INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

H4.SMR/303 - 19

WORKSHOP
GLOBAL GEOPHYSICAL INFORMATICS WITH APPLICATIONS TO RESEARCH IN EARTHQUAKE PREDICTIONS AND REDUCTION OF SEISMIC RISK

(15 November - 16 December 1988)

PREPRINT

DIAGNOSIS OF TIME OF INCREASED PROBABILITY OF STRONG EARTHQUAKES IN DIFFERENT REGIONS OF THE WORLD: ALGORTHM CN

V.I. KEILIS-BOROK & I.M. ROTWAIN

Institute of Physics of the Earth
Academy of Sciences of the U.S.S.R.
Bolshaya Gruzinskaya 10
123 242 Moscow
U.S.S.R.

V. I. Keilis-Borok, I. M. Rotwain.

Diagnosis of Time of Increased Probability of strong earthquakes in different regions of the world; algorithm CN.

Introduction.

This paper summarises the studies, directed at diagnosis of the Time of Increased Probability of a strong earthquake (TIP). Strong earthquake is defined by condition $M \geq M_0$, M being the magnitude, M_0 a numerical threshold. Diagnosis is based on the traits of earthquakes flow in wider magnitude range.

General background of the diagnosis is described in the first paper of this volume (V. I. Keilis-Borok). Let us remind, that the following traits are considered: the intensity of earthquakes flow (level of seismic activity); its temporal variation; clustering of earthquakes in space and time; their concentration in space; their long-range interaction.

These traits are represented by some functions of time defined on the sequence of earthquakes in a territory within a sliding time window.

The aftershocks are eliminated, so that the traits wouldn't be dominated by relatively strong earthquakes. The number of aftershocks is included as one of the parameters of a main shock.

Different functions may correspond to different traits or to the same trait but with different values of numerical parameters (range of magnitudes, length of time window, etc.). At each moment

of time the earthquakes' flow within the considered territory is represented by the vector of values of functions.

The problem of earthquakes prediction can be formulated now as follows: the values of functions at a given moment of time t are known; to find out whether the time interval $(t, t+\tau)$ belongs to a Time of Increased Probability of a main shock or a foreshock with $M \geq M_0$; here τ is a numerical parameter.

The functions are normalized so that their definition can be applied uniformly to the territories of different size and seismicity. The possibility of such normalization is of obvious theoretical interest in connection with the problem of selfsimilarity of earthquakes' flow. On more practical side such normalization, if successful, will allow the uniform analysis of different territories. This is of crucial importance, as the only possibility to accumulate a necessary data base, since just a few strong earthquakes occurred in each single territory during the last decades, when earthquake catalogs became sufficiently complete for our analysis.

Normalization is achieved by the choice of the magnitude range, $\underline{M} \leq M \leq \overline{M}$, for the earthquakes considered.

In definition of some functions earthquakes are counted with equal weight, independently on their magnitude. To normalize these functions, we define $\underline{M}$ by the following condition: the average annual number of earthquakes with $M \geq \underline{M}$ on the territory considered is equal to a constant $n(\underline{M})$, common for all territories. In this way the intensities of earthquakes flows are equalized. This normalization is illustrated on fig. 1.

In determination of other functions the earthquakes are counted with the weight, depending on their magnitude. To normalize these functions, we define magnitude thresholds as (M_0 - c), c being yet another numerical parameter.

Integral traits of the earthquakes flow ("functions").

We consider a sequence of main shocks and foreshocks on a certain territory. Let t_i, M_i - be the time and magnitude of the main shock with the sequence number i, $t_{i+1} > t_i$. Integral traits of earthquakes' flow are defined on a sliding time-window $t-s \le t_i \le t$. The following functions, representing these traits, are considered.

Level of seismic activity.

$$N(t|\underline{M}, s) \quad \ldots\ldots (1)$$

- the number of main shocks with magnitude $M \ge \underline{M}$.

$$\Sigma(t|\underline{M}, \overline{M}, s, \alpha, \beta) = \sum_i 10^{\beta(M_i - \alpha)} \quad \ldots\ldots (2)$$

- the number of main shocks weighted according to M_i. Summation is made on main shocks with $\underline{M} \le M_i \le \overline{M}$.

The value of β is determined by one of the two conditions: that the weight of an earthquake is roughly proportional to the area of the rupture in the source or to its linear dimension. If the energy of earthquake depends on magnitude as $\lg E = A+BM$, these condition imply $\beta = 2B/3$ or $\beta = B/3$ respectively.

For large magnitudes, close to M_0 the energy increase is due

mainly to the increase of the length of the the source (M. Caputo et al, 1974), so that $\beta = B/3$ could be more adequate in both cases. This consideration is ignored here, since the whole analysis is intentionally robust.

$$G(t|M_1, M_2) = 1 - N(t|M_2, s)/N(t|M_1, s); \quad M_1 < M_2 \ldots (3)$$

- the ratio of the number of main shocks in two magnitude ranges (M_1, M_2) and $M \ge M_1$.

Quiescence.

$$q(t|\underline{M}, s, u) = \int \mathrm{POS}[n(\underline{M}) \cdot s - N(t|\underline{M}, s)]\,dt \quad \ldots\ldots (4)$$

- "deficiency" of activity. Here POS indicates that the integral is taken only over positive values of the integrand. Among different kinds of quiescence this function may identify some seismic gaps.

A second measure of quiescence is also considered. Roughly speaking, it is the depth of the last minima of N(t):

$$Q(t|\underline{M}, s) = [N(t_1|\underline{M}, s) - N(t_2|\underline{M}, s)] + [N(t|\underline{M}, s) - N(t_2|\underline{M}, s)] \ldots (5)$$

In definition of function Q only last 15 years are considered, i.e. the time interval from (t-15 years to t). Here t_1 is the time of most recent maximum of the function N(t) and t_2 is the time of a minimum of N(t) between t_1 and t. If there is no intervening minimun in N(t) between t_1 and t, we set $t_2 = t$. If there is no maximum in N(t) within the interval considered, we set $t_1 = t_2$.

Temporal variation of seismicity.

$$L(t|\underline{M}, s) = N(t|\underline{M}, t-t_0) - N(t-s|\underline{M}, t-s-t_0)(t-t_0)/(t-s-t_0) \ldots \quad (6)$$

- deviation from long-term trend. In this case t_0 is the beginning of the catalog. The second term is linear extrapolation of N from (t - s) to t.

$$K(t|\underline{M}, s) = N(t|\underline{M}, s) - N(t-s|\underline{M}, s) \ldots \quad (7)$$

- difference between the number of main shocks at two successive time-intervals (t-s, t) and (t-2s, t-s).

Spatial concentration.

$$Smax(t|\underline{M}, \bar{M}, s, u, \alpha, \beta) = \max[\Sigma(t|\underline{M}, \bar{M}, s, \alpha, \beta)/(N(t|\underline{M}, s) - N(t|\bar{M}, s)] \ldots \quad (8)$$

We take $\beta = 2B/3$, so that the expression in brackets is roughly proportional to the average area of rupture in the source. Maxima of this area within last 3 years is considered.

$$Zmax(t|\underline{M}, \bar{M}, s, u, \alpha, \beta) = \max[\Sigma(t|\underline{M}, \bar{M}, s, \alpha, \beta)/(N(t|\underline{M}, s) - N(t|\bar{M}, s)^{2/3}] \quad (9)$$

We take $\beta = B/3$, so that the expression in brackets is roughly proportional to the average linear dimension of the rupture in the sources. This measure is introduced, because in many laboratory experiments on the rocks fracturing the macrofailure occurrs, when such ratio exceeds a rather universal threshold (S. N. Zhurkov et al, 1978). Maxima of this area within last 3 years is considered.

Clustering of earthquakes:

$$Bmax(t|\underline{M}, s, M_a, e) = \max b_i(M_a, e) \ldots \quad (10)$$

Here $b_i(M_a, e)$ is the number of aftershocks with magnitude $M \geq M_a$ within the period e after the main shock; the maximum value of b_i is taken for the main shocks with $M \geq \underline{M}$ from the time interval (t-s, t).

Many other measures of premonitory clustering are described in the literature; we choose the number of aftershocks b_i, because to our knowledge it is so far the only earthquake's precursor, for which statistical significance is established (G. M. Molchan et al, 1988).

Spatial contrast of activity.

It is represented by simultaneous quiescence and activation in the adjacent territories. Quiescence is diagnosed if $N(t|\underline{M}, s) \leq Nq$ and activation - if $N(t|\underline{M}, s) \geq Na$.

The thresholds Nq and Na are defined by condition that p% of time $N(t|\underline{M}, s) \leq Nq$ and p% of time $N(t|\underline{M}, s) \geq Na$. In other words, these are the p% and (100-p)% quantiles of $N(t|\underline{M}, s)$. Spatial contrast of activity is measured as

$$T_{aq}(t|\underline{M}, s, p) \ldots \quad (11)$$

- the time since the territory under study and some adjacent territory were in opposite states for more than a year.

Long-range interaction of earthquakes.

It is characterized by two phenomena. One of them is suggested by A. G. Prosorov (1982): a strong earthquake of sufficiently large magnitude is followed by some activation of the area, where another strong earthquakes is due. This activation is measured by the magnitude of the earthquakes in this area. Accordingly the earthquakes which occurr within u years after an earthquake with $M \geq M_0$ are defined as long-range aftershocks.

Their maximal magnitude is the measure of a long range interaction:

$$Ml(t|s, M_0, u) = \max M_i \qquad (12)$$

Maximum is taken within the time interval (t-s, t).

Another phenomenon of the long-range ineraction is a simultaneous activation of several territories belonging to the same fault zone or to the same morphostructure of higher rank. Accordingly we consider two functions:

$$NF(t|\underline{M}, s) \qquad (13)$$

- the value of $N(t|M, s)$ in the whole fault zone, which includes the territory under consideration.

$$NR(t|\underline{M}, s) \qquad (14)$$

- the value of $N(t|\underline{M}, s)$ for the morphostructure of a higher rank

which includes the territory under consideretion.

The values of these numerical parameters involved in definition of the functions, are indicated in Table 1.

Formulation of Problem in Terms of Pattern Recognition.

Let us consider a territory divided into regions. An earthquakes' flow in a region K may be represented by the vector

$$P(K, t) = \{p_1(K, t), \ldots, p_m(K, t)\},$$

where p(K, t) is one of the functions listed above. Our problem, again, is the following: knowing P(K, t) to recognize a TIP, that, wether the moment t belongs or das not belong to a TIP, that is whether the probability of a strong ($M \geq M_0$) earthquake in the territory K and in the time interval (t, t+τ) has increased. This is a typical problem of pattern recognition, in our case, of pattern recognition of small samples. We shall use the specific approach to such problems described by I. M. Gelfand et al (1976).

Let us divide the time period into intervals of 2 types (Fig. 2):

D: t_1 years before each strong earthquake;

N: all the remaining time .

The formulation of our problem implies the following hypothesis:

Within some parts of the intervals D the approach of a strong earthquake is expressed in anomalous character of earthquakes' flow; in other words, intervals D overlap with actual TIPs.

Object of recognition is a combination 'region-time' (K, t), described by vector P(K, t).

The rule for the recognition of the TIPs is derived by analysis of the "learning material" that is the samples of the objects D and N.

Our learning material may be mixed; intervals D may include the moments, when the approach of a strong earthquake is not expressed in the vector P(t), i.e. in the traits of the earthquake flow considered here. In other words we know only that at least some moments within D belong to TIPs, but we do not know - which are these moments. To deal with such mixed learning material we use the algorithm of pattern recognition "subclasses"; it was specially designed for similar situation, ecountered in pattern recognition of earthquake-prone areas (I.M.Gelfand et al, 1976).

First t_2 years after each strong earthquake were not used for learning, unless they belong to D; on Fig.2 these years are marked as X. This is done because strong earthquakes may be followed by general activation, not entirely suppressed by elimination of aftershocks. Such activation may be not typical for intervals N and obscure their difference from D.

After the recognition rule is estabilished, however, we will try it for all moments of time, with sufficiently small step. It was never tested, wether elimination of interval X from learning is actually necessary.

Discretization.

According to the algorithm used here the vectors P(K,t) should be represented as a sufficiently short binary vectors. For this purpose the range of values of each function is divided into three intervals, "small", "medium" and "large", or just into two interval, "small" and "large". Then the vector P(K,t) will specify only these intervals for each function but not its more precise value. This discretization leads to the loss of information, hopefuly justified

by the increase of stability and reliability of recognition, as is often the case in the robust exploratory data analysis (J.W.Tukey, 1977).

The thresholds for discretization are determined in such a way, that each interval contains approximately equal number of objects from the learning material. The important advantage of such discretization is a gross reduction of data-fitting. Specifically the discretization is independent on a priory knowledge on what objects in learning material correspond to the eve of a strong earthquake.

Diagnosis of TIPs.

The algorithm for the diagnosis of TIPs, including the values of numerical parameters, is formulated by retrospective analysis of the data on the earthquake country in California and Nevada. It was divided into two regions, shown on Fig.3. The catalogs CIT and NEIS-USGS were applied for the Southern and Northern regions respectively. These catalogs allow to calculate most of the functions for the period from 1938 up to 1985.

Learning material consisted of the following:

Objects D: 15 subclasses, corresponding to two-year periods before each strong earthquake from Table 2. Each subclass included 3 moments of time within these intervals. A common subclass corresponds to earthquakes NN 3 and 5 in Table 2, since they occured in the same area within a mounth.

Objects N: 24 moments from period N, not within 3 years after each strong earthquake.

Aftershocks were identified by the algorithm described by V. I. Keilis-Borok et al (1980) within time interval $T(M_1)$ after a main shock and distance $R(M_1)$ from it. The values of T(M) are given in Table 3. For R(M) we assumed a fixed value 50 km, though more refined definition is available. Strong earthquakes are defined by the threshold M_0 = 6.4. Their list, after elimination of aftershocks, is given in Table 2.

Resolving power of single functions.

According to our procedure of pattern recognition, we estimated for each function which of its values - large, medium or small - are more typical for D and which for N. For this purpose one-dimensional distributions of the values of each function were compared as shown in Table 4. We see, that for objects D the relatively large values of N2, K, Σ, Smax, Zmax, q and Bmax are typical and as well as small values of functions G and N3. However, none of the functions alone is sufficient.

For other functions considered one-dimentional distributions for objects D and N are much closer. These functions were not used in further learning procedure; nor are they shown in Table 4.

Next stage of learning consist of definition of characteristic features of the objects D and N, i.e. of TIPs and non-TIPs for a given region. Below they are called just "features". Features D are defined by condition, that they are sufficienly often encountered in objects D and sufficienly seldom in objects N. Features N are defined by reversed condition. Exact definition of characteristic features is given in I.M.Gelfand et al (1976). These features are given in Table 5 and 6. Each feature indicates a discretized value of a function or a combination of such values for 2 or 3 functions. For example, first feature in Table 5 means, that the value of K is medium or large and value Bmax is large.

Diagnosis of TIPs ("voting").

Each region at a given moment ("each object") has some number $n_D(t)$ of features D and some number $n_N(t)$ of features N. Recognition by algorithms, such as used here, is usually based on the difference $\Delta(t) = n_D(t) - n_N(t)$. This difference was computed for each region with the time step 2 months.

Within a year before each but one strong earthquake there are moments when $\Delta(t) \geq 5$. Exception is the earthquake of 9.2.1941, M = 6.6 in the Northern region.

Accordingly, we may hypothethize the following rule for diagnosis of TIPs: it starts at each moment when $\Delta(t) \geq 5$ and lasts from this moment for 1 year or until a strong earthquake, whichever comes first. Similar rule was assumed in V.I.Keilis-Borok et al(1980) for separate precursors. However, the TIPs, thus diagnosed, would cover 63% of time, which is too much. There would be 9 false alarms. To improve the success-to-failure, score, we may notice, that many false alarms are due to the increase of activity after some strong earthquakes. This increase happened to be not suppressed by elimination of aftershocks. In particular 5 out of 9 false alarms can be distinguished by the large total area of rupture in the sources. We assume for the measure of this area the function: $\Sigma(t|M, s, \alpha, \beta)$, defined by (2), with $\beta = 1$; $\alpha = 5$; $s = 3$; $\underline{M} = M_0 - 1.4$ and, obviously, with no upper limit M. Let us normalize Σ by the area, unlocked during a single strong earthquake. The normalized measure is

$\sigma(t) = 10^{-\beta(Mo - \alpha)} \Sigma(t | \underline{M}, s, \alpha, \beta)$

5 false alarms, mentioned above will be eliminated, if we declare TIPs only when $\sigma < 4.9$. This imposes a limitation on the areas of the fault system, unlocked by main shocks during the last 3 years. Summarily these areas should be less than an area, unlocked by a single earthquake with $M \geq M_0 + 0.7$.

Finally the rule for the diagnostics of the TIPs can be formulated in the following way:

A TIP is announced at a moment t for one year under two conditions:

1. For the features from Tables 5 and 6

$\Delta(t) = n_P - n_N \geq \overline{\Delta}$, *with* $\overline{\Delta} = 5$;

2. $\sigma(t) < \overline{\Sigma}$, *with* $\overline{\Sigma} = 4.9$

Successive TIPs can overlap and extend each other.

A TIP during which a strong earthquake hasn't occurred, is counted, as a false alarm; a strong earthquake, not preceded by a TIP is called a failure-to-predict.

TIPs thus diagnozed are juxtaposed with strong earhquakes on Fig. 4 and in Table 7. TIPs occupy 31% of total time-space. They preceed 13 out of 15 strong earthquakes, with 2 false alarms. Such a good score is predetermined by the fact that all strong earthquakes were allowed for in the learning and in selection of thresholds $\overline{\Delta}$, τ and $\overline{\Sigma}$.

Therefore this rule is mere a hypothesis, which is admittedly obtained by retrospective data-fitting and has to be tested on independent data. We have to test also the stability of this rule to variation of decision, made in the process of data-fitting.

4. Test on continuation of catalog, 1984 - April 1988.

The rule for diagnosis of TIPs, formulated above, is based on the analysis of the earthquake's catalog up to 1983 inclusive. This rule was applied to subsequent seismic history, for which the catalogs were available, i.e. for the period Jan. 1984 - April 1988. The results are also shown on Fig. 4. In Northern region the earthquakes with $M \geq 6.4$ are not reported for this period and no TIPs are diagnosed; exeption is the TIP which starts at May 1986 and so far constitutes a forward diagnosed, yet to be tested by future seismicity. In Southern region catalog PDE lists 3 earthquakes with $M \geq 6.4$, while catalog CIT assigns to them smaller magnitudes. Following is the lists of those earthquakes:

NN	Data	φ, ° N	λ, ° W	Magnitude CIT	Magnitude PDE
1.	1983. 5. 2	36.23	120.26	6,3	6,7
2.	1986. 7.21	37.63	118.43	5,9	6,5
3.	1987.11.24	33.08	115.78	5,8	6,5
4.	1987.11.24	33.01	115.84	6,0	6,7

No such differences in magnitudes of strong earthquakes occur before the first earthquake in this list.

The TIPs diagnosed are also shown on Fig. 4. They preceed each of the 3 earthquakes, listed above (earthquakes 3 and 4 are regarded, as a single event). Average duration of TIP per earthquake is 17 months; which is about the same, as for the learning period.

These results seem positive, in spite of uncerntainty in magnitudes. In further monitoritg it seems preferable to identity strong earthquakes by maximal magnitude among PDE and CIT versions, similar to what was done in studies of separate premonitory patterns (Keilis-Borok et al, 1980).

5. Test on independent data.

The above algorithm was applied to 10 regions of the world listed in Table 8.

The following general ground rules were assumed.

-- TIPs were diagnozed by the conditions (1), (2) with characteristic features from Tables 5 and 6; and with numerical parameters from Table 1.

-- Seismic territories are considered within generally accepted boundaries (see, for instance "Earthquakes in USSR, 1965-1986"). They are drawn according to seismotectonics; we tried to minimize, but coud not avoid, the intersection of boundaries with major lineaments and hence with the clouds of epicenters.

-- The threshold M_0 was chosen in such a way, that the average interval between strong earthquakes is 5 - 7 years, as was the case for California and Nevada. The choice of M_0 was not formalized; this will constitute a significant difficulty in statistical evaluation of results.

-- The catalog of main shocks is obtained by algorithm of identification of aftershocks, with the same numerical thresholds as for California and Nevada.

-- The magnitude thresholds $\underline{M}$, $\overline{M}$, are defined by the rules of normalization of the functions indicated above.

-- Time-scale and, therefore, numerical parameters measured in the units of time were not changed.

In general, these rules provide for practically unambigous change of the parameters in transition to a new territory.

However, the experiment is not strict due to some freedom in the choice of M_0 and in of the boundaries of the regions.

Some particularitie of data processing.

They have to be indicated in order to make our results reproducible.

-- The catalogs Earthquakes in USSR (1965-1986) and Earthquakes in Central Asia and Kasachstan (1979-1986) indicate instead of magnitudes, "energy class K" which is an estimation of energy E of seismic waves, $K = \lg E$, jowls. Magnitudes are given for most of strong earthquakes. If magnitude is not indicated, it was calculated by the formula: $M = 0,55(K - 4)$ (B. Gutenberg and C. F. Richter, 1949).

-- The catalog NEIS-USGS contains the magnitude m_b for most of the earthquakes since 1964, whereas for strong earthquakes the magnitude is given in M_S scale. We recomputed m_b into M_S by formula: $M_S = m_b - 5,34$.

-- The catalog for N. Appalachians (NEIS-USGS) is different before and since 1975. In the second part, 1975-1986, the magnitudes are indicated for most of the earthquakes. The annual magnitude-frequency-of-occurrence statistics since this year suggest that this catalog is acceptably complete for $M \geq 3.3$. In the first part, 1960 - 1974, the magnitudes are indicated only for few strong earthquakes. However, the annual number of earthquakes will be levelled for the whole period 1960-1986, if we assume, that unidentified magnitudes in the first part correspond mainly to the same threshold $M \geq 3.3$. Analysis was done under this assumption. For more details see V. I. Keilis-Borok and I. M. Rotwain (1989a).

-- Similar situation was encountered for two parts of the catalog ENEL for Central Italy, 1950 - 1970 and 1971 - 1986. We assumed that unindentified magnitudes in the first part correspond mainly to $M \geq 3$ in the second part (V. I. Keilis-Borok et al, 1989b).

-- No strong earthquakes occurred during the period investigated in two regions: Trans-Baikal and Southern region of N. Appalachians. Therefore the learning material consists of intervals N only and it would make little sense to discretise the functions in the same way, as for California and Nevada. Therefore we transferred the thresholds of discretization from other regions with the same M_0: from E. Central Asia to Trans-Baikal, and from Northern to Southern region of N. Appalachians. Obviously this leads to no artificial improvement of the result.

The TIPs diagnosed.

They are summarized in Fig. 5. The results seem satisfactory: TIPs occupy on the average about 24% of time in a region and preceed 23 out of 29 strong earthquakes. 17 out of 25 TIPs are followed by a strong earthquake. Particularly important is the fact that the success-to-failures score, obtained with a priori fixed algorithm, is about the same as for California and Nevada, where the algorithm was data-fitted to improve this score.

Stability.

Numerical tests show, that the diagnosis of TIPs is reasonably

stable to variation of the algorithm. Almost the same results are obtained with other values of parameter β in the functions Σ, Smax and Zmax (C. Allen et al, 1983, 1986); these values were assumed the same, as in the previous studies of some separate premonitory seismicity patterns.

Four other regions were also considered with this version of β: the Caucasus, Western Turkmenia, Kamchatka and Kuril islands, but not yet with β accepted here. The results of diagnostics are roughly the same: TIPs preceed 8 out of 11 strong earthquakes and, on the average, occupy 22% of time.

The results are stable also to the change of decisions which are made not in only possible way: choice of the time-period on which M is defined; variation of discretization thresholds; selection of learning material. Illustration of such stability can be found in C. Allen et al (1986). The diagnostics is most sensitive to the boundaries of a region: it has to be tectonically meaningful and not too small).

Second stage of diagnosis: smaller territories.

Diagnostics discribed above has a considerable drawback: a TIP is announced for rather a large region, of linear dimensions by the order of magnitude larger than the predicted source. Thus a problem remains - to localize the prediction, i.e. to assign a TIP to a smaller territory if possible. The approach to this problem here is based on the hypothesis that the process, which leads to a strong earthquake is hierarchichal: it is successively concentrated in the structures of a lower rank. Accordingly by Southern and

Northern regions were subdivided, each into 3 areas shown in the Fig. 3. This division, suggested by C. Allen (C. Allen et al, 1986), is based on tectonics, historical earthquakes and focal mechanisms. The boundaries were drawn so that the areas have common seismotectonic traits. The major active faults are situated within the area and not along their boundaries, countrary to usual morphostructural zonation.

The problem for areas was treated in the same way as for the regions. In each of the areas the flow of main shocks is considered during the TIPs, diagnosed for the whole region. The combination (area, time) is attributed to D, if a strong earthquake occurred in that area, otherwise it is attributed to N (Fig. 2. b). The learning material for D comprised 14 subclasses. For N we used 32 objects, including the cases, where strong earthquake occurred in another area and also the false alarms for a whole region.

The same functions, as for the regions, were considered. According to definition of the functions, they were normalized by seismicity of the areas. That automatically lowered the value of M so that in absolute, unnormalizad scale weaker earthquakes are included into analysis. For some functions M was additionaly lowered to improve the resolution of one-dimensional distributions. All numerical parameters, assumed for the areas, are indicated in Table 9.

Single traits.

One-dimensional distributions of each function for D and N are compared in Table 9. We see, that in contrast to regions, the approach of a strong earthquake within the area is indicated by small values of functions K, Zmax and by large of N3. On the whole,

quite different sets of functions distinguish D from N, that is the TIPs from non-TIPs in regions and areas. Only 4 out of 9 functions are in common.

The features of D and N for areas were obtained by the same algorithm "subclasses". They are given in Tables 10 and 11. In the diagnosis of TIPs it seemed natural to decrease the thresholds: the duration τ of a single TIP, and the area Σ of unlocked part of the fault. We divided these thresholds by the number of areas in a region, which is 3 in our case. The voting margin was selected as $\Delta = 1$. Finally the following rule of diagnosis is formulated:

A TIP in an area is announced for 4 months if:

1. TIP is announced for the region, which includes this area;
2. For the featurs from Table 10 and 11

$\Delta(t) = n_1 - n_2 \geq \bar{\Delta}$, with $\bar{\Delta} = 1$

3. $\sigma(t) < \bar{\Sigma}$, with $\bar{\Sigma} = 1.6$.

The diagnosis of TIPs by this rule can be summarized, as follows: the total space-time volume of TIPs was decreased to 14% with 2 additional failures-to-predict: TIP is not diagnosed before earthquakes of 21.10.1942 and 9.2.1956 in Imperial Valley. These earthquakes were missed due to condition 3: too large is the total area of the faults unlocked. In case of the earthquake of 1942 it was due to preceding strong earthquake. If condition 3 is abandoned, the total time of TIPs will rise up to 23%.

The total value of TIPs could be considered satisfactory. However, TIPs in several areas often occurr simultaneously and/or pass from one area to another. This reduces the practical value of diagnosis. The diagnosis for the areas was not tested yet on independent data and the success-to-failure score should better be improved before such test. Therefore the diagnosis for the areas remains hypothetical so far.

Conclusions.

1. The results of this study support the assertion, that the traits of an earthquakes flow, formalized as a set of functions introduced here, reflect the approach of a strong earthquake.

2. There are indications, that the process, which generates the strong earthquakes is similar, in rather a wide variety of neotectonic environments and of energy ranges. It confirms the concept of academician M. A. Sadovsky (1987) on global selfsimilarity on the dynamics of lithosphere, including seismicity. This seems consequential for the development of the model of earthquakes' occurrence. Also it allows to consolidate the observational base for further studies of earthquakes.

3. The suggested criteria for diagnosis of TIPs could be qualitatively described in the following way: 2 - 3 years before a strong earthquake an activation, broadly defined, occurrs in a region, i. e. in a morphostructure of linear dimensions by the order of magnitude greater than the source of the upcoming earthquake. This activation, is expressed in the large values of N2, Zmax and q; and in large values of K, indicating that activity keeps growing. The activation is preceded by relative quiescens expressed in large q and accompanied by an increase of clustering expressed in large Bmax. The K, Zmax and Bmax functions are most important for doagnosis.

4. Hypothetical, yet untested, continuation of this description is the following: After such activation has been developed and TIP has been diagnosed in a region some specific phenomena are localized within the area, i. e. morphosructure of a small rank, where a strong earthquake will occurr. The activity lags behind the activity of the whole region; this is expressed by small K and Q; If the activity is high, i. e. if it is large, it is due to relatively weak shocks. These phenomena are developed in smaller magnitude range, corresponding to the increase of n(M) up to 6.

5. The algorithm CN is presented here as an empirical regularity, imposing certain liminations on the process of earthquakes' occurrence. We do not claim, that this algorithm is the only possible, nor that it is optimal. Quite the contrary, our impression is that its formulation possibly can be simplified. The fact that it works, seem to justify its further development, as well as further monitoring of TIPs.

References

Algorithm CN and related papers

C. Allen, V. I. Keilis-Borok, I. M. Rotwain, K. Hutton. A set of long-term seismological precursors; California and some other regions. In: Mathematical methods of seismology and geodynamics. M., Nauka, 1986, (in Russian).

C. Allen, K. Hutton, V. I. Keilis-Borok, L. Knopoff, V. G. Kosobokov, I. V. Kuznetsov, I. M. Rotwain. Selfsimilar premonitory seismicity patterns. Abstract of the paper, submitted to XVIII Congress of IUGG, Hamburg, 1983.

O. Dmitrieva, I. Rotwain, V. I. Keilis-Borok, M. De Becker. Premonitory seismicity patterns in a platform region (Ardennes - Rhenish and Brabant massives, lower Rhone graben), 1988, (in press).

A. M. Gabrielov, O. E. Dmitrieva, V. I. Keilis-Borok, V. G. Kosobokov, I. V. Kuznetsov, T. A. Levshina, K. M. Mirzoev, G. M. Molchan, S. Kh. Negmatullaev, V. F. Pisarenko, A. G. Prozoroff, W. Rinehart, I. M. Rotwain, P. N. Shebalin, M. G. Shnirman, S. Yu. Schreider. Algorithms of Long-Term Earthquakes' Prediction. International School for Research Oriented to Earthquake Prediction-Algorithms, Software and Data Handling. Lima, Peru, Septembre 1986.

I. M. Gelfand, Sh. A. Guberman, V. I. Keilis-Borok, L. Knopoff, F. Press, I. Ya. Ranzman, I. M. Rotwain, I. M. Sadovsky. Pattern recognition applied to earthquake epicenters in California. Phys. Earth. Plan. Int. 11, 227-283, 1976.

V. I. Keilis-Borok, L. Knopoff, I. M. Rotwain and C. R. Allen. Intermediate-Term Prediction of Times of Occurence of Strong Earthquakes in California and Nevada. 1988, Nature. (in press).

V. I. Keilis-Borok, L. Knopoff, I. M. Rotwain. Bursts of aftershocks, long-term precursors of strong earthquakes. Nature, v. 283, No. 5744, 259-263. January 17, 1980.

V. I. Keilis-Borok, R. Lamoreaux, C. Johnson, B. Minster. Swarms of main shocks in Southern California. Earthquake Prediction. Res. 1982, v. 1, No. 2, 135-151.

V. I. Keilis-Borok and I. M. Rotwain. Diagnosis of increased probability of strong earthquakes in Northern Appalachean. Computational Seismology, 22, 1989, (in press).

V. I. Keilis-Borok, I. V. Kuznetsov, G. Panza, I. M. Rotwain. Diagnosis of increased probability of strong earthquakes in Central Italy. Computational Seismology, 22, 1989, (in press).

Other papers on earthquake prediction.

C. Allen Geological criteria for evaluating seismicity. Bull. Geol. Soc. Amer., 1975, vol. 86, 1041-1057.

M. Caputo, V. I. Keilis-Borok, T. L. Kronrod, G. M. Molchan, G. F. Parza, A. Piva, V. M. Podgaetskaja, D. Postpischl. Models of earthquake occurrence and isoseismals in Italy. Annali di geofisica, XXYI, 2-3, 1974, 421-444.

F. F. Evison Fluctuations of seismicity before major earthquakes Nature, 1977, vol. 266, 718-762.

B. Gutenberg and C. F. Richter. Seismicity of the Earth and Associated Phenomena. New Jersey. Princeton University press. 1949.

G. M. Molchan, O. E. Dmitrieva and I. M. Rotwain. Statistical analysis of the results of long-term prediction, based on bursts of aftershocks. Computational Seismology, 21, 1988, (in Russian).

A. G. Prozorov. Long - range aftershocks, as the earthquake precursors in Southern California. In. - Mathematic models of the Earth's structure and earthquake prediction. M., Nauka, 20-26, 1982, (in Russian).

I. Ya. Ranzman. Location of earthquakes and morphosrycture of mountain countries, M., Nauka, 1979, (in Russian).

T. Rikitaki Earthquake prediction. Amsterdam. Elsevier, 1976.

M. A. Sadovsky, L. G. Bolckovitinov, V. F. Pisarenko. Deformation of geophysical media and seismic process. M., Nauka, 1987. (in Russian)

G. A. Sobolev and A. D. Zavialov. A concentration criterion for seismically active fault. Earthquake Prediction. Ed. D. Simpson, P. Richard, Wash,. AGU, 1981, 337-380.

M. Wyss, R. E. Habermann. Seismic quiescence precursory to a past and a future Kuril Island earthquake. Pure and Appl. geophys., 1979, vol. 117, No. 6, 1195-1221.

S. N. Zhurkov, V. S. Kuksenko, V. A. Petrov, V. I. Saveljev, U. S. Sultanov Concentration criterion of volume-type fracture in solids. Physical processes in earthquake sources. M., Nauka, 1978, 101-116, (in Russian).

Earthquakes' catalogs

Belgique hypocenters data file 1960-1986. Observatoire Royal de Belgique, Bruxelles, 1987.

Catalogue of earthquakes of Italy. Publication ENEL. Roma, 1980.

European-Mediterraneen hypocenters data file 1979-1985. CSEM, Strasbourg, 1986.

N. V. Kondorskaya, N. V. Shebalin, editors. New catalog of earthquakes on territory of the USSR from most ancient times to 1975, M., Nauka, 1977, (in Russian).

Series: Earthquakes in the USSR, 1962-1983, M., Nauka, 1965-1986, (in Russian).

Series: Earthquakes in Central Asia and Kazachstan, 1979-1983, Dushanbe, Dinish, 1981-1985. Computer file extended up to 1988. (in Russian).

Southern California hypocenters data file 1932-1985. CIT-USGS, USA, 1986.

Worlds Hypocenters data file 1932-1985., USGS-NEIS, USA, 1986.

Figures' Captions

Fig. 1 Normalization of earthquakes' flow.
Thresholds M_1 and M_2 provide for the same average annual number of earthquakes n.

Fig. 2 Definition of intervals D and N.
a) - for regions; b) - for areas.

Fig. 3 Regions and areas in California and adjacent parts of Nevada.
Their boundaries are shown by solid and dashed lines, respectively.
1 and 2 - epicenters, 1938 - 1985, $M \geq 4.5$ and $M \geq 6.4$, respectively.

Fig. 4 TIPs and strong earthquakes in California and adjacent parts of Nevada.
1 - moments of strong earthquakes; 2 - TIPs; 3 - current TIPs.

Fig. 5 TIPs and strong earthquakes in other regions.
Notations are the same, as in Fig. 4.

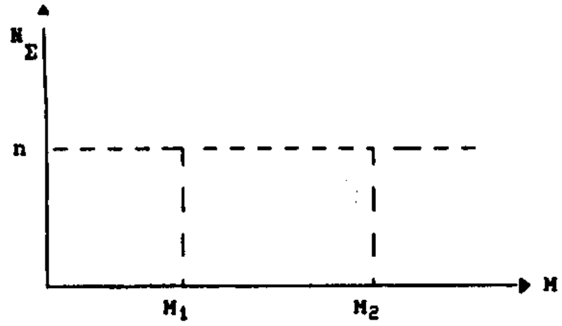


Fig. 1

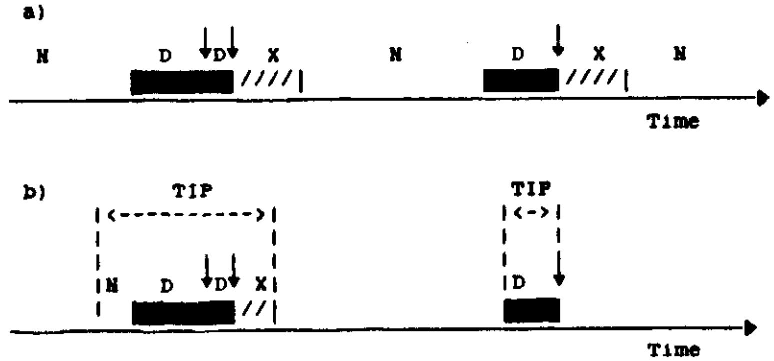
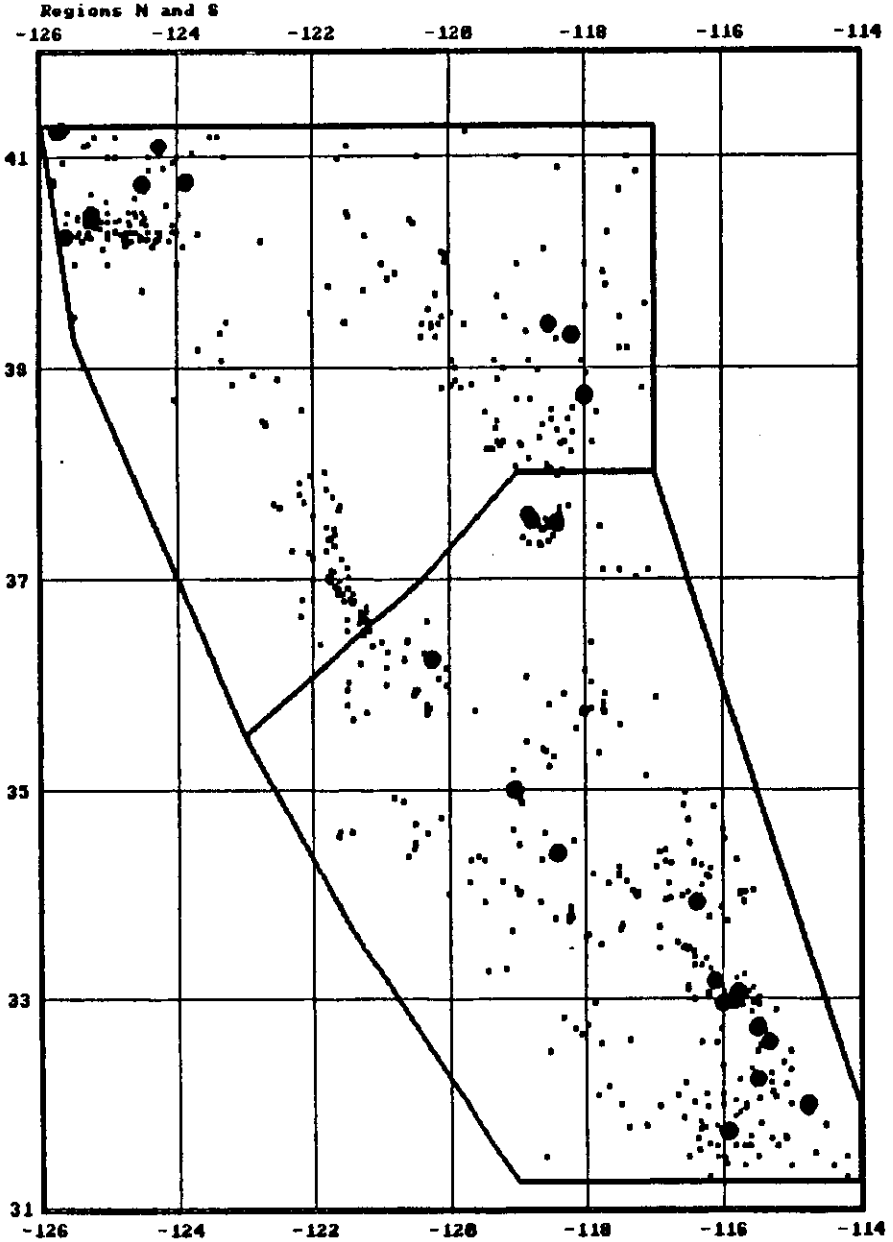


Fig. 2

CN



•-M≥4.5; O-M≥6.4, since 1932

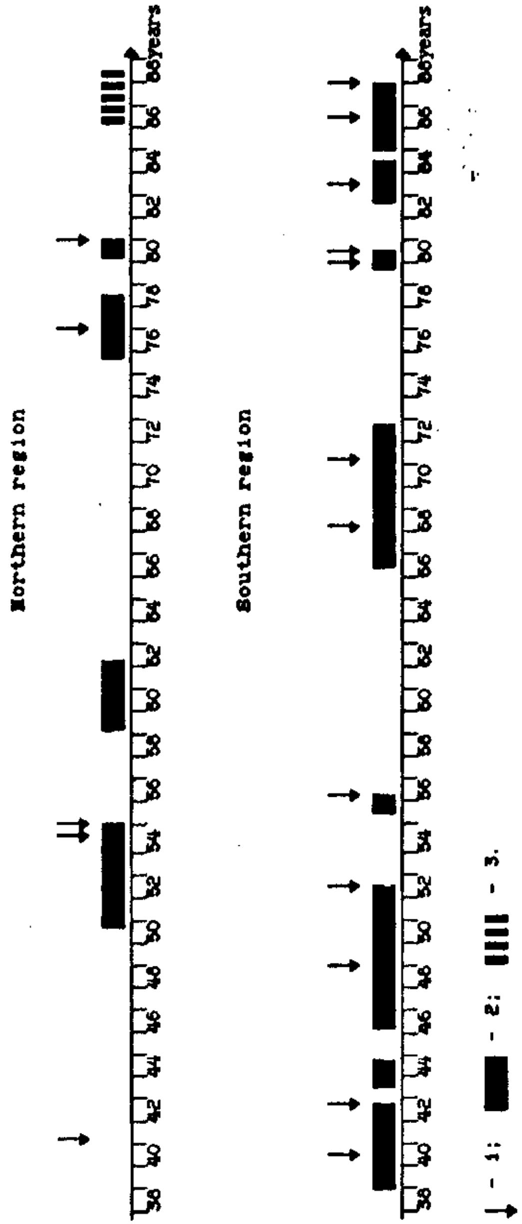


Fig.4

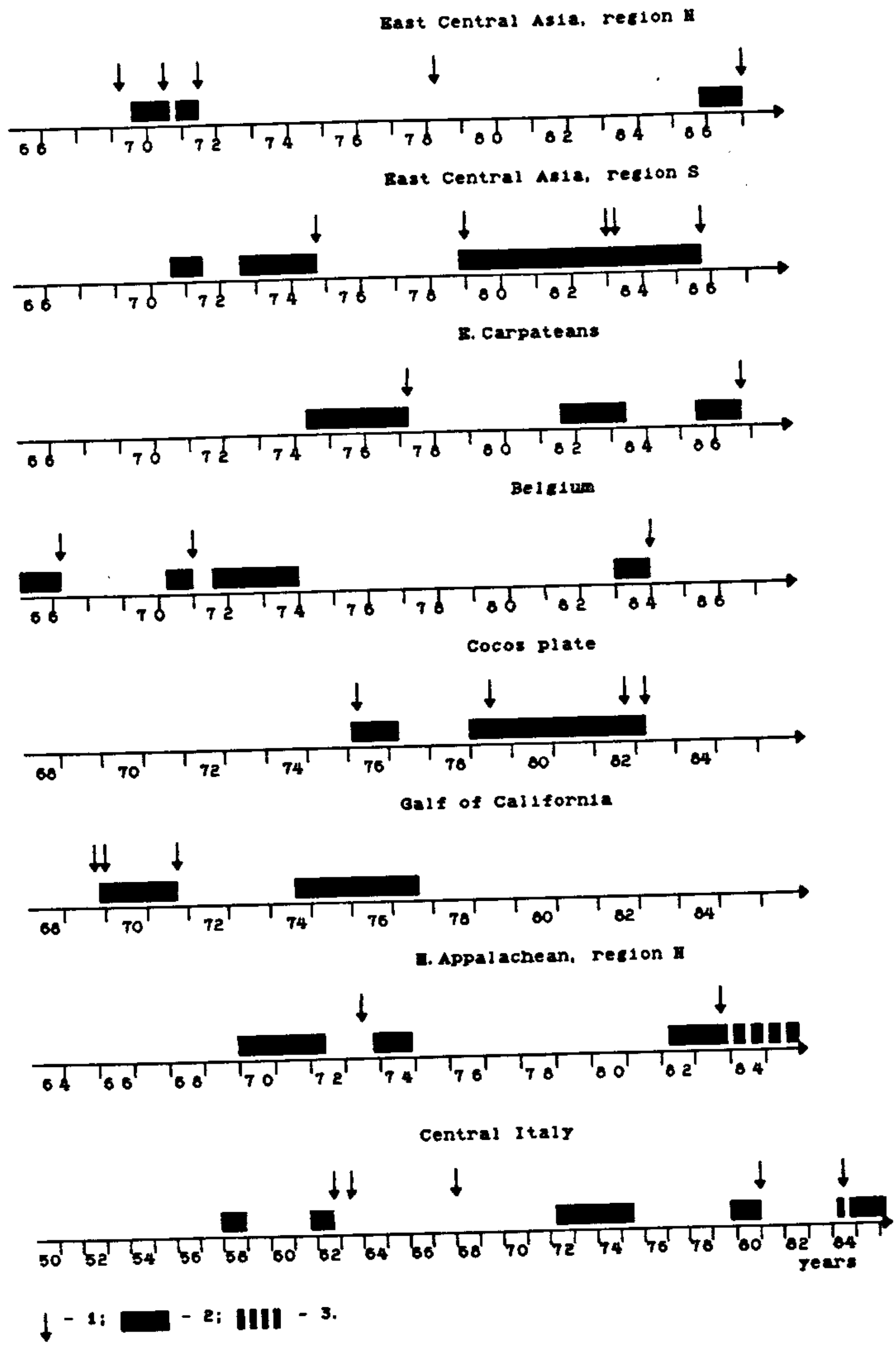

East Central Asia, region N
East Central Asia, region S
E. Carpateans
Belgium
Cocos plate
Galf of California
E. Appalachean, region N
Central Italy
years
- 1; - 2; - 3.
Fig. 5

Table 1

Functions used.

Function*	Def**	n(M) regions	n(M) areas	s, years	Other parameters
1. N1 = N(t|M, s)	(1)	1,4	3	6	
2. N2 = N(t|M, s)	(1)	0,36	1,4	3	
3. K = K(t|M, s)	(7)	1,4	6	2	
4. G = $G(t\|M_1, M_2, s)$	(3)	3; 1,4	3; 1,4	3	
5. Σ = Σ(t|M, M, s, α, β)	(2)	3	6	3	$M = M_o - 0,1$; α = 4,5; β = 1,0
6. Smax = Smax(t|M, M, s, u, α, β)	(8)	3	6	1	u = 3; $M = M_o - 0,1$; α = 4,5; β = 1,0
7. Zmax = Zmax(t|M, M, s, u, α, β)	(9)	3	6	1	u = 3; $M = M_o - 0,1$; α = 4,5; β = 0,5
8. N3 = N(t - u|M, s)	(1)	1,4	1,4	3	u = 7
9. q = q(t|M, s, u)	(4)	1,4	3	6	u = 3
10. L = L(t|M, s)	(6)	3	3	6	
11. Ml = $Ml(t\|s, M_o, u)$	(1)			3	u =1 год
12. Bmax = $Bmax(t\|M, s, M_a, e)$	(10)	-	-	3	M_a = 2,8; e = 48 часов; M = 4,5
13. Q = Q(t|M, s)	(5)	3	6	1	
14. NF1 = NF(t|M, s)	(13)			3	$M = M_o - 1,4$
15. NF2 = NF(t|M, s)	(13)			6	$M = M_o$
16. NR1 = NR(t|M, s)	(14)			3	$M = M_o - 1,4$
17. NR2 = NR(t|M, s)	(14)			6	$M = M_o$
18. Taq = Taq(t|M, s, p)	(11)			1	p = 33%

Notes: * First are given the abbreviated notations, which are used in Tables 4-6 and 9-11.

In paper of V. I. Keilis-Borok et al (1988) other notations are used:

1	2	3	4	5	6	7	8	9	10	11	12	13	16
A1a	A1b	T2	A3	A2	S2	S1	A1c	Q1	T1	I1	C1	Q2	I2.

** "Def" means the number of the definition of a functon in section 2.

Table 2

Strong earthquakes in California and Nevada,

$M \geq 6,4$, 1938 - 1986.

No	Date	φ°, N	λ°, W	M
		Northern region		
1.	9. 2. 1941	40, 5	125, 25	6, 6
2.	6. 7. 1954	39, 42	118, 53	6, 8
3.	25. 11. 1954	40, 27	125, 63	6, 8
4.	16. 12. 1954	39, 32	118, 2	7, 2
5.	21. 12. 1954	40, 78	123, 87	6, 6
6.	26. 11. 1976	41, 29	125, 71	6, 8
7.	8. 11. 1980	41, 11	124, 25	7, 2
		Southern region		
1.	19. 05. 1940	32, 73	115, 5	6, 7
2.	21. 10. 1942	32, 97	116, 0	6, 5
3.	4. 12. 1948	33, 93	116, 38	6, 5
4.	21. 7. 1952	35, 0	119, 02	7, 7
5.	9. 2. 1956	31, 75	115, 92	6, 8
6.	9. 4. 1968	33, 18	116, 12	6, 4
7.	9. 2. 1971	34, 4	118, 4	6, 4
8.	15. 10. 1979	32, 6	115, 32	6, 6
9.	25. 05. 1980	37, 6	118, 82	6, 4
10.	25. 05. 1980	37, 55	118, 78	6, 5

Earthquakes 9 and 10 are regarded, as a single event.

Table 3.

Time-window T(M) for identification of aftershocks.

magnitude	T(M), days
2, 8 - 3, 9	46
4, 0 - 4, 4	91
4, 5 - 5, 4	182
5, 5 - 6, 4	365
> 6, 4	730

Table 4

Functions, used for identification of the TIPs for regions

No in Table 1	Functions	Thresholds for discretization	Frequency of occurrence in learning material, %					
			D			N		
			s	m	l	s	m	l
2.	N2	0	33	-	67	75	-	25
3.	K	-1; 1	13	40	47	50	38	13
4.	G	0,5; 0,67	53	33	13	33	21	46
5.	Σ	36; 71	13	27	60	46	33	21
6.	Smax	7,9; 14,22	6	47	47	50	29	21
7.	Zmax	4,1; 4,6	20	20	60	42	38	21
8.	N3	3; 5	54	31	15	29	25	46
9.	q	0; 12	25	17	58	55	41	5
12.	Bmax	12; 24	20	20	60	50	25	25

s - small; m - middle; l - large value of function.

Table 5

Features D for regions,

$K_1 = 7$, $\tilde{K}_1 = 2$

No	N2	K	G	Σ	Smax	Zmax	N3	q	Bmax
1.		m, l							l
2.								l	
3.							m, l	m, l	m, l
4.						l		m, l	
5.		m, l					s		m, l
6.		s, m				l			m, l
7.		m, l				s, m			m, l
8.		m, l	m, l						m, l
9.					l	l			
10.		s, m		l		l			
11.		m				l			
12.	l	m, s				m, l			
13.		l			s, m				
14.		l			m, l				
15.		l		m, l					
16.		l	s, m						

s - small; m - middle; l - large value of functoin.

A feature of D is encountered in ≥ K_1 subclasses and in ≤ $\tilde{K}_1$ objects N.

Table 6

Features N for regions

$k_2 = 10$, $\tilde{k}_2 = 4$.

No	N2	K	G	Σ	Smax	Zmax	N3	q	Bmax
1.					s, m				s
2.						s, m		s, m	s, m
3.				s, m				s, m	s, m
4.		s, m						s, m	s, m
5.							m, l		s, m
6.					s				s, m
7.		s, m				s, m			s, m
8.	s					s, m			s, m
9.				s, m			m, l	s, m	
10.					s			s, m	
11.						s, m	l		
12.					s		m, l		
13.		s, m			s				
14.		s, m		s					
15.		s				s, m			
16.		s, m	s, m		s, m				
17.		s			s, m				
18.		s		s, m					

s - small; m - middle; l -large value of function

A feature of N is encountered in $\geq k_2$ objects N and in $\leq \tilde{k}_2$ objects D.

Table 7

Strong earthquakes and TIPs for regions

of California and adjacent parts of Nevada.

Start of TIP	Strong earthquake		End of false alarm	Duration of TIP, months
	Date	M		
Northern region				
-	9. 2. 1941	6, 6		failure to predict
1. 11. 1950	6. 7. 1954	6, 8		44
7. 7. 1954	16. 12. 1954	7, 2		5
-	21. 12. 1954	6, 6		failure to predict
1. 5. 1959	-	-	1. 5. 1962	36
1. 11. 1975	26. 11. 1976	6, 8		13
27. 11. 1976	-	-	1. 7. 1978	20
1. 3. 1980	8. 11. 1980	7, 2		8
Southern region				
1. 1. 1939	19. 5. 1940	6, 7		17
20. 5. 1940	21. 10. 1942	6, 5		29
1. 7. 1943	-	-	1. 9. 1944	14
1. 5. 1946	4. 12. 1948	6, 5		31
5. 12. 1948	21. 7. 1952	7, 7		43
1. 9. 1955	9. 2. 1956	6, 8		5
1. 9. 1966	9. 4. 1968	6, 4		19

Table 7 (continuation)

Start of TIP	Strong earthquake Date	M	End of false alarm	Duration of TIP, months
10. 4 1968	9. 2. 1971	6,4		34
10. 2. 1971	-	-	1. 9. 1972	19
1. 10. 1979	15. 10. 1979	6,6		1
16. 10. 1979	25. 5. 1980	6,4; 6,5		7
1. 11. 1982	-	-		30*

* TIP is not diagnosed, since the catalogs available end at April 1985.

Table 8

Test of the algorithm.

Territory	Period of diagnosis	Mo	Strong ea-s within TIPs/total	Duration of TIPs per strong ea-ke, years	%
East of Central Asia					
region N	1966-1987.8	6,4	3/ 5	0,5	11
region S	1966-1987.8	6,4	5/ 5	2,0	43
Trans-Baikal	1966-1984.1	6,4	0/ 0	0	0
Vrancea (E. Carpathean mts.)	1966-1986.7	6,4	2/ 2	3,2	30
Gulf of California	1968-1984.1	6,6	2/ 3	1,7	31
Cocos plate	1968-1984.1	6,5	4/ 4	1,6	38
N. Appalacheans					
region N	1964-1985.1	5,0	1/ 2	3,2	29
region S	1964-1985.1	5,0	0/ 0	0	0
Central Italy	1954-1986.1	5,6	3/ 5	1,2	18
Belgium	1966-1987.1	4,5	3/ 3	1,9	25

Note: Catalogs used are indicated in the text.

Table 9

Functions, used for identification of the TIPs for areas.

No in Table 1	Functions	Thresholds for discretization	Frequency of occurrence in learning material, % D s	m	l	N s	m	l
1.	N1	31; 37	27;	27;	46	34;	44;	22
3.	K	-2; 2	47;	13;	40	31;	38;	31
5.	Σ	24; 43	20;	47;	33	41;	28;	31
7.	Zmax	2,82; 4,4	33;	47;	20	34;	28;	28
8.	N3	2; 4	54;	23;	23	25;	39;	31
10.	L	-6; 4	20;	47;	33	34;	38;	28
11.	M1	3,8; 6,0	27;	40;	33	38;	31;	31
13.	Q	3; 5	40;	27;	33	41;	31;	28
16.	NR1	5; 9	33,	20;	47	41;	25;	34

Notations are the same, as in Table 4.

Table 10

Features D for areas,

$K_1 = 6$, $\tilde{K}_1 = 2$.

No	N1	K	Σ	Zmax	N3	L	M1	Q	NR1
1.			m, l					m, l	l
2.					s				s, m
3.						s, m		l	m, l
4.	l							s, m	m, l
5.		s						l	
6.						m, l	s, m	l	
7.		s, m					s, m	l	
8.						s, m	m, l	l	
9.	s, m						m, l	l	
10.	s, m	s, m						l	
11.				m				s	
12.			m, l			m, l		s, m	
13.				s, m		m, l		s, m	
14.			m, l	m			s, m		
15.	s, m		m, l			m, l			

Notations are the same, as in Table 5.

Table 11

Features N for areas,

$K_2 = 10, \tilde{K}_2 = 3.$

No	Ni	K	Σ	Zmax	N3	L	Mi	Q	NRi
1.					m, l		m, l		s, m
2.				s , l	m, l				s, m
3.				m, l	m, l				s, m
4.		m, l			m, l				s, m
5.	m, l				m, l				s, m
6.					m, l		s, m		s
7.					m, l		s, m	s, m	
8.			s					s, m	
9.	s, m						m, l	s, m	
10.	s, m					s, m		s, m	
11.		m, l			m, l			s, m	
12.	s, m			s , l				s, m	
13.	m							s, m	
14.	s, m	m, l						s, m	
15.					l		s, m		
16.				s , l	m, l		s, m		
17.			s		m, l				
18.		m, l		s , l	m, l				
19.	s, m	m, l			m, l				

Notations are the same, as in Table 6.