Analysis of Detections by the Earthquake Network App between 2017-12-15 and 2020-01-31

(https://doi.org/10.5880/fidgeo.2021.007)

Robert Steed[1], Rémy Bossu[1,2], Francesco Finazzi[3], István Bondár[4], Laure Fallou[1]

1. *European-Mediterranean Seismological Centre (EMSC), c/o CEA, 91297, Cedex, France*
2. *CEA, DAM, DIF, F-91297 Arpajon, France*
3. *Department of Management, Information and Production Engineering, University of Bergamo, Dalmine, Italy*
4. *3Kövesligethy Radó Seismological,Observatory, Geodetic and Geophysical Institute, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences (MTA CSFK GGI KRSZO), Budapest, Hungary.*

1. Licence

2. Citation

When using the data please cite:

Steed, Robert; Bossu, Remy; Finazzi, Francesco; Bondár, István; Fallou, Laure (2021): Analysis of Detections by the Earthquake Network App between 2017-12-15 and 2020-01-31. V. 0.9. GFZ Data Services. https://doi.org/10.5880/fidgeo.2021.007

The data are supplementary material to:

Bossu, R., Finazzi, F., Steed, R., Fallou, L., & Bondár, I. (2021). “Shaking in 5 Seconds!”—Performance and User Appreciation Assessment of the Earthquake Network Smartphone-Based Public Earthquake Early Warning System. Seismological Research Letters. https://doi.org/10.1785/0220210180

Table of contents

3. Data Description

The 'Earthquake Network' (EQN) is an app which detects earthquakes by creating an ad-hoc network of smartphones' accelerometer sensors and provides early warnings for earthquakes via the same smartphone app. Detections are not due to individual smartphone measurements but due to near-simultaneous trigger signals from clusters of smartphones running the app. Therefore, detections are normally located in the closest populated regions to an earthquake's epicentre.

These datasets compare sets of detections with the earthquake parameters published by seismic institutes in order to analyse the performance of the EQN network.

This dataset and analysis are supplementary material to the article Bossu et al. '"Shaking in 5 seconds!" A Voluntary Smartphone-based Earthquake Early Warning System', 2021.

3.1. Sampling method

The EQN app begins to monitor the smartphone's internal accelerometer when the phone is charging and not systematically moving (e.g., being transported) and 2-15 min after the phone's screen has been inactivated.

The app records the acceleration of the device in batches of data of 3 s at the highest allowed frequency (typically 50-100 Hz) and then calculates the variance. The app monitors for changes in variance compared to each smartphone's background variance using a standard control chart technique (s-type control chart). If a vibration is detected, then a trigger is sent to the EQN server in Europe with the location of the phone found via the Android or iOS Location API. It also sends the maximum acceleration seen during the triggering section of the waveform and the delay caused by the analysis of the waveform in milliseconds. However, the starting time of the vibration within the waveform section is not analysed meaning that the time recorded after each trigger will be 0-3 s after the start of strong motion. Since the EQN app collection windows are not synchronized, if the density of recording EQN apps is high enough, then some smartphones will trigger before the others even if they record identical strong motions. There is an additional delay due to the latency of the internet which is unknown but is estimated to be less than 300 ms and probably less than 100 ms, but occasionally it might be larger than 1 s in less developed countries or remote regions with respect to Europe.

Each smartphone sends on average around 30 triggers signals per day that are not related to earthquakes. Anything that knocks or shakes the phone while it is recording, such as its owner or another app's notification could cause one of these false triggers. So, earthquake detection relies on there being an aggregate of triggers that are localized within the same region. For each trigger arriving at the server, the number of triggers within a 0.3 degree radius (≈33 km) and 30 s of the trigger are counted. The number of active EQN apps is also counted within this region. Detection occurs if the following parameter is large enough:

$$S = N_\varepsilon/\varepsilon\lambda - 1, \quad (S1)$$

where N_ε is the number of triggers seen during a ε=30 second period, and

$$\lambda=\exp(\beta_0+\beta_1 \nu_t), \quad (S2)$$

where v_t is the number of active EQN apps at the moment of detection while β_0 and β_1 are tuning parameters. Detections occur when $S>h$ where h is a threshold that varies by region and is calculated so that, on average, there is only one false detection per year per country (*Finazzi*).

Detections are automatically eliminated if the maximum acceleration recorded is very high for 10% of the triggers. These high accelerations are usually caused by phones vibrating due to notifications from other apps. EQN can also make false detections due to strong thunderstorms and so an online service is used to reject detections coincident with such storms.

Once a detection occurs, any subsequent detection within 300 km and 2 min of the first is automatically suppressed.

3.1. Analytical procedure

One dataset (**D1**) contains 550 detections made by EQN between 2017-12-15 and 2020-01-31 in Chile, USA and Italy. Wherever possible, each detection was associated with an earthquake from the parameter catalogue of each country's seismic institute (CSN for Chile, USGS for USA and INGV for Italy). Associations were carried out automatically but also checked manually. It was chosen to analyse events in Chile, USA, and Italy due to the accuracy and completeness of their catalogues. They also have moderate-high seismicity, with a broad range of magnitudes from small events to occasionally large, devastating earthquakes. Importantly, all three regions operate dense seismological station networks that are able to produce accurate event locations and magnitude estimates. An epicentral location inaccuracy of 15 km translates to a seismic phase arrival time change of 2-3 s which can become important in the case of EQN due to its rapid response. All 3 regions also have dense accelerometer networks whose records were used to validate the EQN triggers.

The other dataset (**D2**) contains 134 detections from around the world that could be associated to earthquakes with magnitude ≥ M5 or magnitude ≥ M4.5 in Italy and the USA. This dataset's purpose is to analyse the degree to which EQN offers early warning of earthquakes to its users. All detections were associated to parameters from the USGS earthquake parameter catalogue for consistency. There are 68 detections that are common to the first dataset, although different earthquake parameters will have been used for some of those detections compared to the first dataset.

3.1.1. Association procedure

For the purposes of the analysis, it was important to associate each EQN detection with earthquakes parameters held in an institute's catalogues of events. The following procedure was used for association:

1. Earthquakes were selected from the catalogue from 250 s before the time of the detection until 4 s afterwards.
2. Earthquakes were selected that are also within the association distance defined by each earthquake's magnitude (see association_criteria_distance_mag.png figure).
3. For each earthquake, the arrival time of the P waves at the EQN detection location was estimated using the ak135 model's speed of 8.04 km/s. The events whose P waves arrive within 90 s before the EQN detection and 10 s after the detection were chosen.
4. If multiple earthquakes remained in the selection, then the earthquake of the largest magnitude was chosen as the associated earthquake.

Out of the 550 detections in dataset **D1**, 539 could be associated with earthquakes from the available catalogues. 519 detections matched to a single earthquake, 14 matched to 2 earthquakes, 6 to 3 detections and 11 matched to no earthquakes. Manual investigation of the 11 unexplained detections led to 4 additional events that could be associated with earthquakes. One was associated to a M3.8 earthquake at an unusually large distance of 350 km, and 2 to small magnitude earthquakes (M1.4 and M1.5 located 2 and 8 km from the detection) located through additional investigation by the Seismological Centre of the University of Chile. The fourth was found to be a secondary detection 800 km from epicentre of the March 1st, 2019 Peru M7.0 earthquake. The false detection rate was ~2%.

All of the 68 EQN detections in dataset **D2** could be associated to earthquakes in the USGS catalogue above or equal to M5 or above M4.5 in Italy and the USA. There were also 3 earthquakes that were detected twice by EQN, normally such duplicate detections are suppressed automatically but all 3 earthquakes were large magnitude events (M7.0, M7.5 and M8.0) that led to EQN making detections at distances far from the epicentres. These 3 duplicate detections have been removed from the dataset for clarity.

3.1.2. Causal phase

It has been found that EQN detections can be triggered by either P or S seismic phases. The EQN detections were split heuristically into being caused by P or S phases using the criteria:

Caused by S if (detection delay w.r.t. S > 0 s) & (detection delay w.r.t. P > 6 s). (S3)

Note that distinguishing between P and S phases is less clear within 50 km of the epicentre since both arrive within a short interval of time. In addition, the EQN detections are triggered by strong motion due to the relative insensitivity of the smartphone accelerometers and the P/S phase arrival does not exactly coincide with the onset of motion strong enough to cause a detection.

3.1.3. Calculation of shaking intensities

Intensity predictive equations (IPE) were used to create some columns in the datasets (**D1** and **D2**) and for the analysis of early warning times presented in the article. An IPE predicts the total felt intensity of shaking with respect to hypocentral distance for a given magnitude of earthquake. For a given delay from the origin time of an earthquake, the distance of the S phase from the epicentre can be calculated using the ak135 model and the intensity of shaking for this distance can then be calculated using the IPE. Alternatively, the distance at which the intensity reaches a certain value can be found and then the time at which the S phase passes this distance can be calculated in order to estimate whether there would be time for a warning to be given to people at this intensity.

To convert between hypocentral and epicentral distance:

$$r^2 = d^2 + 4R(R-d)\sin^2(s/2R) \quad (S4)$$

where r is the distance from the hypocentral, d is the hypocentral depth, R is the Earth's radius and s is the epicentral distance.

For most earthquakes, the IPE from Allen et al. 2012 *(Allen)* were used, this formula is only valid for magnitudes >M5 and so we restricted the analysis accordingly.

If r < 50 km

$$\text{Intensity} = 2.085 + 1.428M + 1.402\ln\sqrt{r^2 + R_m^2} \quad \text{(S5)}$$

else

$$\text{Intensity} = 2.085 + 1.428M + 1.402\ln(\sqrt{r^2 + R_m^2}) + 0.078\ln(r/50), \quad \text{(S6)}$$

where r is the hypocentral distance, M is the magnitude of the earthquake and

$$R_m = -0.209 + 2.042\exp(M-5). \quad \text{(S7)}$$

For the Italian earthquakes, the IPE from Tosi et al. 2015 *(Tosi)* was employed for crustal earthquakes (focal depth between 0 and 40 km):

$$\text{Intensity} = -2.15 \log_{10}(r) + 1.03M + 2.31. \quad \text{(S8)}$$

For the western USA, the IPE from Atkinson et al. 2014 *(Atkinson)* was used:

$$\text{Intensity} = 0.309 + 1.864M - 1.672\log_{10}(\sqrt{r^2 + 14^2}) - 0.00219\sqrt{r^2 + 14^2} + 1.77\max(0, \log_{10}(r/50)) - 0.383M \log_{10}(\sqrt{r^2 + 14^2}). \quad \text{(S9)}$$

3.2. Data processing

The Earthquake catalogues of USGS and INGV were downloaded via FDSN using the obspy library (*Krishner*) while the CSN catalogue was provided upon request. Analysis was carried out using the python programming language (www.python.org). Calculations of the P and S seismic phases used the ak135 model (*Kennett*) and were also carried out by the obspy Python library.

Earthquakes were determined to be offshore using the postGIS spatial and geographic extension (https://postgis.net) for the postgresql database and a map of the world's landmasses obtained from the ArcGIS hub (https://arc-gis-hub-home-arcgishub.hub.arcgis.com/).

The last x columns with the prefix 'sm_' contain the results from an analysis of waveforms recorded by the strong motion seismic station closest to each EQN detection (for the 440 detections where it was possible to retrieve waveform data). This analysis and its results is provided by a separate data repository (*steed*).

4. File description

4.1. File inventory

The dataset consists of a zip-file containing a File inventory

File name	Format	Content
2021-007_Steed-et-al_D1-usa-chl-ita	csv	dataset of EQN detections.
2021-007_Steed-et-al_D2-mag-gt-4.5	csv	results of strong motion comparisons.
2021-007_Steed-et-al_association-criteria-distance-mag	png	part of the association criteria used, showing how association distance increased with earthquake magnitude.

4.2. Description of data tables

4.2.1. 2021-007_Steed-et-al_D1-usa-chl-ita.csv

This contains the original set of 550 EQN detections between 2017-12-15 and 2020-01-31 in Chile, Italy and the USA. The 410 events analysed in this dataset are a subset of the events found in this table. The dataset contains the location and time of each EQN detection and the earthquake parameters used for the analysis, as well as many other parameters and results.

The file is a csv file, using ‘,’ delimitated fields and with the field headers on the first row. There are 52 fields which are described in Table 2.

Column header	unit	Description
peakid		Each EQN detection has a random 7-digit numeric id associated.
det_lat	decimal degrees	Latitude where EQN detection occurred (degrees).
det_lon	decimal degrees	Longitude where EQN detection occurred (degrees).
country	iso3166-alpha-3	3 letter Country code, one of {‘chl’,’usa’,’ita’}.
detectiontime	iso 8601 yyyy-mm-dd hh:mm:ss.ss	Date and time of EQN detection (Iso8166 format) (UTC time zone).
detectiontime_local	iso 8601 yyyy-mm-dd hh:mm:ss+/-hh:mm	Localised date and time of EQN detection.
pytz		Time zone of EQN detection.
nighttime	Bolean	Did EQN detection occur between 23h and 7h local time?
signals	integer	The number of signals that caused the EQN detection to trigger.
actives	integer	The number of active EQN apps within 30 km of the detection location.
felt_reports_green	integer	Number of felt reports collected indicating 'green' within 3 min from detection and within a radius of 300 km.
felt_reports_yellow	integer	Number of felt reports collected indicating 'yellow' within 3 min from detection and within a radius of 300 km.
felt_reports_red	integer	Number of felt reports collected indicating 'red' within 3 min from detection and within a radius of 300 km.
notification_time	iso 8601 yyyy-mm-dd hh:mm:ss.ss	When the notification based on felt reports was sent to the Firebase notification service.
notification_delay_from_detection	s	Delay between detection and the notification.
Intensity_strong	binary	If 1, a second alert for strong earthquakes was sent. The time of this second alert is the same of notification_time since it is based on the felt reports.
cat		Seismic catalogue used for earthquake parameters.
num_eq_matches	integer	Number of potentially associated earthquakes in catalogue.
origintime	iso 8601 yyyy-mm-dd	Date and time of associated earthquake in catalogue (UTC).

	hh:mm:ss.ss	
magtype		Type of magnitude for associated earthquake.
magnitude		Magnitude of associated earthquake in catalogue.
eq_lat	decimal degrees	Latitude of associated earthquake in catalogue.
eq_lon	decimal degrees	Longitude of associated earthquake in catalogue.
depth	km	Depth of associated earthquake in catalogue.
separation	km	Separation of EQN detection from epicentre of associated parameters.
detectiondelay	s	Delay between origin time and EQN detection.
P_at_surface_delay	s	Delay for P wave to reach the Earth's surface
offshore	Boolean	Was earthquake offshore?
dist_shore	km	Distance from epicentre to closest point on the shore.
closest_land_lat	decimal degree	Latitude of point of coast closest to the epicentre.
closest_land_lon	decimal degree	Longitude of point of coast closest to the epicentre.
P_at_coast_delay	s	Delay for P wave to reach the closest point on the coastline (and at the surface).
P_on_land_surface_delay	s	Delay for the P wave to read the surface (or the coast if applicable).
detectiondelay_wrt_P	s	Delay between the P wave arriving at the EQN detection location and the detection time.
detectiondelay_wrt_S	s	Delay between the S wave arriving at the EQN detection location and the detection time.
causal_phase	{P,S}	Whether the EQN detection was estimated to have been caused by the P wave or the S wave.
intensity_at_0ld		The predicted intensity of the earthquake for the location of the S wave at the moment of the EQN detection.
intensity_at_5ld		The predicted intensity of the earthquake at locations with a lead time of 5 s with respect to the S wave at the moment of the EQN detection.
intensity_at_10ld		The predicted intensity of the earthquake at locations with a lead time of 10 s with respect to the S wave at the moment of the EQN detection.
intensity_at_15ld		The predicted intensity of the earthquake at locations with a lead time of 15 s with respect to the S wave at the moment of the EQN detection.
sm_net		The seismic network if strong motion accelerometer data was found nearby.
sm_sta		The station name if strong motion accelerometer data was found nearby.
sm_loc		The separation between the strong motion station and the EQN station (km).
sm_unit		One of {'acc','vel','disp'} where acc is acceleration (m/s/s), vel is velocity (m/s) and disp is displacement (m).
sm_sta_lat	decimal degrees	Latitude of strong motion station.
sm_sta_lon	decimal degrees	Longitude of strong motion station.
sm_sta_elv	m	Elevation of strong motion station.
sm_sep_eqn_sta	km	Separation between EQN detection and strong motion station (km).
sm_strongest_motion	m/s/s	Strongest value recorded by station.
sm_strong_motion_	iso 8601	Time of strongest motion (this already accounts for

time	yyyy-mm-dd hh:mm:ss.ss	sm_dt_correction).
sm_strongest_motion_eqn_delay	s	Delay between strong motion and EQN detection (this already accounts for sm_dt_correction).
sm_dt_correction	s	Time correction due to difference in distances of station and EQN detection from the epicentre. A velocity of 8 km/s is used to convert between distance and time.

4.2.2. 2021-007_Steed-et-al_D2-mag-gt-4.5.csv

This contains 134 detections from around the world that could be associated to earthquakes with magnitude ≥ M5 or magnitude ≥ M4.5 in Italy and the USA. There are 68 detections that are common to the first dataset. All detections were associated to parameters from the USGS earthquake parameter catalogue for consistency.

The file is a csv file, using ‘,’ delimitated fields and with the field headers on the first row. There are 36 fields in the dataset and a description of each field can be found in Table 2. Note that the last 16 fields in Table 2 are not in this dataset.

5. References

Allen, T.I.; Wald, D.J. & Worden, C.B. (2012). Intensity attenuation for active crustal regions. J. Seismol. **16**, 409–433. https://doi.org/10.1007/s10950-012-9278-7

Atkinson, G.M.; Worden C.B. & Wald, D.J. (2014). Intensity prediction equations for North America. Bull. Seism. Soc. Am. **104**, 3084–3093. https://doi.org/10.1785/0120140178

Finazzi, F. & Fassò, A. (2017). A statistical approach to crowdsourced smartphone-based earthquake early warning systems. *Stoch. Env. Res. Risk A*. **31**, 1649–1658. https://doi.org/10.1007/s00477-016-1240-8

Kennett, B.L.N.; Engdahl, E.R. & Buland R. (1995). Constraints on seismic velocities in the Earth from travel times, *Geophys J Int*, **122**, 108-124, https://doi.org/10.1111/j.1365-246X.1995.tb03540.x

Krischer, L.; Megies, T.; Barsch, R.; Beyreuther, M.; Lecocq, T.; Caudron, C., & Wassermann, J. (2015). ObsPy: a bridge for seismology into the scientific Python ecosystem. Computational Science & Discovery, **8**(1), 014003, https://doi.org/10.1088/1749-4699/8/1/014003

Steed, R.; Bossu, R.; Finazzi, F.; Bondár, I. & Fallou, L. (2021). Analysis of Strong Motion Waveforms Near the Locations of Detections by the Earthquake Network App in Chile, the USA and Italy. V. 0.9. GFZ Data Services. https://doi.org/10.5880/fidgeo.2021.002

Tosi, P.; Sbarra, P.; De Rubeis, V. & Ferrari, C.. Macroseismic intensity assessment method for web questionnaires. Seism. Res. Lett. 86, 985–990 (2015). https://doi.org/10.1785/0220150127

The seismic catalogue data was derived from the following institutes:

Universidad De Chile. (2013). Red Sismologica Nacional. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/C1

Albuquerque Seismological Laboratory (ASL)/USGS. (1988). Global Seismograph Network (GSN - IRIS/USGS). International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/IU

INGV Seismological Data Centre. (2006, January 1). Rete Sismica Nazionale (RSN). Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy. https://doi.org/10.13127/sd/x0fxnh7qfy