Geophys. J. Int. (1992) **109**, 501–516

Aftershock identification: methods and new approaches

G. M. Molchan and O. E. Dmitrieva

International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Warshavskoye sh., 79, k2, Moscow 113556, *Russia*

Accepted 1991 November 27. Received 1991 November 27; in original form 1991 April 12

SUMMARY

The problem of aftershock identification in earthquake catalogues is studied. Some empirical methods are considered and quantitavely analysed.

Game theory approach is applied to formulate the problem allowing a whole class of optimal methods of aftershock identification. Each method is optimal depending on the goals and gives the best trade-off between the number of missed aftershocks and the number of incorrectly identified ones. Some illustrations of the new approach to the aftershock identification problem are presented.

Key words: aftershock identification, cluster methods, loss function.

1 INTRODUCTION

Seismic events tend to cluster in space and time. Fore- and aftershocks represent the most important type of such clustering. By various estimates the aftershocks contribute about 30–40 per cent to the total number of earthquakes in world catalogues and contain significant information on rupture processes.

Catalogues are usually separated into aftershocks and mainshocks in order to study aftershock properties or the fine structure interaction. The more easily identifiable aftershocks are earlier aftershocks of large earthquakes in the near zone, i.e., where the rate of aftershock occurrence per unit time and space significantly exceeds the background rate. Otherwise the identification of aftershocks is hampered by background seismicity and by overlapping aftershock sequences. The absence of a physical concept of later aftershocks is the main difficulty for aftershock separation. Under these circumstances a great number of aftershock identification techniques exists. However, it is not always clear how the utilization of a particular method affects geophysical conclusions drawn from the data.

Below we shall see that even the ideal model of a single cluster superposed upon the background seismicity permits an infinite number of 'optimal' methods to identify the cluster. The choice of a particular method should depend on the specific goals of the declustering, but in fact this is almost never explicitly discussed.

Therefore it is essential to discuss existing aftershock identification methods together with their goals. As a result, a goal-oriented approach to the problem will be proposed. Our main results are the following: a quantitative analysis of Prozorov's 'dynamic algorithm' of aftershock identification, a mathematical formulation of the problem of cluster identification and its solution, both theoretical and practical.

2 AFTERSHOCK IDENTIFICATION METHODS

Without pretending to give a comprehensive review of aftershock identification methods, we shall consider the most popular techniques.

The ***hand procedure*** is a visual identification of earlier aftershocks. It is usually applied for the analysis of large well-studied earthquakes. The method can use auxiliary information on fault-plane solutions and fault geometry. The method becomes speculatory for later and distant aftershocks, and evidently the procedure needs to be formalized for data processing.

The ***window method*** is the simplest formalization of the hand procedure. Aftershocks of a mainshock (t_0, g_0, M_0) are identified within space–time windows:

$$t_0 < t < t_0 + T, \qquad |g - g_0| < D, \qquad M < M_0, \tag{1}$$

where t, g, M are time, epicentre coordinates and magnitude, respectively. Sometimes the number of aftershocks within a window is required to be significantly greater than the expected number of background events with $M < M_0$.

The catalogue can be analysed either in chronological order or in decreasing order of magnitude. By definition the initial event is considered to be a mainshock. Elimination of each mainshock together with its fore- and aftershocks from the catalogue leads us to the recurrent rule of mainshock identification. If fore- and aftershock events are identified simultaneously, then some uncertainty in mainshock identification arises depending on the choice of mainshock ordering, namely, by time or by magnitude.

Window method thresholds D and T depend on the mainshock magnitude. This dependence is illustrated by Fig.

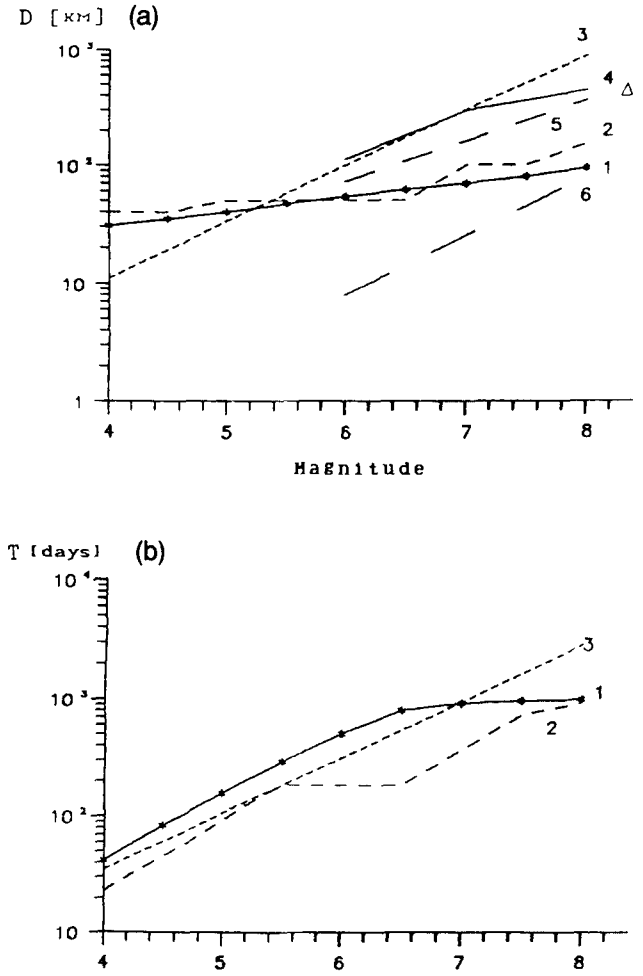

Figure 1. (a) Space and (b) time windows for aftershock identification according to: 1—Gardner & Knopoff (1974); 2—Keilis-Borok *et al.* (1980); 3—Knopoff *et al.* (1982); 4—Prozorov (1986); 5—Molchan & Dmitrieva (by LIR procedure, NOAA catalogue, see text); 6—Utsu & Seki (1954). The triangle marks 1964 Alaska earthquake according to Kanamori (1977).

1 for results from various problems (Gardner & Knopoff 1974; Keilis-Borok, Knopoff & Rotwain 1980; Knopoff, Kagan & Knopoff 1982; Prozorov 1986). On the whole, the spatial window sizes are rather similar. For $M_0 \leq 6$ they are of the order of the accuracy of epicentre location in the world earthquake catalogues. For $M_0 > 6$, the spatial windows are in agreement with our determinations of aftershock zone maximum semi-axis (see below). However, all the thresholds are substantially higher than the estimates by Utsu & Seki (1954) made for the areas of early aftershocks:

$$\log S = M - 4, \qquad M \geq 6 \tag{2}$$

(in Fig. 1 this relation is reduced to the largest semi-axis of aftershock zone under the assumption of its ellipticity with aspect ratio 2:1).

Authors' intentions and the choice of window sizes are to some extent interrelated. Gardner & Knopoff (1974) examined whether the mainshock flow is Poissonian. To do this, they used the largest thresholds D and T. That allows interconnections between events to loosen as much as possible. The next simple operation such as elimination of spatial coordinates (time projection) leads us to a well-known mathematical model: summation of weakly dependent (here thinned) flows of events. Under very general conditions the resulting flow ought to be Poissonian (see, for example, Daley & Vere-Jones 1988).

In earthquake prediction Keilis-Borok *et al.* (1980) used lower time thresholds. This is natural, because for forecasting purposes we need to keep the prediction properties of the catalogue. In turn the influence of remaining aftershocks on prediction requires careful analysis, especially when dealing with the prediction of double earthquakes.

Knopoff *et al.* (1982) tested the hypothesis of equality of slopes of foreshock and aftershock frequency–magnitude curves. Their results show that this hypothesis becomes more acceptable as the size of the time-space windows increases. This may be due to the increased share of background events in samples with larger windows.

The window method is very simple and suitable for data processing, but does not incorporate specific aftershock location features, e.g., offset of the aftershock zone centre with respect to the mainshock epicentre. As can be seen from Fig. 2, the offset for $M_0 = 6.5$ is of the order of 3–30 km, while $D = 50$ km is recommended for the spatial window size in Keilis-Borok *et al.* (1980). The result may be that using moderate spatial windows can lose a lot of aftershocks, thereby producing false alarms in prediction.

Cluster methods define the notion of nearness of events (metric d) and nearness of clusters of events in space and time. In such a way the whole earthquake catalogue can be divided into non-overlapping clusters. The largest event in each cluster is a mainshock by definition and the other events in the same cluster are its fore- and aftershocks. There is a wide variety of near-event definitions, as well as of methods for utilizing physical information.

Formal methods examine earthquakes as homogeneous space–time objects $x = (g, t)$ which should be divided into well-separated groups with high inner concentrations of elements. It is a typical problem of cluster analysis (see for example Granadesikan & Kettenring 1989). One of the oldest methods, Single-Link Cluster analysis (SLC), is now used in seismostatistics (Frohlich & Davis 1985; Davis & Frolich 1991a, b). The SLC procedure connects elements of the original set by a chain of minimum length (length is measured in the metric d) and then removes all the edges longer than d_0. The result is to split the chain into isolated points and clusters. There are some optimal properties of such clustering, e.g., the shortest distance between any two points of different clusters (intercluster distance) is longer than the intracluster distance:

$$d(K) = \max_{x_i \in K} \min_{x_j \in K} d(x_i, x_j)$$

Davis & Frohlich (1991a, b) used the following form of metric d for earthquakes:

$$d^2 = |g_1 - g_2|^2 + c^2 |t_1 - t_2|^2, \tag{3}$$

where c is a parameter. The metric does not depend on magnitude and is invariant under shifts in space and time, which does not seem natural in view of strong spatial

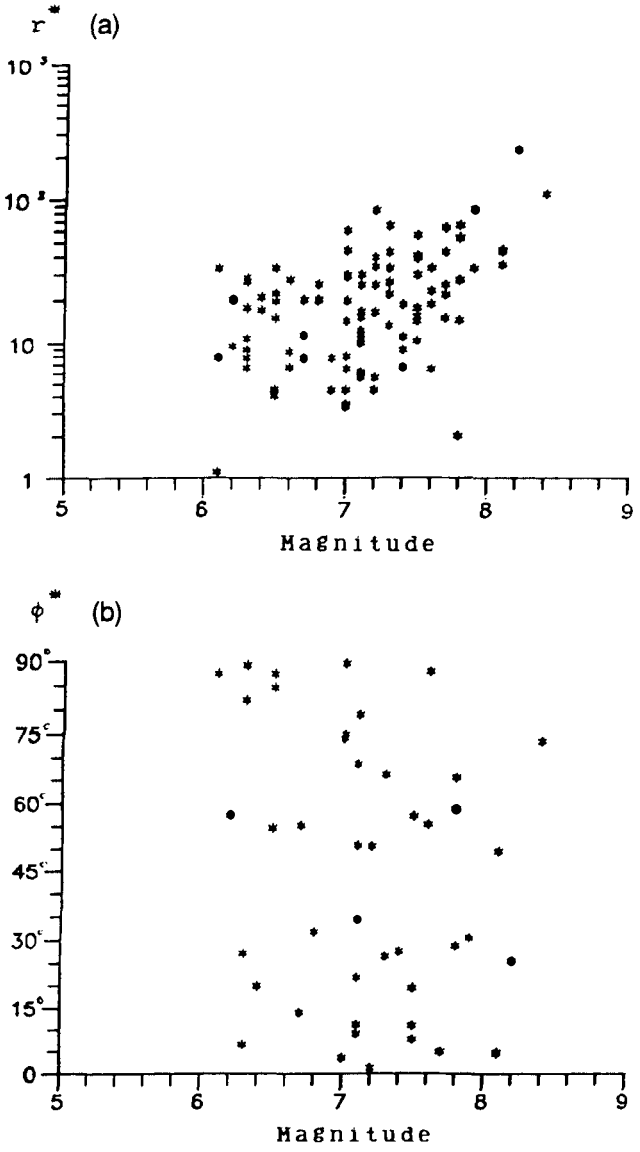

Figure 2. The offset of mainshock epicentre relative to the centre of the aftershock zone as a function of mainshock magnitude $M_0 \geq 6$: (a) distance r^* (km) and (b) the minimum angle φ^* with the longer axis of the aftershock zone. Catalogue NOAA (1964–1980) is used; aftershock zones with aspect ratio ≥2 and with number of aftershocks ≥20 are shown.

inhomogeneity of seismicity. This circumstance should be taken into account when cluster elements are interpreted as foreshocks and aftershocks. Davis & Frohlich (1991a) showed that the SLC method can be equivalent to window methods in an integral sense, i.e. the methods can give the same 'score' S when tested on model examples: S = the fraction of afterevents which are linked to their parent primary event + the fraction of primary events which are correctly identified as such (i.e. not misidentified as aftershocks). Unfortunately the integral index S does not characterize aftershock identification quality for large and small earthquakes separately. The authors did not manage to fit the parameters (c, d_0) so that the SLC method could work for different ranges of mainshock magnitude equally well. This can be seen from the example of the 1964, $M_0 = 8.4$, Alaska earthquake. The SLC method divides later aftershocks of this earthquake into several clusters, two of which are rather large. If the parameters (c, d_0) in the SLC method are fixed and the cut-off magnitude in the catalogue is diminishing then all the events of the region may be identified as a single cluster. This is dissonant to the practice of using low-magnitude aftershocks when determining the aftershock zone. So if the cut-off magnitude diminishes, the SLC method faces two problems: increasing of computer calculations and forced redefining of the notion of event 'nearness' (see Davis & Frolich 1991b).

Non-formal methods attempt to start with a model and then to identify aftershocks in the context of that model. Usually fore- and aftershocks are considered as chains or branching trees of causally connected events with given statistical properties. Let us consider Reasenberg's (1985) method which developed from the work of Savage (1972). Here the window method is used to define a local nearness of events, the spatial threshold d depending on magnitude according to the formula

$$\log d(\text{km}) = 0.4M_0 - 1.943 + k.$$

This relation follows from a simple circular fault model (of radius d) in which the seismic moment for static cracks is defined by Keilis-Borok's formula: $16/7\ \Delta\sigma d^3$ (see Kanamori & Anderson 1975), where $\Delta\sigma$ is the stress drop. Using the empirical estimate $\Delta\sigma = 30$ bar and the moment–magnitude relation, we get the above dependence with $k = 0$.

To link each next event (t, g, M) with an existing cluster we have to evaluate its spatial and temporal proximity to the cluster. Spatial proximity to a cluster is defined by proximity either to the largest event (with magnitude M_*) or to the last event of the cluster (with moment t^*). The values of the spatial proximity parameter 'k' for these two cases are different ($k = 1$ for proximity to the largest event and $k = 0$ for proximity to the last one). Time proximity is defined as follows:

$$t - t^* < \min(\tau_\varepsilon, 10 \text{ days}).$$

The threshold τ_ε taken from Omori's relation depends on time as follows. Let $\lambda(u \mid M_*)$ be Omori's law for a mainshock of magnitude M_* where u denotes the time elapsed since the mainshock. For a Poissonian aftershock flow the time of the next event follows the distribution

$$P\{t > t^* + \tau\} = \exp\left[-\int_0^\tau \lambda(u + t^* - t_* \mid M_*)\, du\right]$$

where t_* is the time of the largest event in the cluster. Using the condition $p(t > t^* + \tau) < \varepsilon$ the value of τ_ε can be obtained:

$$\int_0^\tau \lambda(u + t^* - t_* \mid M_*)\, du = -\ln \varepsilon, \qquad \varepsilon = 0.05.$$

According to this principle each subsequent event is linked with the largest event or with the last one in each cluster which has formed until the current time. Overlapping clusters are joined.

Reasenberg (1985) and Reasenberg & Jones (1989) considered aftershock intensity to be independent of the point in space, i.e., Omori law and the average number of

aftershocks $\Lambda_A(M_0)$ are the same for main events of a given magnitude M_0. In particular, the time–magnitude distribution of aftershocks for California is

$$\lambda(t, M \mid M_0) = 10^{a+b(M_0-M)}(t+t_0)^{-p}, \qquad M < M_0 \tag{4}$$

where $a = -1.76$, $b = 0.91$, $t_0 = 0.05$ (days) and $p = 1.07$.

The cluster procedures are free from *a priori* assumptions on the spatial aftershock distribution structure, so they are able to reveal aftershock migration provided the background seismicity is low. On the whole, the advantages and disadvantages of any method depend on the final goals of research. It can be surmised that in Reasenberg's method and declustered catalogue has to look like an uncorrelated homogeneous Poissonian random field. Unfortunately such goals do not limit the number of false aftershocks.

Modelling and statistical estimation. We mean here statistical methods which do not use subjective declustering for studying fore- and aftershocks (Vere-Jones & Davies 1966; Kagan & Knopoff 1976, 1981; Ogata 1988). In this case the seismic process is described by a model and the parameters involved are estimated by statistical methods, for instance, by means of maximum likelihood. This approach seems to be the most natural, though its possibilities are limited by studying general cluster properties only. Trying to describe the process in detail we are confronted with an estimation problem involving many parameters. As a result, the domain of the maximum likelihood becomes very wide and the problem becomes unstable (Kagan & Knopoff 1976).

3 PROZOROV'S ITERATIVE METHOD

Consider one more method (Prozorov & Dziewonski 1982; Prozorov 1986; Prozorov & Schreider 1986) which plays a significant role in further generalizations. This method is largely a formalization of the hand procedure. It automatically finds the spatial scatter zone for aftershocks and events are identified until their intensity becomes comparable with the local background seismicity.

In Prozorov's method mainshocks are considered in decreasing order of magnitude, obeying time chronology when the magnitudes are equal. For the fixed mainshock (t_0, g_0, M_0), the first thing any method does is to find a portion of statistically significant aftershocks. This is done by using a spatial window S of size $D(M_0)$ (see Fig. 1). An event (t, g, M), $t \geq t_0$, $g \in S$ is identified as an aftershock if

$$n(S\Delta_t) \geq R\lambda_b(S)\,|\Delta_t| \tag{5}$$

where $n(S\Delta_t)$ is the number of events within the $S\Delta_t$ volume, $|\Delta_t|$ is the length of the time interval $\Delta_t = [t_0 + \alpha(t - t_0), t]$, $\lambda_b(S)$ is the expected number of background events within S per unit time, and R is a threshold varied by the author in the range 3 to 100. The quantity $R - 1$ is a kind of the 'signal/noise' ratio, since $n(S\Delta_t)$ includes both aftershocks and background events, while $\lambda_b\,|\Delta_t|$ estimates background seismicity in $S\Delta_t$; α is a parameter within the range $[0, 1]$. The choice $\alpha \neq 0$ is preferable. Identification of an event should depend on the intensity at the current moment. Clearly the quantity $n(S\Delta_t)/|\Delta_t| = \hat{\lambda}_t$ estimates the current intensity better when $0 < \alpha < 1$. The bias for $\alpha \simeq 0$ and the variance for $\alpha \simeq 1$ are large.

Preliminary aftershock identification is terminated as soon as condition (5) fails or when the identification time exceeds the threshold T:

M_0	4	5	6	6.5	8
T (years)	1	2	3	4	5

If the number of preliminary aftershocks is greater than 10, then the aftershock zone S is updated. The new zone S is given by the aftershock concentration ellipse:

$$S_k = \{g : \hat{\rho}^2 = (g - \hat{g}_*)\hat{B}^{-1}(g - \hat{g}_*)' \leq k^2\}$$

where $\{g\}$ are epicentres, $\hat{g}_* = \sum g_i/n$ is the centre of previously identified aftershocks; $\hat{B} = \sum (g_i - \hat{g}_*)'(g_i - \hat{g}_*)/n$ is the empirical covariance matrix for the sample $\{g_i\}$, and k is the dimensionless size of the ellipse. Aftershock identification goes on in several elliptic zones of different sizes k_i. Owing to this, the aftershocks ultimately belong to the pyramid $V = \{S_{k_i}\Delta t_i\}$ where a smaller base is generally combined with a longer time interval Δt_i. Practically V is a cylinder because the time interval Δ_t grows slowly with decreasing zone size k.

If aftershock epicentres are distributed according to a Gaussian distribution, then the size k can be derived from the prescribed confidence level $\varepsilon_k = P\{g \in S_k\}$ of zone S_k. When g_* and B are known, the statistic ρ^2 obeys a χ^2_2 distribution, and so $k^2_\varepsilon = 2 \ln 1/\varepsilon$. Under the Gaussian hypothesis we can also take into account the variation in $\hat{g}$ and $\hat{B}$ (see Molchan & Dmitrieva 1990). Let n be the number of observations on which $\hat{g}$ and $\hat{B}$ are based, then $2\hat{\rho}^2(n-1)/(n-2)$ obeys Fisher's distribution with $(2, n-1)$ degrees of freedom. Therefore we get

$$k^2_\varepsilon = (\varepsilon^{-2/(n-1)} - 1)\frac{(n-1)^2}{n-2}.$$

The values of k typically used (2, 3 and 4) approximately correspond to (0.1, 0.01 and 0.001) confidence levels for $n > 40$. The relation between k and ε is useful in choosing the maximum base of pyramid V. The smaller values of k in Prozorov's method are not substantiated.

Prozorov's procedure takes into account spatial aftershock localization and allows us to assess *a posteriori* (but not to control) the number of false aftershocks. Unfortunately the method is very sensitive to the value of R (Prozorov 1986). High values might lead either to the division of an aftershock sequence into separate clusters of earlier and later events, or many aftershocks might be lost.

The precise distribution of the number of aftershocks (ν) identified by Prozorov's method is given in Appendix A for model examples. Relations between the distribution of ν and the distribution of population in a special Galton–Watson branching process, came to light unexpectedly. On the other hand, this result allows the following practical conclusion: let $\lambda_A(t)$ and λ_b be the intensity of aftershock flow and background seismicity in the vicinity of mainshock epicentre at a moment t. *Then Prozorov's procedure of aftershock identification is soon terminated when* $R_0(t) = [\lambda_A(t) + \lambda_b]/\lambda_b$ *has reached* $R/2$.

The empirical results by Prozorov (1986) corroborate this statement: the decrease of threshold R by a factor of 5 ($R_1 = 100$, $R_2 = 20$) caused the aftershock identification time (t_i) to increase by a factor of 5 or 6 ($t_1 = 1$–3 yr,

$t_2 = 10$–12 yr). In fact, according to the statement, $R_0(t_1) : R_0(t_2) \cong (R_1/2) : (R_2/2)$. Under the Omori law assumptions, $\lambda_A(t)$ and $R_0(t)$ are proportional to t^{-1}. Hence $R_1 : R_2 \cong t_2 : t_1$, in agreement with Prozorov's experience.

The same arguments leads us to a useful estimate of R:

$$R \cong 2R_0(t_0)t_0/t_*$$

where $R_0(t_0)$ is the empirical signal-to-noise ratio for some time t_0, and t_* is a rough estimate of aftershock duration. For instance, when $t_0 = 10$ days, we derive from our data $R_0(t_0) = 10^2$–10^3 for $M_0 \geq 6$ mainshocks. Hence $R = 20$, when $t_* = 10^2$–10^3 days. If the number of background events is proportional to the area of the zone, while the number of aftershocks and the aftershock area are proportional to 10^{bM_0} [see below and (4)], then $R_0(t_0)$ should be weakly dependent on mainshock magnitude. For low M_0 the last conclusion is complicated by the problem of epicentre location accuracy.

4 NEW APPROACHES TO THE AFTERSHOCK IDENTIFICATION PROBLEM

Our basic assumption is that aftershock sequences are finite, the aftershocks concentrate in space and time and are mixed with background seismicity. For this reason an error-free aftershock identification is an impossibility. Larger space–time aftershock windows will capture false events (both background seismicity and aftershocks of different mainshocks n_A^-), while smaller ones will lose true aftershocks (n_A^+). A trade-off between the two kinds of errors ($\Lambda_A^\pm = En_A^\pm$) could be a natural basis for rigorous formulation of the aftershock identification problem.

The total number of aftershock events with $M > M_{\min}$ for all mainshocks of magnitude M_0 seems to be weakly dependent on M_0 for $M_0 > M_{\min} + 2$. [To see this, recall that, according to (4), the number of aftershocks for a mainshock M_0 is proportional to $10^{b\Delta} - 1$, $\Delta = M_0 - M_{\min}$, while the number of magnitude M_0 events is proportional to $10^{-\beta M_0}$. Their product is therefore weakly dependent on M_0, provided $b \cong \beta$ and $M_0 + M_{\min}$ is large.] However, the aftershock sequence of a larger earthquake is more numerous and is more easily identifiable, and hence has higher priority in aftershock studies.

We consider the aftershock identification problem locally. In particular, let us consider an ideal situation where some space–time volume GT contains a mixture of independent background events and the aftershock sequence of a known mainshock (M_0, g_0, t_0). Usually such *a priori* localization is assumed, either explicitly or implicitly, in many informal methods. How is one to identify the aftershocks in this case?

To solve the problem we have to introduce a measure for evaluating the quality of aftershock identification, e.g. a 'loss function' γ which depends on errors of the two kinds and increases in each argument $\gamma = f(\Lambda_A^+, \Lambda_A^-)$. Then the identification problem is reduced to finding the rule that minimizes γ. This localized problem is solved below. Our solution is the most complete in the case of Poissonian background seismicity and aftershock flow.

Two loss functions seem to be reasonable: $\gamma = \alpha\Lambda_A^+ + \beta\Lambda_A^-$ and $\gamma = \max(\alpha\Lambda_A^+, \beta\Lambda_A^-)$.

The respective optimal principles of aftershock identification will be called the *game theory* principle:

$$\alpha\Lambda_A^+ + \beta\Lambda_A^- \Rightarrow \min \tag{6}$$

where α and β are losses for a missed aftershock and for an event identified incorrectly as an aftershock; and the *minimax* principle:

$$\max(\alpha\Lambda_A^+, \beta\Lambda_A^-) \Rightarrow \min. \tag{7}$$

When $\alpha = \beta$ the minimax principle is free of parameters (α, β) and leads to compensation of two kinds of errors: $\Lambda_A^+ = \Lambda_A^-$, that is, the mean number of identified events equals the true value. When $\alpha \neq \beta$ the condition (7) controls the error ratio: $\Lambda_A^+/\Lambda_A^- = \alpha/\beta$. If $\alpha = 0$ ($\beta = 0$) all the events of the region are identified as aftershocks (mainshocks).

Statement 1. Poissonian case

If the background and the aftershock flows are Poissonian with intensities $\Lambda_b(g, t)$ and $\Lambda_A(g, t)$, respectively, then the γ optimal decision rule of aftershock identification for loss functions (6) and (7) takes the form

$$\pi(g, t) = \begin{cases} \textit{aftershock}, & \textit{if } L(g, t) > c, \\ \textit{background event}, & \textit{if } L(g, t) < c, \end{cases} \tag{8}$$

where $L = \Lambda_A(g, t)/\Lambda_b(g, t)$. For the game principle, $c = \beta/\alpha$; in the minimax principle c is given by the equation $\alpha\Lambda_A^+ = \beta\Lambda_A^-$.

In particular, let $T = (0, \infty)$, G is 2-D plane and

$$\Lambda_b(g, t) = \Lambda_b, \qquad \Lambda_A(g, t) = \Lambda_A p(g) f(t) \tag{9}$$

where $\Lambda_A = \Lambda_A(M_0)$ is the average number of events in the aftershock sequence of a mainshock of magnitude M_0, $f(t)$ is the Omori law in normalized form, $t > t_0$:

$$f(t) = (t/t_0)^{-p} \frac{p-1}{t_0}, \qquad t > t_0, \tag{10}$$

and $p(g)$ is a 2-D Gaussian distribution of aftershock epicentres

$$p(g) = (2\pi \det B)^{-1} \exp[-r^2(g - g_x)/2] \tag{11}$$

where $r^2(g) = g'B^{-1}g$ is a quadratic form, g_x is the centre of aftershock dispersion, and B is the covariance matrix of the aftershock sample.

Then the γ-optimal method identifies an even as an aftershock if

$$1/2r^2(g - g_*) + p \ln t/t_0 < c, \qquad t > t_0. \tag{12}$$

In the case of (6),

$$c = c_0 = \ln\left(\frac{\alpha}{\beta}\frac{p-1}{2}\frac{\Lambda_A}{\Lambda_b}\right), \tag{13}$$

where $\Lambda_b = t_0 |S_1| \lambda_b$ is the average number of background events within the dispersion zone $S_1 = \{g : r(g - g_*) < 1\}$ during time period t_0. In the minimax case c is found from the equation

$$c + \ln\left(1 - \frac{c/p}{\exp(c/p) - 1}\right) = c_0 - \ln p. \tag{14}$$

General case

Let background seismicity and an aftershock sequence be represented by a general mixture of two flows with intensities described by (9), where $f(t) \geq 0$, $\int f\,dt = 1$, and

$$p(g) = (2\pi \det B)^{-1}\psi[r(g - g_*)],$$

where $\psi(x) \geq 0$ is a decreasing function normalized by $\int_0^\infty \psi(x)x\,dx = 1$. We restrict ourselves to the class of aftershock identification rules $\mathcal{K}$ which classify as aftershocks the events from the domain

$$\{(g, t): r(g - g_*) < k(t)\} \tag{15}$$

where u is an arbitrary non-negative function. Then rule (8) belongs to $\mathcal{K}$, and is γ optimal in this class with respect to loss functions (6) and (7). In particular, for $\psi = \exp(-x^2/2)$ and f in the form (10) this rule leads to criteria (12–14). Note that:

(i) The algorithim by Reasenberg (1985) is based on the Poissonian distribution of events in space and time. But the first part of the statement and its proof show that under these conditions it is impossible to improve an aftershock identification method by means of dynamical improvements (sequential analysis).

(ii) Procedure (8) corresponds to a broad class of identification rules depending on the goals and *a priori* assumptions for aftershock distributions. The simplest assumptions are given by the intensity parameterization (9). It involves marginal distributions of aftershocks in space (elliptic dispersion) and in time (Omori law). Naturally, the closer model $\Lambda_A(g, t)$ is to reality, the more effective is the procedure. In particular, this model can easily incorporate aftershock migration, diffusion or a more complicated shape of Omori law, provided those phenomena are typical or important for the study of aftershocks in general. (For a proof of statement 1 see Appendix B).

5 PRACTICAL ASPECTS OF THE METHOD

The new approach proposed above is based on the modelling of aftershock and background intensities and on their ratio. We shall call it the local intensity ratio (LIR) method. To use the LIR method, the following parameters of seismicity should be estimated: the background rate around the mainshock Λ_b, the average number of aftershocks Λ_A, the aftershock centre g_* and covariance matrix B, and the parameter p in the Omori law. Prozorov's procedure has encountered these problems, so it would be natural to rely on the experience gained. This consists in iterative refinement of the parameters.

The determination of mainshocks is based on the magnitude ordering: the largest event in the catalogue is always considered as a mainshock. After this mainshock is eliminated from the catalogue together with its fore- and aftershocks then the next mainshock is defined to be the largest event. The background intensity is determined preliminarily from the entire catalogue and then refined using the mainshock catalogue by averaging over temporal and spatial cells depending on the region. The first iteration for each mainshock identifies aftershocks in a preliminary way by any simple method, for example, by means of moderate windows or by the LIR method itself with *a priori* parameter values. For example, the *a priori* parameters could be taken as follows: (1) circular scattering [$B = D(M)I$, where I is the unit matrix] around the mainshock ($g = g_*$); (2) the average number of aftershocks taken from

$$\Lambda_A = \Lambda_A(M_0) \propto 10^{(M_0 - M_{\min})b}$$

[see (4)] where b is the slope parameter for the frequency–magnitude law, and $M_{\min}$ is the cut-off magnitude; and (3) fixed Omori parameter $p = p_0$, say $p_0 = 1.1$, until there are enough aftershocks for its estimation.

Preliminary aftershocks should be significant [the test of significance is standard for the Poissonian distribution of the number of events in the area (e.g., see Molchan & Dmitrieva 1990)]. If there is a sufficient number of preliminary aftershocks ($n_A > 10$) then g_* and B are estimated, Λ_A being set equal to n_A. All the parameters of the aftershock identification procedure (12) are revised after each iteration. Therefore the matrix B is recalculated and the aftershock dispersion centre g_* is found to be displaced relative to the mainshock epicentre. A great number of aftershocks maintains statistical stability of parameters g_* and B. Therefore, diminishing of cut-off magnitude $M_{\min}$ ensures stability of the LIR procedure and enlarges scope for its application.

This procedure requires five to seven iterations to become stable when dealing with mainshocks of $M > 7.5$, but for lower magnitudes three or four iterations are sufficient.

One should be careful at the first step when applying centred windows, as this can cause mistakes in determining the aftershock centre. The elliptic zone of aftershocks corresponding to $\varepsilon = 0.001$ is used to isolate the hypothetical aftershocks from different clusters.

For each mainshsock the LIR procedure identifies significant foreshocks within the aftershock zone as described by Molchan & Dmitrieva (1990). Because of scant foreshock sequences it is impossible to construct a symmetric procedure for their identification.

6 PRELIMINARY ANALYSIS OF THE LIR METHOD

Below we are interested in LIR stability and in how far the properties of minimax aftershocks are consistent with actual experience. Aftershocks are identified by minimax LIR procedure with $\alpha = \beta$. Recall that in this case LIR procedure does not depend on the weights (α, β) and leads to unbiased estimation of $N_A = N_A(M_{\min})$.

We applied the *minimax* LIR method to the NOAA world catagloue (1964–1980, $M \geq 4$, depth < 100 km) and to the regional catalogue (Earthquake Hypocentre Data, California NEIC, File, West US, USGS–NEIC, 1963–3.10.1990, 22°–56°N, 140°–100°W, the total number of earthquakes is 16 380). The LIR method is used when the number of evident aftershocks is not less than 10. In the NOAA catalogue, mainshocks of $M_0 \geq 6$ fulfil this requirement while in the regional catalogue it is true for $M_0 \geq 5$. The results for these mainshocks are presented in Figs 1–5, 8 and 9. We shall summarize the following conclusions.

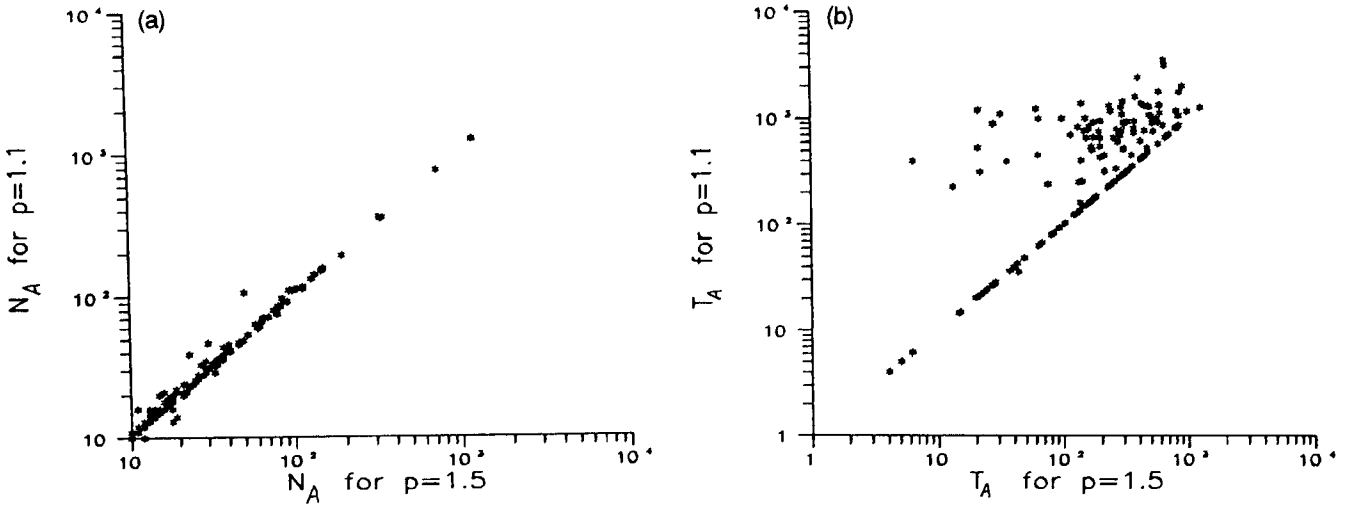

Figure 3. (a) The number N_A and (b) time duration T_A of aftershock sequences identified by the LIR method with the Omori parameter $p = 1.1$ and 1.5 Catalogue NOAA (1964–1980), $M_0 \geq 6$.

(i) Diminishing the Omori parameter p from 1.5 to 1.1 naturally leads to longer aftershock sequences. As follows from Fig. 3(a), the number of events in aftershock sequences does not increase significantly, but much later events appear to be identified as aftershocks. This conclusion is of interest, as it throws light on the stability of the errors $\Lambda_A^{\mp}$.

(ii) The aftershock zones obtained by the minimax LIR procedure are in good agreement with the hand techniques. For example, Fig. 4 presents aftershocks zones for three large ($M_0 \sim 7.5$) South Kuril earthquakes as determined by Balakina (1989) using the ISC data. The zones are in good agreement with 95 per cent elliptic zones, given by the minimax LIR method for the NOAA data.

Fig. 5 presents an aftershock sequence, identified for the 1964 Alaska earthquake ($M_0 = 8.4$). In contrast to the SLC method which divided the aftershock process into several clusters (see above), the LIR procedure automatically identified the entire aftershock sequence. The sequence includes 801 earthquakes with $M \geq 4$ (according to NOAA data); the last aftershock time delay is equal to 942 days; the 95 per cent ellipse bounds the aftershock area $Q =$

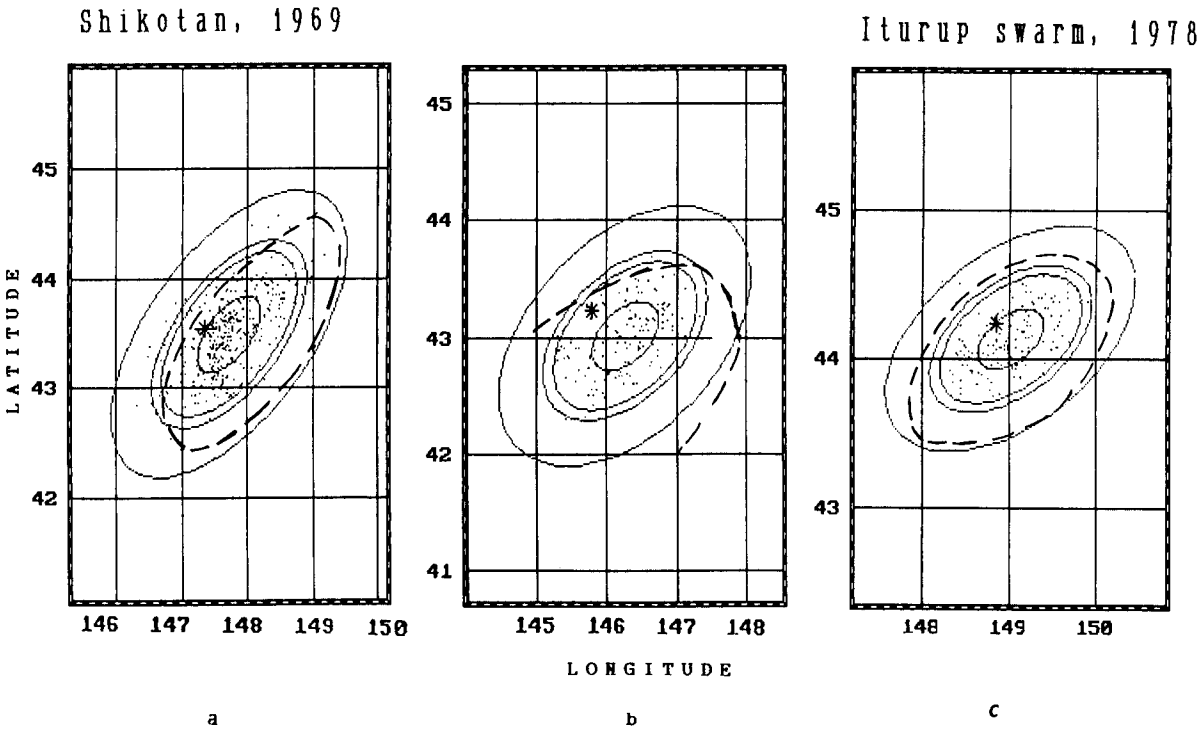

Figure 4. Comparison of the hand procedure (dashed line) with the LIR procedure for South Kuril earthquakes: (a) Shikotan earthquake, 1969 August 11, $M_0 = 7.9$; (*b*) 1973 June 17 earthquake, $M_0 = 7.7$; (c) Iturup swarm, 1978, $M_0 = 7.5$. Mainshocks and aftershocks are mapped by asterisks and dots, respectively. Aftershock zones of 60, 90, 95 and 99 per cent confidence level are shown by solid lines. LIR and hand procedures (Balakina 1989) are based on NOAA and ISC catalogue respectively.

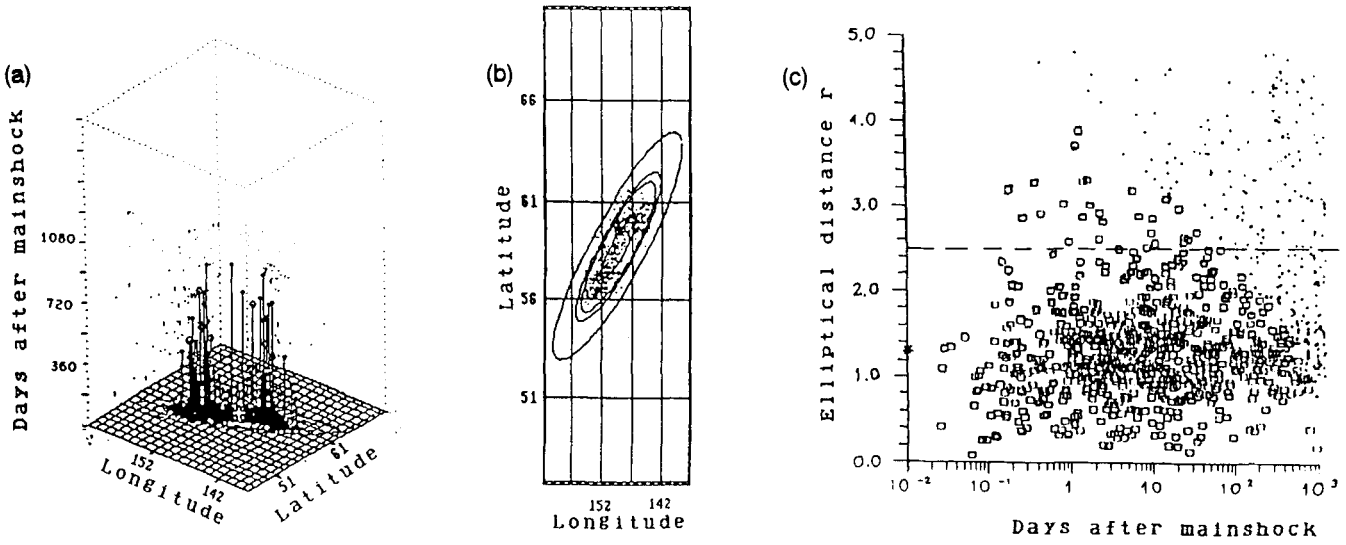

Figure 5. Alaska earthquake of 1964, $M = 8.4$ (NOAA catalogue). (a) Events in space–time: minimax aftershocks are marked by vertical sticks, while background events are marked by dots. (b) Space projection of minimax aftershocks. Aftershocks zones are shown for 60, 90, 95 and 99 per cent confidence levels. (c) Distance–time distribution of seismicity after the mainshock. The distance r of the event g from the aftershock dispersion centre is measured by a dimensionless value: $r^2 = (g - g_*)B^{-1}(g - g_*)$ (see text). Notations: squares—minimax aftershocks, dots—background events, asterisk—mainshock; dashed line corresponds to 95 per cent confidence level [see also (b)].

137 000 km^2 ($\log Q = 5.14$). Kanamori (1977) gave a similar empirical estimate ($\log Q = 5.15$). These examples show that the LIR method is flexible and readily adaptable to local seismicity features and to mainshock magnitude.

The similarity of aftershock zones identified by the LIR procedure to those outlined with hand techniques becomes understandable from Fig. 6 which presents aftershock areas estimated by whole aftershock sequences and by aftershocks taken during 10 days after mainshocks moments. The accordance is unexpectedly good, though in 1/3 of the cases 10 days aftershocks contribute less than 60 per cent to the total volume of minimax aftershock sequence. As they are easily visualized, initial aftershocks are fixed by hand techniques. In fact we see that 10 days aftershocks are sufficient to estimate the future aftershock area.

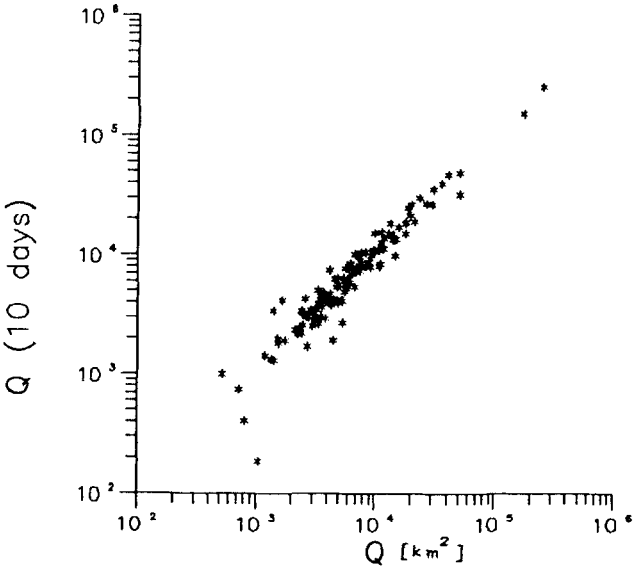

Figure 6. Areas of 95 per cent aftershock zones: for whole aftershock sequences, Q, and for aftershocks of the initial 10 days, Q (10 days) (NOAA catalogue).

(iii) The regional catalogue of the western US (see above) allows to analyse the influence of the cut-off magnitude $M_{\min}$ on the results of aftershock identification. We selected 22 aftershock sequences with mainshocks $M_0 \geq 5.3$ by two versions of the catalogue ($M_{\min} = 3$ and $M_{\min} = 4$). Two contrasting examples, the 1983 ($M_s = 6.7$) Coalinga earthquake and the 1983 ($m_b = 6.2$, $M_s = 7.3$) Borah Peak (Idaho) earthquake, are shown in Figs 7 and 8. When diminishing $M_{\min}$ from 4 to 3 the Coalinga 95 per cent aftershock zone changes its orientation and becomes two times greater in linear size. Such instability of the zone is due to the fact that weak aftershocks here are not in good accordance with the elliptic dispersion model. Conversely, the 95 per cent aftershock zone for the Borah Peak earthquake preserves the orientation and becomes less for $M_{\min} = 3$ due to the increasing of the statistics volume.

These examples illustrate a general situation for the 22 considered aftershock sequences. When changing $M_{\min}$ from 4 to 3, the linear sizes of aftershock zones ($\kappa = \sqrt{Q_3/Q_4}$, where Q_M is the zone area for $M_{\min} = M$) increase/decrease by not more than a factor of 2.2. The values of κ as a function of mainshock magnitude are presented in Fig. 9. The value of κ depends on completeness of registration: cases with deficiency in low magnitudes are specified in Fig. 9. Remark that the shapes of the frequency–magnitude law for aftershock sequences are similar to those for mainshocks within the same areas.

(iv) As has been mentioned, the aftershock concentration ellipse centre is offset relative to the mainshock epicentre (Fig. 2a). The offset increases with mainshock magnitude and is comparable with the source dimension. However, the offset direction relative to the longer aftershock zone axis, φ^*, is complete random (Fig. 2b). Fig. 2(b) presents representative aftershock samples with $N_A \geq 20$ (NOAA catalogue) which form zones with aspect ratio $\geq 2:1$.

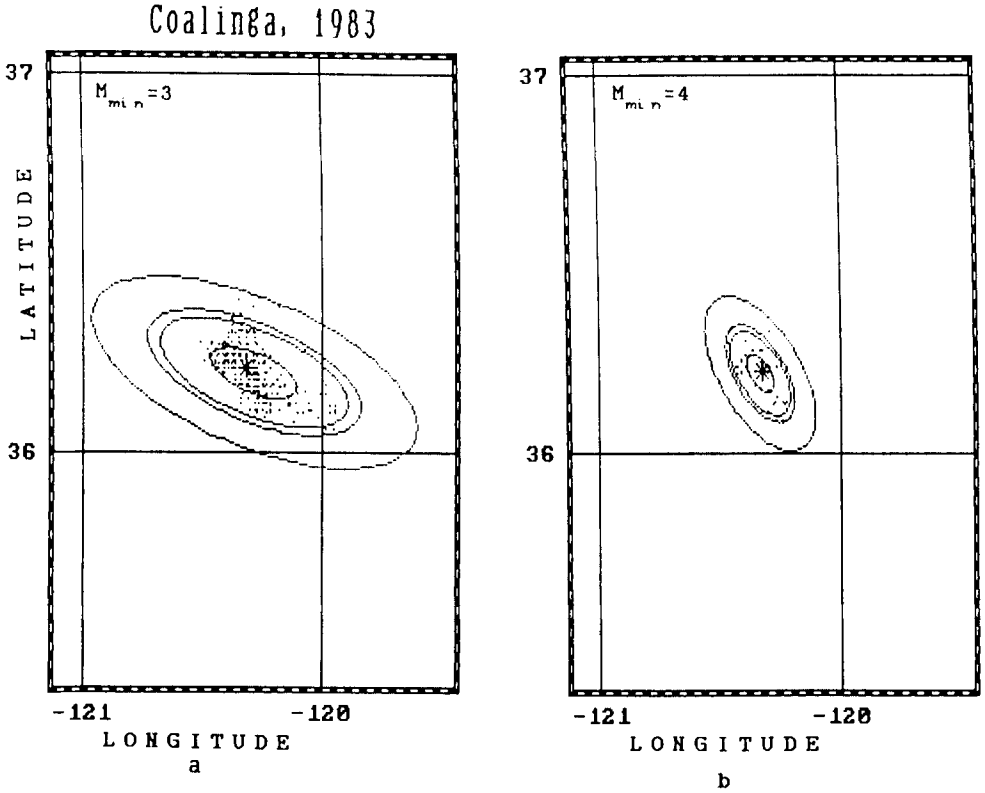

Figure 7. Minimax aftershocks of the Coalinga earthquake (1983, $M_s = 6.7$), identified in the 'West US' catalogue with different cut-off magnitude values: (a) $M_{min} = 3$, number of aftershocks $N_A = 378$; (b) $M_{min} = 4$, $N_A = 48$. Outlined zones correspond to 60, 90, 95 and 99 per cent confidence levels.

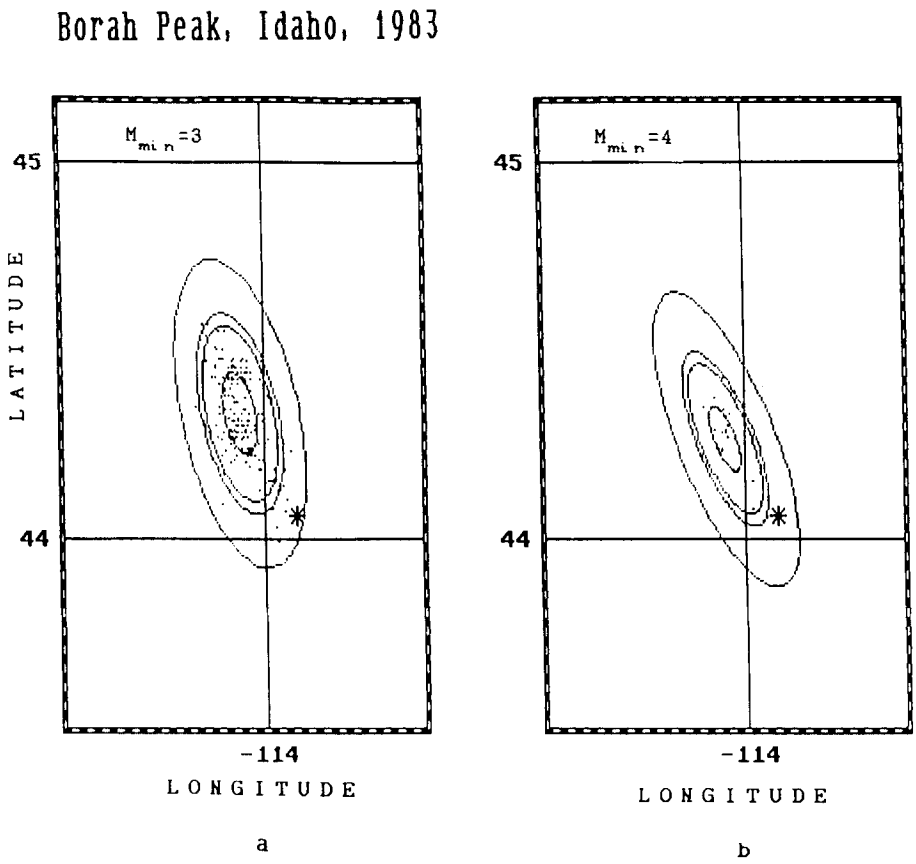

Figure 8. The same as Fig. 7 but for the Borah Peak, Idaho, earthquake (1983, $M_s = 7.3$). (a) $M_{min} = 3$, $N_A = 211$; (b) $M_{min} = 4$, $N_A = 33$.

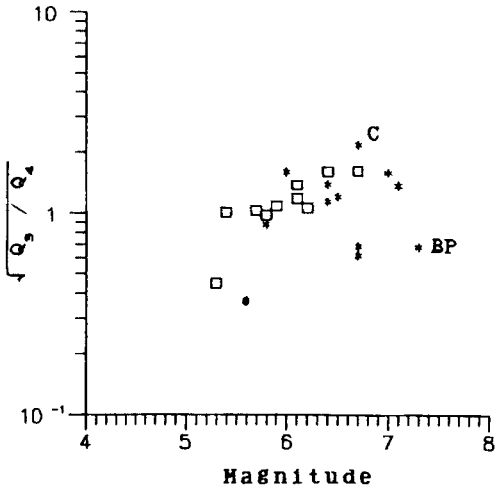

Figure 9. Linear coefficient of extension of the aftershock zone (κ, see text) as a function of mainshock magnitude (West US catalogue). Squares marks aftershock sequences with deficiency in low magnitudes. C and BP mark the Coalinga and Borah Peak earthquakes respectively.

Following Das & Scholz (1981) one might expect that the distribution of φ^* is concentrated near zero, i.e., the mainshock epicentre is predominantly in line with the longer axis of the aftershock zone. Our results obtained by the minimax LIR method are in contradiction with this assumption.

(v) It is a common belief that the number of aftershocks $N_A(M)$ and aftershock zone area $Q_A(M)$ grows exponentially with mainshock magnitude. This is confirmed by Fig. 10 (NOAA catalogue, $M \geq 4$). Figs 10(a) and (b) show the relationship between respective exponent values β_N and β_Q for the two quantities. The number of aftershocks per unit aftershock area for the 95 per cent level (see Fig. 11) shows a slight negative trend with magnitude, $\beta_N - \beta_Q = 0.19$. We estimated $\beta_Q - \beta_N$ under a Poissonian distribution of N_A taking into account the scatter in magnitude ($\sigma_M = 0.2$) and in $\log(N_A/Q_A)$ (see Appendix C). The scatter in $\log(N_A/Q_A)$ is large and should be analysed in detail for each region separately in order to test the interesting hypothesis $\beta_N \cong \beta_Q$ of the self-similarity of the seimic process. So far, we can say that our data for $M = 6$–7 and $M > 7$ are fairly uniformly distributed over the globe, making the decisive effect of any one region on the trend (Fig. 11) unlikely.

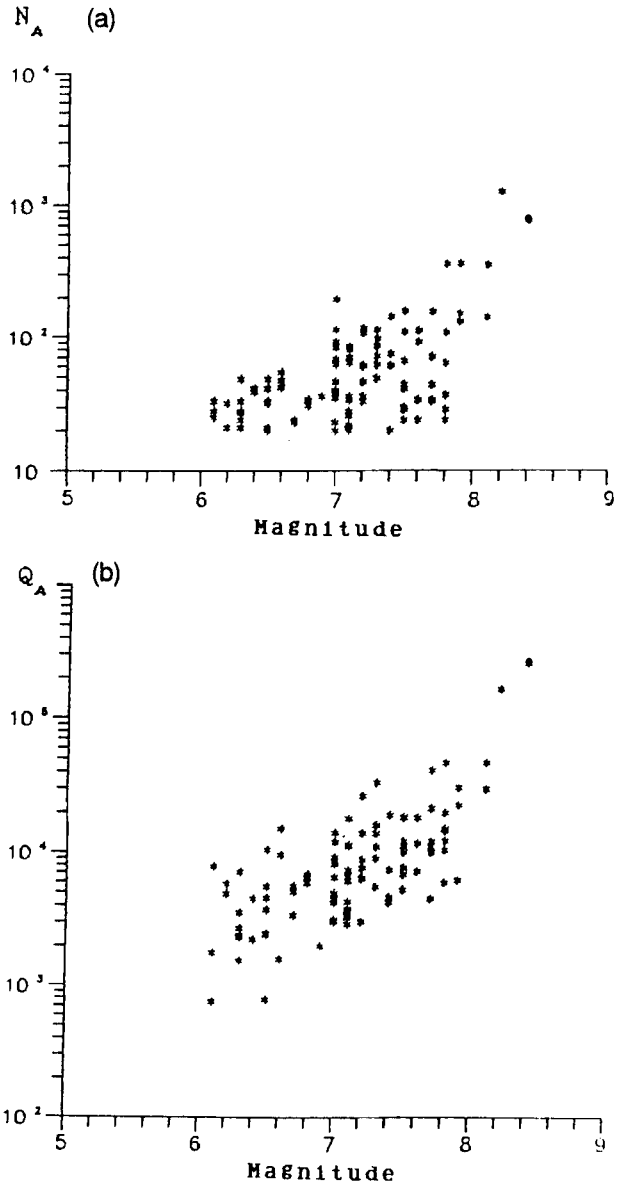

Figure 10. (a) The number of minimax aftershocks N_A (≥ 20) and (b) aftershock zone area Q_A (ellipse of 95 per cent confidence level) as a function of mainshock magnitude (NOAA catalogue, 1964–1980).

7 CONCLUSIONS

The definition of aftershock events, especially of later ones, is fuzzy and often governed by the problem under consideration. Consequently, before comparing different aftershock identification methods we should fix both the goals and the definitions. In the present work we give preference to hand techniques and the well-known relations for aftershock distribution in space and time.

Aftershock identification is formulated as a problem of minimization of a loss function γ associated with the number of missed aftershocks and the number of background events erroneously identified as aftershocks. The loss-function approach produces a set of aftershock identification methods (LIR methods) depending both on the loss function and on the aftershock distribution model assumed. LIR methods are very simple and well adapted to local seismicity. The methods could be of value for idenification of large aftershock sequences localized in space and time. In these cases it is natural to identify fore- and aftershock events simultaneously.

Sequences with high 'signal-to-noise' ratios usually follow larger earthquakes. For this reason, applications of any aftershock identification method are restricted by the mainshock magnitude range. Unfortunately this circumstance is not discussed in the literature. As a result, different phenomena can be confused, say, real aftershock events and earthquakes corresponding to local interaction at

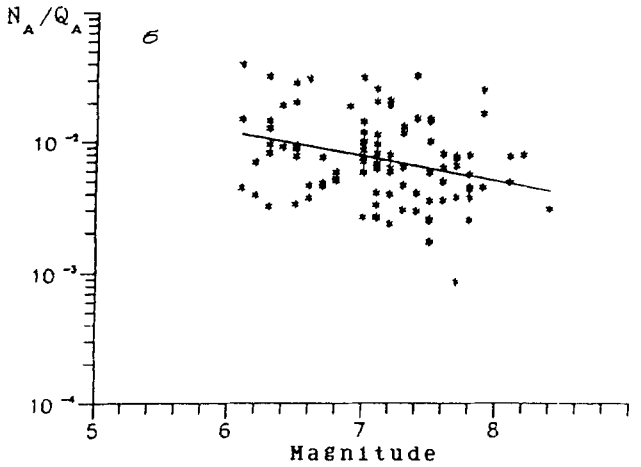

Figure 11. The number of minimax aftershocks per unit of aftershock area (95 per cent) as a function of mainshock magnitude with regression line (notations are the same as in Fig. 9).

the stage of earthquake preparation. But different phenomena require different methods of investigation.

Finally we would like to emphasize unexpected parallels in two totally different problems: aftershock identification and earthquake prediction (see Molchan 1991). In both cases only the loss-function approach permits one to compare different algorithms. The solutions of the two problems appear to have much in common and are derived by means of hypothesis testing theory.

ACKNOWLEDGMENTS

We are very grateful to USGS for use of the US catalogue and also to the editor Dr B. Minster, Dr Y. Kagan and two anonymous referees for their useful recommendations and remarks which significantly improved the manuscript.

REFERENCES

Balakina, L., 1989. Great earthquakes of 1952, 1958, 1969 in the lithosphere of south part of Kuril Arc, *Izvestia Acad. Sci., Phys. Solid Earth*, **2**, 3–16.

Borovkov, A., 1986. *Probability Theory*, Nauka, Moscow.

Daley, D. J. & Vere-Jones, D., 1988. *An Introduction to the Theory of Point Process*, Springer-Verlag, New York.

Das, S. & Scholz, C. H., 1981. Off-fault aftershock clusters caused by shear stress increase?, *Bull. seism. Soc. Am.*, **71**, 1669–1675.

Davis, S. D. & Frohlich, C., 1991a. Single-link cluster analysis, synthetic earthquake catalogues, and aftershock identification, *Geophys. J. Int.*, **104**, 289–306.

Davis, S. D. & Frohlich, C., 1991b. Single-link cluster analysis of earthquake aftershocks: decay laws and regional variations, *J. geophys. Res.*, **96**, 6335–6350.

Feller, W., 1966. *An Introduction to Probability Theory and its Applications*, vol. 2, Wiley, New York.

Frohlich, C. & Davis, S. D., 1985. Identification of aftershocks of deep earthquakes by a new ratios method, *Geophys. Res. Lett.*, **12**, 713–716.

Gardner, J. & Knopoff, L., 1974. Is the sequence of earthquakes in Southern California with aftershocks removed Poissonian? Yes, *Bull. seism. Soc. Am.*, **64**, 1363–1367.

Gnanadesikan, R. & Kettenring, J., 1989. Discriminant analysis and clustering, *Stat. Sci.*, **4**, 34–69.

Kagan, Y. & Knopoff, L. 1976. Statistical search for non-random features of the seismicity of strong earthquakes, *Phys. Earth planet. Inter.*, **12**, 291–318.

Kagan, Y. & Knopoff, L., 1981. Stochastic synthesis of earthquake catalogs, *J. geophys. Res.*, **86**, 2853–2862.

Kanamori, H., 1977. The energy release in great earthquakes, *J. geophys. Res.*, **82**, 2981–2987.

Kanamori, H. & Anderson, L., 1975. Theoretical basis of some empirical relations in seismology, *Bull. seism. Soc. Am.*, **65**, 1073–1095.

Keilis-Borok, V. & Kosobokov, V., 1986. Time of increased probability for great earthquakes of the world, *Comp. Seism.*, **19**, 48–58, eds Keilis-Borok, V. & Levshin, A., Nauka, Moscow.

Keilis-Borok, V., Knopoff, L. & Rotwain, I., 1980. Bursts of aftershocks long term precursors of strong earthquakes, *Nature*, **283**, 259–263.

Knopoff, L., Kagan, Y. & Knopoff, R., 1982. *b*-values for foreshocks and aftershocks in real and simulated earthquake sequences, *Bull. seism. Soc. Am.*, **72**, 1663–1675.

Molchan, G., 1991. Structure of optimal strategies in earthquake prediction, *Tectonophysics*, **193**, 267–276.

Molchan, G. & Dmitrieva, O., 1990. Dynamics of the magnitude–frequency relation for foreshocks, *Phys. Earth planet. Inter.*, **61**, 99–112.

Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, **83**, 9–27.

Prozorov, A., 1986. Dynamic algorithm for removing aftershocks from the world earthquake catalog, *Comp. Seism.*, **19**, 58–62, eds Keilis-Borok, V. & Levshin, A., Nauka, Moscow.

Prozorov, A. & Dziewonski, A., 1982. A method of studying variations in the clustering property of earthquakes: application to the analysis of global seismicity, *J. geophys. Res.*, **87**, 2829–2839.

Prozorov, A. & Schreider, S., 1986. A statistical analisis of positive influence for normal earthquakes in Tian-Shan and Pamir-Alay, *Comp. Seism.*, **19**, 37–47, eds Keilis-Borok, V. & Levshin, A., Nauka, Moscow.

Reasenberg, P., 1985. Second-order moment of Central California seismicity, 1969–1982, *J. geophys. Res.*, **90**, 5479–5495.

Reasenberg, P. & Jones, L., 1989. Earthquake hazard after a mainshock in California, *Science*, **243**, 1173–1176.

Savage, W., 1972. Microearthquake clustering near Fairview Peak, Nevada, and in the Nevada Seismic Zone, *J. geophys. Res.*, **77**, 7049–7056.

Utsu, T. & Seki, A., 1954. A relation between the area of aftershock region and the energy of main shock, *J. seism. Soc. Japan*, **7**, 233–240.

Vere-Jones, D., 1976. A branching model for crack propagation, *Pure appl. Geophys.*, **114**, 711–725.

Vere-Jones, D. & Davies, R., 1966. A statistical survey of earthquakes in the main seismic region of New Zealand. Part II. Time series analysis, *New Zealand J. Geol. Geophys.*, **9**, 251–284.

APPENDIX A: A QUANTITATIVE ANALYSIS OF PROZOROV'S METHOD

Let $\{t_i\}$ be a sequence of events in the area S after the mainshock $t_0 = 0$. Prozorov's method with parameters (α, R) identifies $\{t_1 \cdots t_\nu\}$ as aftershocks if

$$\hat{\lambda}(t_i) \geq R\lambda_b, \qquad 1 \leq i \leq \nu, \tag{16}$$

and the inequality is broken for $i = \nu + 1$. Here

$$\hat{\lambda}(t) = \#\{t_j : \alpha t \le t_j \le t\}/(t - \alpha t), \qquad 0 \le \alpha < 1,$$

is an empirical estimate of the intensity of the point process $\{t_i\}$ for moment t and λ_b is the rate of background seismicity in the area S. The intensity of aftershocks is changed weakly after a short time period. Therefore it is natural to analyse the method in the simplest situation when the intensity of $\{t_i\}$ is constant.

Statement 2

Assume that a mixture of two Poissonian flows with intensities λ_b and λ_A is observed in the area S. Then Prozorov's procedure with parameters (α, R) will classify a random number of events ν with the distribution $P(\nu = n) = p_n(\mu)$, $n \ge 0$, to be given below. Here $\mu = R_0/R$, $R_0 = (\lambda_A + \lambda_b)/\lambda_b$ and λ_A/λ_b is an actual 'signal-to-noise' ratio.

In the case $\alpha = 0$, the random quantity ν has the following distribution:

$$p_n = \frac{[\mu(n+1)]^n}{(n+1)!} e^{-(n+1)\mu} = \frac{e^{1-\mu}}{\sqrt{2\pi}} \frac{(\mu e^{1-\mu})^n}{(n+1)^{3/2}} [1 + O(1)], \qquad n \to \infty. \tag{17}$$

When $\mu > 1$ the probability $p_\infty \neq 0$ is given as a root of

$$p\mu + \ln(1-p) = 0, \qquad 0 < p < 1. \tag{18}$$

In the case $\alpha = 1/2$, the probabilities p_0 and p_1 are given by (17). p_2 and p_3 are

$$\begin{aligned} p_2 &= 0.5\mu \exp(-2\mu)[1 + \mu - \exp(-2\mu)], \\ p_3 &= 0.5\mu \exp(-2\mu)[(1+\mu)/2 + \mu^2/3 - (1 + 2\mu - 4\mu^2)\exp(-2\mu)/2]. \end{aligned} \tag{19}$$

Note that:

(i) It is interesting that (17) is also the distribution of the Poissonian population with parameter μ. By definition this population starts from one element and consists of offspring generated at all steps of reproducing: each element independently generates a random number of elements which obey the Poisson distribution with parameter μ. Each element generates its offspring only once.

(ii) Asymptotics of the type (17) for $\mu = 1$: $p_n \propto n^{-\theta}$, $\theta = 3/2$, are universal for critical Galton–Watson branching processes. It was used by Vere-Jones (1976) to estimate the b-value ($b \cong 0.75$) in the freqeuncy–magnitude relation for rock fracture. An element was associated with a microcrack and the population with a crack, the population size being proportional to the total fracture energy.

(iii) It is convenient to analyse distributions (17, 19) by means of cumulative probabilities $P(\nu \ge n)$ shown in Table A1 for $\alpha = 0$ and 0.5. To derive practical recommendations from this table we should keep in mind that in reality the aftershock intensity λ_A, signal-to-noise ratio $R_0(t) = \lambda_A/\lambda_b + 1$, $\mu = R_0/R$ constantly decreases and that Prozorov's procedure with $\alpha = 1/2$ constantly neglects aftershocks from the earlier half of the current time interval. Therefore, the distributions of ν can characterize the increment of the number of aftershock events under varying conditions.

As follows from Table A1, the distributions of ν with $\alpha = 0$ and 1/2 are close when $R/R_0 > 5$. For $\alpha = 1/2$,

$$P(\nu > 3) \le 5 \text{ per cent}, \qquad R/R_0 > 2.$$

Table A1. The distribution $P\{\nu \ge n\}$. One hundred per cent of the number of aftershocks for Prozorov's method with various values of threshold R (columns a and b correspond to parameter $\alpha = 0$ and $\alpha = 0.5$ respectively).

	Normalized threshold R /R0*											
n	1		2		5		10		15		20	
	a	b	a	b	a	b	a	b	a	b	a	b
1	63	63	39	39	18	18	9.5	9.5	6.4	6.4	4.9	4.9
2	50	50	21	21	4.7	4.7	1.3	1.3	.6	.6	.35	.35
3	42	37	13	10	1.4	1.2	.22	.18	.07	.06	.03	.02
4	37	28	8	5	.5	.3	.04	.02	.008	.005	.003	.002

* $R_0 - 1$ is actual 'signal-to-noise' ratio.

Proof: the case $\alpha = 0$, equation (17)

Let $t_0, t_1, t_2, \ldots$ be the realization of a Poissonian process with intensity $\lambda = \lambda_b + \lambda_A$. According to (16) the number of events ν which were identified after the moment $t_0 = 0$ is equal to n, if

$\lambda t_1 < c_1, \lambda t_2 < c_2, \ldots, \lambda t_n < c_n, \lambda t_{n+1} > c_{n+1}$

where

$c_n = \lambda n/\lambda_b R = \mu n$

and

$P(\nu \geq n) = F_n = Pr\{\lambda t_1 < c_1, \ldots, \lambda t_n < c_n\}$.

Suppose for the moment that $\{c_i\}$ are arbitrary, $c_i < c_{i+1}$. Then

$$F_n = \int_{A_n} \lambda^n \exp\left(-\lambda \sum_1^n x_i\right) d^n x \tag{20}$$

where

$A_n = \{x_i > 0 : x_1 < c_1/\lambda, \ldots, x_1 + \cdots + x_n < c_n/\lambda\}$.

Integrating over x_n leads to

$$F_n = F_{n-1} - e^{-\lambda c_n} \lambda^{n-1} J_{n-1}, \tag{21}$$

$$J_n = \int_{A_n} d^n x. \tag{22}$$

From (21),

$$p_{n-1} \equiv F_{n-1} - F_n = e^{-\lambda c_n} \lambda^{n-1} J_{n-1} \tag{23}$$

then

$F_n = 1 - e^{-\lambda c_1} J_0 - \cdots - e^{-\lambda c_n} J_{n-1} \lambda^{n-1}$.

From (20),

$$\lim F_n/\lambda^n = J_n, \qquad \lambda \to 0. \tag{24}$$

Then from (24) for all $1 \leq p \leq n$

$$J_p - \frac{(c_p)^1}{1!} J_{p-1} + \frac{(c_{p-1})^2}{2!} J_{p-2} - \cdots - (-1)^p \frac{(c_1)^p}{p!} J_0 = 0, \tag{25}$$

where $J_0 = 1$.

Now consider $c_n = \mu n$ and $J_n = \dot{J}_n$ when $\mu = \lambda$. From (22) it follows that

$J_n = (\mu/\lambda)^n J_n$.

We now verify that $\dot{J}_n = (p+1)^p/(p+1)$. Substitute $\dot{J}_n$ in (25). Multiplying by $(p+1)!$ we have

$$\sum_{k=0}^{p+1} C_{p+1}^k (p+1-k)^p (-1)^k = 0$$

or

$$\sum_{k=0}^{p} C_p^k k^{p-1} (-1)^k = 0.$$

The last sum is $(d/dx)^{p-1}(1 - e^x)^p\big|_{x=0}$. It is equal to zero, since $(1 - e^x)^p = \mathcal{O}(x^p)$, $x \to 0$.

Probability $p^* = P(\nu = \infty)$

From the random walk theory (Feller 1966) it follows that $p^* = 0$, when $\mu < 1$. This is proved below. We have

$$p^* = p\left[\max_n (\tau_1 + \cdots + \tau_n) < 0\right]$$

where τ_i are independent with density $e^{-(x+\mu)}$, $x \geq -\mu$ and mean $1 - \mu > 0$. Hence

$$\sum_{n \geq 0} p_n(\mu) = 1, \qquad \mu < 1. \tag{26}$$

When $\mu \neq 1$ the elements of the series decrease faster than the geometrical progression. Thus (26) is true for $\mu = 1$ (because of the continuity of the series), i.e. $p^* = 0$ when $\mu = 1$.

The function $y = \mu e^{-\mu}$ is unimodal with mode $\mu = 1$, i.e., two values of μ correspond to a value of y ($\mu < 1$ and $\mu^* > 1$). By virtue of (26), when $\mu < 1$

$$\mu = \sum_{n\geq 0} p_n(\mu)\mu = \sum_{n\geq 0} [y(\mu)]^{n+1} a_n = \sum_{n\geq 0} [y(\mu^*)]^{n+1} a_n = \sum_{n\geq 0} p_n(\mu^*)\mu^*.$$

Hence

$$\sum_{n\geq 0} p_n(\mu^*) = \mu/\mu^*, \qquad \mu^* > 1. \tag{27}$$

Values μ and μ^* are related by

$$\mu e^{-\mu} = \mu^* e^{-\mu^*}.$$

If $p^* = p(\nu = \infty)$, then $1 - p^* = \mu/\mu^*$ and (27) leads to

$$1 - p^* = \exp(-p^*\mu^*) \quad \text{or} \quad \mu = 1/p^* \ln \frac{1}{1-p^*}.$$

Cumulant function of ν

$$\psi(\theta) = \ln E e^{-\theta\nu} = \sum_{k\geq 1} \frac{(-\theta)^k}{k!} \kappa_k,$$

where κ_k are cumulants of ν. Let $\mu < 1$. From (26) we have

$$e^{\mu} = \sum_{n\geq 0} (\mu e^{-\mu})^n a_n, \qquad a_n = (n+1)^n/(n+1)!$$

and

$$e^{\psi(\theta)+\mu} = \sum_{n\geq 0} (\mu e^{-(\mu+\theta)})^n a_n = e^{\tilde{\mu}(\theta)}$$

if

$$\mu e^{-(\mu+\theta)} = \tilde{\mu} e^{-\tilde{\mu}}, \qquad \tilde{\mu} < 1.$$

Thus

$$\left.\begin{array}{l} \psi(\theta) = \tilde{\mu} - \mu \\ \ln \tilde{\mu}/\mu = \tilde{\mu} - \mu - \theta \end{array}\right| \rightarrow \ln\left(\frac{\psi(\theta)}{\mu} + 1\right) = \psi(\theta) - \theta$$

or

$$\frac{1}{\mu}\psi(\theta) + 1 = E e^{-\theta(\nu+1)}. \tag{28}$$

This relation allows one to obtain cumulants of ν:

$$\mu^{-1}\kappa_k = E(\nu+1)^k.$$

On the other hand (28) is equivalent to the equality $\varphi(\theta) = E e^{-\theta\nu}$:

$$\varphi(\theta) = \exp\{\mu[\varphi(\theta)e^{-\theta} - 1]\} = \sum_{n\geq 0} \frac{\mu^n}{n!} e^{-\mu} (E e^{-\theta(\nu+1)})^n \tag{29}$$

i.e., ν has the same distribution function as for the size of Poissonian population with parameter μ, because

$$\nu = \nu_1 + \cdots + \nu_\pi + \mu$$

where π is the number of elements in the first generation, and ν_i are population sizes for each branch, generated by the first generation elements. Equation (29) expresses this. Thus the correspondence of (17) to the Galton–Watson branching process is established.

The case $\alpha = 1/2$

In this case probabilities p_n are obtained by a straightforward and rather tiresome calculation. The case $n = 0$ is trivial:

$$p_0 = p(\tau_1 > \mu) = e^{-\mu}, \quad \text{where } \tau_i = \lambda(t_i - t_{i-1}).$$

For $n = 1$ we have

$$p_1 = P(\lambda t_1 < \mu,\ \lambda t_2/2 < t_1,\ \lambda t_2/2 > 2\mu) + P(\lambda t_1 < \mu,\ t_1 < t_2/2,\ \lambda t_2 > \mu) = P(\tau_1 < \mu,\ \tau_1 + \tau_2 > 2\mu) = \mu e^{-2\mu}.$$

Similarly

$$p_2 = P[\Omega_2(A_1 \cup A_2 \cup A_3]$$

where

$$\Omega_2 = \{\tau_1 < \mu,\ \tau_1 + \tau_2 < 2\mu\}, \quad A_1 = \{t_3/2 < t_1,\ \lambda t_3/2 > 3\mu\}, \quad A_2 = \{t_1 < t_3/2 < t_2,\ \lambda t_3/2 > 2\mu\}, \quad A_3 = \{t_3/2 > t_2,\ \lambda t_3/2 > \mu\},$$

or

$$p_2 = P[\tau_1 + \tau_2 < \rho,\ \tau_3 > 2\rho - (\tau_1 + \tau_2)] + P(\tau_1 < \rho,\ \rho < \lambda t_2 < 2\rho,\ \tau_3 > \lambda t_2).$$

Subsequential manipulations are based on the independence of τ_1, τ_2 and τ_3, and can be easily carried out by using conditional probabilities.

APPENDIX B: PROOF OF STATEMENT 1

Let (g_i, t_i) be a realization of a mixture of two independent Poissonian flows within GT with intensities $\Lambda_b(g, t)$ and $\Lambda_A(g, t)$. The realization could be obtained as follows. The number of points $\nu = n$ within GT is sampled according to the probability $p_n = \Lambda^n e^{-\Lambda}/n!$, where $\Lambda = \Lambda_b + \Lambda_A$ is the average number of events in GT. Each point is sampled independently according to one of the two probabilities: either $P_b(g, t) = \Lambda_b(g, t)/\Lambda_b$ or $P_A(g, t) = \Lambda_A(g, t)/\Lambda_A$, the choice of the distribution being specified by random guess ξ with probabilities $\mu_b = \Lambda_b/\Lambda$ and $\mu_A = \Lambda_A/\Lambda$ for the outcomes b and A. In other words, given $\nu = n$, we perform n independent Bayesian trials realizing one of the two distributions for the point (g, t): P_b (hypothesis H_b) or P_A (hypothesis H_A). Hence we should construct a decision rule for each trial in favour of one of the hypotheses. As the trials are conditionally independent it is sufficient to consider conditionally independent decision rules.

Let $\pi(g, t)$ be the identification rule for a single point. It refers a point either to 'b' or to 'A' in accordance with some distribution. If α and β are the losses, then the loss for a decision in the game principle is

$$\alpha P\{\xi = A,\ \pi = B\} + \beta P\{\xi = B,\ \pi = A\} = \alpha\mu_A P(\pi = B \mid H_A) + \beta\mu_b P(\pi = A \mid H_b).$$

Thus we derive a linear loss function in the problem of testing hypotheses H_A and H_b. The optimal decision rule for this function is described by the critical acceptance zone for H_A:

$$P_A(g, t)/P_b(g, t) > \beta\mu_b/\alpha\mu_A$$

(Borovkov 1986). In our notation we get

$$L = \Lambda_A(g, t)/\Lambda_b(g, t) \geq \beta/\alpha.$$

The optimal loss is independent of n. Therefore the optimal rule obtained holds in the unconditional case too, i.e. for arbitrary ν. The minimax loss function can be reduced to the optimal testing of hypotheses H_A and H_b with respect to the losses

$$\max[\alpha\mu_A P(\pi = B \mid H_A), \beta\mu_b P(\pi = A \mid H_b)].$$

The problem is traditional for $\alpha = \beta$ (Borovkov 1986). For the case of arbitrary loss function ν, see, for example, Molchan (1991). The case of minimax losses leads to the equation $\alpha\Lambda_A^+ = \beta\Lambda_A^- = \min$. The solution is given in Statement 1.

Consider the general case of non-Poissonian structure. When choosing the critical zone as given by (15) (events within the zone are identified as aftershocks) the average number of missed aftershocks can be written as

$$\Lambda_A^+ = \int\left[\int_{r(g-g_*)>k(t)} \Lambda_A p(g)\, dg\right] f(t)\, dt = \Lambda_A \int_0^\infty \left[\int_{k(t)}^\infty p(r) r\, dr\right] f(t)\, dt.$$

On the other hand, assuming a uniform distribution for background seismicity, the average number of erroneously identified aftershocks is

$$\Lambda_A^- = \lambda_b |S_1| \int_0^\infty k^2(t)\, dt.$$

Thus the problems (6) and (7) are reduced to simple variation problems with respect to $k(t)$:

$$\alpha\Lambda_A^+ + \beta\Lambda_A^- = \min_{k(t)} \quad \text{and in case (7)} \quad \Lambda_A^- = \min_{k(t)} \quad \text{giving} \quad \alpha\Lambda_A^+ = \beta\Lambda_A^-.$$

It is easy to see that the solution is equivalent to (8).

APPENDIX C: ORTHOGONAL REGRESSION WITH INHOMOGENEOUS ERRORS

The regressions presented by Figs 10(a), (b) and 11 fit the following model:

$$y_i = a + bx_i + \sigma_a \varepsilon_i^{(1)} + \sigma_i \varepsilon_i^{(2)}, \qquad z_i = x_i + \sigma_x \varepsilon_i^{(3)},$$

where (y_i, z_i), $i = 1, \ldots, N$ are observations, $(\varepsilon_i^{(1)}, \varepsilon_i^{(2)}, \varepsilon_i^{(3)})$ are uncorrelated errors with the average $(0, 0, 0)$ and unit covariance matrix. The quantity σ_a reflects regional scattering of a, the σ_i are associated with the accuracy of y_i estimation, and σ_x is the accuracy of the x_i measurement. σ_i and σ_x are supposed to be known. Parameters (a, b, σ_a^2) should be estimated under nuisance parameters $\{x_i\}$. Under the Gaussian hypothesis for $(\varepsilon_i^1, \varepsilon_i^2, \varepsilon_i^3)$ the maximum likelihood method leads to the following procedure. Parameters (b, σ_a^2) are estimated by minimization of the functional

$$\chi^2 = 1/N \sum_{i=1}^{N} \left[\frac{(y_i - \hat{a} - bz_i)^2}{\sigma_i^2 + \sigma_a^2 + b^2\sigma_x^2} + \ln(\sigma_i^2 + \sigma_a^2) \right] + \ln \sigma_x^2,$$

where $\hat{a} = \sum_1^N p_i y_i - b \sum_1^N z_i p_i$ is a function of b and σ_a^2:

$$p_k = (\sigma_k^2 + \sigma_a^2 + b^2\sigma_x^2)^{-1} / \sum (\sigma_i^2 + \sigma_a^2 + b^2\sigma_x^2)^{-1}.$$

The quantity of $\hat{a}$ at optimal values of (b, σ_a^2) determines the estimation of a.

The regression $\log Q$ on magnitude (Fig. 10b): in this case $y_i = \log Q_i$, $z_i = M_i$, $\sigma_x = 0.2$ is the magnitude error. If the aftershock zone area Q_i was based on n points, then the variance of $\log Q_i$ is

$$\sigma_i^2 = \frac{(\log e)^2 n}{(n-1)^2},$$

and the bias is $-\log e/(n-1)$ (see Molchan & Dmitrieva 1990). Note the relation of unit aftershock zone area Q_1 with the area of $(1 - \varepsilon)$ confidence level:

$$\log Q^{(1-\varepsilon)} = \log Q_1 + \log \left[(\varepsilon^{-2/(n-1)} - 1) \frac{(n-1)^2}{n-2} \right].$$

When $\varepsilon = 5$ per cent, the second term is approximately equal to $0.7775 + 1.8\,(n-1)^{-1}$.

In the regression of N_A on magnitude we have $y_i = \log N_i$, $z_j = M_i$, $\sigma_x = 0.2$ and $\sigma_i^2 = (\log e)^2/N_i$.

At least, if $y_i = \log(Q_i/N_i)$, $z_i = M_i$, then $\sigma_x = 0.2$ and $\sigma_i^2 \cong 2(\log e)^2/N_i$.

In Molchan & Dmitrieva (1990) we did not take into account regional variation of parameter a ($\sigma_a = 0$), so the average regression line obtained there for $(\log Q_i, M)$ was formal. The correct regression line is the following: $\log Q^{(1-\varepsilon)}(\text{km}^2) = -1.05 + 0.69\,M$, where $\varepsilon = 5$ per cent (see also Fig. 1). It is obtained from data presented in Fig. 10(b) taking into account regional variation of a. However, one should also be careful in using this regression because of its regional averaging.