Humming trains in seismology: an opportunistic source for probing the shallow crust

Laura Pinzon-Rincon[1], François Lavoué[1], Aurélien Mordret[1], Pierre Boué[1], Florent Brenguier[1], Philippe Dales[1,6], Yehuda Ben-Zion[2], Frank Vernon[3], Christopher Bean[4], and Daniel Hollis[5]

[1]Univ. Grenoble Alpes, CNRS, ISTerre, Grenoble, France

[2]Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA

[3]IGPP, University of California, San Diego, La Jolla, CA, USA

[4]Dublin Institute for Advanced Studies, Ireland

[5]Sisprobe, Grenoble, France

[6]Institute of Mine Seismology, USA

July 10, 2020

Abstract

Unveiling the coupling between the atmosphere and the Earth, improving our understanding of the preparation phase of earthquakes and volcanic eruptions, mitigating induced seismic hazard, discovering new natural resources all require improved imaging and monitoring of the top first kilometers of the crust. Passive seismic imaging and monitoring usually relies on blind correlations of long time series of noise. Instead, seismic interferometry applied to opportunistic sources of noise relies on an accurate understanding of noise source mechanisms, on time window and station pair selection, and on specific seismic phases extraction (surface, body). Recently, massive freight trains have been recognized as the most persistent and powerful cultural seismic sources generating tremor equivalent to magnitude 2 earthquakes and detectable up to 100 km distance. In this paper, we discuss the source mechanisms of train tremor and review some basic theory on seismic interferometry applied to opportunistic sources. We finally show two case studies of long-range body- and surface-waves retrieval in the contexts of mineral exploration in Canada and fault zone monitoring in Southern California. This approach of noise recovery to create valuable sources together with disruptive dense data acquisition technologies such as nodes or Distributed Acoustic Sensing will deeply transform our capability to explore and monitor the shallow crust with improved spatial and temporal resolution.

1 Introduction

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., 2018; Green et al., 2017; Fuchs et al., 2018; Inbal et al., 2018) and source modelling (Lavoué et al., 2020). In pioneer studies Nakata et al. (2011); Quiros et al. (2016); Chang et al. (2016) have proposed the idea of using traffic noise and seismic interferometry for both body- and surface-wave imaging. These studies are however limited to local sources of cultural noise and near-surface applications.

In an attempt to reveal the signature of non-volcanic tremors (NVT) along the San Andreas Fault in Southern California, Inbal et al. (2018) discovered tremors that shared a puzzling similarity with NVTs but that were found to be generated by massive freight trains running along the Coachella Valley and detected up to 100 km from the railways. Indeed Brenguier et al. (2019) estimated that the radiated seismic energy from a single 1-km-long freight train travelling through a 10-km-long railway section is equivalent to a magnitude 2 earthquake. By further applying the concepts of seismic interferometry to the correlation of this long-range train-generated noise, Brenguier et al. (2019); Dales et al. (2020) demonstrated the possibility of extracting useful information on the Earth's shallow crust structure and temporal evolution down to a few kilometers depth, thus providing a potential alternative to costly active-source monitoring (Tsuji et al., 2018). This paper reviews basic concepts and shows examples of the application of seismic interferometry to train noise with a special focus on long-range body-waves reconstruction for crustal exploration and monitoring (Fig. 1).

Green's function retrieval through the correlation of a diffused coda or seismic noise recorded at different sensors, also referred to as seismic interferometry, has revolutionized seismology in the last decades (Campillo and Paul, 2003; Shapiro et al., 2005) and led to the publication of more than 2000 papers in the last 15 years. It has been mainly applied to crustal imaging using correlations of the pervasive surface wave noise generated in the oceans in the period range from 1 to 20 seconds. Recent studies have also unveiled the possibility of reconstructing body-waves at global (Poli et al., 2012; Boué et al., 2013) and local scales (Draganov et al., 2009; Nakata et al., 2015; Olivier et al., 2015; Nakata et al., 2016).

A perfect 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). Moving trains are opportunistic sources of noise located at specific locations (railways) at the surface of the Earth and should thus be considered with care for seismic interferometry. Traffic train noise cannot be blindly correlated without considering the effects of non-even source distribution on body-wave reconstruction.

In this paper we first illustrate some typical train noise signal, discuss some recent models of the source mechanisms of train seismic radiations and introduce a map of the predicted spatial extent of useful train noise throughout the contiguous US. Secondly we propose a methodological framework focusing our approach on the concept of stationary phase kernels (Snieder, 2004) and propose a signal processing strategy for applying seismic interferometry to train noise with a focus on long-range body-waves reconstruction. We finally review two recent case studies in the contexts of mineral exploration in Canada and tectonic fault monitoring in Southern California.

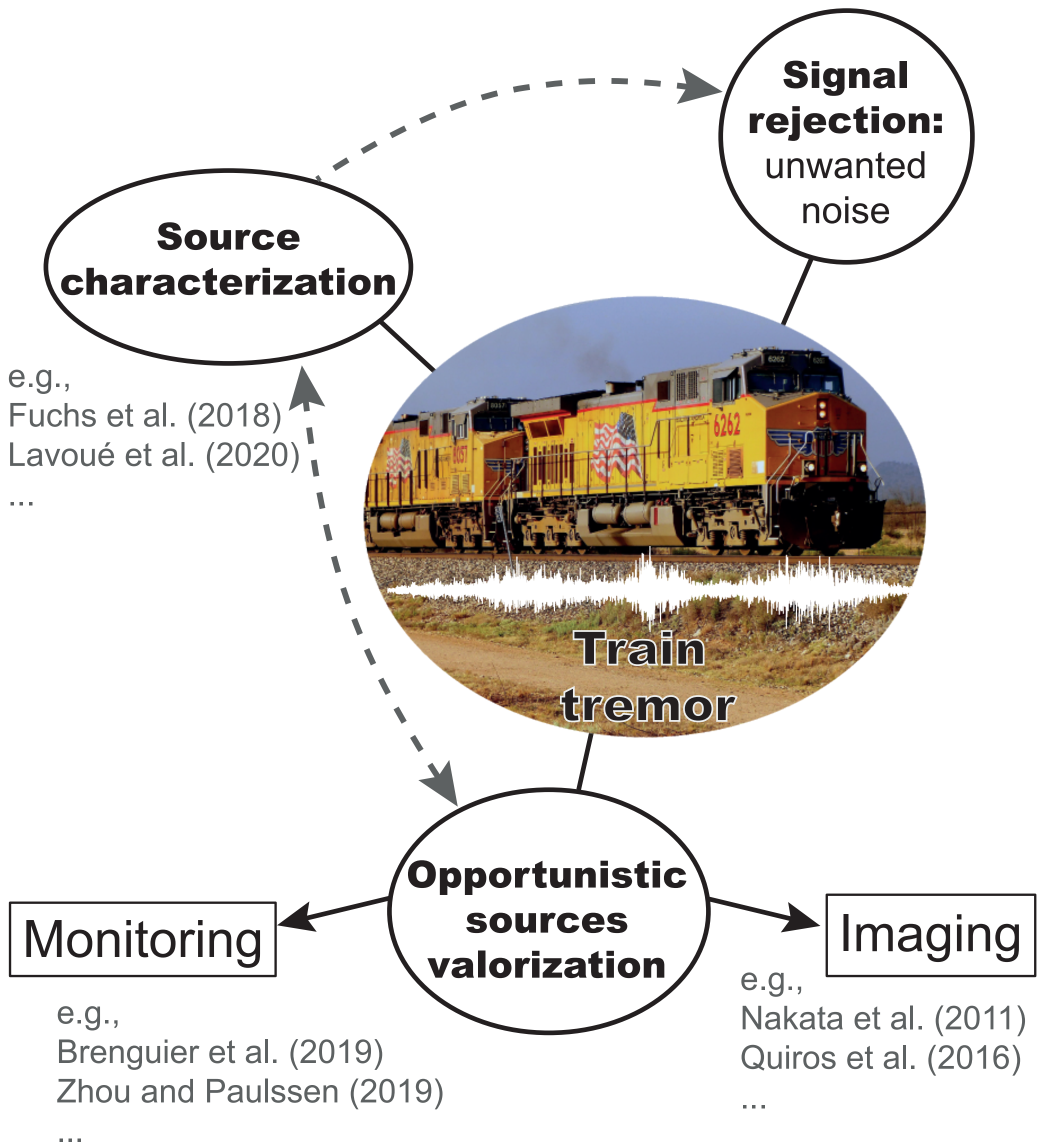

Figure 1: Cartoon showing examples of studies related to train seismic tremors.

2 The sound of trains in the Earth

Massive freight trains generate a seismic wave train that shows a striking similarity with episodic tectonic tremors (Fig. 2 top). As Inbal et al. (2018) report, the confusion can be even more puzzling because train traffic may not show typical cultural diurnal or weekly modulation and typical train speed (25 m/s or 90km/h) is also in the range of reported tectonic tremor migration velocity at depth. Train hum has however a specific spectral signature with clear spectral lines above 1 Hz (Fuchs et al., 2018) illustrated in Fig. 2 for a train signal recorded in Canada about 3 km from the railway (first case study presented below).

The engineering community has studied train-induced ground vibrations thoroughly in order to mitigate their intensity. Several source mechanisms have been proposed (e.g. Connolly et al., 2015) including quasi-static excitation due to axle load, 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 main 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., 2018; Brenguier et al., 2019). It is therefore possible to model train-generated seismic signals by considering only the vertical forces due to loading applied by axles on the railroad ties (referred here to as sleepers) along the railway (Krylov and Ferguson, 1994; Lavoué et al., 2020).

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 thus depend on train geometry (i.e. wagon length and wheel spacing within each wagon), spacing between sleepers, and train velocity. We propose an open-source code to assess the frequency content of a specific train at (https://gricad-gitlab.univ-grenoble-alpes.fr/pacific/publications/2020_Lavoue-et-al_SRL_supplemental-material). It is worth mentioning that for typical massive American freight trains (2 km long, 15 kilotons) we predict that the dominant spectral lines are in the range 1 to 20 Hz which is ideal for crustal body-wave imaging and monitoring (wavelengths not too large and scattering not too strong) (Brenguier et al., 2019).

Our ability to predict the long-range, body-wave Peak Ground Velocity (PGV) of a moving train tremor is crucial to use it for imaging and monitoring with seismic interferometry. Lavoué et al. (2020) propose that train tremor PGV is directly proportional to the wagon weight for a given train length and a square root function of train length for constant wagon weight. It is also discussed that higher train speeds generate higher PGVs. Moreover, the ground stiffness beneath the railways control the high-frequency content and amplitude of the excitation (trains travelling over a stiff soil generate higher-frequency and higher-amplitude signals). This ground stiffness parameter may also reflect the coupling between the rail track and the ground. While this maximum detection distance may be short (a few kilometers) in sedimentary basins due to attenuation and weak excitation, it can reach almost 100 km on a hard-rock substratum. In southern California, for instance, Inbal et al. (2018) observed a freight train tremor signal as far as 90 km away from the railway. At 45 km from the railway they estimate a PGV of about 10^{-7} m/s. By applying a simple correction of intrinsic attenuation and geometrical spreading for body-waves (P), we estimate that the level of PGV for this specific train would be on the order of 5.10^{-6} m/s at 10 km and 10^{-4} m/s at 1 km from the railway located in the Coachella valley. These values are quite low and train signals a these distances might only be recorded using quiet and high sensitivity seismometers. (Brenguier et al., 2019) confirm that these trains from the Coachella Valley can be used for seismic interferometry from a array of geophones (nodes) located as far as 60 km from the railway.

We use this value of 50 km as a typical maximum distance range for detecting train tremors and investigate the spatial spread of detectable train tremor areas throughout the entire contiguous US and southern Canada (Fig. 2)[1]. This map represents the main freight railway routes. The swathes in colors represent high tonnage routes where the thickness

[1] This figure is based on a map published by the US Department of Transportation (https://railroads.dot.gov/sites/fra.dot.gov/files/inline-images/0209.png), built from the (confidential) waybill samples 2010 established by the US Surface Transportation Board.

(100 km), provides an estimate of regions of potential long-range train tremor detections. This map doesn't take into account the reduced detection capacity in urban areas due to intense local noise and in sedimentary basins due to strong attenuation compared to the Southern California reference.

The colors in Fig. 2 represent annual freight tonnage, which is an indication for the number of trains travelling on the rail sections. Assuming average trains with a length of 2 km and a weight of 15 kilotons (according to statistics derived from the public waybill samples, 2018[2]), a tonnage of 100 MT/year corresponds to approx. 18 trains per day. The number of trains per day will condition our capability to stack the reconstructed body waves using seismic interferometry and will also affect the temporal resolution for monitoring applications (see Section 4).

This map highlights the potential of using trains as a source of opportunity with possible applications to Cascade Volcanoes, the San Andreas Fault system in Northern and Southern California, induced seismicity (e.g. Oklahoma) and resource exploration and monitoring (mineral, water).

3 Seismic interferometry with opportunistic sources

Seismic interferometry is a general term that defines all methodologies aiming to create seismic responses from the correlation of seismic signals observed at different receiver locations (e.g., Wapenaar et al., 2010a,b). In the prospect of turning sensors into virtual sources, this concept has been developed in seismology and seismic exploration mostly during the last 20 years based on the pioneering work on random fields or vertical planar wave autocorrelation (Aki, 1957; Claerbout, 1968) and the time-reversal principle in acoustic (Fink, 1997).

When it comes to retrieving a Green's function using the correlation or a equivalent

[2]https://prod.stb.gov/wp-content/uploads/PublicUseWaybillSample2018.zip

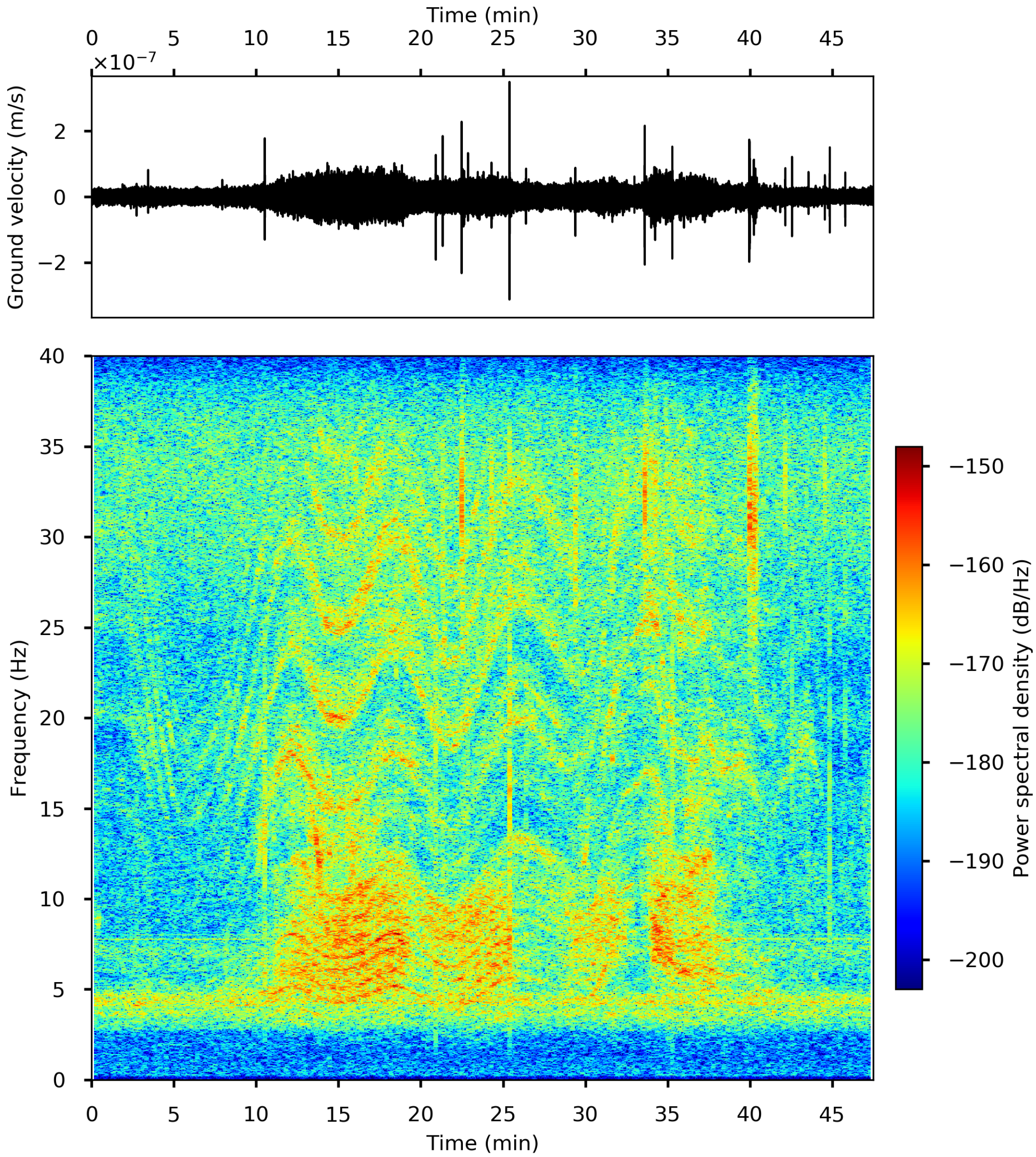

Figure 2: Top: A train tremor recorded 3 km away from a seismic station in Marathon, Canada. Bottom: Its spectrogram showing clear specrtal lines.

operator, the theory mostly relies on either a stationary phase condition (e.g., Snieder, 2004; Roux et al., 2005) and/or an equipartition of modes defining a diffuse fields (e.g., Sánchez-Sesma and Campillo, 2006). The stationary phase condition implies that the correlation

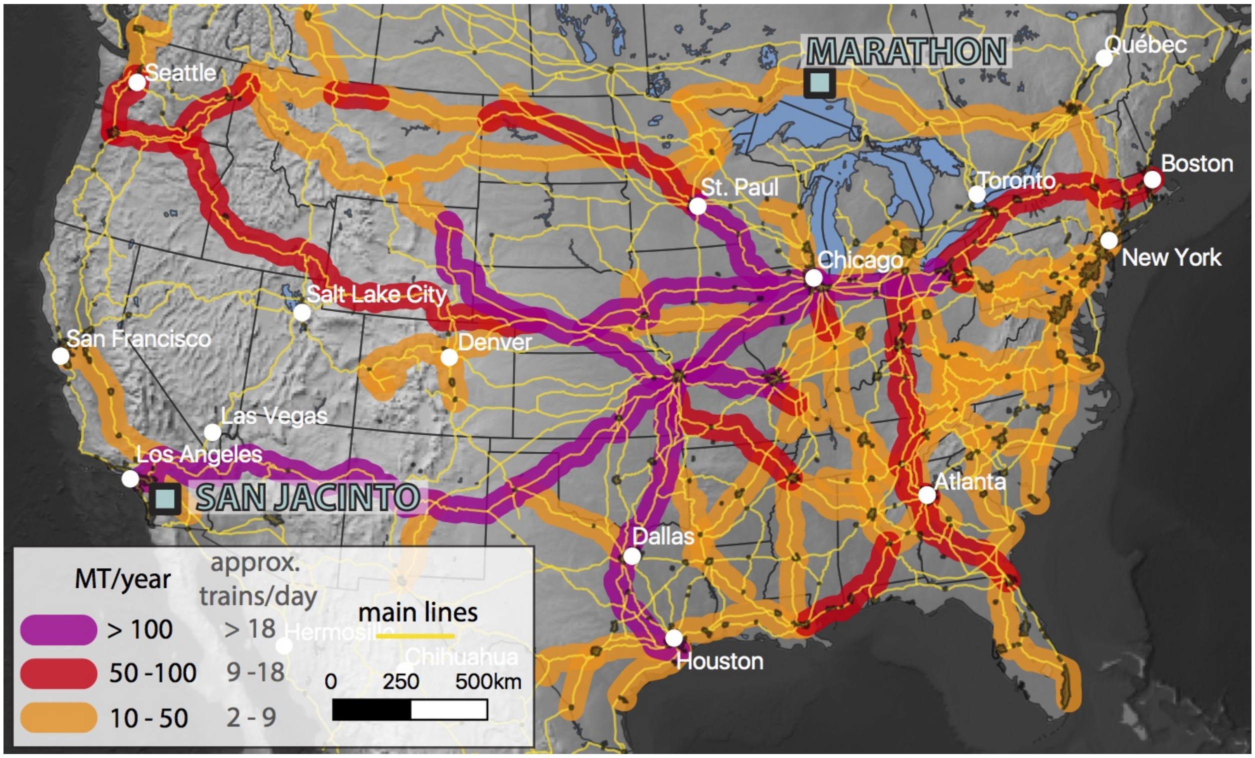

Figure 3: Regions of potential long-range train tremor detection from the main railway route and annual tonnage information in North America. Colors represent annual freight tonnage, which is an indication of the number of trains travelling on the rail sections. Colored lines are 100-km-thick, which is an indication for the distance from which we may detect train-generated signals ($\simeq$ 50 km from the railway, see details in the text).

function convergences towards the Green function requires the presence of sources (or scatterers) inline with the two considered receivers. In a 2D homogeneous medium, these stationary points defined an hyperbolic area, outward from the receiver pair, and which aperture is frequency dependent (the lower frequency the broader source region). Also known as Fresnel zones, these "kernels" correspond to the sensitivity of the correlation to the source location. In 3D and for both surface and body wave retrieval the requirement of equipartition remains and it has been shown that the full Green function retrieval requires sources distributed along arbitrarily shaped surface enclosing the two sensors (e.g., Wapenaar, 2004; Wapenaar and Fokkema, 2006). But even with a clearly dominant distribution of sources at the free surface, several studies successfully investigated the feasibility of retrieving body-waves (e.g., Draganov et al., 2009, 2013), and even specifically using traffic noise (Nakata et al., 2011).

Each of the possible phases (or wave front) included in the Green's function has its own source sensitivity. The main contribution to a particular phase is dominated by sources within its stationary phase area. We can therefore measure a specific phase between two receivers by correlation as soon as a source is located within its stationary phase zone, including at the surface. In the following case study sections we investigate P waves emanating form moving trains and that emerge from the interference between a direct P recorded at a first station, and a PP recorded at a second station after a rebound below the first one (Figure 3b). This interference is possible as soon as seismic sources (trains) comply with the criteria that the arrival time of the PP wave at the second receiver minus the arrival time of the P wave at the first receiver is smaller than the arrival time of the P wave between the two receivers plus/minus a quarter of the dominant period (which is a definition of the stationary phase zone). Note that using somehow controlled sources to retrieve body wave response through interfereometry is very similar to daylight imaging developed by (Schuster et al., 2004) or to virtual source approach discussed by (Bakulin and Calvert, 2006) for borehole imaging.

Train signals represent a very good opportunity for interfereometric studies because we can easily detect and/or predict the source time and location. As soon as a railway is sufficiently close to the targeted area, a single train could illuminate different azimuth and potentially different depth. Figure 3a shows an example of geometry in Marathon (Ontario, Canada) where a railway circumvents a temporary array deployment for ore deposit exploration (detailed in the following section). By selecting stations pairs that are in-lined with specific train location (illustrated for two position by red and blue stars), one can potentially illuminate the structure with on a broad azimuth range. Figure 3b to d are schematics showing different scenarios of interference between a pair of stations : a perfect ballistic interference between a diving P and PP wave (Figure 3b) leading to a directly measurable diving P wave between the two receivers; a classical scenario of a more-or-less scattered wavefield from which we expect some energy to transit between the two receivers from whatever source; and a more problematic interference between two diving wave, or a head wave recorded at the two stations (Figure 3d). The last scenario are sometime referred 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 correlation feature might be useful for imaging if well distinguished from expected diving waves (Dong et al., 2006).

Our idea is to explore the possibility of using a specific data processing workflow, starting with the selection of specific and short time windows including train passage in order to illuminate specific ray paths. This method, which could be extended to any kind of seismic tremors, should help us to extract body-waves between well selected pairs of stations useful for imaging and monitoring studies.

4 Strategy for data processing

The standard noise-correlation workflow typically removes strong transient events such as earthquakes and correlate the entire time series recorded at different sensors (Bensen et al., 2007). In case of specific opportunistic sources such as train traffic we propose a novel workflow based on source characterization, signal and station pairs selection instead as an alternative to blind correlation. By doing so we aim at improving signal to noise ratio of the reconstructed correlation functions and the temporal resolution of monitoring studies. Figure 5 summarizes the five main stages of data processing in comparison to the classical method where continuous data is blindly correlated:

- Source detection and characterization: the first step consists of identifying the opportunistic

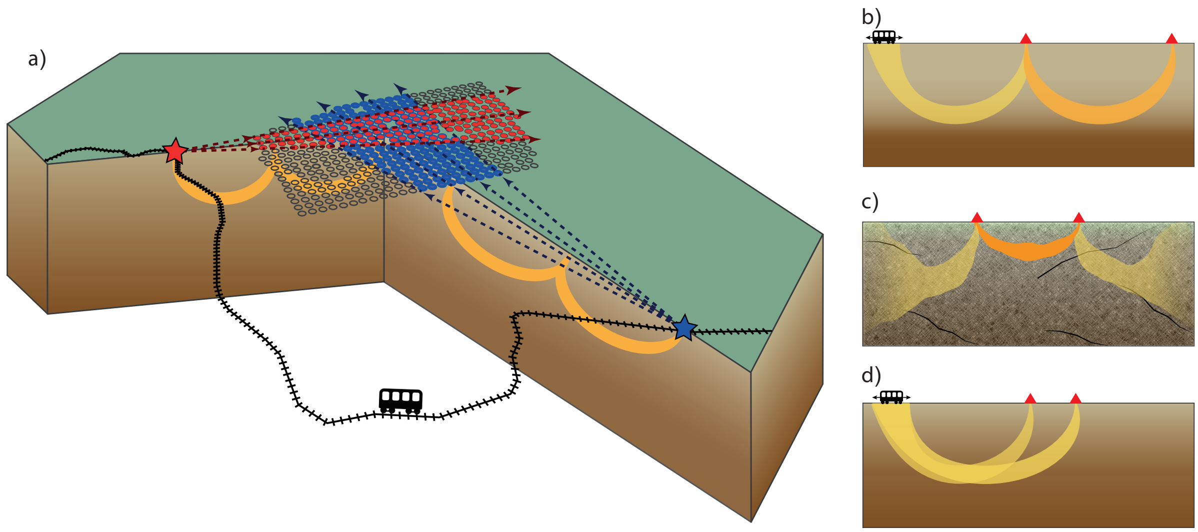

Figure 4: Schematic representation of seismic interferometry for opportunistic sources. (a) A railway surrounding a dense geophone array; an example from Marathon (Canada) deployment. Different train location (stars) allow to illuminate the array with different azimuths. Yellow kernels are schematic views of the propagation a P diving waves. (b-d) 3 different scenarios of wave interference: (b and c) leading to a proper measurement of a P diving wave and (d) leading to a spurious or virtual refraction measurement.

source signature in the continuous data and if possible to locate these sources at least in azimuth. As shown in section 2 the modelling of opportunistic sources helps understanding the temporal and spectral content of the generated wavefield. Standard (STA/LTA) and more advanced techniques such as machine learning (e.g. Seydoux et al., 2016) are used to detect these transient events and array techniques can be used to locate these sources.

- Station pairs selection: Using source location estimates we can apply a spatial selection of station pairs. For a given signal window in time only station pairs located in the stationary phase zone are used 3. During a train passage, the energy emitted by the train travels through an array of sensors from different directions depending on the train position. figure 4a illustrate two train positions at different times (red and blue stars) and the associated selected stations for pair-wise correlations (red and blue dots).

- Compute cross-correlations: after proper time windowing and station pairs selection, we perform cross-correlations.

- Stack(by events, by azimuth): to improve SNR, we can stack the cross-correlations over different events. Especially for cultural sources such as train traffic we can benefit from their repeatability.

- Measurement and analysis: depending on the type of studies different approaches such as travel time measurements can be applied for imaging and monitoring applications.

5 Body- and surface-wave retrieval from correlations of train tremor in the context of mineral exploration

To illustrate one application of train signals to extract body waves in the near sub-surface for imaging purposes, we study a mining exploration block in Marathon, Ontario, Canada (see Fig. 6b). The potential targets are a high concentration of platinum group metals, and minor Cu, hosted in a gabbro intrusion. 1200 seismic stations were deployed in fall 2019 within a backbone array and a dense station line (see Fig. 6b). We recorded 30 days of continuous seismic signals.

Dales et al. (2020) showed that the main source of high-frequency seismic noise in Marathon is freight trains traveling on the railway south-west from the array. They demonstrated that selecting the portions of the noise that correspond to traffic enables to significantly improve the retrieval of body waves compared to correlating the entire noise records. They used only azimuths inline with the dense station line. Here, we generalize the method to all the azimuths to illuminate the medium from a different direction. Following the workflow proposed in section 4, we detect train passages, infer the position and azimuth of the train relative to the array, and carefully select station pairs and time windows for correlation, and finally, we stack by train passage and azimuth.

First, we generate a train catalog. To detect train passages, we use the covariance matrix

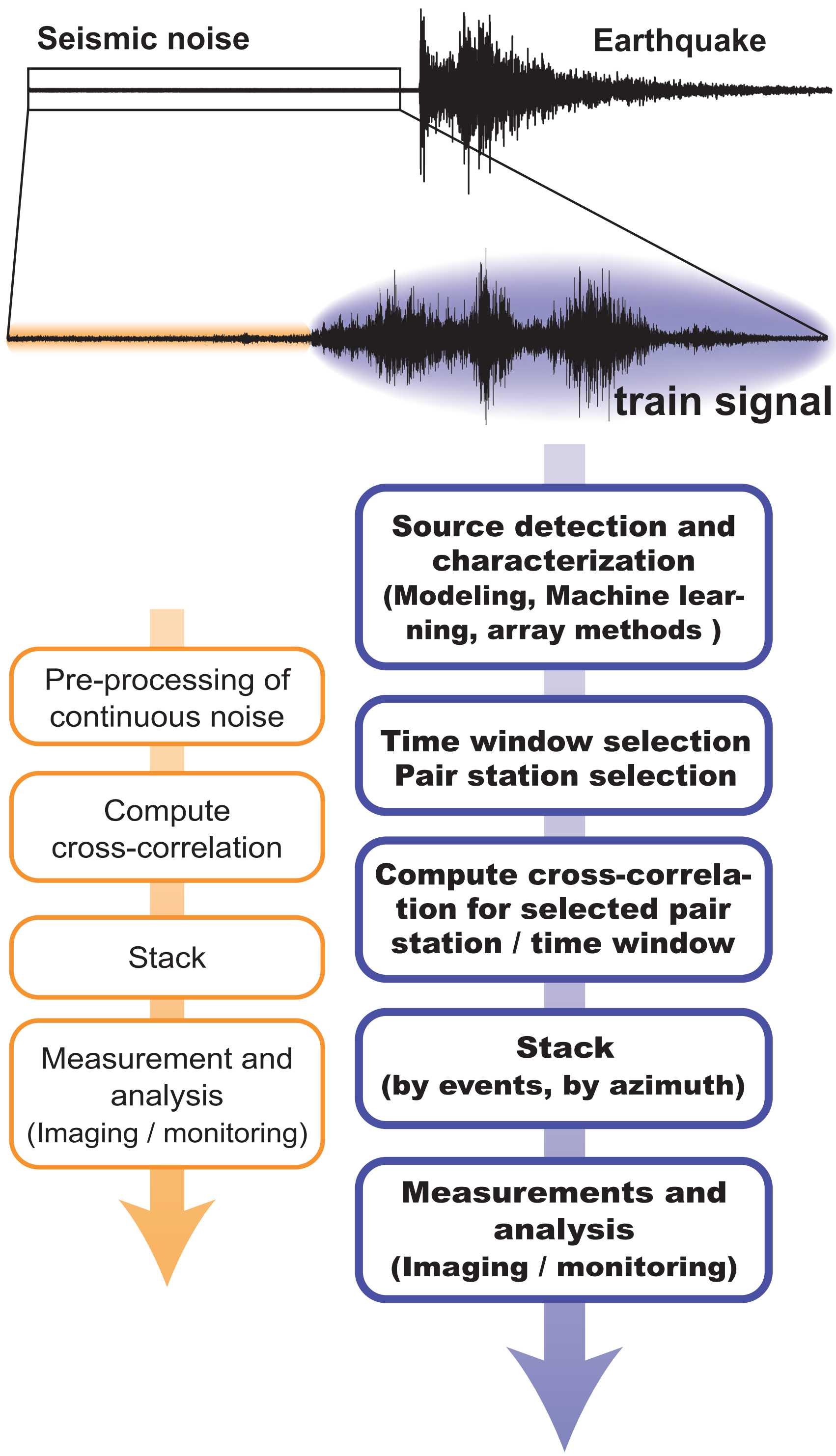

Figure 5: Chart illustrating the processing steps for opportunistic sources (in blue) compared to the standard ambient noise correlation workflow (in orange).

method proposed by Seydoux et al. (2016). The covariance matrix analysis detects emergent signals in the noise, using the spatial coherence of the signals. 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. From these, we retain only single passages (approx. 180 events), i.e., we remove records where the signals generated by several trains overlap. The beamforming technique shows that the array receives energy from each train for a duration of approx. 80 minutes. Second, we extract train signals from the rest of the recording, and we select the station pairs that are inline with the train position. To determine the train position, a 1-minute-long window, beamforming is performed, and we filter between 8 and 16 Hz (Figure 6-d and e, the right side shows 6 beamforming panels for 6 different events at two different times). Each panel corresponds to one-minute time window beamforming and one single train passages. We can see that if we properly select the time window for each event, we have similar azimuths. For each time window, we assume that the main source of energy is the train, and we pick the maximum beam power. We back-projected this signal onto the railway to located the train by minute. Figure 6-b (red and blue cross) shows the train position corresponding to the fist beamforming panel (i.e., one single train). Then, we select station pairs that are inline with the train position for each minute (The red and blue arrow fig. 6b). We apply an azimuthal filter of +/- 5 degrees for each station pairs with respect to the train position. Third, we cross-correlate the selected station pairs by minute without overlapping and for each event (i.e., train passage). We filter between 10 and 18 Hz to avoid surface waves. We stacked the cross-correlations according to their inter-station distances into distance-binned correlation gathers for the selected station pairs (second step).

In the last step, we stack by events for the same train position (i.e, same azimuth). We stack correlation gathers into a reference azimuthal gather. We converge to stable reference stack with 6 train passages. Figure 6-d and e, left side shows the stacked section over 6 train passages, using one-minute data segments. We retrieve two dominant arrivals with an apparent velocity of 3.8 km/s and 7 km/s (yellow and green line, respectively Fig. 6d and e). We suggest that the first arrival is the P-wave and the second one is a S-wave. Figure 6a shows one minute the stack of one-minute cross-correlation for a quiet period (i.e., non-train passage). In comparison with a cross-correlation with regular seismic record, selected stack during train passage allows us to retrieve high-frequency energy even using an array that was initially deployed for a typical passive seismic interferometry.

In the future we plan to use these retrieved sections and both high-frequency surface- and body-waves to map seismic velocity anomalies at different depths.

6 Long-range body-wave retrieval from train tremor correlations for monitoring the San Jacinto Fault Zone

Following the pioneer work of Nakata et al. (2015) and Nakata et al. (2016) on high-frequency body-wave retrieval using dense arrays, Takano et al. (2020), Brenguier et al. (2020) and Zhou and Paulssen (2019) investigated a strategies for monitoring temporal changes of ballistic wave velocities in the aim of improving the depth localisation of stress perturbations at depth. In this section, we illustrate the use of opportunistic seismic sources for passive monitoring applications and revisit the experiment of Brenguier et al. (2019). Here, the goal is to use ballistic P-waves reconstructed from ambient vibrations between two dense arrays to monitor subtle velocity changes at depth within the San Jacinto Fault Zone (SJFZ). Brenguier et al. (2019) showed that using standard ambient noise correlation processing they were able to retrieve high-frequency direct P-waves propagating between the two arrays located at Pinon Flat Observatory (PFO) and in the Cahuilla Reservation (CIR, Fig. 7c). The main sources of these P-waves are the freight trains traveling in the neighboring Coachella Valley, about 30 km to the East-North-East of PFO.

Brenguier et al. (2019) used the full records of ambient noise to obtain stable direct P-wave seismograms. Here we show that, by carefully selecting time-windows where most of the energy is generated by trains, we can improve the quality and spatiotemporal stability of the reconstructed P-waves. As described in Figure 5, the standard three-steps noise correlation computation workflow is replaced by a four-steps one aiming at correlating only the main source of opportunistic energy, i.e., here trains. First, we build a train catalog for the time period of interest (July 22 to August 11 of 2018). To do so, we use three broadband stations (MGE, IDO, and THM of the CI network, Fig. 7c) located along the railway in the Coachella Valley. After band-pass filtering the continuous data between 0.75-5 Hz we slant-stack the envelopes of the continuous seismograms with apparent velocities of ± 95 km/h (dashed blue and orange lines in Fig. 7a), respectively) to detect trains passing through the Fresnel zone (Fig. 7c) and traveling from North to South or South to North, respectively. Once the catalog is completed, we can automatically reject broad time-windows when no train is traveling (large red shaded rectangle in Fig. 7a). In a second step, we cross-correlate the remaining time-windows (large green rectangles), filtered between 3 and 10 Hz, using non-overlapping data segments of 30 min. Then, we stack the cross-correlations according to their inter-station distances into distance-binned correlation gathers. These 30 min correlation gathers are further selected or rejected based on three quality criteria extracted from their respective vespagrams (Davies et al., 1971). The upper panels of the middle row of Figure 7a) show the vespagrams associated with the correlation gathers in the lower panels. The three quality criteria are: 1) $SNR1$, the ratio between the maximum vespagram amplitude in the [0.13-0.2] s/km slowness ([5-7.5] km/s velocity) window (dashed black rectangle in the leftmost vespagram panel, Fig. 7a) and the root-mean-squared (RMS) amplitude of the rest of the vespagram. 2) $SNR2$, the ratio between the maximum vespagram amplitude in the [0.13-0.2] s/km slowness $\times$ [4.5-6] s travel-time window (solid black rectangle in the leftmost vespagram panel, Fig. 7a) and the RMS in the rest of the [0.13-0.2] s/km slowness window.

3) $MaxAmp$, the maximum vespagram amplitude in the [0.13-0.2] s/km slowness $\times$ [4.5-6] s travel-time window. $SNR1$ is used to reject gathers exhibiting phases with apparent velocities different from the expected apparent velocity of a direct P-wave. $SNR2$ is used to reject gathers exhibiting energetic spurious phases with too early or too late arrival time, even though their apparent velocity is correct. $MaxAmp$ is used to reject gathers for which the expected P-wave phase is not enough energetic or for which the energy is too large for a train signal, indicating the detection of an earthquake located in the Fresnel zone (Fig. 7a, top row). For this specific application, we set thresholds such that the conditions $SNR1 \geq 2.5$, $SNR2 \geq 1.5$, and $0.15 \leq MaxAmp \leq 4.0$ must all be fulfilled for a correlation gather to be selected (little green rectangle in Fig. 7a). The actual values for $SNR1$, $SNR2$ and $MaxAmp$ are shown above each correlation gather in Figure 7a). In the last step, we stack the selected correlation gathers into daily gathers and a Reference gather including every selected gather for the whole period of interest (Fig. 7a, bottom row). Ultimately, less than 20% of the full dataset is used for the monitoring measurements (Fig. 7b).

To quantify the improvement of the signals using the opportunistic sources approach, we measure the ratio of SNRs between a reference gather computed with all the data (similar to Brenguier et al. (2019)) and the reference gather from selected train windows shown in Fig. 7a. This operation is performed for each waveform of the gathers and the distribution of the results (Fig. 7d) shows that the opportunistic sources approach improves the SNR of the P-wave signal by more than 25% on average. This has important implications for monitoring. As shown by Silver et al. (2007), the SNR is the main factor controlling the precision of a time delay measurement between two similar waveforms; this precision scales linearly with the SNR. Therefore, carefully selecting train signals before correlation allows us to improve the precision of the monitoring measurements. The 30 min long segments of continuous data used here to discretize the study period could be decreased and adapted even more closely to the train signals, which in turn should allow even larger SNR improvements. This finer processing is beyond the scope of this paper.

The final step of the workflow is to perform the seismic velocity monitoring measurements. Different approaches can be taken. Here we chose to measure the relative time-shift between the seismograms resulting from the slant-stack at 6 km/s of the daily gathers and the reference gather (black and purple traces in Fig. 7b, respectively). We measure the instantaneous time-delay $\delta t(t)$ between the traces in the 3-10 Hz frequency band using the cross-wavelet transform algorithm of Mao et al. (2020). For display purposes, we only show δt values where the amplitude of the P-wave is the largest. Here, the obtained time-shifts are smaller than 0.1% of the propagation time, corresponding to time-shift smaller than 5 ms between the daily and the reference seismograms.

7 Discussion and conclusion

In this paper we discuss the opportunity of massive freight train noise recovery for crustal imaging and monitoring. We show applications to Northern America but the impact of this study is more global. Especially China has the the world's longest high speed railway network (35 000 km) located mostly in the Eastern, most populated part of the country. Passenger trains being lighter than freight trains they generate less energetic tremor and applications might thus be restricted though to near-surface, environmental or engineering studies.

Turning massive freight train noise as a source for imaging and monitoring applications shows potential but is limited to regions neighbouring railways and requires trains travelling at rather high speed. In a more general way, this work provides a workflow for using other more local sources of cultural noise like car and truck traffic, wind farms, but also natural like surf break or even tectonic of volcanic tremor as opportunistic sources of noise.

Even being promising this work raises some challenges that will need to be addressed in the future. Most important is improving our understanding of the spatial sensitivity to structure of the retrieved body- and surface-waves using seismic interferometry and opportunistic sources. Contrarily to active controlled sources, measurements of travel times or temporal travel time perturbations can show sensitivity not only to the structure between the receivers but also between the noise source and the receivers inducing potential misleading interpretations of velocity or velocity change measurements. One drawback from the examples above is also that they use a large number of sensors (hundreds) and thus imply heavy and temporary data acquisition experiments thus limiting the scope of potential applications. A future perspective is to study how permanent, single seismic stations can be used instead of dense arrays. One option would be to temporary deploy dense seismic arrays around permanent seismic stations to help identifying useful phases emanating from noise correlations of opportunistic sources that can in turn be monitored on the long-run using permanent stations only.

Finally one last major perspective is to couple Distributed Acoustic Sensing (Zhan, 2020) data to seismic interferometry for opportunistic sources as described by Dou et al. (2017) for car traffic and near-surface applications. The potential of reconstructing widespread virtual sources along optic fibers from correlations of both short- and long-range opportunistic sources will open the path for endless applications ranging from water resource management in the near surface to earthquake studies at greater depth.

8 Data and Resources

The Marathon dataset will be made publicly available on June 2021. It will either be hosted online or freely sent on external hard disks upon request via the website of passive seismic techniques for environmentally friendly and cost efficient mineral exploration (PACIFIC) (https://www.pacific-h2020.eu). The San Jacinto array data are available on request to Florent Brenguier. The broadband seismic data used in this study originate from the Southern California Earthquake Center, Caltech.Dataset. doi:10.7909/C3WD3xH1.

Maps are made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.

9 Acknowledgments

We acknowledge Nick Arndt, Meysam Rezaeifar, Olivier Coutant, John McBride, Christopher Johnson, Pieter Share and Florian Fuchs, for useful discussions and for supporting data acquisition. We are grateful to the anonymous reviewers for their useful comments.This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 776622 (PACIFIC), from the European Research Council (ERC) under grant agreement No 817803 (FaultScan) and from the Department of Energy (Award DE-SC0016520).

This work was performed in part at the University of California Natural Reserve System Boyd Deep Canyon Desert Research Center Reserve DOI: (DOI: 10.21973/N3V66D). We gratefully acknowledge the Cahuilla Band of Mission Indians Reservation for graciously allowing us to deploy instruments on tribal land.

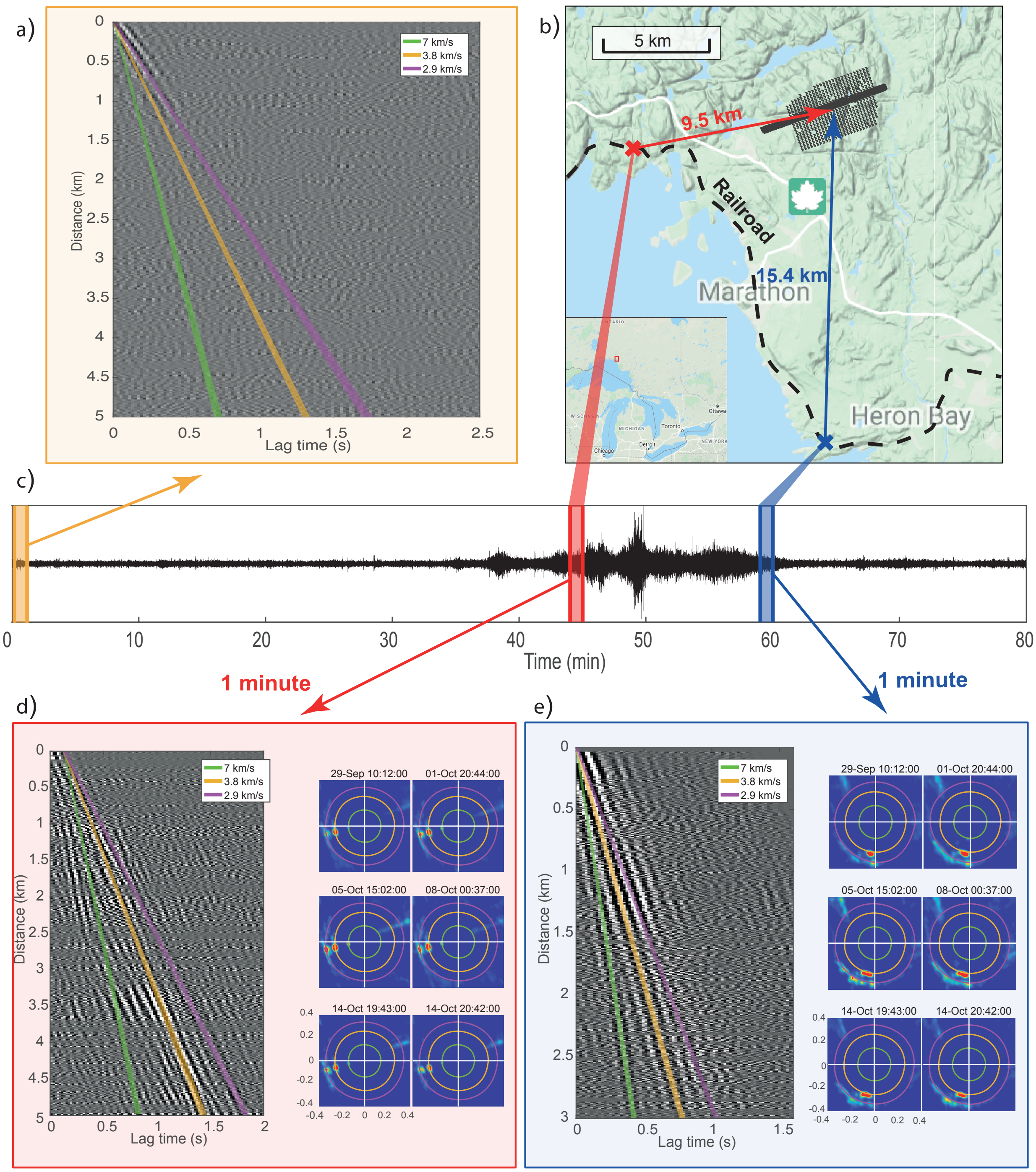

Figure 6: Case studies of train tremor correlations over a dense array at Marathon/Canada. a) One minute cross-correlation for a quiet time period. b)Map of the study zone in the north of Marathon, Ontario, Canada. Grey dots are the 1020 seismic station. The black dashed line is the railroad (CPRS). The red and blue cross are the position studied. c) Train seismic record.d) - e) left: stacked section over 6 train. b) -d) right: 1-minute beamforming panels for 6 train passages. beamforming panels. Left panel: Stack of 6 train passages.

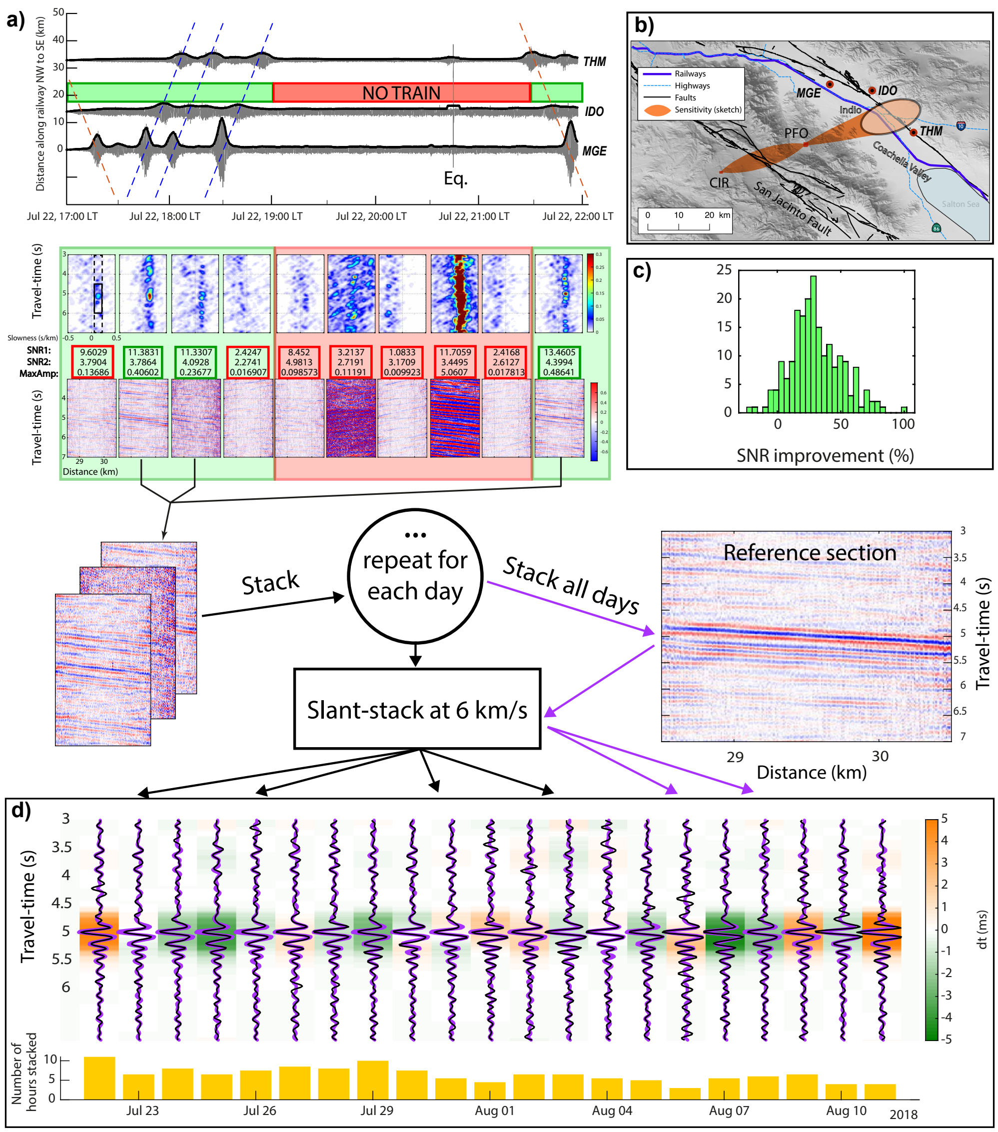

Figure 7: Workflow for monitoring applications (see details in the text). a) Top row: Source detection and characterisation and broad time-windows selection. a) Middle row: Cross-correlation computation and correlation gathers construction for every 30 min-long segments of selected continuous data then correlation gathers selection. a) Bottom row: Stack of 30-min gathers into daily gathers then Reference gather. b) Map of the study area showing the Fresnel zone (orange ellipse) where train signals contribute coherently to the P-waves in the correlations, travelling between PFO and CIR. The railway and the main highway are shown in blue. The active tectonic faults are shown in black. The three broadband stations used for building the train catalogue are show with red and black circles. c) Histogram of the signal-to-noise ratio improvement between the Reference correlation gathers without and with train signal selection. d) Monitoring results: a slant-staked reference gather (purple seismograms) is compared with slant-stacked daily gathers (black seismograms) via cross-wavelet transform to get the instantaneous time-shift between them (shown as the colored background). The bottom histogram shows the number of hours of continuous noise records stacked to obtain the daily correlation gathers.

References

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.

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

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.

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

Brenguier, F., and Coauthors, 2020: Noise-based ballistic wave passive seismic monitoring. part 1: body waves. *Geophysical Journal International*, **221 (1)**, 683–691.

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

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.

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

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.

Dales, P., and Coauthors, 2020: Passive seismic recordings for near surface mineral exploration: The Marathon dataset. *accepted in Seismological Research Letters*.

Davies, D., E. Kelly, and J. Filson, 1971: Vespa process for analysis of seismic signals. *Nature Physical Science*, **232 (27)**, 8–13.

Dong, S., J. Sheng, and J. T. Schuster, 2006: Theory and practice of refraction interferometry.

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.

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

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.

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.

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, doi:10.1785/0120180130, URL https://doi.org/10.1785/0120180130.

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.

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, 2018: 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, doi:10.1785/0220180081.

Mao, S., A. Mordret, M. Campillo, H. Fang, and R. D. van der Hilst, 2020: On the measurement of seismic traveltime changes in the time–frequency domain with wavelet cross-spectrum analysis. *Geophysical Journal International*, **221 (1)**, 550–568.

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.

Nakata, N., P. Boué, F. Brenguier, P. Roux, V. Ferrazzini, and M. Campillo, 2016: Body and surface wave reconstruction from seismic noise correlations between arrays at piton de la fournaise volcano. *Geophysical Research Letters*, **43 (3)**, 1047–1054.

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., 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.

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.

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.

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.

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.

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

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.

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

Silver, P. G., T. M. Daley, F. Niu, and E. L. Majer, 2007: Active source monitoring of cross-well seismic travel time for stress-induced changes. *Bulletin of the Seismological Society of America*, **97 (1B)**, 281–293.

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

Takano, T., F. Brenguier, M. Campillo, A. Peltier, and T. Nishimura, 2020: Noise-based passive ballistic wave seismic monitoring on an active volcano. *Geophysical Journal International*, **220 (1)**, 501–507.

Tsuji, S., K. Yamaoka, R. Ikuta, T. Kunitomo, T. Watanabe, Y. Yoshida, and A. Katsumata, 2018: Secular and coseismic changes in s-wave velocity detected using across in the tokai region. *Earth, Planets and Space*, **70 (1)**, 1–10.

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

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

Zhou, W., and H. Paulssen, 2019: Travel time changes in the groningen gas reservoir by train noise interferometry of borehole data. **2019 (1)**, 1–5, doi:https://doi.org/10.3997/2214-4609.201900671, URL https://www.earthdoc.org/content/papers/10.3997/2214-4609.201900671.