JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 106, NO. B5, PAGES 8699-8711, MAY 10, 2001

A simple and testable model for earthquake clustering

Rodolfo Console and Maura Murru

Istituto Nazionale di Geofisica, Rome, Italy

Abstract. Earthquakes are regarded as the realization of a point process modeled by a generalized Poisson distribution. We assume that the Gutenberg-Richter law describes the magnitude distribution of all the earthquakes in a sample, with a constant *b* value. We model the occurrence rate density of earthquakes in space and time as the sum of two terms, one representing the independent, or spontaneous, activity and the other representing the activity induced by previous earthquakes. The first term depends only on space and is modeled by a continuous function of the geometrical coordinates, obtained by smoothing the discrete distribution of the past instrumental seismicity. The second term also depends on time, and it is factorized in two terms that depend on the space distance (according to an isotropic normal distribution) and on the time difference (according to the generalized Omori law), respectively, from the past earthquakes. Knowing the expected rate density, the likelihood of any realization of the process (actually represented by an earthquake catalog) can be computed straightforwardly. This algorithm was used in two ways: (1) during the learning phase, for the maximum likelihood estimate of the few free parameters of the model, and (2) for hypothesis testing. For the learning phase we used the catalog of Italian seismicity ($M \geq 3.5$) from May 1976 to December 1998. The model was tested on a new and independent data set (January-December 1999). We demonstrated for this short time period that in the Italian region this time-dependent model has a significantly better performance than a stationary Poisson model, even if its likelihood is computed excluding the obvious component of main shock-aftershock interaction.

1. Introduction

The current debates concerning predictability of earthquakes clearly show how this problem is centered on the difficulty of systematically testing the numerous methodologies that over the years have been proposed and sustained by the supporters of prediction [*Geller*, 1996; *Evans*, 1997]. This difficulty starts, sometimes, from the lack of a quantitative and rigorous definition of the precursor concerned and, other times, from the lack of a large enough number of cases upon which to base statistical analyses.

Most earthquake precursors have been observed occasionally and described anecdotally as single cases, or even if they are based on a sufficiently long series of experimental observations, they refer often to past situations. These situations constitute the sort of "retrospective predictions" that are subject to criticism because of the risk of being conditioned by the unconscious trend of their own authors, who want to get some gratifying results from their investigations.

A hypothesis for earthquake forecasting can be objectively tested if it is specified in terms of the conditional occurrence probability of seismic events in a given magnitude and time range and in well-defined geographical areas [*Kagan and Jackson*, 1995; *Jackson*, 1996; *Jackson and Kagan*, 1999]. The considerations based on the subdivision of the space-time-magnitude volume in regions have the drawback that the occurrence probabilities can differ considerably from one region to the bordering one. So the success or the failure of a hypothesis can strongly depend on small errors in the estimation of location and magnitude.

The statistical procedures should take into account the uncertainty in the measurements and the imperfections in the definition of the hypothesis. One way of facing this problem is by considering smaller regions and introducing a smoothing of the probability variations from one region to another. This leads to formulating the hypothesis of occurrence in terms of continuous variables, as function of the distance (in space and time) from earlier seismic events, or of the localization of preceding geophysical anomalies [*Jackson*, 1996]. Regardless of the method and of the data from which the hypothesis is built up, it is generally accepted that its test should be carried out on new and independent data, without any further adjustment of the model parameters [*Rhoades and Evison*, 1989].

Once the earthquake forecast hypothesis (or model) allows the estimation of the expected rate density of seismic events, or the probability of occurrence in location-time-magnitude volumes of finite dimensions, the likelihood of observing a given realization of the process can be computed. A Bayesan approach to the validation of a specific hypothesis consists in evaluating the ratio between the likelihood, computed respectively under the hypothesis to be tested and under a simpler hypothesis of reference, for the same set of observations. A model for earthquake forecasting is acceptable if its likelihood is consistently larger than that coming from a random model (known as a Poisson or null hypothesis) in which the earthquakes occur at the rate determined from the past behavior [*Rhoades and Evison*, 1997]. The ratio between the likelihood of two models (called performance factor of one of the models with respect to the other), being multiplied by the apriori probability that this model is true, gives the new probability that the model is true in light of the observations carried out in the test. The statistical significance of the

Paper number 2000JB900269.
0148-0227/01/2000JB900269$09.00

performance factor can be estimated by means of simulations obtained by the Monte Carlo method [*Rhoades and Evison*, 1997]. However, the creation of hundreds of synthetic catalogs, as would be necessary for obtaining reliable statistical results, and the subsequent computation of the relative likelihood ratios may not be a trivial task for a computer of usual capacity.

There are two well-known features that can be regarded as a deviation of seismicity from a uniform Poisson process: short term-clustering and long-term pseudo-periodicity. These features are reported as possible evidence of a time-dependent rate in earthquake occurrence, but they represent opposite trends with respect to the Poisson constant rate distribution.

In this paper, we deal only with short-term clustering, i.e. the kind of phenomena popularly classified under the denominations of foreshocks, aftershocks, or swarm activity, whose typical duration ranges from few days to a few months. We formulate a simple stochastic model for earthquakes induced by previous earthquakes, based on 23 years of data collected by the Italian Seismological Network. Then we test our model against a Poissonian null hypothesis on a new set of data covering 1 year of observations. In this way the requirement that the hypothesis is formulated before the occurrence of the earthquakes, on which it is tested, is fulfilled.

2. Poisson Model

Earthquakes are regarded as the realization of a point process modeled by a generalized Poisson distribution. Each event is characterized by its location-time-magnitude parameters (x,y,t,m).
The definition of occurrence rate density of the space-time process used here is [*Ogata*, 1998]

$$\lambda(x,y,t,m) = \lim_{\Delta x,\Delta y,\Delta t,\Delta m \to 0} \frac{P_{\Delta x,\Delta y,\Delta t,\Delta m}(x,y,t,m)}{\Delta x \Delta y \Delta t \Delta m}, \tag{1}$$

where Δx, Δy, Δt, and Δm are increments of longitude, latitude, time, and magnitude and $P_{\Delta x\,x\Delta y,\,\Delta t,\,\Delta m}(x,y,t,m)$ is the probability of occurrence of an event in the volume $\{x, x+\Delta x; y, y+\Delta y; t, t+\Delta t; m, m+\Delta m\}$. The dependence of $\lambda(x,y,t,m)$ on the geographical coordinates x and y has to be found in the seismogenic features of each region, and can be obtained from the analysis of the past seismic activity.

With regard to the time dependence, the simplest hypothesis (i.e., a stationary Poisson model) assumes that the point process does not have memory, so any particular event is regarded as unrelated to any other. Consequently, the probability of occurrence of future events is constant irrespectively of the past activity.

For the magnitude distribution we assume the validity of the Gutenberg-Richter law:

$$\mu(x,y,m) = e^{\alpha(x,y)-\beta(x,y)\,m}, \tag{2}$$

where $\mu(x,y,m)$ is the space density of earthquakes of magnitude equal to or larger than m; α and β are characteristic parameters of each seismogenic area, supposed to be approximately independent of time and linked to the well-known a and b parameters by the relations $\alpha = a \ln(10)$ and $\beta = b \ln(10)$, respectively. The maximum likelihood estimate of β, computed from the observation of a set of earthquakes, is [*Utsu*, 1971]

$$\beta = \frac{1}{m_a - m_0}, \tag{3}$$

where m_a is the average magnitude and m_0 is the minimum magnitude of completeness of the data set. In order to avoid the Gutenberg-Richter law, taken without any restriction, yielding finite probability for events of very large magnitude, the distribution could eventually be tapered above an appropriate corner magnitude (see, e.g., *Jackson and Kagan*, 1999).

If β is assumed independent of the space coordinates (while α is allowed to change from point to point), we obtain the decomposition

$$\mu(x,y,m) = \mu_0(x,y)\, e^{-\beta(m-m_0)}, \tag{4}$$

where $\mu_0(x,y) = (x,y,m_0) = e^{\alpha(x,y)-\beta(x,y)m_0}$ is the space density of earthquakes of magnitude equal to or larger than m_0. From (2) or (4), we can compute the space density of earthquakes of magnitude between m_0 and m (increasing with m), and by derivation with respect to m, the density magnitude distribution of earthquakes is obtained:

$$\lambda(x,y,m) = \mu_0(x,y)\, \beta\, e^{-\beta(m-m_0)}. \tag{5}$$

In both (4) and (5) the choice of m_0 is not critical, provided that the set of data is complete above it. In the computation of (5) we prefer to regard the space-density distribution $\mu_0(x,y)$ as a continuous, smooth function of the geographical coordinates (x,y). In the present work, based entirely on instrumental catalogs of the past seismicity, we ignore additional information such as intensity observations of historical earthquakes or surface faulting from prehistoric earthquakes.

The $\mu_0(x,y)$ distribution is obtained by a Gaussian smoothing of the catalog, following the method presented by *Frankel* (1995). First, the region to be analyzed is subdivided in square cells of appropriate size. The number of earthquakes with magnitude greater than m_0, N_k, in each cell, k, is computed. The grid of N_k values is then spatially smoothed by multiplying these values by a Gaussian function with correlation distance d. For each cell k, the smoothed value N_k is obtained from

$$N'_k = \frac{\Sigma_l N_k \exp(-\Delta_{kl}^2 / d^2)}{\Sigma_l \exp(-\Delta_{kl}^2 / d^2)}, \tag{6}$$

where Δ_{kl} is the distance between the centers of the kth and lth cells. In (6), N'_k is normalized to preserve the total number of events. The values of N'_k obtained so far still change abruptly from one cell to another. In order to obtain a continuous function, the single value of $\mu_0(x,y)$ is computed by interpolation between the four cells whose centers surround the point (x,y). The result is given as number of events per cell size in the total time spanned by the catalog. In order to have it in units of number of events per unit time and per unit area, it must be divided by the total time of the input catalog and the area of a cell.

3. Model for Earthquake Clustering

It is widely recognized that the occurrence of a seismic event has influence on the probability of the occurrence of new ones. In a relatively short space-time range this probability can be increased by a factor of hundreds or even thousands of times [*Di Luccio et al.*, 1997]. Of course, the probability of a shock of a given magnitude to be followed by a larger one is smaller than the probability of a shock of the same magnitude to be followed by a smaller event. If the second event of a pair has a size smaller than that of the first one, we speak of a main shock-aftershock pair. If the second event is larger than the first one, we speak of a foreshock-main shock pair. Until now, no method has been discovered for the *a priori* recognition of an earthquake as foreshock, main shock, or single shock, until we can observe the occurrence or nonoccurrence of the following seismic activity. As this terminology is based on empirical definitions, we shall try to avoid its use in the formulation of our clustering model.

The earthquakes that induce further seismicity can be a single "main shock" (in the usual subjective meaning), as in the case of a typical "aftershock" sequence, or more than one "main shock", as in the case of a multiple sequence, or even all the previous events, as in the epidemic model (ETAS) introduced by *Ogata* [1983, 1998]. In this paper we assume the latter case as our hypothesis. Hence the expected resultant rate density of seismic events, taking into account the influence of the previous inducing earthquakes, can be written as

$$\lambda(x, y, t, m) = f_r \lambda_0 (x, y, m) + \sum_{i=1}^{N} H(t - t_i) \lambda_i (x, y, t, m), \tag{7}$$

where f_r is a factor called failure rate (i.e., the ratio between the expected number of independent events and the total number of events) and $\lambda_0(x,y,m)$ represents the background seismicity expressed as in (5). The parameter t_i is the occurrence time of the earthquakes, the total number of which is N; $H(t)$ is the step function; and $\lambda_i(x,y,t,m)$ is the single contribution of the previous earthquakes.

The first and the second terms on the right-hand side of (7) represent the "independent" and the "induced" seismicity, respectively, in probabilistic terms. The rate density corresponding to any earthquake is, in general, constituted by the superposition of both of these two components. In this way, no earthquake is claimed to be fully linked to any other earthquake in particular but rather to all previous events and to the background seismicity, with different weights.

We hypothesize that the contribution of any previous earthquake (x_i,y_i,t_i,m_i) to the occurrence rate density of the subsequent earthquakes is decomposable (for $t>t_i$) into three terms representing the time, magnitude and space distribution, respectively, as:

$$\lambda_i (x, y, t, m) = K h(t - t_i) \beta e^{-\beta(m - m_i)} f(x - x_i, y - y_i), \tag{8}$$

where K is a constant parameter, while $h(t)$ and $f(x,y)$ are the time and space distributions, respectively.

For the time dependence we are adopting in a general way the so-called modified Omori law [*Ogata*, 1983] that is to be applied to any of the subsequent earthquakes with respect to all the previous ones:

$$h(t) = (p-1)\, c^{(p-1)}\, (t+c)^{-p} \qquad (p>1), \tag{9}$$

where c and p are characteristic parameters of the process. It can be recognized that in this formulation we have introduced a normalizing factor, so that it integrates to 1 in time.

As to the dependence on the magnitude m_i of the previous earthquakes, in (8) we adopted the *Reasenberg and Jones* [1989] hypothesis, for which the occurrence rate of subsequent events depends on the magnitude difference with respect to the previous ones. In this respect, the larger number of earthquakes with $m < m_i$ than those with $m > m_i$ reflects the larger probability of main shocks to be followed by main shocks than that of foreshocks to be followed by main shocks.

We model the spatial distribution of the induced seismicity by a function $f(x\text{-}xi,y\text{-}yi)$ that has circular symmetry around the point of coordinates (x_i,y_i) and is normalized to 1, so that K assumes the physical meaning of the expected number of induced events of magnitude equal to or larger than m_i. In polar coordinates (r,θ) this function can be written as

$$f(r,\theta) = \frac{1}{2\pi\sigma^2} e^{-r^2/2\sigma^2}, \tag{10}$$

where r is the distance from the point (x_i,y_i) and σ is the standard deviation in space distribution. *Ogata* [1998] considered different forms of the spatial term and made a statistical comparison of their performance based on the Akaike information criterion. For more complicated forms, the space and time distributions can even be combined, leading to a diffusion function as suggested by *Musmeci and Vere-Jones* [1992].

The decomposition given in (8) implies several important hypotheses. First, we assume that β, K, c and p are independent of the geographical coordinates of the induced (x,y) and the previous event (x_i,y_i). This is, of course, an initial approximation that can be refined if there is enough experience for formulating a more complex situation. Then it must be noted that putting the magnitude m_i of the inducing events in the same term of the magnitude m of the induced events and with the same coefficient β is a strong approximation, which implies the self-similarity of the process over all the ranges of magnitude. With this assumption the probability of an earthquake of magnitude 3.0 after an earthquake of magnitude 4.0 is the same as that of an earthquake of magnitude 5.0 after an earthquake of magnitude 6.0 and so on [*Reasenberg and Jones*, 1989]. It can be easily proven that this assumption leads to instability of the process if the magnitude distribution is not tapered at the tail. Nevertheless, it can be fitted to data for any finite time period, and we use it because of its simplicity, in the context of the parsimony of hypotheses that is a rule all through this work.

Finally, modeling the spatial distribution $f(x,y)$ independently of the magnitude of the inducing earthquake is another strong approximation.

4. Method of Work

The four parameters K, c, p and σ, introduced in section 3, together with the two less critical free parameters β and d, introduced in the context of the Poisson model, represent the set of parameters to be adjusted in order to reach the best fit of the model with real observations. It should be noted that f_r is not a free parameter of the model, as far as its value is constrained by the requirement that the total number of events expected for a particular set of parameters must be equal to the total number of events observed in reality. In this way, during the learning phase the value of f_r is conditioned by the value of all the other parameters and by the value of K in particular. Given a realization of seismic events described by a catalog $\{x_j,y_j,t_j,m_j, j=1,\ldots,N\}$, the correspondent likelihood is computed straightforwardly by the following equation [see, e.g., *Ogata*, 1998]:

$$\ln L = \sum_{j=1}^{N} \ln[\lambda(x_j, y_j, t_j, m_j) V_0] - \int_X \int_Y \int_T \int_M \lambda(x, y, t, m)\, dx\,dy\,dt\,dm, \tag{11}$$

where $V_0 = \int_X \int_Y \int_T \int_M dx\,dy\,dt\,dm$ is a dimensional coefficient equal to the total location-time-magnitude volume spanned by the catalog, so as to give a dimensionless quantity under the logarithm. V_0 multiplied by $\lambda(x_j,y_j,t_j,m_j)$ gives the number of earthquakes that would be counted if the rate density were everywhere and at any time the same as that estimated for the location, time, and magnitude of the jth earthquake. The four-dimension integral at the second member of (11) gives the expected total number of earthquakes under the present hypothesis. Because of the arbitrary choice of V_0 the value of $\ln L$ given by (11) does not have an absolute meaning. It will be useful, however, for the comparison of two models, as explained in the introduction.

The log-likelihood is computed by substitution of (5) and (7) in (11), making use of (8), (9), and (10). The details of this computation are developed in Appendix A. Here we recall that in our algorithms we introduced the approximation that the integral of (11) over the geographical area of study is always equal to, irrespective of how close are the epicentral coordinates of the earthquake concerned to the borders of the area. This is justified by the small value of σ (of the order of 5 *km*) that we use in practice. A more rigorous approach is used by *Ogata* [1998].

In order to implement (6), one has to choose d: the larger d is, the smoother N_k and $\mu_0(x,y)$ will be but also the larger the contamination between different areas will be. In this work, the optimal choice of d is made by trial and error. The catalog is divided into two subcatalogs, each of which spans roughly half of the total time length. Here d is modified until the maximum likelihood is obtained for one of the subcatalogs, using the smoothed seismicity derived from the other subcatalog. Here the seismicity is modeled as a pure Poisson process (K=0), so that no information about the parameters characterizing the induced seismicity is necessary at this stage.

Once a suitable value of d is chosen, the best fit of the whole set of five parameters (β, σ, k, c, and p) is performed on the basis of statistical criteria, changing the value of each parameter until the maximum likelihood of a given realization of the process is obtained. Any method of optimization can be used, provided that it is fast enough to converge to the final solution within a reasonable number of iterations. We found it convenient to make use of our own algorithm based on the Newtonian method of linear parameters estimation, specifically implemented for this problem.

5. Application to Italian Seismicity

The method outlined above has been implemented on the Italian seismological catalog collected by the Istituto Nazionale di Geofisica (ING) from May 27, 1976, to December 31, 1998. We considered the earthquakes contained in a polygonal area surrounding the Italian coasts and borders, within which the National Seismological Network has provided reliable locations for nearly all the events of magnitude 3.5 and larger. Before using the data contained in the catalog, we carried out an analysis of homogeneity, and having found evidence that changes occurred in the magnitude scale, we proceeded with evaluating the significant background rate changes by the magnitude signature algorithm [*Habermann*, 1983]. These changes were corrected by applying suitable magnitude shifts at the proper times within the time period spanned by the catalog [*Wyss et al.*, 1997]. After such magnitude correction, and without carrying out any declustering, the catalog contained 2485 events of magnitude ≥ 3.5 and depth shallower than 70 km (Figure 1). This catalog is characterized by a maximum likelihood value of 0.98 for the b parameter of the Gutenberg-Richter frequency-magnitude relation.

In order to implement the smoothing procedure described in section 4, we covered the area under study with a grid of 100x120 cells, each having a size of 10x10 km, centered on the point 42°N, 13°E.

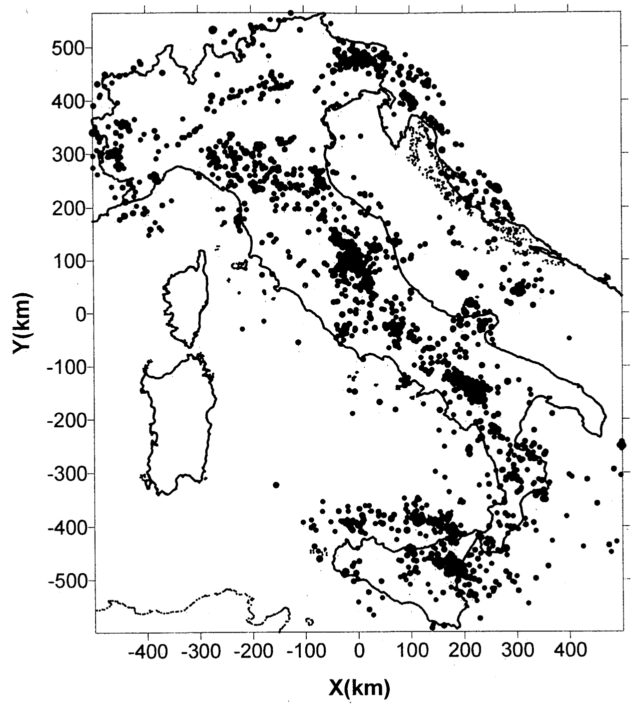


Figure 1. Epicenters of the earthquakes (M≥3.5 and h≤70 km) located by the National Seismological Network in Italy and surrounding areas from May 1976 to December 1998. The origin of the rectangular coordinates is the point 42°N, 13°E.

We started the analysis with the assessment of the value of the parameter d to be used in (6). For doing that, we used the method of comparing the seismicity distribution of two different periods of time, as described in section 4. The periods from May 27, 1976, to June 27, 1986, and from June 28, 1986, to December 31, 1998, were chosen as they contain roughly the same number of earthquakes. The respective contours of the smoothed seismicity obtained using a fixed value of d (26.5 km) are shown in Plate 1. The comparison between Plate 1a and 1b provides evidence of a fairly good similarity in the geographical distribution of the seismicity in the two periods. This supports the hypothesis of the stability in the process of seismic release at this time scale. We are using this hypothesis in modeling the independent component of the seismic activity. As shown in Figure 2, the analysis, carried out using the smoothed seismicity obtained from the former half of the catalog over the latest one, and vice-versa, provided the two most likely values of d in the 25-27 km range. So, we adopted 26 km as the value of d in the following procedures. Plate 2 shows the contours of the smoothed seismicity of the Italian territory obtained using this value of d. The geographical distribution of the expected occurrence rate density so obtained, which clearly merges the features shown in Figures 2a and 2b, represents the model for the time-independent seismicity used in the present work.

The b (or β) parameter is the only one having influence on both the Poisson and the clustering models. We decided to keep its value fixed (b=0.98), as it was obtained by the analysis with the time-independent model. The maximum likelihood estimation of the four parameters that are really significant in the time-dependent model of induced seismicity was performed by the Newtonian estimation procedure mentioned above. The proce-

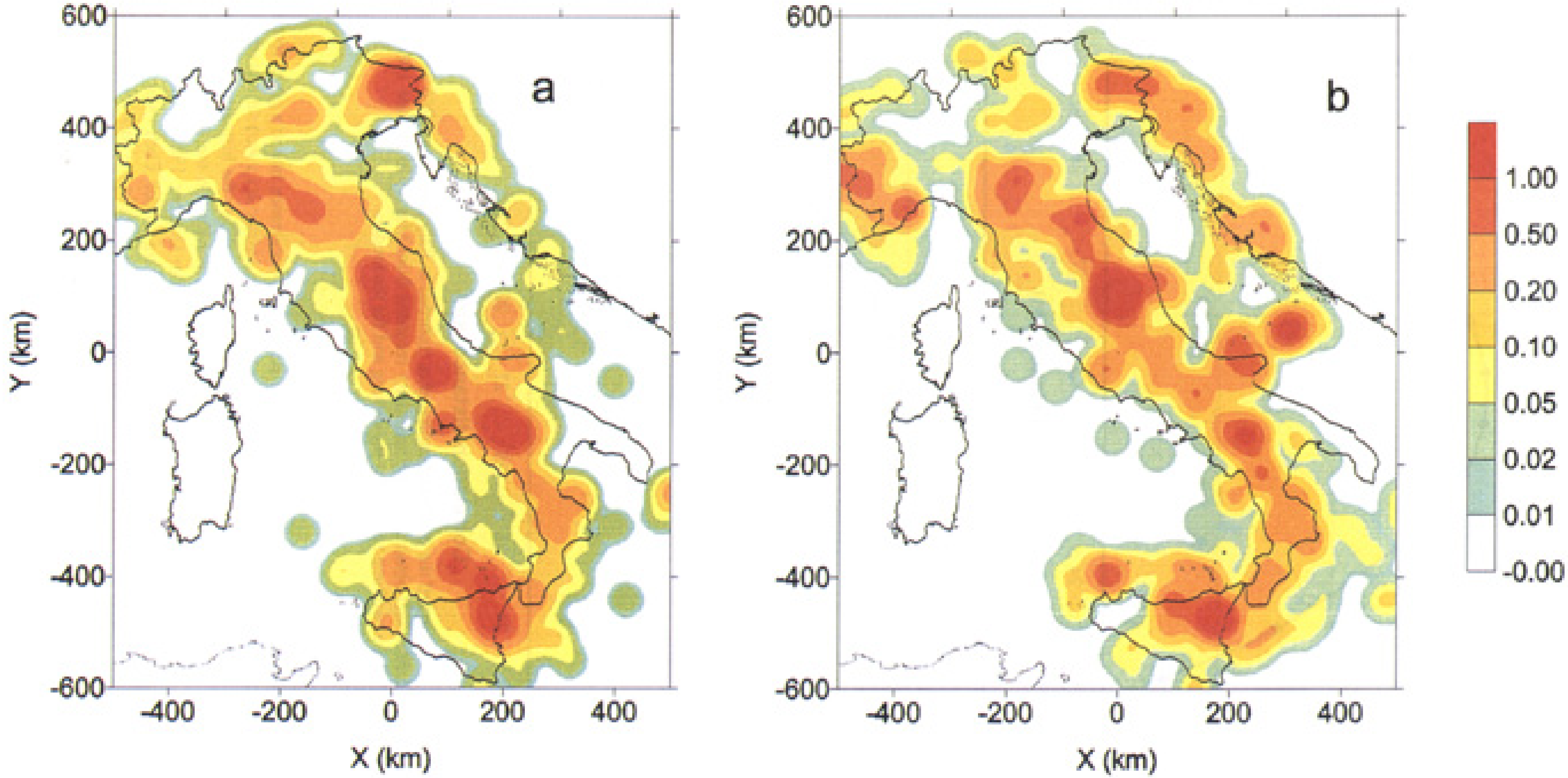
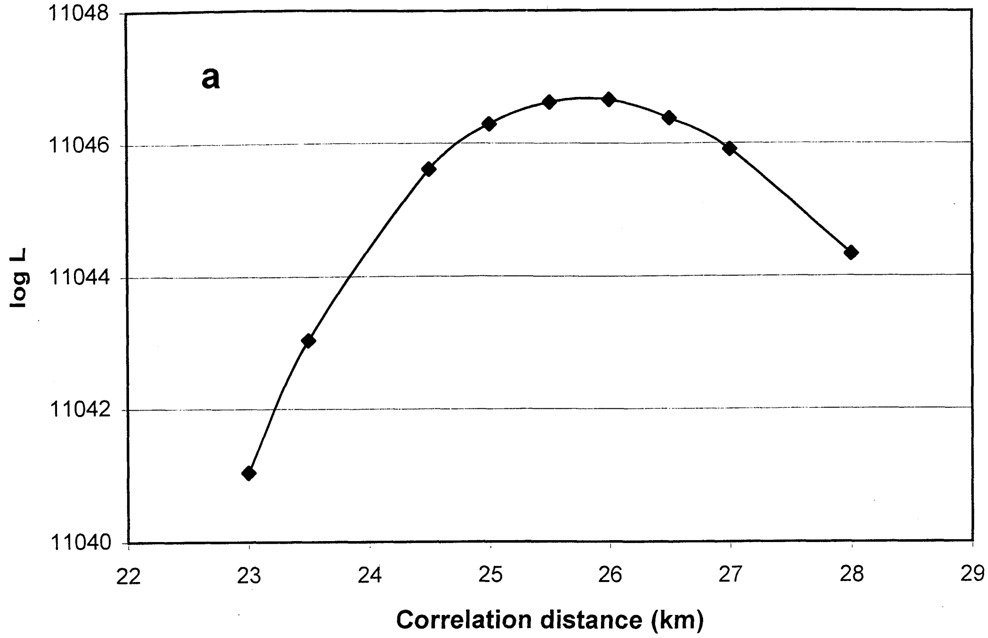


Plate 1. (a) Smoothed seismicity of the Italian territory for the period May 1976 to June 1986, using $d = 26.5$ km as the value of the correlation distance. (b) The same as in Plate 1a for the period June 1986 to December 1998. The color scale gives the average number of earthquakes ($M \geq 3.5$ and $h \leq 70$km) in an area of 100 km^2, over the time period spanned by the respective catalogs. The origin of the rectangular coordinates is in the point 42^0N, 13^0E.



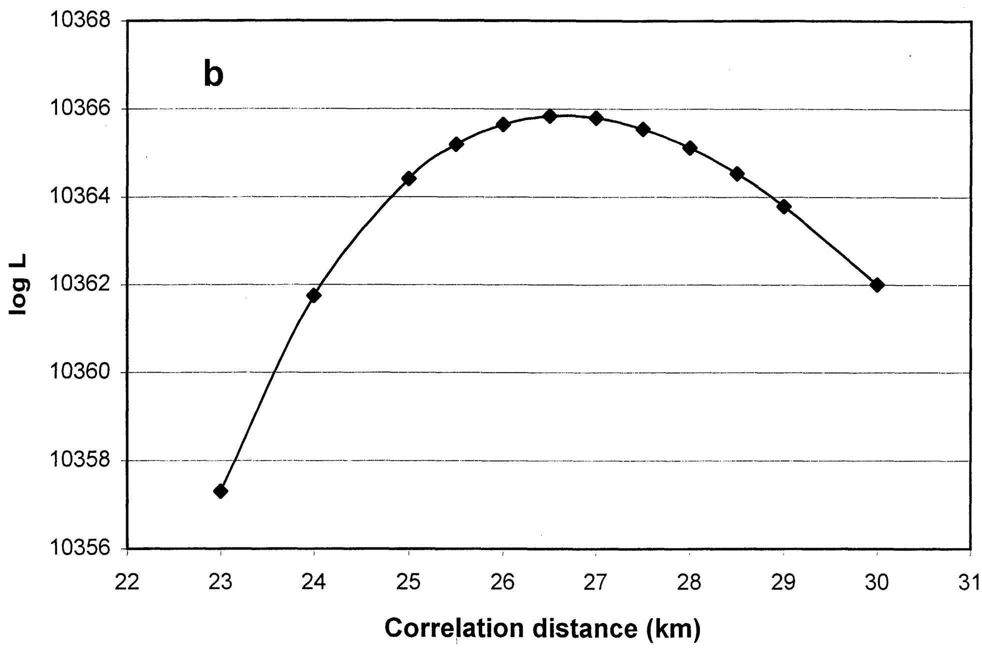


Figure 2. (a) Plot of the log-likelihood of the catalog of earthquakes recorded in Italy from June 1986 to December 1998 under the time-independent model obtained by the seismicity recorded from May 1976 to June 1986. (b) As for Figure 2a, having exchanged the two time periods for the catalog and for the model.

dure converged consistently on the following final values, regardless of the initial guess of the starting values of the parameters used for the best fit:

$$\sigma = 5.2 \text{ km}$$
$$K = 0.0887$$
$$c = 0.0194 \text{ days}$$
$$p = 1.094$$

yielding a failure rate $f_r = 0.523$.

In order to better show how this model works, let us consider the time and the space dependence separately. As to the time dependence, it can be visualized in two ways: either the time history of the occurrence rate in a given rectangular area, or the occurrence rate density $\lambda(x,y,m,t)$ on a given point of coordinates (x,y). Figure 3a shows the histogram of the expected number of earthquakes per day on the whole Italian territory in the period of time from October 1980 to January 1981, in accordance to the clustering model. Each value of the histogram was computed taking into account the past seismicity only. Note that the expected rate before the date of the 1980 Irpinia earthquake is not zero but ranges from 0.16 to 0.20, according to the current level of activity. Figure 3b shows the observed number of earthquakes in the same region and period of time. During this period the occurrence of the November 23, 1980, M_s= 6.9 earthquake of

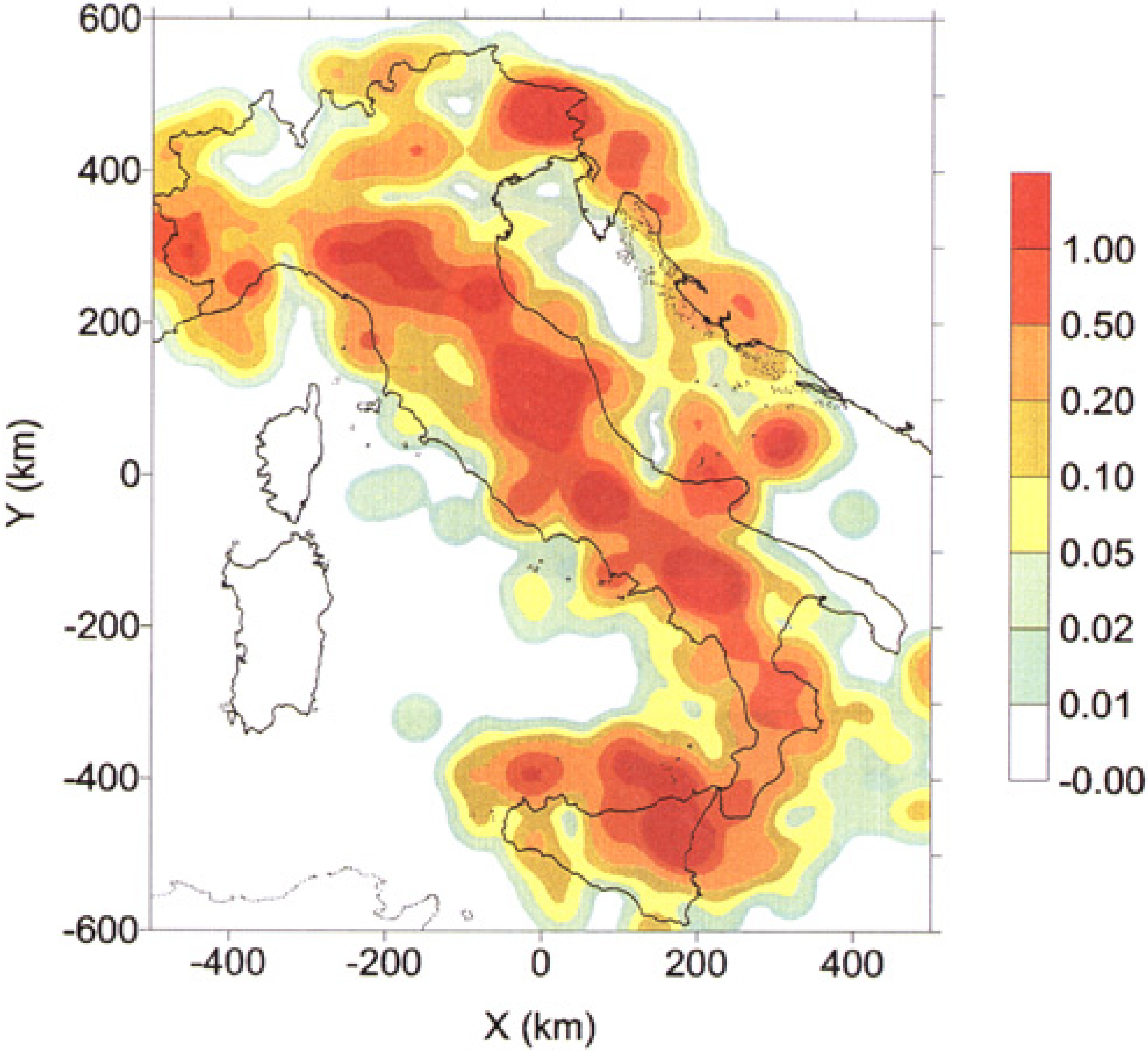


Plate 2. Smoothed seismicity of the Italian territory for the period May 1976 to December 1998, using $d = 26$ km as the value of the correlation distance. The color scale gives the average number of earthquakes ($M \geq 3.5$ and $h \leq 70$ km) in an area of 100 km^2, over the whole time period spanned by the catalog. The origin of the rectangular coordinates is the point 42^{0}N, 13^{0}E.

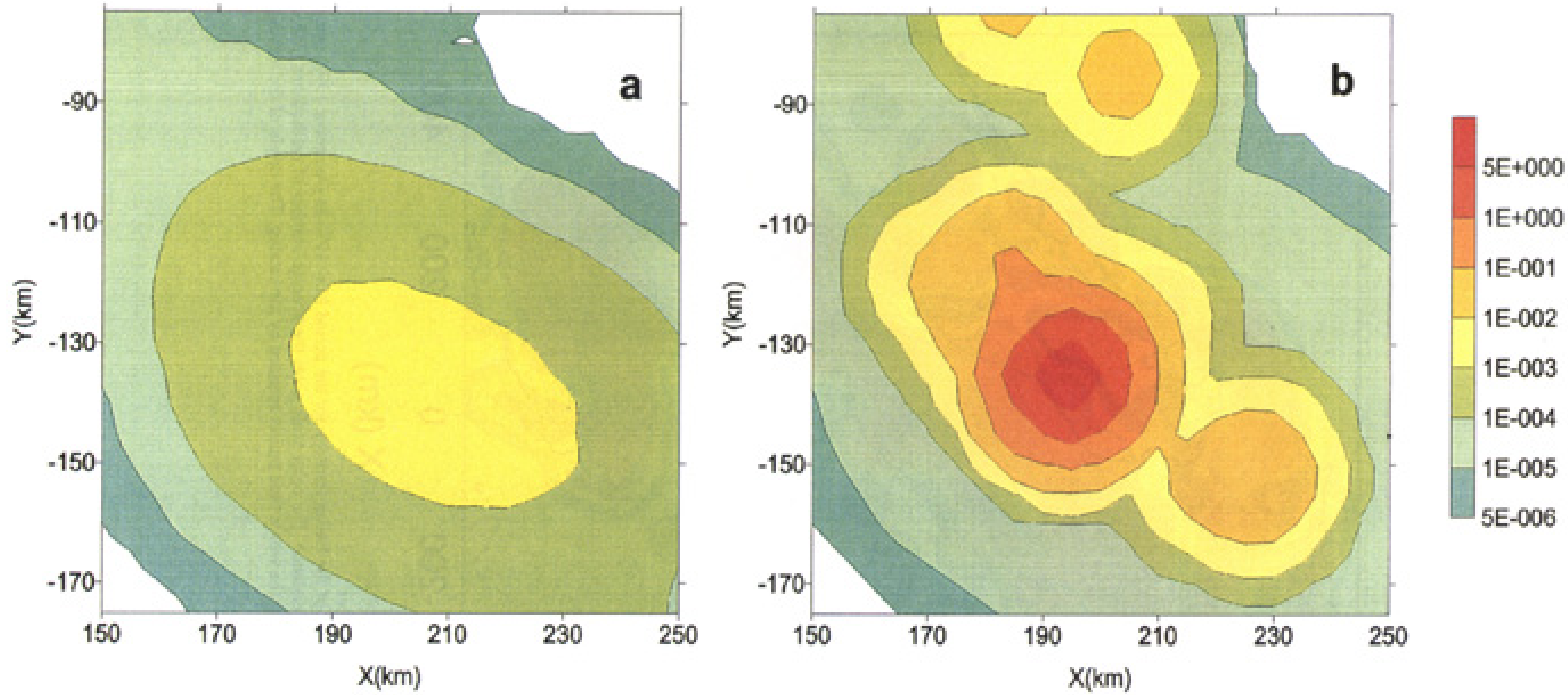
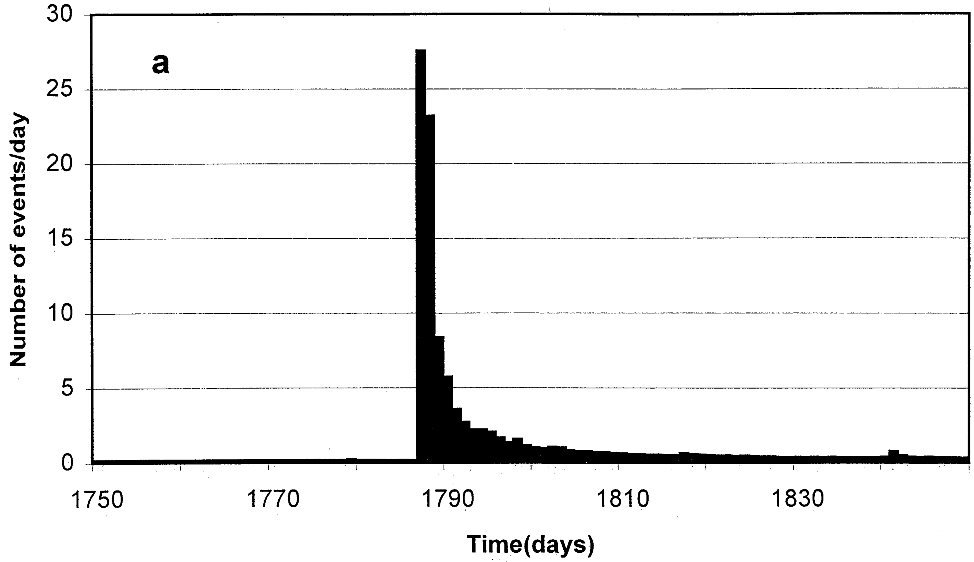


Plate 3. (a) Modeled occurrence rate density (day times 100 km^2)$^{-1}$ in the region surrounding the Irpinia earthquake, at 0000 UTC on November 23, 1980. (b) As in Plate 3a, 24 hours later, after the occurrence of the earthquake. The origin of the rectangular coordinates is the point 42^{0}N, 13^{0}E.



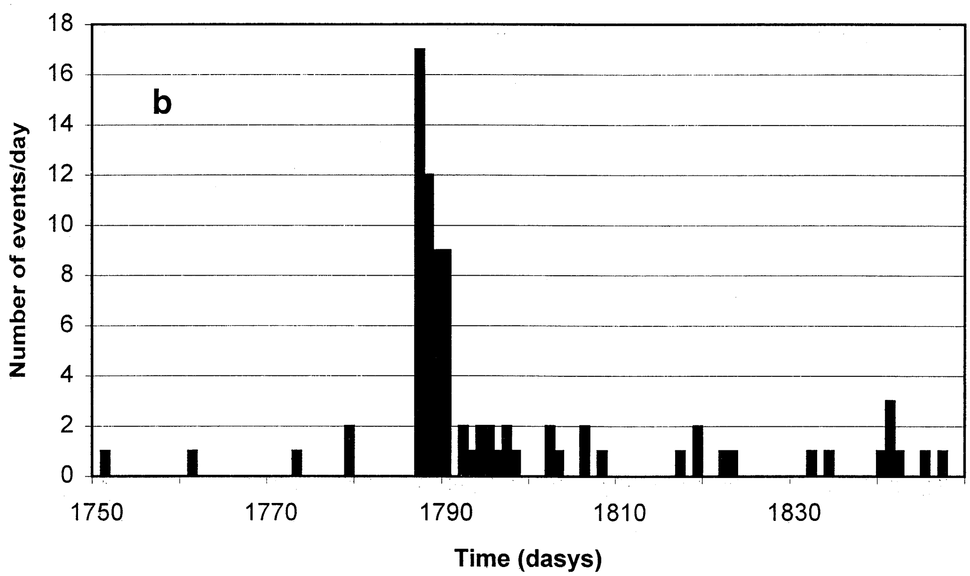


Figure 3. (a) Expected number of earthquakes of magnitude ≥ 3.5 over the whole Italian territory, per day, under the clustering model with the parameters obtained from the best fit, for a period of time including the Irpinia earthquake, M_s= 6.9, occurred in central Apennines on November 23, 1980. (b) Observed number of earthquakes for the same region and the same time period of Figure 3a. For both plots the time scale starts on October 15, 1980 (1750 days after January 1, 1976).

Irpinia in southern Italy dramatically modified the expected rate, in accordance with what has been actually observed. The accordance between the expected and the observed number of events, shown in Figures 3a and 3b, respectively, is pretty good, of course, taking into account the fact that only in the latter, can integer counts be reported. Similar tests, carried out for other large magnitude earthquakes of the catalog, have shown that in most cases the expected occurrence rate is within a difference of 50% from the observed one. Nevertheless, this difference can be occasionally quite larger. Figure 4 shows a continuous plot of the expected occurrence rate density on a point close to the epicenter of the Irpinia earthquake. The values shown by the logarithmic scale indicate an increase of several orders of magnitude as soon as this earthquake occurred. After the main shock the plot exhibits a regular decrease in connection with the fact that no aftershocks of comparable magnitude had been recorded. It is interesting to note that, with the c and p parameters obtained from the best fit the expected rate density is 1 year after the main shock, still at least 1 order of magnitude larger than before.

In order to have a clear vision of the local maxima of the function $\lambda(x,y,m,t)$ in space it was considered useful to plot a slice of $\lambda(x,y,m,t)$ at a given t. As an example of what a plot of this kind looks like, see Plates 3a and 3b, showing the geographical distribution of the modeled occurrence rate density in the region surrounding the 1980 Irpinia earthquake the day before and the day of the occurrence, respectively, of this strong seismic event. Plate 3a can be regarded, actually, as a detail of Plate 2, because the time-independent component of the occurrence rate dominated the overall rate at that particular time (no

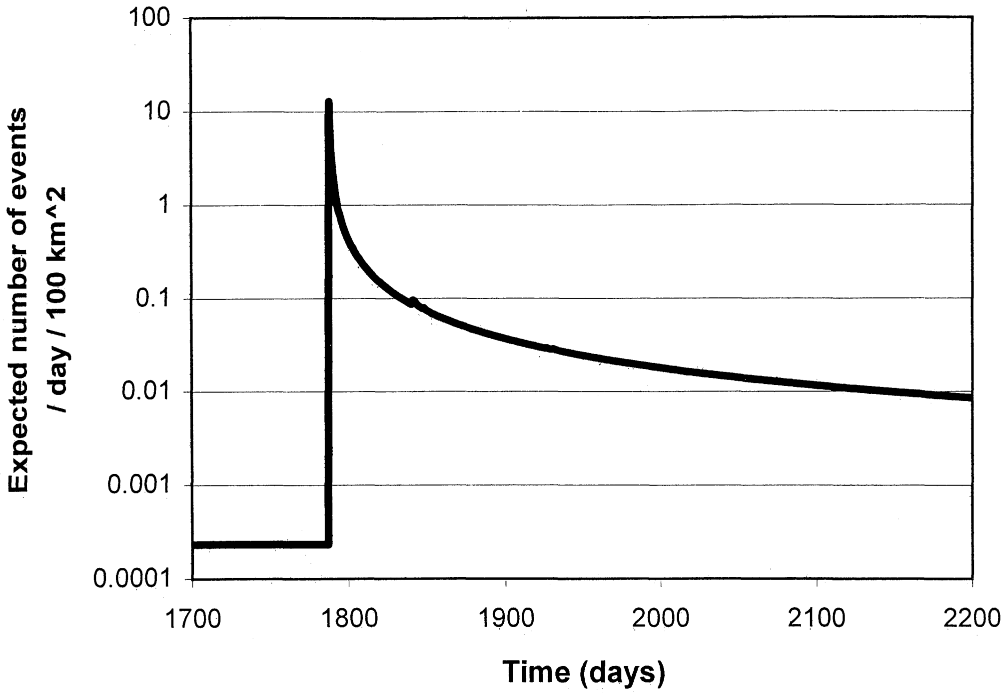


Figure 4. Expected occurrence rate density on a point of the Italian territory close to the epicenter of the Irpinia earthquake. For the beginning of the time scale, see Figure 3.

activity had been observed shortly before the main shock). We are reminded, however, that all the earthquakes in the catalog, including the aftershock series of the 1980 Irpinia earthquake, contributed to density rate reported in Plate 3a, so that it can not be considered a forecast of the future activity in real time. Plate 3b, compared with Plate 3a, clearly shows an increase of the occurrence rate density by a factor as large as 5 orders of magnitude due to the occurrence of the magnitude 6.9 earthquake and its immediate aftershocks. According to the value of σ (5.2 km) adopted in the present clustering model the induced seismicity extends significantly to a maximum radius of some tens of kilometers, as it was observed on the real distribution of aftershocks during the following days.

6. Test of the Hypothesis

As stated in the introduction, an objective test of any forecasting hypothesis should not be carried out on the same data that had been used for the formulation of the hypothesis itself. In this work, we decided to carry out the test on the latest data collected by the Italian Seismological Network, available at the time of the revision of this manuscript, from January 1 to December 31, 1999. This data set consists of 54 epicenters of earthquakes of magnitude ≥ 3.5, within the same geographical area selected for the learning phase. Figure 5 shows the epicentral map of these earthquakes. It could be noted that the number of 54 earthquakes is roughly the half of what would be expected from the average of the time period of 23 years used in the learning phase. This circumstance reflects the unusually low seismic activity observed in Italy in 1999.

The likelihood of the above mentioned catalog lasting 365 days was calculated, in steps of 1 day, both for the clustering and the Poisson model, and their ratio (performance factor) was computed. The performance of the generalized hypothesis during the test period is shown in Figure 6. The final value of the performance factor is 5.25 x 10^{36} and its logarithm is 84.6, as reported in the first column of Table 1, labeled "Total likelihood".

From Figure 6, it is easily seen that the performance factor in-

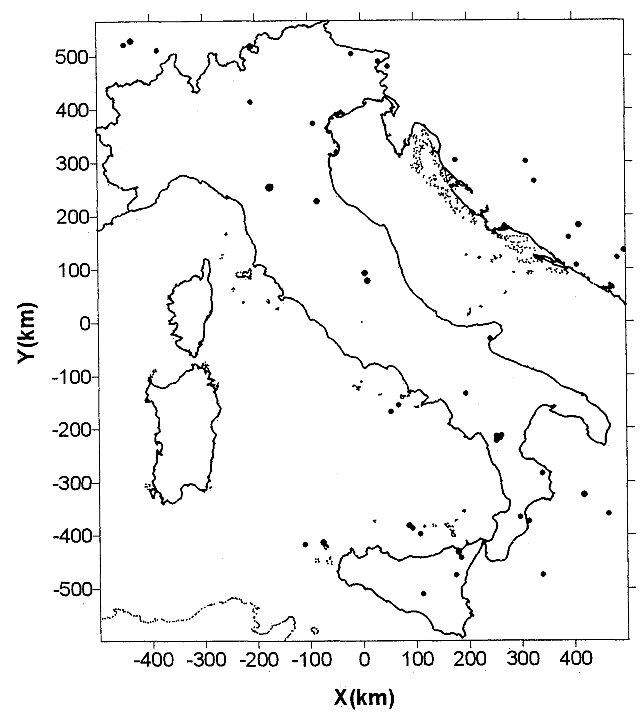


Figure 5. Epicenters of the earthquakes (M≥3.5 and h≤ 70km) located by the National Seismological Network in Italy and surrounding areas from January to December 1999. The origin of the rectangular coordinates the point 42^0N, 13^0E.

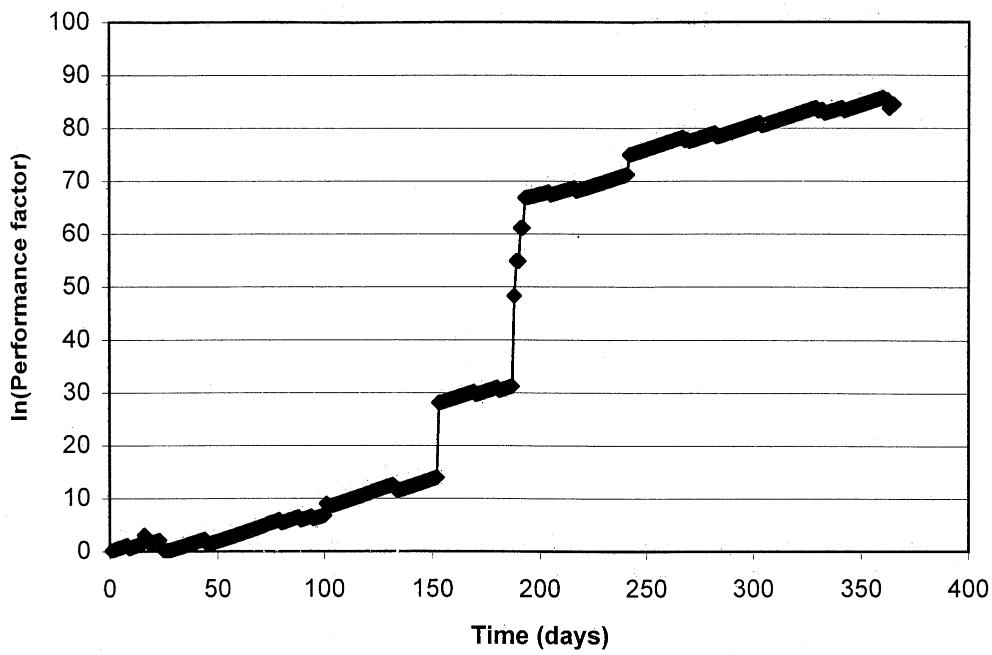


Figure 6. Performance factor of the clustering model against the time-independent Poisson model, computed for the Italian seismicity observed from January to December 1999.

creases steadily whenever there are no earthquakes. This is because according to (11), in absence of preceding earthquakes that can produce an increase of the rate density, the clustering model predicts less events than the Poisson model, and its likelihood is larger. This is reflected in the second column of Table 1 (labeled "Nonoccurrence") showing that in the test period the expected number of earthquakes for the Poisson and the clustering model, computed by the second term of the right-hand side of equation (11), is 109.4 and 68.5, respectively. The difference between these two values (40.9) is equal to the total amount of the steady increases of Figure 6. It can also be noted that the number of 68.5 expected earthquakes according to the clustering model includes 57.2 spontaneous and 11.3 induced events, respectively associated to the first and to the second term of the right-hand side of (7).

Whenever an earthquake not preceded by another one in its neighborhood occurs, the performance factor drops by slight negative steps. On the other hand, the performance factor has a positive step when an event occurs close to a place where a previous one had occurred. This is the case, e.g., at $t = 152$ days and $t = 187$-190 days, corresponding to seismic activity in the southern Tyrrhenian Sea and in the northern Apennines, respectively. The first case refers to an M=4.7 mainshock followed in few hours by two aftershocks of magnitude 3.7 and 3.9, within an epicentral distance of 10 km. The second one is represented by an M=5.4 mainshock followed in the following three days by 4 aftershocks of magnitude ranging between 3.5 and 3.9, within an epicentral distance of 5 km.

The sum of all the positive and negative steps related to each of the 54 earthquakes reported in the 1999 catalog, computed by the first term of the right-hand side of equation (11), is 43.7 and it is reported in the third column of Table 1, labeled "Occurrence". This means that the performance factor achieved by the clustering model with respect to the Poisson model comes in similar proportions from both terms of (11), related to the earthquakes reported in the catalog (occurrence) and to the absence of earthquakes (nonoccurrence).

7. Discussion and Conclusions

In this paper we have shown that given a thoroughly defined hypothesis, it is possible to build up a model that provides the

Table 1. Results of the Hypothesis Test. Italy, January 1 December 31,1999 ($M_{min} = 3.5$)[a]

	Total Likelihood	Nonoccurence	Occurrence
ln L_0	202.6	-109.4	312.0
ln L_1	287.2	-68.5	355.7
ln L_{aft}	-	-	351.7
ln L_{for}	-	40.9	325.9
ln (L_1/L_0)	84.6	40.9	43.7
Performance factor ($=L_1/L_0$)	5.25×10^{36}	5.8×10^{17}	9.5×10^{18}

[a] L_0, likelihood of the catalog under the uniform Poisson hypothesis; L_1, likelihood of the catalog under the clustering hypothesis; L_{aft}, likelihood of the events preceded by larger events in the catalog under the clustering hypothesis; L_{for}, likelihood of the events preceded by smaller events in the catalog under the clustering hypothesis. The "Occurrence" and "Nonoccurrence" components of the log-likelihood are separately computed by the first and second term of the right-hand side of equation (11).

occurrence rate density of earthquakes expected at any point of the location-time-magnitude space, and hence it is possible to test this hypothesis against a real sample of seismicity by statistical methods. In a similar way, a test based on a clustering model has been recently started by *Jackson and Kagan* [1999] for the northwest and southwest Pacific regions of the Earth.

Our hypothesis is a typical example of a fully empirical approach to a problem that cannot be solved in a deterministic way. Giving a physical meaning to the hypothesis would be desirable from a philosophical point of view, and it could also improve the efficiency in modeling the observations. However, at this stage this task seems far from being suitable for practical application.

The qualifying point of the model adopted in this paper is its simplicity, which implies a minimum number of arbitrary assumptions and the relative number of free parameters. It can, for instance, account for a complicate set of seismic features, currently called foreshock and aftershock series, seismic swarms, earthquake sequences, etc., all of which require ad hoc definitions. As a consequence, our model overcomes the necessity of introducing empirical rules such as the so-called "Bath's" law [*Richter*, 1958, p. 69] (that becomes a direct consequence of the "Gutenberg-Richter" law), in line with what was earlier put forward by *Lomnitz* [1966] and *Vere-Jones* [1969].

The drawback of its simplicity is that our model does not account for any space and time variation of the parameters. For this, it would predict a large number of induced events after a large magnitude earthquake in an area or at depths where it could be already known that induced seismicity is unlikely.

Moreover, our model is not capable of adapting itself to circumstances in which the induced activity happens to be lower or higher than the expected one. In principle, this flexibility could be fairly easily incorporated in the algorithm at the cost of making the model more complicated. At this stage, we did not want to do that, because our purpose was to test the model in its most general configuration.

The model presented in this paper is characterized by a short range of interaction due to the negative exponential function chosen for the space distribution in (10), while it would be desirable that a clustering model could reproduce fault interaction even over potentially large distances. In order to achieve this, different space distributions could be tested. In particular, an inverse power law, requiring one more parameter as by *Ogata* [1998], seems suitable for this purpose and will be considered in the development of this work.

With a test carried out on 1 year of nation-wide seismological observations, we have shown that the clustering model works well in comparison with a simple constant-rate Poisson model. This result could appear obvious because it is easy to construct a model that will perform better than a uniform Poisson model for a non declustered catalog. On the other hand, it is also clear that testing on a declustered catalog would remove the possibility of showing the better performance of the clustering model.

What we want to demonstrate is actually that our model can provide good hazard rate estimates, not only for aftershocks following main shocks in the common sense but also for larger-magnitude shocks following a previous seismic activity. This can be accomplished by a more detailed analysis of the first term on the right-hand side of (11). This term is computed by the sum of the logarithm of the rate density for all the earthquakes in the catalog. As shown in Appendix A, each rate density is obtained, for the clustering model, by the contributions of all the previous earthquakes in the catalog. It is possible to separate in this computation the contributions given by the previous earthquakes of magnitude larger than, or equal to, the magnitude of the jth earthquake from those given by the previous earthquakes of smaller magnitude. In this way, we shall obtain two values of likelihood: one related to the events following larger ones (L_{aft}) and the other related to the events following the smaller ones (L_{for}). This is carried out without the need of defining any particular event as aftershock or non aftershock.

Proceeding in this way, for the same data used in the previous tests we obtained, excluding the negative contribution of the expected total number of earthquakes as shown in Table 1, L_{aft} = 351.7 and L_{for} = 325.9, respectively. The latter value, which is still larger than the analogous component of the likelihood obtained from the uniform Poisson model by 13.9 (corresponding to a performance factor of 1.09 x 10^6), proves that the clustering model has a significantly better performance than a stationary Poisson model. This result is true even if the computation of the expected rate density of every earthquake is limited to the preceding earthquakes of smaller magnitude and the difference in the expected number of earthquakes is not taken into account.

Even if our algorithm is still subject to significant improvements, we feel already confident, in agreement with *Jackson and Kagan* [1999] to propose that the model, as a sort of null hypothesis, be used instead of the plain Poisson model in testing the validity of more sophisticated earthquake forecasting hypotheses.

Appendix A

We here build the full occurrence rate density for the model introduced in the main text of the paper. For the meaning of the symbols, refer to the main text.

By substitution of (9) and (10) into (8), and of the latter into (7), together with (5), we have the complete expression of the rate density:

$$\lambda(x,y,t,m) = f_r \mu_o(x,y)\beta e^{-\beta(m-m_o)} + \frac{\beta K}{2\pi\sigma^2}(p-1)c^{(p-1)}\sum_{i=1}^{N} H(t-t_i)(t-t_i+c)^{-p} e^{-\beta(m-m_i)} e^{-r^2/2\sigma^2}, \tag{A1}$$

with $r = [(x\text{-}x_i)^2 + (y\text{-}y_i)^2]^{1/2}$. This expression can be substituted into (11) for computing the log-likelihood of the observation of a set of seismic events $\{(x_j,y_j,t_j,m_j),\ j = 1,\ldots,N\}$ of magnitude equal to or larger than m_o in the time span $[0,T]$.

The first term of (11) is written explicitly as

$$\sum_{j=1}^{N} \ln[\lambda(x_j,y_j,t_j,m_j)V_o]$$

$$= \sum_{j=1}^{N} \ln\{[f_r\lambda_o(x_j,y_j,m_j) + \sum_{i=1}^{j-1}\lambda_i(x_j,y_j,t_j,m_j)]V_o\}$$

$$= \sum_{j=1}^{N} \ln\{[f_r\mu_o(x_j,y_j)\beta e^{-\beta(m_j-m_o)} + K'\sum_{i=1}^{j-1}\frac{e^{-\beta(m_j-m_i)}e^{-r^2/2\sigma^2}}{(t_j-t_i+c)^p}]V_o\}, \tag{A2}$$

where

$$K' = \frac{\beta K}{2\pi\sigma^2}(p-1)c^{(p-1)}.$$

The second term becomes:

$$\int_X\int_Y\int_0^T\int_{m_o}^{\infty}\lambda(x,y,t,m)dxdydtdm$$

$$= T\int_X\int_Y f_r\mu_o(x,y)dxdy$$

$$+K'\int_X\int_Y\int_0^T\int_{m_o}^{\infty}\sum_{i=1}^{N}\frac{H(t-t_i)e^{-\beta(m-m_i)}e^{-r^2/2\sigma^2}}{(t-t_i+c)^p}dxdydtdm. \quad \text{(A3)}$$

The first term of the right-hand side of (A3) represents the expected total number of "spontaneous" events, and the second term represents the expected total number of "induced" events. Notice that in order to compute these quantities it is not necessary to define which particular event is spontaneous and which is induced. The integration over time in the second term of the right-hand side of (A3) can be transformed changing the variable of integration t into t - t_i and the limits of integration of any term of the sum from 0 to $T - t_i$ into t_i to T (while the integration over the space gives approximately 1), thus obtaining:

$$\int_X\int_Y\int_0^T\int_{m_o}^{\infty}\lambda(x,y,t,m)dxdydtdm$$

$$= T\int_X\int_Y f_r\mu_o(x,y)dxdy + K\sum_{i=1}^{N}\int_0^{T-t_i}\frac{e^{-\beta(m_o-m_i)}}{(t+c)^p}dt$$

$$= T\int_X\int_Y f_r\mu_o(x,y)dxdy$$

$$+K\sum_{i=1}^{N}e^{-\beta(m_o-m_i)}[c^{1-p}-(T-t_i+c)^{1-p}],(p>1). \quad \text{(A4)}$$

The rate density for the stationary Poisson model comes from this general formulation for the limit case of f_r = 1 and K = 0, thus obtaining:

$$\ln L_o = \sum_{j=1}^{N}\ln[\lambda_o(x_j,y_j,m_j)V_o]$$

$$-\int_X\int_Y\int_0^T\int_{m_o}^{\infty}\lambda_o(x,y,m)dxdydtdm$$

$$= \sum_{j=1}^{N}\ln\{[\mu_o(x_j,y_j)\beta e^{-\beta(m_j-m_o)}]$$

$$-T\int_X\int_Y\mu_o(x,y)dxdy. \quad \text{(A5)}$$

Acknowledgements. We are particularly grateful to David Rhoades for his useful advice, constructive comments, and suggestions that have led to the preparation of the present paper. We also thank David Vere-Jones for his reviewing the final version of the paper. The development of the methodology was performed by one of the authors (R. Console) during his stage at Earthquake Research Institute of the University of Tokyo, September 22-December 21, 1998.

References

Console, R., Computer algorithms for testing earthquake forecasting hypotheses, internal report, 19 pp., *Earthquake Res. Inst.*, Univ. of Tokyo, Tokyo, 1998.

Di Luccio, F., R. Console, M. Imoto, and M. Murru, Analysis of short time-space range seismicity patterns in Italy, *Ann. di Geofis.*, *40*, 783-798, 1997.

Evans, R. (Ed.), Assessment of schemes for earthquake prediction, *Geophys. J. Int.*, *131*, 413-533, 1997.

Frankel, A., Mapping seismic hazard in the Central and Eastern United States, *Seismol. Res. Lett.*, *66*, 8-21, 1995.

Geller, R.J. (Ed.), Debate on evaluation of the VAN method, *Geophys. Res. Lett.*, *23*, 1291-1452, 1996.

Habermann, R.E., Teleseismic detection in the Aleutian Island arc, *J. Geophys. Res.*, *88*, 5056-5064, 1983.

Jackson, D.D., Hypothesis testing and earthquake prediction, *Proc. Natl. Acad. Sci. U.S.A.*, *93*, 3772-3775, 1996.

Jackson, D.D., and Y.Y. Kagan, Testable earthquake forecasts for 1999, *Seismol. Res. Lett.*, *70*, 393-403, 1999.

Kagan, Y.Y., and D.D. Jackson, New seismic gap hypothesis, Five years later, *J. Geophys. Res.*, *100*, 3943-3959, 1995.

Lomnitz, C., Magnitude stability in earthquake sequences, *Bull. Seismol. Soc. Am.*, *56*, 247-249, 1966.

Musmeci, F., and D. Vere-Jones, A space-time clustering model for historical earthquakes, *Ann. Inst. Stat. Math.*, *44*, 1-11, 1992

Ogata, Y., Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure, *J. Phys. Earth*, *31*, 115-124, 1983.

Ogata, Y., Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, *50*, 379-402, 1998.

Reasenberg, P.A., and L.M. Jones, Earthquake hazard after a mainshock in California, *Science*, *243*, 1173-1176, 1989.

Rhoades, D.A., and F.F. Evison, On the reliability of precursors, *Phys. Earth Planet. Inter.*, *58*, 137-140, 1989.

Rhoades, D.A., and F.F. Evison, The precursory earthquake swarm in New Zealand: Hypothesis test II, *N. Z. J. Geol.Geophys.*, *40*, 537-547, 1997.

Richter, C.F., *Elementary Seismology*, 768 pp., W.H. Freeman, New York, 1958.

Utsu, T., Aftershocks and earthquake statistics (III), *J. Fac. Sc. Hokkaido Univ., Ser. VII, Geophys.*, *3*, 379-441, 1971.

Vere-Jones, D., A note on the statistical interpretation of Bath's law, *Bull. Seism. Soc. Am.*, *59*, 1535-1541, 1969.

Wyss, M., R. Console, and M. Murru, Seismicity rate change before the Irpinia (*M*=6.9) 1980 earthquake, *Bull. Seismol. Soc. Am.*, *87*, 318-326, 1997.

R. Console and M. Murru, Istituto Nazionale di Geofisica, Via di Vigna Murata, 605, I-00143 Rome, Italy. (Console@ingrm.it; Murru@ingrm.it)

(Received November 2, 1999; revised May 15, 2000; accepted June 29, 2000.)