objectif, nous avons développé une méthode de détection et de discrimination des événements sismiques de faible magnitude. Cette approche a été appliquée à l'ensemble du Massif armoricain sur une période allant de 2018 à 2021, ainsi que dans le contexte spécifique d'une étude locale portant sur le séisme de La Laigne survenu le 16 juin 2023.

Dans ce premier chapitre, nous introduisons le contexte géologique et géophysique du domaine intra-plaque du Massif armoricain. Nous examinons l'histoire géologique qui a mené à sa formation. La sismicité de la région, à la fois historique et instrumentale, sera ensuite présentée dans le second chapitre. Les différents facteurs pouvant impacter cette sismicité seront également développés. Le troisième chapitre se concentre sur un état de l'art des méthodes de détection et de discrimination des événements sismiques. Nous explorons à la fois les approches classiques utilisées dans le domaine et les outils modernes tels que l'apprentissage machine (*machine learning*). Les choix et les paramètres qui ont été explorés pour élaborer une méthode adaptée à nos besoins et à la région du Massif armoricain sont présentés en détail dans le quatrième chapitre.

Le cinquième chapitre est consacré à l'application de notre approche sur la sismicité du Massif armoricain durant quatre années de données. Nous présentons les résultats obtenus, en mettant en évidence les nouveaux événements sismiques détectés et en les comparant avec le catalogue initial. Nous étudions également la distribution spatiale des séismes localisés et explorons certains facteurs externes, qui ont été identifiés, permettant de mieux comprendre la sismicité de la région. Dans le sixième chapitre, nous étendons notre étude à la campagne post-sismique installée suite au séisme de La Laigne survenu le 16 juin 2023. Notre méthode a été appliquée aux données recueillies par les stations temporaires installées pour l'occasion. Enfin, dans le dernier chapitre, nous examinons les résultats obtenus et discutons de leur pertinence dans le contexte de notre étude.

En annexe, nous avons inclus des encadrés explicatifs contenant des rappels de base en sismologie. Ils fournissent des notions et définitions fondamentales du domaine afin de faciliter la compréhension de l'ensemble du travail de recherche présenté.

1.2 Le Massif armoricain : un contexte géologique polyphasé

Le Massif armoricain, situé dans l'ouest de la France, est une région qui présente une histoire géologique complexe. Elle englobe une vaste partie de la Bretagne ainsi que des zones adjacentes des régions des Pays de la Loire, de la Nouvelle Aquitaine et de la Normandie. Le Massif armoricain est bordé par la Manche au nord, la mer d'Iroise à l'ouest et le golfe de Gascogne au sud. De plus, le massif s'étend sous les sédiments des bassins méso et cénozoïques adjacents, notamment le Bassin aquitain au sud, le Bassin parisien à l'est et le Bassin de la Manche au nord-ouest. Au cours de sa longue évolution géologique, le Massif armoricain a connu plusieurs orogénèses (cadomienne et varisque), caractérisées par des mouvements tectoniques majeurs, des collisions de plaques et la formation de chaîne de montagnes.

Au fil du temps, les forces érosives de l'eau, du vent et de la glace ont commencé à démanteler cette chaîne de montagnes, réduisant sa hauteur et exposant les structures géologiques, autrefois enfouies, à la surface. Le Massif armoricain est principalement composé de roches métamorphiques telles que des schistes, des quartzites, des granites et des gneiss. Ces roches témoignent des processus géologiques intenses auxquels la

1.2. LE MASSIF ARMORICAIN : UN CONTEXTE GÉOLOGIQUE POLYPHASÉ

région a été soumise. Les contraintes tectoniques élevées ont conduit à des déformations ductiles et fragiles des roches, créant des plis, des failles et des zones de cisaillements complexes.

Afin de comprendre la sismicité dans l'ouest de la France, il est intéressant d'examiner la formation structurale régionale. L'héritage structural potentiel pourra être quantifié. Ainsi, les discontinuités à l'échelle de la croûte, susceptible d'être interprétées comme des failles sismiquement actives, pourront être mises en évidence.

1.2.1 Orogenèse panafricaine (cadomienne), 650 Ma - 540 Ma

Au Protérozoïque Supérieur, deux continents étaient présents au nord : la Laurentia, qui correspond au bouclier canadien, et la Baltica, correspondant au bouclier scandinave. Un troisième continent, plus méridional, appelé Gondwana, était séparé de la Laurentia et de la Baltica par deux vastes océans : l'Iapetus à l'ouest et le Tornquist à l'est.

L'orogénèse cadomienne est caractérisée par l'accrétion de Cadomia (Fig 1.2), qui deviendra ultérieurement l'Armorica, à la fin du Protérozoïque, sur la marge nord du Gondwana (par exemple [Chantraine *et coll.*, 2001](#)). Cette période est marquée par la collision entre un arc volcanique et le continent, résultant de la subduction océanique le long d'une marge continentale active. La fin de l'orogénèse cadomienne se manifeste par la mise en place de granites et une érosion rapide de la chaîne, mis en évidence par le dépôt d'une épaisse série de sédiments terrigènes appelée Briovérien (par exemple [Ballevre *et coll.*, 2013](#)).

Des traces de cette orogénèse sont visibles, à l'heure actuelle, dans la zone Nord-armoricaine sous forme de zones de cisaillement ductiles à savoir les structures Locque-meau-Trégor (LTSZ), La Fresnaye-Coutances (FCSZ) de direction N50-70, et Cancale-Granville (CGSZ) (Fig 1.3A et B).

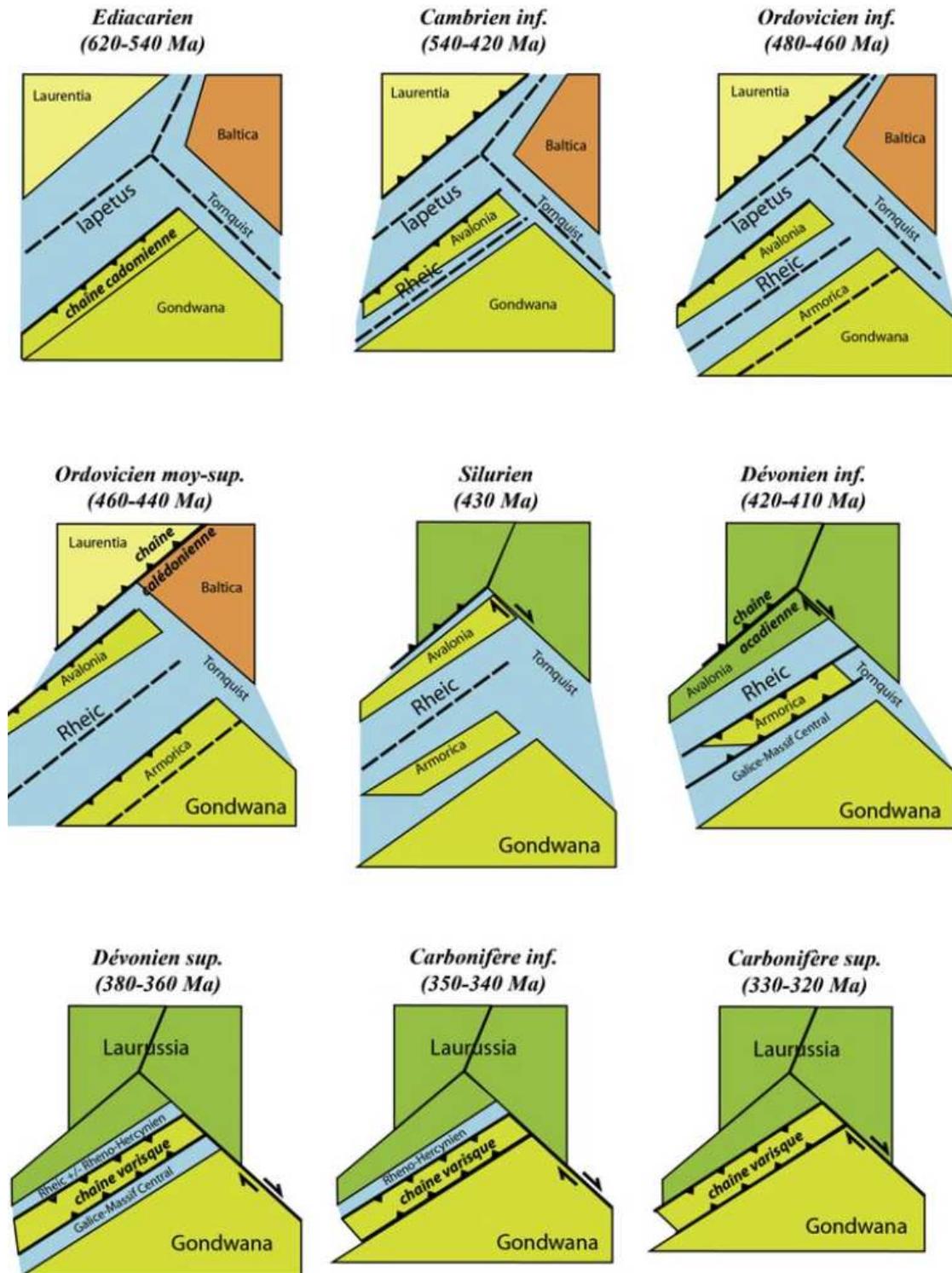


FIGURE 1.2 – Schéma montrant la géométrie et la cinématique des plaques durant le Paléozoïque. La définition des micro-plaques intermédiaires (Avalonia, Armorica), d'origine nord-gondwanienne, restent encore contestées au moins dans leur cinématique, et parfois même dans leur existence (Ballevre et coll., 2013).

1.2.2 Orogenèse varisque (hercynienne), 420 Ma - 290 Ma

L'orogénèse varisque constitue une étape significative dans l'histoire géologique du Massif armoricain, marquant la formation de la chaîne hercynienne et l'acquisition de sa structure actuelle.

Au cours du Cambrien et de l'Ordovicien, l'océan Rhéique s'est formé suite à la séparation de l'Avalonia du bloc Gondwana. Pendant l'Ordovicien supérieur, les deux continents septentrionaux, la Laurentia et la Baltica, sont entrés en collision, donnant naissance à la chaîne calédonienne et au supercontinent, Laurasia. Par la suite, Avalonia est entré en collision avec ce supercontinent lors de l'orogénèse acadienne au Dévonien inférieur, entraînant la fermeture de l'océan Iapetus. Parallèlement, l'océan Galice-Massif central s'est formé, séparant le bloc Gondwana de la plaque Armorica.

Au Dévonien inférieur, la plaque Gondwana a amorcé une migration depuis le Pôle Sud jusqu'au niveau de l'Équateur. Ce mouvement tectonique a eu des conséquences significatives sur la configuration des océans environnants. À la fin de l'Ordovicien et au Silurien, l'océan Rhéique, situé au nord de la plaque Armorica, a été soumis à un processus de subduction, se déplaçant sous la lithosphère armoricaine. Parallèlement, au sud, l'océan Sud Armoricain a été subduit vers le nord.

Ces mouvements de subduction ont entraîné des interactions complexes entre les plaques lithosphériques, conduisant à la fermeture de l'océan Rhéique et à la collision de différentes masses terrestres, formant la Pangée. Cette double subduction, d'abord océanique puis continentale, a eu un impact majeur sur la structuration de la chaîne varisque (Matte, 1986a, 2001). Ce processus a conduit à la formation d'une cordillère symétrique s'étendant sur près de 800 km de largeur et 5 000 km de longueur (Matte, 1986a,b).

Des indices de l'orogénèse varisque peuvent être observés, à l'heure actuelle, à travers les différentes failles présentes dans la région, notamment le Cisaillement Sud-Armoricain (CSA) et le Cisaillement Nord-Armoricain (CNA) (représentés sur la Figure 1.3 par les tracés SANSZ/SASSZ et NASZ, respectivement). Ces failles se caractérisent par une direction moyenne de N90-130 et elles exercent des mouvements horizontaux de cisaillement sur la croûte terrestre.

Dans sa partie occidentale, le CSA est constitué d'une seule branche avec un azimut d'environ N95, mais à mesure que l'on se déplace vers l'est, dès Lorient, il se divise en deux branches distinctes. La branche nord (SANSZ sur la Figure 1.3) s'étend jusqu'à Angers en traversant les Landes de Lanvaux, en conservant un azimut proche de N110, tandis que la branche sud (SASSZ sur la Figure 1.3) s'oriente vers le sud-est en direction de Nantes. Cette branche sud se subdivise également en plusieurs ramifications alors qu'elle progresse vers le seuil du Poitou, qui marque la frontière entre les bassins parisien et aquitain. Elle devient alors moins visible dans la topographie. Les déplacements les plus significatifs se concentrent principalement dans la branche sud du CSA, avec une déformation horizontale cumulée d'environ 300 km (Gumiaux *et coll.*, 2004; Brun *et coll.*, 2001). En revanche, la branche nord présente des déplacements plus modestes, d'au moins 40 km (Jégouzo et Rossello, 1988).

Le CNA (NASZ sur la Figure 1.3) s'étend d'est en ouest de Brest à la Mayenne. Ce cisaillement est moins apparent dans la topographie de la région comparé au CSA. Selon les études, son déplacement ne dépasse pas les 20 km (Ballevre *et coll.*, 2013).

1.2.3 Événements post-hercynien

Pendant le Mésozoïque, la fragmentation progressive de la Pangée a entraîné l'ouverture de plusieurs bassins océaniques, dont l'océan Atlantique et le golfe de Gascogne. Dans ce contexte, le Massif armoricain, en tant que socle continental émergé, a subi une évolution particulière. L'ouverture de l'océan Atlantique (extension permo-triasique, par exemple Chadwick et Evans, 1995; Faure, 1995) a entraîné le développement de bassins, associés à l'activité magmatique et à des réseaux de failles à l'échelle régionale. Au nord, les zones de failles majeures médianes (MCFZ) et d'Aurigny-Ushant (AUFZ), héritées de phases géologiques antérieures, ont été réactivées en tant que failles normales (par exemple Bessin, 2014). À l'ouest, des failles NW-SE, avec, par exemple, la faille de Kerforne (KF) et Le Quessoy (LQF), de direction N140-160, se sont développées en réponse à un deuxième épisode d'extension orientée ENE-WSW (par exemple Chadwick et Evans, 1995).

Entre le Crétacé inférieur et l'Aptien supérieur, le golfe de Gascogne s'est ouvert le long des marges armoricaine et nord-ibérique. La formation d'un rift et l'extension de la croûte terrestre, est donc due au déplacement de l'Ibérie vers le Sud-Ouest. Cette phase de rifting a donné lieu à un relief en forme d'épaulement dans la région du Massif

armoricain. Au fur et à mesure que le Crétacé inférieur progressait, l'épaulement du rift du Massif armoricain s'est aplani, signifiant que la topographie s'est atténuée et s'est nivelée.

Pendant le Cénozoïque, l'histoire géodynamique de l'Europe est marquée par des épisodes de compression résultant de la formation des Alpes et des Pyrénées, principalement dus à la collision entre l'Europe et l'Afrique. À l'Oligocène, une série de rifts d'envergure continentale s'est développée en avant de l'arc alpin. Ce système de rift s'étendait depuis la mer du Nord jusqu'à la fosse de Valence, le long des côtes espagnoles.

1.3 Structure générale du Massif armoricain

La structure actuelle du Massif armoricain est le fruit des différentes orogènes qui ont façonné cette région. Ces événements tectoniques ont donné naissance à des chaînes de montagnes, des failles et des plis, qui ont ensuite été modifiés et érodés au fil du temps. Aujourd'hui, le Massif armoricain présente une topographie variée comprenant des collines, des plateaux et des vallées. Il est caractérisé par la présence de roches magmatiques, métamorphiques et sédimentaires qui témoignent de son histoire géologique complexe.

Situé en tant que marge continentale passive au sein d'un domaine de déformation intra-plaque, la structure du socle est caractérisée par trois familles de failles :

- une direction moyenne de N50, avec par exemple la faille de La Fresnaye-Coutances (FCSZ), d'héritage cadomien ;
- une direction N110-130, avec les Cisaillements Sud-Armoricain (CSA) et Nord-Armoricain (CNA), d'héritage hercynien ;
- et enfin une direction N140-160, avec par exemple la faille Le Quessoy (LQ) (Vigneressée, 1988; Bonnet *et coll.*, 2000), post-hercynienne.

Le Massif armoricain peut être décomposé en quatre grands domaines (Fig 1.4) délimités par les réseaux de failles majeures, visibles dans la topographie.

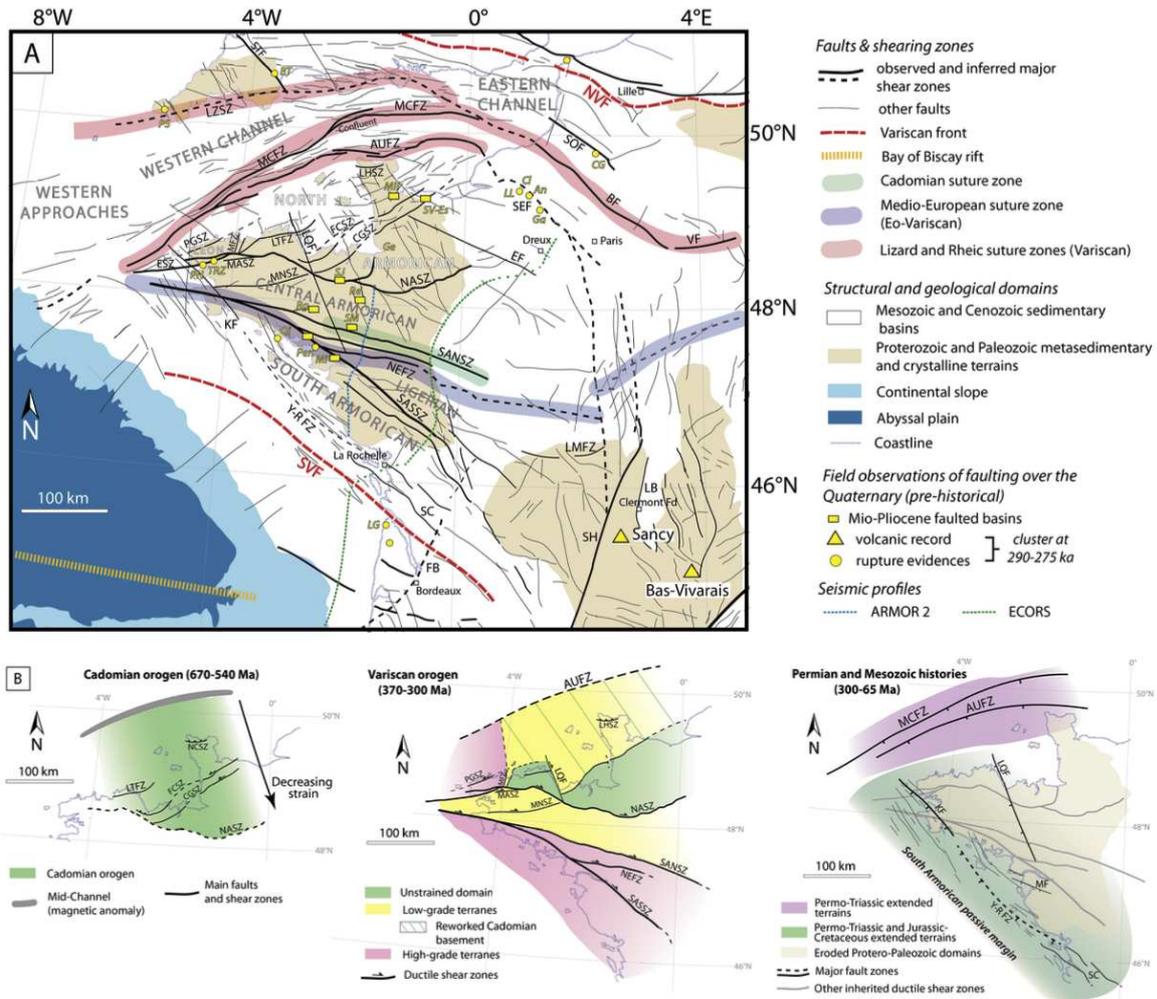


FIGURE 1.3 – (A) Localisation des principales failles de surface dans l’ouest de la France en fonction de leur origine (voir détails dans le texte). (B) Répartition spatiale des principales phases tectoniques et des structures associées aux failles dans l’ouest de la France, avec (1) la ceinture orogéniques cadomienne (2) varisque (modifiée d’après Le Gall et coll., 2021) et (3) les réseaux de failles et de bassins permo-mésozoïques (Beucler et coll., 2021).

1.3. STRUCTURE GÉNÉRALE DU MASSIF ARMORICAIN

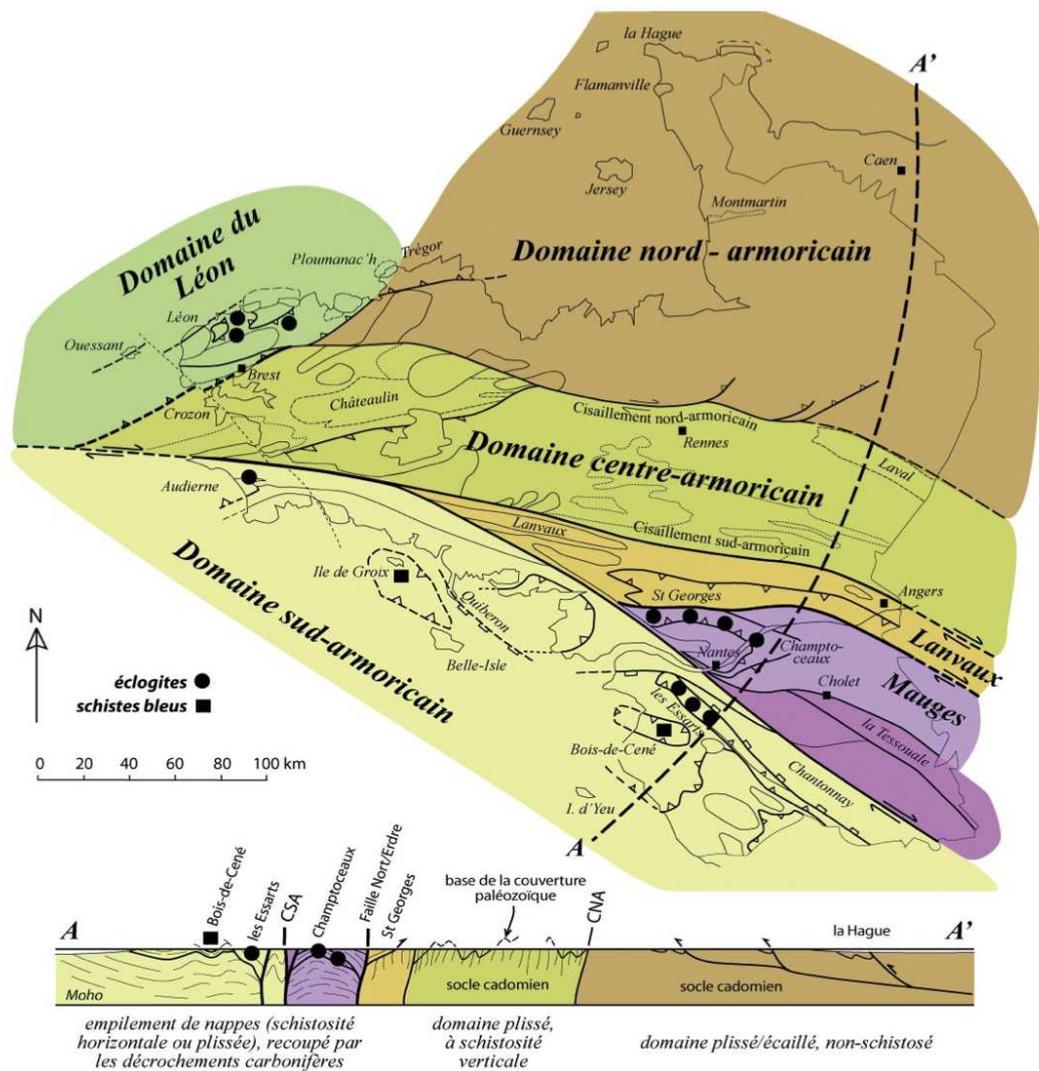


FIGURE 1.4 – Carte et coupe synthétiques du Massif armoricain mettant en évidence le découpage en grandes unités par les zones de cisaillement et l'intensité de la déformation et du métamorphisme varisques (Ballevre et coll., 2013).

1.3.1 Zone Nord armoricaine

Cette zone, située au nord du CNA, s'étend du Cotentin au Trégor est d'origine cadomienne. Cette région est composée de roches sédimentaires et volcano-sédimentaires terrigènes formées au Protérozoïque supérieur. Ces formations sont traversées par des intrusions de granitoïdes qui ont subi un processus de métamorphisme lors de l'orogénèse cadomienne. Ce contexte géologique reflète une phase de collision continentale suivie d'une subduction océanique vers le sud-est, comme en témoignent les études de Brun *et coll.* (2001). En conséquence, la chaîne hercynienne a subi une érosion rapide, donnant lieu à la formation des schistes du Briovérien.

1.3.2 Zone Centrale armoricaine

La zone centrale du Massif armoricain est délimitée au nord par le CNA et au sud par la branche nord du CSA. Dans sa partie orientale, cette zone présente les caractéristiques d'une zone de cisaillement simple dextre (Gumiaux, 2003). Dans cette région, on trouve principalement des roches sédimentaires déformées, qui se présentent sous forme de schistes faiblement métamorphisés. Les dépôts les plus anciens, connus sous le nom de schistes du Briovérien, remontent à environ 570 Ma (Chantraine *et coll.*, 1988). La phase d'extension, qui a eu lieu du Cambrien à l'Ordovicien, se traduit par le dépôt de sédiments continentaux qui se trouvent principalement dans des structures en blocs basculés. La série sédimentaire se poursuit avec des formations telles que les schistes rouges, les grès armoricains et les schistes noirs d'Angers, qui s'étendent jusqu'au Dévonien. Au-delà de cette période, seuls quelques sédiments sont préservés, et sont alignés le long des failles de cisaillement telles que le CNA et le CSA.

1.3.3 Zone Sud armoricaine

Cette zone est située au sud de la branche nord du CSA. Cette région a été fortement impactée par l'orogénèse varisque. Pratiquement toutes les formations sédimentaires, à la fois le socle protérozoïque et la couverture paléozoïque, ont subi un métamorphisme intense, représenté, par exemple, par les faciès en schistes bleus sur l'île de Groix. Cette situation indique une position interne lors de l'orogénèse varisque, en accord avec une forte déformation. La branche sud du CSA a donc eu un rôle prépondérant dans ces déformations comme en témoignent de grandes structures chevauchantes en ciseaux (Ballevre *et coll.*, 2013).

1.3.4 Domaine varisque du Léon

L'origine allochtone du domaine du Léon, formé pendant l'orogénèse varisque, ajoute une complexité supplémentaire à l'histoire géologique de l'Armorique. Son histoire semble être étroitement liée à celle des domaines sud-armoricains. Ce domaine se serait mis en place sous forme de nappes chevauchantes, datées d'environ 440 Ma, et présente un degré de métamorphisme très élevé. Cette unité géologique a subi des cisaillements ductiles et la mise en place de granites au Carbonifère (Cagnard, 2008). Son existence soulève des questions quant à sa relation avec les autres éléments du Massif armoricain et à son origine géodynamique spécifique.

1.4 Apports de la sismologie

Les études sismologiques, au sens tomographique, se consacrent à l'analyse de la propagation des ondes élastiques à travers la Terre. Elles fournissent des informations sur les caractéristiques du milieu traversé en utilisant des approches actives, tels que les profils sismiques, ainsi que des approches passives, qui exploitent le champ d'ondes sismiques ambiant et les données provenant de séismes locaux, régionaux ou lointains. Ces différentes méthodes permettent de construire des modèles tridimensionnels des vitesses des ondes ou des cartes de profondeur de la discontinuité de Mohorovičić (Moho), qui sépare la croûte terrestre du manteau.

1.4.1 Profondeur du Moho

Au début des années 70, les premières investigations sismiques profondes ont été réalisées en France, marquant une avancée majeure dans l'étude de la structure interne de la Terre (Hirn, 1976). Les premières profondeurs du Moho ont donc été déterminées autour de 31 km ($\pm 1,5$ km). Une légère variation de profondeur entre l'ouest et l'est de la région est cependant observée. En effet, entre Brest et Blain, situé au nord de Nantes, la profondeur du Moho était estimée à 31,3 km, tandis qu'elle atteignait 32,5 km entre Blain et Limoges.

À la fin des années 80, de nouveaux profils sismiques ont été établis dans le cadre de l'Étude de la Croûte continentale et Océanique par Réflexion Sismique (ÉCORS), (Bois *et al.*, 1988). Une analyse de ces profils a permis de conclure que la profondeur du Moho était relativement constante autour de 30 km, avec des variations inférieures à 5 km.

À la fin du XXe siècle, de nouveaux profils de sismique réflexion en écoute longue ont été réalisés, notamment les profils Armor 1 et 2, faisant partie du projet GéoFrance3D. Ces profils ont apporté des conclusions supplémentaires aux observations précédentes. Le profil Armor 1 a révélé que la base de la croûte inférieure serait à une profondeur constante entre 30 et 31 km. Le profil Armor 2, quant à lui, traversait la zone de collision hercynienne dans le sud-est du Massif armoricain (Bitri *et coll.*, 1997, 2001). Le Moho serait légèrement plus profond sous les domaines de Champtoceaux et du Centre armoricain (32 km) par rapport à la partie sud du profil (31-32 km).

De 2011 à 2013, des réseaux temporaires de sismomètres à large bande ont été déployés dans le sud-ouest de la France, le long de la côte atlantique française et dans le nord de l'Espagne dans le cadre des expériences PYROPE (Chevrot *et coll.*, 2017) et IBERARRAY (Díaz *et coll.*, 2009). Grâce à cet ensemble de données, Chevrot *et coll.* (2014) a permis de déterminer l'épaisseur apparente de la croûte (Fig 1.5). À partir de mesures de fonctions récepteurs, il est possible de construire des modèles crustaux. La fonction récepteur correspond à la réponse de la structure, près du capteur, soumise à une onde sismique, lors d'un séisme lointain (par exemple Langston, 1977). Elle permet de caractériser les interfaces entre différentes couches, en analysant les variations temporelles et les amplitudes des signaux enregistrés. En pratique, le temps de parcours différentiel de la phase P_mS (la conversion de P à S au Moho) par rapport à l'onde directe P est mesuré, à chaque station. Cette différence de temps peut ensuite être convertie en profondeur apparente du Moho, via un modèle de vitesse. La Figure 1.5 montre les résultats obtenus par cette approche et confirme donc que la profondeur du Moho, dans le Massif armoricain, serait relativement constante entre 28 et 32 km.

1.4.2 Tomographie

La tomographie sismique permet de reconstruire la structure interne de la Terre par l'analyse des vitesses de propagation des ondes sismiques à travers différentes régions du globe. Les variations spatiales de la vitesse de propagation des ondes sismiques peuvent être cartographiées. Une étude de tomographie sismique des ondes P_n en France, modélisant à la fois les variations latérales de vitesse et d'anisotropie, est décrite dans (Judenherc *et coll.*, 1999). Dans un contexte isotrope, la partie Est du Massif armoricain révèle une forte anomalie de vitesse (de +2 à +5%). En revanche, dans un contexte anisotrope, l'analyse de la polarisation des ondes P_n dans le manteau met en évidence des variations de vitesses sismiques de l'ordre de 2 à 4%, avec des directions rapides parallèles aux structures hercyniennes (CSA et CNA). Par la suite, Judenherc (2000) met en évidence un signal distinct dans le manteau supérieur suggérant la présence d'un assemblage de deux lithosphères d'origines différentes. Ceci peut être interprété comme la signature d'une subduction à vergence nord qui s'est achevée au Dévonien. D'autres études s'intéressent à l'anisotropie sismique à partir de la biréfringence des ondes SKS (Judenherc *et coll.*, 2002, 2003). En comparant les temps d'arrivée et les amplitudes des deux composantes polarisées de l'onde SKS , il est possible de quantifier les variations de vitesse des ondes S dans différentes directions et d'en déduire des informations sur

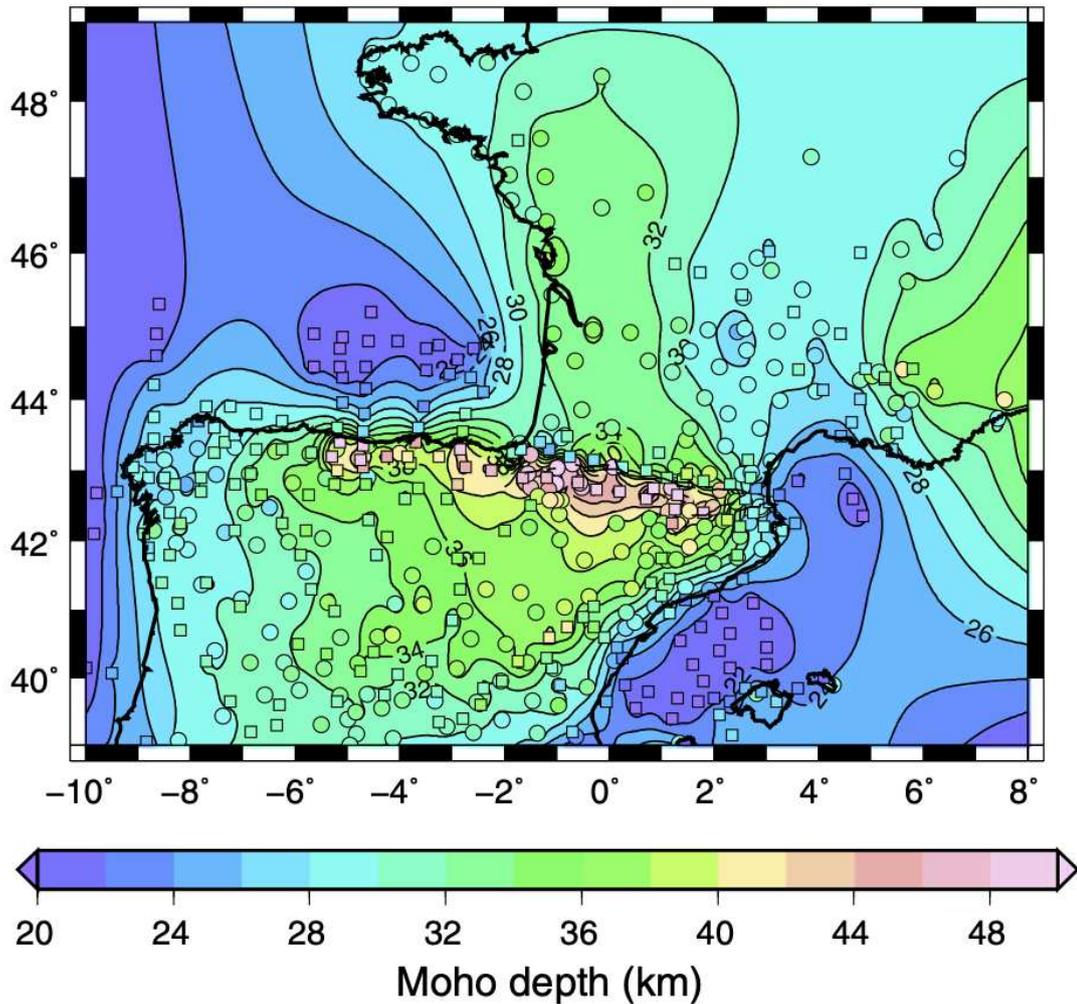


FIGURE 1.5 – Carte de l'épaisseur de la croûte terrestre en France Métropolitaine et Espagne. Cette carte a été obtenue en interpolant les épaisseurs de croûte obtenues à partir des fonctions récepteurs (cercles) et des profils de réflexion/réfraction sismique (carrés) (Chevrot et coll., 2014).

la structure anisotrope du milieu géologique traversé. Dans le sud du Massif armoricain, les résultats sont similaires à l'analyse des ondes P_n . En revanche dans le nord et le centre de la région, les ondes SKS montrent une direction rapide NW-SE, qui ne correspond pas à la déformation hercynienne impliquant un découplage entre la croûte et le manteau.

Plus récemment, une approche non conventionnelle a été développée par Gaudot et coll. (2020) afin de déduire des modèles de vitesse sismique V_s 1D entre des paires de stations. Pour cela, des fonctions de Green empiriques ont pu être déterminées à partir d'inter-corrélations de bruit sismique. Suivant l'idée de Cauchie et Saccorotti (2013) et Panning et coll. (2015), la stratégie proposée repose sur une inversion par chaîne

de Markov Monte Carlo (McMC) du diagramme de dispersion de l'inter-corrélation du bruit afin de récupérer les variations 1D de la V_s en fonction de la profondeur. Enfin, ces modèles sont régionalisés pour obtenir un modèle 3D de la vitesse V_s sur l'ensemble de la région (Fig 1.6). Nous pouvons donc observer que, proche de la surface (4 km), les variations de vitesse semblent être corrélées à la géologie (Fig 1.6.A). Les bassins sédimentaires présentent des anomalies négatives en comparaison avec les affleurements cristallins présentant des anomalies positives (+4%). Ces anomalies positives semblent se propager en dehors des limites des affleurements cristallins, témoignant de l'extension, en profondeur, du socle hercynien sous les sédiments des bassins adjacents. Plus en profondeur, à 16 km, la structure des vitesses montre des différences significatives. Une forte anomalie négative est observée à grande échelle (-10%) sur la partie nord-ouest de la France (Fig 1.6.B). La tranche de 30 km de profondeur présente des similitudes avec les résultats obtenus à 16 km, avec des anomalies de vitesse V_s négatives d'environ 5% dans la région (Fig 1.6.C).

1.4.3 Atténuation des ondes sismiques

L'atténuation de l'amplitude des ondes sismiques est une propriété distinctive des roches, résultant de l'absorption (atténuation anélastique) et de la diffusion (atténuation élastique) des ondes. Cette atténuation peut varier latéralement et fournir des informations sur la présence d'hétérogénéités dans la croûte terrestre.

Lorsque les ondes sismiques traversent différentes formations rocheuses, elles peuvent subir des pertes d'énergie en raison de divers processus. L'atténuation anélastique est due à l'absorption d'énergie par les roches, où une partie de l'énergie est convertie en chaleur. Ce phénomène est souvent favorisé par la présence, par exemple, de minéraux altérés ou de fractures dans la croûte. L'atténuation élastique, quant à elle, est causée par la diffusion des ondes à travers les hétérogénéités du milieu, telles que des variations de densité, de porosité ou de lithologie. Cette diffusion peut disperser l'énergie sismique dans différentes directions.

La variation de l'atténuation sismique peut donc être utilisée pour caractériser les propriétés des roches et identifier les hétérogénéités crustales. Par exemple, des zones présentant une atténuation plus élevée peuvent indiquer la présence de matériaux altérés, de roches fracturées ou de fluides souterrains. En cartographiant ces variations latérales de l'atténuation sismique, il est possible d'obtenir des informations sur la structure et la composition de la croûte terrestre.

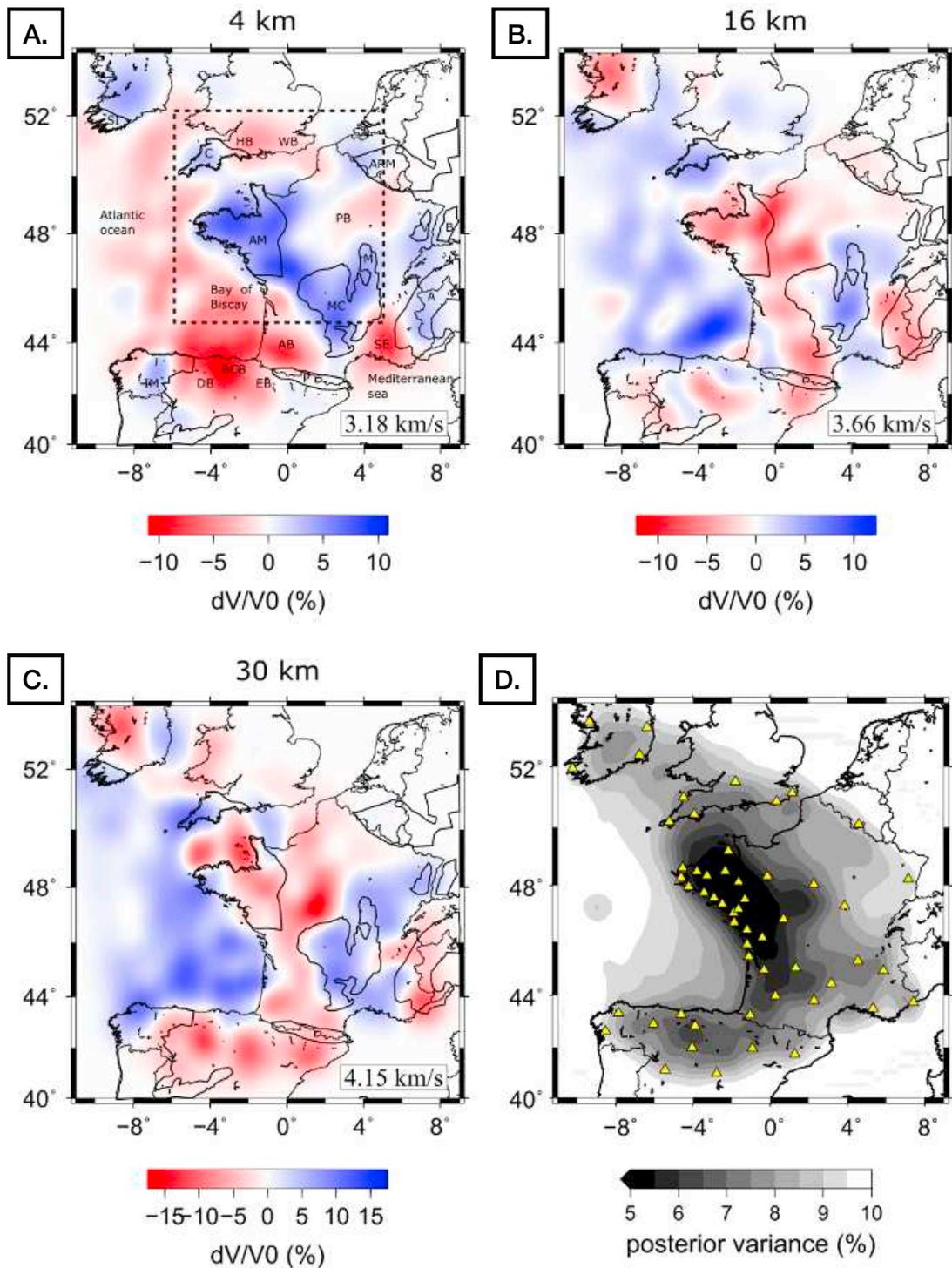


FIGURE 1.6 – Cartes des anomalies V_s isotropes pour trois profondeurs (A) 4 km, (B) 16 km et (C) 30 km. La valeur médiane de la V_s est indiquée en bas à droite de chaque carte. Le rectangle noir en pointillé en (A) indique l'ouest de la France et les régions environnantes. (D) Écart-type a posteriori des paramètres du modèle associés au résultat de la profondeur de 30 km (Gaudot et coll., 2020).

Différentes études montrent une atténuation significativement plus faible, dans le Massif armoricain, par rapport au reste de la France, tant d'un point de vue macro-sismique (Arroucau *et coll.*, 2006; Bakun et Scotti, 2006) que d'un point de vue instrumental (Mazabraud *et coll.*, 2013). Dans une étude récente, menée par Mayor (2016), une faible absorption a été observée dans le Massif armoricain avec une absorption relativement plus forte dans le domaine Centre armoricain (Fig. 1.7), confirmant les observations de Campillo et Plantet (1991). Cette faible atténuation sismique observée dans la région peut entraîner des variations significatives dans la décroissance des amplitudes des ondes S , ce qui peut potentiellement surévaluer les estimations de magnitudes locales.

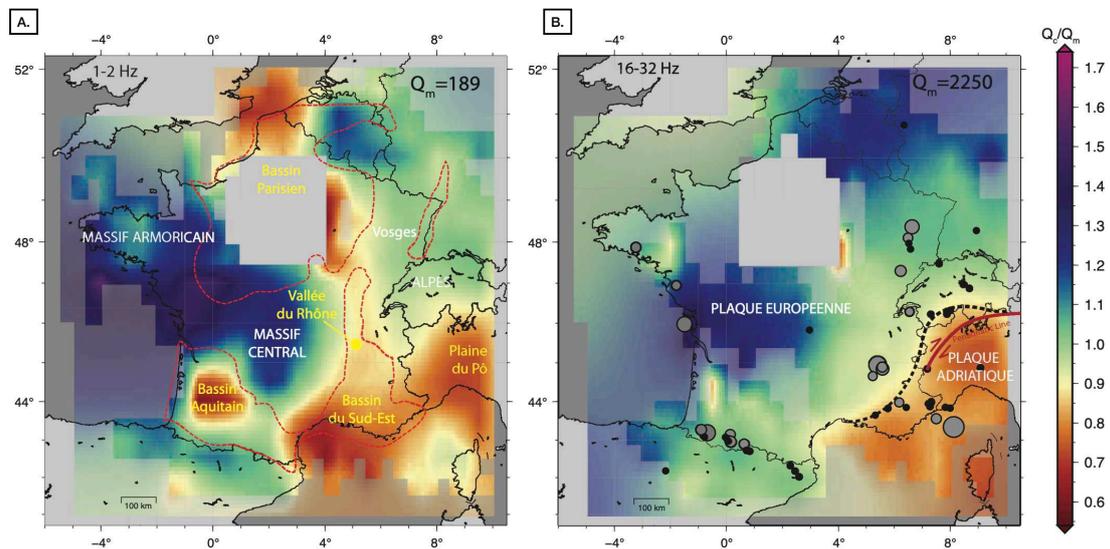


FIGURE 1.7 – (A) Tomographie d'absorption pour la France à 1Hz. Les bassins du Cénozoïque à Mésozoïque de la carte géologique sont indiqués en trait rouge. (B) Tomographie d'absorption pour la France à 24Hz. Les plus forts séismes du catalogue SI-Hex (1962-2009) correspondant à une magnitude supérieure à 4 ont été reportés (points gris). Les séismes historiques de la base SisFrance d'intensité épacentrale supérieure ou égale à VIII (échelle MSK) sont indiqués par les points noirs. La ligne périadriatique (faille décrochante droite), correspondant à la zone de suture entre les plaques européenne et adriatique, est montrée en rouge foncé. La ligne noire en pointillés indique la ligne de contact entre les structures de forte et faible absorption (Mayor, 2016).

Chapitre 2

Sismicité d'une région continentale stable

Sommaire

2.1 Sismicité historique	23
2.2 Sismicité instrumentale	26
2.2.1 Réseaux de stations permanentes	26
2.2.2 Sismicité diffuse	27
2.3 Magnitude de complétude élevée	31
2.4 Sismicité non tectonique	35
2.5 Champs de contrainte	38
2.6 Vers une meilleure compréhension de cette sismicité ? . .	42

La répartition des tremblements de terre, de magnitude supérieure à 4, présente une tendance remarquable à se concentrer dans différentes régions du monde (Fig 2.1). La première zone majeure correspond à la région péripacifique. Cette zone de subduction englobe les contours de l'océan Pacifique ainsi que la boucle des Antilles. Cette zone est particulièrement active en terme d'activité sismique, avec de nombreux séismes puissants se produisant régulièrement.

Une seconde zone où les tremblements de terre sont fréquents se situe le long des dorsales océaniques. Elles s'étendent sur de longues distances sous les océans et sont caractérisées par une intense activité sismique résultant de la divergence des plaques tectoniques. Ces mouvements divergents des plaques créent des accumulations de contraintes qui finissent par se relâcher sous forme de tremblements de terre.

Enfin, on retrouve également une activité sismique significative le long de la frontière entre les plaques eurasienne et africaine, s'étendant depuis les Açores jusqu'à Java. Les forces géologiques en jeu dans cette région résultent de la collision et de l'interaction entre les plaques eurasienne, africaine, arabique et indienne.

Par contraste, certaines zones sont dépourvues d'événements sismiques significatifs (par exemple, l'Afrique de l'Ouest, le Canada ou la Sibérie). Enfin, dans certaines régions, nous pouvons également observer une activité sismique intra-plaque, comme c'est le cas pour l'Australie, pour le Massif armoricain en France ou encore les Appalaches en Amérique du Nord.

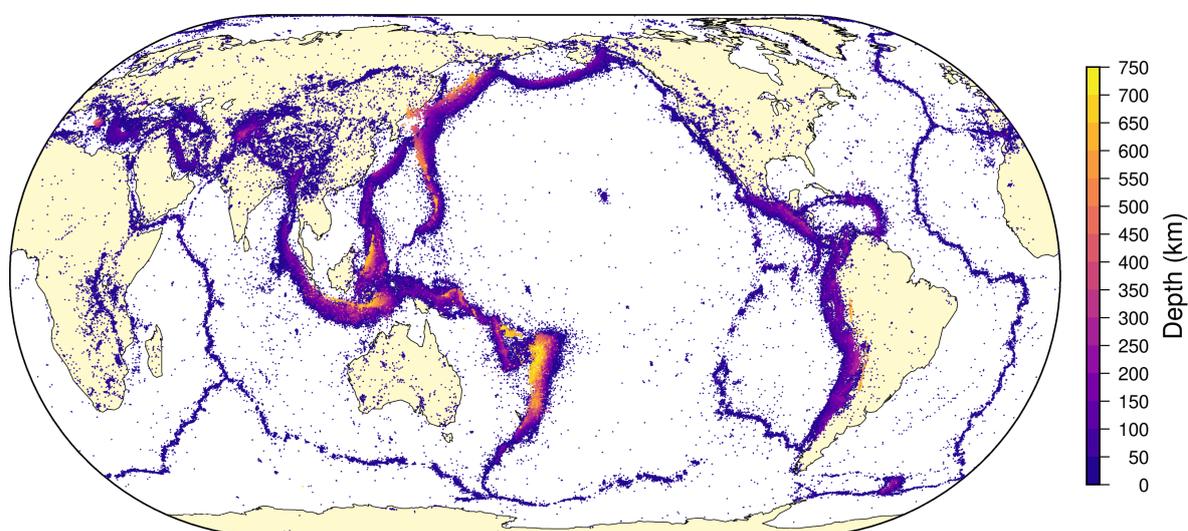


FIGURE 2.1 – Sismicité instrumentale globale de 1964 à 2021 pour les événements sismiques de magnitude supérieure à 4. La couleur des symboles correspond à la profondeur des séismes.

Depuis les débuts de la sismologie moderne, l'attention principale s'est portée aux régions à forte sismicité, qui correspondent généralement aux limites des plaques tectoniques. En conséquence, ces zones ont été de plus en plus étudiées et documentées, ce qui a permis une meilleure compréhension de la dynamique et du comportement sismique associé. Cependant, cette focalisation sur les limites de plaques a longtemps détourné l'attention des sismologues des régions intra-contininentales où la sismicité est faible à modérée. Ces zones sismiquement moins actives ont souvent été considérées comme moins dangereuses du point de vue de l'aléa sismique. Par conséquent, elles ont été relativement moins étudiées par rapport aux zones situées le long des limites de plaques.

Pourtant, avec le temps, il est devenu évident que ces régions intra-continentales présentent un potentiel sismique significatif. Bien que les séismes y soient moins fréquents et généralement de moindre magnitude, l'étude de ces zones offre l'opportunité de découvrir des origines sismiques qui sont généralement masquées par les forces tectoniques dominantes observées le long des limites de plaque. En raison de sa nature et de son histoire géologique, la sismicité dans l'ouest de la France, où se trouve le Massif armoricain et les bassins environnants, peut être classée dans la catégorie intra-plaque ou intra-continentale. Ce chapitre se concentrera donc spécifiquement sur la sismicité de cette région.

2.1 Sismicité historique

Les premiers témoignages de tremblements de terre dans le nord-ouest de la France remontent au Ve siècle, notamment avec le séisme de novembre 469, dans la région d'Angers ([Alexandre, 1990](#)). Jusqu'au milieu du XXe siècle, les archives historiques étaient la principale source d'informations permettant de retracer la sismicité de cette région, étant donné que très peu de preuves géologiques de surface étaient disponibles. Ces témoignages historiques ont été convertis en intensités sismiques en utilisant l'échelle d'intensité macrosismique MSK-64¹.

Les premiers récits décrivant les effets d'un séisme sur la population et les bâtiments sont associés au séisme du 18 avril 577, indiquant une intensité de VI sur l'échelle MSK-64 près de la ville de Chinon. Dans le golfe du Morbihan, lors du séisme de 1286, une intensité maximale estimée à VI, bien que sujette à débat, aurait été atteinte. Des bâtiments ont été sévèrement endommagés voire détruits en raison de cette secousse sismique majeure. Une autre période marquante en termes de dommages sismiques fut l'année 1751 à Nantes, où des bâtiments ont subi des effets similaires (par exemple [Gautier, 1975](#)). Le séisme de Limagne en 1490 a été également un des événements les plus importants en termes d'intensité épiscopale, atteignant VIII sur l'échelle MSK-64. D'autres séismes notables, bien documentés, ont également été enregistrés dans la région. Les événements de Loudun en 1711 et de Lorient en 1725 ont atteint une intensité épiscopale de VII sur l'échelle MSK-64. Le 25 janvier 1799, un tremblement de terre, connu sous le nom de séisme de Bouin, a été ressenti avec une intensité épiscopale estimée à VII-VIII sur l'échelle MSK-64. La Figure 2.2 illustre l'activité sismique historique majeure entre 462 et 1962 dans la région.

1. Plus d'informations sur les échelles d'intensité sont disponibles en Annexe A.1

2.1. SISMICITÉ HISTORIQUE

Ces événements dits historiques témoignent de l'existence d'une activité sismique significative dans le nord-ouest de la France à travers les siècles. Cependant, il est important de noter que la quantité et la qualité des données d'intensité disponibles pour les séismes historiques peuvent varier considérablement. Alors que de nombreuses données d'intensité sont disponibles pour les événements récents, ce n'est pas le cas pour les événements plus anciens. Par exemple, le séisme de 577 à Chinon repose sur une seule valeur d'intensité.

Entre 1861 et 1960, environ une centaine de séismes sont recensés dans le Massif armoricain, marquant l'activité sismique de la région. Parmi ces événements, le séisme de Jersey de 1926 se distingue particulièrement. Des études récentes ont estimé sa magnitude de moment² (M_W , [Hanks et Kanamori, 1979](#)) entre 5,0 et 5,5 ([Amorèse et coll., 2020](#)), ce qui en fait l'un des séismes les plus importants du Massif armoricain au XXe siècle. Ce séisme a été suivi, huit mois plus tard, par un autre tremblement de terre. Bien que sa magnitude soit légèrement inférieure, cet événement sismique est également notable en raison de sa proximité temporelle avec le séisme précédent. Parmi les séismes plus récents, il convient de mentionner la crise sismique de Rouillac en Charente, qui s'est déroulée entre le 18 juillet 1935 et le 24 août 1936 et les séismes du 9 janvier 1931 et du 2 janvier 1959 (par exemple [Gautier, 1975](#)) qui semblent être localisés à proximité de la branche sud du CSA.

La sismicité observée dans le Massif armoricain présente des caractéristiques typiques des régions intracontinentales à faible taux de déformation. Cela se traduit par une période de retour prolongée des séismes majeurs, durant laquelle un événement sismique important ($M_W > 3,5$) survient avec une récurrence apparente d'environ 10 ans.

2. Plus d'informations sur les mesures de magnitudes sont disponibles en Annexe [A.2](#)

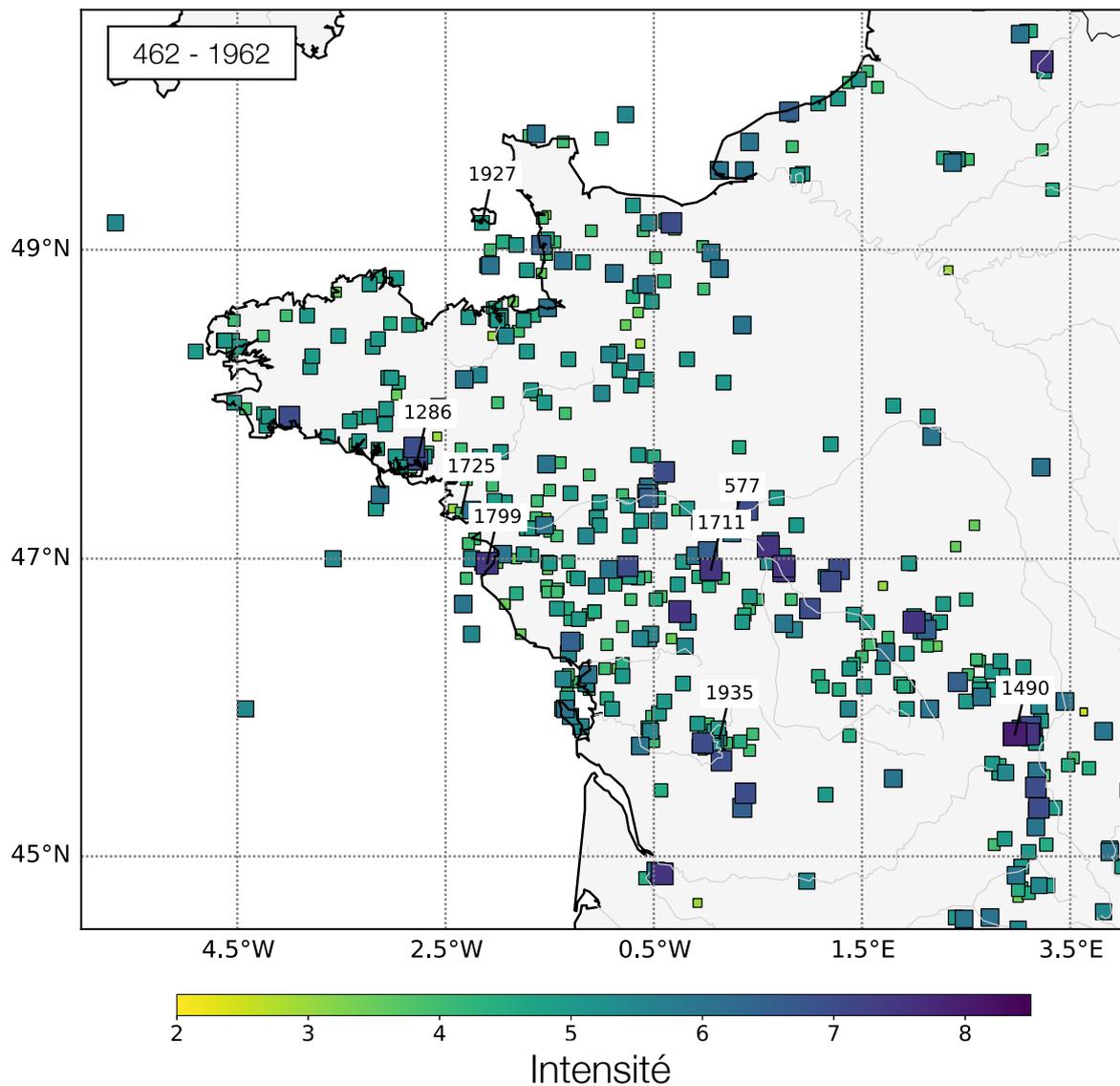


FIGURE 2.2 – Sismicité historique (www.sisfrance.net) entre 462 et 1962 dans le quart Nord-Ouest de la France. L'échelle de couleur et la taille des symboles correspondent à l'intensité de l'échelle MSK-64. Les tremblements de terre majeurs introduits dans le texte sont étiquetés.

2.2 Sismicité instrumentale

La sismicité instrumentale fait référence à l'étude et à l'enregistrement des séismes à l'aide d'instruments de mesure spécifiques appelés sismomètres. Contrairement à la sismicité historique, qui est basée sur des témoignages et des archives, la sismicité instrumentale repose sur des données scientifiques recueillies à l'aide d'instruments de détection et de mesure des mouvements sismiques. La création d'un catalogue fiable de la sismicité instrumentale dans le nord-ouest de la France a débuté dans les années 1960 avec le déploiement du premier réseau sismique par le Commissariat à l'énergie atomique (CEA).

2.2.1 Réseaux de stations permanentes

Le réseau national du Laboratoire de Détection et de Géophysique (LDG) de la DAM-CEA est composé de stations vélocimétriques réparties sur l'ensemble du territoire français. Les stations larges bandes, identifiées par le code RD ([RESIF, 2018](#)) et les triangles bleus sur la Figure 2.3, ont été parmi les premières à être implantées en 1959. Actuellement, le réseau compte une vingtaine de ces stations accessibles via le portail Résif-Epos. Deux d'entre elles sont situées dans le Massif armoricain, à savoir ROSF et MFF. À l'origine, ce réseau avait une mission de surveillance militaire, en particulier dans le cadre de la lutte contre la prolifération nucléaire.

Le Réseau Accélérométrique Permanent (RAP) est chargé de collecter des données de mouvements forts, ce qui permet d'étudier les sources sismiques et d'analyser le mouvement du sol. Dans la partie ouest de la France, on compte sept stations accélérométriques (BOUF, UBVA, UBQP, UBBR, SMFF, QUIF et UNIO). Cependant, ce réseau, identifié par le code RA ([RESIF, 1995](#)) et par les triangles verts sur la Figure 2.3, comprend de nombreuses stations qui sont installées à proximité de zones fortement impactées par l'activité humaine, ce qui entraîne souvent un fort niveau de bruit. Par conséquent, il n'est pas toujours possible de détecter et d'identifier avec précision les arrivées des ondes sismiques locales de faible ou moyenne magnitude.

Le Réseau Large-Bande Permanent (RLBP) est un réseau sismologique spécialement conçu pour surveiller de manière continue le mouvement sismique sur le territoire métropolitain. Il couvre une large bande de fréquences et offre une grande sensibilité pour détecter les événements sismiques lointains ainsi que les petits séismes locaux. Les

institutions universitaires ont progressivement mis en place ce réseau, représenté par des triangles rouges et portant le code FR (RESIF, 1995) sur la Figure 2.3, dans les régions françaises les plus sismiquement actives. Initialement, ce réseau correspondait au RéNaSS (Réseau National de Surveillance Sismique) et, à partir de 2008, les stations large bande du RéNaSS sont devenues RLBP. Entre 2000 et 2010, seules quatre stations ont été installées dans le Massif armoricain. Afin d'accroître le nombre de stations et d'assurer une répartition homogène sur l'ensemble du territoire, la composante large bande du projet national Résif-Epos a été lancée en 2010. Ce projet a permis de renforcer le réseau de stations sismologiques sur tout le territoire français, y compris dans la région du Massif armoricain. Ainsi, entre 2010 et 2022, plus de 25 nouvelles stations large bande ont été installées dans la région, améliorant considérablement la couverture sismologique, comme le montre la Figure 2.3. Grâce à ce nouveau réseau, il est désormais possible d'obtenir une image complète et détaillée des structures crustales et lithosphériques. De plus, il offre une capacité inédite d'enregistrement des petits séismes, jamais atteinte auparavant.

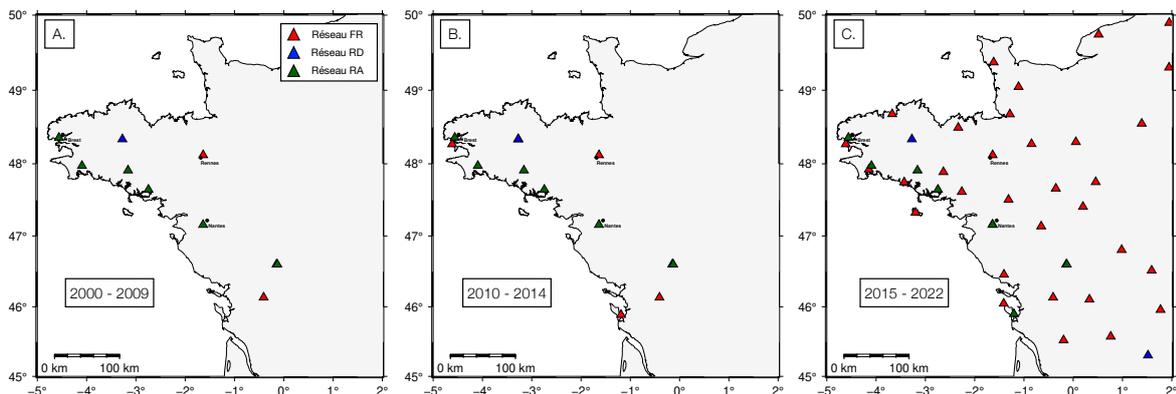


FIGURE 2.3 – Évolution du nombre de stations sismologiques permanentes dans le quart nord-ouest de la France entre 2000 et 2022. Le réseau FR comprend des stations large bande du réseau RLBP (Réseau Large Bande Permanent), le réseau RD comprend les stations large bande du LDG et le réseau RA comprend les stations du RAP.

2.2.2 Sismicité diffuse

En France métropolitaine, la sismicité instrumentale est généralement considérée comme modérée (Fig 2.4.A). À l'est de la France, on observe des séismes s'étendant de la région de Provence jusqu'au Fossé rhénan et aux Ardennes, en passant par l'arc alpin. Les Pyrénées, le Massif central et le Massif armoricain sont également des régions sismiquement actives. Le projet SI-Hex (Cara *et coll.*, 2015), réalisé par le Bureau

2.2. SISMICITÉ INSTRUMENTALE

Central Sismologique Français (BCSF) - Réseau National de Surveillance Sismique (RéNaSS), pour le compte de sept Observatoires des Sciences de l'Univers (OSU), et le CEA, a fourni un premier catalogue comprenant des localisations et des estimations de magnitude pour l'ensemble de la France métropolitaine pour la période 1962-2009. Afin de garantir un catalogue homogène, les magnitudes de moment (M_W) sont calculées à partir de magnitudes locales (M_l) en utilisant différentes conversions, ou directement à partir des formes d'onde pour les événements les plus importants. Pour la période de 1962 à 2009, deux gammes de magnitude ont été prises en compte :

$$M_W = M_{l-LDG} - 0,6 \quad \text{pour } 3,1 \leq M_l \leq 4, \quad (2.1)$$

$$M_W = 0,664M_{l-LDG} + 0,45 \quad \text{pour } M_l \leq 3,1. \quad (2.2)$$

L'étude de [Beucler *et coll.* \(2021\)](#) permet d'étendre ce catalogue pour la période 2010-2020 dans le quart nord-ouest de la France, fournissant ainsi une base de données homogène pour la période 1962-2020. Sur cette période, le catalogue est complété par la sismicité enregistrée et localisée par le BCSF-RéNaSS, dont les magnitudes ont été converties en M_W . Une première conversion en M_{l-LDG} est d'abord réalisée, puis en M_W .

L'élaboration d'un tel catalogue d'événements sismiques constitue la pierre angulaire pour évaluer l'aléa et le risque sismique. En analysant les événements sismiques passés, il est possible de déterminer les zones où l'activité sismique est la plus intense, les failles géologiques les plus actives, ainsi que la magnitude maximale atteignable. Ces informations sont essentielles pour estimer l'aléa sismique, qui évalue la probabilité d'occurrence d'un séisme de certaine magnitude dans une région spécifique sur une période de temps donnée. En utilisant ce catalogue d'événements sismiques, il est également possible d'évaluer le risque sismique, dans la logique de récurrence d'événements le long des failles. Cette évaluation tient compte des populations exposées, des infrastructures critiques et des conséquences potentielles d'un séisme.

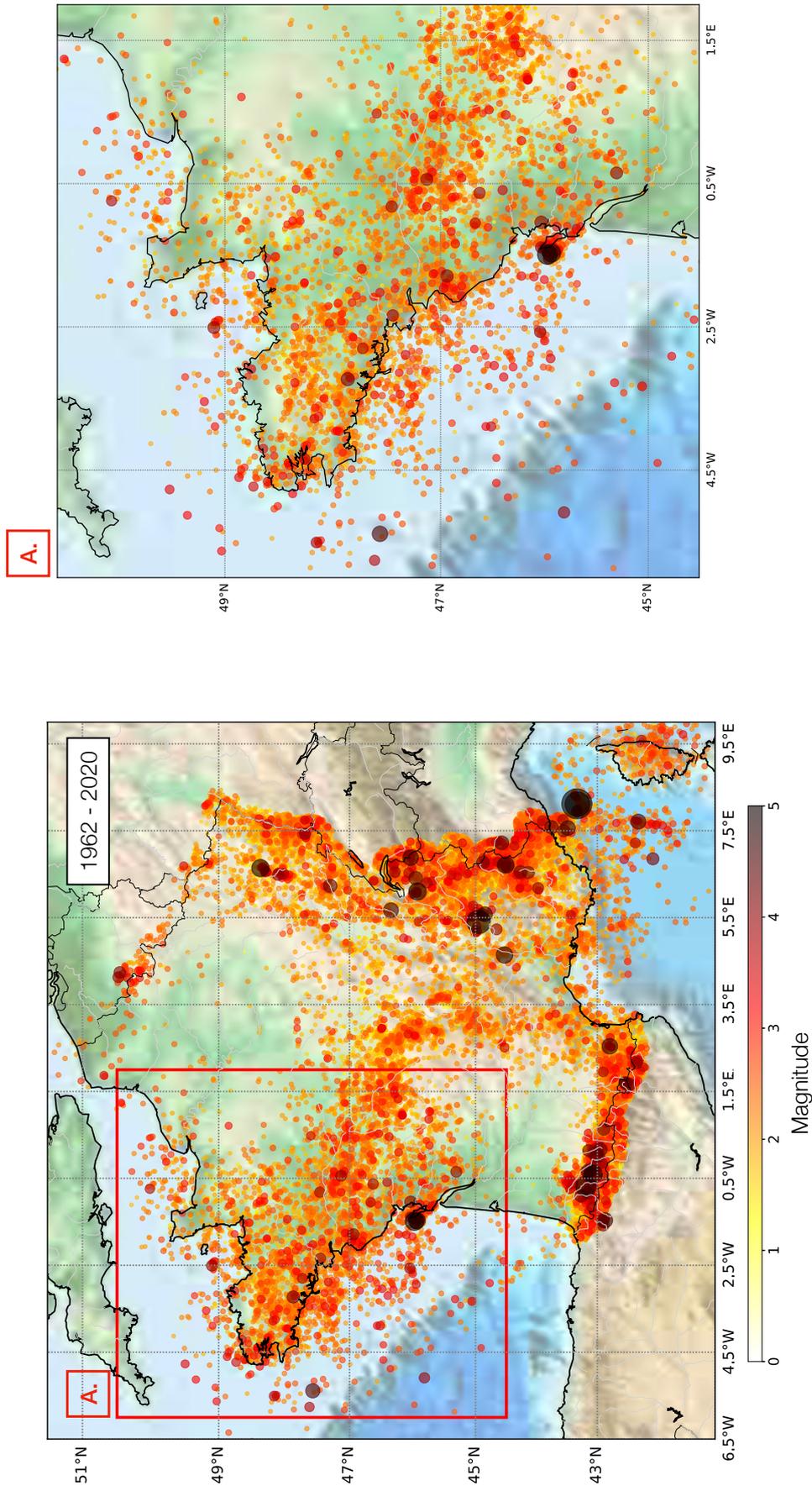


FIGURE 2.4 – Sismicité instrumentale de 1962 à 2020 en France métropolitaine. 76 076 événements sismiques naturels (données d'après : [Cara et coll., 2015](#); [Beucler et coll., 2021](#)). L'échelle de couleur et la taille représentent les magnitudes. Les séismes sont triés par magnitude, les plus fortes magnitudes recouvrant les plus petites.

2.2. SISMICITÉ INSTRUMENTALE

Dans le Massif armoricain, défini par le cadre A. dans la Figure 2.4, un total de 5 927 événements sismiques naturels ont été détectés, entre 1962 et 2020. Ceux-ci représentent une part modeste des séismes enregistrés dans l'ensemble de la France métropolitaine (7,7% du catalogue sur un total de 76 076 événements). Cependant, si nous examinons les magnitudes de ces événements, presque 30% des événements français de magnitudes égale à 3 sont localisés dans le Massif armoricain (Fig 2.5). Cette observation sera discutée dans la section suivante.

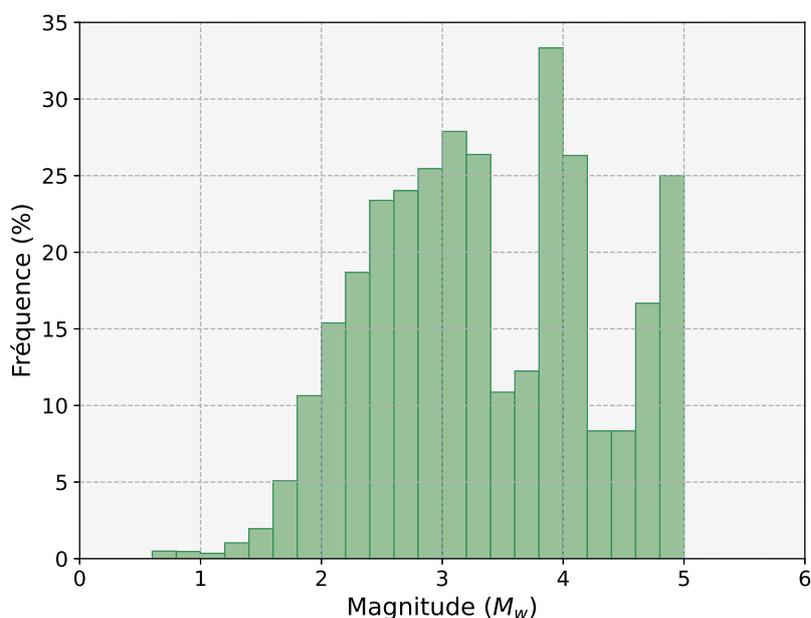


FIGURE 2.5 – Proportion des séismes localisés dans le nord ouest par rapport au nombre total de séismes métropolitains en fonction de la magnitude M_w (bins en magnitude de 0,2).

Depuis les années 1960, la sismicité enregistrée dans la région du Massif armoricain présente une répartition géographique diffuse (par exemple Judenherc *et coll.*, 2002). Les séismes sont principalement concentrés dans les zones où le socle rocheux affleure à la surface, et l'on observe un nombre limité d'événements dispersés sur le plateau continental ainsi que dans les bassins adjacents de Paris et d'Aquitaine. Cependant, contrairement à certaines régions sismiques où les séismes peuvent être alignés le long de structures tectoniques majeures, les événements sismiques dans le Massif armoricain ne semblent pas suivre de schéma clair le long de telles structures (Fig 2.4.A).

Il convient de noter que la concentration la plus importante de séismes dans la région se situe dans la direction des grandes structures varisques, selon une orientation de N120° E. Cette zone sismique s'étend sur une largeur d'environ 100 à 150 km, s'étirant du Finistère jusqu'au Massif central (par exemple Arroucau, 2006). Cependant, il est

intéressant de souligner que ces événements sismiques ne semblent pas être localisés précisément le long de structures tectoniques majeures telles que le CSA ou d'autres failles secondaires.

La répartition diffuse des séismes dans le Massif armoricain suscite des interrogations quant aux mécanismes géologiques responsables de cette activité sismique. Cette caractéristique peut être attribuée à l'histoire géologique complexe de la région, marquée par un réseau étendu de failles héritées qui se chevauchent et sont susceptibles de libérer des contraintes sismiques.

Une autre explication possible de cette répartition diffuse est liée à un biais observationnel dû au nombre limité de stations sismologiques disponibles dans la région. Ce manque de stations peut avoir un impact sur la précision de la localisation des séismes enregistrés. Cependant, le projet RLBP a permis l'installation de stations sismologiques dans la région (Fig 2.3). La précision de la localisation des séismes s'est ainsi améliorée au fil des années. L'incertitude est cependant particulièrement importante en ce qui concerne les profondeurs hypocentrales des séismes. En raison du nombre limité de données disponibles, les analystes étaient souvent contraints de tester différentes valeurs de profondeur *a priori*, ou d'imposer une valeur constante (longtemps fixée à 2, 5, 8, 10, 15 ou 20 km). Dans le catalogue, environ 40% des événements ont des profondeurs hypocentrales fixées à 20 km.

Pendant la période d'observation de 1962 à 2020, on observe une tendance à la récurrence des séismes de plus fortes magnitudes dans des directions proches ou parallèles à celle du CSA. Ces séismes se concentrent notamment dans des régions telles que la Pointe du Raz en Bretagne, le sud de la Bretagne, la Brière, en Vendée et l'île d'Oléron. Cela suggère une possible corrélation entre l'activité sismique et la présence du cisaillement, structure géologique majeure dans la région du Massif armoricain.

2.3 Magnitude de complétude élevée

L'analyse de la période de 1962 à 2020 (Fig 2.6.A) révèle que la sismicité en France métropolitaine se compose d'environ 7% des événements localisés dans la région Nord-Ouest. La Figure 2.6 met en évidence les distributions relatives des magnitudes (M_W) dans la partie nord-ouest de la France (en orange) et dans le reste du pays (en vert). L'analyse du catalogue sismique est effectuée pour deux périodes distinctes, avant et

2.3. MAGNITUDE DE COMPLÉTUDE ÉLEVÉE

après 2018, afin de mesurer le changement significatif dans la couverture de stations, qui a commencé à partir de 2015. Pour la période 1962-2017, nous observons que la majorité des événements situés dans le Massif armoricain ont des M_W comprises entre 1,5 et 2,5. Le déficit d'événements de faibles magnitudes, dû à la faible densité du réseau dans la région, se manifeste clairement avec seulement 5% des événements ayant des magnitudes inférieures à 1,5.

Au cours de la période plus récente (2018-2020), nous constatons une augmentation du nombre de stations sismiques dans l'ensemble de la France métropolitaine (Fig 2.3). Cette augmentation du nombre de stations ne semble en revanche pas être associée à un changement significatif dans la distribution des magnitudes dans le nord-ouest de la France. En effet, près de 85% des magnitudes se situent toujours entre 1,5 et 2,5 dans la région en question. Cependant, une amélioration significative est constatée dans la détection des petits événements sismiques dans le reste de la France (Fig 2.6.B), grâce à l'augmentation du réseau de stations sismologiques.

Cette compréhension de la sismicité de faible magnitude peut toutefois être largement influencée par la méthode de détection des événements naturels utilisée, le choix des seuils de détection, les conditions variables du bruit sismique et les différentes méthodes employées pour estimer la magnitude.

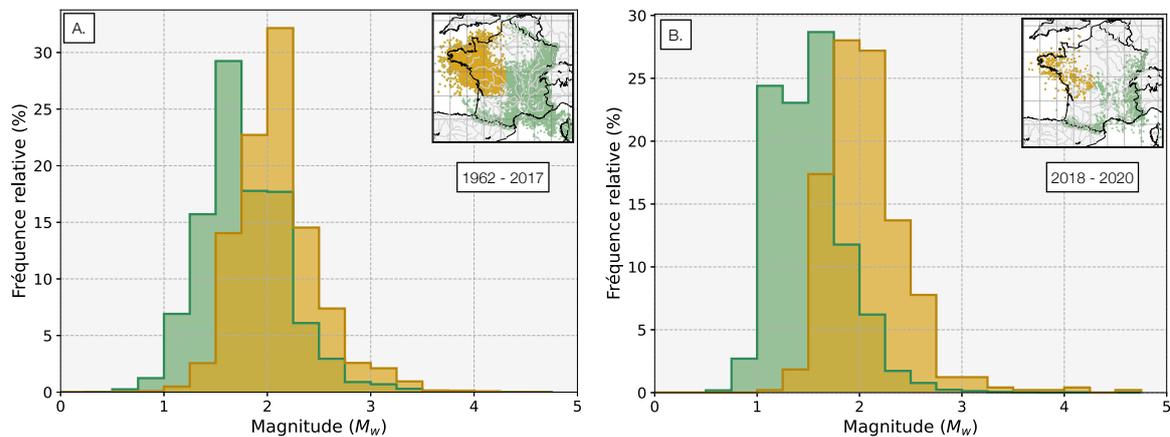


FIGURE 2.6 – Distributions des magnitudes de moment dans la partie ouest de la France (orange) et dans le reste de la France métropolitaine (vert) pendant deux périodes de temps différentes. Les histogrammes oranges sont calculés pour une région comprise entre 2° E et 6° W et 45° N et 51° N, en utilisant des bins de 0,25 (modifiée d'après [Beucler et coll., 2021](#)).

La distribution de la taille des séismes, dans une région donnée, peut également être étudiée à partir de la loi empirique de Gutenberg-Richter. Celle-ci est utilisée pour décrire la décroissance exponentielle du nombre de séismes avec la magnitude (Gutenberg et Richter, 1956),

$$\log_{10} N(M) = A - bM \quad \text{pour } M \geq M_c. \quad (2.3)$$

$N(M)$ représente le nombre cumulé de séismes dont la magnitude est supérieure ou égale à une valeur donnée M_c . La paramètre A décrit le taux de sismicité globale pouvant varier d'une région à une autre. La valeur b caractérise le rapport entre le nombre de grands séismes et de petits séismes et oscille autour de 1 (par exemple Fiedler *et coll.*, 2018). Enfin, la magnitude de complétude, notée M_c , correspond à la magnitude en dessous de laquelle tous les événements sismiques sont supposés être enregistrés dans le catalogue. Cela signifie que tous les séismes de magnitude supérieure ou égale à M_c sont inclus dans l'analyse et la modélisation de la loi de Gutenberg-Richter. Les séismes de magnitude inférieure à M_c peuvent ne pas être tous détectés ou enregistrés en raison des limitations de la géométrie du réseau de stations ou des seuils fixés pour la détection.

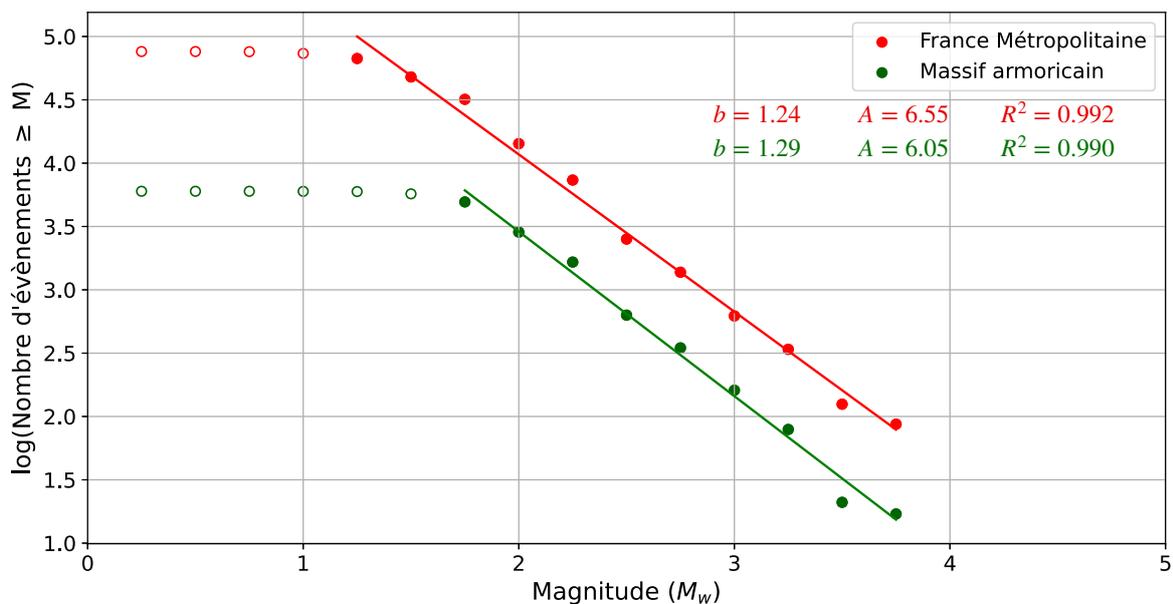


FIGURE 2.7 – Loi de Gutenberg-Richter : en France métropolitaine et dans le Massif armoricain (données d'après Cara *et coll.* (2015) et Beucler *et coll.* (2021), de 1962 à 2020). Les valeurs de A et b de l'équation 2.3 ont été déterminés à partir des cercles remplis.

2.3. MAGNITUDE DE COMPLÉTUDE ÉLEVÉE

La Figure 2.7 présente les lois de Gutenberg-Richter pour la France Métropolitaine et pour le Massif armoricain durant la période 1962-2020. Pour l'ensemble de la France métropolitaine, la magnitude de complétude est d'environ 1,25. En revanche, pour le Massif armoricain, celle-ci est estimée à 1,75. Cette magnitude de complétude plus élevée, dans notre région d'étude, indique qu'il manque une grande partie des petits séismes dans le catalogue d'événements sismiques. Ce manque est principalement dû à la faible densité de stations sismologiques dans la région avant 2010 (Fig 2.3).

Cependant, il convient de noter que notre région d'étude se caractérise par une atténuation significativement plus faible par rapport au reste de la France (Section 1.4.3). Cette particularité peut entraîner une surestimation importante des magnitudes, comme cela a été observé, par exemple, lors du séisme d'Hennebont, avec un écart entre la magnitude locale (M_l) de 5,7 et la magnitude de moment (M_W) de 4,3 (Perrot *et coll.*, 2005). En effet, la magnitude locale est calculée uniquement à partir de l'amplitude des ondes S et de la différence de temps entre les arrivées P et S (Richter, 1958; Savage et Anderson, 1995). La faible atténuation sismique observée dans le Massif armoricain peut entraîner des variations dans la décroissance des amplitudes des ondes S , ce qui peut potentiellement influencer les estimations de la magnitude locale M_l .

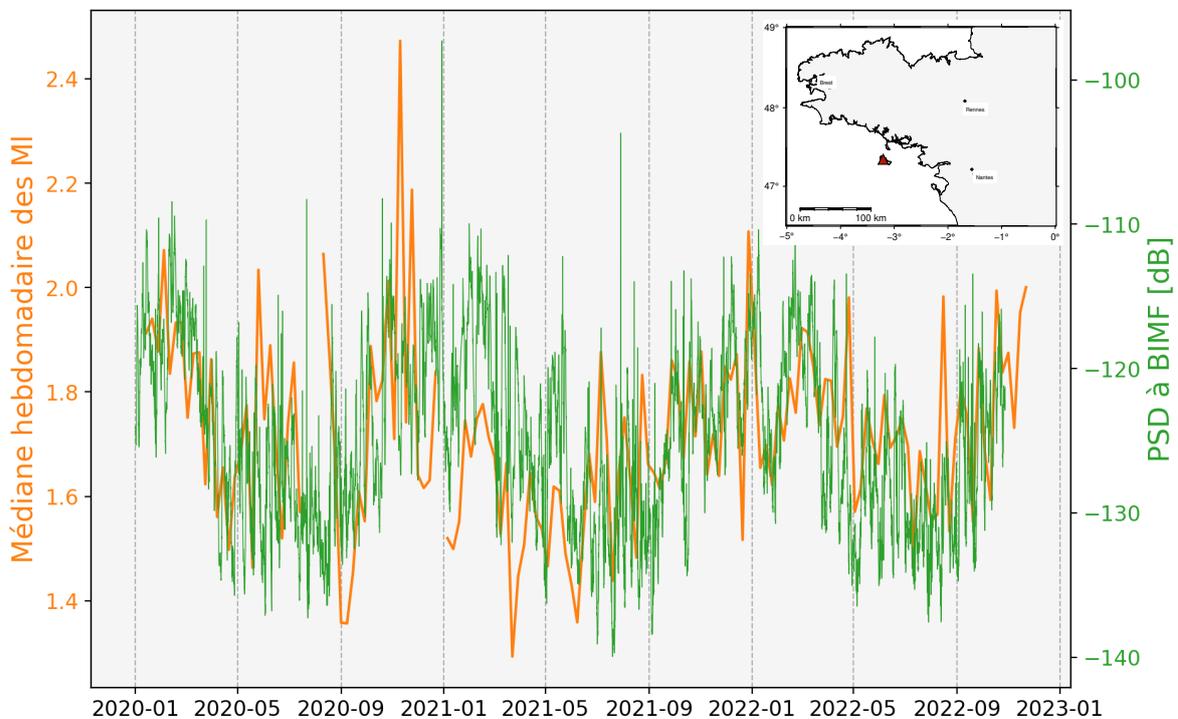


FIGURE 2.8 – PSD (Power Spectral Density) à la station BIMF (triangle rouge sur la carte) de janvier 2020 à janvier 2023 entre 0,1 et 1 Hz, en vert. Distribution de la médiane hebdomadaire des magnitudes (M_l) sur la même période, en orange.

Par ailleurs, la magnitude locale est généralement calculée entre 0,6 et 20 Hz. Cependant, le pic microsismique, caractéristique du bruit ambiant, est compris entre 0,6 et 1 Hz. La Figure 2.8 illustre la Densité Spectrale de Puissance (PSD, *Power Spectral Density*), en vert, enregistrée à la station BIMF, située à Belle-Île au sud de la Bretagne, entre 0,1 et 1 Hz, de janvier 2020 à janvier 2023. Une observation importante réside dans les variations saisonnières du signal sismique enregistré. Nous constatons une diminution du bruit sismique pendant l'été et une augmentation pendant l'hiver. En orange, nous reportons la distribution hebdomadaire des magnitudes (M_l), des séismes détectés dans la région, durant la même période. Une corrélation nette est visible entre les deux courbes. En été, des magnitudes allant jusqu'à 1,3 sont observés alors qu'en hiver, nous observons rarement des séismes de magnitude inférieure à 1,6. Ces résultats suggèrent que le pic microsismique exerce une influence significative sur le calcul des magnitudes locales (M_l).

Dans le cadre du projet SI-Hex, les magnitudes locales ont été converties en magnitudes de moment (M_W) pour tous les événements dont la M_l du LDG est inférieure à 4 (Cara *et coll.*, 2015). Ces variations saisonnières observées à partir des M_l du BCSF-RéNaSS peuvent donc impacter le catalogue 1962-2020. Il est important de prendre en compte ces facteurs pour interpréter les données sismiques et la loi de Gutenberg-Richter dans la région du Massif armoricain.

2.4 Sismicité non tectonique

Avec la densification du réseau de stations, le nombre d'événements sismiques détectés a fortement augmenté au cours de ces dix dernières années. Ceci soulève également la question de la distinction entre les séismes naturels et les signaux d'origine non tectonique. Parmi eux, nous pouvons avoir des séismes d'origine anthropique, le passage de météores ou l'enregistrement d'onde acoustique lors d'une éruption volcanique.

Contrairement aux séismes naturels, qui sont le résultat de forces géologiques internes à la Terre, les séismes anthropiques sont déclenchés par des actions humaines spécifiques. Les principales activités humaines qui peuvent provoquer des séismes anthropiques dans la région comprennent : 1) les tirs de carrières permettant l'extraction de ressources minérales et 2) les explosions de mines sous-marines le long de la côte française, conséquences à long terme de la Seconde Guerre mondiale. Ces séismes anthropiques peuvent varier en intensité et en magnitude, mais sont généralement de magnitudes modérées.

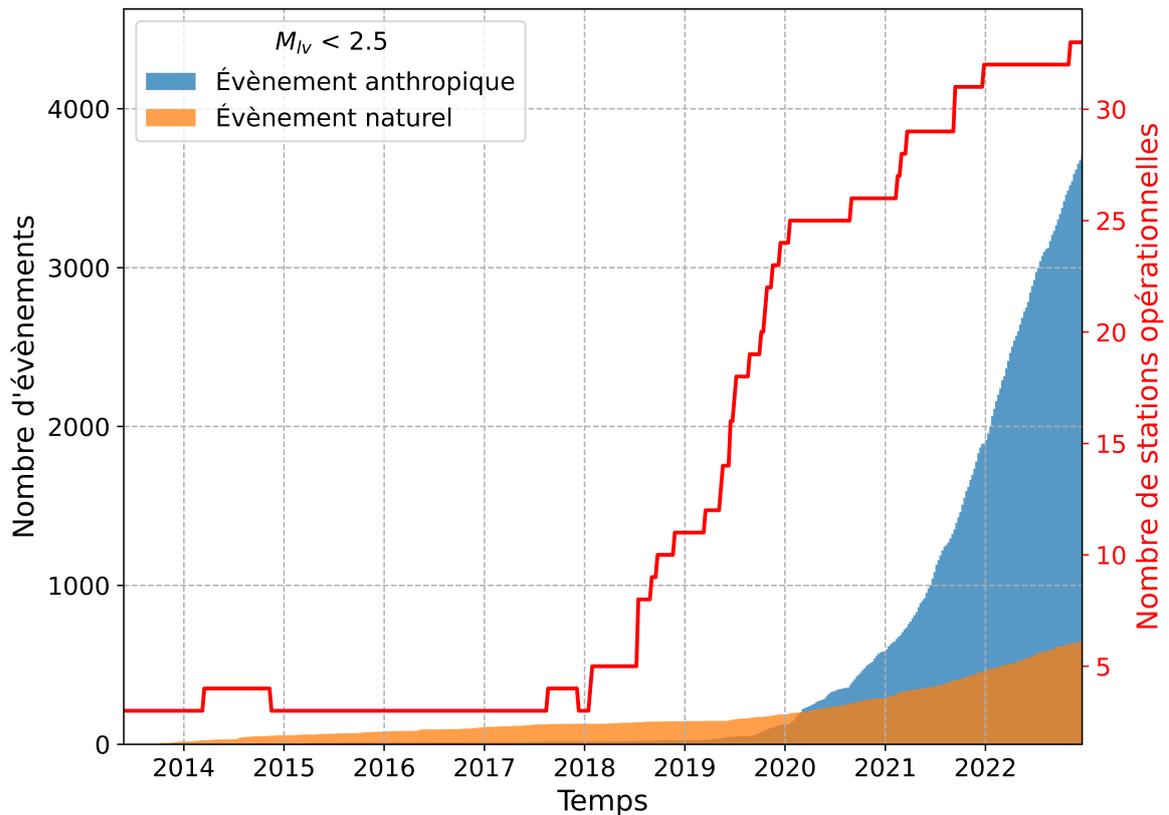


FIGURE 2.9 – Distribution cumulatif des événements naturels (orange) et anthropiques (bleu) de M_l supérieure ou égale à 2,5 en fonction du temps. Le nombre de stations opérationnelles permanentes dans la région est représenté en rouge (réseau FR).

Depuis 2013, ces événements d’origine humaine sont discriminés à la fois par le CEA et le BCSF-RéNaSS. À titre d’exemple, la Figure 2.9 présente deux diagrammes cumulatifs pour les événements naturels et les activités humaines d’une magnitude locale (M_l) inférieure à 2,5, détectés dans la région du Massif armoricain. À partir de mi-2018, grâce au déploiement des stations permanentes dans l’Ouest, la quantité d’événements naturels a augmenté de manière significative. Cependant, cette amélioration de la couverture sismique a également généré une hausse spectaculaire des détections de séismes d’origine anthropique, notamment des tirs de carrières, avec environ 6 fois plus d’événements anthropiques cumulés en 2023 par rapport aux naturels.

L’activité anthropique de la région est principalement liée à la présence abondante de carrières sur son territoire (Fig 2.10). Le Massif armoricain est caractérisé par une diversité géologique comprenant des formations riches en matériaux tels que le granite, le schiste, le calcaire et les grès. La localisation des tirs de carrières enregistrés par le BCSF-RéNaSS (Fig 2.10.B) montre une répartition suivant les frontières des affleurements cristallins avec le Massif armoricain et le Massif central.

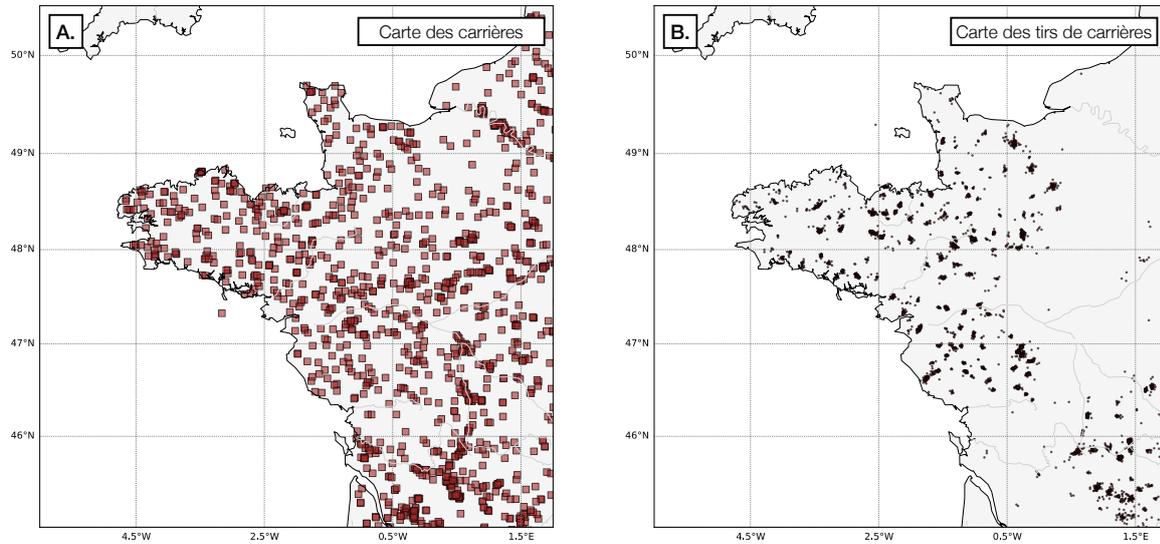


FIGURE 2.10 – Cartes de localisation des (A) des sites de carrière dans la zone d'étude et (B) des tirs de carrières enregistrés du 1er janvier 2013 au 1er juin 2023. La localisation des carrières est basée sur les données de carrières du Bureau de Recherches Géologiques et Minières (BRGM), <http://geoservices.brgm.fr/odmgm>.

Néanmoins, l'exploitation de ces événements sismiques, dont l'épicentre est préalablement connu (la source est en surface et le temps initial parfois bien contraint), offre une opportunité précieuse pour évaluer la performance des méthodes de localisation et d'imagerie sismique. De plus, cela permet d'approfondir notre compréhension des trajectoires des ondes sismiques et de mettre en évidence les variations latérales des vitesses sismiques régionales.

Les sismogrammes peuvent parfois révéler d'autres signaux atypiques qui ne sont pas liés aux séismes. Parmi ces signaux, on peut citer le passage de bolide au-dessus des stations sismologiques. Lorsqu'un météore traverse l'atmosphère terrestre, il peut provoquer une onde de choc et générer un bruit d'explosion audible. Cette onde peut être enregistrée par les capteurs sismologiques. Bien que ces signaux ne soient pas des séismes au sens strict, leur enregistrement par les stations sismologiques peut offrir l'opportunité d'étudier et de mieux comprendre les phénomènes astronomiques, tels que les météores, en utilisant les données sismiques. Un tel événement a eu lieu durant la nuit du dimanche 5 au lundi 6 septembre 2021. Un météore est passé au-dessus de la Bretagne, puis s'est entièrement désintégré dans l'atmosphère. Durant son trajet, plusieurs capteurs ont pu enregistrer son passage (Fig. 2.11). Au moment de l'impact, les ondes sismiques et acoustiques peuvent également être enregistrées par les stations sismologiques.

2.5. CHAMPS DE CONTRAINTE

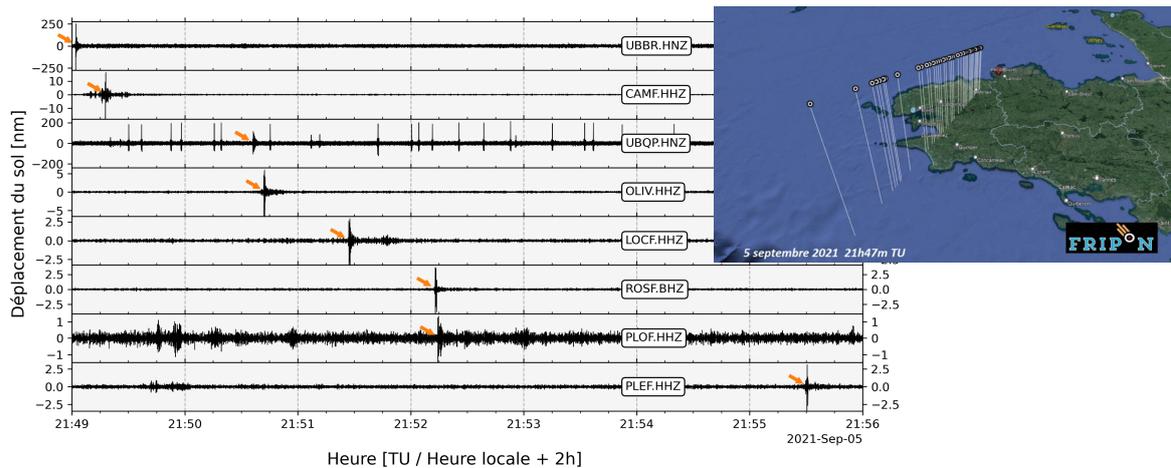


FIGURE 2.11 – Enregistrements sismiques à 8 stations permanentes du quart nord-ouest de la France lors du passage du météore de 5 septembre 2021 à 21h47 (UTC). La carte montre la trajectoire du bolide au-dessus de la Bretagne.

En plus du passage des météores, les sismogrammes peuvent également enregistrer d'autres signaux atypiques, tels que les ondes acoustiques générées par des événements volcaniques. Lorsqu'un volcan entre en éruption, il peut produire des explosions et des émissions de gaz, qui génèrent des ondes sonores dans l'atmosphère. Ces ondes acoustiques peuvent être captées par les stations sismologiques et enregistrées sur les sismogrammes. Au moment de l'éruption du Hunga Tonga, volcan sous-marin du royaume des Tonga au sud-ouest de l'océan Pacifique, du 14 au 15 janvier 2022, des ondes acoustiques ont également pu être enregistrés par les stations sismologiques de la région.

2.5 Champs de contrainte

Depuis ces trente dernières années, 180 mécanismes au foyer³ ont été déterminés dans l'ouest de la France métropolitaine (Nicolas *et coll.*, 1990; Amorèse *et coll.*, 2000; Mazabraud *et coll.*, 2013; Delouis, 2014).

La majorité des mécanismes au foyer anciens sont obtenus à partir de la polarité des premiers mouvements enregistrés par les réseaux permanents à courte période. Les plus récents sont déterminés à partir de l'inversion de la forme d'onde des enregistrements des stations à large bande les plus proches, fournis par les réseaux régionaux plus denses (méthode FMNEAR, Delouis, 2014). La Figure 2.12 montre un ensemble représentatif de mécanismes de la région et permet de mettre en évidence le schéma

3. Plus d'informations sur les mécanismes au foyer sont disponibles en Annexe A.3.

général de la déformation. Ainsi, dans le Massif armoricain, une grande partie des événements majeurs présentent des mouvements de glissement important associés à des composantes normales, avec des orientations des axes P et T variant autour des directions NW-SE et NE-SW, respectivement (Mazzotti *et coll.*, 2021). Les plans nodaux associés sont cohérents, à l'échelle régionale, avec les orientations des failles et des zones de cisaillement décrites.

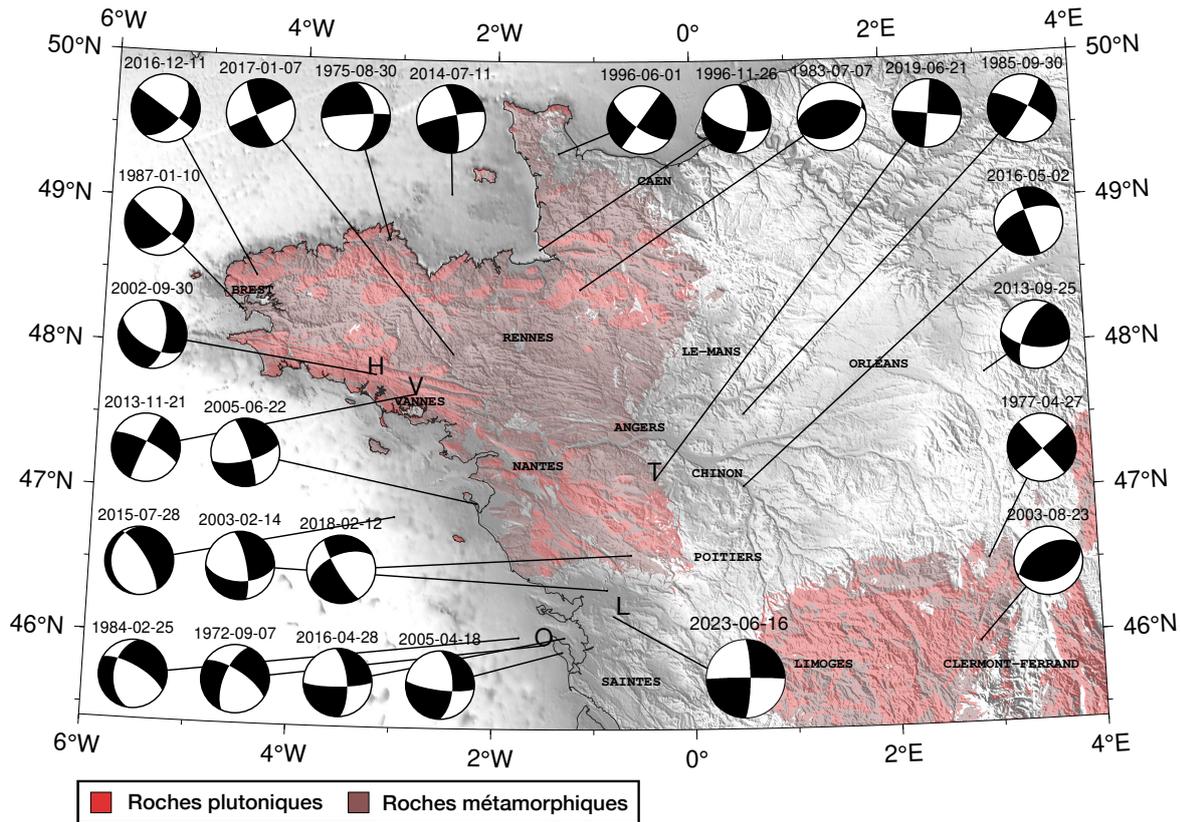


FIGURE 2.12 – Mécanismes au foyer des tremblements de terre les plus importants au cours des 50 dernières années dans l'ouest de la France (extrait de Mazzotti *et coll.*, 2021). Les failles majeures et importantes de la carte géologique numérique de la France (Chantraine *et coll.*, 2003) sont représentées respectivement par des traits pleins gris foncé épais et fins. Les zones métamorphiques et plutoniques sont extraites de la carte géologique numérique de la France (échelle 1/1000 000). Les lettres O, H, V et T correspondent respectivement aux événements d'Oléron, d'Hennebont, de Vannes et du Layon. Le séisme de La Laigne qui a eu lieu le 16 juin 2023 a été rajouté et correspond à la lettre L (modifiée d'après Beucler *et coll.*, 2021).

Depuis les années 1980, le projet *World Stress Map* a apporté une contribution significative à notre compréhension de l'état des contraintes dans la croûte et la lithosphère (Zoback *et coll.*, 1989; Heidbach *et coll.*, 2018). Grâce à ce projet, nous avons pu améliorer considérablement notre connaissance des contraintes présentes dans la majeure partie de la France et de l'Europe occidentale.

Dans ces régions, la contrainte compressive horizontale maximale est généralement orientée NW-SE à NNW-SSE, avec un azimut d'environ 135-160° (Fig 2.13). Cette orientation de la contrainte compressive horizontale maximale à l'échelle continentale est cohérente avec les tenseurs de contraintes locaux, obtenus à partir des inversions des mécanismes au foyer, dans le Massif armoricain.

Les contraintes horizontales associées pourraient résulter de deux sources principales de compression. D'une part, la compression pourrait être causée par l'expansion océanique le long de la dorsale médio-atlantique, où les plaques tectoniques s'éloignent l'une de l'autre, créant une force de divergence. D'autre part, la compression alpine, associée aux mouvements de convergence entre les plaques continentales, pourrait également contribuer à ces contraintes horizontales (Gölke et Coblentz, 1996). Cependant, les vitesses de déformation associées, dans ce contexte intra-plaque, semblent être extrêmement faibles (Frankel *et coll.*, 2012; Craig et Calais, 2014), voire imperceptibles (1 mm/an dans le Massif armoricain Nocquet, 2012). Étant donné l'absence d'accumulation de contraintes actuelles, le mécanisme de déclenchement des séismes à l'intérieur des plaques demeure difficile à appréhender. On se demande quels processus contrôlent le déclenchement et la récurrence temporelle des tremblements de terre à l'intérieur des plaques, certains d'entre eux pouvant atteindre une magnitude importante.

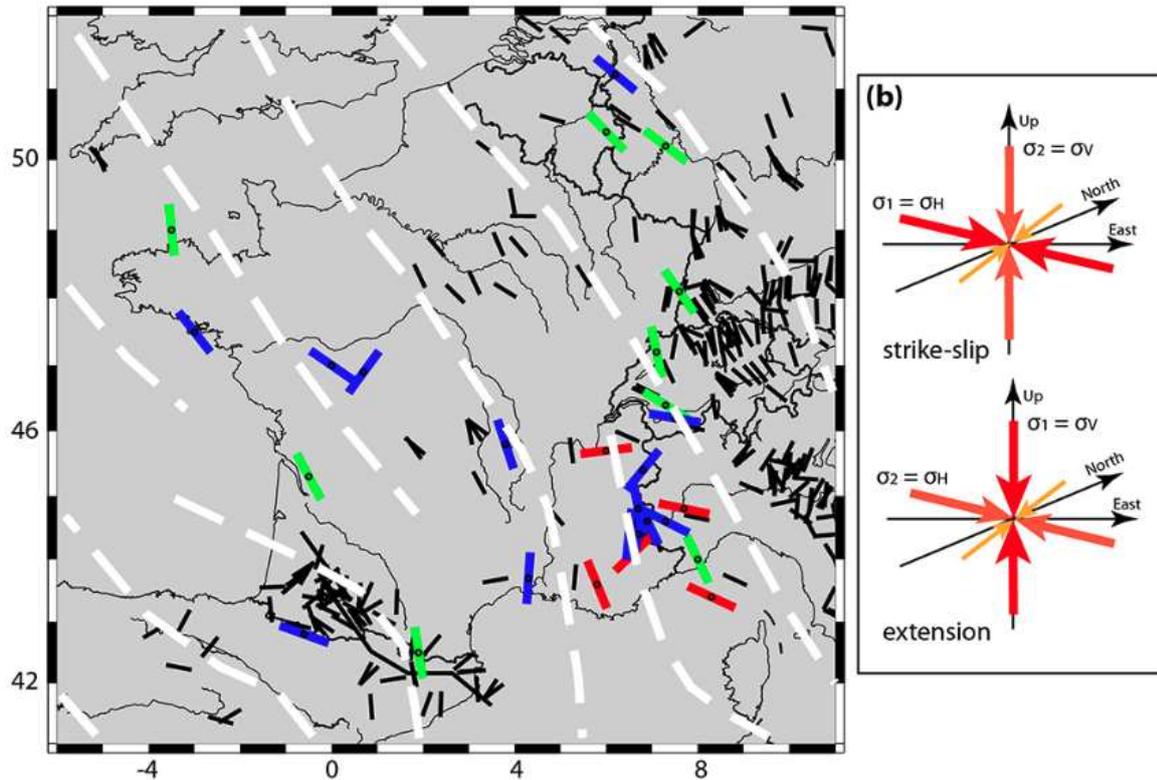


FIGURE 2.13 – Orientation de la contrainte compressive horizontale maximale en France métropolitaine et en Europe de l'ouest. Les tracés pleins noirs montrent les orientations de la contrainte de compression horizontale maximale *in situ* à partir de la carte mondiale des contraintes (World Stress Map, [Heidbach et coll., 2018](#)). Les lignes blanches en pointillés épais montrent l'orientation de contrainte lissée. Les barres vertes, bleues et rouges montrent l'orientation de contrainte à partir d'inversions de mécanismes au foyer. (b) Schéma de l'état de contrainte avec permutations entre σ_1 et σ_2 permettant des oscillations entre les régimes de glissement et d'extension tout en maintenant l'orientation de la contrainte ([Mazzotti et coll., 2020](#)).

2.6 Vers une meilleure compréhension de cette sismicité ?

Le nord-ouest de la France est considéré comme une région continentale stable, située loin des frontières de plaques et soumise à des taux de déformation extrêmement faibles (Nocquet, 2012). Néanmoins, les processus qui contrôlent la déformation, la sismicité, l'aléa sismique et le risque sismique, dans les zones intra-plaques, sont fortement débattus (Stein et Mazzotti, 2007; Clark *et coll.*, 2012; Calais *et coll.*, 2016).

Dans le nord-ouest de la France, le champ de contraintes actuel, déduit des inversions du mécanisme focal des séismes et de quelques mesures de contraintes *in situ*, est caractérisé par une orientation globale de la contrainte horizontale maximale suivant une direction NW-SE à NNW-SSE. Les failles et les zones de cisaillement, résultant des événements tectoniques passés (Section 1.2), sont favorablement orientées pour se rompre dans ce champ de contraintes. En raison des limitations des catalogues de séismes, à la fois instrumentaux et historiques, ainsi que du manque de données sur les failles actives potentielles, il est difficile d'établir une corrélation précise entre un séisme donné et une faille spécifique dans la région nord-ouest de la France.

Bien qu'il n'existe pas de preuve claire de rupture de grandes failles en surface, certaines études spécifiques suggèrent que la sismicité actuelle peut être potentiellement contrôlée par des structures héritées (par exemple Mazzotti et Gueydan, 2018; Kaub *et coll.*, 2021). Dans les régions continentales stables, une corrélation spatiale de premier ordre entre la sismicité et l'héritage structural est observée et discutée depuis les années 1970 (Sykes, 1978; Johnston, 1989). L'héritage structural est généralement défini comme des structures géologiques à l'échelle de la lithosphère associées à des événements tectoniques majeurs passés (généralement du Paléozoïque ou plus anciens) et à des déformations importantes affectant des régions de 10 à plusieurs centaines de km. Au premier ordre, une corrélation spatiale entre la sismicité et l'héritage structural peut également être observée dans le Massif armoricain comme l'illustre la Figure 2.14 (Mazzotti *et coll.*, 2020). L'activité sismique se produit de préférence dans un rayon d'environ 50 km autour des systèmes de failles et le socle varisque : par exemple, la zone de faille sud-armoricaine.

Cependant, cet héritage tectonique ne peut expliquer à lui seul l'ensemble de la sismicité de la région. Plusieurs études suggèrent que l'héritage structural des orogénèses

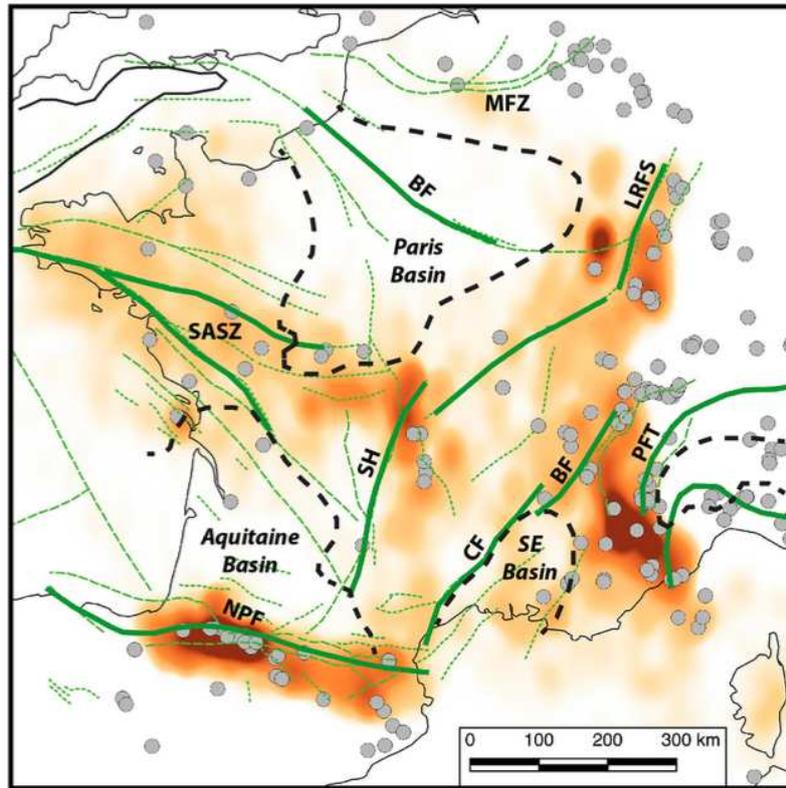


FIGURE 2.14 – Relation entre l'héritage structural et la sismicité en France métropolitaine. Les lignes vertes pleines et pointillées noires indiquent les principales failles et zones de faille crustales (MFZ : Midi Fault Zone ; BF : Bray Fault ; LRFS : Lower Rhine Fault System ; SASZ : South Armorican shear Zone ; SH : Sillon Houiller ; NPF : North Pyrenees Fault ; CF : Cévennes Fault ; BF : Belledonne Fault ; PFT : Pennic Front Thrust). La couleur orange montre la densité relative des séismes instrumentaux ($M_W \geq 2,0$, catalogue SI-Hex). Les lignes continues en pointillés indiquent les principaux bassins sédimentaires. Les cercles gris indiquent les séismes historiques et instrumentaux ($M_W \geq 4,5$) (Mazzotti et coll., 2020).

cadomienne et varisque se combine à des déclencheurs internes ou externes, sans qu'il s'agisse nécessairement d'une relation exclusive. Ces variations locales des contraintes, associées à des facteurs agissant à différentes échelles de temps, peuvent jouer un rôle significatif dans la sismicité des zones continentales stables (Levandowski *et coll.*, 2018).

Parmi ces variations transitoires, l'ajustement isostatique glaciaire, également appelé rebond post-glaciaire, est la réponse de la Terre solide aux charges et décharges de surface associées aux fluctuations des calottes glaciaires, des glaciers. Damon *et coll.* (2022) ont modélisé les ajustements isostatiques liés à la fonte des glaces dans l'Europe occidentale afin d'estimer leur impact en champ lointain sur la stabilité des failles dans l'est du bassin parisien. Les ajustements considérés peuvent induire des perturbations de contraintes suffisamment importantes, de l'ordre de quelques MPa, pour

2.6. VERS UNE MEILLEURE COMPRÉHENSION DE CETTE SISMICITÉ ?

potentiellement réactiver des failles. Cependant, les structures de failles connues dans l'est du bassin de Paris semblent être mal orientées pour une rupture potentielle, et seules quelques-unes d'entre elles peuvent potentiellement être réactivées. Le rebond post-glaciaire est connu pour avoir un impact sur la sismicité dans des régions telles que le Nord-Est des États-Unis, le Canada ou la Norvège (par exemple [Wu, 1998](#); [Bungum et coll., 2005, 2010](#); [Steffen et coll., 2014](#); [Craig et coll., 2016](#)). Dans ces régions, les ajustements isostatiques glaciaires peuvent se traduire par des perturbations de contraintes importantes, jusqu'à quelques dizaines de MPa pouvant être associées à des crises de sismicité. Il est important de noter que certaines de ces régions étaient recouvertes par les glaciers lors de la dernière glaciation, ce qui n'est pas le cas du Massif armoricain. De plus, le nord-ouest de la France est situé entre 500 et 1500 km des calottes glaciaires fennoscandienne et celtique, à environ 500 km des champs de glace alpins, et de 10 à 100 km des glaciers mineurs du Massif central. Bien que l'effet potentiel de ces différentes sources sur la sismicité reste à étudier, il est probablement beaucoup plus faible que celui observé directement au niveau des grandes calottes glaciaires.

En plus du rebond post-glaciaire, l'ajustement flexural et isostatique en réponse à l'érosion et à la sédimentation peut également influencer la sismicité ([Dadson et coll., 2003](#); [Champagnac et coll., 2007](#); [Bilek, 2010](#); [Calais et coll., 2010](#); [Steer et coll., 2014](#)). Ces processus d'érosion et de sédimentation peuvent provoquer des variations locales de contraintes, ce qui peut à son tour affecter la stabilité des failles et potentiellement déclencher des séismes (par exemple [Steer et coll., 2020](#); [Jeandet Ribes et coll., 2020](#)). [Mazzotti et coll. \(2023\)](#) présentent l'impact de l'érosion localisée, à long terme, sur les contraintes crustales et l'augmentation ou l'inhibition de la sismicité dans les régions continentales stables. Leurs modèles indiquent que, dans de nombreux cas, l'érosion localisée à long terme peut favoriser l'instabilité des failles. En effet, après quelques millions d'années, les taux d'érosion lents peuvent entraîner des perturbations des contraintes pouvant aller jusqu'à quelques MPa. Le rôle du climat à long terme et de la dénudation du paysage régional sur la déformation et la sismicité dans les zones intra-plaques n'a fait l'objet que de quelques études axées sur des régions spécifiques telles que le Tennessee oriental ([Gallen et Thigpen, 2018](#)) ou les Pyrénées centrales ([Genti et coll., 2016](#)). Cependant, il existe une corrélation de premier ordre entre les variations spatiales de l'érosion à long terme, la présence de structures héritées à l'échelle crustale/lithosphérique et l'activité sismique. Dans les régions de dépôts, comme les bassins sédimentaires épais (Bassin Parisien et Bassin Aquitain, [Fig 2.4](#)) l'activité sismique est très faible, voire nulle. Les zones présentant des taux d'érosion

plus élevés affichent une activité sismique plus importante (par exemple, les Pyrénées ou les Alpes). Enfin, les zones à faible taux d'érosion présentent de faibles niveaux de sismicité (par exemple, la Bretagne).

Enfin, les fluides peuvent également jouer un rôle important dans la sismicité des zones continentales stables. Les variations transitoires hydrologiques et météorologiques peuvent induire des changements dans les contraintes et la pression des fluides souterrains. Plusieurs études ont montré que des événements sismiques peuvent être déclenchés par des variations de pression des eaux souterraines liées à des phénomènes tels que les précipitations intenses, la fonte des neiges ou les fluctuations saisonnières des niveaux d'eau dans les nappes phréatiques (par exemple [Johnson et McEvilly, 1995](#); [Bollinger *et coll.*, 2007](#); [Leclère et Calais, 2019](#); [Shiddiqi *et coll.*, 2023](#)). Le Massif armoricain comporte énormément de fractures en surface qui pourraient faciliter la circulation de fluides en profondeur. L'interaction entre les mécanismes transitoires hydrologiques et météorologiques et la sismicité dans les zones continentales stables est cependant complexe et dépend de nombreux facteurs, tels que la perméabilité des roches, la géométrie des aquifères et la réponse mécanique des failles aux variations de pression des fluides. Comprendre ces interactions est essentiel pour évaluer les mécanismes de déclenchement des séismes dans ces régions.

Afin d'améliorer la compréhension de la sismicité dans une telle région tectonique et d'établir des liens entre l'activité sismique et des déclencheurs externes, plusieurs facteurs sont essentiels :

- **Un réseau de stations suffisamment dense.** Il est crucial d'avoir un réseau de stations sismiques bien réparties afin de détecter un maximum d'événements sismiques. Le projet national RLBP, initié en 2010, a permis une augmentation significative du nombre de stations dans la région (Figure 2.3 et 2.9), ce qui nous permet d'étudier la sismicité de manière plus approfondie.
- **Une méthode de localisation adaptée à la région.** La méthode développée par [Haugmard \(2016\)](#) est particulièrement pertinente dans ce contexte. Elle ne nécessite pas l'utilisation d'un modèle de structure préétabli, ce qui la rend adaptée à une région où les connaissances sismologiques peuvent être limitées. Cette méthode permet de déterminer simultanément les paramètres du séisme (latitude, longitude, profondeur et temps origine) et les paramètres de propagation du milieu (vitesses des ondes P profondeur du Moho, etc.) et ne repose donc pas sur un modèle de référence fixe. De plus, le temps origine déterminé par

cette approche est directement mis en relation avec la profondeur de l'événement. Cette méthode est donc particulièrement adaptée pour la détection lors de campagne post-sismiques ou pour rechercher les petits séismes avec peu de stations sismologiques.

- **Un catalogue d'événements sismiques naturels le plus complet possible.** Grâce au déploiement de nouvelles stations permanentes Résif-Epos, nous avons constaté une augmentation de la quantité d'événements sismiques naturels détectés dans la région. Cependant, en ce qui concerne la magnitude des événements, nous n'avons pas observé de changement significatif avec l'ajout de ces nouveaux capteurs. Cela suggère que notre catalogue d'événements sismiques n'est pas encore exhaustif dans la région étudiée. Par conséquent, il est essentiel de poursuivre les efforts de détection et d'enregistrement des événements sismiques de faible magnitude afin d'améliorer la couverture du catalogue existant et d'obtenir une vision plus complète de la sismicité régionale.

L'abaissement du seuil de détectabilité des événements sismiques dans la région engendre cependant plusieurs surcoûts, tels que des temps de calcul plus longs et une charge de travail manuelle accrue. Face à ces contraintes, se pose la question de savoir s'il est réellement important de détecter et d'enregistrer ces séismes pour mieux comprendre la sismicité de la région.

Dans les régions sismiquement actives, les séismes de forte magnitude sont des événements peu fréquents. En revanche, les séismes de faible magnitude se produisent plus fréquemment. Par conséquent, une meilleure détection de ces séismes de faible magnitude permettrait d'apporter une plus grande robustesse statistique à l'étude des comportements sismiques à court terme, ce qui contribuerait à une meilleure compréhension de leur origine. Ceci est particulièrement significatif dans une région continentale stable, où les mécanismes responsables de ces tremblements de terre restent difficiles à comprendre. L'étude des petits séismes permettrait également d'établir une cartographie plus précise du comportement sismique de la région, en identifiant potentiellement des répliques ou des essaims qui n'auraient pas été détectés auparavant.

Ainsi, afin de mieux comprendre et caractériser la sismicité naturelle de la région, il est essentiel de développer une méthode de détection robuste et adaptée spécifiquement à l'étude des séismes de faible magnitude. Compte tenu de l'augmentation exponentielle du nombre d'événements anthropiques détectés dans la région, il devient également nécessaire de mettre au point une méthode de discrimination fiable et automatique

permettant de se concentrer sur l'étude des signaux d'origine naturelle.

Pour l'agence nationale, le BCSF-RéNaSS, il est difficile d'adapter leur méthode de détection afin de détecter ces événements sismiques manquants dans la région. Leur mission principale est la surveillance et la détection en quasi temps réel des événements sismiques dans toute la France métropolitaine. Le temps et les efforts supplémentaires nécessaires pour effectuer cette tâche de manière manuelle seraient considérables.

Ainsi, dans le cadre de cette thèse, l'accent a été mis sur le développement d'une telle méthode, impliquant la détection de signaux de faible amplitude et la discrimination des signaux détectés. Dans la suite de ce manuscrit, je vais vous présenter les méthodes de détection et de discrimination développées et utilisées dans la littérature. Ensuite, j'expliquerai les choix que nous avons faits pour élaborer une méthode adaptée à notre région d'étude.

Chapitre 3

Méthodes existantes et apport de l'intelligence artificielle pour la détection des séismes

Sommaire

3.1	État de l'art des méthodes classiques	50
3.1.1	Détection par analyse de l'énergie	50
3.1.2	Détection par corrélation de segments temporels	53
3.1.3	Discrimination des événements sismiques	56
3.2	Apport du <i>machine learning</i> dans l'observation des évé- nements sismiques	58
3.2.1	Introduction - Définitions	58
3.2.2	Détection et identification des phases sismiques	66
3.2.3	Discrimination des événements sismiques	70
3.3	Bilan des approches	71

La détection automatisée des séismes est un problème de longue date en sismologie, les premiers algorithmes ayant été mis au point à la fin des années 1970 (par exemple Swindell et Snell, 1977; Allen, 1978).

L'utilisation de ces méthodes de détection automatique des séismes revêt une importance cruciale en sismologie, offrant la capacité de surveiller et comprendre l'activité et l'aléa sismique de notre planète. Elle permet une détection précoce des séismes, une alerte rapide des populations et une meilleure compréhension de l'aléa sismique. Les

premières méthodes de détection automatisée des signaux sismiques se sont appuyées sur les caractéristiques de ces signaux, telles que leur amplitude, leur fréquence et leur durée. Ces caractéristiques ont été utilisées pour développer des algorithmes qui appliquaient des seuils prédéfinis. Ainsi, les portions du signal continu dépassant ces seuils étaient identifiées comme des zones dignes d'intérêt.

À une époque, les enregistrements sismiques étaient limités aux moments où un séisme était détecté. Au fil des années, avec les progrès de l'informatique et des techniques de traitement du signal, les méthodes de détection automatique se sont affinées. Nous disposons aujourd'hui d'un signal sismique enregistré en continu. Au cours des dix dernières années, l'utilisation de l'apprentissage automatique, notamment des réseaux de neurones, a permis d'améliorer la capacité des systèmes à reconnaître les caractéristiques spécifiques des séismes et à les distinguer du bruit de fond. L'augmentation du nombre de stations, la collecte massive de données sismiques en temps réel et l'utilisation de méthodes avancées d'apprentissage automatique ont permis une détection plus rapide, une localisation plus précise des séismes et une réduction des fausses alertes.

Dans une première partie, je présenterai les méthodes dites « classiques » utilisées pour la détection et la discrimination des événements sismiques. Puis, je réaliserai une introduction sur l'intelligence artificielle et terminerai par la présentation des méthodes de détection et de discrimination utilisant l'apprentissage profond.

3.1 État de l'art des méthodes classiques

3.1.1 Détection par analyse de l'énergie

La méthode STA/LTA est sûrement la méthode la plus utilisée pour la détection automatique d'événements sismiques impulsifs ayant un bon rapport signal sur bruit (par exemple Allen, 1982; Withers *et coll.*, 1998; Trnkoczy, 2009; Sharma *et coll.*, 2010). Sur chaque composante de chaque station, le signal continu est parcouru par deux fenêtres coulissantes : une fenêtre longue *Long Term Average* (LTA) et une fenêtre courte *Short Term Average* (STA) (Fig 3.1). La fenêtre LTA permet de caractériser le bruit de fond par la moyenne du signal sur une longue durée. Parallèlement, la fenêtre STA est sensible aux augmentations rapides d'amplitude, et, par conséquent, à l'arrivée potentielle d'une onde sismique. Ainsi, à chaque instant i , un rapport entre ces deux fenêtres (STA/LTA) est calculé. Les fonctions STA et LTA sont calculées grâce à des relations de récurrence,

$$STA_i(t + \delta t) = STA_i(t) + \frac{1}{\Delta t_S} [u_i^2(t) - u_i^2(t - \Delta t_S)], \quad (3.1)$$

$$LTA_i(t + \delta t) = LTA_i(t) + \frac{1}{\Delta t_L} [u_i^2(t) - u_i^2(t - \Delta t_L)], \quad (3.2)$$

où $u_i(t)$ correspond au signal de la composante i à l'instant t , STA_i à la fenêtre STA, LTA_i à la fenêtre LTA, Δt_S à la durée de la fenêtre STA et Δt_L à la durée de la fenêtre LTA. Si ce rapport dépasse une valeur seuil prédéterminée, cela indique l'arrivée potentielle d'une onde sismique. La méthode STA/LTA peut être difficile à mettre en œuvre car les paramètres de la taille de la fenêtre et le seuil à utiliser peuvent dépendre, lorsqu'on optimise la méthode, de chaque station. Un exemple d'application du STA/LTA est présenté en Figure 3.2.

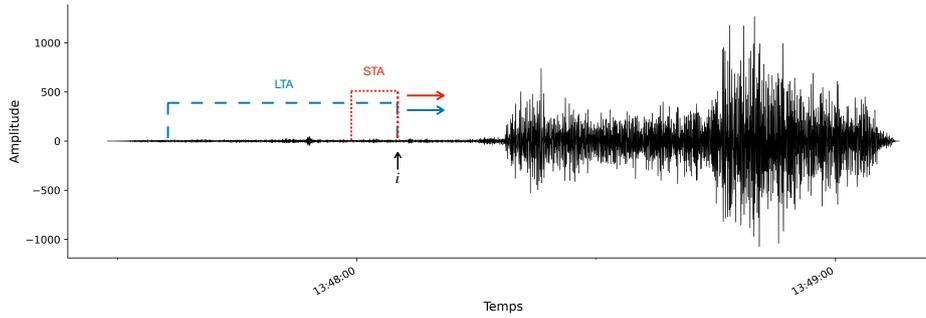


FIGURE 3.1 – Fonctionnement du rapport STA/LTA. Le cadre bleu, en pointillé, représente la fenêtre coulissante LTA. Le cadre rouge, en pointillé, correspond à la fenêtre coulissante STA, parcourant le signal dans la fenêtre LTA. Le rapport entre ces deux fenêtres est réalisé à l'instant i . Le sismogramme correspond au séisme du 20 août 2021 à 13h48 enregistré par la station PLEF (composante verticale), (modifiée d'après Fu et coll., 2020).

L'algorithme STA/LTA peut être appliqué à des traces brutes ou à des fonctions de traces, issues de traitements particuliers, appelées fonctions caractéristiques (CF). Baer et Kradolfer (1987) introduisent la fonction d'enveloppe comme CF et utilisent un seuil dynamique pour détecter les signaux noyés dans le bruit afin d'améliorer le STA/LTA. Parallèlement, Swindell et Snell (1977) ont développé le détecteur Z (Z-detect),

$$Z_{x_n} = \frac{x_n - \mu}{\delta} \quad (3.3)$$

où x_n correspond à une fenêtre STA, μ est la moyenne de la fenêtre STA, et δ est l'écart type de la fenêtre STA. Il présente l'avantage de s'ajuster automatiquement à

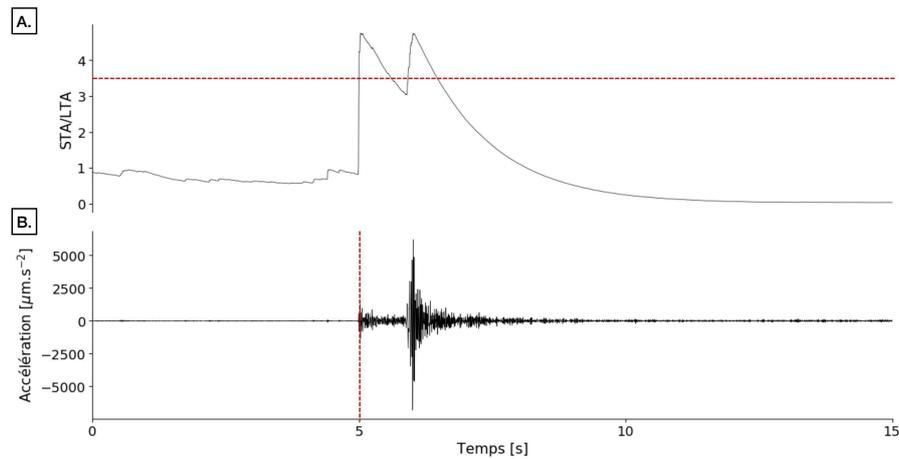


FIGURE 3.2 – Exemple de détection par STA/LTA. (a) Fonction STA/LTA récursif avec, en rouge, le seuil établi à 3,5. La durée de la fenêtre STA est de 1 seconde et la fenêtre LTA de 5 secondes. Le déclenchement correspond à l'intersection entre la courbe de STA/LTA et le seuil (ligne rouge). (b) Sismogramme de l'événement datant du 21 juin 2019 à 20h03, avec, en rouge, la position de la détection.

l'écart de la moyenne. Si le bruit de fond est faible, seule une petite variation du signal est nécessaire pour faire varier la valeur de sortie du Z-detect. Si le bruit de fond est important, une grande variation du signal est alors nécessaire pour faire varier la valeur de sortie du Z-detect.

Ce type d'approches peut également être utilisé pour l'identification automatique des temps d'arrivées des ondes sismiques. Nous pouvons observer dans la Figure 3.2 que la fonction STA/LTA dépasse le seuil fixé à deux reprises, une fois pour l'arrivée de l'onde P puis pour l'arrivée de l'onde S . Baillard *et coll.* (2014) présentent une autre méthode d'identification de phases basée sur la kurtosis. La fonction kurtosis est une mesure statistique utilisée pour évaluer la forme et la distribution des données d'un enregistrement de séisme. Elle quantifie « l'aplatissement » de la distribution des amplitudes des signaux sismiques. En appliquant la fonction kurtosis au signal sismique continu, nous pouvons obtenir des informations sur la présence de séismes distincts et leurs amplitudes. Une valeur de kurtosis élevée indique une distribution plus pointue, ce qui peut suggérer la présence de signaux sismiques impulsifs et concentrés. À l'inverse, une valeur de kurtosis faible indique une distribution plus aplatie, ce qui peut indiquer des signaux plus diffus. Ainsi, l'arrivée des deux phases sismiques P et S peut être identifiée grâce à cette fonction.

Plus récemment, [Frank et Shapiro \(2014\)](#) ont décrit une méthode alternative, reprise par [Beaucé *et coll.* \(2019\)](#), basée sur l'énergie sismique, permettant la détection et la localisation simultanée des événements sismiques. Pour cela, l'approche se base sur la cohérence de l'énergie sismique dans un réseau de récepteurs pour la détection automatique des événements. En pratique, le volume sous la région d'étude est discrétisé en une grille où chacun des points représente un emplacement possible de la source sismique. Chacune de ces sources hypothétiques est associée, à partir d'un modèle de vitesse suffisamment précis, à une collection de temps de parcours des ondes P et S vers chacune des stations. Pour chaque emplacement de source, les formes d'ondes, décalées en fonction de ces différents temps de parcours, sont ensuite sommées. Un seuil de détection est établi. Si ce seuil est dépassé alors une cohérence de l'énergie sismique entre les stations est observée et l'événement sismique est détecté.

3.1.2 Détection par corrélation de segments temporels

Des approches de type *template matching* sont souvent utilisées pour la détection des séismes. Elles permettent de détecter de petits séismes par corrélation entre des formes d'ondes sismiques de référence et le signal continu (par exemple [Gibbons et Ringdal, 2006](#)). Ainsi, des événements sismiques de formes similaires à ceux de référence mais de plus faible amplitude peuvent être détectés.

En pratique, la corrélation mesure la similitude entre deux signaux par translation (temporelle ou spatiale). Ainsi, il est possible de récupérer un retard ou décalage pour lequel nous pouvons superposer les deux signaux (Fig 3.3). La fonction d'inter-corrélation entre les signaux $u(t)$ et $v(t)$ est

$$C_{uv}(\tau) = u(t) \otimes v(t) = \int_{+\infty}^{-\infty} u(t + \tau)v^*(t)dt = \int_{+\infty}^{-\infty} u(t)v^*(t - \tau)dt, \quad (3.4)$$

où τ correspond au délai imposé et v^* représente le complexe conjugué de v . Cette fonction est obtenue en décalant, par exemple, le signal $v(t)$, puis, en multipliant ce signal décalé par $u(t)$, point par point. La somme de l'ensemble des produits est ensuite calculée. Pour chaque pas de décalage, on obtient alors une valeur d'inter-corrélation. Si la valeur est positive, les deux signaux sont en phase, dans le cas contraire, les signaux sont en opposition de phase. Enfin, si le résultat est égal à zéro, il n'y a aucune

3.1. ÉTAT DE L'ART DES MÉTHODES CLASSIQUES

ressemblance entre les signaux. L'inter-corrélation peut être normalisée (équation 3.5), permettant de comparer deux signaux qui n'ont pas les mêmes unités ou deux fonctions d'inter-corrélation.

$$\overline{C_{uv}(\tau)} = \left(\frac{C_{uv}(\tau)}{\sqrt{C_{uu}(0)C_{vv}(0)}} \right), \quad (3.5)$$

où, $C_{uu}(0)$ ($C_{vv}(0)$) correspondent aux auto-corrélations de $u(t)$ ($v(t)$) avec un décalage nul. Les valeurs d'inter-corrélation normalisées peuvent aller de -1 à 1, représentant des pourcentages de similitude. La Figure 3.3 montre deux exemples d'inter-corrélation normalisée entre deux signaux sismiques (Fig 3.3.A et 3.3.C) et entre un séisme et du bruit ambiant (Fig 3.3.B et 3.3.D). Nous observons des différences notables entre ces deux fonctions inter-corrélation. Pour les deux signaux sismiques, la fonction atteint un fort maximum (environ de 0,5) et une symétrie est également observée. Par contraste, la fonction d'inter-corrélation entre le séisme et le bruit ambiant ne dépasse pas 0,1.

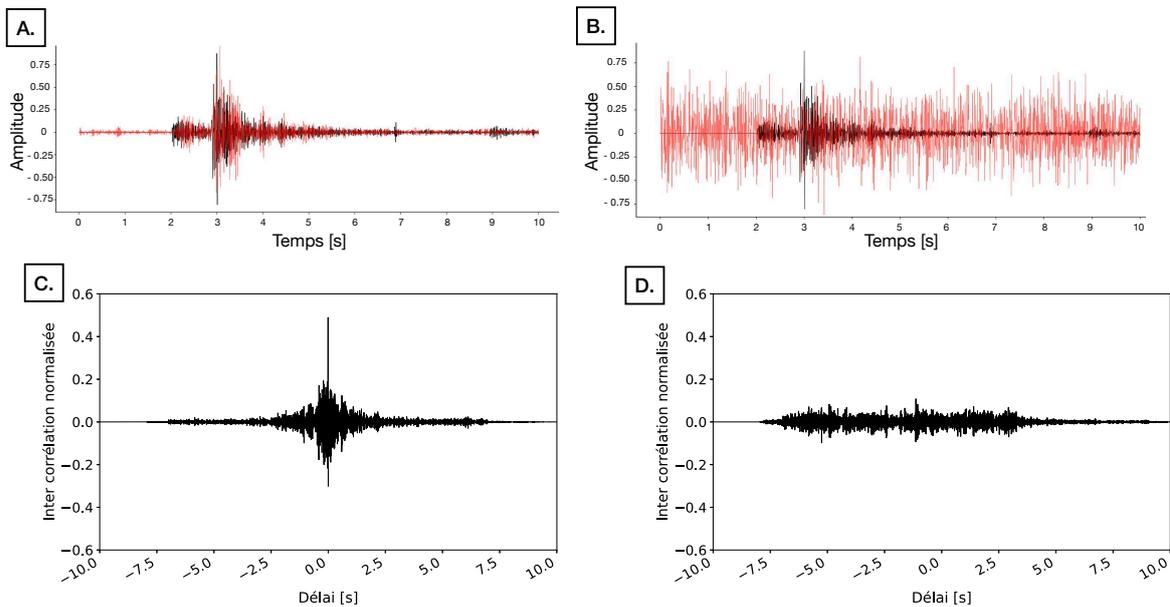


FIGURE 3.3 – Deux exemples d'inter-corrélations. (A) En noir, un sismogramme d'une forme d'onde de référence et, en rouge, un sismogramme d'un séisme présent dans le signal continu. (B) En noir, un sismogramme d'une forme d'onde de référence, en rouge, un sismogramme de bruit. (C) Inter-corrélation entre les deux signaux sismiques présentés en A). (D) Inter-corrélation entre les deux sismogrammes présentés en B).

Cette approche est largement appliquée pour, entre autre, la détection des petits séismes dans une région d'étude (par exemple Israelsson, 1990; Shearer, 1994; Eisner *et coll.*, 2008; Schaff et Waldhauser, 2010), de répliques (par exemple Peng et Zhao,

2009), ou encore pour la relocalisation d'événements sismiques (par exemple [Waldhauser et Schaff, 2008](#)).

Dans [Plenkers *et coll.* \(2013\)](#), la méthode STA/LTA est tout d'abord utilisée pour constituer une base de données d'événements de référence. Puis, pour chaque station, des formes d'onde de référence sont sélectionnées sur la base de 5 événements détectés par STA/LTA. Les formes d'ondes des cinq événements sont normalisées pour chaque composante et chaque station, puis sont sommées individuellement pour chaque composante. Enfin, une série d'inter-corrélations est réalisée afin de détecter les événements sismiques de plus faible magnitude.

Plusieurs approches utilisent la somme des fonctions d'inter-corrélation afin d'augmenter la valeur maximale d'inter-corrélation, par exemple avec [Shelly *et coll.* \(2007\)](#) pour la détection de tremor non-volcanique et d'essaims sismiques ou [Chamberlain *et coll.* \(2018\)](#) avec le développement de EQcarrscan pour la détection et l'analyse de la sismicité répétitive et quasi-répétitive. [Zhang et Wen \(2015\)](#) présentent Match and Locate (M & L), permettant la détection et la localisation par corrélation de segments temporels. Tout d'abord, des événements de référence sont sélectionnés et leurs localisations déterminées. Les localisations potentielles des petits événements sismiques sont représentées dans une région quadrillée (longitude, latitude et profondeur) centrée sur les emplacements du modèle. Les signaux de référence sont inter-corrélés avec le signal continu à chaque station et à chaque composante. À partir des différents temps de parcours des phases sismiques obtenus grâce aux emplacements potentiels, les fonctions d'inter-corrélation alignées sont ensuite sommées pour toutes les stations et toutes les composantes. Si le seuil d'inter-corrélation est dépassé, il s'agit d'une détection possible d'un petit événement et l'emplacement présentant la valeur maximale d'inter-corrélation est considéré comme sa localisation.

Enfin, l'algorithme Fingerprinting and Similarity Thresholding (FAST) ([Yoon *et coll.*, 2015](#); [Bergen et Beroza, 2018](#)) est une approche exploitant les formes d'ondes en les convertissant en une base de données de codes barres binaires. Plutôt que de comparer directement les formes d'ondes, des caractéristiques sont extraites du signal pour condenser chaque forme d'onde en un « code barre » compact. Cette approche permet de conserver uniquement les principales caractéristiques discriminantes du signal. Une empreinte sert d'indicateur unique pour une forme d'onde, ainsi, deux formes d'ondes similaires devraient avoir des empreintes numériques semblables.

3.1.3 Discrimination des événements sismiques

Suite à la détection des événements sismiques, il est essentiel de pouvoir discriminer les différents types de signaux détectés. Une grande diversité de sources peuvent être à l'origine des signaux enregistrés et détectés, allant des séismes naturels aux signaux d'origine anthropique (par exemple, tirs de carrière, explosions en mer), en passant par les tremors tectoniques ou les signaux volcano-tectoniques (par exemple [Rouland et coll., 2009](#)). De plus, les bruits impulsifs naturels ou anthropiques (par exemple, le trafic routier et le passage humain) à proximité des sismomètres peuvent déclencher accidentellement le système et conduire à de fausses alertes (par exemple [Böse et coll., 2014](#); [Wurman et coll., 2007](#)). Chacun de ces types de signaux possède des caractéristiques distinctes qui nécessitent une analyse précise pour une discrimination correcte. La discrimination des événements sismiques est une étape importante car elle permet de filtrer les données et de se concentrer sur les événements d'intérêt spécifique. Dans cette partie, nous explorerons les méthodes et les techniques utilisées pour discriminer avec précision les événements d'origine humaine de ceux d'origine naturelle.

La différence la plus frappante entre les événements naturels et anthropiques est le rapport d'amplitude entre les ondes P et S . Pour les événements naturels, ce rapport est généralement inférieur à 1,0, alors qu'il est presque égal à 1,0 pour les explosions. Cette caractéristique discriminante est largement utilisée afin de classer ces deux catégories de signaux (par exemple [Baumgardt et Young, 1990](#); [Dysart et Pulli, 1990](#); [Allmann et coll., 2008](#); [Pyle et Walter, 2019](#)). [Wang et coll. \(2020\)](#) montrent l'utilisation du rapport d'amplitude P/S grâce à un réseau sismique dense dans un arc volcanique continental près du Mont Saint Helens. Après avoir testé différentes bandes de fréquences, ils observent de meilleurs résultats de classification entre 10 et 18 Hz qu'à des bandes de fréquence plus basses ou plus étroites (Fig 3.4). En effet, l'amplitude des ondes S induites par les explosions diminue par rapport à celle des ondes P pour les fréquences plus élevées.

L'utilisation des ondes de Rayleigh (Rg) a également été expérimentée pour la discrimination (par exemple [O'Rourke et Baker, 2017](#)). Ces ondes sont couramment observées sur les sismogrammes d'événements sismiques d'origine anthropiques et de tremblements de terre tectoniques peu profonds enregistrés à des distances locales. Étant donné que les tirs de carrière se produisent à la surface, [Tibi et coll. \(2018\)](#) tire parti du fait que l'amplitude du mode fondamental des ondes Rg diminue considérablement avec l'augmentation de la profondeur de l'événement. À partir de cette

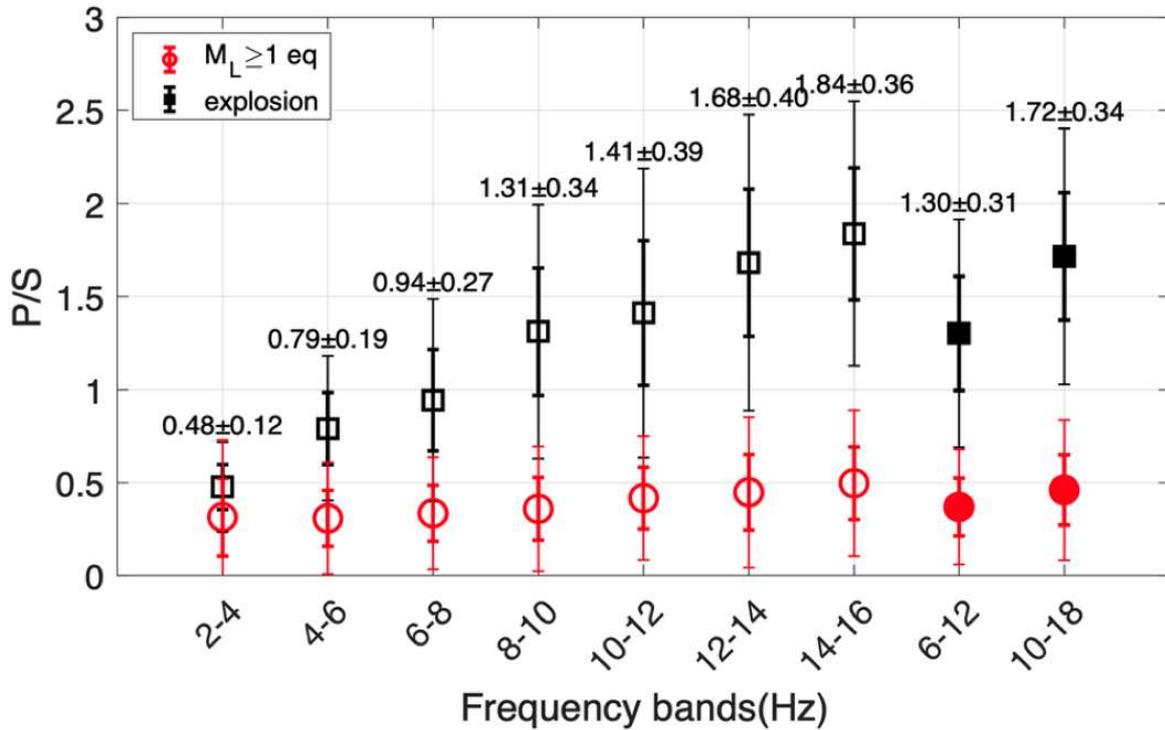


FIGURE 3.4 – Rapport P/S médian pour les explosions (en noir) et les tremblements de terre (en rouge) dans neuf intervalles de fréquence différents : sept intervalles étroits de 2 Hz identiques à ceux de [O'Rourke et Baker \(2017\)](#) et deux intervalles larges (symboles remplis). Les barres d'erreur en gras représentent le SMAD (Scaled Median Absolute Deviation), et les barres fines représentent 2 fois l'écart absolu médian. La médiane \pm SMAD pour les explosions sont étiquetées en haut (d'après [Wang et coll., 2020](#)).

caractéristique, ils proposent un nouveau discriminant de profondeur basé sur les rapports d'amplitude spectrale entre les ondes de Rayleigh et les ondes S . Cette approche a été utilisée pour distinguer les événements peu profonds des séismes tectoniques plus profonds dans la région de l'Utah. Cependant, cette méthode, qui s'appuie sur les caractéristiques des formes d'ondes contrôlées principalement par la profondeur, peut avoir des limites pour la classification des événements anthropiques dans les régions où les événements tectoniques sont peu profonds.

Malgré les performances des méthodes de rapport d'amplitude pour la discrimination, leur application peut être complexe lorsque les rapports dépendent de phases d'ondes sismiques non disponibles (par exemple, à de grandes distances), difficiles à isoler (à de très courtes distances) ou présentes dans des environnements fortement perturbés par du bruit. D'autres approches n'utilisant pas de phases d'ondes isolées ont par conséquent été développées. [Su et coll. \(1991\)](#) constatent que la décroissance de la coda est significativement plus élevée pour les tirs de carrières que pour les trem-

blements de terre pour les basses fréquences. Basé sur ces observations, [Koper *et coll.* \(2016\)](#) utilisent la différence entre la magnitude locale (M_l) et la magnitude de durée (basée sur la longueur de la coda) (M_d) pour discriminer les événements sismiques. Pour une sélection d'événements, la différence $M_l - M_d$ est en moyenne 0,44 unité de magnitude plus petite pour la sismicité provoquée par l'activité anthropique que pour la sismicité tectonique.

3.2 Apport du *machine learning* dans l'observation des événements sismiques

Durant la dernière décennie, l'utilisation du *machine learning* (ML) dans l'observation des événements sismiques offre des avantages significatifs en permettant l'exploitation efficace de grande quantité de données sismiques, améliorant ainsi notre capacité à détecter, classifier et caractériser ces événements d'une manière automatisée. Dans cette section, je vais tout d'abord introduire les notions d'intelligence artificielle et décrire plusieurs algorithmes couramment utilisés. Puis, je présente différentes méthodes de détection et de discrimination des événements sismiques, utilisant des approches de ML.

3.2.1 Introduction - Définitions

Le terme « intelligence artificielle » (IA) désigne un ensemble de techniques utilisées pour développer des machines capables de reproduire les capacités intellectuelles humaines. Deux sous-domaines clés de l'IA sont le *machine learning* (ML, apprentissage machine) et le *deep learning* (DL, apprentissage profond) (Fig 3.5).

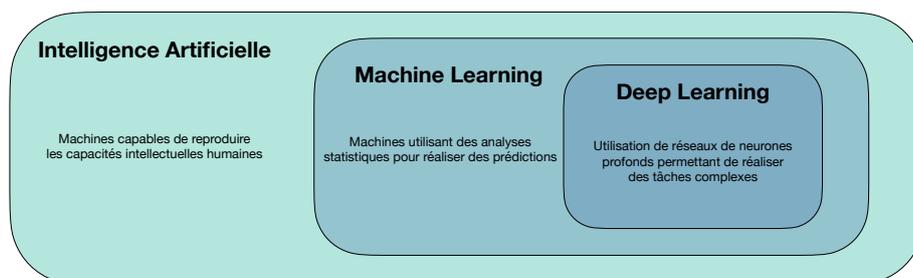


FIGURE 3.5 – Schéma explicatif des termes *Intelligence Artificielle* (IA), *machine learning* (ML) et *deep learning* (DL).

Le *machine learning* est une technologie d'intelligence artificielle permettant aux algorithmes d'apprendre à partir de données sans être explicitement programmées. À partir d'analyses statistiques sur les données, l'algorithme va extraire des motifs permettant d'effectuer des prédictions. La plupart des algorithmes de ML peuvent être regroupés en deux catégories principales : l'apprentissage supervisé et l'apprentissage non-supervisé (Fig 3.6). Si les données disponibles ne sont pas labellisées en fonction de leurs natures, l'apprentissage de la machine se fait de manière autonome et non supervisé. L'apprentissage non supervisé se subdivise en algorithmes de *clustering* et de réduction de dimension, selon que l'on souhaite regrouper les données en catégories sur la base de leur similarité ou simplement réduire les dimensions des données d'entrée. Dans le cas de l'apprentissage supervisé, l'algorithme est « guidé » grâce à des ensembles de données labellisées. Cette catégorie peut être subdivisée en algorithmes de classification, pour la prédiction de classes, et de régression, pour une prédiction quantitative.

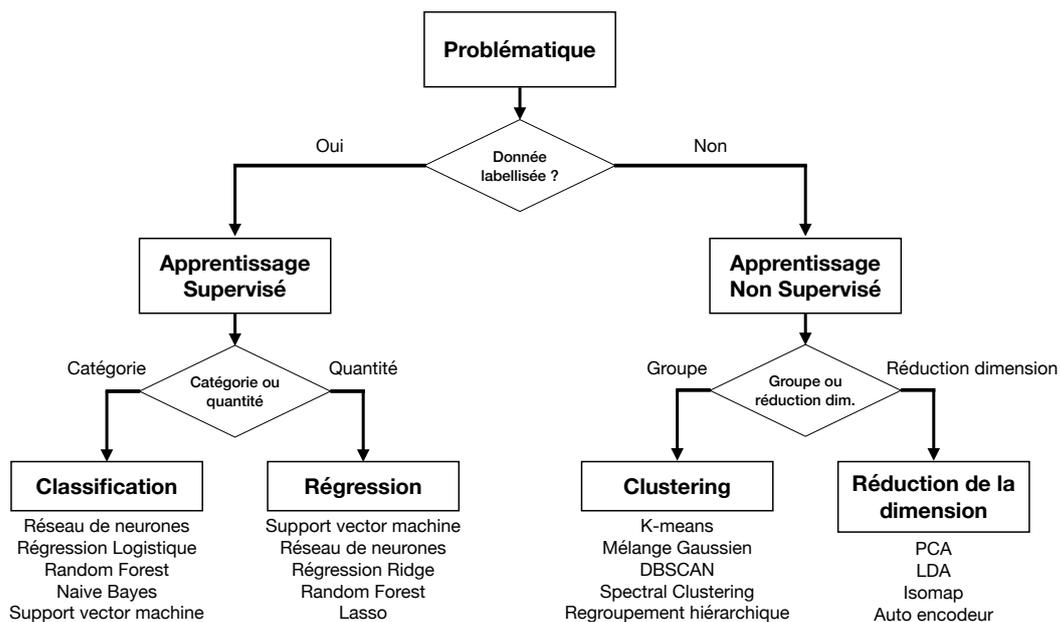


FIGURE 3.6 – Types d’algorithmes de *machine learning* (ML) ou apprentissage machine (modifié d’après Kong et coll. (2018)). L’apprentissage automatique supervisé fonctionne sur des ensembles de données labellisés et a pour objectif de développer des modèles qui prédisent soit une classe soit une valeur quantitative. L’apprentissage automatique non supervisé opère sur des ensembles de données non labellisés dans le but de regrouper les données par similarité ou de réduire la dimension d’ensembles de données d’entrée.

Le *deep learning* est une branche du *machine learning* qui s’inspire du fonctionnement du cerveau humain par l’utilisation de réseaux de neurones profonds (*deep neural network*). Le terme « profond » se réfère à la profondeur du réseau, c’est-à-dire

au nombre de couches qui le composent. Ces réseaux de neurones profonds sont donc composés de couches de neurones interconnectés, traitant les informations de manière hiérarchique. Pour cela, des caractéristiques de plus en plus abstraites sont extraites à mesure que les données traversent le réseau. Alors que le *machine learning* utilise généralement des modèles plus simples, le *deep learning* utilise des architectures de réseaux de neurones profonds, nécessitant davantage de puissance de calcul mais offrant des performances supérieures pour des tâches complexes, telles que la reconnaissance d'images.

Les réseaux de neurones sont les outils classiquement utilisés dans l'intelligence artificielle. Il s'agit d'un algorithme inspiré du fonctionnement du cerveau humain, conçu pour traiter des informations de manière similaire à un réseau de cellules nerveuses interconnectées. Au cœur de l'architecture des réseaux de neurones se trouve le perceptron (aussi appelé neurone), qui est l'élément de base d'un réseau de neurones. Un perceptron prend des entrées, effectue une somme pondérée de celles-ci et applique ensuite une fonction d'activation pour produire une sortie (Fig 3.7). Les fonctions d'activation introduisent une non-linéarité dans les sorties des neurones. Sans cette non-linéarité, le réseau serait essentiellement réduit à une combinaison linéaire de ses entrées, ce qui limiterait sa capacité à représenter des relations non-linéaires entre les variables d'entrée et de sortie. Il existe différentes fonctions d'activation couramment utilisées, telles que la fonction sigmoïde, la fonction ReLU (*Rectified Linear Unit*) et la fonction tangente hyperbolique. Les connexions entre les perceptrons sont pondérées par des valeurs appelées poids. Les poids déterminent l'importance relative des signaux d'entrée et régulent leur contribution à l'activation des perceptrons dans les couches suivantes. Ces poids sont ajustés pendant l'apprentissage du réseau afin d'optimiser ses performances et de minimiser les erreurs.

Un réseau de neurones classique se compose de plusieurs perceptrons organisés en couches, dont les principales sont la couche d'entrée, les couches cachées et la couche de sortie. La couche d'entrée est la première couche du réseau et elle reçoit les données d'entrée (*input*). Chaque neurone dans cette couche représente une caractéristique ou une variable d'entrée et transmet cette information aux neurones de la couche suivante. Les couches cachées, situées entre la couche d'entrée et la couche de sortie, sont responsables du traitement intermédiaire des données. Chaque couche cachée est composée de plusieurs perceptrons qui reçoivent les signaux des perceptrons de la couche précédente et les transmettent aux perceptrons de la couche suivante. C'est pourquoi nous parlons de *Fully connected neural network* (FCNN). Les informations sont propagées

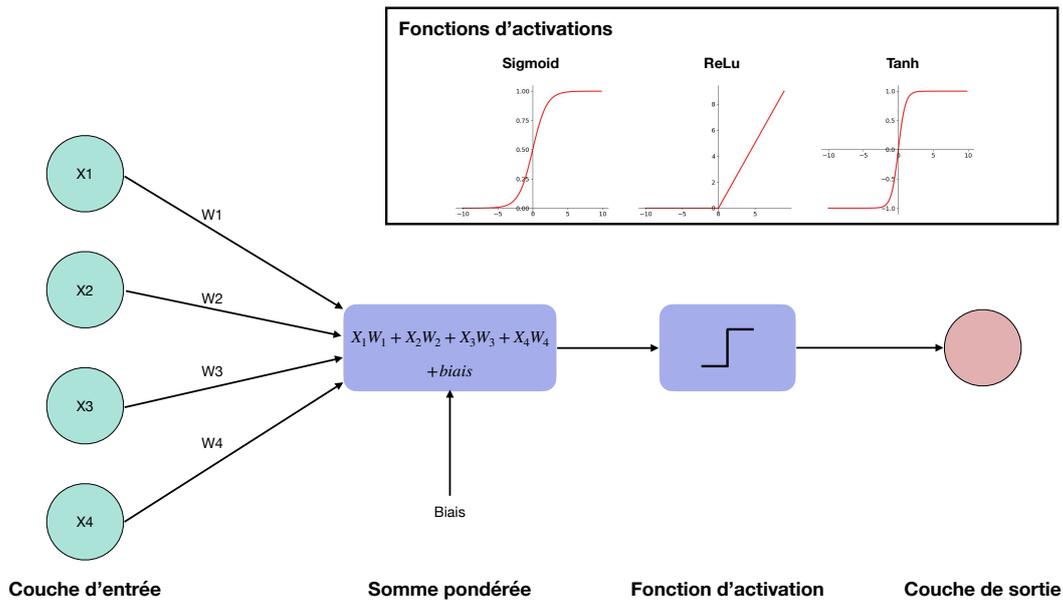


FIGURE 3.7 – Schéma explicatif d'un perceptron et des fonctions d'activation.

de manière itérative à travers ces couches cachées, permettant au réseau de capturer des motifs et des représentations complexes. Enfin, la couche de sortie produit les résultats finaux du réseau (*output*). Chaque perceptron dans cette couche représente une classe ou une valeur de sortie spécifique (si notre problème est soit une classification ou une régression). Les activations de ces perceptrons fournissent les prédictions ou les estimations générées par le réseau en fonction des données d'entrée.

L'entraînement d'un réseau de neurones à partir d'une base de données implique plusieurs étapes clés, notamment la création d'une base « d'entraînement », de « validation » et de « test ». Ces ensembles de données sont utilisés pour former, évaluer et tester les performances du réseau de neurones :

- La base « d'entraînement » (*Training set*) est l'ensemble de données utilisé pour entraîner le réseau de neurones. Les données d'entraînement doivent être représentatives de la tâche que le réseau de neurones doit accomplir. Par exemple, si le réseau doit effectuer une classification d'images, la base d'entraînement doit contenir un large variété d'images labellisées.
- La base de « validation » (*Validation set*) est utilisée pendant le processus d'entraînement pour évaluer les performances du réseau de neurones et ajuster les paramètres du modèle. Elle est utilisée pour mesurer la précision du réseau sur des exemples non vus pendant l'entraînement. La base de validation est essentielle pour détecter et prévenir le surapprentissage (*overfitting*) du réseau, c'est-à-dire

3.2. APPORT DU *MACHINE LEARNING* DANS L'OBSERVATION DES ÉVÉNEMENTS SISMQUES

lorsque le modèle mémorise les données d'entraînement au lieu de généraliser correctement sur de nouvelles données. La base de validation peut être utilisée pour ajuster les hyperparamètres du modèle, tels que le taux d'apprentissage ou le nombre d'itérations (*epochs*), afin d'optimiser les performances du réseau.

- La base de « test » (*Test set*) permet d'évaluer les performances finales du réseau de neurones une fois l'entraînement terminé. Elle doit être complètement séparée de la base « d'entraînement » et de la base de « validation » pour obtenir une évaluation impartiale. La base de test est utilisée pour mesurer l'exactitude et la performance du réseau sur des données réellement inconnues. Elle permet de donner une estimation réaliste des performances du réseau lorsqu'il est confronté à de nouvelles données.

L'architecture d'un réseau de neurones peut varier en fonction de la tâche à accomplir et des spécificités du problème. Par exemple, des architectures telles que les réseaux de neurones convolutionnels (CNN, *Convolutional Neural Network*) sont couramment utilisées pour le traitement d'images, tandis que les réseaux de neurones récurrents (RNN) sont adaptés au traitement de données séquentielles telles que le langage naturel.

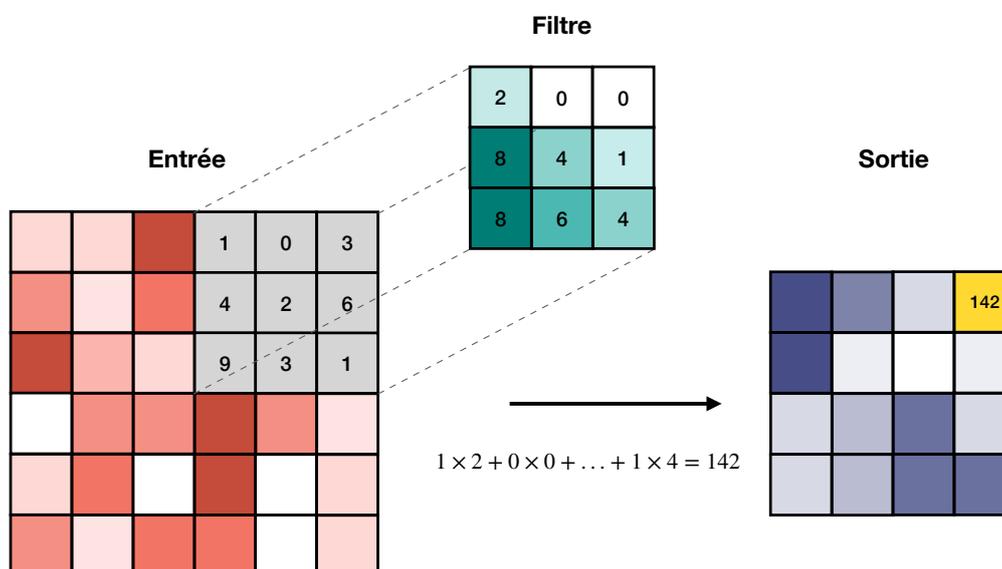


FIGURE 3.8 – Schéma explicatif d'une couche de convolution avec l'application d'un filtre de taille 3x3.

Un réseau de neurones convolutif (CNN) est conçu pour extraire automatiquement des caractéristiques pertinentes à partir d'images en utilisant des opérations de convolution ou de filtrage. Ces filtres sont des matrices de poids qui se déplacent sur l'image

et effectuent des opérations de convolution en multipliant les valeurs de pixels correspondantes, pour une image, et en les sommant. Cette opération permet de détecter des motifs et des caractéristiques spécifiques dans l'image, comme des contours, des textures ou des formes. Un exemple de convolution est montré dans la Figure 3.8 avec un filtre de taille 3×3 sur une image d'entrée de taille 6×6 . En déplaçant le filtre pixel par pixel, la taille de l'image de sortie est donc de 4×4 .

Un CNN est généralement composé de plusieurs couches. Les principales couches comprennent les couches de convolution, les couches de *pooling* et les couches entièrement connectées. Les couches de convolution correspondent donc à la série de filtres appliquée aux données d'entrée. Chaque couche de convolution comprend plusieurs filtres, de tailles différentes, qui sont optimisés par le réseau pendant la phase d'apprentissage. Les couches de *pooling* sont utilisées pour réduire la dimension des données en effectuant une opération d'échantillonnage sur les sorties des couches de convolution. Le *pooling* peut être réalisé en prenant la valeur maximale (*max pooling*) ou la moyenne (*average pooling*) des valeurs dans chaque région (Fig. 3.9).

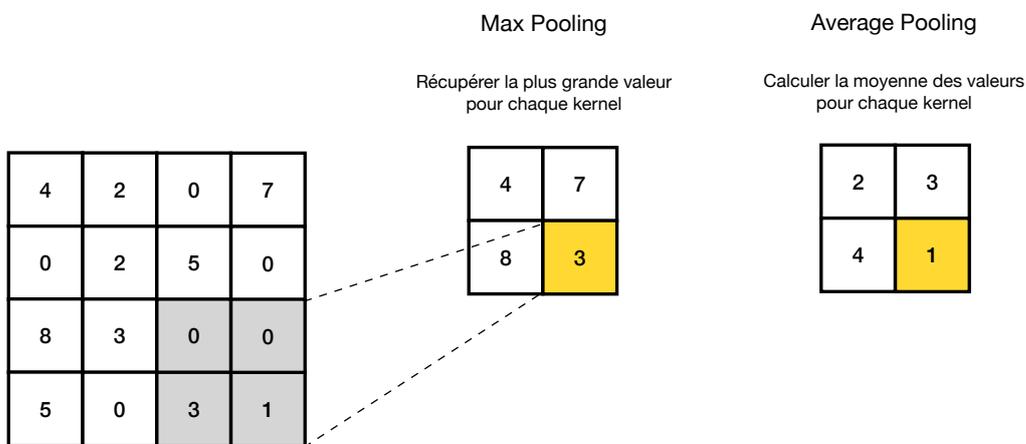


FIGURE 3.9 – Schéma explicatif du *max pooling* et de l'*average pooling*.

La Figure 3.10 illustre l'architecture d'un réseau de neurones convolutifs (CNN) utilisé, ici, pour classifier une image de chat. Lorsque l'image initiale passe à travers le réseau, elle est réduite progressivement en taille à chaque couche de convolution afin d'extraire les caractéristiques discriminantes telles que la forme des oreilles, la couleur de l'animal, et la forme des yeux. Les caractéristiques obtenues par les étapes de convolutions sont ensuite utilisées et intégrées dans un réseau de neurones entièrement connectés. Le rôle de cette partie du réseau est de combiner ces caractéristiques afin de classifier l'image et déterminer si elle représente un chat ou non.

3.2. APPORT DU *MACHINE LEARNING* DANS L'OBSERVATION DES ÉVÉNEMENTS SISMIQUES

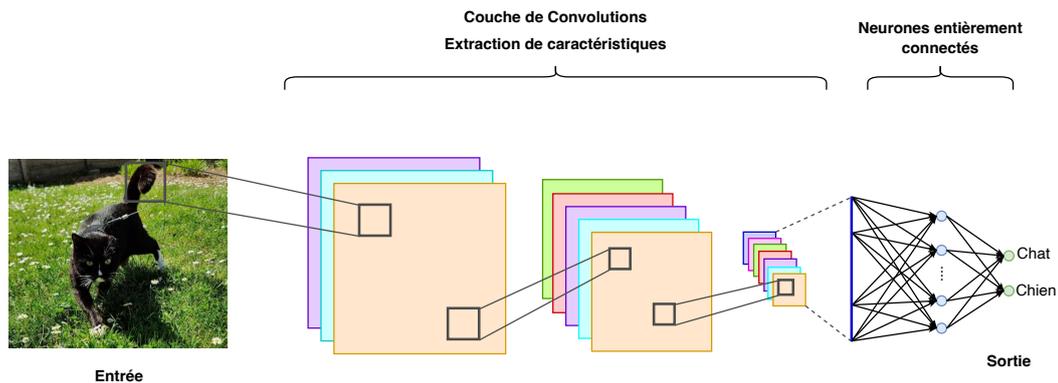


FIGURE 3.10 – Schéma explicatif de l'architecture d'un Réseau de Neurones Convolutif (CNN).

Un autre algorithme de ML couramment utilisé est le *Random Forest* (forêt aléatoire). C'est un algorithme qui combine plusieurs arbres de décision pour effectuer des prédictions. Dans un *Random Forest*, plusieurs arbres de décision sont construits en utilisant différents échantillons de données et différents sous-ensembles des caractéristiques déjà disponibles (Fig 3.11). Chaque arbre est construit de manière indépendante en sélectionnant aléatoirement les données d'entraînement et les caractéristiques à chaque étape. Chaque branche correspond à une condition sur une caractéristique spécifique. L'arbre est construit en divisant récursivement les données d'entrée en fonction de différentes caractéristiques et seuils, jusqu'à ce que des conditions d'arrêt soient atteintes. Lors de la prédiction, chaque arbre de décision du *Random Forest* produit une prédiction individuelle. La prédiction finale est ensuite obtenue par un vote majoritaire (pour la classification) ou une moyenne (pour la régression) des prédictions de tous les arbres.

Lorsque les caractéristiques ne sont pas disponibles pour entraîner le *Random Forest*, ce dernier peut être associé à un réseau GAN (*Generative Adversarial Network*). C'est un type d'algorithme qui se compose de deux parties distinctes : un générateur et un discriminateur. Le générateur a pour rôle de créer de nouvelles données synthétiques, telles que des images, des textes ou des sons, qui ressemblent à celles présentes dans un jeu de données d'entraînement donné. Il apprend à générer ces nouvelles données en utilisant des techniques d'apprentissage automatique, tel que les réseaux de neurones, pour capturer les motifs et les structures présents dans le jeu de données. Le discriminateur, quant à lui, agit comme un classificateur qui évalue si une donnée est réelle (provenant du jeu de données original) ou synthétique (produite par le générateur). Son objectif est d'apprendre à distinguer les données réelles des données

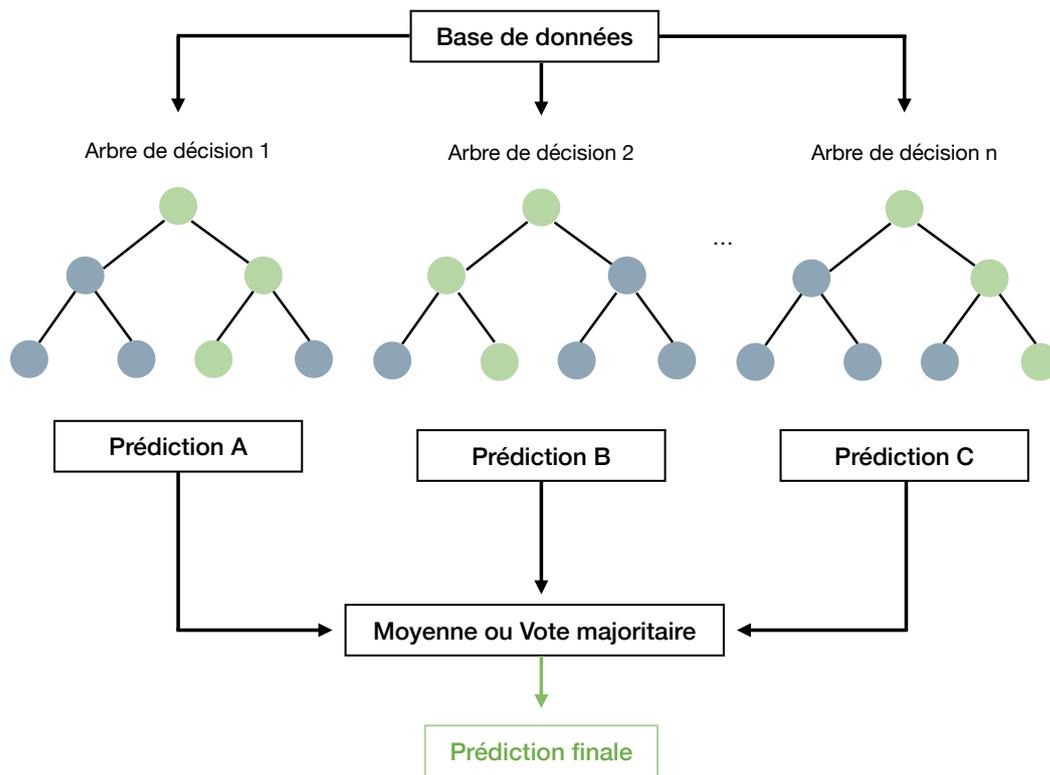


FIGURE 3.11 – Schéma explicatif d'un Random Forest.

synthétiques par l'extraction de caractéristiques. Le générateur et le discriminateur sont entraînés de manière concurrente et itérative. Le générateur tente d'améliorer sa capacité à tromper le discriminateur en générant des données synthétiques de plus en plus réalistes. Le discriminateur, de son côté, est entraîné pour devenir plus efficace dans la distinction entre les données réelles et synthétiques. Celui-ci peut également être utilisé pour extraire automatiquement les caractéristiques des données d'entrée. Ces caractéristiques peuvent ensuite être utilisées en entrée dans un réseau de neurones entièrement connectés ou un *Random Forest* par exemple.

La sismologie est une science riche en données, grâce à l'augmentation du nombre de stations sismologiques permanentes et à l'enregistrement en continu du mouvement du sol. La création et l'utilisation d'algorithmes de *machine learning* connaît une progression rapide au cours de ces 10 dernières années (Fig 3.12). Les premières applications étaient portées pour des problèmes de discrimination et de classification des événements sismiques (par exemple Pulli et Dysart, 1990; Fedorenko *et coll.*, 1999). Les recherches se sont ensuite étendues vers l'utilisation du ML pour améliorer la détection et l'identification des phases sismiques (par exemple Dai et MacBeth, 1995; Zhao et Takano, 1999). À l'heure actuelle, les algorithmes de ML sont entrés dans presque tous

3.2. APPORT DU *MACHINE LEARNING* DANS L'OBSERVATION DES ÉVÉNEMENTS SISMIQUES

les sous-domaines de la sismologie, montrant leur capacité à surpasser les approches classiques pour des tâches telles que la détection des séismes, la classification des événements sismiques ou encore le traitement et l'interprétation des images sismiques (par exemple [Li *et coll.*, 2019](#)). Dans la suite de cette section, nous nous focaliserons sur les différents algorithmes de ML permettant (i) la détection et le pointé automatique et (ii) la classification des événements sismiques.

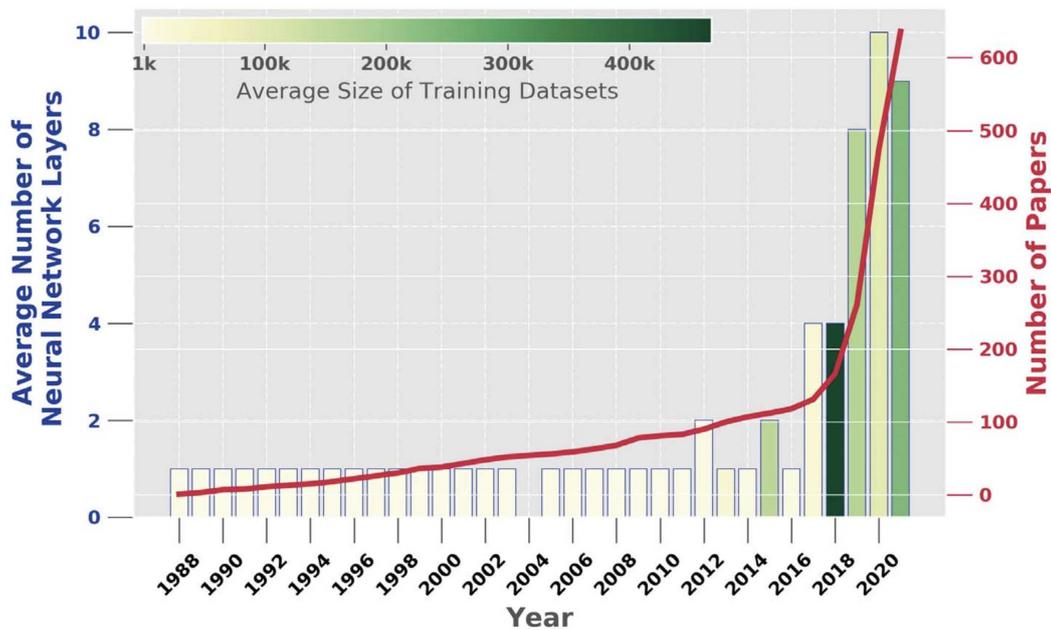


FIGURE 3.12 – Intérêt pour les approches d'apprentissage profond en sismologie. La courbe rouge représente le nombre cumulé d'articles publiés en sismologie entre janvier 1988 et janvier 2022 utilisant des réseaux neuronaux. L'histogramme indique le nombre moyen de couches cachées des réseaux neuronaux / la profondeur des réseaux. L'échelle de couleur représente la taille moyenne de la base de données d'entraînement utilisée pour entraîner ces réseaux ([Mousavi et Beroza, 2022](#)).

3.2.2 Détection et identification des phases sismiques

La reconnaissance des signaux provenant de tremblements de terre parmi le bruit de fond est une tâche difficile. L'utilisation des formes d'ondes comme base de données d'entraînement peut permettre une détection et l'identification rapide des phases sismiques. Dans [Wu *et coll.* \(2018\)](#), un réseau de neurone convolutif (CNN), appelé DeepDetect, est utilisé pour identifier le début et la fin des événements sismiques à partir de sismogrammes. Cet algorithme est conçu pour détecter des séismes de durées différentes. Pour cela ils utilisent une architecture en cascade afin de générer des propositions multi-échelles.

Perol *et coll.* (2018) présentent un algorithme permettant à la fois la détection et la localisation des événements sismiques, appelé ConvNetQuake, à partir de sismogrammes à 3 composantes. Pour cela, leur région d'intérêt a été divisée en six clusters géographiques. La Figure 3.13.A montre les six zones géographiques définies pour cette étude. Ainsi, la base de données d'entraînement est constituée de sept classes : la classe 0 pour le bruit sismique et les classes 1 à 6 pour chaque zone géographique. En pratique les formes d'ondes des événements sont données au réseau de neurones ConvNetQuake composé de huit couches de convolution puis d'une couche de neurones entièrement connectée (Fig. 3.13.B). La couche de sortie correspond aux probabilités de chacune des sept classes. Cet algorithme ne peut donc pas être appliqué à une autre zone géographique sans un nouvel entraînement. Sur la base de cette étude, Dokht *et coll.* (2019) ont apporté des modifications dans le but de pouvoir identifier des séismes dans une zone différente de celle d'entraînement. Pour cela, l'information sur la localisation de l'événement a été retirée et les informations temporelles et spectrales des sismogrammes sont utilisées en entrée afin d'améliorer les performances de détection.

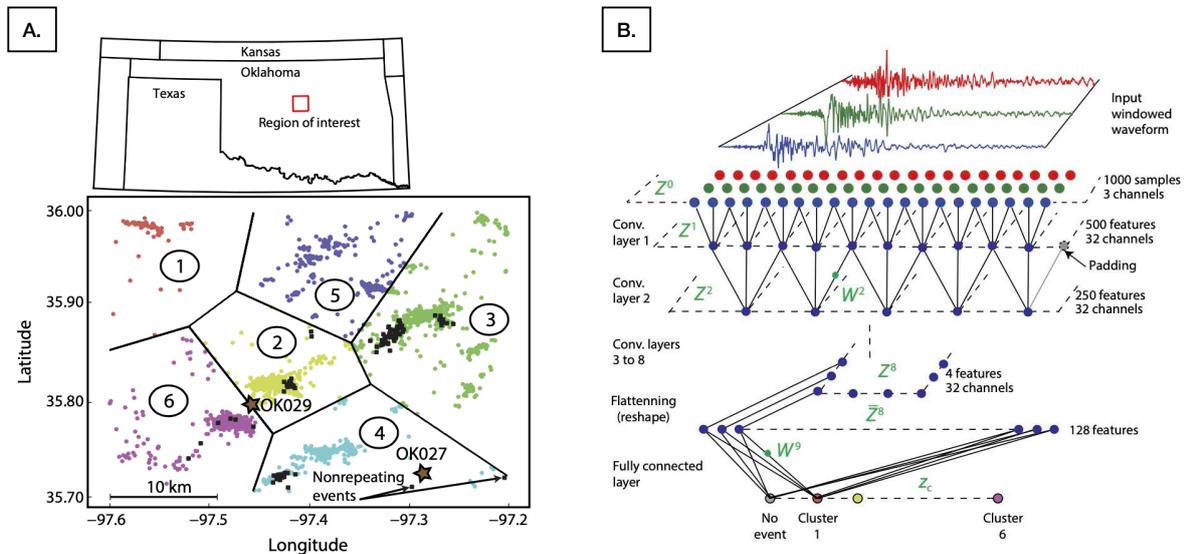


FIGURE 3.13 – (A) Localisation des événements sismiques de la base de données d'entraînement (cercles colorés) et des six clusters géographiques. Les séismes ont été détectés entre le 14 Février 2014 et le 16 Novembre 2016. GS.OK029 et GS.OK027 sont les deux stations qui enregistrent en continu la vitesse du mouvement du sol. Chaque événement est étiqueté avec la zone correspondante. Les lignes noires épaisses délimitent les six zones. Les carrés noirs représentent les événements de l'ensemble de données de test. Deux événements de l'ensemble de données de test sont surlignés parce qu'ils n'appartiennent pas aux mêmes séquences de tremblements de terre et ne se répètent pas. (B) Architecture de ConvNetQuake. L'input est une forme d'onde de 1000 échantillons sur trois composantes (modifié d'après Perol *et coll.* (2018)).

La vitesse de traitement par les algorithmes de ML est un aspect important pour la détection des tremblements de terre en temps réel et l'alerte précoce aux tremblements de terre (*Earthquake Early Warning*, EEW). L'objectif principal de l'EEW est d'avertir rapidement les populations de l'arrivée des ondes sismiques. Le fonctionnement de l'EEW repose sur la détection rapide et précise des premiers signaux sismiques, typiquement quelques secondes de formes d'ondes sismiques provenant d'un petit nombre de stations. Ainsi, à l'aide d'algorithmes d'apprentissage profond, il est possible d'identifier le début des tremblements de terre rapidement et d'éviter les alertes manquées.

Afin de montrer les performances des algorithmes de ML pour la discrimination signal/bruit en temps réel, [Meier et coll. \(2019\)](#) ont comparé plusieurs classificateurs et architectures de réseaux de neurones (FCNN, CNN, RNN et GAN + *Random Forest*) entraînés sur le même ensemble de données. En utilisant des segments de formes d'onde de 3 s, tous ces algorithmes offrent une grande performance de classification. Le CNN et le GAN associé au *Random Forest* atteignent tous deux 99,5% d'*accuracy* sur un ensemble de données de validation.

L'utilisation d'un réseau antagonistes génératifs (GAN) et de *Randoms Forest* a également été utilisé par [Li et coll. \(2018\)](#). Ils ont d'abord entraîné le GAN, à partir de fenêtres de 4 s, afin qu'il puisse reconnaître et produire des ondes *P* synthétiques réalistes. Ensuite, le discriminateur est utilisé comme extracteur automatique de caractéristiques de formes d'ondes. Ces caractéristiques sont enfin utilisées dans un classificateur *Random Forest* pour discriminer les ondes *P* du bruit sismique.

Différents algorithmes de ML permettant l'identification automatiques des phases sismiques ont également vu le jour ces dernières années. À partir de l'algorithme présenté dans [Ross et coll. \(2018\)](#), les temps d'arrivées des phases sismiques peuvent être identifiés. En effet, ce CNN entraîné à partir de 4,5 millions de signaux 3 composantes de 4 secondes, est capable d'identifier la classe dominante de la fenêtre, à savoir : onde *P*, onde *S* ou bruit. Parallèlement, [Zhu et coll. \(2019\)](#) présentent un *Phase Identifier Classifier* basé sur un CNN. La détection est réalisée en 2 étapes : (i) la détection des phases avec l'identification des fenêtres temporelles où des phases sismiques existent puis (ii) la sélection des phases en déterminant les temps d'arrivée grâce à un CNN plus précis. Malgré une base de données composées de peu d'éléments (30 146 phases labellisées), l'algorithme montre des hautes performances de détection avec une *accuracy* de 97,5%.

L'algorithme PhaseNet (Zhu et Beroza, 2019) est devenu une référence et est largement utilisé pour l'identification des temps d'arrivées des phases sismiques directes. Utilisant une architecture modifiée de Ronneberger *et coll.* (2015), ce réseau de neurones profond a été entraîné grâce à des formes d'ondes sélectionnées manuellement dans le nord de la Californie. La sortie du réseau correspond à la probabilité, point par point, de l'apparition des ondes P et S . Il démontre une grande performance de pointé et est largement utilisé par la communauté (par exemple Retailleau *et coll.*, 2022; Park *et coll.*, 2020; Zhou *et coll.*, 2021). La Figure 3.14 montre un exemple d'application de la méthode sur un séisme détecté dans le quart nord-ouest de la France. Pour cet exemple, les pointés manuels réalisés par le BCSF-RéNaSS ont été reportés (lignes vertes et rouges). Nous observons une très bonne cohérence avec les résultats de l'algorithme PhaseNet.

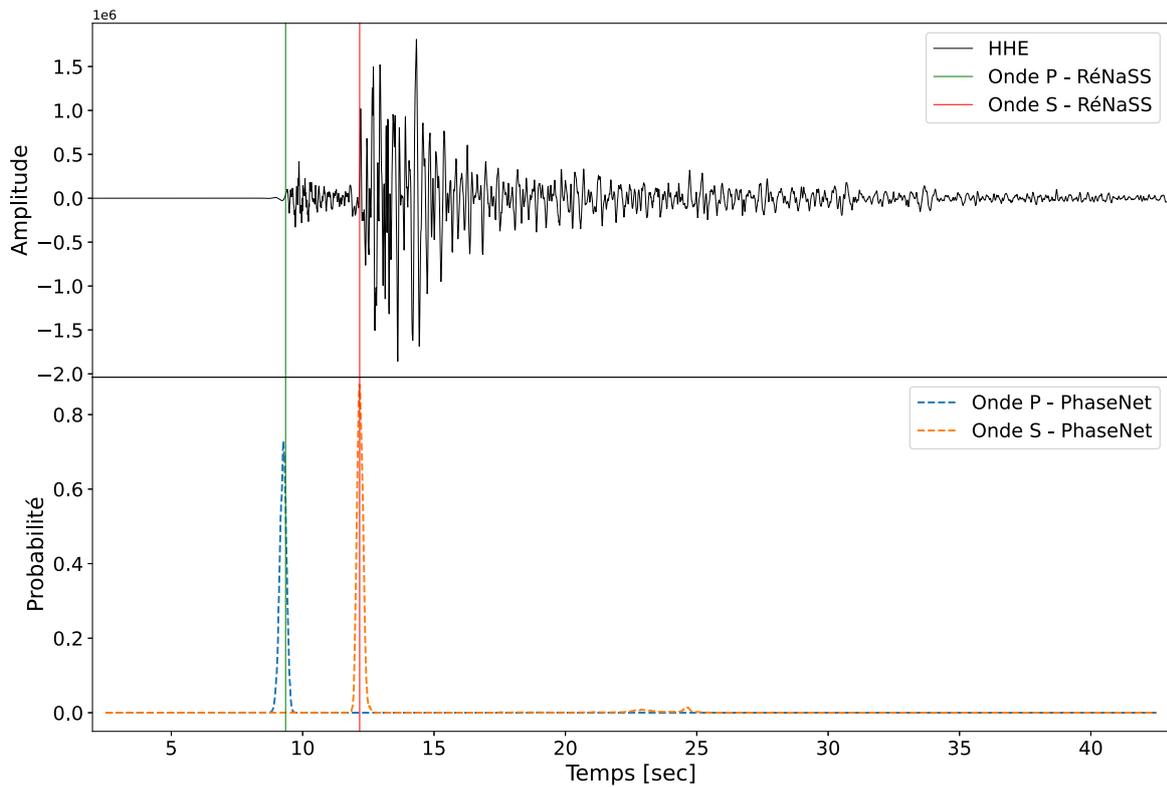


FIGURE 3.14 – Exemple d'application de PhaseNet sur le séisme de La Laigne du 16 Juin 2023 localisé à l'ouest de Niort et enregistré par la station CHIF. La courbe pointillée bleu correspond à l'arrivée de l'onde P et la courbe pointillée orange correspond à l'arrivée de l'onde S . Les lignes vertes et rouges correspondent aux pointés manuels réalisés par le BCSF-RéNaSS pour les ondes P et S respectivement.

Enfin, EQTransformer, développé par [Mousavi *et coll.* \(2020\)](#), permet à la fois la détection et le pointé des ondes P et S sur des séismes locaux (distance épacentrale < 300 km). Pour cela, un réseau de neurones, composé d'une série d'encodages suivie de 3 couches de décodage permet d'obtenir la probabilité d'un séisme, de l'onde P et de l'onde S . Celui-ci a été entraîné à l'aide de l'ensemble de données STanford EArthquake Dataset (STEAD) contenant les formes d'ondes d'environ 850 000 tremblements de terre détectés à l'échelle mondiale ([Mousavi *et coll.*, 2019](#)). Lors de l'application de ce modèle à l'ensemble de « test », un résidu inférieur à 0,5 s a été considéré comme correct. Ainsi, pour la détection de l'onde P , la précision était de 99,0% et pour la détection de l'onde S , de 99,0%.

3.2.3 Discrimination des événements sismiques

Comme présenté dans la section précédente, une grande diversité de sources sismiques peuvent être enregistrées et détectées. En France métropolitaine, nous faisons face à une hausse spectaculaire du nombre de détections de signaux anthropiques grâce à l'amélioration de la couverture sismique. En nous intéressant plus particulièrement à la détection des signaux de faible magnitude, il est donc indispensable de filtrer ces événements afin de se concentrer sur les événements d'intérêt spécifique. Grâce à l'augmentation sans précédent de la quantité de données sismiques disponibles, les caractéristiques propres de chacune de ces sources peuvent être exploitées afin de constituer des bases de données d'entraînement pour les algorithmes de ML. Ainsi, depuis ces dernières années, différentes approches de classification ont pu voir le jour avec, par exemple, [Nakano *et coll.* \(2019\)](#) pour la discrimination des trémors tectoniques, ou [Titos *et coll.* \(2020\)](#) avec les signaux volcano-tectonique. Dans cette partie, nous explorerons les méthodes et les techniques utilisées en ML pour discriminer avec précision les événements d'origine humaine de ceux d'origine naturelle.

[Renouard *et coll.* \(2021\)](#) présentent un *Random Forest* utilisant des caractéristiques de classification basées sur les connaissances d'experts humains. Celui-ci permet de discriminer les tremblements de terre des tirs de carrières dans la région du graben du Rhin Supérieur. Les caractéristiques utilisées comprennent des informations sur la localisation de la source, telles que la latitude, la longitude et la distance par rapport à la ville la plus proche. Des caractéristiques basées sur les signaux sont également prises en compte, notamment la variance des amplitudes dans la transformée de Fourier discrète. Ainsi, grâce à une combinaison de 22 attributs, cette approche hybride

obtient des scores de précision élevés, atteignant 95% pour les tirs de carrières et 96% pour les tremblements de terre. L'utilisation du *Random Forest* est particulièrement efficace pour une utilisation sur une zone spécifique, avec ici l'utilisation des stations temporaires AlpArray (Hetényi *et coll.*, 2018). En revanche, l'extension à d'autres zones géographiques peut s'avérer difficile à partir de ce type d'approche.

Parallèlement, l'utilisation de CNNs a également montré des performances de classification élevées pour la discrimination entre les événements naturels et anthropiques. En effet, Liu *et coll.* (2021) entraînent l'architecture proposée par Perol *et coll.* (2018) grâce une base de données constituée des formes d'ondes à trois composantes de 100 s de tirs de carrières et d'événements naturels. En appliquant ce CNN entraîné sur 8 mois de données, une *accuracy* de 92% est obtenue. Linville *et coll.* (2019) comparent deux types de réseaux de neurones avec l'utilisation d'un CNN et d'un RNN, appelé *Long-Short-Term-Memory* (LSTM), pour la discrimination. La base de données d'entraînement est composée de spectrogrammes de 90 s provenant de stations à trois composantes et à une composante. Ces spectrogrammes ont été calculés pour des fréquences comprises entre 1 et 20 Hz. Les deux types d'architectures obtiennent des scores élevés d'*accuracy* avec plus de 99% pour chacun.

3.3 Bilan des approches

Nous avons examiné différentes approches pour détecter et discriminer les événements sismiques, en présentant à la fois des méthodes classiques et des techniques basées sur l'apprentissage automatique (*machine learning*). Chacune de ces approches présente ses avantages et ses limites, mais certaines se sont avérées plus appropriées pour notre cas d'étude spécifique.

L'approche classique, utilisée par le BCSF-RéNaSS pour surveiller les événements sismiques à l'échelle nationale, est l'analyse STA/LTA (*Short-Term Average/Long-Term Average*). Cependant, cette méthode dépend de plusieurs paramètres, tels que la longueur des fenêtres STA et LTA, qui doivent être adaptées en fonction des caractéristiques des stations sismiques, ainsi que des seuils de détection. La configuration optimale de ces paramètres peut être complexe à déterminer pour notre cas d'étude spécifique. En revanche, les méthodes de type *template matching* ont montré de bonnes performances de détection à différentes échelles d'étude. Ces approches se basent sur la comparaison des formes d'onde sismiques avec des événements de référence préala-

3.3. BILAN DES APPROCHES

blement définis. En identifiant les similitudes entre les signaux enregistrés et les événements de référence, il est possible de détecter les événements sismiques avec une précision satisfaisante. Cette base de données de référence peut être facilement constituée grâce aux événements détectés et localisés par le BCSF-RéNaSS et/ou le CEA dans la région.

D'autre part, les approches basées sur le *machine learning*, telles que PhaseNet et EQTransformer, présentent l'avantage de ne pas nécessiter d'informations *a priori* pour la détection des événements sismiques. Ces méthodes utilisent des réseaux de neurones pour extraire automatiquement les caractéristiques des signaux et les classer en événement ou en bruit. Elles peuvent ainsi s'adapter à différents types de données sismiques sans nécessiter une configuration spécifique des paramètres.

Nous avons donc pris la décision de tester plusieurs approches. Les performances de détection et les contraintes techniques liées à la mise en place de chacune sont également estimées. Par la suite, la méthode ayant les meilleurs résultats de détection sera sélectionnée et adaptée à notre étude.

Chapitre 4

Nouvelle approche pour la détection et la discrimination d'événements sismiques de faible magnitude

Sommaire

4.1	Détection des signaux de faible amplitude	74
4.1.1	Quelle méthode de détection utiliser?	74
4.1.2	Quels paramètres utiliser pour optimiser la détection?	78
4.2	Discrimination des signaux sismiques des fausses détections	82
4.3	Discrimination entre les événements naturels et anthro- piques	90
4.3.1	New CNN-based tool to discriminate anthropogenic from na- tural low magnitude seismic events	91
4.3.2	Complément : Discrimination à l'échelle de la station et du réseau	119
4.4	Lecture des temps d'arrivée des ondes	121

4.1 Détection des signaux de faible amplitude

4.1.1 Quelle méthode de détection utiliser ?

Dans cette étude, notre objectif est d'arriver à identifier les événements sismiques naturels de faible magnitude. Comme expliqué précédemment, diverses méthodes de détection peuvent être utilisées. L'agence nationale BCSF-RéNaSS utilise la méthode STA/LTA pour détecter les séismes, qui sont ensuite inclus dans le bulletin. Nous allons nous baser sur les détections effectuées par le BCSF-RéNaSS dans le quart nord-ouest de la France afin de comparer différentes approches de détection. Étant donné le besoin de réduire la magnitude de complétude, les approches testées devront être capables de détecter ces séismes de magnitude modérée, détectés par le BCSF-RéNaSS, pour être retenues.

Application des approches sur un catalogue d'événements du BCSF-RéNaSS

Pour ce test, nous avons choisi la journée du 17 février 2021 durant laquelle dix événements sismiques ont été détectés par le BCSF-RéNaSS dont huit sont localisés à 5 km de Vannes (56). Les méthodes PhaseNet, EQTransformer (Section 3.2.2 pour plus d'informations sur ces méthodes) et de *template matching* sont testées. Pour les deux algorithmes de *machine learning*, les enregistrements continus provenant de 25 stations sont utilisés. Pour PhaseNet, une détection est définie lorsqu'une onde P est identifiée, avec une probabilité supérieure à 30% (probabilité utilisée par défaut), sur au moins trois stations. Pour EQTransformer, une détection est définie lorsque la probabilité de détection dépasse 30% à au moins trois stations. Pour le *template matching*, nous avons choisi 34 signaux de référence correspondant à des événements sismiques de $M_l > 2,5$ et détectés dans le nord-ouest de la France. Ces signaux de référence sont ensuite comparés grâce à une série d'inter-corrélations normalisées avec les enregistrements continus provenant de 25 stations permanentes. Le choix du seuil de détection est crucial, car un seuil trop élevé conduira à une détection insuffisante des séismes, tandis qu'un seuil trop bas entraînera un grand nombre de fausses détections. L'application à l'ensemble du quart nord-ouest de la France implique que les signaux de référence puissent présenter des variations significatives par rapport aux signaux sismiques présents dans le signal continu. La valeur seuil de 0,2 est donc fixée de manière empirique afin de détecter autant d'événements que possible sans être submergé par de fausses détections. Ainsi,

si la valeur maximale d'inter-corrélation dépasse 0,2 pour au moins trois stations sur une fenêtre d'étude, celle-ci est alors considérée comme une détection.

Comme indiqué sur la Figure 4.1, toutes les approches ont réussi à détecter les dix événements sismiques répertoriés dans le catalogue du BCSF-RéNaSS. Elles ont également détecté des événements sismiques qui ne figuraient pas dans ce catalogue. PhaseNet, par exemple, a permis de détecter 18 événements, dont 14 sont des événements sismiques. Les quatre autres événements correspondent à des fausses détections (signaux impulsifs). De même, EQTransformer a détecté 14 événements, dont 12 sont effectivement des événements sismiques. Enfin, le *template matching* a détecté 42 événements dont 16 sont des séismes.

Il est important de noter que l'utilisation du *template matching* conduit à une détection accrue d'événements non sismiques, correspondant à des signaux impulsifs parasites. En optant pour le *template matching*, cette détection importante d'artefacts nécessite le développement d'un outil capable de différencier automatiquement les séismes du bruit. Ce choix permet cependant d'obtenir la plus grande sensibilité et donc une meilleure détection.

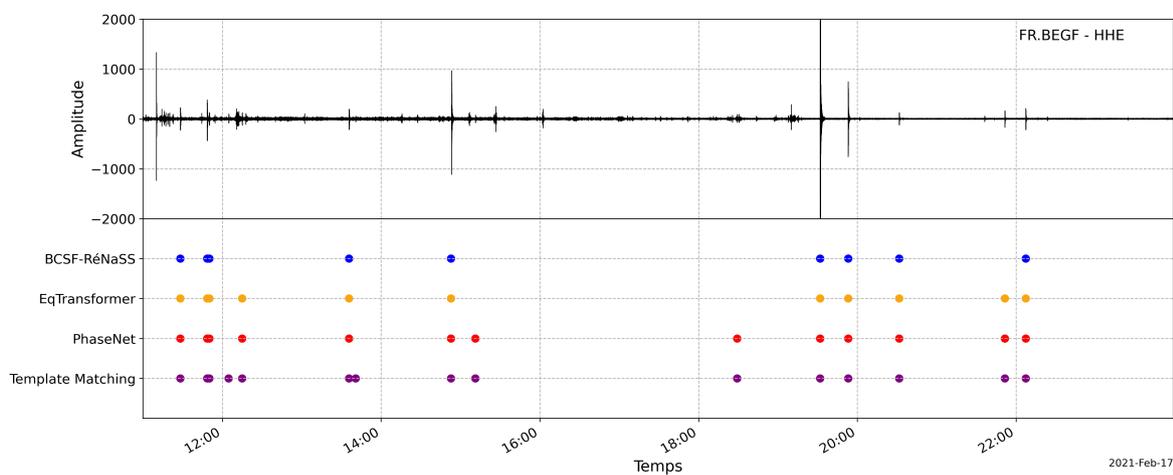


FIGURE 4.1 – Récapitulatif des détections de la journée du 17 février 2021 réalisées par le BCSF-RéNaSS, EQTransformer, PhaseNet et *template matching*. Le signal correspond à l'enregistrement de la journée par la station BEGF, du réseau FR, selon la composante Est-Ouest.

Application sur une plus longue période (2018 - 2020)

Au vu des résultats obtenus grâce au *template matching*, cette approche a été utilisée par [Beucler et coll. \(2021\)](#) afin de mettre en évidence d'éventuels événements sismiques non détectés dans le Massif armoricain. Pour cela, j'ai analysé deux années de données sismiques continues, du 1er août 2018 au 31 juillet 2020. La région d'intérêt est définie par le cercle gris en Figure 4.2. Durant cette période, 252 séismes sont présents dans le catalogue unifié BCSF + CEA dans l'ensemble de la zone et 192 sont situés dans le cercle gris (étoiles rouges avec centres blancs dans la Figure 4.2). Les magnitudes de moment (M_W) sont comprises entre 1,2 et 3,3. À partir de cette base de données, j'ai créé un sous-ensemble de signaux de référence en utilisant les formes d'onde de 40 événements. Pour la détection, seules sept stations, représentées par des triangles verts encadrés, sont utilisées par souci de temps de calcul et en raison du niveau de bruit enregistré pour certaines. Les autres triangles symbolisent des stations utilisées uniquement pour la localisation. Pour chaque composante de chaque station, une série d'inter-corrélation normalisée est réalisée entre tous les signaux de référence et le signal sismique continu sur deux années. Une valeur d'inter-corrélation supérieure à 0,2 à un minimum de trois stations est requise pour présélectionner les fenêtres temporelles qui contiennent potentiellement un événement sismique.

En utilisant ces paramètres, nous avons extrait 1 571 événements sismiques potentiels des données brutes. Après une inspection visuelle et une étape de pointé manuel, il a été déterminé que 318 d'entre eux étaient des tremblements de terre naturels présentant des phases d'ondes P et S bien définies. Presque tous les tremblements de terre signalés dans le catalogue sont également détectés à l'aide de cette approche. Cinq événements ne sont toutefois pas récupérés car ils ne satisfont le critère de détection qu'à deux stations. Un total de 131 nouveaux événements sismiques ont donc été détectés.

Les cercles cyan foncé présents dans la Figure 4.2.B représentent la localisation des séismes notables de ces dernières années. Nous pouvons observer, qu'à première vue, ces 131 séismes ne sont pas localisés au niveau de ces zones. La localisation de ces nouvelles détections semble donc être répartie de manière diffuse sur toute la région d'étude.

Ces 131 événements sismiques naturels ne sont présents dans aucun catalogue de séismes. Par conséquent, nous souhaiterions déterminer si des méthodes telles que PhaseNet et EQTransformer parviendraient à les détecter.

CHAPITRE 4. NOUVELLE APPROCHE POUR LA DÉTECTION ET LA DISCRIMINATION D'ÉVÉNEMENTS SISMIQUES DE FAIBLE MAGNITUDE

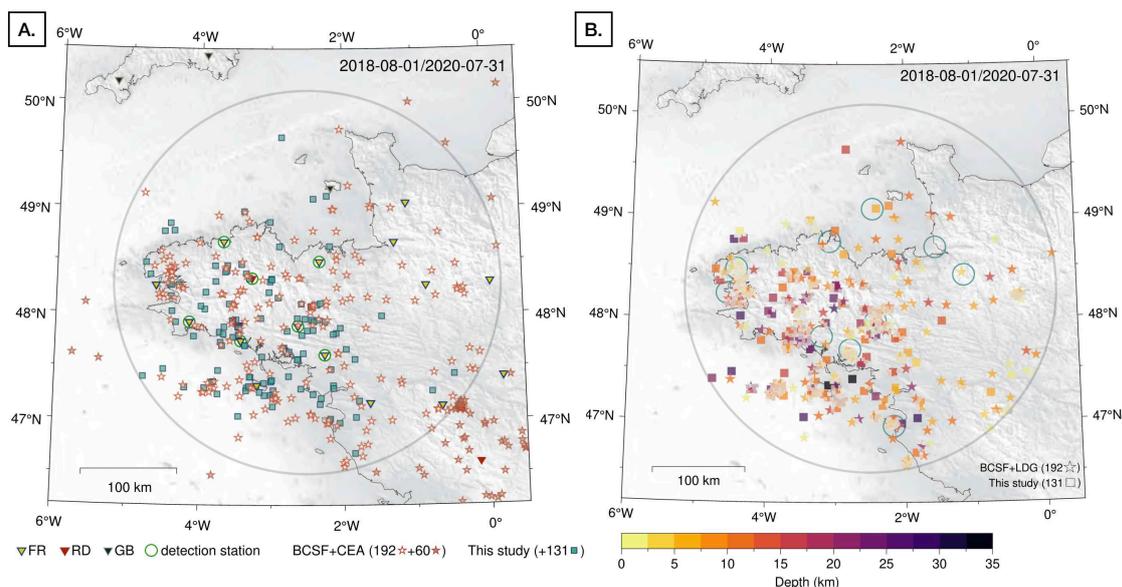


FIGURE 4.2 – Localisation et profondeur de l'ensemble des événements détectés du 1er août 2018 au 31 juillet 2020. La région d'intérêt est représentée par le cercle gris. 192 événements (étoiles) sont rapportés dans le catalogue unifié BCSF + CEA et 131 nouveaux événements (carrés) sont détectés et localisés dans cette étude. (A) Toutes les stations (triangles) sont utilisées pour la localisation. Seulement sept ont été utilisées pour la détection par *template matching*. (B) Représentation de la profondeur des événements. Les cercles en cyan foncé sont localisés au niveau des séismes notables de ces dernières années (Beucler et coll., 2021).

Pour cela, par souci de temps de calcul, nous avons appliqué ces approches seulement sur les formes d'ondes des 131 événements sismiques. À partir du pointé manuel précédemment réalisé, des fenêtres de 5 minutes d'enregistrement, pour chaque station où une onde P ou S a été identifiée, sont récupérées. L'ensemble de ces signaux a ensuite été utilisé comme données d'entrée pour les deux algorithmes, PhaseNet et EQTransformer. Les mêmes seuils de détection ont été utilisés qu'au test précédent. Comme le montre le Tableau 4.1, parmi les 131 séismes, EQTransformer détecte un total de 78 événements sismiques avec aucune fausse détection. PhaseNet quant à lui, détecte 109 événements sismiques.

TABLE 4.1 – Tableau comparatif des détections réalisées par PhaseNet, EQTransformer et *template matching* sur les 131 séismes détectés dans Beucler et coll. (2021).

	PhaseNet	EQTransformer	<i>template matching</i>
Nombre d'événements détectés	109	78	131

4.1.2 Quels paramètres utiliser pour optimiser la détection ?

Nous avons choisi d'adopter le *template matching* en tant que méthode de détection. Cette approche nous permet de bénéficier de l'expertise humaine dans la sélection de signaux de référence et de pouvoir ajuster le seuil de détection selon les besoins. Pour appliquer cette approche, plusieurs paramètres doivent être optimisés.

Dans un premier temps, le seuil de détection joue un rôle crucial dans cette méthode. Généralement, les approches basées sur le *template matching* sont employées pour détecter des répliques ou des essais sismiques (Eisner *et coll.*, 2008; Peng et Zhao, 2009; Schaff et Waldhauser, 2010). Dans de tels cas, les signaux sismiques émanent d'une zone d'étude restreinte et présentent des caractéristiques similaires en termes d'amplitude et de différence de temps entre les arrivées des ondes P et S . Un seuil de détection (ou de similarité) élevé peut être utilisé dans ces situations.

Néanmoins, dans le cadre de notre étude, nous visons à appliquer cette approche à l'ensemble du quart nord-ouest de la France. Cela implique que les stations sismiques sont distantes de plus de 50 km les unes des autres, et les signaux enregistrés peuvent présenter des variations significatives en fonction de la source de l'événement. Le choix du seuil de détection doit tenir compte de cette configuration.

Les Figures 4.3.A et 4.3.B illustrent un exemple d'inter-corrélation normalisée entre deux formes d'ondes provenant d'événements sismiques distincts. Ces événements sont détectés dans des zones géographiques différentes. Le signal noir correspond à un séisme détecté à 42 km au sud de Pau, dans les Pyrénées. Le signal rouge correspond à un séisme détecté à 24 km au nord de Cholet dans le nord-ouest de la France. Bien que les distances épicentrales soient similaires (environ 15 km), les fonctions sources diffèrent. La Figure 4.3.B montre la fonction d'inter-corrélation normalisée. En utilisant directement les formes d'ondes, la valeur maximale d'inter-corrélation est faible, de 0,22 pour cet exemple. Comme présenté dans la Figure 4.4.A, l'inter-corrélation entre deux formes d'ondes induit une gamme de seuils d'inter-corrélation limitée, variant de 0,1 à 0,3.

Dans cette étude, notre objectif est de détecter des formes d'ondes présentant des similitudes avec celles des événements sismiques, sans tenir compte de la fonction source spécifique de chaque événement. En d'autres termes, nous cherchons à repérer une forme d'onde qui montre l'arrivée d'une onde P , suivie d'une onde S très impulsive, puis la décroissance de la coda de l'onde S . Par conséquent, nous avons voulu explorer

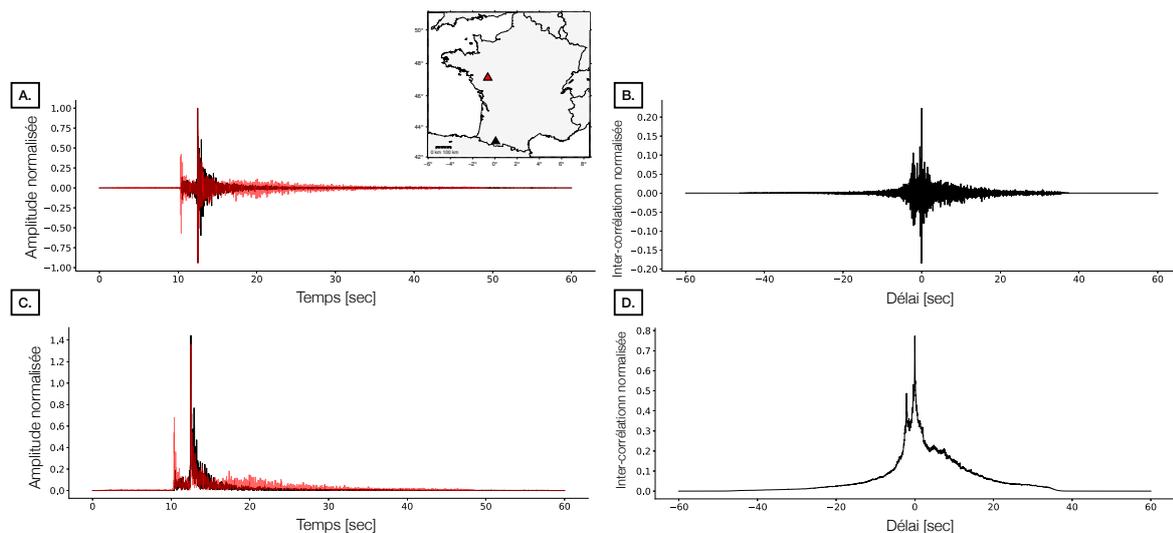


FIGURE 4.3 – Exemples d’inter-corrélations normalisées à partir d’événements sismiques. (A) En noir, un sismogramme d’une forme d’onde de référence d’un séisme détecté le 3 janvier 2022 et enregistré par la station LABF et, en rouge, un sismogramme d’un séisme présent dans le signal continu du 31 décembre 2021 enregistré par CRNF. (B) Inter-corrélation entre les deux signaux sismiques présentés en (A). (C) En noir, l’enveloppe de la forme d’onde de référence du 3 janvier 2022, en rouge, l’enveloppe du signal du séisme du 31 décembre 2021. (D) Inter-corrélation normalisée entre les deux enveloppes présentées en (C).

les possibilités d’augmenter les valeurs d’inter-corrélation. Dans cette optique, l’utilisation de l’enveloppe des formes d’ondes s’avère prometteuse. L’enveloppe du signal est calculée à partir de l’équation 4.1 où x correspond au signal sismique et $H(x)$ à sa transformée de Hilbert (Kanasewich, 1981). En considérant l’enveloppe, nous pouvons préserver les informations relatives à l’arrivée des ondes sismiques.

$$E = \sqrt{x^2 + H(x)^2} \quad (4.1)$$

La Figure 4.3.C montre les enveloppes des deux formes d’ondes sismiques précédemment présentées. La Figure 4.3.D montre la fonction d’inter-corrélation obtenue à partir des enveloppes. Nous pouvons observer que la valeur maximale d’inter-corrélation atteint 0,78. En utilisant l’enveloppe du signal, nous conservons l’information sur l’arrivée des ondes sismiques tout en augmentant les valeurs d’inter-corrélation.

Afin d’illustrer les différentes gammes de valeurs d’inter-corrélations, nous avons réalisé une série d’inter-corrélations entre des événements sismiques de référence et plusieurs types de signaux : (1) bruit sismique ambiant, (2) signaux impulsifs d’origine non sismique (représentant des fausses détections), (3) séismes non détectés par les

4.1. DÉTECTION DES SIGNAUX DE FAIBLE AMPLITUDE

méthodes classiques et afin (4) séismes de référence. Nous avons d’abord effectué ces calculs en utilisant directement les formes d’ondes (Fig 4.4.A), puis en utilisant les enveloppes des signaux (Fig 4.4.B). Nous observons non seulement de plus grandes valeurs d’inter-corrélations mais également une plus grande variabilité de valeurs lorsque nous utilisons les enveloppes plutôt que les formes d’ondes. Par conséquent, la détermination d’un seuil de détection est facilitée.

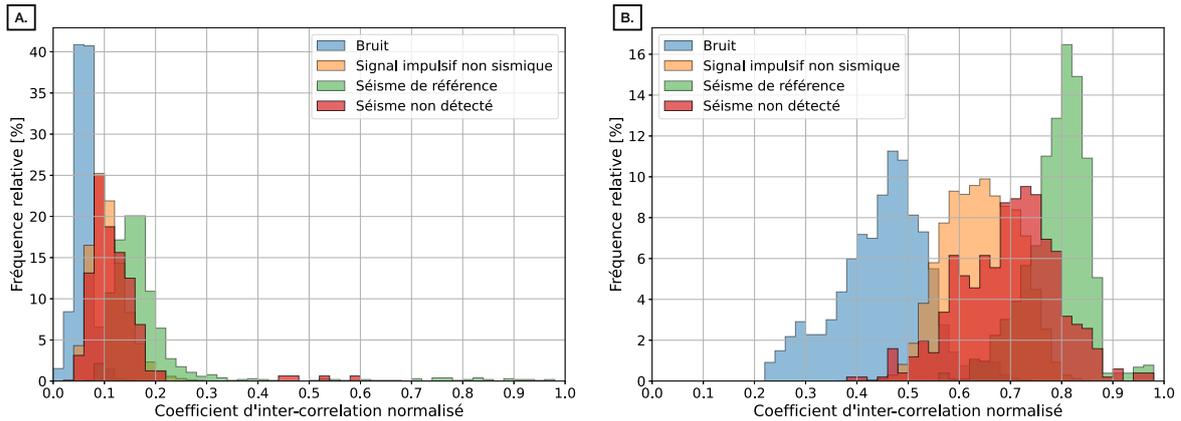


FIGURE 4.4 – *Distribution des valeurs maximales de coefficients d’inter-corrélations normalisées obtenues pour quatre jeux de données. (A) Valeurs maximales d’inter-corrélations obtenues à partir des formes d’ondes. (B) Valeurs maximales d’inter-corrélations obtenues à partir des enveloppes des formes d’ondes.*

Un autre paramètre essentiel concerne le traitement des signaux de référence. Afin d’améliorer le rapport signal sur bruit et d’optimiser la valeur d’inter-corrélation, divers traitements sont appliqués aux signaux de référence. Tout d’abord, le choix de la durée des signaux est déterminé en fonction de la région d’étude. Nous avons décidé d’utiliser des signaux d’une durée de 90 s afin de couvrir toutes les distances épacentrales de la région (environ 600 km), bien que la recherche des signaux de faible amplitude pourraient permettre de réduire cette durée.

Afin d’améliorer le rapport signal sur bruit, le traitement des sismogrammes est indispensable (Fig 4.5). Dans un premier temps, la moyenne de l’amplitude du signal est retirée. Puis, une apodisation est appliquée sur 5% des bords des fenêtres, afin de limiter les effets de bords lors de la procédure de filtrage. Pour isoler, dans le sismogramme, la gamme de fréquences caractéristique des séismes locaux, un filtre passe-haut est appliqué à 2 Hz. Ceci nous permet de nous affranchir de l’influence du bruit océanique, d’amplitude maximale généralement entre 0,05 et 0,5 Hz (par exemple [Beucler *et coll.*, 2015](#)). De plus, les ondes de volume des séismes de magnitude $M_l < 4$, ont généralement des fréquences coins supérieures à 1 Hz (par exemple [Lee](#)

CHAPITRE 4. NOUVELLE APPROCHE POUR LA DÉTECTION ET LA DISCRIMINATION D'ÉVÉNEMENTS SISMIQUES DE FAIBLE MAGNITUDE

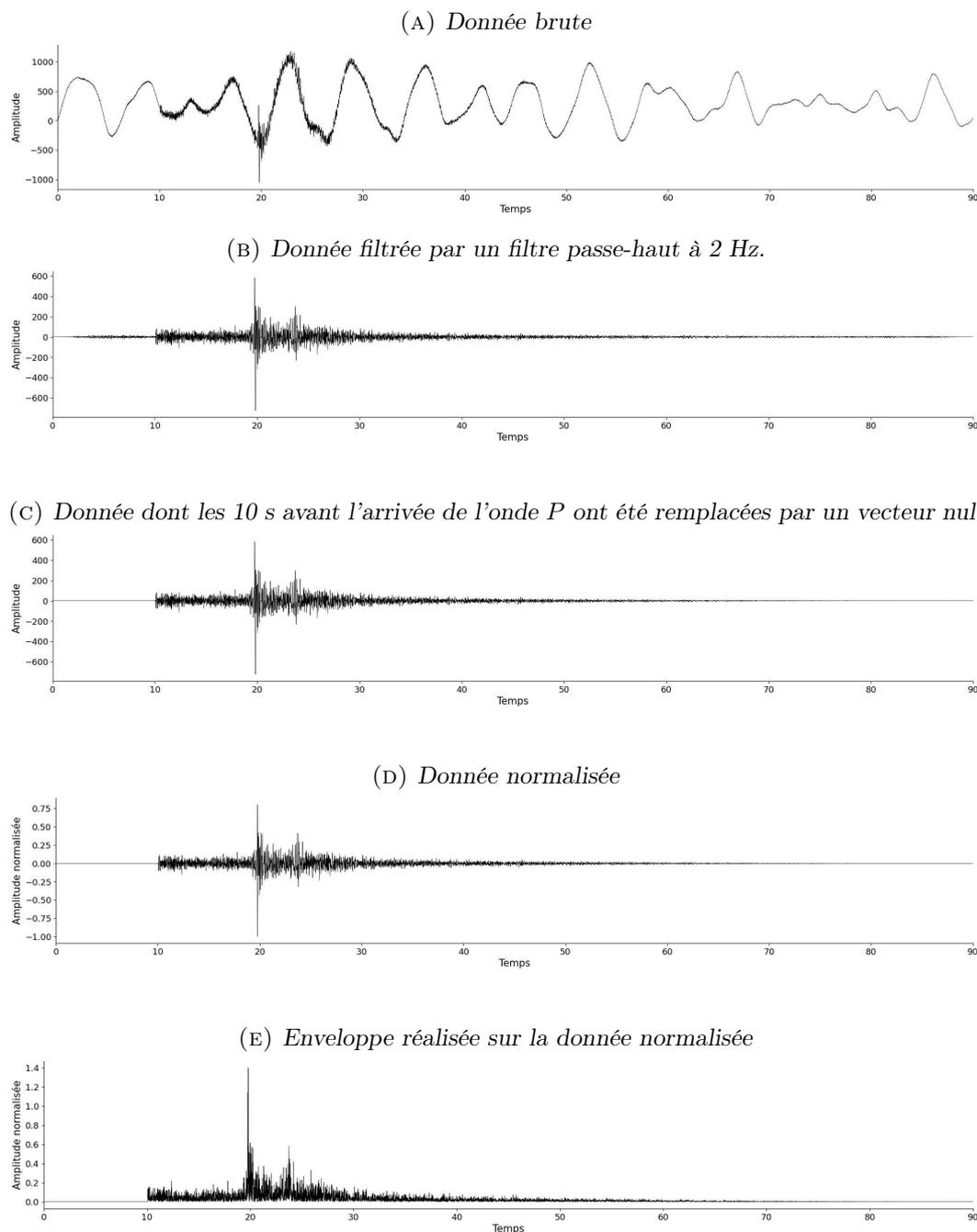


FIGURE 4.5 – *Traitement des données sur un sismogramme : exemple du séisme du 11 Février 2022 à 15h22, heure locale, au sud de Saint-Malo ($M_l = 1.9$) enregistré par la station MTNF dont la distance hypocentrale est de 83 km.*

et coll., 1981). Afin de limiter l'influence du bruit ambiant sur les résultats de l'inter-corrélation, les 10 secondes, avant l'arrivée de l'onde P , sont remplacées par un vecteur nul. Tous les sismogrammes de référence sont ensuite normalisés par leur maximum. Enfin, l'enveloppe de chaque signal est calculée.

La corrélation de segments temporels permet la détection d'événements pouvant être considérés comme sismiques. Les différents traitements décrits dans cette section nous permettent d'optimiser les valeurs d'inter-corrélation. Cependant, la recherche de signaux similaires de très faible énergie peut engendrer un taux de faux événements importants, directement corrélés avec le seuil d'inter-corrélation. Les événements détectés peuvent donc être de différents types : (1) signaux sismiques ou (3) fausses détections représentées par des signaux impulsifs ou des artefacts. Il est essentiel de trouver un approche permettant de discriminer automatiquement ces différents types de signaux.

4.2 Discrimination des signaux sismiques des fausses détections

L'utilisation du *template matching* dans notre région d'étude s'accompagne obligatoirement de la discrimination de signaux impulsifs non sismiques correspondant à des fausses détections. Dans la littérature, il existe peu d'approches permettant la discrimination automatique des fausses détections, la plupart basées sur des approches d'intelligence artificielle. Classiquement, cet exercice est réalisé manuellement par inspection visuelle des détections.

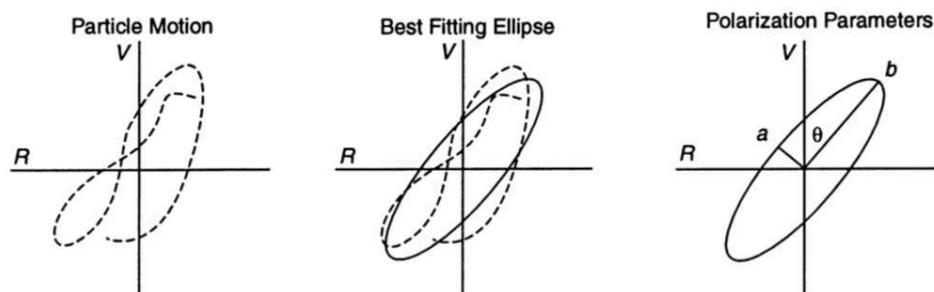


FIGURE 4.6 – Paramètres de la polarisation (modifié d'après Earle, 1999). Les pointillés représentent le mouvement de particule de l'onde P selon le plan radial et vertical. Le mouvement de particule est ensuite assimilé à une ellipse. La direction a correspond à l'axe mineur et la direction b à l'axe majeur. L'angle θ indique l'angle de polarisation.

Nous avons tout d'abord essayé d'analyser la linéarité du mouvement de particule pour mettre en évidence l'arrivée d'une onde sismique dans le signal détecté (par

exemple Schimmel *et coll.*, 2011). Le mouvement du sol, lors de l'arrivée d'une onde sismique, peut être assimilé à une ellipse (par exemple Jurkevics, 1988). Cette ellipse est définie par trois axes, la direction du plus grand axe b est considérée comme la direction de polarisation (Fig 4.6).

La linéarité est

$$\rho = 1 - \left(\frac{a}{b}\right)^{1/2}, \quad (4.2)$$

où a est l'axe mineur de l'ellipse et b , l'axe majeur. Cette linéarité varie de 0, pour une forme circulaire sans direction préférentielle, à 1 pour une énergie polarisée linéaire, correspondant donc à une droite. La direction de polarisation des ondes P étant parallèle à la direction de propagation des ondes sismiques, lors de l'arrivée dans le signal d'une onde P , la linéarité sera alors égale ou proche de 1.

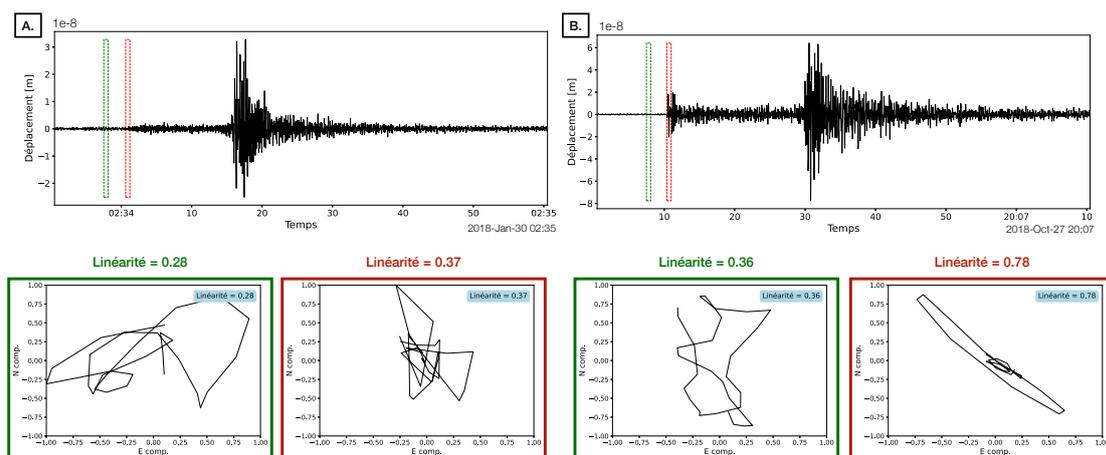


FIGURE 4.7 – Exemples de mouvements de particule de l'onde P calculés à partir des signaux associés (A. et B.). (A) Forme d'onde du séisme datant du 30 Janvier 2018 à 2h36 (UTC) enregistré par la station CHIF. (B) Forme d'onde du séisme datant du 27 octobre 2018 à 20h05 (UTC) enregistré par la station LOCF. Les rectangles rouges correspondent aux fenêtres où le mouvement du particule, au niveau de l'arrivée de l'onde P , a été calculé. Les rectangles verts correspondent aux fenêtres où le mouvement du particule, au niveau du bruit sismique ambiant, a été calculé.

En pratique, j'ai utilisé une fenêtre glissante parcourant le signal et une approche *bootstrap*. Le calcul a été effectué en échantillonnant de manière aléatoire un certain nombre de fois le signal dans la fenêtre d'analyse. La Figure 4.7 montre deux exemples de mouvements de particule calculés à partir de l'arrivée d'une onde P et du bruit ambiant. Nous pouvons observer qu'avec un bon rapport signal sur bruit (Fig 4.7.B), le mouvement de particule calculé à partir de l'onde P forme une ellipse et ainsi la

linéarité est proche de 1. Un rapport signal sur bruit plus faible (Fig 4.7.A) conduit à de faibles valeurs de linéarité qui ne permettent pas de discriminer efficacement les événements sismiques du bruit.

Cette approche ne sera donc pas utilisée pour identifier efficacement les événements sismiques du bruit. En nous basant sur la littérature, l'utilisation de l'apprentissage automatique semblent pouvoir apporter de très bons résultats pour la discrimination des signaux. Nous avons donc choisi de développer un algorithme de *deep learning* pour discriminer les séismes du bruit.

La Figure 4.8 présente quatre exemples d'événements sismiques labellisés par l'agence nationale française BCSF-RéNaSS et quatre signaux impulsifs (traces et spectrogrammes). La discrimination entre ces événements peut être difficile si l'on se base seulement sur les sismogrammes (panneaux de gauche de la Figure 4.8). Pour les événements sismiques (Fig 4.8.A, C, D et G), nous observons une certaine cohérence dans les signaux avec deux arrivées d'ondes puis la décroissance de la coda. En revanche, cette cohérence n'est pas visible dans les signaux impulsifs. Les spectrogrammes (panneaux de droite de la Figure 4.8) des événements sismiques permettent d'identifier clairement deux arrivées d'ondes, avec un fort paquet d'énergie associé à l'arrivée de l'onde *S* impactant, majoritairement, une grande gamme de fréquence. Ce comportement n'est pas observé dans les signaux parasites impulsifs pour lesquels on peut principalement voir des paquets d'énergie apparaissant dans des gammes de fréquences spécifiques à chaque signal.

Les spectrogrammes semblent donc contenir une information « visuelle » suffisante pour nous aider à différencier les événements sismiques du bruit. Cependant, il est particulièrement difficile de développer manuellement un algorithme pour exploiter efficacement ces critères visuels et parvenir à une discrimination précise des événements sismiques. Les variations subtiles et les motifs complexes dans les données sismiques rendent cette tâche difficile et sujette à des erreurs potentielles. Nous avons donc opté pour l'utilisation d'un Réseau de Neurones Convolutifs (CNN), connu pour leur efficacité dans la reconnaissance d'images. Les CNNs étant des algorithmes d'apprentissage supervisés, il est nécessaire de disposer d'une base de données d'entraînement. Elle doit se composer de deux classes de signaux :

- La classe *quake*, correspondant à l'ensemble des événements sismiques détectables dans la région, naturels et anthropiques.
- La classe *noise*, correspondant aux signaux parasites impulsifs qui ne sont pas causés par des tremblements de terre.

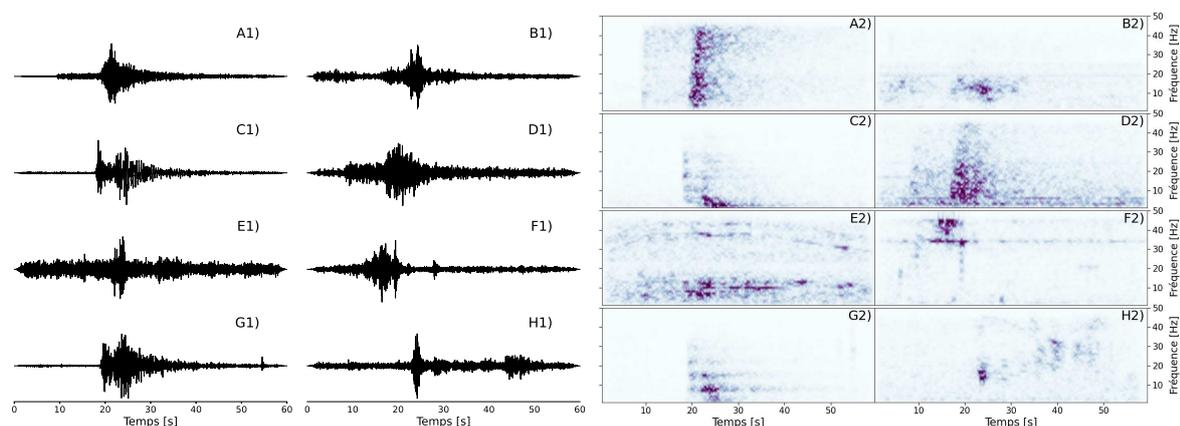


FIGURE 4.8 – Exemples de quatre signaux sismiques (A, C, D et G) et de quatre signaux impulsifs (B, E, F et H), enregistrés par les stations du réseau FR, avec les spectrogrammes associés. Les signaux correspondent aux données brutes normalisées de la composante verticale. Les spectrogrammes sont normalisés par leur maximum.

Pour le jeu de données de la classe « quake », nous utilisons toujours le bulletin du BSCF-RéNaSS entre le 1er janvier 2020 et le 1er juin 2021, pour des événements de $M_l < 3$. Pour chaque événement sélectionné, les formes d'ondes à trois composantes sont collectées aux stations pour lesquelles au moins un pointé d'ondes *P* est documenté dans le bulletin. Seuls les enregistrements des stations du réseau FR sont sélectionnés. Les stations du réseau RD sont exclues en raison de leur fréquence d'échantillonnage (50 Hz au lieu de 100 Hz pour les stations FR). Quant aux stations du réseau RA, elles sont inadaptées à la recherche des petits séismes. La base des données, pour la classe *quake*, se compose de 30 054 enregistrements de signaux naturels (correspondant à 5 172 événements naturels) et de 25 549 enregistrements de signaux anthropiques (correspondant à 4 877 événements anthropiques) enregistrés par 136 stations permanentes. Afin de disposer d'un jeu de données équilibré pour la classe *quake*, nous avons sélectionné aléatoirement, en utilisant une distribution uniforme, 10 500 enregistrements dans chaque classe. Nous obtenons ainsi un ensemble de 21 000 enregistrements correspondant à 4 237 et 4 099 événements naturels et anthropiques (Fig 4.9).

Pour le jeu de données de la classe *noise*, nous avons utilisé les résultats de l'application de *template matching* réalisés dans le quart nord-ouest de la France, entre le 1er août 2018 et le 31 juillet 2020, présentés dans [Beucler et coll. \(2021\)](#). J'ai réalisé une

4.2. DISCRIMINATION DES SIGNAUX SISMIQUES DES FAUSSES DÉTECTIONS

inspection visuelle de l'ensemble des détections par *template matching* afin de constituer notre base de données. Nous obtenons ainsi un ensemble de 21 000 enregistrements de signaux impulsifs composant la classe *noise*.

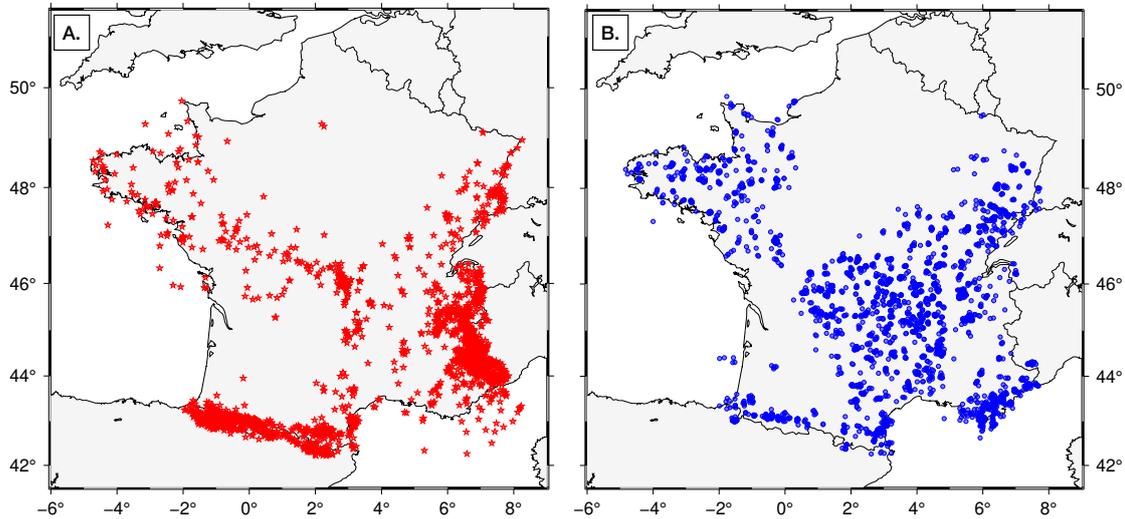


FIGURE 4.9 – Localisation des 8 336 événements constituant la classe *quake* des bases de données d'entraînement, de validation et de test, détectés entre 1er janvier 2020 et le 1er juin 2021 et étiquetés par le BSCF-RéNaSS. (A) 4 237 événements naturels sont représentés par les étoiles rouges. (A) 4 099 événements anthropiques (tirs de carrières et explosions) sont représentés par les cercles bleus.

Le même traitement est appliqué à toutes les formes d'onde. Celles-ci sont échantillonnées à 100 Hz et sont découpées en fenêtres de 60 s. Cette taille de fenêtre limite la distance source-station utilisée pour la discrimination à environ 400 km pour les événements sismiques, ce qui est cohérent avec notre choix de nous concentrer sur les événements $M_l < 3$ dans la région. Afin d'être moins sensible au fenêtrage du *template matching* et à la position des ondes sismiques dans la fenêtre d'analyse, nous autorisons que la fenêtre commence aléatoirement entre 5 et 20 s avant la sélection de l'onde P pour les signaux sismiques. Une apodisation est appliquée sur 5% des bords de la fenêtre. Les signaux sont filtrés à l'aide d'un filtre Butterworth passe-haut avec une fréquence coin de 2 Hz.

Les spectrogrammes sont calculés pour des fenêtres glissantes de 1 s avec un chevauchement de 75 %. Pour chacune des 237 fenêtres temporelles, nous calculons leur transformée de Fourier entre 1 et 50 Hz pour les trois composantes du sismogramme. Pour l'entraînement du réseau de neurones, les données doivent être normalisées. Pour cela, chaque spectrogramme est normalisé par son maximum. Pour chaque sismogramme, nous obtenons ainsi une matrice de $237 \times 50 \times 3$ éléments (temps \times fréquence \times

canaux). Nous disposons donc d'une base de données totale de 42 000 matrices tridimensionnelles correspondant aux trois composantes du sismogramme.

L'architecture du CNN est basée sur celui présenté dans Hourcade *et coll.* (2023) (Fig 4.10). Celui-ci est composé de quatre couches de convolution avec respectivement 18, 36, 68 et 68 filtres pour chaque couche et de trois couches entièrement connectées. Un *max-pooling* 2×2 entre chaque couche de convolution est appliqué pour permettre un sous-échantillonnage en ne conservant que les valeurs maximales. La dernière couche du réseau neuronal est la couche de sortie qui indique les probabilités obtenues avec la fonction d'activation *softmax*.

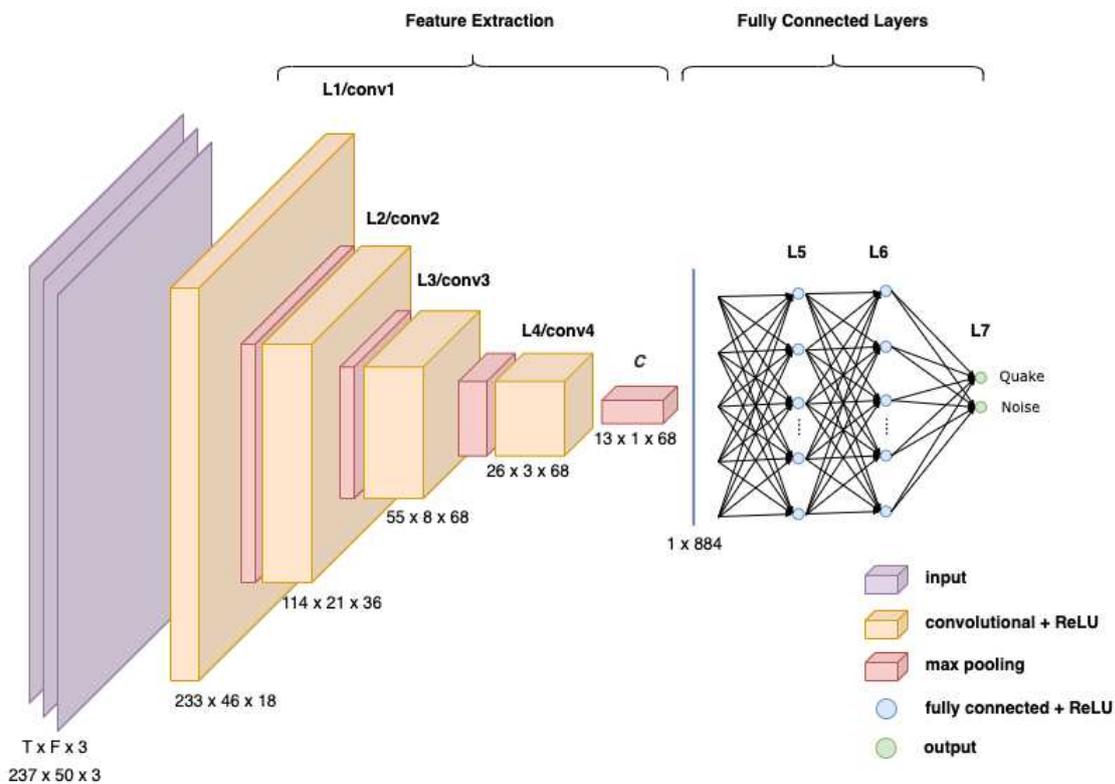


FIGURE 4.10 – Architecture du CNN développé. L'entrée de la matrice correspond aux spectrogrammes à trois composantes de taille 237×50 (temps \times fréquence). Les couches L1-L4 extraient les caractéristiques grâce à l'application d'une série de filtres permettant le sous-échantillonnage de la matrice. Cette extraction de caractéristiques conduit à la matrice C qui est convertie en un vecteur (*flatten*) pour être injectée dans les couches entièrement connectées. La sortie du CNN correspond aux deux probabilités pour les classes : *quake* et *noise* (modifié d'après Hourcade *et coll.*, 2023)

L'ensemble de la base de données est divisée en trois ensembles de données différents, un pour chaque étape de la création du modèle : l'entraînement (80% de la base de données), la validation (10% de la base de données) et le test (10% de la base de données). Nous avons répété l'entraînement du CNN 10 fois (*tenfold cross valida-*

4.2. DISCRIMINATION DES SIGNAUX SISMOLOGIQUES DES FAUSSES DÉTECTIONS

tion) en mélangeant aléatoirement la base de données afin de créer, à chaque itération, une nouvelle répartition des données entre « entraînement », « validation » et « test ». Le modèle final, présenté dans cette étude, correspond au modèle ayant la meilleure performance parmi les 10 entraînements.

Classiquement, pour quantifier les performances d'une classification (Tharwat, 2020; Hourcade *et coll.*, 2023), quatre paramètres sont utilisés, l'*accuracy* (Acc), la précision (Pr), le *recall* (Re) et le F1-score (F1) :

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (4.3)$$

$$Pr = \frac{TP}{TP + FP}, \quad (4.4)$$

$$Re = \frac{TP}{TP + FN}, \text{ et} \quad (4.5)$$

$$F1 = 2 \frac{RePr}{Re + Pr}. \quad (4.6)$$

		Prédiction CNN	
		Quake	Noise
Réalité	Quake	True Positive (TP)	False Negative (FN)
	Noise	False Positive (FP)	True Negative (TN)

FIGURE 4.11 – Exemple de matrice de confusion (modifié d'après Hourcade *et coll.* (2023)). Chaque cellule d'une matrice de confusion contient une estimation de la performance du CNN.

Les vrais positifs (TP) représentent le nombre de *quake* correctement classés par le CNN et les faux négatifs (FN) représentent le nombre de *quake* classés comme *noise*. Les faux positifs (FP) représentent le nombre de *noise* classés comme *quake* et les vrais

négatifs (TN) représentent le nombre de *noise* correctement classés par le CNN (Figure 4.11). L'idéal est d'obtenir aucun FN et aucun FP.

L'*accuracy* (eq 4.3) est le rapport entre la somme de toutes les prédictions correctes et la somme de l'ensemble des prédictions. Elle indique si la discrimination est correcte. La précision (eq 4.4) indique dans quelle mesure l'algorithme est efficace pour la discrimination de ce qui est appelé « positif » (dans notre cas, *quake*). Lorsque la précision est élevée, cela signifie que la majorité des prédictions positives du modèle sont correctes. Le *recall* (4.5) quantifie le ratio des prédictions de classes positives sur l'ensemble des classes positives de la base de données. Lorsque le *recall* est élevé, cela signifie que la majorité des éléments positifs sont bien classés. Le score F1-score (eq 4.6) est une moyenne pondérée de la précision et du *recall*. Ce score tient donc compte à la fois des faux positifs et des faux négatifs.

Le CNN a obtenu d'excellents résultats avec une *accuracy* pour la base de données d'entraînement de 99,16%. Sur l'ensemble des données de « validation » (10% de la base de données, soit 2 100 enregistrements), le modèle final atteint une *accuracy* de 98,22%. Enfin, sur l'ensemble de données « test » (10% de la base de données, soit 2 100 enregistrements), le modèle atteint une *accuracy* de 98,63%.

La discrimination automatique entre les événements sismiques et les fausses détections est donc réalisée à partir de ce CNN entraîné. Les événements classés en « *quake* » peuvent cependant avoir deux origines différentes : (1) naturels (origine tectonique) ou (2) anthropiques (origine humaine). Dans le cadre de cette étude, nous nous intéressons à la caractérisation de l'activité sismique naturelle du nord-ouest de la France. Une seconde étape de discrimination pour séparer ces deux types d'événements sismiques est nécessaire.

4.3 Discrimination entre les événements naturels et anthropiques

La dernière étape, pour l'obtention d'un catalogue de sismicité naturelle, consiste à discriminer les événements naturels et anthropiques. Avec l'augmentation du nombre de détections, due à la densification des réseaux de stations, nous sommes face à une plus grande diversité de formes d'ondes et la distinction entre les événements naturels et anthropiques devient de plus en plus complexe et longue. Lorsqu'il s'agit de tirs réalisés dans une carrière spécifique et enregistrés par une même station, les formes d'ondes des signaux présentent généralement des similarités. La Figure 4.12 présente quatre exemples d'enveloppes de signaux anthropiques provenant de la carrière de La Vraie Croix, enregistrés par la station BEGF. Bien que tous ces signaux proviennent de la même carrière, nous observons néanmoins des variations en termes d'amplitude et de formes d'ondes.

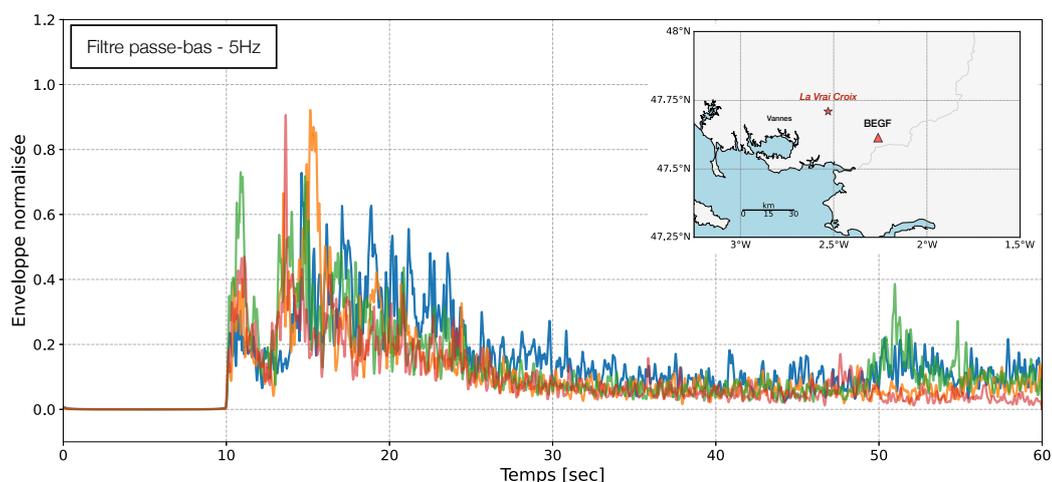


FIGURE 4.12 – Exemples d'enveloppes de formes d'ondes enregistrés par la station BEGF provenant de la carrière située à La Vraie Croix à 34 km.

L'emplacement précis des tirs au sein de la carrière peut créer des variations notables dans les formes d'ondes enregistrées. Les caractéristiques de ces dernières dépendent de plusieurs facteurs, tels que la puissance de l'explosion, les propriétés spécifiques du gisement et la topographie environnante. Par exemple, la puissance de l'explosion peut influencer l'amplitude et la durée des signaux. Chacun de ces éléments contribue à façonner la réponse sismique enregistrée par les stations.

Lorsque nous réalisons une inter-corrélation entre ces différentes enveloppes, le maximum obtenu atteint 0,45. Il n'est donc pas possible d'utiliser, par exemple, un

signal de référence unique par carrière pour discriminer les événements sismiques. Il est donc nécessaire de trouver une solution fiable et efficace pour aider à éliminer tous ces événements sismiques anthropiques du catalogue.

D'après la littérature (Section 3.2.3), la plupart des méthodes de discrimination existante nécessite de connaître les temps d'arrivées des phases sismiques pour pouvoir être appliquée, par exemple, avec les rapports d'amplitude entre les ondes P et S (Pyle et Walter, 2019), l'extraction des caractéristiques et l'implémentation dans un *Random Forest* (Renouard et coll., 2021). Or, nous disposons seulement de temps de détection (ou de maximum de corrélation) et de fenêtres d'analyse sélectionnées autour de ce temps. Nous devons donc développer une méthode adaptée à notre base de données.

4.3.1 New CNN-based tool to discriminate anthropogenic from natural low magnitude seismic events

L'article ci-dessous explique la méthode de discrimination que nous avons développée pour notre problème. Il introduit l'algorithme choisi, dans notre cas, un CNN et la base de données utilisée pour l'entraîner. Il présente ensuite les résultats d'application de l'algorithme sur trois bases de données indépendantes. Sur une base de données de France métropolitaine, qui n'a pas été utilisée lors de la phase d'entraînement, le CNN obtient un score élevé avec une *accuracy* de 98,18%. L'approche est appliquée, sans ré-entraînement, sur un jeu de données correspondant à des événements détectés par l'University of Utah Seismograph Stations (UUSS) aux États-Unis, et celui-ci obtient une *accuracy* de 96,72%. Enfin, le CNN est appliqué sur les données d'une campagne post-sismique qui a eu lieu dans le nord-ouest de la France et obtient une *accuracy* de 100%. Les résultats des différentes classifications sont ensuite discutés. Cet article montre un ensemble de paramètres pouvant influencer les performances de classification (taille de la base de données d'entraînement, magnitude, couverture de station autour d'un événement, etc.).

Nous allons ensuite approfondir certains points sur l'utilisation et la nécessité d'un réseaux de stations pour la discrimination des événements sismiques. En utilisant un seul enregistrement de station, il peut être difficile de déterminer avec précision la nature de l'événement sismique. C'est là que l'apport d'un réseau de stations devient crucial. En disposant de multiples enregistrements provenant de différentes stations, nous disposons d'une meilleure couverture autour de l'événement sismique permettant de mieux distinguer les événements sismiques naturels des anthropiques.

New CNN based tool to discriminate anthropogenic from natural low magnitude seismic events

Céline Hourcade¹, Mickaël Bonnin¹, Éric Beucler¹

¹ Nantes Université, Université d'Angers, Le Mans Université, CNRS, UMR-6112, Laboratoire de planétologie et géosciences, F-44000 Nantes, France

E-mail: celine.hourcade@etu.univ-nantes.fr

Published in *Geophys. J. Int.* : 10 November 2022
<https://doi.org/10.1093/gji/ggac441>

Abstract

With the deployment of high quality and dense permanent seismic networks over the last 15 years comes a dramatic increase of data to process. In order to lower the threshold value of magnitudes in a catalogue as much as possible, the issue of discrimination between natural and anthropogenic events is becoming increasingly important. To achieve this discrimination, we propose the use of a convolutional neural network (CNN) trained from spectrograms. We built a database of labeled events detected in metropolitan France between 2020 and 2021 and trained a CNN with three-component 60 s spectrograms ranging frequencies from 1 to 50 Hz. By applying our trained model on independent French data, we reach an accuracy of 98.2%. In order to show the versatility of the approach, this trained model is also applied on different geographical areas, a post-seismic campaign from NW France and data from Utah, and reaches an accuracy of 100.0% and 96.7% respectively. These tests tend to hypothesise that some features due to explosions compared to earthquakes are widely shared in different geographical places. In a first approach, we propose that it can be due to a contrast in the energy balance between natural and anthropogenic events. Earthquake seismic energies seem to be more continuous as a function of frequency (vertical bands features in a spectrogram) and conversely for explosions (horizontal strips).

Key words: Seismicity and tectonics, Neural networks, fuzzy logic; Machine Learning; Time-series analysis.

1 Introduction

Studying low magnitude earthquakes allows us to better understand natural seismic activity, especially in stable continental regions where it is challenging to explain the occurrence of earthquakes under slow strain-rate conditions. Before 2000's, most of seismic stations operated in triggered mode, recording only events that exceeded an operator-defined detection threshold, while most of present-day stations continuously record the ground motion. This evolution comes at the cost of an increase in the processing capabilities but interestingly offers the possibility to reprocess "old" dataset with new methods to detect events with lower signal-to-noise ratio (SNR), and consequently lower magnitudes seismic events. Denser seismic station networks and high sensitivity modern instruments improve our understanding of global and local seismicity over the last decade (*e.g.* [Levandowski et al., 2018](#); [Beucler et al., 2021](#)). As an illustration of the evolution of the seismic data volume, the size of the IRIS DMC (Incorporated Research Institutions for Seismology Data Management Center) archive has grown from 25 terabytes (TB) in January 2005 to 800 TB in January 2022. In metropolitan France, the number of stations from the FR network ([RESIF, 1995](#)) has increased from 56 in January 2005 to 157 in January 2020. This massive flow of global seismic data implies an important human and temporal cost for the data processing and creation of a catalogue of seismic events (*e.g.* detection, picking of seismic phases, discrimination of events).

One of the major challenges in developing an event catalogue, is the discrimination between natural (tectonic events) and anthropogenic events (caused by human activity, as quarry blasts or military explosions). In metropolitan France, instrumental seismicity is mostly characterized by low magnitude earthquakes ([Cara et al., 2015](#)). The deployment of new permanent stations since mid-2018 ([Beucler et al., 2021](#); [Sylvander et al., 2021](#); [Larroque et al., 2021](#); [Dobre et al., 2021](#)) has conducted to an increase in the number of detected natural and anthropogenic seismic events. However, metropolitan seismicity is now dominated by the detection of anthropogenic events (for instance, see figure 7 in [Beucler et al., 2021](#)), originating from quarry blasting, military training operations or underwater mine explosions along the French coast (*e.g.* [Favretto-Cristini et al., 2020](#)). It is therefore necessary to find an efficient solution to help the discrimination of all these low magnitude seismic events.

Several semi-automatic or automatic approaches have already been developed using, for example, the amplitude ratio between seismic phases ([Dysart et Pulli, 1990](#); [Baumgardt et Young, 1990](#); [McLaughlin et al., 2004](#); [Tibi et al., 2018](#); [Pyle et Walter, 2019](#); [Wang et al., 2020](#)), using spectral content ([Gitterman et al., 1998](#); [Allmann et al., 2008](#)), or studying the coda ([Su et al., 1991](#); [Koper et al., 2016](#)). These types of methods are highly dependent of the SNR and of the epicentral distance (some seismic phases may not be visible at very short distances). However, extracting these information from large datasets can be quite time consuming and the choice of these parameters can be seen as relatively subjective.

To overcome these problems, such as finding relevant features to classify the data, machine learning tools can be implemented. In the last few years, these approaches were widely used for the detection of events ([Yoon et al., 2015](#); [Ross et al., 2018](#); [Mousavi et al., 2020](#)), automatic picking of seismic phase on raw data ([Zhu et Beroza, 2019](#); [Woollam et al., 2019](#); [Pardo et al., 2019](#)), but also for signal classification using features extracted directly from the data ([Del Pezzo et al., 2003](#); [Meier et al., 2019](#); [Renouard et al., 2021](#)), or relying on deep neural networks ([Li et al., 2018](#); [Linville et al., 2019](#)). In the latter, training algorithms with a sufficiently large database allow thoses programs to recognise, like humans, natural objects and to make expert-level decisions.

Among those algorithms, convolutional neural networks (CNNs, a nonlinear classifiers which is a subcategory of neural networks) can achieve high performance recognition and outperform simple linear classifiers. [Linville et al. \(2019\)](#) show that using spectrograms with a frequency content between 1 and 20 Hz as input for a CNN allows to achieve high predictive capability in distinguishing natural from anthropogenic events. If the frequency range used by ([Linville et al., 2019](#)) is well suited for local

to regional moderate magnitude events, in cases of low magnitudes events ($M_l < 2$) occurring in low attenuation medium (such as Armorican Massif, Mayor *et al.*, 2018) a significant part of the recorded signal reaches frequencies up to 50 Hz. We thus propose to extend the spectrogram to frequencies larger than 20 Hz in order to take into account as much as possible the high-frequency spectral content. An additional constraint of the algorithm chosen in Linville *et al.* (2019) is the requirement of *P*-wave picking before the discrimination. By randomly selecting the beginning of the windows studied, we want to overcome this problem and apply the method directly after the detection step. CNNs can moreover automatically extract discriminative features from training data. One of the objectives of this study would be to verify whether these features deduced during training, on a specific geographical region, can be generalized to other regions.

In Section 2, we describe the CNN and the dataset used to train it. Three different applications are presented in Section 3 which illustrates the versatility of the proposed approach. To this end, we used 3 datasets not used for training. These are composed of events located in different geographical areas: Metropolitan France, the United States and a local study in NW France. Finally, in Section 4, we discuss the performance of our discrimination tool and the implications for decision support for discrimination between natural and anthropogenic events.

2 Method

2.1 Targets

The main objective of this article is to describe an alternative discrimination of natural and anthropogenic seismic events, which inherently means low-to-intermediate magnitudes. At the present time, this classification is generally done by experienced analysts. Among criteria used to discriminate seismic events, the most widely used are the source parameters of the events (location, time of origin, magnitude) and the characteristics of the waveforms. Quarry blasts and explosions are generally observed between 9 a.m. and 6 p.m. during working days, whereas natural events can occur at any time of the day and night.

However, some configurations can make discrimination more challenging. For instance, a natural event occurring near the surface and/or near a quarry, can induce signals with characteristics similar to quarry blasts and will very likely be tagged as “anthropogenic event” by an analyst. Low magnitude events produce signals with a low SNR, so differentiation can be complicated. In addition, with the increase in the amount of detections, due to the densification of station networks, we are dealing with a larger diversity of waveforms and the distinction between natural and anthropogenic events is becoming increasingly complex and time consuming.

To illustrate this diversity, Figure 1 presents eight examples of natural and anthropogenic events (traces and spectrograms) labeled by the French National agency BCSF-RéNaSS (*Bureau Central Sismologique Français - Réseau national de surveillance sismique*), after visual inspection. Among them, five are natural events and three are anthropogenic events. The most striking difference between the two categories is the amplitude ratio between *P* and *S*-waves. For natural events (Figure 1A, B, D, F and G) this ratio is generally lower than 1.0 while it almost equals to 1.0 for explosions (Figure 1C, E and H). On the spectrograms (right panels of Figure 1) of the natural events, we can clearly see two arrivals with a strong energy packet associated to the *S*-wave arrival that forms a vertical band that affects the whole frequency range. For the anthropogenic events, energy is generally confined to lowest frequencies and forms horizontal strips. Spectrograms thus seem to bear sufficient “visual” information to help us to discriminate natural from anthropogenic events. Examples for marine explosions and shallow earthquakes are shown in Figure S1.

To achieve high classification performances, and given the fact that spectrograms are images, we have

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

chosen to implement an automatic classifier with a CNN which are “machine learning” approaches that were proven to be particularly efficient for image recognition.

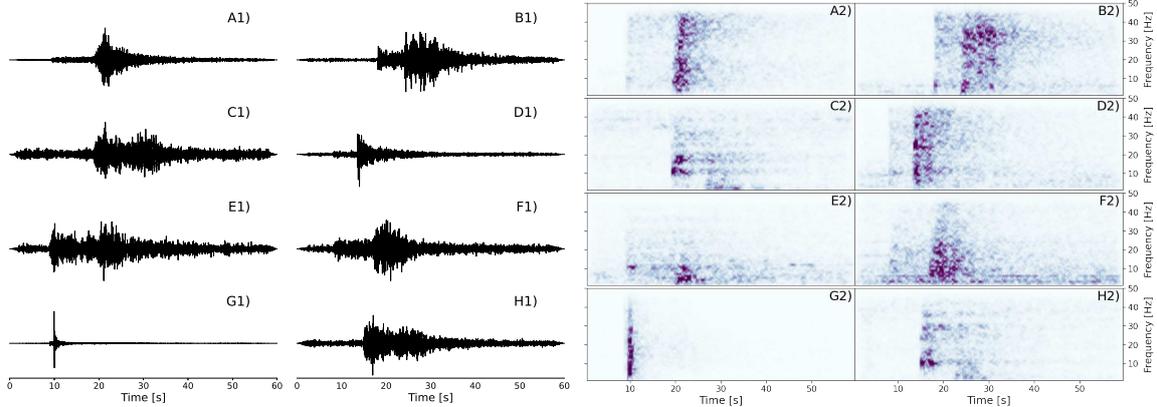


Figure 1: Examples of five natural events (A, B, D, F, G) and three anthropogenic events (C, E, H), recorded by the FR network stations, with the associated spectrograms. The seismic signals correspond to the normalized raw data of the vertical component. The spectrograms are normalized by their maximum. All technical details of the events are in Table S1.

2.2 Convolutional Neuronal Network

The discrimination of the two families of signals is achieved through a supervised learning algorithm based on a CNN. A CNN allows to extract discriminating features through the successive application of hierarchical filters of defined size. The choice of implemented CNN architecture is guided by those presented in the literature (*e.g.* Meier *et al.*, 2019; Linville *et al.*, 2019) and by standard choices made in Deep Learning (*e.g.* Géron, 2019). The first layers correspond to the convolution layers which, thanks to filters of a defined size, allow the extraction of features from the input. Then, these features are implemented in fully connected layers where the number of neurons per layer varies. We tested about ten different architectures by varying the number of convolution layers (from 3 to 4 layers), the number of filters (16 to 128) and the number of neurons in the fully connected layers (80, 128, 256). The training accuracy (see definition in Section 2.4) of these different architectures ranges from 92.83% to 95.63%.

The CNN with the best performances is composed of four convolution layers with respectively 18, 36, 68 and 68 filters for each layer and three fully connected layers (TAB. 1). We apply a 2×2 “max-polling”, between each convolution layer, to allow a sub-sampling by keeping only the maximum values. The final layer of the neural network is the output layer that indicates the probabilities obtained with the “softmax” activation function (FIG. 2). The coefficients of the filters and the weights of the fully connected network are jointly optimized during the training process to maximize the discrimination performance. As input to this CNN, we use normalized spectrograms of size $237 \times 50 \times 3$, corresponding to time (T) \times frequency (F) \times channels. The process to obtain these spectrograms is explained in the next section.

Table 1: Technical details of the CNN. conv: convolution, fc: fully connected, ReLU: rectified linear unit. Convolution layers from L1 to L4 and fully connected layers from L5 to L7.

	L1	L2	L3	L4	L5	L6	L7
Layer Type	conv	conv	conv	conv	fc	fc	fc
Number of filter / neuron	18	36	68	68	80	80	2
Filter size	(5,5)	(3,3)	(3,3)	(2,2)			
Activation	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	softmax

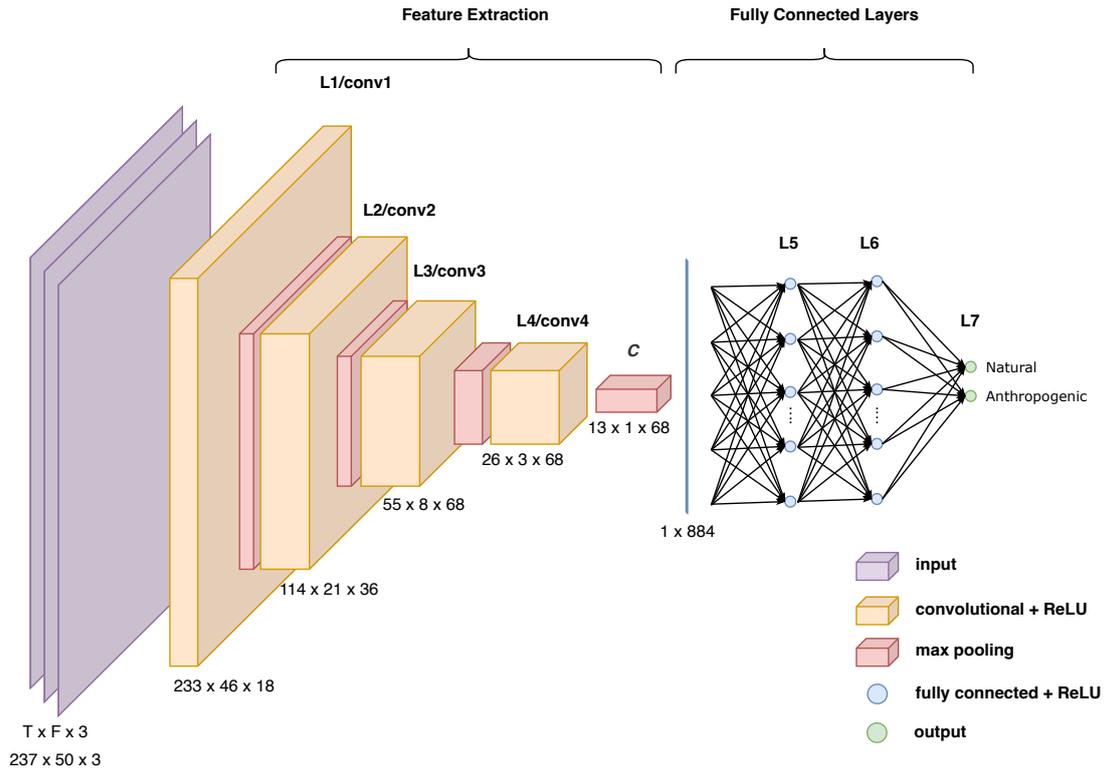


Figure 2: The CNN architecture. The input of the array corresponds to the three-component spectrograms of size 237×50 (time \times frequency). The L1 to L4 layers extract the characteristics thanks to the application of a series of filters allowing the down-sampling of the matrix. This feature extraction leads to the C matrix which is flattened to be injected in the fully connected layers. The CNN output corresponds to the two probabilities for the classes: natural and anthropogenic. See Table 1 for more technical details on the CNN.

2.3 Training dataset

The CNN is trained on a data set composed of elements tagged with two classes: natural and anthropogenic (quarry blast and explosion). To generate this database, we use the BSCF-RéNaSS bulletin between 2020-01-01 and 2021-06-01, for events with $M_l \leq 3$. This bulletin is restricted to metropolitan France extended outside its exclusive economic zone of 20 km. For each selected event, the three-component seismic waveforms are collected at stations from FR network for which at least a P -wave pick is documented in the bulletin.

The dataset is composed of 30,054 natural events records (corresponding to 5,172 natural events) and 25,549 anthropogenic events records (4,877 anthropogenic events) at 136 permanent stations. To have a balanced database between the two classes, we have randomly selected, using a uniform distribution, 25,000 records in each class. This provides a set of 50,000 records that correspond to 5,016 and 4,788 natural and anthropogenic events, respectively. The consequences of different choices for the training dataset are discussed in Section 4.2.

Figure 3 shows the geographical distribution of the stations used to train the CNN. The majority of the stations are grouped in the mountainous areas (Alps in Zone B and Pyrenees in Zone C). In the rest of France, the stations are more evenly distributed.

Figure 4 shows the geographic distribution of natural and anthropogenic events in the database. We observe areas with a high concentration of natural events, such as the Pyrenees (Zone C), the Alps and

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

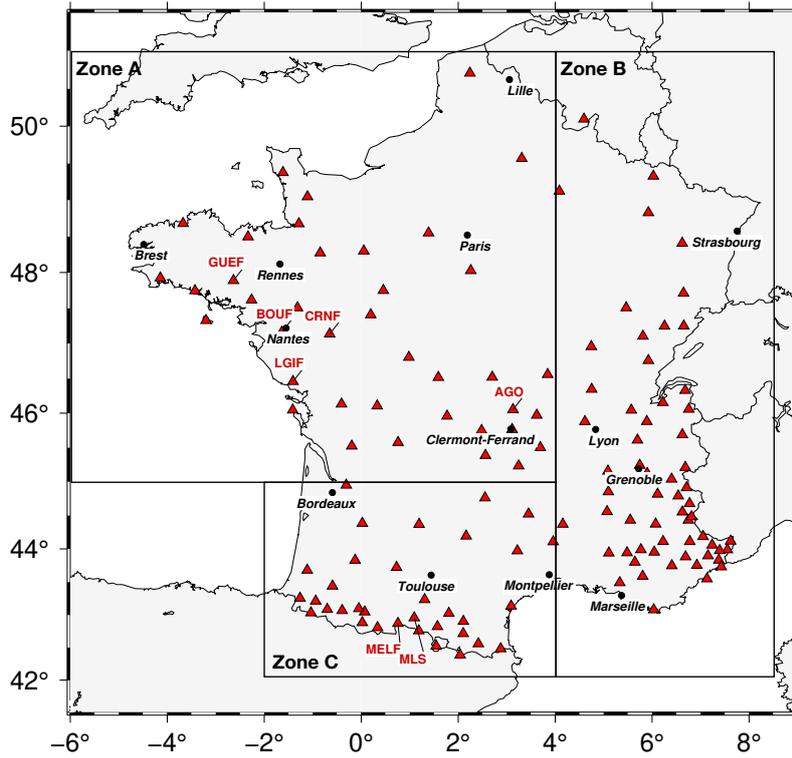


Figure 3: Geographical distribution of the stations used for the training of the CNN and the end-2021 French data application (136 permanent stations of the FR network). The Metropolitan France has been divided into 3 sub-sections. These geographical Zones A, B and C are discussed in Section 4. The number of stations per zone is shown in Table 2. The seven named stations are used as examples for Figure 1.

Table 2: Partition of the data set.

	Zone A	Zone B	Zone C	Total
Natural Events	255	2,612	2,149	5,016
Anthropogenic Events	1,407	2,415	966	4,788
Records	11,126	22,818	16,056	50,000
Stations	40	62	34	136

the Upper Rhine Grabben (Zone B) (TAB. 2). In comparison, in the Armorican Massif and the Massif Central (Zone A), known as stable continental areas, fewer scattered natural events are observed. As expected, we have a few or no earthquakes in the Aquitaine Basin and the Paris Basin, which coincides with the general seismicity of these regions (Cara *et al.*, 2015). For anthropogenic events, the distribution of quarries is very heterogeneous and also related to the geology and tectonic history of the regions. In Zone A, 84.4% of the events correspond to anthropogenic events, whereas for Zones B and C, the ratios are 48% and 31%, respectively (TAB. 2).

Waveforms are sampled at 100 sps (samples per second) and are sliced into 60 s windows. This window size limits the source-station distance used for discrimination to about 400 km which is consistent with our choice to focus on $M_l \leq 3$ event. To mitigate the use of *a priori* information on the P -wave arrival and to make the CNN less sensitive to picking errors or phase mislabelling, we impose that the window starts randomly between 5 and 20 s before the P -wave pick. Indeed, depending on the focal mechanisms, the SNR or the method used (for instance STA/LTA (*e.g.* Allen, 1982) or template matching (*e.g.* Gibbons *et Ringdal*, 2006)), detection time can indeed correspond to P or S -wave onset. This method does not

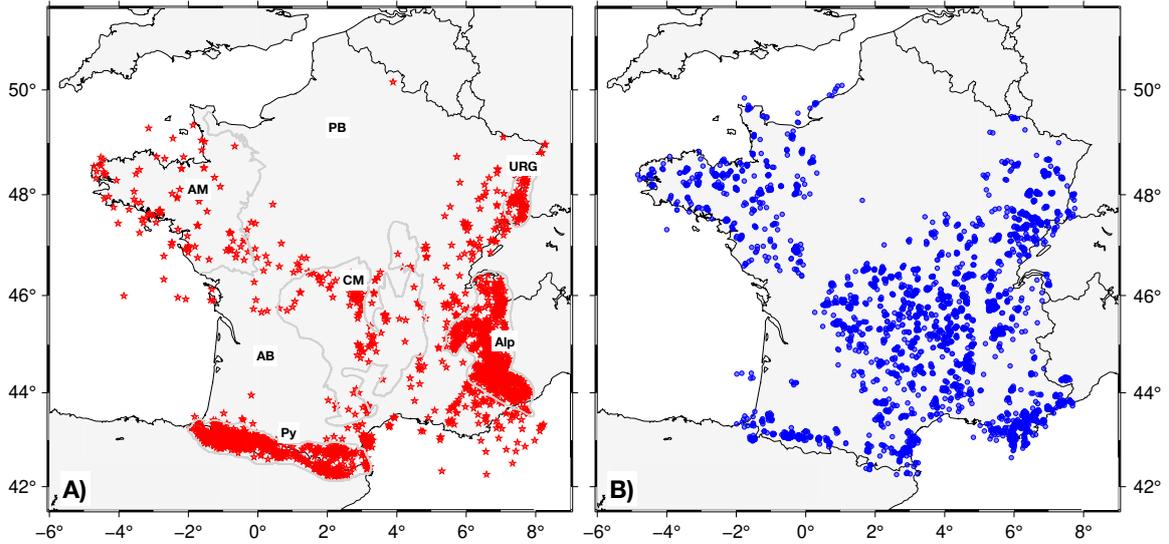


Figure 4: Geographic distribution of training, validation, and test databases composed of 9,804 events detected between 2020-01-01 and 2021-06-01 and labelled by the BSCF-ReNaSS. A) 5,016 natural events are represented by the red stars. The main French geological regions are contoured by the grey lines : PB: Paris Basin, URG: Upper Rhine Grabben, AM: Armorican Massif, CM: Central Massif, AB: Aquitaine Basin, Alp: Alps and Py: Pyrenees. B) 4,788 anthropogenic events (quarry blasts and explosions) are represented by the blue circles.

require an exact and reliable choice of the P -wave onset time.

To standardise the data given to the CNN, waveforms are detrended, tapered (Hann window, 5% of the total window) and filtered with a 2 Hz high-pass Butterworth filter (four-corner, two passes). Spectrograms are computed for 1 s sliding windows with 75% overlap. We end with 237 time windows for which we compute their discrete Fourier Transform between 1 and 50 Hz for the three components of the seismogram. For each seismogram, *i.e.* each source-station couple, we thus obtain $237 \times 50 \times 3$ elements matrices (time \times frequency \times channels).

2.4 Training results

Generally, in machine learning algorithms (*e.g.* James *et al.*, 2013), the whole database is split into three different datasets, one for each stage of the model creation: training (80% of the database), validation (10% of the database) and test sets (10% of the database). The model is trained with a dataset composed of examples used to adjust the parameters results of the model (*i.e.* filter coefficients, weights of the connections between neurons in the fully connected network). At the same time, the trained model is used to predict the observations in a second dataset called the “validation” dataset. The “validation” dataset is used to evaluate the fit to the data during training. The training stops when the accuracy on the “validation” set remains stable over three iterations (epochs). We choose the oldest epoch with the highest validation accuracy as the final model. The “test” dataset is finally used to provide an impartial evaluation of a final model and the results of the classification are presented in the form of a confusion matrix (FIG. 5). An element is categorized as natural or anthropogenic by looking at the highest probability of prediction. We repeated the entire process 10 times (tenfold cross validation) by randomly mixing the database to create the “training”, “validation” and “test” datasets. The final model, presented in this study, corresponds to the model with the best performance among the 10 trainings. Classically, to evaluate the performance of a classification (*e.g.* Tharwat, 2020), we define the accuracy (Acc), precision (Pr), recall (Re) and $F1$ -score ($F1$) as

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (1)$$

$$Pr = \frac{TP}{TP + FP}, \quad (2)$$

$$Re = \frac{TP}{TP + FN}, \text{ and} \quad (3)$$

$$F1 = 2 \frac{RePr}{Re + Pr}. \quad (4)$$

		CNN	
		Natural	Anthropogenic
Reality	Natural	True Positive (TP)	False Negative (FN)
	Anthropogenic	False Positive (FP)	True Negative (TN)

Figure 5: Example of a confusion matrix. Each cell of a confusion matrix contains a performance estimate of the CNN. In our case “Positive” means Natural and conversely “Negative” for Anthropogenic. The term “Reality” means what the National agency BCSF-RéNaSS tagged as Natural and Anthropogenic. If the CNN gives exactly the same results as the Reality, only the main diagonal has non-zero percentage values.

True Positive (TP) is the number of natural events correctly classified by the CNN and False Negative (FN) represents the number of natural events classified as anthropogenic. False Positive (FP) is the number of anthropogenic events classified as natural and True Negative (TN) represents the number of anthropogenic events correctly classified by the CNN.

Since the accuracy (1) is the ratio of the sum of all the true predictions over the amount of all possible predictions, it indicates whether the discrimination is correct. The precision (2) indicates how the algorithm is relevant for the discrimination of what is named “Positive” (in our case, Natural). When the precision is high, it means that the majority of the positive predictions of the model are correct. The recall (3) quantifies the ratio of positive class predictions over all the actual positive classes. When the recall is large, it means that the majority of positive only elements are well classified. The $F1$ -score (4) is a weighted average of precision and recall. This score thus accounts for both false positives and negatives.

The classifier achieved high scores with a training accuracy of 95.63%. The score is computed using all elements independently of the event they are associated with, which is called a station-level prediction. It is possible to obtain an event prediction by combining all the elements at the station-level that correspond to the same event, it is called a network-level prediction. As a first approach, for a given event, a linear average of all station-level probabilities defines the probability at network-level that leads to the event-based classification.

On the “validation” data set (10% of the database, *i.e.* 5,000 records), the final model reaches an accuracy of 94.76% at the station-level and of 95.11% at the network-level. The recall of natural and anthropogenic events is 93.58% and 96.58% respectively. The model incorrectly identifies 6.42% of the natural events and 3.42% of the anthropogenic events, thus, it is more likely to correctly predict the class of anthropogenic events. Finally, on the “test” data set (10% of the database, *i.e.* 5,000 records), the model reaches an accuracy of 94.68% at the station-level and of 95.3% at the network-level.

3 Various discriminations using a unique trained model

To illustrate the relevance of the scheme chosen for the CNN and the versatility of our trained model, we apply it on three different case studies, none of them were used to train it: (i) a dataset composed of events detected in metropolitan France for 5 months, (ii) a revisit of the events detected in the United States, mainly in Utah, and recorded by the University of Utah Seismograph Stations (UUSS) in 2016 (Linville *et al.*, 2019) and (iii) a dataset composed of events recorded during a post-seismic campaign

deployment (Haugmard, 2016). In this section, we discuss the events misclassified by the model and, in order to identify them, we have taken as reference the tags established by the national agencies or established manually in the case of the post-seismic campaign. Some useful information on the dataset (magnitude range and epicentral distances) can be found in Table S3 of the Supplementary Materials.

3.1 Events in metropolitan France in late 2021

The dataset considered is composed of events detected by the French national agency BCSF-RéNaSS between June 1st and November 1st, 2021 in metropolitan France. The same pre-processing steps as described in Section 2.3 are followed. The dataset consists of 9,011 records of natural events (associated to 1,467 natural events) and 11,049 records of anthropogenic events (2,045 anthropogenic events).

The geographical distribution of all the events is represented in Figure 6-A. The natural events are principally concentrated in the Pyrenees and the Alps while anthropogenic events are mainly located in the Massif Central and the Massif Armoricain. No earthquake are recorded in the Paris Basin and the Aquitaine Basin. This distribution is close to the one observed in Figure 4.

Assuming that the discrimination made by national agency is reliable and can be used as reference, Figure 6-D and Figure 6-E shows probabilities obtained from the prediction for natural and anthropogenic events. Figure 6-D corresponds to the probabilities obtained at a station-level, by taking into account each seismic trace independently. We observe that 95.2% of the anthropogenic signals and 89.9% of the natural signals are concentrated on the two edges of the probability range (between 0-0.25 and 0.75-1.0). At the network-level (FIG 6-E), and taking into account all the stations for a given event, we observe that 94.8% of the anthropogenic events and 89.5% of the natural events are concentrated on the two edges of the probability range. This shows that the majority of the data are correctly classified with a high probability and confidence levels. In view of these results, we can establish a confidence level at 75% probability to produce the confusion matrix.

Figure 6-C shows the results of the classification, at a network-level, as a confusion matrix. The CNN model obtains high scores (accuracy of 98.18%) with, in particular, recall for natural and anthropogenic events of 96.86% and 99.12%, respectively. This indicates that, out of the entire database, approximately 3% of the natural events and less than 1% of the anthropogenic events are misclassified. Among the 1,467 natural seismic events classified, 76 are located at less than 1 km depth. Our method thus correctly classified 70 of them (92.10 %). The precision indicates that 98.74% of the predictions of natural events made by the model are correct. The *F1* scores for natural and anthropogenic events are 97.79% and 98.46% respectively. This shows that the model is efficient in classifying natural and anthropogenic events on independent data. We can also observe that the majority of the predictions are made with a high level of probability: 3,382 events (out of the 3,512 in total) are classified by the model with a probability larger than 75%.

Figure 6-B represents the location of the events misclassified by the CNN model. Among 3,512 events, only 64 are not correctly classified by the model, which represents 1.82% of the events. They seem to be located mostly on the eastern border of metropolitan France (Zone B, FIG 3).

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

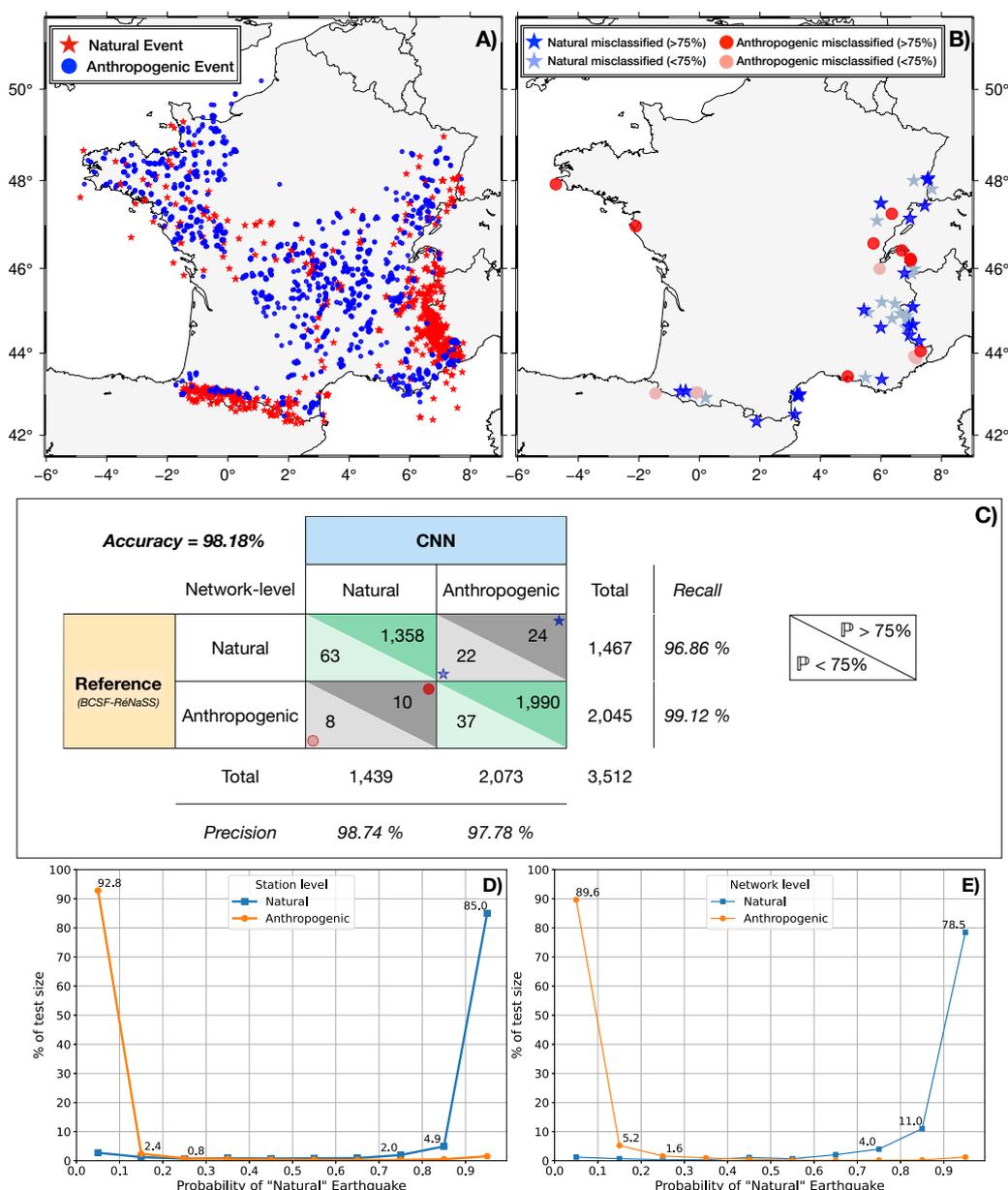


Figure 6: Application on the official French dataset detected between June 1, 2021 and November 1, 2021 from the BCSF-RéNaSS bulletin. (A) Geographic distribution of all seismic events used for this application (1,450 natural and 2,062 anthropogenic events). (B) Geographic distribution of events misclassified by the model. The blue stars correspond to events labeled as “natural” and classified as “anthropogenic” by the model. The red circles correspond to events labeled “anthropogenic” and classified as “natural” by the model. The darker colors correspond to a prediction of more than 75% by the CNN and the lighter colors correspond to a prediction of less than 75% by the CNN. (C) Confusion matrix, for the differentiation between natural and anthropogenic events, obtained from the French dataset. The reference is the discrimination made by BCSF-RéNaSS. The classification is done at the network level. \mathbb{P} is the probability of an event. The symbols correspond to the events located on (B). (D) Normalized histogram of the output probability of being a natural earthquake for natural and anthropogenic earthquake detected between June 1st, 2021 and November 1st, 2021, at a station-level and (E) at a network-level.

3.2 University of Utah Seismograph Stations (UUSS)

The second application corresponds to events detected in the United States, Utah, in 2016. This dataset offers the possibility to test the efficiency of our model in a drastically different context compared to the data used for its construction. The main idea is to check whether the CNN is able to discriminate anthropogenic events with characteristics (*e.g.* source function, power) that can differ from those in France.

Utah is seismically active due to its abundance of faults and fracture zones with, among the most active in Utah, the Wasatch Fault along the Wasatch Front (*e.g.* Arabasz *et al.*, 1980). We use the event solutions including source location of events, labeled as natural or anthropogenic events, detected between 2016-01-01 and 2016-03-01 by University of Utah Seismographic Stations (UUSS). For each event, we recover seismic-waveform data only from three-component broadband permanent seismometers located within 150 km of the earthquake location. *P*-wave times arrival were manually identified to extract the signals. This represents 1,248 records of natural events (288 natural events) and 398 records of anthropogenic events (108 anthropogenic events). We use 49 permanent stations from UU (University of Utah, 1962), WY (University of Utah, 1983), NN (University of Nevada, Reno, 1971) and AE (Arizona Geological Survey, 2007) networks. The geographical distribution of all the events used for this application is represented in Figure 7-A. The natural events are mainly located along active faults in the north-south direction with a clustering at Yellowstone in the north. The anthropogenic events are quarry blasts located on 7 different quarries.

Assuming that the discrimination made by UUSS is reliable and can be used as reference, Figure 7-C shows the results of the classification at the network-level. The CNN model again obtains high results with an accuracy of 96.72%. The percentage is comparable to the results presented for the training phase. The natural and anthropogenic events obtain a recall of 98.61% and 91.67%, respectively. This indicates that less than 2% of the natural events and around 8% of the anthropogenic events are misclassified. The *F1*-scores for the natural and anthropogenic events are 97.76% and 93.83%, respectively. This scores is close to the end-2021 French data application. These very similar results indicate that the CNN seems to discriminate effectively between seismic events in two *a priori* different contexts.

Figure 7-B represents the geographic distribution of the events misclassified by the model. On the 396 seismic events, only 13 are not correctly classified. This represents 3.28% of all the events. Moreover, we observe that 5 of the 9 misclassified anthropogenic events are located in areas where the nearest station is about 100 km away, which can explain this slightly higher percentage of misclassified anthropogenic events compared to the application on the French data.

3.3 Local post-seismic campaign

An earthquake of magnitude $M_W = 3.5$ (Sira *et Schlupp*, 2014) occurred near Vannes (NW France, see Figure 8-A for location) on 2013-11-21T09:53:00 (UTC). Six temporary stations were installed in a region of about 20 km around the main shock to document the aftershocks. Stations (three-component short-period velocimeters with a corner frequency of 1 Hz) were deployed for 26 days (between November 22 and December 17, 2013) and recorded ground displacement velocity with a sampling frequency of 100 sps. During this period, 48 events were detected, thanks to the STA/LTA approach (Haugmard, 2016), close to the temporary station, including 36 natural events and 12 quarry blasts. This represents 173 records of natural events and 66 records of anthropogenic events. The reference labels of these events were identified manually by observing the waveform, the location of the source, the origin time and the depth. Figure 8-A shows the location of the events detected during the post-seismic campaign. Although the dataset is greatly reduced compared to the two other applications, we observe that the 12 quarry blasts originate from the same quarry and that 25 of the 36 natural events form a cluster that corresponds to

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

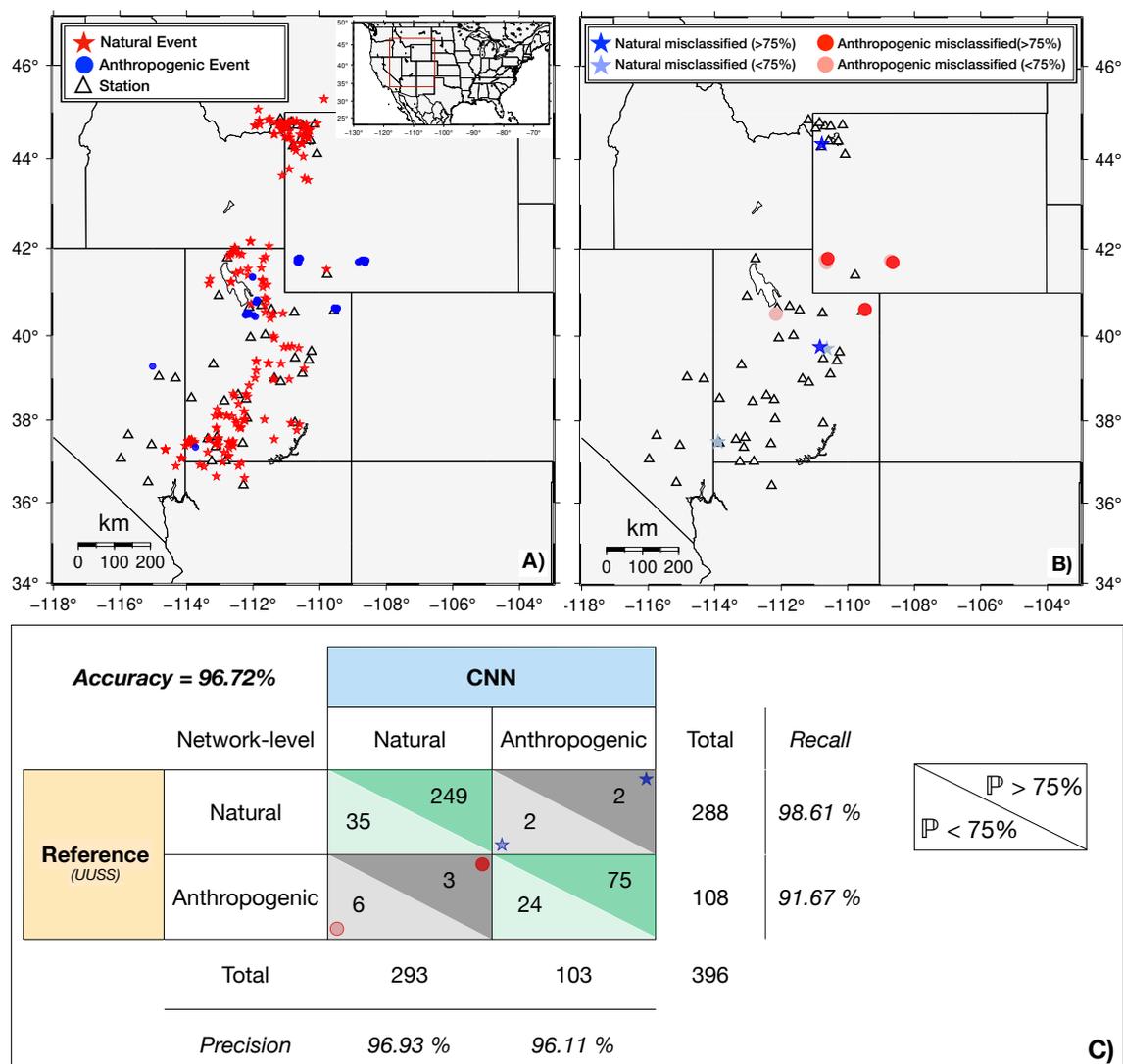


Figure 7: Application on events detected by University of Utah Seismograph Stations (UUSS) between January 1 and March 1, 2016. (A) Geographic distribution of all the seismic events used for the application (288 natural and 108 anthropogenic events). (B) Geographic distribution of the events misclassified by the model. Blue stars correspond to the events labeled as “natural” and classified as “anthropogenic” by the model. Red circles correspond to the events labeled “anthropogenic” and classified as “natural” by the model. The darker colors correspond to a prediction of more than 75% by the CNN and the lighter colors correspond to a prediction of less than 75% by the CNN. (C) Confusion matrix, for the differentiation between natural and anthropogenic events. The classification is done at the network level. \mathbb{P} is the probability of an event. The symbols correspond to the events located on (B).

the location of the main shock.

Figure 8-B shows the results of the classification at a network-level. The model obtained a remarkable score of 100% meaning that all events detected during this post-seismic campaign, have been correctly classified by the model. The high performance of the classifier can be partly explained by the small epicentral distances where the source effect is more important than that of the propagation in the waveform.

CHAPITRE 4. NOUVELLE APPROCHE POUR LA DÉTECTION ET LA DISCRIMINATION D'ÉVÉNEMENTS SISMQUES DE FAIBLE MAGNITUDE

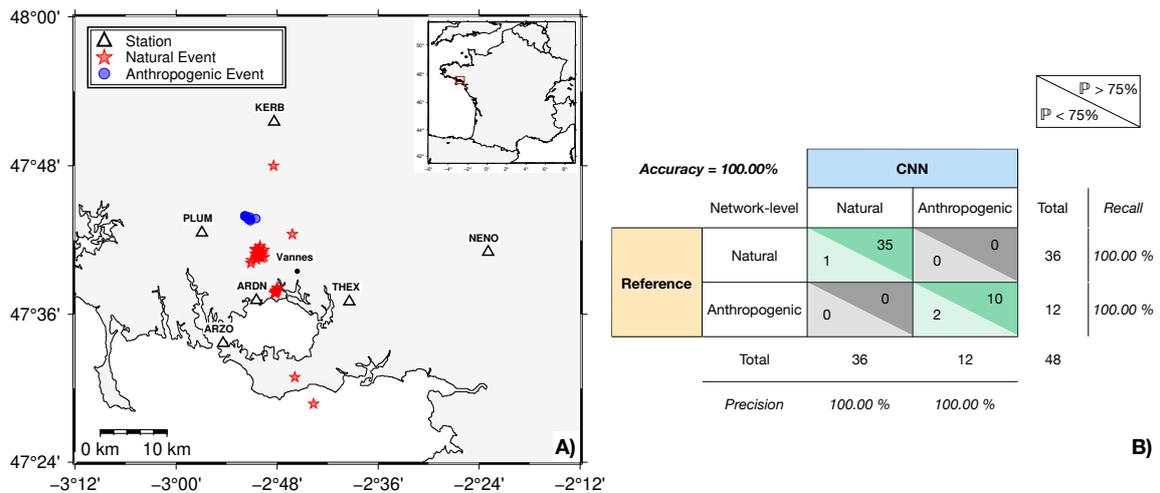


Figure 8: Application on temporary deployed seismic campaign installed near Vannes, in 2013, NW France. (A) Location of the natural (red stars) and anthropogenic events (blue circles) detected between November 22 and December 17, 2013. The triangles represent the temporary stations installed during the post-seismic campaign. (B) Confusion matrix, for the differentiation between the natural and anthropogenic events. The classification is done at the network level. The darker colors correspond to a prediction of more than 75% by the CNN and the lighter colors correspond to a prediction of less than 75% by the CNN.

4 Discussion

4.1 Comparison with other methods

4.1.1 Features or not features?

The use of features extracted from the seismic signal is a widely used technique in machine learning. [Renouard et al. \(2021\)](#), for instance, use an hybrid approach including both human expertise and machine learning algorithms to achieve this objective. They trained a Random Forest algorithm by extracting selected features from the seismic signals and by implementing them in a decision tree. Their algorithm is trained with data recorded by permanent and temporary stations from the AlpArray Seismic Network ([G. Hetényi et al., 2018](#)) between 2016 and 2019 in the Upper Rhine graben area, in northwestern Europe. Their model achieve an accuracy of about 96%. By applying our model only to the Upper Rhine graben area we obtain an accuracy of 97.8%. [Renouard et al. \(2021\)](#) claim that their approach uses human expertise to compute features and thus avoids the “black box” aspect of a classical neural network. However, the number of characteristics itself can be considered subjective and it is also possible that characteristics relevant for discrimination are omitted. Moreover, the use of algorithms such as Random Forest can lead to the use of many features that can be quite region specific and can thus limit the application of the algorithm to these specific areas. As a consequence, the workflow for such method can be quite complex. It is, for instance, required to first, pick the arrival times of the direct and refracted P and S -phases, then, to locate the event and, finally, to extract from the signal the 22 features used by the algorithm to perform its prediction. The CNN, presented in this study, is only based on three component spectrograms that include a seismic event while achieving equivalent accuracy on a similar data set.

4.1.2 What does the spectrogram-based CNN bring?

The use of spectrograms to discriminate natural from anthropogenic events seems to give high classification performances. From Figure 1, we can see that these spectrograms allow to highlight certain

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

characteristic patterns of each of the two classes: i) an energy that mainly impacts the low frequencies for the anthropogenic events (horizontal strips) and ii) an energy that mainly impacts the whole frequency range for the natural events (vertical strips). The use of square filters (TAB. 1) makes it possible to analyse the image without preferential direction and to highlight these vertical and horizontal bands.

We can wonder whether the use of seismograms allows better performance than with spectrograms. We tried to train a CNN with seismograms using the same CNN architecture but with a 1D version of the convolution layers. The input to the model corresponds to a matrix composed of three-components 60 s seismic traces. After ten training tests of this model the accuracy does not exceed 91.63%. It would appear that the use of simple seismic traces doesn't capture spectral patterns and that the use of spectrograms allows for better performance. This statement seems to be confirmed by [Liu *et al.* \(2021\)](#) which also used a CNN to distinguish tectonic from non-tectonic earthquakes. By training their neural network with 100 s three-component waveforms provided by the China Earthquake Network Center (CENC), they achieve 92% of accuracy.

4.1.3 Frequency range and time window width

[Linville *et al.* \(2019\)](#) also used a CNN to differentiate between natural and anthropogenic events. The architecture of their CNN is close to the one used in our study (TAB. 1) with four convolution layers with different number of filters and one fully connected layer. Their neural network is trained with UUSS data recorded between October 2012 and July 2017. They use spectrograms computed on 90 s traces, between 1 and 20 Hz, producing matrices of size $40 \times 48 \times 3$.

By training their CNN with a database of events detected in Utah, they achieve a recall of 99.1% for natural events and of 99.3% for anthropogenic events. These results are comparable but appear to be significantly better than those obtained with our approach. This may be due to the fact that to generate our model, we randomly select waveforms for the training, validation and test stages and not sets of traces associated with a given event. This subtlety necessarily leads to having traces of the same event in at least two of the three data sets used and then to apparently decreasing performances. When we apply our trained model to the UUSS data recorded from January to March 2016, we obtain an accuracy of 96.72% at the network-level with a recall of 98.61% and 91.67% for the natural and anthropogenic events, respectively. Taking into account only the events recorded by stations located at less than 100 km away, this accuracy rises to 97.95% with a recall of 98.61% and 96.11% for the natural and anthropogenic events, respectively. The difference in accuracy may be explained by several parameters like the size of the tested database, the training of the CNN on different geographical areas, the number of stations or the frequency range used for the classification. However, the use of spectrograms with a frequency content between 1 and 20 Hz limits the information usable by the CNN to extract features (such as vertical and horizontal stripes visible in Figure 1), but also limits the seismic noise associated to human activities (traffic, industry).

Moreover, the window used to compute the spectrograms was selected using the *P*-wave time arrival. Thanks to these elements, the training can be simplified and the performance can be higher. The use of spectrograms with a frequency content greater than 20 Hz may be important for the differentiation of low magnitude earthquakes. Moreover, thanks to the data pre-processing, our model can recognize the nature of the event independently of its position in the spectrogram. Different tests were performed by positioning the arrival of the *P*-wave between 5 and 30 s after the beginning of the signal and the prediction results remain unchanged. Our approach can therefore be applied independently of the chosen detection method (STA/LTA, template matching, etc.) by randomly selecting the beginning of the windows studied.

4.2 Impact of the training database on predictions

4.2.1 Effect of variability in the training database

The approach described in this study discriminates with a large efficiency natural from anthropogenic seismic events. For this purpose, we trained a CNN with about 50,000 seismic signals recorded in metropolitan France. Our neural network is thus trained on a relatively small dataset (compared to what can be found in the literature) though achieving high classification performances. To estimate the impact of the size of the database on the performance of the algorithm, we tested to train our CNN only with data coming from one of three distinct geographical areas in France (FIG. 3, TAB. 2).

Data from Eastern France (zone B, FIG. 3) are balanced in terms of ratios between natural and anthropogenic events (namely 2,612 natural events and 2,415 anthropogenic events, TAB. 2) and represent half of the total database (22,818 records out of 50,000 total records). Training only with data from the zone B and applying this trained model on the events detected in Metropolitan France between June 1st and November 1st 2021, we obtain an accuracy of 98.14%. The recall for the natural and anthropogenic events are 96.86% and 99.07%, respectively. These results are very similar to those obtained when we train our CNN with the full database (with a recall for natural and anthropogenic events of 98.86% and 99.12%, FIG. 6-C), indicating that, even with a substantially reduced database, performance remains good.

In contrast, the zones A (northwestern France) and C (Pyrenees) are unbalanced in terms of the proportion between natural and anthropogenic events (TAB. 2). Zone C contains more natural than anthropogenic events (namely 966 anthropogenic events and 2,149 natural events). If we train the CNN with this database and apply it on the French dataset, we obtain an accuracy of 97.5% with a recall for the natural and anthropogenic events of 96.38% and 98.28%. These results show that even with a lower proportion of anthropogenic events, performance remains high. We explain this by the fact that there is a sufficient variety (in terms of spectrogram patterns) of natural and anthropogenic events in the zone C to allow the CNN to be properly trained. On the other hand, the zone A presents significantly more anthropogenic than natural events (namely 255 natural events and 1,407 anthropogenic events). By performing the same procedure as for the other zones, we obtain an accuracy of 89.83% with a recall for the natural and anthropogenic events of 76.2% and 99.6% respectively. For this case, the classification performance collapses for the natural events. This can be explained by the fact that we probably do not have enough variety of natural event waveforms (TAB. 2) to achieve a robust classification. We performed a test to create a unbalanced database including events from the three zones. For this, 90% of the events in the database are anthropogenic and 10% are natural. By training our CNN with this database and applying the trained model on the French application data, we obtain similar results as when using only the zone A for training. It is therefore essential to have a large enough variety of natural events to obtain a good discrimination. However, the recall for the anthropogenic events remains excellent. This is likely due to the fact that, contrary to natural events, anthropogenic events have generally quite similar source functions and a limited range of magnitude (namely between 1.5 and 2.5). A more detailed interpretation would require a specific study and is beyond the scope of this work.

4.2.2 Effect of the training/validation-test ratio

Another drawback of the CNN training process is the selection of the proportion of the “training” database (50,000 records in this study) that is used for the training, validation and testing phases. The results presented in Section 3 were obtained with a proportion of 80% of the database for training and 20% of the database for validation and testing. We present in Figure 9 the performance of the algorithm using different training/validation-test ratios: 80/20 (original ratio), 60/40, 50/50, 40/60 and 20/80.

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

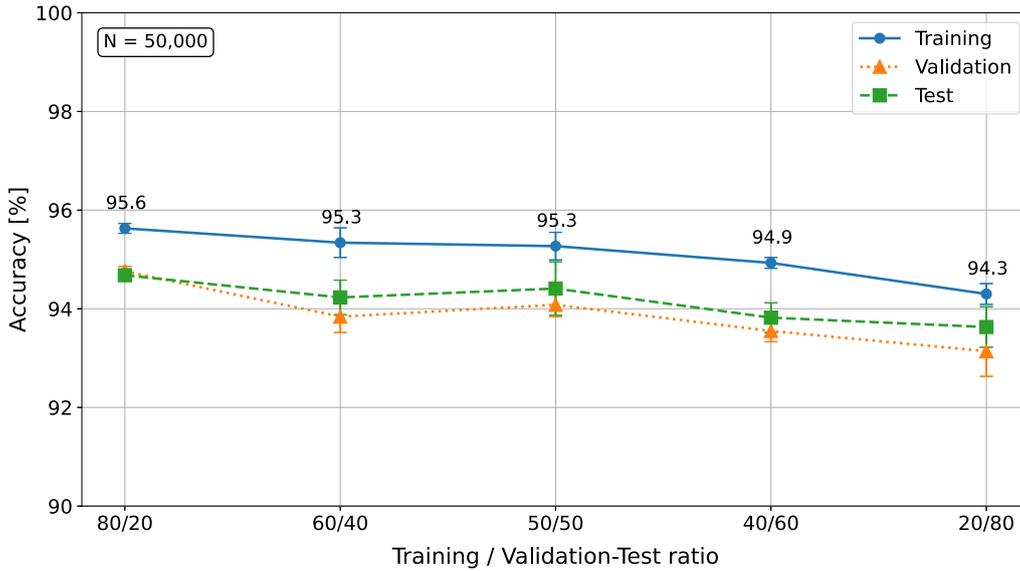


Figure 9: Performance of the CNN, at a station-level, using different Training/Validation-Test ratios. N corresponds to the total number of elements in the database. The accuracy obtained during the training, validation and test phases are represented in blue, orange and green respectively.

Figure 9 shows that our CNN reaches training accuracy of 95.6%, 95.3%, 95.3%, 94.9% and 94.3% for 80/20, 60/40, 50/50, 40/60 and 20/80 ratios, respectively. We thus observe a slight drop in accuracy with the decreasing in the amount of the initial database used for the training phase but, the performance remains high even when using only 20% of the database (10,000 records). We also observe that the CNN achieves better performances (for validation and test) for a ratio of 50/50 compared to the 60/40 ratio. The origin of this observation is not clear and would require more investigation. It is clear from Figure 9 that our initial choice of a 80/20 ratio seems to be optimal for training, validation and testing. We can conclude that the size of the training dataset does not appear to be the primary factor that controls the discrimination performance whereas the diversity of signals in the database seems to be a key factor.

4.3 Impact of catalog quality on predictions

4.3.1 Effect of seismic network geometry

The results presented in this study demonstrate the efficiency of the CNN for the classification of natural and anthropogenic events but also highlight that some events are misclassified by the neural network. Among other parameters, SNR of the data can be a factor impacting the performance of the model. A proxy to address this issue is the magnitude of the events, as the lower the energy the lower the SNR (for a given distance). Figure 10-A shows that, for $M_l < 1$ in the metropolitan France catalog, about 3% of the events are misclassified. Most of the events have $1 \leq M_l \leq 2$ (68.00% of the whole database) and, for these magnitudes, the accuracy reaches 98.45%. Finally, for $M_l > 2$, the classes of all the 106 earthquakes are correctly predicted. Figure 10-B, that shows the average SNR for each event, confirms that the discrimination is better in the highest SNR range. For $SNR < 3$, the accuracy collapses to 93.23% while if the SNR is higher than 15, the accuracy reaches 99.01%. Another parameter that impacts the performance of the model is the epicentral distance. Figure 10-C shows that the further the station is from the event, the lower the accuracy. For distances of more than 150 km, the accuracy drops to 92.63% while for distances of less than 50 km, the accuracy reaches 97.02%. This observation is also confirmed by the application on the data of the post-seismic campaign in NW France (FIG. 8). The

CHAPITRE 4. NOUVELLE APPROCHE POUR LA DÉTECTION ET LA DISCRIMINATION D'ÉVÉNEMENTS SISMIQUES DE FAIBLE MAGNITUDE

average distance between the stations and the events is 16.89 km (TAB. S3) and all these events were correctly classified by our method (FIG. 8-B). We therefore believe that all of these parameters are in competition to obtain a classification.

The lack of stations around an event also has an impact on the efficiency of the method. Figure 7-B that presents the geographical distribution of the misclassified events for the UUSS catalog, shows that 5 of the 9 misclassified anthropogenic events are located at more than 100 km from the nearest station. Taking into account the events only recorded by at least one station within 100 km, the accuracy increases from 96.72% to 97.95%. It can be noticed, in Figure 6-B, that most of the misclassified events are located at the borders of metropolitan France, and mainly in the zone C. Since we have only taken into account the stations of the FR network, for which distribution is limited to the French territory (FIG. 3), there is a lack of stations to the east of the French border. We performed a test on 5 natural events, located in the zone B, that were not correctly classified by the CNN. The stations network CH ([Swiss Seismological Service \(SED\), 1983](#), located in Switzerland) and GU ([University of Genoa, 1967](#), located in Italy) were used to refine the prediction. These new stations are located within 50 km from the events, whereas the stations used in the FR network are more than 60 km away. The addition of these stations for the classification allows to correctly classify these 5 events. This indicates that the epicentral distance is an important parameter affecting discrimination as shown in Figure 10-C. In addition, increasing the number of stations around event can improve the variability of waveforms induced by their radiation pattern.

4.3.2 Effect of the mislabeled events in the database

Finally, a misclassified event can also be simply due to mislabeled events in the database. During the manual inspection of the data for the generation of the confusion matrices we were indeed able to identify some mislabeled events in the database. In Figure 11-A, we present the trace and spectrogram of an event detected on 2019-08-02T20:09:00 UTC labeled as a “quarry blast” and that was classified as a “natural event” at 99.4% by the CNN. This event is undoubtedly a natural event because no quarry blast occurs at such a late hour (10 pm local time). In addition, the spectral characteristics of natural event presented in Section 2.1 are clearly observe here (FIG. 11-A) two seismic wave arrivals that form a vertical band in the spectrogram. In contrast, Figure 11-B shows a signal from a seismic event that occurred on 2021-08-03T09:50:00 UTC. This event is tagged as “natural” in the database but classified at 99.9% as “anthropogenic” by our trained model. The spectrogram of this event is characterised by the impulsive arrival of the *P*-wave and by two horizontal strips along the time axis at frequency between 1 and 30 Hz. Thanks to the time of origin of the event (11:50 am, local time) and to the characteristics of the seismic signal, we can reasonably consider that this event is correctly classified as “anthropogenic” by the CNN. Those examples illustrate the importance of having trusted database for training but shows that a well trained CNN can help to detect, *a posteriori*, inconsistencies in the databases and would allow the reassessment of the associated labels.

5 Conclusion

The objective of this study is to develop a method for automatic discrimination of natural and anthropogenic events in order to reduce the data processing time. Based on recent existing methods, we have developed a CNN to perform this discrimination using spectrograms. We have constituted a training database composed of BCSF-RéNaSS events detected in metropolitan France. By adapting the pre-processing of the data, our CNN is able to recognize the nature of the event independently of the position of the earthquake in the signal and thus to discriminate the events directly after the detection phase. This trained model achieves high classification performance with, for example, an accuracy of 98.18%

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

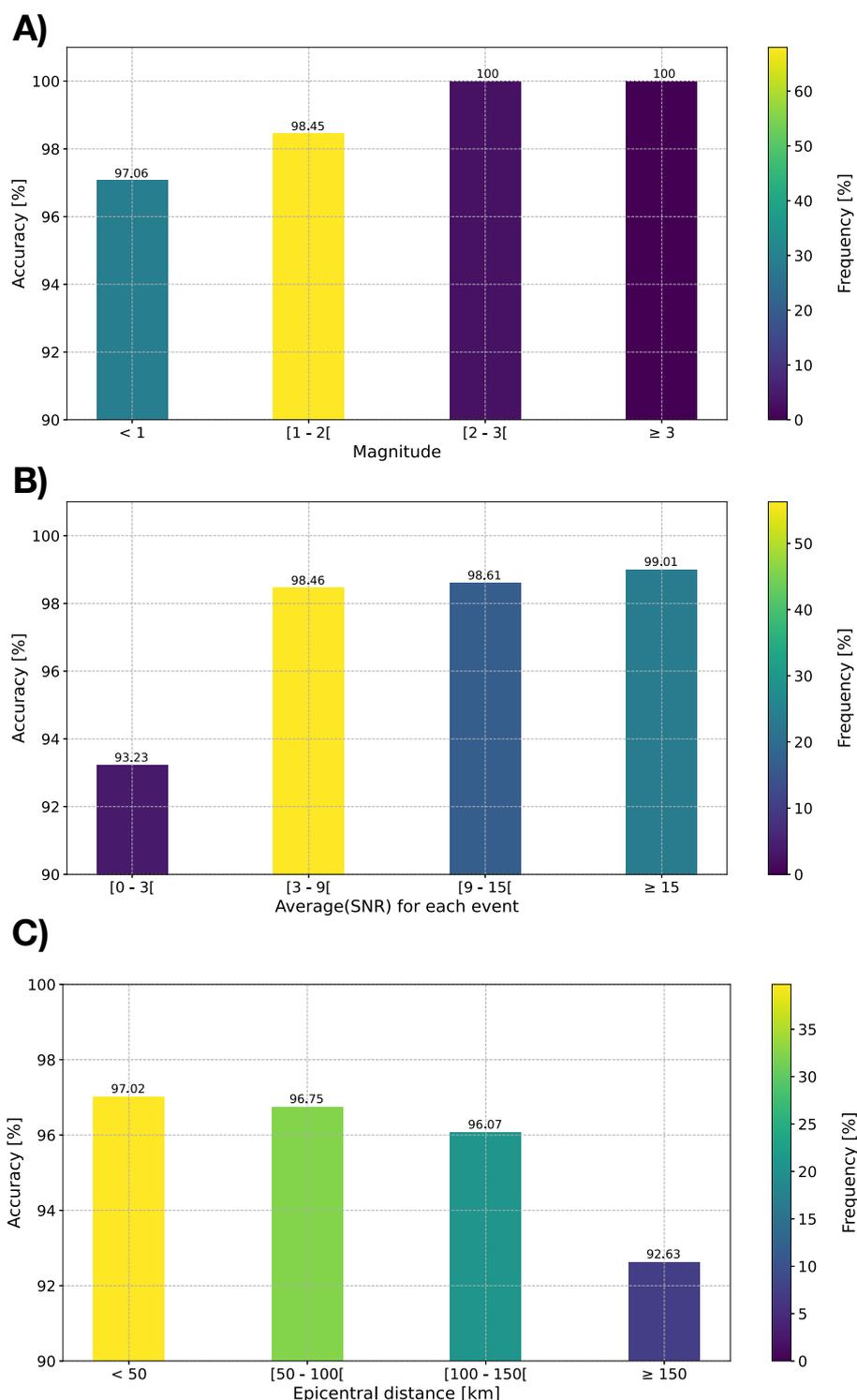


Figure 10: Discrimination performance as a function of magnitude, *SNR* and epicentral distance for events detected in metropolitan France between June 1st and November 1st 2021. (A) Relationship between CNN performance and event magnitude with accuracy marked at the top of the columns. The color corresponds to the proportion of events for each magnitude range. (B) Relationship between the CNN performance and the average *SNR* of the events. The color corresponds to the proportion of events for each *SNR* range. (C) Relationship between CNN performance and epicentral distance with accuracy marked at the top of the columns. The color corresponds to the proportion of records for each distance range.

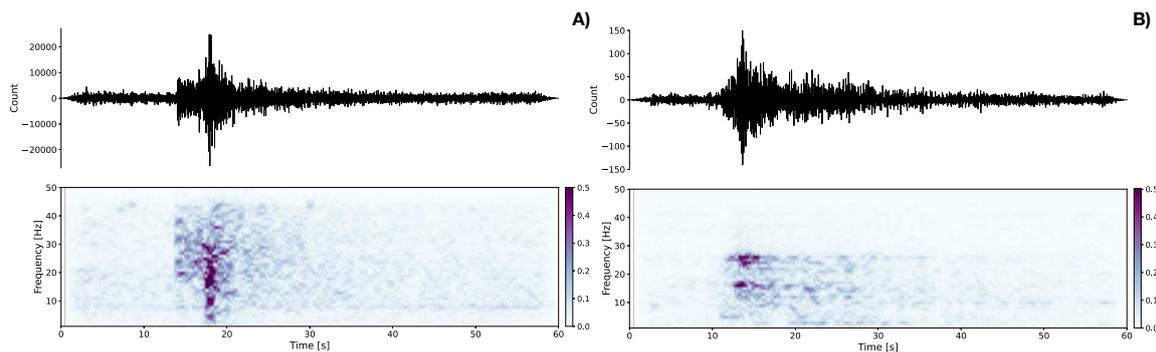


Figure 11: (A) Spectrogram of an earthquake, detected on August 2, 2019 at 22:09, local time, recorded by the VIEF station, at a distance of 25.16 km from the source, labeled "anthropogenic" and classified by the CNN at 99.4% as "natural". (B) Spectrogram of an earthquake, detected on August 3, 2021 at 11:50 am, local time, recorded by the LOUF station, at a distance of 65.52 km from the source, labeled "natural" and classified by the CNN at 99.9% as "anthropogenic".

when applied to independent French data. By applying the model on data from different geographical areas and geological contexts, we show the versatility of our approach. For the USSS application, the model achieved an accuracy of 96.72%. These applications to other geographical areas have shown the robustness of the model in recognizing characteristic patterns in different databases. We observed a certain redundancy of features present in the spectrograms involving a continuous energy in time and discontinuous in frequency for anthropogenic events, represented by a horizontal linearity on the spectrograms. Inversely, for natural events, we observe a discontinuous energy in time and continuous in frequency, represented by a vertical linearity on the spectrograms. These features highlight the relevance of using square filters (TAB. 1) in the CNN, to not focus on one direction in the spectrogram.

This approach can be used as an aid to discriminate, in real time, the detected seismic signals. We can propose an application with human interaction, where the analyst can check only the events with lower probabilities, for example, less than 75%. Moreover, we could highlight mislabeled events present, for example, in the BCSF-RéNaSS bulletin. These mislabeled events can be reclassified by the CNN and thus allow to constitute a catalog as reliable as possible.

Acknowledgments

The authors warmly thank the anonymous reviewers for their constructive criticism and to the editor for their work. This work was only made possible by the manual discrimination performed by BCSF-RéNaSS operators (R. Dretzen, S. Lambotte and M. Grunberg, <http://renass.unistra.fr>).

Résif-Epos is a Research Infrastructure (RI) managed by the CNRS-INSU. It is a consortium of eighteen French research organisations and institutions, included in the roadmap of the Ministry of Higher Education, Research and Innovation. Résif-Epos RI is also supported by the Ministry of Ecological Transition.

Data Availability

This work included data from the permanent seismic networks operated by the French seismological and geodetic network (RESIF), by University of Utah Seismograph Stations (USSS) and temporary stations from a post-seismic campaign deployment near the city of Vannes in 2013. Catalogs and data from the French National Observing Service (BCSF-RéNaSS) and the USSS are available using an International

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

Federation of Digital Seismograph Networks (FDSN) protocol.

The discrimination code is available on Git (DOI: 10.5281/zenodo.7064191). The algorithms are written with Python and data processing is based on Python libraries Obspy ([Krischer *et al.*, 2015](#)), Scipy ([Virtanen *et al.*, 2020](#)) and Numpy ([Harris *et al.*, 2020](#)). We use Tensorflow ([Abadi *et al.*, 2015](#)) for implementations and iterate model parameters using Keras ([Chollet, 2015](#)). Our figures were generated using Matplotlib ([Hunter, 2007](#)) and The Generic Mapping Tools (GMT) ([Wessel *et al.*, 2019](#)).

References

- Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu et X. Zheng (2015). “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”. Software available from [tensorflow.org](https://www.tensorflow.org).
- Allen, R. (1982). “Automatic phase pickers: Their present use and future prospects”. *Bulletin of the Seismological Society of America* 72(6B), S225–S242.
- Allmann, B. P., P. M. Shearer et E. Hauksson (2008). “Spectral discrimination between quarry blasts and earthquakes in southern California”. *Bulletin of the Seismological Society of America* 98(4), 2073–2079.
- Arabasz, W. J., R. B. Smith et W. D. Richins (1980). “Earthquake studies along the Wasatch Front, Utah: Network monitoring, seismicity, and seismic hazards”. *Bulletin of the Seismological Society of America* 70(5), 1479–1499.
- Arizona Geological Survey (2007). “Arizona Broadband Seismic Network”.
- Baumgardt, D. R. et G. B. Young (1990). “Regional seismic waveform discriminants and case-based event identification using regional arrays”. *Bulletin of the Seismological Society of America* 80(6B), 1874–1892.
- Beucler, E., M. Bonnin, C. Hourcade, B. V. Vliet-Lanoë, C. Perrin, L. Provost, A. Mocquet, J. Battaglia, L. Geoffroy, P. Steer, B. L. Gall, J.-M. Douchain, D. Fligiel, P. Gernigon, B. Delouis, J. Perrot, S. Mazzotti, G. Mazet-Roux, S. Lambotte, M. Grunberg, J. Vergne, C. Clément, E. Calais, J. Deverchère, L. Longuevergne, A. Duperret, C. Roques, T. Kaci et C. Authemayou (2021). “Characteristics and possible origins of the seismicity in northwestern France”. *Comptes Rendus. Géoscience* 353(S1), 53–77.
- Cara, M., Y. Cansi, A. Schlupp, P. Arroucau, N. Béthoux, É. Beucler, S. Bruno, M. Calvet, S. Chevrot, A. Deboissy, B. Delouis, M. Denieul, A. Deschamps, C. Doubre, J. Fréchet, S. Godey, O. Golle, M. Grunberg, J. Guilbert, M. Haugmard, L. Jenatton, S. Lambotte, D. Leobal, C. Maron, V. Mendel, S. Merrer, M. Macquet, A. Mignan, A. Mocquet, M. Nicolas, J. Perrot, B. Potin, O. Sanchez, J.-P. Santoire, O. Sèbe, M. Sylvander, F. Thouvenot, J. van der Woerd et K. van der Woerd (2015). “SI-Hex: a new catalogue of instrumental seismicity for metropolitan France”. *Bulletin de la Société Géologique de France* 186(1), 3–19.
- Chollet, F. (2015). “Keras” GitHub. <https://github.com/fchollet/keras>.
- Del Pezzo, E., A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini et S. Scarpetta (2003). “Discrimination of earthquakes and underwater explosions using neural networks”. *Bulletin of the Seismological Society of America* 93(1), 215–223.
- Doubre, C., M. Meghraoui, F. Masson, S. Lambotte, H. Jund, M. B. de Berc et M. Grunberg (2021). “Seismotectonics in Northeastern France and neighboring regions”. *Comptes Rendus. Géoscience* 353(S1), 153–185.
- Dysart, P. S. et J. J. Pulli (1990). “Regional seismic event classification at the NORESS array: seismological measurements and the use of trained neural networks”. *Bulletin of the Seismological Society of America* 80(6B), 1910–1933.

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

- Favretto-Cristini, N., F. Wang, T. Garlan, O. Morio, A. Deschamps, d. ambrois, D. Mercerat, X. Demoulin, E. Beucler, M. Bonnin, P. Cristini et V. Monteiller (2020, December). “Seismo-acoustic wave propagation in the Rade of Hyères (France) generated by counter-mining of explosive devices : risk assessment from the comparison between numerical simulations and real experiments”. In *Forum Acusticum*, Lyon, France, pp. 2115–2120.
- A. Géron (2019). “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems”. *O’Reilly Media, Inc.*
- Gibbons, S. J. et F. Ringdal (2006). “The detection of low magnitude seismic events using array-based waveform correlation”. *Geophysical Journal International* 165(1), 149–166.
- Gitterman, Y., V. Pinsky et A. Shapira (1998). “Spectral classification methods in monitoring small local events by the Israel seismic network”. *Journal of seismology* 2(3), 237–256.
- Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke et T. E. Oliphant (2020, September). “Array programming with NumPy”. *Nature* 585(7825), 357–362.
- Haugmard, M. (2016). *Détermination non-linéaire des paramètres hypocentaux et structuraux: application à la sismicité intracontinentale du Massif armoricain*. Ph. D. thesis, Nantes.
- G. Hetényi, I. Molinari, J. Clinton, G. Bokelmann, I. Bondár, W-C. Crawford, J-X. Dessa, and C. Doubre, W. Friederich, F. Fuchs, D. Giardini, Z. Gráczér, M. Handy, M. Herak, Y. Jia, E. Kissling, H. Kopp, M. Korn, L. Margheriti, T. Meier, M. Mucciarelli, A. Paul, D. Pesaresi, C. Piromallo, T. Plenefisch, J. Plomerová, J. Ritter, G. Rümpker, V. Šipka, D. Spallarossa, C. Thomas, F. Tilmann, J. Wassermann, M. Weber, Z. Wéber, V. Wesztergom, M. Živčić, R. Abreu, I. Allegretti, M-T. Apoloner, C. Aubert, S. Besançon, M. Bès de Berc, D. Brunel, M. Capello, M. Čarman, A. Cavaliere, J. Chèze, C. Chiarabba, G. Cougoulat, L. Cristiano, T. Czifra, E. D’Alema, S. Danesi, R. Daniel, A. Dannowski, I. Dasović, A. Deschamps, S. Egdorf, T. Fiket, K. Fischer, S. Funke, A. Govoni, G. Gröschl, S. Heimers, B. Heit, D. Herak, J. Huber, D. Jarić, P. Jedlička, H. Jund, S. Klíngen, B. Klotz, P. Kolínský, J. Kotek, L. Kühne, K. Kuk, D. Lange, J. Loos, S. Lovati, D. Malengros, C. Maron, X. Martin, M. Massa, F. Mazzarini, L. Métral, M. Moretti, H. Munzarová, A. Nardi, J. Pahor, C. Péquegnat, F. Petersen, D. Piccinini, S. Pondrelli, S. Prevolnik, R. Racine, M. Régnier, M. Reiss, S. Salimbeni, M. Santulin, W. Scherer, S. Schippkus, D. Schulte-Kortnack, S. Solarino, K. Spieker, J. Stipčević, A. Strollo, B. Süle, G. Szanyi, E. Szűcs, M. Thorwart, S. Ueding, M. Valloccchia, L. Vecsey, R. Voigt, C. Weidle, G. Weyland, S. Wiemer, F. Wolf, D. Wolyniec, T. Zieke, AlpArray Seismic Network Team, ETHZ-SED Electronics Lab, AlpArray OBS Cruise Crew, AlpArray Working Group (2018). “The AlpArray seismic network: a large-scale European experiment to image the Alpine orogen”. *Surveys in geophysics* 39, 1009–1033.
- Hunter, J. D. (2007). “Matplotlib: A 2D graphics environment”. *Computing in Science & Engineering* 9(3), 90–95.
- James, G., D. Witten, T. Hastie et R. Tibshirani (2013). *An introduction to statistical learning*, Volume 112. Springer.
- Koper, K. D., J. C. Pechmann, R. Burlacu, K. L. Pankow, J. Stein, J. M. Hale, P. Roberson et M. K. McCarter (2016). “Magnitude-based discrimination of man-made seismic events from naturally occurring earthquakes in Utah, USA”. *Geophysical Research Letters* 43(20), 10–638.

CHAPITRE 4. NOUVELLE APPROCHE POUR LA DÉTECTION ET LA DISCRIMINATION D'ÉVÉNEMENTS SISMIQUES DE FAIBLE MAGNITUDE

- Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron et J. Wassermann (2015). “ObsPy: A bridge for seismology into the scientific Python ecosystem”. *Computational Science & Discovery* 8(1), 014003.
- Larroque, C., S. Baize, J. Albaric, H. Jomard, J. Trévisan, M. Godano, M. Cushing, A. Deschamps, C. Sue, B. Delouis, B. Potin, F. Courboux, M. Régnier, D. Rivet, D. Brunel, J. Chèze, X. Martin, C. Maron et F. Peix (2021). “Seismotectonics of southeast France: from the Jura mountains to Corsica”. *Comptes Rendus. Géoscience* 353(S1), 105–151.
- Levandowski, W., Herrmann, R. B., Briggs, R., Boyd, O., et Gold, R. (2018). “An updated stress map of the continental United States reveals heterogeneous intraplate stress.”. *Nature Geoscience* 11(6), 433–437.
- Li, Z., M.-A. Meier, E. Hauksson, Z. Zhan et J. Andrews (2018). “Machine learning seismic wave discrimination: Application to earthquake early warning”. *Geophysical Research Letters* 45(10), 4773–4779.
- Linville, L., K. Pankow et T. Draelos (2019). “Deep learning models augment analyst decisions for event discrimination”. *Geophysical Research Letters* 46(7), 3643–3651.
- Liu, X., T. Ren, H. Chen et Y. Chen (2021). “Classification of tectonic and non-tectonic seismicity based on convolutional neural network”. *Geophysical Journal International* 224(1), 191–198.
- Mayor, J., P. Traversa, M. Calvet et L. Margerin (2018). “Tomography of crustal seismic attenuation in Metropolitan France: implications for seismicity analysis”. *Bulletin of Earthquake Engineering* 16(6), 2,195–2,210.
- McLaughlin, K. L., J. L. Bonner et T. Barker (2004). “Seismic source mechanisms for quarry blasts: modelling observed Rayleigh and Love wave radiation patterns from a Texas quarry”. *Geophysical Journal International* 156(1), 79–93.
- Meier, M.-A., Z. E. Ross, A. Ramachandran, A. Balakrishna, S. Nair, P. Kundzicz, Z. Li, J. Andrews, E. Hauksson et Y. Yue (2019). “Reliable real-time seismic signal/noise discrimination with machine learning”. *Journal of Geophysical Research: Solid Earth* 124(1), 788–800.
- Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang et G. C. Beroza (2020). “Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking”. *Nature communications* 11(1), 1–12.
- Pardo, E., C. Garfias et N. Malpica (2019). “Seismic phase picking using convolutional networks”. *IEEE Transactions on Geoscience and Remote Sensing* 57(9), 7086–7092.
- Pyle, M. L. et W. R. Walter (2019). “Investigating the effectiveness of P/S amplitude ratios for local distance event discrimination”. *Bulletin of the Seismological Society of America* 109(3), 1071–1081.
- Renouard, A., A. Maggi, M. Grunberg, C. Doubre et C. Hibert (2021). “Toward False Event Detection and Quarry Blast versus Earthquake Discrimination in an Operational Setting Using Semiautomated Machine Learning”. *Seismological Research Letters* 108 (6) 3725–3742.
- RESIF (1995). “RESIF-RLBP French Broad-band network, RESIF-RAP strong motion network and other seismic stations in metropolitan France. RESIF – Réseau Sismologique et géodésique Français. doi:10.15778/resif.fr”.
- Ross, Z. E., M.-A. Meier, E. Hauksson et T. H. Heaton (2018). “Generalized Seismic Phase Detection with Deep Learning”. *Bulletin of the Seismological Society of America* 108(5A), 2894–2901.

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

- Sira, C. et A. Schlupp (2014). “Séisme de Vannes (Morbihan) du 20 novembre 2013 (4,6 ML), synthèse sismologique et étude macrosismique”. *rapport du Bureau Central Sismologique Français, BCSF2014-R2, p.31, ULP/EOST-CNRS/INSU, Strasbourg.*
- Su, F., K. Aki et N. Biswas (1991). “Discriminating quarry blasts from earthquakes using coda waves”. *Bulletin of the Seismological Society of America* 81(1), 162–178.
- Swiss Seismological Service (SED) At ETH Zurich (1983). “National Seismic Networks of Switzerland”.
- Sylvander, M., A. Rigo, G. Sénéchal, J. Battaglia, S. Benahmed, M. Calvet, S. Chevrot, J.-M. Douchain, F. Grimaud, J. Letort et H. Pauchet (2021). “Seismicity patterns in southwestern France”. *Comptes Rendus. Géoscience* 353(S1), 79–104.
- Tharwat, A. (2020). “Classification assessment methods”. *Applied Computing and Informatics.*
- Tibi, R., K. D. Koper, K. L. Pankow et C. J. Young (2018). “Depth discrimination using Rg-to-Sg spectral amplitude ratios for seismic events in Utah recorded at local distances”. *Bulletin of the Seismological Society of America* 108(3A), 1355–1368.
- University of Genoa (1967). “Regional Seismic Network of North Western Italy”.
- University of Nevada, Reno (1971). “Nevada Seismic Network”.
- University of Utah (1962). “University of Utah Regional Seismic Network”.
- University of Utah (1983). “Yellowstone National Park Seismograph Network”.
- Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt et SciPy 1.0 Contributors (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods* 17, 261–272.
- Wang, R., B. Schmandt et E. Kiser (2020). “Seismic discrimination of controlled explosions and earthquakes near Mount St. Helens using P/S ratios”. *Journal of Geophysical Research: Solid Earth* 125(10), e2020JB020338.
- Wessel, P., J. Luis, L. Uieda, R. Scharroo, F. Wobbe, W. Smith et D. Tian (2019). “The generic mapping tools version 6”. *Geochemistry, Geophysics, Geosystems* 20(11), 5556–5564.
- Woollam, J., A. Rietbrock, A. Bueno et S. De Angelis (2019). “Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network”. *Seismological Research Letters* 90(2A), 491–502.
- Yoon, C. E., O. O’Reilly, K. J. Bergen et G. C. Beroza (2015). “Earthquake detection through computationally efficient similarity search”. *Science advances* 1(11), e1501057.
- Zhu, W. et G. C. Beroza (2019). “PhaseNet: a deep-neural-network-based seismic arrival-time picking method”. *Geophysical Journal International* 216(1), 261–273.

New CNN based tool to discriminate anthropogenic from natural low magnitude seismic events

Céline Hourcade¹, Mickaël Bonnin¹, Éric Beucler¹

¹ Nantes Université, Université d'Angers, Le Mans Université, CNRS, UMR-6112, Laboratoire de planétologie et géosciences, F-44000 Nantes, France

E-mail: celine.hourcade@etu.univ-nantes.fr

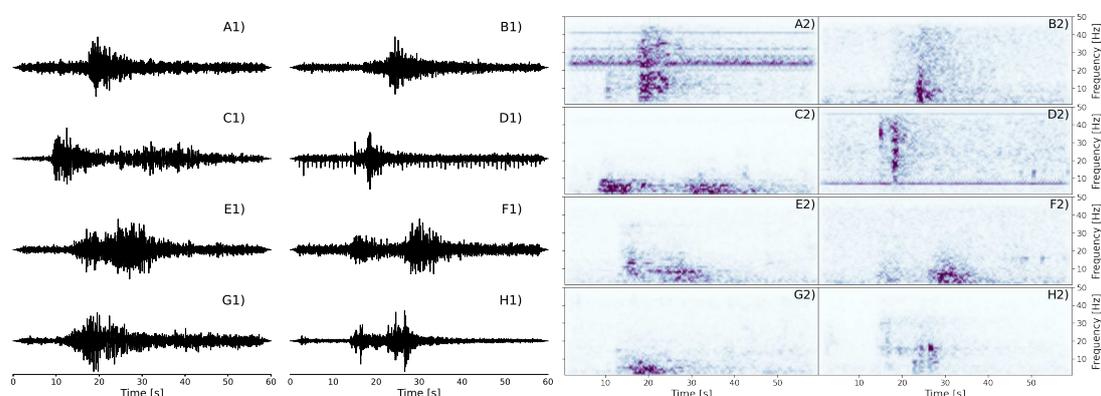


Figure S1: Examples of four natural events (A, B, D, F) and four anthropogenic events (C, E, G, H), recorded by the FR network stations, with the associated spectrograms. The seismic signals correspond to the normalized raw data of the vertical component. The spectrograms are normalized by their maximum. All technical details of the events are in Table S2.

Table S1: Useful information about the records shown in Figure 1, where: d is the source-station distance, z is the depth of the event, SNR the signal to noise ratio and ML_v the local magnitude on the vertical component. The acronyms eq and qb stand for natural event (earthquake) and quarry blast respectively.

	Time (UTC)	Station	d (km)	SNR	ML_v	Latitude (°)	Longitude (°)	z (km)	Label
a)	2021-10-30T12:03:38	GUEF	91.1	11.16	2.1	47.08	-2.46	14.00	eq
b)	2021-10-29T21:43:22	LGIF	51.7	11.27	1.8	46.73	-0.86	2.0	eq
c)	2019-12-12T12:02:23	MLS	60.2	7.68	1.6	42.78	1.79	0.0	qb
d)	2021-10-29T21:43:22	CRNF	47.5	5.56	1.8	46.73	-0.86	2.0	eq
e)	2019-12-11T12:13:29	MELF	86.5	9.62	1.6	42.79	1.81	0.0	qb
f)	2021-10-29T21:43:22	BOUF	75.8	2.85	1.8	46.73	-0.86	2.0	eq
g)	2021-10-29T00:12:04	AGO	7.3	97.74	1.3	46.11	3.08	5.0	eq
h)	2019-12-11T12:13:29	MLS	61.2	13.10	1.6	42.79	1.81	0.0	qb

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

Supplementary Materials

Table S2: Useful information about the records shown in Figure S1, where: d is the source-station distance, z is the depth of the event, SNR the signal to noise ratio and ML_v the local magnitude on the vertical component. The acronyms eq, qb and exp stand for natural event (earthquake), quarry blast and explosion respectively.

	Time (UTC)	Station	d (km)	SNR	ML_v	Latitude (°)	Longitude (°)	z (km)	Label
a)	2021-10-27T19:28:15	LOUF	67	1.83	1.7	48.09	-1.06	16.0	eq
b)	2021-08-16T02:55:53	LABF	44	2.34	0.9	43.10	-0.46	8.0	eq
c)	2021-10-20T08:15:09	JAUF	169	11.38	2.1	43.01	6.00	0.0	exp
d)	2021-10-27T18:49:16	VIEF	20	2.99	0.1	43.00	0.22	11.0	eq
e)	2021-06-01T09:33:07	BSTF	85	2.55	1.6	43.06	5.91	0.0	exp
f)	2021-10-19T18:55:47	TRIGF	104	2.77	0.8	43.87	7.69	0.0	eq
g)	2021-08-09T09:16:24	SAOF	21	3.79	1.4	43.82	7.41	0.0	qb
h)	2021-10-26T09:00:10	COLF	67	6.53	1.4	45.73	4.49	0.0	qb

Table S3: Statistical informations of the three databases used in the application (Section 3). Standard deviations are denoted σ . No magnitude is available for Vannes dataset.

Local magnitude on vertical component (ML_v)										
	Natural					Anthropogenic				
	min	max	mean	median	σ	min	max	mean	median	σ
Resif	0.004	2.99	0.92	0.83	0.52	0.17	2.98	1.41	1.43	0.33
UUSS	-0.08	4.28	1.36	1.36	0.89	0.88	3.02	1.86	1.83	0.43
Epicentral distance (km)										
	Natural					Anthropogenic				
	min	max	mean	median	σ	min	max	mean	median	σ
Resif	0.12	420.1	62.00	46.18	52.31	0.74	384.43	81.39	81.54	44.16
UUSS	0.92	112.31	56.39	59.57	31.19	5.75	110.22	58.50	61.58	27.41
Vannes	3.13	42.31	16.19	14.13	9.15	6.83	52.20	18.77	16.04	9.81

4.3.2 Complément : Discrimination à l'échelle de la station et du réseau

Dans cette section, nous approfondissons certains points abordés dans l'article [Hourcade et coll. \(2023\)](#). Notre attention se porte principalement sur l'utilisation d'un réseau de stations pour optimiser les performances de classification. La Figure 6.B de l'article illustre la localisation des événements mal classés par le réseau de neurones. Nous observons une répartition de ces événements au niveau des frontières de la France métropolitaine. Cette répartition peut être attribuée à une faible couverture de stations autour de ces événements, ce qui peut avoir un impact sur les performances de classification.

Pour rappel, le processus de classification consiste à appliquer le CNN entraîné à chaque enregistrement de chaque station, ce qui est communément appelé une prédiction au niveau de la station (*station-level*). Toutefois, pour obtenir une classification à l'échelle du réseau (*network-level*), nous devons combiner l'ensemble des prédictions provenant de chaque station.

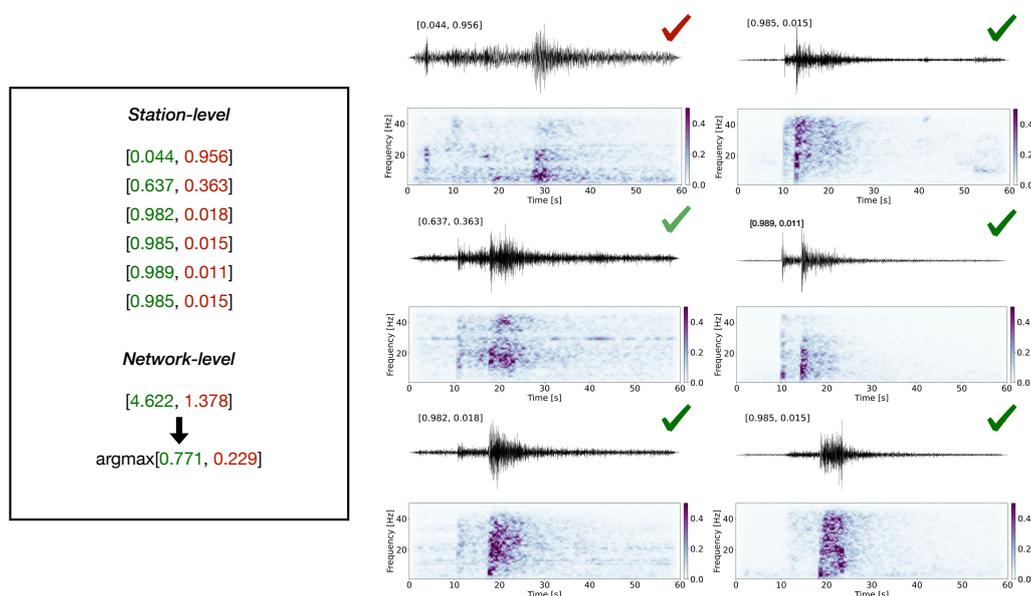


FIGURE 4.13 – Exemple de classification *station-level* et *network-level*. L'événement sismique a été enregistré par six stations. Pour chaque enregistrement nous avons une prédiction. Ces prédictions sont ensuite sommées et une moyenne est réalisée afin d'avoir une prédiction pour l'événement. La couleur des symboles en haut à droite des signaux correspond à la classification du signal. En rouge, le signal est classé comme « anthropique » et en vert, il est classé comme « naturel ». Plus la couleur du symbole est foncée plus la probabilité pour la classe est forte.

4.3. DISCRIMINATION ENTRE LES ÉVÉNEMENTS NATURELS ET ANTHROPIQUES

Nous utilisons donc l'ensemble des prédictions individuelles effectuées pour chaque enregistrement de chaque station afin de prendre une décision globale concernant la classification de l'événement. Cette approche permet de tirer parti de l'ensemble du réseau de stations et d'obtenir une meilleure compréhension des caractéristiques sismiques de l'événement. La Figure 4.13 montre un exemple de passage entre *station-level* et *network-level* à partir de six enregistrements de stations.

Pour investiguer le rôle de la géométrie du réseau dans les résultats de classification, nous avons effectué un test sur cinq événements situés à l'est de la France. Ces cinq événements ont été labellisés comme « naturel », mais le réseau de neurones les a classés comme « anthropique ». Dans le but d'améliorer la couverture azimutale de ces événements, nous avons ajouté les enregistrements des stations suisses (réseau CH, Swiss Seismological Service (SED) At ETH Zurich, 1983) et italiennes (réseau GU, University of Genoa, 1967) (Figure 4.14.A). Grâce à ces stations supplémentaires, ces cinq événements sont maintenant tous correctement classés par le réseau de neurones.

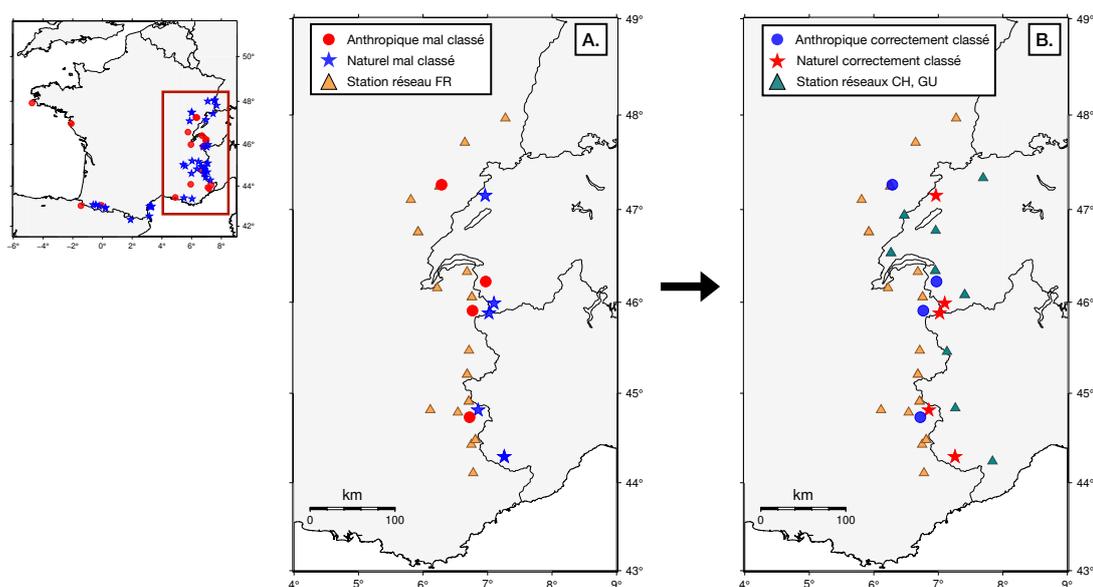


FIGURE 4.14 – Carte de couverture de stations autour de neuf événements sismiques. (A) Exemple d'événements mal classés en utilisant seulement le réseau de station FR. (B) Ces mêmes événements correctement classés en utilisant les réseaux de station FR, GU et CH.

Le même test est effectué sur quatre événements labellisés comme « anthropique », mais classés comme « naturel » par le réseau de neurones (Figure 4.14.B). Pour chacun de ces événements, les probabilités pour chacune des deux classes sont d'environ 50-50. En ajoutant les stations des réseaux GU et CH, ces probabilités passent à environ 90-10 en faveur de la classe « anthropique ». Ces résultats renforcent l'idée que la couverture

des stations est un facteur déterminant dans les performances de classification des événements sismiques.

4.4 Lecture des temps d'arrivée des ondes

Comme nous l'avons montré, notre méthode nous permet de détecter et de récupérer les événements sismiques naturels au sein du signal sismique continu. Afin de localiser ces événements, il est nécessaire de pointer les temps d'arrivée des ondes.

L'enregistrement du mouvement du sol par un sismomètre est le résultat de la convolution de plusieurs fonctions. Cela comprend la fonction source du séisme, qui décrit comment l'énergie est libérée, le diagramme de rayonnement, la dispersion, l'atténuation, les interférences avec les autres phases, le bruit sismique à l'emplacement du sismomètre et les caractéristiques propres au système d'enregistrement. En conséquence, ces différentes composantes se superposent et peuvent créer des formes d'onde extrêmement complexes, notamment dans le cas d'enregistrements locaux et régionaux à haute fréquence.

Les différents types d'ondes sismiques peuvent se manifester sur une période de temps très courte, rendant leur distinction plus difficile. Certains types d'ondes peuvent également se chevaucher, ce qui complique davantage leur identification précise lors de l'analyse des enregistrements sismiques. Par conséquent, il est essentiel de comprendre et de prendre en compte ces facteurs lors de l'interprétation des données sismiques afin de caractériser correctement les événements sismiques et de localiser précisément leurs sources.

Lors de l'identification des phases sismiques¹, il est important de respecter la cohérence à la fois au niveau de chaque station individuelle et au niveau d'un réseau de stations. Si l'arrivée des ondes P et S est proche, alors la distance hypocentrale est faible et aucune phase réfractée n'est enregistrée. À de plus grandes distances, des ondes réfractées peuvent être identifiées et pointées.

1. Plus d'informations sur les différents types d'ondes sismiques sont disponibles en Annexe [A.4](#)

4.4. LECTURE DES TEMPS D'ARRIVÉE DES ONDES

La Figure 4.15 montre un exemple de pointé d'onde P . L'arrivée la plus probable, notée t_A , est située au sein d'un intervalle correspondant à l'arrivée la plus précoce t_E et la plus tardive t_L . Cet intervalle définit donc une incertitude sur le pointé. D'après Diehl et Kissling (2007), t_L peut être défini en identifiant le moment à partir duquel le signal présente une amplitude supérieure à 1,5 fois la moyenne des amplitudes du bruit de fond avant l'arrivée de l'onde (Fig 4.15). En pratique, l'observateur va pointer le temps d'arrivée le plus probable t_A et définir une incertitude autour de ce temps. Puis, la même opération est réalisée pour l'arrivée de l'onde S . Cette procédure est à la fois longue et minutieuse. Pour essayer de faciliter la lecture des temps d'arrivée, j'ai testé plusieurs approches de pointé automatique.

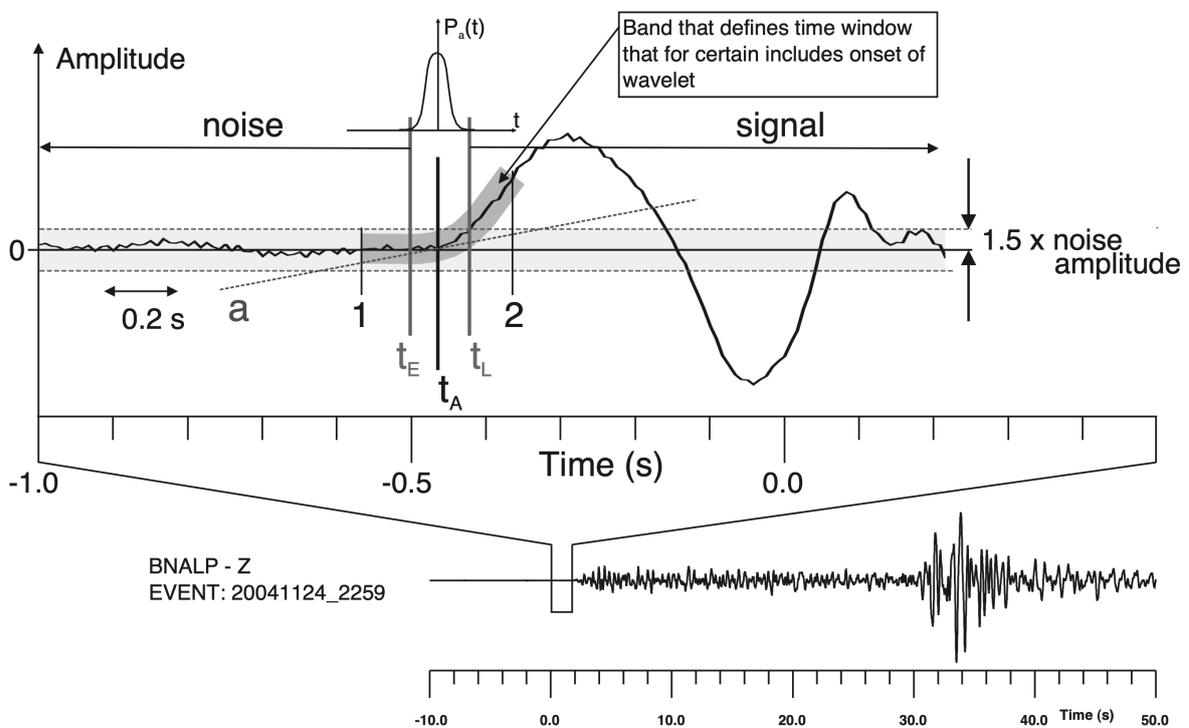


FIGURE 4.15 – Lecture de temps d'arrivée de l'onde P (d'après Diehl et Kissling, 2007). L'arrivée probable la plus précoce est notée t_E et la plus tardive t_L . Le temps d'arrivée la plus probable t_A se trouve au centre de l'intervalle définie par t_E et t_L .

Afin de tester ces approches, nous avons utilisé plusieurs bases de données. La première base de données correspond à une sélection de 99 événements naturels pointés manuellement par les analystes du BCSF-RéNaSS, détectés entre le 1er janvier et le 31 décembre 2021. Pour cette base de données, 568 formes d'ondes, comprenant un pointé d'onde P et S , ont été extraites. La deuxième base de données comprend les 131 nouveaux événements naturels détectés par *template matching* dans le quart nord-ouest de la France, décrit par Beucler *et coll.* (2021) (Fig 4.2). Ces événements sont

pointés manuellement par les analystes du LPG. Pour cette base de données, 1 143 formes d'ondes, comprenant un pointé d'onde P et S , ont été extraites. L'utilisation de la première base de données nous permet de vérifier si nos approches sont capables d'identifier les temps d'arrivée des événements de magnitude modérée (avec une M_d moyenne de 2,3). En revanche, la deuxième base de données nous permet d'estimer si nos différentes approches sont capables de pointer les ondes sismiques associées à des séismes de magnitude inférieure à ceux détectés par le RéNaSS (avec une M_d moyenne de 1,70). Ces événements de plus faible magnitude sont comparables à ceux que nous cherchons à détecter dans cette thèse.

Tout d'abord, la méthode de kurtosis est appliquée sur les deux bases de données. Comme présenté dans la Section 3.1.1, la fonction kurtosis quantifie « l'aplatissement » de la distribution des amplitudes d'un signal sismique. En pratique, celle-ci est calculée à l'aide d'une fenêtre glissante de 2 s. Le traitement du signal pour obtenir la fonction caractéristique est réalisé en suivant la série d'étape présentée dans [Baillard *et coll.* \(2014\)](#) et montrée dans la Figure 4.16.

La kurtosis est utilisée sur les deux bases de données. Sachant que les pointés sont réalisés au niveau de F_4 (Fig 4.16), pour chaque fenêtre d'étude nous disposons d'un pointé d'ondes P et S . L'ensemble de ces pointés peut être comparé à ceux réalisés manuellement.

TABLE 4.2 – *Tableau comparatif des pointés manuels et réalisés par Kurtosis pour différentes bases de données*

Kurtosis	Mode (s)	Médiane (s)
99 événements RéNaSS		
Onde P	0.034	0.14
Onde S	0.101	0.22
131 événements LPG		
Onde P	0.08	0.13
Onde S	0.06	1.24

Les distributions des différences de temps d'arrivée entre les pointés manuels et les pointés réalisés par kurtosis pour les ondes P et S sont illustrées dans la Figure 4.17. Nous observons une asymétrie positive pour toutes ces distributions indiquant que la kurtosis a tendance à surestimer les temps d'arrivée. Afin de retirer les valeurs statistiques extrêmes, les modes, moyennes, écart-types et médianes (Tab 4.2, 4.3 et 4.4) sont calculés, pour toutes les méthodes, entre leur 5e et 95e pourcentile. Cette surestimation

4.4. LECTURE DES TEMPS D'ARRIVÉE DES ONDES

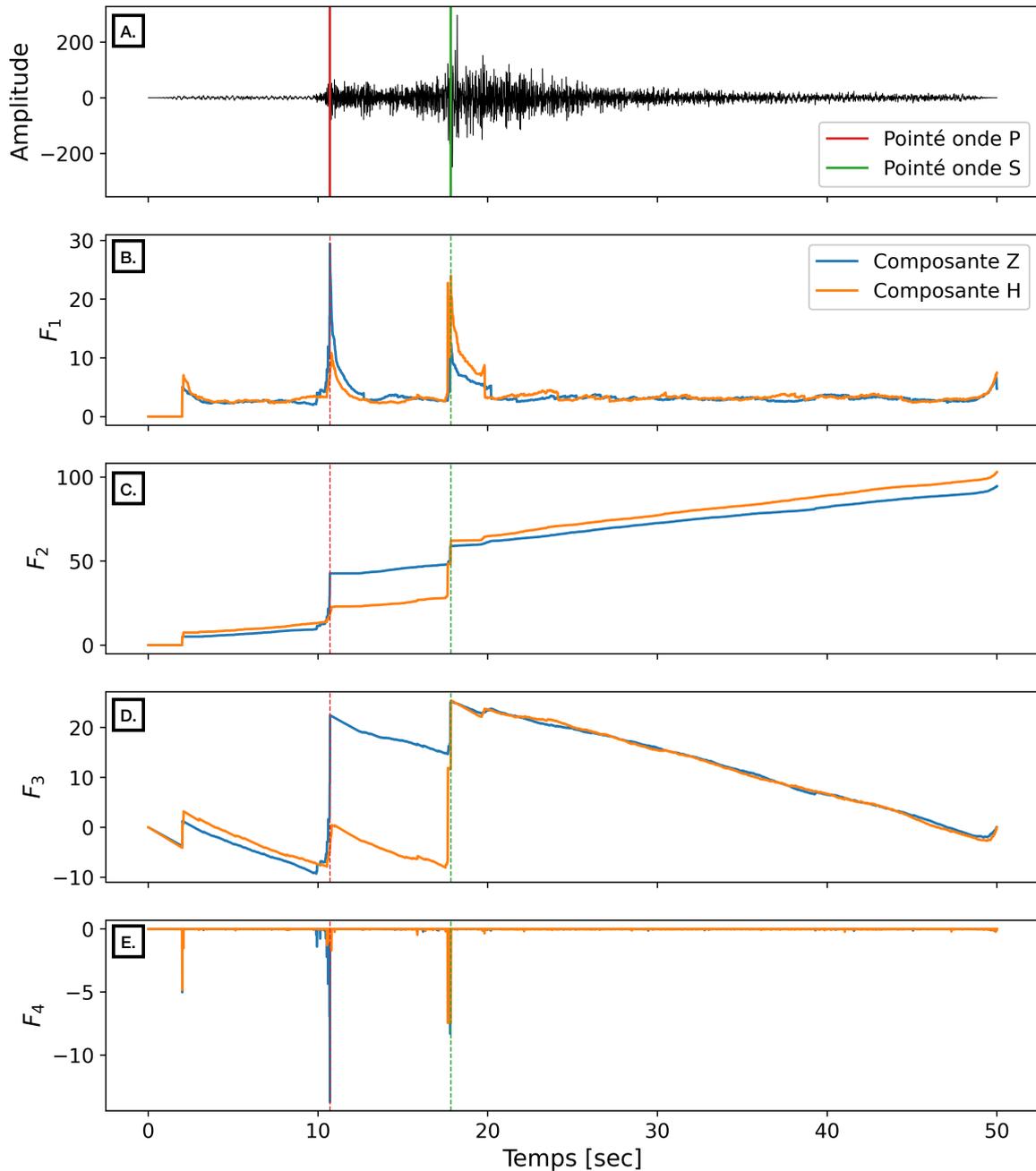


FIGURE 4.16 – Les étapes du calcul de la fonction caractéristique définie à partir de la kurtosis : (A) sismogramme Z filtré ; (B) F_1 correspondant à la fonction kurtosis ; (C) F_2 , retrait de toutes les pentes négatives ; (D) F_3 , correction linéaire ; et (E) F_4 correspond à la fonction caractéristique permettant de localiser les variations rapides d'amplitude de F_3 .

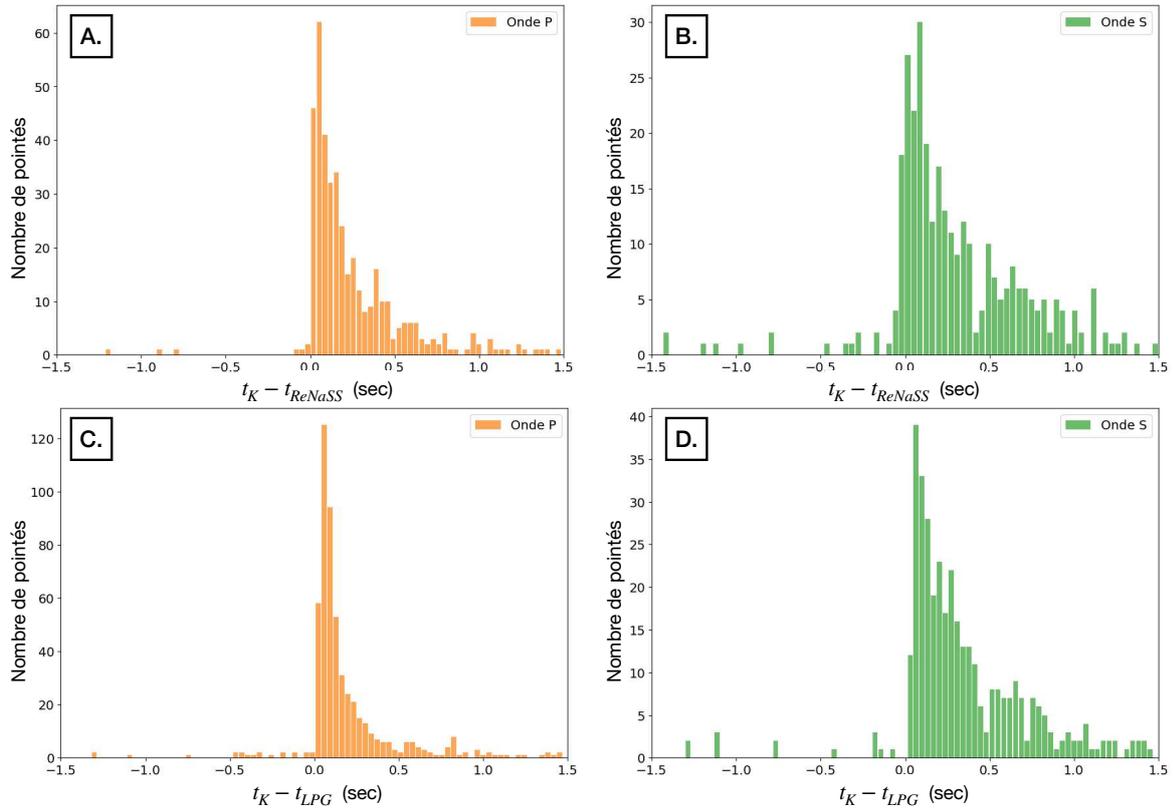


FIGURE 4.17 – Distributions des différences de temps d'arrivée (en secondes) entre les ondes P (A. et C.) et S (B. et D.) réalisées par les analystes du BCSF-RéNaSS (t_{ReNaSS}) ou du LPG (t_{LPG}) et obtenus par kurtosis (t_K).

des temps d'arrivées est confirmée par les modes observées pour les différentes bases de données. Elle peut s'expliquer par le fait que la kurtosis est basée sur la détection d'un changement soudain de l'amplitude de la forme d'onde. Mais l'onde S est souvent contaminée par la coda de l'onde P , ce qui peut dégrader la précision pour identifier l'arrivée de cette onde. L'arrivée de l'onde P peut néanmoins être difficile à pointer, par cette approche, lorsqu'elle est émergente. Par ailleurs, l'avantage de cette approche est qu'il est possible d'obtenir un pointé d'onde P et S sur chaque enregistrement.

Afin d'améliorer la précision des pointés, nous avons donc choisi de tester la méthode PhaseNet (Zhu et Beroza, 2019). Ce réseau de neurones profond a été entraîné grâce à des formes d'ondes sélectionnés manuellement dans le nord de la Californie et montre des performances de pointé élevées. Nous avons appliqué l'approche sur les deux bases de données présentées précédemment, sans ré-entraîner l'algorithme. Un pointé est récupéré si la probabilité de celui-ci dépasse 30%. Les distributions de ces différences de temps d'arrivée entre les pointés manuels et les pointés de PhaseNet pour les ondes P et S sont illustrées dans la Figure 4.18.

4.4. LECTURE DES TEMPS D'ARRIVÉE DES ONDES

Pour les 99 événements naturels de la première base de données, comprenant un total de 568 signaux, 379 ondes P et 392 ondes S ont pu être identifiées. Parallèlement, pour les 131 nouveaux événements naturels de la seconde base de données, comprenant un total de 1 143 signaux, 805 ondes P et 822 ondes S ont pu être pointées. Nous observons des distributions des différences de temps d'arrivées plutôt symétriques, centrées sur zéro. Comparées aux résultats donnés par la kurtosis, les valeurs de moyennes, écart-types et médianes sont considérablement plus faibles (Tab 4.3). Cela indique des pointés plus fiables et en accord avec les temps d'arrivées de référence. Nous pouvons également noter la différence d'écart-type obtenue entre les deux bases de données. Pour les événements BCSF-RéNaSS, les écart-types sont relativement faibles (environ 0.2 s) indiquant des distributions serrées autour de la moyenne. En revanche, pour les événements détectés par le LPG, les écart-types sont plus élevés, ce qui suggère des différences de temps d'arrivées plus importantes.

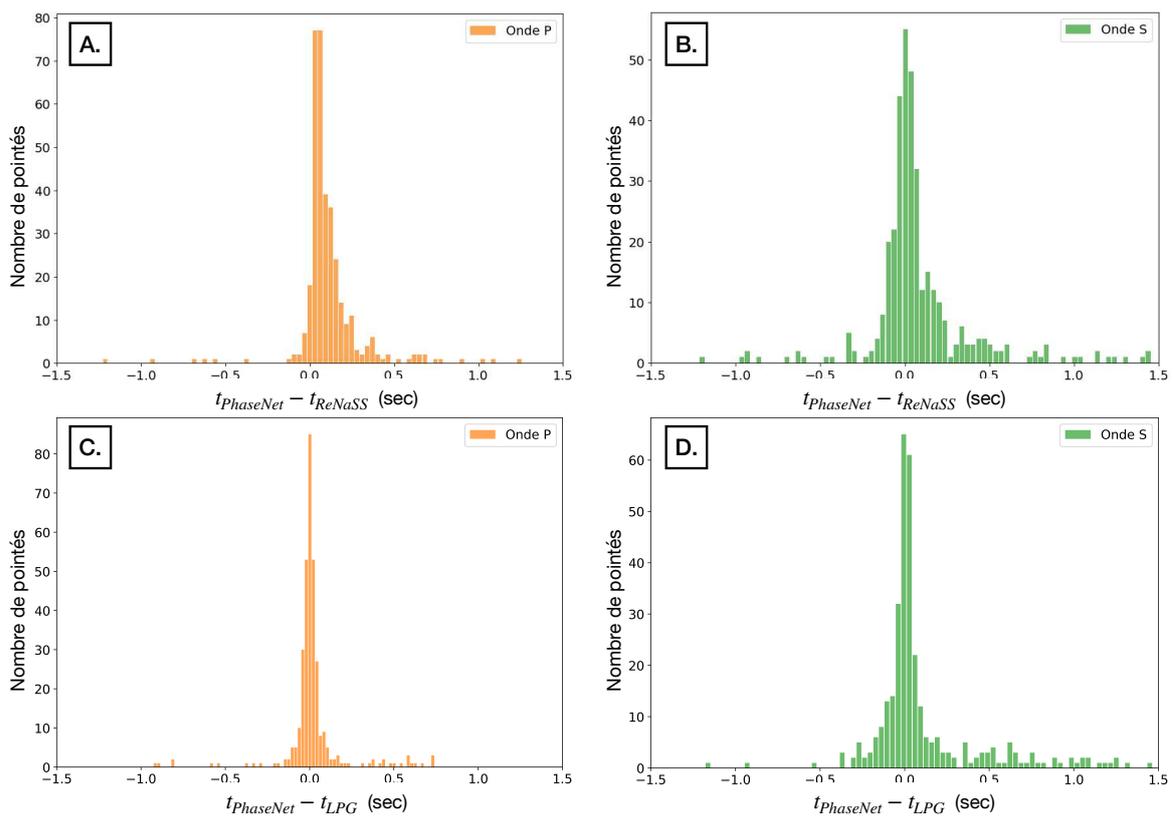


FIGURE 4.18 – Distributions des différences de temps d'arrivée (en secondes) entre les ondes P (A. et C.) et S (B. et D.) réalisés par les analystes du BCSF-RéNaSS (t_{ReNaSS}) ou du LPG (t_{LPG}) et PhaseNet ($t_{PhaseNet}$).

CHAPITRE 4. NOUVELLE APPROCHE POUR LA DÉTECTION ET LA DISCRIMINATION D'ÉVÉNEMENTS SISMIQUES DE FAIBLE MAGNITUDE

TABLE 4.3 – *Tableau comparatif des pointés manuels et réalisés par PhaseNet pour différentes bases de données*

PhaseNet	Moyenne (s)	Ecart-Type (s)	Médiane (s)
99 événements RéNaSS			
Onde P	0.13	0.2	0.07
Onde S	0.04	0.24	0.022
131 événements LPG			
Onde P	1.52	6.71	0.006
Onde S	0.37	3.44	0.02

Cette disparité peut être attribuée au fait que les 131 événements LPG sont de plus faible magnitude, impliquant des formes d'ondes de plus faibles amplitudes et un rapport signal sur bruit réduit. Par conséquent, la méthode PhaseNet peut rencontrer plus de difficultés à identifier les temps d'arrivées sur certains signaux, comme illustre la Figure 4.19.C où le temps d'arrivée de l'onde P n'a pas pu être identifié. Cependant, les valeurs des moyennes sont proches de zéro, indiquant un pointé tout de même précis (Fig 4.19.A et 4.19.B). Il est également important de noter que l'approche PhaseNet identifie seulement les temps d'arrivées des ondes directes (P_g et S_g). Lorsque les ondes réfractées sont visibles sur le sismogramme, nous observons des erreurs dans l'identification des phases (Fig 4.19.D). Pour les deux ensembles de données, environ 70% des signaux ont pu être pointés par PhaseNet alors que 100% des signaux ont été pointés manuellement. Cela indique que cette méthode ne peut donc pas être utilisée de manière entièrement automatique sur notre jeu de données. Cette approche peut être néanmoins utilisée comme un outil d'aide à la décision.

Enfin, nous avons testé une dernière approche : EQTransformer (Mousavi *et coll.*, 2020). Cette approche utilise un réseau de neurones permettant de déterminer trois probabilités : séismes, onde P et onde S . Contrairement à PhaseNet qui a été entraîné seulement sur des signaux de Californie, EQTransformer a été entraîné sur la base de données STEAD contenant environ 850 000 séismes à l'échelle mondiale. Pour ce test, nous avons utilisé les deux bases de données utilisées précédemment. Nous avons choisi le seuil de détection par défaut utilisé par l'approche, ainsi, un pointé est récupéré si la probabilité de celui-ci dépasse 10%.

Les distributions des différences de temps d'arrivée entre les pointés manuels et les pointés de EQTransformer pour les ondes P et S sont illustrées dans la Figure 4.20. Pour les 99 événements naturels de la première base de données, comprenant un total

4.4. LECTURE DES TEMPS D'ARRIVÉE DES ONDES

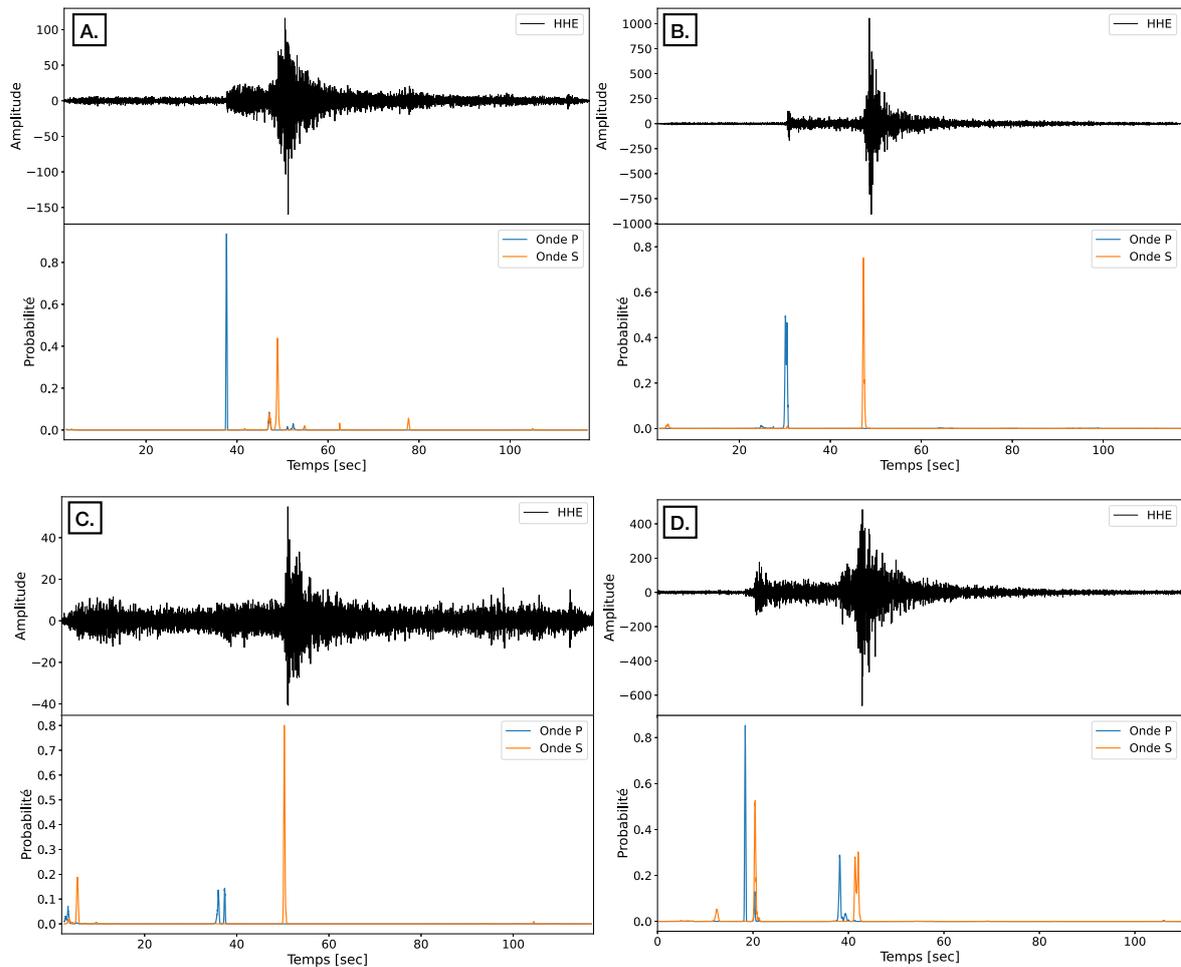


FIGURE 4.19 – Exemples de pointés réalisés par PhaseNet sur quatre formes d’ondes sismiques. (A) Forme d’onde du séisme du 26 Août 2018 à 01h41 (TU) enregistré par la station OLIV. (B) Forme d’onde du séisme du 10 Décembre 2020 à 17h05 (TU) enregistré par la station CHIF. (C) Forme d’onde du séisme du 20 Mai 2019 à 21h04 (TU) enregistré par la station OLIV. (D) Forme d’onde du séisme du 01 Février 2020 à 21h26 (TU) enregistré par la station LOCF.

de 568 signaux, 437 ondes P et 426 ondes S ont pu être identifiées. Parallèlement, pour les 131 nouveaux événements naturels de la seconde base de données, comprenant un total de 1 143 signaux, 509 ondes P et 453 ondes S ont pu être pointés.

Nous observons des distributions des différences de temps d’arrivée semblables aux résultats de PhaseNet. Elles sont symétriques et centrées autour de zéro. Les valeurs de moyennes, d’écart-types et de médianes absolues des différences sont similaires, voire plus faibles lorsqu’appliquées aux 99 événements RéNaSS (Tab 4.4) comparées aux résultats obtenus par PhaseNet. Nous remarquons également une différence dans les écarts-types entre les deux bases de données. Pour les événements RéNaSS, l’écart-

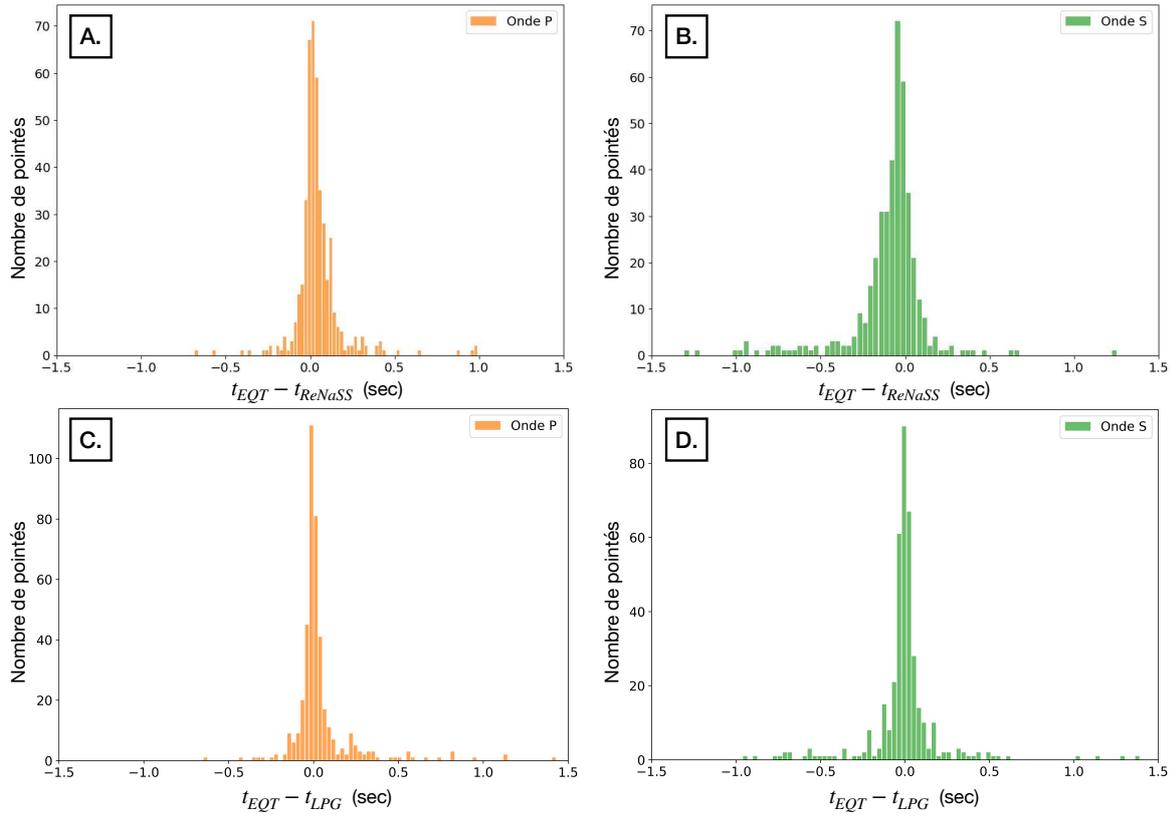


FIGURE 4.20 – Distributions des différences de temps d'arrivée (en secondes) entre les ondes P (A. et C.) et S (B. et D.) réalisés par les analystes du BCSF-RéNaSS (t_{ReNaSS}) ou du LPG (t_{LPG}) et EQTransformer (t_{EQT}).

TABLE 4.4 – Tableau comparatif des pointés manuels et réalisés par EQTransformer pour différentes bases de données

EQtransformer	Moyenne (s)	Ecart-Type (s)	Médiane (s)
99 événements RéNaSS			
Onde P	0.03	0.06	0.02
Onde S	-0.08	-0.05	0.11
131 événements LPG			
Onde P	2.24	6.06	0.002
Onde S	1.5	4.09	-0.001

type est très faible et proche de zéro, ce qui indique une distribution étroite autour de zéro. En revanche, pour les événements LPG, les écarts-types sont plus élevés, ce qui suggère des différences de temps d'arrivée potentiellement plus importantes. Cette différence peut, encore une fois, être attribuée au fait que les 131 événements LPG ont une magnitude plus faible.

4.4. LECTURE DES TEMPS D'ARRIVÉE DES ONDES

Nous pouvons également noter que, pour les 99 événements RéNaSS, sur les 568 signaux, EQTransformer a réussi à pointer environ 75% des formes d'ondes, soit 5% de plus que PhaseNet. Cependant, pour les 131 événements LPG, sur les 1 143 signaux, EQTransformer a pu pointer environ 44% d'entre eux. Étant donné que ces 131 événements, détectés par *template matching*, ont des magnitudes similaires aux événements que nous cherchons à détecter, il n'est donc pas possible d'utiliser cette méthode automatiquement sur les données.

À la lumière de ces résultats, les pointés sont effectués manuellement tout en utilisant PhaseNet comme aide à la décision.

Chapitre 5

Conséquences sur la compréhension de la sismicité du Massif armoricain

Sommaire

5.1	Calcul de la magnitude de durée (M_d)	133
5.2	Application à la sismicité armoricaine de 2018 à 2021	138
5.2.1	Détails des différentes étapes	138
5.2.2	Nouvelle carte de sismicité de 2018 à 2021	146
5.3	Études préliminaires sur l’origine de la sismicité	152
5.3.1	Distribution spatiale de la sismicité	152
5.3.2	Facteurs environnementaux	159
5.3.3	Bilan des éléments permettant de caractériser la sismicité	163
5.4	Conséquences en termes d’observation sismologique	165
5.4.1	Analyse d’un réseau <i>a posteriori</i>	165
5.4.2	Comparaison avec d’autres méthodes de détection	168

Suite aux différents tests que nous avons réalisés (Chapitre 4), nous avons opté pour une approche de *template matching* pour la détection suivie par deux étapes de discriminations par *deep learning* afin de détecter les événements sismiques de faible magnitude dans le Massif armoricain (Fig 5.1). L’ensemble de cette méthode nous permet d’élaborer un catalogue d’événements sismiques naturels avec des événements non détectés par les méthodes classiques. Ces événements sont ensuite localisés par l’approche développée par Haugmard (2016). Afin de comparer ces nouveaux événements sismiques avec le catalogue national du BCSF-RéNaSS, nous devons calculer des magnitudes. Pour cela, nous utilisons la magnitude de durée (M_d).

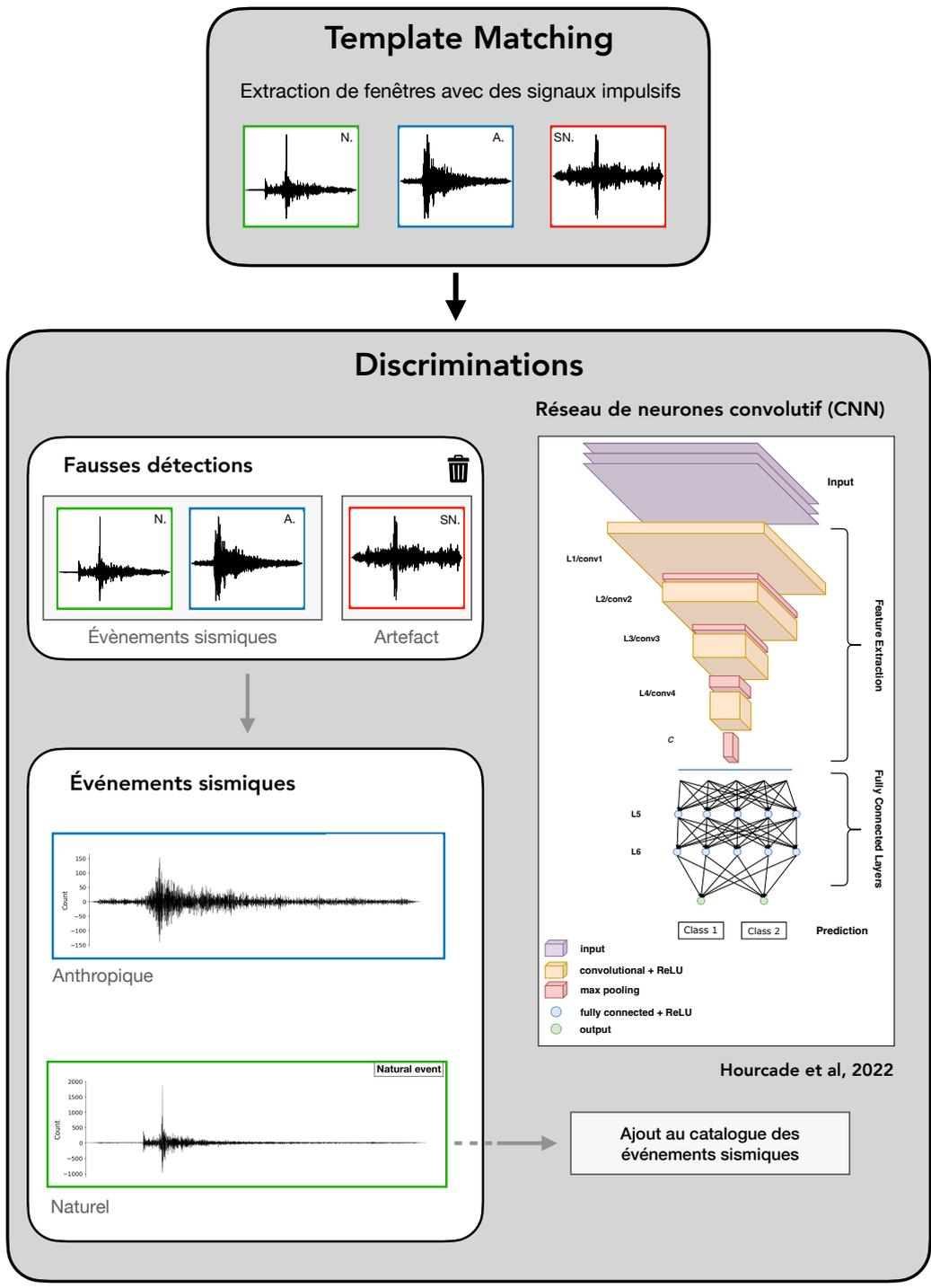


FIGURE 5.1 – Récapitulatif de la méthode de détection et de discrimination suivie. Le signal N correspond à un événement sismique naturel, A anthropique et SN à un signal impulsif parasite.

5.1 Calcul de la magnitude de durée (M_d)

La magnitude de durée est couramment utilisée pour les séismes locaux de magnitude comprise entre 0 et 5. Elle est obtenue à partir de la mesure de la durée de la coda.

Pour calculer la magnitude de durée, nous utilisons l'équation 5.1 (d'après Lee *et coll.*, 1972), où τ représente la durée de la coda de l'onde S et d la distance entre la station et l'événement sismique. La coda fait référence à l'énergie enregistrée après l'arrivée des ondes balistiques et son amplitude décroît de manière exponentielle avec le temps. On considère que la coda se termine lorsque les oscillations du sismogramme retournent à un niveau d'énergie inférieur à deux fois celui enregistré avant le séisme. Les constantes utilisées dans l'équation 5.1 sont théoriquement adaptées à une application en Californie. Cependant, dans le cadre de cette étude, notre objectif est d'estimer une magnitude pour obtenir un ordre de grandeur de l'énergie des événements détectés. Une analyse plus approfondie des magnitudes de ces événements pourra être réalisée ultérieurement.

$$M_d = -0.87 + 2\log_{10}(\tau) + 0.0035d \quad (5.1)$$

Afin de déterminer la durée de la coda de l'onde S , plusieurs approches peuvent être utilisées. Dans une première approche, la décroissance de la coda peut être identifiée à partir de l'enveloppe du signal. L'amplitude du signal avant l'arrivée des premières ondes est enregistrée, puis comparée à l'amplitude après l'arrivée de l'onde S . Lorsque l'amplitude du signal revient à son état de base, généralement environ deux fois l'amplitude du bruit, on estime que la coda se termine. La Figure 5.2 montre un exemple d'application à partir de l'enveloppe d'un signal sismique.

Cependant, cette approche peut manquer de fiabilité. Si le signal sismique est suivi par du bruit impulsif, l'amplitude du bruit ne va pas revenir à son état de base et la magnitude peut être surestimée. Pour palier ce problème, nous appliquons la méthode de Cuvier *et coll.* (prep) permettant une séparation des points qui ne sont pas en accord avec la gaussianité dominante du signal. En se basant sur l'hypothèse que le signal sismique continu suit une distribution gaussienne, nous pouvons discriminer les échantillons perturbés comme étant ceux ne suivant pas cette loi. Un signal gaussien, trié par ordre de croissant, correspond à une distribution appelée fonction *probit* modi-

5.1. CALCUL DE LA MAGNITUDE DE DURÉE (M_D)

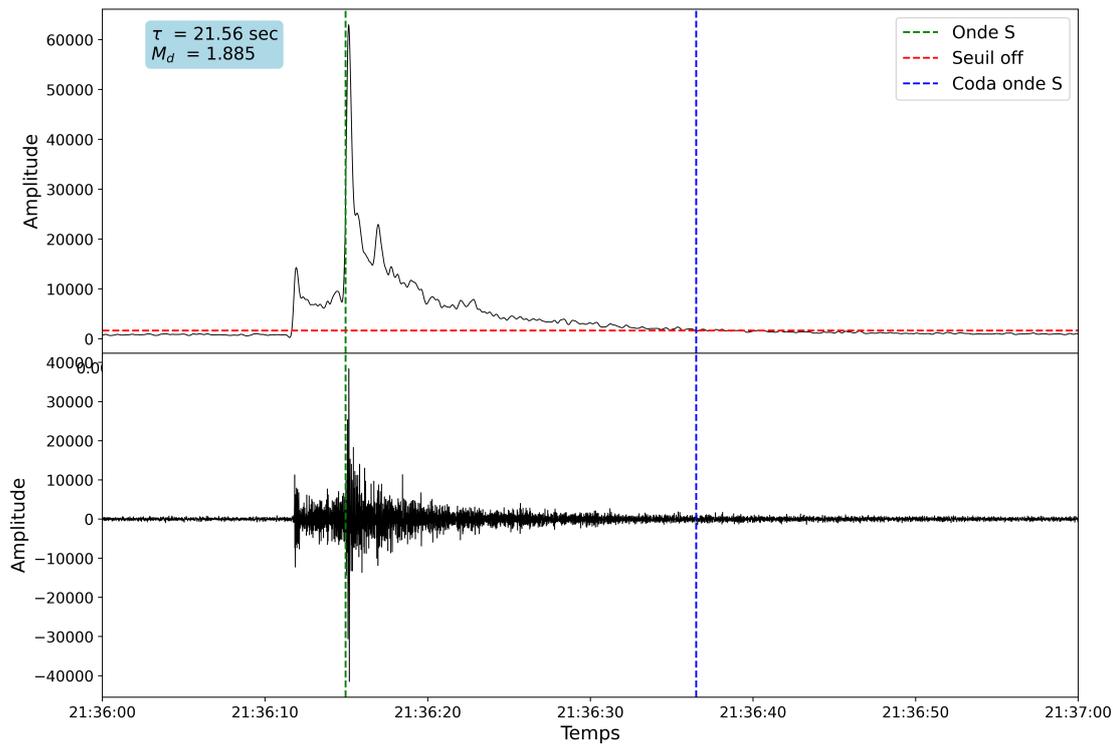


FIGURE 5.2 – Détermination de la durée de la coda de l'onde S à partir de l'amplitude du signal. Le signal du haut correspond à l'enveloppe du signal sismique. Le signal du bas correspond à la forme d'onde de l'événement sismique du 17 février 2021 enregistré par la station GUEF à 21h36. La ligne verte correspond au temps d'arrivée de l'onde S . La ligne rouge correspond au seuil de 2 fois l'amplitude du signal permettant de déterminer la fin de la coda. La ligne bleue correspond à l'intersection entre l'enveloppe du signal et le seuil. Celle-ci représente la fin de la coda de l'onde S .

fiée. À partir d'un signal sismique trié (St), il est donc possible de retrouver les indices des points (St_i) ne faisant pas partie de cette distribution gaussienne. Ces indices définissent l'intervalle $[Q_A, Q_B]$ à l'intérieur duquel les points suivent une distribution gaussienne et à l'extérieur, les points sont considérés comme perturbés.

La Figure 5.3.A montre un exemple d'application pour un signal gaussien. Pour ce cas, l'intervalle $[Q_A, Q_B]$ correspond à l'ensemble du signal échantillonné ($Q_A = 0, Q_B = 2\,000$). La Figure 5.3.B montre un signal gaussien incluant une perturbation similaire à l'arrivée d'une onde sismique. Il n'est ici plus possible de faire correspondre la fonction *probit*, de manière fiable, pour l'ensemble du signal trié. La meilleure adéquation entre les deux courbes est obtenue sur l'intervalle $[Q_A, Q_B] = [30, 1\,970]$. Les lignes oranges en tiretés indiquent à partir de quelle amplitude la distribution sort de l'intervalle $[Q_A, Q_B]$ et révèlent ainsi les points perturbés. Ces points perturbés sont ensuite reportés en rouge sur la Figure 5.3.B3.

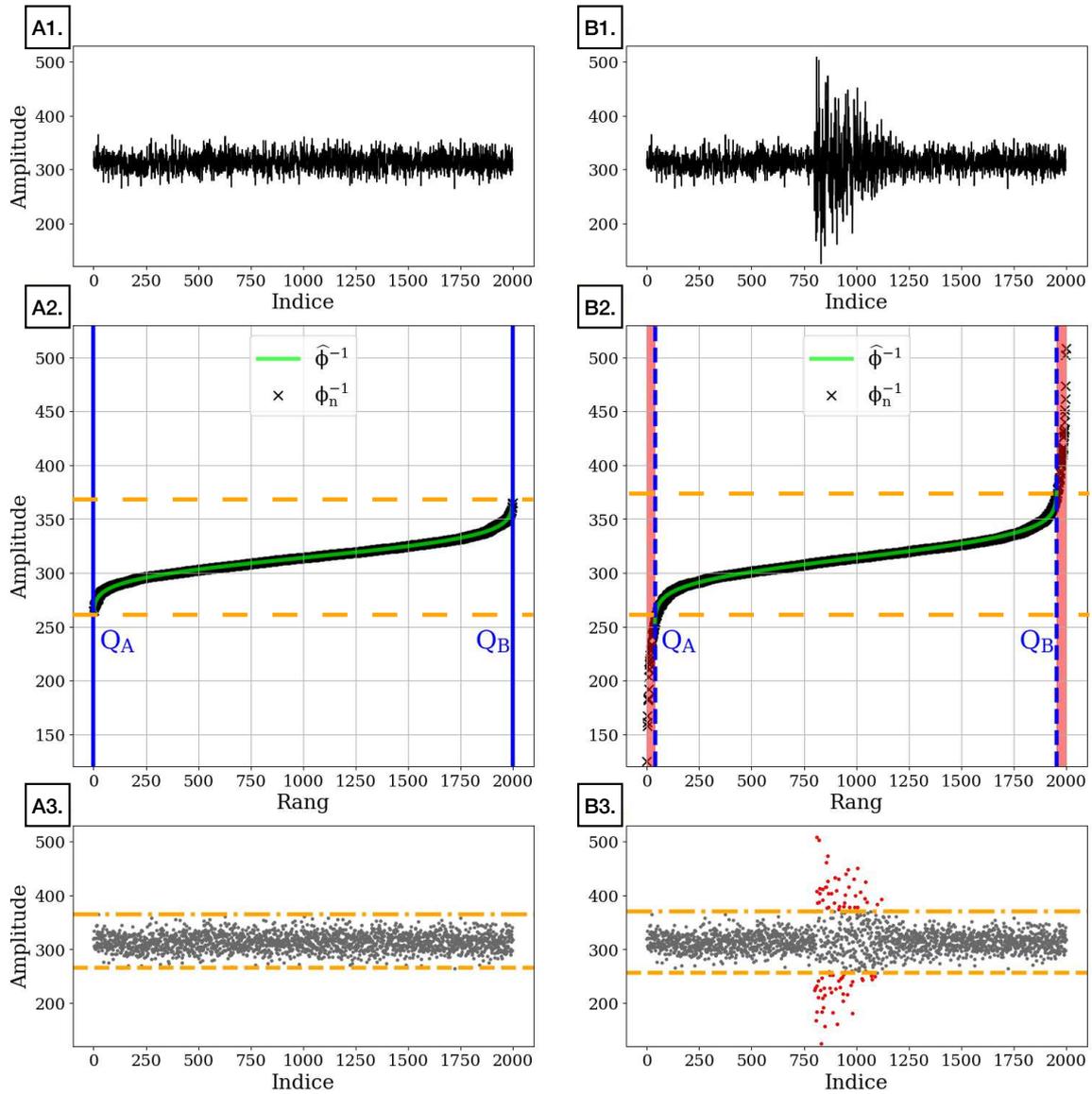


FIGURE 5.3 – Estimation de la déviation de l'hypothèse gaussienne sur deux formes d'ondes synthétiques (A1, B1) suivant la même loi gaussienne. A1 correspond à un signal gaussien, B1 est affectée par une perturbation sur 400 échantillons. A2 et B2 sont les signaux triés par amplitude croissante (courbes noires ϕ_n^{-1}). Les intervalles $[Q_A, Q_B]$ sont obtenus à partir du meilleur ajustement entre ϕ_n^{-1} et $\hat{\phi}^{-1}$ correspondant à la fonction probit. Les tracés oranges montrent à partir de quelle amplitude la distribution sort de cet intervalle $[Q_A, Q_B]$. A3 et B3 illustrent, en gris, les échantillons non perturbés et en rouge les échantillons perturbés.

En appliquant cette approche à une forme d'onde sismique, le changement rapide d'amplitude associé à l'arrivée de l'onde S va modifier la gaussianité du signal sismique continu. Les échantillons de la coda demeurent également statistiquement différents du signal précédent, permettant de les caractériser et d'estimer la durée de cette dernière. En pratique, grâce à une fenêtre de 1 s parcourant le signal, on cherche le moment,

5.1. CALCUL DE LA MAGNITUDE DE DURÉE (M_D)

après l'arrivée de l'onde S , où 100% des points redeviennent gaussiens. Ainsi, la durée de la coda de l'onde S peut être déterminée (Fig 5.4). Cette méthode a pour avantage de ne nécessiter aucun seuil.

Le bulletin du BCSF-RéNaSS fournit des magnitudes locales (M_L) pour les séismes détectés. Afin de pouvoir comparer nos magnitudes avec celles du bulletin, et d'obtenir un catalogue homogène en termes de pointés et de localisations, nous avons procédé à une réidentification manuelle des temps d'arrivées des ondes. Cela nous permet de convertir les magnitudes locales en magnitudes de durée (M_d). La Figure 5.5 montre les magnitudes locales du BCSF-RéNaSS en fonction des magnitudes de durée recalculées. On remarque que les M_d calculées sont généralement sur-estimées par rapport aux M_L publiées dans le bulletin.

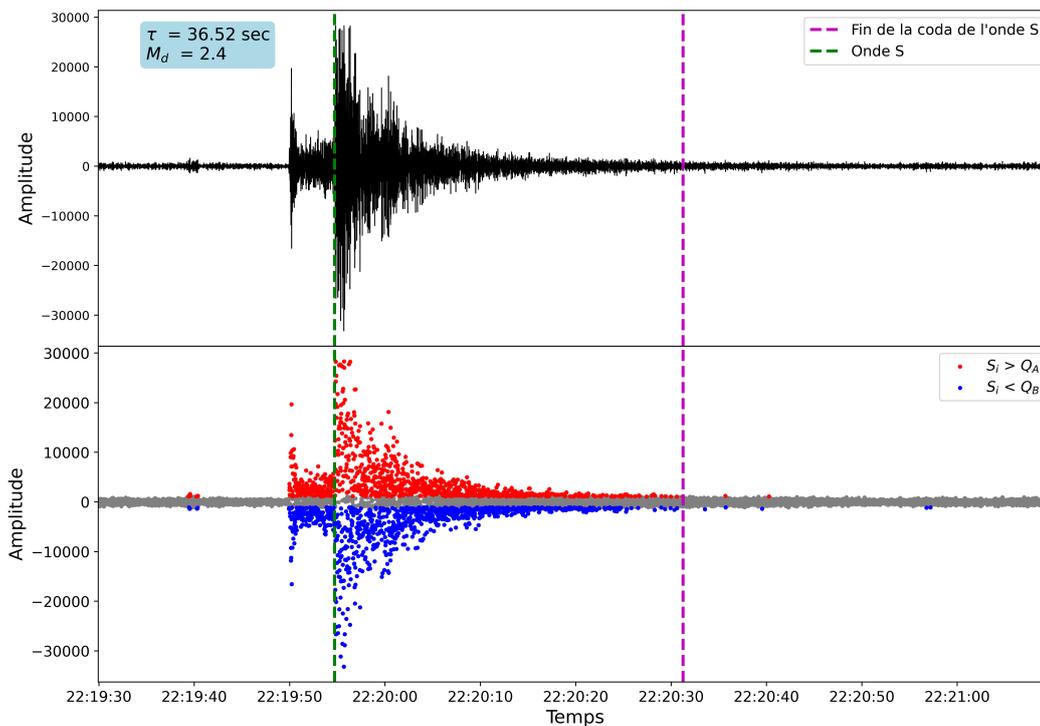


FIGURE 5.4 – Détermination de la durée de la coda de l'onde S à partir de l'analyse de la gaussianité. Le signal du haut correspond à la forme d'onde de l'événement sismique du 21 août 2018 enregistré par la station LOCF à 21h19. Le signal du bas correspond à l'analyse de la gaussianité de la forme d'onde. Les points gris correspondent aux points faisant partie de la distribution gaussienne, à l'intérieur de l'intervalle $[Q_A, Q_B]$. Les points rouges sont les points dont l'indice est supérieur à Q_A et les points bleus correspondent aux points dont l'indice est inférieur à Q_B . Les points bleus et rouges sont donc considérés comme perturbés. La ligne verte correspond au temps d'arrivée de l'onde S . La ligne magenta correspond au seuil à partir duquel, sur une fenêtre de 1 s, tous les points redeviennent gaussiens, définissant la fin de la coda de l'onde S .

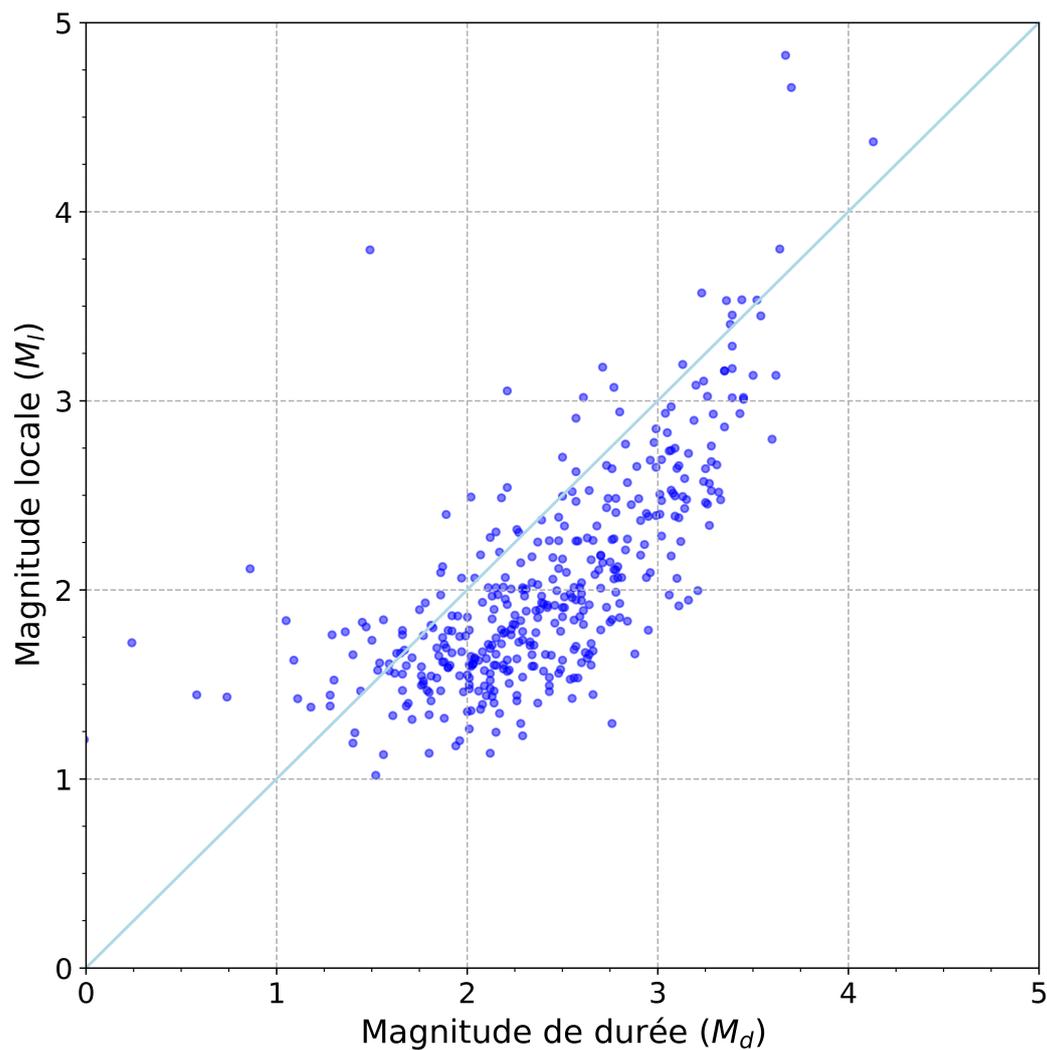


FIGURE 5.5 – Magnitudes de durée M_d comparées aux magnitudes locales (M_l) du bulletin du BCSF-RéNaSS.

5.2 Application à la sismicité armoricaine de 2018 à 2021

5.2.1 Détails des différentes étapes

Détection des signaux de faible amplitude par *template matching*

Afin de détecter les événements sismiques de plus faible amplitude, nous appliquons le protocole sur quatre années de données du 1er janvier 2018 au 31 décembre 2021. Pour ce faire, nous avons constitué un catalogue de signaux de référence en utilisant des signaux d'événements provenant du bulletin du BCSF-RéNaSS. Les événements sélectionnés pour la création du catalogue ont été détectés entre le 1er janvier et le 1er juin 2022 dans le quart nord-ouest de la France. Les formes d'ondes sismiques sur les trois composantes sont collectées à partir des stations du réseau FR pour lesquelles au moins un pointé des ondes P et S est documenté dans le bulletin. Si le rapport signal sur bruit dépasse la valeur de 5, le signal est alors retenu. Les sismogrammes sont redimensionnés et traités selon le protocole présenté dans le chapitre 4.

Le catalogue de référence est constitué de 687 enveloppes de signaux pour chaque composante provenant de 194 événements sismiques. Pour des raisons de temps de calcul, il n'a pas été possible d'utiliser l'intégralité du catalogue pour l'étape d'inter-corrélation. Nous avons sélectionné aléatoirement 100 enveloppes de signaux (33 provenant de la composante Nord-Sud, 33 de la composante Est-Ouest et 34 de la composante verticale). Ces 100 enveloppes de signaux, correspondant à un total de 77 événements sismiques enregistrés par 31 stations, constituent donc notre catalogue de signaux de référence. La Figure 5.6 montre la localisation des 77 événements sismiques pris comme référence.

En pratique, pour cette application, la détection a été effectuée uniquement sur la composante Nord de 25 stations permanentes du réseau FR, en utilisant l'ensemble du catalogue de signaux de référence. Une application sur l'ensemble des composantes pourra être réalisée ultérieurement. Le signal continu de chaque station a été découpé en segments de même taille que les enveloppes de référence (90 s). Ces segments se chevauchent de 25% afin de préserver le plus d'informations. Les inter-corrélations normalisées sont calculées sur les enveloppes des segments temporels. Si la valeur maximale de l'inter-corrélation dépasse le seuil prédéfini (0,75), le segment analysé est retenu.

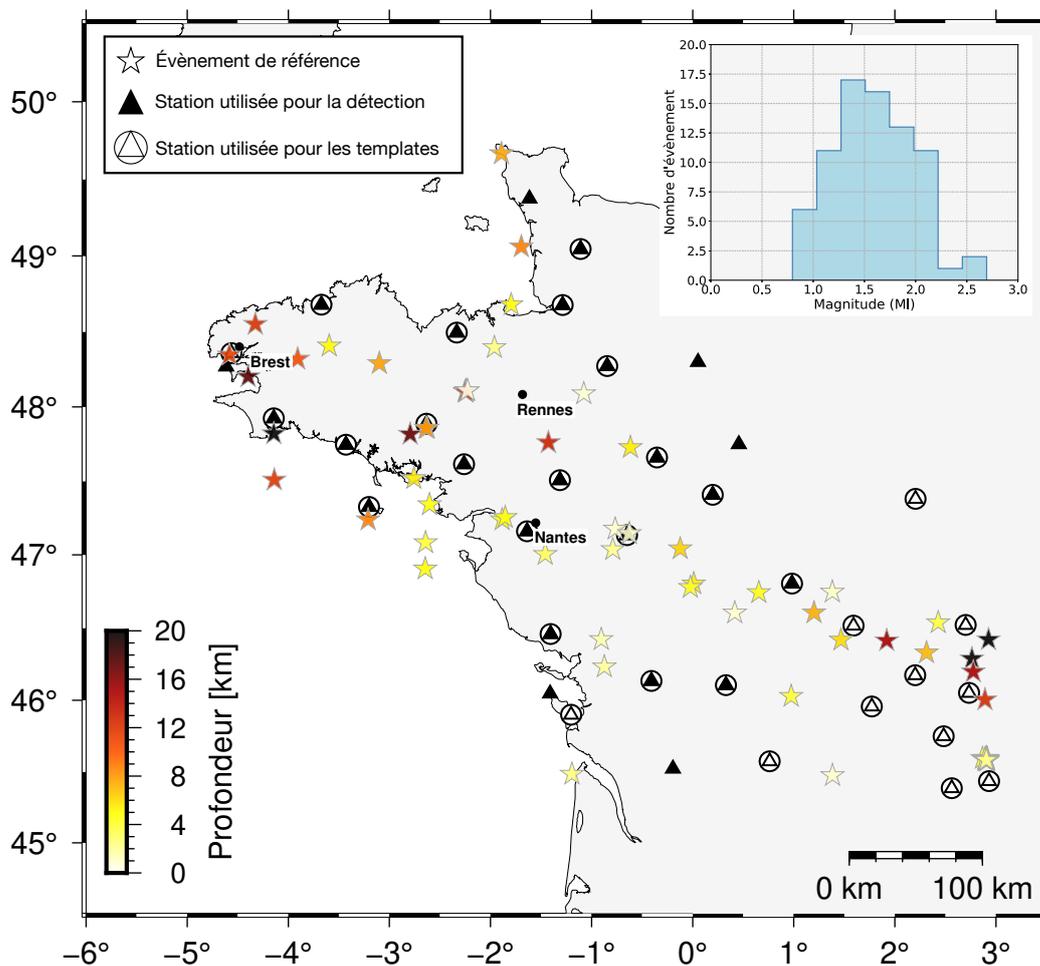


FIGURE 5.6 – Localisation des 77 événements de référence et des stations sismiques utilisées. Les étoiles correspondent aux événements de référence et la couleur est associée à la profondeur. Les triangles noirs correspondent aux stations dont les signaux continus ont été utilisés pour l'étape d'inter-corrélation. Les triangles entourés par un cercle correspondent aux stations dont les formes d'ondes ont été utilisées pour le catalogue de référence.

Lorsqu'un événement sismique se produit, plusieurs stations peuvent enregistrer l'arrivée des ondes sismiques, ce qui crée une cohérence spatiale. Afin de garantir qu'un événement est potentiellement un tremblement de terre, il doit être détecté par plusieurs stations. Pour maximiser la détection des événements sismiques, nous avons utilisé deux approches : (i) utilisation de l'ensemble des stations du quart nord-ouest de la région (Fig 5.7.A) et un nombre minimum de quatre stations pour considérer une détection, (ii) division de la région en quatre sous-réseaux (Fig 5.7.B) et un nombre minimum de trois stations par sous-groupe pour valider une détection. Nous avons choisi cette

seconde approche afin de récupérer les événements sismiques potentiels qui n'auraient pas été détectés par la première approche.

En utilisant toute les stations du quart nord-ouest, 230 147 événements potentiellement sismiques ont été détectés, cet ensemble sera nommé TM_{all} . En utilisant les sous-zones de stations, 206 931 événements potentiellement sismiques ont été détectés, cet ensemble sera nommé TM_{zone} . Ces deux ensembles seront traités de manière distincte afin de pouvoir comparer les résultats *in fine*.

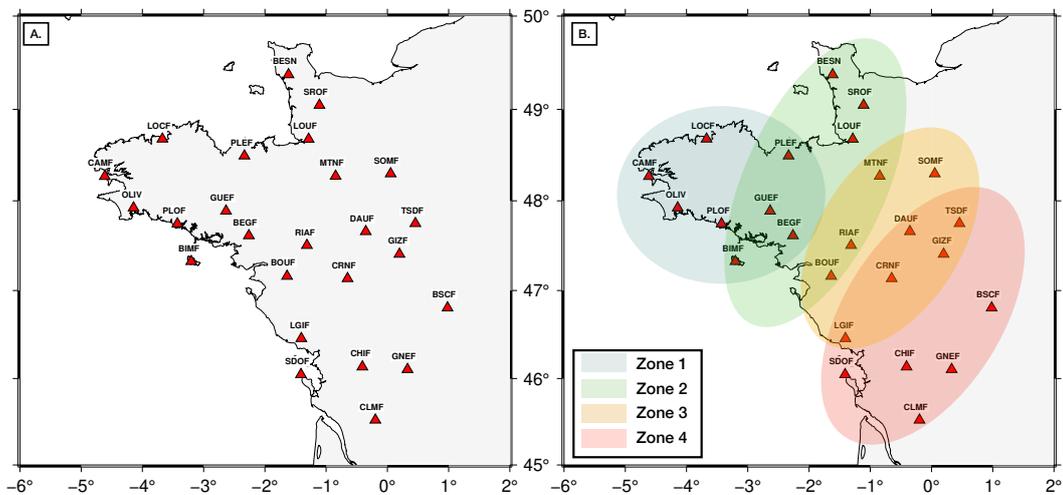


FIGURE 5.7 – Localisation des stations utilisées pour l'inter-corrélation selon les deux approches d'analyses. (A) Toutes les stations sont utilisées et les résultats sont nommés TM_{all} . (B) La région est divisée en quatre zones, correspondant à des sous-ensembles de stations. Les résultats de ce test sont nommés TM_{zone} .

Discrimination des fausses détections

Parmi l'ensemble de ces détections, il existe différents types de signaux. Dans un premier temps, nous allons discriminer les événements sismiques des fausses détections par l'approche décrite en Section 4.2. Ainsi, pour TM_{all} , sur les 230 147 événements détectés par *template matching*, 7 669 sont classés comme « *quake* ». Pour TM_{zone} , sur les 206 931 événements détectés par *template matching*, 22 157 sont classés comme « *quake* ».

Afin de vérifier les performances du réseau neuronal convolutif, nous évaluons les résultats de l'ensemble TM_{all} obtenus entre le 1er janvier et le 31 avril 2021, en effectuant une inspection visuelle des événements détectés. Au cours de cette période, nous avons détecté un total de 2 714 événements. Les résultats de la classification, au *network-level*, sont présentés sous la forme d'une matrice de confusion en Figure 5.8.

		Prédiction CNN		Total	Recall
		Quake	Noise		
Réalité	Quake	603	78	681	88,54 %
	Noise	62	1971	2033	96,95 %
Total		665	2049	2714	
		Précision			
		90,67 %	96,19 %		

Accuracy = 94,8%

FIGURE 5.8 – Matrice de confusion, pour la première étape de discrimination, obtenue sur les détections du 1er janvier au 31 avril 2021, de l'ensemble TM_{all} .

Les performances de classification se révèlent élevées (*accuracy* de 94,8%) avec, en particulier, un *recall* pour la classe *noise* de 96,95%. Cela signifie que moins de 4% des événements de la classe « *noise* » sont mal classés dans l'ensemble de la base de données. Le F1-scores pour la classe « *noise* » est de 96,56%. Ce résultat indique que le réseau est particulièrement efficace pour identifier et exclure les signaux bruités dans des données indépendantes. En revanche, le *recall* pour la classe « *quake* » est de 88,54%. Cette valeur plus faible peut être attribuée à la qualité de la base de données utilisée car, celle-ci est composée de signaux de faible amplitude présentant un rapport signal sur bruit réduit (événements de faible magnitude ou loin des stations). Par conséquent, les événements sismiques mal classés peuvent être difficiles à détecter et à localiser avec précision. Ainsi, les 7 669 événements, pour TM_{all} , et les 22 157 événements, pour TM_{zone} , classés comme « *quake* » sont récupérés pour l'étape suivante.

Discrimination des événements sismiques

La dernière étape correspond à la discrimination entre les événements naturels et anthropiques. Pour cela, la méthode développée en Section 4.3, est utilisée. Les spectrogrammes calculés à l'étape précédente sont réutilisés pour cette discrimination. Ainsi, pour TM_{all} , sur les 7 669 événements, 751 sont classés comme « naturels » par l'algorithme. Pour TM_{zone} , sur les 22 157 événements, 2 859 sont classés comme « naturels ».

Afin de vérifier les performances du réseau neuronal convolutif, nous avons évalué les résultats de l'ensemble TM_{all} obtenus entre le 1er janvier et le 31 avril 2021, en effectuant une inspection visuelle des événements détectés. Au cours de cette période,

603 événements sismiques ont été correctement classés à l'étape 2 (Fig 5.8). Ces 603 événements sont utilisés pour évaluer les performances de classification. Les résultats de la classification, au *network-level*, sont présentés sous la forme d'une matrice de confusion en Figure 5.9.

		Prédiction CNN		Total	Recall
		Natural	Anthropique		
Réalité	Natural	93	3	96	96,87 %
	Anthropique	18	489	507	96,44 %
Total		111	492	603	
Précision		83,78 %	99,39 %		

Accuracy = 96,5 %

FIGURE 5.9 – Matrice de confusion, pour la seconde étape de discrimination, obtenue sur les détections du 1er janvier au 31 avril 2021, de l'ensemble TM_{all} .

Les performances de classification se révèlent élevées (*accuracy* de 96,5%) avec, en particulier, des *recall* pour la classe « anthropique » de 96,44% et pour la classe « naturel » de 96,87%. Les F1-scores pour les classes « naturel » et « anthropique » sont respectivement de 89,85% et 97,89%. Ces résultats indiquent que le réseau est particulièrement efficace pour identifier et exclure les signaux anthropiques dans des données indépendantes.

Ainsi, l'ensemble des séismes classés comme événements naturels sont sélectionnés (les 751 pour TM_{all} et les 2 859 pour TM_{zone}). Après une inspection visuelle de chaque événement, pour l'ensemble TM_{all} , 360 correspondent à des nouveaux événements sismiques qui ne sont pas présents dans les bulletins du BCSF-RéNaSS et du LDG. Pour l'ensemble TM_{zone} , 444 nouveaux événements sont détectés. Parmi ceux-ci, 263 sont en commun avec l'ensemble TM_{all} . Les événements restant pour les deux ensembles correspondent soit à des séismes naturels déjà présents dans les bulletins, soit à des événements anthropiques ou bruités mal classés. La plus forte proportion d'événements mal classés pour l'ensemble TM_{zone} peut être attribuée à l'utilisation d'un minimum de trois stations pour définir une détection. Ainsi, certains événements ne disposaient que de peu de signaux pour réaliser une classification. En comparaison, TM_{zone} offre de meilleurs résultats de classification. Néanmoins, l'ensemble TM_{zone} nous permet de récupérer des séismes qui n'ont pas pu être détectés dans l'ensemble TM_{all} .

Ainsi, l'application de l'ensemble de cette méthode de détection et de discriminations permet de détecter un total de 541 nouveaux événements sismiques naturels non détectés par les agences nationales.

Pointé et localisation des événements sismiques naturels

Pour l'ensemble de ces événements, les temps d'arrivée des ondes sismiques ont été identifiés manuellement, accompagné par PhaseNet en aide à la décision. La localisation de ces événements a été réalisée à partir d'une procédure non-linéaire décrite par Haugmard (2016). Cette méthode permet de déterminer les paramètres hypocentraux du tremblement de terre mais également les paramètres structuraux du milieu, en s'affranchissant ainsi de connaissances *a priori*. Cette méthode de localisation est bien adaptée aux domaines intracontinentaux où les structures crustales sont peu connues, tel que le Massif armoricain. Quatre paramètres hypocentraux sont déterminés : la longitude, la latitude, la profondeur et le temps d'origine du séisme, ainsi que quatre paramètres structuraux : la vitesse des ondes P dans la croûte (V_c) et dans le manteau (V_m), la profondeur du Moho (z_{moho}) et le rapport entre la vitesse des ondes P et S ($\frac{V_p}{V_s}$).

Parmi les 541 événements sismiques naturels, 391 ont pu être pointés à suffisamment de stations pour être localisés. Ces 391 nouveaux événements sismiques naturels correspondent à l'ensemble C_{H2023} dans le reste du manuscrit. La Figure 5.10 présente les localisations des événements sismiques nouvellement détectés. En première approche, la distribution des épicentres dans la région armoricaine semble être dispersée. Bien qu'il y ait quelques regroupements de séismes (comme l'essaim sismique de Vannes et les répliques du séisme du Layon, Fig 5.10), ils ne sont pas localisés le long de tracés de failles cartographiées. La Figure 5.11 montre la carte de densité d'événements sismiques dans la région. Les séismes semblent être principalement localisés selon une direction NW-SE avec une forte densité d'événements au niveau du Morbihan.

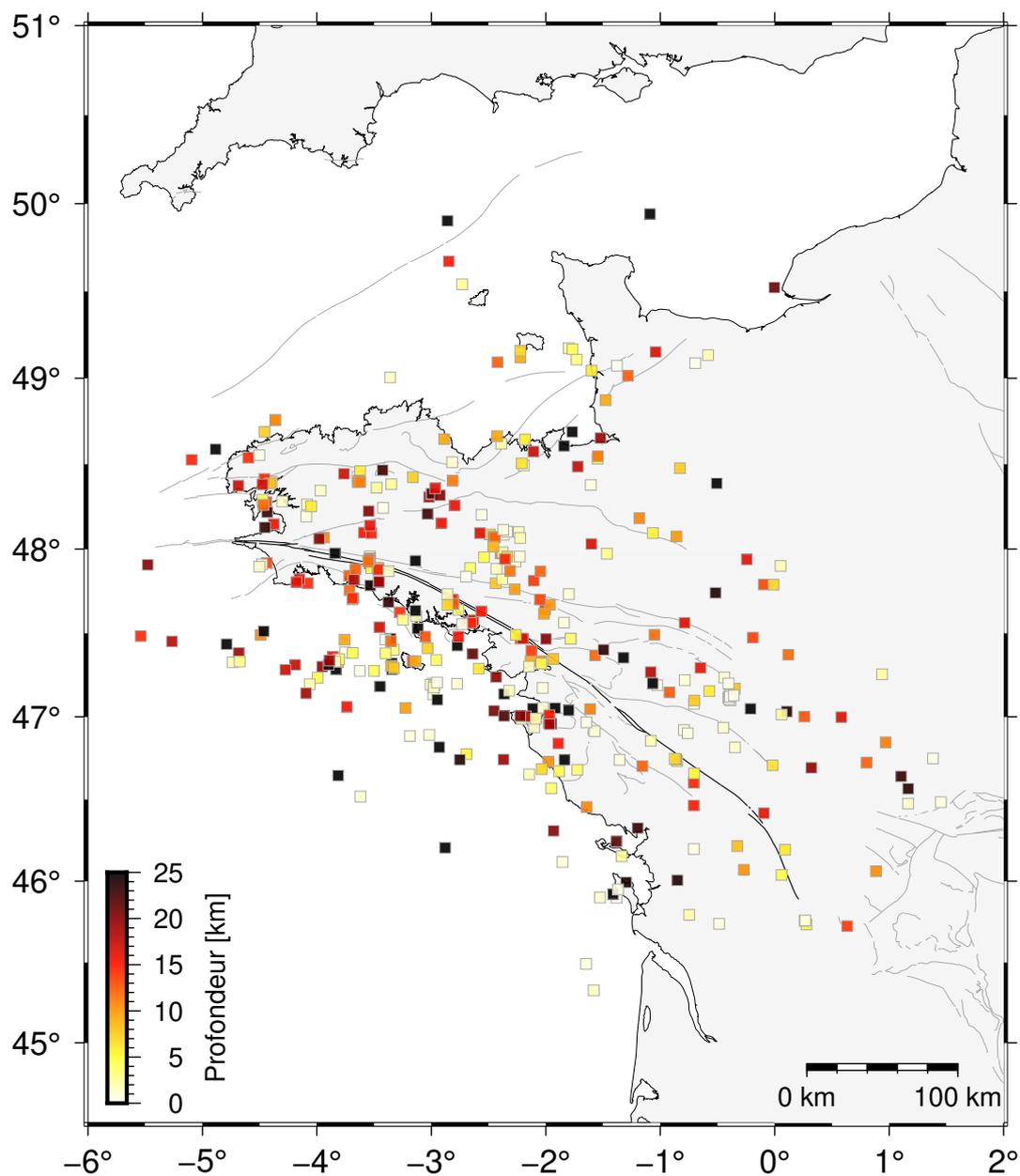


FIGURE 5.10 – Localisation des 391 nouveaux séismes naturels détectés (C_{H2023}). La couleur de chaque événement fait référence à la profondeur. Les tracés gris correspondent aux failles majeures présentes dans la région et le tracé noir représente le Cisaillement Sud Armoricain.

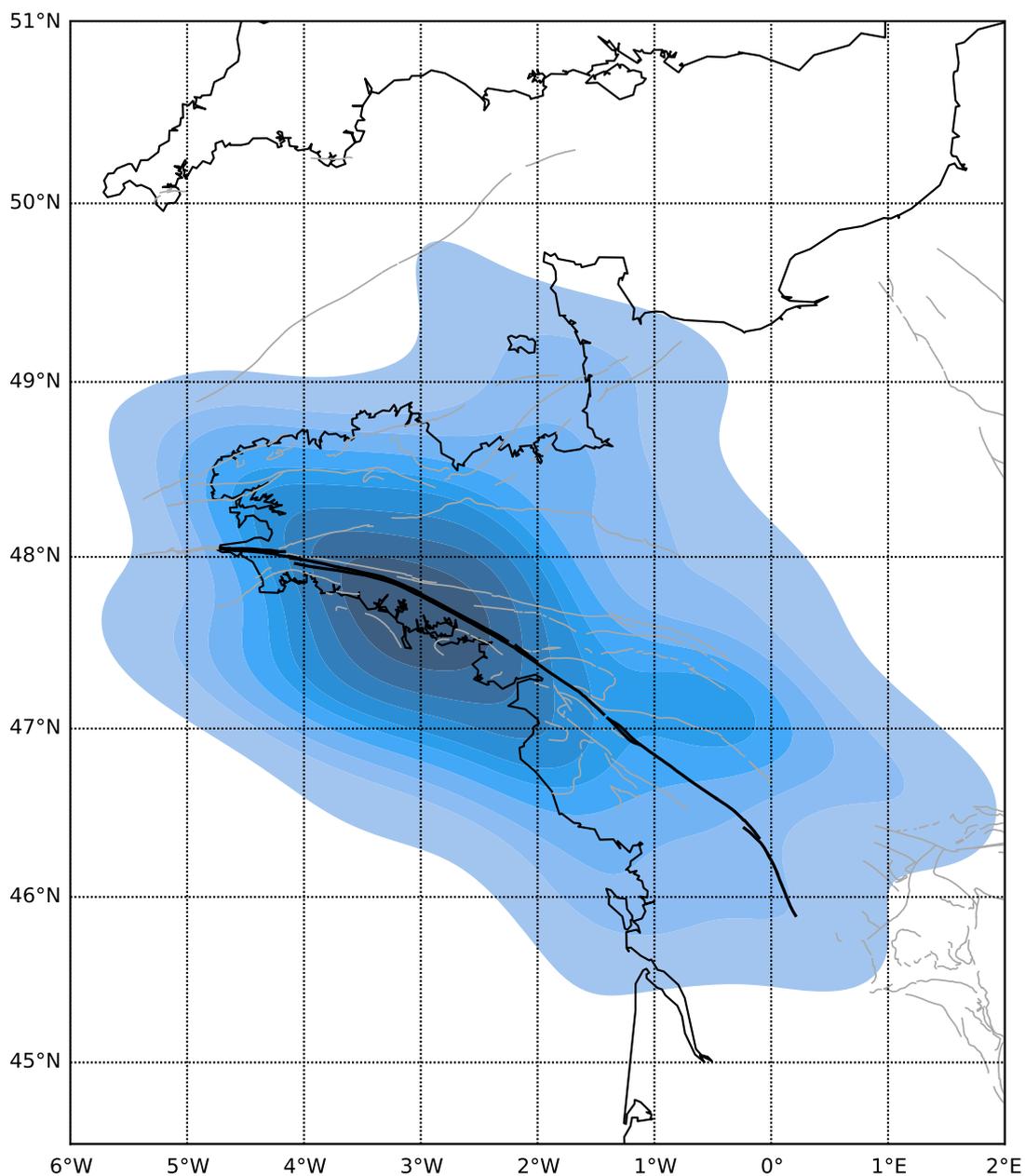


FIGURE 5.11 – Carte de densité des 391 nouveaux séismes naturels détectés (CH_{2023}).

5.2.2 Nouvelle carte de sismicité de 2018 à 2021

L'application des méthodes de détection et de discriminations a permis de détecter 391 nouveaux événements sismiques qui ont été ensuite localisés (Fig 5.10). En combinant ces résultats avec les 412 événements détectés par le BCSF-RéNaSS pendant la même période, appelé C_{BCSF} , nous obtenons un total de 803 événements sismiques naturels détectés dans la région. Afin de comparer efficacement les différents paramètres de localisation (particulièrement en terme de profondeur, et de magnitude), les 412 événements détectés par le BCSF-RéNaSS ont été repointés et relocalisés à partir de l'approche décrite par Haugmard (2016).

Grâce à cette application ($C_{BCSF} + C_{H2023}$), le nombre de séismes naturels détectés et localisés a ainsi presque doublé, passant d'environ 1,9 à 3,8 séismes détectés et localisés par semaine. En prenant en compte l'ensemble des détections réalisées dans le cadre de cette étude (541), nous atteignons un taux de 4,5 séismes par semaine. Pour le reste de la discussion, nous prendrons uniquement en compte les 391 séismes du catalogue C_{H2023} .

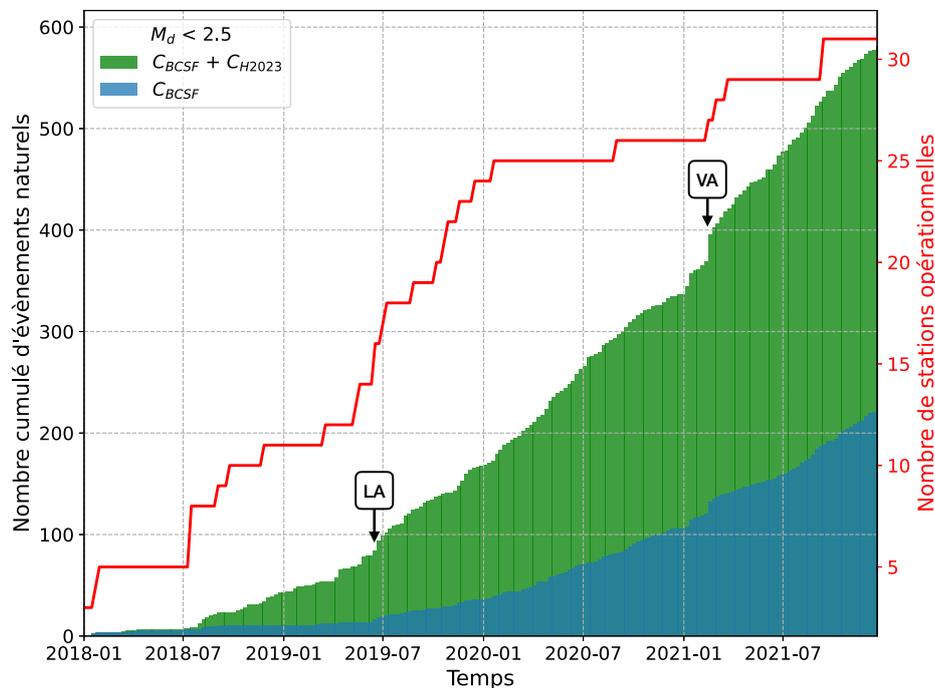


FIGURE 5.12 – Histogramme cumulatif des événements naturels de $M_d \leq 2,5$, avec en bleu les événements du catalogue BCSF-RéNaSS, C_{BCSF} , et en vert $C_{BCSF} + C_{H2023}$, en fonction du temps. Le nombre de stations opérationnelles permanentes, dans la région d'étude, est représentée en rouge (axe de droite). Acronymes des événements : LA (21 juin 2019, Layon), VA (17 au 21 février 2021, Vannes).

La Figure 5.12 montre un histogramme cumulatif des événements naturels de $M_d \leq 2,5$ détectés dans la région. À partir de mi-2018, grâce au déploiement permanent des stations de Résif-Epos dans l’ouest, la quantité d’événements naturels a augmenté de manière significative. Par ailleurs, nous observons plusieurs sauts dans les détections, notamment lors d’un choc principal suivi de répliques (comme le séisme du Layon le 21 juin 2019) ou lors d’un essaim sismique (comme l’essaim de Vannes, du 17 au 21 février 2021). Ces événements particuliers entraînent une augmentation significative des détections. Nous constatons que, pour les détections effectuées par le BCSF-RéNaSS, l’augmentation ne semble pas être linéaire. En revanche, en ajoutant les détections supplémentaires, on observe que cette tendance devient linéaire et peut être interprétée comme une limite de détection.

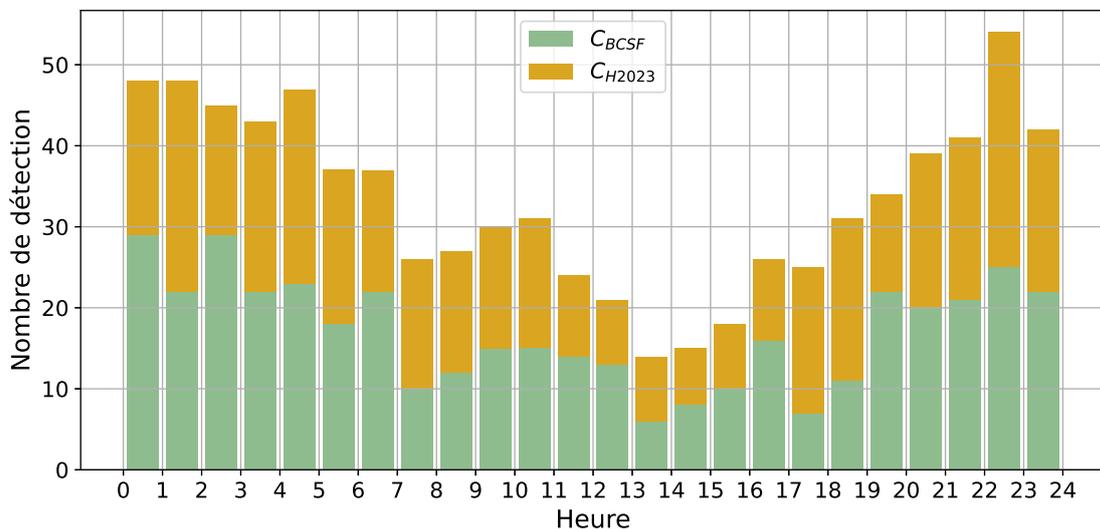


FIGURE 5.13 – Distribution du nombre d’événements sismiques détectés en fonction des heures de la journée avec, en vert, les séismes du catalogue du BCSF-RéNaSS (C_{BCSF}) et en orange, les nouvelles détections (C_{H2023}).

Au sein des détections effectuées par le BCSF-RéNaSS, 64% de ces événements se produisent entre 18 heures et 6 heures du matin (Fig 5.13). Ceci s’explique par la présence de bruits d’origine humaine, pendant la journée (par exemple [Atef et coll., 2009](#); [Hao et coll., 2019](#)). Durant la nuit, le niveau de ce bruit environnemental diminue, ce qui facilite la détection des événements sismiques. En ce qui concerne les détections supplémentaires, environ la même proportion d’événements (62%) se produit pendant la nuit. Les séismes naturels peuvent survenir à n’importe quel moment de la journée, mais nous observons un déficit de détections pendant les heures diurnes, en particulier aux moments où l’activité humaine est intense. Ceci confirme que les bruits d’origines anthropiques ont un impact significatif sur la capacité de détection des séismes de faible

magnitude. Certaines périodes de la journée semblent cependant être comblées pour ces nouvelles détections, comme par exemple, le matin entre 7 heures et 11 heures, comme l'illustre la Figure 5.13. Ces résultats sont extrêmement encourageants pour l'amélioration de la détection des séismes, mais il est important de souligner qu'il reste difficile de se défaire complètement des effets des signaux parasites enregistrés à chaque station.

Parmi l'ensemble C_{BCSF} , seulement 22% des événements sont de magnitude (M_d) inférieure à 2. En revanche, 63% des séismes de l'ensemble C_{H2023} présentent une M_d inférieure à 2. Grâce à notre méthode de détection, la proportion de séismes de faible magnitude a ainsi augmenté. Il y a désormais deux fois plus d'événements de magnitudes inférieures à 2 dans le catalogue. Cette amélioration de la capacité de détection des faibles magnitudes, visible en Figure 5.14, est un résultat positif. Cela signifie que notre méthode est plus sensible à la détection des séismes de faible magnitude. En l'état actuel, il n'est pas possible d'actualiser la Figure 2.6 avec les nouvelles détections car nous ne disposons pas des magnitudes de moment pour ces événements. L'application de cette approche permet donc d'élargir la gamme de magnitudes détectées dans la région.

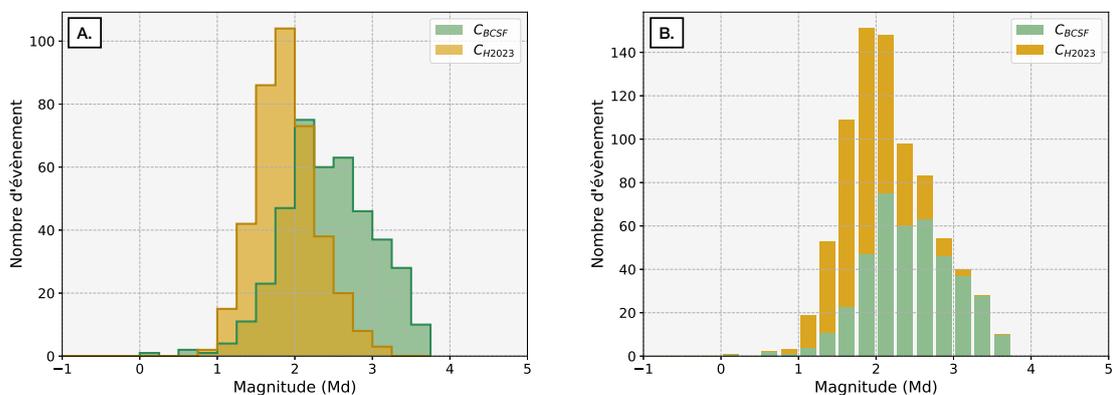


FIGURE 5.14 – Distributions des magnitudes des événements sismiques détectés entre le 1er janvier 2018 et le 31 décembre 2021. (A) Distributions des magnitudes pour les événements détectés dans le cadre de la thèse (C_{H2023} , en orange) et pour les événements présents dans le catalogue du BCSF-RéNaSS durant la même période (C_{BCSF} , en vert). (B) Histogramme cumulé des magnitudes pour les deux ensembles d'événements.

L'ensemble des séismes localisés présentent des profondeurs focales qui s'étendent depuis la surface jusqu'à environ 40 km (Figure 5.15). Cependant, il convient d'être prudent lors de l'interprétation de la profondeur des séismes. Les études antérieures ont montré que le Moho peut se situer à une profondeur d'environ 30 km (par exemple Chevrot *et coll.*, 2014). La profondeur focale des tremblements de terre est, en général,

mal contrainte, particulièrement pour les petits séismes, en raison du réseau de stations dispersé sur l'ensemble de région. Quoi s'il en soit, on observe une diminution du nombre de séismes avec la profondeur. Dans la Figure 5.15, deux maxima sont observés : un premier très superficiel, se situant dans les deux premiers kilomètres, et un second maximum plus profond (entre 10 et 15 km). Ce second maximum est renforcé par l'essai sismique du 17 au 21 février 2021 localisé au nord de Vannes où 12 événements supplémentaires sont détectés et localisés entre 8 et 15 km de profondeur. Cette distribution multimodale des profondeurs des séismes rappelle la distribution obtenue au cours de l'étude de Haugmard (2016) (voir Fig IV.1-10 de Haugmard, 2016) durant laquelle 314 séismes de 2011 à 2013 ont été relocalisés dans le Massif armoricain . La grande proportion d'événements localisés à des profondeurs très superficielles est une autre particularité que l'on observe également dans d'autres régions continentales stables (Klose et Seeber, 2007; Jackson et McKenzie, 2022).

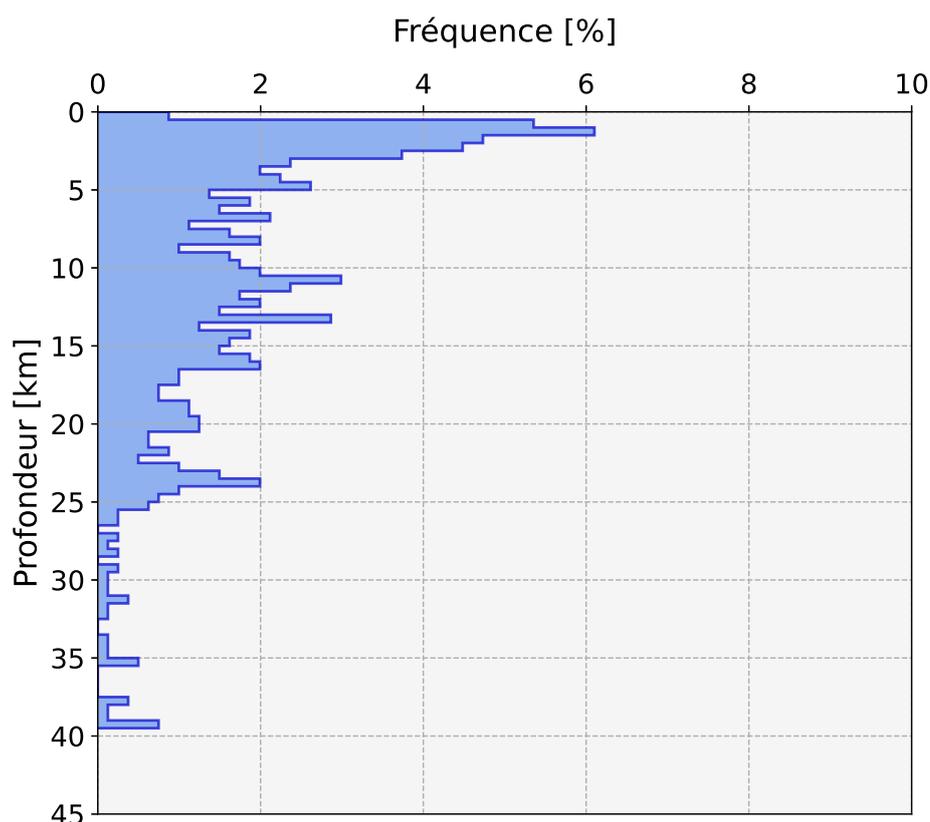


FIGURE 5.15 – Distribution des profondeurs des séismes des 803 séismes naturels détectés du 1er Janvier 2018 au 31 Décembre 2021.

La Figure 5.16 représente la localisation de l'ensemble des événements détectés au cours de ces 4 années d'étude ($C_{BCSF} + C_{H2023}$). Nous remarquons que le caractère diffus de la sismicité de la région semble être confirmé. Bien qu'il y ait quelques regrou-

pements de séismes (comme l'essai sismique de Vannes et les répliques du séisme du Layon, Fig 5.10), ils ne sont pas localisés le long de tracés de failles existantes. Nous pouvons tout de même constater une bande de direction NW-SE d'environ 100 km de large où l'activité sismique est plus importante.

L'application de cette approche sur 4 années de données a donc permis de doubler le nombre de détections dans la région du Massif armoricain. Cette augmentation considérable du nombre d'événements détectés, venant compléter le catalogue existant, offre une occasion précieuse d'approfondir notre compréhension des origines de cette sismicité dans une région continentale stable. En explorant ces nouvelles détections et en appliquant cette approche sur une plus longue période d'étude, nous pouvons rassembler quelques éléments de réponse sur les facteurs responsables de l'activité sismique dans le Massif armoricain.

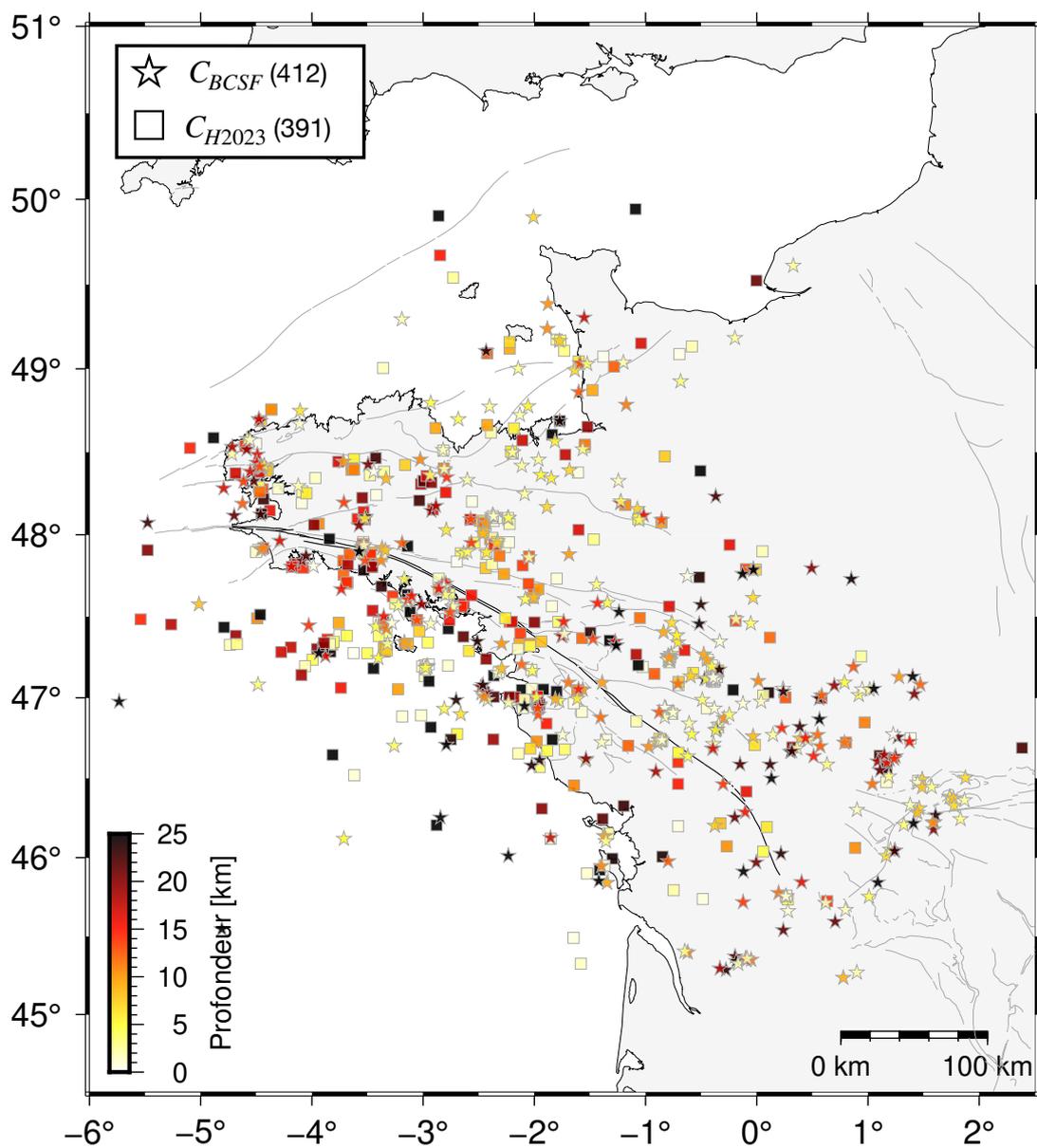


FIGURE 5.16 – Localisation des 803 séismes naturels détectés entre le 1er Janvier 2018 et le 31 Décembre 2021. La couleur de chaque événement fait référence à sa profondeur. Les tracés gris correspondent aux failles majeures présentes dans la région et le tracé noir représente le Cisaillement Sud Armoricain.

5.3 Études préliminaires sur l'origine de la sismicité

Dans la suite de ce manuscrit, nous réalisons une étude préliminaire de certains facteurs pouvant être responsables de l'activité sismique dans le Massif armoricain. Dans un premier temps, nous analysons la distribution spatiale de la sismicité afin d'explorer les liens éventuels entre les structures tectoniques et la localisation des tremblements de terre dans cette zone. Puis, nous étudions les perturbations locales des contraintes, qui peuvent être régies par des facteurs extérieurs tels que les épisodes météorologiques et les variations de marée. Nous étudions comment ces variations peuvent influencer la sismicité dans la région et cherchons à déterminer leur rôle dans la genèse des séismes.

5.3.1 Distribution spatiale de la sismicité

Le nord-ouest de la France fait l'objet d'un fort héritage structural. L'érosion progressive a exposé des structures déformées à la surface. Les roches exposées, telles que les schistes, les quartzites, les granites et les gneiss, témoignent des forces géologiques qui ont influencé la région. De nombreuses failles, visibles et cartographiées en surface, forment une zone d'endommagement majeur de la croûte terrestre. La Figure 5.17 montre un comparatif entre la localisation des failles cartographiées et la distribution de la sismicité de la région. Comme expliqué précédemment, la localisation des tremblements de terre ne semble pas suivre, à première vue, les structures visibles en surface. L'activité sismique semble pour autant principalement localisée au niveau d'une ceinture de 100 km de large, orientée NW-SE et de 600 km de long.

Afin d'analyser la distribution moyenne des événements dans la région et d'identifier si une structure géologique coïncide avec cette sismicité, nous avons appliqué une méthode de régression sur les épicentres des séismes détectés. Pour cela, nous avons pris en compte l'ensemble des événements détectés du 1er janvier 2018 au 31 décembre 2021 (C_{BCSF} et C_{H2023}) en complétant ce catalogue par les séismes détectés, par le BCSF-RéNaSS, du 1er janvier 2022 au 1er mars 2023. Nous avons choisi d'utiliser les événements de 2018-2023 car, durant cette période, nous disposons de la plus grande densité de stations disponibles dans la région impliquant une meilleure localisation et détection des événements de faible magnitude. La Figure 5.18 montre la distribution des épicentres dans le Massif armoricain et les résultats obtenus par la méthode de régression grâce à deux polynômes du second et du troisième degré. Nous observons que ces deux courbes (orange et rouge) coïncident exactement, à l'ouest, avec le CSA puis

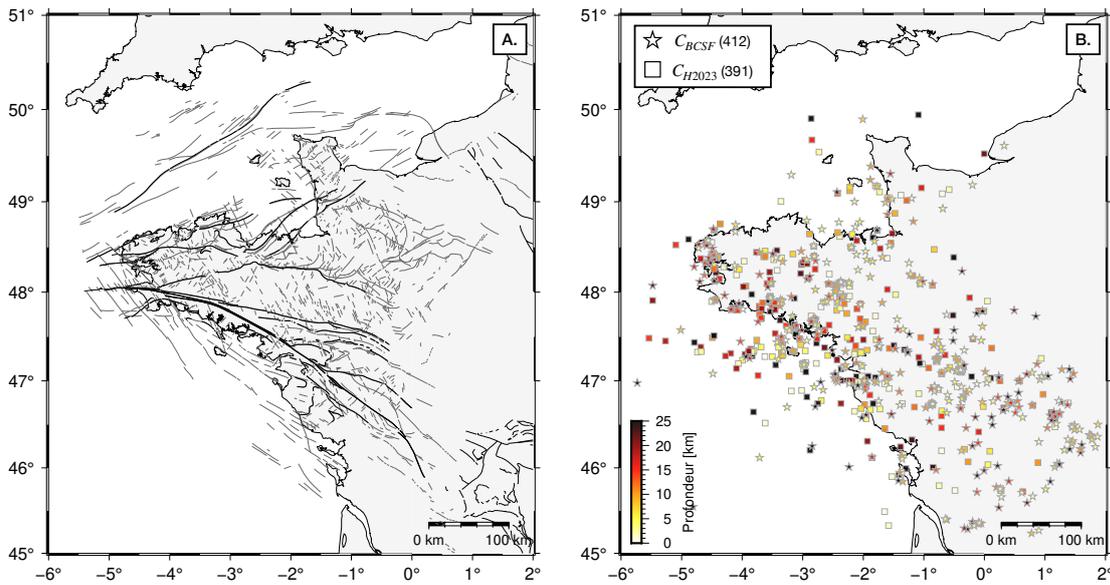


FIGURE 5.17 – Localisation des failles et événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023. (A) Localisation des failles cartographiées de la région. Les tracés noirs correspondent aux failles majeures de la région. (B) Localisation des événements détectés entre le 1er janvier 2018 et le 1er mars 2023 par le BCSF-RéNaSS (C_{BCSF}) et lors de cette étude (C_{H2023}).

passent entre ses deux branches vers l'est. Au niveau de cette zone, la branche sud du cisaillement se subdivise en plusieurs ramifications vers le seuil du Poitou. La présence de cette tendance indique donc que le CSA joue possiblement un rôle significatif dans l'activité sismique observée dans la région. Cependant, il est important de noter que ces polynômes n'expliquent pas la totalité de la distribution des séismes.

Formé durant l'orogénèse varisque, le CSA se caractérise par une direction moyenne de N90-130, exerçant des mouvements horizontaux de cisaillement sur la croûte terrestre. Au fil du temps, les mouvements le long du CSA et l'érosion différentielle ont entraîné des variations topographiques majeures dans la région. L'érosion différentielle se produit lorsque des matériaux ou des formations géologiques sont érodés de manière inégale en raison de différences de résistance, de composition ou de structure. Ce processus peut entraîner la formation de reliefs et de caractéristiques topographiques distinctes dans une région.

La Figure 5.19 montre la répartition géographique des séismes autour du CSA. Nous constatons une distribution intéressante des événements sismiques autour de ce dernier. Plus nous nous rapprochons du CSA, plus le nombre d'événements sismiques augmente. Bien que les séismes ne semblent pas se répartir le long des failles cartographiées, à première vue, une zone de sismicité plus dense est observée autour de ce cisaillement

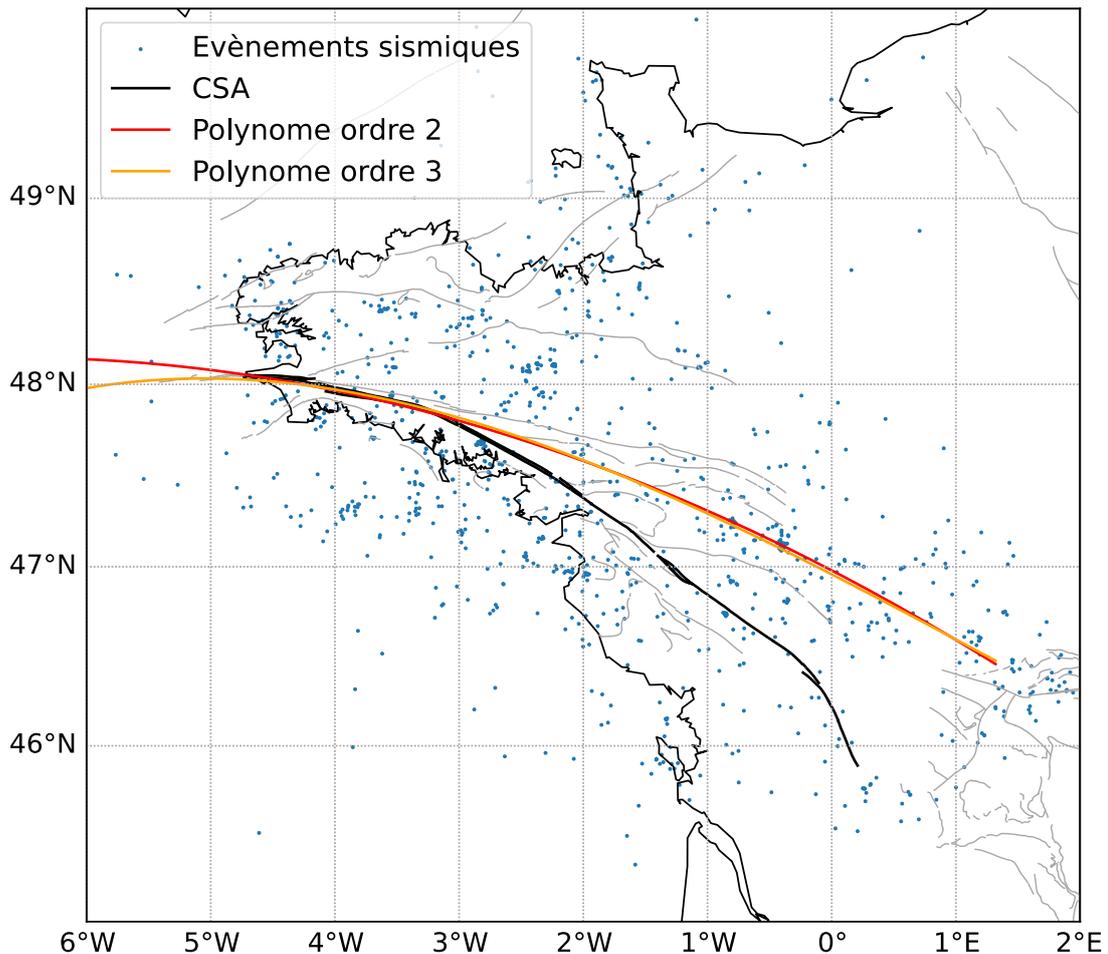


FIGURE 5.18 – Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023. Le tracé noir correspond au CSA et les tracés gris représentent les failles majeures de la région. Les courbes rouge et orange correspondent aux polynômes obtenus à partir de la localisation de l'ensemble des événements.

(Fig 5.11). Une distribution bimodale est observée avec un maximum juste au sud du cisaillement et un second à environ 60 km au nord du cisaillement. Nous remarquons également une forte diminution du nombre de séismes juste au nord du CSA, formant un creux dans la distribution des tremblements de terre.

Les Figures 5.20 et 5.21 présentent la distribution des tremblements de terre sur des sections spécifiques perpendiculaires au CSA. Nous observons des similitudes dans la distribution des épencentres. Dans la Figure 5.20, le maximum semble être localisé au sud du CSA, au niveau de Vannes. Le second maximum semble quant à lui correspondre au tracé CNA. Dans la Figure 5.21, nous remarquons une sismicité notable entre les deux branches du CSA où des ramifications sont cartographiées.

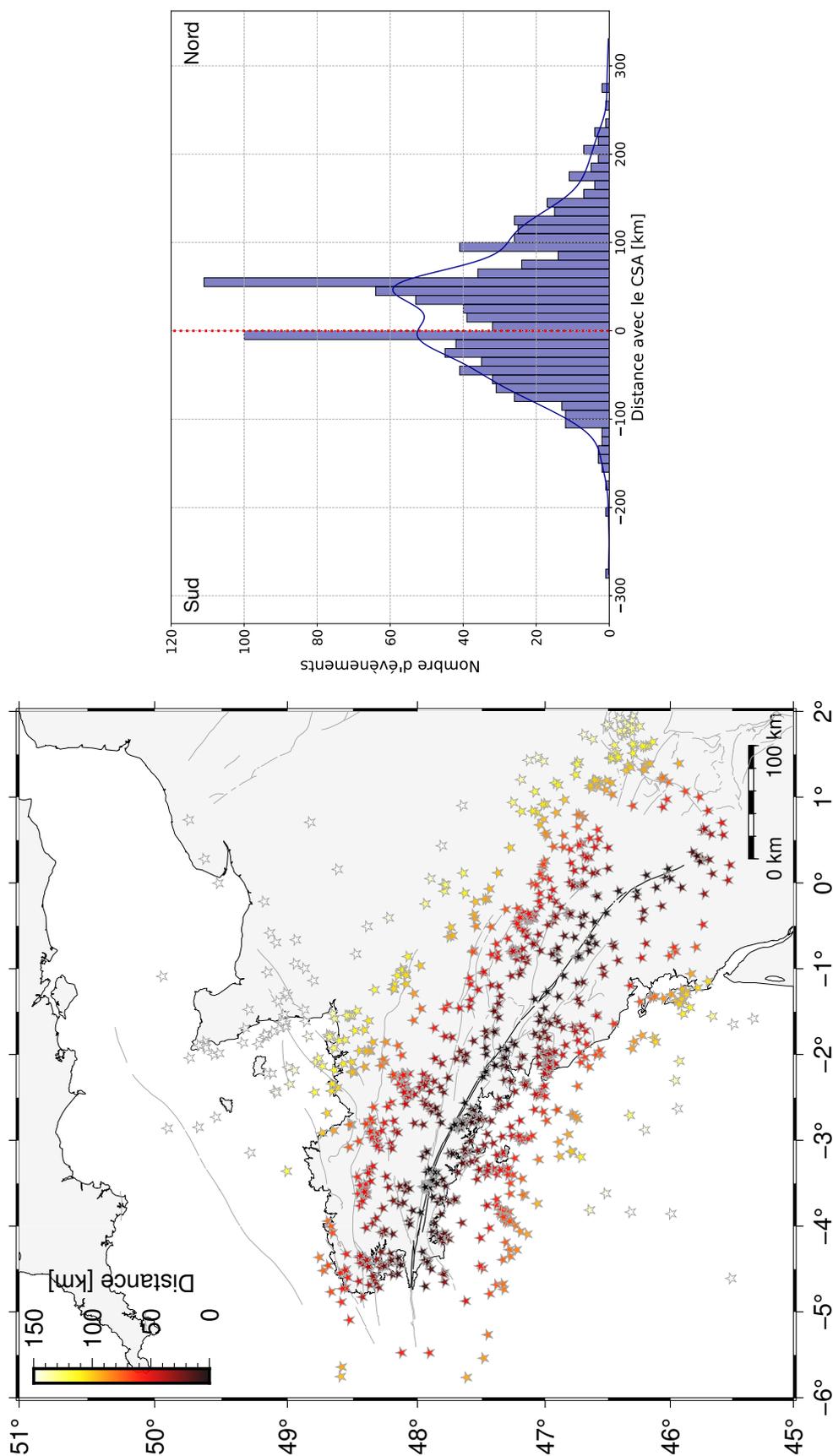


FIGURE 5.19 – Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023. La couleur symbolise la distance de chaque événement avec le CSA, représenté par le tracé noir. L'histogramme représente la distribution du nombre d'événements en fonction de la distance avec le CSA. La courbe bleue correspond au kernel density estimation de la distribution.

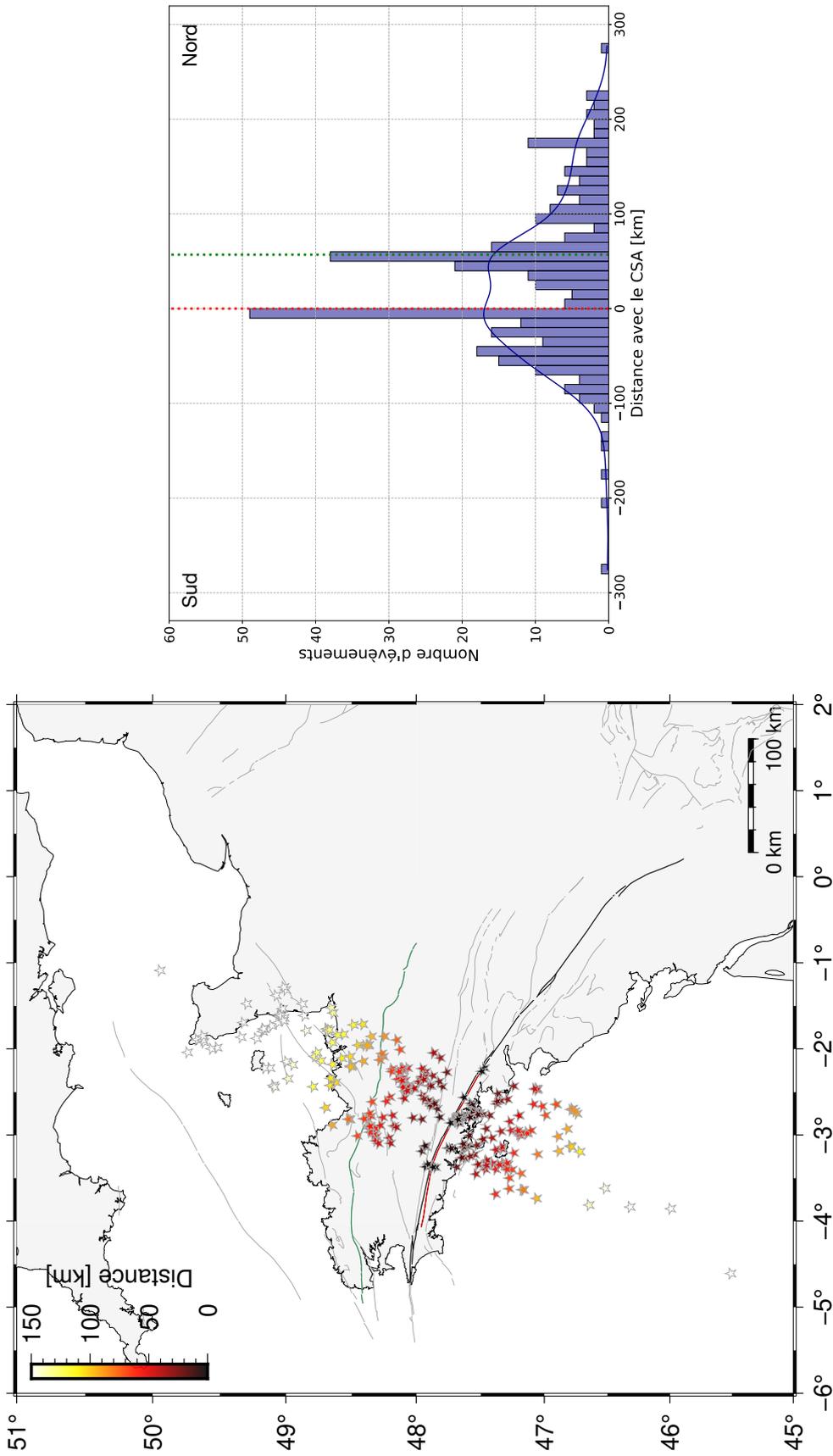


FIGURE 5.20 – Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023 sur une section de 100 km à l'ouest du CSA. La couleur symbolise la distance de chaque événement avec le CSA, représenté par le tracé rouge. Le CNA est symbolisé par le tracé vert. L'histogramme représente la distribution du nombre d'événements en fonction de la distance avec le CSA. La courbe bleue correspond au *kernel density estimation* de la distribution.

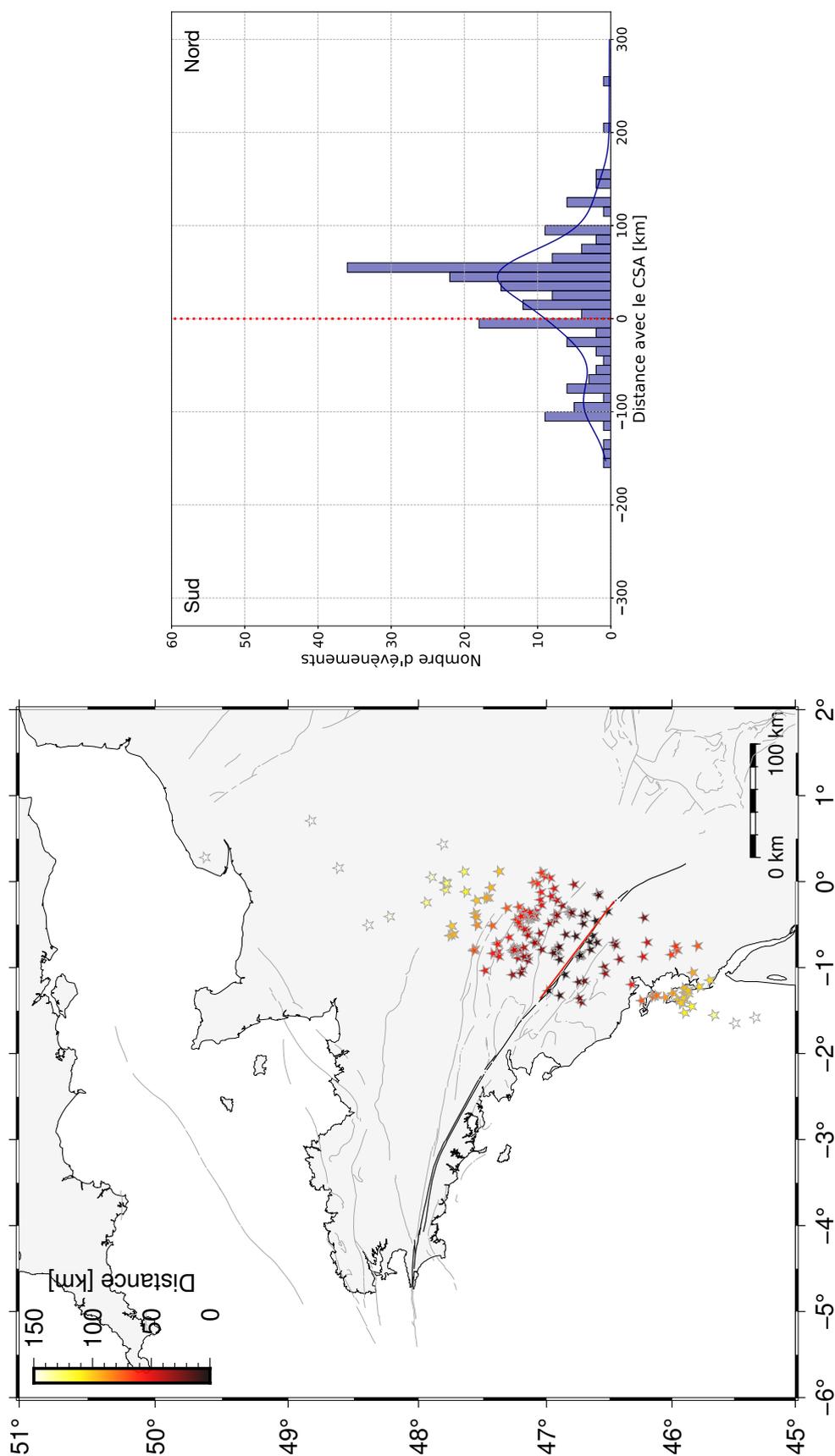


FIGURE 5.21 – Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023 sur une section de 100 km à l'est du CSA. La couleur symbolise la distance de chaque événement avec le CSA, représenté par le tracé rouge. L'histogramme représente la distribution du nombre d'événements en fonction de la distance avec le CSA. La courbe bleue correspond au *kernel density estimation* de la distribution.

Cette répartition des tremblements de terre autour du CSA et du CNA montre l'influence notable de ces structures sur la sismicité de la région. Cependant, cette concentration d'événements sismiques près des cisaillements ne signifie pas nécessairement que tous les séismes sont directement liés à ces structures. D'autres facteurs géologiques et tectoniques peuvent également contribuer à la répartition complexe des séismes dans la région du Massif armoricain.

Nous pouvons nous interroger sur la distribution des séismes le long de failles majeures dans d'autres régions du monde : suit-elle nécessairement un schéma linéaire ou peut être plus diffuse ? Dans plusieurs régions continentales stables, différents comportements sont observés. En Australie, une des régions les plus actives sur le plan sismique se trouve dans le sud-ouest du pays (Denham, 1988). La distribution de la sismicité dans cette région est complexe et semble suivre une tendance NW-SE. Il existe également des zones de sismicité accrues très localisées (Dentith et Featherstone, 2003). Au nord-ouest de la péninsule ibérique, la sismicité instrumentale n'est pas associée aux grandes failles actives et semble par endroit être distribuée sur une large zone, avec une orientation générale de NW-SE (Martín-González *et coll.*, 2023).

Dans la majeure partie de la France et de l'Europe occidentale, la contrainte compressive horizontale maximale présente une orientation générale NW-SE à NNW-SSE (azimut d'environ N135-160°). Cette configuration des contraintes semble être principalement due à la force de poussée de la dorsale médio-atlantique combinée à la résistance à la subduction/collision le long des limites Nubie-Eurasie et Adria-Eurasie (Müller *et coll.*, 1992; Gölke et Coblenz, 1996; Mazzotti *et coll.*, 2020). Cette orientation NW-SE à NNW-SSE à l'échelle continentale de la contrainte de compression horizontale maximale est cohérente avec les tenseurs de contrainte locaux dérivés des inversions des mécanismes au foyer dans la majeure partie de la France et dans notre région d'étude.

L'ensemble de ces éléments ne nous permet pas d'expliquer précisément les origines d'une telle sismicité dans ce domaine continental stable. Nous mettons néanmoins en évidence une plus forte activité sismique localisée Sud-Bretagne, proche du CSA. Il est envisageable que ces séismes puissent s'expliquer également par des perturbations transitoires des contraintes locales ou de la résistance des failles au niveau de cette zone d'endommagement. Ces perturbations pourraient déclencher la libération de l'énergie élastique accumulée au sein des failles.

5.3.2 Facteurs environnementaux

5.3.2.1 Épisodes météorologiques

Comme introduit dans la Section 2.6, la sismicité d'une région peut être reliée à des perturbations locales des contraintes qui peuvent être associées à de nombreux facteurs agissant à différentes échelles de temps. Parmi ces facteurs, une réponse mécanique à de très courts épisodes (jours à mois) transitoires hydrologiques ou météorologiques peut potentiellement expliquer des épisodes sismiques (par exemple [Roeloffs, 1988](#); [Bollinger et coll., 2007](#); [Steer et coll., 2014](#); [Leclère et Calais, 2019](#)). Des changements plus ou moins brusques dans la pression de l'eau souterraine ou dans les niveaux de précipitations peuvent induire des modifications de l'état de contrainte des failles actives, ce qui peut éventuellement déclencher des séismes.

Dans un premier temps, nous nous focalisons sur la saisonnalité des événements sismiques. Bien que la plupart des séismes peuvent se produire à tout moment, il existe des cas où une certaine saisonnalité peut être observée ([Bollinger et coll., 2007](#)). Dans certaines régions, l'activité sismique peut être influencée par des phénomènes saisonniers tels que la fonte des neiges, les changements de pression dans les aquifères ou les variations du niveau d'eau dans les lacs. Ces changements peuvent modifier les contraintes sur les failles et potentiellement déclencher des séismes. De plus, les variations climatiques saisonnières peuvent également jouer un rôle dans l'activité sismique. Par exemple, des régions sujettes aux fortes précipitations saisonnières peuvent connaître des changements dans la distribution de l'eau souterraine, ce qui peut affecter les contraintes sur les failles et augmenter la probabilité de séismes. La Figure 5.22 montre la variation du nombre de séismes, dans l'ensemble du Massif armoricain, au cours de chaque mois entre janvier 2019 et février 2023. Nous observons une légère saisonnalité avec une activité sismique plus faible durant les mois de novembre et décembre. Lorsque nous effectuons la même analyse pour les séismes localisés à une profondeur inférieure à 5 km, la saisonnalité n'est pas, non plus, nettement marquée. Il faut noter que cette diminution du nombre de séismes durant l'hiver peut également être due à une diminution du nombre de détections en raison de l'augmentation du bruit ambiant (activité océanique plus marquée). Les événements sismiques de faible magnitude, qui seraient normalement détectés et enregistrés en d'autres saisons, peuvent ainsi être masqués par le bruit de fond pendant l'hiver.

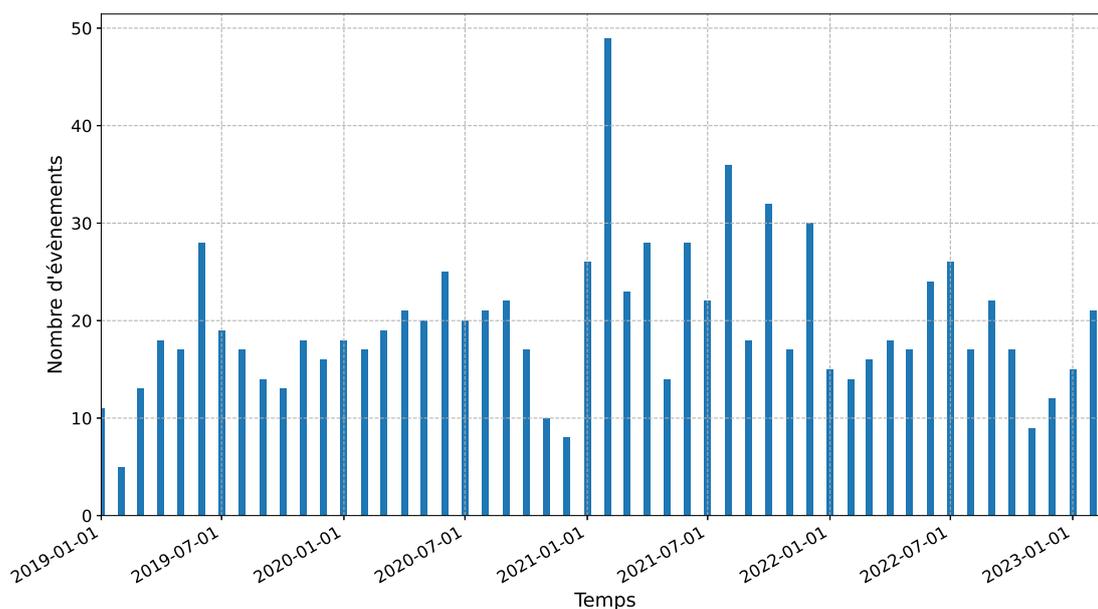


FIGURE 5.22 – Variation du nombre d'événements sismiques naturels pour chaque mois entre Janvier 2019 et Février 2023.

Afin de tenter une autre approche permettant d'illustrer les effets des conditions météorologiques sur l'activité sismique du nord-ouest de la France, nous avons décidé de focaliser notre étude sur des observations locales. Différentes études montrent que le déclenchement d'essai sismique peut être en lien avec des changements de pression interstitielle dus aux mouvements des fluides dans les zones d'activité sismique (par exemple [Kurz et coll., 2004](#)). [Kraft et coll. \(2006\)](#) montrent, par exemple, une corrélation significative entre la sismicité, les précipitations et l'augmentation du niveau des eaux souterraines, dans la zone de l'essai de tremblements de terre de Bad Reichenhall, dans le sud-est de l'Allemagne.

Nous avons donc choisi d'étudier l'essai sismique du 17 au 21 février 2021, proche de Vannes et situé au sud du CSA. Durant cette séquence aucun « choc principal » n'a été enregistré, seule une série d'événements sismiques de magnitude de durée moyenne égale à 1,8 a été détectée et un séisme de plus forte magnitude (M_{lv} de 2,8) a été détecté le 17 février au soir. À l'heure actuelle, aucune explication sur l'origine de cet essai sismique n'est proposée.

Afin de tenter d'analyser le lien avec cette sismicité et la précipitation, nous utilisons les enregistrements disponibles de la station météorologique la plus proche située à Belle-Île à environ 50 km au sud de Vannes. La Figure 5.23 présente les valeurs de précipitations enregistrées à la station météorologique de Belle-Île et la distribution

des événements sismiques situés dans un rayon de 75 km autour de la station. Nous remarquons que les semaines précédant l'essai sismique, une quantité relativement faible de précipitations a été enregistrée à la station. D'autres épisodes de précipitations ont également eu lieu au cours des mois précédant l'essai sismique, mais sans générer, *a priori*, une augmentation notable du nombre de séismes aux alentours de la station météorologique.

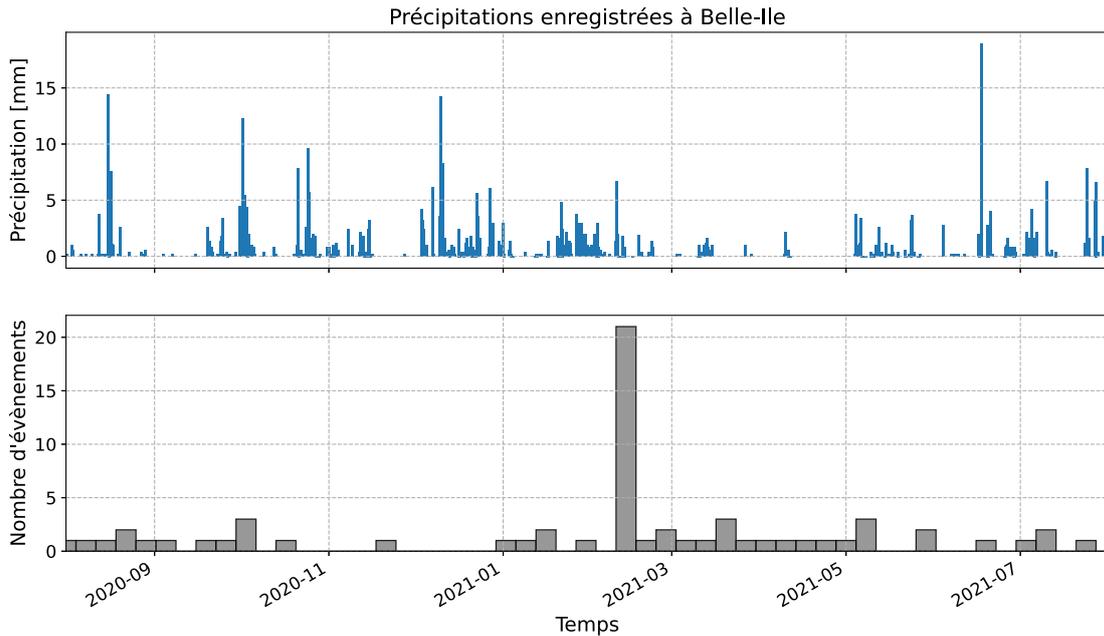


FIGURE 5.23 – Distribution des précipitations enregistrées à la station Belle-Île et événements sismiques naturels détectés dans un rayon de 75 km de la station du 1er Août 2020 au 1er Août 2021.

Ces observations préliminaires suggèrent que la corrélation entre les niveaux de précipitations et l'activité sismique n'est pas évidente, dans cette zone de sismicité restreinte. Cette hypothèse ne doit cependant pas pour autant être exclue. En détectant un plus grand nombre d'événements sismiques dans la région, nous pourrions potentiellement mettre en évidence une certaine activité sismique en lien avec des épisodes transitoires météorologiques en effectuant des analyses statistiques plus approfondies sur d'autres essais sismiques.

5.3.2.2 Forçages périodiques

La marée, en tant que phénomène lié à l'interaction gravitationnelle entre la Terre, la Lune et le Soleil, peut exercer une influence sur certains processus géologiques, y compris, parfois, la sismicité (par exemple [Cochran *et coll.*, 2004](#); [Thomas *et coll.*, 2012](#)).

Cette influence est généralement considérée comme subtile et relativement faible par rapport à d'autres facteurs qui affectent plus directement l'activité sismique. Cependant, dans notre domaine continental stable, ces effets de forçage peuvent être mis en évidence. L'influence potentielle de la marée sur la sismicité est principalement attribuée à l'effet de chargement provoqué par les variations de pression sur les failles et les systèmes de contraintes tectoniques. Ces variations peuvent théoriquement affecter le comportement des failles, modifiant ainsi les conditions de contrainte et augmenter la probabilité de déclenchement des séismes. Des moments spécifiques tels que les marées de vives-eaux (moments où les coefficients de marée sont les plus forts) pourraient présenter une plus grande probabilité de déclenchement de séismes.

Le quart nord-ouest de la France fait l'objet d'un contexte géographique et géologique particulier. Comme expliqué précédemment, le Massif armoricain fait l'objet d'un fort héritage tectonique, impliquant une grande zone d'endommagement en surface. De plus, la côte bretonne est exposée à des marnages significatifs en raison de sa position géographique et de la configuration de ses côtes. Le marnage fait référence à la différence de hauteur entre les marées hautes et basses dans un endroit donné, généralement mesurée par rapport à un point de référence tel que le niveau moyen de la mer. Il représente l'amplitude des variations de niveau de la mer dues à l'attraction gravitationnelle de la Lune et du Soleil, combinée aux effets des forces centrifuges résultant de la rotation de la Terre. C'est au nord-est de la région que se trouvent les marées les plus fortes d'Europe. Dans la baie du Mont-Saint-Michel, ils peuvent atteindre 14 mètres.

Pour étudier l'influence des forces de marée sur la zone d'étude, nous pouvons étudier les positions de masses des sismomètres. La position de masse du sismomètre mesure la tension nécessaire pour garder les masses mobiles dans leur position d'équilibre. La Figure 5.24 montre la position de masse pour deux stations sismologiques (LOCF et SDOF). La station LOCF est localisée au nord de la côte bretonne tandis que SDOF se situe sur la côte sud de la zone. La position de la masse, à chaque station, montre une oscillation d'une période de 6 h qui coïncide avec le signal de marée. De plus, nous observons une oscillation globale d'environ 15 jours qui correspond aux mortes-eaux et vives-eaux.

À certains moments, les signaux sont en opposition de phase (Fig 5.24.A). Ce phénomène est principalement visible, pour ces stations, au moment de mortes-eaux. Ceci est dû à la différence de dynamique océanique entre la Manche au nord et l'océan At-

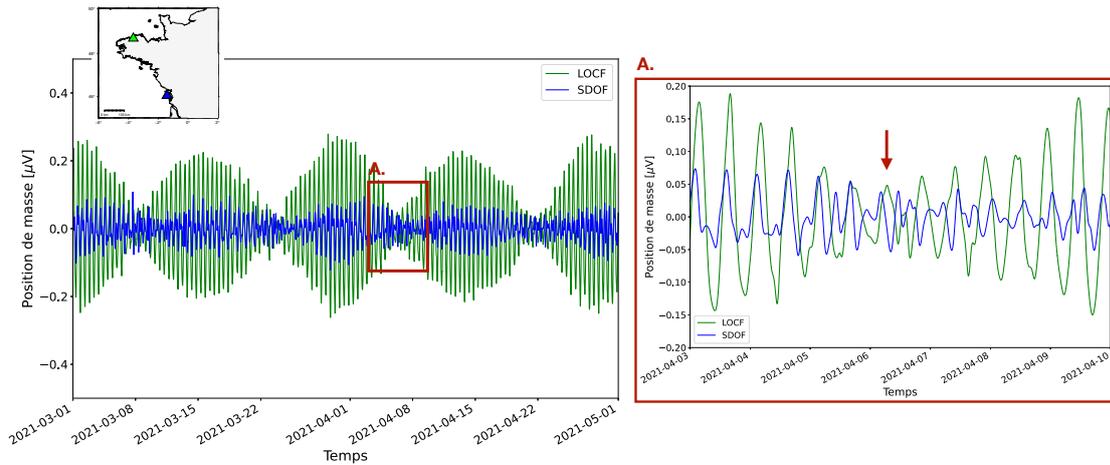


FIGURE 5.24 – Position de masse des stations LOCF et SDOF du 1er mars au 1er mai 2021.

lantique au sud. Ainsi les stations LOCF et SDOF ne subissent pas la marée au même moment. Nous pouvons nous demander si cette légère variation de forces de marée entre le nord et le sud de la région pourrait potentiellement induire une sismicité.

La Figure 5.25 montre la position de masse de la station LOCF et, en rouge, la distribution temporelle des événements détectés en Bretagne. Durant l’année 2020, nous n’observons pas de réelle corrélation entre la sismicité et des moments spécifiques tels que les marées de vives-eaux. Il convient de noter que durant cette année, seule une centaine d’événements ont été détectés dans la région. Une analyse plus approfondie de ces paramètres est nécessaire avec une période d’observation plus longue et plus d’événements sismiques. À une échelle plus locale, nous observons en bleu, la distribution temporelle des événements détectés à moins de 75 km de la station LOCF. Ici encore, nous n’observons pas de réelle corrélation entre la sismicité et des moments spécifiques de l’année.

5.3.3 Bilan des éléments permettant de caractériser la sismicité

Nous avons pu mettre en évidence certains éléments permettant de mieux caractériser la sismicité de notre région d’étude. Dans un premier temps, en observant la distribution des tremblements de terre dans la région, nous avons pu mettre en évidence un zone de forte densité de séismes autour du CSA. Bien que les épicentres des séismes ne soient pas localisés précisément au niveau des tracés des failles, plus nous nous rapprochons du cisaillement, plus l’activité sismique semble importante. De plus,

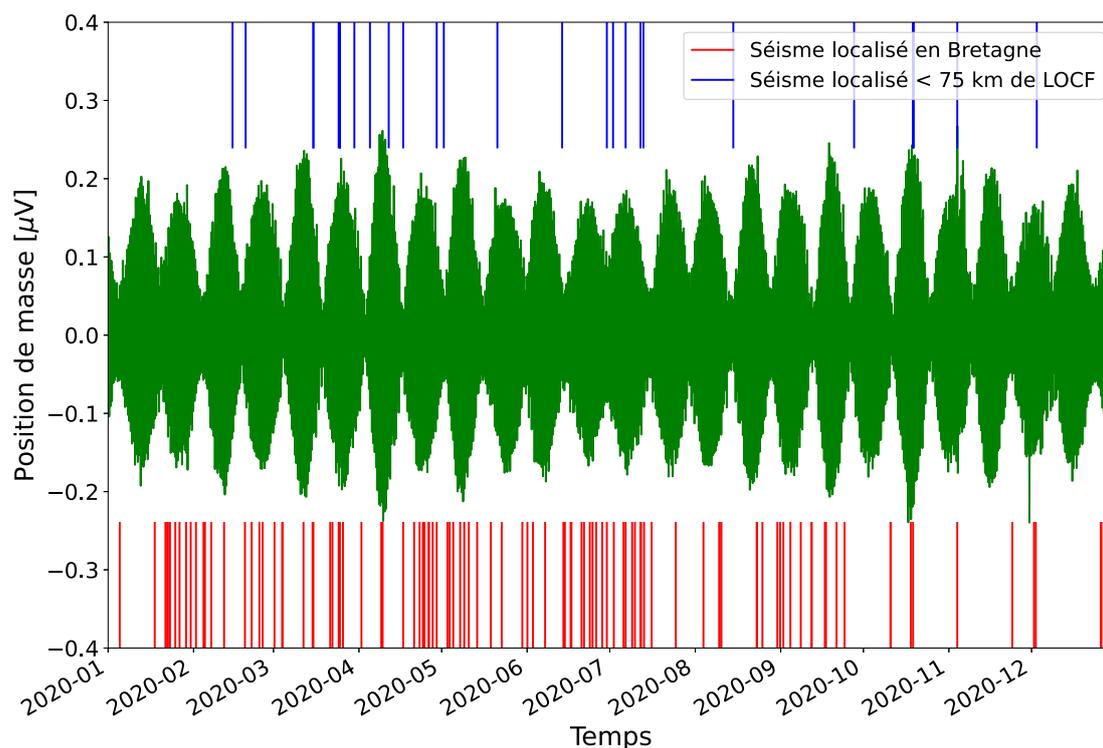


FIGURE 5.25 – Position de masse de la station LOCF du 1er janvier au 31 décembre 2020, en vert. En rouge, la position temporelle des événements sismiques détectés en Bretagne. En bleu, la position temporelle des événements sismiques localisés à moins de 75 km de LOCF.

dans certaines zones, nous observons également une activité sismique notable au niveau du CNA.

Plusieurs facteurs environnementaux peuvent induire une activité sismique. Nous avons tenté d'établir un lien entre le déclenchement de l'essaim sismique de Vannes et les précipitations enregistrés au cours des semaines précédentes. Cette étude ne montre pas de corrélation notable entre ces deux éléments. Il en est de même pour le forçage périodique. Ces différentes hypothèses ne doivent cependant pas être écartées. Ces études préliminaires sont cependant possibles grâce à un catalogue exhaustif de sismicité. Ainsi, en appliquant la méthode de détection et de discrimination sur une plus longue période de temps et en continue, de nouveaux événements sismiques pourront être détectés. Grâce à ces détections, des études plus approfondies sur les perturbations transitoires des contraintes locales ou de la résistance des failles pourront être réalisées.

5.4 Conséquences en termes d'observation sismologique

En marge des considérations géologiques, cette partie fait le point sur les performances du réseau et ses particularités.

5.4.1 Analyse d'un réseau *a posteriori*

L'ensemble des 391 événements ont été pointés manuellement, nous permettant ainsi d'évaluer la qualité des stations sismologiques de la région. En sélectionnant les événements situés à une distance suffisamment proche des stations (moins de 50 km), nous pouvons déterminer : (1) si chacun de ces événements a été détecté par *template matching* à chaque station, et (2) si les temps d'arrivée des ondes *P* et *S* ont pu être identifiés pour chaque station.

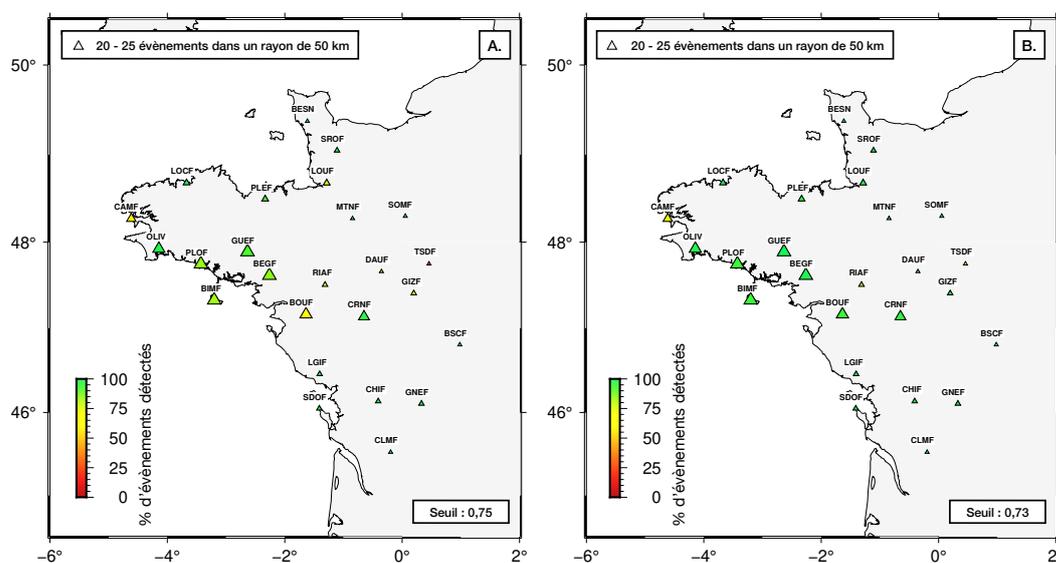


FIGURE 5.26 – Carte des stations utilisées pour la détection par *template matching*. La taille des triangles symbolise le nombre d'événements à moins de 50 km de la station. La couleur correspond au pourcentage d'événements, à moins de 50 km, qui ont été détectés par la station. (A) Résultats obtenus avec le seuil d'inter-corrélation fixé à 0,75. (B) Résultats obtenus avec le seuil d'inter-corrélation fixé à 0,73.

La Figure 5.26 montre les stations utilisées pour l'étape de détection. Ces stations sont symbolisées en fonction du nombre d'événements à 50 km autour d'elles. Une grande proportion d'événements est localisée autour des stations situées au sud de la Bretagne, telles que PLOF, GUEF, BEGF et BIMF. À noter que certaines stations

5.4. CONSÉQUENCES EN TERMES D'OBSERVATION SISMOLOGIQUE

ont été installées durant la période de cette étude, ce qui explique le faible nombre de détections à proximité de ces stations. En appliquant le seuil de détection à 0,75 (Fig 5.26.A), et grâce aux signaux de référence sélectionnés, la majorité des stations ont détecté plus de 75% des événements localisés à moins de 50 km. Nous constatons cependant que les stations BOUF et CAMF ont détecté moins de 75% des événements. Si nous abaissons le seuil à 0,73 (Fig 5.26.B), tous les événements autour de la station BOUF sont détectés. En ce qui concerne la station CAMF, environ 30% des événements restent non détectés. Cela peut être attribué à l'emplacement de la station, qui se trouve à proximité d'une falaise sur la pointe de Crozon (29), entraînant une forte présence de bruit océanique, haute fréquence, perturbant la détection.

Les résultats de détection semblent s'améliorer en abaissant le seuil à 0,73. Cependant, cette modification du seuil n'est pas applicable à l'ensemble des quatre années de données. La Figure 5.27 présente le nombre d'événements détectés en fonction du seuil d'inter-corrélation choisi à partir d'au moins quatre stations. Chaque courbe représente une journée d'analyse. On observe qu'en réduisant le seuil à 0,73, le nombre d'événements détectés par jour est multiplié par 3 par rapport aux résultats obtenus avec un seuil de 0,75. Il est important de noter que le traitement de l'ensemble de ces détections deviendrait excessivement long et nécessiterait un temps de calcul considérable. Par conséquent, les résultats indiquent que le choix du seuil de détection à 0,75 semble le plus approprié pour cette étude, car il permet aux stations de détecter la majorité des événements sismiques autour d'elles.

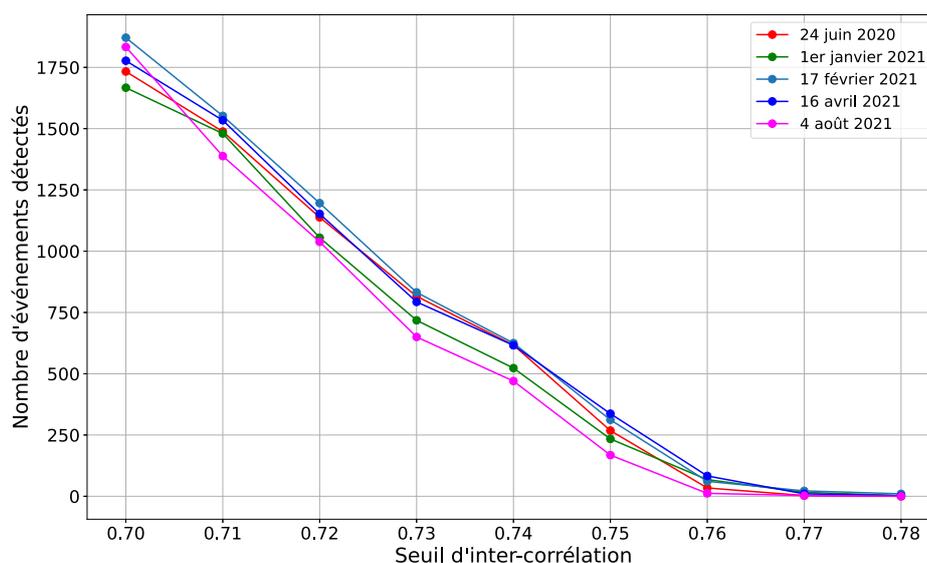


FIGURE 5.27 – Nombre d'événements détectés en fonction du seuil d'inter-corrélation choisi. Chaque courbe correspond à une journée d'analyse.

Un total de 25 stations a été utilisé pour la détection mais 29 stations ont été utilisées pour la localisation. Nous pouvons ainsi étudier le nombre d'événements situés à moins de 50 km des stations pour lequel le temps d'arrivée d'au moins une onde (P ou S) a été pointée manuellement. Dans la Figure 5.28, nous observons des caractéristiques similaires à celles observées pour la détection. Les temps d'arrivées d'une majorité d'événements ont pu être identifiés grâce aux enregistrements des stations les plus proches. Nous observons en revanche, qu'à la station CAMF, environ 30% des événements n'ont pas pu être pointés. Cela peut également provenir du niveau de bruit élevé ne permettant pas d'identifier efficacement l'arrivée des ondes. Ces résultats montrent la qualité des signaux et le niveau de bruit enregistrés aux stations du nord-ouest de la France.

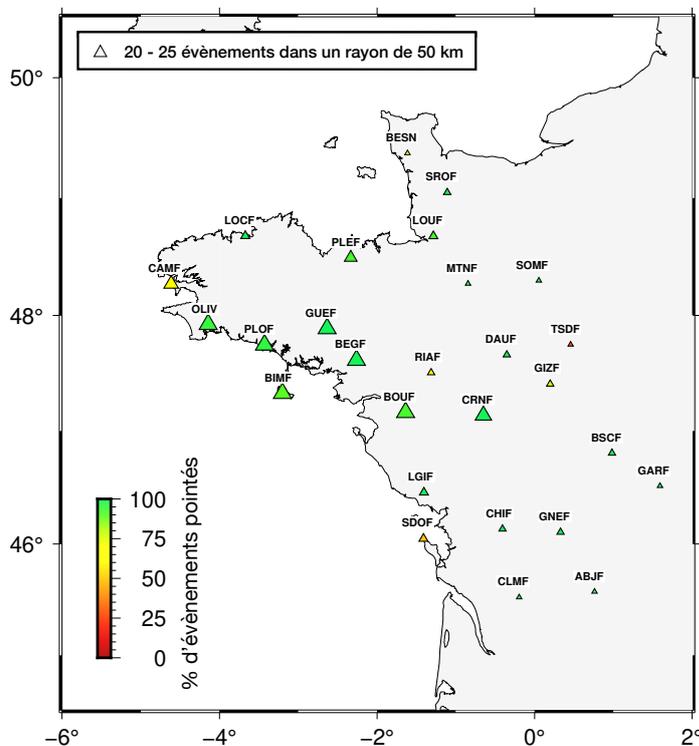


FIGURE 5.28 – Carte des stations utilisées pour la localisation. La taille des triangles symbolise le nombre d'événements à moins de 50 km de la station. La couleur correspond au pourcentage d'événements, à moins de 50 km, qui ont été pointés grâce aux enregistrements de ces stations.

Ces deux paramètres (nombre de détection et nombre de pointé) peuvent être mis en relation. Lorsque la méthode de détection identifie un signal sismique à une station, il est possible que le rapport signal sur bruit soit insuffisant pour permettre une identification précise des temps d'arrivée des ondes à cette station. Cela signifie que même si un signal sismique est détecté, il ne sera pas nécessairement pointé si la qualité

du signal n'est pas bonne. Ainsi, en examinant la relation entre le nombre de fenêtres détectées initialement et le nombre de pointés réalisés, pour chaque événement, nous pouvons mettre en évidence les cas où des signaux sismiques sont détectés mais n'ont pas pu être pointés en raison de la qualité du signal. La Figure 5.29 montre que pour cinq signaux détectés, ils sont associés en majorité à cinq signaux pointés. De plus, le nombre de fenêtres détectées semble être supérieur au nombre de fenêtres pointées. Cela suggère que nous atteignons potentiellement une limite de localisation, où toutes les fenêtres, détectées par la méthode, ne sont pas nécessairement utilisées pour la localisation des événements sismiques.

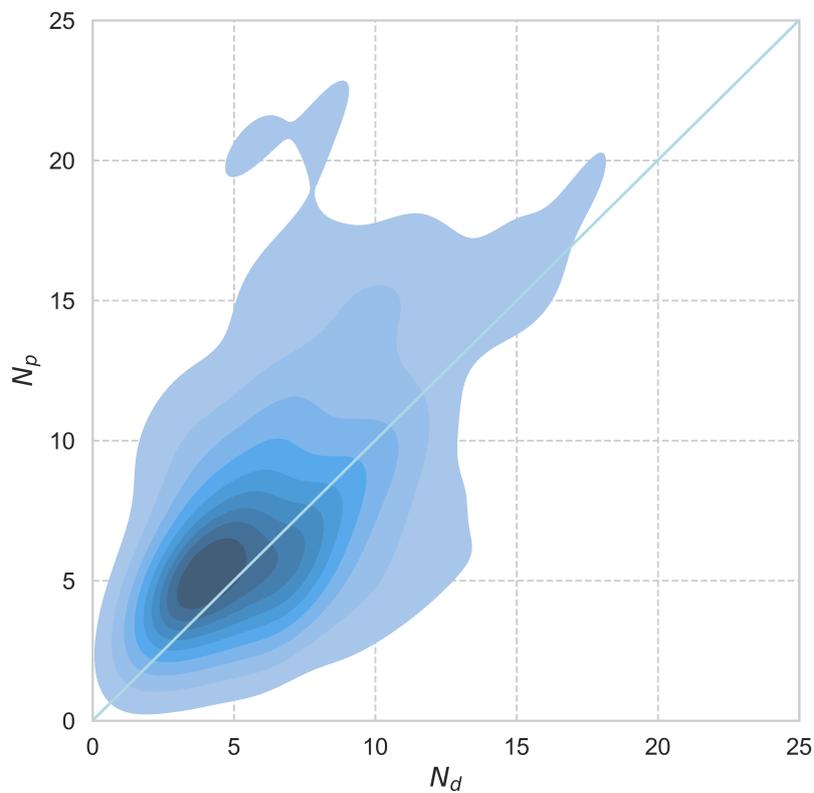


FIGURE 5.29 – Diagramme représentant le nombre de fenêtres détectées par *template matching* (N_d) en fonction du nombre de fenêtres où au moins une onde a été identifiée (N_p).

5.4.2 Comparaison avec d'autres méthodes de détection

Une fois le processus entièrement terminé (détection puis discrimination et enfin pointé), nous pouvons comparer les résultats avec d'autres méthodes qui proposent de réaliser l'ensemble de ces étapes en même temps comme PhaseNet et EQTransformer. Pour cela, nous utilisons les résultats obtenus sur la journée du 17 février 2021

(Fig 5.30). Pour rappel (section 4.1.1), ces deux approches ont détecté respectivement 14 et 12 événements sismiques naturels au cours de cette période. Notre méthode, quant à elle, a détecté 18 séismes.

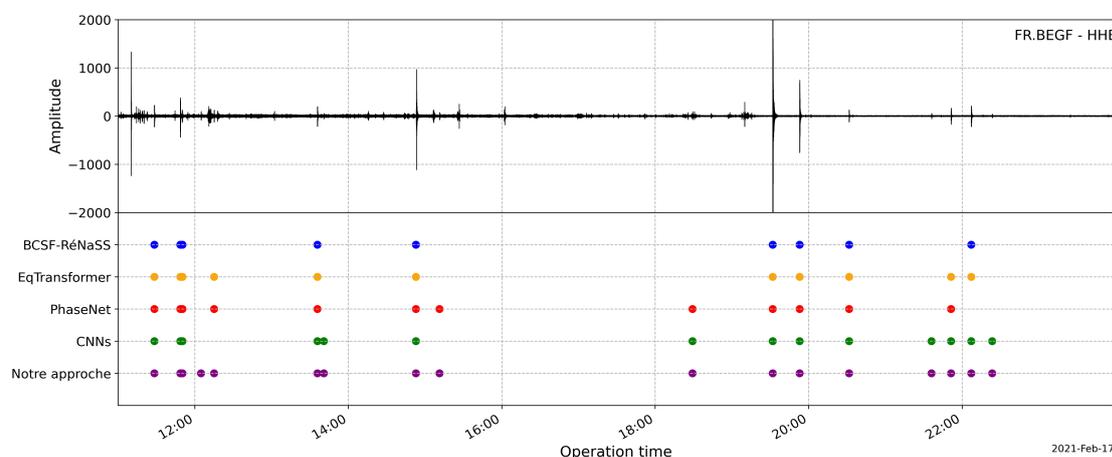


FIGURE 5.30 – Récapitulatif des détections de la journée du 17 Février 2021 réalisées par le BCSF-RéNaSS, EqTransformer, PhaseNet, l'utilisation des deux CNNs seulement et enfin notre approche. Le signal correspond à l'enregistrement de la journée par la station BEGF selon la composante Est-Ouest. Les cercles correspondent aux événements sismiques détectés.

Nous avons également essayé d'appliquer uniquement les étapes de discrimination aux signaux continus. Pour cela, l'ensemble du signal continu, pour chaque station, est découpé en segments de 60 s. Les spectrogrammes sont ensuite calculés pour chaque portion du signal et donnés en entrée au réseau de neurones. La première étape de discrimination des fausses détections permet ainsi de détecter les événements sismiques puis, la seconde étape permet de différencier les événements naturels des anthropiques. Pour un temps de calcul similaire à notre approche en trois étapes (170 minutes sur un total de 25 stations), les approches d'apprentissage profond ont permis de détecter 63 événements, dont 15 étaient des séismes (cercles verts sur la Fig 5.30). Cette proportion élevée de fausses détections est due au fait que l'algorithme de classification entre les séismes et le bruit a été entraîné sur des signaux impulsifs et non sur du bruit ambiant.

En conclusion, notre approche en trois étapes se positionne comme une solution fiable et efficace pour la détection des signaux sismiques de faible magnitude.

Chapitre 6

Application à la campagne post-sismique du séisme de La Laigne

Sommaire

6.1	Le séisme de La Laigne (16 juin 2023, $M_w=4,9$)	171
6.2	Application au signal des stations permanentes	175
6.3	Application au signal des stations temporaires	176
6.3.1	Réseau de stations temporaires post-sismique	177
6.3.2	Détection des répliques	177
6.4	Implications	183

6.1 Le séisme de La Laigne (16 juin 2023, $M_w=4,9$)

Le 16 juin 2023 à 18h38 heure locale (16h38 TU), un séisme d’une magnitude de $M_W = 4,9$ s’est produit au sud de Niort, en France, à une profondeur initialement annoncée de 2,5 km. La zone épacentrale du séisme est située approximativement à mi-chemin entre La Rochelle (Charente-Maritime) et Niort (Deux-Sèvres). Les évaluations de magnitude locale effectuées par le BCSF-RéNaSS indiquent une magnitude M_l de 5,3, tandis que le CEA-DAM estime la magnitude à 6,0. Ces valeurs de magnitude placent ce séisme parmi les plus importants enregistrés pendant la période instrumentale, non seulement dans le quart nord-ouest, mais également dans l’ensemble de la France métropolitaine. Ce séisme a suscité une attention particulière en raison de sa magnitude relativement élevée. Ce dernier a été ressenti dans une grande partie de l’ouest du pays

6.1. LE SÉISME DE LA LAIGNE (16 JUIN 2023, $M_W=4,9$)

(Fig 6.2). Les séismes de cette ampleur peuvent avoir des conséquences significatives sur les infrastructures, l'environnement et la population locale (Fig 6.1). Le mécanisme au foyer, représenté dans le Figure 6.3, semble témoigner d'un mouvement dextre correspondant à une faille verticale E-W ou à un mouvement senestre correspondant à une faille N-S.

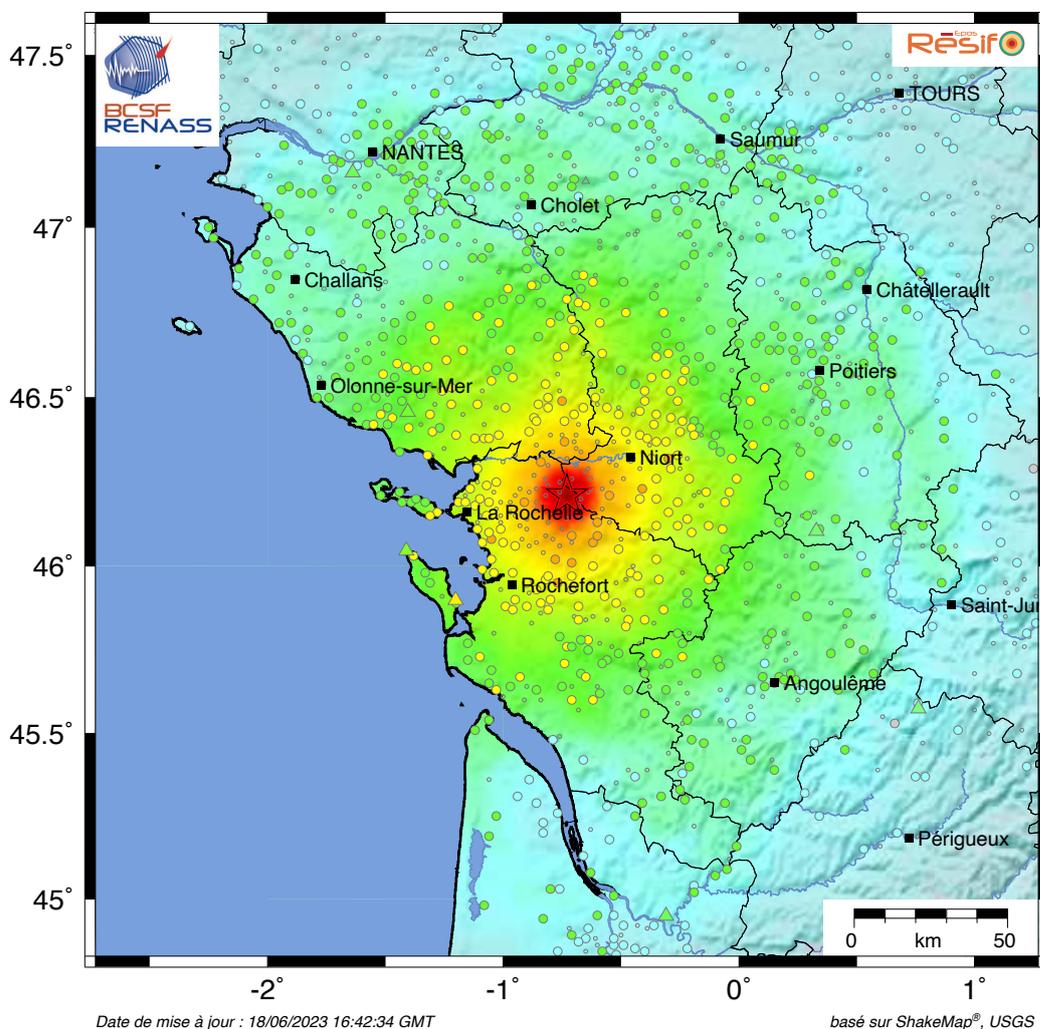


FIGURE 6.1 – *Dégâts causés par le séisme sur les villages aux alentours (communications personnelles : É. Beucler et C. Perrin).*

La région touchée par ce séisme se trouve sur la bordure nord du bassin d'Aquitaine, au sud des reliefs du Massif armoricain. Une couche sédimentaire d'environ 150 à 200 mètres d'épaisseur repose sur un socle composé de roches métamorphiques (micaschistes, gneiss) et volcano-sédimentaires datant du Cambrien (environ 500 Ma) (Fig 6.3). Autour de l'épicentre, on retrouve la faille de « Surgères-Blanzay-Aulnay-Échelle » qui a été cartographiée au sud-est de la zone. Cette faille, probablement décrochante, est représentée par des segments cartographiés qui recoupent les formations jurassiques depuis Mauzé-sur-le-Mignon jusqu'au socle potentiellement en Corrèze. Cependant, d'après le mécanisme au foyer déterminé, cette faille ne semble pas être à l'origine de ce séisme. La connaissance géologique du sous-sol dans la région des Charentes-Maritimes et des régions avoisinantes ne permet pas actuellement de lier ce séisme à une structure tectonique connue.

La nuit qui a suivi cet événement, une réplique d'une magnitude locale de 4,6 s'est produite à 04h27 heure locale (02h27 TU). Au cours des deux jours qui ont suivi le choc principal, une dizaine de répliques ont été détectées dans la région par le BCSF-RéNaSS.

CHAPITRE 6. APPLICATION À LA CAMPAGNE POST-SISMIQUE DU SÉISME DE LA LAIGNE

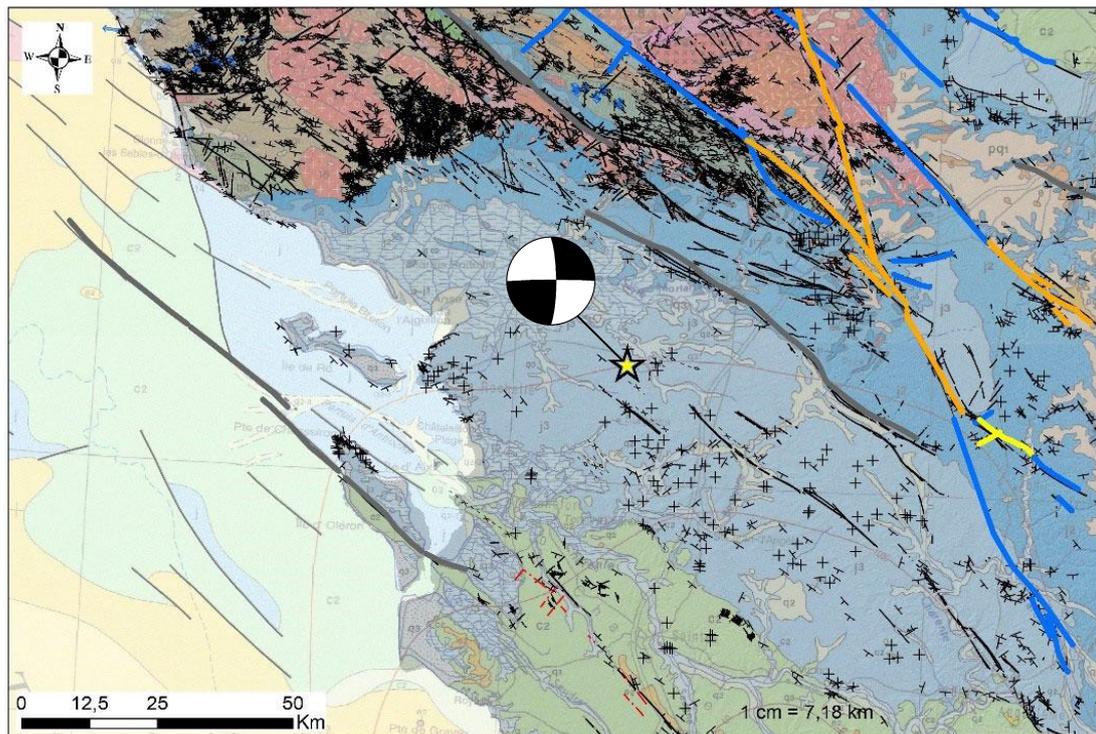


Intensités EMS98	I	II	III	IV	V	VI	VII	VIII	IX	X+
Dégâts potentiels bâtiments vulnérables	aucun	aucun	aucun	aucun	très légers	modérés	quelques effondrements partiels	nombreux effondrements partiels	nombreux effondrements	effondrements généralisés
Dégâts potentiels bâtiments peu vulnérables	aucun	aucun	aucun	aucun	aucun	aucun	très légers	modérés	effondrements partiels	nombreux effondrements
Perception humaine	aucune	très faible	faible	modérée	forte	brutale	très brutale	sévère	violente	extrême

Contribue au calcul	Ne contribue pas au calcul	Type d'observation
○	◦	Intensité d'après données macrosismiques (témoignages internet)
△	◡	Intensité issue d'une conversion à partir du PGA / PGV (donnée instrumentale) <small>Conversion PGA, PGV / intensité basée sur Caprio et al. (2015)</small>

FIGURE 6.2 – Estimation régionale de la secousse du 16/06/2023 18h38, à partir des données macrosismiques et instrumentales. Source : franceseisme.fr

6.1. LE SÉISME DE LA LAIGNE (16 JUIN 2023, $M_W=4,9$)



Légende

**Failles actives base BDFA IRSN
(Jomard et al, 2017)**

Age supposé

- Age syn à post Miocene inf.
- Age syn à post Pliocene
- Age indéterminé
- Faille de la carte neotectonique
(Grellet et al, 1993)

**Carte géologique 1M (BRGM, Chantraine et al. 2003)
Structures issues des cartes géologiques BRGM 1/50.000**

- + Stratification So avec pendage
- + Stratification So à pendage horizontal
- - - Axe synclinal ou synforme

FIGURE 6.3 – Contexte géologique du séisme La Laigne (étoile jaune). Localisation des failles potentiellement actives selon la BDFA IRSN (Jomard et coll., 2017) et des structures tectoniques répertoriées sur les cartes géologiques au 1/50000 (BRGM). Source : BRGM

Afin de mieux comprendre le contexte de ce séisme, des investigations plus ciblées, nécessitant l'installation de stations temporaires autour de la zone épiscopale, ont eu lieu. L'analyse des répliques enregistrées par ces stations temporaires permet de contraindre la localisation et la géométrie du choc principal. En étudiant ces répliques, nous pouvons obtenir des informations sur les caractéristiques de la faille sismique, la propagation des ondes sismiques et les mécanismes physiques à l'origine du déclenchement. Ces données sont essentielles pour améliorer notre connaissance de la sismotectonique de la région afin d'affiner les modèles d'aléas sismiques. Pour se faire, un réseau de treize stations temporaires a été déployé, dans la région épiscopale, dès le 17 juin.

6.2 Application au signal des stations permanentes

Nous avons d'abord appliqué notre approche de détection et de discrimination aux stations permanentes situées autour de l'épicentre du séisme, pendant la période où peu de stations temporaires étaient installées. L'objectif était d'identifier les potentielles répliques qui auraient pu être manquées. Je me suis concentrée sur la période allant du 16 juin 2023 à 15h00 (TU) au 18 juin 2023 à 12h00 (TU). Pour la première étape de détection, nous avons utilisé le même catalogue d'événements de référence que celui présenté dans la Section 5.2 de notre étude. Pour qu'un événement soit considéré comme détecté, il faut à minimum de trois stations ayant dépassé le seuil d'intercorrélation normalisé de 0,75. Au total, 686 événements potentiellement sismiques ont été détectés sur cette période. Pour la deuxième étape, qui vise à différencier les fausses détections des véritables séismes en utilisant notre réseau de neurones entraîné, 94 événements ont été classés comme « *quake* ». Enfin, pour la dernière étape, nous avons utilisé notre réseau de neurones pour distinguer les séismes d'origine anthropique des événements naturels, et 32 événements ont été identifiés comme « naturels ». Après une inspection visuelle, nous avons constaté que 18 événements étaient déjà présents dans le catalogue du RéNaSS, tandis que 14 autres événements correspondaient à des répliques qui n'étaient répertoriées dans aucun catalogue de séismes existant. Parmi ces 14 événements, 11 ont pu être pointés à suffisamment de stations pour être localisés (Fig 6.4).

Les résultats obtenus mettent en évidence l'efficacité de notre approche de détection et de discrimination en temps réel pour l'identification des répliques potentielles lors de cette crise sismique. Une caractéristique importante de notre méthode est qu'aucun

6.3. APPLICATION AU SIGNAL DES STATIONS TEMPORAIRES

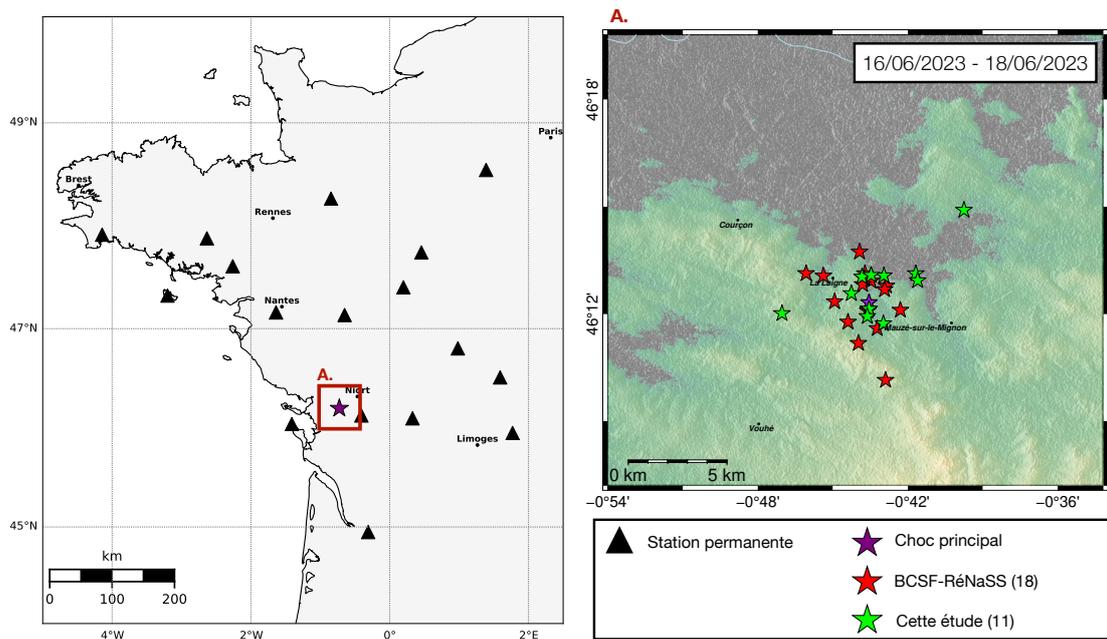


FIGURE 6.4 – Carte de localisation du séisme du 16 Juin 2023 (étoile violette) et des répliques associées (étoiles rouges et vertes) détectés à partir des stations permanentes. Les triangles noirs correspondent aux stations permanentes utilisées pour la détection des répliques.

séisme déjà enregistré pendant cette période n'a été utilisé comme signal de référence pour la détection. Cela souligne la capacité à identifier de manière autonome de nouveaux événements sismiques, dont des répliques de séismes, sans s'appuyer sur des informations préexistantes sur le choc principal. L'ensemble de ces nouvelles répliques détectées ont par la suite pu être ajoutées au catalogue du BCSF-RéNaSS.

6.3 Application au signal des stations temporaires

Après un événement sismique de cette ampleur, des informations précises sur les caractéristiques du séisme et ses répliques éventuelles doivent être récupérées. Pour ce faire, l'installation de stations sismiques temporaires autour de l'épicentre est une pratique courante. Ces stations sont déployées quelques heures après le choc principal, afin de surveiller attentivement la séquence sismique qui suit ce séisme. Ces stations permettent donc de recueillir des données précieuses sur les répliques pour :

- **Étudier la séquence sismique.** L'installation de stations temporaires permet de surveiller et d'étudier cette séquence sismique dans les jours, les semaines, voire les mois qui suivent le choc principal. Cela permet de mieux comprendre le

comportement des failles et leurs géométries et ainsi mieux décrire le mécanisme au foyer.

- **Surveiller l'évolution de la situation.** Les répliques peuvent causer des dommages supplémentaires aux structures déjà affaiblies par le séisme principal. La détection et le suivi des répliques permettent d'estimer le risque résiduel et d'adapter les mesures en conséquence.

6.3.1 Réseau de stations temporaires post-sismique

Grâce à la réactivité d'Éric Beucler, Mickaël Bonnin, Clément Perrin, Damien Frigiel ainsi que Marion Alloncle, un réseau de 13 stations sismologiques temporaires a pu être installé à partir du 17 juin 2023 au matin.

Le réseau est constitué de 12 accéléromètres numériques de la société Sercel (DSU3) et d'un géophone vertical (LC13). Ces capteurs ont été utilisés pour enregistrer le mouvement du sol du 17 au 21 juin 2023, à une fréquence d'échantillonnage de 500 points par seconde. Pour les accéléromètres, le mouvement du sol est enregistré dans les trois directions de l'espace et pour le géophone seulement selon la composante verticale. Lors de l'installation de ces stations, il est crucial de réduire au maximum les sources de bruit ambiant. Cela implique de s'éloigner des sources anthropiques telles que la circulation routière ou les habitations, ainsi que des animaux et d'autres facteurs environnementaux susceptibles de générer des vibrations, comme les variations importantes de température, les racines d'arbres ou les murs. Une attention particulière est portée à l'horizontalité du capteur, vérifiée à l'aide d'une bulle de nivellement, ainsi qu'à son orientation vers le nord géographique (Fig 6.6).

6.3.2 Détection des répliques

Au cours de la période de déploiement du réseau temporaire de La Laigne, nous avons appliqué notre approche de détection et de discrimination. Étant donné que les stations enregistrent le mouvement du sol à une fréquence de 500 points par seconde, le catalogue de signaux de référence que nous avons établi pour l'ensemble du Massif armoricain ne peut pas être utilisé dans ce contexte spécifique. Nous devons donc constituer un nouveau catalogue de signaux en utilisant les enregistrements des stations temporaires. Comme première application, nous avons sélectionné les événements de plus grande magnitude détectés par le BCSF-RéNaSS pendant la période

6.3. APPLICATION AU SIGNAL DES STATIONS TEMPORAIRES



FIGURE 6.5 – Installation de capteurs, exemple de (A) la station LC02 et (B) la station LC03. Les capteurs sont enterrés, horizontalisés et orientés vers le nord grâce à une boussole.

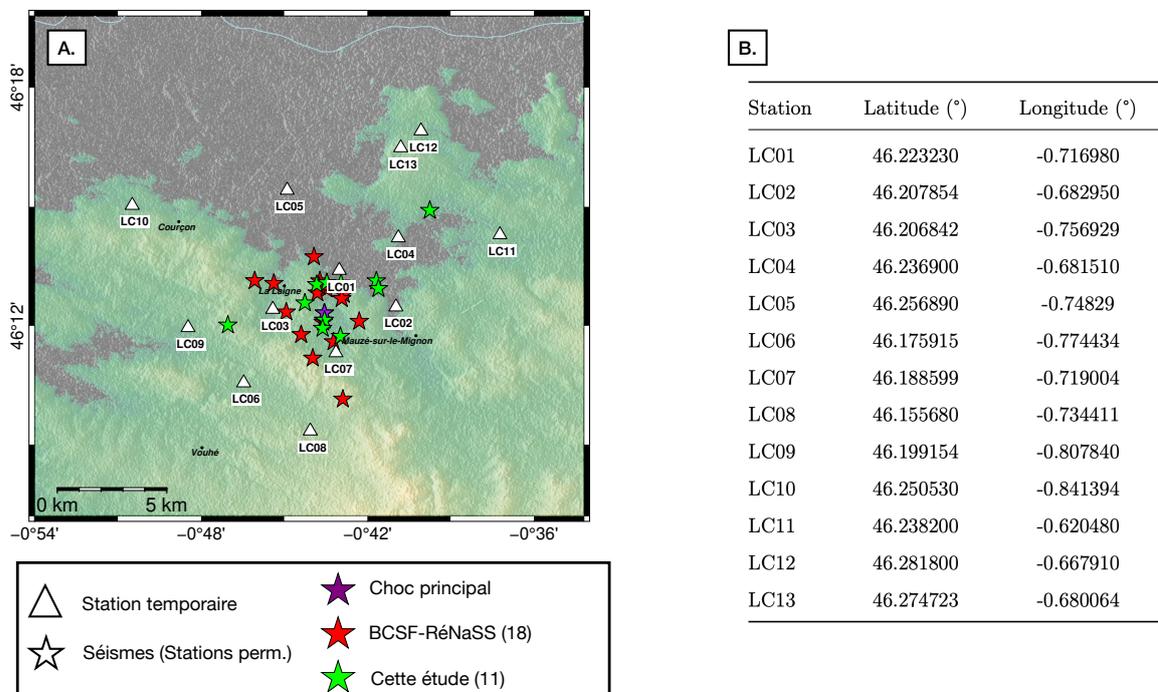


FIGURE 6.6 – Localisation des stations de la campagne post-sismique de La Laigne.

d'enregistrement du réseau temporaire. Ceux-ci correspondent aux séismes survenu le 17 juin à 13h13 (TU), d'une magnitude M_l de 1,87 et à 18h44 (TU), d'une magnitude M_l de 1,66. Ces événements ont été enregistrés par 11 des stations temporaires opérationnelles. Nous avons extrait les formes d'ondes correspondantes, selon la composante Nord, pour constituer notre catalogue de référence. Nous avons appliqué le même traitement que dans la Section 4.1.2, en utilisant des signaux de référence de 10 secondes en raison de la proximité des stations à la source du séisme. Notre catalogue de signaux de référence est donc composé de 22 enveloppes de signaux (11 stations \times 2 événements \times 1 composante).

Nous avons calculé une série d'inter-corrélations normalisées entre chaque signal de référence et le signal continu de la composante Nord des 12 stations. La station LC13, n'enregistrant que selon la composante verticale, n'est pas utilisée pour la détection. Pour qu'un événement soit considéré comme une détection, le maximum d'inter-corrélation doit dépasser 0,75 à au moins 3 stations.

Au cours des 5 jours d'étude, nous avons détecté 773 événements potentiellement sismiques. La deuxième étape consiste à discriminer les signaux impulsifs bruités des événements sismiques. Pour calculer les spectrogrammes, les formes d'ondes ont été décimées à 100 points par seconde afin de correspondre à la taille des données d'entrée du réseau de neurones. Cette étape nous permet également d'évaluer si notre approche de discrimination peut être appliquée à des signaux de fréquence supérieure à 100 Hz et à des stations temporaires potentiellement plus bruitées. Grâce à notre réseau de neurones, nous avons pu classer 173 événements comme « *quake* ». Pendant la période d'application de notre méthode (du samedi au mercredi), aucune activité anthropique n'était présente dans la zone d'implantation des stations. Ainsi, nous n'avons pas appliqué la dernière étape de discrimination. Après inspection visuelle, 105 événements détectés et classés comme sismiques sont bien des séismes naturels.

Les résultats préliminaires obtenus démontrent l'efficacité de notre approche dans le cadre de cette étude locale basée sur seulement deux événements de références. Un nouveau catalogue d'événements de référence a été constitué grâce aux répliques détectées lors de cette première application. Pour cela, nous avons sélectionné les formes d'ondes de quatre répliques détectées. Grâce à notre nouveau catalogue, composé de 43 signaux de référence, nous avons réalisé une seconde série d'inter-corrélations normalisées entre chaque signal de référence et le signal continu de la composante Nord des 12 stations.

Pour cette seconde application, nous avons choisi de regarder les résultats de discrimination au niveau du *station-level*. Chaque signal classé comme « *quake* » par le réseau de neurones sera visualisé. J'ai fait ce choix car la discrimination est réalisée à partir de fenêtres de 60 s, or le signal sismique enregistré par les stations temporaires dure en moyenne 5 s. Nous avons donc de fortes chances qu'un événement sismique soit clairement visible sur une seule station mais que la majorité des fenêtres d'analyses soient bruitées. Ainsi, en visualisant chaque signal classé en « *quake* », nous pouvons récupérer un maximum d'événements sismiques.

Grâce à cette seconde application, 1 192 événements potentiellement sismiques ont été détectés, soit 5 740 fenêtres temporelles (signaux). À partir de notre réseau de neurones entraîné, 2 036 de ces signaux ont été classés comme « *quake* », correspondant à 532 événements. Après inspection visuelle de chaque signal classé comme « *quake* », 198 de ces événements correspondent bien à des tremblements de terre. Ainsi, 93 nouveaux séismes ont pu être détectés en complément des 105 déjà repérés par la première application.

Ces 198 événements sismiques sont donc les répliques de l'événement majeur survenu le 16 juin 2023. L'apparition de ces répliques suit des lois empiriques bien établies. La décroissance du nombre de répliques en fonction du temps peut être décrite par une loi d'Omori (1894),

$$n(t) = \frac{K}{(t + c)^p} \quad (6.1)$$

où K et c sont des constantes, $n(t)$ correspond à la fréquence des répliques par unité de temps t . L'exposant p varie entre 0,9 et 1,5 (Utsu *et coll.*, 1995). Le temps t a pour origine l'instant où s'est produit le choc principal. La Figure 6.7 montre le nombre de séismes en fonction des heures et des jours suivant l'événement du 16 juin de 16h38 (TU). On observe ici une décroissance nette du nombre d'événements avec le temps, coïncidant avec la loi d'Omori présentée.

Sur les 198 événements détectés, seuls 177 ont finalement pu être localisés car les pointés n'étaient pas possibles pour toute les stations.

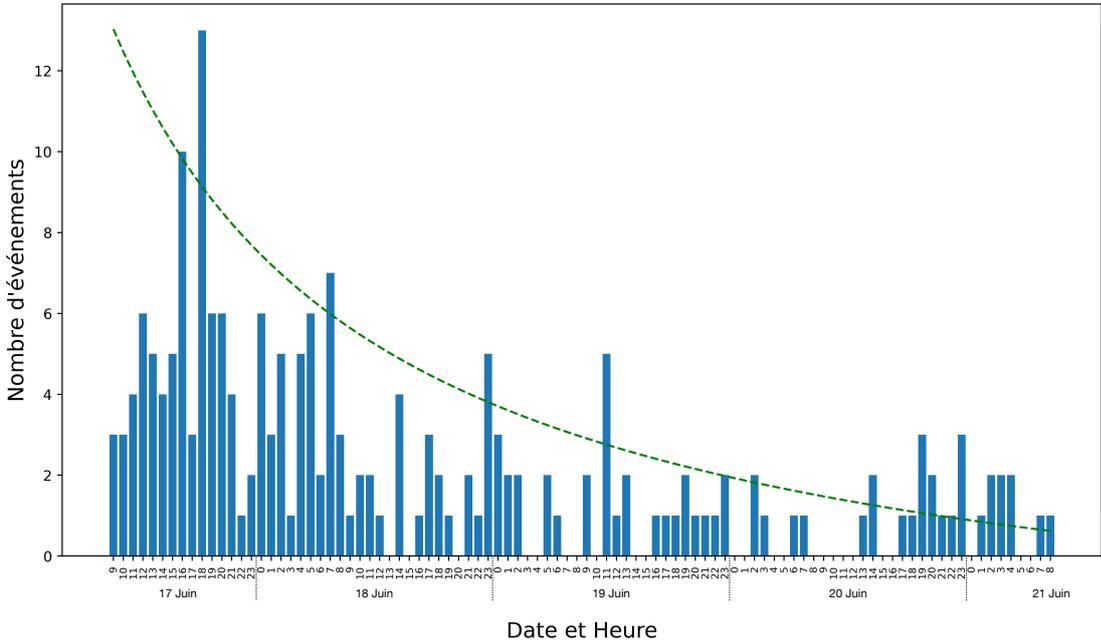


FIGURE 6.7 – Histogramme représentant le nombre de séismes en fonction du temps. La courbe en pointillé représente un exemple de loi d'Omori pouvant être appliquée avec $K = 1,4$, $c = 12$ et $p = 0,95$, d'après (Žalohar, 2018)

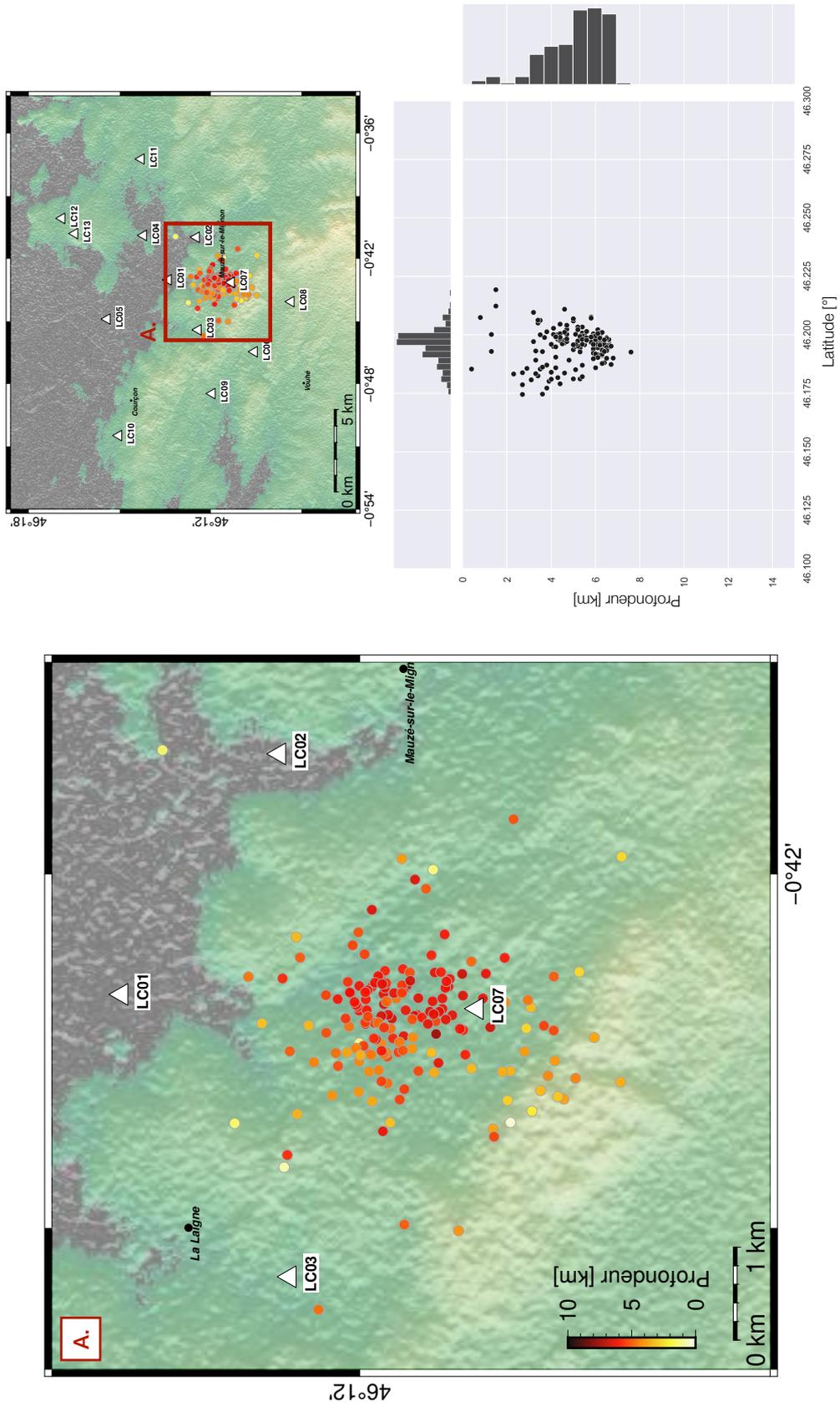


FIGURE 6.8 – Localisation préliminaire des répliques détectées du 17 au 21 juin 2023 et distribution des événements en fonction de la profondeur.

6.4 Implications

Ces répliques sont localisées sur une zone géographique restreinte, comparées aux localisations du RéNaSS (Fig 6.4), sur un rayon de 2 km (Fig 6.8.A). La répartition de ces séismes ne semble cependant pas montrer d'orientation préférentielle.

La localisation de ces répliques ne correspond à aucune faille cartographiée dans la zone. Par conséquent, nous nous interrogeons sur leur origine. La faille « Surgères-Blanzay-Aulnay-Échelle », qui a été cartographiée au sud-est de la zone, est trop éloignée pour être responsable de ces événements. De plus, son orientation ne correspond pas au mécanisme au foyer décrit. Le mécanisme au foyer suggère, le plus vraisemblablement, une faille décrochante avec un plan Est-Ouest et un pendage vertical.

Les répliques sont localisées à des profondeurs allant de 2 à 7 km, la plupart se situant à environ 6 km de profondeur. L'orientation du plan de faille en profondeur semble correspondre à un pendage vertical (Fig 6.8). La profondeur des séismes, dans la région du Massif armoricain, est généralement très superficielle, de l'ordre de 5 à 10 km (Fig 5.15). Le comportement mécanique du sous-sol peut être caractérisé grâce à la distribution des profondeurs des séismes (par exemple [Chen et Molnar, 1983](#); [Bourgeois et coll., 2005](#)). Ce comportement peut être de deux types : fragile et ductile. Un profil rhéologique peut être déterminé, correspondant à la distribution de la résistance du sol en fonction de la profondeur. Ainsi, en détectant un maximum de répliques, la structure crustale pourra être mieux caractérisée.

Notre approche nous a donc permis de détecter 198 répliques du séisme de La Laigne et d'en localiser 177 grâce aux 13 stations temporaires installées autour de l'épicentre. Toutefois, il est encore possible d'améliorer le nombre de détections en adaptant certains paramètres. Une augmentation du nombre de signaux de référence pourrait être envisagée, par exemple en utilisant plus de formes d'ondes de répliques détectées au cours de cette étude. La durée des formes d'ondes et les seuils d'inter-corrélations pourraient également être optimisés. Pour une meilleure application aux données de campagne post-sismique, il est possible d'améliorer la méthode de discrimination. Actuellement, les données d'entrée du réseau de neurones sont basées sur des signaux d'une durée de 60 s. Cependant, lors de l'installation de stations temporaires à quelques kilomètres de l'épicentre du séisme, la durée des signaux sismiques est de l'ordre de quelques secondes. Afin d'optimiser les résultats de la discrimination, il peut être nécessaire de ré-entraîner le réseau de neurones en utilisant une base de données de signaux de plus

6.4. IMPLICATIONS

courte durée. Il est également possible d'améliorer la localisation de ces répliques en utilisant des méthodes telles que l'approche HypoDD (Waldhauser et Ellsworth, 2000), qui permet une localisation plus précise des séismes regroupés, comme les répliques de ce séisme.

Malgré ces possibilités d'amélioration mentionnées, les paramètres que nous avons sélectionnés ont permis de détecter et de discriminer automatiquement 198 répliques. Ceci démontre la capacité de notre approche à identifier les séismes et à les différencier à partir des enregistrements des stations temporaires. Ces résultats prometteurs ouvrent la voie à des études plus approfondies visant à affiner notre méthode et à améliorer encore davantage la détection des événements sismiques.

Conclusion et perspectives

Dans le cadre de cette étude, nous nous sommes intéressés plus particulièrement à l'étude de la sismicité naturelle du Massif armoricain, une région continentale stable. La région nord-ouest de la France se distingue par une sismicité fréquente et de magnitude faible à modérée. Cependant, les origines de cette sismicité sont à l'heure actuelle mal comprises et donc sujettes à débats. Cette région présente des défis uniques pour la caractérisation des événements sismiques naturels dans un tel contexte géologique.

Durant ces dernières années, plusieurs études de cette région, notamment celles menées par [Arroucau \(2006\)](#) et [Haugmard \(2016\)](#), ont permis de répondre à certaines interrogations. Ces études ont introduit une méthode de localisation précise des séismes locaux et régionaux, en s'affranchissant de connaissances *a priori* sur la structure interne de la région. Dans des domaines intracontinentaux où les structures crustales sont peu connues, comme le Massif armoricain, cette méthode de localisation s'avère idéale, particulièrement pour les petits séismes. L'application de cette approche a permis de mettre en évidence le caractère diffus de la sismicité de la région.

Cependant, une question subsistait : est-ce que cette sismicité diffuse n'était pas liée à la faible couverture de stations de l'époque ? Grâce au projet RLBP, le réseau de stations sur l'ensemble du territoire métropolitain a été densifié. Nous avons mis en avant la nécessité de détecter les séismes de faible magnitude afin d'obtenir une cartographie plus fine de la sismicité de la région. Dans cette optique, une nouvelle approche de détection et de discrimination des événements sismiques naturels a été développée.

L'objectif principal était de fournir un catalogue d'événements sismiques naturels le plus complet possible, ce qui contribuera à une meilleure compréhension des phénomènes sismiques spécifiques à cette région. L'application de notre méthode de détection par *template matching* suivie de deux étapes de discriminations par CNN, sur quatre

années de données sismiques, a permis de mettre en évidence plus de 390 nouveaux événements sismiques, doublant ainsi le nombre de détections dans la région par rapport aux bulletins nationaux. La localisation de ces événements montre que les épicentres de ces séismes ne semblent pas être cantonnés aux structures tectoniques majeures du massif et que l'aspect diffus de la sismicité du quart nord-ouest semble se maintenir. Nous avons confirmé l'existence d'une ceinture à forte densité de séismes, large d'environ 100 km et orientée NW-SE sur une distance de 600 km. Cette ceinture de forte densité d'événements semble être corrélée avec la présence des cisaillements Sud et Nord Armoricaïn, comme le montre la distribution de la sismicité autour de ces structures.

Les résultats obtenus lors de l'application sur les données de la campagne post-sismique du séisme de La Laigne, ont démontré l'efficacité et la robustesse de la méthode développée. Durant cette étude, 198 répliques ont pu être détectées autour de l'épicentre du tremblement de terre. Ainsi, la profondeur de ces événements et également la structure du plan de faille sont mieux contraintes.

Ce travail ouvre plusieurs perspectives concernant le développement de la méthode et ces futures applications. Nous avons fait le choix de différents paramètres pour l'élaboration de notre approche. Afin d'affiner les résultats, plusieurs pistes peuvent être envisagées :

- **Application sur l'ensemble des composantes du sismogramme.** Par soucis de temps de calcul, notre approche a seulement été appliquée sur la composante Nord de chacune des stations du Massif armoricaïn. L'application sur l'ensemble des composantes du sismogramme pourra permettre d'affiner les résultats et probablement de détecter plus d'événements sismiques.
- **Adapter le seuil de détection.** Nous avons appliqué un seuil unique à l'ensemble des stations sismologiques de la région. Dans une optique d'application en temps réel, particulièrement en milieu de journée, nous pouvons imaginer utiliser un seuil optimisé en fonction du rapport signal sur bruit de chaque station.
- **Itération des signaux de référence.** Afin d'affiner la détection des événements sismiques dans la région, nous pouvons utiliser les signaux détectés, par exemple, lors d'une première application, comme nouveaux signaux de référence. La variété de signaux de référence serait donc augmentée au fur et à mesure des itérations.
- **Logique spatio-temporelle pour la discrimination.** Afin de conforter les résultats de la discrimination entre les événements sismiques et les fausses détections, nous pouvons suivre la logique spatio-temporelle d'un séisme. À partir

du principe de LOA pour « Localisation par Ordre d'Arrivée », ou AOL pour « *Arrival Order Location* », nous pouvons vérifier la cohérence des données grâce aux temps de détection et à la position des stations et identifier des localisations aberrantes ([Anderson, 1981](#); [Sambridge et Kennett, 1986](#)).

- **Re-entraînement des réseaux de neurones.** Le ré-entraînement des algorithmes d'apprentissage automatique peut être nécessaire dans certains cas. En effet, à l'heure actuelle, nous utilisons des fenêtres d'analyse de 60 s, particulièrement adaptées pour une étude régionale. Pour une étude plus locale, avec des stations installées à quelques kilomètres de l'épicentre d'un séisme, le signal sismique est significativement plus court. Les fenêtres utilisées pour la discrimination seront donc en grande partie composées de bruit sismique pouvant nuire à la classification des signaux. Nous pouvons imaginer une première application de l'approche pour détecter les signaux de plus grande amplitude puis constituer une nouvelle base de données d'entraînement à partir de ces résultats.

En ce qui concerne l'implémentation pratique du code, il serait intéressant de créer des routines permettant de détecter, discriminer et écrire à la fois les données et les résultats dans le format QuakeML ([Schorlemmer *et coll.*, 2011](#)), qui est un standard pour l'échange et la diffusion des données sismologiques. La méthode de discrimination des événements sismiques est, quant à elle, disponible sur GitLab.

En complément à cette méthode de détection, des études approfondies sur les différents facteurs externes et internes pourront ainsi être développées. Dans cet objectif, l'installation de stations multi-instrumentales (gravimètre, piézomètre, pluviomètre, etc.) pourrait permettre une observation multi-échelle et potentiellement une corrélation entre la sismicité et certains facteurs externes. L'utilisation de la méthode HypoDD (l'algorithme de localisation des séismes par double différence, [Waldhauser et Ellsworth, 2000](#)) permettrait une détermination plus fine de la localisation d'événements regroupés, comme pour les répliques du séisme de La Laigne. De plus, l'utilisation combinée de la méthode de détection développée dans le cadre de cette étude et la localisation des séismes permettra d'améliorer la connaissance sur le modèle de vitesse des ondes dans le quart nord-ouest de la France.

L'utilisation future de cette méthode de détection et de discrimination dans des domaines intracontinentaux devrait fournir une vision plus précise de la sismicité. Cette méthode peut s'adapter à de nombreux contextes d'études, de l'échelle locale, grâce à une campagne post-sismique, à l'échelle régionale. Enfin, notre approche a montré

son efficacité et a contribué significativement à améliorer la capacité de détecter et de discriminer les événements sismiques naturels, permettant ainsi une meilleure compréhension des phénomènes sismiques dans notre région.

ANNEXES

Annexe A

Bases de sismologie

A.1 Intensité d'un séisme

L'intensité sismique mesure les effets relatifs observés lors d'un séisme à un endroit donné, contrairement à la magnitude qui quantifie de manière absolue l'énergie libérée par le séisme. Différentes échelles d'intensité ont été développées pour évaluer ces effets. L'ensemble de ces échelles utilise une classification allant de I (non ressenti) à XII (destructions totales).

- Échelle MSK-64 (Medvedev-Sponheuer-Karnik, [Medvedev et coll., 1964](#)) : L'échelle MSK-64 est une échelle macro-sismique développée en Europe de l'Est. Elle évalue les effets sismiques en se basant sur des critères tels que les dommages aux bâtiments, les mouvements des objets, les réactions des personnes, etc.
- Échelle EMS-98 (European Macroseismic Scale, [Grünthal, 1998](#)) : L'échelle EMS-98 est une échelle macro-sismique développée par le Réseau sismologique européen. Elle évalue les effets sismiques en prenant en compte des critères similaires à ceux de l'échelle MSK-64, ainsi que des informations sur les effets environnementaux et les réponses des animaux.
- Échelle de Mercalli : L'échelle de Mercalli est une échelle macro-sismique développée au début du XXe siècle. Cette échelle évalue les effets sismiques en se basant sur les observations des témoins oculaires, les dommages aux structures, les mouvements des objets, etc. Elle peut être utilisée rétrospectivement pour déterminer l'intensité des séismes historiques.

A.2 Magnitude d'un séisme

La magnitude d'un séisme évalue l'énergie libérée par un tremblement de terre à partir de différentes caractéristiques du signal sismique. Il existe plusieurs magnitudes permettant de faire cette mesure :

- Magnitude locale (M_l) : La magnitude locale, également connue sous le nom de magnitude de Richter locale, est une mesure de l'ampleur d'un séisme basée sur les amplitudes des ondes sismiques enregistrées par les stations sismiques proches de l'épicentre. Elle est calculée à partir des enregistrements des ondes de volume.
- Magnitude de durée (M_d) : La magnitude de durée est une mesure de l'énergie libérée par un séisme basée sur la durée du signal sismique. Elle est déterminée en mesurant la durée du signal sur l'enregistrement pour chaque station. Cette mesure de magnitude est couramment utilisée pour les séismes locaux de magnitude comprise entre 0 et 5 (*e.g.* Perrot *et coll.*, 2005; Mandal *et coll.*, 2004).
- Magnitude de moment (M_W , Hanks et Kanamori, 1979) : La magnitude de moment est une mesure plus complète de l'énergie libérée par un séisme. Elle est basée sur la quantité totale d'énergie sismique produite par la rupture de la faille, et reliée au moment sismique. Cette magnitude est calculée à partir de la surface de rupture, de la résistance des roches et du déplacement le long de la faille.

Il existe d'autres types de magnitudes sismiques, tels que la magnitude d'ondes de surface (M_s) qui mesure l'amplitude des ondes de surface enregistrées par les sismographes, et la magnitude d'énergie (M_e) qui évalue la quantité d'énergie totale libérée par un séisme. Cependant, la magnitude de moment (M_W) est généralement considérée comme la mesure la plus fiable et la plus utilisée pour évaluer l'ampleur d'un séisme.

A.3 Mécanisme au foyer

Le mécanisme au foyer d'un séisme décrit la configuration géométrique et les mouvements relatifs des failles qui sont responsables de la libération d'énergie sismique lors d'un séisme. Il modélise la manière dont les roches se déforment et se brisent sous l'effet des contraintes tectoniques accumulées. Le mécanisme au foyer d'un séisme est déterminé à partir de l'analyse des ondes sismiques enregistrées par les stations sismologiques. Les ondes sismiques se propagent dans toutes les directions depuis le point de

rupture où le séisme s'est produit, connu sous le nom d'hypocentre ou foyer sismique. En mesurant les caractéristiques des ondes sismiques, telles que leur vitesse, leur amplitude et leur polarisation, il est possible de déterminer les trois éléments principaux du mécanisme : la direction de la rupture, le sens du déplacement relatif des blocs rocheux et l'angle d'inclinaison de la faille.

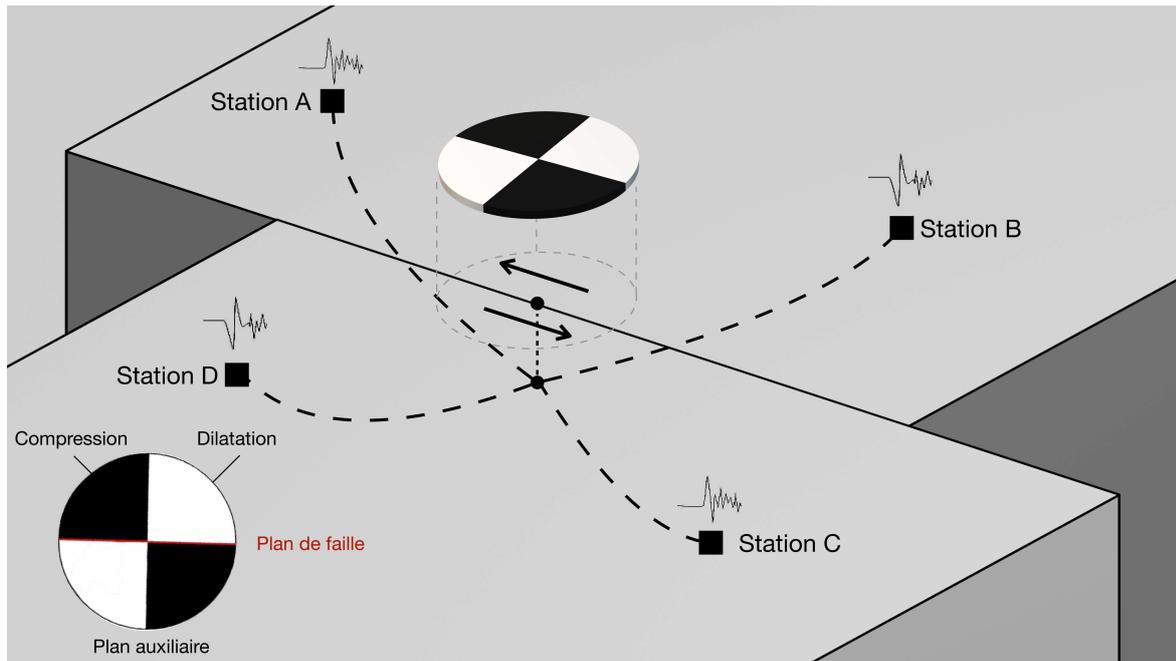


FIGURE A.1 – Bloc-diagramme montrant un plan de faille vertical. A, B, C et D sont les quatre stations sismiques entourant l'épicentre. La polarité des ondes P est représentée pour chaque station.

Les polarités de l'onde P enregistrées peuvent être comparées sur chaque enregistrement des différentes stations. En règle générale, si l'onde P est positive à une station donnée, cela signifie que le mouvement du sol est dans la direction de la station. En revanche, si l'onde P est négative, cela indique que le mouvement est opposé à la direction de la station (Fig A.1).

Le mécanisme au foyer est généralement représenté par un diagramme appelé diagramme de *beachball*, qui présente une projection de la sphère focale montrant les différentes composantes du mouvement des failles. Le diagramme de *beachball* indique les directions des axes de compression et de dilatation, ainsi que la polarité des mouvements dans chaque direction (Fig A.1). Les axes de compression (P) et de dilatation (T) sont perpendiculaires les uns aux autres et représentent les forces responsables du déclenchement du séisme.

A.4 Types d'ondes sismiques

Lors de l'enregistrement d'un séisme, les ondes sismiques sont enregistrées dans un certain ordre. Les premières ondes sont les ondes P , généralement impulsives et préférentiellement enregistrées sur la composante verticale. Les ondes Pg se propagent directement de la source du séisme vers les stations sismologiques sans subir de réfraction. En revanche, les ondes Pn , se réfractent lorsqu'elles rencontrent la discontinuité entre le manteau supérieur et la croûte terrestre (Fig A.2). Ces ondes se propagent ensuite le long de la surface de cette discontinuité avant d'atteindre les stations. Pour une profondeur de la discontinuité de Moho de 30 km et des vitesses sismiques dans la croûte et le manteau de 6 et 8 $km.s^{-1}$, les premières ondes réfractées sont visibles à partir d'une distance de 50 à 75 km. Les ondes Pn sont donc enregistrées 2 à 5 sec après les ondes Pg . À partir d'une distance d'environ 110 km, les ondes réfractées précèdent les ondes directes. Les ondes S , se propageant plus lentement que les ondes P , arrivent en second et sont préférentiellement enregistrées sur les composantes horizontales du sismomètre.

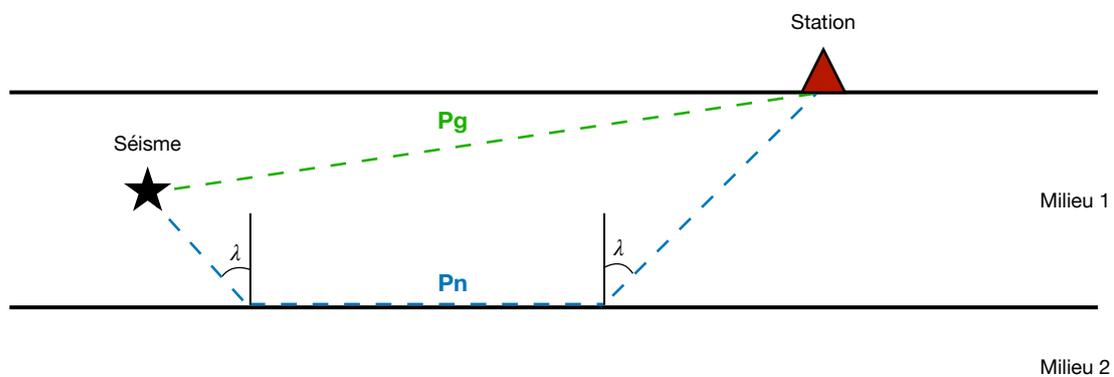


FIGURE A.2 – Schéma explicatif des différents types d'ondes P se propageant dans la croûte du séisme jusqu'à la station sismologique.

Table des figures

1	Réplique du Houfeng Didong Yi.	iv
1.1	Diagrammes montrant les changements de contraintes et séquence de tremblements de terre.	3
1.2	Schéma montrant la géométrie et la cinématique des plaques durant le Paléozoïque.	7
1.3	Localisation des principales failles de surface dans l'ouest de la France.	11
1.4	Carte et coupe synthétiques du Massif armoricain	12
1.5	Carte de l'épaisseur de la croûte terrestre en France Métropolitaine et Espagne.	16
1.6	Cartes des anomalies V_s isotropes.	18
1.7	Tomographie d'absorption pour la France	19
2.1	Sismicité instrumentale globale de 1964 à 2021 pour les événements sismiques de magnitude supérieure à 4. La couleur des symboles correspond à la profondeur des séismes.	22
2.2	Sismicité historique (www.sisfrance.net) entre 462 et 1962 dans le quart nord-ouest de la France.	25
2.3	Évolution du nombre de stations sismologiques permanentes dans le quart Nord-Ouest de la France entre 2000 et 2022.	27
2.4	Sismicité instrumentale de 1962 à 2020 en France métropolitaine.	29
2.5	Proportion des séismes localisés dans le nord ouest par rapport au nombre total de séismes métropolitains en fonction de la magnitude M_W (bins en magnitude de 0,2).	30
2.6	Distributions des magnitudes de moment dans la partie ouest de la France (orange) et dans le reste de la France métropolitaine (vert) pendant deux périodes de temps différentes.	32

TABLE DES FIGURES

2.7	Loi de Gutenberg-Richter en France métropolitaine et dans le Massif armoricain	33
2.8	PSD (Power Spectral Density) à la station BIMF (triangle rouge sur la carte) de janvier 2020 à janvier 2023 entre 0,1 et 1 Hz, en vert. Distribution de la médiane hebdomadaire des magnitudes (M_l) sur la même période, en orange.	34
2.9	Distribution cumulatif des événements naturels et anthropiques de M_l supérieure ou égale à 2,5 en fonction du temps.	36
2.10	Cartes de localisation des (A) des sites de carrière dans la zone d'étude et (B) des tirs de carrières enregistrés du 1er janvier 2013 au 1er juin 2023.	37
2.11	Enregistrements sismiques à 8 stations permanentes du quart nord-ouest de la France lors du passage du météore de 5 septembre 2021 à 21h47 (UTC).	38
2.12	Mécanismes au foyer des tremblements de terre les plus importants au cours des 50 dernières années dans l'ouest de la France.	39
2.13	Orientation de la contrainte compressive horizontale maximale en France métropolitaine et en Europe de l'ouest.	41
2.14	Relation entre l'héritage structural et la sismicité en France métropolitaine.	43
3.1	Fonctionnement du rapport STA/LTA.	51
3.2	Exemple de détection par STA/LTA.	52
3.3	Deux exemples d'inter-corrélations.	54
3.4	Rapport P/S médian pour les explosions et les tremblements de terre dans neuf intervalles de fréquence différents	57
3.5	Schéma explicatif des termes Intelligence Artificielle (IA), <i>machine learning</i> (ML) et <i>deep learning</i> (DL).	58
3.6	Types d'algorithmes de <i>machine learning</i> (ML) ou apprentissage machine.	59
3.7	Schéma explicatif d'un perceptron et des fonctions d'activation.	61
3.8	Schéma explicatif d'une couche de convolution avec l'application d'un filtre de taille 3x3.	62
3.9	Schéma explicatif du <i>max pooling</i> et de l' <i>average pooling</i>	63
3.10	Schéma explicatif de l'architecture d'un Réseau de Neurones Convolutif (CNN).	64
3.11	Schéma explicatif d'un Random Forest.	65
3.12	Intérêt pour les approches d'apprentissage profond en sismologie.	66

3.13	Schéma explicatif de la localisation des événements sismiques de la base de données d’entraînement et de l’architecture de ConvNetQuake. . . .	67
3.14	Exemple d’application de PhaseNet sur le séisme de La Laigne du 16 Juin 2023 localisé à l’ouest de Niort et enregistré par la station CHIF. .	69
4.1	Récapitulatif des détections de la journée du 17 février 2021 réalisées par le BCSF-RéNaSS, EQTransformer, PhaseNet et <i>template matching</i>	75
4.2	Localisation et profondeur de l’ensemble des événements détectés du 1er août 2018 au 31 juillet 2020.	77
4.3	Exemples d’inter-corrélations normalisées à partir d’événements sismiques. . . .	79
4.4	Distribution des valeurs maximales de coefficients d’inter-corrélations normalisées obtenues pour quatre jeux de données.	80
4.5	Traitement des données sur un sismogramme : exemple du séisme du 11 Février 2022 à 15h22, heure locale, au sud de Saint-Malo ($M_l = 1.9$) enregistré par la station MTNF dont la distance hypocentrale est de 83 km.	81
4.6	Paramètres de la polarisation.	82
4.7	Exemples de mouvements de particule de l’onde P calculés à partir des signaux associés.	83
4.8	Exemples de quatre signaux sismiques (A, C, D et G) et de quatre signaux impulsifs (B, E, F et H), enregistrés par les stations du réseau FR, avec les spectrogrammes associés.	85
4.9	Localisation des 8 336 événements constituant la classe <i>quake</i> des bases de données d’entraînement, de validation et de test, détectés entre 1er janvier 2020 et le 1er juin 2021 et étiquetés par le BSCF-RéNaSS. . . .	86
4.10	Architecture du CNN développé.	87
4.11	Exemple de matrice de confusion (modifié d’après Hourcade <i>et coll.</i> (2023)). Chaque cellule d’une matrice de confusion contient une estimation de la performance du CNN.	88
4.12	Exemples d’enveloppes de formes d’ondes enregistrés par la station BEGF provenant de la carrière située à La Vraie Croix à 34 km.	90
4.13	Exemple de classification <i>station-level</i> et <i>network-level</i>	119
4.14	Carte de couverture de stations autour de neuf événements sismiques. .	120
4.15	Lecture de temps d’arrivée de l’onde P	122

TABLE DES FIGURES

4.16	Les étapes du calcul de la fonction caractéristique définie à partir de la kurtosis.	124
4.17	Distributions des différences de temps d'arrivée (en secondes) entre les ondes P (A. et C.) et S (B. et D.) réalisés par les analystes du BCSF-RéNaSS (t_{RNaSS}) ou du LPG (t_{LPG}) et obtenus par kurtosis (t_K).	125
4.18	Distributions des différences de temps d'arrivée (en secondes) entre les ondes P (A. et C.) et S (B. et D.) réalisés par les analystes du BCSF-RéNaSS (t_{RNaSS}) ou du LPG (t_{LPG}) et PhaseNet ($t_{PhaseNet}$).	126
4.19	Exemples de pointés réalisés par PhaseNet sur quatre formes d'ondes sismiques.	128
4.20	Distributions des différences de temps d'arrivée (en secondes) entre les ondes P (A. et C.) et S (B. et D.) réalisés par les analystes du BCSF-RéNaSS (t_{RNaSS}) ou du LPG (t_{LPG}) et EQTransformer (t_{EQT}).	129
5.1	Récapitulatif de la méthode de détection et de discrimination choisie.	132
5.2	Détermination de la durée de la coda de l'onde S à partir de l'amplitude du signal.	134
5.3	Estimation de la déviation de l'hypothèse gaussienne sur deux formes d'ondes synthétiques (A1, B1) suivant la même loi gaussienne.	135
5.4	Détermination de la durée de la coda de l'onde S à partir de l'analyse de la gaussianité.	136
5.5	Magnitudes de durée M_d comparées aux magnitudes locales (M_l) du bulletin du BCSF-RéNaSS.	137
5.6	Localisation des 77 événements de référence et des stations sismiques utilisées.	139
5.7	Localisation des stations utilisées pour l'inter-corrélation selon les deux approches d'analyses.	140
5.8	Matrice de confusion, pour la première étape de discrimination, obtenue sur les détections du 1er janvier au 31 avril 2021, de l'ensemble TM_{all}	141
5.9	Matrice de confusion, pour la seconde étape de discrimination, obtenue sur les détections du 1er janvier au 31 avril 2021, de l'ensemble TM_{all}	142
5.10	Localisation des 391 nouveaux séismes naturels détectés (C_{H2023}).	144
5.11	Carte de densité des 391 nouveaux séismes naturels détectés (C_{H2023}).	145
5.12	Histogramme cumulatif des événements naturels de $M_d \leq 2,5$	146

5.13	Distribution du nombre d'événements sismiques détectés en fonction des heures de la journée avec, en vert, les séismes du catalogue du BCSF-RéNaSS (C_{BCSF}) et en orange, les nouvelles détections (C_{H2023}).	147
5.14	Distributions des magnitudes des événements sismiques détectés entre le 1er janvier 2018 et le 31 décembre 2021.	148
5.15	Distribution des profondeurs des séismes des 803 séismes naturels détectés du 1er Janvier 2018 au 31 Décembre 2021.	149
5.16	Localisation des 803 séismes naturels détectés entre le 1er Janvier 2018 et le 31 Décembre 2021.	151
5.17	Localisation des failles et événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023.	153
5.18	Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023.	154
5.19	Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023.	155
5.20	Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023 sur une section de 100 km à l'ouest du CSA.	156
5.21	Localisation des événements sismiques détectés entre le 1er janvier 2018 et le 1er mars 2023 sur une section de 100 km à l'est du CSA.	157
5.22	Variation du nombre d'événements sismiques naturels pour chaque mois entre Janvier 2019 et Février 2023.	160
5.23	Distribution des précipitations enregistrées à la station Belle-Île et événements sismiques naturels détectés dans un rayon de 75 km de la station du 1er Août 2020 au 1er Août 2021.	161
5.24	Position de masse des stations LOCF et SDOF du 1er mars au 1er mai 2021.	163
5.25	Position de masse de la station LOCF du 1er janvier au 31 décembre 2020, en vert. En rouge, la position temporelle des événements sismiques détectés en Bretagne. En bleu, la position temporelle des événements sismiques localisés à moins de 75 km de LOCF.	164
5.26	Carte des stations utilisées pour la détection par <i>template matching</i>	165
5.27	Nombre d'événements détectés en fonction du seuil d'inter-corrélation choisi.	166
5.28	Carte des stations utilisées pour la localisation.	167

TABLE DES FIGURES

5.29	Diagramme représentant le nombre de fenêtres détectées par <i>template matching</i> (N_d) en fonction du nombre de fenêtres où au moins une onde a été identifiée (N_p).	168
5.30	Récapitulatif des détections de la journées du 17 Février 2021 réalisées par le BCSF-RéNaSS, EQTransformer, PhaseNet, l'utilisation des deux CNNs seulement et enfin notre approche.	169
6.1	Dégâts causés par le séisme sur les villages aux alentours (communications personnelles : É. Beucler et C. Perrin).	172
6.2	Estimation régionale de la secousse du 16/06/2023 18h38, à partir des données macrosismiques et instrumentales. Source : franceseisme.fr . . .	173
6.3	Contexte géologique du séisme La Laigne (étoile jaune). Localisation des failles potentiellement actives selon la BDFA IRSN (Jomard <i>et coll.</i> , 2017) et des structures tectoniques répertoriées sur les cartes géologiques au 1/50000 (BRGM). Source : BRGM	174
6.4	Carte de localisation du séisme du 16 juin 2023 et des répliques associées détectés à partir des stations permanentes.	176
6.5	Installation de capteurs, exemple de (A) la station LC02 et (B) la station LC03. Les capteurs sont enterrés, horizontalisés et orientés vers le nord grâce à une boussole.	178
6.6	Localisation des stations de la campagne post-sismique de La Laigne. . .	178
6.7	Histogramme représentant le nombre de séismes en fonction du temps et une courbe représentant un exemple de loi d'Omori.	181
6.8	Localisation préliminaire des répliques détectées du 17 au 21 juin 2023 et distribution des événements en fonction de la profondeur.	182
A.1	Schématisation d'un mécanisme au foyer	C
A.2	Schéma explicatif des différents types d'ondes P se propageant dans la croûte du séisme jusqu'à la station sismologique.	D

Liste des tableaux

4.1	Tableau comparatif des détections réalisées par PhaseNet, EQTransformer et <i>template matching</i> sur les 131 séismes détectés dans Beucler <i>et coll.</i> (2021).	77
4.2	Tableau comparatif des pointés manuels et réalisés par Kurtosis pour différentes bases de données	123
4.3	Tableau comparatif des pointés manuels et réalisés par PhaseNet pour différentes bases de données	127
4.4	Tableau comparatif des pointés manuels et réalisés par EQTransformer pour différentes bases de données	129

Bibliographie

- Alexandre, P. (1990). “Les séismes en Europe occidentale de 394 à 1259 : nouveau catalogue critique”. *Série Géophysique*.
- Allen, R. (1982). “Automatic phase pickers : Their present use and future prospects”. *Bulletin of the Seismological Society of America* 72(6B), S225–S242.
- Allen, R. V. (1978). “Automatic earthquake recognition and timing from single traces”. *Bulletin of the seismological society of America* 68(5), 1521–1532.
- Allmann, B. P., P. M. Shearer et E. Hauksson (2008). “Spectral discrimination between quarry blasts and earthquakes in southern California”. *Bulletin of the Seismological Society of America* 98(4), 2073–2079.
- Amorèse, D., J. Benjumea et M. Cara (2020). “Source parameters of the 1926 and 1927 Jersey earthquakes from historical, instrumental, and macroseismic data”. *Physics of the Earth and Planetary Interiors* 300, 106420.
- Amorèse, D., A. Walker, J.-L. Lagarde, J.-P. Santoire, P. Volant, M. Font et M. Lecornu (2000). “New seismotectonic data from an intraplate region : focal mechanisms in the Armorican Massif (northwestern France)”. *Geophysical Journal International* 143(3), 837–846.
- Anderson, K. R. (1981). “Epicentral location using arrival time order”. *Bulletin of the Seismological Society of America* 71(2), 541–545.
- Arroucau, P. (2006). *Sismicité du Massif Armoricaïn : relocalisations et interprétation tectonique*. Ph. D. thesis.
- Arroucau, P., A. Mocquet et P. Vacher (2006). “Macroseismic intensity attenuation for Metropolitan France : importance of the epicentral intensity”. *Academie des Sciences Comptes Rendus : Geoscience* 338, 596–605.

-
- Atef, A. H., K. H. Liu et S. S. Gao (2009). “Apparent Weekly and Daily Earthquake Periodicities in the Western United States”. *Bulletin of the Seismological Society of America* 99(4), 2273–2279.
- Baer, M. et U. Kradolfer (1987). “An automatic phase picker for local and teleseismic events”. *Bulletin of the Seismological Society of America* 77(4), 1437–1445.
- Baillard, C., W. C. Crawford, V. Ballu, C. Hibert et A. Mangeney (2014). “An automatic kurtosis-based P-and S-phase picker designed for local seismic networks”. *Bulletin of the Seismological Society of America* 104(1), 394–409.
- Bakun, W. H. et O. Scotti (2006). “Regional intensity attenuation models for France and the estimation of magnitude and location of historical earthquakes”. *Geophysical Journal International* 164(3), 596–610.
- Ballevre, M., V. Bosse, M.-P. Dabard, C. Ducassou, S. Fourcade, J.-L. Paquette, J.-J. Peucat et P. Pitra (2013). “Histoire géologique du Massif armoricain : actualité de la recherche”. *Bulletin de la Société géologique et minéralogique de Bretagne* 500, 5–96.
- Baumgardt, D. R. et G. B. Young (1990). “Regional seismic waveform discriminants and case-based event identification using regional arrays”. *Bulletin of the Seismological Society of America* 80(6B), 1874–1892.
- Beaucé, E., W. B. Frank, A. Paul, M. Campillo et R. D. van Der Hilst (2019). “Systematic detection of clustered seismicity beneath the Southwestern Alps”. *Journal of Geophysical Research : Solid Earth* 124(11), 11531–11548.
- Bergen, K. J. et G. C. Beroza (2018). “Detecting earthquakes over a seismic network using single-station similarity measures”. *Geophysical Journal International* 213(3), 1984–1998.
- Bessin, P. (2014). *Evolution géomorphologique du Massif armoricain depuis 200 Ma : approche Terre-Mer*. Ph. D. thesis, Université Rennes 1.
- Beucler, E., M. Bonnin, C. Hourcade, B. V. Vliet-Lanoë, C. Perrin, L. Provost, A. Mocquet, J. Battaglia, L. Geoffroy, P. Steer, B. L. Gall, J.-M. Douchain, D. Fligiel, P. Gerinon, B. Delouis, J. Perrot, S. Mazzotti, G. Mazet-Roux, S. Lambotte, M. Grunberg, J. Vergne, C. Clément, E. Calais, J. Deverchère, L. Longuevergne, A. Duperret, C. Roques, T. Kaci et C. Authemayou (2021). “Characteristics and possible origins of the seismicity in northwestern France”. *Comptes Rendus. Géoscience* 353(S1), 53–77.

- Beucler, E., A. Mocquet, M. Schimmel, S. Chevrot, O. Quillard, J. Vergne et M. Sylvander (2015). “Observation of deep water microseisms in the North Atlantic Ocean under tide modulations”. *Geophysical Research Letters* 42(2), 316–322.
- Bilek, S. L. (2010). “The role of subduction erosion on seismicity”. *Geology* 38(5), 479–480.
- Bitri, A., J.-P. Brun, J. Chantraine, P. Guennoc, G. Marquis, J.-M. Marthelot, J. Perrin, F. Pivot et C. Truffert (1997). “Structure crustale du bloc cadomien de Bretagne Nord (France) : sismique réflexion verticale et sondage magnétotellurique (projet GéoFrance 3D—Armor)”. *Comptes Rendus de l’Académie des Sciences-Series IIA-Earth and Planetary Science* 325(3), 171–177.
- Bitri, A., J. P. Brun, C. Truffert et P. Guennoc (2001). “Deep seismic imaging of the Cadomian thrust wedge of Northern Brittany”. *Tectonophysics* 331(1-2), 65–80.
- Bois, C. *et coll.* (1988). “Major crustal features disclosed by the ECORS deep seismic profiles”. *Annales de la Société géologique de Belgique*.
- Bollinger, L., F. Perrier, J.-P. Avouac, S. Sapkota, U. Gautam et D. Tiwari (2007). “Seasonal modulation of seismicity in the Himalaya of Nepal”. *Geophysical Research Letters* 34(8).
- Bonnet, S., F. Guillocheau, J.-P. Brun et J. Van Den Driessche (2000). “Large-scale relief development related to Quaternary tectonic uplift of a Proterozoic-Paleozoic basement : The Armorican Massif, NW France”. *Journal of Geophysical Research : Solid Earth* 105(B8), 19273–19288.
- Böse, M., R. Allen, H. Brown, G. Gua, M. Fischer, E. Hauksson, T. Heaton, M. Hellweg, M. Liukis, D. Neuhauser *et coll.* (2014). “CISN ShakeAlert : An earthquake early warning demonstration system for California”. *Early Warning for Geological Disasters : Scientific Methods and Current Practice*, 49–69.
- Bourgeois, O., O. Dauteuil et E. Hallot (2005). “Rifting above a mantle plume : structure and development of the Iceland Plateau”. *Geodinamica acta* 18(1), 59–80.
- Brun, J.-P., P. Guennoc, C. Truffert, J. Vairon *et coll.* (2001). “Cadomian tectonics in northern Brittany : a contribution of 3-D crustal-scale modelling”. *Tectonophysics* 331(1-2), 229–246.

-
- Bungum, H., C. Lindholm et J. Faleide (2005). “Postglacial seismicity offshore mid-Norway with emphasis on spatio-temporal–magnitudal variations”. *Marine and Petroleum Geology* 22(1-2), 137–148.
- Bungum, H., O. Olesen, C. Pascal, S. Gibbons, C. Lindholm et O. Vestøl (2010). “To what extent is the present seismicity of Norway driven by post-glacial rebound?”. *Journal of the Geological Society* 167(2), 373–384.
- Cagnard, F. (2008). “Carte géologique harmonisée du département du Finistère notice géologique-BRGM”. Technical report, RP-56273-FR, 435 p., 12 fig., 2 tab. 3 pl. hors-texte.
- Calais, E., T. Camelbeeck, S. Stein, M. Liu et T. Craig (2016). “A new paradigm for large earthquakes in stable continental plate interiors”. *Geophysical Research Letters* 43(20), 10–621.
- Calais, E., A. Freed, R. Van Arsdale et S. Stein (2010). “Triggering of New Madrid seismicity by late-Pleistocene erosion”. *Nature* 466(7306), 608–611.
- Calais, E., G. Mattioli, C. DeMets, J.-M. Nocquet, S. Stein, A. Newman et P. Rydelek (2005). “Tectonic strain in plate interiors?”. *Nature* 438(7070), E9–E10.
- Campillo, M. et J. Plantet (1991). “Frequency dependence and spatial distribution of seismic attenuation in France : experimental results and possible interpretations”. *Physics of the earth and planetary interiors* 67(1-2), 48–64.
- Cara, M., Y. Cansi, A. Schlupp, P. Arroucau, N. Béthoux, E. Beucler, S. Bruno, M. Calvet, S. Chevrot, A. Deboissy, B. Delouis, M. Denieul, A. Deschamps, C. Doubre, J. Fréchet, S. Godey, O. Golle, M. Grunberg, J. Guilbert, M. Haugmard, L. Jenatton, S. Lambotte, D. Leobal, C. Maron, V. Mendel, S. Merrer, M. Macquet, A. Mignan, A. Mocquet, M. Nicolas, J. Perrot, B. Potin, O. Sanchez, J.-P. Santoire, O. Sèbe, M. Sylvander, F. Thouvenot, J. Van Der Woerd et K. Van Der Woerd (2015, 01). “SI-Hex : a new catalogue of instrumental seismicity for metropolitan France”. *Bulletin de la Société Géologique de France* 186(1), 3–19.
- Cauchie, L. et G. Saccorotti (2013). “Probabilistic inversion of Rayleigh-wave dispersion data : an application to Mt. Etna, Italy”. *Journal of seismology* 17, 335–346.
- Chadwick, R. et D. Evans (1995). “The timing and direction of Permo-Triassic extension in southern Britain”. *Geological Society, London, Special Publications* 91(1), 161–192.

- Chamberlain, C. J., C. J. Hopp, C. M. Boese, E. Warren-Smith, D. Chambers, S. X. Chu, K. Michailos et J. Townend (2018). “EQcorrscan : Repeating and near-repeating earthquake detection and analysis in Python”. *Seismological Research Letters* 89(1), 173–181.
- Champagnac, J., P. Molnar, R. Anderson, C. Sue et B. Delacou (2007). “Quaternary erosion-induced isostatic rebound in the western Alps”. *Geology* 35(3), 195–198.
- Chantraine, J., A. Autran et C. Cavelier (2003). “Carte géologique de la France (version numérique) à l’échelle du millionième”. *BRGM, Orléans*.
- Chantraine, J., J.-J. Chauvel, P. Bale, E. Denis et D. Rabu (1988). “Le Briovérien (Protérozoïque supérieur à terminal) et l’orogénèse cadomienne en Bretagne (France)”. *Bulletin de la Société géologique de France* 4(5), 815–829.
- Chantraine, J., E. Egal, D. Thiéblemont, E. Le Goff, C. Guerrot, M. Ballèvre et P. Guennoc (2001). “The Cadomian active margin (North Armorican Massif, France) : a segment of the north Atlantic Panafrikan belt”. *Tectonophysics* 331(1-2), 1–18.
- Chen, W.-P. et P. Molnar (1983). “Focal depths of intracontinental and intraplate earthquakes and their implications for the thermal and mechanical properties of the lithosphere”. *Journal of Geophysical Research : Solid Earth* 88(B5), 4183–4214.
- Chevrot, S., M. Sylvander et RESIF (2017). “Seismic network X7 :PYROPE PYRenean Observational Portable Experiment (RESIF-SISMOB)”.
- Chevrot, S., A. Villaseñor, M. Sylvander, S. Benahmed, E. Beucler, G. Cougoulat, P. Delmas, M. De Saint Blanquat, J. Diaz, J. Gallart *et coll.* (2014). “High-resolution imaging of the Pyrenees and Massif Central from the data of the PYROPE and IBERARRAY portable array deployments”. *Journal of Geophysical Research : Solid Earth* 119(8), 6399–6420.
- Clark, D., A. McPherson et R. Van Dissen (2012). “Long-term behaviour of Australian stable continental region (SCR) faults”. *Tectonophysics* 566, 1–30.
- Cochran, E. S., J. E. Vidale et S. Tanaka (2004). “Earth tides can trigger shallow thrust fault earthquakes”. *Science* 306(5699), 1164–1166.

-
- Craig, T., E. Calais, L. Fleitout, L. Bollinger et O. Scotti (2016). “Evidence for the release of long-term tectonic strain stored in continental interiors through intraplate earthquakes”. *Geophysical Research Letters* 43(13), 6826–6836.
- Craig, T. J. et E. Calais (2014). “Strain accumulation in the New Madrid and Washash Valley seismic zones from 14 years of continuous GPS observation”. *Journal of Geophysical Research : Solid Earth* 119(12), 9110–9129.
- Cuvier, A., M. Bonnin, É. Beucler et R. Garcia (in prep). “Seismic station monitoring using deviation from the gaussianity”.
- Dadson, S. J., N. Hovius, H. Chen, W. B. Dade, M.-L. Hsieh, S. D. Willett, J.-C. Hu, M.-J. Horng, M.-C. Chen, C. P. Stark *et coll.* (2003). “Links between erosion, runoff variability and seismicity in the Taiwan orogen”. *Nature* 426(6967), 648–651.
- Dai, H. et C. MacBeth (1995). “Automatic picking of seismic arrivals in local earthquake data using an artificial neural network”. *120*, 758–774.
- Damon, A., S. Mazzotti, P. Vernant, R. Vachon, J. Grosset, F. Ego et D. Baumont (2022). “Impact of Far-Field Glacially-Induced Stresses on Fault Stability in the Eastern Paris Basin”. *Available at SSRN 4300077*.
- Delouis, B. (2014). “FMNEAR : Determination of focal mechanism and first estimate of rupture directivity using near-source records and a linear distribution of point sources”. *Bulletin of the Seismological Society of America* 104(3), 1479–1500.
- Denham, D. (1988). “Australian seismicity-the puzzle of the not so stable continent”. *Seismological Research Letters* 59(4), 235–240.
- Dentith, M. et W. Featherstone (2003). “Controls on intra-plate seismicity in southwestern Australia”. *Tectonophysics* 376(3-4), 167–184.
- Díaz, J., A. Villaseñor, J. Gallart, J. Morales, A. Pazos, D. Córdoba, J. Pulgar, J. García-Lobón, M. Harnafi *et coll.* (2009). “The IBERARRAY broadband seismic network : a new tool to investigate the deep structure beneath Iberia”. *Orfeus Newsl* 8(2), 1–6.
- Diehl, T. et E. Kissling (2007). “Users guide for consistent phase picking at local to regional scales”. *Institute of Geophysics, ETH Zurich, Switzerland*.

- Dokht, R. M., H. Kao, R. Visser et B. Smith (2019). “Seismic event and phase detection using time–frequency representation and convolutional neural networks”. *Seismological Research Letters* 90(2A), 481–490.
- Dysart, P. S. et J. J. Pulli (1990). “Regional seismic event classification at the NORESS array : seismological measurements and the use of trained neural networks”. *Bulletin of the Seismological Society of America* 80(6B), 1910–1933.
- Earle, P. S. (1999). “Polarization of the Earth’s teleseismic wavefield”. *Geophysical Journal International* 139(1), 1–8.
- Eisner, L., D. Abbott, W. B. Barker, J. Lakings et M. P. Thornton (2008). “Noise suppression for detection and location of microseismic events using a matched filter”. In *SEG Technical Program Expanded Abstracts 2008*, pp. 1431–1435. Society of Exploration Geophysicists.
- Faure, M. (1995). “Late orogenic carboniferous extensions in the Variscan French Massif Central”. *Tectonics* 14(1), 132–153.
- Fedorenko, Y. V., E. S. Husebye et B. O. Ruud (1999). “Explosion site recognition ; neural net discriminator using single three-component stations”. 113, 131–142.
- Feldl, N. et R. Bilham (2006). “Great Himalayan earthquakes and the Tibetan plateau”. *Nature* 444(7116), 165–170.
- Fiedler, B., S. Hainzl, G. Zöller et M. Holschneider (2018). “Detection of Gutenberg–Richter b-Value Changes in Earthquake Time SeriesDetection of Gutenberg–Richter b-Value Changes in Earthquake Time Series”. *Bulletin of the Seismological Society of America* 108(5A), 2778–2787.
- Frank, W. et N. Shapiro (2014). “Automatic detection of low-frequency earthquakes (LFEs) based on a beamformed network response”. *Geophysical Journal International* 197(2), 1215–1223.
- Frankel, A., R. Smalley et J. Paul (2012). “Significant motions between GPS sites in the New Madrid region : Implications for seismic hazard”. *Bulletin of the Seismological Society of America* 102(2), 479–489.
- Fu, J., X. Wang, Z. Li, H. Meng, J. Wang, W. Wang et C. Tang (2020). “Automatic Phase-Picking Method for Detecting Earthquakes Based on the Signal-to-Noise-Ratio Concept”. *Seismological Research Letters* 91(1), 334–342.

-
- Gallen, S. F. et J. R. Thigpen (2018). “Lithologic controls on focused erosion and intraplate earthquakes in the eastern Tennessee seismic zone”. *Geophysical Research Letters* 45(18), 9569–9578.
- Gaudot, I., É. Beucler, A. Mocquet, M. Drilleau, M. Haugmard, M. Bonnin, G. Aertgeerts et D. Leparoux (2020, 11). “3-D crustal VS model of western France and the surrounding regions using Monte Carlo inversion of seismic noise cross-correlation dispersion diagrams”. *Geophysical Journal International* 224(3), 2173–2188.
- Gautier, M. (1975). “Les tremblements de terre du Massif Armoricain”. *Norvès* 86(1), 189–205.
- Genti, M., J. Chery, P. Vernant et A. Rigo (2016). “Impact of gravity forces and topography denudation on normal faulting in Central–Western Pyrenees : Insights from 2D numerical models”. *Comptes Rendus Geoscience* 348(3-4), 173–183.
- Gibbons, S. J. et F. Ringdal (2006). “The detection of low magnitude seismic events using array-based waveform correlation”. *Geophysical Journal International* 165(1), 149–166.
- Gölke, M. et D. Coblenz (1996). “Origins of the European regional stress field”. *Tectonophysics* 266(1-4), 11–24.
- Grünthal, G. (1998). “European macroseismic scale 1998”. Technical report, European Seismological Commission (ESC).
- Gumiaux, C. (2003). *Modélisation du cisaillement hercynien de Bretagne centrale : déformation crustale et implications lithosphériques*. Géosciences Rennes, Université de Rennes I.
- Gumiaux, C., S. Judenherc, J.-P. Brun, D. Gapais, M. Granet et G. Poupinet (2004). “Restoration of lithosphere-scale wrenching from integrated structural and tomographic data (Hercynian belt of western France)”. *Geology* 32(4), 333–336.
- Gutenberg, B. et C. F. Richter (1956). “Earthquake magnitude, intensity, energy, and acceleration : (Second paper)”. *Bulletin of the seismological society of America* 46(2), 105–145.
- Hanks, T. C. et H. Kanamori (1979). “A moment magnitude scale”. *Journal of Geophysical Research : Solid Earth* 84(B5), 2348–2350.

- Hao, J., J. Zhang et Z. Yao (2019). “Evidence for diurnal periodicity of earthquakes from midnight to daybreak”. *National Science Review* 6(5), 1016–1023.
- Haugmard, M. (2016). *Détermination non-linéaire des paramètres hypocentaux et structuraux : application à la sismicité intracontinentale du Massif armoricain*. Ph. D. thesis, Nantes.
- Heidbach, O., M. Rajabi, X. Cui, K. Fuchs, B. Müller, J. Reinecker, K. Reiter, M. Tingay, F. Wenzel, F. Xie *et coll.* (2018). “The World Stress Map database release 2016 : Crustal stress pattern across scales”. *Tectonophysics* 744, 484–498.
- Hetényi, G., I. Molinari, J. Clinton, G. Bokelmann, I. Bondár, W. C. Crawford, J.-X. Dessa, C. Doubre, W. Friederich, F. Fuchs *et coll.* (2018). “The AlpArray seismic network : a large-scale European experiment to image the Alpine Orogen”. *Surveys in geophysics* 39, 1009–1033.
- Hillis, R. R., M. Sandiford, S. D. Reynolds et M. C. Quigley (2008). “Present-day stresses, seismicity and Neogene-to-Recent tectonics of Australia’s ‘passive’ margins : intraplate deformation controlled by plate boundary forces”. *Geological Society, London, Special Publications* 306(1), 71–90.
- Hirn, A. (1976). “Sondages sismiques profonds en France”. *Bulletin de la Société Géologique de France* 7(5), 1065–1071.
- Hourcade, C., M. Bonnin et É. Beucler (2023). “New CNN-based tool to discriminate anthropogenic from natural low magnitude seismic events”. *Geophysical Journal International* 232(3), 2119–2132.
- Israelsson, H. (1990). “Correlation of waveforms from closely spaced regional events”. *Bulletin of the Seismological Society of America* 80(6B), 2177–2193.
- Jackson, J. et D. McKenzie (2022). “The exfoliation of cratonic Australia in earthquakes”. *Earth and Planetary Science Letters* 578, 117305.
- Jéandet Ribes, L., N. Cubas, H. S. Bhat et P. Steer (2020). “The impact of large erosional events and transient normal stress changes on the seismicity of faults”. *Geophysical Research Letters* 47(22), e2020GL087631.
- Jégouzo, P. et E. Rossello (1988). “La Branche Nord du Cisaillement Sud-Armoricain (France) : un essai d’évaluation du déplacement par l’analyse des mylonites”. *Comptes*

rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre 307(17), 1825–1831.

Johnson, P. et T. McEvilly (1995). “Parkfield seismicity : Fluid-driven?”. *Journal of Geophysical Research : Solid Earth* 100(B7), 12937–12950.

Johnston, A. C. (1989). “The seismicity of ‘stable continental interiors’”. *Earthquakes at North-Atlantic passive margins : Neotectonics and postglacial rebound*, 299–327.

Jomard, H., E. M. Cushing, L. Palumbo, S. Baize, C. David et T. Chartier (2017). “Transposing an active fault database into a seismic hazard fault model for nuclear facilities—Part 1 : Building a database of potentially active faults (BDFA) for metropolitan France”. *Natural Hazards and Earth System Sciences* 17(9), 1573–1584.

Judenherc, S. (2000). *Etude et caractérisation des structures hercyniennes à partir de données sismologiques : le cas du Massif Armoricaïn*. Ph. D. thesis, Université Louis Pasteur-Strasbourg I.

Judenherc, S., M. Granet et N. Boumbar (1999). “Two-dimensional anisotropic tomography of lithosphere beneath France using regional arrival times”. *Journal of Geophysical Research : Solid Earth* 104(B6), 13201–13215.

Judenherc, S., M. Granet, J.-P. Brun et G. Poupinet (2003). “The Hercynian collision in the Armorican Massif : evidence of different lithospheric domains inferred from seismic tomography and anisotropy”. *Bulletin de la Société géologique de France* 174(1), 45–57.

Judenherc, S., M. Granet, J.-P. Brun, G. Poupinet, J. Plomerová, A. Mocquet et U. Achauer (2002). “Images of lithospheric heterogeneities in the Armorican segment of the Hercynian Range in France”. *Tectonophysics* 358(1-4), 121–134.

Judenherc, S., M. Granet, J.-P. Brun, G. Poupinet, J. Plomerová, A. Mocquet et U. Achauer (2002). “Images of lithospheric heterogeneities in the Armorican segment of the Hercynian Range in France”. 358, 121–134.

Jurkevics, A. (1988). “Polarization analysis of three-component array data”. *Bulletin of the seismological society of America* 78(5), 1725–1743.

Kanamori, H. et E. E. Brodsky (2004). “The physics of earthquakes”. *Reports on progress in physics* 67(8), 1429.

- Kanasewich, E. R. (1981). *Time sequence analysis in geophysics*. University of Alberta.
- Kaub, C., L. Geoffroy, L. Bollinger, J. Perrot, P. Le Roy et C. Authemayou (2021). “Is the Machecoul fault the source of the M 6 1799 Vendée earthquake (France) ?”. *Geophysical Journal International* 225(3), 2035–2059.
- Klose, C. D. et L. Seeber (2007). “Shallow seismicity in stable continental regions”. *Seismological Research Letters* 78(5), 554–562.
- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade et P. Gerstoft (2018, 11). “Machine Learning in Seismology : Turning Data into Insights”. *Seismological Research Letters* 90(1), 3–14.
- Koper, K. D., J. C. Pechmann, R. Burlacu, K. L. Pankow, J. Stein, J. M. Hale, P. Robertson et M. K. McCarter (2016). “Magnitude-based discrimination of man-made seismic events from naturally occurring earthquakes in Utah, USA”. *Geophysical Research Letters* 43(20), 10–638.
- Kraft, T., J. Wassermann, E. Schmedes et H. Igel (2006). “Meteorological triggering of earthquake swarms at Mt. Hochstaufen, SE-Germany”. *Tectonophysics* 424(3-4), 245–258.
- Kurz, J. H., T. Jahr et G. Jentzsch (2004). “Earthquake swarm examples and a look at the generation mechanism of the Vogtland/Western Bohemia earthquake swarms”. *Physics of the Earth and Planetary Interiors* 142(1-2), 75–88.
- Langston, C. A. (1977). “Corvallis, Oregon, crustal and upper mantle receiver structure from teleseismic P and S waves”. *Bulletin of the Seismological Society of America* 67(3), 713–724.
- Le Gall, B., C. Authemayou, D. Graindorge, A. Duperret, T. Kaci, A. Ehrhold et T. Schmitt (2021). “Status of basement terranes in orogens : insights from the Cadomian domain of the North Armorican Variscides, Western France”. *Tectonics* 40(5), e2020TC006578.
- Leclère, H. et É. Calais (2019). “A parametric analysis of fault reactivation in the New Madrid Seismic Zone : The role of pore fluid overpressure”. *Journal of Geophysical Research : Solid Earth* 124(10), 10630–10648.

-
- Lee, W. H. K., R. Bennett et K. Meagher (1972). *A method of estimating magnitude of local earthquakes from signal duration*. US Department of the Interior, Geological Survey.
- Lee, W. H. K., W. H. K. Lee et S. Stewart (1981). *Principles and applications of microearthquake networks*, Volume 2. Academic press.
- Levandowski, W., R. B. Herrmann, R. Briggs, O. Boyd et R. Gold (2018). “An updated stress map of the continental United States reveals heterogeneous intraplate stress”. *Nature Geoscience* 11(6), 433–437.
- Li, S., B. Liu, Y. Ren, Y. Chen, S. Yang, Y. Wang et P. Jiang (2019). “Deep-learning inversion of seismic data”. *arXiv preprint arXiv :1901.07733*.
- Li, Z., M.-A. Meier, E. Hauksson, Z. Zhan et J. Andrews (2018). “Machine learning seismic wave discrimination : Application to earthquake early warning”. *Geophysical Research Letters* 45(10), 4773–4779.
- Linville, L., K. Pankow et T. Draelos (2019). “Deep learning models augment analyst decisions for event discrimination”. *Geophysical Research Letters* 46(7), 3643–3651.
- Liu, X., T. Ren, H. Chen et Y. Chen (2021). “Classification of tectonic and non-tectonic seismicity based on convolutional neural network”. *Geophysical Journal International* 224(1), 191–198.
- Long, L. T. (1988). “A model for major intraplate continental earthquakes”. *Seismological research letters* 59(4), 273–278.
- Mandal, P., R. Narsaiah, P. Raju et R. Chadha (2004). “Coda duration magnitude scale of 2001 Bhuj aftershocks, India”. *Current Science*, 520–527.
- Martín-González, F., C. Crespo-Martín, S. Cesca et S. González-Muñoz (2023). “Understanding seismicity and seismotectonics in a stable continental region (NW Iberian Peninsula) : Implications for the nature of intraplate seismicity”. *Global and Planetary Change*, 104177.
- Matte, P. (1986a). “La chaîne varisque parmi les chaînes paléozoïques péri atlantiques, modèle d’évolution et position des grands blocs continentaux au Permo-Carbonifère”. *Bulletin de la Société géologique de France* 2(1), 9–24.

- Matte, P. (1986b). “Tectonics and plate tectonics model for the Variscan belt of Europe”. *Tectonophysics* 126(2-4), 329–374.
- Matte, P. (2001). “The Variscan collage and orogeny (480–290 Ma) and the tectonic definition of the Armorica microplate : a review”. *Terra nova* 13(2), 122–128.
- Mayor, J. (2016). *Variations régionales de l’atténuation sismique en France métropolitaine : observations et modélisation*. Ph. D. thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier.
- Mazabraud, Y., N. Béthoux et B. Delouis (2013). “Is earthquake activity along the French Atlantic margin favoured by local rheological contrasts?”. *Comptes Rendus Geoscience* 345(9-10), 373–382.
- Mazzotti, S., C. Aubagnac, L. Bollinger, K. C. Oscanoa, B. Delouis, D. Do Paco, C. Doubre, M. Godano, H. Jomard, C. Larroque *et coll.* (2021). “FMHex20 : An earthquake focal mechanism database for seismotectonic analyses in metropolitan France and bordering regions”. *Bulletin De La Société Géologique De France* 192(1).
- Mazzotti, S. et F. Gueydan (2018). “Control of tectonic inheritance on continental intraplate strain rate and seismicity”. *Tectonophysics* 746, 602–610.
- Mazzotti, S., H. Jomard et F. Masson (2020). “Processes and deformation rates generating seismicity in metropolitan France and conterminous Western Europe”. *Bulletin de la Société Géologique de France* 191(1).
- Mazzotti, S., X. Vergeron, O. Malcles, J. Grosset et P. Vernant (2023). “Impact of long-term erosion on crustal stresses and seismicity in stable continental regions”. *Geology*.
- Medvedev, S., W. Sponheuer et V. Karník (1964). “Neue seismische Skala Intensity scale of earthquakes, 7. Tagung der Europäischen Seismologischen Kommission vom 24.9. bis 30.9. 1962”. *Jena, Veröff. Institut für Bodendynamik und Erdbebenforschung in Jena* 77, 69–76.
- Meier, M.-A., Z. E. Ross, A. Ramachandran, A. Balakrishna, S. Nair, P. Kundzicz, Z. Li, J. Andrews, E. Hauksson et Y. Yue (2019). “Reliable real-time seismic signal/noise discrimination with machine learning”. *Journal of Geophysical Research : Solid Earth* 124(1), 788–800.

-
- Mousavi, S. M. et G. C. Beroza (2022). “Deep-learning seismology”. *Science* 377(6607), eabm4470.
- Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang et G. C. Beroza (2020). “Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking”. *Nature communications* 11(1), 1–12.
- Mousavi, S. M., Y. Sheng, W. Zhu et G. C. Beroza (2019). “STanford EArthquake Dataset (STEAD) : A Global Data Set of Seismic Signals for AI”. 7, 179464–179476.
- Müller, B., M. L. Zoback, K. Fuchs, L. Mastin, S. Gregersen, N. Pavoni, O. Stephansson et C. Ljunggren (1992). “Regional patterns of tectonic stress in Europe”. *Journal of Geophysical Research : Solid Earth* 97(B8), 11783–11803.
- Nakano, M., D. Sugiyama, T. Hori, T. Kuwatani et S. Tsuboi (2019, 01). “Discrimination of Seismic Signals from Earthquakes and Tectonic Tremor by Applying a Convolutional Neural Network to Running Spectral Images”. *Seismological Research Letters* 90(2A), 530–538.
- Nicolas, M., J. Santoire et P. Delpech (1990). “Intraplate seismicity : new seismotectonic data in Western Europe”. *Tectonophysics* 179(1-2), 27–53.
- Nocquet, J.-M. (2012). “Present-day kinematics of the Mediterranean : A comprehensive overview of GPS results”. *Tectonophysics* 579, 220–242.
- Omori, F. (1894). *On the after-shocks of earthquakes*, Volume 7. The University.
- O’Rourke, C. T. et G. E. Baker (2017). “A spectrogram-based method of Rg detection for explosion monitoring”. *Bulletin of the Seismological Society of America* 107(1), 34–42.
- Panning, M. P., É. Beucler, M. Drilleau, A. Mocquet, P. Lognonné et W. B. Banerdt (2015). “Verifying single-station seismic approaches using Earth-based data : Preparation for data return from the InSight mission to Mars”. *Icarus* 248, 230–242.
- Park, Y., S. M. Mousavi, W. Zhu, W. L. Ellsworth et G. C. Beroza (2020). “Machine-Learning-Based Analysis of the Guy-Greenbrier, Arkansas Earthquakes : A Tale of Two Sequences”. 47.
- Peng, Z. et P. Zhao (2009). “Migration of early aftershocks following the 2004 Parkfield earthquake”. *Nature Geoscience* 2(12), 877–881.

-
- Perol, T., M. Gharbi et M. Denolle (2018). “Convolutional neural network for earthquake detection and location”. *Science Advances* 4(2), e1700578.
- Perrot, J., P. Arroucau, J. Guilbert, J. Déverchère, Y. Mazabraud, J. Rolet, A. Mocquet, M. Mousseau et L. Matias (2005). “Analysis of the Mw 4.3 Lorient earthquake sequence : a multidisciplinary approach to the geodynamics of the Armorican Massif, westernmost France”. *Geophysical Journal International* 162(3), 935–950.
- Plenkers, K., J. R. Ritter et M. Schindler (2013). “Low signal-to-noise event detection based on waveform stacking and cross-correlation : Application to a stimulation experiment”. *Journal of seismology* 17(1), 27–49.
- Pulli, J. J. et P. S. Dysart (1990). “An experiment in the use of trained neural networks for regional seismic event classification”. *Geophysical Research Letters* 17(7), 977–980.
- Pyle, M. L. et W. R. Walter (2019). “Investigating the effectiveness of P/S amplitude ratios for local distance event discrimination”. *Bulletin of the Seismological Society of America* 109(3), 1071–1081.
- Renouard, A., A. Maggi, M. Grunberg, C. Doubre et C. Hibert (2021, 06). “Toward False Event Detection and Quarry Blast versus Earthquake Discrimination in an Operational Setting Using Semiautomated Machine Learning”. *Seismological Research Letters* 92(6), 3725–3742.
- RESIF (1995). “RESIF-RAP French Accelerometric Network”.
- RESIF (1995). “RESIF-RLBP French Broad-band network, RESIF-RAP strong motion network and other seismic stations in metropolitan France. RESIF – Réseau Sismologique et géodésique Français. doi :10.15778/resif.fr”.
- RESIF (2018). “CEA/DASE broad-band permanent network in metropolitan France”.
- Retailleau, L., J.-M. Saurel, M. Laporte, A. Lavayssière, V. Ferrazzini, W. Zhu, G. C. Beroza, C. Satriano, J.-C. Komorowski, O. Team *et coll.* (2022). “Automatic detection for a comprehensive view of Mayotte seismicity”. *Comptes Rendus. Géoscience* 354(S2), 1–18.
- Richter, C. F. (1958). “ELEMENTARY SEISMOLOGY..”.

-
- Roeloffs, E. A. (1988). “Hydrologic precursors to earthquakes : A review”. *Pure and Applied Geophysics* 126, 177–209.
- Ronneberger, O., P. Fischer et T. Brox (2015). “U-Net : Convolutional Networks for Biomedical Image Segmentation”. In N. Navab, J. Hornegger, W. M. Wells, et A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, pp. 234–241. Springer International Publishing.
- Ross, Z. E., M.-A. Meier, E. Hauksson et T. H. Heaton (2018). “Generalized Seismic Phase Detection with Deep Learning”. *Bulletin of the Seismological Society of America* 108(5A), 2894–2901.
- Rouland, D., D. Legrand, M. Zhizhin et S. Vergnolle (2009). “Automatic detection and discrimination of volcanic tremors and tectonic earthquakes : An application to Ambrym volcano, Vanuatu”. *Journal of volcanology and geothermal research* 181(3–4), 196–206.
- Sambridge, M. et B. Kennett (1986). “A novel method of hypocentre location”. *Geophysical Journal International* 87(2), 679–697.
- Savage, M. K. et J. G. Anderson (1995). “A local-magnitude scale for the western Great Basin-eastern Sierra Nevada from synthetic Wood-Anderson seismograms”. *Bulletin of the Seismological Society of America* 85(4), 1236–1243.
- Schaff, D. P. et F. Waldhauser (2010). “One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity”. *Bulletin of the Seismological Society of America* 100(6), 3224–3238.
- Schimmel, M., E. Stutzmann, F. Arduin et J. Gallart (2011). “Polarized Earth’s ambient microseismic noise”. *Geochemistry, Geophysics, Geosystems* 12(7).
- Schorlemmer, D., F. Euchner, P. Kästli, J. Saul, Q. W. Group et coll. (2011). “QuakeML : status of the XML-based seismological data exchange format”. *Annals of Geophysics* 54(1).
- Sharma, B., A. Kumar et V. Murthy (2010). “Evaluation of seismic events detection algorithms”. *Journal of the Geological Society of India* 75(3), 533–538.
- Shearer, P. M. (1994). “Global seismic event detection using a matched filter on long-period seismograms”. *Journal of Geophysical Research : Solid Earth* 99(B7), 13713–13725.

-
- Shelly, D. R., G. C. Beroza et S. Ide (2007). “Non-volcanic tremor and low-frequency earthquake swarms”. *Nature* 446(7133), 305–307.
- Shiddiqi, H. A., L. Ottemöller, S. Rondenay, S. Custódio, V. K. Gahalaut, R. K. Yadav, F. Halpaap et K. Gahalaut (2023). “Seismicity modulation due to hydrological loading in a stable continental region : a case study from the Jektvik swarm sequence in Northern Norway”. *Geophysical Journal International* 235(1), 231–246.
- Soto-Cordero, L., A. Meltzer et J. Stachnik (2018). “Crustal structure, intraplate seismicity, and seismic hazard in the mid-Atlantic United States”. *Seismological Research Letters* 89(1), 241–252.
- Steer, P., L. Jeandet, N. Cubas, O. Marc, P. Meunier, M. Simoes, R. Cattin, J. B. H. Shyu, M. Mouyen, W.-T. Liang *et coll.* (2020). “Earthquake statistics changed by typhoon-driven erosion”. *Scientific reports* 10(1), 10899.
- Steer, P., M. Simoes, R. Cattin et J. B. H. Shyu (2014). “Erosion influences the seismicity of active thrust faults”. *Nature communications* 5(1), 5564.
- Steffen, R., H. Steffen, P. Wu et D. W. Eaton (2014). “Stress and fault parameters affecting fault slip magnitude and activation time during a glacial cycle”. *Tectonics* 33(7), 1461–1476.
- Stein, S. et S. Mazzotti (2007). *Continental intraplate earthquakes : science, hazard, and policy issues*, Volume 425. Geological Society of America.
- Su, F., K. Aki et N. Biswas (1991). “Discriminating quarry blasts from earthquakes using coda waves”. *Bulletin of the Seismological Society of America* 81(1), 162–178.
- Swindell, W. et N. Snell (1977). “Station processor automatic signal detection system, phase I : Final report, station processor software development”. *Texas Instruments Final Report No. ALEX (01)-FR-77 1*.
- Swiss Seismological Service (SED) At ETH Zurich (1983). “National Seismic Networks of Switzerland”.
- Sykes, L. R. (1978). “Intraplate seismicity, reactivation of preexisting zones of weakness, alkaline magmatism, and other tectonism postdating continental fragmentation”. *Reviews of Geophysics* 16(4), 621–688.

-
- Tharwat, A. (2020). “Classification assessment methods”. *Applied Computing and Informatics*.
- Thomas, A., R. Bürgmann, D. R. Shelly, N. M. Beeler et M. Rudolph (2012). “Tidal triggering of low frequency earthquakes near Parkfield, California : Implications for fault mechanics within the brittle-ductile transition”. *Journal of Geophysical Research : Solid Earth* 117(B5).
- Tibi, R., K. D. Koper, K. L. Pankow et C. J. Young (2018). “Depth discrimination using R_g-to-S_g spectral amplitude ratios for seismic events in Utah recorded at local distances”. *Bulletin of the Seismological Society of America* 108(3A), 1355–1368.
- Titos, M., A. Bueno, L. Garcia, C. Benitez et J. C. Segura (2020). “Classification of Isolated Volcano-Seismic Events Based on Inductive Transfer Learning”. 17, 869–873.
- Trnkoczy, A. (2009). “Understanding and parameter setting of STA/LTA trigger algorithm”. In *New manual of seismological observatory practice (NMSOP)*, pp. 1–20. Deutsches GeoForschungsZentrum GFZ.
- University of Genoa (1967). “Regional Seismic Network of North Western Italy”.
- Utsu, T., Y. Ogata *et coll.* (1995). “The centenary of the Omori formula for a decay law of aftershock activity”. *Journal of Physics of the Earth* 43(1), 1–33.
- Vigneressesee, J.-L. (1988). “La fracturation post-hercynienne du Massif armoricain d’après les données géophysiques”. *Géologie de la France* 4, 3–10.
- Waldhauser, F. et W. L. Ellsworth (2000). “A double-difference earthquake location algorithm : Method and application to the northern Hayward fault, California”. *Bulletin of the seismological society of America* 90(6), 1353–1368.
- Waldhauser, F. et D. P. Schaff (2008). “Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods”. *Journal of Geophysical Research : Solid Earth* 113(B8).
- Wang, R., B. Schmandt et E. Kiser (2020). “Seismic discrimination of controlled explosions and earthquakes near Mount St. Helens using P/S ratios”. *Journal of Geophysical Research : Solid Earth* 125(10), e2020JB020338.

- Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore et J. Trujillo (1998). “A comparison of select trigger algorithms for automated global seismic phase and event detection”. *Bulletin of the Seismological Society of America* 88(1), 95–106.
- Wu, P. (1998). “Intraplate earthquakes and post-glacial rebound in eastern Canada and Northern Europe”. *Dynamics of the Ice Age Earth : A Modern Perspective* 603, e628.
- Wu, Y., Y. Lin, Z. Zhou, D. C. Bolton, J. Liu et P. Johnson (2018). “DeepDetect : A cascaded region-based densely connected network for seismic event detection”. *IEEE Transactions on Geoscience and Remote Sensing* 57(1), 62–75.
- Wurman, G., R. M. Allen et P. Lombard (2007). “Toward earthquake early warning in northern California”. *Journal of Geophysical Research : Solid Earth* 112(B8).
- Yoon, C. E., O. O’Reilly, K. J. Bergen et G. C. Beroza (2015). “Earthquake detection through computationally efficient similarity search”. *Science advances* 1(11), e1501057.
- Žalohar, J. (2018). *The Omega-theory : A New physics of earthquakes*. Elsevier.
- Zhang, M. et L. Wen (2015). “An effective method for small event detection : match and locate (M&L)”. *Geophysical Journal International* 200(3), 1523–1537.
- Zhao, Y. et K. Takano (1999). “An artificial neural network approach for broadband seismic phase picking”. 89, 670–680.
- Zhou, P., , W. L. Ellsworth, H. Yang, Y. J. Tan, G. C. Beroza, M. Sheng, R. Chu, et and (2021). “Machine-learning-facilitated earthquake and anthropogenic source detections near the Weiyuan Shale Gas Blocks, Sichuan, China”. 5, 501–519.
- Zhu, L., Z. Peng, J. McClellan, C. Li, D. Yao, Z. Li et L. Fang (2019). “Deep learning for seismic phase detection and picking in the aftershock zone of 2008 Mw7. 9 Wenchuan Earthquake”. *Physics of the Earth and Planetary Interiors* 293, 106261.
- Zhu, W. et G. C. Beroza (2019). “PhaseNet : a deep-neural-network-based seismic arrival-time picking method”. *Geophysical Journal International* 216(1), 261–273.
- Zoback, M. L. (1992). “Stress field constraints on intraplate seismicity in eastern North America”. *Journal of Geophysical Research : Solid Earth* 97(B8), 11761–11782.

Zoback, M. L., M. D. Zoback, J. Adams, M. Assumpcao, S. Bell, E. Bergman, P. Blümling, N. Brereton, D. Denham, J. Ding *et coll.* (1989). “Global patterns of tectonic stress”. *Nature* 341(6240), 291–298.

Titre : Caractérisation de la sismicité naturelle d'une région continentale stable : Application au Massif Armoricaïn

Mots clés : Sismicité de faible magnitude, détection, discrimination, machine learning

Résumé : Les différentes régions continentales stables dans le monde, telles que le Massif armoricain en France, se caractérisent par des taux de déformation très faibles et une sismicité faible à modérée. Devant un forçage tectonique très faible, voire imperceptible, les processus qui gouvernent l'occurrence et la localisation de cette sismicité demeurent mal compris. Dans le Massif armoricain, cela a également été grandement entravée par la couverture limitée du réseau de stations jusqu'en 2015. Au cours des dix dernières années, l'augmentation du nombre de stations dans la région a ouvert la possibilité de détecter les événements sismiques manquants et de mieux caractériser la sismicité de la région. Au-delà de l'aspect purement instrumental, cette compréhension de la sismicité de faible magnitude est fortement influencée par la méthode de détection utilisée pour identifier les événements naturels.

Afin d'améliorer la détection de ces tremblements de terre, nous avons développé et mis en place des outils de détection et de discrimination automatique. Pour la détection, nous avons adopté une approche de type *template matching*, nous permettant d'identifier dans le signal les formes d'ondes similaires à celles d'un séisme. Cependant, cette méthode peut engendrer un taux élevé de faux événements. Pour résoudre ce problème, nous avons créé un outil basé sur le *machine learning*, permettant une discrimination rapide et fiable de ces événements afin de n'extraire que les séismes détectés. Un deuxième outil de *machine learning* a été mis en place pour différencier les événements sismiques naturels (d'origine tectonique) des événements anthropiques (d'origine humaine). L'application de ces outils sur quatre années de données sismiques a apporté des résultats prometteurs. Grâce à cette étude, nous avons pu doubler le nombre d'événements sismiques naturels détectés dans la région. Ce travail de thèse permet donc une cartographie plus fine de la sismicité de la région, essentielle pour une meilleure compréhension de cette sismicité en domaine continental stable. Malgré sa caractéristique diffuse, l'analyse de 4 années du signal sismique montre une plus forte activité sismique localisée Sud-Bretagne, proche des structures majeures telles que le cisaillement sud-armoricain.

Title : Characterisation of the natural seismicity of a stable continental region: Application to the Armorican Massif

Keywords : Low magnitude seismicity, detection, discrimination, machine learning

Abstract : The various stable continental regions in the world, such as the Armorican Massif in France, are characterised by very low deformation rates and low to moderate seismicity. Given that tectonic forcing is very weak, or even imperceptible, the processes that govern the occurrence and location of this seismicity remain poorly understood. In the Armorican Massif, this was also greatly hampered by the limited coverage of the station network until 2015. Over the last ten years, the increase in the number of stations in the region has opened up the possibility of detecting missing seismic events and better characterising the region's seismicity. Beyond the purely instrumental aspect, this understanding of low-magnitude seismicity is strongly influenced by the detection method used to identify natural events.

In order to improve the detection of these earthquakes, we have developed and implemented automatic detection and discrimination tools. For detection, we have adopted a template matching approach, enabling us to identify waveforms in the signal that are similar to those of an earthquake. However, this method can lead to a high rate of false events. To solve this problem, we have created a machine learning-based tool that can quickly and reliably discriminate between these events and extract only the earthquakes that have been detected. Finally, a second machine learning tool was developed to differentiate between natural (tectonic) and man-made (anthropogenic) seismic events. The application of these tools to four years of seismic data has produced promising results. Thanks to this study, we were able to double the number of natural seismic events detected in the region. This thesis work has therefore enabled us to map the seismicity of the region more precisely, which is essential for a better understanding of this seismicity in a stable continental domain. Despite its diffuse nature, a 4-year analysis of the seismic signal shows stronger seismic activity in South Brittany, close to major structures such as the South Armorican shear.