Ⓔ

Assessing the Sensitivity and Accuracy of the MyShake Smartphone Seismic Network to Detect and Characterize Earthquakes

by Qingkai Kong, Sarina Patel, Asaf Inbal, and Richard M. Allen

ABSTRACT

MyShake harnesses private and personal smartphones to build a global seismic network. It uses the accelerometers embedded in all smartphones to record ground motions induced by earthquakes, returning recorded waveforms to a central repository for analysis and research. A demonstration of the power of citizen science, MyShake expanded to six continents within days of being launched and has recorded 757 earthquakes in the first 2 yr of operation. The data recorded by MyShake phones have the potential to be used in scientific applications, thereby complementing current seismic networks. In this article, we (1) report the capabilities of smartphone sensors to detect earthquakes by analyzing the earthquake waveforms collected by MyShake; (2) determine the maximum epicentral distance at which MyShake phones can detect earthquakes as a function of magnitude; and (3) then determine the capabilities of the MyShake network to estimate the location, origin time, depth, and magnitude of earthquakes. In the case of earthquakes for which MyShake has provided four or more phases (*P*- or *S*-wave signals) and an azimuthal gap $<180°$ (21 events), the median ($\pm$ standard deviations) of the location, origin time, and depth errors are 2.7 (± 2.8) km, 0.2 (± 1.2) s, and 0.1 (± 4.9) km, respectively, relative to the U.S. Geological Survey global catalog locations. Magnitudes are also estimated and have a mean error of 0.0 and standard deviation of 0.2. These preliminary results suggest that MyShake could provide basic earthquake catalog information in regions that currently have no traditional networks. With an expanding MyShake network, we expect the event detection capabilities to improve and provide useful data on seismicity and hazards.

Supplementary Content: Figures showing earthquake parameter estimations for the 44 events that have more than four phase pickings to support the article and a collection of selected seismic waveforms recorded by MyShake users.

INTRODUCTION

After more than a century of development, geophysical instrumentation has become increasingly more diversified. High-quality seismic instruments (Havskov and Alguacil, 2016), geodetic instruments (Larson, 2009), and Interferometric Synthetic Aperture Radar (Bürgmann *et al.*, 2000) enable new discoveries and understanding of earthquake physics and active tectonics. In addition, the emergence of various new low-cost and potentially more pervasive sensing technologies provides new ways of detecting earthquakes, collecting additional data to learn about the earthquake process, and potentially making important contributions to seismology (Allen, 2012).

Citizen science has expanded the ability to assess and respond to earthquake hazards. "Did You Feel It?," a U.S. Geological Survey (USGS) earthquake survey platform, collects macroseismic intensity data from Internet users, which are then used to generate intensity maps immediately after earthquakes (Wald *et al.*, 2001; Wald and Dewey, 2005; Atkinson and Wald, 2007). Twitter messages from users who felt an earthquake can be used to detect and characterize events in real time (Earle, 2010; Earle *et al.*, 2010; Sakaki *et al.*, 2010). By monitoring traffic to its website, the European–Mediterranean Seismological Centre can detect and assess the effect of an earthquake within a few minutes (Bossu *et al.*, 2012). Low-cost micro-electro-mechanical systems (MEMS) sensors inside computers or placed in specially installed stand-alone boxes in homes or offices can be used to monitor and study earthquakes (Cochran *et al.*, 2009; Chung *et al.*, 2011; Clayton *et al.*, 2012, 2015; Hsieh *et al.*, 2014; Wu, 2015; Wu *et al.*, 2016; Jan *et al.*, 2018). Distributed acoustic sensing transforms telecommunication fiberoptic cables into seismic arrays, enabling meter-scale recording over kilometers of linear fiber length (Dou *et al.*, 2017; Lindsey *et al.*, 2017).

As more and more people have access to and a need for smartphones these small devices comprise an ever more widespread and dense sensing network around the globe. Seismologists have learned that smartphones can be used in different ways to detect earthquakes. For example, by monitoring when users turn on a specific earthquake application on their phone, earthquakes can be recognized within minutes as clusters of application activity (Bossu *et al.*, 2015, 2018; Steed *et al.*, 2019). The MEMS sensors inside the smartphones that record acceleration have also been shown to be capable of detecting earthquakes (Faulkner *et al.*, 2011; Dashti *et al.*, 2012, 2014; Kong *et al.*, 2015; Finazzi, 2016; Kong, Allen, Schreier, *et al.*, 2016).

MyShake was launched by University of California, Berkeley, in 2016 as a citizen science project. It aims to build a global smartphone seismic network that can be used for research, ultimately contributing to a reduction in earthquake hazards. In the first 2 yr, just under 300,000 people downloaded the MyShake app globally. Now, 2 yr after the launch, there are 40,000 phones with the app installed, and on any given day, ~7000 phones contribute data. The core of MyShake is an artificial neural network, built into the on-phone app, that is trained to recognize earthquake-like movement and distinguish it from everyday human movements and a series of machine learning models that support the confirmation and estimation of the earthquakes (Kong, Allen, Schreier, *et al.*, 2016; Kong, Inbal, *et al.*, 2019). Whenever the phone detects the earthquake-like movement, a real-time message with the trigger location, time, and amplitude is sent to the server for earthquake early warning purposes. At the same time, a 5 min segment of three-component acceleration data is stored on the phone and then uploaded to the MyShake server to be analyzed when the phone connect to power and Wi-Fi. The time series starts 1 min before the trigger is detected and continues for 4 min post-trigger. Data collected by MyShake can be used in various applications. Kong, Allen, and Schreier (2016) show examples of the waveform data recorded by MyShake, illustrating the potential to use them in different seismological applications. Real-time trigger data from MyShake users show earthquake early warning can be done using the smartphones (Kong, Inbal, *et al.*, 2018). The MyShake waveform data can also potentially be used to monitor the structural health state of buildings (Kong, Allen, *et al.*, 2018). In addition, using the MyShake arrays, the system can potentially detect microseismicity and monitor noise in urban areas (Inbal *et al.*, 2019).

In this article, we explore the capabilities of the MyShake smartphone seismic network by mining the archive waveforms recorded to date. Comparing these data with a global earthquake catalog, we explore the detection capabilities of the smartphone network. We explore the sensitivity of the network by determining the distance to which smartphones can detect an earthquake as a function of magnitude. We also show how the waveforms recorded by MyShake phones can be used to estimate basic earthquake parameters, including location, origin time, depth, and magnitude of the earthquake. This illustrates how the MyShake network could be used to monitor earthquake activity in regions of dense populations that currently have no seismic network. Even though the number of users in the MyShake network is large, it still has many limitations and challenges before it can be used as a routine seismic network. These include users losing interest in the app, phones moving around, phones clustered in cities, and so on. We discuss some of these challenges in the Discussion section.

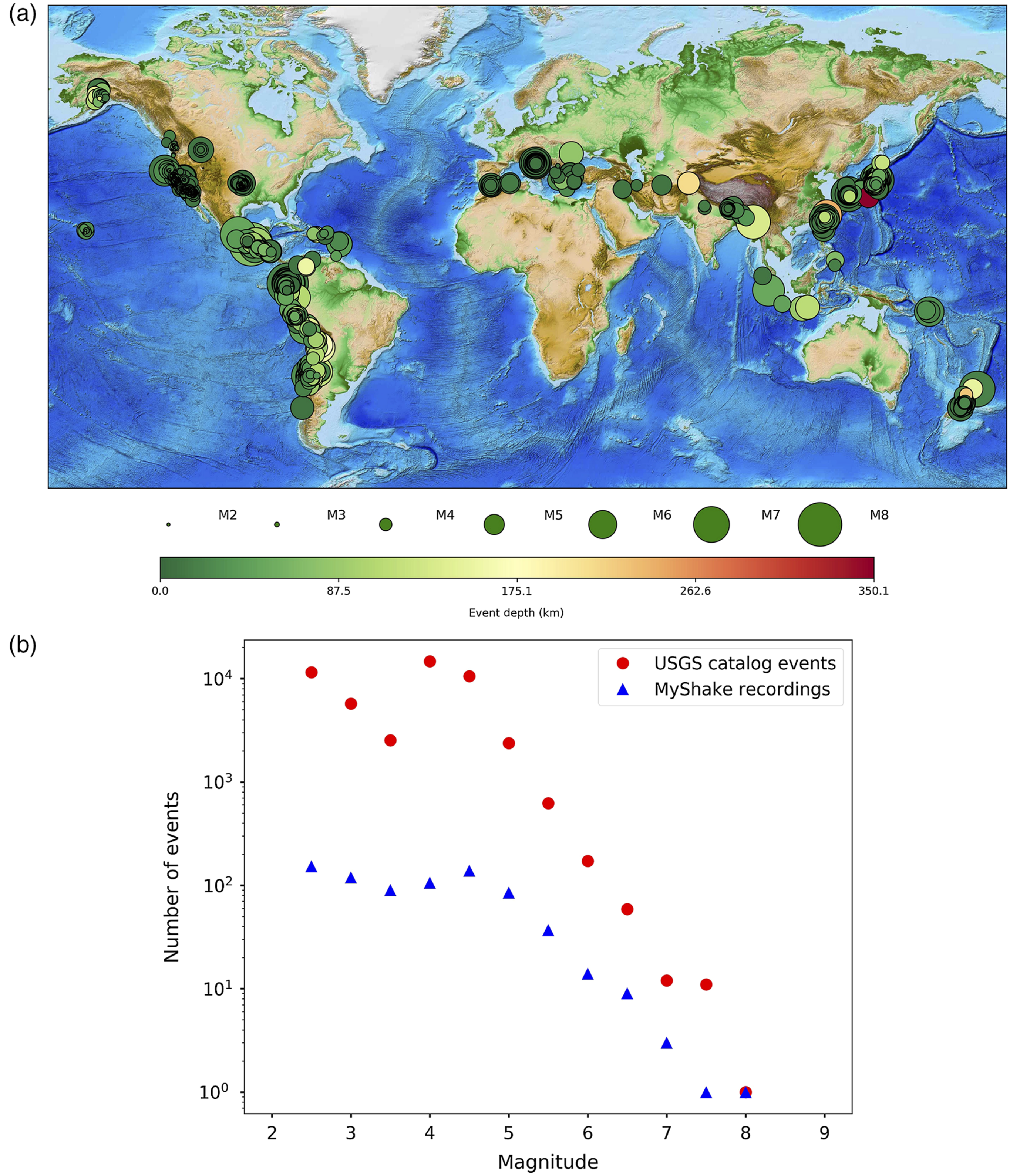
DATA USED

The dataset used in this article comes from global MyShake users. As described in detail by Kong, Allen, Schreier, *et al.* (2016), the MyShake application has a two-stage triggering algorithm. In the first stage, a simple short-term average and long-term average algorithm is used to determine when a previous stationary phone moves (Allen, 1978). An artificial neural network (ANN) algorithm is used to determine whether the movement of the phone is likely caused by an earthquake or by other human activities. When the movements satisfy the ANN algorithm and are determined to be earthquake-like motion, the phone records a 5 min segment of three-component acceleration data that is uploaded to the MyShake servers when the phone is next connected to Wi-Fi and power. An earthquake waveform database is then created from the uploaded waveforms. First we scan for "candidate events" in the USGS Comprehensive Earthquake Catalog (ComCat). For each candidate event, we search the MyShake waveform archive for records within a predefined spatiotemporal window for possible earthquake recordings. Waveforms that meet the requirements of the aforementioned spatial–temporal window are reviewed by a seismologist to filter out those caused by human activities and to remove any bad data (e.g., missing blocks of data). Waveforms that pass all the checks are put into the earthquake waveform database. In the first 2 yr, 757 earthquakes have had at least one recording from a MyShake user.

Figure 1a shows the location of earthquakes for which one or more seismic waveforms (confirmed by a seismologist) were uploaded from MyShake phones. Figure 1b shows the magnitude–frequency relationship of events recorded by MyShake and a comparison with the USGS ComCat. For all the magnitude bins, MyShake records fewer events than the traditional seismic networks. The gap becomes smaller for larger magnitude events, indicating that with increasing magnitude, MyShake's capability to detect events improves, not surprisingly. Waveforms uploaded to our server are three-component acceleration waveforms in 5 min segments (1 min before the trigger and 4 min after), sampled at about 25 Hz. Figure 2 shows epicentral distances of the earthquake waveforms recorded by MyShake users for earthquakes of various magnitudes. As the magnitude of the earthquakes increases, the distance from which smartphones can record useful waveforms also increases. To understand at what range we expect MyShake phones to record earthquake waveforms, we fit an analytic expression to the farthest recordings in each magnitude bin using least-squares regression. For earthquakes of **M** 2.5 to **M** 7.1, using 1000 bootstrap resampling with replacement, we derived the following relationship between the magnitude of the earthquake and the distance in kilometers from which we expect to see recordings from current MyShake users (parameters shown here are the median value of the 1000 bootstrap resampling process):

$$R = 114e^{0.283M} - 216 \ (2.5 \le M \le 7.1), \qquad (1)$$

in which M is the magnitude and R is the epicentral distance for the earthquake. The equation is shown as the red curve in Figure 2, and the light-gray curves are the result of the 1000 bootstrap resampling process. The small inset in Figure 2



▲ **Figure 1.** (a) Earthquakes with one or more useful waveform recordings from MyShake phones in the first two years of operation (12 February 2016 to 12 February 2018). The size of the circle and colors represent magnitude and depth of the earthquake (both magnitudes and locations are from the U.S. Geological Survey [USGS] Comprehensive Earthquake Catalog [ComCat]). (b) Earthquake magnitude–frequency relation for earthquakes detected by MyShake (blue triangles) and in the USGS catalog (red circles). The number of events is measured in 0.5 magnitude bins. The color version of this figure is available only in the electronic edition.

shows the cumulative distribution of the signal-to-noise ratio (SNR) for all the earthquake waveforms from MyShake users. The SNR is calculated on the Y-component by selecting a 2 s window of signal centered on the peak observed ground acceleration (PGA) value and a 2 s window of noise from the beginning of the waveform before the seismic trigger. Using the 2 s window for the signal and noise, we calculate the square of the root mean square amplitude and take the ratio. The 25th, 50th, and 75th percentiles of the SNR are 6.1, 14.6, and 50.9, respectively, for the MyShake recorded earthquake waveforms.

Figure 3 shows examples of six three-component waveforms recorded around the globe by MyShake phones. We show examples from regions that do not have dense seismic networks (additional waveforms are available in the Ⓔ supplemental content). Whereas the X and Y components of the acceleration records are parallel to the short and long directions of the phone screen, the Z component is the direction perpendicular to the phone screen. On most of these waveforms, we can see clear *P* and *S* waves. Because we also archive 1 min of data before a phone trigger, even when the phone triggers on the *S* phase (because of low SNR for the *P* wave), we can often still observe a *P*-wave arrival time.

TIMING AND LOCATION ACCURACY OF SMARTPHONE RECORDS

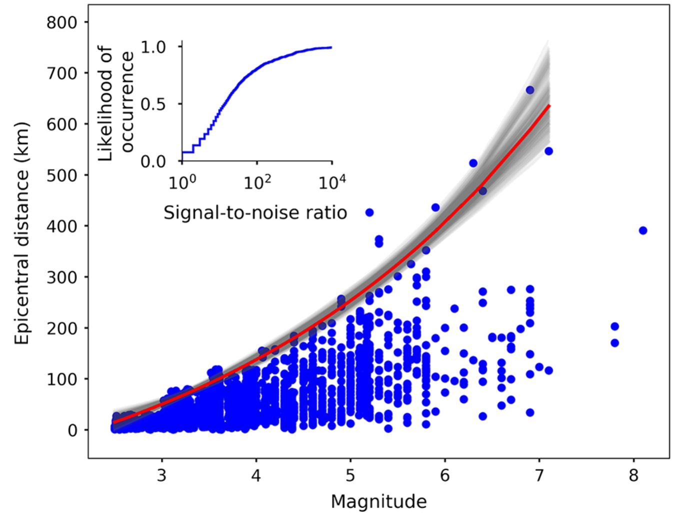
The accuracy of the absolute time and location associated with seismic waveforms is central to the applications and research for which the data can be used. Unlike traditional seismic stations, MyShake sensors are moving around and they do not have continuous Global Positioning System (GPS)–based time. We therefore must develop strategies to improve on the normal timing and location information provided by the phone operating system to the MyShake app.

The internal phone time is insufficiently accurate as a time stamp for seismological applications of the data because it can be off by many seconds (Table 1). Instead, the MyShake app uses network time protocol (NTP) to check in with a remote server and obtain the accurate absolute time. During normal operation, the MyShake app requests an NTP time stamp every hour, and the system stores this information on the backend. The statistics collected include the server time of each query, the roundtrip time for the phone's NTP query to reach and return from the server, and the offset in milliseconds of the server time relative to the phone's internal time when the query was initiated. The total actual offset between the phone's internal time and true NTP time (true offset) was calculated as the recorded offset minus half of the roundtrip query time. We assume that it takes half of the roundtrip time for the query to reach the server.

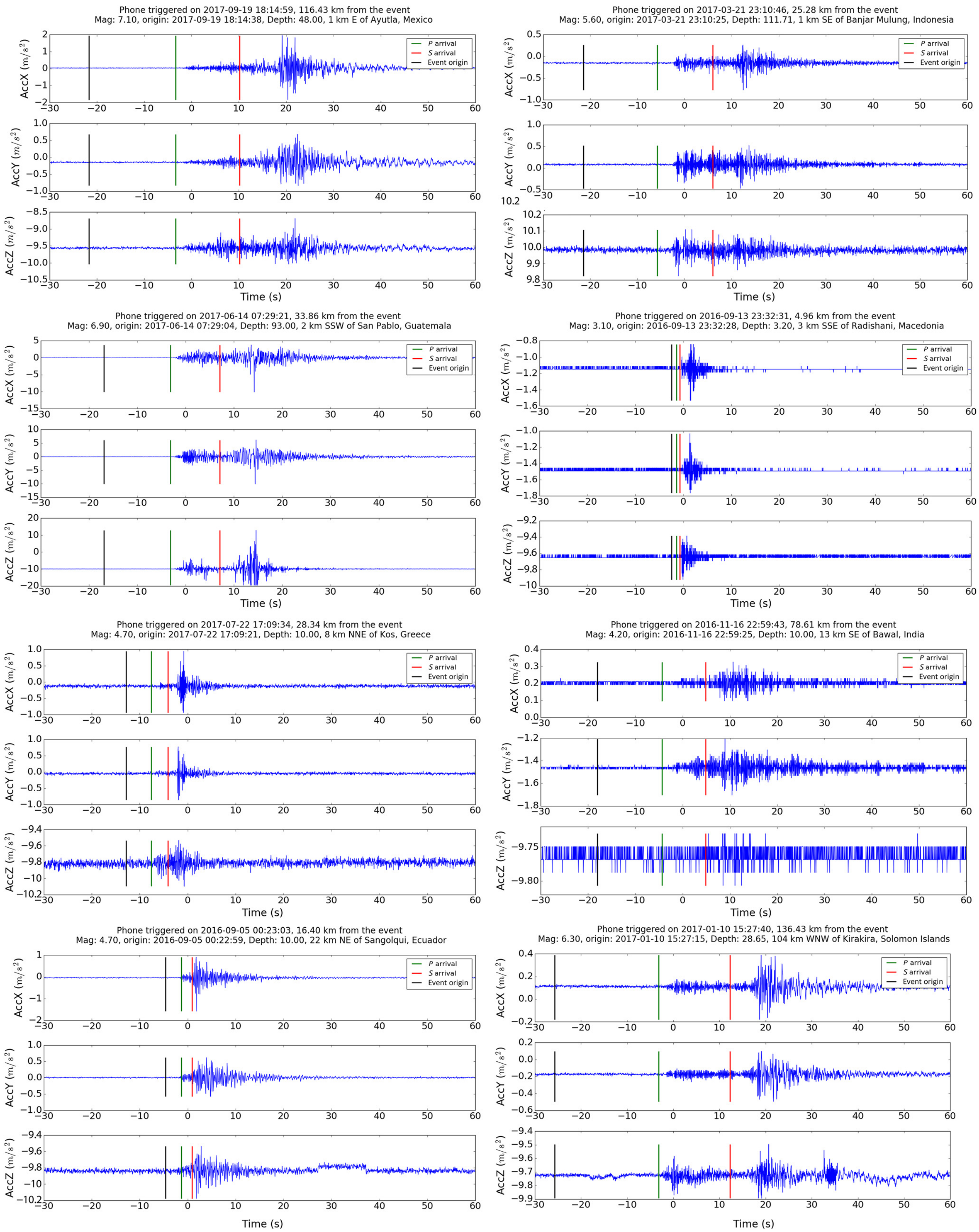
We randomly selected 26 days from our database over a 12-month period beginning in August 2016. These 26 days encompass 6.4 million usable records of a successful NTP query by a phone running MyShake. Table 1 shows the distributions of these time observations. Fifty percent of queries reported an internal time better than 0.723 s, and 75% were better than 1.59 s (true offset). The roundtrip time of an NTP time query is typically an order of magnitude smaller than this; 50% are better than 0.081 s, and 75% are better than 0.131 s.

MyShake does not attempt to correct the internal time of the phone. Instead we use the NTP times to correct the absolute timestamps associated with trigger messages and any recorded waveforms, that is, we apply the calculated true offset time to the internal phone timestamp for the records we collect. Therefore, the accuracy of the timestamp associated with a MyShake waveform is determined by how much drift there has been to the internal phone time between the last NTP query and the time the phone triggers and records a waveform. To assess the accuracy of these timestamps, we calculate the change in the true offset values from one NTP query to the next. Figure 4 shows the distribution of the change in true offset values reported by phones during the 26 days sampled. Fifty percent of the cases have a change in true offset less than 0.027 s, 75% are better than 0.132 s, and 90% are better than 0.503 s (Table 1). The magnitude of the changes increases—and therefore the accuracy of the timestamps decreases—as the time elapsed between NTP queries increases. If queries occur 1–10 min apart, the offset changes by <0.015 s 50% of the time; this metric becomes <0.029 when queries are ~30 min apart and <0.064 s for ~hourly queries.

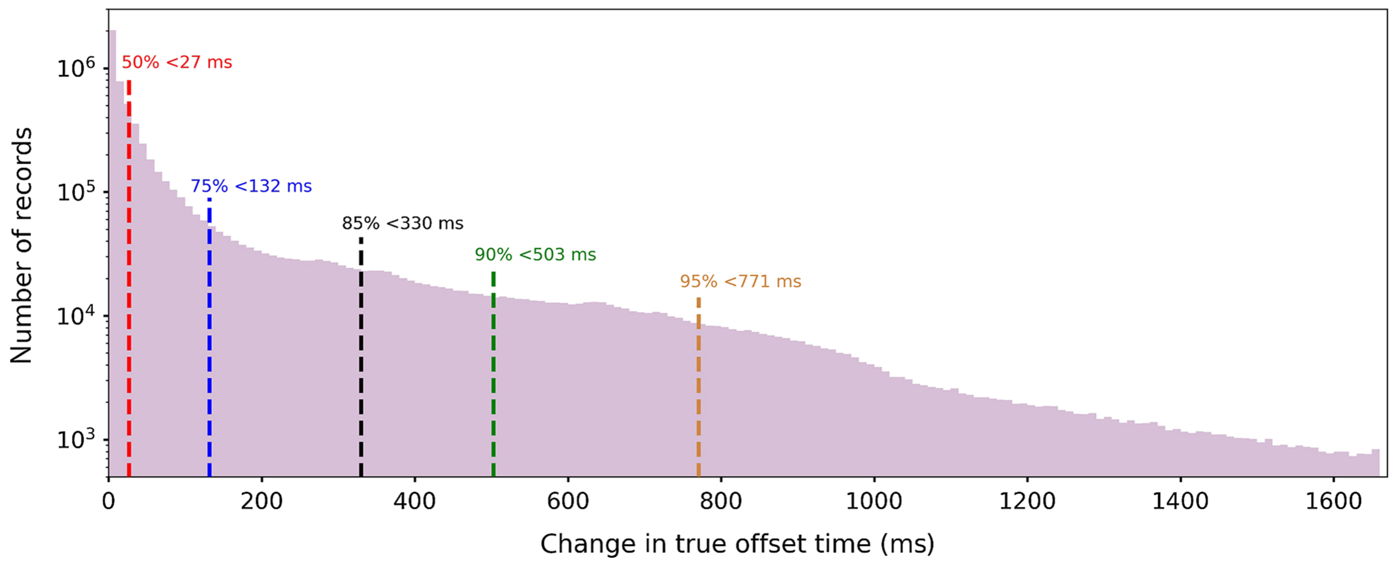
To provide location information, MyShake requests users' permission to collect GPS locations from participating phones. To assess the accuracy of smartphone GPS systems in the



▲ **Figure 2.** Epicentral distance of all MyShake earthquake waveform recordings as a function of magnitude (blue dots). The red curve is equation (1), which approximates the maximum distance to which MyShake phones can trigger and detect earthquakes; the shaded lines are 1000 bootstrap with replacement approximations. We only searched for waveforms corresponding to earthquakes M ≥ 2.5 in the USGS catalog. The inset on the top left is the cumulative distribution of the signal-to-noise ratio (SNR) for all the earthquake recordings measured on one horizontal component (Y component). The color version of this figure is available only in the electronic edition.



▲ **Figure 3.** Example three-component acceleration waveforms from MyShake detections globally. The black line is the event origin time from the USGS catalog; green and red lines are estimated *P* and *S* arrival time using ak135 (Kennett *et al.*, 1995). The zero time on each panel is the time when the phone triggers. E, east; NE, northeast; NNE, north-northeast; SE, southeast; SSE, south-southeast; SSW, south-southwest; WNW, west-northwest. The color version of this figure is available only in the electronic edition.



▲ **Figure 4.** Accuracy of MyShake waveform and trigger timestamps. The histogram shows the change in phone timing offsets extracted from the network time protocol synchronization process. Data are from 26 random days between August 2016 to August 2017. The dashed lines mark the 50th, 75th, 85th, 90th, and 95th percentiles of the distribution. Note that the vertical scale is logarithmic. The color version of this figure is available only in the electronic edition.

MyShake use case, we conducted a test approximating typical conditions for a stationary phone monitoring for an earthquake signal. In a six story building, 10 smartphones were placed on a second floor windowsill facing into a partially sheltered courtyard with a limited view of the sky. We elected to use a windowsill (rather than a location more interior to the building) so we could determine the accurate true location of the phones by determining the location of the side of the building in Google Earth. The phones were left stationary so they could enter into steady mode before being manually prompted to trigger, causing a seismic waveform to be recorded and the location of the phone and waveform reported as it is in normal MyShake operation. Figure 5 shows the distribution of horizontal and vertical (elevation) errors based on 98 triggers, with significant percentiles tabulated in Table 1.

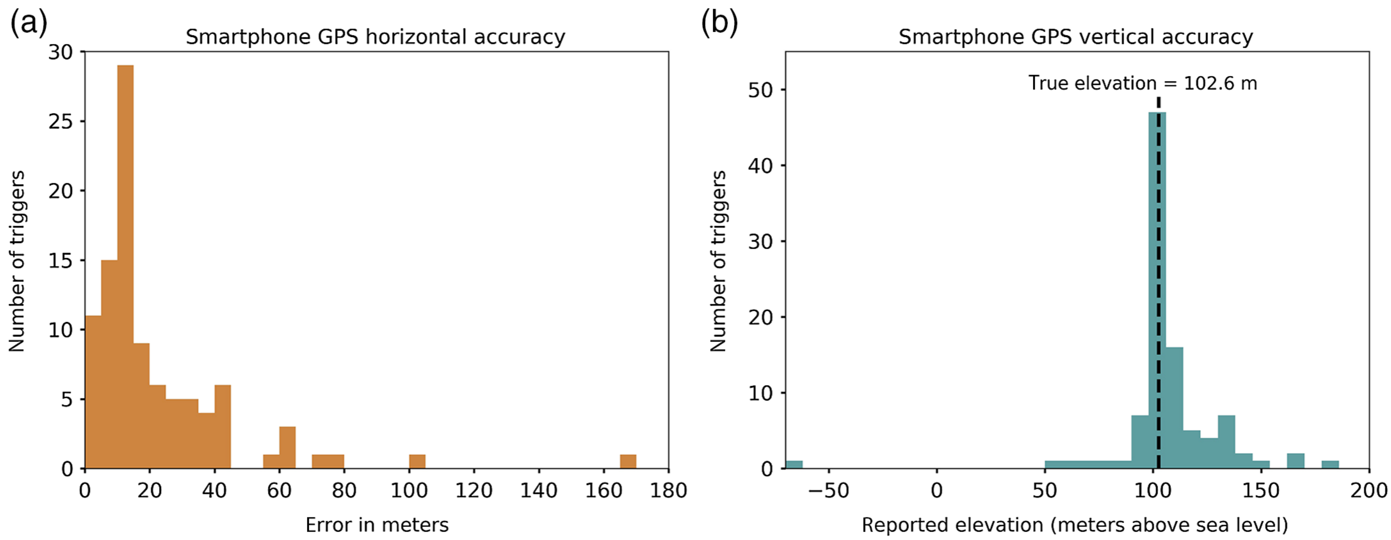
For 50% of triggers, the horizontal location is within 14 m of the true location. For reference, the reported accuracy of basic C/A-code positioning with good sky view is ~5–10 m. It is within 28.8 m 75% of the time in our test and within 43.6 m 90% of the time (Table 1). Whereas a typical home has a footprint 10 m across, office buildings might have a 50 m footprint. This means that the location information is of the same order as the size of buildings, and there is the potential to group waveforms by building or building type to both observe and correct for building amplification factors. The on-phone API providing trigger location information also provides a horizontal accuracy metric. As can be seen in Table 1, our observed errors are consistent with the reported accuracy.

The elevation is within 3.9 m of the true value 50% of the time. The typical floor spacing of a multistory building is ~4 m. Therefore, our results suggest that it is possible to estimate, within plus or minus one story, on which floor a phone was located when it recorded a waveform 50% of the time; 90% of the time, the error is within 34.1 m. This is equivalent to ~8.5 floors. This error is still small enough to allow for a qualitative estimate of whether a phone is located near the bottom, middle, or top of a tall skyscraper building. Such an estimation is useful in identifying cases in which the free oscillations of a

Table 1
Summary of the Location and Timing Accuracy for MyShake Triggers and Waveforms

Measurements	50th Percentile	75th Percentile	90th Percentile
Offset of internal phone clock (s)	0.723	1.59	20.8
NTP query roundtrip time length (s)	0.081	0.131	0.251
Accuracy of waveform and trigger timestamps (s)	0.027	0.132	0.503
Measured horizontal location error (m)	14.0	28.8	43.6
Reported horizontal location accuracy (m)	19.5	20.8	45.3
Measured elevation error (m)	3.9	11.4	34.1

NTP, network time protocol.



▲ **Figure 5.** Accuracy of smartphone location using Global Positioning System (GPS) points reported with MyShake triggers and seismic waveforms, both (a) horizontal and (b) vertical. Ten phones were placed on a second-floor windowsill facing into a partially sheltered courtyard and periodically prompted to collect a spontaneous trigger and record a waveform. The resulting 98 GPS points cluster closely around the true location, both horizontally and vertically. The color version of this figure is available only in the electronic edition.

building produce an amplification effect to the signals MyShake records. The distributions of the horizontal and vertical location errors are shown in Figure 5.

ESTIMATING EARTHQUAKE SOURCE PARAMETERS

We use the earthquake waveforms recorded by MyShake to locate events and estimate the magnitude. In this section, we focus on the accuracy of source parameter estimation for events that have seismic waveforms with good azimuthal coverage, that is when the largest azimuthal gap between stations is <180°. This is the case for 21 events in our dataset. The results of location and magnitude estimation for all events for which we have four or more seismic phases (regardless of the azimuthal coverage) detected are shown in the Ⓔ supplemental content. The locations of these events are also shown in the Ⓔ supplemental content.

METHODS

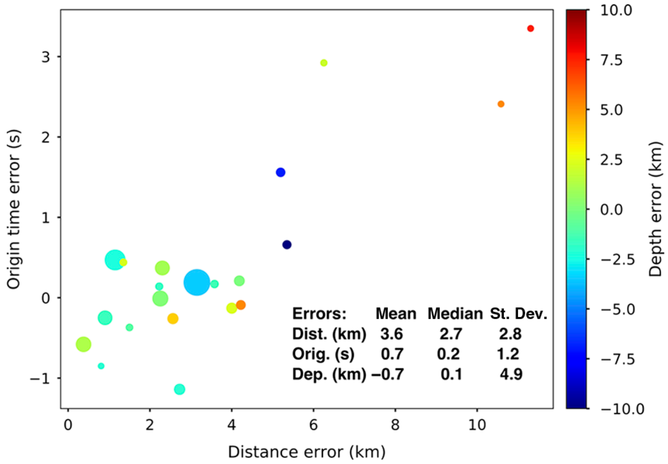
First, we manually pick the *P* and *S* phases. We use Hypoinverse (Klein, 2002) to determine the location, depth, and origin time of the earthquake. Hypoinverse requires an initial location and origin time as the input. For this test, we use the geometric mean of the triggers as the initial location, and the initial origin time is set to 5 s before the first trigger time. We tested a total of 28 homogeneous velocity layer models within Hypoinverse and found that the following model in Table 2 yields good results for most of the events we test. We can only estimate the location and magnitude when there are four or more phase picks (either *P* or *S* or mixed) available.

To estimate the magnitude, we apply the M_L relationship of Bakun and Joyner (1984). This relation estimates local magnitude from the distance, the peak to peak amplitude, and the time span from peak to peak. As shown in Kong, Allen, and Schreier (2016), MyShake recordings typically have larger amplitude than free-field stations at the same epicentral

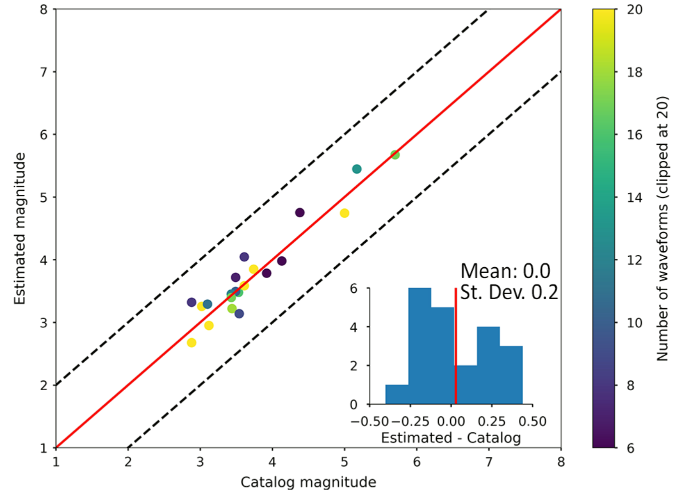
Table 2
Velocity Model Used in the Estimation of the Location, Origin Time, and Depth of the Earthquake Using the Manually Picked *P*-Wave and *S*-Wave Arrivals

Top Depth of the Layer (km)	*P* Velocity (km/s)	*S* Velocity (km/s)
0.0	3.57	2.04
1.5	5.35	3.06
5.1	5.83	3.33
15.0	6.86	3.92
29.0	7.95	4.54

Except for the 3 September M 5.8 2016 Oklahoma event, for which we use the velocity model described in Grandin *et al.* (2017).



▲ **Figure 6.** The distance, origin time, and depth error for the 21 events that have good azimuthal coverage and at least four phase pickings. The mean, median, and standard deviation (st. dev.) are also shown for all the errors. The sizes of the circles indicate how many phase pickings were available for each event, which ranged from 4 to 66. See the errors for all the 44 events that have four or more phase pickings in Ⓔ Figure S1. The color version of this figure is available only in the electronic edition.



▲ **Figure 7.** The magnitude estimates from MyShake compared with those from the USGS ComCat. The mean error is 0.0, and the st. dev. is 0.2. The color of the circle shows the number of waveforms used. The solid red line is the 1-to-1 line, and two black dashed lines show an error of 1 magnitude unit. The magnitude estimates for all the 44 events that have four or more phase pickings are shown in Ⓔ Figure S2. The color version of this figure is available only in the electronic edition.

distance. But the characteristics of the ground motion in terms of amplitude and frequency are preserved well compared with a nearby seismic station. In addition, shake table tests (Dashti *et al.*, 2012; Kong, Allen, Schreier, *et al.*, 2016) show that the sensors inside the smartphones can recover the ground motion very well. Therefore, we scale the PGA amplitude of the MyShake recordings by dividing the observed amplitude by a factor of 1.6. This factor is derived from all the MyShake recordings.

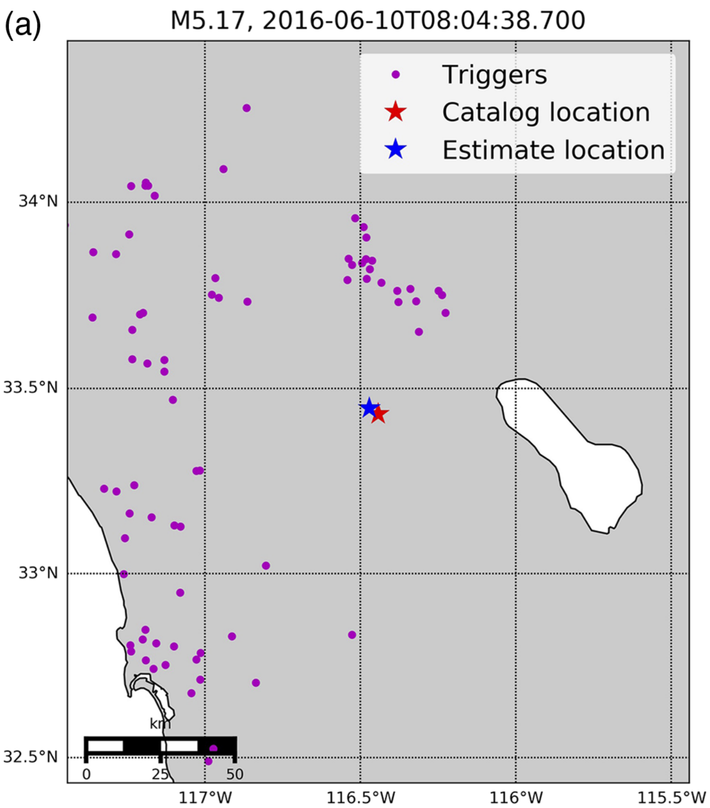
Figure 6 shows the results for the 21 events that have good azimuthal coverage (maximum gap in azimuthal coverage is <180°). The median errors in the location measured as distance from the USGS catalog location in the origin time and the depth are 2.7 km, 0.2 s, and 0.1 km, respectively, with standard deviations of 2.8 km, 1.2 s, and 4.9 km. Most events have distance errors that are <5 km. The larger errors are typically for events with only a small number of phase picks. In this case, three of the five events with distance errors >5 km have only four picks. The MyShake magnitude estimates from these events are compared with USGS catalog magnitudes in Figure 7. The mean and standard deviation of the magnitude error are 0.0 and 0.2, respectively. All these events have the errors <0.5 units. See Ⓔ Figures S1 and S2 for the errors in location, origin time, and magnitude of all 44 events, including the ones have limited coverages in the supplementary content.

We illustrate MyShake source parameter estimation with four events from California, Oklahoma, and Morocco (errors are shown in Table 3). The 10 June M 5.2 2016 Borrego Springs event and the 4 January M 4.4 2018 Berkeley event occurred in locations where a good number of MyShake users were located nearby; both events have maximum azimuthal gap <60° (Fig. 8). The distance, origin time, depth, and magnitude errors for the Borrego Spring event are 3.1 km, 0.19 s, −3.5 km, and 0.28 (estimated−catalog), respectively. Likewise, for the Berkeley event, they are 1.1 km, 0.47 s, −2.4 km, and 0.37, respectively.

Figure 9 shows the results for the 3 September M 5.8 2016 Pawnee, Oklahoma, event and the 15 March M 5.6 2016 event offshore of Morocco. Neither event has phones nearby; all recorded waveforms are at a distance of ≥80 km from the epicenter. The azimuthal gaps in coverage are also large at 150° and 175°. The source parameters are therefore not as good as the ones shown in Figure 8 but are still reasonable. The distance, origin time, depth, and magnitude errors for the Oklahoma event are 7.6 km, 1.74 s, 5.5 km, and −0.38, respectively. For the Morocco event, they are 14.7 km, −0.46 s, 1.0 km, and −0.67, respectively.

DISCUSSION

The use of smartphones as a global seismic network is still a relatively new concept for which we have to determine what the capabilities are, what information this network can provide in addition to the existing more traditional seismic networks, and what challenges the network faces. So far, the successes of the MyShake project include rapid expansion around the globe, rapid increases of instrumentation densification in the aftermath of societally significant earthquakes (both locally and globally), and longevity of the user base as the network has settled into a relatively stable number of 40,000 users. This number is very large compared with the traditional seismic



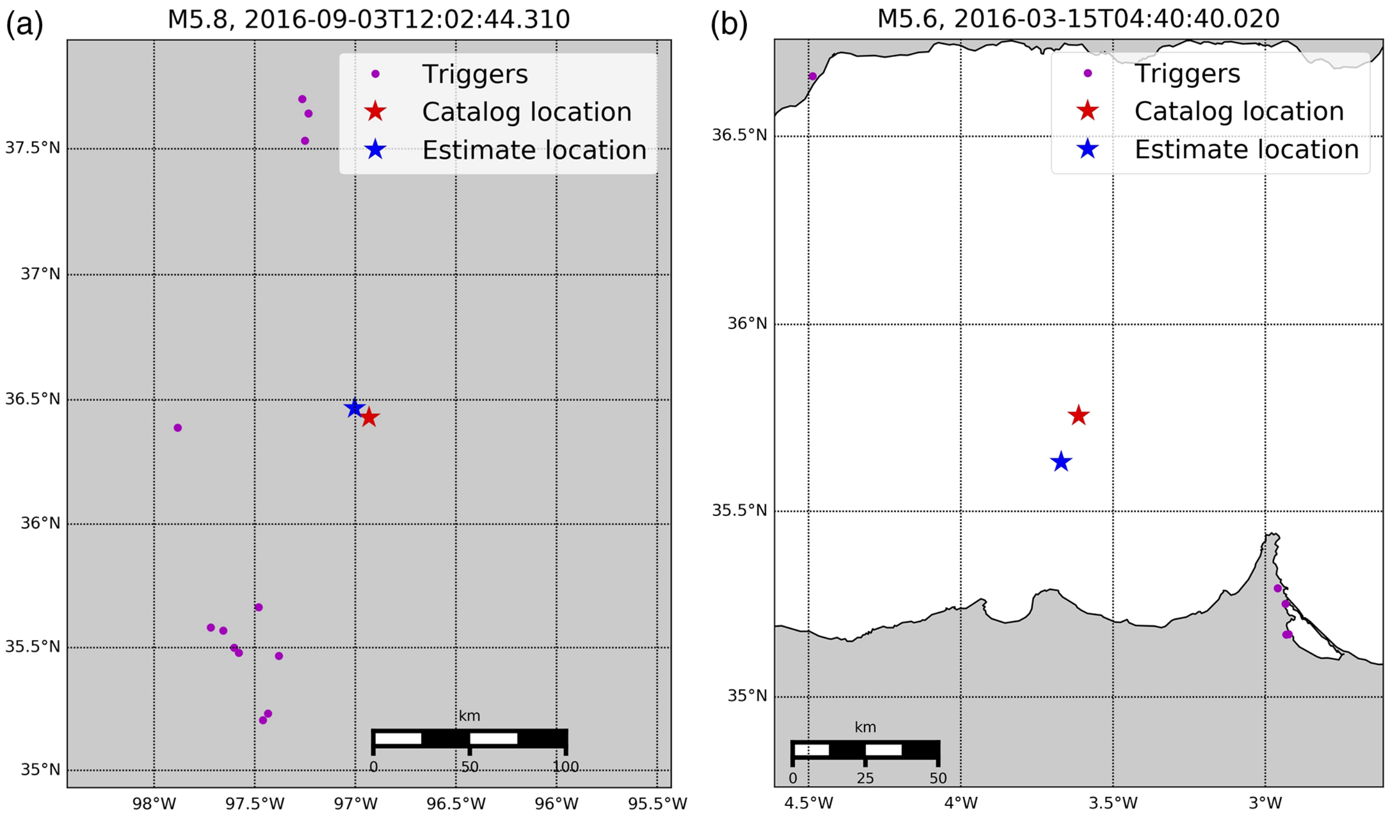
▲ **Figure 8.** Earthquake source parameter estimation. (a) 10 June M 5.2 2016 Borrego Springs event, with maximum azimuthal gap of 56°. (b) 4 January M 4.4 2018 Berkeley event, with maximum azimuthal gap of 17°. The magenta dots are phones that triggered during the earthquake. The estimates MyShake epicentral location is shown as a blue star and the USGS catalog location as a red star. Errors of the estimated source parameters with respect to the catalogs are shown in Table 3. The color version of this figure is available only in the electronic edition.

stations globally, but there are still many challenges to be explored to fully understand the capabilities of this type of network to achieve the routine seismological applications as illustrated in this study. In this discussion, we describe the difficulties and the potential future improvements for this global smartphone seismic network.

First, many elements of the data processing currently still involve human interactions. In our analysis, a seismologist

Table 3
Catalog and Estimation Source Parameters with the Errors for the Four Different Earthquakes

Event Name	Data Source	Origin Time (yyyy/mm/dd hh:mm:ss.sss)	Location	Depth (km)	Magnitude
Borrego Springs, California	Catalog	2016/06/10 08:04:38.700	Latitude: 33.43; Longitude: −116.44	12.3	5.17
	Estimation	2016/06/10 08:04:38.890	Latitude: 33.45; Longitude: −116.47	8.8	5.45
	Error	0.19 s	3.1 km	−3.5	0.28
Berkeley, California	Catalog	2018/01/04 10:39:37.730	Latitude: 37.86; Longitude: −122.26	12.3	4.38
	Estimation	2018/01/04 10:39:38.200	Latitude: 37.84; Longitude: −122.26	9.9	4.75
	Error	0.47 s	1.1 km	−2.4	0.37
Pawnee, Oklahoma	Catalog	2016/09/03 12:02:44.310	Latitude: 36.43; Longitude: −96.93	5.4	5.8
	Estimation	2016/09/03 12:02:46.050	Latitude: 36.47; Longitude: −97.00	10.9	5.42
	Error	1.74 s	7.6 km	5.5	−0.38
Offshore Morocco	Catalog	2016/03/15 04:40:40.020	Latitude: 35.76; Longitude: −3.61	10.0	5.6
	Estimation	2016/03/15 04:40:39.560	Latitude: 35.63; Longitude: −3.67	11.0	4.93
	Error	−0.46 s	14.7 km	1.0	−0.67

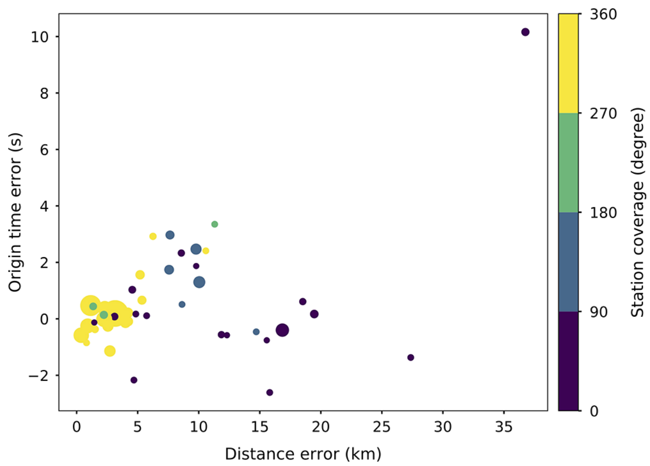


▲ **Figure 9.** Earthquake source parameter estimation. (a) 3 September M5.8 2016 Pawnee, Oklahoma, event, with maximum azimuthal gap of 210°. (b) 15 March M5.6 2016 Morocco event, with maximum azimuthal gap of 185°. Note that this event is beneath the Mediterranean Sea offshore of Morocco. The magenta dots are phones that triggered during the earthquake. The estimates MyShake epicentral location is shown as a blue star and the USGS catalog location as a red star. Errors of the estimated source parameters with respect to the catalogs are shown in Table 3. The color version of this figure is available only in the electronic edition.

needs to review the waveforms to confirm that they are useful earthquake records with a relatively high SNR, to remove the waveforms with data issues (e.g., missing data, spikes), and to manually pick the *P* and *S*-wave arrivals. To build a fully functioning global seismic network, all these steps need to be automated. The very different characteristics of the network compared with traditional networks means that a new suite of processing software is needed.

Second, there are unique challenges to use the MyShake data. The quality of the waveforms is different in ways that are difficult to determine (Kong, Lv, and Allen, 2019). Two different phones at the same location can have very different characteristics depending on their exact physical location, and knowing how to weight the quality of the waveforms is a challenge. There are many events for which we only detect *S*-wave arrivals, that is, in the case when there are many phones but only at larger distances (~100 km or more). Therefore, developing a robust phase picking algorithm for this noisy dataset is important. In addition, phones may be located in different types of buildings, on different floors, and in places where the amplification (i.e., site response) is very different. The response of the buildings, desks and furniture, and so on are all factors, meaning we will require the use of many phones to aggregate the results in an average sense. Further calibration of the amplitude of MyShake recordings against the traditional seismic stations will be helpful. Although the data shown in this study illustrated that MyShake data can be used to detect and characterize earthquakes, improving the location, origin time, and magnitude estimates will require us to better understand these aspects of waveform data quality.

Third, the citizen science nature of the MyShake project brings inherent challenges to seismic network operation. The configuration and density of the smartphone network is constantly changing as users join the network and leave, and individual phones move around during the day. The detection capability is much greater at night than during the day because more phones are steady during the night. MyShake users will tend to be clustered in large cities where populations are concentrated. Therefore, earthquakes in rural areas or far away from population centers can be missed. When earthquakes are detected away from population centers, the station covered has only a narrow azimuthal coverage. This is similar to the situation for regional seismic networks when the earthquake is outside the seismic network. The effect is illustrated in Figure 10, which shows the errors in location and origin time with the azimuthal coverage and number of phases used in



▲ **Figure 10.** The distance, origin time, and station coverage from Myshake network using all the 44 events that have four or more identifiable phases without any coverage filtering. The sizes of the circles indicate how many phase pickings were available for each event, which ranged from 4 to 66. The colors of the circles showing the station coverage for each event. The color version of this figure is available only in the electronic edition.

analysis for all the events that has at least four identifiable phases or more to conduct the calculation. When the azimuthal coverage is limited, there are larger errors. For this type of earthquake, location estimation from traditional regional seismic networks is also poor. One possible solution worth exploring is combining information from traditional seismic stations or complementing the network with permanent low-cost sensors (Cochran *et al.*, 2009; Clayton *et al.*, 2012; Wu, 2015; Nof *et al.*, 2019).

Although the citizen science nature of this project brings with it the limitations described previously, it also brings the possibility of rapid network expansion because there are now almost three billion smartphones in use in the world. To harness a larger fraction of these accelerometers in the future, MyShake has developed a completely redesigned version of the app using a human-centered design methodology (Rochford *et al.*, 2018). The new app brings additional user functionality that we hope will significantly expand the number of users providing data for more earthquakes and further study of the scientific opportunities that MyShake could support.

CONCLUSION

MyShake has now been operating as a global smartphone seismic network for several years. Although there is constant turnover in the people who are running MyShake on their phones, the number of users at any given time has settled to about 40,000 globally. The network has recorded useful seismic waveforms for hundreds of earthquakes with magnitudes from <2 up to M 7.8 and from the surface to 350 km depth. The seismic waveforms recorded by MyShake users show that the MyShake phones could trigger on and detect M3 earthquakes out to 50 km, M5 out to 250 km, and M7 out to 500 km.

The accuracy of the location and timing information determines the scientific uses of the data; 50% of the sampled global dataset has timestamps with accuracies better than 0.03 s. Based on a limited test of phones inside a building, 50% of locations are within 14 m horizontally and 4 m vertically. This makes it possible to estimate in which building and what floor a record is being recorded.

Applying standard regional seismic network techniques to the MyShake data, we can determine the ability of MyShake to characterize events. Using a set of 21 events for which there are four or more *P*- or *S*-wave arrivals and a maximum azimuthal gap <180°, we find that the median location, depth, and origin time error to be 2.7 km, 0.1 km, and 0.2 s, respectively. The mean and standard deviation of the magnitude error are 0.0 and 0.2, respectively. When earthquakes occur beneath urban regions (where there are many MyShake phones), the location errors are smaller.

These preliminary results suggest the potential of the MyShake network to contribute to the seismology community by providing additional data to detect earthquakes and constrain source characteristics. In particular, in locations where there are few traditional seismic stations but dense populations, MyShake could provide valuable data to constrain earthquake hazards. There are many challenges and limitations to address and overcome, but a network such as MyShake can enhance our ability to better understand earthquakes and hazards globally, as well as to engage the public in locations where these earthquakes occur.

DATA AND RESOURCES

The U.S. Geological Survey (USGS) Comprehensive Earthquake Catalog (ComCat) can be accessed at https://earthquake.usgs.gov/fdsnws/event/1. MyShake data are currently archived at Berkeley Seismology Laboratory and use is constrained by the privacy policy of MyShake (see http://myshake.berkeley.edu/privacy-policy/index.html). All websites were last accessed July 2019.

ACKNOWLEDGMENTS

The Gordon and Betty Moore Foundation fund this analysis through Grant Number GBMF5230 to University of California, Berkeley. We thank the MyShake team members, Roman Baumgaertner, Garner Lee, Arno Puder, Louis Schreier, Stephen Allen, Stephen Thompson, Jennifer Strauss, Kaylin Rochford, Doug Neuhauser, Stephane Zuzlewski, Sarina Patel, and Jennifer Taggart, for keeping this project running and growing. All the analysis of this project is done in Python, particularly the ObsPy package (Beyreuther *et al.*, 2010; Krischer *et al.*, 2015). The Advanced National Seismic System (ANSS) catalog data for this study were accessed through the Northern California Earthquake Data Center (NCEDC), doi: 10.7932/NCEDC. We also thank all the MyShake users who contribute to the project. In the analysis of the data, we thank Taka'aki

Taira for providing helpful discussion about the use of Hypoinverse. We also thank editor Brendan Crowell, reviewer Rémy Bossu, and two anonymous reviewers who made this article better.

REFERENCES

Allen, R. M. (2012). Geophysics. Transforming earthquake detection?, *Science* **335,** no. 6066, 297–298.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68,** no. 5, 1521–1532.

Atkinson, G. M., and D. J. Wald (2007). "Did You Feel It?" intensity data: A surprisingly good measure of earthquake ground motion, *Seismol. Res. Lett.* **78,** no. 3, 362–368.

Bakun, W. H., and W. B. Joyner (1984). The ML scale in Central California, *Bull. Seismol. Soc. Am.* **74,** no. 5, 1827–1843.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81,** no. 3, 530–533.

Bossu, R., S. Gilles, G. Mazet-Roux, F. Roussel, L. Frobert, and L. Kamb (2012). Flash sourcing, or rapid detection and characterization of earthquake effects through website traffic analysis, *Ann. Geophys.* **54,** no. 6, doi: 10.4401/ag-5265.

Bossu, R., M. Laurin, G. Mazet-Roux, F. Roussel, and R. Steed (2015). The importance of smartphones as public earthquake-information tools and tools for the rapid engagement with eyewitnesses: A case study of the 2015 Nepal earthquake sequence, *Seismol. Res. Lett.* **86,** no. 6, 1587–1592.

Bossu, R., R. Steed, F. Roussel, M. Landès, A. Fuenzalida, E. Matrullo, A. Dupont, J. Roch, and L. Fallou (2018). App earthquake detection and automatic mapping of felt area, *Seismol. Res. Lett.* **90,** no. 1, 305–312.

Bürgmann, R., P. A. Rosen, and E. J. Fielding (2000). Synthetic Aperture Radar Interferometry to measure earth's surface topography and its deformation, *Annu. Rev. Earth Planet. Sci.* **28,** no. 1, 169–209.

Chung, A. I., C. Neighbors, A. Belmonte, M. Miller, H. H. Sepulveda, C. Christensen, R. S. Jakka, E. S. Cochran, and J. F. Lawrence (2011). The Quake-Catcher network rapid aftershock mobilization program following the 2010 M 8.8 Maule, Chile earthquake, *Seismol. Res. Lett.* **82,** no. 4, 526–532.

Clayton, R. W., T. Heaton, M. Chandy, A. Krause, M. Kohler, J. Bunn, R. Guy, M. Olson, M. Faulkner, M. Cheng, *et al.* (2012). Community Seismic Network, *Ann. Geophys.* **54,** no. 6, doi: 10.4401/ag-5269.

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.

Cochran, E. S., J. F. Lawrence, C. Christensen, and R. S. Jakka (2009). The Quake-Catcher Network: Citizen science expanding seismic horizons, *Seismol. Res. Lett.* **80,** no. 1, 26–30.

Dashti, S., J. D. Bray, J. Reilly, S. Glaser, and A. Bayen (2012). IShake: The reliability of phones as seismic sensors, *Proc. of the 15th World Conference on Earthquake, Engineering*, Lisbon, Portugal, 24–28 September.

Dashti, S., J. D. Bray, J. Reilly, S. Glaser, A. Bayen, and E. Mari (2014). Evaluating the reliability of phones as seismic monitoring instruments, *Earthq. Spectra* **30,** no. 2, 721–742.

Dou, S., N. Lindsey, A. M. Wagner, T. M. Daley, B. Freifeld, M. Robertson, J. Peterson, C. Ulrich, E. R. Martin, and J. B. Ajo-Franklin (2017). Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study, *Sci. Rep.* **7,** no. 1, 11620.

Earle, P. (2010). Earthquake Twitter, *Nat. Geosci.* **3,** 221.

Earle, P., M. Guy, R. Buckmaster, C. Ostrum, S. Horvath, and A. Vaughan (2010). OMG earthquake! Can Twitter improve earthquake response?, *Seismol. Res. Lett.* **81,** no. 2, 246–251.

Faulkner, M., M. Olson, R. Chandy, J. Krause, K. M. Chandy, and A. Krause (2011). The next big one: Detecting earthquakes and other rare events from community-based sensors, *Proc. of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 13–24.

Finazzi, F. (2016). The earthquake network project: Toward a crowdsourced smartphone-based earthquake early warning system, *Bull. Seismol. Soc. Am.* **106,** no. 3, 1088–1099.

Grandin, R., M. Vallée, and R. Lacassin (2017). Rupture process of the Mw 5.8 Pawnee, Oklahoma, earthquake from Sentinel-1 InSAR and seismological data, *Seismol. Res. Lett.* **88,** no. 4, 994–1004.

Havskov, J., and G. Alguacil (2016). *Instrumentation in Earthquake Seismology*, Springer, Dordrecht, The Netherlands, doi: 10.1007/978-1-4020-2969-1.

Hsieh, C.-Y., W. Yih-Min, C. Tai-Lin, K.-H. Kuo, C. Da-Yi, K.-S. Wang, C. Ya-Ting, C. Wen-Yen, L. Wei-Sen, and S.-H. Ker (2014). Low cost seismic network practical applications for producing quick shaking maps in Taiwan, *TAO* **25,** no. 5, 617.

Inbal, A., Q. Kong, W. Savran, and R. M. Allen (2019). On the feasibility of using the dense MyShake smartphone array for earthquake location, *Seismol. Res. Lett.* **90,** doi: 10.1785/0220180349.

Jan, J. C., H.-H. Huang, Y.-M. Wu, C.-C. Chen, and C.-H. Lin (2018). Near-real-time estimates on earthquake rupture directivity using near-field ground motion data from a dense low-cost seismic network, *Geophys. Res. Lett.* **45,** no. 15, 7496–7503.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995). Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.* **122,** no. 1, 108–124.

Klein, F. W. (2002). User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes, *U.S. Geol. Surv. Open-File Rept. 2002-171*.

Kong, Q., R. M. Allen, M. D. Kohler, T. H. Heaton, and J. Bunn (2018). Structural health monitoring of buildings using smartphone sensors, *Seismol. Res. Lett.* **89,** no. 2A, 594–602.

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.

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, e1501055.

Kong, Q., A. Inbal, R. Allen, and J. Strauss (2018). MyShake: Building a global smartphone earthquake early-warning system, *SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists*, 4867–4871.

Kong, Q., A. Inbal, R. M. Allen, Q. Lv, and A. Puder (2019). Machine learning aspects of the MyShake Global Smartphone Seismic Network, *Seismol. Res. Lett.* **90,** no. 2A, 546–552.

Kong, Q., Y. Kwony, L. Schreierz, S. Allen, R. Allen, and J. Strauss (2015). Smartphone-based networks for earthquake detection, *2015 15th International Conference on Innovations for Community Services (I4CS)*, 1–8.

Kong, Q., Q. Lv, and R. M. Allen (2019). Earthquake early warning and beyond: Systems challenges in smartphone-based seismic network, *Proc. of the 20th International Workshop on Mobile Computing Systems and Applications - HotMobile '19*, Santa Cruz, California, ACM Press, New York, New York, 57–62.

Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015). ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.* **8,** no. 1, 014003.

Larson, K. M. (2009). GPS seismology, *J. Geod.* **83,** no. 3, 227–233.

Lindsey, N. J., E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi, and J. B. Ajo-Franklin (2017). Fiber-optic network observations of earthquake wavefields: Fiber-optic earthquake observations, *Geophys. Res. Lett.* **44,** no. 23, 11,792–11,799.

Nof, R. N., A. I. Chung, H. Rademacher, L. Dengler, and R. M. Allen (2019). MEMS Accelerometer Mini-Array (MAMA): A low-cost implementation for earthquake early warning enhancement, *Earthq. Spectra* **35,** no. 1, 21–38.

Rochford, K., J. A. Strauss, Q. Kong, and R. M. Allen (2018). MyShake: Using human-centered design methods to promote engagement in a smartphone-based global seismic network, *Front Earth Sci. Chin.* **6,** 237.

Sakaki, T., M. Okazaki, and Y. Matsuo (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors, *Proc. of the 19th international conference on World Wide Web*, ACM, 851–860.

Steed, R. J., A. Fuenzalida, R. Bossu, I. Bondár, A. Heinloo, A. Dupont, J. Saul, and A. Strollo (2019). Crowdsourcing triggers rapid, reliable earthquake locations, *Sci. Adv.* **5,** no. 4, eaau9824.

Wald, D. J., and J. W. Dewey (2005). Did You Feel It?: Citizens contribute to earthquake science, *U.S. Geol. Surv. Fact Sheet 2005–3016*, doi: 10.3133/fs20053016.

Wald, D. J., L. A. Wald, J. W. Dewey, V. Quitoriano, and E. Adams (2001). *Did You Feel It? Community-Made Earthquake Shaking Maps*, doi: 10.3133/fs03001.

Wu, Y., W. Liang, H. Mittal, W. Chao, C. Lin, B. Huang, and C. Lin (2016). Performance of a low-cost earthquake early warning system (P-Alert) during the 2016 M_L 6.4 Meinong (Taiwan) earthquake, *Seismol. Res. Lett.* **87,** no. 5, 1050–1059.

Wu, Y.-M. (2015). Progress on development of an earthquake early warning system using low-cost sensors, *Pure Appl. Geophys.* **172,** no. 9, 2343–2351.

Qingkai Kong
Sarina Patel
Asaf Inbal
Richard M. Allen
University of California, Berkeley
209 McCone Hall
Berkeley, California 94720 U.S.A.
kongqk@berkeley.edu

Published Online 21 August 2019