Physics of the Earth and Planetary Interiors

Volume 266, May 2017, Pages 1–17

The original publication is available at:

http://www.sciencedirect.com/science/article/pii/S0031920116302801

Pattern recognition approach to the subsequent event of damaging earthquakes in Italy

S. Gentili[a] and R. Di Giovambattista[b]

[a] Istituto Nazionale di Oceanografia e di Geofisica Sperimentale- Centro Ricerche Sismologiche, Udine, Italy

[b] Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143 Roma, Italy

Abstract

In this study, we investigate the occurrence of large aftershocks following the most significant earthquakes that occurred in Italy after 1980. In accordance with previous studies (Vorobieva et al., 1993, Vorobieva, 1999), we group clusters associated with mainshocks into two categories: "type A" if, given a main shock of magnitude *M*, the subsequent strongest earthquake in the cluster has magnitude $\geq M$-1 or type B otherwise.

In this paper, we apply a pattern recognition approach using statistical features to foresee the class of the analysed clusters. The classification of the two categories is based on some features of the time, space, and magnitude distribution of the aftershocks.

Specifically, we analyse the temporal evolution of the radiated energy at different elapsed times after the mainshock, the spatio-temporal evolution of the aftershocks occurring within a few days, and the probability of a strong earthquake.

An attempt is made to classify the studied region into smaller seismic zones with a prevalence of type A and B clusters. We demonstrate that the two types of clusters have distinct preferred geographic locations inside the Italian territory that likely reflected key properties of the deforming regions, different crustal domains and faulting style. We use decision trees as classifiers of single features to characterize the features depending on the cluster type. The performance of the classification is tested by the Leave-One-Out method. The analysis is performed on different time-spans after the mainshock to simulate the dependence of the accuracy on the information available as data increased over a longer period with increasing time after the mainshock.

1. Introduction

Italy is one of the most seismically active Mediterranean regions documented by historical and instrumental catalogues. More than 120,000 people were killed by earthquakes in Italy during the last century (Valensise and Pantosti, 2001). The worst earthquakes in terms of loss of human life were the 1908 Messina (Mw=7.0, Calabria-Sicily Strait), the 1915 Fucino (Mw=6.7, central Apennines), the 1980 Irpinia (Ms=6.9, southern Apennines), and the 1976 Friuli (Mw=6.5, eastern Alps) earthquakes.

A peculiarity of many large earthquakes in Italy is that they are followed by large aftershocks of magnitude similar to the initial quake or even stronger mainshocks. These multiple mainshock sequences, known in literature as doublets or multiplets, occur in many regions where complex fault systems exist. The largest aftershock can occur within seconds to months on the same fault ruptured by the mainshock or in nearby faults.

Multiple mainshock sequences have been observed for the 1968 M 6.1 Belice (Monaco et al., 1996), 1976 M 6.4 Friuli, 1980 M 6.9 Irpinia (Bernard and Zollo, 1989), 1984 M_w 5.9 Val Comino (Milano and Di Giovambattista, 2011), 1997 M_w 5.7 Umbria-Marche (Miller at al., 2004), 2009 M_w 6.1 l'Aquila (Di Luccio et al., 2010), 2012 M_w 5.9 Emilia (Ventura and Di Giovambattista, 2013) and the 2016 M_w 6.5 Amatrice earthquake.

Repeated earthquakes cause accumulated damage to structures; therefore, forecasting the occurrence of the largest aftershocks based on immediate small aftershocks is a challenging task enabling reduced loss of lives.

In this study, we propose an analysis of the earthquake clusters associated with M≥4.5 earthquakes that occurred in Italy from 1980 to today. Given an earthquake of magnitude M≥4.5, we are interested in identifying statistical features based on the spatiotemporal characteristics of seismicity to forecast a subsequent strong earthquake. We classify the clusters of seismicity associated with

mainshocks in which the strong aftershock has a magnitude ≥M-1 as type A and the others as type B.

This paper is organized as follows: section 2 overviews the seismological and geodynamic settings of Italy and the main characteristics of Italian seismicity. Section 3 describes the earthquake catalogue used in this study and the cluster identification method applied. Section 4 provides detailed information on the different seismological precursors analysed and the statistical evaluation of their performance when applied to Italian seismicity. In section 5, we discuss the obtained results and compare them with the existing literature. Finally, we have included an Appendix with a comprehensive description of the decision-tree-based pattern recognition method implemented for the analysis described in this paper.

2. Seismotectonics and Italian Seismicity

Previous papers have shown a clear dependence on stress and faulting styles of fundamental empirical statistical relations such as the Gutenberg-Richter (e.g., Frohlich and Davis, 1993; Kagan, 2002; Schorlemmer et al., 2005), the Omori-Utsu (e.g., Narteau et al., 2009) and Båth's law (Tahir et al., 2012).

In this paper, we briefly describe the tectonic style of the different regions in which we observed a prevalence of type A or B clusters to describe the findings of our research in terms of the influence of the tectonic structures on the occurrence of the largest aftershock of a sequence (see Figure 1).

The Italian region can be divided into several main areas, each characterized by a dominant seismotectonic behaviour: continent–continent convergence (Alps and Dinarides), plate divergence across margins characterized by passive slab sinking (Northern Apennines and Calabrian Arc), plate divergence across a margin previously characterized by lithosphere sinking and afterwards discharged from the subducted slab (Southern Apennines), backarc basins (Northern Tyrrhenian Sea and Southern Tyrrhenian Sea), and transpression (Northern Sicily). Seismicity recorded since 1980 has given new insight into the seismotectonics of the different areas. The most striking features

derived from seismicity recorded since 1980 are that much of the upper crustal seismicity is concentrated along a narrow bend beneath the Apennines and that deep events located in the Southern Tyrrhenian delineate the still active subduction system. Subcrustal earthquakes occur beneath the northern Apennines and have been interpreted as evidence of subduction of the Adriatic lithosphere beneath the Apennine chain (Selvaggi and Amato, 1992).

In the Alps, earthquakes occur mainly within the upper crust, and seismicity is concentrated to the south-east and west, whereas it is sparse and of low magnitude in the central Alps. The Eastern Alps are dominated by active compression normal to the mountain belt (Kiratzi & Papazachos, 1995; Becker, 2000; Vannucci et al., 2004), whereas in the Western Alps focal mechanism solutions are mostly characterized by extension in a direction perpendicular to the trend of the mountain belt (Frechet, 1978; Eva et al., 1997; Sue et al., 1999; Montone et al., 2004).

In the Eastern Alps, two strong earthquakes occurred during the last 40 years (Friuli M=6.5, 6 May 1976 and Kobarid M=5.6, 12 March 1998). The M 6.5 earthquake was followed on September 15 by a strong aftershock (M=6.0) that was classified as a Related Strong Earthquake by Vorobieva and Panza (1993); the cluster was therefore of type A. The 12 March 1998 earthquake was followed by a smaller (M=4.2) aftershock (cluster of type B).

The Apennines chain resulted from the contemporaneous opening of the Tyrrhenian Sea, the eastward migration of a compressive front, and the retreat of the lithospheric plate dipping below the Italian peninsula (Malinverno and Ryan, 1986; Doglioni, 1991, 1995). The Apennines belt consists of two main first-order segments: the Northern Apennines Arc and the NW-SE striking Southern Apennines. Seismological data and recent geodetic studies have revealed that the Apennines are undergoing a NE-trending extension, with seismic deformation rates higher in the southern Apennines (Anderson and Jackson, 1987; Westaway, 1992; Pondrelli et al., 2002; Hunstad et al., 2003).

The seismic activity of the Apennines concentrates in the axial sector of the chain (http://emidius.mi.ingv.it/CPTI/) from NW-trending alignments in the northern Apennines to NNE-trending alignments in Calabria.

The NW-SE striking normal faults that affect the Apennines are considered responsible for the larger historical earthquakes (e.g., in 1688 and 1805, (*Io*) ≥ 10 MCS), early instrumental earthquakes (1962, *Ms*=6.1; 1980, *Ms*=6.9, Valensise and Pantosti, 2001) and recent normal faulting earthquakes.

NE-SW striking faults dissect the Apennines (Oldow et al., 1993; Sorgi et al., 1998), and low-magnitude earthquake swarms cluster along these faults at the tips of the main NW–SE striking faults (Milano et al., 2002, 2008; Valensise and Pantosti, 2001).

In the Central–Southern Apennines earthquakes often occur as multiple mainshock sequences, e.g., Irpinia (1962 and 1980), Abruzzo-Lazio (1984), Potenza (1990), and Umbria Marche (1997). The 2009 L'Aquila sequence was composed of three mainshocks that ruptured distinct NW–SE striking fault segments (Di Luccio et al., 2010) mapped as disjoint segments in the surface but continuous at depth (Calderoni et al., 2012). Also the ongoing 2016 Amatrice seismic activity, very close to l'Aquila in the north, has shown a multiple mainshocks sequence (Gruppo di Lavoro INGV sul terremoto di Amatrice, 2016).

At the front of the southern Apennines and in the Adriatic foreland, major E-W oriented shear zones have been suggested by several authors between latitudes 40.3N and 42.3N, (Di Bucci and Mazzoli, 2003; Valensise et al., 2004, Kastelic et al., 2013). The interaction of these shear zones with the Apennines chain is still debated, and the presence of such a structural element in the source region of the 2002 Mw=5.7 Molise earthquakes focused more interest on their role. The Molise-Gondola shear zone and the Mattinata fault (Milano et al., 2005) are among the best- investigated right-lateral E-W seismogenic faults and have been interpreted as reactivated inherited discontinuities. From a geodynamic point of view, this deformation belt has been interpreted as a right-lateral

transfer zone accommodating higher roll-back velocities in the northern Adriatic slab with respect to the southern part (Doglioni et al., 1994).

3. Earthquake Catalogue and Cluster Detection

3.1 Catalogue

We have analysed Italian seismicity for a period of over 35 years, from 1980 to March 2016. We used the catalogue of Lolli and Gasperini (2006) obtained by merging complementary catalogues: the CSTI that covers the 1981-1996 period (CSTI Working Group version 1.1), the CSI for the period 1997–2002 (Castello et al., 2006), and the Italian Seismic Bulletin (http://bollettinosismico.rm.ingv.it/) for the period January 2003 – December 2004. From 2005 we used data from the Italian Seismological Instrumental and Parametric Data-Base (http://iside.rm.ingv.it/iside/).

All magnitudes were homogenized to M_L (Lolli and Gasperini, 2006).

The catalogue we used is mainly based on data recorded by the Italian Telemetered Seismic Network (ITSN) integrated with those of local networks. ITSN was started after the destructive 1980 Irpinia earthquake and since then has been enlarged until the present configuration, counting of more than 470 stations. Recording was initially analogue; digital acquisition systems became operative after 1984. Particularly in analogue recordings, the immediate aftershocks of mainshocks can be masked by overlapping arrivals of waves from the mainshock and aftershocks, resulting in a lack of early aftershocks in the earthquake catalogues (Kagan, 2004).

3.2 Cluster identification and classification

The first step in the analysis was the choice of the cluster identification method. Many algorithms have been proposed in literature for this purpose: the most widely used are the Reasenberg (1985) algorithm and window-based methods (e.g., Gardner and Knopoff, 1974). Window-based methods define a time interval T and a circular area of radius R, where T and R are functions of the

mainshock magnitude M_m. Among all the earthquakes only the ones within the time-space distance T and R from the mainshock are considered as belonging to the cluster. Several different functions of M_m have been proposed for T and R in the literature (Gardner and Knopoff, 1974; Uhrhammer, 1986; Knopoff, 2000; Lolli and Gasperini, 2003; Gentili and Bressan, 2008; Molchan and Dmitrieva, 1991). The Reasenberg (1985) method, originally developed for California, defines a seismic sequence as a chain of events linked to each other by spatial and temporal windows. The Reasenberg algorithm allows a data-driven selection of events belonging to the cluster in a non-circular area. The main advantages of window-based methods over the Reasenberg (1985) method are that they are less sensitive to losing earthquakes because of failures of the network and catalogue incompleteness (Lolli and Gasperini, 2003) and that they can be applied more easily to a large amount of data. In Italy, Gentili (2010) showed that the window-based Knopoff et al. (2000) method, extended to smaller magnitudes as in Pace et al. (2006), outperformed Reasenberg's method for Italian seismicity from 1994 to 2004. Lolli and Gasperini (2003) showed that a window-based method outperformed Reasenberg's method on seismicity from 1960 to 1996; they hypothesized that a simple window algorithm is more efficient than Reasenberg's method for low quality or incomplete data, whereas Reasenberg's method is preferable when the quality and the completeness of the catalogue is high. Several window-based methods have been applied to Italian seismicity. Lolli and Gasperini (2003) used an empirical formulation for time T and estimated R using Wells and Coppersmith's (1994) scaling relationship. Pace and Peruzza (2006) and Gentili (2010) used an extension of the Knopoff et al. (2000) method, Gentili and Bressan (2008) proposed equations valid for NE Italy analysed by the OGS catalogue (expressed in duration magnitude), and Di Giovambattista and Tyupkin (2000) used the algorithm proposed by Molchan and Dmitrieva (1991) and coded by V. Smirnov. For further details, see also van Stiphout et al. (2012).

Solving the problem of declustering a catalogue by removing the dependent earthquakes is easier than solving the problem of identifying each dependent event of a given cluster. The choice of the

cluster identification algorithm and the variation of the related parameters can have a great impact on the events associated with each cluster.

In this paper, we decided to use a window-based method after several tests and their evaluation with respect to those obtained using single cluster analysis reported in literature. In fact, the quality of the catalogue in the 1980's does not allow the use of a complex algorithm such as Reasenberg's over the entire Italian territory. To estimate the radius R (in km) of the circular area around the mainshock where the cluster is assumed to be located, we used the Uhrhammer (1986) equation, which supplies a quite smaller radius with respect to other methods, such as the Gardner and Knopoff (1974) method (see e.g., van Stiphout et al., 2012). This choice avoids the inclusion of independent earthquakes in a cluster. The Uhrhammer (1986) equation is given by:

$$R = e^{-1.024+0.804M_m} \quad (1)$$

where M_m is the mainshock magnitude. For T (in days), we adopted the Lolli and Gasperini (2003) empirical equation:

$$T = 60 + 60 \cdot (M_m - 4) \quad (2)$$

This last equation supplies a longer T compared with Uhrhammer (1986) for earthquakes reported in the catalogue in the magnitude range [4.5, 6.5], which has the advantage of including late dependent events in the cluster. This procedure failed only in the case of the Gualdo Tadino earthquake that occurred on 3 April 1998, located in a positive stress area at the northwestern edge of the seismogenic volume activated by the 26 September M_w 6.0 mainshock of the 1997 Umbria-Marche seismic sequence that was not included in the cluster. This caused a fictitious cluster around this earthquake with characteristics very different from the others. For this reason, Gualdo Tadino cluster was not included in the analysis. A further selection was made to eliminate earthquakes too far from Italian borders and in offshore areas that were not well monitored.

We decided to use an operational definition of clusters that defines "mainshock" as the first shock of the cluster with a magnitude greater than 4.5 that is not necessarily the largest shock of the

cluster. We adopted this criterion to be able to apply the procedure immediately after the occurrence of a shock with M≥4.5 without waiting to verify if a stronger earthquake followed. We detected 100 earthquake clusters from 1980 to 1 March 2016 (date of the Iside catalogue download).

To classify aftershocks as type A or B, we need at least one aftershock with $M \geq M_m - 1$ for type A.

For B clusters a completeness magnitude M_c inside the cluster $M_c \leq M_m - 1$ is needed to exclude that the cluster could be of type A with large aftershocks missed. Several methods exist in the literature for estimating the completeness magnitude. For a more detailed description of completeness magnitude evaluation, see Mignan and Woessner (2012). The optimal method depends on the individual data set. When a large database with M_c heterogeneities is analysed, the well-known Maximum Curvature Technique (Wiemer and Wyss 2000) can underestimate the completeness magnitude and the EMR method (Entire Magnitude Range method - Woessner and Wiemer, 2005) produces more conservative results. However, the EMR method is especially suitable when a large database is available, whereas Mignan (2011) showed that the Maximum Curvature Technique does not underestimate M_c when considering a local data set with small heterogeneities in M_c. We preferred to apply a rougher estimation of completeness magnitude based on the Maximum Curvature Technique, which is more suitable when few data items are available. We required a minimum of 10 earthquakes (in addition to the mainshock) in the cluster in order to take into account the cluster for completeness magnitude estimation. The number of earthquakes necessary for a stable estimate by these methods is higher than 10; Torman et al. (2014) and Schorlemmer et al. (2003) e.g. used 50 earthquakes; therefore, for cluster with a number of earthquakes between 10 and 50, we validated the completeness magnitude by using a longer time span respect to the cluster duration (three years when available) in order to obtain more stable results. This evaluation does not take into account the increase of completeness magnitude observed immediately after the mainshock (see e.g. Helmstetter et al. 2006), due to waveforms overlapping. However, this is particularly complex in Italy, where in the first hours after a mainshock, temporary stations are added, decreasing the completeness magnitude. Methods for completeness estimation during the

clusters are based on the hypothesis that the seismic network is unchanged during the cluster (see e.g. Hainzl, 2016). We validated the obtained completeness magnitude with an automatic estimate of minimum detectable magnitude based on analysis of the S/N ratio implemented at INGV (Marzorati and Cattaneo, 2016) using the software PQLX (McNamara and Buland, 2004; McNamara and Boaz, 2011).

Our analysis supplies generally higher completeness magnitude respect to these methods, resulting in a more conservative approach. Table 1 shows a selection of detected clusters for which a classification of A or B class was possible (47 clusters), the number of events and the completeness magnitude of each cluster. The error on completeness magnitude has been obtained by 100 bootstrap.

4. Method of Analysis

The method of analysis is based on several features described in detail in section 4.1. The performance of the features was evaluated using a pattern recognition approach described in section 4.2 to develop a classifier described in section 4.3.

4.1 Premonitory seismicity patterns

We tested several precursory seismicity patterns to develop an automatic system to forecast in advance if a cluster is of type A or B. In the past, Vorobieva and Panza (1993) proposed a set of functions representing premonitory phenomena for Italy for which it was necessary to have an observation time (T) of 10 days after the analysed strong earthquake (an earlier version of the method required T=40 days for many functions). However, we verified that in our database of Italian seismicity 71.4% of clusters of type A had the stronger aftershock before the 10-day restriction and their method thus could not be applied. The main idea for the subsequent strong earthquake forecasting of all the functions proposed by Vorobieva and Panza (1993) is the same of M8 and CN algorithms for the preparation phase of the first strong earthquake (see e.g. Peresan et al. 2005). The strong earthquake is considered a critical point (singularity) and, in agreement with

the theory of non-linear dynamical systems, significant variations of observables are expected, like high activity of the system and irregular behaviour, that are symptomatic of the critical state of the system before the strong earthquake. These variations can be seen from the physical point of view as the increased response of the lithosphere to tectonic stress.

We checked Vorobieva and Panza (1993) functions for a shorter and variable T (from 6 hours to 7 days). Additionally, we proposed a set of other functions calculated in the same period and aimed to capture high and irregular earthquakes activity. According to Vorobieva and Panza (1993) only earthquakes with magnitude greater or equal to at least the mainshock magnitude (M_m) minus a given feature-dependent threshold are to be considered. Each function is calculated at time intervals $[s_1,s_2]$, with s_1=1 hour after the mainshock and s_2 increasing to simulate the increasing information with time after a strong earthquake. The precursors analysed by Vorobieva and Panza (1993) for the earthquakes that followed the strong event were the following:

1. *N*: number of events with $M \geq M_m$-3. In this paper, we also analysed the number of events greater than M_m-2 (we call it N_2) to avoid problems connected with completeness on the first hours after the mainshock.
2. *S* (total equivalent source area):

$$S(i) = \sum_i 10^{(m_i - M_m)} \tag{3}$$

   where m_i is the magnitude of the i^{th} event in the selected time interval. S corresponds to the area occupied by the aftershocks normalized to the area occupied by the mainshock. Only aftershocks with magnitude M_m-2 are considered in this function.

3. V_m (cumulative variation of magnitude from event to event):

$$V_m(i) = \sum_i | m_i - m_{i-1} | \tag{4}$$

   All the events with $M \geq M_m$-3 are considered.

4. V_{med} (cumulative variation of average magnitude from day to day):

$$V_{med}(i) = \sum_i |\mu_i - \mu_{i-1}| \quad (5)$$

where μ_i is the average magnitude of the mainshocks on the i^{th} day ($M \geq M_m$-3).

5. V_n (cumulative variation of the number of aftershocks from day to day):

$$V_n(i) = \sum_i |n_i - n_{i-1}| \quad (6)$$

where n_i is the number of aftershocks that occurred on the i^{th} day ($M \geq M_m$-3).

Additionally, Vorobieva and Panza (1993) used the following:

6. N_{for} (number of foreshocks with magnitude greater than M_m-1) in the time interval [s_1, s_2] with s_1=5 years and s_2=3 months before the mainshock origin time and in a radius that is 1.5 times larger than the one considered for the aftershocks. The radius is estimated by Eq. (1).

We did not analyse the function R_{max} proposed by Vorobieva and Panza (1993), the largest distance between the mainshock and the aftershock with magnitude $M \geq M_m$-3, because for M_m=4.5, this feature is very sensitive to noise from spurious background events.

In addition to the analysis of the Vorobieva and Panza (1993) functions, we developed a set of original functions corresponding to seismic precursors; some of them were modified from others existing in the literature, whereas some original ones are proposed for the first time in this paper:

7. Z (linear concentration of aftershocks) is similar to the feature Z of the M8 method for $M \geq 8$ earthquake forecasting (e.g., Kossobokov et al., 1999); the difference is that in this case events with magnitude $\geq M_m$-3 following a strong earthquake are considered instead of mainshocks such as in M8. The function Z is the ratio between the average diameter of the source of the aftershocks divided by the mean distance between aftershocks:

$$Z(i) = \frac{mean(10^{0.69 m_i - 3.22})}{mean(r_{ij})} \quad (7)$$

where r_{ij} is the distance between the generic i^{th} and j^{th} aftershock. We set the starting time s_1=1 hour for the mean value calculation.

8. *SLCum* is the cumulative deviation of S from the long-term trend. The interval $[s_1, s_2]$ is subdivided into smaller intervals $[s_1, s_1 + dt], [s_1, s_1 + 2dt], \ldots [s_1, s_1 + ndt]$ where $s_1 + ndt \le s_2$ and $s_1 + (n+1)dt > s_2$. We chose dt=6 hour and s_1=1 hour. The function *SLCum* is defined as:

$$SLCum(i) = \sum_i abs[S(t_i) - S(t_{i-1}) \frac{i \cdot dt}{(i-1) \cdot dt}] \quad (8)$$

where $t_i = s_1 + i \cdot dt$ and $S(t_i)$ is S calculated on the time interval $[s_1, t_i]$. This feature is sensitive to abrupt variations of S.

9. *SLCum2* is the cumulative deviation of *S* from a sliding window from the long-term trend. In this case, the interval $[s_1, s_2]$ is subdivided into a set of smaller intervals $[s_1, s_1 + dt], [s_1 + dt, s_1 + 2dt], \ldots \quad [s_1 + (n-1)dt, s_1 + ndt]$, and another set $[s_1, s_1 + d\tau], [s_1 + dt, s_1 + dt + d\tau], \ldots \quad [s_1 + (n-1)dt, s_1 + (n-1)dt + d\tau] \quad s_1 + ndt \le s_2$ and $s_1 + (n+1)dt > s_2$. We chose dt=6 hours and $d\tau$=1 hour. *SLCum2* is defined as:

$$SLCum2(i) = \sum_i abs[S([s_1 + (i-1) \cdot dt, s_1 + i \cdot dt]) - S([s_1 + (i-1) \cdot dt, s_1 + (i-1) \cdot dt + d\tau]) \frac{dt}{d\tau}]$$

(9)

where $S[a, b]$ is S calculated over the generic time interval $[a, b]$. This feature is sensitive to abrupt variations of *S*, but differently from *SLCum*, the window does not start at a fixed time close to the mainshock origin time.

10. *Q* (Normalized Radiated Energy). This feature describes the total energy released by the aftershocks normalized to that of the mainshock. Some previous works have studied this feature (Di Giovambattista and Tyupkin, 2002) or its inverse (R_{ES} - Gentili and Bressan 2008).

$$Q(i) = \frac{\sum_i E_i}{E_m} \quad (10)$$

where E_m is the energy of the mainshock and E_i is the energy of the i^{th} aftershock. The energy E (in Joules) of an event of magnitude M is obtained by the equation $Log_{10}(E) = \frac{3}{2}M + 4.8$ (Gutenberg and Richter, 1956).

11. *QLCum* is the cumulative deviation of *Q* from a long-term trend. It is calculated similarly to *SLCum* for the *S* function: the interval $[s_1, s_2]$ is subdivided into smaller intervals $[s_1, s_1 + dt], [s_1, s_1 + 2dt], \ldots [s_1, s_1 + ndt]$, $s_1 + ndt \le s_2$ and $s_1 + (n+1)dt > s_2$. We chose *dt*=6 hours and s_1=1 hour. *QLCum* is defined as:

$$QLCum(i) = \sum_i abs[Q(t_i) - Q(t_{i-1}) \frac{i \cdot dt}{(i-1) \cdot dt}] \quad (11)$$

where $t_i = s_1 + i \cdot dt$ and *Q* is calculated on the time interval $[s_1, t_i]$.

12. *QLCum2* is the cumulative deviation of *Q* from the long-term trend with a sliding window. Analogously to *QLcum* it is calculated similarly to *QLcum2* for the *Q* function: in this case, the interval $[s_1, s_2]$ is subdivided into a set of smaller intervals $[s_1, s_1 + dt], [s_1 + dt, s_1 + 2dt], \ldots [s_1 + (n-1)dt, s_1 + ndt]$, and another set $[s_1, s_1 + d\tau], [s_1 + dt, s_1 + dt + d\tau], \ldots [s_1 + (n-1)dt, s_1 + (n-1)dt + d\tau]$ $s_1 + ndt \le s_2$ and $s_1 + (n+1)dt > s_2$. We chose *dt*=6 hours and *dτ*=1 hour. *QLCum2* is defined as:

$$QLCum2(i) = \sum_i abs[Q([s_1 + (i-1) \cdot dt, s_1 + i \cdot dt]) - Q([s_1 + (i-1) \cdot dt, s_1 + (i-1) \cdot dt + d\tau]) \frac{dt}{d\tau}]$$

(12)

where *Q*[*a, b*] is *Q* calculated on the generic interval [*a, b*]. Analogously to *SLCum2*, this feature is sensitive to abrupt variations of *Q* differently from *QLCum*, the window does not start at a fixed time close to the mainshock origin time.

Additionally, we added four features that depend only on the following mainshock source characteristics:

13. Mainshock location
14. Mainshock magnitude
15. Mainshock depth
16. Mainshock focal mechanism

4.2 Precursor classification and performance evaluation

We applied the pattern recognition approach to develop and test possible classifiers based on precursors able to predict the classification of a cluster. The precursor functions described in section 4.1 supply a numerical response (features) for each cluster. These responses are organized into a vector $\vec{p} = (p_1, \ldots p_n)$, called "pattern", that corresponds to the measured features of the cluster. We used different classifiers for all the features; therefore, the pattern dimension is 1, except for the feature based on cluster positions in the Italian territory, for which the pattern dimension is 2, corresponding to latitude and longitude. Our classes are two: "clusters of type A" and "clusters of type B". In binary classifiers, one class is defined as positive (we chose the A class) or negative (B class).

We decided to use decision trees (e.g., Breiman et al., 1984; Jang et al., 1997) that allow a clear understanding of the characteristics of the precursors depending on the class of the cluster (for details see Appendix A).

To evaluate the performance of the classifier, we adopted the Leave-One-Out (or LOO) method (see appendix A). In terms of accuracy, LOO often results in high variance as an estimator for the test error and a high computational cost. However, in our case we had to address a very limited dataset and it was thus necessary to enlarge the training set as much as possible to capture the variability of

the patterns inside the same class. Because of the small dataset, the large computational cost of the LOO method was not a problem for us to consider.

In estimating the performance of each instance of a binary classifier, four possible outcomes are possible:

- True Positive (Hit): the instance is positive and is classified as positive;
- False Negative (Miss - type II error): the instance is positive and is classified as negative;
- True Negative: the instance is negative and is classified as negative;
- False Positive (False alarm - type I error): the instance is negative and is classified as positive.

Classifier performance was evaluated in terms of the total number of True Positive (TP), False Negative (FN), True Negative (TN) and False Positive (FP) in the test sets defined by the LOO method by using the three functions called *Precision*, *Recall* and *Accuracy* (e.g., Powers, 2011; for definitions see appendix A). These functions are defined in the interval [0, 1]; the best performance is obtained when a feature is 1, the worst when it is 0.

The comparison between different features is shown by the Receiving Operating Characteristics (ROC) graph (Egan, 1975; Swets et al., 2000, Fawcett, 2006 – see Appendix A). Note that ROC graphs are usually not used in earthquake forecasting method performance assessment because they do not account for the fact that earthquakes are clustered in space; Molchan diagrams (Molchan 1991; Molchan and Kagan 1992) are preferred. However, our analysis dealt with distinct clusters and therefore ROC graphs could be applied (for details see Appendix A).

For every selected time interval, a different value of *Precision*, *Recall*, *Accuracy* and a point in the ROC graph were estimated for every precursor.

4.3 Forecasting strong subsequent earthquakes

To develop a classifier able to forecast strong subsequent earthquakes in a cluster, it is necessary to merge together all the information supplied by different features. Many pattern recognition methods

for earthquake forecasting (Vorobieva 1999; Vorobieva and Panza 1993; Kossobokov et al., 1999) simply count the number of precursors supplying positive results, and if the number is over a given threshold, the issue is classified as positive. However, Grandori et al. (1984) showed that a combination of precursors does not always supply more useful forecasting than a simple precursor; lower performance of one precursor can decrease the quality of results, and dependent precursors do not improve the quality of the classification. In this study, we applied an original approach that weights the precursors in a different way depending on their performance to handle this problem. To do this, we used the Informedness concept (Powers, 2011) defined as:

$$Informedness = True\ Positive\ Rate - False\ Negative\ Rate \tag{13}$$

where the True Positive Rate is the ratio between TP and the total number (T) of positive instances, whereas the False Negative Rate is the ratio between FN and the total number (N) of negative instances. Informedness derived from a ROC graph ranges between -1 and 1 and quantifies how informed a predictor is for the specified condition (Powers, 2011). An Informedness of 1 corresponds to the point (0, 1) in the ROC graph (ideal predictor), whereas Informedness <0 corresponds to the lower right triangle (worse than random guessing). For each time interval, we evaluated the precursors performance, estimated the Informedness of each precursor and set to 0 the ones with Informedness <0. We defined the precursors weights as:

$$W_i = \frac{Informedness_i}{\sum_j Informedness_j} \tag{14}$$

where $Informedness_i$ is the Informedness of the i^{th} precursor and W_i is its resulting weight.

We defined the final score Sc of an instance (cluster) as:

$$Sc = \sum_i W_i C_i \tag{15}$$

where C_i is the class assigned to the i^{th} precursor function, 1 if the classifier classifies it as A and -1 if it classifies it as B. *Sc* ranges between -1 and 1; if *Sc*>0 the cluster is classified as A, otherwise it is classified as B.

This approach allowed to also have an estimate of the probability of an A cluster, not only a binary (A or B) answer.

We defined the probability of an A cluster as:

$$\Pr(A) = \frac{Sc + 1}{2} \tag{16}$$

In this way, a higher number of relevant (weight>0) precursors supplying A as the resulting class supplied a higher probability that the cluster was of class A.

5 Results

We characterized the occurrence of the largest subsequent stronger earthquake considering both the information of the mainshock and the time of occurrence and distance of the earthquakes that followed the mainshock.

5.1 Mainshock based features

We analysed the cluster type depending on the location, magnitude, rake and depth of the mainshock. Figure 2 shows the ROC graph obtained by the LOO method for all the features, together with the number of foreshocks. These features had the advantage of being available in a very short time after the mainshock (INGV, e.g., supplies automatic preliminary mainshock, location and magnitude within 100 seconds and reviews a moment tensor solution within a few minutes from the mainshock). However, their performance was generally quite poor.

The only exception was the cluster location, which supplied high True Positive Rate and small False Positive Rate. Figure 3 shows that the Italian territory had a quite homogeneous behaviour in terms of cluster type in the different regions: the North-Eastern Alps and South-Eastern Italy were characterized by type B clusters, and the Central Apennines, Sicily and the Adriatic Sea were prevalently of type A, whereas the Po Valley had an intermediate behaviour. Training on the entire catalogue supplied a decision tree that can be summarized into the following rules:

- If latitude is ≥ 44.1 the class is B
- If latitude is <44.1:
    - If longitude is ≥15.3 the class is B
    - If longitude is <15.3 the class is A

(See the continuous lines in Figure 3.)

We hypothesized another subdivision:

- If latitude is <44.1 and ≥42 and longitude ≥15.3 the class is A.

(See the dashed line in Figure 3.) This choice would have the advantage of dividing the Adriatic Sea from the southern area. However, because of the low statistical relevance of this subdivision (only three clusters were involved), neither the final classifier nor the estimated performance considered this hypothesis.

The poorer performances in Figure 2 were supplied by the feature on the focal mechanism, which considered the rake angle. However, Figure 4 shows that the rake angle appears partially related to the cluster type: most strike slip and inverse faults were of type B. The low performance of the feature was because of normal faults, whose classes were mixed depending on the region where the mainshock occurred. The Central Apennines were mainly characterized by A clusters, whereas southern ones were mainly characterized by B clusters (see Figure 3). For this reason, the previous position-based features had better results and rake based features were no longer be considered in this study.

Other features we considered were depth and mainshock magnitude. The depth distribution of mainshocks may be related to the differences in magnitude between mainshock and subsequent strong earthquakes; however, the accuracy of depth estimation in the early years of the catalogues was too low and this feature should be disregarded.

There was a general hint that the strongest magnitude mainshocks (M_m≥5.8) may correspond to A class clusters. Of five M_m≥5.8 earthquakes, 4 were of type A. The only one belonging to class B was the Irpinia one, in which strong aftershocks may have been hidden in the mainshock because of

the waveform superposition and the analogue recording in the 1980's. Additionally, the Friuli earthquake cluster that occurred in 1976 (M_m=6.4), not included in this database, was of type A and the very recent Amatrice cluster (first strong shock on 24 August 2016 M_m=6), also not included, was of type A. Because of the extremely low number of observations over the threshold, these results cannot be confirmed from a statistical point of view, and a larger dataset should be used. The performance of this feature is affected by the fact that for M<5.8 the two classes are mixed and the developed decision trees may be complex. If the hypothesis that for Mm≥5.8 clusters are all type A will be confirmed in future work, this feature could be used as a strong condition for B class option elimination (if M_m>5.8 the cluster could be automatically classified as A).

We found analogous results that could not be confirmed from a statistical point of view for the feature N_{for}. All the clusters with N_{for}≥2 were of type A. For this feature the False Positive Rate was equal to 0; however, the number of clusters for which N_{for}≥2 was very low (4 clusters). If this result will be confirmed on a larger dataset, this would be a strong condition for B class option elimination (if N_{for}≥2 the cluster could be automatically classified as A).

5.2 Features based on subsequent earthquakes

Other features are based on the ongoing seismic activity after the mainshock. The analysis of the 24 selected type A clusters showed that 67% of the stronger subsequent earthquakes occurred before 10 days. To compare the performance of precursors we selected 10 different time periods (in days: [0, 0.25] [0,0.5] [0,0.75] [0,1], [0, 2] … [0, 7]) and we calculated the values of all the tested precursors for each period.

Because of limitations connected with completeness magnitude and the fact that we could "forecast" a cluster as type A only if the strongest aftershock occurred after the end of the selected period, the number of available clusters was low, ranging from 24 (10 A clusters and 14 B clusters) for the time interval [0,0.25] to 19 (5 A clusters and 14 B clusters) for the time interval [0,7].

Figure 5 shows an example of the performance of a feature (N_2) at different time intervals. The black lines show the values of Precision Recall and Accuracy if N_2 were calculated for the entire period examined. Accuracy should never be lower that the line corresponding to the B clusters percentage; otherwise the classifier would perform worse than one that always classified the clusters as B. All the parameters should be greater than 0.5 (50% correctly classified, i.e., a random classifier for a balanced dataset). For several features, the performance changed with time and for the shortest time of observation they were poorer (e.g., in Figure 5 performance after 6 hours was lower than after 12). This trend was related to the lower number of earthquakes released in a shorter time interval leading to poorer statistics. Another possible reason was the higher percentage of type A clusters for shorter time intervals with respect to longer ones mainly populated by Type B clusters. Imbalanced datasets are known to affect classifier performance. However, performance was lower when the time lapse was increased (in Figure 5 for T>12 hours). This trend could have been caused by several factors, such as the influence of the background seismicity and activation of nearby fault segments, particularly in the case of larger earthquakes, or, more simply, it may have been because of the smaller available database and therefore the lower ability of the classification system to discriminate between classes.

Red lines correspond to the performance if only the information on the first 12 hours was used for all time periods longer or equal to 12 hours. Performance did not decrease as in the case in which all the available information was used. Fluctuations in the values were no longer connected with the feature value but only to the changes in the dataset.

The time at which we obtained best performance changed from one feature to another; for most features the best performance was obtained for time intervals less than or equal to 2 days, whereas for V_m and V_{med} longer time intervals were necessary. For each feature we assumed a maximum s_2 corresponding to the best performance of the feature (hereafter S_2).

Figure 6 shows the ROC graph obtained by the LOO method for different time periods; for shorter time periods, not all of the features could be calculated: *SLcum, SLcum2, QLcum, QLcum2* needed

at least 12 hours, and V_n and V_{med} needed 3 days. Given a feature, its performance is shown only for $s_2 \leq S_2$. From the analysis of Figure 6 we can conclude that the performances of features *Q*, *S*, *SLcum* and *SLcum2* were among the best, depending on the considered period. By considering only functions with Informedness>0 and using all the available dataset we could develop one-node decision trees for each feature based on subsequent earthquakes. In other words, a simple threshold was enough to discriminate between class A and class B.

Table 2 shows the value of S_2 for each feature and the values of the threshold for each feature for $s_2 \leq S_2$ calculated on the entire available dataset. For the V_{med} feature, the class was A if V_{med} was smaller than the corresponding threshold and B if V_{med} was greater or equal. For all the other feature vice-versa. For V_{med} only the threshold for reliable values of the performance (better than random) are shown. When longer time intervals were analysed, the feature was calculated using only the information on the time interval [s_1, S_2]. Note that for s_2=6 hours *SLcum*, *SLcum2*, *QLcum*, and *QLcum2* could not be calculated because of the choice of s_1=1 h and dt=6 hours; for s_2=12 hours *SLcum*≡*SLcum2* and *QLcum*≡*QLcum2.*

5.3 Scoring of the features for future forecasting

In previous papers, a voting approach was applied in which all the features contributed with equal weight and the positive class was assigned if the number of votes was over a given threshold. However, this approach often causes an overweighting of less-relevant features with respect to the most relevant ones.

We decided to develop a classifier that considered the different performances. To this end, we considered the Informedness of each feature (see Figure 7), and we derived the scores of each feature as described in section 4.3. We used the entire available dataset to develop a classifier for future clusters. Using the classification of each decision tree and the corresponding weight it was possible to determine the probability that a given cluster would be of type A (see section 4.3). To

develop a robust classifier, we eliminated the features with Informedness < 0.2 and used the Informedness calculated in the time interval $[s_1,S_2]$ for all the intervals $[s_1,T]$ with $T>S_2$. The feature *QLcum2* was not used in the final classifier because it supplied the best results at a time S_2=12 hours when it coincided with *QLCum*. *SLcum2* was used for T>12 hours. The only feature we used based on mainshock information was the epicentral position (Informedness=0.49) because the others had Informedness<0.2. Table 3 shows the weights we adopted for each feature.

In order to verify that the previous result do not depend on the choice of declustering windows of eq. (1) and (2) we applied, we repeated the whole procedure using Gardner and Knopoff (1974) equations for R (in km) and T (in days):

$$R = 10^{0.1238M_m+0.983} \tag{17}$$

$$T = \begin{cases} 10^{0.032Mm+2.7389} \ if \ Mm \geq 6.5 \\ 10^{0.5409Mm-0.547} \ if \ Mm < 6.5 \end{cases} \tag{18}$$

Gardner and Knopoff (1974) equations supply larger R (for $M_m \leq 6.5$) and T than the ones in eq. (1) and (2), causing the arbitrary introduction of background seismicity. Table 1 and 2 in Electronic Supplements show the comparison between the values of Informedness with both declustering methods. The performances are very similar for all the features with little better performances using the declustering we adopted, confirming the stability of the method. The only feature for which the Informedness never reaches positive values is V_{med}; this feature is therefore very sensitive to declustering method, and should be accurately checked if the method is applied to a different region/catalogue. For all the other features the thresholds are the same of Table 2, confirming the stability of the applied method.

Figure 8 shows a test used on the entire dataset to verify how the data corresponded to the developed model. The assigned probability to have an A cluster was evaluated for all the B (blue curves) and A (red curves) clusters of our dataset. Each cluster is represented by a different curve. It is important to note that (unlike the results presented in Figg. 5, 6 and 7) this is not a reliable estimate of the method performance because the training test and the test set coincided. However,

this application may be useful to outline clusters with an anomalous behaviour with respect to the developed model. Moreover, it provides an estimate of the upper limit of the performance possible with an independent test set. The worst performance was obtained immediately after the mainshock occurrence when only the location of the mainshock could be used (10 out of the 47 clusters were misclassified – see Table 3 in Electronic Supplement). Using an interval of 6 hours, only two A class clusters were misclassified, with a probability of A being 0.43 and 0.45. The total number of clusters available was 24. For intervals greater or equal to 18 hours, we had a wrong classification in only one case: the Reggio Emilia earthquake of 15 October 1996 (Di Giovambattista and Tyupkin 1999), when a B cluster was wrongly classified as A; the probability of A ranged from 0.53 to 0.62, depending on the interval used. The high percentage of right classifications, with the probability values far from 0.5, together with the fact that wrong classifications were close to 0.5, suggested the adoption of a fuzzy approach in which intermediate values of probability, i.e., Prob(A) between 0.4 and 0.6 (see black lines in Figure 8), do not contribute to the classification. Also in this case, we checked the performances of the method using Gardner and Knopoff (1974) declustering. Tables 3 and 4 in the Electronic Supplements show the values of Prob(A) for both methods, in all the considered time intervals for each cluster. Due to the merging of different clusters and to changes in completeness magnitude and due to the larger area and time span considered, there are few differences in the clusters with smallest magnitude mainshock identification; in particular, 2 out of 47 clusters are not detected by the different declustering method and 4 new ones are identified. All these differences concern the analysis of the mainshocks characteristics only, because no one of these 6 clusters has enough productivity/completeness magnitude to be checked for longer time span. For all the other clusters the performances remain very similar, showing once again the stability of the method for different choices of declustering.

6 Summary and Discussion

In this paper, we developed a new method based on pattern recognition for a fast estimate of the probability that, given a strong earthquake, another strong one will follow. We analysed the characteristics of the mainshock in terms of position, magnitude and focal mechanism, together with a spatio-temporal analysis of both foreshocks and seismicity in the first hours after the mainshock. The study was performed by analysing each feature separately, classifying them by a decision tree and evaluating the performances by a ROC graph based on the LOO method. The performance estimates allowed us to develop a new multi-feature classifier in which each feature is weighted depending on its performance. The performance of this new classifier should be evaluated on future earthquakes.

Regarding the features based on mainshock characteristics, good results were obtained by considering the cluster position; in particular, the Central Apennines were characterized by clusters of type A. This may be explained by the presence of fluids in the area. Swarms that are type A clusters have often been assumed to result from fluid intrusion in seismogenetic areas (Hainzl, 2004). In a study on this area, Miller et al. (2004), proposed that aftershocks of the 1997 Col Fiorito cluster were driven by the coseismic release of fluids by ruptures created by the larger events. Zhuang (2015a) and Zhuang et al. (2017) proposed a zonation of Italy into different regions with different seismic characteristics in terms of ETAS parameters by using a weighted likelihood estimator to obtain stable estimates of spatial changes of the parameters (Zhuang, 2015b). The zonation is very similar to our subdivision into A and B regions.

With respect to the focal mechanism, most clusters with strike slip and inverse mechanisms in our database were associated with type B clusters, whereas, for normal faults, the cluster class also depended on the area where it was located (see Figure 4). The dependence of the magnitude of the aftershocks on the rake angle was previously analysed by Zakharova et al. (2013), who found a higher seismic moment ratio (the difference of cumulative moment of aftershocks minus cumulative

moment of foreshocks, all normalized to the mainshock magnitude) for normal faulting earthquakes compared with strike slip earthquakes. These researchers' findings are therefore compatible with our results. However, the focal mechanism has also been related to the *b*-value (Gulia and Wiemer, 2010, Schorlemmer et al., 2005); the highest *b*-values have been found for normal faults, intermediate values for strike-slip faults, and the lowest ones for inverse faults. Because the magnitude of the strongest aftershock is inversely proportional to the *b* value (Shcherbakov and Turcotte, 2004), normal faults should have higher Δm with respect to inverse faults, in contradiction to our findings, in which most clusters associated with inverse faults were of type B, whereas many normal fault clusters were of type A. This contradiction may be explained by the differences in productivity *a* of the clusters (the strongest aftershock magnitude $m^*=a/b$, Shcherbakov and Turcotte, 2004) or to changes in the values of *b* in time throughout the cluster (see e.g., Passarelli et al., 2015). Tahir and Grasso (2015) found a lower aftershock rate for strike slip faults compared with normal faults. This is compatible with our result that most clusters with the mainshock characterized by the strike slip focal mechanism were of type B. Because many different factors affect the value of Δm for the same focal mechanism, we prefer to base our algorithm on the cluster position feature, which supplies better results (see section 5.1).

A relation between the cluster class and the mainshock magnitude M_m (the class is A if $M_m \geq 5.8$) or the number of foreshocks with magnitude $\geq M_m$-1 (the class is A if $N_{for} \geq 2$) was evidenced, although the statistical basis was too low. It may be hypothesized that larger earthquakes activate more complex tectonic structures, such as contiguous faults, or apparently disjoint segments contiguous in depth, increasing in this way the probability of a subsequent strong event. On the other hand, Båth law (Båth, 1965) states that Δm is approximately constant and does not depend on M_m. Several studies have outlined the large variance of Δm (e.g., Shcherbakov and Turcotte, 2004); however, the dependence on M_m is still debated. Tahir et al. (2012), analysing the USGS global earthquake catalogue, found no dependence of Δm on the magnitude of the mainshock for earthquakes in the magnitude range [7, 8.3] (larger than ours).

The analysis of early subsequent earthquakes was performed using functions existing in literature or developed for this paper. Our analysis indicates that the features based on the magnitude of the events (such as the trend of radiated energy and total equivalent source area in time) together with the features based on their number (cluster productivity) supply the best results and provide a robust tool to hypothesize if an earthquake will be followed by a large aftershock. The dependence of Δm on the number of recorded subsequent earthquakes (lower Δm if N is larger) was also found by Chan and Wu (2013) by analysing 706 clusters in Taiwan, even if their results could have been biased by the spatial variation of the completeness magnitude. Gentili and Bressan (2008) related the inverse of Q to Δm based on a small dataset of seismic sequences from 1977 to 2007 in NE Italy.

Our results are based on a limited number of available clusters. Nonetheless, the test by the LOO method and the good capability of the method to discriminate the two classes suggest that our results are not due to limited statistics. The results of this paper may be used to develop improved region-specific hazard estimates for large subsequent earthquakes.

A first partial test of the method resulted from a very recent earthquake, not included in the database, which struck Italy during the writing of this paper: the 24 August 2016 M_L=6 Amatrice earthquake (preliminary data are available). One hour after the mainshock, a strong earthquake (M=5.3) followed the mainshock. The cluster was therefore of type A. Unfortunately, because of the short time between the mainshock and the subsequent strong earthquake none of the features based on subsequent earthquakes could be applied. However, the feature based on the mainshock could supply a first estimate: the cluster happened in the Central Apennines, a type A region based on our results, and the mainshock magnitude was greater than 5.8, compatible with the hypothesis of a type A cluster. However, the number of foreshocks did not supply any information because no foreshock with magnitude $\geq M_m$-1 was recorded before the mainshock, and the mechanism is normal, i.e., a mechanism for which the cluster class is region-dependent.

Acknowledgements

We thank Prof. Rodolfo Console and the unknown reviewers for the useful comments and suggestions.

References

Anderson, H.A., Jackson, J.A., 1987. Active tectonics of the Adriatic region. Geophys J R Astron Soc 91, 937-983.

Båth, M., 1965. Lateral inhomogeneities in the upper mantle, Tectonophysics 2, 483–514.

Becker, A., 2000. The Jura Mountains—An active foreland fold-and-thrust belt?. Tectonophysics, 321, 381– 406, doi:10.1016/S0040-1951(00)00089-5.

Bernard, P., Zollo, A., 1989. The Irpinia (Italy) 1980 earthquake: Detailed analysis of a complex normal fault, J. Geophys. Res., 94, 1631–1648.

Breiman, L., Friedman, J., Olshen, R. A., Stone, C. J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, California.

Castello, B., Selvaggi, G., Chiarabba, C., Amato, A., 2006. Catalogo della sismicità italiana, -CSI 1.0 (1981–2002) (available at http://csi.rm.ingv.it/).

Calderoni, G. , Di Giovambattista, R., Vannoli, P., Pucillo, S., . Rovelli, A. Fault-trapped waves depict continuity of the fault system responsible for the 6 April 2009 Mw 6.3 L'Aquila earthquake, central Italy. Earth and Planetary Science Letters, 323–324 (2012), pp. 1–8 http://dx.doi.org/10.1016/j.epsl.2012.01.003

Chan, C.-H., Wu, Y.-M., 2013. Maximum magnitudes in aftershock sequences in Taiwan Maximum magnitudes in aftershock sequences in Taiwan Journal of Asian Earth Sciences, 73, 409–418.

CSTI Working Group, 2001. Catalogo strumentale dei terremoti Italiani dal 1981 al 1996, Version 1.0. CD-ROM, Clueb, Bologna, also available at:
http://ibogfs.df.unibo.it/user2/paolo/www/ gndt/Versione1_0/Leggimi.htm.

Di Bucci, D., Mazzoli, S., 2003. The October-November 2002 Molise seismic sequence (southern Italy): an expression of Adria intraplate deformation. J. Geol. Soc. London, 160, 503-506.

Di Giovambattista, R., Tyupkin Y. S., 1999. The fine structure of the dynamics of seismicity before M 4.5 earthquakes in the area of Reggio Emilia (Northern Italy). Ann. Geofis., 42, 897– 909.

Di Giovambattista, R., Tyupkin Y.S., 2000. Spatial and temporal distribution of the seismicity before the Umbria-Marche September 26, 1997 earthquakes Journal of Seismology, 4 (2000), pp. 589–598.

Di Giovambattista, R., Tyupkin, Yu.S., 2002. Burst of aftershocks as a manifestation of instability of the earth crust in an area of strong earthquake preparation // European seismological commission (ESC). XXVIII General Assembly. 1-6 September, 2002 Book of Abstracts. P.228.

Di Luccio, F., Ventura, G., Di Giovambattista, R., Piscini, A., Cinti, F.R., 2010. Normal faults and thrusts activated by deep fl uids: The 6 April 2009 MW 6.3 L'Aquila earthquake, Central Italy. J. Geophys. Res. 115 (B06315). doi:10.1029/2009JB007190.

DISS Working Group (2015. Database of Individual Seismogenic Sources (DISS), Version 3.2.0: 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-DISS3.2.0

Doglioni, C., 1991. A proposal of kinematic modelling for W-dipping subductions - possible applications to the Tyrrhenian-Apennines system. Terra Nova 3, 423-434.

Doglioni, C., Mongelli, F., Pieri P., 1994. The Puglia uplift (SE Italy): an anomaly in the foreland of the Apenninic subduction due to buckling of a thick continental lithosphere. Tectonics 13, 1309-1321.

Doglioni, C., 1995. Geological remarks on the relationships between extension and convergent geodynamic settings. Tectonophysics, 252, 253e268.

Egan, J. P., 1975. Signal detection theory and ROC analysis, Series in Cognition and Perception. Academic Press, New York.

Eva, E., Solarino, S., Eva, C., 1997. Stress tensor orientation derived from fault plane solutions in the southwestern Alps. J. of Geophys. Res., 102, 8171–8185.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27,861-874.

Fréchet, J., 1978. Sismicité du Sud-Est de la France, et une nouvelle méthode de zonage sismique. Thèse 3ème cycle, Grenoble, 159p

Frohlich, C., Davis, S.D., 1993. Teleseismic b-values; or, much ado about 1.0. J. Geophys. Res. 98, 631–644.

Gardner, J.K., Knopoff, L., 1974. Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?. Bull. Seismol. Soc. Am. 64, 1363-1367.

Gentili, S., Bressan, G., 2008. The partitioning of radiated energy and the largest aftershock of seismic sequences occurred in the northeastern Italy and western Slovenia, J. Seismol. 12, 343–354.

Gentili, S., 2010. Distribution of Seismicity Before the Larger Earthquakes in Italy in the Time Interval 1994–200. Pure Appl. Geophys., 167, 933–958.

Grandori, G., Guagenti, E., Perotti, F., 1984, Some observations on the probabilistic interpretation of short-term earthquake precursors. Earthquake engineering and structural dynamics, 12, 749-760.

Gruppo di Lavoro CSTI, 2005. Catalogo Strumentale dei Terremoti Italiani dal 1981 al 1996 (Versione 1.1). http://gaspy.df.unibo.it/paolo/gndt/Versione1_1/Leggimi.htm

Gruppo di Lavoro INGV sul terremoto di Amatrice, 2016. Primo rapporto di sintesi sul Terremoto di Amatrice Ml 6.0 del 24 Agosto 2016 (Italia Centrale), doi: 10.5281/zenodo.61121.

Gulia, L., Wiemer, S., 2010. The influence of tectonic regimes on the earthquake size distribution: A case study for Italy. Geophys. Res. Lett., 37, L10305, doi:10.1029/2010GL043066.

Gutenberg, B., Richter, C.F., 1956. Earthquake magnitude, intensity, energy and acceleration. Bull. Seism. Soc. Am., 46: 105-145.

Hainzl, S., 2004, Seismicity patterns of earthquake swarms due to fluid intrusion and stress triggering. Geophys. J. Int., 159, 1090–109.

Hainzl, S., 2016. Rate-Dependent Incompleteness of Earthquake Catalogs. Seismological Research Letters 87: 337-344.

Helmstetter, A.,Y. Y. Kagan, Jackson, D. D. (2006). Comparison of short-term and time-independent earthquake forecast models for southern California, Bull. Seismol. Soc. Am. 96, 90–106.

Hunstad, I., Selvaggi, G., D'Agostino, N., England Clarke, P., Pierozzi, M., 2003. Geodetic strains in peninsular Italy between 1875 and 2001. Geophys. Res. Lett. 30, 1828, doi:10.1029/ 2002GL016447

Jang, L., Sun, C.T. , Mizutani, E., 1997. Neuro-fuzzy and Soft Computing, Prentice-Hall, Englewood Cliffs, NJ.

Jones, L. M., Foreshocks and time-dependent earthquake hazard as- sessment in southern California, Bull. Seismol. Soc. Am., 75, 1669- 1679, 1985.

Kagan, Y. Y., 2002. Seismic moment distribution revisited: I. Statistical results. Geophys. J. Int. 148, 520–541.

Kagan, Y.Y., 2004. Short-Term properties of earthquake catalogs and models of earthquake source. Bull. Seism. Soc. Am., 94: 1207-1228.

Kastelic, V., Vannoli, P., Burrato, P., Fracassi, U., Tiberti, M.M., Valensise, G., 2013. Seismogenic sources in the Adriatic Domain, Mar. Petrol. Geol. 42, 191–213, doi:10.1016/j.marpetgeo.2012.08.002

Kiratzi, A. A., Papazachos, C. B., 1995. Active seismic deformation in the southern Aegean Benioff zone. J. Geodynamics, 19, pp. 65-78

Knopoff, L., 2000. The magnitude distribution of declustered earthquakes in Southern California, Proc. Natl. Acad. Sci. USA 97, 11880–11884.

Kossobokov, V.G., Romashkova, L.L., Keilis-Borok. V.l. and Healy, J.H., 1999. Testing earthquake prediction algorithms: statistically significant advance prediction of the largest earthquakes in the Circum-Pacific, 1992-1997. Phys. Earth and Planet. Inter. 111. 187-196.

Lolli, B., Gasperini, P., 2003. Aftershocks hazard in Italy Part I: Estimation of time-magnitude distribution model parameters and computation of probabilities of occurrence. J. Seismol. 7, 235–257.

Lolli, B., Gasperini, P., 2006. Comparing different models of aftershock rate decay: The role of catalog incompleteness in the first times after main shock. Tectonophysics 423, 43–59.

Marzorati, S., Cattaneo C., 2016. Stima automatica della magnitudo minima rilevabile dalla rete sismica ReSIICO, Quaderni di Geofisica, 136, pp.21

McNamara, D. E., Boaz, R. I., 2011. PQLX: A seismic data quality control system description, applications, and users manual, U.S.Geol. Surv. Open-File Rept. 2010-1292, 41.

McNamara, D. E., Buland, R. P., 2004. Ambient noise levels in the continental United States, Bull. Seismol. Soc. Am. 94, 1517–1527, doi: 10.1785/012003001.

Malinverno, A., Ryan, W.B.F., 1986. Extension in the Tyrrenian Sea and shortening in the Apennines as a result of arc migration driven by sinking of the lithosphere. Tectonics, 5, 227-245.

Mignan, A., 2011. Retrospective on the Accelerating Seismic Release (ASR) hypothesis: Controversy and new horizons. Tectonophysics, 505, 1–16, doi:10.1016/j.tecto.2011.03.010.

Mignan, A., Woessner J., 2012. Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-00180805 Available at http://www.corssa.org.

Milano, G.,Ventura, G., Di Giovambattista , R., 2002. Seismic evidence of longitudinal extension in the southern Apennines chain (Italy): The 1997–1998 Sannio-Matese seismic sequence, Geophys. Res. Lett., 29(20), 2004, doi:10.1029/2002GL015188.

Milano, G., Di Giovambattista, R., 2011. Seismicity at the border between Central and Southern Apennines (Italy): re-evaluation of the early 1984 instrumental earthquake. Tectonophysics, 499 (2011), pp. 92–104 http://dx.doi.org/10.1016/j.tecto.2010.12.008

Milano, G., Di Giovambattista, R., Ventura, G., 2005. Seismic constraints on the presentday kinematics of the Gargano foreland, Italy, at the transition zone between the southern and northern Apennine belts. Geophysical Research Letters, 32, L24308, doi:10.1029/2005GL024604.

Milano G., Di Giovambattista R., Ventura G., 2008. Seismic activity in the transition zone between Southern and Central Apennines (Italy): evidences of longitudinal extension inside the Ortona-Roccamonfina tectonic line. Tectonophysics, 457(1–2):102-110.

Miller, S. A., Collettini, C., Chiaraluce, L., Cocco, M., Barchi, M., Kaus, B. J. P., 2004, Aftershocks driven by a high-pressure CO2 source at depth, Nature, 427, 724–727.

Molchan, G. M., Dmitrieva, O. E., 1991. Identification of aftershocks: Review and new approaches, Comput. Seismol., 24, 19–50 (in Russian).

Molchan, G. M., Kagan, Y. Y., 1992. Earthquake prediction and its optimization, J. Geophys. Res. 97, 4823–4838.

Monaco, C., Mazzoli, S., Tortorici, L., 1996. Active thrust tectonics in western Sicily (southern Italy): The 1968 Belice earthquake sequence, Terra Nova, 8, 372–381.

Montone, P., Mariucci, M.T., Pondrelli, S., Amato A., 2004. An improved stress map for Italy and surrounding regions (Central Mediterranean) J. Geophys. Res., 109, p. 10410

Narteau, C., Byrdina, S., Shebalin, P., Schorlemmer, D., 2009. Common dependence on stress for the two fundamental laws of statistical seismology. Nature, 462, 642–645, doi:10.1038/nature08553.

Oldow, J. S., D'Argenio, B., Ferranti, L., Pappone, G., Marsella, E., Sacchi, M.,1993. Large-scale longitudinal extension in the southern Apennines contractional belt, Italy. Geology, 21, 1123- 1126.

Pace, B., Peruzza , L., Lavecchia , G., Boncio, P., 2006. Layered seismogenic source model and probabilistic seismic-hazard analyses in Central Italy, Bull. Seism. Soc. Am. 96, 107–132.

Passarelli, L., Hainzl, S., Cesca, S., Maccaferri, F., Mucciarelli, M., Roessler, D., Corbi, F., Dahm, T., 2015. Aseismic transient driving the swarm-like seismic sequence in the Pollino range, Southern Italy. Geophys. J. Int., 201, 1553–1567.

Peresan, A., Kossobokov, V. Romashkova, L., Panza, G. F., 2005. Intermediate-term middle-range earthquake predictions in Italy: a review. Earth-Science Reviews, 69, 97-132.

Pondrelli, S., et al., 2002. European-Mediterranean regional centroid-moment tensors: 1997 – 2000, Phys. Earth Planet. Inter., 130, 71 – 101

Powers, D.M.W., 2011. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

Reasenberg, P.,1985. Second-order moment of central California seismicity, 1969-82, J. Geophys. Res., 90, 5479–5495.

A. Rovida, R. Camassi, P. Gasperini e M. Stucchi (a cura di), 2011. CPTI11, la versione 2011 del Catalogo Parametrico dei Terremoti Italiani. Istituto Nazionale di Geofisica e Vulcanologia, Milano, Bologna. DOI: http://doi.org/10.6092/INGV.IT-CPTI11

Schorlemmer, D., Neri, G., Wiemer, S., Mostaccio, A., 2003. Stability and significance tests for b-value anomalies: Example from the Tyrrhenian Sea. Geophys. Res. Lett, 30(16), 1835, doi:10.1029/2003GL017335.

Schorlemmer, D., Wiemer, S., 2005. Microseismicity data forecast rupture area. Nature, 434, 1086, doi:10.1038/4341086a.

Schorlemmer, D., Wiemer, S., Wyss, M., 2005. Variations in earthquake-size distribution across different stress regimes. Nature 437, 539–542.

Selvaggi, G., Amato, A., 1992. Subcrustal earthquakes in the northern Apennines (Italy): evidence for a still active subduction?. Geophys. Res. Lett., 19, 2127–2130.

Shcherbakov, R., Turcotte, D. L., 2004, A Modified Form of Båth's Law. Bull. Seism. Soc. Am., 94, 1968–1975.

Sorgi, C., Deffontaines, B., Hippolyte, J.C., Cadet, J.P., 1998. An integrated analysis of transverse structures in the northern Apennines, Italy. Geomorphology 25, 193–206.

Sue, C., Thouvenot, F., Frechet, J., Trichart, P., 1999. Widespread extension in the core of the western Alps revealed by earthquake analysis. Journal of Geophysical Research, 104, 25611–25622.

Sweets, J. A,, Dawes, R. M., Monahan, J., 2000. Better decisions through science. Scientific American, 283, 82-87.

Tahir, M., Grasso, JR., Amorese D., 2012. The largest aftershock: how strong, how far away, how delayed?. Geophys. Res. Lett., 39: L04301, 10.1029/2011GL050604.

Tahir, M., Grasso, JR., 2015. Faulting Style Controls for the Space–Time Aftershock Patterns. Bulletin of the Seismological Society of America, 105, 5,2480–2497.

Tormann, T., Wiemer, S., Mignan, A., 2014. Systematic survey of high-resolution b value imaging along Californian faults: Inference on asperities J. Geophys. Res. Solid Earth, 119, 2029–2054.

Uhrhammer, R.,1986. Characteristics of Northern and Central California Seismicity, Earthquake Notes, 57(1), 21.

Valensise G., Pantosti D., 2001. The investigation of potential earthquake sources in peninsular Italy: a review, J. Seism., 5, 287–306.

Valensise, G., Pantosti, D., Basili, R. 2004. Seismology and tectonic setting of the 2002 Molise, Italy, earthquake. Earthquake Spectra, 20 (2004), pp. 23–37 http://dx.doi.org/10.1193/1.1756136

Vannucci, G., Pondrelli, S., Argnani, A., Morelli, A., Gasperini, P.. Boschi, E., 2004. An Atlas of Mediterranean Seismicity. Annals of Geophysics, suppl. To Vol. 47, n.1, 247-306, with CD-ROM enclosed.

Van Stiphout, T., Zhuang, J., Marsan, D., 2012. Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-52382934. Available at http://www.corssa.org.

Ventura, G., and Di Giovanbattista, R., 2013. Fluid pressure, stress field and propagation style of coalescing thrusts from the analysis of the 20 May 2012 ML 5.9 Emilia earthquake (Northern Apennines, Italy): Terra Nova, v. 25, p. 72– 78, doi:10.1111/ter.12007

Vorobieva, I.A., 1999. Prediction of a subsequent large earthquake. Physics of the Earth and Planetary Interiors. Vol. 111, p.197-206.

Vorobieva, I.A., Panza, G.F., 1993. Prediction of the Occurrence of Related Strong Earthquakes in Italy. PAGEOPH, vol. 141(1) p.25-41.

Wells, D.L. , Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, Bull. Seism. Soc. Am. 84, 974–1002.

Westaway, R., 1992. Seismic moment summation for historical earthquakes in Italy: tectonic implications. J. Geophys. Res., 97, pp. 15,437–15,464.

Wiemer, S., Wyss, M., 2000. Minimum magnitude of complete reporting in earthquake catalogs: examples from Alaska, the western united states, and japan, Bull. Seismol. Soc. Am., 90, 859–869.

Woessner, J., Wiemer, S., 2005. Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, Bull. Seismol. Soc. Am., 95, doi:10.1785/012040007

Zakharova, O., Hainzl, S., Bach, C., 2013. Seismic moment ratio of aftershocks with respect to mainshocks, Journal of Geophysical Research, 104, 5856–5864.

Zhuang, J., 2015 (a). Detecting spatial variations of earthquake clustering parameters via maximum weighted likelihood estimates, 9th International Workshop on Statistical Seismology (StatSei9) Potsdam, Germany, 14 - 19 June 2015.

Zhuang, J., 2015 (b). Weighted likelihood estimators for point processes. Spatial Statistics, 14 166-178.

Zhuang, J., Guo, Y., Murru, M., Falcone, G., Taroni, M., Console, R., 2017. An extensive study of clustering features of seismicity in Italy during 2005 to 2016. Geophysical Research Abstracts Vol. 19, EGU2017-7433, 2017 EGU General Assembly 2017, Vienna, Austria, 23–28 April 2017.

Table Captions

Table 1: Detected clusters. M_m=mainshock magnitude, Latitude and Longitude=mainshock latitude and longitude, N_{aft}=number of aftershocks in the cluster, M_c=Completeness magnitude, Type=cluster classification, Dm=difference in magnitude between the mainshock (see operative definition of mainshock in the text) and the strongest subsequent earthquake.

Table 2: Values of thresholds and of S_2 for each feature. Thresholds are shown only for $s_2 \leq S_2$ and for Informedness>0.

Table 3: Weights for each feature and time interval for the classification.

Figure Captions

Figure 1: Major Italian earthquakes from CPTI11 (*Rovida et al., 2011)* and seismogenic sources from DISS Working Group (2015). N. Apennines = Northern Apennines; C. Apennines = Central Apennines; S. Apennines = Southern Apennines.

Figure 2: ROC graph for mainshock-based features

Figure 3: Epicentral distribution as a function of the class. Red stars: A class clusters; Blue stars: B class clusters. Rows: area subdivision by decision tree

Figure 4: Rake angle as a function of the difference *Dm* between the mainshock and the strongest subsequent event magnitude. Orange and blue points represent the rakes of the two focal planes. Red rectangles correspond to strike-slip mechanisms, green to normal and blue to inverse.

Figure 5: Feature N_2 performance. Black lines: the feature is calculated for every time interval. Red lines: the feature is calculated only for 6 and 12 hours and kept constant for longer periods.

Figure 6: ROC diagram for the selected time intervals. The time interval of 7 days is not shown because $S_2 \leq 6$ days for all features. The observation time (s_2) is shown at the top of each figure.

Figure 7: Informedness of features for different time intervals.

Figure 8: Prob(A): Probability to have an A cluster. Each line corresponds to a different cluster.

Red lines: type A clusters. Blue lines: type B clusters.

Tables

Date	Time	M_m	Latitude	Longitude	N_{aft}	M_c	Type	Dm
23/11/1980	18:34:52	6.5	40.8	15.37	41	3.0±0.3	B	1.6
15/08/1982	15:09:54	4.5	40.94	15.32	2	-	A	1
29/04/1984	05:02:59	5.2	43.2	12.59	135	2.0±0.1	A	1
07/05/1984	17:49:42	5.8	41.67	13.82	112	2.5±0.1	A	0.6
08/01/1986	00:27:21	4.5	42.69	15.62	11	4.2±0.5	A	0.3
26/04/1988	00:53:45	5.1	42.21	16.66	43	3.1±0.2	A	0.4
05/05/1990	07:21:17	5.6	40.65	15.88	40	2.3±0.1	B	1.1
05/06/1993	19:16:17	4.5	43.12	12.72	73	1.7±0.2	B	2
30/09/1995	10:14:34	5.4	41.79	15.97	34	1.9±0.1	B	1.7
10/10/1995	06:54:22	4.8	44.11	10.00	17	1.8±0.2	B	2
13/04/1996	13:00:22	4.6	46.32	12.59	20	1.3±0.3	B	1.1
15/10/1996	09:56:00	5.5	44.8	10.68	70	2.0±0.2	B	1.4
12/05/1997	13:50:15	4.5	42.76	12.53	157	1.7±0.1	A	1
26/09/1997	00:33:13	5.6	43.02	12.89	550	1.6±0.1	A	-0.2
12/04/1998	10:55:33	5.6	46.28	13.57	24	1.7±0.1	B	1.4
09/09/1998	11:28:00	5.6	40.06	15.95	47	1.9±0.1	B	1.9
29/12/1999	20:42:35	4.6	46.61	10.22	12	1.8±0.7	B	1.9
21/08/2000	17:14:28	4.6	44.77	8.43	35	1.8±0.1	A	-0.2
06/09/2002	01:21:29	5.6	38.38	13.65	164	2.1±0.1	A	1
27/10/2002	02:50:27	4.8	37.77	15.11	116	2.7±0.2	A	0.5
29/10/2002	10:02:24	4.5	37.61	15.14	37	2.4±0.1	A	0.4
31/10/2002	10:32:59	5.4	41.72	14.89	122	1.8±0.1	A	0.1
27/03/2003	16:10:38	4.7	43.09	15.42	143	3.2±0.1	A	-0.7
11/04/2003	09:26:58	4.7	44.76	8.87	15	2.5±0.2	B	1.4
14/09/2003	21:42:53	5	44.26	11.38	208	2.2±0.1	B	1.4
12/07/2004	13:04:06	5.2	46.36	13.64	48	2.2±0.2	B	1.8
24/11/2004	22:59:39	5.2	45.69	10.52	11	2.1±0.3	B	2.5

29/05/2006	02:20:06	4.5	41.8	15.9	51	1.7±0.2	B	1.7
23/11/2006	13:31:54	4.6	35.88	12.68	3	-	A	1
01/03/2008	07:43:13	4.5	44.06	11.25	127	1.9±0.1	A	0.5
23/12/2008	15:24:22	4.9	44.54	10.35	230	2.3±0.1	A	0.5
05/04/2009	20:20:53	4.5	44.23	11.91	45	1.6±0.2	B	1.2
06/04/2009	01:32:40	6.1	42.34	13.38	208	1.4±0.1	A	0.7
07/09/2009	21:26:30	4.6	38.73	14.04	8	-	A	1
23/06/2011	22:02:47	4.5	38.06	14.78	285	1.8±0.2	A	0.6
17/07/2011	18:30:27	4.5	45.01	11.37	11	2.3±0.4	B	1.5
25/01/2012	08:06:37	4.9	44.87	10.51	22	2.1±0.1	B	1.3
27/01/2012	14:53:13	4.9	44.52	10.01	10	2.5±0.4	B	1.8
20/05/2012	02:03:50	5.8	44.9	11.26	424	2.2±0.1	A	0.2
28/08/2012	23:12:15	4.5	38.2	15.73	17	1.6±0.2	A	0.1
25/10/2012	23:05:25	5.2	39.88	16.02	177	1.4±0.1	B	1.6
25/01/2013	14:48:18	4.8	44.16	10.45	125	1.4±0.1	B	1.5
16/02/2013	21:16:09	4.8	41.71	13.57	22	1.2±0.2	B	1.6
21/06/2013	10:33:56	5.1	44.09	10.06	230	1.2±0.1	A	0.6
21/07/2013	01:32:24	4.9	43.51	13.72	73	1.8±0.1	A	0.7
29/12/2013	17:08:43	5	41.4	14.43	75	1.4±0.1	A	0.8
07/04/2014	19:27:00	4.7	44.5	6.71	44	1.2±0.1	B	1.4

Table 1

Feature	S_2	Threshold								
		6 h	12 h	18 h	1 day	2 days	3 days	4 days	5 days	6 days
N	12 h	6.5	7.5							
N_2	12 h	0.5	3.5							
S	12 h	0.016	0.016							
Z	12 h	0.004	0.004							
SLCum	2 days	-	0.1044	0.145	0.145	0.24				
QLCum	1 day	-	3.1	4.1	4.1					
SLCum2	2 days	-	0.1044	0.1044	0.1044	0.1044				
QLCum2	12 h	-	3.1							
Q	12 h	0.002	0.002							
V_m	18 h	5.0	5.6	8.6						
V_{med}	6 days	-	-	-	-	-	-	-	-	0.9
V_n	6 days	-	-	-	-	-	7.5	15.5	15.5	19.5

Table 2

	0 h	6 h	12 h	18 h	1 day	2 days	3 days	4 days	5 days	6 days	7 days
Epicenter	1.00	0.15	0.08	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06
N	0.00	0.18	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.08	0.08
N_2	0.00	0.10	0.12	0.10	0.10	0.10	0.10	0.09	0.09	0.09	0.09
S	0.00	0.12	0.14	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.11
Z	0.00	0.16	0.12	0.10	0.10	0.10	0.10	0.09	0.09	0.09	0.09
SLCum	0.00	0.00	0.09	0.09	0.09	0.11	0.11	0.10	0.10	0.10	0.10

QLCum	0.00	0.00	0.10	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.09
SLCum2	0.00	0.00	0.00	0.09	0.09	0.11	0.11	0.10	0.10	0.10	0.10
QLCum2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q	0.00	0.12	0.14	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.11
V_m	0.00	0.18	0.10	0.10	0.10	0.10	0.10	0.09	0.09	0.09	0.09
V_{med}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.04
V_n	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.05	0.05	0.06	0.06

Table 3.

Figures

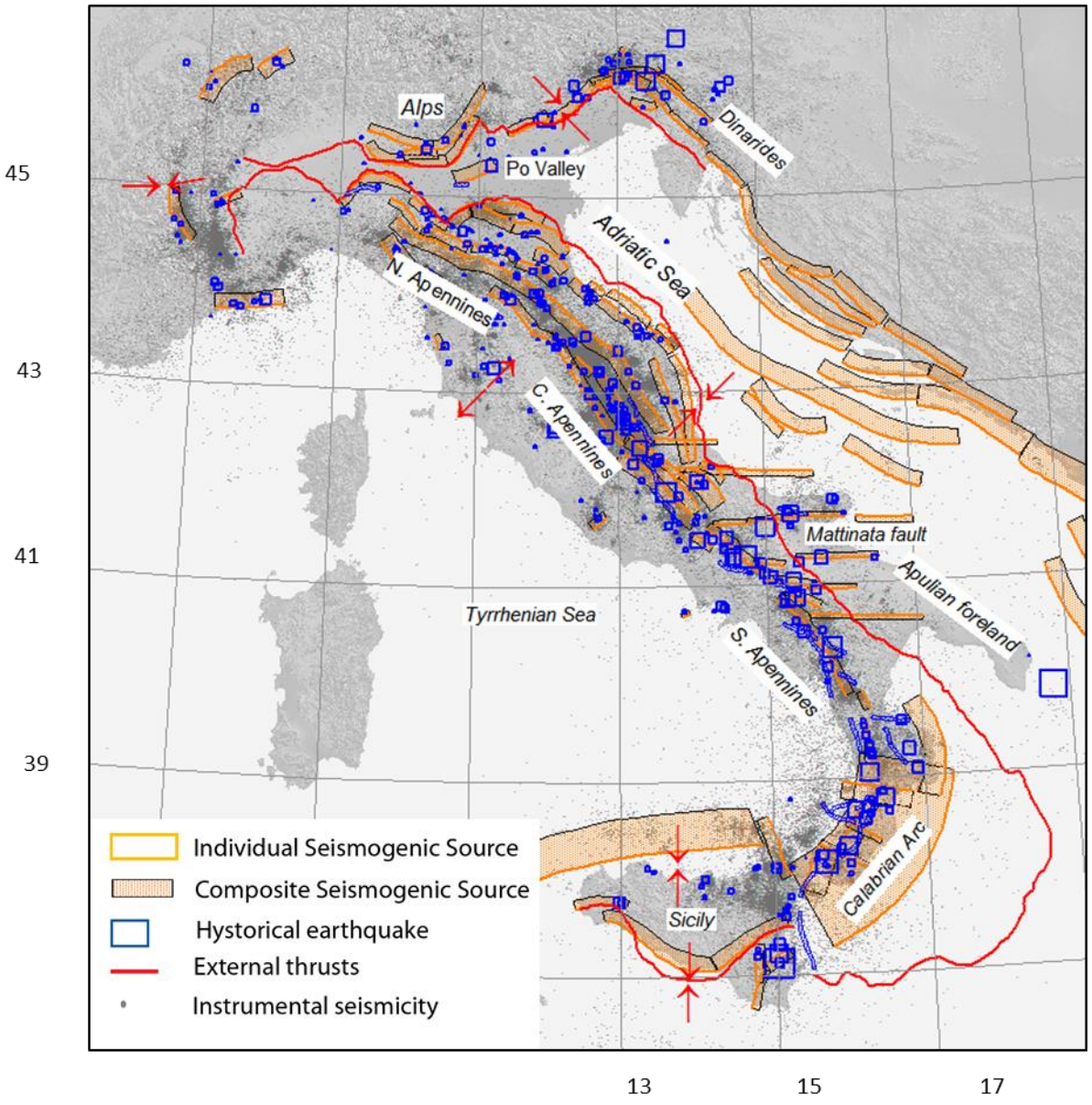


Figure 1.

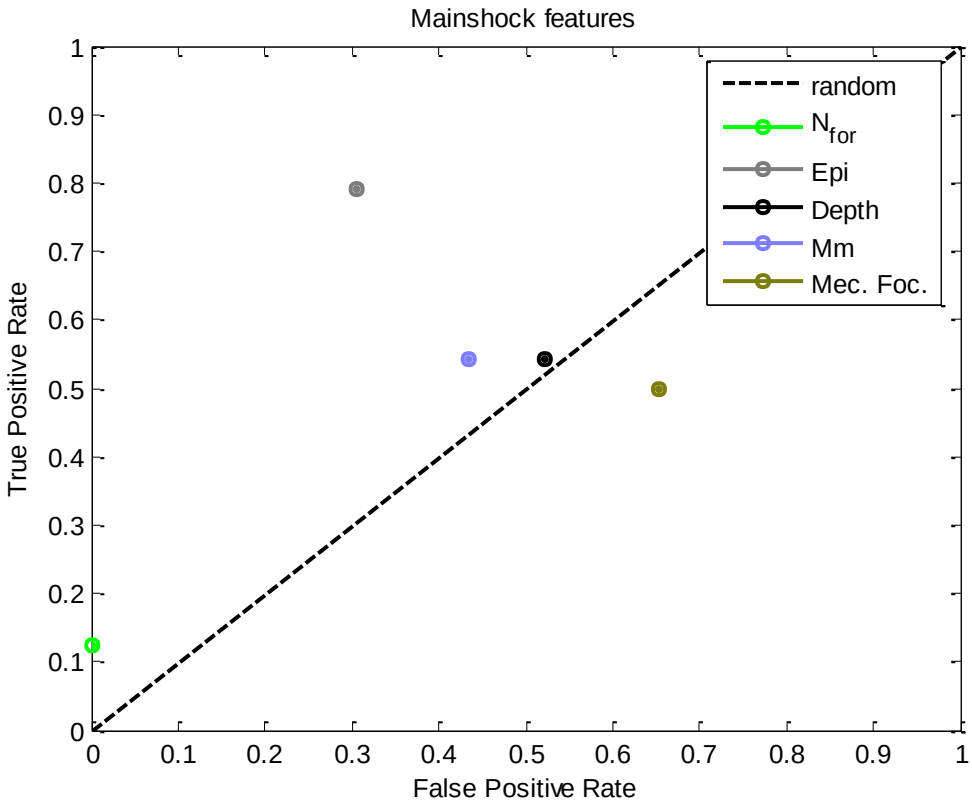


Figure 2.

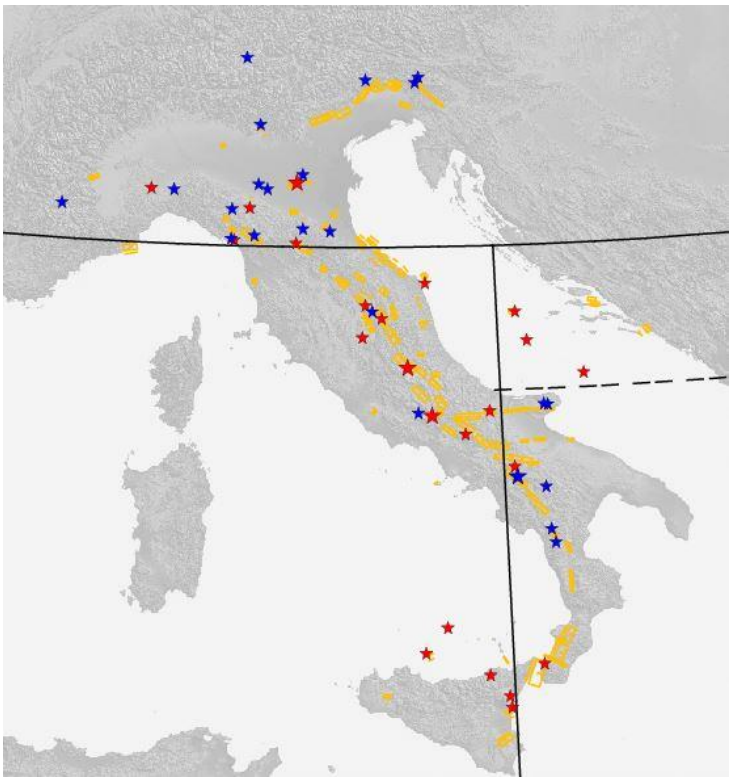


Figure 3.

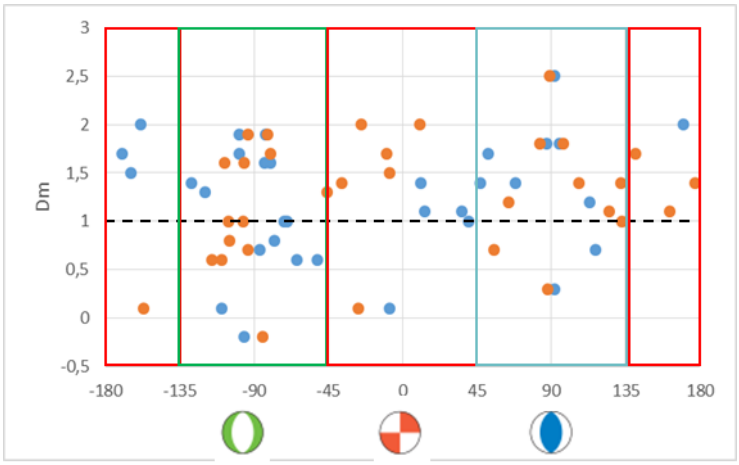


Figure 4.

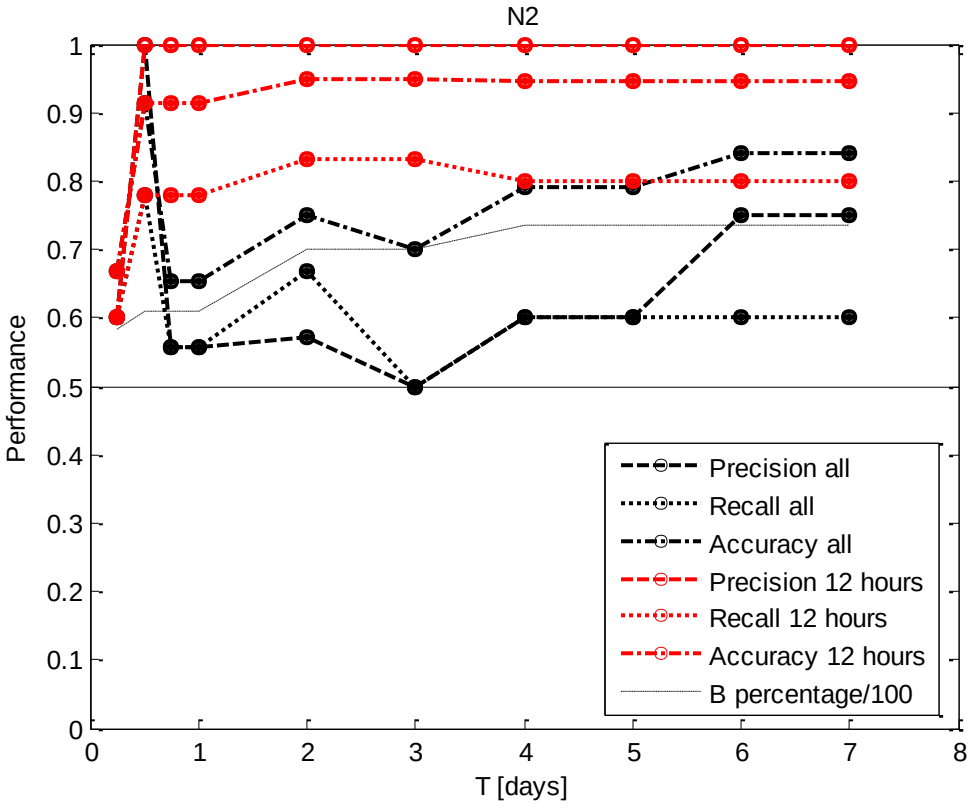
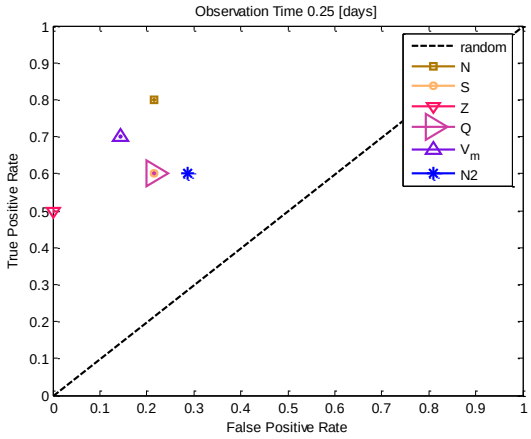
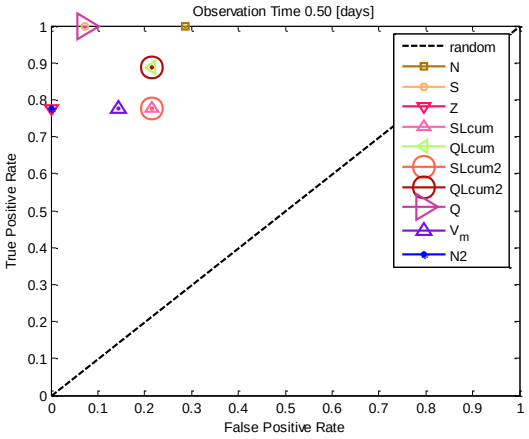


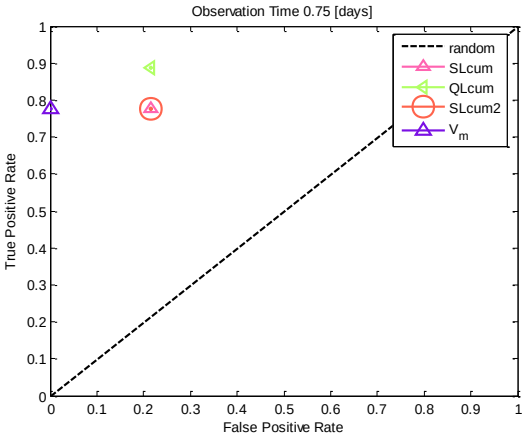
Figure 5.


Observation Time 0.25 [days]
True Positive Rate
False Positive Rate
random
N
S
Z
Q
V_m
N2

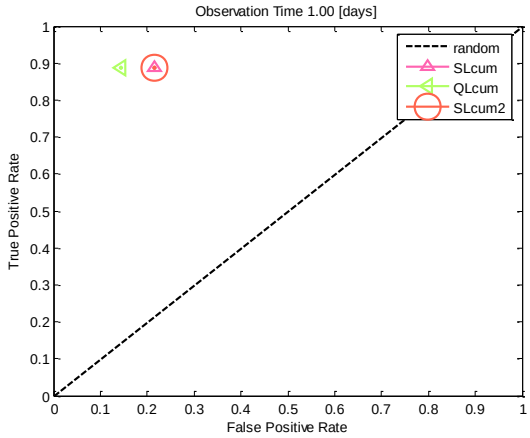
(a)


Observation Time 0.50 [days]
True Positive Rate
False Positive Rate
random
N
S
Z
SLcum
QLcum
SLcum2
QLcum2
Q
V_m
N2

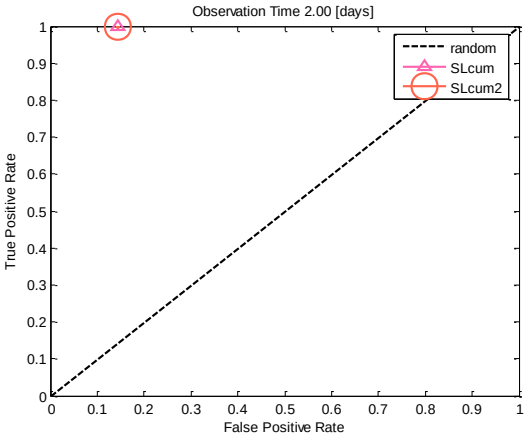
(b)


Observation Time 0.75 [days]
True Positive Rate
False Positive Rate
random
SLcum
QLcum
SLcum2
V_m

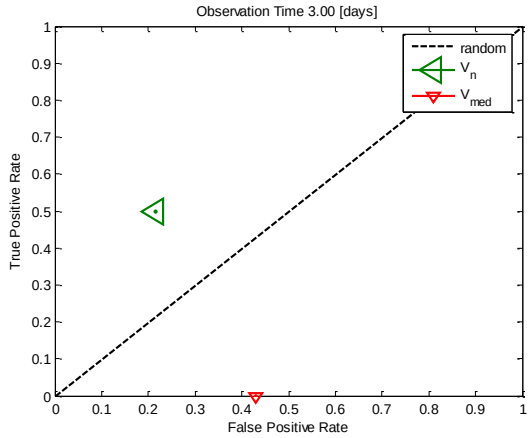
(c)


Observation Time 1.00 [days]
True Positive Rate
False Positive Rate
random
SLcum
QLcum
SLcum2

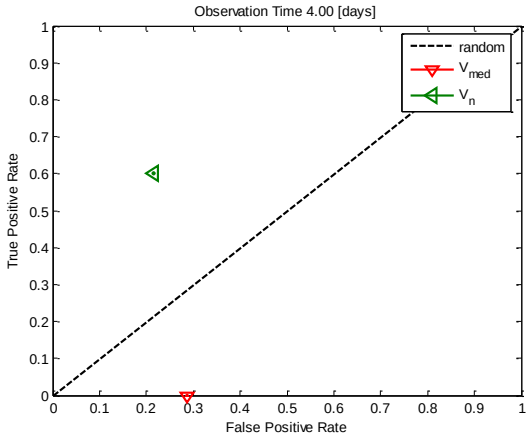
(d)


Observation Time 2.00 [days]
True Positive Rate
False Positive Rate
random
SLcum
SLcum2

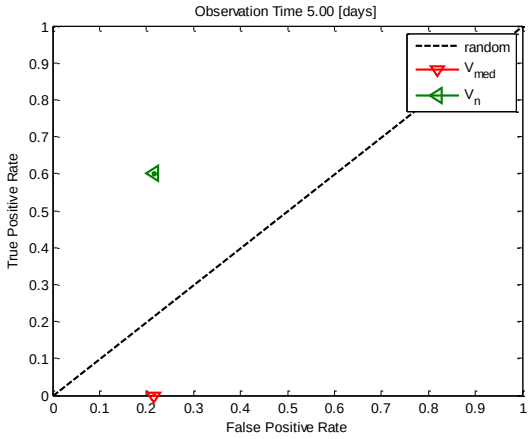
(e)


Observation Time 3.00 [days]
True Positive Rate
False Positive Rate
random
V_n
V_med

(f)


Observation Time 4.00 [days]
random
V med
V n
True Positive Rate
False Positive Rate

(g)


Observation Time 5.00 [days]
random
V med
V n
True Positive Rate
False Positive Rate

(h)

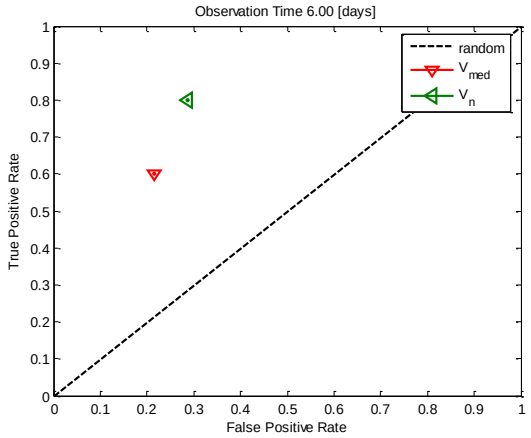

Observation Time 6.00 [days]
random
V med
V n
True Positive Rate
False Positive Rate

Figure 6.

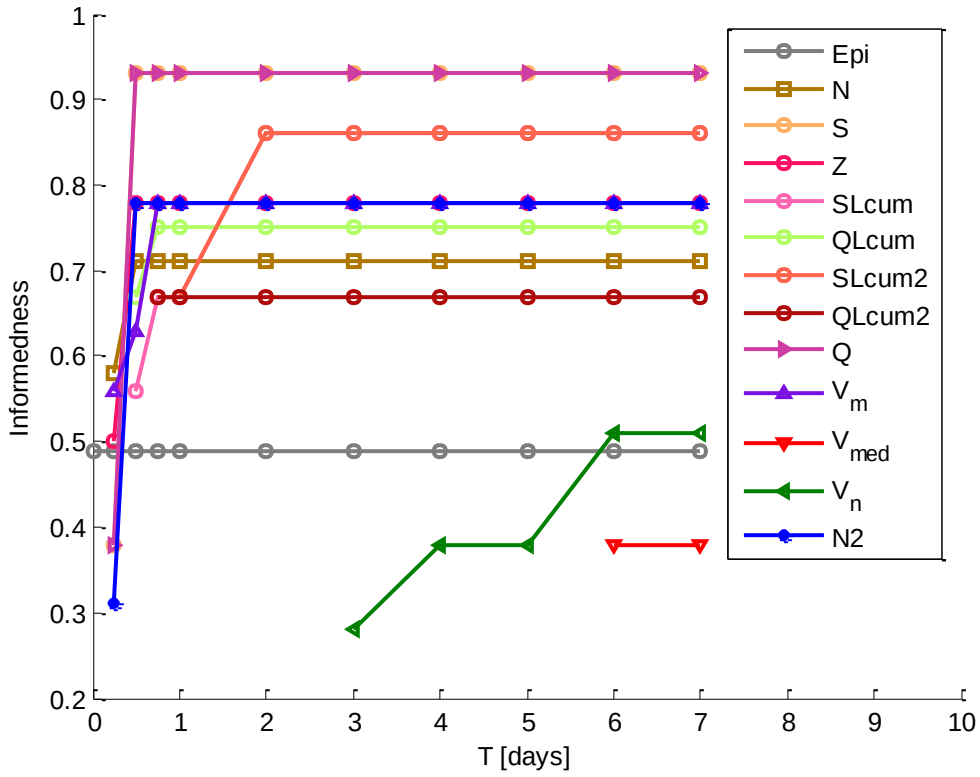


Figure 7.

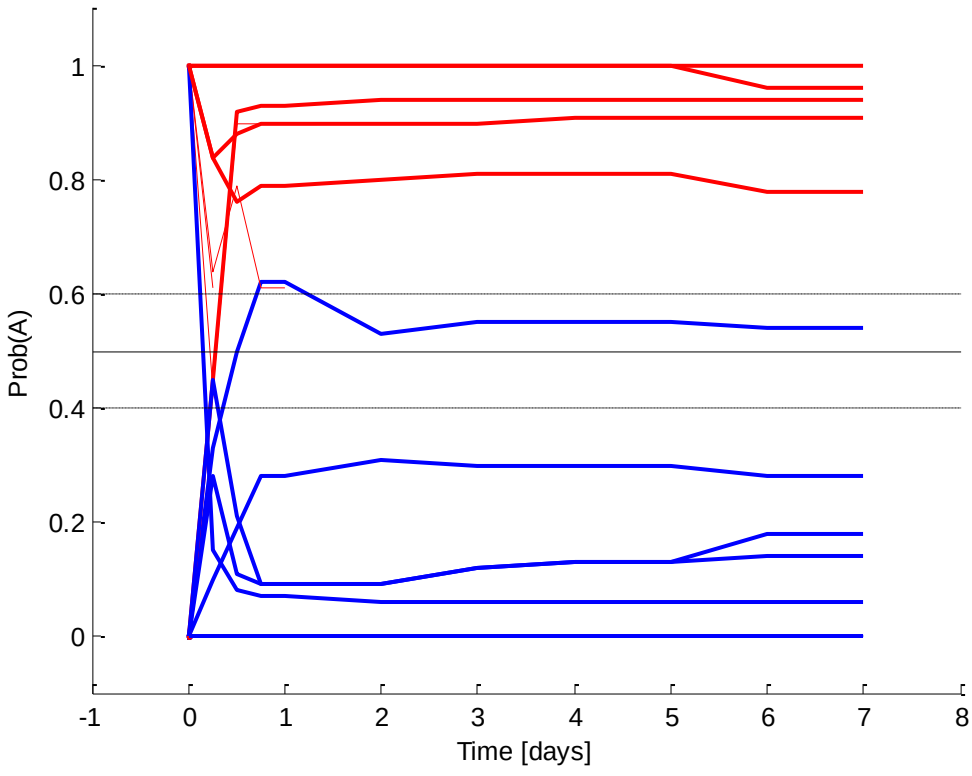


Figure 8.

Appendix A

One possible approach to the classification problem is to build a learning classifier. A learning sample (training set) *P* consists of data $\{(\vec{p}_1, \omega_1), (\vec{p}_2, \omega_2), ...(\vec{p}_n, \omega_n)\}$ where $\vec{p}_i$ (i = 1,…, n) are the patterns corresponding to different classes ω_i (ω_i may be "A" or "B"). A decision tree (e.g., Breiman et al, 1984, Jang et al., 1997) is constructed (trained) by defining repeated splits of subsets of the measurement space *X* into two or more descendant subsets, beginning with *X* itself. Because we have two output classes the decision trees we developed are binary classifiers called "binary decision trees".

The training algorithm we adopted is the following:

1. The training set is processed by a decision node that is a unit that decides how to subdivide the training set in two groups corresponding to the two classes of the problem. If the training set is composed of less than three instances of one of the classes, the node becomes a leaf node labelled with the class with more instances and the training ends for that subset. Otherwise, the training set is split into two disjoint subsets.
2. A new level of 2 nodes (children nodes) is added to the tree. Each subset is assigned to a different child. If one or more subsets are entirely assigned to a class Ω_i, the corresponding children node becomes a leaf node labelled with the class Ω_i and the training ends for that subset.
3. The nodes corresponding to subsets composed of different classes are transformed into decision nodes trained by the corresponding subset patterns.
4. The procedure continues in a recurrent way following the steps 1-3 for all children.
5. When all patterns reach a leaf, the algorithm ends.

Each leaf has a unique path that starts with the root node and ends with the leaf that corresponds to a splitting rule that is a conjunction of all the splitting conditions imposed by the decision-making units along the path.

When the training is ended, the tree structure together with all the conditions imposed by the decision nodes are stored and used as a classification system (classifier). The classifier can be used to classify new instances not used for the training. In our case, given a new cluster, its features can be evaluated and classified with the previously trained corresponding decision tree. The classification algorithm is the following:

Let R be the root node of the tree, a pattern to be classified, and v the current node.

1. Set $v=R$.
2. If v is a leaf node with label Ω_i, then the pattern $\vec{p}$ is assigned to the class Ω_i and the algorithm ends.
3. If v is a splitting node, set $v = v_l$ or set $v = v_r$, where v_l and v_r are the left and right child node, depending on the splitting rule results for that pattern, and go to 2.

The decision trees we developed are usually referred to only one precursor. The resulting decision trees are thus defined by one or more thresholds on the analysed feature to subdivide the two classes.

To evaluate the classifier performance, it is necessary to define a test set, disjoint from the training set, in which each instance (corresponding to a given cluster) has a known class. The classifier output for the instance is subsequently compared with its class. In binary classifiers one class is defined as positive (we chose A class) or negative (B class). Theoretically, both the training set and the test set should be representative of all the characteristics of A and B clusters. In real applications, where only a finite and small set of examples is available, it is well known that learning the parameters of a prediction function and testing it on the same data is a methodological mistake: the final model usually overfits the training data or, in other words, it makes an overly

complex model to explain idiosyncrasies in the data under study that often have some degree of error or random noise. This reduces the generalization capability and therefore the predictive power; on the other hand, the error rate estimate is overly optimistic (lower than the true error rate) because the model would just repeat the labels of the samples that it has just seen.

It is therefore necessary to define an independent test set on which to perform the evaluation of the method. A possible solution to this problem is cross-validation. Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. In the basic approach, called k-fold cross-validation, the training set is split into k smaller sets, the model is trained using k-1 of the folds as training data and the resulting model is validated on the remaining part of the data. The cross-validation process is later repeated k times with each of the k subsamples used exactly once as the validation data. The Leave-One-Out (or LOO) method is a case of k-fold cross-validation in which each learning set is created by taking all the samples except one, the test set being the sample left out. The procedure is repeated for all of the samples (see Fig. 1).

Therefore, for n samples, we have n different training sets and n different test sets. LOO is an exhaustive cross-validation method, i.e., a method that learns and tests on all possible ways to divide the original sample into a training and a validation set. Different from the k-fold cross-validation, n models from n samples are built instead of k models, where $n>k$ and each is trained on n-1 samples rather than (k-1)n/k. Assuming that $k<<n$, LOO is more computationally expensive than k-fold cross validation. In terms of accuracy, LOO often results in high variance as an estimator for the test error. Intuitively, because n-1 of the n samples are used to build each model, models constructed from folds are virtually identical to each other and to the model built from the entire training set. However, in our work the LOO was preferable with respect to other k-fold cross-validation methods because of the small database involved.

In pattern recognition applications, the information on the performance of a binary classifier is stored in a two-by-two matrix called the *confusion matrix* (or *contingency table*) constructed using

the counts of the number of outcomes of the classification (i.e., TP, FN, TN and FP are the numbers of True Positives, False Negatives, True Negatives and False Positives, respectively).

Fig. 2 shows the *confusion matrix*. The numbers along the major diagonal represent the correct decisions made, and the numbers off this diagonal represent the errors—the confusion—between the various classes.

Another common way of evaluating the performance of a classifier is evaluating *Recall*, *Precision* and *Accuracy*.

The *Recall* (also called true positive rate, hit rate, or sensitivity) is defined as

$$\mathrm{Re}\,call = \frac{Positives\ correctly\ classified}{Positives} = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{1}$$

The *Precision* (also called positive predictive value, true positive accuracy or confidence) is defined as:

$$\mathrm{Pr}\,ecision = \frac{Positives\,correctly\,classified}{Classified\ as\ positive} = \frac{TP}{Y} = \frac{TP}{TP + FP} \tag{2}$$

Recall and Precision and their combinations focus only on the positive examples and predictions and neither of them captures any information about how well the model handles negative cases, i.e., TN is disregarded. For this reason, to supply a complete description of the classifier performances it is necessary to also evaluate the Accuracy, which considers the classification of negatives. The Accuracy (also called Rand Accuracy) is defined as:

$$Accuracy = \frac{Correctly\ classified}{All} = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

In pattern recognition applications, the performances are usually visualized by a Receiver Operating Characteristics (ROC) graph (Egan, 1975; Swets et al., 2000). This type of graph is used for organizing and selecting classifiers based on their performance (Fawcett, 2006). These characteristics have become increasingly important in the presence of unbalanced classes (Fawcett, 2006), i.e., cases, such as the one we are dealing with, in which one class has more instances with

respect to the other. ROC does not consider the Precision, but only the Recall (True Positive Rate) and the False Positive Rate (Fallout or Hit Rate) defined as:

$$False\ Positive\ Rate = \frac{Negatives\ incorrectly\ classified}{Total\ negatives} = \frac{FP}{N} = \frac{FP}{FP + TN} \quad (4)$$

In other words, a ROC graph depicts relative trade-offs between benefits (true positives) and costs (false positives). Fig. 3 shows a ROC graph.

A discrete classifier produces a point of coordinates (False Positive Rate, True Positive Rate) in the ROC graph:

- The point (0,0) corresponds to a classifier never issuing a positive classification; therefore, the classifier commits no false positive errors but also gains no true positives. In our tests, if we verified the performance of a constant feature, the decision trees for such features were simply one leaf classified as B because class B (negative) was larger than class A (positive). This classifier thus classifies all the test patterns into class B.
- Point (1,1) represents a classifier unconditionally issuing positive classifications. In our case, we should have had a low *Accuracy* and *Precision* corresponding to this point because of the large number of B samples.
- The point (0,1) represents a perfect classification.
- The point (1,0) represents a classifier that classifies all positive instances in the negative class and all negative instances in the positive class
- The diagonal represents random guessing. Any classifier that appears in the lower right triangle performs worse than random guessing.

One point in ROC space is better than another if it is closer to the point (0,1) - higher true positive rate and/or low false negative rate. Mochan diagrams (Molchan 1991, Molchan and Kagan 1992) are generally preferred in ROC graphs in earthquake forecasting methods evaluation because they account for the fact that earthquakes are clustered in space. In a Molchan diagram the Hit rate is

substituted by the Miss Rate (that is, 1-hit rate, i.e., Inverse Recall) whereas the false positive rate is substituted by the fraction of the space time region occupied by an alarm. This approach is not useful in our analysis because we are interested in predicting only if a strong aftershock will follow inside a cluster and not where or when, as in the strong mainshock forecasting approach. For this reason, we adopted the ROC graph, together with *Precision*, *Recall* and *Accuracy* information, to assess the quality of our classifiers.

Figure Captions

Fig. 1: A schematic representation of the LOO method.

Fig. 2: Confusion matrix.

Fig. 3: The ROC graph.

Figures

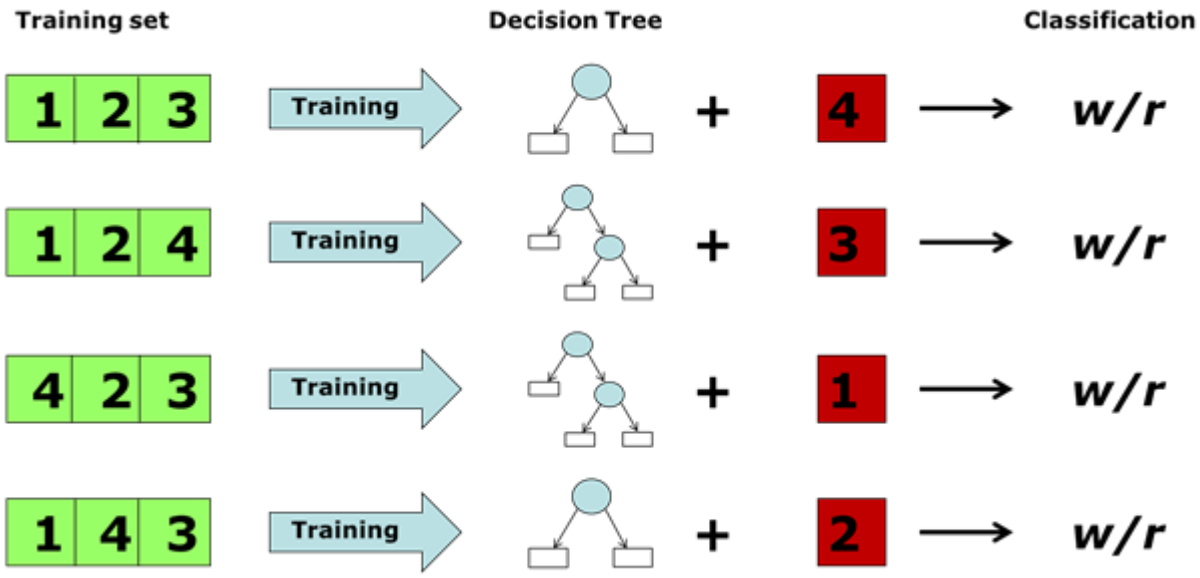


Figure1.

Hypothesized class	True class: p	True class: n
Y	True Positive (TP)	False Positive (FP)
N	False Negative (FN)	True Negative (TN)
Column totals	P	N

Figure 2.

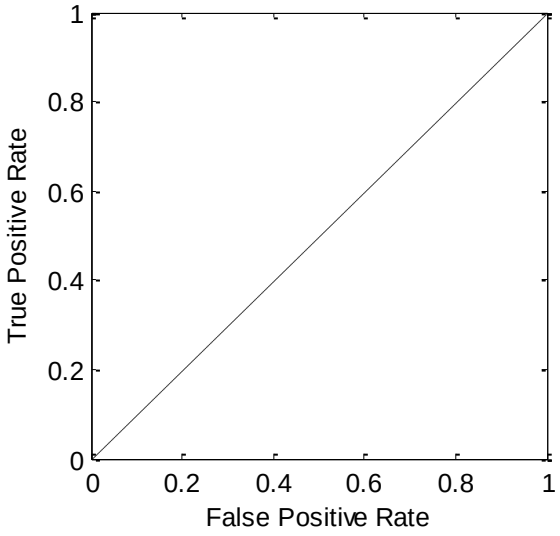


Figure 3.