Understanding seismic waves generated by train traffic via modelling: implications for seismic imaging and monitoring

François Lavoué[1,*], Olivier Coutant[1], Pierre Boué[1], Laura Pinzon-Rincon[1], Florent Brenguier[1], Romain Brossier[1], Philippe Dales[1], Meysam Rezaeifar[2], and Christopher J. Bean[2]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, 38000 Grenoble, France

[2]Dublin Institute for Advanced Studies, School of Cosmic Physics, Geophysics Section, 5 Merrion square, Dublin, Ireland

[*]francois.lavoue@univ-grenoble-alpes.fr

(Accepted version for publication in the *Seismological Research Letters*)

Contents

Abstract **3**

Introduction **4**

Modelling strategy **5**
General approach and workflow 5
Computation of Green functions 6
Train source time functions 7
Characteristic lengths and expected frequencies at play 7
Effect of train geometry on frequency content 8
Effect of the distribution of axle loads over sleepers on frequency content 9

Modelling results **11**
Two end-member cases: single stationary source vs. single moving load 11
Realistic cases (all sleepers, all wheels): effect of sleeper regularity 12

Discussion **13**
Limitations of the modelling strategy 13
Parameters controlling signal amplitudes: implications for seismic detection . . . 14
Variable train speed 16
Wave propagation effects 16
Short vs. long distance signals 16
Effect of seismic properties (velocities and attenuation) 16
Wave partitioning (body vs. surface waves) 17

Conclusions **17**

Data and resources **19**

Acknowledgments **19**

Bibliography **20**

Tables **23**

Figures **24**

Abstract

Trains are now recognized as powerful sources for seismic interferometry based on noise correlation, but the optimal use of these signals still requires a better understanding of their source mechanisms. Here we present a simple approach for modelling train-generated signals inspired by early work in the engineering community, assuming that seismic waves are emitted by sleepers regularly spaced along the railway and excited by passing train wheels. Our modelling reproduces well seismological observations of tremor-like emergent signals and of their harmonic spectra. We illustrate how these spectra are modulated by wheel spacing, and how their high-frequency content is controlled by the distribution of axle loads over the rail, which mainly depends on ground stiffness beneath the railway. This is summarized as a simple rule of thumb that predicts the frequency bands where most of train-radiated energy is expected, as a function of train speed and of axle distance within bogies. Furthermore, we identify two end-member mechanisms –single stationary source vs. single moving load– that explain two types of documented observations, characterized by different spectral signatures related to train speed and either wagon length or sleeper spacing. In view of using train-generated signals for seismic applications, an important conclusion is that the frequency content of the signals is dominated by high-frequency harmonics and not by fundamental modes of vibrations. Consequently, most train traffic worldwide is expected to generate signals with a significant high-frequency content, in particular in the case of trains travelling at variable speeds which produce truly broadband signals. By proposing a framework for predicting train-generated seismic wavefields over meters to kilometers distance from railways, this work paves the way for high-resolution passive seismic imaging and monitoring at different scales with applications to near-surface surveys (aquifers, civil engineering), natural resources exploration, and natural hazard studies (landslides, earthquakes and volcanoes).

Introduction

Trains have recently been recognized as powerful seismic noise sources for imaging (e.g. Nakata et al., 2011; Quiros et al., 2016; Dales et al., 2020) and monitoring (Brenguier et al., 2019) the Earth's subsurface, from tens of meters to kilometers depth. By correlating train-generated noise between sensors, Nakata et al. (2011), Quiros et al. (2016) and Dales et al. (2020) were able to recover body waves that can be processed to reveal subsurface structures, in a similar fashion as from active seismic data. Brenguier et al. (2019) showed that train-generated body waves can be used to illuminate fault zones at a few kilometers depth, thereby allowing *in situ* monitoring of real faults in the field in similar conditions as those achieved in laboratory experiments (e.g. Scuderi et al., 2016). However, the appropriate use of train-generated seismic noise for extracting information about the subsurface still requires a better understanding of their source mechanisms and of their spectral characteristics in the frequency band [1 - 50] Hz of interest for seismological applications targeting the shallow crust.

While there exists detailed modelling studies in the engineering community of how trains generate ground vibrations in the vicinity of the railway (up to $\simeq$ 100 m), including all the complexity of the interactions between trains, rail tracks and ballast (e.g. Kouroussis et al., 2011 or Li et al., 2018b, see also Connolly et al., 2015 for a review), there is a relative lack of modelling for understanding distant seismic observations (Fuchs et al., 2018; Li et al., 2018a; Inbal et al., 2018; Brenguier et al., 2019).

Fuchs et al. (2018) provided many clues for interpreting the signals recorded in a range of distances from the railway (300 to 1500 m). A striking observation is the presence of equidistant spectral lines in the spectrograms, which they explained as being caused by the repeated excitation of stationary sources (which they did not precisely identify) by the successive wheels of the train. They proposed that the fundamental frequency of the resulting harmonic signal may be expressed as $f_1 = V_{train}/\Delta_b$ and controlled by the train speed V_{train} and the bogie distance Δ_b. Their explanations, however, are mainly based on observations and on *rules of thumb* derived from short-distance modelling (Connolly et al., 2015; Degrande and Schillemans, 2001; Ditzel et al., 2001; Wu and Thompson, 2001; Sheng et al., 2003). Moreover, while these propositions seem to well explain most of the observations, they are not entirely satisfactory for explaining the larger spacing between frequency peaks or the Doppler effects that are sometimes observed (e.g. Fuchs et al., 2018, figs 6a,b and S2b). In this study, we further investigate these hypotheses via the simulation of seismic wave propagation.

The interest of this paper is threefold. First, we present an effective strategy for modelling train-generated signals. Without entering into the complexity of train-track interactions, our simple modelling strategy well reproduces the main characteristics of the signals recorded by seismic stations located at moderate distances from the railway (a few hundreds of meters to a few kilometers). Second, we clarify the mechanisms at play in the generation of seismic waves by trains. Although they are well known and

documented in the engineering community (e.g. Krylov and Ferguson, 1994), we feel a need to illustrate these mechanisms in relation to seismological observations. Third, we identify the main parameters that control these mechanisms, and therefore the amplitude and frequency content of the observed signals, and we highlight perspectives for seismic applications.

The outline of the paper is the following: we start by presenting our modelling strategy, based on 3D visco-elastic simulations and on Krylov's model of the quasi-static excitation generated by trains (Krylov and Ferguson, 1994). We detail the design of the source time functions representing the train load applied on the ground and how we expect these functions and their spectral characteristics to control the final results. We then focus on two end-member cases: single moving load vs. single stationary source, and show that they generate spectral signatures that are very similar to the two categories of observations made by Fuchs et al. (2018). Then we show how to reproduce signals similar to the observations, both in the time and frequency domain, when considering realistic cases of entire trains exciting many sources along the railway. We conclude by discussing the parameters that control the amplitudes and frequency content of train-generated signals, and we derive guidelines for detecting these signals in seismic ambient noise.

Modelling strategy

General approach and workflow

Following Krylov and Ferguson (1994) and Li et al. (2018b), we assume that the train load is applied onto the ground surface as point-source, stationary, quasi-static vertical forces acting on the sleepers, located every $\Delta_{sleeper}$ along the rail way, and activated by the passage of the train wheels (Fig. 1). Before going further, we shall specify that we use the term *stationary* in the same sense of *spatially stationary* as Fuchs et al. (2018), meaning that the sleepers are located at fixed positions and therefore constitute stationary sources. Besides, we describe the excitation as *quasi-static* because we only consider the effect of quasi-static pressure of wheel axles onto the track and ground (Krylov and Ferguson, 1994), and disregard dynamic excitations between wheels and track due for instance to wheel or rail irregularities (e.g. Sheng et al., 2003; Lombaert and Degrande, 2009; Galvín et al., 2010a,b; Xia et al., 2010). Finally, the sleepers are small compared to the propagating wavelengths and are assumed to be rigid. Therefore we consider only one series of sources, indicated as a single rail in Fig. 1(b), which amounts to assume that the load of each train axle is applied as a single point source (i.e. the two wheels of each axle act simultaneously on each sleeper). Using source-receiver reciprocity, we simulate the signal that would be generated by a train passing over the sleepers and recorded by a sensor away from the railway as the sum of the signals resulting from a virtual source located at the sensor position and recorded at the sleepers positions (Fig. 1b).

Our workflow is therefore the following:

1. Simulation of wave propagation between one (virtual) source and all (virtual) receivers, using a Dirac delta source time function such as to convolve the resulting impulse responses with any desired source time function at the post-processing stage.

2. Lowpass filter (using a zerophase Butterworth filter) below the maximum frequency allowed by the simulation ($f_{max}^{simulation} = 100$ Hz).

3. Resample at 250 Hz ($> 2f_{max}^{simulation}$ in order to satisfy Nyquist criterion). This is just to facilitate post-processing, given that the original simulation time step is usually of the order of 0.2 ms (5000 Hz) to satisfy the Courant-Friedrichs-Lewy (CFL) condition and ensure the stability of the numerical scheme.

4. Take time derivative to convert the simulated displacement (in m) into ground velocity (in m/s), which is the quantity that is usually recorded by seismic stations. Note that this time derivative has an effect on the high-frequency content of the signals.

5. (optional) Select or mute specific arrivals (e.g. direct P or S, refracted P or S, surface waves).

6. Convolve individual impulse responses with source time functions representing the train passage over each sleeper.

7. Sum all the individual convolved seismograms to get the final seismogram resulting from the contributions of all sleepers.

Steps 6 and 7 can be formalized as

$$v_z(x_r, t) = \sum_{j=1}^{N_S} S_j(t) * G_{zz}(x_r, x_j, t), \tag{1}$$

where $v_z(x_r, t)$ is the vertical component of the ground velocity recorded at receiver location x_r along time t, N_S is the number of sources (sleepers) along the railway, and $*$ denotes time convolution between the source time function $S_j(t)$ of source j and the Green function $G_{zz}(x_r, x_j, t)$ of the medium between the location x_j of the vertical source j and the receiver location x_r.

In the following, we give more details on some stages of this workflow, in particular on the computation of the individual impulse responses (Green functions G_{zz}) and on the design of the source time functions $S_j(t)$ representing the train passage over each sleeper.

Computation of Green functions

The method used to compute the individual impulse responses before convolution and stacking is not critical to our argument and could be performed via a number of different techniques, either (semi)analytical or numerical. In this work we compute these individual

impulse responses using the SEM46 software (Trinh et al., 2019) based on time-domain spectral elements under the visco-elastic approximation and for a maximum frequency of 100 Hz, so as to investigate the generation of high frequencies.

The use of spectral-element simulations presents numerous advantages. For instance, spectral elements enable an accurate localisation of the very densely-spaced virtual receivers and a consistent extraction of the simulated wavefield at these locations, on the same basis of Gauss-Lebatto-Legendre (GLL) polynoms as used in the modelling scheme. Second, the use of source-receiver reciprocity with a very high number ($\simeq$ 20,000) of virtual receivers (actual sources) partly mitigates the computational cost of this fully-numerical method. Needless to say, we only need a short simulation duration (4 s, i.e. the time for waves to travel from the source to the furthest receiver), and we do not simulate the entire duration of the train passage ($\simeq$ 540 s in the examples shown hereafter). Finally, spectral elements enable us to consider media with realistic physical properties (including attenuation) and with arbitrary heterogeneity. In this paper, we restrict ourselves to examples in a homogeneous medium, because we want to focus on the source and it turns out that the main mechanisms can be easily exemplified in such simple models (see Table 1 for the properties considered in our numerical experiments). Further work may investigate the effects of medium heterogeneity in real-scale cases, e.g. in southern California (Brenguier et al., 2019) or in our test sites in Ontario (Canada, Dales et al., 2020) and in the Dublin basin (Ireland, Rezaeifar et al., 2020).

Train source time functions

Characteristic lengths and expected frequencies at play

According to the considered mechanism consisting of sleepers hit by wheels, we may expect two fundamental frequencies, related

1. to the excitation of each source (sleeper) by the successive wheel pairs, spaced every Δ, resulting in a frequency $f_1 = V_{train}/\Delta$, where the spacing Δ may *a priori* be
   - the distance Δ_a between axles within a given bogie,
   - the distance Δ_{b_1} between two bogies of the same wagon (as suggested by Fuchs et al., 2018),
   - the distance Δ_{b_2} between two bogies of consecutive wagons (which is typically shorter than Δ_{b_1}),
   - or the wagon length L_w (as indicated by Krylov and Ferguson, 1994).
2. to the passage of each wheel pair over successive sleepers, resulting in a frequency $f_2 = V_{train}/\Delta_{sleeper}$ called *passage frequency* by Krylov and Ferguson (1994). Note that in this case, we consider the wheels as moving sources which may induce a Doppler effect.

Figure 1 depicts these characteristic lengths within a schematic train and Table 1 indicates the values considered in our numerical experiments.

As a first task of this paper, let us define the source time function describing the passage of a train at a given source (sleeper) location, and clarify which spacing Δ controls the frequency f_1.

Effect of train geometry on frequency content

To define the source time function for a given sleeper, we must consider the passage of each wheel of the train over this particular sleeper. Let us start by considering a simplistic train with only one wheel axle per wagon, i.e. where the only characteristic distance is the wagon length, assumed the same for all wagons. In this case, a simple source time function can be defined as a Dirac comb with regular peaks at the arrival times of each wheel at the source location, with a period L_w/V_{train} related to wagon length L_w and train speed V_{train}. Figure 2 shows such a simple Dirac comb (in greeny yellow), together with its spectrum, corresponding to a 8-wagon train with 1 wheel pair per wagon hitting a single sleeper. The spectrum is harmonic, presenting peaks at frequencies that are multiples of the fundamental frequency $f_1 = V_{train}/L_w$. Because the train has a finite length (a finite number of wagons), the spectrum also presents some frequency leakage which causes the frequency peaks to broaden and not to be as sharp as their time-domain counterparts (this may have its importance later on, when convolving this Dirac comb with more sophisticated functions that might amplify minor peaks). We shall now ask ourselves what happens if we simulate a more realistic train, with 4 wheel axles per wagon (Fig. 1a).

Figure 2 presents in blue the Dirac comb that corresponds to a 8-wagon train with 4 axles per wagon, following the geometry of Fig. 1(a) and the values of Table 1. Intuitively, the time intervals between peaks reflect the distances between the wheels of the train. Less intuitively, we observe in the spectrum that adding more wheels (i.e. more characteristic lengths, see Fig. 1) does not introduce extra fundamental frequencies but *modulates* the original spectrum, as expected from the Fourier transform of a doublet of Dirac peaks (e.g. Krylov and Ferguson, 1994, eq. 18). This is shown in Fig. 2 as black dots for the modulation due to the axle distance Δ_a, whose maxima correspond to frequency $f_a = V_{train}/\Delta_a$ (pink dashed line), and as black dashes for the modulation due to both the axle distance Δ_a and the bogie distance Δ_{b_1} within one wagon, with peaks corresponding to frequency $f_{b_1} = V_{train}/\Delta_{b_1}$ (purple dashed line). Although this description does not fully correspond to the train geometry (as it does not account for the bogie distance Δ_{b_2} between consecutive wagons), note how the black dots and dashes well matches the amplitude of the blue peaks, while cancelling out some of the initial greeny-yellow peaks. As a side note, Krylov and Ferguson (1994) suggested that the zeroes introduced by the modulation effect may serve to suppress vibrations at chosen frequencies, for example at the passage frequency, by designing train and track geometry accordingly.

Another conclusion is that it is the wagon length L_w that controls the main fundamental frequency f_1, and not the interval between bogies as intuited by Fuchs et al. (2018). This leads us to suggest that the high-speed trains generating the signals observed by Fuchs et al. (2018) at AlpArray station A002A, with a spectral line spacing of $\Delta f = f_1 = 1.25$ Hz, are not travelling at a speed of 85 km/h, as the authors suggested based on bogie distance, but rather at 119.25 km/h ($\simeq 120$ km/h), which is therefore the train speed we will consider in our numerical experiments.

These simple Dirac combs provide us with a first understanding of the source time functions and of their spectral characteristics. However, they are not entirely satisfactory, as they assume that the load of each wheel is applied as an instantaneous point source on a single sleeper. As a consequence, assuming simple Dirac combs tends to over-estimate the high-frequency content of the source time functions, and therefore of the resulting seismograms. For more realistic source time functions, we need to account for the fact that the load of each axle is actually distributed over several sleepers, via the elastic response of the rail track (e.g. Krylov and Ferguson, 1994; Li et al., 2018b).

Effect of the distribution of axle loads over sleepers on frequency content

Figure 3 presents examples of functions which may be used to describe the distribution of the load of a single wheel over rail track and sleepers. The first approximation discussed in the previous section is a simple Dirac comb (in blue). More realistically, we may consider that, due to the elasticity of the track, point-source loads like wheels actually affect several sleepers simultaneously. A first way to take this into account is to consider simple weights as given for instance by Paderno (2009, their fig. 4) on 5 sleepers (black line in Fig. 3). Considering a load distributed over a certain width has an effect on the spectrum of the temporal loading function and tends to decrease its high-frequency content (intuitively, the wider the distribution, the lower the frequency content). We can expect to retrieve this effect in the final signal that will result from the convolution of the Green function with a source time function (such as the ones shown in Fig. 2) that will itself be convolved with this loading function.

Alternatively, we may consider a smoothed version of these weights, for instance using a Gaussian function (in green in Fig. 3). It is important here to notice that not only the width but also the shape of the loading function is critical in controlling the high-frequency content of the spectrum: for a comparable width, a Gaussian function will yield much less high frequencies (> 20 Hz) than a triangular function.

The use of *ad hoc* loading functions is therefore dubious and may result in incorrect results in the high-frequency band that is of interest to us. As a consequence, we base our source time functions on a more physical approach and consider as a loading function the (opposite of the) counteracting force exerted by the track on the wheel, which is also the force exerted by the track on the ground via the sleepers, as given by the Euler-Bernoulli elastic beam (E-BEB) model (Krylov and Ferguson, 1994, eqs 2 and 7; Li et al., 2018a,

eqs 2 and 3):

$$P_j(x_i) = 2F_i\, e^{-\beta|x_i - x_j|} \left[\cos(\beta|x_i - x_j|) + \sin(\beta|x_i - x_j|)\right] \frac{\Delta_{sleeper}}{x_0}. \quad (2)$$

In this formulation, the loading function $P_j(x_i)$ corresponds to the vertical force (in N) at sleeper j caused by the axle load of wheel i, and depends on the distance $x_i - x_j$ between wheel i and sleeper j, on the axle load F_i (assumed known for a given train weight), on the sleeper spacing $\Delta_{sleeper}$ (quite standard and assumed to be $\Delta_{sleeper} = 24" = 0.6096$ m in our numerical experiments, see Table 1), on the elastic modulus E (in N/m^2) and cross-sectional momentum I (in m^4) of the rail (assumed known for standard steel rails), and finally on the stiffness α (in N/m^2) of the ground beneath the rail, via the coefficient $\beta = (\alpha/4EI)^{0.25}$ (in m^{-1}) and the total deflection distance $x_0 = \pi/\beta$ (in m, Krylov and Ferguson, 1994). Among these parameters, we identify the ground stiffness α as the main parameter susceptible to vary in real-life situations, depending not only on the actual stiffness of the ground locally, but more generally on the coupling between the rail track, the sleepers and the ground, via the ballast and substratum. Figure 3 shows loading functions given by the E-BEB model with various values of ground stiffness: $\alpha = 10$ MN/m^2 (Li et al., 2018b, in greeny yellow), $\alpha = 61.8$ MN/m^2 (in orange, after Krylov and Ferguson, 1994, who document this as a typical value for British railways), and $\alpha = 800$ MN/m^2 (in red, which matches the weights of Paderno, 2009 and –more importantly– best matches the seismic observations, as we will see later in Fig. 4b,c). The stiffer the ground beneath the rail is, the narrower is the spatial loading function, and the higher the frequencies in the source time functions, and therefore –we expect– in the final seismograms. Moreover, amplitudes are normalised in Fig. 3. In absolute, ground stiffness also has an effect on the magnitude of the force exerted on the ground: the stiffer the ground beneath the rail is, the larger the vertical force exerted by the track on the ground (see Fig. S1).

In conclusion, our final source time function $S_j(t)$ (in orange in Fig. 2) is the result of the convolution of the Dirac comb corresponding to 4 wheel axles per wagon (in blue in Fig. 2) with the time derivative of the loading function $P_j(t)$ given by the E-BEB model:

$$S_j(t) = \frac{dP_j(t)}{dt} * \sum_{i=1}^{N_{axles}} \delta(t - t_{ij}), \quad (3)$$

where t_{ij} is the arrival time of wheel i at sleeper j, and the time derivative is needed to simulate ground velocities instead of displacements.

In summary, just looking in detail at the design and characteristics of this source time function already enables us to draw the following conclusions (or, more precisely, to retrieve the conclusions of Krylov and Ferguson, 1994):

1. The sharp frequency peaks observed by Fuchs et al. (2018) correspond to a fundamental frequency $f_1 = V_{train}/L_w$ controlled by train speed and wagon length L_w,

which is the true spatial periodicity of the train wheels.

2. The spectrum of the source time functions is modulated by the spacing of the wheels within each wagon, in particular by the distance Δ_a between axles within each bogie.

3. The frequency bandwidth of the source time functions is further controlled by the distribution of the load of each wheel axle over the rail track and sleepers, which can be estimated through the E-BEB model and parameterised through an effective *ground stiffness* (which in practice may also reflect the coupling between track and ground): the stiffer the ground (the better the coupling), the narrower the loading function, and the higher the frequency content.

4. Whether we look at ground displacement or velocity matters since the time derivative relating both also has an effect on the spectra.

5. From points 2, 3 and 4, we may derive a rough *rule of thumb* according to which we expect most of the energy of ground velocity spectra to be in the frequency band $[\,0.5\,f_a\;,\;1.5\,f_a\,]$, where $f_a = V_{train}/\Delta_a$ is controlled by the train speed and the distance Δ_a between axles within each bogie. Secondary energy packets can also be expected in subsequent frequency bands of the form $[\,(k-0.5)\,f_a\;,\;(k+0.5)\,f_a\,]$, in between zero-energy notches located at frequencies $(k+0.5)\,f_a$. This is illustrated in Fig. 2(b) where most of the energy of the final source time function (in orange) lies between 7 and 20 Hz, while another packet of energy lies between 20 Hz and 33 Hz. It is further illustrated in the Supplemental Material (Figs S2 and S3).

We expect these characteristics of the source time functions to be reflected in the final seismograms generated by train passages, which we shall now investigate in more details by adding the contributions of all sleepers.

Modelling results

Two end-member cases: single stationary source vs. single moving load

Figure 4(a) shows the seismogram (top), spectrogram (bottom left) and spectrum (right) obtained with a single stationary source (one sleeper only) excited by all wheels of a train with 8x 26.5-m-long wagons, moving at 119.25 km/h ($\simeq$ 120 km/h), as suggested by revisiting Fuchs et al. (2018)'s calculation (see above). Note the zoom on the time axis which makes visible the individual spikes corresponding to the signal of each wheel pair. Since only one sleeper is involved, the spectrum is very similar to the spectrum of the individual source time functions (Fig. 2), with a spectral line spacing related to frequency $f_1 = V_{train}/L_w$, controlled by the wagon length L_w, and a modulation related to frequency $f_a = V_{train}/\Delta_a$, linked to axle distance Δ_a. We indicate frequency $f_2 =$

$V_{train}/\Delta_{sleeper}$ related to sleeper spacing $\Delta_{sleeper}$ on the figure, but of course since there is no sleeper spacing in this case, this frequency does not appear at all. As the source is stationary, we do not observe any Doppler effect either. It is worth noting that these spectral characteristics correspond well to most of Fuchs et al.'s observations, one of which is reproduced in Fig. 4(b) (same as fig. 5a in Fuchs et al., 2018).

On the other hand, Figure 4(d) shows the seismogram, spectrogram and spectrum obtained with a single moving load (one wheel pair only) passing all sleepers at a speed of 105 km/h (the choice of this value shall be justified later on). Here the spectrum and spectrogram are completely different from Fig. 4(a), with a spectral line related to $f_2 = V_{train}/\Delta_{sleeper}$ controlled by the sleeper spacing $\Delta_{sleeper}$, and with a clear Doppler effect typical of a moving source. Interestingly, this spectral signature is very close to another type of observations made by Fuchs et al. (2018) (their figs 6a, 6b and S2b), one of which is reproduced in Fig. 4(e) (same as fig. 6a in Fuchs et al., 2018) and displays a clear Doppler effect. In this case, Fuchs et al. (2018) interpreted the frequency spacing as $\Delta_{f_2} = 2\,f_1$, i.e. as being related to $f_1 = V_{train}/L_w$ and therefore controlled by wagon length L_w (Fuchs et al., 2018, figs 6a,b), while the Doppler frequency shift rather suggests the effect of a moving load. Yet, the frequency spacing $\Delta_{f_2} = 2.54$ Hz suggests that the characteristic length at play is not as short as the sleeper spacing, but may be a longer spatial period related to irregularities of the railway (e.g. unwelded rails). For instance, considering that loads are applied only every 19 sleeper introduces a spatial period of $19\Delta_{sleeper} \simeq 11.6$ m, therefore resulting in a fundamental frequency $f_2' = f_2/19 = 2.52$ Hz (Fig. 4f) which would reasonably explain the observations (Fig. 4e).

These two end-member cases, single stationary source vs. single moving load, therefore enable us to understand the two main mechanisms at play, and nicely relate to the two types of observations made by Fuchs et al. (2018), at least in terms of their spectral characteristics. But they are only end-member, simplistic cases, and the question now is: what happens when considering all wheels exciting all sleepers? How do we generate signals similar to the observations, both in the frequency domain (harmonic spectra) and in the time domain (tremor-like emergent shapes)?

Realistic cases (all sleepers, all wheels): effect of sleeper regularity

Figure 4(f) shows the signal obtained with a realistic wheel geometry (see Fig. 1a) and loads applied every 19 sleeper along the railway, with a regular spacing of $19\Delta_{sleeper} \simeq 11.6$ m, for a train with 6x 23-m-long wagons moving at 105 km/h (such as to try and match the observations in Fig. 4(e), which Fuchs et al. (2018) attributed to a local commuter train, see their table 1)[1]. Surprisingly, the spectrogram is much closer to the case of a single moving load (Fig. 4d, see also Fig. S4c-d), with a clear Doppler effect and a

[1]Here our choice of train parameters aims at presenting the example that best illustrates the underlying mechanisms but we realize that it makes the comparison between Figs 4(a,c) and Figs 4(d,f) not straightforward. For the sake of completeness, we present more results in the Supplemental Material (Fig. S4) for one-to-one comparison.

spectrum dominated by frequency $f_2' = V_{train}/(19\Delta_{sleeper})$, than to the stationary source case (Fig. 4a). However, this result matches well the few observations made by Fuchs et al. (2018) (their figs 6a, 6b and S2b) where the Doppler effect is visible. In this specific case, it is indeed easy to get confused and to interpret the main spectral line spacing in terms of multiples of f_1 but a close inspection of the spectrum indicates that the frequency peaks are much better explained by the moving load mechanism and the frequency $f_2/19$ than by frequency f_1, which barely plays a role here (further tests in this sense are presented in the Supplemental Material, Fig. S4).

The question is then: how can we reproduce seismograms similar to Fig. 4(b), with a tremor-like emergent signal in the time-domain and a spectrum dominated by frequency $f_1 = V_{train}/L_w$, controlled by wagon length L_w?

Figure 4(c) shows the signal obtained with all wheels and all sleepers, with slightly irregular sleepers spaced every $\Delta_{sleeper} = 0.6096 \pm 0.05$ m (i.e. we add a random perturbation to the original spacing, with a maximum perturbation of ± 5 cm, which [very roughly] corresponds to a sleeper half-width). Again, we consider a train with 8x 26.5-m-long wagons, moving at $119.25 \simeq 120$ km/h such that $f_1 = 1.25$ Hz, as observed by Fuchs et al. (2018). A striking observation here is that it is sufficient to assume slightly irregular sleepers to reproduce a signal and spectrum that are very close to most of the observations made by Fuchs et al. (2018) on station A002A (emergent tremor-like time-domain signal and harmonic spectrum dominated by frequency $f_1 = V_{train}/L_w$, controlled by wagon length L_w). While it does not dominate the spectrum anymore, the peak at frequency $f_2 = V_{train}/\overline{\Delta_{sleeper}}$ still appears because of the mean sleeper spacing $\overline{\Delta_{sleeper}} = 0.6096$ m. This f_2-peak is not visible in Fig. 4(b) because it is above the Nyquist frequency of 50 Hz but we expect that it should be seen in seismograms resulting from trains travelling at slower speeds. We actually suspect that it is responsible for the secondary maximum that is sometimes observed around 40 Hz by Fuchs et al. (2018) in the spectra generated by commuter trains (e.g. Fuchs et al., 2018, fig. 3). Finally, we speculate that our modelling parameter of *sleeper regularity* that controls the predominance of one or the other end-member mechanisms (single stationary source vs. single moving load) actually reflects the properties of the railway infrastructure in real cases, i.e. not only the actual regularity of sleeper locations but also their coupling with the ground via the ballast and substratum.

Discussion

Limitations of the modelling strategy

A careful reader will have noticed that our modelling does not quantitatively reproduce the absolute amplitudes of observed ground velocities: while the ground velocities generated by high-speed trains observed at station A002A are of the order of 10^{-6} to 10^{-5} m/s (Fig. 4b,e), the signals simulated in similar configurations have amplitudes of the order of 10^{-8} m/s (Fig. 4c,f). This discrepancy may be partly related to poorly-controlled

parameters such as sensor-ground coupling or possible site effects, or to our choice of medium properties, which consist of low quality factors and high seismic velocities in order to reduce computation costs, while in reality station A002A is located in a sedimentary basin where soft, low-velocity materials may amplify the amplitude of surface waves. Future work will focus on quantitative comparisons with data acquired in better-controlled environments in order to calibrate our synthetic amplitudes.

It may also be argued that this discrepancy between observed and simulated amplitudes could be due to the fact that our modelling does not include all the mechanisms involved in reality. Indeed, by following Krylov's model (Krylov and Ferguson, 1994), we only consider a quasi-static excitation caused by the axle load on the rail track, and we neglect dynamic effects related to other excitation mechanisms due for instance to wheel and rail roughness, rail joints, and sleeper-ground coupling (e.g. Xia et al., 2010). Similarly, we do not account for the generation of horizontal forces due to friction and rail curvature, for the heavier axle loads of locomotives at the front and end of the train, or for the elasticity and spatial extension of the sleepers which actually support two rails while we considered only one (which, again, amounts to assume that the sleepers are rigid and that the load of each train axle is applied as a single point source). In hindsight, we think that the discrepancy between observed and simulated amplitudes is most likely due to the latter point, i.e. to the fact that we inject energy via point sources in the spectral-element simulations, without considering the surface of the sleepers.

Nevertheless, we have shown that our simple modelling considering only quasi-static excitations is able to reproduce – qualitatively at least – the seismic observations, which is also consistent with Krylov and Ferguson (1994)'s statement that the quasi-static approximation is sufficient in the frequency range [0 - 100] Hz. We are confident that this modelling strategy is sufficient to catch the main mechanisms at play and enables us to investigate the general behaviour of generated amplitudes with respect to some parameters. Because we do not reproduce absolute amplitudes accurately, however, we restrict the discussion of amplitudes to a qualitative analysis in the following, and present related results in Supplemental Material (Figs S2 and S3).

Parameters controlling signal amplitudes: implications for seismic detection

In view of using train-generated signals for seismic applications, it is important to assess in which conditions we can expect to detect these signals. Having all the above assumptions and limitations in mind, we may therefore attempt to understand the parameters that control the amplitudes of train-generated signals. According to our approach and assumptions, signal amplitude depends on

- the stiffness of the ground beneath the rail: the magnitude of the vertical force acting on the sleepers increase non-linearly with ground stiffness (except for its

quasi-static component which depends only on axle load, see Fig. S1).

- the length and total weight of the train, with two different effects:
  - via the weight of individual wagons which will reflect on the axle load F_i carried by each wheel: eq. 2 shows that the magnitude of the vertical force acting on the sleepers increase linearly with axle load F_i, and so will the amplitude of the resulting signal.
  - via the number of wagons, as the amplitude of the signals also increase with the number of wheel loads N_{wl}. More precisely, our numerical experiments show that the maximum amplitude of the time-domain signals increases as $\sqrt{N_{wl}}$, while the energy in a given frequency bandwidth increases linearly with the number of wheel axles (see Fig. S2).
- the train speed: faster trains make more noise. More precisely, our numerical experiments show that both the maximum time-domain amplitude and the total signal energy increase supra-linearly with train speed, and so does the energy in frequency bands that evolve proportionally with train speeds, such as $[\,0.5\,,1.5\,]\,f_a$, with $f_a = V_{train}/\Delta_a$ (see Fig. S3). Interestingly, frequency bands of the form $[\,k-0.5\;,\;k+0.5\,]\,f_a$ always contain approximately the same proportion of the spectrum, which brings us back to our *rule of thumb* stating that most of the energy of train-generated signals is expected to be in the $[\,0.5\,,1.5\,]\,f_a$ frequency band controlled by train speed and axle distance Δ_a within each bogie. A direct implication of this for detecting train-generated signals in seismic records is that the considered frequency band should be adapted to the estimated speed of the trains generating these signals.

In conclusion, if we want to maximize the detection of train-generated signals, we should targets portions of railways that enable a good transmission of vibrations to the ground (i.e. with sleepers well coupled to a stiff ground) and where long and/or heavy trains travel at relatively high, and ideally variable, speeds. Moreover, the considered frequency band should be adapted to the estimated speed of the trains generating the signals, targeting the frequency band $[\,0.5\,,1.5\,]\,f_a$ where we expect most of the energy, or higher-frequency bands of the form $[\,k-0.5\;,\;k+0.5\,]\,f_a$, and avoiding narrow frequency bands around the zeroes of the modulated spectrum ($[k+0.5]\,f_a$ and, in a lesser extent, $[k+0.5]\,f_{b_1}$). If the trains travel at variable speeds, their signal becomes more broadband and the effect of these notches might be less critical (see next section). Of course, the detection of these signals in field recordings will also depend on the surrounding noise level, as well as on medium heterogeneity, attenuation and scattering.

Variable train speed

Figure 5 presents the example of a signal generated by a train travelling at a variable speed. When the speed of the train varies, the fundamental frequencies $f_1 = V_{train}/L_w$ and $f_2 = V_{train}/\Delta_{sleeper}$ vary too, as well as their harmonics, which results in *undulating* spectral lines in the spectrograms, and in a *blurring* of the frequency peaks (and notches) in the corresponding spectra, where it is then difficult to identify individual rays. The spectrum is not harmonic anymore but becomes truly broadband as frequency notches disappear, which is actually an advantage for seismic applications based on noise cross-correlations, as distinct spectral rays are difficult to handle by seismic interferometry. Note however that our rule of thumb remains useful: the frequency band $[\,0.5\,,1.5\,]\,f_a$, where f_a can be estimated from a maximum or averaged train speed, still contains most of the energy of the signal. Note also that the variable speed might lead to other prominent maxima in the spectrum, as for instance around 37 Hz in Fig. 5 where a change of train speed happens to occur when the train is closest to the station, which enhances the secondary maximum related to the Doppler effect.

Wave propagation effects

So far we have eluded wave propagation effects on purpose in order to focus on source mechanisms. Of course, once generated, train vibrations propagate as any other seismic wave. For the sake of completeness, we discuss here how wave propagation effects impact train-generated signals. We present associated results in Supplemental Material (Figs S5 and S6).

Short- vs. long-distance signals

Our tests suggest that long-range signals recorded at several kilometers distance have the same properties as the examples we have shown here at short range (300 to 1200 m). In particular, the main characteristics of the spectra obtained at short vs. long distances are nearly identical. The only differences between short- and long-distance signals that we could notice are related to foreseeable wave propagation effects, namely (1) to the expected decay of high frequencies at long distances, and (2) to the relative decay of amplitudes: short-distance signals have a shorter apparent duration because their relative amplitudes are dominated by the effect of the few closest sleepers, while long-distance signals have a longer apparent duration because they see a larger number of sources contributing with an equivalent importance (Figs S5a,b).

Effect of seismic properties (velocities and attenuation)

Obviously, intrinsic attenuation also has an effect on the relative decay of amplitudes and of high frequencies in the spectra. Besides, higher velocities also result in a longer apparent duration of the signal, which can be explained in terms of relative decay of amplitudes due

to geometrical spreading over a smaller number of longer wavelengths. Less intuitively, higher velocities yield lower amplitudes, due to the fact that the amplitude of surface waves is enhanced in soft, low-velocity media. As already mentioned, this may be one of the reasons why the amplitudes of observed train-generated signals are larger than our synthetic amplitudes, obtained for seismic velocities that have been chosen artificially high to mitigate computation costs. Finally, a last effect of seismic velocities relates to the Doppler frequency shift, which will increase if seismic velocities get closer to the (apparent) train speed (i.e. if seismic velocities decrease, or if train speed increases, or if the sensor is aligned with the train trajectory).

Wave partitioning (body vs. surface waves)

We also conducted tests in a medium with a velocity gradient where we selected either the refracted P wave (first arrival) or the surface waves (most energetic arrivals) in the impulse response, before convolving with the train source time functions and stacking. These tests aimed at mimicking long-distance wave propagation in heterogeneous media where high-frequency surface waves would vanish due to strong scattering and attenuation in the shallow subsurface, which is difficult to reproduce in our synthetic simulations. Again, we observe nearly identical spectra in both cases, which suggests that our conclusions, although mostly based on short-distance signals dominated by surface waves, remain valid for long-distance body waves that sample deeper regions of the crust (Brenguier et al., 2019).

Finally, it is worth mentioning that simulations performed in similar configurations but in media with homogeneous vs. gradient velocity structures yield very similar results, suggesting that medium structure plays little role in the patterns of the resulting signals, which complexity mostly comes from the source, i.e. from the train source time functions and from the interferences between waves generated by the many sources involved.

Conclusions

In this paper, we have presented a modelling strategy for simulating seismic waves generated by trains, explained the main mechanisms at play, and derived implications for seismic applications. Our results showed that simple assumptions following Krylov's model of a quasi-static excitation due to axle loads applied as vertical forces distributed on discrete sleepers are sufficient to generate signals that reproduce very well the main characteristics of seismic observations (emergent tremor-like signals with harmonic spectra). Our modelling enabled a detailed understanding of the spectral characteristics of train-generated signals and of their fundamental frequencies. As already noted by Krylov and Ferguson (1994), the exact frequency content of the signals results from a modulation of the initial harmonic spectrum by the wheel spacing within each wagon, and their high-frequency content is further controlled by the distribution of the load of each wheel over the track

and sleepers, which most notably depends on the stiffness of the ground beneath the railway (i.e. on track-ground coupling).

Moreover, we have identified two end-member mechanisms that correspond to the two types of observations made by Fuchs et al. (2018) and depend on sleeper *regularity* (i.e. on the properties of the railway infrastructure in real cases). In the case of perfectly regular sleepers, signals are dominated by the signature of each wheel acting as a single moving load passing over the sleepers. The spectrum then displays peaks corresponding to the fundamental frequency $f_2 = V_{train}/\Delta_{sleeper}$ controlled by train speed and sleeper spacing $\Delta_{sleeper}$, and to its harmonics. In the case of irregular sleepers, signals are dominated by the response of each sleeper acting as a stationary source excited by the passage of the wheels. The spectrum then displays peaks at harmonics of the fundamental frequency $f_1 = V_{train}/L_w$ controlled by train speed and wagon length L_w.

Beyond these technical explanations, we were able to derive important consequences for the future use of train-generated signals for seismic applications. First, it is worth noting that trains generate a very broad and high frequency content [1 - 50 Hz or above], as observed by (e.g.) Fuchs et al. (2018); Brenguier et al. (2019); Dales et al. (2020), because most of the energy actually comes from high-frequency harmonics and not from the fundamental frequencies, which are not necessarily very high ($f_1 \simeq 0.25$ to 5 Hz, depending on train speed and wagon length). To summarize this, we have derived a simple rule of thumb according to which most of the energy radiated by trains should be expected in the range $[\,0.5\,,1.5\,]\,f_a$, where the frequency $f_a = V_{train}/\Delta_a$ is controlled by train speed and by the distance Δ_a between axles within a bogie (usually of the order of $\Delta_a \simeq 1.5$ to 3 m). Higher frequency bands of the form $[\,k - 0.5\ ,\ k + 1.5\,]\,f_a$ are also expected to contain some energy which may prove particularly useful when looking specifically for body waves (e.g. Nakata et al., 2015; Brenguier et al., 2019; Dales et al., 2020). Second, we were able to identify the parameters that control the amplitudes of train-generated signals, and therefore to give guidelines in order to maximize their detection: an ideal seismic source is a long, heavy and fast train traveling on a railway whose sleepers are well coupled to a stiff soil.

Of course, the detection of train-generated signals will also depend on the noise level due to natural or other anthropic noise sources, as well as on medium heterogeneity, attenuation and scattering. While we cannot give general and quantitative rules, we believe that this paper provides elements of understanding and guidelines that will help future case-specific studies to detect and process train-generated seismic signals in ambient noise data (e.g. Nakata et al., 2011; Quiros et al., 2016; Inbal et al., 2018; Li et al., 2018a; Dales et al., 2020), or to design experiments that will explicitly target these signals (e.g. Nakata et al., 2011; Quiros et al., 2016; Brenguier et al., 2019).

Data and resources

Simulations have been performed with the SEM46 code developed in the SEISCOPE Consortium (`http://seiscope2.osug.fr`). Most of the computations were performed using the GRICAD infrastructure (`https://gricad.univ-grenoble-alpes.fr`), which is partly supported by the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche. This study is based on data from the AlpArray Seismic Network (2015) which at the time of publication was not publicly available. Please visit the project page (`http://www.alparray.ethz.ch/en/seismic_network/backbone/data-policy-and-citation`) for more details on data access. All data processing and plotting were done using Python and the ObsPy toolbox (Krischer et al., 2015).

Supplemental Material for this article includes (1) more details on the effect of ground stiffness on the amplitude of the loading force according to the E-BEB model, (2) examples of signals dominated by the signature of moving loads, (3) details on the effect of train length and train speed on the amplitude of the signals discussed in the paper, and (4) examples of wave propagation effects mentioned in the discussion. For the sake of reproducibility, the computer programs developed for the convolutions of Green functions with train source time functions are available at `https://gricad-gitlab.univ-grenoble-alpes.fr/pacific/publications/2020_Lavoue-et-al_SRL_supplemental-material`.

All websites were last accessed on September 2020.

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 (`https://www.pacific-h2020.eu`). We also acknowledge support from the European Research Council under grant No. 817803, FAULTSCAN. ISTerre is part of Labex OSUG@2020 (ANR10 LABX56).

We thank the Editors, an anonymous Reviewer, and Florian Fuchs for their review of the manuscript. We further thank Florian Fuchs for providing us with the timing information of the seismograms used in his paper (Fuchs et al., 2018) and for fruitful discussions at the occasion of the Cargèse workshop on Passive Imaging 2019. We acknowledge the sponsors of the SEISCOPE Consortium in which the SEM46 code has been developed. We thank Ludovic Métivier for his advice on the software and Marwan Irnaka for the use of the Seiscope PyTools.

We acknowledge the use of data from the AlpArray network and would like to thank the AlpArray Seismic Network Team: György HETÉNYI, Rafael ABREU, Ivo ALLEGRETTI, Maria-Theresia APOLONER, Coralie AUBERT, Simon BESANÇON, Maxime BÈS DE BERC, Götz BOKELMANN, Didier BRUNEL, Marco CAPELLO, Martina

ČARMAN, Adriano CAVALIERE, Jérôme CHÈZE, Claudio CHIARABBA, John CLINTON, Glenn COUGOULAT, Wayne C. CRAWFORD, Luigia CRISTIANO, Tibor CZIFRA, Ezio D'ALEMA, Stefania DANESI, Romuald DANIEL, Anke DANNOWSKI, Iva DASOVIĆ, Anne DESCHAMPS, Jean-Xavier DESSA, Cécile DOUBRE, Sven EGDORF, ETHZ-SED Electronics Lab, Tomislav FIKET, Kasper FISCHER, Wolfgang FRIEDERICH, Florian FUCHS, Sigward FUNKE, Domenico GIARDINI, Aladino GOVONI, Zoltán GRÁCZER, Gidera GRÖSCHL, Stefan HEIMERS, Ben HEIT, Davorka HERAK, Marijan HERAK, Johann HUBER, Dejan JARIĆ, Petr JEDLIČKA, Yan JIA, Hélène JUND, Edi KISSLING, Stefan KLINGEN, Bernhard KLOTZ, Petr KOLÍNSKÝ, Heidrun KOPP, Michael KORN, Josef KOTEK, Lothar KÜHNE, Krešo KUK, Dietrich LANGE, Jürgen LOOS, Sara LOVATI, Deny MALENGROS, Lucia MARGHERITI, Christophe MARON, Xavier MARTIN, Marco MASSA, Francesco MAZZARINI, Thomas MEIER, Laurent MÉTRAL, Irene MOLINARI, Milena MORETTI, Anna NARDI, Jurij PAHOR, Anne PAUL, Catherine PÉQUEGNAT, Daniel PETERSEN, Damiano PESARESI, Davide PICCININI, Claudia PIROMALLO, Thomas PLENEFISCH, Jaroslava PLOMEROVÁ, Silvia PONDRELLI, Snježan PREVOLNIK, Roman RACINE, Marc RÉGNIER, Miriam REISS, Joachim RITTER, Georg RÜMPKER, Simone SALIMBENI, Marco SANTULIN, Werner SCHERER, Sven SCHIPPKUS, Detlef SCHULTE-KORTNACK, Vesna ŠIPKA, Stefano SOLARINO, Daniele SPALLAROSSA, Kathrin SPIEKER, Josip STIPČEVIĆ, Angelo STROLLO, Bálint SÜLE, Gyöngyvér SZANYI, Eszter SZŰCS, Christine THOMAS, Martin THORWART, Frederik TILMANN, Stefan UEDING, Massimiliano VALLOCCHIA, Luděk VECSEY, René VOIGT, Joachim WASSERMANN, Zoltán WÉBER, Christian WEIDLE, Viktor WESZTERGOM, Gauthier WEYLAND, Stefan WIEMER, Felix WOLF, David WOLYNIEC, Thomas ZIEKE, Mladen ŽIVČIĆ, Helena ŽLEBČÍKOVÁ.

References

AlpArray Seismic Network (2015). *AlpArray Seismic Network (AASN) temporary component.* AlpArray Working Group. doi:10.12686/alparray/z3_2015. Datacite link: http://data.datacite.org/10.12686/alparray/z3_2015 (last accessed March 2020).

Brenguier, F., Boué, P., Ben-Zion, Y., Vernon, F., Johnson, C., Mordret, A., Coutant, O., Share, P.-E., Beaucé, E., Hollis, D., and Lecocq, T. (2019). Train traffic as a powerful noise source for monitoring active faults with seismic interferometry. Geophysical Research Letters, 46.

Connolly, D., Kouroussis, G., Laghrouche, O., Ho, C., and Forde, M. (2015). Benchmarking railway vibrations — track, vehicle, ground and building effects. Construction and Building Materials, 92:64–81.

Dales, P., Pinzon-Rincon, L., Brenguier, F., Boué, P., Arndt, N., McBride, J., Lavoué,

F., Bean, C. J., Beauprêtre, S., Fayjaloun, R., and Olivier, G. (2020). Virtual sources of body-waves from noise correlations in a mineral exploration context. Seismological Research Letters, 91(4):2278–2286.

Degrande, G. and Schillemans, L. (2001). Free field vibrations during the passage of a Thalys high-speed train at variable speed. Journal of Sound and Vibration, 247(1):131–144.

Ditzel, A., Herman, G., and Drijkoningen, G. (2001). Seismograms of moving trains: comparison of theory and measurements. Journal of Sound and Vibration, 248(4):635–652.

Fuchs, F., Bokelmann, G., and the AlpArray Working Group (2018). Equidistant spectral lines in train vibrations. Seismological Research Letters, 89(1):56–66.

Galvín, P., Romero, A., and Domínguez, J. (2010a). Fully three-dimensional analysis of high-speed train–track–soil-structure dynamic interaction. Journal of Sound and Vibration, 329(24):5147–5163.

Galvín, P., Romero, A., and Domínguez, J. (2010b). Vibrations induced by HST passage on ballast and non-ballast tracks. Soil Dynamics and Earthquake Engineering, 30(9):862–873.

Inbal, A., Cristea-Platon, T., Ampuero, J., Hillers, G., Agnew, D., and Hough, S. E. (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.

Kouroussis, G., Verlinden, O., and Conti, C. (2011). Free field vibrations caused by high-speed lines: Measurement and time domain simulation. Soil Dynamics and Earthquake Engineering, 31(4):692–707.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., and Wassermann, J. (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem. Computational Science & Discovery, 8(1):014003.

Krylov, V. and Ferguson, C. (1994). Calculation of low-frequency ground vibrations from railway trains. Applied Acoustics, 42(3):199–213.

Li, C., Li, Z., Peng, Z., Zhang, C., Nakata, N., and Sickbert, T. (2018a). Long-period long-duration events detected by the IRIS community wavefield demonstration experiment in Oklahoma: Tremor or train signals? Seismological Research Letters, 89(5):1652–1659.

Li, L., Nimbalkar, S., and Zhong, R. (2018b). Finite element model of ballasted railway with infinite boundaries considering effects of moving train loads and Rayleigh waves. Soil Dynamics and Earthquake Engineering, 114:147–153.

Lombaert, G. and Degrande, G. (2009). Ground-borne vibration due to static and dynamic axle loads of InterCity and high-speed trains. Journal of Sound and Vibration, 319(3):1036–1066.

Nakata, N., Chang, J. P., Lawrence, J. F., and Boué, P. (2015). Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry. Journal of Geophysical Research: Solid Earth, 120(2):1159–1173.

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

Paderno, C. (2009). Simulation of ballast behaviour under traffic and tamping process. In 9th Swiss Transport Research Conference (STRC 2009).

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

Rezaeifar, M., Maggio, G., Xu, Y., Bean, C., Lavoué, F., Boué, P., Pinzon-Rincon, L., and Brenguier, F. (2020). Imaging shallow structures in Dublin city using seismic interferometry of seismic waves generated by a train traffic. In EGU General Assembly, number 16678.

Scuderi, M. M., Marone, C., Tinti, E., Di Stefano, G., and Collettini, C. (2016). Precursory changes in seismic velocity for the spectrum of earthquake failure modes. Nature Geoscience, 9:695 EP.

Sheng, X., Jones, C., and Thompson, D. (2003). A comparison of a theoretical model for quasi-statically and dynamically induced environmental vibration from trains with measurements. Journal of Sound and Vibration, 267(3):621–635.

Trinh, P.-T., Brossier, R., Métivier, L., Tavard, L., and Virieux, J. (2019). Efficient time-domain 3D elastic and viscoelastic full-waveform inversion using a spectral-element method on flexible Cartesian-based mesh. Geophysics, 84(1):R75–R97.

Wu, T. and Thompson, D. (2001). Vibration analysis of railway track with multiple wheels on the rail. Journal of Sound and Vibration, 239(1):69–97.

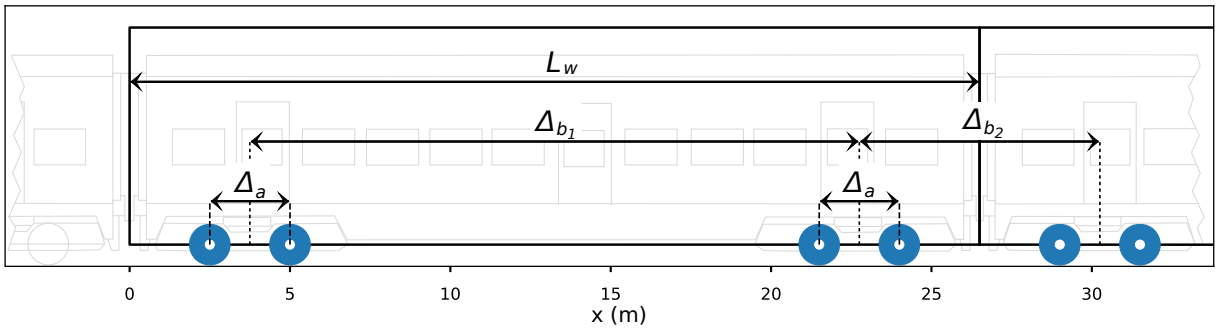
Xia, H., Cao, Y., and Roeck, G. D. (2010). Theoretical modeling and characteristic analysis of moving-train induced ground vibrations. Journal of Sound and Vibration, 329(7):819–832.

Tables

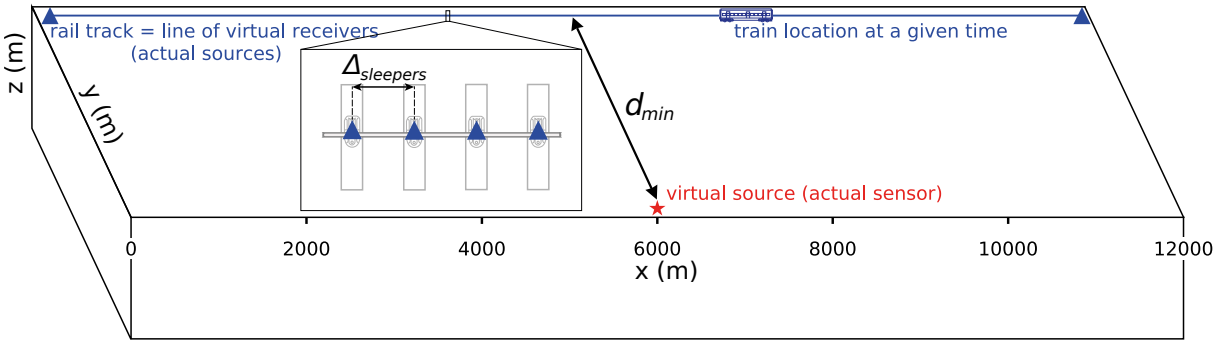
Medium properties
$V_P = 3400$ m/s, $V_S = 2000$ m/s, $\rho = 2600$ kg/m^3, $Q_P = 100$, $Q_S = 50$
ground stiffness beneath railway: $\alpha = 800$ MN.m^{-2}
Train properties
$L_w = 26.5$ m, $\Delta_a = 2.5$ m, $\Delta_{b_1} = 19$ m, $\Delta_{b_2} = 7.5$ m
or $L_w = 23$ m, $\Delta_a = 2.3$ m, $\Delta_{b_1} = 18.5$ m, $\Delta_{b_2} = 4.5$ m
axle load: 17 t (167 kN)
Acquisition and railway geometry
sleeper spacing $\Delta_{sleeper} = 24" = 0.6096$ m
number of sleepers $N_{sleeper} = 19358$
railway length $L_{railway} = 11800$ m $\simeq 12$ km
distance between sensor and railway $d_{min} = 300$ to 1200 m

Table 1: Settings of the numerical experiments presented in the paper. Train properties are taken from Fuchs et al. (2018, Table 1). Note that seismic velocities are deliberately quite high in order to reduce computational costs. Attenuation is also deliberately high in order to better mimic a finite-size medium. We have checked that our main conclusions do not depend on medium properties.

Figures



(a) Train geometry.



(b) Acquisition geometry.

Figure 1: Geometry of a train (a) and of the numerical experiments (b).

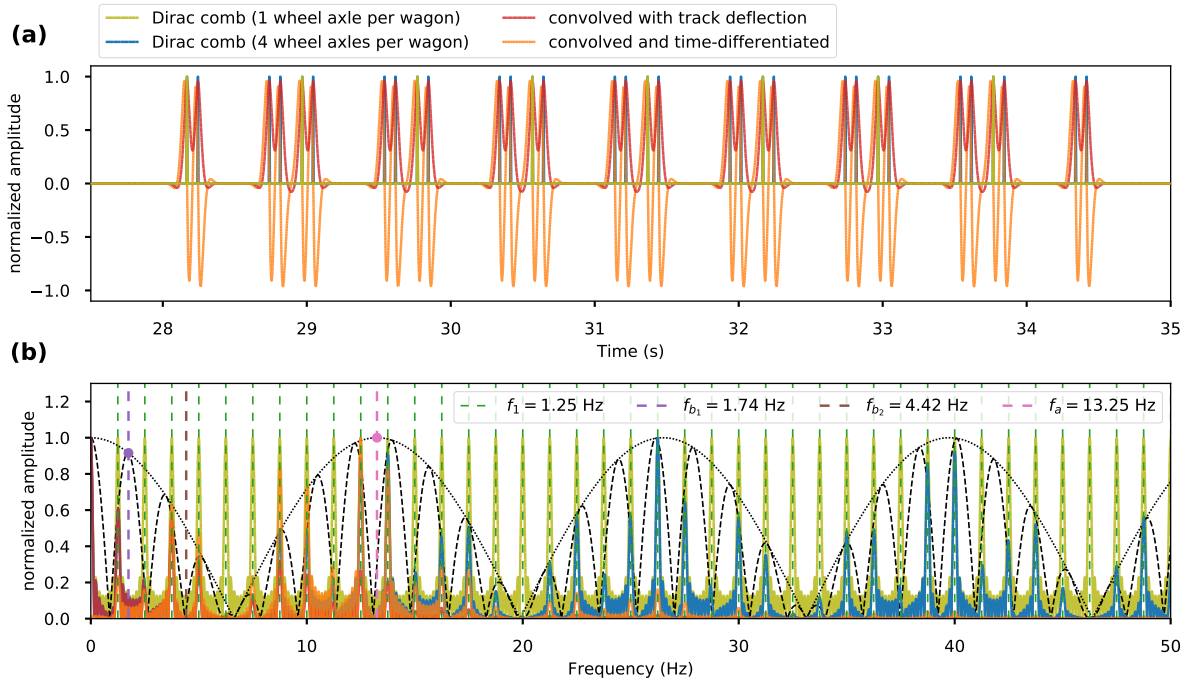


Figure 2: Source time functions (a) and corresponding spectra (b) for a sleeper located at $x = 1$ km along the railway, consisting of pure Dirac combs for 1 and 4 wheel axles per wagon (in greeny-yellow and blue, respectively) and of a 4-axle Dirac comb convolved with a loading function derived from the E-BEB model, assuming a ground stiffness $\alpha = 800$ MN/m^2 (in red, see also Fig. 3). The orange curves correspond to the final source time function obtained after time differentiation of the red curve, in order to simulate ground velocities instead of displacements. We therefore expect the orange spectrum in (b) to be representative of the energy content of the final seismogram obtained by summing the contributions of all sources. In panel (b), black dots and dashes represent the modulation due to axle distance and bogie distance within wagons, respectively. Associated fundamental frequencies are indicated as vertical dashed lines.

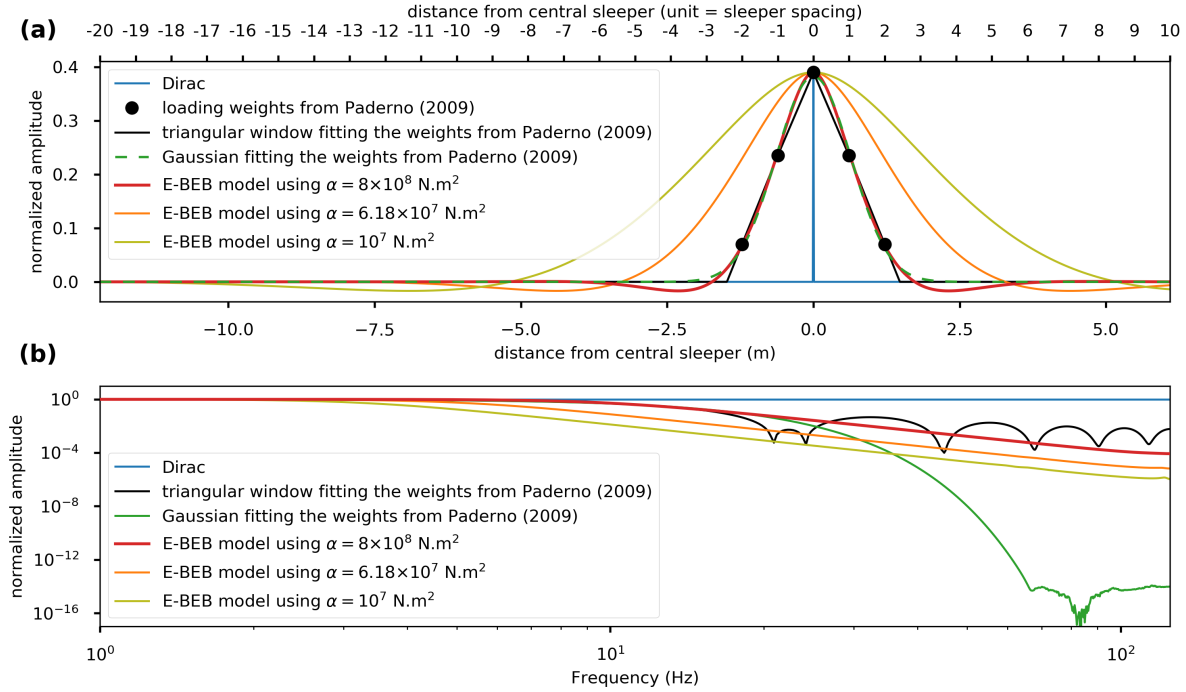
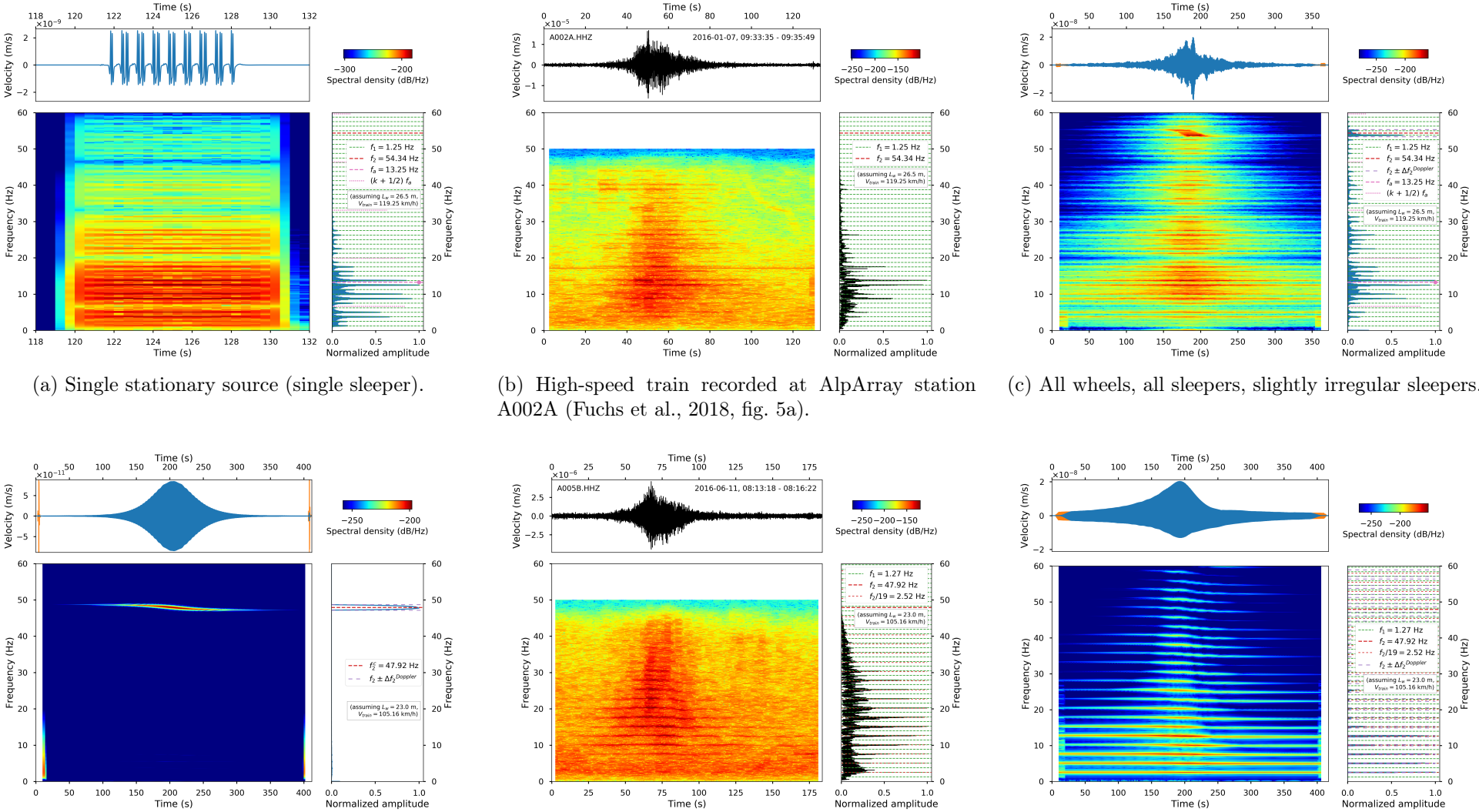


Figure 3: Loading functions representing the distribution of a single wheel load on the track, as a function of distance between the wheel and a given sleeper (a) and corresponding spectra (b), after conversion of the space axis into time using a train speed of 120 km/h.



(a) Single stationary source (single sleeper).

(b) High-speed train recorded at AlpArray station A002A (Fuchs et al., 2018, fig. 5a).

(c) All wheels, all sleepers, slightly irregular sleepers.

(d) Single moving load on perfectly regular sleepers.

(e) Commuter train recorded at AlpArray station A005B (Fuchs et al., 2018, fig. 6a).

(f) All wheels, every 19 sleeper, perfectly regular sleepers.

Figure 4: Seismograms (top of each panel), spectrograms (bottom of each panel) and spectra (right of each panel) resulting from train passages. The first column (panels a,d) corresponds to the two synthetic end-members, the second column (panels b,e) to observations, and the third column to *realistic* synthetic cases involving all sleepers and all train wheels. The first row (panels a, b, c) corresponds to signals dominated by the end-member signature of a single stationary source, while the second row (panels d,e,f) presents signals dominated by the signature of a single moving load. AlpArray data (b,e) are highpass-filtered above 0.5 Hz. Synthetic data are lowpass-filtered below 100 Hz. Spectrograms are computed over time windows of 5 s for observations (b,e) and 20 s for synthetics (a,c,d,f) with a 90% overlap. Synthetic spectra are computed on tapered data (blue seismograms in panels c, d, f) in order to avoid the start and the end of the signal where only some of the train wheels are involved (in orange).

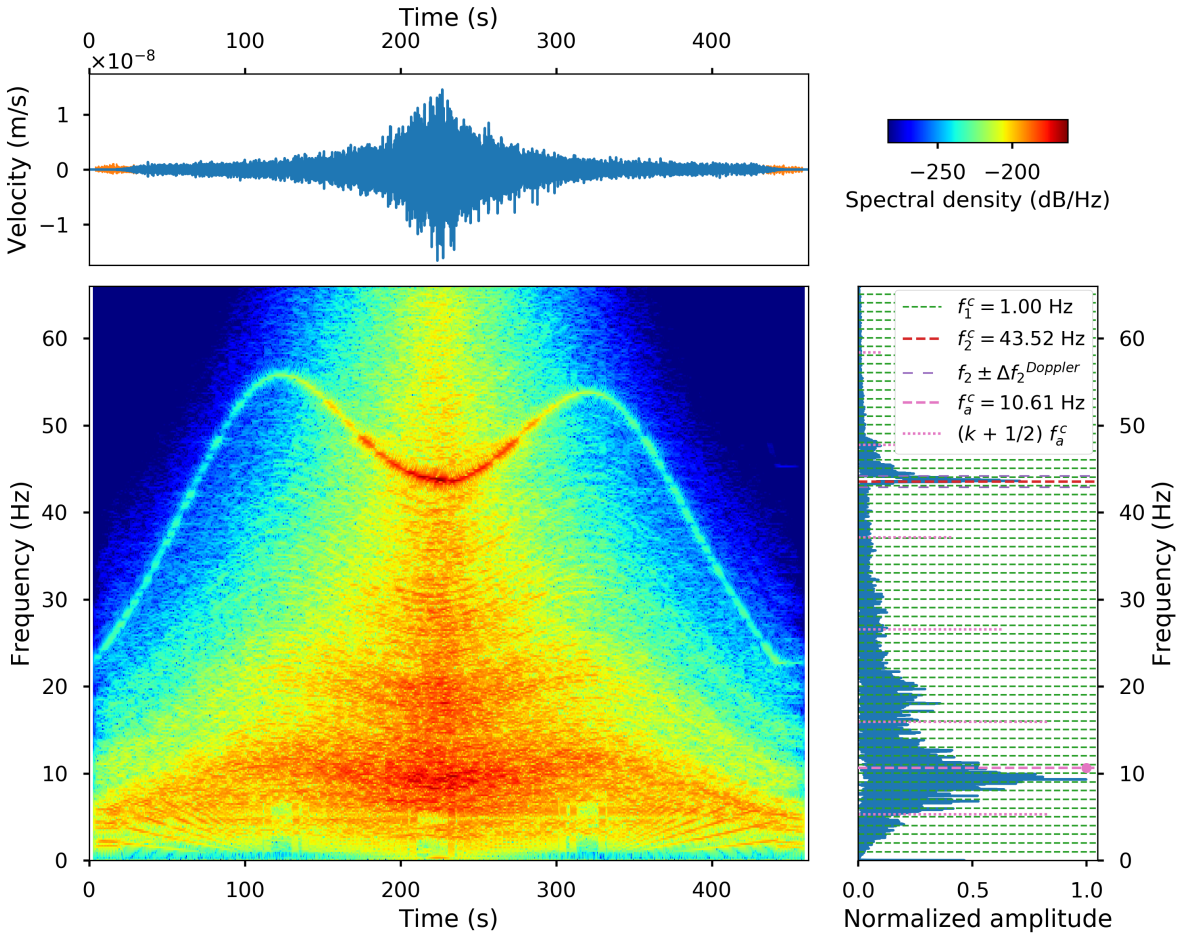


Figure 5: Seismogram (top), spectrogram (bottom) and spectrum (right) resulting from a train travelling over slightly irregular sleepers at a speed increasing from 50 km/h to 120 km/h between 0 and $\simeq$ 120 s, then decreasing to 95 km/h ($\simeq$ 230 s), increasing again to 120 km/h ($\simeq$ 320 s) and finally decreasing to 50 km/h. The spectrogram is computed over time windows of 5 s with a 90% overlap. The spectrum is computed on tapered data (blue seismograms) in order to avoid the start and the end of the signal where only some of the train wheels are involved (in orange in the upper panel).