Laura Pinzon-Rincon, François Lavoué, Aurélien Mordret, Pierre Boué, Florent Brenguier, Philippe Dales, Yehuda Ben-Zion, Frank Vernon, Christopher Bean, Daniel Hollis

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

Laura Pinzon-Rincon1,*, François Lavoué1, Aurélien Mordret1, Pierre Boué1, Florent Brenguier1, Philippe Dales1,6, Yehuda Ben-Zion2, Frank Vernon3, Christopher J. Bean4, and Daniel Hollis5

1 - ISTerre, Univ. Grenoble Alpes, CNRS, Grenoble, France

*Corresponding author: laura-alejandra.pinzon-rincon@univ-grenoble-alpes.fr Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France

2 - Department of Earth Sciences and Southern California Earthquake Center, University of Southern California, Los Angeles, CA, USA

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

4 - Geophysics Section, Dublin Institute for Advanced Studies, Dublin, Ireland

5 - Sisprobe, Grenoble, France

6 - Institute of Mine Seismology, USA

=

September 18, 2020

Abstract

Seismologists are eagerly seeking new and preferably low-cost ways to map and track changes in the complex structure of the top few kilometers of the crust. By understanding it better they can build on what is known regarding important, practical issues. These include telling us whether imminent earthquakes and volcanic eruptions are generating tell-tale underground signs of hazard, about mitigation of induced seismicity such as from deep injection of waste water, how the Earth and its atmosphere couple, and where accessible natural resources are. Passive seismic imaging usually relies on blind correlations within extended recordings of Earth's ceaseless "hum" or coda of well-mixed, small vibrations. In this paper we are proposing a complementary approach. It is seismic interferometry using opportune sources - specifically ones not stationary in time and moving in a well-understood configuration. Its interpretation relies on accurate understanding how these sources radiate seismic waves, on precise timing, on careful placement of pairs of listening stations, and on seismic phase differentiation (surface and body waves). Massive freight trains were only recently recognized as just such persistent, powerful cultural (human activity-caused) seismic sources. One train passage may generate tremor with an energy output of a magnitude 1 earthquake and be detectable for up to 100 km from the track. We discuss source mechanisms of train tremor and review basic theory on seismic interferometry with opportune sources. Finally, we present case studies of body- and surface-wave retrieval as an aid to mineral exploration in Canada and to monitoring of a Southern California fault zone. We believe noise recovery from this new signal source, together with dense data acquisition technologies such as nodes or Distributed Acoustic Sensing, will deeply transform our ability to monitor activity in the shallow crust at sharpened resolution in time and space.

1 Introduction

Vehicle traffic was long seen mainly as a pervasive source of nuisance noise that degrades seismic records (Douze and Laster, 1979). But the recent and intriguing discovery of tremor from trains startled seismologists. Studies soon followed on detection and characterization of these signals (Riahi and Gerstoft, 2015; Li et al., 2018; Green et al., 2017; Fuchs et al., 2018; Inbal et al., 2018) as well as source modeling (Lavoué et al., 2020). Earlier studies Nakata et al. (2011); Quiros et al. (2016); Chang et al. (2016) proposed using traffic noise and seismic interferometry for both body- and surface-wave imaging. These studies were, however, limited to highly local sources of background cultural noise and near-surface applications.

In a fortuitous attempt to gather non-volcanic tremors (NVT) along the San Andreas Fault Zone in Southern California, Inbal et al. (2018) discovered extended tremor sequences that shared puzzling similarities with NVTs. But they traced the new discovery to massive freight trains running along the nearby Coachella Valley. They could detect them as much as 100 km from the rails. Brenguier et al. (2019) calculated that a single 1-km-long freight train rolling through a 10-km-long railway section radiates energy equivalent to a magnitude 1 earthquake. By further using seismic interferometry for correlation of this underfoot train noise Brenguier et al. (2019); Dales et al. (2020) showed it possible to extract useful information on the Earth's crustal structure and temporal changes down several kilometers and that provides a potential alternative to costly monitoring of active sources such as hydraulic thumping or explosives (Tsuji et al., 2018).

This paper reviews basic concepts and examples of the application of seismic interferometry to train tremors. Its special focus is on long-range body-wave retrieval for crustal exploration and monitoring (Fig. 1). The Green's function is the elastic impulse response of the ground between a seismic source and a seismic receiver, i.e., the signature of the ground structure encapsulated by its effects on the velocity and other behavior of a signal as it traveled. Seismic interferometry is often able, by correlating diffuse coda or seismic noise, to retrieve the Green's function between two seismic sensors by turning one sensor into a virtual source. The impact of Green's function retrieval in recent decades is revolutionary (e.g., Campillo and Paul, 2003; Shapiro et al., 2005; Snieder, 2004; Wapenaar, 2004). It spurred publication of at least 2000 seismology papers in the last 15 years. One payoff from seismic interferometry and Green's function retrieval is improved crustal imaging through correlation of 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 (e.g., 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 application of Green's function retrieval and seismic interferometry requires correlation of either a fully diffused seismic wavefield or noise signals generated from all around the studied region, including at depth (Wapenaar, 2004). In practice these demands are never met. Seismologists must live with or find work-arounds for partial reconstructions and potentially biased wave travel times (Snieder et al., 2006; King and Curtis, 2012). Moving trains are welcome, opportune sources of noise on well-mapped railways. It is essential that they be rigorously assessed for seismic interferometry. Train traffic noise cannot be blindly correlated without considering the effects of irregular source distribution on body-wave retrieval.

In this paper we first describe typical train noise signals, discuss recent models of mechanisms that create train seismic radiations, and provide a map of the predicted extent of useful train noise in the contiguous US. Second, we propose a methodological framework focusing our approach on the stationary zones (geographical area where we observe constructive interferences when cross-correlating signals between two stations, Snieder, 2004) and propose a signal processing strategy for applying seismic interferometry to train noise with focus on long-range body-wave retrieval. We finally review two recent case studies regarding mineral exploration in Canada and tectonic fault monitoring in Southern California.

2 The sound of trains in the Earth

As noted, massive freight trains generate seismic waveforms with striking similarity to episodic tectonic tremors. These may be from such events as slow-slip fault motion (Fig. 2 top). As Inbal et al. (2018) report, the identity of the sources as manmade 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 also in the range of reported tectonic tremor migration velocity at depth. Train hum has however a distinct 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, see Section 5).

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 among trains, tracks, and ground. In a recent study, Lavoué et al. (2020) showed that excitation due to axle loads is the main mechanism producing the spectral characteristics of seismic signals 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). 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) along the railway (Krylov and Ferguson, 1994; Lavoué et al., 2020).

Lavoué et al. (2020) conclude that the spectral lines arise from complex interactions of periodic loads through the regularly spaced wheels on the even more evenly separated 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. We provide an open-source code to assess the frequency response of a specific train (see link in Data and Resources). With most trains, the dominant spectral lines are expected in the 1 to 20 Hz range, 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).

Our ability to predict the long-range, body-wave Peak Ground Velocity (PGV) of a moving train tremor – the physical motion in the medium as signals go through it - is crucial to image formations and monitoring any changes 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 is a square-root function of train length for constant wagon weight. Faster trains also generate higher PGVs. Moreover, the ground stiffness beneath railways controls high-frequency content and amplitude of excitation (trains traveling across 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. While maximum detection distance may be limited (a few kilometers) in sedimentary basins due to attenuation and weak excitation but, again, 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 from as far as 90 km from the railway. At 45 km from the railway, they estimated a PGV of about 6 × 10−8 m/s. By applying a simple correction for intrinsic attenuation and geometrical spreading for body (P) waves, we now estimate that the level of PGV for a specific Coachella Valley train would be of the order of 5 × 10−7 m/s at 10 km and 5 × 10−6 m/s at 1 km. These values are quite low. Even high-sensitivity seismometers may record such train signals at long distances only in quiet environments. Nevertheless, Brenguier et al. (2019) demonstrated that Coachella Valley train noise supported seismic body-wave interferometry - with data recorded by an array of geophones (nodes) - as much as 60 km from the railway (see Section 6).

At shorter 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. But it should be possible to extract body waves from the correlations of train tremors by stacking data from several passages (see Section 5). 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.

Observation from previous studies (Inbal et al., 2018; Brenguier et al 2019) persuade us that 50 km is a typical maximum distance range for detecting tremors generated from large North American freight trains. That led us to look into the spatial extent of detectable train tremor in the entire contiguous US plus southern Canada (Fig. 3). The map displays the main freight railway routes. The swathes in colors represent high tonnage routes. Their width (100 km) is a rough guide to potential long-range train tremor detection scope. This map does not take into account the reduced detectability of signals in urban areas due to intense local noise and in sedimentary basins with strong attenuation compared to the Southern California-Coachella Valley reference.

Noteworthy is that the Coachella Valley, a stretch of Sonoran Desert northwest of the Salton Sea, is a singularly apt place to find practical uses for ground vibrations of massive trains. Its Union Pacific RR tracks are a prime corridor to and from the ports of Los Angeles and Long Beach, the Western Hemisphere's busiest seaport complex. Dozens of trains go through daily. The average length is more than 2.5 km with more than 100 cars, often including multiple engines front and back. Rail enthusiasts visit to make, and often to post on YouTube, mesmerizing videos of the immense steel caravans rumbling by (see Data and Resources).

Annual freight tonnage (Fig. 3) is a proxy for the number of trains travelling on rail sections. Assuming an average train length of 2 km and a weight of 15 kilotons (according to statistics derived from the public waybill samples, 20181), a tonnage of 100 MT/year corresponds to about 18 trains per day. The number of trains per day will affect ability to stack the reconstructed body waves from seismic interferometry. It also affects the temporal resolution needed in monitoring applications (see Section 6). This map highlights the potential of using trains as sources of opportunity. Potential application may be in Cascade volcanoes, the Southern California's San Andreas Fault system, induced seismicity (e.g. Oklahoma), and resource exploration and monitoring (minerals, water).

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

3 Seismic interferometry with opportune sources

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 turn sensors into virtual sources, this concept has been refined in seismology and seismic exploration, mostly in the last 20 years, based on the pioneering studies of random fields or vertical planar wave autocorrelation (Aki, 1957; Claerbout, 1968) and the time-reversal principle in acoustics (Fink, 1997).

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., 2005) and/or an equipartition of modes defining a diffuse field (e.g., Sánchez-Sesma and Campillo, 2006). The stationary phase condition implies that the correlation function's convergence towards the Green function requires the presence of sources (or scatterers) in line with two carefully placed receivers. In a 2D homogeneous medium, these stationary points define a hyperbolic area, outward from the receiver pair, with an aperture that is frequency dependent (the lower the frequency, the broader the calculated source region). Also known as Fresnel zones, these "kernels" are clues to the reliability of the correlation's implied source locations. In 3D and for both surface and body wave retrieval, the requirement of equipartition remains. Full Green function retrieval demands sources evenly distributed, along an arbitrarily shaped surface enclosing the two sensors (e.g., Wapenaar, 2004; Wapenaar and Fokkema, 2006). However, even with a clearly dominant distribution of sources at the free surface, several studies confirmed the feasibility of retrieving body waves (e.g., Draganov et al., 2009, 2013), and even explicitly using traffic noise (Nakata et al., 2011).

Each of the possible phases (or wave types) included in the Green's function has its own source sensitivity. The main contributors to a particular phase are sources within its stationary phase area. We can therefore measure a specific phase between two receivers by correlation of a source within its stationary phase zone including the surface. The following case studies investigated P waves from moving trains, and emerging 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, Fig. 4b).

Useful interference occurs if the seismic sources (trains) satisfy the stationary phase criterion:

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

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. Note that using somehow 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. For train signals, we need to characterize the source and of course take into account that the sources are moving.

One reason train signals are practical for interferometric studies is that we can easily detect, or learn in advance, that a train is coming. If a railway is sufficiently close to a targeted area, a single train's motion could illuminate many azimuths and potentially different depths. Figure 4a shows an example of geometry in Marathon (Ontario, Canada). There a railway essentially surrounds a temporary array put in to assess an ore deposit (detailed in the following section). 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. Figure 4b to 4d 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 4b) 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 4c); See also a more problematic interference between two diving wave, or a head wave recorded at the two stations (Figure 4d). 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).

We decided to try to illuminate specific ray paths by using a data processing workflow, starting with the selection of short time windows including specific train passages. This presumably could be extended to any kind of seismic tremors and should help extract body waves between well-selected pairs of stations useful for imaging and monitoring studies.

4 Strategy for data processing

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 we propose 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. This approach is illustrated is sections 5 and 6 for imaging and monitoring applications, respectively. Figure 5 summarizes the five main stages of our data processing in comparison to the classical method of continuous blind data correlation.

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 section 2, 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.

- 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. Figure 4a illustrates 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.

- Stack (by events, by azimuth): To improve SNR, we stack the cross-correlations over different events. Cultural sources such as train traffic have the advantage of reliability and frequent repetition.

- Measurement and analysis: Depending on the type of studies, various approaches such as travel time measurements can enhance imaging and monitoring applications.

5 Body- and surface-wave retrieval from correlations of train tremors applied to mineral exploration

We investigated a region near Marathon, Ontario, Canada (see Fig. 6b) where potential targets include a high concentration of platinum group metals and minor Cu in a gabbro intrusion. Reconstruction of high-frequency body-waves from train noise correlations was of significant interest. A reason is such signals' sharp sensitivity to seismic velocity contrasts at depth, offering a clear path to imaging geological boundaries. In the fall of 2018 we 1020 seismic stations in a backbone array and a dense station line (see Fig. 6b). We recorded 30 straight days of seismic signals.

Dales et al. (2020) showed that the main generators of high-frequency seismic noise in Marathon are freight trains to the southwest. They reinforced earlier evidence that by selectively using train traffic noise, one retrieves body waves better than does correlating a more extended or full noise record. Dales et al. (2020) stacked correlations over 1 month, selecting all periods during which the ambient noise came from the direction aligned with a dense W-E line of sensors. Their study is illustrative but the results did not allow them to perform 3D imaging. We moved a step further by separating and binning noise azimuths for virtual source retrieval in different directions. Following the workflow proposed in section 4, 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.

A more detailed workflow follows:

- We first generated a catalog of train passages with the covariance matrix method proposed by Seydoux et al. (2016). This method uses the spatial coherence of the signals to detect emergent signals in the noise. We applied the procedure to the entire data set day by day and detected 207 train passages in 30 days. We retained for study approx. 180 events after skipping overlapping trains. Beamforming concluded that the array receives energy from each train for approx. 80 minutes.

- Second, we extracted train signals from the rest of the recording, and selected station pairs in line with train positions. To determine position, we did beamforming within 1-minute-long windows using data filtered between 8 and 16 Hz (Fig. 6-d and e, the right side shows 6 beamforming panels for 6 different events at two different times). Each panel corresponds to a one-minute beamforming time window and one single train passage. We saw that with properly selected time windows for each event, we got a tight and usable ranges of azimuth. We assumed that the main source of energy was the train and noted the maximum beam power. We back-projected this signal onto the railway to locate each train minute by minute. Figure 6-b (red and blue cross) shows the train position from the first beamforming panel (i.e., one single train). We then selected station pairs that are in line with the train position for each minute, always taking the station closest to the railway as a virtual source (red and blue arrows in Fig. 6b). We applied an azimuthal filter of +/- 5 degrees for each station pair with respect to the train position.

- Third, we cross-correlated the selected station pairs minute by minute without overlap and for each event (i.e., train passage). Filters excluded signals outside 15 to 40Hz to avoid surface waves. We stacked cross-correlations according to their inter-station distances and collected them in distance-binned correlation gathers for the selected station pairs (second step). In contrast, Figure 6a shows the stack of one-minute cross-correlation for a quiet period (i.e., non-train passage), highlighting the absence of coherent wave propagation in this rather high-frequency window.

- In the fourth and last step we stacked events sharing the same train azimuth. We stacked these correlation gathers into a reference azimuthal gather. We converged on a stable reference stack from 6 train passages. We showed that by applying the workflow explained in section 4 we only needed one minute of data and stacking over the 6 events to retrieve body-waves. Figures 6-d and e, left side, show the stacked section over 6 train passages with one-minute data segments.

We retrieved two dominant arrivals with an apparent velocity of 3.8 km/s and 7 km/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. One can use both P- and S-waves plus high-frequency surface waves jointly for imaging the subsurface. We need further analysis to assess the types of body-waves (direct, refracted) and how one can use velocity variations in the azimuth's function for 3D imaging.

6 Retrieving long-range body waves from train-tremor correlations to monitor the San Jacinto Fault Zone

Following studies by Nakata et al. (2015) and Nakata et al. (2016) of high-frequency body- wave retrieval using dense seismic receiver arrays Takano et al. (2020), Brenguier et al. (2020) and Zhou and Paulssen (2020) explored ways to monitor temporal changes of ballistic wave velocities. In this section, we employ opportune seismic sources to passively monitor temporal changes and revisit the experiment of Brenguier et al. (2019). Here, the goal was 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, parallel to the San Andreas and part of the same fault complex).

Brenguier et al. (2019) showed that standard ambient noise correlation processing can retrieve high-frequency direct P-waves that traveled between two arrays, one at Piñon Flat Observatory (PFO) and the other on the Cahuilla Reservation (CIR, Fig. 7). The main sources of these P-waves were Coachella Valley freight trains traversing the Coachella Valley about 30 km to the East-North-East of PFO. Brenguier et al. (2019) used full records of ambient noise to obtain stable direct P-wave seismograms. We showed that, by carefully selecting time-windows where most of the energy is generated by trains, the quality and spatiotemporal stability of the reconstructed P-waves rose. As described in Figure 5, the standard three-step noise correlation computation workflow was replaced by a four-step procedure to correlate only the main source of opportune energy i.e., here, trains. Our workflow:

- First, we built our train catalog for the period of interest (July 22 to August 11 of 2018) using three broadband stations (MGE, IDO, and THM of the CI network, Fig. 7b) near the railway in the Coachella Valley.) After band-pass filtering the continuous data between 0.75-5 Hz we slant-stacked the envelopes of the continuous seismograms with apparent velocities of plus or minus ~95 km/h (dashed blue and orange lines in Fig. 7c). This procedure detected trains passing through the stationary phase zone (Fig. 7b) both North to South and South to North.

- With the catalog in hand, we sorted broad time-windows, those with and without train tremors (large green and red shaded rectangle in Fig. 7c, respectively).

- In the third step, to analyze the dense nodal array data, we cross-correlated the selected time-windows, (green rectangles in Fig 7c) and 8a)i), filtered between 3 and 10 Hz and using non-overlapping data segments of 30 min.

- We next stacked the cross-correlations, according to their inter-station separations, into distance-binned correlation gathers. We highlight only correlation gathers for the causal part of the correlations (from PFO to CIR), for a time-window centered at the P-wave arrival time. These 30 min correlation gathers were further pruned based on three quality criteria seen in their vespagram, indicators of the different waves' apparent velocities observed in the correlation gathers. (Davies et al., 1971). Figure 8a)ii shows the vespagrams associated with the correlation gathers in the upper panels (Fig. 8a)i). The three quality criteria were: 1) SNR1, the ratio between the maximum vespagram amplitude in the [0.13-0.2] s/km slowness (inverse of velocity [5-7.5] km/s velocity) window (dashed black rectangle in the leftmost vespagram panel, Fig. 8a)ii 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 × [4.5-6] s travel-time window (solid black rectangle in the leftmost vespagram panel, Fig. 8a)ii and the RMS in the rest of the [0.13-0.2] s/km slowness window. 3) MaxAmp, the peak vespagram amplitude in the [0.13-0.2] s/km slowness × [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 arrival times that are either too early or too late, even though their apparent velocity is correct. We used the MaxAmp criteria to reject gathers for which the expected P- wave phase is not energetic enough or is too large for a train signal, indicating the detection of an earthquake located in the Fresnel zone (Fig. 8a)i). For this specific situation, we set thresholds to be sure the conditions SNR1 ≥ 2.5, SNR2 ≥ 1.5, and 0.15 ≤ MaxAmp ≤ 4.0 were met for a correlation gather to be selected (green boxes in Fig. 8a)iii). The actual values for SNR1, SNR2 and MaxAmp are shown below each vespagram in Figure 8a.

- In the last step, we stacked the selected correlation gathers into daily gathers and a reference gather including every selected gather for the whole period of interest (Fig. 8a)vi). Ultimately, we used less than 20% of the full dataset for the monitoring measurements (Fig. 8b).

To quantify the improvement of the signals using the opportune sources approach, we measured the ratio of SNRs between a reference gather computed with all the data (Fig. 8c)i, similar to Brenguier et al. (2019) and the reference gather from selected train windows shown in Fig. 8a)iv. We performed this operation for each waveform in the gathers.

The results (Fig. 8c)ii) show that the opportune source concentration improves the SNR of the P-wave signal by an average of more than 25%. This has important implications for monitoring. As Silver et al. (2007) showed, the SNR is the main factor controlling the precision of a time delay measurement between two similar waveforms; such 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 final process is still ahead of us but the methodology proved effective.

-The final step of the workflow was to measure seismic velocities. Different approaches were available. We chose to measure 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 red traces in Fig. 8b, respectively). We measured 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). Although a time-delay was determined for each sample of the waveform, we only show δt values for the direct P-wave. Here, the time-shifts we found are shorter than 0.1% of the propagation time, corresponding to time-shift shorter than 5 ms between the daily and reference seismograms. These time-shifts can be translated into relative velocity changes with the relation $\delta v/v = -\delta t/t$, using the absolute travel-time t of the slant-stacked P-wave. We obtained velocity changes on the order of ±0.1%. The meaning of these values is difficult to know because it will take a lot more work to understand the exact sensitivity of the reconstructed P-wave and the different trade-offs among the source and structure sensitivities (see Fig. 4). We plan to estimate 3D spatial sensitivity kernels for these retrieved travel time perturbations and correct for shallow, environmental velocity changes. Thus, we shall see whether we will soon be observing and locating any places where changes in seismic velocity at a few kilometers depth occur within the San Jacinto Fault Zone.

7 Discussion and conclusions

We see great opportunity for exploiting any available massive freight train noise recovery to improve crustal imaging and monitoring dramatically. We describe applications to North America but our conclusions have global ramifications, 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. The latter are lighter than freight trains and generate less energetic tremors to be sure, but applications might be found in near-surface environmental or engineering studies.

For all its potential, to put heavy freight train noise to work for seismic imaging and monitoring reasons is of course limited to regions near railways. It also requires trains traveling at rather high speed. But generally, this paper presents a workflow for using other and more local sources of cultural noise, including car and truck traffic, wind farms, and natural sources such as surf break or tectonic, volcanic tremor, as opportune sources of useful seismic noise.

Although promising, this work poses important, practical challenges that the field must confront. Most important is to improve understanding of the retrieved body and surface waves' spatial sensitivity to crustal structures when combining 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 not only to the structure between the receivers but contrarily can also blur the overall picture due to interference in areas between the noise source and the receivers. This latter downside may induce misleading interpretations of velocity or velocity change measurements.

A drawback of examples in this paper is that they used so many sensors (hundreds). Train vibrations cost seismologists nothing but recording them is not yet easy. One solution to overcome these limitations is to find a way to use permanent, single seismic stations instead of costly temporary arrays. One potential initial approach is to deploy dense but temporary seismic arrays around permanent seismic stations. This may help to identify useful phases emanating from noise correlations of opportune sources. A hope is that we learn enough to extract the needed information on a long-term basis with permanent stations alone.

One additional major advance would be to couple Distributed Acoustic Sensing data (Zhan, 2020) to seismic interferometry with opportune sources, as described by Dou et al. (2017) for car traffic and near-surface applications. This indicates potential for reconstructing widespread virtual sources along fiber-optics from correlations of both short- and long-range opportune sources. Success will open the path to many applications including water resource management in the near-surface and earthquake studies at greater depth.

8 Data and Resources

The Marathon dataset will be released in June 2021. It will either be hosted online or freely sent on external hard disks upon request via the website for 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.

Open-source codes reproducing Lavoué et al. (2020)'s results are available at https://gricad-gitlab.univ-grenoble-alpes.fr/pacific/publications/2020_Lavoue-et-al_SRL_supplemental-material.

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

Figure 3 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, which we could unfortunately not access directly.

Coachella Valley train video can be found at https://www.youtube.com/watch?v=pE0LYuf7_F8

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. 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). We are grateful to the reviewers for their useful comments.

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.

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.

Cheng, Y., Ben-Zion, Y., Brenguier, F., Johnson, C. W., Li, Z., Share, P. E., ... & Vernon, F. (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.

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: 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.

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

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.

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, 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.

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 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 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 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 , 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 . 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 . Geophysics, 71 (4), SI33–SI46, doi:10.1190/1.2213955.

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

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

Zhou, W., and H. Paulssen, 2020: Compaction of the Groningen gas reservoir investigated with train noise. Geophysical Journal International, doi:10.1093/gji/ggaa364.

List of Figures

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

Figure 2: Top: A train tremor recorded 3 km away from a seismic station in Marathon, Canada. Bottom: Spectrogram showing clear spectral lines oscillating as train speed varies.

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 (sections with annual tonnage < 10 MT/year are ignored). Colored lines are 100-km-thick, which is an indication for the distance from which we may detect individual train tremors (50 km from the railway, see details in the text).

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

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

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 stations. 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 trains. d) -e) right: 1-minute beamforming panels for 6 train passages.

Figure 7: Detection of train passages in the Coachella Valley. a): Layout of the dense nodal arrays used for the monitoring of the San Jacinto Fault. b) Map of the study area showing the Fresnel zone (yellow

ellipse) where train signals contribute coherently to the P-waves in the correlations, travelling between PFO and CIR. The orange areas show the zones of sensitivity to the structure for a wave reconstructed by interferometry between the two arrays. 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 shown with red and black circles. c) The train catalog is built by slant-stacking the envelops of the train tremors. The colored rectangles show the location of the Fresnel zone along the railway, green for a time-window with trains, red for a time-window without train and rejected from further analysis.

Figure 8: Workflow for monitoring applications. a) i: Cross-correlation computation and correlation gathers construction for every 30 min-long segments of selected continuous data from the dens arrays. a) ii: Vespagrams of the correlation gathers used for the 30-min window selection. The black rectangles in the leftmost panel are used to measure the different selection criteria. a) iii: The three selection criteria associated to each 30-min window. The red boxes are rejected, the green boxes are kept for the next step. a) iv: Stack of the selected 30-min gathers into daily gathers, then the daily gathers into the Reference gather. b) Monitoring results. The bottom histogram shows the number of hours of continuous noise records stacked to obtain the daily correlation gathers. c) i: Stack of every 30-min windows without train selection, similar to the section shown in Brenguier et al. (2019). c)ii: Histogram of the signal-to-noise ratio improvement between the Reference correlation gathers without and with train signal selection

Authors:

Laura Pinzon-Rincon: laura-alejandra.pinzon-rincon@univ-grenoble-alpes.fr, Université602Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France.

François Lavoué: francois.lavoue@univ-grenoble-alpes.fr, Université Grenoble Alpes, ISTerre,604CS 40700, 38058 GRENOBLE Cedex 9, France

Aurélien Mordret: aurelien.mordret@univ-grenoble-alpes.fr, Université Grenoble Alpes,606ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France

Pierre Boué: pierre.boue@univ-grenoble-alpes.fr, Université Grenoble Alpes, ISTerre, CS60840700, 38058 GRENOBLE Cedex 9, France

Florent Brenguier: florent.brenguier@univ-grenoble-alpes.fr, Université Grenoble Alpes,610ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France

Philippe Dales: pdales@gmail.com, Institute of Mine Seismology 19 Jeanine Street Sudbury612Ontario, P3B 0E5 Canada

Yehuda Ben-Zion: benzion@usc.edu, Department of Earth Sciences and Southern California Earthquake Center, University of Southern California, Zumberge Hall, Room 109, USC Earth Sciences, Los Angeles, CA 90089-0740

Frank Vernon: flvernon@ucsd.edu, IGPP, Scripps Institution of Oceanography, UC San617Diego, 9500 Gilman Dr #0225, La Jolla, CA 92093-0225,CA, USA

Christopher J. Bean: chris.bean@dias.ie, Geophysics Section, Dublin Institute for Advanced Studies, 5 Merrion Square North, Dublin 2, D02 Y006, Ireland

Daniel Hollis: dan.hollis@sisprobe.com, Sisprobe, IMAG building, office #230, 700 avenue Centrale, 38401 Saint Martin d'Héres, FRANCE