Lateral Variations Across the Southern San Andreas Fault Zone Revealed from Analysis of Traffic Signals at a Dense Seismic Array

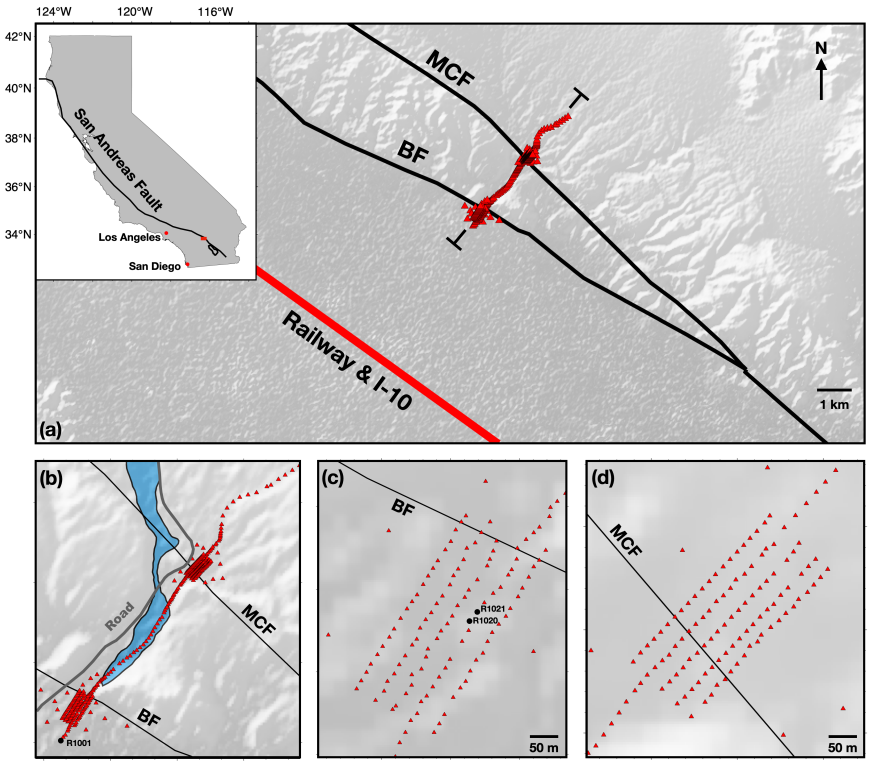
Hao Zhang[1], Haoran Meng[2], and Yehuda Ben-Zion[1]

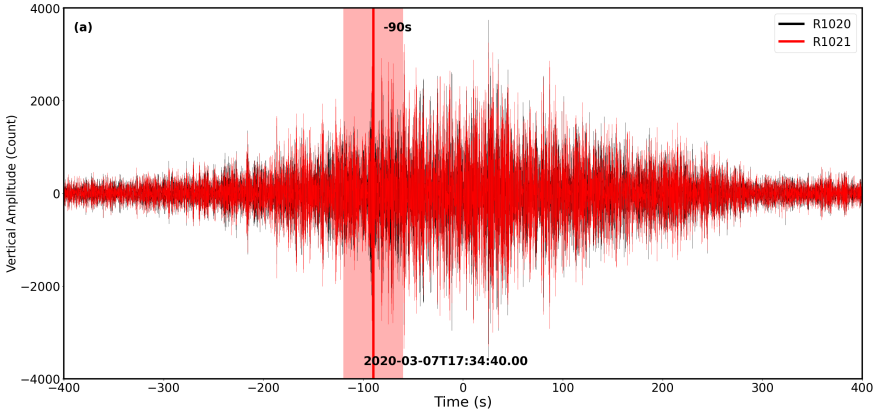
[1]University of Southern California
[2]Southern University of Science and Technology

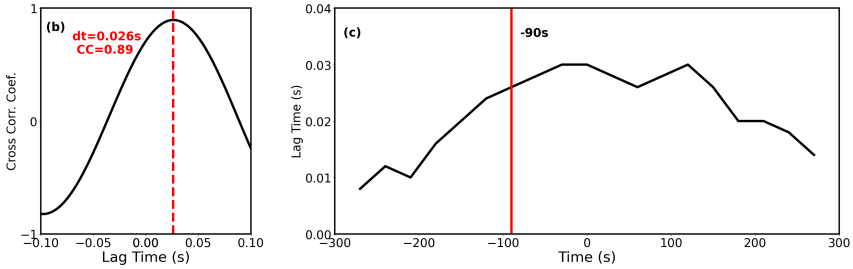
March 26, 2023

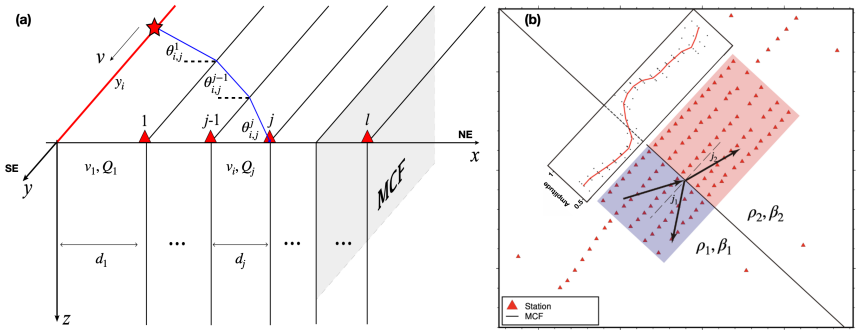
Abstract

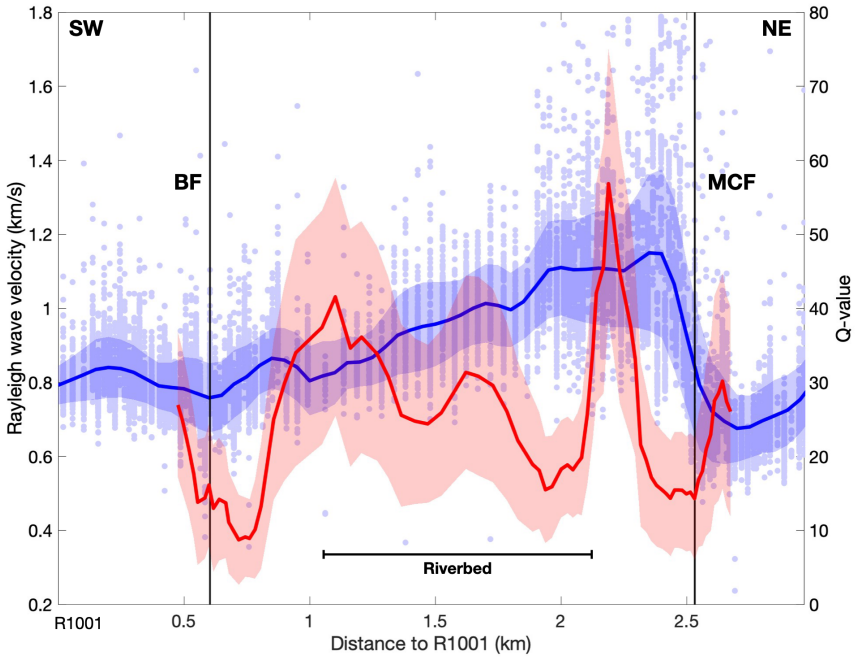
We image the shallow seismic structure across the Southern San Andreas Fault (SSAF) using signals from freight trains and trucks recorded by a dense nodal array, with a linear component perpendicular to SSAF and two 2D subarrays centered on the Banning Fault (BF) and Mission Creek Fault (MCF). Particle motion analysis in the frequency band 2-5 Hz shows that the examined traffic sources can be approximated as moving point sources that primarily induce Rayleigh waves. Using several techniques, we resolve strong lateral variations of Rayleigh wave velocities across the SSAF, including 35% velocity reduction across MCF towards the northeast. Additionally, we derive Q-values and find strong attenuation around the BF and MCF. We further resolve 10% mass density reduction and 45% shear modulus decrease across the MCF. These findings suggest that the MCF is the main strand of the SSAF in the area with important implications for seismic hazard assessments.











Lateral Variations Across the Southern San Andreas Fault Zone Revealed from Analysis of Traffic Signals at a Dense Seismic Array

Hao Zhang[1], Haoran Meng[2], and Yehuda Ben-Zion[1,3]

[1]University of Southern California, Los Angeles, California, USA.
[2]Southern University of Science and Technology, Shenzhen, China.
[3]Southern California Earthquake Center, Los Angeles, California, USA.

Key Points:

- We detect frequent seismic signals from rail and road traffic in a dense array across the southern San Andreas fault zone.
- We use the traffic signals to image shallow structural properties across the Banning and Mission Creek fault strands.
- The resolved velocity and density contrasts across the Mission Creek fault suggest it is the main active strand of the SSAF in the area.

Corresponding author: Haoran Meng, menghr@sustech.edu.cn

Abstract

We image the shallow seismic structure across the Southern San Andreas Fault (SSAF) using signals from freight trains and trucks recorded by a dense nodal array, with a linear component perpendicular to SSAF and two 2D subarrays centered on the Banning Fault (BF) and Mission Creek Fault (MCF). Particle motion analysis in the frequency band 2-5 Hz shows that the examined traffic sources can be approximated as moving point sources that primarily induce Rayleigh waves. Using several techniques, we resolve strong lateral variations of Rayleigh wave velocities across the SSAF, including 35% velocity reduction across MCF towards the northeast. Additionally, we derive Q-values and find strong attenuation around the BF and MCF. We further resolve 10% mass density reduction and 45% shear modulus decrease across the MCF. These findings suggest that the MCF is the main strand of the SSAF in the area with important implications for seismic hazard assessments.

Plain Language Summary

Imaging the internal structure of fault zones is essential for understanding earthquake properties and processes. Here we utilize seismic data generated by trains and trucks in the Coachella valley and recorded by a dense seismic array to image the subsurface structure of two main strands of the Southern San Andreas Fault (SSAF). Several types of analyses allow us to resolve seismic velocities, attenuation coefficients, and mass density across the entire San Andreas Fault zone. The results show a clear contrast in physical properties across the Mission Creek strand of the SSAF, highlighting the presence of a bimaterial fault interface and suggesting that it is the main strand of SSAF. The research opens up possibilities for using common rail and road traffic signals to derive high resolution imaging results of subsurface seismic properties at other locations.

1 Introduction

Earthquake fault zones have geometrical and material heterogeneities that reflect their past history and can strongly affect future earthquakes and seismic motion generated by the faults (e.g., Stierman, 1984; Ben-Zion, 2008). The Southern San Andreas Fault (SSAF) has not experienced a large earthquake over the past 300 years and is considered to pose a significant seismic hazard (Field et al., 2014). Various studies attempted to derive seismic velocity models for the SSAF (e.g., Shaw et al., 2015; Ajala et al., 2019), but they lack resolution on internal fault zone components such as sharp bimaterial interfaces and damage

zones. Imaging these features, as well as properties of the top structure which may be compared with geological information, require dense seismic arrays that cross the fault zone of interest (e.g., Ben-Zion et al., 2015; Share et al., 2020; Qiu et al., 2021).

The SSAF in the Coachella valley has two major strands - the Mission Crack Faults (MCF) and the Banning Fault (BF) - and it is a matter of debate which one is the primary active strand of the SSAF (Jones et al., 2008; Blisniuk et al., 2021). In the present study, we use seismic data recorded by a dense temporary nodal array across the BF and MCF near the Thousand Palms Oasis Preserve in the Coachella Valley (Figure 1) to image the subsurface properties of the SSAF in the area. During the ∼1 month deployment, only a small number of local earthquakes occurred near the array (Share et al., 2022), necessitating the use of other signals for detailed seismic imaging. Among such signals, seismic waves generated by cars and trains are used increasingly in imaging and monitoring studies due to their high reproducibility and simple source features (Fuchs et al., 2017; Meng et al., 2021; Pinzon-Rincon et al., 2021; Jiang et al., 2022; Rezaeifar et al., 2023). The reproducibility of traffic-generated signals also makes them suitable for monitoring temporal changes of seismic velocities (Brenguier et al., 2019; Sheng et al., 2022).

In the following sections, we use seismic signals generated by vehicle traffic in the Coachella valley to image the shallow internal structure of the SSAF below the dense nodal array. Utilizing waveforms generated by freight trains and trucks with high signal-to-noise ratio (SNR), we derive Rayleigh wave velocities and Q-values across both the BF and MCF, as well as the contrast of mass density across the MCF. While the analyses indicate significant contrasts of material properties across the MCF, only minor variations are observed across the BF. The results support the view that the MCF is more likely the main strand of the SSAF in the area (Blisniuk et al., 2021).

2 Methods

2.1 Seismic Network and Data

This study utilizes data from 322 nodes deployed across the BF and MCF strands of the SSAF in the Coachella Valley, California (Figure 1). The data was collected continuously from March 3 to April 13, 2020, and recorded using Zland 3-component 5 Hz nodes with a sampling rate of 500 Hz (Share et al., 2020). The array had a quasilinear profile with more than 100 nodes that crossed both the BF and MCF, perpendicular to their surface

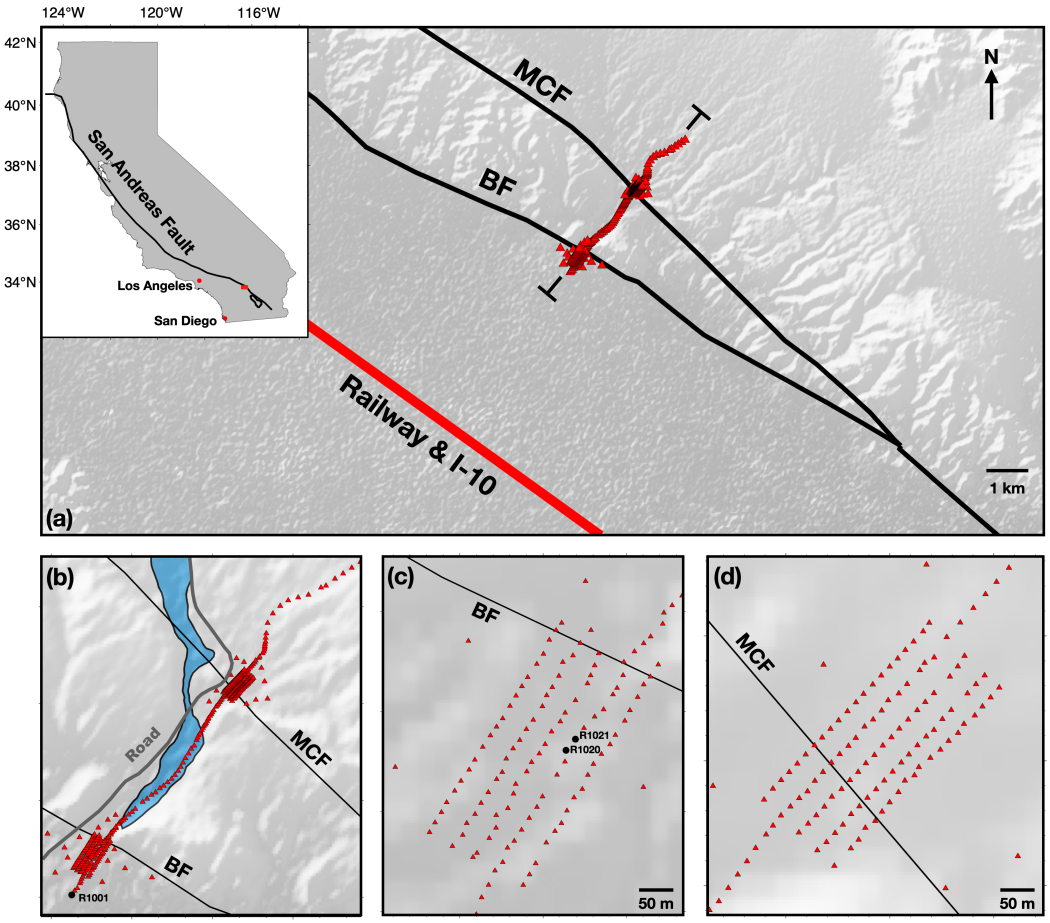


Figure 1. **(a)** Location map of the Coachella Valley showing the dense seismic nodal array employed in this study (red triangles), fault strands (black lines), highway I-10 and adjacent railway (bold red line). The inset in the top-left corner provides a regional view of California with the San Andreas Fault, Los Angeles, San Diego and the study area (red rectangle). **(b)** Location of a local road (gray curve) along the array and a riverbed between the two strands (blue shading). **(c,d)** Zoomed-in views of two 2D sub-arrays centered on the Banning Fault (BF) and Mission Creek Fault (MCF), respectively. Example nodes R1020 and R1021 are denoted by black dots in (c).

traces, with internode spacings that varied from 15 m near the fault traces to 50 m away from them. The linear array connected two 2D subarrays, each containing more than 100 nodes with a 20 m interstation spacing and a 1 km aperture, centered on the BF and MCF. During the data acquisition period, the nodal array recorded about 100 local earthquakes and prevalent seismic signals generated by traffic.

The majority of the traffic-induced signals were from local roads near the array and showed strong energy in the frequency band of 5-35 Hz (Figure S1). However, these signals were usually generated by small vehicles and could only be recorded by a small subset of the array due to geometrical spreading and strong attenuation effects. To image the subsurface structure beneath the entire array, we utilize here much stronger signals from truck traffic on the Interstate 10 (I-10) highway, located approximately 5 km southwest of the BF, and from freight trains on the railway adjacent to the I-10. The highway and railway (jointly referred to as the route) run nearly parallel to the BF and MCF, as shown in Figure 1a. To separate the target signals of freight trains and trucks (together referred to as vehicles) from other anthropogenic and environmental sources, including local cars, air-traffic events, and wind-induced signals (Meng et al., 2019; Johnson et al., 2019; Díaz et al., 2022), we use a bandpass filter and only analyze data in the frequency range of 2 to 5 Hz. A typical vehicle signal has a strong spindle-like symmetry both in the time series and spectrogram, with a duration of hundreds of seconds (Figure 2a and Figure S1). To ensure robustness, 29 events with $SNR > 5$ are manually selected and analyzed further in this study.

2.2 Tracking the motion of vehicles

The recorded traffic-induced signals generally last for several minutes, during which vehicles can move a few kilometers. As the sizes of vehicles are much smaller than the wavelengths (200-500 m) of interest and source-receiver distances, we approximate vehicles as moving point sources with a uniform radiation pattern according to the far-field approximation. We neglect the Doppler effects as the speed of vehicles ($\leq$ 35 m/s) is much less than the seismic wave velocity ($\geq$ 700 m/s). This is confirmed in the following analyses. To determine the location of a vehicle, we back-project waves within the two 2D arrays to source locations associated with the route.

Depending on the wave propagation direction, the time delay between the target node and the reference node (referred to as lag time) varies in time. Thus lag times can be used

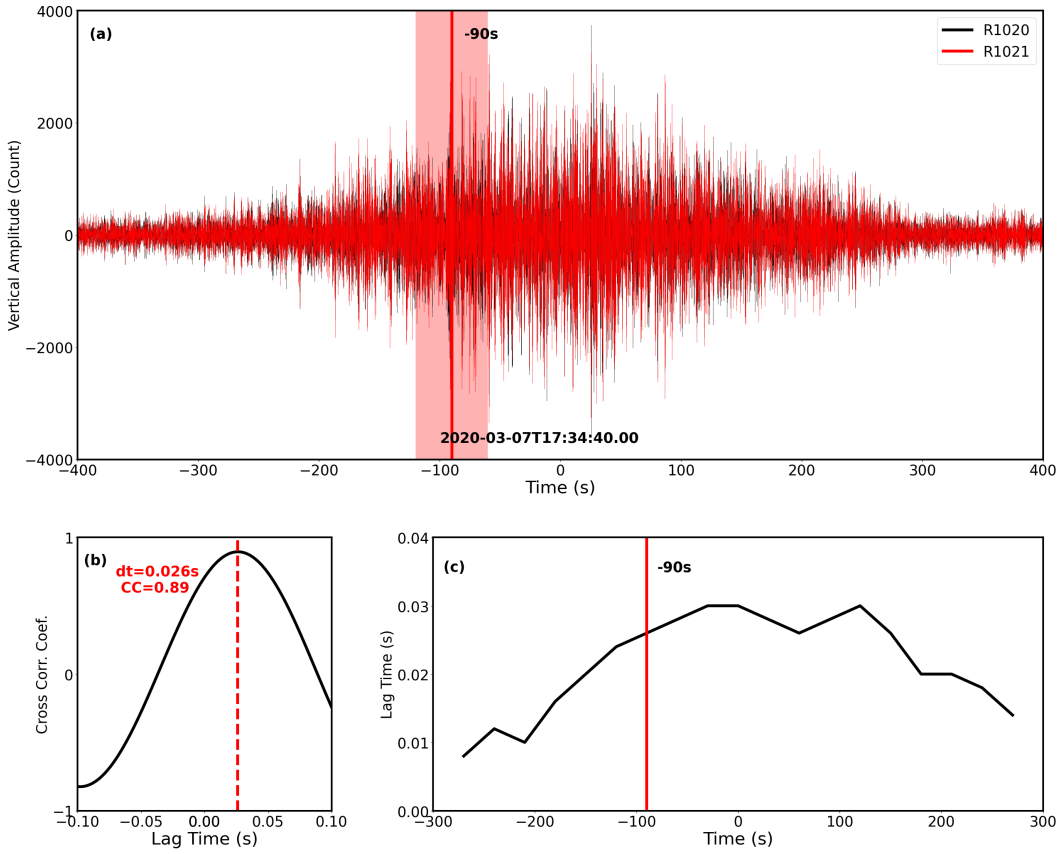


Figure 2. **(a)** Vertical waveforms of a vehicle event E1 on March 7th, 2020, around 17:34:40 recorded by example nodes R1020 and R1021 (see Figure 1c). The seismic records are band-pass filtered at 3-5 Hz with a 4th-order Butterworth filter. An example time window centered at -90s with a length of 60 seconds is shown with a vertical solid line and red transparent area used to compute the time delay. **(b)** Cross-Correlation Coefficient (CC) of waveforms recorded by R1020 and R1021 during the example 60 s time window shown in (a). The CC reaches a maximum (0.86) with a lag time of 0.026 s. **(c)** The resloved lag time from different moving windows increases as the train approaches the intersection of the railway and line connecting nodes R1020 and R1021 and then decreases as it moves away.

to track the movement of vehicles. Figure 2a shows vertical waveforms for a traffic-induced event that was recorded on March 07, 2020, by nodes R1020 and R1021 (marked as black dots in Figure 1c). The event center time is used as the reference zero time. To measure the lag time δ between nodes i and j for any given time t, we first calculate the cross-correlation function (Figure 2b)

$$C_{i,j}(\tau) = \frac{\int_{t_b}^{t_e} w_i(t) w_j(t-\tau) \mathrm{d}t}{\sqrt{\int_{t_b}^{t_e} w_i(t)^2 \mathrm{d}t} \cdot \sqrt{\int_{t_b}^{t_e} w_j(t)^2 \mathrm{d}t}} \tag{1}$$

between recorded waveforms w_i and w_j within the time window $[t_b, t_e] = [t - 30s, t + 30s]$, and take the time corresponding to the maximum value of cross-correlation function as the lag time

$$\delta = \arg\max_{\tau} C_{i,j}(\tau) \tag{2}$$

We use a sliding time window to measure the change in lag times between R1020 and R1021 during the entire event. As shown in Figure 2c, the lag time increases as the vehicle approaches the origin and then decreases as the vehicle moves away.

This sliding window analysis can be further utilized to resolve the wave propagation direction in a 2D array. With an array of n nodes, there are $n(n+1)/2$ node pairs. We calculate the Cross-correlation Coefficients (CC, the maximum value of cross-correlation functions) and lag times for all node pairs. To ensure robustness, only node pairs with $CC \geq 0.7$ are selected for each time window (Figure S2). Since the source-receiver distance is much larger than the apertures of the 2D arrays, we use a plane wave assumption that the wave propagation direction is uniform in the 2D array. The horizontal slowness of wave propagation $\underline{s}$ can be represented as

$$\underline{s} = -\underline{\underline{X}}^g \underline{\Delta} \tag{3}$$

where $\underline{\underline{X}} = [\underline{x_1}, ..., \underline{x_k}]^T \in \mathbb{R}^{k \times 2}$ are the locations of k node pairs and $\underline{\Delta} = [\delta_1, ..., \delta_k]^T \in \mathbb{R}^k$ are the lag times of those node pairs ($\underline{\underline{X}}^g$ is the generalized inverse matrix of $\underline{\underline{X}}$). This equation implies that the wave travels in the opposite direction of the gradient of the lag time. The Rayleigh wave velocity is determined by the inverse of the slowness.

Since previous studies (e.g., Blisniuk et al., 2021) show that the structure around the MCF is more complex compared to the BF, we determine the locations of vehicles by using back projection only from the 2D array around the BF and ignoring the bending of the wave propagation from vehicles to these 2D array. Using a 60-second long sliding window with a 30-second overlap, we resolve changes in the wave propagation directions over time at the

BF array. In each 60-second window, the directional change of the waves propagating from the vehicle to nodes is less than 5°, so we regard the vehicle as a static point in each time window. To evaluate the uncertainties, we apply a bootstrap procedure with 100 samplings. For each sampling, 75% of the total node pairs are randomly selected to resolve the vehicle's location. The mean value and the one-standard deviation of all samplings are considered as the final location and its uncertainty. For the example event, the vehicle's location relative to the origin increases from about -7 km at -180 s to 4 km at 180 s, demonstrating that the vehicle moves from the northwest to the southeast along the route at a speed of 108 km/h. This analysis also provides clear evidence that the signal detected is induced by traffic and not other events, such as tremors that have similar waveform characteristics (Li et al., 2018; Inbal et al., 2018).

2.3 A vertical layered model

The ongoing tectonic deformation and occurrence of earthquakes along faults modify the rock properties in the surrounding volume, including the shear modulus, wave velocity, attenuation coefficient, and mass density (e.g., Ben-Zion & Sammis, 2003; Allam et al., 2014; Qiu et al., 2021). To focus on lateral variations across the two main strands of the SSAF in the study area, we use the vertical layered model illustrated in Figure 3a. We define a Cartesian coordinate system with its origin at the point where the vehicle route and the linear array profile intersect. The x-axis has its positive direction along the linear array toward northeast and the y-axis has its positive direction along the highway or railway toward southeast. A recent analysis by Vavra et al. (2021) suggested that the BF and MCF have non-vertical dips in the study area below 1 km. However, the frequency band analyzed in this study (2-5 Hz) is primarily sampling the top 100 m of the subsurface (Figure S3), and hence the dipping effect on our results is negligible.

In the employed vertical layered model, nodes in the linear array divide the half-space into 113 layers with interfaces parallel to the MCF (i.e., the $y-z$ plane). The width of the j-th layer d_j is given by the distance between the $(j-1)$-th node and the j-th node, except for d_1, which is the distance between the first node (R1001) and the route. The value of d1 (about 4.5 km) is much larger than subsequent layer widths (20 to 50 m). The medium properties are assumed homogeneous in each layer, and we also assume smooth variations of properties other than across fault surfaces. We therefore consider reflections only on the interfaces associated with the BF and MCF.

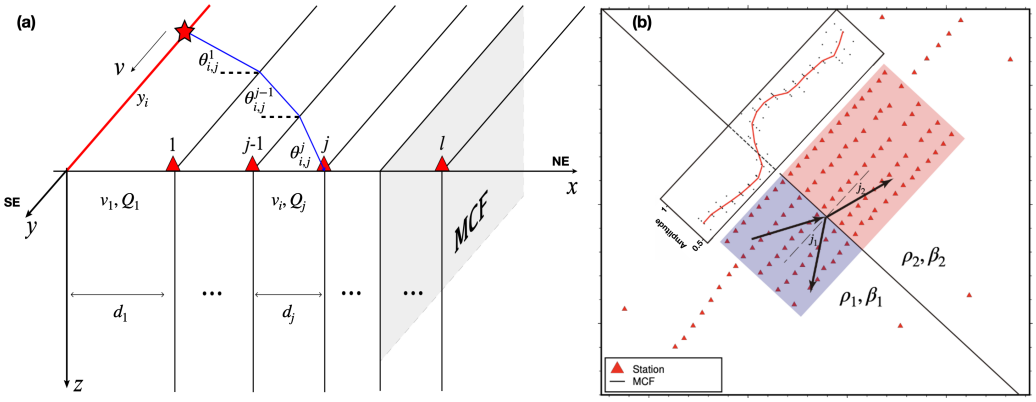


Figure 3. **(a)** A vertical layered model employed in the analysis. We divide the half space to 113 vertical layers with 112 nodes on the interfaces. In each layer, the Rayleigh wave velocity and Q-value are homogeneous. The interfaces are parallel to MCF shown by the gray parallelogram. The distance between the first node (R1001) and the vehicle route (bold red line) is denoted as d_1. The widths of subsequent layers are given by the distances (e.g. d_j) between nodes. The blue line illustrates wave propagation from the vehicle (red star) to the j-th node at the i-th time window. **(b)** Illustration of a Rayleigh wave front (black arrows) propagating to the northeast across the MCF (the black straight line) with an incidence angle j_1. The incident wave is reflected with an angle j_1 and transmitted with an angle j_2. The inset along the 2D subarray shows the amplitude of the vertical amplitude of Rayleigh wave across the MCF. Each black dot represents the amplitude at a node in the 2D array and the red curve gives the mean absolute amplitude on nodes in lines that are parallel to the MCF.

2.4 Derivation of shallow Rayleigh wave velocities

Extracting specific phases from the waveforms is challenging due to the long duration of traffic-induced signals that mask the time difference between body waves and surface waves. However, the surface waves excited by sources such as moving vehicles on rail and road dominate the energy recorded by the surface seismographs (Meng et al., 2021). This is also confirmed by particle motion analysis (Text S1, Figures S4 and S5).

We use $\theta^k_{i,j}$ to denote the incident angle at the k-th interface of the ray path from the vehicle to the j-th station in the i-th time window (Figure 3a). With the known Rayleigh wave velocities in first $(j-1)$-th layers $v_1, ..., v_{j-1}$, the location of a vehicle within the i-th time window y_i can represented as:

$$y_i \approx \sum_{k=1}^{j-1} d_k \cdot \tan\theta^k_{i,j} \tag{4}$$

The values of $\theta^k_{i,j}$ can be solved using Snell's law $\sin\theta^k_{i,j}/v_k = \sin\theta^{j-1}_{i,j}/v_{j-1}$. The lag time between the $(j-1)$-th and j-th nodes within the i-th time window can then be represented as

$$\delta_{i,j} = \sum_{k=1}^{j} \frac{d_k}{v_k \cdot \cos\theta^k_{i,j}} - \sum_{k=1}^{j-1} \frac{d_k}{v_k \cdot \cos\theta^k_{i,j-1}} \approx \frac{d_j \cdot \cos\theta^j_{i,j}}{v_j} \tag{5}$$

The Rayleigh wave velocity in the j-th layer can be solved (using again Snell's law) from

$$v_j = \frac{d_j}{\sqrt{\delta^2_{i,j} + d^2_j \sin^2\theta^{j-1}_{i,j}/v^2_{j-1}}} \tag{6}$$

For simplicity, we assume the first two layer share the same Rayleigh wave velocity, which can be solved as $d_2 \cdot \cos\theta^1_{i,2}/\delta_{i,2}$, where $\theta^1_{i,2} = \arctan(y_1/d_1)$ is the angle between the ray from the vehicle to the first node and x-axis. The Rayleigh wave velocities for subsequent layers can be solved inductively.

2.5 Inversion of amplitudes to Q-values

As seismic waves propagate, their amplitudes attenuate because of geometrical spreading, intrinsic attenuation due to inelastic processes (e.g. internal friction), and scattering due to small-scale heterogeneities. The intrinsic attenuation and scattering effects are quantified jointly by the dimensionless parameter Q defined as:

$$\frac{1}{Q(\omega)} = -\frac{1}{\pi}\frac{\Delta A}{A} \tag{7}$$

where ω and A represent the frequency and amplitude of the analyzed waves, respectively, and ΔA is the attenuation of amplitude in a cycle.

As demenstrated above, a moving vehicle can be approximated by a point source in a short time window that generates surface waves. We calculate the Root Mean Square (RMS) amplitudes of waveforms recorded by the linear array and normalized by the maximum RMS among all nodes. The normalized RMS amplitude of wave motion recorded by the j-th node within the i-th time window at frequency f can be expressed as:

$$A_{i,j} = \frac{A_0}{\sqrt{r_{i,j}/r_0}} \cdot \exp(-\pi f t^*_{i,j}) \tag{8}$$

where A_0 is the amplitude at a reference distance r_0 and $r_{i,j}$ is the distance between the vehicle and the node. The term $\sqrt{r_{i,j}/r_0}$ is used to correct the effect of geometrical spreading. The attenuation factor $t^*_{i,j}$ is the integrated value of $1/Q$ along the ray path. In the assumed vertical layered model (Figure 3a), this can be discretized as

$$t^*_{i,j} = \sum_{k=1}^{j} \frac{d_k}{Q_k v_k \cdot \cos\theta^k_{i,j}} = \sum_{k=1}^{j} \frac{t^*_k}{\cos\theta^k_{i,j}} \tag{9}$$

where $t^*_k = d_k/Q_k v_k$, and Q_k is the Q-value for the k-th layer. Comparing the amplitudes of a traffic event signal recorded by two neighboring node $j-1$ and j, we have

$$\ln\frac{A_{i,j}}{A_{i,j-1}} = -\frac{\pi f}{\cos\theta^j_{i,j}} t^*_j + \pi f \sum_{k=1}^{j-1} \left(\frac{1}{\cos\theta^k_{i,j-1}} - \frac{1}{\theta^k_{i,j}}\right) t^*_k + \frac{1}{2}\ln\frac{r_{i,j-1}}{r_{i,j}} \tag{10}$$

For m time windows and l nodes, we can build an overdetermined linear system with $m \times (l-1)$ equations. The unknown $t^*_1, ..., t^*_l$ can then be solved along with $Q^*_1, ..., Q^*_l$ for a given frequency f.

3 Results

3.1 Classification of traffic event sources

Estimating the speed of vehicles can lend support to the reliability of the results and offer valuable information about the types of traffic event sources. We use the data generated by the examined 29 traffic events to estimate their speed by applying the tracking method in section 2.2. The estimated speeds are listed in Table S1 and their distribution is shown in Figure S6. The results consist of two main clusters with velocities below 70 km/h and above 100 km/h, corresponding to freight trains and trucks, respectively. This is consistent with direct estimates of velocities conducted during a field observation near the highway and railway. We do not observe a significant difference in amplitude between waveforms induced by these two types of sources. The amplitudes are combined effects of the weight and speed of vehicles, along with the coupling between railway/highway to the ground, and can not be used alone to separate between freight trains and trucks.

3.2 Shallow Rayleigh wave velocities

We apply the method described in section 2.4 to 150 time windows from 29 observed freight train and truck events. To ensure robustness, we only consider the resolved velocity of a layer between two nodes for further analysis if the waveforms recorded by these two nodes have a CC value greater than 0.7 within the analyzed window. The quality controlled results are shown in Figure 4 as blue dots; we average these values to estimate the Rayleigh wave velocity along the linear array and use ± one standard deviation as the measurement uncertainties. Since the locations of vehicles have a good coverage of the route, heterogeneities parallel to the fault direction (if any) should be averaged. The result shown in Figure 7 is over 2-5 Hz; we also perform a similar analysis to bandpass filtered waveforms over 2-3 Hz, 3-4 Hz, and 4-5 Hz and obtain similar resolved Rayleigh wave velocities (Figure S7). We do not observe clear dispersion effect in this frequency range, which corresponds to a depth range of 50-100 m (Figure S3).

The results show clear horizontal variations of Rayleigh wave velocities along the linear array across the BF and the MCF. Shallow Rayleigh wave velocities fluctuate slightly around 0.8 km/s near the BF and gradually increase to 1.1 km/s on the SW side of the MCF. After crossing the MCF, the Rayleigh wave velocity drops abruptly by about 35% to 0.7 km/s on the NE side, indicating that the MCF is a sharp bimaterial interface in the top crust. In contrast, the BF does not appear to be associated with a significant velocity contrast in the shallow crust.

3.3 Low Q-values around the BF and MCF

We first normalize the RMS amplitudes of each event on the linear array by the maximum amplitude among all nodes and then take an average for all 29 events. To suppress the site effects of different nodes, we smooth the amplitude along the linear profile using a moving average with a length of 3 nodes. We observe a decrease in recorded amplitudes from SW to NE along the linear array profile, a trend that persists after applying a distance correction to remove the effects of geometrical spreading (Figure S8). Given the sudden increase in amplitude across the MCF (Figure 2b), which is not caused by attenuation, we apply the method described in section 2.5 separately to the two sides of the MCF and then combine the results together to obtain Q-values along the entire linear array. The final

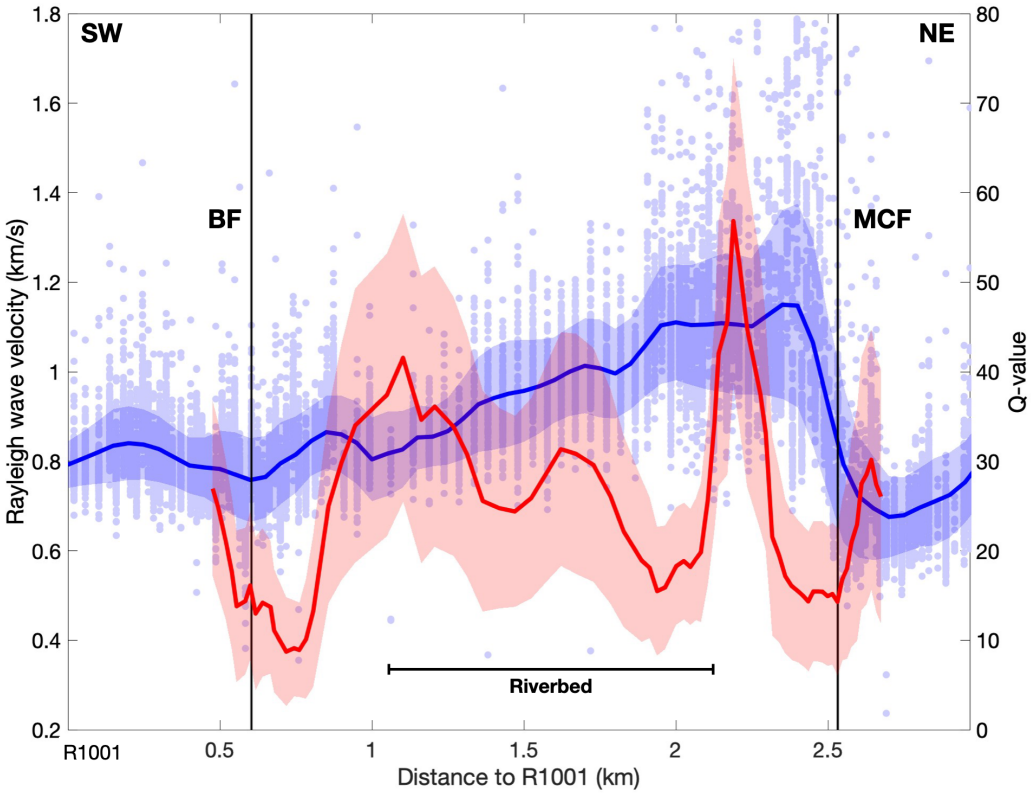


Figure 4. **(a)** The resolved velocities of Rayleigh waves (bold blue curve) and Q-values (bold red curve) along the linear array (starting from sensor R1001), both with shading for $\pm$ one standard deviation of the measurement uncertainty. Each blue dot represents the velocity estimated using the waveform in a given time window of a traffic event. The average values of dots with $CC > 0.7$ provide estimates for the velocities along the linear array. The resloved shallow Q-values across the SSAF range from 10 to 40, with low values around the two fault strands suggesting the presence of possible damage zones. The location of the Banning Fault (BF) and Mission Creek Fault (MCF) are marked by the black lines.

results are shown in Figure 4, along with uncertainties derived from measurement error in the Rayleigh wave velocities used in the inversion.

The resolved Q-values range from 10 to 40 with a mean value of 20. These values are consistent with previous derivations of attenuation factors of the San Jacinto fault zone from analysis of traffic signals within a similar frequency range (Meng et al., 2021; Zhao et al., 2023). As expected, The damage zones around the two strands are characterized by strong attenuation, with Q-values as low as 10, in agreement with results based on modeling of trapped waves in several fault and rupture zones (Peng et al., 2003; Lewis & Ben-Zion, 2010; Qiu et al., 2017; Qin et al., 2021). We also observe low Q-values in the riverbed between the BF and the MCF, which is consistent with the strong attenuation in the unconsolidated sediments of the shallow structure (Figure 1b).

4 Discussion and Conclusions

We present a detailed analysis of subsurface seismic properties across the Banning and Mission Creek strands of the Southern San Andreas Fault in the Coachella Valley, CA, using traffic-induced seismic signals recorded by a dense nodal array. The resolved Rayleigh wave velocities exhibit slight variations across the Banning Fault, gradual increases from BF to MCF, and a significant reduction of 35% across the MCF on its northeastern side (Figure 4). We observe a steep change in the amplitude of the wavefield across the MCF due to the interaction of the wavefield with the property contrast across the fault, providing evidence of a sharp bimaterial interface in the shallow structure of the MCF. Our study also reveals that the Q-values fluctuate across the entire fault zone, ranging from 10 to 40, consistent with values previously derived from studies of attenuation coefficients of shallow and/or fault zone materials.

The 35% reduction in velocity across the MCF is consistent with previous studies of velocity contrasts across faults at shallow depths in other segments of the San Andreas Fault (SAF). For instance, Ben-Zion et al. (1992) and Lewis et al. (2007) found shallow P-wave velocity contrasts of 15% and 50% across the SAF at Parkfield and south of Hollister, respectively. The observed bimaterial interface across the MCF may also correspond to the boundary of the fault with a low velocity damage zone, as found for several locations along the San Jacinto fault (e.g., Qiu et al., 2017; Qin et al., 2021). However, the limited extent of the linear array on the NE side of the MCF precludes a determination of whether

the observed Rayleigh wave velocity contrast across the MCF corresponds to an interface between two crustal blocks or between the fault and a damage zone.

The results summarized in Figure 7 can be used to derive changes of the mass density and shear rigidity across the MCF. For a narrowly bandpass waveform (2-5 Hz), the amplitude ratios of the incident wave A_0, reflected wave A_R and transmitted wave A_T are given by (Aki & Richards, 2002):

$$\frac{A_R}{A_0} = \frac{\rho_1 v_1 \cos j_1 - \rho_2 v_2 \cos j_2}{\rho_1 v_1 \cos j_1 + \rho_2 v_2 \cos j_2} \tag{11}$$

$$\frac{A_T}{A_0} = \frac{2\rho_1 v_1 \cos j_1}{\rho_1 v_1 \cos j_1 + \rho_2 v_2 \cos j_2} \tag{12}$$

where ρ, v and j represent the mass densities, wave velocities, and angles, respectively, and subscripts 1 and 2 mark the incident (SW) and transmission (NE) sides of the wavefield. Considering the phase change of the reflected wave on the interface, the amplitude at the node next to the interface from the SW side is $A_1 = A_0 - A_R$. Thus, the contrast of amplitude at nodes next to the interface on the NE and SW sides should be

$$\frac{A_T}{A_1} = \frac{\rho_1 v_1 \cos j_1}{\rho_2 v_2 \cos j_2} \tag{13}$$

To analyze data with equation (13), we first divide the 2D array around the MCF to two parts separated by the fault interface and then derive wave velocities and incident/transmission angles on the different sides (Figure 2b). The obtained incident and transmission velocities are 1.08 km/s and 0.73 km/s, while the incident and transmission angles are 15° and 10°. These values satisfy the relations

$$\frac{v_2}{v_1} \approx 0.68, \ \frac{\sin j_2}{\sin j_1} \approx 0.65 \tag{14}$$

The results correspond well to Snell's law and support the stability of the analysis. The density contrast across the bimaterial interface is

$$\frac{\rho_2}{\rho_1} = \frac{A_1 v_1 \cos j_1}{A_2 v_2 \cos j_2} \approx 0.9 \tag{15}$$

The relations between the derived velocities and mass densities are consistent with the empirical relationship of Brocher (2005) based on borehole data in California and ultrasonic laboratory measurements. For a Poisson solid, the phase velocity of Rayleigh wave $v \approx 0.92\beta$ where $\beta = \sqrt{\mu/\rho}$ is the shear wave velocity with being the shear rigidity. Assuming that the contrast of shear wave velocities across the MCF is 35%, same as for Rayleigh waves (Figure 4), along with the 0.9 density ratio in equation (15), the ratio of the shallow shear

moduli across the MCF is estimated to be

$$\frac{\mu_2}{\mu_1} = \frac{\beta_2^2 \rho_2}{\beta_1^2 \rho_1} \approx 0.53 \tag{16}$$

While the MCF is associated in the top crust with strong changes of seismic velocities, Q-values, and mass densities, the changes across the BF are minor. These observations suggest that the MCF is the main active strand of the SSAF in the area, rather than the BF, in agreement with previous studies by Blisniuk et al. (2021) and Vavra et al. (2021). Earthquake ruptures on the MCF to the northwest are expected to produce less shaking in the heavily populated area south of the San Bernardino and San Gabriel mountains relative to comparable ruptures on the BF that is closer to the population centers.

The subsurface location of the MCF is well resloved by the strong changes in material properties and agrees with the mapped surface trace of the fault. In contrast, the subsurface location of the BF is not well constrained by our observations and may be offset from the surface trace. Additional studies are needed to better constrain the location of the BF below the surface and to image the depth variations of the SSAF structure in the study area at great depths. Several methods could be used to achieve these goals, including utilizing surface and fault zone phases derived from the ambient seismic noise (e.g., Hillers et al., 2014; Mordret et al., 2019; Zigone et al., 2019), using body waves generated by heavy vehicles (Brenguier et al., 2019), and analyzing the scattered wavefield below the array (Touma et al., 2022). Some of these analyses will be conducted in follow-up studies to further enhance our understanding of the subsurface structure of the Southern San Andreas Fault in the Coachella Valley.

Acknowledgments

We thank Wei Wang for useful discussions. This study is supported by the U.S. Department of Energy Office of Science (Award ED-SC0016520). Haoran Meng is supported by the Shenzhen Natural Science Foundation (K23436305) and Research Startup Fund (Y01436128).

Open Research

The nodal array data and associated metadata (Vernon et al., 2020) are freely available for download from the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC) via `https://doi.org/10.7914/SN/YA_2020`.

References

Ajala, R., Persaud, P., Stock, J. M., Fuis, G. S., Hole, J. A., Goldman, M., & Scheirer, D. (2019). Three-dimensional basin and fault structure from a detailed seismic velocity model of Coachella Valley, Southern California. *Journal of Geophysical Research: Solid Earth*, *124*(5), 4728–4750. doi: 10.1029/2018JB016260

Aki, K., & Richards, P. G. (2002). *Quantitative seismology.*

Allam, A. A., Ben-Zion, Y., Kurzon, I., & Vernon, F. (2014). Seismic velocity structure in the Hot Springs and Trifurcation areas of the San Jacinto fault zone, California, from double-difference tomography. *Geophysical Journal International*, *198*(2), 978-999. doi: 10.1093/gji/ggu176

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics*, *46*(4). doi: 10.1029/2008RG000260

Ben-Zion, Y., Katz, S., & Leary, P. (1992). Joint inversion of fault zone head waves and direct P arrivals for crustal structure near major faults. *Journal of Geophysical Research: Solid Earth*, *97*(B2), 1943-1951. doi: 10.1029/91JB02748

Ben-Zion, Y., & Sammis, C. G. (2003). Characterization of Fault Zones. *Pure and Applied Geophysics*, *160*(3), 677-715. doi: 10.1007/PL00012554

Ben-Zion, Y., Vernon, F. L., Ozakin, Y., Zigone, D., Ross, Z. E., Meng, H., ... Barklage, M. (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophysical Journal International*, *202*(1), 370-380. doi: 10.1093/gji/ggv142

Blisniuk, K., Scharer, K., Sharp, W. D., Burgmann, R., Amos, C., & Rymer, M. (2021). A revised position for the primary strand of the Pleistocene-Holocene San Andreas fault in southern California. *Science Advances*, *7*(13). doi: 10.1126/sciadv.aaz5691

Brenguier, F., Boué, P., Ben-Zion, Y., Vernon, F., Johnson, C., Mordret, A., ... Lecocq, T. (2019). Train Traffic as a Powerful Noise Source for Monitoring Active Faults With Seismic Interferometry. *Geophysical Research Letters*, *46*(16), 9529–9536. doi: 10.1029/2019GL083438

Brocher, T. M. (2005). Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust. *Bulletin of the Seismological Society of America*, *95*(6), 2081-2092. doi: 10.1785/0120050077

Díaz, J., DeFelipe, I., Ruiz, M., Andrés, J., Ayarza, P., & Carbonell, R. (2022). Identifi-

cation of natural and anthropogenic signals in controlled source seismic experiments. *Scientific Reports*, *12*(1), 3171. doi: 10.1038/s41598-022-07028-3

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., ... Zeng, Y. (2014). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—The Time-Independent Model. *Bulletin of the Seismological Society of America*, *104*(3), 1122-1180. doi: 10.1785/0120130164

Fuchs, F., Bokelmann, G., & the AlpArray Working Group. (2017). Equidistant Spectral Lines in Train Vibrations. *Seismological Research Letters*, *89*(1), 56–66. doi: 10.1785/0220170092

Hillers, G., Campillo, M., Ben-Zion, Y., & Roux, P. (2014). Seismic fault zone trapped noise. *Journal of Geophysical Research: Solid Earth*, *119*(7), 5786-5799. doi: 10.1002/2014JB011217

Inbal, A., Cristea-Platon, T., Ampuero, J., Hillers, G., Agnew, D., & 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. doi: 10.1785/0120180130

Jiang, Y., Ning, J., Wen, J., & Shi, Y. (2022). Doppler effect in high-speed rail seismic wavefield and its application. *Science China Earth Sciences*, *65*(3), 414-425. doi: 10.1007/s11430-021-9843-0

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. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8519-8539. doi: 10.1029/2018JB017151

Jones, L. M., Bernknopf, R., Cox, D., Goltz, K., James amd Hudnut, Mileti, D., Perry, S., ... Wein, A. (2008). The ShakeOut Scenario: U.S. Geological Survey Open-File Report 2008-1150 and California Geological Survey Preliminary Report 25. *U.S. Geological Survey*.

Lewis, M. A., & Ben-Zion, Y. (2010). Diversity of fault zone damage and trapping structures in the Parkfield section of the San Andreas Fault from comprehensive analysis of near fault seismograms. *Geophysical Journal International*, *183*(3), 1579-1595. doi: 10.1111/j.1365-246X.2010.04816.x

Lewis, M. A., Ben-Zion, Y., & McGuire, J. J. (2007). Imaging the deep structure of the San Andreas Fault south of Hollister with joint analysis of fault zone head and

direct P arrivals. *Geophysical Journal International*, *169*(3), 1028-1042. doi: 10.1111/j.1365-246X.2006.03319.x

Li, C., Li, Z., Peng, Z., Zhang, C., Nakata, N., & Sickbert, T. (2018). Long-Period Long-Duration Events Detected by the IRIS Community Wavefield Demonstration Experiment in Oklahoma: Tremor or Train Signals? *Seismological Research Letters*, *89*(5), 1652-1659. doi: 10.1785/0220180081

Meng, H., Ben-Zion, Y., & Johnson, C. W. (2019). Detection of random noise and anatomy of continuous seismic waveforms in dense array data near Anza California. *Geophysical Journal International*, *219*(3), 1463-1473. doi: 10.1093/gji/ggz349

Meng, H., Ben-Zion, Y., & Johnson, C. W. (2021). Analysis of Seismic Signals Generated by Vehicle Traffic with Application to Derivation of Subsurface Q-Values. *Seismological Research Letters*, *92*(4), 2354–2363. doi: 10.1785/0220200457

Mordret, A., Roux, P., Boué, P., & Ben-Zion, Y. (2019). Shallow three-dimensional structure of the San Jacinto fault zone revealed from ambient noise imaging with a dense seismic array. *Geophysical Journal International*, *216*(2), 896-905. doi: 10.1093/gji/ggy464

Peng, Z., Ben-Zion, Y., Michael, A. J., & Zhu, L. (2003, 12). Quantitative analysis of seismic fault zone waves in the rupture zone of the 1992 Landers, California, earthquake: evidence for a shallow trapping structure. *Geophysical Journal International*, *155*(3), 1021-1041. doi: 10.1111/j.1365-246X.2003.02109.x

Pinzon-Rincon, L., Lavoué, F., Mordret, A., Boué, P., Brenguier, F., Dales, P., ... Hollis, D. (2021). Humming Trains in Seismology: An Opportune Source for Probing the Shallow Crust. *Seismological Research Letters*, *92*(2A), 623–635. doi: 10.1785/0220200248

Qin, L., Share, P.-E., Qiu, H., Allam, A. A., Vernon, F. L., & Ben-Zion, Y. (2021). Internal structure of the San Jacinto fault zone at the Ramona Reservation, north of Anza, California, from dense array seismic data. *Geophysical Journal International*, *224*(2), 1225-1241. doi: 10.1093/gji/ggaa482

Qiu, H., Ben-Zion, Y., Catchings, R., Goldman, M. R., Allam, A. A., & Steidl, J. (2021). Seismic imaging of the Mw 7.1 Ridgecrest earthquake rupture zone from data recorded by dense linear arrays. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2021JB022043. doi: 10.1029/2021JB022043

Qiu, H., Ben-Zion, Y., Ross, Z., Share, P.-E., & Vernon, F. (2017). Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array. *Geophysical Journal International*, *209*(3), 1369-1388. doi: 10.1093/gji/ggx096

Rezaeifar, M., Lavoué, F., Maggio, G., Xu, Y., Bean, C. J., Pinzon-Rincon, L., ... Brenguier, F. (2023). Imaging shallow structures using interferometry of seismic body waves generated by train traffic. *Geophysical Journal International*, *233*(2), 964-977. doi: 10.1093/gji/ggac507

Share, P.-E., Qiu, H., Vernon, F. L., Allam, A. A., Fialko, Y., & Ben-Zion, Y. (2022). General Seismic Architecture of the Southern San Andreas Fault Zone around the Thousand Palms Oasis from a Large-N Nodal Array. *The Seismic Record*, *2*(1), 50–58. doi: 10.1785/0320210040

Share, P.-E., Tábořík, P., Štěpančíková, P., Stemberk, J., Rockwell, T. K., Wade, A., ... Ben-Zion, Y. (2020). Characterizing the uppermost 100 m structure of the San Jacinto fault zone southeast of Anza, California, through joint analysis of geological, topographic, seismic and resistivity data. *Geophysical Journal International*, *222*(2), 781-794. doi: 10.1093/gji/ggaa204

Shaw, J. H., Plesch, A., Tape, C., Suess, M. P., Jordan, T. H., Ely, G., ... Munster, J. (2015). Unified Structural Representation of the southern California crust and upper mantle. *Earth and Planetary Science Letters*, *415*, 1-15. doi: 10.1016/j.epsl.2015.01.016

Sheng, Y., Mordret, A., Sager, K., Brenguier, F., Boué, P., Rousset, B., ... Ben-Zion, Y. (2022). Monitoring Seismic Velocity Changes Across the San Jacinto Fault Using Train-Generated Seismic Tremors. *Geophysical Research Letters*, *49*(19), e2022GL098509. doi: 10.1029/2022GL098509

Stierman, D. J. (1984). Geophysical and geological evidence for fracturing, water circulation and chemical alteration in granitic rocks adjacent to major strike-slip faults. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 5849-5857. doi: 10.1029/JB089iB07p05849

Touma, A., R.and Aubry, Ben-Zion, Y., & Campillo, M. (2022). Distribution of seismic scatterers in the san jacinto fault zone, southeast of anza, california, based on passive matrix imaging. *Earth and Planetary Science Letters*, *578*, 117304. doi: 10.1016/j.epsl.2021.117304

Vavra, E., Qiu, H., Chi, B., Share, P.-E., Allam, A., Morzfeld, M., ... Fialko, Y. (2021). Seismo-Geodetic Investigations of Subsurface Properties of the Southern San Andreas Fault. In *AGU Fall Meeting* (p. S51D-03).

Vernon, F., Share, P.-E., Ben-Zion, Y., Fialko, Y., & Allam, A. (2020). *Southern San*

Andreas Fault Zone. International Federation of Digital Seismograph Networks. doi: 10.7914/SN/YA_2020

Zhao, Y., Nilot, E. A., Li, B., Fang, G., Luo, W., & Li, Y. E. (2023). Seismic Attenuation Extraction From Traffic Signals Recorded by a Single Seismic Station. *Geophysical Research Letters*, *50*(3), e2022GL100548. doi: 10.1029/2022GL100548

Zigone, D., Ben-Zion, Y., Lehujeur, M., Campillo, M., Hillers, G., & Vernon, F. L. (2019). Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays. *Geophysical Journal International*, *217*(2), 879-893. doi: 10.1093/gji/ggz069

Figure 1.

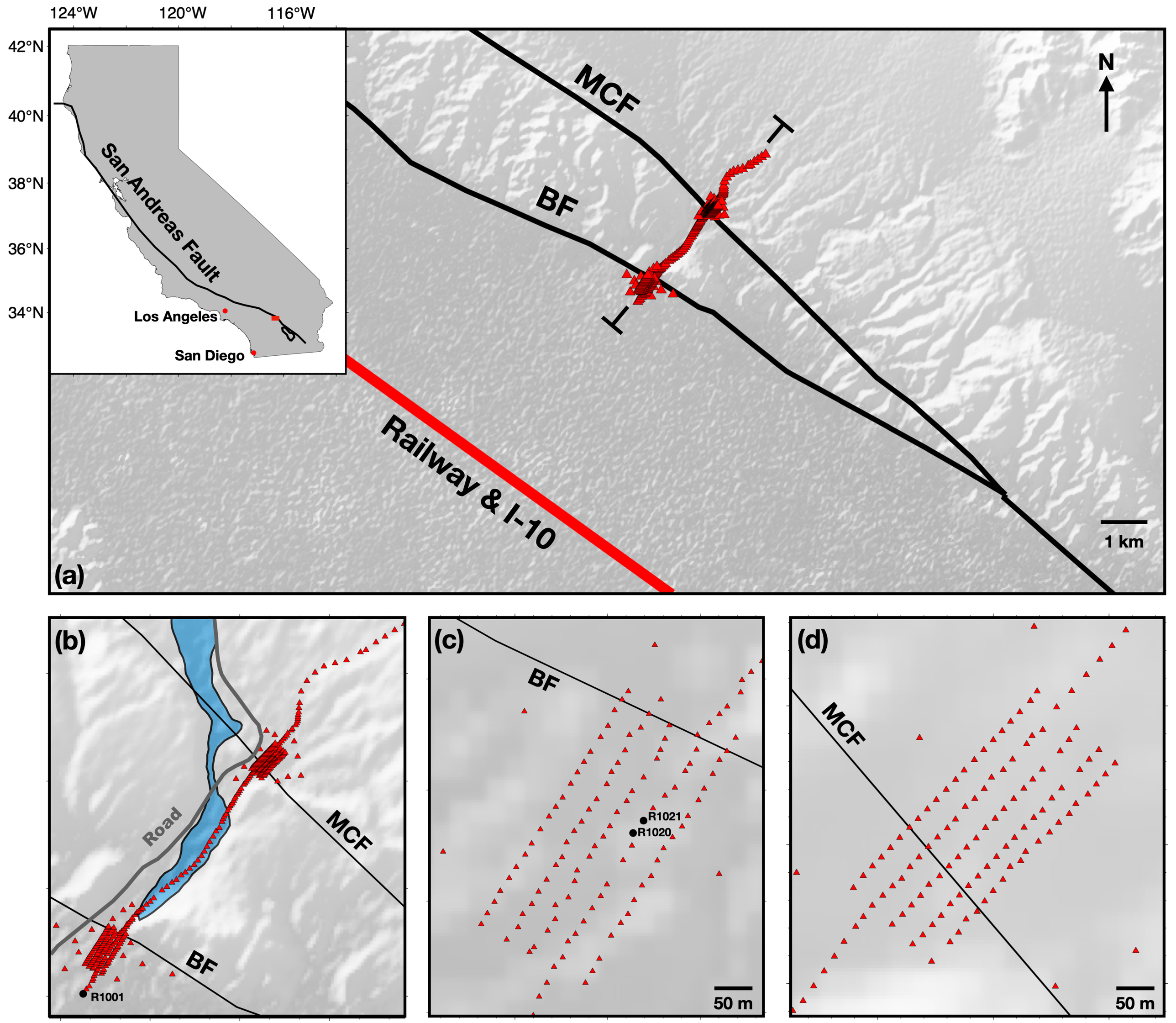
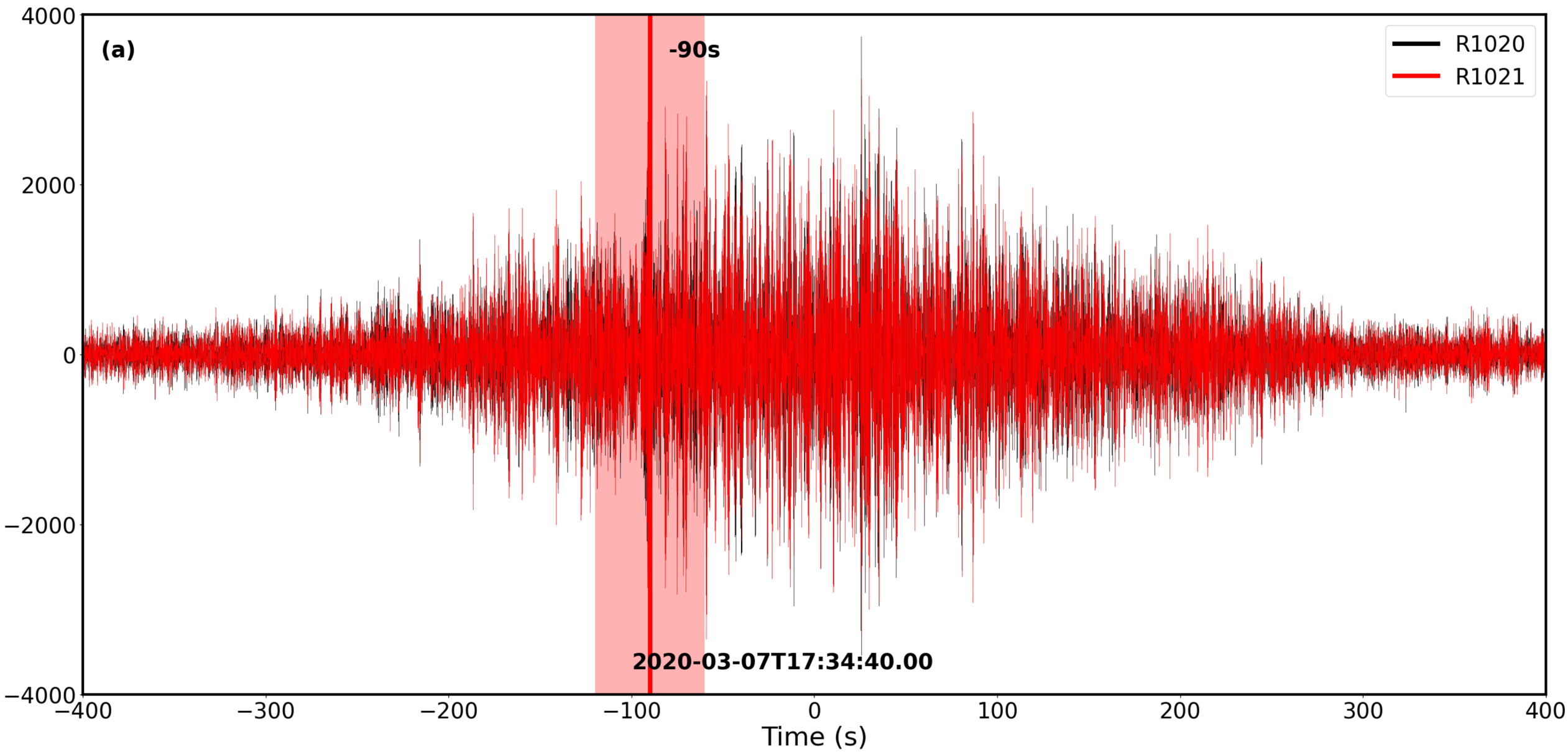
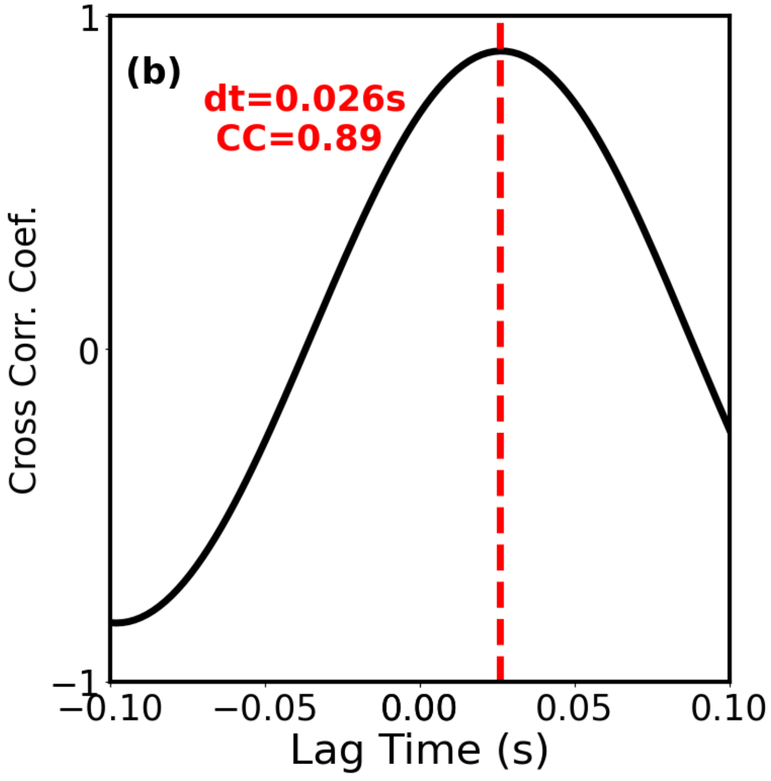

124°W
120°W
116°W
42°N
40°N
38°N
36°N
34°N
San Andreas Fault
Los Angeles
San Diego
MCF
BF
T
T
N
Railway & I-10
1 km
(a)
(b)
Road
MCF
BF
R1001
(c)
BF
R1021
R1020
50 m
(d)
MCF
50 m

Figure 2.


(a)
-90s
R1020
R1021
2020-03-07T17:34:40.00
Vertical Amplitude (Count)
4000
2000
0
−2000
−4000
−400
−300
−200
−100
0
100
200
300
400
Time (s)


(b)
dt=0.026s
CC=0.89
Cross Corr. Coef.
1
0
−1
−0.10
−0.05
0.00
0.05
0.10
Lag Time (s)

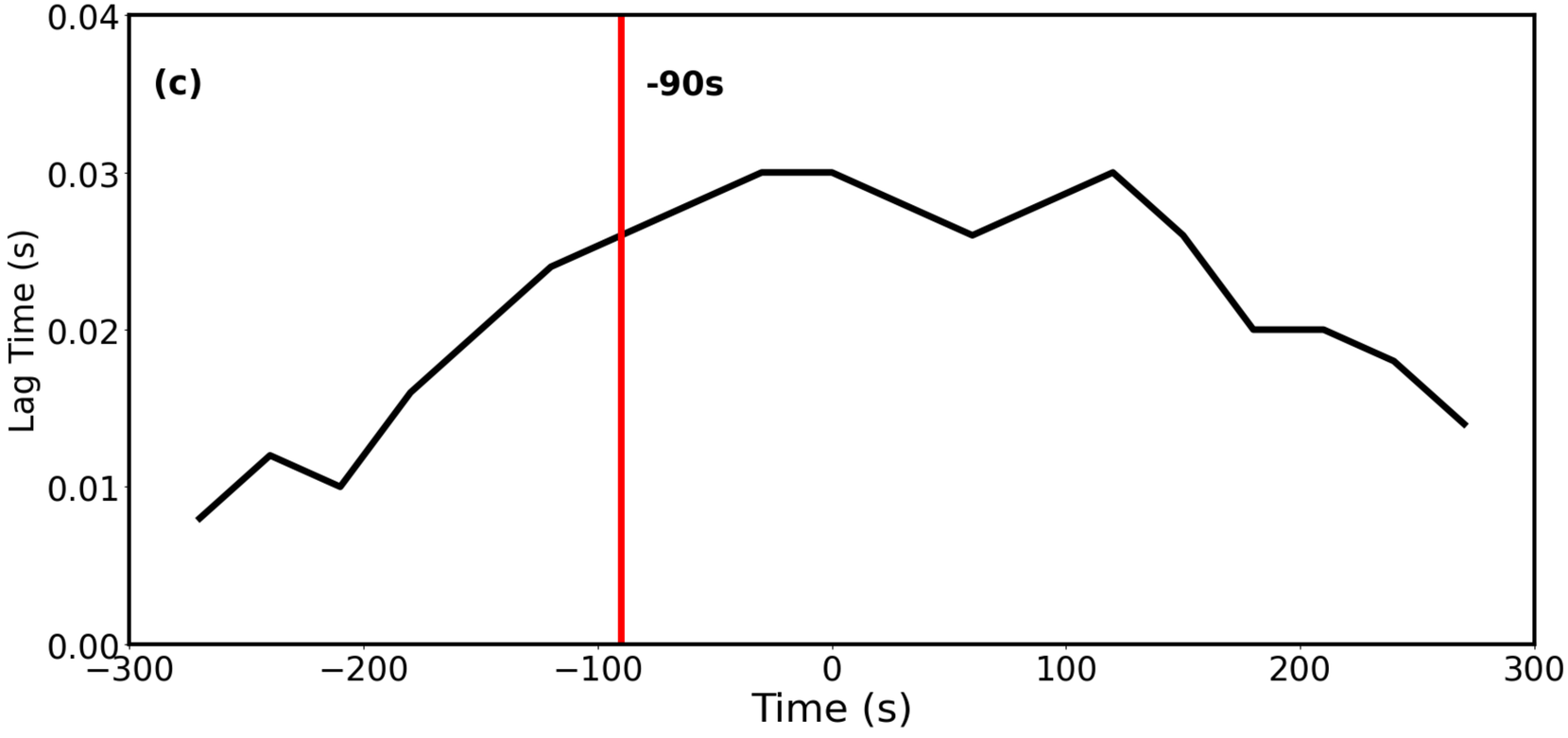
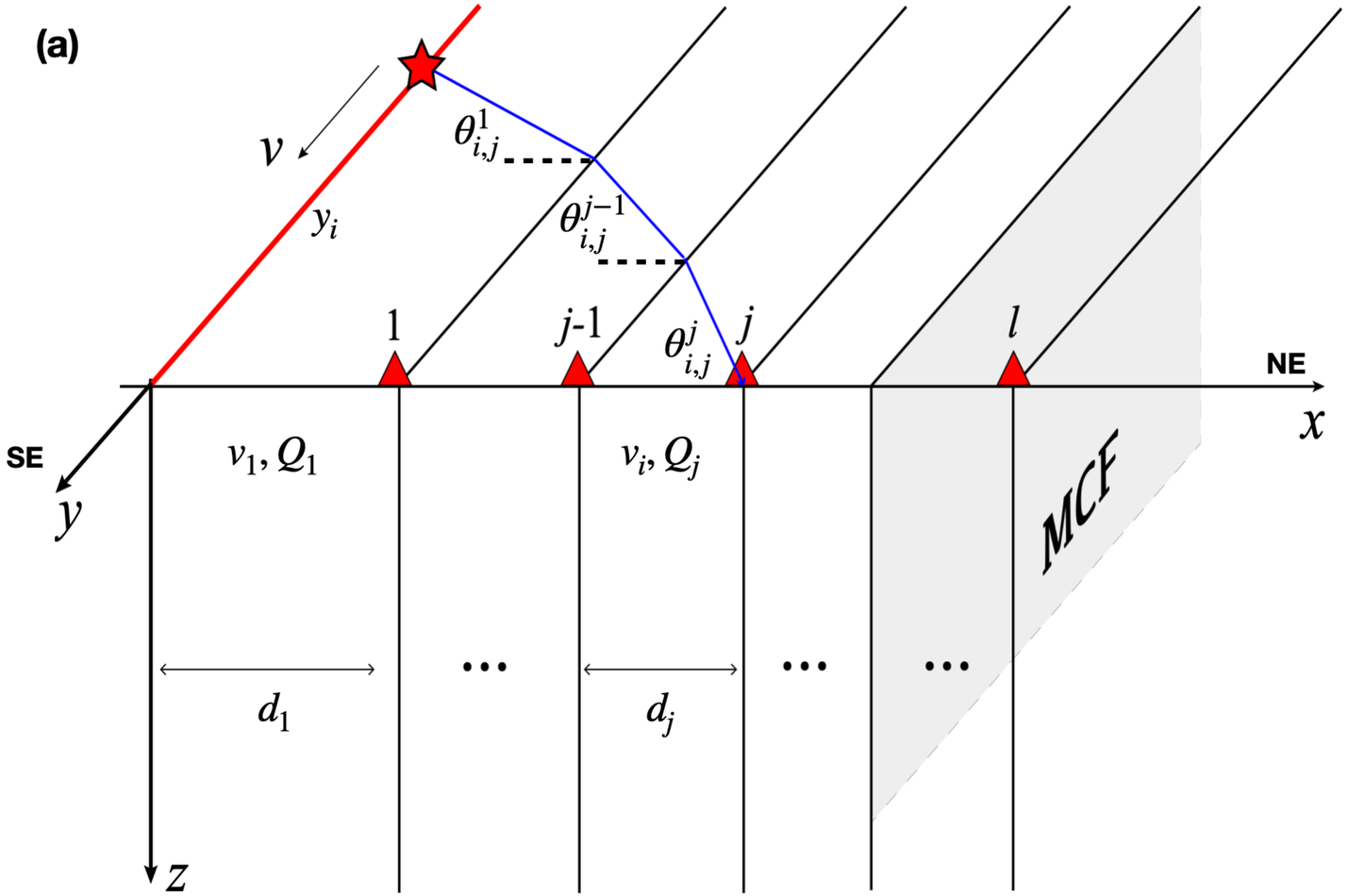

(c)
-90s
Lag Time (s)
0.04
0.03
0.02
0.01
0.00
−300
−200
−100
0
100
200
300
Time (s)

Figure 3.


(a)
v
y_i
$\theta^1_{i,j}$
$\theta^{j-1}_{i,j}$
$\theta^j_{i,j}$
1
j-1
j
l
NE
x
SE
y
z
v_1, Q_1
v_i, Q_j
MCF
d_1
d_j

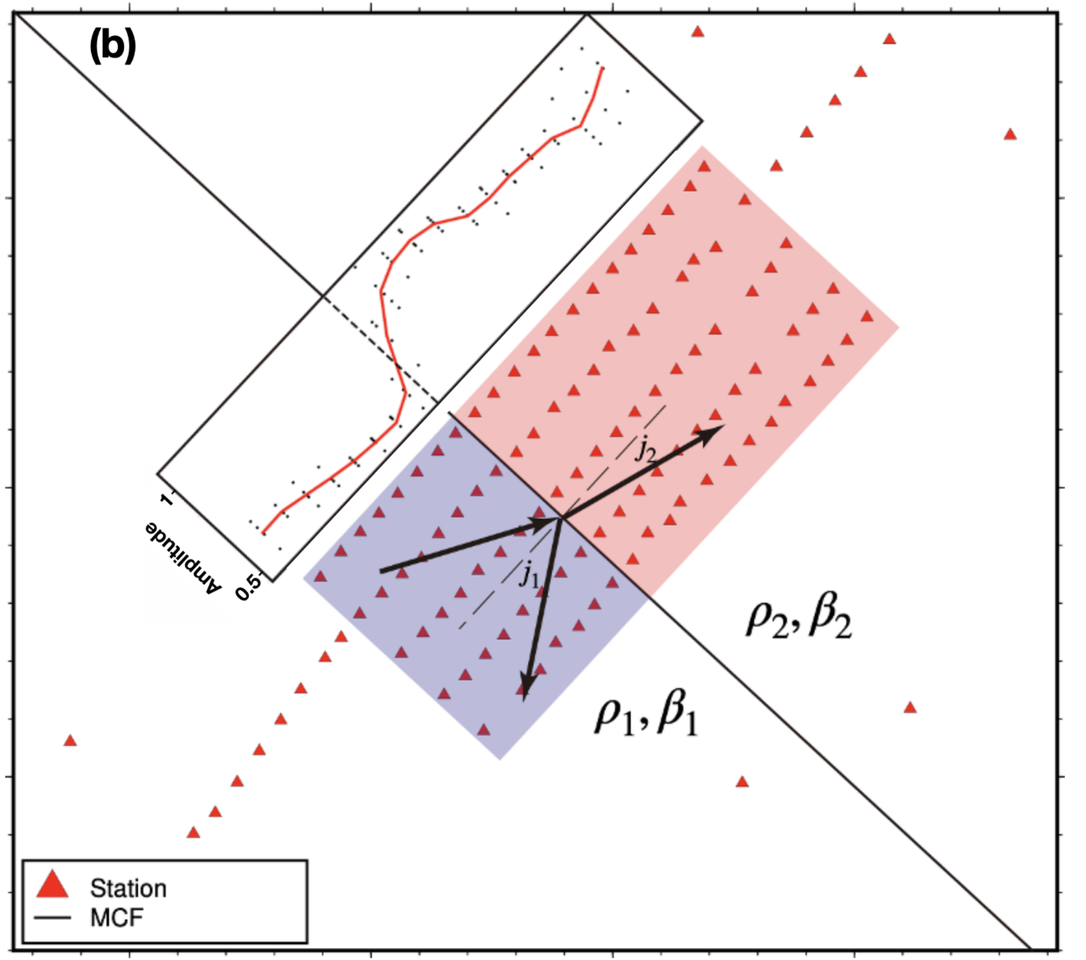
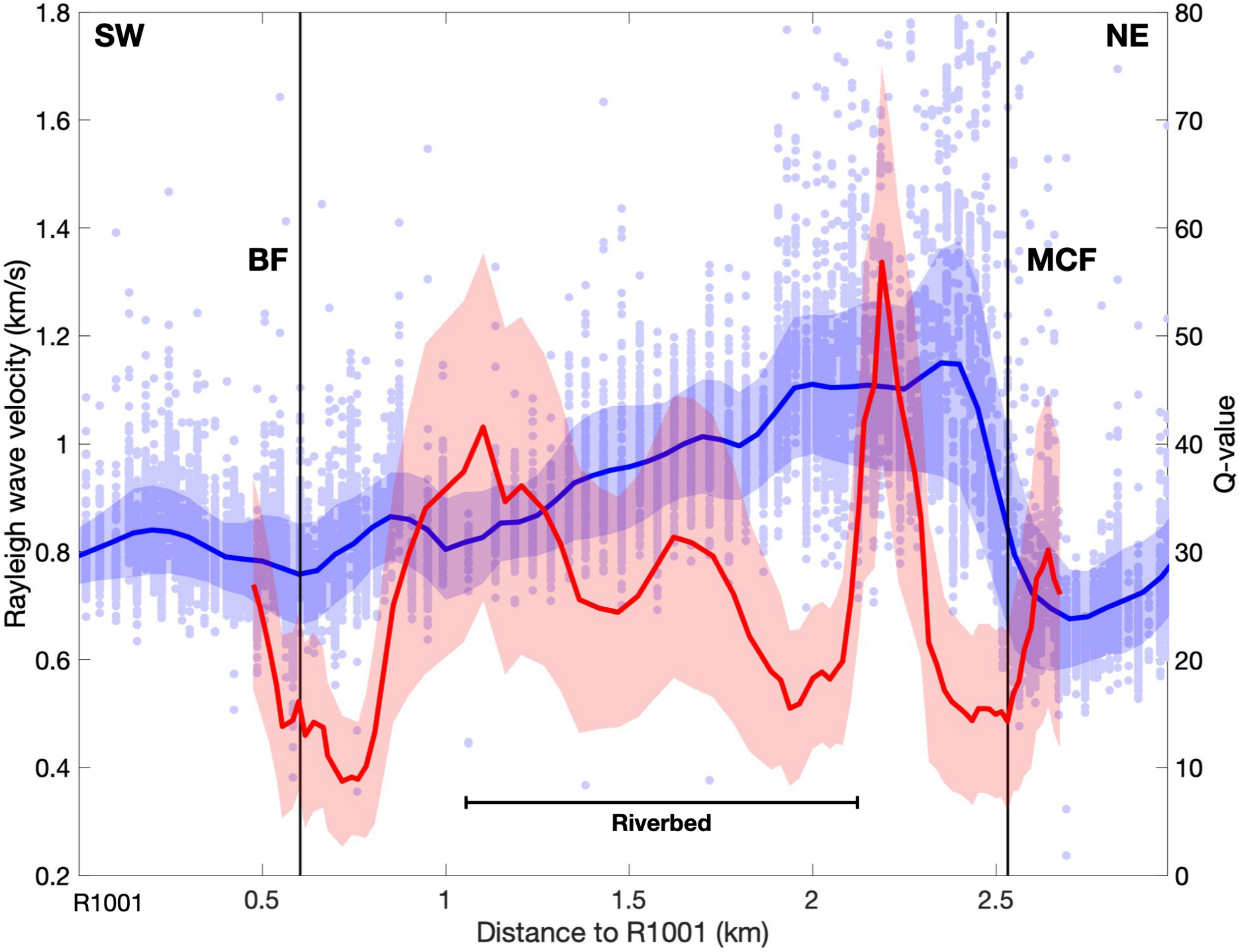

(b)
Amplitude
1
0.5
j_2
j_1
ρ_2, β_2
ρ_1, β_1
Station
MCF

Figure 4.


SW
NE
BF
MCF
Riverbed
Rayleigh wave velocity (km/s)
Q-value
Distance to R1001 (km)
R1001
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
80
70
60
50
40
30
20
10
0
0.5
1
1.5
2
2.5

Lateral Variations Across the Southern San Andreas Fault Zone Revealed from Analysis of Traffic Signals at a Dense Seismic Array

Hao Zhang[1], Haoran Meng[2], and Yehuda Ben-Zion[1,3]

[1]University of Southern California, Los Angeles, California, USA.
[2]Southern University of Science and Technology, Shenzhen, China.
[3]Southern California Earthquake Center, Los Angeles, California, USA.

Key Points:

- We detect frequent seismic signals from rail and road traffic in a dense array across the southern San Andreas fault zone.
- We use the traffic signals to image shallow structural properties across the Banning and Mission Creek fault strands.
- The resolved velocity and density contrasts across the Mission Creek fault suggest it is the main active strand of the SSAF in the area.

Corresponding author: Haoran Meng, menghr@sustech.edu.cn

Abstract

We image the shallow seismic structure across the Southern San Andreas Fault (SSAF) using signals from freight trains and trucks recorded by a dense nodal array, with a linear component perpendicular to SSAF and two 2D subarrays centered on the Banning Fault (BF) and Mission Creek Fault (MCF). Particle motion analysis in the frequency band 2-5 Hz shows that the examined traffic sources can be approximated as moving point sources that primarily induce Rayleigh waves. Using several techniques, we resolve strong lateral variations of Rayleigh wave velocities across the SSAF, including 35% velocity reduction across MCF towards the northeast. Additionally, we derive Q-values and find strong attenuation around the BF and MCF. We further resolve 10% mass density reduction and 45% shear modulus decrease across the MCF. These findings suggest that the MCF is the main strand of the SSAF in the area with important implications for seismic hazard assessments.

Plain Language Summary

Imaging the internal structure of fault zones is essential for understanding earthquake properties and processes. Here we utilize seismic data generated by trains and trucks in the Coachella valley and recorded by a dense seismic array to image the subsurface structure of two main strands of the Southern San Andreas Fault (SSAF). Several types of analyses allow us to resolve seismic velocities, attenuation coefficients, and mass density across the entire San Andreas Fault zone. The results show a clear contrast in physical properties across the Mission Creek strand of the SSAF, highlighting the presence of a bimaterial fault interface and suggesting that it is the main strand of SSAF. The research opens up possibilities for using common rail and road traffic signals to derive high resolution imaging results of subsurface seismic properties at other locations.

1 Introduction

Earthquake fault zones have geometrical and material heterogeneities that reflect their past history and can strongly affect future earthquakes and seismic motion generated by the faults (e.g., Stierman, 1984; Ben-Zion, 2008). The Southern San Andreas Fault (SSAF) has not experienced a large earthquake over the past 300 years and is considered to pose a significant seismic hazard (Field et al., 2014). Various studies attempted to derive seismic velocity models for the SSAF (e.g., Shaw et al., 2015; Ajala et al., 2019), but they lack resolution on internal fault zone components such as sharp bimaterial interfaces and damage

zones. Imaging these features, as well as properties of the top structure which may be compared with geological information, require dense seismic arrays that cross the fault zone of interest (e.g., Ben-Zion et al., 2015; Share et al., 2020; Qiu et al., 2021).

The SSAF in the Coachella valley has two major strands - the Mission Crack Faults (MCF) and the Banning Fault (BF) - and it is a matter of debate which one is the primary active strand of the SSAF (Jones et al., 2008; Blisniuk et al., 2021). In the present study, we use seismic data recorded by a dense temporary nodal array across the BF and MCF near the Thousand Palms Oasis Preserve in the Coachella Valley (Figure 1) to image the subsurface properties of the SSAF in the area. During the ∼1 month deployment, only a small number of local earthquakes occurred near the array (Share et al., 2022), necessitating the use of other signals for detailed seismic imaging. Among such signals, seismic waves generated by cars and trains are used increasingly in imaging and monitoring studies due to their high reproducibility and simple source features (Fuchs et al., 2017; Meng et al., 2021; Pinzon-Rincon et al., 2021; Jiang et al., 2022; Rezaeifar et al., 2023). The reproducibility of traffic-generated signals also makes them suitable for monitoring temporal changes of seismic velocities (Brenguier et al., 2019; Sheng et al., 2022).

In the following sections, we use seismic signals generated by vehicle traffic in the Coachella valley to image the shallow internal structure of the SSAF below the dense nodal array. Utilizing waveforms generated by freight trains and trucks with high signal-to-noise ratio (SNR), we derive Rayleigh wave velocities and Q-values across both the BF and MCF, as well as the contrast of mass density across the MCF. While the analyses indicate significant contrasts of material properties across the MCF, only minor variations are observed across the BF. The results support the view that the MCF is more likely the main strand of the SSAF in the area (Blisniuk et al., 2021).

2 Methods

2.1 Seismic Network and Data

This study utilizes data from 322 nodes deployed across the BF and MCF strands of the SSAF in the Coachella Valley, California (Figure 1). The data was collected continuously from March 3 to April 13, 2020, and recorded using Zland 3-component 5 Hz nodes with a sampling rate of 500 Hz (Share et al., 2020). The array had a quasilinear profile with more than 100 nodes that crossed both the BF and MCF, perpendicular to their surface

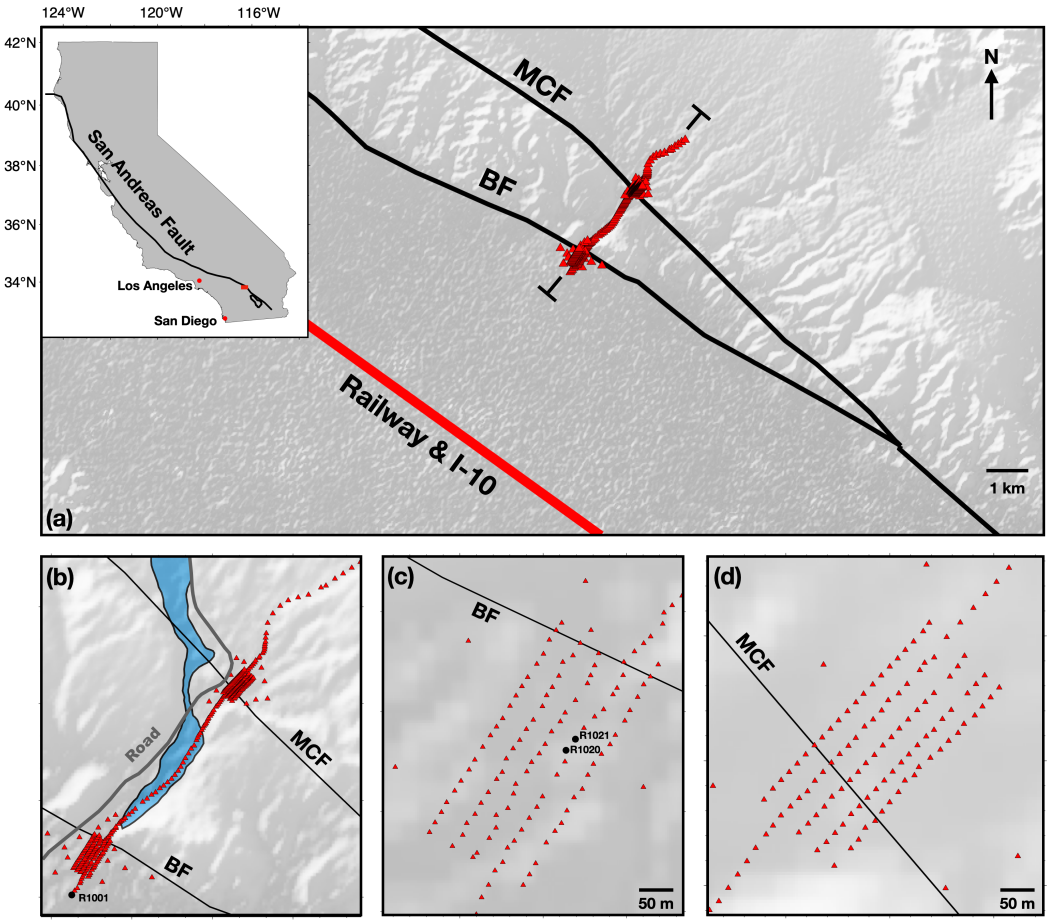


Figure 1. **(a)** Location map of the Coachella Valley showing the dense seismic nodal array employed in this study (red triangles), fault strands (black lines), highway I-10 and adjacent railway (bold red line). The inset in the top-left corner provides a regional view of California with the San Andreas Fault, Los Angeles, San Diego and the study area (red rectangle). **(b)** Location of a local road (gray curve) along the array and a riverbed between the two strands (blue shading). **(c,d)** Zoomed-in views of two 2D sub-arrays centered on the Banning Fault (BF) and Mission Creek Fault (MCF), respectively. Example nodes R1020 and R1021 are denoted by black dots in (c).

traces, with internode spacings that varied from 15 m near the fault traces to 50 m away from them. The linear array connected two 2D subarrays, each containing more than 100 nodes with a 20 m interstation spacing and a 1 km aperture, centered on the BF and MCF. During the data acquisition period, the nodal array recorded about 100 local earthquakes and prevalent seismic signals generated by traffic.

The majority of the traffic-induced signals were from local roads near the array and showed strong energy in the frequency band of 5-35 Hz (Figure S1). However, these signals were usually generated by small vehicles and could only be recorded by a small subset of the array due to geometrical spreading and strong attenuation effects. To image the subsurface structure beneath the entire array, we utilize here much stronger signals from truck traffic on the Interstate 10 (I-10) highway, located approximately 5 km southwest of the BF, and from freight trains on the railway adjacent to the I-10. The highway and railway (jointly referred to as the route) run nearly parallel to the BF and MCF, as shown in Figure 1a. To separate the target signals of freight trains and trucks (together referred to as vehicles) from other anthropogenic and environmental sources, including local cars, air-traffic events, and wind-induced signals (Meng et al., 2019; Johnson et al., 2019; Díaz et al., 2022), we use a bandpass filter and only analyze data in the frequency range of 2 to 5 Hz. A typical vehicle signal has a strong spindle-like symmetry both in the time series and spectrogram, with a duration of hundreds of seconds (Figure 2a and Figure S1). To ensure robustness, 29 events with $SNR > 5$ are manually selected and analyzed further in this study.

2.2 Tracking the motion of vehicles

The recorded traffic-induced signals generally last for several minutes, during which vehicles can move a few kilometers. As the sizes of vehicles are much smaller than the wavelengths (200-500 m) of interest and source-receiver distances, we approximate vehicles as moving point sources with a uniform radiation pattern according to the far-field approximation. We neglect the Doppler effects as the speed of vehicles ($\leq$ 35 m/s) is much less than the seismic wave velocity ($\geq$ 700 m/s). This is confirmed in the following analyses. To determine the location of a vehicle, we back-project waves within the two 2D arrays to source locations associated with the route.

Depending on the wave propagation direction, the time delay between the target node and the reference node (referred to as lag time) varies in time. Thus lag times can be used

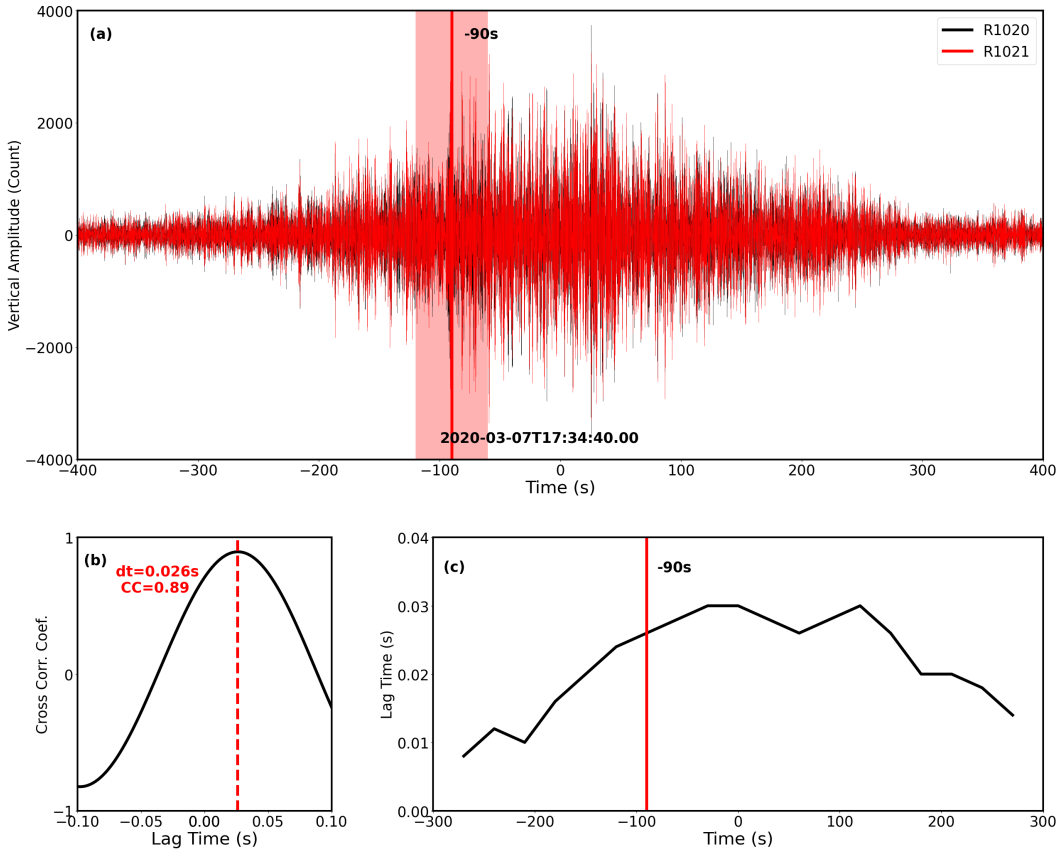


Figure 2. **(a)** Vertical waveforms of a vehicle event E1 on March 7th, 2020, around 17:34:40 recorded by example nodes R1020 and R1021 (see Figure 1c). The seismic records are band-pass filtered at 3-5 Hz with a 4th-order Butterworth filter. An example time window centered at -90s with a length of 60 seconds is shown with a vertical solid line and red transparent area used to compute the time delay. **(b)** Cross-Correlation Coefficient (CC) of waveforms recorded by R1020 and R1021 during the example 60 s time window shown in (a). The CC reaches a maximum (0.86) with a lag time of 0.026 s. **(c)** The resloved lag time from different moving windows increases as the train approaches the intersection of the railway and line connecting nodes R1020 and R1021 and then decreases as it moves away.

to track the movement of vehicles. Figure 2a shows vertical waveforms for a traffic-induced event that was recorded on March 07, 2020, by nodes R1020 and R1021 (marked as black dots in Figure 1c). The event center time is used as the reference zero time. To measure the lag time δ between nodes i and j for any given time t, we first calculate the cross-correlation function (Figure 2b)

$$C_{i,j}(\tau) = \frac{\int_{t_b}^{t_e} w_i(t) w_j(t-\tau) \mathrm{d}t}{\sqrt{\int_{t_b}^{t_e} w_i(t)^2 \mathrm{d}t} \cdot \sqrt{\int_{t_b}^{t_e} w_j(t)^2 \mathrm{d}t}} \tag{1}$$

between recorded waveforms w_i and w_j within the time window $[t_b, t_e] = [t - 30s, t + 30s]$, and take the time corresponding to the maximum value of cross-correlation function as the lag time

$$\delta = \arg\max_{\tau} C_{i,j}(\tau) \tag{2}$$

We use a sliding time window to measure the change in lag times between R1020 and R1021 during the entire event. As shown in Figure 2c, the lag time increases as the vehicle approaches the origin and then decreases as the vehicle moves away.

This sliding window analysis can be further utilized to resolve the wave propagation direction in a 2D array. With an array of n nodes, there are $n(n+1)/2$ node pairs. We calculate the Cross-correlation Coefficients (CC, the maximum value of cross-correlation functions) and lag times for all node pairs. To ensure robustness, only node pairs with $CC \geq 0.7$ are selected for each time window (Figure S2). Since the source-receiver distance is much larger than the apertures of the 2D arrays, we use a plane wave assumption that the wave propagation direction is uniform in the 2D array. The horizontal slowness of wave propagation $\underline{s}$ can be represented as

$$\underline{s} = -\underline{\underline{X}}^g \underline{\Delta} \tag{3}$$

where $\underline{\underline{X}} = [\underline{x_1}, ..., \underline{x_k}]^T \in \mathbb{R}^{k \times 2}$ are the locations of k node pairs and $\underline{\Delta} = [\delta_1, ..., \delta_k]^T \in \mathbb{R}^k$ are the lag times of those node pairs ($\underline{\underline{X}}^g$ is the generalized inverse matrix of $\underline{\underline{X}}$). This equation implies that the wave travels in the opposite direction of the gradient of the lag time. The Rayleigh wave velocity is determined by the inverse of the slowness.

Since previous studies (e.g., Blisniuk et al., 2021) show that the structure around the MCF is more complex compared to the BF, we determine the locations of vehicles by using back projection only from the 2D array around the BF and ignoring the bending of the wave propagation from vehicles to these 2D array. Using a 60-second long sliding window with a 30-second overlap, we resolve changes in the wave propagation directions over time at the

BF array. In each 60-second window, the directional change of the waves propagating from the vehicle to nodes is less than 5°, so we regard the vehicle as a static point in each time window. To evaluate the uncertainties, we apply a bootstrap procedure with 100 samplings. For each sampling, 75% of the total node pairs are randomly selected to resolve the vehicle's location. The mean value and the one-standard deviation of all samplings are considered as the final location and its uncertainty. For the example event, the vehicle's location relative to the origin increases from about -7 km at -180 s to 4 km at 180 s, demonstrating that the vehicle moves from the northwest to the southeast along the route at a speed of 108 km/h. This analysis also provides clear evidence that the signal detected is induced by traffic and not other events, such as tremors that have similar waveform characteristics (Li et al., 2018; Inbal et al., 2018).

2.3 A vertical layered model

The ongoing tectonic deformation and occurrence of earthquakes along faults modify the rock properties in the surrounding volume, including the shear modulus, wave velocity, attenuation coefficient, and mass density (e.g., Ben-Zion & Sammis, 2003; Allam et al., 2014; Qiu et al., 2021). To focus on lateral variations across the two main strands of the SSAF in the study area, we use the vertical layered model illustrated in Figure 3a. We define a Cartesian coordinate system with its origin at the point where the vehicle route and the linear array profile intersect. The x-axis has its positive direction along the linear array toward northeast and the y-axis has its positive direction along the highway or railway toward southeast. A recent analysis by Vavra et al. (2021) suggested that the BF and MCF have non-vertical dips in the study area below 1 km. However, the frequency band analyzed in this study (2-5 Hz) is primarily sampling the top 100 m of the subsurface (Figure S3), and hence the dipping effect on our results is negligible.

In the employed vertical layered model, nodes in the linear array divide the half-space into 113 layers with interfaces parallel to the MCF (i.e., the $y-z$ plane). The width of the j-th layer d_j is given by the distance between the $(j-1)$-th node and the j-th node, except for d_1, which is the distance between the first node (R1001) and the route. The value of d1 (about 4.5 km) is much larger than subsequent layer widths (20 to 50 m). The medium properties are assumed homogeneous in each layer, and we also assume smooth variations of properties other than across fault surfaces. We therefore consider reflections only on the interfaces associated with the BF and MCF.

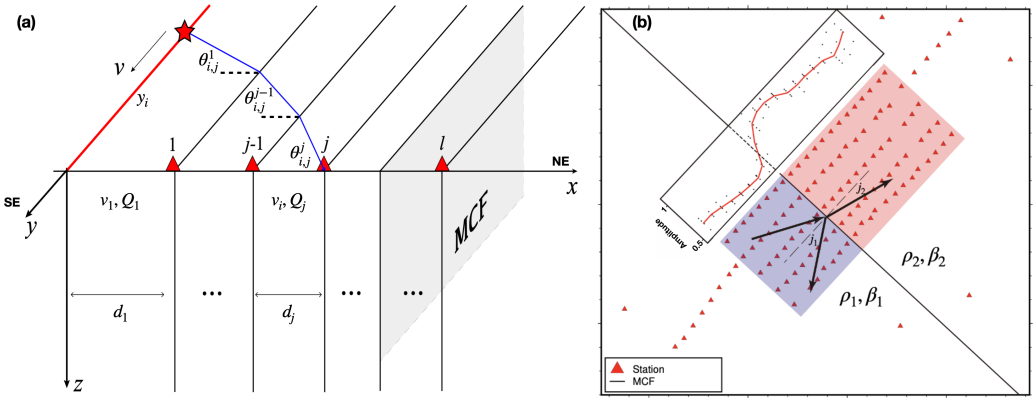


Figure 3. **(a)** A vertical layered model employed in the analysis. We divide the half space to 113 vertical layers with 112 nodes on the interfaces. In each layer, the Rayleigh wave velocity and Q-value are homogeneous. The interfaces are parallel to MCF shown by the gray parallelogram. The distance between the first node (R1001) and the vehicle route (bold red line) is denoted as d_1. The widths of subsequent layers are given by the distances (e.g. d_j) between nodes. The blue line illustrates wave propagation from the vehicle (red star) to the j-th node at the i-th time window. **(b)** Illustration of a Rayleigh wave front (black arrows) propagating to the northeast across the MCF (the black straight line) with an incidence angle j_1. The incident wave is reflected with an angle j_1 and transmitted with an angle j_2. The inset along the 2D subarray shows the amplitude of the vertical amplitude of Rayleigh wave across the MCF. Each black dot represents the amplitude at a node in the 2D array and the red curve gives the mean absolute amplitude on nodes in lines that are parallel to the MCF.

2.4 Derivation of shallow Rayleigh wave velocities

Extracting specific phases from the waveforms is challenging due to the long duration of traffic-induced signals that mask the time difference between body waves and surface waves. However, the surface waves excited by sources such as moving vehicles on rail and road dominate the energy recorded by the surface seismographs (Meng et al., 2021). This is also confirmed by particle motion analysis (Text S1, Figures S4 and S5).

We use $\theta^k_{i,j}$ to denote the incident angle at the k-th interface of the ray path from the vehicle to the j-th station in the i-th time window (Figure 3a). With the known Rayleigh wave velocities in first $(j-1)$-th layers $v_1, ..., v_{j-1}$, the location of a vehicle within the i-th time window y_i can represented as:

$$y_i \approx \sum_{k=1}^{j-1} d_k \cdot \tan\theta^k_{i,j} \tag{4}$$

The values of $\theta^k_{i,j}$ can be solved using Snell's law $\sin\theta^k_{i,j}/v_k = \sin\theta^{j-1}_{i,j}/v_{j-1}$. The lag time between the $(j-1)$-th and j-th nodes within the i-th time window can then be represented as

$$\delta_{i,j} = \sum_{k=1}^{j} \frac{d_k}{v_k \cdot \cos\theta^k_{i,j}} - \sum_{k=1}^{j-1} \frac{d_k}{v_k \cdot \cos\theta^k_{i,j-1}} \approx \frac{d_j \cdot \cos\theta^j_{i,j}}{v_j} \tag{5}$$

The Rayleigh wave velocity in the j-th layer can be solved (using again Snell's law) from

$$v_j = \frac{d_j}{\sqrt{\delta^2_{i,j} + d^2_j \sin^2\theta^{j-1}_{i,j}/v^2_{j-1}}} \tag{6}$$

For simplicity, we assume the first two layer share the same Rayleigh wave velocity, which can be solved as $d_2 \cdot \cos\theta^1_{i,2}/\delta_{i,2}$, where $\theta^1_{i,2} = \arctan(y_1/d_1)$ is the angle between the ray from the vehicle to the first node and x-axis. The Rayleigh wave velocities for subsequent layers can be solved inductively.

2.5 Inversion of amplitudes to Q-values

As seismic waves propagate, their amplitudes attenuate because of geometrical spreading, intrinsic attenuation due to inelastic processes (e.g. internal friction), and scattering due to small-scale heterogeneities. The intrinsic attenuation and scattering effects are quantified jointly by the dimensionless parameter Q defined as:

$$\frac{1}{Q(\omega)} = -\frac{1}{\pi}\frac{\Delta A}{A} \tag{7}$$

where ω and A represent the frequency and amplitude of the analyzed waves, respectively, and ΔA is the attenuation of amplitude in a cycle.

As demenstrated above, a moving vehicle can be approximated by a point source in a short time window that generates surface waves. We calculate the Root Mean Square (RMS) amplitudes of waveforms recorded by the linear array and normalized by the maximum RMS among all nodes. The normalized RMS amplitude of wave motion recorded by the j-th node within the i-th time window at frequency f can be expressed as:

$$A_{i,j} = \frac{A_0}{\sqrt{r_{i,j}/r_0}} \cdot \exp(-\pi f t^*_{i,j}) \tag{8}$$

where A_0 is the amplitude at a reference distance r_0 and $r_{i,j}$ is the distance between the vehicle and the node. The term $\sqrt{r_{i,j}/r_0}$ is used to correct the effect of geometrical spreading. The attenuation factor $t^*_{i,j}$ is the integrated value of $1/Q$ along the ray path. In the assumed vertical layered model (Figure 3a), this can be discretized as

$$t^*_{i,j} = \sum_{k=1}^{j} \frac{d_k}{Q_k v_k \cdot \cos\theta^k_{i,j}} = \sum_{k=1}^{j} \frac{t^*_k}{\cos\theta^k_{i,j}} \tag{9}$$

where $t^*_k = d_k/Q_k v_k$, and Q_k is the Q-value for the k-th layer. Comparing the amplitudes of a traffic event signal recorded by two neighboring node $j-1$ and j, we have

$$\ln\frac{A_{i,j}}{A_{i,j-1}} = -\frac{\pi f}{\cos\theta^j_{i,j}} t^*_j + \pi f \sum_{k=1}^{j-1} \left(\frac{1}{\cos\theta^k_{i,j-1}} - \frac{1}{\theta^k_{i,j}}\right) t^*_k + \frac{1}{2}\ln\frac{r_{i,j-1}}{r_{i,j}} \tag{10}$$

For m time windows and l nodes, we can build an overdetermined linear system with $m \times (l-1)$ equations. The unknown $t^*_1, ..., t^*_l$ can then be solved along with $Q^*_1, ..., Q^*_l$ for a given frequency f.

3 Results

3.1 Classification of traffic event sources

Estimating the speed of vehicles can lend support to the reliability of the results and offer valuable information about the types of traffic event sources. We use the data generated by the examined 29 traffic events to estimate their speed by applying the tracking method in section 2.2. The estimated speeds are listed in Table S1 and their distribution is shown in Figure S6. The results consist of two main clusters with velocities below 70 km/h and above 100 km/h, corresponding to freight trains and trucks, respectively. This is consistent with direct estimates of velocities conducted during a field observation near the highway and railway. We do not observe a significant difference in amplitude between waveforms induced by these two types of sources. The amplitudes are combined effects of the weight and speed of vehicles, along with the coupling between railway/highway to the ground, and can not be used alone to separate between freight trains and trucks.

3.2 Shallow Rayleigh wave velocities

We apply the method described in section 2.4 to 150 time windows from 29 observed freight train and truck events. To ensure robustness, we only consider the resolved velocity of a layer between two nodes for further analysis if the waveforms recorded by these two nodes have a CC value greater than 0.7 within the analyzed window. The quality controlled results are shown in Figure 4 as blue dots; we average these values to estimate the Rayleigh wave velocity along the linear array and use ± one standard deviation as the measurement uncertainties. Since the locations of vehicles have a good coverage of the route, heterogeneities parallel to the fault direction (if any) should be averaged. The result shown in Figure 7 is over 2-5 Hz; we also perform a similar analysis to bandpass filtered waveforms over 2-3 Hz, 3-4 Hz, and 4-5 Hz and obtain similar resolved Rayleigh wave velocities (Figure S7). We do not observe clear dispersion effect in this frequency range, which corresponds to a depth range of 50-100 m (Figure S3).

The results show clear horizontal variations of Rayleigh wave velocities along the linear array across the BF and the MCF. Shallow Rayleigh wave velocities fluctuate slightly around 0.8 km/s near the BF and gradually increase to 1.1 km/s on the SW side of the MCF. After crossing the MCF, the Rayleigh wave velocity drops abruptly by about 35% to 0.7 km/s on the NE side, indicating that the MCF is a sharp bimaterial interface in the top crust. In contrast, the BF does not appear to be associated with a significant velocity contrast in the shallow crust.

3.3 Low Q-values around the BF and MCF

We first normalize the RMS amplitudes of each event on the linear array by the maximum amplitude among all nodes and then take an average for all 29 events. To suppress the site effects of different nodes, we smooth the amplitude along the linear profile using a moving average with a length of 3 nodes. We observe a decrease in recorded amplitudes from SW to NE along the linear array profile, a trend that persists after applying a distance correction to remove the effects of geometrical spreading (Figure S8). Given the sudden increase in amplitude across the MCF (Figure 2b), which is not caused by attenuation, we apply the method described in section 2.5 separately to the two sides of the MCF and then combine the results together to obtain Q-values along the entire linear array. The final

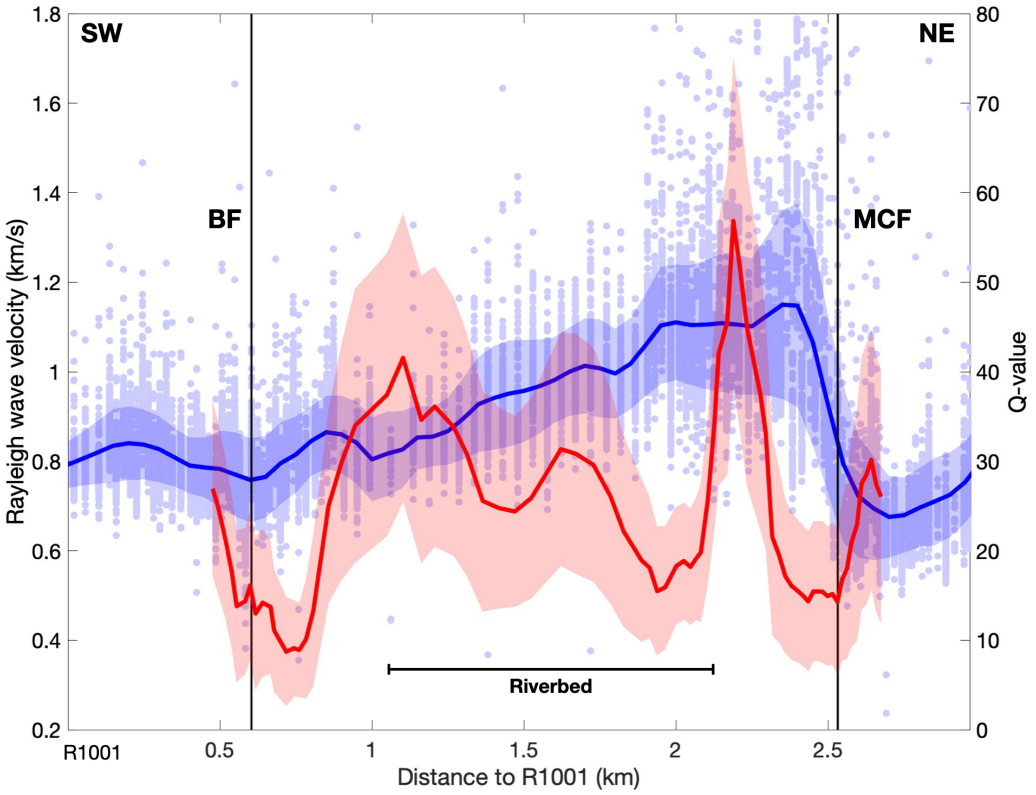


Figure 4. **(a)** The resolved velocities of Rayleigh waves (bold blue curve) and Q-values (bold red curve) along the linear array (starting from sensor R1001), both with shading for $\pm$ one standard deviation of the measurement uncertainty. Each blue dot represents the velocity estimated using the waveform in a given time window of a traffic event. The average values of dots with $CC > 0.7$ provide estimates for the velocities along the linear array. The resloved shallow Q-values across the SSAF range from 10 to 40, with low values around the two fault strands suggesting the presence of possible damage zones. The location of the Banning Fault (BF) and Mission Creek Fault (MCF) are marked by the black lines.

results are shown in Figure 4, along with uncertainties derived from measurement error in the Rayleigh wave velocities used in the inversion.

The resolved Q-values range from 10 to 40 with a mean value of 20. These values are consistent with previous derivations of attenuation factors of the San Jacinto fault zone from analysis of traffic signals within a similar frequency range (Meng et al., 2021; Zhao et al., 2023). As expected, The damage zones around the two strands are characterized by strong attenuation, with Q-values as low as 10, in agreement with results based on modeling of trapped waves in several fault and rupture zones (Peng et al., 2003; Lewis & Ben-Zion, 2010; Qiu et al., 2017; Qin et al., 2021). We also observe low Q-values in the riverbed between the BF and the MCF, which is consistent with the strong attenuation in the unconsolidated sediments of the shallow structure (Figure 1b).

4 Discussion and Conclusions

We present a detailed analysis of subsurface seismic properties across the Banning and Mission Creek strands of the Southern San Andreas Fault in the Coachella Valley, CA, using traffic-induced seismic signals recorded by a dense nodal array. The resolved Rayleigh wave velocities exhibit slight variations across the Banning Fault, gradual increases from BF to MCF, and a significant reduction of 35% across the MCF on its northeastern side (Figure 4). We observe a steep change in the amplitude of the wavefield across the MCF due to the interaction of the wavefield with the property contrast across the fault, providing evidence of a sharp bimaterial interface in the shallow structure of the MCF. Our study also reveals that the Q-values fluctuate across the entire fault zone, ranging from 10 to 40, consistent with values previously derived from studies of attenuation coefficients of shallow and/or fault zone materials.

The 35% reduction in velocity across the MCF is consistent with previous studies of velocity contrasts across faults at shallow depths in other segments of the San Andreas Fault (SAF). For instance, Ben-Zion et al. (1992) and Lewis et al. (2007) found shallow P-wave velocity contrasts of 15% and 50% across the SAF at Parkfield and south of Hollister, respectively. The observed bimaterial interface across the MCF may also correspond to the boundary of the fault with a low velocity damage zone, as found for several locations along the San Jacinto fault (e.g., Qiu et al., 2017; Qin et al., 2021). However, the limited extent of the linear array on the NE side of the MCF precludes a determination of whether

the observed Rayleigh wave velocity contrast across the MCF corresponds to an interface between two crustal blocks or between the fault and a damage zone.

The results summarized in Figure 7 can be used to derive changes of the mass density and shear rigidity across the MCF. For a narrowly bandpass waveform (2-5 Hz), the amplitude ratios of the incident wave A_0, reflected wave A_R and transmitted wave A_T are given by (Aki & Richards, 2002):

$$\frac{A_R}{A_0} = \frac{\rho_1 v_1 \cos j_1 - \rho_2 v_2 \cos j_2}{\rho_1 v_1 \cos j_1 + \rho_2 v_2 \cos j_2} \tag{11}$$

$$\frac{A_T}{A_0} = \frac{2\rho_1 v_1 \cos j_1}{\rho_1 v_1 \cos j_1 + \rho_2 v_2 \cos j_2} \tag{12}$$

where ρ, v and j represent the mass densities, wave velocities, and angles, respectively, and subscripts 1 and 2 mark the incident (SW) and transmission (NE) sides of the wavefield. Considering the phase change of the reflected wave on the interface, the amplitude at the node next to the interface from the SW side is $A_1 = A_0 - A_R$. Thus, the contrast of amplitude at nodes next to the interface on the NE and SW sides should be

$$\frac{A_T}{A_1} = \frac{\rho_1 v_1 \cos j_1}{\rho_2 v_2 \cos j_2} \tag{13}$$

To analyze data with equation (13), we first divide the 2D array around the MCF to two parts separated by the fault interface and then derive wave velocities and incident/transmission angles on the different sides (Figure 2b). The obtained incident and transmission velocities are 1.08 km/s and 0.73 km/s, while the incident and transmission angles are 15° and 10°. These values satisfy the relations

$$\frac{v_2}{v_1} \approx 0.68, \ \frac{\sin j_2}{\sin j_1} \approx 0.65 \tag{14}$$

The results correspond well to Snell's law and support the stability of the analysis. The density contrast across the bimaterial interface is

$$\frac{\rho_2}{\rho_1} = \frac{A_1 v_1 \cos j_1}{A_2 v_2 \cos j_2} \approx 0.9 \tag{15}$$

The relations between the derived velocities and mass densities are consistent with the empirical relationship of Brocher (2005) based on borehole data in California and ultrasonic laboratory measurements. For a Poisson solid, the phase velocity of Rayleigh wave $v \approx 0.92\beta$ where $\beta = \sqrt{\mu/\rho}$ is the shear wave velocity with being the shear rigidity. Assuming that the contrast of shear wave velocities across the MCF is 35%, same as for Rayleigh waves (Figure 4), along with the 0.9 density ratio in equation (15), the ratio of the shallow shear

moduli across the MCF is estimated to be

$$\frac{\mu_2}{\mu_1} = \frac{\beta_2^2 \rho_2}{\beta_1^2 \rho_1} \approx 0.53 \tag{16}$$

While the MCF is associated in the top crust with strong changes of seismic velocities, Q-values, and mass densities, the changes across the BF are minor. These observations suggest that the MCF is the main active strand of the SSAF in the area, rather than the BF, in agreement with previous studies by Blisniuk et al. (2021) and Vavra et al. (2021). Earthquake ruptures on the MCF to the northwest are expected to produce less shaking in the heavily populated area south of the San Bernardino and San Gabriel mountains relative to comparable ruptures on the BF that is closer to the population centers.

The subsurface location of the MCF is well resloved by the strong changes in material properties and agrees with the mapped surface trace of the fault. In contrast, the subsurface location of the BF is not well constrained by our observations and may be offset from the surface trace. Additional studies are needed to better constrain the location of the BF below the surface and to image the depth variations of the SSAF structure in the study area at great depths. Several methods could be used to achieve these goals, including utilizing surface and fault zone phases derived from the ambient seismic noise (e.g., Hillers et al., 2014; Mordret et al., 2019; Zigone et al., 2019), using body waves generated by heavy vehicles (Brenguier et al., 2019), and analyzing the scattered wavefield below the array (Touma et al., 2022). Some of these analyses will be conducted in follow-up studies to further enhance our understanding of the subsurface structure of the Southern San Andreas Fault in the Coachella Valley.

Acknowledgments

We thank Wei Wang for useful discussions. This study is supported by the U.S. Department of Energy Office of Science (Award ED-SC0016520). Haoran Meng is supported by the Shenzhen Natural Science Foundation (K23436305) and Research Startup Fund (Y01436128).

Open Research

The nodal array data and associated metadata (Vernon et al., 2020) are freely available for download from the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC) via `https://doi.org/10.7914/SN/YA_2020`.

References

Ajala, R., Persaud, P., Stock, J. M., Fuis, G. S., Hole, J. A., Goldman, M., & Scheirer, D. (2019). Three-dimensional basin and fault structure from a detailed seismic velocity model of Coachella Valley, Southern California. *Journal of Geophysical Research: Solid Earth*, *124*(5), 4728–4750. doi: 10.1029/2018JB016260

Aki, K., & Richards, P. G. (2002). *Quantitative seismology.*

Allam, A. A., Ben-Zion, Y., Kurzon, I., & Vernon, F. (2014). Seismic velocity structure in the Hot Springs and Trifurcation areas of the San Jacinto fault zone, California, from double-difference tomography. *Geophysical Journal International*, *198*(2), 978-999. doi: 10.1093/gji/ggu176

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics*, *46*(4). doi: 10.1029/2008RG000260

Ben-Zion, Y., Katz, S., & Leary, P. (1992). Joint inversion of fault zone head waves and direct P arrivals for crustal structure near major faults. *Journal of Geophysical Research: Solid Earth*, *97*(B2), 1943-1951. doi: 10.1029/91JB02748

Ben-Zion, Y., & Sammis, C. G. (2003). Characterization of Fault Zones. *Pure and Applied Geophysics*, *160*(3), 677-715. doi: 10.1007/PL00012554

Ben-Zion, Y., Vernon, F. L., Ozakin, Y., Zigone, D., Ross, Z. E., Meng, H., ... Barklage, M. (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophysical Journal International*, *202*(1), 370-380. doi: 10.1093/gji/ggv142

Blisniuk, K., Scharer, K., Sharp, W. D., Burgmann, R., Amos, C., & Rymer, M. (2021). A revised position for the primary strand of the Pleistocene-Holocene San Andreas fault in southern California. *Science Advances*, *7*(13). doi: 10.1126/sciadv.aaz5691

Brenguier, F., Boué, P., Ben-Zion, Y., Vernon, F., Johnson, C., Mordret, A., ... Lecocq, T. (2019). Train Traffic as a Powerful Noise Source for Monitoring Active Faults With Seismic Interferometry. *Geophysical Research Letters*, *46*(16), 9529–9536. doi: 10.1029/2019GL083438

Brocher, T. M. (2005). Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust. *Bulletin of the Seismological Society of America*, *95*(6), 2081-2092. doi: 10.1785/0120050077

Díaz, J., DeFelipe, I., Ruiz, M., Andrés, J., Ayarza, P., & Carbonell, R. (2022). Identifi-

cation of natural and anthropogenic signals in controlled source seismic experiments. *Scientific Reports*, *12*(1), 3171. doi: 10.1038/s41598-022-07028-3

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., ... Zeng, Y. (2014). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—The Time-Independent Model. *Bulletin of the Seismological Society of America*, *104*(3), 1122-1180. doi: 10.1785/0120130164

Fuchs, F., Bokelmann, G., & the AlpArray Working Group. (2017). Equidistant Spectral Lines in Train Vibrations. *Seismological Research Letters*, *89*(1), 56–66. doi: 10.1785/0220170092

Hillers, G., Campillo, M., Ben-Zion, Y., & Roux, P. (2014). Seismic fault zone trapped noise. *Journal of Geophysical Research: Solid Earth*, *119*(7), 5786-5799. doi: 10.1002/2014JB011217

Inbal, A., Cristea-Platon, T., Ampuero, J., Hillers, G., Agnew, D., & 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. doi: 10.1785/0120180130

Jiang, Y., Ning, J., Wen, J., & Shi, Y. (2022). Doppler effect in high-speed rail seismic wavefield and its application. *Science China Earth Sciences*, *65*(3), 414-425. doi: 10.1007/s11430-021-9843-0

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. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8519-8539. doi: 10.1029/2018JB017151

Jones, L. M., Bernknopf, R., Cox, D., Goltz, K., James amd Hudnut, Mileti, D., Perry, S., ... Wein, A. (2008). The ShakeOut Scenario: U.S. Geological Survey Open-File Report 2008-1150 and California Geological Survey Preliminary Report 25. *U.S. Geological Survey*.

Lewis, M. A., & Ben-Zion, Y. (2010). Diversity of fault zone damage and trapping structures in the Parkfield section of the San Andreas Fault from comprehensive analysis of near fault seismograms. *Geophysical Journal International*, *183*(3), 1579-1595. doi: 10.1111/j.1365-246X.2010.04816.x

Lewis, M. A., Ben-Zion, Y., & McGuire, J. J. (2007). Imaging the deep structure of the San Andreas Fault south of Hollister with joint analysis of fault zone head and

direct P arrivals. *Geophysical Journal International*, *169*(3), 1028-1042. doi: 10.1111/j.1365-246X.2006.03319.x

Li, C., Li, Z., Peng, Z., Zhang, C., Nakata, N., & Sickbert, T. (2018). Long-Period Long-Duration Events Detected by the IRIS Community Wavefield Demonstration Experiment in Oklahoma: Tremor or Train Signals? *Seismological Research Letters*, *89*(5), 1652-1659. doi: 10.1785/0220180081

Meng, H., Ben-Zion, Y., & Johnson, C. W. (2019). Detection of random noise and anatomy of continuous seismic waveforms in dense array data near Anza California. *Geophysical Journal International*, *219*(3), 1463-1473. doi: 10.1093/gji/ggz349

Meng, H., Ben-Zion, Y., & Johnson, C. W. (2021). Analysis of Seismic Signals Generated by Vehicle Traffic with Application to Derivation of Subsurface Q-Values. *Seismological Research Letters*, *92*(4), 2354–2363. doi: 10.1785/0220200457

Mordret, A., Roux, P., Boué, P., & Ben-Zion, Y. (2019). Shallow three-dimensional structure of the San Jacinto fault zone revealed from ambient noise imaging with a dense seismic array. *Geophysical Journal International*, *216*(2), 896-905. doi: 10.1093/gji/ggy464

Peng, Z., Ben-Zion, Y., Michael, A. J., & Zhu, L. (2003, 12). Quantitative analysis of seismic fault zone waves in the rupture zone of the 1992 Landers, California, earthquake: evidence for a shallow trapping structure. *Geophysical Journal International*, *155*(3), 1021-1041. doi: 10.1111/j.1365-246X.2003.02109.x

Pinzon-Rincon, L., Lavoué, F., Mordret, A., Boué, P., Brenguier, F., Dales, P., ... Hollis, D. (2021). Humming Trains in Seismology: An Opportune Source for Probing the Shallow Crust. *Seismological Research Letters*, *92*(2A), 623–635. doi: 10.1785/0220200248

Qin, L., Share, P.-E., Qiu, H., Allam, A. A., Vernon, F. L., & Ben-Zion, Y. (2021). Internal structure of the San Jacinto fault zone at the Ramona Reservation, north of Anza, California, from dense array seismic data. *Geophysical Journal International*, *224*(2), 1225-1241. doi: 10.1093/gji/ggaa482

Qiu, H., Ben-Zion, Y., Catchings, R., Goldman, M. R., Allam, A. A., & Steidl, J. (2021). Seismic imaging of the Mw 7.1 Ridgecrest earthquake rupture zone from data recorded by dense linear arrays. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2021JB022043. doi: 10.1029/2021JB022043

Qiu, H., Ben-Zion, Y., Ross, Z., Share, P.-E., & Vernon, F. (2017). Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array. *Geophysical Journal International*, *209*(3), 1369-1388. doi: 10.1093/gji/ggx096

Rezaeifar, M., Lavoué, F., Maggio, G., Xu, Y., Bean, C. J., Pinzon-Rincon, L., ... Brenguier, F. (2023). Imaging shallow structures using interferometry of seismic body waves generated by train traffic. *Geophysical Journal International*, *233*(2), 964-977. doi: 10.1093/gji/ggac507

Share, P.-E., Qiu, H., Vernon, F. L., Allam, A. A., Fialko, Y., & Ben-Zion, Y. (2022). General Seismic Architecture of the Southern San Andreas Fault Zone around the Thousand Palms Oasis from a Large-N Nodal Array. *The Seismic Record*, *2*(1), 50–58. doi: 10.1785/0320210040

Share, P.-E., Tábořík, P., Štěpančíková, P., Stemberk, J., Rockwell, T. K., Wade, A., ... Ben-Zion, Y. (2020). Characterizing the uppermost 100 m structure of the San Jacinto fault zone southeast of Anza, California, through joint analysis of geological, topographic, seismic and resistivity data. *Geophysical Journal International*, *222*(2), 781-794. doi: 10.1093/gji/ggaa204

Shaw, J. H., Plesch, A., Tape, C., Suess, M. P., Jordan, T. H., Ely, G., ... Munster, J. (2015). Unified Structural Representation of the southern California crust and upper mantle. *Earth and Planetary Science Letters*, *415*, 1-15. doi: 10.1016/j.epsl.2015.01.016

Sheng, Y., Mordret, A., Sager, K., Brenguier, F., Boué, P., Rousset, B., ... Ben-Zion, Y. (2022). Monitoring Seismic Velocity Changes Across the San Jacinto Fault Using Train-Generated Seismic Tremors. *Geophysical Research Letters*, *49*(19), e2022GL098509. doi: 10.1029/2022GL098509

Stierman, D. J. (1984). Geophysical and geological evidence for fracturing, water circulation and chemical alteration in granitic rocks adjacent to major strike-slip faults. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 5849-5857. doi: 10.1029/JB089iB07p05849

Touma, A., R.and Aubry, Ben-Zion, Y., & Campillo, M. (2022). Distribution of seismic scatterers in the san jacinto fault zone, southeast of anza, california, based on passive matrix imaging. *Earth and Planetary Science Letters*, *578*, 117304. doi: 10.1016/j.epsl.2021.117304

Vavra, E., Qiu, H., Chi, B., Share, P.-E., Allam, A., Morzfeld, M., ... Fialko, Y. (2021). Seismo-Geodetic Investigations of Subsurface Properties of the Southern San Andreas Fault. In *AGU Fall Meeting* (p. S51D-03).

Vernon, F., Share, P.-E., Ben-Zion, Y., Fialko, Y., & Allam, A. (2020). *Southern San*

Andreas Fault Zone. International Federation of Digital Seismograph Networks. doi: 10.7914/SN/YA_2020

Zhao, Y., Nilot, E. A., Li, B., Fang, G., Luo, W., & Li, Y. E. (2023). Seismic Attenuation Extraction From Traffic Signals Recorded by a Single Seismic Station. *Geophysical Research Letters*, *50*(3), e2022GL100548. doi: 10.1029/2022GL100548

Zigone, D., Ben-Zion, Y., Lehujeur, M., Campillo, M., Hillers, G., & Vernon, F. L. (2019). Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays. *Geophysical Journal International*, *217*(2), 879-893. doi: 10.1093/gji/ggz069

Supporting Information for "Lateral Variations Across the Southern San Andreas Fault Zone Revealed from Analysis of Traffic Signals at a Dense Seismic Array"

Hao Zhang[1], Haoran Meng[2] and Yehuda Ben-Zion[1,3]

[1]University of Southern California, Los Angeles, California, USA.
[2]Southern University of Science and Technology, Shenzhen, China.
[3]Southern California Earthquake Center, Los Angeles, California, USA.

Contents of this file:

1. Text S1
2. Figures S1 – S8
3. Table S1

Text S1: Particle motion analysis of waveforms

To clarify the character of observed signals, we first apply particle motion analysis using data from the event shown in Figure 2. We improve the SNR by shifting waveforms recorded by a 2D array on a BF using offsets determined by slowness solved with a method described in section 2.3. The aligned waveforms are then linearly stacked with equal weights. Figure S4 shows different components of the stacked waveform and particle motion diagrams within a 1-second window around reference zero time. The horizontal polarization is linear and aligns with wave propagation direction while vertical-radial particle motion follows a retrograde elliptical shape, indicating that the waveform is a Rayleigh wave. This analysis is repeated for additional events with various moving speeds and results show similar particle motion features (Figure S5).

Figure S1. **(a)** The spectrogram of vertical waveform recorded by node R1001 from 2020-03-07T17:15:00 to 2020-03-07T17:45:00. The vertical light blue stripes with energy from 2-35 Hz are traffic events on local road (Figure 1b). The horizontal strip with energy from 2-5 Hz, marked by the red rectangle, corresponds to a freight truck on the highway I-10. **(b)** The corresponding waveform bandpass filtered from 1 to 50 Hz. **(c)** The same waveform but bandpass filtered from 1 to 5 Hz. **(d)** The waveform of local car events (within the red rectangle in (b)) recorded by the linear array.

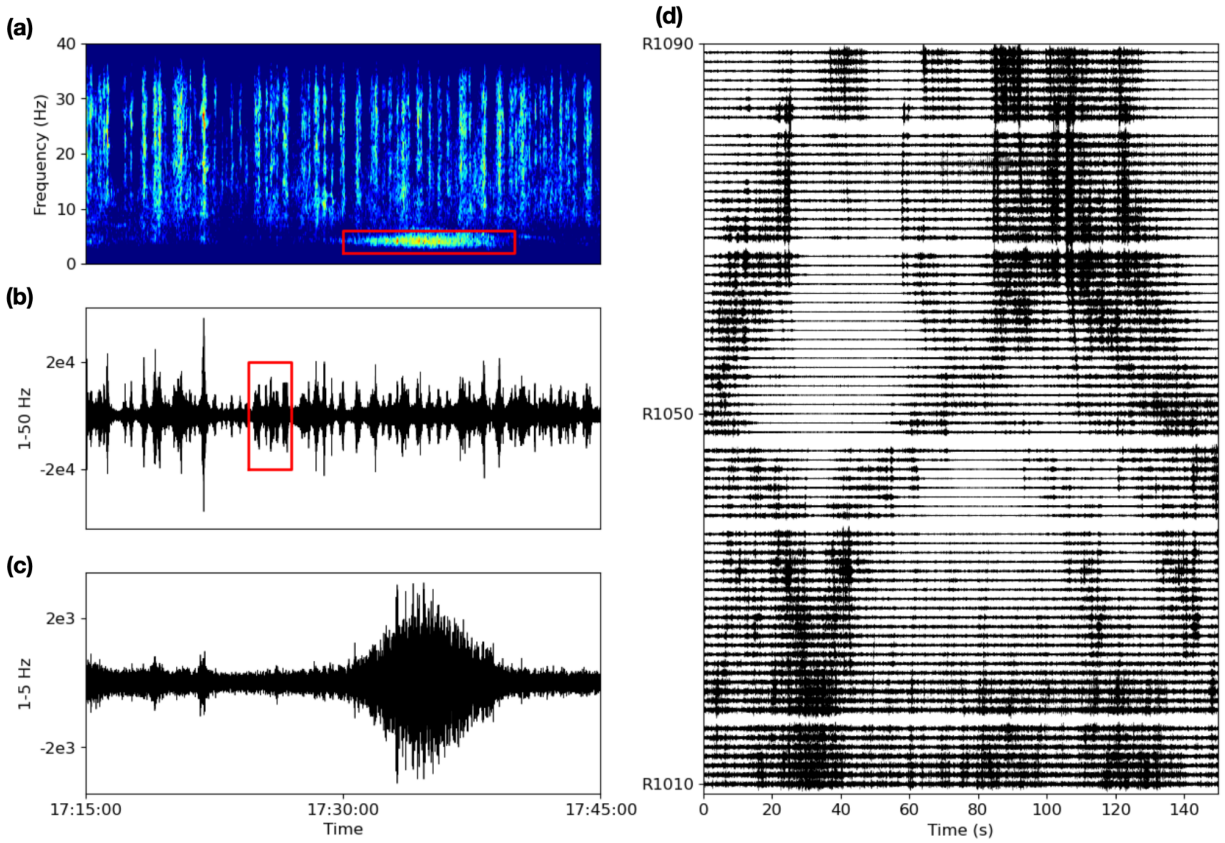


Figure S2. Tracking the moving vehicle source in event E1 using 2D arrays by resolving the propagation direction of Rayleigh wave. **(a,b)** Lag time distribution and wave propagation directions for the example time window in Figure 2a on the 2D subarrays centered on MCF and BF, respectively. Lag times are computed using waveform cross-correlations of a reference node and its neighbors. The Rayleigh wave propagation direction is determined by the opposite direction to the gradient of the lag times at the neighbors of all reference nodes at the center. **(c)** The location of the vehicle, denoted by the red star, is determined by the intersection of the railway and the wave propagation direction. The corresponding timing is determined by subtracting from the reference time the wave propagation time from source to node. The reference center of the railway denoted by the red dot is the closest point to the center of the subarray on BF. **(d)** The azimuth from the vehicle to the two subarrays on MCF and BF, and the distance from the vehicle to the center of the railway. The estimated speed of the traffic is 100 km/h.

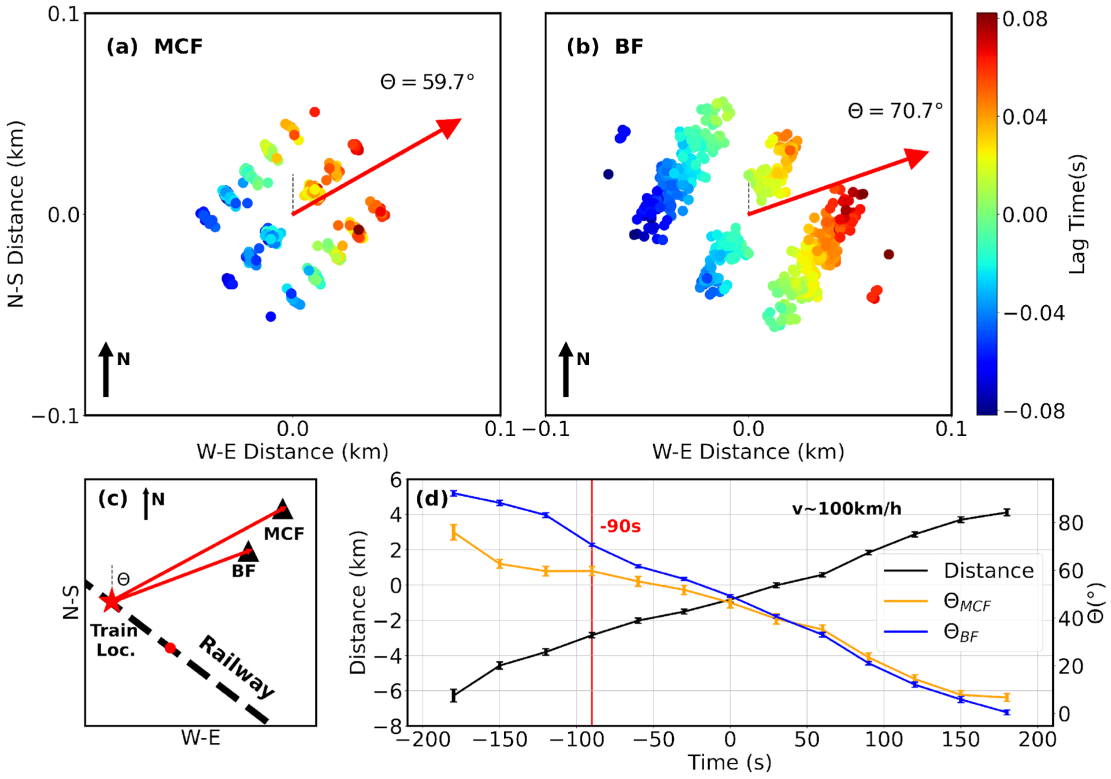


Figure S3. Sensitivity kernels of Rayleigh waves at 3 Hz (gray), 4 Hz (blue) and 5 Hz (red). The kernels are calculated using a 1D average version of the CVM-S4.26-M1 model in the study area, which includes a geotechnical layer representing seismic properties from surface to about 350m depth.

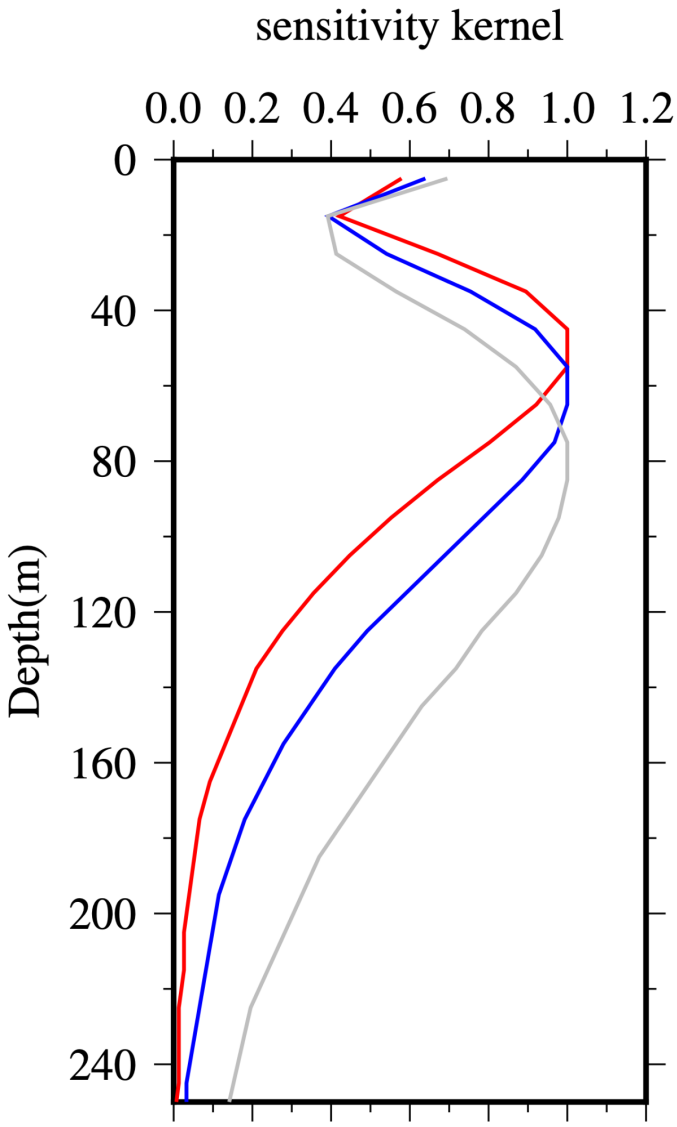


Figure S4. Particle motion analysis for the event in Figure 2 using the waveforms recorded by the 2D subarray in the frequency range 2-5 Hz. The waveforms are aligned and stacked to improve the SNR. The upper left and right panels show, respectively, the stacked waveforms in the vertical (Z), north-south (NS), east-west (EW), radial (R), and tangential (T) directions. The lower panels present particle motions for different pairs of components color-coded by time. The wave propagation direction is indicated by a red arrow in the horizontal particle motion diagram.

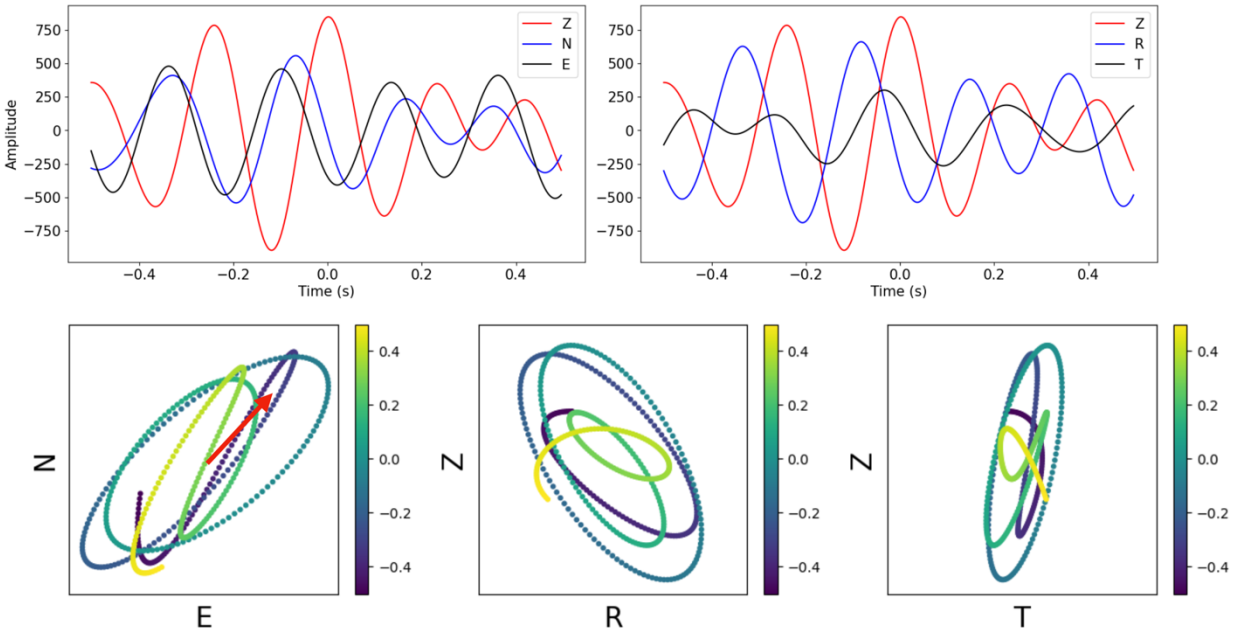


Figure S5. Particle motion analysis for Traffic event 20, similar to Figure S4.

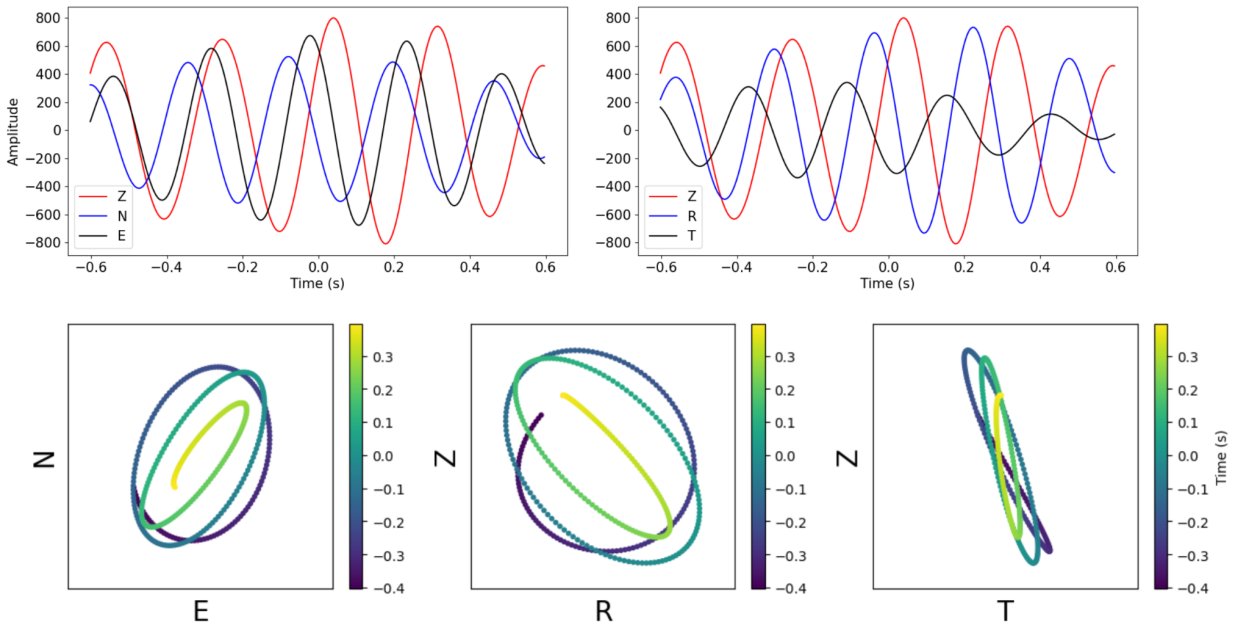


Figure S6. The distribution of speeds for the 29 analyzed traffic events. Fifteen signal sources have speeds lower than 85 km/h and are likely to be freight trains. The remaining fourteen events have speeds higher than 85 km/h and may be trucks on the highway.

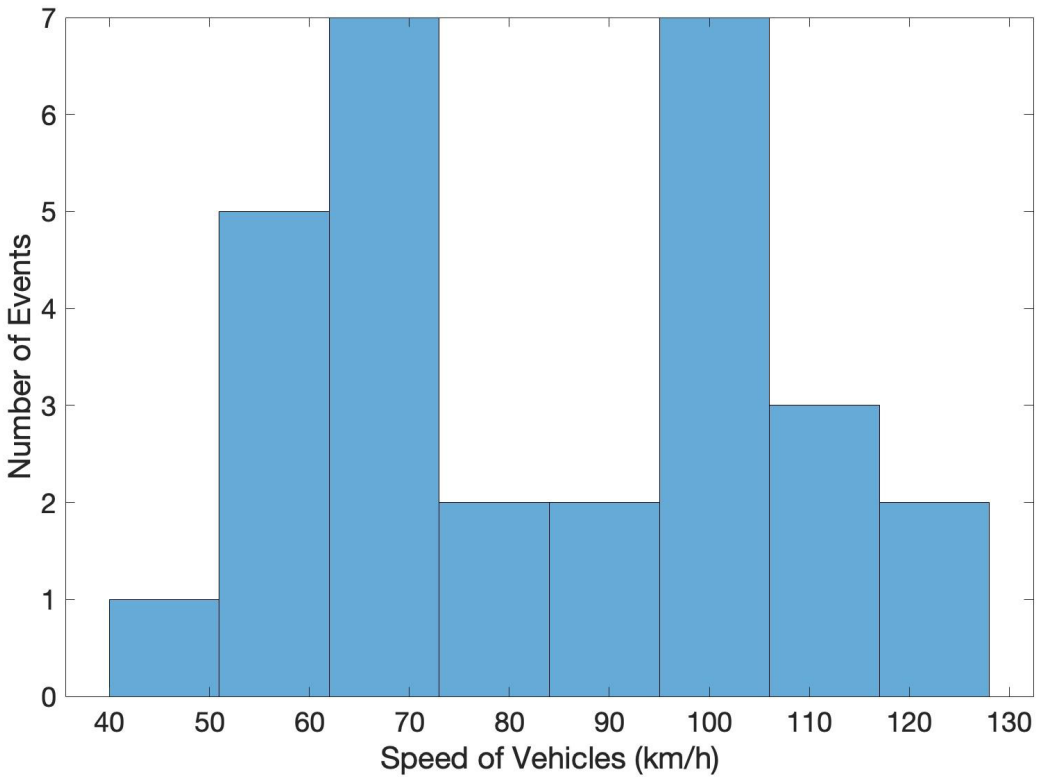


Figure S7. Rayleigh wave velocities resolved for different frequency bands. There is no clear dispersion observed in the frequency range of 2-5 Hz.

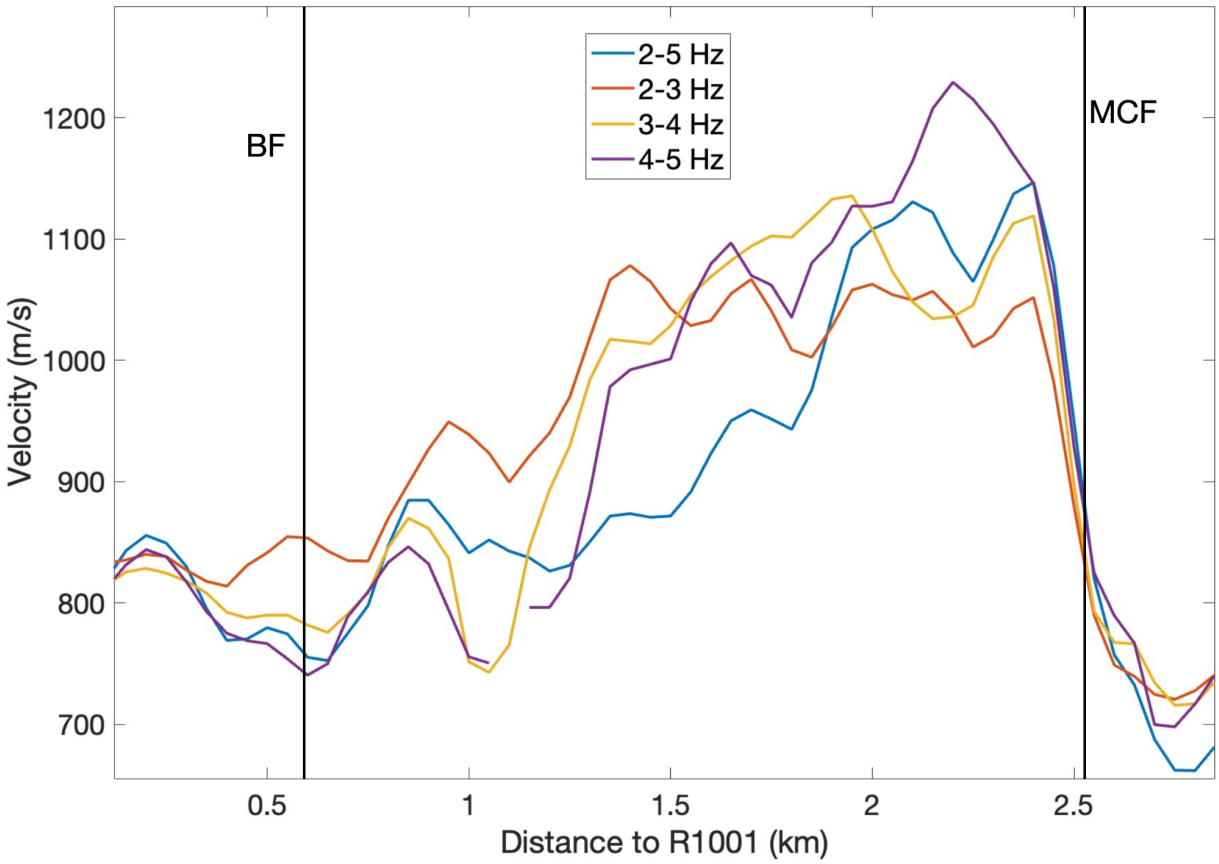


Figure S8. The RMS of signals at different nodes with (red curve) and without (black curve) correction of geometrical spreading. Each curve is normalized by the maximum RMS among all nodes.

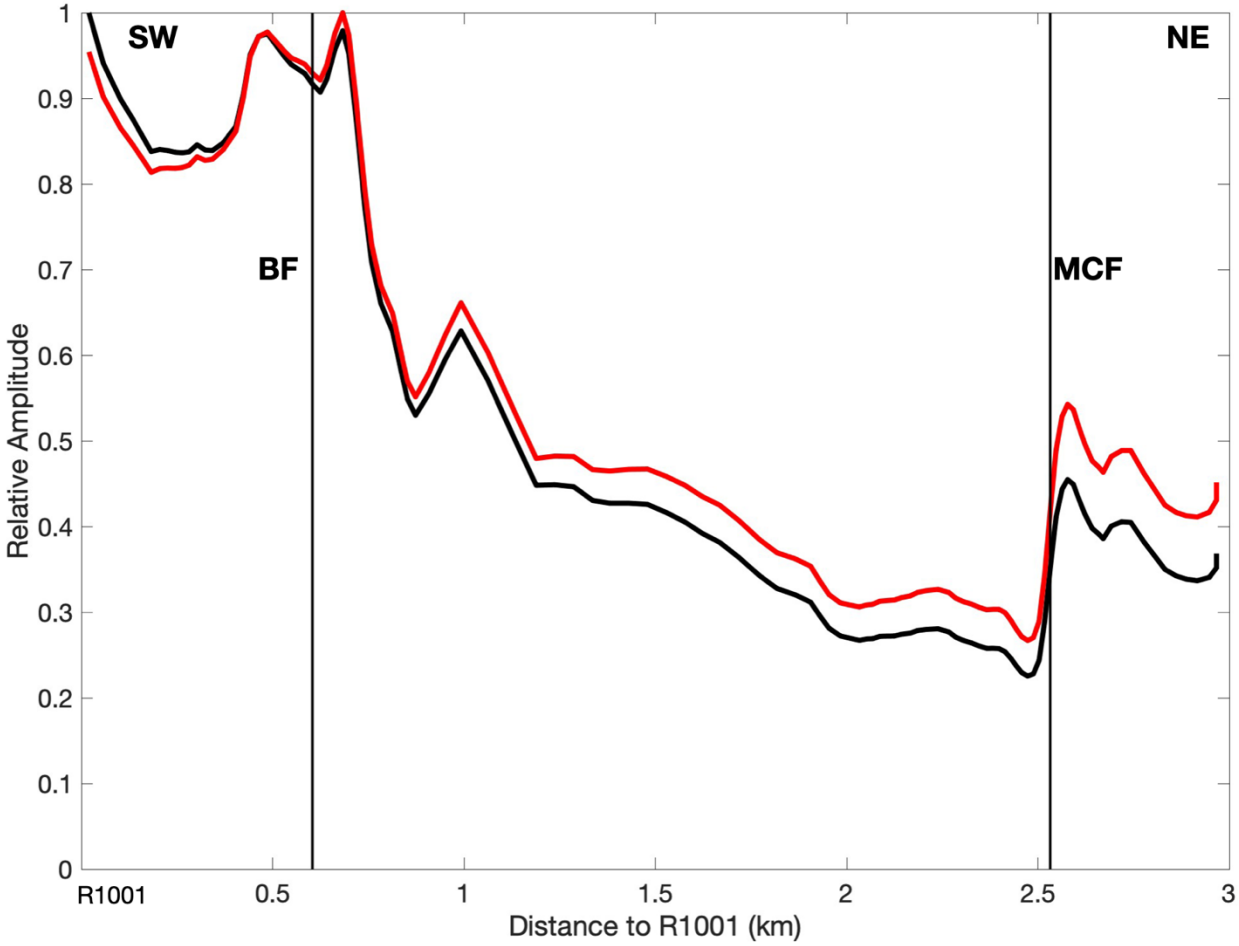


Table S1. Time and speed of 29 analyzed freight train or truck events on the railway or I-10.

Event Index	Time	Speed (km/h)
1	2020-03-06T11:00:30	61.14
2	2020-03-06T12:52:10	88.52
3	2020-03-06T13:56:40	72.56
4	2020-03-06T19:45:00	70.21
5	2020-03-06T20:42:30	94.78
6	2020-03-06T22:54:10	65.37
7	2020-03-07T00:19:10	96.01
8	2020-03-07T02:45:40	96.53
9	2020-03-07T08:31:40	65.82
10	2020-03-07T00:23:10	61.53
11	2020-03-07T09:53:00	56.96
12	2020-03-07T10:37:00	96.99
13	2020-03-07T14:46:40	82.23
14	2020-03-07T17:34:40	107.27
15	2020-03-07T19:51:20	100.74
16	2020-03-07T21:09:40	121.41
17	2020-03-08T02:37:30	61.23
18	2020-03-08T04:20:20	67.30
19	2020-03-08T07:05:30	104.54
20	2020-03-08T09:26:00	42.51
21	2020-03-08T10:43:50	67.36
22	2020-03-08T11:08:00	82.44
23	2020-03-08T11:30:00	96.49
24	2020-03-08T12:10:00	55.92
25	2020-03-08T15:53:20	105.54
26	2020-03-08T18:46:00	115.69
27	2020-03-08T20:20:00	116.91
28	2020-03-08T20:44:00	72.24
29	2020-03-08T22:45:20	113.90