Geophysical Journal International

Geophys. J. Int. (2023) **233,** 964–977
Advance Access publication 2022 December 23
GJI Seismology

https://doi.org/10.1093/gji/ggac507

Imaging shallow structures using interferometry of seismic body waves generated by train traffic

M. Rezaeifar [1,*] F. Lavoué,[2] G. Maggio,[1] Y. Xu,[1,3] C. J. Bean,[1] L. Pinzon-Rincon,[2] S. Lebedev [1,3] and F. Brenguier [2]

[1] *School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin* 2, *Ireland. E-mail: mhmr20@gmail.com*
[2] *Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, Université Gustave Eiffel, ISTerre,* 38000 *Grenoble, France*
[3] *Department of Earth Sciences, Bullard Laboratories, University of Cambridge, Cambridge CB*30*EZ, UK*

Accepted 2022 December 21. Received 2022 November 9; in original form 2022 February 25

SUMMARY

Train-induced vibrations act as potential powerful high-frequency source for imaging subsurface with higher resolution than typical ambient noise interferometry. In this study, we present results of seismic interferometry applied on three days of railroad traffic data recorded by an array of seismographs along a railway in Dublin, Ireland. Our virtual shot gathers show significant surface and body wave energy that could be used for seismic interferometry. Reflection sections obtained with our interferometry approaches applied on selected time windows of train-induced vibrations is consistent with nearby borehole data and an active seismic profile. The consistency of the results given by these approaches confirms that train-generated vibrations represent a valuable source of signal for high-resolution subsurface imaging. Furthermore, our results show spurious arrivals that are due to the train geometry and also the cross-correlation approach that needs consideration for body wave interferometry studies.

Key words: Body waves; Seismic interferometry; Seismic noise; Crustal structure.

1 INTRODUCTION

Although active seismic reflection techniques provide high-resolution images of the subsurface (Salisbury *et al.* 2007; Malehmir *et al.* 2012), they are expensive, the methodology can be environmentally destructive (since explosives are often used to achieve an adequate signal-to-noise [S/N] ratio), and some places can be inaccessible to active sources. This brings new opportunities and motivation to develop and introduce new, more efficient high-resolution passive seismic methods.

Seismic interferometry, a technique that uses the correlation between two stations to retrieve the Green's function of the medium, is a passive seismic technique that potentially can overcome some of the above-mentioned limitations and offer an independent source of information to active seismic, although not at the same high resolution (Snieder 2004; Roux *et al.* 2005; Shapiro *et al.* 2005; Curtis *et al.* 2006; Draganov *et al.* 2007, 2009; Xu *et al.* 2021; Hariri Naghadeh *et al.* 2021). Seismic interferometry based on interstation correlations of ambient noise has developed into a standard tool for exploring and monitoring the Earth's interior (Shapiro *et al.* 2005; Brenguier *et al.* 2008; Fichtner 2015; Liu *et al.* 2021; Pinzon-Rincon *et al.* 2021; Hariri Naghadeh *et al.* 2021). In general, this method is based on the extraction of the surface wave contribution to the seismic wavefield from the cross-correlation of seismic noise between the station pairs (Snieder 2004; Roux *et al.* 2005; Shapiro *et al.* 2005; Curtis *et al.* 2006; Roux 2009; Liu *et al.* 2021; Xu *et al.* 2021). Another application of active seismic interferometry is to extract body waves and retrieve the Earth's reflection response from cross-correlations of seismic noise recordings (Draganov *et al.* 2007, 2009; Nakata *et al.* 2011; Brenguier *et al.* 2019; Dales *et al.* 2020; Liu *et al.* 2021; Pinzon-Rincon *et al.* 2021; Hariri Naghadeh *et al.* 2021). Compared to surface and body wave extraction and reflection retrieval is a much greater challenge because ambient noise is typically dominated by surface wave energy (Draganov *et al.* 2007, 2009; Panea *et al.* 2014; Nakata *et al.* 2015; Dales *et al.* 2020; Liu *et al.* 2021; Pinzon-Rincon *et al.* 2021; Chamarczuk *et al.* 2021; Hariri Naghadeh *et al.* 2021). In theory, the images produced with body wave seismic interferometry techniques are comparable to those produced with conventional reflection seismic data (Artman 2006; Draganov *et al.* 2009; Panea *et al.* 2014; Chamarczuk *et al.* 2021; Hariri Naghadeh *et al.* 2021). Nakata *et al.* (2011) and Quiros *et al.* (2016) demonstrated the potential of exploiting train-generated body waves using seismic interferometry to characterize the seismic wavefield and to image subsurface structures.

In this study, we demonstrated the potential of train-induced vibration for subsurface imaging using body wave seismic interfrometry. We applied Common-Mid-Point Cross-Correlation (CMP-CC)

*Now at: Geophysics Section, School of Cosmic Physics, Dublin Institute for Advanced Studies, 5 Merrion Square North, Dublin 2, Ireland.

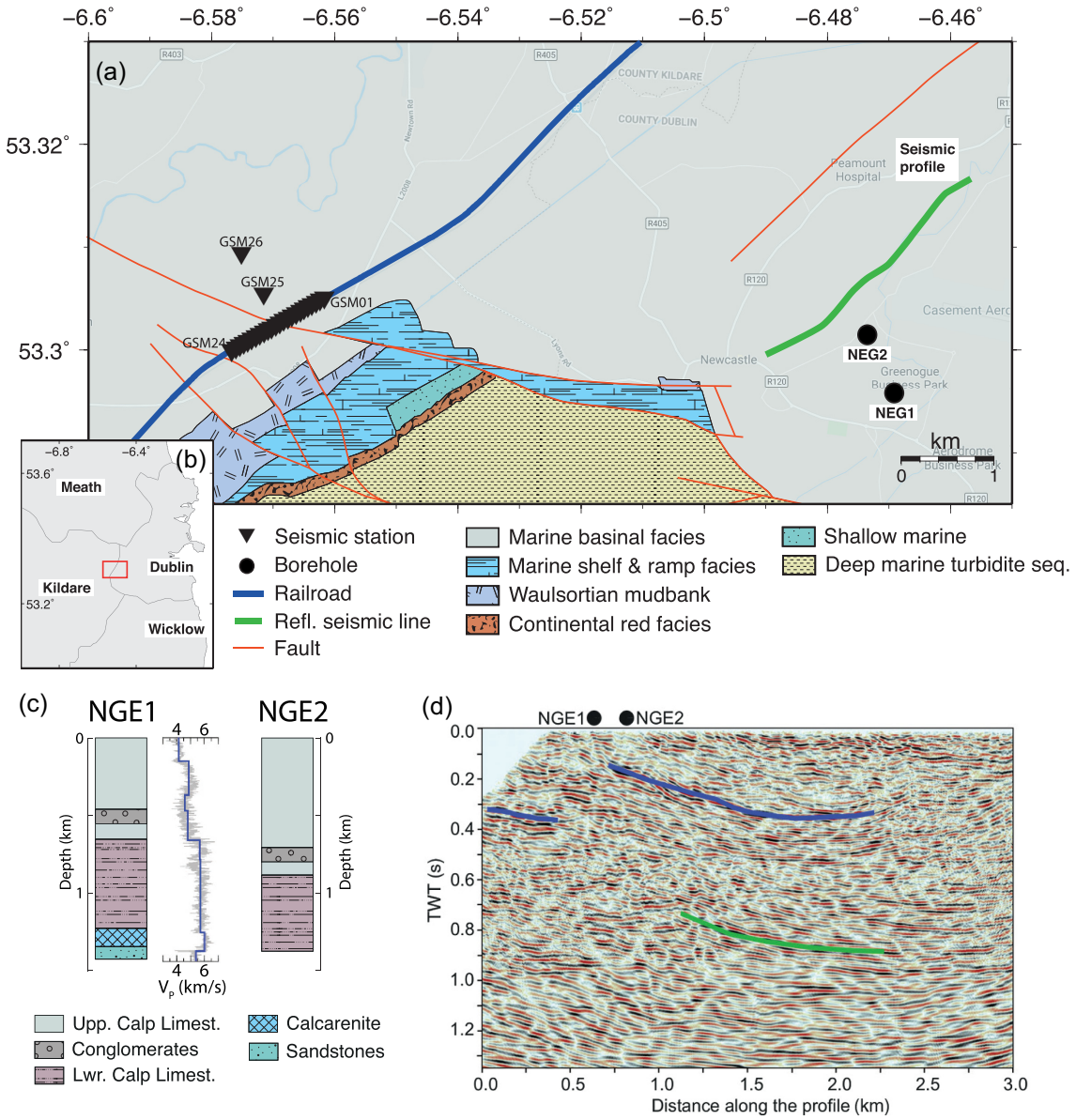

Figure 1. (a) Location of the seismic stations (black triangles) deployed along a railroad (bold blue line) located at the southern margin of the Dublin Basin. Geology map derived from Geological Survey Ireland (GSI) 1:100,000 bedrock map. Also shown are an active reflection seismic profile (bold green line), with the same orientation as the railroad, and boreholes NGE1 and NGE2 (black circles). (b) Inset map shows survey location (red rectangle) relative to Dublin city. (c) Lithostratigraphies from boreholes NGE1 and NGE2 with the sonic log (grey line) and average interval velocities (blue line) for NGE1. (d) Seismic reflection profile with interpretation of the main reflectors from Licciardi & Piana Agostinetti (2017). The location of the boreholes and reflection profile is on map (a).

approach as proposed in Quiros *et al.* (2016) with some modifications specifically relating to how time windows are chosen when applying the imaging procedure. The major difference of our approach compared to the previous studies is that we selected the time windows when the train is at each station to apply cross-correlation instead of the time windows when the train is approaching or receding from the array. Consequently, this specific time window selection allows us to consider the train source as an active source to image the subsurface structures. We tested this technique on simple simulated data and then applied it on field data recorded by a profile of seismometers deployed along a railroad located at the southern margin of the Dublin Basin (Fig. 1).

In addition to the CMP-CC approach, we applied and tested autocorrelation approach on the same selected time windows as used for cross-correlation. Autocorrelation results could help us to validate the CMP-CC results and to better understand the possible artefact caused by processing steps or source itself..

2 DATA

In 2019 November–December, a pilot test seismic array consisting of 24 seismic stations (equipped with three-components 1-Hz Lennartz LE-3Dlite sensors) was deployed with 50 m interstation spacing parallel and very close to the railroad (at ∼5 m distance from the train track) across the southern margin of the Dublin Basin (length of the along track array 1150 m; Fig. 1). Two additional stations were also deployed in a line perpendicular to the centre of

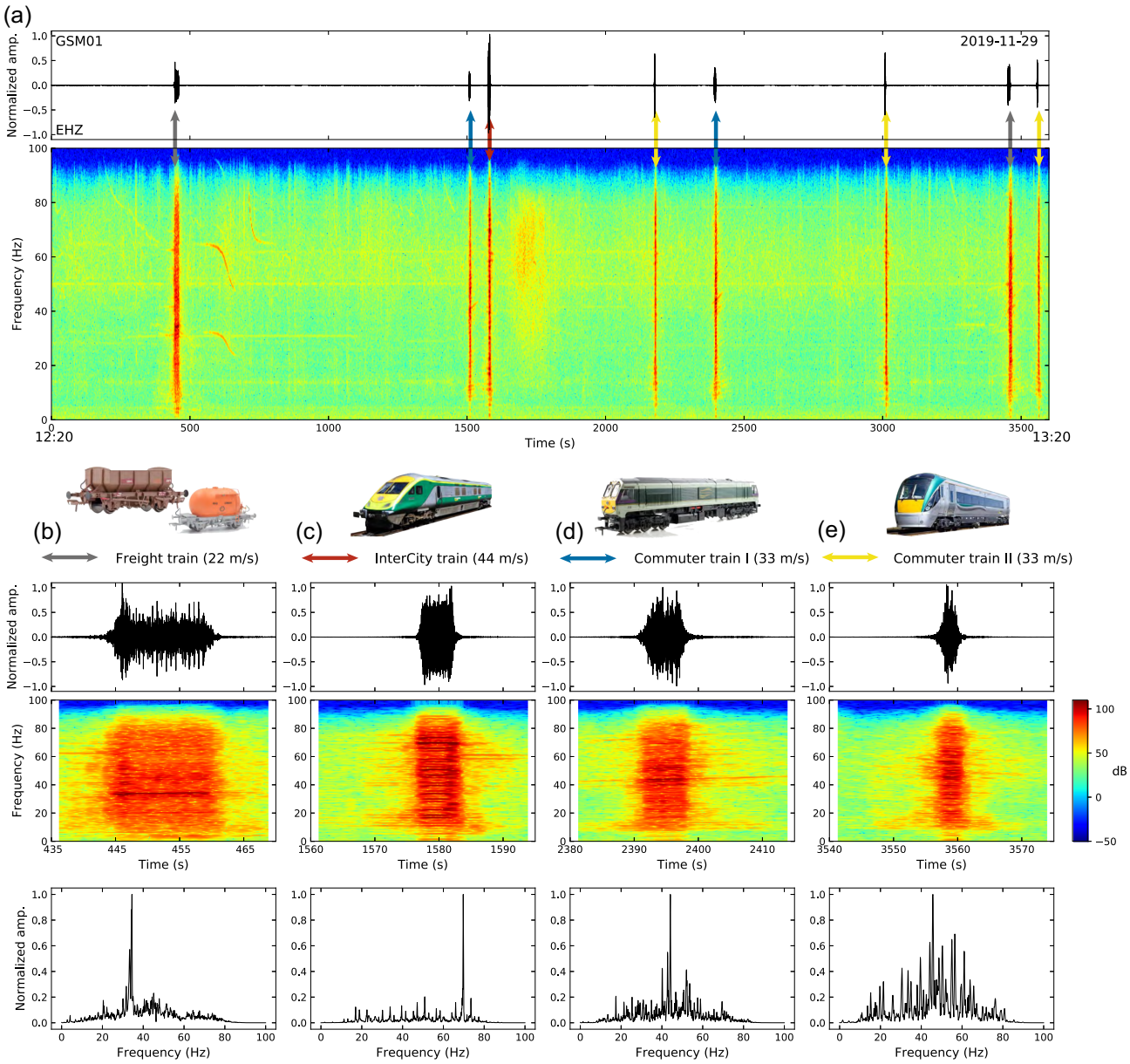

Figure 2. (a) Ground vibration waveform (upper panel) and spectrogram (lower panel) for a 1-hr-long signal recorded at station GSM01 at 5 m distance from the train track. Coloured vertical arrows mark train signals, with colours indicating train types (https://www.irishrail.ie/about-us/iarnrod-eireann-fleet, last accessed: 2020 September). (b)–(e) Detailed view of Fourier analysis of individual train signals. Top row: ground vibration waveform for different train types. Middle row: spectrograms. Bottom row: amplitude spectra of the traces. Spectrograms were calculated with time windows of 5 s and 90 per cent overlap. Note the frequency cutoffs toward 100 Hz (due to the 200 Hz data sampling rate) and below 2 Hz (high-pass filtering). The main difference between Commuter train I and II is the locomotive type. However, number of carriage might also differ.

the array at ∼350 and ∼900 m distance from the railroad (Fig. 1) to help us better understand the train-generated signal as explained by Lavoué *et al.* (2021). Due to the high dynamic range of the instruments, there was no clipping of the waveforms for being so close to the track.

The site was chosen for the following reasons: (1) it has one of the largest volumes of train traffic in the area, approximately 100 trains per day, (2) it crosses the main basin-bounding fault zone almost perpendicularly, (3) there are deep boreholes and active seismic reflection data close to the area and (4) we were able to record the GPS tracks of some of the recorded trains.[1]

For the data acquisition, we used a sampling rate of 0.005 s (200 Hz) corresponding to a Nyquist frequency of 100 Hz. Using the pilot test array, we acquired data continuously for 72 hr between 2019 November 29 and December 2 with good data recovery as only

[1] During the deployment, we recorded the train GPS tracks for some of the trains using a handheld GPS on board (e.g. for few trains, someone travelled with the train—middle carriage—to record the GPS track).

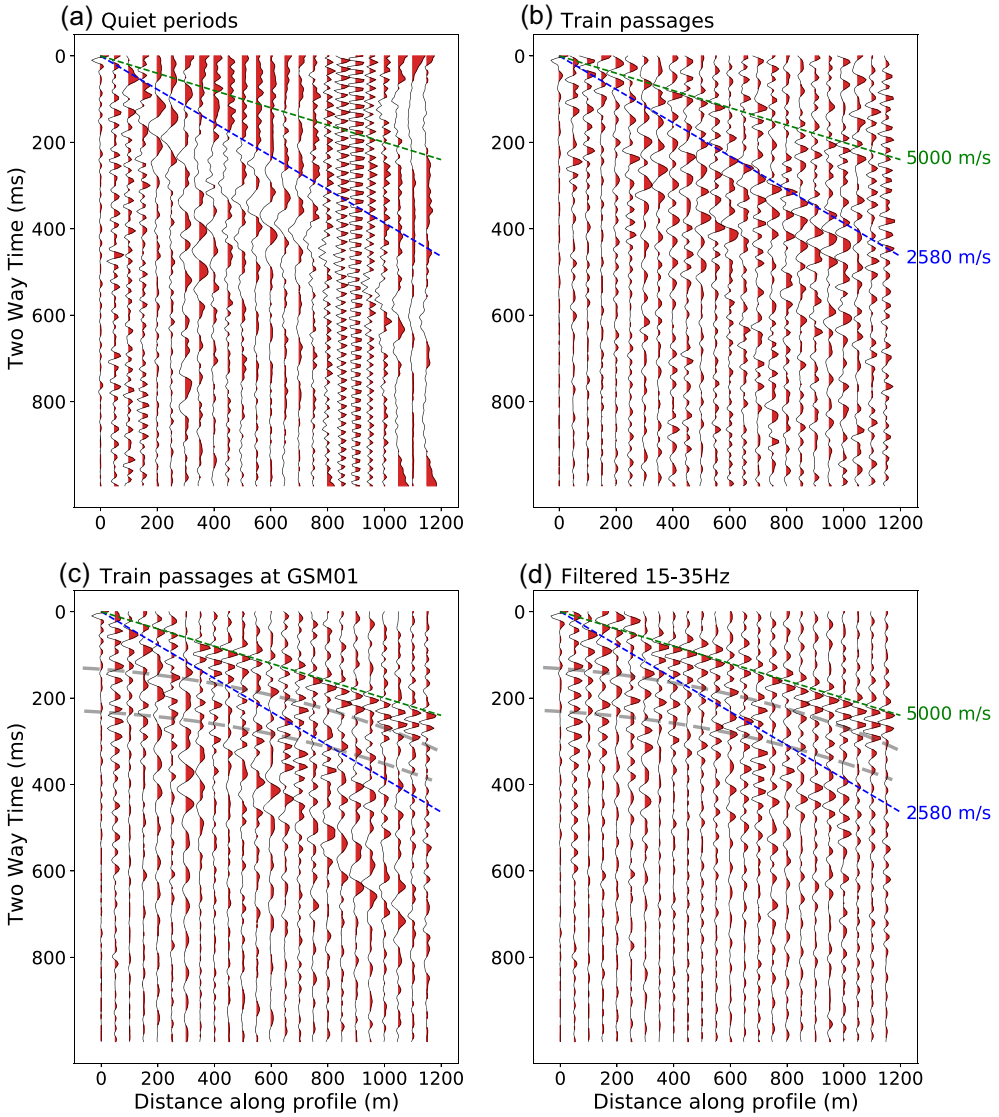

Figure 3. Virtual shot gathers (positive correlation lags) generated by cross-correlation of all the traces with receiver no. 1. Various pre-processing steps were applied (see the text for more details). Different panels illustrate the results of cross-correlation and stacking for (a) quiet period when there is no train passage, (b) the time windows when the train is passing through the array and (c) 4 s time windows around the time when the train is at the first station (GSM01) location. (d) Similar time windows as in (c) but with 15–35 Hz bandpass filtering to reduce the surface wave effect, while (a)–(c) are bandpass filtered between 2 and 35 Hz. The apparent velocity of the *P* and *S* waves are ∼4500–4900 and ∼2300–2600 m s^{-1}, respectively. Dashed lines show indicate the possible reflectors.

two stations failed (the farthest station from the railroad -GSM26- failed after 40 hr and one of the stations in the parallel line -GSM11- failed after 60 hr of data recording).

3 INSPECTION OF RECORDINGS

The 72-hr continuous data were divided into hourly sections and filtered to remove the non-zero mean value. While inspecting the raw records, trains were identified based on the temporal length of the vibrations at each station and the velocity of the signal as it traversed the array. The records contained 172 and 175 clear signals for trains travelling southbound and northbound, respectively, which represent a total of 347 events averaging 115.6 per day. Fig. 2(a) shows an example of an hour-long record that contains signals corresponding to different train types.

Analyzing the recorded signal, we were able to distinguish between different train types (e.g. freight, commuter and intercity trains) and different train lengths (e.g. 3–4, 5–6, 7–9 and 10–11 wagons). Figs 2(b)–(e) show examples of different train signals detected by investigating the raw trace (top row), together with their spectrograms (middle row) and their amplitude spectra (bottom row). These train signals are comparable to the ones obtained by Quiros *et al.* (2016) in a similar configuration. They are characterized by clear spectral lines in the spectrograms, related to train speed and geometry (Fuchs *et al.* 2018; Liu *et al.* 2021; Lavoué *et al.* 2021). The spectra shows that these train signals, that are recorded by the sensors deployed very close to the railway (∼5 m), have substantial energy over the entire recorded frequency range (from 2 to 80 Hz, these limits being imposed by our high-pass filtering and sampling rate, respectively), with sharp peaks associated to the spectral lines visible in the spectrograms. In comparison, in the train signals detected by Quiros *et al.* (2016) further away from the rail road (∼300 m), most of the energy lies below 50 Hz because high frequencies are attenuated with distance. Note that a Doppler

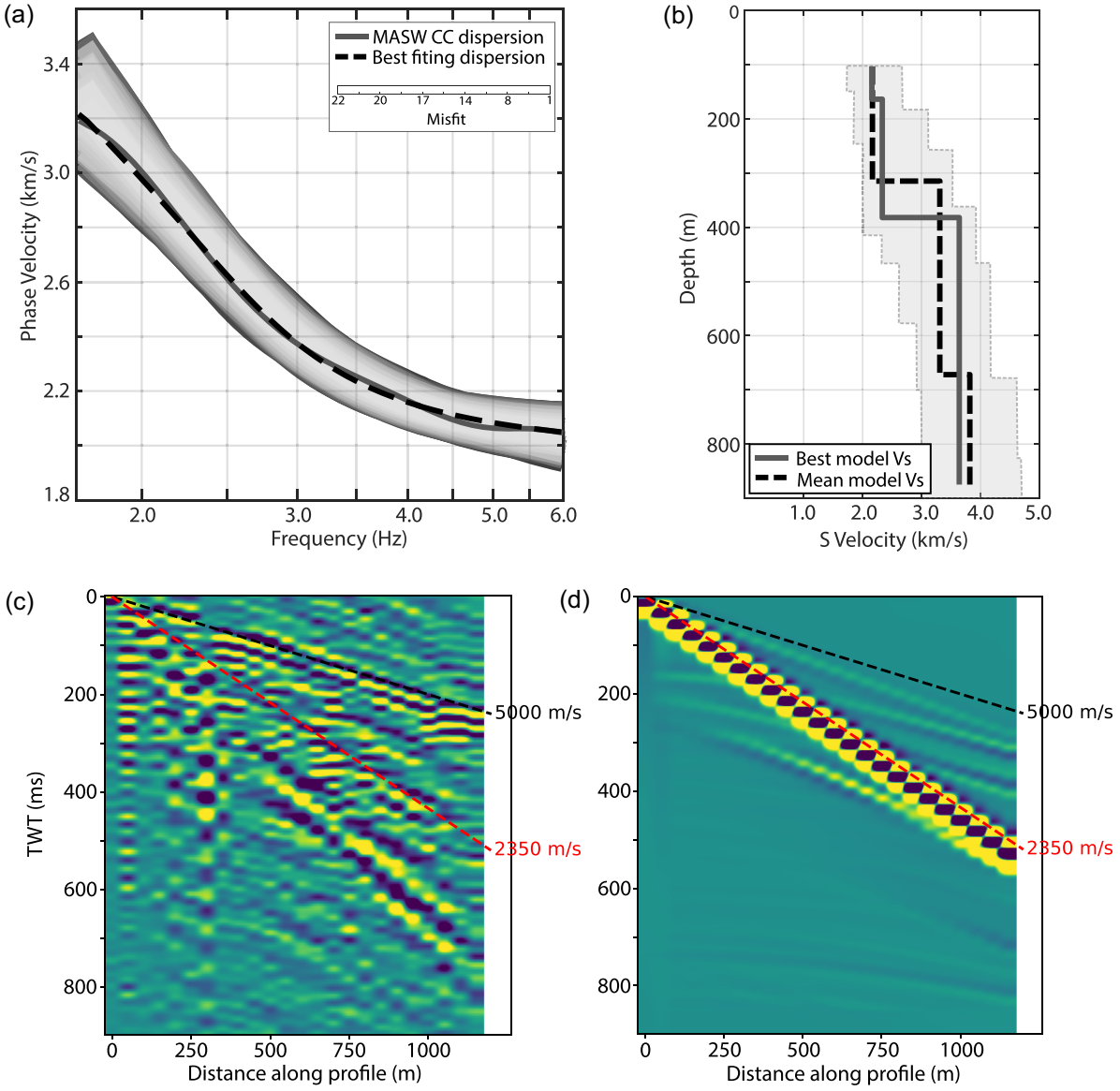

Figure 4. (a) Frequency-dependent phase velocity obtained for bandpass filtered (2–15 Hz) virtual-shot-gather data with source (e.g. train) at station GSM01. (b) 1-D *S*-wave velocity model obtained from the inversion in the frequency range 2–6 Hz of the dispersion curve. The black line is the estimated shear wave velocity. (c) Virtual shot gather from the real data for the time windows when the train is at the first station. (d) Simulated shot data using the velocity model obtained from MASW (black line in panel b) and a 20 Hz single force Ricker wavelet source at the location of the first station. Panels 'a' and 'b' are modified after Maggio (2022).

effect is visible in the spectrograms, with higher frequency content present at earlier times (i.e. approaching train) compared to later times (i.e. receding train). However, for some of the trains, this effect is reversed probably due to the train accelerating (Fig. 2).

4 SEISMIC INTERFEROMETRY

As the initial interferometric processing, in the time domain, we computed the positive lags of the cross-correlations for each station considered as a virtual source, what we here refer to as a virtual shot gathers, for different time periods (i.e. quiet periods vs. trains, cars, or other possible noise sources; Fig. 3). The procedure is to select one station (e.g. first station to the north -GSM01-) and cross-correlate its signal against all other stations, which result in 23 cross-correlations for the selected station. We tested several frequency bandpass filters and amplitude normalization routines for pre-processing the records, and applied spectral whitening and one-bit amplitude normalization (i.e. sign-bit) to the raw recordings prior to the correlation to recover stronger linear and quasi-linear arrivals (Bensen *et al.* 2007; Draganov *et al.* 2007, 2009; Quiros *et al.* 2016).

4.1 Virtual shot gather

The cross-correlation was computed as follow (Fig. S1, Supporting Information): (1) based on the expected subsurface velocities and array length we decided to use 2 s segments for the correlation, hence we split the hourly segments into in 2 s segments and extracted selected time window. (2) After detrending, tapering and amplitude normalization, we applied spectral whitening and then

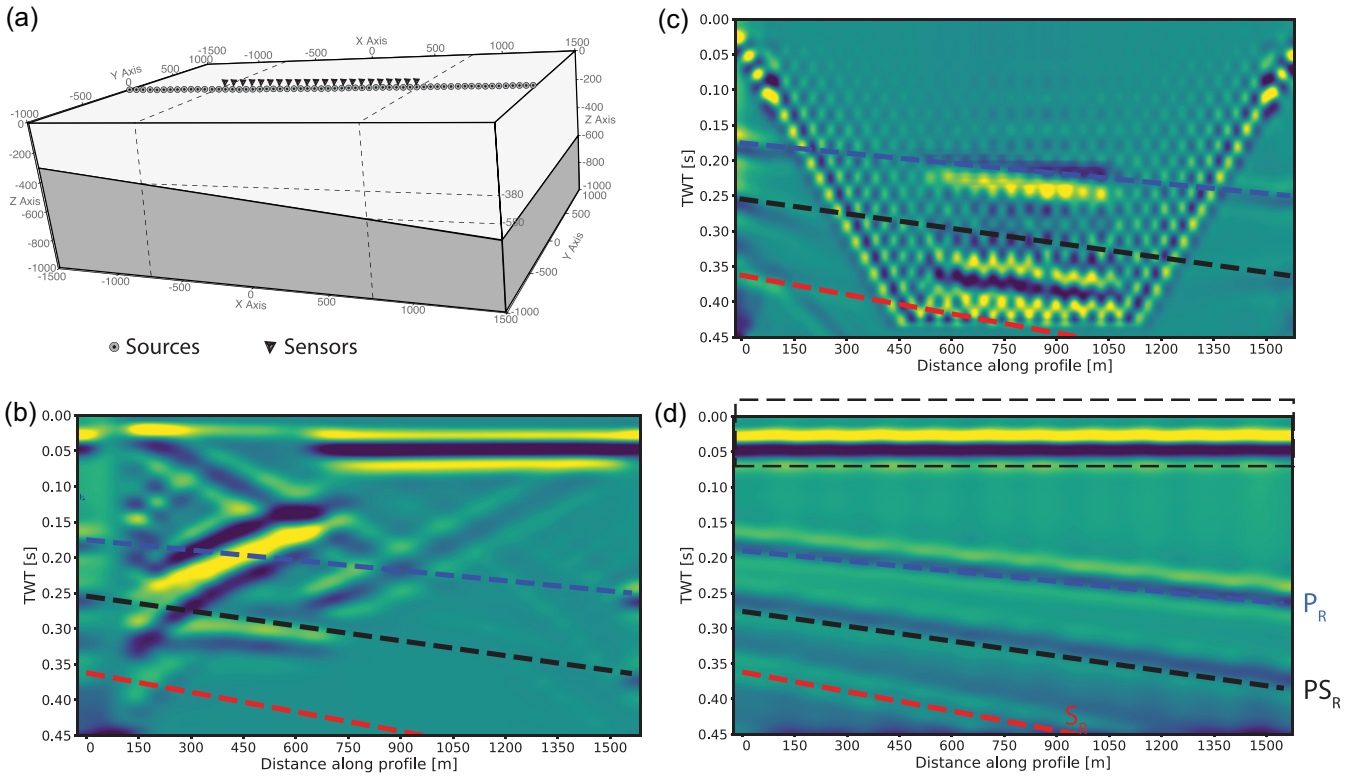

Figure 5. (a) 3-D dipping layered model and acquisition geometry used as inputs for the train signal simulation. Triangles show the representative location of 24 sensors with 50 m spacing and circles indicate the representative location of 190 sources with 15 m interval. (b) Resulting seismic section from the CMP-CC approach when the source is located at the beginning or at the end of the array, but not in between. (c) Results when CMP-CC is applied to all data. (d) Resulting seismic from the CMP-CC approach when the source is inside the array located at each station along the array. P_R, PS_R and S_R correspond to the TWTs of *P*-, converted *P*-to-*S* and *S*-wave reflections from the interface, respectively. The strong signal close to 0.0 time (masked with the rectangle) is a cross-correlation artefact and will be discussed in the interpretation section.

cross-correlate first station's trace against all other stations. (3) In the final step, we stack (linear stacking) all of these cross-correlated traces (e.g. virtual-shot-gather sections) from different time windows.

Fig. 3 shows typical virtual shot gathers (positive lags only) corresponding to different time windows. Each panel varies in terms of pre-processing steps. Fig. 3(a) was obtained by excluding train passages and keeping the remaining noise ('background' seismic energy), whereas in Fig. 3(b), we specifically selected time windows corresponding to train passages (e.g. approximately an average of 40 s time window for when each train is inside the array).

Finally, Figs 3(c) and (d) resulted from a narrower selection, a small time window (4 s) when the train locomotive (e.g. first carriage) is at the station considered as a virtual source. Fig. S2 (Supporting Information) illustrates an example of different selected time windows to derive virtual shot gathers presented in Fig. 3 and further to derive the final CMP section. The latter approach of selecting a narrower time window (Figs 3c and d), results in much clearer body waves signals compared to Figs 3(a) and (b). Additionally, bandpass filtering between 15 and 35 Hz further enables one to mitigate against surface wave arrivals, and to enhance the direct *P* wave with a velocity of $V_p \sim 4000$–5000 m s^{-1} in the virtual shot gathers (Fig. 3d).

In addition to what we interpret as clear direct *P* and surface waves on the virtual gathers (Figs 3c and d), there are other arrivals that exhibit curvature suggestive of a hyperbolic moveout. One of these hyperbolic events correspond to a reflector at approximately 0.18 s vertical two-way time (TWT) or ∼350 m below the surface, and a normal moveout (NMO) velocity of approximately ∼4500–4700 m s^{-1} (see the hyperbolic event marked with dashed line in Fig. S3, Supporting Information).

4.2 *S*-wave velocity from surface waves

Besides body waves, the virtual shot gathers in Fig. 3 clearly revealed that trains are also a good source of high-frequency surface waves, with a frequency content much higher (2–16 Hz) than what is usually retrieved from classical microseism interferometry (typically <4 Hz). This provides the opportunity for high-resolution surface wave imaging of the near subsurface (Quiros *et al.* 2016). With this in mind, we applied a routine cross-correlation interferometry technique to estimate a 1-D *S*-wave velocity model from the dispersion curves derived from the correlations (Maggio 2022).

Fig. 4(a) shows the frequency-dependent phase velocity (dispersion curve) estimated from the virtual shot gather corresponding to station GSM01 for a frequency range of 2–15 Hz. We picked the phase velocity of the fundamental mode (black line in Fig. 4a) and used these measurements to invert for the *S*-wave velocity as a function of depth using a Multi-channel Analysis of Surface Waves (MASW) algorithm (Park *et al.* 2007). The inverted *S*-wave interval-velocity distribution obtained from the phase velocity measurements is shown by the black line in Fig. 4(b).

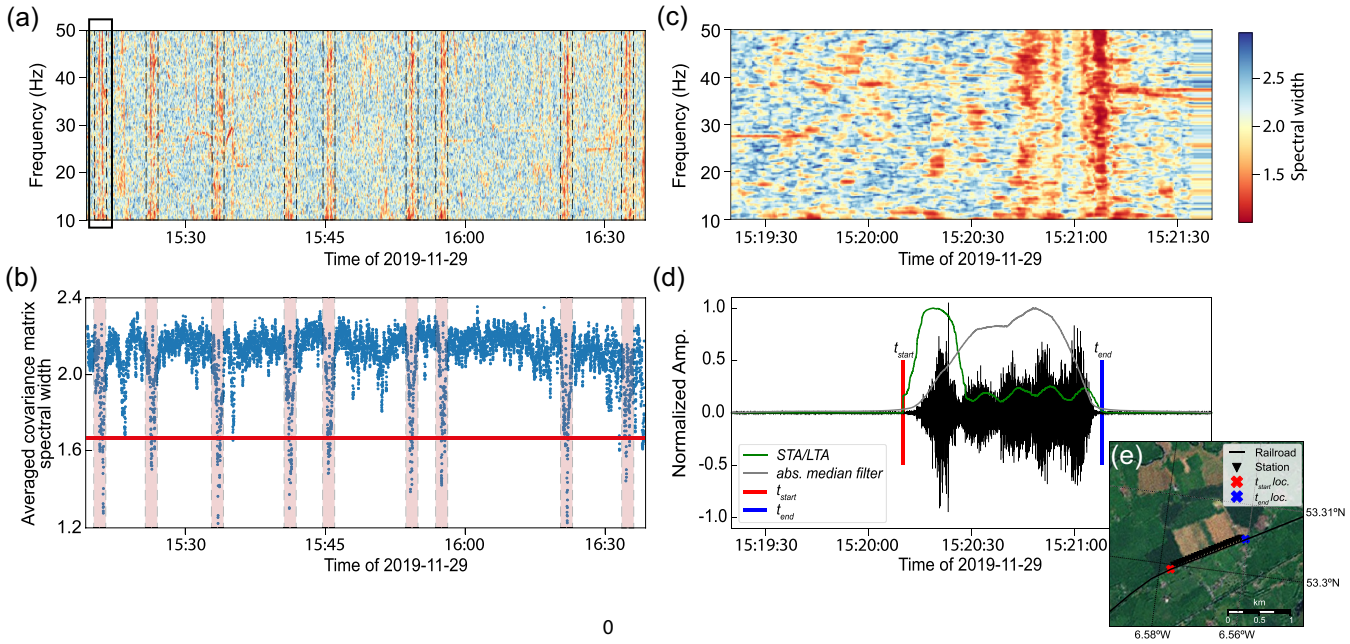

Figure 6. (a) Spectral width of the array covariance matrix as a function of time and frequency for 1-hr-long data filtered between 10–40 Hz for the first three stations (GSM01–GSM03). (b) Frequency average of the covariance matrix spectral width from panel (a) as a function of time. The red line is the detection threshold corresponding to a spectral width of 1.7 for train passages, and coloured segments indicate the detection periods. (c) Zoom-on the first 140 s of panel (a) (black rectangle in panel a) corresponding to a train passage. The colour scale indicates the spectral width (same scale for panels a and c). (d) Stacked signal of all the 24 stations for the same example train as in panel (c), with corresponding STA/LTA and median-filter values used for train signal detection. We picked t_{start} (red thick line) and t_{end} (blue thick line) as the beginning and end of the time window when the train is inside the array. (e) Aerial view of the deployment. The red and blue crosses denote the train locations extracted from GPS data for the picked t_{start} and t_{end} times, respectively.

The *S*-wave velocities indicated by the surface wave inversion qualitatively agrees with the *P*-wave velocity of ∼4500–4900 m s^{-1} derived from the virtual shot gather and the recorded borehole sonic log (assuming Poisson ration of $\sigma = 0.25$ and $V_p/V_s \approx 1.7$). Furthermore, our *S*-wave velocity model resolves several layers, with gradually changing velocity that could be correlated to the reflectors visible in the virtual shot gathers (Figs 3c and d).

Accordingly, we used this inverted *S*-wave interval-velocity distribution along with a Poisson's Ratio of 0.25 to calculate *P*-wave velocity and density for a simple 3-D model. We then simulated a shot using this 3-D model for a 20 Hz single force Ricker wavelet source at the location of the first station in an effort to help better understand the virtual shot gathers. Fig. 4(d) illustrates the resulting shot gather for the simulated data which is comparable to the virtual shot gather for the train data (Figs 4c, and 3c and d).

The comparison between the results in Figs 3 and 4 suggests that selecting specific, small enough, time windows (e.g. 4 s) when the train is around the station, produce the strongest and most useful body and surface wave energy that could be used for seismic interferometry. Simulated shot gather in a good agreement with the virtual shot gather (Figs 4c and d) reveals that most of the *P*-wave reflected signals from the shallow structures are limited to the top 0.4–0.5 s (TWT) and the deeper part is mostly contaminated with *S*- and surface wave energy. Additionally, comparison between Figs 3(c) and (d) document that applying a 15–35 Hz bandpass filter prior to computing the correlations suppresses the surface wave contribution and enhances body waves recovery. In practice, to be on the safe side, we used an even higher bandpass filter (17–40 Hz) for the reflection imaging.

5 REFLECTION IMAGING

Based on our virtual-shot-gather results, we used a different approach to that of Nakata *et al.* (2011) and Quiros *et al.* (2016) to select the data that goes into the cross-correlation step. Nakata *et al.* (2011) use all data available, while Quiros *et al.* (2016) explored using the approaching and receding trains prior to applying slightly different versions of seismic interferometry. Our approach, called CMP-CC for short (Common-Mid-Point Cross-Correlation), is to select the time windows when the train is located at each station, to generate CMP gathers from the cross-correlation of different station pairs and consequently generate a stacked reflection seismic profile and directly image subsurface structures. For this purpose, we applied the following steps: (1) for the 4 s time windows when the train is located at each station, we applied one-bit normalization, amplitude scaling, spectral whitening and bandpass filtering for all the traces. (2) Then we cross-correlate that station/trace with all the other stations/traces. (3) We sorted the pairs that target the same midpoint based on interstation distance and (4) made a virtual CMP gather that has an interval equal to half of the interstation distance. (5) We applied NMO correction to the virtual CMP gather, and (6) stacked all the traces from this virtual CMP gather to have a single trace representing this midpoint location. (7) Selecting different target midpoints and corresponding cross-correlated stations through the profile enables us to image the underlying structures. As discussed before, the main processing sequence includes amplitude scaling, NMO correction, and stacking. However, to optimize the technique, to find appropriate parameters, and the best pre- and post-processing steps, we first applied and tested the modified CMP-CC approach on synthetic signals simulated in a model mimicking the real situation.

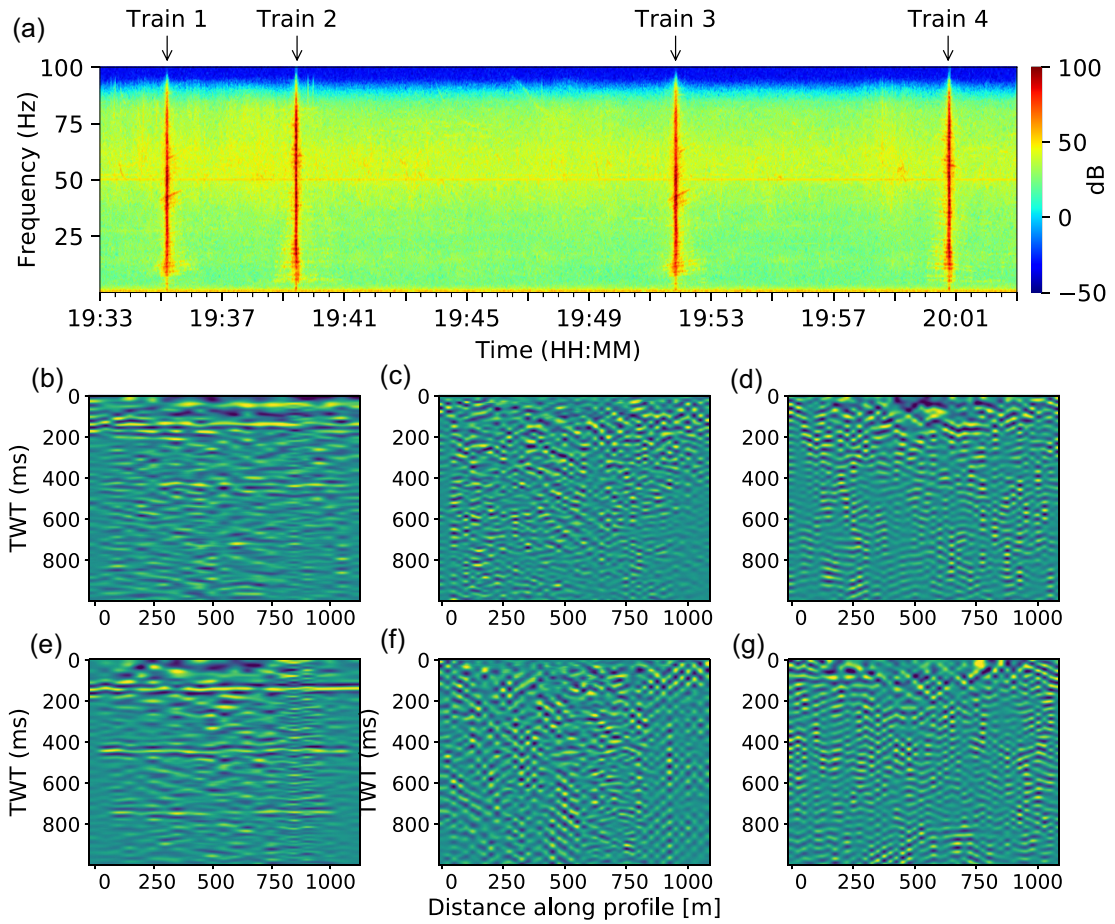

Figure 7. (a) Spectrogram of the raw seismogram at station GSM01 in the period 2019-11-29T19:33:00-20:03:00. The spectrogram is shown in decibel scale. Train signals at 19:35:12, 19:39:27, 19:51:50 and 20:00:45 show a wide frequency band spanning nearly the entire frequency range of 1–100 Hz. (b)–(d) CMP-CC sections for single Commuter train 1 passage for the time windows (b) when trains are approaching and receding the array, (c) when trains are inside and passing through the array and (d) when train is at each station. (e)–(g) CMP-CC sections for single Intercity train 2 passage for the time windows (e) when train is approaching and receding the array, (f) when train is inside and passing through the array and (g) when train is at each station, respectively.

5.1 Application of the CMP-CC approach on simulated data

We derived one simulation using a moving point source (25 Hz single force source, with a Ricker wavelet source time function, moving with 33 m s^{-1} speed), for simulating signals on 24 sensors with the same geometry as for real data. To simulate wave propagation, we located point sources on the surface with 33 m spacing and shoot them one after another every 1 s. Since we do not have the true 3-D model of the area, we use a simple two-layer model with a dipping interface and parameters (e.g. ρ, $V_p$and V_s) similar to those derived from the available geological information and borehole data. Although there might not exist a dipping interface in the real ground for this study area, we intentionally used a dipping layer for the simulation to be able to investigate any probable artefact. We then applied similar CMP-CC processing steps described in the previous paragraph with a bandpass filter of 17–40 Hz to different time windows (in this case, different source sets) of the simulated data. Additionally, in the final step we applied Automatic Gain Control (AGC) to the CMP section to reduce the effect of very shallow artefacts and also enhance weaker signals in the deeper part of the seismic section. Comparing the resulting CMP section with the input layered model, enabled us to find the most appropriate parameters and time windows for the cross-correlation and CMP-CC processing.

The comparison of the synthetic simulation results in Fig. 5 clearly demonstrates the constructive effect of selecting appropriate time windows for the CMP-CC approach. Comparison between Figs 5(b) and (c) suggest that when we apply CMP-CC approach on the whole time windows when the source is inside the array, rather than just using the time windows when the source is located either at the beginning or at the end of the array, we observe the signal corresponding to the *P*-wave energy reflected from the interface (Fig. 5c). However, as expected from shot-gather results (e.g. compare Figs 3b and d), using the appropriate smaller time windows when the source is located at each recording sensor, we observe even clearer reflection signals corresponding to the *P*-, *P*-to-*S* conversion and *S*-wave energy reflected from the interface (P_R, PS_R and a weaker signal marked as S_R, respectively; Fig. 5d).

Results associated with times (i) when the sources are only offend the linear, but not inside the array and (ii) when the sources are inside the array but not time localized to individual seismic stations exhibit strong artefacts that could be misinterpreted (Fig. 5d). Part of these artefacts are related to the source geometry (as we

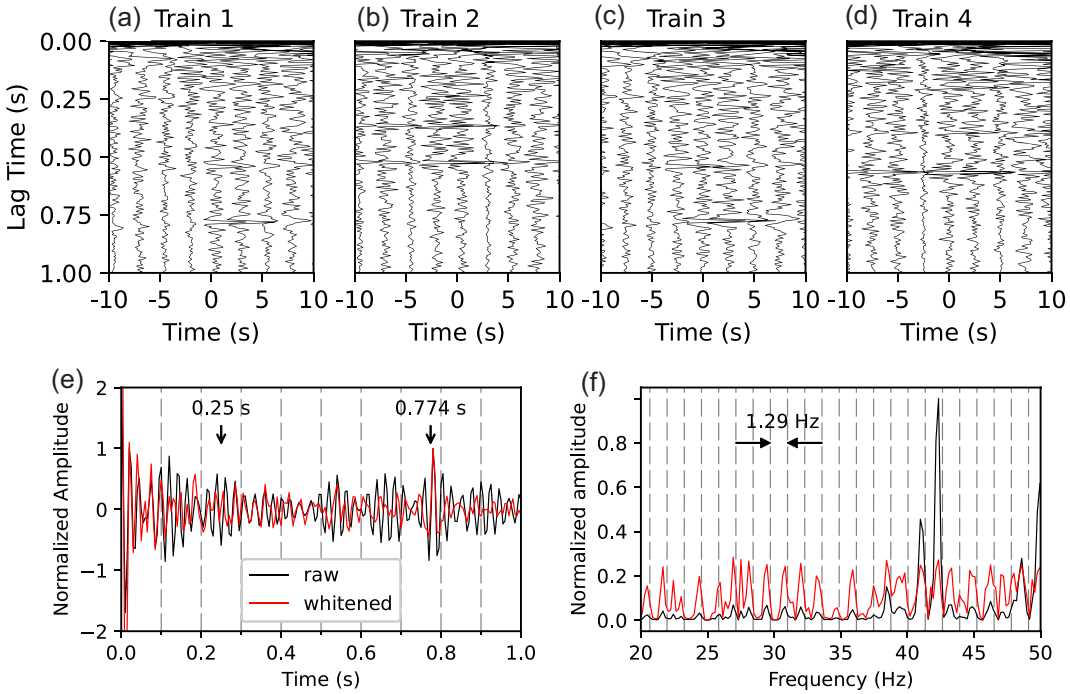

Figure 8. (a)–(d) Autocorrelograms of train signals at station GSM01 for different trains in Fig. 7(a). The autocorrelations are computed every 2.5 s with a 5-s window. Only the autocorrelograms in [−10, 10] s around the train arrivals are shown. Trains 1 and 3 are categorized as Commuter trains and trains 2, 4 are categorized as Intercity trains [see panels (c)–(e) in Fig. 2]. (e) Pre-stack autocorrelogram before (black) and after (red) spectral whitening for the first train at station GSM01. The strong signal at 0.774 s and very short-period oscillations at 0.25 s are marked. The two autocorrelograms are normalized by the amplitude of the signal. (f) Amplitude spectrum of the autocorrelograms for the first train at station GSM01. Vertical dashed lines denote the multiple frequencies of 1/0.774 Hz. Both are normalized by its peak amplitude, but the whitened one is scaled to 0.3 for illustrative purposes.

will see later for real data) and to potential cross-terms between phases in the correlations (e.g. cross-correlating the combination of direct, surface and reflected waves; see Fig. S4, Supporting Information).

5.2 Time window selection

Initial seismic interferometry and synthetic simulation results discussed in the previous sections suggested that we need to select a short time window when the train is located at each station position along the array to obtain a clearer image. In order to select the appropriate time windows corresponding to train passages, we first tested the covariance matrix method proposed by Seydoux *et al.* (2016) to locate the trains in time and space (i.e. as a detector). This method detects signals embedded in the noise in the time–frequency domain, using the width of the distribution of the eigenvalues of the array covariance matrix as an estimation of the level of spatial coherency of the wavefield. First, we filtered the data between 10–40 Hz to avoid other noise sources, and then we computed the covariance matrix for the first three stations in the north (see Fig. 1 for the location). Using more sensors causes difficulties for train detection because coherency decreases when the train passes through the array. Fig. 6(a) shows the detected trains within one hour of recorded data. As illustrated in Fig. 6(b), during a train passage the spectral width decreases (indicating an increase of the wavefield coherency), which can be used as a proxy for train detection. Accordingly, we averaged the spectral width in frequency to avoid false detection, and established an arbitrary detection threshold of 1.7 to detect the train signals (Fig. 6b). A train passage is detected when the spectral width is lower than 1.7 for a time window longer than 1 min, however the picked time windows are not accurate enough for our CMP-CC approach [compare panels (d) and (c) in Fig. 6].

To this end, we subsequently stacked the traces of all 24 stations over the detection periods identified by the covariance matrix method, then we derived a smooth time series by median filtering of the absolute amplitudes of the stacked signal, and finally we applied the short-time-average through long-time-average trigger (STA/LTA) on these filtered amplitudes to pick the exact start and end times of the window that we consider to correspond to a train passage over the array (Fig. 6d). As a validation of this procedure, Fig. 6(e) illustrates the actual train location (from its GPS track) for the start and end times picked for this example train passage, which correspond well to the beginning and to the end of the array. We further used these start and end times to calculate the speed and direction of the trains, and to infer the specific times when the trains are passing each station along the array.

5.3 CMP-CC section of a single train passage

To investigate more the effect of an appropriate time window selection in our interfrometry approach, using the extracted time windows from the previous section, we applied the CMP-CC approach on two different train categories [e.g. Commuter train versus InterCity train, trains 1 and 2 in Fig. 7, and panels (d) and (e), and (c) in Fig. 2, respectively]. Following the procedure described and tested on synthetics in the previous sections, we cut the traces in 4 s windows around the time when a given train is at a given station (e.g. 4 s

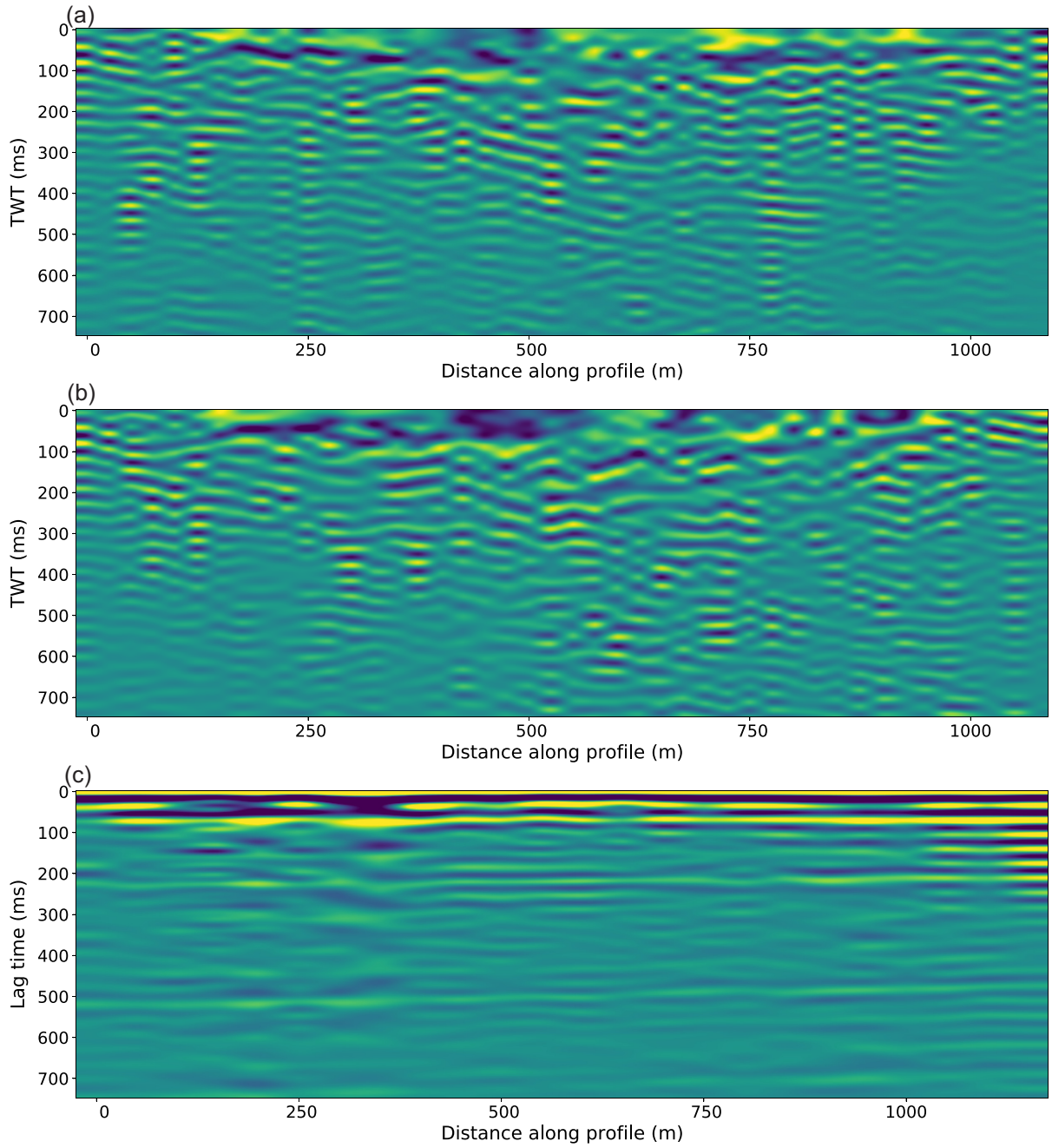

Figure 9. Resulting sections from the CMP-CC approach for (a) all the time windows when trains are inside the array but we do not considered any specific virtual source location (e.g. time windows between t_{start} and t_{end} in Fig. 6d) and (b) selective time windows when the trains are at each station (e.g. 4 s around t_{start} in Fig. 6d). (c) Resulting sections from the autocorrelation method applied on the selected time windows corresponding to train passages.

around t_{start} for cross-correlating GSM01 with all the other stations), split the window in 2 s segments with 50 per cent overlap, and for each segment we cross-correlated the considered station with all the others. Accordingly, we applied the CMP-CC processing steps for all the Mid-Points to generate a stacked seismic section. We applied a 17–40 Hz bandpass filter on the data (based on the frequency content of different train signals) and performed NMO correction using an averaged 1-D velocity model extracted from MASW (assuming Poisson ration of $\sigma = 0.25$ and $V_p/V_s \approx 1.7$; Fig. 4b) and from the sonic log acquired in the nearby borehole [NGE1, Fig. 1; see also Fig. S3a (Supporting Information) for the v_{NMO} profile]. We then applied AGC scaling to enhance S/N at depth, as practiced in active seismic.

Fig. 7 shows results for the InterCity train (b)–(d) and for the Commuter train (e)–(g), obtained by applying the CMP-CC approach either on time windows when the train is at the beginning or end of the array (b) and (e), or the whole time window (≈30–40 s) when the train is inside the array but we do not considered the train location in relation to the stations (c) and (f), and or a small time window (4 s) when the train is at a specific location along

the array (d) and (g). As expected from the synthetic simulation, results obtained by cross-correlating time windows when the virtual sources (e.g. trains) are inside the array (Figs 7c and f) and are located at each station (Figs 7d and g) show less artefacts compared to when the trains are located only at the beginning or end of the array, as proposed by Quiros *et al.* (2016, Figs 7b and e). The application of the CMP-CC approach on real data confirms that the artefacts are due to cross-terms in the synthetic test (as illustrated in Fig. S4, Supporting Information), but also show another effect due to the source time function of real trains. It should be noted that we could not see this effect in the synthetic test because we assumed a simple Ricker wavelet for this test, while actual train source time functions are much more complex (see Lavoué *et al.* 2021). The results obtained by correlating time windows when the trains are located only at the beginning or end of the array are expected to suffer from strong artefacts due both to cross-terms (as illustrated in Fig. S4, Supporting Information) and to the periodicity of the train source time functions. Note however that the CMP approach suppresses some of the high-frequency signals due to NMO correction and cross-correlation of the wide angle (far offset) stations. Furthermore, as illustrated in Fig. S3 (Supporting Information), after NMO correction, some of the reflected arrivals are not fully aligned, which could be due to the presence of a shallow weathering layer and/or to the use of an NMO correction where the higher-order terms are neglected (Yilmaz 2001; Quiros *et al.* 2016). These reasons could explain the lower resolution and the unrealistic anomalies at zero time due to the NMO stretching in the CMP sections in Figs 7(c) and (d), and (f) and (g). Consequently, to see clearer reflector signals in Figs 7(c) and (d), and (f) and (g), we need to stack the sections over several train passages.

5.4 Autocorrelation of train signals

In addition to the cross-correlation approach, we applied and tested autocorrelation on the train signals to investigate more the train signal interferometry on each individual station over time for imaging subsurface interfaces. The autocorrelation was computed as follows: we first selected appropriate time windows as discussed for the CMP-CC approach and split these time windows into 5-s-long segments with 50 per cent overlap. After detrending and tapering, each segment was transformed to the frequency domain by fast Fourier transform (FFT) and the Autocorrelation was computed through its power spectrum. We then whitened the power spectrum using three-point running absolute average normalization. The normalized power spectrum is then converted back to the time domain to obtain a pre-stack autocorrelogram.

We applied these processing steps to 20-s-long time windows when the four trains shown in Fig. 7(a) are at the first station (e.g. station GSM01) and calculated pre-stack autocorrelograms of that station for the four different trains (Fig. 8). A closer look at the pre-stack autocorrelated signals reveals a clear variation with respect to the trains/sources for the deeper reflectors (i.e. a dependency on the different trains considered as sources). Even though the four trains all have spike-like signals (time-domain spikes in Fig. S5, Supporting Information), the lag time of the arrivals varies in relation to train speed and geometry (Lavoué *et al.* 2021). For trains 1 and 3 (Commuter trains travelling North at ∼100–110 $\mathrm{km\,h^{-1}}$), the arrivals appear at 0.18, 0.52 and ∼0.774 s, while for trains 2 and 4 (Intercity trains travelling South at ∼160–170 $\mathrm{km\,h^{-1}}$), the lag times are 0.17, 0.45 and 0.52, respectively.

The temporal variations of the autocorrelations suggest that the characteristics of the source plays an important role in the autocorrelograms, which may affect the retrieval of the true reflection signals. Examination of the spectrum of the autocorrelograms (Fig. 8f, and Fig. S5, Supporting Information) reveals a series of regularly spaced frequency spikes inherited from the harmonicity of the raw train signals (Fig. 2), which itself naturally follows from the harmonicity of the train source time functions (Lavoué *et al.* 2021). In this particular case, the frequency spacing between spikes is 1.29 Hz, a fundamental frequency of the form $f_1 = v/L$ that is related to train speed v and wagon length L (Lavoué *et al.* 2021), and that is exactly the reciprocal of the latest lag time of the signal in the autocorrelogram (0.774 s, Fig. 8e, and Fig. S5, Supporting Information). This frequency spacing of 1.29 Hz is consistent with an Irish commuter train with 23.5-m-long carriages[2] travelling at ∼110 $\mathrm{km\,h^{-1}}$ (as estimated from our picked start and end times), which strongly suggests that this 'reflection signal' is more likely to be caused by the harmonicity of the train source time function rather than by a real subsurface reflector. Furthermore, our results in Fig. 8(e), and Fig. S5 (Supporting Information) illustrate some other very short-period oscillations in the time-domain autocorrelograms at $t \sim 0.25$ s that could be related to the frequency peak at $f \sim 42$ Hz, which would correspond to the *sleeper passage* frequency mentioned in Lavoué *et al.* (2021) ($f_2 = v/\Delta$, related to train speed v and sleeper spacing Δ). Autocorrelation results enable us to get more insights into this source-related effect and could also explain the horizontal artefacts in CMP-CC results at times of the form L/v: at ∼0.72 s for Commuter trains and ∼0.53 s for InterCity trains (Figs 7b and d).

6 FINAL STACKED RESULTS AND DISCUSSION

Based on the results obtained in the previous sections, we applied CMP-CC processing on the time windows when trains are at a specific virtual source along the array (Fig. 9b), and also when trains are inside the array but we do not considered any specific virtual source location (Fig. 9a). The resulting stacked CMP sections (Figs 9a and b) show coherent events down to 0.8 s (TWT). The most prominent events are the reflectors previously identified in Fig. 3 at about 0.18, 0.42 and ∼0.7 s. The time differences (Δt) suggest that these reflectors cannot be multiples and are related to subsurface structures.

Consequently, we applied autocorrelation processing to 20-s-long time windows when trains are at each station and stacked the autocorrelograms for all the detected trains to build an autocorrelation section (Fig. 9c). Similar to the CMP-CC sections, the autocorrelation result shows coherent events down to 0.8 s (lag time) where most prominent events are the reflectors identified at about 0.10–0.18, 0.40–0.50 and ∼0.7 s (Fig. 9c).

Although results of the previous sections document artefacts in the pre-stack autocorrelograms (Fig. 8) and single train CMP sections (Fig. 7), we do not see clear artefacts in the final stacked sections. Selecting the appropriate time window, stacking over different trains (i.e. different source time functions), and probably the multi-offset capability of the CMP-CC approach, constructively reduce the amplitude of artefacts in the final stacked sections (Fig. 9). However, special care needs to be taken when interpreting the coherent events in the final results.

[2]https://en.wikipedia.org/wiki/Coaching_stock_of_Ireland, last accessed: 2021 September,

Prominent reflector events from seismic interferometry methods could be associated with the contact between Palaeozoic sedimentary rocks (i.e. between different types of limestones) based on the interpretation of the nearby boreholes data. Compared to the available borehole data at NGE1 and NGE2 (Fig. 1c), the two shallower reflectors at ∼0.18 and 0.42 s could correspond to lithological boundaries (e.g. between Upper and Lower Calp and with the top of a sandstone layer found in the drilling records).

The 2-D reflection seismic section, together with Licciardi & Piana Agostinetti's (2017) interpretation of the main reflectors (Fig. 1d), also documents two reflectors at approximately 0.3 and 0.8 s (TWT) that correspond to the top of the Lower Calp Grainstone and to the Upper-to-Lower Calp transition, respectively (Licciardi & Piana Agostinetti 2017). These two reflectors are consistent with those in our CMP-CC and autocorrelation sections and could be correlated with the reflectors at 0.18 and 0.42 s (Fig. 9). However, possible lithology changes should be considered due to a ∼3 km distance between boreholes and the seismic profile location and the study area, as it is clear from the changes in the lithostratigraphies of the two boreholes within a small distance.

Although we observe some deeper reflectors in CMP-CC section, we do not interpret them because we know, based on our synthetic simulation tests, that the deeper part is presumably contaminated by *S*-wave related signals. Furthermore, the resolution is lower in the deeper part due to short array aperture and short recording time (Quiros *et al.* 2016). In principle, shallow reflectors correspond to illumination at relatively wide angles (Quiros *et al.* 2016). Ideally, to image deeper structures, we would need sources distributed beneath the recording array (Draganov *et al.* 2006; Quiros *et al.* 2016). However, since the train source is restricted to the surface and most of the energy propagates horizontally, increasing the aperture of the array and the length of the recording time period might allow the recovery of deeper structures (Quiros *et al.* 2016).

7 CONCLUSION

The signals generated by train traffic proved to be valuable for a range of near-surface applications of seismic interferometry. Useful body (mostly *P* wave) and surface wave energy can be recovered for imaging of the near-surface and upper-crustal structure. Although in this study the high-resolution part of the reflection image is limited to the top 1.5 km of the subsurface, the depth range of the imaging could be increased by recording and stacking for longer time periods and by using array with wider apertures.

The virtual shot gathers generated from cross-correlation interferometry applied on the railroad traffic exhibit robust dispersive surface wave (Rayleigh) energy at frequencies higher than those obtained from conventional interferometry. We used the surface wave data and inverted the dispersion measurements to obtain an *S*-wave velocity profile that is generally consistent with the body waves retrieved from the train-generated signals.

Our results document a successful application of the CMP-CC and autocorrelation methods on the retrieved body waves from the train-generated signals, but also highlight the need for a dedicated processing for using the 'noise' from railroads as an effective source for body-wave interferometry. If used without careful consideration, both the autocorrelation and cross-correlation methods can result in non-physical, spurious reflectors caused first by cross-terms in the correlations and second by the periodicity of the train source time functions. Applying the CMP-CC approach on specific time windows (when the trains are at specific virtual sources) and using different train types enables us to reduce the artefacts. The CMP-CC approach is more efficient than the autocorrelation method for suppressing artefacts, likely because it benefits from multi-offset illumination and stacking. In the cross-correlation approach, artefacts due to cross-terms can be removed by carefully selecting the signal to be correlated: to obtain the shot gather for a given train and a given virtual source, we recommend to correlate only the short time window that corresponds to the moment when the train is located at this virtual source, thereby mimicking the configuration of an active seismic experiment (i.e. turning trains into seismic vibrators). On the other hand, artefacts related to the properties of the train source time functions can easily be identified if the characteristics (geometry and speed) of the trains are known, and can be mitigated by stacking the shot gathers obtained for several train passages, as long as trains with different geometries and/or speeds travel over the railway. Following this methodology, both the CMP-CC and the autocorrelation methods image consistent subsurface structures that are consistent with the prior geological and geophysical information. Train-generated vibrations are thus a useful source of signal for body wave interferometry for imaging the subsurface in the immediate vicinity of a railway. Further investigations are required in order to extend the methodology for cases where the sensors and area of interest are located further away from the railroad.

ACKNOWLEDGMENTS

This research emanates from PACIFIC—Passive seismic techniques for environmentally friendly and cost-effective mineral exploration—which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 776622. We also acknowledge support from the European Research Council under grant no. 817803, FAULTSCAN. Seismic instruments were provided by the Dublin Institute for Advanced Studies (DIAS) and authors wish to thank all DIAS staff for their invaluable work, technical support and helping us with the deployment. We are very grateful to Michal Malinowski, the editor, and Fern Storey, the assistant editor, for their critical assessments during the supportive review process that helped to improve the manuscript. We also thank two anonymous reviewers whose thoughtful comments and recommendations substantially improved the manuscript. Obspy package (Krischer *et al.* 2015), Python (Van Rossum & Drake 1995) and Generic Mapping Tools (Wessel *et al.* 2013) were used to read and process the data and to create the presented figures.

DATA AVAILABILITY

The data can be accessed upon request before 2023 January from GFZ pool (Rezaeifar & Bean 2022) and would be publicly available afterwards.

REFERENCES

Artman, B., 2006. Imaging passive seismic data, *Geophysics,* **71**(4), SI177–SI187.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *J. geophys. Int.,* **169**(3), 1239–1260.

Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N.M., Nadeau, R.M. & Larose, E., 2008. Postseismic relaxation along the san andreas

fault at parkfield from continuous seismological observations, *Science,* **321**(5895), 1478–1481.
Brenguier, F. *et al.*, 2019. Train traffic as a powerful noise source for monitoring active faults with seismic interferometry, *Geophys. Res. Lett.,* **46**(16), 9529–9536.
Chamarczuk, M., Malinowski, M., Draganov, D., Grant, A., Asgharzadeh, M. & Urosevic, M., 2021. Characterization of drilling-related noise and curvelet-based evaluation of seismic-interferometric reflections for imaging of iron-bearing formations in pilbara, western australia, *J. geophys. Int.,* **226**(1), 377–404.
Curtis, A., Gerstoft, P., Sato, H., Snieder, R. & Wapenaar, K., 2006. Seismic interferometry-turning noise into signal, *Leading Edge,* **25**(9), 1082–1092.
Dales, P. *et al.*, 2020. Virtual sources of body waves from noise correlations in a mineral exploration context, *Seismol. Res. Lett.,* **91**(4), 2278–2286.
Draganov, D., Wapenaar, K. & Thorbecke, J., 2006. Seismic interferometry: reconstructing the earth's reflection response, *Geophysics,* **71**(4), SI61–SI70.
Draganov, D., Wapenaar, K., Mulder, W., Singer, J. & Verdel, A., 2007. Retrieval of reflections from seismic background-noise measurements, *Geophys. Res. Lett.,* **34**(4) , doi:10.1029/2006GL028735.
Draganov, D., Campman, X., Thorbecke, J., Verdel, A. & Wapenaar, K., 2009. Reflection images from ambient seismic noise, *Geophysics,* **74**(5), A63–A67.
Fichtner, A., 2015. Source-structure trade-offs in ambient noise correlations, *J. geophys. Int.,* **202**(1), 678–694.
Fuchs, F. & Bokelmann, G., the AlpArray Working Group, 2018. Equidistant spectral lines in train vibrations, *Seismol. Res. Lett.,* **89**(1), 56–66.
Hariri Naghadeh, D., Bean, C.J., Brenguier, F. & Smith, P.J., 2021. Retrieving reflection arrivals from passive seismic data using radon correlation, *J. geophys. Eng.,* **18**(2), 177–191.
Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. Obspy: a bridge for seismology into the scientific python ecosystem, *Comput. Sci. Discov.,* **8**(1), 014003. doi:10.1088/1749-4699/8/1/014003.
Lavoué, F., Coutant, O., Boué, P., Pinzon-Rincon, L., Brenguier, F., Brossier, R., Dales, P., Rezaeifar, M. & Bean, C.J., 2021. Understanding seismic waves generated by train traffic via modeling: implications for seismic imaging and monitoring, *Seismol. Soc. Am.,* **92**(1), 287–300.
Licciardi, A. & Piana Agostinetti, N., 2017. Sedimentary basin exploration with receiver functions: seismic structure and anisotropy of the dublin basin (ireland), *Geophysics,* **82**(4), KS41–KS55.
Liu, Y., Yue, Y., Luo, Y. & Li, Y., 2021. Effects of high-speed train traffic characteristics on seismic interferometry, *J. geophys. Int.,* **227**(1), 16–32.
Maggio, G., 2022. *Multi-environment seismic ambient noise imagery in Ireland,* Ph.D. thesis, School of Earth Science, University College Dublin.
Malehmir, A., Durrheim, R., Bellefleur, G., Urosevic, M., Juhlin, C., White, D.J., Milkereit, B. & Campbell, G., 2012. Seismic methods in mineral exploration and mine planning: A general overview of past and present case histories and a look into the future, *Geophysics,* **77**(5), WC173–WC190.
Nakata, N., Snieder, R., Tsuji, T., Larner, K. & Matsuoka, T., 2011. Shear wave imaging from traffic noise using seismic interferometry by cross-coherence, *Geophysics,* **76**(6), SA97–SA106.
Nakata, N., Chang, J.P., Lawrence, J.F. & Boué, P., 2015. Body wave extraction and tomography at long beach, california, with ambient-noise interferometry, *J. geophys. Res.: Solid Earth,* **120**(2), 1159–1173.
Panea, I., Draganov, D., Almagro Vidal, C. & Mocanu, V., 2014. Retrieval of reflections from ambient noise recorded in the mizil area, romania, *Geophysics,* **79**(3), Q31–Q42.
Park, C.B., Miller, R.D., Xia, J. & Ivanov, J., 2007. Multichannel analysis of surface waves (masw)—active and passive methods, *Leading Edge,* **26**(1), 60–64.
Pinzon-Rincon, L. *et al.*, 2021. Humming trains in seismology: an opportune source for probing the shallow crust, *Seismol. Soc. Am.,* **92**(2A), 623–635.
Quiros, D.A., Brown, L.D. & Kim, D., 2016. Seismic interferometry of railroad induced ground motions: body and surface wave imaging, *Geophys. Suppl. Mon. Not. R. Astron. Soc.,* **205**(1), 301–313.
Rezaeifar, M. & Bean, C.J., 2022. *Dublin Pilot Test Train Experiment, Phase 1, GFZ Data Services. Other/Seismic Network*. https://doi.org/10.14470/8D7564413901.
Roux, P., 2009. Passive seismic imaging with directive ambient noise: application to surface waves and the san andreas fault in Parkfield, CA, *J. geophys. Int.,* **179**(1), 367–373.
Roux, P., Sabra, K.G., Gerstoft, P., Kuperman, W. & Fehler, M.C., 2005. P-waves from cross-correlation of seismic noise, *Geophys. Res. Lett.,* **32**(19) , doi:10.1029/2005GL023803.
Salisbury, M., Snyder, D. & Goodfellow, W.D., 2007. Application of seismic methods to mineral exploration, *Mineral deposits of Canada: A synthesis of major deposit types, district metallogeny, the evolution of geological provinces, and exploration methods: Geological Association of Canada, Mineral Deposits Division, Special Publication,* **5,** 971–982.
Seydoux, L., Shapiro, N.M., de Rosny, J., Brenguier, F. & Landès, M., 2016. Detecting seismic activity with a covariance matrix analysis of data recorded on seismic arrays, *J. geophys. Int.,* **204**(3), 1430–1442.
Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* **307**(5715), 1615–1618.
Snieder, R., 2004. Extracting the green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E,* **69**(4), 046610 , doi:10.1103/PhysRevE.69.046610.
Van Rossum, G. & Drake, F.L. Jr, 1995. *Python Tutorial, Centrum voor Wiskunde en Informatica Amsterdam,* The Netherlands.
Wessel, P., Smith, W.H., Scharroo, R., Luis, J. & Wobbe, F., 2013. Generic mapping tools: improved version released, *Eos, Trans. Am. geophys. Un.,* **94**(45), 409–410.
Xu, Y., Lebedev, S., Meier, T., Bonadio, R. & Bean, C., 2021. Optimized workflows for high-frequency seismic interferometry using dense arrays, *J. geophys. Int.,* **227**(2), 875–897.
Yilmaz, Ö., 2001. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data,* Society of Exploration Geophysicists.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Schematic representation of the data processing steps for CMP-CC applied on synthetic and real recorded data.

Figure S2. Example of two minutes long records normalized by the max amplitude. Red rectangles illustrate different selected time windows for cross-correlation for (a) quiet period when there is no train passage, (b) the time windows when the train is passing through the array and (c) 4-s time windows around the time when the train is at the first station (GSM01) location.

Figure S3. (a) 1-D velocity model extracted from borehole sonic log for NMO correction. (b) Virtual shot gather (positive lags) for station GSM01 with no NMO correction and (c) corrected gather using the V_{nmo} shown in (a). Arrow and dashed lines indicate the hyperbolic event at ∼0.18 s.

Figure S4. (a) Simple model with a single flat reflector located at 600 m depth used for a simple simulation to formulate the relation between seismic interferometry and extracted body waves. (b) Recorded signal for sensor-A located at 50 m distance to source and (c) recorded signal for sensor-B located at 150 m distance to source. (d) Output of cross-correlation of the two signals. R_P and R_S show corresponding times for *P* and *S* waves reflected from the reflector. The amplitudes are exaggerated.

Figure S5. Each rows show suspicious reflection signal in the autocorrelograms of different trains in Fig. 9. Left: pre-stack autocorrelogram before (black) and after (red) spectral whitening. The strong signal related to *f*1 and very short-period oscillations related to *f*2 are marked (please refer to the text for more details) for different trains based on their speed extracted from the timetable. The two auto-

correlograms are normalized by the amplitude of the signal. Right: amplitude spectrum of the autocorrelograms. Vertical dashed lines denote the multiple frequencies of $f1$. Both are normalized by its peak amplitude, but the whitened one is scaled to 0.3 for illustrative purposes.