

Body waves from train noise correlations: potential and limits for monitoring the San Jacinto Fault, CA

Quentin Higuieret¹,¹ Yixiao Sheng,¹ Aurelien Mordret¹,¹ Florent Brenguier¹,¹ Pierre Boué,¹ Andreas Fichtner²,² Frank Vernon,³ Lion Krischer,⁴ Dan Hollis,³ Coralie Aubert¹ and Yehuda Ben-Zion⁵

¹*Institut des Sciences de la Terre, Université Grenoble Alpes, 38610, Gières, France. E-mail: quentin.higuieret@univ-grenoble-alpes.fr*

²*ETH Zurich, Department of Earth Sciences, Institute of Geophysics, 8092, Zurich, Switzerland*

³*Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92037, USA*

⁴*Mondiac Ltd., 8005, Zurich, Switzerland*

⁵*Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089, USA*

Accepted 2024 November 12. Received 2024 October 16; in original form 2024 May 14

SUMMARY

A large portion of the stress release on seismic faults remains silent and undetected, requiring the development of novel observation techniques. Measuring traveltimes perturbations from the correlation of ambient seismic noise at different stations is a well-known approach to assess temporal changes in seismic velocities, which can provide insights into hydrologic, tectonics and volcanic dynamic processes. In this work, we study the specific case of a *P*-wave phase retrieved from the correlation of freight train noise in Southern California and evaluate its potential to detect localized velocity changes along the San Jacinto Fault. We use a full waveform modelling approach to simulate this *P*-wave interference and further assess its sensitivity to the position of the train source, near-surface velocity changes and localized velocity changes in the fault zone. Our results show that the uncertainty in trains location can induce large traveltimes biases which can be mitigated by averaging over many trains. Our results also highlight the weak sensitivity of these correlation *P* waves to near-surface velocity changes, while they show significant sensitivity to localized changes at depth. This modelling highlights the potential of monitoring traveltimes perturbations of this ballistic *P*-wave interference to detect hidden slow-slip events on the San Jacinto Fault, particularly in identifying subtle velocity anomalies associated with fault zone changes that may otherwise go unnoticed by conventional seismic monitoring techniques.

Key words: Numerical modelling; Seismic interferometry; Wave scattering and diffraction.

1 INTRODUCTION

Passive seismic velocity monitoring is widely applied in a variety of fields, including earthquake studies, non-destructive testing, groundwater monitoring, volcanic activity research and geothermal activities (Sens-Schönfelder & Wegler 2008; Brenguier *et al.* 2008; Taira *et al.* 2014; Rivet *et al.* 2014; Obermann *et al.* 2015; Larose *et al.* 2015; Mao *et al.* 2022; Sánchez-Pastor *et al.* 2023; Müller *et al.* 2023). By measuring traveltimes perturbations in the coda of noise correlations in different frequency ranging from typically 0.1 to a few Hz, a number of studies have observed seismic velocity drops and recoveries during and following large earthquakes (Wegler & Sens-Schönfelder 2007; Brenguier *et al.* 2008, 2014; Sens-Schönfelder & Wegler 2011; Taira *et al.* 2015; Okubo *et al.* 2024). However, a large portion of the stress accumulation along faults is released by aseismic processes such as slow-slip events,

and it remains challenging to detect these events because they barely produce seismic waves and detectable displacements at the Earth's surface.

Seismic velocities are known to be sensitive to strain perturbations (Yamamura *et al.* 2003; Sens-Schönfelder *et al.* 2019; Mao *et al.* 2019; Sheng *et al.* 2024), so there is potential to use passive seismic velocity monitoring and the sensitivity of seismic waves to seismogenic regions at a few kilometres depth to detect silent strain transients. Due to the long travel path of scattered waves, the extended spatial sensitivity of coda waves allows detecting weak homogeneous velocity perturbations (Obermann *et al.* 2013; Kanu & Snieder 2015; Margerin *et al.* 2016). However, coda wave interferometry makes it challenging to detect weak localized velocity changes such as those produced by slow transient fault movements.

In recent years, high-frequency (>1 Hz) ballistic body waves, particularly *P* waves, have been successfully retrieved from the

correlation of anthropogenic noise induced by car and train traffic at distances of a few to tens of kilometres (Nakata *et al.* 2015; Brenguier *et al.* 2019; Sheng *et al.* 2023). These ballistic *P* waves dive down to a few kilometres depth and show a localized sensitivity at depth because of their relatively high-frequency content, making them potentially effective for detecting velocity changes within fault zones close to seismogenic depths (Sager *et al.* 2021; Sheng *et al.* 2022).

This study focuses on the specific case of a *P*-wave phase retrieved from the correlation of freight train tremors in the San Jacinto Fault Zone (SJFZ) region (Inbal *et al.* 2018; Brenguier *et al.* 2019; Sheng *et al.* 2022, 2024). In a recent work, Sheng *et al.* (2022) identified a static traveltime change of this *P*-wave phase occurring in early 2014 that they interpreted as the effect of a slow-slip event on the locked portion of the Anza gap. In this work, we use numerical simulations to assess potential biases that could affect the measurement of traveltime perturbations for this *P* wave. We model body waves in noise correlations using full waveform simulations (van Manen *et al.* 2006; Tromp *et al.* 2010; Hanasoge 2013; Fichtner 2014; Sager *et al.* 2021) and compare these with real noise correlations obtained from an array of 300 seismometers deployed in 2022 at the Piñon Flat Observatory near Anza (Brenguier *et al.* 2020). Additionally, we use data from the permanent stations II.PFO and AZ.FRD located across the San Jacinto Fault which are part of the AZ network (Vernon 1982) and the II network (Scripps Institution of Oceanography 1986).

Two primary sources of bias are considered in this analysis: (1) the position of the moving freight train, the source of the *P* waves, and (2) the influence of near-surface velocity perturbations compared to those occurring at depth within the fault zone. The assessment of these biases is critical for improving the accuracy of traveltime measurements. Our results confirm that the traveltime perturbations induced by the moving source position is sufficiently small to preclude this bias as being at the origin of the 2014 static change observed by Sheng *et al.* (2022) and also show that the relatively small sensitivity to near-surface perturbations can explain the lack of seasonal velocity changes observed by Sheng *et al.* (2022).

2 NOISE CORRELATIONS ACROSS THE SAN JACINTO FAULT

The SJFZ is a highly active seismic region in Southern California. It hosts the trifurcation zone of the San Jacinto Fault (Fig. 1), responsible for generating over 10 per cent of the region's earthquakes (Ross *et al.* 2017). In 2018, a dense array of 300 stations (vertical 5 Hz geophones, network code 9K, Brenguier *et al.* 2018) was deployed at Piñon Flat for one month to evaluate the possibility of retrieving body waves in noise correlations for stations located across the SJFZ (Brenguier *et al.* 2019). In 2022 April, a similar array of three-component instruments was deployed at the same location for a longer acquisition duration of 3 yr to track velocity changes across the fault with improved accuracy (network code 7V, Brenguier *et al.* 2025).

In Fig. 2, we show the vertical-to-vertical (ZZ) noise cross-correlations between this array and a broad-band permanent station (FRD, network code AZ) for one month of data in 2022 October. All seismograms are downsampled to 40 Hz. To enhance the signal-to-noise ratio, we stack the correlations using distance bins of 10 m. Initially, each trace is detrended and filtered within the frequency bands of interest, either (0.5–1 Hz) or (4–6 Hz). Cross-correlations are then performed on 30-min time windows. Each trace undergoes

clipping at 3 standard deviations to mitigate spikes and other large-amplitude signals like earthquakes, followed by whitening within the target frequency band and 1-bit normalization to emphasize the phase information. Finally, the cross-correlation is derived using the cross-coherence between the two traces.

For the frequency range [4–6] Hz, we observe a body wave with an apparent velocity of 6 km s⁻¹ on the causal side of the correlations (Fig. 2, bottom panel). Arrivals on the causal part of the correlations correspond to the presence of a noise source on the array's east side. This source was previously identified as freight trains travelling in the Coachella Valley (Inbal *et al.* 2018; Brenguier *et al.* 2018; Sheng *et al.* 2022). For the frequency range of 0.5–1 Hz (Fig. 2, top panel), the correlograms display a strong body-wave arrival in the acausal part, with a moveout of approximately 5 km s⁻¹ (red arrow). This microseism body wave has been previously described (Roux *et al.* 2005; Zhang *et al.* 2009) and attributed to pressure oscillations at the ocean bottom originating from wind-induced swell interactions. However, the mechanism generating these low-frequency arrivals is not fully understood, and their source appears to be unstable over time. Therefore, in the following, we will focus our synthetic correlation modelling solely on the body waves generated by the train, as their generation process is better understood and their source is more stable. This approach will enhance our ability to assess the sensitivity of these arrivals to seismic velocity perturbations at depth, with a focus on monitoring the San Jacinto Fault.

3 MODELLING BODY WAVES IN NOISE CORRELATIONS

In order to compute synthetic noise correlations, we follow the strategy of Sager *et al.* (2021), based on previous work from, for example, Tromp *et al.* (2010), Hanasoge (2013) and Fichtner (2014). We consider spatially uncorrelated sources located only at the Earth's surface (freight trains). In order to evaluate the noise correlation wavefield between receivers 1 (virtual source) and 2, we follow this generic numerical scheme:

- (i) We simulate the generating wavefield by injecting an impulse function at the reference station (receiver 2) and record the wavefield at the noise source positions (blue patches in Fig. 1).
- (ii) We time-reverse the wavefield recorded at the noise source positions and convolve it with the power-spectral density distribution of the synthetic noise source.
- (iii) We simulate the propagation of this time-reversed, convolved wavefield and record it at receiver 1 to form the synthetic correlation function between receivers 1 and 2.

Following this methodology, the correlation function is formally defined as follows:

$$C_{ij}(x_1, x_2) = \int \underbrace{G_{i,n}(x_1, \xi)}_{\text{Correlation}} \underbrace{[G_{j,m}^*(x_2, \xi) S_{nm}(\xi)]}_{\text{Generating wavefield}} d\xi \quad (1)$$

This equation describes the cross-correlation resulting from wave propagation between two receivers at positions x_1 and x_2 . We have $S_{n,m}(\xi)$ as the power-spectral density distribution that describes the initial energy pattern of the ambient noise sources. The indices m and n are used to specify the orientation of the ambient noise sources that are being modelled where n corresponds to the noise source influencing the wavefield at x_1 and m corresponds to the source affecting the wavefield at x_2 . The Green's function $G_{i,n}(x_1, \xi)$ and its complex conjugate $G_{j,m}^*(x_2, \xi)$ characterize the medium's response

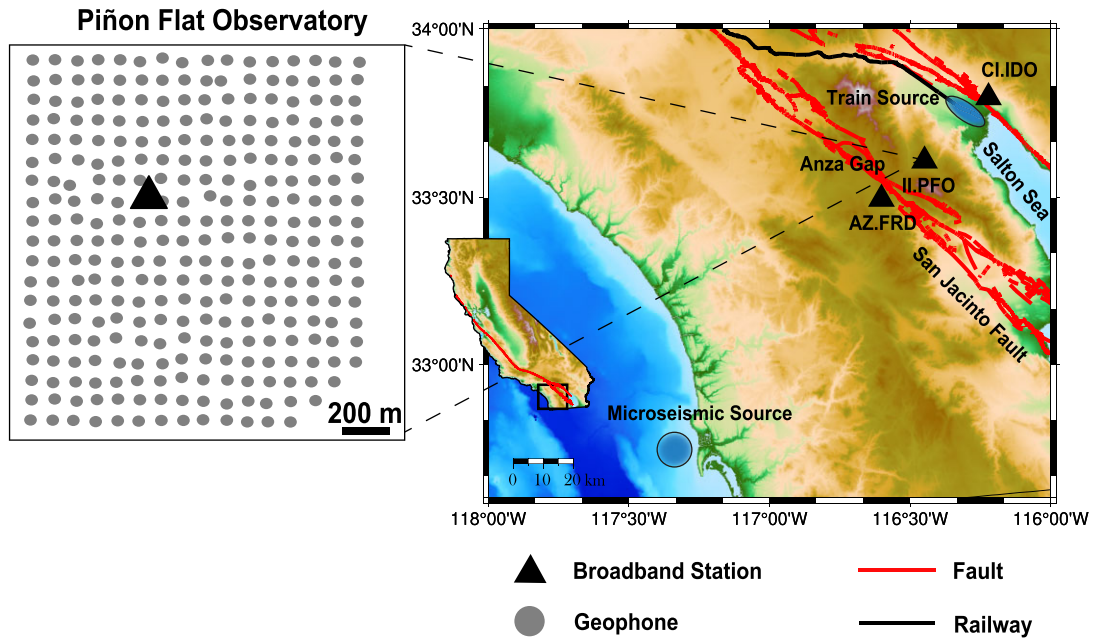


Figure 1. Map of the study area. The position of the seismic stations used in the study is shown with black triangles. The blue circle in the southwest highlights the location of the wind-driven noise source in the ocean. The black line shows the location of the freight train railways in the Coachella Valley, and the blue ellipse in the northeast delimits the position of the train noise source. The inset depicts the map of the dense nodal array of 300 geophones.

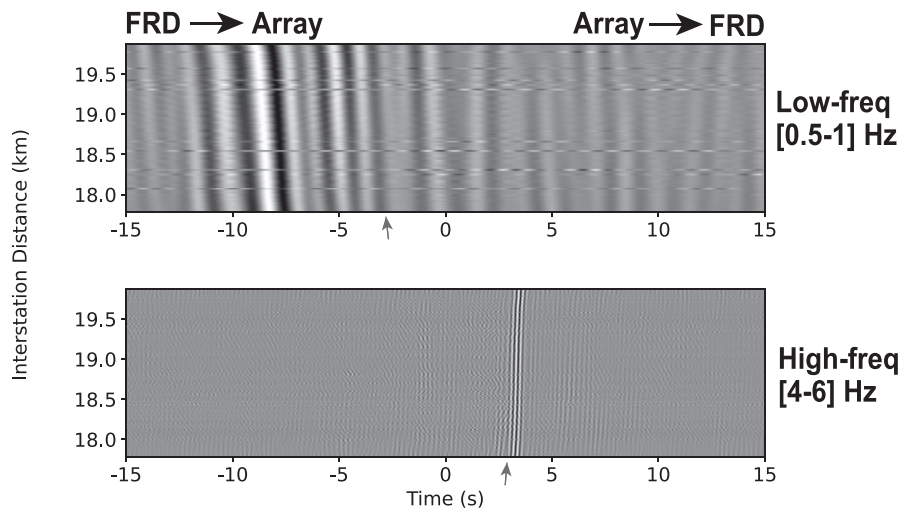


Figure 2. Causal and anticausal stacked vertical-to-vertical noise correlations between the array and station AZ.FRD using distance-averaged bins. Each bin is 10 m wide. The small red arrow indicates the arrival of the main ballistic body wave.

at x_1 due to an impulse source in the n -direction. With the complex conjugation indicating a reversal in the time domain. The indices i and j denote the component directions of the wavefield at receivers located at positions x_1 and x_2 , respectively.

In our study, we specifically focus on body waves within correlation functions by employing a ‘ P -wave windowing’ approach that isolates the ballistic P -wave first arrivals in the generating wavefield. This is achieved by applying a time window centered on the theoretical P -wave arrival time, with a window length of 1.2 s. By focusing on this specific window, we effectively suppress surface waves and mitigate cross-term interactions between P , S and surface waves. Given the complexity of separating different wave types and their sources in natural settings, our approach deliberately simplifies the correlation functions compared to those derived

from actual data. This simplification is crucial because it ensures that each interference is linked to one single target, helping us better understand what we observe in the real data. By selecting only specific phases, we make our analysis clearer and more effective in identifying subtle velocity changes at a given depth of investigation. We concentrate exclusively on ZZ correlations and employ vertical forces as sources.

In the following, we attempt to reproduce the noise correlations shown in Fig. 2, where the primary source at high frequency (4–6 Hz) is generated by freight trains running in the Coachella Valley. The selection of the 4–6 Hz frequency band is based on the prior study of Sheng *et al.* (2022) where clear body-wave energy emitted from trains was observed in this range. We use the spectral-element solver (Salvus Afanasiev *et al.* 2019) for the wavefield modelling.

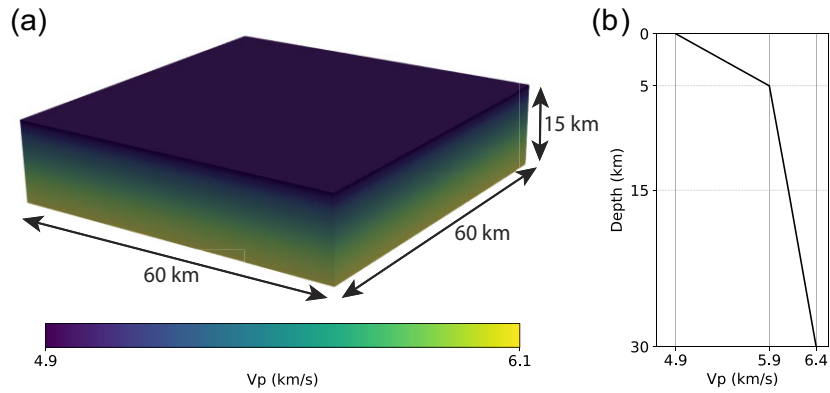


Figure 3. (a) 3-D representation of the V_p (P -wave velocity) mesh configuration. Dimensions of the mesh are detailed with the corresponding measurements. (b) Velocity profile from the model used in both this study and Sager *et al.* (2021).

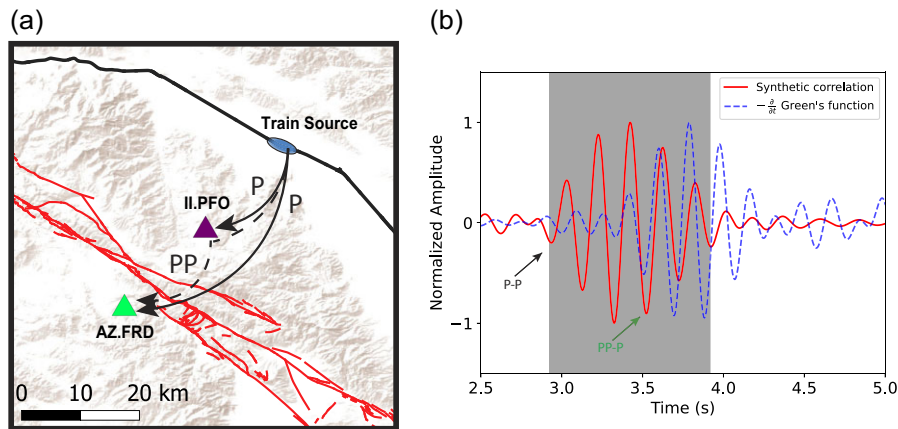


Figure 4. (a) Map view of the locations of the virtual source (II.PFO) in purple and the receiver (AZ.FRD) in green. The Gaussian patch represents the elongated noise source of the train along the railway. The black thick line represents the railway crossing the Coachella valley. The red line indicates the SJFZ. The solid black arrow depicts the direct P -wave path between the source and both stations, while the dashed black arrow represents the PP phase. (b) Comparison between synthetic correlation (red) and the negative time derivative of the Green's function (blue dashed line). The solid black arrow indicates the arrival of the early P - P phase, while the solid green arrow denotes the PP - P phase. The grey window delimits the time window used to compute the data-independent structure sensitivity kernels.

Table 1. Modelling parameters.

Parameter	Value
f_{\max} (Hz)	6
Elements per wavelength	1.5
Tensor order	1
Absorbing boundaries size (number of wavelengths)	1

To mimic train signals, we use localized noise sources with spatial extent and amplitude following a Gaussian-shaped distribution, with a standard deviation of 2.5 km along the railway direction and 0.5 km perpendicular to it. The maximum of this distribution is aligned with the two seismic stations between which we simulate the correlation. The source time function for the train tremor is a Ricker wavelet with a maximum frequency of 6 Hz and a central frequency of 5 Hz. The receivers, II.PFO and AZ.FRD situated on the surface, are separated by 19 km. In this scenario, we assume that the long-lasting train tremors are uncorrelated in time, which is a limiting approximation due to computational constraints, as a single forward run on 64 cores already requires 6 hr to complete. More work should be done considering a more realistic train source (Lavoué *et al.* 2020; Pinzon-Rincon *et al.* 2021). For the velocity

model and mesh characteristics, we use the parameters detailed in Fig. 3. Virtual station II.PFO is located 30 km away from the train source (Fig. 4b). The specific parameters employed in the modelling are summarized in Table 1.

Fig. 4(b) shows the synthetic correlation compared to the synthetic Green's function between the two stations. In the case of train noise correlations, we observe an early arrival at around 3 s, preceding the direct P wave of the Green's function. We interpret this early arrival (black arrow Fig. 4b) as the result of the correlation of the two direct P waves propagating between the source and the two stations. We call this correlation phase a P - P (P 'minus' P) phase. The latter part of the pulse at around 3.8 s represents a PP - P interaction resulting from the reconstructed correlation between a P phase at station II.PFO and a PP phase at station AZ.FRD reflecting below station II.PFO. This phase better aligns with the Green's function between the two stations, evidenced by a similar arrival time. The observed discrepancies in earlier phases arise from the inclusion of both spatial and temporal distribution of the noise sources, deviating from the traditional Green's function retrieval.

In traditional seismic interferometry, correlations between two stations should ideally reflect the Green's function, which represents

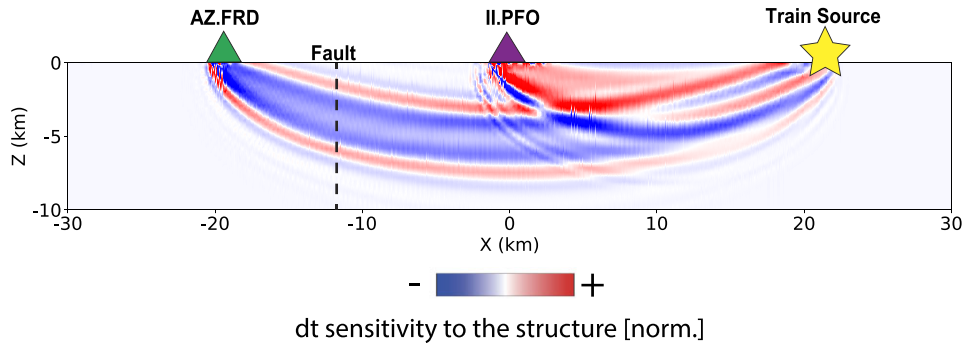


Figure 5. 2-D sensitivity kernel for traveltime shifts for two station pairs: II.PFO (purple triangle)-AZ.FRD (green triangle) for the train source (4–6 Hz) case (yellow star). The SJFZ is represented using a black dash line.

the ground response between the two points, assuming a homogeneous distribution of noise sources. However, as depicted in Fig. 4, our method employs temporally and spatially localized sources, which modify the behaviour of the correlation functions by capturing more than just the simple impulse response of the ground between stations. As a result, the correlation data often exhibit unique seismic arrivals not present in the corresponding Green’s function. These unique arrivals are not artefacts, but rather real seismic phases arising from cross-terms between different wave types, including P and S waves. While the ballistic arrivals are well understood, the cross-terms remain less clear but contain meaningful information about the subsurface (Mikesell *et al.* 2008).

To investigate the first body-wave arrival interactions in terms of change in the structure, we follow the same procedure as Sager *et al.* (2021). We create an adjoint wavefield that derives from a new simulation, where the adjoint source time function is made from the arrival of interest (grey window in Fig. 4). Then, by calculating sensitivity kernels corresponding to the interferences (correlation) between the first forward and the adjoint wavefield, we visualize areas of the model that influence the waveform. Fig. 5 shows a 2-D sensitivity kernel for traveltime shifts using stations II.PFO and AZ.FRD. We observe a single strong P wave emanating from the train source and reaching both stations. The sensitivity kernel highlights a broad distribution along the P -wave path between the source and both stations. As the sensitivity crosses the fault zone, it alternates between negative (blue) and positive (red) values, with the strongest lobe showing negative sensitivity. The colours in the kernel indicate how modifications in P -wave velocity impact the seismogram. In areas of positive sensitivity, an increase in velocity would reduce traveltimes, whereas in regions of negative sensitivity, a decrease in velocity would lead to increased traveltimes. This alternating sensitivity pattern along the fault suggests a complex interaction with the structure. However, the predominant influence within the fault region is negative sensitivity, implying that velocity decrease in the fault vicinity would mainly result in increased traveltimes along the P -wave path.

4 COMPARING SYNTHETIC AND OBSERVED CORRELATIONS

This section compares the synthetic correlations with the observed correlograms between the dense array of 300 stations at Piñon Flat Observatory and the permanent station AZ.FRD (Fig. 1). For this analysis, we initially evaluated the synthetic correlation functions

along a specific line of receivers because the alignment of these receivers simplifies the interpretation and visualization of the sensitivity kernels. We chose a line of receivers aligned with stations II.PFO and AZ.FRD, maintaining the same interstation distance as the actual positions of the nodes in the array with respect to AZ.FRD. We notice a match between the observed and synthetic first arrival P wave, with a moveout of approximately 6 km s^{-1} (Fig. 6a). This correct moveout of the train source is an important result, as it demonstrates the accuracy of our modelling approach in capturing the key characteristics of the body-wave propagation generated by the train source. We can also observe weak secondary arrivals in the synthetic correlations that likely correspond to S waves, surface waves, or cross-talks between P waves and unmuted later arrivals (Fig. 6b). The missing energy in the observed correlograms could be attributed to several factors, such as attenuation, scattering and the complex topography between the Coachella Valley and the SJFZ, which are not fully accounted for in our simulations. Additionally, subsurface heterogeneities and differences in the source distribution in the real medium can lead to energy loss in the observed data. These factors, combined with the imperfect retrieval of the Green’s function, contribute to the discrepancies between the observed and for later arrivals.

5 SOURCE VARIABILITY: EFFECTS ON THE TRAVELTIME OF BODY WAVES

Monitoring with noise correlations necessitates sources that, when averaged over specific time intervals, consistently emit seismic waves from stable positions (Hadziioannou *et al.* 2009). In practice, the spatial variability of noise sources affects the traveltimes of body waves and can mask traveltime perturbations induced by genuine velocity changes in the medium. We model a moving train by placing different elongated Gaussian sources along the actual trajectory of the Coachella Valley railway. The train catalogue used to compute the observed correlations was created using a permanent station close to the railway (CI.IDO, Fig. 1) to detect train activity (Sheng *et al.* 2022). Each train detection has an uncertainty in the train position, inducing an apparent time-shift between correlation functions from different trains. We do not consider in our modelling the variations in speed or weight of the trains, which could further vary the radiated seismic energy and potentially influence the correlation P -waves traveltimes. Here, we only quantify how the uncertainty in train position maps to the uncertainty in relative traveltime of the P -body wave phase in the correlation function between II.PFO and AZ.FRD.

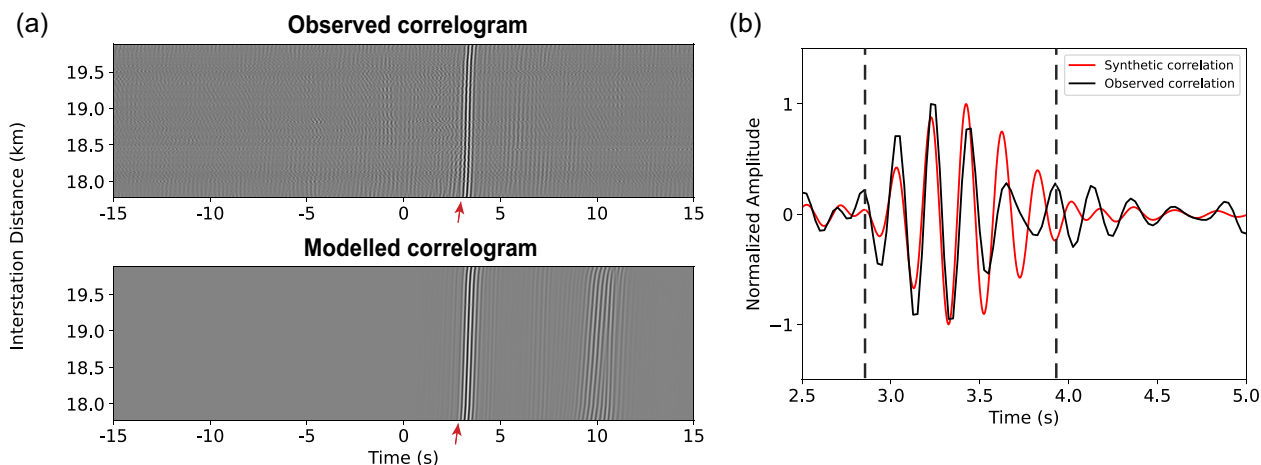


Figure 6. (a) A comparative display of observed and synthetic correlograms in the frequency range (4–6 Hz), with the top section displaying the observed correlogram and the bottom section showing the modelled one. (b) Comparison with synthetic and observed correlation for station pair II.PFO-AZ.FRD in the lower frequency range (4–6 Hz). The black dashed line delimits the window used to compute traveltime measurements.

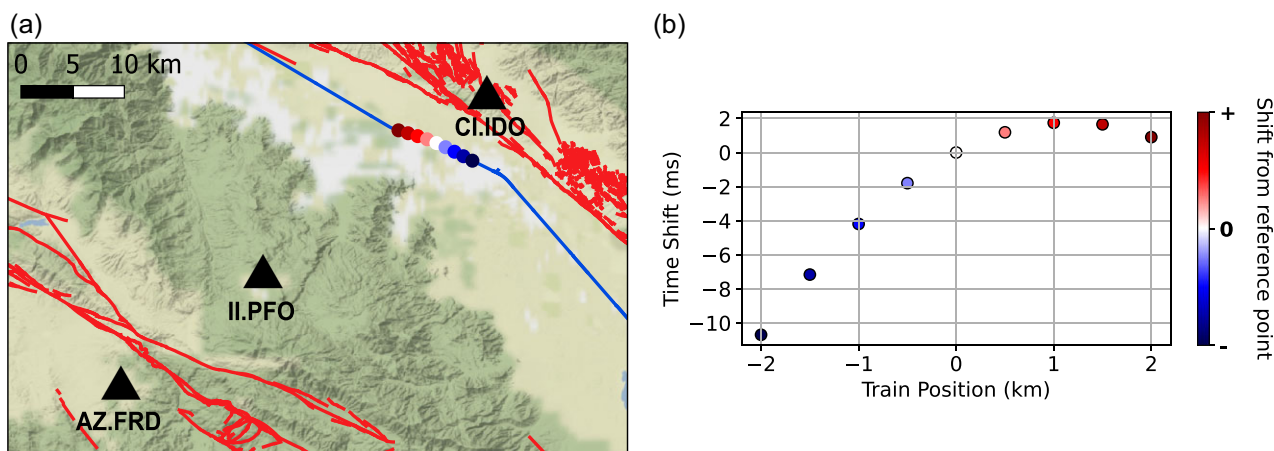


Figure 7. (a) A map illustrating the two stations used for cross-correlation, AZ.FRD and II.PFO. Red lines depict the fault in the area, while the coloured points indicate the train's position along the railway. The railway corresponds to the solid black line. (b) Time-shift measurements on a window including P - P and PP - P wave interaction for different train positions along the railway.

We use a similar model as before, where the central part of the Gaussian distribution, measuring 2.5 km along the railway direction and 0.5 km perpendicular to it, is shifted from -2 to 2 km relative to the alignment of stations II.PFO, AZ.FRD and CI.IDO. To measure time delays (δt), we determine the slope of the cross-spectrum between the reference cross-correlation and the shifted cross-correlation as a function of frequency (Poupinet *et al.* 1984) on a window duration delimited by the vertical black dashed lines, that include both P - P and PP - P wave interaction in Fig. 6(b). As displayed in Fig. 7(b), an uncertainty of 1 km in the source position causes a time-shift variation as large as 4 ms on the first body-wave arrival. Sheng *et al.* (2022) measure a time-shift of 6 ms occurring in 2014 February that they interpret as being associated with a slow slip along the San Jacinto Fault at depth. They use δt measurements averaged over 30 trains, which reduces δt variability to just a few tenths of milliseconds. This observed variability is consistent with the time-shift variations modelled here. These results highlight the need to locate trains as accurately as possible along the railway (less than 1 km accuracy) to reduce the scattering of δt measurements.

However, the step change in δt observed by Sheng *et al.* (2022) implies a systematic source location discrepancy of more than 1 km before and after 2014. Such a substantial location shift seems improbable, suggesting that the observed step change may not be solely attributed to source position variability.

6 SENSITIVITY OF TRAVELTIME SHIFT TO VELOCITY CHANGES IN THE SHALLOW SUBSURFACE AND THE FAULT ZONE

In this section, we study how the direct body waves retrieved from train noise correlations across the SJFZ are sensitive to environmental changes in the near-surface or localized velocity changes along the SJFZ, with the goal of detecting fault-related activity using seismic interferometry. Our study goes deeper into the analysis of the bias in traveltime perturbation measurements induced by moving sources, such as trains, providing a general framework to assess

Table 2. Comparison of time-shift measurements between near-surface and fault zone perturbations, using ZZ component of synthetic correlation function between station pair II.PFO and AZ.FRD filtered between 4 and 6 Hz.

	Near surface	Fault zone
Velocity decrease	−1 per cent (0–100 m)	−1 per cent (0–1 km)
δt	3.09 ± 1.40 ms	1.68 ± 0.90 ms
$\delta t/t$	0.1 per cent	0.06 per cent

uncertainties when measuring traveltime perturbations on ballistic waves retrieved from specific noise correlations.

To probe the sensitivity of body waves to modifications in the near-surface structure, we implemented a unit test in our model, applying a uniform velocity decrease of 1 per cent throughout the first 100 m of the near-surface. This test model is designed to evaluate how the sensitivity scales with near-surface perturbations rather than to mimic realistic conditions directly. Since we cannot distinguish between *PP-P* and *P-P* interactions in the observed correlation (Fig. 6b), we measure the time-shift in a window (grey window in Fig. 6b) that includes both phases in the synthetic correlations. We measure a traveltime perturbation of 3.09 ± 1.40 ms (Table 2). Assuming the *P*-wave traveltime is approximately 3 s, we find that a 100 m thick velocity decrease of 1 per cent results in a perturbation in $d\tau/t$ of about 0.1 per cent, which is 10 times smaller than the actual velocity perturbation at the surface. This result demonstrates that *P* waves (direct and cross-terms) from train noise correlations are weakly sensitive to near-surface perturbations. This behaviour is opposite to the sensitivity of correlation surface waves, both ballistic and within coda waves, the main phases used in interferometry-based monitoring, which is amplified in the near-surface. This provides a better understanding of why seasonal velocity changes, typically observed in many noise-based monitoring studies, were not detected by Sheng *et al.* (2022).

We also investigate how a velocity change along the SJFZ between stations II.PFO and AZ.FRD affects the reconstructed high-frequency *P*-wave traveltime. In this case, we introduce a 1 per cent velocity decrease in a 0.5–1 km wide, 1 km thick and 1 km deep rectangular parallelepiped centred on the San Jacinto Fault. This scenario is designed to assess the sensitivity of our measurements to localized velocity changes in the fault zone, with the depth extent of the simulated velocity change limited to the top 1 km. At greater depths, seismic velocities are less sensitive to stress perturbations due to the high lithostatic pressure closing cracks (Yamamura *et al.* 2003). The analysis aims to evaluate how sensitive the traveltime perturbations are in relation to the 6 ms traveltime shift observed by Sheng *et al.* (2022) in 2014, which was attributed to fault zone processes.

We find a time-shift of 1.68 ± 0.90 ms (Table 2). Using a *P*-wave traveltime of 3 s, the analysis shows that a very localized velocity decrease in a fault zone induces a perturbation in $\delta t/t$ of approximately 0.06 per cent. This low value of traveltime perturbation indicates that, while the correlation *P* waves can be sensitive to localized changes, it is unlikely that such perturbation could be detected with the $d\tau$ measurement from a single train; especially given the uncertainties introduced by near-surface changes (see previous section). To achieve a $d\tau$ measurement accuracy of 0.06 per cent induced by a real local change it would be necessary to average the measurements over many trains, which will also result in reducing the temporal resolution of the measurements. Our results tend to corroborate that the $d\tau$ anomaly detected by Sheng *et al.* (2022) is induced by a large-scale volumic change rather than a localized one.

7 DISCUSSION AND CONCLUSION

This study compares synthetic and observed correlations from localized noise sources, specifically focusing on their *P*-wave phases. By building on the framework developed by Fichtner (2014), Sager *et al.* (2021) and Sheng *et al.* (2022), we were able to reproduce ballistic *P* waves emerging from high-frequency train noise correlations with good fidelity. The good agreement between the synthetic and observed first-arrival *P* waves demonstrates that our synthetic models are approaching a realistic representation of train noise correlations, at least within the studied frequency range of 4–6 Hz.

Initially, we examined the sensitivity of traveltime shifts to train position variations along the railway. This analysis showed that while shifts in train location could impact $d\tau$ measurements, the step change observed by Sheng *et al.* (2022) equivalent to a 6 ms time step would imply a systematic source position discrepancy of more than 1 km. Such a large location shift seems unlikely, suggesting that the observed step change is not solely due to source position variability.

We then examined how $d\tau$ measurements respond to velocity perturbations in the fault zone and the near-surface environment. Although these perturbations are detectable, they are small: a 1 per cent localized velocity decrease in the fault zone resulted in a 0.06 per cent change in traveltime, while a 1 per cent decrease in the top 100 m of the near-surface caused a 0.1 per cent shift. While these values could theoretically match the magnitude of the changes reported by Sheng *et al.* (2022), the actual velocity reductions required to produce such shifts would be unusually large. This suggests that other scenarios or a combination of factors, such as simultaneous variations in both the fault zone and near-surface layers, may have contributed to the observed time-shift.

Despite these promising results, limitations remain in our modelling approach. We utilized a 1-D velocity model, which fails to account for the strong lateral and depth heterogeneities present in the region (Jiang *et al.* 2021). In fault zones like the SJFZ, where significant heterogeneity and reduced velocities can alter the sensitivity of noise correlations, it is essential to consider the effects of localized changes within the fault zone itself. For instance, if a velocity change occurs specifically within the fault zone, it could impact the noise correlation sensitivity as suggested by van Dinther *et al.* (2021). The velocity changes obtained by Sheng *et al.* (2022) represent a volumetric alteration, which likely contributed to its detectability by traveltime perturbation measurements. Our synthetic tests indicate that such changes cannot arise solely from a small volume around the fault, underscoring the need for models that can account for these broader variations. To fully capture the behaviour of noise correlations in fault zones, more sophisticated models incorporating 3-D heterogeneities and anisotropy will be necessary.

In an ideal case, the correlation should reflect the Green's function and be sensitive only to the medium between the two stations. By 'ideal case', we refer to a scenario where the noise sources are perfectly random and uncorrelated in both space and time, allowing the noise correlation to converge exactly to the Green's function. However, this is rarely achieved in practice, especially when spatially localized sources, such as trains, are involved. Those sources are isolated in space and time, which causes the sensitivity of the correlation function to be influenced by both the source and the source–receiver path, preventing the correlation from converging to the Green's function. From a monitoring perspective, we are

interested in identifying transient or permanent velocity changes. In our case study, we cannot directly infer perturbations only between the two stations, because the sensitivity of the correlation P waves is spread between the source and the two receivers in a large volume and in a complex manner. To better localize the source of the velocity perturbation, it is therefore necessary to combine measurements from several receiver pairs in order to resolve the spatial sensitivity ambiguity. By adding more station pairs, we introduce additional sensitivity kernels that provide more paths across the study area. These multiple kernels can be summed constructively, enhancing the spatial coverage and sensitivity to velocity changes. The ability to integrate information from kernels across multiple station pairs enhances our capability to pinpoint perturbations, suggesting that a network of stations, rather than isolated pairs, could significantly improve our understanding and monitoring of velocity changes.

While our study focused on well-identified high-frequency body-wave sources, it would be tempting to extrapolate it to low-frequency natural sources of body waves such as the one identified by Roux *et al.* (2005). This leap is not as straightforward as it seems. Modelling microseismic sources, such as those generated by ocean swell and wind–swell interactions, is challenging due to the complex physics involved, uncertainties in the seafloor structures, and the spatial and temporal variabilities of the source processes. This results in a limited understanding of how these body waves appear and how they influence the resulting seismic correlation wavefield. Unlike anthropogenic sources like trains, microseismic noise sources are less stable over time, adding further uncertainty to traveltime perturbation measurements and limiting the replicability of these findings in different regions.

Our work provides a general framework for assessing uncertainties in traveltime perturbation measurements, especially when using moving noise sources like trains. This framework offers important insights into the bias induced by such sources, while also highlighting the potential for generalization to other stationary noise sources, with broader applications in areas such as near-surface monitoring, CO₂ sequestration and nuclear waste storage.

The specific site conditions at the SJFZ play a significant role in our study's results. The proximity of the railway and the dense seismic array provided a unique opportunity to explore these noise sources, but replicating this setup at other fault zones may prove difficult. While we believe the methodology developed here has broader applicability, especially using other stationary sources like wind turbines or industrial noise, replicating this precise setup in different fault zones may present challenges. Similar arrays have already been deployed at strategic locations, such as the SAFOD project, where data were collected from a dense borehole array (Zoback *et al.* 2011). Future work will be necessary to explore how these methodologies might be applied across different geological and seismic settings.

In conclusion, this study offers valuable insights into the use of body-wave noise correlations for seismic velocity monitoring, particularly in fault zones. Continued advancements in modelling techniques and the deployment of high-density seismic networks will be crucial to fully harnessing the potential of this method for subsurface monitoring and understanding fault-related processes.

ACKNOWLEDGMENTS

This study was funded by the European Research Council (ERC) under grant agreement no. 817803 (FAULTSCAN).

DATA AVAILABILITY

The seismic waveform data used in this study are available via IRIS. For the Anza Seismic Network, the data are archived as described by Frank Vernon (1982). Data for the Global Seismograph Network are detailed by the Scripps Institution of Oceanography (1986). Data from the San Jacinto FaultScan Project are forthcoming and will be available through Brenguier *et al.* (2025). For downloading continuous seismic data and performing seismic interferometry, we employed the pycorr software package by Boué & Stehly (2022).

REFERENCES

- Afanasyev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D.A., Knepley, M.G. & Fichtner, A., 2019. Modular and flexible spectral-element waveform modelling in two and three dimensions, *Geophys. J. Int.*, **216**(3), 1675–1692.
- Boué, P. & Stehly, L., 2022. *Pycorr (software)*, <https://doi.org/0.5281/ZENODO.6793401> (Accessed November 2023).
- Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N.M., Nadeau, R.M. & Larose, E., 2008. Postseismic relaxation along the san andreas fault at parkfield from continuous seismological observations, *Science*, **321**(5895), 1478–1481.
- Brenguier, F., Campillo, M., Takeda, T., Aoki, Y., Shapiro, N.M., Briand, X., Emoto, K. & Miyake, H., 2014. Mapping pressurized volcanic fluids from induced crustal seismic velocity drops, *Science*, **345**(6192), 80–82.
- Brenguier, F. *et al.*, 2019. Train traffic as a powerful noise source for monitoring active faults with seismic interferometry, *Geophys. Res. Lett.*, **46**(16), 9529–9536.
- Brenguier, F. *et al.*, 2020. Noise-based ballistic wave passive seismic monitoring. Part 1: body waves, *Geophys. J. Int.*, **221**(1), 683–691.
- Brenguier, F., Vernon, F. & Ben-Zion, Y., 2018. 2018 FaultScan San Jacinto Fault dense array experiment, https://doi.org/10.7914/SN/9K_2018.
- Brenguier, F., Vernon, F. & Ben-Zion, Y., 2025. San Jacinto Fault-Scan Project, <https://doi.org/10.7914/WQVG-2272> (Accessed November 2023).
- Fichtner, A., 2014. Source and processing effects on noise correlations, *Geophys. J. Int.*, **197**(3), 1527–1531.
- Hadziioannou, C., Larose, E., Coutant, O., Roux, P. & Campillo, M., 2009. Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: Laboratory experiments, *J. acoust. Soc. Am.*, **125**(6), 3688–3695.
- Hanasoge, S.M., 2013. The influence of noise sources on cross-correlation amplitudes, *Geophys. J. Int.*, **192**(1), 295–309.
- 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, *Bull. seism. Soc. Am.*, **108**(6), 3511–3527.
- Jiang, E., Liu, K.H., Gao, Y., Fu, X. & Gao, S.S., 2021. Spatial variations of upper crustal anisotropy along the san jacinto fault zone in southern california: constraints from shear wave splitting analysis, *J. geophys. Res.: Solid Earth*, **126**(4), e2020JB020876, doi:10.1029/2020JB020876.
- Kanu, C. & Snieder, R., 2015. Time-lapse imaging of a localized weak change with multiply scattered waves using numerical-based sensitivity kernel, *J. geophys. Res.: Solid Earth*, **120**(8), 5595–5605.
- Larose, E. *et al.*, 2015. Environmental seismology: what can we learn on earth surface processes with ambient noise?, *J. appl. Geophys.*, **116**, 62–74.
- Lavoué, F. *et al.*, 2020. Understanding seismic waves generated by train traffic via modeling: implications for seismic imaging and monitoring, *Seismol. Res. Lett.*, **92**(1), 287–300.
- Mao, S., Campillo, M., van der Hilst, R.D., Brenguier, F., Stehly, L. & Hillers, G., 2019. High temporal resolution monitoring of small variations in crustal strain by dense seismic arrays, *Geophys. Res. Lett.*, **46**(1), 128–137.

- Mao, S., Lecointre, A., van der Hilst, R.D. & Campillo, M., 2022. Space-time monitoring of groundwater fluctuations with passive seismic interferometry, *Nat. Commun.*, **13**(1), 4643.
- Margerin, L., Planès, T., Mayor, J. & Calvet, M., 2016. Sensitivity kernels for coda-wave interferometry and scattering tomography: theory and numerical evaluation in two-dimensional anisotropically scattering media, *Geophys. J. Int.*, **204**(1), 650–666.
- Mikesell, T., Wijk, K., Calvert, A. & Haney, M., 2008. The virtual refraction: useful spurious energy in seismic interferometry, *Kasper van Wijk*, **74**, doi:10.1190/1.3095659.
- Müller, J., van Laaten, M., Eulenfeld, T. & Wegler, U., 2023. Coseismic drop of seismic velocity caused by the 2023 Turkey–Syria earthquakes, *Geophys. J. Int.*, **234**(3), 2465–2472.
- Nakata, N., Chang, J.P., Lawrence, J.F. & Boué, P., 2015. Body wave extraction and tomography at long beach, *J. geophys. Res.: Solid Earth*, **120**(2), 1159–1173.
- Obermann, A., Planès, T., Larose, E., Sens-Schönfelder, C. & Campillo, M., 2013. Depth sensitivity of seismic coda waves to velocity perturbations in an elastic heterogeneous medium, *Geophys. J. Int.*, **194**(1), 372–382.
- Obermann, A., Kraft, T., Larose, E. & Wiemer, S., 2015. Potential of ambient seismic noise techniques to monitor the St. Gallen geothermal site (Switzerland), *J. geophys. Res.: Solid Earth*, **120**(6), 4301–4316.
- Okubo, K., Delbridge, B.G. & Denolle, M.A., 2024. Monitoring velocity change over 20 years at parkfield, *J. geophys. Res.: Solid Earth*, **129**(4), e2023JB028084, doi:10.1029/2023JB028084.
- Pinzon-Rincon, L. et al., 2021. Humming trains in seismology: an opportune source for probing the shallow crust, *Seismol. Res. Lett.*, **92**(2A), 623–635.
- Poupinet, G., Ellsworth, W.L. & Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: an application to the calaveras fault, *J. geophys. Res.: Solid Earth*, **89**(B7), 5719–5731.
- Rivet, D., Brenguier, F., Clarke, D., Shapiro, N.M. & Peltier, A., 2014. Long-term dynamics of Piton de la Fournaise volcano from 13 years of seismic velocity change measurements and GPS observations, *J. geophys. Res.: Solid Earth*, **119**(10), 7654–7666.
- Ross, Z.E., Hauksson, E. & Ben-Zion, Y., 2017. Abundant off-fault seismicity and orthogonal structures in the San Jacinto fault zone, *Sci. Adv.*, **3**(3), e1601946, doi:10.1126/sciadv.1601946.
- Roux, P., Sabra, K.G., Gerstoft, P., Kuperman, W.A. & Fehler, M.C., 2005. P-waves from cross-correlation of seismic noise, *Geophys. Res. Lett.*, **32**(19), doi:10.1029/2005GL023803.
- Sager, K., Tsai, V.C., Sheng, Y., Brenguier, F., Boué, P., Mordret, A. & Igel, H., 2021. Modelling P waves in seismic noise correlations: advancing fault monitoring using train traffic sources, *Geophys. J. Int.*, **228**, 1556–1567.
- Scripps Institution of Oceanography, 1986. Global Seismograph Network - IRIS/IDA, <https://doi.org/10.7914/SN/II> (Accessed November 2023).
- Sens-Schönfelder, C. & Wegler, U., 2008. Passive image interferometry and seasonal variations of seismic velocities at merapi volcano, in *Seismic Interferometry: History and Present Status*, Vol. **26**, Society of Exploration Geophysicists, doi:10.1190/1.9781560801924.
- Sens-Schönfelder, C. & Wegler, U., 2011. Passive image interferometry for monitoring crustal changes with ambient seismic noise, *Comp. R. Geosci.*, **343**(8), 639–651.
- Sens-Schönfelder, C., Snieder, R. & Li, X., 2019. A model for nonlinear elasticity in rocks based on friction of internal interfaces and contact aging, *Geophys. J. Int.*, **216**(1), 319–331.
- Sheng, Y. et al., 2022. Monitoring seismic velocity changes across the san jacinto fault using train-generated seismic tremors, *Geophys. Res. Lett.*, **49**(19), e2022GL098509, doi:10.1029/2022GL098509.
- Sheng, Y. et al., 2023. Seeking repeating anthropogenic seismic sources: implications for seismic velocity monitoring at fault zones, *J. geophys. Res.: Solid Earth*, **128**(1), e2022JB024725, doi:10.1029/2022JB024725.
- Sheng, Y. et al., 2024. In situ velocity–strain sensitivity near the san jacinto fault zone analyzed through train tremors, *Geophys. Res. Lett.*, **51**(15), e2024GL110113, doi:10.1029/2024GL110113.
- Sánchez-Pastor, P. et al., 2023. Steam caps in geothermal reservoirs can be monitored using seismic noise interferometry, *Commun. Earth Environ.*, **4**(1), 1–9.
- Taira, T., Bürgmann, R., Nadeau, R.M. & Dreger, D.S., 2014. Variability of fault slip behavior along the San Andreas Fault in the San Juan Bautista Region, *J. geophys. Res.: Solid Earth*, **119**(12), 8827–8844.
- Taira, T., Brenguier, F. & Kong, Q., 2015. Ambient noise-based monitoring of seismic velocity changes associated with the 2014 Mw 6.0 South Napa earthquake, *Geophys. Res. Lett.*, **42**(17), 6997–7004.
- Tromp, J., Luo, Y., Hanasoge, S. & Peter, D., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.*, **183**(2), 791–819.
- van Dinther, C., Margerin, L. & Campillo, M., 2021. Implications of laterally varying scattering properties for subsurface monitoring with coda wave sensitivity kernels: application to volcanic and fault zone setting, *J. geophys. Res.: Solid Earth*, **126**(12), e2021JB022554, doi:10.1029/2021JB022554.
- van Manen, D.-J., Curtis, A. & Robertsson, J.A.O., 2006. Interferometric modeling of wave propagation in inhomogeneous elastic media using time reversal and reciprocity, *Geophysics*, **71**(4), SI47–SI60.
- Vernon, F., 1982. ANZA Regional Network, <https://doi.org/10.7914/WQVG-2272> (Accessed November 2023).
- Wegler, U. & Sens-Schönfelder, C., 2007. Fault zone monitoring with passive image interferometry, *Geophys. J. Int.*, **168**(3), 1029–1033.
- Yamamura, K., Sano, O., Utada, H., Takei, Y., Nakao, S. & Fukao, Y., 2003. Long-term observation of in situ seismic velocity and attenuation, *J. geophys. Res.: Solid Earth*, **108**(B6), 2002JB002005, doi:10.1029/2002JB002005.
- Zhang, J., Gerstoft, P. & Shearer, P.M., 2009. High-frequency P-wave seismic noise driven by ocean winds, *Geophys. Res. Lett.*, **36**(9), doi:10.1029/2009GL037761.
- Zoback, M., Hickman, S. & Ellsworth, W. & the SAFOD Science Team, 2011. Scientific drilling into the san andreas fault zone - an overview of safod's first five years, *Sci. Drill.*, **11**, 14–28.