united nations educational, scientific and cultural organization

international atomic energy agency

the
abdus salam
international centre for theoretical physics

H4.SMR/1330-12

" Sixth Workshop on Non-Linear Dynamics and Earthquake Prediction"

15 - 27 October 2001

Pattern Recognition: Algorithhms and Applications to Geophysical Problems

V. Keilis-Borok, A. Soloviev

International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences Moscow, Russia www.mitp.ru

I. INTRODUCTION

Let a set of objects, phenomena or processes is considered. Certain information (for example results of measurements) is available about each element of the set, and there is some feature, possessed only by a part of the elements. If possessing this feature by an element does not present evidently in the information available, then a problem arises to distinguish elements that possess this feature. This problem could be solved by construction a model on the basis of mechanical, physical, chemical or other scientific laws which could explain the connection between the available information and the feature under consideration. But in many cases the construction of such model is difficult or practically impossible. In this case it is natural to apply pattern recognition methods.

1.1 Examples of Problems to Apply Pattern Recognition Methods

Recognition of earthquake-prone areas (for example Gelfand et al., 1976). A seismic region is considered. The problem is to determine in the region the areas where strong (with magnitude $M \geq M_0$ where M_0 is a threshold specified) earthquakes are possible. The objects are the selected geomorphological structures (intersections of lineaments, morphostructural knots, etc.) of the region. The possibility for a strong earthquake to occur near the object is the feature under consideration. The available information is the topographical, geological, geomorphological and geophysical data on the objects.

The problem as the pattern recognition one is to divide the selected structures into two classes:

- structures where earthquakes with $M \geq M_0$ may occur;
- structures where only earthquakes with $M < M_0$ may occur.

Intermediate-term prediction of earthquakes (for example Keilis-Borok and Rotwain, 1990). A seismic region is considered. The problem is to determine for any time t will a strong (with magnitude $M \geq M_0$ where M_0 is a threshold specified) earthquake occur in the region within the period $(t, t + \tau)$. Here τ is a given constant. The objects are moments of time. The occurrence of a strong earthquake is the feature under consideration. The available information is the values of functions on seismic flow calculated for the moment t.

The problem as the pattern recognition one is to divide the moments of time into two classes:

- moments, for which there is (or will be) a strong earthquake in the region within the period $(t, t + \tau)$;
- moments, for which there are not (or will not be) strong earthquakes in the region within the period $(t, t + \tau)$.

Recognition of strata filled with oil. The strata encountered by a borehole are considered. The problem is to determine what do the strata contain: oil or water. The objects are the strata. The filling of the strata with oil is the feature under consideration. The geological and geophysical data on the strata are the available information.

The problem as the pattern recognition one is to divide the strata into two classes:

- strata, which contain oil;
- strata, which contain water.

Medical diagnostics. A specific disease is considered. The problem is to diagnose the disease by using results of medical tests. The objects are examined people. The disease is the feature under consideration. The available information is the data obtained through medical tests.

The problem as the pattern recognition one is to divide examined people into two classes:

- people who have the disease;
- people who do not have it.

1.2 General Formulation of the Pattern Recognition Problem

One may give the general abstract formulation of the problem of pattern recognition as follows.

The set $W = \{ \mathbf{w}^i \}$ is considered, where objects $\mathbf{w}^i = (w_1^i, w_2^i, \ldots, w_m^i)$, $i = 1, 2, \ldots$ are vectors with real (integer, binary) components. Below these components will be called functions.

The problem is to divide the set W into two or more subsets, which differ in certain feature or according to clustering themselves.

There are two kinds of pattern recognition problems and methods:

- classification without learning;
- classification with learning.

1.3 Classification without Learning (Cluster Analysis)

The set W is divided into groups (clusters, see Fig. 1) on the basis of some measure in the m-dimensional space $w_1, w_2, \ldots, w_m$.

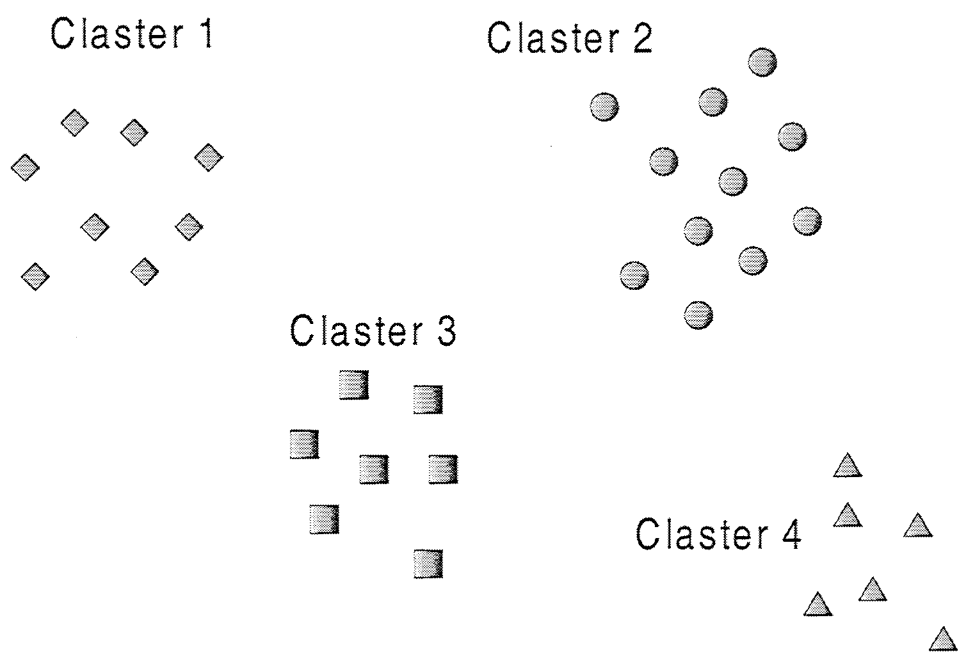

FIGURE 1 Clustering of objects in two-dimensional space

Denote $\rho(\mathbf{w}, \mathbf{v})$ a distance between two m-dimensional vectors $\mathbf{w} = (w_1, w_2, \ldots, w_m)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_m)$.

To define classification and to estimate at the same time its quality the special function is introduced. The best classification gives the extremum of this function.

Examples of the functions. Let W is a finite set. The following two functions can be used.

$$J_1 = \frac{(K-1)\sum_{k=1}^{K} \rho_k}{2\sum_{k=1}^{K-1} \sum_{j=k+1}^{K} \rho_{kj}} \Rightarrow \min$$

$$J_2 = \frac{1}{K}\left(\sum_{k=1}^{K} \rho_k - \frac{2}{K-1} \sum_{k=1}^{K-1} \sum_{j=k+1}^{K} \rho_{kj} \right) \Rightarrow \min$$

Here K is the number of groups,

$$\rho_k = \frac{2}{m_k(m_k - 1)} \sum_{i=1}^{m_k - 1} \sum_{s=i+1}^{m_k} \rho(\mathbf{w}^i, \mathbf{w}^s),$$

$$\rho_{kj} = \frac{1}{m_k m_j} \sum_{i=1}^{m_k} \sum_{s=1}^{m_j} \rho(\mathbf{w}^i, \mathbf{v}^s),$$

m_k, m_j are the number of objects in the group numbered k and in the group numbered j respectively; $\mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^{m_k}$ are the objects of the group numbered k; $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^{m_j}$ are the objects of the group numbered j.

After the groups are determined the next problem can be formulated: to find common feature of objects, which belong to the same group.

1.4 Classification with Learning

If it is a priori known about some objects to what groups (classes) they belong, then this information can be used to determine classification for other objects.

As a rule the set W is divided into two classes, say D and N.

The a priori examples of objects of each class are given. They are called the learning set W_0:

$W_0 \subset W$,

$W_0 = D_0 \cup N_0$.

Here D_0 is the learning set (the a priori examples) of objects belonging to class D, N_0 is the learning set of objects belonging to class N.

The learning set W_0 is used to determine a priori unknown distribution of objects of the set W_0 between the classes D and N.

The result of the pattern recognition is twofold:

- the rule of recognition; it allows to recognize which class an object belongs to knowing the vector $\mathbf{w}^i$ describing this object;
- the actual division of objects into separate classes according to this rule (Fig. 2):

$W = D \cup N$

or if there are objects with undefined classification then

$W = (D \cup N) \cup U$.

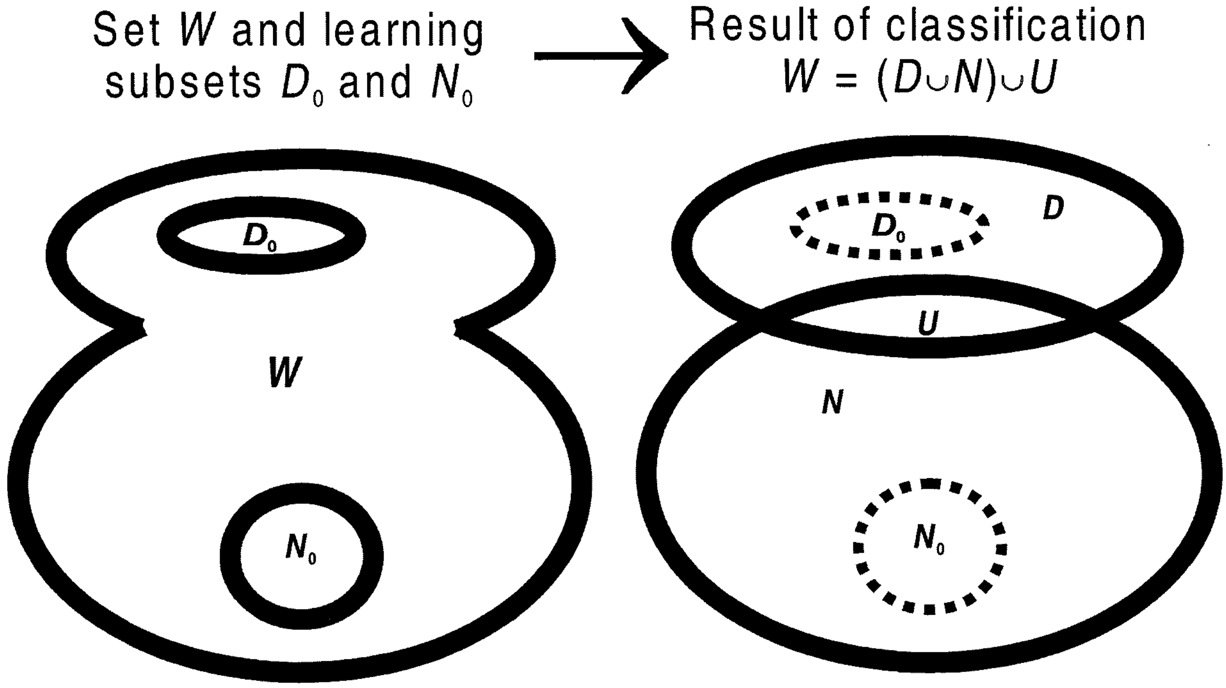

FIGURE 2 Classification with learning

Analysis of the obtained rule of recognition may give information for understanding the connection between the feature, which differs the classes *D* and *N*, on one hand and description of objects (components of vectors $\mathbf{w}^i$) on another.

II. EXAMPLES OF ALGORITHMS

Some algorithms used to solve problems of classification with learning are described below.

2.1 Statistical Algorithms

These algorithms are based on the assumption that distribution laws are different for vectors from classes D and N (see Fig. 3). The samples D_0 and N_0 are used to define the parameters of these laws.

The recognition rule includes calculating an estimation of conditional probabilities for each object $\mathbf{w}^i$ to belong to class D (P_D^i) and N (P_N^i). Classification of the objects according to these probabilities is performed as follows:

$$\mathbf{w}^i \in D, \text{ if } P_D^i - P_N^i \geq \varepsilon,$$
$$\mathbf{w}^i \in N, \text{ if } P_D^i - P_N^i < -\varepsilon,$$
$$\mathbf{w}^i \in U, \text{ if } -\varepsilon \leq P_D^i - P_N^i < \varepsilon,$$

where $\varepsilon \geq 0$ is a given constant.

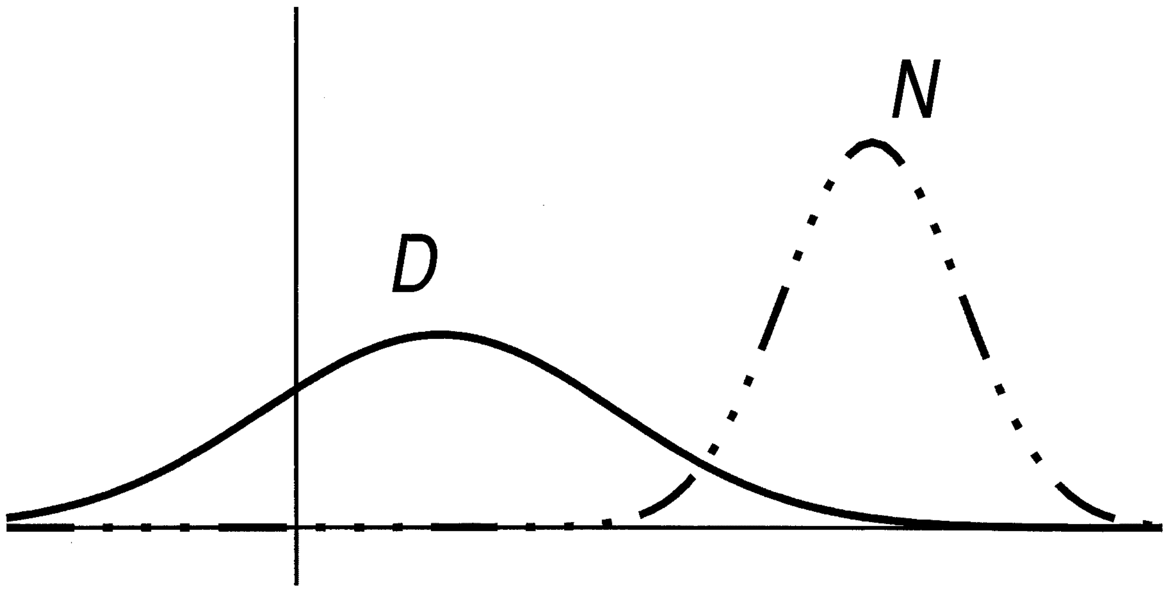

FIGURE 3 Different distribution laws for classes D and N

Bayes algorithm. This is an example of a statistical algorithm. According to Bayes formula

$$P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in D)\, P(\mathbf{w} \in D) = P(\mathbf{w} \in D | \mathbf{w} = \mathbf{w}^i)\, P(\mathbf{w} = \mathbf{w}^i) \tag{1}$$

It follows from (1) that

$$P_D^i = P\left(\mathbf{w} \in D \middle| \mathbf{w} = \mathbf{w}^i\right) = \frac{P\left(\mathbf{w} = \mathbf{w}^i \middle| \mathbf{w} \in D\right) P(\mathbf{w} \in D)}{P\left(\mathbf{w} = \mathbf{w}^i\right)}.$$

Similarly

$$P_N^i = P(\mathbf{w} \in N | \mathbf{w} = \mathbf{w}^i) = \frac{P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in N) P(\mathbf{w} \in N)}{P(\mathbf{w} = \mathbf{w}^i)}.$$

Estimations of probabilities in the right side of these relations are given by following approximate formulae, in which the samples D_0 and N_0 are used:

$$P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in D) \approx P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in D_0),$$

$$P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in N) \approx P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in N_0),$$

$$P(\mathbf{w} = \mathbf{w}^i) \approx P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in D_0)\, P(\mathbf{w} \in D) + P(\mathbf{w} = \mathbf{w}^i | \mathbf{w} \in N_0)\, P(\mathbf{w} \in N).$$

Probability $P(\mathbf{w} \in D)$ is a parameter of the algorithm and has to be given,
$P(\mathbf{w} \in N) = 1 - P(\mathbf{w} \in D)$.

2.2 Geometrical Algorithms

In these algorithms surfaces in the space $w_1, w_2, \ldots, w_m$ are constructed to separate classes D and N (see Fig. 4).

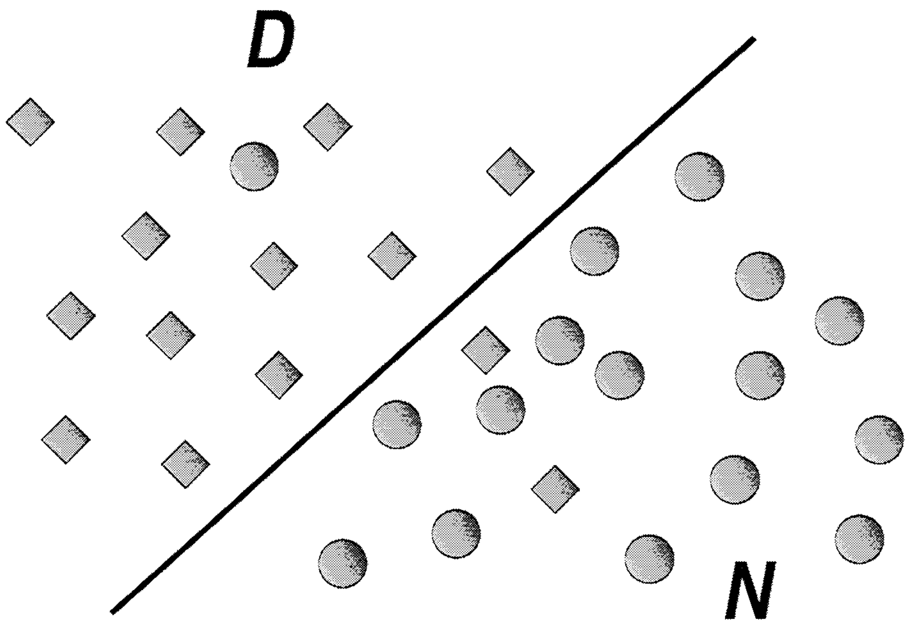

FIGURE 4 Separation of objects from classes D (rhombs) and N (circles) in two-dimensional space by a straight line.

Algorithm Hyperplane. This is an example of a geometrical algorithm.

The hyperplane $P(\mathbf{w}) = a_0 + a_1 w_1 + a_2 w_2 + \ldots + a_m w_m = 0$ is constructed in the space $w_1, w_2, \ldots, w_m$ to separate the sets D_0 and N_0 by the best way. It means that some function on the hyperplnane has to have extremum value.

The example of the function is

$$J(a_0, a_1, \ldots, a_m) = \sum_{i=1}^{n_1} P(\mathbf{w}^i) - \sum_{i=1}^{n_2} P(\mathbf{v}^i) \Rightarrow \max.$$

Here $\mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^{n_1}$ are objects of D_0, $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^{n_2}$ are objects of N_0.

The recognition rule is formulated as follows:

$$\mathbf{w}^i \in D, \text{ if } P(\mathbf{w}^i) \geq \varepsilon,$$
$$\mathbf{w}^i \in N, \text{ if } P(\mathbf{w}^i) < -\varepsilon,$$
$$\mathbf{w}^i \in U, \text{ if } -\varepsilon \leq P(\mathbf{w}^i) < \varepsilon,$$

where $\varepsilon \geq 0$ is a given constant.

2.3 Logical Algorithms

In these algorithms characteristic traits of classes D and N are searched using the sets D_0 and N_0. Traits are boolean functions on $w_1, w_2, \ldots, w_m$. The object $\mathbf{w}^i$ has the trait, if the value of the corresponding function, calculated for it, is ***true***, and does not have the trait, if it is ***false***. A trait is a characteristic trait of the class D, if the objects of the set D_0 have this trait more often than the objects of the set N_0. A trait is a characteristic trait of the class N, if the objects of the set N_0 have this trait more often than objects of the set D_0.

Using the searched characteristic traits the recognition rule is formulated as follows:

$$\mathbf{w}^i \in D, \text{ if } n_D^i - n_N^i \geq \Delta + \varepsilon,$$
$$\mathbf{w}^i \in N, \text{ if } n_D^i - n_N^i < \Delta - \varepsilon,$$
$$\mathbf{w}^i \in U, \text{ if } \Delta - \varepsilon \leq n_D^i - n_N^i < \Delta + \varepsilon.$$

Here n_D^i and n_N^i are the numbers of characteristic traits of classes D and N, which the object $\mathbf{w}^i$ has, Δ and $\varepsilon \geq 0$ are given constants.

Logical algorithms are useful to apply in cases then the numbers of objects in sets D_0 and N_0 are small.

As a rule logical algorithms are applied to vectors with binary components. An example of logical algorithm is the algorithm CORA-3. It is applied to geophysical problems in particular to the problems of recognition of earthquake-prone areas and intermediate-term prediction of earthquakes. The detailed description of this algorithm can be found in Gelfand et al. (1976) and is given below.

III. PRELIMINARY DATA PROCESSING

As it was mentioned above some pattern recognition algorithms (for example CORA-3) do classify the vectors with binary components. Therefore, if the set W initially consists of vectors with real components (functions) then prior to an algorithm application, the coding of objects in the form of vectors with binary components has to be carried out. To do it, the characteristics should be discretized, i.e. their intervals of values should be represented as the union of disjoint parts. Each of these parts is given accordingly by the value of a component of a binary vector or by the combination of values of its several components.

After discretization the data become robust. For example if a range of some function is divided into three parts only three gradations for this function ("small", "medium", "large") are used after the discretization instead of its exact value. Do not regret the loss of information. This makes results of recognition stable to variations of data.

3.1 Discretization

Let us consider some component (function) w_j of vectors (objects), which form the set W. Let the range of the function variation is limited with the numbers x_0^j and x_f^j ($x_0^j < x_f^j$). The procedure of discretization for the function w_j consists of dividing the range of its variation into k_j intervals by thresholds of discretization (Fig. 5):

$$x_1^j, x_2^j, \ldots, x_{k_j-1}^j \quad \left(x_0^j < x_1^j < x_2^j < \ldots < x_{k_j-1}^j < x_f^j\right)$$

Assume that the value w_j^i of the function numbered j of the object numbered i belongs to the interval numbered s, if $x_{s-1}^j < w_j^i \le x_s^j$, where $x_{k_j+1}^j = x_f^j$. In a process of discretization we substitute the exact value of the function by the interval, which contains this value.

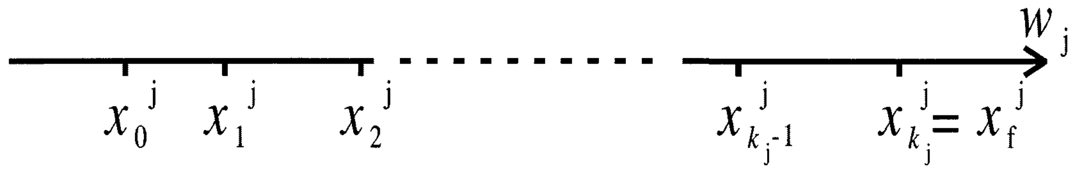

FIGURE 5 Discretization of the function w_j.

Usually we divide the range of function variation into two intervals ("small" and "large" values) or into three intervals ("small", "medium" and "large" values).

Thresholds of discretization can be introduce manually on the basis of various considerations for the nature of the given function.

The other way to define the thresholds is to compute them so as to make the numbers of objects within each interval (x_{s-1}^j, x_s^j), $s = 1, 2, \ldots, k_j$, are roughly equal to each other. In this case one has to specify the number of intervals k_j only. Then the thresholds of discretization may be calculated by using a special algorithm. All objects together or only objects of D_0 and N_0 can be considered. This type of discretization is called here and below as ***objective*** or ***automatic***.

Our purpose is to find such intervals where values of the function w_j for objects from one class occur more often than for objects from another class.

How informative is the function w_j in a given discretization can be characterized as follows.

1. Let us compute for each interval $(x_{s-1}^j,\ x_s^j)$ the numbers P_s^D and P_s^N (s = 1, 2, ..., k_j), which give for the sets D_0 and N_0 respectively the percent of objects, for which the value of the function w_j falls within the interval numbered s.

Let us denote $P_{\max} = \max\limits_{1\le s\le k_j}\left|P_s^D - P_s^N\right|$.

In other words P_s^D and P_s^N are empirical histograms of the function w_j for the sets D_0 and N_0, and $P_{\max}$ is the maximal difference of these histograms.

The larger is $P_{\max}$, the more informative is the function w_j.

Functions for which $P_{\max} < 20\%$ are usually excluded.

2. Let $k_j = 3$. Let us denote:

$$M_D = \frac{\left|P_2^D - P_1^D\right| + \left|P_3^D - P_2^D\right|}{\left|P_3^D - P_1^D\right|},$$

$$M_N = \frac{\left|P_2^N - P_1^N\right| + \left|P_3^N - P_2^N\right|}{\left|P_3^N - P_1^N\right|}.$$

If P_s^D changes monotonously with s, $M_D = 1$; the larger is M_D, more jerky is P_s^D. This is clear from Fig. 6. Similar statements are true for M_N, P_s^N.

The smaller are M_D and M_N, the better is the discretization of the function w_j. Functions with both $M_D, M_N \ge 3$ are usually excluded.

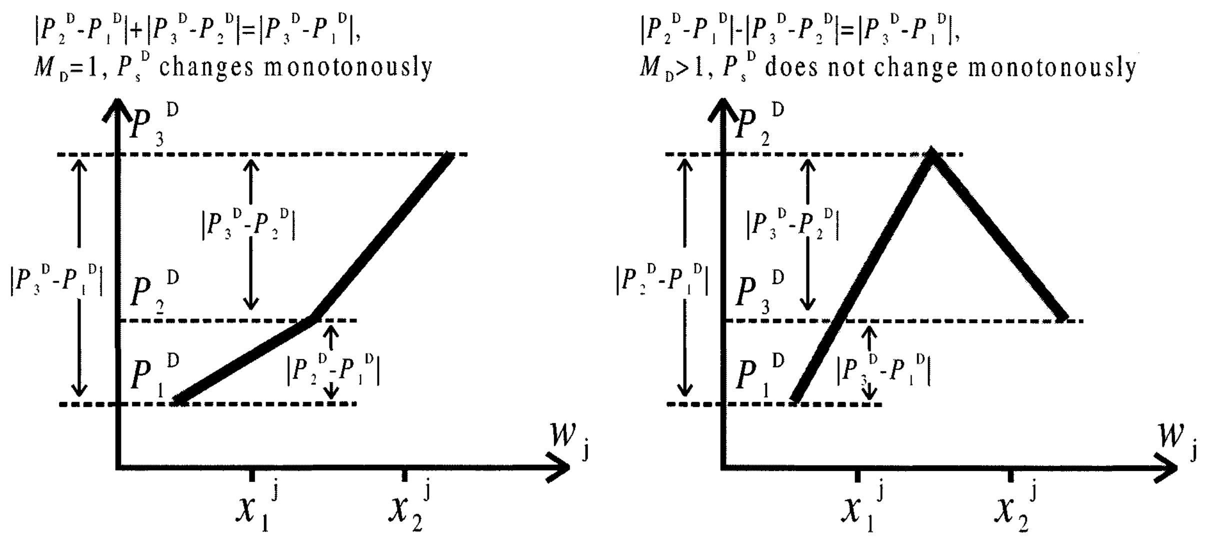

FIGURE 6 Monotonous and non-monotonous changing of P_s^D

3. Samples D_0 and N_0 are often marginally small, so that their observed difference may be random. Therefore the relation between functions P_s^D and P_s^N after discretization should be not absurd according to the problem under consideration, though they may be unexpected indeed.

3.2 Coding

With discretization thresholds determined, a procedure of coding of vectors $\mathbf{w}^i$ into the form of binary vectors is undertaken. For coding only the functions selected at the stage of discretization are considered. At the stage of coding l_j components of binary vectors are determined for the function w_j. Number l_j depends on the number of thresholds as well as on the type of coding procedure applied to the function w_j.

For coding the following two procedures are used. In the case of ***I*** ("impulse") procedure $l_j = k_j$, i.e. the number of binary vector components allocated for the coding of the function w_j is equal to the number of intervals into which the range of its variation is divided after discretization.

Let us denote as $\omega_1, \omega_2, ..., \omega_{lj}$ the values of binary vector components, which code the function w_j. If the value w_j^i of the function w_j for the object numbered i falls within the s-th interval of its discretization, i.e. $x_{s-1}^j < w_j^i \leq x_s^j$, then we set

$\omega_1 = \omega_2 = ... = \omega_{s-1} = 0, \omega_s = 1, \omega_{s+1} = 0 = ... = \omega_{lj} = 0.$

In the case of ***S*** ("stair") procedure $l_j = k_j - 1$, i.e. the number of binary vector components, allocated for the coding of a function, is equal to the number of the thresholds of discretization. If the value w_j^i for the object numbered i falls within the s-th interval of its discretization, then we set

$\omega_1 = \omega_2 = ... = \omega_{s-1} = 0, \omega_s = \omega_{s+1} = ... = \omega_{lj} = 1.$

Below the case when the codes of the function w_j are constructed for $k_j = 3$ is considered.

If the value w_j^i belongs to the first interval ($x_0^j < w_j^i \leq x_1^j$) ***I***-coding has the form: 100. ***S***-coding for the same value w_j^i has the form: 11. For the second interval ($x_1^j < w_j^i \leq x_2^j$) the codes are 010 (***I***-method) and 01 (***S***-method). For the third interval ($x_2^j < w_j^i \leq x_3^j$) they are 001 and 00 respectively.

Discretization and coding procedures transform the set of vectors $W = \{ \mathbf{w}^i \}$, $i = 1, 2, ..., n$, which correspond to all objects into a set of vectors with l binary components. Here $l = \sum' l_j$, where summation is implemented only over the functions left after discretization.

Thus, discretization and coding transform the initial problem in the form of the classification within the finite set of l-dimensional vectors with binary components. These vectors will be called objects of recognition.

IV. ALGORITHM CORA-3

Algorithm CORA-3 operates in two steps:
- selection of characteristic traits (*learning*);
- *voting*.

4.1 Learning

The sets of characteristic traits for classes D and N are constructed at this step on the basis of sets D_0 and N_0.

Traits. Matrix

$$\mathbf{A} = \begin{Vmatrix} i_1 & i_2 & i_3 \\ \delta_1 & \delta_2 & \delta_3 \end{Vmatrix}$$

is called by a trait. Here i_1, i_2, i_3 are the natural numbers such as $1 \le i_1 \le i_2 \le i_3 \le l$ and δ_1, δ_2, δ_3 are equal to 0 or to 1.

We say that the object, which is the binary vector $\boldsymbol{\omega}^i = (\omega_1^i, \omega_2^i, ..., \omega_l^i)$, has the trait **A** if

$$\omega_{i_1}^i = \delta_1, \quad \omega_{i_2}^i = \delta_2, \quad \omega_{i_3}^i = \delta_3.$$

Characteristic traits. Let $W' \subseteq W$. We shall denote by $K(W', \mathbf{A})$ the number of objects $\boldsymbol{\omega}^i \in W'$, which have the trait **A**.

The algorithm has four free parameters $k_1, \bar{k}_1, k_2, \bar{k}_2$, which may take integer non-negative values. While the values of the free parameters are specified, the notion of characteristic traits is introduced.

The trait **A** is a characteristic trait of class D if

$K(D_0, \mathbf{A}) \ge k_1$ and $K(N_0, \mathbf{A}) \le \bar{k}_1$.

The trait A is a characteristic trait of class N if

$K(N_0, \mathbf{A}) \ge k_2$ and $K(D_0, \mathbf{A}) \le \bar{k}_2$.

Parameters k_1 and k_2 are called by selection thresholds for characteristic traits of classes D and N respectively. Parameters $\bar{k}_1$ and $\bar{k}_2$ are called by the contradiction thresholds for characteristic traits of classes D and N.

Equivalent, weaker, and stronger traits. The number of characteristic traits of each class may be large enough. Among them groups of traits, which occur on the same learning objects of their class, may be. There is no reason to include all traits from such group in the final list.

Let $\Omega(\mathbf{A})$ be a subset of the set W consisting of the objects, which have the trait **A**. Let, also, $\mathbf{A}_1$ and $\mathbf{A}_2$ be two characteristic traits of class D. We say that the trait $\mathbf{A}_1$ is weaker than the trait $\mathbf{A}_2$ (or $\mathbf{A}_2$ is stronger than $\mathbf{A}_1$), if

$\Omega(\mathbf{A}_1) \cap D_0 \subset \Omega(\mathbf{A}_2) \cap D_0$ and $(\Omega(\mathbf{A}_2) \cap D_0) \backslash (\Omega(\mathbf{A}_1) \cap D_0) \ne \varnothing$.

In other words it means that all objects from D_0, having $\mathbf{A}_1$, possess also $\mathbf{A}_2$. At the same time there is at least one object from D_0, which, having the trait $\mathbf{A}_2$, does not have $\mathbf{A}_1$.

A similar definition is introduced for characteristic traits of class N. Let $\mathbf{A}_1$ and $\mathbf{A}_2$ be two characteristic traits of class N. Then the trait $\mathbf{A}_1$ is weaker than the trait $\mathbf{A}_2$ (or $\mathbf{A}_2$ is stronger than $\mathbf{A}_1$), if

$\Omega(\mathbf{A}_1) \cap N_0 \subset \Omega(\mathbf{A}_2) \cap N_0$ and $(\Omega(\mathbf{A}_2) \cap N_0) \backslash (\Omega(\mathbf{A}_1) \cap N_0) \ne \varnothing$.

If two characteristic traits $\mathbf{A}_1$ and $\mathbf{A}_2$ of class D are both found in the same objects of the set D_0 i.e.

$\Omega(\mathbf{A}_1) \cap D_0 = \Omega(\mathbf{A}_2) \cap D_0$,

we call $\mathbf{A}_1$ and $\mathbf{A}_2$ as equivalent.

Similarly, characteristics traits $\mathbf{A}_1$ and $\mathbf{A}_2$ of class N are called equivalent if

$\Omega(\mathbf{A}_1) \cap N_0 = \Omega(\mathbf{A}_2) \cap N_0$.

The lists of characteristic traits of classes being formed as a result of the learning step by definition include no any trait, which is weaker than any trait in the list of its class. Only one trait (selected first) is included from each group of equivalent ones to the final list.

Thus, the learning step results in the set of q_D characteristic traits of class D and the set of q_N of ones of the class N. These sets containing no weaker or equivalent traits in relation to any one from the same set.

4.2 Voting and Classification

The second step of the algorithm involves voting and classification. For each object $\omega^i \in W$ the number n_D^i of the characteristic traits of class D, which the object has, the number n_N^i of ones of class N, and the difference $\Delta_i = n_D^i - n_N^i$ are calculated.

Classification is performed by the following way.

Class D (the set D) is formed from the objects ω^i, for which $\Delta_i \geq \Delta$. The objects, for which $\Delta_i < \Delta$, are included in class N (the set N).

Here Δ is a parameter of the algorithm as well as k_1, $\bar{k}_1$, k_2, and $\bar{k}_2$.

This recognition rule corresponds to $\varepsilon = 0$ in the description of logical algorithms given above.

4.3 Algorithm CLUSTERS

Algorithm CLUSTERS is the modification of algorithm CORA-3 (Gelfand et al., 1976). It is applied in the case when the set D_0 consists of S subsets (subclasses):

$D_0 = D_0^1 \cup D_0^2 \cup \ldots \cup D_0^S$,

and it is known a priori that each subclass has at least one object of class D but some objects of the set D_0 may belong to class N.

At the learning step algorithm CLUSTERS differs from CORA-3 in the following.

First, by definition a subclass has a trait if at least one object among those, which belong to this subclass, has this trait.

The trait $\mathbf{A}$ is a characteristic trait of class D if

$K^S(D_0, \mathbf{A}) \geq k_1$ and $K(N_0, \mathbf{A}) \leq \bar{k}_1$.

Here $K^S(D_0, \mathbf{A})$ is the number of subclasses, which have the trait $\mathbf{A}$.

Second, the definition of the weaker and equivalent traits for characteristic traits of class D changes to the following.

A characteristic trait $\mathbf{A}_1$ of class D is weaker than a characteristic trait $\mathbf{A}_2$ of this class if any subclass having the trait $\mathbf{A}_1$ has also $\mathbf{A}_2$, and there is at least one subclass, which has the trait $\mathbf{A}_2$ but does not have the trait $\mathbf{A}_1$. Traits $\mathbf{A}_1$ and $\mathbf{A}_2$ are equivalent if they are found in the same subclasses.

Algorithm CLUSTERS forms the sets of characteristic traits of classes D and N like CORA-3.

The step of voting and classification is the same as in algorithm CORA-3.

V. ALGORITHM HAMMING

Another algorithm applied to geophysical problems is algorithm HAMMING (Gvishiani and Kosobokov, 1981). There are also other possible applications of this algorithm (for example Keilis-Borok and Lichtman, 1981).

The application of this algorithm consists also in two steps.

5.1 Learning

At the first step (learning) for each component ω_k ($k = 1, 2, ..., l$) of binary vectors the following values are calculated:

$q_D(k|0)$ - the number of objects of the set D_0, which have $\omega_k = 0$,

$q_D(k|1)$ - the number of objects of the set D_0, which have $\omega_k = 1$,

$q_N(k|0)$ - the number of objects of the set N_0, which have $\omega_k = 0$,

$q_N(k|1)$ - the number of objects of the set N_0, which have $\omega_k = 1$.

Then the relative number of objects, which have this component equal to 1, is determined for the set D_0:

$$\alpha_D(k|1) = \frac{q_D(k|1)}{q_D(k|0) + q_D(k|1)}$$

and for the set N_0:

$$\alpha_N(k|1) = \frac{q_N(k|1)}{q_N(k|0) + q_N(k|1)} \cdot$$

Then components of a binary vector $K = (\kappa_1, \kappa_2, ..., \kappa_l)$, which is called as ***kernel of class D***, are determined as follows

$$\kappa_k = \begin{cases} 1, \text{ if } \alpha_D(k|1) \geq \alpha_N(k|1), \\ 0, \text{ if } \alpha_D(k|1) < \alpha_N(k|1). \end{cases}$$

Values of the components of the kernel of class D are more "typical" for the objects of the set D_0 than for the objects of the set N_0. The calculation of the kernel **K** completes the first step of applying the algorithm.

NOTE: It may be more reliable to eliminate the components, for which $|\alpha_D(k|1) - \alpha_N(k|1)| < \varepsilon$, where ε is a small positive constant.

5.2 Voting and Classification

The voting and the actual classification are carried out at the second step. The voting consists of calculating for each object a Hamming's distance ρ_i to the kernel of class D. It is calculated by the formula:

$$\rho_i = \sum_{k=1}^{l} |\omega_k^i - \kappa_k|.$$

Classification is performed as follows.

Class D (the set D) is formed from the objects ω^i, for which $\rho_i \le R$.
The objects, for which $\rho_i > R$, are included in class N (the set N).
Here R is a parameter of the algorithm.
Hamming's distance may be calculated considering the weights of the components

$$\rho_i = \sum_{k=1}^{l} \left|\omega_k^i - \kappa_k\right| \xi_k .$$

Here $\xi_k > 0$ ($k = 1, 2, ..., l$) are the weights associated to the components of the binary vectors. The weights may be assigned intuitively or computed by the formula:

$$\xi_k = \frac{\left|\alpha_D(k|1) - \alpha_N(k|1)\right|}{\max_k \left|\alpha_D(k|1) - \alpha_N(k|1)\right|}$$

where maximum is taken among the components used in the given run of the algorithm.

VI. EVALUATION OF THE CLASSIFICATION RELIABILITY

Reliability of results of recognition is evaluated by several methods including control experiments, statistical analysis of the established classification and other techniques. These tests are necessary to be sure in the obtained results. It is especially important in the case of small samples D_0 and N_0. The tests illustrate - how reliable are the results of the pattern recognition. However they do not provide a proof in the strict statistical sense if the learning material is small.

The following simplest tests are useful.

1. To save the part of objects from W_0 for recognition only, not using it in learning.
2. To check the conditions: $D_0 \subset D$, $N_0 \subset N$.

NOTE: Sometimes these conditions are not valid because the sets D_0 and N_0 are not "clear" enough. For example in the case of recognition of earthquake-prone areas objects of D_0 are structures where epicenters of earthquakes with $M \geq M_0$ are known and objects of N_0 are structures where epicenters of such earthquakes are not known. Objects of N_0 may belong to the class D, because in some areas earthquakes with $M \geq M_0$ may be possible, though yet unknown. Objects of D_0 may belong to the class N due to the errors in the catalog (in epicenters and/or magnitude).

The examples of some other tests are listed below. These tests include some variation of the objects, used components of vectors, numerical parameters etc. The test is positive if the results of recognition are stable to these variations. Since the danger of selfdeception is not completely eliminated by these tests the design and implementation of new tests should be pursued.

6.1 Using a Result of Classification as a Learning Set (RLS Experiment)

This experiment is an attempt to repeat the established classification $W = D \cup N$, using the resultant sets D and N as the new learning sets instead of D_0 and N_0. We usually consider this experiment as successful if not more than 5% of the total number of objects are classified in the experiment differently comparing with their initial classification. The "physical" idea of the experiment is rather obvious and natural: if our classification is correct then such changing of learning material should not change the result of classification. The use of CORA-3 algorithm enables usually to repeat the initial classification, if making the experiment we assume $\bar{k}_1 = \bar{k}_2 = 0$ and specify sufficiently small values for thresholds k_1 and k_2. Thus if CORA-3 algorithm is used then the experiment should be executed with non-zero thresholds $\bar{k}_1$ and $\bar{k}_2$. For instance, $\bar{k}_1 = \bar{k}_2 = 1$, or $\bar{k}_1 = \bar{k}_2 = 2$, or $\bar{k}_1$ and $\bar{k}_2$ have the same values as in the initial classification. In the case of $\bar{k}_1 = \bar{k}_2 = 0$ the substantial information is carried with maximum values of k_1 and k_2, under which the initial classification can be repeated.

In the case of any algorithm used to obtain the initial classification, it's advisable to repeat it in making the experiment by using HAMMING algorithm. We consider success of RLS experiment as the necessary condition for the classification obtained to pretend to be the problem solution. In this sense RLS experiment is obligatory to check the reliability of the classification.

6.2 Stability Testing (ST) Experiments

These experiments generalize RLS experiment. Their goal is to obtain the initial classification $W = D \cup N$, using the various subsets $D_0' \subseteq D$, $N_0' \subseteq N$ as D_0 and N_0 learning sets. The experiment is considered successful if the initial classification is rather stable while we change learning material. Usually we accept the result if not more than 10% of the total number of objects change their classification in the result of the experiment. The choice of the subsets D_0' and N_0', which are used as the learning sets to play the role of D_0 and N_0 in the experiment, has to be in sufficiently natural way. For instance, in the case of recognition of earthquake-prone areas the region can be divided into two geographical parts, southern and northern or western and eastern. The subsets D_0' and N_0' then may be formed correspondingly from objects of the sets D and N fallen within one of the parts selected in the region. The subsets D_0' and N_0' may be also selected on the basis of the voting results in the initial classification. If algorithm HAMMING is used, the objects $\mathbf{w}^i \in D$, remaining in the set D with smaller value of R, may be assigned to D_0'. The objects from the set N, which remain in the set N with greater value of R, are assigned to the new learning set N_0'. If CORA-3 or CLUSTERS algorithm is used, objects from the set D, which remain in the set D with greater value of Δ, can be assigned to D_0'. Analogously the objects $\mathbf{w}^i \in N$, remaining in the set N with smaller value of Δ, can be assigned to N_0'.

Successful results of ST experiments, especially in the case when D_0' does not contain the whole D_0 set, are convincing arguments for the validity of the established classification. At the same time the success of this experiment with any choice of D_0' and N_0' can not be considered as necessary condition for the result of recognition to be valid.

6.3 Sliding Control (SC) Experiment

This experiment is designed for establishing classifications on the basis of the learning sets $(D_0 \backslash \mathbf{w}^i)$ and $(N_0 \backslash \mathbf{w}^{i+n_1})$, $i = 1, 2, ..., \max(n_1,n_2)$. The idea of SC experiment is very clear. We just want to check weather classification of the objects belonging to the learning set is stable while they are excluded from the learning set. The first variant discards the objects $\mathbf{w}^1 \in D_0$ and $\mathbf{w}^{1+n_1} \in N_0$, the second variant resets them but discards the objects $\mathbf{w}^2 \in D_0$ and $\mathbf{w}^{2+n_1} \in N_0$, etc. If one of the sets D_0 or N_0 (with a smaller number of objects) has already all its objects discarded once, we proceed only with the other set. With applying CLUSTERS algorithm, the whole subclasses are excluded in turn from the set D_0.

Formal criteria of success of the experiment is small value of the ratio $\frac{m_D}{|D_0|}$ or $\frac{m_D + m_N}{|D_0| + |N_0|}$. Here m_D and m_N show how many objects of D_0 and N_0 respectively change classification after they were eliminated from learning. We usually consider SC experiment as successful if not above 20% of objects in each of D_0 and N_0 sets change their classification while neglecting.

This experiment is very similar to the well-known "Jack-Knife" procedure, under which each variant discards only one object, first from D_0, and then from N_0. On the other hand SC is preferable because it needs executing less variants of classification.

6.4 Voting by the Set of Equivalent Features (VSEF)

This experiment is applied only if classification is obtained by CORA-3 or CLUSTERS algorithms. In both cases the result of classification depends, generally speaking, on the choice of traits from the groups of equivalent. The chosen traits then are included in the sets of characteristic ones resulted from learning. VSEF experiment pursues the goal to evaluate how much the classification obtained is stable toward such choice. Let us denote by u_{Dj}^i the number of traits, which the object $\mathbf{w}^i$ has from the group of ones equivalent to j-th trait of class D. Analogous u^i_{Nj} is the number of traits, which the object $\mathbf{w}^i$ has from the group of ones equivalent to j-th trait of class N. Let us define on the bases of numbers u^i_{Dj} and u^i_{Nj} the numbers of "votes" in favor of classes D and N respectively. They are defined by the formulas

$$u^i_D = \sum_{j=1}^{p_D} \frac{u^i_{Dj}}{p^j_D}, \quad u^i_N = \sum_{j=1}^{p_N} \frac{u^i_{Nj}}{p^j_N} .$$

Here p^j_D is the total number of traits equivalent to j-th trait of class D, p^j_N - the number of traits equivalent to j-th trait of class N. In calculation of both numbers p^j_D and p^j_N j-th trait itself is obviously included. In the experiment the set D is formed from the objects, which satisfy the condition u^i_D - $u^i_N \geq \Delta$ and the rest of objects forms the set N.

We usually assume that the experiment is successful if this classification differs from the original one in less than 5% of objects. According to the construction the success of this experiment has to be considered as necessary condition of the validity of the result if CORA-3 or CLUSTERS algorithm is applied.

6.5 Experiments with Random Data

These experiments (Gvishiani and Kosobokov, 1981) are designed for estimating probability of a classification error and for verifying the classification to be non-random. Let us consider the sequence of intermixed problems. An intermixed problem is formulated on the basis of initial one by a random choice of n_1 objects from given n objects of the set W and also by a random choice n_2 objects from the rest of n - n_1 objects of the set W. These two new random learning sets we symbolize as D_0' and N_0'. Coding of the objects in the form of binary vectors remains the same for an intermixed problem as it is in the real one. In other words it means that we preserve the relationship between the characteristics, which organic one to the set W as a whole.

Therefore $C_n^{n_1} C_{n-n_1}^{n_2} = n!/n_1!n_2!(n-n_1-n_2)!$ of intermixed problems may be defined. On the next stage a pattern recognition algorithm is applied to an intermixed problem. This way the classification $W = D \cup N$ based upon the learning sets D_0' and N_0' is obtained in the given intermixed problem. In the random case we require the condition that $|D|$ is not greater than the number of objects in the set D obtained in the initial classification.

Assume that F of intermixed problems have been formed and f_1 among them succeeded to include $D_0' \subseteq D$. Then f_1/F ratio may be used as the measure of the result to be non-random. If the values of f_1/F are small it obviously means that, it is complicated to obtain a random result of the same quality as the real one. In this sense the small values of f_1/F speak for the fact that the real result obtained is non-random. On the other hand it cannot of course be used as necessary condition to proceed with the classification.

As it is shown by Gvishiani and Kosobokov (1981) under some natural additional requirements classifications in intermixed problems offer to define the upper estimate of classification error probability for the original problem. This upper estimate is calculated by the formula

$$\overline{p} = \overline{|N|}/n - \overline{v_D}/n_1 .$$

Here $\overline{|N|}$ is the mean number of objects allocated to class N in the intermixed problems: $\overline{v_D}$ - the mean number of objects from sets D_0' allocated to N for different intermixed problems.

It is obvious that a small value of $\overline{p}$ is the argument for the validity of the classification in the original problem. If the estimation results in a sufficiently large value ($\overline{p}$ > 0.5), it is advisable to analyze again the statement of the original problem. For instance, large value $\bar{p}$ can be related with an insufficiently large size of D_0. On the other hand one should remember that $\bar{p}$ gives only the upper estimation of the error probability. In fact might really have much smaller value than $\bar{p}$.

6.6 Result Replication Experiments

These experiments are the attempts to replicate the obtained result by altering the solution procedure starting with some intermediate stage. The application of another pattern recognition algorithm is used in the simplest example of such experiment. In other words classification was established by performing CORA-3 algorithm, then, using that same coding of objects, an attempt may be made to repeat the classification by applying HAMMING algorithm. This experiment is usually considered as satisfactory one if not more than 20% of objects change their classification.

When application of a simpler algorithm results in repeating almost entirely the initial classification, its validation rises, of course. On the other hand replication of the classification by another algorithm cannot be considered, of course, as the necessary condition for the result to be valid.

The set of used components of binary vectors may be changed. In particular this may include elimination of each used component in turn.

An attempt may be also made to repeat the classification altering discretization thresholds for the functions describing the objects. Corresponding changes in coding of the objects should be also made. New functions may be included in the description of the objects. Then by replication of all subsequent stages of the problem consideration, a new classification is established and its comparison with the initial is made.

VII. APPLICATION OF PATTERN RECOGNITION METHODS TO GEOPHYSICAL PROBLEMS

Application of the pattern recognition methods to the problems of earthquake-prone areas determination and intermediate-term earthquake prediction is considered below.

7.1 Recognition of Earthquake-prone Areas

The problem under consideration is to determine in the region the areas where strong (with magnitude $M \geq M_0$ where M_0 is a threshold specified) earthquakes are possible. The basic assumption is that strong earthquakes associate with morphostructural nodes, specific structures that are formed around intersections of the fault zones. This gives possibility to apply the pattern recognition approach.

The nodes are considered as objects of recognition. They are identified by means of the morphostructural zonation method and described by characteristics determined on the basis of the topographical, geological, geomorphological and geophysical data. When these characteristics are measured the objects are represented by vectors, components of which are values of the characteristics.

The problem as the pattern recognition one is to divide the vectors into two classes: vectors D (Dangerous) and vectors N, which represent correspondingly the nodes where earthquakes with $M \geq M_0$ may occur and the nodes where only earthquakes with $M < M_0$ may occur. Application of the pattern recognition algorithms requires a learning set of vectors, for which we know *a priori* the class they belong to. The learning set is formed on the basis of the data on seismicity observed in the region. It consists of vectors D_0 and N_0 representing correspondingly the nodes where strong earthquakes occurred and the nodes, which are far from the known epicenters of such earthquakes.

Formulation of the Problem and the Main Stages of Its Investigation

Let M_0 is the selected threshold for earthquake magnitude. The problem of earthquake-prone areas determination is informally formulated as follows. To separate the territory of the region under consideration into two parts: the area **D** where epicenters of earthquakes with magnitude $M \geq M_0$ may be situated and the area **N** where only epicenters of earthquakes with magnitude $M < M_0$ are possible.

The first question, which arises concerning this formulation of the problem, is how to select the region and the magnitude threshold M_0. The following criteria could be recommended to do this selection:

- at least 10-20 epicenters of earthquakes with $M \geq M_0$ have to be known in the region under consideration;
- the environs of the known epicenters of earthquakes with $M \geq M_0$ cover not overwhelming part of the region territory (otherwise the problem has no sense: the whole region belongs to part **D**);
- the region has to be enough uniform to possible causes of origin of earthquakes with $M \geq M_0$.

These criteria establish some dependence between size of the region and the threshold M_0. For instance, for M_0 = 5.0-6.0 linear sizes of the region are hundreds of kilometers; for M_0 = 7.0-7.5 - thousands of kilometers; for M_0 = 8.0 - tens of thousands of kilometers.

The following examples illustrate selection of the regions and corresponding thresholds M_0: Italy, M_0 = 6.0 (Caputo et al., 1980); California, M_0 = 6.5 (Gelfand et al., 1976); the Pacific Coast of the South America, $M_0 = 7.75$ (Gvishiani and Soloviev, 1984); the whole Pacific seismic belt, $M_0 = 8.0$ (Gvishiani et al., 1978).

The existing experience shows that pattern recognition methods and algorithms may be applied to investigate earthquake-prone areas (Gelfand et al., 1972, 1973, 1974a, 1974b, 1976; Gvishiani et al., 1978, 1987; Caputo et al., 1980; Zhidkov and Kosobokov, 1980; Gvishiani and Kosobokov, 1981; Kosobokov, 1983; Gvishiani and Soloviev, 1984; Cisternas et al., 1985; Gorshkov et al., 1987; Zhidkov et al., 1975).

After selection of the region and threshold magnitude M_0 the objects of recognition should be defined in the region.

First time pattern recognition methods were applied to earthquake-prone areas studies in the Pamirs and Tien Shan by Gelfand et al. (1972). Since then several important studies for determination of earthquake-prone areas have been developed by pattern recognition methods. In principal three types of objects are used in such recognition studies: plane areas, points and their vicinities and segments of linear structures.

Morphostructural knots are an example of area objects. They are the most fractional areas characterized by the most active tectonic movements. The boundaries of the knots are determined by using geological and geomorphological criteria and data of field investigations. Necessary condition to use knots as objects of recognition is the fact that all epicenters of earthquakes with $M \geq M_0$ known in the region are located inside knots (according to the accuracy of their determination). If it is so the further study is based on the assumption that all epicenters of earthquakes with $M \geq M_0$ are located inside morphostructural knots. Geomorphological basis of this assumption is formulated by Ranzman (1979). Pattern recognition algorithms are applied in this case to classify the finite set of morphostructural knots of the region into two classes: **D**-knots, inside which earthquakes with $M \geq M_0$ may occur and **N**-knots, inside which only earthquakes with $M < M_0$ are possible. According to this classification the area **D** is formed as the part of the territory of the region covered by **D**-knots and the area **N** is the rest of the region. The examples of using morphostructural knots for earthquake-prone areas investigations can be found in Gelfand et al. (1972, 1973, 1974a).

Disadvantage of this approach to the choice of objects is the fact that determination of boundaries of morphostructural knots is a difficult problem. The necessary data are not always simply available. This is especially true for not well studied regions. In such cases it is recommended to use the intersections of axes of morphostructural lineaments as the point objects of recognition.

Selection of lineament intersections as objects of recognition is caused by the fact that blocks, lineaments and knots, which are the basic elements of morphostructural zoning are characterized by successively increasing tectonic activity (Ranzman, 1979). This selection is also based on the hypothesis about clustering of epicenters of sufficiently strong earthquakes to intersections of morphostructural lineaments (Gelfand et al., 1974b). This hypothesis seems to be highly truthful for a region if the following two conditions are valid. The distance from all well determined epicenters of earthquakes with $M \geq M_0$ known in the region to the nearest intersection does not exceed r. The area covered by circles of radius r with centers in all intersections of the region is a small part of the total area of the region.

By application of pattern recognition algorithms the finite set of lineament intersections is separated into two classes: **D**-intersections, near of which earthquakes with $M \geq M_0$ may occur, and **N**-intersections, near of which only earthquakes with $M < M_0$ are possible. According to this classification the area **D** is formed as the part of the territory of the region covered by vicinities of **D**-intersections. The area **N** is the rest of the region. As a rule a circle with a center in an intersection is understood as the vicinity of the intersection. The circle radius R is selected so that all well determined epicenters of earthquakes with $M \geq M_0$ known in the region are located within the circles.

The example of objects defined on linear structures is segments of recent active geological faults. In this case linear structures are divided into the set of segments of about equal length. By means of pattern recognition algorithms the finite set of segments is separated into two classes: **D**-segments, in the vicinities of which earthquakes with $M \geq M_0$ may occur, and **N**-segments, near of which only earthquakes with $M < M_0$ are possible. A vicinity of a segment is usually a zone with the segment as an axis line or a circle around a center of the segment. In the last case a radius usually is about half of the segment length. According to this classification the area **D** is formed as the part of the region covered by vicinities of **D**-segments and the area **N** is the rest of the region.

Segments of linear structures as objects were used for recognition of earthquake-prone areas in California (Gelfand et al., 1976), in the whole Pacific Ocean seismic belt (Gvishiani et al., 1978), and in the Western Alps (Cisternas et al., 1985). In the first case the basic linear structure was the San-Andreas fault. In the second case it consisted of the axes of the deep-water trenches or (in their absence) the bottom lines of the continental slopes situated along the Pacific Ocean belt. In the third case the segments of linear structures, forming a neotectonic scheme were considered.

Pattern recognition algorithms with learning are applied to investigate the problem. Selection of the learning subjects D_0 and N_0 depends of course on the type of the objects of recognition. In case of area objects, usually all those, inside which known earthquake epicenters with $M \geq M_0$ do exist, are included in D_0. The subset N_0 is either constituted by all other objects from W or belonging to it are objects from W, which have no known epicenters of earthquakes with $M \geq M_0 - \delta$ ($\delta > 0$) where δ is usually considered to be equal to 0.5. It is necessary to emphasize that in both cases set N_0 is not "pure" in the sense that among its objects can be those, inside which earthquakes with $M \geq M_0$ are possible. Moreover, in the first case where $N_0 = W \setminus D_0$ the problem indeed is to locate such objects. Such specific fussy type of learning material represents a specific difficulty in locating possible earthquake-prone areas by pattern recognition techniques.

If the period of seismic observations in a region under consideration is long enough then it is possible to assume that the percentage of objects to be reclassified from N_0 into class D is relatively small. Let us note that introducing $\delta > 0$ to include in N_0 set only the objects, which have no epicenters of earthquakes with $M \geq M_0 - \delta$ reduces the above percentage. It is natural to require the condition $D_0 \subseteq D$. In other words all known places of strong earthquakes should be recognized. If the number of objects in the set D_0 is large enough then it may be used not entirely as the learning set and the part of its objects may be left for examination to verify the reliability of decision rule obtained.

In the case of point objects, the set D_0 is constituted by the objects situated at a distance from the known epicenters of earthquakes with $M \geq M_0$ not exceeded a fixed threshold r. Value of r is a function of M_0: for instance, for earthquake-prone areas recognition in the Eastern part of the Central Asia ($M_0 = 6.5$) value of r has been chosen 40 km (Zhidkov and Kosobokov, 1980), for the Pacific coast of South America ($M_0 = 7.75$) r = 100 km (Gvishiani and Soloviev, 1984) etc. Value of r has to satisfy the following condition: all sufficiently well located epicenters of earthquakes with $M \geq M_0$ in the region have to be at a distance not above r from the nearest object. The set N_0 is either constituted by the remaining objects or by the objects not included in that part of the region covered with the circles of radius r_1 ($r_1 \geq r$) and the centers in the epicenters of known earthquakes with $M \geq M_0 - \delta$ ($\delta > 0$). In this case the set N_0 also can contain objects that potentially are objects of class D.

Let us emphasize the following difficulty: the same epicenter may be at a distance not exceeded r from several objects. Therefore, among these objects, which belong to the set D_0 by definition some objects may be reclassified into class N. The algorithm CLUSTERS, which takes into account this specific feature of the set D_0 is used to overcome this difficulty. In case when the necessity arises of applying CLUSTERS algorithm, instead of $D_0 \subseteq D$ condition, it is quite natural to require the following condition: each known earthquake with $M \geq M_0$ has an object of class D at a distance not exceeded r from its epicenter.

In the case of objects in the form of segments of linear structure, the set D_0 is constituted by segments, on which the epicenters of earthquakes with $M \geq M_0$ are projected. The set N_0 consists either of all the rest segments ($N_0 = W \setminus D_0$) or contains the segments from W, which have no the joint borders with the segments belonging to D_0. The other example of N_0 is the set of the segments, on which no projection of an epicenter of earthquake with $M \geq M_0 - \delta$ ($\delta > 0$) takes place. Since in the case of linear structures one epicenter, as a rule, produces only one object to be included in the set D_0, it is quite natural to require the resulting classification satisfying the condition $D_0 \subseteq D$.

For applying pattern recognition algorithms the objects defined in the region have to be described in the uniform set of geological-geomorphological and geophysical characteristics. As far as the earthquake-prone area problem is considered it is natural to use the characteristics, which describe the intensity of recent tectonic activity in the objects' vicinities. Experience obtained in recognition of earthquake-prone areas enables us to establish the most typical of such characteristics. Among them are:

- the characteristics related to the description of topography inside the area of the object and in its vicinity such as maximum altitude above see level H_{max} inside the object's area, altitude range ΔH; dominating combination of geomorphological structures in the object's vicinities, percentage of the object's area with existing Paleogene Quaternary sediments, etc.;
- the characteristics related to the schemes of neotectonic and morphostructural lineaments of the region, maps of recent faults etc. such as number of lineaments forming the object, the highest rank of lineament among those which form the object, etc.;
- the characteristics related to description of gravity field anomalies inside the area of the object.

In case of area objects the used notion of "area" is obvious. In case of point objects - the "area" is a circle of the fixed radius for all the objects and with the center in the object. In case of objects being in the form of linear structures' segments - the "area" is a circle of the fixed radius for all the segments and with the center in the middle of the segment. Area objects may under the above definition have various values of area. Thus area of an object area may be used as a characteristic for the object description.

All available information related directly or indirectly to seismicity description may be in principle used while the object characteristics are selected. The key condition for a characteristic to be introduced is the possibility to measure uniformly its values for all the objects in the region. After measuring the values of selected characteristics for all the objects the set of corresponding vectors $\mathbf{w}^i = \{w_1^i, w_2^i, ..., w_m^i,\}$, $i = 1, 2, ..., n$, can be constructed. Here m is the total number of characteristics; n - the total number of objects in the set W; w_k^i - the value of k-th characteristic measured for i-th object.

Pattern recognition algorithms, used to investigate the problem, do classify the vectors with binary components. Therefore, prior to an algorithm application, the discretization of the characteristics and the coding of objects in the form of vectors with binary components has to be carried out.

The next step of the problem investigation is applying a pattern recognition algorithm to obtain classification $W = D \cup N$ where D is the set of objects, in which known and future epicenters of earthquakes with $M \geq M_0$ may be located; N is the set of objects where only epicenters of earthquakes with $M < M_0$ may occur. As it was pointed out above the resulting classification should satisfy the condition $D_0 \subseteq D$ in cases of area objects and objects in the form of linear structures' segments. In case of point objects this condition was reformulate in the way that in the vicinity of each epicenter of an earthquake with $M \geq M_0$ an object belonging to the set D should be present. In order to avoid trivial classification the following condition for resulting classification has to be introduced:

$$|D| \leq \beta\,|W|.$$

Here as everywhere $|D|$ and $|W|$ denote the numbers of objects in the sets D and W respectively; β ($0 < \beta < 1$) is the real number, which gives a priori upper evaluation of spatial pattern of the region's seismicity as far as earthquakes with $M \geq M_0$ are concerned. The value of the parameter β has to be definite prior to the application of pattern recognition algorithms. The choice of this value should proceed from the expert evaluation of all geological, seismological and other available information concerning the region under investigation.

The quality and reliability of the classification obtained are verified by control experiments. The successful ones produce arguments in favor of the fact that the resulting classification represents well the actual division of objects into the classes related and not to possible prone-areas of earthquakes with $M \geq M_0$. The verification of the classification obtained also follows from the comprehensive analysis of the result involving additional information not applied in the problem tackling. Of course the most important in this sense are non-instrumental epicenters data. Let there are some objects, which are able to be among D_0 set but not included in learning material and only classified. Therefore their classification as objects of the set D is a convincing argument in favor of the classification reliability. Unfortunately we operate with small samples. But the volume of D_0 set offers, as a rule, no possibility for the checking of the result in such a way.

If necessary control experiments were unsuccessful, then the classification obtained cannot be taken even as a possible solution of the problem. In this case we should return to the procedure of recognition. Furthermore sometimes even the region under consideration and threshold of considered magnitudes M_0 have to be corrected. The necessity of returning to the earlier stages of the recognition procedure may arise not only due to the results of verification. It may occur at any stage of classification if the results appear to be unsatisfactory.

Summarizing the following stages of the recognition of earthquake-prone areas may be formulated.

1. Definition of the region under consideration and the threshold of magnitude of strong earthquakes in this region for investigation of earthquake-prone areas.
2. Choice of the type of objects of recognition and their definition in the region. This stage requires detailed geological, geomorphological, and seismotectonic observations. By using these data a scheme of lineaments as the basis for definition of objects has to be constructed.
3. Construction of learning material $W_0 = D_0 \cup N_0$ for "dangerous" (D_0) and "non-dangerous" (N_0) classes.
4. Selection of object description characteristics and measurement of their values.
5. Discretization and coding of the characteristics.
6. Application of a recognition algorithm to obtain classification $W = D \cup N$ into dangerous (D) and non-dangerous (N) objects.
7. Control experiments. Evaluation of the reliability of classification obtained.
8. Definition of **D** and **N** plane areas in the region on the basis of the classification $W = D \cup N$.
9. Geological and geomorphological interpretation of the obtained division of the region into dangerous and non-dangerous zones and their recognized features.

The key among the above problem tackling stages is the validation of the results and evaluation of their reliability. This stage is also the most time consuming but should be carried out very detailly and carefully.

After the definition of **D** and **N** areas on the region's territory it is advisable to do a statistical analysis of the locations of the known epicenters of earthquakes with $M < M_0$ relative to the located areas (as e.g. in Kosobokov and Soloviev, 1983). The result of such comparison can lead, in principle, to the conclusion that the obtained **D** and **N** areas are actually earthquake-prone areas for earthquakes with $M < M'_0$ where M'_0 is a smaller magnitude threshold than M_0.

Recognition of earthquake-prone areas for the Western Alps

The problem of recognition of places in the Western Alps where earthquakes with $M \geq 5.0$ may occur (Cisternas et al., 1985) is briefly considered below.

The objects are the intersections of the morphostructural lineaments obtained as the result of the morphostructural zoning of the Western Alps. The scheme of the morphostructural zoning of the Western Alps and the objects are shown in Fig. 7. The total number of objects in the set W is 62. The problem is to classify these objects into two classes: objects where earthquakes with $M \geq 5.0$ may occur (class D) and objects where earthquakes with $M \geq 5.0$ may not occur (class N).

Table 1 contains the list of characteristics, which describe the objects. The components of vectors $\mathbf{w}^i$ are the values of these characteristics.

The epicenters of earthquakes with $M \geq 5.0$ or $I \geq 7$ (I is maximum macroseismic intensity) are shown in Fig. 7 by dark circles with years of occurrence. The learning set D_0 of class D consists of 14 objects, near which instrumental epicenters of earthquakes with $M \geq 5.0$ are known (earthquakes in the 1900-1980 period): 3, 12, 13, 14, 20, 30, 31, 35, 40, 41, 42, 44, 51, 57. The objects (1, 5, 6, 8, 53, 55, 56, 58, 60, 61), which have historic earthquake epicenters (events prior to 1900) with $I \geq 7$, were not included both in D_0 and N_0 learning sets. These objects and objects 18, 19, which are located near the epicenter of 1905, were voted only. The remaining 36 objects constituted the learning set N_0 of class N.

The following characteristics (Table 1) ought to be considered as the most informative: maximum altitude H_{max}, altitude gradient $\Delta H/l$, the portion of the soft (quaternary) deposits Q, the highest rank of the lineament in the intersection R_h, distance to the nearest second rank lineament ρ_2. For all these functions $P_{max} > 20\%$.

Coding of all the functions, except the combinations of relief types (Table 1), was performed by S-method with the thresholds given in Table 1. Their values have been obtained by the method of objective discretization. Functions describing relief pattern need no additional discretization and coding since they take values 1 (yes) or 0 (no).

Value of β, which gives a priori upper evaluation of spatial pattern of the region's seismicity for earthquakes with $M \geq 5.0$, was estimated as 0.6. Therefore classifications with $|D| \leq 0.6\,|W|$ were considered only.

Algorithm CORA-3 was applied with the following values of its parameters: $k_1 = 3, \bar{k}_1 = 2, k_2 = 11, \bar{k}_2 = 1$, and $\Delta = 0$. The selected sets of characteristic traits of classes D and N (D- and N-traits) are given in Table 2. The traits are given in the table as conjunctions of inequalities in the values of the object description characteristics.

The obtained classification of the objects is shown in Fig. 7: 34 objects are attributed to class D, and 28 objects are attributed to class N. All the objects of the learning set D_0 are classified as objects of class D. The number of objects of N_0, classified as objects of class D, is roughly 30% of the their total number in N_0.

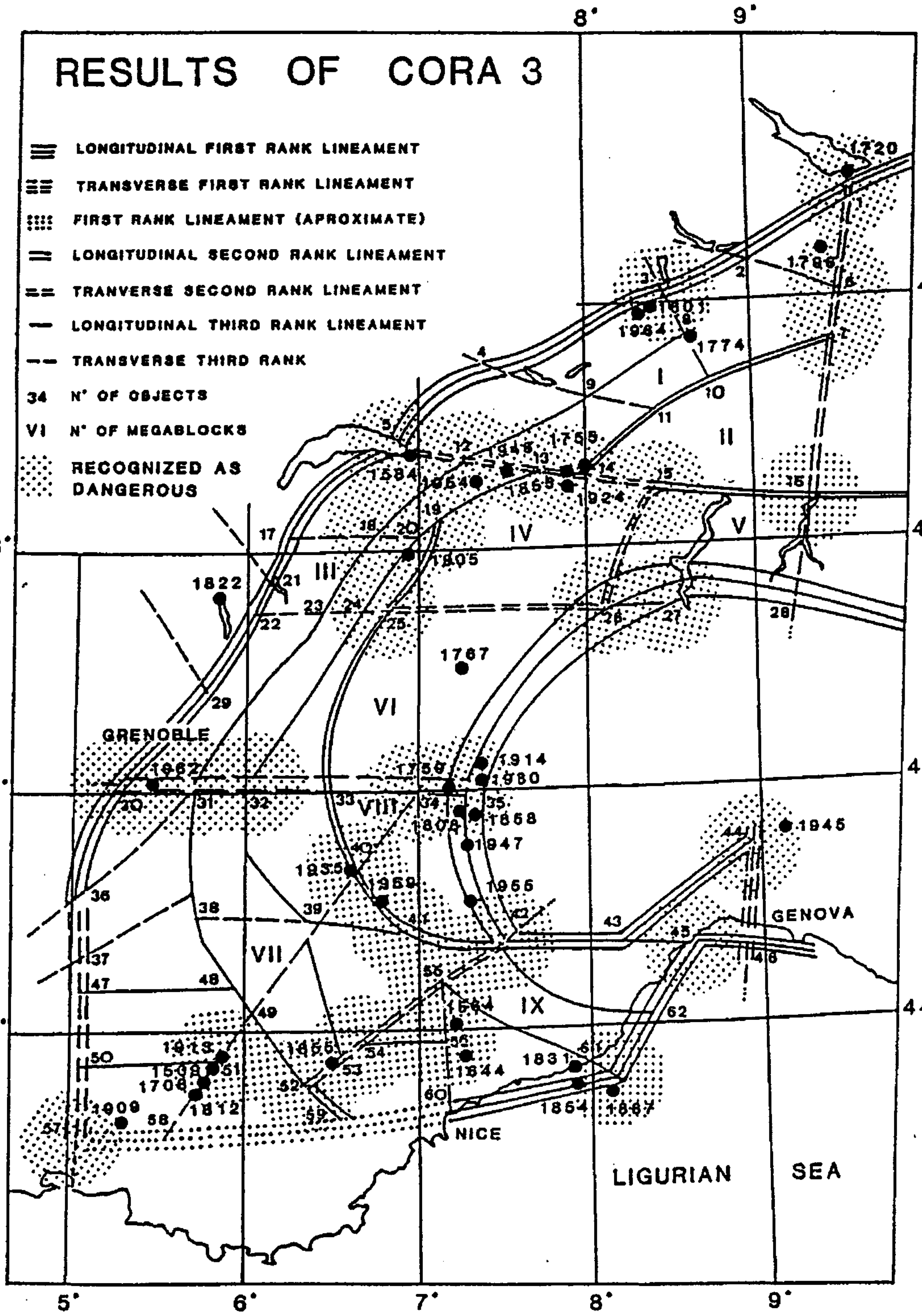

FIGURE 7 The morphostructural scheme of the Western Alps and the result of recognition.

TABLE 1 Characteristics of objects of the Western Alps

Functions	Discretization thresholds	
	first	second
Maximum altitude H_{max}, *m*	2686	4807
Minimum altitude H_{min}, *m*	325	-
Altitude in the lineament intersection point H_0, *m*	490	900
Distance between points where H_{max} and H_{min} are measured *l*, *km*	32	42
$\Delta H = H_{max} - H_{min}$, *m*	2500	-
Altitude gradient $\Delta H/l$, *m/km*	51	91
Combinations of relief types (yes, no)		
mountain slope/mountain slope (m/m)		
mountain slope/plain (m/p)		
mountain slope/piedmont/plain (m/pd/p)		
mountain slope/piedmont (m/pd)		
piedmont/plain (pd/p)		
The portion of the soft (quaternary) deposits *Q*, %	10	-
The highest rank of the lineament in the intersection R_h	1	2
Number of lineaments forming the intersection n_l	2	-
Number of lineaments in the circle of radius 25 *km* N_l (3 thresholds)	2	3, 4
Distance to the nearest intersection ρ_{int}, *km*	20	31
Distance to the nearest first rank lineament ρ_1, *km*	0	32
Distance to the nearest second rank lineament ρ_2, *km*	0	40
Maximum value of Bouguer anomaly B_{max}, *mGal*	-82	-8
Minimum value of Bouguer anomaly B_{min}, *mGal*	-145	-85
$\Delta B = B_{max} - B_{min}$, *mGal*	45	65
$\overline{B} = (B_{max} + B_{min})/2$, *mGal*	-110	-44
$HB = 0.1\ H_{max}\ [m] + B_{min}\ [mGal]$	153	-
Number of Bouguer anomaly isolines N_B	4	7
Number of closed Bouguer anomaly isolines N_{BC}	1	-
Minimum distance between two Bouguer anomaly isolines with values divided by 10 *mGal* $(\nabla B)^{-1}$, km	2	3

TABLE 2 Characteristic traits selected by algorithm CORA-3 for recognition of objects of the Western Alps

#	Q, %	n_l	N_l	ρ_1, *km*	ρ_2, *km*	ΔB, *mgl*	$(\nabla B)^{-1}$, km
	D-traits						
1				≤32		≤65	≤2
2				>0		≤65	≤2
3				≤32	0	≤65	
4			>3		0	≤65	
5			>4			>45	≤3
6					>0; ≤40	>45	
7		2		>32		>45	
8		2		>32			≤3
9		>2	≤3				≤2
10	>10		>3		≤40		
	N-traits						
1						≤45	>2
2					>0	≤45	
3		2				≤45	
4					>40	≤45	
5					>40		>2
6		2			>40		
7		2	≤3		>0		
8		2		0			

7.2 Intermediate-term Prediction of Earthquakes

The pattern recognition methods were used to develop the intermediate-term earthquake prediction algorithm CN (Keilis-Borok and Rotwain, 1990). This algorithm was initially applied to California-Nevada region and is called algorithm CN.

Objects of Recognition

The objects are moments of the time. These moments are described by the functions defined in the lecture "Functions on Catalogs ..." (Kossobokov and Novikova, 2001). The selection of the moments and the forming of the learning sets D_0 and N_0 are described below.

If the earthquake catalog of some region covers the time from t_0 to T_k the three types of time periods can be defined between t_0 and T_k:

- periods, which precede strong earthquakes, - periods *D*;
- periods, which follow strong earthquakes, - periods *X*;
- periods, which are not connected with strong earthquakes, - periods *N*.

The formal definition can be formulated as follows.

Let $t_1, t_2, \ldots, t_m$ ($t_0 < t_1 < t_2 < \ldots < t_m < T_k$) be the moments of strong earthquakes of the region under consideration. Here strong earthquakes are the main shocks with magnitude $M \geq M_0$, where M_0 is a given threshold.

Periods *D* are time intervals from $t_i - \Delta t_D$ to t_i ($i = 1, 2, \ldots, m$).

Periods *X* are time intervals from t_i to $t_i + \Delta t_X$ out of periods *D*.

Periods N are intervals from t_0 to T_k which remain after exclusion of all periods D and X.

Here $i = 1, 2, ..., m$; Δt_D and Δt_X are given constants.

Example of periods D, X, and N is shown in Fig. 8. The moments t_i, t_{i+1}, t_{i+2}, and t_{i+3} in this figure are the moments of four strong earthquakes.

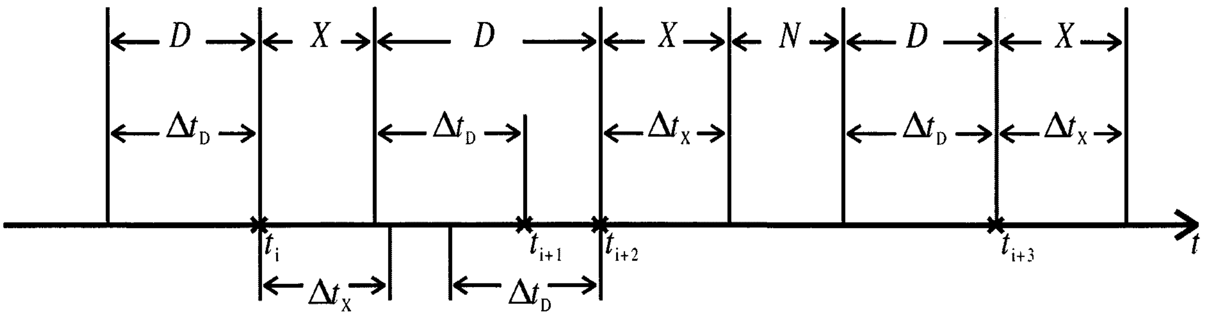

FIGURE 8 Periods D, N, and X.

Moments of time are considered as objects of recognition. For time period from t_0 to T_k three types of moments are defined: D_0, N_0, and X.

Moments D_0 (the set D_0) are the moments before strong earthquakes. For each strong earthquake with origin time t_i the interval from $t_i - \Delta t_D$ to $t_i - \delta t$ is divided into k equal parts of the length $\Delta t_2 = \Delta t_1/k$ where $\Delta t_1 = \Delta t_D - \delta t$. Here $\delta t \geq 0$ and k are specified so to satisfy the relationship $\delta t << \Delta t_2$.

Moments D_0 are the moments

$$t_i^j = t_i - \Delta t_D + j\Delta t_2$$

where $j = 0, 1, 2, ..., k$. The moments D_0, which are earlier than the origin time t_{i-1} of the preceding strong earthquake, are eliminated (see Fig. 9b).

Moments N are selected within periods N with equal steps, unless there is not specific reason to do otherwise.

Moments N_0 (the set N_0) are selected from moments N to be regularly distributed among them. The number of moments N_0 is usually selected about the same as the number of moments D_0.

Moments X are selected from periods X with step Δt_2.

Subclasses

Among the moments D_0 subclasses are formed. One subclass includes moments D_0, which precede the same strong earthquake.

Let t_{i-1} and t_i are origin times of two consecutive strong earthquakes. If $t_i - t_{i-1} > \Delta t_D$ then the subclass connected with the strong earthquake numbered i consists of the following moments D_0:

$$t_i^j = t_i - \Delta t_D + j\Delta t_2$$

where $j = 0, 1, 2, ..., k$. If $t_i - t_{i-1} \leq \Delta t_D$ then only moments t_i^j, which are after t_{i-1} ($t_i^j > t_{i-1}$), are included in the subclass.

In Fig. 9a the subclass connected with the strong earthquake occurred at time t_{i-1} consists of three moments D_0: t_{i-1}^0, t_{i-1}^1, and t_{i-1}^2. The subclass, connected with the strong earthquake, occurred at time t_i, consists also of three moments D_0: t_i^0, t_i^1, and t_i^2.

In Fig. 9b the subclass, connected with the strong earthquake, occurred at time t_{i-1}, consists also of three moments D_0: t_{i-1}^0, t_{i-1}^1, and t_{i-1}^2, and the subclass, connected with the strong earthquake, occurred at time t_i, consists only of two moments D_0: t_i^1 and t_i^2.

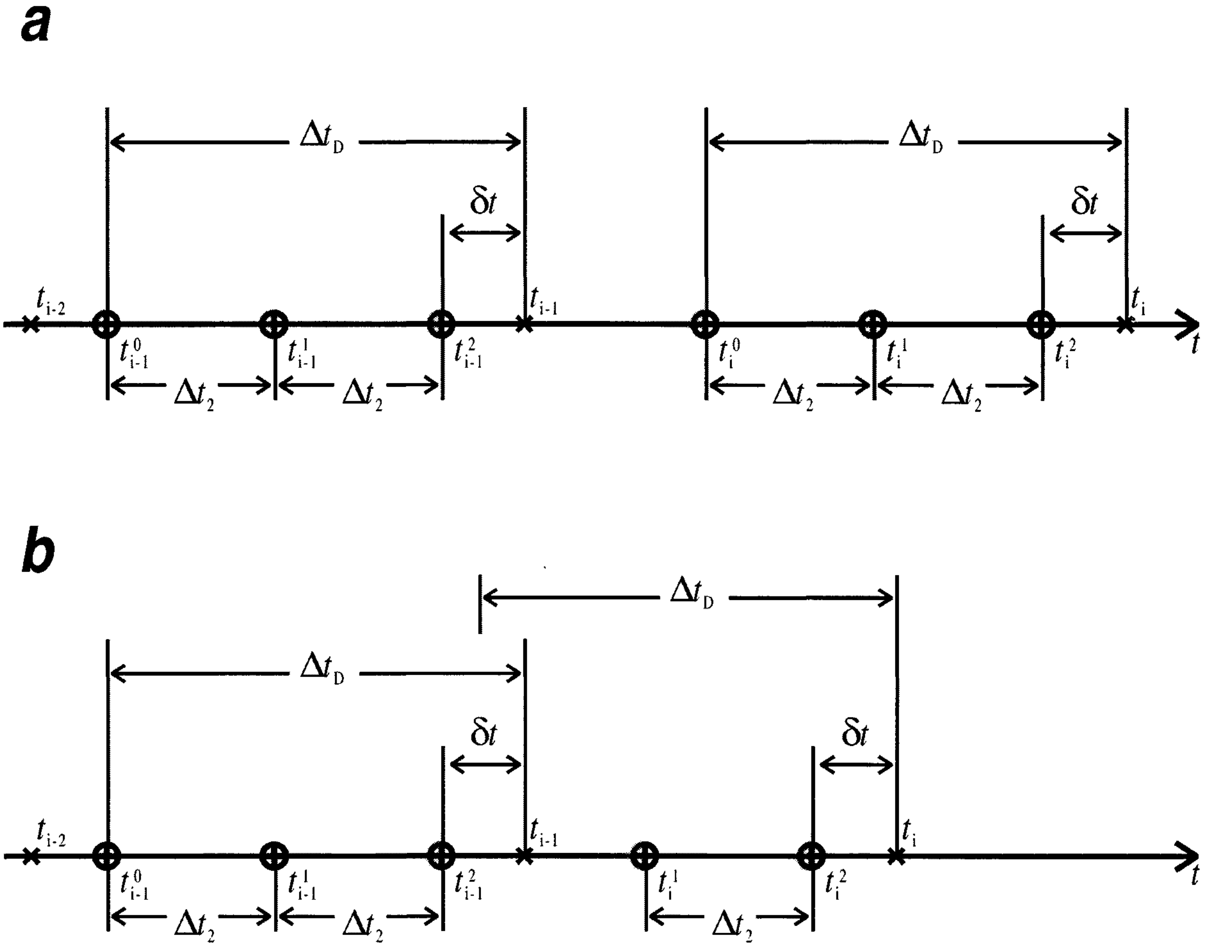

FIGURE 9 Moments D_0 (marked by $\oplus$), $k = 2$.

<u>*Algorithm CN*</u>

The earthquake catalog of the Southern California for the time period 1938-1984 was used to determine the learning set. The threshold magnitude for the strong earthquakes was $M_0 = 6.4$. Table 3 contains the thresholds for discretization of the functions on the earthquake flow, calculated for these moments. The coding was performed by *S*-method with these thresholds.

The algorithm CLUSTERS was applied to obtain the characteristic traits of classes *D* and *N*. These traits are listed in Table 4. The parameters had the following values: $k_1 = 7, \bar{k}_1 = 2, k_2 = 10, \bar{k}_2 = 4$. The moments defined for the Southern California are classified by using these traits and $\Delta = 5$. If a moment t is attributed to class *D* then this moment is considered to belong to a period of the time of increased probability (TIP) of a strong earthquake. Formally if t is attributed to class *D* then a TIP is diagnosed from t to $t + \tau$ where τ is a given constant. For the Southern California $\tau = 1$ year was used.

TABLE 3 Thresholds for discretization of functions on the earthquake flow (Southern California)

Function	Thresholds	
N2	0	-
K	-1	1
G	0.5	0.67
SIGMA	36	71
Smax	7.9	14.2
Zmax	4.1	4.6
N3	3	5
q	0	12
Bmax	12	24

TABLE 4 Characteristic traits of classes *D* and *N* obtained by algorithm CLUSTERS for the moments of the Southern California
(traits of the algorithm CN)

Traits *D*	N2	K		G		SIGMA		Smax		Zmax		N3		q		Bmax	
1		0															0
2															0		
3												0		0		0	
4											0			0			
5		0										1				0	
6			1								0					0	
7		0									1					0	
8		0		0												0	
9									0		0						
10			1				0				0						
11		0	1								0						
12	0		1							0							
13			0						1								
14			0					0									
15			0			0											
16			0		1												

Traits *N*	N2	K		G		SIGMA		Smax		Zmax		N3		q		Bmax	
1									1							1	
2											1						1
3							1								1		1
4			1												1		1
5												0			1		1
6								1									1
7			1								1						1
8	1										1						1
9							1					0			1		
10								1							1		
11											1		0				
12								1				0					
13			1					1									
14			1			1											
15		1									1						
16			1		1				1								
17		1							1								
18		1					1										

REFERENCES

Caputo,M., V.Keilis-Borok, E.Oficerova, E.Ranzman, I.Rotwain, and A.Solovjeff (1980). Pattern recognition of earthquake-prone areas in Italy. *Phys. Earth Planet Int.*, 21: 305-320.

Cisternas,A., P.Godefroy, A.Gvishiani, A.I.Gorshkov, V.Kosobokov, M.Lambert, E.Ranzman, J.Sallantin, H.Soldano, A.Soloviev, and C.Weber (1985). A dual approach to recognition of earthquake prone areas in the western Alps. *Annales Geophysicae*, **3**, 2: 249-270.

Gelfand,I.M., Sh.Guberman, M.L.Izvekova, V.I.Keilis-Borok, and E.Ia.Ranzman (1972). Criteria of high seismicity determined by pattern recognition. In A.R.Ritsema (ed.), *The Upper Mantle. Tectonophysics*, **13** (1-4): 415-422.

Gelfand,I.M., Sh.I.Guberman, M.P.Zhidkov, M.S.Kaletskaya, V.I.Keilis-Borok, and E.Ja.Ranzman (1973). Transfer of high seismicity criteria from the East of Central Asia to Anatolia and adjacent regions. *Doklady Academii Nauk SSSR*, **210**, 2: 327-330 (in Russian).

Gelfand,I.M., Sh.I.Guberman, M.P.Zhidkov, M.S.Kaletskaya, V.I.Keilis-Borok, E.Ja.Ranzman, and I.M.Rotwain (1974a). Identification of sites of possible strong earthquake occurrence. II. Four regions of Asia Minor and South Eastern Europe. In V.I.Keilis-Borok (ed.), *Computer Analysis of Digital Seismic Data.* Moscow, Nauka: 3-40 (Comput. Seismol.; Iss. 7, in Russian).

Gelfand,I.M., Sh.I.Guberman, M.P.Zhidkov, V.I.Keilis-Borok, E.Ja.Ranzman, and I.M.Rotwain (1974b). Identification of sites of possible strong earthquake occurrence. III. The case when the boundaries of disjunctive knots are unknown. In V.I.Keilis-Borok (ed.), *Computer Analysis of Digital Seismic Data.* Moscow, Nauka: 41-65 (Comput. Seismol.; Iss. 7, in Russian).

Gelfand,I.M., Sh.A.Guberman, V.I.Keilis-Borok, L.Knopoff, F.Press, I.Ya.Ranzman, I.M.Rotwain, and A.M.Sadovsky (1976). Pattern recognition applied to earthquake epicenters in California. *Phys. Earth Planet. Inter.*, **11**: 227-283.

Gorshkov,A.I. G.A.Niauri, E.Ya.Ranzman and A.M.Sadovsky (1987). Use of gravimetric data for recognition of places of possible occurance of strong earthquakes in the Great Caucasus. In V. I. Keilis-Borok and A. L. Levshin (eds), *Theory and Analysis of Seismological Information*, Allerton Press Inc, New York: 117-123 (Comput. Seismol., volume 18).

Gvishiani,A.D., A.V.Zelevinsky, V.I.Keilis-Borok, and V.G.Kosobokov (1978). Study of the violent earthquake occurrences in the Pacific Ocean Belt with the help of recognition algorithms. *Izvestiya Acad. Sci. USSR. Physics of the Earth*, 9: 31-42 (in Russian).

Gvishiani,A.D. and V.G.Kosobokov (1981). On foundations of the pattern recognition results applied to earthquake-prone areas. *Izvestiya Acad. Sci. USSR. Physics of the Earth*, 2: 21-36 (in Russian).

Gvishiani,A.D., and A.A.Soloviev (1984). Recognition of places on the Pacific coast of the South America where strong earthquakes may occur. *Earthq. Predict. Res.*, 2: 237-243.

Gvishiani,A., A.Gorshkov, V.Kosobokov, A.Cisternas, H.Philip, and C.Weber (1987). Identification of seismically dangerous zones in the Pyrenees. *Annales Geophysicae*, 5B(6): 681-690.

Keilis-Borok,V.I. and A.J.Lichtman (1981). Pattern recognition applied to presidential elections in the United States, 1860-1980: role of integral social, economic and political traits. *Proceedings of US National Ac. Sci.*, 78, 11: 7230-7234.

Keilis-Borok,V.I. and I.M.Rotwain (1990). Diagnosis of Time of Increased Probability of strong earthquakes in different regions of the world: algorithm CN. *Phys. Earth Planet. Inter.*, **61**: 57-72.

Kosobokov,V.G. (1983). Recognition of the sites of strong earthquakes in East Central Asia and Anatolia by Hamming's method. In V.I.Keilis-Borok and A.L.Levshin (eds), *Mathematical Models of the Structure of the Earth and the Earthquake Prediction*, Allerton Press Inc, New York: 78-82 (Comput. Seismol., volume 14).

Kosobokov,V.G., and A.A.Soloviev (1983). Disposition of epicenters of earthquakes with $M \geq 5.5$ relative to the intersection of morphostructural lineaments in the East Central Asia. In V.I.Keilis-Borok and A.L.Levshin (eds), *Mathematical Models of the Structure of the Earth and the Earthquake Prediction*, Allerton Press Inc, New York: 75-77 (Comput. Seismol., volume 14).

Kosobokov,V.G., and O.Novikova (2001). Functions on Catalogs: Measures of Activity, Clustering, etc. Sixth Workshop on Non-Linear Dynamics and Earthquake Prediction, 15-27 October 2001, Trieste: the Abdus Salam ICTP.

Ranzman,E.Ia. (1979). *Places of Earthquakes and Morphostructures of Mountain Countries*. Nauka, Moscow (in Russian).

Zhidkov,M.P., I.M.Rotwain, A.M.Sadovsky (1975). Recognition of places where strong earthquakes may occur. IV. High-seismic intersections of lineaments of the Armenian upland, the Balkans and the Aegean Sea basin. In V.I.Keilis-Borok (ed.), *Interpretation of Seismology and Neotectonics Data*, Moscow, Nauka: 53-70 (Comput. Seismol.; Iss. 8, in Russian).

Zhidkov,M.P., and V.G.Kosobokov (1980). Identification of the sites of possible strong earthquakes - VIII: Intersections of lineaments in the east of Central Asia. In V.I.Keilis-Borok (ed.), *Earthquake Prediction and the Structure of the Earth*, Allerton Press Inc, New York: 31-44 (Comput. Seismol., volume 11).