united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

H4.SMR/1150 - 10

Fifth Workshop on Non-Linear Dynamics
and Earthquake Prediction

4 - 22 October 1999

Some Statistical Problems in Earthquake Prediction and Seismicity

G. Molchan

International Institute of Earthquake Prediction
Theory and Mathematical Geophysics
Russian Academy of Sciences
Moscow 113556, Russian Federation

Pure appl. geophys. 149 (1997) 233–247
0033-4553/97/010233-15 $1.50 + 0.20/0

Pure and Applied Geophysics

Earthquake Prediction as a Decision-making Problem

G. M. MOLCHAN[1]

Abstract—In this review we consider an interdisciplinary problem of earthquake prediction involving economics. This joint research aids in understanding the prediction problem as a whole and reveals additional requirements for seismostatistics. We formulate the problem as an optimal control problem: Possessing the possibility to declare several types of alerts, it is necessary to find an optimal strategy minimizing the total expected losses. Losses include costs both for maintaining alerts and for changing alert types; each successful prediction prevents a certain amount of losses; total expected losses are integrated over the semi-infinite time interval. The discount factor is included in the model. Algorithmic and exact solutions are indicated.

This paper is based on the recent results by MOLCHAN (1990, 1991, 1992).

Key words: Earthquake prediction, prediction objective, prediction error diagram, hazard function, Bellman equation.

1. Introduction

Earthquake prediction is usually understood as a physical prediction, that is, deterministic localization of future strong events in time and space. At the same time, practical applications in intermediate-term and short-term predictions are based on stochastic features. This is reflected in statistical characteristics of prediction as well as in methods of interpretation of alarms (KEILIS-BOROK and ROTWAIN, 1990; KEILIS-BOROK and KOSSOBOKOV, 1990; NISHENKO, 1989). Therefore practical use of prediction constitutes an important part of the general problem.

In AKI's (1989) opinion, the general problem of prediction, including decisions and practical actions, must be considered separately by geophysicists and users (for example, economists). Similarly, these two parts of the problem were separated by KANTOROVICH and KEILIS-BOROK (1977; see also for short version in SADOVSKY, 1986). Unfortunately, we deal in practice with a number of prediction methods (algorithms) that are not of a very high quality. The number of such algorithms grows with time, complicating the situation. In fact, these algorithms are decision functions calling (or not calling) an alarm at a given point or in a region. In this

[1] International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Warshavskoye sh. 79 k. 2, Moscow 113556, Russia.

situation, two nonequivalent methods can lead (and do lead) to contradictory results. It is impossible to avoid this difficulty by choosing "the best" method in the frame of prediction physics (see below). Moreover, it is important to understand objective principles in prediction algorithms. These principles are vague because they are not clearly realized by the authors themselves or are based on artificial efficiency criteria (see below). Thus an expert is compelled to work with a system of ready (perhaps contradictory) decisions in no way associated with applications. Finally, in splitting the prediction problem, geophysicists do not know whether their results are sufficient for applications.

Here we make an attempt to develop a qualitative analysis of the prediction problem as a whole. The principle notion in this analysis is the prediction *objective*. The multiplicity of predictions turns from an obstacle to a favorable base to choose the best decision. Below we investigate two models of loss functions. The first model is important for most practical prediction algorithms; it is useful in the research stage of prediction (we are now just in this stage). The second model roughly simulates prediction economics. In both cases we find the structure of predictions that optimizes loss functions under conditions of prediction information $\mathcal{I}(t)$ of a very general nature. We found that in complicated cases optimal prediction is based on two entities: first, the conditional (relative to $\mathcal{I}(t)$) earthquake flow rate and second, the matrix of transition probabilities for the states of information $\mathcal{I}(t)$ in consecutive time intervals. Most recent investigations involve evaluation of the first entity, using combinations of predictors. As far as we know, the second entity has not yet been studied, though numerous descriptions of preparatory processes before strong earthquakes can constitute a base for their study.

This paper reviews recent results by the author (MOLCHAN, 1990, 1991, 1992). In the first section we consider the academic prediction type with two alert states, yes/no. The second section presents the analysis with an arbitrary number of alert states. Optimizing mean (discounted) losses associated with prediction, we obtain a Bellman-type equation. This part of the paper helps to clarify which statistical parameters are useful in the problem of prediction as a whole.

2. The Simplest Prediction Problem

2.1. Prediction Errors Diagram

Recently the author (MOLCHAN, 1990, 1991) presented the problem of comparing prediction methods for stationary point processes (that is, a sequence of strong earthquakes in a region). The problem was solved using two prediction parameters: $\hat{n}$, the rate of failures-to-predict (the number of missed events divided by the total number of events in the time interval $T \gg 1$), and $\hat{\tau}$, the rate of time alarms (the total time of alarms divided by T when $\gg 1$).

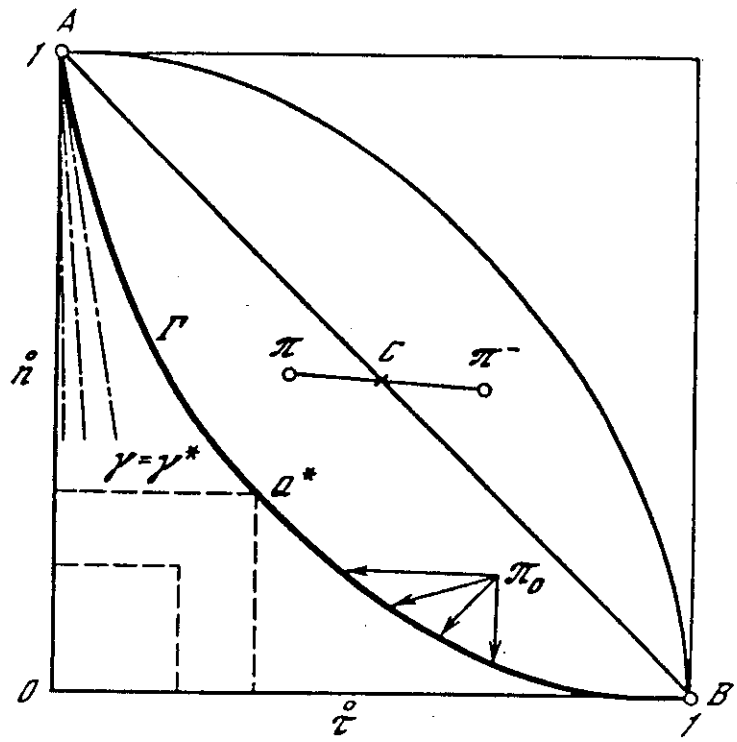

Figure 1

Error set $(\hat{n}, \hat{\tau})$ for prediction strategies based on the same information. (A) Optimist strategy. (B) Pessimist strategy. (AB) Results of a random guess. (C) The center of symmetry; π and π^- are a forecast and its antipodal forecast. Γ is the diagram of optimal prediction errors. Arrows indicate a better forecast relative to the strategy π_0. Dashed lines are contours of the loss function $\gamma = \max(\hat{n}, \hat{\tau})$. Q^* are errors of the minimax strategy, $\hat{n} = \hat{\tau}$. Dash-dotted lines are contours of losses $\nu = \hat{\tau}/(1 - \hat{n})$.

Let us agree on the type of information $\mathcal{J}(t)$ available at the moment t for the prediction of events in the point process. In practice $\mathcal{J}(t)$ can include catalogs of events of various magnitudes in the region under study, data on physical fields, and data on predictors in some time intervals $(t - t_i, t - \tau_i)$, where τ_i is the delay of the ith data type. In the simplest case the observer uses information $\mathcal{J}(t)$ and makes the decision $\pi(t)$: to declare or not to declare an alarm in the time interval $(t, t + \Delta)$ where Δ can equal the period of information renewal. The set of decisions $\{\pi(t)\} = \pi$ is called the prediction *strategy*. In practice the strategy is defined by the method or by the prediction algorithm. It is useful to consider class π of strategies where decisions can be made with some probabilities, that is, after an additional test of coin-tossing type with outcome probabilities $(p, 1 - p)$ depending on $\mathcal{J}(t)$. In practice deterministic solutions are usually preferred, where $p = 0$ or $p = 1$.

Any two strategies π_1 and π_2 of the type considered can be combined into a new strategy that independently uses π_1 or π_2 with probabilities q and $1 - q$ in each time interval Δ. This leads to a mixture of parameters $(\hat{n}, \hat{\tau})_i$ of initial strategies with the same weights q and $1 - q$. Hence the set $G = \{(\hat{n}, \hat{\tau})_\pi\}$ of errors corresponding to various strategies is convex if these strategies are based on the same information $\mathcal{J}(t)$ (Figure 1).

By definition, the set G belongs to the unit square $(0 \leq \mathring{n}, \mathring{\tau} \leq 1)$. It contains points $(1, 0)$ and $(0, 1)$ and, by convexity of G, the diagonal $\mathring{n} + \mathring{\tau} = 1$. The first point stands for the widespread optimist strategy in which an alarm is never declared. The second point corresponds to the total pessimist strategy in which the continuous alarm is kept. Points on the diagonal $\mathring{n} + \mathring{\tau} = 1$ correspond to the strategy of a random guess in which an alarm is declared with probability p independent of $\mathcal{J}(t)$.

The set G has the center of symmetry $(1/2, 1/2)$, because every prediction corresponds to the antipodal predition π^- where an alarm and not-an-alarm change places and errors $(\mathring{n}, \mathring{\tau})$ are replaced by $(1 - \mathring{n}, 1 - \mathring{\tau})$. Therefore all points of G above the diagonal $\mathring{n} + \mathring{\tau} = 1$ correspond to strategies constructed by negation of nontrivial strategies with $\mathring{n} + \mathring{\tau} < 1$. However, only nontrivial strategies at the lower boundary Γ of the set G are important. The boundary Γ connects the points $(1, 0)$ and $(0, 1)$. It is monotonic and convex downward due to the properties of G.

The points of Γ are incomparable, that is, if $\mathring{\tau}_1 < \mathring{\tau}_2$, then $\mathring{n}_1 \geq \mathring{n}_2$. For any point $(\mathring{n}, \mathring{\tau}) \in G$ there exists another point $(\mathring{n}_1, \mathring{\tau}_1) \in \Gamma$ where $\mathring{n}_1 < \mathring{n}$, $\mathring{\tau}_1 < \mathring{\tau}$, which corresponds to a better prediction. Therefore, there exists a minimum set of best and incomparable strategies among their total set. The number of these strategies is infinite, they are described by the error curve Γ.

The curve Γ is sufficient for the choice of the best strategy in the following problem. Suppose that the long-term losses associated with prediction can be expressed in terms of $\mathring{n}$ and $\mathring{\tau}$, that is, as a function $\gamma = \gamma(\mathring{n}, \mathring{\tau})$ increasing in its arguments. If sets $\{(\mathring{n}, \mathring{\tau}): \gamma < u\}$ are convex for any level u then the point Q^* where the contour line $\gamma = \gamma^*$ is touching Γ corresponds to the strategy minimizing γ (Figure 1).

Thus the general prediction problem (minimization of losses γ) in a class of loss functions $\gamma = \gamma(\mathring{n}, \mathring{\tau})$ is split into two independent problems. The first problem is the construction of the loss function γ, which falls in the area of economics or other studies. The second problem is the derivation of the curve Γ using the information $\mathcal{J}(t)$; here the physics of the seismic process is applied. In the latter case a union of points $(\mathring{n}, \mathring{\tau})$ should be analyzed; this union is generated by a variation of parameters in an algorithm and by applying various algorithms based on identical information to predict events of a certain magnitude. The boundary of the convex hull of these points serves as an estimate of the diagram Γ.

Efficiency of prediction is often measured by two parameters, $e_1 = (1 - \mathring{n})/\mathring{\tau}$ (GUSEV, 1976) and $e_2 = 1 - \mathring{n} - \mathring{\tau}$ (FENG *et al.*, 1985). Clearly, the most effective strategy is obtained with the minimum loss function $\gamma_1 = 1/e_1$ in the first case with $\gamma_2 = 1/e$ in the second case. Therefore the optimist strategy with errors $(1, 0)$ is the most effective in the first case. Indeed, γ_1 contour lines form a bundle of straight segments with the center $Q^* = (1, 0)$. The same strategy is the least effective in the second case because $e_2 = 0$. Thus the functions γ_1 cannot be used to measure the efficiency. Moreover, an attempt to choose a universal strategy from a continuum

of incomprable strategies is unsuccessful by itself. The loss function γ_2 is certainly useful for research purposes; however, the choice of γ falls, in general, out of prediction physics.

2.2. The Optimal Prediction Strategy

Consider the hazard function $r(t)$ that is, in other words, the conditional (relative to the information $\mathcal{I}(t)$) rate of predicted events

$$r(t) = P\{\text{there exists an event in } (t, t+\Delta) \mid \mathcal{I}(t) = u\}/\Delta = r_u.$$

We denote by λ the unconditional rate, that is, the number of strong events per unit time. Let us discretize time with the step Δ and denote by $n(t)$ the event indicator in the interval $(t, t+\Delta)$; if the interval Δ contains at least one event then $n(t) = 1$, otherwise $n(t) = 0$.

Statement 1. If the flow $(n(t), \mathcal{I}(t))$ is stationary and ergodic then there exists such threshold r^* depending on the loss function γ that the optimal prediction strategy is declaring an alarm every time when $r(t) > r^*$. In rare cases in which the relation $r(t) = r^*$ has a nonzero probability, an alarm is selected with some probability p^*.

If Q^* is the point where the contour line $\gamma = \gamma^*$ touches the error curve Γ, then the threshold r^* is expressed in terms of the common derivative of Γ and $\gamma = \gamma^*$ at that point

$$r^* = -\lambda \frac{d\mathring{n}}{d\mathring{\tau}}(Q^*).$$

If one of the curves is not differentiable at Q^*, then the derivative is changed for the slope of any common line of support for Γ and $\gamma = \gamma^*$ at Q^*. In the important case of linear loss function $\gamma = \alpha\lambda\mathring{n} + \beta\mathring{\tau}$ the threshold is $r^* = \beta/\alpha$, because the contour $\gamma = \gamma^*$ is a straight line; therefore it is a line of support with the slope $-\beta/\lambda\alpha$. This case was studied by LINDGREN (1985) and Ellis (1985). We can interpret α as prevented losses when the prediction is successful and β as the cost of maintaining an alarm per unit time; then γ stands for total losses per unit time.

2.3. Minimax Strategy

Another important case of the prediction strategy involves the loss function $\gamma_0 = \max(\mathring{n}, \mathring{\tau})$. The optimization of prediction leads in this case to the minimax strategy with $\mathring{n} = \mathring{\tau} = \min$. This strategy is useful when the loss ratio β/α in the linear function is unknown, so that an observer prefers the worst case $\beta/\lambda\alpha$ in the following sense:

$$\max_{\alpha,\beta} \min_{\pi} \frac{\alpha\lambda\mathring{n} + \beta\mathring{\tau}}{\alpha\lambda + \beta} = \min_{\pi} (\gamma_0).$$

This leads to a certain stability of the minimax strategy (MOLCHAN, 1990, 1991). Therefore it is no wonder that the minimax prediction principle is employed in practice, though inadvertently. For example, the algorithm CN (KEILIS-BOROK and ROTWAIN, 1990) yields $\mathring{n} \approx \mathring{\tau} = 25$–$30\%$, on average, for all the regions considered in the prediction of events with $M \geq 6.4$. Similarly, $\mathring{n} \approx \mathring{\tau} = 33\%$ for the M8 algorithm (KEILIS-BOROK and KOSSOBOKOV, 1990) in the prediction of events with $M \geq 7.5$ in the Circum-Pacific belt (private discussions with I. Rotwain and A. Khokhlov). Some simple precursors have a similar arithmetic mean of prediction errors, $(\mathring{n} + \mathring{\tau})/2$ (see, for example, the energy precursor by NARKUNSKAYA and SHNIRMAN, 1993). That means that the modern intermediate-term collective precursors (CN, M8, etc.) extend the geography of applications rather than leading to a higher quality of prediction. This hypothesis requires careful review.

The seismic gap hypothesis has recently been used in long-term forecasting (WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES, 1988). In this case the information used $\mathcal{J}(t)$ is the elapsed time since the last large event on a certain fault or plate boundary. Therefore the optimal strategy is defined by the rule:

$$\pi(\mathcal{J}(t) = u) = \text{alert} \quad \text{if} \quad r_u = \dot{F}(u)/(1 - F(u)) \geq r_* \tag{1}$$

where F is interevent time distribution.

If F has mean m and variance σ^2, and belongs to the Weibull, Gamma, Log-Normal type with a reasonable ratio $\sigma/m \in (0.25, 0.6)$, then the optimal minimax rule takes a simple explicit form:

$$\pi(\mathcal{J}(t) = u) = \text{alert} \quad \text{if} \quad u \geq k \cdot m, \quad k \simeq 0.75.$$

In addition, the errors $\mathring{n}, \mathring{\tau}$ are similar to the M8 algorithm errors i.e. $\mathring{n} \simeq \mathring{\tau} \leq 0.35$ (MOLCHAN, 1991).

The rule (1) is not used by the Working Group (WGCP, 1988), therefore its forecasting becomes more vulnerable to criticism. The recent discussion of the seismic gap hypothesis by KAGAN and JACKSON (1991) actually raises the issue of the distribution of F. Considering that earthquake times show clustering, it is necessary to use those distributions for which r_u is U-shaped or

$$F(u) = 1 - \exp\left(- \int_0^u U(\tau)\, d\tau \right)$$

where $U(\tau) \geq 0$, $U(\tau) \to \infty$ with $\tau \to 0$ and $\tau \to \infty$. Then the optimal alert times within an interevent period form two intervals: $(0, u_1)$ and (u_2, ∞). The first alert interval is a reaction on the clustering while the second one is in agreement with the gap hypothesis.

3. *Prediction with Multiphase Alerts*

The prediction model considered above is sufficiently general and yields a simple optimal strategy. It clearly divides the domain of activity into two parts: one is the province of geophysics (estimation of the hazard function); the other is related to economics (for example, estimation of the loss ratio β/α.) In the case of the linear loss function, the process $(n(t), \mathcal{J}(t))$ can even be nonstationary. The rejection of stationarity leads to a dependence of the threshold r^* on time. Indeed, the obtained prediction is optimal under linear losses per unit time both on the interval Δ (local optimality) and on the entire time axis (global optimality).

The simplest model considered is suitable for many types of practical forecasts that involve only two alert states: that is, where an alarm is declared or called off. However, a real alert must be multiphase, as a rule, because different degrees of hazard require different systems of protective measures (SADOVSKY, 1986). Hence we modify the prediction model by introducing multiphase alerts and generalized linear losses.

Let us assume that an observer can declare any alert from a given set of alerts $(A_0, A_1, \ldots, A_m)$ using the information $\mathcal{J}(t)$. The cancellation of an alert is included in the set; it is A_0. We also assume that every alert A_i requires cost β_i per unit time and that α_i is the prevented loss per one successful prediction. In particular, $\alpha_0 = \beta_0 = 0$ for A_0.

We assume that any change of alerts leads to loss c_{ij}, $0 \le c_{ij} < \infty$. The case $c_{ij} = \infty$ means that the change $A_i \to A_j$ is forbidden. For example, the population can be evacuated only after solving transportation problems. Whereas some of the protective measures require an ordering of corresponding alert types, other protective measures can be carried out in parallel. A block of such parallel measures is considered as a single measure in our model.

Nonzero c_{ij} values result in stability of alert sequences because they prevent fast alternating of alerts. However, the introduction of c_{ij} complicates the problem; locally, optimal decisions are not globally optimal in this case. Examples of this kind were discussed by MOLCHAN and KAGAN (1992).

Denote by z_t the losses associated with the decision $\pi(t)$; prevented losses enter there with the minus sign. Let us consider the total losses associated with the prediction strategy on the semiaxis $t > 0$ relative to the initial moment $t = 0$ with time factor ρ

$$Z_\pi = \sum_{k \ge 0} z_{k\Delta} \exp(-\rho \cdot k\Delta) = \sum_{k \ge 0} z_{k\Delta} \theta^k$$

$$\theta = \exp(-\rho\Delta). \tag{2}$$

Z_π is called discounted losses in the theory of optimal control (HOWARD, 1960; ROSS, 1970). In practical problems the factor ρ can stand for the efficiency of capital investments. Mathematically, the introduction of ρ allows us to consider the

problem on a finite interval of order $1/\rho$, escaping difficulties due to boundary effects when stationary prediction methods are studied.

The loss function now as the mean total discounted losses, that is, the prediction goal is the minimization of

$$S = EZ_\pi \to \min_\pi . \tag{3}$$

Consider the case in which the change of alert types does not lead to additional losses, that is, $c_{ij} = 0$ over all i and j.

Statement 2. Let $c_{ij} = 0$, $i, j = 1, 2, \ldots, m$, *and let* $\pi(t)$ *depend only on the stationary information sequence* $\mathcal{J}(t)$. *Then the optimal strategy is such that*

$$\pi(t) = A_{j^*} \tag{4}$$

where the subscript j^* *realizes the minimum*

$$\min_j (\beta_j - \alpha_j r(t)) = S[r(t)] \tag{5}$$

for the current value of the hazard function $r(t)$.

Remark 1. The function $S[r]$ in equation (5) is the convex polygonal envelope of the system of straight lines $y = \beta_j - \alpha_j r$ (Figure 2a). Let $P_0(0, 0)$, $P_1(r_1, y_1), \ldots, P_k(r_k, y_k)$ be the vertices of the polygon $S(r)$ that are ordered in r, $0 < r_1 < \cdots < r_k < r_{k+1} = \infty$, and let $j(n)$ be the number of the straight line $y = \beta_j - \alpha_j r$ with the pair of vertices P_n and P_{n+1}. The prediction strategy (4), (5) means that there exist $k \le m$ hazard levels $r(t):\{r_i\}$ such that the alert with $j^* = j(n)$ is always declared in the interval $r(t) \in (r_n, r_{n+1})$ (see Figure 2b). A number of alerts $\{A_i\}$ can be cost ineffective; such is the alert A_3 in Figure 2a.

2. The quantity (5) defines minimum conditional mean losses per unit time in the interval $(t, t + \Delta)$ under the given information $\mathcal{J}(t)$. Hence the strategy (4), (5) is simultaneously global and local. It does not depend on time factor ρ and is the generalization of the prediction strategy for two-phase alerts with the linear loss function discussed above.

To study the general case we introduce the following notion. We say that the process $(n(t), \mathcal{J}(t))$ *has* M property if the information sequence is a Markov process, that is,

$$P\{\mathcal{J}(t + \Delta) = v | \mathcal{J}(t) = u; \mathcal{J}(s), \forall s < t\} = P\{\mathcal{J}(t + \Delta) = v | \mathcal{J}(t) = u\} = P_{uv} \tag{6}$$

and that

$$E\{n(t) | \mathcal{J}(s), s \le t\} = E\{n(t) | \mathcal{J}(t)\} = r(t)\Delta. \tag{7}$$

Conditions (6) and (7) hold when the information $\mathcal{J}(t)$ contains all past data on predictors up to the moment t and all prehistory of the process $n(\cdot)$. In other

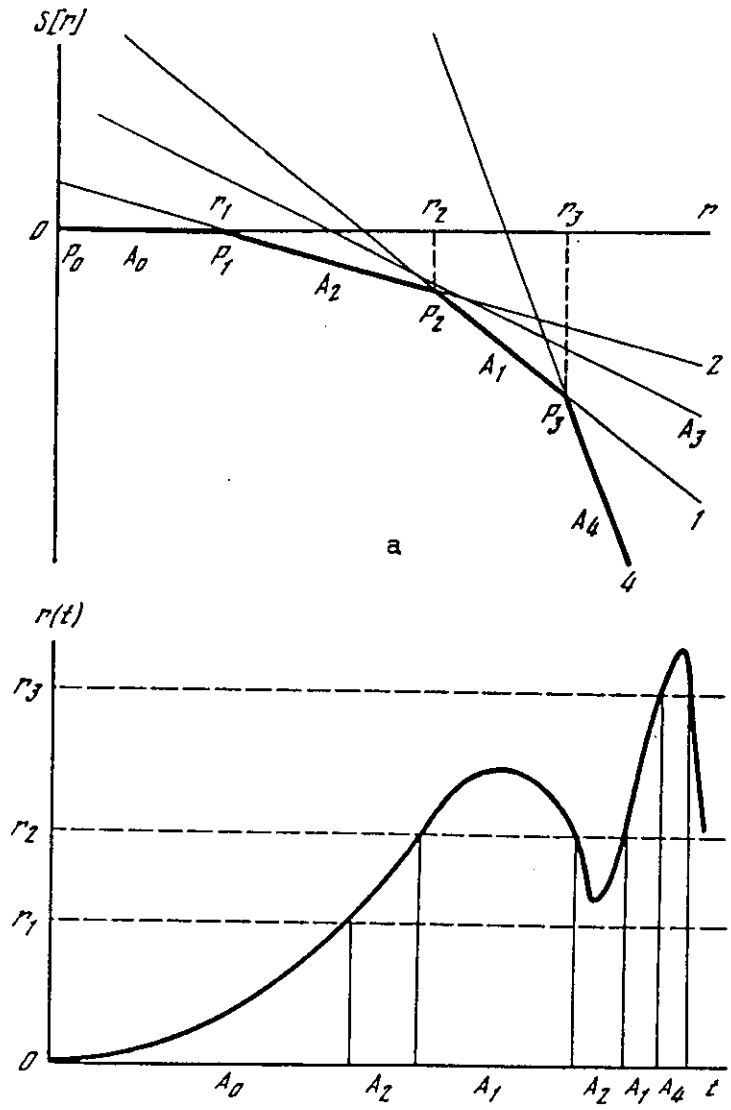

Figure 2

Optimal multiphase alert with zero losses for the change of phases. (a) Optimal mean losses per unit time $s(r)$ as a function of hazard level r; $s(r)$ is the envelope of straight lines $\beta_i - \alpha_i r$ indexed by alerts A_i; P_α are vertices of the envelope; r_α are hazard levels for the change of alerts A_i; the alert A_3 is not cost effective. (b) Hazard function $r(t)$ and optimal alert $A_{j(t)}$ as a time function under conditions of Figure 2a.

words, the past $\{\mathcal{J}(t), n(s), s < t\}$ is measurable relative to $\mathcal{J}(t)$. If the sequence $n(t)$ and a physical process $x(t)$ used to predict $n(t)$ has a finite memory, that is, a finite correlation interval J, then (6) and (7) are true for the information sequence $\mathcal{J}(t) = \{n(s), x(s), t - J < s < t\}$; its dimensionality is less as compared with the case of infinite memory.

Adding new requirements to the description of the process $(n(t), \mathcal{J}(t))$ we can

extend the class of decisions; $\pi(t)$ can depend on $\mathcal{I}(t)$ and past decisions $\pi(s)$, $s < t$. This dependence can be stochastic.

Statement 3. Assume that $(n(t), \mathcal{I}(t))$ *is stationary and has M property and that decisions* $\pi(t)$ *depend on* $\mathcal{I}(t)$ *and* $\{\pi(s), s < t\}$. *Then*

a) The optimal multiphase alert with parameters $[\alpha_i, \beta_i, c_{ij}]$ *in problem* (3) *exists and can be chosen stationary, that is, decisions* $\pi(t)$ *depend on t only in terms of the current information* $\mathcal{I}(t) = u$ *and current alert* $\pi(t - \Delta) = A_i$.

b) Minimal mean losses (3), $S^*(u, i)$, *under initial conditions* $\mathcal{I}(0) = u$ *and* $\pi(-\Delta) = i$, *are defined by the equation*

$$S(u, i) = \min_j [c_{ij} + \beta_j \Delta - r_u \Delta \alpha_j + \sum_v P_{uv} S(v, j)]. \tag{8}$$

We assume for simplicity that the set of states $\mathcal{I}(t)$ *is countable.*

c) If $\theta \in (0, 1)$ *or* $\rho > 0$ *then equation* (8) *has a unique solution. This solution can be found by an iterative procedure*

$$S^{(0)}(u, i) = 0; \quad S^{(n+1)} = T_0 S^{(n)} \tag{9}$$

where T_θ *is the operator defined by the right-hand side of* (8) *with the domain of functions* $S(u, i)$. *The error of the nth iteration is*

$$|S^{(n)} - S^*| < \theta^n (1 - \theta)^{-1} L$$

where

$$L = \max_{ij: c_{ij} < \infty} |c_{ij} + b_j \Delta + \alpha_j|$$

d) Under conditions $\mathcal{I}(t) = u$ *and* $\pi(t - \Delta) = A_i$ *the optimal solution is* $\pi(t) = A_{j^*}$ *where the subscript* $j^* = j(u, i)$ *minimizes the right-hand side of* (8).

Remarks. 1) Equation (8) is of Bellman type in the theory of optimal control (HOWARD, 1960; ROSS, 1970). In our case the control parameter j enters the loss function rather than transition matrix $[P_{uv}]$. 2) The recurrence (9) leads to the set of functions $S_k = S^{(N-k)}$, $k = 0, \ldots, n$, which are optimal mean discounted losses in the intervals $(k\Delta, N\Delta)$, $k = 0, 1, \ldots, N - 1$. The sequence of subscripts $j_k^*(u, i)$ minimizing (8) with $S = S_k$ defines the sequence of optimal decisions in intervals $(k\Delta, (k + 1)\Delta)$ under information states $(\mathcal{I}(t), \pi(t - \Delta))$, $t = k\Delta$.

The described algorithm is also suitable for optimization of total losses in the interval $(0, N\Delta)$ when $\theta = 1$, that is, without the time factor ρ. Unfortunately, the optimal prediction strategy for a finite time interval is nonstationary in the case $[c_{ij}] \neq 0$.

Let us consider two examples.

Renewal process. NISHENKO (1989) used the following model to predict charac-

teristic earthquakes. Interevent intervals are independent and have distribution $f(x)$; the information is the time $\mathcal{I}(t) = u$ elapsed from the last event. This model satisfies conditions (6) and (7). The hazard function is

$$r_u = \frac{F(u+\Delta) - F(u)}{\Delta(1 - F(u))} \simeq \frac{\dot{F}(u)}{1 - F(u)}$$

and the transition matrix P_{uv} is such that only two transitions from the state u are possible, one to $u + \Delta$ with the probability $1 - r_u\Delta$ (no events) and the other two 0 with the probability $r_u\Delta$ (an earthquake occurred). Therefore equation (8) takes the form

$$S(u, i) = \min_j [c_{ij} + \beta_j\Delta - \alpha_j r_u\Delta + \theta(S(u+\Delta, j)(1 - r_u\Delta) + S(0, j)r_u\Delta)].$$

Cyclic Poisson process. To describe a sequence of catastrophic events VERE-JONES (1978) used a model of Poisson process with periodic rate, $\lambda(t) = \lambda(t + T)$. Clearly, the information takes the form $\mathcal{I}(t) = t(\mathrm{mod}\ T)$. Therefore conditions (6) and (7) are true. Though this model is nonstationary, Statement 3 still holds. Equation (8) takes the form

$$S(u, i) = \min_j [c_{ij} + \beta_j\Delta - \alpha_j\lambda(u)\Delta + \theta S(u+\Delta, j)].$$

We also add the obvious condition of periodicity $S(u, i) = S(u + T, i)$ and $T = N\Delta$.

Despite the simplicity of these examples, the optimal prediction cannot be obtained in an explicit form if $[c_{ij}] \neq 0$, even in the case of two-phase alerts. The case $[c_{ij}] \neq 0$ involves the hazard function r_u as well as the matrix of transition probabilities for information states $\mathcal{I}(t)$ in constitutive time intervals. The practical estimation of this matrix P_{uv} is complicated and has not yet been stated. Difficulties in estimation of the matrix $[P_{uv}]$ depend on the type of information sequence $\mathcal{I}(t)$ and on detailing of the phase of its states. Above we concluded that the increase in the number of predictors leads to new application areas rather than two predictions of higher quality. In other words, it is sufficient to use the information phase space of a small number of dimensions.

Optimization of mean loss rate. The limit case of problem (3) when $\rho \to 0$ ($\theta \to 1$) stands for the situation in which the loss function takes the form of total expected losses per unit time, that is,

$$\gamma_\pi = \lim_{n\to\infty} \inf E\frac{z_1 + \cdots + z_n}{n\Delta} \to \min_\pi . \tag{10}$$

Problem (10) is interesting for its theory, rather than its applications. Its loss function is associated with the prediction strategy on the entire time semi-axis, whereas the prediction interval in (3) is of order $1/r$. Furthermore, problem (3) is

difficult for numerical analysis as is partly. For results in this area see (MOLCHAN, 1992, MOLCHAN and KAGAN, 1992).

4. Problems

4.1. Performance of Prediction Algorithms

Intermediate-term prediction methods that have been recently developed solve, in fact, the theoretical problem: Is the prediction at all possible? Therefore these methods mostly reduce to the simplest two-phase alert and are characterized by errors $(\hat{n}, \hat{\tau})$. As noted before, it is not sufficient to know these errors to compare various algorithms at the research stage or when economical data are unrefined. The situation improves if any algorithm A is characterized by an error curve similar to error diagram Γ (Section 2). The problem arising is that any algorithm is a complex of prediction methods that yields only one method after specifying the vector Θ of internal parameters of the algorithm. Varying the essential parameters we obtain the error set $(\hat{n}, \hat{\tau})_\Theta$. The lower bound of its convex hull is the error curve Γ_A representing the capability of the algorithm for prediction with the information chosen.

Curves Γ_A are important because they are estimates for the diagram Γ describing the limit prediction capability of the information. Curves Γ_A are also useful in the qualitative comparison of algorithms. For example, curves Γ_A for two algorithms (see Figure 3) intersect at an intermediate point (end points are always the same). When $\hat{\tau}$ is small, the curve Γ_{A_1} is under Γ_{A_2}. Hence algorithm A_1 is preferable in applications with great values of $\beta/\alpha\lambda$ where β is the cost rate of alert and $\alpha\lambda$ is the loss rate from failures-to-predict (β and $\alpha\lambda$ are defined in Section 2).

Note that the calculation of the Γ_A curve is a time-consuming procedure, estimates of Γ_A being affected as they are by the amount of available data and nonuniqueness of $\hat{\tau}$-definition in a case of time-space prediction (MOLCHAN, 1991).

4.2. Estimation of the Hazard Function

The present study indicates that statistical estimation is necessary for the hazard function r_u and the probability transition matrix P_{uv} for information states on consecutive time intervals. Note that the hazard function depends on information states $\mathcal{I}(t) = u$ rather than on time. At present the corresponding statistics are collected for separate, simplest predictors. Estimates of r_u using any predictor combinations (SOBOLOV *et al.*, 1991) are still very rough, because they ignore the conditional dependence of predictors on future events.

The real use of the hazard function r_u and matrix $[P_{uv}]$ requires the strict selection of the most informative predictors and economic discretization of their

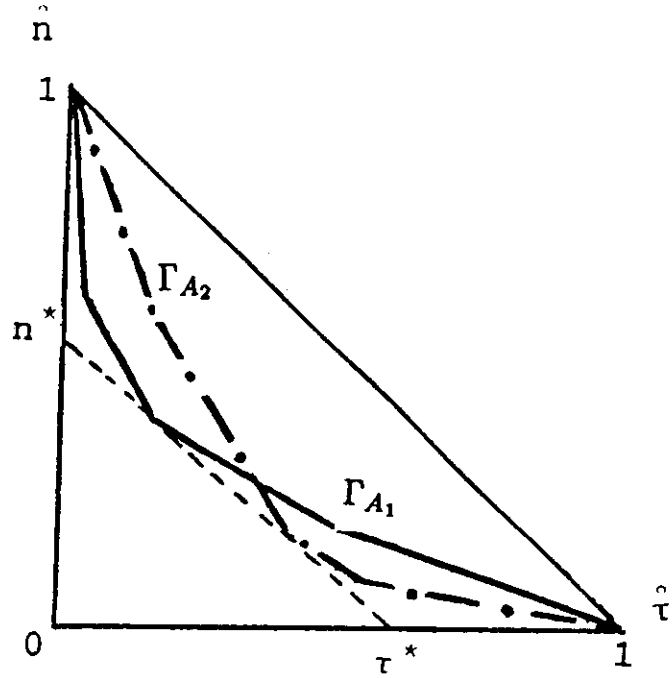

Figure 3
Comparison of algorithms by error diagrams Γ_A. Solid line and dash-dotted line are error diagrams for two algorithms. The line (n^*, τ^*) is the common line of support for curves Γ_{A_1} and Γ_{A_2}. If $\gamma = \lambda\alpha\hat{n} + \beta\hat{\tau}$ is the loss function and $\beta/\alpha\lambda > n^*/\tau^*$ then algorithm A_1 is preferable because it yields lesser losses.

phase space. This is probably a real problem in intermediate-term prediction which is effectively based on a narrow range of predictors, that is, those with energy parameters (Zhurkov's criterion, b value, and others) and phenomena of quiescence and activation. On the other hand, some simple precursors have the same prediction errors as the more universal (in space) collective precursors like CN and M8.

An important example of the estimation of r_u is statistical modeling of earthquake catalogs dating back to HAWKES (1971) and KAGAN (1973). The models involve clustering and for this reason were successfully adapted for intermediate-term prediction by incorporating precursor S like activation and quiescence (OGATA, 1988; KHOKHLOV and KOSSOBOKOV, 1992). Modeling a sequence of earthquakes in the space $X = \{\text{time } t, \text{ magnitude } M, \text{ location } g\}$ is equivalent to assigning a conditional probability density of event x, given a prehistory of the sequence, $\mathcal{J}(t)$, that is, the point process $\{t_i, M_i, g_i\}$ is determined by the hazard function $r_u(x)$ with $u = \mathcal{J}(t)$. The most popular model of $r_u(x)$ has the form

$$r_u(x) = \lambda_0(x) + \sum_{x_i \in \mathcal{J}(t)} \Phi(x \mid x_i) \tag{11}$$

where $\lambda_0(x)$ is the main-shock rate and $\Phi(x \mid x_i)$ the rate of aftershocks of the first generation (primary aftershocks) due to event x_i. The function $\Phi(x \mid x_0)$ is frequently factorizable:

$$\Phi(t, M, g \mid t_0, M_0, g_0) = c(M_0)\varphi_1(t - t_0)\varphi_2(M)[\varphi_3((g - g_0)/r(M_0))r^{-d}(M_0)]$$

where φ_i, $i = 1, 2, 3$ are normalized distribution laws of primary aftershocks over time, magnitude and space, respectively, that is, $\int \varphi_i = 1$ and d is the dimension of the spatial coordinate g. The statistical properties of the primary aftershocks are not known, therefore the parameterization of φ_i rests on known statistical properties of aftershocks: φ_1 is Omori's law, φ_2 is the Gutenberg relation, φ_3 is a Gaussian distribution with linear scale parameter $r(M_0) \propto 10^{M/2}$ (OGATA, 1988; KAGAN, 1991).

A simple transfer of these laws to primary aftershocks is not entirely justified. Attention should therefore be directed to the recent work by KHOKHLOV and KOSSOBOKOV, 1992), where φ_1 has a Gaussian shape and the spatial scale $r(M_0) \propto 10^{M/4}$. The work specifically aims at prediction and shows significantly better prediction than the M8 algorithm.

One paradox of the model (11) is that it is successfully used in prediction in an unusual form: an alarm is declared for very large values of r_u (as response to activation) and even for very low values (as response to quiescence).

We conclude by noting that making the seismicity-responsive model more complicated puts greater demands on the accuracy of r_u and $[P_{u,v}]$. When dealing with prediction involving two kinds of alarm it was necessary to be able to estimate the hazard function with good precision in the vicinity of a single threshold value. With n types of alarm and $[C_{ij}] = 0$, there appear "n" threshold values r_u. Lastly, when $[C_{ij}] \neq 0$, the complete structure of the hazard function needs to be known. However, practice calls for simple and reliable solutions. For this reason an active dialogue is needed between seismologists and economists to discuss realistic typical prediction problems.

REFERENCES

AKI, K. (1989), *Ideal Probabilistic Earthquake Prediction*, Tectonophys. *169*, 197–198.

ELLIS, S. P. (1985), *An Optimal Statistical Decision Rule for Calling Earthquake Alerts*, Earthq. Pred. Res. *3*, 1–10.

FENG, DE YI, JING PING GU, MING ZHOU LIN, SHAO XIE XU, and HUE JUN YU (1985), *Assessment of Earthquake Hazard by Simultaneous Use of the Statistical Method and the Method of Fuzzy Mathematics*, Pure and Appl. Geophys. *126*, 982–997.

GUSEV, A. A. (1976), *Indicator earthquakes and prediction.* In *Seismicity and Deep Structure of Siberia and Far East* (in Russian) (Nauka, Novosibirsk 1976) pp. 241–247.

HOWARD, R. A., *Dynamic Programming and Markov Processes* (New York 1960) 136 pp.

HAWKES, A. G. (1971), *Spectra of Some Self-exciting and Mutually Exciting Point Processes*, Biometrica *58*, 83–90.

KAGAN, Y. (1973), *On a Probabilistic Description of the Seismic Regime*, Fizika Zemli *4*, 110-123.

KAGAN, Y. Y., and JACKSON, D. D. (1991), *Seismic Gap Hypothesis: Ten Years After*, J. Geophys. Res. *96*, 21419–21431.

KANTOROVICH, L. V., and KEILIS-BOROK, V. I., *Economics of Earthquake Prediction*, Proceedings of UNESCO Conference on Seismic Risk, Paris, 1977.

KEILIS-BOROK, V. I., and ROTWAIN, I. M. (1990), *Diagnosis of Time of Increased Proability of Strong Earthquakes in Different Regions of the World: Algorithm CN*, Phys. Earth and Planet. Inter. *61*, 57–72.

KEILIS-BOROK, V. I., and KOSSOBOKOV, V. G. (1990), *Premonitory Activation of Earthquake Flow: Algorithm M8*, Phys. Earth and Planet. Inter. *61*, 73–83.

KHOKHLOV, A. V., and KOSSOBOKOV, V. G. (1992), *Seismic Flux and Major Earthquakes in the Northwestern Pacific* (in Russian), Doklady RAN, 325, No. 1, 60–63.

LINDGREN, G. (1985), *Optimal Prediction of Level Crossing in Gaussian Processes and Sequences*, Ann. Probab. *13*, 804–824.

MOLCHAN, G. M. (1990), *Strategies in Strong Earthquake Prediction*, Phys. Earth and Planet. Inter. *61*, 84–98.

MOLCHAN, G. M. (1991), *Structure of Optimal Strategies of Earthquake Prediction*, Tectonophys. *193*, 267–276.

MOLCHAN, G. M., *Models for optimization of earthquake prediction* (in Russian). In *Problems in Earthquake Prediction and Interpretation of Seismological Data* (Vychislitelnaya Seismologiya *25*) (Nauka, Moscow 1992) pp. 7–28.

MOLCHAN, G. M., and KAGAN, Y. Y. (1992), *Earthquake Prediction and its Optimization*, J. Geophys. Res. *97*, 4823–4838.

NARKUNSKAYA, R. S., and SHNIRMAN, M. G., *On an algorithm of earthquake prediction.* In *Computational Seismology and Geodynamics*, 22/23 (English Transl.) (Am. Geophys. Union 1993).

NISHENKO, S. P. (1989), *Circum-Pacific Earthquake Potential: 1989–1999*, USGS, Open File Report 89–86, 126 pp.

ROSS, M., *Applied Probability Models with Optimization Applications* (San Francisco 1970) 198 pp.

OGATA, Y. (1988), *Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes*, J. Am. Statist. Assoc. *83* (401), Applications, 9–26.

SADOVSKY, M. A. (ed.), *Long-term Earthquake Prediction Methodology* (in Russian), (Nauka, Moscow 1986) 128 pp.

SOBOLEV, G. A., CHELIDZE, T. L., ZAVYALOV, A. D., SLAVINA, L. B., and NICOLADZE, V. E. (1991), *Maps of Expected Earthquakes Based on Cobmination of Parameters*, Tectonophys. *193*, 255–265.

VERE-JONES, D. (1978), *Earthquake Prediction—A Statistician's View*, J. Phys. Earth. *26*, 129–146.

WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES (1988), *Proabilities of Large Earthquakes Occurring in California on the San Andreas Fault*, U.S. Geol. Surv. Open File Rep. 88–398, 66 pp.

Appendix : PEPI-1990

Physics of the Earth and Planetary Interiors, 61 (1990) 84–98

3. The renewal process

3.1. Examples of strategies

Let us consider the simplest stochastically periodic point process which has independent and equally distributed time intervals τ_i, i.e. the renewal process with distribution function $f(x) = \Pr\{\tau < x\}$. It includes both an absolutely random Poisson sequence $F'(x) = \lambda \exp(-\lambda x)$, and a predictable periodic sequence $F'(x) = \delta(x - x_0)$. These extremities present a range of possibilities in prediction of the renewal processes.

A random renewal process seems to be an appropriate model for rare events in some local catalog (a typical object for a practically interesting prediction), or to reflect the main elements of the 'seismic cycle' for a large tectonic region. At the same time, the well-known mathematical results on superimposition of the renewal processes show (see, for example, Daley and Vere-Jones (1988)) that the total catalog of rare local events can have all the properties of Poisson flow in time. In such a case, events from different places should be in weak dependence.

The spread index is a useful parameter of the process

$$J = \lim_{t \to \infty} \frac{\operatorname{Var} \nu(t)}{E\nu(t)} = \frac{\operatorname{Var} \tau}{(E\tau)^2} \qquad (13)$$

where $\nu(t)$ is the number of events in the interval $(0, t)$ and Var is the dispersion. Deviations of J from unity reflect on average the tendency of a point process of 'clustering' ($J > 1$), and to periodicity ($J < 1$, 'scattering'), as related to the Poisson chaos.

The γ-optimal strategies for renewal processes are defined by sets

$$\Delta = \{x : r(x) = F'(x)/[1 - F(x)] > u\}$$

Δ is the interval for the unimodal hazard function $r(x)$. It is unbounded, $\Delta = (x, \infty)$, if $r(x)$ is an increasing function (the so-called property of 'positive ageing'; this term is used in the description of endurance failures in materials). As is known (Cox and Lewis, 1966), in this case $J < 1$, so when events tend to scattering, announcement of an alarm should be delayed. For decreasing function $r(x)$ ('negative ageing'), $\Delta_{opt} = (0, x)$, i.e. it is preferable to give alarm at the beginning of a prediction cycle if events in the process tend to clustering.

In earthquake prediction there is a tendency to reduce the duration of the alarm. However, there are some exceptions: prediction in Kamchatka, based on seismic gaps (Fedotov et al., 1977), and the method of long-range aftershocks (Prozorov and Rantsman, 1972). Thus, if density $F'(x)$ is unimodal, with a well-expressed peak, its short vicinity seems a natural choice for an alarm set. Generally, such a decision is not correct, as an alarm after the modal point $F'(x)$ will be quickly cancelled if $F'(x)$ decreases sufficiently rapidly. Otherwise, the announcement of unbounded alarm is not always to the detriment of a prediction strategy. Some examples of optimal alarm sets are given below.

3.1.1. Example 1

F is a uniform distribution on $[0, 2m]$. Then

$$r(x) = (2m - x)_+^{-1}, \quad J = 1/3$$

The minimax interval, i.e. the γ_d-optimal strategy is the set $\Delta_{opt} = (km, \infty)$, where km is the point of 'golden section' of $[0, 2m]$: $k = 3 - \sqrt{5} \simeq 0.76$ (an unexpected optimal property of the 'golden section'!). The errors are $\mathring{n} = \mathring{\tau} \simeq 38\%$.

3.1.2. Example 2

F is a γ-distribution with average m and index J:

$$F'(x) = cx^{1/J - 1} \exp(-x/Jm), \quad x \geq 0$$

TABLE 1

Parameters of minimax strategy for γ-distribution

	Strategy							
	$\Delta = \{x : x < km\}$		$\Delta = \{x : x > km\}$					
Spread index $J = \sigma^2/m^2$ [a]	2	1	0.4	0.3	0.2	0.1	0.05	0.02
k	0.71	0.69	0.72	0.73	0.74	0.77	0.81	0.87
Prediction errors $\mathring{n} = \mathring{\tau}$ (%)	40	50	39	36	31	25	21	19

[a] m, average; σ^2, dispersion.

This distribution is unimodal, and is almost Gaussian with $J \ll 1$. The hazard function $r(t)$ is monotonic, thus the γ-optimal alarm sets are $\Delta = (0, x)$ for $J > 1$ and $\Delta = (x, \infty)$ for $J < 1$. $\varkappa = 1$ corresponds to the unpredictable Poisson process; here condition B is incorrect: $r(x) = m^{-1}$, $x \geqslant 0$. Thus the optimal open sets Δ could take any form and contain any point.

Table 1 demonstrates prediction possibilities by means of γ_d-optimal (minimax) strategy for γ-distribution. The minimax strategy has the following form:

$$\Delta = \{x: \operatorname{sgn}(1 - J) \cdot [x - k(J)m] > 0\}$$

Dimensionless parameter $k(J)$ turned out to be unexpectedly stable: over a wide range of index $J \in [0.05,2]$, the value of $K(J) = 3/4 \pm 0.06$.

An example can demonstrate stability not only of the strategy but also of the prediction results. We approximate a γ-distribution by a uniform distribution with the same average and value of J. It is possible for $J = 1/3$. We use the γ_d-optimal strategy $\Delta = (0.76m, \infty)$ in this approximation (see example 1) for prediction in the true model F. The pseudo-optimal strategy leads to errors $\mathring{n}_* \approx 40\%$ and $\mathring{\tau}_* = 34.6\%$. The average error is $(\mathring{n}_* + \mathring{\tau}_*)/2 = 37\%$, against true minimax errors $\mathring{n} = \mathring{\tau} = 36\%$. The agreement is very good for such a crude approximation of true distribution.

3.2. *Prediction on the San Andreas fault*

Nishenko and Buland (1987) analysed 53 intervals τ_i in 14 regions with high seismicity. Out of three types of distributions $F_G(x)$, Normal, Weibull and Lognormal, preference was given to the third type. The following model is accepted for inter-event intervals for tectonic regions $\{G\}$

$$\tau = m_G \exp\left(v_G \xi - v_G^2/2\right) \tag{14}$$

where ξ is a standard normal random value $N(0, 1)$. $m_G = E\tau_G$ and v_G^2 is the variation of $\ln \tau$.

The spread index for this model is

$$J = \exp\left(v_G^2\right) - 1$$

The WGCEP (1988) applied model (14) for prediction of the strongest earthquakes on the San Andreas fault. Table 2 contains data on eight of the fault segments. Prediction is determined by conditional probabilities

$$R_\delta = P\{\nu(\tau_-, \tau_- + \delta) \geqslant 1 \mid \tau \geqslant \tau_-\}$$
$$= [F(\tau_- + \delta) - F(\tau_-)]/[1 - F(\tau_-)]$$

of occurrence of an event on segments G for forthcoming (after 1988) $\delta = 30$ yr; τ_- is the time since the last event in G. According to the WGCEP calculations the most dangerous of the denoted eight segments is now the Parkfield (P) region, where $R_\delta > 0.9$ against $R_\delta < 0.4$ for the other segments (see Table 2).

Using this method of comparing the fault segments almost always Parkfield is diagnosed as the most dangerous. Actually, for any segment except P, $R > 0.9$ for $0 \leqslant \tau \leqslant 800$ yr. Moreover, the R value in Parkfield can reach the level of 0.5 only in the impossible case of quiescence for 4000 yr.

TABLE 2

Conditional probability R_δ of major earthquakes along segments of the San Andreas fault, 1988–2018

Fault segment	1	2	3	4	5	6	7	k	k [a]	10
San Francisco Peninsula	90	7	1906	196	0.38	0.176	2032	0.75	8.8	0.30
Santa Cruz Mountains	35	6.5	1906	136	0.44	0.185	2007	0.74	8.4	0.30
Parkfield	30	6	1966	21	0.9	0.056	1983	0.81	37	0.19
Cholame	55	7	1857	159	0.3	0.281	1972	0.72	4.8	0.34
Carrizo	145	8	1857	296	0.1	0.137	2082	0.76	13	0.28
Mojave	100	7.5	1857	162	0.3	0.168	1979	0.75	9	0.29
San Bernardino Mountains	100	7.5	1812	198	0.2	0.36	1951	0.7	3.4	0.36
Coachella Valley	100	7.5	1680	256	0.4	0.09	1880	0.78	19	0.25

1, length (km); 2, expected magnitude; 3, date of most recent event; 4, expected recurrence time (yr); 5, R_δ; 6, spread index, J; 7, start of alarm; 8, 9, k and k [a] are time alarm thresholds; 10, Errors $\mathring{n} = \mathring{\tau}$.

[a] Except for the last four columns, the data are from WGCEP, (1988).

Otherwise, the R-test causes a steady alarm in P, but no alarms in other segments.

Let us consider the prediction in model (14) from this viewpoint. The hazard function $r(x)$ of the lognormal distribution is unimodal, with $r(0) = r(\infty) = 0$. Thus optimal alarm sets form finite intervals $\Delta = \{x : km_G < x < k_* m_G\}$. For minimax strategies, normalized thresholds Δ, k and k^* are presented in Table 2. Upper thresholds k^* are determined by the choice of the model $F(x)$ and are practically unlimited. They change within the range 700–5000 yr. That is no longer important because if upper thresholds are substituted for ∞, then prediction errors differ by 0.001 for observed values of $J \leqslant 0.4$. k threshold values and prediction errors $\epsilon = \mathring{n} = \mathring{\tau}$ proved to be stable in the sense of the model choice. Functions $k(J)$ and $\epsilon(J)$ stay almost unchanged when a lognormal distribution was replaced by Weibull or γ distribution, with $\Delta = \{x : x > k(F)\}$ (compare Tables 1 and 2).

The above circumstances justify the choice of a minimax strategy according to which five out of eight fault segments should be in a state of alarm by 1988, in particular, the region P—since 1983—and the Coacella Valley (CV)—since 1880 (!)—although according to R_δ, alarm is impossible in CV. However, as related to the minimax strategy, the concentration of U.S. scientific efforts in region P is justified. According to the minimax strategy, a maximum of successes $\delta m_G^{-1}(1 - \mathring{n})$ for the considered period δ is expected in region P.

The analysis becomes more complicated if we take into account absolute losses for the whole region G, i.e., losses of the type

$$\gamma_a = \alpha_G m_G^{-1} \mathring{n} + \beta_G \mathring{\tau}$$

Concretization of α_G, β_G is not unambiguous. We can suppose that

$$\alpha_G = \alpha_M (I);\ \alpha_G = \tilde{\alpha}_M m_G (II);\ \beta_G = \beta_M |G|$$

where subscript M shows the dependence of parameters on the magnitude of predicted events.

Losses from one failure-to-predict of magnitude M are proportional to area $Q(M)$ of the destruction zone if the destruction objects are homogeneous and uniformly distributed in space. If we account for recurrence frequency $p(M)$ then

$$\alpha_M \propto \int_{\Delta M} Q(M) p(M)\, dM$$

Usually $Q(M) \propto 10^{b_1 M}$, $p(M) \propto 10^{-bM}$ and $b_1 \simeq b$. Thus $\alpha_M \propto |\Delta M|$.

Evidently, model $\alpha_G - II$ can occur in insurance. Let the insurance rates be independent of location on the fault. Then insurance payments are proportional to time. However, the period between failures-to-predict is proportional to m_G. Thus we can assume that compensation for an object's destruction is also proportional to m_G. Consequently, $\alpha_G \propto m_G$. However, these arguments do not account for the efficiency of investments.

For methodical purposes we assume $\alpha_M = \alpha = \tilde{\alpha}_M$, $\beta_M = \beta$; $k = \beta/\alpha$. The minimum of loss function γ_a is determined by the value

$$\gamma_* = \alpha_G m_G^{-1} \left\{ 1 - \int_0^\infty [r_0(u) - km_G |G|]_+ \times [1 - F_0(u)] \right\} du$$

where $x_+ = (x + |x|)/2$, F_0 is the distribution of τ with $E\tau = 1$, i.e.

$$F_0(x) = \Phi[\ln(x/v) + (v/2)]$$

$\Phi(x)$ is a standard Gaussian distribution, and $r_0(x)$ is a hazard function for $F_0(x)$. Value γ_* is an increasing function of parameter k:

$$\frac{d}{dk}\gamma_* = \alpha |G| \mathring{\tau} \geqslant 0 \tag{15}$$

Moreover,

$$\gamma_* = \alpha_G^* \begin{cases} k|G|[1 + o(1)] & k = \beta/\alpha \ll 1 \\ 1/m_G & k = \beta/\alpha \geqslant r_0^*/(m_G|G|) \end{cases} \tag{16}$$

where

$$r_0^* = \max r_0(x)$$

Hence we can obtain a qualitative conclusion of optimal losses, if we use real data on the segments' dimensions and periods m_G (see Table 2). We summarize our conclusions as follows. Under condition $\alpha_G = \alpha$, in the Parkfield region expected losses are minimal if $\beta/\alpha < k_1$ and maximal if $\beta/\alpha > k_2 (k_1 < k_2)$. Under other conditions $\alpha_G = \alpha m_G$, losses in Parkfield are not greater than in other regions with any α/β.

For both models, when $\alpha_G = \alpha$ the simplest 'optimistic' strategy (no alarm) is γ_a-optimal if $\beta/\alpha \geqslant \max r_0^*/|G| m_G$. However, the simplest pessimistic strategy (endless alarm) cannot be γ_a-optimal in any segment if $\beta > 0$.

4. Tail behavior for recurrence time distribution

There are practically no data on the tail of distribution of τ for strong events. Some attempts have been made to overcome difficulties by means of models. Kagan and Knopoff (1987b) considered an earthquake as an event of 'dumping' of the accumulated stress σ_t in a seismic region when σ_t reaches the critical state σ. The τ-distribution is simulated by a random walk

$$\sigma(t+\Delta) = \sigma(t) + \Delta\xi(t),\ \sigma(0) = x,\ \sigma(t) < \sigma \tag{20}$$

where ξ_t is the Wiener process with average ct and dispersion d^2t. The non-zero average stress rate $c > \sigma$ can be related to plate tectonics; the diffusion part of $\Delta\sigma$ can be associated with random events of 'loading' or 'dumping' which are realized by creep and by weak seismicity inside/outside the region. When the critical state σ is achieved at a random moment τ, the stress drops to the initial level $\sigma(0)$ and the system starts to function anew.

As is already known (Skorohod, 1964), the described renewal model has the recurrence time distribution

$$F'(t\,|\,c) = \left[\,|\sigma - x|/d\sqrt{(2\pi)}\right] t^{-3/2} \times \exp\left\{-\tfrac{1}{2}\left[\,|\sigma - x|/\sqrt{(t)} - c\sqrt{(t)}\right]^2/d^2\right\},\quad t > 0 \tag{21}$$

The function distribution $F(t\,|0)$ has a self-similar character of attenuation with $t \gg 1$, i.e. $F'(t) \propto t^{-3/2}$. Among distributions of τ with fixed average value $E\tau = (\sigma - x)/c$ the distribution $F(t\,|\,c)$ is the nearest to $F(t\,|0)$ in the entropy distance: $\mathrm{dist}(F, F_0) = \int(\log \mathrm{d}F/\mathrm{d}F_0)\,\mathrm{d}F$. Self-similarity F with exponent $-3/2$ or proximity to this self-similarity are the main arguments in favour of model (20). It turns out that this property disappears when natural limits are set; the accumulated stresses in the region cannot drop below a certain zero threshold. The modified model is described by the relation

$$\sigma(t+\Delta) = \max[0,\ \sigma(t) + \Delta\xi(t)],\quad \sigma(0) = x,\ \sigma(t) < \sigma \tag{22}$$

Models of type (22) occur in the theory of stochastic storage processes. However, the author could not find the τ-distribution for the described case in the literature. The τ-distribution can be derived as follows. Let

$$V(t\,|\,x) = P\{\sigma(s) < \sigma \text{ for all } s \in (0, t);\ \sigma(0) = x\} = P(\tau > t)$$

Standard methods of the Markov process (Skorohod, 1964) show that $V(t, x)$ satisfies the parabolic equation

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x} + 1/2d^2\frac{\partial^2}{\partial x^2}\right)V = 0,\ t > 0,\ x \in (0, \sigma$$

with boundary conditions

$$\frac{\partial}{\partial x}V(t, 0) = 0;\ V(0, x) = 1,$$

$$x \in (0, \sigma);\ V(t, \sigma) = 0$$

By means of the Laplace transform

$$Z_\lambda(x) = \int_0^\infty \mathrm{e}^{-\lambda t}V(t, x)\,\mathrm{d}t = (1 - E\,\mathrm{e}^{-\lambda\tau})/\lambda$$

the problem is reduced to an ordinary differential equation

$$\lambda Z_\lambda(x) - 1 = \left(cD + \frac{d^2}{2}D^2\right)Z_\lambda(x),\ D = \frac{d}{dx}$$

$$Z_\lambda(\sigma) = 0 = Z_\lambda(0)$$

Hence

$$\Phi_x(\lambda) = E\,\mathrm{e}^{-\lambda\tau} = \frac{\lambda_+ \exp^{(-\lambda_- x)} - \lambda_- \exp^{(-\lambda_+ x)}}{\lambda_+ \exp^{(-\lambda_- \sigma)} - \lambda_- \exp^{(-\lambda_+ \sigma)}} \tag{23}$$

where

$$d\cdot\lambda_\pm = c/d \pm \sqrt{\left[2\lambda + (c/d)^2\right]}$$

Equation (23) defines all τ-moments. It can be inversed according to Jordan's lemma, i.e. the τ-distribution density can be expressed by poles of $\Phi_x(\lambda)$ in the semi-plan $\mathrm{Re}\,\lambda < 0$. As a result we have the following.

Statement 4

(1) In model (22), with $c \geqslant 0$, the distribution density τ has an exponential asymptotic, $F'(t) = 0\,[\exp^{(-\rho_1 t)}]$ To be more precise

$$F'(t) = \sum_{n \geqslant 1} a_n \exp(-\rho_n t) \tag{24}$$

where

$$\rho_n \equiv \rho_n(\sigma) = \begin{cases} 1/2(c/d)^2\left[(x_n/k)^2 + 1\right], & c > 0 \\ 1/2(d/\sigma)^2\pi^2(n - 1/2)^2, & c = 0 \end{cases} \tag{25}$$

Here $k \equiv k(\sigma) = c\sigma/d^2$ and $x_n \equiv x_n(k)$ are positive roots of the equation

$$k \sin x/x = -\cos x,\ k \geqslant 0$$

i.e.

$$x_n = \pi(n - 1/2) + k\cdot O(1/n)$$

Coefficients

$$a_n = \operatorname*{Res}_{\lambda = -\rho_n} \Phi_x(\lambda)$$

In particular, if $c = 0$ then

$$a_n = (-1)^{n-1}(d/\sigma)^2\pi(n-1/2) \times \cos[\pi(n-1/2)x/\sigma]$$

Series (24) converges fast because $a_n = O(n)$ and $\rho_n = O(n^2)$, $n \to \infty$.

(2) In model (22) there is a useful representation for the recurrence time

$$\tau = \sum \epsilon_n \tau_n \tag{26}$$

where $\{\epsilon_n, \tau_n\}$ are jointly independent random variables,

$$P\{\tau_n > t\} = \exp(-\rho_n t)$$

and $\epsilon_n = 0$ or 1 with probability p_n or $1 - p_n$, where

$$p_n = \rho_n(\sigma)/\rho_n(x)$$

($\rho_n(\circ)$; see (25)). If $x = 0$ then $\epsilon_n \equiv 1$. (Representation (26) corresponds to infinite factorization of the characteristic function τ, based on zeroes and poles $\Phi_x(\lambda)$). In models (20) and (22), where $x = 0$, the distribution of τ is asymptotically normal with $k = c\sigma/d^2 \to \infty$, i.e.

$$P\{\sqrt{(k)}(\tau c/\sigma - 1) < 1\} \to (2\pi)^{-1/2}\int_{-\infty}^{t} \exp^{(-u^2/2)}\, du$$

(3) In model (22), where $x = 0$, the average recurrence time is

$$E\tau = \sigma/c[1 - (1 - e^{-2k})/2k]$$

and the spread index

$$J = 1 - 2\frac{2(k-1)^2 + 1 - (2k+3)\, e^{-2k}}{(2k - 1 + e^{-2k})^2} \simeq \frac{4k-5}{(2k+1)^2},\ k \to \infty$$

decreases as the function k from $J(0) = 2/3$ to 0 with $k \to \infty$, i.e. $J \leqslant 2/3$. (In model (20), $x = 0$, $E\tau = \sigma/c$ and $J = k^{-1} \geqslant 0$).

(4) In model (22), where $x = 0$, the hazard function

$$r(t) = \rho_1 - \left|\frac{\rho_1 a_2}{\rho_2 a_1}\right|(\rho_2 - \rho_1)\exp^{[-(\rho_2-\rho_1)t]},\ t \gg 1$$

i.e. the γ-optimal alarm sets $\Delta = \{t : r(t) > u\}$, and $u \ll \rho_1$ is unbounded. This is not true for model (20).

APPENDIX : Tectonophysics (1991)

Tectonophysics, 193 (1991) 267–276
Elsevier Science Publishers B.V., Amsterdam

G.M. MOLCHAN

Stability of minimax strategy

We consider a special case of prediction in which the current information $\xi(t)$ consists of quiescence duration τ_- alone, i.e. of the time that has elapsed since the last strong event. A prediction of this kind arose in Nishenko (1989) for the Circum-Pacific belt.

Let $F(x)$ be the distribution of time τ between strong events. The hazard function for the information $\xi(t) = \tau_-$ is then:

$$r(t) = F'(t)/[1 - F(t)], \quad t \geq 0 \cdot \tag{13}$$

if the time of last event $\{t_i < t\}$ is taken as the origin. The right-hand side is well known as a hazard function in the analysis of survival data (Cox and Oakes, 1984).

In engineering research the Weibull distribution is usually employed for F:

$$F(x) = 1 - \exp[-(\lambda x)\alpha], \quad \alpha > 0 \tag{14}$$
$$p\lambda = \Gamma(1 + 1/\alpha)/m$$

It also is the basic one for seismic risk analysis in the engineering literature (Cornell, 1986), although a possible alternative is the gamma distribution:

$$F(x) = c\int_0^x u^{\alpha-1} \exp(-\alpha u/m)\, du \tag{15}$$

It is not however considered physically natural. Prediction of characteristic earthquakes is treated in Nishenko (1989) by using the lognormal model F:

$$F(x) = \Phi\left(\alpha^{-1} \ln(x/m) + \alpha/2\right) \tag{16}$$

where Φ is the standard Gaussian distribution and $\alpha^2 = \mathrm{Var}(\ln \tau)$. The symbol m in the above distributions denotes the mean of τ. According to Nishenko (1989), typical values of the coefficient of variation $\rho = \sigma/m$ (σ^2 is the variance of τ) belong to the interval 0.25–0.6. These estimates are based on historical earthquakes and paleo-data. For this reason we will keep in mind the values $\rho < 1$ mentioned above.

Function (13) is extremely sensitive to the choice of F. However, when treated within the framework of the loss function approach the prediction is much more stable. This is borne out by the following calculations.

The Weibull and gamma distributions behave similarly, having a single mode when $\sigma/m < 1$. The hazard functions (13) are increasing; for this reason γ-optimal strategies for any loss function reduce to declaration of alarm when the "quiescence" τ_- exceeds the threshold $k_\gamma m$. The density $F'(x)$ decreases when $\sigma/m > 1$. So does $r(x)$, consequently a γ-optimal strategy is to declare an alarm at once after a large event and to persist with it until the threshold value $\tau = k_\gamma m$, unless the alarm is removed by a next event. The monotony of $F'(x)$ makes models (14, 15) of little practical interest for $\sigma/m > 1$.

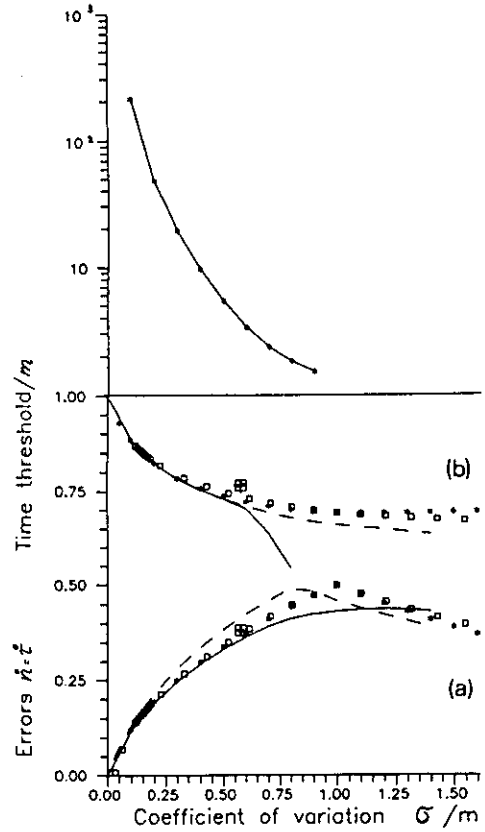

Fig. 2. Parameters of minimax strategies for four $F(x)$ models: Weibull (squares); gamma (asterisks); lognormal (solid line = two thresholds, dashed line = single threshold); uniform (+), (a) is prediction results, $\gamma = \max(\hat{n}, \hat{t})$, (b) is normalized alarm thresholds, lower (k_γ) and upper ($\bar{k}_\gamma$), for model (14); m is the mean, σ^2 the variance of $F(x)$

The lognormal density and its hazard function have a single mode for all $\sigma/m > 0$. For this reason a γ-optimal alarm is declared only in the finite interval: $k_\gamma < \tau_-/m < \bar{k}_\gamma$.

The thresholds k_γ of minimax strategy for models (14–16) are shown in Fig. 2. They practically coincide for distributions (14, 15) when $\sigma/m < 1$, differing by 0.01 or 1.5% at most. Practically coincident with these is the lower threshold k_γ for

a lognormal distribution for $\sigma/m < 0.6$, the deviation being $|\Delta k_\gamma| < 0.03$. The upper threshold $\bar{k}_\gamma$ in the same range $\sigma/m < 0.6$ is large, $\bar{k}_\gamma > 3.3$, and makes little contribution to the prediction errors. The threshold $\bar{k}_\gamma$ is entirely dependent on the problematical behaviour of model F at infinity. For this reason it is natural in the lognormal model to consider a slightly simpler γ-optimal strategy involving only a lower alarm threshold $\tilde{k}_\gamma$. (The value of $\tilde{k}_\gamma$ is again found from the condition $\hat{n} = \hat{\tau}$.) The threshold is consistent with k_γ for (14, 15) in the entire interval $\sigma/m < 1$: $|\Delta k_\gamma| < 0.03$.

For the practically interesting interval $\sigma/m = 0.25$–0.6 the thresholds k_γ and $\tilde{k}_\gamma$ are generally very stable: 0.75 ± 0.05 ($k_\gamma = 0.76$ for a uniform distribution). These results point to a stable structure of the minimax prediction strategy.

The second type of stability is relevant for prediction results, i.e., for the values of the loss $\gamma = \max(\hat{n}, \hat{\tau})$; see Fig. 2. The values of γ for optimal strategies involving a single alarm threshold differ within 0.02 or 3% for $\sigma/m \leq 0.6$.

Figure 3 presents error curves Γ for three F models (14–16) with the coefficients of variation $\sigma/m = 0.25$, 0.5, and 0.75. The curves are consistent for models (14, 16) in the range $2\hat{n} > \hat{\tau}$.

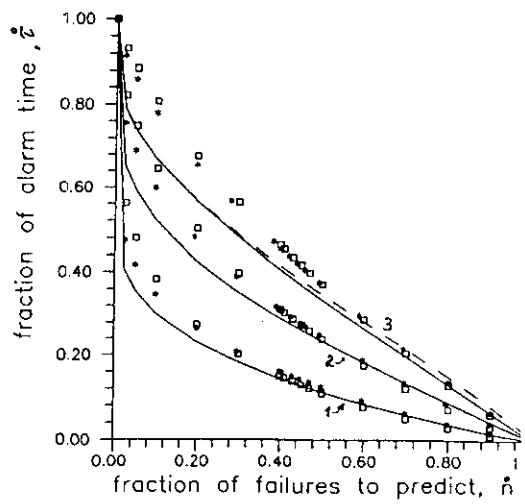

Fig. 3. Error curves $\Gamma = (\hat{n}, \hat{\tau})$ for three models of $F(x)$: Weibull (squares); Gamma (asterisks); Lognormal (solid line = two thresholds, dashed line = single threshold); and the three values of σ/m: $1 = 0.15$, $2 = 0.5$, $3 = 0.75$.

This indicates that the prediction, its structure and the results are similar for models (14, 15) and a broad class of loss functions.

Discussion

(1) Statement 1 gives an idea of the optimal structure of predictions that are local in space. Locally optimal predictions for natural loss functions form an optimal prediction in the whole space.

There is a fundamental difference between the Kolmogorov-Wiener prediction concept and the one considered above. In the first case prediction deals with the best extrapolation of time series values (e.g. earthquake energy), whereas in the second case it refers to the best temporal localization of certain singularities of this series (e.g. large peaks). In other words we have a "vertical prediction" in the classical case and a "horizontal prediction" in the second case. The hazard function $r(t)$ is well known as an instantaneous optimal predictor of a point process in the classical sense: $r(t)$ gives the unbiased estimate of the number of events on the interval dt with minimal variation. Statement 2 shows that $r(t)$ is also the main component of the optimal prediction of the second type. As it turned out, the latter is related to the theory of testing statistical hypotheses. This explains some analogies of the present paper with the mentioned theory.

(2) The results of the present work are not only pure theoretical because hazard function estimation does not seem to be hopeless today. This is confirmed by attempts of seismologists to estimate probability gain, $PG(\xi)$, for a set of precursors $\xi = (A_1, \ldots, A_n)$. If $\{A_i\}$ were monitored within the intervals $(t - T_i, t)$, then the value $PG(\xi)$ just corresponds to the normalized hazard estimator for the current moment t:

$$PG(\xi) = P(N(t+\delta) - N(t) > 0 \mid \epsilon_1, \ldots, \epsilon_n)/P_E$$
$$= r(t)/\lambda$$
$$P_E = [P(N(\tau+\delta) - N(t) > 0)]$$

where ϵ_i are precursor indicators in the interval $(t-T, t)$, i.e., $\epsilon_i = 1$ if A_i occurred, and $\epsilon_i = 0$ otherwise. The following assumption is often used

to calculate $PG(\xi)$. Let $\epsilon_i, \ldots, \epsilon_n$ be conditionally independent, given $N(\delta t)$, then:

$$\left[PG(\xi)^{-1} - P_E\right]/(1 - P_E) = q\Pi\left[\left(PG_i(1)^{-1} - P_E\right)/(1 - P_E)\right] \qquad (17)$$

where multiplicators:

$$PG_i(1) = P\left[N(\delta t) > 0 \mid \epsilon_i = 1\right], \quad i = 1, \ldots, m$$

are related to those precursors which have occurred in $(t-T, t)$. The factor:

$$q = \Pi\left[\left(PG_i^{-1}(0) - P_E\right)/(1 - P_E)\right], \quad i = m+1, \ldots, n$$

is related to non-occurred precursors in $(t-T, t)$, where:

$$PG_i(0) = (1 - PG_i(1)\,p_i)/(1 - p_i)$$

and p_i unconditional probability of precursor A_i: $Pr(\epsilon = 1) = p_i$. For true precursors $PG_i(0) < 1$. Thus computations with $q \equiv 1$ (Aki, 1981; Grandori and Grandori, 1983) lead to overestimation of the value $PG(\xi)$. Conditionally dependent values $\{\epsilon_i\}$ result in a similar effect. Hence, prediction methods based on formula (17) (Sobolev et al., 1989), are sensitive to quantity and quality of precursors $\{A_i\}$. The paper of Zhuang et al. (1989) contains an attempt of expert estimation of $r(t)$, taking into account correlations between precursors.

(5) What is the analogy of Γ_M for a collection of point objects $\{g\} = G$. The answer is not unique. Take the example of total linear loss:

$$\gamma = \int \alpha(g)\lambda(g)\mathring{n}(g)\,dg + \int \mathring{\tau}(g)\beta(g)\,dg = \bar{\alpha}\Lambda\mathring{n}(G) + \bar{\beta}\,|G|\,\mathring{\tau}(G)$$

where $\beta(g)$ is the loss for an alarm per unit space–time, $|G|$ is the area of G, and $\mathring{n}(G)$, $\mathring{\tau}(G)$ is analogous to the errors $\mathring{n}$, $\mathring{\tau}$. Putting $\alpha(g) = \alpha$, $\beta(g) = \beta$, we have the ordinary mean characterstics $\mathring{n}$, $\mathring{\tau}$:

$$\mathring{n}(G) = \int_G \lambda(g)\mathring{n}(g)\,dg/\Lambda$$

$$\mathring{\tau}(G) = \int_G \mathring{\tau}(g)\,dg/|G|$$

where $\Lambda = \int \lambda(g)\,dg$.

The model $\beta(g) = \beta$ is acceptable at the research stage, if strong events are a priori permissible everywhere in G. Otherwise, $\mathring{\tau}(G)$ can be made infinitely small by adding aseismic areas. Hence, when characterizing the method, the choice of G must be well grounded. One sometimes uses a different way, putting $\beta(g) = c\hat{\lambda}(g)$, where $\hat{\lambda}(g)$ is well smoothed epicentre density for strong earthquakes. Then we derive for $\mathring{\tau}(G)$ an alternative version:

$$\mathring{\tau}(G) = \int_G \mathring{\tau}(g)\hat{\lambda}(g)\,dg \Big/ \int_G \hat{\lambda}(g)\,dg$$

which makes sense on the research stage of prediction, until there is no special agreement on the boundaries of G.

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.

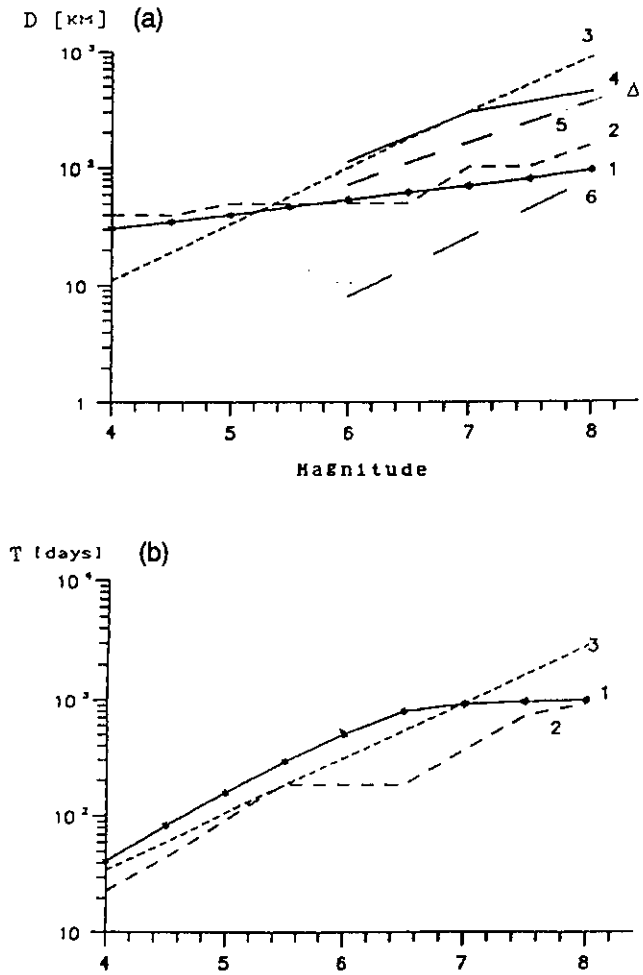

Figure 1. (a) Space and (b) time windows for aftershock identification according to: 1—Gardner & Knopoff (1974); 2—Keilis-Borok *et al.* (1980); 3—Knopoff *et al.* (1982); 4—Prozorov (1986); 5—Molchan & Dmitrieva (by LIR procedure, NOAA catalogue, see text); 6—Utsu & Seki (1954). The triangle marks 1964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

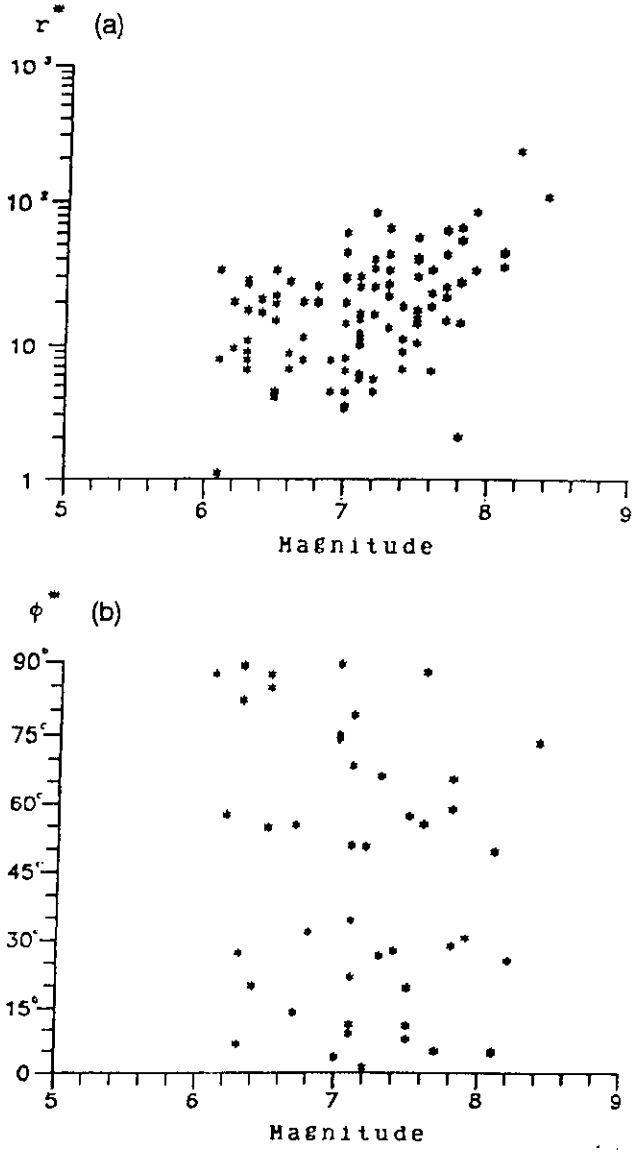

Figure 2. The offset of mainshock epicentre relative to the centre of the aftershock zone as a function of mainshock magnitude $M_0 \geq 6$: (a) distance r^* (km) and (b) the minimum angle φ^* with the longer axis of the aftershock zone. Catalogue NOAA (1964–1980) is used; aftershock zones with aspect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

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

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

$$n(S\Delta_t) \geq R\lambda_b(S)\,|\Delta_t| \qquad (5)$$

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 $M_0 \neq M_{min}$ 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.

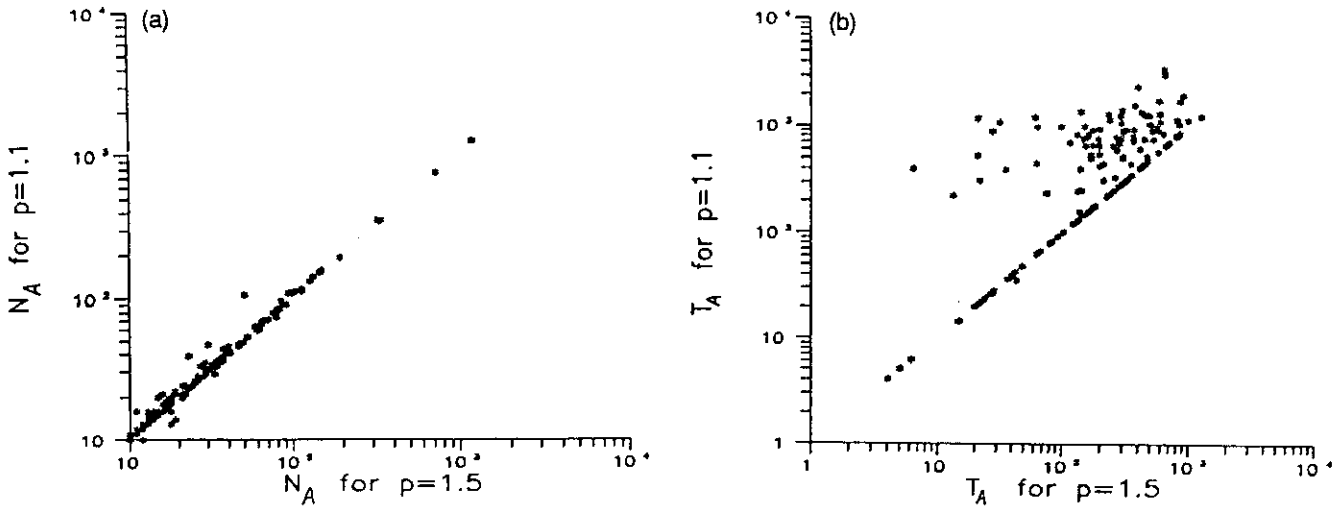

Figure 3. (a) The number N_A and (b) time duration T_A of aftershock sequences identified by the LIR method with the Omori parameter $p = 1.1$ and 1.5 Catalogue NOAA (1964–1980), $M_0 \geq 6$.

(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 $Q =$

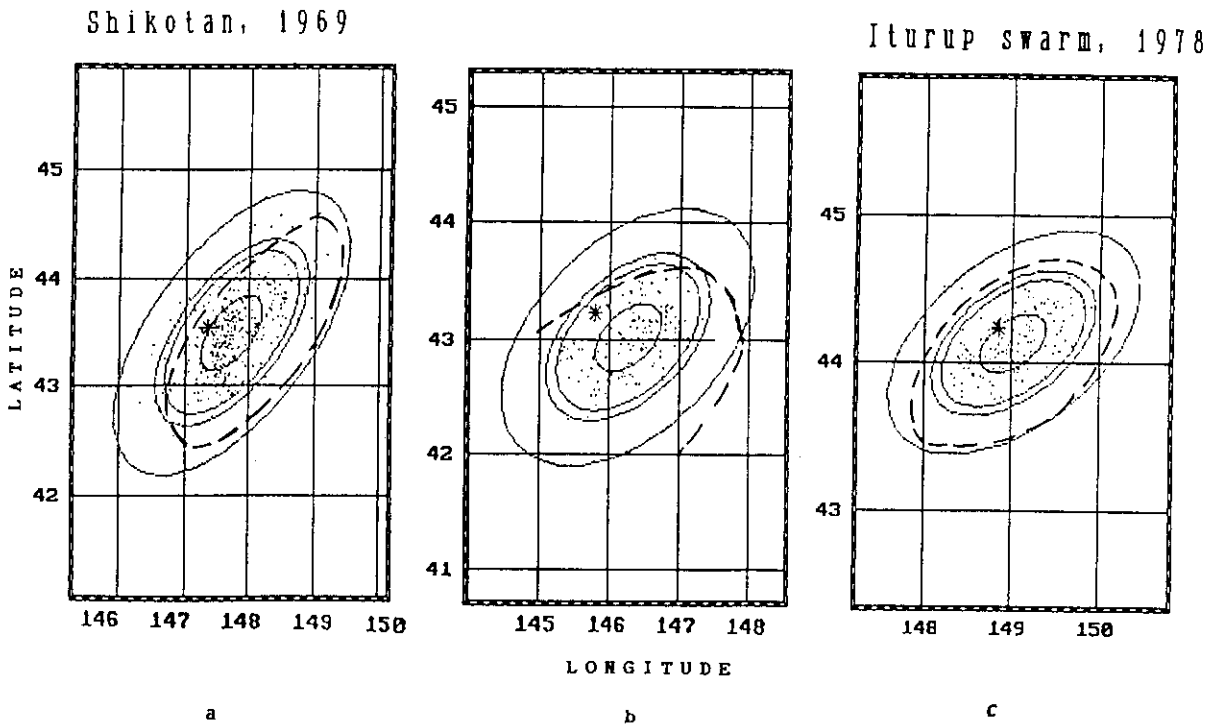

Figure 4. Comparison of the hand procedure (dashed line) with the LIR procedure for South Kuril earthquakes: (a) Shikotan earthquake, 1969 August 11, $M_0 = 7.9$; (*b*) 1973 June 17 earthquake, $M_0 = 7.7$; (c) Iturup swarm, 1978, $M_0 = 7.5$. Mainshocks and aftershocks are mapped by asterisks and dots, 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.

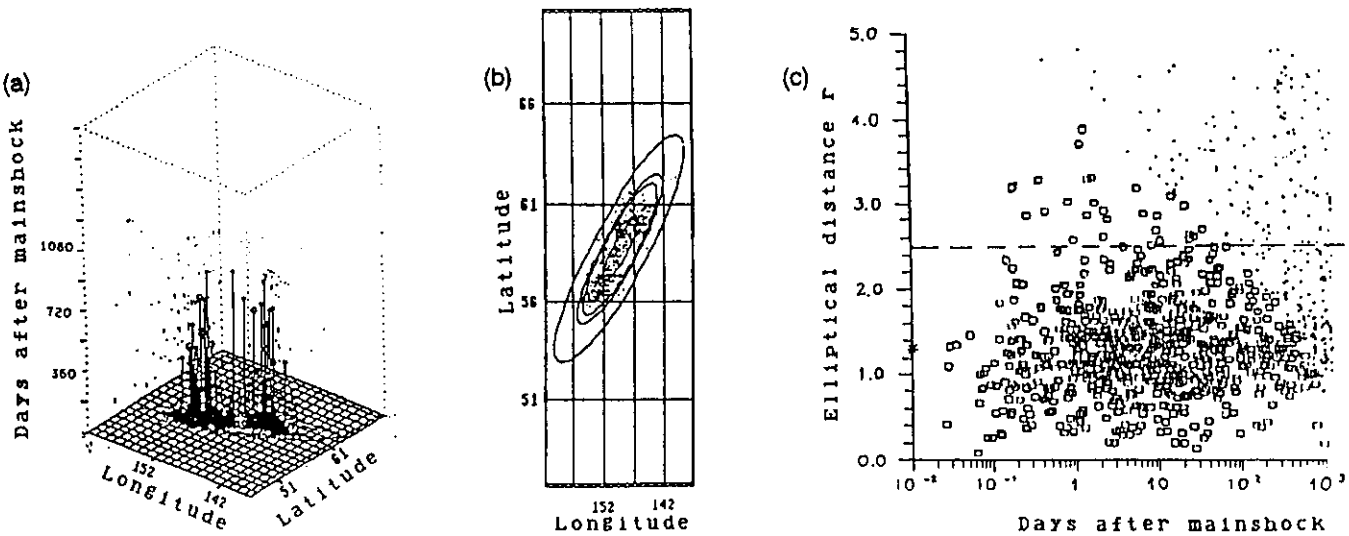

Figure 5. Alaska earthquake of 1964, $M = 8.4$ (NOAA catalogue). (a) Events in space–time: minimax aftershocks are marked by vertical sticks, while background events are marked by dots. (b) Space projection of minimax aftershocks. Aftershocks zones are shown for 60, 90, 95 and 99 per cent confidence 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.

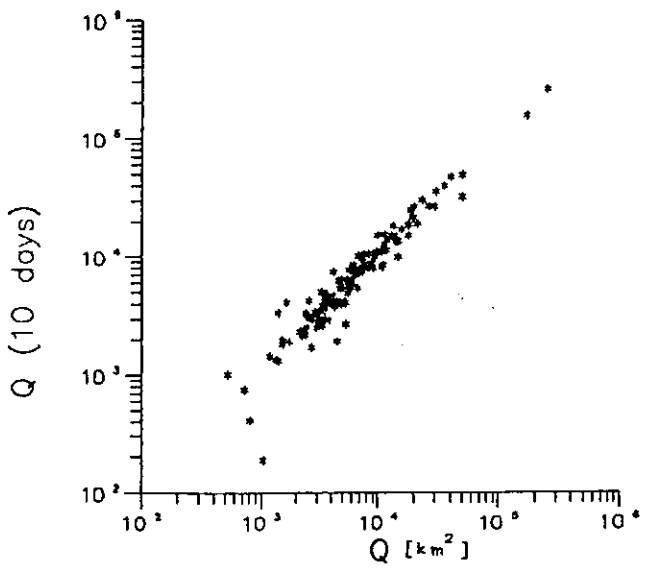

Figure 6. Areas of 95 per cent aftershock zones: for whole aftershock sequences, Q, and for aftershocks of the initial 10 days, Q (10 days) (NOAA catalogue).

(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 $\geq 2:1$.

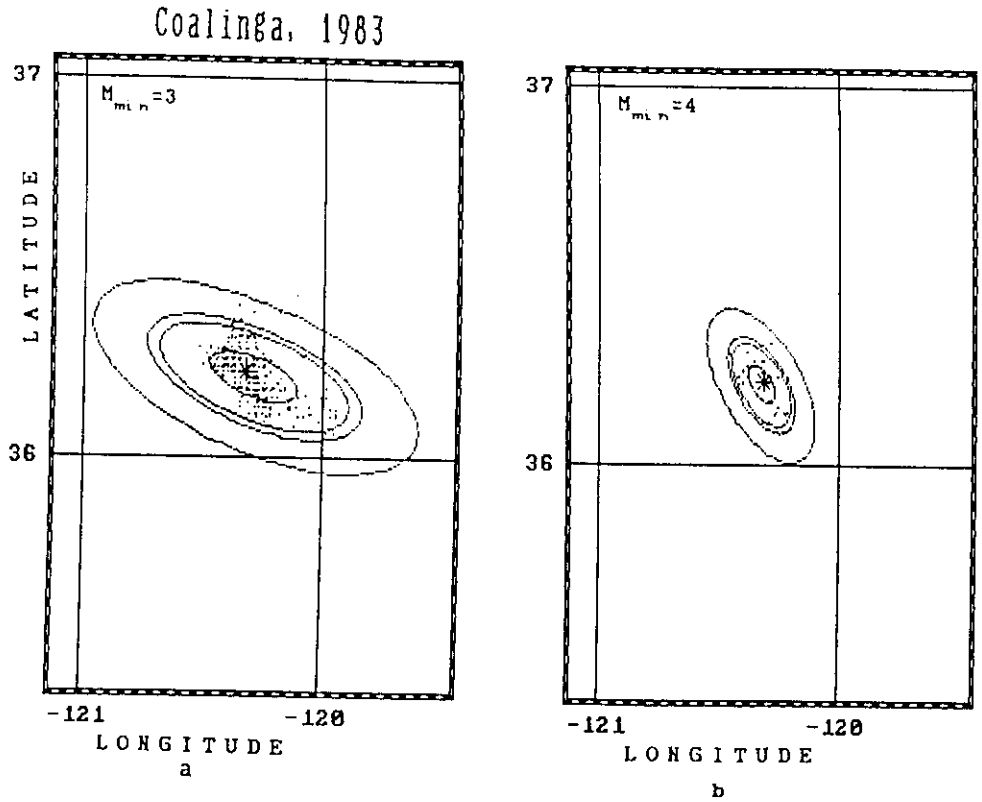

Figure 7. Minimax aftershocks of the Coalinga earthquake (1983, $M_s = 6.7$), identified in the 'West US' catalogue with different cut-off magnitude values: (a) $M_{min} = 3$, number of aftershocks $N_A = 378$; (b) $M_{min} = 4$, $N_A = 48$. Outlined zones correspond to 60, 90, 95 and 99 per cent confidence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

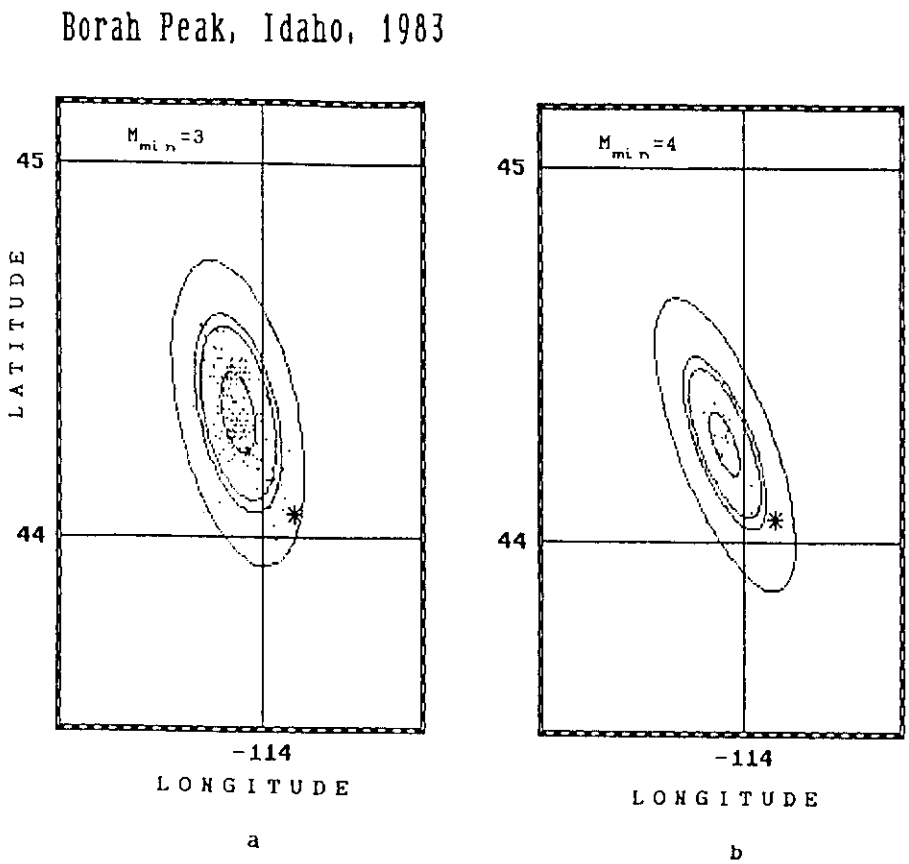

Figure 8. The same as Fig. 7 but for the Borah Peak, Idaho, earthquake (1983, $M_s = 7.3$). (a) $M_{min} = 3$, $N_A = 211$; (b) $M_{min} = 4$, $N_A = 33$.

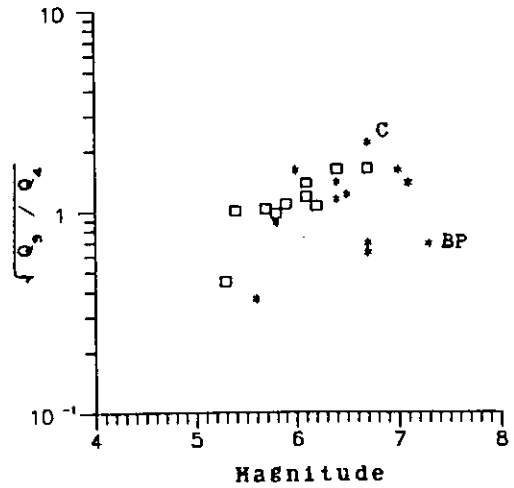

Figure 9. Linear coefficient of extension of the aftershock zone (κ, see text) as a function of mainshock magnitude (West US catalogue). Squares marks aftershock sequences with deficiency in low magnitudes. C and BP mark the Coalinga and Borah Peak earthquakes respectively.

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.

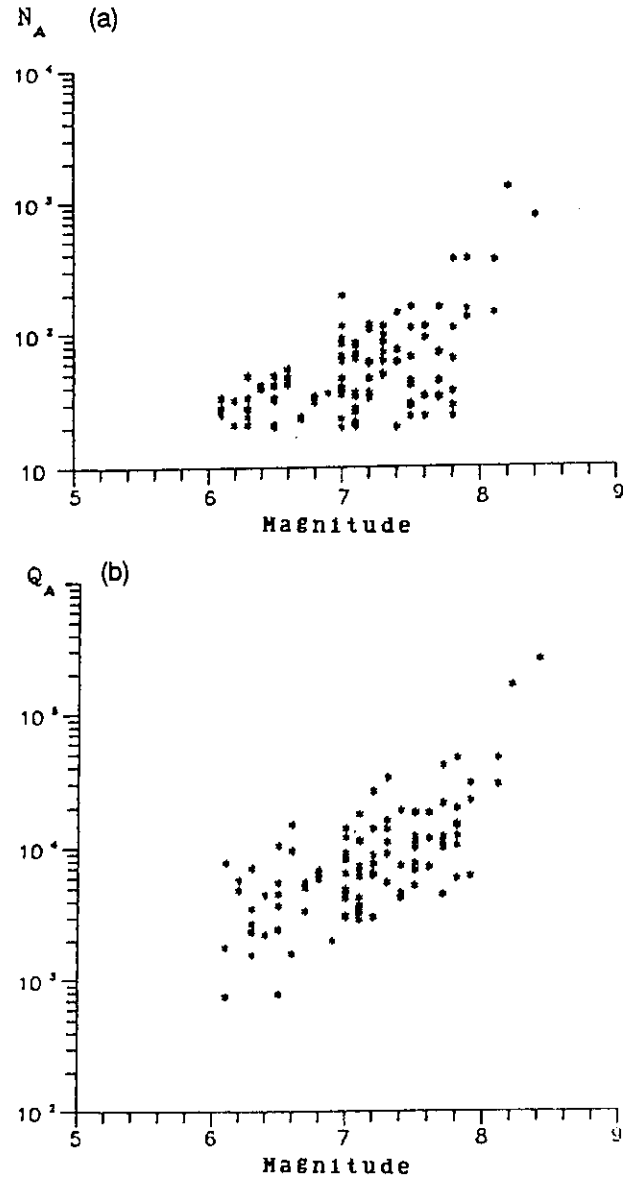

Figure 10. (a) The number of minimax aftershocks N_A (≥ 20) and (b) aftershock zone area Q_A (ellipse of 95 per cent confidence level) as a function of mainshock magnitude (NOAA catalogue, 1964–1980).

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

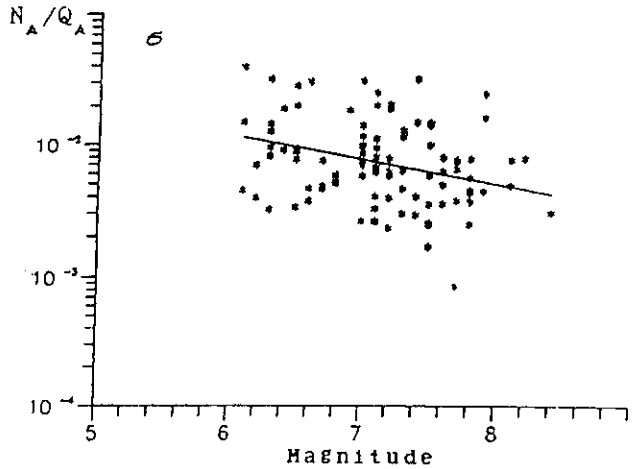

Figure 11. The number of minimax aftershocks per unit of aftershock area (95 per cent) as a function of mainshock magnitude with regression line (notations are the same as in Fig. 9).

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 $p_\infty \ne 0$ 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 /Ro*											
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

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

where

$$c_n = .n/\lambda_b R = \frac{\mu n}{\lambda}$$

and

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

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

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

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

$$J_n = (\mu/\lambda)^n \dot{J}_n.$$

We now verify that $\dot{J}_p = (p+1)^p/(p+1)!$. 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 $(d/dx)^{p-1}(1 - e^x)^p|_{x=0}$. 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 $\mu < 1$. Indeed ; We have

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

where τ_i are independent random variables with density $e^{-(x+\mu)}$, $x \geq -\mu$ and mean $1 - \mu > 0$. Hence (Feller 1966) $p^* = 0$ i.e

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

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

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

if

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

Thus

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

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 for $\varphi(\theta) = E e^{-\theta\nu}$:

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

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

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

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.

ELSEVIER

Physics of the Earth and Planetary Interiors 111 (1999) 229–240

PHYSICS OF THE EARTH AND PLANETARY INTERIORS

Immediate foreshocks: time variation of the b-value [1]

G.M. Molchan *, T.L. Kronrod, A.K. Nekrasova

International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Warshavskoye sh.79, kor.2, Moscow 113556, Russian Federation

Abstract

We compare b-values in the frequency–magnitude law for the foreshocks and long-term seismicity. Foreshocks are considered in two non-overlapping time intervals, preceding a main shock: (i) last hours; (ii) last days. Statistical analysis of the regional and global earthquake catalogs shows, that in either of these intervals b-value drops by a half, compared to long-term seismicity. This extends the previous results by Molchan and Dmitrieva [Molchan, G., Dmitrieva, O., 1990. Dynamics of the magnitude–frequency relation for foreshocks. Phys. Earth Planet. Inter. 61, 99–112], who have found such a drop a few hours before a mainshock. Our results concern a statistical regularity, not necessarily seen for each individual sequence of foreshocks.

Keywords: Frequency–magnitude relation; Foreshocks; Aftershocks; Statistical analysis

1. Introduction

Molchan and Dmitrieva (1990) have studied temporal variations of the magnitude–frequency relation for foreshocks during some hours–days before the mainshock. The events we mean are those which precede the mainshock in the future aftershock area. For simplicity, we shall call them immediate foreshocks. The essence of the difficulty of the study of foreshocks consists of the available quality worldwide catalogs which report all events with magnitude $M \geq 4$–5. Complete foreshock sequences are rare with these cutoff magnitudes, and most of them consist of few events, so that a straightforward estimate of the b-value is out of the question. Nevertheless, if earthquake preparation processes are similar for the places where fore- and aftershocks are possible, then the problem as formulated is meaningful from the statistical point of view (see Section 2). Molchan and Dmitrieva (1990) showed that the b-value for foreshocks (b^-) decreases during 10 days preceding the mainshock and is 50% of the stationary b-value (b_0) during the last hours. This can be of interest for modeling the little known final phase in the preparation of a large earthquake. Recently, Ogata et al. (1995) studied short-term earthquake prediction to confirm, for Japanese data, that a large foreshock of $M > M_{ms} - 0.45$ has an increased probability of occurrence during the last few hours before a magni-

* Corresponding author. Fax: +7-095-310-70-32; E-mail: molchan@mitp.ru

[1] In memoriam A.G. Peozorov.

0031-9201/99/$ - see front matter
PII: S0031-9201(98)00163-0

tude M_{ms} mainshock. This qualitative conclusion is a weaker form of the result as to a decreasing b^-/b_0 ratio.

Quantitative conclusions as to temporal variations of b^-/b_0 may be affected by the raw data used and by the particular technique used for identification of clustered events. It is for this reason that we present an additional analysis for temporal variations of b^-/b_0 in the 10-day interval. To do this, we supplement the 1964–1980 NOAA catalog used previously with 17 years of more data. In addition, we also make use of regional catalogs for southern California, Japan, and a standardized catalog of North Eurasia. Lastly, we use a more accurate technique for identification of clustered events.

Identification of clusters is of obvious importance for our problem. In the previous analysis, we used the mainshock catalog produced by the procedure of aftershock identification of Prozorov (1986) (see also Molchan and Dmitrieva, 1992). This method identified only aftershock events, hence, foreshock sequences found in the reduced catalog could be either contaminated by events of different clusters or, v.v., some of the foreshocks could be lost as false aftershocks. In the present work, fore- and aftershocks of a mainshock are identified simultaneously. To identify aftershocks, we use the minimax procedure due to Molchan and Dmitrieva (1992). This method maintains a trade-off between the expected number of missed aftershocks and the expected number of erroneously identified ones in an aftershock sequence. The aftershocks zones are used for identification of immediate foreshocks (see Section 3). The data and calculations of b^-/b_0 are discussed in Sections 4 and 5, respectively.

As a whole, the conclusion by Molchan and Dmitrieva (1990) can be refined as follows: the b^-/b_0 ratio decreases to the value 0.5 a day before the mainshock.

For convenience, some results and elements of the previous analysis by Molchan and Dmitrieva (1990, 1992) are recalled below.

2. The problem and methodology

To analyze immediate foreshocks, we have to make the best use of earthquake catalog data, i.e., to take into consideration all the magnitudes M, and not only those which had been completely reported. Thus, we shall describe a normalized frequency–magnitude law in region G by the function

$$f(M) = p_G(M)10^{-b_G M}Z_G^{-1}, \qquad M_- \le M \le M_+ \tag{1}$$

where (M_-, M_+) is a range of magnitudes, Z_G is a normalizing constant such that $\int f(M)\mathrm{d}M = 1$, b is the slope, the function $p(M) \le 1$ measures the incompleteness of smaller events: $p(M) = 1$ for the range of completely reported magnitudes $M \ge M_0$. For the present analysis, the following assumptions are important: (a) the function $p(M)$ does not depend on the type of event being recorded in the magnitude range (M_-, M_+): it may be either a main event, a foreshock, or an aftershock.

The requirement assumes symmetry in the recording of foreshocks and aftershocks. This is evidently violated when a catalog incorporates data supplied by a local network deployed after a large earthquake.

Under condition (a), it is natural to assume that: (b) the distribution of type (1) is correct for immediate foreshocks within. In other words, the function $p_G(M)$ remains the same for foreshocks; M_+ is the mainshock magnitude M_{ms}, and only the parameter $b_G^- = b_G^-(\tau)$ is a function of time (τ is the time before mainshock).

Our goal is to analyze temporal variations of $\theta = b_G^-/b_G$ within some hours–days before mainshocks. To be more exact, we consider here temporal variations of θ within three logarithmically equal time intervals preceding a mainshock:

$$(-10, -1) \text{ days}, \ (-1, -0.1) \text{ days},$$

$$(-0.1, 0.01) \text{ days}. \tag{2}$$

The choice of the ranges is determined by Omori's law: aftershock occurrence is approximately constant over time on the log scale; the same may be presumed for foreshock rate as well (see for example, Ogata et al., 1995).

For a practical analysis of the θ-value, it is necessary to make one more assumption: (c) θ-values do not depend upon the location of foreshock sequence,

i.e., $\theta = \theta_\Delta$ is a function of the $\Delta\tau$ time interval and, probably, of the mainshock magnitude only.

The b_G-value describes the relation between smaller and larger events in the region, while θ reflects relative changes in this relation during the rearrangement of crustal stress before the mainshock. Thus, assumption (c) expresses a certain similarity pattern of such processes in the media where fore- and aftershocks occur; it is a similarity, since θ does not depend upon local seismicity parameters.

Thus, assumption (c) imposes certain limitations on the medium. Taking into account the data selected for the analysis, we actually limit ourselves to media in which mainshocks are accompanied by a sufficiently big number of aftershocks ($n_\Lambda \geq 10$) and are preceded by foreshocks. According to Mogi (1963), earthquakes of the foreshock–mainshock–aftershock type can occur only in 'weakly homogeneous' media.

Now we recall the statistical procedure employed to estimate θ.

The function $p_G(M)$ and the b_G-value are long-term local characteristics. They can be easily estimated from the whole earthquake catalog in region G. Thus, when estimating θ_Δ we may suppose that $p_G(M)$ and b_G are given, i.e., the foreshock b-value in a region G_j is $b_j^- = \theta b_j$, where $\theta = \theta_\Delta$ and $b_j = b_{G_j}$.

Suppose we have m foreshock sequences with magnitudes

$$\left\{M_1^i, \cdots M_{n_i}^i\right\}, \qquad i = 1, \cdots m;\ n_i \geq 1$$

in a fixed time range $\Delta\tau$ before the mainshocks. According to our assumptions, the elements of each sample obey the distribution with the density

$$p_i(M)10_\Delta^{-b_i\theta M} Z_i^{-1}(\theta_\Delta), \qquad M_- < M < M_i^+,$$

where M_i^+ is the mainshock magnitude for the ith foreshock sequence and b_i is identical with b_G if the ith sequence belongs to the region G. The estimate for θ_Δ is found from the moment-type equation

$$\sum_{i=1}^{m} \hat{\mu}_i\left(\overline{M}_i(\theta) - \hat{M}_i\right) = 0,$$

where $\hat{M}_i$ is the arithmetic mean of $\{M_k^{(i)},\ k = 1, \ldots, n_i\}$,

$$\overline{M}_i(\theta) = \int_{M_-}^{M_i^+} x p_i(x) 10^{-b_i\theta x} \mathrm{d}x Z_i^{-1}(\theta)$$

is the theoretical average magnitude for the same sample, and

$$\hat{\mu}_i = n_i b_i / N, \quad N = \sum_{i=1}^{m} n_i$$

are weights proportional to sample size.

Molchan and Dmitrieva (1990) found that for short foreshock samples, the standard deviation of $\hat{\theta}_\Delta$ takes the form

$$\sigma_\theta \cong \theta_\Delta N^{-1/2}\kappa \tag{3}$$

where κ depends on $\theta\hat{b}$ and $\hat{M}^+ - M_-$. Here $\hat{b}$ and $\hat{M}^+$ are the arithmetic means of b_i and M_i^+, respectively. In particular, if $\theta b_i \cong 0.5$ and $\hat{M}^+ - M_- = 2$ then $\kappa \cong 2$.

3. Foreshock identification

We are interested in *immediate foreshocks*, i.e., events that precede a large earthquake later than 10 days before and occur in the future aftershock area. In the context of the present study, it is not necessary to identify all such events. It is enough to have reliable sequences with the total number of events sufficient for estimating the parameter θ (see Eq. (3)). The above considerations determine the following method for foreshock identification.

Suppose we know the aftershock zone Q (see below). The criterion that there are foreshocks in a space–time volume $v = Q \times \Delta\tau$ is based on the fact that the expected average number of events (foreshocks plus background) in the volume v must be significantly greater than the long-term average $\bar{n}_v$. If mainshocks and foreshocks are considered to be Poissonian, the solution to the problem is well-known (see e.g., Bolshev and Smirnov, 1983). Specifically, suppose the long-term average $\bar{n}_v$ has been well-estimated. Then we say that foreshocks occur with confidence level $(1 - \varepsilon)$, if $\bar{n}_v < n(\varepsilon)$. Here $n(\varepsilon)$ is a function of the observed number of events n_v in v.

For the significance level $\varepsilon = 5\%$ applied here and $n_t \le 10$, the values of $n(\varepsilon)$ are as follows:

n_t	1	2	3	4	5	6	7	8	9
$n(\varepsilon)$	0.05	0.35	0.82	1.37	1.97	2.61	3.29	3.98	4.70

When $n_t > 10$, the quantity $n(\varepsilon)$ is well approximated by the expression:

$$n(\varepsilon) \cong \left(\sqrt{n_t + d} - 0.5\psi_\varepsilon\right)^2 - d,$$

where $\psi_\varepsilon = 1.6448$ is $(1-\varepsilon) = 95\%$-quantile of the Gaussian distribution and $d = 0.25$ or $d = 0.375$ for $n < 20$ and $n \ge 20$, respectively.

Aftershock zone Q is fitted by an ellipse. That zone is easily found from statistical considerations by assuming a Gaussian scatter of the aftershock epicentres. Let (g_0, R) be the centre and dispersion matrix, respectively, for the aftershocks, while the quadratic form

$$r^2(g, g_0) = (g - g_0) R^{-1} (g - g_0)$$

defines the elliptic metric. Then, a region of the form

$$S_k = \{g : r(g, g_0) < k\}$$

is assumed to be the aftershock zone with the confidence coefficient

$$\alpha(k) = \mathrm{Prob}(\xi \in S_k).$$

Here, ξ is a 2D Gaussian vector with parameters (g_0, R). If (g_0, R) are known exactly, then $k^2 = -2\ln(1-\alpha)$. If, on the other hand, they are based on N_a independent observations, then

$$k^2 = (N_a - 1)\left((1-\alpha)^{-2/(N_a-1)} - 1\right) \times \left(1 + (N_a - 2)^{-1}\right)$$

(see Molchan and Dmitrieva, 1992). In practice, we determined aftershock zones with $N_a > 10$ and $\alpha = 95\%$. In that case $k = 3.1$, 2.7, 2.6, and 2.4, if $N_a = 10$, 20, 30, and ∞, respectively.

Aftershock identification is based on the minimax principle, which equalizes the expected number of missed events in an aftershock sequence and the expected number of background events that have been erroneously identified as aftershocks. For more details, see Appendix A.

4. Data

The main source of foreshock sequences for the present study is the PDE catalog of NEIC/USGS, which is freely available in the Internet. We used the period 1964–1997 and the depth range $H \le 100$ km. The cutoff magnitude for this catalog was generally $M_- \cong 4.0$. However, $M_- = 2.5$–3.5 for some space–time volumes. This concerns, in particular, the Caribbean and Hellenic arcs and the part of Europe from Gibraltar to Italy. By definition, the quantity $M = \max(m_b, M_S)$ was used as the working magnitude. When these two magnitudes m_b and M_S were not given, the first of those available in the series $\{M_L, M_p\}$ was taken for M. The notations m_b, M_S and M_L are the standard ones; M_p means that the earthquake size was based on individual stations or on data of a regional network.

Additional aftershocks were taken from regional catalogs: catalog SCSN for southern California (available in the Internet at SCEC-DC), 1935–1994, $M = M_L$, $M_- = 3$ and for Japan 1983–1994, $M = M_{JMA}$, $M_- = 2.8$–3. [2] For oceanic regions the last catalog is less representative than PDE. For this reason, the JMA catalog was used for the onshore part of the region only. Generally however, when a main event occurred in several catalogs, we preferred the foreshocks in the regional catalog. Also, we used the Standardized Catalog of North Eurasia for the area of the former Soviet Union (UNIC), 1957–1997, $M = M_{LH}$ (Kondorskaya et al., 1997).

It should be emphasized that the nonuniformity of the working magnitude M is not a serious drawback for a worldwide analysis of the time behaviour of θ. This is so because the ratio $\theta = b_G^-/b_G$ is essentially a spatially local characteristic that is invariant under linear transformations of the magnitude scale. For checking purposes, we use the work of Gusev (1991) to show in Fig. 1 the regression curves of M_L, M_{LH} and M_{JMA} vs. $\max(m_b, M_S)$. From these curves, it is

[2] JMA Catalogue, Global Hypocenter Data Base CD ROM, Version 3.0, USGS/NEIC, Denver, CO, USA.

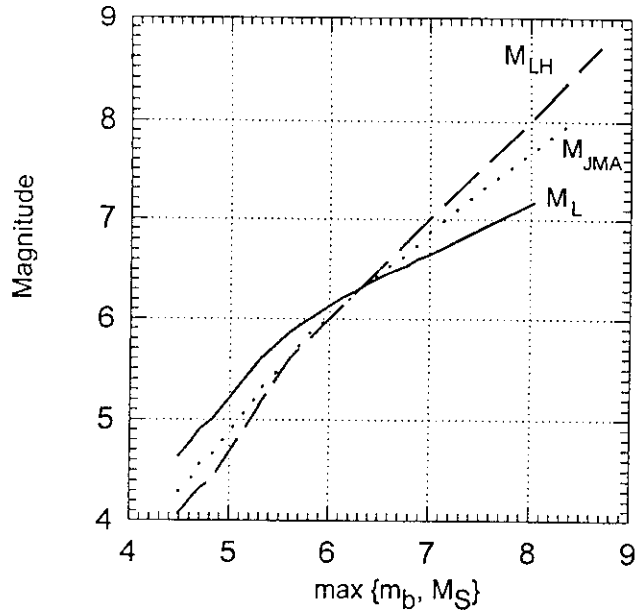

Fig. 1. Magnitude relations (modified from Gusev, 1991): M_L, M_{JMA} and M_{LH} vs. $\max(m_b, M_S)$.

apparent that some distortions in θ are possible for foreshock sequences measured on the M_L, M_{LH} scale, when a sequence includes events of $M > 6$. There are a few such foreshock sequences in the regional catalogs: two in SCSN and four in UNIC. We have excluded them from our analysis.

5. Estimation of local $p_G(M)$, b_G parameters

The fu ction $p_G(M)$ and the b_G-value are considered as long-term characteristics of the site G where the mainshock occurred. Therefore, they are estimated from the catalog for the longest period possible. The period is determined by requiring stable seismicity rates for different magnitudes at a given location. Based on these considerations, we used our seismotectonic global regionalization with elements given by region G (Fig. 2). The time period T was divided into two or three segments, depending on the region G and the particular catalog used, assuming $p_G(M)$ and b_G to be constant in such segments (Table 1).

The incompleteness parameter for $p_G(M)$, $M \in (M_-, M_0)$ was estimated for each magnitude at a step of $\Delta = 0.1$, with at least 50 events to be used for the estimate. That requirement defines the cutoff magnitude M_- uniquely. The other cutoff magnitude gives the beginning of log linearity in the frequency–magnitude relation. The values of M_- and M_0 for each region are shown in Fig. 3.

Our estimate of the b-value, b_G, is based on maximum likelihood estimates of the parameter b in the exponential distribution $10^{-bM}/\int_{\Delta M} 10^{-bx}\mathrm{d}x$, $M \in \Delta M = (M_0, M_+)$ for grouped data at a constant spacing δ ($\delta = 0.1$ or 0.2). The upper limits M_+ for all regions are also shown in Fig. 3. The estimate for $p_G(M)$ is $\hat{p}_G(M) = N_G(M)/\hat{N}_G(M)$, where $N(M)$ is the observed number of events with magnitude $M \pm \Delta/2$ and $\hat{N}_G(M) = \hat{\alpha}10^{-\hat{b}M}$ as its estimated value. The weight functions p_G were estimated in two ways, with and without aftershocks. The estimates of b_G are shown in Fig. 4 with Fig. 5 displaying some examples of estimates of p_G. In these examples both estimates of p_G are practically identical. However, the overall preference should be given to that based on mainshocks, since some of the earlier aftershocks may be missed when occurring in rapid succession. As a result, assumption (a) in Section 2 may sometimes be violated.

The estimates of $(p_G(M), b_G)$ for the Hellenic arc should be viewed with some caution, for two reasons. First, the seismicity involves many foreshock sequences (see Fig. 2) and, secondly, the PDE catalog is inhomogeneous in magnitude for that region, the events with missing m_b and M_S being 50% of the total number.

6. Results

We intend to study the dynamics of the b-value for foreshocks occurring within 10 days of the mainshock. The dynamics of b^- is characterized by the variation of θ within three logarithmically equal time intervals (2): $\Delta_1 = (-10, -1)$, $\Delta_2 = (-1, -0.1)$, and $\Delta_3 = (-0.1, -0.01)$ days. The worldwide and the three regional catalogs contain 216 *small* sequences identified for this purpose with the number of foreshocks equal to $N_f < 20$ in the 10-day interval and 19 *large* sequences, i.e., ones with $N_f \geq 20$. Fig. 2 shows the mainshocks preceded by these sequences; 95% of the mainshocks with $N_f <$

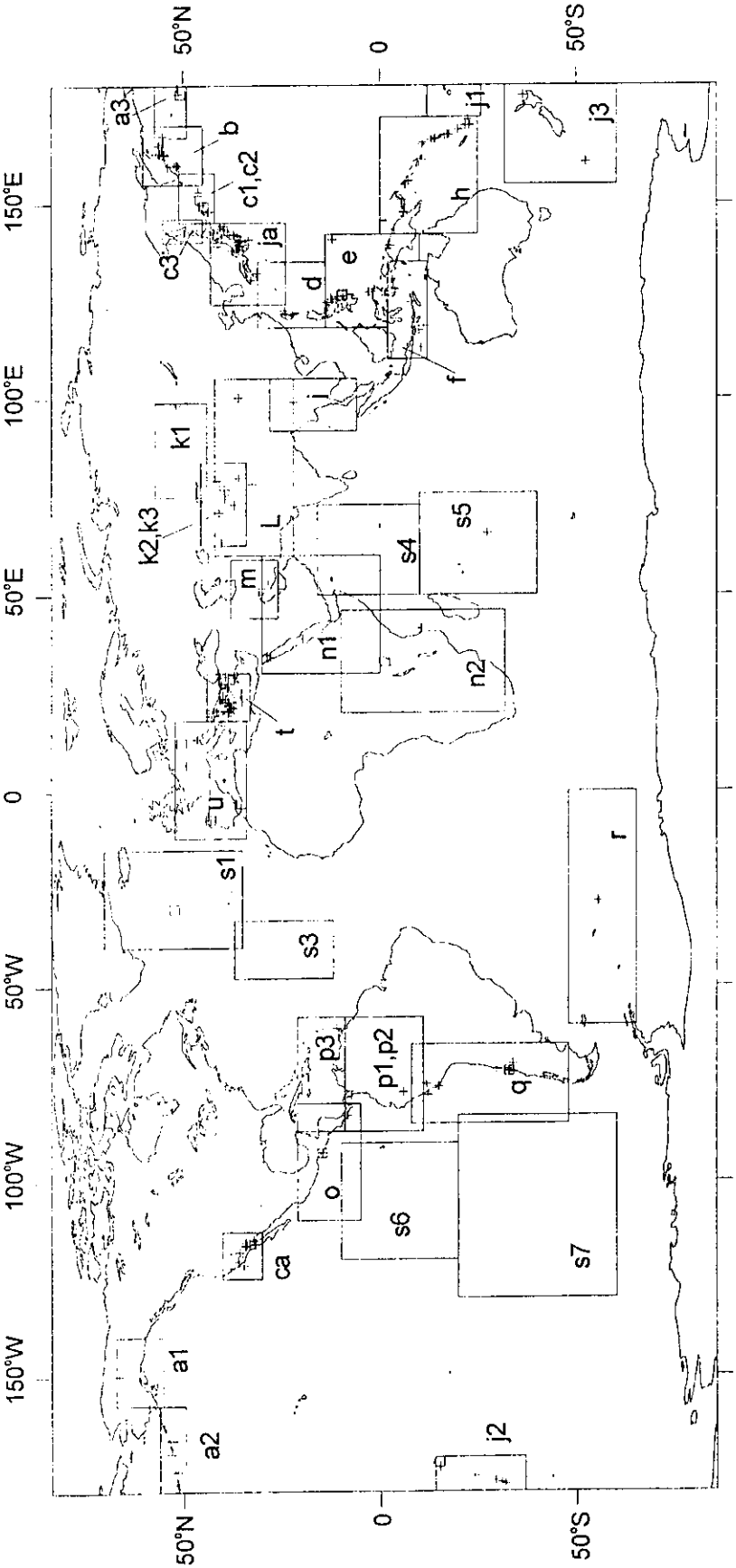

Fig. 2. Seismotectonic regionalization and mainshocks of the earthquake clusters used in the analysis of θ. Notation: ($s3$) index of a regionalization element (for full name see Table 1); (cross) mainshock with $N_f < 20$; (square) mainshock with $N_f \geq 20$.

Table 1
Space–time partition of seismic catalogs for estimation of $(p_G(M), b_G)$ (see main text)

Index	Time period and catalogue	Region
A1		Alaska
A2		Western Aleutians
A3		Eastern Aleutians
B1	UNIC	Kamchatka
B2	1991–1997	
C1	UNIC	Kuril arc
c2	1991–1997	
c3	UNIC	Sakhalin
ja1	JMA, 1982–1987	Japan Islands
ja2	JMA, 1988–1990	
ja3	1964–1980	Japan
ja4	1981–1997	
d1	1964–1994	Taiwan
d2	1995–1997	
e1	1964–1979	Philippines and Guinea
e2	1980–1994	
e3	1995–1997	
f1	1964–1979	Java to Band Sea
f2	1980–1994	
f3	1995–1997	
g1	1964–1982	Sumatra
h1	1964–1994	Solomon and New Hebrides arcs
h2	1995–1997	
I1	1964–1994	Indo-Birman ridges
I2	1995–1997	
j1		Law islands
j2		Tonga and Kermadec
j3	1981–1997	New Zealand
k1	UNIC	Altai and Sayany
k2	UNIC, 1957–1977	Middle Asia (former SU)
k3	UNIC, 1978–1990	
L1	1964–1973	Central Asia
L2	1974–1994	
L3	1995–1997	
m1	1981–1997	Zagros
m2	1964–1980	
n1	1966–1997	Red Sea to Owen fault
n2	1966–1997	S. Africa rifts
ca	SCSN	S. California
o1	1981–1997	Central American arc
o2	1964–1980	
p1	1981–1997	Bolivian Andes and Caribbean arc
p2	1964–1980	
p3	1986–1997	Caribbean arc
q1	1964–1991	Andes
q2	1992–1994	
q3	1995–1997	
r		S. Sandwich arc
s1	1981–1997	N. Atlantic (Lat. > 35°)
s2	1964–1980	
s3		N. Atlantic (Lat. < 35°)
s4		Central Indian ridge
s5		Triple junction in Indian Ocean
s6		Cocos plate
s7		Nazca plate
t1	1964–1980	Greece and Hellenic arc
t2	1981–1988	Hellenic arc
t3	1989–1997	
t4	1981–1988	S. Continental Greece
t5	1989–1997	
u1	1964–1975	Europe
u2	1976–1981	
u3	1982–1987	

First column: index of the partition element mapped in Fig. 2.
Second column: catalog (PDE implied if omitted) and respective time interval (years).
Third column: name of seismotectonic region.

20 belonging to the magnitude range 4.9–7.8. The large sequences have mainshocks of $M \in (4.6–7.9)$.

We eliminated 68 clusters from the main calculation of θ; in these, the largest foreshock (22 cases) or the largest aftershock occurring within 10 days (50 cases) was different from the mainshock magnitude by 0.1. Foreshocks are difficult to identify in such clusters due to magnitude uncertainties. Besides, many such clusters can be treated as swarms, which are defined as containing several strong events of roughly the same size. In Section 4, we have rejected another set of six regional clusters.

One important conclusion made by Molchan and Dmitrieva (1990) is that $\theta = b^-/b \cong 0.5$ within Δ_3 and $\theta \leq 0.9$ within the two other intervals. Our new analysis updates this result primarily for the Δ_2 interval.

We begin by considering *small foreshock sequences* ($N_f < 20$) with all the exceptions mentioned above. Small sequences are the most typical among all foreshocks, hence, provide a more faithful representation of the statistical patterns proper to immediate foreshocks, since they involve all of the earthquake-generating area (see Fig. 2). It is also a well-known fact that foreshocks are followed by aftershocks of their own (Molchan and Dmitrieva, 1990). For this reason, the dynamics of θ for large foreshock sequences bears the imprint, not only of the particular location involved, but also of patterns

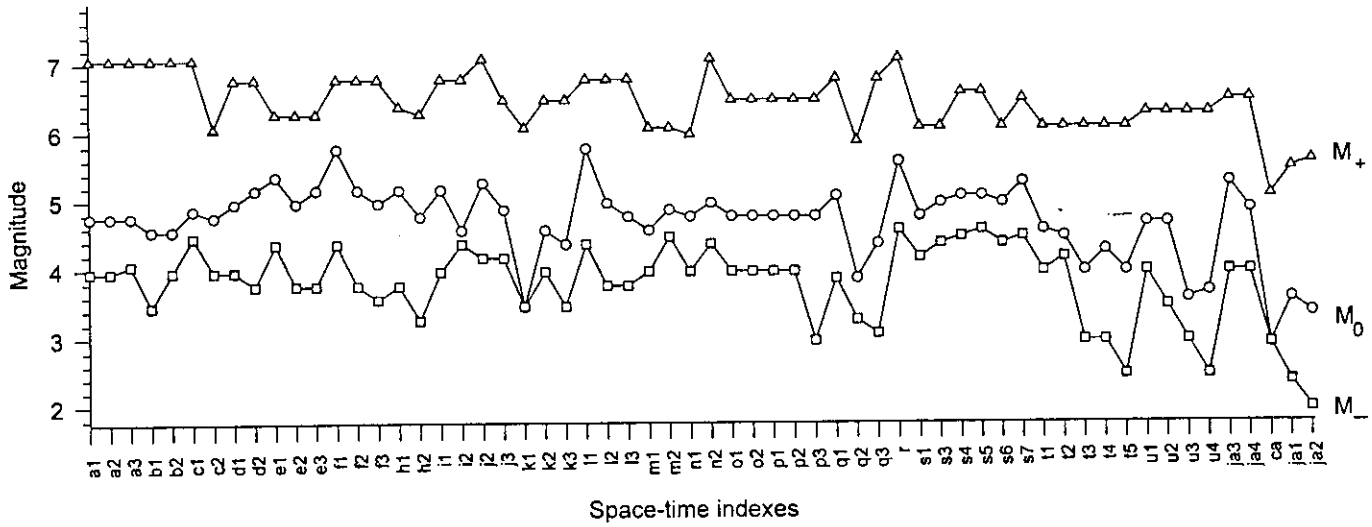

Fig. 3. Magnitude cutoff values M_-, M_0, M_+ (see the text) used for regional frequency–magnitude relations (regions are as in Fig. 2).

peculiar to a mixture of immediate foreshocks and aftershocks.

To sum up, if $N_f < 20$, then the estimates of θ for Δ_1, Δ_2, Δ_3 are as follows:

θ:	0.79 ± 0.08	0.49 ± 0.09	0.51 ± 0.09
n_{ms}:	96	65	85
n_{for}:	288	164	165

(4)

Here $n_{ms}(\Delta)$ is the number of mainshocks with statistically significant foreshocks, and $n_{for} = n_{for}(\Delta)$ is the total number of foreshocks in the Δ-interval. The scatter in θ is consistent with the rough estimate of standard deviation for the estimate of θ as given by Molchan and Dmitrieva (1990).

The θ-estimates (4) are stable enough under changes in the definition of a small foreshock sequence, rejection of data and under errors resulting from the choice of b_G and weights p_G. Below we examine these variations in more detail.

The requirement $N_f < 20$ does not rule out 19 events in a foreshock sequence occurring in one of the intervals Δ_i, $i = 1,2,3$. We require that these

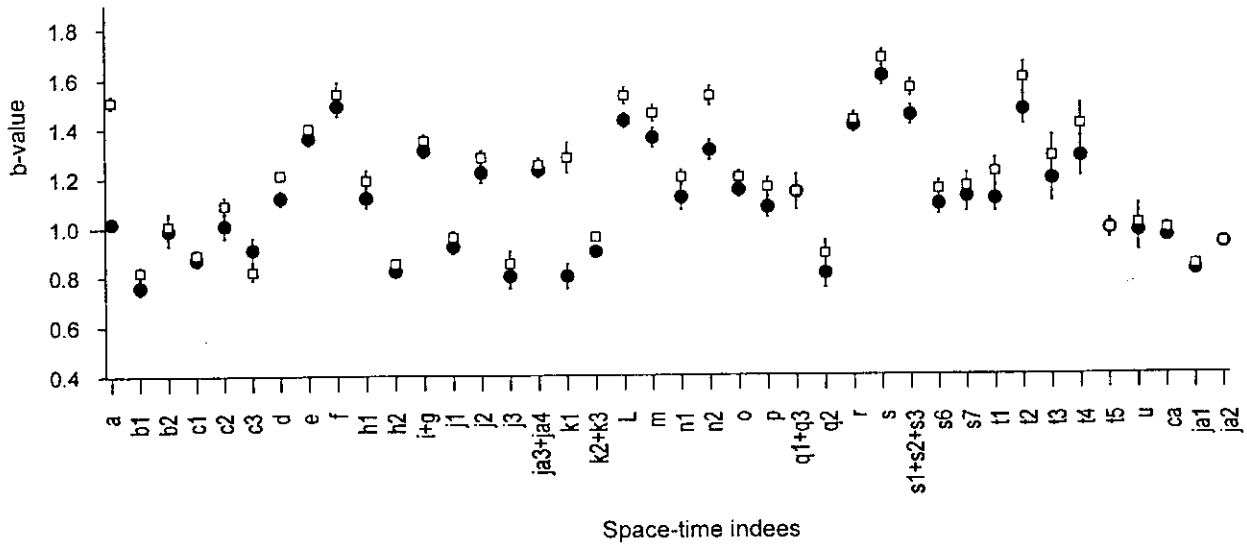

Fig. 4. b-value with standard deviation for the regions given in Fig. 2; (filled circle) b-value for mainshocks; (square) b-value for all events.

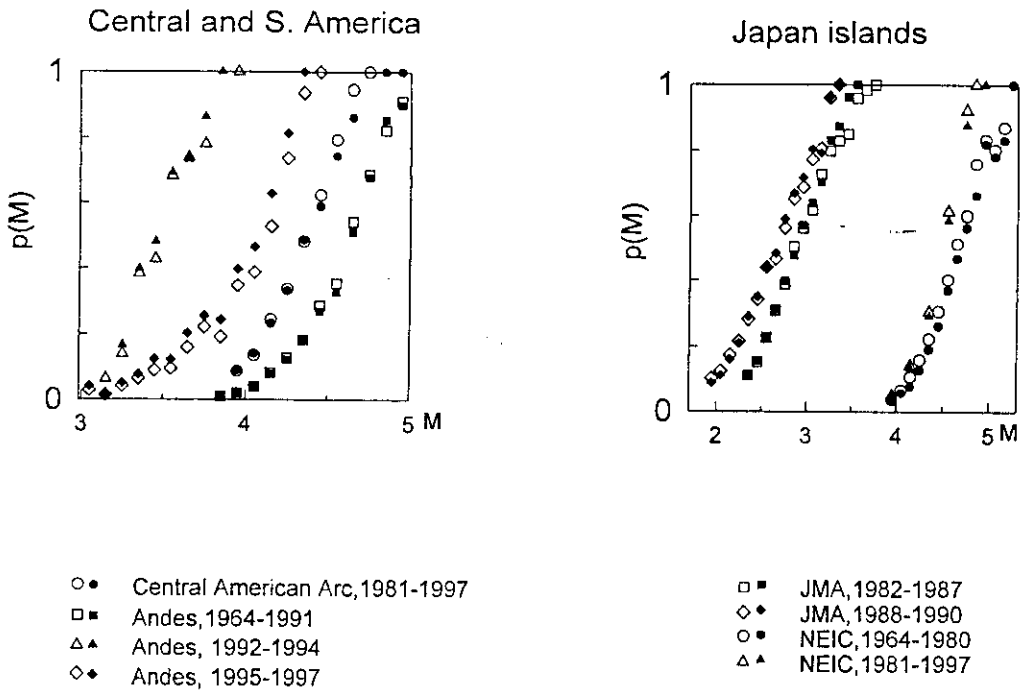

Fig. 5. Examples of the weight function $p_G(M)$ for two types of catalog: a catalog with aftershocks eliminated (filled symbols) and the complete catalog (open symbols).

numbers $N_f(\Delta_i)$ be smaller than 10. In that case the estimates of $\theta(\Delta_i)$ will vary within 0.01.

Let us narrow the foreshock magnitude range in order to use only completely reported events, $M < M_0$ In that case $p(M) \equiv 1$, the foreshock data are reduced by a half, but this does not significantly affect the estimates of θ:

θ: 0.64 0.39 0.45

(compare with (4)).

Let us lift the prescription to reject doubtful foreshock sequences. Namely, we add to the main variant the 50 clusters in which the largest aftershock is 0.1 below the mainshock, and include Greece and six events from the regional catalogs that had been eliminated on account of nonlinearities in the M_L, M_{LH} scales. In this case, the variations of θ do not exceed 0.02:

θ:	0.81 ± 0.06	0.50 ± 0.08	0.49 ± 0.08
n_{for}:	463	273	241

(5)

Replace the regional estimates of the b-value (b_G) by local ones obtained for a mainshock of magnitude M in a circle of radius $R = \max(10^{0.5M-1}, 100)$ km. Then the θ-estimates (4) will be modified within the original errors:

θ:	0.74 ± 0.09	0.52 ± 0.12	0.53 ± 0.12
n_{for}:	199	112	115

This variant uses those areas of mainshocks where the b-values are reliable enough ($\sigma_b < 0.1$).

6.1. Large foreshock sequences ($N_f \geq 10$)

We have no actually observed samples of large foreshock sequences such as to provide a reliable time behaviour of θ in intervals (2). For this reason, we are forced to treat the 18 selected sequences with $N_f \geq 20$ (see Fig. 2) as a statistical population. The population has turned out to be inhomogeneous as to the time behaviour of θ, involving as it does four foreshock sequences localized in the Izu Peninsula, Japan. That area is known to exhibit some peculiarities, because it is a junction of three global tectonic plates: the Eurasian, the North American, and the Philippine plate (Yoshioka et al., 1993). The esti-

mate of $\theta(\Delta_2)$ is 2.5 times that for small foreshock sequences:

$$\theta(\Delta_1) = 0.98 \pm 0.24, \quad n_{for} = 68, \quad M > M_0$$
$$\theta(\Delta_2) = 1.26 \pm 0.31, \quad n_{for} = 57, \quad M > M_0$$
$$\theta(\Delta_2) = 1.29 \pm 0.07, \quad n_{for} = 560, \quad M > M_-$$

Here and below, we quote estimates of θ for $n_{for}(\Delta_i) > 20$ alone, those of $\theta(\Delta_i)$ for the entire range $M > M_-$ being only given when they are consistent with the estimates of θ based on the events of completely reported magnitudes, $M > M_0$. The bulk of the large foreshock sequences (14) is much closer to the set of small sequences with respect to the estimate of θ:

$$\theta(\Delta_1) = 0.75 \pm 0.10, \quad n_{for} = 244, \quad M > M_0$$
$$\theta(\Delta_2) = 0.71 \pm 0.19, \quad n_{for} = 62, \quad M > M_0$$
$$\theta(\Delta_2) = 0.78 \pm 0.12, \quad n_{for} = 128, \quad M > M_-$$

The relative increase in $\theta(\Delta_2)$ for large foreshock sequences compared with the smaller ones was to be expected because of the presence of foreshock-induced aftershocks. However, these changes are abnormally high for the Izu Peninsula.

Molchan and Dmitrieva (1990) quote seven examples of well-studied large foreshock sequences taken from Papazachos et al. (1967), Suyehiro and Seiya (1972), Gettrust et al. (1981), Jones et al. (1982) and Comte et al. (1986). The estimates for these sequences are $\theta \approx 53 \pm 0.08$ in time intervals of 1 to 3 days before the mainshock and $\theta \approx 0.83 \pm 0.18$ within $\Delta = (-10,0)$ days. This is in good agreement with the estimates of θ based on the short foreshock sequences.

As a whole, the values of $\theta(\Delta_1)$ and $\theta(\Delta_2)$ for large foreshock sequences are below 1; the rare exceptions localized in the Izu Peninsula constitute an object of special interest for further study.

6.2. θ as a function of magnitude

Fixed time intervals $\Delta\tau$ were used to examine the time behaviour of θ. Assuming a time similarity for the preparation of earthquakes for different magnitudes, it follows that one and the same interval will correspond to different phases in the preparation of the mainshock, depending on the mainshock magnitude. To obviate this difficulty, we list in Table 2 estimates of θ in the same intervals Δ_i, but for different narrow magnitude ranges, namely, 5.0–6.0, 5.6–6.6, and 6.2–7.2. These ranges are made to overlap to provide enough data for estimation; the estimate of θ was derived from all foreshock sequences, i.e., from the same used sample as that to get (5).

Table 2
θ as a function of mainshock magnitude (see the text for the notation)

ΔM	$\Delta\tau$ (days)	n_{ms}	n_{for}	$\theta \pm \sigma_\theta$
5.0–6.0	0.1–0.01	37	73	0.35 ± 0.18
	1–0.1	31	104	0.47 ± 0.16
	10–1.0	56	150	0.81 ± 0.13
5.6–6.6	0.1–0.01	53	101	0.41 ± 0.13
	1–0.1	37	99	0.52 ± 0.14
	10–1.0	55	156	0.82 ± 0.11
6.2–7.2	0.1–0.01	46	98	0.49 ± 0.12
	1–0.1	41	100	0.47 ± 0.11
	10–1.0	46	124	0.78 ± 0.11

Table 2 demonstrates that the time behaviour of θ for small foreshock sequences is practically independent of the magnitude range and is similar to the main variant (4).

The above conclusion is at variance with the Berg (1968) result that the parameter b^- for immediate foreshocks increases with mainshock magnitude, M_{ms}. In addition, for the case $M_{ms} \leq 6.5$, the foreshock magnitudes automatically belong to the linear part in the relations of M_L, M_{LH} and M_{JMA} vs. $\max(m_b, M_S)$. Therefore, one is free to use any magnitude scale in the estimation of θ. Now since the time behaviour of θ remains stable under changes in M_{ms}, this provides an extra corroboration that the diversity of magnitudes only affects a little the problem considered here.

7. Conclusions

We have studied temporal variations of normalised b-value for the foreshocks, $\theta = b^-/b_0$ during the last 10 days before a mainshock. Our main result is that $\theta \cong 0.5$ during the periods 0.1 day and 0.1–1 day prior to a mainshock. This precisely worded result is based on the hypothesis that θ does

not depend on location, so that we pooled together more than 200 small foreshock sequences worldwide. The result is rather robust. It is slightly affected by diversity of the mainshock magnitudes (in the range 5–7); neither is it affected by diversity of earthquake catalogs and magnitude scales. Generally speaking, a decreasing b-value before a mainshock is no surprise for seismologists, though their attitude to this phenomenon is not unambiguous. The problem is not so transparent, because different researchers consider different time–space magnitude ranges. Therefore, the effect is not always observed. It is noted that an earthquake may be attended by increasing background seismicity that does not affect b^- (Asada, 1982). The result may depend on the catalog and on the estimation method. To take an example, according to Comte et al. (1986), $b^- = 0.88$ for 1985, Chile, $M = 7.8$, earthquake, while according to our estimation $b^- = 1.23$ for 10-day foreshocks of this earthquake. Methods of fore- and aftershock identification may also affect the estimates. Most of decreasing b^- examples concern 'preshocks', i.e., events occurring some years or months before a mainshock. Sometimes it is possible to estimate the several-days b^--value using local network data (see above). Our result is substantially different. It is statistical (i.e., it is true in general), and reflects temporal variations of the b^--value for foreshocks with high magnitudes ($M \geq 3$–4) and within extremely restricted time intervals (days–hours).

Acknowledgements

We are grateful to Prof. V.I. Keilis-Borok and to Prof. P. Suhadolc for careful reading of the manuscript and useful comments. The present work was supported by the grant (97-05-65 817) from the Russian Foundation for Basis Research and by the National Science Foundation (Grant EAR-9423818).

Appendix A. Aftershock identification

Molchan and Dmitrieva (1992) noticed the uncertainty of aftershock identification problem: a researcher can arbitrarily regulate the trade-off between the two kinds of errors. Error of the first kind defines the average number of background events erroneously identified as aftershocks, N_B^-, while the error of the second kind accounts for the average number of lost aftershocks, N_A^-. For this reason, the declustering problem should be supplemented by formulating the goals set up by the researcher. For example, the minimax method minimizes the quantity $\max(N_A^-, N_B^-)$, or equalizes and minimizes the two errors N_A^- and N_B^-.

This method is based on the following assumptions.

(1) The aftershock intensity in space and time has the form $\Lambda_A(g,t) = \Lambda_A\, p(g) f(t)$ where $f(t)$ is a normalized Omori law:

$$f(t) = \min\left(1, (t/t_0)^{-1}(1-1/p)/t_0\right), \qquad t > 0,$$

with $p = 1.1$ and $t_0 = 1$ day, and $p(g)$ is the Gaussian distribution of aftershocks in space:

$$p(g) = (2\pi \det R)^{-1} \exp\left(-r^2(g-g_*)/2\right)$$

Here, $r^2(g) = g' R^{-1} g$ is a quadratic form with covariance matrix R, g_* is the centre of scattering. Values of g_* and R are dependent on the sequence.

(2) The intensity of mainshocks (background seismicity) is constant in the vicinity of g_*: $\Lambda_B(g,t) = \Lambda_B$.

The minimax method identifies an event (g,t) as an aftershock if the ratio Λ_A/Λ_B is greater than a threshold, i.e.,

$$1/2\, r^2(g-g_*) + p(\max(0, \ln(t/t_0)) < c, \quad t > 0$$

where $c = -\ln x$ and x is defined as the root of the equation:

$$x = \rho(1 - x^{1/p}), \qquad 0 \leq x \leq 1,$$

with $\rho = 2p(\pi \det R t_0 \Lambda_B)/\Lambda_A$. Obviously, $c \cong \ln(1 + 1/\rho)$ when $p \sim 1$.

The procedure realizing the minimax method consists of the following steps.

(i) *Sampling a mainshock.* At the initial step the largest event in the catalog is declared as a mainshock. Eliminating the largest event together with its fore- and aftershocks from the catalog gets us back to the initial situation.

(ii) *Iterative aftershock identification.* At the first iteration, aftershocks are found roughly, for instance, by a window method. If the number of aftershocks

found at the first stage is not small, $n_A \geq 10$, then we estimate the dispersion parameters g_* and R. On the basis of the estimated parameters, the minimax procedure performs the next iteration of the aftershock sequence. The minimax procedure is repeated iteratively until the parameters $\hat{g}_*$ and $\hat{R}$ become stable.

References

Asada, T. (Ed.), 1982. Earthquake Prediction Techniques. Their Application in Japan, University of Tokyo Press, 250 pp.

Bolshev, L.N., Smirnov, N.V., 1983. Tables for Mathematical Statistics. Nauka, Moscow, 415 pp. (in Russian).

Berg, E., 1968. Relation between earthquake foreshocks, stress and mainshocks. Nature 219, 1141–1143.

Comte, D., Eisenberg, A., Lorca, E., Pardo, M., Ponce, L., Saragoni, R., Singh, S.K., Suarez, G., 1986. The 1985 Central Chile earthquake: a repeat of previous great earthquakes in the region?. Science 233, 449–452.

Gettrust, J.F., Hsu, V., Helsley, C.E., 1981. Pattern of seismicity preceding the Petatlan earthquake of 14 March 1979. Bull. Seismol. Soc. Am. 71, 767–770.

Gusev, A.A., 1991. Intermagnitude relationships and asperity statistics. PAGEOPH 136 (4), 515–527.

Jones, L.M., Wang, D., Xu, S., Fitch, T.J., 1982. The foreshock sequence of the February 4, 1975 Haicheng earthquake (M = 7.3). J. Geophys. Res. 87 (B6), 4575–4584.

Kondorskaya, N.V., Gorbunova, I.V., Kireev, I.A., Lagova, N.A., Storchak, D.A., Khrometskaya, E.A., 1997. Analysis of the unified catalogue of earthquakes of northern Eurasia. J. Earthquake Prediction Res. 6 (1), 51–73.

Mogi, K., 1963. Some discussions on aftershocks, foreshocks and earthquake swarms. Bull. Earthquake Res. Inst. Tokyo Univ. 41, 615–658.

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

Molchan, G., Dmitrieva, O., 1992. Aftershock identification: methods and new approaches. Geophys. J. Int. 109, 501–516.

Ogata, Y., Utsu, T., Katsura, K., 1995. Statistical features of foreshocks in comparison with other earthquake clusters. Geophys. J. Int. 121, 233–254.

Papazachos, B., Delibasis, N., Liapis, N., 1967. The time distribution for foreshocks. G. Annali di Geofizica 20, 22–29.

Prozorov, A.G., 1986. Dynamic algorithm for removing aftershocks from the world earthquake catalog. In: Keilis-Borok, V. (Ed.), Computational Seismology. Nauka, Moscow, Vol. 19 pp. 58–62.

Suyehiro, S., Seiya, H., 1972. Foreshocks and earthquake prediction. Tectonophysics 14, 219–225.

Yoshioka, S., Yabuki, T., Sagiya, T., Tada, T., Matsu'ura, M., 1993. Interplate coupling and relative plate motion in the Tokai district, central Japan, deduced from geodetic data inversion using ABIC. Geophys. J. Int. 113, 607–621.

Bulletin of the Seismological Society of America, Vol. 87, No. 5, pp. 1220–1229, October 1997

Multi-Scale Seismicity Model for Seismic Risk

by George Molchan, Tatiana Kronrod, and Giuliano F. Panza

Abstract For a general use of the frequency-magnitude (FM) relation in seismic risk assessment, we formulate a multi-scale approach that relies on the hypothesis that only the ensemble of events that are geometrically small, compared with the elements of the seismotectonic regionalization, can be described by a log-linear FM relation. It follows that the seismic zonation must be performed at several scales, depending upon the self-similarity conditions of the seismic events and the linearity of the log FM relation, in the magnitude range of interest. The analysis of worldwide seismicity, using the Harvard catalog, where the seismic moment is recorded as the earthquake size, corroborates the idea that a single FM relation is not universally applicable. The multi-scale model of the FM relation is tested in the Italian region.

Introduction

In the last decades, increasing attention has been paid to seismic hazard (Cornell, 1968; Working Group, 1995) and seismic risk assessment (Molchan *et al.*, 1970; Caputo *et al.*, 1974; Keilis-Borok *et al.*, 1984). The generally accepted methodology for risk assessment includes the following interrelated steps: (1) seismic zoning, that is, the identification of potential earthquake source zones; (2) construction of a seismicity model; (3) construction of a spatial model of strong-motion effects as a function of event location and size; (4) risk assessment based on models (1) through (3), that is, estimation of the probability for a given effect (peak ground acceleration at a site, total economic losses, or the number of injured people in an area), to exceed a fixed threshold during a time interval T.

We will deal with the second step, that is, with the construction of a model for the sequence of main events (not aftershocks). The usual description of long-term seismicity (T = 50 to 100 yr) is based on the assumption that, in a given region, the earthquakes follow a random distribution (Poisson hypothesis) and the Gutenberg–Richter (GR) law. More specifically, it is assumed that the numbers $N(\Delta)$ of main events in the elementary cells $\Delta = dg\ dM\ dt$ (g is spatial coordinate, M is magnitude, t is time) are statistically independent and follow the Poisson distribution with mean $\langle N(\Delta)\rangle = n(g,M)\Delta$, where

$$\log n = a - bM, \quad M \in (M_-, M_+) \qquad (1)$$

and each of the parameters (a,b,M_+) has a different space variable dependence. The quantity M_- represents the threshold for damaging earthquakes, while M_+ is treated as the maximum magnitude.

This model is generally considered satisfactory for small and moderate earthquakes, while for the largest events, some authors actually suggest a formal smoothing of relation (1), truncated in the high magnitudes range. For example, the Kulbak principle of maximum entropy leads to the model (Main and Burton, 1984; Kagan, 1991, 1994)

$$\log n = a - bM - 10^{\theta(M-M_0)}, \quad \theta = {}^3/_2, \qquad (2)$$

with parameters a, b, and M_0. In (2), $\log n$ rapidly decreases around M_0, so that it can be treated as the *effective* maximum magnitude (an analog of M_+).

In other approaches, the seismicity rate $n = n(M)$ around M_+ is transformed into one or several peaks that are supposed to describe the rate of characteristic earthquakes (Schwartz and Coppersmith, 1984) or that of their "cascades," that is, multiple segment earthquakes (Working Group, 1995). The time behavior of the characteristic events is modeled as a nonpoissonian renewal process, and, very importantly, the events are related to a whole fault rather than to a point, as it is usually assumed in application of (1).

The question of what model is preferable has been the subject of lively debate; for example, see the discussion in Wesnousky (1994, 1996) and Kagan (1994, 1996). The debate seems to us to reflect the contradiction between two paradigms that recently appeared in seismology. One paradigm, formulated by Bak and Tang (1989), is related to a new insight into the dynamics of the Earth's crust. They treat the seismic process as a dissipative dynamic system that follows the mechanism of Self-Organized Criticality (SOC), and this implies that (1) is valid in a wide range of M, with a possibly universal parameter **b**. The other paradigm is based on observations and is related to the Characteristic Earthquake (CE) concept (Schwartz and Coppersmith, 1984). The CE has, by definition, the largest possible magnitude for the fault considered and a significantly higher rate of occurrence than the one predicted by (1).

A compromise is adopted in the recent, conceptually

important work on seismic hazard for southern California (Working Group, 1995). Each zone is assumed to have randomly distributed earthquakes with a universal **b**-value, **b** = 1, plus characteristic earthquakes on specific faults; additionally, "cascades" of characteristic events can happen as well. Unfortunately, the time dependence of the large earthquakes in this model is described by a large number of parameters.

The leading idea of this article is that in a specific seismogenic zone the SOC paradigm can be applied with some limitations, depending upon the zone itself. Roughly speaking, the (scaling) relation (1) holds for those events whose linear size (l_M) is small compared with the "physical linear size" of the zone (L), that is, $l_M \ll L$. It follows that the description of the recurrence of events, with size $M \in \Delta M$, in a point g, based on the GR law, should be made by finding a zone (containing the point) with the appropriate dimensions and by relating the considered events to that zone rather than to the point. Therefore, depending upon the physical features of the area under study, several levels (scales) of seismic zoning may be needed rather than a single one. Each zoning scale must match the size range of the considered seismic events, ΔM. This scale is due, not to certain features in the spatial distribution of earthquakes of different magnitudes (Woo, 1996), but to event self-similarity conditions within a unit (an elementary area) of the regionalization and to the linear representation of log $n(M)$ in the range ΔM. This idea is illustrated schematically in Figure 1: on a small (detailed) scale (L_1), the area is divided into 10 seismogenic parts where the GR law is satisfied in the range $[M_-, M_1]$, while on a larger scale (L_2), there are three macrozones where the GR law is satisfied in the range $[M_1, M_2]$. The b-value depends on the zoning unit and on the magnitude range only, while the differences in the **b**-value in a point for different zoning scales must indicate a change in the self-similarity conditions for events of different size. This approach largely overcomes the contradictions between the two paradigms, in fact, the earthquakes in the range $[M_1, M_2]$ may turn out to be characteristic events for zones of the first level (L_1), while following the GR law in zones of the next larger level (L_2). However, in the framework of the SOC paradigm, we cannot usually predict the shape of the tail in the magnitude distribution; therefore, the occurrence rate for the largest magnitudes may remain unknown. The log $n(M)$ relation in Figure 1, for the maximum scale, (L_3) or ($M > M_2$), can have an unknown nonlinear shape.

The condition $l_M \ll L$ (possibly in the weaker form $l_M \leqq L$) is not new in seismology (Caputo *et al.*, 1973), and it appears in the recent articles by Pacheco *et al.* (1992), Romanowicz (1992), and Rundle (1993) who examine, for the global seismicity, the departure of log $n(M)$ from a straight line, in the range of large magnitudes. The departure from linearity for spatially unbounded seismogenic zones has been explained by the saturation effect of the transverse dimension of the fault, as a result of the finite thickness of the brittle zone. If one recalls the usual difficulties with cata-

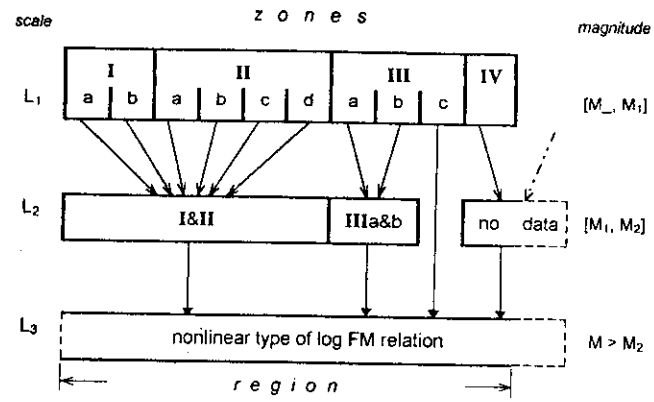

Figure 1. Diagram of the multi-scale representation of the FM relation: log FM is linear in each zone with scale L_1 or L_2 in the range of M: $[M_-, M_1]$ or $[M_1, M_2]$, respectively. **a**- and **b**-values depend on the zone; at the same time, **a** depends on the position g in the zone. Dotted lines indicate the possibility to extend the zone. The zone identification is related to the Italian region (see Figs. 2 and 3).

logs—small amount of homogeneous data and saturation of all magnitude scales—a statistical substantiation of the above effect is not easy and has been disputed by Kagan (1997).

Here (1) we discuss the idea behind the multi-scale model $n(g, M)$; (2) we present statistical arguments against the universality of the parameter **b**, using the Harvard global centroid moment tensor catalog of earthquakes; and (3) we illustrate our multi-scale model on Italy, where a unique 1000-yr historical catalog is available. A full multi-scale analysis of Italian seismicity can be found in Molchan *et al.* (1996).

Multi-Scale Representation of the Frequency-Magnitude Relation

Seismic risk estimation requires that the seismicity within a set of seismogenic zones is modeled in the best possible way (strictly speaking, the term "best" cannot be defined, since the risk problem has many targets and a model shows an integral effect). So far, the commonly accepted tool to deal with the problem is the GR law and the Poisson hypothesis. The latter assumption permits to consider separately individual seismogenic zones. The choice of a zone is influenced by seismotectonic and geological considerations that are used to provide evidence that the zone is homogeneous with respect to a number of parameters, in particular, the **b**-value. Fixing a zone implicitly introduces a characteristic scale (L) related to the spatial structure of the dominant fault system in the zone and to the physical conditions there, L may be determined by one of several quantities that ultimately control the earthquake size in that zone: for example, the fault length, the thickness of the brittle layer, and the

typical distance between rare (compared to the timescale T) events in the zone.

Crustal dynamics is frequently treated as a nonlinear process close to a critical state [e.g., Turcotte (1995) and the references therein]. The idea finds its theoretical support in the Bak *et al.* (1988) model in which, starting from any initial state, a dynamic system with many degrees of freedom will, by itself, attain a critical state—the SOC phenomenon. Usually these processes involve phenomena such as fractality, self-similarity, and power-law relations with "universal" scaling exponents. The GR law belongs to these laws when the size of the earthquakes is expressed in terms of the seismic moment or of the energy. Therefore, in seismology, the SOC paradigm is essentially based on the GR law. However, the appearance of a characteristic scale, L, and the consideration of a limited and fixed zone can violate the self-similarity and the universality of the scaling exponents, at least for $l_M \geqq L$.

Since a satisfactory mathematical modeling for the lithosphere dynamics is missing—we are still in the pre-equation era—we can support our previous considerations only using some analogies, taken from other fields of science.

In a deterministic framework, narrowing the observation of a homogeneous fractal set to a restricted volume S, we can merely get an accidental idea of the statistics of large clusters, since the characteristic cluster (comparable in size to the volumes) can be found in- or outside of the observation volume.

In a probabilistic framework (fractal random set), there is a difference between conditional and unconditional distributions (see Palm measures in the theory of point processes). Namely, the conditional distribution of a cluster K, under the condition that K belong to a fixed area or that the convex hull of K contains a fixed point g_0, and the unconditional one are different. For example, when the time interval between two subsequent Poissonian events contains a point fixed beforehand, the mean length of this interval is twice the mean length of the unconditional case [a well-known paradox for Poissonian processes (Feller, 1966)].

In physics, the Kolmogorov theory of well-developed turbulence (Landau and Lifshitz, 1959) is based on the hypothesis of self-similarity for turbulent pulsations and successive transmission of energy from larger pulsations to smaller ones, and it defines the so-called inertial range of scales, r, in which turbulence scaling is assumed; namely,

$$\mathrm{Re}^{-3/4} \ll r/L \ll 1, \tag{3}$$

where L is the external scale, $\mathrm{Re} = LV/\nu$ is the Reynolds number (V is the characteristic velocity and ν the molecular viscosity).

Consequently, keeping to the standpoint of nonlinear dynamics and treating an earthquake as a spatial object rather than a point, one may assume that the GR law is valid for the earthquakes with linear dimensions, l_M, much smaller than the characteristic scale, L, of the seismic zone considered (hypothesis **A**).

Risk analysis is concerned with damaging ($M \geqq 3.5$) and therefore relatively large earthquakes, thus hypothesis **A** is the analog of the right-hand side of (3). On the other end, a seismological analog of the left-hand side of (3) is of theoretical interest, as described by Aki (1987), who shows a significant departure of log $n(M)$ from a straight line for $M < 1.5$.

In a seismic zone, several characteristic scales can be identified. Caputo *et al.* (1973) distinguish three critical magnitudes ($M_1 < M_2 < M_3$) that can be observed in the statistical properties of earthquake occurrence:

- up to M_1, the source area of an event is small compared with the geometrical dimensions of the main tectonic faults in the zone;
- for $M > M_2$, a rupture penetrates the entire crust, so that the earthquake size can only be increased by increasing the source length, l (the earthquake source has lost one degree of freedom because of the saturation with depth); and
- for $M > M_3$, an earthquake can occur within a single isolated zone only by simultaneous slippage on several faults.

Pacheco *et al.* (1992) and Okal and Romanowicz (1994) give estimates for the saturation of the earthquake size with depth: $M_2 = 6.0$ for mid-ocean ridges (MOR) where the down-dip width of the seismic source zone, w, varies from 10 to 15 km, and $M_2 = 7.5$ for shallow earthquakes in subduction zones (S) where w is about 60 km.

The above-mentioned critical magnitudes indicate the existence of different conditions for the self-similarity of source zones. This may affect the scaling laws, that is, the **b**-values. For this reason, if a zone has several characteristic scales, one would expect the log frequency-magnitude (FM) relation to be piecewise linear, and then the parameters $M_\pm$ in (1) control the size range of the events, ΔM, for which the self-similarity conditions are fulfilled. For instance, if $l \gg w$, then two ranges of linearity of log FM are possible: (M_-, M_1) and (M_2, M_3). In practice, the interval between the two ranges may degenerate to a point, in order to fit the log FM relation when few data are available. Starting from some magnitude, say M_3, the self-similarity conditions are no longer valid for a single seismic zone, and then relation (1) can break down for large M because of the hypothesis **A**.

The idea of the characteristic earthquake (Schwartz and Coppersmith, 1984) is an important attempt to forecast the form of the FM relation for a fault segment when the linear relation (1) is no longer applicable. This idea, as advocated in its orthodox form (Wesnousky, 1994), runs into serious difficulties (Working Group, 1995; Kagan, 1996). On the other hand, the alternative solution (2), based on a formal device, borrowed from information theory, rather than on the physics of the phenomenon, involves an arbitrary choice of a function ψ of the energy E, $\psi(E)$, or of the magnitude

M, $\psi = \psi(10^{\theta M})$ with $\theta = 3/2$. If we fix the mean value of ψ, the principle of maximum entropy in combination with the GR law leads to a new form of the FM relation:

$$\log n(M) = a - bM - \lambda\psi(10^{\theta M}), \qquad (4)$$

where a, b, and λ are parameters. If one takes into account the relation of the earthquake energy with the size of the rupture zone, the case $\psi(E) = E^{\rho}$ ($\rho = \frac{2}{3}$) can be considered as well, and since the statistical estimation of the exponent ρ has a very weak resolution (Kagan, 1991), many other models like (4) can be claimed to fit observed data.

As a rule, to predict the frequency of earthquakes using a linear log FM relation, the larger the event, the greater must be the geometrical dimension of the zone, so that it is of the appropriate scale level for (1) to be valid in the ΔM of interest. However, the zone-broadening process has a limit, since physical factors will interfere with the self-similarity conditions for large events (e.g., Caputo *et al.*, 1973). A hierarchical analysis is then reasonable in which the seismicity is described by several GR laws for several scales and magnitude ranges ΔM (see Fig. 1 and the last section). In this hierarchical analysis, the smaller events may be less useful for the prediction of the occurrence of the larger events.

There are serious obstacles to the use of conventional statistical techniques in the estimation of the maximum magnitude. The statistical technique (Pisarenko *et al.*, 1996) is based on the parametrization of $n(M) = n(g, M)$ for large M, but the SOC paradigm does not permit to predict the recurrence of very large events in the same zone used to predict the smaller events. The multi-scale representation of the FM relation gives, in the best case, a piecewise linear representation of $\log n(g, M)$, which is not universal and is dependent on point g.

Using theoretical arguments, we can infer that the variability of the estimated b-value should not necessarily be explained by appealing to criticisms of the magnitude scales involved and to the poor quality of the data (Kagan, 1994). There is merely a drawback in the current interpretation of the **b**-value determined considering restricted areas. The parameter **b** is representative only for a definite scale range. It is therefore necessary to show that the parameter **b** is not universal also from a statistical point of view.

Variation of the b-value: Global Seismicity

Kagan (1994), from the study of global seismicity, assumes that the **b**-value is universal: **b** = 1 for all events and **b** = 0.75 for mainshocks (excluding aftershocks). Any variation of the **b**-value is treated by Kagan (1994) as an artifact due to the small size of the samples considered and to the fact that the magnitudes used in regional studies are inhomogeneous and have not a clear physical meaning. On the other end, using the same data analyzed by Kagan (1997), we provide here several examples of statistical comparison in major seismic provinces that indicate that the b-value is not universal.

Data. For the purposes of seismic risk determination, it is natural to consider shallow earthquakes (with focal depth less than 70 km), and a homogeneous catalog, reporting a physically meaningful measurement of the earthquake size, is needed. The available possibilities are rather limited. There is a short global centroid moment tensor (CMT) catalog (CMTS, 1995) that reports the scalar seismic moment M_0 (dyne-cm) or the moment magnitude $M_w = \frac{2}{3}(\log M_0 - 16.1)$. As of 30 April 1995, the catalog contains 12,417 events, with depth $H \leqq 70$ km, and is, in our estimate, complete for $M_w \geqq 5.75$, $\geqq 5.55$, $\geqq 5.45$ starting from 1977, 1982, and 1987, respectively.

The absence of smaller events in this catalog does not permit the use of refined techniques of aftershock identification (Molchan and Dmitrieva, 1992); therefore, the aftershocks have been identified by the window method. The spatial radius, R, and the time duration of the aftershocks sequence, T, are as follows (Molchan and Dmitrieva, 1992):

M_w	5.5–6.5	6.5–7.0	7.0–7.5	7.5–8.0	≧8.0
R (km)	50	60	70	100	200
T (years)	1	2	2	2	2

Technique. The elimination of aftershocks lends more credence to the Poisson hypothesis in the estimation and comparison of the parameters **a** and **b** in the GR law. The solution of this problem for arbitrary grouping of the data, over magnitude and time, is given in Molchan and Podgaetskaya (1973) and summarized in the Appendix, where the hypothesis H_b of equality of the **b**-value in several samples that obey (1) is tested using the generalized Pearson test, π. The probability ε of excedence of the observed value π_{obs} under the hypothesis H_b gives the significance level of H_b. The hypothesis H_b is doubtful, when ε is small (i.e., $\varepsilon \leqq 5\%$).

Examples. In the time–magnitude intervals in which it is complete, the CMT catalog contains 6776 events, of which 4832 are mainshocks (71%). With this amount of data, we can analyze credibly the **b**-value for the major seismotectonic features only. In this case, we have $l \gg w$, so that the critical characteristic scale is the down-dip width of the fault zone, i.e., $L = w$ for small events and $L = l$ for the large ones. We begin with a well-known example.

1. *Subduction zones and mid-ocean ridges.* According to Okal and Romanowicz (1994), widely different values of w characterize subduction (S) and MOR zones: 60 and 10 km, respectively. The earthquakes with M_w between 5.8 and 6.5 are "small" (there is no saturation along the down-dip width of the zone) for the S zones, and "large" for the MOR zones (there is saturation). The differences in self-similarity conditions for the two source zones do affect the **b**-value.

Table 1 (row 1) contains the **b**-value estimates based on all events (b_+) and on the mainshocks (**b**) for S and MOR zones. The differences in the **b**-value are so large that neither any estimation techniques nor various methods of aftershock identification can remove the effect. The difference is expressed quantitatively by the significance, ε, of the hypothesis H_b.

2. *Subduction zones: two magnitude ranges.* Events with $M_w \geqq 7.5$ are large for an S zone (Pacheco *et al.*, 1992); therefore, the equality of the **b**-value in the ranges $M_w \leqq 7.5$ and $M_w \geqq 7.5$ is doubtful. The conclusion is corroborated statistically in Pacheco *et al.* (1992), who used a combined worldwide catalog for the period 1900 to 1989. The CMT data also indicate a significant change in **b** for $M_w \geqq 5.55$ [see Table 1 (row 2)].

The significant difference in the **b**-value between the S and MOR zones is not exceptional; in fact, we show that both zones are internally inhomogeneous with respect to the **b**-value. In the next two examples, we identify subzones by employing strictly seismotectonic arguments without any preliminary data analysis.

3. *Mid-ocean ridges: two subzones.* The MOR zones are segments of rift zones that are cut by transform faults and dominated by pure strike-slip movement. There are two seismogenic transform faults in the Mid-Atlantic Ridge (MAR) that are abnormal for their linear size ($L \approx 2000$ km in both cases): the Azores-Gibraltar (AG) ridge and the Equatorial (E) fault. We compare the **b**-value for the union (Σ) of the transform faults in AG and E and for its complement (MAR\Σ) in the Mid-Atlantic Ridge zone. Table 1 (row 3) shows that the **b**-values are different at the significance level $\varepsilon \approx 5\%$: $\mathbf{b} \approx 1$ for AG and E and $\mathbf{b} \approx 1.3$ for its complement in the MAR zone. $\varepsilon < 5\%$ if we compare the Σ zone ($\mathbf{b} = 0.97$) with its complement in the MOR zone ($\mathbf{b} = 1.25$), where the number of data is 547 instead of 107.

4. *Island arcs: two subzones.* Following Kronrod (1985), we consider the island arcs in the Northwest Pacific, from Alaska to Taiwan, divided into two sets by their tectonic characteristics: volcanic arcs (V): Aleutians-Commander Is., Kuril Is., Ryu-Kyu, Izu-Bonin, Marianas; geosynclinal arcs (GS): Alaska, Calgary coast, Kamchatka, Japan, and Taiwan. The CMT catalog corroborates an earlier inference by Kronrod (1985), based on the pre-1975 worldwide catalog data, about the existence of a significant difference in the **b**-value for these subduction zones [see Table 1 (row 4)].

5. *Subduction zone: three ranges of depth.* The distribution of centroid depths, H_c, in the CMT catalog has two distinct peaks at 10 and 15 km, and a fuzzier one at 33 km. Because of the difficulties inherent in the determination of H_c, the values $H_c = 10$ km or $H_c = 15$ km are just markers (used during different time periods) of shallowness. For this reason, we consider the following division of the H_c scale: up to 15 km, from 16 km to 33 km, and from 34 km to 70 km. This grouping divides all data into three roughly equal parts. In view of the effect of saturation along the down-dip

Table 1
Global Seismicity: **b**-Value Comparison

	Zone		Magnitude Range, M_w	All Events		Mainshocks		
				N	b_+	N	$\mathbf{b} \pm \Delta b$*	ε†
1	Subduction zone, S‡		≧5.88	1761	0.98	1233	0.88 ± 0.05	
								<0.05%
	Mid-ocean ridges, MOR‡		≧5.88	313	1.49	298	1.47 ± 0.16	
2			5.55–7.56	3012	0.95	1927	0.80 ± 0.04	
	Subduction zone, S‡							<4.9%
			7.57–8.90	36	1.72	32	1.50 ± 0.70	
	Mid-Atlantic							
3	Ridge,	Σ = AG & E¶	5.45–8.00	71	1.0	61	0.97 ± 0.25	
								<5.4%
	MAR§	MAR\Σ	5.45–8.00	111	1.38	107	1.30 ± 0.22	
4	Island	V	5.45–8.90	519	1.08	350	0.97 ± 0.10	
								<0.5%
	arcs‖	GS	5.45–8.90	329	0.89	212	0.75 ± 0.10	
5	Subduction	$H_c \leqq 15$	≧5.45	965	1.05	639	0.93 ± 0.10	
		$16 \leqq H_c \leqq 33$	≧5.45	792	0.80	487	0.63 ± 0.10	<0.05%
	zone‖	$34 \leqq H_c \leqq 70$	≧5.45	689	0.92	459	0.83 ± 0.11	

*$(\mathbf{b} - \Delta b, \mathbf{b} + \Delta b)$ is the 95% confidence interval for the **b** value. The estimate b_+ is not supplied with a confidence interval because the data are correlated.

†Significance level of the hypothesis H_b: all **b** are equal.

‡S and MOR zones, as in Kagan (1997), include the following Flinn–Engdahl seismic regions (Young *et al.*, 1996): *S* (1, 5–8, 12–16, 18–24, 46); MOR (4, 32, 33, 40, 43–45).

§MAR is zone 32 in Flinn–Engdahl seismic regions (Young *et al.*, 1996).

¶Geographical limits of the zones. AG = (35.6° N, 40.0° N) × (−60° W, −29.7° W), E = (2.1° N, −3.4° E) × (−12.0° W, −31.4° W).

‖See text for the subzone symbols.

width of the zone, we eliminate large (for an S zone) events, that is, events with $M_w > 6.5$, and compare the **b**-values for the three ranges of depth. Table 1 (row 5) shows that the confidence level for the hypothesis of equal **b**-values in the three ranges of H_c is extremely low, $\varepsilon < 0.05\%$. The difference is mainly due to the lower **b**-value obtained for $H_c \in$ [16 to 33] km. This fact is difficult to interpret from a physical point of view, because the depths of shallow earthquakes in the CMT catalog are occasionally incorrect; however, the assumption of a universal **b**-value permits any formal grouping of the data. Therefore, these results can be viewed as another confirmation that in the S zone the **b**-value is not constant.

These five examples show statistically significant variations in the **b**-value for $M_w \geqq 5.55$. The actual **b**-values can depend on the magnitude type used; however, these variations are generally consistent with similar conclusions reached using other magnitude scales. Our results contradict the statement by Kagan (1997) that "there is no statistically significant variation of the **b**-value for all seismic regions except for the mid-ocean ridge systems." The Kagan's conclusion can be explained by the fact that the choice of the CMT catalog dramatically lowers the resolution capability of the tests. Therefore, we use here a technique of great flexibility (see Appendix), and we consider an expanded data set—the starting date of the catalog considered depends on magnitude, and the final date is 8 months later than in Kagan (1997).

FM Relation for Italy

To illustrate the multi-scale representation of the FM relation, we chose Italy as a test region, since there is available a unique historical mainshock catalog that covers a time interval of about 1000 yr (Stucchi *et al.*, 1993) and that is particularly suitable for the analysis of the recurrence of the large events ($M > 5$). The other difficulties inherent with regional earthquake catalog remain; thus, while having a gain in the time span and energy range, compared with the CMT data, we lose in quality, since the earthquake size is usually represented by different magnitudes of different accuracy and representativity. This puts stringent requirements on the data-processing techniques and may hamper the interpretation of any result.

For the period 1900 to 1995, we use a catalog that we have labeled the *Current Catalog of Italy* (CCI, 1994), where the dominant magnitudes are local, M_L, duration, M_D, and macroseismic, M_I. The magnitude used in the analysis is the first available in the sequence: M_L, M_D, M_I. Magnitude $M = M_L$ is grouped in intervals of 0.2 to 0.3, while magnitudes $M \neq M_L$ are grouped in intervals of 0.5. The time boundaries of complete reporting for M are adopted depending upon the location, the value, and type of M. The aftershocks are identified using a flexible minimax approach developed by Molchan and Dmitrieva (1992). We use aftershock areas (of confidence level 95%) to identify different, interacting seismogenic zones, and the technique for the estimation and comparison of **a** and **b** in the GR law is described in the Appendix. The comparison of the **b**-value determined in several different areas is employed as an extra argument in favor of zone broadening or narrowing.

The value $M_I = 7.3$ defines the maximum magnitude observed in Italy during the last 1000 yr. For this reason, the representation of the FM relation for risk purposes is relevant for shallow events with $M \in$ [3.5 to 7.5]. Therefore, a non-trivial FM representation can involve no more than two or three scale levels, and a possible multi-scale model of this type is presented formally in Figure 1. In what follows, we limit our attention to space scale l, rather than to physical scale L, since the available data do not allow its definition.

The largest space scale (l_3) is controlled by plate tectonics and by the size of the area under study. According to Lort (1971), the Adriatic region is regarded as an African promontory at the plate boundary between Eurasia and Africa. The boundary is well marked in Figure 2b by the largest earthquakes ($M > 6.3$), and it justifies to keep the region as a whole, with $l_3 \approx 1500$ km.

Since the data ($M > 6.3$) in Figure 2b span about 1000 yr, and moderate events are well dispersed over the entire region (Fig. 2a), the alignment of epicenters (at least for central and southern Italy) cannot be casual. Hence, the kernel standardized technique recently suggested by Woo (1996) to smooth maps of earthquake activity $\mathbf{a}(g, M)$ calls for careful handling.

The intermediate space scale (l_2 = 400 to 500 km) is in part controlled by the geometry of the plate boundary. From a tectonic point of view, Italy is conventionally divided into four zones (Boriani *et al.*, 1989): (I) Alpine compression zone, (II) Northern Apennine Arc, (III) Calabrian arc, and (IV) Sicily with a possible continuation toward Tunisia. The **b**-zones of space scale level l_2 are represented in Figure 2b where **b**-zones I and II are lumped in a single one, because the **b**-values for the larger events are similar. From now on, **a**-zone and **b**-zone will indicate zones with a postulated constant value of **a** or **b**, respectively.

The lowest space scale (l_1 = 200 to 300 km) for zones having constant **b**-values is determined by hypothesis **A**: $l_M \ll l_1$. We use the aftershock zone linear size (L_{aft}) to estimate l_M. According to our analysis, for $M \in$ [4 to 6], the typical values of L_{aft} are 20 to 60 km, while for $M \in$ [5 to 7], a few observations give L_{aft} = 100 to 140 km; for $M \geqq 3.5$, a further splitting of the **b**-zones generally leads to a poorer resolution of the **b**-value. Some seismotectonic and statistical considerations have led us to define 10 **b**-zones of level 1 (Fig. 3), composed by elements of the seismotectonic regionalization of Italy, recently developed by GNDT (1992). Each element of the GNDT zoning includes seismogenic structures of definite kinematics type (see Fig. 3) and has typical dimensions of 40 to 130 km by 20 to 30 km, comparable with L_{aft}. Most GNDT zones contain a small number of instrumental data with $M \geqq 3.5$; it is therefore impossible

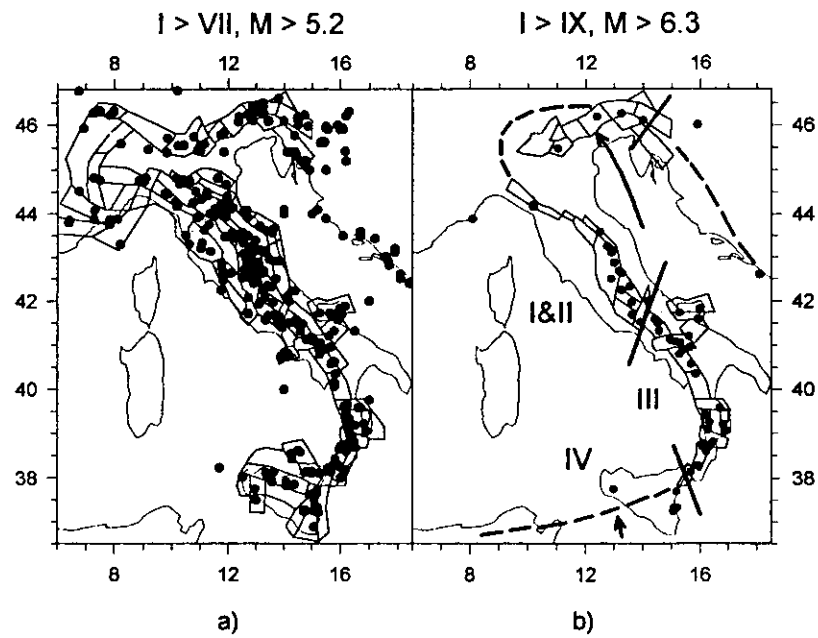

Figure 2. Space distribution of mainshocks (solid circles) for the period 1000 to 1980 according to Stucchi *et al.* (1993) and seismogenic zones after GNDT (1992). Right panel: solid segments mark the boundaries between **b**-zones of scale l_2; the indicated GNDT zones forms the hypothetical seismogenic area for events with $M > 6.3$; dashed line is the sketch of the plate boundary between Eurasia and Africa in the Italian region (Lort, 1971); arrows show the direction of motion relative to Eurasia.

to subdivide them, when they are used as **a**-zones of scale level 1.

Table 2 presents the estimated **b**-values for two levels: level 1 is appropriate for $M \in [3.5 \text{ to } 5]$, while level 2, for $M \in [5 \text{ to } 7]$. The historical data are used for the **b**-value estimates of level 2 only. The data indicate that the magnitude range [5 to 6] can be considered as an intermediate one. Although the events with $M \in [5 \text{ to } 6]$ essentially control the **b**-value in the entire range [5 to 7] (see Appendix), we have statistically significant difference in the **b**-value for the two scale levels. The physical and statistical nature of the variation in the **b**-value, reported in Table 2, calls for a special study.

Conclusion

This article is an attempt to derive from the SOC paradigm corollaries relevant to seismic risk problems (e.g., Main, 1995; Woo, 1996). Our conclusions are these:

- Given a seismic zone, the conventional description of seismicity puts conditions on the scale and the magnitude range to be considered when representing the FM relation in log linear form. Hypothesis **A** leads to a multi-scale representation of the FM relation preserving the log linear character. In this case, the query of Wesnousky (1994) "The Gutenberg–Richter or characteristic earthquake distribution, which is it?" could be answered: both. Large events can themselves form a statistical population having a GR law in a zone with the appropriate scale.
- The adoption of hypothesis **A** may reduce the number of parameters needed to describe the recurrence of the larger events, but, at the same time, hypothesis **A** reduces our statistical ability to estimate M_{max}.
- Using the CMT catalog, a significant worldwide variation in the **b**-value has been found for $M_w > 5.5$, and this fact justifies the search for geometrical/physical factors that cause the regional variations in **b**.
- The analysis of the seismicity in Italy shows that the multi-scale approach can be used where the size range of damaging earthquakes is large ($\Delta M > 4$) and catalogs with a large amount of historical data are available.
- The multi-scale approach calls for an understanding of seismicity at different space–time scales.

Appendix

Seismic risk studies involve different groupings over magnitude, depending on space and time. This considerably complicates the statistical estimation and comparison of the parameters in the GR law, for a set of nonintersecting volumes in location-magnitude-time space. These problems are considered in the nearly inaccessible article by Molchan and Podgaetskaya (1973), which we summarize below. The technique is essential even for earthquake catalogs containing the scalar seismic moment, given in the form of an exponent and a two-digit prefactor, leading to nonuniform grouping over M_w.

Estimation of (a, b). Let $\{N_i\}$ be the sample of mainshocks in an area G in nonintersecting time–magnitude cells $\Delta T_i \Delta M_i$. Taking into consideration the GR law, we assume that N_i are independent and that they follow the Poisson distribution with parameter Λ_i (mean value of N_i); that is,

$$\Lambda_i = \Delta T_i \int_{\Delta M_i} 10^{a+bM} dM.$$

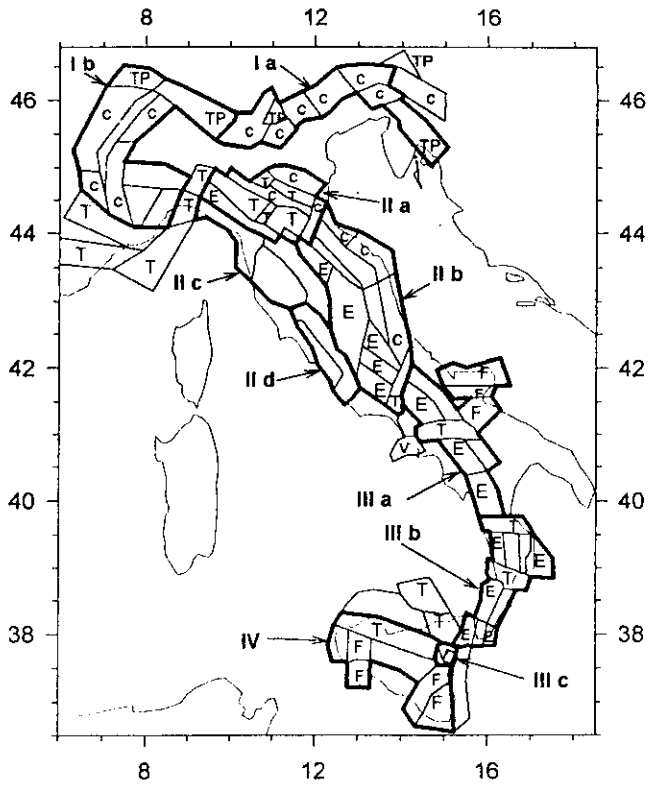

Figure 3. Seismogenic zones (solid line) from GNDT (1992), and **b**-zones of scale l_1 (bold line). Notation for GNDT zones: C, compressional areas; F, areas of fracture in the foreland zones; T, transition areas; TP, areas of transpression; V, volcanic areas. Notation for **b**-zones of scale l_2: Ia, southern branch of Eastern Alps; Ib, Western Alps; IIa, Northern Apennines; IIb, Central Apennines and Ancona zone; IIc, Tuscany; IId, Roma comagmatic zone; III, Calabrian arc [(a) northern branch and Gargano zone, (b) Centre, (c) Etna]; IV, Sicily.

Table 2

FM Multi-scale Model for Italy (the **b**-value of each zone depends on space scale level and magnitude range)

Zones of Scale Level I (Fig. 3)	T*: 1900–1993 $3.5 \leqq M < 5.0$ N*	**b** ± Δb†	Zones of Scale Level II (Fig. 2b)	T: 1000–1993 $5 \leqq M \leqq 7$ N	**b** ± Δb
Ia,b; IIa,b,d‡	1491	0.89 ± 0.15			
			I&II	184	1.07 ± 0.13
IIc	169	1.32 ± 0.24			
IIIa	224	0.65 ± 0.16			
			IIIa,b	59	≤0.65§
IIIb	178	1.00 ± 0.20			
IIIc¶	189	1.02 ± 0.20		log FM is not linear	
IV	151	0.76 ± 0.20	IV&?‖	data are not complete	

*T is the time period covered by the catalog; N is the total number of used main events.

†($\mathbf{b} - \Delta b$, $\mathbf{b} + \Delta b$) is 95% confidence interval for the **b**-value.

‡All five zones are uniform with respect to the **b**-value; therefore, we show only the **b**-value estimate for one of them (Ia, $N = 275$) plus the total number of events in their union.

§We give a qualitative **b**-value estimate because of lack of data.

¶Etna volcanic zone: observed $M_{max} = 5.2$. It is an interesting example of a small **b**-zone ($l = 45$ km) that contains the 95%-aftershocks area related to the mainshocks with $M < M_{max}$.

‖Sicily zone must apparently be extended beyond the boundary of the studied area, toward North Africa.

The estimate of (**a**, **b**) is given by the point where the log likelihood of the data, $\mathcal{L}(G)$, reaches the maximum. Here

$$\mathcal{L}(G) = \sum_{i=1}^{k} \mathbf{l}(N_i \mid \Lambda_i),$$

where $\mathbf{l}(n \mid \Lambda) = n \ln\Lambda - \Lambda$.

The conditional log likelihood for N_i given the statistics $N = \Sigma N_i$, namely,

$$\mathcal{L}(G|N) = \mathcal{L}(G) - \mathbf{l}(N \mid \sum \Lambda_i),$$

is a function of **b** only. The statistics $\mathcal{L}(G|N)$ is approximately gaussian. This permits one to define a more exact distribution of $\mathcal{L}(G|N)$ using the Edgeworth expansion and six moments of $\mathcal{L}(G|N)$. Knowing the distribution of $\mathcal{L}(G|N)$, one can find a confidence interval for **b** (Cox and Hinkley, 1974).

Explicit formulas for **b** exist only in the following theoretical situation: $\Delta T_i \equiv T$, $\Delta M_i \equiv \Delta$, $M_+ = \infty$. The formulas are

$$\hat{b} = \begin{cases} \Delta^{-1} lg[1 + N / \sum N_i(i-1)], & \Delta > 0 \quad \text{(Kulldorf, 1961)} \\ N lge / \sum (M_i - M_-), & \Delta = 0 \quad \text{(Aki, 1954)} \end{cases} \tag{A1}$$

The distribution of Aki (1954) estimator is known exactly. Namely, $\hat{\mathbf{b}}/\mathbf{b} = 2N/\chi^2_{2N}$, where the random variable χ^2_f has the χ^2 distribution with f degrees of freedom. It follows that $\hat{\mathbf{b}}$ has the bias $\Delta\mathbf{b} = \langle\hat{\mathbf{b}} - \mathbf{b}\rangle \approx \mathbf{b}/N$ and the standard deviation $\sigma_b = \mathbf{b}/\sqrt{N}$. The modified estimator $\hat{\mathbf{b}}^* = (1 - 1/N)\hat{\mathbf{b}}$ reduces the variance σ_b^2 and removes the bias.

The explicit formulas (A1) show that the contribution of the statistics of M_i in **b** estimates is proportional to $N_i M_i$. Since N_i is proportional to 10^{-bM_i}, the maximum likelihood (MLH) estimate of **b**, based on uniformly sampled data, depend rather weakly on the large events. Consequently, when the zone size and the magnitude range involved in the statistical estimation of a **b** are mismatched, the MLH estimate of **b** may represent the log FM relation correctly only among the smaller events of the range (events of different magnitudes are here assumed to be completely reported for the same time span).

Comparison of the (a, b) Parameters. Let us consider d nonintersecting volumes $V_a = (G_a T_a M_a)$ with their grouped

data (G_a is a spatial area, T_a a time period, M_a a magnitude range). In order to compare $\mathbf{b}_a$ or $(\mathbf{a}, \mathbf{b})_a$ for different areas G_a, one usually has to test one of the following two hypotheses: all the **b**-values for the volumes V_a are equal, whereas the **a**-values may be arbitrary (hypothesis H_b: $\mathbf{b}_1 = \ldots = \mathbf{b}_d$), or all $(\mathbf{a}, \mathbf{b})$ pairs for volumes V_a are identical (hypothesis $H_{a,b}$: $\mathbf{b}_1 = \ldots = \mathbf{b}_d$, $\mathbf{a}_1 = \ldots = \mathbf{a}_d$). The general alternative to the hypotheses H_b and $H_{a,b}$ is H: $(\mathbf{a}, \mathbf{b})_\alpha$, $\alpha = 1, \ldots, d$ are arbitrary; that is, there are no restrictions on **a** and **b**. The hypothesis H_i, $i = \mathbf{b}$ or $i = (\mathbf{a}, \mathbf{b})$, against H is tested using the generalized Pearson statistics

$$\pi_i = -2\left[\max_{H_i} \mathcal{L}_\Sigma - \max_{H} \mathcal{L}_\Sigma\right],$$

where $\mathcal{L}_\Sigma = \sum_{a=1}^{d} \mathcal{L}(G_a)$ and the maximum is taken over the parameters $(\mathbf{a}, \mathbf{b})_a$ with due account of the relevant hypothesis H_i and H. The values of $(\mathbf{a}, \mathbf{b})_a$ that provide max $\mathcal{L}_\Sigma$ under the hypothesis H_i are maximum likelihood estimates in the general case.

The asymptotic theory of statistical hypothesis testing asserts (Cox and Hinkley, 1974) that H_i should be rejected in favor of H, if $\pi_i > u(f_i, \gamma)$. Here $u(f, \gamma)$ is the quantile of the χ^2_f-distribution of level γ, $f_b = d - 1$, $f_{a,b} = 2(d - 1)$. In this case, the probability of rejecting H_i when it is in fact true is approximately equal to $\varepsilon = 1 - \gamma$.

When the Aki (1954) estimators are valid, a test of the hypothesis H_b for the two regions G_1 and G_2 can be based on the exact distribution of $\hat{\mathbf{b}}_1/\hat{\mathbf{b}}_2$; namely,

$$\hat{\mathbf{b}}_1/\hat{\mathbf{b}}_2 = (\mathbf{b}_1/\mathbf{b}_2) \cdot (N_1/N_2) \cdot (\chi^2_{2N_2}/\chi^2_{2N_1}),$$

where the χ^2 variables are independent and N_i is the total number of events in region G_i. Therefore, if $\mathbf{b}_1 = \mathbf{b}_2$, the ratio $\hat{\mathbf{b}}_1/\hat{\mathbf{b}}_2$ for the two regions follows the F-distribution (Utsu, 1971).

Acknowledgments

We are grateful to P. Scandone for very useful discussions about the seismic zonation of Italy, to our reviewers R. Console and A. McGarr for constructive criticism, to V. Ez, I. Rotwain, and V. Kossobokov for careful reading of the manuscript and useful comments, and to our colleagues F. Vaccari, O. Dmitrieva, and A. Nekrasova for various services rendered. This research was made possible by Contracts CNR 93.02492.PF54, 94.00193.CT05, and 94.01703.PF54, and MURST (40% and 60%), and funds by NSF (Grant EAR-9423818) and INTAS (Grant 94-232).

References

Aki, K. (1954). Maximum likelihood estimate of b in the formula logN = a − bM and its confidence limits, *Bull. Earthquake Res. Inst. Tokyo Univ.* **43**, 237–239.

Aki, K. (1987). Magnitude-frequency relation for small earthquakes: a clue to the origin of f_{max} of large earthquakes, *J. Geophys. Res.* **92**, 1349–1355.

Bak, P. and C. Tang (1989). Earthquake as a self-organized critical phenomenon, *J. Geophys. Res.* **94**, 15635–15637.

Bak, P., C. Tang, and K. Wiesenfeld (1988). Self-organized criticality, *Phys. Rev. A* **38**, 364–374.

Boriani, A., M. Bonafede, G. B. Piccardo, and G. B. Vai (Editors) (1989). *The Lithosphere in Italy*, Academia Nazionale dei Lincei, Roma.

Caputo, M., V. I. Keilis-Borok, T. L. Kronrod, G. M. Molchan, G. F. Panza, A. Piva, V. M. Podgaetskaja, and D. Postpishl (1973). Models of earthquake occurrence and isoseismals in Italy, *Ann. Geofis.* **26**, nos. 2-3, 421–444.

Caputo, M., V. I. Keilis-Borok, G. M. Molchan, T. L. Kronrod, D. Panza, A. Piva, V. M. Podgaetskaja, and D. Postpishl (1974). The estimation of seismic risk for Central Italy, *Ann. Geofis.* **27**, nos. 1–2, 349–365.

CCI (1994) Current Catalog of Italy PFG-ING, Open data file PFGING.DAT, size = 840620, date 94.03.19.11:16.

CMTS (1995). *The Harvard Centroid Moment Tensor Solutions Catalog*, 1 Jan 1977–30 Apr 1995, Harvard University, Dept. Earth and Planet. Sci., Open Data File.

Cornell C. A. (1968). Engineering seismic risk analysis, *Bull. Seism. Soc. Am.* **58**, 1583–1609.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*, Chapman & Hall, London, 550 pp.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York.

GNDT (1992). Gruppo Nazionale per la Difesa dai terremoti, GL Seismotettonica (zone sis-mogenetiche), Modello seismotettonico del territorio italiano, versione aggiornata al Maggion 1992.

Kagan, Y. Y. (1991). Seismic moment distribution, *Geophys. J. Int.* **106**, 123–134.

Kagan, Y. Y. (1994). Observational evidence for earthquakes as a nonlinear dynamic process, *Physica D* **77**, 160–192.

Kagan, Y. Y. (1996). Comment on "The Gutenberg-Richter or characteristic earthquake distribution, Which is it? by S. G. Wesnousky, *Bull. Seism. Soc. Am.* **86**, 274–285.

Kagan, Y. Y. (1997). Seismic moment-frequency relation for shallow earthquakes: regional comparison, *J. Geophys. Res.* **102**, 2835–2852.

Keilis-Borok, V. I., G. M. Molchan, A. Ch. Koridze, T. L. Kronrod, and O. D. Gotsadze (1984) An insurance-oriented pilot estimation of seismic risk for rural dwellings in Georgia, *Geneva Association, Etudes et Dossiers*, **77**, 85–109. Full version (in Russian): *Comput. Seism.* **17**, 85–109. Full version (in Russian): *Comput. Seism* **17**, 58–67, Moscow, Nauka.

Kronrod, T. L. (1985). Parameters of seismicity in tectonically similar regions. *Comput. Seism.* **18**, 154–164, Moscow, Nauka.

Kulldorf, G. (1961). Contributions to the theory of estimation from grouped and partially grouped samples. Almqvist & Wiksell, Stockholm-Göteborg-Uppsala, 144 p.

Landau, L. D. and E. M. Lifshitz (1959). *Fluid Mechanics*, Pergamon, New York.

Lort, J. M. (1971). The tectonics of the Eastern Mediterranean, *Rev. Geophys.* **9**, 189–216.

Main, I. G. (1995). Earthquakes as critical phenomena: implications for probabilistic seismic hazard analysis, *Bull. Seism. Soc. Am.* **85**, 1299–1308.

Main, I. G. and P. W. Burton (1984). Information theory and the earthquake frequency-magnitude distribution, *Bull. Seism. Soc. Am.* **74**, 1409–1426.

Molchan, G. M. and O. E. Dmitrieva (1992). Aftershock identification: methods and new approaches, *Geophys. J. Int.* **109**, 501–516.

Molchan, G. M. and V. M. Podgaetskaya (1973). Parameters of global seismicity, *Comput. Seism.* **6**, 44–66, Moskow, Nauka.

Molchan, G. M., E. V. Vilcovich, and V. I. Keilis-Borok (1970). Seismicity and principal seismic effects, *Geophys. J. R. Astr. Soc.* **21**, 323–335.

Molchan, G. M., T. L. Kronrod, and G. F. Panza (1996). Hazard oriented multiscale seismicity model: Italy, UNESCO and International Centre for Theoretical Physics, IC/96/23, Miramare-Trieste, preprint.

Okal, E. A. and B. A. Romanowicz (1994). On the variation of b-values with earthquake size, *Phys. Earth Planet. Interiors* **87,** 55–76.

Pacheco, S. F., C. H. Scholz, and L. R. Sykes (1992). Changes in frequency-size relationship from small to large earthquakes, *Nature* **355,** 71–73.

Pisarenko, V. F., A. A. Lyubushin, V. B. Lysenko, and T. V. Golubeva (1996). Statistical estimation of seismic hazard parameters: maximum possible magnitude and related parameters, *Bull. Seism. Soc. Am.* **86,** 691–700.

Romanowicz, B. (1992). Strike-slip earthquakes on quasi-vertical transcurrent faults: inferences for general scaling relations, *Geophys. Res. Lett.* **19,** 481–484.

Rundle, J. B. (1993). Magnitude-frequency relations for earthquakes using a statistical mechanical approach, *J. Geophys. Res.* **98,** 21943–21949.

Schwartz, D. P. and K. S. Coppersmith (1984). Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones, *J. Geophys. Res.* **89,** 5681–5698.

Stucchi, M., R. Camassi, and G. Monachesi (1993). Il catalogo di lavoro del GNDT. CNR GNDT GdL "Macrosismica", GNDT Int. Rep., Milano, 89 pp.

Turcotte, D. L. (1995). Chaos, fractals, nonlinear phenomena in Earth sciences. U.S. National Report to IUGG 1991–1994. *Rev. Geophys.*, 341–343.

Utsu, R. (1971). Aftershocks and earthquakes statistics (III), *J. Fac. Sci.* Hokkaido University, **3,** 379–441.

Wesnousky, S. G. (1994). The Gutenberg-Richter or characteristic earthquake distribution, Which is it? *Bull. Seism. Soc. Am.* **84,** 1940–1959.

Wesnousky, S. G. (1996). Reply to Yan Kagan's "Comment on 'The Gutenberg-Richter or characteristic earthquake distribution, which is it?'," *Bull. Seism. Soc. Am.* **86,** 286–291.

Woo, G. (1996). Kernel estimation methods for seismic hazard area source modeling, *Bull. Seism. Soc. Am.* **86,** 353–362.

Working Group on California Earthquake Probabilities (1995). Seismic hazard in southern California: probable earthquakes, 1994 to 2024, *Bull. Seism. Soc. Am.* **85,** 379–439.

Young, J. B., B. W. Presgrave, H. Aichele, D. A. Wiens, and E. A. Flinn (1996). The Flinn–Engdahl regionalisation scheme: the 1995 revision, *Phys. Earth Planet. Interiors* **96,** 221–300.

International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Warshavskoye sh., 79, kor.2
Moscow 113556, Russia
(G.M., T.K.)

Dipartimento di Scienze della Terra
via Weiss, 4
34127 Trieste, Italy
(G.F.P.)

International Center for Theoretical Physics
SAND Group
P.O. Box 586
34100 Trieste, Italy
(G.M., T.K., G.F.P.)

Manuscript received 15 February 1996.