IDENTIFICATION OF EARTHQUAKE CLUSTERS THROUGH A NEW SPACE-TIME-MAGNITUDE METRIC

Renata Rotondi [1], Antonella Peresan [2], Stefania Gentili [2] and Elisa Varini [1]

[1] CNR - Istituto di Matematica Applicata e Tecnologie Informatiche, Via Bassini 15, Milano (I), (e-mail: reni@mi.imati.cnr.it, elisa@mi.imati.cnr.it)

[2] Istituto Nazionale di Oceanografia e Geofisica Sperimentale, Udine (I), (e-mail: aperesan@inogs.it, sgentili@inogs.it)

ABSTRACT: We consider a metric that exploits the statistical properties of seismicity to quantify the correlation between earthquakes in a given catalogue. The method is based on nearest-neighbour distance between pairs of events in a combined space-time-magnitude domain and allows us to identify and analyse seismic clusters. We exemplify results from selected major earthquakes (i.e. Colfiorito 1997, L'Aquila 2009 and Emilia 2012), showing that the method can deal with data of different quality. Moreover, we show that this data-driven approach permits to disclose possible correlations and complex features in the internal structure of the identified clusters.

KEYWORDS: generalized distance, spanning tree, average leaf depth, aftershock/foreshock

1 Generalized earthquake distance

Earthquake clustering is an essential aspect of seismicity with signatures in space, time, and size domains that provide key information on earthquake dynamics. Clustering in space is shown by the concentration of earthquakes along regional fault networks. Clustering in time is primarily, but not only, associated with the increase of seismic activity immediately after large earthquakes leading to *aftershock sequences*. Despite the overall agreement on the existence of multiple types of clusters (denoted through various terms as *swarms, bursts,* etc.), a formal definition of seismic clusters is lacking, so as a unique method to identify them.

The main declustering algorithms present in the literature are divided into deterministic and stochastic: the former identify clustered events according to their membership to a user-specified time-space window, whereas the latter fit a branching model (e.g. versions of the ETAS model) to the data and, on the basis of the estimated intensity function, obtain the probabilities that each event is a background or a triggered event. Most of the studies on aftershocks concern major earthquakes characterized by prominent aftershock series clearly emerging from the background seismicity, whereas the behaviour of aftershock sequences related to small-to-medium size events is largely unsettled. We consider a method based on the distribution of the nearest-neighbour distances between pairs of earthquakes in a combined space-time-magnitude domain, and which exploits the statistical properties of seismicity. This method provides significant results also for small-medium size earthquakes.

Bak *et al.* (2002) analysed the distributions $F(t)$ of the waiting time between earthquakes occurring in California by varying the M magnitude threshold and the linear size L of the studied area; they noted that the curves collapse into a unified law when the coordinates are properly rescaled, that is, when t is replaced by $t\,10^{-bM}L^{d_f}$ and $F(t)$ by $t^{\alpha}F(t)$. The parameters α, b, and d_f are derived from well-known statistical properties of seismicity: the Omori-law exponent for aftershocks, the b value of the Gutenberg-Richter law and the fractal dimension of epicentres, respectively. Following this approach Baiesi & Paczuski (2004) proposed a metric which measures the correlation between earthquakes in terms of the Gutenberg-Richter (GR) distribution for the magnitude $F(M) \sim 10^{-bM}$ and the fractal appeareance of earthquake epicentres. Let us consider a catalogue which provides, for each event i, its occurrence time t_i, epicentre (θ_i, ϕ_i), and magnitude M_i. Given the earthquake j, its distance from any preceding earthquake i is defined as:

$$\eta_{ij} = t_{ij}\ (r_{ij})^{d_f}\ 10^{-bM_i}$$

where $t_{ij} = t_j - t_i$ is the interoccurrence time; r_{ij} is the spatial distance between the epicentres; d_f is the fractal dimension of the epicentres and b is the parameter of the GR law in the studied area. The event i^* that corresponds to the nearest-neighbour distance, $\eta_j^* := \min_i \eta_{ij}$, is called the *parent* of the j event, whereas the event j is called an *offspring* of i. By connecting each event with its nearest-neighbour one obtains a time-oriented tree where each event has a unique parent, and also may have multiple offspring. Since the larger is the distance η_{ij}, the weaker is the link between the events i and j, by removing the weak links (that is distances exceeding a threshold η_0) we identify the clusters of events. The *main* shock is the largest magnitude event, earthquakes observed prior to (later than) the main shock are considered as *foreshocks* (*aftershocks*). To fix the critical threshold η_0, Zaliapin & Ben-Zion (2013) analysed the distribution of η^* and the joint distribution of magnitude-normalized time and space components (T,R) of the nearest-neighbour distance η^*, being $T_{ij} := t_{ij}\,10^{-bM_i/2}$ and $R_{ij} := r_{ij}^{d_f}\ 10^{-bM_i/2}$. They simulated these distributions for three point processes: (i) homogeneous Poisson marked process, (ii) single self-excited aftershock series generated by

Omori law, (iii) ETAS model that combines the first two models. Their study shows that both the distributions of η^* and (T,R) are unimodal in the first two models whereas they are bimodal in the third case. This suggests approximating the distribution of η^* through a mixture of two Gaussian distributions $F(\eta^*) = \omega\, N(\eta^*;\mu_1,\Sigma_1) + (1-\omega)\, N(\eta^*;\mu_2,\Sigma_2)$ and choosing the threshold η_0 as the value that equalizes the densities of the two estimated Gaussian distributions: $N(\eta_0;\mu_1,\Sigma_1) = N(\eta_0;\mu_2,\Sigma_2)$.

The tree structure of the clusters may be quite different. For identification of the cluster type it is useful to introduce the concept of *vertex depth*, that is the minimal number d of links that connects a given vertex (earthquake) to the tree root (the first earthquake in the cluster). The *average leaf depth* $< d >$ - the vertex depth averaged over the tree leaves (vertices with no children) - provides a scalar measure, which characterizes the tree structure and turns out to be in agreement with the statistical properties of the clusters and with the local physical characteristics of the crust. For instance, a linearly shaped tree has much larger depth than a spray-shaped tree with the same number of leaves; the first tree type is named *swarm-type* and is mainly associated with relatively thin seismogenic zones manifested by shallow seismicity, whereas the second type indicates a *burst*-like behaviour corresponding to relatively thick seismogenic zones with higher seismicity depth (Zaliapin & Ben-Zion, 2013). We apply this methodology to data drawn from two different Italian seismic catalogues with the goal to estimate the sensitivity of the cluster detection method to the various databases, and to explore the main features of selected clusters.

2 Clustering of seismic catalogues

We consider two databases for the Italian seismicity: the historical catalogue CPTI15 (Rovida *et al.*, 2016) and the instrumental catalogue, hereinafter denoted as INGV. CPTI15 includes 4584 earthquakes of moment magnitude $M_w \geq 4$ occurred in the time period (1000-2014) in the entire Italian territory. The INGV database covers over 35 years, from 1981 to March 2016; it corresponds to the catalogue of Lolli & Gasperini (2006), updated from 2005 using the data from the Italian Seismological Instrumental and parametric Data-basE (http://iside.rm.ingv.it/iside/). The earthquake size is measured by local magnitude M_L, and it can be considered as complete at least from $M_L \geq 3$. Hence we compare results from a longer, but less complete database (CPTI15) with those from a shorter but richer database (INGV). In particular we consider the set of the earthquakes occurred in the central Apennines region, an area located between Lat (40°; 46°) and Long (10°; 15°). One of the most critical features of seismic catalogues is their incompleteness, that is, the difficulty of recording events of low magnitude (especially aftershocks) in historical age. To limit this problem we start our analysis from 1950 so that also the historical catalogue CPTI15 includes a significant portion of the secondary shocks that happen after a strong earthquake. Moreover, in order to equalize the magnitude thresholds used for the two catalogues, we convert $M_w = 4$ into M_L, using available relations, and we get $M_L = 3.7$ up to 2004 and $M_L = 3.9$ from 2005.

As a first step we compute the nearest-neighbour distance η^* between each pair of events of the two databases, choosing the values given in Nekrasova *et al.* (2011) for the parameters: $b = 1.0$, $d_f = 1.4$. Figure 1 shows the empirical distribution of η^*, in log_{10}-scale, for CPTI15 and INGV catalogues. In both cases the distribution is bimodal: the left mode, corresponding to shorter distances (strong links), represents the clustered seismic activity, whereas the right mode refers to the background seismicity. As expected CPTI15 catalogue is less complete than INGV one, particularly the clustered part; therefore the background part is dominant. The crossing point of the two estimated Gaussian densities is approximately the same in the two data sets: in particular we have $\log_{10}\eta_0 = -4.10$ for CPTI15 and $\log_{10}\eta_0 = -4.03$ (up to 2004) and $\log_{10}\eta_0 = -4.23$ (from 2005) for INGV catalogue.

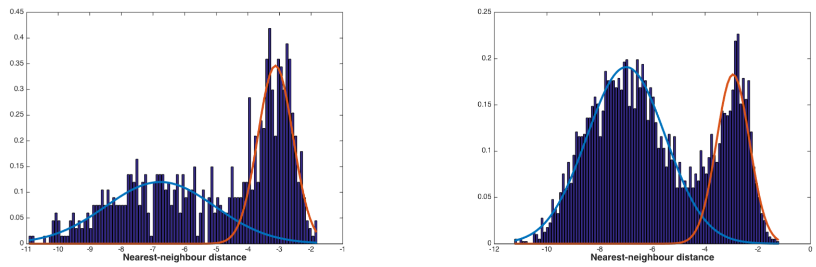

Figure 1. *Histogram of the nearest-neighbour distance* $\log_{10}\eta^*$ *and estimated Gaussian densities for events of* $M_w \geq 4$ *drawn from CPTI15 (left) and of* $M_L \geq 3$ *drawn from INGV (right) catalogues respectively.*

Once fixed the threshold η_0, the algorithm identifies several clusters; the most prominent ones, reported in both data sets, correspond to the 1997/09/26 Colfiorito earthquake (M_w 5.97), the 2009/04/06 L'Aquila earthquake (M_w 6.29), and the 2012/05/20 (M_w 6.09) Emilian plain earthquake.

Figure 2 shows the clusters associated with the Emilian plain earthquake obtained from the two databases: the figures represent the time vs distance distribution of the events, with links between *fathers* and their respective *offspring*. Despite some differences, mostly related to magnitude and location

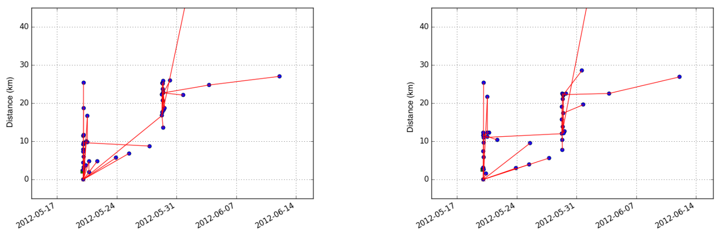

Figure 2. *Emilian plain 2012 cluster - Distance vs. time distribution of the clustered events drawn from CPTI15 (left) and from INGV (right): blue circles correspond to aftershocks, green squares to foreshocks, red lines to parent links.*

differences, the two clusters have a substantially similar structure; analogous considerations apply to the sequences related to L'Aquila and Colfiorito earthquakes. This allows us to conclude that the method is quite robust with respect to the database used, which permits expanding the analysis to a broader time interval, using CPTI15, or to a larger magnitude range, using INGV. Thus we deepen the analysis by lowering the threshold magnitude to $M_L \geq 3$ and examining events of the INGV catalogue.

Figure 3 shows the clusters obtained for Colfiorito 1997 and L'Aquila 2009 earthquakes: top pictures show the distance vs time distributions with the links between *fathers* and *offspring*, whereas bottom pictures show the time vs magnitude distributions of the two sequences. The Colfiorito cluster has 268 shocks and the tree branches out into 14 levels with average leaf depth $< d >= 4.64$; in contrast, the cluster related to L'Aquila earthquake has 281 shocks, spreaded out into 9 levels with average leaf depth $< d >= 2.05$. Comparing the two clusters we can observe that the Colfiorito sequence is more complex in space and develops in multiple generationss, whereas the L'Aquila sequence is more concentrated and is mostly composed of the first-generation aftershocks.

In conclusion we note that in all clusters we have analysed by the nearest-neighbour method:

• there is a space migration of the secondary shocks (offspring), which move away from the epicentre of the main shock; this is highlighted by the vertical development of the tree structure;

• after a period of decreasing seismicity, a reactivation is observed, that is a second strong event with its own tree structure.

These appear to be common features for major earthquakes in the study region, as evidenced by the ongoing complex Amatrice-Norcia sequence started in August 2016.

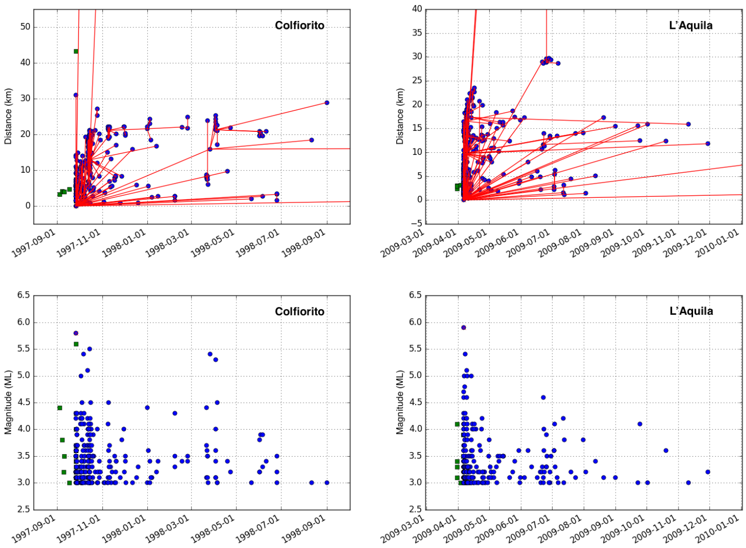

Figure 3. *(top panel) Tree clusters associated with Colfiorito 1997 earthquake (left) and L'Aquila 2009 earthquake (right). (bottom panel) Magnitude vs. time representation of the same sequences. Symbols are as in Figure 2*

References

BAIESI, M., & PACZUSKI, M. 2004. Scale-free networks of earthquakes and aftershocks. *Physical Review E*, **69**, 066106.

BAK, P., CHRISTENSEN, K., DANON, L., & SCANLON, T. 2002. Unified scaling law for earthquakes. *Physical Review Letters*, **88**(17), 178501.

LOLLI, B., & GASPERINI, P. 2006. Comparing different models of aftershocks rate decay: The role of catalog incompleteness in the first times after main shock. *Tectonophysics*, **423**, 43–59.

NEKRASOVA, A., KOSSOBOKOV, V., PERESAN, A., AOUDIA, A., & PANZA, G.F. 2011. A multiscale application of the unified scaling law for earthquakes in the central Mediterranean area and Alpine region. *Pure Appl. Geophys.*, **168**, 297–327.

ROVIDA, A., LOCATI, M., CAMASSI, R., LOLLI, B., & GASPERINI, P. (EDS). 2016. *CPTI15*. Istituto Nazionale di Geofisica e Vulcanologia. doi:http://doi.org/10.6092/INGV.IT-CPTI15.

ZALIAPIN, I., & BEN-ZION, Y. 2013. Erthquake clusters in southern California II: Classification and relation to physical properties of the crust. *J. Geophys. Res.: Solid Earth*, **118**, 2865–2877. doi:10.1002/jgrb.50178.