



HAL
open science

Towards structural imaging using seismic ambient field correlation artefacts

Lise Retailleau, Gregory C. Beroza

► **To cite this version:**

Lise Retailleau, Gregory C. Beroza. Towards structural imaging using seismic ambient field correlation artefacts. *Geophysical Journal International*, 2021, 225, pp.1453-1465. 10.1093/gji/ggab038 . insu-03590054

HAL Id: insu-03590054

<https://insu.hal.science/insu-03590054>

Submitted on 26 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Towards structural imaging using seismic ambient field correlation artefacts

Lise Retailleau^{1,2,3} and Gregory C. Beroza¹

¹*Department of Geophysics, Stanford University, Stanford, CA, USA. E-mail: retailleau@igpp.fr*

²*Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France*

³*Observatoire Volcanologique du Piton de la Fournaise, Institut de physique du globe de Paris, F-97418 La Plaine des Cafres, France*

Accepted 2021 January 22. Received 2020 December 8; in original form 2020 January 3

SUMMARY

Correlations of the ambient seismic field recorded by seismic stations carry information about the wave propagation between the stations. They also contain information about the ambient field—both the source of the ambient field, and sources of scattering that contribute to it. The waves that comprise the ambient field are subject to scattering due to the heterogeneous Earth, which can generate supplementary arrivals on the correlation functions. We use these effects to locate sources of signals linked to scattering. For this analysis, we use correlation functions computed from continuous signals recorded between 2013 and 2015 by a line of seismic stations in Central California. We identify spurious arrivals on the Vertical to Vertical and Transverse to Transverse correlation functions and use array analysis to map the source of scattering, which is linked to strong structural variations in the Coast ranges and at the border of the Great Valley.

Key words: Seismic noise; Surface waves and free oscillations; Wave scattering and diffraction.

1 INTRODUCTION

The seismic ambient field and the correlation functions derived from it have been used widely and successfully in a variety of contexts: from imaging the structure of the Earth to monitoring its evolution with time. After appropriate processing, the correlation of ground motion recorded by two seismic stations approximates the Green's function, which describes the wave propagation between the two stations (Derode *et al.* 2003; Weaver & Lobkis 2004). These signals are particularly useful in areas where earthquake data is scarce. This capability has led to numerous new images of the structure of the Earth at all scales, including the local scale for the Long beach data set (Lin *et al.* 2013a), the regional scale for California (Shapiro *et al.* 2005; Lee *et al.* 2014), the continental scale for North America (Yang *et al.* 2007) and the global scale for the deep Earth (Boué *et al.* 2013; Lin *et al.* 2013b; Retailleau *et al.* 2020).

For the Green's function to be recovered correctly from the correlation process, the ambient seismic field must be equipartitioned (Weaver & Lobkis 2004; Sánchez-Sesma & Campillo 2006; Gouedard *et al.* 2008). This implies a random distribution of uncorrelated sources or strong and pervasive scattering of the waves due to heterogeneities to randomize the wavefield (Roux *et al.* 2005; Wapenaar & Fokkema 2006; Wapenaar *et al.* 2006). Sources of the ambient field at periods longer than a few seconds are mostly linked to the interaction between the ocean and the solid Earth. Although they are generated at parts of the Earth's surface that are covered

by oceans (Longuet-Higgins 1950; Hasselmann 1963; Webb 1998; Nishida *et al.* 2000; Stutzmann *et al.* 2000), they are observed globally and are referred to collectively as microseisms due to their small amplitude. Multiple studies have shown that the ambient field correlations depend on its sources (e.g. Tsai 2009; Harmon *et al.* 2010). Different pre-processing schemes have been developed to reduce the bias due to an uneven distribution of ambient field sources (Bensen *et al.* 2007; Ritzwoller & Feng 2018).

The dependence of the correlation results on the ambient field distribution means that they contain both information on the wave propagation between stations and on the source field. Some studies have limited the effects of heterogeneous source distribution by suppressing the effects of persistent non-diffuse sources to improve their correlation functions (Zheng *et al.* 2011; Liu *et al.* 2019). Other studies drop the assumption that the correlation function converges to the Green's function and instead use it directly to obtain the medium and the source distribution simultaneously (Yao & Van Der Hilst 2009; Tromp *et al.* 2010; Fichtner *et al.* 2016; Sager *et al.* 2017). Finally, correlations have been used to analyse the source distribution itself. The most obvious manifestation of the effects of noise source heterogeneity on the correlation functions is the asymmetry of the causal and anticausal parts of the correlation functions (Stehly *et al.* 2006; Yang & Ritzwoller 2008; Ermert *et al.* 2016). Another manifestation of non-homogeneous sources is the occurrence of spurious arrivals on the correlations (Snieder 2006). Several methods have used these arrivals to locate strong and

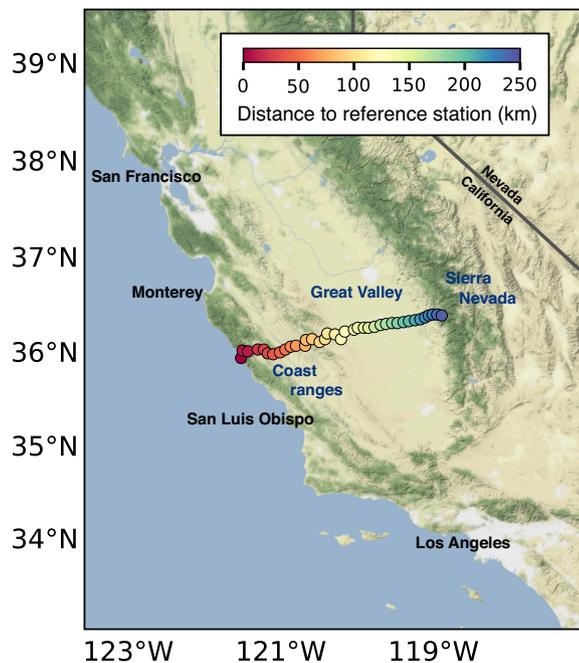


Figure 1. Map of California and the stations used in this study. The colour, from red to blue represents the increasing distance from the westernmost station.

isolated microseism sources generated by different mechanisms. Sources have been linked to ocean waves (Retailleau *et al.* 2017), long period volcanic tremor (Zeng & Ni 2010; Kawakatsu *et al.* 2011), scattering from a magma chamber (Ma *et al.* 2013) and other more enigmatic sources, like the 26 s microseism observed in the Guinea Gulf (Shapiro *et al.* 2006).

Ma *et al.* (2013) extracted a spurious arrival from correlations computed in Peru and associated it with a strong scatterer in the volcanic arc, potentially linked to a low velocity magma chamber and estimated the size of the magma chamber based on the arrival they observed. Any sharp heterogeneity has the potential to scatter incident waves originally generated by ocean sources, and in doing so to act as a secondary source from which the scattered wavefield will propagate. In this paper, we extract spurious arrivals and use them to locate strong scatterers linked to sharp heterogeneities. Scattering is usually analysed from the coda of earthquakes (Margerin *et al.* 2009). Scattering analysis has been used to extract properties of the Earth at different scales (Kaneshima 2016) and understand its behaviour (Planes *et al.* 2014; Obermann *et al.* 2016; Sato 2019). The complicated characteristic of the coda, and possible complications from multiple scattering lead to the use of stochastic models of Earth structure to explain the coda. In our case, spurious arrivals are linked to single scattering dissociated from the direct arrivals which means they can be attributed deterministically to a specific location.

We first discuss the effects of the ambient seismic source distribution on the correlations and the emergence of spurious arrivals in an homogeneous medium with synthetic signals. We then discuss our data set and its Central California setting. We use data recorded by a line of 38 sensors crossing the Great Valley [fig. 1, CCSE (2015)]. The stations recorded signals continuously for two years between 2013 and 2015. This line of sensors is an ideal setup to track direct velocity variations due to propagation across Central California by correlating the signals of the most western station to

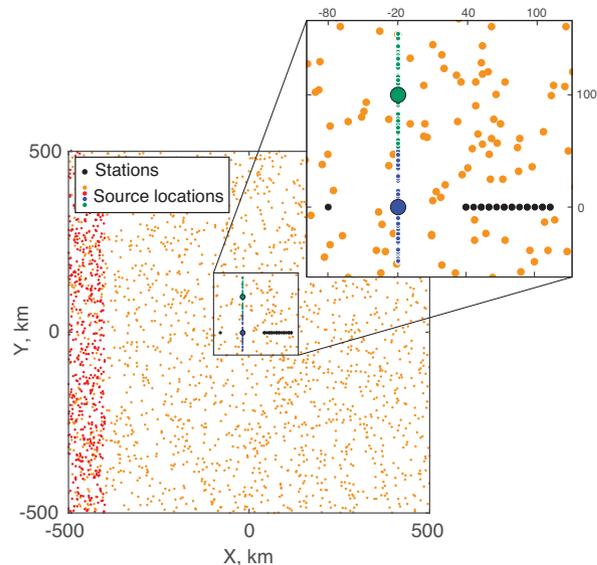


Figure 2. The black dots represent the stations. The coloured dots represent sources in different configurations. Orange represents the homogeneous distribution, where sources are distributed randomly all over the area. Red sources are distributed randomly on one side of the area, simulating an asymmetry of the source distribution. Blue and green represent two localized sources, either at an isolated location (large dot) or along a line (small dots).

the signals of the other stations. We describe the methodology we use to extract spurious arrivals and locate their sources on synthetic data and apply it on the Central California data. We finally use the process to image strong scatterers, and discuss their origins.

2 CORRELATION RECONSTRUCTION AND AMBIENT NOISE SOURCE DISTRIBUTION

At different locations, ambient field sources have a different influence on the recovered correlations (Campillo & Roux 2015; Retailleau *et al.* 2017). The Green's function emerges in the correlations when the ambient field sources have propagation paths that go through both receivers, that is the sources are located in the end-fire lobes. In that case, the signals recorded by the two receivers are similar, and the time delay reflects the propagation between them. In the end-fire lobes, the arrival time of different sources on the correlations do not vary much, so the different arrivals contribute coherently to the correlation. The sources that are not located in the end-fire lobes, and thus don't reach the two receivers along one path, tend to interfere destructively in the correlation function because their arrival times vary rapidly.

We illustrate these ideas with a synthetic test in a homogeneous medium. We generate simple wavelets of 5 s period and 3 km s^{-1} propagation velocity at source locations defined by different geographic settings, and recorded by a group of stations (Fig. 2). The station configuration consists of a reference station (isolated black dot at $x = -80, y = 0$ in Fig. 2) aligned with a set of 20 stations (at $y = 0$ and $x \in [40, 120]$ in Fig. 2), represented in black in Fig. 2. We correlate the signal at the reference station with the signals at the other stations for different source configurations (Figs 3 and 4). The resulting correlations reflect the effects of the source distribution.

The reference station is crucial as its data is correlated to all the stations. Its location to the closest station should be large enough

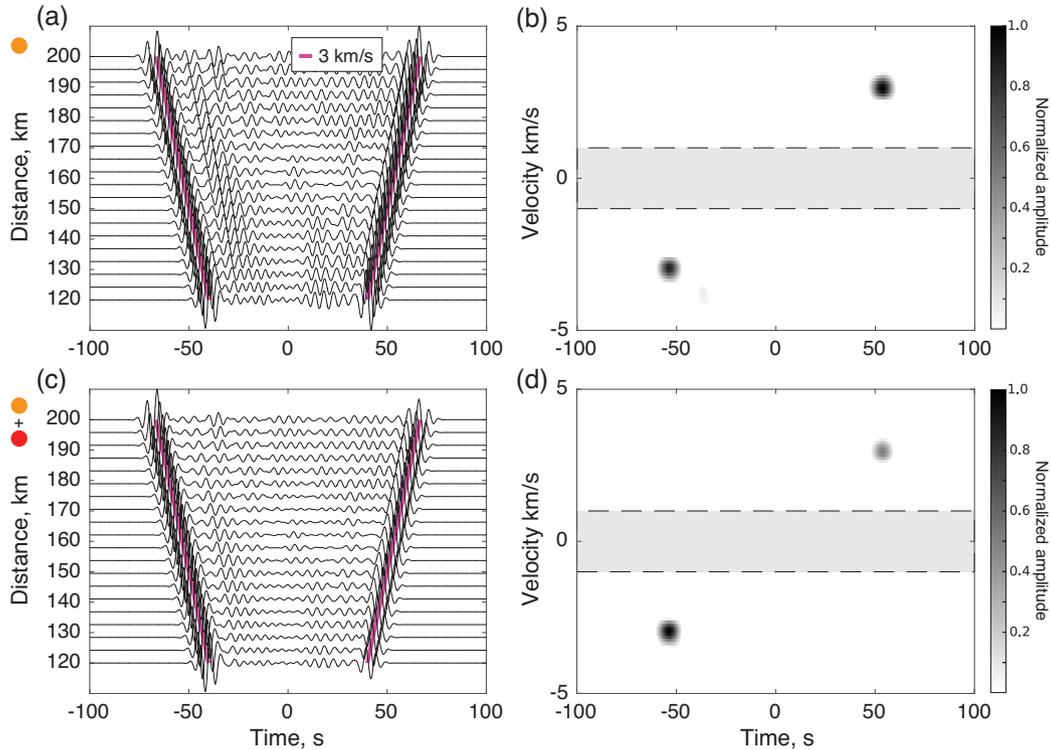


Figure 3. (a, c) Correlations between the reference and the other stations as a function of distance for (a) the homogeneous distribution (orange) and (c) homogeneous and asymmetric distribution (orange and red). (b, d) Normalized vespigrams obtained with the correlations in (a) and (c), respectively. The grey rectangles represent non tested velocities.

to be able to discriminate the direct arrival from the spurious arrival. We require a distance between the reference station and the expected source of signal of at least a wavelength to discriminate the direct arrival from the spurious arrival. For example at 5-s period and assuming a velocity of 3 km s^{-1} , the wavelength, and hence minimum allowable separation, is 15 km. In this study, we carefully observe the correlation signals to be sure that the signals extracted correspond to spurious arrivals.

If the source distribution is homogeneous, only the sources in the end-fire lobes interfere constructively and contribute to the signal. We create a homogeneous source distribution with 2000 sources spread randomly around the stations (orange dots in Fig. 2). As expected, this homogeneous source distribution leads to symmetric direct arrivals and similar causal and anti-causal wavelet amplitudes with an apparent velocity of 3 km s^{-1} (Fig. 3a). We compute the vespigram of the correlations (Rost & Thomas 2002) to extract the velocity and average arrival time of the waves at the stations. We compute the vespigram in two velocity windows: $[-5, -1] \text{ km s}^{-1}$ for the arrivals coming from the reference station to the other (primarily in the anticausal part) and $[5, 1] \text{ km s}^{-1}$ for the other direction (primarily in the causal part). As expected, the vespigram (Fig. 3b) highlights two arrivals at negative and positive times corresponding to the anticausal and causal arrivals on the correlations. The velocity extracted is 3 km s^{-1} , which is the source wavelet velocity. The correlation process thus recovers the wavelet well in this idealized homogeneous case.

If sources in one of the end-fire lobes are stronger, or if there are more of them, then the corresponding amplitude is larger on the correlation. To illustrate the effect of this heterogeneous source distribution, we add 500 sources to the homogeneous distribution on one side of the area (red dots in Fig. 2). As expected, the correlations

corresponding to waves coming from the reference station (anticausal side) have larger amplitudes (Fig. 3c) than the arrivals on the causal part of the correlations. This can also be observed on the vespigram, where the anticausal arrival is stronger than the causal arrival. However, the extracted velocity is still correct (Fig. 3d) for both arrivals on the causal and anticausal parts, and thus for end fire lobes with and without the added sources. This simple test agrees with different studies that have shown that a heterogeneous source distribution can still lead to a correct extraction of wave speed and arrival time, which would not lead to biased tomographic results (e.g. Yao & Van Der Hilst 2009; Tsai 2009).

If a source is not located in the end-fire lobes but is sufficiently strong, then its precursory arrival may not be suppressed by destructive interference during the correlation processing (Snieder 2006). We test the effects of different strong sources by adding 50 sources at specific locations (Fig. 2 and inserts Fig. 4).

When the strong source (or set of sources at the same location) is aligned with the stations (large blue dot in Fig. 2 and insert on Fig. 4a), a spurious arrival appears on the anticausal part of the correlations (Fig. 4a) with the velocity of the waves (Fig. 4b), supplementary to the direct arrivals. The arrival time corresponds to the arrival time from the source to the reference station minus the arrival time from the source to the other stations. Because the sources and stations are aligned, the propagation differences are only linked to the interstation distances. For this reason, the slope corresponds to the slope of the waves coming from the reference station, whose arrival times also vary with the distance between stations.

On the other hand, if the source is not aligned with the stations (large green dot in Fig. 2 and in the insert of Fig. 4c), the arrivals do not reflect the wave velocity (Figs 4c and d). The arrival times

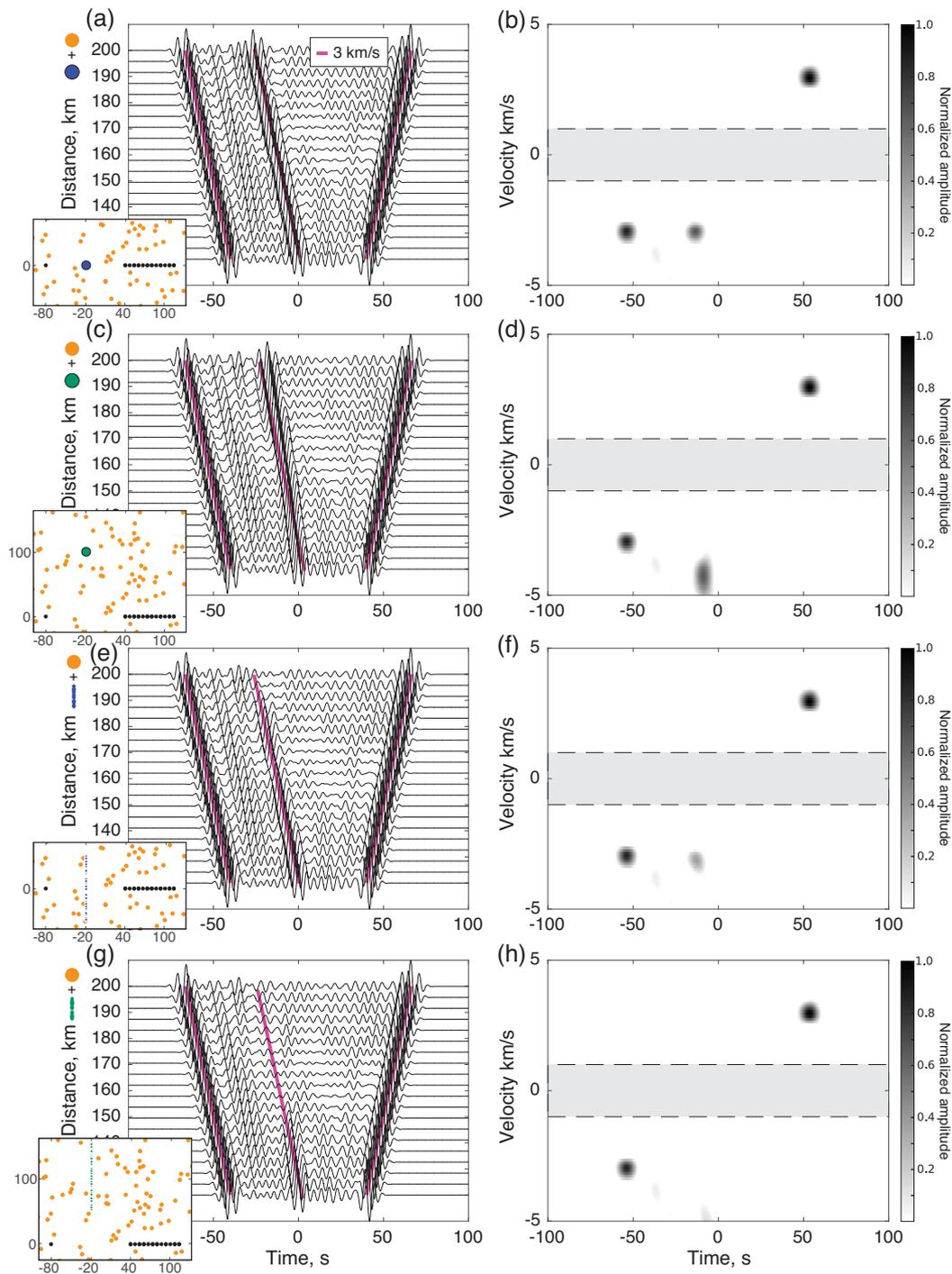


Figure 4. (a, c, e, g) Correlations between the reference station and other stations as a function of distance for the homogeneous distribution (orange) to which a concentration of 50 sources is added (a) at a single location aligned with the stations (large blue dot), (c) at a single location not aligned with the stations (large green dot), (e) along a line centred on the line of stations (small blue dots), (g) along a line not centre on the line of stations (small green dots). The magenta lines represent the 3 km s^{-1} and -3 km s^{-1} slopes. The inserts show a zoom of each source configuration around the stations and the concentrated sources (from Fig. 2). (b, c, f, h) Normalized vespagrams obtained with the correlations in a, c, e, g, respectively. The grey rectangles represent non-tested velocities.

only vary with the difference in distance from the source to the non-reference stations. In the non aligned isolated source case, the difference in distances from the source to the stations is smaller than the distances between the stations, leading to a higher apparent velocity. The arrivals thus appear faster than the wavelet velocity (around 4.5 km s^{-1} , instead of 3 km s^{-1}). This indicates that the

arrival time (and thus slope of different arrivals) contains information about the source location. If the slope corresponds to the velocity of the wavelet, its source is in line with the stations.

A strong scatterer acts as a secondary source and its effects are the same as a real source, as presented in this section. We expect to extract scattering effects linked to structural heterogeneities that

may not be concentrated in one location. For this reason we consider the case of scattering where the source is not a single point but spread along a line (blue and green lines in Fig. 2). This mimics a line of scatterers, for example linked to an interface that might represent a geologic boundary, such as a basin edge. If the line of scatterers is centred on the line of stations (blue line of sources in Fig. 2 and in the insert on Fig. 4e), the spurious arrival remains consistent with the source wavelet (Figs 4e and f), though the signal is smaller. In case of a line of sources away from the station line (green line of sources in Fig. 2 and in the insert on Fig. 4g); however, the arrivals are not coherent enough between the different station pairs and the spurious arrival does not appear on the vespagram (Figs 4g and h). This difference can be anticipated from the hyperbola of arrival times arising from sources located at different locations [see fig. 3 in Retailleau *et al.* (2017)]. If source locations vary along the Y -axis close to zero (blue), their arrival times on the correlations do not vary rapidly. That is not the case when the locations get farther from zero, implying strong destructive interactions among these sources.

These observations imply that the apparent velocity of the spurious arrival reflects the location of its source. We can be confident that a source that generates spurious arrivals with the same velocity as the direct arrival is in line with the stations. Hence, it is possible to use the velocity of the direct arrival as a detector for spurious arrivals linked to sources in line with the stations. Sources off line are more difficult to locate because they have weaker constructive interference. Moreover, a source away from the line of stations also implies wave propagation different from the propagation between stations (e.g. linked to model variations) and would result in a biased location estimate.

To overcome the influence of strong localized sources, passive imaging studies stack correlations over long times. Ocean microseism sources can be expected to move, so stacking averages out their signals. That is not the case for sources of scattered waves which remain at the same location. The arrivals thus stay consistent and, unlike ocean microseism sources, are enhanced by stacking. For these reasons, scatterer-generated spurious arrivals are a convenient way to localize structural effects. The line of stations installed in California is a good case study because it crosses several prominent structural boundaries at nearly right angles.

3 DATA SET AND METHODS

3.1 Correlation data set

We download the available 2 yr of 3-component continuous records for the 38 aligned stations of the Central California Seismic Experiment (CCSE 2015, fig. 1) from the Incorporated Research Institutions for Seismology (IRIS, <http://www.iris.edu/mda>) data services using **obspy** (Krischer *et al.* 2015). The correlation workflow follows Retailleau *et al.* (2017). We correct the time-series for instrument response for all daily records, window the time-series into 4-hr records, and suppress those with strong signals. We compute the coherence between all stations for these 4-hr records, stack the resulting correlation functions over all time periods, and rotate the signals from the (Z,E,N) components into the (Z,R,T) directions. We filter the signals between 5 and 10 s period because this period band is sensitive to the uppermost ten kilometers of the crust and has a high signal-to-noise ratio.

Fig. 5 shows the correlations obtained between the westernmost station with all of the others for (a) the vertical to vertical, or ZZ component and (b) the Transverse to Transverse, or TT component of the correlation tensor. The anticausal parts (negative time) correspond to waves travelling from the reference station to the others, that is waves traveling from the ocean towards the continent. The causal part corresponds to the waves traveling from the continent towards the ocean.

The correlation functions reflect the complexity of the structure of Central California. The Great Valley is a forearc basin with much lower velocity compared to the surrounding Coast ranges and Sierra Nevada (Lee *et al.* 2014). This low velocity layer leads to the generation of strong higher mode Rayleigh waves (Nayak & Thurber 2020). These arrivals appear clearly in the correlations with stations in the Great Valley (starting at distances around 120 km on Fig. 5a).

The correlations are strongly asymmetric, especially on the ZZ component (Fig. 5a), where clear signals are only observable on the anticausal part of the correlations. This is an indication that most of the ambient field sources are located in the nearby Pacific Ocean, as expected since the ocean interactions with the solid Earth dominate the seismic ambient field in this period band (5–10 s). For our line of receivers, most of the ocean sources are located in the end-fire lobe (for the anticausal part of the correlations), meaning that the Green's function should not be too strongly biased in arrival time. Indeed, the ocean sources located in the end fire lobes contribute directly to the main anticausal arrival. Because of their location in the end fire region, these ocean microseisms do not generate spurious arrivals that could be mistaken for scattering. Moreover, concerning the ocean sources outside of the end fire lobe, as seen in the previous section, sources far from the station line (Y far from zero) generate arrivals that will tend to interact destructively (Figs 4g and h). Finally, even assuming there is only one very localized and strong source, as seen in the previous section, the slope of its arrival is large compared to the Green's function (Figs 4c and d), and for that reason is not likely to bias the scattering detection, which searches arrivals of the direct arrival velocity. In summary, the ocean microseism sources to the west are the primary source of the ambient field in our data set, and do not generate spurious sources that will be extracted by our scatterer detector.

Since the line of stations crosses tectonically active interfaces, seismicity could also generate spurious arrivals and bias our scattering interpretations. This is unlikely since only the creeping section of the San Andreas Fault has appreciable seismicity and all of it is small, and hence deficient in low frequency energy. Moreover, during the processing of the correlations, segments of data with strong signals are dismissed (signals of amplitude greater than four times the standard deviation of the daily signals). This processing removes most of the seismicity with the potential to corrupt the correlations. If some seismicity energy remains in the 5–10 s period band and generates the observed spurious arrivals, shorter period correlations should exhibit stronger spurious energy. To confirm that this is not the case we filtered the correlations in two period bands, 2–7 s and 8–13 s (Fig. S1). The correlations filtered in the 2–7 s period band do not exhibit stronger spurious arrivals on the ZZ and TT components. For these reasons we are confident that the signals we observe are not linked to seismicity. With all this in mind, we note that both components exhibit several strong and systematic precursory arrivals that we interpret as due to scatterers.

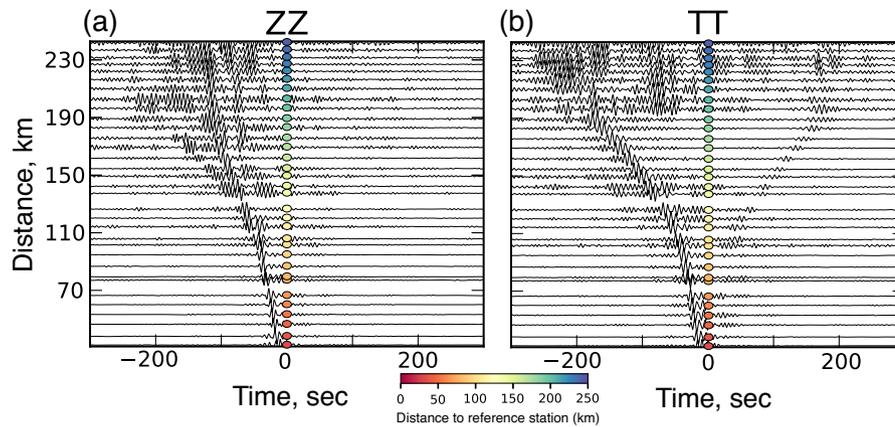


Figure 5. (a) Vertical to vertical (ZZ) and (b) transverse to transverse (TT) correlation functions as a function of distance between the westernmost station and the other stations. The colour, from red to blue represent the increasing distance of the different stations to the westernmost/reference station.

3.2 Testing source localization

Retailleau *et al.* (2017) used slant stacking to associate spurious arrivals with their source location and to locate microseism sources. In this paper, we locate scatterers using a similar process, by associating the arrivals on correlations obtained from the different station pairs. Following the logic of the previous sections, we extract spurious arrivals with a velocity that corresponds to the Green's function approximated by the correlation, to search for sources of scattering along the line of stations.

We first detail the steps of the method with the synthetic signals generated using a homogeneous source distribution to which a strong source is added, in the first case aligned with the stations (Figs 6a–d) and in the second case not aligned with the stations (Figs 6e–h) from the stations. We choose to place the supplementary source closer to the reference station (reference station at $x = -80$ km and source at $X = -65$ km, Figs 6a and e) than presented in Section 2 to simulate a less favourable case, which is also more similar to some of the real observations presented in later sections. Figs 6(c) and (g) represent the resulting correlation functions and Figs 6(b) and (f) the corresponding vespagrams. As expected, a source in line with the stations but closer to the reference station results in a signal closer to the main arrival and the same wave velocity (Figs 6b and c). The second source (Fig. 6e) generate an arrivals much earlier because it is located at $Y = 100$ km, and is still quite far from the reference station (Fig. 6e). Because this source is not aligned with the stations, the apparent velocity of the generated signals is higher than the velocity of the waves (Fig. 6f).

For the location process, we use the wave propagation velocity 3 km s^{-1} to search for sources of signals along a line from the reference station to the more distant stations with a spacing of 2 km (Figs 6a and e). For each potential source along the line, we first shift each correlation by the difference between the propagation time from the source to the reference station and the propagation time from the source and each station. Fig. 6 shows steps of the process with synthetic data for the on-line and off-line sources. We first consider the set-up where the strong isolated source of signals aligned with the stations (Figs 6a and d). Each source time arrival is simply the propagation time from the reference station location (same as the potential source) to itself (zero) minus the propagation time from the reference station location (same as the potential source) to the other stations (distance between the stations divided by the wave velocity 3 km s^{-1}). We shift the signals along the

corresponding arrival on 40sec windows (black curves on Fig. 6d for a tested source at $X = -65$ km) and then stack these signals and compute the envelope (red curve on Fig. 6d). The amplitude of the potential source at the tested location is the stack amplitude at $t = 0$ s. At $X = -65$ km along the testing line, the supplementary source is clearly highlighted after the velocity shift (Fig. 6d at $t = 0$ s).

We follow this process for all the potential source locations and obtain a probability of sources as a function of location (blue line on Fig. 6i). The final result along the line of potential source (blue on Fig. 6g) highlights the direct arrival on the anticausal part ($X = -80$ km), the source of the spurious arrival ($x = -65$ km) and the direct arrival on the causal part (starting at $X = 40$ km). We also applied this process for a case where the reference stations and the other stations are not distant from each other (Fig. S2). We thus simulate a case similar to our data set setting, also adding a noise source distribution asymmetry by using more sources on the reference side of the area. We are able to extract the location of the supplementary source in that setting. The Figure also shows that the influence of the direct arrival of the causal part on the results is more spread because the reference station and other stations are closer to each other. This influence is still dominated by the strong source.

As mentioned in Section 2, if the source of the spurious arrival is not located along the line of stations, its apparent velocity is faster than the wave velocity (Figs 4c, d and 6e, f). After a shift with the wave velocity of 3 km s^{-1} (Fig. 6f), no energy appears at the potential location $X = -65$ km. The total result along the line (green dashed line Fig. 6i), presents a small energy around $X = -35$ km which is very small compared to the result obtained for an aligned source. This confirms that the process is dominated by sources aligned with the stations.

Fig. 6(i) displays the source location results for a supplementary strong source aligned with the stations (blue curve), and one that is not aligned with the stations (dashed green curve). Not surprisingly, the results are identical except for the supplementary source location. As designed, the method correctly extracts the aligned source location and discards the mis-aligned source. As the potential source reaches the stations ($X = 40$ km), some energy appears on the stack. This is the contribution of the direct arrival of the stations closer to the reference station than the potential source. In this configuration the time-shift directly corresponds to the propagation between the reference station and the stations closer than the potential source. In these correlations the supplementary source contributes to the

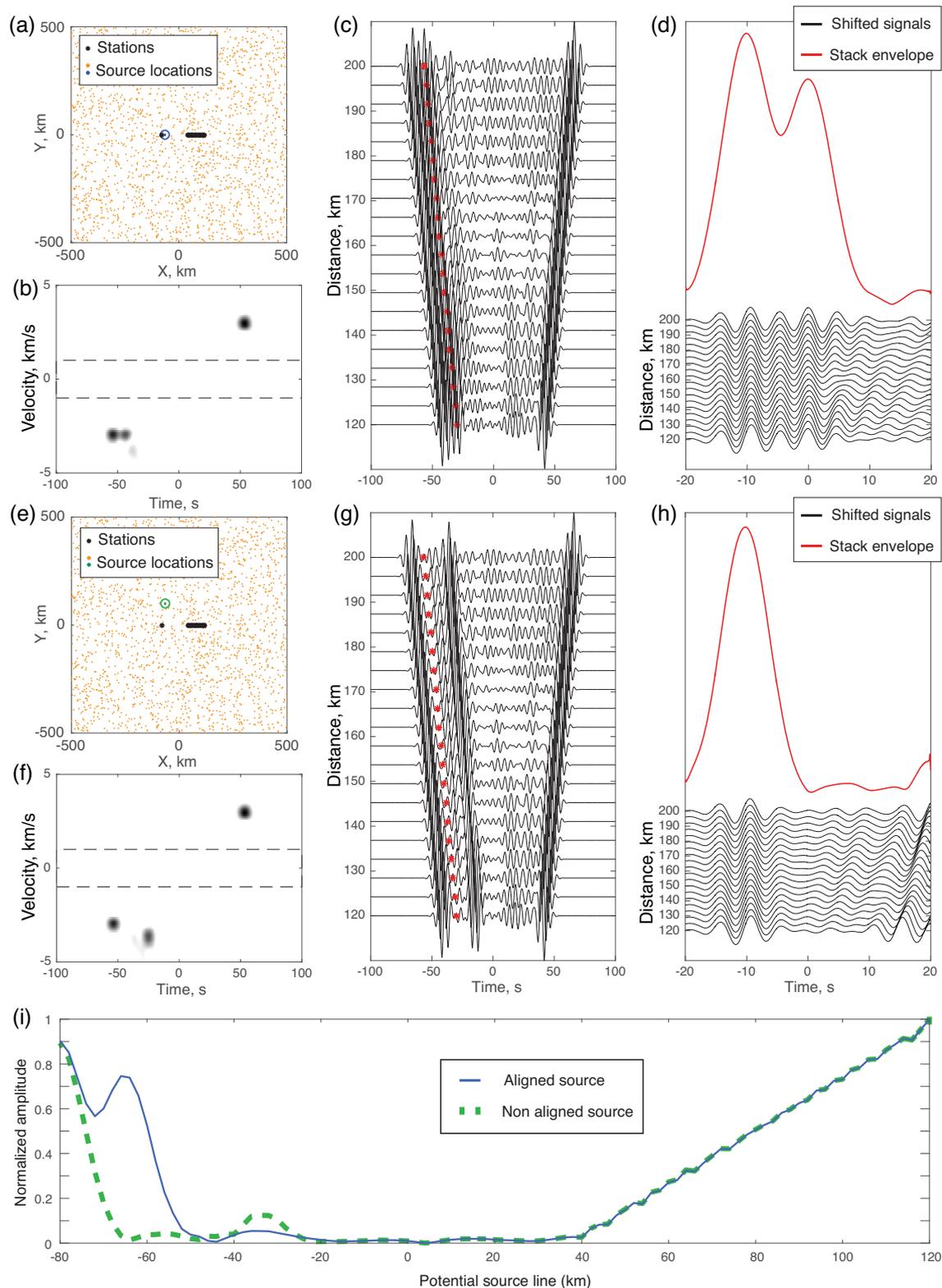


Figure 6. (a, e) Station and source configuration for a homogeneous source distribution to which is added a strong source (50 sources) (a) aligned with the stations ($X = -65$ km and $Y = 0$ km) and (e) not aligned with the stations ($X = -65$ km and $Y = 100$ km). (b, f) Corresponding vespagrams. (c, g) Corresponding correlation signals. The red stars represent the expected time arrival of the supplementary source along the line. (d, h) Shift and stack for the (d) supplementary source from (a) and (h) supplementary source from (e). (i) Location result along the search line for the aligned source case (blue) and the mis-aligned source case (dashed green).

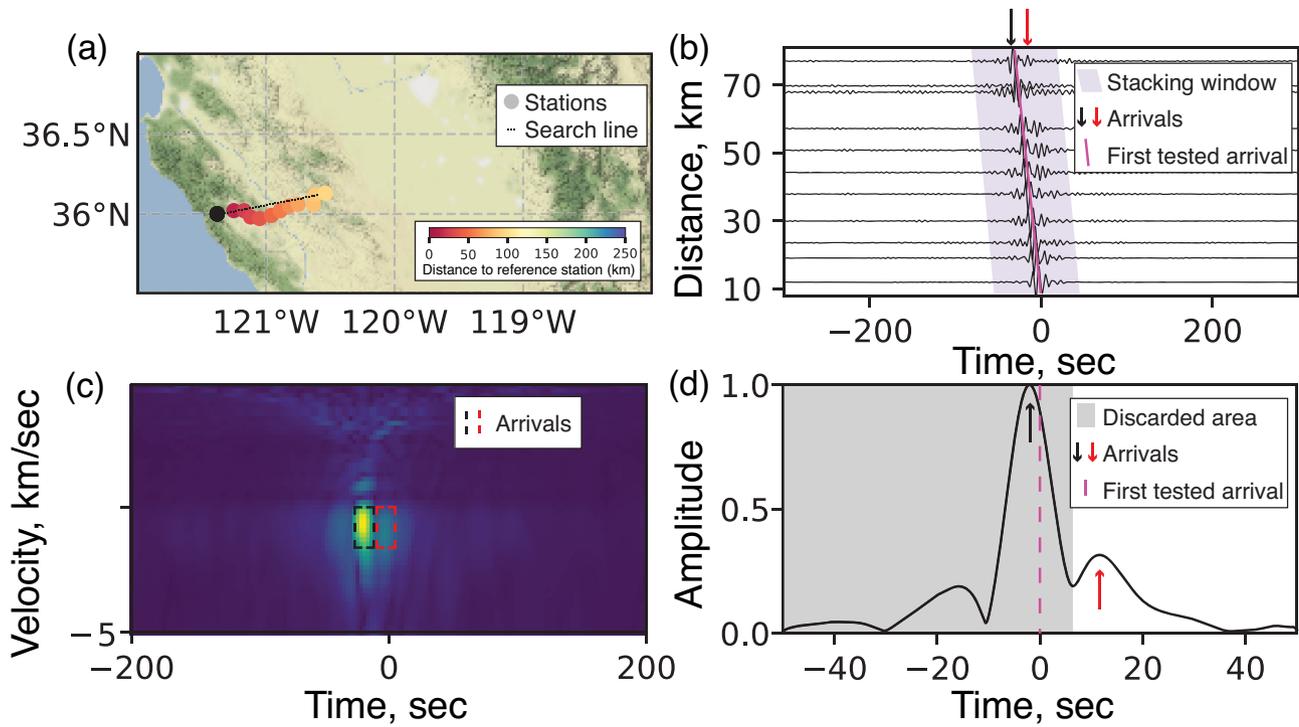


Figure 7. (a) Stations used for the location process. The black station is the reference. The dotted line represents the 2 km spaced locations of the search. (b) Correlations between the reference station and all others. The black arrow highlights the direct arrival and the red arrow marks the spurious arrival that we attribute to scattering. The pink line highlights the expected arrivals for a potential source located at the reference station. The purple patch represents the window along which the stacking is performed for this source. (c) Vespagram computed from the correlations. (d) Shifted stack from (b).

direct arrival of the causal part. For our California setting, most of the sources are located on the western side of the line of stations and the direct arrival on the causal part is very small. As a consequence, the contribution to the source search is mostly due to scattering and not the direct arrival. If there were strong energy on the causal part of the direct arrival, the reference stations and the other stations should be arranged to surround the sources of signals.

3.3 Localizing the scattering source

This area of study is very complex and the line of stations crosses several geological units. The correlation functions computed along the line show the effects of these structures with clear velocity variations of the main arrivals, both on the ZZ and TT components (Fig. 5). Locating scattering energy using the entire line of stations could be hindered by these strong velocity variations. For this reason, we separate the line into several segments to avoid biasing our observations by the strong velocity variations that occur between the Coast Ranges and the Great Valley. This allows us to identify the spurious arrivals visually and check their accurate extraction. Future studies could use varying velocities extracted from the move-out of the direct arrival or velocities obtained from earth models. The scatterers we locate by using supplementary arrivals also generate signals in the main arrival, in the more classic way they are observed. However, in this case, since we look for arrivals aligned with the stations, and assume single scattering, their signals are dominated by the direct propagation and difficult to extract. In 2-D setups, both methods could be combined to extract the scatterers.

Fig. 7 illustrates the different steps of the location process using the ZZ correlations computed from the Central California data, following the process presented in the previous section with synthetics.

Fig. 7(a) represents the segment of stations used, in the westernmost part. We choose to not use all the stations together due to strong velocity variations that interfere with a straightforward analysis. Nevertheless, the station density and number of stations considered allows us to image scatterers along the line. Fig. 7(b) represents the correlations as a function of distance to the reference station (black dot in Fig. 7a). The direct arrival (black arrow) is preceded by a clear spurious arrival (red arrow). We extract this arrival to localize its source, that is the scattering location.

We constrain our search to scattering sources along the line that follows the stations. The station spacing is around 10 km and we search scattering sources along a line of 2 km spacing. We add *a priori* information to the search by selecting the velocity. As discussed in the previous section, arrivals corresponding to sources away from the line tend to interact destructively and their apparent velocity increases as they are located farther from the line. We select the velocity from the direct arrival by computing a vespagram from the correlations (Fig. 7b), as in Figs 3 and 4. The vespagram (Fig. 7c) highlights the direct arrival (black square) and a secondary arrival (red square), which corresponds to our spurious arrival.

The location of an expected source can be extracted by combining the spurious arrivals observed on the different correlation functions. We search for scattering (source of a signal) along the line from the reference station to the most distant station. For each potential source location along the line, we select the arrival time window (100 s) centred at the difference between the expected arrivals from the source to the reference station and each of the other stations. Using such a long time window is not necessary at this frequency, but we use it here to illustrate the process. The pink line in Fig. 7(b) shows the expected arrival for a trial source at the

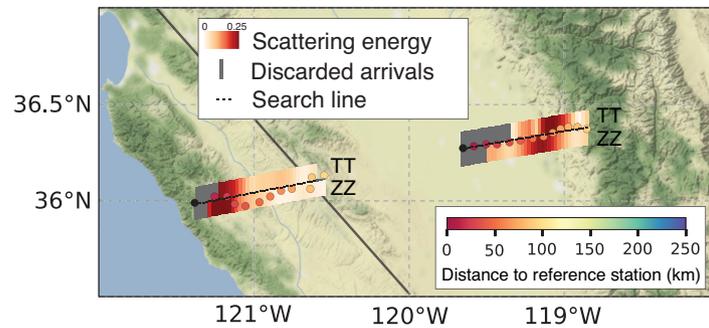


Figure 8. Location results using a segment in the west of the line, and one in the east for the ZZ and the TT components of the correlation tensor. The black dotted lines represent each location of the search. The coloured dots represent the stations used for the location. The black line represents the San Andreas fault.

location of the reference station and the shaded patch represents the corresponding stacking window. As expected, a source at the reference station location contributes to the expected propagation between the stations, and thus expected its time arrival corresponds to the direct arrival.

We stacked the signals and computed the envelope of the result (black curve in Fig. 7d, obtained with the same process as the synthetic results Fig. 6i) for this first potential source. The amplitude at $t = 0$ s (corresponding to the pink dashed line in Fig. 7d), which corresponds to the exact difference between the expected arrivals from the source to the reference station and the average of the stations (here the source is located at the reference station), is then extracted as the source amplitude. Because we do not want to extract the amplitude that corresponds to the propagation between the stations, we discard the strongest arrival. The grey patch in Fig. 7(d) shows the discarded window. We perform the same processing for locations along the line of the stations to determine the location of strongest energy. We normalize the amplitudes by the energy of the direct arrival. This curve Fig. 7(d) is our scattering energy. We represent it with colour on the map Fig. 8 (west segment ZZ). Similarly to Fig. 7(d), the locations of discarded arrivals are represented in grey.

4 RESULTS AND DISCUSSION

We extract scattering signals along two segments of the line separately, for two components of the correlation tensor, ZZ and TT. The first segment crosses the Coast Ranges and the other segment crosses the Eastern part of the Great Valley and the Western Sierra Nevada (Fig. 8). For the western part of the line (Figs 7 and 8), a clear arrival is observed on the ZZ component of the correlations (red arrow on Fig. 7b) as well as on the vespagram (red square on Fig. 7c). The arrival is obviously spurious, rather than part of the true Green's function because, at shorter distance, the arrivals cross the 0 time arrival and appear on the causal part instead of the anticausal part. After slant-stacking along the arrivals, a clear, spurious secondary arrival emerges (red arrow on Fig. 7d). After processing the signals of the TT component similarly to the ZZ component, strong scattering energy linked to spurious arrivals also emerges in Fig. 8 for both ZZ and TT. The arrival time of the direct arrivals on the correlations do not show any velocity variation (Fig. 7b or shorter distances in Fig. 5); however, a spurious arrival still appears. Jiang *et al.* (2018) performed tomography with a data set containing the stations used in this study; however, their study does not highlight a sharp variation of velocity at 7 s period,

which would coincide with our observation. This indicates that our scattering analysis could provide complementary information to tomographic analysis, as imaging using transmitted waves has different properties than imaging using scattered waves. The scatterer is located in the area of the Santa Lucia Range, close from the transition between Granitic and metamorphic rocks to marine sediments (Page *et al.* 1998), and close to the Rinconada Fault. Several high resolution 3-D tomography analyses were performed on Central California data sets in the Parkfield earthquake, using ambient noise correlations, earthquakes, control shots, quarry blasts and/or low-frequency earthquake observations (Zeng *et al.* 2016; Lippoldt *et al.* 2017; Zeng & Thurber 2019). These studies show velocity variations in the area where we extract a scatterer (northwest of these studies areas), although the variations are smoothed in those models.

We perform a search over potential scatterers on the eastern part of the line of stations, still using ZZ and the TT components (Fig. 8). We proceed as previously described, although we use two different velocities along the segment because of the large increase of velocity in the Sierra (larger distances in Fig. 5). We compute the vespagrams and extract the velocities for two sections of the lines, in the Great Valley and in the Sierra. For each tested source location, we use the Great Valley velocity to the west of the potential source location and the Sierra velocity to the east. The result (Fig. 8) clearly shows a scattering signal at the boundary between the Great Valley and the Sierra on both ZZ and TT. This observation coincides with the strong velocity variation observed by Jiang *et al.* (2018) East of the Great Valley at 7 s period. The scattering effect, however, seems to not be located at exactly the same place for the ZZ and the TT components, showing different effects of the structure on the Rayleigh and the Love waves.

We only extract weak scattering signals in the eastern part of the west segment that crosses the San Andreas Fault (Fig. 8). This is somewhat unexpected since the San Andreas is a major fault and could be expected to generate strong scattering. To explore this further, we focus our scattering analysis on the part of the line that crosses the San Andreas Fault. We analyse the Rayleigh wave arrival and the Love wave arrival, by using the ZZ and TT components. We use a shorter segment than the other tests to avoid the strong velocity variation entering the Great Valley and focus on the fault. We plot the stacked source amplitude at the reference station (Fig. 9a) to observe the spurious arrivals and compare the two components. As described in the method section, we discard the energy associated with the direct arrival. Figs 9(a) and (b) (west) show that a weak scattering signal emerges on the TT component but not on the ZZ component. This suggests that using different components of the correlation

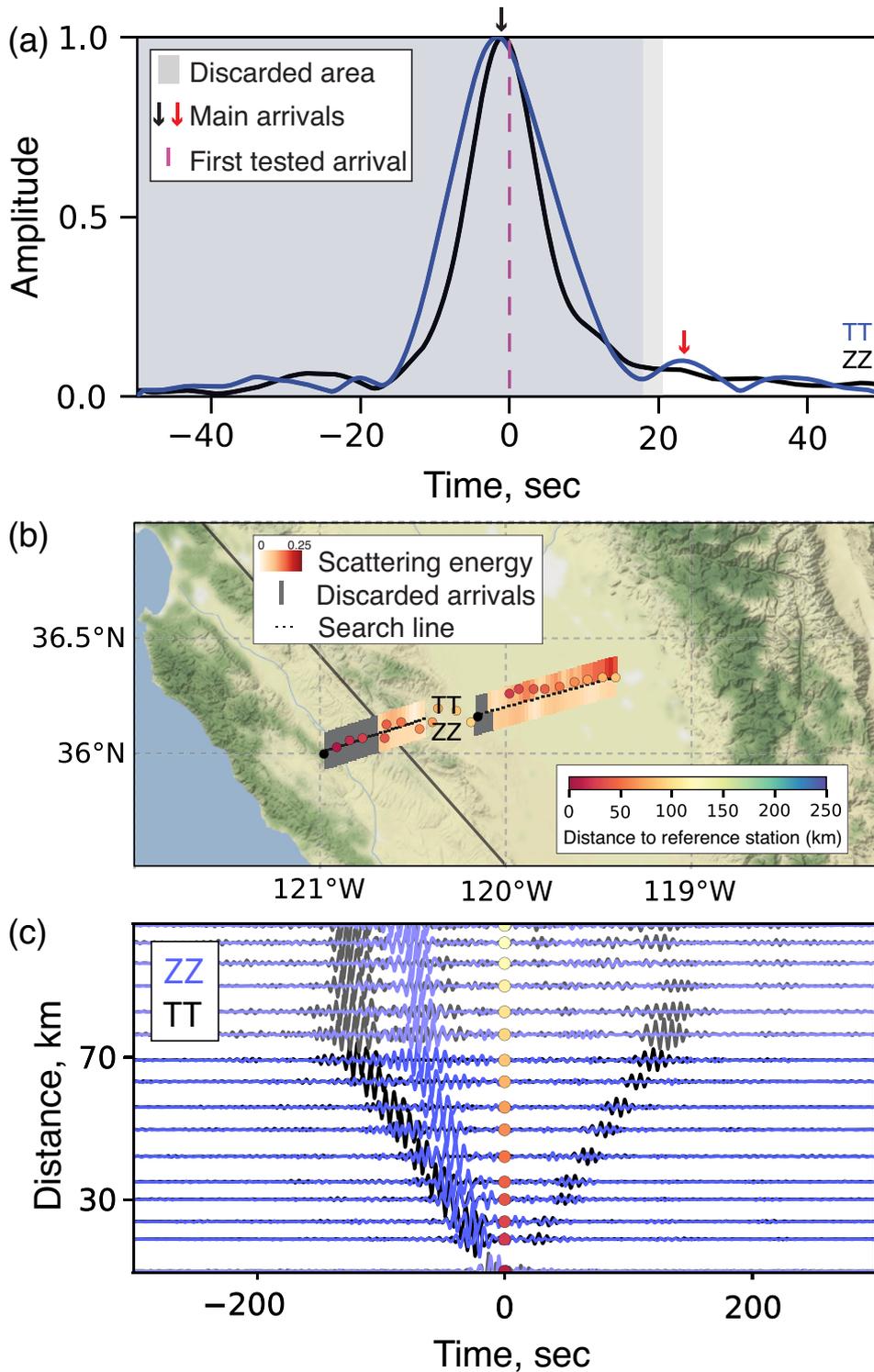


Figure 9. (a) Stacks of the correlations for a source located at the reference station for the ZZ (in black) and the TT (in blue) components on the western segment in b. The transparent patches show the discarded arrivals. (b) Location results across the San Andreas Fault and in the Great Valley using the ZZ and TT component of the correlations. The black dots represent the actual source locations tested. The black line represents the San Andreas Fault. Black dotted lines represent each location of the search. The coloured dots represent the stations used for the location. (c) Correlation functions used to extract scattering signals in the Great Valley. The correlations used for location in the Great Valley are represented with darker colours (distances from 10 to 70 km.).

functions lead to different observations of scattering effects, which is not unexpected given that different waves have different scattering patterns. Previous tomography studies show velocity variations in this zone between 5 and 10 s of period (Zeng *et al.* 2016; Lippoldt *et al.* 2017; Zeng & Thurber 2019). The low scattering extraction could be linked to the interface shape. This area may be too complex and the wave field could be subject to multiple scattering that our method is not set up to extract.

Finally, we test if the Great Valley shows the expected low scattering (Fig. 9b). The ZZ component confirms the low scattering effects, but the TT component seems to show some scattering signals, although there is not one dominant signal. This is due to the bias of the strong direct arrival on the causal part, which does not exist on the ZZ component. This influence of the causal part of the correlations is also showed in the synthetic examples (Fig. 6i and Fig. S2). Fig. 9(c) clearly shows the strong variation of amplitude symmetry between the causal and anticausal parts of the correlations on the ZZ and the TT components of the correlation tensor. While Pacific ocean microseism sources dominate the Rayleigh waves on the ZZ component, the Love waves observed on the TT component are either generated by sources coming from the opposite direction, or are more influenced by scattering from heterogeneity. As seen in Fig. 8, the eastern border of the Great Valley strongly scatters the signals. This implies that the direct arrivals in the anticausal part of the correlations is quite strong. As a consequence, as seen in Section 3.2, the direct arrival dominates the source extraction in the absence of strong spurious arrivals. This probably explains the signals that can be observed in the Great Valley, especially since this signal is stronger on the TT component, whose causal direct arrival is stronger than the ZZ signal.

5 CONCLUSIONS

Correlation functions computed between seismic stations contain information about the propagation between those stations but also information about the ambient field surrounding the stations. Strong sources located between the stations can generate spurious signals whose arrivals are earlier than the theoretical arrival of the fastest waves. Scatterers act as secondary sources that generate apparent local sources of waves that are originally excited by ocean microseisms. These scatterers can lead to spurious arrivals in ambient-field correlation functions that can be used for imaging after appropriate analysis. Their location indicates strong interfaces that can complement conventional tomography by illuminating the locations of strong lateral boundaries. We used a line of stations in Central California to extract scattered signals associated with the Rinconada Fault, the San Andreas Fault and the interface between the Great Valley and the Sierra Nevada. We also find that Rayleigh and Love waves are influenced differently by scattering effects. We were able to localize sources of scattering due to dense in-line station spacing.

3-D propagation modeling will be crucial in future work to assess the generations of these signals in the different components of the correlation tensor and their usefulness for imaging. Indeed, the ZZ and TT components show different behaviour, and spurious signals can also be observed on non-diagonal components of the correlation tensor. 3-D simulations will also permit to assess depths and 3-D propagation effects. If future developments in seismic monitoring allow such sampling in two dimensions, we can anticipate that it will be possible to reconstruct the geometry of structures that lead to scattering in detail. This method could be particularly useful with the development of large-N arrays and fiber optic measurements

(distributed acoustic sensing). Finally, this study presents an analysis of surface waves propagating in a regional setting, but there is no reason not to use spurious body wave phases for global scale imaging as well (Wang & Tkalčić 2020).

ACKNOWLEDGEMENTS

The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used to access the waveforms, related metadata and/or derived products used in this study (<http://www.iris.edu/mda>). The IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the National Science Foundation under Cooperative Agreement EAR-1261681. LR thanks Pierre Boué who provided the scripts that served as a basis for the computations used for this analysis. This research was supported by Pacific Gas and Electric and by the Southern California Earthquake Center (Contribution No. 10008). SCEC is funded by NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047. The authors thank the editor and two anonymous reviewers for their useful comments.

REFERENCES

- Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A.L., Lin, F., Moschetti, M., Shapiro, N. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**(3), 1239–1260.
- Boué, P., Poli, P., Campillo, M., Pedersen, H., Briand, X. & Roux, P., 2013. Teleseismic correlations of ambient seismic noise for deep global imaging of the earth, *Geophys. J. Int.*, **194**(2), 844–848.
- Campillo, M. & Roux, P., 2015. Crust and lithospheric structure - seismic imaging and monitoring with ambient noise correlations, in *Treatise on Geophysics*, 2nd edn, pp. 391–417, ed. Schubert, G., Elsevier.
- CCSE, 2015. Central California seismic experiment, caltech, Other/Seismic Network.
- Derode, A., Larose, E., Tanter, M., De Rosny, J., Tourin, A., Campillo, M. & Fink, M., 2003. Recovering the greens function from field-field correlations in an open scattering medium (I), *The J. acoust. Soc. Am.*, **113**(6), 2973–2976.
- Ermert, L., Villaseñor, A. & Fichtner, A., 2016. Cross-correlation imaging of ambient noise sources, *Geophys. J. Int.*, **204**(1), 347–364.
- Fichtner, A., Stehly, L., Ermert, L. & Boehm, C., 2016. Generalised interferometry – I. Theory for inter-station correlations, *Geophys. J. Int.*, **208**(2), 603–638.
- Gouedard, P., *et al.*, 2008. Cross-correlation of random fields: mathematical approach and applications, *Geophys. Prospect.*, **56**(3), 375–393.
- Harmon, N., Rychert, C. & Gerstoft, P., 2010. Distribution of noise sources for seismic interferometry, *Geophys. J. Int.*, **183**(3), 1470–1484.
- Hasselmann, K., 1963. A statistical analysis of the generation of microseisms, *Rev. Geophys.*, **1**(2), 177–210.
- Jiang, C., Schmandt, B., Hansen, S.M., Dougherty, S.L., Clayton, R.W., Farrell, J. & Lin, F.-C., 2018. Rayleigh and s wave tomography constraints on subduction termination and lithospheric foundering in central California, *Earth planet. Sci. Lett.*, **488**, 14–26.
- Kaneshima, S., 2016. Seismic scatterers in the mid-lower mantle, *Phys. Earth planet. Inter.*, **257**, 105–114.
- Kawakatsu, H., Yamamoto, M., Kaneshima, S. & Ohkura, T., 2011. Comment on a persistent localized microseismic source near the Kyushu island, Japan by Xiangfang Zeng and Sidao Ni, *Geophys. Res. Lett.*, **38**, L17307.
- Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.*, **8**(1), 014003.

- Lee, E.-J., Chen, P., Jordan, T.H., Maechling, P.B., Denolle, M.A. & Beroza, G.C., 2014. Full 3-d tomography for crustal structure in southern California based on the scattering integral (SI) and adjoint-wavefield methods, *J. geophys. Res.*, **119**(8), 6421–6451.
- Lin, F.-C., Li, D., Clayton, R.W. & Hollis, D., 2013a. High-resolution 3d shallow crustal structure in long beach, California: application of ambient noise tomography on a dense seismic array, *Geophysics*, **78**(4), Q45–Q56.
- Lin, F.-C., Tsai, V.C., Schmandt, B., Duputel, Z. & Zhan, Z., 2013b. Extracting seismic core phases with array interferometry, *Geophys. Res. Lett.*, **40**(6), 1049–1053.
- Lippoldt, R., Porritt, R.W. & Sammis, C.G., 2017. Relating seismicity to the velocity structure of the San Andreas fault near Parkfield, CA, *Geophys. J. Int.*, **209**(3), 1740–1745.
- Liu, X., Beroza, G.C. & Nakata, N., 2019. Isolating and suppressing the spurious non-diffuse contributions to ambient seismic field correlations, *J. geophys. Res.*, **124**(9), 9653–9663.
- Longuet-Higgins, 1950. A theory of the origin of microseisms, *Phil. Trans. R. Soc. Lond., A*, **243**(857), 1–35.
- Ma, Y., Clayton, R.W., Tsai, V.C. & Zhan, Z., 2013. Locating a scatterer in the active volcanic area of southern Peru from ambient noise cross-correlation, *Geophys. J. Int.*, **192**(3), 1332–1341.
- Margerin, L., Campillo, M., Van Tiggelen, B. & Hennino, R., 2009. Energy partition of seismic coda waves in layered media: theory and application to pinyon flats observatory, *Geophys. J. Int.*, **177**(2), 571–585.
- Nayak, A. & Thurber, C.H., 2020. Using multi-component ambient seismic noise cross-correlations to identify higher mode Rayleigh waves and improve dispersion measurements, *Geophys. J. Int.*, **222**(3), 1590–1605.
- Nishida, K., Kobayashi, N. & Fukao, Y., 2000. Resonant oscillations between the solid earth and the atmosphere, *Science*, **287**(5461), 2244–2246.
- Obermann, A., Planès, T., Hadziioannou, C. & Campillo, M., 2016. Lapse-time-dependent coda-wave depth sensitivity to local velocity perturbations in 3-D heterogeneous elastic media, *Geophys. J. Int.*, **207**(1), 59–66.
- Page, B.M., Coleman, R.G. & Thompson, G.A., 1998. Overview: late Cenozoic tectonics of the central and southern coast ranges of California, *Bull. geol. Soc. Am.*, **110**(7), 846–876.
- Planes, T., Larose, E., Margerin, L., Rossetto, V. & Sens-Schoenfelder, C., 2014. Decorrelation and phase-shift of coda waves induced by local changes: multiple scattering approach and numerical validation, *Waves Rand. Complex Media*, **24**(2), 99–125.
- Retailleau, L., Boué, P., Stehly, L. & Campillo, M., 2017. Locating microseism sources using spurious arrivals in intercontinental noise correlations, *J. geophys. Res.*, **122**(10), 8107–8120.
- Retailleau, L., Boué, P., Li, L. & Campillo, M., 2020. Ambient seismic noise imaging of the lowermost mantle beneath the North Atlantic Ocean, *Geophys. J. Int.*, **222**(2), 1339–1351.
- Ritzwoller, M.H. & Feng, L., 2018. Overview of pre-and post-processing of ambient noise correlations, in *Seismic Ambient Noise*, pp. 144–187, eds Nakata N., Gualtieri L. & Fichtner A., Cambridge Univ. Press.
- Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.*, **40**(3), 1008.
- Roux, P., Sabra, K.G., Kuperman, W.A. & Roux, A., 2005. Ambient noise cross correlation in free space: theoretical approach, *J. acoust. Soc. Am.*, **117**(1), 79–84.
- Sager, K., Ermert, L., Boehm, C. & Fichtner, A., 2017. Towards full wave-form ambient noise inversion, *Geophys. J. Int.*, **212**(1), 566–590.
- Sánchez-Sesma, F.J. & Campillo, M., 2006. Retrieval of the Green's function from cross correlation: the canonical elastic problem, *Bull. seism. Soc. Am.*, **96**(3), 1182–1191.
- Sato, H., 2019. Power spectra of random heterogeneities in the solid earth, *Solid Earth*, **10**(1), 275–292.
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.
- Shapiro, N.M., Ritzwoller, M. & Bensen, G., 2006. Source location of the 26 sec microseism from cross-correlations of ambient seismic noise, *Geophys. Res. Lett.*, **33**(18), doi:10.1029/2006GL027010.
- Snieder, R., 2006. The theory of coda wave interferometry, *Pure appl. Geophys.*, **163**(2–3), 455–473.
- Stehly, L., Campillo, M. & Shapiro, N., 2006. A study of the seismic noise from its long-range correlation properties, *J. geophys. Res.*, **111**(B10), doi:10.1029/2005JB004237.
- Stutzmann, E., Roullet, G. & Astiz, L., 2000. Geoscope station noise levels, *Bull. seism. Soc. Am.*, **90**(3), 690–701.
- Tromp, J., Luo, Y., Hanasoge, S. & Peter, D., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.*, **183**(2), 791–819.
- Tsai, V.C., 2009. On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophys. J. Int.*, **178**(3), 1555–1564.
- Wang, S. & Tkalčić, H., 2020. Seismic event coda-correlation: toward global coda-correlation tomography, *J. geophys. Res.*, **125**(4), doi:10.1029/2019JB018848.
- Wapenaar, K. & Fokkema, J., 2006. Greens function representations for seismic interferometry, *Geophysics*, **71**(4), SI33–SI46.
- Wapenaar, K., Slob, E. & Snieder, R., 2006. Unified greens function retrieval by cross correlation, *Phys. Rev. Lett.*, **97**(23), 234301.
- Weaver, R.L. & Lobkis, O.I., 2004. Diffuse fields in open systems and the emergence of the Green's function (I), *J. acoust. Soc. Am.*, **116**(5), 2731–2734.
- Webb, S.C., 1998. Broadband seismology and noise under the ocean, *Rev. Geophys.*, **36**(1), 105–142.
- Yang, Y. & Ritzwoller, M.H., 2008. Characteristics of ambient seismic noise as a source for surface wave tomography, *Geochem. Geophys. Geosyst.*, **9**(2), doi:10.1029/2007GC001814.
- Yang, Y., Ritzwoller, M.H., Levshin, A.L. & Shapiro, N.M., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.*, **168**(1), 259–274.
- Yao, H. & Van Der Hilst, R.D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to se Tibet, *Geophys. J. Int.*, **179**(2), 1113–1132.
- Zeng, X. & Ni, S., 2010. A persistent localized microseismic source near the Kyushu Island, Japan, *Geophys. Res. Lett.*, **37**(24), doi:10.1029/2010GL045774.
- Zeng, X. & Thurber, C., 2019. Three-dimensional shear wave velocity structure revealed with ambient noise tomography in the Parkfield, California region, *Phys. Earth planet. Inter.*, **292**, 67–75.
- Zeng, X., et al., 2016. 3-D P- and S-wave velocity structure and low-frequency earthquake locations in the Parkfield, California region, *Geophys. J. Int.*, **206**(3), 1574–1585.
- Zheng, Y., Shen, W., Zhou, L., Yang, Y., Xie, Z. & Ritzwoller, M.H., 2011. Crust and uppermost mantle beneath the north China Craton, northeastern china, and the sea of Japan from ambient noise tomography, *J. geophys. Res.*, **116**(B12), doi:10.1029/2011JB008637.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1 (a, c) vertical to vertical (ZZ) and (b, d) transverse to transverse (TT) correlation functions as a function of distance to the westernmost station and the other stations. The correlations are filtered in the (a, b) 2–7 s period band and (c, d) 8–13 s period band.

Figure SD2 Location process on synthetic with a station and spurious source configuration close to the CCSE line. (a) Station (black line) and source combination for the synthetic test close to the data configuration. 1000 sources are spread over the all area as the homogeneous distribution (orange dots). 1000 sources are spread on the western side of the area to simulate the source distribution asymmetry (red dots). 50 identical sources in $x = -65$ km are used to simulate a strong scatterer/source of energy. Contrarily to the other examples, the station line is continuous and the source is not

isolated. (b) Resulting correlations functions along the line. The red dots represent the expected time arrivals of the strong source. (c) Location result along the search line which highlights the main arrival and the supplementary source at $x = -65$ km.

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