Detection and removal of coherent anthropogenic noise from passive seismic data

Ethan Williams and Eileen Martin

ABSTRACT

We analyze the impact of identifying and removing coherent anthropogenic noise on synthetic Green's functions extracted from ambient noise recorded on a shallow trenched, dense, linear distributed acoustic sensing (DAS) array. Low-cost, low-impact urban seismic surveys are possible with ambient noise recorded by DAS, which uses dynamic strain sensing to detect seismic waves incident to a buried fiber optic cable. However, ambient noise data recorded in urban areas include coherent, time-correlated noise from near-field infrastructure such as cars and trains passing the array, in some cases causing artifacts in estimated Green's functions and yielding potentially incorrect surface wave velocities. Based on our comparison of several methods, we propose an automated, real-time data processing workflow to detect and reduce the impact of these events on data from a dense array in an urban environment. We show the effect of removing such unwanted noise on estimated Green's functions from ambient noise data recorded in Richmond, CA in December 2014 and Fairbanks, AK in August 2015.

INTRODUCTION

Urban seismic surveys are an essential tool for many areas of geoscience and civil engineering, including the design of earthquake-resistant structures and the quantification of seismic hazard in cities. However, conventional seismic surveys are all but impossible in the urban environment because of the high impact of active sources and the difficulty of deploying a sizable receiver array. Recent experiments have successfully shown the applicability of ambient noise interferometry on recordings made with distributed acoustic sensing (DAS) as an alternative to conventional surveys that is both low-cost and low-impact (Martin et al., 2015; Daley et al., 2013; Ajo-Franklin et al., 2015). In a DAS survey, an interrogator unit regularly transmits a coherent laser pulse along a fiber optic cable network and records the Rayleigh backscatter intensity, which is converted into an approximate seismogram. The ambient noise recorded using DAS can then be correlated to extract synthetic Green's functions through seismic interferometry, as has been successfully done by Martin et al. (2015) and Ajo-Franklin et al. (2015).

One important barrier to the wide application of DAS surveys in urban areas is the presence of coherent transportation-related noise from sources such as cars or trucks

passing near the array. During correlation, these events correlate with themselves to produce artifacts in Green's functions (Martin et al., 2015). In order to improve Green's function estimates in these environments, there have been efforts to identify and automatically down-weight high amplitude noise (Lindsey, 2016). Here, we build on these methods more specifically to identify strong moving sources of noise related to transportation. First, we propose a real-time processing workflow that develops a catalog of times where strong transportation-related noise has been recorded. Our method utilizes the short-term average over long-term average (STA/LTA) algorithm from the field of earthquake detection and takes advantage of the density of the data recordings to minimize the number of false triggers. Second, we demonstrate a method for down-weighting these events during interferometry that improves the strength and coherency of Green's functions estimates on the test data. The methods discussed in this paper have been developed for use with real-time traffic monitoring and the removal of near-array transportation-related noise in the DAS ambient noise survey on Stanford campus that began in September 2016.

CHARACTERIZING NOISE SOURCES

Several sources of noise contribute to any ambient noise recording. Generally, incoherent noise from microseism dominates at low frequencies and incoherent anthropogenic noise from freeways and other infrastructure dominates at high frequencies. Additionally, recordings often include signals from earthquakes and coherent noise from infrastructure near the array, such as passing cars.

Seismic interferometry posits that the correlation of a diffuse wavefield in a medium will return the impulse response of that medium (Wapenaar et al., 2010). Applied to the Earth, this means that the cross-correlation of ambient noise recorded on different channels in an array with that recorded on a single channel produces an approximate Green's function, which represents the response of the Earth to a receiver acting as a virtual source. Greens functions constructed from ambient noise can be used to monitor changing properties in the subsurface or for tomographic investigations. The presence of coherent, time-correlated noise in an ambient noise recording violates the diffuse wavefield assumption and results in artifacts in the estimated source response that resemble multiple virtual sources.

Ambient noise recorded using DAS in Richmond, CA in 2014 and in Fairbanks, AK in 2015 contained visually identifiable coherent noise from passing traffic (Ajo-Franklin et al., 2015; Martin et al., 2016). Incoherent noise from urban infrastructure far from the recording array is the primary noise source used for high frequency ambient noise interferometry because it is a diffuse component of the ambient spectrum (e.g. Nakata et al. (2011)), but coherent noise from traffic sources near the recording array is the primary cause of artifacts. This is because traffic passing the array is in two ways time-correlated. First, surface waves generated by a passing vehicle have visually identifiable group moveout across the array at traffic speed, giving the wave package associated with any given car a temporally-correlated velocity relationship.

Second, surface waves generated by the vehicle do not originate at the source channel but rather at road bumps and other features that lie along the array, causing additional virtual sources within the response estimate. An example recording of a vehicle passing the Fairbanks array is shown in Figure 1.

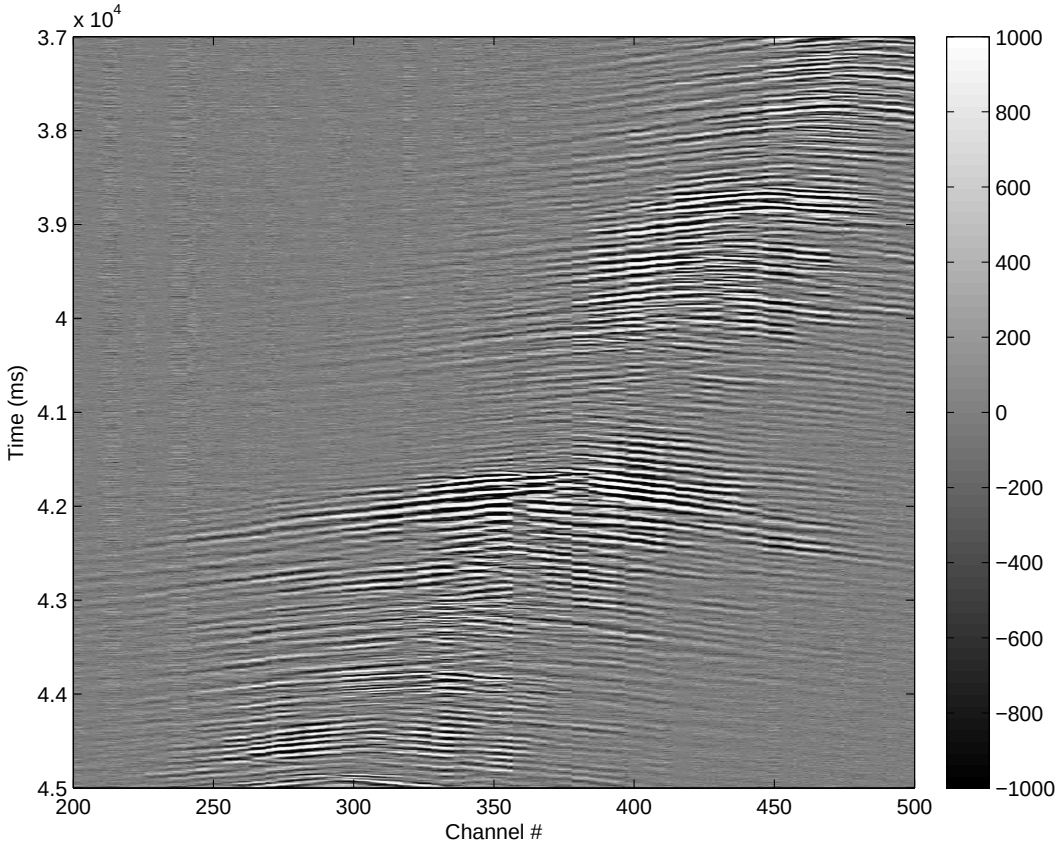


Figure 1: Example recording showing a car driving past a linear segment of the trenched DAS array in Fairbanks, AK. Channel spacing is 1 m. [**CR**]

The geometry of relevant roads and their distance from the array determine the properties of the coherent transportation-related noise recorded. In Fairbanks, the closest cars travelled on a road very near and parallel to the array, so that their traffic speed was easily identifiable as the group move out of the ground roll at approximately 25 m/s. In Richmond, the closest cars travelled on a road slightly farther from and oblique to the array, so that the traffic speed was difficult to discern but the surface waves were still highly coherent. The following methods were designed on both data sets, and we present suggestions for modifying proposed processing workflow to best suit recordings with different geometries and noise intensities.

COMPARISON OF METHODS

In the earthquake detection literature, three principal methods for event detection exist: (1) autocorrelation, (2) waveform comparison methods, and (3) short-term average over long-term average (STA/LTA) (Withers et al., 1998).

One prominent feature of data recorded with DAS is optical noise spikes at random intervals over several channels due to vibration of the laser interrogator unit. The autocorrelation method for event detection entails partitioning each channel's record into short-time windows and cross-correlating all possible combinations of these windows. For data recorded with DAS, optical noise correlates with itself, creating artificially high correlation coefficients and making transportation-related events difficult to distinguish. Even more importantly, this method can only operate on a single channel, so performing this method across an array of several thousand channels proves to be too slow to run in real time with ordinary computational resources.

While waveform comparison methods have been shown to be the quickest and most effective method of identifying earthquakes (Yoon et al., 2015), waveform comparison tests are impossible to implement for data recorded with DAS because there are no catalogs of transportation-related event recordings to build a database. Because of DAS's low signal-to-noise ratio compared to conventional seismic recording instruments and because the traces themselves are only derived indirectly from strain rate approximations, ambient noise recorded with DAS cannot be compared with conventional data.

The STA/LTA method effectively detects impulsive events with high signal-to-noise ratio by comparing the squared amplitudes of data summed over two moving windows of differing lengths. Equation 1 takes two parameters—STA window length (ℓ_{sta}) and LTA window length (ℓ_{lta})—which need to be set manually based on the signal-to-noise ratio and other parameters of the data. The long-term window finds the average amplitude of the data for some number of samples ℓ_{lta} before sample i, which is representative of the background, and the short-term window finds the average amplitude of the data for a shorter number of samples ℓ_{sta} before sample i, which is representative of the signal at i. When amplitude increases significantly at some iteration of sample i, the STA/LTA ratio increases. Thus, in conventional earthquake detection, STA/LTA is a trigger-based algorithm, meaning that the user pre-sets a threshold STA/LTA value and when the STA/LTA ratio calculated on a seismogram exceeds this value, an event is cataloged. The STA/LTA method is advantageous for real-time event detection simultaneously across channels on a large recording array because it can be easily vectorized across channels and it is often implemented using a recursive algorithm, making it the cheapest detection method by far.

$$\frac{STA}{LTA}(i) = \frac{\sum_{j=i-\ell_{sta}}^{i} |u(j)|^2}{\sum_{j=i-\ell_{lta}}^{i} |u(j)|^2} \tag{1}$$

PROCESSING WORKFLOW

STA/LTA

We first import a single 60 s file and apply basic pre-processing steps, including a de-spiking tool to reduce the number of optical noise spikes. Because our data were recorded at $f_s = 1$ kHz, we lowpass filter the data and downsample by a factor of 10, which improves computational speed without reducing the quality of the STA/LTA image. We then calculate STA/LTA on all channels of the file for the complete record length using a recursive algorithm (Equation 2) that approximates the actual STA/LTA ratio very closely.

$$\begin{aligned} STA(i) &= \frac{|u(j)|^2}{\ell_{sta}} + (1 - \frac{1}{\ell_{sta}}) \times STA(i-1) \\ LTA(i) &= \frac{|u(j)|^2}{\ell_{lta}} + (1 - \frac{1}{\ell_{lta}}) \times LTA(i-1) \end{aligned} \tag{2}$$

We found that $\ell_{sta} = 0.5$ s and $\ell_{lta} = 10$ s were effective parameter choices for data recorded in both Richmond and Fairbanks. Smaller ℓ_{sta}/ℓ_{lta} reduces the event-to-background contrast, whereas higher ℓ_{sta}/ℓ_{lta} reduces the sharpness of event arrivals (see Figure 2).

Window check

After calculating STA/LTA on the file and storing these values, we apply two threshold-based tests to determine whether events are present. First, we zero all STA/LTA values below the STA/LTA threshold ($\tau_{sta/lta}$) in order to improve the STA/LTA signal-to-noise ratio (i.e., the contrast of events from background noise). As with ℓ_{sta} and ℓ_{lta}, $\tau_{sta/lta}$ must be chosen by trial and error on a short data sample (5-10 minutes). Specifically, a threshold value that is less than the average STA/LTA value of all visually identifiable transportation-related events but greater than the average STA/LTA value of background intervals between events is preferred.

Second, we calculate the square of sums for each time sample on a running time window across all channels. For each file, this yields a single vector containing the short-term energy of the STA/LTA ratio averaged over the all channels at any given

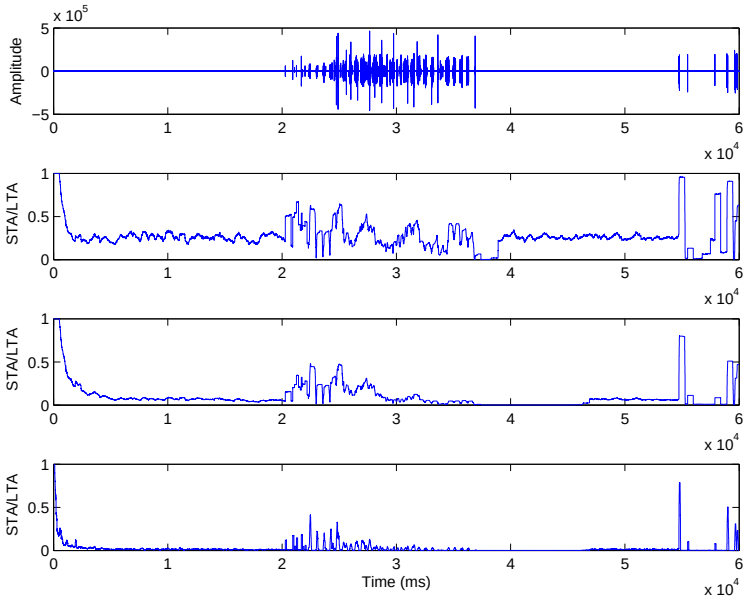
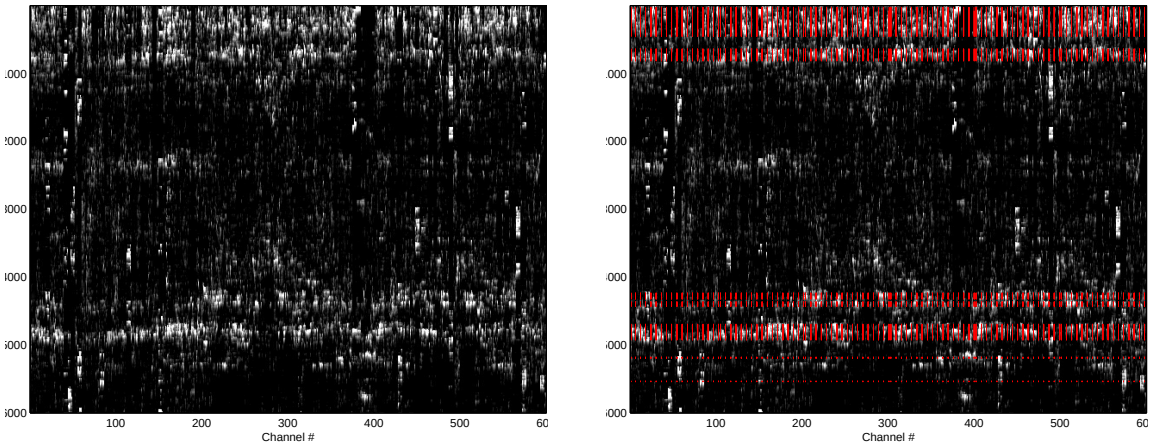


Figure 2: Examples of different ℓ_{sta} and ℓ_{lta} parameter choices. Data must be padded with $\ell_{lta} - 1$ samples to prevent the spike at the beginning, which occurs because the LTA window is only partially filled. From the top, (a) the original trace (raw data), (b) $\ell_{sta} = 0.5$ s, $\ell_{lta} = 2$ s, (c) $\ell_{sta} = 0.5$ s, $\ell_{lta} = 10$ s, (d) $\ell_{sta} = 0.1$ s, $\ell_{lta} = 10$ s. [**CR**]

time, which we denote $\epsilon_{sta/lta}$. Values of $\epsilon_{sta/lta}$ that exceed an energy threshold τ_ϵ are marked as events, producing a binary event catalog of record length with 1 at time samples where a transportation-related event has been detected and 0 otherwise. The energy threshold τ_ϵ must also be set by experimentation on a small dataset. Because this is the most sensitive parameter for distinguishing true events from both background and optical noise, τ_ϵ performs better when designed on as large a sample dataset as possible. In particular, τ_ϵ may need to change with the balance of traffic and background noise intensities during night time and rush hour in order to perform most effectively. For the Fairbanks dataset, we found $\tau_{sta/lta} = 3$ and $\tau_\epsilon = 500$ to be effective threshold values.

On the Richmond dataset, where the geometry of the array relative to traffic results in no discernible group moveout of vehicle-induced surface waves, the short-term energy sum can be applied with a horizontal window that examines all channels at the same time sample, because the phase moveout of events at apparent surface wave velocities is effectively instantaneous relative to the sample rate $f_s = 1 - 2$ kHz (see Figure 3).



(a) (b)

Figure 3: (a) The STA/LTA image of an example 60 s file from the Richmond dataset. (b) Shading shows the region included in the final catalog, superimposed over (a). [**CR**]

The Fairbanks dataset, however, showcases one of the potential challenges of this detection method because the geometry of the array relative to traffic results in a group moveout at low velocity (approx. 25 m/s). This means that a horizontal window will only ever include a portion of the event, and the sensitivity of $\epsilon_{sta/lta}$ will be reduced because channels of background noise will be averaged with any real event. Additionally, a down-weighting filter designed on the resulting event catalog will remove a significant quantity of usable data. To address this challenge, we correct

the STA/LTA matrix for linear moveout (LMO) at the approximate traffic speed before calculating $\epsilon_{sta/lta}$ so that events are parallel to the moving time window and have a stronger signature. This is similar to slant stacking.

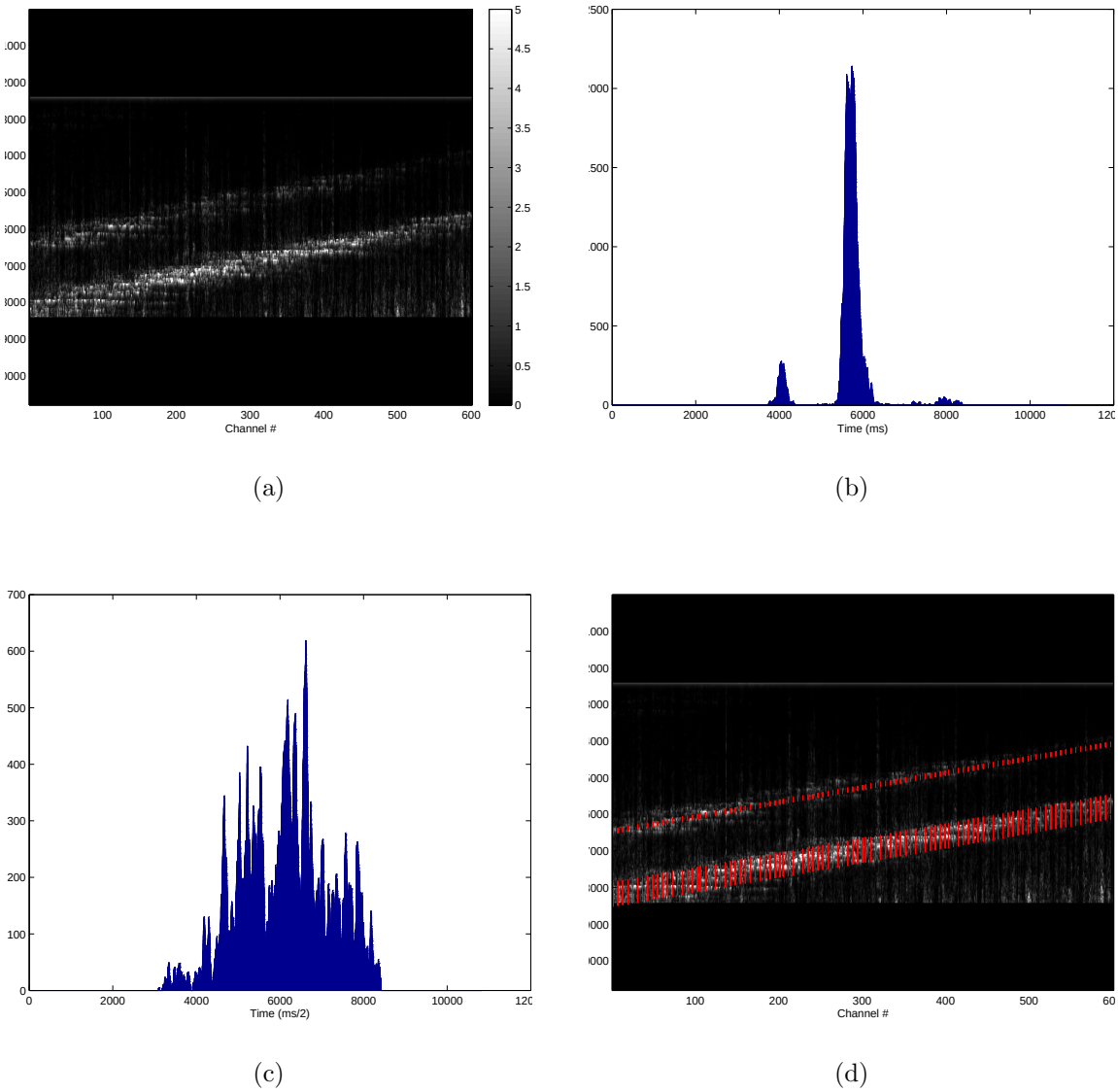
Because traffic can be traveling in two directions on any given road, $\epsilon_{sta/lta}$ must be calculated twice with LMO in either direction. We distinguish the favorable direction (which catalog to use, i.e., the direction in which most cars are traveling in any file) using the properties of the energy distribution with time. For a file containing a single car (a unimodal energy distribution with time), the kurtosis of $\epsilon_{sta/lta}$ will be less for the favorable direction because the window will sum across a flat event as compared to an event with significant move out. For files containing multiple cars (a multimodal energy distribution with time), the skewness of $\epsilon_{sta/lta}$ will be greater for the favorable direction for a similar reason. When an LMO correction has been applied during event detection, a direction variable must be passed with the event catalog in order to design a down-weighting filter with the correct shape. If a file records cars traveling in multiple directions at the same time (e.g., Figure 5), this method selects the direction with the strongest events, and either will catalog the entire interval where events in the opposite direction occur or will not successfully identify these events, depending on their relative strength.

Down-weighting

When calculating cross-correlations on the data, we use a simple down-weighting scheme to remove events using the catalogs which were produced by running the tools outlined above in real time as the data were recorded. First, a data file is imported along with its event catalog and basic processing is done, including de-spiking (Martin et al., 2015), temporal normalization, and spectral whitening (Bensen et al., 2007). Second, a filter is applied to the data which is zero at every time sample corresponding to a 1 in the event catalog and one at every time sample corresponding to a 0 in the event catalog, except at the edge of events, where it follows a Gaussian taper. This taper prevents sharp transitions in the down-weighted file and should ensure that imperfect detection parameters which only identify portions of events are still effective at reducing their contribution to the final correlations. When a LMO correction has been applied during event detection, the direction variable associated with the event catalog is used to extrapolate the filter from one channel to the next. Figure 6 shows an example of a down-weighting filter applied to a file from the Fairbanks dataset.

RESULTS ON THE TEST DATA

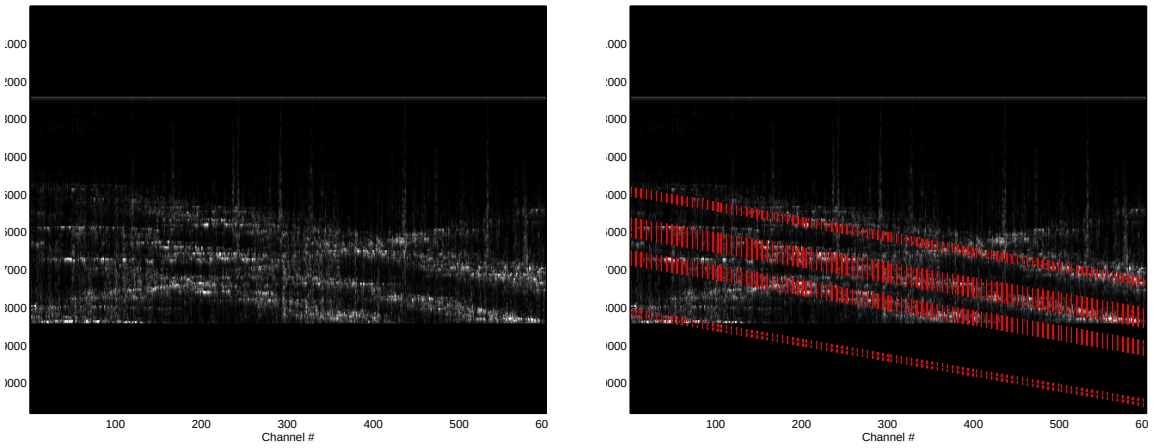
We calculated cross-correlations on 40 minutes of ambient noise from the Fairbanks dataset following the methods of Martin et al. (2015) and Bensen et al. (2007), with the goal of showing the effectiveness of our proposed processing workflow. Lindsey (2016) showed that manual identification and removal of transportation related noise



(a) (b)

(c) (d)

Figure 4: (a) The STA/LTA image of an example 60 s file from the Fairbanks dataset. (b) The $\epsilon_{sta/lta}$ calculated with LMO correction in the correct direction. (c) The $\epsilon_{sta/lta}$ calculated with LMO correction in the incorrect direction. (d) Shading shows the region included in the final catalog superimposed over (a). [**CR**]



(a) (b)

Figure 5: (a) The STA/LTA image of an example 60 s file from the Fairbanks dataset where multiple events cross. (b) Shading shows the region included in the final catalog superimposed over (a). [**CR**]

improved extracted Greens functions significantly. Using a smaller test dataset and less refined pre-processing tools, we were able to show a decrease in the strength of artifacts in cross-correlation images and a small increase in the degree of convergence of the Rayleigh wave fundamental mode associated with the virtual source channel (see Figure 7). More refined workflows, such as in Lindsey (2016), have the potential to improve the image further.

CONCLUSIONS

Overall, the use of an automated processing workflow for identifying and removing coherent anthropogenic noise from ambient noise data recorded by a dense DAS array improved the quality of extracted Greens functions. Our proposed method calculates the STA/LTA ratio on each file, uses a running window check to pick out events, and down-weightes these events in the final correlations. The method is adaptable to different recording geometries, and we present one example of a variation of the method that corrects for apparent group moveout at road speeds.

This processing workflow is most effective when applied to recordings with low to moderate levels of traffic-related noise, because it is unable to recover significant high-quality ambient noise data from recordings capturing more than 3-4 vehicles or where vehicles traveling in two directions frequently cross the array at the same time. The method is able to recover the most usable data from recordings capturing 1-3 vehicles traveling in the same direction or where vehicles traveling in two directions do

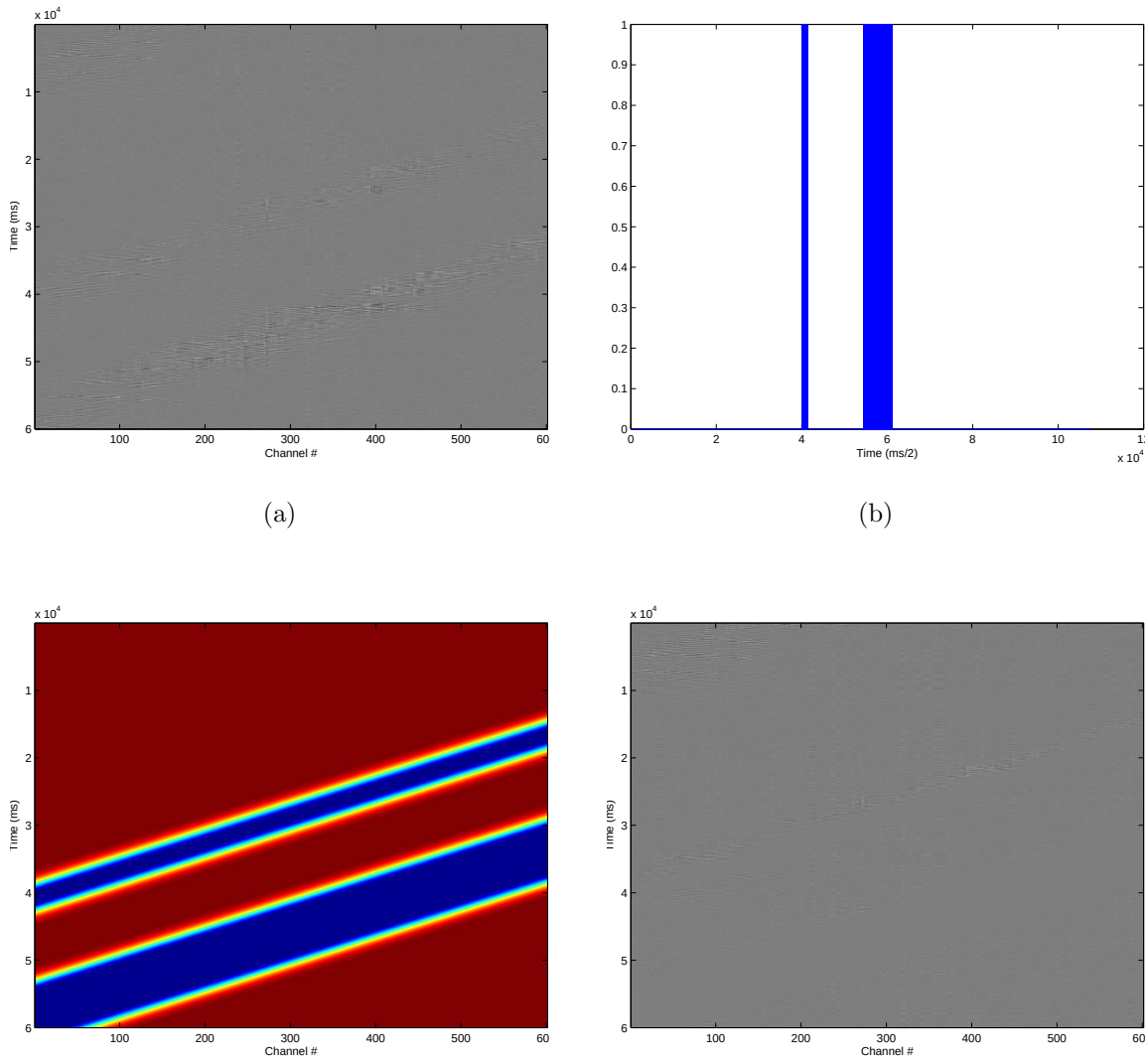
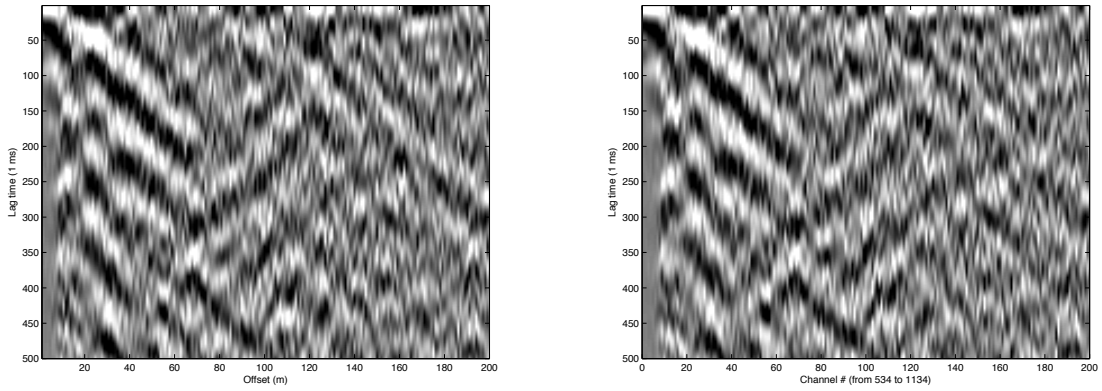


Figure 6: Effect of downweighting on an example 60 s file from Fairbanks. (a) The raw data. (b) The binary catalog determined as above. (c) The downweighting filter designed using the event catalog derived from (b) and the direction of the LMO correction. (d) The data after downweighting. [**CR**]



(a) (b)

Figure 7: Cross-correlations on 40 mins of ambient noise data recorded by the trenched DAS array in Fairbanks, AK. The left image (a) was created using the complete recordings, and the right image (b) was created using recordings where transportation-related events had been down-weighted. [**CR**]

not cross the array at the same time. This extends the amount of daytime recordings that we are able to use in interferometry without generating strong artifacts. By increasing the fraction of recorded data that can be used in interferometry, we are increasing the potential of this method for time-lapse monitoring of the near-surface because reliable Greens functions can be extracted more frequently, providing a higher temporal resolution in monitoring studies.

Future development of this method has the potential to include a more sophisticated filter design process which is able to down-weight cars that cross paths along the array, as well as the consideration of metrics which determine whether an entire file needs to be thrown out without following the entire processing workflow. This method will be applied in a survey currently taking place on Stanford campus in order to test the potential of ambient noise recorded with DAS to monitor traffic and to develop an event catalog that can be used in future processing work.

ACKNOWLEDGMENTS

We would like to thank Biondo Biondi for many helpful discussions about this project. We would also like to thank the SERDP Permafrost Thaw Experiment team for providing the datasets used in this report. The data collection was supported by the US Department of Defense under SERDP grant RC-2437 "Developing Smart Infrastructure for a Changing Arctic Environment Using Distributed Fiber-Optic Sensing Methods." The team, led by PI Jonathan Ajo-Franklin (Lawrence Berkeley Lab) and co-PI Anna Wagner (US Army Corps of Engineers CRREL), includes Kevin Bjella

(CRREL), Tom Daley (LBL), Shan Dou (LBL), Barry Freifeld (LBL), Nate Lindsey (LBL/UC Berkeley), Michelle Robertson (LBL), and Craig Ulrich (LBL). Eileen Martin's work is supported in part by the Department of Energy Computational Science Graduate Fellowship, provided under grant number DE-FG02-97ER25308. Ethan Williams's work is supported by a grant from the Stanford Earth Summer Undergraduate Research (SESUR) Program.

REFERENCES

Ajo-Franklin, J., N. Lindsey, S. Dou, T. M. Daley, B. M. Freifeld, E. R. Martin, M. Robertson, C. Ulrich, and A. Wagner, 2015, A field test of distributed acoustic sensing for ambient noise recording. SEG Extended Abstracts.

Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Y, 2007, Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements: Geophysical Journal International, **169**, 1239–1260.

Daley, T. M., B. M. Freifeld, J. Ajo-Franklin, and S. Dou, 2013, Field testing of fiber-optic distributed acoustic sensing (das) for subsurface seismic monitoring: The Leading Edge, **June 2013**, 699–706.

Lindsey, N., 2016, Seismic interferomerty with distributed acoustic sensing for near-surface monitoring of critical infrastructure. Seismological Society of America.

Martin, E. R., J. Ajo-Franklin, S. Dou, N. Lindsey, T. M. Daley, B. M. Freifeld, M. Robertson, A. Wagner, and C. Ulrich, 2015, Interferometry of ambient noise from a trenched distributed acoustic sensing array. SEG Extended Abstracts.

Martin, E. R., N. J. Lindsey, S. Dou, J. B. Ajo-Franklin, A. Wagner, K. Bjella, T. M. Daley, B. Freifeld, M. Robertson, and C. Ulrich, 2016, Interferometry of a roadside das array in fairbanks, ak. SEG Extended Abstracts (to appear).

Nakata, N., R. Snieder, T. Tsuji, K. Larner, and T. Matsuoka, 2011, Shear wave imaging from traffic noise using seismic interferometry by cross-coherence: Geophysics, **76**, SA97–SA106.

Wapenaar, K., D. Draganov, R. Snieder, X. Campman, and A. Verdel, 2010, Tutorial on seismic interferometry: Part 1 – basic principles and applications: Geophysics, **75**, 75A195–75A209.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo, 1998, A comparison of select trigger algorithms for automated global seismic phase and event detection: Bulletin of the Seismological Society of America, **88**, 95–106.

Yoon, C. E., O. O'Reilly, K. J. Bergen, and G. C. Beroza, 2015, Earthquake detection through computationally efficient similarity search: Science Advances, **1**, e1501057–e1501057.