Detection of random noise and anatomy of continuous seismic waveforms in dense array data near Anza California

Haoran Meng [1], Yehuda Ben-Zion[1] and Christopher W. Johnson[2]

[1]*Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089–0740, USA. E-mail: haoranme@usc.edu*
[2]*Scripps Institution of Oceanography, Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, CA 92093, USA*

SUMMARY

We develop a methodology to separate continuous seismic waveforms into random noise (RN), not random noise (NRN) produced by earthquakes, wind, traffic and other sources of ground motions, and an undetermined mixture of signals. The analysis is applied to continuous records from a dense seismic array on the San Jacinto fault zone. To detect RN signals, we cut hourly waveforms into non-overlapping 1 s time windows and apply cross-correlations to separate RN candidates from outliers. The cross-correlation coefficients between different RN candidates fall into a tight range (i.e. 0.09–0.35), while cross-correlation coefficients of RN candidates with NRN signals (e.g. seismic or air-traffic events) are lower. The amplitude spectra of RN candidates have a well-defined level, while the amplitude spectra of other signals deviate from that level. Using these properties, we examine the amplitude spectra of moving time windows and cross-correlation coefficients with RN templates in each hour. The hourly RN is quasi-stationary and the results cluster tightly in the parameter space of cross-correlation coefficients and L2 norm deviations from the mean spectra of RN candidates. Time windows with parameters in this tight cluster are identified as RN, windows that deviate significantly from the RN cluster are identified as NRN and windows with values in between are identified as mixed signals. Several iterations on each hourly data are used to update and stabilize the selection of RN templates and mean noise spectra. For the days examined, the relative fractions of RN, NRN and mixed signals in local day (night) times are about 26 (42 per cent), 40 (33 per cent) and 34 per cent (25 per cent), respectively.

Key words: Fourier analysis; Time-series analysis; Seismic noise; Site effects; Wave propagation; Wave scattering and diffraction.

1 INTRODUCTION

Seismic ground motion is generated by a variety of mechanisms including tectonic sources such as earthquakes and tremor (e.g. Shelly *et al.* 2007; Ross *et al.* 2017), interactions of ocean waves with the solid Earth (e.g. Gerstoft & Tanimoto 2007; Hillers *et al.* 2013), storms and wind shaking of structures, trees and other obstacles (e.g. Tanimoto & Valovcin 2015, Johnson *et al.* 2019) and various anthropogenic sources such as trains, cars, airplanes, helicopters and wind turbines (e.g. Eibl *et al.* 2015; Salvermoser *et al.* 2015; Neuffer & Kremers 2017; Inbal *et al.* 2018; Meng & Ben-Zion 2018a). It is generally assumed that continuous waveforms processed to remove impulsive sources (e.g. Bensen *et al.* 2007) consist primarily of a random wavefield, referred to as the ambient seismic noise, which is suitable generally for imaging and monitoring of sub-surface structures (e.g. Shapiro *et al.* 2005; Brenguier *et al.* 2008; Zigone *et al.* 2015). Extracting accurate Green's functions from the ambient seismic noise requires a stronger condition

that the wavefield is fully diffuse (e.g. Weaver 1982; Shapiro *et al.* 2000; Sánchez-Sesma *et al.* 2008). Demonstrating that a wavefield is diffuse requires analysis with techniques of the type developed by Margerin *et al.* (2009), Sens-Schonfelder *et al.* (2015) and Liu & Ben-Zion (2016).

In this paper, we develop techniques based on cross-correlations and the amplitude spectra of moving time windows to separate wave packets with random signals from waveform sections having structured signals. The results help to clarify general aspects of observed seismic waveforms, and the separated random/not-random signals can be used in various additional analyses involving detection of small seismic events, imaging and monitoring sub-surface structures, etc. The study employs data of a dense seismic array with 1108 vertical component 10 Hz geophones at the Sage Brush Flat (SGB) site on the Clark branch of the San Jacinto fault zone (SJFZ) southeast of Anza, California (the red box, Fig. 1a). The array recorded continuously at 500 Hz from 2014 May 7 to 2014 June 13 in an area of about 600 m × 600 m (Fig. 1b) with nominal
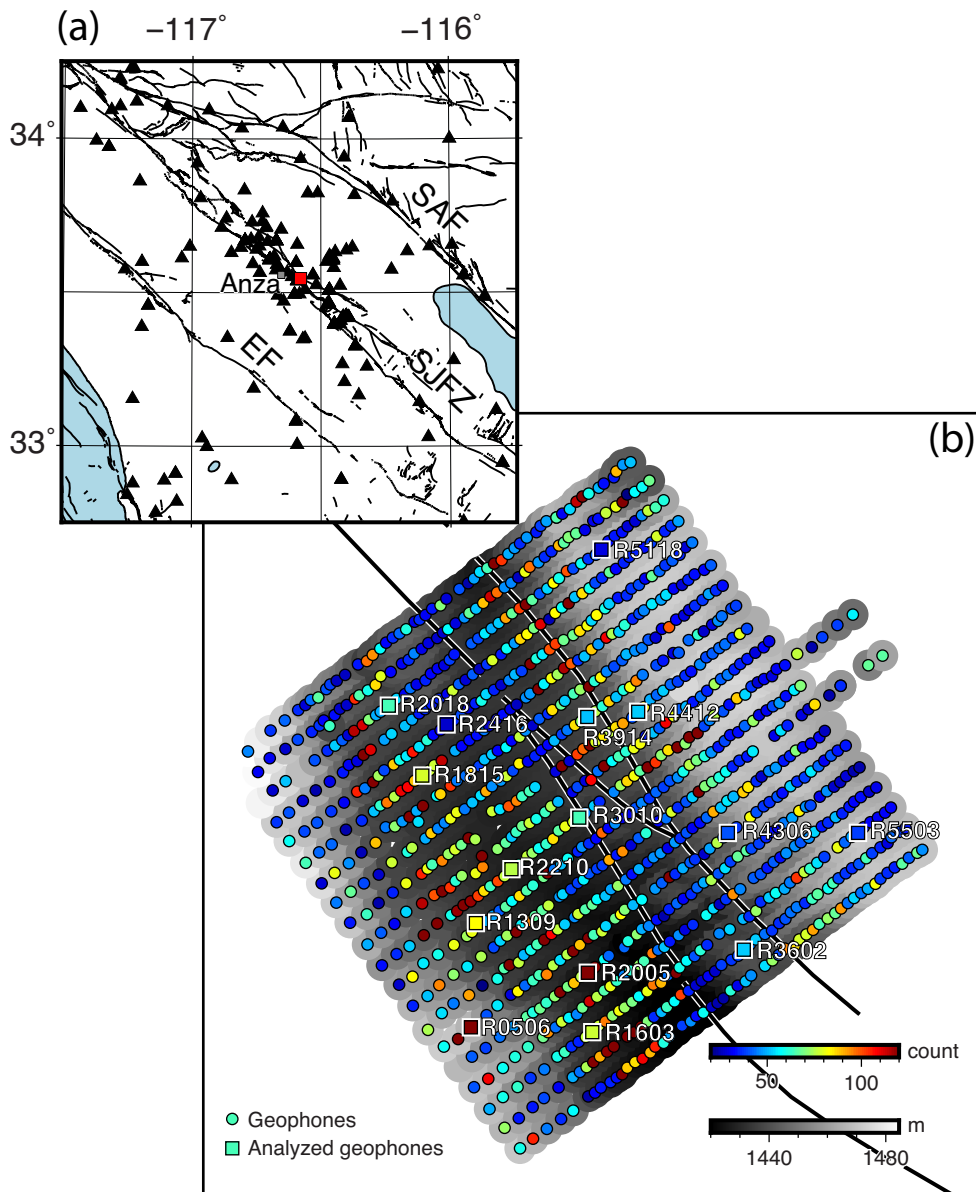
Figure 1. (a) Regional map showing the dense array on the San Jacinto fault zone (the red square) with the regional seismic stations (the black triangles) in southern California. (b) The coloured circles show the median absolute amplitude of waveforms recorded at the array stations on Julian day 146, 2014. The geophones with analysed data are marked by the squares. The background grey colours represent topography.

sensor spacing of 10 m normal to the fault and 30 m along-strike (Ben-Zion *et al.* 2015). The dense array data have been used for detailed imaging of subsurface structures at different depth ranges (Hillers *et al.* 2016; Qin *et al.* 2018; Mordret *et al.* 2019) and detection of microearthquakes (Meng & Ben-Zion 2018b; Gradon *et al.* 2019).

As part of an effort to understand the main contributions to ground motion observed at the SGB site, Meng & Ben-Zion (2018b) analysed waveforms generated by airplanes and helicopters, and found that air-traffic events occupy >7 per cent of the recorded wavefield. Johnson *et al.* (2019) showed that wind interactions with obstacles above the surface produce ground velocities larger than those expected to be generated by earthquakes with magnitudes M ≤ 1.5 about 6 per cent of the day. The air-traffic and wind-related events produce earthquake- and tremor-like waveforms that stand out clearly from the seismic noise field recorded at the site. Less

frequent sources of seismic signals that are visually distinguishable from the noise are genuine earthquakes and local car traffic (Fig. 2).

Inspection of waveforms at the SGB site shows a variety of emergent and impulsive signals occurring throughout the day. Quantitative analysis of the main classes of waveforms can improve the understanding of the observed ongoing ground motion and provide useful information for refined seismic imaging and detection studies. Towards this goal, we use the dense SGB data to separate Random-Noise (RN) signals from wave packets containing Not-Random-Noise (NRN) data produced by various sources such as air and car traffic, wind and earthquakes. To identify wave packets that may be labelled as RN, we use 1000 1-s time windows for each hour and select those with the lowest root-mean-squared (RMS) amplitude as initial candidate RN wave packets. The selection of RN signals is refined iteratively using cross-correlations and the
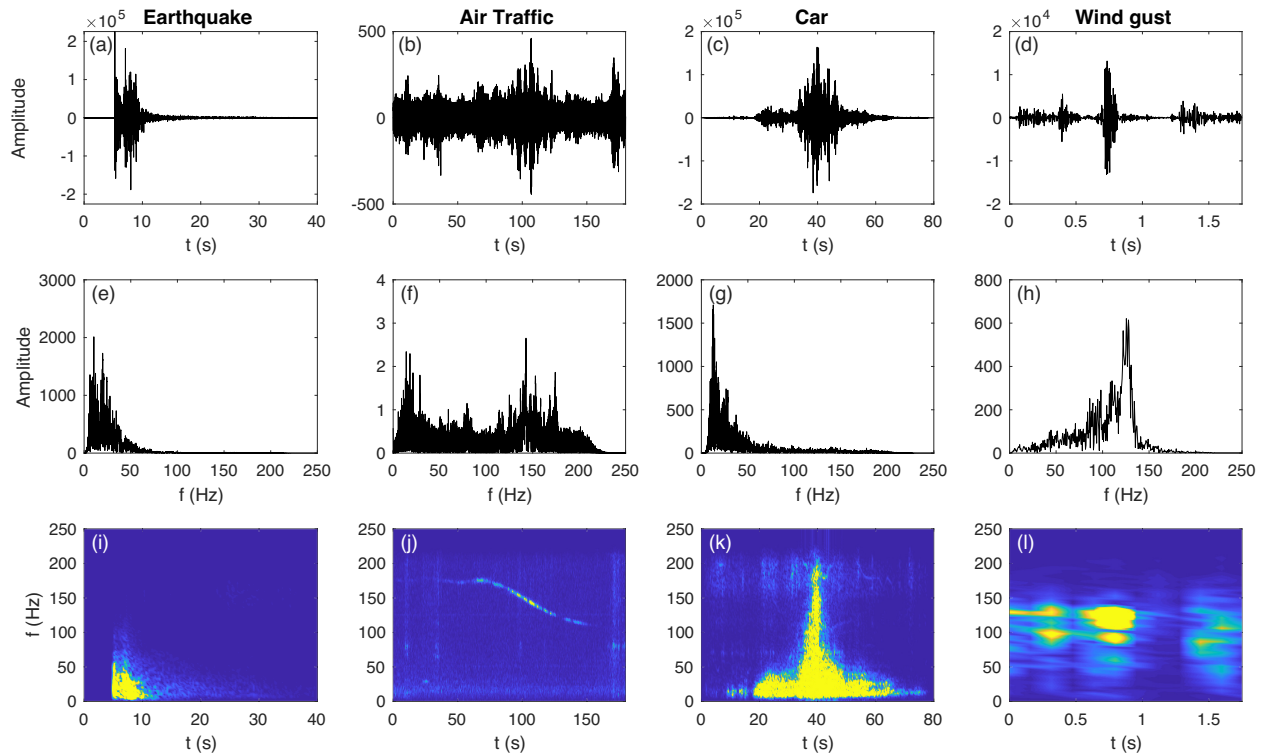
Figure 2. Waveforms in count generated by (a) a local M 2.29 earthquake, (b) an air-traffic event, (c) a car event and (d) a tree-shaking event driven by wind gust. (e)–(h) Corresponding spectral amplitudes of (a)–(d). (i)–(l) Corresponding spectrograms of (a)–(d).

distribution of the amplitude spectra. The method allows us to separate each 1-s interval of the continuous waveforms into RN signals, NRN with higher amplitude in time domain or some structure in the frequency domain, and mixed signals where neither the RN nor NRN features are dominant. The RN windows can be used to obtain better estimates of noise spectra, monitor hourly and longer changes of the background noise level, derive structural information and measure the degree to which the noise field is diffuse. The RN windows detected by our method can also be used as training data sets for machine learning analyses of phase picking and signal classification. The time windows containing NRN signals may be further analysed to select templates with specific signals such as those generated by air/car traffic, wind, earthquakes or tectonic tremor for detection and classification of such signals.

2 ANALYSIS

Seismograms are analysed using cross-correlations of waveform segments and a classification algorithm to extract parameters in time and frequency domains. Data recorded by station R3010 (Fig. 1) is used to illustrate the method of separating continuous seismic waveforms into RN, NRN and a mixture of the two. Station R3010 is in the centre of the array and has good data quality to demonstrate the method. Several iterations are implemented to update and stabilize the RN templates, mean noise spectra and separation results. The methodology is summarized in a workflow diagram (Fig. 3). We analyse the seismic records in short periods, $T_S = 1$ hr, and cut the waveforms into 1 s non-overlapping windows, T, with 500 samples each to stably estimate the amplitude spectra and cross-correlation coefficients. The window length T should be adjusted for other data with different sampling rates. The mean and linear trend are

removed and waveforms are high-pass filtered at $2/T$ (2 Hz for this study) to include at least two cycles in each window.

To investigate potential techniques and parameters that can separate RN and NRN signals, we first generate synthetic RN by Fourier transforming a low amplitude 500 s waveform recorded at night on sensor R3010, replacing the phase information with a uniform distribution of values from 0 to 2π, and performing an inverse Fourier transform on the amplitude spectrum and randomized phase (Figs 4a and b). Next, we extract a local M2.29 earthquake and an air-traffic event recorded by sensor R3010 to use as example NRN waveforms (Figs 4c and d). The waveforms are cut into 1 s non-overlapping windows to obtain 500 synthetic RN windows, 25 earthquake windows and 50 air-traffic event windows (the grey boxes in Figs 4c and d).

The maximum of the absolute value of cross-correlation coefficients (MACCs) is computed for all window pairs. Fig. 4(e) shows the distributions of MACCs for RN–RN windows (grey), RN–earthquake windows (blue) and RN–air-traffic windows (red). The MACCs of RN–RN windows have a well-defined interval between 0.09–0.35, while the distributions of RN–NRN systematically shift to intervals with lower median values demonstrating a simple statistical difference between the RN and NRN waveforms. The median of MACCs for these target windows is the first extracted time domain parameter to differentiate the RN and NRN signals. Fig. 5 shows that the distributions and standard deviations (STDs) for earthquake–earthquake windows (green), air-traffic–air-traffic windows (purple) and earthquake–air-traffic windows (brown) are spread over a broader range than RN–NRN windows. This indicates that the NRN windows contain both very similar and highly dissimilar waveforms. For example, the spectrogram of the air-traffic event (Fig. 2j) has a changing peak frequency due to the Doppler effect (Meng & Ben-Zion 2018a) that results in high MACCs for similar
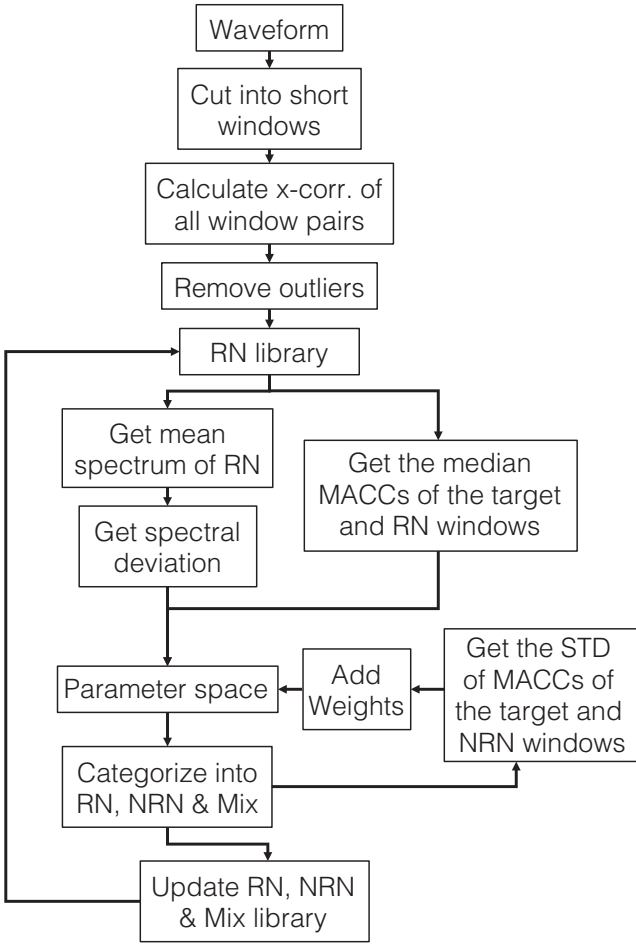
```
┌──────────────────────┐
│       Waveform       │
└──────────────────────┘
           │
┌──────────────────────┐
│  Cut into short      │
│  windows             │
└──────────────────────┘
           │
┌──────────────────────┐
│  Calculate x-corr. of│
│  all window pairs    │
└──────────────────────┘
           │
┌──────────────────────┐
│   Remove outliers    │
└──────────────────────┘
           │
┌──────────────────────┐
│     RN library       │
└──────────────────────┘
```

Figure 3. A workflow diagram summarizing the main steps of the method. The workflow includes multiple iterations to update and stabilize RN, not RN and the mixture.

window pairs. The STDs of the MACCs for NRN–NRN windows are significantly higher than the STDs of MACCs of RN–NRN windows (Fig. 5b). This feature is used as the second extracted time domain parameter to differentiate the RN and NRN waveforms.

To adopt the method to continuous seismic records, we assume that the RN is quasi-stationary (i.e. statistically time invariant) within a short period of 1 hr. This implies that the amplitude spectrum of RN only changes slightly within a short period T_S. Supporting Information Fig. S1 shows that the RN is generally quasi-stationary within time intervals of about 1 hr, although the noise level varies from day to night. In the frequency domain, the spectra of RN and NRN windows deviate significantly from each other. The L2 norm of the deviation of an amplitude spectrum of a given target window and the mean amplitude spectra of RN candidates provides a third extracted frequency domain parameter to differentiate between RN and NRN windows. These three parameters in time and frequency domains are used to classify time windows of RN.

The corresponding three parameters extracted in time and frequency domains are shown in Fig. 6 for the 1 hr waveform from sensor R3010 in Julian day 146 from 14:00 to 15:00 (local time). The data contain frequent impulsive signals throughout the hour, and the median MACCs of target and RN windows generally show a decrease during the strong bursts. In the following, we describe

how we iteratively build libraries of RN, NRN and mixture signals to separate the continuous waveforms into different classes.

The sets of RN, NRN and the undetermined mixtures are denoted by N, S and M, where

$$N = \{i : i \subseteq \text{indexes of all RN windows}\},$$
$$S = \{i : i \subseteq \text{indexes of all NRN}\},$$
$$M = \{i : i \subseteq \text{indexes of undetermined mixtrues}\}. \quad (1)$$

These three sets satisfy

$$N \cup S \cup M = U = \{i : i \text{ is an integer}, 1 \leq i \leq T_S/T\},$$
$$N \cap S = N \cap M = S \cap M = \emptyset, \quad (2)$$

where T_S/T is 3600 in this study. The element i is linked to the waveform of ith window and its MACCs with other windows. Since windows with a small RMS value are more likely to be RN, we first build an initial RN library N by taking the indexes of 1000 windows with the lowest RMS. The initial NRN library S contains indexes of the 1000 windows with larger RMS than the others. The initial library of the undetermined mixture, which is a transition from RN to NRN, is the complement of the N and S sets.

To avoid redundant computations, we calculate the MACCs for all pairs of the 3600 windows. The MACC of the ith and jth windows is defined as

$$C_{ij} = \max\left(\text{abs}\left(\frac{f_i(-t) * f_j(t)}{\text{STD}(f_i(t)) \cdot \text{STD}(f_j(t))}\right)\right), \quad (3)$$

where $f_i(t)$ and $f_j(t)$ are the waveforms of the ith and jth windows, $0 \leq t \leq T$, $\text{STD}(f_i(t))$ is the STD of all the data points of the ith window. For the ith window, we compute the median of the MACCs of the target window and RN windows:

$$C_i^{\text{MDN}} = \text{median}\,(C_{in})|_{n \in N, n \neq i}, \quad (4)$$

and the STD of MACCs of the target window and NRN windows:

$$C_i^{\text{STD}} = \text{STD}\,(C_{in})|_{n \in S, n \neq i}. \quad (5)$$

The outliers are removed by excluding windows with $C_i^{\text{STD}} \geq 1.1 * \text{median}(C_j^{\text{STD}})$, $C_i^{\text{MDN}} \geq 1.1 * \text{meidan}(C_j^{\text{MDN}})$ and $C_i^{\text{MDN}} \leq 0.9 * \text{median}(C_j^{\text{MDN}})$, where $1 \leq j \leq 3600$. The method is not sensitive to the initial RN library. All libraries are updated iteratively and stabilize quickly. The iteration stops when the total number of changed elements in libraries N and S from two consecutive iterations are < 0.5 per cent $\cdot T_S/T$. The C_i^{MDN} and C_i^{STD} for each window of the target waveform in Fig. 6(a) are presented in Figs 6(b) and (d), respectively. The waveform of each window is cosine tapered with a width of 5 per cent on each end to stably estimate the amplitude spectrum using Fourier transform. Fig. 6(c) shows the L2 norm deviations from the mean spectra of RN windows.

Results of the classification algorithm performed in the parameter space of the spectral deviation and C_i^{MDN} are shown in Fig 7. Each point in Fig. 7(a) represents a time window and the colours denote the normalized data points density ρ_i, which is computed by the total number of data points within a rectangle area centred at the target points with a domain size of $0.2 \cdot \text{STD}$ along each axis. The hourly RN is quasi-stationary so the results should cluster tightly (with high ρ_i) in the parameter space. Fig. 7(b) shows the same diagram but coloured by the C_i^{STD}. The windows with smaller C_i^{STD} are more likely to be RN. To use both ρ_i and C_i^{STD}, Fig. 7(c) uses a
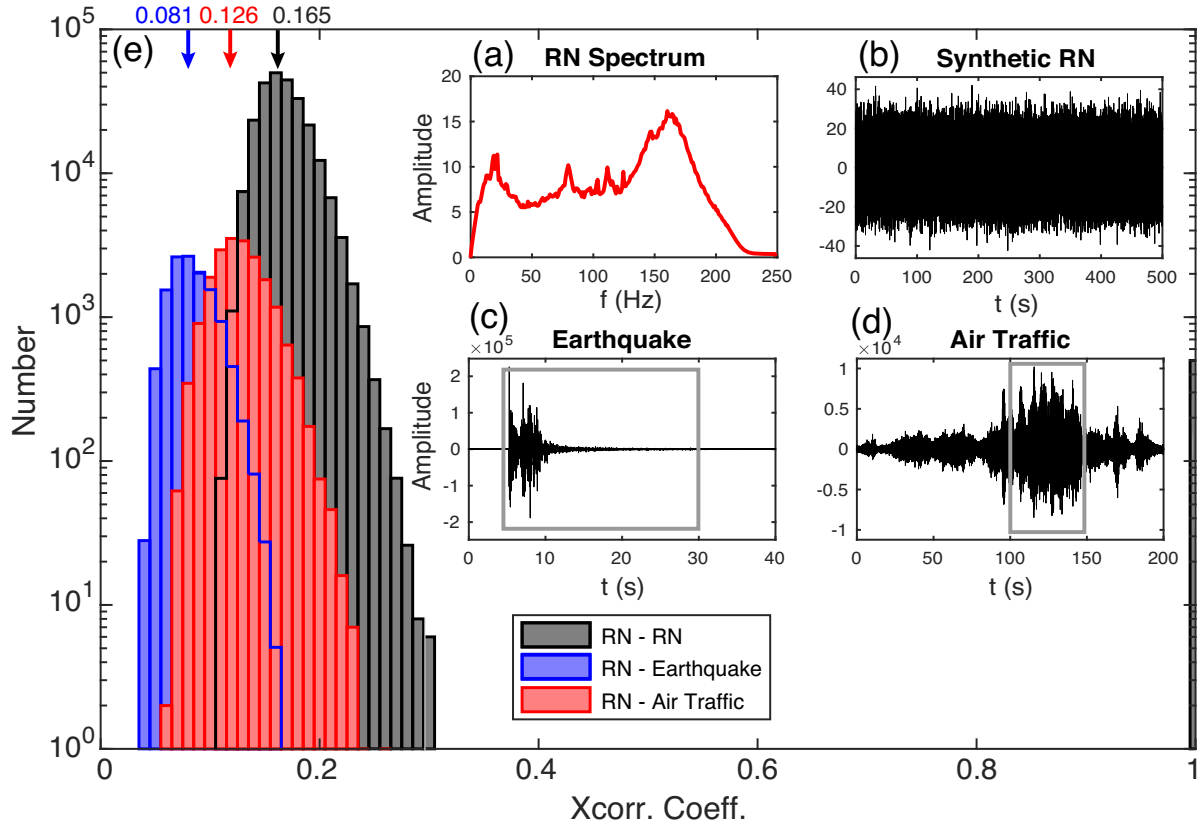
Figure 4. (a) Amplitude spectrum of a 500 s waveform at night with low amplitudes. (b) Synthetic RN waveform using the amplitude spectrum in (a). (c) Waveform of a local M 2.29 earthquake recorded by the geophone R3010 in the centre of the array. (d) Waveform of an air-traffic event recorded by the same geophone. (e) Histograms of the MACCs of RN–RN windows, RN–earthquake windows and RN–air-traffic windows. The median values for each distribution are shown above. The grey boxes show the NRN waveform segments used for the cross-correlation calculation.
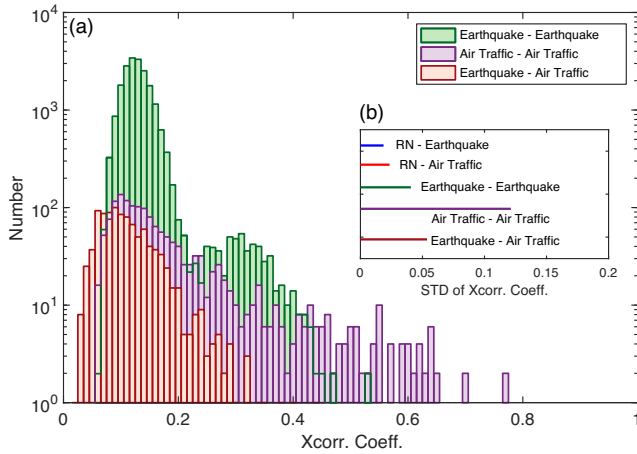

Figure 5. (a) Histograms of the MACCs of earthquake–earthquake windows, air-traffic–air-traffic windows and earthquake–air-traffic windows. (b) Standard deviations of histograms for different window pairs in (a) and Fig. 3(e). The standard deviation of the NRN–NRN windows is significantly larger than the standard deviation of RN–NRN windows.

weighted density as

$$\rho_i^W = \frac{\dfrac{\rho_i}{C_i^{\mathrm{STD}}}}{\max\left(\dfrac{\rho_j}{C_j^{\mathrm{STD}}}\right)}. \tag{6}$$

This combines C_i^{MDN}, C_i^{STD} and spectral deviations extracted from the time and frequency domains, and, therefore, is more sensitive to the difference between RN and NRN windows. Time windows with parameters in the tight cluster with high weighted density ρ_i^W (coloured by dark red in Fig. 7c) are identified as RN, time windows that deviate significantly from the RN cluster are identified as NRN and windows with values in between are identified as a mixture. The contours in Fig. 7(c) show the criteria to determine RN, NRN and the mixture. The corresponding sets are then updated as

$$N = \left\{i : \rho_i^W \geq 0.45\right\},$$
$$S = \left\{i : \rho_i^W \leq 0.15\right\},$$
$$M = \left\{i : 0.15 < \rho_i^W < 0.45\right\}. \tag{7}$$

With the updated RN library N, we recalculate the mean amplitude spectrum of RN and then update the spectral deviations for each window. The C_i^{MDN} and C_i^{STD} are updated correspondingly with the new libraries N and S. This process is repeated for a few iterations until the three libraries are stable. The solid lines in Supporting Information Fig. S2 show the mean amplitude spectra of RN, for example, waveforms (Fig. 6a) in different iterations. As a validation step, Fig. 7(d) presents the histograms of MACCs of RN–RN windows, NRN–NRN and mixed–mixed windows. The MACCs of RN–RN windows span a range of 0.09–0.35, while the MACCs of not NRN–NRN windows span a much broader range. These results are consistent with the histograms of the MACCs of the synthetic RN, earthquake and air traffic shown in Figs 4(e) and 5(a). The
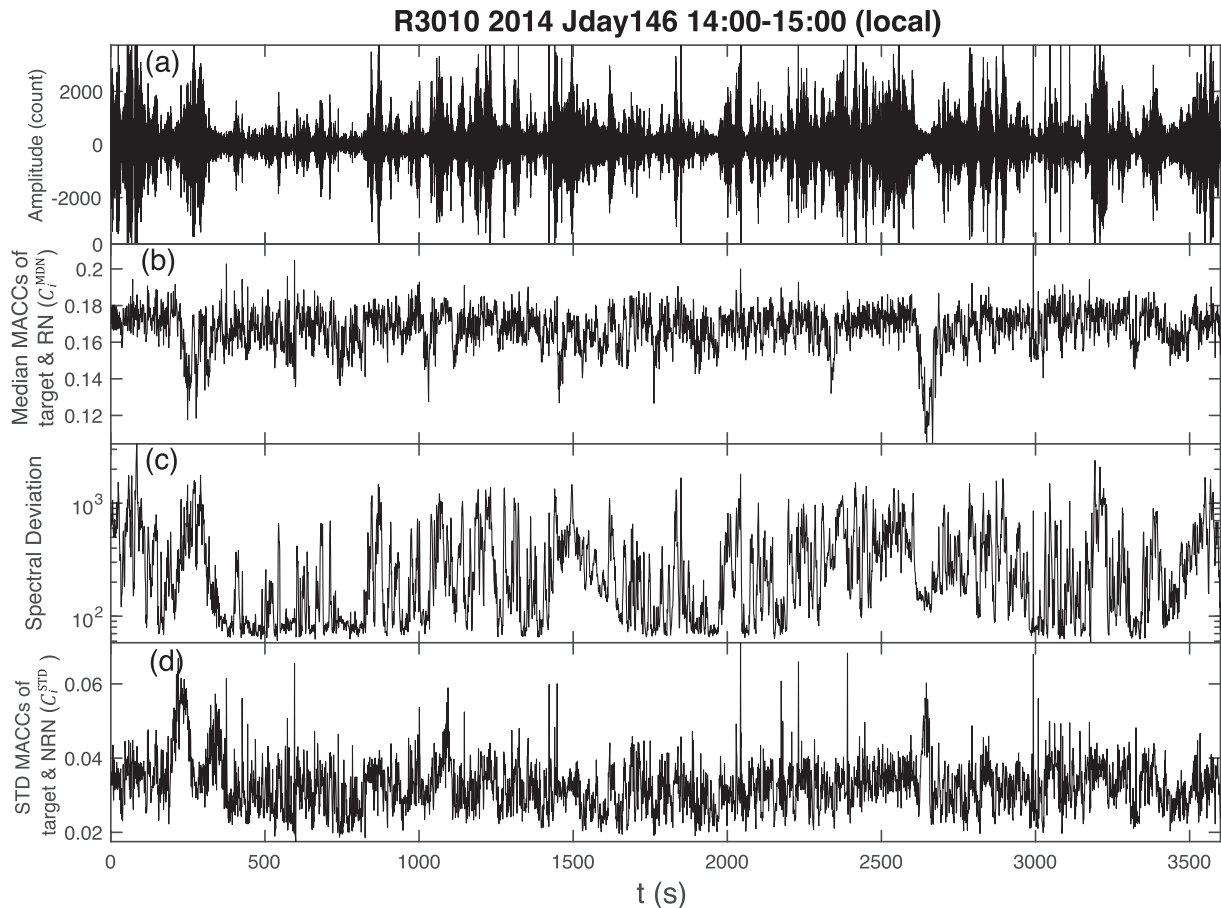
R3010 2014 Jday146 14:00-15:00 (local)

Figure 6. (a) Example raw waveform recorded by sensor R3010 on Julian day 146, 2014 from 14:00 to 15:00 (local time). (b) The median of MACCs of detected RN windows and target waveforms (C_i^{MDN}). (c) L2 norm deviations from the mean amplitude spectra of RN windows. (d) The STD of MACCs of NRN windows and target waveforms (C_i^{STD}).

thresholds in eq. (7) should be adjusted to minimize the difference of distribution, if the histograms of the MACCs of the detected RN (e.g. Fig. 7d) and the synthetic RN (e.g. Fig. 4e) are considerably different.

3 ANATOMY OF RECORDED WAVEFORMS

The waveforms and corresponding spectrograms of two example hours recorded by geophone R3010 on Julian day 146 from 14:00 to 15:00 (local time) coloured, red, black and blue for RN, NRN and mixed signals, respectively, are shown in Fig. 8. The earthquakes detected by Meng & Ben-Zion (2018a) are highlighted by the green vertical lines and occur in windows of NRN. The air-traffic events from 200 to 400 s and 1000 to 1200 s also fall in windows of NRN and have clear Doppler effect in the spectrogram (Fig. 8b). The RN windows have signals with lower amplitudes. The NRN windows have either larger amplitudes or clear signatures that stand out from the background in the spectrograms. However, the windows with lower amplitudes are not necessarily RN as exemplified by the time window from 2550 to 2650 s marked with the red horizontal bar in Fig. 8(a) and is identified as NRN due to three overtones in the spectrogram. The analyses for night-time hours show similar results with more time windows defined as RN. During the daytime RN and NRN waveforms are found to occupy 25.9 and 34.1 per cent of the time. During night-time, the RN fraction increases to 41.6 per cent

and the NRN fraction decreases to 25.2 per cent (Figs 8c and d). As an additional demonstration of detecting NRN, Supporting Information Fig. S3 shows the waveforms and spectrograms of an example hour (Julian day 146 from 5:00 to 6:00 local time), and a zoom-in view of an M 2.86 local earthquake with an epicentral distance of 176 km recorded by geophone R3010. The entire earthquake waveform from the first arrival to the end of the clear coda is identified as NRN in Supporting Information Fig. S3c, substantiating the sensitivity of the developed methodology.

Fig. 9 illustrates the discussed features in more detail using time windows of only 100 s. The daytime data show NRN with large amplitudes and strong energy around 150 Hz (Figs 9a and b). The periods of RN contain lower spectral amplitudes as expected and are not continuous for the 100 s duration presented. In general, the energy of NRN spans a large spectral range and the transition from RN to NRN is smooth. No hard threshold is determined to differentiate the RN from NRN, and the mixture label is used to classify the superposition of RN and weak NRN during the transitions.

Many NRN windows show significant energy around 150 Hz (Fig. 8) and the signals possibly originate from the atmosphere or very local sources since high-frequency seismic waves attenuate quickly within distances of 20–30 m (Meng & Ben-Zion 2018a). Aircrafts can generate high-frequency acoustic waves with little intrinsic attenuation in the air that couple to the ground and recorded by the seismometers (Meng & Ben-Zion 2018b). However, the Doppler effect of sources moving in the air is not observed in most
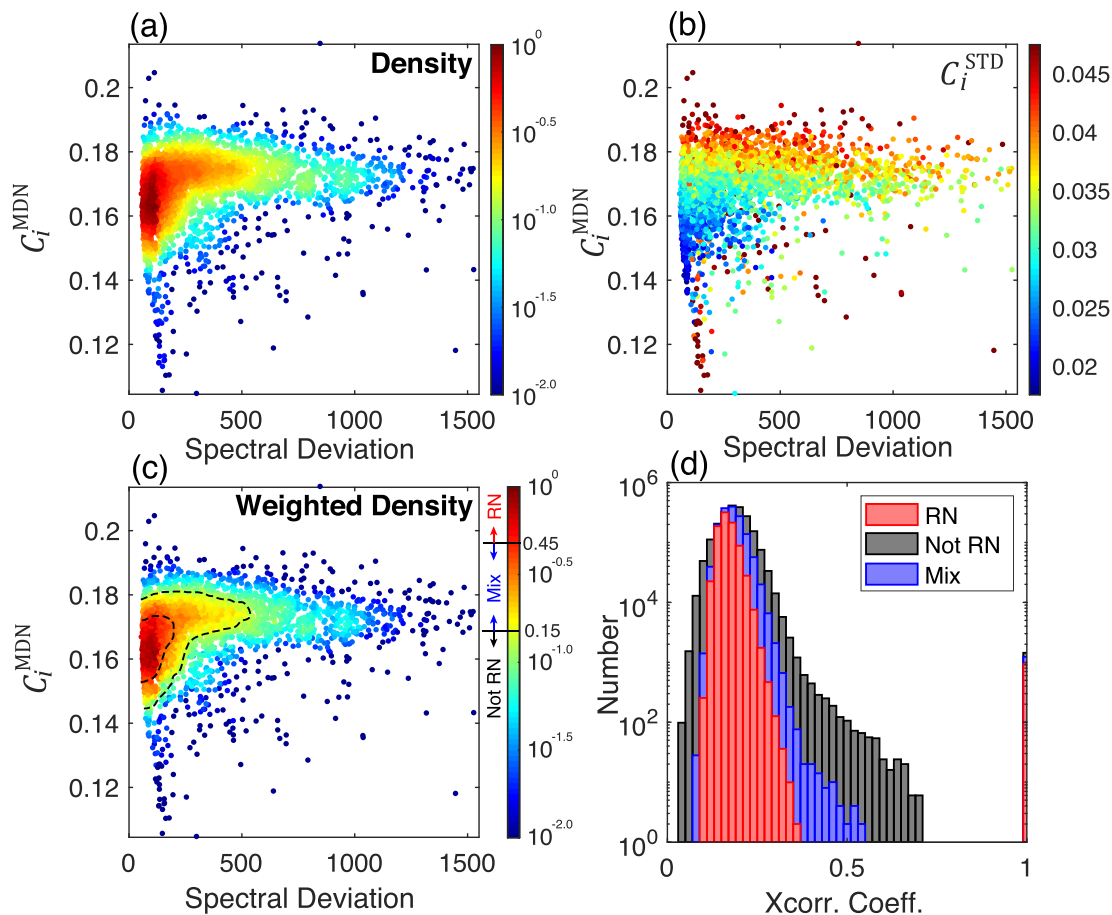
Figure 7. (a) A 2-D histogram of the example waveform (Fig. 6a) in the parameter space of C_i^{MDN} and spectral deviation. The colours denote the normalized data density. (b) Same as (a) but coloured by C_i^{STD}. (c) Same as (a) with colours denoting the normalized data density weighted by the inverse of the C_i^{STD}. The contour lines show the criteria to determine RN, not RN and mixture signals. (d) histograms of MACCs of RN–RN windows, NRN–NRN and mixed–mixed windows.

of the NRN windows indicating the vibrations are possibly generated by very local sources on the ground. The spectra for a geophone located near a tall tree have this high-frequency signal (Supporting Information Fig. S4) corroborating that some such sources originate from objects on the surface. Supporting Information Fig. S4 shows a clear propagation effect from geophone R2005 to its nearby seismometers. The corresponding spectra at R2005 has a peak around 125 Hz and the high-frequency energy attenuates within 20 m. The vibrations of the vegetation are likely driven by wind (Johnson *et al.* 2019) and the peak frequency at 125 Hz may be associated with instrumental effect or the coupling of the tree next to the sensor to the ground, with possible generation of small failure events in the top crust. At the SGB site, many of the NRN windows with significant energy around 150 Hz are likely wind-generated ground motions if air traffic is not observed in the spectrogram.

4 DISCUSSION

The developed methodology uses waveforms from a single seismic station at a time, to examine the statistical characteristics of data segments in time and frequency domains, and separate the records into RN, NRN and undetermined mixture. Seismic records consist of RN, which may be associated partly with classical diffuse noise, and NRN containing earthquakes and ground motions produced by atmospheric and anthropogenic activities such as air

traffic, lightning, cars, trains and wind shaking of structures, vegetation and other obstacles above the surface. Small earthquakes can be detected by multiple techniques such as template matching or beamforming. Unlike tectonic activities, NRN produced by atmospheric and anthropogenic activities can depend strongly on various aspects of the site including the distribution and types of obstacles above the ground and attenuation in the very shallow crust. NRN is, therefore, not a distinct repeating signal, making it unfavourable for similarity-based detection techniques.

The presented waveform classification method is stable and not sensitive to the initial criteria of building and updating the libraries three classes of waveforms. The method jointly examines the amplitude spectra of moving time windows and the MACCs with RN and NRN templates from hourly waveforms. The hourly RN is quasi-stationary (Supporting Information Fig. S1) and the results group in the parameter space of the median MACC and L2 norm deviations from the mean spectra of RN candidates (Fig. 7). Time windows with parameters in this tight cluster are identified as RN, windows that deviate significantly from the RN cluster are identified as NRN and windows with values in between RN and NRN are identified as undetermined mixtures. Several iterations are performed on each hourly data to update the RN templates and mean noise spectra until the libraries stabilize. We tested an initial RN library by including the indexes of 500 windows with the smallest RMS without excluding outliers and the results quickly converge in four iterations and
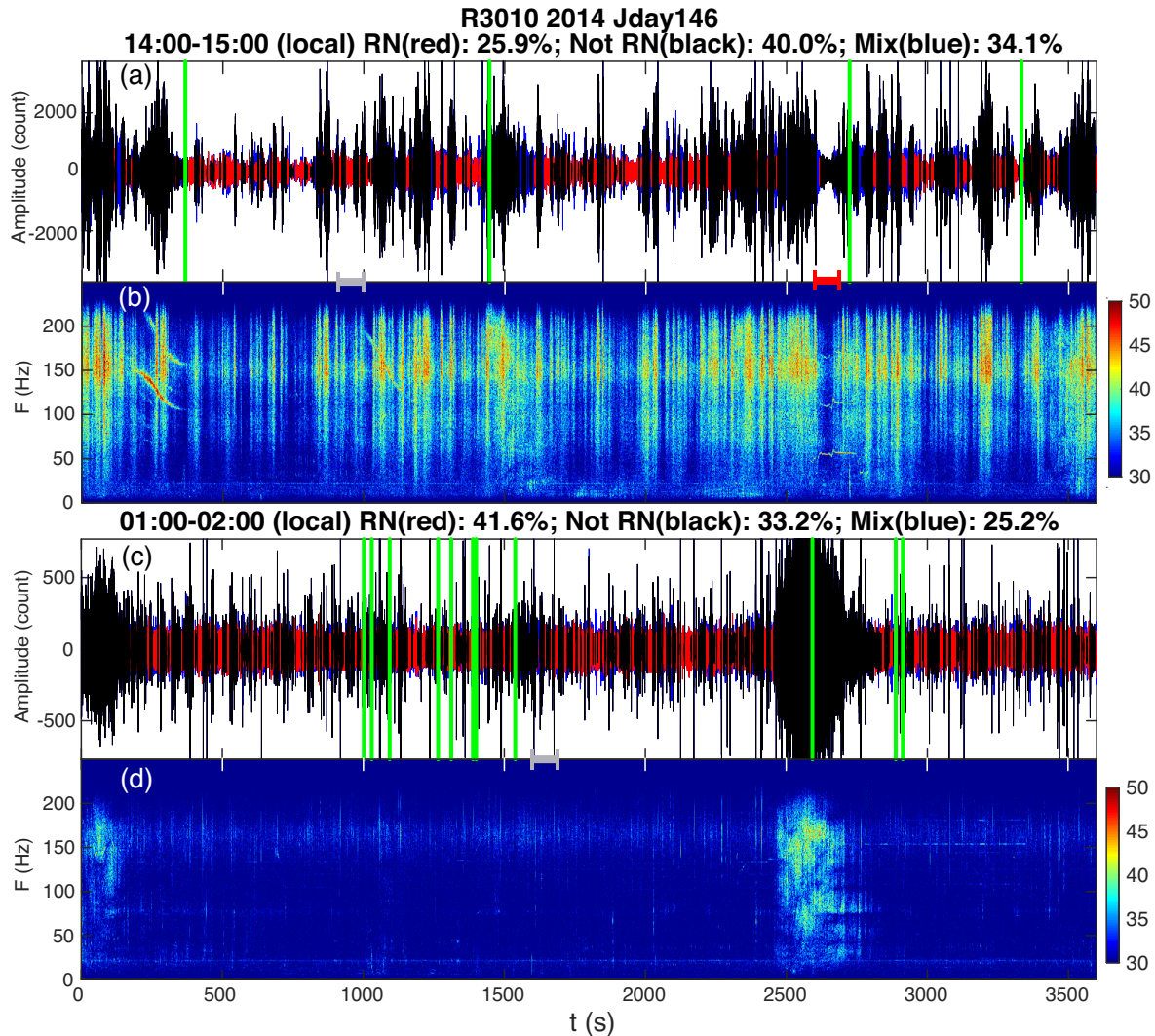
R3010 2014 Jday146
14:00-15:00 (local) RN(red): 25.9%; Not RN(black): 40.0%; Mix(blue): 34.1%

Figure 8. (a) Example raw waveform recorded by R3010 on Julian day 146, 2014 from 14:00 to 15:00 (local time). The RN, NRN and mixture windows are coloured in red, black and blue. (b) Corresponding spectrogram of the waveform in (a) with colours denoting power (dB). The data marked by the red horizontal bar between (a) and (b) illustrate that windows with lower amplitudes are not necessarily RN, as evidenced by the three overtones in the spectrogram. (c) Example raw waveform recorded by sensor R3010 at day 146 from 01:00 to 02:00 (local time). (d) Corresponding spectrogram of waveform in (c). The green vertical lines in (a) and (c) mark the small earthquakes detected by Meng & Ben-Zion (2018b).

are almost identical to those shown in Fig. 8. The results are not sensitive to the criteria in eq. (7) and perturbing the lower and upper thresholds by 0.1 produces changes <1 per cent in the fraction of time covered by RN and NRN. The results are also not sensitive to the domain size used to compute the data point densities in Fig. 7, which is $0.2 \cdot STD$ along each axis. Variations of this choice from 0.1 to $0.4 \cdot STD$ produce no more than 0.6 per cent changes in the fraction of time covered by RN and NRN (Supporting Information Fig. S5).

By detecting RN windows for each hour, we are able to estimate the hourly mean spectral amplitude of the RN (Fig. 10). The amplitude spectra change gradually from 1 hr to the next, supporting the notion that the RN is quasi-stationary. Since there are more atmospheric and anthropogenic activities in daytime hours, the energy of scattered wavefield is also higher. Peaks at 20, 80 and 150 Hz are consistently seen in the spectrograms, indicating various local resonators including vegetation, ground structures and sedimentary layers. The hourly mean spectral peaks at 125 Hz in Supporting

Information Fig. S6 and 110 Hz peak in Supporting Information Fig. S7 may be associated, respectively, with instrumental effects and/or coupling of the tree and antenna tower near sensors to the ground. Fig. 11 summarizes daily records by the hourly percentage of RN, NRN and the mixture for (a) Julian day 146 for all analysed geophones and (b) all available days for a single geophone. The curves in Fig. 11(b) are smoother and the uncertainties are slightly larger than those in Fig. 11(a), but the overall structures are consistent. The percentage of the day occupied by NRN decreases significantly from ~50 per cent in daytime to ~30 per cent at night. The results are consistent with the occurrence of the air-traffic events (Meng & Ben-Zion 2018b) and wind-driven vibration of vegetation and local structures with diurnal fluctuations that increase at the SGB site by mid-day local time and subside during the evening hours (Johnson *et al.* 2019).

It is important to note that since ~35 per cent of the day is covered by RN on average, 7 per cent of the day is occupied by air-traffic events and 6 per cent by wind-related events, we can now classify
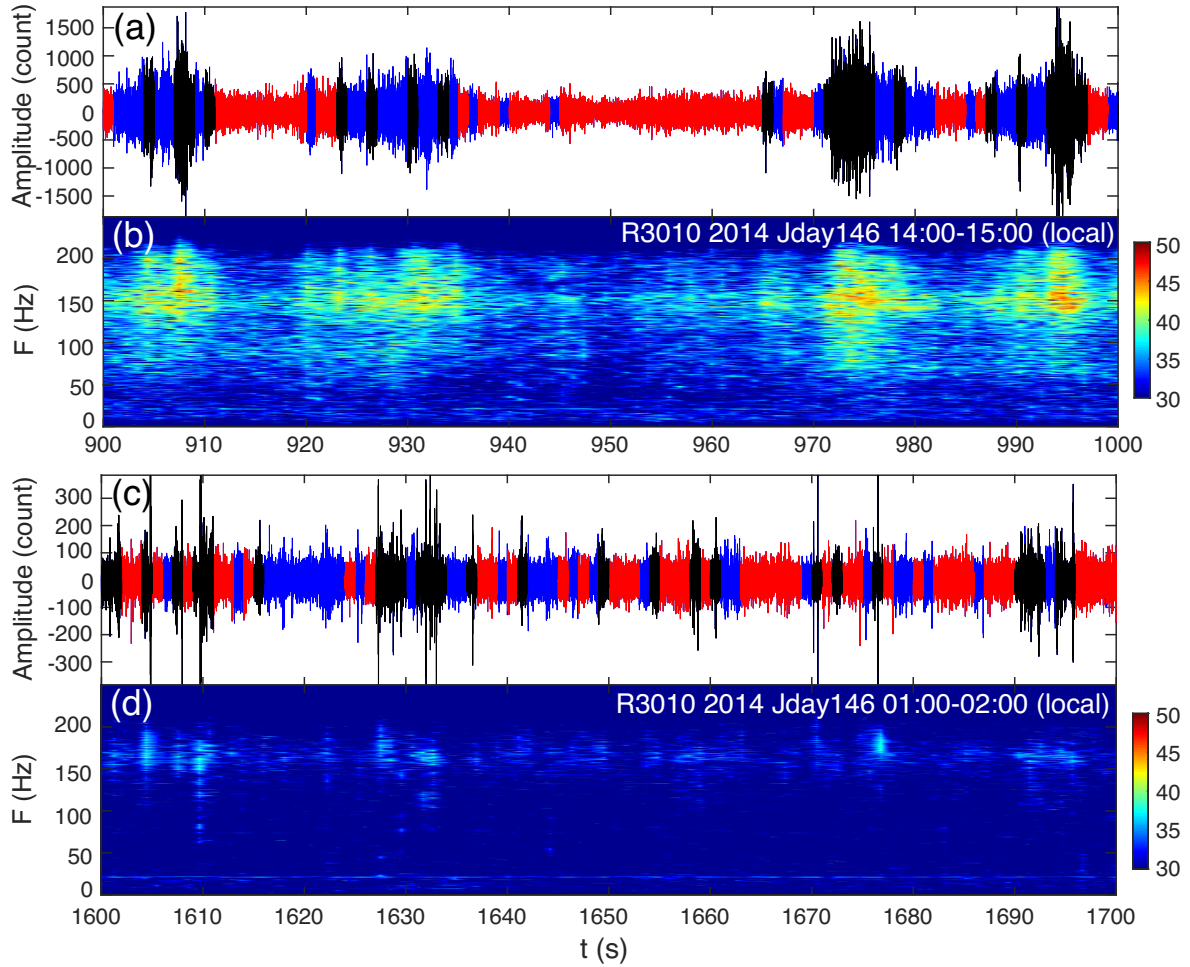
Figure 9. (a) and (b) Zoom-in views of two 100 s windows highlighted by the grey horizontal bars in Figs 8(a) and (b). (c) and (d) Corresponding spectrograms of (a) and (b). The RN, NRN and the mixture waveforms are coloured in red, black and blue, respectively.
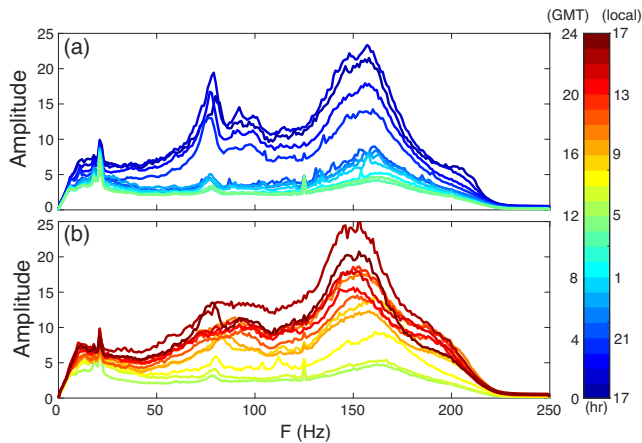
Figure 10. Hourly mean spectral amplitudes of the detected RN of the waveform recorded by sensor R3010 during Julian day 146, 2014, with colours denoting the hours. (a) Results for 0:00–12:00 GMT (17:00—5:00 local time). (b) Results for 12:00–24:00 GMT (5:00–17:00 local time).

∼48 per cent the continuous waveforms at the SGB site. By excluding the NRN and mixture windows, the RN windows can be used to stably estimate better and use for imaging and monitoring studies noise spectra with windows excluding NRN signals. The method
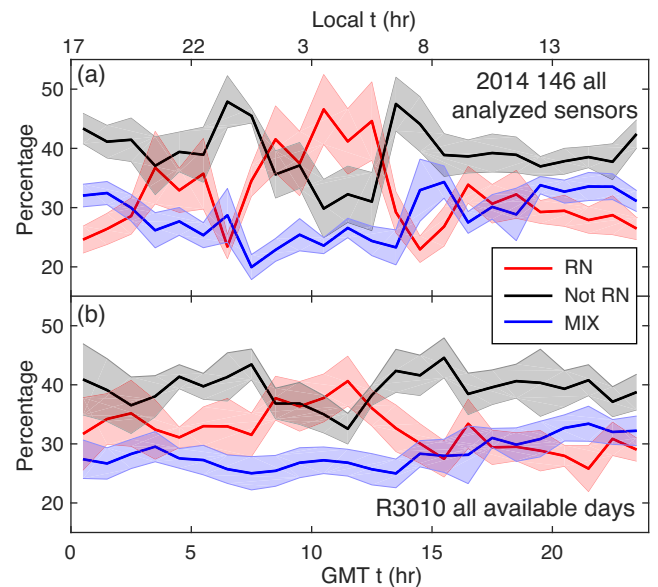
Figure 11. (a) Hourly mean percentage of RN, NRN and mixed signals for all analyzed geophones in Fig. 1(b) for Julian day 146, 2014. (b) Hourly mean percentage of RN, NRN and mixed signals for all available days at geophone R3010. The shaded regions show the corresponding standard deviations.

can be generalized to three-component records by redefining C_i^{MDN}, C_i^{STD} and spectral deviation, and using the results of the different components for deriving refined structural information (e.g. via H/V ratios). Estimating the extent to which wavefields in different time windows are diffused requires additional analyses (e.g. Margerin *et al.* 2009; Sens-Schonfelder *et al.* 2015; Liu & Ben-Zion 2016). This topic will be addressed in a follow-up study. The NRN covers on average 40 per cent of the day and analysing these signals further will provide improved understating of common sources producing the observed ongoing ground motion. The RN windows detected by our method can also provide a good training data set for studies using machine learning for phase picking and signal classification.

ACKNOWLEDGEMENTS

REFERENCES

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, *Geophys. J. Int.,* **169**(3), 1239–1260.

Ben-Zion, Y. *et al.*, 2015. Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone, *Geophys. J. Int.,* **202**, 370–380, .

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.

Eibl, E.P., Lokmer, I., Bean, C.J., Akerlie, E. & Vogfjörd, K.S., 2015. Helicopter vs. volcanic tremor: characteristic features of seismic harmonic tremor on volcanoes, *J. Volcanol. Geotherm. Res.,* **304**, 108–117.

Gerstoft, P. & Tanimoto, T., 2007. A year of microseisms in southern California, *Geophys. Res. Lett.,* **34**, L20304.

Gradon, C., Moreau, L., Roux, P. & Ben Zion, Y., 2019. Analysis of surface and seismic sources in dense array data with match field processing and Markov Chain Monte Carlo sampling, *Geophys. J. Int.,* **218**, 1044–1056.

Hillers, G., Ben-Zion, Y., Landès, M. & Campillo, M., 2013. Interaction of microseisms with crustal heterogeneity: a case study from the San Jacinto fault zone area, *Geochem. Geophys. Geosyst.,* **14**, 2063–2545.

Hillers, G., Roux, P., Campillo, M. & Ben-Zion, Y., 2016. Focal spot imaging based on zero lag cross correlation amplitude fields: application to dense array data at the San Jacinto fault zone, *J. geophys. Res.,* **121**, 8048–8067.

Inbal, A., Cristea-Platon, T., Ampuero, J.P., Hillers, G., Agnew, D. & Hough, S.E., 2018. Sources of long-range anthropogenic noise in southern California and implications for tectonic tremor detection, *Bull. seism. Soc. Am.,* **108**(6), 3511–3527.

Johnson, C.W., Meng, H., Vernon, F. & Ben-Zion, Y., 2019. Characteristics of ground motion generated by wind interaction with trees, structures, and other surface obstacles, *J. geophys. Res.*.

Liu, X. & Ben-Zion, Y., 2016. Estimating correlations of neighboring frequencies in ambient seismic noise, *Geophys. J. Int.,* **206**, 1065–1075.

Margerin, L., Campillo, M., Van Tiggelen, B.A. & Hennino, R., 2009. Energy partition of seismic coda waves in layered media: theory and application to Pinyon Flats Observatory, *Geophys. J. Int,* **177**(2), 571–585.

Meng, H. & Ben-Zion, Y., 2018a. Characteristics of airplanes and helicopters recorded by a dense seismic array near Anza California, *J. geophys. Res.,* **123**, 4783–4797.

Meng, H. & Ben-Zion, Y., 2018b. Detection of small earthquakes with dense array data: example from the San Jacinto fault zone, southern California, *Geophys. J. Int.,* **212**(1), 442–457.

Mordret, A., Roux, P., Boué, P. & Ben-Zion, Y., 2019. Shallow 3-D structure of the San Jacinto Fault zone revealed from ambient noise imaging with a dense seismic array, *Geophys. J. Int.,* **216**, 896–905.

Neuffer, T. & Kremers, S., 2017. How wind turbines affect the performance of seismic monitoring stations and networks, *Geophys. J. Int.,* **211**, 1319–1327.

Qin, L., Ben-Zion, Y., Qiu, H., Share, P.E., Ross, Z.E. & Vernon, F.L., 2018. Internal structure of the San Jacinto fault zone in the trifurcation area southeast of Anza, California, from data of dense seismic arrays, *Geophys. J. Int.,* **213**, 98–114.

Ross, Z.E., Hauksson, E. & Ben-Zion, Y., 2017. Abundant off-fault seismicity and orthogonal structures in the San Jacinto fault zone, *Sci. Adv.,* **3**, e1601946.

Salvermoser, J., Hadziioannou, C. & Stähler, S., 2015. Structural monitoring of a highway bridge using passive noise recordings from street traffic, *J. acoust. Soc. Am.,* **138**, 3864–3872.

Sánchez-Sesma, F.J., Pérez-Ruiz, J.A., Luzón, F., Campillo, M. & Rodríguez-Castellanos, A. 2008. Diffuse fields in dynamic elasticity. *Wave Motion,* **45**, 641–654.

Sens-Schönfelder, C., Snieder, R. & Stähler, S.C., 2015. The lack of equipartitioning in global body wave coda. *Geophys. Res. Lett.,* **42**(18), 7483–7489.

Shapiro, N.M., Campillo, M., Margerin, L., Singh, S.K., Kostoglodov, V. & Pacheco, J., 2000. The energy partitioning and the diffusive character of the seismic coda, *Bull. seism. Soc. Am.,* **90**, 655–665.

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.

Shelly, D.R., Beroza, G.C. & Ide, S., 2007. Non-volcanic tremor and low-frequency earthquake swarms, *Nature,* **446**(7133), 305–307.

Tanimoto, T. & Valovcin, A., 2015. Stochastic excitation of seismic waves by a hurricane, *J. Geophy. Res.,* **120**(11), 7713–7728.

Vernon, F., Ben-Zion, Y. & Hollis, D., 2014. Sage Brush Flats Nodal Experiment. International Federation of Digital Seismograph Networks. Other/Seismic Network.

Weaver, R.L., 1982. On diffuse waves in solid media, *J. acoust. Soc. Am.,* **71**(6), 1608.

Zigone, D., Ben-Zion, Y., Campillo, M. & Roux, P., 2015. Seismic tomography of the southern California plate boundary region from noise-based Rayleigh and Love waves, *Pure appl. Geophys.,* **172**, 1007–1032.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. RMS amplitude of all detected RN windows at station R3010 in Julian day 146, 2014. The hourly RN is quasi-stationary.
Figure S2. Mean amplitude spectra of the RN templates in different iterations for the two example hours shown in Figs 8(a) and (c). The spectra converge quickly in four iterations.
Figure S3. (a) Example raw waveform recorded by sensor R3010 in Julian day 146, 2014 from 5:00 to 6:00 (local time). The RN, NRN and the mixture windows are coloured in red, black and blue, respectively. (b) Corresponding spectrogram of waveform in (a). The colours denote power (dB). (c) and (d) Zoom-in views of two 200 s windows highlighted by the grey horizontal bars in (a) and (b). The green vertical lines in (a) and (c) highlight the small earthquakes detected by Meng & Ben-Zion (2018b).

Figure S4. Waveforms generated by vibration of a tree next to station R2005. (a) The self-normalized waveforms recorded by R2005 and 10 nearby stations in a linear profile perpendicular to the fault. The sensor spacing is about 10 m. (b) Corresponding self-normalized amplitude spectra.

Figure S5. Changes in the fraction of time covered by RN, NRN and mixture caused by the variations of domain size used to compute the data points densities in Fig. 7.

Figure S6. Hourly mean spectral amplitude of detected RN in the waveform recorded by geophone R2005 underneath a tree in Julian day 146, 2014, with colours denoting the hours. (a) Results for 0:00–12:00 GMT (17:00–5:00 local time). (b) Results for 12:00–24:00 GMT (5:00–17:00 local time).

Figure S7. Hourly mean spectral amplitude of the detected RN of the waveform recorded by R0506 beneath an antenna tower for each hour on Julian day 146, 2014, with colours denoting the hours. (a) Results for 0:00–12:00 GMT (17:00–5:00 local time). (b) Results for 12:00–24:00 GMT (5:00–17:00 local time).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.