Earthquake Prediction and Its Optimization

G. M. Molchan

International Institute of Earthquake Prediction Theory, Russian Academy of Science, Moscow, Russia

Y. Y. Kagan

Institute of Geophysics and Planetary Physics, University of California, Los Angeles

This paper continues the work by Molchan (1990, 1991b), who has considered earthquake prediction as a problem of the optimization of a certain loss function \(\gamma\). Function \(\gamma\) is defined by specific social, economic, and geophysical goals. This problem can be fully solved for the loss function \(\gamma\) dependent on only two parameters: fraction of alarm time, \(\tau\), and fraction of failures to predict, \(\nu\). For such a loss function the tasks of geophysicists and decision makers are clearly defined and can be separated; the requirements of the prediction algorithm can be completely formulated in terms of hazard function. We also consider a more complex model of the loss function, introducing a finite number of various alarms. For each of these alarms, \(\gamma\) depends on three parameters; in addition to \(\tau\) and \(\nu\), we take into account the total number of alarms, \(\lambda\). For this model we find strategies which optimise the losses for each time unit (locally optimal). Contrary to the simple case of \(\gamma(\tau, \nu)\), such strategies may not be optimal globally (for the entire time interval). We determine the conditions under which the locally optimal strategy becomes globally optimal. We illustrate our results using the model of a short-term earthquake prediction for central California.

We emphasize that for this complex case, the tasks of geophysicists and decision makers in prediction are intertwined. Prediction and mitigation of many other natural disasters may benefit from the alert strategies discussed here.

1. INTRODUCTION

The appearance of several different algorithms and methods of earthquake prediction (see, for example, McCann et al. [1979], Bakun and Lindh [1985], Heaton [1985], Kagan and Knopoff [1987], Reasenberg and Jones [1989], Sobolev et al. [1989], Keilis-Borok and Rotwain [1990], Keilis-Borok and Kossobokov [1990], Toksoz et al. [1990], and Nishenko [1991]) calls for a new comprehension of prediction and ways of its implementation and utilisation. Aki [1989, p. 197] formulates ideal prediction as follows: "[T units of time (years)] before target earthquake with given magnitude and location, we would tell people that the probability of the occurrence of this earthquake is one per [T years]." Aki [1989, p. 198] defines, therefore, the task of a geophysicist in earthquake prediction as estimating objectively the probability of occurrence of an earthquake with a specified magnitude, place, and time window under the condition that a particular set of precursory data was observed." In this program it is not clear whether the ideal forecast is useful for the solution of certain social and economic problems. These problems are purposely left outside the purview of physical scientists [Aki, 1989]. However, consideration of the problem of earthquake prediction viewed as a whole is always beneficial since it may contribute to development of rational strategies for earthquake mitigation and to analysis of the stability of these strategies.

Molchan [1990, 1991b] has attempted to analyse earthquake prediction algorithms from the point of view of prediction goals. Social, economic, and scientific prediction goals are formulated as a problem of minimisation of a certain function \(\gamma\), which depends on two parameters: fraction of alarm time, \(\tau\), and fraction of events missed by an algorithm, \(\nu\). Two similar parameters to evaluate the prediction's effectiveness have been discussed by Kagan and Knopoff [1987] and Keilis-Borok et al. [1988]. Depending on the circumstances, we call function \(\gamma\) either a loss function or a goal function. If we characterise our prediction by these two parameters only, the structure of optimal decisions is fully described [Molchan, 1990, 1991b]. In this case the general problem of prediction may be subdivided into two completely independent parts: one related to the prediction user (presentation of the loss function \(\gamma\), and the other related to geophysics. The latter subproblem consists of compiling the error curve \(\Gamma = (\tau, \nu)\) for the set of the "best" prediction strategies which use available information, and of evaluating conditional probabilities [Aki, 1989] in real time for very small space-time windows (see more in section 2). In such a program, algorithms of prediction could be considered as one of many possible ways to estimate the error curve \(\Gamma\).

The situation becomes more complex if we consider other characteristics of the prediction. In section 3, we consider several types of earthquake alerts which are characterised by three parameters: \(\tau, \nu, \lambda\). We find locally optimal strategies, which, contrary to the first case in which the loss function is dependent on only two parameters, \(\gamma(\tau, \nu)\), are useful only under specific conditions. We consider in detail a special case when a locally (in time) optimal strategy is globally optimal. We show that this special case is a generalisation of the result of Ellis [1985] and Lindgren [1985].

In section 4, we illustrate our results using the multidimensional model of a short-term earthquake prediction for central California [Kagan and Knopoff, 1987]. We also use a model of the cyclic Poisson process of earthquake occurrence...
discussed by Vere-Jones [1978] to show an application of some techniques developed in previous sections and outline new approaches to the solution of the prediction problem.

Finally, in the last section (section 5) we discuss possible applications of our results. We show that in a general case the prediction problem requires an active dialog between a physicist and an economist/decision maker. The goal of the dialog is to establish real requirements for each side. Since both time-dependent earthquake hazard mitigation measures and statistical methods of earthquake prediction are currently in their infancy, our discussion in several places below is, of necessity, tentative.

Although in our considerations we emphasize earthquake prediction as a major application domain of the developed techniques, it is clear that prediction and mitigation of many other natural disasters [Oaks, 1988] may benefit from the alert strategies discussed here. The general framework of prediction errors and optimisation of losses accompanying these errors should be applicable to such catastrophic natural phenomena as earthquakes, volcano eruptions [Tilling, 1989; Brantley, 1990], landslides, and hurricanes. We hope that our contribution will stimulate the development of quantitative prediction methods as well as a discussion of appropriate and optimal disaster mitigation measures.

2. OPTIMIZATION OF PREDICTION FOR TWO PARAMETERS

De Mare [1980] and Lindgren [1985] were probably the first scholars to point out a fundamental difference between Kolmogorov-Wiener’s idea of prediction of temporal sequences and prediction of the occurrence time of a catastrophe. In De Mare’s and Lindgren’s interpretation, the catastrophe corresponds to a (high) level crossing by some continuous stochastic process. In the first (Kolmogorov-Wiener’s) interpretation, our goal is extrapolation of each sequence value with the least squares errors (vertical prediction), whereas in the latter case it is important to forecast the moment of a disaster (horizontal prediction). In the horizontal prediction case, the error in forecasting a value of a stochastic variable is less important; we are more interested in a compromise between the total length of an alarm period and the number of failures to predict. Forecasting air temperature is an example of the “vertical” prediction, whereas predicting the moment of adverse weather (like frost, etc.) is an example of the “horizontal” prediction. The difference between the vertical and the horizontal predictions is especially obvious in the case of stochastic point processes (earthquakes or explosive volcanic eruptions [Tilling, 1989]) which are temporary sequences of δ functions. Murphy [1985], as well as Winkler and Murphy [1985, and references therein] discuss probabilistic weather predictions and statistical decision strategies based on these forecasts. Some solutions developed therein may be useful for understanding statistical decision making in earthquake prediction problems.

Molchan [1990, 1991b] has formulated a problem of comparing different methods (strategies, algorithms) of earthquake prediction. For this purpose two criteria have been used:

\[
\tau = \text{total time of alarms}/T, \quad \nu = \text{fraction of failures to predict}
\]

\[
(\text{the number of missed events})/(\text{the total number of events in time interval } T), \text{ for } T \gg 1.
\]

A two-dimensional criterion \((\tau, \nu)\) only partially orders all of the possible prediction methods according to their efficiency. Fortunately, however, the set \(G\) of all points on the \((\tau, \nu)\) plane corresponding to various methods of prediction and using the same predictive information \(I(t)\) is always convex (Figure 1). The set \(G\) has a center of symmetry \((1/2, 1/2)\) and two limit points \((1, 0)\) and \((0, 1)\). The first point stands for total "pessimism" (continuous alarm), and the second point corresponds to total "optimism" (no alarm ever). The lower boundary \(\Gamma\) of the set \(G\) connects the points \((1, 0)\) and \((0, 1)\) and corresponds to the minimum set of the "optimal" strategies. It means that for each point on the error curve \(\Gamma\), (1) there is a strategy with larger errors \((\tau, \nu)\), and (2) there is no strategy for which both errors \(\tau\) and \(\nu\) are smaller than those for the optimal strategy. The upper boundary of set \(G\) corresponds to such forecasts when an alert is called if the predicted danger is minimal, and vice versa (antipodal forecasting). All strategies available using a particular prediction algorithm (set \(G\)) are situated between the error curve and the antipodal curve (Figure 1). In section 4 we review the application of the two-dimensional loss function in more detail.

These simple considerations, which are typical of statistical decision theory (see, for example, Chernoff and Moses [1959]), allow us to formulate the prediction problem as a description of error curves \(\Gamma\) corresponding to a certain sequence of information \(I(t)\). Indeed, let us formulate the economic, social, or geophysical problem of earthquake prediction as a problem of minimisation of losses which are characterised by a certain functional \(\gamma\). These losses are determined only by errors \(\tau\) and \(\nu\), and the functional \(\gamma\) strictly increases with each of the arguments \((\gamma(\tau, \nu)\) is a loss function). If sets \(\{\tau, \nu : \gamma < u\}\) are convex for each \(u > 0\), then
the optimal strategy which minimizes \( \gamma \) is completely determined by the error curve \( \Gamma \) [Molchan, 1990, 1991b]. More specifically, if \( p_0 = (r, \nu) \) is a point of tangency of \( \Gamma \) to the isoline \( \gamma = \gamma_0 \), then \( \gamma_0 \) defines the optimal level of losses, and coordinates of \( p_0 \) define errors of the optimal prediction. The optimal prediction operates according to the following algorithm: an alarm is called each time when

\[
\frac{r(t)}{\mu} > \frac{\text{df} / \text{dr}}{(p_0)},
\]

where \( \text{df} / \text{dr} \) is a derivative of \( \Gamma \) at the point \( p_0 \), \( r(t) \) is a conditional rate of predicted event based on the information \( I(t) \), i.e.,

\[
r(t) = \text{Prob}\{\text{an event in } (t, t+\Delta)|I(t)\}/\Delta,
\]

and \( \mu \) is an average number of events per unit of time (rate of occurrence).

In engineering and statistical literature \( r(t) \) is usually called the hazard function or conditional intensity function; if \( I(t) \) consists of all the earthquake catalog data up to moment \( t \), the ratio \( r(t)/\mu \) in geophysics is called the "predictive ratio" [Kagan and Knopoff, 1977] or the "risk enhancement factor" [Vere-Jones, 1978]. Aki [1981] called a similar ratio "probability gain" for the case when \( I(t) \) includes information on all precursors found in interval \( (t-\tau, t) \). The importance of the ratio \( r(t)/\mu \) is determined by the fact that \( \log[r(t)/\mu] \) can be interpreted as the amount of Shannon's information about future events that an observer gains through \( I(t) \).

In the important case of linear losses,

\[
\gamma = \alpha \mu \nu + \beta \tau,
\]

where \( \alpha \) is prevented loss per one successful prediction, equivalently the differences in losses for predicted and unpredicted events, and \( \beta \) is a cost of an alarm per one unit of time. The threshold in (1) is defined without \( \Gamma \), since in the point of contact of two curves (\( \Gamma \) and \( \gamma = \gamma_0 \)) their derivatives coincide. Hence (1) can be replaced by condition

\[
r(t) > \beta/\alpha.
\]

Equation (4) was published simultaneously by Ellis [1985] and by Lindgren [1985].

Molchan [1991b] introduced minimax strategies which have an important role from the point of view of a solution stability. These strategies optimize function \( \gamma = \text{max}(r, \nu) \); therefore they have a property \( r = \nu = \text{min} \). Many researchers, often without rational justification, try to reach an approximate equality of errors \( r \) and \( \nu \) during optimization of algorithms for earthquake prediction. According to results of Keilis-Borok and Rotwain [1990], these errors are of the order of 25-30% each for their long-term prediction algorithm.

Gusev [1974] tried to define the efficiency of earthquake prediction by the parameter \( \eta \):

\[
\eta = (1 - \nu)/\tau.
\]

A similar quantity has been often used by other researchers (see, for example, Kagan and Knopoff [1987]). The quantity (5) by itself cannot be used, however, for development of an optimal alarm strategy. Indeed, we can interpret \( \eta^{-1} \) as \( \gamma \), and all isolines of this function form a bundle of lines through point \( (r, \nu) = (0,1) \) of curve \( \Gamma \) (see Figure 1). This means that the trivial "optimist strategy" or strategies which are close to the optimist strategy are maximally efficient (see also section 4).

Function \( r(t) \) is easy to calculate, if information \( I(t) \) available at time \( t \) is defined only by the time \( t \) elapsed since the last large earthquake on a particular fault or a fault segment. In this case, if the distribution of interevent times is known, complete optimization of the prediction and a computation of curves \( \Gamma \) is possible [Molchan, 1990, 1991b]. For example, predictions by McCann et al. [1979] and Nishenko [1991] are based on information \( I(t) = t \); analysis of various aspects of these efforts is presented by Molchan [1990] and Kagan and Jackson [1991].

3. ADDITIONAL PREDICTION CRITERIA

The approach discussed in the previous section has a certain defect. If we have two strategies with the same errors \( (r, \nu) \), the strategy which has fewer connected intervals of alarm is preferable from the practical point of view, since most earthquake mitigation measures require an initial period during which valuable resources are spent but no mitigation effect is produced. Therefore it is important to consider a model of alarm systems which go beyond the scheme of two parameters.

3.1. Model

Systems of possible earthquake mitigation efforts have not been sufficiently analyzed (see, for example, Pate-Cornell, [1982] and Dmitrieva et al. [1987, and references therein]). We are interested here in that part of the mitigation measures which is actuated after alarm declaration. Formalizing the situation, we subdivide alarm into several phases (alarms of different kinds). Each of these alarms induces an application of a complex of various measures; therefore one alarm may include several other alarms in various combinations.

Similar introduction of alarms (alerts) of different levels is discussed, for example, by Bakun [1988] for the Parkfield prediction experiment, by Brantley [1990] for eruption of Redoubt volcano, and by Pinatubo Volcano Observatory Team [1991]. The duration of an alarm should be defined by decision strategy. The alarm phases can be changed according to certain rules. In a general situation, alarms can be subdivided into phases which should strictly follow in a certain order, required, for example, by technological or social reasons (for example, deployment of civil defense units, shutting down dangerous processes at plants and factories, partial evacuation, etc.). On the other hand, within each alert phase there might be measures which can be carried out either simultaneously or in any order. For simplicity's sake, below we consider two such alarm rules: (1) alarm phases are strictly ordered, i.e., \( i \)-th phase is always preceded by \( (i-1) \)-th phase, \( i \neq 0 \) (i = 0 corresponds to absence of alarm). The decision on a particular occasion can influence future decisions, and, therefore, the anticipation of these future phases of alarm may influence the choice of a current action. Winkler and Murphy [1985] call such decisions sequential. They [Winkler and Murphy, 1985, pp. 515-517] discuss an example of sequential (two-day) strategy which outperforms the strategy based on daily choices. Otherwise, (2) the alarm phases can be in arbitrary order.
As an example of alarms where ordering is imposed by a priori considerations, let us consider possible earthquake alerts at a nuclear power station. Suppose the highest, the i-th level, of alert involves evacuating the personnel from the station; this cannot be accomplished before potentially dangerous processes are safeguarded. Thus an (i-1)-th phase may involve shutting down the reactor; this phase may for technological reasons also require an implementation of some additional preparatory measures. These preliminary alarm phases require time and expense, but they must be performed regardless of cost and earthquake risks, before the i-th alert is called on. The execution of preliminary alerts may be canceled or modified if the probability of a strong earthquake diminishes. As another example, before deployment of civil defense units we need to mobilize and transport them to the place of a forecasted disaster. In section 3.4 below, we consider an additional example of ordered alarms.

A transition from one phase to another (i-th) requires certain initial expenditures $B_i \geq 0$. The transition from $i \neq 0$ to $i = 0$ (calling off an alert) is cost-free. For most earthquake hazard mitigation measures, the initial phase of an alarm should require a time interval (lead time) for implementation. As a first approximation and for simplicity, we take this initial stage to be instantaneous. Alarm maintenance costs $\beta_i$ per unit of time. An introduction of an alarm with accompanying mitigation measures decreases the losses due to a predicted earthquake by a quantity $\alpha_i$.

Any prediction strategy $\pi$ leads to total losses $S_\pi(t)$ (the losses include expenses for prediction and prevented losses taken with a minus sign) for the observation period $t$. The problem is to optimize mean losses per one unit time unit,

$$S = \lim \inf \frac{1}{t} S_\pi(t), \quad t \to \infty,$$

by choosing a prediction strategy $\pi$.

We discretize time in units $\Delta \ll \mu^{-1}$. The number of events $n_t$ in the interval $(t, t+\Delta)$ can be taken to be zero or 1. At moment $t$ an observer has information $I(t)$ about various geophysical fields and earthquake catalogs in intervals

$$J_t = (t - T_j, t - h_j),$$

where $j$ is a type of data, and $h_j$ is a time delay between occurrence of different events and their registration in $J_t$. A prediction strategy $\pi$ at moment $t$ is based on information $I(t)$ and previous decisions $\pi_s = \{\pi_s, t - t_0 \leq s < t\}$. If $I(t)$ and $\Pi_t$ are fixed, decision $\pi_t$ may be deterministic: $\pi_t = i$ (to call i-th alarm), $i = 0, ... p$, or otherwise decision $\pi_t$ may be stochastic. A stochastic decision requires that we conduct an experiment with possible random choices $0, ... p$ and probabilities equal to $p_i(I(t), \Pi_t)$. An outcome of such an experiment determines a stochastic decision.

If $\{n_t, \pi_t\}$ form a stationary ergodic process, the limit (6) is reached with probability 1, and

$$S = \sum_{i=1}^{p} \left( \lambda_i B_i + \tau_i \beta_i - \mu_i \alpha_i \right)$$

$$= \sum_{i=1}^{p} \left( \lambda_i B_i + \tau_i \beta_i + \nu_i \mu_i \alpha_i \right) - \sum_{i=1}^{p} \alpha_i \mu_i,$$

where $\lambda_i$ is the intensity of alarms of the i-th kind, i.e., transition moments to the i-th phase of the alarm; $\tau_i$ is the fraction of i-th alarm time; $\mu_i$ is the rate of occurrence of strong earthquakes predicted by i-th alarms; and $\nu_i = 1 - \mu_i/\mu$ is the fraction of events which were not predicted by i-th alarms.

If $p = 1$, we obtain the loss function $\gamma$:

$$\gamma = B \lambda + \alpha \nu \mu + \beta r.$$

This function for $B = 0$ transforms into (3).

3.2. Locally Optimal Strategy for Strictly Ordered Alarms

For the stationary ergodic process $\{n_t, \pi_t\}$ the quantity is

$$S = E[S_\pi(t + \Delta) - S_\pi(t)]/\Delta,$$

where $E$ is a symbol of mathematical expectation. Therefore we call strategy $\pi_t$ locally optimal, if it minimizes average losses for any moment of time $t$.

Statement 1. Let us consider the model with strictly ordered alarms. Then locally optimal strategy for the above model declares alarm of the i-th kind (after an alarm of the (i-1)-th kind) if

$$r(t) > \frac{(B_i/\Delta + \beta_i)/\alpha_i}{\beta_i}, \quad \text{for } i = 1, ... p,$$

and

$$\frac{(\alpha_i - \alpha_{i-1})r(t)}{\beta_i} \geq B_i/\Delta + (\beta_i - \beta_{i-1})$$

for $i = 2, ... p$, (11)

and keeps the i-th alarm if

$$r(t) < \frac{\beta_i}{\alpha_i}, \quad \text{for } i = 1, ... p,$$

and

$$\frac{(\alpha_{i+1} - \alpha_i)r(t)}{\beta_i} \leq B_{i+1}/\Delta + (\beta_{i+1} - \beta_i),$$

for $i = 1, ... p - 1$. (12)

In all other cases, the i-th alarm is called off (zeroth level). In particular, for $p = 1$ an alarm is called as soon as

$$r(t) \geq \frac{(B/\Delta + \beta)/\alpha}{\beta/\alpha},$$

and is canceled if

$$r(t) \leq \beta/\alpha.$$ (14)

Proof. See Appendix A.

Figure 2 illustrates statement 1. Line a defines, for various levels of $r(t) = r$, average local losses per one unit of time during the i-th alarm; line b defines the same losses when the i-th alarm is changed to the (i+1)-th alarm. Therefore, a locally optimal strategy chooses a decision which leads to minimum losses in the i-th phase of an alarm: if $r(t) \in (r_1, r_2)$, the alarm continues; if $r(t) > r_2$, the alarm switches to the (i+1)-th phase; and if $r(t) < r_1$, the alarm is called off. Thus, for each alarm phase we have two thresholds: one for calling it on, and the other, lower threshold to call off the alarm. A barrier of the next (i+1)-th alarm is overcome only when expected losses are immediately compensated for in the next unit of time as compared with continuation of the i-th alarm. Such a strategy is similar to the strategy of descent from a mountain without using a map: we should move downward; a local barrier is surmounted only if it is smaller than the length of averaging. Moreover, to be op-
Fig. 2. Local optimization of prediction for strictly ordered alarms. The diagram shows the decision in the i-th phase of alarm: line a is for \( y = \beta_i - \alpha_i r \), and line b is for \( y = B_{i+1}/\Delta + \beta_{i+1} - \alpha_{i+1} r \). Intervals \( r(t) \): (0, \( r_1 \)), (\( r_1, r_2 \)), and (\( r_2, \infty \)) correspond to cancellation of i-th alarm, its continuation, and transition to (i+1)-th alarm, respectively.

In our case, in order to declare a new type of alarm, it is necessary to obtain a certain amount of information, so that conditional probability \( r(t) \Delta \) of event occurrence in interval \( \Delta \) increases by amount \( B/\alpha \). The economic gain for a correct prediction, \( \alpha \), should be larger than the initial cost of an alarm, \( B/\alpha < 1 \). Otherwise, the best strategy is the "optimist" strategy (no alarm). The necessary condition for cancellation of an alarm is a decrease of the hazard function below the \( \beta/\alpha \) level. In the short-term seismicity model, discussed in section 4, the hazard function \( r(t) \) has very sharp peaks. Therefore, the locally optimal strategy should be reasonably close to the globally optimal one.

For the whole interevent period [Molchan, 1990]. In other words, on the half axis \( z = t - t_i > 0 \) it is possible to show time moments when alarms need to be declared (alarm set). This set is based on information \( I(t_i) \) and \( \Pi(t_i) \); the set is reconstructed after each event \( t_{i+1} \) (the moment when information \( I(t) \) is updated). Information \( I(t_i) = I_i \) yields a conditional probability of interevent interval

\[
F_-(x) = \text{Prob}\{t_{i+1} - t_i < x|I_i\}.
\]

Information \( \Pi(t) \) is a set of alarm decisions made before time \( t_i \), in particular, whether the last event has been successfully predicted:

\[
\pi(t_i) = \begin{cases} 
1 & \text{if last event } t_i \text{ was predicted,} \\
0 & \text{if last event } t_i \text{ was not predicted.} 
\end{cases}
\]

In this situation the risk function in interval \((t_i, t_{i+1})\) is defined by

\[
r(x) = F'_-(x)/[1 - F_-(x)], \quad \text{for } x = t - t_i,
\]

[\text{Molchan, 1990}].

The next statement illustrates a complexity of the decision rule which optimises losses for the entire interval \((t_i, t_{i+1})\) and is based on information \( I(t) \) for \( p = 1, \alpha > B \), in the interval \((t_i, t_{i+1})\) is defined as follows: the alarm is called for those moments \( t = t_i + x \), when

\[
x \in A = \sum_{k=1}^{n}(a_k, b_k),
\]

where

\[
n \leq \min(\alpha/B, l + 1),
\]

and \( l \) is the number of solutions for the equation \( r(x) = \beta/\alpha \). The points \( a_k \) and \( b_k \) are

\[
0 \leq a_1 < b_1 < a_2 < \ldots < b_k \leq b_x,
\]

where \( b_x \) is the minimum value of \( x \) for which \( F_-(x) = 1 \), and

\[
r(a_k) = \beta/\alpha - B(k - \epsilon), \quad \text{for } k = 2, \ldots, n,
\]

(for \( k = 1, (21) \) is true if \( a_1 \neq 0 \), and

\[
r(b_k) = \beta/\alpha, \quad \text{for } k = 1, \ldots, n - 1,
\]

(for \( k = n, (22) \) is true if \( b_n < b_x \)). In (21) \( \epsilon = 1 \) if \( a_1 = 0 \) and \( \pi(t_i) = 1 \); in other cases \( \epsilon = 0 \).

The number of alarms \( A \) having properties (18)–(22) is finite. The alarm for which the functional

\[
S(A) = \sum_{k=1}^{n} [\int_{a_k}^{b_k} [\beta(1 - F_-(x)) - \alpha F'_-(x)] dx + B \int_{a_k}^{a_{k+1}} (k - \epsilon) dF_-(x)], \quad a_{n+1} = \infty,
\]

reaches minimum is optimal.
Proof. See Appendix B. We illustrate this case with the following examples:

1. Let \( r(x) \) be strictly decreasing from infinity to zero (Figure 3a). (This corresponds to clustering of events.) Then the following strategies are optimal: \( A_1 = (0, z_0) \), \( r(z_0) = \beta/\alpha \), i.e., locally optimal strategy (4); or \( A_2 = \emptyset \) (empty set), i.e., optimist strategy. The first case \( (A_1) \) is valid if \( \pi(t_i) = 1 \) or if \( \pi(t_i) = 0 \) and

\[
C_1 : \quad B < -\int_0^{z_0} \left[ \beta (1 - F_+ (x)) - \alpha F'_- (x) \right] dx \quad (24)
\]

In other cases \( A_2 = \emptyset \), i.e., any alarm is not cost-effective. Therefore, if \( r(x) \) decreases, an alarm is possible only in the beginning of the interearthquake time span. It is cost-effective only if alert declaration leads to small expenditures: if \( \pi(t_i) = 1 \), the cost is zero, since we just continue the previously initiated alert; if \( \pi(t_i) = 0 \), the cost \( B \) is determined by the condition \( C_1 \).

2. Let \( r(x) \) be strictly increasing from infinity to zero (Figure 3b). (This corresponds to a product wear-out in engineering applications or to seismic gaps, i.e., to quasiperiodicity of events.) In this case the optimal strategies are

\[
A_1 = (\pi_1, \infty) = \{ x : r(x) \geq \beta/((\alpha - B)) \},
\]

\[
r(x_1) = \beta/((\alpha - B)),
\]

or \( A_3 = (0, \infty) \), i.e., the pessimist strategy. The second case \( (A_3) \) is valid if \( \pi(t_i) = 1 \) and

\[
C_2 : \quad B > \int_0^{z_0} \left[ \beta (1 - F_+ (x)) - (\alpha - B) F'_- (x) \right] dx \quad (26)
\]

This means that if the cost of alarm is high, it is cost-effective to continue the alert from a previous period, even if the values of \( r(x) \) are small \( (r(x) < \beta/((\alpha - B))) \). In other cases the condition \( A_1 \) prevails. For proof see Appendix B.

3. The risk function \( r(x) \) is U-type function (combination of examples 1 and 2). In this case there are the following optimal strategies:

For \( \pi(t_i) = 1 \):

\[
A_1 = (0, \infty)
\]

and

\[
A_2 = (0, z_0) + (y_1, \infty).
\]

For \( \pi(t_i) = 0 \):

\[
A_3 = (0, \infty),
\]

\[
A_4 = (0, x_0) + (y_2, \infty) \quad \text{if} \quad (\alpha > 2B). \quad (27)
\]

In the above equations \((x_k, y_k)\) where \( x_k < y_k \), are the roots of the equation \( r(x) = \beta/((\alpha - kB)) \). The effectiveness of any alert strategy again depends on the ratio of \( B \) and losses due to the risk of continuation of the alert in the domain of low values of \( r(x) \). The conditions for the alarm continuation are easily obtained as inequalities for \( \tilde{S}(A_i) \) (see (23)).

From these examples it becomes clear that we need to know the structure of the hazard function \( r(x) \) in much more detail than for the case discussed in section 2, where it is sufficient to know only the time moments when \( r(x) \) intersects a fixed threshold \( r_c \) (see (1)).

3.4. Optimal Strategy for Disordered Alarms

Globally optimal strategies in section 3.1 can be successfully described if the initial cost of alarms is zero \((B_i = 0)\), and alarms are not ordered, i.e., alarms can be declared in an arbitrary order or combination. The information on which the predictions are based can include any data, not just an earthquake catalog.

On plane \((r, y)\) let us consider a set of points \( U \), which satisfy relations (Figure 4)

\[
y \leq 0,
\]

\[
r \geq 0,
\]

\[
y \leq \beta_i - \alpha_i r \quad i = 1, \ldots, p \quad (28)
\]

\( U \) is a convex polygonal domain, since it is the result of the intersection of half planes. Let
The idea of statement 3 is clear after statement 1, where the choice is between the i-th and the (i+1)-th alarms. Lines $y = \beta_i - \alpha_i r$ correspond to the mean local losses per unit time on the condition that the i-th alarm is called when $r(t) = r$. Therefore, the boundary of $U$ corresponds to minimum local losses in state $r(t) = r$, when there is a choice between several different kinds of alarms. Statement 3 is nontrivial, because a locally optimal solution is at the same time globally optimal. As we see from (30), there is certain alarm order in the optimal strategy according to the value of the hazard function. This order is determined by parameters $(\alpha_i, \beta_i)$.

Strategy (30) uses no more than $p$ levels of the hazard function, whereas in the cases discussed earlier (sections 3.2 and 3.3), optimal strategies are based on the whole structure of the hazard function. Therefore, the stability of solution (30) is possibly superior to the solutions analysed in sections 3.2 and 3.3. The difference between globally and locally optimal strategies is not only due to the condition $\{B_i \neq 0\}$. Suppose alarms 1, 2, ... $p$ are strictly ordered, but the optimal strategy (30) in the unordered case has an inversion like 3,2,1,4,... Then if $r(t)$ is small, to declare alarm 3 we would need at least two units of time for calling alarms 1 and 2. Therefore, strict ordering of alarms even for $\{B_i = 0\}$ may create nonzero losses if certain alarms are declared. In its turn, this may decrease the efficiency of a locally optimal strategy.

To illustrate the difference between ordered and disordered alerts, let us consider a hypothetical example. Suppose three earthquake mitigation measures are proposed: (1) closing a bridge to all traffic, (2) evacuating the community, and (3) the combination of measures 1 and 2. The values of the parameters are $\alpha_1 = 50,000$, $\beta_1 = 1,000$ per day, $\alpha_2 = 100,000$, $\beta_2 = 5,000$, $\alpha_3 = \alpha_1 + \alpha_2$, and $\beta_3 = \beta_1 + \beta_2$. If evacuation can be accomplished by routes other than the bridge, the alarms are disordered. Making a plot similar to Figure 4, we find that the optimal strategy is as follows: at $r_1 \geq 2\%$ per day, we should carry out measure 1 and implement measure 3 when $r_2 \geq 5\%$. However, if the bridge is needed for evacuation, alerts are ordered a priori. In this case we should start implementation of measure 3 at $r \geq 4\%$. If probability $r$ is between 2% and 5%, the latter strategy is less effective than the former combination of measures. Most probably, the population cannot be evacuated immediately, so measure 2 has to be carried out before measure 3 in the case of "ordered" alerts. This should further degrade the performance of the latter strategy.

4. Examples (Seismicity Models and Prediction)

4.1. Short-Term Prediction Model

We illustrate our results using the multidimensional self-exciting model of seismicity for central California [Kagan and Knopoff, 1987]. The prediction in the model is based only on the earthquake catalog for central California (CALNET catalog) [Marks and Lister, 1980, and references therein]. In this model, all earthquakes above a cutoff level $M_0$ are subdivided into background (main) and cluster (dependent) events. Former shocks form a stationary Poisson process with a rate $\mu_0$ (earthquakes/dayxkilometers) distributed uniformly along the San Andreas fault ($SA$). Each background event $(t_0, \sigma_0, M_0)$, where $t_0$ is time, $\sigma_0$ is location, and $M_0$ is the seismic moment, may generate, independently from other events, a Poisson sequence of earthquakes with the rate $\phi(M_0)(t - t_0, \sigma - \sigma_0, M)$. This sequence is concentrated around the background event (earthquakes of the first generation). These new events, in their turn, may produce earthquakes of second generation according to the same rules, etc. The events of all generations form a cluster of events which is normally interpreted as a foreshock, main shock, aftershock sequence.

In this model, the hazard function $r(t)$ can be calculated...
analytically if we know all of the earthquakes \((t_i, z_i, M_i)\) which have occurred by the time \(t\):

\[
\frac{r(t|z, M)}{\mu_0} = \mu_0 \phi(M) + \sum_i \psi(M_i)(t - t_i, z - z_i, M), \quad \forall \sigma \in \mathcal{S}, A,
\]

where \(\phi(M)\) is the seismic moment distribution (see more discussion in the work of Kagan and Knopoff [1987]). Function \(\psi\) depends on three parameters which are evaluated by the maximum likelihood analysis [Kagan and Knopoff, 1987; Kagan, 1991]. The rate \(r\) in (31) is for any earthquake with a moment larger or equal to \(M_i\); the actual size of a predicted event is selected according to \(\phi(M)\) distribution.

In our calculations we use the part of the CALNET catalog (1971-1977) which covers the San Andreas fault in central California. Only events with a local magnitude greater than or equal to 1.5 have been used \((M_e = 1.5)\). The length of the fault spanned by the catalog is 363 km, \(\mu_0\) is estimated to be 0.0058 earthquakes/(day\times kilometers). In our predictions we use the hazard function (31) and assume that the information contained in the catalog becomes available in real time. In terms of (2) and (7) it means that \(\Delta \to 0\) and \(h \to 0\).

Figures 5-7 show \(r(t)/\mu_0\) as a function of time and space. Figure 5 shows \(r(t)/\mu_0\) at the location of Stockdale Mountain (near Parkfield, California). The vertical lines in the top diagram indicate that the probability rises immediately following each of several small events. These sharp peaks are the functions of \(t^{-3/2}\). The height of each spike depends on the seismic moment of the triggering event and its distance from Stockdale Mountain (see more details in the work of Kagan and Knopoff [1987]).

The threshold rate \(r_e\) is also normalised by the background Poisson rate \(\mu_0\):

\[
r_e = \mu_0(1 + \kappa) .
\]

In Figure 5 the threshold rate parameter is taken to be...
\( \kappa = 10 \). Figure 6 shows a blowup of Figure 5 for a 5-day interval including the Stockdale Mountain event with local magnitude 4.6. The Stockdale event had a small "foreshock," although the conditional probability had dropped below the threshold before the main event. In the few minutes following the main event, the probability per unit of time of another event exceeded the background Poissonian rate by 400,000 times. The earthquake risk remained more than 10 times the Poissonian rate for several days.

Figure 7 shows a plot of the ratio \( \tau/\mu_0 \) as a function of time and space during 7 years and over 363 km of the San Andreas fault. The southern (\( z = 0 \)) end of the region is near Cholame (California), while the northern end (\( z = 363 \, \text{km} \)) is north of San Francisco. The stippled areas indicate predicted rates that exceed the threshold rate (32) with \( \kappa = 10 \). Temporal and spatial variations in seismicity are clearly visible. Due to a large-scale plot, details of the space-time hazard function are not obvious, but Figure 2 of Kagan and Knopoff [1987] displays part of the hazard map on a much smaller scale (7 days \times 10 \, \text{km}). The horizontal line in Figure 7 at about \( z = 26 \, \text{km} \) shows the location of Stockdale Mountain (see Figure 5). The arrow at \( t = 1461 \) days points to the time of occurrence of the Stockdale Mountain earthquake (see Figure 6).

In Table 1 we demonstrate the calculation of \( \Gamma \) for the prediction algorithm proposed by Kagan and Knopoff [1987]. Various points in the top part of the table (titled "One Threshold") are obtained by changing parameter \( \kappa \) of (32) and measuring \( \tau \) and \( \nu \) for obtained prediction maps (such as Figure 7). We measure \( \nu \) as the fraction of earthquakes which fall into nondangerous areas (open areas in Figure 7), and \( \tau \) is the fraction of the time-space stippled areas (see Figure 7) as compared to the total time-space area for the catalog (7 years \times 363 \, \text{km}).

We also calculate the average number of alarms \( \lambda \) per one point of the San Andreas fault (last column in Table 1). Unlike \( \tau \) and \( \nu \), the number \( \lambda \) is not normalised in these calculations. We cannot use for normalisation purposes a magnitude-frequency relation since it is defined for a unit of area, and \( \lambda \) is defined for a point in space. We could, however, normalise \( \lambda \) using intensity-frequency or maximum acceleration-frequency distributions. Then \( \lambda \) would represent a fraction of alarms compared to the number of destructive earthquake ground motions affecting an average
Fig. 7. Earthquake hazard map for a 363-km segment of the San Andreas fault in central California. The CALNET catalog for 1971-1977 has been used. The time scale is the same as in Figure 5. The prediction point for Figure 5 is at the 26-km coordinate in this figure. The stippled areas indicate predicted rates that exceed the threshold rate. The threshold rate parameter is $\log_{10} t = 1$.

The prediction strategy shown in line 1 of Table 1 has the maximum efficiency $\eta$, but it fails to predict almost all earthquakes. However, this strategy has an advantage of having the minimum number of alarms $\lambda$ per year because of a very large threshold ($\kappa = 10^4$) level used in calculations. For this threshold, only relatively few strong earthquakes trigger an alarm. The alarm rate $\lambda$ reaches the maximum for $\kappa$ close to 1.0, i.e., for the threshold level which is only twice as large (see (32)) as the background rate. For smaller thresholds, $\lambda$ again decreases since alarm areas merge. In principle, if we take $\kappa \to 0$, the whole catalog area should become an alarm zone ("pessimist" strategy; see section 2). This does not happen in Table 1 because of high spatial inhomogeneity of seismicity along the San Andreas fault. In the top part of the plot (Figure 7) which corresponds to the vicinity of San Francisco, there were relatively few earthquakes during 1971-1977; hence this area is not covered by alarms even for very low values of $\kappa$.

In Figure 8 we plot the set of errors $(\tau, \nu)$ displayed in Table 1. As one should expect from a short-term prediction, failure to predict errors $\nu$ decrease very fast as $\tau$ increases to a few percent of the total time-space. One would expect the curve to be linear from about the point corresponding to entry 11 in Table 1 to the point $(0,1)$, since the short-term prediction algorithm will not be much more advantageous than a random guess for large time intervals. The strategies corresponding to the dashed line in the plot are obtained by declaring nondangerous space-time areas in Figure 7 to be dangerous with probability $p$, $0 \leq p \leq 1.0$. The observed improvement in the predictive power of the method for larger values of $\tau$ is, therefore, in our opinion, an artifact of the averaging method: calculating predicted events, we have assumed that seismicity has only short-term fluctuations and is uniform along the San Andreas fault. Calling, for example, a 7-year alarm in the southern half of the catalog area and no alarm in the northern half could result in a significant advantage in $(\tau, \nu)$ errors without using any time-dependent information. Thus, to calculate adequately errors $\tau$ and $\nu$, our model of seismicity should take into account long-term, long-range variability of earthquake occurrence.

In order to illustrate what improvement in prediction can be achieved using the results of section 3.2, we calculate $\tau$, $\nu$, and $\lambda$, applying two thresholds. The results are displayed in the bottom part of Table 1; these thresholds are separated by a slash. The first threshold parameter, $\kappa_1$, is used for declaration of an alarm, and its value is set at 10 times
TABLE 1. Parameters of Short-Term Prediction System

<table>
<thead>
<tr>
<th>Prediction Strategy</th>
<th>Threshold Parameter $\log_{10}(\alpha)$</th>
<th>Fraction of Failures to Predict $\nu \cdot 100%$</th>
<th>Fraction of Alarm Time $\tau \cdot 100%$</th>
<th>Prediction Efficiency $(1 - \nu)/\tau$</th>
<th>Number of Alarms per Year $\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One Threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5.0</td>
<td>99.5</td>
<td>&lt;0.0001</td>
<td>5280</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>4.5</td>
<td>95</td>
<td>0.0005</td>
<td>3500</td>
<td>2.1</td>
</tr>
<tr>
<td>3</td>
<td>4.0</td>
<td>97</td>
<td>0.002</td>
<td>2130</td>
<td>3.4</td>
</tr>
<tr>
<td>4</td>
<td>3.0</td>
<td>93</td>
<td>0.009</td>
<td>758</td>
<td>4.9</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>88</td>
<td>0.04</td>
<td>284</td>
<td>5.8</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>81</td>
<td>0.2</td>
<td>93.9</td>
<td>6.4</td>
</tr>
<tr>
<td>7</td>
<td>0.5</td>
<td>77</td>
<td>0.6</td>
<td>51.2</td>
<td>6.5</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>74</td>
<td>0.9</td>
<td>20.2</td>
<td>6.6</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>71</td>
<td>1.4</td>
<td>20.9</td>
<td>6.6</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>68</td>
<td>2.1</td>
<td>15.2</td>
<td>6.5</td>
</tr>
<tr>
<td>11</td>
<td>1.0</td>
<td>62</td>
<td>3.9</td>
<td>9.9</td>
<td>6.3</td>
</tr>
<tr>
<td>12</td>
<td>1.6</td>
<td>51</td>
<td>9.1</td>
<td>5.4</td>
<td>5.4</td>
</tr>
<tr>
<td>13</td>
<td>2.0</td>
<td>41</td>
<td>15.4</td>
<td>3.9</td>
<td>4.5</td>
</tr>
<tr>
<td>14</td>
<td>2.6</td>
<td>25</td>
<td>28.4</td>
<td>2.6</td>
<td>3.0</td>
</tr>
<tr>
<td>15</td>
<td>3.0</td>
<td>17</td>
<td>38.8</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>Two Thresholds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5/4</td>
<td>98</td>
<td>0.0009</td>
<td>2360</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>4/3</td>
<td>93</td>
<td>0.009</td>
<td>774</td>
<td>3.3</td>
</tr>
<tr>
<td>3</td>
<td>3/2</td>
<td>88</td>
<td>0.04</td>
<td>286</td>
<td>4.6</td>
</tr>
<tr>
<td>4</td>
<td>2/1</td>
<td>81</td>
<td>0.2</td>
<td>94.0</td>
<td>5.4</td>
</tr>
</tbody>
</table>

that of the second threshold parameter, $\kappa_2$, which controls cancellation of the alarm, $\kappa_1 = 10\kappa_2$. Such values of two thresholds would correspond to certain costs of $B$, $\beta$, and $\alpha$. It can be expected that the values of $\tau$ and $\nu$ should be close to those obtained in the top part of Table 1 for $\kappa = \kappa_2$. On the other hand, the value of $\lambda$ is even smaller than the one obtained for $\kappa = \kappa_1$. Obviously, some alarm zones which were initiated when $r(t)$ exceeded the appropriate $r_1$ were not terminated when $r(t)$ fell below $r_1$, but instead they merged with other zones, since we now use a lower threshold $r_2$ to call off the alarm. Therefore, we observe significant improvement in the parameter set $(\tau, \nu, \lambda)$ of the predictive algorithm. These calculations are made again for $\Delta = 0$, $h = 0$. In practical predictions we would need to calculate the conditional intensity function for various values of $\Delta$, especially when the lead time $h$ (see (7)) is significant. These calculations can be made by extensions of the methods discussed by Kagan and Knopoff [1977, 1987] and by Kagan [1991]; the last two papers evaluate the loss of information due to a nonzero time delay.

In order to demonstrate the relation between the values of alert parameters, the resulting thresholds, and the losses, let us assume that there is only one alert level, and an earthquake to be predicted has the magnitude $M = 5.5$. Taking the b value of the magnitude-frequency law to be equal to 1.0, for a fault segment of about 25 km we obtain the Poisson rate $\mu = 10^{-8}$ earthquakes/min (eq/min). There is a certain difficulty in using the results of Table 1 for a straightforward threshold calculation: the values of $\lambda$ have been computed here for an ideal case when the information about earthquakes is available immediately after the end of coda, and protective measures can be instantly carried out (see more discussion by Kagan and Knopoff [1987]). Thus in Table 1 even those alarms which are separated by less than 1 min are taken as distinct alerts. (The duration of an earthquake with $M_L = 1.5$ is of the order of 20 s, increasing for stronger events to a few minutes.) We could have calculated $\lambda$ for the case of alerts having a "dead" time, i.e., alerts separated by less than a certain time interval $\Delta t$ are considered as one alert, and $\Delta t$ may be related to the time scale of protective measures; but since we presently lack data on $\Delta t$ size, we refrained from doing this. In our discussion in section 3.2, we assumed that the time is discretized in units of $\Delta$; thus we take in this example $\Delta = 1$ min.

As an illustration, we take the cost of initiating an alarm $B = 9$, the cost of sustaining an alarm per min $\beta = 1.01$, and the prevented loss due to the alarm $\alpha = 1$,000,000; since only ratios of these quantities enter (13) and (14), $B$, $\beta$, and $\alpha$ can be multiplied by an arbitrary factor. Applying the above formulas, we obtain the threshold for initiating the alert, $r_1 = 1.001 \times 10^{-4}$ eq/min, and the threshold for terminating it, $r_2 = 1.01 \times 10^{-4}$ eq/min. These thresholds correspond to line 3 in the bottom part of Table 1. Using (9), and the values of errors from Table 1 plus the above mentioned values of expenses, we calculate an average loss per minute of $\gamma = 0.928 \times 10^{-2}$ for strategy 3, and of $\gamma = 0.941 \times 10^{-2}$ and $\gamma = 1.021 \times 10^{-2}$ for strategies 2 and 4, respectively. Thus the loss is at minimum for line 3, as expected. If we take $B = 0$, the threshold levels for the top part of Table 1 may be obtained (see (4)), and the losses can be easily calculated using (3). In a real-life situation, many additional factors like a spatial distribution of ground motion, the magnitude-frequency law, etc., are to be taken into account.
4.2. Cyclic Poisson Process

Another example of a statistical model of seismicity is the cyclic Poisson process applied by Vere-Jones [1978] for description of Kawasumi's historical earthquake sequence in the South Kwanto area of Japan during the years 818-170. Although the physical significance of this model is not clear [Vere-Jones and Ozaki, 1982], the cyclic model is of interest as an example of long-term prediction. The model is a Poisson process with a rate

$$
\mu(t) = A \exp[\rho \cos(\omega t + \phi)],
$$

where according to Vere-Jones [1978]

$$
A = 0.021, \quad \rho = 1.06, \quad \omega = 0.0914, \quad \text{and} \quad \phi = 0.5.
$$

In this case the hazard function is \( \tau(t) = \mu(t) \), and the error curve is calculated as follows

$$
\nu = J_\mu(\tau)/\mu(0),
$$

where

$$
J_\mu(\tau) = \int_0^\tau \exp(\rho \cos \pi x) dx.
$$

The curve \((\tau, \nu)\) is shown in Figure 8. The optimal alarm set which, using a fixed level of error \( \tau = \tau_0 \), minimizes \( \nu \) (min\( \nu \)) is given by (35) consists of the segments

$$
\left( -\pi \tau_0 - \phi + 2\pi k, \quad \pi \tau_0 - \phi + 2\pi k \right),
$$

\( k = 0, \pm 1, \ldots \)

The a priori quality of prediction in the model is not very good. However, the minimax strategy leads to errors \( \tau = \nu \approx 35\% \). Other models which use an interevent time distribution with a coefficient of variation (standard deviation/mean) that equals 0.6 [Nishenko, 1991] lead to similar prediction errors if we use the minimax strategy and minimal information, namely, time after the last large event [see Molchan, 1991b].

The cyclic Poisson model allows for an effective solution of a globally optimal strategy for one alarm type and nonzero cost \( B \) for the alarm initiation. The strategy is constructed nontrivially by methods of dynamic and linear programming [Bellman, 1957]. The solution demonstrates that introducing the initial cost in the loss function leads to complicated strategies even for the simplest seismicity models.

In a long-term prediction one should not deal with an average loss for unit time (6), but rather with overall losses \( S_\tau \) in the interval \((t_0, \infty)\). These losses should be depreciated to the initial moment \( t_0 \) taking into account the discount factor \( \zeta \). The factor \( \zeta \) can be understood as a coefficient of effectiveness of capital investments: losses \( u \) in the moment \( t \) are equivalent to losses \( u \exp[-\xi(t - t_0)] \) incurred in the initial moment \( t_0 \). Let us find the strategy which optimizes average losses \( S_\tau \) for the model in section 3.1. For strategy \( \pi(t) \), depreciated losses in \((t, \infty)\) are

$$
S_{\tau, \pi(t-\Delta)} = u_t + \xi u_{t+\Delta} + \xi^2 u_{t+2\Delta} + \ldots
$$

where \( u_t \) are losses in the interval \((s, s + A)\); \( \xi = \exp(-\xi\Delta) \); and the subscript \( \pi(t-\Delta) \) indicates boundary conditions in the moment \( t \): the alarm was on \((\pi = 1)\) or off \((\pi = 0)\).

Let \( S_{\tau} = E[S_{\tau, \pi}] \) be the optimal expected discounted cost with the initial condition \( \pi(t-\Delta) = \epsilon \). The solution \( \pi(t) = 1 \) leads to loss \( B\xi^2 + \beta\Delta \), which follows from the declaration of an alarm \((\epsilon = 0)\) and its continuation. Another solution, \( \pi(t) = 0 \), leads to average loss \( \mu(t)\Delta \), which follows from failure to predict the event \((t, t + A)\). Taking into account (38), for a stochastic Polygon process with an arbitrary time-dependent rate \( \mu_t \), we obtain for \( S_{\tau, \pi} \)

$$
S_{\tau, \pi} = \min \left( B\xi^2 + \beta\Delta + \xi \mu(t), \alpha \mu(t) \right),
$$

where \( \tau = t_0, \Delta, 2\Delta, \ldots \), and \( \epsilon = 0, 1 \). If \( \mu_t \) is periodic with the period \( T = N\Delta \), then \( S_{\tau, \pi} = S_{\tau, \pi_t} \). Therefore we can consider (39) as a system of equations with \( 2N \) unknowns. The system has a solution. Indeed, the right-hand side of (39) determines a continuous mapping \( Q \) of the \( 2N \)-dimensional hypercube

$$
\left\{ S_{\tau, \epsilon} \in [0, L], \quad k = 0, \ldots, N - 1 \right\},
$$

with a side \( L = \max[BB\xi^2 + \beta\Delta, \alpha]/(1 - \xi) \) into itself. (We take into account that \( \mu(t) \leq 1 \)). The mapping \( Q \) is contractive:

$$
|Q S_{\tau, \epsilon} - Q S_{\tau, \epsilon'}| < \xi \max |S_{\tau, \epsilon} - S_{\tau, \epsilon'}|,
$$

which follows from inequality
\[ |\min(x_1, x_2) - \min(y_1, y_2)| < \max_i |x_i - y_i|. \]  

(42)

The above inequality should be applied to the right-hand side of (39). Therefore solution (39) is unique and can be found by an iterative procedure:

\[ S_{i+1}^{(n)} = 0, \quad S_{i+1}^{(n+1)} = Q_{i+1}^{(n)} \rightarrow S_{i+1}, \quad n \rightarrow \infty. \]  

(43)

If we know \( S_{i+1} \), it is easy to find the optimal strategy \( \pi(t) \). According to (39), if we are in state \( \pi(i-1) = \epsilon \), the optimal strategy is \( \pi(t) = 1 \) if \( S_{i+1} \) coincides with the first term of the right-hand side of (39), and \( \pi(t) = 0 \) otherwise.

In the simplest case of the uniform Poisson process, \( \mu = \mu \), (39) can be solved explicitly and leads to strategies which are independent of time: (1) an "optimist" strategy \( \pi(t) = 0 \), if \( \mu < \beta/\alpha \); (2) a "pessimist" strategy \( \pi(t) = 1 \), if \( \mu > (\beta + \beta)/\alpha \); and (3) for an intermediate case \( \beta/\alpha \leq \mu \leq (\beta + \beta)/\alpha \), a "conservative" strategy \( \pi(t) = \equiv (\pi(\mu) \equiv (\pi(\mu) - 0) \right) \) that does not change the initial decision. The approach, discussed in this section, is applied for solution of more general prediction problems [Molchan, 1991b, 1992a].

5. DISCUSSION

We have discussed problems of earthquake prediction together with problems of prediction application. The situation is relatively simple when prediction goals can be formulated as optimization of a loss function of two variables \( \gamma(\tau, \nu) \). If we select function \( \gamma(\tau, \nu) \), then various methods of prediction proposed in geophysics can be compared and evaluated. From the practical point of view, a domain of application of most prediction methods is unnecessarily narrow, since if an alarm rule is formulated as a final result, we are given in effect only one point \( (\tau, \nu) \) of the error curve. However, any proposed method of prediction commonly has an internal parameter(s), \( \theta \). Modifying these parameters, we can obtain in most cases the entire error curve \( \{\pi(\tau), \nu(\nu)\} \). The strategies corresponding to this curve significantly expand the range of prediction application, as well as allow us to make more complete comparisons of prediction methods [Molchan, 1990, 1991b].

The problem of prediction becomes more difficult if we use a more complex loss function (8). This function takes into account various phases of alarms, their ordering, and the number of alarms connected in time. The globally optimal decisions in this case do not coincide with locally optimal ones. These decisions require a detailed analysis of the whole stochastic structure of hazard function \( \gamma(t) \) together with parameters of the goal function. Therefore, a more important task is to find simple quasi-optimal global decisions. In other words, separation of duties in prediction between geophysicists and economists or decision makers becomes more questionable in this situation. Therefore, a new chain of requirements appears: a more complicated goal function necessitates better knowledge of the hazard function, which, in turn, leads to instability of obtained solutions. Hence, cooperation between specialists in various fields becomes more essential.

We find that one of these complex subproblems is, however, relatively close to the problem of the simple loss function \( \gamma(\tau, \nu) \) discussed in section 2: the finite number of alarms which can be called in arbitrary order and without any initial costs. In this case a locally optimal prediction is at the same time globally optimal, the strategy being a natural generalization of simple alarm systems introduced by Ellis [1985] and by Lindgren [1985]. That prediction uses only a finite number of levels for hazard function \( \gamma(t) \), and therefore it may be useful in practical applications [Bakun, 1988; Brantley, 1990]. Before such applications are possible, several additional problems need to be solved. We need to know \( \gamma(t) \) and economic parameters \( \alpha_i \) and \( \beta_i \). We feel that general methods for evaluation of the function \( \gamma(t) \) are not sufficiently developed at present.

In conclusion, we would like to note that the methods and strategies discussed in this paper may be used in prediction and mitigation of natural disasters other than earthquakes. Actually, results of section 3.3, in which only minimal information about the one-dimensional process is used, may be more applicable, for example, to prediction of volcano eruptions [Tilling, 1989; Brantley, 1990; Pinatubo Volcano Observatory Team, 1991] than to forecasting such a multidimensional process as earthquake occurrence.

APPENDIX A: PROOF OF STATEMENT 1

Losses \( \delta S \) in time interval \( \Delta \) are

\[ \delta S = \sum_{i=1}^{p} [\tau_{i-1} = i - 1] [\tau_i = i] (B_i + \beta_i \Delta - \alpha_i n_t) \]

\[ + \sum_{i=1}^{p} [\tau_{i-1} = i] [\tau_i = i] (\beta_i \Delta - \alpha_i n_t), \]  

(41)

where \([.]. \) here and below is a logical function: \([A] = 1 \) if \( A \) is true, and \([A] = 0 \) if \( A \) is false. From (41) average losses are

\[ E(\delta S) = \sum_{i=1}^{p} B_i E[\tau_{i-1} = i - 1] [\tau_i = i] \]

\[ + \sum_{i=1}^{p} E[\tau_i = i] \beta_i \Delta - \sum_{i=1}^{p} \alpha_i E[\tau_i = i] n_t, \]  

(42)

where \( E \) is a symbol of mathematical expectation (averaging). The decision \( \tau_t \) depends on information \( I(t) \equiv \tau_t \); therefore using conditional expectations, we obtain

\[ E[\tau_t = i] n_t = E[\tau_t = i] E[n_t | I_t] \]

\[ = E[\tau_t = i] \tau_t \Delta. \]  

(43)

Substituting (A3) in (A2) and using the relation

\[ [\tau_t = i] = ([\tau_{t-1} = i - 1] + [\tau_{t-1} = i]) [\tau_t = i], \]

\[ i = 1, \ldots, p, \]  

(44)

which was already used implicitly in (A2), we obtain

\[ E(\delta S) = \sum_{i=1}^{p} E[\tau_{i-1} = i - 1] \]

\[ \times \{[\tau_t = i] (B_i + \beta_i \Delta - \alpha_i n_t \Delta) + \]

\[ [\tau_t = i - 1] (\beta_i \Delta - \alpha_i n_t \Delta) \} \]

\[ + E[\tau_{t-1} = p][\tau_t = p] (\beta_p - \alpha_p n_t \Delta), \]  

(45)

where \( \alpha_0 = \beta_0 = 0 \).
If decisions \( \pi_t \) are based on data \( I_t \) and decisions \( \pi_{t-1} \) and \( \pi_t \) are of a stochastic nature (see section 3.1), then we average in (A5) first by all outcomes of choosing \( \pi_t \) when \( \{I_t, \pi_{t-1}\} \) are fixed. Then \( \pi_t = i \) in (A5) can be replaced by its conditional average

\[
E[(\pi_t = i) | \pi_{t-1}, I_t] = g_i(\pi_{t-1}, i).
\tag{A6}
\]

It is obvious that

\[
0 \leq q_t(j, i) \leq 1, j = i - 1, i;
\tag{A7}
\]

and all other cases \( q_t(j, i) = 0 \).

Therefore

\[
\frac{\partial S}{\partial a_k} = \beta \left( 1 - F_-(a_k) \right) - \alpha F'_-(a_k) = 0.
\tag{B5}
\]

Conditions (B4) and (B5) are true for \( a_k \) and \( b_k \) if these points are inside the interval \((0, b_k)\). We obtain (21) and (22), since \( F_-(a_k) < 1 \) and \( F_-(b_k) < 1 \). From estimates

\[
0 \leq r(a_k) = \beta [\alpha - B(n - \epsilon)] < \infty
\tag{B6}
\]

(r \( \neq \infty \) because of its continuity), it follows that \( \alpha - B(n - \epsilon) > 0 \) or

\[
n \leq \alpha / B + \epsilon.
\tag{B6}
\]

Let the equation \( \tau(x) = c \) have a finite number of solutions

**APPENDIX B: PROOF OF STATEMENT 2**

Let

\[
A = \sum_{i=1}^{n}(a_i, b_i)
\tag{B1}
\]

be a set of intervals on half axis \( z \geq 0 \). If we consider \( A + t_i \) as an alarm set in cycle \( (t_i, t_i+1) \), then the intersection \( A \cap (0, L), L = t_i+1 - t_i \) determines the alarm set which has been declared in this time interval. The losses are

\[
S_i = \beta \int_{0}^{L} xA(z) dz + B \sum_{k+1}^{n}(k - \epsilon) xA(a_k, a_{k+1})(L) - \alpha xA(L),
\tag{B2}
\]

where \( xA(z) = 1 \) if \( z \in A \), and \( xA(z) = 0 \) for the opposite case; \( \epsilon = 1 \) if \( \pi(t_i) = 1 \) and \( a_1 = 0 \). Otherwise, \( \epsilon = 0 \).

Let \( F_-(z) \) be a conditional distribution of \( L \) given information \( I(t_i) \) and \( \pi(t_i) \). Then

\[
E\left( \int_{0}^{L} xA(z) dz | I(t_i) \right) = \int_{0}^{\infty} \int_{0}^{z} xA(u) dF_-(z)
\tag{B3}
\]

Therefore, conditional average \( \bar{S} = E[S_i | I(t_i), \pi(t_i)] \) has form (23). Let \( \text{Prob}\{L \geq b\} = 0 \); then (20) is true, i.e., intervals \( (a_k, b_k) \) are nonzero and nonintersecting, and \( b_k < b \), \( k < n \). If \( \{a_k, b_k, i = 1, \ldots, n\} \) minimize \( \bar{S} \), then for \( 1 < k < n \)

\[
-\frac{\partial S}{\partial a_k} = \beta \left( 1 - F_-(a_k) \right) - \alpha F'_-(a_k) = 0,
\tag{B4}
\]

and for \( 1 \leq k < n \)

\[
-\frac{\partial S}{\partial b_k} = \beta \left( 1 - F_-(b_k) \right) - \alpha F'_-(b_k) = 0.
\tag{B5}
\]
for \( c = \beta / (\alpha - kB) \). Then from (B4)-(B6) it follows that there is a finite number of choices for intervals \( \{ (a_i, b_i) \} \).

The best solution for \( A \) is that which minimizes \( S \), defined by (23).

As an illustration, let us consider example 2 in section 3.3, when \( f(x) \) is strictly increasing. In such a case the equation \( r(x) = c \) has only one solution; hence \( n \leq 2 \). If \( r(\varepsilon_k) = \beta / (\alpha - kB) \), then \( b_1 = \varepsilon_0 \) or \( b_1 = \infty \), and \( a_1 = 0 \) or \( a_1 = \varepsilon_1 > b_1 \). We select \( A \) out of four possibilities:

\[
A_1: (0, \varepsilon_0), \emptyset, (\varepsilon_1, \infty), (0, \infty).
\]

From (23) we derive the following inequalities

\[
S(A_1) \geq S(A_2) = 0 > S(A_3),
\]

where

\[
S(A_3) = \int_{\varepsilon_1}^{\infty} \left[ \beta \left( 1 - F_{-}(u) \right) - (\alpha - B)F_{-}^{r}(u) \right] du.
\]

Comparing \( S(A_3) \) with

\[
S(A_4) = \int_{\varepsilon_1}^{\infty} \left[ \beta \left( 1 - F_{-}(u) \right) - (\alpha - B)F_{-}^{r}(u) \right] du - B\varepsilon,
\]

we find that \( A_4 = (0, \infty) \) is optimal; then

\[
B\varepsilon \geq \int_{\varepsilon_1}^{\infty} \left[ \beta \left( 1 - F_{-}(u) \right) - (\alpha - B)F_{-}^{r}(u) \right] du.
\]

\( A_4 \) starts from \( a_1 = 0 \); hence \( \varepsilon = \pi(t_i) \). If \( \pi(t_i) = 0 \), inequality \( (B9) \) is contradictory. Therefore \( A_4 = (0, \infty) \) is optimal only if \( \pi(t_i) = 1 \) and \( B\varepsilon \) is true with \( \varepsilon = 1 \). Otherwise, the alert \( A_5 \) is optimal. The analysis of examples 1 and 3 in section 3.3 is similar.

**Appendix C: Proof of Statement 3**

Let \( B_i = 0 \); alarms can be called in arbitrary order. According to (A1) and (A2) conditional average losses per a time interval \( (t, t + \Delta) \), given past information, are

\[
\delta S(t) = \sum_{i=1}^{p} E[\sigma_i = i] (\beta_i - \alpha_i r_i) \Delta.
\]

Contrary to the case \( B \neq 0 \), \( \delta S(t) \) in (C1) depends only on decision \( \pi_t \) and does not depend on \( \pi_{t-1} \). Therefore, optimizing (C1), we obtain the globally optimal decision, since for any function \( \phi_i \)

\[
\min \{ \pi_i \} \sum_{\pi_i} \phi_i(\pi_i) = \sum_{\pi_i} \min \phi_i(\pi_i).
\]

Previously, we optimised the functions \( \sum \phi_i(\pi_i, \pi_{i-1}) \) for which (C2) is not satisfied. Let us take

\[
[\pi_i = i] = q_i(\omega), \\
\beta_i - \alpha_i r_i = f_i(\omega).
\]

It is clear that

\[
\sum q_i(\omega) \leq 1, \quad q_i(\omega) \geq 0;
\]

and

\[
\sum q_i(\omega)f_i(\omega) \geq \min[f_1(\omega), \ldots, f_p(\omega), 0].
\]

Let \( \Omega_1 = \{ \omega : f_i(\omega) = \min[f_1(\omega), \ldots, f_p(\omega), 0] \} \), and \( \chi_i(\omega) \) is the indicator of the set \( \Omega_i \). Then inequality (C4) reaches equality if \( q_i(\omega) = \chi_i(\omega) \), meaning that (C4) reaches the minimum if \( q_i(\omega) = \chi_i(\omega) \), i.e.,

\[
\pi_t(\omega) = i, \omega \in \Omega_i.
\]

We find the solutions of (C6) through examination of Figure 4. Let \( \pi_t(\omega) = r, f_i(\omega) = \beta_i - \alpha_i r_i \), function

\[
y(r) = \min(\beta_i - \alpha_i r_i, \ldots, \beta_p - \alpha_p r_i), 0
\]

is a boundary of polygon

\[
U = \{ (r, y) : \gamma > 0, \rho \leq 0, \beta_i - \alpha_i r_i \leq \rho, i = 1, \ldots, p \}
\]

in plane \( (r, y) \). Set \( \Omega_i \) corresponds to that linear part of the boundary of \( U \) which is formed by function \( y = \beta_i - \alpha_i r_i \). Projection of this boundary part on axis \( r(t) \) gives us an interval \( \Delta_i \). Therefore, the \( i \)-th alarm is declared if \( r(t) \subset \Delta_i \).

**Acknowledgments.** This work has been partly performed under the auspices of the Environmental Protection Agreement 02.09-18 between the United States and the former USSR and was supported in part by grant EAR 88-04883 of the National Science Foundation and in part by a grant from the National Science Foundation through Cooperative Agreement EAR-890126 and USGS Cooperative Agreement 14-08-0001-A0890 to the Southern California Earthquake Center (SCEC). One of the authors (G.M.M.) wishes to thank the U.S. Geological Survey and UCLA for their hospitality during his visit to the United States. The authors thank V. I. Keilis-Borok (Russian Academy of Sciences) and D. D. Jackson (UCLA) for valuable discussions. We are grateful to the associate editor, P. Spudich, reviewer P. A. Reasenberg (USGS), and especially to an anonymous referee for their useful remarks which significantly improved the manuscript. Publication 3648, Institute of Geophysics and Planetary Physics, University of California, Los Angeles. Publication 4, SCEC.

**References**


G. M. Molchan, International Institute of Earthquake Prediction Theory, Varshavskoe shosse 79/2, Russian Academy of Science, Moscow 113556, Russia.

(Received June 11, 1990; revised May 15, 1991; accepted December 10, 1991.)