Stoch Environ Res Risk Assess
DOI 10.1007/s00477-014-0873-8

Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs

Giada Adelfio · Marcello Chiodi

Received: 20 November 2013 / Accepted: 25 March 2014

Abstract An estimation approach for the semi-parametric intensity function of a class of space-time point processes is introduced. In particular we want to account for the estimation of parametric and nonparametric components simultaneously, applying a forward predictive likelihood to semi-parametric models. For each event, the probability of being a background event or an offspring is therefore estimated.

Keywords Nonparametric estimation · Forward predictive likelihood · ETAS model · Point process · Earthquakes

1 Introduction

To describe and interpret the features of realizations of space-time point processes (e.g. seismic data, fires data, diseases data) a reliable estimation of the conditional intensity function is necessary. In exploratory contexts or to assess the adequacy of a specific parametric model, some kind of nonparametric estimation procedure could be useful, though in some fields (e.g. seismological one) predictive properties of the estimated intensity function are pursued.

In particular in such processes where the reproduction or some epidemic activity can be modelled, prediction of the basic reproductive rate is often complicated by the presence of triggered events, superimposed to the persistent background component. For instance in the seismological process, earthquake clusters, formed by the main event of each sequence, its foreshocks and its aftershocks, may complicate the statistical analysis of the background seismic activity that might be related to changes in the tectonic field. Since the persistent background activity prevails, in large time scale, over the aftershock activity, location of large earthquakes may be forecasted starting from the analysis of the background seismicity, for which removal of temporal cluster members may be a crucial issue.

Indeed if we want to predict large earthquakes in presence of clusters of aftershocks, earthquake clusters may complicate the statistical analysis of the background seismic activity. Because of the different seismogenic features controlling the kind of seismic release of clustered and background seismicity (Adelfio et al.2006), to describe the seismicity of an area in space, time and magnitude domains, it could be useful to study separately the features of independent events and those of the strongly correlated ones.

Zhuang et al. (2002) proposed a stochastic method associating to each event a probability to be either a background event or an offspring generated by other events, based on the ETAS model (Epidemic Type Aftershocks-Sequences model; Ogata 1988) for clustering patterns: a random assignment of events generates a thinned catalog, where events with a higher probability of being mainshocks are more likely included, and a inhomogeneous Poisson process is used to model their spatial intensity. This procedure identifies the two complementary subprocess of the seismic process: the background subprocess and the cluster (or offspring) subprocess.

In previous papers (Adelfio 2010; Adelfio et al.2010) we proposed a technique to find out the two main components of seismicity, i.e. the background seismicity and the triggered one.

G. Adelfio (✉) · M. Chiodi
Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy
e-mail: giada.adelfio@unipa.it

Published online: 12 April 2014

Adelfio et al. (2010) presented a seismic sequences detection technique based on MLE of parameters, that identifies the conditional intensity function of a model describing the seismic activity as a clustering-process, like ETAS model. In Adelfio (2010) nonparametric methods are used to estimate the intensity function of a space-time point process and clustering results are interpreted by a second-order diagnostic approach (Adelfio and Schoenberg 2009; Adelfio and Chiodi 2009). Console et al. (2010) proposed a stochastic method associating to each event a probability to be either a background event or an offspring generated by other events; Marsan and Lenglin (2008) used the concept of cascade triggering without using models; Diaz-Avalos et al. (2013) used also a nonparametric approach to check the separability of a point process.

A probabilistic clustering approach, providing an uncertainty about an object's class membership, can be provided by latent clustering analysis (Fraley and Raftery 2002). This is a very flexible approach, in the sense that both simple and complicated distributional forms can be used for the observed variables within clusters, although restrictions can be imposed on the parameters to obtain more parsimony and formal tests can be used to check their validity.

The basic latent class cluster model is given by:

$$P(\mathbf{y}|\theta) = \sum_{j=1}^{S} \pi_j P_j(\mathbf{y}|\theta_j)$$

where $P(\cdot)$ is obtained as a mixture of classes-specific densities $P_j(\cdot)$, given the clusters parameters θ_j, $\mathbf{y}$ is the observed variables, S the number of clusters and π_j the prior probability of membership in cluster j.

In this paper, in an analogous way we want to classify events according to their probability of being a background or an offspring event, estimating the space-time intensity of the generating point process of the different components by mixing nonparametric and parametric approaches.

Therefore, we propose an estimation of the space-time intensity of a branching-type point process that is usually characterized by these different components, that accounts simultaneously for the estimation of parametric and non-parametric ones, applying a forward predictive likelihood estimation approach to semi-parametric models (Chiodi and Adelfio 2011).

In sect. 2 some formal definitions of point processes are recalled. A new method for nonparametric estimation is introduced in Sect. 3; the simultaneous approach for non-parametric and parametric estimation is proposed in Sect. 4 with an application to the ETAS model for earthquake description, while final remarks and future developments are presented in Sect. 5.

2 Intensity function in point processes and branching-type model

Point process is a random collection of points, each one representing the time and space coordinates of a single event.

Let $Z^d = T \times S^{d-1}$ be a general $d-$dimensional closed region, with T the time domain and S^{d-1} a two or three dimensional space. Any analytic space-time point process is uniquely characterized by its associated *conditional intensity function* (Daley and Vere-Jones 2003) defined as the frequency with which events are expected to occur around a particular location in time and space, conditional on the prior history $\mathcal{H}_t$ of the point process up to time t, i.e.:

$$\lambda(\mathbf{z}) = \lambda(t, \mathbf{s}|H_t) = \lim_{\Delta t, \Delta \mathbf{s} \to 0} \frac{\mathrm{E}[\mathrm{N}([\mathrm{t}, \mathrm{t} + \Delta \mathrm{t}] \times [\mathbf{s}, \mathbf{s} + \Delta \mathbf{s}])|\mathrm{H_t}]}{\Delta t \Delta \mathbf{s}} \tag{1}$$

where H_t is the space-time occurrence history of the process up to time t, $\Delta t, \Delta \mathbf{s}$ are time and space increments, $\mathrm{E}[\mathrm{N}([\mathrm{t}, \mathrm{t} + \Delta \mathrm{t}] \times [\mathbf{s}, \mathbf{s} + \Delta \mathbf{s}])|\mathrm{H_t}]$ is the history-dependent expected number of events occurring in the volume $\{[t, t + \Delta t) \times [\mathbf{s}, \mathbf{s} + \Delta \mathbf{s}]\}$. Generally, intensities $\lambda(\mathbf{z})$ depend on some unknown parameters.

2.1 Branching point processes

In probability theory, a branching process is a Markov process in which each individual in the $n-th$ generation produces some random number of individuals in the $(n+1)-th$ generation, according to a probability distribution that does not vary from individual to individual. Branching processes are used to model reproduction phenomena. These models have been recently considered for the description of different applicative fields: biology (Caron-Lormier et al. 2006), demography (Jagers and Klebaner 2000; Johnson and Taylor 2008), epidemiology (Becker 1977; Balderama et al. 2012), wildfires distribution and size (Schoenberg et al. 2003; Juan et al. 2012).

In general, the conditional intensity function of the branching model is defined as the sum of a term describing the large-time scale variation (spontaneous activity or background) and one relative to the small-time scale variation due to the interaction with the events in the past (induced activity or offsprings):

$$\lambda_{\boldsymbol{\theta}}(t, \mathbf{s}|\mathcal{H}_t) = \mu f(\mathbf{s}) + \tau_{\boldsymbol{\phi}}(t, \mathbf{s}) \tag{2}$$

with $\boldsymbol{\theta} = (\boldsymbol{\phi}, \mu)'$, the vector of parameters of the induced intensity $(\boldsymbol{\phi})$ together with the parameter of the background general intensity (μ), $f(\mathbf{s})$ the space density, and $\tau_{\boldsymbol{\phi}}(t, \mathbf{s})$ the induced intensity, given by:

$$\tau_{\boldsymbol{\phi}}(t,\mathbf{s}) = \sum_{t_j<t} v_{\boldsymbol{\phi}}(t-t_j, \mathbf{s}-\mathbf{s}_j).$$

In such models, we have to simultaneously estimate the different components of the intensity function (large-time scale and small-time scale). If the large-time scale component $\mu f(\mathbf{s})$ in (2) is known, the parameters $\boldsymbol{\phi}$ can be usually estimated by Maximum Likelihood method. In applications, the large-time scale component $\mu f(\mathbf{s})$ is usually estimated trough non parametric techniques, like kernel estimators.

2.2 Kernel estimator for intensity function

Given n observed events $\mathbf{z}_1, \mathbf{z}_2 \ldots, \mathbf{z}_n$ in a d-dimensional closed region, the kernel estimator of the unknown intensity f (Silverman 1986; Wand and Jones 1994) in a generic point $\mathbf{z} \in \mathbb{R}^d$ is:

$$\hat{f}_{\boldsymbol{\Sigma}}(\mathbf{z}) = \sum_{i=1}^{n} K(\mathbf{z}-\mathbf{z}_i, \boldsymbol{\Sigma}) \tag{3}$$

where $K(\cdot,\cdot)$ is a multivariate kernel function centered at observed points and $\boldsymbol{\Sigma}$ is a matrix of smoothing constants. A common choice for $K(\cdot,\cdot)$ is the normal multivariate density; in this case, and if $\boldsymbol{\Sigma}$ is diagonal, the kernel function is defined by the superposition of separable kernel densities. To take into account highly variable patterns in a space region, variable smoothing matrices $\boldsymbol{\Sigma}_i$ can be more suitable (Terrell and Scott 1992).

In general, in kernel approaches, the smoothing parameters are set by external choices, or by cross-validation techniques. Indeed, the usual maximization of the likelihood with respect to the smoothing parameters, as known, would produce bandwidths of length zero and degenerate intensities only on the observed points. Therefore, for nonparametric estimation we propose the use of an estimation procedure based on the subsequent increments of likelihood obtained adding an observation one at a time, reported in the next section.

3 Forward predictive likelihood (FLP)

Suppose that in a space-time point process the intensity function $\lambda(\cdot)$ depends on a set of parameters $\boldsymbol{\psi}$, such that $\lambda(\mathbf{z}, \boldsymbol{\psi})$.

Let denote by $\hat{\boldsymbol{\psi}}(H_{t_k}) \equiv \hat{\boldsymbol{\psi}}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_i, \ldots, \mathbf{z}_k)$ a generic estimator of $\boldsymbol{\psi}$, based on observations until t_k.

Assume that a realization of the process is observed in the space region $\Omega_{\mathbf{s}}$ and the time interval $(T_0; T_{max})$. The log-Likelihood for the point process, given the k observed values $\mathbf{z}_i$ and computed using the estimator $\hat{\boldsymbol{\psi}}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_i, \ldots, \mathbf{z}_k)$ is:

$$\log L(\hat{\boldsymbol{\psi}}(H_{t_k}); H_{t_k}) = \sum_{i=1}^{k} \log \lambda(\mathbf{z}_i; \hat{\boldsymbol{\psi}}(H_{t_k})) - + \int_{T_0}^{T_{max}} \int_{\Omega_{\mathbf{s}}} \lambda(\mathbf{z}; \hat{\boldsymbol{\psi}}(H_{t_k}))\, d\mathbf{s}\, dt \tag{4}$$

As mentioned in Sect. 2.2, the ML estimation can not be directly used in a semi-parametric context: in fact, for example, considering the intensity (2), which contains a component that is usually estimated nonparametrically, the likelihood (4) would be maximized putting all the mass on the observed points.

In this paper, we use the method proposed in Chiodi and Adelfio (2011) that measures the ability of the observations and estimation until t_k to give information on the next observation.

Let $\hat{\boldsymbol{\psi}}(H_{t_k})$ be a vector of estimators, that could include smoothing constants in a semi-parametric context, based on the observed history up to t_k. Let $\log L(\hat{\boldsymbol{\psi}}(H_{t_k}); H_{t_{k+1}})$ be the likelihood computed on the first $k+1$ observations, but using the estimates based on first k, defined as:

$$\log L(\hat{\boldsymbol{\psi}}(H_{t_k}); H_{t_{k+1}}) = \sum_{i=1}^{k+1} \log \lambda(\mathbf{z}_i; \hat{\boldsymbol{\psi}}(H_{t_k})) - + \int_{T_0}^{t_{k+1}} \int_{\Omega_{\mathbf{s}}} \lambda(\mathbf{z}; \hat{\boldsymbol{\psi}}(H_{t_k}))\, d\mathbf{s}\, dt \tag{5}$$

For example, in equation (5), $\lambda(\mathbf{z}_{k+1}; \hat{\boldsymbol{\psi}}(H_{t_k}))$ could be the intensity of the $(k+1)th$ point estimated by a kernel method using the centers given by the previous k points.

Then, we use the difference between (4) and (5) to measure the *predictive information* of the first k observations on the $k+1$-th as:

$$\begin{aligned}
\delta_{k,k+1}(\hat{\boldsymbol{\psi}}(H_{t_k}); H_{t_{k+1}}) \equiv \\
&= \log L(\hat{\boldsymbol{\psi}}(H_{t_k}); H_{t_{k+1}}) \\
&\quad - \log L(\hat{\boldsymbol{\psi}}(H_{t_k}); H_{t_k}) \\
&= \sum_{i=1}^{k+1} \log \lambda(\mathbf{z}_i; \hat{\boldsymbol{\psi}}(H_{t_k})) \\
&\quad - \int_{T_0}^{t_{k+1}} \int_{\Omega_{\mathbf{s}}} \lambda(\mathbf{z}; \hat{\boldsymbol{\psi}}(H_{t_k})) d\mathbf{s} dt \\
&\quad - - \sum_{i=1}^{k} \log \lambda(\mathbf{z}_i; \hat{\boldsymbol{\psi}}(H_{t_k})) \\
&\quad - \int_{T_0}^{t_k} \int_{\Omega_{\mathbf{s}}} \lambda(\mathbf{z}; \hat{\boldsymbol{\psi}}(H_{t_k})) d\mathbf{s} dt \\
&= \\
&= \log \lambda(\mathbf{z}_{k+1}; \hat{\boldsymbol{\psi}}(H_{t_k})) \\
&\quad - \int_{t_k}^{t_{k+1}} \int_{\Omega_{\mathbf{s}}} \lambda(\mathbf{z}; \hat{\boldsymbol{\psi}}(H_{t_k})) d\mathbf{s} dt.
\end{aligned} \tag{6}$$

This leads to a technique similar to cross-validation, but applied only to the future observations: in fact, each contribution $\delta_{k,k+1}$ is based only on the past observations $t_1, \ldots, t_k$.

Therefore, given n the number of observations, we choose $\widetilde{\boldsymbol{\psi}}(H_{t_k})$ which maximizes:

$$FLP_{k_1,k_2}(\hat{\boldsymbol{\psi}}) \equiv \sum_{k=k_1}^{n-1} \delta_{k,k+1}, \quad (7)$$

where k_1 is a fixed constant, for example $k_1 = \left[\frac{n}{2}\right]$.

The quantity in (7) can be used also to compare different kinds of intensity estimates obtained by considering the optimized values of the quantities $FLP_{k_1,k_2}(\boldsymbol{\psi})$.

In this paper, we use the measure defined in (7) to estimate the nonparametric component of models like (2). In previous applications (Chiodi and Adelfio 2011), on the basis of the measure in (7), we observed that the bandwidths estimated by FLP approach produced better kernel estimates (in terms of MISE) of space-time intensity functions than classical methods.

The following theorem proves that a martingale can be obtained from the quantity in (6): this result can be useful to study its theoretical asymptotic distributional properties.

Theorem 1 Let N be a point process on $R^2 \times R$, such that $\mathbf{z} \in R^2 \times R$ and $\lambda(\mathbf{z}; \boldsymbol{\psi}(H_{t_m}))$ its conditional intensity function up to time t_m. Let us define

$$\delta_{m,m+1} = \log \lambda(\mathbf{z}_{m+1}; \boldsymbol{\psi}(H_{t_m})) - \\ + \int_{t_m}^{t_{m+1}} \int_{\Omega_S} \lambda(\mathbf{z}; \boldsymbol{\psi}(H_{t_m})) d\mathbf{s} d\mathbf{t}$$

as a measure of the predictive information on the first m observations on the $(m+1)-$th and

$$I_{m+1} = \exp\left[- \int_{t_m}^{t_{m+1}} \int_{\Omega_S} \lambda(\mathbf{z}; \boldsymbol{\psi}(H_{t_m})) d\mathbf{s} d\mathbf{t} \right]$$

and $i_{m+1} = \log I_{m+1}$. Hence $\frac{\exp[\delta_{m,m+1}]}{I_{m+1}} - i_{m+1}$ is a martingale process.

Proof

$$\begin{aligned} & E\left[\frac{\exp[\delta_{m,m+1}]}{I_{m+1}} - i_{m+1} \middle| H_{t_m}\right] \\ & = E\left[\frac{\lambda(\mathbf{z}_{m+1}; \psi(H_{t_m}))}{I_{m+1}} I_{m+1} - i_{m+1} \middle| H_{t_m}\right] \\ & \approx E\left[E\left[N(\mathbf{z}_{m+1}, \mathbf{z}_{m+1} + \Delta\mathbf{z} \middle| H_{t_{m+1}})\right] - i_{m+1} \middle| H_{t_m}\right] \\ & = E\left[E[N(\mathbf{z}_{m+1}, \mathbf{z}_{m+1} + \Delta\mathbf{z} | H_{t_m})] - i_{m+1} \middle| H_{t_{m+1}}\right] \\ & = E\left[(\lambda(\mathbf{z}_m; \boldsymbol{\psi}(H_{t_m})) + i_{m+1}) - i_{m+1} \middle| H_{t_{m+1}}\right] \\ & = \lambda(\mathbf{z}_m; \boldsymbol{\psi}(H_{t_m})) = \frac{\exp[\delta_{m-1,m}]}{I_m} - i_m \end{aligned}$$

□

4 Alternating estimation of components

In order to estimate the different components of a space-time branching model (2), we here propose a simultaneous estimation of nonparametric and parametric components of a branching-type model. In other words, we alternate the standard parametric likelihood method, to estimate the parameters of the offsprings component, with the FLP approach, used just to compute the smoothing parameters $\boldsymbol{\Sigma}$ in (3) of the background intensity nonparametric estimation. Further, the proposed mixed procedure estimates the probability of each event to belong to one of the model components, given the class specific parameters, according to a latent cluster model with two possible groups.

Given a set of n events occurred in a fixed space-time region, and set $v = 1$, let $\hat{f}_{\boldsymbol{\Sigma}^{(0)}}(x, y)$ be a starting estimation of the background intensity, obtained by kernel estimators, with default values for the bandwidth $\boldsymbol{\Sigma}^{(0)}$. The $v-th$ iteration of the simultaneous estimation of the nonparametric and parametric components proceeds as follows:

1. Get the ML estimator $\hat{\boldsymbol{\theta}}^{(v)}$ of the parameters of the model $\boldsymbol{\theta} = (\boldsymbol{\phi}, \mu)'$, maximizing the whole likelihood (4) and compute the values $\lambda_{\hat{\boldsymbol{\theta}}^{(v)}}(t_i, x_i, y_i | \mathcal{H}_{t_i}) = \hat{\mu}^{(v)} \hat{f}_{\boldsymbol{\Sigma}^{(v-1)}}(x_i, y_i) + \tau_{\hat{\boldsymbol{\phi}}^{(v)}}(t_i, x_i, y_i)$, $i = 1, \ldots, n$.
2. Estimate $\rho_i^{(v)} = \frac{\hat{\mu}^{(v)} \hat{f}_{\boldsymbol{\Sigma}^{(v-1)}}(x_i, y_i)}{\lambda_{\hat{\boldsymbol{\theta}}^{(v)}}(t_i, x_i, y_i | \mathcal{H}_{t_i})}$, $i = 1, \ldots, n$, for each point of the data set, on the basis of the estimated parameters. $\rho_i^{(v)}$ is used as a vector of weights for the nonparametric estimation of the background intensity and is an estimation of the probability to belong to the background group.
3. Estimate an optimal smoothing parameter $\boldsymbol{\Sigma}^{(v)}$ of the kernel estimator, through the FLP approach, that is maximizing (7) and holding fixed $\tau_{\hat{\boldsymbol{\phi}}^{(v)}}(t_i, x_i, y_i)$, $i = 1, \ldots, n$.
4. Update the estimation of the background intensity $\hat{f}_{\boldsymbol{\Sigma}^{(v)}}(x_i, y_i)$, through weighted kernel estimator with weights $\rho_i^{(v)}$, $i = 1, \ldots, n$.
5. Update v and start a new iteration, until some convergence rule is reached. Convergence is judged comparing the values of model components in consecutive iterations, checking also the increase in the overall likelihood function.

4.1 ETAS model a particular branching point process

A branching process for earthquake description, widely used in seismological context, is the Epidemic Type Aftershocks-Sequences (ETAS) model (Ogata 1988).

The ETAS conditional intensity function can be written, starting from model (2), as follows:

$$\lambda_{\theta}(t, \mathbf{s}|\mathcal{H}_t) = \mu f(\mathbf{s}) + \sum_{t_j < t} v_{\phi}(t - t_j, \mathbf{s} - \mathbf{s}_j | m_j) \quad (8)$$

with m_j the magnitude of the j-th event and $v_{\phi}(t - t_j, \mathbf{s} - \mathbf{s}_j|m_j) = g(t - t_j|m_j)\ell(x - x_j, y - y_j|m_j)$. Therefore, in the ETAS model, the background seismicity is assumed to be stationary in time, while the occurrence rate of aftershocks at time t, following the earthquake of time t_j and magnitude m_j, is described by the following parametric model:

$$g(t - t_j|m_j) = \frac{\kappa \; e^{(\alpha-\gamma)\;(m_j - m_0)}}{(t - t_j + c)^p}, \quad \text{with } t > t_j \quad (9)$$

where κ is a normalizing constant, c and p characteristic parameters of the seismic activity of the given region; p is useful for characterizing the pattern of seismicity, indicating the decay rate of aftershocks in time.

For the spatial distribution, conditioned to magnitude of the generating event, the following distribution is often used:

$$\ell(x - x_j, y - y_j|m_j) = \left\{ \frac{(x - x_j)^2 + (y - y_j)^2}{e^{\gamma\;(m_j - m_0)}} + d \right\}^{-q} \quad (10)$$

It relates the occurrence rate of aftershocks to the mainshock magnitude m_j, through the parameters α, γ that measure the influence on the relative weight of each sequence; m_0 is the completeness threshold of magnitude, i.e. the lower bound for which earthquakes with higher values of magnitude are surely recorded in the catalog, d and q are two parameters related to the spatial influence of the mainshock.

The simultaneous estimation of the background intensity and the triggered intensity components of a Epidemic type model is a crucial statistical issue.

Although parametric models are widely used, their estimation has many disadvantages, often related to the definition of a reliable mathematical model from the geophysical theory (Choi and Hall 2001) and to the sensitivity of statistical estimates to the composition of the space-time region under study (Choi and Hall 2001).

A computationally efficient procedure to maximize the expected complete data log-likelihood function, based on the expectation-maximization algorithm is introduced in Veen and Schoenberg (2008).

While the first component $f(\cdot)$ of models like (8) is generally estimated by nonparametric techniques, $\boldsymbol{\theta}$ is usually estimated by ML approach. In particular, in kernel-type approaches either a fixed (Vere-Jones 1992) or adaptive kernel smoothing method with Gaussian kernel (Zhuang et al. 2002) can be used. However, while our approach has the big advantage of being only semi-parametric, Zhuang et al. (2002) proposed a purely parametric estimation method, also estimating the probability for each event of being a background event ($\rho_i, i = 1, \ldots, n$) in order to provide a random classification of events and obtain a thinned catalog, that includes events with a bigger probability of being mainshock, which spatial intensity is described by inhomogeneous Poisson process.

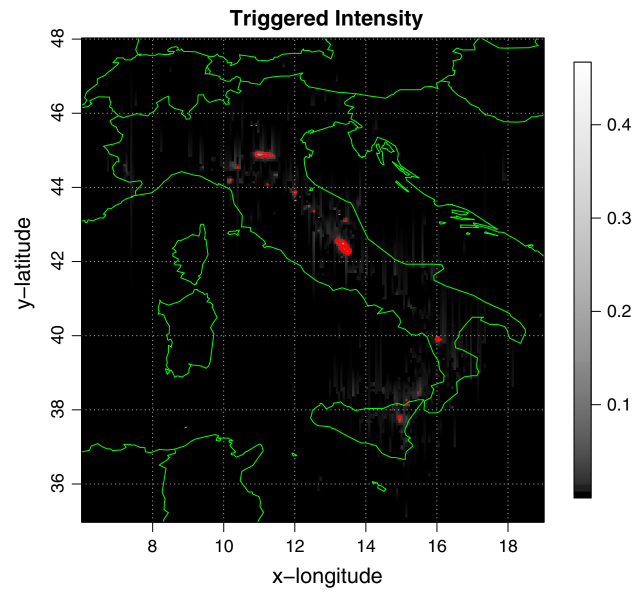

Fig. 1 Estimated triggered intensity of the Italian seismicity (2005–2013) with magnitude threshold 2 by FLP approach.

In our algorithm, according to Console et al. (2010), we use ρ_i as weights for the kernel estimation of the background seismicity to get a simultaneous estimate of the intensity components of the ETAS model (8).

As an example of application, we apply the proposed approach to the catalog of the Italian seismic events recorded from 2005 to 2013, with three different thresholds of magnitude (2, 2.2, 2.5) that identify 20894, 13748, 6886 events, respectively. The estimate of the triggered intensity function and the background component for threshold 2 using the FLP approach are reported in Figs. 1 and 2. The corresponding plots by using the ETAS estimates with fixed bandwidth for the background component are reported in Figs. 3 and 4. The FLP estimates of the background seismicity seem to be more realistic than the used fixed-bandwidth-ETAS model, as a consequence of a reduced smoothing effect, estimating an intensity function more coherent with some known tectonic structures. The plots corresponding to magnitude thresholds 2.2 and 2.5 are not reported for brevity and since they do not suggest different conclusions.

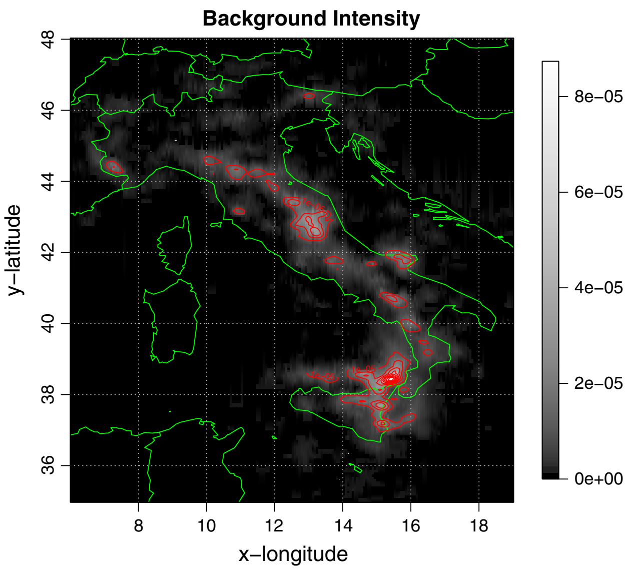

Fig. 2 Estimated background intensity of the Italian seismicity (2005–2013) with magnitude threshold 2 by FLP approach.

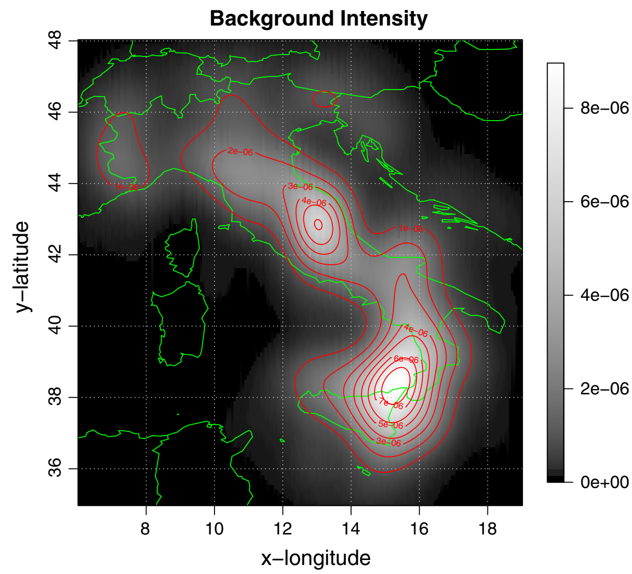

Fig. 4 Estimated background intensity of the Italian seismicity (2005–2013) with magnitude threshold 2 by fixed-bandwidth approach.

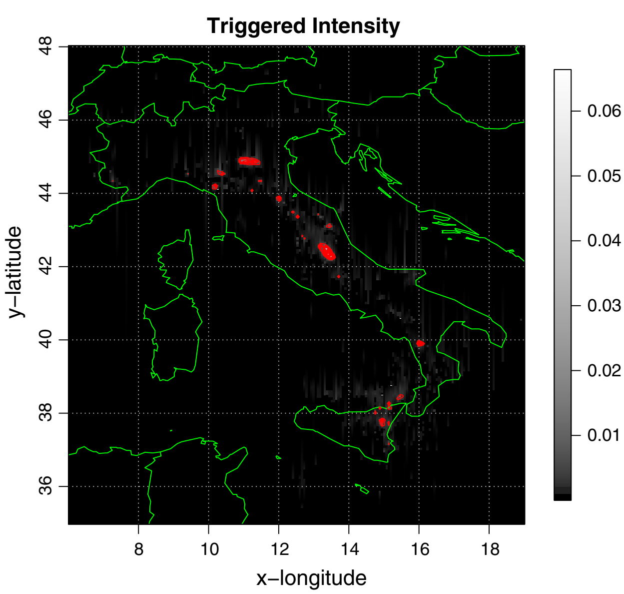

Fig. 3 Estimated triggered intensity of the Italian seismicity (2005–2013) with magnitude threshold 2 by fixed-bandwidth approach.

It is interesting to note the correspondence between the high peaks of the estimated background and triggered intensity functions for Italy and some focal areas of the Italian seismicity in the observed period, like L'Aquila and Reggio Emilia, where two big sequences of events occurred in 2009 and 2012, respectively, and Mt Etna Volcano, where a quite continuous activity is recorded.

The proposed approach seems to perform much better (both in terms of AIC, estimates and diagnostic results) than the usual fixed-bandwidth-ETAS estimates.

The comparison (based on the AIC values obtained at each iteration of the algorithm) between the two methods to estimate model (8) are reported in Figs. 5, 6 and 7, for each of the used magnitude thresholds. These results easily suggest the outperforming behavior of the FLP approach, independently of the magnitude threshold.

5 Remarks and future developments

The proposed simultaneous estimation of nonparametric and parametric components is a very flexible procedure, that accounts for predictive properties of the estimated intensity. Moreover, in a latent class model context, it estimates, for each event, the probability of being a background event or a triggered one in a branching-type model.

An interesting point of the considered estimation approach can be discussed analyzing the obtained results: indeed, the estimated model seems to follow properly the observed intensity of the observed area, characterized by highly variable changes both in space and in time. In other words, because of its flexibility, the estimation approach seems to provide a good fitting to local space-time changes, crucial to analyze possible correlation between the estimated intensity function and particular distributions of some structural features (i.e. geological structures) of the studied region.

In terms of performance, the proposed method produces AIC values considerably better than other approaches that choice the smoothing parameters according to different procedures.

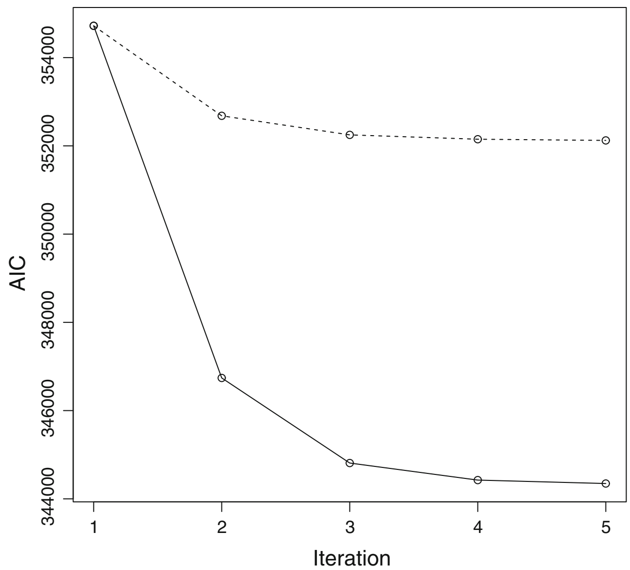

Fig. 5 Comparison between the AIC of the model (8) estimated by FLP (*solid line*) and fixed bandwidth (*dotted line*) for the Italian seismicity (2005–2013) with magnitude threshold 2.0.

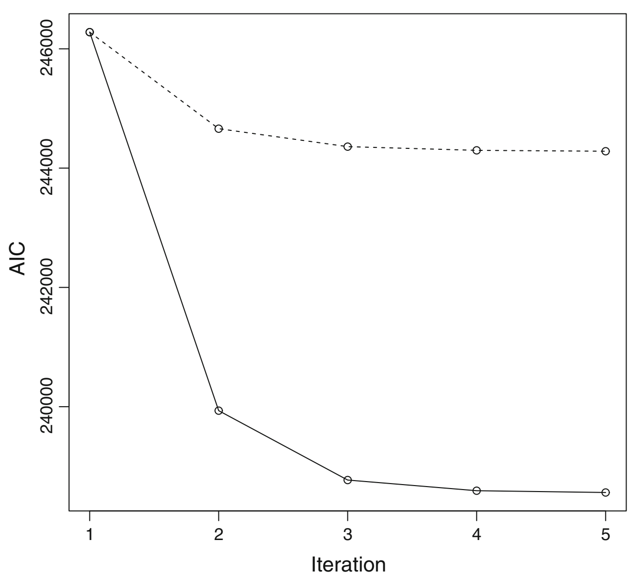

Fig. 6 Comparison between the AIC of the model (8) estimated by FLP (*solid line*) and fixed bandwidth (*dotted line*) for the Italian seismicity (2005-2013) with magnitude threshold 2.2.

For future work, we are developing an R package that provides tools for the comprehension and analysis of space-time data. In particular the package will allow to estimate both only time and space-time processes, making also possible the estimation of subset of parameters, together with computation of profile likelihood, diagnostics and graphical tools.

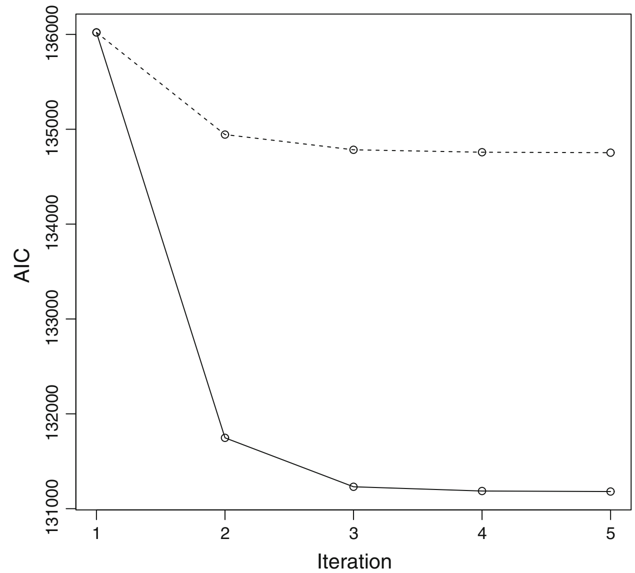

Fig. 7 Comparison between the AIC of the model (8) estimated by FLP (*solid line*) and fixed bandwidth (*dotted line*) for the Italian seismicity (2005–2013) with magnitude threshold 2.5.

Moreover, anisotropic kernel with variable smoothing parameters will be introduced to take into account more realistic situations, with very variable observed intensity in space and time.

Acknowledgments We would like to thank the Editor and the two referees for their fruitful and insightful comments. This paper has been partially supported by the Grant of University of Palermo (Italy): "2012-ATE-0332-FFR 2012-2013-Metodi statistici per dati spazio-temporali applicati all'analisi, monitoraggio e previsione ambientale" G. Lovison.

References

Adelfio G, Chiodi M, De Luca L, Luzio D, Vitale M (2006) Southern-Tyrrhenian seismicity in space-time-magnitude domain. Ann Geophys 49:1245–1257

Adelfio G, Chiodi M (2009) Second-order diagnostics for space-time point processes with application to seismic events. Environmetrics 20:895–911

Adelfio G, Schoenberg FP (2009) Point process diagnostics based on weighted second-order statistics and their asymptotic properties. Ann Inst Stat Math 61:929–948

Adelfio G (2010) An analysis of earthquakes clustering based on a second-order diagnostic approach. Data analysis and classification. Springer, Berlin, pp 309–317

Adelfio G, Chiodi M, Luzio D (2010) An algorithm for earthquake clustering based on maximum likelihood. Data analysis and classification. Springer, Berlin, pp 25–32

Adelfio G, Chiodi M (2011) Kernel intensity for space-time point processes with application to seismological problems. In: Fichet S (ed) Classification and multivariate analysis for complex data structures. Springer, Berlin, pp 401–408

Balderama E, Schoenberg FP, Murray E, Rundel PW (2012) Application of branching point process models to the study of invasive red banana plants in Costa Rica. JASA 107(498):467–476

Becker N (1977) Estimation for discrete time branching processes with application to epidemics. Biometrics 33(3):515–522

Caron-Lormier G, Masson JP, Menard N, Pierre JS (2006) A branching process, its application in biology: influence of demographic parameters on the social structure in mammal groups. J Theo Biol 238:564–574

Chiodi M, Adelfio G (2011) Forward Likelihood-based predictive approach for space-time processes. Environmetrics 22:749–757

Choi E, Hall P (1999) Nonparametric approach to analysis of space-time data on earthquake occurrences. J Comput Graph Stat 8(4):733–748

Choi E, Hall P (2001) Nonparametric analysis of earthquake point-process data. Lect Notes-Monograph Ser 36:324–344

Console R, Jackson DD, Kagan YY (2010) Using the ETAS model for catalog declustering and seismic background assessment. Pure Appl Geophys 167:819–830

Daley DJ, Vere-Jones D (2003) An introduction to the theory of point processes. Springer, New York

Diaz-Avalos C, Juan P, Mateu J (2013) Similarity measures of conditional intensity functions to test separability in multidimensional point processes. Stoch Environ Res Risk Assess 27(5):1193–1205

Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis and density estimation. J Am Stat Assoc 97:611–631

Gardner J, Knopoff L (1974) Is the sequence of earthquakes in southern California, with aftershock removed, poissonian? B Seimol Soc Am 64:1363–1367

Jagers P, Klebaner FC (2000) Population-size-dependent and age-dependent branching processes. Stoch Process Appl 87:235–254

Johnson RA, Taylor JR (2008) Preservation of some life length classes for age distributions associated with age-dependent branching processes.Stat Probabil Lett 78:2981–2987

Juan P, Mateu J, Saez M (2012) Pinpointing spatio-temporal interactions in wildfire patterns. Stoch Environ Res Risk Assess 26(8):1131–1150

Marsan D, Lenglin O (2008) Extending earthquakes' reach through cascading. Science 319:1076–1079

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Association 83:9–27

Resenberg P (1985) Second-order moment of central california seismicity, 1969–1982. J Geophys Res 90(B7):5479–5495

Schoenberg FP, Peng R, Woods J (2003) On the distribution of wildfire sizes. Environmetrics 14:583–592

Silverman B (1986) Density estimation for statistics and data analysis. Chapman and Hall, London

Terrell GR, Scott DW (1992) Variable Kernel density estimation. Ann Statist 20(3):1236–1265

Veen A, Schoenberg FP (2008) Estimation of space-time branching process models in seismology using an EM-type algorithm. JASA 103(482):614–624

Vere-Jones D (1992) Statistical methods for the description and display of earthquake catalogues. In: Walden A and Guttorp P (eds) Statistics in the environmental and earth sciences, Edward Arnold, London, pp. 220–236

Wand MP, Jones MC (1994) Multivariate plugin bandwidth selection. Comput Stat 9:97–116

Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97:369–379