ARTICLE IN PRESS

Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
Journal of Statistical Planning and Inference ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

ELSEVIER

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Non-parametric smoothing of spatio-temporal point processes

Carlo Grillenzoni*

IUAV, Department of Planning, Santa Croce 1957, Venice 30135, Italy

Received 16 January 2003; accepted 7 September 2003

Abstract

This paper develops adaptive non-parametric modelings for earthquake data. Non-parametric techniques are particularly suitable for space–time point processes, however they must be adapted to deal with the non-stationarity of seismic phenomena. By this we mean changes in the spatial and temporal pattern of seismic occurrences. A set of non-parametric tests, kernel density and regression estimators are proposed to study the space–time evolution of earthquakes. The implied solutions, by respecting the unidirectional nature of time and minimizing prediction errors, are naturally oriented to forecasting. An extensive application to the Northern California Earthquake Catalog (NCEC) data-set, starting from 1930, illustrates and checks the approach.

Keywords: Bandwidth; California Earthquakes; Exponential Weighting; Kernel density and regression; Recursive tests; Space–time non-stationarity

1. Introduction

In essence, a space–time point process is a random collection of points whose coordinates provide the instant and the location of an event. These events may be binary or may be *marked*, in such a case an additional variable is involved. Unlike continuous space–time processes, which can be represented as multi-indexed sequences of random variables $\{Z_{st}\}$, in a point process the indices themselves constitute part of the process, namely $\{(\boldsymbol{s}, t, Z)_k\}$ (see Schoenberg et al., 2002). Forecasting the spatial location $\boldsymbol{s}$, the occurrence time t and the mark Z of a future event is the main target in many fields of research.

Statistical methods for point processes have been extensively developed in the context of seismology. Their findings, however, may be useful for other disciplines, such as

* Tel.: +39-41-2572112; fax: +39-41-5140403.
E-mail address: carlog@iuav.it (C. Grillenzoni).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.09.030

economy (modeling birth-death of firms) and epidemiology. In studying earthquakes, the main concern is in estimating the *intensity* and *strength* of the phenomenon. By this it is meant the frequency with which events are expected to occur at a particular point and their mark in neighboring points. Such estimates may provide useful indications on the location and the size of future occurrences.

A number of sophisticated models have been provided for representing the intensity function. Many of these are extensions of models only developed in the temporal dimension. The two major approaches are the *self-exciting* scheme, for events which tend to be clustered, and the *self-correcting* scheme, for regular and inhibitory point patterns. The space–time extension of the self-exciting scheme of Hawkes has been pursued by Vere-Jones (1992), Rathbun (1993) and Ogata (1998); recent refinements are proposed by Zhuang et al. (2002). The resulting models can represent swarms of earthquakes due to aftershock activity.

Structural models have proved their effectiveness over short periods of time and in homogeneous areas, where common physical causes are operating. When fitted to more general data sets, however, they tend to provide estimates which lie outside plausible ranges and are sensitive to the inclusion of new data (see Rathbun, 1993). This is due to the fact that seismic phenomena are not only characterized by strong randomness, but also are intrinsically non-stationary in space and time. By this we mean both spatial migration of seismic sites and time-variability in the rate of occurrence of significant events. Such pattern changes usually produce anomalous observations which perturb parameter estimates.

A practical alternative to structural models is represented by *non-parametric* smoothing methods. Density estimators can easily manage space and time variables simultaneously and can measure non-stationary intensity functions. Similarly, regression smoothers can estimate the strength function, which has not received much attention in the literature. The Kernel approach, in particular, is very flexible and through the *sequential* implementation (Grillenzoni, 2000) it can manage time-varying functions without using time as an explicit variable.

Non-parametric estimation of the intensity function has already been pursued in seismology, see Vere-Jones (1992), Bailey and Gatrell (1995) and Stock and Smith (2002). However, the proposed solutions do not consider the temporal dimension or do not respect its *unidirectional* nature, because treat time as an additional spatial axis. As a consequence, they estimate the present value of the functions by also including future events. As we shall see, time-irreversibility has also important implications on the criteria for selecting optimal bandwidths.

As regards non-parametric regression, it may also be a useful alternative to Kriging techniques for interpolating surfaces. Its advantages are computational simplicity and adaptability to time-varying problems; on the contrary, inclusion of time in Kriging smoothers is a difficult task. At the pure spatial level, Yakowitz and Szidarovszky (1985) showed superiority of Kernel regression with respect to Kriging; but more recently, Hobert et al. (1997) have mitigated this conclusion.

The article is organized as follows: To better involve non-specialized readers, Section 2 provides a general description of the data-set which regards the application of the paper. Section 3 investigates space–time dependence tests based on point pattern

analysis. This is a necessary step before discussing any modeling issue. The core section of the paper, Section 4, develops sequential weighted kernel estimators. Finally, Section 5 discusses specialized aspects of space–time smoothing, such as recursive implementation and forecasting.

2. Data description

Seismic data may be considered as the realization of a marked space–time point process. This is defined by a sequence of $(c+2)$-variate random variables $\{\boldsymbol{s}_k, t_k, Z_k\}$ ordered by time; in particular: $k \in \Im^+$ is the index (it belongs to positive integers); $t \in \Re^+$, $t_k > t_{k-1}$ is the time variable (unidirectional); $\boldsymbol{s} \in \Re^c$, $c = 1, 2, 3$ is the spatial variable (c is its dimension); $Z \in \Re_{[0,10]}$ is the magnitude of an event on the Richter scale. If $c = 3$, we have the full 3D space, therefore $\boldsymbol{s}' = [x, y, z]$ are the coordinates of the epicenters: x is the latitude, y is the longitude (both are usually expressed in degrees) and z is the depth (in meters or km). On the other hand, for $c = 1$ one has the projection on a particular axis.

To better involve non-specialized readers, we introduce in advance the data set concerning the application. The Northern California Earthquake Data Center (NCEDC) is a joint project of the University of California Berkeley (UCB) and the United States Geological Survey (USGS). It is a long-term archive and distribution center for seismological and geodesic data for Northern and Central California; at the Internet address `http://quake.geo.berkeley.edu/` an on-line `ftp` service is available for retrieving such data in ASCII format.

Most general catalog is that of the UCB Seismological Laboratory, which includes events even of magnitude $Z_k = 0.1$ on the Richter scale, starting on year 1910. Unfortunately, these data are *incomplete*, in that from 1910 to 1930 only 17 events were available; in other periods, only events of magnitude 2.5 or higher were recorded; finally, the depth coordinate z was not computed before 1984. For reasons of homogeneity, our analysis only regards data in the period Jan 1931–Dec 2000, with minimum magnitude 2.5 and without the depth component. As a consequence, $\boldsymbol{s}' = [x, y] \in \Re^2$ and the sample size N became 17868.

Figs. 1 and 2 provide descriptive plots of the seismic data. Fig. 1 shows the geographic location of epicenters for increasing levels of magnitude of the events. Fig. 2 performs the same exercise for the occurrence time. It may be noted that events tend to cluster at both levels, although clustering in space is much more apparent. In the next section we formally address issues of statistical tests.

It is well known that location of earthquakes tends to follow tectonic faults, which may be geologically identifiable. In this context, geographic prediction becomes a relatively simple task. In Fig. 1(c) one may note the presence of at least 4 different spatial clusters which surround the Sacramento Valley. The nearly absence of epicenters in this area makes it similar to an hurricane eye.

Forecasting the occurrence time is a more challenging issue. At a given location, a typical feature is that big earthquakes are preceded by a quiescent time and are followed by a swarm of smaller events. The matter, however, is not so simple because according

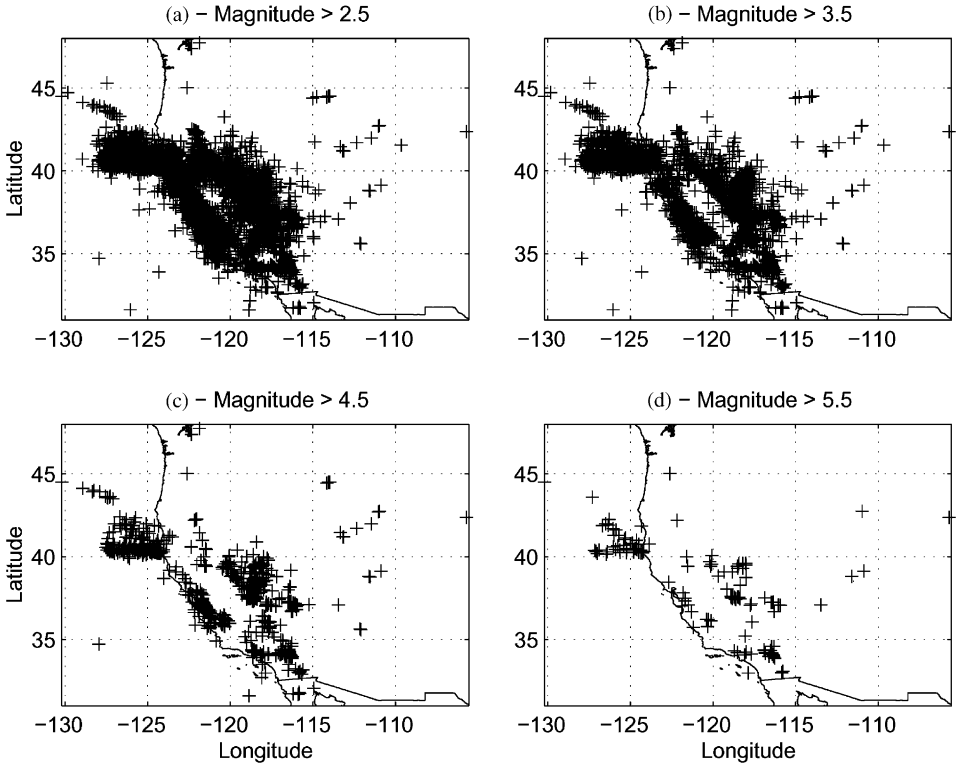

Fig. 1. Geographic location of epicenters of seismic data for different levels of magnitude. Latitude and longitude are expressed in degrees.

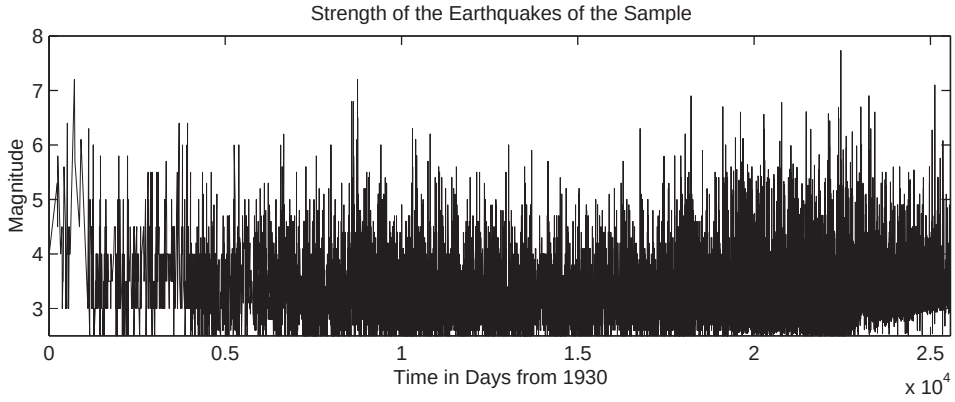

Fig. 2. Plot of event strength over time for the selected sample, $N = 17868$. Magnitude is on the Richter scale and time is in number of days from Dec 31, 1930.

to the self-exciting theory, a significant earthquake increases the rate of occurrence in the short term, but according to the *strain-release* theory it decreases the rate in the medium term. Bolt and Schoenberg (2000) have encompassed the two dynamics in their SELC model, which concerns restricted areas.

For the needs of subsequent analysis the variables $(x, y; t)$ should be expressed in decimal units. Time was then expressed in number of days from Dec 31, 1930 and latitude and longitude was transformed in meters, using the Universal Transverse Mercator (UTM) projection. This transformation was also required because the area of investigation is wide and calculating distances on coordinates in degrees leads to distortion. The UTM system can be viewed as a Cartesian scheme, in which the x-axis is the equator and the y-axis is a meridian passing from a specified point. In our case we selected the meridian passing through the city of San Francisco.

3. Space–time dependence tests

In this section we consider the problem of testing for space–time dependence in seismic data. This is an important topic in the analysis of point processes and is preliminary to any modeling attempt. A peculiar feature of earthquakes is that they exhibit two opposite type of dependence: one leading to clustering (on a small scale and aftershock activity) and the other leading to spacing out of events (on a wider scale and main shocks). In Figs. 1 and 2, one may appreciate existence of this feature both at the spatial and temporal level; it is certainly interesting to test what kind of dependence scheme tends to prevail.

NN analysis: Dependence analysis has been well developed in spatial and temporal contexts separately. For example, at spatial level the *nearest neighbor* (NN) analysis examines the distances between each point and the closest point to it, and then compares these to expected values for a sample of points from a complete random pattern. The basic test statistic is the average distance

$$\hat{d}_N = \frac{1}{N}\sum_{k=1}^{N} d_k, \quad d_k = \min_{h\neq k}\left[d_{kh} = \sqrt{(x_k - x_h)^2 + (y_k - y_h)^2}\right],$$

where $\{x_k, y_k\}$ are spatial coordinates of the kth event, and d_{kh} is the Euclidean distance between points k and h.

In order to treat time dependent events and their possible non-stationary patterns, one can use the sequential weighted mean

$$\hat{d}_t = \left(\sum_{k:t_k\leqslant t} \lambda^{t-t_k}\right)^{-1} \sum_{k:t_k\leqslant t} \lambda^{(t-t_k)}\, d_k w_k,$$

$$d_k = \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}, \tag{1}$$

where the concept of nearest neighbor now regards the time dimension only. In statistic (1) the factor $\lambda \in (0, 1)$ gives more weight to recent observations, whereas the weights

$w_k \geqslant 0$ give more or less weight to the distances, according to the magnitude of the corresponding events. A suitable choice is $w_k = Z_k Z_{k-1}/\bar{Z}_t^2$, where $\bar{Z}_t = t^{-1}\sum_{k=1}^{t} Z_k$, because $\{w_k\}$ would move around 1.

Alternatively, one may consider the nearest neighbor approach with respect to *all* events that occurred before t, and including time in the distance measure. This choice is motivated by viewing the data-set as organized in a 3-way array, where time is the vertical dimension. Moreover, using the adaptive weights

$$\tilde{w}_{kh} = Z_k Z_h/\tilde{W}_t, \quad \tilde{W}_t = \sum_{k:t_k \leqslant t} \lambda^{t-t_k} Z_k Z_h \Big/ \sum_{k:t_k \leqslant t} \lambda^{t-t_k},$$

where $t_h < t_k$, it is easy to verify that the statistic (1) becomes

$$\hat{D}_t = \left(\sum_{k:t_k \leqslant t} \lambda^{t-t_k} Z_k Z_H\right)^{-1} \sum_{k:t_k \leqslant t} (\lambda^{t-t_k} Z_k Z_H) D_{kH},$$

$$D_{kH} = \min_{h<k} \left[D_{kh} = \sqrt{(x_k - x_h)^2/\sigma_x^2 + (y_k - y_h)^2/\sigma_y^2 + (t_k - t_h)^2/\sigma_t^2}\right], \tag{2}$$

where H is the value of h corresponding to the minimum of D_{kh} and σ_x^2, σ_y^2, σ_t^2 are the variances of the coordinates. Notice that standardization makes homogeneous, and therefore summable, spatial and temporal distances.

The value of the discounting factor λ depends on the memory of the process; estimates in the next section indicate, on average, 0.955. This value is moderate, but such that the sequence λ^k decays rapidly. Application of statistics (1) and (2) with $\lambda = 0.96$ to the NCEC data provided the results in Fig. 3. Probability distributions of $\{\hat{d}_t, \hat{D}_t\}$, under the null hypothesis of random pattern, were derived with Monte Carlo simulations based on 1000 replications and a Kernel smoother. Such distributions are nearly Gaussian with means and variances that depend on the choice of λ and on spatial and temporal boundaries of the data-set shown in Figs. 1 and 2.

Analyzing Fig. 3, we may note that statistic (1) shows local independence of events, whereas statistic (2) confirms persistence of space–time dependence. These results are not contradictory because they come from qualitatively different methods. In particular, statistic (1) by working on temporally ordered data, is sensitive to events that belong to different seismic sites, and these may easily be independent. A common feature of the two statistics is their high value at the beginning (where the catalog is less complete), which may mean existence of an inhibitory pattern. Instead, in the remaining part of the sample, smaller distances prevail which are a symptom of clustering. Finally, in comparing Figs. 2 and 3(a), one may note that non-significant values of (1) occur in periods of "low" seismic activity.

ST correlation: It is interesting to compare these results with those of existing tests for the analysis of point processes (e.g. Odland, 1988, Chap. 7). The most simple of these calculates the correlation coefficient between space and time distances of each event with respect to the previous ones. Given the distances $Ds_{kh} = [(x_k - x_h)^2 +$

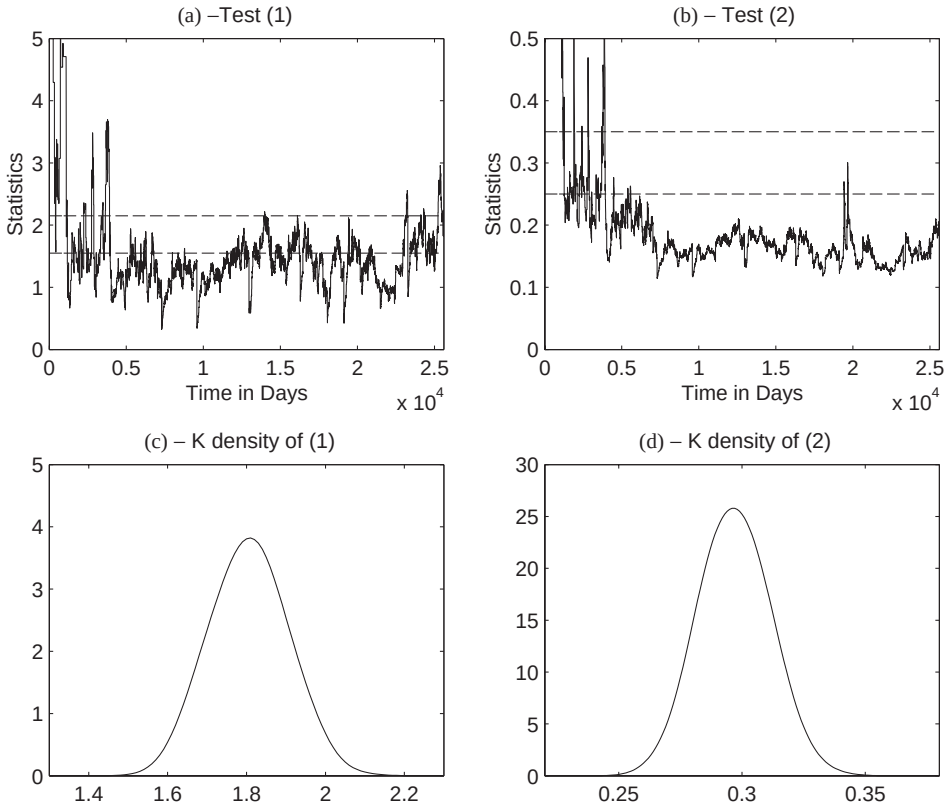

Fig. 3. (a,b) Time path of statistics (1) and (2) with $\lambda = 0.96$, together with 99% critical regions obtained from their simulated distributions under the null (c,d).

$(y_k - y_h)^2]^{1/2}$, $Dt_{kh} = |t_k - t_h|$ the weighted covariance is given by

$$\Gamma_{st} = \frac{2}{(n-1)^2} \sum_{k=2}^{n} \sum_{h:t_h < t_k} Ds_{kh}\, Dt_{kh}\, w_{kh}, \tag{3}$$

where the weights $w_{kh} = Z_k Z_h / \bar{Z}_n^2$ keep account of the magnitude of the events.

Given the large number of observations in the NCEC data-set and the need to check the time-variability of statistics, we applied (3) to sub-samples of size $n = N/70$ and to the data of each year. Numerical results are reported in Fig. 4: as we see, there is evidence for a significant correlation of space and time distances, although results on data per year are more stable.

Knox test: The Knox (1964) approach is used to test for the existence of a significant cluster *within* certain spatial and temporal bounds: Ds^*, Dt^*. First, it counts the number of point pairs in a 2×2 contingency table, then it calculates the P-value under the

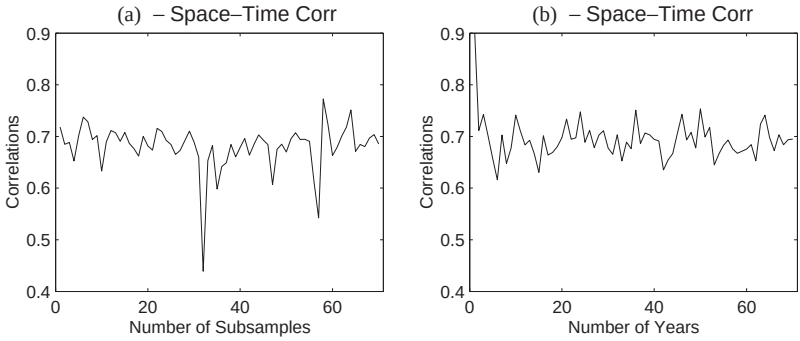

Fig. 4. Space–time correlations computed on sub-samples (a) and data per year (b).

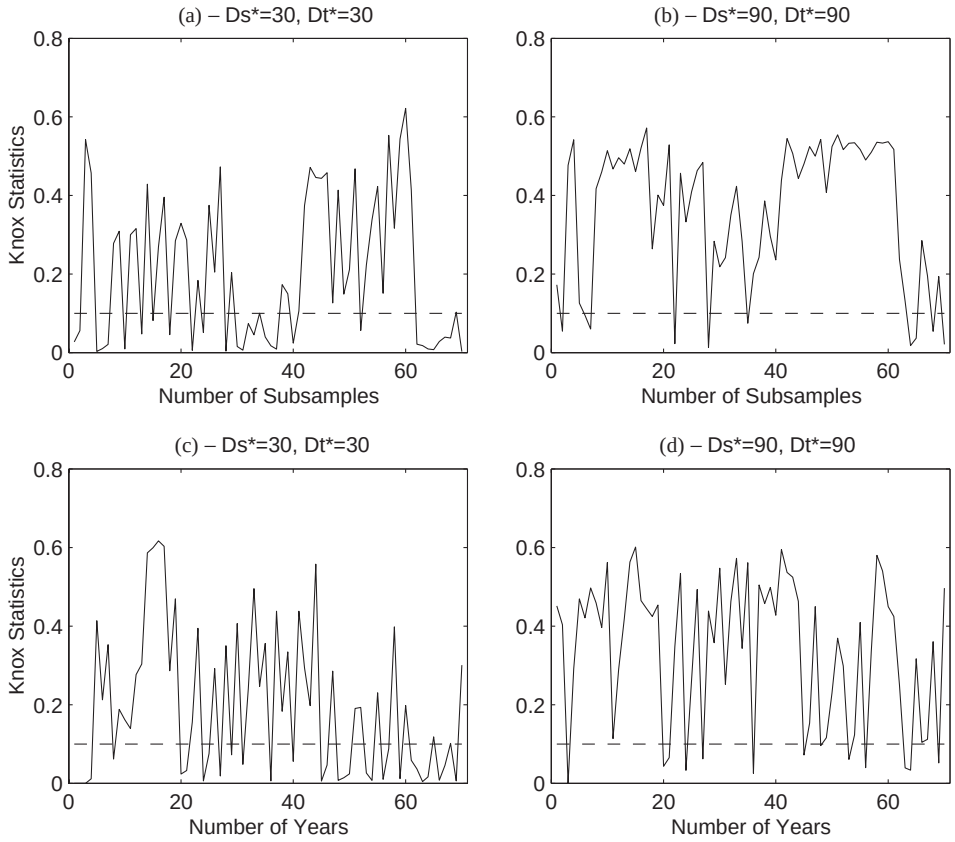

Fig. 5. Time-path of statistic (4) computed on sub-samples (a,b) and data per year (c,d), and for small (a,c) and large (b,d) distances. Ds^* is in km, Dt^* is in number of days and the critical P-value (dotted line) is 10%.

assumption of random pattern as generated by a Poisson model:

$$Pv = \mathrm{P}(x \geqslant n_{11} \mid \lambda_0) = 1 - \sum_{x=0}^{n_{11}-1} (\mathrm{e}^{-\lambda_0} \lambda_0^x / x!),$$

$$\lambda_0 = [(n_{11} + n_{12}) \cdot (n_{11} + n_{21})]/n_0, \tag{4}$$

where n_{11} is the number of pairs in the first cell, i.e. such that $(Ds_{kh} \leqslant Ds^*_{kh}) \cap (Dt_{kh} \leqslant Dt^*)$, and where distances are defined as in (3). The quantities n_{12}, n_{21} are the number of pairs in off-diagonal cells, and n_0 is the total number of pairs in the table. In this approach, it is clear that the null hypothesis of complete random pattern is rejected for *large* values of (4).

As in formulas (1)–(3), we weighted pair counts by the magnitude of the corresponding events; moreover, we applied (4) to sub-samples of size $n = N/70$ and to the data of each year. Results are reported in Fig. 5: as we may see, there is a significant time-variability of statistics, and space–time independence may occur within small bounds: $Ds^* < 30$ km and $Dt^* < 30$ days.

Following Ogata (1998) and Pievatolo and Rotondi (2000), given a main shock of size $4 \leqslant Z \leqslant 6$, the spatio-temporal window which may include events due to aftershock activity (which therefore are dependent), has radius $Ds^* = 50$ km and span $Dt^* = 100$ days. In Fig. 5(a) these limits are considerable lower and independence of events may be encountered. Dropping events which fall within the window, is a complex and controversial issue, discussed in Ogata (1998).

In conclusion, all methods developed in this section confirm existence of a positive relationship between occurrence time and spatial location of earthquakes. The self-exciting theory seems then supported by data and this legitimates further modeling attempts, parametric as well as non-parametric.

4. Adaptive non-parametric smoothers

The statistical properties of a point process are entirely characterized by its joint density $F(\boldsymbol{s}, t, Z)$. The intensity function is defined in conditional form as $f(\boldsymbol{s}, t \mid \mathscr{H}_t)$, where $\mathscr{H}_t = \{(\boldsymbol{s}_k, t_k, Z_k), t_k \leqslant t\}$ is the history up to time t. In the self-exciting approach $f(\cdot)$ is modeled as $\mu(\boldsymbol{s}) + \sum_k \nu(\|\boldsymbol{s} - \boldsymbol{s}_k\|, t - t_k)$, where $\mu(\cdot)$ is the background rate and $\nu(\cdot)$ is the *infectivity* function. Because $\mu(\cdot)$ does not depend on t, it may be modeled as a stationary Poisson process. What distinguishes the various self-exciting approaches is the parametric form of $\nu(\cdot)$, see Ogata (1998).

Given the difficulty in finding a satisfactory expression for $\nu(\cdot)$ and problems in fitting it to data, one may resort to the non-parametric approach for estimating $f(\cdot)$. Main problem is the definition and the treatment of time, which has not received much attention in the literature. For example, Bailey and Gatrell (1995, Chap. 4) and Stock and Smith (2002) applied Kernel smoothing to seismic data as if events occurred at the same instant, obtaining *static* risk maps. Inclusion of time, however, is useful for depicting changes in the pattern of the phenomenon, so as to provide more fresh information.

Density: Kernel density estimators can measure the intensity of a point process in space, namely the expected rate of occurrences per unit area. Adaptation of the Nadaraya-Watson smoother to time-varying situations can be obtained as in (1), by considering the estimator in sequential form and then weighting observations:

$$\hat{f}_t(x,y) = \frac{1}{\kappa_1\kappa_2}\left(\sum_{k:t_k\leqslant t}\lambda^{t-t_k}\right)^{-1}\sum_{k:t_k\leqslant t}\lambda^{(t-t_k)}K_1\left(\frac{x_k-x}{\kappa_1}\right)K_2\left(\frac{y_k-y}{\kappa_2}\right)w_k, \quad (5)$$

where $K_{1,2}(\cdot)$ are the kernel functions and $\kappa_{1,2}$ are their bandwidths. With respect to static smoothers, the weights $\lambda^k/\sum_k \lambda^k$ have the same role as the term $1/N$, whereas the weights w_k keep account of the relative strength of each event. Typical form is $w_k = Z_k/\bar{Z}_t$, where $\bar{Z}_t$ is the sample mean computed at time t. Alternatively, using the adaptive mean $\tilde{Z}_t = \sum_k \lambda^{t-t_k}Z_k/\sum_k \lambda^{t-t_k}$, the estimator (5) becomes

$$\tilde{f}_t(x,y) = \frac{1}{\kappa_1\kappa_2}\left(\sum_{k:t_k\leqslant t}\lambda^{t-t_k}Z_k\right)^{-1}\sum_{k:t_k\leqslant t}(\lambda^{t-t_k}Z_k)\,K_1\left(\frac{x_k-x}{\kappa_1}\right)K_2\left(\frac{y_k-y}{\kappa_2}\right). \quad (6)$$

In the application of non-parametric methods to time-varying problems, there is the tendency to use kernel smoothers even for the time component, namely to replace the term $\lambda^{(t-t_k)}$ with the two-sided filter $K_3[(t_k-t)/\kappa_3]/\kappa_3$ (e.g. Robinson, 1989). This is not correct because it involves future observations, whereas present ones only depend on past events. However, to see the relationship of estimator (5) with a 3-variate smoother, let us consider the approximation $\sum_k \lambda^k \approx 1/(1-\lambda)$, and replace the kernels $K(z)$ with $\lambda^{|z|}/C$, where $C = (1+\lambda)/(1-\lambda)$ is a normalizing constant. With these changes, the estimator (5) becomes

$$\hat{f}(t,x,y) = (1-\lambda_1)\frac{(1-\lambda_2)}{(1+\lambda_2)}\frac{(1-\lambda_3)}{(1+\lambda_3)}\sum_{k:t_k\leqslant t}\lambda_1^{(t-t_k)}\lambda_2^{|x_k-x|}\lambda_3^{|y_k-y|}w_k. \quad (7)$$

Design of smoothing parameters $0<\lambda_2,\lambda_3<1$ is simpler than that of $0<\kappa_1,\kappa_2<\infty$; however, in the multivariate spatial context they induce a marked diamond-shape profile in the density estimate.

Optimal design of the smoothing coefficients $\lambda,\kappa_1,\kappa_2$ in (5)–(7) might be obtained with classical cross-validation methods. Given the unidirectional nature of time, however, a more sensible approach is to focus on the *predictive* performance of kernel estimates. In principle, given the impulsive nature of the density function of a point process, one could refer to the criterion $\sum_k [1-\hat{f}_{k-1}(x=x_k,y=y_k)]$, where $\hat{f}_{k-1}(\cdot)$ is the density estimate based on the first $k-1$ observations. In practice, given a suitable grid for t,x,y we consider a more smoothed loss function

$$V_1(\lambda,\kappa_1,\kappa_2) = \sum_t\sum_x\sum_y [\hat{f}_t(x,y) - h_{t+m}(x,y)]^2, \quad (8)$$

where $h_{t+m}(\cdot)$ is the bivariate histogram of relative frequencies computed on weighted events in the interval $[t+1,t+m]$, with m moderate. Seismologist believe that a value

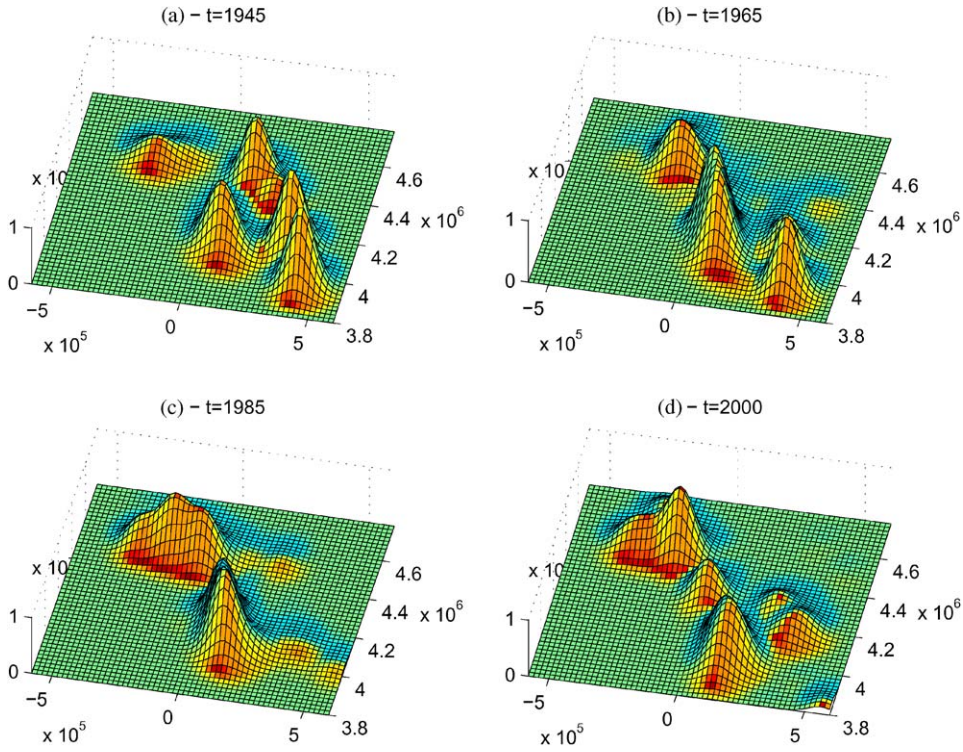

Fig. 6. Kernel estimates of the intensity at four instants of time. They are generated with (6) and $\lambda = 0.97$, $\kappa = 43$ km. Spatial coordinates are in the UTM system; the y-axis passing through $x = 0$ is the meridian of San Francisco.

$m = 100$, namely 3 months, is sufficient to rise independence in significant earthquakes, see Ogata (1998) and Pievatolo and Rotondi (2000).

Minimization of criterion (8) on the NCEC data-set was computationally demanding; for this reason we restricted analysis to a subregion bounded by longitudes $(-128, -116)$ and latitudes (33,43), centered on San Francisco city. Using filter (6) with Gaussian kernels, $\kappa_1 = \kappa_2$, and a $70 \times (60 \times 50)$ regular grid for $t,(x, y)$, we obtained the values $\hat{\lambda} = 0.97$, $\hat{\kappa} = 43$ km. Next, with filter (6) we generated the estimates in Fig. 6. These show a significant space–time variability of the intensity function, although four persistent clusters can be recognized.

Regression: Typical regression analysis for point processes consists of evaluating the dependence of the magnitude on space and time components. This provides the expected value of the magnitude of future events or may measure the diffusion of the strength of past events. As for the intensity function, this is useful for building maps of seismic risk. In this context, we assume a model of the type

$$Z_k = g_k(x_k, y_k) + e_k, \quad e_k \sim \mathrm{IID}\,(0, \sigma^2), \tag{9}$$

where $g_k(\cdot)$ is a time-varying non-linear response function. For this scheme, the kernel estimator corresponding to method (6) is given by

$$\hat{g}_t(x,y) = \frac{\sum_{\{k:t_k \leqslant t\}} \lambda^{(t-t_k)} K_1[(x_k - x)/\kappa_1] K_2[(y_k - y)/\kappa_2] Z_k}{\sum_{\{k:t_k \leqslant t\}} \lambda^{(t-t_k)} K_1[(x_k - x)/\kappa_1] K_2[(y_k - y)/\kappa_2]}. \tag{10}$$

It is useful to compare model (9) with $Z_k = g(x_k, y_k, t_k) + e_k$, that explicitly includes time as a regression variable. Although the two schemes are related, model (9) is more general, because it admits heterogeneity and evolution of the regression function $g(\cdot)$. This is the main reason that motivates the implementation with the forgetting factor λ. It follows that a suitable estimator for $g(x, y, t)$ can be obtained from (10) by replacing the weights $\lambda^{(t-t_k)}$ with the kernel $K_3[(t_k - t)/\kappa_3]$.

Finally, one may consider the general model $Z_k = g_k(x_k, y_k, t_k) + e_k$ and try to estimate it with filter (10), after the inclusion of the kernel $K_3(\cdot)$ in the numerator and the denominator. This component has the same role as the weights $\lambda^{(t-t_k)}$, therefore there should be no substantial difference in the performance of the two filters. Indeed, empirical applications provided a medium difference of about 0.023 on the Richter scale, which is less than 1%. Hence the model is redundant.

Although filter (10) has fundamentally a role of interpolator, it may provide information on the mark of future events at any point. To enhance its forecasting vocation, one should design its smoothing coefficients with a sequential cross-validation criterion. Denoting with $\hat{g}_{k-1}(x, y)$ the estimate (10) based on the first $k - 1$ observations, the loss function may be outlined as follows

$$V_2(\lambda, \kappa_1, \kappa_2) = \sum_{k=2}^{N} [Z_k - \hat{g}_{k-1}(x = x_k, y = y_k)]^2, \tag{11}$$

where $\hat{g}_{k-1}(x_k, y_k)$ provides the prediction of the magnitude at time t_k and location (x_k, y_k), for which the actual value Z_k is available.

Minimization of criterion (11), under the same conditions as Fig. 6, provided the estimates $\hat{\lambda} = 0.94$, $\hat{\kappa} = 65$ km, which are similar to those of the density case. With these values and the filter (10) with Gaussian kernels, we generated the surfaces in Fig. 7 at four instants of time. As a comment, one may note a marked space–time variability of the regression function and a significant dependence of the magnitude on the axis South-East, North-West.

Some interesting remarks can be done in comparing Figs. 6 and 7. For example, intensity of seismic activity usually declines rapidly in space, whereas the same does not hold for the diffusion of strength. Moreover, there are locations of relatively high intensity that are not associated with a comparable large strength (as in the area of San Francisco at the center of the images). This might suggest the existence of an inverse relationship, or at least independence, between the two seismic features. Coherently, the theory of gradual strain-release claims that the larger the number of events, the smaller their magnitude. In designing space–time maps of seismic risk, both intensity and strength functions should be considered; the first one indicates the location of an

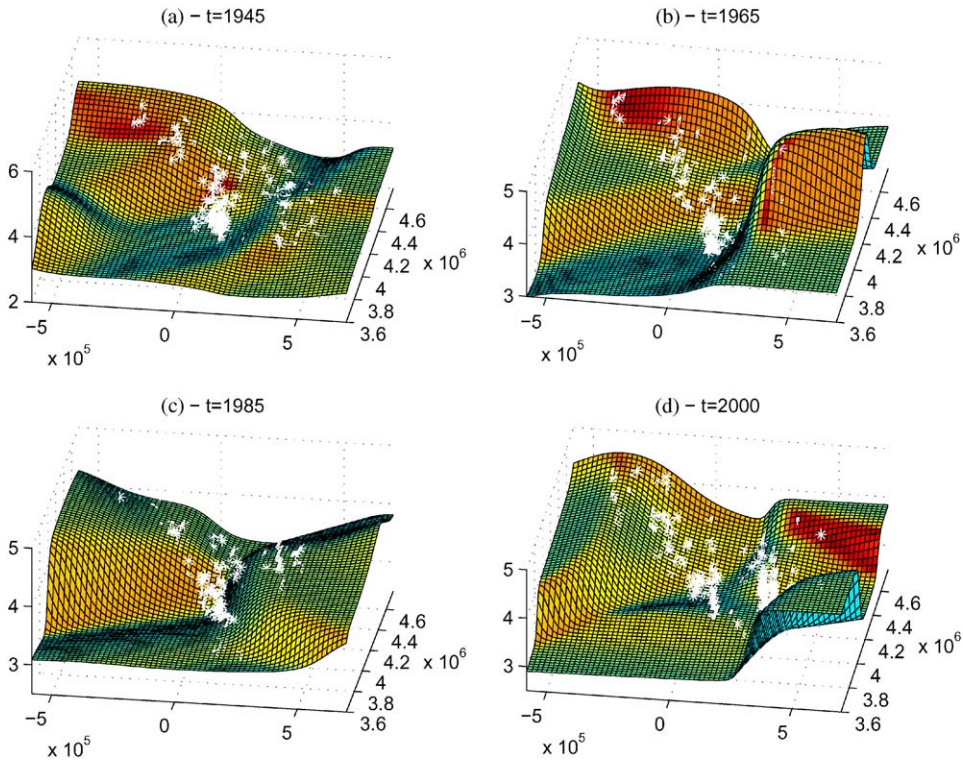

Fig. 7. Regression estimates (10) obtained with $\lambda = 0.94$, $\kappa = 65$ km, at four instants, together with recent observations. View is from South-East to North-West.

earthquake, the second one estimates the associated magnitude. As a result, several classes of seismic sites can be outlined.

5. Special aspects of space–time smoothing

Recursions: The importance of implementing filters (5)–(11) with the weighting sequence $\lambda^{(t-t_k)}$, rather than with a Kernel function, emerges when data are equally spaced in time. For point processes, this regularity condition can be achieved by aggregating data in weekly, monthly or yearly series. More precisely, given a suitable spatio-temporal grid for (x, y, t), seismic events must be classified in the cells of the resulting 3-way array and their strengths cumulated as

$$Z_{xyt} = \sum_{\{k:(t-1<t_k\leqslant t),(x-1<x_k\leqslant x),(y-1<y_k\leqslant y)\}} Z_k.$$

These provide a measure of the total stress that occurs at a given space–time location.

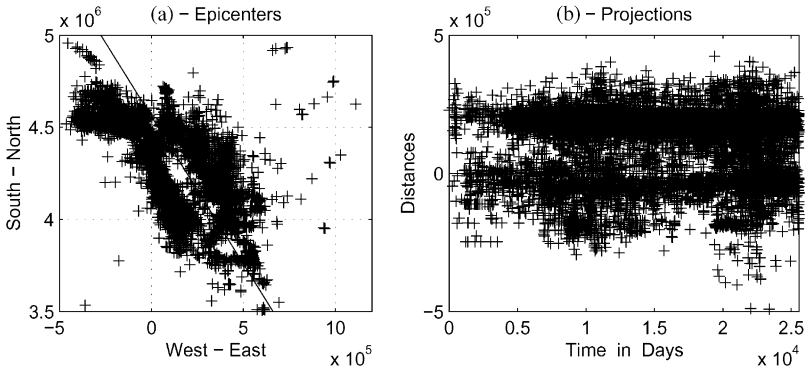

Fig. 8. (a) Epicenters of earthquakes expressed in UTM coordinates together with the axis of projection; (b) Distances of epicenters from the Sacramento axis.

A problem in this approach is that magnitudes on the Richter scale are expressed in logarithms, therefore they are not additive. This problem is not suffered by the energy (E) which is related to the magnitude (M) in the following way: $\log_{10} E = c + 1.5M$. Hence, by means of simple transformations one can build a regular series for the strength variable. Once this goal is achieved, using the approach of Grillenzoni (2000), filters (5) and (10) can be arranged in *recursive* form as

$$\hat{f}_t(x,y) = \lambda \hat{f}_{t-1}(x,y) + \frac{(1-\lambda)}{\kappa_1\kappa_2} \sum_{k:t-1<t_k\leqslant t} K_1\left(\frac{x_k - x}{\kappa_1}\right) K_2\left(\frac{y_k - y}{\kappa_2}\right) w_k,$$

$$\hat{\varepsilon}_t(x,y) = [Z_{xyt} - \hat{g}_{t-1}(x,y)],$$

$$\hat{g}_t(x,y) = \hat{g}_{t-1}(x,y) + \frac{(1-\lambda)}{\kappa_1\kappa_2 \hat{f}_t(x,y)} \sum_{k:t-1<t_k\leqslant t} K_1\left(\frac{x_k - x}{\kappa_1}\right) K_2\left(\frac{y_k - y}{\kappa_2}\right) \times \hat{\varepsilon}_t(x,y), \tag{12}$$

where $\hat{\varepsilon}_t(x,y)$ is the prediction error of total strength at location (x,y).

Computational advantages of this framework are apparent, and the role of the forgetting factor λ in tracking non-stationary patterns is enhanced. A result that cannot be achieved by using a kernel function for the time component. Prediction errors have special importance because can be used in designing smoothing parameters, by minimizing the loss function

$$V_3(\lambda,\kappa_1,\kappa_2) = \sum_t \sum_y \sum_x [\hat{\varepsilon}_t(x,y)]^2. \tag{13}$$

This is similar to a sequential cross-validation criterion, where the omitted observation is always the present one.

Projections: With frames such as those in Figs. 6 and 7, one can build dynamic graphs (movies) for viewing the entire evolution of earthquakes. To have a synthetic

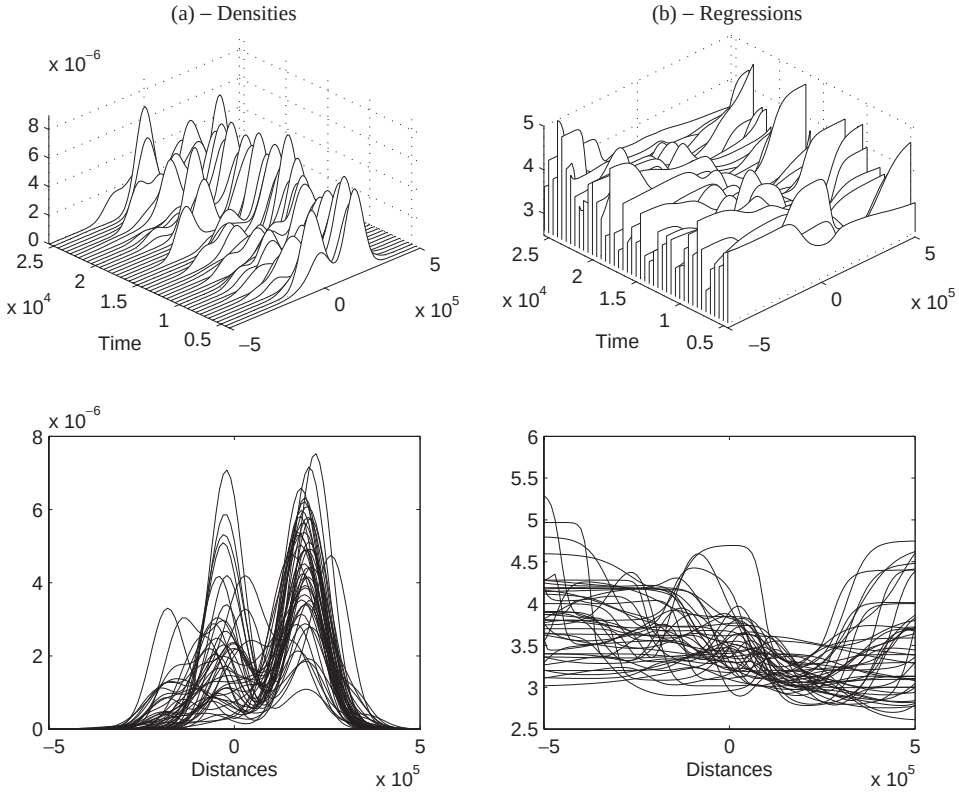

Fig. 9. Plot of density (a) and regression (b) functions of earthquake data projected on the Sacramento axis. Estimates are performed every 500 days, and the view is from North-East to South-West.

representation of the phenomenon, one may work with projections of the epicenters on a suitable axis $\aleph$, as suggested in Ogata (1998). In practice, one can substitute the spatial coordinates (x_k, y_k) with the Euclidean distance $p_k = d(P_k, \aleph)$ of any epicenter P_k from the defined axis.

In our application we selected the axis which runs in the middle of the Sacramento Valley in direction NW–SE, see Fig. 8(a). As a convention, we attributed negative sign to the coordinates of epicenters placed on the right of $\aleph$, see Fig. 8(b). This is the contrary of that adopted by the UTM system, which gives negative sign to the x-coordinate of the points placed on the left of San Francisco meridian.

Smoothers for data represented in this form can be obtained from (5)–(10) by replacing the functions $K_1(\cdot), K_2(\cdot)$ with $K[(p_k - p)/\kappa]$. Using a time interval of 500 days and the cross-validation estimates $\hat{\lambda} = 0.95$, $\hat{\kappa} = 45$ km, we obtained the densities and the regressions in Fig. 9. Both estimates confirm that Sacramento axis is a watershed between two distinct seismic zones, the most active of which is on the Pacific Coast. In

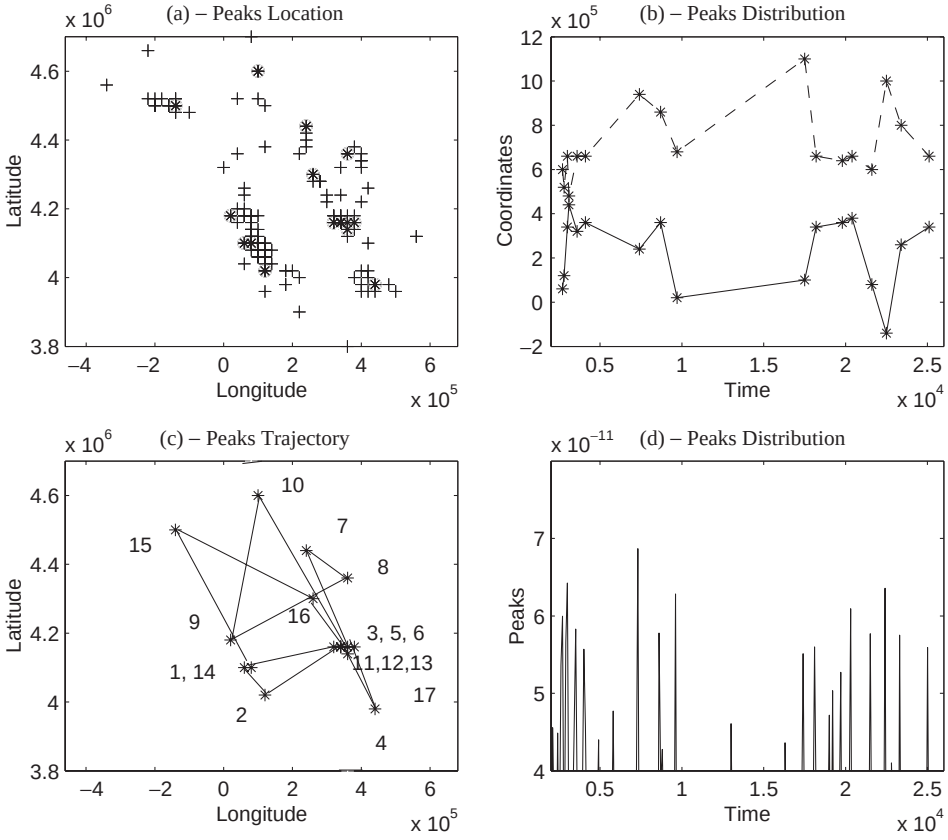

Fig. 10. Space–time evolution of peak intensity: (a) location of the peaks of $\hat{f}_t(x, y)$ estimated every year (+), (*) denotes most significant peaks; (b) time evolution of the coordinates of significant peaks; (c) spatial migration of significant peaks; (d) temporal distribution and size of significant peaks.

particular, Fig. 9(b) shows a sort of "alternation" in the peaks of the functions $\hat{f}_t(p)$, $\hat{g}_t(p)$ of two groups as time goes on.

Peaks migration: Monitoring the space–time evolution of peak intensity of earthquakes may be useful to understand the dynamics of tectonic faults. In practice, this may be accomplished through the pattern analysis of the modal values of the kernel densities $\hat{f}_t(x, y,)$. Choi and Hall (2000) have provided statistical conditions that must be fulfilled for such estimation; in particular, these require a sort of square integrability of the underlying probability functions.

In our application densities were estimated every 365 days, and their modes were plotted in Fig. 10(a). Obviously, it is sensible to investigate the space–time trajectory only for the most significant ones. Now, given the threshold $\psi = 5/10^{11}$, only 17 peaks exceeded this value and their locations are plotted in Fig. 10(b–d). Fig. 10(c), in

particular, shows that migration mainly follows the NW–SE direction (i.e. the Sacramento axis), and three distinct seismic sites can be identified.

From Fig. 10(d), which plots peak values against time, one may identify *quiescent* periods. These arise when seismic activity goes below a given threshold intensity ψ. In this investigation, also strength/magnitude should be taken in account, but estimators (5)–(7) actually do so through the weights w_k. With respect to $\psi = 5/10^{11}$, the longest quiescent period occurred from 1957 to 1977.

Forecasting: Smoothing methods developed in the previous section are useful for building maps of seismic risk. While the density (6) provides information on the location of events, the regression function (10) estimates their magnitude. Therefore, assuming quasi-stationarity of the point process, prediction of future events may proceed in two steps: first find $(x^*, y^*) = \arg\max \hat{f}_t(x, y)$ from (6); then calculate $Z^* = \hat{g}_t(x^*, y^*)$ from (10). Prediction of the occurrence time t^* can be obtained from an autoregressive model on intervent periods r_k:

$$r_k = (t_k - t_{k-1}) = h_k(r_{k-1}) + u_k, \quad u_k \sim \text{IID}\,(0, \sigma_u^2).$$

This can be estimated non-parametrically by the filter

$$\hat{h}_k(r) = \left[\sum_{i=2}^{k} \lambda^{(k-i)} K\left(\frac{r_{i-1} - r}{\kappa}\right)\right]^{-1} \sum_{i=2}^{k} \lambda^{(k-i)} K\left(\frac{r_{i-1} - r}{\kappa}\right) r_i \tag{14}$$

and the prediction becomes $t^* = t_k + \hat{h}_k(r = r_k)$. Notice that purely interpolating models, such as $r_k = h(k) + u_k$, are not useful in forecasting.

The last remark suggests that a general approach to forecasting should be based on autoregressive models for every component of the seismic phenomena:

$$\begin{aligned} Z_k &= g_k^Z(Z_{k-1}) + e_k^Z, \\ x_k &= g_k^x(x_{k-1}, y_{k-1}) + e_k^x, \\ y_k &= g_k^y(y_{k-1}, x_{k-1}) + e_k^y, \\ t_k &= g_k^t(t_{k-1}) + e_k^t, \end{aligned} \tag{15}$$

where for latitude and longitude we have assumed existence of dependence. This could be assumed also for other components, but for spatial coordinates this is a crucial point, in view of the issue of peaks migration.

Notice that each equation of system (15) can be estimated with filter (14). This, in turn, can be written in recursive form (12) because index k provides a regular sequence. In this case, the optimization criterion (13) can be expressed in terms of the prediction errors $\hat{\varepsilon}_k^v(\lambda, \kappa) = [v_k - \hat{g}_k^v(v = v_{k-1})]$, for any component $v = Z, x, y, t$. Analysis of statistical properties can be found in Grillenzoni (2000).

6. Conclusions

This paper has developed non-parametric techniques for analyzing space–time point processes. A major advantage is that they do not require specification of structural

models, which are suitable only when data come from homogeneous seismic zones. On the contrary, data often concern wide geographical areas and long periods of time, where there are changes in spatial location of events and variability in the rate of occurrences. In these cases, parametric models do not fit well, which can be reflected by parameter estimates that lie outside admissible ranges. Non-parametric smoothing methods have been designed as exploratory tools of data analysis; however, they have great flexibility and adaptability. In this paper we have developed sequential and recursive solutions which are suitable in the case of non-stationarity point processes. We have proved their efficacy in building dynamic risk maps and we have outlined their potential use in forecasting.

References

Bailey, T.C., Gatrell, A.C., 1995. Interactive Spatial Data Analysis. Longman, Essex, UK.

Bolt, B., Schoenberg, F., 2000. Short-term exciting, long-term correcting models for earthquake catalogs. Bull. Seismol. Soc. Amer. 90, 1–13.

Choi, E., Hall, P., 2000. On the estimation of poles in intensity functions. Biometrika 87, 251–263.

Grillenzoni, C., 2000. Nonparametric regression for nonstationary processes. J. Nonparametric Statist. 12, 265–282.

Hobert, J.P., Altman, N.S., Schofield, C.L., 1997. Spatial analysis of the fish species richness of Adirondack lakes: applications of geostatistics and nonparametric regression. J. Amer. Statist. Assoc. 92, 846–854.

Knox, E.G., 1964. The detection of space–time interactions. Appl. Statist. 13, 25–29.

Odland, J., 1988. Spatial Autocorrelation. Sage, Beverly Hills, CA.

Ogata, Y., 1998. Space–time point-process models for earthquakes occurrences. Ann. Inst. Statist. Math. 50, 379–402.

Pievatolo, A., Rotondi, R., 2000. Analyzing the intervent time to identify seismicity phases: a Bayesian approach to the multiple-changepoint problem. Appl. Statist. 49, 543–562.

Rathbun, S.L., 1993. Modeling marked spatio-temporal point patterns. Bull. Internat. Statist. Inst. 55, 379–396.

Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change. Springer, Berlin, pp. 253–264.

Schoenberg, F.P., Brillinger, D.R., Guttorp, P., 2002. Point processes, spatial-temporal. In: El-Shaarawi, A., Piegorsch, W. (Eds.), Encyclopedia of Environmetrics, Vol. 3. Wiley, New York, pp. 1573–1577.

Stock, C., Smith, E., 2002. Comparison between seismicity models generated by different kernel estimations. Bull. Seismol. Soc. Amer. 92, 913–922.

Vere-Jones, D., 1992. Statistical methods for the description and display of earthquake catalogs. In: Walden, A., Guttorp, P. (Eds.), Statistics in the Environmental and Earth Sciences. Arnold, London, pp. 220–246.

Yakowitz, S.J., Szidarovszky, F., 1985. A comparison of Kriging with nonparametric regression methods. J. Multivariate Anal. 16, 21–53.

Zhuang, J., Ogata, Y., Vere-Jones, D., 2002. Stochastic declustering of space–time earthquake occurrences. J. Amer. Statist. Assoc. 97, 369–380.