Train traffic as a seismic noise source for imaging the shallow crust with seismic interferometry

Laura Alejandra Pinzon Rincon

Laura Alejandra Pinzon Rincon. Train traffic as a seismic noise source for imaging the shallow crust with seismic interferometry. Earth Sciences. Université Grenoble Alpes [2020-..], 2021. English. NNT : 2021GRALU034 . tel-03794052

HAL Id: tel-03794052
https://theses.hal.science/tel-03794052v1

Submitted on 3 Oct 2022

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Terre Solide (CETSOL)
Arrêté ministériel : 25 mai 2016

Présentée par

Laura Alejandra PINZON RINCON

Thèse dirigée par **Florent BRENGUIER**,enseignant chercheur
et co-encadrée par **Pierre BOUé**,Enseignant-chercheur, UGA

préparée au sein du **Laboratoire Institut des Sciences de la Terre**
dans **l'École Doctorale Sciences de la Terre de lEnvironnement et des Planètes**

Le trafic ferroviaire comme source de bruit sismique pour l'imagerie de la croûte peu profonde par interférométrie sismique

Train traffic as a seismic noise source for imaging the shallow crust with seismic interferometry

Thèse soutenue publiquement le **10 décembre 2021**,
devant le jury composé de :

Monsieur Florent BRENGUIER
PHYSICIEN, Université Grenoble Alpes, Directeur de thèse
Monsieur Gregor HILLERS
PROFESSEUR, Helsingin Yliopisto , Rapporteur
Madame Elena KOZLOVSKAYA
PROFESSEUR, Oulun Yliopisto , Rapporteure
Madame Hanneke PAULSSEN
PROFESSEUR ASSISTANT, Universiteit Utrecht , Examinatrice
Monsieur Xander CAMPMAN
DOCTEUR EN SCIENCES, Shell, Examinateur
Monsieur Stéphane GARAMBOIS
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes, Président

Acknowledgements

I would like to thank all the people who, in one way or another, participated and supported me during my Ph.D training.

First of all, I would like to thank my supervisors Florent and Pierre for allowing me to work with them and for all the advice and guidance they gave me during these last three years. Second, I would also like to thank all the participants of the PACIFIC project (François, Phill, Sophie, Meysam, Yihe, Chris, Nick, Noelie, Rosemary, ... etc.). Thank you all for helping me in my first moments in academic life; thank you for your fruitful scientific discussions.

Likewise, I would like to thank the whole ”isterre team,” who are too many to name one by one but. Thank you for accompanying me in the day-to-day of these years, for the long talks, the coffees, and the beers when the days were hard to handle. To my colocs, thank you for being my second family, for putting up with me when I was not at my best. Without you, the challenging moments of these last years would not have been bearable. Thank you guys for all the nice adventures we had during this period of time !

Finally, but just as important, thank my family. Ustedes siempre han sido el motor que motivo a ir lo más lejos posible, a esforzarme, a superarme y a buscar ser la mejor version de mi. No es sorprendente pues aprendí de los mejores de ustedes, de mis padres, de mis tíos y tías y mi primo y hermano y hermana. Este logro es mío, pero no hubiera sido posible sin ustedes y para mí este logro se siente nuestro. Para los peques, espero poder ser también un ejemplo como ustedes lo fueron para mí, espero que esto les permita ver nuevos caminos y nuevas formas de ver y explorar el mundo. Gracias a todos por acompañarme, desde lejos por apoyarme. Los amo. Quisiera agradecer especialmente a Henri, que si dios no te da hijos el diablo te da sobrinos. Y menos mal, que haces parte de mi vida tío, porque gracias a ti he sobrepasado mis limites y he descubierto un mundo que no conocía, gracias por siempre motivarme a ir más lejos, ha hacerme más preguntas. Gracias por ser parte de mi vida.

Abstract

Seismologists eagerly seek new and preferably low-cost ways to map the complex structure of the top few kilometers of the crust. Passive seismic imaging appears as a novel, low-cost, and environmentally-friendly approach for exploring the sub-surface in the mining context. Usually, passive seismic interferometry relies on blind correlations within long time series of seismic noise or coda waves. In this thesis, we propose a complementary approach: seismic interferometry using opportune sources, specifically moving sources that are not stationary in time. This new approach relies on an accurate understanding of the seismic source's mechanism, a careful signal time-window and station pairs selection, and seismic phase identification (surface and body waves).

Massive freight trains were only recently recognized as a persistent, powerful cultural (human activity-caused) seismic source. For example, one train passage may generate a tremor with an energy output equivalent of a magnitude M1 earthquake and be detectable for up to 100 km from the tracks. Thus, these train signals can be considered an opportune seismic source for passive seismic interferometry because they are readily available, detectable, repeatable, and generate high-frequency broadband energy. To illustrate this novel method's potential, we show a case study in a mineral exploration context at the Marathon site, Ontario, Canada.

The Marathon dataset consists of 30 days of continuous seismic data recorded by a dense array of 1020 1-component geophones (nodes). First, the sources of ambient seismic noise, including train signals, are identified and characterized, and the source mechanisms of train signals are discussed. Second, we developed the theory of seismic interferometry applied to opportune sources and designed a novel workflow to process and analyze this kind of data. By doing so, body- and surface-wave propagating between pairs of stations carefully oriented are retrieved. Finally, surface waves arrivals are picked to generate a 3D seismic velocity model of the Marathon area near-surface. We discuss the pros and cons of the method compared to more standard approaches with the help of numerical modeling, specifically focusing on the potential for body-wave imaging and the retrieval of azimuthal anisotropy. We showed that by correlating train tremors, we retrieved high-frequency arrivals with higher SNR than using the standard method while using fewer data. Thus, the high-frequency surface waves allowed us to do near-surface imaging. Far from being restrained to near-surface imaging, this new way of analyzing opportune seismic sources can be applied in various contexts and scales using natural or man-generated signals.

Keywords: Seismic interferometry, near-surface imaging, Mineral exploration, train seismic signals, Marathon deposit.

Résumé

Les sismologues sont à la recherche de nouvelles méthodes, de préférence peu coûteuses, pour cartographier la structure complexe des quelques kilomètres supérieurs de la croûte. L'imagerie sismique passive apparaît comme une approche nouvelle, peu coûteuse et respectueuse de l'environnement pour explorer le sous-sol dans le contexte minier. Typiquement, l'interférométrie sismique passive repose sur des corrélations de longues séries temporelles de bruit sismique ou d'ondes coda. Dans cette thèse, nous proposons une approche complémentaire : l'interférométrie sismique utilisant des sources opportunes, plus précisément des sources mobiles qui ne sont pas stationnaires dans le temps. Cette nouvelle approche repose sur une compréhension précise du mécanisme de la source sismique, une sélection minutieuse de la fenêtre temporelle du signal et des paires de stations, et l'identification de la phase sismique (ondes de surface et de volume).

Les trains de marchandises massifs n'ont été reconnus que récemment comme une source sismique culturelle (d'origine humaine) puissante et persistante. Par exemple, le passage d'un train peut générer une énergie équivalente à un tremblement de terre de magnitude 1 et être détectable jusqu'à 100 km des voies. Ainsi, ces signaux de train peuvent être considérés comme une source sismique appropriée pour l'interférométrie sismique passive car ils sont facilement disponibles, détectables, répétables et génèrent une énergie large bande à haute fréquence. Pour illustrer le potentiel de cette nouvelle méthode, nous présentons une étude de cas dans un contexte d'exploration minière sur le site de Marathon en Ontario, Canada.

Le jeu de données de Marathon consiste en 30 jours de données sismiques continues enregistrées par un réseau dense de 1020 géophones à 1 composante. Tout d'abord, les sources de bruit sismique ambiant, y compris les signaux de train, sont identifiées et caractérisées, et les mécanismes de source des signaux de train sont discutés. Ensuite, nous avons développé la théorie de l'interférométrie sismique appliquée aux sources ponctuelles et conçu un nouveau workflow pour traiter et analyser ce type de données. Grace à ce traitement, les ondes de volume et de surface se propageant entre des paires de stations soigneusement orientées sont récupérées.

Enfin, les ondes de surfaces sont sélectionnées pour générer un modèle de vitesse sismique 3D de la zone proche de la surface de Marathon. Nous discutons des avantages et des inconvénients de la méthode par rapport aux approches plus standard en utilisant la modélisation numérique, et en nous concentrant spécifiquement sur le potentiel d'imagerie des ondes de volume et de récupération de l'anisotropie azimutale. Nous avons montré qu'en corrélant les tremblements de train, nous avons récupéré les arrivées à haute fréquence avec un rapport signal/bruit plus élevé qu'en utilisant la méthode standard tout en utilisant moins de données. Ainsi, les ondes de surface à haute fréquence nous ont permis de faire de l'imagerie proche de la surface. Loin d'être limitée à l'imagerie proche de la surface, cette nouvelle méthode d'analyse des sources sismiques opportunes peut être appliquée dans une variété de contextes et d'échelles en utilisant des signaux naturels et artificiels.

Mots-clés: Interférométrie sismique, imagerie proche de la surface, signaux sismiques de train, exploration minière, gisement de Marathon.

Contents

Introduction

Passive seismic interferometry

In seismology, standard imaging methods rely on recording earthquakes or man-made controlled sources to explore the Earth's interior structure, composition, and dynamics. However, depending on the location or the size of the imaged target, artificial sources may be too expensive or too cumbersome to use and often lack the low-frequency content needed to reach deep structures. On the other hand, earthquakes happen only in specific regions and at random times, which is inconvenient for imaging tectonically stable areas, especially when using temporary seismic networks.

The alternative to using complex fields such as diffuse fields and/or seismic noise to extract information of the Earth structures was suggested in earlier studies (Aki, 1957; Claerbout, 1968). Aki (1957) suggested a theoretical approach of spectral correlation in a 2D homogeneous medium. He proposed the spatial auto-correlation theory (SPAC) currently used to measure dispersive properties of surface waves. Claerbout (1968) proposed that the cross-correlation of two receivers at the surface is equivalent to a reflection response at one receiver if there was a source at the other receiver.

The laboratory experiments in acoustic medium done by Fink (1997) propagates signals from a source, records it, revers, and then re-emits in the medium; the re-emitted field is perfectly focused and can be considered as a virtual source. These experiments made it possible to draw an equivalence between time-reversal experiments and the complex fields in seismology by a mathematical analogy between correlation and convolution by a time-reversed signal. Assuming that the signals recorded between two stations are uncorrelated enough, seismic interferometry can extract the coherent signals corresponding to the deterministic information between the two stations, thus, if we have one source S and two receivers A and B. Then, the correlation $C(x_A, x_B)$ is equivalent to having a source at x_A, recorded in S where it was time-reversed and re-emitted to be recorded in x_B.

Defining the Green's function as the impulse response of the medium, in other words, is the waveform that would be recorded at point B if an impulse source were generated at point A. Further theoretical studies proved that the Greens function could be retrieved from the average cross-correlation of a diffuse acoustic noise field in a close medium or with a homogeneous distribution of sources (Weaver and Lobkis, 2001; Roux et al., 2004). Later, Derode et al. (2003) retrieved the Green's function of a heterogeneous open medium by cross-correlating a scattered wavefield based on time-reversal symmetry. Wapenaar (2004) provided a formal demonstration of the Green's function retrieved by cross-correlation based on reciprocity in the elastic case. Snieder (2004) showed the convergence of the noise correlation function to the Green's function for the surface wave, assuming an isotropic distribution of scattered waves aligned with the stations' azimuth. This concept, also called stationary phase zone, can be defined as the regions in which a source disturbance will cause constructive interference, resulting in the reconstruction of one or more phases. The stationary phase zone will be discussed later. In general, this approach can be described by the following equation Gouedard et al. (2008):

$$\frac{d}{dt}C_{A,B}(\tau) \simeq G_{A,B}(\tau) - G_{A,B}(-\tau) \tag{1}$$

The equation 1 expresses that the derivation of the cross-correlation $C_{A,B}$, between A and B converges to the Green's function of the medium between these two points $G_{A,B}(\tau)$ (also called causal Green's function) and the opposite of the Green's function between B and A: $G_{B,A}(-\tau) = G_{A,B}(-\tau)$ (the anti-causal Green's Function). In theory, these two Green's functions should be identical according to the principle of reciprocity.

In the Earth, the background seismic vibrations can be considered as diffuse wavefields. It has been also shown empirically (e.g., Campillo and Paul, 2003; Shapiro and Campillo, 2004; Wapenaar, 2004; Sabra et al., 2005; Roux et al., 2005a) that the cross-correlation $C_{A,B}$ of seismic noise at two points A and B converges to the Green's function. This approach, also called passive seismic interferometry, is sketched in an explanatory diagram (figure 1). The schematic shows two stations, A and B (red circles), recording seismic noise (black traces); this noise is generated by different sources (black wiggle arrows) at different locations and at different frequencies. These noise signals' correlation ($\otimes$) results in a surface wave (red trace) propagating between both stations. Positive times (or causal part) correspond to the wave traveling through the medium from A to B, and negative times correspond to the propagation from B to A. The difference between the causal and acausal parts is caused by the characteristics of seismic noise

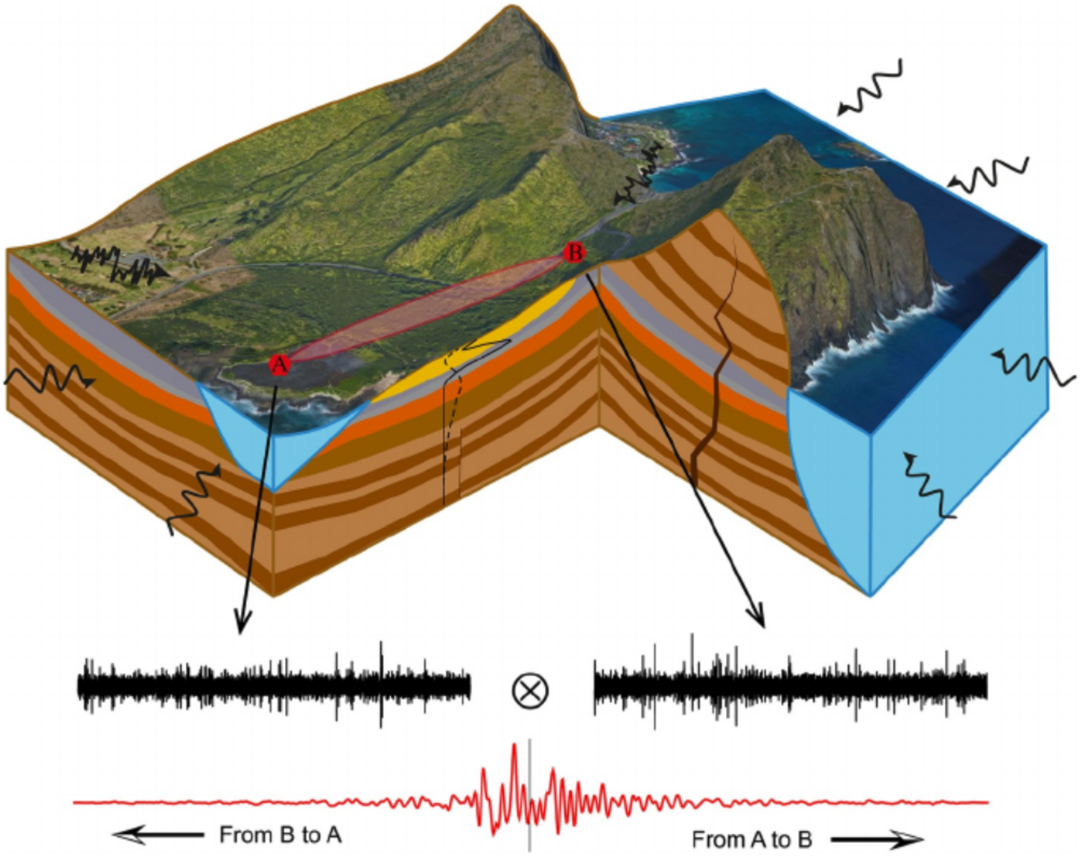

Figure 1: Schematic cross correlation-based Green's function retrieval. The correlation of the seismic noise recorded at two stations, A and B (black traces), results in a seismogram representing the surface wave propagating between the two stations (red trace; A to B corresponds to the positive times, B to A corresponds to the negative times). For this particular pair of stations, this retrieved Green function is laterally sensitive to the area indicated by the red ellipse. By cross-correlating different pairs of stations, different areas of the study site will be covered. The black curves schematically display the depth sensitivities of these surface waves in the cross-section. The solid curve illustrates a typical short-period sensitivity; the dashed curve represents the sensitivity curve for a longer period. The incoming seismic noise (black-wiggled arrows) is generated, for instance, by the ocean and along the coast for the longer periods (above 1 s), and in urban areas, along roads (traffic noise) and trails for the shorter periods (below 1 s). Modified from Grobbe et al. (2021).

that we will discuss below.

We can define seismic noise as the Earth's permanent and weak background vibrations in the absence of earthquakes and active sources. Any transient forcing applied on the Earth's surface sufficiently energetic to produce seismic waves is considered seismic noise. These vibrations were initially considered to be undesirable (noise) and of no use. However, with seismic instruments' technical progress, the theoretical study of this so-called seismic noise and its use by interferometry has increased steadily over the last twenty years. The seismic noise is originated by either natural or anthropogenic seismic sources characterized by their location and frequency content.

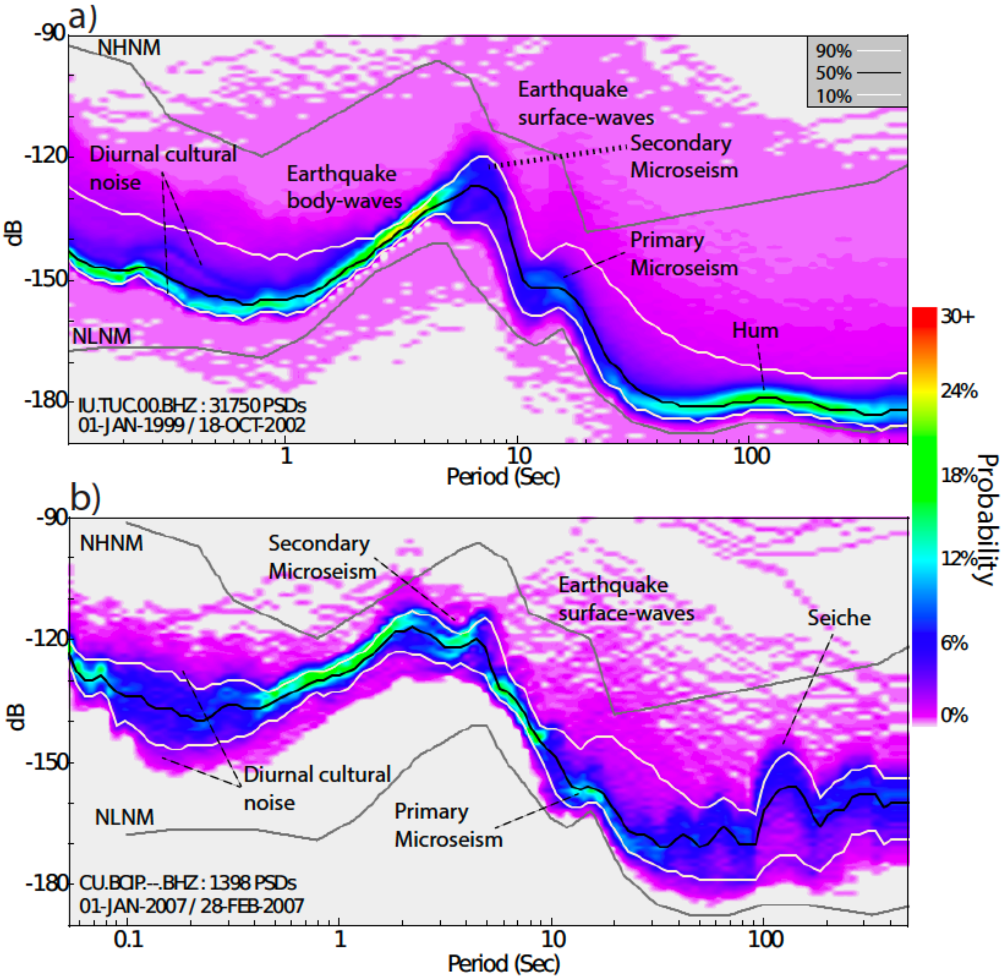

Figure 2: Major components of the seismic ambient noise spectrum. Red to green colors indicate the highest probability seismic ambient noise power levels at each period. Baseline statistics shown as solid lines (10th% lower white line, 50th% black line and 90th% upper white line). New low noise model (NLNM) and new high noise model (NHNM) shown as gray lines Peterson et al. (1993). a) PSDPDF computed using data from a GSN station in Tucson, Arizona (IU.TUC.00.BHZ, 31,750 PSDs, 01-JAN-99-18-OCT-02). b) PSDPDF computed using data from a GSN station on Isla Barro Colorado, Panama (CU.BCIP.00.BHZ, 1,398 PSDs, 01-JAN-07-28-FEB-07). Modified from McNamara and Boaz (2019).

At long-period (20s), the signals are generated by ocean swell generated by storm-forced shoreward directed winds that reach coastlines; we referred to this mechanism as ”Hum” (e.g., Rhie and Romanowicz, 2004; Webb, 2007; Ardhuin et al., 2015). At intermediate periods (1 - 30 s), seismic noise is generated by ocean waves explained by two mechanisms. The primary microseism is generated by the interference of ocean waves with shallow ocean bottom topography, and the secondary microseismic is generated by the interference of pairs of ocean wave trains (e.g., Longuet-Higgins, 1950; Kedar et al., 2008). At low periods, the seismic noise is generally linked to human activities such as traffic, machinery energy, etc.; this is commonly called cultural noise. The sources of cultural noise are mainly located at the surface of the Earth, and present temporal variations related with in human activity (Bonnefoy-Claudet et al., 2006). The sources that can generate cultural noise are extremely diverse. Generally, these sources are recorded a few kilometers from the station because they have a low forcing compared with the mechanisms at long periods (hum or microseismic noise); in addition, its high-frequency content attenuates rapidly with distance. Different studies have been conducted to understand and characterize theses sources as traffic (Riahi and Gerstoft, 2015; Lavoué et al., 2020), massive events as concerts or sportive games (Boese et al., 2015; Diaz et al., 2017; Green et al., 2017), storms (Fan et al., 2019; Retailleau and Gualtieri, 2019) and wind (Johnson et al., 2019), among other sources.

Figure 2 shows the Probability Density Function of Power Spectral Density (PSDPDF) of ambient noise recorded by two broadband stations located at different positions (details in the figure caption). We distinguish several physical mechanisms coming from several different sources characterized by their frequency content. Other than the signal generated by earthquakes, we see that human activity generates seismic noise at low periods. In an intermediate period, the secondary microseismic pick is observable as well, and at long periods the records are dominated by the noise generated by the hum.

Seismic noise, although continuous in time, is dependent on the locations and frequency of the seismic sources. Therefore, the theoretical assumptions necessary to retrieve Green's function are no longer fully respected, causing different effects. One effect of having a non-homogeneous distribution of sources is the asymmetry of noise correlations functions. In a perfect case, the sources are uniformly distributed (figure 3-a), the energy between two stations is the same in both directions, and the causal (positive times) and anti-causal (negative times) Green's functions are perfectly symmetric.

When the sources are anisotropically distributed (figure 3-b and c), and without multiple scattering, the cross-correlation is not symmetric anymore. The different amplitude of the causal or anti-causal part depends on the locations of the most energetic source. For example, if the sources are more energetic on side 1 of the studied medium, as shown in figure 3-b, the causal part (from station 1 to station 2) will present a greater amplitude than the anti-causal part (Stehly et al., 2006).

Besides having an isotropic distribution of sources, sources can be distributed in the stationary phase zone to fully retrieved the Green's Function. The stationary phase condition implies that the correlation function's convergence towards the Green function requires the presence of sources (or scatters) in line with two carefully placed receivers.

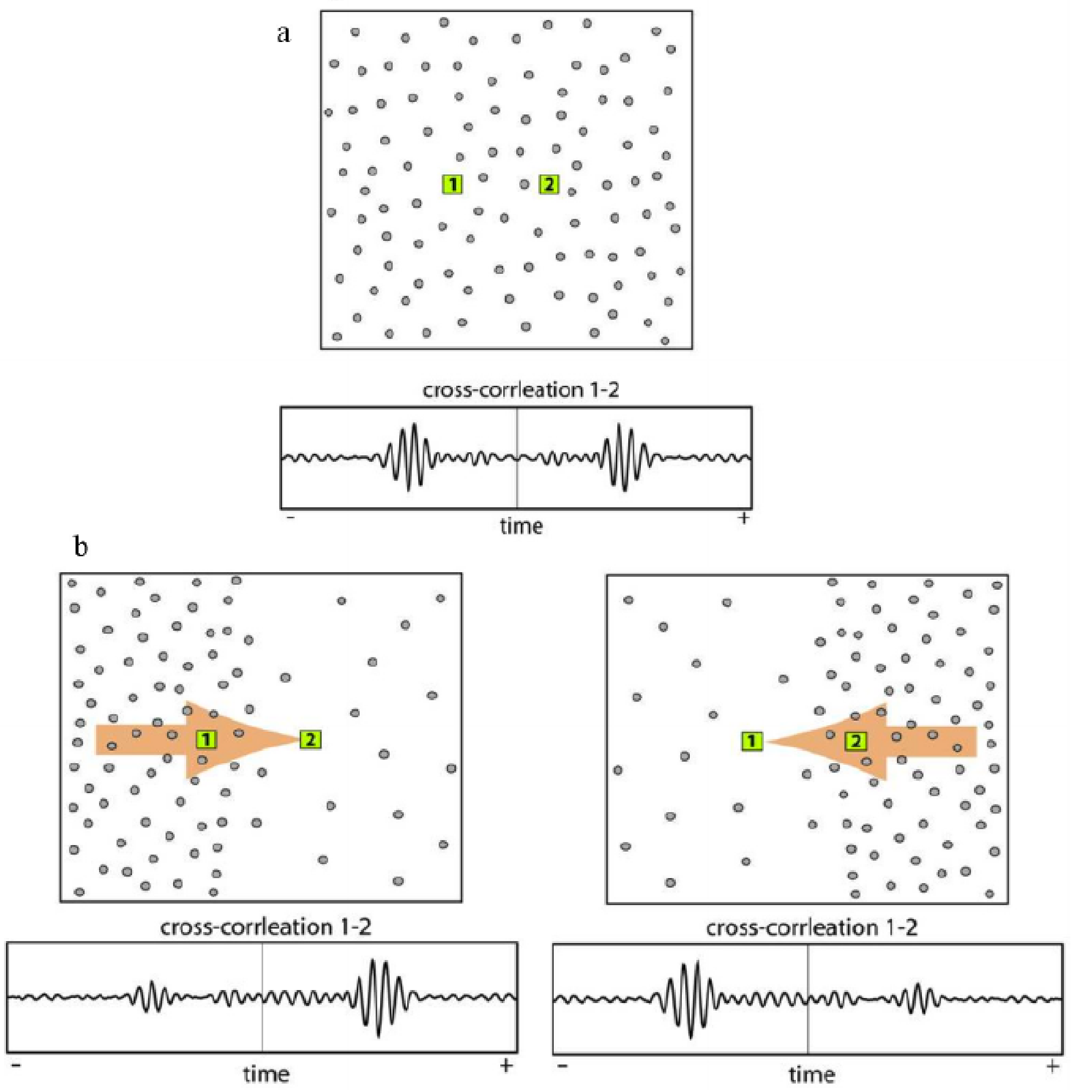

Figure 3: Schematic illustration of the effect of inhomogeneous noise sources distribution on the degree of symmetry of cross correlation. (a) Symmetric cross correlation between 1 and 2 obtained when the sources of noise are evenly distributed. (b) Asymmetric cross correlation (but symmetric travel times) associated with a nonisotropic distribution of sources. Modified from Stehly et al. (2006).

Each of the possible phases (or wave types) included in the cross-correlation function could be sensitive to various sources regions. The main contributors to a particular phase are sources within its stationary phase area. Therefore, we can measure a specific phase between two receivers by correlating a source within its stationary phase zone, including the surface. For example for P waves emerge from the interference between a direct P recorded at the first station and a PP (redirected once by a buried layer or formation edge) recorded by a second station after a rebound below the first one. One can do the same with S waves.

Constructive summation interference occurs if the seismic sources satisfy the stationary phase criterion:

$$\Delta t = t_{pp} - t_p \leq t_{green} \pm \frac{T}{4} \tag{2}$$

Where t_{pp} is the arrival time of the PP wave at the second receiver; t_p is the arrival time of the P wave at the first station t_{green} is the arrival time of the P wave between the two receivers, and T is the dominant period. This can be generalized for phases (e.g., Colombi et al., 2014; Roux et al., 2005b) and all cross-terms between two phases that share same ray parameter (slowness) (e.g., Pham et al., 2018).

In a 2D homogeneous medium, these stationary points define an hyperbolic area outward from the receiver pair, with a frequency-dependent aperture, the lower the frequency, the broader the calculated source region (e.g., Roux et al., 2005b; Snieder, 2004). In 3D and for both surface- and body-wave retrieval, the full Green's function retrieval demands sources that are evenly distributed along a more complicated shape enclosing the two sensors (e.g., Wapenaar, 2004; Wapenaar and Fokkema, 2006).

The assumptions of a diffuse wavefield or homogeneous distribution of noise sources are not generally verified when discussing seismic noise. Several studies showed that when Green's function retrieval is broken, several effects can be created. Phase and amplitude changes could be generated (e.g., Kimman and Trampert, 2010; Seats et al., 2012) and spurious arrivals can be present in the cross-correlation function (e.g., Halliday et al., 2008; Snieder et al., 2006). It was also showed that travel times measured on the seismic noise correlations can be affected by the distribution of noise sources (Tsai, 2009; Froment et al., 2010).

Body-wave retrieval is another critical problem. Body waves arrivals can be weak or absent in the correlation functions due to the heterogeneous distribution of high-frequency sources (e.g., Forghani and Snieder, 2010). The stationary phase zones are smaller for body waves than for surface waves (Forghani and Snieder, 2010) and a proper reconstruction of the body-wave requires a full 3D distribution (i.e., at surface and depth). This configuration is almost unrealistic. Besides the lack of availability of high-frequency sources, their constraint location at the Earth's surface, their limited radiation, and the poor understanding of these sources' excitation mechanisms make that the seismic passive seismic interferometry commonly use only the surface waves.

Sager et al. (2018, 2021) showed that the cross-correlation function energy and waveform are sensitive to both the structure between the sensors and the structure between the source and the sensors for surface and body waves. For example, figure 4 shows the structure sensitivity for two different source distribution cases. The structure sensitivity can be seen as the medium's regions through which virtual waves are traveling; these regions can be located between the sensors or between the sensor and the sources. Thus, it can be seen that source distribution strongly affects the sensitivity between the sensors and the source. For example, case 2 (figure 4-2) is less sensitive to the regions between the source and the sensor, and consequently, the correlation function is closer in phase and amplitude to the Green's function between the two sensors without however converging entirely to the Green's function.

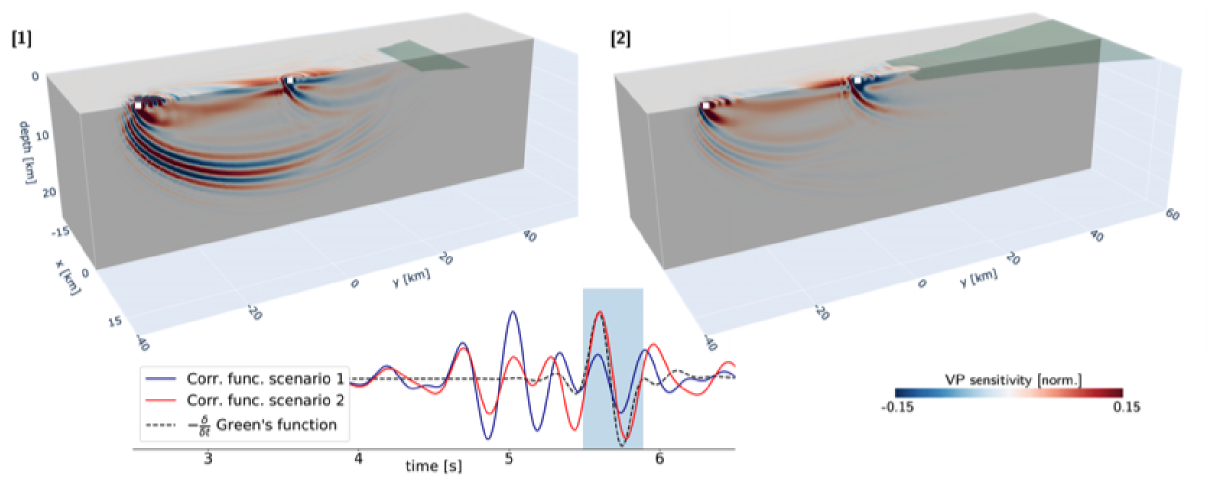

Figure 4: Structure sensitivity kernels for PP-P waves velocity (top panel). We define two source configurations (dark gray regions at the surface of each domain): 1) a small source covering the center region of the stationary zone and 2) a large source covering the main region with positive contributions with a minimum distance of 10km to the right station. The resulting correlation functions, band-pass filtered between 2 and 4Hz (bottom left), are windowed around the PP-P interaction. A Gaussian source time function with a half-width of 5 times the simulation time step was used to generate wavefield. For visualization purposes, the box does not outline the full computational domain. Modified from Sager et al. (2021).

Since the seismic noise is not equipartitioned, the noise sources are not homogeneously distributed, the correlation function obtained by passive seismic interferometry is only a Green's function approximation. Different approaches have been developed in order to converge as much as possible to Green's function. Some classical signal processing approaches like One-bit normalization or whitening have been developed (Bensen et al., 2007), and alternative approaches like interferometry by deconvolution (e.g., Snieder and Safak, 2006; Vasconcelos and Snieder, 2008a,b), or interferometry by multi-dimensional deconvolution (e.g., Wapenaar et al., 2008; Wapenaar and van der Neut, 2010; Wapenaar et al., 2011) among others.

Despite the difference between the green function and the correlation function, numerous studies have shown that approximate Green's functions, or cross-correlation functions, can be useful to extract information about the medium. Campillo and Paul (2003) and Shapiro and Campillo (2004) showed first that the cross-correlation of coda waves or seismic noise records results on the approximate surface wave part of the Green's functions with similar characteristics to actual earthquake signals. Since then, this technique has been applied in a wide variety of contexts to monitoring and imaging applications.

Passive seismic interferometry has been applied to monitor temporal changes in the Earth. Velocity changes precursory to volcanic eruptions (e.g., Brenguier et al., 2008; Mordret et al., 2010) or related to earthquakes or slow-slip event (e.g., Wegler and Sens-Schönfelder, 2007; Froment et al., 2013; Rivet et al., 2011; Chen et al., 2010) have been observed using seismic noise. Seasonal changes due to thermo-elastic effect (e.g., Meier et al., 2010) have been also observe among other monitoring applications.

Seismic noise correlation has also numerous applications in Earth imaging at different scales. Surface-wave tomography has been applied successfully at the crustal and lithospheric scales (e.g., Sabra et al., 2005; Kang and Shin, 2006; Shapiro et al., 2005; Lin et al., 2008; Ritzwoller et al., 2011a; Nishida et al., 2008; Verbeke et al., 2012) at near subsurface or smaller structures (e.g., Wapenaar, 2004; Picozzi et al., 2009; Behm and Snieder, 2013; Brenguier et al., 2007) and landslides (e.g., Renalier et al., 2010). The anisotropy has also been studied using seismic noise (e.g., Huang et al., 2010; Ritzwoller et al., 2011b). Recent studies have also unveiled the possibility of reconstructing body-waves at global (e.g., Poli et al., 2012; Nishida, 2013; Boué et al., 2013) and local scales (Draganov et al., 2009; Nakata et al., 2015; Olivier et al., 2015; Nakata, 2016).

Passive seismic imaging in industrial contexts

In recent years, with the emergence of dense industrial networks (nodes or ocean bottom cable), passive seismic interferometry has crossed academic barriers and has been applied in numerous industrial applications. As a result, this approach emerged as a novel, low-cost, and environmentally friendly method for exploring the sub-surface, providing access to the most remote and environmentally sensitive locations on Earth.

Numerous studies have succeeded in imaging the near surface using arrivals retrieved from seismic noise including oil and gas industry (e.g., Draganov et al., 2009; Mordret et al., 2013a; Lin et al., 2013b; de Ridder and Dellinger, 2011; Chmiel et al., 2019; Spica et al., 2018) and geothermal industry (e.g., Penumadu and Park, 2005; Jousset et al., 2016). Several noise-based monitoring applications successfully applied seismic interferometry in different fields (e.g., De Ridder and Biondi, 2013; Arts et al., 2013; Obermann et al., 2015; Hillers et al., 2015; Zhang et al., 2016)

Mineral exploration uses geophysical surveys (electric, magnetic, gravimetric, etc.), geological models, and diamond drilling combined to identify targets for further investigation. However, active seismic methods are rarely employed in the mineral industry because they are too expensive and require an active source, which can be challenging to use in remote or environmentally sensitive areas. Passive seismology is an economically and environmentally interesting alternative that has been applied in different mining contexts(e.g., King and Luo, 2009; Roots et al., 2017; Dales et al., 2017; Olivier et al., 2015; Polychronopoulou et al., 2020; Chamarczuk et al., 2021). However, passive surveys cannot always reach the resolution required to image the ore-bodies explored by the mining industry because most passive seismology applications retrieved low-frequency surface waves. High-frequency body waves remain more useful for mining contexts because of their sharp sensitivity to seismic velocity perturbations at depth. Unfortunately, these arrivals also prove difficult to extract from passive data sets due to the lack of high-frequency body wave sources. Despite the difficulties to reconstruct body-waves from seismic noise, several studies target body-waves retrieving for imaging applications (e.g., Ruigrok et al., 2011; Vidal et al., 2014; Nakata, 2016; Retailleau et al., 2020) and monitoring applications (Brenguier et al., 2019). Here, we explore seismic noise sources suitable for passive seismic interferometry in a mineral exploration context to aim for body-wave retrieval and high-resolution surface-wave imaging.

Outline of the thesis

Passive interferometry relies on the hypothesis of homogeneous distribution of noise sources to reconstruct Green's function; however, this hypothesis is never verified in practice. Therefore, different data processing methods have been developed in seismology to get as close as possible to Green's function. This thesis is a paradigm change and proposes instead to focus on using appropriate seismic noise sources for passive interferometry. For this, we will develop the method in the Marathon deposit case study (Ontario, Canada). We will apply our alternative method to retrieve high-frequency energy for imaging the shallow crust using train-generated seismic noise in this mineral context. This thesis manuscript is divided into three main parts.

In the first part, Chapters 1 and 2, the area studied is presented. First, we present the geological and geophysical overview of the region. Then, we will present the dataset, and the seismic noise records will be analyzed, describing the characteristics of seismic noise in the region, especially the train signals. Finally, these signals will be presented by discussing the seismic observation in the Marathon region and numerical modeling.

The second part focuses on the methodological framework. The strategy of processing the seismic signal of the train by interferometry is proposed. Finally, a detailed application to the Marathon case is presented, focusing on high-frequency recovery.

The third part is devoted to the imaging application. We will use the arrivals recovered in the second part to apply standard surface wave imaging methods to provide a high-resolution 3D shear wave velocity model of Marathon. Finally, we will discuss the pros and cons of our methods by comparing the results with standard cross-correlation methods and numerical modeling.

Chapter 1

PACIFIC project and Marathon deposit site

Contents

This chapter introduces the PACIFIC project, and my contribution to the project. The PACIFIC project seeks to develop new techniques in passive seismology applied to shallow surface imaging. Its principal scientific motivations and the project context will be presented later. Then, we will present the Marathon deposit pilot test where this thesis was conducted. The site is a mining exploration block in Marathon, Ontario, Canada, with potential targets of interest of high concentrations of Cu-PGM (platinum group metals) hosted in a gabbro intrusion. A more detailed geological description and the previous geophysical studies of the regions will be shown. Finally, The dataset at Marathon will be presented.

1.1 PACIFIC project

1.1.1 PACIFIC project overview

PACIFIC (Passive seismic techniques for environmentally friendly and cost-efficient mineral exploration) is a research project of which my Ph.D. is part. It received €3.2 million worth of funding from the European Union to develop passive imaging techniques in mineral exploration. The project was launched in June 2018, and It is set to run for 42 months. To find information, follow this link `https://www.pacific-h2020.eu`.

A new world-class ore deposit must be discovered every two years to satisfy the global metals and mineral demand. However, this task is becoming increasingly difficult because easily accessible deposits have been exhausted. As a result, mineral exploration is taking place at greater depths and in remote locations (ARTTIC, 2018).

PACIFIC consortium will conduct fundamental and applied passive seismic imagery research, which will have the accuracy and resolution needed for the minerals industry while at the same time having a relatively low environmental impact (ARTTIC, 2018). The passive seismic technique uses arrays of seismic receivers to record ambient seismic noise and retrieve surface waves used for imagery. These waves attenuate rapidly with depth, and the method currently lacks the resolution and depth penetration needed to image ore deposits directly. Neverless, it provides background information and can acquire a broad-brush image of regions under exploration. But to be of more practical application in the mineral industry, the method must be able to image small objects, with high resolution, at depths of several hundred meters. Two mineral test sites will be studied to develop these techniques, one in Canada and the other in Sweden.

In PACIFIC project, academia, geological survey, and industry will closely collaborate to achieve these objectives. The academic partners are Université de Grenoble Alpes (UGA) and Dublin Institute for Advanced Studies (DIAS), Geological Survey Ireland (GSI). The industrial panthers are Sisprobe (a company in the passive seismic technique), STILLWATER Canada INC, a global metal mining group (SIBSTIL), and Beowulf Mining Plc (BEOW) exploration and development company. The link to an explicative video about the project `https://www.youtube.com/watch?v=7uUl9Wh86gQ`.

1.1.2 My role in the project

My Ph.D. took place in this project, investigating the first test site at Marathon Canada. The first goal of my thesis was to develop a processing method to extract high-frequency arrivals from seismic noise in a mineral context and then use them to image the region and contribute to improving the geological understanding of the area. To do so, I worked jointly with academic partners of the project in UGA and DIAS and with industry partners as Sisprobe and Stillwater. The material presented in this script concert my work and the work resulting from these collaborations.

1.2 Marathon mineral deposit - Geological overview

Marathon mineral deposit, one of the two PACIFIC pilot test sites, is studied in this thesis. It is located 10km north of the town of Marathon, Ontario, Canada, along the north shore of Lake Superior (Figure 1.1-a). The site was chosen for several reasons: the close relation with STILLWATER company, the previous tests of the surface-wave passive seismic method, the geology and mineralization are all correctly known. In the following section, Marathon's geological and geophysical background will be exposed.

Figure 1.1: Marathon test site a) Map of North America. Marathon site (red square) is located north of Superior Lake near Marathon town, Ontario, Canada. b) Aerial photo of Marathon deposit c) Ground photo of the studied region where Marathon environmental conditions are exposed (forest region).

The Marathon test site is located 10km north of Marathon town, Ontario, Canada, along the north shore of Lake Superior. Marathon deposit is a platinum group metals (PGM) exploration block owned by Stillwater Canada Inc. The deposit contains platinum group metals (PGM) hosted in gabbro intrusion, which Stillwater Canada Inc owns. The exploration block is part of the Coldwell Alkaline Complex.

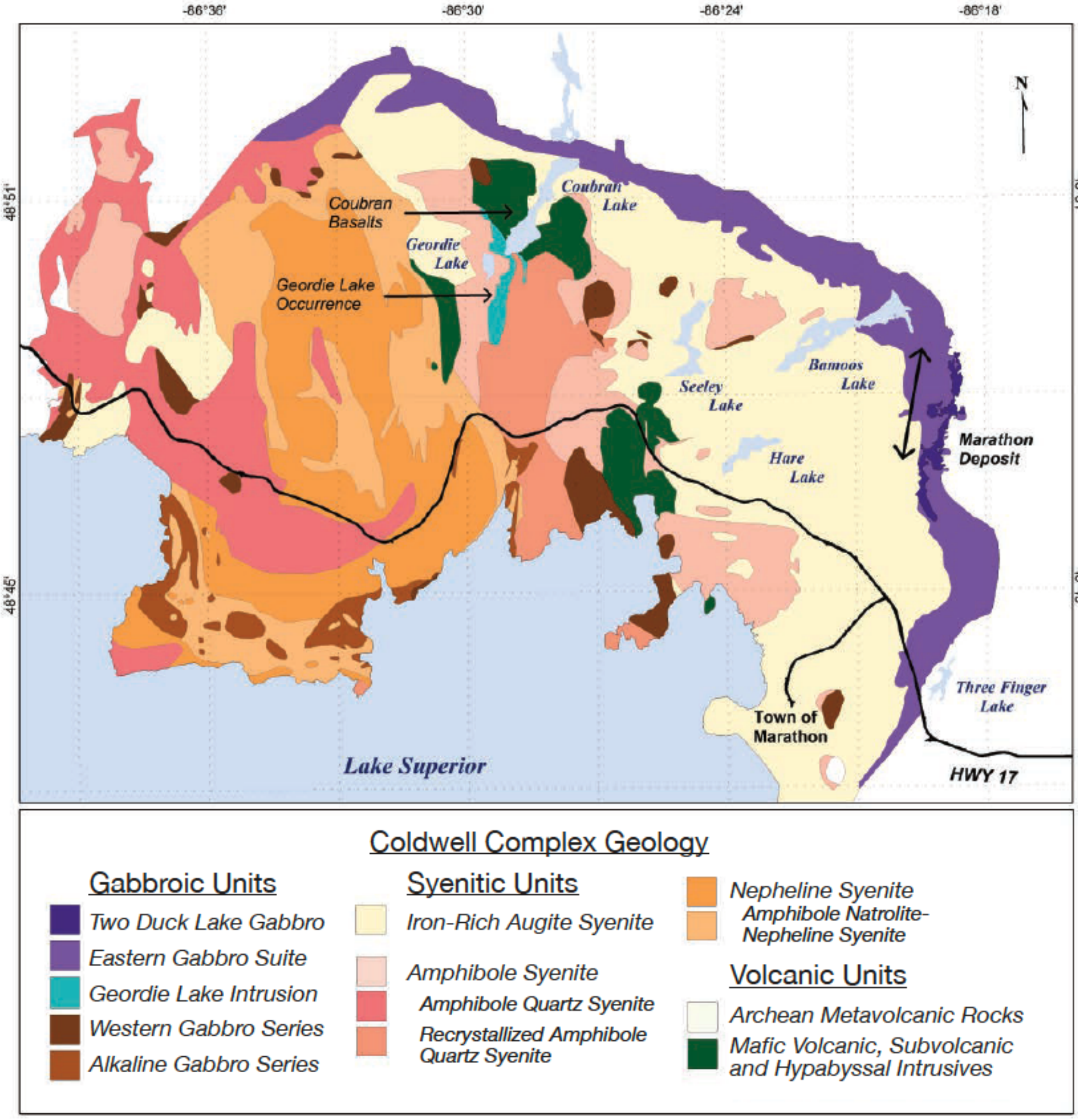

Figure 1.2: Geology map of the Coldwell Alkaline Complex by the staff of Marathon PGM Corp. and Stillwater Canada Inc. The geological units are differentiated by colors, as explained at the bottom of the image. Modified from Good et al. (2015).

The Coldwell Complex is part of the Keweenawan igneous rocks that were emplaced around and in the vicinity of the Great Lakes segment of the Midcontinent Rift System (1108-1105Ma) (Davidson Jr, 1982; Klasner et al., 1982; Weiblen, 1982; Good et al., 2015). Figure 1.2 shows the geology of the Coldwell Complex, including complex lithology of syenite, gabbroic, and volcanic units. Marathon deposit is hosted by the Two Duck Lake intrusion, located at the eastern margin of the Port Coldwell alkalic complex (see figure 1.2).

The Two Duck Lake gabbro is the host rock for the Marathon mineralization deposit; it is undeformed, unmetamorphosed, except for local alteration. Mineralization is a 4-km-long and 250-m-thick unit of the Marathon Series. The Marathon Series is the latest of three magmatic series that make up the 1- to 2-km-thick Eastern Gabbro Suite, which wraps around the eastern and northern margin of the Coldwell Alkaline Complex (Good et al., 2015). The Two Duck Lake intrusion, as described in this study, is broadly composed of the olivine gabbro and ferrogabbro units it is approximately 1500m long with a maximum thickness of 130m (Dahl et al., 2001). Thus, the Marathon Cu-PGE sulfide deposit has been classified as a gabbro-associated contact-type deposit. The size and grade of the Marathon deposit resource are 97.4 million metric tons (Mt) at 0.27% Cu, 0.75 ppm Pd, 0.23 ppm Pt, and 0.09 ppm Au (Puritch et al., 2009).

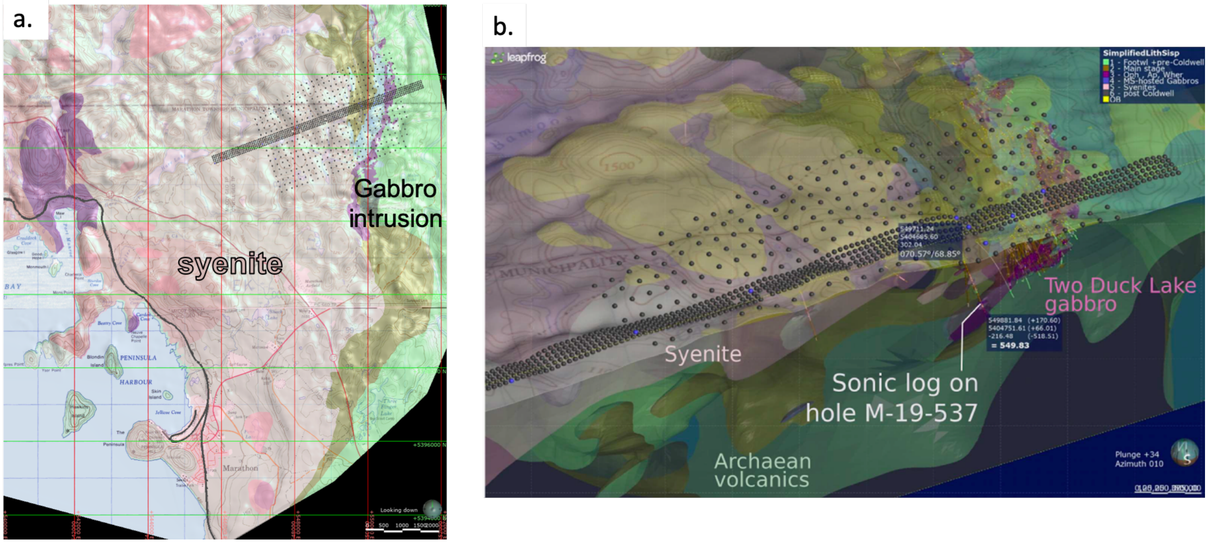

Figure 1.3: a) Geological map of Marathon site, two main units are shown: the syenite in pink and the gabbro (green, blue, yellow, and Violet). The colors represent different types of gabbro. b) Cross-section through the drill-hole-constrained 3D geological model for the Marathon deposit, showing the location of hole M-19-537 where a sonic log was measured (see details in the text). Most drilling extends less than 400m from the surface. Modified from Beard et al. (2021).

This region has an economic interest, given the geological composition. As shown, the region has been well explored using geological and geochemical methods. But, the gabbro geometry and, more generally, the deeper geology is not well constrained. Thus, drilling has been done, providing only 1D information of the first hundreds of meters. Figure 1.3 shows the zoom of the geological map in the Marathon site at the surface and the geological 3D background obtain from geochemical studies and the geological interpretations.

1.3 Geophysical studies

Marathon deposit has been studied for several years due to its economic importance. Although some of these studies are not public, in this work we will present two public studies carried out by Sisprobe and STILLWATER Canada INC with the intervention of The University of Western Ontario (UWO) and at Memorial University Newfoundland (MUN).

1.3.1 Sonic logs

The sonic studies measure the acoustic properties in the laboratory or in-situ using borehole sonic logs. This supports the ore seismic response calibration and a better understating of the geological features. At Marathon deposit, two measurements were made, one in-situ acoustic log and the other at the laboratory using surface samples and drillcore. Here, werecapitulate the log results presented by Beard et al. (2021).

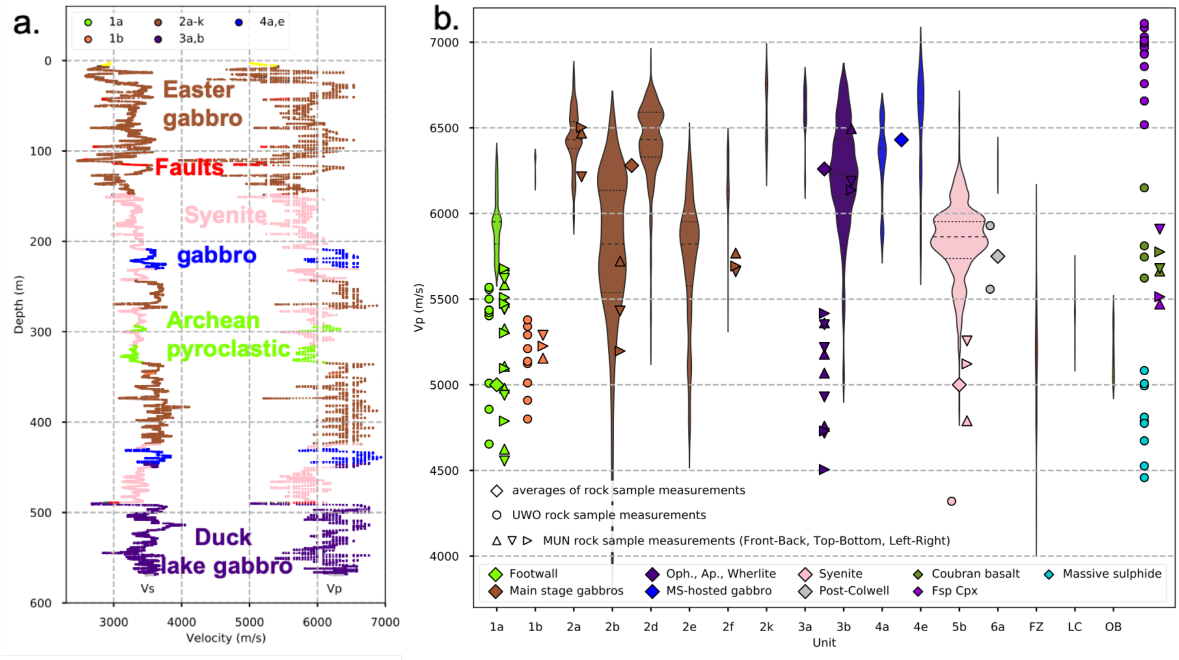

Figure 1.4: a) Sonic logs for hole M-19-537, Vp and Vs values at depth categorized with respect to lithology. b) Violin plots of the sonic log Vp measurements for each identified lithology, with median and interquartile ranges shown with dashed lines within the violins. Triangles and diamonds are velocity measurements made on surface samples and drill core at the University of Western Ontario (UWO) and at Memorial University Newfoundland (MUN). Modified from Beard et al. (2021).

Figure 1.4 summarize both in-situ and laboratory measurements. Panel a shows the Vp-velocity and Vs - velocity values at depth for one single hole in Marathon deposit (see figure 1.3). It can be observed that the different gabbro structures have similar velocities and the faults result in lower Vp velocities. Laboratory Vp measures are shown in figure 1.4-b. It can be seen that the velocity of several units overlaps for two approaches, showing a poor contrast in the region. Hence, the seismic characterization will be more challenging. Therefore, due to the similar velocity of all gabbro units, seismic methods cannot directly image the mineralization but rather the gabbro intrusion.

1.3.2 Passive seismic test

A first noise test at the Marathon site took place in 2017 by Sisprobe to characterize ambient seismic noise in the region. The results shown here are modified from Chmiel et al. (2017) report. 31 vertical component stations were deployed in Marathon for one month from 11 September 2017 to 09 October 2017 (Figure 1.5-a).

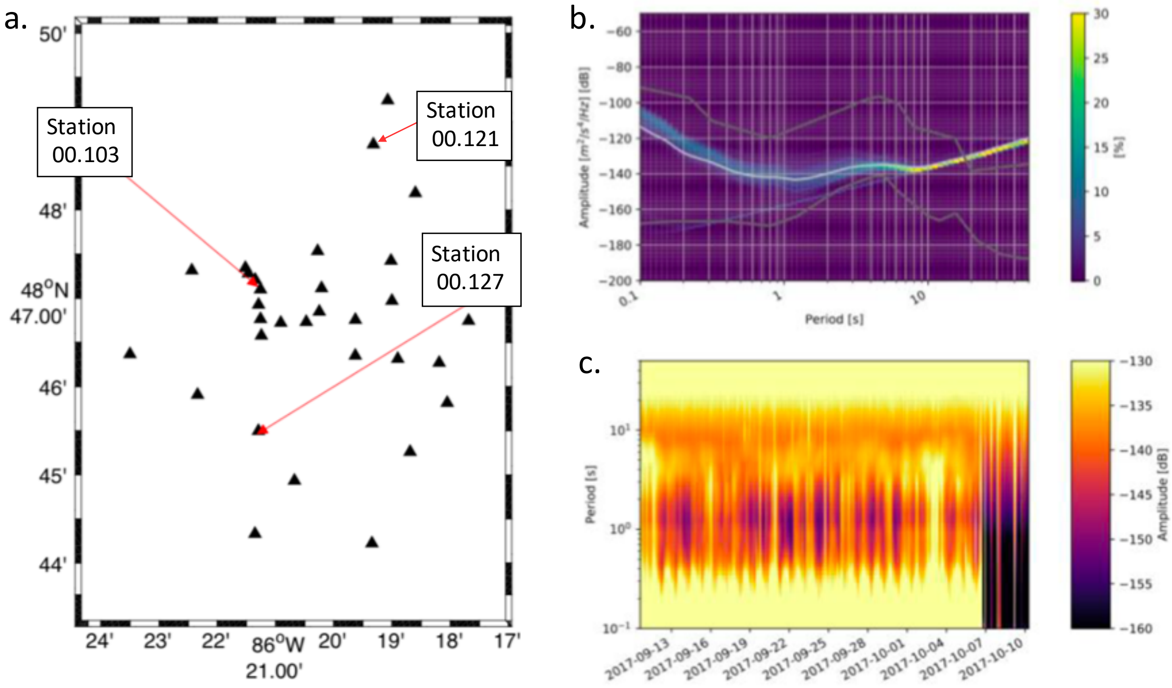

Figure 1.5: a) Marathon (Canada) noise test deployment. b) Probabilistic Power Spectral Density Gray lines represent the standard new low noisemodel (NLNM) (Peterson et al., 1993). c) Spectrogram for station 00.127 from 2017-09-11 to 2017-09-10 using one hour segment without overlap . Modified from Chmiel et al. (2017).

They conclude that the noise level was relatively high, and the quality of the data is good. Figure 1.5-b shows the Probabilistic Power Spectral Density and the spectrogram of the station

1.5-c indicating a high level of high-frequency noise. The pseudospectral density functions show that the microseismic noise energy is correctly recorded up to a period of 10 seconds; after this value, the noise is generated by the instrument and is not of seismic origin.

Then beamforming analysis was performed on daily records. Figure 1.6 shows a beamforming result averaged over the month. High-frequency bands show noise coming from Lake Superior (South-West) and low-frequency noise coming from North Atlantic (North-East). These sources are stable over time. The Rayleigh wave is weakly dispersive in the area. Therefore, they only observed the fundamental mode of the Rayleigh waves.

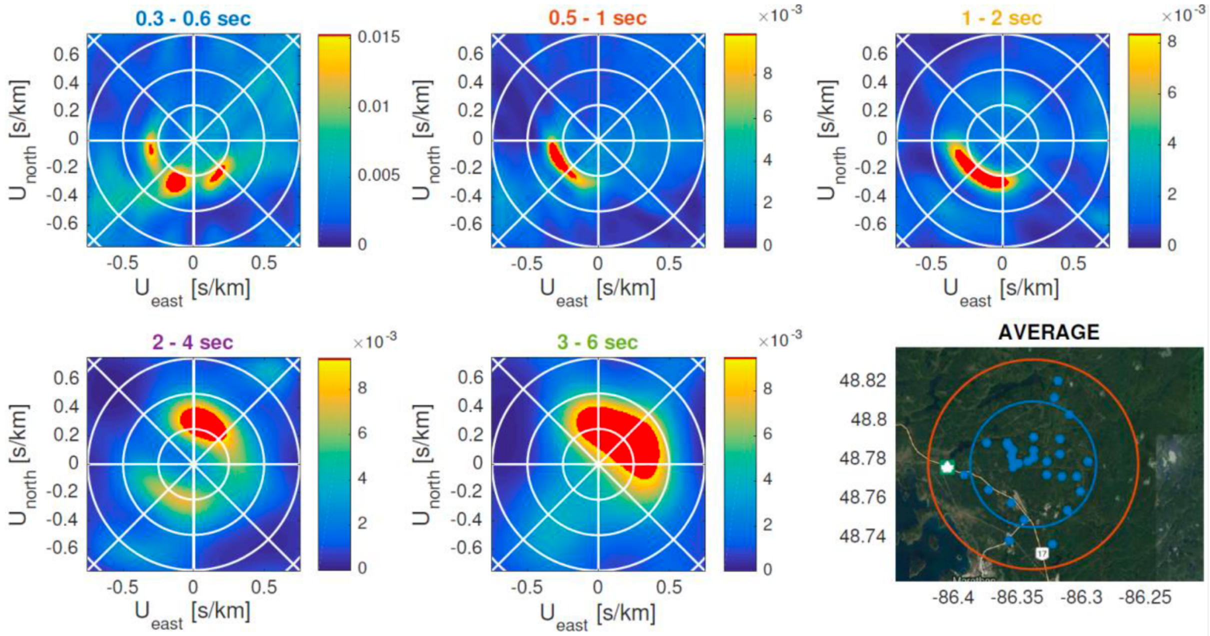

Figure 1.6: Average noise beamforming analysis for the entire duration of the survey. Five periods bands are studied. Warm colors represent high power amplitudes. Blue dots on the map (lower right panel) show the stations used. Modified from Chmiel et al. (2017).

Correlation functions are computed from daily ambient noise records. Signals are first downsampled to 50Hz and normalized in the frequency domain. Figure 1.7 b and c show correlations between a reference station (00.121) versus other stations sorted by increasing distance and filtered in two different period-bands ([1-10]s and [0.05-1]s respectively). The propagation of Rayleigh waves is visible across the array, even for important offsets. As observed from the beamforming, the strong directionality of the original ambient noise results in an asymmetry of the reconstructed wave-front and the cross-correlations show good SNR.

Distance-binned gather was constructed by averaging seismic section by bins of fixed distance intervals every 200m (Figure 1.7-a). The Rayleigh wave data shows fundamental mode, which is weakly dispersive. This test showed that the seismic noise was suitable for passive imagery.

Noise cross-correlations showed a good signal-to-noise ratio and propagation of the Rayleigh wave fundamental mode with weak dispersion. This study provides a promising overview of the noise characteristics (correct distribution and relatively high-frequency); They used these seismic noise records to apply seismic interferometry, allowing them to retrieve surface waves. However, to explore the gabbro intrusion, a more dense array is needed it. In the frame of PAFICIC project, a dense array was designed and deployed. In the following section, it will be described.

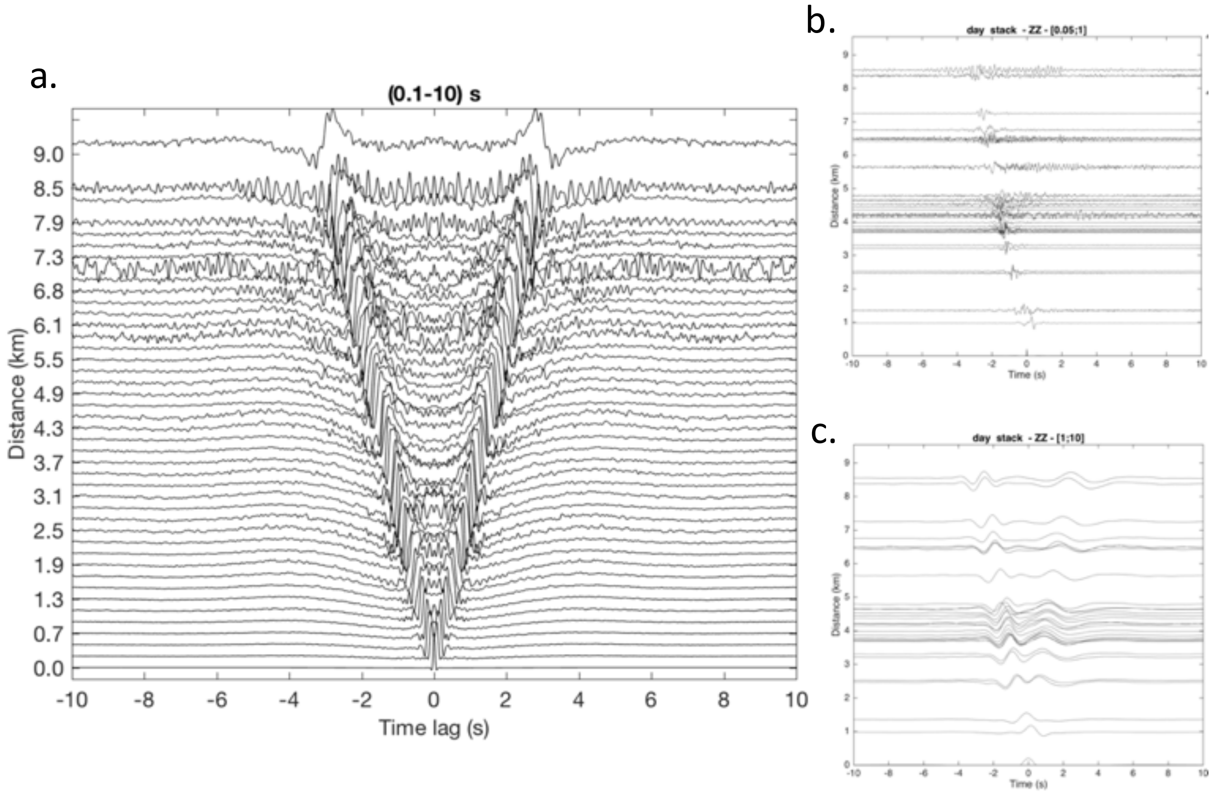

Figure 1.7: a)Stacked section of noise cross-correlations in [0.1 - 10]s for Rayleigh waves. b) - c) Noise cross-correlations for station 00.121 and all others. b) long periods [1 - 10]s. c) an short periods [0.05 - 1]s. Modified from Chmiel et al. (2017).

1.4 Marathon dataset

The marathon site corresponds to a mineral exploration block. The potential targets are bodies with high concentrations of platinum group metals (Cu-PGM) hosted in a Gabbro intrusion (Figure 1.3-b). These bodies have irregular shapes up to 200m in length and 25m thick. It is unlikely to be able to image these structures with passive seismology methods because of their size. Nerveless, the thickness and dip of the intrusion and the lower contact geometry are largely unknown. Provide this background information is valuable for the mining exploration industry.

Surface waves are often used for tomography in passive seismic. However, these waves lack the resolution needed to image localized targets at depth as needed in Marathon deposit. Instead, body waves can image small objects with high resolution at depths of several hundred meters at least. In addition, body waves and high-frequency surface waves retrieved from ambient seismic noise could improve spatial resolution to better detect geological structures' geometry. However, high-frequency arrival retrieving is challenging because low-frequency energy dominates the seismic noise records.

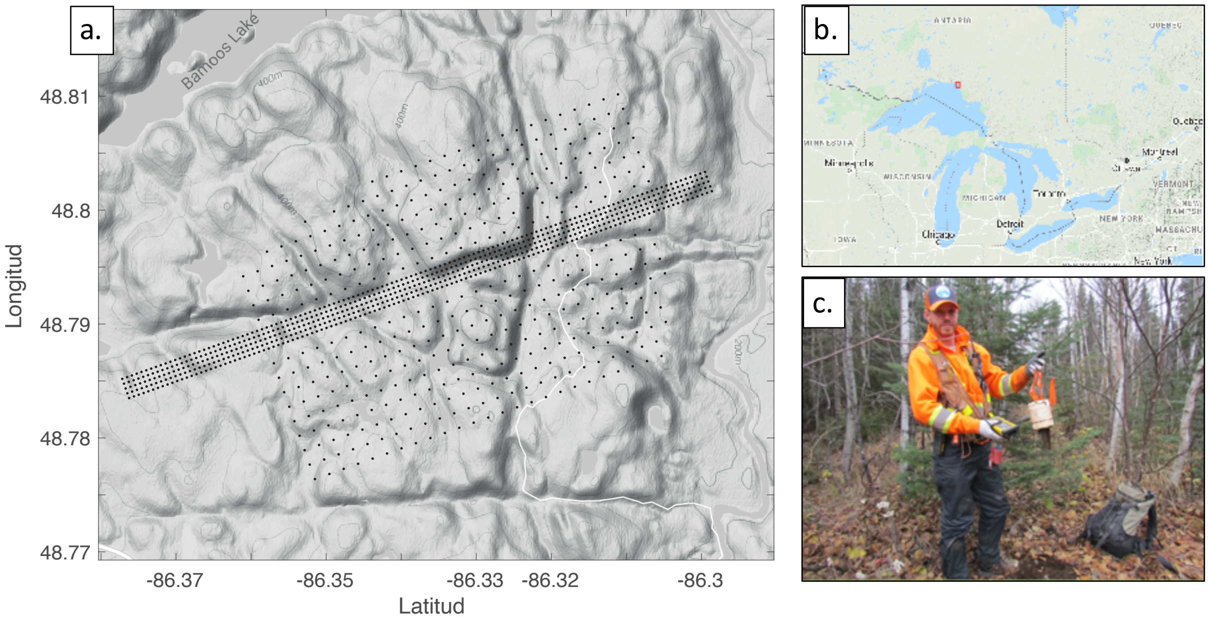

Figure 1.8: a) Grid design for the Marathon site. b) Map of North America. The site test is represented by the red rectangle. c) Deployment work at Marathon test site.

Based on the previous passive study (noise sources) and the geological requirement (gabbro's geometry), the grid was designed to develop a technique appropriate for greenfield exploration. The grid was composed of two overlapping grids, a 416-sensor array and a 609-sensor profile line. The figure 1.8-a shows the theoretical position (black dotes). The first grid is a regular array comprising 416 sensors spaced 150m apart in all directions. For the rest of the manuscript, this array will be called a sparse array. The array's extent was sufficient to cover the Marathon Deposit and the model down-dip extension to the West (see figure 1.3-a). The second grid was a profile line with a 200m width, a 6040m length, and a sensor spacing of 50m. The profile line consisted of 609 sensors (called dense line here). The noise direction was determined from an early noise test performed in 2017 (section 1.3.2, Chmiel et al. (2017)). According to these results, both grids designs were configured along Superior Lake at 250 deg to the West (Modified from STILLWATER and Sisprobe (2019)).

The sensors used for the survey were the ZLand Gen2 one component (vertical direction) with a 10Hz range. The nodes are compacts and contain a power supply convenient for passive deployments in isolated and difficult access sites such as Marathon deposit. Figure 1.9 shows the instrumental response of the sensors. The red line represents the corner frequency. Therefore, the technical details are in the table below:

Parameter:	**Description:**
Co-ordinate system	UTM WGS84 Zone 16N
Acquisition Schedule	Continuous
Sample rate	4ms
Pre-Amp Gain	36dB
Anti-alias Filter	Linear
Low Cut Filter	None
Sensor used	Internal 10hz DTCC SOLO

Table 1.1: Table of ZLand Gen2 Parameters

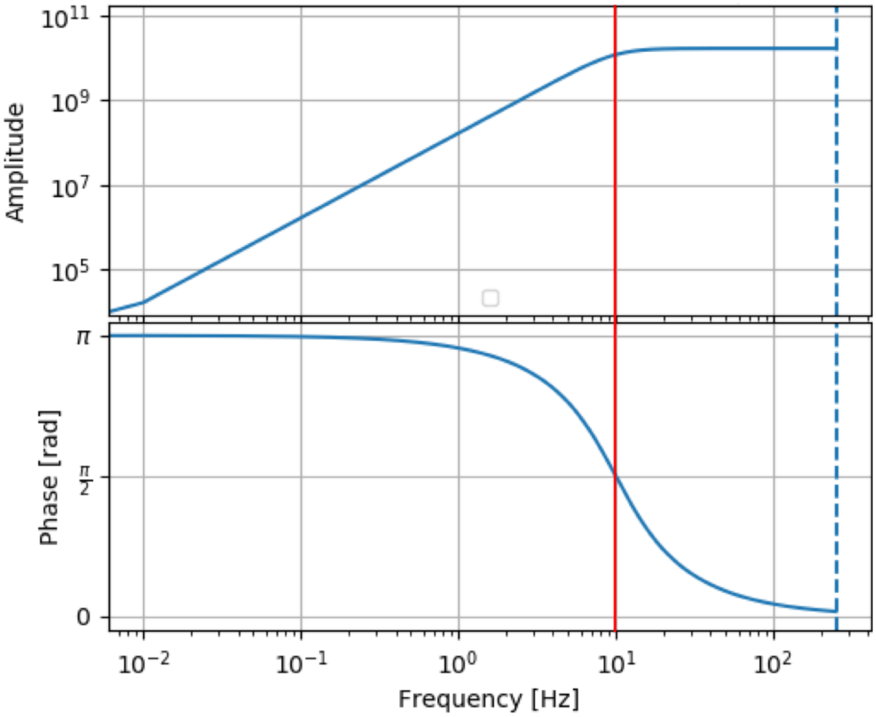

Figure 1.9: Transfer function in for ZLand Gen2 sensor. In red is the corner frequency.

The deployment of the sensors started on September 17 and was completed by September 21, 2018. This allowed for a recording time of 30 days at 250Hz. The sensors were retrieved from October 22 to October 26. Of the 1024 sensors deployed, six were not recovered, either because they were disturbed by animal movement or buried under a rockslide. In addition, some sensors (in both arrays) could not be placed at the designed grid location or within a distance tolerance of <20m because of environmental constraints as lakes or unattainable places. Most of the node's real positions were noted. For those whose position was not noted, the theoretical position will be used since the differences are minor (of the order of a few tens of meters).

Once the sensors were retrieved, the data was successfully downloaded and shipped to Sisprobe, UGA, and DIAS for analysis by late November. Then, the data were downsampled at 50Hz, and 125Hz to be used during my Ph.D. research.

In the next chapter, wewill describe the recorded noise signals and their frequency and temporal characteristics. Then, the seismic noise sources will be explored to select the most adapted for Marathon geological context.

Chapter 2

High frequency seismic noise in Marathon

Contents

2.1 Introduction

Passive seismic interferometry successfully retrieved surface waves in general up to 1Hz and mainly originated by ocean swell activity. This gives rise to a large class of methods that can be regrouped under a generic name of ambient noise surface-wave tomography (Nakata et al., 2019). Notwithstanding the good results and applications using passive surface waves for subsurface imaging, only a few studies retrieved passive body waves at this scale. Body-waves reconstruction is of significant interest because it can provide additional information about the medium than the information provided by surface waves.

To obtain images with better resolution and even to extract body waves at a local scale (first $\sim$ 10km) using passive seismic interferometry, it is necessary to use high frequency noise sources. Various authors have extracted reflected body-waves (e.g., Draganov et al., 2007; Panea et al., 2014; Olivier et al., 2015) or refracted waves (Roux et al., 2005a; Nakata et al., 2015) using ambient noise to image the sub-surface. Dong et al. (2006), Mikesell et al. (2009) and Snieder et al. (2006) estimated physical parameters from non-physical refractions resulting from spurious terms after the cross-correlation. However, the interpretation and extraction of body-waves was not straightforward and the studies were carried out in populated or industrial regions where sources of high-frequency noise sources were a priori present and strong.

At local scale the sources of high-frequency seismic noise (>1Hz) are either anthropogenic originated such as traffic, windmills, industrial machinery, etc (Bonnefoy-Claudet et al., 2006) or natural as near-coast breaking ocean waves, rivers, lakes and winds (Zhang et al., 2009; Poppeliers and Mallinson, 2015; Xu et al., 2017). These sources are not uniformly distributed in time, space or frequency, and due to their high-frequency content, they attenuate fast with distance. A good understanding of sources mechanism and their limits is needed prior to proper utilization. Due to the diverse detection techniques, it is possible to locate and discriminate them from the rest of the seismic noise records (Kong et al., 2019; Meng et al., 2019).

To obtain high-resolution imaging of the sub-surface, high-frequency seismic noise sources are needed, and their understanding is essential. Therefore to image Marathon deposit, first a detailed noise sources characterizations are needed. Thus, we will explore the anthropogenic and natural sources of the region, their frequency and temporal characteristics, and their advantages and disadvantages for imaging the shallow crust. Subsequently, train signals will be presented, showing them as the primary high-frequency energy source in the region. Finally, through observations and modeling, train-generated signals will be studied in detail to describe their source mechanism and signals properties.

2.2 Seismic noise in Marathon

Based on the previous passive seismic test in Marathon, we expect two sources of noise: the Lake Superior and the North Atlantic sea. These sources mainly generate low-frequency noise (<1Hz). However, to image the gabbro intrusion, which has a size of approximately 600m, we need to retrieve energy with frequencies greater than 5Hz, and thus high-frequency sources that allow us to reach is frequency.

Bonnefoy-Claudet et al. (2006) showed that the high-frequency noise is mainly generated by human activity. An a priori look for potential noise sources in the region was performed before a thorough analysis of the seismic records. Marathon town, the aerodrome, a mine site, and the railway and highway traffic are the potential cultural noise sources in the region. As shown in the figure 2.1, Marathon town is located on the north of Lake Superior north and at 9km southwest of the center of the array. The town area is 170.54 km^2 and the population is 3,273 people (Marathon, 2019). The aerodrome is situated 5km northeast of Marathon town; It is little crowded, mainly used for helicopter flights. Due to the size of the town and the aerodrome, we do not expect them to be a primary source of cultural noise in the region. Blasting events from an active mine site located 30km Southeast of the array are promising sources. Besides these sources, the Canadian Pacific Railway (CPRS) and the trans-Canada primary highway produce significant vehicle traffic.

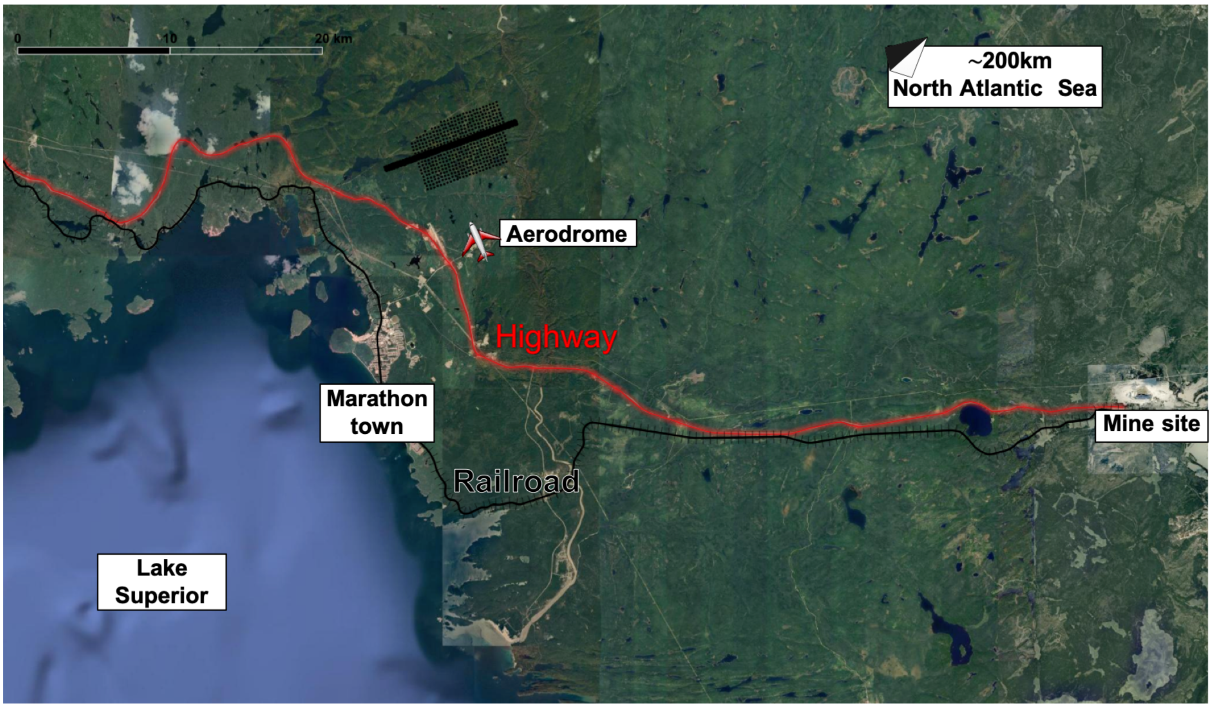

Figure 2.1: Map of sources of noise at Marathon. Gray dots are the 1020 seismic stations. The black dashed line is the railroad (Canadian Pacific Railway), and the continuous red line represents the Canadian highway (HWY).

We can look forward to natural sources of high-frequency seismic noise, Roux et al. (2018) showed that trees could behave as a locally resonant material creating high-frequency seismic noise. As shown in figure 1.1, the studied zone is a forest region, and the interaction between the tree, and the wind could be considered as a natural source of high-frequency noise.

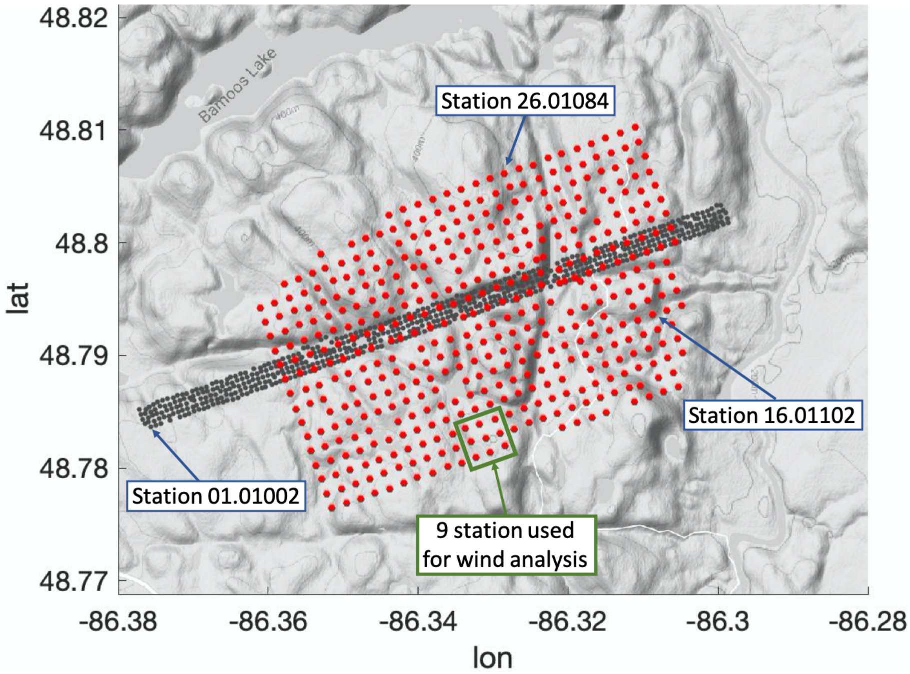

Figure 2.2: Array for the Marathon site. Gray dots are the 1020 seismic stations. In red is the array used for the beamforming analyses. The 3 stations showed by the arrow are the station used for the PSD analysis.

The raw data was examined to quantify and understand the characteristics of seismic sources of ambient noise. First, a quality check is performed based on spectrograms and Probability Power Spectral Density (PPSD).

PPSD estimates the statistical distribution of the seismic ambient noise at a given station in the function of the period and the amplitude. To calculate the PPSD, a discrete Fourier Transform is applied to the data using one-hour window without overlap, the Power Spectral Density (PSD) computed as the normalized square of the displacement spectrum, and the instrument response removed (Peterson et al., 1993). Then the probability density function is generated from PSD.

Figure 2.3 show the PPSD and the spectrograms at 3 different stations located as in figure 2.2. These three examples represent three different cases in Marathon. First, station 01.01002, figure 2.3-a and d show high amplitudes for periods between [0.1 - 10]s. This station is the nearest station to the railway and the highway. The high-frequency seismic noise is mostly due to vehicle traffic and the lake. Station 15.01102 is farthest from the railway, the high-frequency noise is still present, but the energy is attenuated with distance, so the amplitude is less important than the precedent case. Finally, station 26.01084 had a problem during the recording, and only the

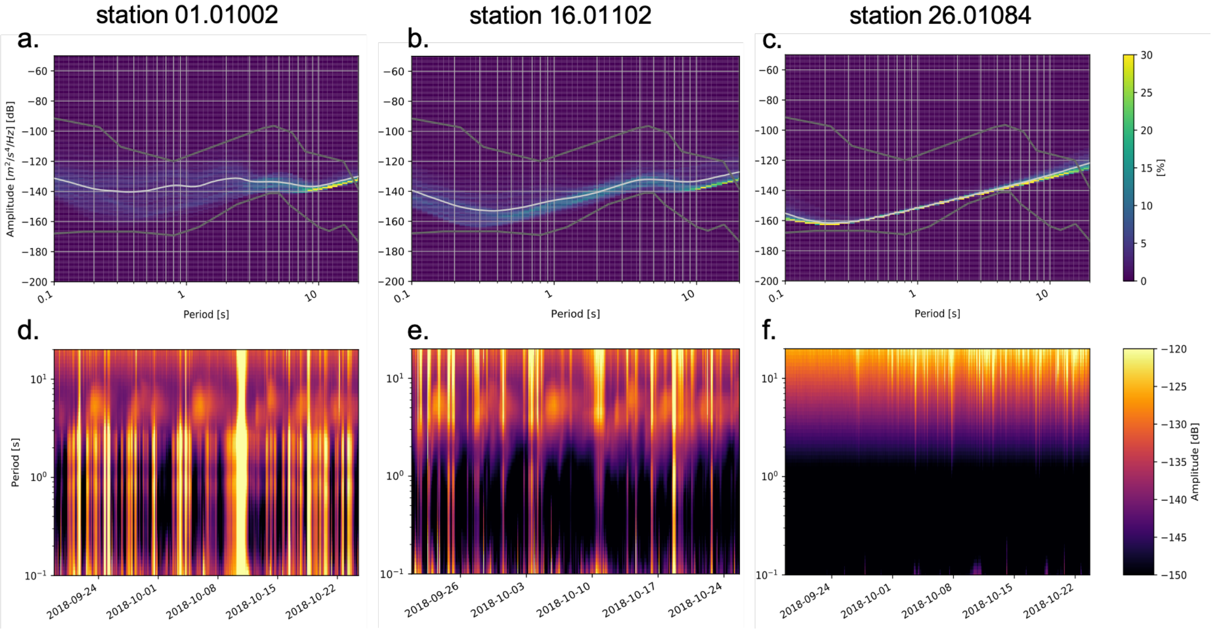

Figure 2.3: Probability Power Spectral Density (PPSD) and spectrograms computed for the 3 stations showed in figure 2.2. PPSD, first line, green to yellow colors indicate the highest probability power as a function of period. Gray lines represent the standard new low noise model (NLNM) (Peterson et al., 1993). Spectrograms, second line, are computed for one month of data using one hour segment without overlap.

electronic noise was recorded. The station will not be used for the rest of the study.

A beamforming analysis was performed to analyze the seismic noise sources distribution in time and space as well. Beamforming is an array method that estimates the coherent portion of seismic wave energy that propagates over the seismic array. Assuming a plane wave arrival, the signal at the station i is described as:

$$a_i(t) = a_{ref}(t - \tau_i) + n_i(t) \tag{2.1}$$

Where $a_{ref}(t - \tau_i)$ is the reference signal and $n_i(t)$ the incoherent noise and t the time. Knowing the stations position r_i, the delay between the stations τ is defined as $\tau_i = ur_i$; Where u is the slowness vector. The N records, $a_i(t)$, are shifted to a specific slowness and back azimuth. The time-shifted traces are then summed, producing the beam trace b(t):

$$b(t) = \frac{1}{N}\sum_{i=a}^{N} a_i(t + \tau) = a_{ref}(t) + \frac{1}{N}\sum_{i=a}^{N} n_i(t + \tau_i) \tag{2.2}$$

Suppose the single-station recordings are appropriately shifted for a particular back azimuth and slowness. In that case, all signals with the matching back azimuth and slowness will sum constructively finding the direction of propagation (e.g., Rost and Thomas, 2002, 2009). The frequency-wave-number (f-k) beamforming analysis implement this approach in the frequency domain.

In order to have high-resolution beamforming, the beam is deconvolved by the array response function (ARF). The ARF is the array response to a single monochromatic wave propagating over the array. The ARF power spectrum is separated into two parts. The beam main lobe (actual signal) is controlled by the array aperture and the beam frequency range (Nakata et al., 2019). The ARF deconvolves the beam power to have high-resolution beamforming (to decrease the width of the main beam lobe and decrease the contribution from beam side lobes).

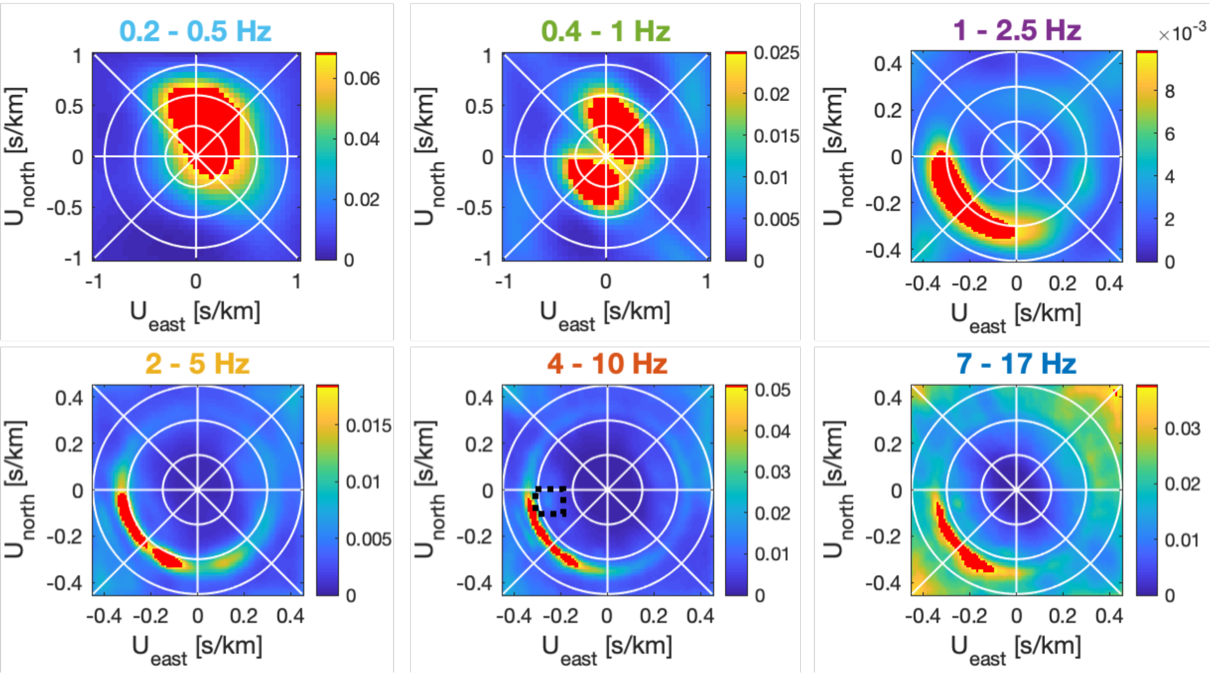

Figure 2.4: Beamforming averaged over one day of acquisition, the 1 October 2018. It is computed in 6 different frequency bands. The amplitude spectra in the horizontal slowness domain. Warm colors represent high power amplitudes.

The beamforming method was applied to the raw data sampled at 50Hz. Only sparse array stations were used (red dotes figure 2.2). we decided to use this array to have a regular grid and conserve a good resolution while decreasing computation time using fewer sensors. The data was cut into 1-hour time windows and the beamforming computed in the spectral domain. Then, the beam is averaged over one day and one month.

Figure 2.4 shows the averaged beamforming for the 1st of October in 6 different frequency bands from 0.2Hz to 17Hz. The higher frequency was chosen base on the spatial resolution of the array and the Nyquist frequency (25Hz, here). Even if the interest here is to extract high-frequency energy (>1Hz), lower frequency bands up to 0.2Hz were analyzed to characterize the noise in a general way. The figure 2.5 shows the beamforming averaged over 30 days (from 22 September to 22 October 2018) for the same frequency bands.

Beamformings (daily and monthly) show that the low-frequency energy, [0.2 - 0.5]Hz, is coming from the Northeast. This is due to the microseismic activity generated by the North Atlantic sea, as was also shown in the pilot test at Marathon (chapter 1, section 1.3.2). The daily beamforming also indicates the energy coming from the North Atlantic sea for the [0.4 -1]Hz frequency band that coexists with the energy coming from the Superior lake Southwest direction. However, the sea energy is not a daily constant for the [0.4 -1]Hz band; when the daily beamforming is averaged, the apparent amplitude is reduced, and the Rayleigh wave is not clear anymore. For the high-frequency bands (>1Hz), the Rayleigh wave is visible for a good azimuthal rage. The surface wave has an apparent velocity of 3.4km/s and illuminates the array from the Est to the South. Base on the potential sources at Marathon, this arrival could be generated by several mechanisms as the train and highroad traffic and the swell from the Superior lake. A second arrival is briefly seen for the [4 -10]Hz and [7 -17]Hz frequency bands (figure 2.4 black square). This arrival comes from the West of the array and arrives with a velocity of 3.7km/s. Given their velocity and frequency content, we suppose that it is an S-wave or a higher morde surface-wave originated by the traffic in Marathon.

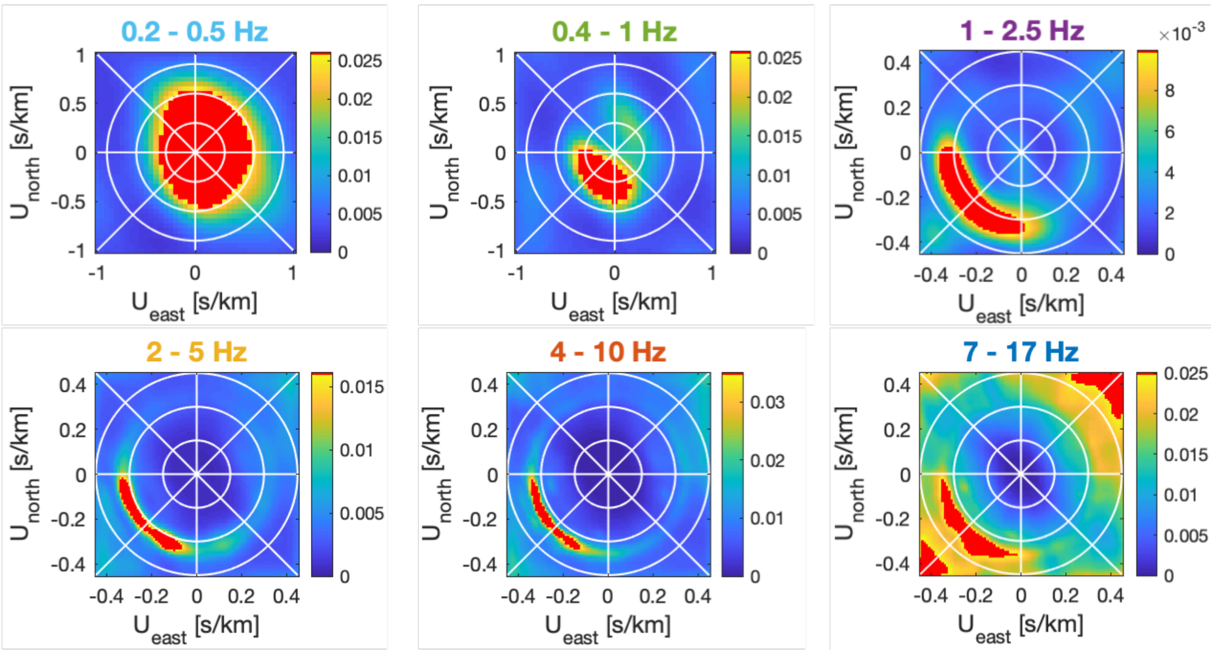

Figure 2.5: Beamforming averaged over the 30 days of acquisition, computed in 6 different frequency bands. The amplitude spectra in the horizontal slowness domain. Warm colors represent high power amplitudes.

2.3 Sources of high frequency noise at Marathon

In order to quantify the spectral characteristics of the possibles sources of noise in Marathon, we explored the noise records of individual stations. By doing, so we expect to describe the seismic noise characteristics in the time and frequency domain and therefore find sources of high-frequency noise, which can be used for imaging. We inspect the spectrograms of all stations during the 30 days of the deployment. Here, we will only present two representative examples for one day of data. These examples enable us to explain the spectral and temporal aspects of the principal sources of seismic noise in the region.

Figure 2.6 show the seismic trace, the spectrogram and the spectrum of the 21 October for 2 stations: 01.01002 and 16.011102 (as showed in figure 2.2). The data is pass-band filtered below 20Hz. Spectrograms are computed over time windows 1 minute with 70% overlap. For both stations, the high-frequency content is important and dominates the records, which is promising. We also note different particular features in the spectrograms and the seismic traces. These features will be discussed in detail in the next section, especially blast event (red dashed rectangles) and vehicle noise (blue dashed rectangles). Low-frequency energy (<2.5Hz) will cover as well.

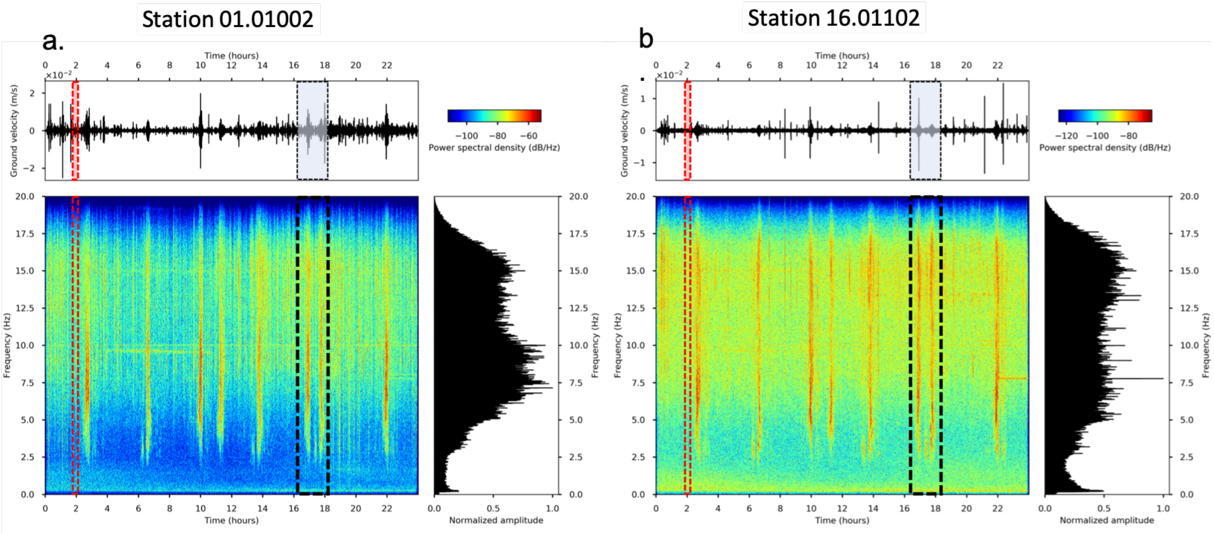

Figure 2.6: Spectral analyses for 21 October 2018. (Top of each panel) Seismograms, (bottom of each panel) spectrograms, and (right of each panel) spectra resulting from 1-day data of station 01.01002 and 16.01202 (see figure 2.2) The data is low-pass filtered at 20Hz. Spectrograms are computed over time windows 1 minute with 70% overlap. Red dashed rectangles represent 2 minutes during a blast events occurred (zoom in figure 2.7). Blue dashed rectangles represent 2 hours of data during two trains passages were recorded.

2.3.1 Superior Lake

Xu et al. (2017) showed several examples of seismic noise generated by lakes around the world. They demonstrated that lakes produced microseisms at periods of [0.5 - 2.0]s where primary and secondary peaks that are often simultaneously observed. The energy engender by lakes is smaller than in oceans because of the less energetic wave action in lakes. It propagates as short-period Rayleigh waves with a rapidly attenuation with distance (e.g., Myers et al., 1999).

The Superior Lake is one of the biggest lakes in the world with a surface of 82000 km^2 and waves between 0.3 and 1.3m high. The previous studies in Marathon showed that it is the main source of seismic noise in the region. New beamforming analyses (figures 2.5 and 2.4) confirm that the presence of microseismic energy from the Superior Lake. The energy dominates the frequency bands of [0.4 - 1] and [1 - 2.5]Hz, generating surface waves. Furthermore, Teodor et al. (2021), and Sisprobe did ambient noise surface wave tomography at Marathon, showing that Lake's noise is an efficient source of seismic noise for low-frequency retrieving. Even so, the lake seismic energy is not high-frequency enough to generate body waves for subsurface imaging. The spectra (figure 2.6) show the presence of energy below 2.5 Hz; however, this energy is relatively low amplitude compared with the energy between [3 - 20]Hz. Hence, the Lake is not the most accurate source of seismic noise in the region for high-resolution imaging of the shallow sub-surface.

2.3.2 Mine blast

Blasting is a method for fragmenting or moving rock in open-pit mines. The signal caused by blasting events release energy in the form of seismic waves, which are similar to the microsismicity signal wave-forms (Li et al., 2018a). Barrick Hemlo Mine is an active open gold mining site situated 30km away from the array. In this active mining environment blasting events are expected. Thanks to one of our partners in the PACIFIC project, we had access to the blasting timetables during the deployment. Two blasting sections were done every day, including weekends, with an average of 3 blast events per section.

Figure 2.7 shows the spectral and temporal characteristics of one blasting section (i.e., series of blasting events) the 21 October. Three blast events were detected by the sensor 01.01002 (figure 2.7-a). The blast events influence seismic noise at frequencies between [1 - 20]Hz but during short time windows ($\approx 2 seconds$). Conversely, the sensor 01.01002 does not detect energy from the blast events neither in temporal or spectral-domain even if it is closer to the mine (figure 2.7-b). This difference could be due to the local background noise that hides the blast event.

It is concluded that blast events are not appropriate in the Marathon case for the following reasons. First, there are few events per day. Second, the mine is too far from the studied zone, and it is not always possible to detect the blast events. Third, the blats are only coming from one azimuth that is not optimal for seismic interferometry.

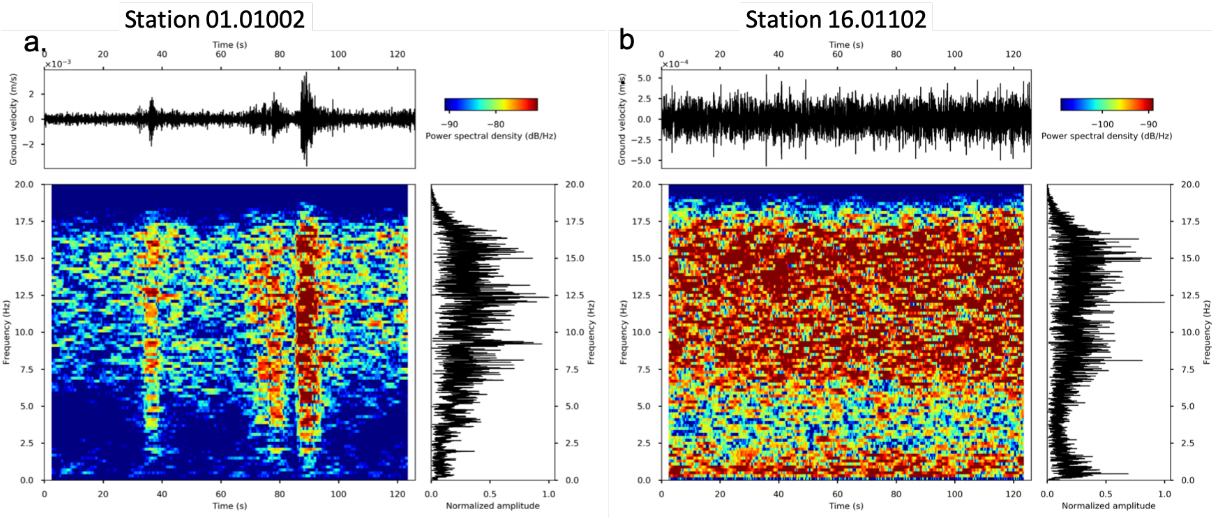

Figure 2.7: Spectral analyses during blast events. (Top of each panel) Seismograms, (bottom of each panel) spectrograms, and (right of each panel) spectra are resulting from 2 minutes (01h59-02h01) data the 21 October 2018 (station 01.01002 and 16.01202). The data is low-pass filtered at 20Hz. Spectrograms are computed over time windows 5 s with 90% overlap.

2.3.3 Wind

The ground motions generated by wind may cover significant portions of seismic records with wave-form amplitudes similar to or larger than those produced by microseismicity, i.e., magnitude < 2 (Withers et al., 1996). The turbulent wind flow can generated seismic energy in contact with the Earth's surface. The interaction with objects at the surface, for example trees, can also generate high-frequency tremor-like signals (e.g., Roux et al., 2018; Johnson et al., 2019). Wind-generated ground motions near trees have significant spatial variability, over tens of meters modulating the recording amplitude in the function of wind velocity and distance from local structures (Johnson et al., 2019). Frankinet et al. (2020) demonstrated that two linear laws relate noise level and the wind velocity at wind velocity lower and greater than 6m/s.

To evaluate the interaction between the wind trees and the seismic noise in Marathon and the eventual generation of high-frequency signals, we analyzed the correlation between the wind velocity and the Peak ground velocity (PGV) as done by Johnson et al. (2019). First, we obtained the wind speed from 22 September to 22 October. The closest meteorological station is located at the aerodrome of Marathon (4km away from the array). Therefore, the data is sampled per hour and available in The Environment and Climate Change Canada database (`https://climate.weather.gc.ca/historical_data/search_historic_data_e.html`). Then, we used the 30 days of seismic records sampled at 50Hz for 9 sensors located at the south of the array. The sensors were selected because of their close position to the meteorological station (see green square figure 2.2). Next, we calculated the PGV for 5 minutes moving windows with 1-minute overlap for each sensor separately, and finally, we computed the median of the nine sensors.

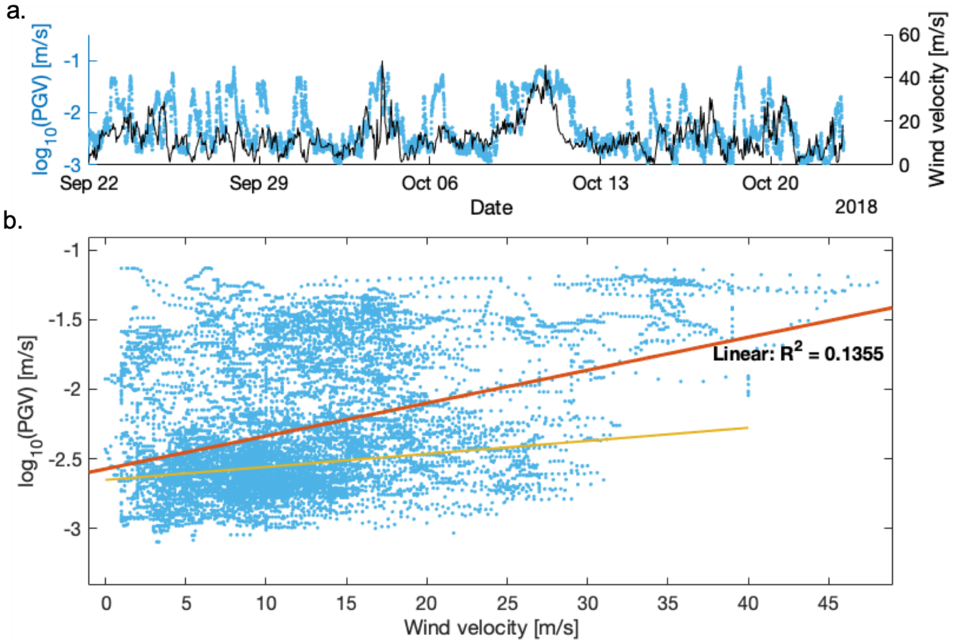

Figure 2.8: Peak ground velocity (PGV) compare to the wind velocity for one month of data. (a) The blue line represents the decadic logarithm of the PGV averaged for 9 sensors during 30 days (green square figure 2.2). Data for each sensor were processed separately by moving averaging PGV (see details in the text). Wind velocity per hour is shown in black. (b) Blue dots represent the wind velocity versus the decadic logarithm of the PGV. The orange line is a linear regression of this data, and the yellow line is the linear regression for PGV $< 2.5m/s$.

Figure 2.8-a shows the averaged PGV (blue dots) compared to the 1-hour sampled wind speed (black line). A linear log correlation between PGV and wind velocity is clear for low wind velocity values (<20m/s). Otherwise, the correlation is unclear for greater velocity values except for two days (from 2 to 4 October and 10 to 13 October). Thus, the greater PGV values are not only generated by the wind but from others sources of seismic noise. Figure 2.8-b shows the PGV against the wind velocity (blue dots). The orange line represents the data's linear regression with a Pearson coefficient of 0.1355, which indicates a low correlation between both parameters. For PVG values smaller than 2m/s, a linear tendency is visible (yellow line) as been

described before by Frankinet et al. (2020). Even if we showed a correlation between the seismic noise and the wind velocity, this mechanism produces local noise sources inside the array, which does not meet the appropriate conditions for seismic interferometry. So this source should be considered mainly as a source that pollutes the records.

2.3.4 Vehicle noise from the highway

Trucks or vehicles moving along highways induce vibrations into the ground generating high-frequency seismic signals (>1Hz). These signals can be detected and discriminated from the rest of the seismic records using its seismic properties (e.g., Riahi and Gerstoft, 2015; Diaz et al., 2017). In addition, several studies used continuous seismic records dominated by traffic noise to apply seismic interferometry to successfully extract surface waves (Behm, 2017; Chang et al., 2016) and both surface and body-waves (Nakata et al., 2011) for imaging purposes.

Marathon's highway is located west to the south of the array, and the closest sensor is 1.3km away from the highway. This highway is relatively calm, and the transit is mostly due to cars and mining trucks. It is important to note that large and heavy trucks produce more seismic noise than standard cars. The Annual Average Daily Traffic (AADT) for Marathon's highway, defined as the average twenty-four-hour, two-way traffic for the period 1 January to 31 December, is 2,250 for 60,9km railway long in 2016 (Traffic-Office, 2016). Knowing that this value is for 60km of highway and not just the portion near the array, we can deduce that the AADT value at Marathon is considerably lower.

The averaged number of passages per day is not insignificant, but the vehicle-generated signals are not perceptible by all sensors. For example, figure 2.9 shows the difference in the temporal and spectral domain for 2 sensors regarding vehicles-generated signals (black dashed rectangles). One sensor is located near to the highway 1.3km away (left panel), and the other one is 4km away from the closest portion of the highway (right panel), as shown in the figure 2.2. The spectrogram (left dashed rectangle) shows high-frequency energy between [7 - 20]Hz shaped like vertical spectral lines for the station close to the highway. Typical traffic-generated signals are observed in the temporal domain; their signature is a tremor-like signal with a small amplitude and a duration of approximately 1 minute. However, only one signal is detected when the sensor is farther, most likely because it is produced by trucks or heavy vehicles. The other traffic-generated signals are not clear anymore because they are not energetic enough, and the produced energy decreases rapidly with distance.

For this reason, the Northeast part of the array does not receive energy from vehicle traffic, and this source can not be used directly for seismic interferometry in the Marathon case. Nevertheless, other powerful traffic-generated seismic signals are discernible around the entire array, train-generated signals. The following section will describe the train signals observations, and then the source mechanism will be explained.

2.3.5 Train seismic signals: primary source of high-frequency noise in Marathon

Vehicle traffic has long been recognized as a pervasive source of noise detrimental to the quality of seismic records (Douze and Laster, 1979). In recent years, the intriguing tremors generated by trains startled seismologists and gave rise to a number of publications related to signal detection and characterization (Riahi and Gerstoft, 2015; Li et al., 2018b; Green et al., 2017; Fuchs et al., 2018; Inbal et al., 2018) and source modeling (Lavoué et al., 2020). In pioneer studies Nakata et al. (2011), Quiros et al. (2016), and Chang et al. (2016) have proposed the idea of using seismic records dominated by traffic noise and to passive seismic imaging for near-surface applications. They proved that the presence of traffic noise in the seismic records helps to retrieve high-frequency signals from passive seismic interferometry. By further applying the concepts of seismic interferometry to the correlation of this long-range train-generated noise, Brenguier et al. (2019) demonstrated the possibility of extracting useful information on the Earth's shallow crustal structure and temporal changes down to a few kilometers depth.

As shown before, high-frequency energy in the range of [1 - 17]Hz comes from the Southwest in the Marathon area. Several sources of cultural and natural noise are located in this direction, but none are stronger enough to be perceived by all the array except train traffic. Trains passages generate signals energetic enough to be detectable for the entire array. The railroad is located from the Southeast to the West of the array, similar to the highway's location (see figure 2.1). This rail line is a rail freight transport only used for transport cargo. Thus, only freight trains, in large part hauling filled containers, are crossing the region at any time of the day.

Figure 2.9 shows the temporal and spectral characteristics of the train-generated signals observed at Marathon. Two hours of records of 22 October were explored in the [0 - 20]Hz frequency band. During this period, two trains traveled along the array. In the temporal domain, the train signals are characteristic of their tremor-like signals, and in the spectral domain, the train produces broad-band signals from 2.5 to 20Hz. Note that the data used for this figure was sampled at 25Hz, for this reason, the maximum observed frequency is 20Hz. Horizontal spectral lines (figure 2.9-c) are also remarkable during train passages, the modeling in the following section will help to explain this behavior. Even if the power spectral density and the temporal amplitude decrease with distance, it is essential to note that the signals are observable with a greater amplitude for both sensors (panel a and b), which is the case for the whole array. Equally important, trains tremors are recorded for more than 20 minutes. Thus, we can conclude that freight trains are the primary source of high-frequency noise in Marathon, generating promising high-frequency seismic noise for long periods several times per day and perceivable for all the array.

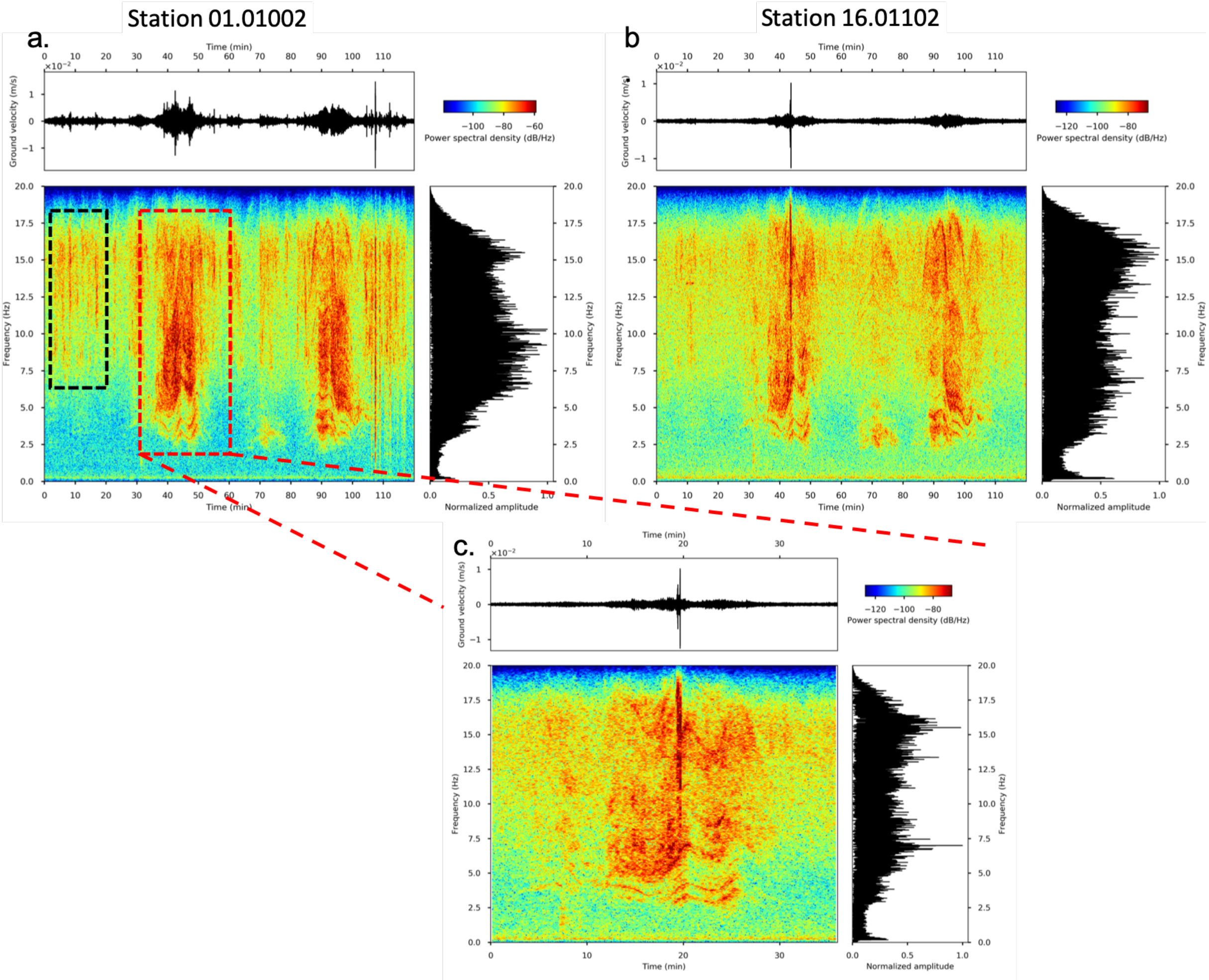

Figure 2.9: Spectral analyses during train and car passages. (Top of each panel) Seismograms, (bottom of each panel) spectrograms, and (right of each panel) spectra are resulting from 2 hours (16h-18h) data the 21 October 2018 (station 01.01002 and 16.01202). The data is low-pass filtered at 20Hz. Spectrograms are computed over time windows 20 s with 70% overlap. a) black dashed square: car passages, red dashed square: train passage. Figure c represents a zoom in the data during a train passage for the first station.

To characterizing the wavefield during a train passage, we applied 3-minute long window beamforming. Figure 2.10 shows a train passage record at station 01.01002 and 8 beamforming panels computed using the sparse array for the 3 minutes of data filtered between [7 - 17]Hz.

Three coherent arrivals are seen, one at 3.1km/s considered as a surface wave. The high apparent velocities at 3.7km/s and 7.1km/h suggest body waves (S- and P-waves corresponding). It can be seen how the array receives coherent energy from a different azimuth (following the railway from West to East) over time, conforming train's position. Thus, the energy propagates on a particular path (train position - array), allowing a vast illumination and providing additional information about the medium. Note that the power amplitude varies for each beamforming panel, showing the greater amplitudes for panels 2 and 5. To explain the origins of amplitude variation in train signals, we will present train seismic signal modeling in the next section.

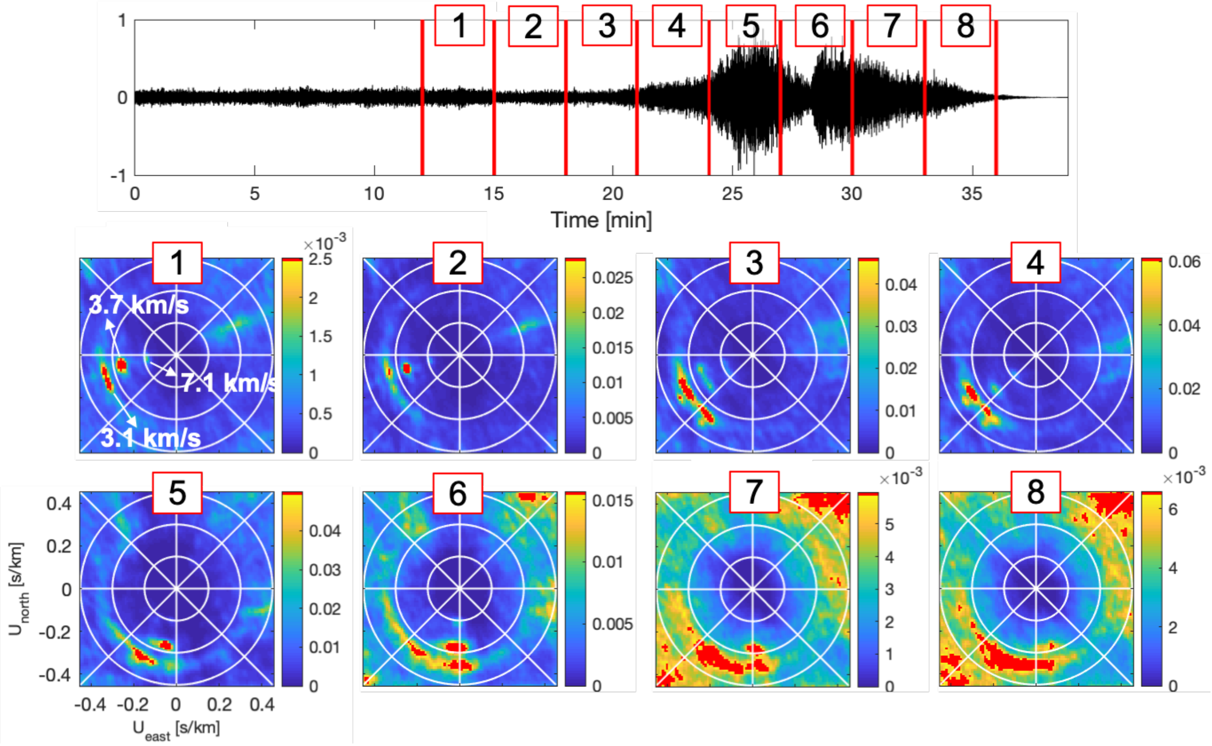

Figure 2.10: Beamforming during a train passage. Upper panel: normalized train tremor at station 01.01002. Lower panels: 3-min window beamforming corresponding to the numbered sections in the recording. They are computed using the sparse array and filtered between [7 - 17]Hz. Warm colors represent high power amplitudes; note that amplitude is variable for each panel. In white, the apparent velocities of high amplitude arrivals (beamforming 1).

To see if the behavior of train signals is common to several passages, we studied several train passages for data sampled at 50Hz and not filtered. Figure 2.11 shows 6 different train passages for the same sensor (01.01002). The train-generated signal is similar in the temporal domain and dominates the records. To analyze the similarity in view of the array, we beamformed a one-minute window for each train separately, using the sparse array and filtering the data between [1 - 17]Hz. The array is illuminated from similar azimuths for different train passages showing the repeatably of the source. These results conclude that trains passages are the only significant source of high-frequency surface waves and body wave energy. Besides, train passages are a repeatable seismic source of cultural noise (approximately 6 train passages per day).

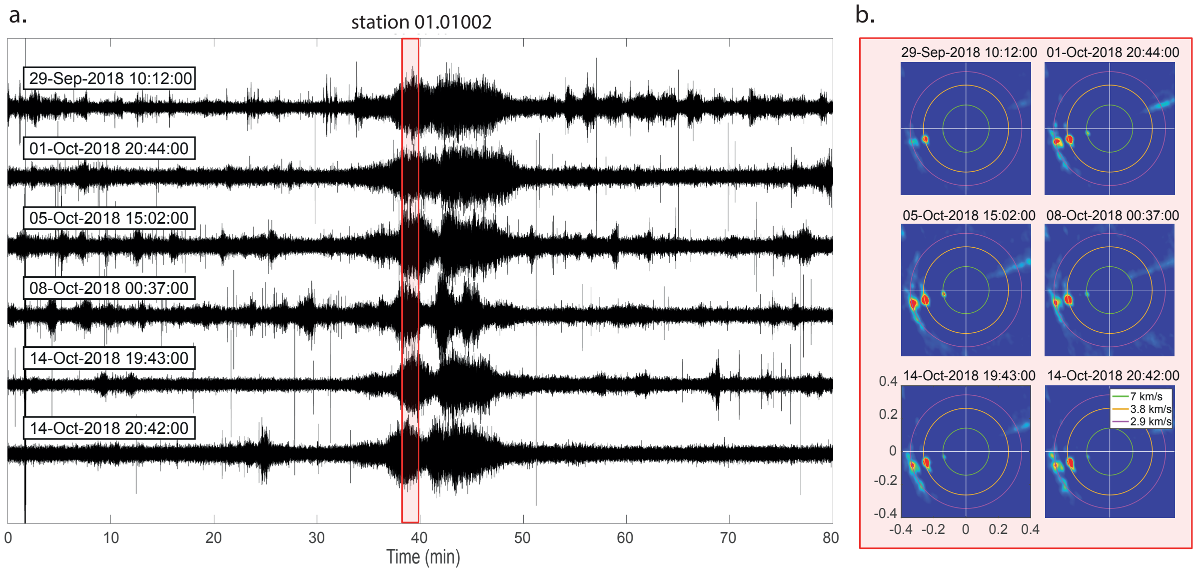

Figure 2.11: a) 6 train tremors recorded at station 01.01002. b) 1-minute window beamforming for six train passages, corresponding to the red square in the Right panel. The beamformings are computed using the sparse array and filtered between [7 - 17]Hz. The amplitude spectra in the horizontal slowness domain and warm colors represent high power amplitudes.

2.4 Train signals characterization

After realizing that the primary high frequency (>1Hz) noise source in Marathon is the freight trains passing nearby the array, we decided to understand these signals from a seismic point of view by modeling them. This section material is modified from Pinzon-Rincon et al. (2021) section 1 and inspired by the exchanges and collaborations with François Lavoué about train modeling summarized in Lavoué et al. (2020) within the framework of the PACIFIC project.

Massive freight trains generate seismic wave-forms with a striking similarity to episodic tectonic tremors (figure 2.9 top panels). As Inbal et al. (2018) reported, the identity of the sources as man-made was not obvious because freight train traffic often lacks cultural diurnal or weekly modulation, and typical train speed (25 m/s or 90km/h) is in the range of reported tectonic tremor migration velocity at depth. However, train hum has a distinct signature with clear spectral lines above 1Hz (Fuchs et al., 2018), as illustrated in Figure 2.14 for a train signal recorded in Marathon about 3km from the railway.

The engineering community has studied train-induced ground vibrations thoroughly to damp them and mitigate potential hazards. Several source mechanisms are under study (e.g., Connolly et al., 2015) including quasi-static excitation due to axle loads and dynamic interactions between trains, track, and ground. In a recent study, Lavoué et al. (2020) showed that the quasi-static excitation due to axle loads is the primary mechanism explaining the spectral characteristics of seismic signals observed at intermediate to long distances from the railway (from hundreds of meters to tens of kilometers Fuchs et al., 2018; Inbal et al., 2018; Li et al., 2018b; Brenguier et al., 2019).

Lavoué et al. (2020) conclude that the spectral lines arise from the complex interactions of periodic loads from the regularly spaced wheels on the also regularly spaced sleepers. The frequencies of these spectral lines depend on train geometry (i.e., train car length and wheel spacing within each car), spacing between sleepers, and train velocity. Nevertheless, simple assumptions are sufficient to reproduce the main features of observed signals and enable us to understand the complexity of seismic signals produced by trains.

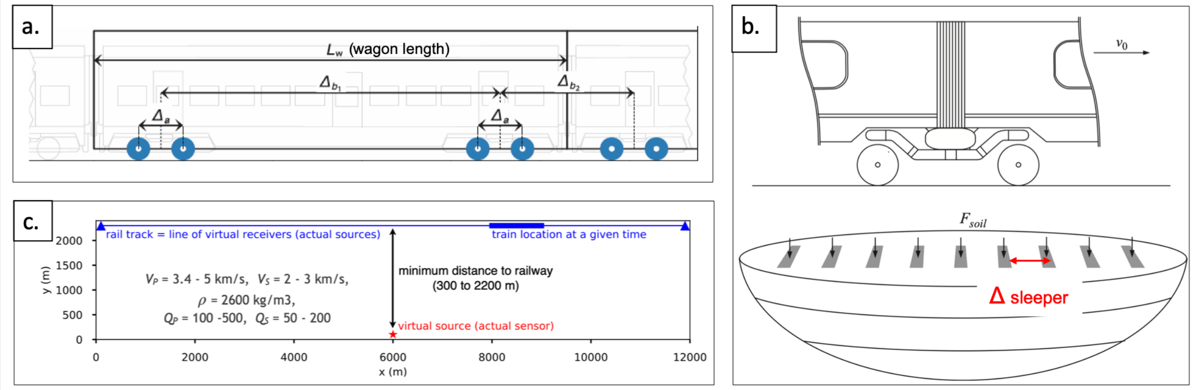

Figure 2.12: Train simulations geometry. a) Geometry of a train and b) Schematic representation of the forces applied by a train. c) Geometry of the numerical space experiment with the values used for the simulations. a) and b) modified from Lavoué et al. (2020).

One may then model train-generated seismic signals by considering only the vertical forces due to loading applied by axles on the railroad ties (commonly called sleepers, figure 2.12-b) and activated by the passage of the train wheels (Krylov and Ferguson, 1994; Lavoué et al., 2020). Therefore, we consider only one single rail, assuming that a load of each train axle is applied as a single point source (i.e., the two wheels of each axle act simultaneously on each sleeper). Using source-receiver reciprocity, we simulate the signal that would be generated by a train passing over the sleepers and recorded by a sensor away from the railway as the sum of the signals resulting from a virtual source located at the sensor position and recorded at the sleepers positions (2.12-c). A Dirac source time function is used in SEM46 spectral-element software for the simulations (Trinh et al., 2019) in the visco-elastic approximation in a homogeneous medium (the values are in the figure 2.12-c).

Two mechanisms explain well the observations made by Fuchs et al. (2018) and Li et al. (2018b): the single stationary source (sleepers) and the moving load (wheels). When considering irregular sleepers, signals are dominated by the signature of the sleepers acting as stationary sources, and we then retrieve a fundamental frequency $f_1 = v_{train}/L_w$ controlled by train speed and wagon length L_w and the modulation of the spectrum is according to the train geometry. Figure 2.13-a, shows the spectrogram with a narrow frequency spacing that corresponds very well to most of the observations made by Fuchs et al. (2018) (figure 2.13-b). When perfectly regular sleepers are considered, signals are dominated by the signature of the train wheels acting as moving loads. This generates a harmonic signal with a high fundamental frequency $f_2 = v_{train}/\Delta sleeper$ (usually ∼50Hz), related to train's speed v_{train} and sleeper spacing $\Delta sleeper$ and a clear Doppler effect. Figures 2.13- c and d show a Doppler effect from synthetic and observed data, respectively. Doppler effect is caused by a characteristic length; it is not the sleeper spacing but a longer spatial period of the railway, which results in a lower frequency f_2. In this case, the length of rails themselves explains the frequency spacing, emitting seismic waves at the end of unwelded rails. f_2 depends on the geometry of the railway and not of the train, and the underlying mechanism is fundamentally different, typical of a moving load.

The two signatures coexist in real signals, and the dominance of one mechanism over the other depends on source regularity. Although, sleeper regularity should be understood as the coupling between the sleepers and the ground via the ballast and substratum (i.e., an effective parameter) The spectral lines generally related to frequency f_1 are predicted to be in a range of about 3 to 20Hz, which is ideal both for high-frequency surface wave tomography of the near subsurface and for crustal body-wave imaging and monitoring (wavelengths not too large and scattering not too strong, Brenguier et al., 2019). On the other hand, the energy-related to frequency f_2 has a higher frequencies content, which is of particular interest for the retrieval of high-frequency body waves (Figure 2.14-b).

Lavoué et al. (2020) also observed that the spectrum changes with variable train velocity as in figure 2.14-b. The spectrum becomes broadband as frequency notches disappear. Broadband signals are advantageous for passive seismic applications because distinct spectral rays are difficult to handle by seismic interferometry. However, at short distances, small-amplitude body waves might be barely visible in the raw data, either because surface waves hide them or because they are below the ambient noise level. Using train tremors for seismic interferometry thus depends both on detection limits (instrument sensitivity and local noise level) and on reliably recognizable features in train signals. Open-source codes reproducing Lavoué et al. (2020) results and the examples presents here are available at `https://gricad-gitlab.univ-grenoble-alpes.fr/pacific/publications/2020_Lavoue-et-al_SRL_supplemental-material`.

At Marathon, the observations match to the models proposed by Lavoué et al. (2020). Figure 2.14-a shows a train signal recorded at 3km from the railway. The spectral lines related to fundamental frequencies and their harmonics are observed and the oscillations produced by the train speed variations as well. Thus, the broadband signals produced by train tremors in Marathon are a perfect source of noise that can be used to extract high-frequency signals from ambient noise recordings.

2.5 Conclusion

Given the modeling results presented before and train's seismic observations (Fuchs et al., 2018; Inbal et al., 2018; Li et al., 2018b), we know that trains are one of the most powerful and persistent seismic sources of cultural noise. Brenguier et al. (2019) estimated that the radiated seismic energy from a single 1-km-long freight train traveling through a 10-km-long railway section assuming a vertical force induced by the loading of each wheel is equivalent to a magnitude 1 earthquake.

However, trains are not comparable with sources used in active seismic exploration, such as vibroseis trucks, air guns, or explosives, among others. Active sources have specific signals that are well known in the time and frequency domain. The source type, the number of shots, and location are fully controlled and chosen depending on the investigation depth and the studied region extension (Dobrin, 1953; Huenges and Ledru, 2011).

Train signals can not be as considered an active seismic source because trains are moving and extended sources for which we do not have full control. Meanwhile, it been demonstrated by modeling and seismic observation that trains generate high-frequency energy up to 60Hz and can be detectable up to 100km away (e.g., Lavoué et al., 2020; Inbal et al., 2018; Pinzon-Rincon et al., 2021). Furthermore, trains are sources located at specific locations (railways) at the surface of the Earth, of which we know the source mechanism. Thus, trains are considered an opportune source because they are readily available, detectable, repeatable, and generate high-frequency broadband energy. As we can distinguish the train signals from the rest of the noise records, it is possible to categorize them and use them to take advantage of all their benefits for imaging purposes using seismic interferometry.

A Green's function retrieval using seismic interferometry requires the correlation of either a fully diffused seismic wavefield or of noise signal generated by sources distributed all around the studied region, including at depth (Wapenaar, 2004). In practice these conditions are never fulfilled, leading to partial reconstructions and potentially biased arrivals (Snieder et al., 2006; King and Curtis, 2012). For opportune sources of noise such as freight trains tremors usual, blind correlations of long time series of noise for passive seismic interferometry should be considered with care.

When we blindly correlated traffic train noise without considering the effects of non-even source distribution we can not consider the correlation function as the Green's function. In the next chapter, we will present a methodological framework focusing our approach on the stationary zones and propose a signal processing strategy for applying seismic interferometry for opportune sources in general, and then the application to train noise with a focus on high-frequency retrieval in the context of mineral exploration.

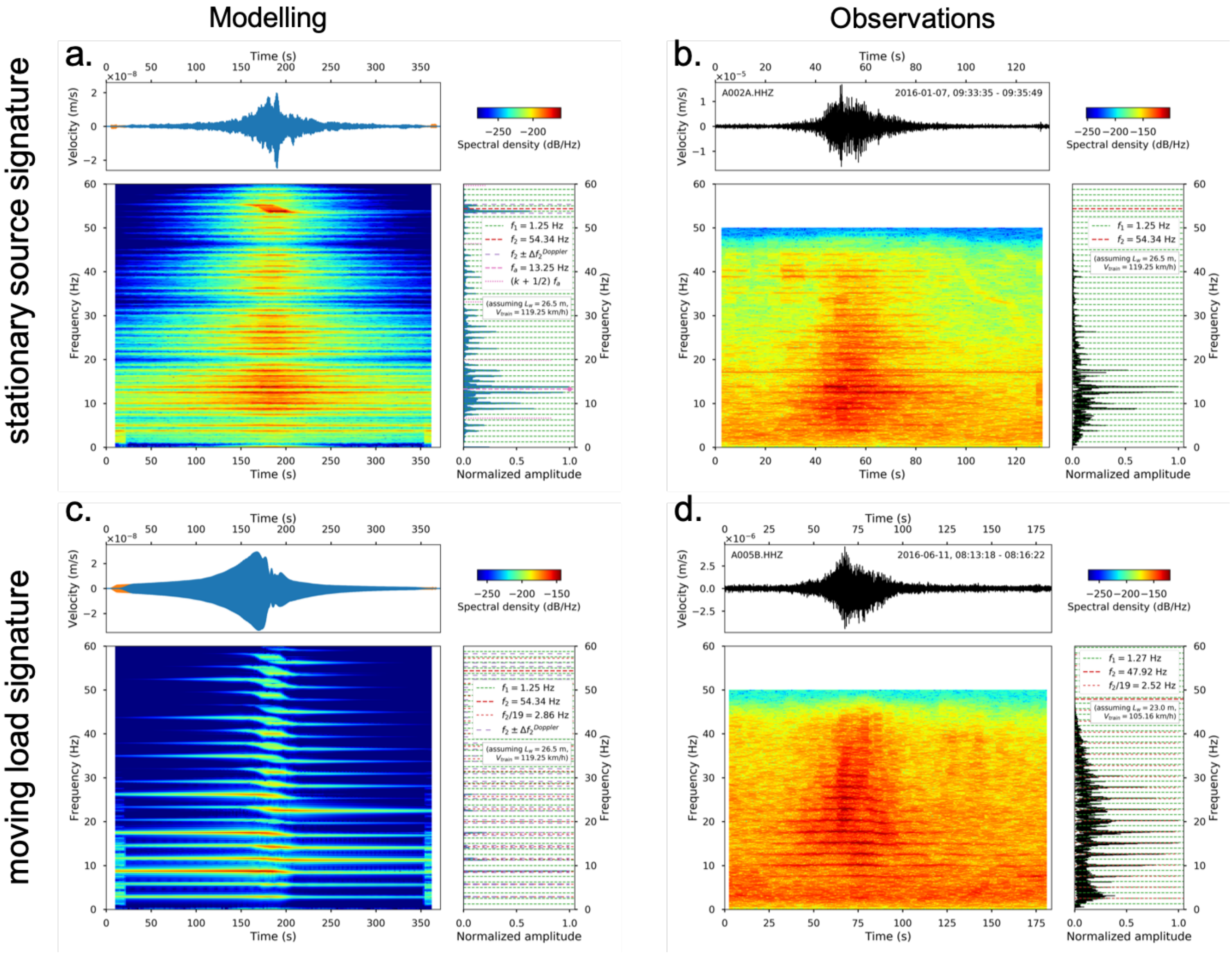

Figure 2.13: (Top of each panel) Seismograms, (bottom of each panel) spectrograms, and (right of each panel) spectra resulting from modeling train passages and observations. The observation are from AlpArray data as shown in Fuchs et al. (2018). Spectrograms are computed over time windows of (b and d) 5 s for observations and (a and c) 20 s for synthetics with 90% overlap. a) and b) Signals dominated by the end-member signature of a single stationary source of the modeling and observations, respectively. c) and d) Signals dominated by the signature of a single moving load of the modeling and observations, respectively. Modified from Lavoué et al. (2020).

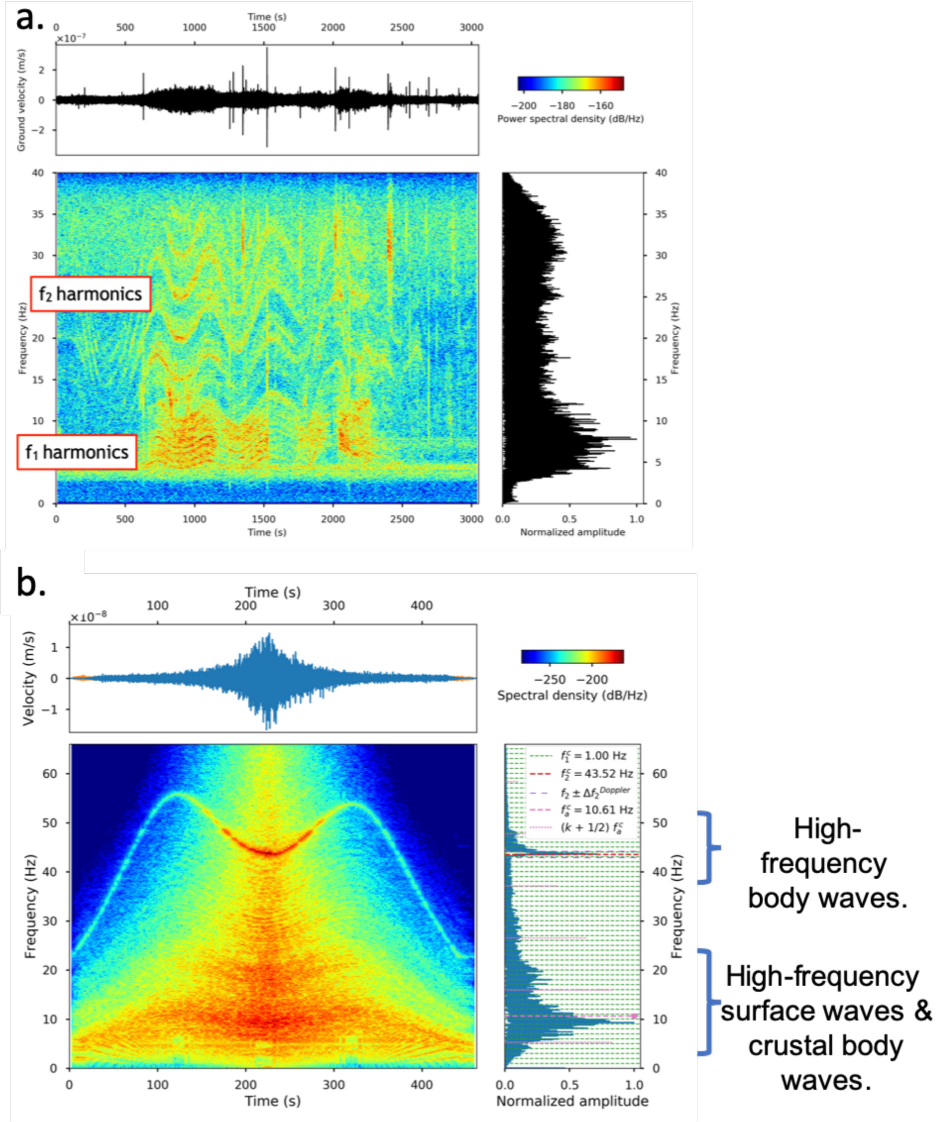

Figure 2.14: (Top) Seismogram, (bottom) spectrogram, and (right) spectrum. a) Train tremor recorded 3km away from a seismic station in Marathon, Canada. Spectrogram showing clear spectral lines oscillating as train speed varies controlled by the two mechanism (f_1 and f_2). b) Modeling resulting from a train traveling over slightly irregular sleepers at a speed variable (increasing and decreasing) from 50 to 120km/h. The spectrogram is computed over time windows of 5 s with a 90% overlap. Modified from Pinzon-Rincon et al. (2021) and Lavoué et al. (2020).

Chapter 3

Alternative processing method using train signals

Contents

3.1 Introduction

The material presented in this chapter has been written based on the Pinzon-Rincon et al. (2021) sections 2, 3, and 4.

Seismic interferometry is a general term embracing all methodologies aiming to infer seismic responses from the correlation of seismic signals observed at multiple receiver locations (e.g., Wapenaar et al., 2010a,b). To retrieve a Green's function using the correlation or an equivalent operator the theory heavily relies on either a stationary phase condition (e.g., Snieder, 2004; Roux et al., 2005b) and/or an equipartition of modes defining a diffuse field (e.g., Sánchez-Sesma and Campillo, 2006). However, even with a dominant distribution of sources at the free surface, several studies confirmed the feasibility of retrieving the Greens function, especially for waves body-waves (e.g., Draganov et al., 2009, 2013), and even explicitly using traffic noise (Nakata et al., 2011). Using controlled sources to retrieve body-wave response through interferometry is similar to daylight imaging developed by Schuster et al. (2004) or to the virtual source approach discussed by Bakulin and Calvert (2006) for borehole imaging.

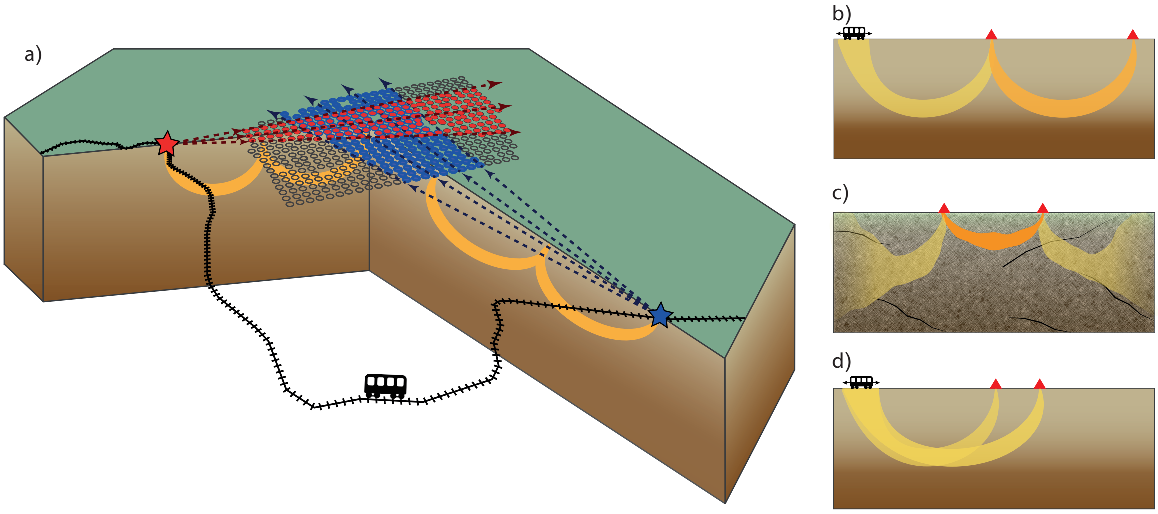

Figure 3.1: Schematic representation of seismic interferometry for opportune sources. (a) A railway surrounding a dense geophone array; an example from the Marathon deployment. 2 different train locations (stars) allow for the illumination of the array with different azimuths. Yellow kernels are schematic views of the propagation of diving P waves. (b–d) Three different scenarios of wave interference: (b,c) leading to a proper measurement of a diving P wave and (d) leading to a spurious or virtual refraction measurement. Modified from Pinzon-Rincon et al. (2021).

Usually, passive seismic interferometry relies on blind correlations of long-time series of noise for imaging and monitoring purposes. Instead, we introduce an alternative processing method for opportune seismic noise sources based on an accurate understanding of noise source mechanisms, time window, station pair selection, and specific seismic phases extraction (surface or body waves). Under this approach, we present the general workflow and then the detailed application at Marathon using train signals as an opportune source for passive seismic interferometry.

Train signals are practical for interferometric studies because we can easily detect or learn that a train is coming in advance. A single train's motion could illuminate many azimuths and potentially different depths if a railway is sufficiently close to a targeted area. Figure 3.1-a shows an example of geometry in Marathon (Ontario, Canada). A railway essentially surrounds a temporary array to assess an ore deposit (detailed in the previous chapters). By selecting station pairs aligned with train locations (illustrated for two positions by red and blue stars), one can potentially illuminate the ore body from a broad azimuth range. Figures 3.1-b to 3.1-d are schematics of several P-wave interference scenarios, each with a pair of stations. They offer a perfect ballistic interference between a diving P and PP wave (Figure 3.1-b) leading to a directly measurable diving P wave between the two receivers; and a classical scenario of a scattered wavefield from which we expect some random source energy to transit between the two receivers (Figure 3.1-c). See also a more problematic interference between two diving waves or a head wave recorded at the two stations (Figure 3.1d). Instances of this last scenario are sometime regarded as spurious correlations or virtual refractions (Dong et al., 2006; Snieder et al., 2006; Mikesell et al., 2009). Although not included within the impulse response between the two stations, this last correlation feature might be useful for imaging if it is well distinguished from expected diving waves (Dong et al., 2006).

3.2 Preliminary studies using train signals at Marathon

As shown in the last chapter 2 the primary source of high-frequency seismic noise in Marathon is freight trains traveling on the railway southwest from the array; in addition, previous studies performed with the same dataset by Dales et al. (2020) reinforce this suggestion. These studies evaluated the seismic noise and concluded that high-frequency seismic noise originated mainly from trains and vehicles passing near the array. They selectively stacked cross-correlation during periods where vehicles and trains pass near the array, limiting the selection to azimuths in line with the dense line array. They created common midpoint (CMP) gathers computed by inter-station distance binning and stacked cross-correlation for all station pairs along the dense line that share a CMP (midpoint is at the center of the dense line). The study demonstrated that selecting the portions of the noise recordings that correspond to train passages significantly improves the retrieval of body waves compared to correlating the entire noise stream.

Figure 3.2 compares periods without noise directionality (also called quiet periods) and periods during which the ambient noise came from the West (when trains and vehicles are aligned with the dense line). The first colon shows the virtual common midpoint gathers of quiet periods filtered between [10 - 30]Hz and [20 - 60]Hz; not explicit body wave retrieval is observed. By constants, the gathers at the left colon show a dominant arrival with apparent velocities of 3600m/s suggests a head S waves arriving with about a 20° angle of incidence and either reflected or refracted P-wave at 7000m/s.

The study is illustrative, but the results did not allow them to perform 3D imaging. Therefore, we moved a step further by separating and binning noise azimuths for virtual source retrieval in different directions. Hence, wedecided to try to illuminate specific ray paths by using a new data processing workflow. We detected train passages, inferred the positions and azimuths of the trains relative to the array, carefully selected station pairs and time windows for correlations, and finally stacked by train passage and azimuth.

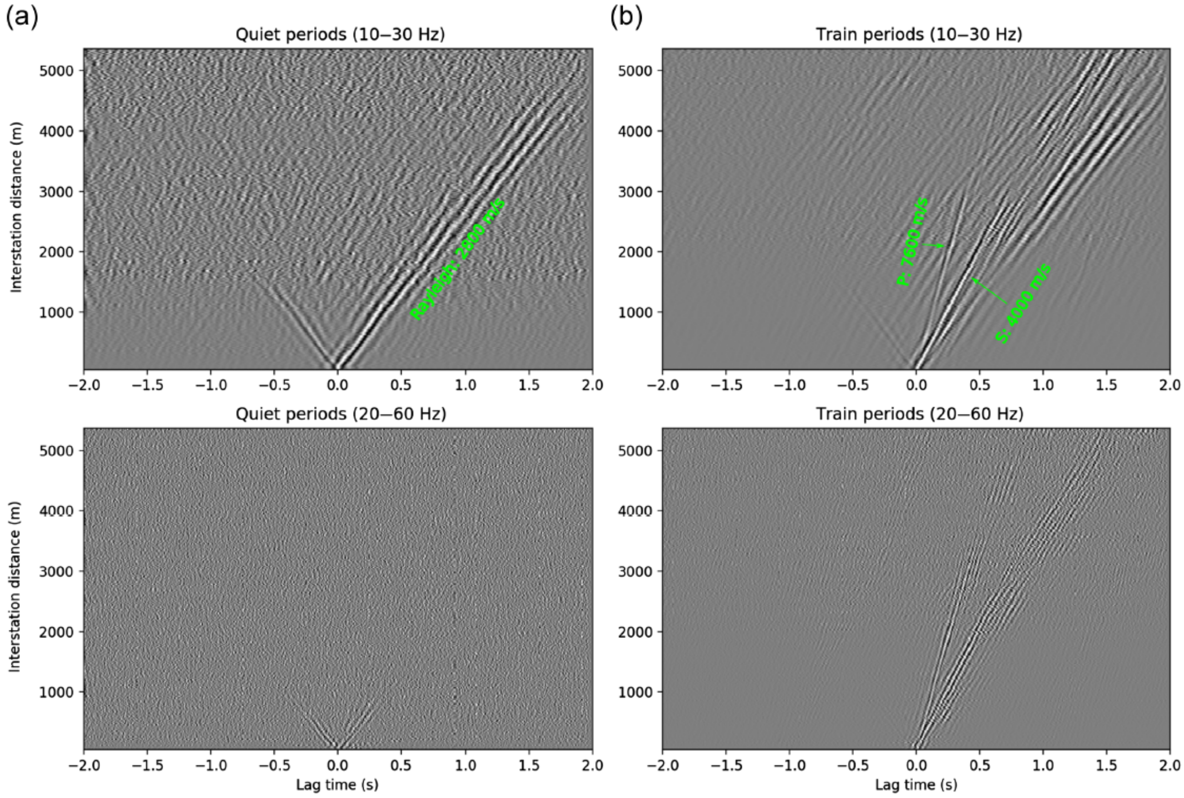

Figure 3.2: Virtual common midpoint (CMP) gathers for the dense line selectively stacking only. (a) CCFs for “quiet” time periods in which there is no obvious noise directionality and (b) periods when rail or highway traffic is passing the dense line. Note the relative dominance of the faster body-wave arrivals for the train periods compared to the surface-wave dominance for the quiet periods. These CMP gathers are computed by distance binning (20mbins) and stacking CCFs for all station pairs along the dense line that share a CMP. All panels use the same midpoint, which is near the center of the dense line. The CCFs are arranged so causal lag-times correspond to energy traveling up the array from southwest to northeast. Modified from Dales et al. (2020).

3.3 Seismic interferometry with opportune sources

Standard noise-correlation workflow typically removes strong transient events such as earthquakes and then correlates the entire remaining time series recorded at different sensors (Bensen et al., 2007). With opportune sources, including train traffic, wepropose a novel workflow. It includes source characterization with signal and station pair selections as alternatives to blind correlation. We thus aim to improve the signal-to-noise ratio (SNR) of the reconstructed correlation functions and the temporal resolution of monitoring studies. Each step of this approach is illustrated in the following sections. Figure 3.3 summarizes the five main stages of our data processing in comparison to the classical method of continuous blind data correlation.

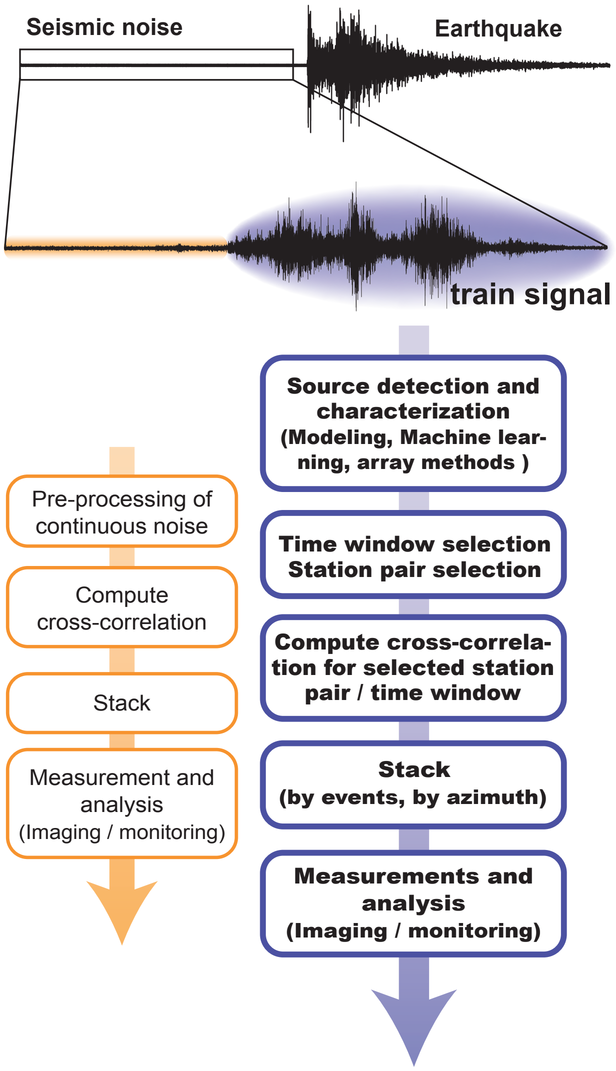

Figure 3.3: Chart illustrating the processing steps for opportune sources (in blue) compared with the standard ambient noise correlation workflow (in orange). Modified from Pinzon-Rincon et al. (2021)

The workflow's steps:

- Identify opportune source signatures in the continuous data and, if possible, locate these sources perhaps by distance but at least in azimuth. As shown in chapter 2-section 2.4, the modeling of opportune sources helps reveal the temporal and spectral content of the generated wavefield. Standard (short-time average window / long-time average window) and more advanced techniques (e.g., Meng et al., 2019; Kong et al., 2019) detect these transient events. Array processing techniques (e.g., Cheng et al., 2020) can be used to locate their sources. For example, figure 3.1-a illustrates two train positions at different times (red and blue stars) and the associated selected stations for pair-wise correlations (red and blue dots).
- Station pair selection: with source location estimates in mind we can narrow down the options for station pairs. For a given signal time window we use only station pairs for which the train source is in a stationary phase zone. During a train passage, the energy carried by its seismic signal reaches an array of sensors from a range of directions.
- Compute cross-correlations after proper time windowing and station pairs selection.
- Measurement and analysis: depending on the type of studies, various approaches such as travel-time measurements can enhance imaging and monitoring applications.

This workflow could be extended to any kind of seismic tremors, natural or human origin. It should help to extract high-frequency energy between well-selected pairs of stations useful for imaging and monitoring purposes. In the following sections, we will develop each step of this method using train tremors in Marathon. During a train passage, the energy emitted by the train illuminates the array from different directions depending on the train's position. We will study all these different paths separately to obtain additional information about the medium, retrieve high-frequency energy and increase the data quality.

3.3.1 Source detection and characterisation

The first step of the workflow presented beforehand is the source characterization. Modeling proposed by Lavoué et al. (2020) showed that characteristic lengths of the train system (train and railway) modulate the spectral signatures of train signals. They show that trains produce broadband signals between [1 - 50]Hz, and long, heavy, and fast trains generate more noise. The results were reproducible, with observations at Marathon, where trains up to 4km long and

18000 tons weight travel near the studied zone. Marathon represents the perfect configuration to use train signals as an opportune source of noise for passive seismic interferometry. More details about train-generated signals from a seismological perspective are described in chapter 2 section 2.4.

Then the train schedules are needed to separate the train periods from the rest of the seismic records. We contacted the Canadian Pacific railway company, the transportation company in charge of the railway in Marathon, to have the information of train passages at Marathon. But, after several unsuccessful attempts, they weren't allowed to share the information with an externals person for security reasons. Finally, seismological methods had to be used to analyze to detect the train tremors and create a catalog.

In order to scan the complete noise records efficiently (1018 nodes one single-component continually recording during 30 days, sampled at 50Hz), weapplied the covariance matrix method proposed by Seydoux et al. (2015). This method uses the spatial coherence of the wavefield to detect in time and frequency the emergent signals embedded in the continuous seismic noise.

Seydoux et al. (2015) defined the Fourier spectra of seismic records as $[u_1(f,t), u_2(f,t), .., u_N(f,t)]$ where N is the number seismic station, f the frequency and t the time. The covariance matrix is defined as the average of the cross-spectral matrix:

$$C(f,t) = \langle u(f,t)u^\dagger(f,t)\rangle_{\Delta t} = \frac{1}{M}\sum_{m=0}^{M-1} u(f,t + m\delta t/2)u^\dagger(f,t + m\delta t/2) \tag{3.1}$$

where † denotes Hermitian transpose. M represent the overlapping data time segments of duration δt. The matrix can be decomposed on the basis of eigenvectors associated with its eigenvalues $\lambda_i(f,t)$.

Seydoux et al. (2015) showed that the uncorrelated noise eigenvalue spectrum decays steadily. It slowly converges to a flat distribution with an increasing number of subwindows. They sort eigenvalues in descending order to obtain the covariance matrix spectrum and define the covariance matrix spectral width (typical width of distribution of the eigenvalues $\lambda_i(f,t)$) as:

$$\sigma(f,t) = \frac{\sum_{i=0}^{N}(i-1)\lambda_i(f,t)}{\sum_{i=0}^{N}\lambda_i(f,t)} \tag{3.2}$$

In this approach, the spectral width is a proxy for the number of independent sources of noise. As shown in figure 3.4-b, the spectral width is high for incoherent seismic noise. On the contrary, a low spectra width value represents a spatially coherent signal produced by a single localized source. For example, figure 3.4-c shows a signal that corresponds to an earthquake

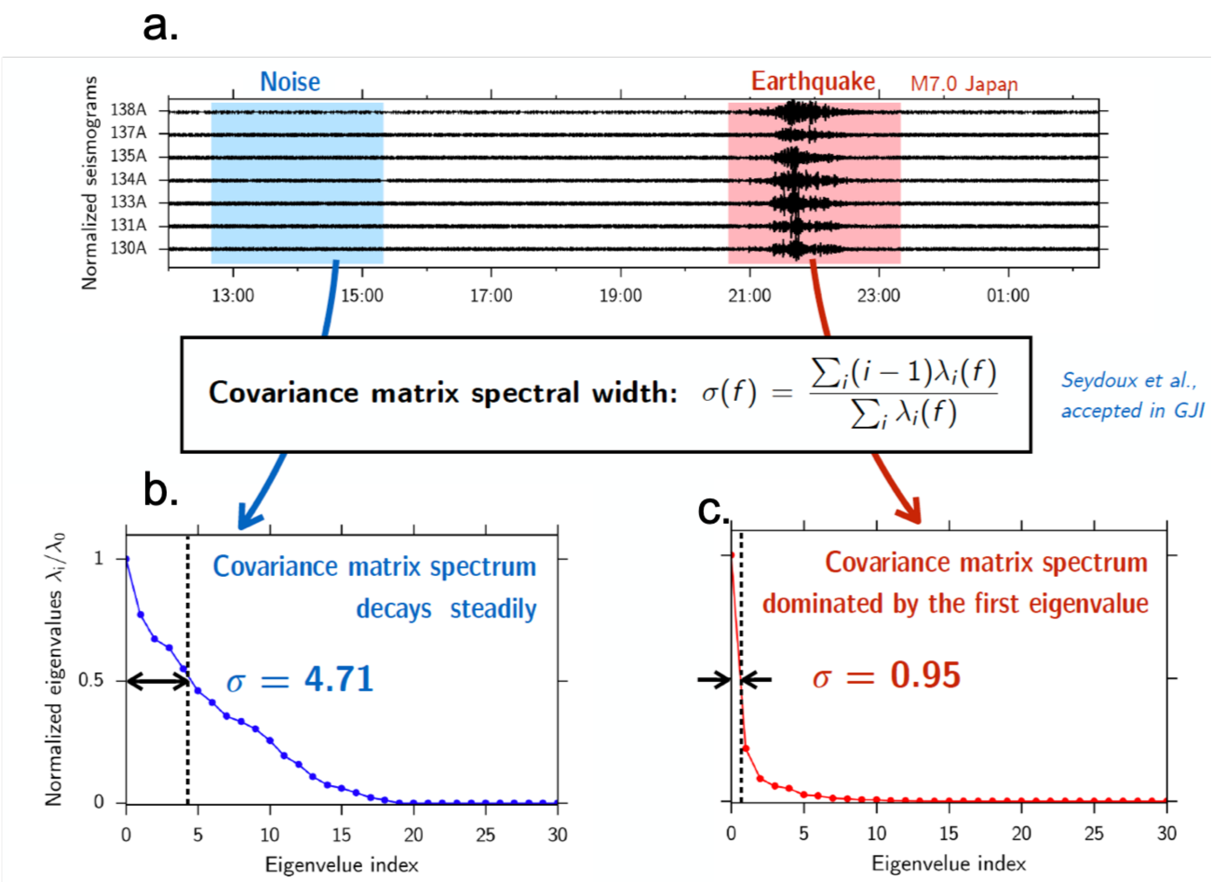

Figure 3.4: Example of covariance matrix spectra computed from real records a) Covariance matrices were computed at f = 0.07Hz. b) Covariance matrix spectrum for noise. c) Covariance matrix spectrum for earthquake signal. Values of the covariance matrix spectral width σ are indicated with vertical dashed lines. Modified from Seydoux et al. (2015).

(i.e., a single source).

The covariance matrix method was applied to Marathon dataset to detect the train signals. Thirty days (from September 22 to October 22, 2018) were processed using only the geophones in the sparse array (figure 2.2) to decrease the computation time. The data were processed daily as follows:

- Parameterization: the selection parameters M and δt are base on a trade-off between a statistically robust estimation (M) and a good time resolution (δt). δt is linked to the lowest frequency-resolved. Here, the higher period is 0.5s. Hence, we can use 3s-long subwindows to have more than two signal oscillations in every subwindow. High values of M estimate more robustly the covariance matrix. To reasonably estimate the covariance matrix while keeping a maximal temporal resolution. We used 10 overlapping time windows.

- Pre-processing: spectral whitening, temporal normalization were applied to each overlapping time window to compensate for the non-stationary of seismic signals and to disregard any amplitude information as Seydoux et al. (2015) advise. Data is also filtered in the train's frequency band [2 - 23]Hz.

- Covariance matrix calculation: first the Fourier transform and then the cross-spectral matrices are computed. Following the equation 3.1 the covariance matrix is calculated by averaging over a set of M (M=10).

- Matrix decomposition: the covariance matrix is decomposed in eigenvalues values, and their distribution (the spectral width) is computed as a function of frequency and time as described in the equation 3.2.

- Detection: The spectral width is used as a detection parameter (Seydoux et al., 2016). First, it is averaged in frequency. We averaged between [2 - 20]Hz (train characteristic frequencies) to decrease the other sources' contributions, especially low-frequency sources. Then, a spectral width detection threshold for the detection is established. The spectral width tends to 1 when one single source dominates the records. However, trains are moving complex sources with a characteristic spectral width >2.5. Therefore, we determine a detection threshold of 4 for the spectral width (red line figure 3.5-b) all periods with a spectral width higher than 4 are rejected. This value is chosen based on the median averaged width; the threshold is lower than the median averaged to reject background noise but great enough to avoid missing trains with a high spectral width. Trains are not punctual sources; thus, not only intervals below the detection threshold are retained but with a duration longer than 15 minutes (figure 3.5-b pink boxes).

In figure 3.5, an example of one-day record, is shown. The dashed rectangles are periods where the spectral width decreases between 2Hz to 25Hz (indicating an increase of the wavefield coherency). Thus these periods correspond to train passages. We detect eight events (i.e., train passages) using frequency averaged spectral width (figure 3.5-b). Figure 3.5-a shows other periods with low spectral width. At low frequency (<3Hz), these low values are related to the noise produced by the Superior lake, and the periods between [4 - 10]Hz are originated by the highway traffic. By applying this detection method to the entire data set day by day, we detected the passage of 207 trains over the 30 days of recording.

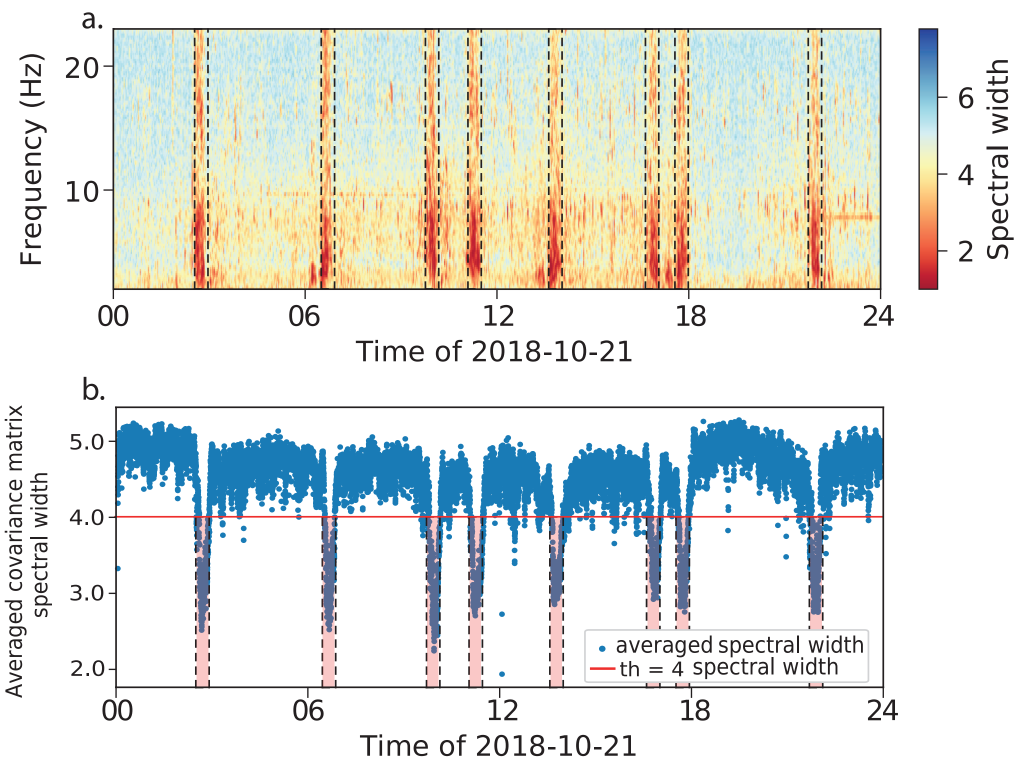

Figure 3.5: Example of one-day train detection. a) Covariance matrix spectral width as a function of time and frequency for 1-day-long data filtered between 2-23 Hz. The color scale indicates the spectral width. The dashed rectangles correspond to a train passage b) Frequency-average of the covariance matrix spectral width from panel (a) as a function of time. The red line is the detection threshold corresponding to a spectral width of 4 for train passages, and colored segments indicate the detection periods.

3.3.2 Time window selection and station-pair selection

The catalog created with the covariance matrix provides periods where the trains are passing near the array. However, to apply the workflow presented in section 3.3 a more precise time detection is needed to localize the train. To explore the type of seismic waves present during train passages, a frequency-wave-number (f-k) beamforming analysis was performed, using the method described in chapter 2, equation 2.2.

Beamforming analysis aims to localize the sources as much precise as possible, simulating a source in active seismic exploration. The active source of seismic waves, either explosions, air guns, or seismic vibrator (vibroseis), are located at a precise and known position during the acquisition. But since that trains are moving and extended sources, their localization is an emergent problem. Therefore, the time window length is the principal concern. If the window is too long, the source is not located in a precise position anymore, and it can not be considered and processed as described before; but if the window is too short, the signal can not propagate enough to do passive seismic interferometry.

In considering this trade-off, 1-minute-long windowing was applied. Assuming an average train speed of 90km/s, the train travels approximately 1.5km per window; freight trains also have a minimum length of 1km. Considering these characteristics, the train travels 3km per window. Hence, we can assure that the train is in the stationary phase zone for all the window lengths and applied workflow.

First, all events (i.e., train passages) were divided into 1-minute-long windows. Then, f-k beamforming was performed for each time window individually at the frequency range of [8 - 18]Hz and without overlapping. We decided to use the only station in the sparse array (figure 2.2) to decrease the computation time while keeping a good resolution in frequency and space (inter-station distance and aperture, respectively). We decided to apply first the covariance matrix to the full record and then the beamforming only to train records to reduce the computation time. It's important to note that scan all the data using beamforming methods was too expensive.

Figure 3.6 shows beamforming results by the minute for one train. The beam amplitude is plotted in the slowness domain. We observe several lobes of coherent energy moving (warms colors) from west to east in the function of time. These lobes represent a specific coherent arrival. We distinguished three different arrivals: one at 2.9km/s that corresponds to a surface-wave, a second one at 3.8km/s that we suggest corresponds to the S-wave, and finally, the lobe at 7km/h that correspond to the P-wave and is not clear. The P-wave amplitude varies notably with time, to the point of not being visible for certain time windows.

Beamforming analysis provided additional information about the type of waves that trains produce, their velocities, and their positions. According to this new information, two new characteristics were added to the catalog: the train direction and the number of simultaneous passages. The train direction is inferred using the displacement of the surface wave during the train passage. Two direction are defined for the rest of the study: **WE** (West to Est, as in the fig 3.6) or **EW** (Est to West). This method enables us to have the number of simultaneous train passages as well. When two trains pass simultaneously, the coherent energy of one wave arriving at the array comes from two different directions. Such simultaneous train passages are also detected and added to the catalog.

A final catalog including the train detection, direction, and simulations passages is generated. Figure 3.7 shows the results of the final train catalog. Two hundred seven train passages were detected, but only 180 events were retained for this study after skipping overlapping trains and bad quality trains (during too-noise periods and/or non-energetic trains).

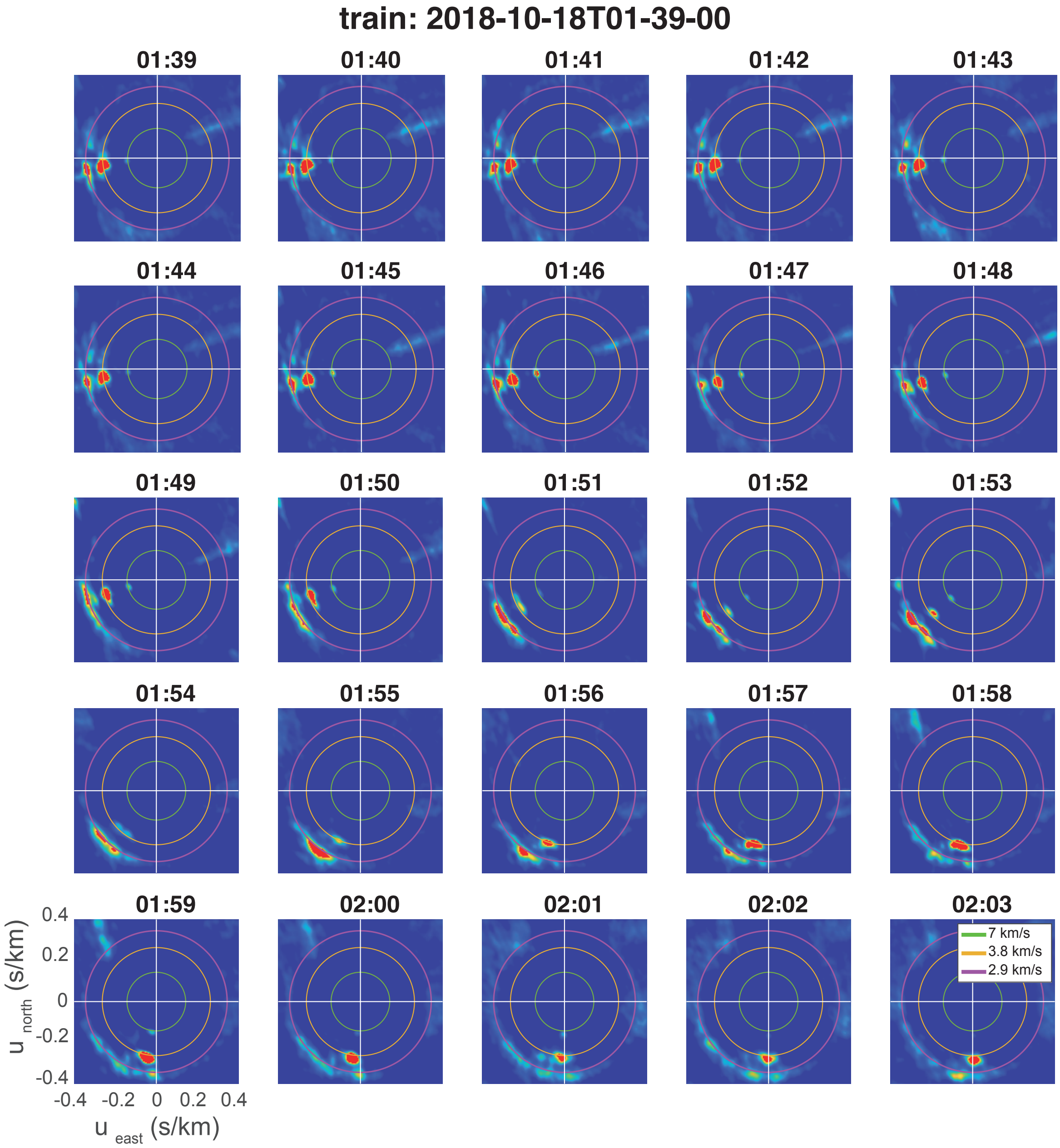

Figure 3.6: Beamforming analysis during a train passage: amplitude spectra in the horizontal slowness domain computed between [8 - 18]Hz. Each panel represents the beamforming result for one minute window of a train signal. Greens circles represent a velocity of 7km/s, the orange circles a velocity of 3.8km/s, and violet ones a velocity of 2.9km/s, which corresponds to velocities of P, S, and surface waves, respectively. Warm colors represent high amplitudes and so a detection.

To determine the train position, we used the 1-minute-long windows beamforming computed before. We assumed that the primary energy source was the train and the maximum beam power generated by them. Therefore, we pick the maximum beam power of the S-wave of each beamforming panel to infer their azimuth. To do the picking, we applied three different approaches in order to evaluate the accuracy of the picking and separating the events by direction.

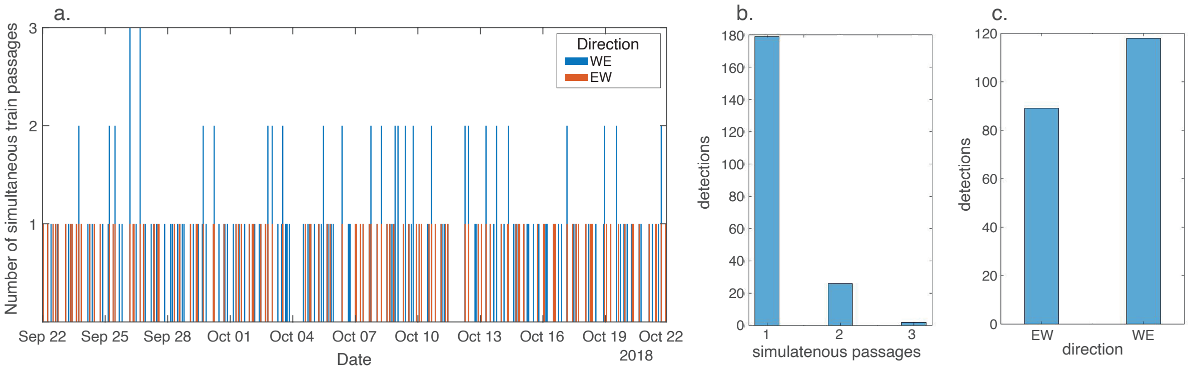

Figure 3.7: Train catalog. a) Train passages detection in time during one month. Each bar represents one event. The color describes the direction and the bar's height the number of trains passing simultaneously. b) Histogram of simultaneous train passages. b) Histogram of train direction.

First, we chose three events randomly per direction, and we manually pick the maximum as shown in the figure 3.8 by the green lines. The dashed lines are the averages of each of the three events. From this preliminary analysis we concluded that the azimuth is similar for each event at each minute. Then, we applied the second method, an automatic picking of the maximum (for each train at each minute) to examine the entire catalog. Figure 3.8 light blue and red lines show the averaged automatic picking and the associated error bars. For both directions (WE and EW), the uncertainties are low when trains are close to the array (minutes 8 - 25) but increase for the earlier and the later times. We observe an odd behavior for the azimuth at the south ($< 220°$). WE trains are supposed to travel from higher azimuths (West) to lower azimuths (Est). However, after 25 minutes, the azimuth increase. The decrease can be translated as a ”train turn back” that is not logical. The same behavior is observed for EW trains at the first minutes when trains are at the est. This error comes from the back azimuth beamforming uncertainties and because the train is farther from the array at the south est and the energy arriving is lower. A constrained automatic picking is applied based on the averaged azimuth resulting from the second method to improve the azimuth measurement. The maximum is only searched in the surroundings azimuths ($\pm 20°$ averaged azimuth). Figure 3.8 dark blue and red lines show the results of the final method. We observed a decrease in of the uncertainties proving an enhancement of the azimuth measurement. This proved that the azimuths have closer values for different trains, and thus It is possible to stack all the events by azimuth.

However, the last 5 minutes of the WE trains and the first ten minutes of the EW trains

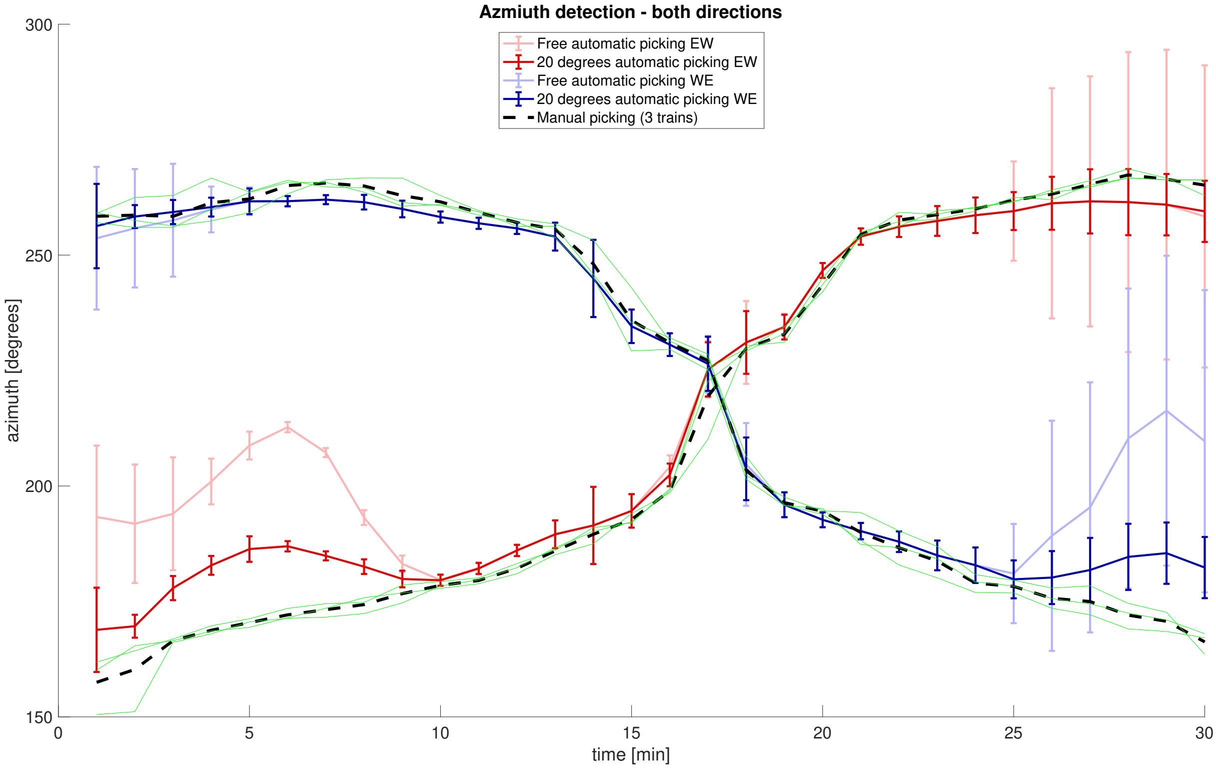

Figure 3.8: Train azimuth picking in function of time. Green lines are the manually picking of 3 events and the dashed lines the associated averaged. Light blue and red lines are the free automatic picking averaged with the error bar. Dark blue and red lines are constrained automatic picking averaged (using ±20° constrain)

continue to show irregular behavior. Therefore, we decide not to use these minutes. For the trains going WE 25 minutes are used and for the EW trains 20 minutes. For the following section, WE and EW events are processed separately. Only WE the trains going to be exposed to facilitate the explanations. Nerveless, the same workflow is applied for EW trains, and similar results are obtained.

Then having the railway coordinates and the azimuth by minute, we back-project the signal onto the railway (figure 3.9-3 red crosses), which gives us the position of the train by minute. When the train is on the western part of the railway, the back-projection cannot rely exclusively on the azimuthal information derived from beamforming. Instead, we assume an average train speed (based on Canadian train regulation) to locate the train. Figure 3.9 show the example of 3 location inferred from the beamforming panels next to them.

To clarify the terms, one minute of train record is equivalent to an azimuth (infer from beamforming) and to a train position (infer from the azimuth):

$$\textbf{1 minute} \iff \textbf{azimuth} \iff \textbf{train position}$$

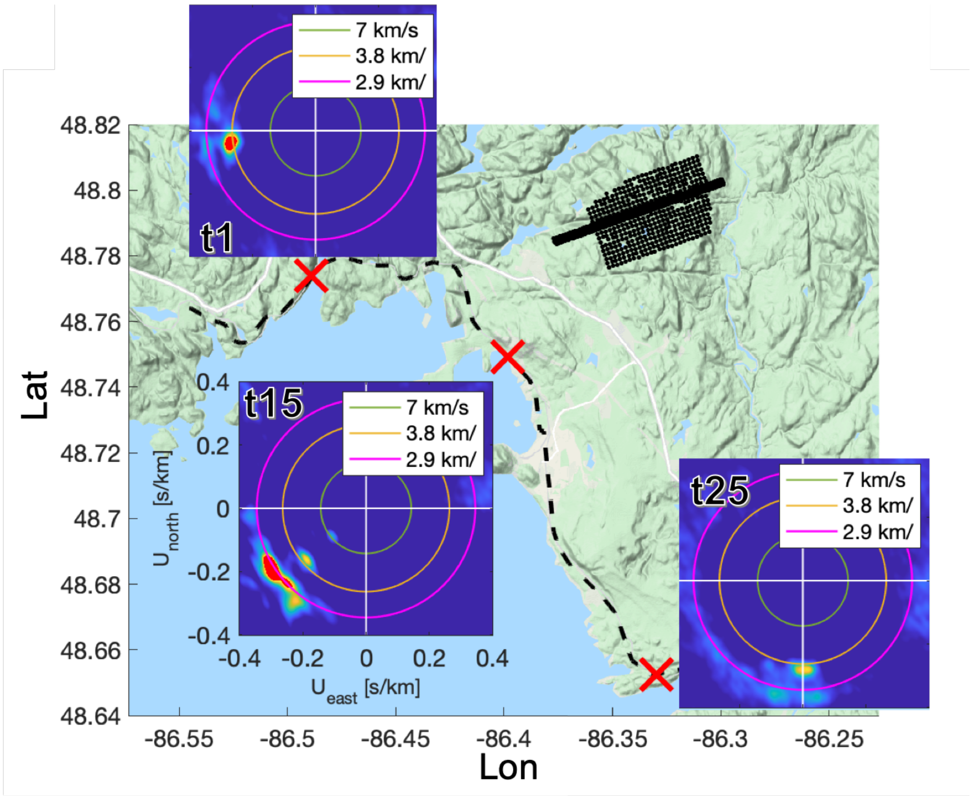

Figure 3.9: Train location schematization. Map of the study zone, gray dots are the seismic stations, the black-dashed line is the railroad, and red crosses are three examples of train positions with the corresponding beamforming panel. The position is computed by back-projecting the maximum beam power onto the railway

Next, we selected station pairs that will be used for the correlations based on train position. For this purpose, we assume a train length of 1km and the azimuth extracted from beamforming as the train's center for all events. Train length is quite conservative (freight trains length in Marathon is up to 4km), but allow us to stay in the stationary phase zone. Then, we select a segment of the railway equivalent to a train of 1km long, and we choose all the stations pair in line with this segment. Figure 3.10 lower left panels to illustrate three train positions (red line) at different minutes (i.e., azimuth) and the associated selected stations for pair-wise correlations (blue dots) for one single source (red dot at the center of the array). Figure 3.10 right panel shows the station selected during all the passages for the virtual source at the center of the array (red dot).

The disadvantage of the station pair selection is that not all the station pairs are used. For example, at Marathon 519690, station pairs are possible $(N*(N/2))$ with N: number of sensors, but only $\approx$ 5.7% are retained. Figures 3.11-a and 3.11-b show the number of station pairs selected and the percent of station pairs selected regarding the total possible station pair. We

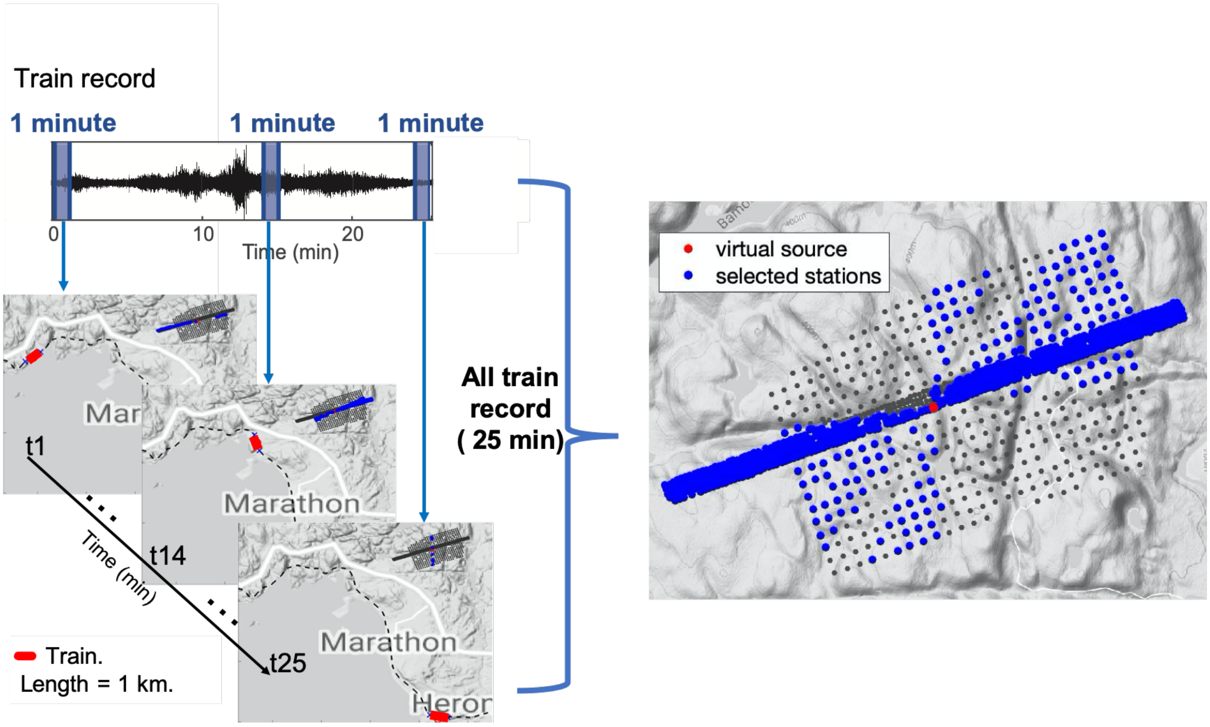

Figure 3.10: Chart illustrating the station selection reconstructed in time. Upper left panel: seismic train record. Lower left panels: map of the station selected by azimuth (i.e., minute) in function of the train position. The red lines represent the 1km train, the grey dots are the seismic stations, and the blue dots station selected for each position. Right figure: final map of all the stations chosen for the entire record (i.e., 25 minutes) for one virtual source (red dot).

also observed a lack of station in the direction Northwest to Southeast as shown in figure 3.11-c by the stations colored in blue (< 300 stations selected). The lack of selected stations is due to the railway position with respect to the array and the poor path illumination in this direction.

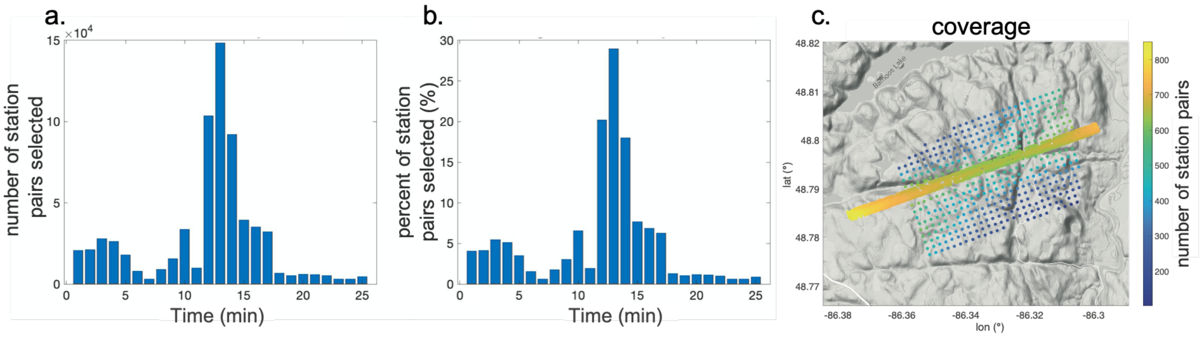

Figure 3.11: a) Number of the station pair selected by minute b) Percent of station pair selected regarding the total number of station pair c) Coverage array map. The color bar represents the number of times that a station will be used for the cross-correlation computation.

3.3.3 Cross-correlation

The following step is the cross-correlation computation for the selected time windows and station pairs. We computed a power-normalized cross-correlation (or cross-coherence) in the frequency domain, given by:

$$C(B, A, \omega) = \sum_t \frac{u_t(B, \omega) u_t^*(A, \omega)}{|u_t(B, \omega)||u_t(A, \omega)|} \tag{3.3}$$

Where, $u_t(AorB, \omega)$ is the seismic noise at the station A or B at the time t in the frequency domain ω. $*$ is the complex conjugate. $C(B, A, \omega)$ is the cross-coherence between sensor A and sensor B that provided the approximated Greens Function between A and B.

Only the train periods were used for the cross-correlation computation. For the previous steps, data sampled at 50Hz was employed; however, as shown before, trains generated energy between [3 - 40]Hz in Marathon, and the body-waves retrieved by Dales et al. (2020) are between [10 - 60]Hz. Thus, train signals are extracted from the raw data at a sampling frequency of 125Hz.

Then, 67 train passages of 25 minutes were divided into a 1-minute window without overlapping and correlated using cross-coherence (equation 3.3). Each event was processed separately, taking the station closer to the train position as the virtual source. Only causal lag times will correspond to energy traveling from the train to the station pairs, and only them will be used for the rest of the study.

Figure 3.12-a shows an example of a single broadband cross-correlation function between two sensors located in the dense-line and separated by 2.5km. Figure 3.12-c shows its spectrum. The frequency content, mainly between 1 to 30Hz, was expected due to the train tremor frequency content. Figures 3.12-b and 3.12-d show the filtered cross-correlation in a low-frequency range [1 - 15]Hz and high-frequency range [15 - 30]Hz, respectively. At low frequency, the surface wave arrival is retrieved. At high frequency, the P-wave is observed, and a secondary arrival is retrieved. This arrival could be a higher mode of a surface wave or an S-wave.

3.3.4 Stack

After the cross-coherence computation, single cross-correlations are stacked to increase the data quality. Since trains are repetitive noise sources, we can benefit by stacking trains at the same position (i.e., same azimuth).

First, we stacked cross-correlations according to their inter-station distances and collected them in distance-binned correlation gathers for the selected station pairs at one azimuth. Then,

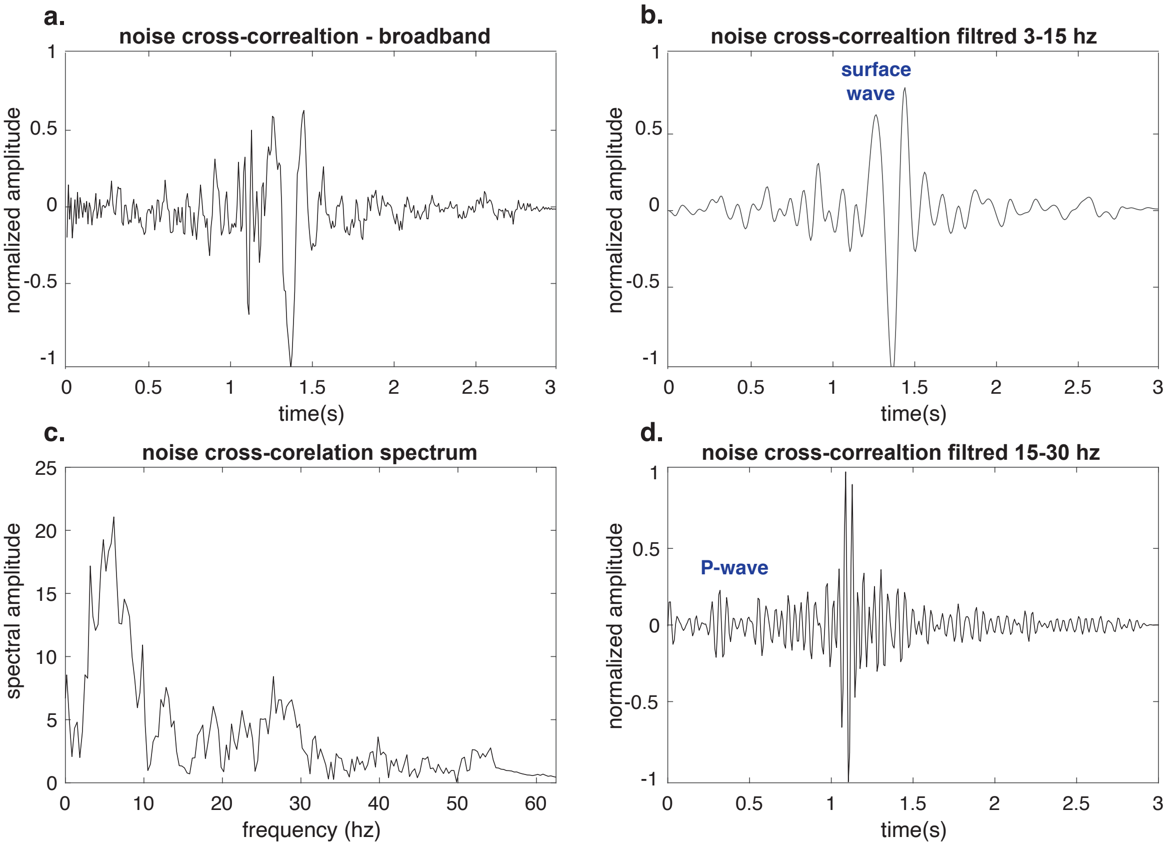

Figure 3.12: Examples of noise cross-correlation at inter-station distance of 2.5km. a) Cross-coherence computed using 1-minute-long window. c) Corresponding noise correlation spectrum, b) Cross-correlation filtered between [1 - 15]Hz, a surface wave is observed. d) Cross-correlation filtered between [15 - 30]Hz, a P-wave is retrieved.

we stacked windows whereby the train was in the same position (i.e., same azimuth). To do so, we used six train passages. The events were selected during periods with low background noise and high train tremor amplitude.

Figure 3.13-c and 3.13-d show the distance-binned stacked section over six trains passages at 2 train positions. Three dominant arrivals are retrieved. We suggest that the first arrival (7km/s) is a P-wave, and the second one (2.8km/s) is probably a mix of S- and surface waves higher modes and the surface wave (2.9km/s). In contrast, Figure 3.13-b shows the stack of one-minute cross-correlation for a quiet period (i.e., non-train passage), highlighting the absence of coherent body-wave propagation for this rather high frequency range, a weak surface wave is barely perceived.

We observe that only one minute of data stacked over a few events (6 train passages) is necessary to retrieve body waves. Thus, by applying the workflow presented here, we converged on a stable reference. However, the final goal is to image the subsurface in the region, and this

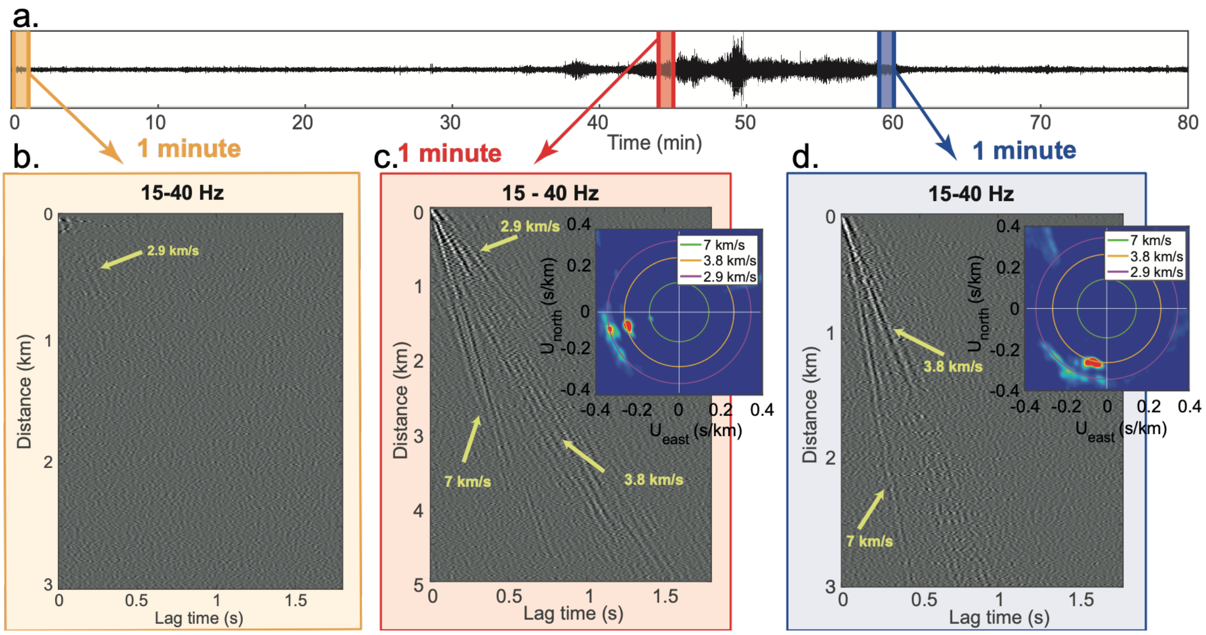

Figure 3.13: Example of distance-binned gather. a) Train seismic record. b) 1 min cross-correlation for a quiet period (i.e, non-train passages). c) - d) Stacked section over six trains passages at 2 train positions, using one-minute data segments.

stacked gather does not allow us to have 3D information about the medium. Therefore, looking more finely at the respective virtual sources gathers is needed to extract the ballistic wave arrival times. For this, we stack all WE train passages (62 events) for each station.

To analyze the stack convergence, SNR is measured in the function of stacked number trains was done. We define the SNR as the peak amplitude in the signal window divided by the standard deviation amplitude in the noise window. First, we compute the P-wave SRN for each event separately. Figures 3.14-a and 3.14-b show the SNR distribution of all trains passages for 2 virtual sources. The quality of all trains is similar, the SNR distribution varies between 1 and 4, but there is no evident difference between the events. Then, we stacked the SNR distribution in chronological order and measured the SNR. (figure 3.14-b and -c). We randomly stacked the trains to avoid the artifacts created with a linear order for the stacking. We repeated the random stack 300 times, measured the SNR, and averaged it. Results are shown in figures 3.14 c and f. Similar behavior is remarkable for the stacked distribution; we show that the stacking order does not affect the SNR distribution.

Typically, cross-correlation SNR grows approximately as the square root of the time series. SNR of train cross-correlation stabilizes after a stack of approximately 15 events. Given that similar signals are correlated and stacked, the SNR of trains correlations does not follow the same standard noise correlation SNR behavior. The same arrivals are summed, and a the SNR is fast converged. The addition of new time series does not make a large difference. However, we decided to use the complete stacked data for the subsequent studies to improve the quality of the data as much as possible.

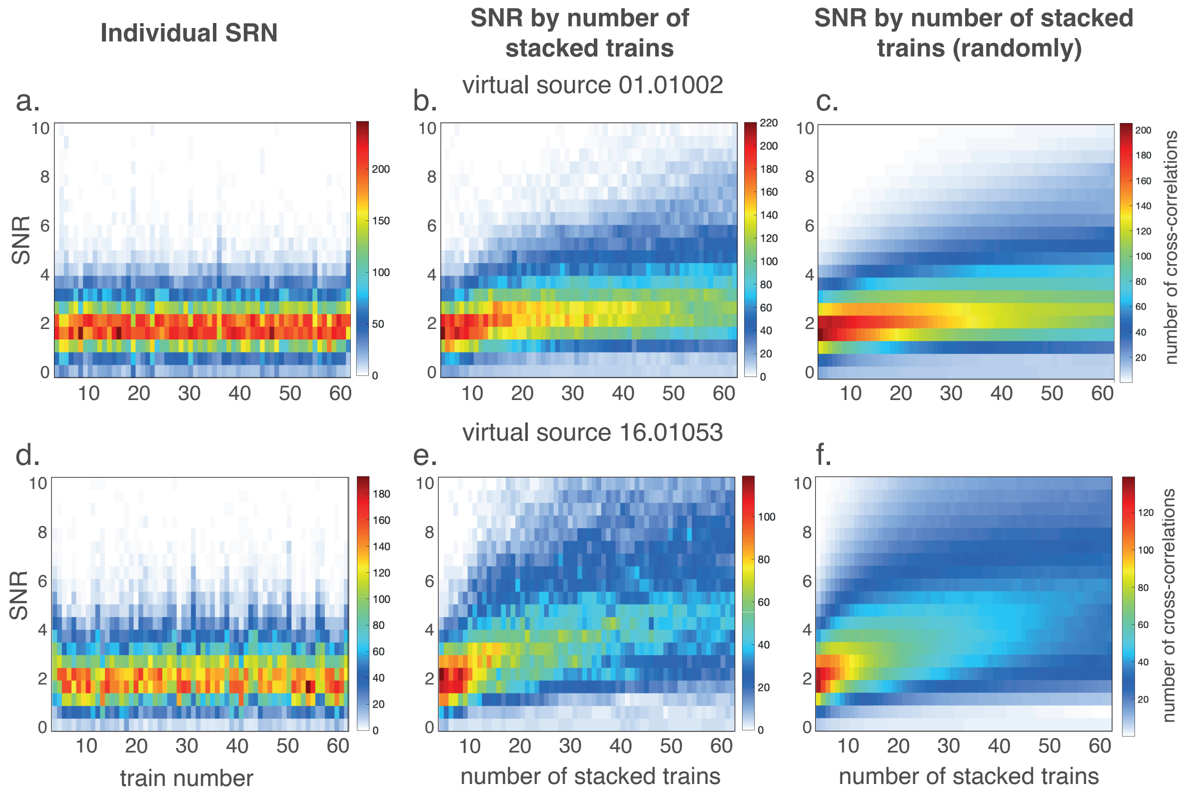

Figure 3.14: SNR analyse in function of stacked trains. The fist row: virtual source station 01.01002 (located at West of the dense-line). The second row: virtual source station 16.01054 (located at the south of the spare array). a)-d) Individual SNR by train passage. b)-e) SNR by number of stacked events. c)-f) SNR by number of stacked events (randomly stacked).

3.3.5 Measurement and analysis

In this section, we will only show the waves retrieved for virtual sources gathers after the workflow application. The measurement an analysis process for imaging will be developed in detailed in the chapter 4.

After the stacking step, the virtual source gathers by azimuth showed several arrivals in the function of the frequency range studied. However, to perform 3D imaging, singles virtual shot gathers are necessary. These gather are created by stacking train passages (i.e., events) that illuminates the same station pair.

Figure 3.15 shows the gathers of the virtual source located at the center of the array (as shown in the figure 3.10, red dot). The gather a is filter between [2 - 15]Hz after stack. A clear

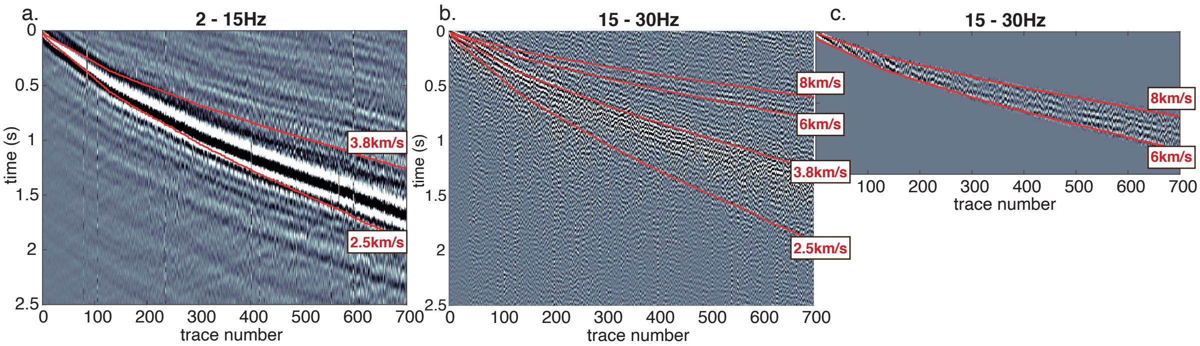

Figure 3.15: Virtual shots gather constructed from the stack of 62 train passages. Trace numbers are sorted by increasing distance from the virtual source with a not constant step. The red lines designate travel times with constant velocities, as indicated in the withe boxes. a) Gather filtered between [3 - 15]Hz, a clear surface wave propagates. b) Gather filtered between [15 - 30]Hz, two arrivals are retrieved an S-wave or higher mode of the surface wave and a P-wave. c) Gather b muted between 6–8km/s.

surface wave propagates between 2.5 and 2.8km. The seismic source ringing generates parallel arrivals that do not pass by zero. Even if the train is located in short time windows, it is still a moving source that continuously generates noise. Figure 3.15-b shows the gather filtered in a high-frequency range, [5 - 30]Hz, to avoid surface waves. Below 15Hz, only surface-wave arrivals are seen. We retrieved two dominant arrivals with a velocity of 3.8km/s and 7km/s. There are uncertainties, but we suggest that the first arrival is a P-wave, and the second one is probably a mix of S- and surface waves. To highlight the P-wave, the gather is muted between [6 – 8]km/s (Fig. 3.15-c). The P-wave is propagating especially at long distances.

3.4 Computational considerations

The particularities of the signal processing carried out in each step are explained in the previous sections, but the computational characteristics have not been discussed yet. The computational cost and the effective data storage were not negligible and became a critical issue in the thesis due to the large volumes of data and computations.

The raw data was composed of 30 days of continuous records for 1018 nodes. It was unnecessary to use data with a high sampling frequency to inspect it, knowing that seismic noise is generally low frequency. We decided to download the raw data and downsampled it at 50Hz. The dataset was organized by daily files and entirely represented precisely 50GB.

The 50Hz sampled data allow us to compute the first steps of the analysis correctly (spectrograms, beamforming, and covariance matrix) while keeping relatively fast computation and decreasing the storage memory used. The sparse array was used to applied array methods (beamforming and covariance matrix analysis) for the same computational reason for the array methods. The covariance matrix was not an expensive computation; however, to accelerate the computation, the covariance matrix was computed using daily data, and it was calculated in parallel using 5 CPUs per computation. The computation time of each computation was 4 hours, and in total, 60 CPUs were needed generating 927Gb of data.

The beamforming was more expensive than the covariance matrix; for this reason, the 1-minute window beamforming was only applied to train data detected with covariance matrix sampled at 50Hz and representing 166G. The computation time for one 1-minute window was $\sim$25 min using one CPU.

207 trains of 30 minutes were detected with the covariance matrix, and 1-minute window beamforming was applied. As a result, 6210 windows were beamformed for a total computational time of $\sim$117 days ($\sim$168300 minutes), generating 812Gb of data. Parallel computing made it possible to reduce the computation time. However, the number of CPUs that could be used simultaneously varied depending on the availability of the laboratory's computation server, so this step took a considerable amount of time.

Train modeling showed that the frequency content of train signals is high (up to 60Hz). Therefore, to perform the cross-correlation, we used data sampled at 125Hz, increasing the dataset size and computational time. The train dataset after a proper selection (180 retained events) was extracted, representing 150Gb.

The cross-correlations were parallel computed by 64CPUs using pycorr SOFTWARE pycorr, and the cross-correlation (for selected station and time window) and stacked dataset per station equates to 421G. However, it is important to note that the correlation computed represents only 5% of the total possible station pairs.

Most of the computations presented here were performed using the GRICAD an HPC platform, and associated storage infrastructure (`https://gricad.univ-grenoble-alpes.fr`), which is supported by Grenoble research communities.

3.5 Conclusion

In this chapter, we showed how we can use massive freight train noise recovery to retrieved high-frequency waves between sensors in regions neighboring railroads. Furthermore, this approach of noise recovery to create valuable virtual sources could be applied for several other seismic noise sources and in different contexts improving spatial and temporal resolution.

The application of this method, specifically in Marathon, confirmed its effectiveness in extracting relatively high-frequency energy (for typical seismic noise applications) from train signals. To quantify the improvement in arrival retrieving, we compared virtual source gathers from standard cross-correlation with the method developed here. First, blind correlations were computed following the standard ambient noise interferometry workflow (figure 3.3 orange workflow). Then, the continuous data (30 days, 1018 nodes) was cut into 30 minute long segments, and the cross-coherence (equation 3.3) was computed without overlap. Finally, the cross-correlations are stacked, and the causal and acausal parts are averaged.

Figure 3.16-a and 3.16-c show the virtual source gathers reconstructed from train periods and from blind correlations respectively for a virtual source located at the center of the array (figure 3.10 red dot). The gathers are filtered between [3 - 15]Hz, and the Rayleigh wave propagation along distance (increasing with the trace number) is clear in both cases. The blind correlation has more data than train gather because all the virtual receivers are used. However, the quality of the train data is higher, figure 3.16-a shows the SNR distribution for both gathers; the blue for the blind correlations and the violet for the train correlations. These distributions show the improvement of the data quality with our method almost doubling the SNR median.

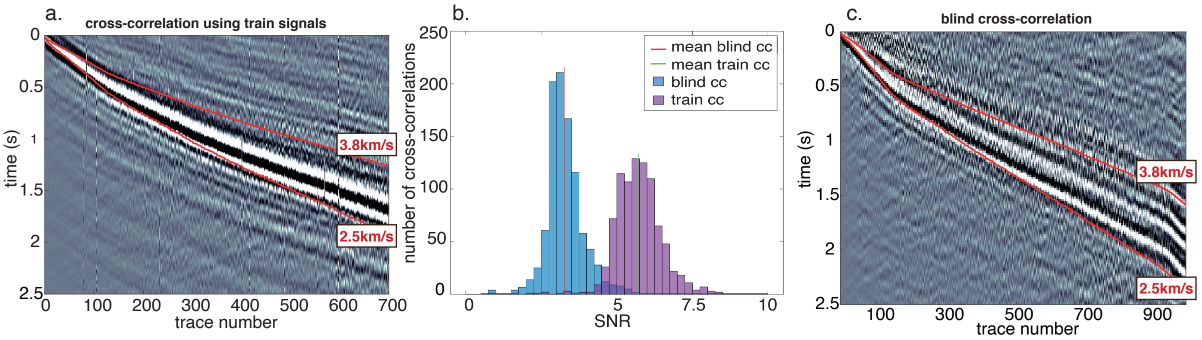

Figure 3.16: Virtual shots gather reconstructed from a) the stack of 62 train passages and c) from blind cross-correlation of one month of continuous records. Trace numbers are sorted by increasing distance from the virtual source with a not constant step. The gathers are filtered between [3 - 15]Hz and retrieve an evident surface wave. The red lines designate travel times with constant velocities, as indicated in the withe boxes. b) Histogram of SNR distribution in blue for the blind cross-correlation and violet for the train periods. The red line corresponds to the mean SNR value for blind correlations and the green line to the mean value for train periods.

Dales et al. (2020) showed in previous studies that body waves in Marathon could only be retrieved by selecting all time periods for which the ambient noise comes from a direction aligned with the W-E dense line of sensors (figure 3.2). This study is thus illustrative, and their results do not allow to perform 3D imaging. Here, we intend to move one step further by separating different noise azimuths for virtual source retrieval in different directions. The noise correlations sections might thus be of lower quality because single virtual source gathers are studied but ultimately may allow performing 3D imaging for both surface and body waves.

The next chapter will explore these arrivals by extracting travel times from the coherent wave-forms to apply seismic noise-based imaging in this area. With the aim of imaging different geological structures

Chapter 4

Surface-wave imaging at Marathon deposit using train signals

Contents

4.1 Introduction

We showed in the previous chapter that it is possible to retrieve promising coherent arrivals by selectively cross-correlation of train signals. However, the utilization of these arrivals for imaging purposes is not straightforward because we are somewhere between two propagation regimes that have very different response to seismic interferometry: the passive seismic interferometry and the interferometry of active and controlled sources. In the case of the passive seismic interferometry, it has been shown that the cross-correlation of diffuse wavefields two between the stations converges to the Green's function (e.g., Weaver and Lobkis, 2001; Snieder, 2004; Wapenaar, 2004). On the other hand, in active seismology, the explicit knowledge of the seismic source allows converging to Green's function by deconvolution of the signals source function. However, in our case, trains are large, extended, and not controlled sources; we are not in the case of a diffuse wavefield or a fully controlled source; thus, Green's function retrieval assumptions are break.

Despite the numerous studies showing the consequences of not converging to the Green's function, fundamental-mode surface waves and their travel times can be extracted from the cross-correlation function and be sufficiently relatable to the Green's function travel times for imaging applications (Fichtner and Tsai, 2019). Therefore, we decide to apply conventional surface waves tomography methods to image the subsurface on a local scale, assuming that the energy retrieved from train signals is close enough to the Green's Function and neglecting the effect generated by the propagation between the train and the first sensor.

Since the first tomography obtained with seismic noise (Shapiro and Campillo, 2004) until today, numerous studies have proven the efficacy to obtain velocity models from retrieved surface waves. The surface waves are sensitive to the S-wave structures at depth, depending on their frequency, and the inversion of these surface wave dispersion curves allow us to infer S-wave structures at depth which is a common problem in seismology (Dorman and Ewing, 1962).

At Marathon, the surface wave dominates the cross-correlation functions at low frequency (<10Hz). Though the body waves are present in the cross-correlation functions, it remains of weak amplitude and quality to use them for body-wave tomography; we decide to use surface waves to image the Marathon deposit.

In this chapter, we will present the imaging the implementation results of a depth inversion of surface waves obtained from train correlations. First, we measured Rayleigh wave phase velocities for each virtual source gather using the eikonal approach, from which we derive 2D phase velocity maps for frequencies in the range of [1.5 - 9.5]Hz by averaging the virtual source gather measurement. Next, we invert the phase velocity maps using the Neighborhood Algorithm inversion. In order to improve the depth inversion, we implement a second linear 1-D inversion taking as an initial model the resulting model of the Neighborhood Algorithm inversion. Finally, we obtained a 3D Vs model of Marathon deposit by inverting each cell's local 1-D Vs profile. To evaluate our approach, we compare the train cross-correlation with both synthetics and standard cross-correlations. To investigate Green's function convergence for train signals, we do numerical modeling in a simplified case of Marathon, to which we will also apply the Eikonal tomography.

On the other hand, we will use standard seismic noise correlations obtained from the same dataset to apply Eikonal tomography and analyze the anisotropy in the Marathon deposit. These comparisons will allow us to judge the limits of the train signals' cross-correlation workflow and the utilization of the eikonal tomography. Therefore, we will evaluate the eikonal tomography limitation using the synthetic data.

4.2 Eikonal tomography

The eikonal tomography is a surface wave tomography method proposed first by Lin et al. (2009) and then generalized by Lin and Ritzwoller (2011). It has been applied at regional scale (e.g., Lin and Ritzwoller, 2011; Qiu et al., 2019) and local scale (e.g., Mordret et al., 2013b; Lin et al., 2013a). The eikonal tomography is an array-based method which use wave-fronts tracking to directly and locally estimate phase velocities by applying the wave equation.

The basis of the method is the eikonal equation (Wielandt, 1993):

$$\frac{1}{c_i(\omega)^2} = |\nabla\tau(r_i, r)|^2 + \frac{\nabla^2 A_i(r)}{A_i(r)\omega^2} \tag{4.1}$$

Where: $\tau(r_i, r)$ is the travel time between a point r and the virtual source r_i. c is the phase velocity in function of the frequency ω and A is the spectral amplitude. When the amplitude varies sufficiently smoothly, or the frequency is high enough, the second term on the right of equation 4.1 is much smaller than the first term and it can be neglected (Lin et al., 2009). Thus, the phase velocity c can be described as:

$$\frac{1}{c_i(\omega)} \cong \nabla\tau(r_i, r) \tag{4.2}$$

Where, ∇ denotes the local gradient ($\nabla = \frac{\partial x}{\partial} + \frac{\partial y}{\partial}$). This equation can by applied to each individual virtual source i and the final velocity map is constructed by averaging the individual computations.

Lin et al. (2009) proposed a 3 step workflow to apply the eiknonal tomography:

1. **Phase travel time measurement:** computed across the array centered on the virtual source. A waveform selection based on a set of quality criteria is also performed to improve the measurement.
2. **Gradient computation:** travel time is interpolated on a regular grid and spatial gradient computed using the equation 4.2. The magnitude of the gradient is approximated to the local phase slowness. The direction of the gradient represents the direction of propagation of the geometrical ray.
3. **Statistical average:** for each spatial node in the grid the local phase speeds and wave path directions are compiled and averaged using all the virtual sources in the array.

4.3 Eikonal tomography implementation

The eikonal tomography was implemented to the virtual gathers reconstructed by selective cross-correlation and stack train tremors (detail in chapter 3). As shown in figure 4.1 the surface-wave propagates in the individual virtual sources gathers with good quality. The coherence of the wavefield is also evident in the figure 4.1-b. The figure shows snapshots of the wave-front propagating in the array away from the virtual source (red dot). We now seek to track this wave front separately for each virtual source.

4.3.1 Waveform selection and phase travel times measurement

Before measuring the phase travel time, the correct waveform has to be properly isolated. To do so, a preprocessing workflow is applied by virtual source.

First, the data is filtered, and the traces with an SNR $< SNR_{threshold}$ are rejected. The SNR is estimated as the ratio between the maximum absolute value in the window of interest and the noise's standard deviation (here, the signal after 2.5 seconds). Correlations with inter-station distances smaller than 0.5km are rejected to avoid near-source effects, and inter-station

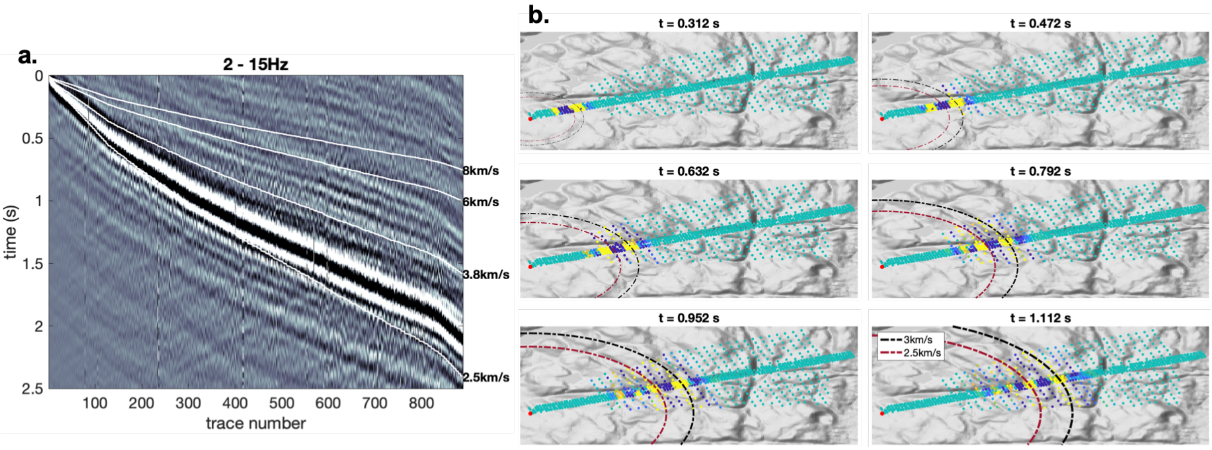

Figure 4.1: a) Virtual shot gather of station 01.01002, red dot panels b. The shot gather is reconstructed from the stack of 60 train passages using all the azimuths and filtered between [2 - 15]Hz. Trace numbers are sorted by increasing distance from the virtual source with a not constant step. The white lines indicate travel times with constant velocities as shown in the figure. b) Snapshots of the surface wave-front from the same virtual source at six different propagation times. The virtual source is the red dot in all panels. Data is filtered between [2 - 15]Hz. The red and black lines indicate the wave-front traveling at 2.5km/s and 3km/s.

distances larger than 6km are rejected as well to preserve the data quality. Figure 4.2 shows the workflow implementation for the virtual source 01.01002 at 4.5Hz. Panel 1 shows the filtered gather keeping only traces with SNR > 4.

Second, the gather is muted around the targeted wave. The move-out muted window is defined by $t_{up} = d/v_{mute} - 0.1s$ for the upper bound and $t_{low} = d/v_{mute} + 0.4s$ for the lower bound, d being the inter-station distance. Figure 4.2-2 show the muted gather using $v_{mute} = 2.4km/s$. We observe some traces which are not in phase with the other traces (red circle figure 4.2-2). This could be due to the coordinates uncertainties (details in chap. 1) or the record quality of some sensors. To prevent futures problems, these traces are rejected, as explained in the next steps.

Third,the envelopes of the continuous seismograms are slant-stacked with apparent velocities of 3km/s (figure 4.2-3).

Fourth, the resulting slant-stacked trace is cross-correlated with the virtual source gather after the second step (i.e., filtered and muted). The maximum amplitude of the cross-correlation envelope was picked and a linear trend fitted to the maximum data. This step aims to achieve a phase continuity with increasing trace number as much as possible; thus, the phase measurement is the most accurate. Figure 4.2-3 shows the maximum amplitude (dotes) and the fitted curve is in black.

Fifth, the time lag outlier values are rejected. Outliers are defined as any lag time value greater or smaller than the trend value $\pm 0.05s$. Figure 4.2-4 shows the maximum amplitude selected in cyan blue and the rejected in red. It can be seen that the red values match the erroneous traces in figure 2. Finally, figure 4.2-5 presents the proper virtual source gather that will be used for the phase travel time measurement.

The parameters such as filter frequencies f1 and f2, muted velocity, and $SNR_{threshold}$ are variables and depend on the frequency at which the phase velocity is measured (here called central frequency f_c). The following table describes the parameters.

Central frequency f_c (Hz)	**f1 - f2 (Hz)**	**mute velocity v_{mute} (km/s)**	$SNR_{threshold}$
1.5	0 - 3	2.2	4
2.5	1 - 4	2.2	4
3.5	2 - 5	2.2	4
4.5	3 - 6	2.4	5
5.5	4 - 7	2.4	5
6.5	5 - 8	2.4	5
7.5	6 - 9	2.5	5
8.5	7 - 10	2.5	5

Table 4.1: Waveform selection preprocessing parameters

After a proper waveform selection the phase velocity is extracted. The frequency time analysis described by Lin et al. (2008) is applied, where, the spectral phases of a surface wave single-mode φ is defined as:

$$\varphi(\omega) = -wt_c + 2n\pi + \varphi_0 \tag{4.3}$$

Where, ω is the frequency, φ_0 the initial phase term, and $2n\pi$ is the intrinsic 2π phase ambiguity. Thus, the phase travel time t_c can be written as:

$$t_c(\omega) = \frac{-\varphi(\omega) + 2n\pi + \varphi_0}{\omega} \tag{4.4}$$

The exact value of the initial phase term is not needed since the phase travel-time is used to compute a gradient; thus, φ_0 can be neglected. The spectral phase is computed and unwrapped, and the phase travel time is calculated using the equation 4.4 for each central frequency by averaging the time travel for the frequency range of $f_c - 0.5Hz$ to $f_c + 0.5Hz$. Figure 4.3 shows an example of phase travel time measurement at 4.5Hz for the virtual source 01.01002. Panel

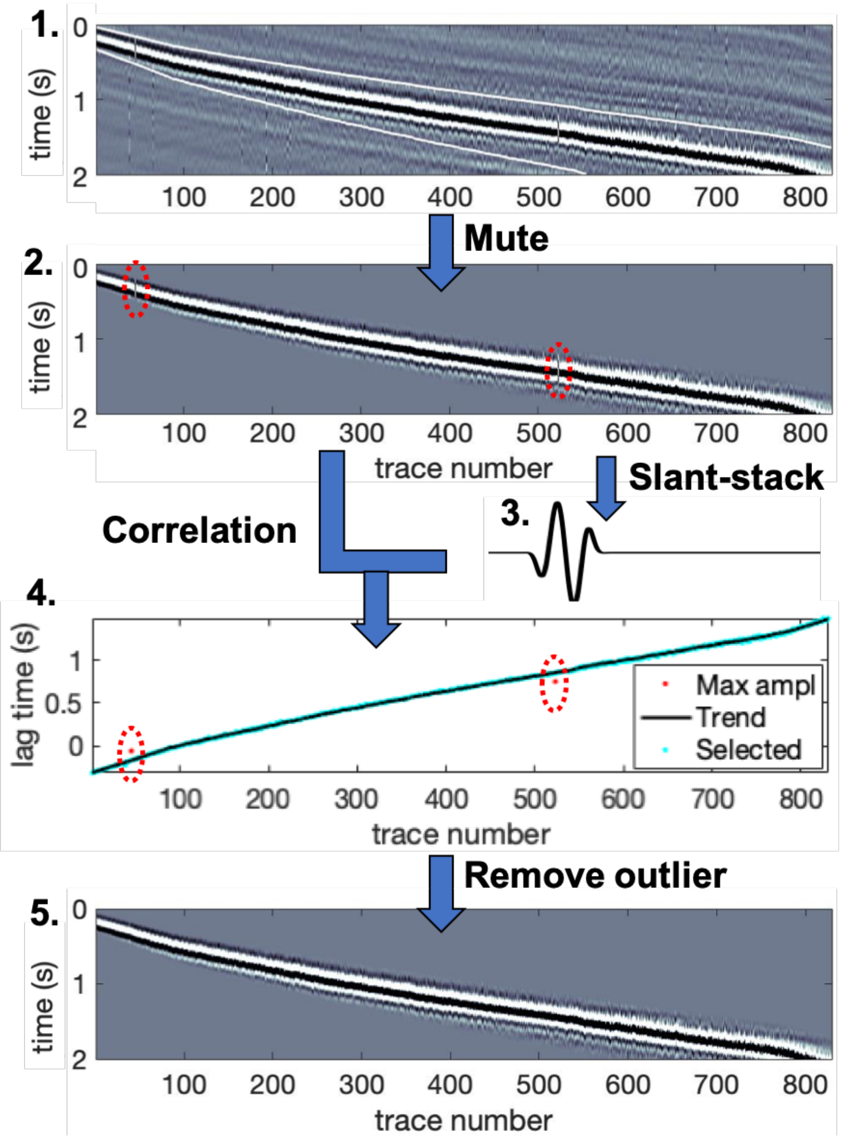

Figure 4.2: Waveform selection preprocessing workflow, example, for $f_c = 4.5Hz$ and station 01.01002 (see text for details). From top to bottom: 1. virtual shot gather filtered between [3 - 6.5]Hz. 2. Gather muted around the surface wave. 3. Slant-stack computation. 4. Correlation between the muted gather and the slant stack and Outlier removal. 5. Final virtual shot gather.

a shows the unwrapped phase, and then panel b shows the travel time in seconds. Figure 4.3-c shows the travel-time measurements superimposed on the virtual gather (red dots). To compare, the maximum amplitude of the envelope is also computed (green dots), but this approach is less accurate to pick phase velocity.

4.3.2 Interpolation of phase travel-times

After properly tracking the waveform across the array for each virtual source, the spatial gradient of the travel time can be computed. The gradient computation process is implemented following

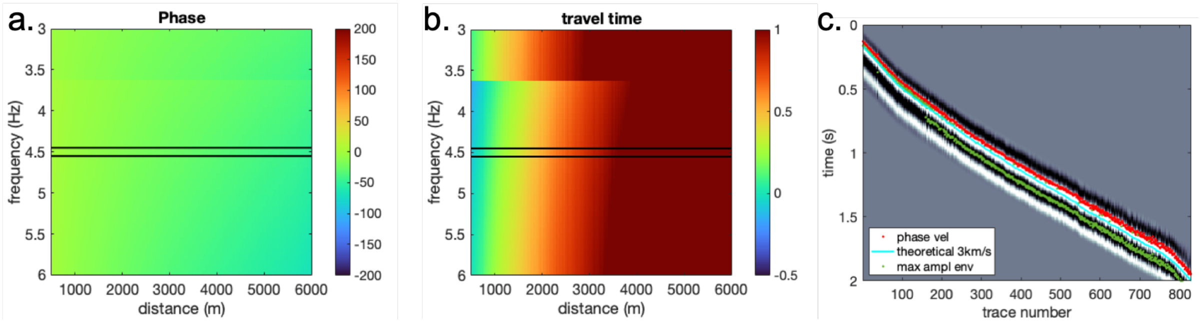

Figure 4.3: Phase travel time measurement workflow, example for $\omega = 4.5Hz$ and station 01.01002. a) Spectral phase of the correlations between [3-6]Hz and b) Travel-time both figures as a function of the frequency and the inter-station distance. c) Pre-processed virtual shot gather sorted in by increasing distance, red dotes are the phase velocities computed following equation 4.4, green dots are the envelop maximum amplitude, and cyan blue line represent the velocity at 3km/s.

the next nine steps. Figure 4.4 shows an example of the implementation for the virtual source 01.01001 at 4.5Hz (the same case show for the waveform selection and phase travel times measurement).

1. The phase travel time is measured (preliminary step). Figure 4.4-1 (the red star is the virtual source)

2. First interpolation: the travel time measurements obtained at station locations are interpolated onto a regular spatial grid using the spline 2d method (details below). Figure 4.4-2 shows the interpolation.

3. Gradient computation and quality control: the gradient is calculated in 2 dimensions (figure 4.4-3). The outlier values (values smaller than $2.5e^{-4}s/m$ values greater than $5e^{-4}s/m$) are rejected.

4. Curvature computation and quality control: to avoid spurious oscillations, the gradient's curvature (Laplacian of the travel-time surface) is computed (figure 4.4-4). Areas with values greater than $4e^{-7}s^2/m^2$ are removed.

5. Second interpolation: applied after outlier rejection (figure 4.4-5).

6. Interpolation difference: we computed the difference between the two interpolations (steps 2 and 5) to see the numerical errors. Figure 4.4-6 shows errors of the order 100 picoseconds with is negligible.

7. Second spatial gradient computation: using the second interpolation, the gradient is computed a second time, as shown in figure 4.4-7. The outlier values are removed as in the 3 steps.

8. Slowness: the slowness is obtained by computing the squared gradient, and measurements from locations that are not surrounded by at least four locations with measurements are removed (figure 4.4-8).

9. Velocity maps: individual velocity maps are computed by virtual source by taking the inverse of the slowness (figure 4.4-9).

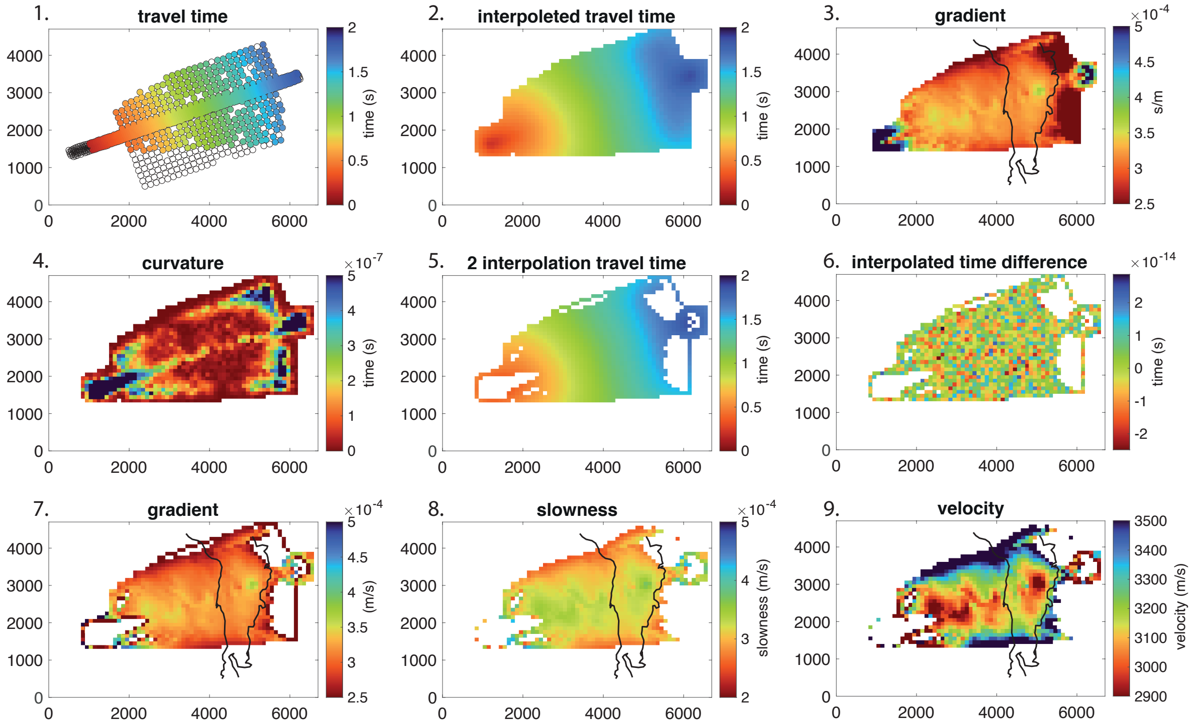

Figure 4.4: Eikonal tomography's procedure for virtual source 01.1002 the red start (see text for details). The black line is the Gabbro structure at the surface (fig. 3, 7-9). Travel times are in seconds (fig. 1, 2, 3, and 6). Gradient and slowness panels in s/m (fig. 3,7,8), and the velocity in m/s (fig. 9). Figure 6 shows the difference between first and second interpolation (numerical noise).

The interpolation method applied for the phase travel time is a spline-in-tension interpolation scheme developed by Wessel and Bercovici (1998). The technique creates an ”elastic surface” that passes smoothly between the data points (here, the travel time at the station's locations). The surface has a flexural rigidity controlled by a ”tension coefficient” applied at the boundaries. It ranges between 0 and 1 and represents the strain energy resulting from the tension relative to the total strain energy of the surface (Wessel and Bercovici, 1998). The effects of this tension coefficient will be discussed later in the text. The grid used for the interpolation is a regular grid of 100x100m. This value is a good trade-off; greater distances will create regions without data generating uncertainties for the interpolation, and a grid with a smaller distance will have several measurements at one spatial node, and interpolation will smooth the data too much, hiding velocity changes. The outlier values for the gradient and the curvature are chosen based on the prior geological information of the shear velocity at Marathon.

Additional quality control is implemented based on the number of the virtual receiver. Because of the selective cross-correlation method, each virtual source has a different number of virtual receivers (i.e., number of traces). Good coverage is needed with enough phase travel time measurement to apply the gradient computation workflow, especially the interpolation. For this reason, virtual sources with less than 150 virtual receivers are not taken into account. For the rest of the virtual sources, the workflow is applied separately at a given central frequency.

4.3.3 Computing final phase velocity map

By calculating the gradient, the individual phase velocity maps are obtained. However, these maps are noisy because of biases and errors induced during the phase travel time measurement and the interpolation. In addition, each virtual source covers different regions due to the variant station coverage. To have a more complete and less noisy phase velocity map we applied a statistical average for each spatial node (i.e., cell in the grid) as proposed by Lin et al. (2009). They introduce the velocity distribution V(r) as:

$$V(r) = \frac{1}{N(r)} \sum_{i=1}^{N(r)} v_i(r) \tag{4.5}$$

Where $s_i(r)$ is the individual velocity and N(r) the number of measurements at every spatial node (r). The standard deviation of the mean slowness σ is also introduced as:

$$\sigma^2 = \frac{1}{N(r)(N(r)-1)} \sum_{i=1}^{N(r)} (v_i(r) - V(r))^2 \tag{4.6}$$

Before the final phase velocity map calculation, an outlier rejection in four steps is performed. First, the mean velocity by the virtual source is computed, and virtual sources with mean velocity values out of ±1.5 the standard deviation of virtual source mean velocities. Figure 4.5-1 shows mean velocity distribution by virtual source (blue cross) for a central frequency of 4.5Hz and the conserved virtual sources (red dots).

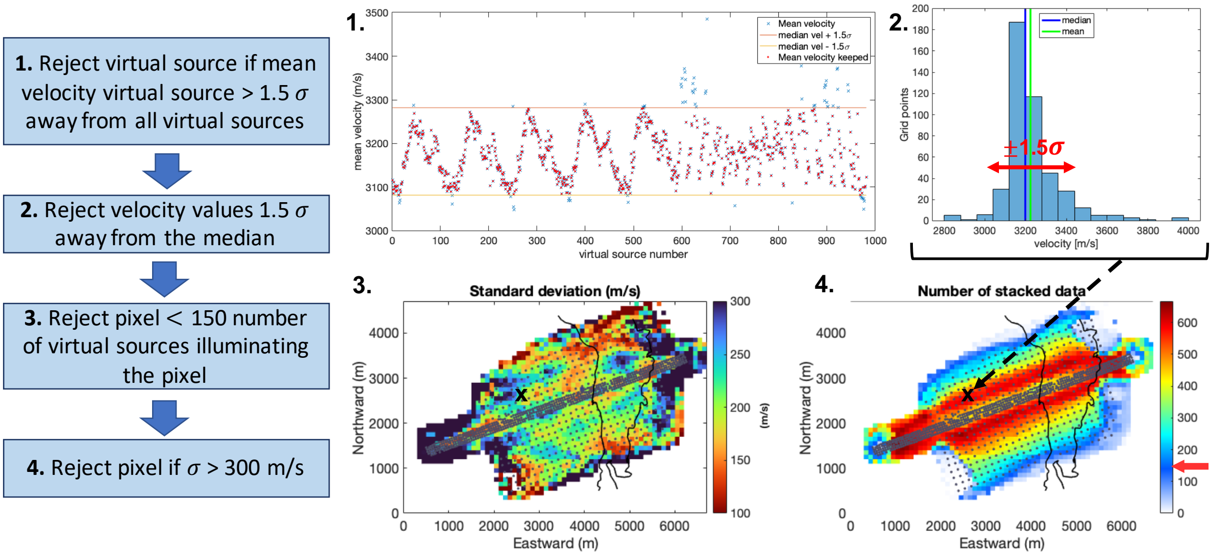

Figure 4.5: Statistical quality control workflow, example at 4.5Hz (see text for details). 1. Mean velocity by virtual source. Lines are ± standard deviation, blue cross represent the mean velocity per virtual source and the red dots are the values conserved. 2. Standard deviation distribution for one pixel (black cross figs. 3 and 4). 3. Map of the standard deviation ω. 3. Map of the number of stacked data. The black line is the gabbro structure at the surface. Gray dots are the nodes (figs. 3 and 4)

Second, the velocity distribution at each spatial node is analyzed, and values greater or smaller than $\pm 1.5\sigma$ are rejected. Figure 4.5-2 shows the velocity distribution at the node located at the black cross in panels 3 and 4.

Third, the spatial node for which the standard deviation is higher than 300 m/s are rejected because these values are not reliable data (figure 4.5-3). Finally, the nodes with less than 150 measurements are removed as well because we do not have enough data to do an statistical approach (figure 4.5-4). These two steps help to reduce side effects by mainly removing values at the borders of the array. The impacts of train coverage are evidenced by the low number of data in the Northwest Southeast direction as it can see the figure 4.5-4.

4.3.4 Phase velocity models

The different workflows beforehand explained are applied for central frequencies from 1.5 to 8.5Hz. Figure 4.6 presents the surface wave phase velocity maps of Marathon deposit obtained with the eikonal tomography at eight different frequencies. Three main structures are discernible: two low-velocity regions at the East and West and fast velocity anomalies (yellow regions) in the center East, crossing the studied area. In addition, an apparent rotation of the strike of the high-velocity anomaly at depth is also discerned.

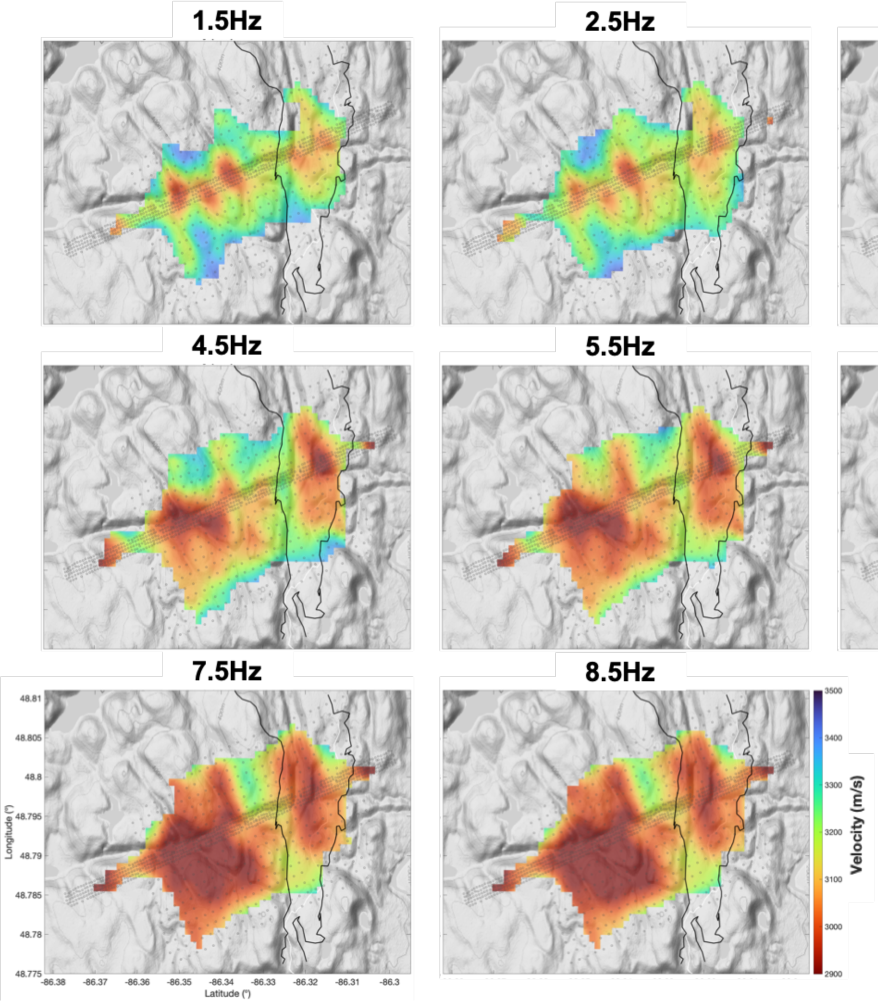

Figure 4.6: Final eikonal tomography phase velocity maps from 1.5 to 8.5Hz. The black line is the Gabbro structure at the surface. Gray dots are the nodes. The same color bar was used for all maps.

4.4 Eikonal tomography implementation with synthetic data

4.4.1 Synthetic dataset

A synthetic dataset was created to assess the methods developed in this work and evaluate the viability of the eikonal tomography in Marathon geological context. These synthetic tests will also allow us to understand better the train virtual shot gathers content.

The synthetic dataset consist of a virtual shot gathers for the same array that in Marathon's case. The virtual shot gathers are simulated by modeling a source at the virtual source locations using SEM46 software (Trinh et al., 2019). SEM46 is a time-domain spectral-element method under the visco-elastic approximation. To simulate wave propagation, we used a model with three homogeneous geological units corresponding to the Syenite, the Gabbro, and the Footwall (see the characteristics in the table 4.2). This is a simplified model based on the prior geological information at Marathon, where the Gabbro is dipping to the West (see figure 4.7). The values of Qp and Qs are 200 and 100, respectively. It is important to note that the velocities contrasts used for the modeling are higher than the described in the sonic studies in order to facilitate the interpretation while we conserve realistic values.

Geological structure	V_P **[m/s]**	V_S **[m/s]**	**Density** $[kg/m^3]$
Syenite	5740	3295	5740
Footwall	6000	3350	2870
Gabbro	6380	3615	3070

Table 4.2: Seismic properties of the geological units considered in the synthetic models, inferred from sonic log measurements.

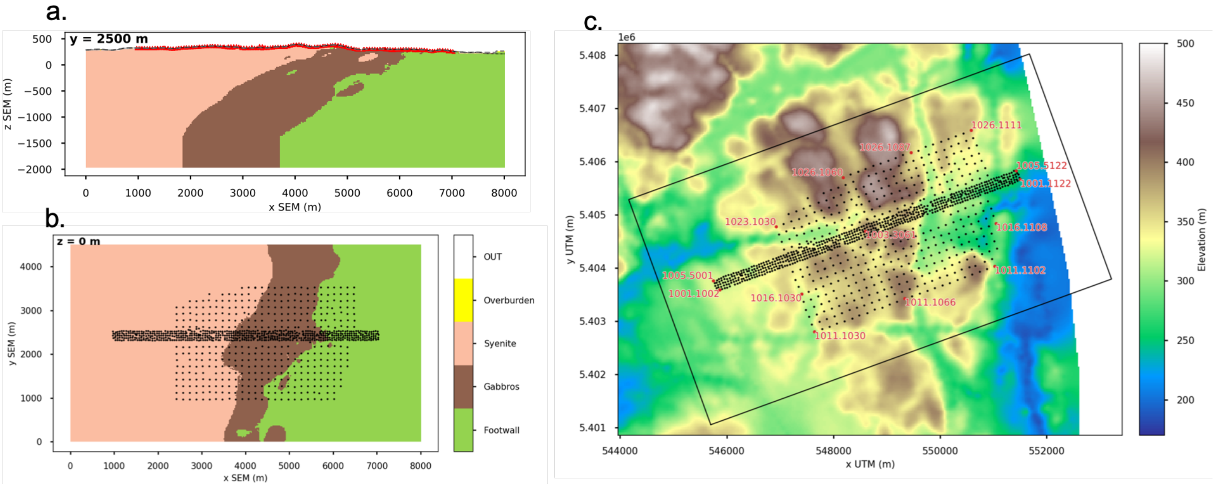

Figure 4.7: Geological units of the synthetic models. a) vertical cross-section along the dense line (red triangles represents the sensors). b) horizontal cross-section at elevation z = 0m (sea level). c) Acquisition geometry of the numerical experiments. Black dots represent the sensors, and the red dots are the 14 sources used for the synthetic data. The black rectangle represents the simulation domain. Courtesy François Lavoué

The shot gathers were simulated for 14 source positions (red dots figure 4.7) located in the extremities of the array to have full illumination of the medium while an approachable computational time. A 4s Dirac delta source time function was used as a source. The resulting seismograms are low-pass filtered at 50Hz and sampled at 250Hz.

Figure 4.8 shows the synthetic shot gathers for the source located at the receiver 1001.1002.

The gathers are band-pass filtered between [3 - 15]Hz (panel a) and between [15 - 30]Hz (panel b and c). The surface wave traveling at 3.5km/s is dominant at low frequency. At higher frequencies, the P-wave and a mix between an S-wave and higher modes of the surface waves are traveling at 6km/s and 3.8km/s, respectively. The velocities and frequency ranges at which these waves propagate are similar to the waves retrieved by cross-correlation of train signals. Therefore, we could suggest that the arrivals retrieved from cross-correlation tend to converge to the green function. However, this hypothesis must be taken with caution because we used a simplified model of Marathon geological structures, and the comparison is not straightforward.

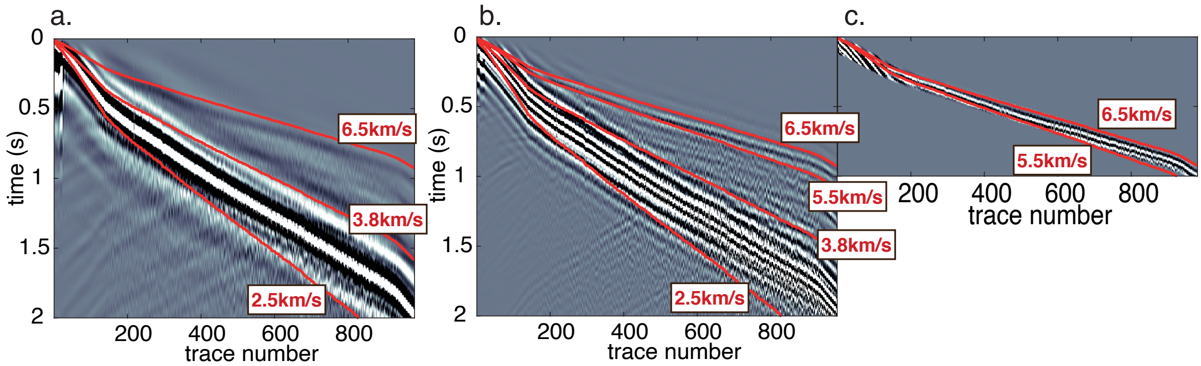

Figure 4.8: Synthetic shot gathers for the source at the position of the receiver 1001.1002 (figure 4.7-c). Trace numbers are sorted by increasing distance from the virtual source with a not constant step. The red lines designate travel times with constant velocities, as indicated in the withe boxes. a) Gather filtered between [3 - 15]Hz, a clear surface wave propagates. b) Gather filtered between [15 - 30]Hz, two arrivals are retrieved an S-wave or higher mode of the surface wave and a P-wave. c) Gather muted between 5.5–6.5km/s to visualize the P-wave better.

To understand the numerical wavefield, figure 4.9 shows snapshots of seismic wave propagation at different times for the same source that before and the normalized and clipped seismograms for a receiver at the center of the dense line. It can be seen the different waves propagating in the medium. Remarkably, the body wave recorded is a direct wave, and it does not travel deep in the medium because there is no velocity gradient in the model. Besides, the body-waves reflections have a low amplitude, making them undetectable and therefore unusable even in a model with higher velocity contrast than expected in the region.

This modeling is theoretical and simplified, but it guides us to realize the feasibility of obtaining certain results in Marathon deposit. Therefore, we can conclude that it is too challenging to obtain virtual body-wave reflections in this context due to the poor velocity contrasts. In addition, the virtual P- wave retrieved from real data is probably a direct P-wave sensitive to the shallow surface.

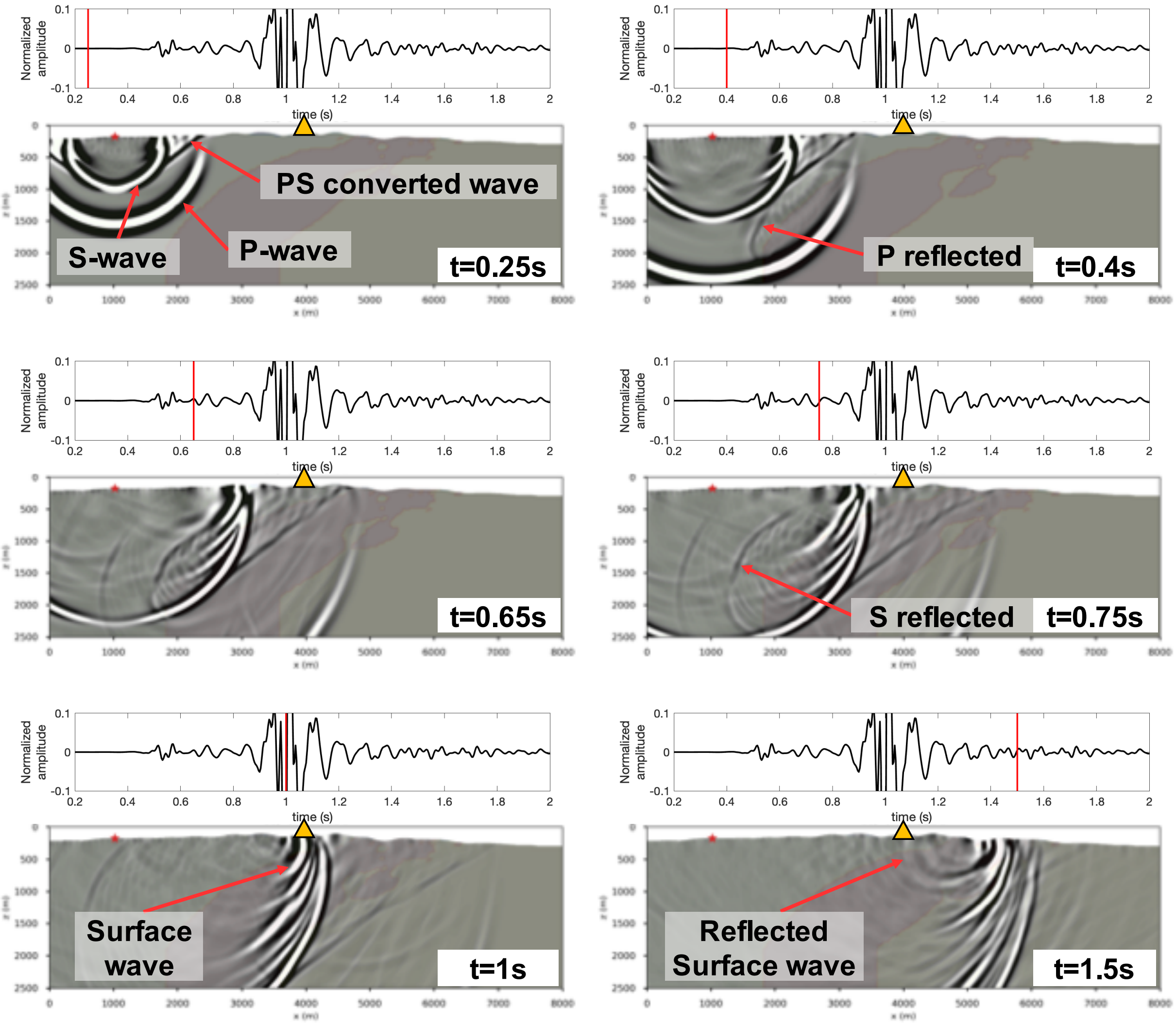

Figure 4.9: Numerical modeling results for a source located at the red start. (bottom panels) Snapshots of seismic wave propagation at 0.25, 0.4, 0.65, 0.75, 1 and 1.5s. The dark violet structure corresponds to the Gabbro. (Top panels) seismic record for the station located a the yellow triangle. The trace is filtered between [5 - 40]Hz, normalized, and clipped at ±0.1. The red line represents the snapshot time. Courtesy François Lavoué.

4.4.2 Eikonal tomography

In order to evaluate the efficiency of the eikonal tomography approach, the same data processing explained and applied to the real data (section 4.3) was applied to the synthetic data. First, the surface wave is selected, and phase travel times are measured for one central frequency at 4.5Hz. Second, the phase travel times are interpolated onto the same spatial grid (100x100m). Finally, the final phase velocity map is obtained by stacking all 14 sources. The final velocity map (figure 4.10) indicates 3 main velocity zones: two low velocity structures (at $\approx 3000m/s$) and a high velocity anomaly (at $\approx 3250m/s$). These structures coincide with the geological units of the models; the faster anomaly corresponds to the gabbro (black line) and the low-velocity structures to the Syenite and the Footwall. It is important to note that the velocities obtained with the eikonal tomography are the Rayleigh wave velocities and not the S-velocity V_S; Rayleigh wave velocity represents $\approx 85\% V_S$ which is coherent with the velocity values measured.

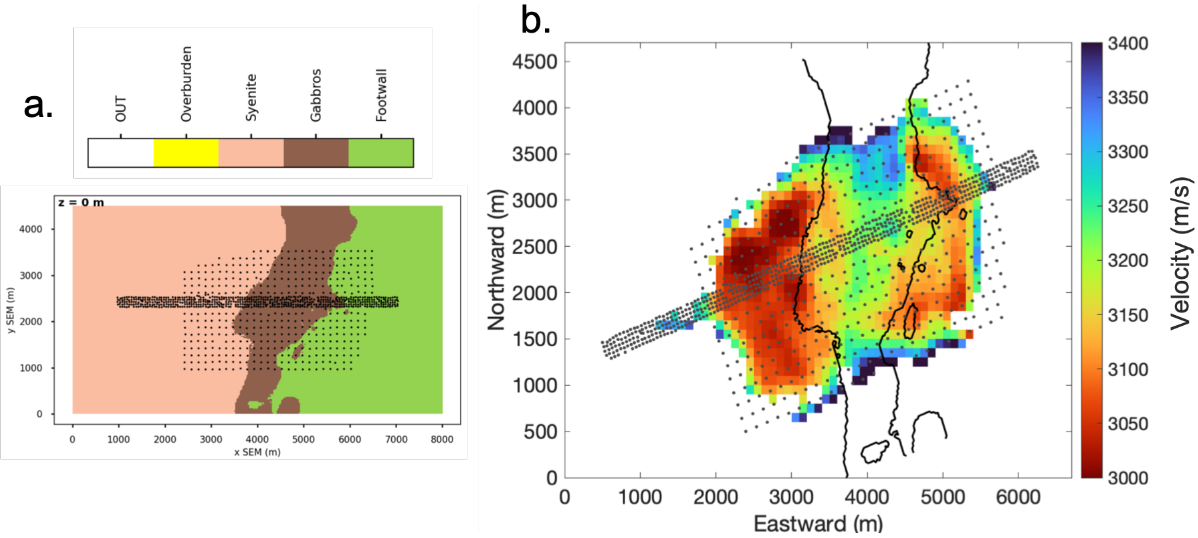

Figure 4.10: a) Geological synthetic models horizontal cross-section at elevation z = 0m(sea level). b) Eikonal tomography at 4.5Hz for the synthetics data, Courtesy François Lavoué. The black line represent the synthetic Gabbro unit at the surface. Gray dots are the nodes. Courtesy François Lavoué

The high-velocity anomaly does not match precisely with gabbro for several reasons. First, the black line here represents the gabbro at the surface, and the phase velocity map is sensitive to the structure at depth; thus, the comparison can not be made directly. Second, the eikonal tomography approach computes the phase velocity using a statistical approach to decrease the errors; however, only 14 sources are used for the simulations; therefore, error reduction is less efficient. Third, the spatial discretization of the measurement is defined by inter-station distance ($\approx 150m$ in the sparse array), creating a measurement ambiguity of this order of magnitude. Finally, the interpolation smooths the measured producing increasing the ambiguity. We also observe that small structures ($\approx 30m$) of gabbro can not be imaged using the eikonal tomography at this frequency because of the resolution limits. Thus, only major gabbro intrusion can be detected for both synthetics and real data.

The synthetic data was also used to evaluate the effect of the spline-in-tension interpolation. As explained before, this interpolation is based on a tension coefficient. We will use the synthetic data to evaluate the effects of this coefficient. To do so, we apply the same processing by varying the tension coefficient from 1-1e^{-15} to 0.2. Figure 4.11 shows the finals phase velocity maps for eight different values of tension coefficient. We notice that for high tension coefficient values (panels a, b, and c), the "elastic surface" used for the interpolation is too flat and smoothing phase travel time too much. Consequently, the travel time variations related to changes in the medium will be suppressed, and we will lose the information of some structures; as shown in figure 4.11 panels a, b, and c. the gabbro intrusion is not apparent anymore. Additionally, the side effects strongly produce high velocities at the array boundaries at high tension coefficient values.

On the contrary, when the tension coefficient value tends to zero (figure 4.11 panels f, g, and h), the "elastic surface" used for the interpolation is too flexible, and it creates artificial travel time differences that are not related with the geological structures. However, the side effects are less critical. The intermediate values of the tension coefficient (figure 4.11 panels d and e) seem to be the more accurate; they do not create artificial structures nor smooth too much the measures by removing structures. Therefore, a tension coefficient value of $1 - 1e^{-6}$ was used for the synthetics and the read data.

The synthetic data allow us to detect the limits of the Eikonal tomography at the Marathon case and test the tension coefficient to better applied the technique. Besides, even a simplified model of Marathon deposit modeling allows us to be guided in the compression of that wave propagation.

4.5 Depth inversion

The eikonal tomography allows us to have phase velocity maps at different frequencies where geological structures are discernible. However, to construct a 3D model, the phase velocity maps need to be inverted. The relationship G between the physical parameters in the model m and the observables d has to be established; mathematically, this is described as d = Gxm. The inverse problem consists in estimating the model parameters m from the observed data d_{obs}, expressed as m = F(d_{obs}). There are two main families of inversion methods: deterministic and probabilistic. Deterministic methods are based on the local linearization of the direct problem, while probabilistic methods perform a global exploration in the parameter space.

Since surface waves are more sensitive to V_S velocities in the medium, the inverse problem in our case consists in obtaining a 3D V_S model from the measurements of the phase velocity. There is not a linear solution to obtain the velocity model, $V_S(z)$ can not be express in terms of the dispersion curves. Additionally, the solution is not unique, which means several different models

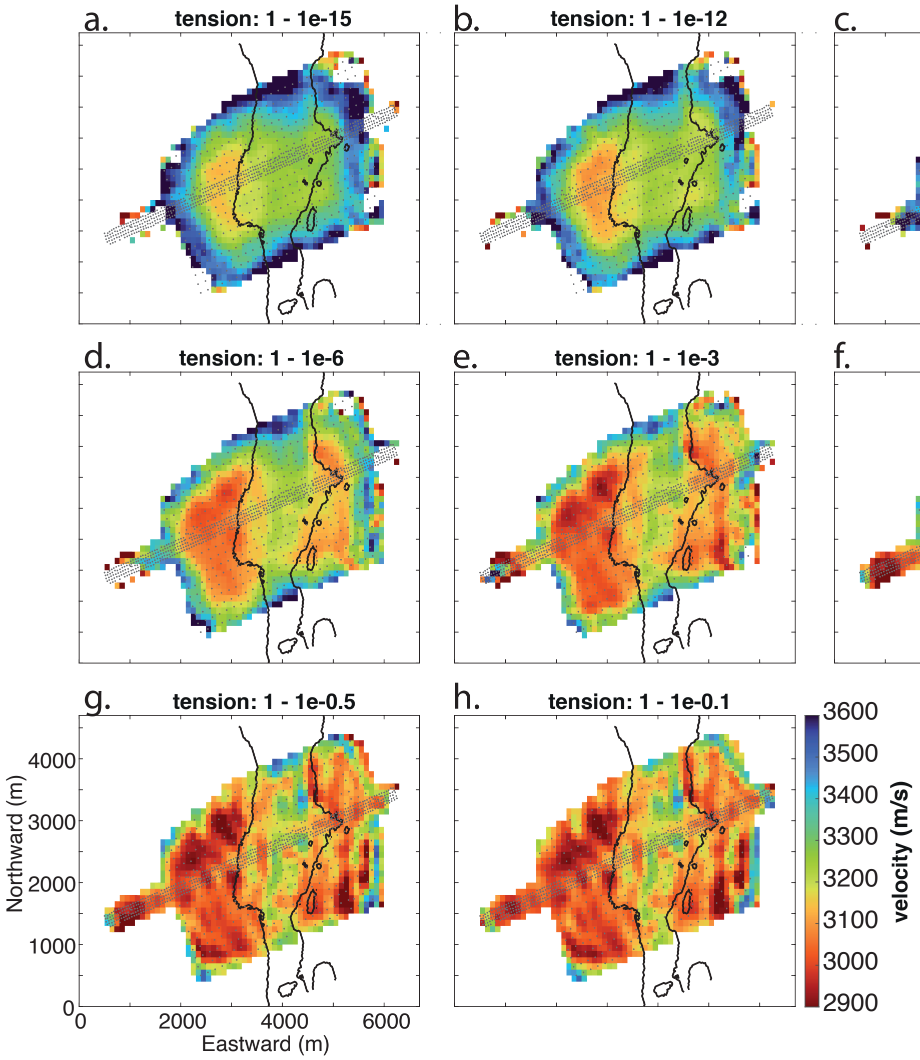

Figure 4.11: Eikonal tomography at 4.5Hz for the synthetics data using different values of tension coefficients (value at the top of each panel). The black line represents the Gabbro instruction. Gray dots are the nodes.

can describe very similar wave propagation generating similar dispersion curves. This problem is well suited to probabilistic inversions as the Monte Carlo method. Monte Carlo inversion consists of randomly sampling the model space to determine the a posteriori probability density of the solutions (Tarantola, 2005).

The different phase velocity maps at one single spacial node can be seen as a phase dispersion curve. To evaluate the local velocity model (1D layered model), these curves need to be inverted. We decided to use the Neighborhood Algorithm (NA) developed by Sambridge (1999) for the inversion problem. NA is a Monte Carlo global direct-search technique aimed to find an ensemble of models that sample the promising data-fitting regions of model-space.

NA algorithm makes an inversion in several steps. First, it generates randomly distributed models n_i in the model space. Then, Voronoi cell mesh is created (the model space is partition into regions close to each model). Next, the theoretical dispersion curve is computed using Herrmann and Ammon (2004) routines for each model. The misfit between the theoretical and the observed dispersion curve is defined as the area of the theoretical dispersion curve outside the measured dispersion curve area, and its uncertainties, normalized by the area of the measured dispersion curve (Mordret et al., 2014). The misfit is associated with the corresponding Voronoi cell, and the n_c cells with the better misfits are conserved. Then, each conserved Voronoi cell is resampled into n_s subspace, and new modes are generated inside each subcell. This procedure is repeated n_n times. This first random distribution is designed to explore a wide range of model space. For then, focus only on the most promising zones of model space and thus converge more rapidly to a global minimum.

The model parametrization requires the thickness and the elastic proprieties in each layer. Following Mordret et al. (2014) approach, we only invert V_S velocity at depth. Other parameters as V_P and density ρ are not considered as free parameters but they are defined as a function of V_S thus: $V_P = 1.16VS + 1.36$ (Castagna et al., 1985) and $\rho = 1.74V_p^{0.25}$ (Brocher, 2005).

Besides, the parameterization of the 1-D Velocity profile is not done in function of layers and the corresponding velocities V_S, but it is parametrized by overlaying cubic B splines; this allows us to decrease the number of parameters to be inverted conserving a good characterization in depth. A cubic spline basis function is defined by the weights of each spline and the relative spacing between the knots. The knots spacing can also be inverted to less control the inversion. To do so, we vary the spacing spline knots by raising them at the power of p, keeping the first four knots and last four knots set to zero forces the velocity to be nonzero at the surface. Finally, we have seven parameters, six weights of each spline (S1, S2, S3, S4, S5, and S6), and the power p.

Figure 4.12 shows the model parametrization approach. Panel a shows two the cubic spline basis created with six cubic splines (continuous lines) and the corresponding knots (stars). The first basis (black lines and red stars) is generated using a power p=2 and the second one (gray lines and green stars) uses p=1. The weighted cubic splines are generated from the cubic spline basis and the weights S1, S2, S3, S4, S5, and S6 (values detailed in the figure 4.12 caption). Figure 4.12-b illustrated the example of the weighted cubic splines for p=2. Finally, the weighted cubic splines are overlapped, and the velocity model is parameterized as shown in the figure 4.12-c for the two cubic spline basis. The exploration value range of the splines weights and p used for the inversion are detailed in Table 4.3.

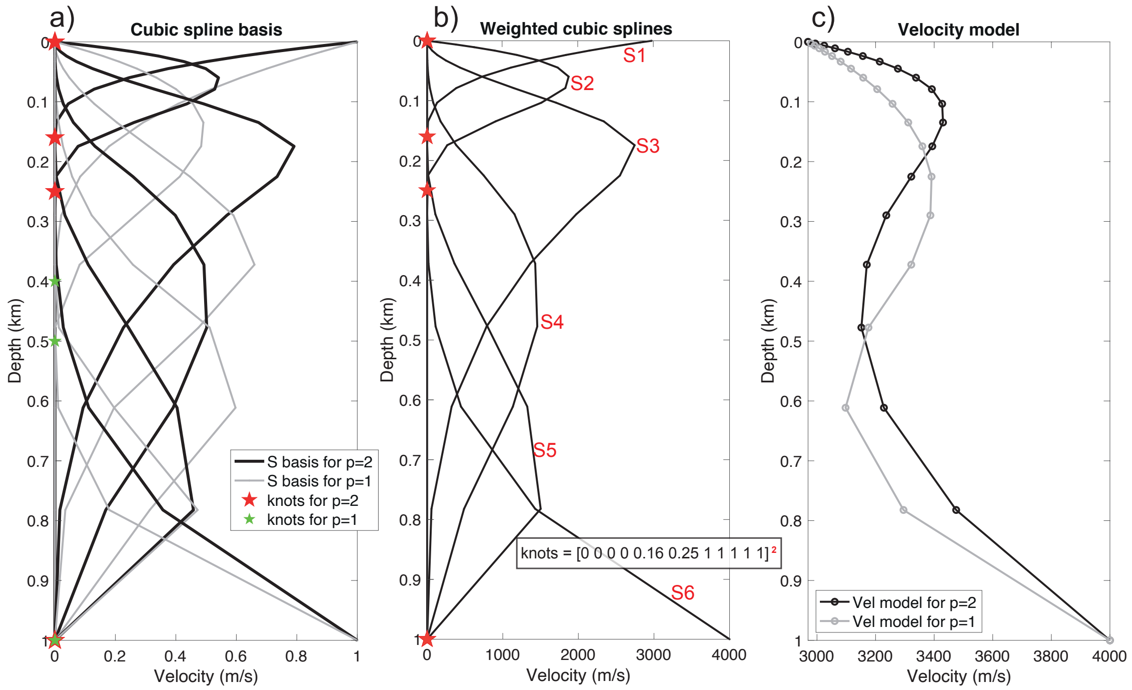

Figure 4.12: Model parameterization for two values of p. The power p for the spline knots controls the depth extent. The seven parameters for inversion are in red (S1, S2, S3, S4, S5, S6, and p). a) Cubic spline basis (continuous lines) for two parameters of p and knots positions (stars). b) Weighted cubic spline for p =1. The weigh used are: S1=2.9695km/s, S2=3.4590km/s, S3=3.4699km/s, S4=2.9km/s, S5=3.2771km/s, S6=4km/s. c) Resulting 1D velocity model for both examples.

The NA inversion provides a group of 1D velocity models for each spatial node and their corresponding misfit. However, since the velocity variations at the marathon deposit are weak, we decided to improve the inversion by including a new step: the linear inversion. To do so, we applied the Rayleigh-wave phase linear inversion method developed by Haney and Tsai (2017). This finite-element method seeks to refine an initial shear-wave velocity model by perturbing the initial model iteratively to reach a model with an acceptable misfit.

The linear inversion method assume that (1) Poisson's ratio and density are fixed throughout the inversion and (2) P-wave velocity and density are fixed (Haney and Tsai, 2017). Under this assumption a linear relation between phase velocity and shear-wave velocity for any given mode can be described as:

$$\frac{\delta c}{c}(f) = K(f,z)\frac{\delta\beta}{\beta} \tag{4.7}$$

Where, $\frac{\delta c}{c}(f)$ are the phase velocity variations depending on the frequency, K is the sensitivity kernel and $\frac{\delta\beta}{\beta}$ are the relative shear velocity perturbations.

To regularize the inversion, a weighted damped least-squares method is used. The data and model covariance matrices are chosen as define in Garofalo et al. (2016). The data covariance matrix is diagonal and described as:

$$C_d(i,i) = \sigma_d(i)^2 \tag{4.8}$$

Where $\sigma_d(i)$ is the standard deviation of the data and i is the phase velocity measurement. The model covariance matrix is described define as:

$$C_m(i,j) = \sigma_m^2 exp(-|z_i - z_j|/d) \tag{4.9}$$

Where σ_m is the model standard deviation, z_i and z_j the depth of the top and the elements ith and jth and d the correlation length. With these matrices, we proceed to apply the total inversion algorithm (Tarantola and Valette, 1982; Muyzert, 2007) by solving the following system:

$$\begin{bmatrix} C_d^{1/2}K \\ C_m^{1/2} \end{bmatrix} \frac{\delta\beta}{\beta} = \begin{bmatrix} C_d^{-1/2} \\ 0 \end{bmatrix} \frac{\delta c}{c} \tag{4.10}$$

The inversion is computed, generating a new model (or updated model). Then, the updated model is perturbed, and we repeat the inversion computation until the stopping criterion is attended. Here the stopping criterion is defined as chi-squared value (Gouveia and Scales, 1998):

$$\chi^2 = (\frac{\delta\beta}{\beta})^T C_d^{-1} (\frac{\delta\beta}{\beta})/F \tag{4.11}$$

Where F is the number of measurements, the iteration is repeated until the chi-squared value falls within $[\chi_{low} - \chi_{high}]$ window or the maximum allowed number of iterations is reached.

4.6 Depth inversion implementation

The depth inversion implementation was carried out in two stages NA inversion and linear inversion. The parametrization and the implementation for Marathon case will be described in the following section. The phase velocity maps used for the inversion are obtained from the Eikonal method; they were measured between 1.5 and 8.5, every 1Hz. Since not all the spatial nodes contain the same number of measurements, we decide to conserve only spatial nodes with at least six-velocity measurements.

4.6.1 Neighborhood Algorithm parametrization

We invert the local dispersion curves at depth following the forward modeling method propose by Mordret et al. (2014) and using the subroutines from the Computer Programs in Seismology package (Herrmann and Ammon, 2004).

The allowed ranges for the spline basis are described in the table 4.3, they were chosen based on the sonic log values studies and a priori geological knowledge from Marathon deposit (chapter 1, section 1.3.1). The NA was run with n_i= 5001 (initial models). Then, the best 300 Veronoi cells conserved (n_c=300). These sub-cells are resampled $n_s = 3$ times. And the full procedure is repeated 12 times ($n_c = 12$).

Parameter	**Allowed range**
S1	[2000 - 3200] m/s
S2	[2500 - 3800] m/s
S3	[2700 - 4000] m/s
S4	[2700 - 4100] m/s
S5	[2700 - 4100] m/s
S6	[2700 - 4100] m/s
p	[1-5]

Table 4.3: Ranges Allowed for the NA inverted parameters

4.6.2 Linear inversion parametrization

We follow the inversion method and the algorithms described by Haney and Tsai (2017) which were adapted to our case and modified to decrease the computational time.

We generate the initial model by averaging the 100 best models obtained from the Neighborhood Algorithm inversion and interpolating it in depth with a regular step of 30m. By doing so, we ensure to be neighborhood to the final model, and thus the linear inversion is done in a more restrained space. The model covariance matrix is formed using a correlation length d=200m, and the standard deviation factor equals 2.5. It is important to note that the model standard deviation is given as σ_f times the median of the standard data deviations. The stopping criterion χ_{low} and χ_{high} are established at 1 and 1.5, respectively. Therefore, the maximum iteration number is 5; since the initial model is neighborhood to the final model, we do not need many iterations to converge to a minimum.

The linear inversion is evaluated by individual node; when the initial model reaches the χ stopping criterion, the inversion obtained with NA method is conserved as the final inversion.

To evaluate the improvement of the inversion, we estimate the variance improvement between the initial model (obtained from NA inversion) and the final model after the linear inversion.

4.6.3 Results

Figure 4.13 summarizes the results for each inversion at 2 different cells (stars figure 4.16) which are part of different velocity structures. The first line of the figure shows the NA inversion distribution. We can point out that the 1D velocity distributions for small misfits (warm colors) are too extended for the first 100m showing that the model is not well constrained for the first hundred meters. Given that the maximum frequency analyzed is 9.5Hz, the resolution at the shallow surface (<100m) can not be correctly constrained. Besides, from 100m, the problem is correctly constrained to 1km depth.

Since the parameterization used for the NA inversion was a spline basis which is not flexible enough to detail all the anomalies at Marathon where the velocity contrasts are poor. The second inversion was carried out to improve the velocity models to highlight the anomalies by approaching the "real model." We remained that the initial model for the linear inversion is the average of the 100 best models from NA inversion. The figure 4.13 second line shows the results from the linear inversion at the same two points. For point 1272, the linear inversion does not change the final model for this drastically; the misfit reduction between the NA and the linear inversion is 15% (as shown in the variance reduction map figure 4.14-b). For the second example, the variance reduction is approximately 60%, improving the final 1D model. The kernel sensitivity (figures 4.13 i and j) show the linear inversion sensitivity between depth and the period.

Figure 4.15 shows four horizontal slices of the S-wave velocity model at 160m, 300m, 500m, and 700m depth. Blue regions represent fast velocities, and red colors low-velocity regions. The shallower slice at 160m shows three structures two low structures at the extremities and one fast at the center. As we increase the depth, the medium is more homogeneous, and there is no clear fast velocity structure anymore after 500m. We can distinguish two artifacts probably created by the numerical procedure. The first one is a red cell located at the south est (approximately at lat:-86.315°, lon:48.795°). This artifact can also be seen in the error map (figure 4.14-a) as it has a very large value (>80m/s) that exceeds the median misfit values.

The second artifact is located at the zero value regions in the reduction misfit map southwest (figure 4.14-b). These values correspond to cells where the second inversion was not performed since stopping criterion values were reached directly with the NA inversion. These results present

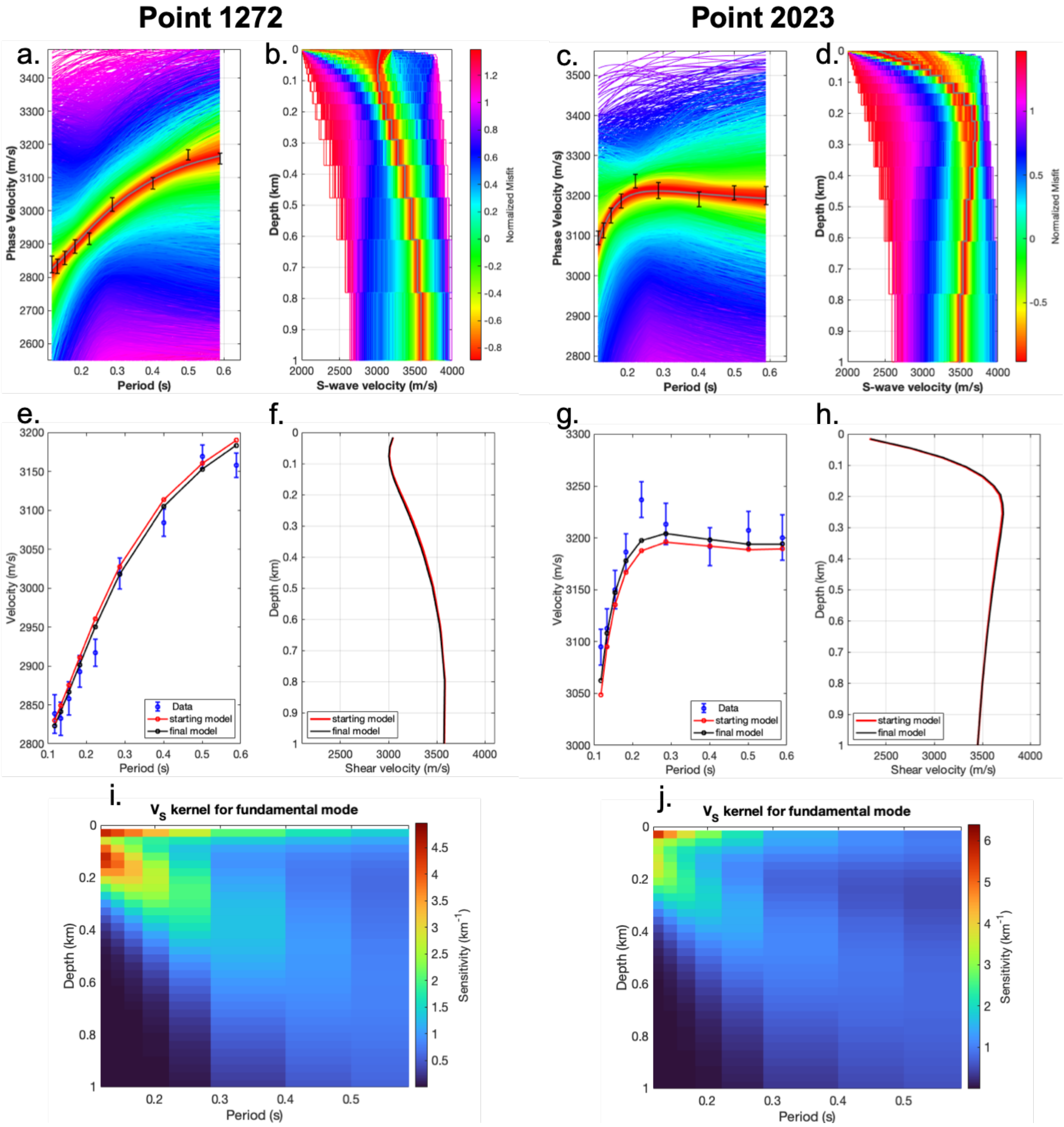

Figure 4.13: Examples of 1D NA and linear inversions at 2 locations (stars figure 4.16). First line: NA inversion. Second line: linear inversion. e) and g) Dispersion curves. Blue dots are measurement with error bars, black lines are the initial models (average of 100 best models of NA inversion), and red lines are inverted dispersion curves from linear inversion. b) and d) Shear velocity depth models from NA, the color represent the misfit. The gray line represents the best model. f), and h) Shear velocity depth model black lines are the initial models (average of 100 best models of NA inversion), and red lines are inverted dispersion curves from linear inversion. i) and j) Sensitivity kernels of the linear inversion for the Rayleigh wave.

a difference in the inversion methodology that causes a sharp velocity change which is not physical but purely numerical. The sharp velocity contrast is clear in the figure 4.15 slice

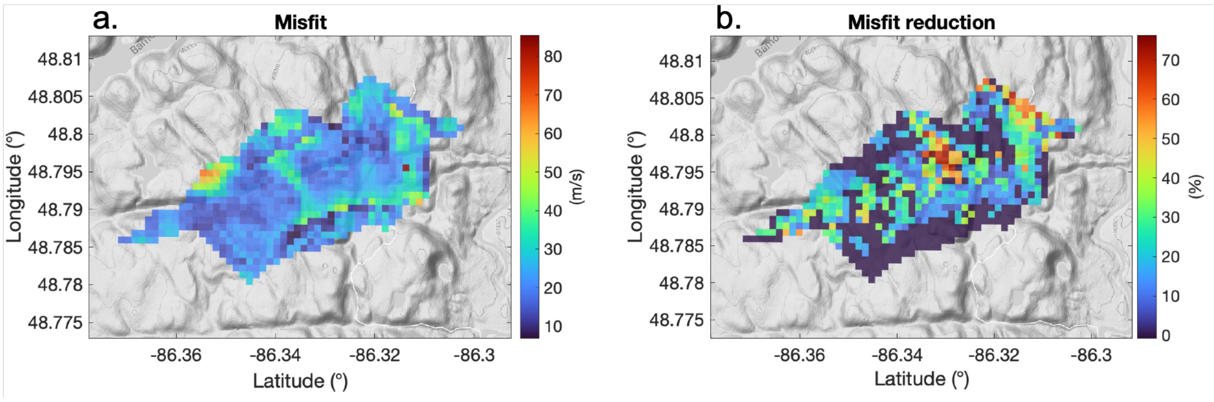

Figure 4.14: a) Misfit map of the final inversion at each cell. b) Map of the misfit reduction between the NA inversion and the linear inversion. Note that the cell misfits reduction equal to zero are regions for which the linear inversion did not occur (see detail in the text).

at 700m. In addition to these two numerical artifacts, the model eastern and western ends corresponding to the dense line are unreliable. These regions in the model are not constrained anymore, and the edge effects mainly control them. We assume that the rest of the structures are linked to a physical signature and can be trusted.

Figure 4.16 show five vertical profiles in the velocity model. These profiles were chosen, taking into account previous geological information on the medium. Generally, the profiles show a homogeneous medium as expected in Marathon deposit due to poor velocity contrasts. Even so, some structures are visible. The profile AA' which follows the dense line, shows a fast structure dipping the first meters and then stabilizes horizontally after 500m, and there is an inversion in the velocities. The profile BB' located at supposed the gabbro intrusion position shows a high velocity as expected for gabbro unities. The EE' profile is homogeneous, and in DD' profile, we see an inversion in the velocity at depth as for the AA' profile. Finally, the CC' profile shows two layers, the fast one at depth.

4.7 Geological interpretation

Marathon deposit presents three main geological units: the Syenite, the footwall, and the gabbro intrusion. The gabbro intrusion consists of different types of gabbros; our method does not allow us to differentiate between them because their S-velocity are similar, so we will assume one gabbro unite that is discernible using seismic noise. The final S-wave velocity model obtains by cross-correlating train signals and extracting surface wave present structures comparable with the geological units in the region.

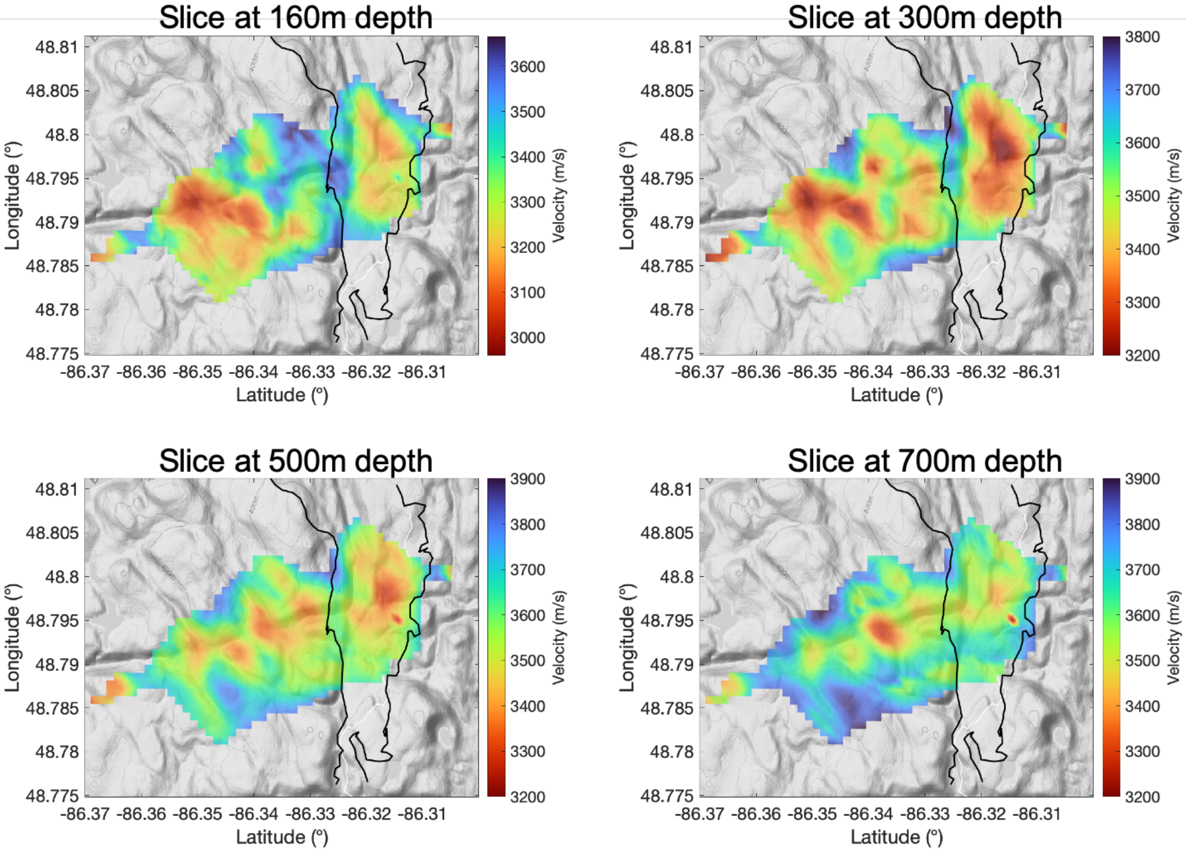

Figure 4.15: Horizontal slices through the S-wave velocity model at four different depths as shown in each figure title. Note that the color bar varies for each slice to accentuate the velocity contrast. The black line is the gabbro structure at the surface.

Figure 4.17 shows the geological interpretation of one horizontal and two vertical slices of the final S-wave velocity model. The Gabbros are defined by faster S-wave velocities (⩾3450m/s), .and the archaean granitoids defined by slow S-wave velocities (≈3200m/s). In figure 4.17-s regions below the dense line edges were omitted for interpretation because the edge effects are strong in this area. The pick dashed line represents the contact between gabbro and the footwall, and the black dashed line the upper contact with the Syenite that is consistent with geological observations (e.g., Good et al., 2015). We can notice that the lower gabbro contact (figure 4.17-a pink line) has an important dipping angle compare with the upper contact. The upper contact (black dashed line) has a sharp dipping angle for the first 500m after the dipping angle is less strong; The slices CC' shoes that deeper than 500m the high-velocity zone is located to the south west. The black line (Syenite-Gabbro contact) in the figure 4.17-a shows a slow velocity anomaly at depth; this zone might correspond to syenite structure as defined by drilling. Thus, the S-velocity model agrees with the area's previous knowledge and the data obtained by drilling. It also provides new clues about the complexity of the structure in-depth, which does

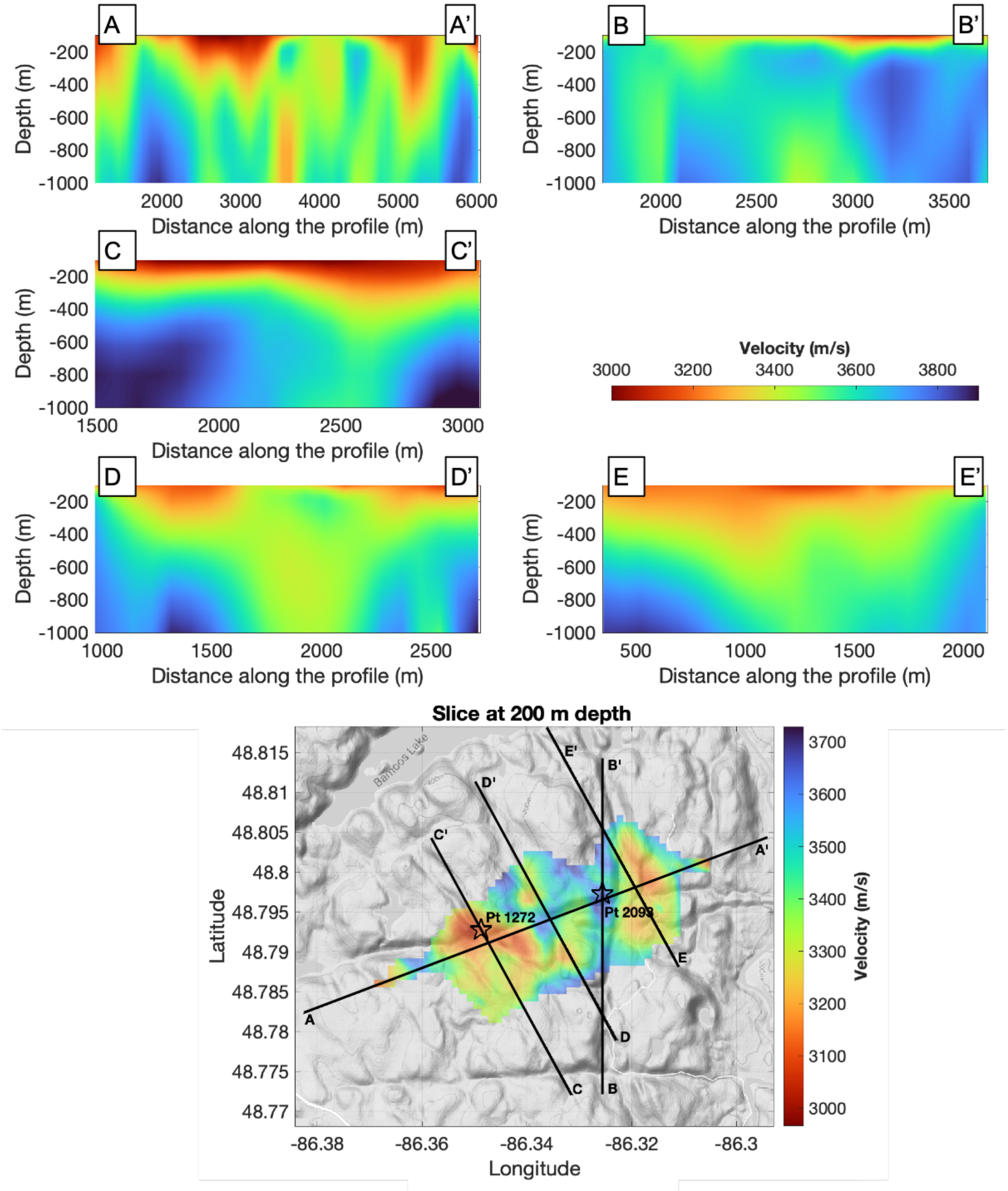

Figure 4.16: Upper panels: five vertical slices across the 3D velocity from 100m to 1km depth. The locations of the slices are presented by the black in the lower figure. Lower panel: horizontal slice at 250mdepth. The stars represent two cell position for which the 1D inversion is detailed. Note that the colorbar for the vertical slices and the horizontal slice is different.

not only present three blocks (as in the modeling) but a greater complexity and presence of the low-velocity structures in the gabbro intrusion.

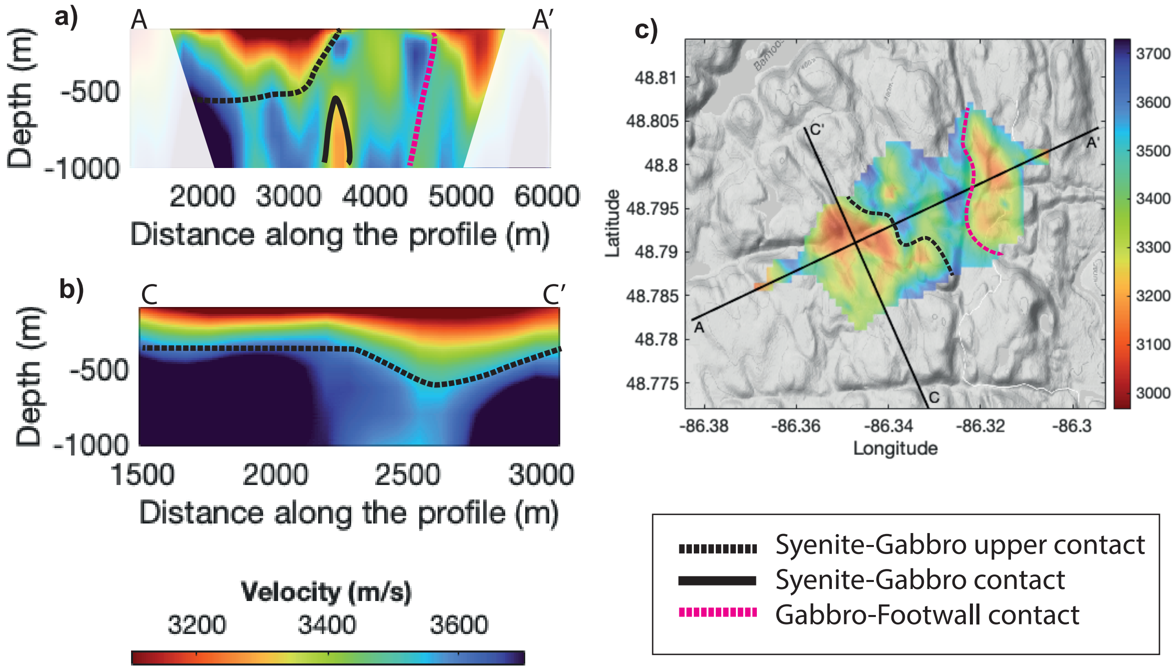

Figure 4.17: Vertical and horizontal slices across the 3D velocity with the geological features interpreted

4.8 Comparison with continuous data cross-correlation

This section will compare the difference between the cross-correlation of train periods using the processing proposed in this thesis and the standard correlations computed following the standard cross-correlation approach. The standard cross-correlations were computed using the 30 days of continuous data sampled at 50Hz. The data was cut in segments of 30 minutes long, the cross-coherence computed each segment, then stacked by the station, and the causal and acausal parts were averaged. Finally, the virtual source gathers were reconstruct using the individually stacked cross-correlation as shown in the figure 3.16 (previous chapter).

4.8.1 Eikonal Tomography

To compare the effect of both approaches on the eikonal tomography, we applied to the standard cross-correlation the same the eikonal tomography processing as for the train data. It is important to note that only the $SNR_{threshold}$ parameter for the data rejection was changed to 3 because standard cross-correlation has a lower SNR distribution. This $SNR_{threshold}$ value allows us to keep the data's correct quality while conserving enough traces to track the wavefront.

Figure 4.18 upper panels show the phase velocity maps obtained from one virtual shot gathered. Figure 4.18-a is resulting from the standard cross-correlation and figure 4.18-b from train periods correlation. It can be seen that the phase velocity map from standard cross-correlation is considerately noisy, and there is not a clear anomaly that could be interpreted as a geological structure. Otherwise, for the train period phase velocity map, a high anomaly is perceivable at the expected position of the gabbro. Even if the phase velocity map of the train period is still relatively noisy and it is necessary to stack the measurement from all virtual sources and do the statistical rejection, we can already see that the standard correlations lower quality data has a direct implication in tomographic methods as the eikonal tomography.

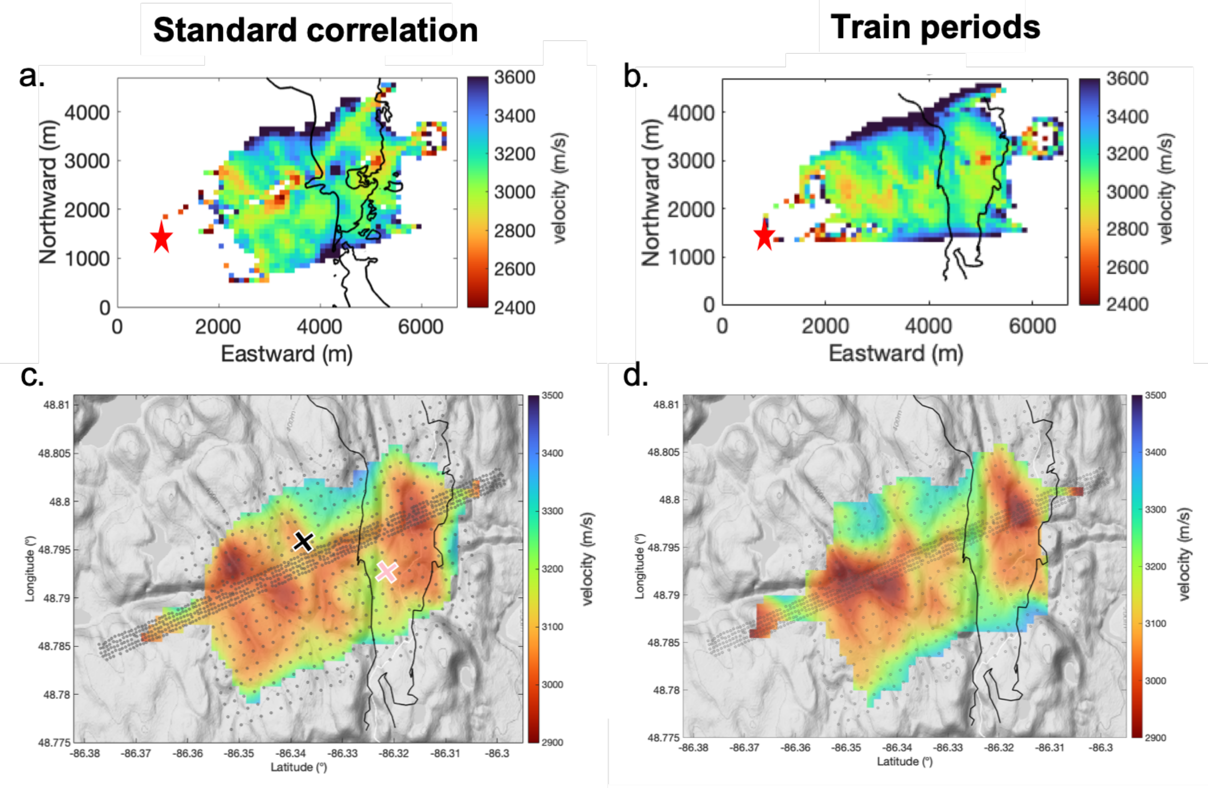

Figure 4.18: Eikonal tomography at 4.5Hz. Upper panels: phase velocity maps for a single virtual source represented as a red star red start resulting from blind cross-correlation (a) and from cross-correlation of train periods (b). Lower panels: final phase velocity maps obtained using all the virtual sources resulting from blind cross-correlation (c) and from cross-correlation of train periods (c). The black line represent the Gabbro instruction. Gray dots are the nodes.

Then we compare the final velocity maps (figure 4.18 lower panels). The final maps are similar for both cases; the eikonal approach converges to similar structures after the stack over all virtual sources. The high-velocity anomaly related to the gabbro intrusion is observed in both cases. However, 4.18-c obtained from standard correlations presents a low-velocity artifact at the dense line probably related to the nether data quality. We observe that although we have fewer data and less azimuthal coverage for train signals, the contrast and the structure are a little more visible than for standard cross-correlations because of the quality of the train correlations.

4.8.2 Anisotropy

Besides allowing us to obtain topographic maps of isotropic seismic velocities, the Eikonal tomography also brings the possibility to access the azimuthal anisotropy of surface waves in a relatively easy way. By applying the principle of Eikonal tomography described in the section 4.2, we obtain an estimate of the local phase velocity (the amplitude of gradient of the travel time) as well as the local direction of propagation of the wave (the direction of the gradient of the travel time). The computation of the equation 4.2 for every virtual source provides the slowness direction that represents the local wave propagation, and their amplitude which is equal to the local phase velocity (Lin and Ritzwoller, 2011; Mordret et al., 2013a). By combining these individual measurements for each station, the phase velocity variations as a function of azimuth can be estimated. However, to have an accurate estimation of the anisotropy, a good azimuthal discretization is needed. The cross-correlations from trains signals are constraints in azimuth because only station pair inline with train position are used reducing the azimuthal illumination. For this reason, we will only use the standard correlations to study the anisotropy in the Marathon deposit.

Figure 4.19-a, shows the example of three virtual sources (green, red, and blue triangles) and the corresponding local wave propagation direction, i.e., the travel time gradient (arrows with the same colors). The thick lines represent the local phase velocity and direction of wave propagation inferred from the three virtual sources at a single spatial node of the model (black square figure 4.19-a). Then, using the relationship between the local phase velocities and local propagation directions (explained below), we compute the azimuthal anisotropy at each point of the grid.

Smith and Dahlen (1973) showed that the relation between the azimuth, ψ, and surface-wave phase velocity c is in the form of an even order sinusoid with a 180° and 90° periodicity. This relation is frequency-dependent and is valid for a slightly anisotropic medium. This relation represents as a truncated Fourier series analysis of the data is defined as:

$$c(\psi) = c_0 + \frac{A'}{2c_0} cos[1(\psi - \phi_1)] + \frac{B'}{2c_0} cos[2(\psi - \phi_2)] + \frac{C'}{2c_0} cos[3(\psi - \phi_3)] + \frac{D'}{2c_0} cos[4(\psi - \phi_4)] \quad (4.12)$$

Where c_0 is the average phase velocity for one station, and ψ angle is measured positive

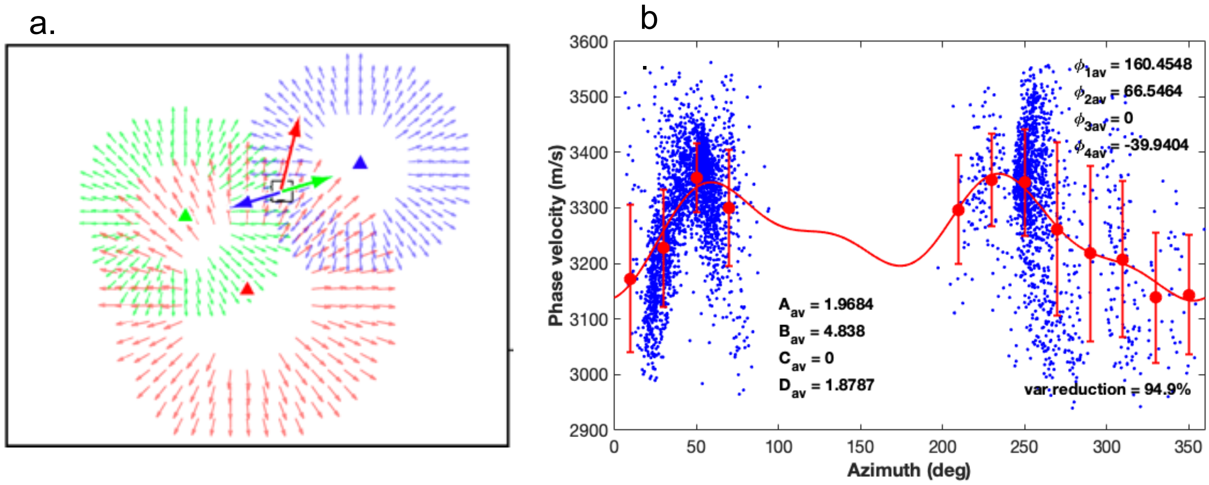

Figure 4.19: a) Schematic diagram illustrating the anisotropy phase velocity measurements ($c(\psi)$) at one spatial node of the model (small black square) and the related to 3 surrounding stations or virtual sources (colored triangles). Each colored vector field shows the velocity field of the waves emitted by the corresponding colored virtual source. The vectors in the particular cell are enlarged for clarity. Modified from Mordret et al. (2013a). b) Azimuthal distribution of the phase velocity at 4.5Hz for the spatial node represented as a black cross in the figure 4.18-a. The blue dots are the phase velocity measurements, and the red dots represent the phase velocity averaged over 20° with the corresponding error bars. The red curve best fits the 1ψ , 2ψ, and 4ψ azimuthal variation for the averaged velocity measurements. The values of the fitted parameters and the variance reduction are shown.

clockwise from north. $A = A'/c_0$, $B = B'/c_0$,$C = C'/c_0$ and $D = D'/c_0$ are the peak-to-peak relative amplitude of the 1ψ, 2ψ, 3ψ and 4ψ terms and ϕ_1, ϕ_2, ϕ_3, and ϕ_4, define the orientation of the fast axes for the the 1ψ, 2ψ, 3ψ and 4ψ terms, respectively. 1ψ represents the isotropic therm of the velocity. However, this component is not physical at a local point because the velocity is constant locally. Each 1ψ and 3ψ violates the reciprocity principle, and they can not be considered as physical parameters (Lin and Ritzwoller, 2011). The odds components 2ψ and 4ψ are the anisotropic terms.

To obtain the phase velocity in function on the azimuth, we follow the study proposed by Lin and Ritzwoller (2011). First, we use the final phase velocity measurements (i.e., after the statistical selection explained in the section 4.3.3). Then, we extract the phase velocity and the azimuth for each spatial node (i.e., cell of the medium) and the eight surrounding cells; if one of the surrounding cells does not contain data, the spatial node is rejected. The blue dots in the figure 4.19-b represent an example of the phase velocity distribution in function of the azimuth for a cell located at the black cross figure 4.18-a. Next, the measurement are binned each 20° (figure 4.19-b red dots). Measurements are rejected if the data does not span at least four times 45° quadrants. This is done in order to conserve only robust observations for the 2ψ, which need

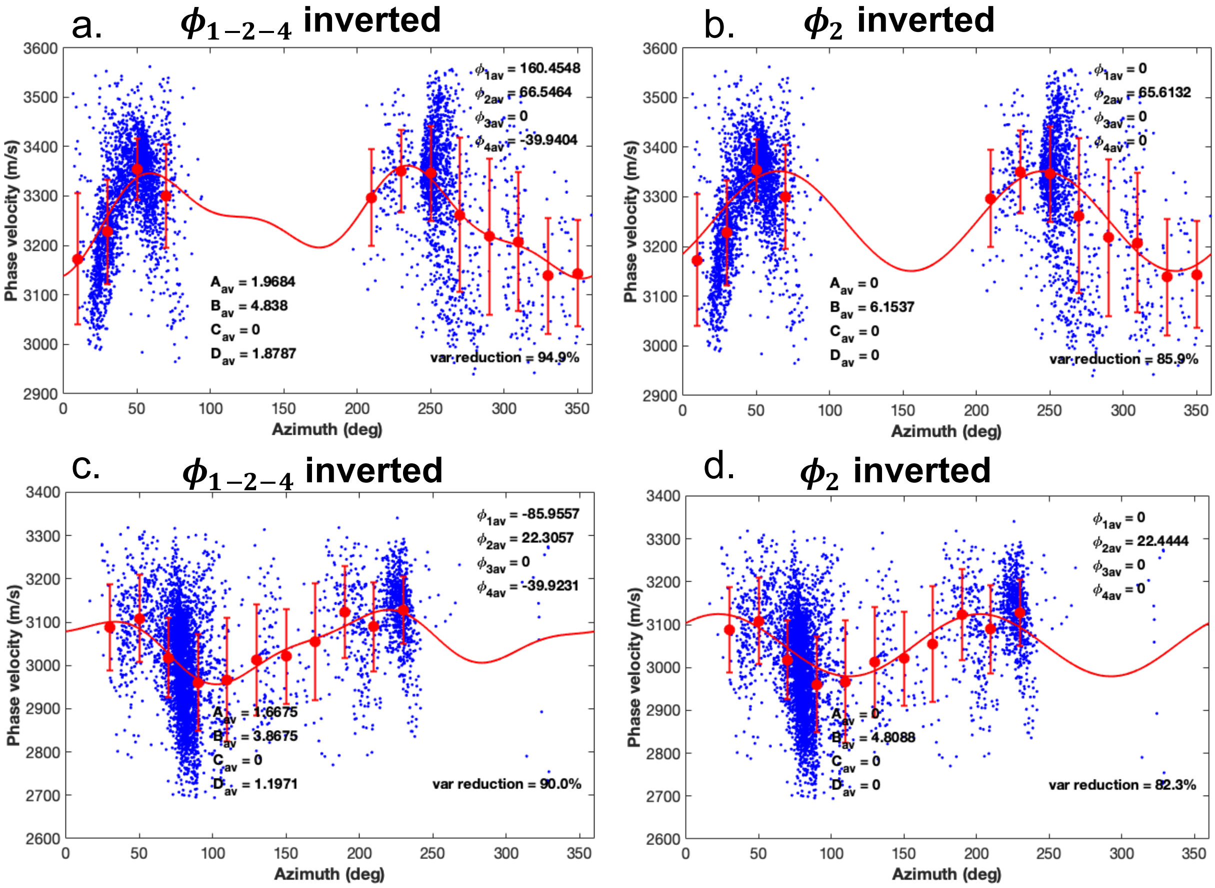

Figure 4.20: Azimuthal distribution of the phase velocity at 4.5Hz for the spatial node represented as (a-b) a black cross in the figure 4.18-a and (c-d) a pink cross in the figure 4.18-a. The blue dots are the phase velocity measurements, and the red dots represent the phase velocity averaged over 20° with the corresponding error bars. The red curve best fits inverting (a and c) 1ψ , 2ψ, and 4ψ therms and (d and d) ψ therm. The values of the fitted parameters and the variance reduction are shown for each case.

enough measurements to cover at least 180°.

It is possible to invert 1ψ, 2ψ and 4ψ therms or only the anisotropy therm 2ψ. Figure 4.20 shows two examples for each inversion at two different cells. Adding the 1ψ and 4ψ term to the inversion does not improve the fitting notable, and the 2ψ value remains the same for both cases. Therefore, we decided to fit the binned velocities with the weighted least-square only for the 2ψ (figure 4.19-bred line).

The misfit is evaluated and defined as the standard deviation between the measured and the predicted phase speeds. For interpretation, only the spatial nodes a misfit smaller than 30 m/s and a variance reduction greater than 70% are conserved. Figure 4.21 shows the misfit and variance reduction map (a and b respectively), the red arrows show the limit values for which we consider the azimuthal values reliable measurement (i.e., sufficient constrain).

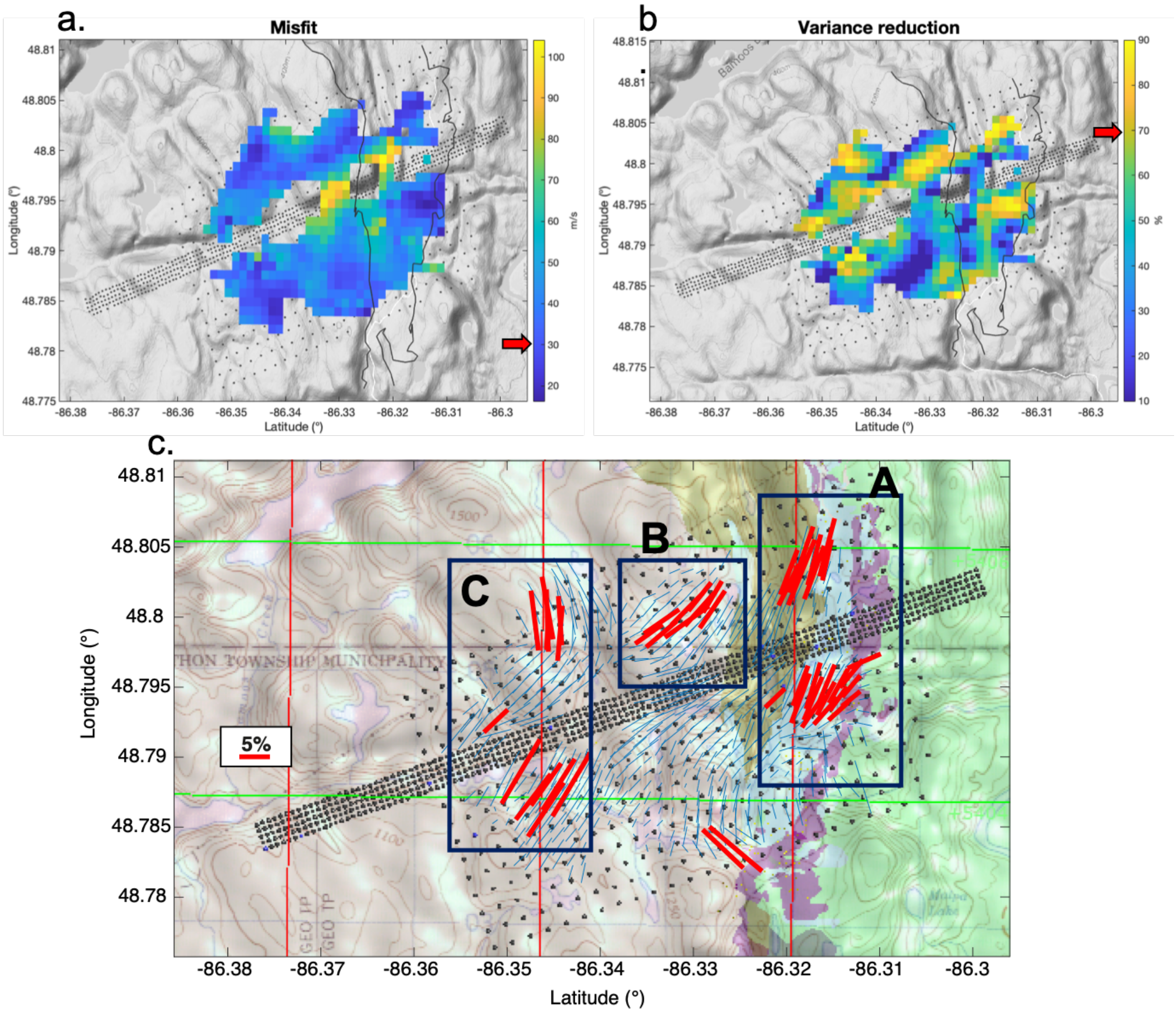

Figure 4.21: a) Misfit map of the 4.5Hz surface wave azimuthal anisotropy inversion. b) Variance reduction map of the 4.5Hz surface wave azimuthal anisotropy inversion. c) Geological map at Marathon with 2ψ azimuthal anisotropy at 4.5Hz. The 2ψ amplitude and fast direction are summarized by the orientation and the length of the bars. Blue bar represents all the 2ψ azimuthal anisotropy measurement, and red bars are the reliable measurement (see details in the text).

Figure 4.21-c shows the geological map of Marathon at the surface and the fast axis and amplitude of the 2ψ anisotropy represented by the orientation and the length of the bars obtained. The anisotropy is evaluated at 4.5Hz which is equivalent in depth to ≈200m (as shown in the kernels figure 4.13-i and -j). The blue bar represents all anisotropy measurements, and the red bars are reliable measurements (i.e., regions with a small misfit and high variance reduction).

From the 2ψ anisotropy, we discern three regions (the three rectangles in the figure 4.21-c). At rectangle A the anisotropy orientation is North-south, which follows the gabbro intrusion orientation. We suggest that the anisotropy in this region is related to the gabbro intrusion. In rectangle B, the anisotropy direction changes, presenting now a northeast-southeast direction. This direction could be due to the Gabbro-Syenite contact that alters the structures at depth. Finally, zone C is entirely in the Syenite. The anisotropy orientations in this region are correlated to the topography at the surface. We can not make a straightforward interpretation; however, topography at the surface may be linked to compression mechanisms at depth that would also alter the geological structures, generating this anisotropy direction for some regions of the Syenite unit.

4.8.3 Ambient noise surface wave tomography (ANSWT) results from Sisprobe

Sisprobe carried out the results present in this subsection; the objective of including them in this manuscript is to compare the imaging results generated by the standard cross-correlation approach and the cross-correlation of train periods. Therefore, we will first briefly present the data process and then the results.

The cross-correlations used for the ANSWT were the standard cross-correlation described earlier in the section. Phase and group velocity dispersion curves were computed for each correlation using automatic algorithms (FTAN algorithm (Mordret et al., 2015) for the group velocities, and the algorithm developed by Ekström et al. (2009)). 70% of the dispersion curves were conserved after the quality control. The dispersion curves are regionalized into a regular grid of 111.2m by 73.5m cells (in North and East directions) as described in Mordret et al. (2013b). Finally, group and phase velocities were jointly inverted at every single cell using NA algorithm (see section 4.5). The NA inversion parameters used are described in the table 4.4. Note that a constant model was defined for this inversion, and the cubic splines basis was superimposed to create velocity anomalies.

Parameter	**Allowed range**
Velocity in the first layer	[2000- 3100] m/s
Velocity in the last layer	[2900 - 4100]
S1	[50 - 750] m/s
S2	[300 - 1150] m/s
S3	[350 - 1550] m/s
S4	[100 - 2000] m/s
S5	[1000 - 2300] m/s

Table 4.4: Ranges Allowed for the NA inverted parameters used by sisprobe

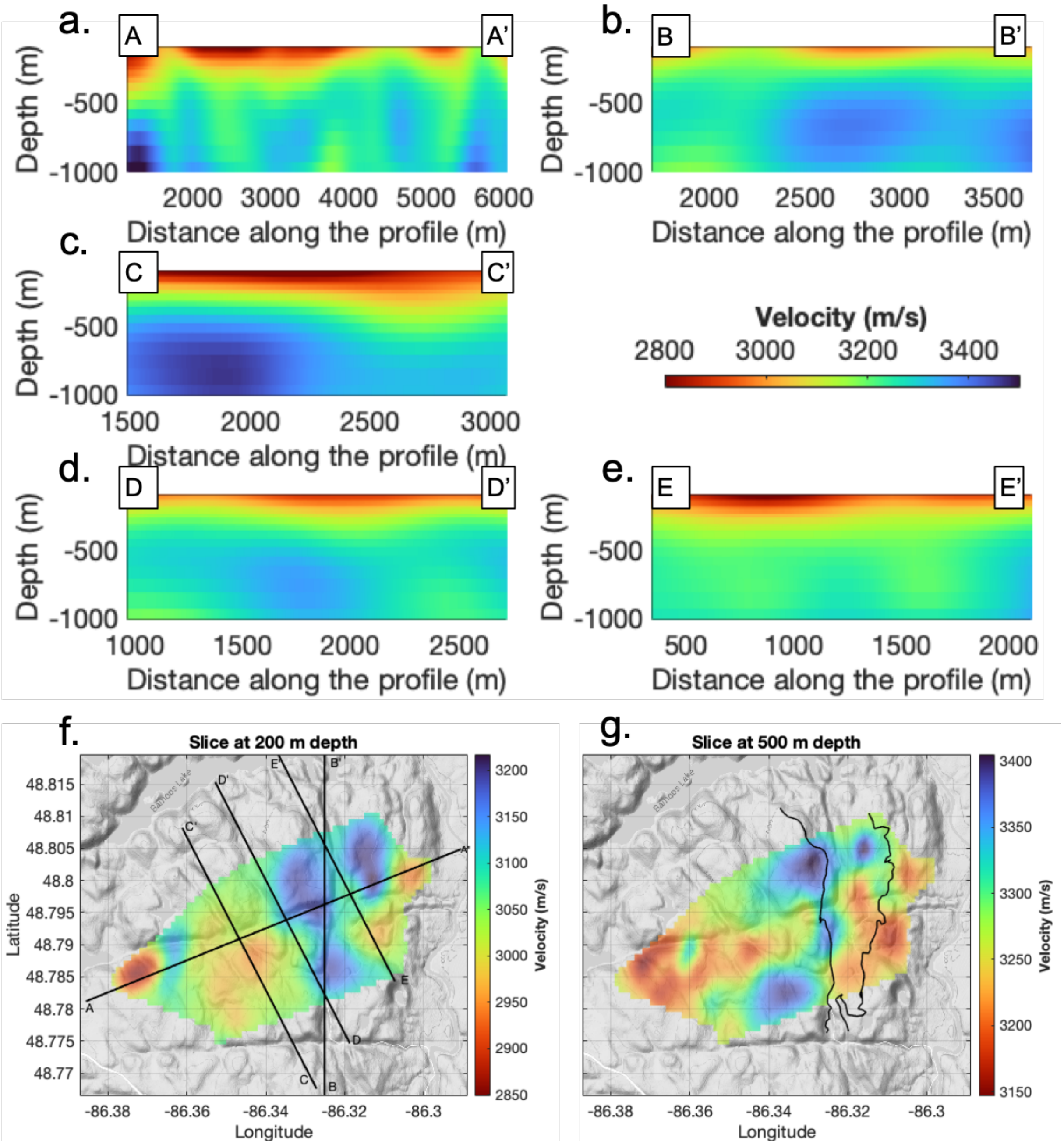

Figure 4.22: a) to e) Vertical slices across the 3D velocity from 100m to 1km depth. The black line presents the locations of the slices in the figure f. f) and g)Horizontal slices through the S-wave velocity model at 200m and 500m depth as shown in each figure's title. Note that the color bar varies for each slice to accentuate the velocity contrast. g) The black line is the gabbro structure at the surface.

The V_S model was generated by averaging the 300 best models (with the lowest misfits). Figure 4.22 shows two horizontal slices of the S-wave velocity model at 200m and 500m depth and the same five vertical profiles that showed for our approach (figure 4.16).

We observed comparable structures. The profile AA' shows for both cases a high-velocity anomaly at the west (A') and a low-velocity zone appears at 500m depth. The BB' profile presents a fast structure located at the North and the CC' to the south; both profiles BB' and CC' from both approaches show similar fast velocity anomalies. Since profile EE' is very homogeneous, it is difficult to highlight any structure. On the other hand, the DD' profiles do not correspond; there is an inversion between the fast and slow zones. Although the structures are relatively similar and consistent, it should be noted that the absolute value of the velocities is significantly different (≈300m/s).

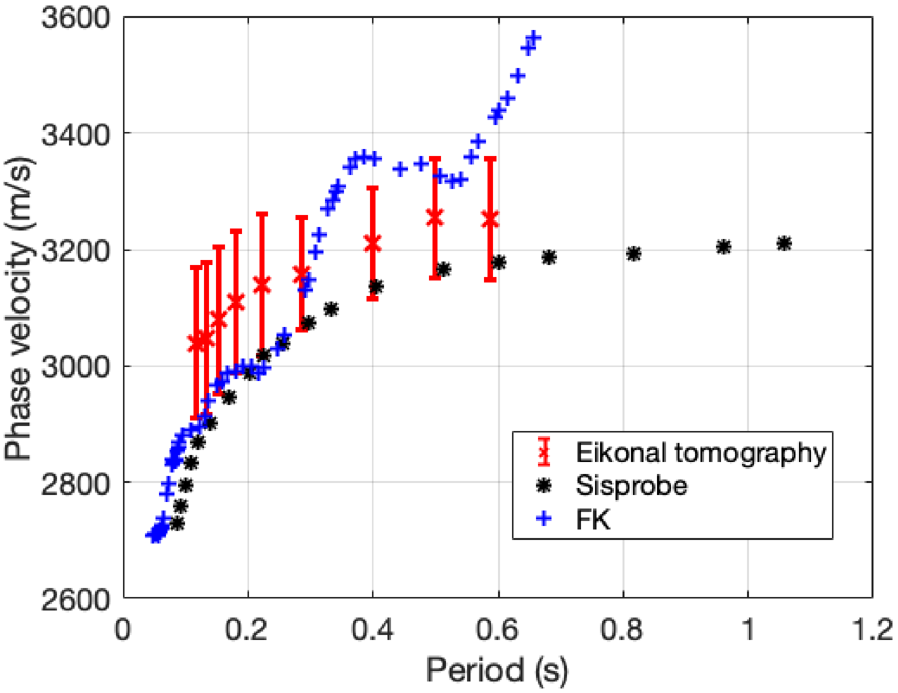

Figure 4.23: Averaged phase dispersion curve obtained using three different approaches. In red dispersion curve obtained from the eikonal tomography. In black, the dispersion curve obtained by Sisprobe and in blue applying FK analysis (see details in the text).

To understand these differences, we plotted the average dispersion curves obtained using different methods. We compare the dispersion curves obtained using train signals correlations and the standard correlation. For the standard cross-correlation, two methods were used the automatic approach picking applied by Sisprobe and FK approach. The frequency-wavenumber (F-K) analysis was implemented by Teodor et al. (2021). They made the FK analysis of sub-arrays in the dense line using the same correlation dataset. They computed an average dispersion curve considering only the Rayleigh wave propagation's fundamental mode. Figure 4.23 shows the average dispersion curves obtained with the three methods. We observe that the dispersion curves obtained by Sisprobe and our method are consistent, but our dispersion curve is shifted to 100m/s. On the other hand, the curve obtained with the FK method is consistent with the other curves at low periods (<0.3s) but significantly different at large periods. We notice the strong dependency of phase velocity values on the used method in a general way. Specifically, the velocity differences at the tomography resulting from both methods could be due to the method used to obtain the phase velocity. Moreover, the differences are accentuated by the different smoothing parameters for the velocity maps construction and inversion methods.

In any case, the general structures obtained with both approaches are indicators of physical structures at depth related to geological units. Furthermore, both approaches converge to similar regions of fast velocity related to the gabbro intrusion, which has a sharp dipping angle ($\approx 60°$) the first 500 m, and the angle decreases drastically. However, we should be cautious with the direct analysis of absolute velocity values because they are method-dependent.

4.9 Conclusion

In this chapter, we apply common surface-wave tomographic methods to surface waves retrieved using train-generated seismic signals. We first sought to obtain a 3D shear wave velocity of the region under study. For this, we applied Eikonal tomography; we obtained phase velocity maps that we then inverted using a probabilistic approach followed by a linear inversion. This workflow allowed us to obtain a robust model of the studied region, which agrees with the geological background and the drills information. We also provided new information about the geometry of the Gabbro intrusion, more specifically on the upper contact and the dipping angles of both regions.

Using synthetic data, even in a simpler context than expected in Marathon, helped us to understand the content of our virtual shot gathers. The synthetic data showed that obtaining reflected waves is very challenging in a context like Marathon, where velocity contrasts are very slight. The numerical study also allows us to tune some Eikonal tomography parameters and analyze the resolution limits of the method in the case of Marathon, thus confirming that Eikonal tomography was an adequate method for viewing structures in a context such as Marathon.

This chapter also analyzed the pros and cons of the train-generated signal approach for seismic imaging. For this, we compared the correlations and final topographies obtained with our approach and with a more classical passive seismic interferometric approach. Using the Eikonal approach, we observed that trains signals shot gathers allow us to obtain a phase velocity map using a single virtual source, whereas standard correlations shot gathers need to stack over several virtual sources to converge to a constructive image. However, in both approaches, robust and consistent phase maps were obtained, concluding that the use of train signals to obtain virtual shot gathers allows us to increase the quality of the data, which is then reflected in a decrease of the measurement errors when common imaging methods are applied.

On the other hand, using local noise sources such as trains restricts the illumination of the region. In Marathon, the railroad covers approximately 130°, which allows having a correct illumination range; however, this restricts certain studies that need a better illumination, such as those of anisotropy for which it was necessary to use the standard correlations. In a more general way, these limits for imaging application are case dependent (depends on the source and the imaging target)

Conclusion and perspectives

Marathon deposit

This research work aimed to perform high-resolution imaging at the Marathon mineral deposit using passive seismic techniques. To do so, we use 30 days of continuous seismic noise recorded by 1020 vertical component geophones. First, we explored the distribution of noise sources around the array and characterized them in time and frequency to find the most suitable for high-resolution imaging. We found that the most persistent and powerful source of seismic noise in Marathon was the freight trains passing nearby the array. Using numerical modeling and observations of train seismic signals at Marathon, we described and characterized the source mechanisms, concluding that long, heavy, fast trains generate more energy. Train-generated energy is the high frequency range between [2 - 30]Hz, and train's speed variation generates broadband energy in Marathon.

Considering that trains are moving and extended sources, we can not apply standard passive seismic interferometry. Thus we developed a methodological framework focusing on the stationary zones and propose a signal processing strategy focusing on high-frequency retrieval. This workflow is based on five steps. First, we identified the train in the continuous data, and we located them. Using array methods (beamforming and covariance matrix), we created a detailed catalog of train passages and localized them by the minute. Second, we divided the train signals into one-minute windows, and we selected station pairs for which the train is in a stationary phase zone for each one-minute window. Third, we computed cross-correlations after proper time windowing and station pairs selection. Fourth, we stacked the cross-correlations over different events (i.e., trains). Finally, we reconstructed virtual source gather and retrieved three dominant arrivals; we suggest the arrivals are a P-wave, a mix of S- and surface waves higher modes, and the surface wave.

The next step was to apply the Eikonal tomography to the surface-wave by virtual shot gathers. To do so, we measured the phase velocity; we interpolated the measurement to a regular grid to then calculate the phase velocity by virtual shot gathers. Finally, we averaged the individual velocity maps to obtain the final phase velocity maps. Subsequently, the phase velocities are inverted in two steps. We first used a probabilistic inversion and then explored the neighborhoods of the more probable models with a linear inversion. The inversion of these phase velocity maps yielded a 3D shear velocity model that correlates very well with the region's geological structures and drill data while providing additional information on the Gabbro intrusion's dipping angle.

In order to evaluate the arrival time interpolation and efficiency of the Eikonal tomography, we compared our method with numerical models computed in a simplified version of the Marathon geology. Numerical modeling allows us to constrain the interpolation parameters and better understand the wavefield in a close but simplified Marathon deposit model. To judge our approach, we also compared our results with the standard passive seismic interferometry results. We found that individual phase velocity maps obtained by the train correlation are less noisy, allowing the recovery of sharper images. However, final velocity maps (averaged of all virtual sources) converged to a similar result. Since the standard correlations are noisier than train cross-correlation, the individual phase velocity maps are noisier, too. Nevertheless, we converge to a similar final result because we stack more data (more stations pairs per virtual source and more virtual sources). Besides, standard correlations use all station pairs improving the azimuthal coverage compare with our method. We, therefore, studied the azimuthal anisotropy of Rayleigh waves with the Eikonal tomography technique. We noticed a strong correlation between the gabbro intrusion and the directionality of the anisotropy in the gabbro. In the Syenite, the anisotropy is controlled mainly by topography. Finally, we compared the 3D models we obtained with the one obtained by Sisprobe using standard correlation. We noticed similar structures between the two models; however, a difference in the absolute velocity value is present ($\approx$300 m/s) due to the phase velocity measurement and inversion methods.

Generally, we implement an alternative method to use freight train signals. Trains are available, detectable, repeatable sources that generate high-frequency broadband energy for imaging in a mineral exploration context. We extracted high-frequency surface waves with a higher SNR than standard correlations and used them to obtain a 3D velocity model. We showed that it is possible to retrieve body waves using selective correlations contrary to the standard use of seismic noise. Although these P-waves may be observable in the individual virtual shot gathers (figures 3.16 and 4.24), their use for seismic imaging is not straightforward. The P-waves could not be used at Marathon for several reasons. First, P-waves have a lower quality

(SNR) and lower amplitude than surface waves, making it harder to extract and isolate from the rest of the record. Second, the source ringing is not random noise that hampers the precise analyses of the P-wave (as shown for the last two red traces in figure 4.24). Third, the spatial coherency of the P-waves is poor. Moreover, at short distances, the body waves are hidden by the surface waves; at distances shorter than 2km, it is not possible to discern the P-wave (first red traces in figure 4.24). Because of all these reasons, it was not possible to pick the P-arrival. However, advanced denoising methods could be explored in order to pick the P-wave for later use in classical imaging applications providing additional information about the geological interfaces at Marathon.

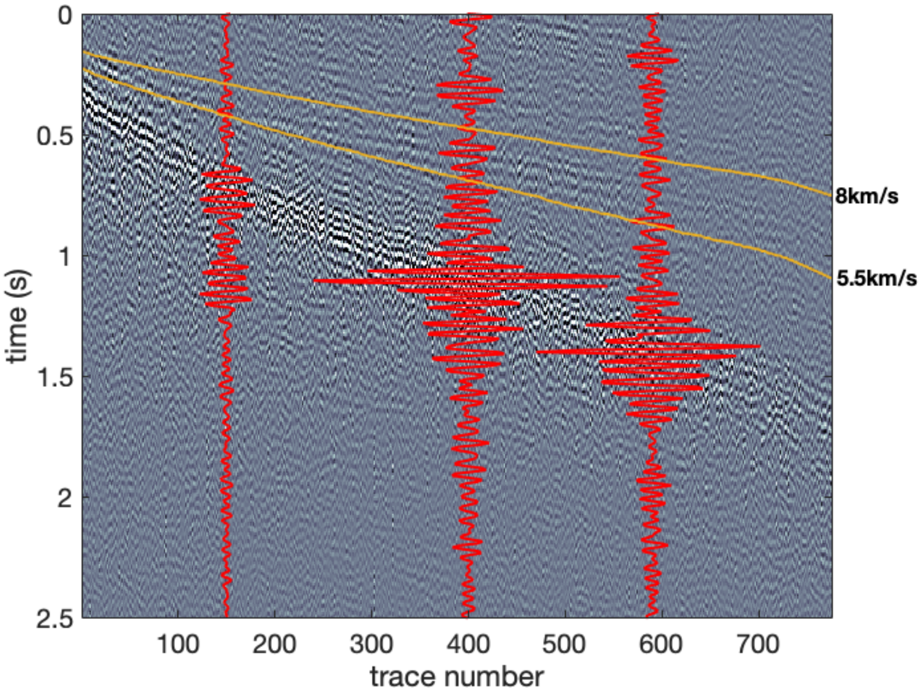

Figure 4.24: Virtual shot gather of station 01.01002 (figure 2.2). It is constructed from the stack of 60 train passages using all the azimuths and filtered between [15 - 30]Hz. Trace numbers are sorted by increasing distance from the virtual source with a not constant step. The yellow lines indicate travel times with constant velocities. We show three traces (red lines) at different inter-station distances to highlight the poor quality of the direct P-arrival.

In addition to the body waves, other waves were reconstructed using train signals. Figure 4.25-a shows the virtual shot gather for the dense line array reconstructed by stacking the five lines in the array. In this figure, we observe some coherent energy arriving with apparent negative velocities (red arrows). This arrivals were also observed for each line (figure 4.25-b) and in the synthetic data (figure 4.25-c). We suggest that their arrival could be related to reflections of surface waves on the contact geological structures and/or reflected by the topography. The surface waves were already used, but these reflections could be complementary as they would better delimit the contrasts between the geological blocks on the surface. However, a deeper study to determine the origin of these arrivals is essential before they can be used.

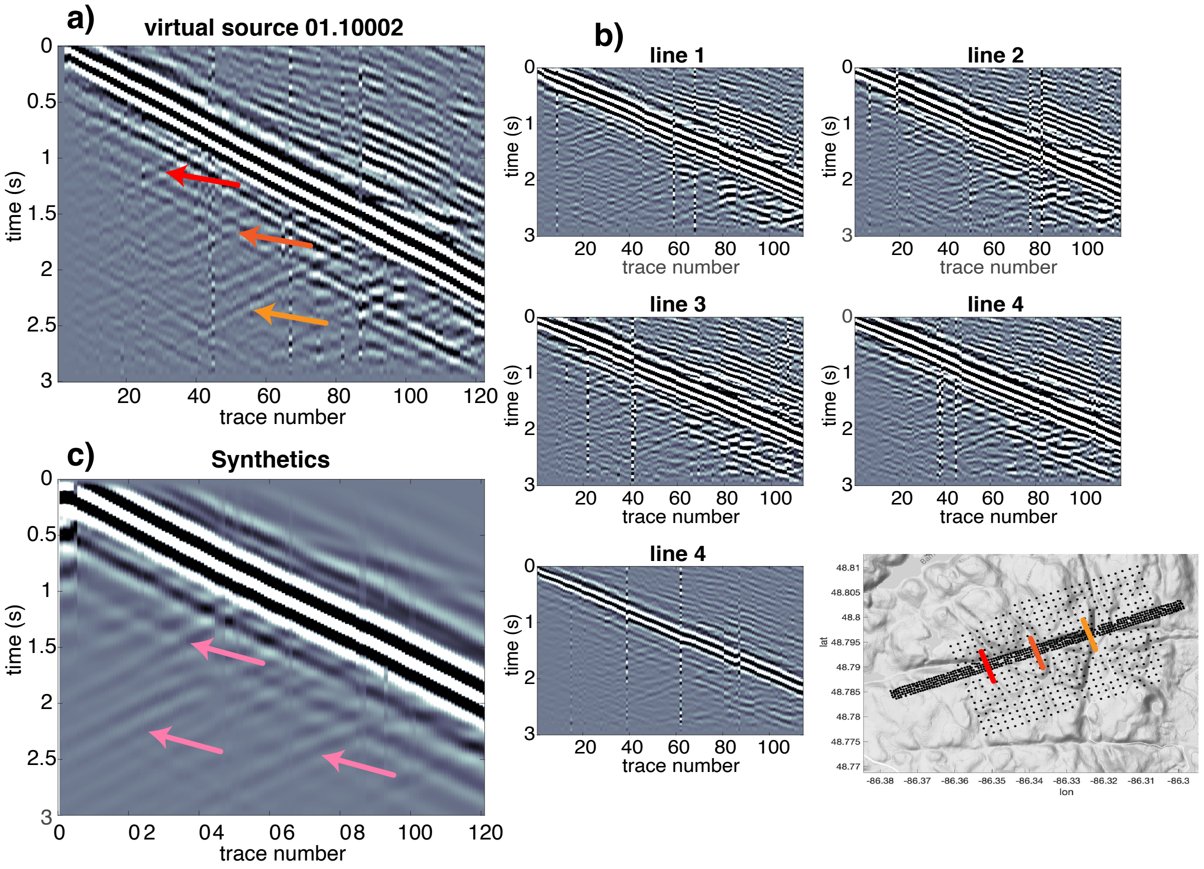

Figure 4.25: Virtual shot gathers of station 01.01002 for the dense line array (figure 2.2). a) Virtual shot gather reconstructed by stacking the five lines in the array. b) Virtual shot gather for each dense line as shown in the title of each panel. c) Synthetic virtual shot gather reconstructed by stacking the five lines in the array. a),b), and c) Trace numbers are sorted by increasing distance from the virtual source with a not constant step. The data was filtered between [5 - 12]Hz. a) and c) show the suggested surface waves reflections (highlighted by the arrows) for both real (a) and synthetic data (b). The colored lines in the map show reflection locations, the color corresponds to the arrow colors in panel a.

Train signals and others opportune sources of noise

In this work, we propose a processing strategy for applying seismic interferometry to train-generated signals, and it was successfully applied for seismic imaging in a mineral exploration context. Besides, train signals and the workflow applied here can be used in a different context. For example, we use freight and passenger train generated signals to imaging shallow structures in Dublin basin by Rezaeifar et al. (2020), and ballistic P waves were extracted using train signals for monitoring purposes at San Jacinto Fault by Brenguier et al. (2019); Pinzon-Rincon et al. (2021).

However, train tremors utilization for seismic interferometry thus depends both on detection limits (instrument sensitivity and local noise level) and on reliably recognizable features in train signals. Therefore, our ability to predict the long-range body-wave peak ground velocity (PGV) of a moving train tremor is crucial to image formation and monitoring any changes with seismic interferometry. Train signal modeling showed that the train geometry (wagon weight and train length) and the train's velocity control the PGV Lavoué et al. (2020). Moreover, the ground stiffness beneath railways controls high-frequency content and amplitude of excitation (trains traveling across a rock or stiff soil generate higher-frequency and higher-amplitude signals). This ground stiffness parameter may also reflect a coupling between the rail track and the ground.

Observation from previous studies Inbal et al. (2018); Brenguier et al. (2019) persuades us that 50km is a typical maximum distance range for detecting tremors generated from large North American freight trains. Although maximum detection distance may be limited (a few kilometers) in sedimentary basins due to attenuation and weak excitation, and it can reach almost 100km on a hard-rock substratum. This led us to look into the spatial extent of detectable train tremors in the entire contiguous United States plus southern Canada (figure 4.26). The map[1] displays the main freight railway routes. The swathes in colors represent high tonnage routes. Their width (100km) is a rough guide to the potential long-range train-tremor detection scope. This map does not consider the reduced detectability of signals in urban areas due to intense local noise or in sedimentary basins with strong attenuation. Annual freight tonnage [2] (figure 4.26) is a proxy for the number of trains traveling on rail sections. Assuming an average train length of 2km and a weight of 15kt, a tonnage of 100 Mt/yr corresponds to about 18 trains per day. The number of trains per day will affect the ability to stack the reconstructed body waves from seismic interferometry.

This map highlights the excellent opportunity to exploit any available massive freight train noise recovery to improve crustal imaging and monitoring dramatically. Potential applications may be in Cascade volcanos, southern California's San Andreas fault system, induced seismicity (e.g., Oklahoma), and resource exploration and monitoring (minerals and water). Moreover,

---

[1] The map was made with Natural Earth, free vector and raster map data are available at `https://www.naturalearthdata.com`.

[2] Figure is based on a map published by the U.S. Department of Transportation (`https://railroads.dot.gov/sites/fra.dot.gov/files/inline-images/0209.png`), built from the waybill samples 2010 established by the U.S. Surface Transportation Board.

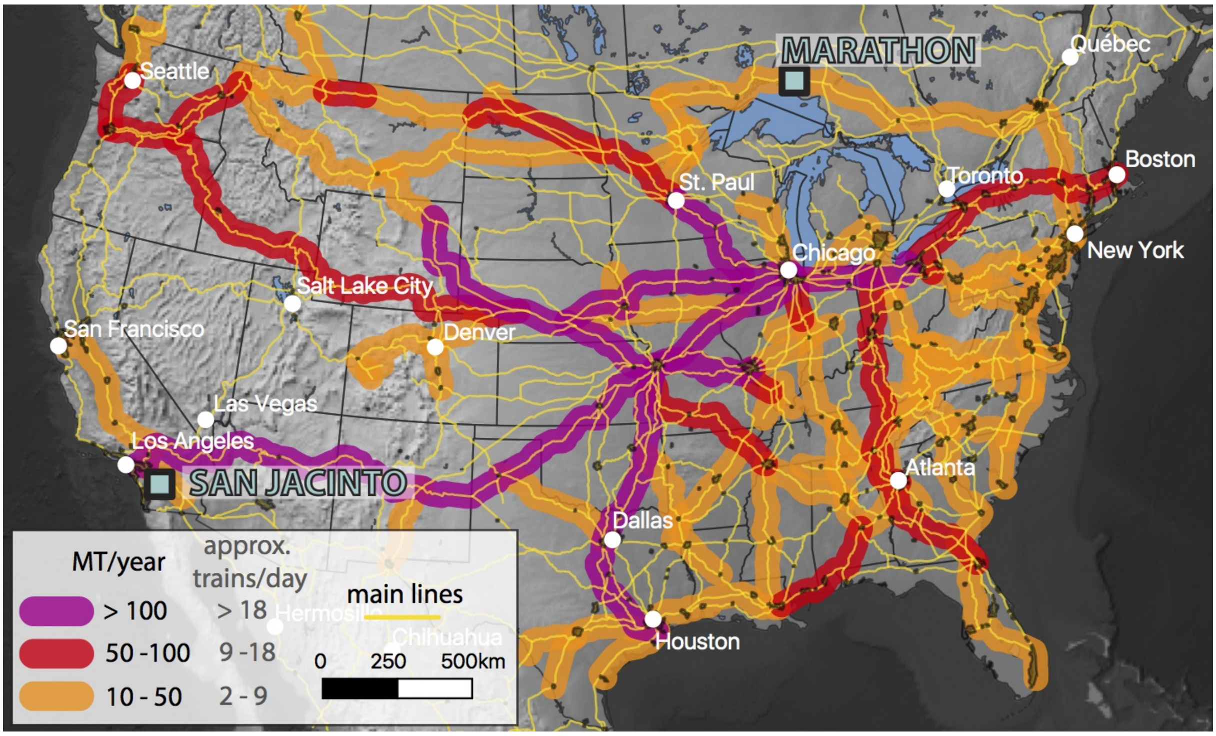

Figure 4.26: Regions of potential long-range train-tremor detection from the main railway route and annual tonnage information in North America. Colors represent annual freight tonnage, which is an indication of the number of trains traveling on the rail sections (sections with annual tonnage < 10 Mt/yr are ignored). Colored lines are 100km thick, indicating the distance from which we may detect individual train tremors 50km from the railway (see details in the text). Modified from Pinzon-Rincon et al. (2021)

distributed acoustic sensing data (Zhan, 2020) can be coupled to seismic interferometry of opportune sources as described by Dou et al. (2017) for car traffic and near-surface applications. This study indicates the potential for reconstructing widespread virtual sources along fiber optics from correlations of both short- and long-range opportune sources that open the path to many applications.

Beyond these regional applications in North America, trains can be used globally, especially in such countries and regions as China, Europe, Japan, and India. All have large freight railway systems, often with high-speed passenger lines too. Passengers trains are lighter than freight trains and generate less energetic tremors; thus, they can be used for near-surface environmental or engineering studies. However, the utilization of train signals as a source of seismic noise for imaging and monitoring reasons is limited to regions near the railway and requires trains traveling at relatively high speeds.

Although promising, this work poses significant, practical challenges that the field must confront. The most important is to improve understanding of the retrieved body and surface waves' spatial sensitivity to crustal structures for seismic interferometry with opportune sources. In contrast to actively controlled and placed sources, measurements of travel times or temporal travel-time perturbations using more irregular sources can improve sensitivity to the structure between the receivers and blur the overall picture due to interference between the noise source and the receivers. This latter downside may induce misleading interpretations of velocity or velocity change measurements.

In a general way, the workflow presented here in this work can be generalized to other local sources of cultural noise, including car and truck traffic, wind farms, and natural sources such as surf break or tectonic or volcanic tremor, which can be considered opportune sources for seismic interferometry.

Bibliography

Aki, K., 1957: Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute, University of Tokyo*, **35 (3)**, 415–456.

Ardhuin, F., L. Gualtieri, and E. Stutzmann, 2015: How ocean waves rock the earth: Two mechanisms explain microseisms with periods 3 to 300 s. *Geophysical Research Letters*, **42 (3)**, 765–772.

Arts, R., X. Zhang, A. Verdel, D. Santonico, J. Meekes, R. Noorlandt, B. Paap, and V. Vandeweijer, 2013: Experiences with a permanently installed seismic monitoring array at the co2 storage site at ketzin (germany).-a status overview. *Energy Procedia*, **37**, 4015–4023.

ARTTIC, 2018: Pacific website. PACIFIC consortium, URL https://www.pacific-h2020.eu.

Bakulin, A., and R. Calvert, 2006: The virtual source method: Theory and case study. *Geophysics*, **71 (4)**, SI139–SI150.

Beard, C., N. Arndt, F. Lavoué, and S. Beauprêtre, 2021: Sonic logs and passive seismic imaging at the marathon pge-cu deposit, ontario. *Core-Log-Seismic Integration Workshop.*

Behm, M., 2017: Feasibility of borehole ambient noise interferometry for permanent reservoir monitoring. *Geophysical Prospecting*, **65 (2)**, 563–580, doi:https://doi.org/10.1111/1365-2478.12424.

Behm, M., and R. Snieder, 2013: Love waves from local traffic noise interferometry. *The Leading Edge*, **32 (6)**, 628–632.

Bensen, G., M. Ritzwoller, M. Barmin, A. L. Levshin, F. Lin, M. Moschetti, N. Shapiro, and Y. Yang, 2007: Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, **169 (3)**, 1239–1260.

Boese, C. M., L. Wotherspoon, M. Alvarez, and P. Malin, 2015: Analysis of Anthropogenic and Natural Noise from Multilevel Borehole Seismometers in an Urban Environment, Auckland, New Zealand. *Bulletin of the Seismological Society of America*, **105 (1)**, 285–299, doi:10.1785/0120130288.

Bonnefoy-Claudet, S., F. Cotton, and P.-Y. Bard, 2006: The nature of noise wavefield and its applications for site effects studies: A literature review. *Earth-Science Reviews*, **79 (3-4)**, 205–227.

Boué, P., P. Poli, M. Campillo, H. Pedersen, X. Briand, and P. Roux, 2013: Teleseismic correlations of ambient seismic noise for deep global imaging of the Earth. *Geophysical Journal International*, **194 (2)**, 844–848, doi:10.1093/gji/ggt160.

Brenguier, F., N. M. Shapiro, M. Campillo, V. Ferrazzini, Z. Duputel, O. Coutant, and A. Nercessian, 2008: Towards forecasting volcanic eruptions using seismic noise. *Nature Geoscience*, **1 (2)**, 126–130.

Brenguier, F., N. M. Shapiro, M. Campillo, A. Nercessian, and V. Ferrazzini, 2007: 3-d surface wave tomography of the piton de la fournaise volcano using seismic noise correlations. *Geophysical research letters*, **34 (2)**.

Brenguier, F., and Coauthors, 2019: Train traffic as a powerful noise source for monitoring active faults with seismic interferometry. *Geophysical Research Letters*, **46 (16)**, 9529–9536.

Brocher, T. M., 2005: Empirical relations between elastic wavespeeds and density in the earth's crust. *Bulletin of the seismological Society of America*, **95 (6)**, 2081–2092.

Campillo, M., and A. Paul, 2003: Long-range correlations in the diffuse seismic coda. *Science*, **299 (5606)**, 547–549, doi:10.1126/science.1078551.

Castagna, J. P., M. L. Batzle, and R. L. Eastwood, 1985: Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. *geophysics*, **50 (4)**, 571–581.

Chamarczuk, M., M. Malinowski, and D. Draganov, 2021: 2d body-wave seismic interferometry as a tool for reconnaissance studies and optimization of passive reflection seismic surveys in hardrock environments. *Journal of Applied Geophysics*, **187**, 104 288.

Chang, J. P., S. A. L. de Ridder, and B. L. Biondi, 2016: High-frequency Rayleigh-wave tomography using traffic noise from Long Beach, California. *Geophysics*, **81 (2)**, B43–B53, doi:10.1190/geo2015-0415.1.

Chen, J. H., B. Froment, Q. Y. Liu, and M. Campillo, 2010: Distribution of seismic wave speed changes associated with the 12 may 2008 mw 7.9 wenchuan earthquake. *Geophysical Research Letters*, **37 (18)**.

Cheng, Y., and Coauthors, 2020: An automated method for developing a catalog of small earthquakes using data of a dense seismic array and nearby stations. *Seismological Society of America*, **91 (5)**, 2862–2871.

Chmiel, M., A. Mordret, D. Hollis, and N. Arndt, 2017: Stillwater, marathon pilot test for passive seismic imaging. Tech. rep., Sisprobe.

Chmiel, M., and Coauthors, 2019: Ambient noise multimode rayleigh and love wave tomography to determine the shear velocity structure above the groningen gas field. *Geophysical Journal International*, **218 (3)**, 1781–1795.

Claerbout, J. F., 1968: Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, **33 (2)**, 264–269, doi:10.1190/1.1439927.

Colombi, A., L. Boschi, P. Roux, and M. Campillo, 2014: Green's function retrieval through cross-correlations in a two-dimensional complex reverberating medium. *The Journal of the Acoustical Society of America*, **135 (3)**, 1034–1043, doi:10.1121/1.4864485.

Connolly, D., G. Kouroussis, O. Laghrouche, C. Ho, and M. Forde, 2015: Benchmarking railway vibrations — track, vehicle, ground and building effects. *Construction and Building Materials*, **92**, 64–81, doi:10.1016/j.conbuildmat.2014.07.042.

Dahl, R., D. H. Watkinson, and R. P. Taylor, 2001: Geology of the two duck lake intrusion and the marathon cu-pge deposit, coldwell complex, northern ontario. *Exploration and Mining Geology*, **10 (1-2)**, 51–65.

Dales, P., P. Audet, G. Olivier, and J.-P. Mercier, 2017: Interferometric methods for spatio temporal seismic monitoring in underground mines. *Geophysical Journal International*, **210 (2)**, 731–742, doi:10.1093/gji/ggx189.

Dales, P., and Coauthors, 2020: Virtual Sources of Body Waves from Noise Correlations in a Mineral Exploration Context. *Seismological Research Letters*, **91 (4)**, 2278–2286, doi:10.1785/0220200023, URL https://doi.org/10.1785/0220200023.

Davidson Jr, D. M., 1982: Geological evidence relating to interpretation of the lake superior basin structure. *See Wold & Hinze*, **1982**, 5–14.

De Ridder, S., and B. Biondi, 2013: Daily reservoir-scale subsurface monitoring using ambient seismic noise. *Geophysical Research Letters*, **40 (12)**, 2969–2974.

de Ridder, S., and J. Dellinger, 2011: Ambient seismic noise eikonal tomography for near-surface imaging at valhall. *The Leading Edge*, **30 (5)**, 506–512.

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

Diaz, J., M. Ruiz Fernandez, P. Sánchez-Pastor, and P. Romero, 2017: Urban seismology: On the origin of earth vibrations within a city. *Scientific Reports*, **7**.

Dobrin, M. B., 1953: *Introduction to Geophysical Prospecting.* American Association for the Advancement of Science.

Dong, S., J. Sheng, and G. T. Schuster, 2006: Theory and practice of refraction interferometry. *SEG Technical Program Expanded Abstracts 2006*, Society of Exploration Geophysicists, 3021–3025.

Dorman, J., and M. Ewing, 1962: Numerical inversion of seismic surface wave dispersion data and crust-mantle structure in the new york-pennsylvania area. *Journal of Geophysical Research*, **67 (13)**, 5227–5241.

Dou, S., and Coauthors, 2017: Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study. *Scientific reports*, **7 (1)**, 1–12.

Douze, E., and S. Laster, 1979: Seismic array noise studies at roosevelt hot springs, utah geothermal area. *Geophysics*, **44 (9)**, 1570–1583.

Draganov, D., X. Campman, J. Thorbecke, A. Verdel, and K. Wapenaar, 2009: Reflection images from ambient seismic noise. *Geophysics*, **74 (5)**, A63–A67.

Draganov, D., X. Campman, J. Thorbecke, A. Verdel, and K. Wapenaar, 2013: Seismic exploration-scale velocities and structure from ambient seismic noise ($>$ 1 hz). *Journal of Geophysical Research: Solid Earth*, **118 (8)**, 4345–4360.

Draganov, D., K. Wapenaar, W. Mulder, J. Singer, and A. Verdel, 2007: Retrieval of reflections from seismic background-noise measurements. *Geophysical Research Letters*, **34 (4)**.

Ekström, G., G. A. Abers, and S. C. Webb, 2009: Determination of surface-wave phase velocities across usarray from noise and aki's spectral formulation. *Geophysical Research Letters*, **36 (18)**.

Fan, W., J. J. McGuire, C. D. de Groot-Hedlin, M. A. Hedlin, S. Coats, and J. W. Fiedler, 2019: Stormquakes. *Geophysical Research Letters*, **46 (22)**, 12 909–12 918.

Fichtner, A., and V. C. Tsai, 2019: Theoretical foundations of noise interferometry. *Seismic Ambient Noise*, N. Nakata, L. Gualtieri, and A. Fichtner, Eds., Cambridge University Press.

Fink, M., 1997: Time reversed acoustics. *Physics today*, **50 (3)**, 34–40.

Forghani, F., and R. Snieder, 2010: Underestimation of body waves and feasibility of surface-wave reconstruction by seismic interferometry. *The Leading Edge*, **29 (7)**, 790–794, doi:10.1190/1.3462779.

Frankinet, B., T. Lecocq, and T. Camelbeeck, 2020: Wind-induced seismic noise at the princess elisabeth antarctica station. *The Cryosphere Discussions*, 1–16.

Froment, B., M. Campillo, J. Chen, and Q. Liu, 2013: Deformation at depth associated with the 12 may 2008 mw 7.9 wenchuan earthquake from seismic ambient noise monitoring. *Geophysical Research Letters*, **40 (1)**, 78–82.

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

Fuchs, F., G. Bokelmann, and the AlpArray Working Group, 2018: Equidistant Spectral Lines in Train Vibrations. *Seismological Research Letters*, **89 (1)**, 56–66, doi:10.1785/0220170092, URL https://doi.org/10.1785/0220170092.

Garofalo, F., and Coauthors, 2016: Interpacific project: Comparison of invasive and non-invasive methods for seismic site characterization. part i: Intra-comparison of surface wave methods. *Soil Dynamics and Earthquake Engineering*, **82**, 222–240, doi:https://doi.org/10.1016/j.soildyn.2015.12.010.

Good, D., R. Epstein, K. McLean, R. Linnen, and I. Samson, 2015: Evolution of the main zone at the marathon cu-pge sulfide deposit, midcontinent rift, canada: spatial relationships in a magma conduit setting. *Economic Geology*, **110 (4)**, 983–1008.

Gouedard, P., and Coauthors, 2008: Cross-correlation of random fields: Mathematical approach and applications. *Geophysical prospecting*, **56 (3)**, 375–393.

Gouveia, W. P., and J. A. Scales, 1998: Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research: Solid Earth*, **103 (B2)**, 2759–2779.

Green, D. N., I. D. Bastow, B. Dashwood, and S. E. Nippress, 2017: Characterizing broadband seismic noise in central london. *Seismological Research Letters*, **88 (1)**, 113–124.

Grobbe, N., and Coauthors, 2021: A multi-hydrogeophysical study of a watershed at kaiwi coast (o'ahu, hawai'i), using seismic ambient noise surface wave tomography and self-potential data. *Water Resources Research*, **57 (4)**, e2020WR029 057.

Halliday, D., A. Curtis, and E. Kragh, 2008: Seismic surface waves in a suburban environment: Active and passive interferometric methods. *The Leading Edge*, **27 (2)**, 210–218.

Haney, M. M., and V. C. Tsai, 2017: Perturbational and nonperturbational inversion of rayleigh-wave velocities. *Geophysics*, **82 (3)**, F15–F28.

Herrmann, R., and C. Ammon, 2004: Surface waves, receiver functions and crustal structure. *Computer programs in seismology, version*, **3**, 30.

Hillers, G., S. Husen, A. Obermann, T. Planès, E. Larose, and M. Campillo, 2015: Noise-based monitoring and imaging of aseismic transient deformation induced by the 2006 basel reservoir stimulation. *Geophysics*, **80 (4)**, KS51–KS68.

Huang, H., H. Yao, and R. D. van der Hilst, 2010: Radial anisotropy in the crust of se tibet and sw china from ambient noise interferometry. *Geophysical Research Letters*, **37 (21)**.

Huenges, E., and P. Ledru, 2011: *Geothermal energy systems: exploration, development, and utilization*. John Wiley & Sons.

Inbal, A., T. Cristea-Platon, J.-P. Ampuero, G. Hillers, D. Agnew, and S. E. Hough, 2018: Sources of Long-Range Anthropogenic Noise in Southern California and Implications for

Tectonic Tremor Detection. *Bulletin of the Seismological Society of America*, **108 (6)**, 3511–3527.

Johnson, C. W., H. Meng, F. Vernon, and Y. Ben-Zion, 2019: Characteristics of ground motion generated by wind interaction with trees, structures, and other surface obstacles. *Journal of Geophysical Research: Solid Earth*, **124 (8)**, 8519–8539, doi:10.1029/2018JB017151.

Jousset, P., and Coauthors, 2016: Seismic tomography in reykjanes, sw iceland. *European geothermal congress 2016*.

Kang, T.-S., and J. S. Shin, 2006: Surface-wave tomography from ambient seismic noise of accelerograph networks in southern korea. *Geophysical Research Letters*, **33 (17)**.

Kedar, S., M. Longuet-Higgins, F. Webb, N. Graham, R. Clayton, and C. Jones, 2008: The origin of deep ocean microseisms in the north atlantic ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **464 (2091)**, 777–793.

Kimman, W., and J. Trampert, 2010: Approximations in seismic interferometry and their effects on surface waves. *Geophysical Journal International*, **182 (1)**, 461–476.

King, A., and X. Luo, 2009: Methodology for tomographic imaging ahead of mining using the shearer as a seismic source. *GEOPHYSICS*, **74 (2)**, M1–M8, doi:10.1190/1.3074334.

King, S., and A. Curtis, 2012: Suppressing nonphysical reflections in Green's function estimates using source-receiver interferometry. *Geophysics*, **77 (1)**, Q15–Q25, doi:10.1190/geo2011-0300.1.

Klasner, J., W. Cannon, and W. Van Schmus, 1982: The pre-keweenawan tectonic history of southern canadian shield and its influence on formation of the midcontinent rift. *Geological Society of America Memoirs*, **156**, 27–46.

Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, 2019: Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, **90 (1)**, 3–14.

Krylov, V., and C. Ferguson, 1994: Calculation of low-frequency ground vibrations from railway trains. *Applied Acoustics*, **42 (3)**, 199–213, doi:10.1016/0003-682X(94)90109-0.

Lavoué, F., O. Coutant, P. Boué, L. Pinzon-Rincon, F. Brenguier, P. Dales, M. Rezaeifar, and C. J. Bean, 2020: Understanding seismic waves generated by train traffic via modelling: implications for seismic imaging and monitoring. *submitted to Seismological Research Letters*.

Li, C., Z. Li, Z. Peng, C. Zhang, N. Nakata, and T. Sickbert, 2018a: Long-Period Long-Duration Events Detected by the IRIS Community Wavefield Demonstration Experiment in Oklahoma: Tremor or Train Signals? *Seismological Research Letters*, **89 (5)**, 1652–1659.

Li, X., Z. Li, E. Wang, Y. Liang, B. Li, P. Chen, and Y. Liu, 2018b: Pattern recognition of mine microseismic and blasting events based on wave fractal features. *Fractals*, **26 (03)**, 1850 029.

Lin, F.-C., D. Li, R. W. Clayton, and D. Hollis, 2013a: High-resolution 3d shallow crustal structure in long beach, california: Application of ambient noise tomography on a dense seismic array. *Geophysics*, **78 (4)**, Q45–Q56.

Lin, F.-C., M. P. Moschetti, and M. H. Ritzwoller, 2008: Surface wave tomography of the western united states from ambient seismic noise: Rayleigh and love wave phase velocity maps. *Geophysical Journal International*, **173 (1)**, 281–298.

Lin, F.-C., and M. H. Ritzwoller, 2011: Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure. *Geophysical Journal International*, **186 (3)**, 1104–1120.

Lin, F.-C., M. H. Ritzwoller, and R. Snieder, 2009: Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*, **177 (3)**, 1091–1110.

Lin, F.-C., V. C. Tsai, B. Schmandt, Z. Duputel, and Z. Zhan, 2013b: Extracting seismic core phases with array interferometry. *Geophysical Research Letters*, **40 (6)**, 1049–1053, doi: 10.1002/grl.50237.

Longuet-Higgins, M. S., 1950: A theory of the origin of microseisms. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **243 (857)**, 1–35.

Marathon, T., 2019: Marathon town website. esolutions Group, URL https://www.marathon.ca/.

McNamara, D. E., and R. I. Boaz, 2019: Visualization of the seismic ambient noise spectrum. *Seismic Ambient Noise*, N. Nakata, L. Gualtieri, and A. Fichtner, Eds., Cambridge Univ. Press, 1–29.

Meier, U., N. M. Shapiro, and F. Brenguier, 2010: Detecting seasonal variations in seismic velocities within los angeles basin from correlations of ambient seismic noise. *Geophysical Journal International*, **181 (2)**, 985–996.

Meng, H., Y. Ben-Zion, and C. W. Johnson, 2019: Detection of random noise and anatomy of continuous seismic waveforms in dense array data near Anza California. *Geophysical Journal International*, **219 (3)**, 1463–1473, doi:10.1093/gji/ggz349.

Mikesell, D., K. van Wijk, A. Calvert, and M. Haney, 2009: The virtual refraction: Useful spurious energy in seismic interferometry. *Geophysics*, **74 (3)**, A13–A17, doi:10.1190/1.3095659.

Mordret, A., A. Jolly, Z. Duputel, and N. Fournier, 2010: Monitoring of phreatic eruptions using interferometry on retrieved cross-correlation function from ambient seismic noise: Results from mt. ruapehu, new zealand. *Journal of Volcanology and Geothermal Research*, **191 (1-2)**, 46–59.

Mordret, A., M. Landès, N. Shapiro, S. Singh, and P. Roux, 2014: Ambient noise surface wave tomography to determine the shallow shear velocity structure at valhall: depth inversion with a neighbourhood algorithm. *Geophysical Journal International*, **198 (3)**, 1514–1525.

Mordret, A., D. Rivet, M. Landès, and N. M. Shapiro, 2015: Three-dimensional shear velocity anisotropic model of piton de la fournaise volcano (la réunion island) from ambient seismic noise. *Journal of Geophysical Research: Solid Earth*, **120 (1)**, 406–427, doi:https://doi.org/10.1002/2014JB011654.

Mordret, A., N. M. Shapiro, S. Singh, P. Roux, J.-P. Montagner, and O. I. Barkved, 2013a: Azimuthal anisotropy at valhall: The helmholtz equation approach. *Geophysical Research Letters*, **40 (11)**, 2636–2641.

Mordret, A., N. M. Shapiro, S. S. Singh, P. Roux, and O. I. Barkved, 2013b: Helmholtz tomography of ambient noise surface wave data to estimate scholte wave phase velocity at valhall life of the field. *Geophysics*, **78 (2)**, WA99–WA109.

Muyzert, E., 2007: Seabed property estimation from ambient-noise recordings: Part 2—scholte-wave spectral-ratio inversion. *Geophysics*, **72 (4)**, U47–U53.

Myers, S. C., W. R. Walter, K. Mayeda, and L. Glenn, 1999: Observations in support of rg scattering as a source for explosion s waves: Regional and local recordings of the 1997 kazakhstan depth of burial experiment. *Bulletin of the Seismological Society of America*, **89 (2)**, 544–549.

Nakata, N., 2016: Near-surface S-wave velocities estimated from traffic-induced Love waves using seismic interferometry with double beamforming. *Interpretation*, **4 (4)**, SQ23–SQ31, doi:10.1190/INT-2016-0013.1.

Nakata, N., J. P. Chang, J. F. Lawrence, and P. Boué, 2015: Body wave extraction and tomography at long beach, california, with ambient-noise interferometry. *Journal of Geophysical Research: Solid Earth*, **120 (2)**, 1159–1173, doi:10.1002/2015JB011870.

Nakata, N., L. Gualtieri, and A. Fichtner, 2019: *Seismic ambient noise.* Cambridge University Press.

Nakata, N., R. Snieder, T. Tsuji, K. Larner, and T. Matsuoka, 2011: Shear wave imaging from traffic noise using seismic interferometry by cross-coherence. *Geophysics*, **76 (6)**, SA97–SA106, doi:10.1190/geo2010-0188.1.

Nishida, K., 2013: Global propagation of body waves revealed by cross-correlation analysis of seismic hum. *Geophysical Research Letters*, **40 (9)**, 1691–1696, doi:10.1002/grl.50269.

Nishida, K., H. Kawakatsu, and K. Obara, 2008: Three-dimensional crustal s wave velocity structure in japan using microseismic data recorded by hi-net tiltmeters. *Journal of Geophysical Research: Solid Earth*, **113 (B10)**.

Obermann, A., T. Kraft, E. Larose, and S. Wiemer, 2015: Potential of ambient seismic noise techniques to monitor the st. gallen geothermal site (switzerland). *Journal of Geophysical Research: Solid Earth*, **120 (6)**, 4301–4316.

Olivier, G., F. Brenguier, M. Campillo, R. Lynch, and P. Roux, 2015: Body-wave reconstruction from ambient seismic noise correlations in an underground mine. *Geophysics*, **80 (3)**, KS11–KS25, doi:10.1190/geo2014-0299.1.

Panea, I., D. Draganov, C. Almagro Vidal, and V. Mocanu, 2014: Retrieval of reflections from ambient noise recorded in the mizil area, romania. *Geophysics*, **79 (3)**, Q31–Q42.

Penumadu, D., and C. B. Park, 2005: Multichannel analysis of surface wave (masw) method for geotechnical site characterization. *Earthquake engineering and soil Dynamics*, 1–10.

Peterson, J., and Coauthors, 1993: Observations and modeling of seismic background noise. Tech. rep., US Geological Survey.

Pham, T.-S., H. Tkalčić, M. Sambridge, and B. L. Kennett, 2018: Earth's correlation wavefield: Late coda correlation. *Geophysical Research Letters*, **45 (7)**, 3035–3042.

Picozzi, M., S. Parolai, D. Bindi, and A. Strollo, 2009: Characterization of shallow geology by high-frequency seismic noise tomography. *Geophysical Journal International*, **176 (1)**, 164–174.

Pinzon-Rincon, L., and Coauthors, 2021: Humming trains in seismology: an opportune source for probing the shallow crust. *Seismological Society of America*, **92 (2A)**, 623–635.

Poli, P., M. Campillo, H. Pedersen, L. W. Group, and Coauthors, 2012: Body-wave imaging of earth's mantle discontinuities from ambient seismic noise. *Science*, **338 (6110)**, 1063–1065.

Polychronopoulou, K., A. Lois, and D. Draganov, 2020: Body-wave passive seismic interferometry revisited: mining exploration using the body waves of local microearthquakes. *Geophysical Prospecting*, **68 (1-Cost-Effective and Innovative Mineral Exploration Solutions)**, 232–253.

Poppeliers, C., and D. Mallinson, 2015: High-frequency seismic noise generated from breaking shallow water ocean waves and the link to time-variable sea states. *Geophysical Research Letters*, **42 (20)**, 8563–8569.

Puritch, E., D. Orava, T. Armstrong, A. Yassa, R. Gowans, I. Wislesky, and C. Jacobs, 2009: Technical report on the updated mineral resource estimate and feasibility study for the marathon pgm-cu project, marathon, ontario, canada. p&e mining consultants inc. *P&E Mining Consultants Inc.*

Qiu, H., F.-C. Lin, and Y. Ben-Zion, 2019: Eikonal tomography of the southern california plate boundary region. *Journal of Geophysical Research: Solid Earth*, **124 (9)**, 9755–9779.

Quiros, D. A., L. D. Brown, and D. Kim, 2016: Seismic interferometry of railroad induced ground motions: body and surface wave imaging. *Geophysical Journal International*, **205 (1)**, 301–313, doi:10.1093/gji/ggw033.

Renalier, F., D. Jongmans, M. Campillo, and P.-Y. Bard, 2010: Shear wave velocity imaging of the avignonet landslide (france) using ambient noise cross correlation. *Journal of Geophysical Research: Earth Surface*, **115 (F3)**.

Retailleau, L., P. Boué, L. Li, and M. Campillo, 2020: Ambient seismic noise imaging of the lowermost mantle beneath the north atlantic ocean. *Geophysical Journal International*, **222 (2)**, 1339–1351.

Retailleau, L., and L. Gualtieri, 2019: Toward high-resolution period-dependent seismic monitoring of tropical cyclones. *Geophysical Research Letters*, **46 (3)**, 1329–1337, doi: 10.1029/2018GL080785.

Rezaeifar, M., G. Maggio, Y. Xu, C. Bean, F. Lavoué, P. Boué, L. Pinzon-Rincon, and F. Brenguier, 2020: Imaging shallow structures in dublin city using seismic interferometry of seismic waves generated by train traffic. *EGU General Assembly Conference Abstracts*, 16678.

Rhie, J., and B. Romanowicz, 2004: Excitation of earth's continuous free oscillations by atmosphere–ocean–seafloor coupling. *Nature*, **431 (7008)**, 552–556.

Riahi, N., and P. Gerstoft, 2015: The seismic traffic footprint: Tracking trains, aircraft, and cars seismically. *Geophysical Research Letters*, **42 (8)**, 2674–2681, doi:10.1002/2015GL063558.

Ritzwoller, M. H., F.-C. Lin, and W. Shen, 2011a: Ambient noise tomography with a large seismic array. *Comptes Rendus Geoscience*, **343 (8-9)**, 558–570.

Ritzwoller, M. H., F.-C. Lin, and W. Shen, 2011b: Ambient noise tomography with a large seismic array. *Comptes Rendus Geoscience*, **343 (8-9)**, 558–570.

Rivet, D., M. Campillo, N. M. Shapiro, V. Cruz-Atienza, M. Radiguet, N. Cotte, and V. Kostoglodov, 2011: Seismic evidence of nonlinear crustal deformation during a large slow slip event in mexico. *Geophysical Research Letters*, **38 (8)**.

Roots, E., A. J. Calvert, and J. Craven, 2017: Interferometric seismic imaging around the active Lalor mine in the Flin Flon greenstone belt, Canada. *Tectonophysics*, **718**, 92 – 104, doi:https://doi.org/10.1016/j.tecto.2017.04.024.

Rost, S., and C. Thomas, 2002: Array seismology: Methods and applications. *Reviews of geophysics*, **40 (3)**, 2–1.

Rost, S., and C. Thomas, 2009: Improving seismic resolution through array processing techniques. *Surveys in Geophysics*, **30 (4)**, 271–299.

Roux, P., W. Kuperman, and Coauthors, 2004: Extraction of coherent wavefronts from ocean ambient noise. *J. Acoust. Soc. Am*, **116 (4)**, 1995–2003.

Roux, P., K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, 2005a: P-waves from cross-correlation of seismic noise. *Geophysical Research Letters*, **32 (19)**, doi:10.1029/2005GL023803.

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

Roux, P., and Coauthors, 2018: Toward seismic metamaterials: the metaforet project. *Seismological Research Letters*, **89 (2A)**, 582–593.

Ruigrok, E., X. Campman, and K. Wapenaar, 2011: Extraction of p-wave reflections from microseisms. *Comptes Rendus Geoscience*, **343 (8)**, 512 – 525, doi:https://doi.org/10.1016/j.crte.2011.02.006.

Sabra, K. G., P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, 2005: Extracting time-domain green's function estimates from ambient seismic noise. *Geophysical Research Letters*, **32 (3)**, doi:10.1029/2004GL021862.

Sager, K., C. Boehm, L. Ermert, L. Krischer, and A. Fichtner, 2018: Sensitivity of seismic noise correlation functions to global noise sources. *Journal of Geophysical Research: Solid Earth*, **123 (8)**, 6911–6921, doi:https://doi.org/10.1029/2018JB016042.

Sager, K., V. Tsai, Y. Sheng, F. Brenguier, P. Boué, A. Mordret, and H. Igel, 2021: Modeling p waves in seismic noise correlations: Advancing fault monitoring using train traffic sources. *Geophysical Journal International.*

Sambridge, M., 1999: Geophysical inversion with a neighbourhood algorithm—i. searching a parameter space. *Geophysical journal international*, **138 (2)**, 479–494.

Sánchez-Sesma, F. J., and M. Campillo, 2006: Retrieval of the green's function from cross correlation: the canonical elastic problem. *Bulletin of the Seismological Society of America*, **96 (3)**, 1182–1191.

Schuster, G., J. Yu, J. Sheng, and J. Rickett, 2004: Interferometric/daylight seismic imaging. *Geophysical Journal International*, **157 (2)**, 838–852.

Seats, K. J., J. F. Lawrence, and G. A. Prieto, 2012: Improved ambient noise correlation functions using welch's method. *Geophysical Journal International*, **188 (2)**, 513–523, doi: 10.1111/j.1365-246X.2011.05263.x.

Seydoux, L., N. Shapiro, J. de Rosny, and M. Landes, 2015: A spatial coherence analysis of seismic wavefields based on array covariance matrix: Application to one year of the usarray data. *AGU Fall Meeting Abstracts*, Vol. 2015, S34B–04.

Seydoux, L., N. M. Shapiro, J. de Rosny, F. Brenguier, and M. Landès, 2016: Detecting seismic activity with a covariance matrix analysis of data recorded on seismic arrays. *Geophysical Journal International*, **204 (3)**, 1430–1442.

Shapiro, N. M., and M. Campillo, 2004: Emergence of broadband rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, **31 (7)**, doi:10.1029/2004GL019491.

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

Smith, M. L., and F. Dahlen, 1973: The azimuthal dependence of love and rayleigh wave propagation in a slightly anisotropic medium. *Journal of Geophysical Research*, **78 (17)**, 3321–3333.

Snieder, R., 2004: Extracting the green's function from the correlation of coda waves: A derivation based on stationary phase. *Phys. Rev. E*, **69**, 046 610, doi:10.1103/PhysRevE.69.046610.

Snieder, R., and E. Safak, 2006: Extracting the building response using seismic interferometry: Theory and application to the millikan library in pasadena, california. *Bulletin of the Seismological Society of America*, **96 (2)**, 586–598.

Snieder, R., K. Wapenaar, and K. Larner, 2006: Spurious multiples in seismic interferometry of primaries. *Geophysics*, **71 (4)**, SI111–SI124, doi:10.1190/1.2211507.

Spica, Z., M. Perton, N. Nakata, X. Liu, and G. C. Beroza, 2018: Shallow vs imaging of the groningen area from joint inversion of multimode surface waves and h/v spectral ratios. *Seismological Research Letters*, **89 (5)**, 1720–1729.

Stehly, L., M. Campillo, and N. Shapiro, 2006: A study of the seismic noise from its long-range correlation properties. *Journal of Geophysical Research: Solid Earth*, **111 (B10)**.

STILLWATER, C. I., and Sisprobe, 2019: Deliverable d3.1: deployment complete. Tech. rep., PACIFIC consortium.

Tarantola, A., 2005: *Inverse problem theory and methods for model parameter estimation*. SIAM.

Tarantola, A., and B. Valette, 1982: Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics*, **20 (2)**, 219–232.

Teodor, D., C. Beard, L. A. Pinzon-Rincon, A. Mordret, F. Lavoué, S. Beaupretre, P. Boué, and F. Brenguier, 2021: High-frequency ambient noise surface wave tomography at the marathon pge-cu deposit (ontario, canada). Tech. rep., Copernicus Meetings.

Traffic-Office, 2016: Provincial highways annual average daily traffic (aadt). Tech. rep., Ministry of Transportation.

Trinh, P.-T., R. Brossier, L. Métivier, L. Tavard, and J. Virieux, 2019: Efficient time-domain 3d elastic and viscoelastic full-waveform inversion using a spectral-element method on flexible cartesian-based mesh. *Geophysics*, **84 (1)**, R61–R83.

Tsai, V. C., 2009: On establishing the accuracy of noise tomography travel-time measurements in a realistic medium. *Geophysical Journal International*, **178 (3)**, 1555–1564.

Vasconcelos, I., and R. Snieder, 2008a: Interferometry by deconvolution: Part 1—theory for acoustic waves and numerical examples. *Geophysics*, **73 (3)**, S115–S128.

Vasconcelos, I., and R. Snieder, 2008b: Interferometry by deconvolution: Part 2—theory for elastic waves and application to drill-bit seismic imaging. *Geophysics*, **73 (3)**, S129–S141.

Verbeke, J., L. Boschi, L. Stehly, E. Kissling, and A. Michelini, 2012: High-resolution rayleigh-wave velocity maps of central europe from a dense ambient-noise data set. *Geophysical Journal International*, **188 (3)**, 1173–1187.

Vidal, C. A., D. Draganov, J. Van der Neut, G. Drijkoningen, and K. Wapenaar, 2014: Retrieval of reflections from ambient noise using illumination diagnosis. *Geophysical Journal International*, **198 (3)**, 1572–1584.

Wapenaar, K., 2004: Retrieving the elastodynamic green's function of an arbitrary inhomogeneous medium by cross correlation. *Physical review letters*, **93 (25)**, 254 301.

Wapenaar, K., D. Draganov, R. Snieder, X. Campman, and A. Verdel, 2010a: Tutorial on seismic interferometry: Part 1—basic principles and applications. *Geophysics*, **75 (5)**, 75A195–75A209.

Wapenaar, K., and J. Fokkema, 2006: Green's function representations for seismic interferometry. *Geophysics*, **71 (4)**, SI33–SI46, doi:10.1190/1.2213955.

Wapenaar, K., E. Ruigrok, J. van der Neut, and D. Draganov, 2011: Improved surface-wave retrieval from ambient seismic noise by multi-dimensional deconvolution. *Geophysical Research Letters*, **38 (1)**.

Wapenaar, K., E. Slob, R. Snieder, and A. Curtis, 2010b: Tutorial on seismic interferometry: Part 2—underlying theory and new advances. *Geophysics*, **75 (5)**, 75A211–75A227.

Wapenaar, K., and J. van der Neut, 2010: A representation for green's function retrieval by multidimensional deconvolution. *The Journal of the Acoustical Society of America*, **128 (6)**, EL366–EL371.

Wapenaar, K., J. van der Neut, and E. Ruigrok, 2008: Passive seismic interferometry by multidimensional deconvolution. *Geophysics*, **73 (6)**, A51–A56.

Weaver, R. L., and O. I. Lobkis, 2001: Ultrasonics without a source: Thermal fluctuation correlations at mhz frequencies. *Physical Review Letters*, **87 (13)**, 134 301.

Webb, S. C., 2007: The earth's 'hum'is driven by ocean waves over the continental shelves. *Nature*, **445 (7129)**, 754–756.

Wegler, U., and C. Sens-Schönfelder, 2007: Fault zone monitoring with passive image interferometry. *Geophysical Journal International*, **168 (3)**, 1029–1033.

Weiblen, P. W., 1982: Keweenawan intrusive igneous rocks. *Geological Society of America Memoir*, **156**, 57–82.

Wessel, P., and D. Bercovici, 1998: Interpolation with splines in tension: a green's function approach. *Mathematical Geology*, **30 (1)**, 77–93.

Wielandt, E., 1993: Propagation and structural interpretation of non-plane waves. *Geophysical Journal International*, **113 (1)**, 45–53.

Withers, M. M., R. C. Aster, C. J. Young, and E. P. Chael, 1996: High-frequency analysis of seismic background noise as a function of wind speed and shallow depth. *Bulletin of the Seismological Society of America*, **86 (5)**, 1507–1515.

Xu, Y., K. D. Koper, and R. Burlacu, 2017: Lakes as a source of short-period (0.5–2 s) microseisms. *Journal of Geophysical Research: Solid Earth*, **122 (10)**, 8241–8256.

Zhan, Z., 2020: Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas. *Seismological Research Letters*, **91 (1)**, 1–15.

Zhang, J., P. Gerstoft, and P. M. Shearer, 2009: High-frequency p-wave seismic noise driven by ocean winds. *Geophysical Research Letters*, **36 (9)**.

Zhang, Y., T. Planes, E. Larose, A. Obermann, C. Rospars, and G. Moreau, 2016: Diffuse ultrasound monitoring of stress and damage development on a 15-ton concrete beam. *The Journal of the Acoustical Society of America*, **139 (4)**, 1691–1701.