Earthquakes clustering based on maximum likelihood estimation of point process conditional intensity function

Giada Adelfio∗, Marcello Chiodi†, Luciana De Luca†, Dario Luzio†
adelfio@dssm.unipa.it, chiodi@unipa.it, deluca@unipa.it, luzio@unipa.it

∗Dipartimento di Scienze Statistiche e Matematiche “Silvio Vianelli” (DSSM)
†Dipartimento di Chimica e Fisica della Terra (CFTA)
1. Abstract

In this paper we propose a clustering technique to separate and find out the two main components of seismicity: the background seismicity and the triggered one. The method here proposed assigns each earthquake to a cluster or to the set of background events according to the intensity function. Parameters are estimated by maximum likelihood methods, iteratively changing the assignment of the events. This technique develops nonparametric estimation methods and computational procedures for the maximization of the point process likelihood function.
2. Main features of the method

- Many of the underlying assumptions of the used model, come from the Ogata’s ETAS model (Epidemic Type Aftershocks-Sequences model; [Ogata(1988)], [Ogata(1998)]).
- The intensity function of the corresponding point process (section 5) is composed by two parts: one relative to the background seismicity and one to the induced seismicity depending from $k$ clusters.

  - **Background seismicity**
    * The background seismicity is stationary in time.
    * The background seismicity is estimated smoothing the main events, by a nonparametric kernel method.

  - **Induced seismicity**
    * The induced seismicity in each cluster is described by a spatial term estimated only through the points belonging to that cluster, by a nonparametric kernel method.
The induced seismicity in each cluster is also described by a time dependent term, modelled by the modified Omori law ([Utsu(1961)]).

- Different parameters in the model account for different aspects of the seismicity: e.g. the Omori law parameters or the parameters that explain the different relative weights of the clustered and the background seismicity.

- Parameters are estimated maximizing the likelihood of the process.

- The proposed method of **clustering (or de-clustering)** alternates movements of points and estimates of parameters until convergence is achieved.
3. **Type of events**

- We suppose to have a catalog of \( n \) events, with a fixed value \( m_0 \) of the magnitude completeness threshold.

- \( \mathcal{P}_{k+1} \) is a partition of the whole set of the events in \( k + 1 \) sets: one relative to the background seismicity and \( k \) relative to clusters. We then distinguish three types of events:
  1. \( n_0 \) isolated points
  2. \( k \) mainshocks
  3. \( n_j \) points belonging to the \( j \)th cluster \((j = 1, 2, \ldots, k)\).

  \[ \Rightarrow \text{ Of course: } \sum_{j=1}^{k} n_j + n_0 = n. \]
4. **Aim of the method**

- Find a Partition $\mathcal{P}_{k+1}$
- Estimate the vector of parameters $\theta$ of the intensity function.

1. For a given partition $\mathcal{P}_{k+1}$, we find the estimate $\hat{\theta}$ which maximizes the likelihood function;
2. For a value $\hat{\theta}$, we try to change the partition $\mathcal{P}_{k+1}$, moving single points, in order to increase the likelihood function;
3. Steps 1 and 2 are alternated until a convergence criterion is satisfied.
5. The Intensity Function

The starting point is the definition of the conditional intensity function of the considered point process, for a generic space-time point $x, y, t$:

$$
\lambda(x, y, t) = \lambda_t \mu(x, y) + \gamma \sum_{j=1}^{k} g_j(x, y) \frac{\exp[\alpha(m_j - m_0)]}{(t - t_{0j} + c)^p}
$$

where:

- $\mu(x, y)$ is the background space intensity function;
- $\lambda_t$ is the weight of the background seismicity;
- $g_j(x, y)$ is the space intensity function of the cluster $j$;
- $\gamma$ is the weight of the clustered seismicity;
- $t_{0j}$ is the time of the first event of the cluster $j$;
- $m_j$ is the magnitude of the mainshock of the cluster $j$;
- $m_0$ is the threshold magnitude;
- $\alpha$ is a parameter measuring the influence of the mainshock magnitude on the relative weight of each sequence (or cluster);
- $c, p$ are the parameters of the modified Omori law (we also consider the possibility to obtain different estimates for each cluster).
6. **Estimation of the spatial components of the seismicity**

- Both the background and the clustered spatial seismicity are estimated by a kernel method.
  - The background spatial seismicity \( \mu(x, y) \) is computed using the \( n_0 \) isolated points and the \( k \) mainshocks.
  - The spatial seismicity for each cluster \( g_j(x, y) \) \((j = 1, 2, \ldots, k)\), is computed using the \( n_j \) points belonging to the \( j \)th cluster.

The *kernel* estimator is given by:

\[
\hat{f}(x, y) = \frac{1}{n h_x h_y} \sum_{i=1}^{n} K \left[ \frac{(x - X_i)(y - Y_i)}{h_x h_y} \right]
\]

(2)

where \( K(x, y; X_i, Y_i, h_x, h_y) \) is the generic kernel function and \( \mathbf{h} = (h_x, h_y) \) is the vector of the smoothing constants.

The kernel estimator can be obtained as the sum of \( n \) surfaces corresponding to the single observations: the kernel function defines the shape of these surfaces, whereas the smoothing constants (estimated by the formula \( h_{opt} = 1.06 A n^{-1/5} \) with \( A = \min\{\text{standard deviation}, \text{range-interquartile}/1.34\} \), ([Silverman(1986)])) measure their dispersion.
7. The Likelihood Function

The log likelihood for the space-time point process, given the observed values \((x_i, y_i, t_i)\) of the space-time coordinates of the \(i\)th event \((i = 1, 2, \ldots, n)\), is:

\[
\log L = \sum_{i=1}^{n} \log \lambda(x_i, y_i, t_i) - \int_{T_0}^{T_{\text{max}}} \int_{\Omega_{xy}} \lambda(x, y, t) \, dx \, dy \, dt
\]

Using the intensity function in (1), and given a partition \(\mathcal{P}_{k+1}\), the log likelihood is given by:

\[
\log L(\theta; x, y, t, \mathcal{P}_{k+1}) = \\
= \sum_{i=1}^{n} \log \left\{ \lambda_t \mu(x_i, y_i) + \sum_{j=1}^{k} \frac{g_j(x_i, y_i) \exp \left[ \alpha (m_j - m_0) \right]}{(t_i - t_{0j} + c)^p} \right\} - \\
- (T_{\text{max}} - T_0) \int_{\Omega_{xy}} \lambda_t \mu(x, y) \, dx \, dy - \\
- \sum_{j=1}^{k} \left\{ (T_{\text{max}} - t_{0j} + c)^{1-p} - c^{1-p} \right\} \exp \left[ \alpha (m_j - m_0) \right] \int_{\Omega_{xy}} g_j(x, y) \, dx \, dy
\]

where:
\(\theta\) is the vector of parameters: \((\lambda_t, \gamma, c, p, \alpha)\);

\((T_0 - T_{\text{max}})\) is the observed period of time.
8. Finding a candidate cluster

For each unit $U_h$, $(h = 1, \ldots, n)$ either an isolated or a clustered point, we firstly find the cluster $r_h$ which maximizes the conditional intensity; approximately this is obtained comparing the $k$ contributions in the sum:

$$\sum_{j=1}^{k} g_j(x_h, y_h) \frac{\exp [\alpha(m_j - m_0)]}{(t_h - t_{0j} + c)^p}$$

and assigning each unit $U_h$ to the cluster $r$ that maximizes:

$$I(t_h > t_{0r}) g_r(x_h, y_h) \frac{\exp [\alpha(m_r - m_0)]}{(t_h - t_{0r} + c)^p}$$

$r = 1, 2, \ldots, k$ (4)
9. Likelihood changes due to changes of partition

- When the partition changes from $\mathcal{P}_{k+1}$ to $\mathcal{P}_{k+1}^*$ ($\Delta(\mathcal{P}_{k+1})$) we examine the change in $\log L(\theta; x, y, t, \mathcal{P}_{k+1})$ ($\Delta_P \{\log L(\theta; x, y, t, \mathcal{P}_{k+1})\}$).

- Schematically, kinds of change in $\Delta(\mathcal{P}_{k+1})$, due to the movement of a single unit, are:
  - type A $\Delta_{(h,0,r)} \Rightarrow$ unit $U_h$ moves from background seismicity to cluster $r$;
  - type B $\Delta_{(h,r,0)} \Rightarrow$ unit $U_h$ moves from cluster $r$ to the set of background seismicity;
  - type C $\Delta_{(h,r,s)} \Rightarrow$ unit $U_h$ moves from cluster $r$ to cluster $s$.

- We must compute $\Delta_P \{\log L(\theta; x, y, t, \mathcal{P}_{k+1})\}$ for each kind of moving (A,B and C).
  - Eventually, we could use a first order approximation not reported here for brevity.

- The movement of the generic unit $U_h$ can also cause a change in the mainshock of the clusters involved in the movement.

- The $\Delta_P \{\log L(\theta; x, y, t, \mathcal{P}_{k+1})\}$ values are computed for each possible change in partition, induced by the movement of a unit $U_h$, $h = 1, 2, \ldots, n$.

- In this computation $\hat{\theta}$ is supposed to remain unchanged.
10. The algorithm of clustering

The technique of clustering that we propose leads to an intensive computational procedure, implemented by software R ([R Development Core Team(2005)])

1. Starting steps:
   - We start from a partition found by a window-based method (similar to a single-linkage procedure).
     - Briefly, the starting method puts events in a cluster if in it there is at least an event inside a window of $\delta_s$ units of space and $\delta_t$ units of time.
     - $\delta_s, \delta_t$ are given as input.
   - Clusters with a minimum fixed number of elements are found out: the number $k$ of clusters is determined.
   - The first partition $\mathcal{P}_{k+1}$ is then completed with the set of isolated points, constituted by the $n_0$ points not belonging to clusters.

2. Estimation of the space seismicity (see section 6).

3. Maximum Likelihood Estimation of parameters:
   - In (1) it is possible to assume either common Omori law parameters $c$ and $p$ over all cluster or variable $c_j$ and $p_j$ in each cluster (this could depend on the available catalog): as default, we consider the second type parametrization.
   - An iterative simplex method is used for the maximization of the likelihood (3).
   - $\hat{\theta}$ is the value of the MLE.
4. Finding a better partition $\mathcal{P}_{k+1}$:

- For each unit $U_h$, either an isolated or a clustered point, the best candidate cluster $r_h$ is found, according to the rule in (4).
- Different kinds of movements are tried (type A, B or C, according to the definition given in section 9).
- Points are assigned to the best set of events (best in the sense of Likelihood).
- Points are moved from clusters to background (and vice versa) if their movement increases the current value of the likelihood (3).
  - In these last steps, the likelihood (3), is computed using the current value of $\hat{\theta}$.

5. If no point is moved the algorithm stops, otherwise $\mathcal{P}_{k+1}$ is updated and the algorithm come back to step 2.

6. On the basis of the final partition and the final values of the estimates, the vector of estimated intensities for each point is computed.
11. Advantages of the procedure

- The procedure here proposed returns a plausible separation of the components of seismicity and clusters that can be reasonably interpreted.
- The space pattern of the induced seismicity does not depend on fixed parametric model but is estimated by non parametric methods, using only the points of each cluster.
- Different kind of parametrization allows to take into account for different assumptions about the seismicity of an area (e.g. Omori law parameters).
- Estimates of the likelihood and of the intensity function for each point are returned.
12.  **Critical aspects of the procedure**

The optimization steps can be improved in the future:

- Computational burden of the algorithm.
- Very high increments of the likelihood are achieved immediately after the first iteration.
- A certain dependence of the convergence of the iterative algorithm on some initial choices.
- The number of clusters is determined in the first step, and it can only decrease in the steps of the algorithm.
- Only single movements of units are checked.
13. Conclusion

This method could be the basis to carry out an analysis of the complexity of the seismogenetic processes relative to each sequence and to the background seismicity, separately. Indeed parameters that control the way in which strain energy is released could be strongly different through clusters and isolated events; this could be also observed in big differences in the parameter estimates of a phenomenological model applied to sets of earthquakes relative to the two different processes.
14. Example of application and results

Structural map of the central Mediterranean Sea
Contour plot of the estimated space density, more numerous clusters (more than 10 earthquakes) and isolated events, found through the proposed clustering method applied to the sicilian catalog (1988-2002)
Space-time plot of clusters and isolated events of the sicilian catalog (1988-2002)
Results relative to the estimation of the model with $c_j$ and $p_j$ varying in each cluster and the model with common $c$ and $p$ for all clusters.

<table>
<thead>
<tr>
<th></th>
<th>$p$</th>
<th>$c$</th>
<th>$\alpha$</th>
<th>$\gamma$</th>
<th>$AIC^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model with variable Omori law parameters</td>
<td>0.7973</td>
<td>0.9548</td>
<td>0.9556</td>
<td>0.6892</td>
<td>0.0153</td>
</tr>
<tr>
<td>Model with common Omori law parameters</td>
<td>0.8854</td>
<td>0.0</td>
<td>0.1211</td>
<td>3.6253</td>
<td>0</td>
</tr>
</tbody>
</table>

* these values are not completely comparable because the partition may change in each iteration and so the number of clusters.
Riferimenti bibliografici


