

On the Feasibility of Using the Dense MyShake Smartphone Array for Earthquake Location

by Asaf Inbal, Qingkai Kong, William Savran, and Richard M. Allen

ABSTRACT

MyShake is a growing smartphone-based network for seismological research applications. We study how dense array analysis of the seismic wavefield recorded by smartphones may enhance microearthquake monitoring in urban environments. In such areas, the microearthquake signal-to-noise ratio on smartphones is not well constrained. We address this issue by compiling a seismic noise model for the Los Angeles (LA) metropolitan area using over 500,000 seismograms recorded by stationary phones running MyShake. We confirm that smartphone noise level is reduced during nighttime, and identify strong noise sources such as major traffic highways, the LA airport, and the Long Beach seaport. The noise analysis shows that stationary smartphones are sensitive to human-induced ground motions, and therefore smartphone-derived seismograms may be used to infer the elastic properties of the shallow subsurface. We employ array backprojection analysis on synthetic data to estimate what fraction of LA's smartphone user population is required to install MyShake to enable the location of events whose induced ground motions are below the smartphone noise level. We find that having 0.5% of LA's population download the MyShake app would be sufficient to accurately locate $M > 1$ events recorded during nighttime by stationary phones located at epicentral distances < 5 km. Currently, the MyShake user coverage in LA is approaching a value that will allow us to locate events whose magnitude is near the regional catalog's magnitude of completeness.

Supplemental Content: Description of numerical simulations.

INTRODUCTION

The earthquake catalog is the most important data product provided by a seismic monitoring system. The quality of the earthquake catalog can be estimated from its magnitude of completeness M_{cut} defined as the magnitude cutoff above which all events recorded by the network are registered in the catalog. It is well established that earthquake-size population follows a power-law distribution, according to which earthquake counts increase by a tenfold for each unit drop in magnitude. Hence, reducing the magnitude of completeness by 1

magnitude unit is expected to yield a tenfold reduction in the time interval required for effective seismic monitoring.

Two factors determine the seismic network's monitoring capabilities: station density and ambient noise levels. In southern California, the highest station density is found within the Los Angeles (LA) metropolitan area. Anthropogenic activity in that area generates signals that may mask arrivals from local $M < 2$ earthquakes. As a result, M_{cut} in the LA basin is ~ 2.2 , one unit larger than M_{cut} of less well-instrumented faults in rural areas surrounding LA (Hutton *et al.*, 2010).

Array processing may be used to enhance detection and improve hypocentral location accuracy from measurements of slowness associated with energy impinging on the array (Rost and Thomas, 2002; Meng and Ben-Zion, 2018). If a detailed velocity model is available, time reversal may be used to focus the energy onto the source (Larmat *et al.*, 2009; Chambers *et al.*, 2014; Inbal *et al.*, 2015, 2016; Nakata and Beroza, 2016), thus facilitating hypocenter determination in noisy environments. In the source imaging procedure, the array effectively filters low-incident incoming surface waves generated primarily by shallow noise sources, which leads to an improvement in the signal-to-noise ratio (SNR; Inbal *et al.*, 2015). However, obtaining high-resolution images of micro-earthquake sources requires that the wavefield at the 1–10 Hz frequency band be well resolved, and thus should be sampled at interstation distances smaller than a few hundred meters. To date, near-fault dense array data at this resolution are very scarce.

The MyShake application developed at University of California (UC) Berkeley, utilizes the principle of crowdsourcing to record vast amounts of seismic waveform data from private smartphones running the application, effectively transforming them into a seismic monitoring system (Kong, Allen, Schreier, and Kwon, 2016; Kong, Allen, and Schreier, 2016). Since it was launched in 2016, more than 296,000 people installed MyShake on their smartphones, with 40,000 actually running the application, and about 6000 users who contribute waveform data to our database on a daily basis.

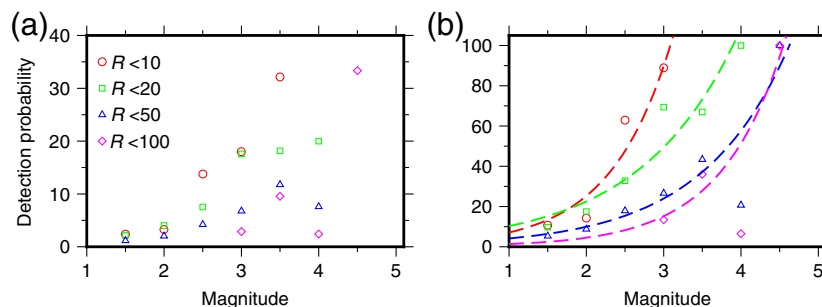
A single smartphone will classify a signal as a likely earthquake signal when the following three conditions are met: (1) the smartphone has been stationary for at least 30 min;

(2) the ratio between the short-time average (STA) to the long-time average (LTA) amplitudes of one of the channels exceeds some pre-defined threshold coded into the application; and (3) an artificial neural network (ANN) detection algorithm installed on the smartphone determines the signal has earthquake characteristics (for more details on the detection procedure, see Kong, Allen, Schreier, and Kwon, 2016; Kong, Allen, and Schreier, 2016). Any smartphone with a working application will also broadcast state-of-health messages every two hours. It is quite common that a smartphone will not issue a triggering message when an earthquake occurs nearby because condition 1 is not met. We find that, on average, the ANN will label an incoming signal as a tectonic event for about 20% of the signals that pass the STA/LTA criteria. Once condition 3 is met, a smartphone will transmit a 5 min window containing three-component ground acceleration sampled at 25 samples per second starting 1 min prior to the trigger time. These accelerations are subsequently archived in our database.

The system was originally designed to detect moderate to large ($M_w > 5$) earthquakes and sends early warning alerts based on smartphone-derived triggers. Version 2.0 of the application, released in March 2018, also enables remote initiation of continuous recording on selected groups of smartphones connected to an external power source for periods of up to about 30 min. In light of this development, our objective here is to assess the potential of the MyShake array for microearthquake monitoring, especially in sparsely monitored urban areas. Earthquake monitoring can be roughly divided into two tasks, the first is the earthquake detection and the second is the determination of the hypocentral parameters and magnitude. We examine the usage of a dense smartphone array for continuous monitoring of seismic activity via a source imaging approach. We focus on hypocentral determination using primarily S -wave energy assuming the signal has been positively identified, and estimate how well earthquakes recorded by a large number (> 100) of phones can be located.

SINGLE-PHONE TRIGGERING PROBABILITIES

We adopt an empirical approach for estimating the detection probability of a single smartphone. For this purpose, we query the MyShake archive for active phones around the time of $M > 2$ events in California. We define a spatial window that ranges out to 100 km away from the epicenter, and a temporal window that includes the propagation time to the furthest smartphones. Active phones within this spatiotemporal window are divided into one of these three groups: nonstationary, stationary (and therefore ready to trigger), and triggered on earthquake. The groups are then binned in concentric areas with increasing size focused around the epicenter. For each magnitude bin, we compute the detection probabilities from

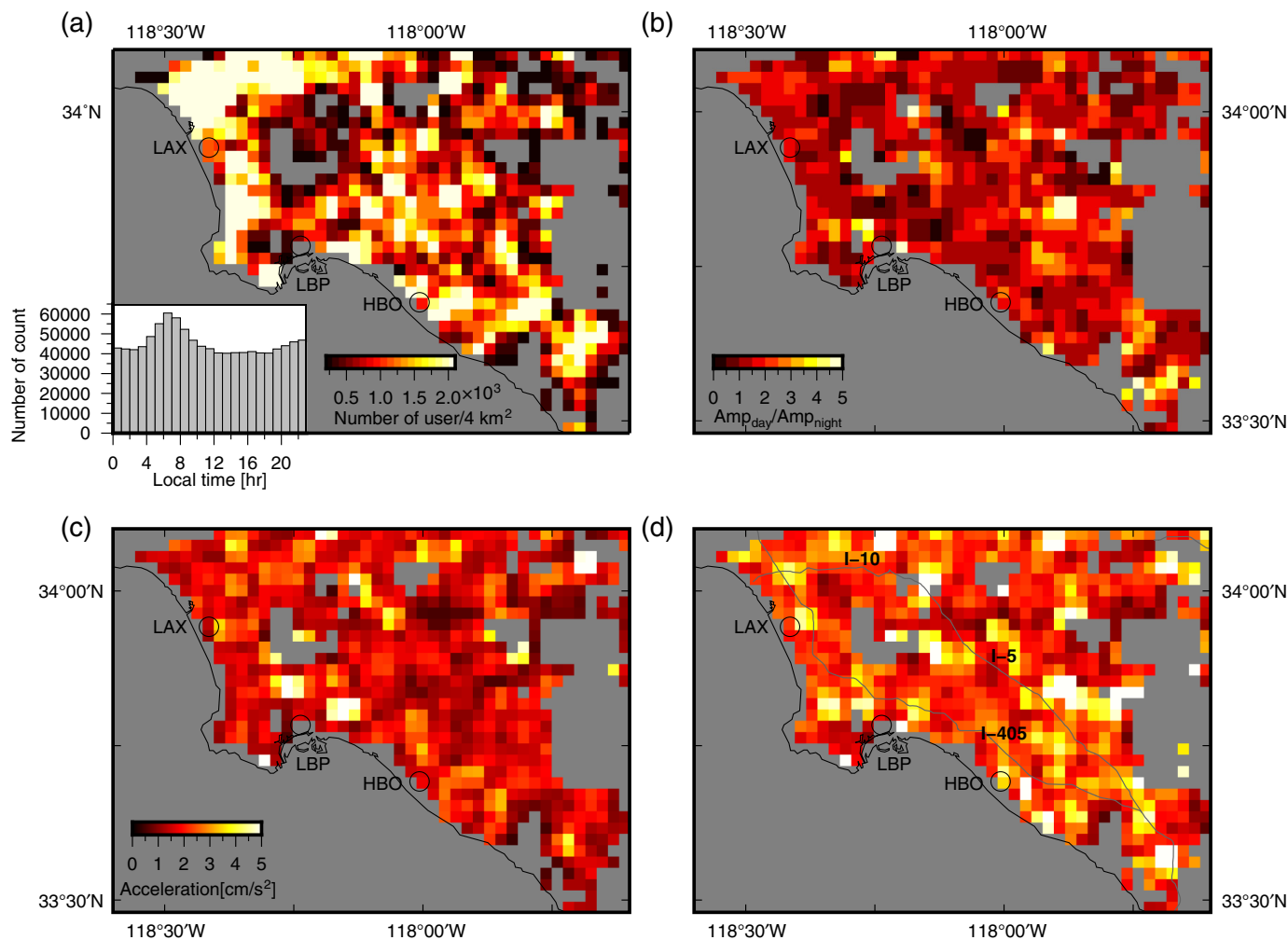


▲ **Figure 1.** Detection probability as a function of magnitude and epicentral distance. Circles, squares, triangles, and diamonds are for users located less than 10, 10–20, 20–50, and 50–100 km away from the epicenter, respectively. (a) Stationary and nonstationary smartphones. (b) Stationary smartphones. Dashed curves are polynomial fit to the data. The color version of this figure is available only in the electronic edition.

the ratio between the detection counts to the count of available phones in each distance bin during the expected arrival time of P - or S waves. We do not discriminate between phones triggered by P - or S -wave arrivals since the most of the phones recording $M < 3$ events were likely triggered by the S wave.

The results are presented in Figure 1, which shows the detection probabilities as a function of magnitude and epicentral distance. We first examine the detection probability of the entire user population including nonstationary and stationary phones. The number of nonstationary smartphones is much larger than stationary ones and nonstationary smartphones cannot issue triggering messages. As a result, the detection probabilities for the entire population are generally below 40% (Fig. 1a). For practical purposes, the measure of detections only on stationary phones is more important for estimating the monitoring capabilities of the MyShake array. This metric is shown in Figure 1b. The results suggest that stationary phones are generally effective for monitoring of $M > 3$ events at short-epicentral distance. A stationary phone is likely (probability $> 90\%$) to detect $M \approx 3$ events occurring at a distance of < 10 km, and $4 < M < 5$ events as far as 100 km from the epicenter. However, the estimate for $M > 4$ events relies on a small number of detection data and therefore exhibits larger scatter relative to smaller magnitude bins.

The results suggest that S waves due to events with $M < 2$ are not likely to trigger individual smartphones running the MyShake app, even at short-epicentral distances (probability $< 20\%$). Indeed, the signal from small magnitude events is well below the average noise level of most smartphones (Kong, Allen, Schreier, and Kwon, 2016). This is the result of factors that reduce the SNR on the smartphones: high-noise floor of microelectromechanical sensors (-55 dB in the 1–10 Hz band vs. -120 dB for traditional seismometers), high level of motion induced by cultural activity in the vicinity of the smartphones, poor coupling between the smartphones and the underlying strata, and limited dynamic range due to the 16-bit digitizer used in the smartphone. In addition, the ANN algorithm designed to discriminate between tectonic and



▲ **Figure 2.** Noise model for the Los Angeles (LA) basin. (a) Distribution of 5 min waveform recorded between 2016 and 2017 totaling 500,000 seismograms. (Inset) Trigger counts as a function of hour of day are shown. (b–d) 3.2–6.4 Hz noise amplitudes within the LA basin. (b) Ratio between day- and nighttime amplitudes, (c) nighttime amplitudes, and (d) daytime amplitudes. Gray curves are for highways with large traffic volumes, with the highway number denoted in black. LAX, Los Angeles International Airport; LBP, Long Beach port; HBO, Huntington Beach Oilfield. The color version of this figure is available only in the electronic edition.

nontectonic signals was trained on a dataset of strong-motion seismograms due to events with $M > 4.5$; and hence, it is biased toward detection of moderate to large magnitude events.

Next, we examine how earthquake location might be improved using dense array analysis. First, we derive a smartphone-based noise model for the LA metropolitan area. We then compare this model to recorded amplitudes of small earthquakes ($M < 3.5$) to establish a frequency band that enhances the smartphone's SNR for array processing purposes.

SMARTPHONE-BASED NOISE MODEL FOR THE LA METROPOLITAN AREA

The LA basin contains over 5000 registered MyShake users. Assuming the percentage of smartphone users in LA is similar to the percentage of users in North America, MyShake users constitute about 0.1% of smartphone users in LA (see [Data and Resources](#)). Between 2016 and 2017, MyShake users in the LA

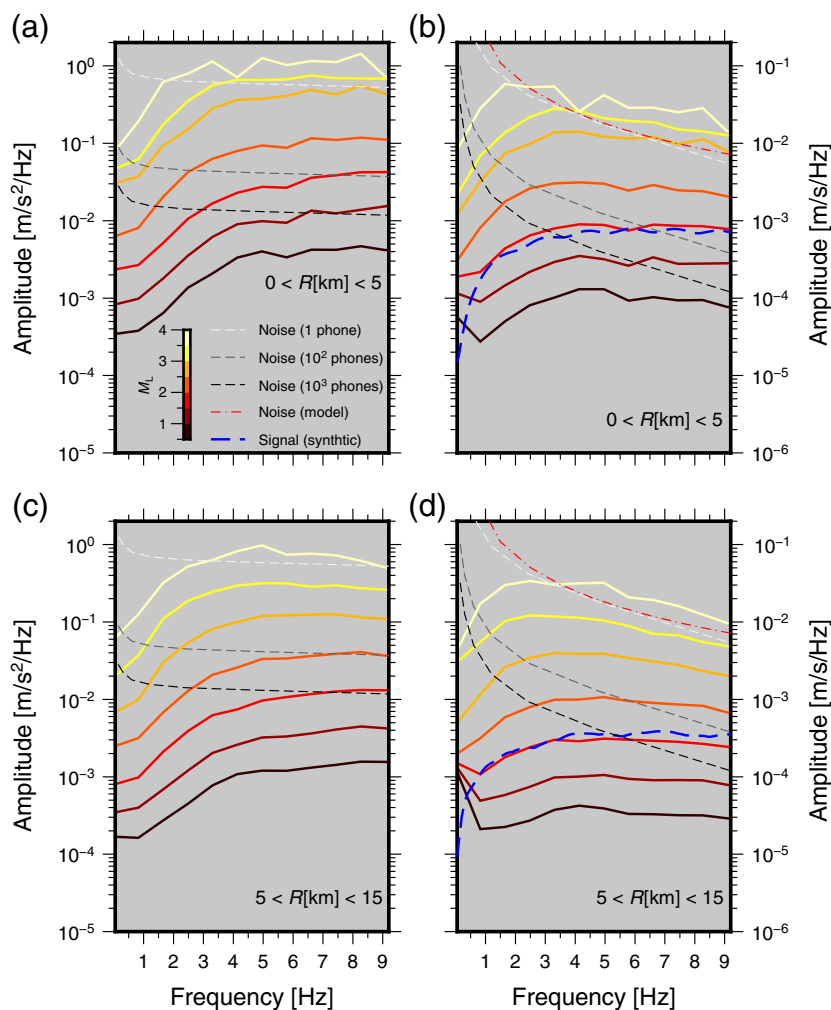
basin contributed approximately 500,000 5 min, three-component waveform data segments, which we use to derive a noise model. Noise is defined here as coherent or incoherent ground motion induced by a nontectonic source. We bin the users in 4 km² cells, extract a 1 min pre-trigger window from each seismogram and take its fast Fourier transform (FFT). We separate between waveform data acquired during day- (6 a.m.–8 p.m.) and nighttime (8 p.m.–6 a.m.) hours, and stack the spectral amplitudes in each cell during these time windows.

Figure 2 presents the user density distribution, and the averaged 3.2–6.4 Hz horizontal ground-motion amplitudes. This range of frequencies is selected based on analysis of the smartphones' SNR (Fig. 3). As a result of increased human activity, the triggered waveform data counts peak during morning hours (Fig. 2a). This indicates that the MyShake detection algorithm tends to mislabel human-induced signals as earthquake signals during these times. Anthropogenic activity also induces ground motions whose daytime amplitudes are on

average a factor of 1.6 higher than nighttime amplitudes. However, in some areas associated with major traffic highways, daytime ground-motion amplitude is much larger than $1.6\times$ nighttime amplitude. For example, daytime horizontal ground-motion amplitudes along Interstate I-5 and Interstate I-405, near the LA international airport, and near the Long Beach seaport are three to four times higher than nighttime amplitudes. Some areas are characterized by relatively high-noise levels that are persistent throughout day- and nighttime. For example, the portion of Interstate I-405 running along Huntington Beach and Long Beach is also very noisy during nighttime. It is possible that regular activity (i.e., not confined to daytime hours) in oilfields lying along that portion of the highway increases the ambient noise levels on the smartphones.

To estimate the SNR for $M < 4$ earthquakes in the LA basin, we use ground motions recorded by the Southern California Seismic Network (SCSN) from 3684 events that occurred within between 2008 and 2018. For each seismogram, we take the FFT of a 1.5 s time window centered on the (manually picked) S -wave arrival time and bin the spectra in concentric regions around the epicenter. We then stack the spectra of the channel containing peak ground motion in 0.5 magnitude bins. Figure 3 presents the averaged horizontal acceleration and velocity spectra within epicentral of up to 15 km away from $0.5 < M < 4$ events, and the average nighttime noise levels, computed using the results presented in Figure 2c. To estimate the expected improvement in SNR introduced using a dense array, we assume the noise between the array's elements is uncorrelated, and scale the noise curves by a factor equal to the square root of the number of channels in the array (Rost and Thomas, 2002). We find that the signal due to a $1.5 < M < 2.0$ recorded by a 100-smartphone array located at < 5 km from the epicenter exceeds the ambient nighttime noise level.

The integration of the acceleration data to velocity brings about an increase of the high-frequency fall-off rate. This trend is observed for both the seismic source spectra and the smartphone noise spectra; however, the SNR in velocity space is larger than in acceleration space. This is because the seismic source acceleration spectra are flat above the corner frequency, whereas the smartphone acceleration noise spectra decay with increasing frequency (Fig. 3a). In velocity space, the target frequency band (1–10 Hz) is above the corner frequency of the smartphone noise but below the corner frequency of earthquakes with $M < 2$ (Fig. 3b). Thus, integrating the accelerations prior to performing array backprojection analysis is



▲ **Figure 3.** Earthquake and smartphone spectra of (a,c) horizontal acceleration and (b,d) velocity. Solid curves are averaged earthquake S -wave spectra recorded at epicentral distances smaller than (a,b) 5 km and (c,d) between 5 and 15 km. Light thick gray dashed curve is for the average smartphone noise amplitude. Dark gray and black thick dashed curves are for theoretical noise amplitudes within arrays containing 100 and 1000 smartphones, respectively. Dashed-dotted curve in (b,d) is fit to the single-smartphone noise spectra. Dashed thick curves in (b,d) is for the spectra of a synthetic source with M_w 2.0. The color version of this figure is available only in the electronic edition.

expected to enhance microearthquake detectability. The horizontal velocity and acceleration ground-motion amplitude due to an $M \approx 3$ event recorded by a single smartphone located less than 5 km away from the epicenter is expected to exceed the smartphone's > 3 Hz noise level by roughly a factor of 3. Earthquakes with $M < 2.5$ will generally not be detected by a single smartphone even at short-epicentral distances (see the [Single-Phone Triggering Probabilities](#) section).

However, the distribution of smartphone users in urban areas offers the opportunity to sample the seismic wavefield at unparalleled density, thus potentially enhancing the detection capabilities. Our investigation of SNR within LA suggests that increasing the density and analyzing > 3 Hz velocity data may improve the ability to locate $M \approx 2$ earthquakes. Yet, it is

not clear what density of users is required for locating an event which induces ground motions whose amplitude is below the smartphone's noise level. Because dense smartphone array data for small magnitude ($M \approx 2$) are not yet available, in the next section, we attempt to address this issue using synthetic tests incorporating the smartphone-derived noise model.

ARRAY PROCESSING SCHEME

We explore the possibility of locating sources in backprojection images constructed from stacked envelopes, using an approach that is similar to the one used by Inbal *et al.* (2015, 2016). Our objective is to accurately determine the location of $M < 2$ earthquakes, and we therefore focus on S -wave energy whose SNR is generally larger than P -wave energy. The ground velocity envelope $s(t)$ is computed by filtering the traces between 3.2 and 6.4 Hz using a fourth-order Butterworth filter, tapering the analyzed window, squaring, and smoothing using a 8-point (0.32 s) running mean window. The envelope stack computed for an array consisting of n elements is defined here as

$$S_j(t) = \sum_{i=1}^n S_{NE}(t + \tau_{ij}), \quad (1)$$

in which τ_{ij} is the S -wave travel time between a source at grid-point j and a sensor at location i , and the subscript NE denotes the sum of the north–south and east–west enveloped seismograms. The location is determined by scanning over a range of possible sources beneath the array. The grid of potential sources is discretized at 250 m in the horizontal and vertical dimensions. In practice, for each time window we compute $S(t)$ for each grid point, select the maxima of each stack, and assign the source location to the grid point with the maxima of the ensemble of stack maximum. For a set of l grid points spanning the possible source locations, and a backprojection time window containing m samples, the selected source location is

$$S_s(t) = \max_{j \in [l]} (\max_{k \in [m]} (S_{jk})). \quad (2)$$

In this study, we use synthetic smartphone data to explore what phone densities are needed to locate events. In addition, our scheme assumes that the smartphone clock does not suffer from instrumental drifts reducing timing accuracy. The MyShake app updates the smartphone clock every hour by communicating with a server hosting a network time protocol. Our tests show that this results in absolute timing errors (at the end of the hour) that may be as large as 2 s, a magnitude that will prohibit array-based processing. Thus, the smartphone's clock will need to be calibrated at a higher rate to allow array processing.

To estimate timing errors for continuously recording collocated phones, we conducted a series of tap tests, in which a group of phones were placed on a table and recorded a single source produced by tapping on the table. We find that the

relative timing error between these phones is of the order of one sample (0.04 s). To estimate the effect of this timing error on the backprojection-derived locations, we use a local 1D velocity model for the LA basin (Shaw *et al.*, 2015), and compute the arrival-time differences at a 1000-element array with a 10 km aperture for sources at various depths. We assume that the epicentral distance is smaller than the array's aperture. If the source depth is larger than about twice the array's aperture, it will exert a larger influence on the mean travel-time differences than the source's horizontal location. To resolve the location using backprojection, the travel-time differences between the array's elements must be smaller than the phone's timing error. For example, 56% and 25% of the travel-time differences are larger than three samples for sources at depths of 5 and 10 km, respectively. From the distributions of travel-time differences, we find that relative timing errors of the order of 0.04 s will increase locations' uncertainties by roughly 3 and 1 km in the vertical and horizontal directions for sources located at a depth of 5 km, and by roughly 4 and 2 km in the vertical and horizontal directions for sources located at a depth of 10 km.

SYNTHETIC TESTS

We carry out a series of synthetic tests whose objective is to evaluate the performance of the array processing scheme. For this purpose, we simulated broadband (up to 30 Hz) ground motions due to a point source embedded in an elastic homogeneous half-space (see the $\text{\textcircled{E}}$ supplemental content), neglecting inelastic attenuation. At the short epicentral distances considered here, our simulations give rise to direct P - and S -wave amplitudes that nicely match the observations. This is illustrated in Figure 3b,d, which shows the match between the azimuthally averaged S -wave synthetic spectra generated from a simulation of M_w 2.0, and the mean raw spectra. The effects of inelastic attenuation become more severe at epicentral distances larger than about 10 km, in which the synthetic spectral amplitude at ~ 9 Hz is roughly twice as large as the observations (Fig. 3d). We therefore confine the maximum epicentral distance in the synthetic tests to be 5 km. The volume onto which data are backprojected measures 10 km in the vertical and horizontal dimensions, and 10 and 5 km in the vertical and horizontal dimensions for tests conducted with 100- and 1000-smartphone arrays, respectively. We focus on S -wave energy recorded on the horizontal channels. For the short epicentral distances sampled here, the time window we backproject may contain both P - and S -wave energy, but we consider the effects of P -wave energy on the backprojection images negligible. This is a valid assumption given that P -wave amplitudes in the horizontal channels are small compared to S -wave amplitude. In addition, backprojection is done using S -wave velocity model, so that P -wave energy is not expected to stack as coherently as S -wave energy.

We simulate noise whose attributes match the LA smartphone velocity noise model (Fig. 3b), by generating colored noise whose amplitude as a function of frequency is

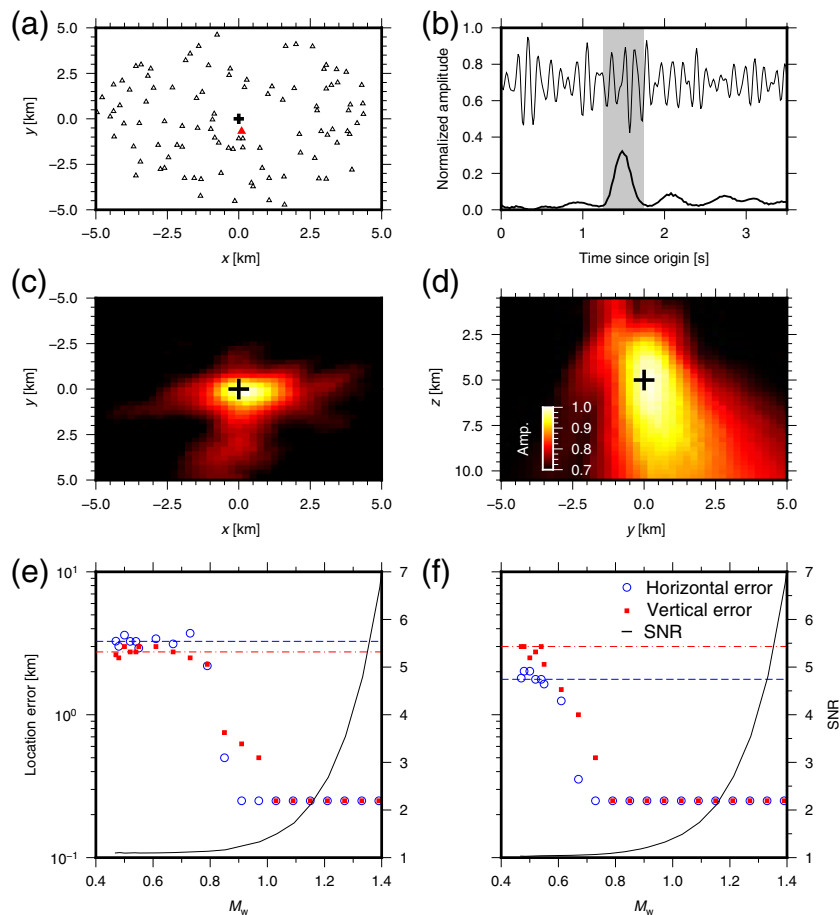
$$A(f) = \alpha f^{-\beta}, \quad (3)$$

in which α and β are coefficients whose value is equal to 0.2 and 1.5, respectively (dashed–dotted curve in Fig. 3b,d). We conduct 100 tests for each input magnitude using arrays consisting of 100 and 1000 sensors uniformly distributed around the epicenter. The noise is assumed to be (1) uncorrelated between smartphones and (2) uncorrelated between channels. The former assumption is reasonable given the average spacing between sensors. To verify that noise between horizontal channels is uncorrelated we cross correlated the pre-trigger, collocated horizontal accelerations used to construct the noise model (see the [Smartphone-Based Noise Model for the LA Metropolitan Area](#) section). The mean of the maximum cross-correlation values is 0.06, which validates our assumption.

The results of the synthetic tests are presented in Figure 4. The distribution of phones in one of the synthetic tests containing 100 receivers is shown in Figure 4a. The synthetic seismograms consist of both horizontal channels with simulated colored noise (an example trace shown by the thin line in Fig. 4b). Figure 4c,d presents the images constructed by backprojecting a 6 s window containing both P and S waves due to an M_w 1.03 using a 100-smartphone array. An example of a stack of envelopes in which P -wave amplitudes are negligible relative to S -wave amplitudes is shown in Figure 4b.

The images formed from synthetic data focus around the horizontal input source location, with some smearing occurring primarily along the depth axis. This is a common artifact of the backprojection method we apply and is the result of the limited aperture of the array and noise present in the data. The location error, defined as the difference between the input location and the maxima of the backprojection image, is about 250 m in the horizontal and vertical directions. This level of accuracy may be improved by backprojecting higher frequency energy onto a finer grid.

If the velocity model is perfectly known, using only a small number of smartphones located at an epicentral distance < 5 km will allow us to accurately determine the location of earthquakes whose magnitude is below M_{cut} of the LA basin. If the phones' timing errors are of the order of one sample (0.04 s), then location uncertainties are expected to be less than 5 km. This range is comparable to the absolute source location uncertainties that result from traditional network-



▲ **Figure 4.** Synthetic source imaging of an M_w 1.03 and location errors as a function of magnitude. (a) Distribution of receivers for one test including 100 smartphones uniformly distributed around a synthetic source. (b) Normalized amplitude as a function of time. Thin line is for velocities filtered between 3.2 and 6.4 Hz at a site nearest to the epicenter (triangle in a). Bold curve is for the amplitude of the stack at the grid point corresponding to the maxima of the backprojection shown in (c) and (d). Gray rectangle denotes the S -wavetrain. (c) Backprojected maximal stack amplitude as a function of horizontal location at 5 km depth. (d) Backprojected maximal stack amplitude as a function of depth. Cross section is along $x = 0$. Black cross is for the input source location. Amplitudes are normalized by the maxima of the backprojection volume. (e,f) Location errors as a function of magnitude for synthetic tests employing (e) 100 and (f) 1000 sensors. Values shown are median differences between the input and output location (defined as the maxima in the backprojection image) from 100 synthetic tests. Circles and squares indicate the error in the horizontal and vertical directions, respectively. Horizontal and vertical location error for input data containing only colored noise are shown with the dashed and dashed–dotted curves, respectively. Solid lines are for signal-to-noise ratio (SNR) of the stack at the grid point corresponding to the synthetic source location (indicated by the bold curve in at the bottom of b). The color version of this figure is available only in the electronic edition.

wide location approach (for a network whose average station spacing is about 5 km). We further evaluate the effect of absolute timing errors on the backprojection-derived locations by randomly shifting the input waveforms using time delays that are normally distributed with standard deviation equal to two

samples (0.08 s, a factor of 2 larger than the absolute timing error we estimated from the tap test presented in the [Array Processing Scheme](#) section). We find that this timing error magnitude tends to reduce the resolution on the hypocenter's location (i.e., increases smearing) but does not affect the location of the images' maxima.

Figure 4e,f presents the median horizontal and vertical location errors as a function of magnitude from 100 synthetic tests using 100 and 1000 phones, respectively. The medians are computed after discarding output locations at the boundary of the grid. The expected location error for data containing only noise is shown by the dashed curves in Figure 4e,f. The horizontal location error of the noise-only dataset for 1000-smartphone arrays is smaller than that of the 100-smartphone array (dashed blue curve in Fig. 4e,f), which is the result of the smaller horizontal dimension of grid used for backprojecting the 1000-smartphone array data.

Using 100 stationary smartphones distributed within a radius of 5 km away from the epicenter (i.e., densities of 1.27 users/km²) allow us to determine the location of $M_w \approx 1.0$ events occurring at depths < 10 km. The location uncertainties derived from 100 simulations correspond to < 2 and < 4 km in the horizontal and vertical directions (filled squares and open circles in Fig. 4e), respectively. Increasing the user density by a factor of 10 allows us to accurately determine the location of events as small as $M_w \approx 0.8$. However, smaller magnitude events are not likely to exceed the array's noise level even for configurations containing 1000 smartphones. The smartphones acceleration data are stored in half-precision format, which further results in poor resolution on very weak accelerations.

Figure 4e,f shows the SNR of the stacked waveforms, defined here as the ratio between the median amplitude in a 0.5 s window containing the *S*-wavetrain (denoted by the gray rectangle in Fig. 4b) to the median amplitude in a 0.5 s window containing simulated noise outside the *S*-wave window. The SNR is important for developing a continuous earthquake detection strategy. We find that events that were well located using a 1000-smartphone array (Fig. 4f) correspond to stack with SNR ≈ 1.1 . This suggests that an STA/LTA event detector applied to the stacked waveform data might miss events with $M_w < 1.0$ recorded by very dense arrays. The SNR's dependency on magnitude is similar for arrays of 100 and 1000 smartphones. However, increasing the array's density enhances focusing and results in sharper backprojection images. Thus, one possible detection approach is to rely on the statistical properties of the images rather than on the raw waveforms or the SNR of $S_i(t)$ (the grid point assumed to contain the source). Developing a robust array-based classification scheme will be part of our future work.

The average population density in the LA metropolitan area is 2300 people/km², roughly 60% of which use smartphones on a regular basis (see [Data and Resources](#)). Our records indicate that >60% of the active phones are also stationary for intervals longer than 30 min (Kong *et al.*, 2018). Thus, a fraction as small as 0.5% of the population is required

to install MyShake for us to be able to locate $M \approx 1$ occurring beneath the city during nighttime. We note that approximately 0.1% of smartphone users have already downloaded the MyShake app (although not all of the users continue to run the app). In real-life applications, we expect the monitoring system to continuously backproject data recorded by the array and scan the stacked data for earthquake signals.

DISCUSSION

Our analysis of smartphone earthquake triggers suggests that $M > 3$ are likely to be detected by a stationary smartphone located at epicentral distances smaller than 10 km, and that $M > 4$ events may be detected by a stationary smartphone located several tens of kilometers from the epicenter (Fig. 1). Under favourable SNR conditions, events with magnitude as small as $M = 2.5$ may be recorded by a nearby smartphone (Kong, Allen, and Schreier, 2016). The current single-phone detection capabilities rely on an ANN that was designed to detect seismic arrivals due to events with $M > 4.5$. However, the detection statistics may improve once the ANN algorithm is trained on the entire MyShake waveform database (since 2016), so as to enhance the detectability of events with $M < 3$. We expect that this will improve MyShake's detection capability to a level comparable with performance of local networks in seismically active, sparsely instrumented areas such as Nepal (Pandey *et al.*, 1999). Smartphone data provided by the large number of users in the metropolitan areas residing within these regions may be used for characterizing active faults, thus contributing to seismic hazard mitigation.

We derive a smartphone-based noise model for the LA metropolitan area, which allows us to assess the performance of the MyShake smartphone array operating there. The 1–10 Hz smartphone-recorded noise level is dependent on the time of day and the location within the city, a pattern that is frequently observed in urban environments (e.g., Bonnefoy-Claudet *et al.*, 2006; Groos and Ritter, 2009). Previous studies analyzed the SNR on low-cost sensors (Evans *et al.*, 2014; Kong, Allen, and Schreier, 2016), derived the likelihood of detecting events as a function of their magnitude in a manner similar to the one presented here (Clinton and Heaton, 2002), and described and evaluated the performance of a low-cost sensor network for monitoring strong earthquake motions (e.g., Lawrence *et al.*, 2014; Clayton *et al.*, 2015). Our analysis of the SNR on low-cost sensors installed in smartphones differs from the SNR analysis presented in these studies mainly because the MyShake user density allows us to estimate noise amplitudes across our study area at a finer resolution than has been obtained thus far. In addition, whereas previous studies attempted to assess whether data from low-cost sensors may be used for providing earthquake early warning alerts, here we focus on the earthquake monitoring capabilities of the MyShake array. Using array-based techniques for earthquake location purposes is more computationally challenging than relying on phase picks within a network. However, array analysis may enhance monitoring of small earthquakes, even

in situations when the $\text{SNR} < 1$, as is frequently the case for $M < 2$ recorded on smartphones. Our estimates may be used for future installments of urban dense seismic networks, and for analyzing the recording capacity of the permanent network deployed in the LA basin. The SCSN coverage is densest in the LA metropolitan area, and yet its magnitude of completeness there is about one unit larger than within rural regions in southern California. The detection capabilities may be significantly improved if dense array data are available (Inbal *et al.*, 2015, 2016). Assessing the advantages from future dense deployments requires that the SNR is well constrained in space and time. Currently, the microearthquake signal amplitude within the LA basin is well constrained (Fig. 3), but the noise level is not. The noise maps presented in this study may be used for estimating the density of low-cost sensors that are needed to be deployed to meet a predefined earthquake detection criteria.

We find that some areas in the LA basin exhibiting high-noise levels are correlated with known sources of noise (highways, airport, seaport), whereas other areas exhibit high-noise levels during day- and nighttime. The correlation with known sources of noise along the Huntington oilfield is marginal. However, it is possible that at least part of the 1–10 Hz seismic motions recorded by smartphones in that area is the result of ongoing injection in adjacent oilfields. That smartphones are not insensitive to human-induced seismic ground motion suggests that, similar to studies using high-quality seismic data, the ambient 1–10 Hz noise field recorded on a smartphone array may be used to image shallow subsurface heterogeneities (Bowden *et al.*, 2015; Hillers *et al.*, 2016) or recover the shallow inelastic attenuation profile (Fuchs and Bokelmann, 2018; Inbal *et al.*, 2018; Li *et al.*, 2018).

Results from synthetic tests suggest that locating microearthquakes with the dense MyShake smartphone array is feasible. This suggests that smartphone data may be used in conjunction with data from high-quality sensors to improve the detection of small earthquakes in poorly instrumented cities such as Kathmandu. Data from a 100-smartphone array with a 10 km aperture allow us to recover the location of $M \approx 1.0$ synthetic sources (if they occur beneath the array during nighttime). Increasing the array's density by a factor of 10 allows us to recover the location of events as small as $M \approx 0.8$. Our method focuses energy envelopes onto the source and is hence relatively robust against small errors in the velocity model (Beskardes *et al.*, 2018). However, scattering from small-scale heterogeneities or at layer interfaces may introduce artifacts to backprojection images. This phenomenon has been reported for arrays located at teleseismic distances (Yue *et al.*, 2017) but has not been considered in this study. For the range of magnitude and epicentral distance considered here, it is unlikely that the coherent stack of any scattered phase will have an amplitude larger than the stack of the direct S -wave phase, given that both the scattered and direct phases are included in the same analyzed time window. The effect of scattering in that case might be to produce secondary focal spots in the image, thus increasing its ambiguity.

We are currently analyzing 3.2–6.4 Hz energy, a range that is selected based on our SNR analysis (Fig. 3b) and are limited by the Nyquist frequency of our data (12.5 Hz). Increasing the sampling rate on the smartphones will allow us to examine even higher frequencies, but will result in very large data volumes further complicating the backprojection analysis presented here. In addition, increasing the sampling rate tends to increase the smartphone's temperature, which might limit our recording capacity. At the range of source–receiver distances considered here, refining the grid of potential sources below 250 m dramatically increases the computational cost. Certain optimization techniques and usage of graphical processing units may reduce the cost of carrying the backprojection analysis at a finer level of discretization.

Version 2.0 of the MyShake app allows us to record continuously on selected phones plugged into power. The availability of continuous data provided by our users will allow us to test the ideas presented in this article. In particular, continuous backprojection of data from over 5000 registered users in the LA area will allow us to identify sources of permanent noise to estimate the level of coherency within the array, and to test whether earthquakes with $M < M_{\text{cut}}$ may be located using the MyShake smartphone data.

CONCLUSIONS

We use data from active phones located around $M > 2$ events in California to assess the on-phone detection capabilities. We find that there is about 90% probability of detecting an $M \approx 3$ event on a stationary smartphone located < 10 from the epicenter, or of detecting a moderate event ($4 < M < 5$) by a smartphone located < 100 km from the epicenter.

We derive a smartphone-based noise model for the LA metropolitan area. Noise amplitudes show some degree of correlation with known anthropogenic noise sources such as major highways, the LA international airport, and the Long Beach seaport.


We estimate the amplitude of $M < 3.5$ earthquakes occurring within the LA area. The level of the signal and the empirical noise model we derive are used to estimate the smartphone density required to locate $M < 1.5$ earthquakes occurring within the LA basin using array backprojection techniques.

We find that densities of 1.27 users/km² uniformly distributed within 5 km from an $M \approx 1.0$ synthetic source are required to accurately recover the input source location.

Given the density of smartphone users in LA, a fraction as small as 0.5% of the population is required to install MyShake to enable backprojection derived locations of earthquakes whose magnitude is below the SCSN catalog magnitude of completeness in LA.

We are currently collecting continuous smartphone data that will be used to better characterize sources of high-frequency noise and to compile a smartphone-based earthquake catalog for the LA basin.

DATA AND RESOURCES

Data recorded by MyShake are currently archived at Berkeley Seismological Laboratory and are constrained by the privacy policy of MyShake (see <http://myshake.berkeley.edu/privacy-policy/index.html>, last accessed June 2018). For information about access to the data for research purposes, contact rallen@berkeley.edu. Seismic data used in this study were recorded and maintained by the Southern California Seismic Network (SCSN; doi: [10.7914/SN/CI](https://doi.org/10.7914/SN/CI)). North America smartphone users statistics are based on the Ericsson Mobility Report at <https://www.ericsson.com/en/mobility-report> (last accessed June 2018). 

ACKNOWLEDGMENTS

A. Inbal's work at the University of California, Berkeley, was supported by a postdoctoral fellowship for seismological research from the Israeli Minister of Energy and Water. This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (Award Numbers OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. This work is also part of the "Extending the Spatiotemporal Scales of Physics-based Seismic Hazard Analysis" Petascale Computing Resource Allocations (PRAC) allocation support by the National Science Foundation (Award Number OCI-1440085). The MyShake project is funded by the Gordon and Betty Moore Foundation through Grant Number GBMF5230 to UC Berkeley.

REFERENCES

Beskardes, G. D., J. A. Hole, K. Wang, M. Michaelides, Q. Wu, M. C. Chapman, K. K. Davenport, L. D. Brown, and D. A. Quiros (2018). A comparison of earthquake backprojection imaging methods for dense local arrays, *Geophys. J. Int.* **212**, no. 3, 1986–2002, doi: [10.1093/gji/ggx520](https://doi.org/10.1093/gji/ggx520).

Bonnefoy-Claudet, S., F. Cotton, and P.-Y. Bard (2006). The nature of noise wavefield and its applications for site effects studies: A literature review, *Earth Sci. Rev.* **79**, 205–227, doi: [10.1016/j.earscirev.2006.07.004](https://doi.org/10.1016/j.earscirev.2006.07.004).

Bowden, D. C., V. C. Tsai, and F. C. Lin (2015). Site amplification, attenuation, and scattering from noise correlation amplitudes across a dense array in Long Beach, CA, *Geophys. Res. Lett.* **42**, no. 5, 1360–1367, doi: [10.1002/2014GL062662](https://doi.org/10.1002/2014GL062662).

Chambers, K., B. D. Dando, G. A. Jones, R. Velasco, and S. A. Wilson (2014). Moment tensor migration imaging, *Geophys. Prospect.* **62**, no. 4, 879–896, doi: [10.1111/1365-2478.12108](https://doi.org/10.1111/1365-2478.12108).

Clayton, R. W., T. Heaton, M. Kohler, M. Chandy, R. Guy, and J. Bunn (2015). Community seismic network: A dense array to sense earthquake strong motion, *Seismol. Res. Lett.* **86**, no. 5, 1354–1363, doi: [10.1785/0220150094](https://doi.org/10.1785/0220150094).

Clinton, J. F., and T. H. Heaton (2002). Potential advantages of a strong-motion velocity meter over a strong-motion accelerometer, *Seismol. Res. Lett.* **73**, no. 3, 332–342, doi: [10.1785/gssrl.73.3.332](https://doi.org/10.1785/gssrl.73.3.332).

Evans, J. R., R. M. Allen, A. I. Chung, E. S. Cochran, R. Guy, M. Hellweg, and J. F. Lawrence (2014). Performance of several low-cost

accelerometers, *Seismol. Res. Lett.* **85**, no. 1, 147–158, doi: [10.1785/0220130091](https://doi.org/10.1785/0220130091).

Fuchs, F., and G. Bokelmann (2018). Equidistant spectral lines in train vibrations, *Seismol. Res. Lett.* **89**, no. 1, 56–66, doi: [10.1785/0220170092](https://doi.org/10.1785/0220170092).

Groos, J. C., and J. R. R. Ritter (2009). Time domain classification and quantification of seismic noise in an urban environment, *Geophys. J. Int.* **179**, no. 2, 1213–1231, doi: [10.1111/j.1365-246X.2009.04343.x](https://doi.org/10.1111/j.1365-246X.2009.04343.x).

Hillers, G., P. Roux, M. Campillo, and Y. Ben-Zion (2016). Focal spot imaging based on zero lag cross-correlation amplitude fields: Application to dense array data at the San Jacinto fault zone, *J. Geophys. Res.* **121**, no. 11, 8048–8067, doi: [10.1002/2016JB013014](https://doi.org/10.1002/2016JB013014).

Hutton, K., J. Woessner, and E. Hauksson (2010). Earthquake monitoring in southern California for seventy-seven years (1932–2008), *Bull. Seismol. Soc. Am.* **100**, no. 2, 423–446, doi: [10.1785/0120090130](https://doi.org/10.1785/0120090130).

Inbal, A., J. P. Ampuero, and R. W. Clayton (2016). Localized seismic deformation in the upper mantle revealed by dense seismic arrays, *Science* **354**, no. 6308, 88–92, doi: [10.1126/science.aaf1370](https://doi.org/10.1126/science.aaf1370).

Inbal, A., R. W. Clayton, and J.-P. Ampuero (2015). Imaging widespread seismicity at mid lower crustal depths beneath Long Beach, CA, with a dense seismic array: Evidence for a depth-dependent earthquake size distribution, *Geophys. Res. Lett.* **42**, no. 15, 6314–6323, doi: [10.1002/2015GL064942](https://doi.org/10.1002/2015GL064942).

Inbal, A., T. Cristea-Platon, J. Ampuero, G. Hillers, D. Agnew, and S. E. Hough (2018). Sources of long-range anthropogenic noise in southern California and implications for tectonic tremor detection sources of long-range anthropogenic noise in southern California, *Bull. Seismol. Soc. Am.* **108**, no. 6, 3511, doi: [10.1785/0120180130](https://doi.org/10.1785/0120180130).

Kong, Q., R. M. Allen, and L. Schreier (2016). MyShake: Initial observations from a global smartphone seismic network, *Geophys. Res. Lett.* **43**, no. 18, 9588–9594, doi: [10.1002/2016GL070955](https://doi.org/10.1002/2016GL070955).

Kong, Q., R. M. Allen, L. Schreier, and Y.-W. Kwon (2016). MyShake: A smartphone seismic network for earthquake early warning and beyond, *Sci. Adv.* **2**, no. 2, e1501, 055–e1501, 055, doi: [10.1126/sciadv.1501055](https://doi.org/10.1126/sciadv.1501055).

Kong, Q., A. Inbal, R. M. Allen, L. Qin, and A. Puder (2018). Machine learning aspects of the MyShake global smartphone seismic network, *Seismol. Res. Lett.* **90**, no. 2A, 546–552, doi: [10.1785/0220180309](https://doi.org/10.1785/0220180309).

Larmat, C. S., R. A. Guyer, and P. A. Johnson (2009). Tremor source location using time reversal: Selecting the appropriate imaging field, *Geophys. Res. Lett.* **36**, no. 22, L22304, doi: [10.1029/2009GL040099](https://doi.org/10.1029/2009GL040099).

Lawrence, J. F., E. S. Cochran, A. Chung, A. Kaiser, C. M. Christensen, R. Allen, J. W. Baker, B. Fry, T. Heaton, D. Kilb, *et al.* (2014). Rapid earthquake characterization using MEMS accelerometers and volunteer hosts following the M 7.2 Darfield, New Zealand, earthquake, *Bull. Seismol. Soc. Am.* **104**, no. 1, 184–192, doi: [10.1785/0120120196](https://doi.org/10.1785/0120120196).

Li, C., Z. Li, Z. Peng, C. Zhang, N. Nakata, and T. Sickbert (2018). Long-period long-duration events detected by the IRIS community waveform demonstration experiment in Oklahoma: Tremor or train signals?, *Seismol. Res. Lett.* **89**, no. 5, 1652–1659, doi: [10.1785/0220180081](https://doi.org/10.1785/0220180081).

Meng, H., and Y. Ben-Zion (2018). Detection of small earthquakes with dense array data: Example from the San Jacinto fault zone, southern California, *Geophys. J. Int.* **212**, no. 1, 442–457, doi: [10.1093/gji/ggx404](https://doi.org/10.1093/gji/ggx404).

Nakata, N., and G. C. Beroza (2016). Reverse time migration for micro-seismic sources using the geometric mean as an imaging condition, *Geophysics* **81**, no. 2, KS51–KS60, doi: [10.1190/geo2015-0278.1](https://doi.org/10.1190/geo2015-0278.1).

Pandey, M., R. Tandukar, J. Avouac, J. Vergne, and T. Héritier (1999). Seismotectonics of the Nepal Himalaya from a local seismic network, *J. Asian Earth Sci.* **17**, nos. 5/6, 703–712, doi: [10.1016/S1367-9120\(99\)00034-6](https://doi.org/10.1016/S1367-9120(99)00034-6).

Rost, S., and C. Thomas (2002). Array seismology: Methods and applications, *Rev. Geophys.* **40**, no. 3, 1008, doi: [10.1029/2000RG000100](https://doi.org/10.1029/2000RG000100).

Shaw, J. H., A. Plesch, C. Tape, M. P. Suess, T. H. Jordan, G. Ely, E. Hauksson, J. Tromp, T. Tanimoto, R. Graves, *et al.* (2015).

Unified structural representation of the southern California crust and upper mantle, *Earth Planet. Sci. Lett.* **415**, 1–15, doi: [10.1016/j.epsl.2015.01.016](https://doi.org/10.1016/j.epsl.2015.01.016).

Yue, H., J. C. Castellanos, C. Yu, L. Meng, and Z. Zhan (2017). Localized water reverberation phases and its impact on backprojection images, *Geophys. Res. Lett.* **44**, no. 19, 9573–9580, doi: [10.1002/2017GL073254](https://doi.org/10.1002/2017GL073254).

Asaf Inbal¹
Department of Geophysics
Tel Aviv University
Klausner Street 17, Ramat-Aviv
Tel Aviv 69978, Israel
asafinbal@tauex.tau.ac.il

Qingkai Kong
Richard M. Allen
Berkeley Seismological Laboratory
University of California
McCone Hall, 215 Haviland Path #4760

Berkeley, California 94703 U.S.A.
kongqk@berkeley.edu
rallen@berkeley.edu

William Savran
University of Southern California
3651 Trousdale Parkway
Los Angeles, California 90089 U.S.A.
wsavran@gmail.com

Published Online 10 April 2019

¹ Also at Berkeley Seismological Laboratory, University of California, McCone Hall, 215 Haviland Path #4760, Berkeley, California 94703 U.S.A.