Using a Time-Based Subarray Method to Extract and Invert Noise-Derived Body Waves at Long Beach, California

Jorge C. Castellanos\(^1\), Robert W. Clayton\(^1\), and Alan Juarez\(^2\)

\(^1\)Seismological Laboratory, Department of Earth and Planetary Sciences, Caltech, Pasadena, California, USA, \(^2\)Department of Earth Sciences, University of Southern California, Los Angeles, California, USA

Abstract  The reconstruction of body waves from the cross-correlation of random wavefields has recently emerged as a promising approach to probe the fine-scale structure of the Earth. However, because of the nature of the ambient noise field, the retrieval of body waves from seismic noise recordings is highly challenging and has only been successful in a few cases. Here, we use seismic noise data from a 5,200-node oil-company survey to reconstruct body waves and determine the velocity structure beneath Long Beach, California. To isolate the body wave energy from the ambient noise field, we divide the entire survey into small-aperture subarrays and apply a modified double-beamforming scheme to enhance coherent arrivals within the cross-correlated waveforms. The resulting beamed traces allow us to identify clear refracted \(P\) waves traveling between different subarray pairs, which we then use to construct a high-resolution 3D velocity model of the region. The inverted velocity model reveals velocity variations of the order of 3% and strong lateral discontinuities caused by the presence of sharp geologic structures such as the Newport-Inglewood fault (NIF). Additionally, we show that the resolution that is achieved through the use of high-frequency body waves allows us to illuminate small geometric variations of the NIF that were previously unresolved with traditional passive imaging methods.

1. Introduction

Detailed knowledge of the subsurface velocity structure is essential for predicting ground motion and, thus, earthquake-hazard assessment. Over the past few decades, significant advancements in imaging the Earth's interior have been possible with the advent of ambient noise cross-correlation. This method goes beyond the spatial limitations of classical earthquake seismology and uses the Earth's background vibrations (i.e., ambient noise field) recorded at a pair of synchronous seismic stations to reconstruct Green's function between the two stations (Campillo & Paul, 2003; Lobkis & Weaver, 2001; Shapiro & Campillo, 2004). Both the robustness and practicality of this technique have allowed the seismological community to investigate the mechanical properties of the shallow earth on regional (e.g., Sabra et al., 2005; Shapiro et al., 2005) and continental scales (e.g., Lin et al., 2008; Villasenor et al., 2007; Yang et al., 2007). However, as the number of high-density seismic networks continues to increase, it is being revealed that near-surface velocities can have large lateral contrasts that result in peak ground acceleration variations of a factor of 5 over a horizontal length scale of less than a kilometer (Clayton et al., 2012). This scale of variations motivates our necessity to push the limits of modern seismic imaging and explore the use of new generation instrumentation and processing techniques to resolve the fine-scale velocity structure of the Earth.

During the first half of 2011, an unprecedentedly dense seismic network was deployed in Long Beach, California, area as part of a petroleum industry survey. This network consisted of more than 5,200 vertical velocity sensors that were distributed over a \(7 \times 10\) km area at the outboard transition from continental southern California to the Inner Borderland, across the Newport-Inglewood fault system (Figure 1). While the general purpose of this survey was to perform conventional active-source imaging, Lin et al. (2013) showed that high-frequency (0.5–4 Hz) ambient noise surface waves recorded by the network could also be used to construct a detailed 3D shear-wave velocity model for the top 800 m of the crust. The clear correlation of the surface wave velocity model with the known geologic structure proved that passive seismic data can constrain the subkilometer-scale material properties of the subsurface. More recent work has shown...
that this type of industry-based instrumentation can further be used to investigate local microseismicity (e.g., Li et al., 2018), the mechanics of active fault zones (e.g., Inbal et al., 2016), and diverse characteristics of the ambient seismic wavefield (e.g., Spica et al., 2018).

As the use of dense seismic networks grew more common, it also became evident that it is possible to extract body waves from the cross-correlation of the ambient noise field (Olivier et al., 2015; Roux et al., 2005; Spica et al., 2018). This topic has attracted considerable attention from the geophysical community since body waves penetrate deep into the Earth and can probe the seismic structure with higher resolution than surface waves (e.g., Nakata et al., 2015, 2016). However, because ambient noise energy propagates mainly horizontally (Aki & Richards, 2002; Shapiro & Campillo, 2004; Sabra et al., 2005), reconstructing body waves from seismic noise recordings has proven to be a difficult task (Nakata & Nishida, 2019; Snieder & Larose, 2013). In a recent study, Nakata et al. (2015) used a 2,500-station array adjacent to the Long Beach survey to extract body waves and performed the first-ever body wave tomography using pure ambient noise recorded at the ground surface. This investigation, however, used a series of selection filters on individual waveforms to isolate the body wave energy, which, based on a set of quality control criteria, only allowed the use of a small portion of the entire data set (about 35%).

In this study, we extend the work of Nakata et al. (2015) and use the Long Beach network to extract ambient noise body waves and map the velocity structure beneath the array. First, we validate the existence of coherent body wave arrivals in the cross-correlated waveforms by generating a stacked record section of the Long Beach data set, inverting for a 1D velocity model, and computing a synthetic wavefield. We then divide the entire array into several small-aperture subarrays and apply an array processing technique to retrieve first-arrival body waves traveling between the different groups of stations. Finally, we make reliable
traveltime measurements and perform a 3D traveltime tomography to resolve the $P$ wave velocity variations of the uppermost part of the crust.

2. Ambient Noise Correlation and Body Waves

For the Long Beach survey, Lin et al. (2013) built a complete set of more than 5,200 virtual sources (i.e., one for each physical receiver) by cross-correlating the receiver at the source location and every other receiver in the array. The processing scheme used to generate such data set followed the method of Bensen et al. (2007) (without the temporal normalization step) and yielded more than 13.5 million ambient noise cross-correlations. Aside from being able to retrieve clear fundamental Rayleigh waves, Lin et al. (2013) showed that weak yet coherent body waves existed in the Long Beach correlograms and that they could be readily discern above the noise after stacking over all receiver pairs into discrete offset bins (Figure 2). These arrivals, however, are not prominent enough to be reliably identified in individual traces. In this study, we use the same gathers and array processing tools to extract body waves traveling between small groups of stations.

A close inspection into the first 5 seconds of the stacked correlated waveforms reveals that body waves start propagating at approximately time zero with velocities larger than 1.7 km/s (Figure 3A). As in the work of Nakata et al. (2015), we find that the apparent wave speed of the first arrival increases with offset, which suggests that the reconstructed wave corresponds to a refracted or diving $P$ wave. Moreover, we observe two distinct pulses arriving shortly after the first arrival that have slightly different horizontal slownesses and, hence, different propagation paths (Figure S1). To determine the nature of these arrivals, we convert the $P$ wave portion of the stacked cross-correlations into the $Tau-P$ domain and use this representation to invert for a smoothly variant 1D velocity model (Figure 4 Shearer, 2019). We then use the inverted velocity profile to perform a 2D finite difference simulation (Li et al., 2014) and resolve whether these arrivals are the result of simple 1D wave propagation or due to large lateral velocity variations within the array. Our analysis shows that a simple 1D model is sufficient to replicate all three arrivals and that they correspond to a diving $P$ wave, a $PP$ wave, and a direct $P$ wave that is propagating close to the surface (Figures 3B and 3C). The strong resemblance between the synthetic and observed waveforms proves the correlation’s convergence to Green’s function and confirms the presence of coherent body wave energy in the Long Beach correlograms.

3. Double Beamforming and Body Wave Extraction

Although $P$ waves are observable in the stacked correlograms, the generation of a high-resolution velocity model requires a complete characterization of the different wave propagation properties within the seismic survey. To this end, we apply a slightly modified double-beamforming (DBF) scheme to extract first-arrival $P$ waves propagating between different sections of the array. The DBF method (Boué et al., 2013) combines classical slant stack processing (Rost & Thomas, 2002; Thorson & Claerbout, 1985; Weber et al., 1996) on a source and receiver array to reconstruct a beam or eigenray between the center-point of the two groups of stations. In short, this technique entails finding the appropriate slowness, $u$, and azimuthal direction, $\theta$, of a given phase simultaneously on both sides by applying a systematic delay-and-sum to all the recordings. Following the notation of Nakata et al. (2016), the double-beam space, $B$, at a given time, $t$, can be constructed by scanning the slowness and azimuth domains at the source and receiver locations through the computation of

$$B(u_s, \theta_s, u_r, \theta_r, t) = \frac{1}{NN_t} \sum_{x_s} \sum_{y_s} \sum_{x_r} \sum_{y_r} C(x_s, y_s, x_r, y_r, t = \tau_s(x_s, y_s, u_s, \theta_s) + \tau_r(x_r, y_r, u_r, \theta_r)).$$

where the subscripts $s$ and $r$ refer to the sources and receivers, respectively, $N$ is the number of source and receivers, $C$ is the cross-correlation function between the spatial points $x$ and $y$ (e.g., eastward
Figure 3. Observed and synthetic wavefields windowed around the expected P-wave time of arrival. (A) Record section of the first 5 seconds of the stacked cross-correlations. The thick black lines mark 3 different reference velocities that approximate the variations in the apparent wave speed of the first-arrival P-wave as a function of offset. (B) Synthetic wavefield computed by 2D finite difference modeling using the inverted 1D velocity model (black profile in Figure 4B). The frequency range of the waveforms is the same as Figure 2. (C) Snapshots of the 2D velocity wavefield at different times (1.5 s, 2.5 s and 3.5 s). The explosion at the surface marks the location of the source. Note that attenuation was not considered in the wavefield simulation.

and northward), and \( \tau \) is the relative time lag from a reference point, which, for a 2D seismic array, can be defined as

\[
\tau(x, y, u, \theta) = u(x - x^c) \sin(\theta) + u(y - y^c) \cos(\theta),
\]

wherein the coordinates \( x^c \) and \( y^c \) represent the center of an array. Because of the large number of traces that are generally involved in the construction of a double-beam, the improvement on the signal-to-noise ratio is significant and, therefore, ideal to bring out the weak body waves that are buried within the correlograms noise.

As in every array-processing problem, it is desirable to explore the complete model space to find the optimal pairs of slownesses and directional azimuths that would result in the most energetic and coherent beam. However, due to the large number of instruments in the Long Beach survey, scanning through the 5D space of the double-beam for every possible source and receiver array configuration is computationally very demanding and basically impractical. To overcome this challenge, we apply two modifications to the traditional DBF method. First, we assume that, because of the relatively short interstation distances of the array, the diving P-waves do not suffer large deviations from their direct raypaths. This assumption allows us to collapse the directional azimuth dimensions, \( \theta_s \) and \( \theta_r \), by fixing their optimal values to the azimuth between the geographic center of the source and receiver arrays. Synthetic tests using the real Long Beach station geometry indicate that this simplification is reasonably valid even for lateral velocity contrasts as large as 20\% (Figure S2). Second, instead of scanning through the remaining slowness and time space,
we compute the time-frequency spectra of each individual record via an $S$ transform (Stockwell et al., 1996) and, from this representation, extract the timing of every energy peak that is ±0.2 seconds within the predicted arrival time of the diving $P$ wave (Figure 5). The length of the tolerance time window is chosen based on the fact that we are focusing on body waves that are mostly higher than 3 Hz (Nakata et al., 2015) and we only want to include a maximum of three wavelets to avoid severe cycle skipping. Subsequent analyses of our DBF results suggests that most of the shallow crustal velocity variations of Long Beach can be captured with this window length and that its broadening does not change the first-order distribution of our traveltime observations (Figure S3). As an aside, it is worth noting that here we do not use nor retain any frequency information that is contained in the $S$ transform spectrograms as we opt to degrade their frequency resolution in order to enhance their temporal resolution and obtain finer traveltime picks (Mansinha et al., 1997).

After collecting the time of arrival of every potential refracted $P$ wave, we cast the entire double-beam calculation as a linear inverse problem and simultaneously solve for the source and receiver side slownesses that best predict the traveltime data set via weighted least-squares. In other words, we minimize the prediction error $E = e^T W e$ and solve,

$$m = (G^T W G)^{-1} G^T W d,$$

where $m$ is a vector containing the source-side and receiver-side slowness, $G$ is a sensitivity matrix that relates the source and receiver array geometries to the traveltime data, $W$ is a weighting matrix, and $d$ is a vector containing the traveltime picks. To weight the inversion, we use the absolute amplitude of the picked energy contours of the $S$ transformed correlograms. It is important to point out that the amplitudes of the $P$ wave signals largely depend on the local noise distribution and are only used here to prioritize the alignment of energetic arrivals rather than the one of weak and random oscillations within the traces. This last step turns our processing scheme into something that is equivalent to finding the source-side and receiver-side velocities that best maximizes the energy of the stacked waveforms around the predicted time of arrival of the refracted $P$ wave.

To implement the DBF scheme described above, we first downsample the data to a 0.02 s sampling and band-pass filter every available correlogram between 2 and 8 Hz. We then loop through every virtual source and construct source arrays by taking their 29 nearest neighbors. Then, for each source array, we loop through every other station that is not part of the source group and take their 29 nearest neighbors to construct the receiver arrays (Movie S1). Each source and receiver array is, therefore, composed by 30 stations each, which results in a total of 900 Green’s functions for each subarray pair. The number of stations that are involved in the construction of the subarrays is determined empirically based on the tradeoff that, to better capture the small-scale velocity variations within the survey, we want to use the least amount of instruments in the DBF and still be able to observe clear body wave arrivals (Figure S4). As a quality control, we only apply the DBF to source and receiver groups in which both array radii were below 900 m and their geographic centers are at least 1 km apart. The first condition is set to ensure waveform strong waveform coherency between the traces while the second condition is set to allow some time separation between
Figure 5. Examples of traveltime picking of individual traces. (A) Map of the Long Beach array configuration. The red lines mark the shortest path between the virtual source (station S) and the receivers (stations R1 and R2) involved in the computation of the correlograms shown in (B) and (C). (B) Cross-correlated waveforms between stations S and R1 together with its time-frequency spectra computed via the $S$ transform. The dashed red line on top of the waveforms marks the predicted arrival time of the $P$ wave. The star markers in the lower panel depict all of the energy peaks identified in the time-frequency spectra and the red box delimits the time window used to identify the potential $P$ wave arrivals. The energy peaks in the spectrogram are determined using a watershed algorithm (Meyer, 1994). (C) Same as (B) but for the correlogram between stations S and R2.

the body and surface wave arrivals. Moreover, to assure that the reconstructed beams are statistically significant, we apply a bootstrapping method to calculate 95% confidence limits, and choose to only retain beams in which their confidence region is smaller than 0.3 km$^2$/s$^2$ using a total of 500 picked traveltime resamples. The size of the threshold for the confidence region is determined based on a visual inspection of the quality of the beams and the overall bootstrapping results. Figure 6 shows an example of the modified double-beamforming scheme for a single source-receiver array configuration.

A unique aspect of our processing scheme is that the size and geometry of the source and receiver arrays changes every time a beam is constructed at different locations across the survey. This particularity causes that the resolution of the DBF, which largely depends on the size of the subarrays, to vary spatially such that its ability to extract wavelets of specific frequencies is not always the same. To prevent spatial aliasing, the general rule in beamforming is that the aperture of the array must be at least $>1$ wavelength than the longest period of interest and the average interstation spacing must be $<0.5$ wavelengths (Brenguier et al., 2015; Nakata et al., 2016; Roux et al., 2008). For our case, the Long Beach survey has an average interstation distance of 100 m and, as stated above, all of our source and receiver arrays are composed of 30 stations each. This geometry causes the average aperture of most subarrays to be at around 800 m (Figure S6). Now, assuming that the slowest body wave propagates as a perfectly plane wave with an apparent velocity of 1.5 km/s, the minimum size of the subarrays capable of extracting $>3$ Hz seismic arrivals is $\sim$450 m. This result allows us to conclude that our choice subarray geometries is optimal for extracting the body wave arrivals that are buried within the Long Beach correlograms and that such configuration would technically allow us to beamform up to 7.5 Hz waves without being spatially aliased.

The application of the modified DBF to the entire Long Beach data set yielded approximately 12 million unique beams that now show enhanced diving $P$ waves. As an example, Figure 7 presents snapshots of the constructed wavefield before and after the application of DBF. A clear body wave propagating away from the virtual source in all directions can be seen after the waveforms have been coherently stacked. It is interesting to notice how the wavefield is not completely spherical, which indicates the existence of lateral
Figure 6. Double-beamforming process for a single source-receiver array configuration. (A) Map showing the entire Long Beach survey together with a source array (red circles) and a receiver array (gray circles). The red radius around the source array marks the 1-km distance threshold imposed in our stacking scheme. For this particular source-receiver configuration, the source radius is of 0.32 km, the receiver radius is of 0.42 km, and the distance between the arrays’ geographic centers is of 2.12 km. (B) Entire source and receiver side velocity space of the DBF. The yellow star indicates the best-fitting velocity pair obtained in the least-squares inversion. The black dots correspond to the best-fitting velocity pairs obtained in the bootstrapping process and the red box marks their 95% confidence limits. For this case, the area of the 95% confidence region is 0.09 km²/s². (C) Empirically derived probability density function of the source-side velocity (SSV) and the receiver-side velocity estimations (RSV) calculated from the bootstrapped results. (D) Beamed trace (black waveforms) plotted behind one of the 900 original correlograms used to build the beam (red waveforms). Note how the P wave is essentially unidentifiable in the original correlogram and how its signal-to-noise ratio improves dramatically after applying DBF. The retrieved phase corresponds to a diving P wave propagating between the center of the source and receiver arrays. The collection of all the individual traveltime picks involved in the computation the source-side and receiver-side velocities for this particular group of stations is shown in Figure S5.

variations in the elastic properties of the medium. To further illustrate this point, we build two orthogonal refraction profiles by stacking all available beams along two corridors that are 0.5 km wide in a 50-m offset bin (Figure 8). Aside from showing clear diving P waves regardless the direction of propagation, both profiles show a prominent step in the traveltime that is typical of a horizontal discontinuity in a buried layer. For the case of the N-S profile, the velocity jump is coincident with the surface trace of the Newport-Inglewood fault. This observation supports the premise that we are observing real features of the data and that the constraints imposed in the DBF are flexible enough to capture the variations in the structural properties of the Long Beach crust.

4. Traveltime Measurements and Body Wave Tomography

After we successfully apply the DBF to every possible source-receiver configuration, we pick the arrival times of the beamed P waves. For this matter, we first construct a set of reference traces by stacking all available beams in a 50-m offset bin (Figure 9A). We then manually identify the diving P wave in every reference trace and extract their waveform shape so that they can be used as templates to automatically find their time of arrival in individual beams with similar offsets. To extract the waveform shape of these arrivals, we use the seismogram decomposition method of Juarez and Jordan (2018). This technique takes the time-frequency spectra of a given record computed by the S transform and systematically isolates all apparent arrivals using
Figure 7. Wavefield emitted by a virtual source. The location of the virtual source is marked by the yellow star. The lag-times are shown on the upper left corner of each panel. The top panels show snapshots of the 2–8 Hz band-passed wavefield observed at each station. The middle panels show snapshots of the beamed wavefield observed at each station (Movie S2). Strong body waves are now visible propagating away from the virtual source. The lower panels show snapshot of a synthetic wavefield using the inverted velocity model (black profile in Figure 4B). Note how the observed wavefield is not completely spherical as in the synthetic case.
Figure 8. Examples of $P$ wave refractions from a N-S line (A) and a E-W line (B). The profiles are formed by stacking all available beams along corridors that are 0.5 km wide (green areas in C) in a 50-m offset bin. The location of the virtual sources are marked by the yellow stars. Note how there exist a clear jump in the N-S profile that is coincident with the Newport-Inglewood fault (thick red line in all panels), and shorter jump in the far end of the E-W profile that might be associated with a smaller fault (thick blue line in Panel B).

localized Gaussian filters around every local maxima in the power spectrum. It then integrates every filtered spectra with time and separately transforms them back into the time domain, thus yielding a finite set of waveforms that are localized in time and frequency. As an example, Figures 9B–9D present the waveform decomposition process for a single reference trace and shows how one can recover the input trace after summing all of its decomposed elements. It is important to note that, for this particular reference trace, the temporal resolution of the $S$ transform is sufficient to isolate the diving $P$ wave from the $PP$ and direct $P$ waves. Although the latter may not be the case for shorter off-set traces, the waveform decomposition scheme ultimately allows us to retrieve the characteristic shape of the first-arrival whether they consist of a single diving $P$ wave or a superposition of different $P$ wavelets.

Once we build a library of $P$ wave templates, we apply the same phase decomposition scheme to all individual beams and, for each one of them, retain the timing of the isolated phase that is within the expected time of arrival of the diving $P$ wave and best correlates with its respective waveform template. To show the performance of our picking method, Figure 10 presents the distribution of the traveltime differences between the picked time and the theoretical arrival (from the inverted 1D model) for two virtual sources located in opposite sides of the survey. The spatial coherency of the traveltime anomaly maps, together with the general high correlation coefficients (CCs) obtained in the picking process, subjectively attest to the reliability of our measurements and suggest that their smooth fluctuations are likely to be associated with variations in the local velocity structure. Moreover, there is a clear north-south velocity dichotomy with the transition just on the surface trace of the Newport-Inglewood fault that is revealed by the northern virtual source. It
Figure 9. Stacked record section of beams (A) and waveform decomposition process of a single reference trace (B–D). (A) Stacked record section showing how the DBF has successfully isolated the body wave energy and suppressed any other arrival that is not arriving with a slowness of a refracted P wave. The dashed colored lines mark the time window where the higher mode surface waves were present before the DBF. (B) Reference trace (black waveform) and reconstructed trace (red waveform) that is obtained after linearly adding all of the decomposed waveforms (all traces in D). The reference trace corresponds to the stacked beam in the 7-km offset bin (red arrow in A). (C) Time-frequency spectra of the reference trace computed via the S transform. The star markers indicate the arrivals identified in the power spectrum and the numbers indicate their order in time. (D) Decomposed traces sorted by time. A total of 27 apparent phases are identified in the reference trace spectra, with the diving P wave being the most prominent amongst all. The green trace alone is what is used as a template to pick the refracted P wave arrival in individual beams that have a 7-km offset.

is also worth noticing that stations in which a low correlation coefficient was obtained are in agreement with beams that have a low signal-to-noise ratio (SNR) and regions where traveltime measurements have sudden and unrealistic deviations from their neighboring stations. The majority of these random measurements appear to be correspondent to beams that are constructed from either large-offset correlations or from stations that are located in the vicinity of the fault. Furthermore, a general analysis of the CC and SNR distribution of different virtual sources reveals that higher-quality beams are mostly constructed from subarrays that are located on similar sides of the fault (Figure S7). This observation suggests that there must exist some zone of structural complexity along the fault that may be acting as a barrier to wave propagation.

Because of the extensively large number of traveltime measurements that are available to estimate the 3D velocities, solving the structural problem using a regularized inversion, which involves calculating and storing the data sensitivity (Fréchet) kernel for each measurement, is an intense computational task. For this reason, we opt to implement an iterative backprojection method (Hole, 1992) that maps, on a ray-by-ray basis, traveltime anomalies into slowness perturbations along the ray paths until the data are satisfied. To parameterize the model, we use a regularly spaced 70-m grid that extends 8 km (easting) × 12 km (northing) × 2.5 km (depth) and the inverted 1D velocity profile as a starting model. We then trace every ray through the model using the fast sweeping method (Zhao, 2005) and average the perturbations applied to each model element from all the rays that are influenced by that model element (Figure S8). Once the entire model is updated, we retrace the rays, compute new theoretical traveltimes, and remap the new measurements into slowness perturbations. This process is iterated until the updates to the model start to become negligible. It is important to note that the DBF that was applied to the traces will cause the resulting model to be inherently smooth. Figure 11 shows the horizontal and vertical slices of the obtained velocity model after five iterations of the backprojection method. To ensure robustness in the construction of this model, we only used traveltimes measurements that had a correlation coefficient larger than 0.5 with their respective waveform templates and a signal-to-noise ratio larger than 2. This restriction allowed the use of 11,033,733 rays which, based on the ray coverage maps, we find that can resolve velocities down to a depth of 2 km.
Figure 10. Traveltime measurements for two virtual sources localized on opposite sides of the survey. The left panels show the distribution of the traveltime differences between the picked time and the theoretical arrival. The middle panels show the normalized correlation coefficient between each picked phase with their respective waveform template. The right panels show the signal-to-noise ratio of the beams. Here, we define the SNR by the ratio between the peak amplitude within a window containing the refracted $P$ wave arrival to the root-mean-square of the amplitudes before this window. The locations of the virtual sources are marked by the yellow stars. The red lines depict the major strands of the Newport-Inglewood fault and the black dashed lines delimit the regions where traveltime measurements appear to be random.

To assess the uncertainty of our estimated velocities, we first generate a set of one-iteration tomograms by bootstrapping through the virtual sources. We then follow a similar approach to the one that is used in traditional Eikonal tomography and, for each point in our model, estimate a mean slowness, $s_0$, and a standard deviation, $\sigma_{s_0}$, from the distribution of slownesses, $s_i$:

$$s_0 = \frac{1}{n} \sum_{i=1}^{n} s_i,$$

(4)

$$\sigma_{s_0}^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (s_i - s_0)^2,$$

(5)

at the center of the survey (Figure S9). A collection of traveltime anomaly maps filtered with the imposed quality criteria is shown in Figure S10 of the supporting information.
where \( n \) is the number of virtual sources (Lin et al., 2009). Once these quantities are obtained, we simply compute the uncertainty of the velocity, \( \sigma_v \), via

\[
\sigma_v = \frac{1}{\sigma_0^2} \sigma_s.
\]  

After resampling the virtual sources 50 times and applying such process, we find that the uncertainties of our model are below 5 m/s and that they tend to be largest at the edges of the array (Figure 12). Note that these uncertainties are related to random within our traveltime measurements and not the tomographic inversion itself.

5. Results and Discussion

The tomographic model shows velocity variations that are, in general, consistent with the surface wave results of Lin et al. (2013). As expected, the most prominent feature within our model is the Newport-Inglewood fault zone, which shows up as a high velocity anomaly (probably as a result of strain...
Figure 12. Depth slices of the velocity uncertainties obtained after bootstrapping the virtual sources 50 times. The depth of the horizontal slices is given at the upper right corner of each panel (0.14, 0.49, and 0.98 km).

focusing; e.g., Papaleo et al., 2017) that emerges at depths of around 500 m. Although the average velocity structure of this feature has been imaged in previous tomography studies (e.g., Chang et al., 2016; Lin et al., 2013), the resolution that is now achieved with high-frequency body waves allow us to illuminate some of its geometric variations. Figure 11 shows a comparison between different cross-sections of our velocity model cut perpendicular to the main fault trend. From this comparison, we find that there are structural differences amongst the three imaged segments of the fault. For the northwestern strand, the fault trace appears to be mostly vertical and well defined at shallow depths (Figures 13C and 13D). In contrast, cross-sections across the southeastern strand suggest that the fault extends to deeper depths and might even have a small SW dipping component that grows as it approaches Signal Hill (Figures 13F and 13G; see Figure S11 for elevation map). These observations are consistent with reconstructed maps of the internal structures of Los Angeles basin that suggest that the primary strand of the Newport-Inglewood fault (also known as the Cherry Hill fault) extends nearly vertical down to depths of 1,100–1,500 m and, at greater depths, may dip as much as 60° (Wright, 1991). Near the central strand, we find that the fault structure is even more complicated and unidentifiable within our velocity model (Figure 13E). This complexity may arise since, at this location, the Cherry Hill fault is known to branch out near the surface through a series smaller-scale fault strands that bound a recently uplifted wedge (Figures 13B–13E).

Although the general agreement between our tomographic model with known geologic features attest to the reliability of ambient noise body waves to characterize subsurface velocity structure, we can take advantage of the large station density of the survey to qualitatively validate our results in a different way. Figure 14 shows the effects of velocity variations on a Mb 2.5 earthquake wavefront propagating through the Long Beach array. From this image, it is clear how the structural variations within the area introduce complexities into the wavefront and can even cause a 180° phase change in the S wave across the Newport-Inglewood fault (Movie S2). The partitioning of the wavefront is most extreme in the southeastern strand of the fault, precisely where our velocity model reveals the existence of a prominent and elongated fast velocity anomaly (associated to the fault itself) adjacent to a slow velocity zone that extends to a depth of about 1 km (Figure 11A). Moreover, it is important to note that the fast velocities of the fault also causes a slight increase in the wavelength of the surface wave packet, which may have some implications when quantifying the frequency response of the basin at a local scale. The fact that our velocity model captures the structural features that introduce these complexities into the wavefield supports the reliability of our estimations, and highlights the relevance of characterizing the fine-scale velocity variations of the shallow crust.
Figure 13. Fault-perpendicular cross-sections of the inverted velocity model. (A) Map showing the Long Beach survey with the location of the profiles. (B) Cross-sections illustrating the general structural pattern across different segments of the Newport-Inglewood fault zone (modified from Wright, 1991). The thick black lines in A show the location of the two profiles (X-X' and Y-Y'). (C–G) Vertical cross-sections of the inverted velocity model (colored profiles in A). The red arrows on top of each profile indicate the location of the surface trace of the Newport-Inglewood fault. The black dashed lines in each profile mark our interpretation of the fault. The inferred fault geometry of Wright (1991) across profile X-X' is overlain on profile E.

Figure 14. Wavefield emitted by the Carson Mb2.5 earthquake at 4 seconds after its origin time (A). Note how the earthquake surface waves clearly break up along the Newport-Inglewood fault. The dashed black box in the zoomed panel is the same as in Figure 11A and delimits a region where the wavefield is being most advanced and stretched by the high velocities of the fault. The velocity perturbation at 0.98 km depth of this small area is shown in (B).
The availability of high-resolution velocity models is essential in many geophysical applications and paves the way for more complete interpretation of the Earth’s geologic structures. For this study case, the results are mostly relevant to seismic hazard studies of Los Angeles basin since they can be incorporated into the standard velocity models for the region and, hence, be used in earthquake ground motion predictions. In addition to its role to seismic hazard, our velocity model can be used in an exploration context to derive a set of static corrections that can be applied to seismic reflection analysis and used to improve active-source processing. Static corrections as such have been derived from the surface wave velocity model of Lin et al. (2013) and found to be effective in enhancing P wave reflective signals (Holllis et al., 2018). Nonetheless, the statics derived from our P wave velocity model should be even more accurate since they span a larger depth range and are not converted from phase and shear wave velocities.

6. Conclusions

We used a high-density oil company survey in Long Beach, California, to demonstrate that body waves can be extracted from the ambient noise field and subsequently inverted to produce a high-resolution velocity model of the subsurface. First, we confirmed the existence of coherent body wave energy in the noise correlograms by computing a synthetic wavefield and predicting every prominent arrival that is present in an average record section of the entire data set. We then retrieved refracted P waves propagating through different sections of the survey by dividing the Long Beach network into small-aperture subarrays and applying a double-beamforming (DBF) technique that simultaneously solves for the optimal source-side and receiver-side stacking velocities. Profiles of P wave refractions along particular corridors of the survey after applying DBF revealed that there exist prominent velocity jumps that are likely associated with regional geological features such as the Newport-Inglewood fault. After extracting clean refracted P waves, we measured their absolute traveltime by generating waveform templates and using them to identify their time-of-arrival in individual beams. Once the traveltime measurements were made, we applied an iterative backprojection tomography to map the traveltime measurements into velocity perturbations and solve for the 3D velocity structure beneath the survey.

The estimated 3D velocities are in general agreement with previous tomographic results of the area and correlate with the main structural features of the region. A close analysis of the Newport-Inglewood fault system reveals that there are structural differences amongst the three imaged segments of the fault. The northwestern strand of the fault seems to extend to depths of around 500 m and appears as a well-imaged vertical velocity discontinuity. On the other hand, the southeastern strand of the fault extends to deeper depths (>1 km) and appears to have a subtle southwest dipping component that grows as the fault approaches Signal Hill. The central strand of the fault is not apparent within our velocity model, which we believe is caused by the structural complexities of the fault zone. In short, the results of this work confirmed the reliability of ambient noise body waves to characterize the seismic structure and showed that, in spite of the processing challenges of retrieving them, they can provide unique constraints that are otherwise unattainable with current passive imaging methods.

Acknowledgments

The data used in this study is the property of Signal Hill Petroleum Inc., and permission from them is required to access it. We gratefully acknowledge Dunzhu Li for providing the ambient noise cross-correlations. We gratefully thank Nori Nakata and an anonymous reviewer for their careful and constructive suggestions. The figures presented in this paper were made using the Generic Mapping Tools v.4.5.9 (https://soest.hawaii.edu/gmt). The final P wave velocity model can be downloaded from: https://doi.org/10.22002/D1.1293. The traveltime measurements used to generate the velocity model can also be downloaded from that site.

References

Aki, K., & Richards, P. G. (2002). Quantitative seismology. San Francisco: W.H. Freeman and Company.
Bensen, G., Ritzwoller, M., Barmen, M., Levshin, A., Lin, F., Moschetti, M., et al. (2007). Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, 169(3), 1239–1260.
Boué, P., Roux, P., Campillo, M., & de Caquerey, B. (2013). Double beamforming processing in a seismic prospecting context. *Geophysics*, 78(3), V101–V108.
Brenguier, F., Kowalski, P., Ackerley, N., Nakata, N., Boué, P., Campillo, M., et al., Roux, P., Ferrazzini, V., Villemuneve, N., Shapiro, N. M., & Chaput, J. (2015). Toward 4D noise-based seismic probing of volcanoes: Perspectives from a large-n experiment on piton de la fournaise volcano. *Seismological Research Letters*, 87(1), 15–25.
Campillo, M., & Paul, A. (2003). Long-range correlations in the diffuse seismic coda. *Science*, 299(5606), 547–549.
Chang, J. P., de Ridder, S. A., & Blondi, B. L. (2016). High-frequency Rayleigh-wave tomography using traffic noise from Long Beach, California. *Geophysics*, 81(2), B43–B53.
Clayton, R. W., Heaton, T., Chandy, M., Krause, A., Kohler, M., Bunn, J., et al. (2012). Community seismic network. *Annals of Geophysics*, 54(6), 738–747.
Hole, J. (1992). Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research*, 97(B5), 6553–6562.
Holllis, D., McBride, J., Good, D., Arndt, N., Brenguier, F., & Olivier, G. (2018). Use of ambient noise surface wave tomography in mineral resource exploration and evaluation. *SEG technical program expanded abstracts* 2018 (pp. 1937–1940). Anaheim, California: Society of Exploration Geophysicists.
Inbal, A., Ampuero, J. P., & Clayton, R. W. (2016). Localized seismic deformation in the upper mantle revealed by dense seismic arrays. *Science*, 354(6308), 88–92.

Jennings, C., & George, J. (1994). Fault activity map of california and adjacent areas, with locations and ages of recent volcanic eruptions, scale 1: 750 000. *Divisions of Mines and Geology, Geologic Data Map*, 6.

Juarez, A., & Jordan, T. (2018). Optimization of data functionals for full-3D tomography. In *American Geophysical Union, Fall Meeting 2018, abstract #S51A-07*, Washington, D.C.

Li, D., Helmberger, D., Clayton, R. W., & Sun, D. (2014). Global synthetic seismograms using a 2-D finite-difference method. *Geophysical Journal International*, 197(2), 1166–1183. https://doi.org/10.1093/gji/ggu050

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClenahan, J. (2018). High-resolution seismic event detection using local similarity for large-n arrays. *Scientific Reports*, 8(1), 1646.

Lin, F.-C., Li, D., Clayton, R. W., & Hollis, D. (2013). 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., Moschetti, M. P., & Ritzwoller, M. H. (2008). Surface wave tomography of the Western United States from ambient seismic noise: Rayleigh and love wave phase velocity maps. *Geophysical Journal International*, 178(1), 281–298.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R. (2009). Eikonal tomography: Surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*, 177(3), 1091–1110.

Lokhis, O. I., & Weaver, R. L. (2001). On the emergence of the Green’s function in the correlations of a diffuse field. *The Journal of the Acoustical Society of America*, 110(6), 3011–3017.

Mansinha, L., Stockwell, R., & Lowe, R. (1997). Pattern analysis with two-dimensional spectral localisation: Applications of two-dimensional s transforms. *Physica A: Statistical Mechanics and its Applications*, 239(1-3), 286–295.

Meyer, F. (1994). Topographic distance and watershed lines. *Signal Processing*, 38(1), 113–125.

Nakata, N., Boué, P., Brenguier, F., Roux, P., Ferrazzini, V., & Campillo, M. (2016). Body and surface wave reconstruction from seismic noise correlations between arrays at Piton de la Fournaise volcano. *Geophysical Research Letters*, 43, 1047–1054. https://doi.org/10.1002/2015GL066997

Nakata, N., Chang, J. P., Lawrence, J. F., & Boué, P. (2015). Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry. *Journal of Geophysical Research: Solid Earth*, 120, 1159–1173. https://doi.org/10.1002/2015JB011870

Nakata, N., & Nishida, K. (2019). Body wave exploration. In *Seismic Ambient Noise* (p. 239). Cambridge, England: Cambridge University Press.

Olivier, G., Brenguier, F., Campillo, M., Lynch, R., & Roux, P. (2015). Body-wave reconstruction from ambient seismic noise correlations in an underground mine. *Geophysics*, 80(3), KS11–KS25.

Papaleo, E., Cornwell, D. G., & Rawlinson, N. (2017). Seismic tomography of the North Anatolian fault: New insights into structural heterogeneity along a continental strike-slip fault. *Geophysical Research Letters*, 44, 2186–2193. https://doi.org/10.1002/2017GL072726

Rost, S., & Thomas, C. (2002). Array seismology: Methods and applications. *Reviews of Geophysics*, 40(3), 2–1.

Roux, P., Cornuelle, B. D., Kuperman, W., & Hodgkiss, W. (2008). The structure of raylike arrivals in a shallow-water waveguide. *The Journal of the Acoustical Society of America*, 124(6), 3430–3439.

Roux, P., Sabra, K. G., Gerstoft, P., Kuperman, W., & Fehler, M. C. (2005). P-waves from cross-correlation of seismic noise. *Geophysical Research Letters*, 32, L19303. https://doi.org/10.1029/2005GL023803

Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W., & Fehler, M. C. (2005). Surface wave tomography from microseisms in Southern California. *Geophysical Research Letters*, 32, L14311. https://doi.org/10.1029/2005GL023155

Shapiro, N. M., & Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31, L07614. https://doi.org/10.1029/2004GL019491

Shapiro, N. M., Campillo, M., Strehl, L., & Ritzwoller, M. H. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, 307(5715), 1615–1618.

Shaw, J. H., Plesch, A., Tape, C., Suess, M. P., Jordan, T. H., Ely, G., et al. (2015). Unified structural representation of the Southern California crust and upper mantle. *Earth and Planetary Science Letters*, 415, 1–15.

Shearer, P. M. (2019). *Introduction to seismology*. Cambridge, England: Cambridge University Press.

Snieder, R., & Laroze, E. (2013). Extracting Earth’s elastic wave response from noise measurements. *Annual Review of Earth and Planetary Sciences*, 41, 183–206.

Spica, Z. I., Nakata, N., Liu, X., Campman, X., Tang, Z., & Beroza, G. C. (2018). The ambient seismic field at Groningen gas field: An overview from the surface to reservoir depth. *Seismological Research Letters*, 89(4), 1450–1466.

Stockwell, R. G., Mansinha, L., & Lowe, R. (1996). Localization of the complex spectrum: the S transform. *IEEE Transactions on Signal Processing*, 44(4), 998–1001.

Thorson, J. R., & Claerbout, J. F. (1985). Velocity-stack and slant-stack stochastic inversion. *Geophysics*, 50(12), 2727–2741.

Vilasenor, A., Yang, Y., Ritzwoller, M. H., & Galiart, J. (2007). Ambient noise surface wave tomography of the Iberian Peninsula: Implications for shallow seismic structure. *Geophysical Research Letters*, 34, L11304. https://doi.org/10.1029/2007GL030164

Weber, M., Davis, J., Thomas, C., Krüger, F, Scherbaum, F., Schittenhardt, J., & Könnig, M (1996). The structure of the lowermost mantle as determined from using seismic arrays. In E. Boschi, G. Ekström, A. Morelli, (Eds.), *Seismic Modeling of the Earth’s Structure* (pp. 399–442). Rome: Istil. Naz. Di Geophys.

Wright, T. L. (1991). Structural geology and tectonic evolution of the Los Angeles basin, California: Chapter 3: Part 1.

Yang, Y., Ritzwoller, M. H., Levshin, A. L., & Shapiro, N. M. (2007). Ambient noise Rayleigh wave tomography across europe. *Geophysical Journal International*, 168, 259–274. https://doi.org/10.1111/j.1365-246X.2006.03203.x

Zhao, H. (2005). A fast sweeping method for Eikonal equations. *Mathematics of Computation*, 74(250), 603–627.