Adinolfi, G. M., Cesca, S., Picozzi, M., Heimann, S., Zollo, A. (2019): Detection of weak seismic sequences based on arrival time coherence and empiric network detectability: an application at a near fault observatory. - *Geophysical Journal International*, *218*, 3, pp. 2054—2065.

DOI: http://doi.org/10.1093/gji/ggz248

Geophysical Journal International

Geophys. J. Int. (2019) **218,** 2054–2065 doi: 10.1093/gji/ggz248
GJI Seismology

Detection of weak seismic sequences based on arrival time coherence and empiric network detectability: an application at a near fault observatory

G.M. Adinolfi,[1] S. Cesca,[2] M. Picozzi,[1] S. Heimann[2] and A. Zollo[1]

[1]*Department of Physics "Ettore Pancini", University of Naples Federico II, Naples, Italy. E-mail: adinolfi@fisica.unina.it*
[2]*GFZ German Research Centre for Geosciences Potsdam, Potsdam, Germany*

Accepted 2019 May 29. Received 2019 May 15; in original form 2018 May 26

SUMMARY

Microseismic monitoring is a primary tool for understanding and tracking the progress of mechanical processes occurring in active rock fracture systems. In geothermal or hydrocarbon fields or along seismogenic fault systems, the detection and location of microseismicity facilitates resolution of the fracture system geometry and the investigation of the interaction between fluids and rocks, in response of stress field perturbations. Seismic monitoring aims to detect locate and characterize seismic sources. The detection of weak signals is often achieved at the cost of increasing the number of false detections, related to transient signals generated by a range of noise sources, or related to instrumental problems, ambient conditions or human activity that often affect seismic records. A variety of fast and automated methods has been recently proposed to detect and locate microseismicity based on the coherent detection of signal anomalies, such as increase in amplitude or coherent polarization, at dense seismic networks. While these methods proved to be very powerful to detect weak events and to reduce the magnitude of completeness, a major problem remains to discriminate among weak seismic signals produced by microseismicity and false detections. In this work, the microseimic data recorded along the Irpinia fault zone (Southern Apennines, Italy) are analysed to detect weak, natural earthquakes using one of such automated, migration-based, method. We propose a new method for the automatic discrimination of real vs false detections, which is based on empirical data and information about the detectability (i.e. detection capability) of the seismic network. Our approach allows obtaining high performances in detecting earthquakes without requiring a visual inspection of the seismic signals and minimizing analyst intervention. The proposed methodology is automated, self-updating and can be tuned at different success rates.

Key words: Computational seismology; Earthquake source observations; Induced seismicity; Seismicity and tectonics; Statistical seismology; Earthquake detection.

INTRODUCTION

The increasing demand for fast and automated seismic monitoring routines for earthquake hazard management and industrial applications stimulated in recent years the development of new, unsupervised and accurate methods for the detection and location of microseismicity. These techniques are developed to process large seismic data sets and detect low magnitude events. Typically, when processing microseismic data, seismicity detectors are prone to a certain number of false detections, which results from amplitude anomalies in seismic records, either because of instrument malfunctioning or due to simultaneous incoherent signals recorded at different stations. Obviously, it is desirable to minimize the percentage of false detections, and this percentage can be used to evaluate the detection algorithm performance. Thus, automated unsupervised seismicity detectors face the challenge of discriminating coherent seismic signals produced by weak seismic sources from comparable coherent seismic noise related ones, reducing the percentage of false detections.

Among different automated detectors, migration-based methods (also called coherence-based methods) have undergone a considerable development in recent years (e.g. among others, Cesca & Grigoli 2015). Using the full waveform information, these methods are based on the migration techniques extensively adopted in reflection seismic: the observed seismic waveforms are delayed and stacked to maximize the coherence of selected wave packets recorded at different stations. The migration-based methods have been used at various scales and context: from global seismology (Kruger & Ohrnberger 2005; Ekstrom 2006; Maercklin *et al.* 2012) to local microseismic monitoring (Whiters *et al.* 1999; Gharti *et al.*

2010; Grigoli *et al.* 2013; Zeng *et al.* 2014), but also for real-time applications in earthquake early warning (e.g. Picozzi *et al.* 2011; Meng *et al.* 2014). In all these methods, the first step towards the earthquake source location and detection is the computation of *characteristic* functions (CFs) that are non-linear combinations of seismic signals, which allows depicting changes of energy, frequency or polarization of the original signals with respect to the background noise. A variety of CFs have been used for earthquake detection purposes. The ratio between signal average amplitudes computed for windows of different length, generally referred as short-term average/long-term average (STA/LTA, Freinberger 1962; Allen 1978, 1982) is the most applied, but various alternative CFs in which higher-order statistics such as skewness and kurtosis of these functions (Saragiotis *et al.* 2002; Gentili & Michelini 2006; Kuperkoch *et al.* 2010; Baillard *et al.* 2013) have been also proposed (see Cesca & Grigoli 2015, for an overview). According to some authors, CFs based on signal energy offer better performances than STA/LTA in situations of low signal-to-noise ratio (SNR) or intense seismic activity (Grigoli *et al.* 2016). Independently on the adopted CFs, migration-based approaches detect and locate earthquakes by shifting in time and stacking the chosen CFs, or their STA/LTA ratio, for all the possible source locations at all stations. For the correct earthquake source location, the stacked waveforms provide the highest coherence. Migration based methods show several advantages: (1) phase identification and picking are not required, (2) high robustness against noisy data, (3) full automation, (4) fast processing and (5) high detection rate. Despite these advantages, waveforms stacking and coherence techniques show also some important drawbacks. Indeed, their performance strongly depends on the chosen detection threshold. In particular, these approaches exploit the coherence values as indicators for the detection quality, and the larger is the coherency the more likely the detected event is true one. Furthermore, the coherence value depends on the event magnitude. Large events are recorded by many stations with high-SNR, thus many traces contribute to the stacking of CFs, and the overall coherence is large. Weak events are instead recorded by few stations, and the coherence decreases. On the other hand, in order to recover weaker events and reduce the completeness magnitude of the catalogue, earthquake detectors are generally set with low coherence thresholds. While decreasing this threshold typically results in a high number of detections, it is well known that the results are certainly contaminated by higher numbers of false detections; requiring a post-processing phase to check and/or classify detected signals and to guarantee that each detection corresponds to a seismic event. Since the discrimination between real and false detections is a challenging task to accomplish for large data sets, it became a priority to implement strategies to facilitate this task and to improve the performances of detection algorithms.

In this work, we present an application of a coherence-based detection and location method applied to two seismic sequences recorded at a near-fault observatory along the Irpinia fault zone (Fig. 1), Southern Apennines (Italy), and the implementation and validation of a new method for the automatic discrimination of real against false detections, which considers empirical data and includes implicit information about the detectability of the seismic network. The detectability of a seismic network depends on several factors: distribution of stations, density of the network, site conditions, instrumental characteristics and post-processing methods of recorded data (Schorlemmer & Woesssner 2008). The approach that we propose is based on the idea that, while coherence is a robust metric for event detection and location, the information about the number and geometry of stations can be used for improving the discrimination of real and false events.

DATA

The Irpinia Fault System is a complex, fault structure area with extensional kinematics. It is characterized by high seismic potential and destructive earthquakes. The most recent large event, the *M*s 6.9, 1980 Irpinia earthquake, occurred along NW–SE striking faults with complex rupture characterized by three main episodes, producing about 3000 fatalities and severe damage (Bernard & Zollo 1989, Fig. 1). On the 3 April 1996, a M_L 4.9 earthquake occurred within the epicentral area of the 1980 earthquake; it showed a normal fault mechanism (Cocco *et al.* 1999) and was the largest one since 1980.

The seismicity in this area has been monitored since 2005 by a near-fault seismic observatory, the Irpinia Seismic Network (ISNet). ISNet is a dense seismic network made up of 32 stations, each equipped with three-component ground acceleration and velocity (short and broad-band) sensors. ISNet covers an area of about 100 km x 70 km, monitors the Campania-Lucania Apennines, in Southern Italy, and is operated by the Department of Physics of the University of Naples 'Federico II'.

Instrumentally recorded seismicity observed during the last 10 yr is comprised of low magnitude earthquakes ($M < 3$), mostly confined within a volume bounded by subparallel, NW–SE trending normal faults, following the Apenninic chain direction, as the segments of the fault system activated during the 1980 Irpinia earthquake (Bernard & Zollo 1989). The seismicity is distributed uniformly in the uppermost 15 km of the crust, with normal to normal-strike focal mechanisms, evidencing a NE–SW extension, consistent with the regional stress field in the Southern Appenines (De Matteis *et al.* 2012; Adinolfi *et al.* 2015). Seismicity frequently occurs in the form of microseismic sequences or swarms that last a few days with weak maximum magnitudes ($M < 3$). These repeated microearthquake sequences originate along zones of weakness and are characterized by colocated events with the same focal mechanism (Stabile *et al.* 2012). The seismic sequences suggest a similar stress loading-unloading mechanism spanning a range of spatial scales and distributed along the segmented normal fault system in the Southern Apennines, also related to the internal local stress variations and/or fluid migration along the fault zone near the geometrical barrier (Stabile *et al.* 2012).

In this study, we examine two microseismic sequences, both occurred along the Irpinia fault zone near the village of Lioni (Fig. 1). The first occurred on 2–3 August 2011 with a M_L 2.7 main shock happened on August 2nd at 20:14:17 (hereinafter, SS1). It consisted of 25 events with a minimum magnitude equal to M_L 0.3. The second sequence began with an event of M_L 2.8 on 16 July at 18:01:24 (hereinafter, SS2) and finished on 17 July 2017. It was composed of 16 earthquakes with a minimum magnitude equal to M_L 0.8 (http://isnet-bulletin.fisica.unina.it/cgi-bin/isnet-events/isnet.cgi).

In order to enhance the number of detected foreshocks and aftershocks, continuous seismic data streams were analysed for both seismic sequences for a period of 10 d, extending from 5 d before to 5 d after the occurrence of each main shock. We used continuous waveforms recorded at three-component velocimeters (1 Hz short-periods and 40 s broad-bands sensors) with a sampling frequency of 125 Hz. SS1 and SS2 data are acquired by 21 and by 10 seismic stations, respectively.

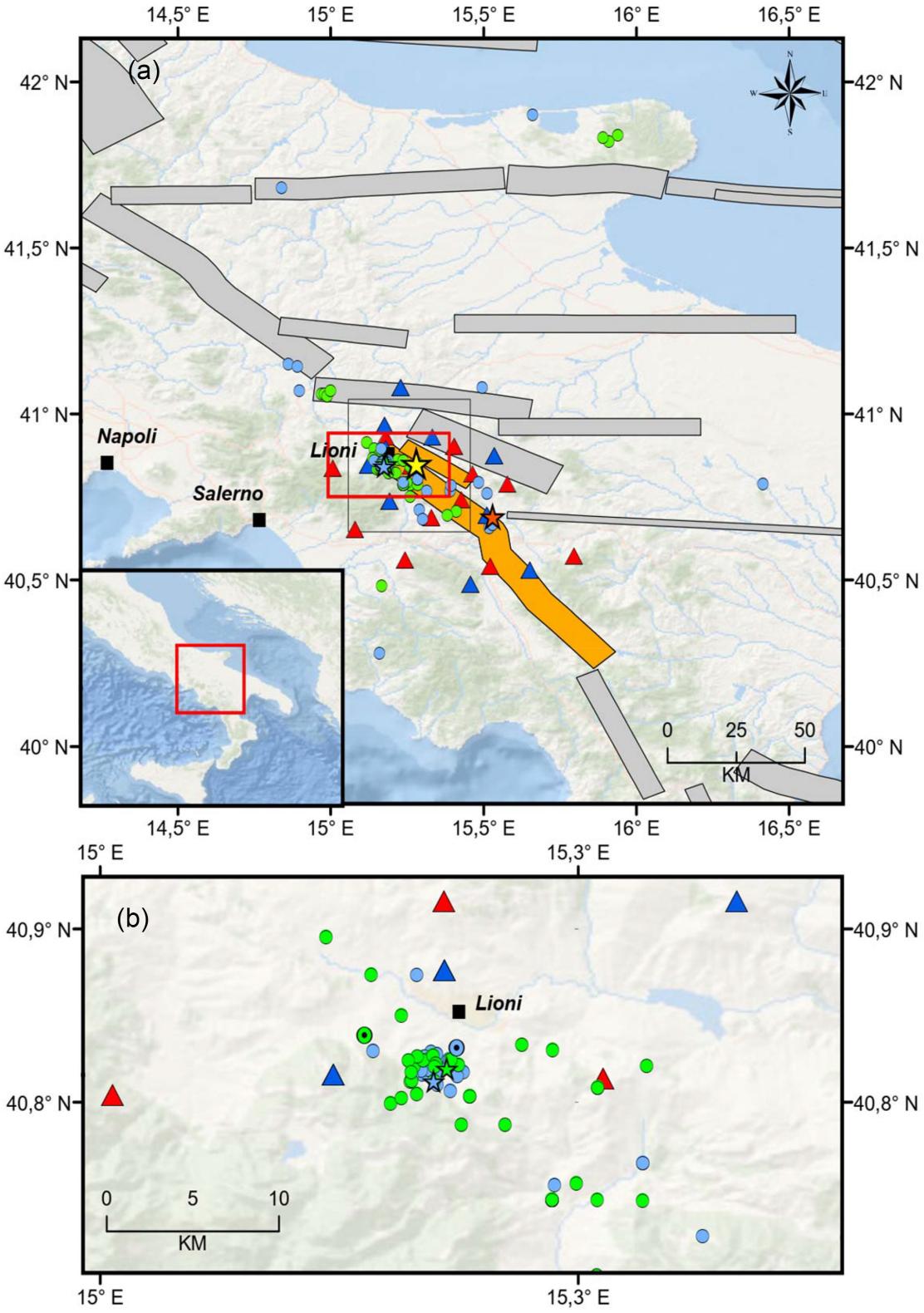

Figure 1. Epicentral map of the earthquakes analysed in this study. Green circles and star (mainshock) refer to SS1, light blue circles and star (mainshock) to SS2. The Irpinia Seismic Network (ISNet) is indicated with red and blue triangles. SS1 was recorded by seismic stations indicated in blue; SS2 was recorded by seismic stations indicated in red and blue. Yellow star refers to epicentral location of the 23 November 1980, Ms 6.9, Irpinia earthquake and orange star refers to the epicentral location of 3 April 1996, M_L 4.9 earthquake. Seismogenic sources related to Irpinia Fault System are indicated in orange; potential sources for earthquakes larger than M 5.5 of surrounding areas are indicated in grey (Database of Individual Seismogenic Sources, DISS, Version 3.2.1, DISS Working Group 2018). Black square shows the location of volumetric grid of potential hypocentres explored in this work. (b) Enlarged map of the red rectangle presented in (a). Earthquakes evidenced in (b) are used to calculate the scaling factor of the two seismic sequences.

METHOD

We used a recently developed detection and location algorithm (Lassie, https://gitext.gfz-potsdam.de/heimann/lassie, Heimann *et al.* 2017). Lassie exploits a migration-based technique that depicts coherent (*P* and *S* wave) arrivals at different stations for detecting and locating earthquakes (Matos *et al.* 2018; Lopez-Comino *et al.* 2017). In a pre-processing step, traveltimes for different seismic phases were computed for a 3-D grid (with size 20 × 20 × 20 km and 2 km grid spacing) of possible locations using the CAKE tool (Heimann *et al.* 2017) and using a local 1-D velocity model (Matrullo *et al.* 2013).

The detection process includes following steps (Figs 2a–d and S3):

(1) First, continuous waveforms are cut into overlapping time windows (width equal to 20 min with 20 per cent overlap). Waveforms of each time windows are used to compute the characteristic functions. In particular, each three-component recording is band-pass filtered using a 4th order Butterworth filter (1–15 Hz) and squared. The squared components are combined by summation into a single trace, which is proportional to the seismic energy. Next, the resulting trace is smoothed by convolving it with a Hann window. The length of the window should roughly correspond to the duration of the transient wave packets of interest (i.e. we used a length of 10 s). Finally, the signal is normalized by its moving average of length 50 s. This processing results in a smooth positive characteristic function (CF) for each station, which is sensitive to transient increases of seismic energy. We adopted a smooth CF to reduce the spatial and temporal sampling. This choice improves the computational performance and allows to process high sampling rate and large data set quickly (Lopez-Comino *et al.* 2017).

(2) The 3-D grid is scanned iterating over all the possible sources. For each location, CFs are shifted in time according to the traveltime corrections and stacked to form a so-called image function contribution (IFC):

$$IFC(k,t) = \sum_i CF_i(t + \Delta\tau_i(k)), \quad (1)$$

where k is the grid node index, t is the time sample and $\Delta\tau$ is the travel time correction for the ith station and kth node. The IFC is computed for each spatial location and time step. Multiple IFCs, for example derived by the detection of different seismic patterns and/or seismic phases, may be weighted and combined to build a global image function (i.e. an indirect measure of signals coherence). In this study, we worked with a single CF pre-processing scheme, but we used two different travel time corrections in the stacking, one for *P* and one for *S* phases, resulting in two distinct IFCs as defined in eq. (1). We combined the two IFCs using summation with equal weight. *P* and *S* traveltimes corrections generally improve the performance of the algorithm (Heimann *et al.* 2017; Lopez-Comino *et al.* 2017), even if comparable detection results can be obtained using only one seismic phase with smoothed CF (i.e. *S* wave, the most energetic phase; as shown in Table S1).

(3) For detection purposes, we considered the time dependence of the coherence. In particular, the events detection is performed by analysing the image function, which is a time-series built by considering at each time step the highest coherence value in the spatial grid. Hence, the events detection is achieved by searching for local maxima in the image function, whereas an event is declared whenever the image function is greater than a threshold value. Of course, in relative terms, higher coherence values reflect both a coherent seismic energy increase recorded at a larger number of sensors and a better SNR. However, the range of possible coherence values strongly depends on the network geometry, number of sensors and noise conditions. Therefore, the detection threshold is generally manually defined evaluating the real/false detections and their relating detection (coherence) value.

In this study, we carried out little adjustments to the Lassie code to get additional information. In particular, to automatically let it to discriminate between real from false detections, we implemented an empirical approach based on an assimilation principle. Our approach exploits the distribution of the interdistances of seismic stations for a set of real detections and leads us to define a seismic network detectability, which is in turn applied to new data. The network detectability is in general function of different parameters: seismic network density (i.e. number of stations and their areal distribution), geometrical characteristic of the seismic network, site conditions that account for velocity, attenuation model, seismic noise, waveforms characteristic that account for the magnitude and frequency content and time. The detectability is a time-dependent property, mainly due to possible changes in operational conditions with time (i.e. different number of operating stations, network density and geometrical characteristic) that the seismic networks could experiment and variations in the environmental seismic noise conditions. Finding a theoretical relation or expression that accounts for all these variables is a difficult task, since it would require also an accurate knowledge of the medium. For this reason, we decided to exploit directly the empirical data characteristics to assimilate information on the network capability of detecting earthquakes. To this purpose, we investigated the relation between the number and the mutual distances of stations contributing to real detections and the relative coherence values.

We define as 'triggered' the seismic stations that mostly contribute to determine the coherence value for a seismic event. To do this, we single out the stations having a CF amplitude greater or equal than the average CF value computed for all stations. In other words, triggered stations correspond to those mostly contributing to the stack of CFs and, consequently, to the final value of the IFC. In the case of detections of earthquakes, as expected, we observe that seismic stations located close to the epicentre are associated to larger CF amplitudes than those at greater distance. Hence, for each detection, we defined the triggered stations interdistance (TSI; see Supporting Information) as the mean value of their distances from their spatial barycentre. Larger magnitude events are typically well-recorded at many seismic stations, even at large epicentral distances; and therefore, we obtain for them large TSI values. On the contrary, smaller magnitude events are recorded with good SNR only by few stations near the epicentre, leading to small TSI. In the case of false detections due to natural/anthropic seismic sources or instrumental malfunctioning, simultaneous spurious signals may lead to high IFC, but the distribution of triggered stations can be spatially sparse and not clustered like in case of earthquakes. Hence, while the coherence value associated to a false detection can be comparable the one of weaker earthquakes, we expect that the TSI for false detection will be in general larger (i.e. the triggered stations for a false detection are not spatially clustered). The TSI depends on earthquake location with respect to seismic network and its geometry, and is affected by the ambient noise level too. The TSI is therefore a parameter potentially capable of providing useful information about the network performance. In the following, we illustrate how the TSI can be used as additional parameter towards the automated discrimination of real and false detections, especially when approaching the

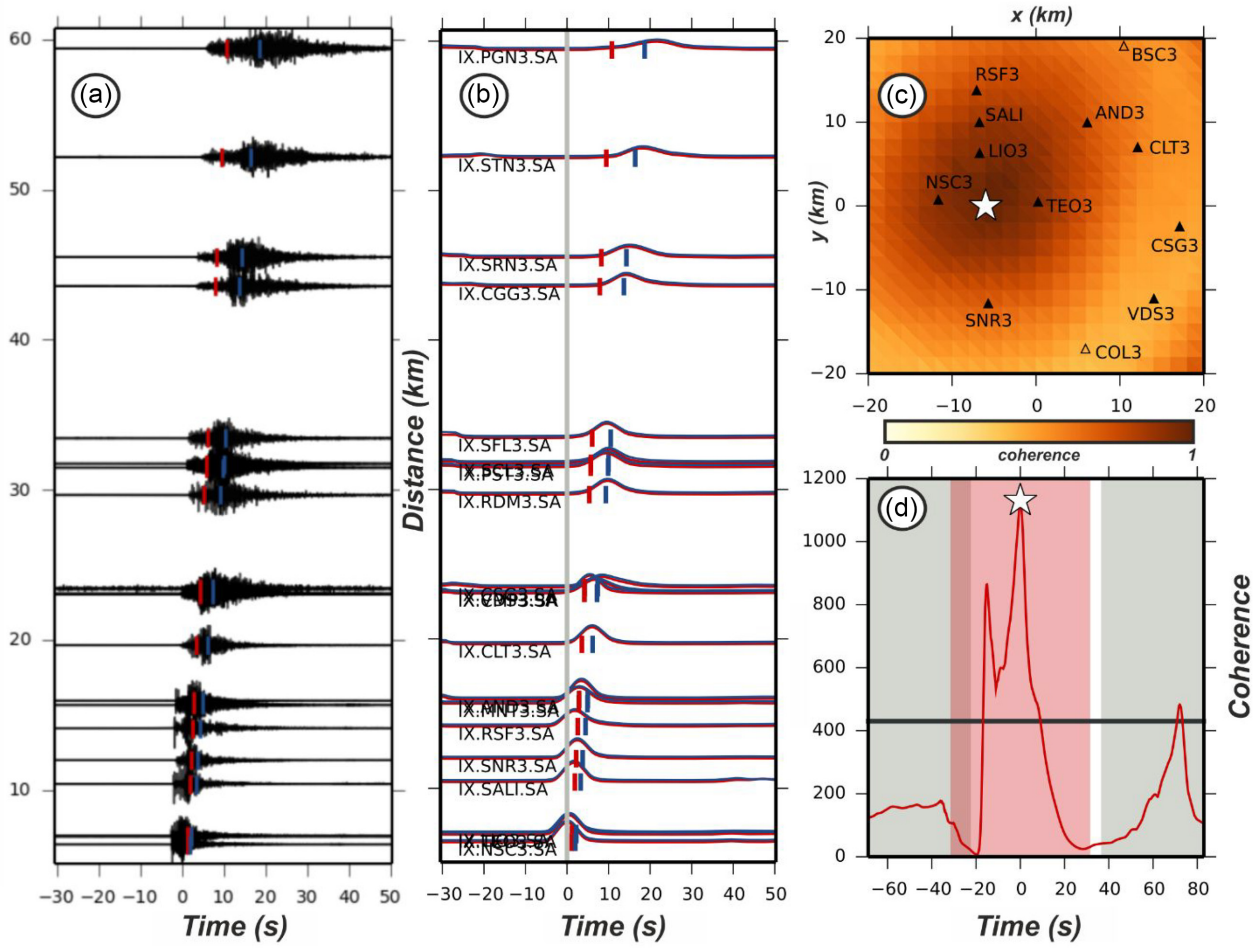

Figure 2. Example of detection belonging to SS1. The earthquake (M_L 2.7) occurred on the 2011-08-02 at 20:14:17 (UTC) nearby the Lioni village. (a) Waveforms used for detection, sorted by hypocentral distance. (b) Characteristic functions (normalized amplitude) calculated for each station. They are corrected with *P*-wave velocity (red lines) and *S*-wave velocity (blue lines) for the travel time and stacked to obtain the image function (d). The markers indicate the best fit of synthetic arrival time for *P* and *S* phase. (c) Stack of the coherence map for the search region with seismic stations (black triangle) and event detected (white star). The colour bar shows coherence values. (d) Image function corresponding to the best fit of source position along a processing time window centred on the origin time of the detected earthquake. The white star indicates the detected event above a fixed threshold value (black line).

coherence threshold, where real and false detections mostly overlap (Fig. 3 and Figs S1 and S2).

We first took into consideration the Lassie's detections for the second earthquakes sequence (SS2) and analysed the TSI and coherence. In particular, a binned correlation matrix between coherence and TSI was calculated, estimating for each bin the real detection frequency, which we define as the number of real detections divided by the number of total detections (i.e. real plus false). A detection is defined 'real' or 'false' after a visual inspection of the recordings by an operator. We selected a grid sampling of 10 units for the coherence and 1 km for the TSI. Finally, the binned correlation matrix was spatially interpolated for computing correlation values also at those bins for which detections were not available (see Section 'Results and Discussion'). The derived coherence–TSI detection matrix, which is obtained from the analysis of a training data set (hereinafter SS2) and accounts for the real detections performance of the considered seismic network (i.e. it considers its geometry and data quality) can be used to discriminate true/false detections in new data sets. Indeed, real detections can be automatically identified by extracting all the detections that fall into the coherence–TSI bins with values larger than a desired threshold.

In order to carry out a performance analysis and to validate our method, for both the considered data sets all the detections were visually inspected and in the case of earthquakes their location and magnitude were estimated. We point out that in this study we did not use the preliminary locations provided by Lassie, which are obtained by the maximum value of coherence and the travel-time stacking after the *P*- and *S*- wave velocities corrections. The reason for this choice is that our main goal here was to accomplish the earthquakes detection being as fast and accurate as possible. Hence, we decided to adopt a smooth CF which allows good performance in detecting energetic signals at the cost of a larger location uncertainties. Our idea is that, once the detection of earthquakes is carried out with the Lassie's configuration used in this study, a refined analysis on the data set of trusted events with fine-tuned CFs capable of better identifying *P* and *S* waves arrival times can be easily implemented.

Therefore, for the location of earthquakes occurred inside the ISNet seismic network, we manually picked the *P* and *S* wave

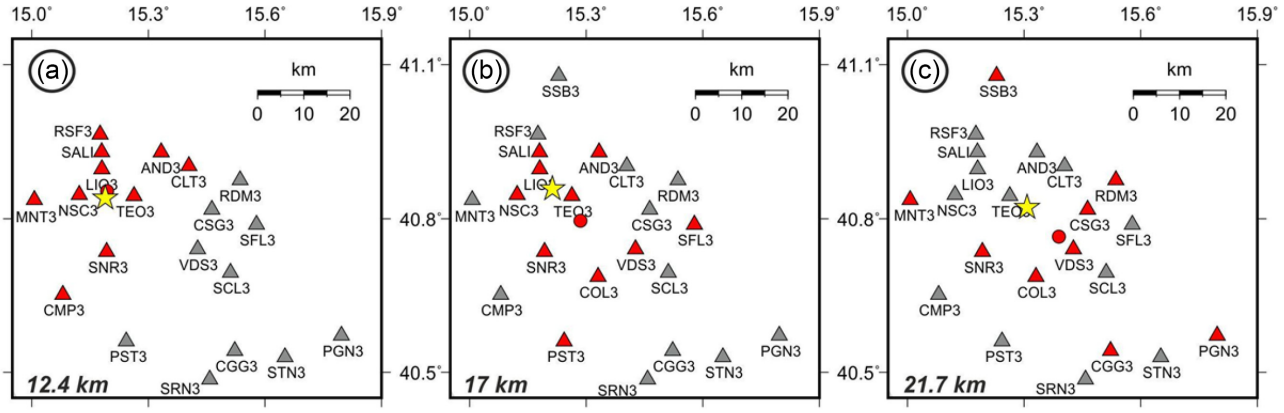

Figure 3. Plot of triggered (red triangles) and not-triggered (grey triangles) stations, their barycentre (red circles) with the location of detected event (yellow star) for the example shown in Fig. 2 (a), for a case of real detection (b) and false detection (c) of comparable coherence value. For each case, the 'TSI' values are indicated. More details are reported in the supporting information.

arrivals and we used the NonLinLoc software (Lomax *et al.* 2000) with a local velocity model (Matrullo *et al.* 2013). The local magnitude was estimated following Bobbio *et al.* (2009). The hypocentre parameter and local magnitude for the earthquakes outside the network were extracted from the INGV Bulletin web page (http://cnt.rm.ingv.it/en).

RESULTS AND DISCUSSION

Full waveform detections for both SS1 and SS2 were carried out with a threshold value (IFC) equal to 200. We adopted such rather low threshold value aiming to obtain large numbers of detections which can be analysed in terms of TSI. For SS1, 147 detections were found, with 51 real earthquakes after visual inspection (Figs 4a and S4a). For these real detections, the coherence ranges between 265 and 1235, while the TSI ranges from 12.4 and 22.7 km. For SS2, we visually identified 52 real events out of 316 detections (Figs 4b and S4b). The coherence of real detections varies here between ∼200 and ∼700 and the TSI between 12.0 and 23.5 km. The detected seismicity is not limited to the Lioni seismic sequences, but also includes the background seismicity occurred inside and outside the seismic network.

As shown in Fig. 4(a), the selection of an IFC threshold equal to 400 would lead to avoiding all false events, but at the price of reducing the number of detected events to only 13. A smaller IFC threshold equal to 300 would instead introduce some false detections. We observe a 'transitional' zone between 300 and 400 in which both real and false detections are concomitant due to low SNR or coherent noise at near stations. These results confirm that an event-discrimination strategy based only on coherence thresholds is not optimal to classify real/false events and it would force the operators to look for a compromise between the need of detecting the microseismicity with the lowest possible magnitude and the need of avoiding the interpretation of false detections as real earthquakes.

Figs 5(a) and (b) shows the coherence–TSI matrix computed for the SS2 data set. The contour lines represent the real detection frequency (RDF, i.e. the number of real detections divided by the number of total detections) as a function of coherence and TSI. The coherence grows as TSI grows; so, once the TSI is calculated, the coherence threshold can be used for discriminating among real and false events, or to identify transitional zones that need special attention. The coherence–TSI matrix provides empirical thresholds for the discrimination of real/false events with different confidence levels. Our aim was to verify if such a matrix could be used as a tool for an automatic real event discrimination. To this purpose, we tested this strategy on the SS1 dataset, using as discriminator the real detection coherence–TSI matrix obtained for SS2. Before running the discrimination analysis, we must consider that any change in the number and geometry of operating stations could impact on the coherence– TSI matrix, and therefore on the discrimination results. Indeed, if the number of triggered stations changes, considering that the IFC functions (eq. 1) are calculated by stacking a different number of waveforms, the coherence level threshold should change too. To correct for this effect, we select a couple of earthquakes from the two seismic sequences SS1 and SS2 having similar magnitude and location (M_L 1.0, Lioni on 16 July 2017, at 18:18:55 and M_L 1.0, Lioni on 2 August 2011, at 20:38:50; Fig. 1b) and that were recorded by a different number of stations. Therefore, we scaled the coherence values of SS1 (i.e. the validation dataset) by a factor equal to the coherences ratio of the two considered earthquakes (i.e. the correcting factor is 0.5; Fig. S5). We estimated such correcting factors for three more suitable earthquake couples, obtaining a mean value of 0.5.

The coherence–TSI matrix defined for SS2 is then applied to the normalized values of SS1. We tested the performance of our approach for different RDF values as threshold. We compared the success rate of our automatic and empirical discriminator for RDF equal to 0.8, 0.5 and 0.3 against the effective number of real detections from visual inspection (number of real detections, correctly identified, divided by the total number of real detections for a considered RDF threshold as percentage). Setting a RDF > 0.8, we identified 31 detections, with 30 real detections and only 1 false detection (Fig. 6a). For a RDF > 0.5, we found 38 detections, with 35 real detections and 3 false detections (Fig. 6b). For a RDF > 0.3, we obtained 48 detections, with 40 real and 8 false detections (Fig. 6c). These results show that our approach select empirically real earthquakes from false detections with a 96.8, 92.1 and 83.3 per cent of success rate, respectively. We obtained a percentage increase in the number of automated earthquake detections equal to 43 and 76 per cent compared to the INGV and the ISNet catalogues in the first case (RDF > 0.8), to 66 and 106 per cent in the second case (RDF > 0.5), and to 90 and to 135 per cent in the third case (RDF > 0.3). These

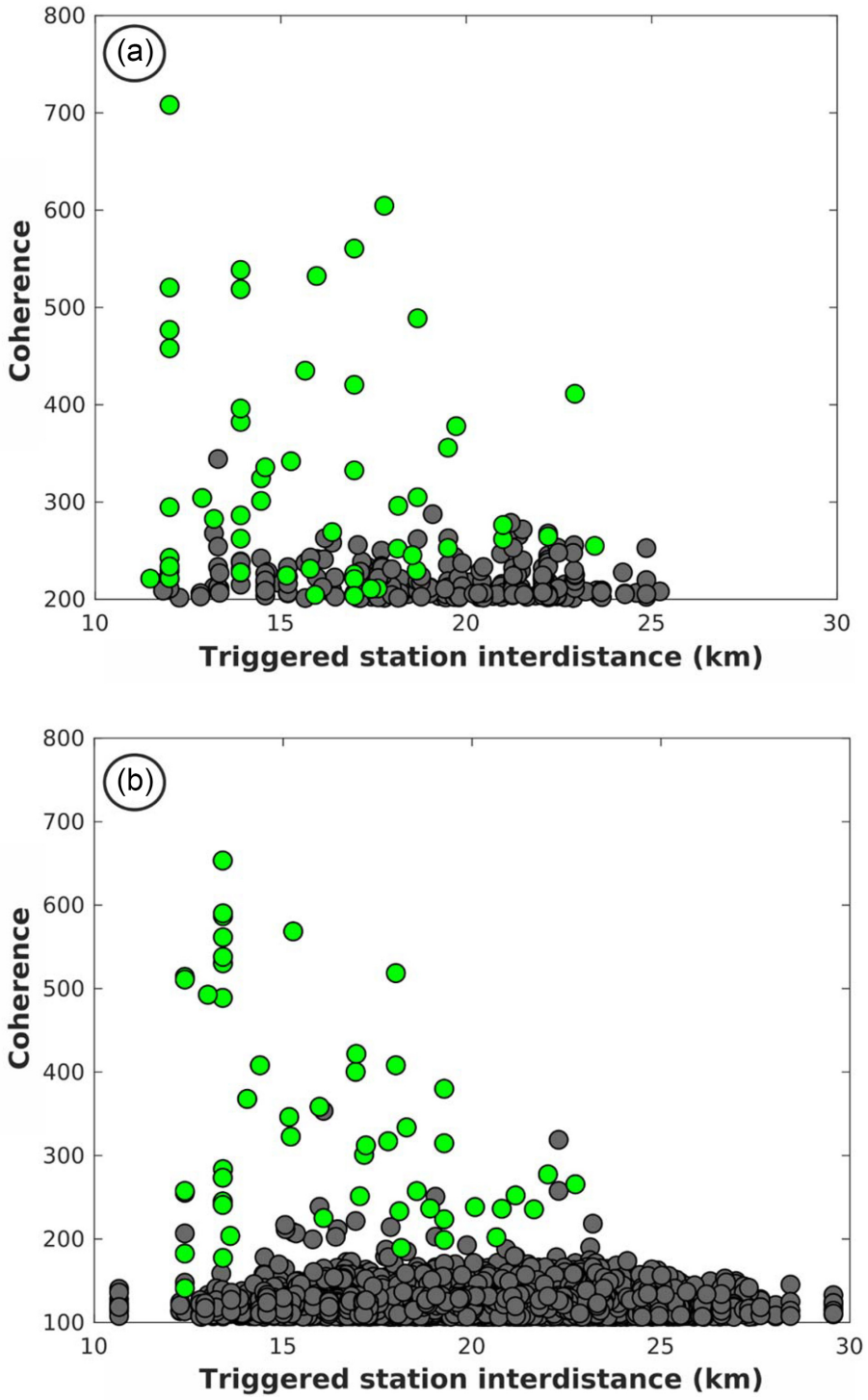

Figure 4. Coherence versus Triggered station interdistance (TSI) for SS2 (a) and for SS1 (b). Real detections are indicated in green and false detections are indicated in grey. In (b) the coherence values are multiplied for a scaling factor equal to 0.5.

results demonstrate the potential of our approach for the automatic classification of large detection catalogues (Table S2).

Considering the magnitude and the hypocentres of the detected earthquakes, our results agree with the ISNet performance estimated by Stabile *et al.* (2013). Although the present network configuration is slightly different from that considered by Stabile *et al.* (2013), we detected and located earthquakes at Lioni with a minimum magnitude (M_L) of 0.3. Other detected earthquakes, probably with smaller magnitude, could not be located because they did not meet the minimum location criteria (3 *P*-wave + 1 *S*-wave arrival times), in accordance with the detection and location thresholds indicated for ISNet by Stabile *et al.* (2012).

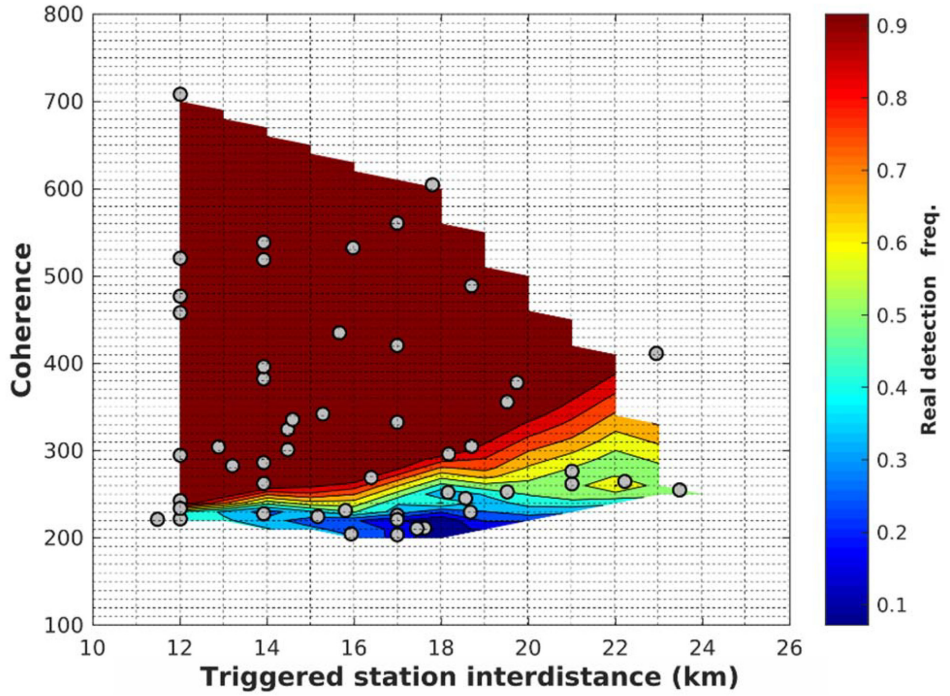

Figure 5. Coherence–TSI matrix calculated for SS2. Bins used for the linear interpolation are indicated. The linear interpolated matrix is shown with real detections used (grey circles). The colours refer to different values of the frequency of real detections calculated for each bin.

The coherence–TSI matrix derived using a training dataset represents an empirical measure of the network detectability. Its usage for the processing of new data sets is only conditioned to the seismic network characteristics of the training dataset. If we look at the trend of isolines, we see that the coherence grows as TSI grows (Fig. 5b); this is due to the attenuation experimented by the seismic wavefield. As an effect, small earthquakes will be recorded by few stations, located close to epicentre, while large earthquakes will be recorded by a larger number of stations with higher interdistances. Our approach empirically defines the range of station interdistance at which the analysed seismicity is well recorded by a specific network configuration.

If the seismic network experiences a variation of its operating conditions (e.g. change in the number or in the distribution of seismic stations), the coherence-based detection method can be easily recalibrated by choosing a new threshold, suitable for the modified geometry and numbers of stations. A further useful aspect of the coherence–TSI matrix is that different thresholds can be selected while analysing the earthquake detections results, modulating the chance of detecting false events and the ratio of real/false detections. If an operator wishes to reduce or exclude false detections, a high threshold can be selected; while, if it is necessary to focus the attention on smaller earthquakes that are located at the limit of the detectability of seismic network, a smaller threshold can be selected. If the geometrical configuration of the seismic network is constant and the ambient noise and the physics proprieties of the medium are in first approximation stationary, the coherence–TSI matrix will depend only on the seismicity characteristics (location and magnitude distributions).

The approach proposed in this study is data- and network-driven, in the sense that the method performance is driven by the available data and the network geometry. Furthermore, the approach is assimilative, in the sense that the method detection performance will improve in time with the increase of the data set used for training the system. Indeed, the coherence–TSI matrix can be easily updated with new recorded data and real/false detection results so that its power in discriminating between the real/false detections, especially at 'transitional zone', will increase with time.

One of the problems shown by coherence-based earthquake detectors is related to earthquakes occurring at the edge or outside the network. As shown in Fig. 7, the real detections show two distinct trends with respect to the coherence. The earthquakes occurring inside the seismic network show coherences that linearly grow with magnitudes with a lower gradient with respect to earthquakes occurring outside the network (i.e. these latter show a higher slope). Even if characterized by larger magnitudes, earthquakes occurring outside of the network have *P*- and *S*-waves signals with lower SNR than the within network events due to the larger epicentral distances and the larger seismic phases attenuation, which leads to low coherences and high stations interdistance. For instance, considering the real detection of SS1, earthquakes located at the border or outside the seismic network show a TSI above 20 km. Therefore, combining the information of coherence, magnitude and TSI, it is possible to estimate a range of stations interdistance and coherence values to further discriminate earthquakes occurring inside or outside the network.

CONCLUSIONS

A coherence-based earthquake detector was applied to two microseismic sequences ($M_L <= 2.8$) at the Irpinia Fault System (Southern Italy). The earthquakes were detected with a migration-based technique using coherent *P*- and *S*-wave arrivals recorded by the dense seismic network operating in Irpinia and assuming a local, 1-D velocity model. A new strategy was proposed to automatically discriminate between real and false detections. The triggered

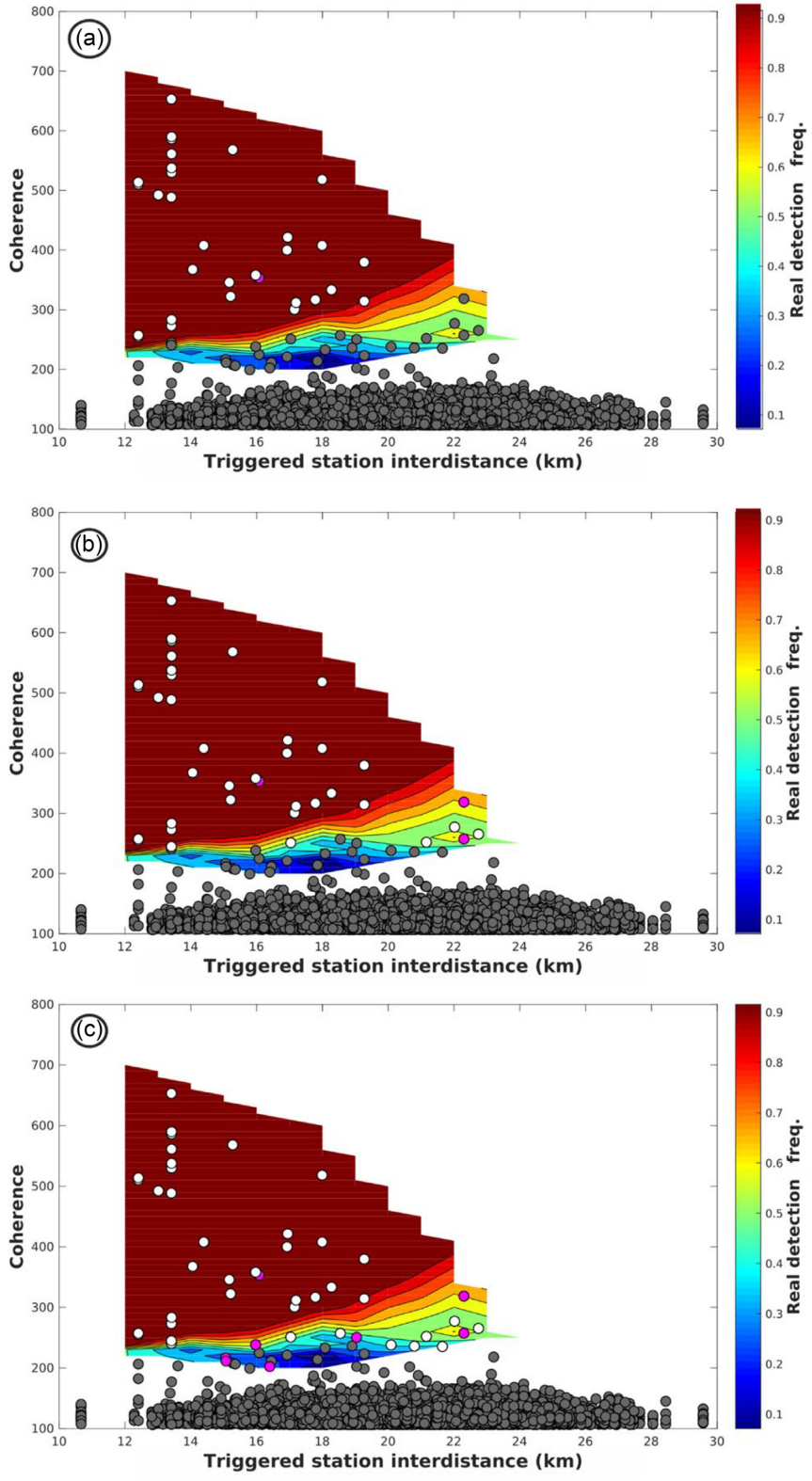

Figure 6. Results for SS1 data using threshold larger than 0.8 (a), 0.5 (b) and 0.3 (c). The coherence–TSI matrix used for the selection of real/false detections is shown. Grey circles refer to false detections; white circles refer to real detections, correctly selected for the relative threshold; fuchsia circles refer to false detections, incorrectly selected as real earthquakes.

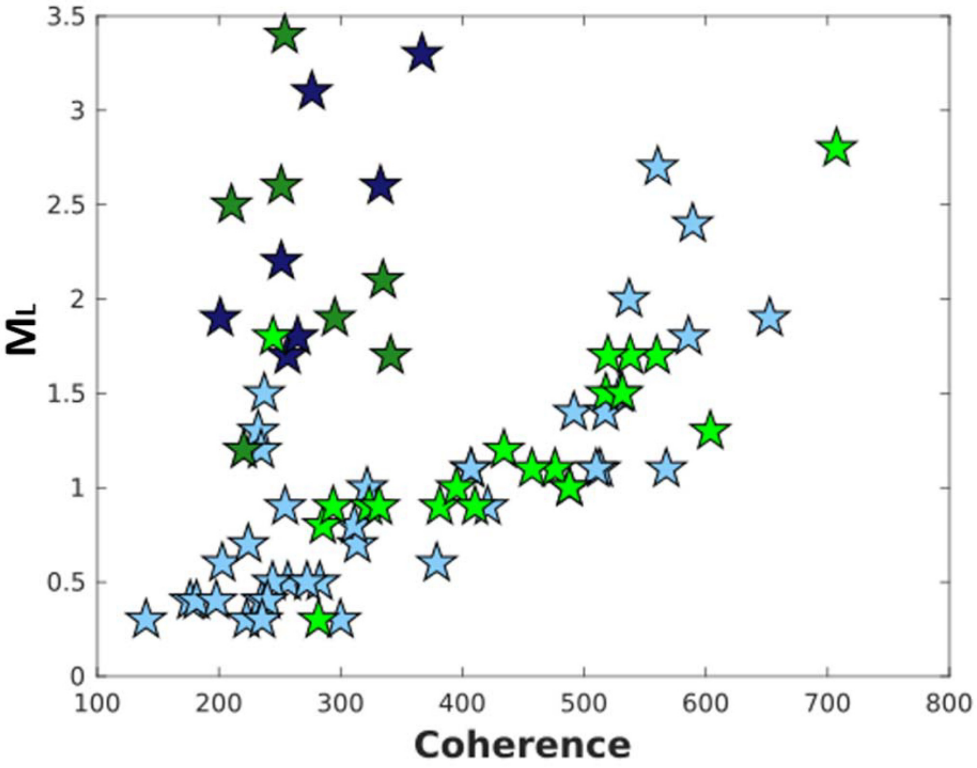

Figure 7. Local magnitude versus coherence for SS1 (green stars) and SS2 (blue stars). Light-blue and light-green stars showing a linear trend with a smaller slope correspond to earthquakes occurred inside the seismic network. Dark-green and dark-blue stars refer to earthquake located outside the network.

station interdistance (TSI) information was included in the earthquake detection strategy; a coherence–TSI matrix was empirically calculated and applied to real data. Our results show that the proposed approach empirically selects real earthquakes from overall detections with 96.8, 92.1 and 83.3 per cent of success rate, corresponding to a real detection frequency (RDF) greater than 0.8, 0.5 and 0.3, respectively.

The new proposed detecting strategy show several advantages:

(1) Based on empirical data, the proposed tool is data-driven and automatic in selecting real/false detections without the requirement of visual inspection of the waveforms.

(2) Selecting different RDFs, a false alarm versus missed events threshold may be selected according to temporal changes in noise level or network performance, giving the opportunity to be more/less conservative in reducing the number of possible false detections.

(3) The detection results derived using different geometric configurations of the seismic network can be corrected to be compared to one other.

(4) The coherence-triggered station interdistance matrix can be easily updated and refined with the incoming of new datasets.

The limit of our empirical procedure is related to the 'resolution' of the coherence-triggered station interdistance matrix, which especially in the initial phases is influenced by the characteristics of the training dataset. In fact, assuming the ambient noise and the geometric characteristics of the seismic network constant in time, the method resolution depends on how much the data analysed is statically representative of the seismicity in the considered seismogenic volume. Therefore, statistically representative dataset are required; for instance, they can be obtained updating the coherence-triggered stations interdistance matrix whenever new detection results are available.

Our empirical procedure can be improved in the next development combining different types of information or new coherence-based matrices (i.e. coherence–magnitude matrix) with the aim to correctly identify the earthquake occurrence, further reducing the number of false detections.

ACKNOWLEDGEMENTS

We thank editor M. Schimmel for his constructive and fruitful comments. We thank two anonymous Reviewers for their valuable suggestions which improved the manuscript.

The work was supported by Compagnia di San Paolo and University of Naples, Federico II, Italy, Programme STAR L1-2014. The work was also financed by Italian Ministry of Economic Development and Polytechnic and Basic Sciences School, University of Naples, Federico II, 2017-2018 Operational Agreement.

REFERENCES

Adinolfi, G.M.,De Matteis, R., Orefice, A., Festa, G., Zollo, A., de Nardis, R. & Lavecchia, G. , 2015. The September 27, 2012, ML 4.1, Benevento earthquake: a case of strike-slip faulting in Southern Apennines (Italy), *Tectonophysics,* **660,** 35–46.

Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.,* **68,** 1521–1532.

Allen, R.V., 1982. Automatic phase pickers: their present use and future prospects, *Bull. seism. Soc. Am.,* **72,** 225–242.

Baillard, C., Crawford, W.C., Ballu, V., Hibert, C. & Mangeney, A., 2013. An automatic kurtosis-based P-and S-phase picker designed for local seismic networks, *Bull. seism. Soc. Am.,* **104**(1), 394–409.

Bernard, P. & Zollo, A., 1989. The Irpinia (Italy) 1980 earthquake: detailed analysis of a complex normal faulting, *J. geophys. Res.: Solid Earth,* **94**(B2), 1631–1647.

Bobbio, A., Vassallo, M. & Festa, G., 2009. Local Magnitude estimation for the Irpinia Seismic Network, *Bull. seism. Soc. Am.* **99,** 2461–2470.

Cesca, S. & Grigoli, F., 2015. Full waveform seismological advances for microseismic monitoring, In *Advances in Geophysics,* Vol. **56,** pp. 169–228, Elsevier.

Cocco, M., Chiarabba, C., Di Bona, M., Selvaggi, G., Margheriti, L., Frepoli, A. & Campillo, M., 1999. The April 1996 Irpinia seismic sequence: evidence for fault interaction, *J. Seismol.,* **3**(1), 105–117.

De Matteis, R., Matrullo, E., Rivera, L., Stabile, T.A., Pasquale, G. & Zollo, A., 2012. Fault delineation and regional stress direction from the analysis of background microseismicity in the southern Apennines, Italy, *Bull. seism. Soc. Am.,* **102**(4), 1899–1907.

DISS Working Group, 2018. Database of Individual Seismogenic Sources (DISS), Version 3.2.1: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. http://diss.rm.ingv.it/diss/, Istituto Nazionale di Geofisica e Vulcanologia; doi:10.6092/INGV.IT-DISS3.2.1.

Ekstro¨m, G., 2006. Global detection and location of seismic sources by using surface waves, *Bull. seism. Soc. Am.,* **96**(4A), 1201–1212.

Freiberger, W.F., 1962. An approximate method in signal detection Technical Report No. 12, Department of the Army, Division of Applied Mathematics, Brown University.

Gentili, S. & Michelini, A., 2006. Automatic picking of P and S phases using a neural tree, *J. Seismol.,* **10**(1), 39–63.

Gharti, H.N., Oye, V., Roth, M. & Kühn, D., 2010. Automated microearthquake location using envelope stacking and robust global optimizationAutomated microearthquake location, *Geophysics,* **75**(4), MA27–MA46.

Grigoli, F., Cesca, S., Krieger, L., Kriegerowski, M., Gammaldi, S., Horalek, J. & Dahm, T., 2016. Automated microseismic event location using master-event waveform stacking, *Sci. Rep.,* **6,** 25744.

Grigoli, F., Cesca, S., Vassallo, M. & Dahm, T., 2013. Automated seismic event location by travel-time stacking: an application to mining induced seismicity, *Seismol. Res. Lett.,* **84**(4), 666–677.

Heimann, S. *et al.*, 2017. Pyrocko—an open-source seismology toolbox and library, *GFZ Data Services*, Potsdam, doi:10.5880/GFZ.2.1.2017.001.

Krüger, F. & Ohrnberger, M., 2005. Tracking the rupture of the M w = 9.3 Sumatra earthquake over 1,150 km at teleseismic distance, *Nature,* **435**(7044), 937.

Küperkoch, L., Meier, T., Lee, J. & Friederich, W. EGELADOS Working Group, 2010. Automated determination of P-phase arrival times at regional and local distances using higher order statistics, *Geophys. J. Int.,* **181**(2), 1159–1170.

Lomax, A., Virieux, J., Volant, P. & Thierry, B.C., 2000. Probabilistic earthquake location in 3D and layered models: introduction of a Metropolis–Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location,* pp. 101–134, eds Thurber, C.H. & Rabinowitz, N., Kluwer Academic, Amsterdam.

López-Comino, J.A., Cesca, S., Heimann, S., Grigoli, F., Milkereit, C., Dahm, T. & Zang, A., 2017. Characterization of hydraulic fractures growth during the Äspö Hard Rock Laboratory experiment (Sweden), *Rock Mech. Rock Eng.,* **50**(11), 2985–3001.

Maercklin, N., Festa, G., Colombelli, S. & Zollo, A., 2012. Twin ruptures grew to build up the giant 2011 Tohoku, Japan, earthquake, *Sci. Rep.,* **2,** 709.

Matos, C., Custódio, S., Batlló, J., Zahradník, J., Arroucau, P., Silveira, G. & Heimann, S., 2018. An active seismic zone in intraplate west Iberia inferred from high resolution geophysical data, *J. geophys. Res.: Solid Earth,* **123**(4), 2885–2907.

Matrullo, E., De Matteis, R., Satriano, C., Amoroso, O. & Zollo, A., 2013. An improved 1-D seismic velocity model for seismological studies in the Campania–Lucania region (Southern Italy), *Geophys. J. Int.,* **195**(1), 460–473.

Meng, L., Allen, R. & Ampuero, J.-P., 2014. Application of seismic array processing to earthquake early warning, *Bull. seism. Soc. Am.,* **104**(5), 2553–2561.

Picozzi, M., Milkereit, C., Fleming, K., Cakti, E. & Zschau, J., 2011. A generalized zero-lag cross-correlation approach for Rapid Earthquake Localization (REL): the example of the Istanbul Megacity Rapid Response System, *J. Seismol.,* **15,** 557–578.

Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI-S/K: A robust automatic seismic P phase arrival identification scheme, *IEEE Trans. Geosci. Rem. Sens.,* **40**(6), 1395–1404.

Schorlemmer, D. & Woessner, J., 2008. Probability of detecting an earthquake, *Bull. seism. Soc. Am.,* **98**(5), 2103–2117.

Stabile, T.A., Iannaccone, G., Zollo, A., Lomax, A., Ferulano, M.F., Vetri, M.L.V. & Barzaghi, L.P., 2013. A comprehensive approach for evaluating network performance in surface and borehole seismic monitoring, *Geophys. J. Int.,* **192**(2), 793–806.

Stabile, T.A., Satriano, C., Orefice, A., Festa, G. & Zollo, A., 2012. Anatomy of a microearthquake sequence on an active normal fault, *Sci. Rep.,* **2,** 410.

Withers, M., Aster, R. & Young, C., 1999. An automated local and regional seismic event detection and location system using waveform correlation, *Bull. seism. Soc. Am.,* **89**(3), 657–669.

Zeng, X., Zhang, H., Zhang, X., Wang, H., Zhang, Y. & Liu, Q., 2014. Surface microseismic monitoring of hydraulic fracturing of a shale-gas reservoir using short-period and broadband seismic sensors, *Seismol. Res. Lett.,* **85**(3), 668–677.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. The same as in Fig. 2, but for another example of real detection. (c) Map of spatial coherence variation for the search region with seismic stations (black triangle) and event detected (white star).

Figure S2. The same as in Fig. 2, but for an example of false detection. (c) Map of spatial coherence variation for the search region with seismic stations (black triangle) and event detected (white star).

Figure S3. The same as in Fig. 2, but with the map spatial coherence variation (c) for the search region with seismic stations (black triangle) and event detected (white star).

Figure S4. Coherence versus Triggered stations interdistance (TSI) for SS2 (a) and for SS1 (b). Real detections are indicated in green and false detections are indicated in grey. In (b) the coherence values are multiplied for a scaling factor equal to 0.53. Earthquakes that did not met the minimum location criteria (3 *P*-wave + 1 *S*-wave arrivals) are evidenced; for these earthquakes, location and magnitude estimate could not be calculated.

Figure S5. Coherence versus Triggered stations interdistance (TSI) for SS1. Real detections are indicated in light green (not scaled values), green (scaled values using a correcting factor = 0.5) and false detections are indicated in grey. Since the number and geometry of operating stations are different, we scale the coherence values of the target SS1 for a factor equal to the ratio of the coherences for the two earthquakes with same location and magnitude.

Table S1. Earthquake detections assuming both *P* and *S* phases (evidenced in yellow; as calculated in our work) and assuming only *S* phases (evidenced in green) for the calculus of CF. A period of three days was considered for a test. Number of detections, number and interdistance of 'triggered' stations are reported.

Table S2. Earthquakes detected and located in this study for SS1, evidenced in light blue, and for SS2, evidenced in yellow. For each earthquake, it is reported if INGV or ISNet detected the earthquakes. For earthquakes that did not met the minimum location criteria (3 *P*-wave + 1 *S*-wave arrivals), earthquake location and magnitude estimate could not be calculated, in accordance with the detection and location thresholds indicated in Stabile *et al.* (2013) for the same area.