CrossQuake: A Cross-Correlation Code for Detecting Small Earthquakes in the Frequency Domain

Carlos Ramírez Piña[1], Christian R. Escudero[2], J. A. Hernández-Servín[1](✉),
and Gerardo León Soto[3](✉)

[1] Universidad Autónoma del Estado de México, Toluca, Estado de México, Mexico
joseph_servin@uaemex.mx
[2] Centro Universitario de la Costa, Universidad de Guadalajara, Puerto Vallarta, Jalisco, Mexico
[3] Instituto de Investigaciones en Ciencias de la Tierra, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, Mexico
gleon@umich.mx

Abstract. A cross correlation code to detect seismic waveforms in the frequency domain is presented. A known seismic event (master event) is correlated with data corresponding to a one-day of continuous recording. The master event is shifted along the one-day recording and cross correlation in the frequency domain is computed. Earthquake candidates are selected by defining a threshold in the cross correlation coefficient. Transforming the selected candidates back to the time domain, a candidate seismic signal is obtained. Waveforms are linked to an automatic phase picker, location algorithm, double-difference relocation code, and a coda magnitude calculator. The code is tested on a catalog of 293 master events in western Mexico. The dataset analyzed was a temporal Passcal experiment (CODEX). CODEX array consisted of 22 broadband stations located around the Colima volcanic complex. The code was able to identify 1003 events ranging in coda magnitude from 1.2 to 3.8. The code is now public available in the GitLab repository https://gitlab.com/cramirezp/crossquake.git.

Keywords: seismic signals · cross correlation · automatic detection

1 Introduction

Automatic detection of seismic waveforms is a routine in seismic observatories worldwide. Numerical codes have been successfully implemented in order to identify earthquake signals. The most common algorithm to detect seismic events is the short-time-average/long-time-average trigger (STA/LTA) algorithm [1]. The algorithm continuously estimates the average amplitude of a seismic signal in two windows of different lengths. The short time window (STA) is sensitive to changes in the signal amplitude, while the long time window (LTA) provides

R. Silhavy and P. Silhavy (Eds.): CSOC 2023, LNNS 722, pp. 488–497, 2023.
https://doi.org/10.1007/978-3-031-35311-6_49

information about the background seismic noise. Signal detection is made when the ratio, STA/LTA, exceeds a pre-set value. In order to detect specific seismic signals, the STA/LTA algorithm requires the calibration of parameters, e.g., window length, signal filters, trigger threshold. In this way, the major limitation of the algorithm STA/LTA is the time-consuming process of tunning such parameters for the correct identification of specific seismic signals. This is an issue when identification of small magnitude earthquakes is required.

Since tectonic earthquakes are generated on geological faults, their epicentral distributions are bounded to regions where active faults are present. Geological active faults are the boundary of two crustal blocks moving with respect to each other. Even though tectonic stresses induce motion of the tectonic blocks, their common boundary (the fault itself) often keeps locked by friction. During this process, strain is being accumulated along the fault until a threshold is reached. At this point, the fault yields, and a fraction of the accumulated strain energy is released, giving rise to an earthquake, and the cycle repeats itself. This stick-and-slip behavior of a fault suggests that the source of seismic events is nearly similar when they are nucleated from a nearby region of a fault. It is then expected that earthquakes with close hypocenters within a fault have similar waveforms [2].

Under the assumption that cross correlation of two earthquakes with nearby initial rupture points is high, it is possible looking for hidden events in a large time window [3]. By cross correlating waveforms of a well located earthquake (which we call master event) along a time window, it is possible screening for similar waveforms. This method has been applied successfully in the time domain [3,4], however, the procedure is time consuming. In the frequency domain, cross correlation reduces to an arithmetic multiplication and processing times are significantly reduced.

2 Methodology

It is proposed a methodology to detect waveforms similar to seismic signals previously detected by a seismic network. These known signals are called *master events* or *templates*. The proposed methodology has five steps: Identification of candidates by cross correlation, extraction of candidate events, previous localization of detected events, relocation of located events, and coda magnitude computation.

2.1 Identification of Candidate Events

It is assumed that a catalog of master events and its associated waveforms (templates) are available. A master event catalog may downloaded from available international seismological catalogs or created by manual selection. Waveforms are available from open repositories such as SAGE (Seismological Facility for the Advancement of Geoscience) [5]. It is also assumed that continuous recordings of the seismic stations arranged by day are available.

The method only considers the vertical component of each master event, $u(t)$ and day template, $v(t)$. Each time series $v(t)$ is divided in 19 disconnected time intervals, $v_i(t)$, that cover $v(t)$. A fast Fourier transform is applied to each waveform template, $u(t)$, and each time interval, $v_i(t)$, to obtain the frequency spectra $U(f)$ and $V_i(f)$, respectively. Additionally, there were defined 18 windows covering the intersections of the $v_i(t)$ windows to check if an event is between adjacent windows. In order to reduce processing time, a decimation of the data was performed. It was taken into account only the even values of the time series.

In the time domain, the cross correlation of $u(t)$ and $v_i(t)$, $C_i(t)$, is defined by

$$C_i(t) = u(t) * v_i(t) = \int\limits_{-\infty}^{\infty} u(\tau) v_i(t+\tau)\, d\tau \tag{1}$$

The corresponding cross correlation in the frequency domain, $C_i(f)$, is given by

$$C_i(f) = U^*(f) V_i(f) \tag{2}$$

Each frequency window $V_i(f)$ is cross correlated as a data streaming over the template $U(f)$. Cross correlations greater than an specified threshold are marked.

It was performed a normalization with the Cauchy-Schwarz inequality

$$\left| \int_{-\infty}^{\infty} u^*(\tau) v_i(\tau) d\tau \right|^2 \leq E_u E_u \tag{3}$$

$$\Rightarrow |w(t_0)|^2 = \left| \int_{-\infty}^{\infty} u^*(\tau - t_0) v_i(\tau) d\tau \right|^2 \leq E_v E_u = E_v E_u \tag{4}$$

for t_0 arbitrary. Then, $|w(t)| \leq \sqrt{E_u E_v}$ and the normalized cross correlation of $u(t)$ and $v_i(t)$ is given by

$$z(t) = \frac{u(t) \otimes v_i(t)}{\sqrt{E_u E_v}} \tag{5}$$

Once that similar waveforms to the master events have been identified, the next step is extracting the candidate waveforms in the time domain. Each candidate is written in the time domain in Seismic Analysis Code (SAC) format [6]. In order to detect the onset of each candidate, $v_i(t)$, the arrival time is estimated applying an automatic picker (e.g. APK of SAC tools). If three or more stations have close arrival times, the signal is cut in the three channels for each station and saved in an individual directory for further analysis.

2.2 Extraction of Candidate Events

Candidate events were classified by waveform template and extracted from the time series in three channels. An automatic peaker to mark the onset of the seismic signal and its quality was applied and this information recorded for further analysis. Wavefronts and picking information is recorded in an individual directory by candidate event. This procedure is still semiautomatic, the analyst needs to check whether the wavefronts and pickings are appropriate before processing.

2.3 Previous Location

As a first step, the event parameters (i.e., latitude, longitude, depth, and origin time) of grouped events are estimated using the Hypoinverse software [7]. Hypoinverse reads the timing of the onset of the seismic signal (beginning of the P phase) and the quality of the arrival to estimate the event parameter. It is assumed that a unidimensional model of velocities is known. Additionally, the latitude and longitude of the seismic stations is provided as a separate file. Hypoinverse provides a set of events parameters for a trial depth. CrossQuake code tests trials depths until Hypoinverse returns a minimized error in hypocenter depth. The event parameters that minimizes the uncertainty in the hypocentral depth are taken as previous location.

2.4 Relocation

The results of the previous location are relocated by a double difference algorithm with HypoDD software [8]. HypoDD relocates nearby events comparing their arrival times and correlations coefficients. CrossQuake code generates automatically the input files for HypoDD. One file containing the travel times of pair of earthquakes in their common stations and a second file containing the cross correlations of pairs of earthquakes at common stations.

2.5 Coda Magnitude Computation

Coda magnitude is computed defining the end of the event at certain time, t, where the wave amplitude is a given factor over the noise amplitude. The travel time of the coda waves, t_c, is given by

$$t_c = t - t_O \tag{6}$$

where t_O is te origin time. The beginning of the signal, t_P, is taken at the onset of the initial P phase. Then the coda duration of an earthquake is defined by

$$t_{\text{coda}} = t - t_P \tag{7}$$

Combining (6) and (7), the coda duration is expressed by

$$t_{\text{coda}} = t_c - (t_P - t_O) \tag{8}$$

The coda magnitude scale has the form

$$M_C = a \log(t_{\text{coda}}) + br + c \tag{9}$$

where a, b, c are constants and r is the hypocentral distance [9]. Since coda magnitude computation needs the duration of the event (in seconds), a threshold in the frequency spectrum is defined. Assuming noise power is smaller than the event signal power, a bandpass filter is defined with dynamical corners ranging from 1 to 35 Hz. Each signal the end of the coda is determined by analyzing variations in signals by sections of 2000 points. If the average is near zero, the beginning of that section is identified as the end of the coda.

3 Results

CrossQuake code is applied for detecting small magnitude earthquakes in western Mexico. Western Mexico is a complex tectonic environment involving continental rifting [10,11], fragmentation of the overriding plate, tearing of a subducting slab, and an unusual magmatism [10]. In this region the oceanic Rivera and Cocos plates subduct beneath the continental North American plate along the Middle America Trench (MAT) (Fig. 1 left). The Trans Mexican Volcanic Belt (TMVB) is the volcanic arc associated with the subduction of the Rivera and Cocos plates in central Mexico. The west of the TMVB is characterized by a triple rift junction: the Tepic-Zacoalco, the Chapala, and the Colima rifts (Fig. 1 right). The Tepic-Zacoalco rift is the southern boundary of the TMVB with the Jalisco Block (Fig. 1 right). The Jalisco Block (JB) is bounded in the north by the Tepic-Zacoalco rift and to the east by the Colima rift (Fig. 1 right). The Jalisco and Michoacan blocks are separated by the Colima rift (CR). As a consequence, large stresses are deforming the crust in western Mexico causing earthquakes. This makes the region a good candidate to test the proposed methodology.

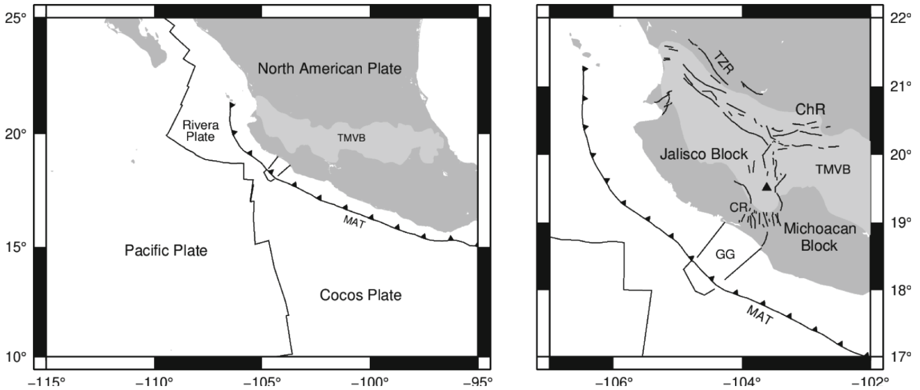

Fig. 1. Left: Tectonic setting for western Mexico. The light gray zone labelled by TMVB stands for Trans Mexican Volcanic Belt. The solid line with triangles is the MAT (Middle-America Trench). Triangles represent the direction of the subduction. Right: Main tectonic features of the area of study. The black lines are the main geological faults in the region. The black triangle represents the Colima volcano TZR = Tepic-Zacoalco Rift, ChR = Chapala Rift, CR = Colima Rift, GG = El Gordo Graben.

The dataset was provided by the temporal experiment Colima Deep Experiment (CODEX) [12]. CODEX experiment consisted of 22 broadband seismic stations around the Colima volcanic complex recording data between 2006 and 2008 at a sampling rate of 100 Hz (see Fig. 2). A seismic catalog with 321 seismic events [13] to extract the master events was used. This catalog was processed by the seismological observatory of the University of Guadalajara. Waveforms of the master events and the continuous recording by day were downloaded from the repositories of IRIS [12].

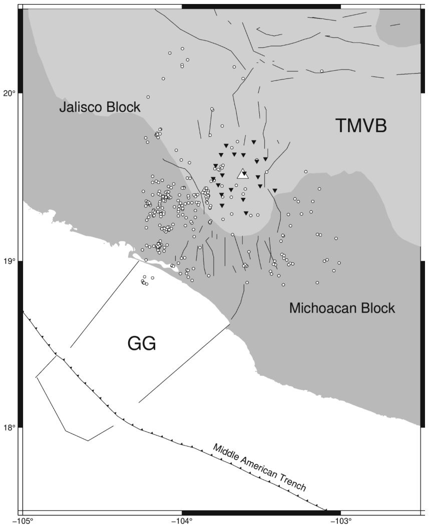

Fig. 2. Map showing the seismic stations of CODEX experiment (inverted black triangles) and the epicenters of master events (open circles). White triangle represents the Colima volcano.

Identification of Candidate Events. In the input parameters of the code, a cross correlation threshold of 0.8 was set as similarity criterium. The results of the code applied to the vertical component of the master events and the recordings of the seismic stations lead to 18253 signals with the specified requirements. 9307 were identified as candidates of seismic signals, and 8946 were identified as noise (see Fig. 3). After grouping similar candidates, a catalog of 907 seismic events was obtained.

Previous Localization. For the previous localization of candidate events, it was assumed the unidimensional velocity model proposed by Gomez *et al.* [14] for western Mexico. Trial depths were varied from 0 to 100 km. The solution which minimized the vertical uncertainty was taken as previous location. 906 events were located successfully. The results are shown in Fig. 4.

Relocation. The previous location was improved applying a double difference relocation algorithm [8]. HypoDD code was used combining catalog data and cross correlation. 692 earthquakes were relocated. Results are shown in Fig. 5.

Fig. 3. Examples of detected signals. (a) Signals classified as seismic waveforms. (b) Signals identified as noise.

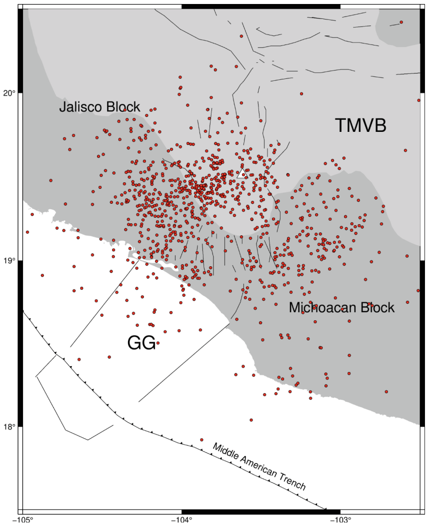

Fig. 4. Results of location process. Epicenters of previous locations are indicated by red points.

Coda Magnitude Computation. Coda magnitude was determined with the coda magnitude parameters reported by Zamora *et al.* [15]:

$$M_c = 1.87 \log_{10}(t_{\text{coda}}) - 0.86 \tag{10}$$

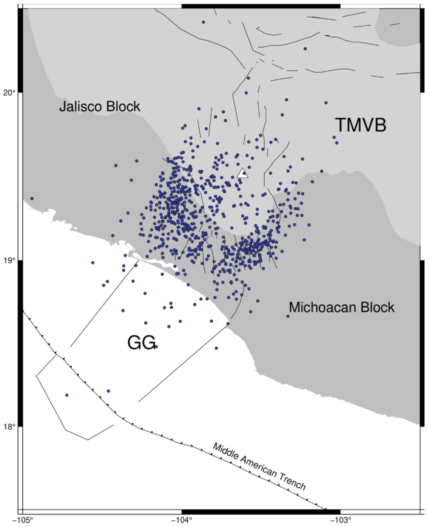

Fig. 5. Results of relocation process. Epicenters of relocated events are indicated by blue points.

Distribution of magnitudes are usually reported by the Gutenberg-Richter relation [16]:

$$\log_{10} N(M) = a - bM \tag{11}$$

where $N(M)$ is the number of earthquakes with magnitude equal or greater than M. Figure 6 shows the Gutengerg-Richter relation for the coda magnitude results. Note that a linear relation of $\log_{10}$ with Mc is clear between magnitudes 2.3–3.0. Lineality of $\log_{10} N(M)$ against M is lost for events smaller than $M_c = 2.3$. This may be interpreted in terms of the resolution of the seismic array instead of a lack of seismicity.

4 Discussion

It is presented CrossQuake, a code to detect seismic events based on data from a given catalog. This catalog can be obtained from a seismological service. Waveforms of the catalog are used as master events to detect similar waveforms in larger time series. Waveforms are available from the SAGE consortium and CrossQuake code is able to download the available data.

CrossQuake code compares daily time series to master events in the frequency domain. Then it extracts candidate earthquake waveforms and save them in individual directories. Given a unidimensional model of velocities, a previous location of the selected waveforms is estimated (Hypoinverse). These locations are refined by a double-difference algorithm (HypoDD). The code also estimates the event parameters provided some extra information. Once the final locations

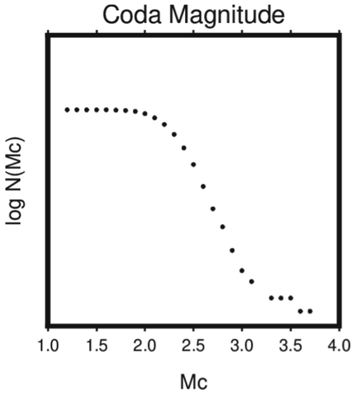

Fig. 6. Gutenberg-Richter relation for coda magnitudes.

are obtained, a coda magnitude calculator is applied to the located earthquakes provided a set of parameters for the coda magnitude.

The main hypothesis is that nearby earthquakes nucleated by the same geological fault have similar waveforms. Then taking a catalog with a number of seismic events, it was generated an expanded catalog.

The code is applied to western Mexico as a case of study. Western Mexico is in an extensional regime where the Jalisco and Michoacan blocks diverge to generate the Colima rift (Fig. 1). This extension causes fracture in the crust as the walls of the rift. The results of the code show that the epicenters of the relocated events are well correlated with these tectonic features (Fig. 5).

The coda magnitude results plotted in Fig. 6 have a linear trend for earthquakes between 2.3 and 3.0 Mc. Although the Gutenberg-Richter relation implies a proportional relation of $\log_1 0N(M)$ against M, deviations of small magnitude earthquakes from lineality are expected because very small earthquakes are still not detected. Deviations from lineality for earthquakes greater than 3.0 Mc may be an indicative of a limited time recording of the network array. CODEX experiment was in operation just for 18 months. We can conclude that the code generates a good Gutenberg-Richter relation for the analized data set.

CrossQuake code and additional information is now public available at the GitLab repository: https://gitlab.com/cramirezp/crossquake.git.

References

1. Allen, R.V.: Automatic earthquake recognition and timing from single traces. Bull. Seismol. Soc. Am. **68**, 1521–1532 (1978)
2. Geller, R.J., Mueller, C.S.: Four similar earthquakes in Central California. Geophys. Res. Lett. **7**(10), 821–824 (1980)
3. Gibbons, S.J., Ringdal, F.: The detection of low magnitude seismic events using array-based waveform correlation. Geophys. J. Int. **165**, 149–166 (2006)

4. Gibbons, S.J.: The optimal correlation detector? Geophys. J. Int. **228**, 355–1365 (2021)
5. Seismological Facility for the Advancement of Geoscience (SAGE). https://www.iris.edu/hq/sage
6. Goldstein, P., Snoke A.: SAC Availability for the IRIS Community, Incorporated Institutions for Seismology Data Management Center Electronic Newsletter (2005)
7. Klein, F.W.: Hypocenter Location Program HYPOINVERSE. US Geological Survey Open File Report 02-171, Version 1.0 (2002)
8. Waldhauser, F., Ellsworth, W.L.: A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California. Bull. Seismol. Soc. Am. **90**, 1353–1368 (2000)
9. Havskov, J., Ottemöller, L.: Routine Data Processing in Earthquake Seismology. Springer, Heidelberg (2010). https://doi.org/10.1007/978-90-481-8697-6
10. Luhr, J.F., Nelson, S.A., Allan, J.F., Carmichael, S.E.: Active rifting in Southwestern Mexico: manifestations of an incipient eastward spreading-ridge jump. Geology **13**, 54–57 (1985)
11. Allan, J.F.: Geology of the Northern Colima and Zacoalco Grabens, Southwest Mexico: Late Cenozoic rifting in the Mexican Volcanic Belt. Geol. Soc. Am. Bull. **97**, 473–485 (1986)
12. West, M.: The Colima Deep Seismic Experiment: Imaging the Magmatic Root of Colima Volcano [Data set]. International Federation of Digital Seismograph Networks (2006)
13. Gutierrez, Q.J., Escudero, C.R., Núñez-Cornú, F.J.: Geometry of the Rivera-Cocos Subduction Zone inferred from local seismicity. Bull. Seismol. Soc. Am. **105**, 3104–3113 (2015)
14. Gómez-González, J.M., Mendoza, C., Sladen, A., Guzmán-Speziale, M.: Kinematic source analysis of the 2003 Tecomán, México, earthquake (Mw 7.6) using teleseismic body waves. Bol. Soc. Geol. Mex. **62**, 249–262 (2010)
15. Zamora-Camacho, A., Espíndola, J.M., Reyes-Dávila, G.: The 1997–1998 activity of Volcan de Colima, Western Mexico: some aspects of the associated seismic activity. Pure Appl. Geophys. **164**, 39–52 (2007)
16. Gutenberg, B., Richter, C.F.: Frequency of earthquakes in California. Bull. seism. Soc. Am. **34**, 185–188 (1944)