

A Python Code for Detecting True Repeating Earthquakes from Self-Similar Waveforms (FINDRES)

Monica Sugan^{*1} , Stefano Campanella², Alessandro Vuan², and Nader Shakibay Senobari³

Abstract

Seismic data are generally scrutinized for repeating earthquakes (REs) to evaluate slip rates, changes in the mechanical properties of a fault zone, and accelerating nucleation processes in foreshock and aftershock sequences. They are also used to study velocity changes in the medium, earthquake physics and prediction, and for constraining creep rate models at depth. For a robust detection of repeaters, multiple constraints and different parameter configurations related to waveform similarity have been proposed to measure cross-correlation values at a local seismic network and evaluate the location of overlapping sources. In this work, we developed a Python code to identify REs (FINDRES), inspired by previous literature, which combines both seismic waveform similarity and differential *S-P* travel time measured at each seismic station. A cross-spectral method is applied to evaluate precise differential arrival travel times between earthquake pairs, allowing a subsample precision and increasing the capacity to resolve an overlapping common source radius. FINDRES is versatile and works with and without *P*- and *S*-wave phase pickings, and has been validated using synthetic and real data, and provides reliable results. It would contribute to the implementation of open-source Python packages in seismology, supporting the activities of researchers and the reproducibility of scientific results.

Cite this article as Sugan, M., S. Campanella, A. Vuan, and N. Shakibay Senobari (2022). A Python Code for Detecting True Repeating Earthquakes from Self-Similar Waveforms (FINDRES), *Seismol. Res. Lett.* **93**, 2847–2857, doi: [10.1785/0220220048](https://doi.org/10.1785/0220220048).

Introduction

Repeating earthquakes (REs) are families of two or more events, usually of small magnitude, having the same source rupture, location, and geometry, at different times. They represent recurring seismic energy releases from the same single fault patch and are commonly found on creeping plate boundaries, where asperities are loaded by aseismic slip (e.g., Nadeau and Johnson, 1998; Chen and Lapusta, 2009; Igarashi, 2010). REs are observed in different tectonic settings involving strike-slip, normal, and reverse faults (e.g., Chen *et al.*, 2008; Duverger *et al.*, 2018). They are helpful in investigating and measuring temporal and spatial variations in fault creep rates (Chen *et al.*, 2007; Uchida and Bürgmann, 2019), slip directly on the fault independently of geodetic measurements, and changes in the mechanical properties of a fault zone (Chaves *et al.*, 2020). They have also been observed during water injection experiments, providing images of the geometry and kinematics of transient slip patches (Bourouis and Bernard, 2007).

Many studies identify REs using waveform correlation or coherence (e.g., Nadeau and McEvilly, 1999; Igarashi *et al.*, 2003; Schaff and Waldhauser, 2005; Lengliné and Marsan,

2009; Uchida and Bürgmann, 2019). When using the cross-correlation (CC) approach, the frequency band and the length of the time window used in the calculation strongly impact the identification of the candidate repeaters (Gao and Kao, 2020). Long-time windows and broad frequency bands, together with the highest CC values, help to ensure that the identified events truly overlap, but at the same time, too-strict criteria may miss REs, despite the identical slip patch (e.g., Lengliné and Marsan, 2009; Gao *et al.*, 2021). On the contrary, when insufficiently strict criteria are used, the associated events could be nearby triggered earthquakes rather than true repeaters (Uchida, 2019).

For this reason, robust REs detection using multiple constraints and different parameter configurations related to waveform similarity and physics-based methods have been proposed (i.e., Chen *et al.*, 2008; Lengliné and Marsan,

1. National Institute of Oceanography and Applied Geophysics—OGS, Udine, Italy, <https://orcid.org/0000-0002-1247-3193> (MS); 2. National Institute of Oceanography and Applied Geophysics—OGS, Sgonico, Italy; 3. University of California Riverside, Riverside, California, U.S.A.

*Corresponding author: msugan@ogs.it

© Seismological Society of America

2009; Ellsworth and Bulut, 2018; Shakibay Senobari and Funning, 2019; Gao *et al.*, 2021). Using different approaches, such methods require waveform similarity and at least 50% seismic source overlap to confirm REs (Waldhauser and Ellsworth, 2002).

To identify REs, we follow the approach proposed in Chen *et al.* (2008) and Shakibay Senobari and Funning (2019). The method is based on the evidence that if two events have the same source mechanism and their location overlaps, the seismic waveform at each station should be very similar, as the difference between the *S*- and *P*- travel-time arrivals (ΔS -*P*). Using ΔS -*P* has the advantage of preventing problems with station timing biases or uncertainties in the events' origin time (Chen *et al.*, 2008).

In particular, we developed an open-source Python code named FINDRES (Python Code for Detecting true Repeating Earthquakes from Self-similar Waveforms), including NumPy (Harris *et al.*, 2020), SciPy (Virtanen *et al.*, 2020), ObsPy (Beyreuther *et al.*, 2010; Krischer *et al.*, 2015), Multitaper (Prieto, 2022), and Matplotlib (Hunter, 2007) dependencies. The efficiency of ObsPy has been widely demonstrated in different open-access packages for matched-filter detection methods (Chamberlain *et al.*, 2018; Vuan *et al.*, 2018) and REs detection (e.g., Hotovec-Ellis and Jeffries, 2016), among others.

In the next section, we provide a detailed description of the code workflow and describe the architecture of the input data. FINDRES is open-source and available through a public GitHub repository, where usage and additional information are provided.

The code is validated using both synthetic and real datasets. For evaluating FINDRES accuracy and testing, we calculate synthetics by wavenumber integration and use some real data REs in the Northern San Francisco Bay (California) provided by Shakibay Senobari and Funning (2019).

Material and Methods

Figure 1 illustrates the FINDRES flow chart. An initial catalog of candidate repeating earthquakes (CRE) is needed as a starting point to search for possible REs, together with the associated seismic waveforms acquired by a pool of seismic stations. An inventory file of the seismic stations and a suitable 1D *P*- and *S*-velocity model (V_P , V_S) for the area under investigation are also required. Optionally, associated *P* and *S* picks provided in one supported format can be used.

In the simplest case, CRE can be defined by analyzing earthquake interevent distances using a high-quality catalog and selecting as CRE the events close enough in space, in which the term “enough” depends basically on the station configuration and the quality of the starting catalog available. More complete and sophisticated criteria can be adopted to define CRE, such as the ones based on CC analysis (e.g., template matching), the output of waveform coherency, and so forth.

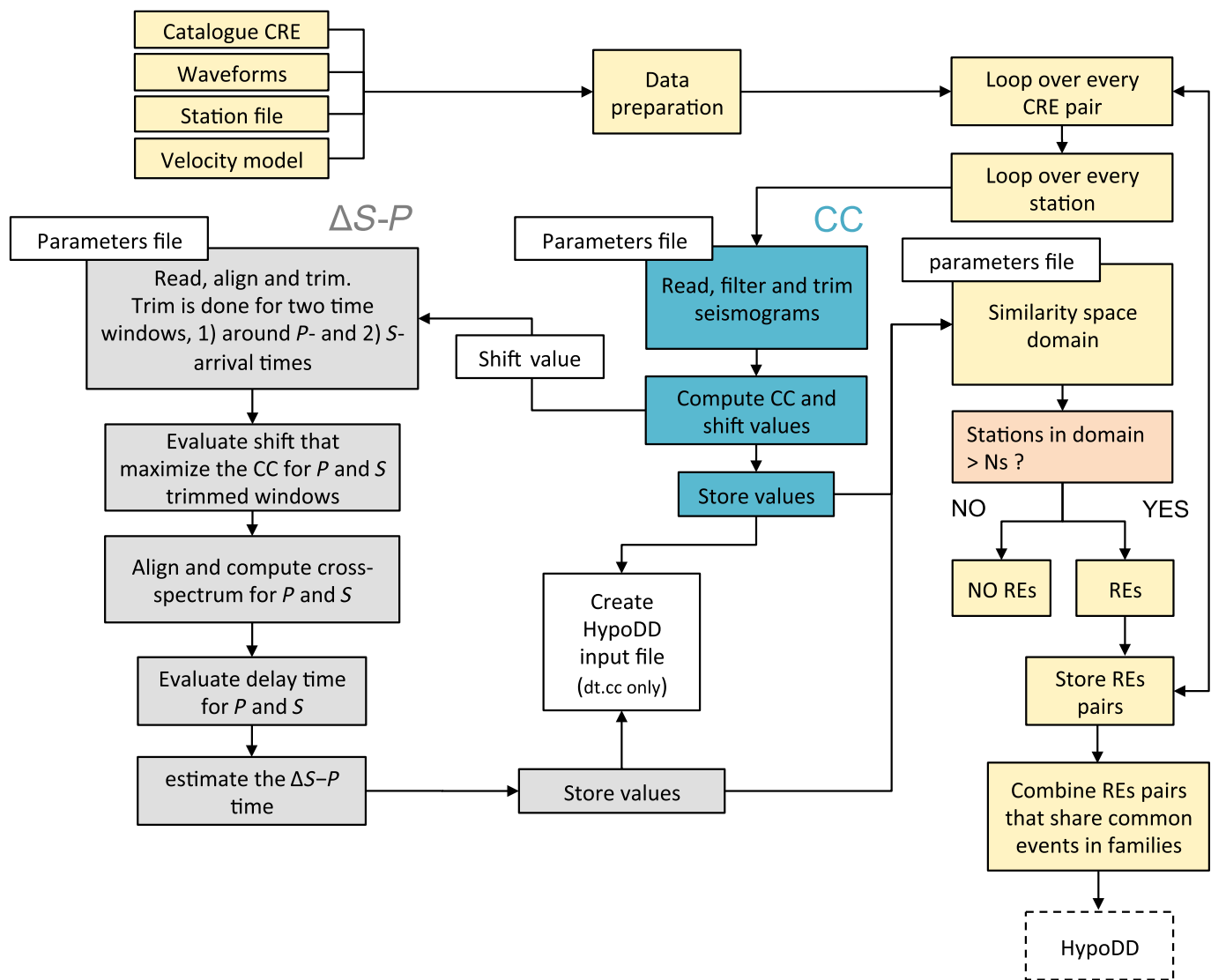
Whatever method is applied, the CRE catalog should be converted to a specific format. The code uses a modified version of the ZMAP format (Wiemer, 2001), in which the 11th and 12th columns of the file are reserved for the event's name (its index in a parent catalog) and the path to the waveform file. This indirection level provides a means to work with multiple catalogs and arbitrary filenames. After CRE identification, the associated seismic waveform and station metadata should be collected. Seismic data preparation consists of trimming CRE waveforms from continuous streams by using *P*- and *S*-wave theoretical travel times through a suitable 1D velocity model or associated picks. The waveforms are cut to include a pre and postevent time window; during later analyses, select shorter chunks of data centered around the *P* or *S* picks are used.

After input data preparation, the code explores every CRE pair, using CC values and differential arrival times of *P* and *S* waves. The code uses an early stopping strategy, and if the CC criteria fail, it will skip the subsequent test. Waveforms pairs that pass both tests are kept valid, and if we count enough stations, we consider them repeaters. We define a family of REs as a connected component of the graph where the nodes are events, and two nodes are connected if they are found to be a repeater doublet. In the end, FINDRES outputs the lists of events found together, optionally with the files needed for relocalization and graphics. Because the algorithm iterates over each possible event pair, its time complexity is quadratic depending on the size of the starting catalog. Nonetheless, if one can divide the latter into separate clusters, then these can be processed independently. Therefore, in the case of large-scale analyses, the recommendation is to unpack the catalog of CRE into smaller ones (e.g., exploiting candidates' location or preliminary CC analyses) and leveraging the input system of FINDRES to process them in parallel.

Numerical parameters for the CC and cross-spectrum analysis, together with the definition of thresholds to be used for REs identification, should be set in a YAML parameters configuration file, whereas the paths of the catalog and of the other input files are passed to the program via a command-line interface. An example of the numerical parameters file is provided for documentation purposes and is extensively commented on. The choice to use the popular YAML data-serialization format for both input and output was made, given its readability and comprehensive support.

CC values

The CC values for each CRE pair at the available seismic stations are first evaluated for an appropriate time window, including a pre-event interval before *P* phase and an interval after the *S* phase. This choice allows considering both *P* and *S* phases for stations at different epicentral distances without fixing a priori the time window length. For each station, *P* and *S*



phases can be obtained by reading picks from a bulletin in standard formats or calculating the theoretical travel-time arrivals by a suitable earth model when P and S picks are not available (e.g., Li *et al.*, 2020). As for the first solution, we implemented the readings for Hypoellipse (Lahr, 1999), Hypoinverse (Klein, 2002), NonlinLoc (Lomax *et al.*, 2000), and QuakeML format, whereas theoretical travel-time calculations are made using the Java TauP Toolkit routines (Crotwell *et al.*, 1999). We also include the possibility of using an ObsPy picker to identify P -arrival times instead of the theoretical values.

When reading phase picks from a detected event in a bulletin, S phases are often missing because S -arrival travel times can be difficult to recognize and mark. To overcome this difficulty, because we want to consider arrival travel times for both P and S phases, we proceeded as follows: when there is only a P pick in the bulletin, the S pick is approximately determined by supposing that S -wave arrivals precede the maximum of the signal envelope after the P pick. As a further

Figure 1. FINDRES flow chart. In yellow, we show the blocks related to data preparation and repeating earthquakes (REs) identification; in blue and gray the blocks related to the cross-correlation (CC) and S - and P -travel-time arrival (ΔS - P) processing procedures, respectively. The color version of this figure is available only in the electronic edition.

constraint, we force the maximum envelope to be found in a specific time window, centered using the P - and S -arrival time criteria ($t_S - t_P$), taking into account the epicentral distance D and average crustal P - and S -wave velocities (V_P and V_S).

After the P - and S -arrival travel-time definition, the trimmed CRE waveforms are finally cross-correlated in the time domain, using an appropriate frequency band, as suggested by Uchida (2019).

The user can implement the code, including different options for the correlation method (e.g., Gao and Kao, 2020) or other features.

The frequency bandwidth is usually fixed based on the source dimension (r) and corner frequency (f_c). For recognizing overlapping events with approximately the same source size, we must account for frequencies (f) $f \geq V_s/4r$, in which r is the source radius. When a circular source radius is assumed, r can be estimated using the following equation (e.g., Eshelby, 1957):

$$r = ((7/16)(M_0/\Delta\sigma))^{1/3}, \quad (1)$$

in which M_0 is the seismic moment, and $\Delta\sigma$ is the uniform stress drop. Although M_0 is generally well estimated, $\Delta\sigma$ can be poorly constrained. Its incorrect assumption can erroneously estimate the seismic source radius and classify close events as repeaters and vice versa. The upper bound of the bandwidth should instead be lower than the corner frequency to reduce the possibility of excluding events with different rupture patterns (i.e., slip directivity) but overlapping source areas compatible with REs (Uchida, 2019). The corner frequency increases as the magnitude and seismic source radius decrease. In general, even the detectable distance, which corresponds to the critical distance beyond which the event of a given magnitude can be detected, decreases with magnitude (e.g., Li *et al.*, 2020). Commonly, the corner frequency is computed using spectral ratios (Andrews, 1986) or empirical Green's functions (Hough and Dreger 1995) or simplified relationships using dynamic rupture models (e.g., Brune, 1970; Madariaga, 1976).

Because the CRE catalog can have widespread magnitude ranges (corresponding to different source radii and corner frequencies), we consider defining different band-pass filters, specified using the input parameter file. The first step of the analysis performs the CC, and when a threshold CC value is reached (e.g., 0.7–1.0), the procedure calculates the ΔS - P by applying the cross-spectrum methods at each seismic station.

Cross spectrum for ΔS - P estimation

To measure precise differential ΔS - P arrival times between CRE pairs, we apply the cross-spectral method by following the approach described in Poupinet *et al.* (1984) and using the ObsPy port (Krischer *et al.*, 2015). The spectral method is preferred because it allows subsample precision to resolve minimal source separation.

Data processing at a single station requires a prealignment of the waveforms for each pair of events based on the relative shift for which a maximum CC value is obtained. This step is essential, and signal padding is also required to have the same number of samples for the signals under process.

To compute the cross spectrum, we selected two different time intervals around P - and S -phase arrival travel times for each station and used the Multitaper module developed in Python (Prieto, 2022). Cross-spectrum parameters are set in the parameter file. As described in the previous section, P

and S window lengths are tailored based on P and S travel times.

The delay times for both P and S waves are calculated by the best-fitting slope of the phase spectrum for points with coherence above a specific value. The value can be tailored based on the case study but should be high enough to guarantee high coherence in a specific frequency band (e.g., 0.85 or above).

To constrain further, we can extend the frequency range beyond the maximum frequency used during the CC and specify a minimum number of points above the coherence threshold to evaluate fitting. Then, the relative difference between P and S delay times is taken to estimate the ΔS - P time between the events.

Similarity space domain

The similarity space domain identifies earthquake pairs on the same asperity at different times. After the CC and cross-spectrum analyses are performed, we need to define CC and ΔS - P thresholds indicating REs that share at least 50% of the seismic source. Between these thresholds, the similarity space domain took place, and the number of seismic stations inside it can be evaluated for a proper REs identification (Chen *et al.*, 2008).

Although the CC threshold can be fixed by choosing a good value, the ΔS - P threshold varies according to the seismic source radius. We provide a utility to explore various settings to get an average value for the ΔS - P domain expected for REs pair, based on magnitude ranges and the adopted $\Delta\sigma$ and velocity model.

The selected CC and ΔS - P thresholds that define the similarity space domain must be specified in the parameters file, together with the minimum number of seismic stations (N_s) required. Because the source radius and N_s vary based on magnitude ranges, the code enables selecting ΔS - P thresholds and N_s for different magnitude ranges. The selected N_s depends on the available stations, network configuration, and the size of the events.

Sampling frequency, number of stations, magnitude range, and hypocentral distance are related, and results strictly depend on the data quality. The code itself can be used with various configuration parameters to perform some tests before approaching the problem. These tests can provide indications on the setup of parameters.

The procedure loops for all CRE pairs and stores the resulting REs in a list. All REs pairs that share common events are grouped in families in the final step. Different parameter configurations can be used to perform stress tests and evaluate the stability of the identified REs.

We can relocate REs families with multiple events by applying a double-difference relocation algorithm. For this reason, as a by-product, we also generate an input file to be used with HypoDD code (Waldhauser and Ellsworth, 2000). To develop

it, we follow the technique of [Chen et al. \(2008\)](#), in which the relative times for P and S (called tt_P and tt_S , respectively) are derived from the S - P time (S_mP) and V_P/V_S ratio:

$$tt_P = S_mP / ((V_P/V_S) - 1), \tag{2}$$

$$tt_S = S_mP / (1 - (V_S/V_P)). \tag{3}$$

Using this approach, relocated events' absolute locations may not be remarkably accurate, but their relative position is improved ([Chen et al., 2008](#)), and multiple REs can be classified.

When using HypoDD, we recommend using the singular value decomposition (SVD). It is the standard method for solving the least-squares problem. Other choices include the conjugate gradients method ([Paige and Saunders, 1982](#)), which is used when SVD is not computationally feasible and sacrifices accuracy for speed ([Waldhauser and Ellsworth, 2000](#)). The number of REs in a family is usually less than 100, so SVD, which is suitable for a smaller number of events, is appropriate.

Application to Synthetic and Real Data

We apply FINDRES to a synthetic dataset to observe the CC and ΔS - P estimated on event pairs of known seismic sources. Synthetic seismograms are generated using a frequency–wave-number code ([Herrmann, 2013](#)) and a simple 1D velocity model (Table 1).

We investigate the CC values and the differences between the theoretical and calculated ΔS - P for pairs of earthquakes with the same focal mechanism (strike, dip, rake: 242°, 40°, 80°, respectively), depth (10 km), M_0 (2.24×10^{11} N · m, corresponding to an M_w 1.5) and corner frequency (25 Hz), assuming a stress drop of 1 MPa. We generate synthetic seismograms at epicentral distances of about 15 km and the azimuth of about 226°, at a sampling rate of 100 Hz.

We simulate pairs of events characterized by interevent distances (d) of 5, 10, 50, and 100 m. The spaces between the events are defined based on the size of the seismic source to investigate their overlap (>50% for d of about 5 and 10 m, 50% for d of about 50 m, and <50% for d of about 100 m). Assuming a stress drop of 1 MPa, the corresponding source radius using specified M_0 is approximately 46 m, using equation (1).

FINDRES is applied to synthetics pairs. The experiment is used to validate the code and does not try to mimic any real case or evaluate limits due to the network configuration related to magnitude–distance couples. Figure 2 shows an example of earthquakes characterized by interevent distance (d) of about 10 m. CC is estimated using a 1–25 Hz band-pass filter, suitable

TABLE 1
Velocity Model Used to Generate Synthetic Seismograms

Layer Depth (km)	V_P (km/s)	V_P/V_S
0.00	4.10	1.84
2.00	5.10	1.77
7.86	5.79	1.79
17.80	6.34	1.73
42.00	8.00	1.78

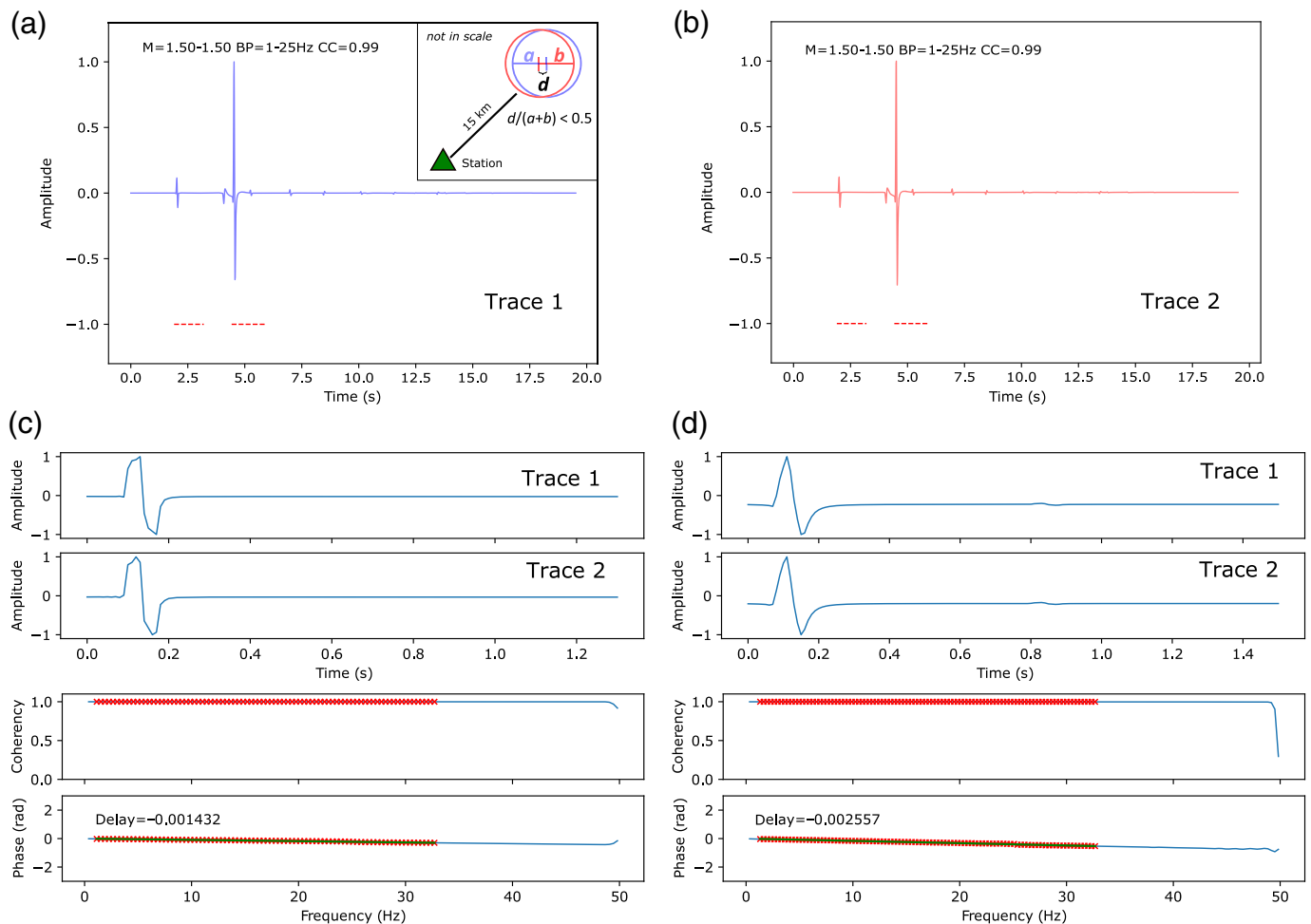
to recognize possible nonoverlapping events and not exceeding the corner frequency.

Although the CC threshold can be fixed by choosing a conservative value (e.g., 0.9), the ΔS - P threshold varies according to the magnitude of the events, stress drop, and seismic radius. For synthetics, using M_w 1.5, stress drop of about 1 MPa, the velocity model in Table 1, and event-station distances of 15 km, the ΔS - P threshold defining the spatial similarity domain is estimated at approximately 0.00575 s. This value is less than the single sample accuracy achievable using 100 Hz data. FINDRES results are shown in the similarity space domain in Figure 3a, in which the agreement between theoretical and calculated ΔS - P can be observed. We also investigate how noise, incorporated in the synthetic signal, can modify the correlation and coherence-based S - P time, adding different signal-to-noise ratio (SNR) levels of random noise to the synthetic data (SNR 2, 3, and 5). SNR is calculated using the root mean square (rms), in which signal and noise are defined by the rms amplitude in a window 1 s after and 3 s before the P -arrival time, respectively. We can observe that both CC and ΔS - P lose accuracy, but still, the values are within the selected similarity space domain (Fig. 3b–d) up to SNR 2. The synthetic also shows a demonstration of closely located but not overlapping events (green symbols in Fig. 3). The green symbols correspond to pair of earthquakes with high-CC values (CC > 0.90 in Fig. 3a–d), located outside the similarity space domain (gray box), and therefore rejected by the procedure. In this case, it is evident that the evaluation of the delta S - P is essential to correctly identify or discard REs.

To validate FINDRES using a real dataset, we applied the code to two REs families identified by [Shakibay Senobari and Funning \(2019\)](#) in the Northern San Francisco Bay Area (Table 2).

For these events, we use waveform data, metadata, and picks from the [Northern California Earthquake Data Center \(NCEDC\) \(2014\)](#).

All the parameters and the data used for the test are described in the FINDRES GitHub repository, together with



all the necessary information to reproduce the results. In particular, we used a 1–20 Hz band-pass filter to evaluate the CC values, whereas the delay times for *P* and *S* waves are estimated using the cross-spectrum plot versus frequency. Delay times are obtained by fitting the slope of the points with coherence above 0.88 for a portion of a signal centered on *P*- and *S*-arrival travel times (Fig. 4). The lower bound of the CC used for repeated earthquake detections in the time series is about (0.90). This is not a rule. When using phase coherence, we can choose to use a slightly lower value (0.88) to account for possible numerical instability and ripples at certain frequencies. In general, we recommend choosing a value greater than 0.85 and less than or equal to 0.90.

For source radius estimation, we use 3 MPa as stress drop, the M_w – M_0 conversion, and the velocity model of [Shakibay Senobari and Funning \(2019\)](#). We established the similarity space domain using 0.9 for CC and 0.007 s for ΔS –*P*, and set N_s 3 to declare REs.

Using FINDRES and the HypoDD code, we obtained consistent results with the published ones ([Shakibay Senobari and Funning, 2019](#)). Figure 5 shows the similarity space domain for the pairs with id 21128020 and 71439381 in Table 2, and

Figure 2. FINDRES applied to synthetic seismograms for pair of earthquakes characterized by interevent distance (*d*) of about 10 m. We generate synthetic seismograms with an epicentral distance of about 15 km and the azimuth of about 226° for earthquakes pairs with the same focal mechanism (strike, dip, rake: 242°, 40°, 80°). The inset in (a) shows a sketch (not in scale) of the geometry used for the generation of synthetics with *d* = 10 m and seismic source radii (*a* and *b*). Example of CC evaluation showing (a) trace 1 and (b) trace 2, and (c,d) cross-spectrum analysis for *P* and *S* waves time window, respectively. (a,b) CC value (0.99) is calculated for the Z component, waveforms are band-pass filtered between 1 and 25 Hz, and amplitudes are normalized. The dotted red lines represent the time windows used in the cross spectrum for *P* and *S* waves. We chose 0.1 s before and 1.2 s after the *P* pick and 0.1 s before and 1.4 after the estimated *S* arrival travel time. (c) The delay times for *P* and (d) *S* waves are estimated using the cross-spectrum plot versus frequency. Delay times are obtained by fitting the slope of the points with coherence above 0.88. Finally, the difference between relative *P* and *S* delay times is used to estimate ΔS –*P*. The color version of this figure is available only in the electronic edition.

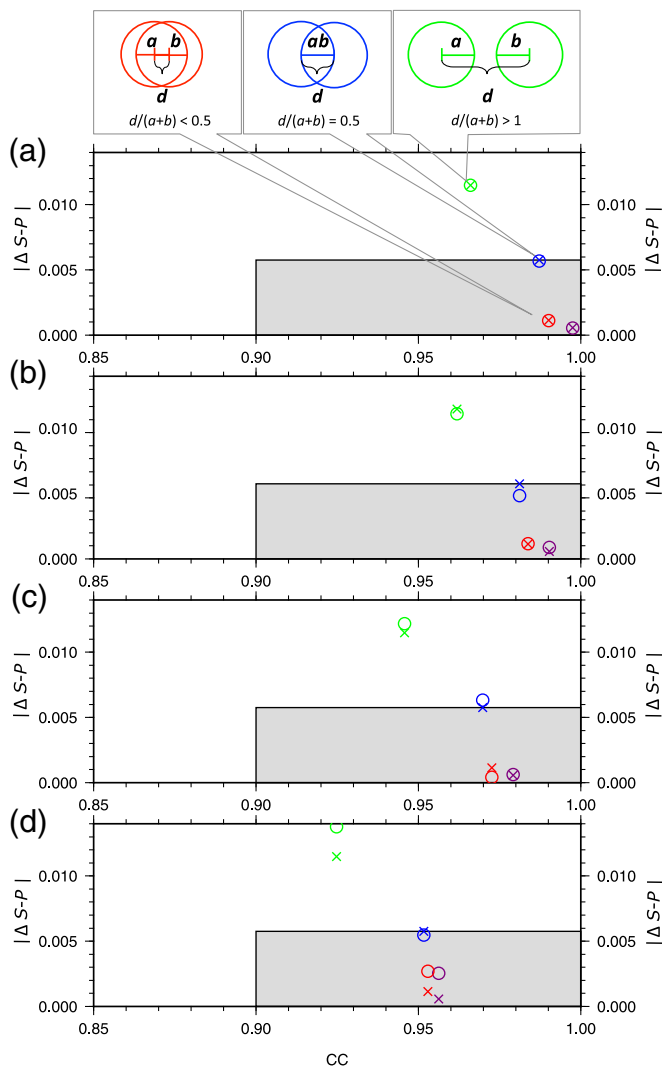


Figure 3. Similarity space domain for synthetic seismograms: CC values and absolute $\Delta S-P$ using theoretical (crosses) and estimated (circles) values. Each symbol in the figure corresponds to only one station and one different pair of events. We generate synthetic seismograms with an epicentral distance of about 15 km and the azimuth of about 226° for earthquakes pairs with the same focal mechanism (strike, dip, rake: 242° , 40° , 80°), characterized by four different interevent distances d considering seismic source overlapping ($d = 5, 10$, and 50 m) and non-overlapping case ($d = 100$ m). Purple: $d = 5$ m; red: $d = 10$ m; blue: $d = 50$ m; green: $d = 100$ m. The gray box highlights the similarity space domain estimated for the M_w 1.5 earthquake pairs (see [Application to Synthetic and Real Data](#) for details). In the figure, the seismic source overlap that characterized the similarity space domain is expressed in terms of interevent distance (d) and seismic source radii (a and b). Overlapping REs are characterized by $d/(a+b) < 0.5$, nonoverlapping events by $d/(a+b) > 1$. Similarity space domains for synthetic seismograms (a) without noise and with random noise and different signal-to-noise ratio (SNR) levels are shown, (b) SNR about 5, (c) SNR about 3, and (d) SNR about 2. The color version of this figure is available only in the electronic edition.

Figure 6 the final HypoDD location for the corresponding REs family (Id_REs 1).

Conclusions

We developed a code that helps identify true REs from self-similar waveforms. REs represent an intriguing subject because they are able to provide insights on the fault strength and the interseismic slip independently of geodetic observations. Moreover, small magnitude earthquakes repeated in space and time could be associated with aseismic slip in foreshocks and aftershock sequences and thus, indicative of variations in tectonic loading and possible creeping.

We describe here the FINDRES Python code implementation that combines seismic waveform similarity using CC and differential $S-P$ travel times.

The code can be applied to different types of earthquakes. The data preparation and setup of input parameters must be selected appropriately considering the magnitude ranges, corner frequencies, and network configurations. The seismometer characteristics and the sampling frequency limit the frequency band available for analysis, that is, we cannot investigate REs if the corner frequency is higher than the Nyquist frequency. [Uchida \(2019\)](#), among others, shows frequency and magnitude ranges for different-scale seismic events suitable for the detection of repeating earthquakes.

The code itself can be used with various configuration parameters to perform some preliminary tests. These tests can help in the setup of the input parameters and approaching the problem.

FINDRES is versatile and works with and without P - and S -wave phase pickings. A suitable velocity model for the area under investigation is required in the second case. The procedure is also applicable to seismic catalogs built by nonstandard waveform-based location methods, for which P and S picks are not always available.

Finally, the code is validated using wavenumber integration synthetic seismograms, and a real case study is reproduced to confirm REs families in Northern California.

Data and Resources

Seismic waveform and picks phases can be accessed through the Northern California Earthquake Data Center (NCEDC), doi: [10.7932/NCEDC](https://doi.org/10.7932/NCEDC). In particular, we used the Northern California Seismic Network (NCSN) catalog and Phase in Hypoinverse format, available at <https://ncedc.org/ncedc/catalog-search.html> (last accessed January 2022). Wavenumber integration seismograms are calculated using Computer Programs in Seismology ([Herrmann, 2013](#)). FINDRES code and data used in this work are available at <https://github.com/msugan/FINDRES> (last accessed June 2022).

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

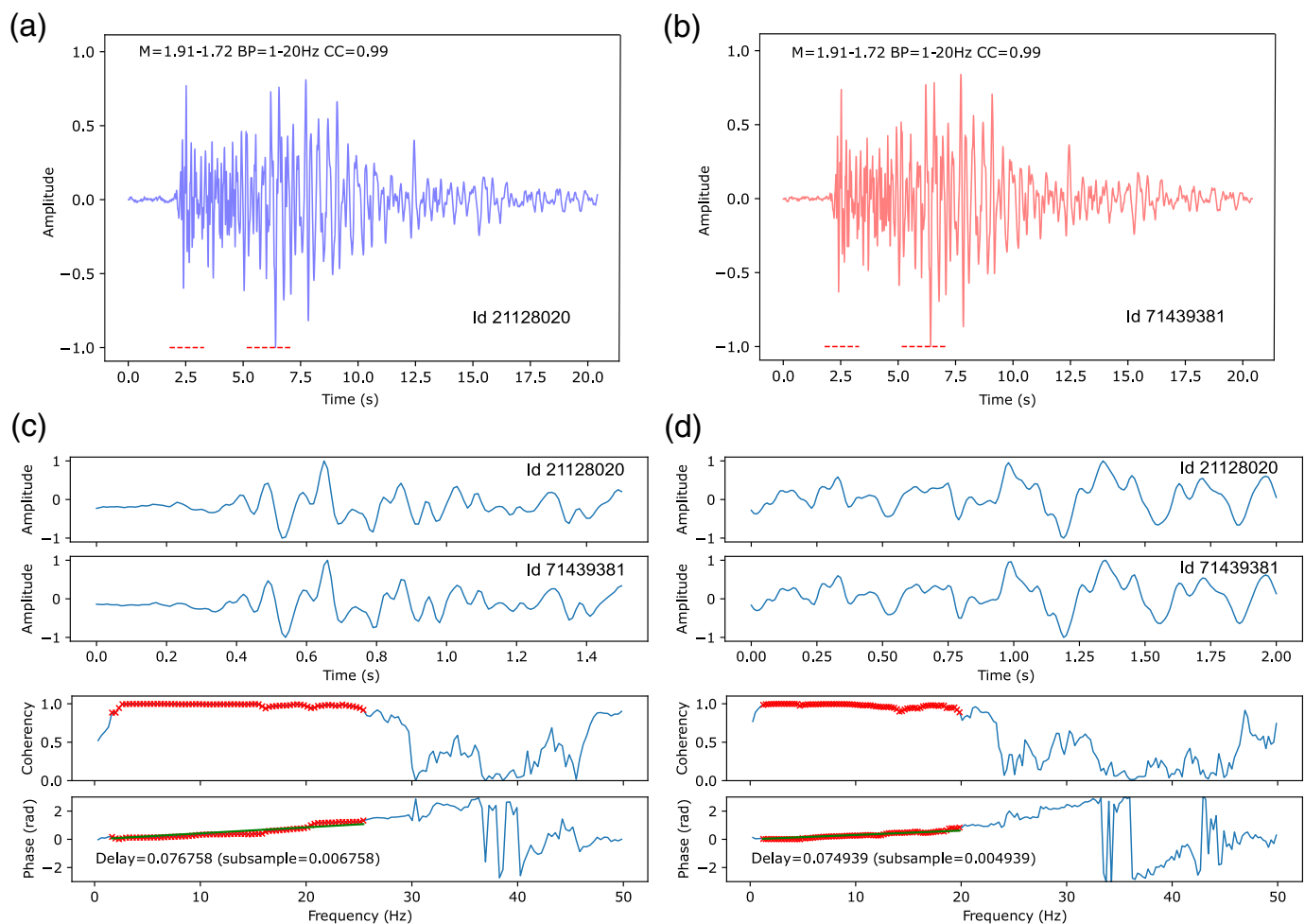


Figure 4. FINDRES applied to a real dataset (event id 21128020 and 71439381 in Table 2); NTYB seismic station. Example of CC evaluation showing the seismic waveform of (a) 21128020, and (b) 71439381 events, (c,d) cross-spectrum analysis for *P* and *S* waves time window, respectively. (a,b) the CC value (0.99) is calculated for the Z component, waveforms are band-pass filtered between 1 and 20 Hz, amplitudes are normalized. The dotted red lines represent the time windows used in the cross

spectrum for *P* and *S* waves. We choose 0.2 s before and 1.3 s after the *P* pick and 0.2 s before and 1.8 after the estimated *S* arrival travel time. The delay times for (c) *P* and (d) *S* waves are estimated using the cross-spectrum plot versus frequency. Delay times are obtained fitting the slope of the points with coherence above 0.88. Finally, the difference between relative *P* and *S* delay times is used to estimate $\Delta S-P$. The color version of this figure is available only in the electronic edition.

TABLE 2

Two REs Families (Id_REs 0, 1) (from Shakibay Senobari and Funning, 2019)

Id_REs	Id_NCEDC	Latitude (°)	Longitude (°)	Date (yyyy/mm/dd)	Time (hh:mm:ss.ss)	Magnitude
0	122842	-122.99767	38.88830	1988/08/25	21:48:30.40	1.87
	484038	-122.99550	38.88750	1996/11/08	07:52:19.60	2.15
	21442564	-122.99617	38.88733	2005/03/01	10:01:21.00	2.08
	72388871	-122.99117	38.88750	2015/01/30	06:48:41.18	2.04
1	128170	-122.76884	38.54000	1988/12/07	06:47:34.21	2.04
	21128020	-122.76817	38.54183	2000/10/02	00:12:38.37	1.91
	71439381	-122.77167	38.54550	2010/07/30	05:57:34.30	1.72

Associated values (latitude, longitude, date, time, and magnitude) for each event are taken from NCEDC (2014).

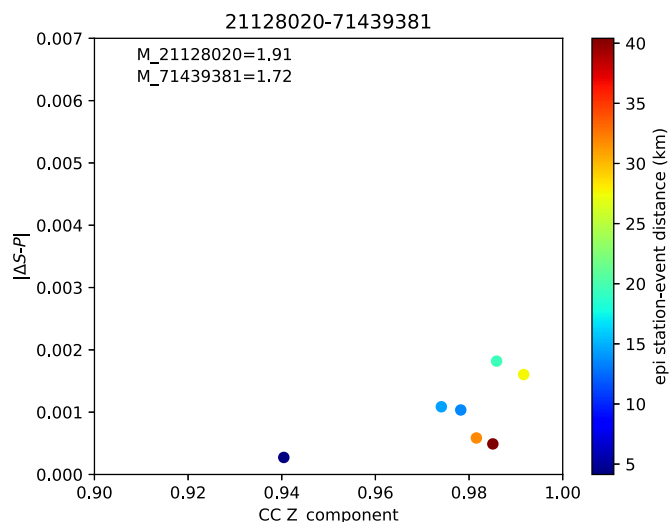


Figure 5. FINDRES applied to a real dataset; (a) similarity space domain obtained for REs pair (event id 21128020 and 71439381 in Table 2). Each point in the graph corresponds to a seismic station, colored according to the epicentral distance. The color version of this figure is available only in the electronic edition.

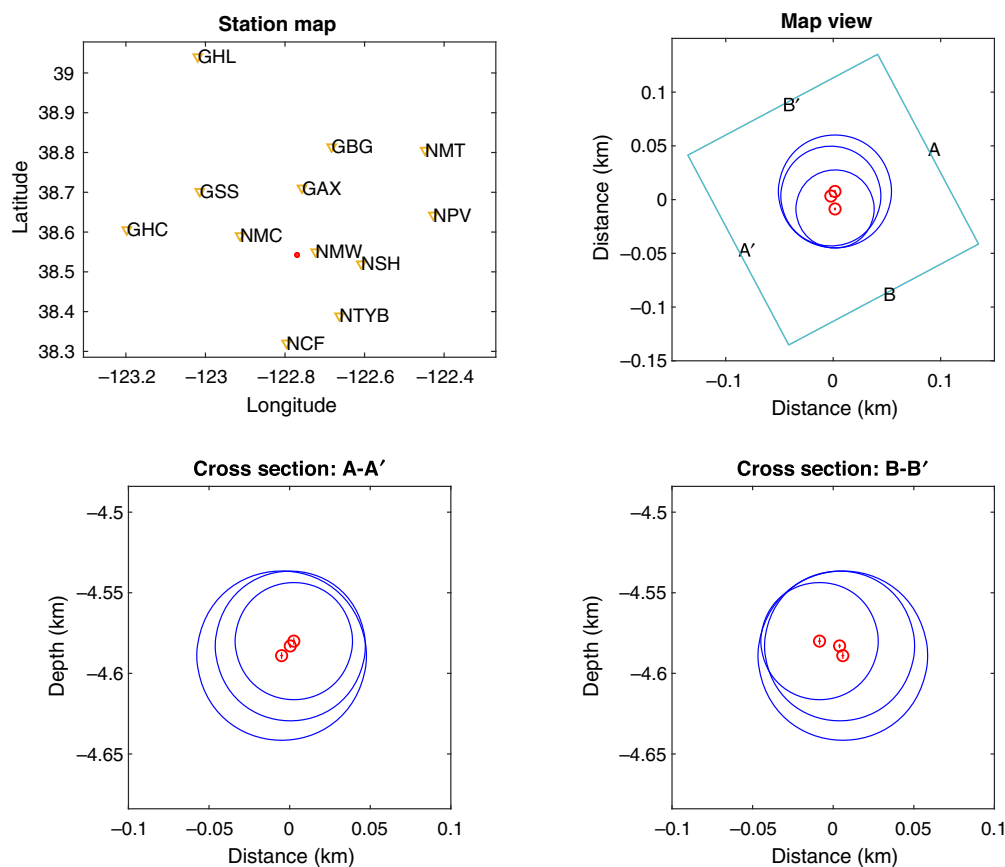


Figure 6. FINDRES applied to a real dataset; HypoDD location obtained for REs family (Id_REs 1 in Table 2), after the analysis of all the pair combinations. Red points show earthquake locations and the blue circles show the associated seismic source radii. The color version of this figure is available only in the electronic edition.

Acknowledgments

The authors would like to thank the Associate Editor, Calum J. Chamberlain, and one anonymous reviewer for their constructive comments. Software implementation is partly supported by the project “Real-time earthquake risk reduction for a resilient Europe” (RISE), funded by the European Union’s Horizon 2020 research and innovation program under Grant Agreement Number 821115, and by Istituto Nazionale di Oceanografia e di Geofisica Sperimentale-OGS and Consorzio Interuniversitario per il Calcolo Automatico dell’Italia Nord Orientale (CINECA) under High Performance Computing - Training and Research for Earth Sciences (HPC-TRES) program award number 2021-01. Nader Shakibay Senobari acknowledges support from U.S. Geological Survey (USGS) Award G20AP00092 and National Science Foundation (NSF) Award ID 2103976. The authors thank the Python and ObsPy communities for sharing open-source codes and ideas, which can be used and implemented to enrich existing seismological codes.

References

- Andrews, D. J. (1986). Objective determination of source parameters and similarity of earthquakes of different size, in *Earthquake Source Mechanics*, American Geophysical Union, Washington D.C., 259–268, doi: [10.1029/GM037p0259](https://doi.org/10.1029/GM037p0259).
- 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, doi: [10.1785/gssrl.81.3.530](https://doi.org/10.1785/gssrl.81.3.530).
- Bourouis, S., and P. Bernard (2007). Evidence for coupled seismic and aseismic fault slip during water injection in the geothermal site of Soultz (France), and implications for seismogenic transients, *Geophys. J. Int.* **169**, 723–732, doi: [10.1111/j.1365-246X.2006.03325.x](https://doi.org/10.1111/j.1365-246X.2006.03325.x).
- Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.* **75**, 4997–5009.
- Chamberlain, C. J., C. J. Hopp, C. M. Boese, E. Warren-Smith, D. Chambers, S. X. Chu, K. Michailos, and J. Townend (2018). EQcorrscan: Repeating and near repeating earthquake detection and analysis in Python, *Seismol. Res. Lett.* **89**, no. 1, 173–181, doi: [10.1785/0220170151](https://doi.org/10.1785/0220170151).
- Chaves, E. J., S. Y. Schwartz, and R. E. Abercrombie (2020). Repeating earthquakes record fault weakening and healing

- in areas of megathrust postseismic slip, *Sci. Adv.* **6**, no. 32, eaaz9317, doi: [10.1126/sciadv.aaz9317](https://doi.org/10.1126/sciadv.aaz9317).
- Chen, K. H., R. M. Nadeau, and R.-J. Rau (2007). Towards a universal rule on the recurrence interval scaling of repeating earthquakes? *Geophys. Res. Lett.* **34**, L16308, doi: [10.1029/2007GL030554](https://doi.org/10.1029/2007GL030554).
- Chen, K. H., R. M. Nadeau, and R.-J. Rau (2008). Characteristic repeating microearthquakes on an arc-continent collision boundary zone: The Chihshang fault of eastern Taiwan, *Earth Planet. Sci. Lett.* **276**, nos. 3/4, 262–272, doi: [10.1016/j.epsl.2008.09.021](https://doi.org/10.1016/j.epsl.2008.09.021).
- Chen, T., and N. Lapusta (2009). Scaling of small repeating earthquakes explained by interaction of seismic and aseismic slip in a rate and state fault model, *J. Geophys. Res.* **114**, no. B1, doi: [10.1029/2008JB005749](https://doi.org/10.1029/2008JB005749).
- Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.* **70**, no. 2, 154–160, doi: [10.1785/gssrl.70.2.154](https://doi.org/10.1785/gssrl.70.2.154).
- Duverger, C., S. Lambotte, P. Bernard, H. Lyon-Caen, A. Deschamps, and A. Nercissian (2018). Dynamics of microseismicity and its relationship with the active structures in the western Corinth Rift (Greece), *Geophys. J. Int.* **215**, 196–221, doi: [10.1093/gji/ggy264](https://doi.org/10.1093/gji/ggy264).
- Ellsworth, W. L., and F. Bulut (2018). Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks, *Nature Geosci.* **11**, 531–535, doi: [10.1038/s41561-018-0145-1](https://doi.org/10.1038/s41561-018-0145-1).
- Eshelby, J. D. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems, *Proc. Math. Phys. Sci.* **241**, 376–396.
- Gao, D., and H. Kao (2020). Optimization of the match-filtering method for robust repeating earthquake detection: The multisegment cross-correlation approach, *J. Geophys. Res.* **125**, e2020JB019714, doi: [10.1029/2020JB019714](https://doi.org/10.1029/2020JB019714).
- Gao, D., H. Kao, and B. Wang (2021). Misconception of waveform similarity in the identification of repeating earthquakes, *Geophys. Res. Lett.* **48**, no. 13, e2021GL092815, doi: [10.1029/2021GL092815](https://doi.org/10.1029/2021GL092815).
- Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.* (2020). Array programming with NumPy, *Nature* **585**, 357–362, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- Herrmann, R. B. (2013). Computer programs in seismology: An evolving tool for instruction and research, *Seismol. Res. Lett.* **84**, no. 6, 1081–1088, doi: [10.1785/0220110096](https://doi.org/10.1785/0220110096).
- Hotovec-Ellis, A. J., and C. Jeffries (2016). Near real-time detection, clustering, and analysis of repeating earthquakes: Application to Mount St. Helens and Redoubt Volcanoes—Invited, *Presented at Seismological Society of America Annual Meeting*, Reno, Nevada, 20 April.
- Hough, S. E., and D. S. Dreger (1995). Source parameters of the 4/23/92 Joshua Tree, California, earthquake and its aftershocks: Empirical Green's function analysis of TERRA scope and GEOS data, *Bull. Seismol. Soc. Am.* **85**, 1576–1590.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* **9**, no. 3, 90–95.
- Igarashi, T. (2010). Spatial changes of inter-plate coupling inferred from sequences of small repeating earthquakes in Japan, *Geophys. Res. Lett.* **37**, L20304, doi: [10.1029/2010GL044609](https://doi.org/10.1029/2010GL044609).
- Igarashi, T., T. Matsuzawa, and A. Hasegawa (2003). Repeating earthquakes and interplate aseismic slip in the northeastern Japan subduction zone, *J. Geophys. Res.* **108**, no. B5, 2249, doi: [10.1029/2002JB001920](https://doi.org/10.1029/2002JB001920).
- Klein, F.W. (2002). User's guide to HYPOINVERSE-2000: A Fortran program to solve for earthquake locations and magnitudes, *U.S. Geol. Surv. Profess. Pap.* 2002, Rep. 02-17 1–123.
- 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**, doi: [10.1088/1749-4699/8/1/014003](https://doi.org/10.1088/1749-4699/8/1/014003).
- Lahr, J. C. (1999). HYPOELLIPSE: A computer program for determining local earthquake hypocentral parameters, magnitude, and first-motion pattern (Y2K Compliant Version), *U.S. Geol. Surv. Open-file Rept.* 99-023, Golden, Colorado, Paper and On-line Editions, 112 pp.
- Lengliné, O., and D. Marsan (2009). Inferring the coseismic and post-seismic stress changes caused by the 2004 Mw = 6 Parkfield earthquake from variations of recurrence times of microearthquakes, *J. Geophys. Res.* **114**, no. B1, doi: [10.1029/2008JB006118](https://doi.org/10.1029/2008JB006118).
- Li, L., J. Tan, B. Schwarz, F. Staněk, N. Poiata, P. Shi, L. Diekmann, L. Eisner, and D. Gajewski (2020). Recent advances and challenges of waveform-based seismic location methods at multiple scales, *Rev. Geophys.* **58**, e2019RG000667, doi: [10.1029/2019RG000667](https://doi.org/10.1029/2019RG000667).
- Lomax, A., J. Virieux, P. Volant, and C. Berge-Thierry (2000). Probabilistic earthquake location in 3D and layered models, in *Advances in Seismic Event Location. Modern Approaches in Geophysics*, C. H. Thurber and N. Rabinowitz (Editors), Vol. 18, Springer, Dordrecht, The Netherlands, doi: [10.1007/978-94-015-9536-0_5](https://doi.org/10.1007/978-94-015-9536-0_5).
- Madariaga, R. (1976). Dynamics of an expanding circular fault, *Bull. Seismol. Soc. Am.* **66**, 639–666.
- Nadeau, R. M., and L. R. Johnson (1998). Seismological studies at Parkfield VI: Moment release rates and estimates of source parameters for small repeating earthquakes, *Bull. Seismol. Soc. Am.* **88**, no. 3, 790–814.
- Nadeau, R. M., and T. V. McEvilly (1999). Fault slip rates at depth from recurrence intervals of repeating microearthquakes, *Science* **285**, 718–721, doi: [10.1126/science.285.5428.718](https://doi.org/10.1126/science.285.5428.718).
- Northern California Earthquake Data Center (NCEDC) (2014). Northern California Earthquake Data Center, *UC Berkeley Seismological Laboratory*, Dataset, doi: [10.7932/NCEDC](https://doi.org/10.7932/NCEDC).
- Paige, C., and M. Saunders (1982). LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.* **8**, 43–71, doi: [10.1145/355984.355989](https://doi.org/10.1145/355984.355989).
- Poupinet, G., W. L. Ellsworth, and J. Frechet (1984). Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California, *J. Geophys. Res.* **89**, no. B7, 5719–5731.
- Prieto, G. A. (2022). Multitaper: A multitaper spectrum analysis package in Python, *Seismol. Res. Lett.* **93**, no. 3, 1922–1929, doi: [10.1785/0220210332](https://doi.org/10.1785/0220210332).
- Schaff, D. P., and F. Waldhauser (2005). Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seismol. Soc. Am.* **95**, no. 6, 2446–2461, doi: [10.1785/0120040221](https://doi.org/10.1785/0120040221).
- Shakibay Senobari, N., and G. J. Funning (2019). Widespread Fault Creep in the Northern San Francisco Bay Area revealed by multi-station cluster detection of repeating earthquakes, *Geophys. Res. Lett.* **46**, 6425–6434, doi: [10.1029/2019GL082766](https://doi.org/10.1029/2019GL082766).

- Uchida, N. (2019). Detection of repeating earthquakes and their application in characterizing slow fault slip, *Prog. Earth Planet Sci.* **6**, 40, doi: [10.1186/s40645-019-0284-z](https://doi.org/10.1186/s40645-019-0284-z).
- Uchida, N., and R. Bürgmann (2019). Repeating earthquakes, *Ann. Rev. Earth Planet. Sci.* **47**, 305–332, doi: [10.1146/annurev-earth-053018-060119](https://doi.org/10.1146/annurev-earth-053018-060119).
- Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. (2020) SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nat. Methods* **17**, no. 3, 261–272.
- Vuan, A., M. Sugan, G. Amati, and A. Kato (2018). Improving the detection of low-magnitude seismicity preceding the Mw 6.3 L'Aquila earthquake: Development of a scalable code based on the cross-correlation of template earthquakes, *Bull. Seismol. Soc. Am.* **108**, no. 1, 471–480, doi: [10.1785/0120170106](https://doi.org/10.1785/0120170106).
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, *Bull. Seismol. Soc. Am.* **90**, no. 6, 1353–1368.
- Waldhauser, F., and W. L. Ellsworth (2002). Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations, *J. Geophys. Res.* **107**, no. B3, 2054, doi: [10.1029/2000JB000084](https://doi.org/10.1029/2000JB000084).
- Wiemer, S. (2001). A software package to analyze seismicity: ZMAP, *Seismol. Res. Lett.* **72**, no. 3, 373–382, doi: [10.1785/gssrl.72.3.373](https://doi.org/10.1785/gssrl.72.3.373).

Manuscript received 15 February 2022

Published online 12 July 2022