Supporting Information for

The unlocking process of the 2016 Central Italy seismic sequence

Sugan M.[1], Campanella S.[1], Chiaraluce L.[2], Michele M.[2], Vuan A.[1]

[1] National Institute of Oceanography and Applied Geophysics – OGS (Italy)

[2] Istituto Nazionale di Geofisica e Vulcanologia (Italy)

Contents of this file

TEXT S1 to S4

Introduction

A full description of the methods used in the article is provided for:

TEXT S1 - *Input catalog*

TEXT S2 - *Template matching*

TEXT S3 - *Clustering*

TEXT S4 - *Repeating earthquakes*

TEXT S1

Input catalog

The events initially recorded and located by the Italian National Network and Seismological Data Centre (INGV Seismological Data Centre 2006; ISIDe Working Group, 2007) have been relocated in absolute terms using NonLinLoc code (Lomax et al., 2000), based on a nonlinear inversion method. We used the same 1D gradient velocity model and setup used by Chiaraluce et al. (2017); we also included station corrections to counteract the effects of using an oversimplified velocity model. These preliminary locations already have good quality factors, as shown in Michele et al. 2019, with most events (~66%) included in the A and B quality classes (see Figure S1 in the supplementary material).

To further maximize the quality of the templates' catalog in terms of hypocentral location resolution, we apply a double difference (Waldhauser and Ellsworth, 2000) scheme taking only absolute travel times. In the relative relocation process, we used the same setup proposed by Michele et al., 2020 for relocating the 2016 aftershocks sequence. We kept all the M_L>1.5 events of the 2016 sequence together with the previous events to be relocated. In doing this, we additionally constrain the depth of the whole system (see Figure S2). We selected a well-constrained subset of events to obtain more robust quantitative estimates of coordinates location errors of the final catalog we used as input for the template matching. We relocated them using the full covariance matrix and singular value decomposition (SVD) instead of the weighted least squares (LSQR) method (see Waldhauser and Ellsworth, 2001).

TEXT S2

Template matching

Template matching is run to search eight years (2009-2016) of continuous data exploiting approximately 23,000 well-located earthquakes in Central Italy. Codes are rewritten from Vuan et al. (2018) to improve the performance and scalability, evaluate background seismicity, and analyze clustering before the 2016-2017 sequence in Central Italy.

The technical improvements needed to address massive computations involved: a) performance: ≈200% speedup in single-threaded mode, near-linear scaling using multiple threads. GPU support with further performance improvements: 50 templates per second per node with 4 GPU (NVIDIA V100), and higher speedups possible using longer signals. Faster post-processing thanks to AVRO/Parquet data serialization, b) usability: CLI, logging, input and output handling, arbitrary signal length, template duration (per channel basis), and sampling rate, c) robustness & correctness: better error handling, among fixed bugs: negative normalization in Obspy cross-correlation routine, drop multiple detections within template length, more stable magnitude estimation (missing data, template data and other bug fixes), detections at beginning or end of the signal, full usage of data (template/signal traces matching before processing), d) maintainability: less code (-80%)

and dependencies, enhanced readability (refactoring into functions, meaningful variable nomenclature), modularity, easier deployment via registered PyPI package.

Template matching is applied to daily three-component continuous waveforms covering the 2009/01/01-2016/08/24 time window. Seismic data from 2009 to 2016 are collected for 37 stations of the INGV seismic network (Figure 1 in the manuscript). We resample waveform data to 20 Hz and apply a 3-8 Hz bandpass filter. Templates are trimmed using a 5 s data window, starting 2.5 s before the theoretical S wave arrival, computed using the ObsPy port (Krisher et al., 2015) of the Java TauP Toolkit routines (Crotwell et al., 1999) and a suitable 1D-model (a modified version of Carannante et al., 2013). We adopt Kurtosis-based tests to evaluate the signal-to-noise ratio of templates (Baillard et al., 2014), avoiding unwanted signals in the matching technique (Vuan et al., 2018, Vuan et al., 2020).

A match, or detection, is a peak above a given threshold, set to 0.4, in the average of the stacked correlograms. In post-processing, some detections have been dropped based on the ratio between the average cross-correlation and the noise baseline level, estimated via the daily Median Absolute Deviation (MAD) of the correlograms. Keeping only the detections with a high ratio proved a robust method to exclude artifacts and false detections. The threshold for this ratio, defined after a visual inspection of some examples of detected events, was set to 18 times the MAD.

Time windows of 6 s are selected. Within each one, the template for which the normalized correlation coefficient is the greatest is taken to determine the event location and magnitude (e.g., Kato et al., 2012). In synthesis, in declaring a detection, we use very restrictive criteria: a) the average cross-correlation must be greater or equal than 0.4, b) it must be also greater or equal than 18 times the MAD, and c) at least 8 channels must have cross-correlation greater or equal to 0.4.

The location of the small events in the augmented catalog strictly depends on the quality of the input catalog locations and associated errors. Due to the high resolution of the starting catalog, we decided to keep the new detections co-located with the templates.

This choice of not relocating the new events is partly justified by the area dimensions (100 x 100 km), the seismic station's inter-distance relatively to the area under study, and the reduced number of earthquakes for which it is possible to obtain a more refined location. Ross et al. (2019), Simon et al. (2021), and Cabrera et al. (2022), by using different relocation tools, demonstrated that only a small portion of events from template matching

could be relocated (on average less than 20%). Moreover, relocation techniques based on limited frequency-band envelopes (e.g., Vuan et al., 2017) cannot improve the location of small magnitude events when the network coverage is sparse. These considerations led us to co-locate the new events at the respective template position.
To improve the robustness of the magnitude assessment, we removed the outliers in the pool of used channels (e.g., Ross et al., 2019).

Text S3
Clustering
A nearest-neighbor approach (Zaliapin and Ben-Zion, 2016) performs a statistical analysis of the augmented catalog to separate the background seismicity from the clusters. The nearest-neighbor method computes the time-space distance η between pairs of earthquakes. Rescaled time (T) and distance (R) between an event *i* and its parent *j* are normalized by the magnitude of *j* and expressed as:

$$T_{ij}=t_{ij}10^{-pbm_i/2};\ R_{ij}=r_{ij}^{d}10^{-(1-p)bm_i/2} \quad (1)$$

Where *p* is a weight parameter, *b* is the Gutenberg-Richter value, *m is* the magnitude of the *i* event, *t* and *r* are the time and distance between the two earthquakes, respectively, and *d* is the fractal dimension. We fixed *p*=0.5, *b*=0 (Zaliapin and Ben-Zion (2020) justifies to use *b*=0 for small events) and *d*=1.6. Thus, η, the generalized distance between pairs of earthquakes, is formulated as:

$$\mathrm{Log}\ \eta_{ij}=\log R_{ij}+\log T_{ij} \quad (2)$$

Subsequently, clusters are classified into swarms, mainshock-aftershock, and foreshock-mainshock sequences following the criterion proposed by Ogata and Katsura (2012). The mainshock is defined as the strongest earthquake in a cluster, and all the seismic events occurring before are pre-shocks. All pre-shocks are set foreshocks when the magnitude gap between the largest pre-shock and the mainshock is greater than 0.5. Unlike, a swarm-like sequence has pre-shocks with similar magnitudes (the difference is smaller than 0.5).

Further indications on clustering are also provided by comparing the covariance with the moment ratio variations. The covariance of interevent times (e.g., Kagan and Jackson, 1991) define the level of temporal clustering as close to 0 for periodic seismicity and greater than 1 for temporally clustered sequences. The seismic moment ratio (Vidale and Shearer, 2006) values provide indications when most of the seismic moment can be associated with a single event (values close to 1) or when no prevailing event is found (values relative to 0).

The covariance and moment ratio are evaluated using rolling windows of a variable number of events (50 to 150 with a step of 10). By shifting the rolling windows of 1 event and averaging the results, we can also derive the associated uncertainties.

Text S4

Repeating earthquakes

We search for repeating earthquakes, families of two or more events with nearly identical waveforms, locations, geometry, and magnitudes that repeatedly rupture the same fault patch at different times (e.g., Uchida 2019). We analyze the overall cross-correlation output of the template matching procedure, looking for couples of events characterized by a mean cross-correlation value >=0.90. Then we group events sharing common events and compile a list of candidate repeating earthquakes (CRE). Events span in magnitude range from about 0.4 to 2.6

This dataset is further investigated using a Python code for detecting true repeating earthquakes from self-similar waveforms that combines seismic waveform similarity by using cross-correlation (CC) and differential S-P travel times (Shakibay Senobari and Funning 2019; Sugan et al., 2022). Precise differential ΔS–P arrival times between CRE pairs are obtained by applying the cross-spectral method described in Poupinet et al. (1984). The spectral method is preferred since it allows sub-sample precision to resolve minimal source separation. We explore different CC and ΔS–P thresholds indicating possible RE that share at least 50% of the seismic source.

The code needs as input a CRE catalog, the associated seismic waveforms, the associated P and S picks, if any, and a simple 1D velocity model.

We use the original seismic waveforms sampled at 100Hz, and the travel-time phase file used to localize the templates. When this information is missed (e.g., for the new detections obtained with template matching), we use the theoretical arrival times calculated using a suitable velocity model or perform automatic picking using STA/LTA strategies (Carannante et al., 2013).

We explore different frequency configurations, considering up to four magnitude ranges. We set a minimum of three stations in the similarity space domain to declare a RE (Table 1). Cross-correlation values are calculated for a time window starting 0.5 seconds before the P arrival phase and ending 12 seconds after the S phases. Cross-spectrum is evaluated for P and S wave time windows with a length of about 1.4 s and 1.8 s, respectively.

We found a maximum number of 46 RE for configuration #1 and a minimum number of 6 RE for configuration #6. They are almost all doublets, characterized by short inter-even times (less than 10 hours); only one RE exceeds this value, with an inter event time of about 23 days (RE events occurred on 18 May and 10 June 2009).

In Figure 3, we show the results using configuration #5. The final RE’s location for each doublet is obtained using the mean value of the events' latitude, longitude, and depth.

		Conf#1	Conf#2	Conf#3	Conf#4	Conf#5	Conf#6
	SSD	***CC 0.95***			***CC 0.97***		
M	***ΔS–P (sec)***	BP(Hz)	BP(Hz)	BP(Hz)	BP(Hz)	BP(Hz)	BP(Hz)
0.5	***0.004***	1-30	1-35	1-40	1-30	1-35	1-40
1.0	***0.006***	1-25	1-30	1-35	1-25	1-30	1-35
1.5	***0.009***	1-20	1-25	1-30	1-20	1-25	1-30
2.0	***0.01***	1-15	1-20	1-25	1-15	1-20	1-25
	RE	46	37	28	17	7	6

Table 1 – *ΔS–P and CC* values used for #6 different configurations to define the similarity space domain (SSD). Different band-pass filters (BP) for different magnitudes (M) ranges are explored. For each configuration, the corresponding number of RE is shown.