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 $\gg 1$).

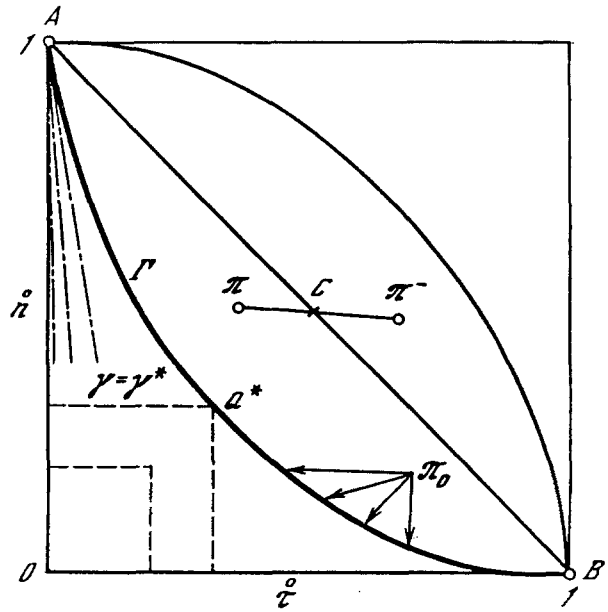

Figure 1
Error set $(\mathring{n}, \mathring{\tau})$ for prediction strategies based on the same information. (A) Optimist strategy. (B) Pessimist strategy. (AB) Results of a random guess. (C) The center of symmetry; π and π^- are a forecast and its antipodal forecast. Γ is the diagram of optimal prediction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

Let us agree on the type of information $\mathcal{I}(t)$ available at the moment t for the prediction of events in the point process. In practice $\mathcal{I}(t)$ 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 $\mathcal{I}(t)$ 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 $\mathcal{I}(t)$. 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 $(\mathring{n}, \mathring{\tau})_i$ of initial strategies with the same weights q and $1 - q$. Hence the set $G = \{(\mathring{n}, \mathring{\tau})_\pi\}$ of errors corresponding to various strategies is convex if these strategies are based on the same information $\mathcal{I}(t)$ (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 $\mathcal{I}(t)$.

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 $\mathcal{I}(t)$; 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 $\mathcal{I}(t)$ 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:

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

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:

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

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 $(n(t), \mathcal{I}(t))$ 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 $\mathcal{I}(t)$. 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, 196u; 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* $\mathcal{I}(t)$. *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 $k \leq m$ 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 $\mathcal{I}(t)$. 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 $(n(t), \mathcal{I}(t))$ *has* M property if the information sequence is a Markov process, that is,

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

and that

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

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

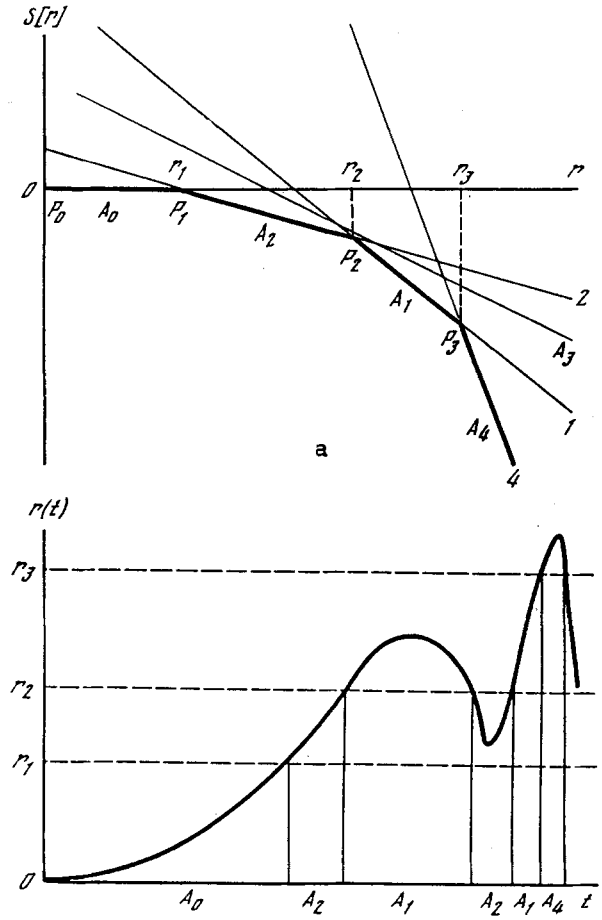

Figure 2

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 $\{\mathcal{I}(t), n(s), s < t\}$ is measurable relative to $\mathcal{I}(t)$. 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 $\mathcal{I}(t) = \{n(s), x(s), t - J < s < t\}$; its dimensionality is less as compared with the case of infinite memory.

Adding new requirements to the description of the process $(n(t), \mathcal{I}(t))$ 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*

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

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 $(\mathring{n}, \mathring{\tau})$. 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 $(\mathring{n}, \mathring{\tau})_\Theta$. 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 $\mathring{\tau}$ 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 $\mathring{\tau}$-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 $\mathscr{I}(t) = u$ 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

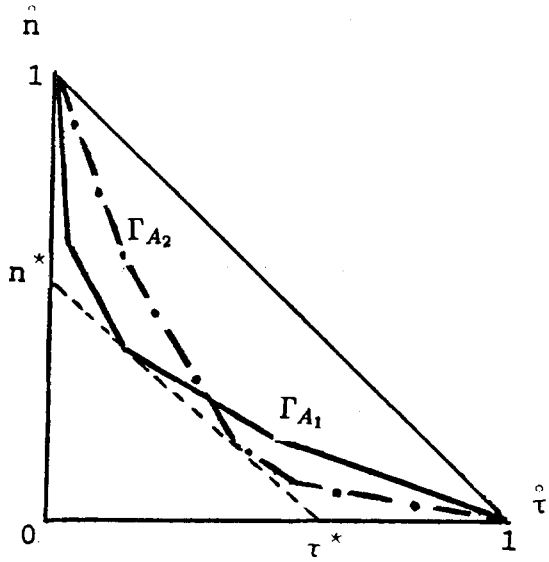

Figure 3
Comparison of algorithms by error diagrams Γ_A. Solid line and dash-dotted line are error diagrams for two algorithms. The line (n^*, τ^*) is the common line of support for curves Γ_{A_1} and Γ_{A_2}. If $\gamma = \lambda\alpha\hat{n} + \beta\hat{\tau}$ is the loss function and $\beta/\alpha\lambda > n^*/\tau^*$ 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, $\mathcal{I}(t)$, that is, the point process $\{t_i, M_i, g_i\}$ is determined by the hazard function $r_u(x)$ with $u = \mathcal{I}(t)$. The most popular model of $r_u(x)$ has the form

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

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.