Nonparametric analysis of the time structure of seismicity in a geographic region

Graciela Estévez-Pérez ([1]), **Henrique Lorenzo-Cimadevila** ([2]) **and Alejandro Quintela-del-Río** ([3])

([1]) *Departamento de Matemáticas, Facultad de Ciencias, Universidad de A Coruña, Spain*
([2]) *Departamento de Ingeniería de los recursos naturales y medio ambiente, Escuela de Ingeniería Técnica Industrial, Universidad de Vigo, Spain*
([3]) *Departamento de Matemáticas, Facultad de Informática, Universidad de A Coruña, Spain*

Abstract
As an alternative to traditional parametric approaches, we suggest nonparametric methods for analyzing temporal data on earthquake occurrences. In particular, the kernel method for estimating the hazard function and the intensity function are presented. One novelty of our approaches is that we take into account the possible dependence of the data to estimate the distribution of time intervals between earthquakes, which has not been considered in most statistics studies on seismicity. Kernel estimation of hazard function has been used to study the occurrence process of cluster centers (main shocks). Kernel intensity estimation, on the other hand, has helped to describe the occurrence process of cluster members (aftershocks). Similar studies in two geographic areas of Spain (Granada and Galicia) have been carried out to illustrate the estimation methods suggested.

Key words *nonparametric estimation – hazard function – intensity function – clustering – dependent data*

1. Introduction

The problem of searching for stochastic models to describe the sequence of occurrence times of earthquakes from some geographic region is of great interest to seismologists. In effect, a detailed analysis of such process might reveal new aspects of the pattern of occurrence of earthquakes, and suggest important ideas on the mechanism of earthquakes.

The development of detailed stochastic models to describe the list of origin times or equivalently that of time intervals between consecutive earthquakes is quite recent. Vere-Jones (1970) surveys some of the stochastic models (clustering models and stochastic models for aftershock sequences) proposed in the literature and describes their behavior in several data sets. More specifically, Udias and Rice (1975) propose the gamma distribution to describe the series of time intervals between consecutive shocks. These authors also deal with hazard and intensity functions. Other more recent models include the trigger models (Lomnitz and Nava, 1983), the Epidemic-Type Aftershock Sequence (ETAS) model (Ogata, 1988), whose extensions can be seen in Ogata (1998), or refinements of Hawkes' (1971) self-exciting point process model, which describes spatial-temporal patterns in a catalog.

However, standard models applied to seismic data do not always fit the data well. In part, this is because parametric models are usually only well suited for a sequence of seismic events that have similar causes. Moreover, parametric models can be insensitive to poorly-fitting events, which often are at least as interesting as well-fitting ones (see Ogata, 1989).

In this article, we suggest nonparametric methods for analyzing seismic data. These methods do not require formulation of structural models, so they are not affected by the deficiencies noted above. They involve several different approaches to nonparametric estimation of the hazard and intensity functions of point processes that evolve with time. This enables us to split up and analyze the occurrence of temporal processes of earthquakes within a region without constraining them to having predetermined properties. We argue that nonparametric methods for the analysis of earthquake data are valuable supplements to more conventional parametric approaches, especially as tools for exploratory data analysis.

The objective of our analysis is to show two statistical tools (hazard and intensity functions) which could help to describe the whole cycle of seismic activity in a region without imposing predetermined conditions on this activity. That is, our analysis is based on the information provided by the data and on the universally accepted assumption of temporal grouping of earthquakes. The hazard function is used to confirm this grouping and characterize the occurrence process of main shocks. On the other hand, the aftershock sequences (clusters) have been studied by means of the intensity function.

The paper is organized as follows: in Section 2 we describe the occurrence process of earthquakes in terms of its evolution in time. Section 3 introduces the nonparametric estimator of hazard function and Section 4 contains the analysis of seismic activity of two Spanish geographic regions using the nonparametric methods beforehand mentioned.

2. The occurrence process of earthquakes

Earthquakes can be represented by point events in a five-dimensional space-time-energy continuum (φ_i, λ_i, h_i, t_i, M_i) where φ_i and λ_i are the latitude and longitude of the epicenter, h_i the depth of the focus, t_i the origin time and M_i the magnitude. A complete statistical analysis of earthquakes must consider the distributions and correlations of these five parameters. This would involve handling a five-dimensional series, which generally constitutes a very complex problem. The starting point of such problem is the consideration of the one-dimensional series of occurrence times $\{t_i\}$. To give a precise meaning to this time series, its space, time and magnitude boundaries must be specified. Obviously, these boundaries will be chosen according to the objectives of the study: to characterize different seismic areas in the same time period, to analyze several seismic series in a particular region, etc.

Space specifications define the volume from which the population of earthquakes, represented by the time series $\{t_i\}$ is taken. This may be done in practice by specifying an area A and a lower limit in depth H. Since detectability is limited, a lower limit of magnitude M_0 must also be specified. This limit is a function of the station distribution and sensitivity, and defines the lowest magnitude for which all events from anywhere in the bounded region can be detected. A bounded set of earthquakes will then be a series $\{t_i\}$ defined for a certain area A, depth H and lower limit of magnitude M_0, such that, it would include all shocks originating from inside the defined volume of magnitude larger than M_0 for a particular time interval from t_0 to t_n.

Once a bounded set of time occurrence $\{t_i\}$ is established, a series can be constructed with the time intervals between consecutive earthquakes $\{\Delta t_i\}$, such that $\Delta t_i = t_i - t_{i-1}$ for $i = 1, ..., n$. The distribution of the values of these intervals is of great interest to specify the time structure of the seismicity of a region.

The simplest statistical model to fit a series of occurrence times of earthquakes is the Poisson process, under which the time intervals between consecutive events are exponentially distributed. This model presupposes independence of the events, so that the occurrence of one earthquake is not influenced by that of previous ones, which is very far from reality. In effect, several authors (Knopoff, 1964; Lomnitz,

1966; Vere-Jones, 1970; Udias and Rice, 1975) have found, for different populations of earthquakes, that the Poisson fit to the time series of microearthquakes is very poor, especially for active periods. The deviation from the model is principally due to the existence of a much larger number of small intervals than expected. The reason for such a large number of small intervals is that the earthquakes happen forming clusters, that is, one main shock is followed and/or preceded by a stream of smaller shocks, called aftershocks and/or precursors (Lomnitz and Hax, 1967; Vere-Jones and Davies, 1966) produced in the same general focal region.

Therefore, some authors (*e.g.*, Vere-Jones, 1970; Hawkes, 1971 and many others) have defined the occurrence process of earthquakes in terms of two components: i) a process of cluster centers; and ii) a subsidiary process defining the configuration of the members within a cluster. The final process is taken as the superposition of all the clusters. Several possible models for these processes can be found in the literature, for example the compound Poisson processes (Vere-Jones and Davis, 1966), trigger models (Vere-Jones and Davies, 1966; Lomnitz and Nava, 1983) or epidemic-type models (Hawkes, 1971; Lomnitz, 1974; Ogata and Akaike, 1982; Ogata, 1988 and their references). All these models assume structural conditions on the occurrence of earthquakes, as for example, that the process of cluster centers is stationary and Poissonian. In Section 4.1, we will show that this hypothesis is not very likely either for particular cases. A drawback of these approaches is that they depend very much on the models, and so are subject to the instability and goodness of fit problems noted in Section 1.

Because the standard parametric models do not always fit the earthquake time series well, we suggest making use of nonparametric analysis techniques. As mentioned in the previous section, nonparametric methods for analyzing the distribution of time intervals between consecutive earthquakes will be considered, obtaining another attempt to describe temporal behavior of an earthquake series in a geographic region.

3. Kernel estimation of hazard function

The distribution of time intervals between consecutive earthquakes can be characterized using the hazard function, which is defined by

$$r(t) = \lim_{\Upsilon t \to 0} \frac{P(\Delta t < t + \Upsilon t \mid \Delta t \geq t)}{\Upsilon t}$$

where Δt is the random variable that measures the time between consecutive shocks and $P(\cdot \mid \cdot)$ indicates the conditional probability. This function can also be written as

$$r(t) = \frac{f(t)}{1 - F(t)}$$

with $f(\cdot)$ and $F(\cdot)$ being the density and distribution functions of Δt, respectively. Thus, $r(t)\Upsilon t$ can be interpreted as the approximate probability of a shock in the time interval $(t, t + \Upsilon t)$ given that the immediately preceding one happened at time 0. In other words, it is considered the instantaneous hazard earthquake occurrence at time t.

Nonparametric estimation of hazard function started with Watson and Leadbetter (1964 a,b) who introduced the kernel estimator (see (3.1)), and from that time on many papers on this topic have appeared in the literature (see *e.g.*, Hassani *et al.*, 1986 for a survey). Most of this literature is based on the assumption of independence of the data. However, in the case of microearthquake studies, for instance, the hypothesis of dependence, that is, the existence of causality or interaction between the occurrence times of shocks, is more suitable (Rice and Rosenblatt, 1976). Papers on dependent hazard estimation include Sarda and Vieu (1989), Vieu (1991), Estévez and Quintela (1999, 2002) and many others.

Throughout this paper we will use the kernel estimator of the hazard function (Watson and Leadbetter, 1964 a,b) because it has been studied in great detail in dependence contexts. It is defined as

$$r_h(t) = \frac{f_h(t)}{1 - F_h(t)} \tag{3.1}$$

499

where

$$f_h(t) = \frac{1}{nh} \sum_{i=1}^{n} K\!\left(\frac{t - \Delta t_i}{h}\right)$$

and

$$F_h(t) = \int_0^t f_h(x)dx = \frac{1}{n} \sum_{i=1}^{n} H\!\left(\frac{t - \Delta t_i}{h}\right)$$

are the kernel estimators of $f(t)$ and $F(t)$, respectively; $K(\cdot)$ is a kernel function, $H(t) = \int_{-\infty}^{t} K(x)dx$, and $h = h(n) \in R^+$ the smoothing parameter or bandwidth. This parameter is crucial in the estimation method since the shape of the resulting estimator varies greatly according to its value.

Estévez and Quintela (1999, 2002) proposed two versions of the cross-validation method to select the bandwidth h when the data are dependent: the global and the local choice. The first one selects only one h, searching the lowest global estimation error. That is, if the global estimation error is

$$MISE(h) = E\int\left(r_h(x) - r(x)\right)^2 dx$$

the value of cross-validation bandwidth used by these authors is h_{cv}, that satisfies $h_{cv} = \underset{h}{\text{argmin}}\, CV(h)$, with

$$CV(h) = \int r_h^2(x)dx -$$

$$-\frac{2}{n}\sum_{i=1}^{n} \frac{f_h^{-i}(X_i)}{\left(1 - F_h^{-i}(X_i)\right)\left(1 - F_n(X_i)\right)} \qquad (3.2)$$
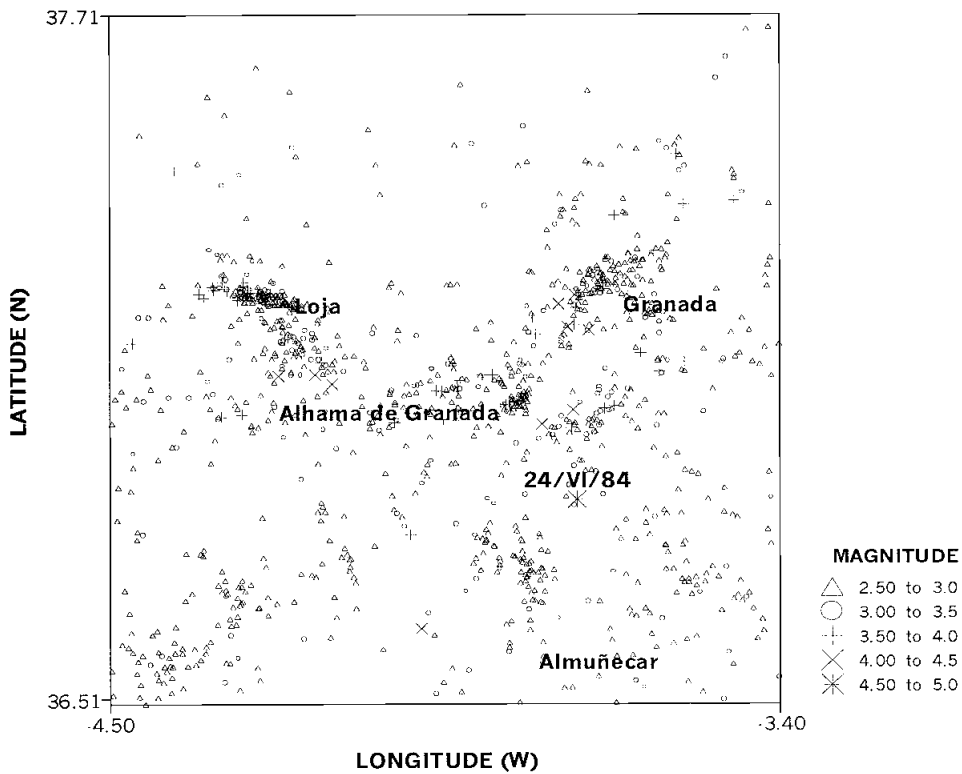
Fig. 1. Seismicity of Granada in the period 1983-1999.

an estimation of global error $MISE(h)$. The functions $f_h^{-i}(\cdot)$, $F_h^{-i}(\cdot)$ in (3.2) are kernel estimations of density and distribution functions modified to take into account the dependence of the data and $F_n(\cdot)$ is the empirical distribution function. Note that $MISE(h)$ depends on $r(\cdot)$ that is an unknown function and hence it is necessary to estimate it.

On the other hand, the local cross-validation method takes a different $h_{cv,x}$ for each estimation point x, obtaining the lowest estimation error for each x. In this case, $h_{cv,x} = \underset{h}{\operatorname{argmin}} CV_x(h)$ with $CV_x(h)$ an estimator of local error $MISE_x(h) = E(r_h(x) - r(x))^2$. We will illustrate these two approaches in the sections that follow.

4. Examples

In this section we try to study and compare the seismic activity of two Spanish geographic regions: Granada in the southeast and Galicia in the northwest of Spain.

4.1. *Granada earthquakes data*

The data-set to study is a group of micro-earthquakes recorded during the period 1983-1999 with epicenters between 36.5° to 37.7°N and 3.5° to 4.5°W and magnitude larger than or equal to 2.5 on the Richter scale. The scatter plot of epicenters (fig. 1) for the $n = 1117$ shallow shocks (focal depths of less than 30 km) illustrates the seismicity of the region studied.

Basic descriptive plots of the event magnitudes are given in fig. 2a,b. The first one presents the cumulative number of shocks with time with the associated magnitudes. The second one represents the distribution of magnitudes in this region. Note in particular the straight line fit in (b), which indicates that the Gutenberg-Richter
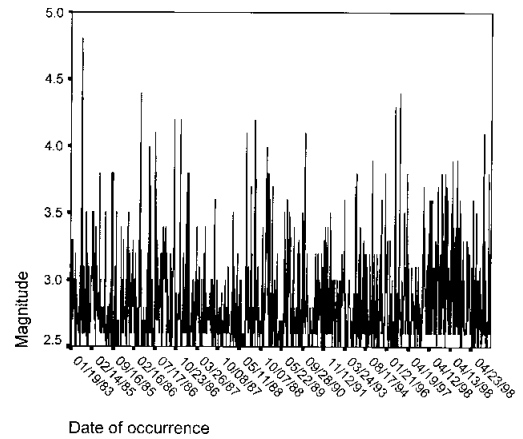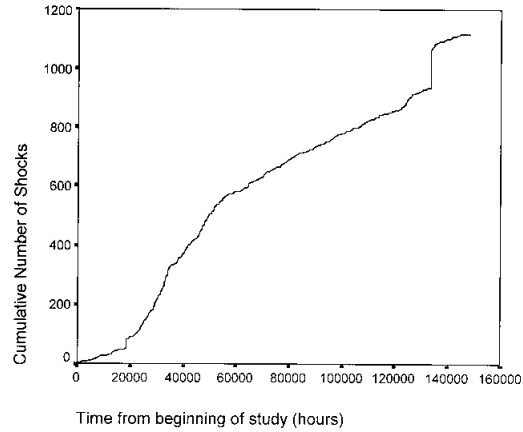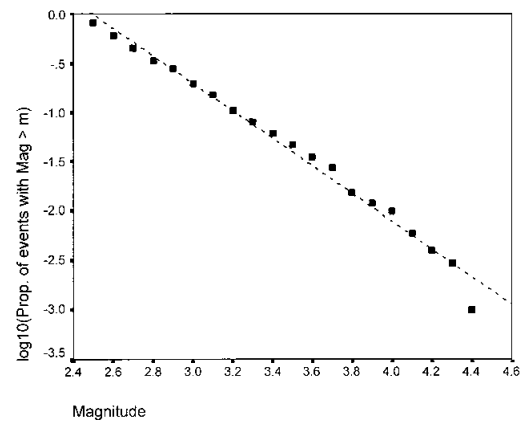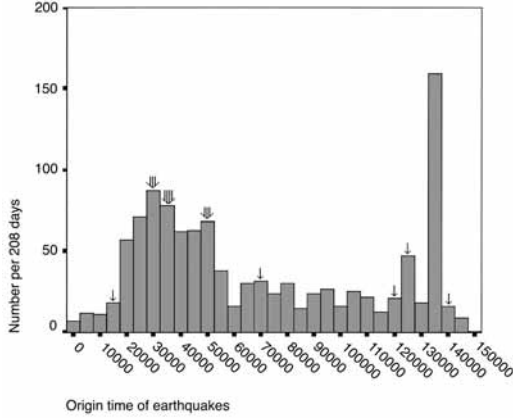

Fig. 2a,b. Basic descriptive plots of the seismic data: a) cumulative number of shocks and plot of the magnitudes versus the occurrence times; b) cumulative distribution of magnitudes.

501

Fig. 3. Time history of seismic activity during the period 1983-1999 given in number of shocks per 208 days. Arrows indicate the occurrence of one or more earthquakes with magnitude $M \geq 4$.
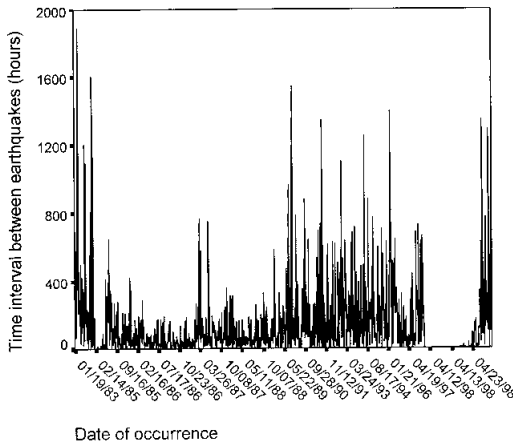
Fig. 4. Sequence of time intervals between consecutive earthquakes $\{\Delta t_i\}_{i=1}^{1116}$.

(1944) law of magnitude frequency holds for these data. This relation suggests that the catalog includes essentially all the events of this magnitude that occurred.

The histogram for the origin times in hours $\{t_i\}_{i=1}^{1117}$ (fig. 3) and the sequence graph for the

time intervals between consecutive earthquakes $\{\Delta t_i\}_{i=1}^{1116}$ (fig. 4) show a prominent trend towards clustering, particularly clear in April, 1998 with more than 100 events per 5 days. Note that large time intervals (Δt_i) indicate slack periods and times close to 0 show strong clustering. Arrows pointing downward at the upper part of the histogram mark the occurrence of one or more earthquakes (the number of bars indicates the number of shocks) with magnitude equal to or larger than 4 on the Richter scale.

The average number of earthquakes per day is 0.18 but the activity, as is shown by figs. 2a, 3 and 4, is not uniform with time. Figure 2a shows five time periods whose seismic activity seems fairly different. Combining the information provided by this graph and fig. 4, we can distinguish: a quite active time span from 1985 to 1989, a slack span from 1990 to 1995, and finally, strong seismic activity in April 1998. Thus, it is obvious that the process $\{\Delta t_i\}_{i=1}^{1116}$ is not stationary. A possible cause of the lack for stationarity could be that the studied time period is quite short because, as Ogata (1988) points out, «stationary models are considered here in the prior belief that such geophysical activity for a long time span should be stationary».

Concerning the hypothesis of dependence of the data, the theory of runs and the analysis of autocorrelations establish that this hypothesis is verified. In effect, the run test up and down gives a p-value of 0.039, the run test above and below the Median produces $2.81 \cdot 10^{-18}$ and the autocorrelation function in fig. 5 gives autocorrelations significantly different from zero.

The previously mentioned non-stationarity makes the estimation of hazard function difficult, however, the transformed process $y_j = \ln\Delta t_{j+1} - \ln\Delta t_j$; $j = 1, \ldots, n$, which is stationary as shown in fig. 6, solves the problem. In effect, as

$$y_n = \ln\Delta t_{n+1} - \ln\Delta t_n$$

and the random variable $\ln\Delta t_{n+1}$ conditioned to $\Delta t_1, \ldots, \Delta t_n (\ln\Delta t_{n+1} | \Delta t_1, \ldots, \Delta t_n)$ has the same distribution as $(y_n + \ln\Delta t_n) | \Delta t_1, \ldots, \Delta t_n$, that is

$$\ln\Delta t_{n+1} | \Delta t_1, \ldots, \Delta t_n \stackrel{d}{\equiv} (y_n + \ln\Delta t_n) \,\big|\, \Delta t_1, \ldots, \Delta t_n$$
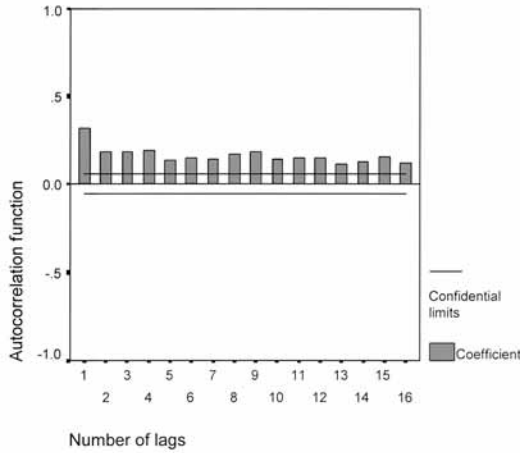
502

Fig. 5. Sample autocorrelation function for $\{\Delta t_i\}_{i=1}^{1116}$.
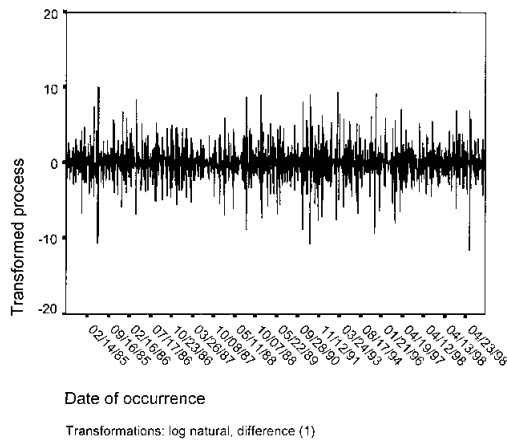

Transformations: log natural, difference (1)

Fig. 6. Sequence graph of transformed process $\{y_j\}$.

the hazard function of time intervals between consecutive earthquakes, given that the last registered time interval was Δt_n, is estimated using the hazard function estimation of $y_n(r_{h,1}(\cdot))$.

So, first we compute

$$r_{h,1}(y) = \frac{f_{h,1}(y)}{1 - F_{h,1}(y)} = \frac{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right)}{1 - \frac{1}{n}\sum_{i=1}^{n} H\left(\frac{y - y_i}{h}\right)}$$

with

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{if } u \in [-1, 1] \\ \\ 0 & \text{if } u \notin [-1, 1] \end{cases}$$

the Epanechnikov kernel and h calculated in the same way as in Estévez and Quintela (1999). Then, by means of a translation and a change of variable, we estimate the hazard function of time intervals between earthquakes, given that the last one has happened at time $t_n \equiv 17/12/1999$ at 22:31:59.06 h and $\Delta t_n = e^{7.5} = 1807.65$ h, obtaining the curve in fig. 7.

The shape of this curve (high hazard for small times and low hazard in other cases) suggests a prominent clustering. Moreover, the last time interval, $\Delta t_n = e^{7.5} = 1807.65$, plays an important role in the estimation process: if, for example, Δt_n was equal to $e^{3.95} = 51.93$ h or $e^{5.25} = 190.57$ h (the two last shocks could be closer), the instantaneous hazard earthquake occurrence at time t_n (dotted line and dashed line, respectively) would be higher.

The main drawback of this method is that we can only estimate the hazard for time t_n.

Now, let us suppose that the time series $\{\Delta t_i\}$ is stationary, that is, that we have not detected the lack of stationarity. In such case, we would estimate the hazard function of $\{\Delta t_i\}$ using

$$r_h(t) = \frac{f_h(t)}{1 - F_h(t)}$$

for $t > 0$ and h the cross-validation bandwidth (Estévez and Quintela, 1999). By comparing this curve (fig. 8) with the previous ones, we can conclude that both methods engender very

503
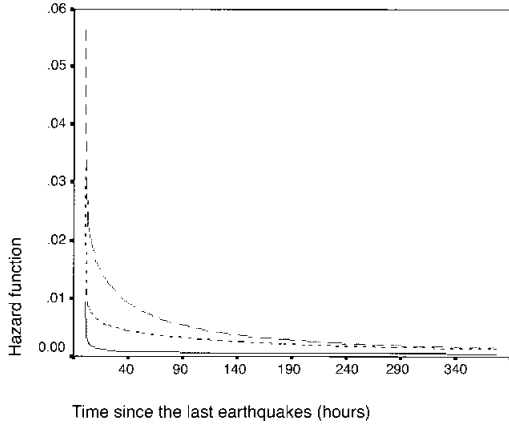
Fig. 7. Conditional hazard at t_n (solid line). Hazard estimation for $\ln\Delta t_n = 3.95$ (dashed line) and for $\ln\Delta t_n = 5.25$ (dotted line).
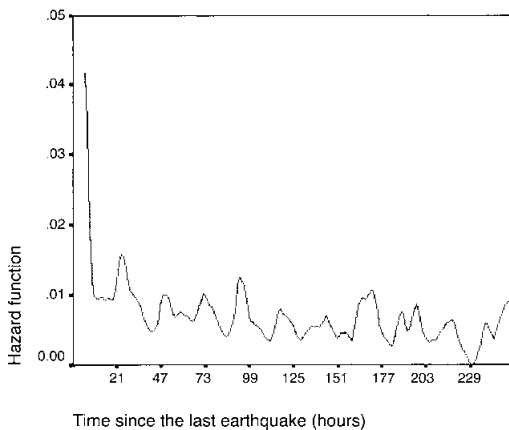
Fig. 8. Hazard estimation under stationarity.

similar estimations when $\ln\Delta t_n$ is not an outlier. So, if we are interested in estimating the hazard at a particular time t_n, the first procedure, which takes into account the information provided by Δt_n, should be used. However, to estimate the hazard function globally, the stationary could be ignored and the second procedure, which is easier, would be used.

In both cases the estimated hazard function suggests the natural strong clustering of the occurrence times of shocks.

By defining a cluster as a set of earthquakes originating from a relatively small volume and separated in time by intervals smaller than a fixed duration («cluster length»), and cluster center as a representative of the cluster (the first shock, for instance), the occurrence process of earthquakes is taken as the superposition of all the clusters.

In this problem we have considered clusters with «cluster length» of less than 120 h, obtaining a set of independent cluster centers. That is, after removing aftershocks and precursors, the remaining shocks can be taken as independent events. A length of 120 h was chosen because it was the smallest value producing independent shocks, that is, removing the effect of the aftershocks. Figure 9a,b presents the sequence graphs of the cluster sizes (fig. 9a) and the magnitudes of the centers (fig. 9b).

By comparing fig. 9a,b with fig. 3, we observe that the stronger earthquakes occur at the start of or within small clusters. In other words, the great clusters are formed by events of small magnitude.

The largest cluster will be studied later.

We begin by analyzing the occurrence process of cluster centers by means of the distribution of time intervals between them $\{\Delta t_i'\}_{i=1}^{345}$. These events happen independently (the p-values of runs test above and below the median, and up and down are 0.553 and 0.522, respectively) but they do not have exponential distribution (the p-value of the Kolmogorov-Smirnov test for goodness of fit is less than 0.01). Thus, such centers do not form a stationary Poisson process contrary to what Vere-Jones and Davies (1966), Vere-Jones (1970) and Udias and Rice (1975) used. Since several tests for goodness of fit show that any known model of distribution seems suitable to describe the distribution of $\{\Delta t_i'\}_{i=1}^{345}$, we resort to nonparametric methods to estimate this distribution. In particular, we use the kernel estimator of hazard function (3.1). Figure 10 presents this estimation for $\{\Delta t_i'\}_{i=1}^{345}$ when the bandwidths have been globally selected (see Estévez and Quintela, 1999) (solid line) and locally selected

504
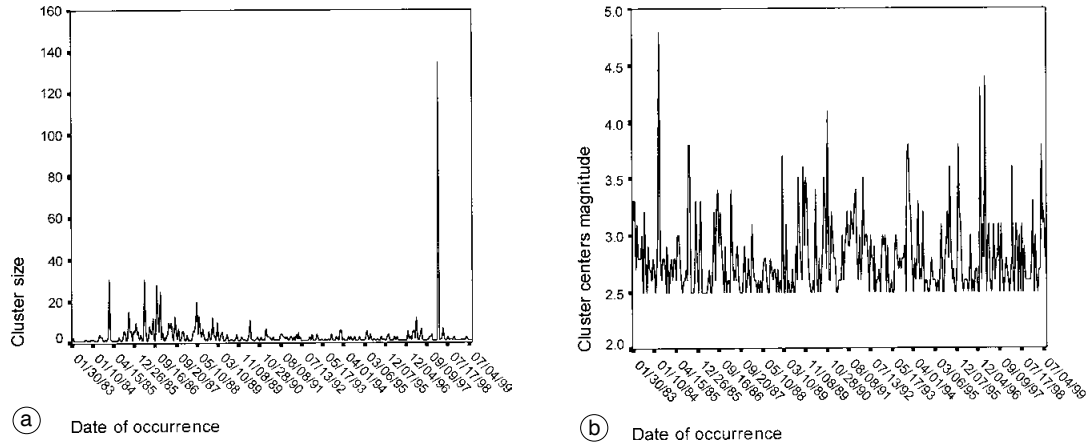
Fig. 9a,b. a) Sequence graph of cluster sizes (a total of 346); b) sequence graph of magnitudes of cluster centers.
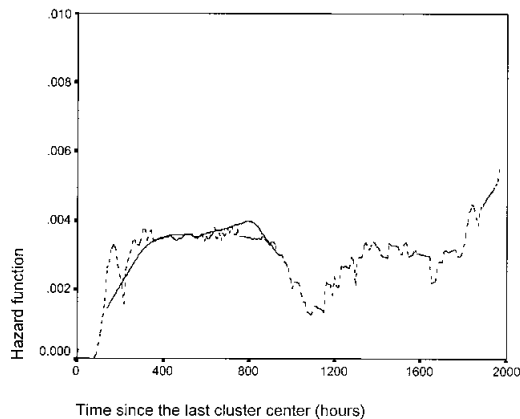
Fig. 10. Hazard estimation of time intervals between cluster centers with global bandwidth (solid line) and local bandwidths (dashed line).

(see Estévez and Quintela, 2002) (dashed line). The low values of hazard function and the small fluctuations throughout indicate that clusters happen with a low occurrence rate and are almost constant with time.

The occurrence process of the members within the largest cluster is formed by 133 events which happened from April 11 to April 30, 1998. Their epicenters are concentrated in a small area close to Loja, southwest of Granada (Spain). Although the number of shocks per day was large (6.65), none exceeded magnitude 3.9 on the Richter scale.

The sequence graph of time intervals between consecutive cluster members $\{\Delta t_i^{''}\}_{i=1}^{132}$, shown in fig. 11, indicates that such process is not stationary and the last span $\Delta t_{132}^{''}$ is large compared to the other ones.

Therefore, it will not be possible to estimate the distribution of time intervals between cluster members.

The behavior of the cluster will be studied using the process of origin times $\{t_i^{''}\}_{i=1}^{133}$ and its intensity function. This function gives, for each time t, $\lambda(t)$ the average number of earthquakes per unit of time, t hours after the main shock (cluster center). The advantage of using the intensity function is two-fold: firstly place is not affected by the lack of stationarity of the time intervals between shocks. Secondly, it clearly shows the evolution of the cluster with time, and therefore, it is suitable for comparing the seismic activity of several geographic regions. In this work, we propose to estimate the intensity function by means of a kernel estimator, which
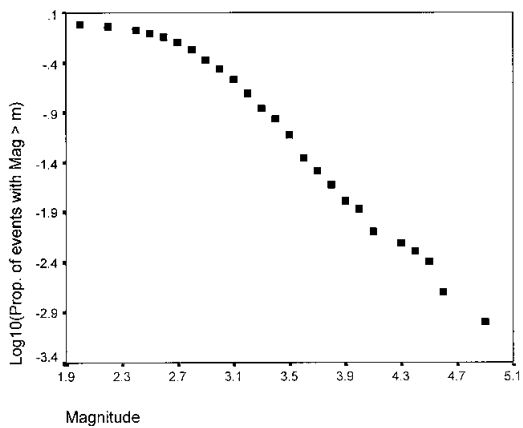
505
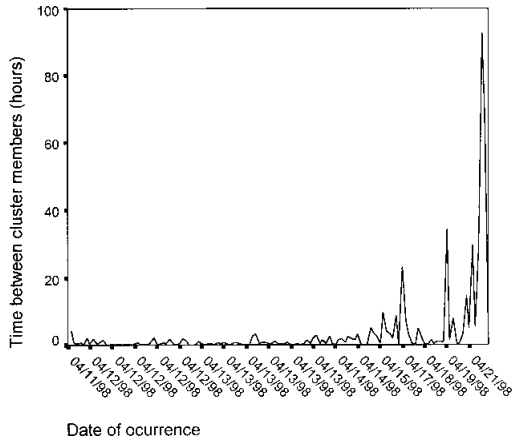
Fig. 11. Sequence graph of time intervals between consecutive cluster members.
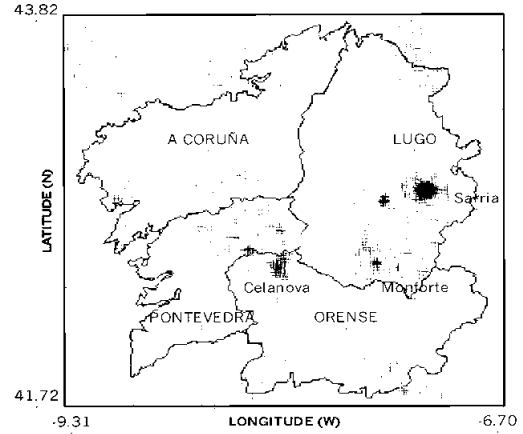
Fig. 13. Seismicity of Galicia in the period 1987-2000.
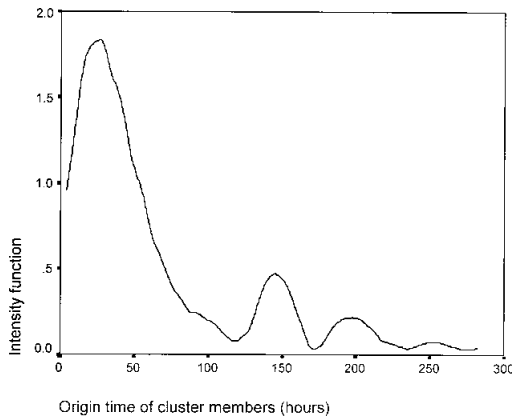
Fig. 12. Intensity estimation of the largest cluster.

Fig. 14. Cumulative distribution of magnitudes.

is defined as

$$\hat{\lambda}(t) = \frac{1}{h} \sum_{i=1}^{n} K\left(\frac{t - t_i''}{h}\right), \quad \forall t \in R^+$$

(Wand and Jones, 1995; Choi and Hall, 1999 and their references), where $K(\cdot)$ is the kernel function and h the bandwidth, as in (3.1). For this example, the estimated curve appears in fig. 12, showing a clear increase in the number of shocks at the

beginning of the cluster until the time with highest intensity (approximately at $t = 30$ h) and then a decrease.

4.2. *Galicia earthquakes data*

Seismic activity increased in this region during the 1990's. The data analyzed correspond to the 978 earthquakes occurring from January

506

15, 1987 to May 3, 2000. Their epicenters, scattered throughout Galicia (fig. 13), show small spatial groupings at high risk areas: Sarria, Monforte and Celanova.

The magnitude distribution of this region is shown in fig. 14, which supports the Gutenberg-

Richter law of magnitude frequency. The histogram of origin times $\{t_i\}_{i=1}^{978}$, the sequence of time intervals between consecutive shocks $\{\Delta t_i\}_{i=1}^{977}$ and the graph of the cumulative number of shocks with their magnitudes are shown in fig. 15a-c. These graphs show a quiet period until

(a)

(b)

(c)
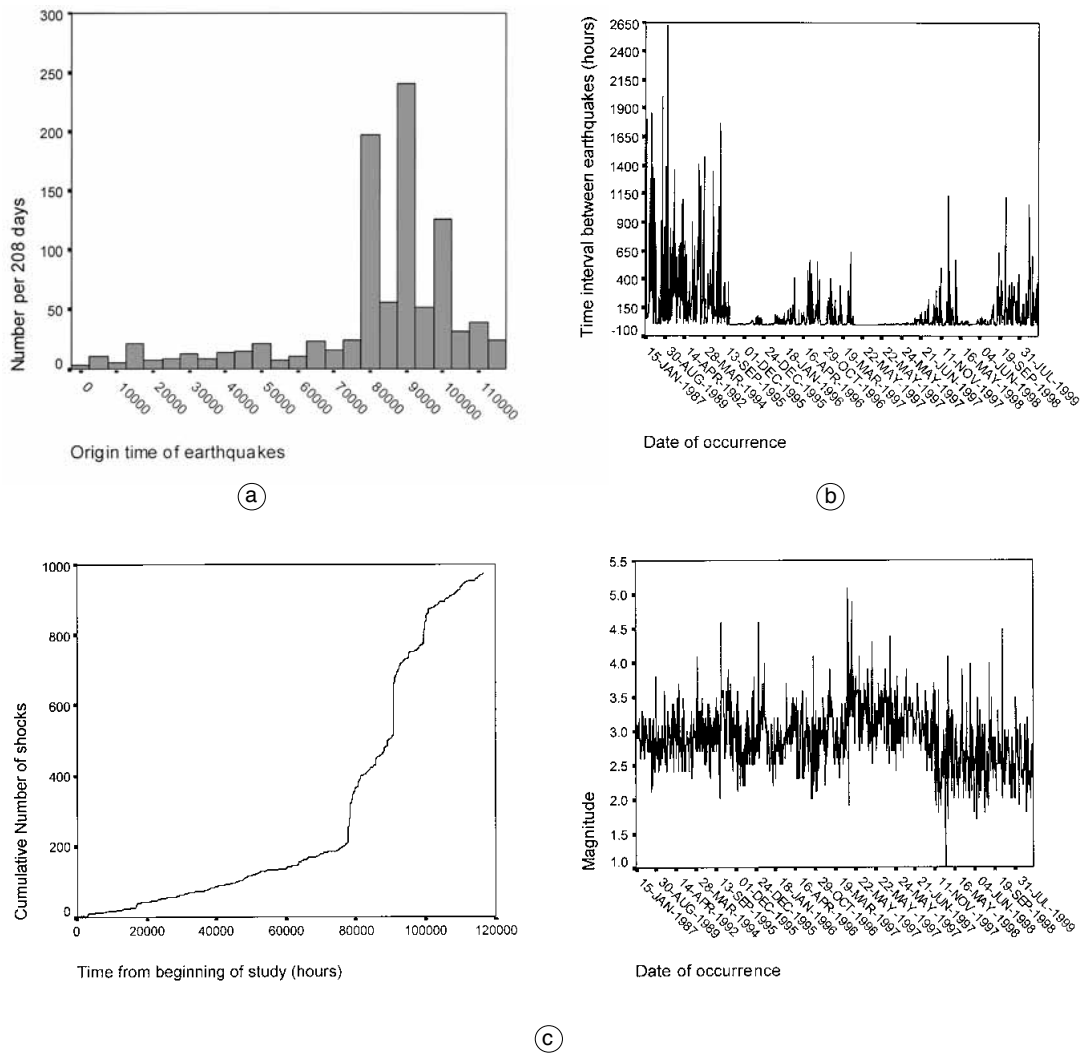
Fig. 15a-c. a) Histogram of origin times $\{t_i\}_{i=1}^{978}$, number of shocks per 208 days; b) sequence graph of time intervals between consecutive earthquakes $\{\Delta t_i\}_{i=1}^{977}$; c) cumulative number of shocks with sequence graph of magnitudes.

507

September 1995 and three temporal groupings in December 1995, May 1997 and May-June 1998, which are related to the strongest earthquakes.

The hazard function estimation (fig. 16), which was constructed as in the previous example, presents a common shape: it decreases
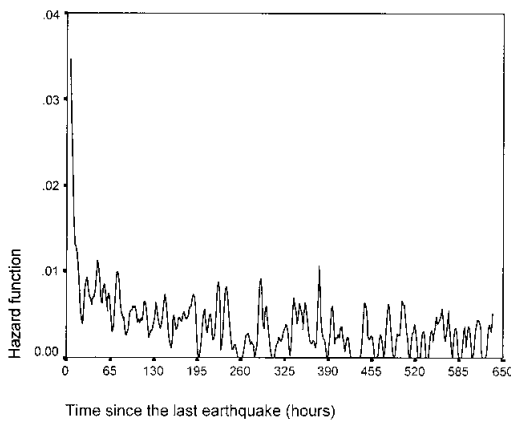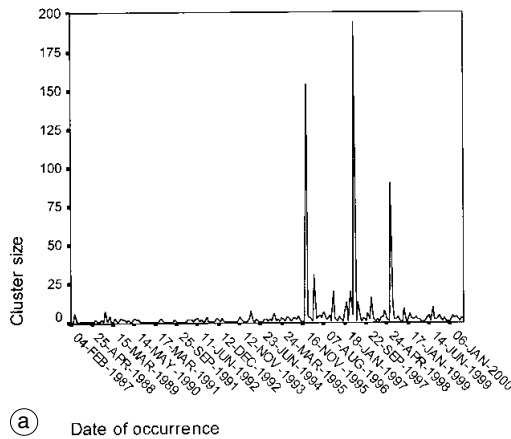

Fig. 16. Hazard estimation of time intervals between consecutive shocks.

suddenly in the first hours and then it fluctuates around a small risk. This shape confirms the well known fact that earthquakes form clusters.

Similarly, as in the Granada case, 206 clusters with «cluster length» less than 144 h have been formed. The sequence of sizes and magnitudes of cluster centers (fig. 17a,b) show three important clusters, the first one beginning with a big shock (4.7 on the Richter scale). Figures 17a,b and 15c also suggest that the strongest earthquakes are related to small clusters, that is, such events are not followed or preceded by many earthquakes.

The hazard estimation of time intervals between cluster centers (fig. 18) suggests small and stable risk of clusters in time. However, as in the Granada case, the distribution of time intervals is not exponential either (the p-value of the Kolmogorov-Smirnov test for goodness of fit is 0.0006).

Finally, the occurrence of cluster members in each grouping was analyzed using the intensity function. The first cluster (fig. 19a) is composed of 147 shocks, whose epicenters are contained in a small area close to Sarria (Lugo). The intensity estimation indicates that: i) the cluster begins with a high number of shocks (the first of magnitude 4.6); ii) there is another grouping of
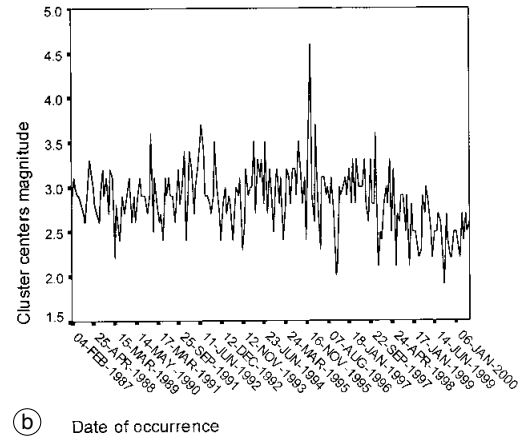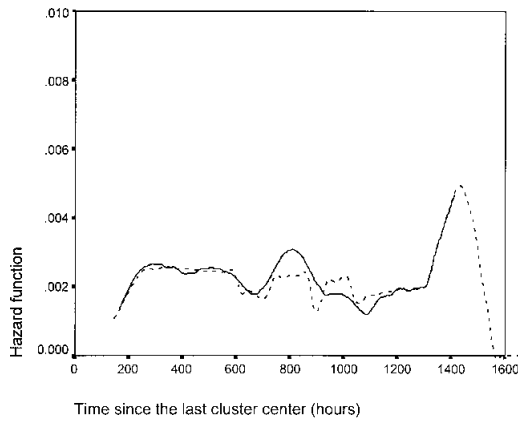

Fig. 17a,b. a) Sequence graph of cluster sizes (a total of 206); b) sequence graph of magnitudes of cluster centers.

508

Fig. 18. Hazard estimation of time intervals between center clusters with global bandwidth (solid line) and local bandwidths (dashed line).

events coinciding with another earthquake of magnitude 4.6, and iii) the intensity function is low in the remainder of the range.

The second cluster (fig. 19b) consists of 190 earthquakes, whose epicenters are also near Sarria (Lugo). The shape of the intensity function reflects the behavior of the cluster members: a small group of precursors warns that important shocks are to arrive (one of magnitude 5.1 and another of 4.9) and then there is a sequence of aftershocks.

Finally, the third cluster (fig. 19c) involves 79 events, with epicenters in Celanova (Orense). This is the weakest cluster (few shocks of low magnitude). Its estimated intensity function shows fluctuations around quite small values the whole time.
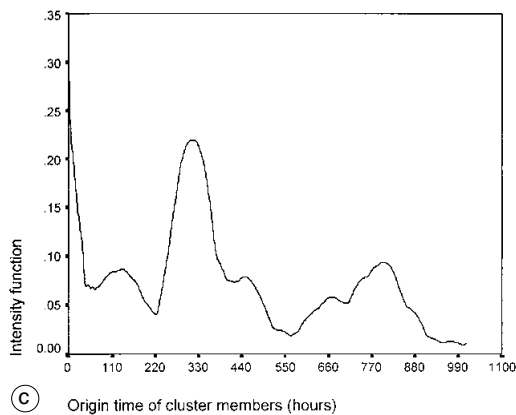
5. Conclusions

In this work we show how the hazard and intensity functions represent another way of describing the temporal structure of seismic activity in a geographic region. Kernel estimation of hazard function has confirmed what Vere-Jones (1970), Udias and Rice (1975) and many others have noted: earthquakes have the tendency
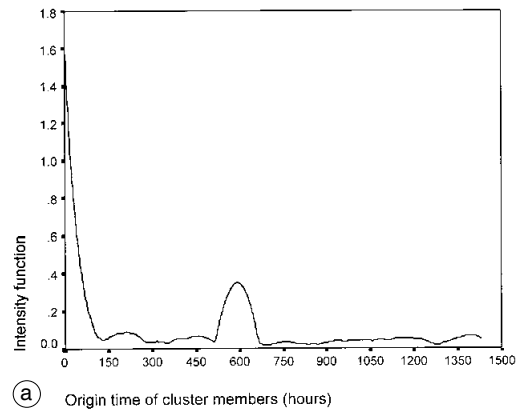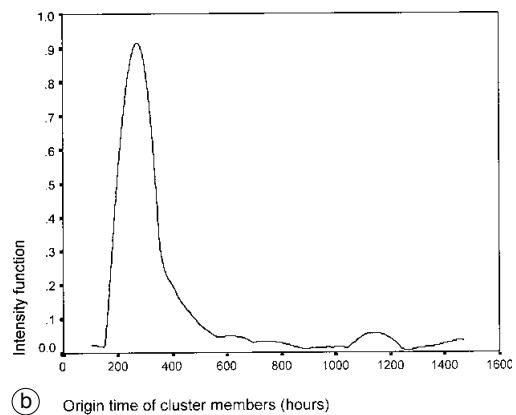
Fig. 19a-c. Intensity estimations for important clusters: a) December 1995; b) May 1997, and c) May-June 1998.

509

to group. The occurrence process of these groups was also studied by means of the hazard function and each important cluster has been described using the intensity function.

As we argued, a major advantage of nonparametric methods is that they do not require formulation of structural models, which are often well suited only for data that have closely related seismic causes. Another novelty is that we take into account the possible dependence of the data to estimate the distribution of time intervals between earthquakes, which has not been considered in most statistics studies on seismicity.

In reference to the results obtained on the seismic activity of Granada and Galicia, we may point out that: due to the effect of aftershocks, the hazard function of time intervals between consecutive shocks in an earthquake catalog will present a shape as that shown in fig. 8 and 16. Its greater or smaller above zero is an indicator of the intensity of seismic activity of each region. In Granada and Galicia the maximum values of hazard are similar when shocks with magnitude from 2 to 5 are considered.

Next, to analyze the seismic activity of the Granada and Galicia region we formed clusters lasting less than or equal to 120 and 144 h, respectively. In both cases, a sequence of main shocks mutually independent with low rate of occurrence and almost constant with time has been obtained. Thus, consecutive earthquakes far away from 120 or 144 h do not present interaction. The processes of main shocks seem very similar for both regions.

Concerning the greatest clusters, although the intensities take different shapes and values (see fig. 12 and 19a-c), there is a clear predominance of shocks at the beginning of clusters. However, the Granada cluster is shorter and intense and in the Galicia region the aftershocks are more aloof in time. Moreover, the clusters of Galician earthquakes present intensity functions with different shapes. This characteristic suggests the use of intensity function, of its shape in particular, to characterize and compare different seismic regions. Note this characterization is possible thanks to the use of nonparametric methods, which do not predetermine the shape of such curves.

It is important to point out that although our conclusions contradict the belief that Granada has more seismic activity than Galicia, seismicity in the latter has grown much in recent years.

Acknowledgements

REFERENCES

CHOI, E. and P. HALL (1999): Nonparametric approach to analysis of space-time data on earthquake occurrences, *J. Comput. Graphic. Stat.*, **8** (4), 733-748.

ESTÉVEZ, G. and A. QUINTELA (1999): Nonparametric estimation of the hazard function under dependence conditions, *Commun. Stat. Theory Methods*, **28** (10), 2297-2331.

ESTÉVEZ, G. and A. QUINTELA (2002) : Estimación no paramétrica de la función de riesgo: aplicaciones a sismología, *Questiió*, **25** (3), 437-477.

GUTENBERG, R. and C.F RICHTER (1944): Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.,* **34**, 185-188.

HASSANI, S., P. SARDA and P. VIEU (1986): Approche non paramétrique en théorie de la fiabilité, *Rev. Stat. Appl.*, **35** (4), 653-676.

HAWKES, A.G. (1971): Point spectra of some mutually exciting point processes, *J. R. Stat. Soc., Ser. B.*, **33**, 438-443.

KNOPOFF, L. (1964): The statistics of earthquakes in Southern California, *Bull. Seismol. Soc. Am.*, **54**, 1871-1873.

LOMNITZ, C. (1966): Statistical prediction of earthquakes, *Rev. Geophys.*, **4**, 377-393.

LOMNITZ, C. (1974): *Global Tectonics and Earthquakes Risk* (Amsterdam: Elsevier).

LOMNITZ, C. and A. HAX (1967): Clustering in aftershock sequences in The Earth Beneath the Continents, *Geophys. Monogr. 10, Am. Geophys. Union.*

LOMNITZ, C. and F. A. NAVA (1983): The predictive value of seismic gaps, *Bull. Seismol. Soc. Am.*, **73**, 1815-1824.

OGATA, Y. (1988): Statistical models for earthquakes occurrences and residual analysis for point processes, *J. Am. Stat. Ass.,* **83** (401), 9-27.

OGATA, Y. (1989): Statistical model for standard seismicity

and detection of anomalies by residual analysis, *Tectonophysics,* **169**, 159-174.

OGATA, Y. (1998): Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Mathem.,* **50**, 379-402.

OGATA, Y. and H. AKAIKE (1982): On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes, *J. R. Stat. Soc., Ser. B*, **44**, 102-107.

RICE, J. and M. ROSENBLATT (1976): Estimation of the log survival function and hazard function, *Sankhya, Ser. A*, **36**, 60-78.

SARDA, P. and P. VIEU (1989): Empirical distribution function for mixing random variables. Application in nonparametric hazard estimation, *Statistics*, **20** (4), 559-571.

UDIAS, A. and J. RICE (1975): Statistical analysis of microearthquakes activity near San Andres Geophysical Observatory, Hollister, California, *Bull. Seismol. Soc. Am.*, **65**, 809-828.

VERE-JONES, D. (1970): Stochastic models for earthquake occurrence, *J. R. Stat. Soc.*, *Ser. B*, **32**, 1-62.

VERE-JONES, D. and R.B. DAVIES (1966): A statistical survey of earthquakes in the main seismic region of New Zealand. Part 2. Time Series Analyses, *N. Z. J. Geol. Geophys.*, **3** (9), 251-284.

VIEU, P. (1991): Quadratic errors for nonparametric estimates under dependence, *J. Multivariate Anal.*, **39**, 324-347.

WAND, M.P. and M.C. JONES (1995): *Kernel Smoothing* (Chapman & Hall), 167-168.

WATSON, G.S. and M.R. LEADBETTER (1964a): Hazard analysis I, *Biometrika*, **51**, 175-184.

WATSON, G.S. and M.R. LEADBETTER (1964b): Hazard analysis II, *Sankhya, Ser. A*, **26**, 110-116.